Molecular investigation of cellular immunity in marsupials

Sabine Flenady
B. Biomed. Sc., B. Biomed. Sc. Hon. (1st Class)

A thesis submitted for the degree of Doctor of Philosophy

in accordance with the guidelines of CQUniversity Australia

School of Science, Engineering and Health
CQUniversity Australia

August 2012

Certificate of Authorship and Copyright statement

Certificate of Authorship of Thesis

The work contained in this thesis has not been previously submitted either in whole or in part for a degree at CQUniversity or any other tertiary institution. To the best of my knowledge and belief, the material presented in this thesis is original except where due reference is made in text.

Signature of Candidate. ……

Date. ……

Copyright Statement

This thesis may be freely copied and distributed for private use and study, however, no apart of this thesis or the information contained therein may be included in or referred to in any publication without prior written permission of the author and/or any reference fully acknowledged.

Signature of Candidate. ……..

Date. ……..

Abstract

The T cell signalling cascade is an important biochemical event in the adaptive immune system. In this study, various molecules involved in the T cell biology of marsupials were characterized. Investigation of possible differences in the molecular sequence of these genes was undertaken to explain why marsupials appear to respond differently to other mammals in their reaction to challenge by pathogens. Prior to this study it has been generally accepted that the adaptive immune system in marsupials may not function in the same manner as in other mammals.

Reverse transcription polymerase chain reactions, rapid amplification of cDNA ends, and quantitative polymerase chain reactions were employed in this study to obtain sequences for key molecules of the marsupial T cell signalling cascade. All putative amino acid sequences were investigated for structurally important motifs using various bioinformatics tools. Tertiary structure predictions were carried out using the open reading frame for the key molecules in the marsupial T cell signalling cascade. For the ZAP-70 and Lck molecules, commercially available human antibodies were trialled using a Western Blot method in a crude cell lysate prepared from *Onychogalea fraenata* spleen, liver and gut node and other tissues. This was undertaken to investigate possible species cross reactivity. A species-specific antibody for IL-2 was designed and successfully trialled in the *Macropus eugenii* tissue lysate prepared from thymus. This contributes an important new immunological tool for marsupial research.

This work demonstrated conclusively that the marsupial mRNA derived sequences of various T cell molecules were similar to their eutherian counterparts. It was shown that in marsupials the putative amino acid sequences were comparable to those of other mammals and many of the structurally important motifs were conserved. An important exception was the marsupial TCRα molecule which appeared to be significantly different in the putative amino acid sequence when compared to that of other mammals. It was also found that the marsupial IL-2 protein appeared to be functional since all structural motifs important for biological activity were identified and were at least partially conserved indicating a potential similarity. What this means in terms of marsupial immune competency is not yet completely understood. As a result of this study it is now hypothesized that the marsupial immunological adaptive immune system is as sophisticated as that of other mammals.

Table of Contents

Table of Contents

Table of contents

Table of contents

Table of contents

Table of contents

Table of contents

Table of contents

Table of contents

List of Tables

Chapter 2 – Materials and Methods

Chapter 3 – The diprotodontic T cell signalling unit and the corresponding receptor

List of Tables

List of Tables

List of Tables

Chapter 6 – The diprotodontic cytokines Interleukin-2 (IL-2), Interleukin-17 (IL-17) and the regulatory T cell surface marker Foxp3

List of Figures

Chapter 1 – Introduction and Literature Review

Chapter 2 – Materials and Methods

Chapter 3 - The diprotodontic Tcell signalling unit and the corresponding receptor

Chapter 4 – The diprotodontic co-receptor and co-stimulator to the T cell receptor

List of Figures

Chapter 5 - The diprotodontic signalling molecules – TCRζ, ZAP-70 and Lck

Chapter 6 - The diprotodontic cytokines Interleukin-2 (IL-2), Interleukin-17 (IL-17) and the regulatory T cell surface marker Foxp3

Chapter 7 - Research summary, concluding discussion and future work

Acknowledgements

I would like to thank Professor Jennelle Kyd for her invaluable advice and her courage to take me on in a very late stage of my candidature. Without her, this thesis would have not been written. Her tireless commitment and endless guidance and support will never be forgotten. I also wish to thank the members of Professor Kyd's team at the Capricornia Centre for Mucosal Immunology, especially Dr. Ajay Krishnamurthy for his encouraging words and Jessica Browne for lighting up my day when the writing became hard going.

I cannot forget my long-standing friend and colleague Dr. Noel Sammon with whom I went through my undergraduate studies at CQUniversity Australia. He provided countless hours of his time to read through thesis drafts and to coach me in the English language. When things were difficult he would often be the only person that understood. I am grateful for his time and advice throughout our studies together especially in my PhD.

This has been a long journey which I would have not been able to undertake without the support of my husband Barry and my son Daniel. My husband has been very supportive for the seven years it took for me to get through my undergraduate studies including honours and finally my PhD. My son Daniel has known no other mother but the one that always buried her head in a text book or in scientific papers but I hope he learned that perseverance pays off and not to throw the towel in the ring too early.

I must not forget my friends Alana and Robert Evans. A thank you to Robert who took time out to read through the early drafts of the thesis and gave me directions as to how to apply the English language more effectively and taught me how to organize my work more efficiently. Alana, what can I say! Your humor and personality made me laugh in my darkest hours.

A very grateful THANK YOU to all the above, you played a very important part in the completion of this thesis.

Statement of contribution by others

Chapter 3 to 6 inclusive

Sequencing reactions were sent to the Australian Genome Research Facility, Brisbane, Queensland for processing on a paid consultancy basis.

The molecules characterized in *M. domestica* were obtained from a spleen and thymus cDNA library provided by Associate Professor Dr. Katherine Belov, Sydney University, NSW.

Cell Pellet and RNA from stimulated *M. eugenii* lymphocytes and RNA from ConA stimulated *M. eugenii* lymphocytes as well as a spleen and thymus were donated by Dr. L. J. Young, CQUniversity Australia, Rockhampton, Queensland.

Spleen, mesenteric lymph nodes and lung tissue from *O. fraenata* were donated by Dr. L. J. Young, CQUniversity Australia, Rockhampton, Queensland.

The Alma Street Veterinary Hospital, Rockhampton performed a necropsy on a euthanized, diseased *O. fraenata* and and donated the tissues to the Young laboratory at CQUniversity Australia, Rockhampton, Queensland.

Chapters 3 and 4

RNA from *L. hirsutus* was isolated and synthesized into cDNA by Krista Howard, Honours student at CQUniversity Australia, Rockhampton, Queensland.

Chapter 5

Primer ST3F1 was donated by Amy Suthers, PhD student at CQUniversity Australia, Rockhampton, Queensland.

List of publications and Conference presentations

Published Papers.

Young, L. J., Cross, M. L., Duckworth, J. A., **Flenady, S**. & Belov, K. 2011. Molecular identification of interleukin-2 in the lymphoid tissues of the common brushtail possum, *Trichosurus vulpecula*. *Dev & Comp Immunol,* 2011.

Conference Presentations

2010

Flenady, S. - Oral presentation. T cell signalling in endangered marsupials. Expression of TCRζ, Lck and ZAP-70 in the endangered bridled nailtail wallaby. Presented at the Australian Health and Medical Congress (AHMRC), 14 – 18th November, 2010, Melbourne, Australia.

2009

Flenady, S. - Poster. Functional Investigation of Marsupial T cell Responses. Presented at the Institute for Resource Industries and Sustainabilities, (IRIS), 3rd December, 2009, Rockhampton, Australia.

2008

Flenady, S. - Poster. The T Cell Receptor Repertoire in the endangered marsupial *Onychogalea fraenata* (Bridled Nailtail Wallaby). Presented at ComBio 21 – 25th September, 2008, Canberra, Australia.

2007

Flenady, S. - Poster. Characterization of immune receptor molecules in the endangered marsupial *Onychogalea fraenata* (bridled nailtail wallaby). Presented at ComBio 22 – 26th September, 2007, Sydney, Australia.

Abbreviations

°C	degree Celsius
µ	micro
3’	3 prime end of cDNA sequence
5’	5 prime end of cDNA sequence
aa	amino acid
AMVRT	Avian Myoblastosis virus reverse transcriptase
AP-1	adaptor protein-1
APC	antigen presenting cell
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BCA	Bicinconinic acid
BLAST	basic local alignment search tool
BLAT	(**B**asic Local Alignment Search Tool)-**L**ike **A**lignment **T**ool
BLSOSUM	Blocks of Amino Acid Substitution Matrix
bp	base pair
BSA	bovine serum albumin
BTLA	B- and T cell attenuator
Ca	calcium
CD	Cluster of differentiation
CD28RE	CD28 response element
CDART	Conserved Domain Architecture Retrieval Tool
cDNA	complementary DNA
CDR	complementarity determining region

Abbreviations

CIP	calf intestine phosphatase
CLUSTAL**W**	weighted
ConA	concanavalinA
CT	cytoplasmic tail
CTLA-4	cytotoxic T lymphocyte antigen-4
DAB	3,3'-diaminobenzidine
DAG	diacylglycerol
DEPC	Diethylene Pyrocarbonate
DFTD	Devil Facial Tumor Disease
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	Dithiothreitol
e-value	expectation value
EC	extracellular
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic Reticulum
ERK	Extracellular Signal-Regulated Kinase
EXPASY	Expert Protein Analysis System
Foxp3	Forkhead box protein 3
Fyn (SLK)	Serine/Threonine like kinase
GADS	Grb2 related adapter protein
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
gDNA	genomic DNA
GITR	glucocorticoid-induced TNFR (tumor necrosis factor receptor)

Grb2	growth factor receptor-bound protein2
h	hour
Havana	Human and Vertebrate Analysis and Annotation project
HRM	high resolution melt analysis
ICOS	Inducible T-cell Co-stimulator
IFN	interferon gamma
IG	immunoglobulin
IG-V	immunoglobulin-variable
IL-2	Interleukin-2
IP_3	Inositol triphosphate
ITAM	immune tyrosine activation motif
ITK	inducible tyrosine kinase
IUCN	International Union for Conservation of Nature
kDa	kilo Dalton
L	litre
LAT	linker of activated T cells
LB	Luria Bertani
Lck	Lymphocyte associated kinase
LT	lymphotoxin
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight mass spectrometry
MAPK	Mitogen activated protein kinase
mg	milligram
MHC	Major Histocompatibility complex
min	minute
mL	millilitre

MLR	mixed lymphocyte response
$MnCl_2$	Manganese chloride
mpIL-2	marsupial IL-2 antibody
mRNA	messenger RNA
Nck	non-catalytic region of tyrosine kinase adaptor protein
NFAT	nuclear factor of activated T cells
NFκB	nuclear factor kappa-light-chain-enhancer of activated B cells
ng	nanogram
NMR	nuclear magnetic resonance spectrometry
Oct-1	octamer transcription factor-1
PBMC	Peripheral blood mononuclear cell
PBS	phoshate buffered saline
PCR	polymerase chain reaction
pdb	Protein Data Bank
PHA	phytohaemagglutinin
pI	isoelectric point
PI-3k	phosphoinositol-3 kinase
Phyre	Protein Homology/analogy Recognition Engine
PKC	Protein Kinase C
PLCγ1	Phospholipase C gamma 1
pMHC	peptide Major Histocompatibility complex
PSIpred	Protein structure prediction server
PTK	Protein Tyrosine Kinase
qPCR	quantitative polymerase chain reaction
RACE	Rapid amplification of cDNA ends

Ras	rat sarcoma
RIPA	radio immunoprecipitation assay
RNA	Ribonucleic acid
rpm	revolution per minute
RT-PCR	Reverse Transcription Polymerase chain reaction
s	second
SLP-76	SH_2 containing lymphocyte protein of 76kDa
SNP	Single nucleotide polymorphism
S.O.C.	super optimal catabolite repression medium
SA-PMPS	Streptavidin MagneSphere Paramagnetic Particles
SCOP	Structural Classification of Proteins
SDS	Sodium dodecyl sulfate
SHP	SH_2 domain-bearing protein tyrosine phosphatase
Syk	Spleen tyrosine kinase
TBE	Tris-borate EDTA
TBS	Tris buffered saline
TBS-T	Tris buffered saline and Tween
TCR	T cell receptor
TdT	terminal deoxynucleotydil transferase
Tec	tyrosine kinase expressed in hepatocellular carcinoma
TEMED	Tetramethylethylenediamine
T_h	T helper
TGF	Transforming growth factor
T_m	melt Temperature
TM	Transmembrane

Abbreviations

TNF	Tumor necrosis factor
TNFR	Tumor necrosis factor receptor
T_{reg}	regulatory T cell
UCSC	University of California, Santa Cruz
UT	untranslated
UTF	untranslated forward
UTR	untranslated reverse
V	volt
VAV	guanidine nucleotide exchange factor
VH/VL	variable heavy/variable light
ZAP-70	zeta chain associated protein of 70 kDa
α	alpha
β	beta
γ	gamma
δ	delta
ε	epsilon
ζ	zeta
κ	kappa

Amino Acid description

Amino Acid	3-letter abbreviation	1-letter abbreviation	Polarity of side chain	Charge of side chain
Alanine	Ala	A	nonpolar	neutral
Arginine	Arg	R	polar	positive
Asparagine	Asn	N	polar	neutral
Aspartic acid	Asp	D	polar	negative
Cysteine	Cys	C	polar	neutral
Glutamic acid	Glu	E	polar	negative
Glutamine	Gln	Q	polar	neutral
Glycine	Gly	G	nonpolar	neutral
Histidine	HIs	H	polar	neutral
Isoleucine	Ile	I	nonpolar	neutral
Leucine	Leu	L	nonpolar	neutral
Lysine	Lys	K	polar	positive
Methionine	Met	M	nonpolar	neutral
Phenylalanine	Phe	F	nonpolar	neutral
Proline	Pro	P	nonpolar	neutral
Serine	Ser	S	polar	neutral
Threonine	Thr	T	polar	neutral
Tryptophan	Trp	W	nonpolar	neutral
Tyrosine	Tyr	Y	polar	neutral
Valine	Val	V	nonpolar	neutral

(Source: Lehninger Principles of Biochemistry, David L. Nelson, 2008)

The genetic code

1st letter	Second Letter: U	Second Letter: C	Second Letter: A	Second Letter: G	3rd letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

(Source: taken from (http.//www.google.com.au/imgres?imgurl=http.//biology.kenyon. edu/)

Chapter 1 – Introduction and Literature Review

1.0 Introduction and Literature Review

Pathogens continually evolve and adapt to avoid detection by the host's defence mechanisms and therefore two major systems, the innate and adaptive immune systems, have evolved to combat these pathogens in all mammals. A function of the adaptive immune system is the recognition of antigens by the T cell receptor (TCR) and major histocompatibility complex (MHC) molecules. This process leads to the activation of the T cell signalling cascade. In humans and other mammals many of the molecules involved in this cascade have been characterized and their functions elucidated (Davis, 2002, Gong *et al.,* 2001). However, up until the turn of the last century marsupials were thought to be immunologically immature (Belov *et al.,* 2002). The difference between metatherians and eutherians is the degree of development at birth. Marsupials have a shorter gestation period and their neonates have immature lymphoid organs with pouch young being unable to elicit an immune response for some time after birth (Baker and Miller, 2007, Belov *et al.,* 2002, Ashman *et al.,* 1975). While some of the molecules involved in the T cell signalling cascade have since been characterized in marsupials, some of the receptors, signalling molecules and cytokines tied to that cell signalling cascade were still unknown. Studies conducted in the second half of the last century reported differences in the T cell responses between marsupials and eutherians (Baker *et al.,* 1999, Waring *et al.,* 1978, Ashman *et al.,* 1977, Ashman *et al.,* 1976, Ashman *et al.,* 1975). It was further discovered that marsupials were wildlife reservoirs for diseases such as Ross River fever caused by the Ross River Virus (RRV), and infections caused by *Mycobacteria bovis* and *M. avium* (Old and Deane, 2005, Boyd *et al.,* 2001, Buddle and Young, 2000).

Because some marsupial species are endangered, and frequently succumb to these diseases in captivity, the development of immunological reagents that would identify asymptomatic animals is highly desireable. In order to do so it is first necessary to understand the marsupial immune system. However prior to this study little work has been done to characterize the immune molecules important in the disease processes. Marsupials represent an important branch of mammalian evolution. Based on genetic

data, the marsupial divergence from the most recent common ancestor can be placed at 173 – 193 million years ago (Woodburne *et al.,* 2003, Tyndale-Biscoe, 1973). This makes them ideal subjects for the study of evolutionary links in the development of the adaptive immune system and to identify changes in that system over time. For these reasons this study was undertaken to characterize the key molecules of the adaptive immune system in marsupials. The marsupials *Monodelphis domestica*, *Macropus eugenii, Onychogalea fraenata, Trichosrurus vulpecula* and *Lagrochestes hirsutus* were chosen as appropriate species for this study.

Monodelphis domestica (South American gray short tailed opossum), a didelphid, was the first marsupial to have its genome sequenced (Mikkelsen *et al*., 2007). Because so few characterizations of marsupial molecules have been conducted, the *M. domestica* genome published at UCSC (http://genome.ucsc.edu/) and the predicted sequences for many molecules deposited in Genbank (http.//www.ncbi.nlm.nih.gov/genbank/) were used in the present study as positive controls. *M. domestica* lacks a pouch thus exposing the neonates which are easily accessed for studies into the development of the immune system (Wang *et al.,* 2003). This marsupial species is unique in that it can repair a damaged spinal cord and can develop melanomas by exposure to UV light which makes it an invaluable research animal (Mladinic *et al*., 2010, Chan *et al*., 2002). This animal has also been used for comparative genetics studies (Samollow, 2006). For those reasons *M. domestica* was included in the present study as a model species for the investigation of immune molecules in Australian marsupials.

The Australian marsupial *Macropus eugenii* (tammar wallaby), a macropod, has been used for a number of years as a biomedical research model in anatomical and physiological studies (Griffiths *et al.,* 1993, Langer, 1984), reproductive studies (Renfree *et al.,* 2011, Hickford *et al.,* 2009) and genetics research (Marshall Graves and Westerman, 2002). *M. eugenii* was the first Australian marsupial to have its genome sequenced and this was published by Renfree *et al.*, (2011). This species was included in this study since it represented the most studied marsupial species and is the main focus for genetics and comparative immunology research in Australia.

A close relative of *M. eugenii* is *Onychogalea fraenata* (bridled nailtail wallaby) which is an endangered marsupial native to Central Queensland, Australia. Prior to this study no molecular characterizations have been published for any of the immune molecules in this animal. This species was chosen because, in a captive environment, it appears to resist infection by different pathogens. This makes this species interesting and novel to investigate in respect of whether or not specific differences in its immunology can be detected when compared to *M. eugenii* and *M. domestica*.

Lagorchests hirsutus (rufous hare wallaby) is a small macropod which appears to readily succumb to mycobacterial infections (Buddle and Young, 2000) in contrast to *O. fraenata* which appears to be more resistant to such infection. Some molecular studies have been carried out for *L. hirsutus* but the main focus of research was predominantly concentrated on immunohistochemical studies (Young and Deane, 2003, Young *et al.,* 2003). This species was chosen for part of this study for comparison with *O. fraenata* and *M. eugenii*. It was considered that because of this species' susceptibility to mycobacterial infections, a sequence comparison of one of the key molecules with that of *O. fraenata* may possibly reveal differences in some of the important structural motifs.

Trichosurus vulpecula (Australian brushtail possum) was included in this study when an opportunity for collaborative research presented itself with a New Zealand research group into the immune status of that marsupial. This animal is native to Australia but a large introduced population exists in New Zealand. *T. vulpecula* is a vector of *Mycobacterium* spp and is of great concern in the cattle industry in that country.

1.1 Adaptive Immunity

The adaptive immune system responds to a pathogen that has evaded the innate (inborn) immune system (Medzhitov and Janeway, 1997). One of the main features of the adaptive immune system is the differentiation of self from non-self. This is determined by the avidity of the interactions between T cell receptors (TCRs) on T lymphocytes, specific to antigens, and major histocompatibility complex (MHC)/antigen peptides (pMHC) presented in the antigen presenting groove of the MHC on the antigen presenting cell

(APC) (Jiang and Chess, 2009, Sehgal and Berger, 2000). The adaptive immune response also ensures immunological memory which is achieved through clonal expansion and lymphocyte differentiation of antigen-specific lymphocytes. Thousands of copies of a particular antigen receptor on the cell's outer surface recognize a specific antigen. The diversity of antigen receptors is created by a process known as somatic recombination or V(D)J recombination (Matzinger, 2002). This mechanism allows a few genes to generate a large number of different antigen receptors which are then expressed on each individual lymphocyte thus enabling them to respond to antigen (Krangel *et al.,* 1998). An important functional component in the adaptive immune system is the release of the immune-modulatory molecule IL-2 which has been termed the master regulator of the adaptive immune system (Sharma *et al.,* 2011) .

T cell differentiation of naïve T lymphocytes into effector and memory subsets is one of the necessary events of T cell mediated immunity which is an essential function of the adaptive immunity (Fig.1.1).

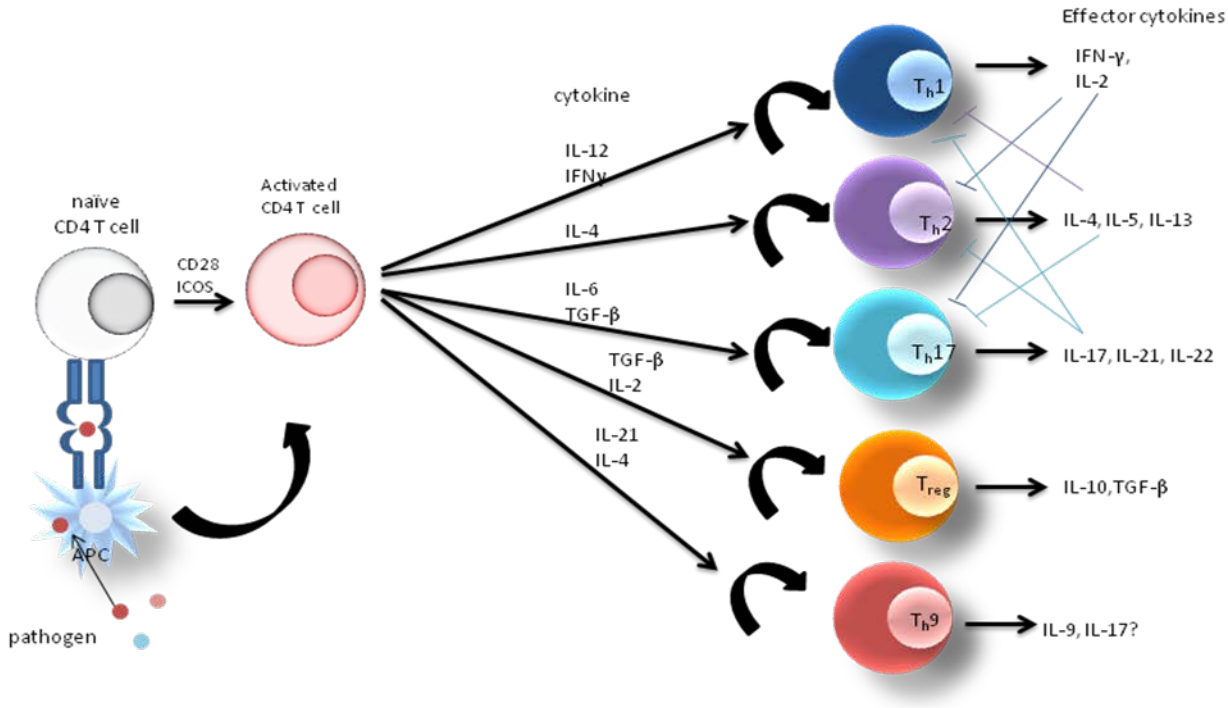


Figure 1.1. T cell sub-populations of T_h1, T_h2, T_h17, T_h9 and T_{reg} lymphocytes. Adapted from (Brand, 2009).

A detailed characterization of the different T lymphocyte phenotypes, their function, and their pathways of differentiation of T cell subpopulations has been carried out in human cells (Appay *et al.,* 2008). However in marsupials no such in-depth investigation has been undertaken to date. Those T cell sub-populations in marsupials have been identified by the characterization of receptors which identify a certain cell type. A morphological investigation of marsupial T-lymphocytes revealed a similarity to eutherian T lymphocytes but they were not identical to human T-lymphocytes (Young and Deane, 2005). T helper T lymphocytes are subsets of αβT-cells and play a role in controlling and regulating the immune system (Rolland and O'Hehir, 1999). T helper subsets express different profiles of cell surface molecules and are characterized by a distinguishing cytokine milieu.

T_h1 cells comprise the main T cell sub-population involved in the cellular immune response and host defense and are characterized by the production of pro-inflammatory

cytokines such as IFN-γ, IL-2 and lymphotoxin-α (LTα) (Spellberg and Edwards, 2001). An *in silico* prediction exists for the IFNγ gene in *M. eugenii*, however no expression studies have to date been successful (Wong *et al.,* 2006). The identification of LTα and LTβ in marsupials indicated that marsupials have the capacity to deal with intracellular pathogens. This is consistent with what has been observed in captivity (Harrison and Deane, 2000, Harrison and Deane, 1999). Interleukin-12 has been identified in *M. domestica* with an *in silico* method, however no expression studies have yet confirmed the prediction (Wong *et al.,* 2006).

The T_h2 T cell subset is mainly involved in the humoral immune response and host defense against extracellular parasites (Charles *et al*., 2010). They differentiate from a naïve CD4 cell in the presence of the cytokine IL-4 which has been elucidated in marsupials (Young, 2010). The marsupial IL-4 appeared to have a 32% sequence homology to human at the putative amino acid level. However structurally important motifs were conserved indicating that the function of the molecule might also be conserved (Young, 2010). This cell population was not relevant to this study.

While T_h1 cells are important for protection against viruses and intracellular bacteria, T_h2 cells direct immunity to extracellular parasites at the mucosal surface. A third T cell subset, T_h17, only recently discovered in humans constitutes the link between innate and adaptive immunity (van Beelen *et al.,* 2007). T_h17 cells appear to be developmentally distinct from the T_h1 and T_h2 subsets and appear to play key roles mainly in autoimmune diseases in humans (Brand, 2009, Emamaullee *et al.,* 2009, Hofstetter *et al.,* 2009). T_h17 cells produce IL-17A and IL-17F, and it has been suggested that T_h17 polarized cells may mediate the regression of established tumors (Muranski *et al.,* 2008, Martin-Orozco *et al.,* 2009). It was evident in human and mouse studies that the T_h17 sub-population is induced by IL-6, IL-1β and TGF-β (Manel *et al.,* 2008, Ouyang *et al.,* 2008, Volpe *et al.,* 2008). Despite the fact that IL-17 was identified in the wallaby genome (Renfree *et al.,* 2011), no sequence motifs have been identified in IL-17 and no immunohistochemical

investigations have been conducted. The elucidation of the IL-17 sequence will therefore further contribute to our knowledge of the marsupial immune system and will indicate how sophisticated it is in comparison to other mammals.

T_h17 cells have their antagonists in the regulatory T cell population. T_{regs} play an active role in immune tolerance and can be recognized by their cell surface receptors. T_{regs} are peripheral $CD4^+$ T lymphocytes which express CD25 on their cell surface and play an essential role in suppressing damaging immune responses (Sakaguchi *et al.,* 1995). Initially it was thought that this T cell was a variant of CD4 cells but it was soon evident that naturally occurring $CD4^+CD25^+T_{reg}$ cells were important for homeostasis and play a role in preventing immune responses to auto-antigens (Sakaguchi *et al.,* 2008). T_{reg} cells can suppress CD4 and CD8 T lymphocytes, B lymphocytes, NK cell, dendritic cells and neutrophils, hence limiting both the innate and adaptive immune responses (Azuma *et al.,* 2003, Fallarino *et al.,* 2003, Lewkowicz *et al.,* 2006).

T_{regs} participate in regulating graft rejection (Walsh *et al.,* 2004) but may also suppress anti-tumor responses and prevent the elimination of some infections such as *Pneumocystis carnii* and some multicellular parasites (Suvas and Rouse, 2006, Mittrucker and Kaufmann, 2004). T_{regs} express the Forkhead box protein 3 transcription factor which is essential for the development and function of thymic and peripheral $CD4^+CD25^+$ T_{reg} cells. A fragment of Foxp3 is annotated in the wallaby genome. The expression of this molecule in *M. eugenii* and *O. fraenata* was investigated in the current study to determine whether or not the full T cell repertoire exists in these animals.

1.2 Marsupial immunology

Immunohistochemical studies of marsupial tissues have indicated that T and B- cells could be defined by using antipeptide antibodies (Coutinho *et al.,* 1995, Hemsley *et al.,* 1995, Jones *et al.,* 1993). Use of a polyclonal anti-human CD3- and a monoclonal anti-human CD5 antibody to the Pan T cell markers CD3 and CD5 also enabled marsupial

immunologists to identify T lymphocytes in different tissues. This included gut-associated lymphocyte tissue (GALT), mesenteric lymph nodes, and tissue beds of spleen, thymus and bronchus-associated lymphoid tissue (BALT) (Old and Deane, 2001, Old and Deane, 2002b). By using these species cross-reactive antibodies, T lymphocytes were identified in adult *T. vulpecula, Pseudocheirus peregrinus* (ringtail possum), *M. domestica* (Jones *et al.,* 1993) and *Phascolarctos cinereus* (koala) (Canfield *et al.,* 1996, Hemsley *et al.,* 1995). By employing a monoclonal antibody from a conserved peptide sequence of human CD5 antigen, the cell surface marker CD5 was identified in *M. eugenii*, *Potorous longipes* (long-footed potoroo), *Potorous tridactylus* (long nosed potoroo) and *L. hirsutus* (rufous hare wallaby) (Young and Deane, 2003). The identification of T lymphocytes in the lymphoid tissue of *Isodoon macrourus* (northern brown bandicoot) was also demonstrated by using a mouse anti-CD3 antibody (Cisternas and Armati, 2000). In addition, rabbit anti-CD3 polyclonal antibodies were used to identify the same cell population in the more distantly related *Didelphis albiventris* (Brazilian white bellied opossum) (Coutinho *et al.,* 1995). The monotreme *Ornithorhynchus anatinus* (platypus) was also investigated with anti-human CD3 and CD5 antibody resulting in the identification of T lymphocytes in spleen, thymus and lymph nodes (Connolly *et al.,* 1999). This indicated that T lymphocytes could be located in marsupial tissues but no distinction could be made between the different T cell subsets except that CD3 and CD5 in human lymphoid tissue were associated with a T_h1 T cell sub-population.

It has been stated that marsupials have diminished T cell responses, such as an almost non-existent mixed lymphocyte response, and a delayed response time in skin graft rejections. They were therefore termed inferior in their immunological capacity compared to other mammals (Stone *et al.,* 1997b, Stone *et al.,* 1996). This was reported by Jurd (1994) who called the metatherian mammals 'primitive'. Marsupial immune-competence has been assessed in model animals such as *Setonix brachyurus* (quokka), *M. domestica* and *M. eugenii* in early marsupial research (Waring *et al.,* 1978, Brozek *et al.,* 1992, Stone *et al.,* 1996) . Cell mediated immune responses were studied both *in vivo* by observing

responses to graft rejection, and *in vitro* by culturing lymphocytes with mitogens (Old and Deane, 2000). It was further reported that skin grafts were successful in young immunologically underdeveloped *S. brachyurus* (Waring *et al.,* 1978). No further experiments were carried out to investigate the mechanism of skin graft acceptance in different juvenile age groups. Similarly, the transplantation of thymus from 30 day and 40 day postpartum young were reported to be successful in the same species (Ashman *et al.,* 1975). Old and Deane (2000) reported a low stimulation index in a mixed lymphocyte culture from marsupial pouch young while Baker *et al*. 1999 documented a rise in the proliferative response in aged animals. It appears that the argument concerning the *in vitro* stimulation and mixed lymphocyte reaction in marsupials remains unresolved.

As research continued over the years into the immune competency of marsupials it was found that although suffering from serious mycobacterial infections some did not succumb easily to the disease (Buddle and Young, 2000). As a vector for *Mycobacterium bovis, T. vulpecula* poses a threat to the economic viablility of the cattle industry in New Zealand. *Mycobacterium avium* also poses a great risk to endangered marsupials in captivity. Buddle and Young (2000) linked the inability to wall off the mycobacterial infection sites to deficiencies in the cellular immunity of marsupials. The same argument was presented by researchers investigating mycobacterial infection in *Dendrolagus matschiei* (Matschie's tree kangaroo). These animals have a documented history of *M. avium* infections and there is no evidence of any immunosuppressive retroviral infection in that species (Montali *et al.,* 1998). This differes from *P. cinereus* which often suffers from koala immunodeficiency virus (KIDS) and chlamydia and was shown to have retroviral activity in its genome (Hanger *et al.,* 2000).

Other marsupials such as *O. fraenata* were observed to resist infection with *Toxoplasma gondii* and *Echinococcus granulossus* in captivity (Turni and Smales, 2001). A detailed study of *E. granulosus* in *M. eugenii* found differences in this macropod's response to the infection when compared to *Ovis aries* (sheep) (Barnes *et al.,* 2007). *M. eugenii* did not show the conventionally observed cellular recruitment, granuloma organization and

fibrosis. It was therefore concluded that the T_h1 (cell mediated) immune response in *M. eugenii* was inefficient (Barnes *et al.,* 2007).

P. cinereus (koala) was investigated in relation to antibody production in response to soluble antigens and particulate antigens. When compared to *Oryctolagus cuniculus* (rabbit) the response was very slow (Wilkinson *et al.,* 1994, Wilkinson *et al.,* 1992), however when these experiments were repeated in an *in vitro* lymphocyte assay the responses were very similar to those of other eutherian mammals (Wilkinson *et al.,* 1994).

By the turn of the twentieth century a few contradictory statements were found in the literature in regard to marsupial immunology. While the majority of researchers maintained that marsupial immune systems were under-developed, Stone *et al.,* suggested otherwise. Harrison and Wedlock (2000) took an interesting position stating that marsupials and eutherian mammals share similarities, but also have substantial differences such as the presence of two thymuses in the macropods. A distinct difference in the B cell biology of marsupials where delayed secondary antibody responses are dominated by IgM rather than the obligatory class-switch from IgM to IgG was also reported. So to was the lack of T cell dependent carrier effects in secondary antibody responses (Dean and Cooper, 1998). It was further hypothesized that the cytokines will hold the key to the functioning of the marsupial immune system (Harrison and Wedlock, 2000).

During the process of marsupial genome sequencing, researchers came to the conclusion that marsupials may not be so different to the eutherian mammals (Samollow, 2008, Belov *et al.,* 2007). *M. domestica* was the first marsupial to have its genome sequenced (Mikkelsen *et al.,* 2007) followed *by M. eugenii,* an Australian macropod (Renfree *et al.,* 2011). The latest marsupial genome sequencing project was that of *Sarcophilus harrisii* (Tasmanian devil) infamous for the terrible Devil Facial Tumor Disease (DFTD), a transmissible tumor that has brought this species to the brink of extinction

(Murchison *et al.,* 2012). It was hoped that the elucidation of the genome would hold the answer to the survival of that species.

A characterization of key elements in the adaptive immune system using the marsupial genomes would identify any differences in the molecular makeup of marsupial genes compared to the same genes in other mammals. Nucleotide sequences, intron/exon structures, gene ontogeny and the ability to infer structure and functionality would quickly indicate any differences to other mammalian sequences and show significant modifications or similarities. Investigating key molecules of the T cell signalling cascade at the message (mRNA) level would therefeore add to the debate as to whether marsupials are as sophisticated in their immunological makeup as are other mammals.

The diagram in Fig. 1.1 compares the elements of the T cell signalling cascade of humans with what was known in marsupials before the present study was undertaken.

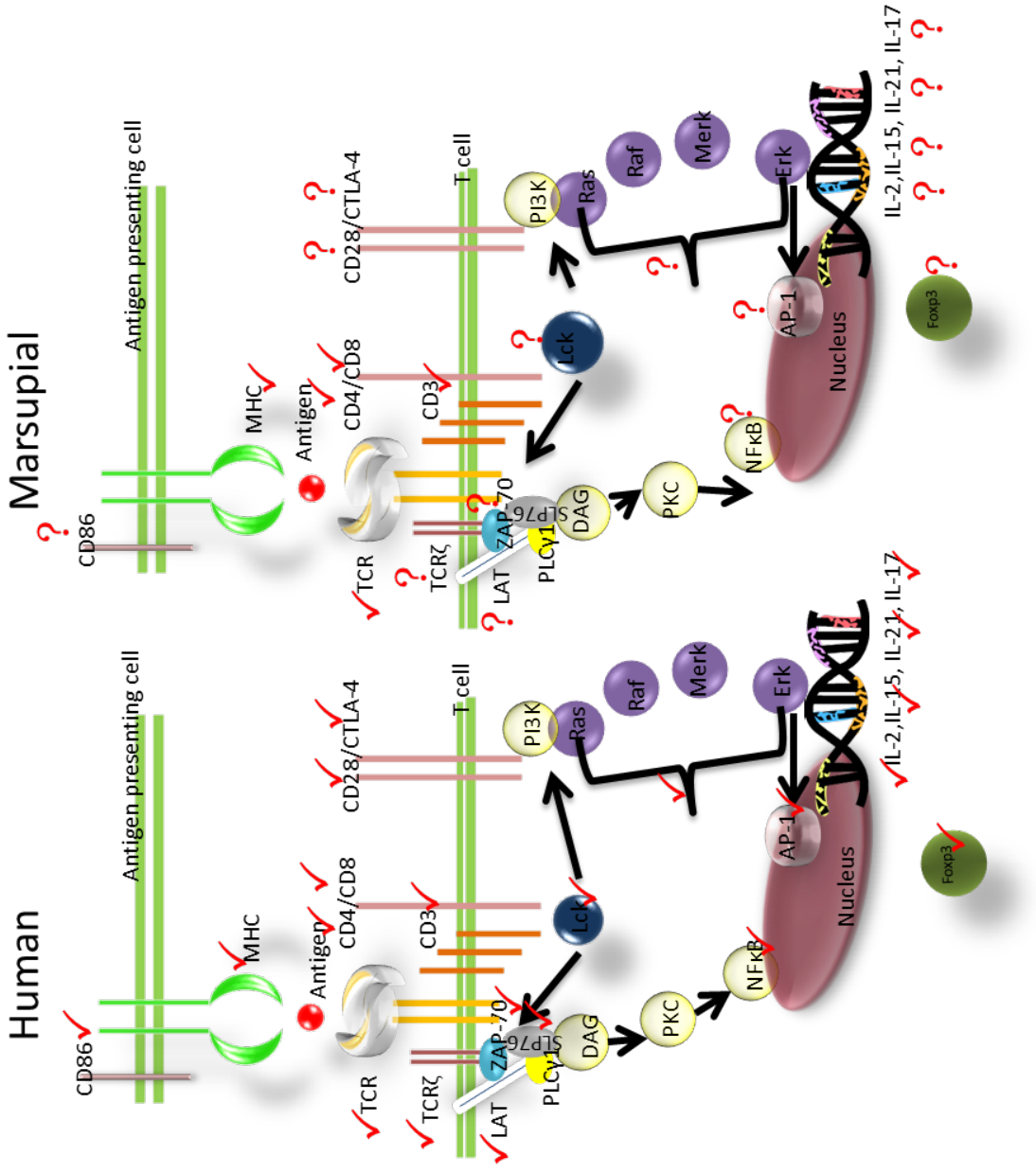


Figure 1.2. Comparison of the known components of the T cell signalling cascade between human and marsupials.

1.3 T cell signalling cascade

The T cell signalling cascade is initiated when an antigen binds the antigen binding groove of the major histocompatibility complex molecule located on an antigen presenting cell. This event triggers the TCR located on T lymphocytes. This is the first stimulus that will initiate the T cell signalling cascade in humans and other mammals. The activation of the TCR promotes a number of signalling cascades that determine cell fate and the transcription of immune regulatory molecules (cytokines) that regulate cell survival, proliferation and differentiation of T lymphocytes (Dranoff, 2004).

An early event in T cell signalling is the phosphorylation of the immunoreceptor tyrosine-based activation motifs (ITAMs) on the cytosolic side of the TCR/CD3 complex by lymphocyte protein-tyrosine kinase (Lck). Once Lck and other Src family tyrosine kinases are activated through phosphorylation, the zeta-chain associated protein kinase (ZAP-70) is recruited to the TCR/CD3 complex (Burbach *et al.,* 2007). This event activates ZAP-70 which in turn promotes recruitment and phosphorylation of adaptor and scaffold proteins found downstream from ZAP-70 (Alonso *et al.,* 2003). ZAP-70 phosphorylates the SH_2 domain containing leukocyte protein of 76kDa (SLP-76) which recruits the guanine nucleotide exchange factor (VAV) and the non-catalytic region of tyrosine kinase adaptor protein (Nck), the Grb2 related adaptor protein (GADS) and the IL-2 inducible tyrosine kinase (ITK) (Qi and August, 2007). Transient molecules such as phospholipase (PLCγ1) produce second messengers such as diacylglycerol (DAG) and inositol trisphosphate (IP_3). DAG activates the protein kinase C-theta (PKCθ) and the Mitogen activated kinase (MAPK)/Extracellular Signal-Regulated Kinase (Erk) pathways, both promoting transcription factor nuclear factor kappa beta (NF-κB) activation (Okkenhaug *et al.,* 2004). The release of calcium from the endoplasmic reticulum (ER) is triggered through IP_3, and through the activation of calcium bound calmodulin, calcineurin is activated. This promotes the transcription of the IL-2 gene through the transcription factor nuclear factor of activated T lymphocytes (NFAT).

The activation of the transcription factor NFκB is dependent on stimulation of the TCR and co-stimulation from CD28 (Davis, 2002). This cascade is negatively regulated in order to

keep hyperactivation of immune responses under control. The cytotoxic T-lymphocyte antigen-4 (CTLA-4) negatively regulates T cell activation (McCoy and Le Gros, 1999). Once T lymphocytes are activated, CTLA-4 is rapidly endocytosed thus removing it from the cell surface. ZAP-70 initiates the process of active release and translocation of CTLA-4 to the membrane after TCR signalling. CTLA-4 antagonizes T cell function through the inhibition of CD28 signalling by competing for the CD80 and CD86 ligands which are located on the antigen presenting cell (Gough *et al.,* 2005, Lin *et al.,* 1998).

T lymphocytes are important in cell mediated immunity which is a part of the adaptive immune system. It is this part of the immune system that is involved in the defense against intracellular pathogens (Anderton, 2006). It is also involved in graft rejection, and in the mixed lymphocyte response, either performed *in vitro* (Waldmann, 2006, Rich and Rich, 1974) or during the sensitization phase, where T lymphocytes recognize allo-antigens expressed on cells of the foreign graft and proliferate in response (Kreijveld *et al.,* 2008).

1.4 Key molecules of the T cell signalling cascade

1.4.1 The Major Histocompatibility Complex

In the human T cell signalling cascade the initiation of the signal begins from when the MHC captures and internalizes an antigen. The MHC molecule is a cell surface glycoprotein on the surface of an antigen presenting cell (APC) and determines compatibility of donors in transplant medicine as well as the susceptibility of an individual to an autoimmune disease (Gerloni *et al.*, 2004).

The MHC gene family comprises class I and II regions which encode antigen presenting molecules, and a class III region containing immune and non-immune genes. The chicken has one of the smallest class II regions with 19 genes (Kaufman *et al.,* 1999), while the human class II region contains 264 genes (Horton *et al.,* 2004, The MHC sequencing consortium, 1999).

In most other mammals, MHC class III is located between classes I and II. In *R. norvegicus* (rat) the class I genes are adjacent to the antigen processing genes (Kelley *et al.,* 2005) while in *Sus scrofa* (pig) the class II genes are separated from the remainder of the MHC (Joly *et al.,* 1998). Amphibian and bird MHC molecules appear to be similar to the human gene organization with three closely linked regions (Ohta *et al.,* 2006). However in teleost fish the class II loci are divided between two linkage groups distinct from the linkage group containing the class I loci (Bingulac-Popovic *et al.,* 1997). The MHC of *M. domestica* contains 114 genes, 87 of which are shared with humans (Belov *et al.,* 2006). The genotypic variation of the marsupial MHC lies between eutherian mammals and birds, but the gene organization of the marsupial MHC appears to be closer to that of non-mammals. An investigation by Belov *et al.* (2006) indicated that the MHC class I genes amplified within the class II region, creating a unique class I/II region not seen in any other animal. Recent studies indicate that there has been a shift in the *M. eugenii* MHC away from the core MHC due to retroviral activity (Siddle *et al.,* 2011). However, the *M. eugenii* MHC is predicted to consist of 129 putative functional genes, which puts this marsupial between *M. domestica* and human in terms of the size of the MHC molecule (Siddle *et al.,* 2011). Although there appears to be a difference in the organization of the MHC genes between vertebrates, and even between mammals, the function appears to be conserved.

The human MHC has great diversity and is highly polymorphic, acting almost like a fingerprint, and in a mixed population there are no two individuals, except for identical twins, with the same set of MHC genes. However there are species, such as *Acinonyx jubatus* (Cheetah) (Castro-Prieto *et al.,* 2011), *Castor fiber* (Eurasian beaver)(Babik *et al.,* 2005) and *Ailuropoda melanoleuca* (giant panda)(Zhu *et al.,* 2007) where low MHC diversity has been reported. Low MHC diversity was identified as the contributing factor in the near demise *of S. harrisii* which appears to be threatened by a transmissible tumor involved in devil facial tumour disease (DFTD) (Siddle *et al.,* 2007).

The MHC molecule plays a major part in transplant biology. In a transplant situation the MHC molecules themselves are antigens and incite an immune response in the recipient leading to transplant rejection (Janeway, 2001).

Another molecule important in graft rejection and in the T cell signalling cascade is the co-receptor CD86 located on the antigen presenting cell.

1.4.2 CD86

It is well documented that T lymphocytes require co-stimulatory signals for optimal activation and clonal expansion. CD28 is thus far the best characterized co-stimulator associated with the known ligands CD80 and CD86 (van Rijt *et al.,* 2004, Nakajima *et al.,* 1997). CD86 is a type I membrane protein and a member of the immunoglobulin superfamily. CD86 is involved in various immune responses including graft rejection, and is stimulated via the CD28 receptor. This system has a major role in regulating inflammation in autoimmune diseases and in graft vs. host disease (Nolan *et al.,* 2009).

1.4.3 The T cell receptor alpha /beta (TCR αβ) and CD3 complex

The T cell receptor, like the MHC molecule, is a disulfide linked dimeric receptor located on the surface of T lymphocytes which have an antigen binding cleft located between the α and β chains. These two chains have specificity for antigen but no signalling capacity which is consistent with the short cytoplasmic tail of this molecule (Holst *et al.,* 2008). TCR signalling commences with the activation of protein tyrosine kinases that are mediated by the Src kinases Lck and Fyn, the 70kDa zeta-associated protein kinase, and members of the Tec kinase family (Nel and Slaughter, 2002). Signalling is induced by the phosphorylation of tyrosine residues in these molecules which leads to the activation of specific signalling molecules (Weiss and Littman, 1994). The signalling task falls upon a complex called the CD3 complex, an accessory molecule to α which is composed of six invariant polypeptide chains (Abram and Lowell, 2007). The assembly of the CD3 complex proteins is responsible for the transport of the receptor to the cell surface where the TCR binds to peptide MHC (pMHC) thus activating the signalling cascade.

The TCR and the glycoproteins of the CD3 complex are developed and expressed in late thymic ontogeny (Reinherz *et al.,* 1986). Interspecies sequence comparisons of structural elements of the mammalian TCRs found that they exhibit clear homologies to the human TCR especially in the constant domains (Marchalonis *et al.,* 1996).

The TCR of humans is composed of a 51kDa alpha and a 43kDa beta chain which is similar to the TCRs of other species (Acuto *et al.,* 1983). Both chains contain variable regions and constant domains (Cole *et al.,* 2007). The V domain contains the complementarity determining regions (CDRs). The number of CDRs vary between the alpha and beta chains (Haynes and Wu, 2007). The CDR1 and 2 regions bind to the MHC, while the CDR3 domain contains a single binding site for an antigen (Haynes and Wu, 2007). Further evidence of a conserved structure was found when the sequences of the *O. aries* TCRα chain was elucidated. This indicated that the primary structure of the immunoglobulin (Ig) domain, transmembrane domain, and cytoplasmic C gamma domains were conserved compared to human (Hein *et al.,* 1991). The same was found when the sequence of the *S. scrofa* TCR was revealed (Thome *et al.,* 1993).

It appears that the CDR3 region is well conserved, not only among mammals but also in birds and teleost fish, indicating that this region experienced selective pressure in vertebrate evolution (Partula *et al.,* 1996, Gobel *et al.,* 1994). The hinge region in the TCR alpha chain varies greatly from species to species indicating that it is a biologically important domain. The TCRα chain sequence has been characterized in many mammals and vertebrates including the monotremes which have a sequence identity of 37% in the constant region compared to other mammals (Belov et al., 2004). The alpha chain of the TCR has also been elucidated in the marsupials *T. vulpecula* (Zuccolotto *et al.,* 2000) and *M. domestica* (Baker *et al.,* 2001). The TCRα sequences of those marsupials encompass all the important structural motifs found in other mammals.

The TCRβ chain, like the α chain, can be identified by the hypervariable CDR3 region which is the site of V(D)J recombination and antigen contact (Freeman *et al.,* 2009). The alpha chain only undergoes a VJ recombination which may explain the reduced conservation in the alpha chain among mammals. The recombination process is facilitated by the RAG proteins RAG-1 and RAG-2 both of which have been characterized in *M. domestica* (Miller and Rosenberg, 1997).

From a sequence alignment it can be determined that the constant domains of the TCRβ gene are highly conserved in a number of mammals such as *H. sapiens*, Mus *musculus*

(mouse), *Rattus norvegicus* (rat) and *Oryctolagus cuniculus* (rabbit), while the teleost fish and avian sequences are conserved among each other but show greater diversity to the mammals. The known marsupial sequences are highly conserved to other mammalian sequences in the CDR3 regions, but differ at the N-terminal end of the molecule.

A sequence homology of ~30% was observed between the amphibian TCRβ sequences compared to mammalian sequences (Chretien *et al.,* 1997). Anurians also have a conserved TCRβ chain that contains elements involved in molecular interactions with the α chain as well as some of the polypeptide chains that make up the CD3 complex (Fellah *et al.,* 1993). Elements such as the FG loop, the beta core-peptide, and the F-loop were found in all mammalian sequences. The TCRβ chain of *T. vulpecula* had a 67% identity to the human β1 sequence. Reference to a partial sequence of *M. eugenii* TCRβ was made however the sequence has not been published to date (Zuccolotto *et al.,* 2000). The beta chain has also been characterized for *M.domestica* and a phylogenetic analysis revealed that the marsupial TCR constant domains of both TCRα and TCRβ group together to form a sister group to other mammals (Baker *et al.,* 2001).

The CD3 complex consists of eight polypeptide chains which are expressed on the cell surface in an orderly fashion after being assembled in the endoplasmic reticulum(ER) (San José *et al.,* 1998). The polypeptide chains form dimers by pairwise interaction. CD3ε forms a dimer with either CD3γ or CD3δ. Once formed, they assemble with the TCR chains and on completion the ζ unit joins the receptor. The whole molecule is then transported from the ER to the cell surface, thus making the TCRζ chain the rate limiting step of this process (Jose *et al.,* 1998).

The CD3ε and TCRζ chains are important in transmitting a signal from the TCR since the cytoplasmic tail of the receptor is too short to transduce the signal. Therefore the gene organization of both of these molecules is important in order to elucidate any possible differences in function. The human CD3ε consists of nine exons. Three of these encode the junction of the signal peptide and mature protein and were found to be very small containing only 21, 15 and 18 base pairs respectively (Clevers *et al.,* 1988b). These small

exons have only been found in *H. sapiens* and *M. musculus* where the latter has only two small exons.

The CD3ε chain is one of the most important chains of the CD3 complex and has been investigated in many vertebrates. The molecule has also been investigated in the marsupial *M. eugenii* showing a high sequence homology to humans (>50%) and other mammals, especially in the ITAM region (Old *et al.,* 2001). However no other marsupial CD3ε molecule has been elucidated therefore, prior to the present study, no comparison could be made between the marsupials and other mammals.

Other subunits of the CD3 complex, such as CD3γ and CD3δ, have also been characterized in marsupials. The CD3γ chain was defined in *M. eugenii* (Harrison *et al.,* 2003a), while the CD3δ chain was characterized in *Isoodon macrourus* (northern brown bandicoot), *M. eugenii* and *M. domestica* (Baker *et al.,* 2005).

1.4.3.1 The T cell receptor gamma/delta (TCRγδ)

γδ T lymphocytes are a small subset of T lymphocytes that carry a distinct T cell receptor on their surface. This receptor consists of a single γ and a single δ chain. This group of T lymphocytes is a minor cell population in the peripheral blood but is found abundantly among intestinal intraepithelial lymphocytes (Holtmeier and Kabelitz, 2005). These cells can be placed between the adaptive immune and innate immune response (Born *et al.,* 2006) since they re-arrange TCR genes similar to TCRαβ T lymphocytes and also develop a memory phenotype. On the other hand, restricted TCRs can be used as pattern recognition receptors (Morita *et al.,* 2000). γδ T lymphocytes have shown distinct myeloid characteristics by presenting as professional phagocytes which make them an ancient lymphocyte in an evolutionary sense (Wu *et al.,* 2009). In contrast to the αβ T lymphocytes the γδ T lymphocytes are either MHC restricted or can bind free antigen which is similar to the function of immunoglobulins (Hayday, 2000).

Both chains of this TCR have been characterized in marsupials, specifically *I. macrourus, M. eugenii* and *M. domestica* (Baker *et al.,* 2005, Harrison *et al.,* 2003b). The investigation of

these chains in marsupials indicated that the structural features are conserved compared to humans and other eutherians.

A unique T cell receptor that appeared to share similarities with the TCRδ chain has been identified in the marsupials *I. macrourus*, *M. domestica* and *M. eugenii* which are distantly related to each other (Parra *et al.,* 2007), and in the monotreme *O. anatinus* (Wang *et al.,* 2011b). This TCR appeared to have pre-joined V, D and J segments in germline DNA as opposed to somatic recombination (Parra *et al.,* 2009). The newly found T cell receptor was named mu for marsupial and has two isoforms TCRμ1.0 and TCRμ2.0. The TCRμ2.0 appeared to be analogous to the TCRδ isoform described in *Ginglymostoma cirratum* (nurse shark) which also contains a double V domain (Criscitiello *et al.,* 2006). Further research indicated that although TCRμ appears to be similar to the IgNAR, a shark IgH chain isotype that contains two V domains, TCRμ is not orthologous to the shark molecule (Criscitiello *et al.,* 2010, Parra *et al.,* 2007). Whether this newly found TCR binds MHC or binds free antigens like the γδTCR is unknown at present. The signalling capacity of that TCR is also unknown.

1.4.4 The Co-receptors CD4, CD8, CD28 and CTLA-4

Co-receptors are molecules that are expressed on the cell surface and either bind ligands to effect signalling or transmit via adaptor molecules through cytoplasmic domains that bind signalling motifs (Kirkbridge *et al.,* 2005). Co-receptors such as CD4 and CD8 require serine or threonine residues in cytoplasmic domains in order to transduce a signal. CD4 and CD8 are some of the best-characterized proteins of the adaptive immune system. The expression of CD4 and CD8 on T lymphocytes is mutually exclusive (Wang and Bosselut, 2009). $CD4^+$ and $CD8^+$ T lymphocytes are restricted by MHC class II and class I respectively (Meuer *et al.,* 1984). Co-receptors are often associated with the TCR which is unable to transduce and regulate the signal into the cell (Kirkbridge *et al.,* 2005).

1.4.4.1 The CD4 co-receptor

Human CD4 is a 55kDa glycoprotein consisting of nine introns and ten exons, expressed on the surface of MHC class II restricted T helper cells (Zhang *et al.,* 2008, Hanna *et al.,* 1994). CD4 is a member of the immunoglobulin superfamily and has four immunoglobulin domains of which D_1 and D_3 resemble immunoglobulin variable domains located at the N-terminal end of CD4 (Ryu *et al.,* 1990). D_2 and D_4 resemble the constant regions of immunoglobulin molecules (Zhang *et al.,* 2008, Wu *et al.,* 1997). CD4 also contains a joining (J)- like region, a third extracellular domain, a membrane spanning region which is homologous to the MHC class II β-chains, and a cytoplasmic domain that contains phosphorylated tyrosine's (Maddon *et al.,* 1985). CD4 interacts with the MHC class II molecule via the β_2 domain, while specific residues in the cytoplasmic domain interact with the signalling molecule lymphocyte specific antigen (Lck). The interaction of CD4 with the MHC and the T cell receptor makes it an essential molecule responsible for T cell activation and it increases the affinity of thymocytes with antigen presenting cells (Marrack *et al.,* 1983, Reinherz and Schlossman, 1980).

CD4 has been characterized in several vertebrates including some fish species such as *Hippoglossus hippoglossus* (Atlantic halibut) (Patel *et al.,* 2009) and *Carassius auratus langsdorfii* (ginbuna crucian carp) (Nonanaka *et al.,* 2008). It was found that in those species the Lck binding site is conserved the same as in mammals. CD4 was also characterized in the marine mammal *Delphinapterus leucas* (Beluga whale) which has a similar gene organization to that of human and mouse CD4 molecules. Differences in the secondary structure of CD4 in *D. leucas* were detected thus indicating possible differences in T cell responses and activation (Romano *et al.,* 1999). CD4 has also been elucidated in the marsupials *M. eugenii* and *M. domestica* (Duncan *et al.,* 2007). Secondary structure variations were found in these animals. In marsupials, a cysteine substitution prevents the building of a disulphide bridge in D1 of CD4. The consequences of these biochemical changes are unknown. It appears that the binding sites on CD4 that bind the MHC class II molecule are intact in marsupials. It is therefore expected that, in marsupials, antigen recognition functions in a similar manner to human CD4. Since CD4 controls important T

cell functions, it was important to investigate this molecule in *O. fraenata* due to the documented ability of that species to efficiently combat pathogens (Konig *et al.,* 2002, Turni and Smales, 2001).

1.4.4.2 The CD8 co-receptor

CD8, like CD4, is a heterodimeric co-receptor to the T cell receptor. It consists of a disulphide bonded alpha and beta chain and binds an MHC class I molecule. Human CD8 is a member of the immunoglobulin superfamily containing an immunoglobulin variable extracellular domain (Parnes *et al.,* 1985) which is connected via a thin stalk to the cell membrane (Gao and Jakobsen, 2000). CD8 also contains a hydrophobic transmembrane domain (Sukhatme *et al.,* 1985) and a cytoplasmic tail that contains an Lck binding motif similar to that of CD4 (Kim *et al.,* 2003). CD8 was first discovered as a cell surface marker in mice and was used to distinguish between $CD8^+$ cytotoxic T lymphocytes and $CD4^+$ T helper cells (Littman *et al.,* 1985). CD8 can exist as a heterodimer (CD8αβ) as well as a homodimer (CD8αα). It is involved in cytotoxic T lymphocyte co-activation due to increased antigen sensitivity, and stabilizes the TCR/pMHC interaction (Devine *et al.,* 2000).

The CD8 molecule binds to a distinct region in the MHC class I molecule and induces TCR/pMHC/CD8 interactions. In humans, the main interface between CD8αα and pMHC class I lies between CD8 residues 51-55 and the pMHC residues 223-29 in the α3 domain which forms the CD8 binding loop (Currier *et al.,* 2002).

CD8 is involved in the immune response to infections caused by *M. bovis* and *M. avium* (Pollock *et al*., 2005). Endangered marsupials, such as *L. hirsutus,* have been shown to suffer from *M. avis* infection in captivity. They are unable to wall off infection sites resulting in the formation of secondary lesions leading to severe disease states (Buddle and Young, 2000). This process is linked to cell mediated immunity and specifically to CD8. There is evidence from mouse models that MHC class I restricted $\alpha\beta TCR^+CD8^+$ T lymphocytes play a major role in protection from mycobacterial infections (Smith and Dockrell, 2000, Canaday *et al.,* 1999).

Observations by zookeepers and animal handlers suggest that the marsupial *O. fraenata* is less susceptible to mycobacterial infections than are other macropods kept in captivity. The elucidation of CD8 in *M. eugenii* and *M. domestica* further advanced the knowledge of the T cell biology in marsupials (Duncan *et al*., 2008). However, since those two species are only distantly related to each other, the elucidation of CD8 in *O. fraenata* would possibly show variances that may explain the observed differences in disease susceptibility between species.

1.4.4.3 The CD28 co-receptor

CD28 delivers the second signal required in the activation of T lymphocytes while the first signal is delivered through the TCRαβ/CD3 complex (Alegre *et al.,* 2001). The combined signals from CD28 and TCRαβ lead to the expression of the cytokine interleukin-2 through a CD28 response element (CD28RE) in the IL-2 promoter. CD28 stabilizes the IL-2 mRNA and enhances T cell survival through a stabilization motif in the 3' un-translated region (Hehner *et al.,* 2000, Boise *et al.,* 1995). CD28 binds the ligands CD80 and CD86, but when a T cell interacts with an MHC molecule without the surface expression of either CD28/CD80 or CD86 the T cell is said to be anergic or unresponsive (Schwartz, 2003).

The proteins CD28, CTLA-4, ICOS and BTLA belong to the family of co-stimulatory receptors and each contains a single V-type extracellular immunoglobulin domain, a transmembrane domain, and an intracytoplasmic domain. The human CD28 is a 44kDa glycoprotein that contains a proline rich region that binds to the SH_3 domain in molecules such as Itk and Tec at the N-terminal. The C-terminal end of CD28 binds the signalling molecule Lck and transports it to the lipid raft of the immune synapse (Evans *et al.,* 2005).

CD28, like the other molecules discussed so far, belongs to the immunoglobulin superfamily and contains an important hexapeptide MYPPPY motif for binding the CD80 and CD86 ligands (Yang *et al.,* 2003a). A YMNM motif found in the cytoplasmic domain of human CD28 is the binding site for phosphatidylinositol 3-kinase (PI3K) and Grb2 (Harada *et al.,* 2003). CD28 exists as a disulphide linked homodimer due to an extracellular interchain disulphide bond which lies in the linker region (Lazar-Molnar *et al.,* 2006)

connecting the immunoglobulin variable region and the transmembrane domain (Aruffo and Seed, 1987).

CD28 crosslinking under physiological conditions alone does not result in a significant induction of transcription nor in any functional response of a T cell. It has therefore been argued that the role of CD28 is a mere TCR signal transduction amplifier (Michel *et al.,* 2001).

A partial annotation of the CD28 co-receptor for *M. eugenii* is in the ensembl database. There are 453bp in the database but the start site and the 3' end of the open reading frame are missing. It was therefore important in this study to complete the sequence in order to identify structural motifs, especially the motifs that play a role in the regulation of T cell responses.

1.4.4.4 Cytotoxic T- lymphocyte antigen-4 (CTLA-4) - the negative regulator

CLTA-4 is a glycoprotein (33-37kDa) consisting of four exons and is also known as CD152. Human CTLA-4 is viewed as having a strand topology where the top surface is formed by the CDR regions known as CDR1, 2 and 3 that resemble the V-alpha domains. CTLA-4 is able to dimerize the B7 binding site away from the dimerization interface thus allowing it to bind two divalent B7 molecules. It is the CDR3 region that plays the major role in the binding of the B7 molecule (Ostrov *et al.,* 2000).

The human CTLA-4 sequence contains a short cytoplasmic tail with two potential Src homology (SH_2) domain binding sites centered at tyrosine residues 165 and 182, spaced by a proline-rich stretch. There appears to be an association between CTLA-4 and phosphatidylinositol-3 kinase (Schneider *et al.,* 1995), and between CTLA-4 and the SH_2 domain-containing protein (SHP) -2 phosphatase (Marengere *et al.,* 1997). In humans the expression of CTLA-4 on the cell surface is dependent on phosphorylation of a tyrosine residue (Y^{165}) in the cytoplasmic domain of CTLA-4 (Baroja *et al.,* 2000). Phosphorylation of this residue prevents the interaction of CTLA-4 (Bradshaw *et al.,* 1997, Chuang *et al.,* 1997, Shiratori *et al.,* 1997) with the clathrin-associated AP-2 internalization adapter and hinders CTLA-4 internalization (Owen and Evans, 1998). The CTLA-4 dimer interface in the

complex is formed by residues that are positioned strategically within the different strands in the C- terminal end thus burying some of the access area (Schwartz, 2001).

CD28 and CTLA-4 play a role in T_h1/T_h2 differentiation. It has been reported that ligation of CD28 during TCR stimulation *in vitro* promotes the differentiation of T_h2 cells (Rulifson *et al.,* 1997). Higher concentrations of antigen may favor T_h2 development but little is known about the mechanism by which CTLA-4 exerts its inhibitory function. It has been suggested that CTLA-4 might 'steal' B7 ligands making them unavailable to bind CD28, thus reducing T cell responses (Fallarino *et al.,* 2006).

Studies indicate that CTLA-4 disrupts stimulatory signalling complexes by competing with CD28 for binding the B7 isoforms, and promotes the assembly of inhibitory signalling complexes (Schwartz *et al.,* 2001). Unlike CD28, which is expressed on resting T lymphocytes, CTLA-4 is not detected on the cell surface until 24 hours after activation. In addition, CTLA-4 exhibits a 10 to 100 times greater affinity for the B7 isoforms than for CD28 (Bluestone, 1997). The balance between the opposing signals elicited by CD28 and CTLA-4 is central to the regulation of T cell responsiveness and homeostasis (Bluestone, 1997).

M. eugenii CTLA-4 is annotated in the ensembl database. However unlike CD28 it appears that the full transcript is in the ensembl database. This facilitated the design of primers to investigate the expression of CTLA-4 in various marsupial tissues in this study.

1.5 Signalling molecules

Signalling molecules activate specific receptors on the cell membrane and through this event either a conformational change is triggered or a secondary messenger molecule transmits a signal into the cell thus prompting a response.

1.5.1 T cell receptor zeta chain (TCRζ)

The TCRζ chain, or CD247, is a component of the TCR complex that plays a critical role in the assembly, transport and signal transduction of the TCR complex. This molecule is required for the expression of the TCR/CD3 complex as well as transporting the fully

assembled complex from the Endoplasmic Reticulum to the Golgi apparatus (Chowdhury *et al.,* 2005, Geisler *et al.,* 1989). In the absence of the zeta chain, transportation of the TCR/CD3 complex is greatly diminished. The TCRζ is a mediator between the binding of an antigen and the transduction of a signal leading to T cell activation (Duchardt *et al.,* 2007).

The human TCRζ gene is composed of eight exons, separated by distances of between 0.7 and 8 kb (Weissman *et al.,* 1988). The spliced mRNA product of the human gene is 1492kb long, comprising a 492bp coding domain and a 3'-untranslated region (3'-UTR) spanning 906 bp (Lanier *et al.,* 1989). The TCRζ chain forms a disulfide-linked dimer and contains a triplicated immunoreceptor tyrosine activation motif (ITAM). All ITAMs consist of a pair of YXXL (where Y denotes the amino acid tyrosine, X is any amino acid residue, and L denotes the amino acid leucine) sequences separated by seven or eight amino acid residues (Irving *et al.,* 1993). The tyrosine (Y) and leucine (L) residues are required for the functional activity of the ITAM motif in the cytoplasmic tail of the molecule (Irving *et al.,* 1993). The zeta homodimer is the last unit to join the complex and the extracellular domain (EC) of this subunit regulates receptor assembly (Minguet *et al.,* 2008).

The short EC domain of TCRζ consists of nine amino acids. The length, but not its primary amino acid sequence, is highly conserved. This was demonstrated in *Gallus gallus* (chicken) where only four of the nine amino acids are conserved (Gobel and Bolliger, 1998). The highest identity (81%) between chicken and mammalian ζ-chain sequences is found in the transmembrane (TM) domain. This includes the cysteine residue involved in the dimerization of the ζ-chain, and a negatively charged residue which is a putative site of interaction with TCR chains (Gobel and Bolliger, 1998). The EC is the main functional unit of the zeta chain and, when reduced from the customary nine amino acids to two amino acids, the receptor assembly is inhibited.

There is no annotation for the *M. eugenii* TCRζ in the ensembl database, and this molecule has not been characterized in any marsupial. The molecular characterization of the marsupial TCRζ chain was reported for the first time in this study.

1.5.2 TCRζ chain and the immunoreceptor tyrosine activation motif (ITAM)

Recognition of foreign antigens by lymphocytes initiates a cascade of biochemical steps which lead to cellular activation (Bakker and van der Merwe, 2002). A specialized motif found in antigen receptors are the immunoreceptor tyrosine based activation motifs (ITAMs) which recruit other tyrosine kinases such as ZAP-70 (Borroto *et al.,* 1999). Specifically, once the signalling cascade is activated by the engagement of the TCR with the pMHC the tyrosine kinase Lck also becomes activated and in turn phosphorylates the intracellular portions of the CD3 complex. The most important member of the CD3 family is the TCRζ chain to which the signalling molecule ZAP-70 binds. The tandem SH_2-domains of ZAP-70 are engaged by the doubly phosphorylated ITAMs of TCRζ, which together position ZAP-70 to phosphorylate the transmembrane protein Linker of Activated T lymphocytes (LAT) (Deindl *et al.,* 2007).

This ITAM motif is responsible for the signal transduction ability of the CD3 complex. When activation occurs, phosphorylation of the zeta chain leads to the recruitment of kinases and adaptor proteins that serve as docking sites to organize multi-protein complexes (Methi *et al.,* 2007). This results in the phosphorylation of several substrates such as Cγ1 (phospholipase C gamma1), MAPK (mitogen activated protein kinase), ERK1 (extracellular signal regulated kinases), and ERK2. It also activates gene transcription factors such as NFκB (nuclear factor kappa beta) resulting in the expression of IL-2 (Livolsi *et al.,* 2001).

1.5.3 Zeta associated protein of 70 kDa (ZAP-70)

ZAP-70 belongs to the Syk family protein tyrosine kinases (PTKs), is important for the selective activation of T lymphocytes through its interaction with the zeta chain of the TCR/CD3 complex, and is necessary for thymocyte development (Hanks and Quinn, 1991). ZAP-70 is composed of 14 exons which encode a 70kDa molecule comprised of three functional domains, two SH_2 domains arranged in tandem at the amino-terminus, and a tyrosine kinase domain at the carboxyl terminus. The two tandem N-terminal SH_2 domains precede an extended interdomain region which is followed by the PTK domain at

the C-terminal end of the molecule (Brdicka *et al.,* 2005). The interdomain A (IA) consists of a coiled-coil structure bringing the two SH_2 domains into close proximity thus creating a binding site for one of the phosphotyrosine residues within the ITAM motif. ITAM binding induces a conformational change in the structure of the interdomain A region.

The structure termed interdomain B (IB) follows the SH_2 domains and precedes the kinase domain of ZAP-70. This domain is important in regulating the kinase activity of ZAP-70 and contains a number of conserved phosphotyrosines. The activation loop contains the regulatory sites for this molecule. Those sites contain two important phosphotyrosines which are the sites for tyrosine phosphorylation by the lymphocyte specific kinase (Lck) (Jin *et al.,* 2004b). The glycine rich P-loop is known to be important for nucleotide and substrate binding, catalysis, and regulation of kinase activity. This loop forms the upper side of the ATP-binding cleft and is the catalytic domain of ZAP-70 (Zoller *et al.,* 1997).

Prior to 2006, ZAP-70 was considered to be associated only with T lymphocytes and natural killer cells. However Crespo *et al*. (2006) demonstrated that this molecule is also expressed in mice pro-B, pre-B and $CD19^+$ splenic B lymphocytes. It was subsequently confirmed that ZAP-70 is expressed in all developing subsets of B lymphocytes as well as in re-circulating B lymphocytes indicating that this molecule is also involved in the humoral immune response (Fallah-Arani *et al.,* 2008).

ZAP-70 participates in guiding signalling molecules towards the immunological synapse. This is necessary for the gathering of cell membrane molecules and their positioning in the synapse. The absence of ZAP-70 causes the microtubule network to become disoriented, and signalling molecules will not accumulate in the synaptic zone thus causing T lymphocytes to become non-functional (Blanchard *et al.,* 2002). ZAP-70 has been identified as an indispensable link in directing the signalling cascade towards the nucleus and affecting the gene transcription of the cytokine IL-2 and therefore T cell activation (Blanchard *et al.,* 2002). The identification of ZAP-70 in marsupials was undertaken in this study to provide an important step in the elucidation of the biochemical T cell signalling cascade.

1.5.4 Lymphocyte specific kinase (Lck)

The Lymphocyte specific kinase associates with the cytoplasmic tail of the cell surface receptors CD2, CD4, CD5, CD8, CD44, CD45 and CD122, and assists with the signal transduction from the T cell receptor (Isakov and Biesinger, 2000). It also associates with other protein kinases, the most important for this study being ZAP-70. Association of the TCR with peptide antigen-bound MHC complex facilitates the interaction of CD4 and CD8 with MHC class II and class I molecules respectively, and thereby recruits the associated Lck to the vicinity of the TCR/CD3 complex. By phosphorylating tyrosine residues within the ITAMs in the cytoplasmic tail of the CD3ε chains and other CD3 subunits, Lck initiates the TCR/CD3 signalling pathway (Briese and Willbold, 2003, Isakov and Biesinger, 2000). These interactions require the chelated metal zinc, which is critical for T cell development and activation (Kim *et al.,* 2003) as well as for binding the short cytoplasmic domains of CD4 and CD8α (Huse *et al.,* 1998, Turner *et al.,* 1990).

Human Lck is a 56kDa protein consisting of 13 exons and the N-terminal is myristoilated and palmitoylated to form a 'unique domain' which connects the molecule to the plasma membrane of the cell (Briese and Willbold, 2003). This molecule is a member of the src-type tyrosine kinase family and consists of three functional domains, an NH_2 terminal domain, a regulatory SH_2 domain, and an SH_3 domain. Both SH_2 and SH_3 domains are independently folded modules of 60-100 amino acid residues (Eck *et al.,* 1994).

Lck also has a kinase domain with a short C terminal tail (Eck *et al.,* 1994) and is involved in T cell and IL-2 receptor signalling (Briese and Willbold, 2003). The SH_3 domain binds to an internal proline rich region which modulates the enzymatic kinase activity in the early phase of the activation of the TCR signalling cascade (Romir *et al.,* 2007).

Preceding the SH_2 and SH_3 domains is a unique domain which serves as a membrane anchor and plays a role in the function and specificity of the SH_2 and SH_3 domains but is the least conserved domain (Carrera *et al.,* 1995).

The interaction of Lck with the CD4 or CD8 molecules occurs through the N-terminal region. In humans the amino acid residues 417 to 429 of CD4 and 10 to 23 of Lck are

important for the non-covalent interaction between the two molecules (Rudd *et al.*, 1988, Ravichandran and Burakoff, 1994). In humans, residues C^{420} and C^{422} of CD4 and residues 20 and 23 of Lck are critical for this interaction (Ravichandran and Burakoff, 1994). The conserved CxCP cysteine motif within the cytoplasmic tails of CD4 and CD8, and the CxxC motif of the Lck unique domain are crucial for the signal complex formation and allow the regulated release of CD4 for internalization (Huse *et al.,* 1998). Lck is an example of a metal- dependent co-folding of two protein sequences where a short polypeptide chain and a small less conserved protein domain can fold together to mediate protein interactions (Kim *et al.,* 2003). Further investigation in humans is required to understand this co-folding and exactly how the CD4-Lck complex is disrupted to allow CD4 internalization.

A partial Lck fragment is also annotated in the ensembl database but, prior to this study, no detailed investigation of this molecule in a marsupial had been conducted.

1.6 Cytokines and Forkhead box protein 3 (Foxp3)

1.6.1 Marsupial cytokine research

Cytokines are small secretory molecules that take care of intercellular communication and belong to the category of signalling molecules. These molecules are immune modulatory glycoproteins affecting changes in the immune system (Gooding, 1992).

Research into marsupial cytokines only began late last century to increase the number of cross-reacting reagents and to deduce the sophistication of the marsupial immune system. Tumor necrosis factor (TNF) was characterized in *T. vulpecula* (Wedlock *et al.,* 1996) and in *M. eugenii* (Harrison *et al.,* 1999). The sequence and secondary structure of this pro-inflammatory cytokine was compared to its human counterpart and it was found that there were distinct structural similarities between marsupials and eutherians. The secondary structure agreed with the human TNF secondary structure which suggested that protein conformation is conserved in both metatherian and eutherian molecules.

In humans, TNF and lymphotoxin are cytotoxic proteins with similar biological activities and which share a 30% amino acid homology. Lymphotoxin-α and β (LTα/β) are found on activated T_h1 cells. Both LT-α (Harrison and Deane, 2000) and LT-β were characterized in *M. eugenii* at the molecular level. All structurally important motifs were identified including the eutherian LT-α 3'UTR AU-rich region (Harrison and Deane, 2000). At the genomic level it was found that the gene organization of all three molecules was similar in marsupials compared to eutherian mammals (Harrison and Wedlock, 2000).

Interleukin-1β (IL-1β), another pro-inflammatory cytokine, was tested in *M. domestica* for serological cross reactivity with *H. sapiens* and *M. musculus* and yielded a negative result (Brozek and Ley, 1991). IL1-β was cloned and sequenced in *T. vulpecula* and a recombinant construct proved to be biologically active. Bovine IL-1β was subsequently tested on *T. vulpecula* but had no biological activity (Wedlock *et al.,* 1999).

The interferon family is secreted by a variety of cell types. The most prolific manufacturer of interferon type I (IFN-I) proteins is the plasmacytoid dendritic cell (McKenna *et al.,* 2005). Some family members of this cytokine have been sequenced in *M. eugenii* and the monotreme *Tachyglossus aculeatus* (short-beaked echidna) (Harrison *et al.,* 2003c). A number of IFN type I genes were identified in marsupials which is consistent with the number found in eutherians. However, only a few type I IFN genes were found in the monotremes indicating a long evolutionary distance between the two clades.

Interleukin-10 has been isolated from LPS stimulated and *M. bovis* infected *T. vulpecula* alveolar macrophages. Sequence identity of the putative amino acids was $\geq$ 50% compared with eutherians. Structurally important motifs in the marsupial IL-10 have been identified, and were found to be conserved in both marsupials and other mammals (Wedlock *et al.,* 1998). Interleukin-5 was identified in *M. eugenii* and *Sminthopsis macroura* (stripe-faced dunnart) with a sequence identity of 48 – 63% to eutherian mammals (Hawken *et al.,* 1999). A compound microsatellite was identified within the second intron of the *M. eugenii* IL-5 gene and was also found in *Wallabia bicolor* (swamp wallaby), *Dendrolagus bennettianus* (tree kangaroo), *S. macroura*, *M. domestica*, *T.*

vulpecula and *P. cinereus* making it a possible tool for the study of population genetics (Hawken *et al.,* 1999).

Leukemia Inhibitory Factor has been identified in *T. vulpecula* (Cui and Selwood, 2000), *Sminthopsis crassicaudata* (fat-tailed dunnart), and the monotreme *Tachyglossus aculeatus* (Australian Echidna) (Cui *et al.,* 2001). The functionally important motifs were conserved in these species.

So where were the important interleukins that drive the T_h1 and T_h2 response in the adaptive immune system? An attempt was made to show that interferon γ (IFNγ) and IL-4 exist in marsupials by trialling an anti-bovine IL-4 and IFNγ antibody to label these molecules in the lymphocytes of *P. cinereus* and *T. vulpecula* (Higgins *et al.,* 2004). It was determined by flow cytometry that PMA (phorbol 12-myristate 13-acetate) stimulated cells of *P. cinereus* and *T. vulpecula* both showed cross reactivity to an anti-bovine IFNγ, but no cross reactivity was observed for IL-4. The sequence of IL-4 was reported by Young (2010). This result may support the theory that marsupials lack the important T_h1 immune response, and possibly the T_h2 immune response (Higgins *et al*., 2004). Whether the cell mediated immunity of marsupials is functional or not remained unresolved.
Twenty three key immune genes in *M. domestica* were identified through data mining. This *in silico* identification of immune genes in that species inferred, for the first time, that the marsupial immune system is as sophisticated as that of other eutherians (Wong *et al.,* 2006).

1.6.1.1 Interleukin-2 (IL-2)

Interleukin-2 is a T cell growth factor and was first identified in humans in a cell culture of mixed lymphocytes and named blastogenic factor (BF) (Morgan *et al.,* 1976). A quantitative assay for T cell growth factor (TCGF) based on the ability of IL-2 to induce proliferation of T lymphocytes in culture was developed by (Smith *et al*., 1980b, Smith *et al*., 1980a). Interleukin-2 is produced by activated T lymphocytes and causes activated B lymphocytes and natural killer cells to proliferate.

Interleukin-2 has been characterized in many vertebrates. This cytokine belongs to a family of interleukins that bind the same receptor, the common gamma chain (γ_c). This cytokine is tissue and species specific and is usually only up-regulated in tissue from sick animals or mitogen stimulated tissues making it difficult to detect. Members of this family include IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21. To date only IL-4 in *M. eugenii* (Young, 2010) and IL-2 in *T. vulpecula* (presented as part of this study) (Young *et al.,* 2011) from this family has been characterized in marsupials. Preliminary characterizations of IL-15 and IL-21 were conducted in this study.

The avian IL-2 appears to have a length of 140 amino acids with a 21 amino acid long signal peptide (Zhou *et al.,* 2005). By comparison, the IL-2 of the marine mammals *D. leucas* and *Halichoerus grypus* (grey seal) has lengths of 155 amino acids and 156 amino acids respectively. Their identity to other mammals was >50% in both cases. In contrast, the avian IL-2 amino acid sequence has an identity of >60% to other birds and 30% identity to IL-15 in mammals (St-Laurent *et al.,* 1999), and 29% identity to *T. vulpecula*. Human IL-2 has a length of 153 amino acids, including the signal peptide, as was the case for the avian IL-2. Human IL-2 also has structural similarities to IL-15. The differences between the two cytokines are found primarily in the adaptive immune responses. While IL-2 is responsible for the maintenance of regulatory T lymphocytes, IL-15 is necessary for the support of surviving CD8 memory T lymphocytes (Waldmann *et al.,* 2001).

Mouse studies have shown that there is a decline in the proliferation of activated T lymphocytes in the pathogenesis of *M. avium* infection. This was shown to be due to the inability of T lymphocytes to produce or respond to IL-2 (Mannering and Cheers, 2002). The inability to isolate IL-2 in marsupials for such a long period of time led to the argument that these animals might not have the ability to fight mycobacterial infections. However, observations of captive marsupials by zoo keepers did not support that hypothesis. *Dasypus novemcinctus* (nine-banded armadillo) is the model species for the study of *M. leprae* infections. Prior to 2005 there were no species specific immunological reagents available for *D. novemcinctus* and marsupials. The isolation of IL-2 in *D. novemcinctus* advanced the species as a translational model (Adams *et al.,* 2005). One of

the aims of the present study was the characterization of the marsupial IL-2 and the design of a marsupial specific IL-2 antibody.

1.6.1.2 Interleukin-17 (IL-17)

In humans and mice, IL-17(A) is a 155 amino acid long disulphide-linked homodimeric glycoprotein with a molecular mass of 35kDa (Kolls and Linden, 2004). Each of the homodimeric chains has a weight of 15-20kDa. IL-17 contains a signal peptide spanning 23 amino acids which is followed by 123 amino acids, a characteristic of the IL-17 family. The IL-17 family members have four cysteine residues that form two disulphide bonds, a characteristic of the cysteine knot family (Yao *et al.,* 1995b). IL-17 bears no resemblance to any other known interleukins or structural domains.

The pro-inflammatory cytokine IL-17 induces differentiation and migration of neutrophils through the production of the cytokine IL-8. Protective immunity against pathogens, including mycobacteria, is provided by IL-17 (Matsuzaki and Umemura, 2007). IL-17's pro-inflammatory capacity induces the release of antimicrobial peptides, matrix metalloproteinase, chemokines, and cytokines which in turn affect the expansion of neutrophils (Xu and Cao, 2010). IL-17 is also expressed by γδ T lymphocytes and natural killer cells thus suggesting a link between the innate and adaptive immune systems (Korn *et al*., 2009).

1.6.1.3 Forkhead box protein 3 (Foxp3)

Human Foxp3 consists of 11 exons and is a specific marker of natural T regulatory cells (nT_{regs}) and adaptive or induced T regulatory T lymphocytes (iT_{regs}). T_{reg} cells expressing Foxp3 are critical in the transfer of immune self-tolerance (Shevach, 2000). Foxp3 contains a Forkhead domain (FKH) that mediates a direct interaction with the transcription factor NFAT (Wu *et al.,* 2006). T_{regs} express high levels of CD257, $CD26L^{+}$, $CTLA\text{-}4^{+}$, $GITR^{+}$, $ICOS^{+}$ and $CD127^{low}$ (Fontenot *et al.,* 2003). They are also hypo-responsive to TCR stimulation *in vitro* and fail to proliferate or produce activation-induced cytokines such as IL-2 or IFNγ (Fontenot *et al.,* 2003, Hori *et al.,* 2003, Schubert *et al.,* 2001). On the other hand, T_{regs} proliferate in response to homeostatic signals and then inhibit the

proliferation and cytokine production of naïve $CD4^+$ T lymphocytes (Fontenot *et al.,* 2005b).

Recent research has found that T_{regs} may be able to alter their function depending on the cytokine milieu. T_h17 cells are pro-inflammatory, are produced under similar conditions to iT_{regs}, and change under the influence of TGFβ and IL-6 (or IL-21) (Curotto de Lafaille and Lafaille, 2009). Evidence suggests that Foxp3 is restricted to αβ T lymphocytes and is linked to suppressor activity even without the expression of CD25 (Fontenot *et al.,* 2005b). Natural T_{regs} are $CD4^+CD25^+$ T lymphocytes that develop and migrate from the thymus to perform key roles in immune homeostasis (Takahashi *et al.,* 1998). Adaptive T_{regs} are non-regulatory $CD4^+$ T lymphocytes that acquire CD25 expression outside the thymus.

The immune-suppressive mechanism of T_{reg} cells is not presently understood. It has been found that IL-9, IL-10, and TGFβ are immune-suppressive soluble factors, while cell mediated regulation is controlled by the co-stimulatory molecules such as CTLA-4 (Curotto de Lafaille and Lafaille, 2009).

Prior to this study a T_{reg} population had not been identified in marsupials.

1.7 Aims and Objectives

The argument, prevalent in the literature that marsupials differ in their immunological character compared to other mammals is largely based on differences observed in their cellular immune responses. In this study a molecular characterization of the key molecules of the T cell signalling cascade which drives the cellular immune response was undertaken. This study tested the hypothesis that interleukin-2, the master regulator of the adaptive immune system, and the key signalling molecules are expressed in marsupials.

The specific aim of this study was to find evidence that the adaptive immune system of marsupials is in fact as competent as that of other mammals. The following steps were undertaken.

1. Characterize the sequences of the key signalling motifs of the T cell receptor α and β chains and predict a feasible structure of those motifs. Investigate the CD3ε subunit in the endangered marsupials *O. fraenata* and *L. hirsutus* by determining the sequence and structurally important motifs. Determine the expressed sequence of *M. domestica* CD3ε to confirm the predicted sequence.
2. Characterize the sequences of the co-receptors CD4 and CD8 and the co-stimulators CD28 and CTLA-4 in order to show that the mechanism of anergy and T cell activation is supported by the same molecules that drive this mechanism in other mammals. Determine feasible homology structures for all mature putative protein sequences by prediction modelling.
3. Characterize the sequence of the TCRζ, Lck and ZAP-70 signalling molecules of the T cell signalling cascade in marsupials. Trial human specific antibodies to Lck and ZAP-70 in order to investigate possible species cross reactivity in a marsupial. Develop a python shell for structure modelling of the putative ZAP-70 protein. Determine feasible homology structures for the TCRζ and Lck mature putative protein sequences by prediction modelling.
4. Characterize the sequence of the cytokine interleukin-2 in a marsupial and determine the putative homology structure by prediction modelling. Design a marsupial specific antibody to IL-2 and investigate its ability to identify the protein in a crude cell lysate. Determine the promoter region by data mining the ensembl database. Find evidence of a T_{reg} population in a marsupial by characterizing the sequence of the T_{reg} surface marker Foxp3, and find evidence of a T_h17 population by characterizing IL-17, the cytokine associated with that T cell sub-population.

Chapter 2

General Materials and Methods

Chapter 2 – General Materials and Methods

2.0 General Materials and Methods

There are two parts in this General Materials and Methods chapter. Part 1 describes the methodology for the molecular work, and Part 2 describes the methodology for the protein studies. These materials and methods are common to all experiments in the thesis. Methods that are specific to each of the molecules are stated in the Materials and Methods section of the relevant chapters.

2.1 Part I - Molecular Studies

2.1.1 Animal tissues

Tissues from four different species of marsupials, *M. domestica, O. fraenata, M. eugenii* and *L. hirsutus* were used in this study. *O. fraenata t*issues originated from two animals. The first animal was covered by animal ethics number 97042 from the Macquarie University Animal Care and Ethics Committee and gut nodes, mesenteric lymph nodes, liver and spleen tissues from that animal were donated by Dr. L. J. Young, CQUniversity Australia, Rockhampton, Queensland. The second animal was a young healthy male euthanized after a serious accident, and was obtained opportunistically from the Alma Street Veterinary Hospital, Rockhampton, Queensland.

The *M. eugenii* was covered by animal ethics number 97042 issued by the Macquarie University Animal Care and Ethics Committee, and spleen, mesenteric lymph node and thymus tissues from that animal were donated by Dr. L. J. Young.

The *L. hirsutus* was covered by animal ethics number 97042 issued by the Macquarie University Animal Care and Ethics Committee, and spleen and lymph node tissues from that animal were donated by Dr. L. J. Young.

2.1.2 Primer design

Degenerate consensus primers were designed for the most conserved regions within the genes which were investigated. Conservation was elucidated using the vertebrate sequences available in Genbank and the predicted *M. domestica* sequence from the UCSC databank (Kent *et al.,* 2002), prior to the publication of the wallaby genome. That genome was subsequently used as a tool for primer design whenever a gene annotation was available. Sequence conservation was detected by aligning mammalian amino acid (aa) and nucleotide (nt) sequences with the sequences from other vertebrates and then comparing them with the annotated *M. domestica* amino acid sequence in CLUSTALW2 using a Blossum62 substitution scoring matrix (Thompson *et al.,* 1994). Primers were subsequently designed from nucleotide alignments targeting identified, conserved regions of the annotated *M. domestica* sequence. Degenerate bases in three positions at either the 5' or the 3' end of the primers' oligonucleotide sequences were avoided to prevent non-specific binding (Pan *et al.,* 2007). Degenerate primers were analyzed using the OligoAnalyzer (IDT) tool from Integrated DNA Technologies (http.//www.idtdna.com/) and PCR Primer Stats (SMS) to confirm melt Temperatures and possible impeding secondary structures of the selected primers. Care was taken to exclude degenerate bases in the primer design, and melt Temperatures of primers were kept between 50°C and 60°C. These primers were then trialled on cDNA libraries made from *M. domestica* thymus and spleen.

2.1.3 Total RNA isolation by Tri-Reagent

O. fraenata spleen was treated with Tri-Reagent, homogenized, and centrifuged to inactivate nucleases. RNA was then extracted with chloroform. The supernatant was treated with isopropanol to precipitate the RNA. Subsequent wash steps with 95% and 75% ethanol and centrifugation at 4°C resulted in an RNA pellet. The pellet was air dried for 10 min and then re-suspended in 50 µL of nuclease-free water and stored at -20°C.

2.1.3.1 Total RNA isolation using SV Total RNA Isolation System™ (Promega, Madison, USA)

One mL of RNA Lysis Buffer was dispensed into a nuclease-free tube and weighed. Approximately 30 mg of tissue was added to the Buffer and manually homogenized as quickly as possible. The tube was weighed again and the tissue mass calculated by substraction. In accordance with the manufacturer's instruction, 175 µL of the tissue lysate was transferred to a 1.5 mL microcentrifuge tube. The remaining lysate was frozen at -20°C for later use. Three hundred and fifty microlitres (350 µL) of RNA Dilution Buffer was added to the tissue lysate and mixed by inverting four times. The lysate was incubated on a 70°C heat block for two and a half min. and then centrifuged for 10 min at 13,000 x g. After centrifugation, 200 µL of 95% ethanol was added and mixed by pipetting. The mix was transferred to a Spin Column Assembly and centrifuged at 13,000 x g for one min. The liquid in the collection tube was discarded, a wash solution was added to the spin basket, and the spin column assembly was centrifuged at 13,000 x g for one min. For every sample, a DNase incubation mix was prepared by mixing 40 µL of Core Buffer, 5 µL $MnCl_2$, and 5 µL of DNase I enzyme. The reaction was mixed by pipetting and kept on ice as per the manufacturer's instruction. The 50 µL of DNase treatment was added directly to the membrane of the spin column and incubated for 15 min at 23°C. Immediately after incubation, 200 µL of DNase Stop Solution was added to the spin basket and centrifuged at 13,000 x g for one min. Once centrifuged, 600 µL of RNA Wash Solution was added and again centrifuged at 13,000 x g for two min. The spin basket was transferred to a 1.5 mL Elution Tube and 100 µL of nuclease-free water was added to the membrane of the spin basket. The Elution Tube assembly was then centrifuged at 13,000 x g for one min. The spin basket was discarded and the purified RNA was stored at -20°C.

The total RNA was verified by an ethidium bromide gel and spectrophotometric measurement. Two microlitres (2 µL) of the total RNA were reverse transcribed and the resultant cDNA was used for subsequent Polymerase Chain Reactions.

2.1.3.2 mRNA isolation using FastTrack MAG

Messenger RNA (mRNA) was isolated from total RNA with a FastTrack® MAG mRNA isolation kit (Invitrogen, Carlsbad-USA) using oligo dT-conjugated magnetic beads to isolate the polyA$^+$ RNA from total RNA (Morrissey and Collins, 1989).

An aliquot of 100 μL of Binding Buffer B6 was added to the chilled total RNA and placed on a heat block at 68°C. RNase-free water equal to the volume of Binding Buffer B6 was added to the total RNA sample and the reaction was placed on ice. FastTrack® MAG Beads were mixed by pipetting up and down and 20 μL of the beads were added to an RNase-free tube. The tube was inserted into a magnetic separator for one min and the resulting liquid was discarded. This procedure was repeated three times. The tube was again inserted into the magnetic separator, the resulting liquid pipetted out and the total RNA sample and the heated Binding Buffer B6 was added. The tube was then removed from the magnetic separator and the beads were re-suspended in the solution by pipetting gently up and down. The sample was incubated at 68°C for three min, and then transferred to a rotator for 10 min at ambient temperature. The Eppendorf tube was again inserted into the magnetic separator until the beads were clearly separated (~one min) and the supernatant was saved. The wash buffer W7 was immediately added to the beads and the tube was re-inserted into the magnetic separator. When the beads were separated the wash buffer was removed and discarded, and the procedure was repeated three times. Ten microlitres (10 μL) of RNase-free water was added because the starting volume was less than 50 μg of total RNA. The tube was removed from the magnetic separator, incubated at 37°C for three min, re-inserted into the magnetic separator until the beads were clearly separated, removed again and the supernatant was saved. This procedure was repeated and the supernatant from both separations were combined to form the isolated mRNA. The yield was determined using a spectrophotometer and resulted in ~0.3% recovery of mRNA from total RNA [(mRNA yield (μg/μL) = A_{260} x 0.04 μg/μL RNA x Dilution factor)].

2.1.3.3 mRNA isolation with PolyATract® mRNA (Promega, Madison, USA)

Ten micrograms (10 µg) of total RNA was brought to a final volume of 500 µL in RNase-free water in a 1.5 mL Eppendorf tube. This was heated at 65°C for 10 min. Three microlitres (3 µL) of Biotinylated-Oligo (dT) probe and 13 µL of 20 X SSC were added to the heated RNA and gently mixed. The RNA was then incubated at ambient temperature until completely cooled (~seven min). A stock solution was prepared by combining 30 µL of 20 X SSC with 1.170 mL of RNase-free water resulting in the required 1.2 mL of sterile 0.5 X SSC solution. A second stock solution was prepared by combining seven microlitres (7 µL) of 20 X SSC with 1.393 mL of RNase-free water resulting in the required 1.4 mL of 0.1 X SSC.

The Streptavidin MagneSphere Paramagnetic Particles (SA-PMPs) were rinsed three times with equal volumes of 0.5 X SSC and used within the manufacturer specified timeframe of 30 min after washing. The particles were re-suspended, and clumped particles were discarded. One tube of SA-PMPs (1 mg/mL in PBS, 1 mg/mL BSA and 0.02% sodium azide) was re-suspended by flicking the bottom of the tube until complete dispersion was achieved. Capture took place by insertion of the tube in the magnetic stand until the SA-PMPs were collected at the side of the tube (30 s). The supernatant was removed and the SA-PMPs were washed three times with 0.5 X SSC (300 µL per wash), each time capturing them using the magnetic stand and carefully removing the supernatant. The washed SA-PMPs were re-suspended in 100 µL of 0.5 X SSC. The entire contents of the annealing reaction were added to the tube containing the washed SA-PMPs. The annealing reaction, together with the SA-PMPs, was incubated at ambient temperature for 10 min, gently mixing the tube by inverting every 1.5 min. The SA-PMPs were captured using the magnetic stand, and the supernatant removed without disturbing the SA-PMP pellet. The supernatant from this step was saved until the amount of mRNA was verified. The particles were then washed four times with 0.1 X SSC (300 µL per wash) by gently flicking the bottom of the tube until all the particles were re-suspended. After the final wash, as much supernatant as possible was removed without

disturbing the SA-PMPs. The final SA-PMP pellet was gently re-suspended in 100 µL of RNase-free water by flicking the tube. The SAPMPs were captured again magnetically and the supernatant containing the eluted mRNA was transferred to a new RNase-free tube. The elution step was repeated by re-suspending the SA-PMP pellet in 150 µL of RNase-free water.

The isolated mRNA was verified on an ethidium bromide gel. A spectrophotometric analysis was carried out for all mRNA isolations by measuring the absorbance ratio of A_{260}/A_{280} using a Qubit ®Fluorometer 1.0.

The mRNA was then reverse transcribed as described in section 2.1.5.

2.1.4 Qubit® 1.0 Fluorometer RNA quantitation

The reagents were prepared as per manufacturer's instructions. Dilutions of 1/10, 1/50, 1/100, 1/1,000 and 1/10,000 of the isolated RNA were prepared and read at 260nm, 280nm and 320nm absorbance to determine if any phenol contamination was present. The ratio of the A_{260}/A_{280} absorbance was obtained to indicate the purity of the RNA.

2.1.5 Reverse Transcription

cDNA was prepared from total RNA and mRNA. A commercially available Reverse Transcription Kit (Promega, Madison-USA) was used according to the manufacturer's instructions. In brief, RNA was incubated at 72°C to denature any secondary structures and then incubated with Avian Myeloblastosis Virus Reverse Transcriptase enzyme (AMVRT), ribonuclease (RNase) inhibitor, deoxynucleoside triphosphates (dNTPs), buffer, magnesium chloride ($MgCl_2$), oligodT primers and nuclease-free water at 42°C for 55 min, followed by a 5 min incubation at 95°C to inactivate the AMVRT enzyme. The resulting cDNA was stored at -20°C until required. A polymerase chain reaction was carried out with CD3ε primers to check the integrity of the cDNA.

2.1.5.1 Reverse Transcription using Superscript^{TM}II RT

The following components were added to a nuclease-free Eppendorf tube:
2 µL RNA or mRNA (~ 2ng) together with 1 µL Oligo(dT)$_{12-18}$ primers and 1 µL dNTP Mix (10mM each) were added to a nuclease-free eppendorf tube and made to 12 µL volume with nuclease-free water. The mixture was heated to 65°C for 5 min and then chilled on ice. The content of the tube was collected by brief centrifugation and 4 µL of 5 X First-Strand Buffer, 0.1 M DTT, and 1 µL of RNaseOUT^{TM} were added to the tube. The contents of the tube was mixed gently and incubated at 42°C for 2 min. One microliter (1 µL = 200 units) of SuperScript^{TM}II RT was added and mixed by pipetting gently up and down. The mixture was then incubated at 42°C for 50 min. The reaction was inactivated by heating at 70°C for 15 min. The reaction was then chilled on ice and 1 µL of RNaseH was added to the tube and incubated ate 37°C for 20 min and then stored at -20°C.

2.1.6 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

RT-PCR was performed to determine the molecular characteristics of each molecule. The PCR reaction contained 1.0 µL DNA, 10 x iTaq PCR buffer (Bio-Rad, Gladesville, Australia), a final concentration of 2 mM $MgCl_2$ (Bio-Rad, Gladesville, Australia), 0.2 mM dNTPs (Bio-Rad, Gladesville, Australia), 0.2 µM forward and reverse primers (Geneworks, Hindmarsh, Australia) and 0.1 U Taq polymerase - iTaq (Bio-Rad, Gladesville, Australia). This was brought to a 20 µL final volume with nuclease-free water. For the negative control, the DNA template was substituted with nuclease-free water. PCR reactions were performed using either a Bio-Rad iCycler (Bio-Rad, Gladesville, Australia) and/or a G-Storm (Geneworks, Hindmarsh, Australia) thermocycler.

2.1.7 Genomic DNA extraction

A Promega genomic DNA isolation kit (Promega, Madison, USA) was used to extract DNA from a *M. eugenii* thymus. One hundred and twenty microlitres (120µL) of 0.5 M EDTA solution was added to 500 µL of Nuclei Lysis Solution in a 1.5 mL Eppendorf centrifuge tube. The mixture was chilled on ice.

The thymus tissue was ground in liquid nitrogen and dispensed into a 1.5 mL Eppendorf tube. Six hundred microlitres (600 µL) of the EDTA/Nuclei Lysis Solution were added to the ground tissue. Seventeen and a half microlitres (17.5 µL) of 20 mg/mL Proteinase K was added in accordance with the manufacturer's instruction. The mixture was incubated at 55°C on a shaking platform for three h. The sample was vortexed once every hour.

Two hundred microlitres (200 µL) of Protein Precipitation Solution was added to the sample and the mixture was vortexed vigorously at high speed for 20 s. The sample was chilled on ice for 5 min and was then centrifuged for four min at 14,000 x g. After the protein precipitated the supernatant containing the DNA was removed and transferred into a clean 1.5 mL Eppendorf tube containing 600 µL of isopropanol at room temperature.

The solution was mixed by inversion until the white thread-like strands of DNA formed a visible mass. The mixture was then centrifuged for one min at 14,000 x g at ambient temperature. The DNA was visible as a small white pellet. The supernatant was decanted, 600 µL of 70% ethanol at ambient temperature was added and the tube was inverted several times to wash the DNA. The solution was again centrifuged for one min at 14,000 x g at ambient temperature.

The ethanol was aspirated and the tube was inverted on clean absorbent paper and air dried for 12 min. One hundred microlitres (100 µL) of DNA rehydration solution was added to the tube and the DNA was rehydrated by incubating at 65°C for 1 h. The solution was mixed periodically by gently tapping the tube as per the manufacturer's instruction.

For the second extraction, the DNA was incubated overnight at 4°C before storing the gDNA at 4°C.

2.1.8 Polymerase Chain Reaction

2.1.8.1 Cycling conditions

The temperature and time of each cycling step are set out in the following templates:

Template 1

Cycle Number	Temperature	Time	Cycling Step
1	94°C	2 min	Denaturation
35	94°C 50°C 72°C	30 s 50 s 1 min	Denaturation Annealing Extension
1	72°C	10 min	Final Extension
Final well Temperature 10°			

Template 2

Cycle Number	Temperature	Time	Cycling Step
1	94°C	2 min	Denaturation
5	94°C 65°C 72°C	30 s 50 s 1 min	Denaturation Annealing Extension
5	94°C	30 s	Denaturation
	60°C	50 s	Annealing
	72°C	1 min	Extension
30	94°C	30 s	Denaturation
	55°C	50 s	Annealing
	72°C	1 min	Extension
1	72°C	10 min	Final Extension
Final well Temperature 10°C			

Template 3

Cycle Number	Temperature	Time	Cycling Step
1	95°C	3 min	Denaturation
35	95°C 55°C 72°C	30 s 1.20 min 1.40 min	Denaturation Annealing Extension
1	72°C	20 min	Final Extension
Final well Temperature 10°C			

The above cycling conditions were adjusted according to factors that included, but were not limited to, the melt Temperature of the primers, composition of the nucleotide sequence, the length of the transcript, and the nature of the polymerase.

2.1.8.2 Gel electrophoresis

Gels were electrophoresed using a mini-sub Gel tank at 100 V for 45 min in 1 X Tris borate ethylenediaminetetra acetic acid (TBE) buffer. Amplicon sizes were estimated using a pGem marker (Promega, Madison, USA), Bio Lab 100 bp DNA ladder (Bio Lab, Australia) or the Bioline Hyperladder™ II (Bioline, Alaxandria, Australia). The Hyperladder was also used to estimate the quantity of DNA present in the samples for sequencing purposes.

All PCR products were visualized on a 1.5% agarose gel in 1 X TBE buffer containing a final concentration of 0.03 µg/mL ethidium bromide. Gels were visualized on a Bio-Rad GelDoc™ EQ system (Bio-Rad,Gladesville, Australia) and images were captured using the Quantity One software package (Bio-Rad, Segrate, Italy) and Vilber Lourmat Gel imaging system using a Quantum Capture software package (Fisher BioTec, Wembley, Australia).

Products of correct size were excised from the gels using a sterile No. 10 scalpel blade. The PCR products were separated from the agarose gel matrix using the Promega Wizard SV Gel and PCR clean up kit (Promega, Madison, USA) and prepared for sequencing.

2.1.8.3 DNA purification of PCR products from an agarose gel matrix

The DNA was purified from the agarose matrix with the Promega Wizard SV Gel and PCR clean up kit (Promega, Madison, USA). Excised gel slices were weighed and a Membrane Binding solution was added at the ratio of 10 µL of solution to 10 mg of agarose gel slice. The mixture was vortexed and incubated at 60°C for 10 min. The dissolved gel mixture was transferred into a SV Minicolumn assembly and subjected to two washes with a membrane wash solution. After the second wash, the SV Minicolumn was transferred to a clean 1.5mL micro centrifuge tube and the DNA eluted from the column with 15 µL – 30 µL of nuclease-free water (dependent on quantity of DNA present).

The clean-up procedure was verified by gel electrophoresis using a Hyperladder II (Bioline, Alexandria, Australia) for quantitation. The amplified DNA was sent for sequencing to the Australian Genome Research Facility (AGRF- http.//www.agrf.org.au/) in a 1.5 mL Eppendorf tube with the required concentration of 6 ng DNA template and 9.6 pmol of primer in 12 µL final volume of nuclease-free water. For the sequencing reaction, the ABI Big Dye Terminator Version 3.1 chemistry was used and analyses were undertaken on an AB3730x sequencer.

2.1.9 Rapid amplification of cDNA ends (RACE)

RACE cDNA was made from spleen and node tissue from *O. fraenata* and from Phythaemagglutinin (PHA) and ConcanavalinA (ConA) stimulated lymphocytes from *M. eugenii* using a GeneRacer™ kit (Invitrogen – Carlsbad, USA) in accordance with manufacturer's instructions. Seventy percent (70%) and 95% ethanol were prepared prior to starting the procedure. The first step in making RACE DNA is the dephosphorylation of the total RNA. A 10 µL reaction containing 5 µL total RNA, 10 X calf intestine phosphatase (CIP) buffer, RNAse Out, calf intestine phosphatase (CIP) and Diethylene Pyrocarbonate (DEPC) water was prepared on ice. The reaction was incubated at 50°C for 1 h after centrifugation and was placed on ice. This step was followed by an RNA precipitation by adding 90 µL of DEPC water and 100 µL of phenol-chloroform. After centrifugation, the top phase was transferred to a new 1.5 mL Eppendorf tube to which 2 µL of 10 mg/mL mussel glycogen, 10 µL of 3 M sodium acetate and 220 µL of 95% ethanol were added. The mixture was placed on dry ice for 10 min and subsequently centrifuged for 20 min at 4°C. The resultant supernatant was removed without disturbing the pellet. A volume of 500 µL of 70% ethanol was added to the pellet and centrifuged at maximum speed for 2 min at 4°C. The position of the pellet was noted and the ethanol removed by pipetting. Another centrifugation at 16,000 x g collected the remaining fluid, which was subsequently removed and the pellet was air dried for 1.5 min at ambient temperature. The pellet was then re-suspended in 7 µL of DEPC water.

The Cap Structure of the mRNA was then removed. A 10 µL reaction containing 7 µL of dephosphorylated RNA (from previous step), 1 µL of 10 X Tap buffer, 1 µL of RNaseOut and 1 µL of TAP was mixed and incubated at 37°C for 1 h. The enzyme was inhibited by placing the reaction on ice. Another RNA precipitation was performed in the same manner resulting in 7 µL of de-capped mRNA, and this was followed by ligating the RNA Oligo.

The 7 µL of dephosphorylated and decapped RNA was added to the pre-aliquot of lyophilized GeneRacer™ RNA Oligo, pipetted up and down and centrifuged briefly to collect the fluid in the bottom of the tube. This reaction was then incubated at 65°C to relax the secondary structure of the RNA, chilled on ice (~2 min), and briefly centrifuged. A solution containing a 10 X Ligase buffer (1 µL), 10 mM Adenosine Triphosphate (ATP) (1 µL), RNAseOut (1 µL) and the enzyme T4 RNA ligase (1 µL) was added to the 7 µL of RNA and incubated at 37°C for 1 h. The reaction was then briefly centrifuged and placed on ice to inhibit the enzyme activity and a third RNA precipitation was carried out with the pellet being re-suspended in 10 µL of DEPC water. The resultant RNA was then reverse transcribed using the AMV Reverse Transcription kit.

One microlitre (1 µL) of primer and 1 µL of the dNTP mix were added to the 10 µL of RNA and incubated at 65°C for 5 min. The reaction was then placed on ice for 2 min and briefly centrifuged. A solution containing 5 X RT buffer (4 µL), the enzyme AMV RT (1 µL), RNAseOUT (1 µL), and nuclease-free water (2 µL) was added to the 12 µL of ligated RNA and primer mix. This was well mixed and incubated at 45°C for 1 h and then incubated at 85°C for 15 min. This reaction was briefly centrifuged and stored at -20°C.

Dilutions of 1/10 and 1/50 were made from the obtained RACE cDNA and the validity of the cDNA was checked by RT- PCR with CD3ε primers. The amplicons from this PCR were visualized on a 1.5% agarose gel. The obtained product was excised using a sterile No. 10 scalpel blade and the DNA was purified using the Promega Wizard SV Gel and PCR clean-up kit (Promega, Madison, USA). The obtained product from the DNA purification was sent to the Australian Genome Research Facility (AGRF) for sequencing.

RACE primers, or gene-specific primers, were designed in order to perform RACE-PCR by applying the same principles as for RT-PCR primer design (Chapter 2, section 2.1.2) using the obtained gene sequence from RT-PCR.

The adaptor primers for the respective RACE ends are listed in Table 2.1.

Table 2.1. RACE primers supplied in Invitrogen RACE kit

Primer	Sequence 5'-3'	bases	T_m
GeneRacer™ 5' Primer	5' CGACTGGAGCACGAGGACACTGA-3'	23	74°C
GeneRacer™ 5' Nested Primer	5' GGACACTGACATGGACTGAAGGAGTA-3'	26	78°C
GeneRacer™ 3' Primer	5'-GCTGTCAACGATACGCTACGTAACG-3'	25	76°C
GeneRacer™ 3' Nested Primer	5'-CGCTACGTAACGGCATGACAGTG-3'	23	72°C

The RACE DNA was used to amplify the 5' and 3' ends of the genes. The 5' regions were investigated for start sequences (atg) as well as for leader signals and other functionally important motifs. The 3' ends of the genes were investigated for the stop sequences (tga, taa), RNA retention motifs, polyadenylation signals and polyA tails.

2.1.9.1 Cycling conditions for RACE-PCR

Template 1

Cycle Number	Temperature	Time	Cycling Step
1	94°C	2 min	Denaturation
35	94°C 50°C 68°C	30 s 50 s 1 min	Denaturation Annealing Extension
1	68°C	10 min	Final Extension
Final well Temperature 10°C			

Template 2

Cycle Number	Temperature	Time	Cycling Step
1	94°C	2 min	Denaturation
35	94°C 50°C 68°C	30 s 50 s 1 min	Denaturation Annealing Extension
1	68°C	10 min	Final Extension
Final well Temperature 10°C			

2.1.9.2 Rapid amplification of 3' cDNA ends

Total RNA (1-5 µg) or mRNA (50 ng) and DEPC water were combined to give a final volume of 11 µL in a 0.5 mL Eppendorf tube. Four microlitres (4 µL) of RNA in 7 µL of DEPC water was used to make 3' RACE from *M. eugenii* ConA and PHA stimulated lymph cells. One microlitre (1 µL) of 10 µM ATP solution was added and the reaction was then mixed and collected at the bottom of the tube by brief centrifugation. The reaction was heated to 70°C for 10 min and chilled on ice for at least 1 min. The content of the tube was again centrifuged and a solution containing a 10 X PCR buffer (2 µL), 25 mM $MgCl_2$ (2µL), 10 mM dNTPs (1 µL) and 0.1 M DTT (2 µL) was added. The reaction was gently mixed and centrifuged. The mixture was equilibrated to 42°C for 3 min. One microlitre (1 µL) of Super Script II RT was added and the reaction was incubated at 42°C in a heat block for 50 min. The enzyme was inactivated by incubating the reaction at 70°C for 15 min. The reaction was again chilled on ice then briefly centrifuged to collect the contents of the tube. One microlitre (1 µL) of RNaseH was added to the contents which were then mixed and incubated for 20 min at 37°C before storing the reaction at -20°C. A 1/20 dilution of the resultant RACE DNA was verified using CD3eUTF and the 3' Universal Amplification Primer (UAP) (5'CUACUACUACUAGGCCACGCGTCGACTAG TAC-3') supplied by the kit (3' RACE kit invitrogen, Carlsbad, USA). The Abridged Universal Amplification Primer (AUAP) (5'-GGCCACGCGTCGACTAGTAC-3') was used for a nested strategy.

2.1.10 Cloning of PCR and RACE-PCR products

2.1.10.1 Media (LB Agar)

Nutrient agar was prepared by adding 2 g of Bacto-Tryptone (Sigma, Australia), 1 g of Bacto-Yeast (Sigma, Australia) and 2 g of Sodium Chloride (NaCl) (Astral, Australia) to 200 mL of dIH_2O. Once the ingredients were dissolved, 3 g of Bacto-Agar (Sigma, Australia) was added and the mixture was autoclaved. After cooling to 55°C, 1 mL of 10 mg/mL kanamycin (Invitrogen, Carlsbad, USA) was added aseptically and the media was dispensed into standard Petri dishes.

2.1.10.2 Luria Bertani Broth (LBroth)

LBroth was prepared by mixing 20 g of LB (Sigma) and 1.5 g of Bacteriological Agar (Sigma, Australia) with 1 L dIH_2O. This was heated in a microwave oven to dissolve the agar, dispensed into 100 mL Schott bottles, autoclaved, and subsequently stored at 4°C. As required, 10 mL aliquots of the broth were warmed to 37°C in 30 mL McCartney bottles and 500 µL of 10 mg/mL kanamycin (Invitrogen, Carlsbad, USA) was added to each aliquot before inoculation with *Escherichia coli*.

2.1.10.3 Transformation

Purified PCR products as described in section 2.1.8.3 were used for insertion into the pCR™4-TOPO® TA vector purchased from Invitrogen, Carlsbad, USA and executed in accordance with the manufacturer's instructions. The reaction mix was incubated for 5 min at room temperature (25°C) and immediately put on ice. Transformation of the *E. coli* cells was carried out by using 2 µL of the cloning reaction which was incubated with the competent cells for 10 min. The competent cells were heat-shocked for 30 s at 42°C and placed on ice. Two hundred and fifty microlitres (250 µL) of super optimal catabolite repression (S.O.C.) medium were added to the cells, the tubes were capped tightly, and incubated on a shaker incubator at 200 rpms for 1 h at 37°C. Aliquots of 20 µL and 50 µL of the transformed cell mixture were spread together with 20 µL of S.O.C medium on

media plates amended with 500 µL of 10 mg/mL of kanamycin. The plates were incubated overnight at 37°C.

One colony was selected from each of the 20 µL and 50 µL spread plates and put into 10 mL of LBroth containing 50 µg/mL of kanamycin. The inoculated LBroth was then incubated on a shaker incubator at 150 rpm overnight at 37°C.

The resulting plasmids were prepared for purification using a plasmid DNA preparation SV Wizard DNA clean-up kit (Promega, Maddison, USA).

2.1.10.4 Plasmid DNA preparation

The Promega Wizard Plus Miniprep DNA purification system (Promega, Madison, USA) was used in accordance with the manufacturer's instructions. One milliliter (1 mL) aliquots of overnight culture were spun sequentially in the same tube at 10,000 x g for 5 min at ambient temperature and the supernatant was discarded after each spin. Two hundred and fifty microlitres (250 µL) of Cell Resuspension Solution was added and the cells were completely re-suspended by vortexing or pipetting. Two hundred and fifty microlitres (250 µL) of Cell Lysis Solution was added and mixed by inverting four times until the cell suspension cleared (~4 min). Ten microlitres (10 µL) of Alkaline Phosphatase was added and mixed by inverting the tube four times. The mixture was incubated for 3 min, and 350 µL of Neutralization Solution was then added and mixed immediately by inverting the tube four times. The bacterial lysate was then centrifuged at top speed for 10 min at ambient temperature.

The cleared lysate was transferred to a spin column assembly, centrifuged and washed twice using a wash solution. The spin column assembly was centrifuged again in order to dry any carry-over ethanol before eluting the plasmid DNA in 100 µL of nuclease-free water.

2.1.10.5 Plasmid digest

Five microlitres (5 µL) of DNA was incubated for 2 h at 37 °C with Buffer H, 5% EcoRI and nuclease free water in a 20 µL reaction mix. The digested product was visualized on a 2% agarose gel on a Vilber Lourmat Gel imaging system using the Quantum Capture software package (Fisher BioTec, Wembley, Australia).

2.1.10.6 Sequencing of PCR products and plasmid DNA

Sequencing reactions contained 0.8 pmol/µL of primer in a 12 µL reaction volume. The amount of cDNA was determined by investigating the verification gel for clean-up products run against a quantitative marker (Hyperladder II, Bioline, Australia) for the respective genes. Eight nanograms (8 ng) (if product size was 100-200 bp) or 75 ng (if product size was >800bp) of PCR products, and approximately 1000 ng of plasmid DNA were sent to AGRF at the University of Queensland, St. Lucia, Brisbane for sequencing. Sequencing was carried out with the Big Dye Terminator Chemistry Version 3.1 on the AB3730x sequence platform.

2.2 Bioinformatics tools

2.2.1 CLUSTALW2

CLUSTALW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) was used to show where the obtained gene sequence aligned compared to the sequences of other species. Some sequences needed to be aligned manually using the JAVA applet supplied by CLUSTALW. The Jalview function allowed the alignment of sequence fragments to their proper positions. For amino acid residue alignments a BLOSUM62 (BLOcks of Amino Acid SUbstitution Matrix) matrix, which is a scoring alignment between evolutionary divergent protein sequences, was employed and default gap penalties were used (Henikoff and Henikoff, 1992).

2.2.2 The Basic Local Alignment Search Tools (BLAST and BLAT)

The Basic Local Alignment Search Tools (http://blast.ncbi.nlm.nih.gov/) (Altschul and Gish, 1996, Altschul *et al.,* 1990) BLASTn, BLASTx, BLAST2 and BLAT were used to investigate the sequences obtained from RT-PCR and RACE-PCR.

2.2.2.1 BLASTn

BLASTn is a nucleotide query and determines short matches between the sequences by heuristic methods. After the seeding was complete, BLAST constructs local alignments, which were used to investigate low complexity sequences indicated by lower case letters.

2.2.2.2 BLASTx

This function compares translational products of the nucleotide query sequence to a protein database. BLASTx translates in all six reading frames showing significant statistics for hits to different frames. This program was used to identify the correct reading frame of novel nucleotide sequences and was verified using EXPASY's translate function.

2.2.2.3 BLAST2

BLAST2 was used for the comparison of novel sequences not yet deposited in Genbank. Similarly to BLASTn, one or more sequences can be compared to one another and the e-values for the queried sequences were calculated.

2.2.2.4 BLAT

The BLAT function (http://genome.ucsc.edu/) (Kent *et al*., 2002) can be found in the UCSC genome browser and maintains an index of the entire genome in memory. BLAT is designed to find sequences containing at least 25 bases with 95% or greater length similarity.

2.2.3 Expert Protein Analysis Systems (EXPASY)

The scientific software tools in the Swiss Institute of Bioinformatics database (http://web.expasy.org/) which were used in this study are detailed in sub-sections 2.2.3.1 to 2.2.3.10.

2.2.3.1 Translate tool

The translate tool (http://web.expasy.org/translate/) translates nucleotide (DNA/RNA) sequences to a putative protein sequence in six reading frames and was used to verify the results from BLASTx.

2.2.3.2 ProtParam

This tool (http://web.expasy.org/protparam/) was used to compute physical and chemical parameters for a given peptide sequence in order to understand the hydropathy pattern of the peptide (Moller *et al.,* 1999). This was especially useful for the design of the interleukin-2 antibody sequence. This tool was used to ascertain the molecular weight, theoretical pI, amino acid composition, atomic composition and extinction coefficient (Bjellqvist *et al.,* 1993, Geer *et al.,* 2002).

2.2.3.3 Signal IP-4.0

The Signal IP-4.0 program (http.//www.cbs.dtu.dk/services/SignalP/) is part of the Center for Biological sequence analysis suite sequences (Bendtsen *et al.,* 2004, Petersen *et al.,* 2011). This tool was used to predict the presence and location of signal peptide cleavage sites in amino acid.

2.2.3.4 Protein Domain database (ProDom)

ProDom (http.//prodom.prabi.fr/prodom/) is a protein domain family database available through the EXPASY server. This database and software was used to search for protein domain families for comparison with the results obtained from SMART (see sub-section 2.2.4).

2.2.3.5 C-mannosylation

NetCGlyc 1.0 (http.//www.cbs.dtu.dk/services/NetCGlyc/) was used for the prediction of C-mannosylation characterized by the attachment of an alpha-mannopyranose to a tryptophan (WxxWW) via a cysteine-cysteine link in mammalian proteins. NetCGlyc 1.0 predicts, with 93% accuracy, positive and negative C-mannosylation sites (Julenius, 2007). All sequences obtained in this study were inspected for C-mannosylation sites.

2.2.3.6 O-linked glycosylation

The NetOglyc server (http.//www.cbs.dtu.dk/services/NetCGlyc/) (Julenius, 2007) was used to predict mucin type GalNac O-glycosylation sites (glycosylated serine or threonine residues) found in coil or turn regions in the putative amino acid sequences obtained in this study.

2.2.3.7 Protein glycation

The NetGlycate 1.0 server (http.//www.cbs.dtu.dk/services/NetGlycate/) (Johansen *et al.,* 2006) was used to predict glycation of ε amino groups of lysines in the putative amino acid sequences obtained in this study.

2.2.3.8 N-linked glycosylation

This tool (http.//www.cbs.dtu.dk/services/NetNGlyc/) was used to show the consensus sequence for an N-linked glycosylation (N-Xaa-S/T where Xaa is not P, N = asparagine, S = serine, T = threonine, P = proline) event in the putative amino acid sequences in this study. Glycosylation is an important post-translational modification which influences protein folding, protein solubility, antigenicity, biological activity as well as cell-cell interactions (Gupta *et al.,* 2004).

2.2.3.9 Phosphorylation prediction

The NetPhos server (http.//www.cbs.dtu.dk/services/NetPhos/) (Blom *et al.,* 1999) was used for predicting putative phosphorylation sites at serine, threonine and/or tyrosine

residues in the putative amino acid sequences obtained in this study. Putative phosphorylation and kinase phosphorylations were investigated with the NetPhos and NetPhosK programs. NetPhosK detected kinase specific protein phosphorylation sites in the putative amino acid sequences obtained in this study.

2.2.3.10 Disulphide bond prediction

Significant disulfide bonds were predicted using the DiANNA disulphide bond prediction server (http://clavius.bc.edu/~clotelab/DiANNA/) (Ferre and Clote, 2005).

2.2.4 SMART (**S**imple **M**odular **A**rchitecture **R**esearch **T**ool)

This search tool (http.//smart.embl-heidelberg.de/) (Letunic *et al.,* 2012) was used for the identification and annotation of genetically mobile domains, the analysis of domain architecture, internal repeats, and protein family domains (PFAM). The domains were annotated with respect to phyletic distribution, functional class, tertiary structures and functionally important residues.

2.2.4.1 CDART (*Conserved* **D**omain **A**rchitecture **R**etrieval **T**ool)

CDART is part of the National Centre for Biotechnology Information (NCBI) database (http.//www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi) and displays the functional domains that make up a given protein sequence. It also lists proteins with similar domain architectures and retrieves proteins that contain particular combinations of domains (Geer *et al.,* 2002). This tool was used to verify the SMART results when domains returned low e-value readings with the SMART tool. It was also employed to determine domains such as SH_2, SH_3 or Ig-like domains from the putative protein sequences obtained in this study.

2.2.5 Primary structure prediction

The primary structure, which is the putative amino acid sequence, is the basis for a secondary structure prediction that calculates the positions of loops and helices.

2.2.5.1 2ZIP-Leucine Zipper prediction

The primary structure prediction program 2ZIP (http.//2zip.molgen.mpg.de/)was used to predict Leucine zippers which are dimerization domains occurring mostly in regulatory proteins (Bornberg-Bauer *et al.,* 1998).

2.2.6 Secondary structure prediction

A secondary structure prediction aims to predict the local secondary structure of proteins based on the putative amino acid sequence. The prediction programs assign alpha helices, beta strands and/or beta turns to regions of the amino acid sequences. The secondary structure also refers to the interactions that occur between the carbon, oxygen and amine (NH) groups on amino acids in a polypeptide chain. The tools used are detailed in sub-sections 2.2.6.1 and 2.2.6.2.

2.2.6.1 Protein Structure Prediction Server (PSIPRED)

The Protein Structure Prediction Server (http.//bioinf.cs.ucl.ac.uk/psipred/) was used to predict alpha helices, coiled coils and beta strands by assigning numerical values to determine the probability of either structures occurring (Bryson *et al.,* 2005). This server also predicted transmembrane topology and protein fold recognition. The MEMSAT2 program on the server was used to determine transmembrane helix topology. The accuracy of the method is ~80% for predicting the topology of all-helical transmembrane proteins (Bryson *et al.,* 2005).

2.2.6.2 Secondary structure prediction through NetSurfP

This program (http://www.cbs.dtu.dk/services/NetSurfP/) gives probability values for the different structures such as alpha helices, beta strands and coils (Petersen *et al.,* 2009), and these values were used for comparison with the results obtained by the PSIPRED program.

2.2.7 Tertiary structure prediction (Homology modelling)

The tools used for homology modelling are detailed in sub-sections 2.2.7.1 to 2.2.7.8.

2.2.7.1 I-TASSER

The I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used to obtain feasible models for the molecules where the full open reading frame was obtained. Some of the models presented in this thesis were modelled a*b initio* because no templates were available in the database. 3D models in the I-TASSER server are built on multiple-threading alignments by the Local Meta-Threading-Server (LOMETS) and iterative TASSER assembly simulations (Roy *et al.,* 2010, Wu and Zhang, 2010).

2.2.7.2 RaptorX

RaptorX (http.//raptorx.uchicago.edu/about/) is a structure prediction program that aligns distantly related proteins that have limited sequence profiles. This program consists of four modules - single-template threading, alignment quality assessment, multiple-template threading and the fragment-free approach (Peng and Xu, 2011). RaptorX was used to model some of the structural motifs found within the putative amino acids sequences obtained in this thesis.

2.2.7.3 3D Jigsaw

This server (http.//bmm.cancerresearchuk.org/~3djigsaw/) builds 3D structures using as a template identifier in the HMM program (Soding, 2005). This program is a comparative modelling program where gaps and missing amino acid residues are closed and all models are finally recombined (Bates *et al.,* 2001). A genetic algorithm (GA) is at the heart of this program. GA's are well known for their powerful optimization techniques, and have been used in a number of protein modelling efforts ranging from *ab initio* folding models to model-building by homology (Offman *et al.,* 2006). After 10 rounds of conversion, all models are ranked using the highest sequence identifier and the top five models are returned. The models can be selected either in auto mode or in interactive mode. The

latter allows the user to select the models according to their coverage, sequence identification and domains. Both methods were used in this study depending upon which was the most appropriate in the circumstances. This program was also used to verify the the architecture of structural motifs obtained using RaptorX.

2.2.7.4 Modweb

Modweb (http.//modbase.compbio.ucsf.edu/) is a server for comparative protein structure modelling and was used in this study on all putative amino acid sequences that encompassed the full open reading frame. Modweb depends on Modpipe for its functionality (Eswar *et al.,* 2003). The structural templates used to build the models are extracted from the pdb database. Significant alignment (e-value better than 1.0) covering at least 30 amino acid residues are selected for modelling. Comparative modelling is achieved by satisfying the spatial restraints similar to the Modeller program described sub-section 2.2.7.6. The resulting models were evaluated by assigning scores and the highest scoring models were used in this study. Model selection is based on a normalized discrete optimized protein energy (DOPE) score as well as on the alignment with the best sequence identity.

2.2.7.5 Swiss-Model

The SWISS-MODEL program (http.//swissmodel.expasy.org/) is a web based integrated service committed to protein structure homology modelling and is available from the EXPASY server. Several functions are available. A fully automated model is suited for cases where the target template is highly conserved (>50% sequence identity) and this model was used throughout this thesis. An interactive model is used when the sequence identity is less than 50%. This was employed in this thesis for the TCRα partial sequence in order to confirm the position of some important residues in an Ig-fold predicted with RaptorX.

Model quality evaluation is determined by the QMEAN and Density plots. The estimated residue error is visualized using a colour gradient from blue to red (Fig. 2.1a). Blue

indicates more reliable regions and red are potentially unreliable regions. The Z-score is normalized to a mean of 0 and a standard deviation of 1. The Z-score therefore directly indicates how many standard deviations the model's QMEAN score differs from expected values (Fig. 2.1 a, b). Models of low quality are expected to have strong negative Z-scores for QMEAN. Large negative values correspond to red regions while positive values correspond to blue regions and therefore to homologous structures.

Density plots were calculated for all reference models. The background distribution is marked in red. This plot is a 'projection' of the model quality diagram. The number of reference models used in the calculation is shown at the bottom of the plot under the x-axis (Fig. 2.1c).

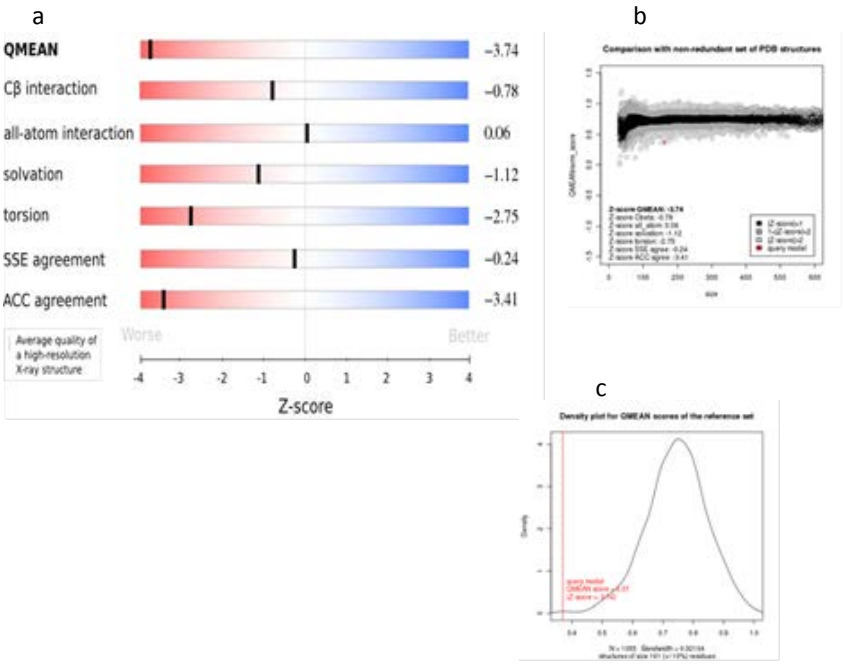


Figure 2.1. Example of statistical representations of models developed by the SWISS-MODEL program. (a) Z-score slider indicating QMEAN, (b) comparison of target model with pdb database, (c) density plot.

The molecular graphics viewer Jmol (http.//jmol.sourceforge.net/) (Hanson *et al.,* 2010) was used repeatedly to investigate the predicted structures and the different plots generated by the SWISS-MODEL program.

2.2.7.6 Modeller

Modeller (http.//salilab.org/modeller/) creates three-dimensional protein structures for comparative protein structure modelling. This program was used in this thesis because it allows the user to input the alignment of sequences to be modelled with known related structures. The output is a model 3D structure that satisfies the spatial arrangement of the amino acid sequences. These can operate on distances, angles, dihedral angles, pairs of dihedral angles and other spatial features defined by atoms or pseudo atoms. These values are taken from known related structures and their alignment with the target sequence (Eswar *et al.,* 2006, Sali and Blundell, 1993). In this thesis, a Python shell was specifically created to obtain a custom model which was than compared to other models. This is fully described in Chapter 5, section 5.2.

2.2.7.7 The **P**rotein **H**omology/analog**Y** **R**ecognition **E**ngine V 2.0 (Phyre2)

Phyre2 (http.//www.sbg.bio.ic.ac.uk/phyre2/)(Kelley and Sternberg, 2009) was also used for homology modelling of putative amino acid sequences in this study. This software aligns two sequences of interest and uses that alignment to generate a homologus 3D model.

Phyre2 uses a library of known protein structures taken from the Structural Classification of Proteins (SCOP) database. This database can be expanded with structures that were deposited later in the Protein Data Bank (pdb) (Berman *et al.,* 2000, Murzin *et al.,* 1995). Three secondary structure prediction programs are used in Phyre2-Psi-Pred, SSPro and JNet (Cole *et al.,* 2008, Pollastri *et al.,* 2002, McGuffin *et al.,* 2000). This makes Phyre2 a very comprehensive modelling program that was used to compare structures determined by I-TASSER and Modweb.

2.2.7.8 Ligand binding site prediction

The identification of ligand-binding sites is important because proteins perform their function on ligands, and ligands in turn regulate the protein. Protein structures are mostly

solved in the absence of ligands. It is therefore important to identify amino acids that are capable of binding ligands and which consequently have a biological effect. This program was used throughout this study to determine whether or not the putative amino acids are able to bind to other structures and are thus potentially functional proteins.

The Imperial College London's 3DLigandSite-Ligand binding site prediction Server (Wass *et al.,* 2010) (linked from Phyre2) was used to map ligand sites on predicted structures. The University of Michigan's I-TASSER server also has a ligand binding site prediction server. Both programs were used to predict the ligand binding sites and, where possible, the predictions were compared with each other.

2.3 Phylogenetic Analyses

Phylogenetic analyses were carried out using the Mega5 program (Tamura *et al.,* 2011). Two types of phylogenetic trees were generated; a Maximum likelihood phylogenetic tree and a Neighbor-Joining tree. Both trees were based on the Poisson distribution model which is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space (Gullberg, 1997).

2.3.1 Maximum likelihood tree

Amino acid sequences were collected in FASTA formats from Genbank and deposited in MEGA5. The CLUSTALW function within the MEGA5 program was used to align the sequences which were used in the determination of a Maximum likelihood tree.

The tree with the highest likelihood has the lowest negative logarithmic-transformed values. Likelihood is regarded as statistically consistent. Reliability in an inferred tree is achieved using a bootstrap test which is a re-sampling technique where trees are reconstructed by repeat sampling numerous times. If the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered to be correct (Felsenstein, 1985). In this study a 1,000 bootstrap test was applied.

2.3.2 Neighbor-Joining tree

Amino acid residues were collected in FASTA format from Genbank and deposited in MEGA5. A CLUSTALW alignment was carried out within the MEGA5 suite of programs and 1,000 bootstraps were applied for the purpose of stringency when selecting the Neighbor-Joining program.

The Neighbor-Joining technique is a polynomial-time algorithm suited to large datasets and bootstrapping. Neighbor-Joining is a bottom-up clustering method which uses the distance between each pair of taxa (Didelot, 2010, Saitou and Nei, 1987). The Dayhoff algorithm described in section 2.3.2.1 was used where sequences were identical.

2.3.2.1 Dayhoff algorithm

Where identical sequences were included, such as in the comparison of functional or structural important motifs, a Dayhoff algorithm (Dayhoff *et al.,* 1978) contained within the MEGA5 program was used. Gamma distribution values produced by this program were useful in this study.

2.4 Genomic DNA analyses

2.4.1 Promoter Scan

Promoter Scan (http.//www-mas.cit.nih.gov/molbio/proscan/analysis.html) is designed to find putative eukaryotic PolII promoter sequences in primary sequence data. This program was used to identify the IL-2 gene promoter region in the relevant ensembl genome database. For a sequence to qualify as a promoter region, certain transcriptional elements have to be contained within that sequence (Prestridge, 1995). A cut-off score of 53% was used in this study to identify a promoter sequence. Promoter Scan reported the sequence range in which the putative promoter was found. The TATA box was determined and estimation of the transcription Start Site (TSS) was made. Significant signals, such as transcriptional elements, were also reported. The results from Promoter

Scan were verified with Genscan in order to find the START site of the gene as well as the TATA box.

2.4.2 Genscan

Genscan (http.//genes.mit.edu/GENSCAN.html) predicts gene structures in vertebrate genomic sequences. This program was used in this study to detect the boundary between exon-3 and exon-4 in the IL-2 gene. This allowed a comparison of the human exon/intron boundaries with the marsupial exon/intron boundaries.

2.5 Ensembl and UCSC Genome Browsers

The wallaby genome had not been published when this study commenced. It was subsequently deposited in the ensembl database by Renfree *et al.* (2011). The *M. domestica* genome (Mikkelsen *et al.,* 2007) is also in ensembl although this genome was originally published in the UCSC genome browser (http.//genome.ucsc.edu/) (Dreszer *et al*., 2012, Fujita *et al*., 2011, Kent, 2002). The *M. domestica* genome was used to search for conservation of genes between marsupials and other vertebrates. With this information, consensus primers were designed in this study to permit identification of the gene in question at the molecular level. *M. domestica* is not closely related to the other species investigated in this study, hence the use of the consensus method. Following its publication, the wallaby genome was used in this study for comparison with the expressed gene sequences determined from RNA and mRNA in the species investigated.

2.6 Semi-quantitative Expression studies

A Nanodrop 2000c spectrophotometer (Thermo scientific, Australia) was used to quantify the cDNA obtained from RNA and mRNA. cDNA from *O. fraenata* and *M. eugenii* tissues was normalized against the housekeeping gene Glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

2.7 Real Time PCR – Quantitative PCR (qPCR) for Interleukin-2

2.7.1 Primer design

The primers were selected by inspecting a sequence alignment and using the secondary structure prediction to avoid unsuitable areas. The secondary structure of the primer pair was inspected with OligoanalyzerdT for heterodimer and hairpin formations. The location of the primer pair used in the HRM experiment was located in exon-3 (tcactaaagaactg aaccctgtggc/T_m 58.9°C) and exon-4 (gtcactacgcctccaaga tgagg/T_m 59.1°C) spanning 152 bp. The gene interleukin-2 was reported to have a splice variant where exon-2 was missing. The primers were therefore positioned outside that region.

2.7.1.1 Real Time Polymerase Chain Reaction (qPCR)

The qPCR protocol must start with an initial incubation step of 5 min at 95°C to activate the HotStar Taq Plus DNA Polymerase. The tubes containing the 25 µL reaction volume were loaded into a 36 tube rotor on a Rotorgene Q thermocycler (Qiagen, Doncaster, Vic, Australia) ensuring that the negative control was loaded into the rotor after the samples had been placed. The fluorescence data acquisition commenced at the combined annealing/extension step after the initial denaturation. Care was taken to select the correct fluorescence channel (green for SYBR green). Reactions containing 2.5 µL of cDNA and a HRM SYBR green Master mix were brought to a final volume of 25 µL with nuclease free water. Primers were used at a final concentration of 0.2 µM. The specificity of the amplicon for each reaction was verified by examination of the corresponding dissociation curve and visualization on a 1% agarose gel matrix. For all reactions, a denaturation cycle of 5 min at 95°C, 40 annealing cycles of 15 s at 95°C, 15 s at 50°C and 1 min at 72°C, and one extension cycle of 10 min at 72°C were carried out. Data acquisitions were performed with the Rotorgene Q software.

2.7.2 **H**igh **R**esolution **M**elt analysis (HRM)

SYBR Green I technology was utilized for all quantitative PCR reactions using the HRM kit (Qiagen, Doncaster, Australia). Baseline limits were set as suggested by manufacturer. The threshold was set to lie in the middle of the exponential phase of the amplification plot so that efficiency values truly reflected the reaction dynamics at the C_t. All efficiency values were determined with the Rotorgene Q analysis software. Data resulting from reactions that did not reach the threshold within the first 40 cycles (C_t = 40) were discarded from the analysis. cDNA dilutions of 1/100 and 1/1,000 were made and a qPCR run was set up in triplicate and the samples were normalized against the housekeeping gene GAPDH.

2.8 Part II Protein studies

Western Blots were carried out for ZAP-70, Lck and IL-2 to ascertain whether or not the protein could be identified in a crude cell lysate. The procedures are described in the following sub-sections 2.8.1 to 2.8.4.

2.8.1 Protein Extraction

The animal tissue was manually homogenized with a sterile pestle in a 1.5 mL Eppendorf tube. The weight of the tissue was determined by substracting the tare of the tube from its loaded weight. The following additives were then dispensed into a sterile 10 mL tube and placed on ice:

3 mL of Ice cold Radio Immunoprecipitation Assay buffer (RIPA) per gram of tissue.

100 µL of protease inhibitor (Sigma –Protease inhibitor cocktail 1) per mL of RIPA buffer.

10 µL of phosphatase inhibitor (Sigma –Phosphatase inhibitor cocktail 2) per mL of RIPA buffer.

The homogenized tissue was transferred to the 10 mL tube which was rotated on a MACSmix™ Tube Rotator (Miltenyi Biotec, Sydney, Australia) at 12 rpm for 1 h at 4°C. The tube was then spun in a bench top centrifuge at 14,000 x g for 10 min. The supernatant was aspirated and dispensed into a clean Eppendorf tube leaving the tissue debris behind.

The supernatant was spun again to remove any remaining debris, then aspirated, and a 50 µL aliquot was dispensed into a 0.6 mL tube for storage at -20°C.

2.8.2 Bicinchoninic acid assay (BCA)

The BCA assay was conducted with a Pierce ® BCA Protein Assay Kit. Bovine serum albumin standards (BSA) were prepared in sterile Eppendorf tubes by diluting 2 mg/mL Albumine Standard from that kit with 1 X Phosphate buffered saline (PBS). Dilutions to suit a working range of 20 to 2,000 µg/mL as specified by the manufacturer were prepared in triplicate and are listed in Table 2.2.

Table 2.2. Dilution scheme for Eppendorf tube protocol for BCA assay.

Tube	Volume of Diluent	Volume and Source of BSA	Final BSA concentration (µg/mL)
A	0	300 µL of Stock	200
B	125 µL	375 µL of Stock	1500
C	325 µL	325 µLof Stock	1000
D	175 µL	175 µL of vial B dilution	750
E	325 µL	325 µL of vial C dilution	500
F	325 µL	325 µL of vial E dilution	250
G	325 µL	325 µL of vial F dilution	125
H	400 µL	100 µL of vial G dilution	25
I	400 µL	0	Blank

The preparation of the BCA working reagent (WR) was carried out in accordance with the following manufacturer's instructions:

(#standards + #unknowns) x (#replicates) x (volume of WR per sample) = total V WR required. (9 standards + 3 samples) x (2 replicates) = 12 x 200 = 2400 µL= 2.4 mL WR. Two millilitres (2.0 mL) of the WR was required for each sample. The working reagent was prepared by mixing 50 parts of BCA reagent A with 1 part of BCA reagent B (50.1, Reagent A.B) as per manufacturer's instruction i.e. Δ 2,500 µL(A) + 50 µL(B).

Samples were diluted 1/5, 1/20 and 1/50 and 2.0 mL of the working reagent was added to each tube and mixed. The tubes were incubated at 37°C for 30 min. All tubes were cooled to room temperature. A Nanodrop spectrophotometer (Thermofisher, Australia) was used

to measure the absorbance at 562 nm for dilutions of 1/5, 1/20 and 1/50. The average was taken and the concentration determined in µg/µL.

2.8.2.1 Qubit® protein assay

A Quant iT (Life Technologies, Invitrogen, Victoria, Australia) working solution was prepared by diluting 2 µL Quant iT protein reagent in 198 µL of Quant iT buffer (1/200). Standards to be used to calibrate the fluorometer were prepared by diluting the supplied kit standards in Quant iT working solution in accordance with the manufacturer's instructions. The protein samples were also diluted in the Quant iT working solution. The dilution series are shown in Table 2.3. All diluted standards and protein samples were vortexed for 2-3 s and then incubated for 15 min at ambient temperature. The Qubit® fluorometer was calibrated using the prepared standards. After calibration of the fluorometer, the concentrations of the protein samples were determined by selecting 'protein assay' from the fluorometer menu.

Table 2.3: Standards for fluorometer calibration, and protein sample dilutions for Qubit protein assays.

	Kit standard or protein sample volume	Quant iT working solution volume	Final volume
Kit Standard 1 (0 ng/µL)	10 µL	190 µL	200 µL
Kit Standard 2 (200 ng/µL)	10 µL	190 µL	200 µL
Kit Standard 3 (400 ng/µL)	10 µL	190 µL	200 µL
Protein sample	2 µL	198 µL	200 µL

2.8.3 Protein Gels

An APS/TEMED solution was prepared by dissolving 1 polymerization tablet (Amresco, Astral Scientific, Sydney, Australia) in 1 mL of autoclaved milliQ water. Five millilitres (5 mL) of Next Gel solution (Amresco, Astral Scientific, Sydney, Australia) was mixed with 90 µL of the APS/TEMED polymerization solution. Two glass plates were filled to the top with gel solution, the comb was inserted, and the glass plates were overlaid with saturated Butanol. These polyacrylamide gels were allowed to polymerize at ambient temperature for 1 h before refrigerating.

As an alternative method, polyacrylamide gels were made *de novo.* These gels contained 30% Acrylamide/Bis solution (BioRad, Gladesville, Australia), 4 X resolving Gel Buffer (Appendix 7), 5 µL of Tetramethylethylenediamine (TEMED) per 10 mL of gel solution, and 10% APS solution. Two glass plates were filled to the top with gel solution, the comb was inserted, and the glass plates were overlaid with saturated Butanol. These polyacrylamide gels were allowed to polymerize at ambient temperature for 1 h before refrigerating.

In order to achieve greater resolution of the extracted proteins, a resolving gel and a stacking gel were used. The resolving gel, made as detailed above, was dispensed between two glass plates and filled to the three quarter level using a transfer pipette, and was overlaid with saturated Butanol. The resolving gel was allowed to set at ambient temperature for 1 h. The Butanol was then rinsed off with milliQ water and the resolving gel was dried with Kim wipes before the stacking gel was poured on top. The stacking gel was made with 30% Actylamide/Bis, 4 X stacking gel buffer, 10% APS solution, and 5 µL of TEMED per 10 mL of gel solution. The comb was then inserted and the stacking gel was overlaid with Butanol. The stacking gel was allowed to polymerase at ambient temperature and the combined gel was then wrapped in plastic and refrigerated until required.

Prior to use, the comb was removed and the wells were rinsed with milliQ water before the gels were transferred to a Mini Protean gel tank (BioRad, Gladesville, Australia). The protein samples, together with Laemmli loading buffer containing 2 M Dithiothreitol (DTT) (BioRad, Gladesville, Australia) was heated at 55°C for 15 min. The gel wells were then loaded with the protein/loading buffer mixture (40 µg/well). One well was loaded with a Precision Plus Protein™ Standard (BioRad, Gladesville, Australia). The gel was then electrophoresed at 100 V for 2 h in 1 X SDS running buffer (Amresco, Astral Scienctific, Sydney, Australia), followed by 75 V for 1 h. The gel was then stained with Coomassie Brilliant Blue (BioRad, Gladesville, Australia) for 1.5 h on an orbital shaker and de-stained overnight in a de-staining solution (Appendix 7) on an orbital shaker.

SYPRO Ruby stain (Invitrogen, Carlsbad, USA) was used as an alternative to Coomassie Brilliant Blue. After the proteins were electrophoresed, 100 mL of fixing solution was poured on the gel which was gently agitated on an orbital shaker for 30 min. The fixing solution was poured off and replaced with 60 mL of SYPRO® Ruby gel stain. The gel was agitated again on an orbital shaker overnight. The gel was transferred to a new container and 100 mL wash solution was poured on the gel and agitated for 30 min on an orbital shaker.

The gel was visualized with the Vilber Lourmat Gel imaging system using the Quantum Capture software package (Fisher BioTec, Wembley, Australia).

2.8.4 Western Blotting

2.8.4.1 Protein transfer to membrane

After the protein gel was electrophoresed, it was removed from the glass plates and the wells were trimmed away. The gel was immersed in transfer buffer for approximately 20 min. Filter papers were soaked in transfer buffer for at least 30 s. The PVDF membrane was moistened in methanol for 30 s, rinsed in milliQ water for 2 min, then placed in transfer buffer for 5 min. The foam pads were also soaked in transfer buffer. The stack was then assembled in a cassette from the cathode end to the anode end starting with the foam pad, filter paper, gel, membrane, filter paper and foam pad making sure that none of the components dried out. The cassette was then closed and immediately placed in the cassette holder facing the cathode with the membrane facing the anode. Transfer buffer, a stirrer bar, and an ice pack were added to the tank and the gel was electrophoresed for 80 min at 100 V.

Once transfer was complete the stack was disassembled and the membrane was placed into 50 mL blocking solution (low-fat powdered milk in Tween Buffered Saline -Tween (TBS-T)) and agitated on an orbital shaker for 1 h at ambient temperature. The primary antibody was diluted in the blocking solution, added to the blot, and incubated for 1 h at 4°C on an orbital shaker (Table 2.4). The primary antibody solution was discarded and the

blot incubated for 30 min with the secondary antibody diluted in TBS-T. The membrane was then washed twice for 10 min each in TBS-T and then twice for 5 min each in Tris buffered saline (TBS). The membrane was incubated with 3,3′,5,5′-Tetramethylbenzidine (TMB) for up to 20 min depending on the gene being investigated. In most instances 10 min was sufficient incubation time.

Table 2.4. Antibodies and Dilutions used in Western Blots and Dot Blots.

Name of Antibody	Primary Antibody	Secondary Antibody	Blocking %	Dilution
β-Actin	✓		5%	1:200
Anti-Rabbit HRP		✓	5%	1:2500
ZAP-70	✓		3% and 5%	1:100 and 1:2000
Anti-goat mouse IgG_1HRP		✓	3% and 5%	1:2000
mpIL-2	✓		3%	1:500
Anti-Rabbit HRP		✓	3%	1:2500

All protein samples were investigated using the β-Actin antibody (Santa Cruz, sc-130657, dil. 1.200) in order to ascertain an approximate protein expression profile. Jurkat whole cell lysate (Santa Cruz – Biotechnology, USA, sc-2204) served as positive control in the Western Blots for ZAP-70 and Lck.

2.8.4.2 Dot Blot

Two microlitres (2 µL) of peptide and horse serum was spotted onto a PVDF membrane which was then dried before proceeding. Non-specific sites were blocked by soaking the membrane in 0.5% non-fat milk in TBS-T for 1 h. The Dot Blots were incubated with the primary antibody (0.2 µg/µL). Dilutions of 1/100, 1/1,000, 1/10,000 in blocking buffer were incubated with the blot for 1 h. The blots were washed 3 times with TBS-T for 5 min each time. The blots were then incubated for 30 min with the secondary HRP conjugated antibody Anti-Rabbit IgG (W401B, Promega, Australia) diluted to 1/2,500 in blocking buffer as per the manufacturer's recommendation. The blots were washed in TBS-T for 15 min, then twice for 5 min each, followed by a final wash in TBS for 5 min. The blots were then incubated with a 3,3'-Diaminobenzidine (DAB) substrate.

2.8.4.3 Isotype control

Isotype controls were used for all primary antibodies. Isotype control antibodies have no specificity for target cells within a particular experiment. They serve to confirm the specificity of primary antibody binding that is not a result of cellular protein interactions. Isotype controls were used for both Dot Blots and Western Blots. An isotype control for the rabbit IgG contained Rabbit IgG + anti-IgG, Protein + anti IgG, Protein + PBS incubated with primary antibody, Horse serum (HS) + rabbit IgG (=negative control). Protein was spotted on PVDF membrane in single dots of 1 µL.

Chapter 3

The diprotodontic T cell signalling unit and the corresponding receptor

Chapter 3 – The diprotodontic T cell signalling unit and the corresponding receptor

Chapter 3 – The diprotodontic T cell signalling unit and the corresponding receptor

3.0 Abstract

The T cell receptor (TCR) is unable to transduce a signal without the immune receptor tyrosine activation motifs (ITAMs) of the oligomeric CD3 complex. In the metatherians, the TCRα and β chains have only been characterized in *M. domestica* and *T. vulpecula,* and the CD3ε molecule has only been characterized in *M. eugenii*. Consequently, prior to the present study, no comparison of this gene sequence with that of other metatherians was possible.

RT-PCR and RACE-PCR were used for comparative analyses of the CD3ε molecule in the two endangered marsupials *O. fraenata* and *L. hirsutus,* and in *M. domestica.* The same methods were used to investigate the TCRαβ chains in *O. fraenata* and *M. eugenii.* Structure predictions of important motifs of the marsupial TCRαβ and the CD3ε molecule were carried out with prediction programs.

It was found that the transmembrane domain of the marsupial TCR α chain was composed of different amino acid residues to that of other mammals thus suggesting a different architecture. This was subsequently supported by a structure prediction program. CD3ε was found to be highly conserved in the marsupial species included in this study. Some differences in the stalk region of the marsupial CD3ε molecule were observed when compared with that of eutherians. However, the substituted amino acid residues showed no change in the charge distribution or polarity of the stalk region.

It was shown that both the TCRβ the CD3ε chains are highly conserved among marsupial species, and between marsupials and eutherians. However, the marsupial TCRα chain, which is the functional unit of the receptor, differed considerably from the eutherian TCR α chain. Notwithstanding that difference, it is expected that this part of the marsupial T cell signalling cascade functions in a manner similar to that of eutherian mammals.

Chapter 3 – The diprotodontic T cell signalling unit and the corresponding receptor

3.1 Introduction

The T cell receptor-CD3 complex (TCR-CD3) is important for T cell development and function. Structurally important motifs within this complex are responsible for the initiation of a signal that ultimately leads to the transcription of cytokines such as interleukin-2 (IL-2) (Call *et al.,* 2004). Activation of signalling molecules occurs by ligand-binding of receptors, and in the case of T lymphocytes by the assembly of the TCR-CD3 and the interaction between them. The CD3 complex consists of a number of polypeptide proteins (γδεζη) and the α and β chains of the T cell receptor. The CD3 chains serve two major functions; the surface expression of the receptor (Clevers *et al.,* 1988a), and the activation of the T cell signalling cascade (Klausner and Samelson, 1991, Clevers *et al.,* 1988a).

Specific amino acid residues are responsible for each of these functions and are located in strategic regions within the CD3ε molecule (Xu *et al.,* 2008). Structural motifs within CD3ε, such as the ITAM motif, are responsible for signal transduction through a phosphorylation event. This leads to a conformational change and the exposure of a proline-rich region (PxxDY) capable of binding SH_3 domains (Kesti *et al.,* 2007). The associated functions of the PxxDY motif have only been investigated in detail in the last decade in human and mouse species (Aitio *et al.,* 2008, Mingueneau *et al.,* 2008). Motifs in the cytoplasmic domain of this molecule are involved in T cell differentiation (Brodeur *et al.,* 2009) and motifs responsible for the activation of the T cell receptor have been investigated in more detail in the last decade (Kim *et al.,* 2010).

In marsupial species, the CD3ε chain has been characterized at the molecular level only in *M. eugenii* (Old *et al.,* 2001). Whether or not important structural motifs are conserved in other marsupial species was not known prior to the present study. Nevertheless an anti-human CD3 antibody has been used in the past on marsupial tissues and was found to be cross-reactive due to the paratope of the human CD3ε sequence matching that of the *M. eugenii* sequence (Old *et al.,* 2001, Jones *et al.,* 1993).

The anti-human CD3 antibody has been used successfully to detect CD3 in the tissues of a number of marsupials including *Phascogale calura* (red- tailed phascogale) (Old *et al.,* 2006), *Isoodon macrourus* (northern brown bandicoot) (Old and Deane, 2002a), *Sminthopsis macroura* (stripe-faced dunnart) (Old *et al.,* 2003), *Macropus giganteus* (grey kangaroo) (Old and Deane, 2001) and *M. eugenii* (Old and Deane, 2003). However, while this indicated the presence of T lymphocytes in marsupial tissues it could not confirm the presence of functionally important motifs and their conservation among those species.

A comparative investigation of the CD3ε chain is important since it highlights the evolution of the gene. In order to detect sequence changes within the marsupial clade compared to other mammals it was necessary to investigate this gene in macropods other than *M. eugenii*, such as *O. fraenata* and *L. hirsutus* and in the distantly related marsupial *M. domestica.*

The two invariant α and β chain of the TCR are linked by disulphide bridges but also contain specific amino acid residues that are involved in the dimerization as well as the localization of the receptor on the membrane (Arnaud, 1997). The chains of the T cell receptor have been characterized in *T. vulpecula* (Australian silver brushtail possum) and *M. domestica* (Zuccolotto *et al.,* 2000, Baker *et al*., 2001) but not in *M. eugenii* or *O. fraenata*.

Two important motifs are contained within the TCRα chain; the transmembrane region responsible for the assembly and surface expression of the receptor, and the connecting peptide which ensures a close proximity of the receptor to the co-receptor CD8. These assure signal strength and control antigen responsiveness (Mallaun *et al.,* 2008, Backstrom *et al.,* 1996).

The 14 amino acid long FG loop (F = phenylalanine, G = glycine) which lies in the constant domain of the TCR beta chain is the structural link between the CD3 subunits and the receptor during the signalling event (Degermann *et al.,* 1999b, Wang *et al.,* 1998b). The FG loop is important in the biological function of the beta chain.

In this study prominence was given to investigation of the FG loop structure to determine its conservation and possible changes in charge distribution that may alter the biological activity of the molecule. This study also investigated other important structural motifs within the cDNA sequence, their conservation, and secondary structure. Where possible, models of the marsupial molecules CD3ε, the TCRα chain and the TCRβ chain were constructed.

3.2 Aims and Objectives

The aims of the experiments reported in this Chapter were to investigate the marsupial TCR and CD3ε molecules, and in particular:

- To characterize the CD3ε chain in the endangered marsupials *O. fraenata* and *L. hirsutus*.
- To confirm the existing gene predictions for CD3ε in the ensembl (http.//www.ensembl.org) and UCSC (http.//genome.ucsc.edu/) databases for *M. domestica*.
- To investigate the nucleotide sequences of CD3ε, TCRα and TCRβ for possible polymorphisms and to explore in particular the putative amino acid sequences for any apparent differences in their structural and functional motifs.
- To determine the position of the marsupial clade compared to other mammals employing phylogenetic analyses.
- To undertake phylogenetic analyses of functional motifs in order to determine divergent evolution within these motifs.
- To conduct protein modelling from the putative amino acid sequences and to deduce possible amino acid residue interactions through ligand binding.

3.3 Specific Materials and Methods

The following materials and methods were unique to this chapter. General materials and methods are detailed in Chapter 2.

3.3.1 RNA, mRNA and cDNA

RNA was isolated from spleen, liver, gut node and thymus of *O. fraenata* as described in Chapter 2, section 2.1.3. mRNA was isolated from total RNA, and cDNA was synthesized as described in Chapter 2, sections 2.1.3.2, and 2.1.5.

RNA and mRNA extracted from *M. eugenii* spleen were donated by Dr. L. J. Young. RNA from stimulated *M. eugenii* lymphocytes was also donated by Dr. Young for further processing.

RNA, mRNA and cDNA were quantified as described in Chapter 2.1.4. All sequences obtained by PCR from cDNA were cloned as detailed in Chapter 2, section 2.1.10. Three clones were prepared for sequencing in the forward and reverse direction as detailed in Chapter 2, section 2.1.10.6.

A cDNA library from *M. domestica* thymus and spleen was donated by Dr. K. Belov and was used for elucidation of the *M. domestica* CD3ε gene as described in section 3.3.3. cDNA was also synthesized from RNA isolated from *L. hirsutus* spleen by K. Howard.

3.3.2 Reverse Transcription

RNA and mRNA were reverse transcribed as described in Chapter 2, section 2.1.5. The transcribed cDNA was verified with glucose-6-phosphate dehydrogenase (G6PD) and trialled with CD3ε.

3.3.3 Primer design

3.3.3.1 CD3ε, TCRα and TCRβ

Primers were designed as detailed in Chapter2, section 2.1.2. The primer sequences are presented in Table 3.1.

Table 3.1. Primer sequences used to deduce the open reading frames of CD3ε, and the partial open reading frames of TCRα and TCRβ in three marsupial species including the source tissue.

Gene	Forward primer 3' – 5'	Reverse primer 3' – 5'	T_m	Size	Species and source tissue
CD3ε	gaaataaacccaccaaaccctg (5' untranslated region)	ccaggctggaagtggaggg (3' untranslated region)	53°C/63°C	~800bp	*O. fraenata* (spleen, thymus) *L. hirsutus* (spleen) *M. domestica* (cDNA library)
	atgcatttggaagctctctggactgtg (START)	gatggctctctgattcaggccagcata (STOP)	61°C/61°C	~567bp	*O. fraenata* (spleen, thymus) *L. hirsutus* (spleen) *M. domestica* (cDNA library)
	atgcagttgggatctctctggaccgt	catggctctctgattcaggccagcatac	63°C/62.5°C	~600bp	*M.domestica* (cDNA library)
		caggaaacagctatgac (M13*) 3' end	47°C	1000bp	*M. domestica* (cDNA library)
TCRα	tgcctcttcacmgaytttgactc	ccacagmagmagcgtcatgarcagg	53°C/62.9°C	308bp 295bp	*O. fraenata* (thymus) *M. eugenii* (PHA stim. Lφ)
TCRβ	gccacwctggtctgtgtggccacaggc	ggtwacmccacaatctgcytttcccc	67°C/61°C	834bp 862bp	*O. fraenata* (spleen, thymus) *M. eugenii* (PHA stim. Lφ)

*M13 reverse primer melt Temp. 47°C was used in the cDNA library of *M. domestica*

3.3.3.2 Polymerase Chain Reaction (PCR), cloning and sequencing

Generally used RT-PCR templates, concentrations of the PCR mixes, and preparation of the RT-PCR and RACE-PCR products are detailed in Chapter 2 sections 2.1.6, 2.1.8 and 2.1.10. Table 3.2 shows the PCR templates used in this Chapter.

RACE primers were designed as outlined in Chapter 2, section 2.1.2 from the initial sequence derived from the RT-PCR products. RACE-PCRs were performed to obtain the 5' and 3' ends of the molecules as described in Chapter 2 section 2.1.9.

Table 3.2. PCR and RACE –PCR templates used for amplification of CD3ε, TCRα and TCRβ from three marsupial species.

Species	Gene of interest	PCR template*	RACE PCR template
O. fraenata *L. hirsutus* *M. domestica*	CD3ε	No.3 No.3 No.3	---
O. fraenata *M. eugenii*	TCRα	Nos.1 and 2 Nos.1 and 2	No. 2 No. 2
O. fraenata *M. eugenii*	TCRβ	No.1 No.1	Touchdown (60-50°C) ---

*These numbers relate to the templates specified in General Materials and Methods (Chapter 2).

3.3.4 Amplification of 3' end of CD3ε from *M. domestica* cDNA library

A cDNA library constructed from *M. domestica* thymus and spleen tissue using the λZAPII vector was used to amplify the 3' end of *M. domestica* CD3ε to confirm the predicted coding domain of the molecule. A primer containing the start site of the CD3ε molecule was paired with a primer complementary to the M13 universal sequencing site in λZAPII. The annealing temperature of the RT-PCR reaction was 55°C and a Platinum® HiFi Taq DNA polymerase (Invitrogen, Carlsbad, USA) was used. The product was excised and prepared for direct sequencing. When it was verified that the correct product was amplified, it was cloned as described in Chapter 2, section 2.1.10.6.

3.3.5 Phylogeny

A phylogenetic analysis was undertaken for CD3ε using Mega5. Phylogenetic analyses were undertaken for specific structurally important motifs in the TCRα and TCRβ sequences with a Dayhoff algorithm in the Mega5 program.

3.3.6 Bioinformatics

All of the bioinformatics tools detailed in Chapter 2 were used to elucidate putative domain structures, putative glycosylation sites, possible disulphide bonds, putative amino acid sequences, putative secondary and tertiary structures, and homology searches.

3.4 Results

3.4.1 RNA and mRNA isolation

The RNA and mRNA concentrations determined with a Qubit® Fluorometer and verified with a Nanodrop spectrophotometer (Thermofisher, Australia) as detailed in Chapter 2 are listed in Table 3.3. The yield of mRNA from total RNA was approximately 0.3%.

Table 3.3. Concentrations of RNA, mRNA and cDNA for *O. fraenata* and concentrations of cDNA for *M. eugenii,* in various tissues.

O. fraenata				*M.eugenii*
Tissue	Total RNA Concentration	mRNA Concentration	cDNA Concentration	cDNA Concentration
Spleen	2.0 µg/µL	0.318 µg/µL	0.535 µg/µL	
Liver	4.920 µg/µL	0.036 µg/µL and 0.550 µg/µL	0.548 µg/µL	
Gut Node	0.22 µg/µL	0.4 µg/µL	0.460 µg/µL	
Thymus	0.56 µg/µL	0.202 µg/µL	0.510 µg/µL	
PHA stim. Lϕ	*			0.445 µg/µL
ConA stim.Lϕ	*			0.432 µg/µL

*(donated by Dr. Lauren Young)

3.4.2 CD3 epsilon (CD3ε)

3.4.2.1 CD3ε - Homology

A homology search revealed a high conservation among the marsupials. The expressed *M. domestica* sequence had an identity to *M. eugenii* of 79% at the nucleotide level and 74% at the amino acid level. *O. fraenata* and *L. hirsutus* sequences each had an identity to *M. eugenii* of 93% at the nucleotide level and 85% and 87% respectively at the amino acid level. *O. fraenata* and *L. hirsutus* sequences had an identity to each other of 91% at the nucleotide level and 82% at the amino acid level. For the first time, a comparison could be made between the predicted and the expressed sequences of *M. domestica* CD3ε. A

BLASTx search showed a sequence identity of 95% between the predicted and expressed amino acid sequences of the *M. domestica* CD3ε molecule which indicates that the annotation might not be accurate.

The conservation of the marsupial CD3ε amino acid sequences to the human CD3ε amino acid sequence varies among the species. The expressed *M. domestica* sequence had a 54% identity, while the predicted sequence had a 55% identity to the human sequence. The *O. fraenata* and *L. hirsutus* sequences had identities of 58% and 55% respectively to the human sequence. Since the amino acid sequence identity was above 50%, it can be hypothesized that the tertiary structure may be similar to that of the human structure. The identity percentages of the CD3ε nucleotide and amino acid sequences in the marsupial species are shown in Table 3.4.

Further examination of the results revealed that, in the three marsupial species, some of the amino acid sequences were of low complexity resulting in a lower protein identity percentage. Twenty three low complexity amino acid residues were found in *O. fraenata*, 34 were found in *L. hirsutus* and 6 were found in *M. domestica*. The low complexity sequences in the three marsupial species occurred in the helix region of the CD3ε molecule. In *O. fraenata* and *L. hirsutus* the low complexity sequences commenced prior to the helix region and extended beyond that motif indicating that this region was not modelled with a high confidence level.

Table 3.4. Homology search results for the *M. domestica*, *O. fraenata* and *L. hirsutus* CD3ε nucleotide and amino acid sequences, their identities and their respective e-values.

CD3ε Homology search results for *M. domestica*, *O. fraenata* and *L. hirsutus*					
Species	To	Nucleotide	e-value	Amino acid	e-value
M. domestica (582bp, 193aa)	*M. eugenii*	79%	2e.142	74%	3e.95
	O. fraenata(*)	77%	2e.133	68%	3e.66
	L. hirsutus (*)	79%	1e.147	72%	3e.74
	S. harisii	79%	6e-142	74%	7e.89
	C. jacchus	77%	5e.60	57%	4e.65
	M. fascicularis	77%	3e.57	55%	1e.60
	H. sapiens	76%	5e.54	54%	1e.60
	F. catus	75%	5e.48	53%	2e.58
	S. scrofa	75%	2e.47	55%	2e.56
	M. musculus	75%	8e.45	50%	6e.53
	R. norvegicus	73%	1e.42	50%	1e.55
	C. familiaris	73%	5e.41	49%	8e.56
	A. mexicanum	69%	7e.14	41%	1e.37
	G. gallus	67%	3e.13	45%	8e.34
O. fraenata (567bp, 188aa)	*M. eugenii*	93%	0.0	85%	2e.93
	S. harisii	83%	3e-165	79%	3e.87
	M. domestica (predicted)	79%	4e.131	67%	1e-73
	M. domestica (expressed)	77%	2e.133	68%	8e.82
	L. hirsutus	91%	0.0	82%	2e.102
	C. jacchus	82%	1e.79	59%	3e.55
	M. fascicularis	81%	3e.75	59%	5e.53
	H. sapiens	81%	6e.72	58%	1e.54
	A. platyrhynchos	86%	2e-13	48%	9e.30
	M. musculus	78%	5e.60	53%	1e.46
	C. familiaris	77%	6e.59	53%	7e.54
	S. scrofa	77%	3e.56	56%	2e.48
	R. norvegicus	75%	5e.54	54%	3e.47
	B. taurus	76%	7e.52	55%	7e.53
	F. catus	76%	3e.50	54%	1e.52
	A. mexicanum	69%	2e.14	43%	1e.29
	X. laevis	77%	4e.04	37%	2e.21
L. hirsutus (579bp, 192aa)	*M. eugenii*	93%	0.0	87%	3e.88
	S. harisii	82%	3e.165	79%	2e.76
	M. domestica (predicted)	78%	1e-132	68%	9e-65
	M. domestica (expressed)	79%	1e-147	73%	3e.92
	C. jacchus	81%	4e.74	-----	------
	M. fascicularis	80%	2e.72	56%	9e.41
	H. sapiens	78%	1e.68	55%	4e.40
	F. catus	75%	4e.49	53%	5e.41
	S. scrofa	76%	3e.50	51%	2e.34
	M. musculus	77%	1e.55	53%	5e.35
	R. norvegicus	74%	1e.49	46%	1e.32
	B. taurus	75%	5e.48	51%	3e.36
	C. familiaris	77%	2e.58	50%	3e.36
	G. gallus	-----	-----	43%	2e.21

(*) carried out with BLAST 2 program, for Accession numbers see Appendix 3A.

3.4.3.2 CD3ε - Structural motifs

Structural motifs such as leader sequences, disulphide bridges, the stalk region, ITAM and a helix and polyproline motif were detected in all three marsupial CD3ε sequences and are listed in Table 3.5.

When the CD3ε motif sequences in *O. fraenata, L. hirsutus, M. domestica* and *M. eugenii* were compared with those of the human, it was found that the least conserved motif was the leader sequence in *L. hirsutus*. The identity percentages of the putative leader sequences in *L. hirsutus, O. fraenata, M. eugenii* and *M. domestica* were 48%, 52%, 52% and 71% respectively.

The polyproline motif sequences in *O. fraenata*, *L. hirsutus, M. eugenii* and *M. domestica* were highly conserved between those four marsupial species. The conservation percentage of each of those species compared to human was 100% except for *L. hirsutus* which was 92%. This high conservation level indicates that the polyproline motif in those marsupials is important in the function of the molecule.

In the stalk motif, the putative cysteine residues were conserved across the four marsupial species and human. In the marsupials, the stalk motif identity to the human sequence ranged from 67% to 70% except in *M. domestica* which had a 50% identity. In the marsupials, the transmembrane region identity to the human transmembrane region ranged from 59% to 77%. The endoplasmic retention signal motif identity to the human endoplasmic retention signal motif ranged from 75% to 88%. In each of the four marsupials, the helix motif identity to the human helix motif was 75%. The ITAM motif identity to the human ITAM motif ranged from 80% to 93%.

When all the motifs of the marsupial sequences were compared with each other it was found that there was 100% identity between *O. fraenata* and the published sequence of *M. eugenii*. Both *L. hirsutus* and *M. domestica* were also highly conserved with identity percentages above 75%.

Table 3.5. Comparison of CD3ε structural motifs in three macropod and one didelphid species to the human CD3ε sequence. The conservation percentages of the structural motifs were calculated in relation to *M. eugenii* and the *H. sapiens* putative amino acid sequences found in Genbank under Accession numbers AY028923 and NM_000733.3 respectively.

Species	Functional motifs	Conservation %	
		M. eugenii	*H. sapiens*
O. franeata (not yet deposited in Genbank)	MHLEALWTVVGFCLLSACVWG (Leader sequence)	100%	52%
	CLLSAC/CEGC (disulfide bridge/stalk motif)	100%	70%
	MDVLTVAGIVIADVFITLGVLLLVYYW (TM region)	100%	70%
	RGGGGGGR (Helix motif)	100%	75%
	RPPPVPNPDYEP (Polyproline motif/Nck bind. site)	100%	100%
	YEPIRKGQRDLYAGL (ITAM motif)	100%	93%
	DLYAGLNQ (ER retention motif)	100%	88%
L. hirsutus (not yet deposited in Genbank)	MHLEALWTVVGF**C**QLSA**C**VWG (Leader sequence)	95%	48%
	CQLSA**C/C**EG**C** (disulfide bridge/stalk motif)	90%	67%
	MDVLTVAGIVIADVFITLGVLLLVYYW (TM region)	100%	77%
	RGGGGGGR (Helix motif)	100%	75%
	RPPPVPNPDYDP (Polyproline motif)/Nck bind. site	92%	92%
	YDPIRKGQQDLYAGL (ITAM motif)	86%	80%
	DLYAGLNH (ER retention motif)	88%	75%
M. domestica (not yet deposited in Genbank)	MQLGSLWTVLGFFLLSA**C**VWG (Leader sequence)	76%	71%
	FLLSAC/**C**HG**C** (disulfide bridge/stalk motif)	80%	50%
	MGVLTVAGIIIADVFITLGVLILVYHW (TM region)	85%	59%
	RGGGAGGK (Helix motif)	75%	75%
	RPPPVPNPDYEP (Polyproline motif/Nck bind. site)	100%	100%
	YEPIRKGQRELYAGL (ITAM motif)	93%	86%
	ELYAGLNQ (ER retention motif)	88%	75%
M. eugenii (Genbank Accession No. AY028923)	MHLEALWTVVGFCLLSACVWG (Leader sequence)	100%	52%
	CLLSAC/CEGC (disulfide bridge/stalk motif)	100%	70%
	MDVLTVAGIVIADVFITLGVLLLVYYW (TM region)	100%	70%
	RGGGGGGR (Helix region)	100%	75%
	RPPPVPNPDYEP (Polyproline motif/Nck bind. site)	100%	100%
	YEPIRKGQRDLYAGL (ITAM motif)	100%	93%
	DLYAGLNQ (ER retention motif)	100%	88%
H. sapiens (Genbank Accession No. NM_000733.3)	MQSGTHWRVLGLCLLSVGVWG (Leader sequence)	52%	100%
	CLLSVG/CENC (disulfide bridge/stalk motif)	70%	100%
	MDVMSVATIVIVDICITGGLLLLVYYW (TM region)	70%	100%
	RGAGAGGR (Helix motif)	75%	100%
	RPPPVPNPDYEP (Polyproline motif)/Nck bind. site	100%	100%
	YEPIRKGQRDLYSGL (ITAM motif)	93%	100%
	DLYSGLNQ (ER retention motif)	88%	100%

TM = transmembrane, ER = endoplasmic retention

3.4.3.3 CD3ε – Domain structure

In *O. fraenata*, *L. hirsutus* and *M. domestica* a putative Ig-like domain and an ITAM motif were identified. The e-values in Table 3.6 show the high sequence conservation of the Ig-like domain and the ITAM motif indicating that these domains were conserved in the three marsupials.

Table 3.6. Significant e-values of important structural motifs in *O. fraenata*, *L. hirsutus* and *M. domestica.*

Structural domains of CD3ε in three marsupials and the respective e-values		
Species	Domain	e-value
O. fraenata	Ig-like domain	3.93e-01
L. hirsutus		1.36e-01
M. domestica		1.36e-01
O. fraenata	ITAM	1.68e-03
L. hirsutus		1.34e-02
M. domestica		2.54e-03

A putative signal peptide of 21 amino acid residues in length was predicted in the open reading frame of each of the three marsupial sequences. This was similar to the signal peptides found in other mammals. The prediction program indicated that > 99% probability of the sequence qualified it as a signal peptide. These results are listed in Table 3.7.

Table 3.7. Signal peptide length and cleavage probability as predicted by SignalIP 4.0 for *O. fraenata*, *L. hirsutus* and *M. domestica.*

Predicted CD3ε signal peptide length and probability in three marsupials		
Species	Signal peptide length	Probability
O. fraenata	21 amino acids	99.9%
L. hirsutus	21 amino acids	99.5%
M. domestica	21 amino acids	99.9%

3.4.3.4 CD3ε - Glycosylation and glycation sites (non-enzymatic glycosylation)

Single putative O-linked glycosylation sites were predicted in *O. fraenata* at position 153, and in *L. hirsutus* and *M. domestica* at position 157. No putative N-linked glycosylation sites were predicted in any of the three marsupial CD3ε chains.

Putative glycation sites were predicted only in *O. fraenata* and *M. domestica* (Table 3.8). It appeared that the CD3ε sequence in *L. hirsutus* was devoid of any putative glycation sites thus indicating a possible difference in post-translational modification. Conventional prediction methods are not sufficiently stringent to produce accurate data. Detection of glycated proteins with high confidence levels requires the method of MALDI-TOF-MS (Meltretter and Pischetsrieder, 2008) or high-resolution mass spectrometry (Stefanowicz *et al.,* 2010), therefore the values produced in this study with N-glycate have to be treated with caution. Glycation or non-enzymatic glycosylation can be a measure of age as well as a diabetic marker. Most of the putative glycation sites were found in the oldest animal, *M. domestica,* while no glycation sites were found in the youngest animal *L. hirsutus*. This was an observation only since no experimental data were available.

Table 3.8. Positions of glycated lysine residues in *O. fraenata*, *L. hirsutus* and *M. domestica* CD3ε chains and their probabilities.

CD3ε glycation sites and probabilities for three marsupials		
Species	Positions of glycated lysine residues	Probability
O. fraenata	44	82.1%
	52	51.6%
L. hirsutus	None predicted	
M. domestica	49	84.6%
	57	74.9%
	88	91.4%
	139	74.2%
	156	55.8%

3.4.3.5 CD3ε - Phosphorylation sites

Putative phosphorylation sites were predicted in the three marsupial species however the location and frequency of those sites varied between the species. Three serine, two threonine and five tyrosine phosphorylation sites were predicted in the *O. fraenata* CD3ε sequence. The *L. hirsutus* sequence contained four serine, no threonine, and four tyrosine predicted phosphorylation sites. In *M. domestica,* there were six serine, one threonine and four tyrosine predicted phosphorylation sites. The positions of the predicted phosphorylation sites are shown in Table 3.9.

Table 3.9. Predicted phosphorylation sites and their positions in the CD3ε sequence in *O. fraenata*, *L. hirsutus* and *M. domestica.*

Predicted phosphorylation sites and their positions in the CD3ε molecules of three marsupial species			
Species	Serine	Threonine	Tyrosine
O. fraenata	23 33 85	45 57	31
L. hirsutus	35 37 73 90		79
M. domestica	35 37 73 90 145	71	79

Putative phosphokinase binding sites were predicted in all three species investigated. In *O. fraenata* this predicted binding site was located at position 85 with a probability of 75%. In *L. hirsutus* it was located at position 90 with a probability of 76%, and in M. domestica it was located at position 138 with a probability of 85%.

3.4.3.6 CD3ε - Disulphide bond predictions

Disulphide bonds are important for stabilizing regions like the stalk region in the CD3ε molecule. Two putative disulphide bonds were found in the *O. fraenata* sequence of which one was an intrachain disulphide bond. *L. hirsutus* showed only one putative disulphide bond which was an interchain bond, while in *M. domestica* three putative disulphide bonds were predicted one of which was an intrachain disulphide bond. Predicted disulphide bonds which were identified using the DiANNA server are shown in Table 3.10.

Table 3.10. Disulphide bond predictions carried out with DiANNA showing residue number, type of bonds, and probabilities in *M. domestica*, *O. fraenata* and *L. hirsutus* CD3ε sequences.

Predicted disulphide bond residues, types and probabilities in three marsupials				
Species	Residue number	Interchain	Intrachain	Interchain probability
M. domestica	18-85	✓		high
	46-105	✓		high
	106-109		✓	
O. fraenata	41-103	✓		high
	13-18		✓	
L. hirsutus	45-85	✓	None predicted	high

3.4.3.7 CD3ε - Primary sequence and secondary structure predictions

The open reading frame of *O. fraenata* CD3ε consisted of 567bp which translated into 188 amino acid residues (Fig. 3.1). The open reading frame of *L. hirsutus* consisted of 579bp which translated into 192 amino acids (Fig. 3.2) and the open reading frame of *M. domestica* consisted of 582bp which translated into 193 amino acids (Fig. 3.3). The secondary structures, such as alpha helices and beta strands, for each species are shown in the same figures. The endoplasmic retention signal found in eutherian mammals at positions 174-183 was conserved in the three marsupial sequences at the same positions.

Differences in strand and helix distributions were detected within the three marsupial sequences. The alpha helix which almost spanned the distance of the leader sequence in *O. fraenata* was shorter than the helix in the leader sequences of *L. hirsutus* and *M. domestica.* Despite its length, this helix does not qualify as a transmembrane helix. This alone may not be of significance. Additional beta sheets as found in *O. fraenata* and *L. hirsutus* might strengthen the tertiary structure due to increased hydrogen bonding. Additional alpha helices were found in *O. fraenata* and *M. domestica*. The second helix in *O. fraenata* and *L. hirsutus* appeared to be an amphipathic alpha helix, which stabilizes helix-helix packing, unlike the second helix in *M. domestica*. This indicates that in *O. fraenata* and *L. hirsutus* the area where the second helix is located interacts with polar residues. The third helix in *M. domestica* is located in the ITAM domain of the molecule. This was unusual since some of the residues could be buried in a turn therefore being

inaccessible to signalling moieties. Examination of the sequence of the helix (QRELYA; Q = glutamine, R = arginine, E = glutamic acid, L= leucine, Y = tyrosine, A = alanine) found that most amino acid residues in this structure prefer a helix conformation and are polar in nature. Three of the residues are hydrophilic and three are hydrophobic thus balancing the ability to interact with an aqueous environment. It appeared that although some residues might be buried in an alpha helical turn, sufficient exposure to the outside environment occurs.

A transmembrane helix was predicted for the *M. domestica* putative amino acid sequence at position 113 reaching as far as position 135. In contrast, the *O. fraenata* prediction indicated the presence of two transmembrane helices. One helix was located in the area of the signal peptide, and the second transmembrane helix was at positions 109 to 131 in the stalk region. This prediction appeared to be accurate since the stalk region interacts with CD3δ and the amino acid residues need to be accessible. The *L. hirsutus* putative amino acid sequence appeared to have only one predicted transmembrane helix located at positions 112 to 134. The apparent differences in the secondary structure would influence the tertiary structure of these molecules.

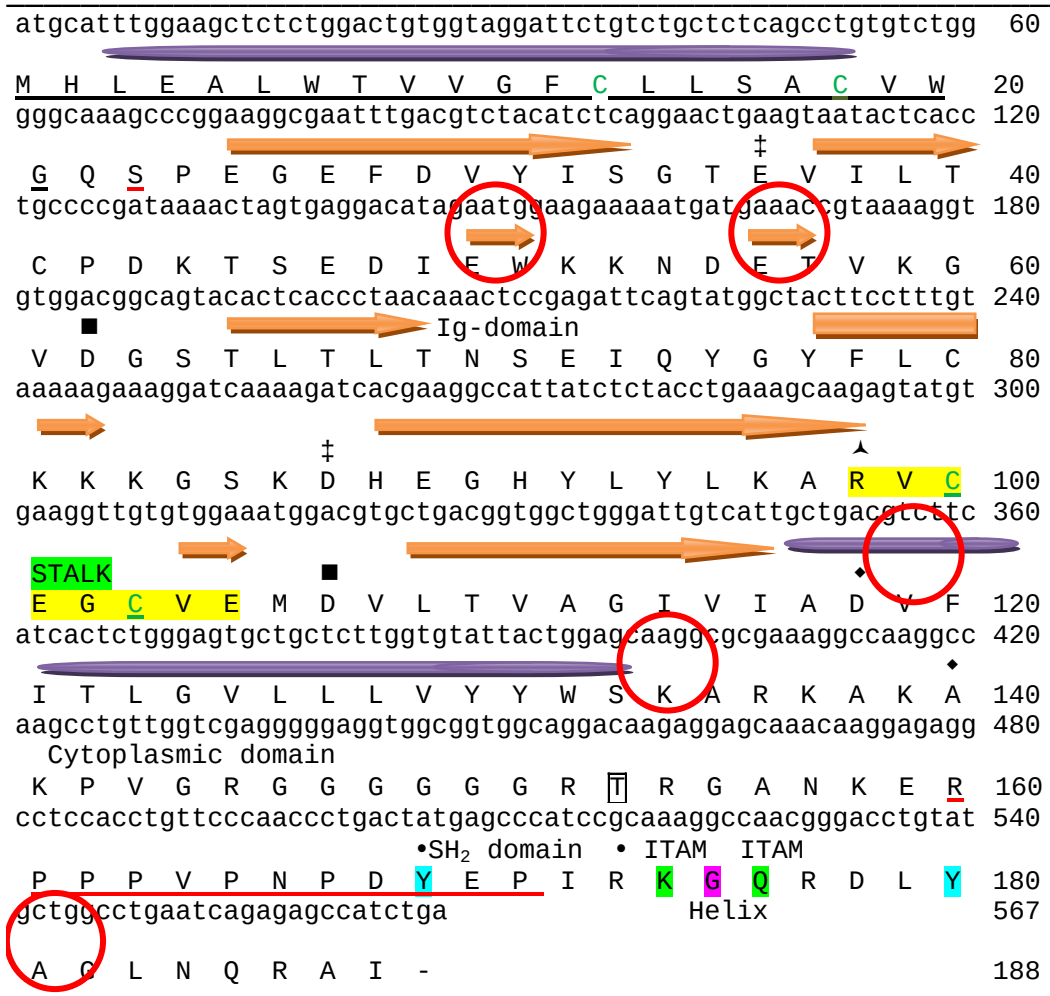


Figure 3.1. *O. fraenata* CD3ε primary sequence and secondary structure prediction. Black underlined = signal peptide. C = functionally important cysteines. S = putative serine phosphorylation site. ‡ = Ig domain. ⅄ = C-terminal end of β sheet. Yellow highlight = stalk region. ■ = putative Zinc finger domain. ◆ = Extracellular domain. T = Threonine glycosylation. Red underlined = proline rich motif. • = SH_2 domain. K and Q = ITAMs motif. G = helix motif. Y = phosphorylated tyrosines. Amino acid residues 167-183 = ARAM (antigen recognition activation motif). Amino acid residues 174-183 = endoplasmic retention signal. = Transmembrane helix, = Strand. = differences in secondary structure compared with *M. eugenii*.

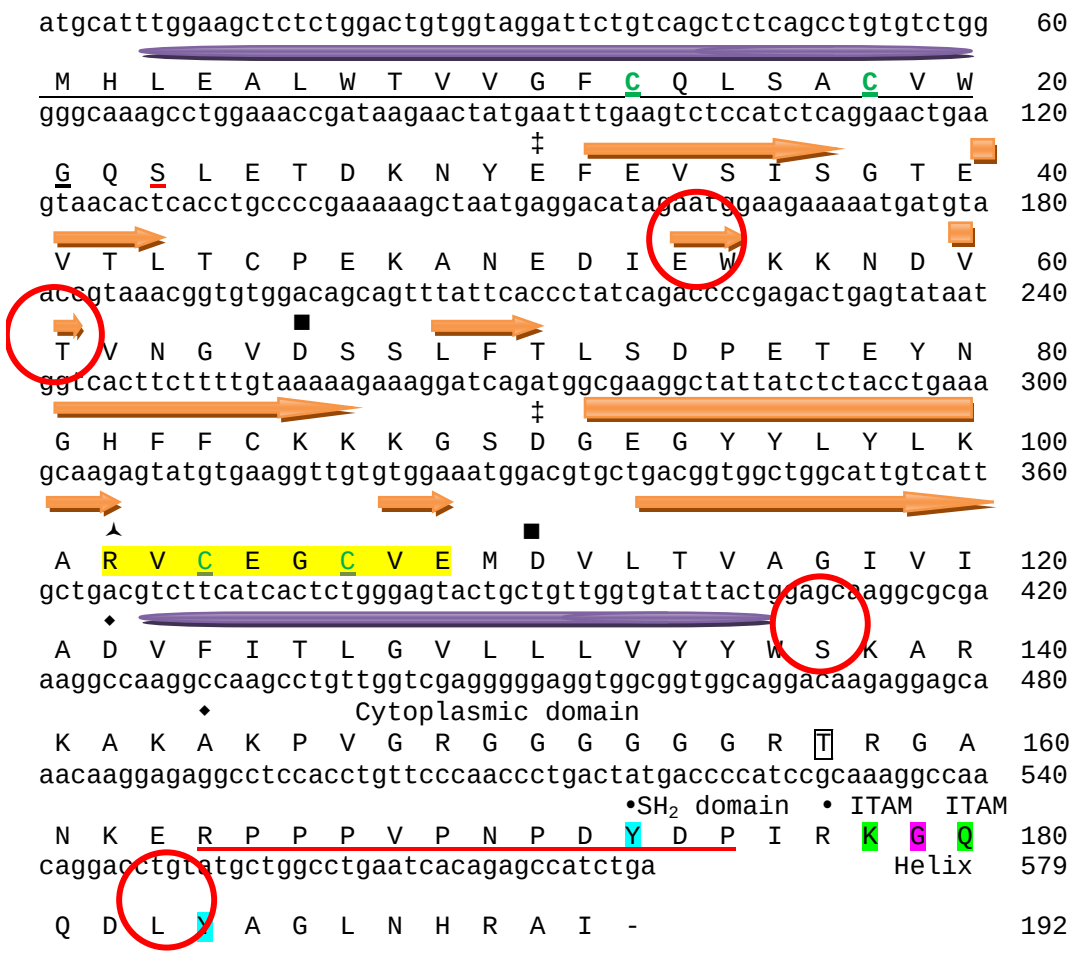


Figure 3.2. *L. hirsutus* CD3ε primary sequence and secondary structure prediction. Black underlined = signal peptide. C = functionally important cysteines. S = putative serine phosphorylation site. ‡ = Ig domain. ⅄ = C-terminal end of β sheet. Yellow highlight = stalk region. ■ = putative Zinc finger domain. ◆ = Extracellular domain. T = Threonine glycosylation. Red underlined = proline rich motif. • = SH_2 domain. K and Q = ITAMs motif. G = helix motif. Y = phosphorylated tyrosines. Amino acid residues 167-183 = ARAM (antigen recognition activation motif). Amino acid residues 174-183 = endoplasmic retention signal. = Transmembrane helix, = Strand. = differences in secondary structure compared with *M. eugenii*.

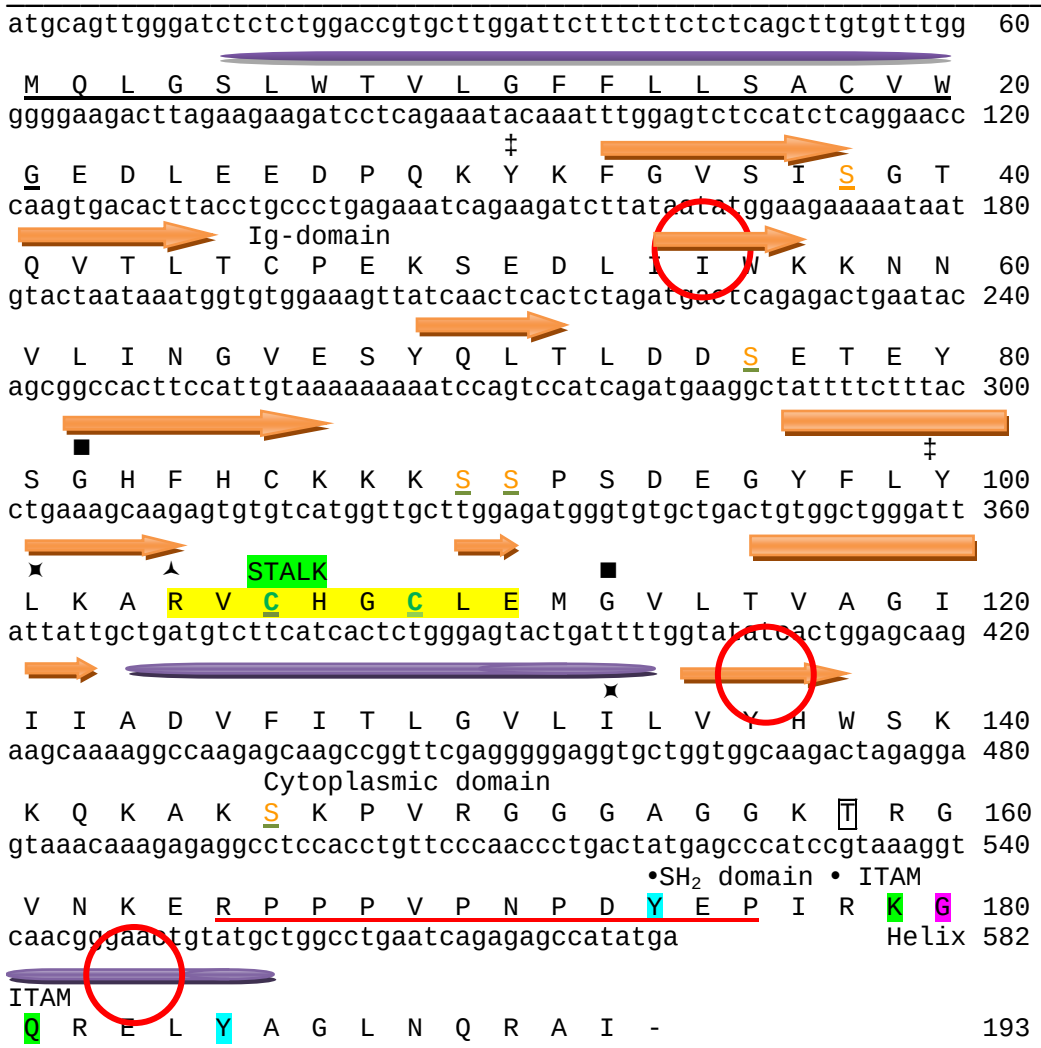


Figure 3.3. *M. domestica* CD3ε primary sequence and secondary structure prediction. Black underlined = signal peptide. C = functionally important cysteines. S = putative serine phosphorylation site. ‡ = Ig domain. ⅄ = C-terminal end of β sheet. Yellow highlight = stalk region. ■ = putative Zinc finger domain. ◆ = Extracellular domain. T = Threonine glycosylation. Red underlined = proline rich motif. • = SH_2 domain. K and Q = ITAMs motif. G = helix motif. Y = phosphorylated tyrosines. Amino acid residues 167-183 = ARAM (antigen recognition activation motif). Amino acid residues 174-183 = endoplasmic retention signal. = Transmembrane helix, = Strand. = differences in secondary structure compared with *M. eugenii*.

3.4.3.8 Tertiary structure modelling

Homology modelling produced a slightly different model for each marsupial CD3ε sequence investigated thus demonstrating that the differences in the secondary structure could impact on the tertiary structure of the CD3ε molecule. The models all had the same orientation. A structure prediction was also carried out for the published CD3ε sequence of *M. eugenii* for comparison with the other two macropods, *O. fraenata* and *L. hirsutus*. The models that resulted in the highest C-score (confidence score) are the most probable and are shown in Fig. 3.4.

It appeared that the beta sheets were all in similar areas although the differences in length, and the addition of smaller beta sheets in the secondary structure, clearly made a difference to the predicted folding of the protein.

O. fraenata	*L. hirsutus*	*M. eugenii*	*M. domestica*
C-score = -3.530	C-score = -3.065	C-score = -3.3	C-score = -3.28

Figure 3.4. Most probable models of the CD3ε chain produced by I-TASSER in (a) *O. fraenata* (b) *L. hirsutus* (c) *M. eugenii* d) *M. domestica*.

3.4.3.9 CD3ε - Phylogenetic analysis

A phylogenetic analysis carried out for all known CD3ε sequences confirmed the high sequence homology of the CD3ε sequences in *O. fraenata, L. hirsutus* and *M. domestica* with those known sequences (Table 3.4). The evolutionary history was inferred with a Neighbor-Joining phylogenetic tree because, statistically it is the most consistent, and

generates a true tree with high probability. The optimal tree had a sum of branch lengths of 5.51. The branch lengths were different for each of the three marsupial species. *M. domestica* had the longest branch length. This indicated that, although *O. fraenata* and *L. hirsutus* are closely related to each other, there are differences in their sequences. This is consistent with the BLAST search. The percentage of replicate trees in the bootstrap test (1,000 replicates) in which the associated taxa are clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (Fig. 3.5). The analysis involved 26 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 128 positions in the final dataset.

O. fraenata and *L. hirsutus* are clustered together, with *M. domestica* being slightly more distant. The marsupial clade is located between the primates and the birds indicating that, although the marsupial CD3ε gene is highly conserved, the marsupials are only distantly related to most eutherians.

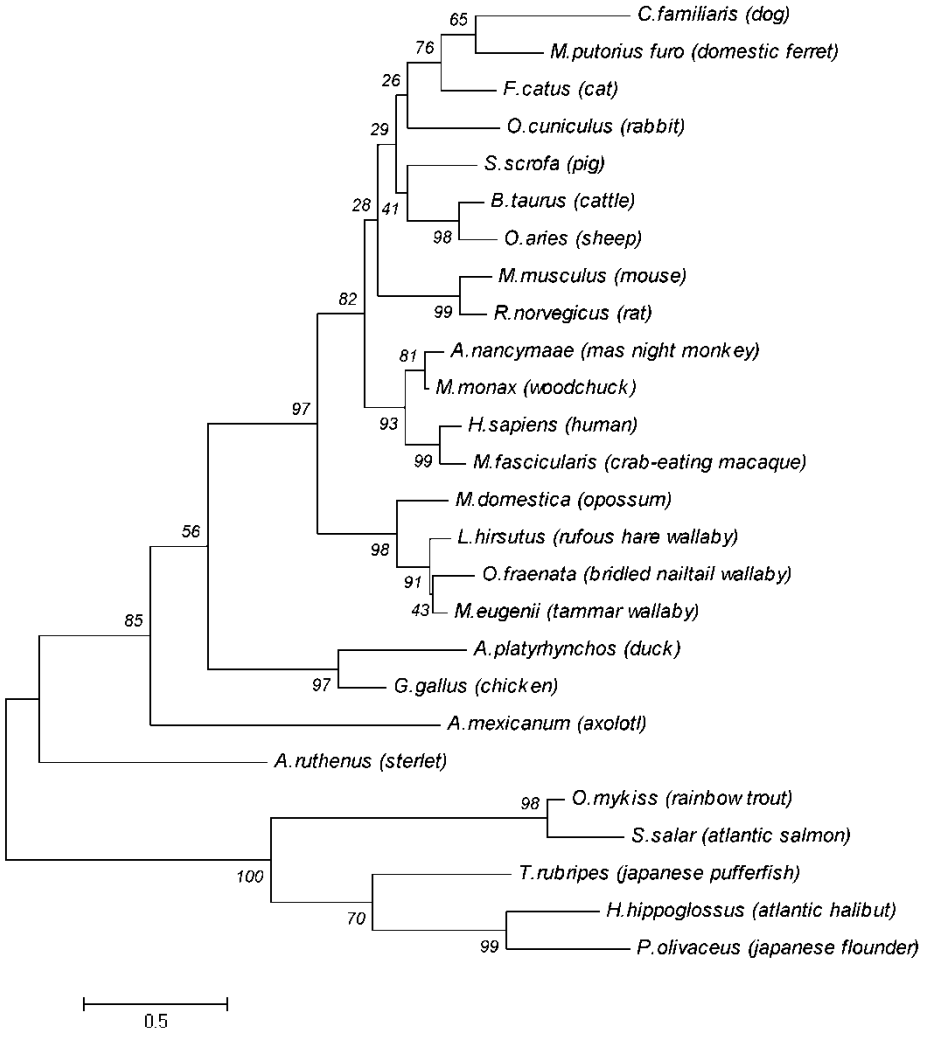


Figure 3.5. Neighbor-Joining phylogenetic tree for CD3ε (including the marsupials *M. eugenii*, *O. fraenata*, *L. hirsutus* and *M. domestica*). Branch lengths determine the evolutionary relationship between the species.

3.4.3.10 CD3ε - Gene organization

The exon/intron boundaries of the CD3ε sequences in *M. eugenii* and *M. domestica,* which were investigated within the ensembl database, showed differences in the length of the exons when compared to humans and other mammals (supporting data are in Appendix 3A). The exon which is missing from the marsupial CD3ε chain has been crossed in the schematic of the human exon map (Fig. 3.6).

The lengths of the introns and exons are indicated in the schematic for the human and *M. domestica* CD3ε chains. The *M. eugenii* and *M. domestica* CD3ε sequences, which were identified in the ensembl database, extended over a number of scaffolds. A clear identification of the introns in *M. eugenii* was not achieved since there were a number of unidentified bases designated 'NNN' in ensembl. The human CD3ε gene consists of nine exons while the predicted *M. domestica* CD3ε gene appeared to consist of six exons. According to ensembl, the *M. eugenii* CD3ε sequence also consisted of six exons. Due to the close phylogenetic relationship of *O. fraenata* and *M. eugenii* it was assumed that the CD3ε gene in *O. fraenata* also consists of six exons. According to the ensembl database, the *M. domestica* exon-1 consists of 49bp, exon-2 contains 30bp, exon-3 consists of 219bp, exon-4 has 165bp, exon-5 has 47bp and exon-6 has 57bp. A comparison of the exon structures of the human and the four marsupial CD3ε genes revealed that the human exon-3 was missing in *M. eugenii* and in *O. fraenata*, while *L. hirsutus* and *M. domestica* both had a very small exon-3 consisting of only four and five amino acids. All four marsupial species were missing nine amino acids in exon-4 compared to the human CD3ε chain (Fig. 3.6, supporting data are in Appendix 3A).

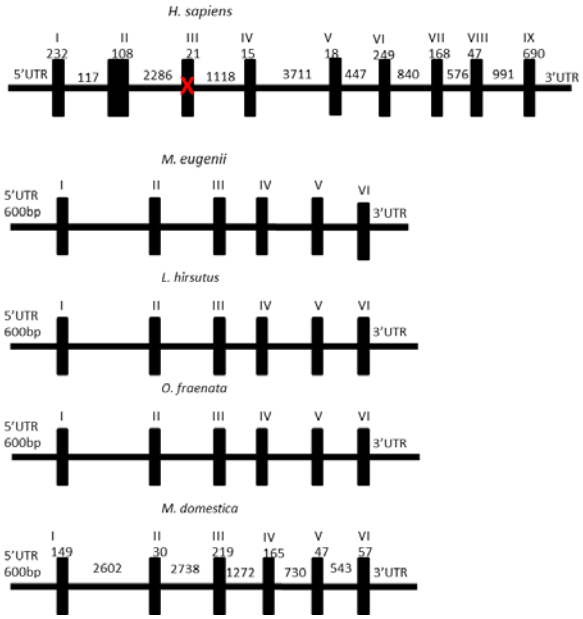


Figure 3.6. Schematic for *M. eugenii*, *L. hirsutus*, *O. fraenata* and *M. domestica* CD3ε genes depicting exonic and intronic sequences. ✕ indicates exon missing in marsupials. Exons are indicated by roman numerals. The arabic numerals indicate the number of base pairs in the intronic sequences for human and *M. domestica*.

3.4.4 T cell receptor alpha (TCRα)

The T cell receptor alpha chain was characterized in two marsupial species *M. eugenii* and *O. fraenata*. Important structural motifs were identified and investigated for amino acid substitutions and sequence homology. Comparisons were made with the published sequences of *T. vulpecula* and *M. domestica*.

3.4.4.1 TCRα - Homology

A homology search across the putative protein and the nucleotide sequences of *O. fraenata* and *M. eugenii* indicated a conservation of >75% and >55% respectively between those species. However the TCRα molecule was only 40% conserved at the putative amino acid level in the two marsupials when compared to other mammals (Table 3.11).

Low complexity sequences were detected thus lowering the identity score at the amino acid level. The *O. fraenata* sequence had 11 amino acid residues of low complexity while the *M. eugenii* sequence was devoid of low complexities. The nucleotide sequence was highly variable in the region amplified and therefore the amino acid percentage was much lower than expected. Up to 14% of sequence gaps were detected in some of the alignments generated by the BLAST algorithms, again lowering the identity percentage values. The sequence conservation at the nucleotide level was the highest between *O. fraenata* and *T. vulpecula* at 81%, while at the amino acid level the conservation was only 66% and was associated with the high sequence variation between those two species. The identity level between *O. fraenata* and *M. domestica* was 77% at the nucleotide level and 55% at the amino acid level also due to high sequence variations. The *M. eugenii* sequence had an 80% identity to the *T. vulpecula* sequence at the nucleotide level and 64% identity at the amino acid level. Comparison of *O. fraenata* and *M.eugenii* to other mammals such as *H. sapiens*, *R. norvegicus* and *M. mulatta* revealed insufficient similarities to give an e-value that was of statistical significance at the nucleotide level. Those three mammalian species, although recognized at the nucleotide level, did not return any e-values for the amino acid sequence and were not recognized at the amino acid level in the BLASTx search due to degeneracy in the genetic code (Table 3.11).

Table 3.11. Homology search results for the *O. fraenata* and *M. eugenii* partial TCRα nucleotide and amino acid sequences, their identities, and their respective e-values. The values are calculated over 295 bp/97 aa length of the partial sequence.

TCRα homology search results for *O. fraenata* and *M. eugenii*					
Species	To	Nucleotide	e-value	Amino acid	e-value
O. fraenata	*T.vulpecula*	81%	1e-71	66%	2e-37
	M.domestica	77%	2e-43	55%	5e-27
	H.sapiens	-------	-------	43%	3e-17
	M.mulatta	-------	-------	40%	3e-17
	R.norvegicus	-------	-------	40%	1e-16
M. eugenii	*T.vulpecula*	80%	1e-65	64%	6e-35
	M.domestica	75%	7e-36	55%	4e-29
	H.sapiens	-------	-------	42%	6e-17
	R.norvegicus	-------	-------	41%	9e-16
	M.mulatta	-------	-------	41%	9e-18

------ = not recognized by the BLAST algorithms.

3.4.4.2 TCRα - Structural domains and motifs

The Protein family (Pfam) domain DUF1968 was predicted in the partial sequences of *O. fraenata* and *M. eugenii* from position 1 to position 68 in *O. fraenata* and from position 1 to position 62 in *M. eugenii.* The statistically important e-values for this domain were 2.40e-13 for *O. fraenata* and 2.40e-12 for *M. eugenii*. This domain is found in most mammalian T cell antigen receptors and adopts an immunoglobulin-like beta-sandwich fold.

The structurally important transmembrane region and connecting peptide motifs were also identified in *O. fraenata* and *M. eugenii* TCRα sequences and when compared to the Genbank sequences of *M.domestica*, *T. vulpecula* and *H. sapiens* variations were detected. The transmembrane regions of both *O. fraenata* and *M. eugenii* had identity percentages of 77% to *T. vulpecula,* 80% to *M. domestica* but only 44% to human. The connecting peptide of both *O. fraenata* and *M. eugenii* had identity percentages of 100% to *T. vulpecula,* 87% to *M. domestica* but only 75% to human (Table 3.12).

Table 3.12. Sequences and conservation percentages of functional motifs of the TCRα chain in *O. fraenata* and *M. eugenii* compared to *T. vulpecula, M. domestica and H. sapiens.*

Functional motifs for TCRα				
Species	Functional motif	Conservation %		
		T. vulpecula	*M. domestica*	*H. sapiens*
O .fraenata	FLRVIFLKT(Transmembrane region)	77%	80%	44%
	FETDKDLN (Connecting peptide)	100%	87%	75%
M. eugenii	FLRVIFLKT(Transmembrane region)	77%	80%	44%
	FETDKDLN (Connecting peptide)	100%	87%	75%
M. domestica	VLRIIFLKT (Transmembrane region)	88%	100%	44%
	FETDRDLN (Connecting peptide)	87%	100%	75%
T. vulpecula	FLRIIFLKT (Transmembrane region)	100%	88%	44%
	FETDKDLN (Connecting peptide)	100%	87%	75%
H. sapiens	GFRILLLKV (Transmembrane region)	44%	44%	100%
	FETDTNLN (Connecting peptide)	75%	75%	100%

3.4.4.3 TCRα - Glycosylation and glycation sites

A single putative O-linked glycosylation site was predicted in the *O. fraenata* TCRα sequence at position 24, while in *M. eugenii* two putative O-linked glycosylation sites

were identified at positions 4 and 19. In *O. fraenata,* the threonine at position 24 only just exceeded the threshold with a probability of 50.7%. This might not be sufficient for it to function as a glycosylation site. In *M. eugenii,* the threonine at position 4 had a probability of 58.8%. However the threonine at position 19 just exceeded the threshold with a probability of 50.5% and therefore may not function as a glycosylation site.

Two putative glycation sites were predicted at positions 44 and 82 in the partial sequence of *O. fraenata* compared to four putative glycation sites at positions 39, 46, 60 and 77 in the *M. eugenii*. It therefore appeared that *M. eugenii* had a higher capacity to non-enzymatically bind sugars. However validation of the results requires more sophisticated mass spectrometry analysis which was outside the scope of this study.

Four putative N-linked glycosylation sites were predicted in both *O. fraenata* and *M. eugenii* TCRα sequences. The putative N-linked glycosylation site had a probability of 73.58% and was located at position 19 in O*. fraenata* and had a probability of 64.2% at position 12 in *M. eugenii*. The NPTV site was identified in both species where a proline residue occurred just after an asparagine residue. This made it highly unlikely that the asparagine was glycosylated due to conformational constraints. This reduced the number of possible glycosylation sites to three in each species (Table 3.13).

Table 3.13. Putative N-linked glycosylation site sequences, positions and probabilities in TCRα of *O. fraenata* and *M. eugenii*.

Putative N-linked glycosylation sites in TCRα of *O. fraenata* and *M. eugenii*			
Species	Sequence	Position	Probability
O. fraenata	NGTN	19	73.58%
	NPTV	22	62.78% ✕
	NSSC	46	55.43%
	NISL	88	61.80%
M. eugenii	NTSG	12	64.2%
	NPTV	17	69.06% ✕
	NISL	83	62.06%

(The proline-containing site is shown in red).

3.4.4.4 TCRα - Phosphorylation sites

The *O. fraenata* partial TCRα sequence had six predicted serine, two predicted threonine and one predicted tyrosine phosphorylation sites. The *M. eugenii* partial TCRα sequence had the same number of predicted serine phosphorylation sites, one predicted threonine and one predicted tyrosine phosphorylation site. The positions of these sites within the partial TCRα sequences are shown in Table 3.14.

In both *O. fraenata* and *M. eugenii* the predicted phosphorylated serine, threonine and tyrosine residues were all above the threshold of 50%. The predicted serine residue in position 9 of *M. eugenii* had the highest probability of 98.6%, while the predicted tyrosine residue in *O. fraenata* had the lowest probability of 54.7% (Table 3.14).

Table 3.14. Positions and probabilities of predicted amino acid phosphorylation sites in the partial TCRα sequences of *O. fraenata* and *M. eugenii.*

Predicted phosphorylation sites of TCRα in *O. fraenata* and *M. eugenii*					
O. fraenata			*M. eugenii*		
Amino Acid	Position	Probability	Amino Acid	Position	Probability
Serine	14	93.4%	Serine	9	98.6%
	33	64.6%		14	81.8%
	35	79.2%		28	75.9%
	38	97.8%		30	93.2%
	48	98.2%		33	87.2%
	77	66.6%		72	66.6%
Threonine	68	54.7%	Threonine	75	69.1%
	80	69.1%			
Tyrosine	36	86.2%	Tyrosine	59	97.3%

A phosphokinase binding site in the TCRα molecule was predicted for both species. This binding site was located in *O. fraenata* at position 101, and in *M. eugenii* at position 96 with probabilities of 76% and 77% respectively.

3.4.4.5 TCRα - Disulphide bond predictions

A single putative disulphide bond was predicted in the partial TCRα sequences in both *O. fraenata* and *M. eugenii*. In both species the predicted disulphide bridge appeared to begin at C^{5} (C=cysteine) and connect to C^{69}.

3.4.4.6 TCRα - Primary sequence and secondary structure prediction

The *O. fraenata* partial sequence of 308 nucleotides translated into 102 putative amino acids, while for *M. eugenii* the partial sequence of 295 nucleotides translated into 96 putative amino acids.

The secondary structure prediction indicated that the distribution of helices and beta sheets in the TCR α chain differed between the two species. *O. fraenata* had four beta sheets and one helix and the *M. eugenii* contained three beta sheets and two alpha helices over the same length of sequence. Antigenic binding sites were found in the TCRα sequences of both species (Figs. 3.7 and 3.8).

Figure 3.7. *O. fraenata* partial TCRα chain primary sequence and secondary structure prediction. N-linked glycosylation sites are highlighted in yellow underlined. **C** = Disulfide bridge. O = Antigenic binding sites. Marks the differences in structure compared to *M. eugenii*. Green highlight = the connecting peptide. Blue highlight = the transmembrane region. = Transmembrane helix. = Strand.

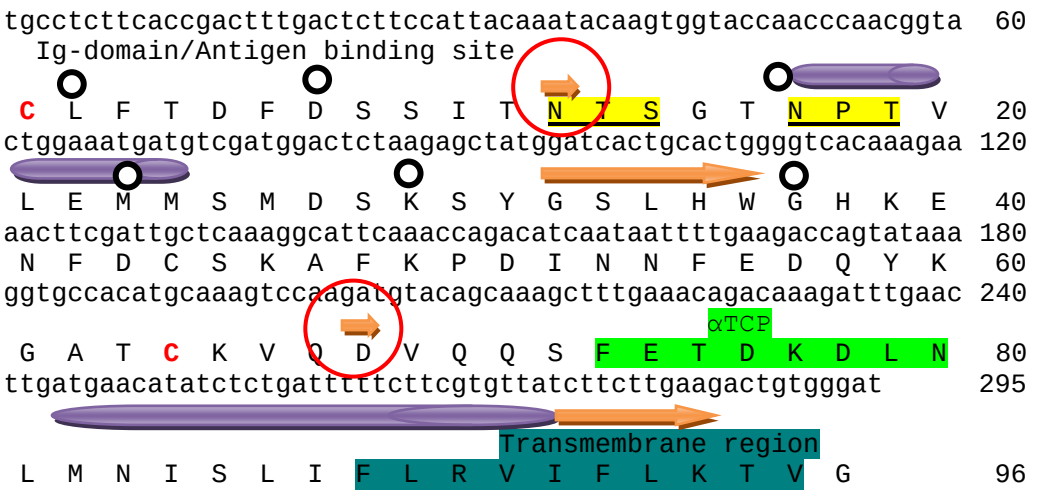


Figure 3.8. *M. eugenii* partial TCRα chain primary sequence and secondary structure prediction. N-linked glycosylation sites are highlighted in yellow underlined. **C** = Disulfide bridge. O = Antigenic binding sites. Marks differences in structure compared to *O. fraenata*. Green highlight = the connecting peptide. Blue highlight = the transmembrane region. = Transmembrane helix. = Strand.

It was found that the first two beta strands in the *O. fraenata* TCRα sequence were quite long compared to the single beta strand found in *M. eugenii*. In the *M. eugenii* sequence an alpha helix was found following a very short beta strand which was absent in the *O. fraenata* sequence. This may indicate different fold properties compared to other marsupial TCRα sequences.

3.4.4.7 TCRα - Structure modelling

The homology models constructed from the *O. fraenata* and *M. eugenii* partial sequences of the TCR alpha chain, together with the positions of two phenylalanine residues that are in close proximity to each other, are illustrated in Fig. 3.9.

The models indicated that the ring structures of the phenylalanine residues in both species were in very close proximity. In *O. fraenata* they point towards each which may indicate steric hindrance. In *M. eguenii* the two phenylalanine residues point away from each other indicating a lack of steric hindrance which facilitates a better folding pattern.

The alignment of the *M. eugenii* phenylalanine residues may be due to the different biochemical environment surrounding those residues.

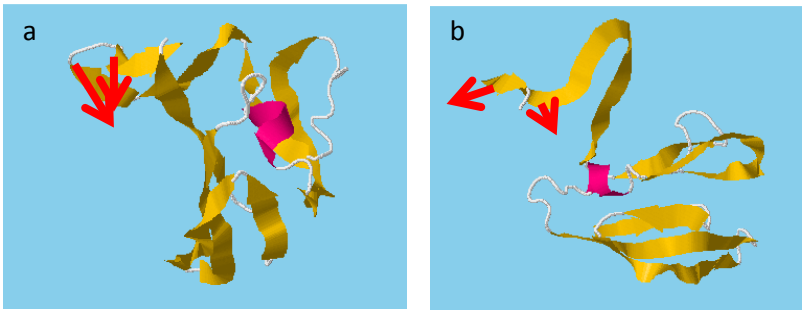


Figure 3.9. Putative tertiary structures of TCRα. (a) *O. fraenata* - arrows indicate the direction of the ring structure of the phenylalanine residues. (b) *M. eugenii* - arrows indicate the direction of the ring structure of the phenylalanine residues.

3.4.4.8 TCRα - Phylogenetic analysis – connecting peptide

The evolutionary history of the marsupial TCRα connecting peptide was inferred by using the Maximum Likelihood method based on the Dayhoff matrix model to produce a phylogenetic tree.

3.4.4.8.1 TCRα - Phylogenetic analysis – Connecting peptide of TCRα

The program produced a phylogenetic tree with the highest logarithmic scale likelihood (-131.22). The percentage of trees in the resampling process in which the associated taxa are clustered together is shown next to the branches. The evolutionary rate difference among sites had a gamma distribution of 0.18 (5 categories). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 10 amino acid sequences. There were a total of 10 positions in the final dataset (Fig. 3.10).

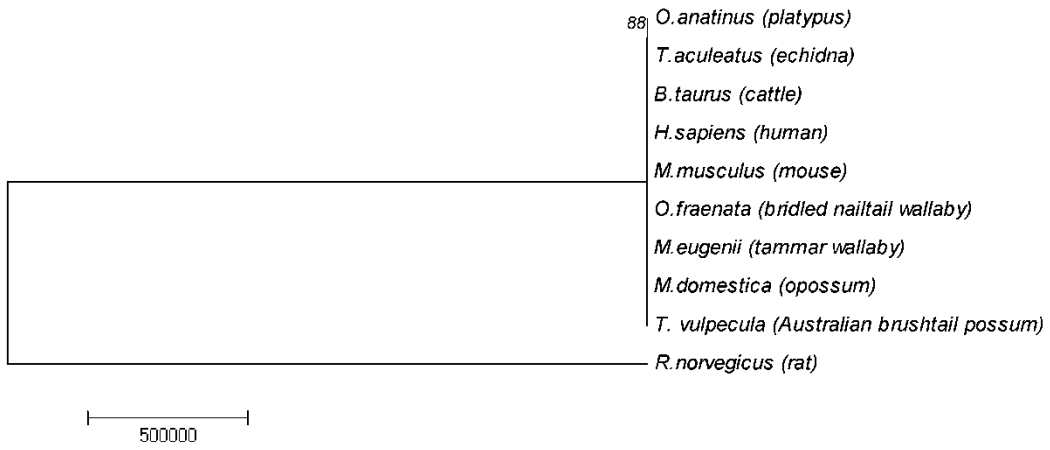


Figure 3.10. Maximum likelihood phylogenetic tree for connecting peptide of the marsupial TCRα gene. A Dayhoff algorithm was used for this motif because of the high sequence conservation.

3.4.4.8.2 TCRα - Phylogenetic analysis - Transmembrane region of TCRα

The program produced a phylogenetic tree with the highest logarithmic scale likelihood (-11,470.38). The percentage of trees in which the associated taxa are clustered together is shown next to the branches. The evolutionary rate difference among sites had a gamma distribution of 200.0000 (5 categories). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 10 amino acid sequences. There were a total of 9 positions in the final dataset (Fig. 3.11).

The high conservation of the transmembrane region in both *M. eugenii and O. fraenata* is clearly shown in this tree. It was evident that the marsupial sequences were closely related to the monotreme sequences. Unexpectedly, the monotremes were embedded within the marsupials in this sequencing motif. This is unusual, since *M. domestica* and *T. vulpecula* are not close relatives and neither of them are closely related to the monotremes.

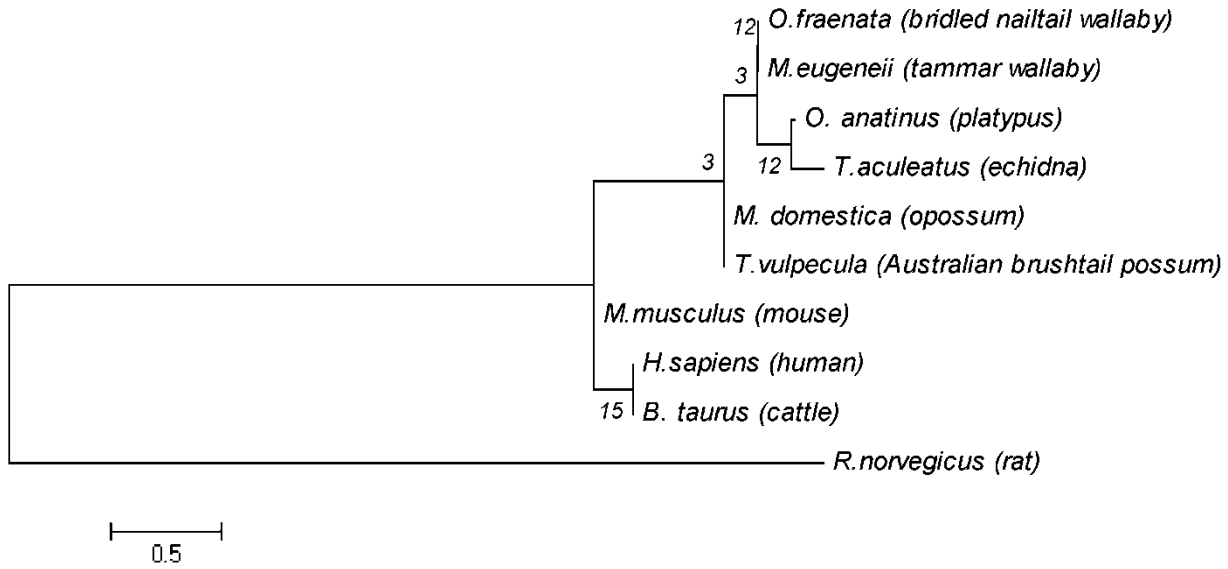


Figure 3.11. A Maximum likelihood phylogenetic tree of the transmembrane region found in the marsupial TCRα gene using a Dayhoff algorithm.

3.4.5 T cell receptor beta chain (TCRβ)

The full TCRβ constant domain was characterized in *O. fraenata* and a partial sequence of the domain was amplified in *M. eugenii*. Important functional motifs were identified and investigated for amino acid substitutions and sequence homology.

3.4.5.1 TCRβ - Homology

The *M. eugenii* partial sequence of the TCRβ chain had a sequence identity of 87% to *T. vulpecula* at the nucleotide level and 91% at the amino acid level. The *O. fraenata* sequence also had a high identity percentage at the nucleotide level to *T. vulpecula* (83%) and an 84% sequence identity at the amino acid level. The identity percentage to the sequence of *M. domestica* was lower at the amino acid level for both *M. eugenii* and *O. fraenata* (84% and 75% respectively) (Table 3.15).

Both *M. eugenii* and *O. fraenata* showed an e-value of 0 to the *T. vulpecula* sequence indicating a significant match between the two. It appeared that the marsupial sequences did not have a high homology at the nucleotide level to mammalian species such as *M. musculus*, *R. norvegicus*, *S. scrofa* and *M. mulatta* and, in some cases at the amino acid level (Table 3.15).

Table 3.15. Homology search results for the *M. eugenii* and *O. fraenata* partial TCRβ nucleotide and amino acid sequences, their identities, and their respective e-values. The values are calculated over the 818bp/186 aa length of the partial sequence.

TCRβ sequence homology search results for *O. fraenata* and *M. eugenii*								
	M. eugenii				*O. fraenata*			
Species	Nucleotide	e-value	Amino acid	e-value	Nucleotide	e-value	Amino acid	e-value
T. vulpecula	87%	0.0	91%	2e-122	83%	0.0	84%	2e-109
M. domestica	86%	0.0	84%	3e-106	78%	3e-173	75%	5e-89
H. sapiens	73%	2e-80	67%	2e-83	73%	2e-60	61%	1e-66
M. musculus	72%	1e-81	64%	3e-78	-----	-----	57%	6e-62
R. norvegicus	72%	5e-81	64%	1e-78	-----	-----	-----	-----
S. scrofa	-----	-----	67%	3e-81	-----	-----	59%	2e-66
M. mulatta	-----	-----	65%	3e-80	-----	-----	59%	7e-64

----- = not recognized by the BLAST algorithms.

3.4.5.2 TCRβ - Structural domains and motifs

A putative IGc1 domain and a predicted transmembrane domain were identified in the partial sequences of both *O. fraenata* and *M. eugenii*. The IGc1 domain had an e-value of 2.04e-18 in the *O. fraenata* sequence and 1.09e-18 in the *M. eugenii* sequence. The predicted e-value of the transmembrane domain was too low to be identified.

A high identity score at the putative amino acid level was found in the loop motif sequences of both species. Of the three important structural motifs which were identified, the F and G-loops had the highest identity percentage at the putative amino acid level. The loop structures showed varying degrees of conservation. The Cβ-loop was the lowest conserved structure with 50% identity, while the F loop was the highest with 92% identity. The G-loop had a conservation percentage identity of 88%. The structural regions in the TCRβ chain of *O. fraenata* and *M. eugenii* were compared to the sequences of *H. sapiens, T. vulpecula* and *M. domestica* and the results are shown in Table 3.16.

Table 3.16. Conservation percentages of F-loop, Cβ elbow loop, and G-loop of the TCRβ chain in *O. fraenata*, *M. eugenii*, *T. vulpecula*, *M. domestica* and *H. sapiens.*

TCRβ structural motifs and their respective conservation percentages				
Species	Functional motif sequence	Conservation %		
		T. vulpecula	*M. domestica*	*H. sapiens*
O. fraenata **(not yet deposited in Genbank)**	KNSFRCQVLFHGI (F-loop) GENETW (Cβ elbow loop) PITRNVSDQIWE (G-loop)	92% 66% 83%	77% 83% 75%	62% 50% 58%
M. eugenii **(not yet deposited in Genbank)**	KNSFRCQVLFHGI (F-loop) GENETW (Cβ elbow loop) PITQNVSDQIWG (G-loop)	92% 66% 100%	77% 83% 92%	62% 50% 58%
L. hirsutus **(not yet deposited in Genbank)***	KNSFRCQVLFHGI (F-loop) EENEIW (Cβ elbow loop) PITQNVSDQIWG (G-loop)	92% 66% 100%	84% 83% 92%	62% 50% 58%
T. vulpecula **(Accession Number AF133098)**	KNSFRCQVLFNGI (F-loop) SENEPW (Cβ elbow loop) PITQNVSDQIWG (G-loop)	100% 100% 100%	92% 67% 92%	62% 66% 58%
M. domestica **(Accession Number AY014506)**	KNNFRCQVLFNGI (F-loop) EENETW (Cβ elbow loop) PVTQNVSDQIWG (G-loop)	92% 66% 92%	100% 100% 100%	62% 50% 66%
H. sapiens **(Accession Number AY232284)**	RNHFRCQVQFYGL (F-loop) SENDEW (Cβ elbow loop) PVTQIVSAEAWG (G-loop)	62% 66% 58%	62% 50% 66%	100% 100% 100%

*(characterized by K. Howard)

3.4.5.3 TCRβ - Glycosylation and glycation sites

Three putative O-linked glycosylation sites were predicted in the partial *O. fraenata* TCRβ sequence. These sites were located at positions 26, 29 and 34 with probabilities of 54%, 58% and 52%, respectively. *M. eugenii* had seven putative O-linked glycosylation sites located at positions 4, 9, 13, 16, 21 and 82 with probabilities of 63%, 65%, 60%, 60%, 60%, 54% and 51% respectively.

Five putative glycation sites were predicted in the *O. fraenata* partial TCRβ sequence compared to six putative sites in *M. eugenii*.

Five putative N-linked glycosylation sites were predicted in the *O. fraenata* partial TCRβ sequence at locations 7, 92, 128, 134 and 142 with probabilities of 61.39%, 48.72%, 58.17%, 72.88% and 69% respectively. Three putative N-linked glycosylation sites were predicted in the *M. eugenii* partial TCRβ sequence at locations 79, 115 and 129 with probabilities of 47.77%, 58.63%, and 74.1% respectively. The sites located at positions 92 in *O. fraenata* and 79 in *M. eugenii* did not reach the probability threshold of 50% and therefore were not recognized as viable sites (Table 3.17).

Table 3.17. Predicted N-linked glycosylation sites, their locations, and their probabilities in the *O. fraenata* and *M. eugenii* partial TCRβ sequences.

Putative N-linked glycosylation sites in *O. fraenata* and *M. eugenii* TCRβ			
Species	Sequence	Location	Probability
O. fraenata	NYSE	7	61.39%
	NFSR	92	48.72%
	NETW	128	58.17%
	NLTK	134	72.88%
	NVSD	142	69.00%
M. eugenii	NFST	79	47.77%
	NETW	115	58.63%
	NVSD	129	74.10%

Note: The sequences that did not reach the probability threshold are shown in red.

3.4.5.4 TCRβ - Phosphorylation sites

Ten predicted serine phosphorylation sites were found in *O. fraenata* and nine in *M. eugenii*. Two predicted threonine phosphorylation sites were found in both species

but only one putative tyrosine phosphorylation site was predicted in *M. eugenii.* The *O. fraenata* TCRβ chain did not have a putative tyrosine phosphorylation site. The serine in *O. fraenata* at position 97 had a probability of 50.3%. This is marginally above the threshold and this residue may not be phosphorylated. A summary of the phosphorylation sites is presented in Table 3.18.

Table 3.18. Predicted serine, threonine and tyrosine amino acid phosphorylation sites in the TCRβ sequences of *O. fraenata* and *M. eugenii* including their position and probabilities.

Phosphorylation sites of TCRβ in *O. fraenata* and *M. eugenii*						
Species	Serine	Probability	Threonine	Probability	Tyrosine	Probability
O. fraenata	20 39 79 85 94 97 100 105 115 144	99.6% 99.7% 98.2% 99.2% 82.2% 50.3% 53.7% 92.8% 98.9% 75.0%	29 74	93.3% 81.8%	147	82.4%
M. eugenii	26 66 72 87 92 102 143 146 186	99.7% 95.3% 99.2% 61.0% 92.8% 98.9% 53.4% 73.8% 86.6%	7 16	92.1% 89.6%	none	none

3.4.5.5 TCRβ - Disulphide bond prediction

A single putative disulphide bond was predicted for the *O. fraenata* TCRβ partial sequence at position C^{53} connecting with C^{153}. This prediction had a probability of 99.7%. A single putative disulphide bond was also predicted for the *M. eugenii* TCRβ partial sequence at position C^{40} connecting with C^{105}, a long bond spanning 65 amino acids.

3.4.5.6 TCRβ - Primary sequence and secondary structure prediction

A secondary structure prediction determined the distribution of helices and beta sheets for both *O. fraenata* and *M. eugenii* TCRβ chains (Figs.3.12 and 3.13). The confidence

levels of the helix and strand predictions at >80% were high for both species (Appendix 3A.1, 3A.3 and 3A.4).

In *O. fraenata,* a product of 576bp was amplified translating into 191 putative amino acids, while for *M. eugenii* a product of 561bp was amplified translating into 186 putative amino acids. A polyA sequence was detected in both species in the 3' region. It appeared that this was not the actual polyA tail since no polyadenylation signal was detected prior to the polyA sequence.

The helix in the extracellular domain in *O. fraenata* was much shorter than in *M. eugenii*. A helix preceding the F-loop in *O. fraenata* was not found in *M. eugenii.* There were different distributions of beta strands in the F-loop and in the Cβ-elbow loop between *O. fraenata* and *M. eugenii*, which suggested that the molecules may have different fold properties.

Transmembrane helices are stable structures in cell membranes and can be predicted with the three stage model of membrane protein folding (Cuthbertson *et al.,* 2005). The transmembrane helices were predicted with two different programs and since the predictions were not always accurate they were investigated for their hydrophobicity. The *O. fraenata* TCRβ sequence had a predicted transmembrane helix at positions 160 to 182 and in *M. eugenii* the helix was located at positions 155 to 177. The *M. eugenii* and *O. fraenata* transmembrane helices revealed an inside helix cap and an outside helix cap, both with a length of 6 amino acids followed by an inside loop.

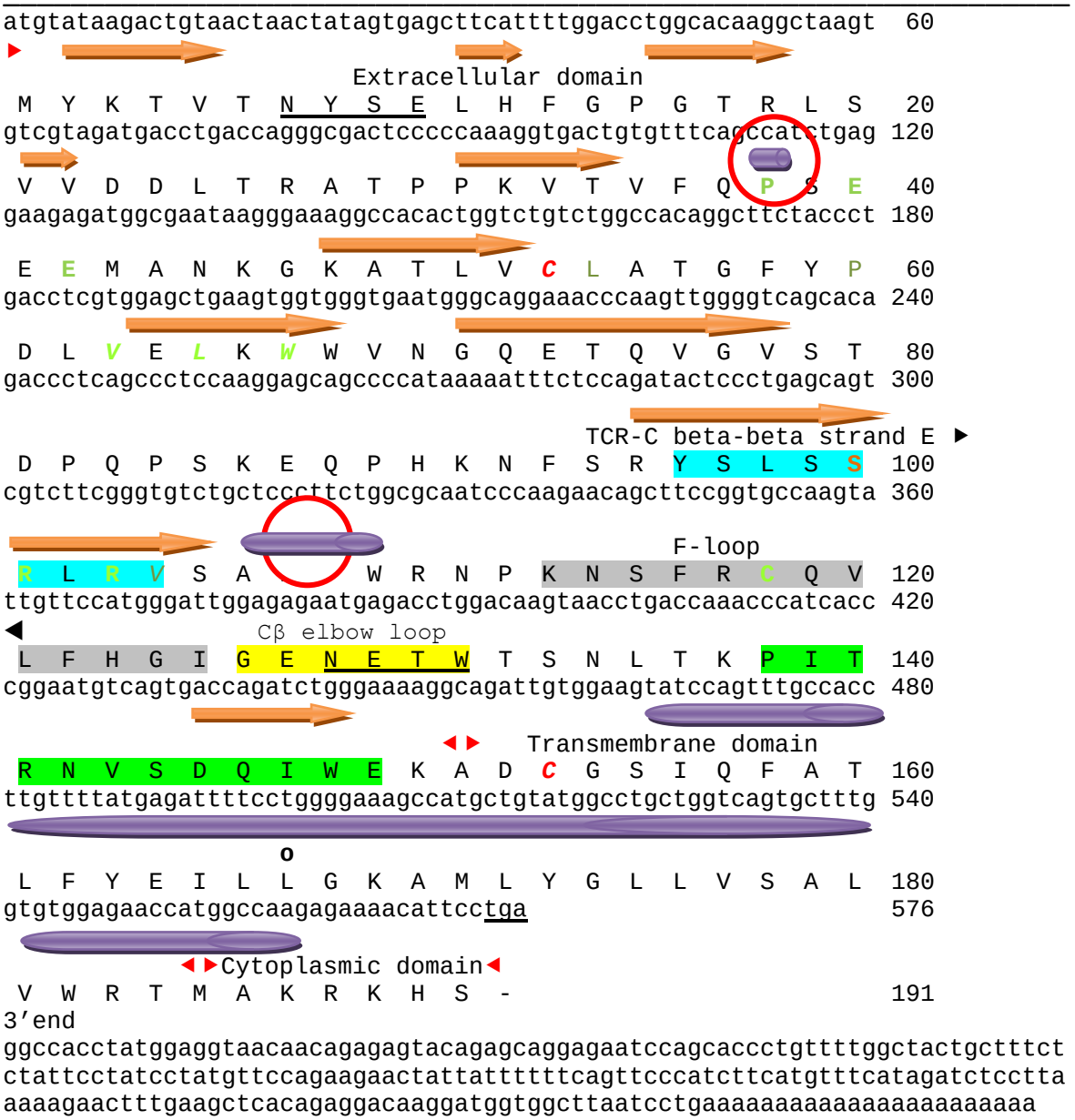


Figure 3.12. *O. fraenata* TCRβ chain primary sequence and secondary structure prediction. **P =** residue important in the loop architecture. — = β strand characteristic residues pointing inside the sheet. — = amino acids involved in domain-domain interaction. Black underlined amino acids indicate a glycosylation site. C = disulphide bridge. Letters in italics show hydrophobic residues buried in the domain core. Bold letters are conserved residues. ▶◀ = involved in TCRα – TCRβ interactions. ◀ ▶ = the start and end of a domain. Highlighted in yellow = Cβ loop. Highlighted in gray = F-loop. Highlighted in light blue = TCR-C beta-beta strand E. o = conserved leucine motif involved in signal transduction. = Transmembrane helix. = Strand.

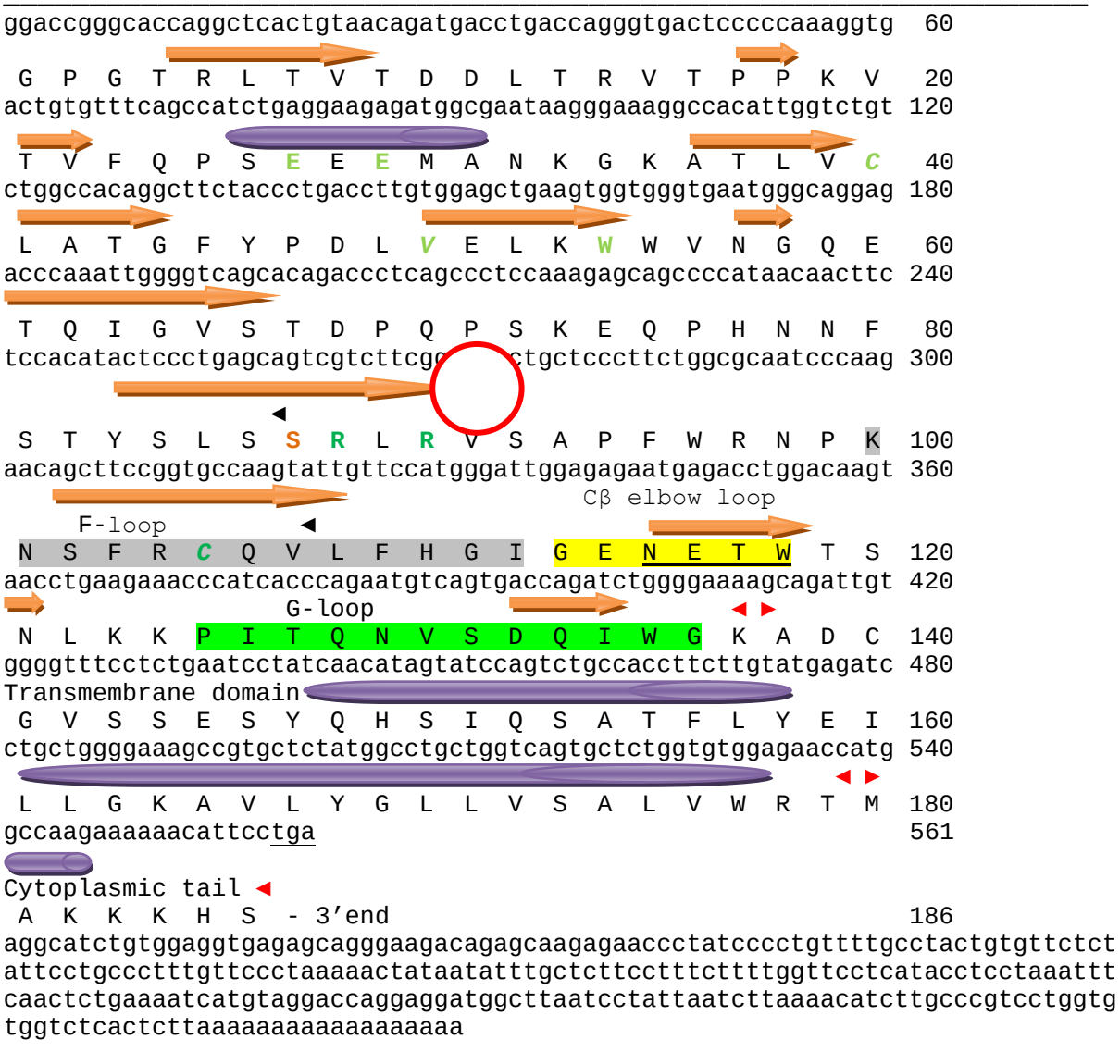


Figure 3.13. *M. eugenii* TCRβ chain partial primary sequence and secondary structure prediction. **P=** residue important in the loop architecture. — = β strand characteristic residues pointing inside the sheet. — = amino acids involved in domain-domain interaction. Black underlined amino acids show a glycosylation site. C = disulphide bridge. Letters in italics show hydrophobic residues buried in the domain core. Bold letters are conserved residues. ▶◀ = involved in TCRα – TCRβ interactions. ◀ ▶ = the start and end of a domain. Highlighted in yellow = Cβ loop. Highlighted in gray = F-loop. Highlighted in light blue TCR-C beta-beta strand E. o = conserved leucine motif involved in signal transduction. = Transmembrane helix. = Strand.

3.4.5.7 TCRβ - Phylogenetic analysis - FG-loop

The Neighbor-Joining tree produced with a Dayhoff algorithm showed a branch length sum of 6.26. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. The branch length is shown in the same units as those of the evolutionary distances used to infer the phylogenetic tree and therefore the tree is drawn to scale. The evolutionary distances are in the units of the number of amino acid substitutions per site. The analysis involved 18 amino acid sequences, and all positions containing gaps and missing data were eliminated leaving a total of 19 positions in the final dataset (Fig. 3.14).

The marsupial clade is clustered together and is located between the human and the amphibian *Ambystonum mexicanum*.

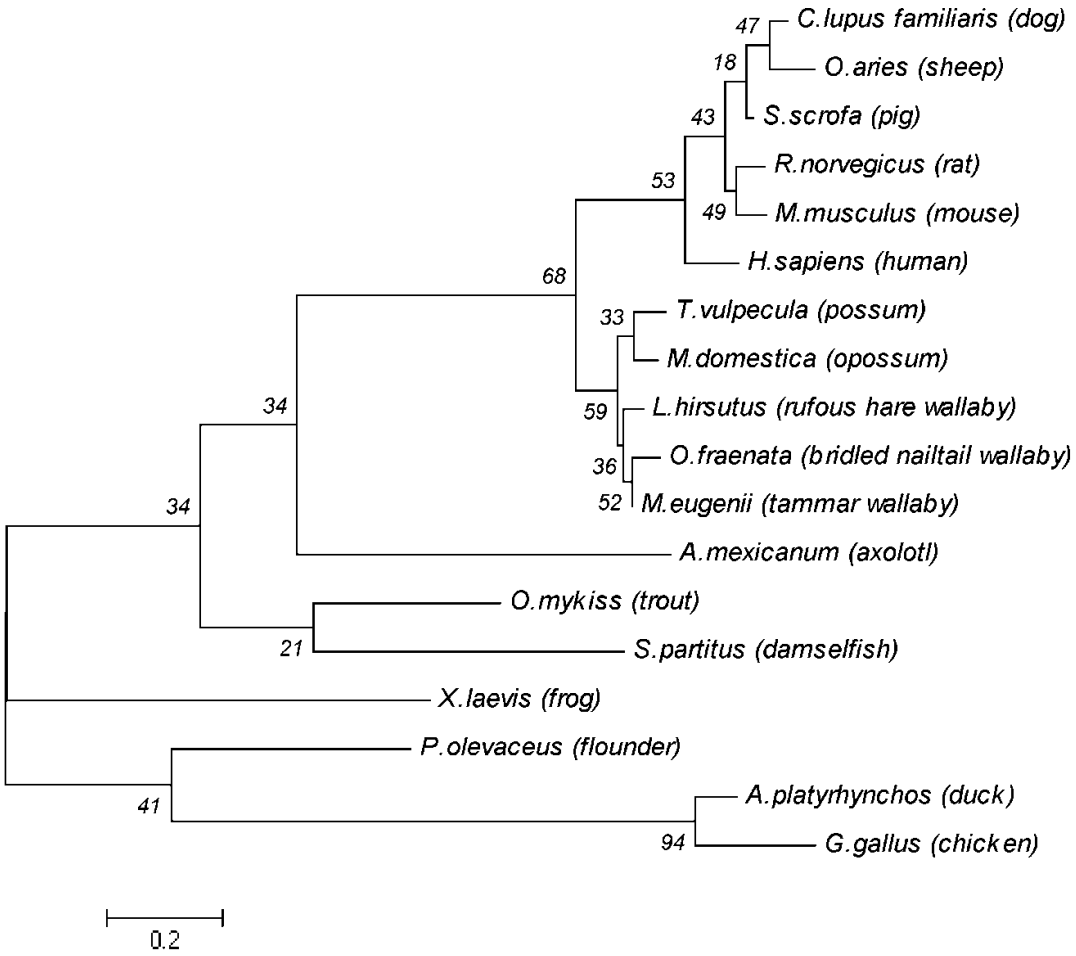


Figure 3.14. Neighbor-Joining tree using the Dayhoff algorithm for the FG loop of the marsupial TCRβ. Branch lengths give an indication of the evolutionary relationship between the species.

3.5 Discussion

The sequence of the CD3ε chain in *M. eugenii* has been published (Old *et al.,* 2001). In the present study, this molecule was characterized in *O. fraenata, L. hirsutus* and *M. domestica* for the first time thus allowing a comparison of this molecule's structural motifs. It is evident from the results that CD3ε is a highly conserved molecule in those species as well as in other mammals. The four marsupials are located in the evolutionary tree between the primates and avian species. It is important to extend our knowledge of the marsupial immune system since marsupials are extant species and can provide an insight into changes of the immune system which have occurred over time. It also provides an opportunity to investigate why some of these animals are more disease resistant than others by allowing comparison of important immune system molecules within the marsupial clade to those of humans.

Until this study was undertaken, the *M. domestica* CD3ε sequence was only predicted (Genbank Accession No. XM_001380690). In this study a number of differences were found between the *M. domestica* CD3ε prediction in ensembl (http.//www.ensembl.org/) and the sequence of the expressed gene. An alignment between the predicted and the expressed sequence was carried out (supporting data are in Appendix 3A) and it appeared that the predicted sequence was missing five amino acids of which three are situated within the Ig-domain of the molecule. Ig-domains are found in the CD3 complex chains of all species indicating that this region is responsible for the formation of side-to-side configured dimers such as CD3εγ and CD3εδ (Clements *et al.,* 2006).

The difference in intron length between marsupial and human CD3ε genes would have resulted from evolutionary changes over time. Intron-exon structures may therefore offer a means of gaining evolutionary information, however these matters are still the subject of debate (Yandell *et al.,* 2006). Since the TCR interaction sites in the marsupial CD3ε molecule are little different to that of other mammals it appears that the function of the molecule is not impaired. All important amino acid residues necessary for the function of

the molecule are in locations in the molecule similar to those of humans thus suggesting that the marsupial CD3ε, despite some differences, functions in the same manner.

Examination of the secondary structures showed that, while the structures in *O. fraenata* and *L. hirsutus* were almost identical to each other, they differed to that of *M. eugenii,* and were markedly different to the structure of *M. domestica*. The most interesting difference found was an extra helix in the cytoplasmic domain of the molecule in *M. domestica* (Fig. 3.3). The beta sheet in the extracellular domain appeared to be larger and an additional beta sheet was found in the transmembrane region of the *M. domestica* CD3ε molecule when compared with the sequences of *O. fraenata* and *L. hirsutus*. This may indicate that the folding profile of *M. domestica* CD3ε is different to that of the other marsupials and other mammals. Given the evolutionary distance between *M. domestica* and the other marsupials and mammals, its CD3ε structure may reflect an older mammalian version of that molecule. The number of structural differences increased when the marsupial CD3ε molecule was compared to that of bird and fish species.

The CD3 epsilon molecule contains a proline-rich region (PRR) motif which appeared to be highly conserved in the marsupial CD3ε sequences. The proline-rich motif is also a SH_3 binding domain containing a consensus sequence of xPPxP (x =any amino acid and P = proline) (Pawson, 1995) through which binding to the non-catalytic region of the tyrosine kinase adaptor protein (Nck) occurs. It is this ligand binding that causes a conformational change in the cytoplasmic domain of the CD3ε chain and exposes the proline-rich region (PRR) (Manolios *et al.,* 1997). Mouse studies have indicated that this motif functions in the degradation of the TCRζ chain resulting in the down regulation of TCR expression on the cell surface (Mingueneau *et al.,* 2008). This is an important function in T cell regulation and the high conservation of this proline-rich motif in marsupials suggests that the same processes occur in the marsupial CD3ε molecule. This motif in *M. eugenii*, *O. fraenata* and *M. domestica* was shown to be 100% conserved compared to humans indicating this to be of high importance in the functionality of the marsupial CD3ε.

Proline is an interesting amino acid since it consists of a side chain that folds back toward the backbone and the amide position (Williamson, 1994). This means that the conformation of the proline backbone is restricted due to its bulkiness. This restricts the sequence conformation preceding the proline residue (Hurley *et al.,* 1992). Proline cannot act as a hydrogen bond donor due to the amide proton being replaced by a CH_2 group (Williamson, 1994). The PRR found in CD3ε is non-repetitive and it is thought to be involved in signal transduction (Ren *et al.,* 1993). Examination of the secondary structure prediction of all the marsupial CD3ε chains investigated in this study revealed that no helices were located in the PRR motifs. This may indicate that the number of proline residues inhibit helix formation due to the bulkiness of their side chains. Due to the high conservation of this motif in the marsupial CD3ε sequences it can be assumed that the function of this motif is the same as in other mammals where a similar number of prolines have been found.

The CD3ε chains of *O. fraenata*, *L. hirsutus* and *M. domestica*, were not N-linked glycosylated, unlike those of *Marmota monax* (woodchuck), *Bos taurus* (cattle) (Hagens *et al.,* 1996), *Anas platyrhynchos* (duck) (Kothlow *et al.,* 2005), *Ambystoma mexicanum* (axolotl) (Fellah *et al.,* 1993) and some fish species (Ishiguro *et al.,* 1990, Bello *et al.,* 2009). The N-terminal end of the CD3ε in the three marsupial species differed in the number of acidic residues. *M. domestica* had five, *O. fraenata* had three, and *L. hirsutus* had four such acidic residues. The N-terminal end of this molecule was quite diverse and differs considerably in length in different species. For example, this study found that the metatherian N-terminal consisted of 7 residues in total which is two residues longer than the N-terminal sequence of fish (Bello *et al.,* 2009). The overall acidity of the N-terminal end is important in exposing the PRR motif in the cytoplasmic domain therefore affecting a conformational change in the molecule (Manolios *et al.,* 1997). Investigation of the acidity of the N-terminal in the marsupial CD3ε molecule by determining the side chains of the amino acid residues within the motif revealed an overall acidity. It was therefore concluded that the PRR motif would be exposed through a conformational change.

However this assumption will have to be confirmed with site-directed mutagenesis studies9.

The CD3ε sequences of the marsupial species which were studied were compared with those of other mammals, fish, and amphibians. This comparison revealed a high level of conservation in the transmembrane and cytoplasmic domains as well as in parts of the extracellular domain and this is reflected in the phylogenetic tree (Fig 3.5).

The marsupial CD3ε sequence included a small microdomain containing the CxxCxE motif (C=cysteine and E=glutamate) which is involved in the interaction with the TCR and other CD3 chains. The cysteine residues in this motif stabilizes the secondary structure and flank a 30 amino acid variable motif which forms an immunoglobulin-like domain within the motif (Kirkham *et al.,* 1996). This indicated that the marsupial CD3ε also interacts with the TCR in similar fashion to other mammalian CD3ε chains.

The level of phosphorylation and control of protein-protein interactions are a pre-requisite for intracellular signal transduction (Wandless, 1996). These interactions are mediated by small protein sub-domains such as the src homology 2 (SH_2) domains (Cohen *et al.,* 1995) which have become the hallmark of proteins involved in intracellular signal transduction (Pawson, 1995). The SH_2 domains bind to phosphorylated tyrosine residues such as those found in the CT domain of the CD3ε molecule. The consensus sequence (YxxL/I) (Y = tyrosine, L = leucine and I = Isoleucine) identifies the SH_2 domains that are contained within the ITAM motif of the CD3ε molecule. The tyrosine residues Y^{174} and Y^{185} in *M. domestica*, Y^{169} and Y^{180} in *O. fraenata,* and Y^{172} and Y^{184} in *L. hirsutus* appeared to be the putative phosphorylation sites. Human studies have indicated that these two tyrosine residues are phosphorylated upon TCR-ligand binding (Ishiguro *et al.,* 1990). This phosphorylation event is the beginning of the signalling cascade and leads to thymocyte differentiation (Ishiguro *et al.,* 1990). In this study the molecular investigation of the marsupial CD3ε molecule confirmed the immunohistochemical surveys into T cell populations in marsupial tissues which indicated T cell differentiation and therefore a

highly conserved function (Old and Deane, 2003, Old and Deane, 2002b, Old and Deane, 2000).

A study undertaken by Dave (2009) examined the expression of the proline, ITAM, and the endoplasmic retention motifs at the different developmental stages of T lymphocytes in mice and human. It was found that all those motifs are necessary for the T lymphocytes to progress in their development, and to determine the degree of TCR expression on the cell surface. It was also shown that it is the PRR that is important for VDJ gene rearrangement. This study demonstrated that all of these motifs are conserved in *O. fraenata*, *L. hirsutus*, *M. eugenii* and *M. domestica* thus indicating that the functionality of CD3ε is also conserved in those species. It therefore appeared that the first phase of signal transduction was also intact.

CD3ε is only a small part of the whole TCR-CD3 complex. In order for the CD3ε molecule to fulfill its functions as a signalling molecule some amino acid residues are required to interact with the receptor. For those residues to interact with the receptor, the receptor itself has to be assembled. Important motifs within the receptor, such as the transmembrane region and the core-peptide motif of the alpha chain together with important loops in the beta chain of the receptor. These fulfill key tasks in receptor assembly and subsequently in T cell activation, and are discussed in the following paragraphs.

This study found that the transmembrane region of the TCRα molecule was present in all the marsupial cDNA sequences which were investigated. However, when compared to the human transmembrane sequence, the conservation of 44% in each of the species was quite low (Table 3.7). This is also reflected in the phylogenetic tree (Fig. 3.9) that was constructed to show the evolutionary rate relationship of the marsupial sequences to other mammals.

The human transmembrane region of TCRα contains two positively charged amino acid residues, arginine (R) and lysine (K). It has been reported that these two amino acids are

critical in receptor assembly (Call *et al.,* 2002, Manolios, 1995, Manolios *et al.,* 1991). The arginine residue interacts with the TCRζ chain and plays a major role in T cell receptor assembly. On the other hand, the lysine residue interacts with the CD3δ part of the CD3 complex (Ali *et al.,* 2006) and its deletion alters the charge distribution and prevents the assembly of the receptor complex (Arnaud *et al.,* 1996). Both of these residues were shown to be conserved in the *O. fraenata* and *M. eugenii* TCRα sequences hence the assembly mechanism of the receptor appeared to be functioning in the same manner as in other mammals.

The secondary structure prediction of the marsupial partial TCRα chain carried out with PSIpred (Bryson *et al.,* 2005) (Appendix 3B) indicated that there are differences between the two closely related species *O. fraenata* and *M. eugenii*. It also indicated that the two beta sheets found in the antigenic binding sites of *O. fraenata* stabilized the structure of TCRα in that species. The prediction algorithm awarded a higher confidence level to the *O. fraenata* secondary structure than to that of *M. eugenii*. However this assumption needs to be substantiated by X-ray crystallography.

Another interesting feature found in the marsupial TCRα sequence was the amino acid residue arrangement. In humans, a poly-leucine motif ILLL found within the transmembrane region is critical for the functioning of the molecule. However the marsupial transmembrane region is devoid of this motif. The activity of the transmembrane region is dependent on the poly-leucine motif since it exposes the molecule to lipid interactions. The spatial arrangement of this motif is also important for its function (Ali *et al.,* 2006). In this study it was found that the structural integrity of the poly-leucine motif was not conserved in marsupials. Two phenylalanine residues were found in the marsupial sequence, one of which substitutes for the leucine found in the human transmembrane region. The side chain of phenylalanine is bulkier due to the ring structure of this amino acid. Molecular modelling in mouse studies have shown that the transmembrane region assumes a secondary structure which indicates that the positively charged lysine and arginine residues sit at opposite faces with the hydrophobic leucine

domain between them (Ali *et al.,* 2006). The phenylalanine substitution found in marsupial species may interfere with the topography and spacing of the charged groups. This is important for the functioning of the transmembrane region due to the ring structure of the phenylalanine residue. Modelling of the marsupial TCRα chain indicated that the two phenylalanine residues do indeed compete for space in *O. fraenata.* However the model constructed with the same parameters for *M. eugenii* indicated that the two residues face away from each other and therefore do not compete for space (Fig 3.10). This may mean that the receptor in *O. fraenata* does not work as efficiently as in *M. eugenii* and in other mammals. Resolution of this paradox will require investigation of the torsion on these residues to confirm their alignment and any steric hindrance.

Human studies have shown that the charged residues within the transmembrane region permit the TCRα molecule to exist either in a stabilized form in the membrane or as a soluble molecule (Shin *et al.,* 1993). The charged residues are also the link to the CD3 complex. This study found no evidence of a soluble form in marsupials. Further investigations are needed to solve the structural integrity of this motif in metatherian mammals.

The connecting peptide (αCPM) is another motif located in the transmembrane domain of the TCRα. The αCPM plays an important role in the positive selection of $CD8^+$ and $CD4^+$ T lymphocytes and connects the transmembrane and Ig domains (Backstrom *et al.,* 1996). This motif also plays a role in the interaction of the receptor with the CD3 complex by shaping a precise structure which is involved in the dimerization of the receptor (Arnaud *et al.,* 1997). Deletion of this motif has shown that the receptor does not interact with the CD3 complex to the full extent and signal transduction is interrupted (Backstrom *et al.,* 1997). Seven residues within the connecting peptide are responsible for the interaction with the CD3 complex. The seven main residues of the connecting peptide FETD*X*NLN (human sequence, F = phenylalanine, E = glutamic acid, T = threonine, D = aspartic acid, *X* is any amino acid, N = asparagine, L = leucine), are conserved in a number of mammals and the first four residues are conserved in marsupials. However the NLN portion of the

sequence is replaced with DLN in marsupials and is associated with a substitution of asparagine (N) by aspartic acid (D). Both amino acids are polar but asparagine prefers to be on the surface of proteins and therefore exposed to an aqueous environment (Betts and Russell, 2003). Asparagine contains an amino group instead of one of the oxygen molecules found in aspartate and therefore lacks a negative charge. Both amino acid residues are known for their involvement with both the active sites and binding sites of proteins. This substitution does not appear to hinder the functionality of the molecule in marsupials since both amino acid residues have similar responsibilities in biological systems and one can carry out the function of the other (Betts and Russell, 2003).

Also located in the human transmembrane region are the amino acid residues leucine and phenylalanine which are conserved in a number of species (supporting alignment data are in Appendix 3B). Both of these residues are important for the expression of the TCRα chain as well as for the association with the CD3 complex (Bhatnagar *et al.,* 2003). This study showed that in the species sequence alignment the phenylalanine residue was conserved in *O. fraenata* and *M. eugenii* but not in *M. domstica.*

The leucine was shown to be conserved in the *M. domestica* TCRα sequence while *M. eugenii* and *O. fraenata* had an isoleucine in place of leucine, and *T. vulpecula* had a methionine instead of the leucine. Given the close relationship of leucine and isoleucine, it was assumed that the function of both might either be conserved or be similar. Since methionine has quite different biochemical properties to leucine its function in the connecting peptide may not be conserved in *T. vulpecula*. Mutational studies will shed further light on the functionality of the marsupial TCRα chain.

The TCRα chain interacts with distinct regions of the TCRβ chain (Appendix 3C). One of these regions is located in the TCR-C beta-beta strand E in humans with the sequence YCLSSRLRVSA compared to YSKSSRKRV(I)SA in *O. fraenata* and *M. eugenii*. The cysteine of the human sequence is substituted with a serine in the marsupial motif. Similar substitutions have been reported in other receptors resulting in reduced receptor activity

(Okamoto *et al.,* 2009). If reduced receptor activity due to such substitutions in marsupials can be verified it would support the documented diminished T cell responses in marsupials (Stone *et al*., 1998, 1997a). Since the TCR-C beta-beta strand E motif is also an interaction site with the TCRα chain, this substitution could mean a diminished interaction with the TCRα. The interaction between the α and β chains of the receptor facilitates the necessary conformational changes to prepare for signal transduction.

It has been reported that the Cβ CP (transmembrane) domain is an important region involved in signal transduction in the β chain of the TCR. Mutational studies in human cell lines have shown that a signalling defect occurs downstream from the zeta associated protein of 70kDa (ZAP-70), and it appears that the calcium ion influx is diminished in that circumstance (Backstrom *et al.,* 1997). It was also reported that the transmembrane domain interacts directly with the zeta chain of the CD3 complex (Backstrom *et al.,* 1996). It is the glutamine (Q) residue of the Cβ CP that is important in signal transduction. In this study the TCRβ sequence alignment (Appendix 3C) showed that the glutamine residue was not conserved in the marsupials and other mammals where the sequence had been characterized. *M. eugenii* and *T. vulpecula* had a histidine (H) instead of a glutamine, while in *O. fraenata* the Cβ CP region appeared to be missing. No investigation was carried out into possible polymorphisms or splice variants in *O. fraenata.* Future investigations of this may explain why this region appeared to be absent in the *O. fraenata* TCRβ clones.

The amino acid glutamine is often found at binding and active sites of proteins (Betts and Russell, 2003). In the present study it was found that the sequence alignment (Appendix 3C) in a number of marsupial and other mammals had the glutamine substituted with histidine, arginine (R), lysine (K) or serine (S). Histidine, like glutamine, is a polar amino acid and is also required at protein binding sites. The substitution of glutamine with arginine is found in birds and some fishes in the Cβ CP domain but this does not appear to affect the functionality of the TCR molecule in those species. However, arginine will also interact at active sites of proteins, although it is more vigorous in its binding capacity than

histidine and glutamine and therefore builds very strong bonds (Waksman *et al.,* 1992). It was therefore concluded that the active binding site of the marsupial TCRβ chain protein in this region identified in this study is conserved, even though several substitutions may occur at the active site of that protein.

In humans, the FG-loop is a hydrophobic region composed of the residues tryptophan (W^{225}), leucine (L^{219}) and proline (P^{232}). These form the base of the FG loop which connects with L^{119} and valine (V^{122}) at the start of the constant domain of the beta chain (Hennecke *et al.,* 2000) and regulates the αβ-T cell development (Touma *et al.,* 2006). This study found that, in marsupials, these residues are conserved but L^{219} is substituted for isoleucine. The structure of the FG loop produces a cavity large enough to accommodate a single Ig- domain (Wang *et al.,* 1998a). In humans and mice, the CD3ε subunit has an acidic residue to basic residue ratio of 2:1 (Gold *et al.,* 1987). This ratio was also found in the *O. fraenata* and *M. eugenii* CD3ε sequence in this study. CD3ε has a non-glycosylated single Ig-like extracellular domain with a length of 70 amino acid residues in *O. fraenata* and *M. eugenii.* This contrasts with the single Ig-like domain in humans which has 87 amino acid residues and has the correct complementary charge to the cavity which is created by the FG-loop (Touma *et al.,* 2006, Ghendler *et al.,* 1998). It appeared that the same structural components and biochemical properties were conserved in the sequences of *O. fraenata* and *M. eugenii* in this study.

The cysteine in the TCRβ sequence of humans, located at position 247 (Appendix 3C), forms an interchain disulfide bridge with the α chain. The F strand cysteine (C^{212}) (Appendix 3C) forms the intra-domain disulfide bridge with the β strand cysteine (C^{147}) of the constant domain (Sasada *et al.,* 2002). These disulphide bridges were found to be conserved in *O. fraenata* and *M. eugenii,* and a similar interaction between the TCRα and TCRβ chains was expected.

Of lesser importance is a region known as the elbow region in the TCRβ constant domain. In humans, this region was reported to be dispensable in the development of TCRαβ

bearing T lymphocytes, but was required for the NK T cell sub-set (Degermann *et al.,* 1999a). Neither the *O. fraenata* nor the *M. eugenii* TCRβ chains were investigated to this extent. Consequently it is not known whether or not this motif has the same function in those species. In humans, the transmembrane domain in the beta chain contains a number of polar residues which are highly conserved (Kunjibettu *et al.,* 2001). This was also found to be the case in *O. fraenata* and *M. eugenii* in this study (Appendix 3C). A lysine residue (K^{271}) in the transmembrane domain of the human TCRβ chain is a key amino acid responsible for the assembly with the CD3 complex and the expression of the receptor complex on the cell surface (Alcover *et al.,* 1990). Substitutions of this amino acid residue have shown that while it does not affect the formation of the T cell receptor dimer it does affect the cell surface expression of the molecule (Alcover *et al.,* 1990). In this study, an alignment of all known TCRβ chain sequences found that the residues of importance were contained within the *O. fraenata* and *M. eugenii* sequences. It was also shown that the important amino acid residues for receptor assembly and dimerization were conserved thus implying that the functional aspect of the molecule is conserved in those species.

3.6 Conclusion

In this study it was found that despite the evolutionary distance of marsupials from humans the relatively high conservation found in the CD3ε chain and the TCRβ molecule in *M. eugenii* and *O. fraenata* indicate that they are not so different in their immunological makeup. It was evident from the results that the CD3ε and TCRβ chains are highly conserved between the marsupial clade and other mammals. The CD3ε chain was determined for the first time in *L. hirsutus* and *M. domestica*. The functional motifs and domains were conserved and by investigating their biochemical makeup of the molecules it was shown that the interactions between the TCRβ chain and the CD3ε molecule were similar to those of other mammals. These sequences were characterized for the first time in *O. fraenata* and the sequences for TCRα and β chains were determined for the first time in *M. eugenii* and *O. fraenata*. The results of the study indicated that the TCRα chain

in those species may function differently to that of other mammals. The additional phenylalanine residue in the TCRα chain found in those species was only four residues removed from another phenylalanine resulting in a different structure which may indicate steric hindrances in this molecule in marsupials.

Mutational studies in humans carried out by Ali *et al.,* (2006) indicated that the spatial arrangement within this motif, and the presence of certain charged amino acid residues surrounding the motif, are crucial for its function. The area of this motif in the *O. fraenata* sequence differed from that in humans and further studies are needed to understand its function in that marsupial species. This was outside the scope of the present project.

Chapter 4

The diprotodontic co-receptors and co-stimulators to the T cell receptor

Chapter 4 – The diprotodontic co-receptors and co-stimulators to the T cell receptor

4.0 Abstract

There are various TCR co-receptor and TCR co-stimulatory molecules that are important for T cell signal transduction. In this study five molecules (CD4, CD8, CD28, CTLA-4 and CD86) that are central to the T cell signal transduction pathway in humans and mice were characterized in several macropods and a didelphid. The partial sequence of CD4 was characterized in *O. fraenata* using an RT-PCR approach. The same technique was used to characterize the open reading frames of the CD8α and β chains in *O. fraenata* and *L. hirsutus*. A sequence fragment of the MHC co-receptor CD86 was identified in the didelphid *M. domestica*. A RACE-PCR approach was used to characterize the TCR co-stimulator CD28 in *M. eugenii* and *M. domestica*. The CD28 antagonist CTLA-4, which is responsible for the regulation of T cell responsiveness, was also characterized with the RT-PCR technique. Comparative analyses were carried out for all of the mRNA transcripts of structurally important motifs. Comparative structure predictions were carried out for CD28, CD8α and β, and CTLA-4 using the Modweb platform. Key findings of this study were the absence of the D4 domain in the *O. fraenata* CD4 sequence. A comparison of the marsupial CD8 sequences with other mammalian CD8 sequences revealed differences in the number of post-translational modifications highlighting possible differences in function. The marsupial sequences of CD28 and CTLA-4 exhibited high conservation when compared with other mammalian and non-mammalian sequences suggesting similarity in function.

4.1 Introduction

Peripheral T cell populations and secondary lymphoid organs are distinguished by the mutually exclusive expression of either the CD4 or CD8 T cell receptors. This is maintained throughout the post-thymic life of T lymphocytes where two distinct T cell lineages are distinguished (Bosselut, 2004). CD4 T lymphocytes are MHC II restricted and define the T helper cell sub-populations, whereas CD8 T lymphocytes are MHC class I restricted and

confer cytotoxicity to target cells such as cells infected by mycobacteria (Wang and Bosselut, 2009). The selection of either a CD4 or a CD8 lineage occurs after the rearrangement of the TCRα and β genes in thymocytes, and then only in cells whose TCRs recognize MHC ligands (Starr *et al.,* 2003). Activation of T lymphocytes occurs through the engagement of the T cell receptor, the CD28 located on the T cell, and the pMHC and B7 (CD80/CD86) family members located on the antigen presenting cell. However if CD28 is not expressed on the cell surface the cell adopts a state of anergy. Two mechanisms are considered to be important for anergy; one with a direct inhibitory effect on CD28 signalling, and the other with an indirect effect on cell-cycle progression via stimulation of growths factors such as IL-2 (Becker *et al.,* 1995, Beverly *et al.,* 1992, Jenkins, 1992). T cell receptor-antigen coupled signals are amplified by transduction through CD28 and are opposed by the CD28 antagonist CTLA-4 (Lenschow *et al.,* 1996). T lymphocytes must therefore progress through the cell cycle in order to escape the fate of anergy (Wells *et al.,* 2001). This is called the 'two-signal' model where both an activation signal and a recognition signal are necessary to avoid an anergic state (Baxter and Hodgkin, 2002, Chambers, 2001, Lafferty *et al.,* 1978).

The four important molecules CD4, CD8, CD28 and CTLA-4 are at the centre of this 'two-signal' model for T cell activation. The literature reports differences in the T cell responses of marsupials when compared to humans and other mammals (Meyer-Lucht *et al.,* 2008, Stone *et al.,* 1996). Marsupial T lymphocytes appear to not react with co-cultured T lymphocytes of other species, thereby demonstrating no mixed lymphocyte responses (MLR) (Stone *et al.,* 1998, Stone *et al.,* 1997a, Stone *et al.,* 1997b). Although no culturing of marsupial cells was undertaken in the present study, the involvement of these four molecules in the MLR was addressed by investigating differences in sequence and therefore in structure and possibly function. The question then arose as to whether or not the co-receptors and modulators are expressed in *O. fraenata, L. hirsutus, M. eugenii* and *M. domestica*. This question was partially answered when the CD4 and CD8 receptors were characterized in *M. eugenii* (Duncan *et al.,* 2007, Duncan *et al.,* 2008). While

infections with mycobacteria lead to the activation of both $CD4^+$ and $CD8^+$ T cell responses, it is the $CD8^+$ T lymphocytes that play a major role in the control of such infections (Smith and Dockrell, 2000). Therefore differences in structural motifs that are implicated in the control of mycobacterial infections may compromise the immune response to such pathogens. While the *M. eugenii* sequences of CD4 and CD8 are available in Genbank (Accession Numbers EF490599 and EU152103 respectively) a characterization of these two molecules has not been reported in *O. fraenata*. Anecdotal reports by animal handlers suggested that *O. fraenata* has a greater degree of resistance to pathogens than other marsupial species. In order to determine the validity of those reports it was first necessary to characterize sequence motifs in both the CD4 and CD8 molecules which are responsible for antagonistic responses to pathogens. This was undertaken in this study.

CD4 and CD8 operate in conjunction with the T cell signalling modulatory molecules CD28 and the cytotoxic T lymphocyte antigen-4 (CTLA-4). They are both members of the immunoglobulin superfamily with an extracellular IgV-like domain (Linsley *et al.,* 1995). A hexapeptide (MYPPPY) is found in the complementarity determining region 3 in the IgV fold of CD28 and CTLA-4. The localization of this motif in the solvent-exposed region and its conservation in a number of taxa suggests the possible presence of a ligand-binding epitope (Srinivasan *et al.,* 2001).

As previously described, CD28 and CTLA-4 are important in the 'two-signal' model since they either accelerate or block the signal thus making them modulators of the T cell activation mechanism. The 'two-signal' theory is important in transplant science where, in marsupials, it is evident that there is a different response to skin grafts when compared to humans and other mammals (Stone *et al.,* 1997a, Stone *et al.,* 1997b). In recent times both CD28, CTLA-4, and their respective ligands CD80 and CD86, have been implicated in graft rejection, and have also been investigated for therapeutic applications to prevent graft rejection (Goldstein, 2011, Habicht *et al.,* 2007). Differences in their structure may explain the different responses to skin grafts in marsupials. CD86 is part of the B7-CD28/CTLA-4 pathway and provides signals for the activation and survival of T

lymphocytes by acting as a ligand for CD28 and CTLA-4. The effects of CD86 are mainly seen in the early events of T cell activation of naïve T lymphocytes when the pathway to anergy or stimulation is determined. The MYPPPY motif is the binding site that ties CD28, CTLA-4 and CD86 together.

Characterization of the important sequencing motifs within the CD4 and CD8 molecules in marsupials is necessary to further our understanding of how some species resist bacterial infections while others succumb to them. Prior to this study CD28 and CTLA-4, the two molecules responsible for the T cell activation signal, had not been characterized in any marsupial species.

4.2 Aims and Objectives

The aims of the experiments reported in this Chapter were.

- To characterize important sequencing motifs of the co-receptor CD4 in *O. fraenata* for comparison with the same structures in the published *M. eugenii* sequence.
- To characterize the co-receptors CD8α and CD8β in *O. fraenata* and *L. hirsutus* for comparison with the same structures in the published *M. eugenii* sequence.
- To characterize the T cell modulators CD28 in *O. fraenata* and CTLA-4 in O. fraenata and *M. eugenii* for comparison with the expressed and annotated sequences in the ensembl database (http.//www.ensembl.org) for *M. eugenii*.
- To characterize, at least in part, the molecule CD86 located on the antigen-presenting cell, and to establish the binding site to the immune modulators CD28 and CTLA-4 in *M. domestica*.
- To investigate the putative amino acid sequences of the *O. fraenata* CD4, CD8, CD28 and CTLA-4 genes for any apparent differences in their structural and functional motifs when compared to *M.eugenii* and *M. domestica*.
- To characterize CD8 in *L. hirsutus* to investigate the documented susceptibility of this species to mycobacterial infections.

- To undertake phylogenetic analyses of the functional motifs of the open reading frames of CD8, CD28 and CTLA-4 in order to show evolutionary changes which have occurred in the genes or in the structurally important motifs.
- To conduct protein modelling from the putative amino acid sequences to deduce possible amino acid residue interactions and to determine a probable tertiary structure for the molecules in marsupials.

4.3 Specific Materials and Methods

The following materials and methods were unique to this chapter. General materials and methods are detailed in Chapter 2.

4.3.1 RNA, mRNA and cDNA

RNA was isolated from spleen, liver, gut node and thymus of *O. fraenata*. mRNA was isolated from total RNA, and cDNA was synthesized as described in Chapter 2. The CD4 sequence in *O. fraenata* was characterized from a gut node, while the molecules CD8 and CTLA-4 were characterized from a spleen of this species. The sequences obtained for *M. eugenii* originated from stimulated lymphocytes donated by Dr. L. J. Young. All sequences were cloned, and three clones were sequenced in the forward and reverse direction as outlined in Chapter 2, section 2.1.10.

4.3.2 Primer design

4.3.2.1 CD4, CD8α, CD8β, CD28, CTLA-4 and CD86

Primers were designed as detailed in section 2.1.2. The primer sequences used are presented in Table 4.1.

Table 4.1. Primer sequences used to characterize the partial sequence of CD4 in *O. fraenata*, the open reading frames of CD8α, CD8β in *O. fraenata* and *L. hirsutus*, the open reading frame of CD28 in *M. eugenii,* a partial sequence in *M. domestica*, the open reading frames of CTLA-4 in *O. fraenata* and *M. eugenii*, and the partial sequence of CD86 in *M. domestica*.

Gene/ Species	Forward primer 3' – 5'	Reverse primer 3' – 5'	T_m	Size	Primer ID
CD4 *O. fraenata*	gactcggggatgtacttctgtgaggtggaaga	gtgagaaaaagacctgccagtgt	68°C/56°C	1009bp	*CD4F/CD4ABR*
	gtgttcaaggtgacagcca	ggggacactggcaggtc	55.4°C/54 °C	1094bp	*CD4AF.2/CD4DR*
CD8α *O. fraenata* *L. hirsutus*	atgggctccctcttggct gtacgatccc (Start)	ttaagcatatctctctgatgggcc agcc (Stop)	66°C/61°C	729bp	*BCD8aF/BCD8aR*
CD8β *O. fraenata* *L. hirsutus*	atggctcagcctctgcccattcag (Start)	ctatttcacgacgtggrgccgagc tac (Stop)	61°C/63°C	613bp	*BCD8bF/BCD8bR*
CD28 *M. eugenii*	gattttggtcaaacagccacatttgc (Exon-2)	ggagtcatgttcatgtagtcacta tg (Exon-4)	56°C/56°C	463bp	*CD28F/CD28R*
	cgactggagcacgaggacactga (3'RACE primer)	caaacggatatttagttctgtaaa attg (Exon-2)	74°C/53°C	~500bp	*TR28F/3'RACE*
	caattttacagaagtaaatatccgtttg (Exon-2)	gctgtcaacgatacgctacgtaa cg (5'RACE primer)	53°C/76°C	~300bp	*TR28F/3'RACE*
CTLA-4 *M. eugenii*	rcattctccttcctgacattccgag (5' UTR)	ctatccctctttgcaccactcc (3'UTR)	57°C/57°C	742bp	*CTLA4-UTF/ CTLA4-UTR*
O. fraenata	attgatgctcactctcacagg	tcaaagtctgggcaaggttc	54.6°C/ 55.2°C	125bp	*CTLAfexp2/* CTLArexp2*
CD86 *M.domestica*	gcattgttcaacgggactgtagacctg	ctggattatccttgttcagaagag cagg	61°C/60°C	410bp	*O86F/O86R*

* Expression primer

Primer combinations CD4AF.2/CD4DR and CD4CF/CD4ABR were used to determine partial sequences of the *O. fraenata* CD4 molecule and to produce 1009bp and 1094bp products. An overlap was created between the two products to permit verification of the CD4 sequence.

4.3.1.2 Polymerase chain reaction (PCR), cloning and sequencing

Generally used RT-PCR templates, concentrations of the PCR mixes, and preparation of the RT-PCR and RACE-PCR products are detailed in Chapter 2 sections 2.1.6, 2.1.8 and 2.1.10. The PCR and RACE-PCR templates used in this chapter are listed in Table 4.2.

RACE primers were designed as outlined in Chapter 2, section 2.1.2 from the initial sequence derived from the RT-PCR products. RACE-PCRs were performed to obtain the 5' and 3' ends of the molecules as described in Chapter 2 section 2.1.9.

Table 4.2. PCR and RACE-PCR templates used for the amplification of CD4, CD8α, CD8β, CD28, CTLA-4 and CD86 in *O. fraenata, M. eugenii, L. hirsutus* and *M. domestica*.

PCR and RACE-PCR templates for CD4, CD8α, CD8β, CD28, CTLA-4 and CD86			
Species	Gene	PCR template number	RACE-PCR template number
O. fraenata	CD4	No. 3	-
O. fraenata *L. hirsutus*	CD8α	No. 3	-
O. fraenata *L. hirsutus*	CD8β	No. 3	-
M. eugenii	CD28	No. 2	Nos. 1 and 2
M. eugenii *O. fraenata*	CTLA-4	No. 3	-
M. domestica	CD86	No. 3	-

4.3.1.3 Phylogeny

A phylogenetic analysis using Mega5 was undertaken for all genes after the full coding domains were determined. The method is detailed in Chapter 2, section 2.3.

4.3.1.4 Bioinformatics

All of the bioinformatics tools detailed in Chapter 2 were used to elucidate putative domain structures, putative glycosylation sites, possible disulphide bonds, putative amino acid sequences, putative secondary and tertiary structures, and homology searches.

4.4 Results

4.4.1 CD4

A partial sequence of the CD4 molecule was characterized in *O. fraenata* and compared to the published sequences of *M. eugenii* and *M. domestica*.

4.4.1.1 CD4 - Homology

The BLAST homology search showed that over the 862bp partial sequence the *O. fraenata* CD4 nucleotide had a 97% identity to the *M. eugenii* nucleotide sequence and a 77% identity to the *M. domestica* nucleotide sequence. The predicted 292 amino acid sequence in *O. fraenata* had an identity of 80% to *M. eugenii* and 53% to *M. domestica*. The homology of the *O. fraenata* CD4 amino acid sequence was low (between 23% and 38%) compared to the CD4 sequences in other vertebrates due to a sequence gap found in domain 4 of the *O. fraenata* CD4 (Table 4.3).

Table 4.3. Homology search results for the nucleotide and amino acid sequences and their respective e-values for the *O. fraenata* partial CD4 molecule. The values are calculated over the 862bp/292aa length of the partial sequence.

Homology search result for CD4 in *O. fraenata*				
Species	Nucleotide	e-value	Amino acid	e-value
M. eugenii	97%	0.0	80%	3e.141
M. domestica	77%	1e.95	53%	4e.66
B. taurus	---	---	37%	3e.33
C. familiaris	---	---	38%	8e.31
F. catus	---	---	38%	1e.30
C. hircus	---	---	36%	3e.30
O. aries	---	---	36%	6e.30
R. norvegicus	---	---	34%	3e-28
S. scrofa	---	---	38%	4e.27
O. cuniculus	---	---	32%	1e.26
I. punctatus	---	---	23%	0.71

For Accession numbers see Appendix 4A. --- = not recognized by the BLAST algorithms.

4.4.1.2 CD4 - Domain structure

The partial CD4 sequence in *O. fraenata* contained two putative Ig domains with e-values of 4.38e+00 and 2.31e+00 respectively. Within the two predicted Ig-domains, putative Igc2 and Igv domains were identified with e-values of 4.15e+00 and 1.12e+03 respectively. These e-values indicated that the occurrence of the domain structures was of high probability.

4.4.1.3 CD4 - Glycosylation and glycation sites

Four predicted O-linked glycosylation sites were found in the partial *O. fraenata* sequence. The putative glycosylated threonines were at positions 4, 6, 165 and 167. Two putative N-linked glycosylation sites NVT and NET were found in the partial *O. fraenata* CD4 sequence, and there were predicted glycated residues at positions 49, 50, 56, 103,143, 150, 161, 200, 251 and 255. However validation of this result requires more sophisticated mass spectrometry or Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOFF) analysis which was outside the scope of this study.

4.4.1.4 CD4 – Phosphorylation sites

Eleven serine, six threonine and two tyrosine phosphorylation sites were predicted in the partial CD4 sequence of *O. fraenata*. The probabilities of the putative phosphorylation sites ranged from 56.5% to 99.6%. The positions of these sites and their probabilities are shown in Table 4.4.

Only a single putative N-linked phosphorylation site was predicted in *O. fraenata*. The putative N-linked phosphorylation site in domain 4 which is observed in the published *M. eugenii* CD4 sequence is absent in *O. fraenata*.

Table 4.4. Predicted serine, threonine and tyrosine phosphorylation sites, their positions and probabilities in the partial *O. fraenata* CD4 sequence.

Serine		Threonine		Tyrosine	
Position	Probabilities	Position	Probabilities	Position	Probabilities
8	58.9%	6	87.1%	10	95.7%
23	84.4%	75	82.5%	95	90.8%
72	92.1%	153	91.9%		
141	98.2%	230	89.5%		
157	99.1%	232	93.1%		
170	62.9%	253	69.3%		
203	98.2%				
206	99.6%				
231	98.8%				
247	56.5%				
249	91.2%				

4.4.1.5 CD4 - Primary sequence and secondary structure prediction

The partial CD4 sequence of *O. fraenata* consisted of 876bp that translated into 292 amino acids. A frame shift was found when translating the nucleotide sequence into the putative amino acid sequence. A premature stop codon, marked *, was found in domain 4 of CD4. This was verified in three clones and consequently a sequencing error was considered to be improbable. The sequence following the stop codon was in a different reading frame. A sequence gap was found in the *O. fraenata* CD4 sequence in domain No. 4 preceding the joining region No. 4 when compared with other mammalian sequences (Appendix 4A). If this is an actual sequence gap then this molecule will not fold in the same manner as other mammalian CD4 molecules. This may be evidence that the *O. fraenata* CD4 molecule is a soluble form of the receptor and not the membrane bound type. Despite a number of attempts, the full coding domain of *O. fraenata* CD4 could not be obtained and further attempts were abandoned. However important sequence motifs relevant to the function of the molecule in the *O. fraenata* partial sequence were elucidated in the present study.

A secondary structure prediction which was carried out for the partial sequence of *O. fraenata* is illustrated in Fig. 4.1. For the purpose of comparison, a secondary structure prediction was also carried out for the *M. eugenii* sequence which had been characterized previously by Duncan *et al.* (2007) (data in Appendix 4A). The variances in the secondary structures were mainly in the length of the beta sheets and may not affect the overall functionality of the molecule, although this will have to be verified once the full coding domain is determined in *O. fraenata*.

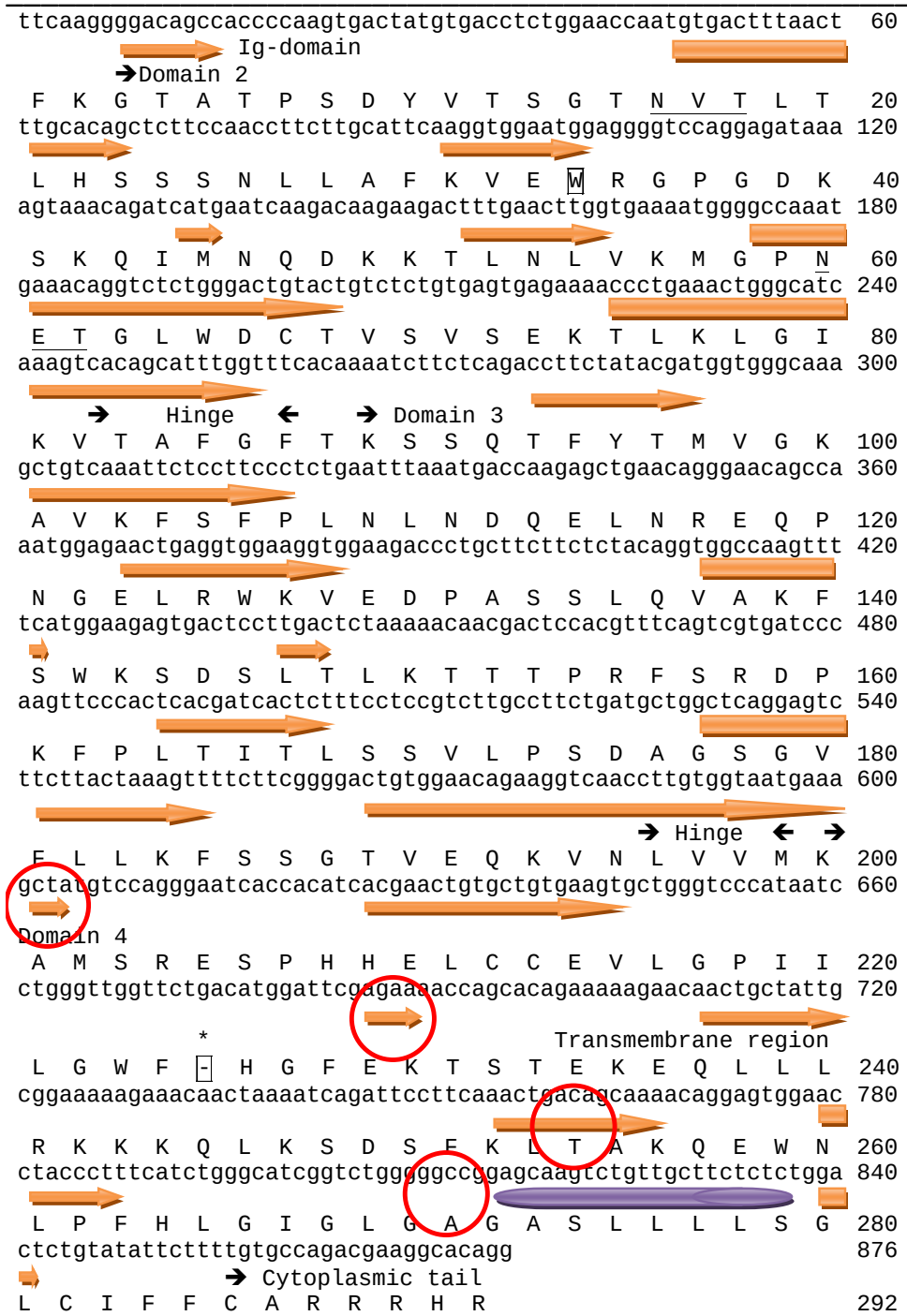


Figure 4.1. *O. fraenata* CD4 partial primary sequence and secondary structure prediction. Underlined = glycosylation sites. ➔ = domain regions. ◯ = differences compared to *M. eugenii* CD4. * = premature stop codon. ⬭ = Transmembrane helix. ➡ = Strand.

4.4.2 CD8 alpha (CD8α)

The open reading frames of the CD8α molecule were characterized in *O. fraenata* and *L. hirsutus* and these were compared to the published sequences of *M. eugenii* and *M. domestica*.

4.4.2.1 CD8α -Homology

The BLAST homology search indicated that the *O. fraenata* CD8α sequence had a 97% identity to the *M. eugenii* sequence at the nucleotide level and a 95% sequence identity at the amino acid level. The BLAST search for the *L. hirsutus* CD8α chain showed a sequence identity of 98% at the nucleotide level and 96% identity to *M. eugenii* at the amino acid level. The *L. hirsutus* CD8α had a sequence identity of 97% at the nucleotide level and 96% at the amino acid level to *O. fraenata*. In contrast, the *O. fraenata* and *L. hirsutus* CD8α putative amino acid sequences had a sequence homology of only 45% and 46% respectively to the human amino acid sequence (Table 4.5).

Low complexity amino acid sequences were found in the BLAST homology search in both *O. fraenata* and *L. hirsutus*. The *O. fraenata* sequence contained 38 such residues while *L. hirsutus* had 36. This significantly lowered the amino acid identity percentages in both species.

Table 4.5. Homology search results for the *O. fraenata* and *L. hirsutus* CD8α nucleotide and amino acid sequence, their identities and respective e-values.

CD8α homology search results for *O. fraenata* and *L. hirsutus*								
	O. fraenata				*L. hirustus*			
Species	Nucleotide		Amino acid		Nucleotide		Amino acid	
	Identities	e-value	Identities	e-value	Identities	e-value	Identities	e-value
M. eugenii (*)	97%	0.0	95%	5e-118	98%	0.0	96%	5e-120
L. hirsutus (*)	97%	0.0	96%	2e-97	100%	0.0	100%	2e-97
T. truncatus	97%	1e-11	48%	2e-38	97%	1e-11	47%	3e-37
F. catus	85%	5e-36	48%	4e-37	85%	1e-37	46%	2e-38
C. l. familiaris	87%	9e-27	50%	2e-39	87%	4e-25	50%	1e-38
M. domestica	76%	1e-150	64%	8e-66	77%	7e-149	64%	1e-66
H. sapiens	94%	5e-11	45%	3e-33	94%	5e-11	46%	9e-34
S. scrofa	93%	1e-18	44%	4e-33	93%	1e-18	48%	2e-39
S. sciureus	66%	1e-18	47%	6e-42	66%	2e-17	47%	2e-39
S. hispidus	66%	2e-15	42%	3e-29	71%	8e-15	50%	7e-40
B. taurus	65%	3e-07	48%	6e-42	64%	4e-05	46%	5e-40

For Accession numbers see Appendix 4B. (*) = BLAST2 algorithm was used.

4.4.2.2 CD8α - Domain structure

A leader sequence was predicted in both *O. fraenata* and *L. hirsutus* CD8α chains. For both species the leader sequence had a predicted cleavage site between amino acid residues 25 and 26. In both sequences, a predicted Igv domain at positions 44 to 124 and a predicted transmembrane domain at positions 190 to 212 were detected.

4.4.2.3 CD8α - Glycosylation and glycation sites

Neither the *O. fraenata* nor the *L. hirsutus* CD8α chain sequence contained putative N-linked glycosylation sites; however eight putative O-linked glycosylation sites were predicted within the CD8α hinge region of *O. fraenata* and seven in *L. hirsutus.* The O-linked glycosylation sites had probabilities ranging from 51% to 72.4%. Six of these glycosylation sites were found to be in the same position in both species.

Six putative glycation sites were predicted in the CD8α chains for *O. fraenata* while five sites were predicted for *L. hirsutus*. In both marsupials, the probabilities ranged from 49% to 85.45%. The probability of the glycated lysine residue at position 36 in *O. fraenata* was 49% which was below the threshold of 50% thus making the probability unreliable (Table 4.6). However validation of this result requires more sophisticated mass spectrometry or

Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOFF) analysis which was outside the scope of this study.

Table 4.6. Predicted O-linked glycosylation and glycation sites and their probabilities in in the CD8α molecule in *O. fraenata* and *L. hirsutus.*

O-linked glycosylation and glycation sites for CD8α in *O. fraenata* and *L. hirsutus*				
Species	O-linked glycosylation sites	Probability	Glycation sites	Probability
O. fraenata	144	62.0%	36	49%
	145	63.1%	68	66.5%
	146	66.3%	151	60.1%
	147	68.2%	177	52.7%
	154	72.4%	181	85.3%
	155	71.7%	233	73.2%
	160	67.6%		
	161	63.3%		
L. hirsutus	145	61.6%	68	66.8%
	146	63.3%	153	82.5%
	147	65.2%	182	52.7%
	154	68.1%	186	85.4%
	155	67.4%	238	54.2%
	160	64.1%		
	163	51.0%		

4.4.2.4 CD8α – Phosphorylation sites

A number of putative phosphorylation sites were predicted in both *O. fraenata* and *L. hirsutus* CD8α sequences. The *O. fraenata* sequence had the least number of putative phosphorylation sites when compared with *L. hirsutus*. Six serine, five threonine and two tyrosine phosphorylation sites were predicted in the *O. fraenata* sequence. In the *L. hirsutus* sequence, seven serine, four threonine and four tyrosine phosphorylation sites were predicted. The variance in the number of putative phosphorylation sites in the *O. fraenata* and *L. hirsutus* CD8α chains indicated a possible difference in post-translational modification in those species (Table 4.7).

The *O. fraenata* sequence, unlike the *L. hirsutus* sequence, was devoid of tyrosine residues. This suggested that the *O. fraenata* molecule does not have the capability to

phosphorylate tyrosines. This would indicate a different post-translational modification for *O. fraenata* and could influence the functionality of the molecule.

Table 4.7. Predicted amino acid phosphorylation sites in the CD8α chain, their locations and their confidence levels in *O. fraenata* and *L. hirsutus*.

Predicted phosphorylation sites in CD8α for *O. fraenata* and *M. eugenii*						
Species	Serine	Confidence level	Threonine	Confidence level	Tyrosine	Confidence level
O. fraenata	42	99.7%	104	75.8%	190	98.5%
	55	82.1%	145	58.1%	240	84.5%
	101	98.0%	147	93.2%		
	103	90.6%	155	64.2%		
	216	98.7%	161	87.0%		
	237	83.5%				
L. hirsutus	42	99.7%	104	75.8%	119	80.8%
	55	90.5%	147	94.7%	120	85.4%
	82	54.9%	155	64.2%	190	98.5%
	101	98.0%	161	87.0%	240	84.5%
	103	90.6%				
	216	98.7%				
	237	83.5%				

4.4.2.5 CD8α - Disulphide bond prediction

Four disulphide bonds were predicted in the *O. fraenata* and *L. hirsutus* CD8α chains. For both species, the predicted disulphide bonds were in the same positions. Comparison of those results with the published sequences of *M. eugenii* and *M. domestica* indicated that these bonds are highly conserved among marsupials. Conservation of those bonds is also evident in all other mammals (alignment in Appendix 4A).

4.4.2.6 CD8α - Primary sequence and secondary structure prediction

It was found that the *O. fraenata* CD8α chain consisted of 726bp which translated into 241 amino acids. The *L. hirsutus* CD8α chain had the number of base pairs as the *O. fraenata* CD8α sequence (Figs. 4.1 and 4.2).

The sequence had a common pattern of one leucine for every seven amino acid residues. Therefore it was assumed that the putative leader sequence contained a possible leucine zipper. The amino acid residues in the predicted leader sequence were mainly

hydrophobic, which is a prerequisite for a leucine zipper. The leucine zipper is necessary for the dimerization of the molecule. There were eleven beta sheets and three alpha helices in both the *O. fraenata* and the *L. hirsutus* sequences. The lengths of the beta sheets and alpha helices varied between the two species, however there appeared to be no change in structure. The $p56^{lck}$ binding site was identified in both sequences and is fully conserved when compared to the sequence of humans.

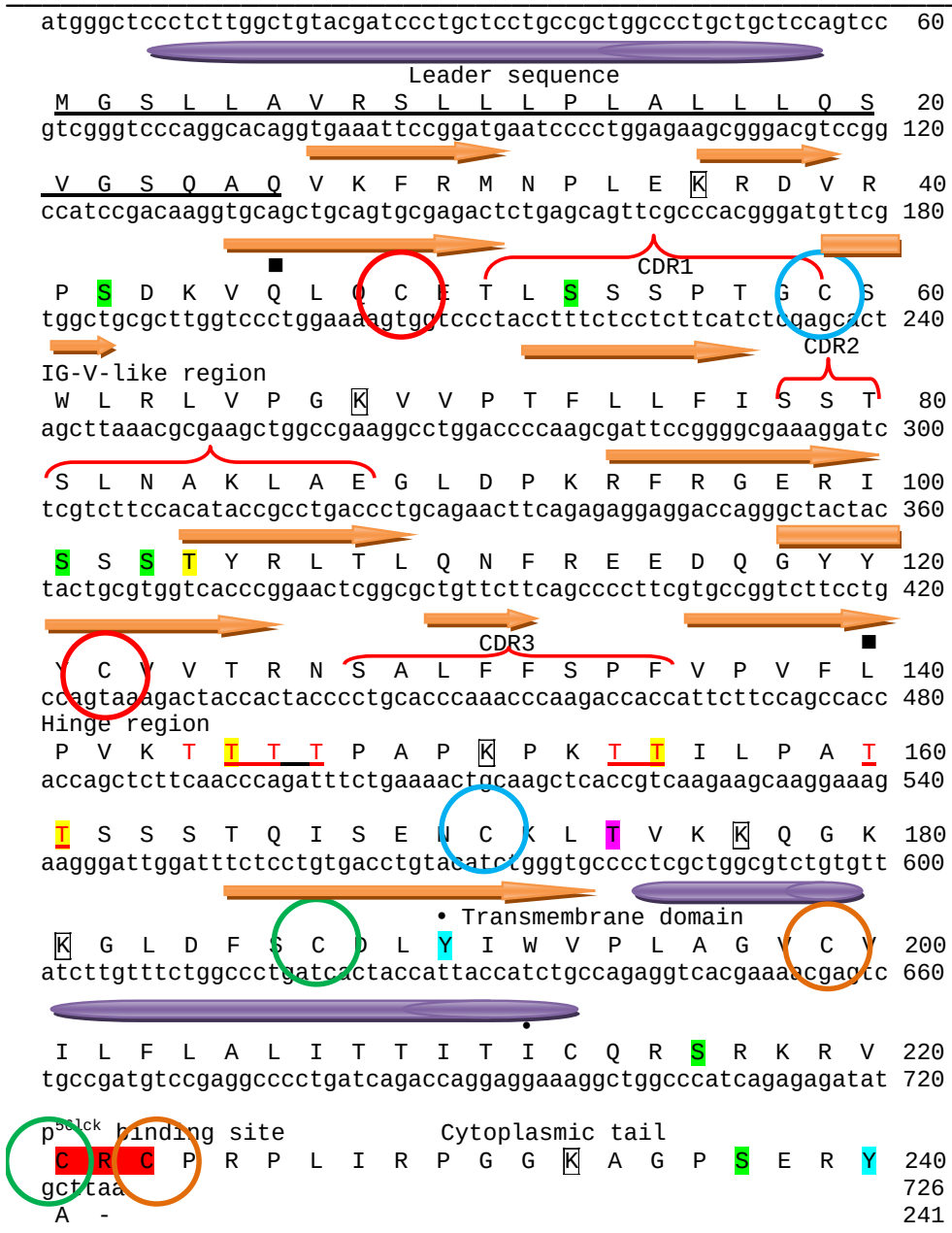


Figure 4.2. *O. fraenata* CD8α primary sequence and secondary structure prediction.
Underlined in bold = Signal peptide. K = putative glycation sites. S = Serine phosphorylation sites. T = threonine phosphorylation site. Y = tyrosine phosphorylation sites. T = PKC binding site. • = Transmembrane helix. ■ = IgSF-domain. CRC = $p56^{lck}$ binding site. T = marks the putative O-linked glycosylation sites. Coloured circles = disulphide bond formations.
= Transmembrane helix. = Strand.

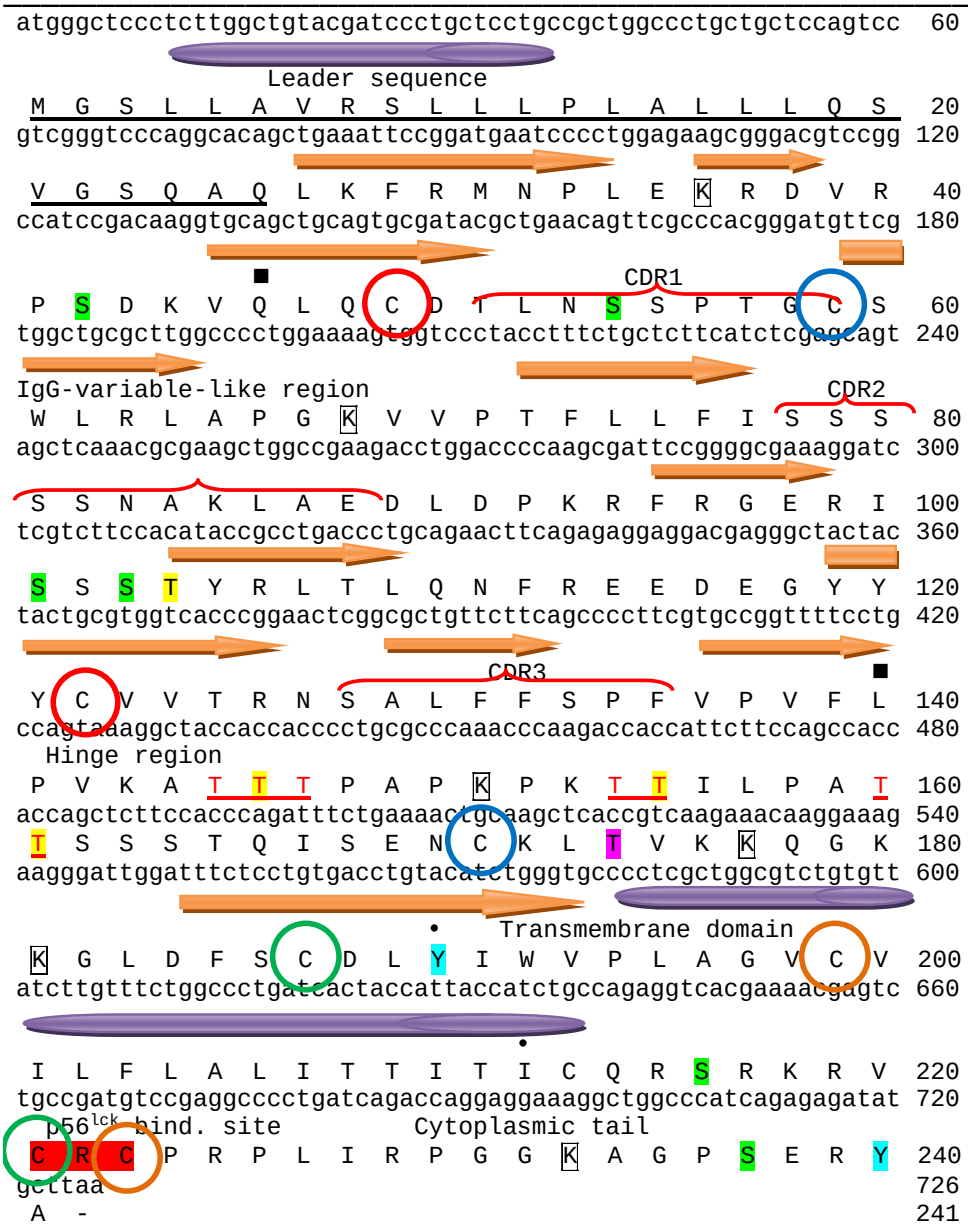


Figure 4.3. *L. hirsutus* CD8α primary sequence and secondary structure prediction.
Underlined in bold = Signal peptide. K = putative glycation sites. S = Serine phosphorylation sites. T = threonine phosphorylation site. Y = tyrosine phosphorylation sites. T = PKC binding site. • = Transmembrane helix. ■ = IgSF-domain. CRC = $p56^{lck}$ binding site. T = marks the putative O-linked glycosylation sites. Coloured circles = disulphide bond formation.
= Transmembrane helix. = Strand.

4.4.2.7 CD8α - Tertiary structure and ligand binding predictions

A tertiary structure prediction was carried out for the CD8α sequences of *O. fraenata, L. hirsutus*, *M. eugenii* and *M. domestica* and the results are shown in Fig.4.4. There were differences in the loops and β-strand positions of the CD8α chains but overall it appeared that the structure was conserved between the species (Fig. 4.4a). However the putative ligand binding sites of the CD8α chain differed between the four marsupial species (Fig. 4.4b). Whether or not the function is conserved despite the differences in the accessibility of the ligands is yet to be determined.

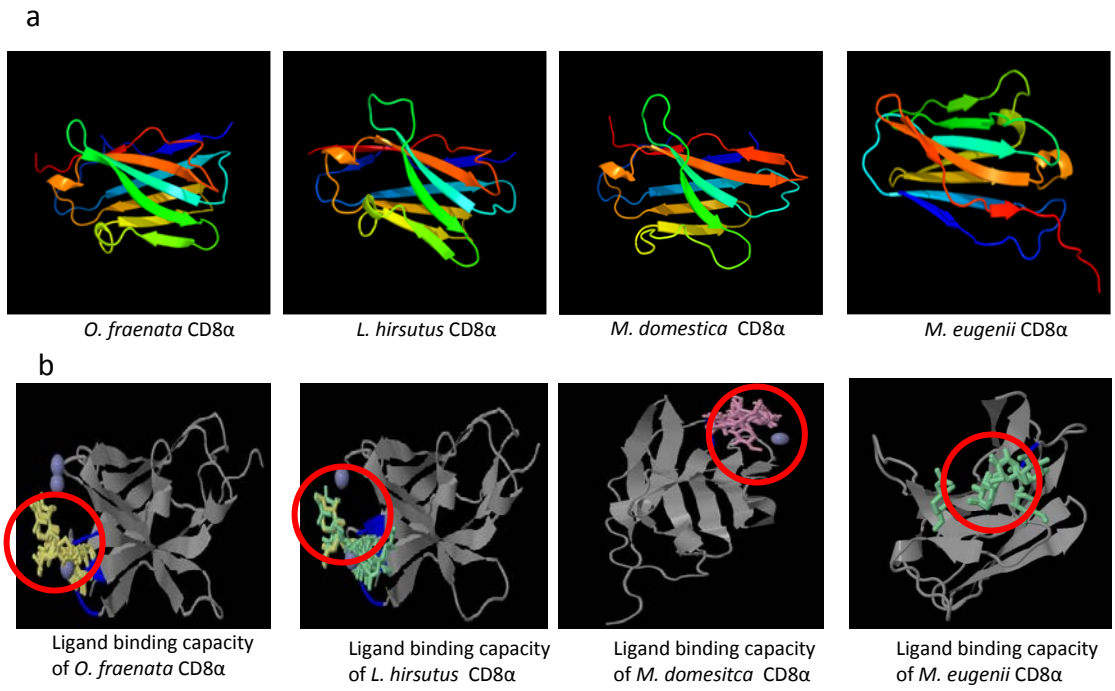


Figure 4.4. (a) Tertiary structure predictions for CD8α using the Phyre2 protein structure prediction server for *O. fraenata, L. hirsutus*, *M. domestica* and *M. eugenii.* (b) Ligand binding capacity of the CD8α chain using the 3D Ligand server for *O. fraenata*, *L. hirsutus*, *M. domestica* and *M. eugenii.*

The number of ligands bound by the CD8α chain and the amino residues making contact with the ligands in the four species is shown in Table 4.8. It is clear from the table that *M. domestica* and *M. eugenii* have different abilities to bind ligands. The number of residues putatively binding a ligand is reduced in *M. domestica*.

Table 4.8. Predicted ligand binding sites showing residue positions and residues in the CD8α of *O. fraenata*, *L. hirsutus*, *M. domestica* and *M. eugenii* using the 3D Ligand server.

Ligand binding capacity of CD8α for four marsupials			
Species	Residue position	Residue	Predicted binding site (contacts)
O. fraenata	57	Threonine	11
	58	Glycine	6
	125	Threonine	8
	126	Arginine	13
	127	Asparagine	6
L. hirsutus	57	Threonine	11
	58	Glycine	6
	125	Threonine	8
	126	Arginine	13
	127	Asparagine	6
M. domestica	57	Threonine	3
M. eugenii	56	Serine	4
	57	Threonine	5

The Swiss-Model prediction server from EXPASY was used to compare the *O. fraenata* and *L. hirsutus* CD8α putative amino acid sequences with models within the pdb database. This comparison indicated that both CD8α molecules in both species compared well with other CD8α structures (Fig. 4.4 a, c). The model of best fit was the pdb structure 2q3a which was annotated as a dimer. The CD8α structures of both of *O. fraenata* and *L. hirsutus* could only be predicted as a single chain by the software.

Both the Swiss-Model server and the Modweb server were used to compare the structure 2q3a, which in the pdb database was shown to be the *M. mulatta* CD8α chain, with the *O. fraenata* and *L. hirsutus* CD8α chains. The Z score slider indicated that for both *O. fraenata* and *L. hirsutus* the structure predictions were moderately accurate. Both CD8α chains were found to have good Z-scores in the non-redundant comparison (Fig. 4.5 b, d). The amino acid residues at positions 28 to 145 were modelled with high probability while the amino acid residues at positions 1 to 27 and at positions 146 to 241 could not be modelled with any degree of accuracy by the software.

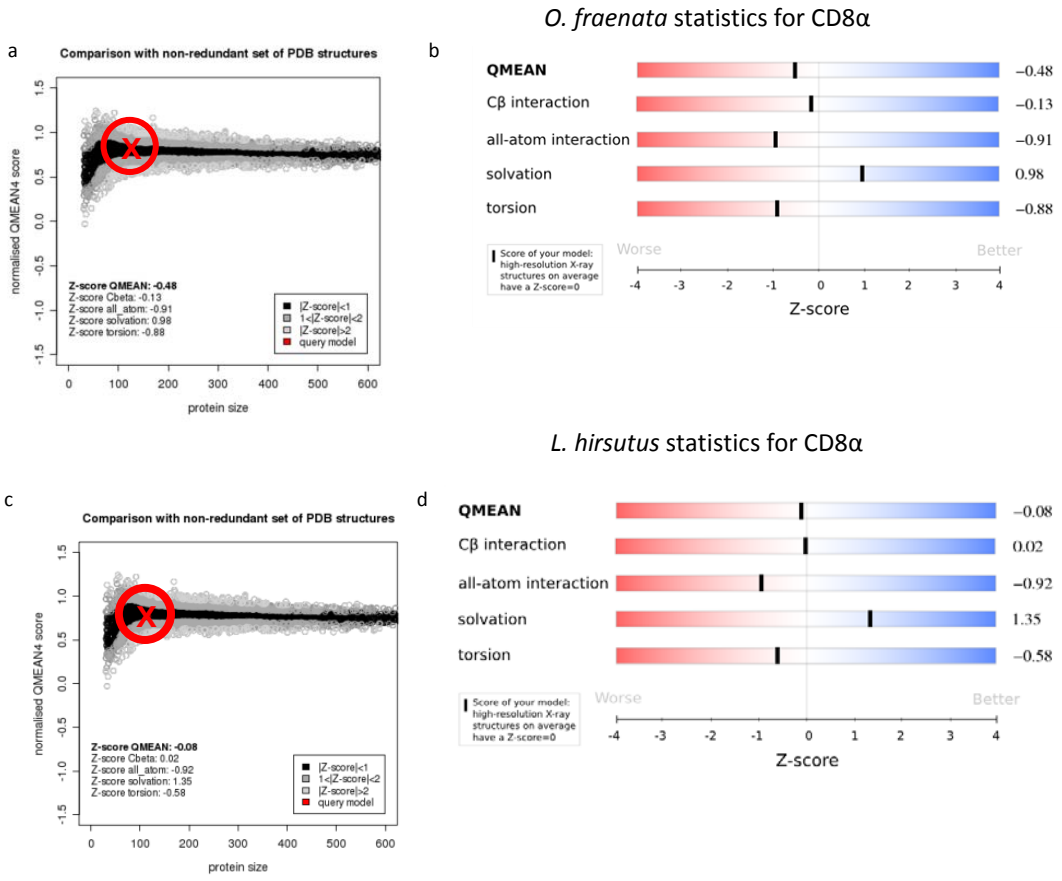


Figure 4.5. Comparison of *O. fraenata* and *L. hirsutus* CD8α structures with the pdb structure. (a) (X) = *O. fraenata* CD8α sequence compared to homologous structures in the pdb database. (b) Z- score slider indicating the quality of the *O. fraenata* CD8α model showing a good QMEAN. (c) (X) = *L. hirsutus* CD8α sequence compared to homologous structures in the pdb database. (d) Z-score slider indicating the quality of the *L. hirsutus* CD8α model showing a good QMEAN.

4.4.2.8 CD8α - Phylogenetic analysis

The evolutionary history was inferred with a Neighbor-Joining phylogenetic tree. The optimal tree had a sum of branch lengths of 6.28. The branch lengths were similar for the three macropods *O. fraenata, M. eugenii* and *L. hirsutus*, however the didelphid *M. domestica* had the longest branch length which indicated that, in an evolutionary sense, it is further removed from the other marsupials. This would be expected since this animal is

not a close relative of the macropods and a greater sequence variation should occur. The percentage of replicate trees in the bootstrap test (1,000 replicates) in which the associated taxa are clustered together are shown next to the branches (Fig. 4.6). The marsupial species were clustered together but the distance between them and the primates, including *H. sapiens,* was significant. This was verified in the alignment shown in Appendix 4B. The phylogenetic tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the tree. This analysis involved 28 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 138 positions in the final dataset (Fig. 4.6).

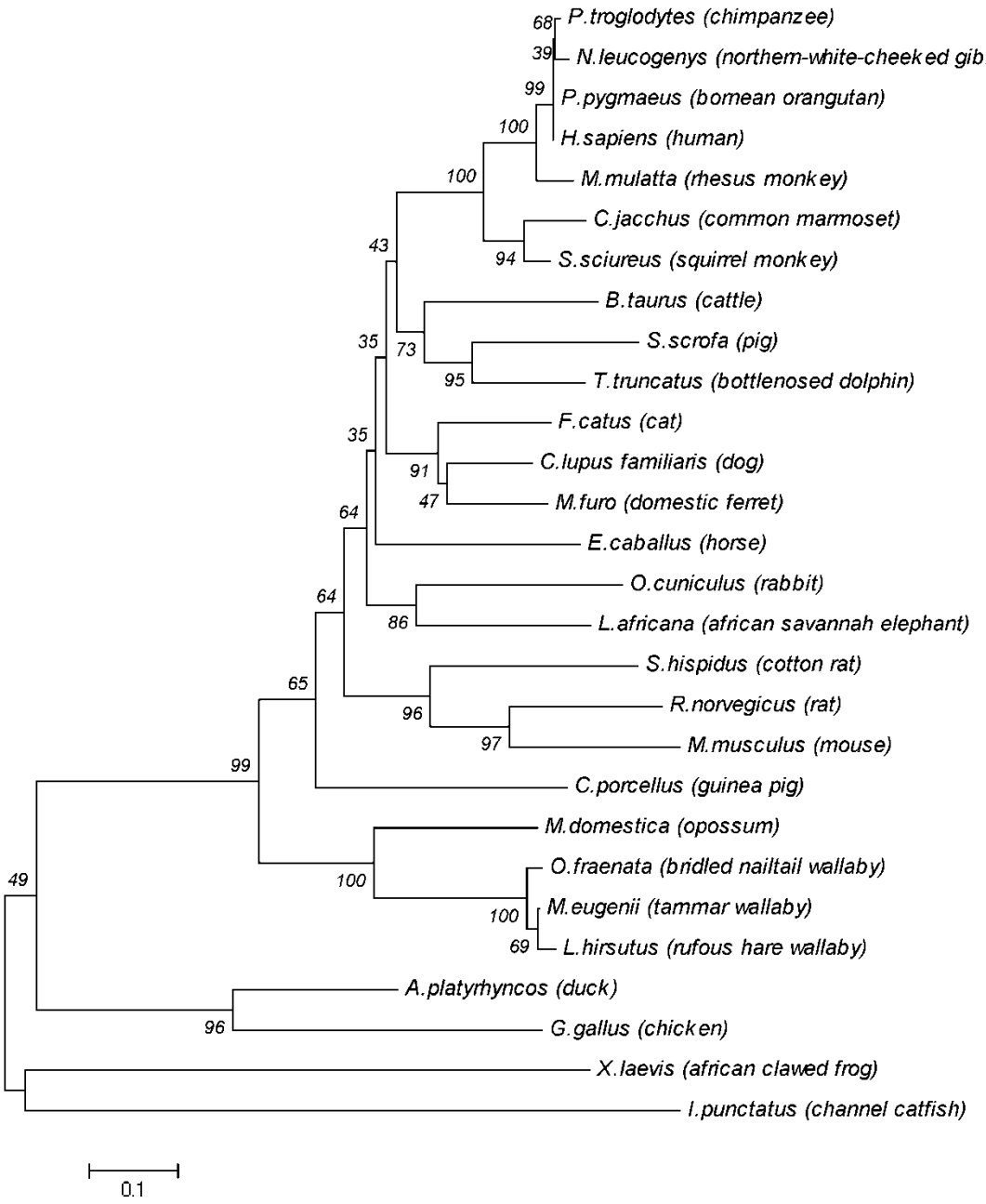


Figure 4.6. Phylogenetic tree for CD8α (including the marsupials *M. eugenii, O. fraenata*, *L. hirsutus* and *M. domestica*). Branch lengths give an indication of the evolutionary relationship between the taxa.

4.4.3 CD8 beta (CD8β)

The CD8β chain was characterized in O*. fraenata* and *L. hirsutus* and compared with the published sequences of *M. eugenii* and *M. domestica*.

4.4.3.1 CD8β - Homology

The homology search conducted with the BLAST algorithm revealed a close relationship between the macropods (≥94% identity at both the nucleotide and amino acid levels), while the conservation to other mammals was much lower (≤71% identity at the nucleotide and amino acid levels). The sequences of *O. fraenata* and *L. hirsutus* both had a 69% identity *to M. domestica* at the nucleotide level, while at the amino acid level the identity percentages were 44% and 57% respectively. A low complexity sequence of 32 amino acid residues was found in both *O. fraenata* and *L. hirsutus* CD8β chains thus lowering the identity scores at the amino acid level (Table 4.9).

Table 4.9. Homology search results for the *O. fraenata* and *L. hirsutus* CD8β nucleotide and amino acid sequences, their identity and respective e-values. The values are calculated over the 786 786 bp/192aa length of the partial sequence.

CD8β homology search results for *O. fraenata* and *L. hirsutus*								
	O. fraenata				*L. hirsutus*			
Species	Nucleotide		Amino acid		Nucleotide		Amino acid	
	Identities	e-value	Identities	e-value	Identities	e-value	Identities	e-value
O. fraenata	100%	0.0	100%	0.0	99%	0.0	96%	3e-116
L. hirsutus	99%	0.0	96%	3e-116	100%	0.0	100%	0.0
M. eugenii	96%	0.0	95%	2e-105	96%	0.0	94%	1e-107
F. catus	71%	2e-08	44%	3e-24	71%	9e-07	38%	2e-25
S. scrofa	71%	1e-10	37%	5e-18	-----	-----	35%	1e-17
T. truncates	70%	1e-11	40%	4e-23	70%	1e-11	39%	7e-24
M. domestica	69%	1e-68	44%	5e-26	69%	5e-67	57%	7e-55
H. sapiens	69%	3e-06	44%	9e-24	68%	1e-04	43%	1e-24
R. norvegicus	69%	9e-07	36%	3e-21	69%	4e-05	35%	8e-22
M. musculus	-----	-----	38%	8e-24	-----	-----	38%	1e-24
C. porcellus	-----	-----	37%	2e-25	-----	-----	36%	2e-26
G. gallus	-----	-----	32%	1e-08	-----	-----	32%	9e-08
A. mexicanum	-----	-----	-----	-----	-----	-----	30%	1e-08
X. laevis	-----	-----	-----	-----	-----	-----	28%	2e-08

For Accession numbers see Appendix 4C. ---- = not recognized by the BLAST algorithms.

4.4.3.2 CD8β - Domain structure

Predicted Ig domains were identified in both *O. fraenata* and *L. hirsutus* at positions 23 to 130 with e-values of 3.51e-08. In both species, a predicted sub-domain called Igc2 lies at positions 29 to 188 within that predicted Ig domain and this sub-domain had an e-value of 2.20e-01. Predicted Igv domains (variable region) were located at positions 33 to 113 with e-values of 3.71e-01. Predicted transmembrane domains at positions 166 to 188 were also identified. A predicted signal peptide cleavage site was located at position 15 in the CD8β sequences of both species.

4.4.3.3 CD8β - Glycosylation and glycation sites

Five predicted O-linked glycosylation sites were found in the mature CD8β chain of both *O. fraenata* and *L. hirsutus.* Four predicted glycation sites were also identified in the CD8β chains of both species. No predicted N-linked glycosylation sites were detected in either species. The locations of the predicted O-linked glycosylation and glycation sites are shown in Table 4.10. However, validation of the glycated residues requires more sophisticated mass spectrometry or Matrix-assisted laser desorption/ ionization-time of flight mass spectrometry (MALDI-TOFF) analysis which was outside the scope of this study.

Table 4.10. Predicted O-linked glycosylation and glycation sites in the CD8β chains of both *O. fraenata* and *L. hirsutus,* and their positions and probabilities.

O-linked glycosylation and glycation sites in CD8β of *O. fraenata* and *L. hirsutus*				
Species	O-linked glycosylation sites	Probability	Glycation sites	Probability
O. fraenata	135	56%	32	92%
	140	56%	141	72%
	143	63%	147	73%
	144	64%	203	78%
	152	52%		
L. hirsutus	137	56%	32	92%
	142	56%	143	80%
	145	62%	149	73%
	146	64%	205	78%
	154	52%		

4.4.3.4 CD8β - Phosphorylation sites

There were a number of predicted phosphorylation sites within the *O. fraenata and L. hirsutus* CD8β sequences. Predicted phosphorylation sites for *O. fraenata* included twelve serine, seven threonine and two tyrosine residues while *L. hirsutus* had twelve serine, six threonine and one tyrosine predicted phosphorylation sites. These positions together with their confidence levels are shown in Table 4.11.

The confidence levels of the predicted phosphorylation sites were mostly above 90%. Notable exceptions included the serine in positions 104 and 106 both of which had a confidence percentage of 59%. This low confidence level suggests that those two sites may not be phosphorylated.

Table 4.11. Predicted phosphorylation sites of CD8β in *O. fraenata* and *L. hirsutus,* their positions, and their respective confidence levels.

Phosphorylation sites of CD8β in *O. fraenata* and *L. hirsutus*						
Species	Serine	Confidence level	Threonine	Confidence level	Tyrosine	Confidence level
O. fraenata	18	95%	30	92%	76	93%
	22	87%	75	94%	92	82%
	43	73%	87	98%		
	44	76%	95	97%		
	45	99%	140	90%		
	59	91%	144	97%		
	62	91%	152	98%		
	88	99%				
	91	99%				
	94	94%				
	104	59%				
	136	99%				
L. hirsutus	18	95%	30	92%	94	82%
	22	87%	77	64%		
	43	73%	97	97%		
	44	74%	142	90%		
	45	98%	146	97%		
	59	91%	154	97%		
	62	90%				
	90	99%				
	93	99%				
	96	94%				
	106	59%				
	138	99%				

4.4.3.5 CD8β - Disulphide bond prediction

Five cysteines were located in the CD8β sequence of both *O. fraenata* and *L. hirsutus*. The *O. fraenata* CD8β chain formed two disulphide bonds while the *L. hirsutus* sequence formed three possible bonds. These bonds were formed between C^{38} and C^{113} and between C^{152} and C^{167}. The DiANNA server also predicted a possible bond for *O. fraenata* between C^{165} and C^{176} but it appeared, after visual inspection of the sequence, that a bond between C^{150} and C^{165} or between C^{150} and C^{176} was more likely. Identifying the correct bonding of the disulphide bridge is important to define the tertiary structure. To detect disulphide bonds, an NMR (nuclear magnetic resonance spectrometry) or mass spectrometric approaches are more accurate methods, but this was outside the scope of this project.

The disulphide bond formation and the probability of connectivity for both species are detailed in Table 4.12.

Table 4.12. Predicted disulphide bond locations in CD8β and their connectivity probabilities for *O. fraenata* and *L. hirsutus.*

Predicted disulphide bonds in CD8β for *O. fraenata* and *L. hirsutus*		
Species	Position	Probability
O. fraenata	38-111	99.7% (**)
	165-176	99.0% (**)
L. hirsutus	38-113	99.7% (**)
	38-167	99.7%(*)
	152-167	99.5%(**)

(*) = not predicted to form a disulphide bridge. (**) = predicted to have connectivity.

4.4.3.6 CD8β - Primary sequence and secondary structure prediction

The mature CD8β protein of *O. fraenata* and *L. hirsutus* consisted of 613bp which translated into 189 amino acids. The start site could not be determined, however the putative signal peptide cleavage site was found and this permitted the characterization of the mature molecule. The putative protein sequence was used for a secondary structure prediction to obtain the order of alpha helices and beta strands in the CD8β chains of the two marsupial species (Figs. 4.6 and 4.7). There was a similar distribution of alpha helices

and beta strands in the two CD8β chain sequences. Only a single alpha helix was detected in the partial putative leader sequence of both the CD8β chains.

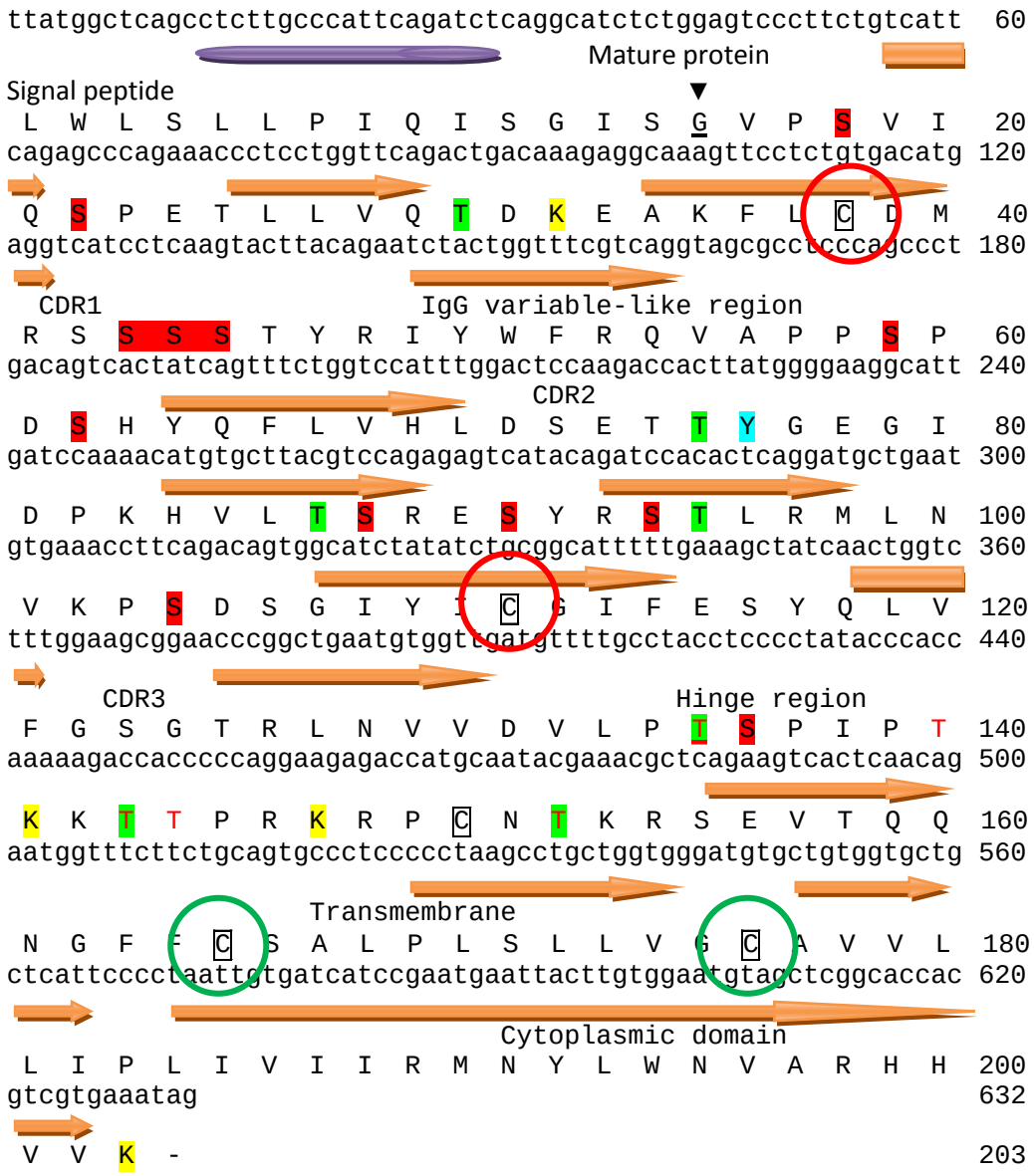


Figure 4.7. *O. fraenata* CD8β primary sequence and secondary structure prediction. G = start of the mature protein. S = Serine O-linked glycosylation sites. K = glycation sites. T = Threonine O-linked phosphorylation. C = disulphide bridges. Y = phosphorylated tyrosine. = Transmembrane helix. = Strand.

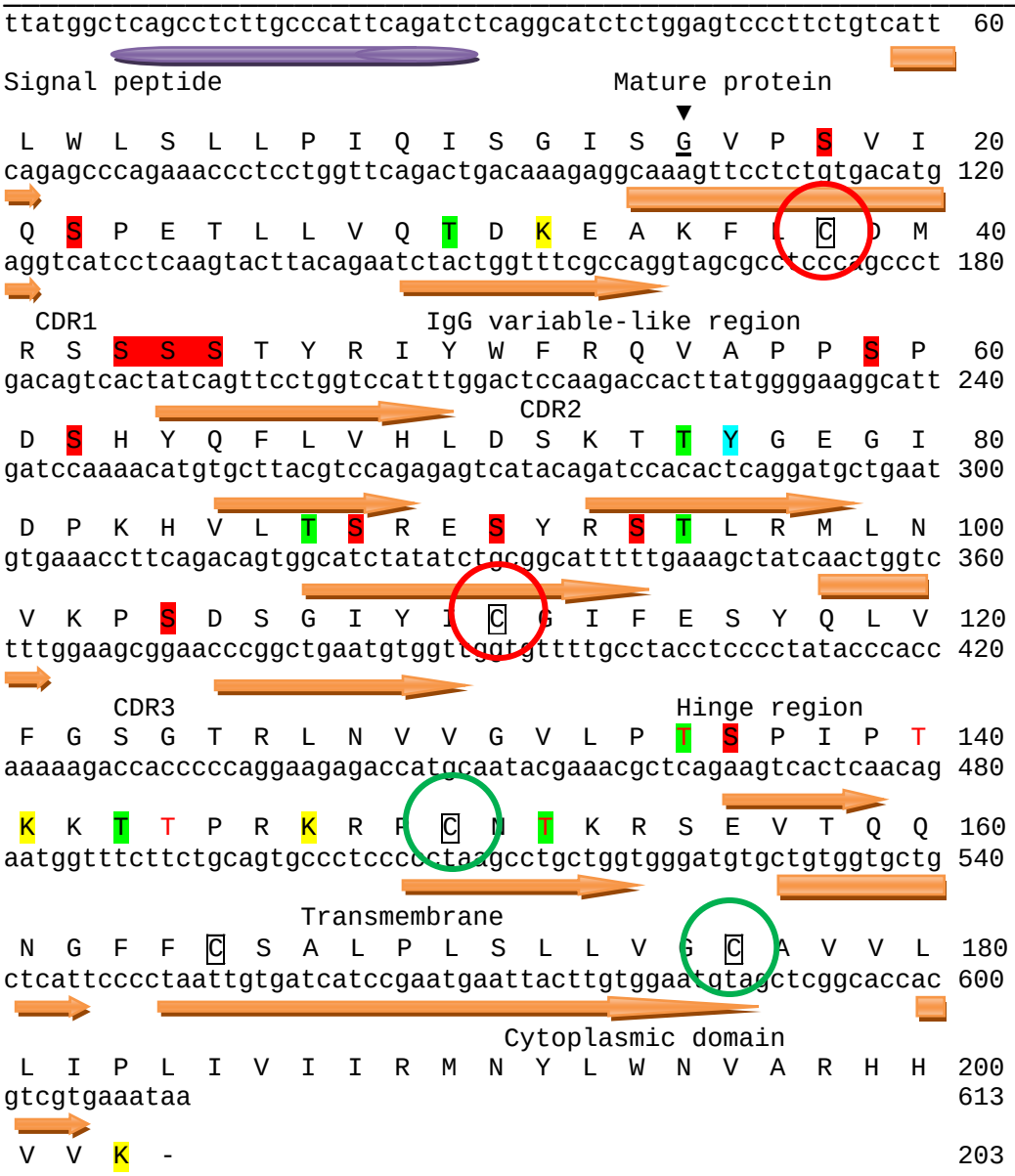


Figure 4.8. *L. hirsutus* CD8β primary sequence and secondary structure prediction.
G = start of the mature protein. S = Serine O-linked glycosylation sites. K = glycation sites.
T = Threonine O-linked phosphorylation. C = disulphide bridges. Y = phosphorylated tyrosine.
= Transmembrane helix. = Strand.

4.4.3.7 CD8β - Tertiary structure and ligand binding predictions

The tertiary structure and the ligand binding prediction were determined for the mature CD8β chains in *O. fraenata* and *L. hirsutus* and also for the published sequences of *M. eugenii* and M*. domestica*. The results are shown in Fig. 4.9 (a, b). The tertiary structures appeared to be very similar in the four marsupial CD8β chains but a closer examination of the ligand binding capacity showed differences between the species (Fig. 4.9b). The Swiss Model platform was used to compare the results with the pdb structure 3dmm_1, which is an immune co-receptor complex. This particular model is annotated as a hetero-tetramer but the marsupial sequences could only be modelled as a single chain by the software.

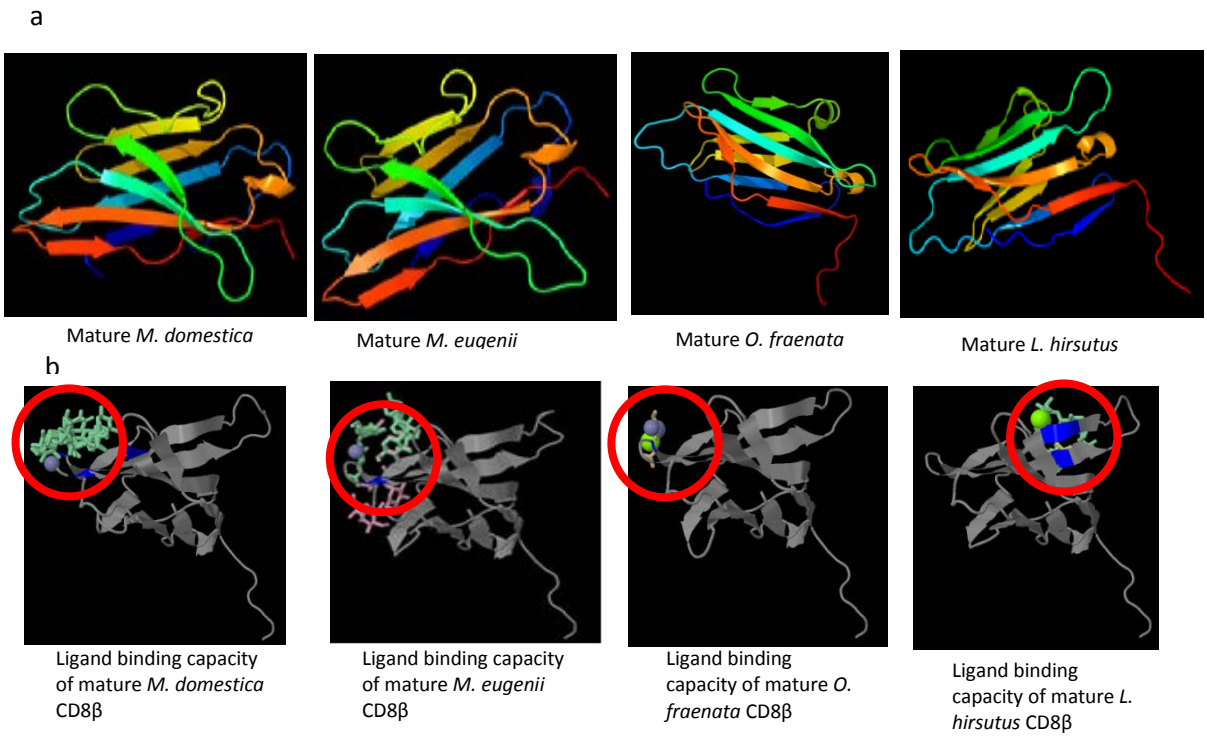


Figure 4.9. Prediction of the mature CD8β chains in *M. domestica, M. eugenii, O. fraenata* and *L. hirsutus.* (a) Tertiary structures. (b) Ligand binding capacities. The areas of the CD8β molecule involved in the ligand binding are circled in red.

The Z-scores for the *O. fraenata* and *L. hirsutus* CD8β chains were -3.29 and -1.96 respectively (Fig. 4.10 (a,c)). The model predictions for the two species were very similar in the values for the Cβ interaction, all-atom interaction and solvation. However the torsion value was significantly different between *O. fraenata* and *L. hirsutus* and this is reflected in the lower QMEAN value. This is illustrated in Fig. 4.10.

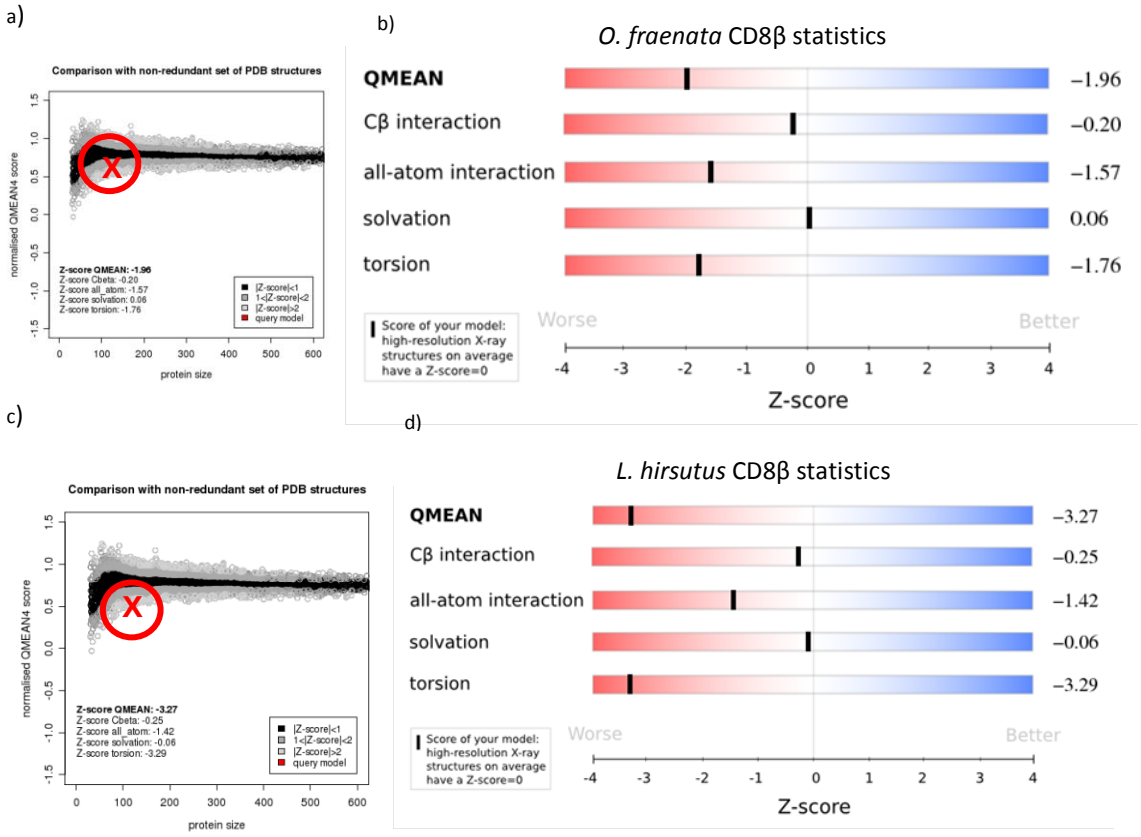


Figure 4.10. Comparison of *O. fraenata* and *L. hirsutus* CD8β structures with the pdb structures. (a) X = *O. fraenata* CD8β sequence compared to homologous structures in the pdb database. (b) Z- score slider indicating the quality of the *O. fraenata* CD8β model showing the QMEAN. (c) X = *L. hirsutus* CD8β sequence compared to homologous structures in the pdb database. (d) Z-score slider indicating the quality of the *L. hirsutus* CD8β model showing the QMEAN.

Unlike the Swiss-Model program, the Modweb software utilizes the Modeller 9.10 platform. This program was used to compare the *O. fraenata* and *L. hirsutus* CD8β chains with the pdb structure 2atpB (crystal structure of a CD8αβ heterodimer). The Modweb software was able to model the *L. hirsutus* CD8β residues at positions 19 to 132 with a reliability of 44%. In contrast, the same segment in *O. fraenata* could only be modelled with 26% reliability. The Modweb program identified the model 2atpB from the pdb database as the closest related structure to the *O. fraenata* and *L. hirsutus* CD8β models.

The data shown in Table 4.13 strengthens the ligand prediction shown in Fig. 4.8(b). There were differences in the number of residues involved in ligand binding in all four marsupial CD8β chains and this is pictorially illustrated in Fig. 4.8(b). The tertiary structures were all similar to each other thus indicating that the function of the molecule is most likely conserved in all four species. The *O. fraenata* and *L. hirsutus* CD8β sequences showed little difference to the *M. eugenii* sequence, but when compared to the *M. domestica* CD8β molecule the difference in their ligand binding ability was significantly lower. Therefore the ability of the macropod CD8β to bind with important ligands differs to that of the didelphid. This indicates that the CD8β chain in the didelphid *M. domestica* is more active than the CD8β chain in the macropods due to the greater number of available binding sites (Table 4.13).

Table 4.13. Predicted ligand binding sites of the mature protein CD8β. The residues, their positions and the predicted number of binding sites are shown for *O. fraenata*, *L. hirsutus*, *M. domestica* and *M. eugenii*.

Ligand binding capacity of CD8β in four marsupials			
Species	Residue position	Residue	Predicted binding site (contacts)
O. fraenata	72	Serine	5
	74	Threonine	3
L. hirsutus	72	Serine	5
	74	Threonine	3
M. domestica	51	Tyrosine	20
	54	Tyrosine	8
	76	Glutamine	11
	78	Arginine	9
	121	Isoleucine	10
M. eugenii	31	Tyrosine	5
	56	Glutamine	5

4.4.3.8 CD8β - Phylogenetic analysis

The evolutionary history was inferred using a Neighbor-Joining tree. The optimal tree had a branch length sum of 8.20. The percentages of replicate trees in which the associated taxa were clustered together in the bootstrap test (1,000 replicates) were shown next to the branches. The marsupial species cluster together and are situated between the rodents and the birds. There is a clear distinction between the macropods *M. eugenii*, *O. fraenata* and *L. hirsutus* and the didelphid *M. domestica*. This was expected since *M. domestica* is only a distant relative of the macropods. The branch length of the macropod species was short indicating a close relationship between them. On the other hand, the *M. domestica* branch is longer than that of the macropods thus indicating a greater evolutionary distance between those groups. The evolutionary distances were computed using the Poisson correction method and the units are the number of amino acid substitutions per site.

The phylogenetic tree is drawn to scale, and the branch lengths are in the same units as those of the evolutionary distances used to infer the tree. This analysis involved 25 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 133 positions in the final dataset (Fig. 4.11).

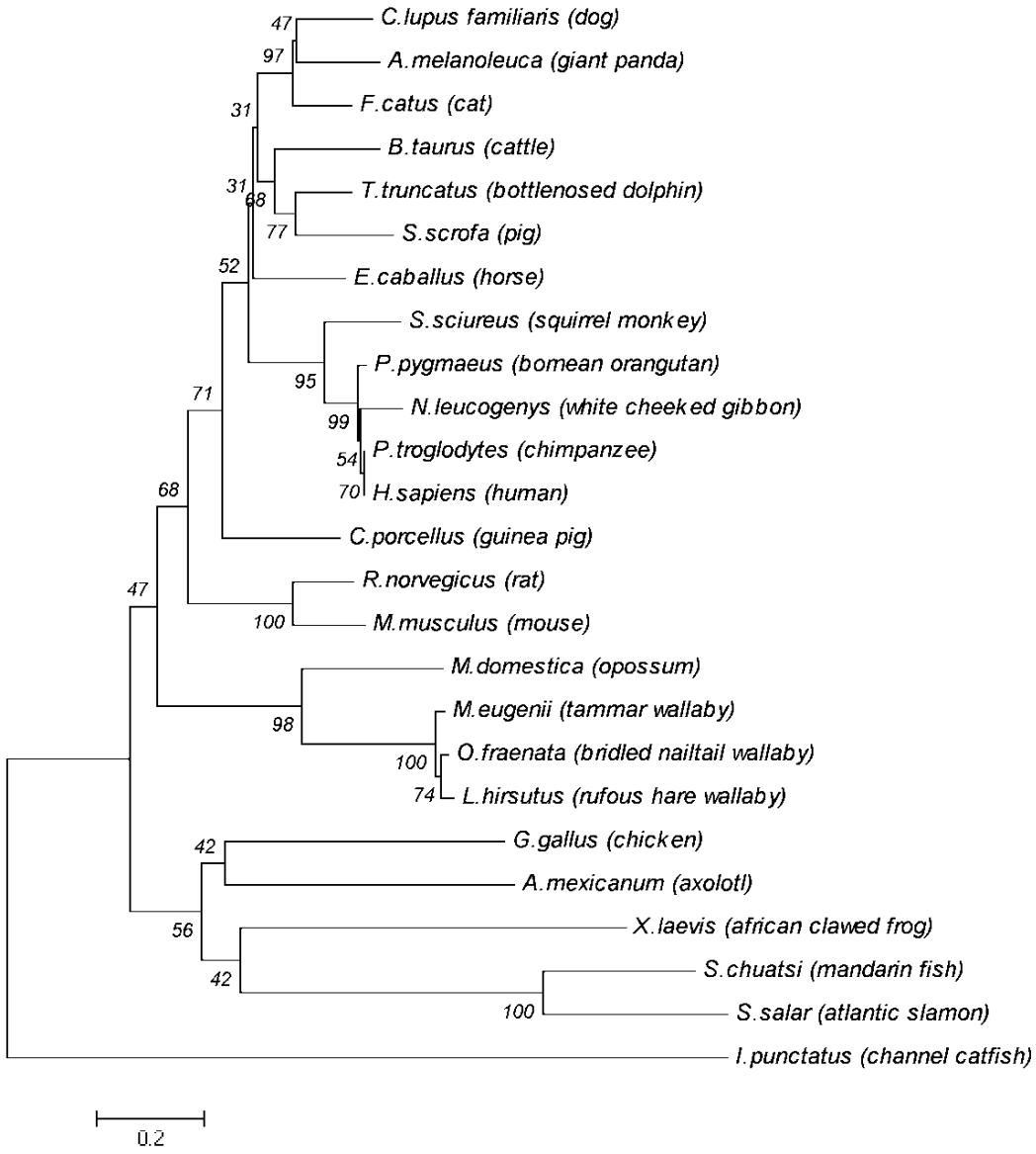


Figure 4.11. Neighbor-Joining phylogenetic tree of the CD8β chain. Short branch lengths of the macropods *M. eugenii*, *O. fraenata* and *L. hirsutus* indicates the close relationship between those species. *M. domestica* has a longer branch length and is therefore more distantly related.

4.4.4 CD28

A product of the *M. eugenii* CD28 molecule was obtained by RT-PCR and portions of the 5' and 3' ends of the sequence were subsequently elucidated with RACE-PCRs. The same strategy was unsuccessful for *O. fraenata.* A polymerase chain reaction was carried out for *O. fraenata* CD28 and an amplicon of the correct size was visualized. However the product was faint and a cloning procedure failed to produce a clone with the correct insert. A number of optimizations were carried out but the sequence of CD28 could not be determined in *O. fraenata*. Consequently the following sub-sections relate only to the *M. eugenii* CD28 molecule.

4.4.4.1 CD28 - Homology

A BLAST homology search showed that the *M. eugenii* CD28 had an 84% identity at the nucleotide level and 65% at the putative amino acid level to the predicted CD28 sequence of *M. domestica*. The sequence homology of *M. eugenii* CD28 to the human sequence was 77% at the nucleotide level and 54% at the putative amino acid level. Interestingly, the *M. eugenii* CD28 sequence had a 70% identity to the chicken CD28 sequence at the nucleotide level and 51% identity at the putative amino acid level. The closeness of the identities in those three species, which are far removed from each other in an evolutionary sense, indicates that the CD28 molecule has an important function in the immune system. Overall, the amino acid identity percentages were much lower than the nucleotide percentages thus indicating gap penalties and low complexity sequences. The results of the homology search are shown in Table 4.14.

Table 4.14. Homology search results for the *M.eugenii* CD28 nucleotide and amino acid sequences and their identities and respective e-values.

CD28 homology search results for *M. eugenii*				
Species	Nucleotide	e-value	Amino acid	e-value
M. domestica (predicted)	84%	4e-170	65%	6e-100
H. sapiens	77%	1e-30	54%	1e-71
G. gallus	70%	3e-19	51%	2e-57
M. fascicularis	69%	8e-52	54%	3e-71
E. maximus	69%	1e-49	55%	6e-74
G. camelopardalis	68%	3e-57	58%	3e-83
B. taurus	68%	7e-53	57%	3e-82
R. unicornus	68%	5e-55	58%	7e-78
M. mulatta	67%	6e-54	56%	2e-69
S. caffer	67%	1e-50	-------	-------
B. bonasus	67%	4e-49	-------	-------
E. grevyi	67%	4e-49	57%	3e-77
E. zebra hartmannae	67%	4e-49	57%	3e-77
O. aries	67%	4e-49	57%	2e-81
E. asinus somalicus	67%	2e-46	56%	1e-76
M. monax	67%	2e-40	55%	6e-74
O. cuniculus	65%	2e-33	56%	3e-78
R. norvegicus	65%	3e-32	55%	1e-69
X. silurana	-------	-------	32%	1e-21

For Accession numbers see Appendix 4D. ----- = not recognized by the BLAST algorithms

4.4.4.2 CD28 - Domain structure

The *M. eugenii* CD28 sequence contained a putative signal peptide spanning 18 amino acids beginning at position 1 (Methionine). The putative signal peptide was followed by a predicted Ig-like domain starting at positions 36 to 115. This region had an e-value of 2.45e-01. The sequence also contained a predicted transmembrane domain at positions 154 to 176.

4.4.4.3 CD28 - Glycosylation and glycation sites

The *M. eugenii* CD28 sequence contained six predicted N-linked glycosylation sites. In contrast, the *M. domestica* partial sequence had only four putative N-linked glycosylation sites. Two putative O-linked glycosylation sites at positions 144 and 147 were also contained within the *M. eugenii* CD28 sequence. Predicted glycation sites were found at positions 4, 47, 97, 128, 184 and 199. Position 162 in the sequence was one of the six predicted N-linked glycosylation sites. However the confidence level was only 11%, and

the location of the motif in the cytoplasmic domain made it highly unlikely that it was a functional N-linked glycosylation site (Table 4.15).

Table 4.15. Predicted O- and N-linked glycosylation and glycation sites in the *M. eugenii* CD28 and their confidence levels.

Predicted O- and N-linked glycosylation and glycation sites in *M. eugenii* CD28									
O-linked glycosylation sites	Confidence level	N-linked glycosylation sites and sequences		Confidence level		Glycation sites	Confidence level
144	57.2%	40	NGTH	74%	++	4	95.5%
147	54.0%	54	NCSV	61%	+	47	90.3%
		61	NKTV	58%	+	97	91.8%
		70	NMSI	68%	++	128	76.7%
		74	NQTD	61%	+	184	93.4%
		98	NGTI	70%	++	199	62.2%
		162	NMTP*	11%	-----		

+ = below 65%. ++ = above 65%. * = highly unlikely, located in the cytoplasmic domain. ----- = no signal strength.

Validation of the glycated residues required more sophisticated mass spectrometry or Matrix-assisted laser desorption/ ionization-time of flight mass spectrometry (MALDI-TOFF) analysis which was outside the scope of this study.

4.4.4.4 CD28 - Disulphide bond prediction

Four disulphide bonds were predicted in the *M. eugenii* CD28 sequence together with their connectivity (Table 4.16). Since two half-cysteines are in involved in a disulphide bond it was assumed that the CD28 sequence, which contains eight cysteines, would form four disulphide bonds. Since unpaired cysteine residues can lead to protein agglomeration due to the thiol-thiol interaction, it was assumed that all cysteine pairs naturally would form a disulphide bond. However the prediction program only showed a single possible disulphide bond for the *M. eugenii* CD28 sequence. This may be the reason why the *M. eugenii* CD28 molecule could only be modelled as a monomer.

Table 4.16. Predicted disulphide bonds in *M. eugenii* CD28, their locations and their corresponding sequences.

Predicted disulphide bonds in *M. eugenii* CD28	
Residue number	Corresponding sequences
41-113	VATLSNYO**C**D – TDIYF**C**KIEFM *
45-207	SCNYI**C**DKTPT - PTNKL**C**LPYAP
67-87	SLEV**C**FVYVN - MEDFN**C**SVNFD
162-178	VVAAL**C**VLAFY - VTFFN**C**WLKIK

* = highest probability. **C** = cysteine residues.

4.4.4.5 CD28 - Phosphorylation sites

Five serine, six threonine and three tyrosine predicted phosphorylation sites were identified in the *M. eugenii* CD28 sequence by the EXPASY suite of programs. The confidence levels of the predicted phosphorylation sites ranged from 52% to 92%. Their positions are shown in Table 4.17.

Table 4.17. Predicted phosphorylation sites in *M. eugenii* CD28, their positions and confidence levels.

M. eugenii CD28 phosphorylated sites and their confidence levels					
Serine		Threonine		Tyrosine	
Position	Confidence level	Position	Confidence level	Position	Confidence level
55	97%	48	90%	70	65%
80	71%	50	52%	123	86%
81	99%	74	72%	192	76%
148	81%	147	64%		
150	92%	196	98%		
		203	74%		

4.4.4.6 CD28 - Primary sequence and secondary structure prediction

The open reading frame of *M. eugenii* CD28 contained 666bp that translated into 333 amino acids. In *M. domestica* this gene was partially amplified obtaining 466bp that translated into 155 amino acids.

The putative *M. eugenii* CD28 protein sequence was used for a secondary structure prediction to obtain the order of helices and strands. A transmembrane helix was found at position 160 with the sequence WAVVAALCVLAFYVLLMTVTFFNCW in the transmembrane region. This transmembrane helix appeared to be a conserved structure,

however the sequence was not conserved between *M. eugenii* and *M. domestica*. The functionally important MYPPPY motif was conserved in both species.

Twelve beta sheets were detected in the partially amplified *M. domestica* CD28 sequence mentioned above. When these were compared with the beta sheets in the equivalent section of the *M. eugenii* sequence, variations in the sheet lengths were detected (Figs. 4.12 and 4.13).

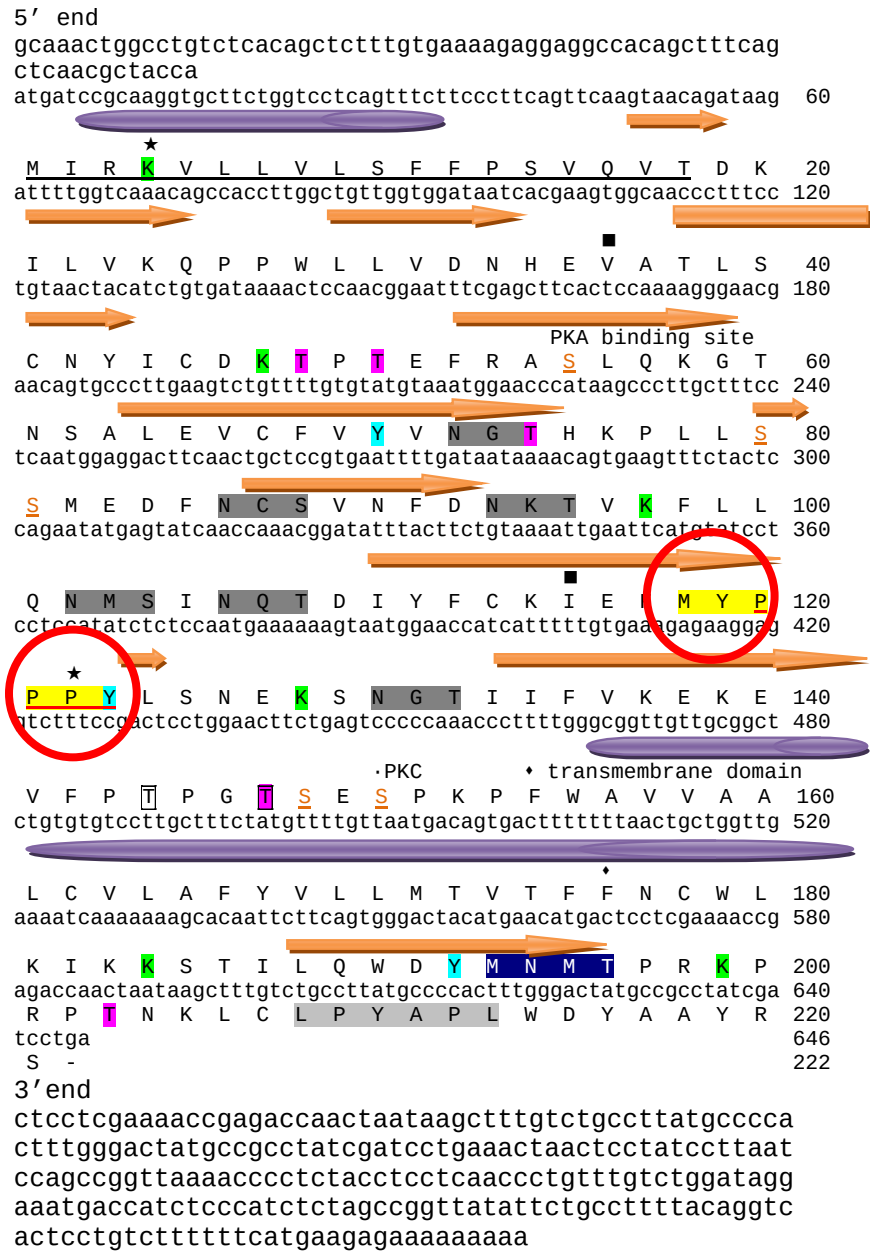


Figure 4.12. *M. eugenii* CD28 primary sequence and secondary structure prediction. NGT = N-linked glycosylation sites. K = Glycation sites. S = Serine phosphorylation sites. T = O-linked glycosylation sites. T = threonine phosphorylation sites. Y = tyrosine phosphorylation sites. • = PKC binding site. ■ = Ig like domain, ITIM = immunoreceptor tyrosine-based inhibitory motif. ♦ = Transmembrane domain. ★ = AARP2CN similar to GTP binding domain. NMT = not a possible N-linked glycosylation site. ○ = MYPPPY binding site. = Transmembrane helix. = Strand.

Figure 4.13. *M. domestica* CD28 partial primary sequence and secondary structure prediction. NGT = N-linked glycosylation sites. = Transmembrane helix. = Strand. Highlighted in yellow = MYPPPY binding motif.

4.4.4.7 CD28 - Tertiary structure and ligand binding predictions

The beta strands predicted in the secondary structure were clearly seen in the predicted tertiary structure (Fig. 4.14a). The predicted ligand binding sites of *M. eugenii* CD28 are illustrated in Fig. 4.14b and the residues involved in ligand binding as predicted by the I-TASSER program are shown in Table 4.18.

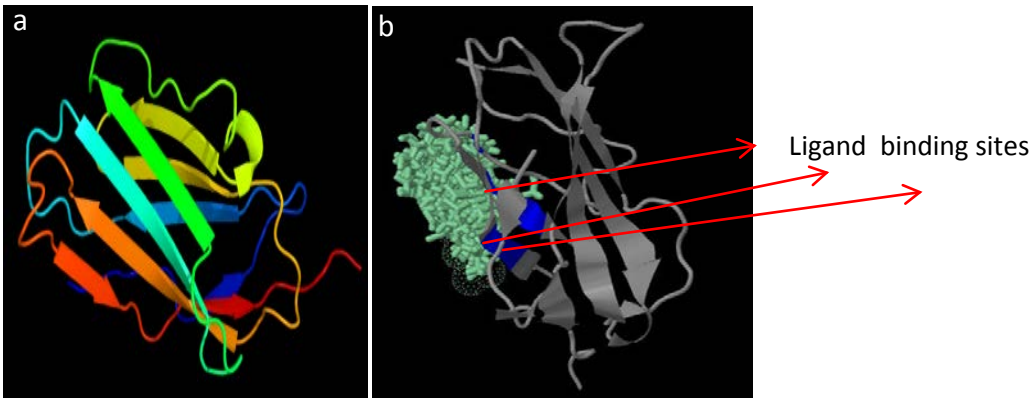


Figure 4.14. M. eugenii CD28 tertiary structure and ligand binding sites predicted with the I-TASSER program. (a) Tertiary structure of *M. eugenii* CD28. (b) Ligand binding sites of *M. eugenii* CD28 (dark blue).

Table 4.18 *M. eugenii* CD28 predicted ligand binding sites, positions, and amino acid identities.

M. eugenii CD28 ligand binding sites		
Species	Residue position	Amino Acid identity
M. eugenii	38	Threonine
	39	Leucine
	90	Asparagine
	97	Lysine
	98	Phenylalanine
	99	Leucine

The Z-score slider in Fig. 4.15 (a) indicated the location of the model in relation to CD28 structures solved by X-ray crystallography. The quality of the *M. eugenii* CD28 model, evidenced by the QMEAN statistics, was better than that of any other molecular model reported in this thesis. The Modweb program, which was used to investigate the *M. eugenii* CD28 sequence, showed that the structure prediction for that sequence was very close to other predicted structures found in the pdb database. The QMEAN was located between -1 and zero which indicated that the prediction was highly probable. When normalized by the program, the *M. eugenii* CD28 model was shown to be closely positioned to the known CD28 structures of other species (Fig. 4.13b).

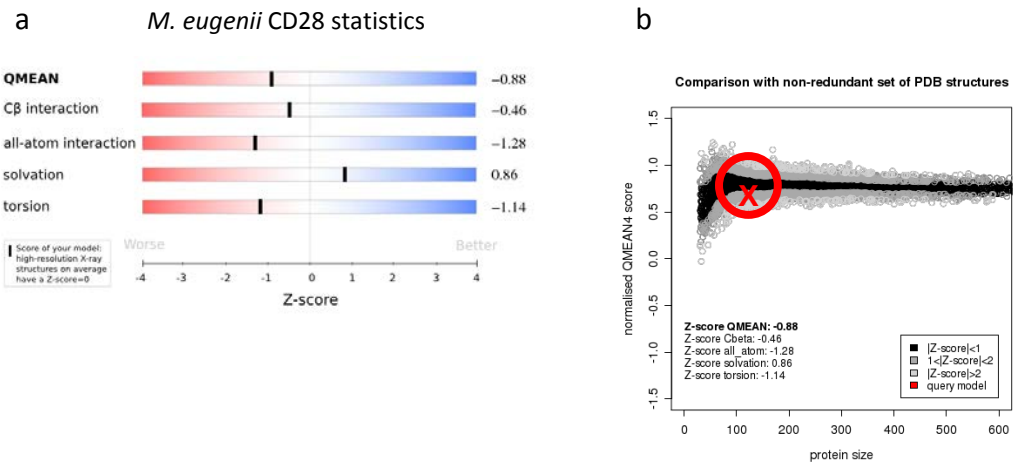


Figure 4.15. *M. eugenii* CD28 statistics. (a) Z-score slider for CD28. (b) Comparison of *M. eugenii* CD28 model (X) to other mammalian structures in the pdb resolved using X-ray crystallography.

4.4.4.8 CD28 - Phylogenetic analysis

The phylogenetic analysis indicated that the *M. eugenii* CD28 molecule was well removed from *M. domestica* in an evolutionary sense. The branch length suggested that *M. eugenii* was more closely related to rodents, cats and dogs rather than to *M. domestica*. This is an unusual result since *M. eugenii* and *M. domestica*, although not closely related, are both marsupials and it would be expected that they would be closer together on the phylogenetic tree. This apparent anomaly may be due to sequence variation in the 5' end of the molecule and the Ig-domain (Appendix 4C), or the predicted *M. domestica* CD28 sequence may have been incorrectly annotated.

The optimal phylogenetic tree with a sum of branch length of 8.18 is shown in Fig. 4.16. The percentages of the replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the tree. The analysis involved 22 amino acid sequences and all positions containing gaps and missing data were eliminated. There were a total of 152 positions in the final dataset.

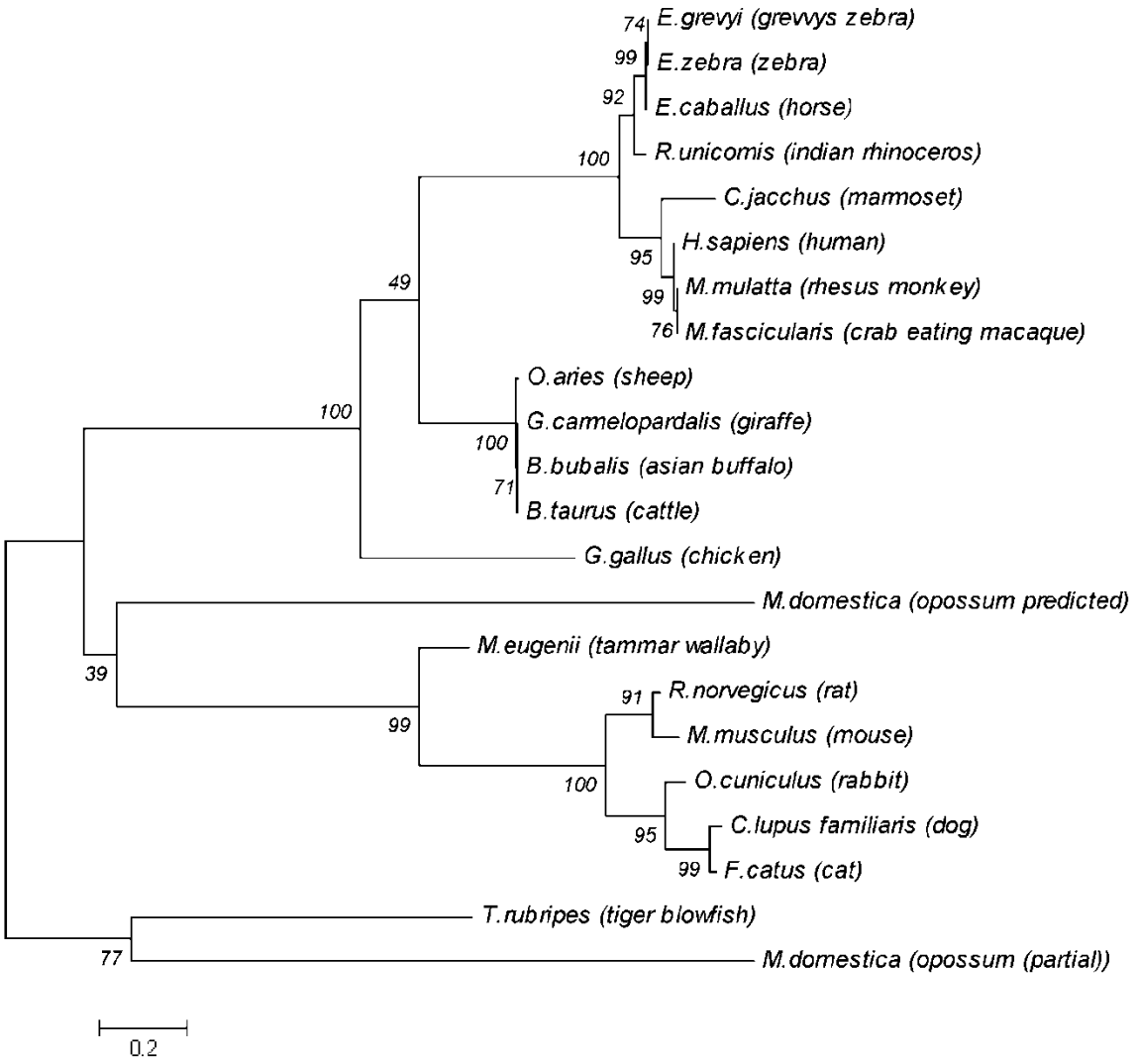


Figure 4.16. Neighbor-Joining phylogenetic tree for CD28. Branch lengths of the marsupials *M. eugenii* and *M. domestica* show great variation indicating the evolutionary distance between the two species.

4.4.5 Cytotoxic T lymphocyte antigen-4 (CTLA-4)

CTLA-4 was characterized in *O. fraenata* and *M. eugenii* and a partial sequence was characterized in *M. domestica*.

4.4.5.1 CTLA-4 - Homology

A homology search was carried out with the BLAST2 program to compare the sequences of *O. fraenata* and *M. eugenii* to each other, while the BLAST algorithm was used to compare the predicted *M. domestica* sequence to that of the macropods and other vertebrate species. This investigation revealed that the *M. eugenii* CTLA-4 sequence had an 88% identity to the predicted *M. domestica* sequence and a 97% identity to that of *O. fraenata*. Homology percentages to other species are shown in Table 4.19.

Table 4.19. Homology search results for the *O. fraenata* and *M. eugenii* CTLA-4 nucleotide and amino acid sequences, their identities and their respective e-values.

CTLA-4 homology search results for *O. fraenata* and *M. eugenii*								
	O. fraenata				*M. eugenii*			
Species	Nucleotide	e-value	Amino acid	e-value	Nucleotide	e-value	Amino acid	e-value
M. domestica (predicted)	86%	0.0	81%	4e-133	88%	0.0	83%	1e-136
C. porcellus	76%	3e-57	62%	4e-96	84%	0.91	62%	1e-95
F. catus	73%	3e-102	66%	4e-104	74%	1e-107	69%	2e-102
M. mulatta	73%	1e-94	64%	2e-100	73%	5e-105	65%	1e-100
M. monax	73%	2e-97	66%	3e-104	74%	1e-106	66%	4e-105
M. nemestrina	73%	5e-93	------	------	73%	2e-103	------	------
C. familiaris	73%	7e-97	69%	1e-102	73%	2e-103	70%	2e-103
P. abelii	72%	2e-92	65%	1e-103	73%	3e-102	65%	1e-102
H. sapiens	72%	2e-97	64%	5e-103	73%	1e-100	58%	4e-64
C. jacchus	72%	2e-84	65%	2e-102	73%	1e-100	71%	4e-92
S. scrofa	72%	1e-88	67%	3e-100	73%	1e-99	68%	1e-100
A. melanoleuca	72%	2e-91	65%	4e-102	73%	6e-98	65%	1e-101
L. Africana	72%	9e-96	67%	1e-105	74%	1e106	70%	1e-105
N. leucogenyx	72%	4e-94	65%	4e-103	73%	2e-103	65%	1e-102
O. aries	71%	5e-86	68%	7e-100	72%	4e-94	67%	3e-99
B. bubalis	71%	2e-86	67%	3e-100	72%	6e-92	67%	1e-99
B. taurus	71%	8e-84	67%	7e-100	72%	3e-89	66%	3e-99
G. gallus	65%	6e-16	44%	2e-59	65%	5e-17	47%	1e-59

For accession number see Appendix 4D. ------ = not recognized by BLAST algorithms.

4.4.5.2 CTLA-4 - Domain structure

CTLA-4 of *O. fraenata* and *M. eugenii* consisted of a predicted signal peptide, an Ig-like domain and a predicted transmembrane domain. The Ig-like domain of *O. fraenata* was predicted with an e-value of 2.63e-01, while the Ig-like domain of *M. eugenii* was predicted with an e-value of 4.48e-02. This indicated that they both had a high probability.

4.4.5.3 CTLA-4 - Glycosylation and glycation sites

One predicted O-linked glycosylation site was found in both the *O. fraenata* and *M. eugenii* CTLA-4 sequences at position 40. Three putative glycation sites were predicted at positions 12, 111, and 192 in *O. fraenata*, while the *M. eugenii* sequence showed five putative glycation sites at positions 12, 17, 50, 111 and 192. Two N-linked glycosylation sites at positions 63 and 145 were predicted with probability values of 66% and 80% respectively in both *O. fraenata* and *M. eugenii* (Table 4.20).

Table 4.20. Predicted O- and N-linked glycosylation and glycation sites in *O. fraenata* and *M. eugenii* CTLA-4 and their respective confidence levels.

Predicted O- and N-linked glycosylation and glycation sites for CTLA-4 in *O. fraenata* and *M. eugenii*						
Species	O-linked glycosylation sites	Confidence levels	N-linked glycosylation sites and sequences	Confidence levels	Glycation sites	Confidence levels
M. eugenii	40	63%	63 NKKT 145 NGTQ	65% 79%	12 17 50 111 192	95.6% 83.1% 85.2% 92.4% 92.8%
O. fraenata	40	63%	63 NKTT 145 NGTQ	66% 80%	12 111 192	95.8% 84.5% 92.8%

4.4.5.4 CTLA-4 - Phosphorylation sites

Three serine and three tyrosine phosphorylation sites were predicted in *O. fraenata* and *M. eugenii* CTLA-4. Only one phosphorylated threonine site was predicted in *M. eugenii,* and none were predicted in *O. fraenata* (Table 4.21).

Table 4.21. Predicted phosphorylation sites of CTLA-4 and their respective probabilities in *O. fraenata* and *M. eugenii*.

Predicted phosphorylation sites of CTLA-4 in *O. fraenata* and *M. eugenii*				
	O. fraenata		*M. eugenii*	
Amino acid	Location	Probabilities	Location	Probabilities
Serine	48 62 122	77% 79% 94%	6 48 122	61% 94% 69%
Threonine	none	none	62	85%
Tyrosine	127 140 201	56% 63% 84%	127 140 201	56% 63% 84%

4.4.5.5 CTLA-4 - Disulphide bond prediction

Seven cysteine residues were predicted in the *O. fraenata* CTLA-4 sequence and six were predicted in the *M. eugenii* sequence. This indicated that the possible disulphide bond distribution was different in the two species. The locations of the putative disulphide bonds for each species are shown in Table 4.22.

The third putative disulphide bond in the CTLA-4 molecule in *O. fraenata* contained 19 amino acids while that of *M. eugenii* contained 45 amino acids. The difference in the lengths of the third disulphide bond may indicate different fold properties of the molecule in the two species (Table 4.22).

Table 4.22. Positions and probabilities of predicted disulphide bonds in CTLA-4 in *O. fraenata* and *M. eugenii.*

Predicted disulphide bonds in *O. fraenata* and *M. eugenii*		
Species	Amino Acid residue positions	Probability
O. fraenata (7 cysteines)	21-129 57-157 84-103	99.6% 69.1% 88.5%
M. eugenii (6 cysteines)	21-129 57-157 84-129	99.7% 99.8% 87.2%

4.4.5.6 CTLA-4 - Primary sequences and secondary structure prediction

The open reading frames of both *O. fraenata* and *M. eugenii* CTLA-4 sequences consisted of 669bp that translated into 223 amino acids. The *M. domestica* partial CTLA-4 sequence

had 210bp which translated into 70 amino acid residues. An intracellular localization motif was found in all three CTLA-4 sequences.

Fourteen strands and a single helix were found in the putative secondary structure of *O. fraenata*, and 13 strands and two helices were found in *M. eugenii*. There is a clear difference in the secondary structures of the *O. fraenata* and the *M. eugenii* sequences. While *M. eugenii* has a PKC binding site at position 6 of the putative amino acid sequence, the *O. fraenata* sequence is devoid of this binding site. The *M. eugenii* sequence showed an alpha helix beginning at position 21, but this is missing in the *O. fraenata* sequence. There are also differences in the number of beta sheets that are contained within the sequences. The effect of this on the structures of the *O. fraenata* and *M. eugenii* CTLA-4 molecule is not known. While the lengths of the beta sheets might not have a great impact on the functionality of the molecule, the extra helix in the *M. eugenii* sequence will change the topography of that molecule. Even if the helix is buried, the molecule will be bulkier and will have a different torsion. This appeared to be reflected in the structure prediction (Figs.4.20 and 4.21).

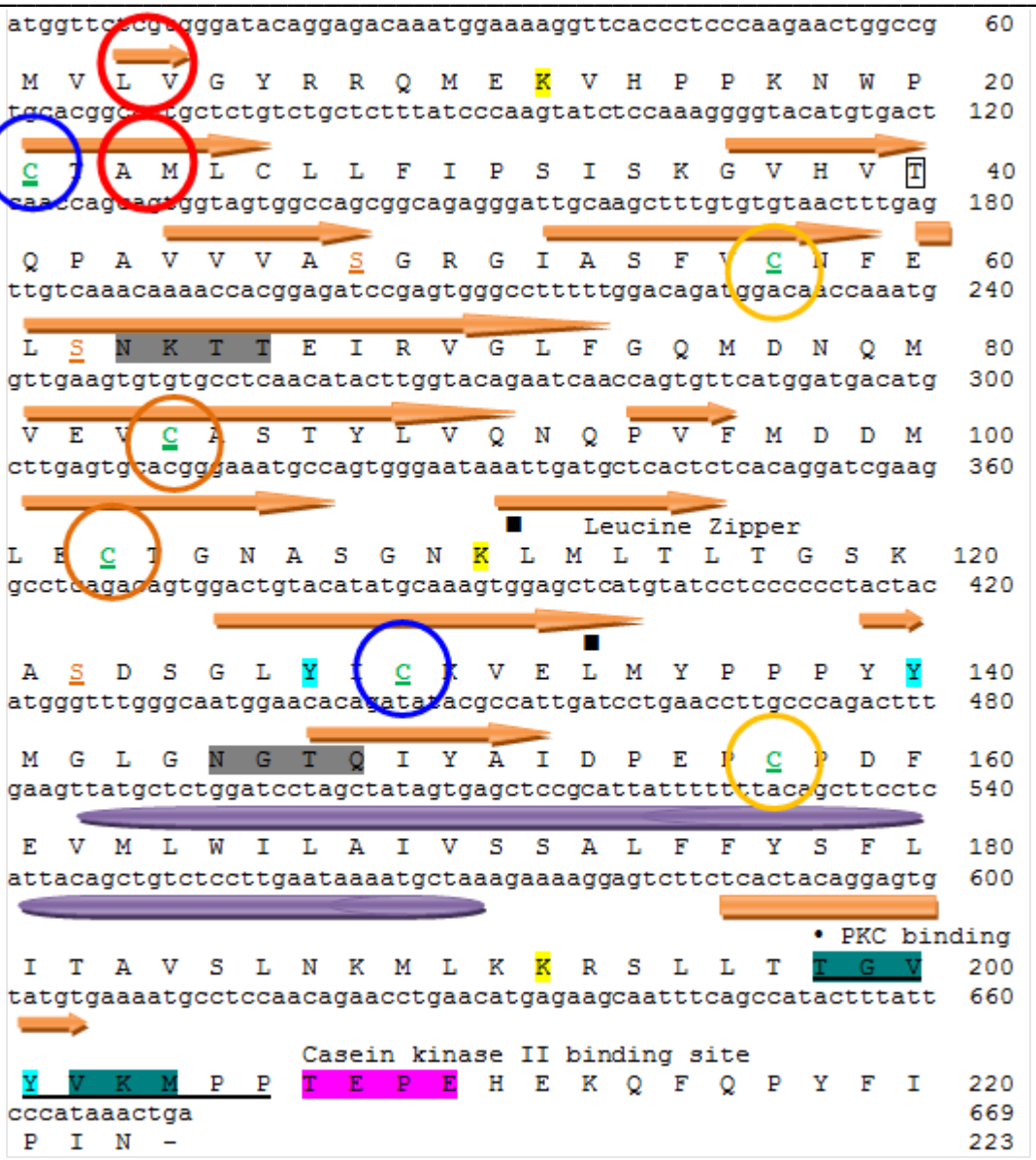


Figure 4.17. *O. fraenata* CTLA-4 primary sequence and secondary structure prediction. K = glycation sites. NGT = N- linked glycosylation sites. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites, C = disulphide bonds. TGVYVKM = intracellular localization motif. = differences to the *M. eugenii* sequence. Other circles = disulphide bonds. ■ = Leucine zipper. TEPE = Casein kinase II binding site. = Transmembrane helix. = Strand.

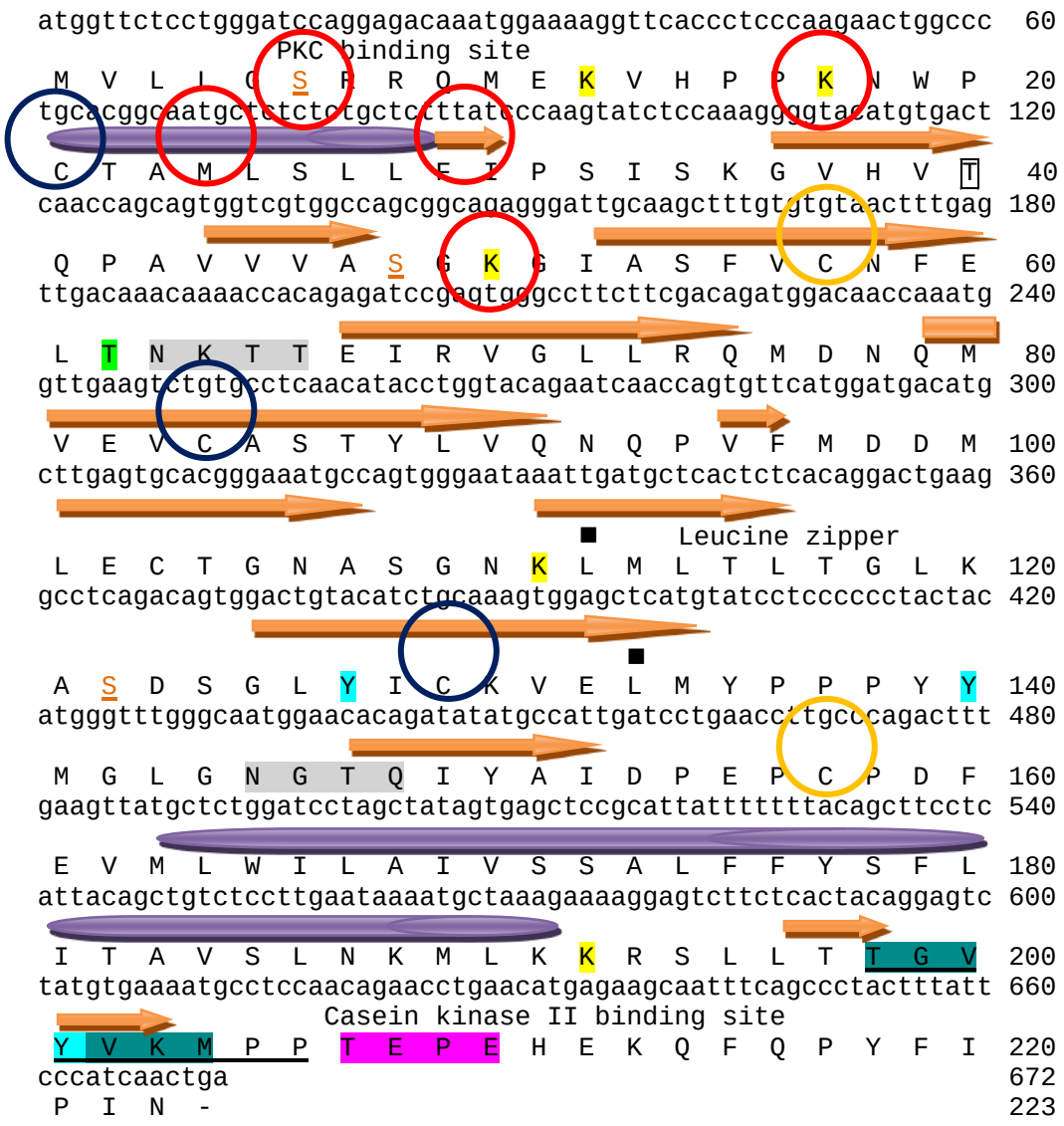


Figure 4.18. *M. eugenii* CTLA-4 primary sequence and secondary structure prediction. K = glycation sites. NGT = N- linked glycosylation sites. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. TGVYVKM = intracellular localization motif. = differences to the *M. eugenii* sequence. Other circles = disulphide bonds. ■ = Leucine zipper. TEPE = Casein kinase II binding site. = Transmembrane helix. = Strand.

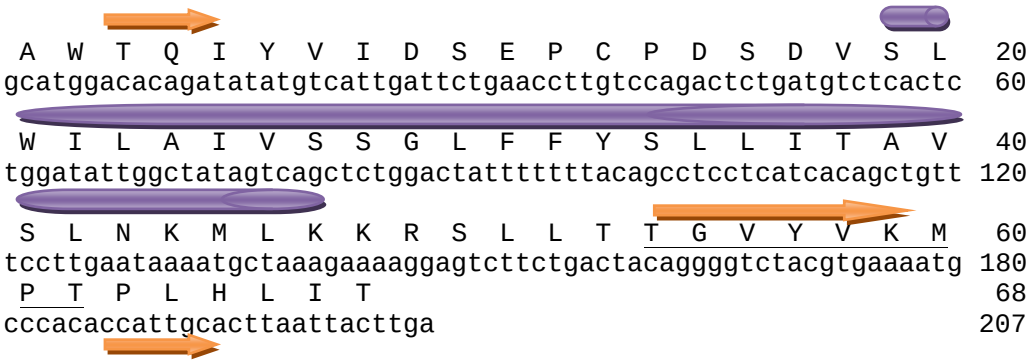


Figure 4.19. *M. domestica* CTLA-4 partial sequence.
TGVYVMPPT = intracellular localization motif. = Transmembrane helix.
= Strand.

4.4.5.7 CTLA-4 - Tertiary structure and ligand binding site predictions

The tertiary structures of the *O. fraenata* and *M. eugenii* CTLA-4 molecules predicted by I-TASSER appeared to be similar to each other (Fig. 4.20 a, c). The ligand binding abilities and their positions in the CTLA-4 sequence, also predicted by I-TASSER, were identical in both species (Fig. 4.20 b, d, and Table 4.23).

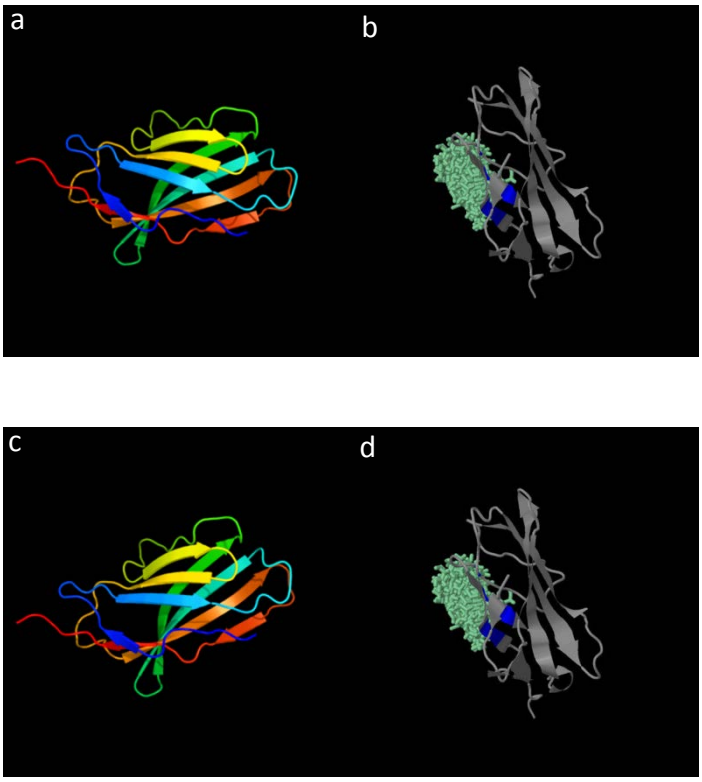


Figure 4.20. CTLA-4 tertiary structures and ligand binding capacities for *O. fraenata* and *M. eugenii* predicted by I-TASSER. (a) *O. fraenata* tertiary structure. (b) *O. fraenata* ligand binding sites. (c) *M. eugenii* tertiary structure. (d) *M. eugenii* ligand binding sites.

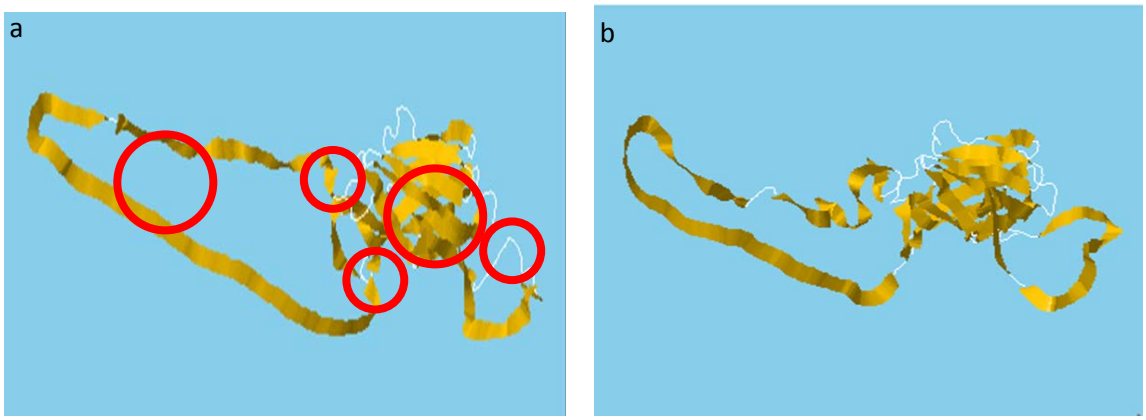
Table 4.23. Predicted ligand binding sites, the residues involved and their positions in CTLA-4 in *O. fraenata* and *M. eugenii.*

Predicted ligand binding sites in CTLA-4 in *O. fraenata* and *L. hirsutus*			
Species	Residue position	Residue	Predicted binding site (contacts)
O. fraenata	54	Serine	31
	55	Phenylalanine	16
	106	Asparagine	45
	113	Methionine	39
	115	Threonine	37
M. eugenii	54	Serine	31
	55	Phenylalanine	16
	106	Asparagine	45
	113	Methionine	39
	115	Threonine	37

4.4.4.8 CTLA-4 - Structure modelling

Homology protein modelling was undertaken with the putative amino acid sequences of *M. eugenii* and *O. fraenata* CTLA-4. In the case of *O. fraenata*, the model prediction carried out with RaptorX was based on the 1hzh template from the pdb database, while the template for *M. eugenii* was 3osk. The quaternary structure of 1hzh is annotated as a hetero-tetramer but the final model for the *O. fraenata* CTLA-4 molecule was calculated as a single chain (Fig. 4.21). The model produced for *M. eugenii* with I-TASSER was also a single chain, but the Modweb program modelled the *M. eugenii* CTLA-4 as a dimer with high probability (Fig. 4.22).

There are significant differences in the models of the *O. fraenata* and *M. eugenii* CTLA-4 molecules produced by RaptorX. It is clearly shown that some coils are in different positions and the beta sheets are at different locations. This indicates that although the tertiary structure was predicted as similar the homology modelling program was able to differentiate the difference in the topography of the CTLA-4 molecule in *O. fraenata* and *M. eugenii* (Fig. 4.21).



Homology model of *O. fraenata* CTLA-4 Homology model of *M. eugenii* CTLA-4

Figure 4.21. RaptorX homology model of CTLA-4. (a) *O. fraenata.* (b) *M. eugenii*. Differences are indicated with red circles.

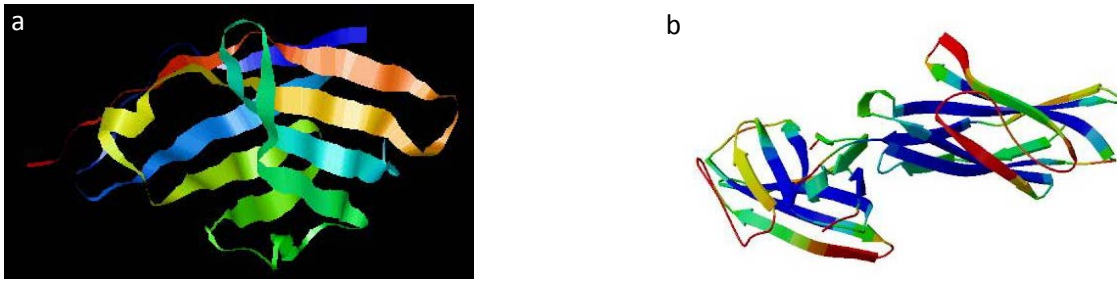


Figure 4.22. CTLA-4 homology models for *O. fraenata* and *M. eugenii*. (a) I-TASSER modelled *O. fraenata* as a single chain. (b) Modweb modelled *M. eugenii* as a dimer.

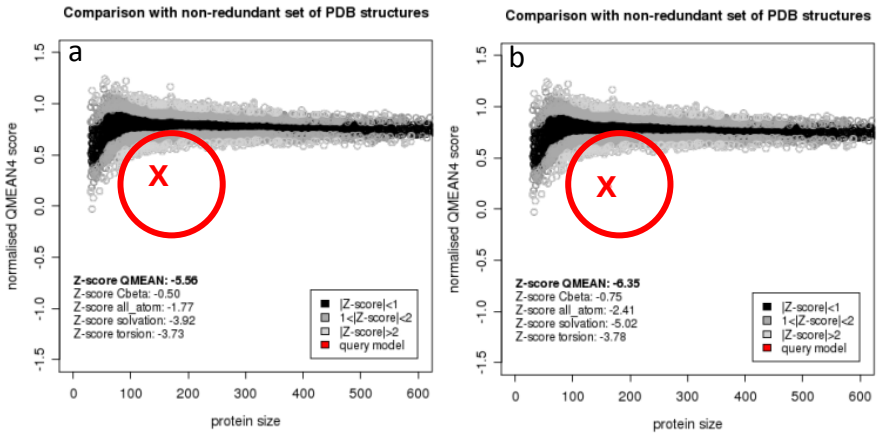


Figure 4.23. Homology models of *O. fraenata* and *M. eugenii* CTLA-4 compared to other mammalian structures in the pdb solved by X-ray crystallography. (a) X = *O. fraenata* structure (b) X = *M. eugenii* structure.

All statistical values, including their Z-scores, were measured against scores obtained from high-resolution experimental structures of similar size by the SWISS-MODEL program (Table 4.24). Only the mature protein was used in the modelling procedure.

Table 4.24. Statistical values of CTLA-4 model predictions by the Swiss-Model program for *O. fraenata* and *M. eugenii* showing the Z-scores and raw scores against the prediction markers.

Statistics for CTLA-4 model of *O. fraenata* and *M. eugenii*				
	O. fraenata		*M. eugenii*	
Prediction markers	Raw score	Z score	Raw score	Z-score
C_beta interaction energy	-48.36	-0.50	-124.53	-0.85
All-atom pairwise energy	-1677.01	-1.77	-5416.92	-2.41
Solvation energy	11.29	-3.92	-19.09	-5.02
Torsion angle	6.35	-3.73	-60.59	-3.78
QMEAN4 score	0.292	-5.56	0.729	-6.35

4.4.4.9 CTLA-4 - Phylogenetic analysis

The evolutionary history of CTLA-4 was inferred using a Neighbor-Joining tree. The optimal tree has a branch length sum of 2.02. The percentages of replicate trees in which the associated taxa are clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The predicted *M. domestica* sequence was used in this

analysis and it was shown that the marsupial clade form a distinct group separated from the monotreme *O. anatinus* and mammals such as *L. africana*, *C. familiaris*, *F. catus* and *S. scrofa*. The branch length varies between the marsupials species. The *O. fraenata* and *M. eugenii* branch lengths are shorter than the branch length of *M. domestica* indicating that the macropods are further removed from the didelphid. Only a minor distance separated the two macropods *O. fraenata* and *M. eugenii* indicating a close evolutionary relationship. The analysis involved 32 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 158 positions in the final dataset (Fig. 4.24).

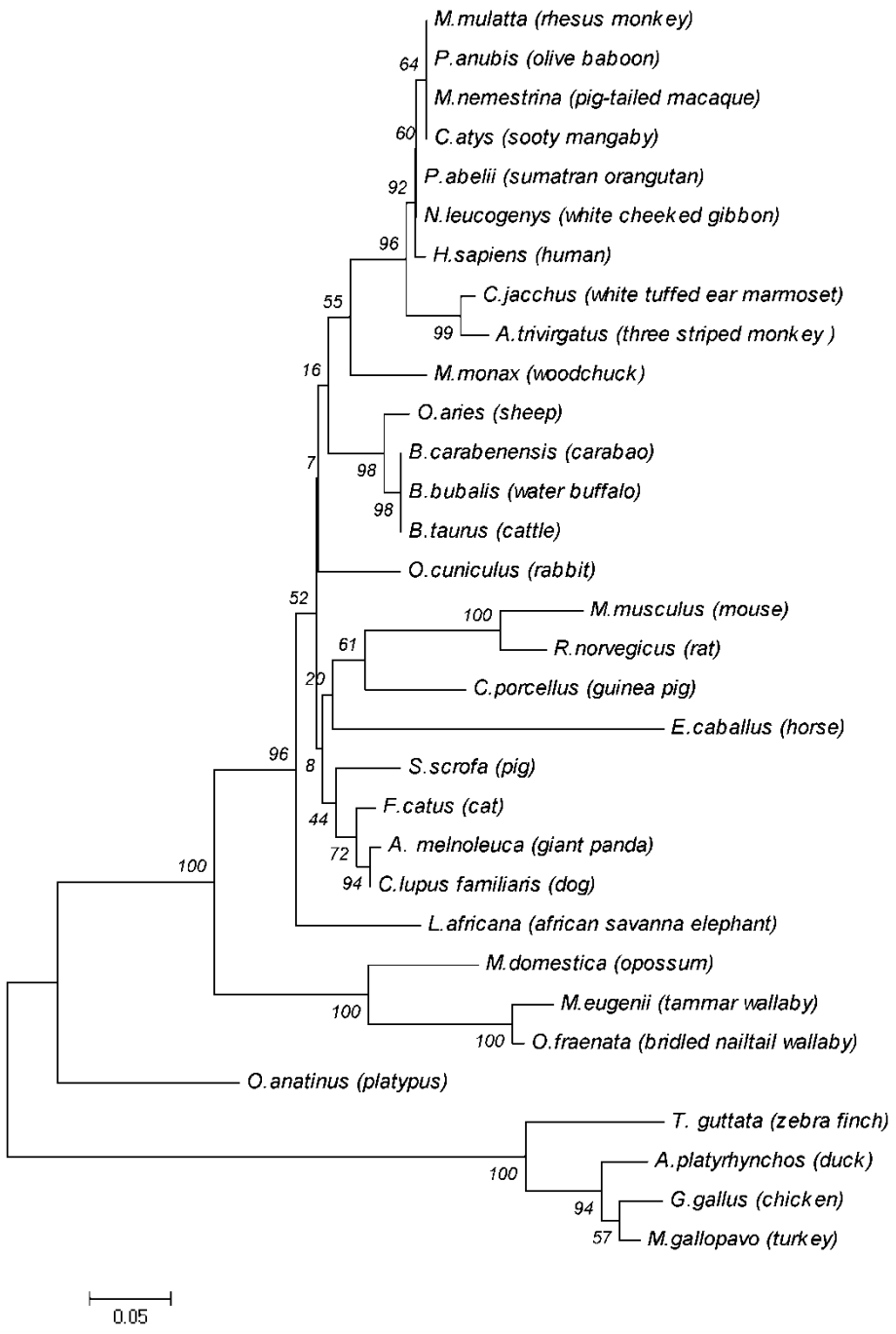


Figure 4.24. Phylogenetic Neighbor-Joining tree for known CTLA-4 sequences. Branch lengths indicate the evolutionary relationship between the taxa.

4.4.6 CD86

A partial sequence of CD86 was characterized only in the didelphid *M. domestica*.

4.4.6.1 CD86 - Homology

A homology search carried out with the BLAST algorithm indicated that the *M. domestica* CD86 partial nucleotide sequence of 381bp which translated into 127 amino acids had an identity of 74% to *H. sapiens* at the nucleotide level and 45% identity at the amino acid level. Identity percentages to other mammals are shown in Table 4.25.

Table 4.25. Homology search results for the partial *M. domestica* CD86 nucleotide and amino acid sequence, its identity and respective e-values. The values are calculated over the 391bp/127aa length of the sequence.

CD86 homology search results for *M. domestica*				
Species	Nucleotide	e-value	Amino acid	e-value
B. taurus	78%	2e-07	48%	5e-36
C. jacchus	76%	0.17	43%	2e-32
H. sapiens	74%	0.58	45%	5e-35
S. scrofa	73%	0.048	52%	3e-37
L. Africana	73%	0.17	46%	1e-30
A. melanoleuca	71%	3e-05	47%	9e-34
F. catus	71%	0.001	47%	8e-33
M. monax	63%	0.048	53%	5e-38

Accession numbers are in Appendix 4D.

4.4.6.2 CD86 - Domain structure

The domain architecture investigated with the SMART program indicated a putative Ig-like domain from positions 2 to 79 with an e-value of 5.02e-02 in the partial sequence. A putative IGc2 domain was also predicted with an e-value of 5.31e+01.

4.4.6.3 CD86 – Glycosylation sites

Three glycosylation sites were predicted in the *M. domestica* CD86 partial sequence. The glycosylation sites were predicted at positions 51, 56 and 113 with confidence levels of 79%, 53% and 61% respectively. A fourth putative glycosylation site was predicted at position 123 but the confidence level for this site was only 36% and it was therefore disregarded.

4.4.6.4 CD86 - Phosphorylation sites

Two serine and 2 tyrosine phosphorylation sites were predicted in the *M. domestica* partial CD86 amino acid sequence. The putatively phosphorylated serine sites were found at position 16 with a 94% probability and at position 83 with an 83% probability. Two tyrosine sites were predicted at position 49 with a probability of 67% and at position 55 with a 63% probability.

A PKA binding site was predicted at position 94 with a 66% probability.

4.4.6.5 CD86 - Primary sequence and secondary structure prediction

The CD86 molecule is the ligand for cell surface proteins CD28 and CTLA-4. This molecule was found to be composed of a membrane-distal receptor binding domain and a membrane-proximal constant immunoglobulin domain.

The secondary structure prediction was carried out to elucidate the distribution of alpha helices and beta strands within CD86. Eleven beta strands were found in the *M. domestica* CD86 partial sequence and no transmembrane helices were detected in this partial sequence (Fig. 4.26).

Figure 4.25. *M. domestica* CD86 partial sequence and secondary structure prediction. NRT = N-linked glycosylation sites. Y = tyrosine phosphorylation sites. ▶◀ = Ig domain. ▶◀ = IGc2 domain. S = serine phosphorylation sites. . = PKA binding site. → = Strand.

4.5 Discussion

The two co-receptors to the T cell receptor, CD4 and CD8 have been characterized in *M. eugenii* and *M. domestica* by (Duncan *et al*., 2008, Duncan *et al*., 2007). In the present study a further characterization of this molecule was conducted in *O. fraenata* that allowed for a comparison of structurally important motifs between *O. fraenata* and its close relative *M. eugenii*. The *O. fraenata* CD4 sequence contained a sequence gap in domain 4 which was found to be absent in *M. eugenii* and eutherian mammals suggesting that this CD4 could be a soluble form of the receptor. It has been reported that a soluble form of the glycoprotein CD4 which lacks the transmembrane and cytoplasmic domains does exist but is not naturally occurring (Deen *et al.,* 1988). It is usually manufactured by transfection of mammalian or insect cells with vectors encoding forms of the CD4 gene that lack the two domains (Deen et al., 1988, Fisher et al., 1988, Smith et al., 1987). In the present study it was found that the *O. fraenata* CD4 sequence continued after a premature stop codon in a different reading frame encoding the transmembrane and cytoplasmic domains. This suggested that the *O. fraenata* CD4 molecule may have different fold properties to that of other marsupials and mammals. Different transcripts of the CD4 gene can be found in the ensembl database (http.//www.ensembl.org) which suggests that variations in the CD4 molecule are not rare occurrences. Splenic mouse macrophages display a CD4 marker on their cell surface that is truncated in the 3' end of the coding domain (Moore *et al*., 1992). This is similar to the truncation observed in *O. fraenata* and supports the finding in the present study. It will be important to investigate the frame shift found in three clones of the *O. fraenata* CD4 molecule further in order to determine if this sequence translates into a functional protein and if it may represent a splice variant of the more common CD4 gene.

CD8 has also been characterized in *M. eugenii* and *M. domestica* by (Duncan et al., 2008). In the present study this molecule was investigated in *O. fraenata* and *L. hirsutus*, two other macropod species which appeared to have opposing immunological competency with regard to mycobacterial infections. Like other marsupials, *L. hirsutus* is reported to

readily succumb to mycobacterial infections (Montali et al., 1998, Gaynor et al., 1990, Peet et al., 1982) while anecdotal evidence from animal handlers suggests that *O. fraenata* is less susceptible to such infections. The literature is silent as to possible reasons for the differences in this macropod's response to mycobacterial infections. It is noted that in humans, $CD8^+$ T lymphocytes have been implicated in the control of intracellular mycobacterial infections (Lewinsohn *et al.,* 2003, Smith and Dockrell, 2000). In the present study, characterization of the CD8α and CD8β chains in *O. fraenata* and *L. hirsutus* was conducted to detect differences in the sequences of the two molecules that may explain the different responses to mycobacterial infection in those two species.

It was found in this study that the number of putative O-linked glycosylation sites was the most obvious difference between the marsupial CD8α hinge region and the human hinge region. Five O-linked glycosylated threonine residues were found in the human CD8α sequence compared to eight such sites in *O. fraenata* and seven in *L. hirsutus*. A sequence alignment revealed that the eight O-linked glycosylation sites in *O. fraenata* is the highest number of such sites found in any vertebrate. O-linked glycosylation sites are comprised of negatively charged sialic acids (Devine *et al.,* 1999). These negatively charged residues add mass and interact with positively charged amino acid residues on the lateral surface of the variable domain of the TCRα chain. These glycosylation sites are also involved in the dimerization of CD8α with CD8β (Fares, 2006). The difference in the number of the O-linked glycosylation sites may also be associated with a change in structure and may either inhibit or favour the dimerization process. In any case, this post-translational modification may change the folding properties of the marsupial CD8α chain. Another post-translational modification is the N-linked glycosylation site. A sequence alignment revealed that all mammals except *O. fraenata, M. eugenii, L. hirsutus* and *M. domestica* have an N-linked glycosylation site. N-linked glycosylation alters proteolytic resistance, changes protein solubility and stability, changes local structure and immunogenicity (Lis and Sharon, 1993), and influences protein folding (Kim *et al.,* 2009). In summary, a higher number of O-linked glycosylation sites, as found in *O. fraenata* and *L. hirsutus,* may

indicate a better ability to modulate the biological activity of the molecule. However functional studies, which were outside the scope of this project, are needed to confirm this assumption. The predicted folding pattern of the marsupial CD8α was shown to have good QMEAN scores. This indicates that the prediction was of high quality and that the structure of the marsupial molecules compared well with those of other CD8α protein structures found in the protein database. Consequently, the absence of N-linked glycosylation may not have affected the structure of the marsupial CD8α chain.

In this study it was found that the marsupial CD8α molecule binds a number of external ligands including a zinc ion at distinct residues. In humans, this event transmits a signal inside the cell which facilitates the interaction with signalling proteins downstream. The CXC motif is part of a zinc finger domain where one cysteine in the cytoplasmic domain and a corresponding cysteine in the N-terminal end of the $p56^{lck}$ molecule interact with a single zinc atom (Moore *et al.,* 2005, Bosselut *et al.,* 1999). This motif was identified in both *O. fraenata* and *L. hirsutus* CD8α sequences and is conserved. Zinc ions emulate the action of cytokines and are essential for the biological activity of transcription factors (Beyersmann and Haase, 2001). This suggested that the marsupial CD8α molecule may transfer a similar biological activity to its transcription factors. However, this assumption is based on high sequence conservation and a high QMEAN score for the structure. Further investigation is therefore warranted to confirm the biological activity of the transcription factors.

In most mammals, the CDR loops of CD8α are the main regions that interact with the MHC class I molecule and are important for the biological function of CD8α (Li *et al.,* 1998). In human CD8α, L^{73}, N^{75} and T^{76} which are located on the tip of the β turn, form the bioactive core of the DE loop (Li *et al.,* 1998). By comparing the marsupial DE loop sequences, as determined in this study, with other mammalian DE loop sequences it was apparent that the sequences of the marsupial DE loops are unique. It has been reported that, apart from its role in CD8α bioactivity, the DE loop is involved in graft versus host responses (Li *et al.,* 1998).

The apparent differences in the DE loop together with the absence of the N-linked glycosylation sites found in this study may affect the bioactivity of the marsupial CD8α molecule. However, the conservation of the p^{56}lck binding motif and the high homology of the structure prediction suggests that the marsupial CD8α molecule is not dissimilar to that of other mammalian CD8α molecules.

The CD8α molecule dimerizes with the β chain of the CD8 co-receptor. It was found in this study that the mature CD8β protein in marsupials contained a V-like domain followed by a well conserved J-like region. The J-like region is typical of the CD8β chain and is not found in the CD8α chain (Chida *et al.,* 2011). The FGXG motif (F = phenylalanine, G = glycine and X = any amino acid) in the β-G strand is found in the V region of CD8β where it creates a 'β-bulge' which is necessary for VH/VL dimerization (Chothia *et al.,* 1985). It was found in the present study that the FGXG motif is conserved in the *O. fraenata* and *L. hirsutus* CD8β chains but not in *M. domestica*. It appeared that the CD8β chains in *O. fraenata* and *L. hirsutus*, as in humans, will also dimerize with the CD8α chain, however the implications of this for the didelphid *M. domestica* are not known. It was found that in *M. domestica*, the phenylalanine residue (F) is substituted with a serine (S). This is unusual since it has been reported that phenylalanine is usually only substituted with other aromatic or hydrophobic amino acids (Betts and Russell, 2003). Phenylalanine is a non-reactive amino acid and therefore is rarely involved in protein function but plays a role in substrate recognition of SH_3 and WW domains as part of a polyproline binding site (Macias *et al.,* 2002). Serine can be located either within a protein or on its surface. The substitution of phenylalanine with serine in *M. domestica* appears to be benign and no structural hindrance was expected in its CD8β chain.

It was found that the hinge region containing the glycosylated threonine sites followed the FGXG motif in *O. fraenata* and *L. hirsutus*. As in the CD8α chain, the number of O-linked glycosylation sites in the CD8β chain differed between marsupials and other mammals. *O. fraenata* and *L. hirsutus* had five O-linked glycosylation sites while *M. domestica* had seven such sites. This contrasts with the four glycosylated residues in

humans. O-linked glycosylation occurs in post-translational modifications which are initiated in the Golgi-apparatus and takes place instead of tyrosine phosphorylation, thus affecting the signalling process (Van den Steen *et al.,* 1998). As previously described, the marsupial alpha chain of the CD8 molecule also has a different number of O-linked glycosylation sites. The effects of those differences in the CD8α and CD8β chains could influence the functionality of the marsupial CD8 molecule. Neither the alpha chain nor the β chain of the CD8 molecule had any N- linked glycosylation sites. This is in contrast to *M. musculus* (mouse), *Felis catus* (cat), *Ailuropoda melanoleuca* (giant panda) and the primates which all have a single N-linked glycosylation site in the CDR2 region of the CD8β chain.

The connecting peptide (CP) region of CD8β in *O. fraenata*, *L. hirsutus*, *M. eugenii* and *M. domestica* is a lysine rich region and could be a suitable site for proteolytic cleavage. The connecting peptide core in the human CD8β sequence is less pliable due to the proline residues found in this domain, and has an extended conformation which allows contact with MHC class I molecules (Fellah *et al.,* 2002). The present study found a similar number of proline residues in the *O. fraenata* and *L. hirsutus* CD8β chain and it is therefore expected that the function of this motif in those species is the same as in humans.

The cytoplasmic tails (CT) of both the CD8α and β chains are important in signal transduction due to the location of the lck binding site within that region. The marsupial CT domain in CD8α is 30 amino acids long which is the same length as in humans. However the marsupial CT domain in the CD8β chain is 14 amino acid residues long compared to 19 residues in humans. It has been found that, in humans, the CT regions play an important role for the activation of $CD8^+$ T lymphocytes by binding the signalling molecule $p56^{lck}$ via a zinc ion complex (Arcaro *et al.,* 2001, Bosselut *et al.,* 2000, Zamoyska, 1994). The lck binding site is conserved in the marsupial CD8β chain. Although its effect on the signalling capacity of the marsupial CD8α and β CT regions is not known, a ligand binding prediction for the marsupial molecules indicated the presence of a zinc binding site which activates signal transduction.

Differences in the number of O- and N-linked glycosylation sites and the lengths of the cytoplasmic tails were observed in the marsupial CD8β chain compared to the CD8α chain. However, the conservation of important structural motifs in the marsupial CD8 molecule suggests a similar function to that of other mammalian CD8 molecules.

In contrast to the CD4 and CD8 receptors, the co-stimulators responsible for T cell signalling and activation had not been characterized in any marsupial prior to this study. CD28 and CTLA-4 were characterized in *O. fraenata*, *M. eugenii, L. hirsutus* and *M. domestica* for the first time in this study to support the hypothesis that all the key elements of a functional T cell signalling cascade are expressed in marsupials. The TCR co-stimulators CD28 and CTLA-4 are members of the immunoglobulin superfamily and are glycoproteins expressed on mature T lymphocytes (Hansen *et al.,* 1980). The ensembl database contains an annotation for both CD28 and CTLA-4 in *M. eugenii*. However, when the expressed *M. eugenii* CD28 sequence was compared to its annotated sequence it was found that the annotated sequence was a partial sequence only (supporting data are Appendix 4C). The annotation of the CTLA-4 sequence in *M. eugenii* was confirmed by the expressed sequence of the gene.

In this study it was found that the hexapeptide MYPPPY, which is essential for the interactions between CD80, CD86 and CTLA-4, was conserved in the *M. eugenii* CD28 sequence. In humans and other mammals this motif is located on the tips of a "Y" (Margulies, 2003) and a sequence alignment showed that the *M. eugenii* motif is located in the same position. In this study, the binding motif (YMNMTPR) for the p85 subunit of phosphoinositide 3-kinase (PI 3-kinase) which delivers the second signal that regulates cell growth by interaction with the cytoplasmic tail of CD28 or CTLA-4 (Pages *et al.,* 1994, Prasad *et al.,* 1994) was found in the *M. eugenii* CD28 sequence. The YMNMTPR motif in the marsupial CD28 sequence is conserved when compared with that of other mammals. However the expressed partial sequence of *M. domestica* differed in the last amino acid of this motif where the arginine (R) was replaced with a glutamate (Q).

The marsupial CD28 sequence contained 8 cysteine residues, 5 of which were located in the IgV-like region. This is in contrast to 5 cysteine residues in total found in other mammals (supporting data are in Appendix 4C). An extracellular interchain disulphide bond, residing in the linker region connecting the IgV domain with the transmembrane domain (Lazar-Molnar *et al.,* 2006), is responsible for the heterodimerization of the CD28 molecule in humans (June *et al.,* 1990). Although the unpaired cysteine residue responsible for the dimerization (Lenschow *et al.,* 1996) is conserved in the *M. eugenii* sequence, the marsupial CD28 molecule was modelled as a monomer.

The secondary structure of the *M. eugenii* CD28 appeared to be similar to that of the human secondary structure, but there was a notable difference in the tertiary structure prediction. The beta strands and the helices had a similar distribution in the *M. eugenii* CD28 when compared to the human CD28 suggesting a similarity in structure. However tertiary homology modelling did not produce similar structures despite the modelling statistics indicating a good match to other pdb structures.

The sequence homology between human CD28 and CTLA-4 is 20%, while the marsupial sequence homology between those two molecules is 28%. This suggests that the gene duplication theory (Balzano *et al*., 1992) may also apply to the marsupials CTLA-4 and CD28 sequences.

It has been reported that the human CTLA-4 sequence has a short cytoplasmic tail with two potential src homology (SH_2) domain binding sites centered at Y^{165} and Y^{182}, spaced by a proline-rich stretch (Baroja *et al.,* 2000). In this study it was confirmed that those two residues are conserved in *O. fraenata* and *M. eugenii* (supporting data in Appendix 4D). Human Y^{165} is involved in the expression of CTLA-4 on the cell surface (Baroja *et al.,* 2000) and the conservation of this residue in the two marsupial species suggested a similar function.

Once T cell activation occurs, an increase in the levels of CTLA-4 mRNA and protein takes place. Baroja *et al.* (2000) reported that the expression of CTLA-4 occurred only in

stimulated tissue. However, in this study it was demonstrated that the molecule was expressed in stimulated lymphocytes in *M. eugenii*, and in un-stimulated healthy thymus and spleen tissue of *O. fraenata*.

This study showed that the intracellular domain of the marsupial CTLA-4 contained the YVKM motif and that it was highly conserved in *M. eugenii* and *O. fraenata* (supporting data are in Appendix 4D). According to Qureshi (2012) the YVKM motif controls localization and trafficking of the CTLA-4 molecule. In the human CTLA-4 sequence, the tyrosine residue contained within that motif binds adaptor proteins 1 and 2 (AP-1 and AP-2) in its un-phosphorylated state and, since it is conserved in the two marsupial species, this binding may also occur in them.

The predicted tertiary structure of CTLA-4 and the residues that have the ability to bind ligands appeared to be identical in both *M. eugenii* and *O. fraenata*. As expected, the *O. fraenata* CTLA-4 could only be modelled as a single chain which is the resolved structure of this molecule. However, homology modelling showed that *M. eugenii* CTLA-4 could only be successfully modelled as a dimer (Fig. 4.19). This structural difference may be attributable to the different forms of CTLA-4. The membrane-bound isoform functions as a homodimer while the soluble isoform has been shown to function as a monomer (Oaks and Hallett, 2000). It may be possible that one of the marsupial molecules is a different isoform of the CTLA-4 molecule, however this is an assumption which will have to be substantiated by functional studies which were outside the scope of this study. It has also been documented that the soluble form of CTLA-4 is associated with autoimmune diseases in humans and inhibits the mixed lymphocyte reaction (MLR) (Berry *et al.,* 2008, Oaks and Hallett, 2000). Stone *et al*. (1996) reported that marsupials have no mixed lymphocyte reaction.

Since the soluble form of the CTLA-4 molecule is responsible for inhibition of the MLR it is possible that the monomeric form found in marsupials is the soluble form of the receptor which was indicated by the structure prediction in *O. fraenata*. However, further

investigations need to be conducted to confirm the existence of two different isoforms of this molecule in marsupials.

Since homology modelling showed that the *O. fraenata* CTLA-4 is a single chain structure, and since this is one of the resolved states of that molecule, the single chain structure may explain the reported resistance of *O. fraenata* to mycobacterial infections as discussed earlier. *M. eugenii* is reported to be more susceptible to such infections, and homology modelling of its CTLA-4 molecule in this study showed that it was a dimer. *O. fraenata* and *M. eugenii* are the only marsupial species for which CTLA-4 has been fully characterized. If the CTLA-4 molecules in other marsupial species are shown to be dimers then the resistance displayed by *O. fraenata* may well be attributed to that single chain structure.

Human studies have established that the interaction of CTLA-4 and CD28 with their natural ligands CD80 and CD86, together with TCR signalling, enhances the production of interleukin-2, thereby completing the T cell signalling cascade (Bugeon and Dallman, 2000). In this study, the amplification of a partial sequence of the *M. domestica* CD86 molecule showed conclusively that CD86 is expressed in marsupials, and demonstrated that the communication link between T lymphocytes and antigen presenting cells in marsupials is similar to that in other mammals.

4.6 Conclusion

This study showed that all of the functional and structural motifs discussed in this chapter are conserved in the expressed genes which were investigated in marsupials. Identifying structural motifs from sequence information presents challenges since many different amino acid sequences are compatible with the same secondary structure. By making comparisons with other secondary and tertiary structures that have been resolved by X-Ray crystallography, and by reference to the literature, it was possible to annotate most of the functional motifs contained within the sequences discussed in this chapter. By using Modeller 9.10, a program that allows manipulation of the sequences by the user, it

was possible to build a structure from the putative amino acid sequence and to compare it with homologous structures deposited in the protein database (pdb). Inspection of the structures obtained from the modelling process found that all the genes investigated in this chapter had homologous counterparts. It also appeared that the marsupial structures were very similar to other mammalian structures and, consequently, it is concluded that the functionality of the genes investigated in marsupials in this chapter is not dissimilar to that of other mammals.

One exception was the predicted CTLA-4 structure in *O. fraenata*. While this molecule was modelled as a dimer in *M. eugenii,* the *O. fraenata* CTLA-4 could only be modelled as a monomer. The monomeric form of this molecule has been linked to the soluble form of CTLA-4 and is implicated in either disease resistance or susceptibility depending on its expression kinetics (Toussirot *et al*., 2009). It may therefore be possible that this monomeric structure may contribute to the reported resistance of *O. fraenata* to mycobacterial infections as discussed in this chapter. It may also be possible that the monomeric structure has implication for the diminished mixed lymphocyte reaction reported in marsupials (Stone *et al*., 1996).

The characterization of the CD28 molecule in *M. eugenii,* and the subsequent comparison with the annotated sequence in the ensembl database, revealed shortcomings in the annotation and demonstrated that the annotated sequence is incorrect when compared with the expressed sequence.

Chapter 5

Signalling molecules – TCRζ, ZAP-70 and Lck

Chapter 5 – Signalling molecules – TCRζ, ZAP-70 and Lck

5.0 Abstract

The signalling molecules, T cell receptor zeta chain (TCRζ), zeta associated protein of 70kDa (ZAP-70) and the lymphocyte specific kinase (Lck) have been characterized in many mammals, but no characterization of these molecules has taken place for any marsupial prior to the present study.

The sequences of these molecules were obtained using RT-PCR and RACE PCR strategies in *M. eugenii*, *O. fraenata* and *M. domestica* either in full or as fragments. The sequences were investigated for important structural motifs that show similarity to the known functional regions found in human and other vertebrates. Where possible, a comparison with the published wallaby genome was carried out with the result that sequence gaps were identified in the genome, which in turn identified further areas of research. Structure modelling was conducted for all molecules where the open reading frame was obtained. For the molecule ZAP-70 in particular, a customized python shell was written in order to obtain the most probable structure. The main findings of this chapter were the discovery of sequence differences in the transmembrane domain of the marsupial TCRζ chain and the catalytic loop of the marsupial ZAP-70 molecule. The sequence of Lck was found to be highly conserved between marsupials and eutherian mammals. Of interest was the absence of a signal peptide in the *M. eugenii* TCRζ chain. In contrast, the *O. fraenata* TCRζ chain showed a predicted N-terminal signal peptide as is observed in all other mammals. A human anti-ZAP-70 and anti-Lck antibodies were trialled for cross reactivity in *M. eugenii* tissue but both antibodies failed to recognize the marsupial ZAP-70 and Lck proteins.

5.1 Introduction

Signalling molecules are characterized by the presence of SH_2 and SH_3 domains. The SH_2 domains transmit intracellular signals by mediating protein-protein interactions and exert their effects by recognizing phosphotyrosine residues. When the TCR engages an antigen

receptor, protein tyrosine kinases dimerize and trans-phosphorylate each other. The phosphotyrosine sites thus created recruit SH_2–containing proteins that in turn mediate downstream signal transduction (Yu and Schreiber, 1994).

CD3ε, together with the T cell receptor zeta (TCRζ) are the molecules that ready the cell for signal transduction. TCRζ is required for the expression of the TCR/CD3 complex as outlined in Chapter 1 (section 1.4.1.2.5.1) and in the present study this molecule was identified in the macropods *O. fraenata* and *M. eugenii,* and in the didelphid *M. domestica*. The activation of T lymphocytes through the antigen receptor triggers a number of intracellular signalling events which result in the transcription of the IL-2 gene in the nucleus of T lymphocytes (Okoye *et al.,* 2007). This has not been described in marsupials prior to this study.

The TCRζ chain is the molecule that initiates the signalling cascade through a unique motif called the immunoreceptor tyrosine activation motif (ITAM) located in the cytoplasmic tail of the molecule. The ITAMs interact with the tandem SH_2 domains of the zeta associated protein of 70 kDa (ZAP-70) which phosphorylates the transmembrane protein Linker of Activated T lymphocytes (LAT) (Deindl *et al.,* 2007). ZAP-70 has been identified as an indispensable link in directing the signalling cascade towards the nucleus and affecting the gene transcription of the cytokine IL-2 and therefore T cell activation (Blanchard *et al.,* 2002). The molecular identification of ZAP-70 in the marsupials in the present study provides an important step in the elucidation of the biochemical T cell signalling cascade.

ZAP-70 promotes the phosphorylation of TCRζ and the association of TCRζ with Lck (Ashe *et al.,* 1999). In double positive thymocytes ($CD4^+/CD8^+$) the phosphorylation of tyrosines in the ITAM motifs by Lck is dependent on the presence of the ZAP-70 protein but independent of the kinase activity of ZAP-70 (Ashe *et al.,* 1999). Lck controls T cell activation by amplifying the signal through the CD4 co-receptor and enhances the activation of ZAP-70 by phosphorylating distinct residues in the interdomain B and the activation loop of ZAP-70 (Dong *et al.,* 2010). In order to show that the marsupial T cell signalling cascade involves key molecules it was necessary to show the expression of Lck,

a non-receptor tyrosine kinase, and identify the important domains that include SH_2, SH_3 and the tyrosine kinase domains. The Lck transduced signal is required for T lymphocyte development and for antigen-dependent activation of mature T lymphocytes. Lck is known to drive T lymphocytes to reach the $CD4^+/CD8^+$ double positive stage (Molina *et al.,* 1992). It is also required for efficient antigen-induced T cell activation (Glaichenhaus *et al.,* 1991). Due to its association with the short cytoplasmic tails of CD4 and CD8, Lck phosphorylates the TCR-ζ chain (Iwashima *et al.,* 1994). This interaction is an important step in the T cell signalling cascade. Identification of the residues in Lck that interact with CD4, CD8 and TCRζ in marsupials demonstrated the sophistication of the marsupial adaptive immune system.

5.2 Aims and Objectives

The aims of the experiments reported in this chapter were to characterize the signalling molecules TCRζ, ZAP-70 and Lck in *O. fraenata, M. eugenii* and *M. domestica* and in particular:

- To identify structural motifs such as post-translational modifications and to model a realistic structure from the putative amino acid sequences of TCRζ, ZAP-70 and Lck.
- To determine whether or not a Western Blot procedure is able to identify the marsupial proteins ZAP-70 and Lck in a crude cell lysate using human specific antibodies for the first time.

5.3 Specific Materials and Methods

The following materials and methods were unique to this chapter. General materials and methods are detailed in Chapter 2.

In this chapter the TCRζ and ZAP-70 molecules were characterized in *O. fraenata, M. eugenii* and *M. domestica*. The Lck molecule was characterized in *O. fraenata* and *M. eugenii* only.

5.3.1 RNA, mRNA and cDNA

RNA was isolated from spleen, liver, lung, gut node and thymus tissues of *O. fraenata* as described in Chapter 2, section 2.1.3. mRNA was isolated from total RNA, and cDNA was synthesized as described in Chapter 2, sections 2.1.3.2, and 2.1.5. RNA, mRNA and cDNA were quantified as described in Chapter 2, section 2.1.4.

The TCRζ and Lck sequences in *O. fraenata* were obtained from spleen tissue, while ZAP-70 was amplified from thymus tissue. The sequences for *M. eugenii* originated from stimulated lymphocytes donated by Dr. L. J. Young. All such sequences were cloned as outlined in Chapter 2, section 2.1.10. Three clones were prepared for sequencing in the forward and reverse direction as described in Chapter 2, section 2.1.10.6.

5.3.2 Primer design

5.3.2.1 TCRζ, ZAP-70 and Lck

Primers were designed as detailed in Chapter 2, section 2.1.2. The primer sequences used and their respective melt temperatures for TCRζ are presented in Table 5.1.

Table 5.1. Primer sequences for TCRζ used to elucidate the sequence of the TCRζ gene in *O.fraenata, M. eugenii* and *M. domestica*.

Primer	Species specificity	Source tissues	Primer Sequence	T_m
CD3zFc	*M. domestica*	spleen, thymus	ctcttcmtmtaygghgtcatyvtcacngc	60.8°C
CD3zRc		cDNA library	actccattacagtcttgacagatggcaaaac	60.3°C
TCR2_F1	*M. domestica,*		atgcaattcctttccacagaggcccag	61°C
TCR2_R1	*O. fraenata,*	spleen	ttaacggggaggcaggggctg	60°C
	M. eugenii	PHA stim.ϕ		
B247F	*O. fraenata*	spleen	(3' RACE primer). gaggcaaaggaaatgatgtcctgtac	58°C
B247R			5' RACE primer). gagttgattctggtcctgttggtag	58°C
TRR	*O. fraenata*	spleen	5' RACE primer).	
	M. eugenii	PHA stim.ϕ	cttctctgttttcctcccatctctg	56.9°C
TRF			3' RACE primer). cagagatgggaggaaaacagagaag	56.9°C
TCRzSTART	*O. fraenata*	spleen	atgaagtggaaggggattgttatc	54.8°C
TCRzSTOP	*M. eugenii*	PHA stim.ϕ	catgcagcccctgcctccccgttaa	67.1°C
TTCRzF	*O. fraenata*	spleen	cagagatgggaggaaaacagagaag	58°C
TTCRzR	*M. eugenii*	PHA stim.ϕ	cttctctgttttcctcccatctctg	58°C
TCRzSTOP	*O. fraenata*	spleen	gttgttcttgttaacggggaggcag	59°C
	M. eugenii	PHA stim.ϕ		
TCRzexF	Expression study primers	spleen	cctcttcatttatggagtcatcatcac	57°C
TCRzexR	for *O. fraenata*		cggtttctgtaggattcttccttctc	58°C

Primers for ZAP-70 are presented in Table 5.2 together with their respective melt temperatures, RACE primers and expression primers.

Table 5.2. Primer sequences for ZAP-70 used to elucidate the sequence of the ZAP-70 gene in *O. fraenata, M. eugenii* and *M. domestica*.

Primer	Species specificity	Source tissues	Primer Sequence	T_m
ZAPFc ZAPRc	*M. eugenii*	PHA stim.ϕ	gccaggcmcctcaggtggagaagctyattgc gcttcccacatggtgactccrtagctcc	69°C 65°C
OZF_2 OZ-R1	*M. domestica*	thymus cDNA library	atgccvgayccgcggcgcacctg tcaggcactggccacctcttgtgtttgag	68°C 64°C
70TSQ_1	*O. fraenata*	thymus	5'RACE primer. cgtagctgccttgttccttcctg	59°C
70TSQ_F	*M. eugenii*	PHA stim.ϕ	3'RACE primer. gcctctgcacaagtttctggccgc	62.9°C
ZAPSTART ZAPSTOP	*M. eugenii*	PHA stim.ϕ	atgccagatgcagctgcccatttgc tcaggcaggagtagccccctctgattg	61°C 64°C
RBZAPF-1	*O. fraenata*	thymus	3' RACE primer. catatgccagtgcttctactgctac	58°C
RBZAPR-1	*O. fraenata*	thymus	5'RACE primer. caggggtgtatccatcagagttgagggtgtc	66°C
RBZAPF-2	*O. fraenata*	thymus	3'RACE primer. cgaaatgttcttctggtcaaccagcac	60°C
RBZAPR-2	*O. fraenata*	thymus	5'RACE primer. gtcctggttgatgaggtagtggtagacag	63°C
ST3F1 [(1)]	*O. fraenata and M. eugenii*	thymus PHA stim.ϕ	3'RACE primer. gtgacagggatctggctgcaagaaatg	63.5°C
BZAPexF BZAPexR	Expression studies in *O. fraenata*	thymus	cctgagggcacaaagtttgacacc cagggagtgtgggagcagcagtcac	59°C 64°C

[(1)] designed by A. Suthers.

Primers for the lymphocyte specific kinase (Lck) are presented in Table 5.3 together with their respective melt temperatures (T_m).

Table 5.3. Primer sequences for Lck used to amplify the Lck gene in both *O. fraenata* and *M. eugenii*.

Primer	Species specificity	Source tissue	Primer Sequence	T_m
LckF LckR	*O. fraenata*	spleen	gactggatggaaaayatygacgtgtg gccctcygtggchgtgaagaagtc	57.9°C 63.4°C
TRLCKF	*O. fraenata* *M.eugenii*	spleen PHA stim.ϕ	ccatcaagtcagatgtctggtcttttggc	61°C
LckSTART LckSTOP	*O. fraenata* *M. eugenii*	spleen PHA stim.ϕ	atgggctgctcctgcagctccagc tcatggctggggctggtactggccctc	64°C 67°C
Lckexpf Lckexpr	*O. fraenata*	thymus, spleen, axial node	gcacgctcccaatgaggaatggctctgac ccattttccattttgctccaggatc	66°C 56°C

5.3.2.2 Polymerase chain reaction, cloning and sequencing

Generally used RT-PCR templates, concentrations of the PCR mixes, and preparation of the RT-PCR and RACE-PCR products are detailed in Chapter 2 sections 2.1.6, 2.1.8 and 2.1.10. The PCR templates used to amplify TCRζ, ZAP-70 and Lck are listed in Table 5.4.

Table 5.4. PCR and RACE-PCR templates used for TCRζ, ZAP-70 and Lck in *O. fraenata*, *M. eugenii* and *M. domestica.*

Species	Gene of interest	PCR template	RACE PCR template
O. fraenata	TCRζ	No. 3	-
M. eugenii		No. 3	No. 2
M. domestica		No. 3	-
O. fraenata	ZAP-70	Nos. 1 and 2	No. 2
M. eugenii		Nos. 1 and 2	No. 2
O. fraenata	Lck	No. 1	Touchdown (60-50°C)
M. eugenii		No. 1	---

The templates are described in Chapter 2, section 2.1.8.1. The obtained products were visualized and prepared as detailed in Chapter 2, sections 2.1.8.3 and were subsequently cloned as described in Chapter 2, section 2.1.10. The obtained plasmids were prepared and enzyme digested to verify that the gene of interest was contained within the plasmids. Sequencing reactions were prepared as detailed in Chapter 2, section 2.1.10.6 in accordance with the specifications of the Australian Genome Research Facility (AGRF).

RACE primers were designed as outlined in Chapter 2, section 2.1.2 from the initial sequence derived from the RT-PCR products. RACE-PCRs were performed to obtain the 5' and 3' ends of the molecules as described in Chapter 2 section 2.1.9.

5.3.2.3 Phylogeny

A phylogenetic analysis using Mega5 was undertaken for all genes after the full coding domains were determined. The method is detailed in Chapter 2, section 2.3.

5.3.2.4 Bioinformatics

All of the bioinformatics tools detailed in Chapter 2 were used to elucidate putative domain structures, putative glycosylation sites, possible disulphide bonds, putative amino acid sequences, putative secondary and tertiary structures, and homology searches.

5.3.3 Protein extraction and protein concentration

Protein extraction was carried out as described in Chapter 2, section 2. 8.1. The concentration of the protein was determined by BCA assay using a Nanodrop

spectrophotometer (Thermo Fisher Scientific, Scoresby, Victoria, Australia) and a Qubit®fluorometer 1.0 as outlined in Chapter 2, sections 2.8.2 and 2.8.2.1 respectively.

5.3.3.1 Protein Gels

The extracted proteins were electrophoresed on a 12% SDS-PAGE gel at concentrations of 30, 40 and 50 µg/lane as described in Chapter 2, section 2.8.3. Staining was carried out with either Coomassie Blue or SYPRO® Ruby.

5.3.4 Western Blots

5.3.4.1 Western Blot ZAP-70

An amino acid alignment between *H. sapiens* and *M. eugenii* was investigated for sequence identity in order to trial a human ZAP-70 antibody. The alignment (Appendix 5B) shows the sites of the different available antibodies. A ZAP-70 (G4) antibody together with an HRP conjugated anti-mouse IgG secondary antibody (Santa Cruz Biotechnology, USA) was used for detection of the marsupial ZAP-70 protein. The human Jurkat cell lysate (Santa Cruz Biotechnology, USA, sc-2204) was used as a positive control. The Western Blot was carried out as outlined in Chapter 2 sections 2.8.4.

5.3.4.2 Western Blot Lck

The Lck antibody (2102): sc-13 from Santa Cruz Biotechnology (USA) was used in a Western Blot together with a HRP conjugated secondary antibody goat-anti rabbit IgG (Promega, Madison, USA) for the detection of the marsupial Lck protein. The human Jurkat cell line was used as a positive control. The Western Blot was carried out as outlined in Chapter 2, sections 2.8.4.

5.3.5 Expression studies

The concentration of the cDNA was maintained across the different tissue types used (spleen, liver, thymus, gut node) and was normalized against the house keeping gene GAPDH.

5.3 Results

5.3.1 Protein extraction and determination of concentration

The protein extracted from the different tissues and their concentrations are shown in Table 5.5. The extracted proteins were also used in subsequent experiments.

Table 5.5. Protein concentrations determined by BCA assay in *O. fraenata*, *M. eugenii* and *H. sapiens.*

Species	Tissue	Protein concentration
O. fraenata	Lung	36.83 µg/µL
	Spleen	38.80 µg/µL
	Thymus	11.5 µg/µL
	Axial Node	26.0 µg/µL
	Gut Node	18.0 µg/µL
M. eugenii	Thymus	10.44 µg/µL
H. sapiens	PBMC(*)	4.25 µg/µL

(*) PBMC = peripheral mononuclear cell. *H. sapiens* PBMCs were donated by Dr. L. J. Young.

5.3.2 Molecular characterization

5.3.2.1 Homology - T cell receptor zeta (TCR ζ)

A BLAST homology search revealed that the *M. eugenii* TCRζ sequence had an 89% sequence identity to the predicted sequence of *M. domestica* at the amino acid level, while at the nucleotide level the identity percentage was 88% with an e-value of 2e-158. The closest identity at the nucleotide level was to the *Macaca mulatta* (rhesus monkey) sequence with 80% identity an e-value of 7e-108, and at the amino acid level had an 83% identity with an e-value of 1e-87. The *O. fraenata* TCRζ nucleotide sequence had an 87% identity to the predicted *M. domestica* sequence and 79% to *M. mulatta*. At the amino acid level, *M. domestica* had an 81% identity to *M. mulatta*. Identity percentages to different species for TCRζ are shown in Table 5.6. The identity percentage between the two macropods *O. fraenata* and *M. eugenii* was 82% at the nucleotide level and 87% at the amino acid level.

Table 5.6. Homology search result for the *M. eugenii* and *O. fraenata* TCRζ nucleotide and amino acid sequences, their identities, and their respective e-values.

Homology search for *M. eugenii* and *O. fraenata* TCRζ chain								
	M. eugenii (***)				*O. fraenata (***)*			
Species	Nucleotide	e-value	Amino acid	e-value	Nucleotide	e-value	Amino acid	e-value
M.eugenii	100%	------	100%	------	82%	------	87%	------
M. domestica	88%	2e-158	89%	2e-88	87%	2e-154	89%	5e-88
M. mulatta	80%	7e-108	83%	1e-87	79%	3e-119	81%	2e-94
M. fascicularis	80%	7e-108	83%	1e-86	79%	3e-119	81%	5e-94
P. abelii	79%	1e-99	82%	3e-82	78%	5e-110	81%	4e-90
H.sapiens	79%	1e-99	82%	6e-82	77%	4e-105	79%	6e-90
E. caballus	------	------	75%	2e-73	77%	6e-65	73%	5e-82
O. cuniculus	78%	5e-97	78%	1e-80	76%	7e-102	77%	1e-88
S. scrofa	78%	5e-97	78%	4e-80	76%	9e-101	77%	4e-88
M. musculus	77%	4e-97	76%	2e-76	75%	4e-92	74%	6e-81
C. griseus	------	------	67%	6e-59	75%	5e-79	78%	1e-73
B. taurus	75%	1e-80	76%	2e-77	74%	3e-88	75%	2e-85
O. aries	74%	3e-75	75%	1e-76	73%	3e-81	74%	5e-84
G. gallus	71%	9e-56	67%	1e-59	71%	1e-60	63%	4e-67
A. carolinensis	67%	4e-28	59%	5e-51	68%	5e-41	61%	1e-62
X. laevis	------	------	50%	1e-43	------	-------	49%	3e-45
S. salar	------	------	47%	2e-27	------	-------	47%	9e-30
O. mykiss	------	------	44%	3e27	-------	-------	42%	6e-28
I. punctatus	------	------	43%	1e-26	-------	-------	44%	8e-27
O. niloticus	------	------	40%	9e-25	-------	-------	38%	3e-25
H. hippoglossus	------	------	38%	1e-17	-------	-------	38%	6e-18
D. rerio	------	------	39%	3e-19	-------	-------	36%	7e-22

------ = not recognized by BLAST, (***) sequence reported for the first time

5.3.2.2 TCRζ - Domain structure

The *M. eugenii* TCRζ molecule appeared to have no putative leader sequence, while the *O. fraenata* sequence had a predicted putative leader cleavage site at position 21. The result for the *M. eugenii* TCTRζ molecule is unusual since the TCRζ chain is a single-pass type I membrane protein and usually needs a signal sequence for integration into the endoplasmic reticulum. Other domains were identified and are listed in Table 5.7. Three ITAM motifs were identified in the TCRζ sequence in both species.

Table 5.7. Predicted structural domains, their positions, and their respective e-values within the TCRζ sequences of *M. eugenii* and *O. fraenata.*

Structural domains in the TCRζ chain of *M. eugenii* and *O. fraenata*						
	M. eugenii			*O. fraenata*		
Domain	Begin	End	e-value	Begin	End	e-value
PFAM	15	47	3.50e-19	28	60	1.20e-18
Transmembrane	20	38	-------	31	53	-------
ITAM	56	76	7.70e-03	69	89	7.70e-03
ITAM	94	115	1.32e-01	107	128	1.52e-01
ITAM	127	147	1.90e-02	140	160	1.90e-02

------ = not a significant e-value

5.3.2.3 TCRζ - Glycosylation and glycation sites

There were no putative O-linked glycosylation sites observed in any of the marsupial TCRζ sequences. At amino acid position 104, a single N-linked glycosylation site was predicted in the *M. eugenii* and *O. fraenata* TCRζ chains. The prediction for the N-linked glycosylation sites in both species carried a probability of 69%.

A number of putative glycation sites were observed in the marsupial TCRζ chains and these are shown in Table 5.8. However, validation of these results requires more sophisticated mass spectrometry or Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOFF) analysis. This was outside the scope of this study.

Table 5.8. Predicted glycated lysine residues in the TCRζ chains of *M. eugenii*, *O. fraenata* and *M. domestica* showing their positions and confidence levels .

Glycated lysine residues		
Species	Position	Confidence level
M. eugenii	44	74.9%
	47	93.8%
	102	79.1%
	124	84.5%
	138	89.6%
O. fraenata	2	91.9%
	57	73.4%
	60	93.6%
	137	84.3%
	151	89.6%
M. domestica	122	83.1%
	126	80.5%

5.3.2.4 TCRζ – Phosphorylation sites

Both *O. fraenata* and *M. eugenii* sequences contained predicted phosphorylation sites. The *M. eugenii* TCRζ sequence contained five serine and three tyrosine phosphorylation sites which are the ITAMs. The *O. fraenata* sequence contained four serine, one threonine and four tyrosine phosphorylation sites. The confidence levels of these predictions ranged from 64% to 97% for *M. eugenii,* and from 59% to 92% for *O. fraenata*. The positions of these putative phosphorylation sites are shown in Table 5.9.

Table 5.9. Predicted serine, threonine and tyrosine phosphorylation sites in *M. eugenii, O. fraenata* and in the partial sequence of *M. domestica*.

Distribution of phosphorylation sites					
M. eugenii		*O. fraenata*		*M. domestica (partial sequence)*	
Amino Acid	Position	Amino Acid	Position	Amino Acid	Position
Serine	10 49 63 95 110	Serine	23 62 76 123	Serine	36 98
Threonine	none	Threonine	108	Threonine	32 83
Tyrosine	70 109 141	Tyrosine	83 110 122 154	Tyrosine	57 85 97 129

5.3.2.5 TCRζ - Disulphide bonds

No predicted disulphide bonds were found in the TCRζ chains in any of the three marsupial species, *O. fraenata, M. eugenii* and *M. domestica*.

5.3.2.6 TCRζ - Primary sequence and secondary structure prediction

The *M. eugenii* TCRζ chain consisted of 459bp which translated into 152 amino acids. The *O. fraenata* TCRζ chain consisted of 498bp which translated into 165 amino acids. The putative amino acid sequences are shown above the nucleotide sequences in Figs. 5.1, 5.2 and 5.3. A central transmembrane helix segment with the sequence GILFIYGVIIT was detected. No leader sequence was predicted in the topology search thus confirming the leader sequence prediction. The secondary structure of each species is shown above the

putative amino acid sequence in Figs. 5.1, 5.2, and 5.3. There was a difference in the number of alpha helices between the *M. eugenii* and the *O. fraenata* TCRζ chains. The *M. eugenii* sequence contained eight helices while the *O. fraenata* sequence contained ten helices. How this affects the structure of the molecules is not known. An mRNA instability motif was found in the *M. eugenii* TCRζ 3' end.

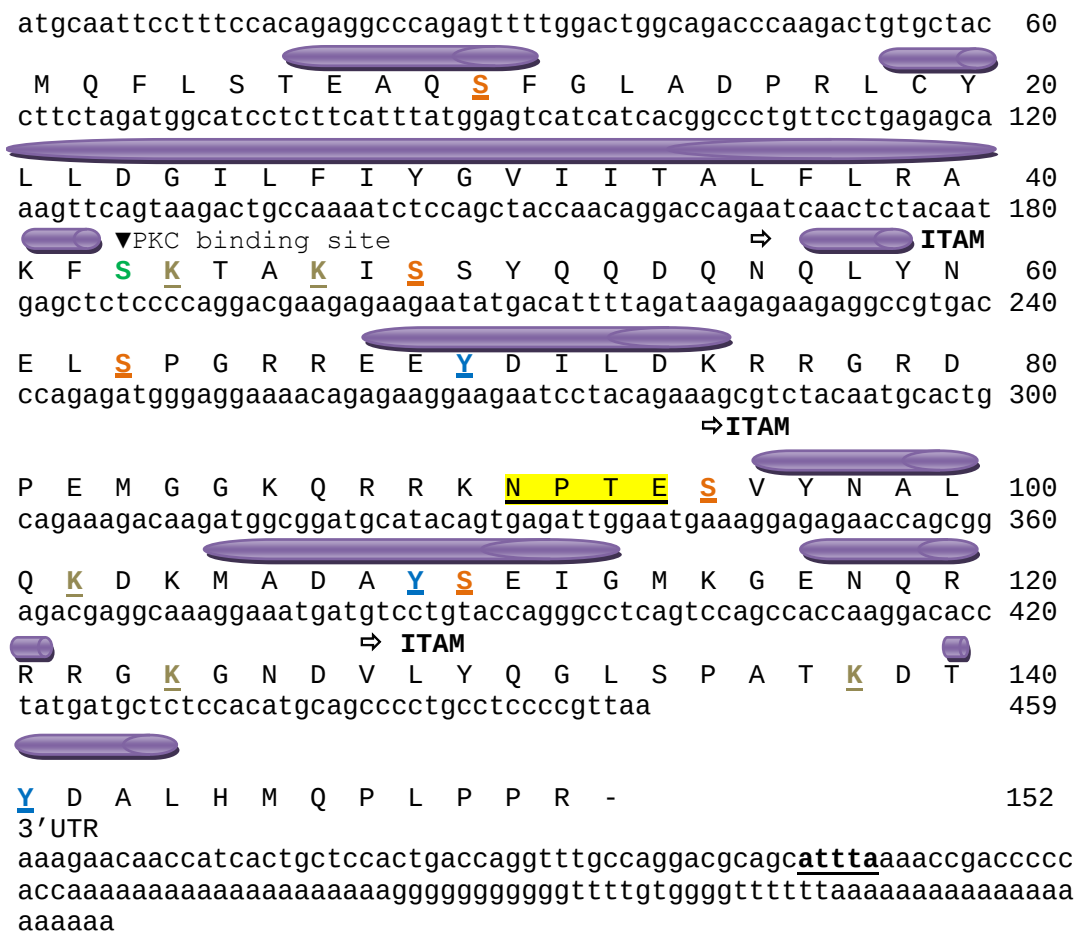


Figure 5.1. *M. eugenii* TCRζ chain primary sequence and secondary structure prediction. Highlighted in yellow and black underlined = single N- linked glycosylation site. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. ▼ = PKC binding site. ⇨ = ITAM domains. K = glycation sites. Bold and underlined is the mRNA instability (**attta**) motif. = Transmembrane helix.

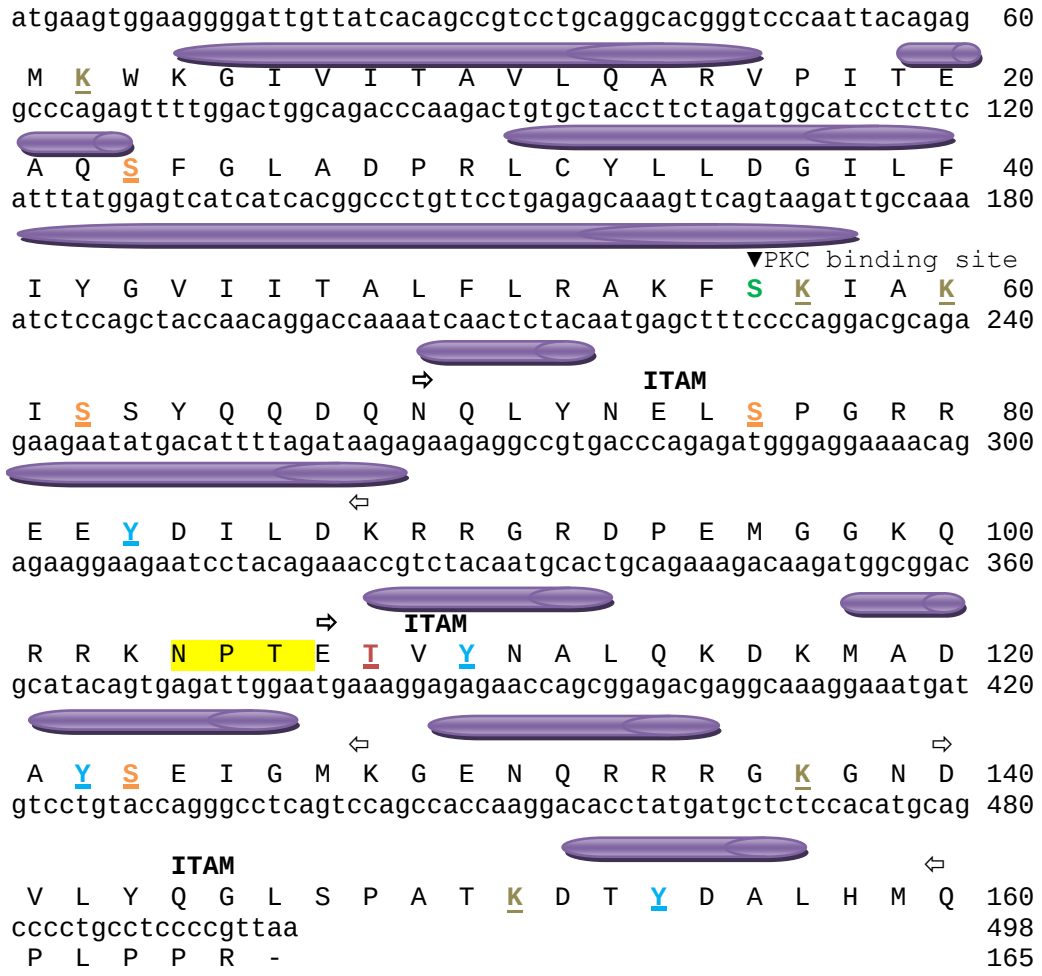


Figure 5.2. *O. fraenata* TCRζ chain primary sequence and secondary structure prediction. Highlighted in yellow and black underlined = single N- linked glycosylation site. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. ▼ = PKC binding site. ⇨ = ITAM domains. K = glycation sites. T = Threonine phosphorylation sites. Bold and underlined is the mRNA instability (**attta**) motif.
= Transmembrane helix.

Figure 5.3. *M. domestica* TCRζ partial primary sequence and secondary structure prediction. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. T = Threonine phosphorylation sites. ▼ = PKC binding site. K = glycation sites. = Transmembrane helices.

5.3.2.7 TCRζ - Phylogenetic analysis

A Neighbor-Joining phylogenetic tree revealed a clustering of the marsupial clade and a close phylogenetic relationship between the *M. eugenii* and *O. fraenata* TCRζ chains (Fig. 5.4). As expected, it appeared that the *M. domestica* sequence was further removed from the two macropods since the homology was lower. The optimal tree is shown with a sum of branch length = 2.18. The percentage of replicate trees in which the associated taxa are clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. The tree in Fig. 5.4 is drawn to scale with branch lengths in the same units as those of the evolutionary distances. This particular analysis involved 27 amino acid sequences where all positions containing gaps and missing data were eliminated leaving 135 positions in the final data set.

The marsupial clade is a sister clade to the avian clade as shown in the phylogenetic tree (Fig. 5.4).

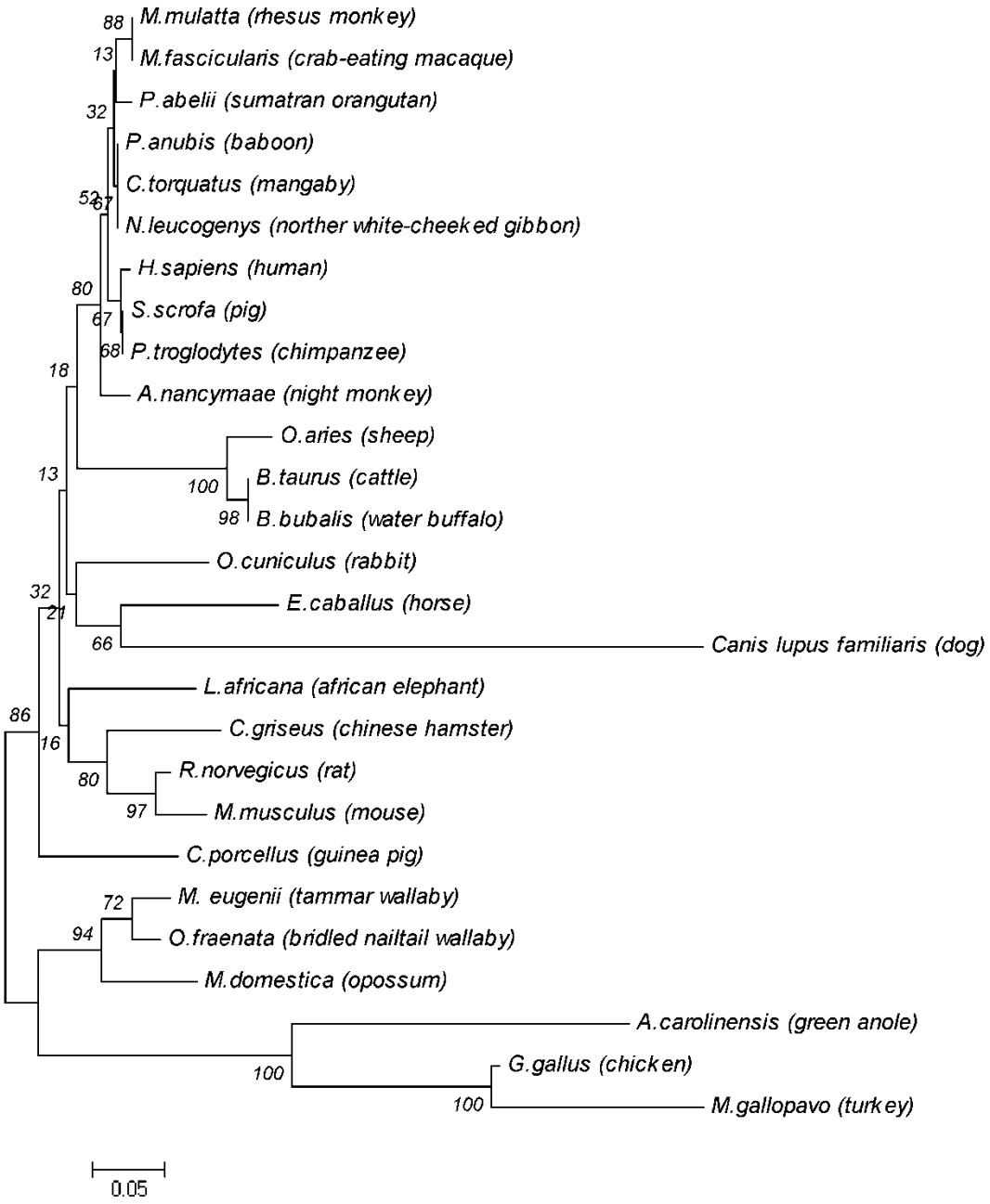


Figure 5.4. TCRζ Phylogenetic analysis by Neighbor-Joining method.

5.3.2.8 TCR ζ- Structure modelling

Homology modelling indicated that both *O. fraenata* and *M. eugenii* TCRζ structures are closest to the structure with the pdb identifier 2hacA (Fig.5.5).

Figure 5.5. pdb structure 2hacA

Homology model predictions carried out with a number of different programs showed that the region with the highest confidence level was located between amino acid residues 28 and 60. The Phyre program prediction indicated a hydrogen bonded turn at amino acid residue 29 and two bends at residues 58 and 59 in the 2hacA model but this could not be confirmed in the *O. fraenata* and *M. eugenii* sequences.

The Swiss model homology prediction executed through EXPASY indicated a low Q-mean Z-score of -4.98 for the *M. eugenii* TCRζ chain (Fig. 5.6 (a) and (b)). The homology model 2hacA showed an 81.25% sequence identity to the *M. eugenii* TCRζ chain. The quaternary structure is a dimer and it appeared that the *M. eugenii* TCRζ putative protein could also be modelled as a dimer. However, the model illustrated in Fig. 5.6 (a) falls outside the mean shown in the figure as **x**. The Z-score slider confirmed the very low Z-score readings (Fig. 5.6 (b)). This result is the same as the *O. fraenata* TCRζ chain, consequently only the *M. eugenii* result is shown.

a) b)

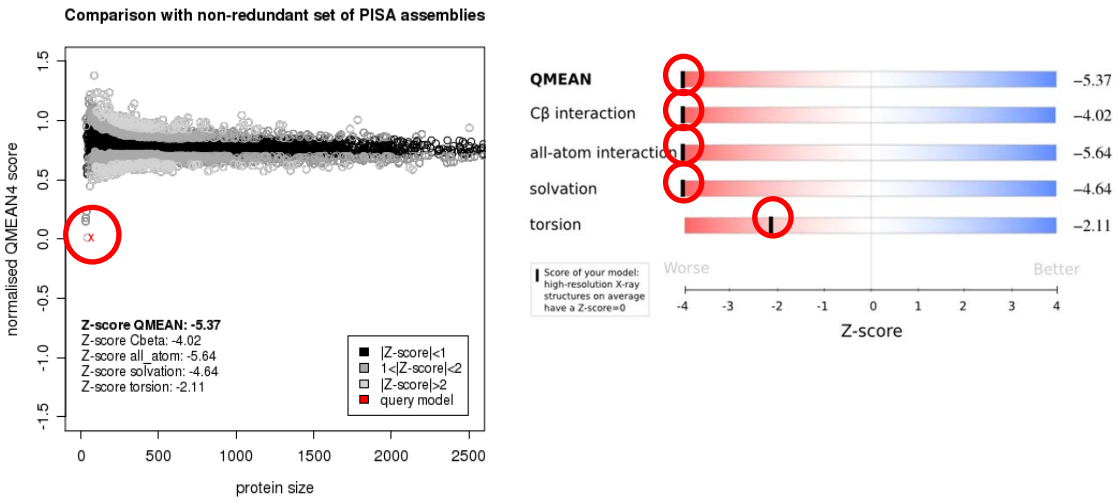
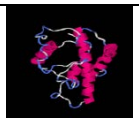
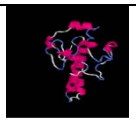
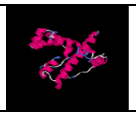


Figure 5.6. QMean Z-score for both *O. fraenata* and *M. eugenii* TCRζ chains produced by Swiss Model. Red circles indicate the low values.

Five homology models identified with the I-TASSER program for the *M. eugenii* and *O. fraenata* TCRζ putative protein sequence showed C-scores above -4 except for model 1 which had a distinctly higher value of -3.77 (Table 5.10).

Table 5.10. Homology models of TCRζ for *O. fraenata* and *M. eugenii*. The model with the highest C-score was selected as the most probable and is highlighted in red.

Homology models of TCRζ for *O. fraenata* and *M. eugenii*							
Species	Model No.	Model	C-score	Species	Model No.	Model	C-score
O. fraenata	1		-3.77	*M. eugenii*	1		-3.88
	2		-4.40		2		-4.08
	3		-4.56		3		-3.94
	4		-4.74		4		-4.27
	5		-4.75		5		-4.85

The I-TASSER program predicted model No. 1 in both cases as having the highest estimated accuracy of 0.31±0.10 (TM-score). Models 1, 4, and 5 had similar topologies. Models 2 and 3 did not conform to the secondary structure predictions shown in Figs. 5.1 and 5.2.

The Modweb program which uses modeller 9.10 as a platform constructed the structures directly from the putative protein sequences. The models with the highest confidence values for *M. eugenii* and *O. fraenata* were selected and are shown in Figure 5.7.

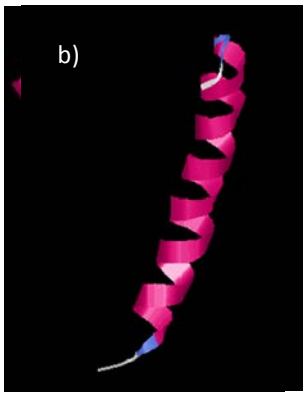


Figure 5.7. The homology models of the putative TCRζ proteins produced by Modweb (a) *O. fraenata* and (b) *M. eugenii.*

The 3D-Jigsaw program identified alternative homology models for the *O. fraenata* and *M. eugenii* TCRζ chains. A different arrangement of the helices was observed in the models and these are shown in Figure 5.8. In addition, coils and loops which were not visible in the Modweb generated models were detected in Fig. 5.8.

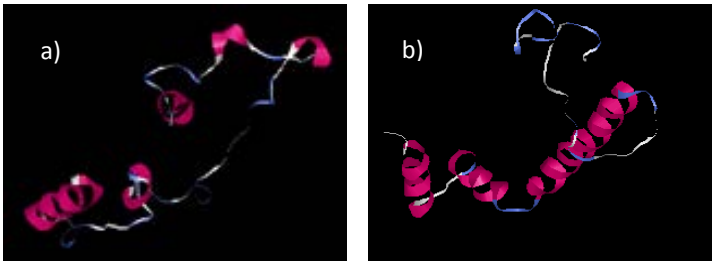


Figure 5.8. Alternative TCRζ homology models produced by 3D-Jigsaw for (a) *O. fraenata* and (b) *M. eugenii*.

Ligand connectivity studies could not be conducted for *M. eugenii* and *O. fraenata* TCRζ due to there being no data in the ligand database for this protein. However, binding sites determined by the I-TASSER program are illustrated in Figure 5.9. The ligand binding sites in the *M. eugenii* sequence are much bulkier than those in the *O. fraenata* sequence.

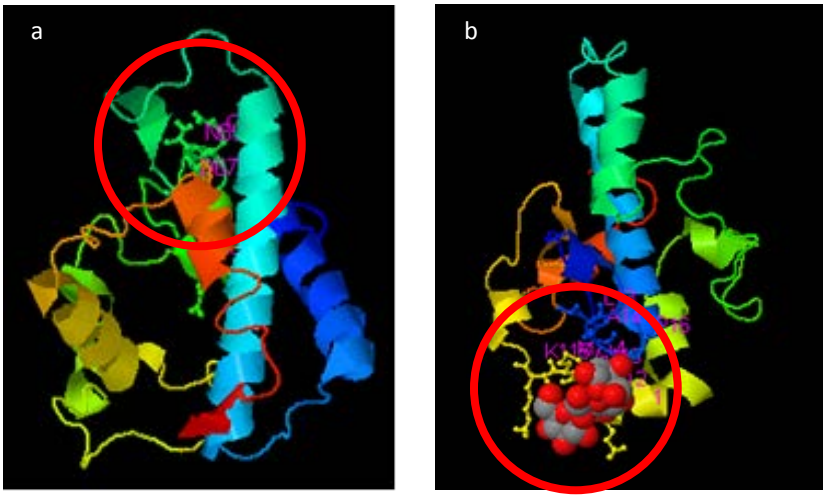


Figure 5.9. TCRζ putative binding sites identified with I-TASSER (a) *O. fraenata* and (b) *M. eugenii*.

Sixty two percent (62%) of both the *O. fraenata* and *M. eugenii* TCRζ chains were predicted as being disordered and hence did not adopt a stable structure. These areas are of interest since they have a close relationship with protein expression and functionality. In the case of the marsupial TCRζ chain this indicated that the fold property of this protein is different to that of other TCRζ proteins deposited in the pdb databank.

5.3.2.9 TCRζ - Semi-quantitative expression studies

The image of the 2% agarose gel in Fig. 5.10 revealed a high intensity of expression of the TCRζ gene in *O. fraenata* spleen and thymus tissue, and much lower intensity in lung and axial node tissue. This may indicate that the TCRζ gene is upregulated in the spleen and to a lesser extent in the thymus, and is perhaps down-regulated in the lung and the axial node.

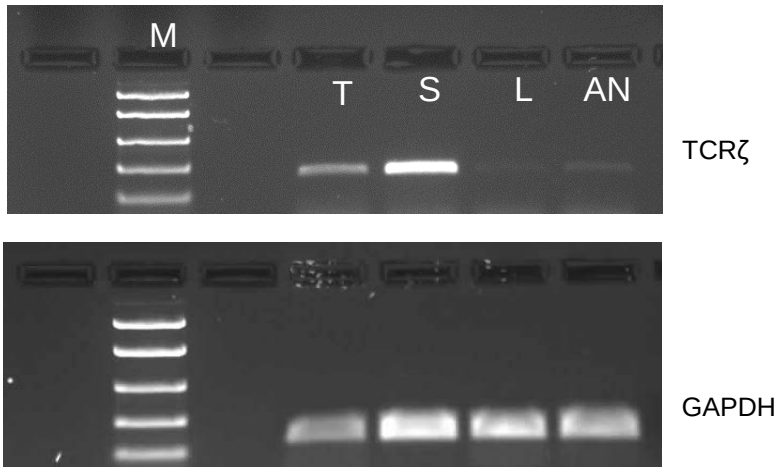


Figure 5.10. Image of 2% agarose gel showing the expression of the TCRζ gene in *O. fraenata* against the housekeeping gene GAPDH (T = thymus, S = spleen, L = lung, AN = axial node).

5.3.3 Zeta associated protein of 70 kDa (ZAP-70)

5.3.3.1 ZAP-70 - Homology

A BLAST homology search for the *M. eugenii* ZAP-70 sequence showed an 87% sequence identity to the predicted sequence of *M. domestica* with the closest sequence being that of *S. scrofa* which had a sequence identity of 78%. E-values from the BLAST search were 0.0 for both of these sequence identities. The *O. fraenata* partial sequence showed an 88% identity to *M. domestica* at the nucleotide level and 93% identity at the amino acid level over the 338 amino acids in the partial sequence. Other identity percentages and e-values for various mammalian species compared to the *M. eugenii* open reading frame and the *O. fraenata* partial sequence are shown in Table 5.11.

Table 5.11. Homology search results for the *M. eugenii* and *O. fraenata* ZAP-70 nucleotide and amino acid sequences, their identity and respective e-values. For the O. fraenata partial sequence the values were calculated over the 1016bp/338aa length of the partial sequence.

Homology search for ZAP-70								
Species	*M. eugenii*				*O. fraenata* (partial sequence 1016 bp, 338 aa)			
	nt	e-value	aa	e-value	nt	e-value	aa	e-value
M. eugenii	100%	0.0	100%	0.0	100%	0.0	93%	0.0
M. domestica	87%	0.0	89%	0.0	88%	0.0	90%	0.0
S. scrofa	78%	0.0	82%	0.0	79%	0.0	80%	2e-173
P. abelii	78%	0.0	74%	0.0	-----	-----	68%	9e-134
O. cuniculus	77%	0.0	82%	0.0	78%	0.0	80%	8e-171
N. leucogenys	77%	0.0	79%	0.0	78%	0.0	81%	6e-172
H. sapiens	77%	0.0	79/5	0.0	78%	0.0	80%	7e-171
A.melanoleuca	77%	0.0	80%	0.0	79%	0.0	79%	2e-167
M. mulatta	77%	0.0	82%	0.0	78%	0.0	-----	-----
B. taurus	77%	0.0	81%	0.0	78%	0.0	80%	4e-171
C. griseus	77%	0.0	76%	0.0	78%	0.0	79%	1e-166
R. norvegicus	77%	0.0	81%	0.0	78%	0.0	79%	4e-169
E. caballus	77%	0.0	------	------	79%	0.0	81%	1e-170
M. musculus	77%	0.0	81%	0.0	78%	0.0	80%	1e-169
C. jacchus	77%	0.0	81%	0.0	77%	0.0	80%	6e-171
T. guttata	75%	0.0	------	------	73%	0.0	74%	1e-156
C.l. familiaris	74%	0.0	86%	0.0	79%	0.0	80%	1e-172
M. gallopavo	74%	0.0	75%	0.0	73%	0.0	74%	1e-158
H. hippoglossus	73%	5e-123	62%	0.0	-----	-----	60%	1e-124
X. laevis	71%	0.0	73%	0.0	72%	2e-176	72%	6e-151
D. rerio	70%	2e-96	64%	0.0	-----	-----	63%	2e-131

nt = nucleotide, aa = amino acid, ----- = not recognized by BLAST algorithms.

Table 5.12 lists the results of the BLAST homology search for the partial sequences of the macropod *L. hirsutus* and the didelphid *M. domestica*. A comparison of the partial *M. domestica* sequence to the predicted sequence showed multiple differences between them (supporting data in Appendix 5B). This may be due to polymorphisms in the gene however no investigations into polymorphisms were carried out in this study.

Table 5.12. Homology search results showing sequence identities for the *L. hirsutus* and *M. domestica* partial ZAP- 70 sequences at the nucleotide and putative amino acid levels and the respective e- values.

Homology search for *L. hirsutus* and *M. domestica* ZAP-70								
Species	*L. hirsutus* (partial sequence 732nt, 243aa)				*M. domestica* (partial sequence 819nt, 272aa)			
	nt	e-value	aa	e-value	nt	e-value	aa	e-value
M. domestica (predicted)	89%	0.0	92%	3e-130	89%	0.0	90%	6e-144
S. scrofa	79%	0.0	90%	1e-115	79%	0.0	90%	4e-143
P. abelii	84%	1e-83	-----	-----	83%	6e-107	78%	1e-88
O. cuniculus	79%	0.0	88%	4e-115	78%	0.0	90%	8e-142
N. leucogenys	79%	0.0	90%	1e-115	78%	0.0	90%	1e-142
H. sapiens	79%	0.0	88%	2e-114	78%	0.0	90%	8e-142
A.melanoleuca	79%	0.0	86%	4e-112	78%	0.0	89%	4e-139
M. mulatta	-----	-----	-----	-----	-----	-----	90%	6e-143
B. taurus	78%	0.0	90%	1e-114	78%	0.0	90%	2e-141
C. griseus	78%	0.0	90%	1e-111	78%	0.0	91%	3e138
R. norvegicus	78%	0.0	92%	1e-113	77%	0.0	91%	2e-140
E. caballus	79%	0.0	88%	8e-116	79%	0.0	90%	1e-142
M. musculus	78%	0.0	90%	3e-114	78%	0.0	91%	9e-141
C. jacchus	77%	0.0	90%	4e-115	77%	0.0	90%	4e-142
T. guttata	75%	1e-177	84%	4e-109	74%	0.0	90%	9e-130
C.l. familiaris	79%	0.0	86%	4e-115	79%	0.0	89%	3e-142
M. gallopavo	-----	-----	86%	4e-109	-----	-----	92%	3e-132
H. hippoglossus	82%	1e-88	76%	2e-79	78%	1e-96	86%	9e-101
X. laevis	72%	9e-123	82%	2e-100	71%	8e-137	89%	6e-125
D. rerio	75%	3e-78	75%	1e-87	72%	3e-78	84%	3e-108

nt = nucleotide, aa = amino acid, ----- = not recognized by BLAST.

5.3.3.2 ZAP-70 - Domain structure

A variety of domains were identified by SMART in the ZAP-70 open reading frames in *M. eugenii*, *O. fraenata*, *L. hirsutus* and *M. domestica* and their locations and respective e-values are shown in Table 5.13. The principal domains identified were the SH_2 and tyrosine kinase domains.

Table 5.13. Predicted domain structures identified by SMART together with locations in the sequence and e- values in *M. eugenii*, *O. fraenata*, *L. hirsutus* and *M. domestica*.

Domain structures			
Species	Domain structure	Location	e-value
M. eugenii	SH_2	8-93	1.68e-25
	SH_2	161-245	5.30e-23
	Tyrosine kinase	336-591	3.03e-105
	Internal repeat	4-102	3.14e-12
O. fraenata	SH_2	1-68	6.92e-04
	Tyrosine kinase	159-337	4.09e-29
L. hirsutus	STYKc	121-140	1.25e-04
M. domestica	SH_2	1-71	5.04e-06
	STYKc	162-177	1.15e-03

SignalP-4.0 predicted a single spike that indicated a signal peptide with a cleavage site at amino acid residue 21. However, the predicted site did not reach the threshold at 50% but held at 45%. This was not high enough to identify it as a signal peptide even though the open reading frame had a number of putative N-linked glycosylation sites which require a signal peptide.

5.3.3.3 ZAP-70 - Glycosylation and glycation sites

The ZAP-70 sequence of *M. eugenii* contained a single putative C-mannosylated site at position 528 with a confidence level of 61.8%. Three possible O-linked glycosylation sites were predicted at positions 263, 265 and 615. Only position 265 with a confidence level of 59.1% was above the threshold of 50%. A number of possible glycation sites were predicted in the *M. eugenii* (27)*, O. fraenata* (24)*, L hirsutus* (12) *and M. domestica* (15) ZAP-70 sequences.

Four putative N-linked glycosylation sites were predicted in the *M. eugenii* and *O. fraenata* sequences of ZAP-70, while *M. domestica* and *L. hirsutus* had three predicted N-linked glycosylation sites. The positions, the confidence levels and the signal strengths of these sites are shown in Table 5.14.

Table 5.14. Predicted N-linked glycosylation sites in ZAP-70 their positions, sequences, confidence levels, and signal strengths in *M. eugenii*, *O. fraenata*, *L. hirsutus* and *M. domestica.*

Predicted N-linked glycosylation sites in ZAP-70				
Species	Position	Sequence	Confidence level	Signal strength
M. eugenii	66	NGTY	71.06%	++
	256	NASA	57.32%	+
	280	NDTL	60.27%	++
	296	NKSQ	69.36%	++
O. fraenata	79	NASA	60.9%	+
	103	NDTL	63.6%	++
	119	NKSQ	71.1%	++
	406	NFSD	54.6%	+
L. hirsutus	41	NASA	62.1%	+
	65	NDTL	64.5%	++
	81	NKSQ	72.1%	++
M. domestica	82	NASA	59.9%	+
	106	NDTL	62.2%	++
	122	NKSQ	70.0%	++

5.3.3.4 ZAP-70 – Phosphorylation sites

Several putative phosphorylation sites including nineteen serines, five threonines and fourteen tyrosines were predicted in *M. eugenii*. The partial sequences of O. *fraenata,* and *M. domestica* each had six serine, four threonine, and eight tyrosine predicted phosphorylation sites. *L. hirsutus* had six serine, three threonine, and four tyrosine predicted phosphorylation sites. The positions of these putative phosphorylation sites are shown in Table 5.15.

Table 5.15. Predicted amino acid phosphorylation sites within the ZAP-70 sequences of *M. eugenii*, *O. fraenata*, *L. hirsutus* and *M. domestica*.

Predicted amino acid phosphorylation sites in ZAP-70			
Species	Serine positions	Threonine positions	Tyrosine positions
M. eugenii	14 16 42 106 117 121 169 179 181 201 218 298 311 315 318	130 227 282 289 492	46 69 164 198 221 288 313 317 355
O. fraenata	41 121 134 138 194 338 339 341 345 408 423	50 105 112 315	44 111 136 140 178 295 313 314 356 390
L. hirsutus	83 96 100 103 156 228	12 67 74	6 73 98 102
M. domestica	5 7 124 137 141 144 197	53 108 115	24 47 114 139 143 181

In addition to the putative amino acid phosphorylation sites, a number of putative O-linked glycosylation sites were also predicted in the four marsupial ZAP-70 sequences and are shown in Table 5.16. The confidence levels of the predicted O-linked glycosylation sites ranged from the threshold level of 50% up to 61%.

Table 5.16. Predicted O-linked glycosylation their positions and confidence levels sites in the ZAP-70 sequences of four marsupial species.

Species	Position	Confidence level
M. eugenii	263	50.5%
	269	52.9%
	615	56.8%
O. fraenata	86	57.2%
	88	55.9%
	92	58.9%
L. hirsutus	46	53.6%
	48	59.9%
	50	58.6%
	54	61.4%
	60	52.5%
	67	52.1%
M. domestica	87	52.6%
	89	58.9%
	91	57.7%
	95	60.5%
	101	51.5%
	108	51.1%

5.3.3.5 ZAP-70 - Disulphide bonds

The *M. eugenii* open reading frame contained fifteen predicted cysteines. Six of these cysteines formed predicted disulphide bonds. Twelve predicted cysteines which formed six predicted disulphide bonds were found in the partial sequence of *O. fraenata*. Four predicted cysteines which formed two predicted disulphide bonds were found in the *L. hirsutus* partial sequence, and six predicted cysteines which formed three predicted disulphide bonds were found in *M. domestica*. Both the *M. eugenii* and *O. fraenata* ZAP-70 sequences had cysteine residues that bind a Zinc ligand. The location of the predicted disulphide bridges and the status of the cysteines involved are shown in Table 5.17.

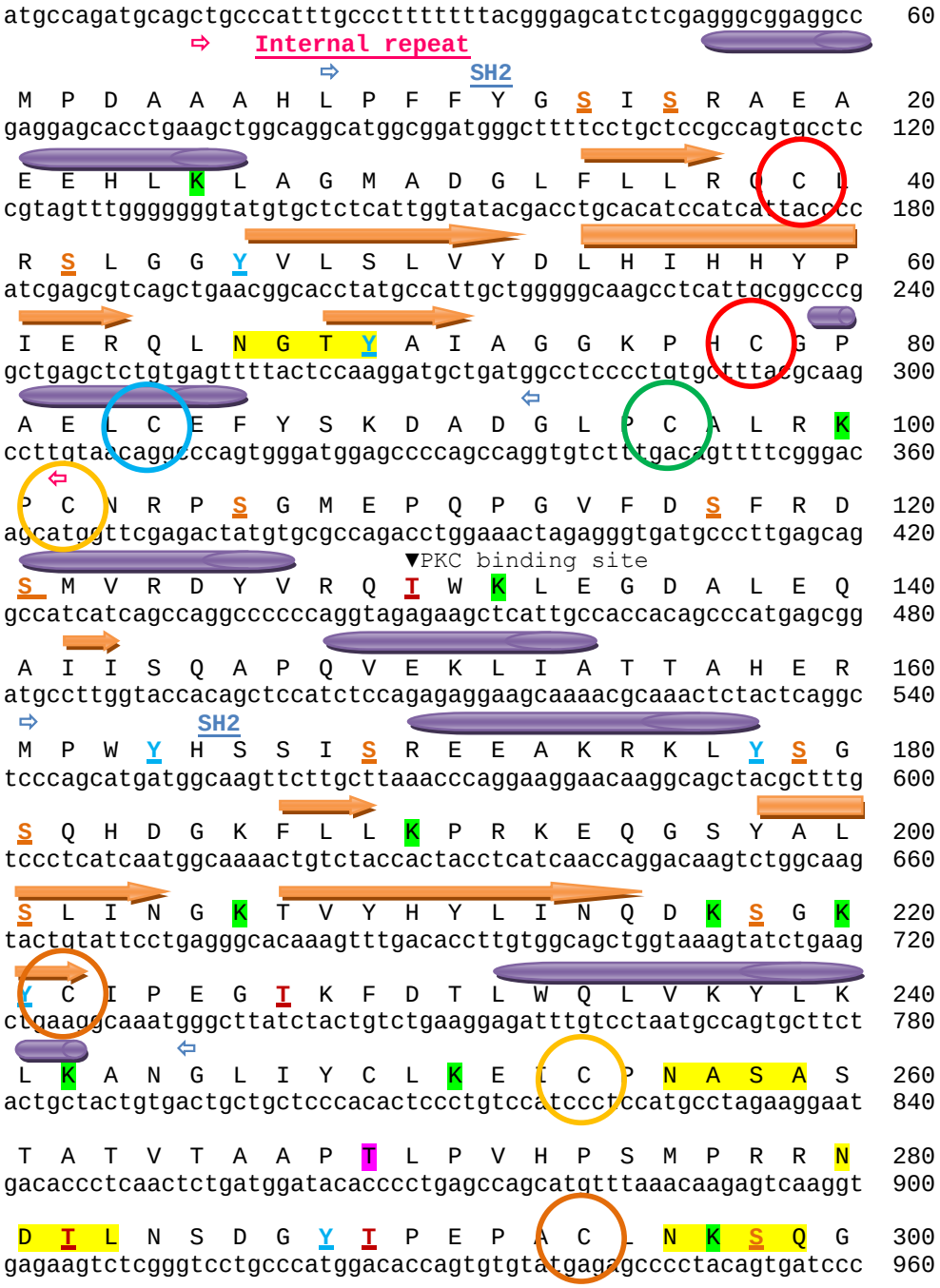
Table 5.17 Predicted disulphide bonds, their connectivity and locations in the ZAP-70 sequences of four marsupial species.

Predicted disulphide bonds in the ZAP-70 sequence of four marsupial species			
Species	Disulphide bond	Location	Cysteine type
M. eugenii	FLLRQCLRSLG-GGKPHCGPAEL	39-78	free
	GPAELCEFYSK - DPPPECPPNMY	84-562	½ cysteine binds free cysteine
	ADGLPCALRKP - SGNFGCVRKGV	96-349	½ cysteine binds free cysteine
	ALRKPCNRPSG - CLKEICPNASA	102-254	½ cysteine binds free cysteine
	ALRKPCNRPSG - CLKEICPNASA	222-294	½ cysteine binds free cysteine
	RIIGVCKAEAL - TLMKKCWIYKW	403-573	free
	WYAPECINYRK - YRKFSCQSDVW	508-516	Binds zinc ligand
O. fraenata	KSGKYCIPEGT - CLKEICPNASA	45-77	½ cysteine binds free cysteine
	NGLIYCLKEIC - SGNFGCVRKGV	72-172	free
	RIIGVCKAEAL - TLMKDCWIYKW	226-396	Binds zinc ligand
	WYAPECINYRK - DRPPECPPDMY	331-385	ligand binding cysteines (Zn)
L. hirsutus	KSGKYCIPEGT - CLKEICPNASA	7-39	free
	DGLIYCLKEIC - SGNFGCVRKGV	34-134	free
	RIIGVCKAEAL - KGGGPCKQCRG	188-218	free
	GPCKQCRGAAA - RRKKFCAPXXX	221-241	½ cysteine
M. domestica	KSGKYCIPEGT - NGLIYCLKEIC	48-75	½ cysteine-free cysteine
	CLKEICPNASA - SGNFGCVRKGV	80-175	free
	TPEPACLNKSQ - SNVVSCCTKWP	120-265	free
	RIIGVCKAEAL - NVVSCCTKWPW	229-266	free

Yellow highlights = zinc ligand binding cysteine residue.

5.3.3.6 ZAP-70 - Primary sequence and secondary structure prediction

The *M. eugenii* ZAP-70 (Fig. 5.11) coding domain was found to be 1,854bp long which translated into 617 amino acids compared to 619 amino acids in the *H. sapiens* sequence. The alpha helices and beta strands were distributed differently within the marsupial ZAP-70 molecule. There were 18 beta strands in the open reading frame of *M. eugenii* and 27 beta strands in the partial sequence of *O. fraenata* (Fig. 5.12). Seven beta strands were found in *L. hirsutus* (Fig. 5.13) and 10 beta strands were found in *M. domestica* (Fig. 5.14). In contrast, 19 helices were found in *M. eugenii*, and in the partial sequences, 12 were found in *O. fraenata*, six in *L. hirsutus* and five in *M. domestica*. The length of the helices, and the lengths of the beta strands differed in the ZAP-70 molecules of each species. The effects of this on the folding properties of the molecules are unknown.



Cont. page 227

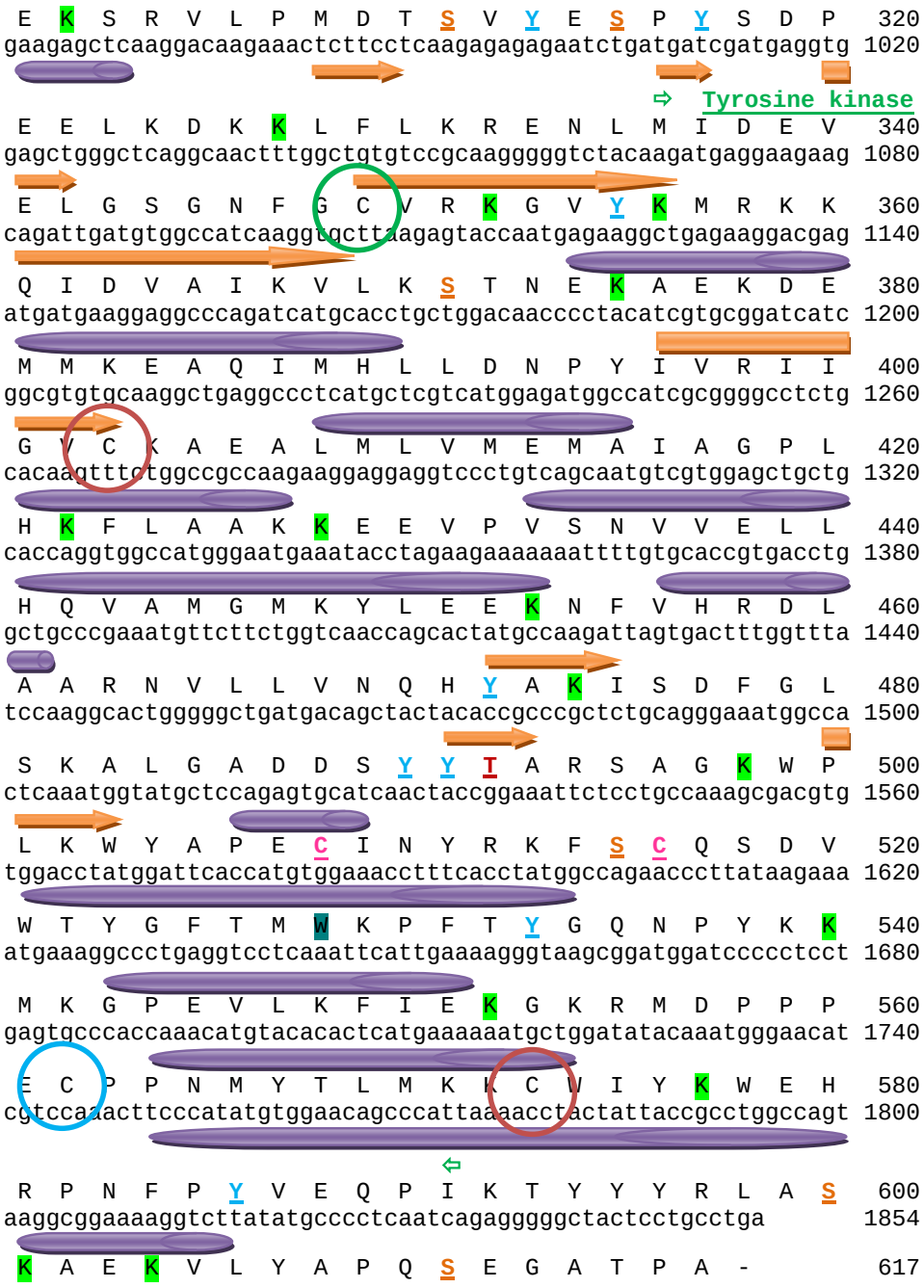
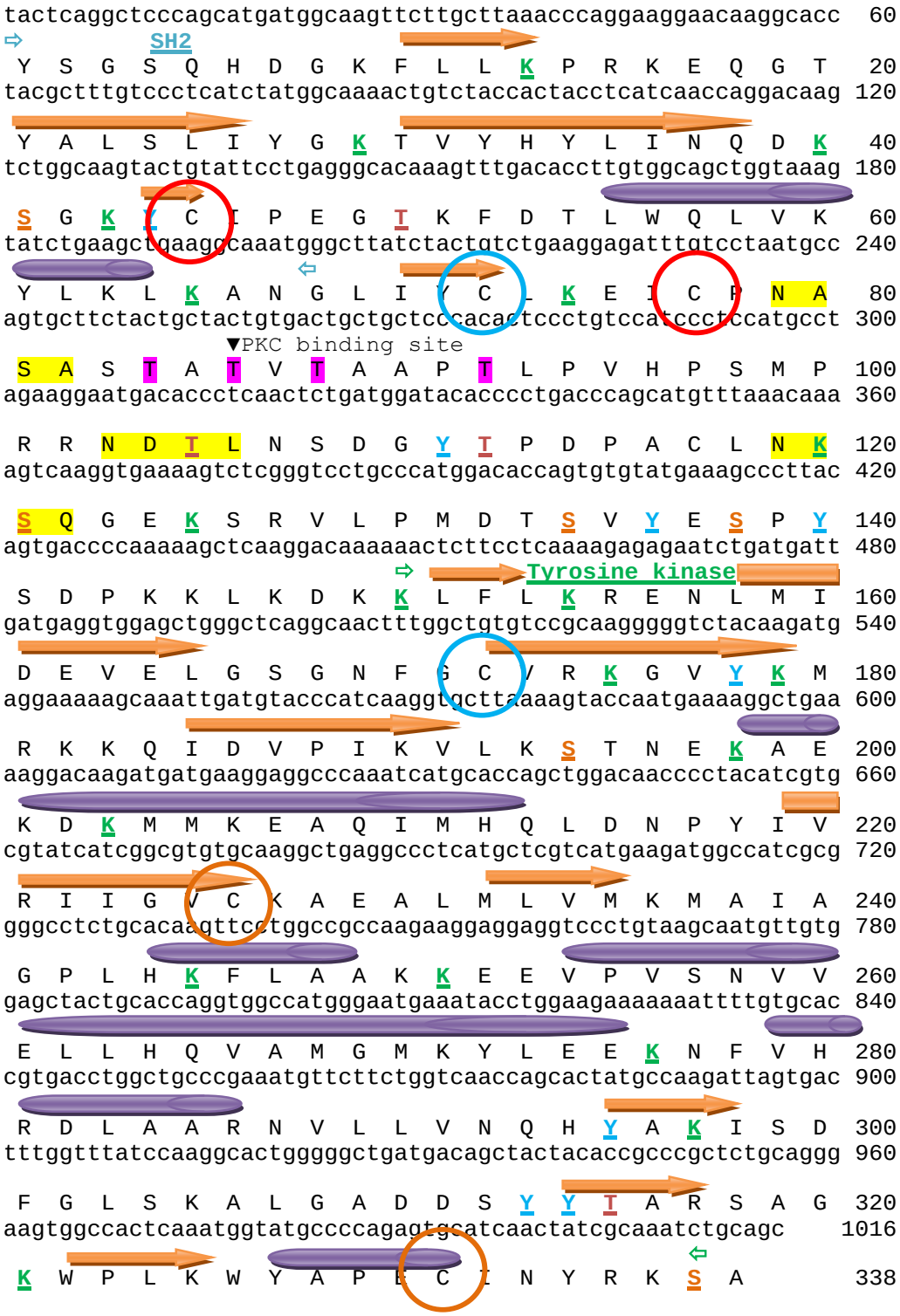


Figure 5.11. *M. eugenii* ZAP-70 primary sequence and secondary structure prediction.
K = glycation sites. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. T = threonine phosphorylation sites. Highlighted in yellow = N-linked glycosylation sites. W = C- mannosylation site. T = threonines are O-linked glycosylation sites. Different colours = disulphide bridges. C = ligand binding cysteine.▼= PKC binding sites.
= Transmembrane helix. = Strands.



Cont. page 229

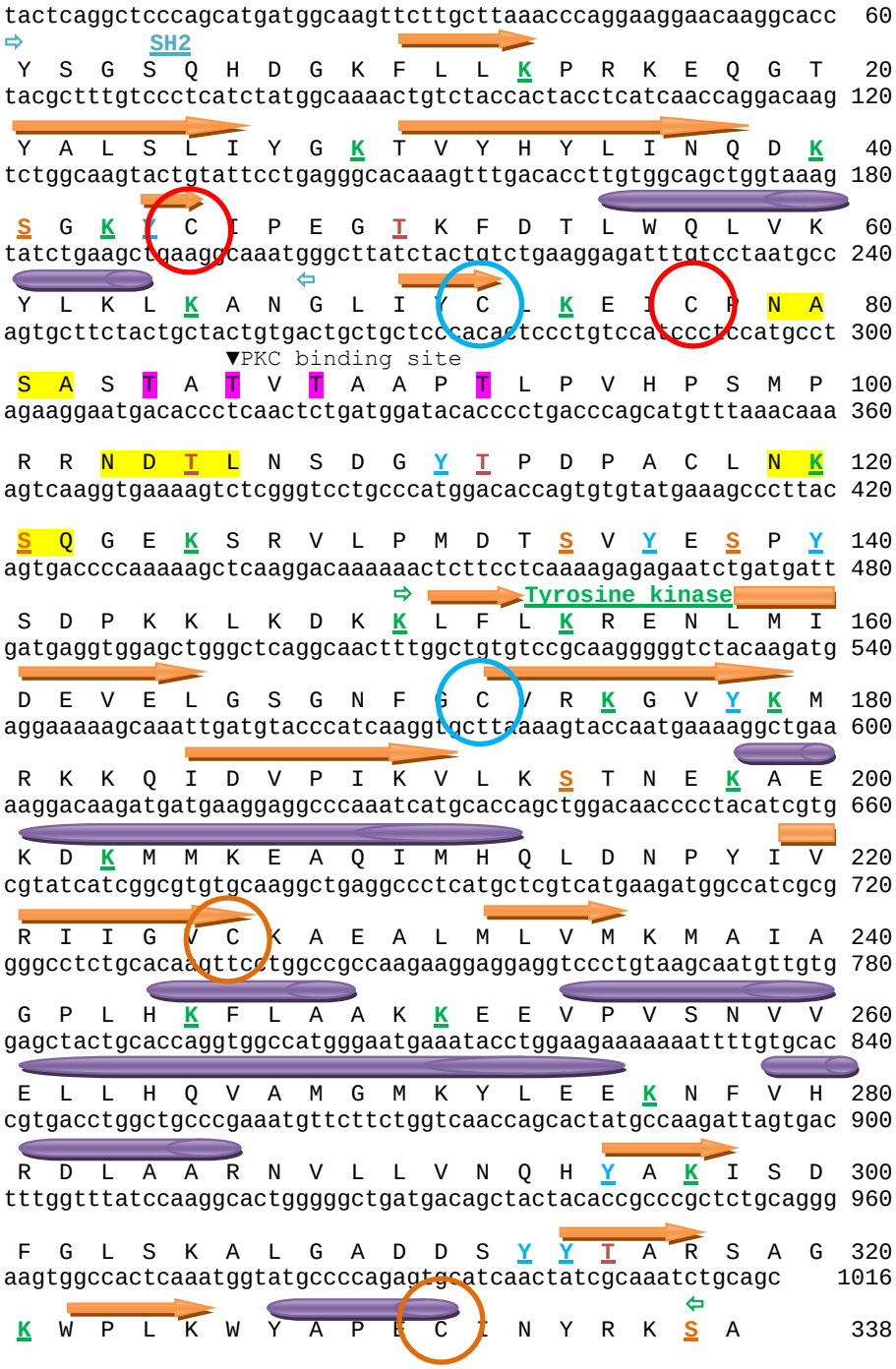


Figure 5.12. *O. fraenata* partial ZAP-70 primary sequence and secondary structure prediction. K = glycation sites. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. T = threonine phosphorylation sites. Highlighted in yellow = N-linked glycosylation sites. T = threonines are O-linked glycosylation sites. Different colours = disulphide bridges. C = ligand binding cysteine. ▼ = PKC binding sites. = Transmembrane helix. = Strands.

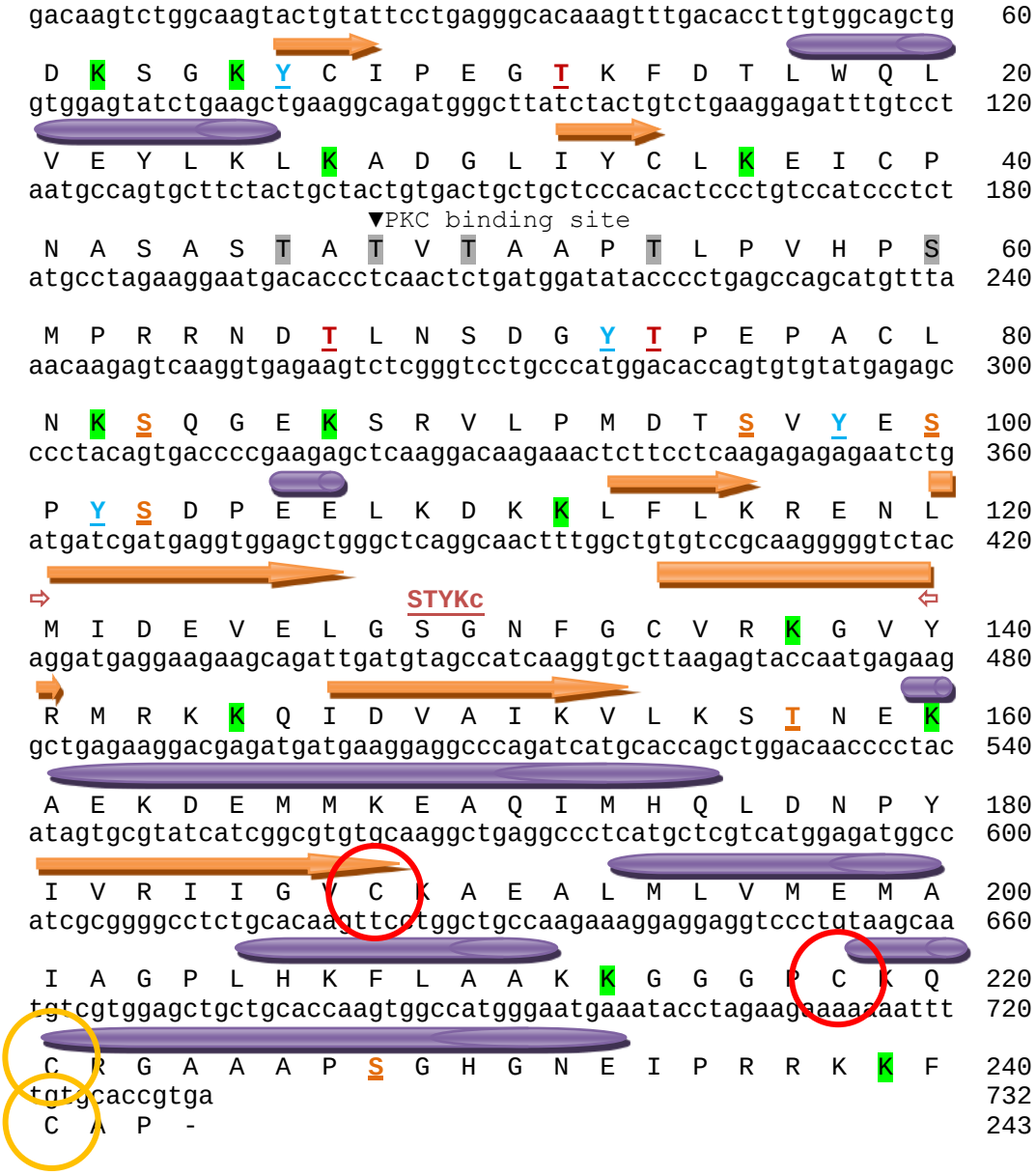


Figure 5.13. *L. hirsutus* ZAP-70 partial primary sequence and secondary structure prediction. K = glycation sites. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. T = threonine phosphorylation sites. Highlighted in yellow = N-linked glycosylation sites. W = C- mannosylation site. T = threonines are O-linked glycosylation sites. Different colours = disulphide bridges. ▼ = PKC binding sites.

= Transmembrane helix. = Strands.

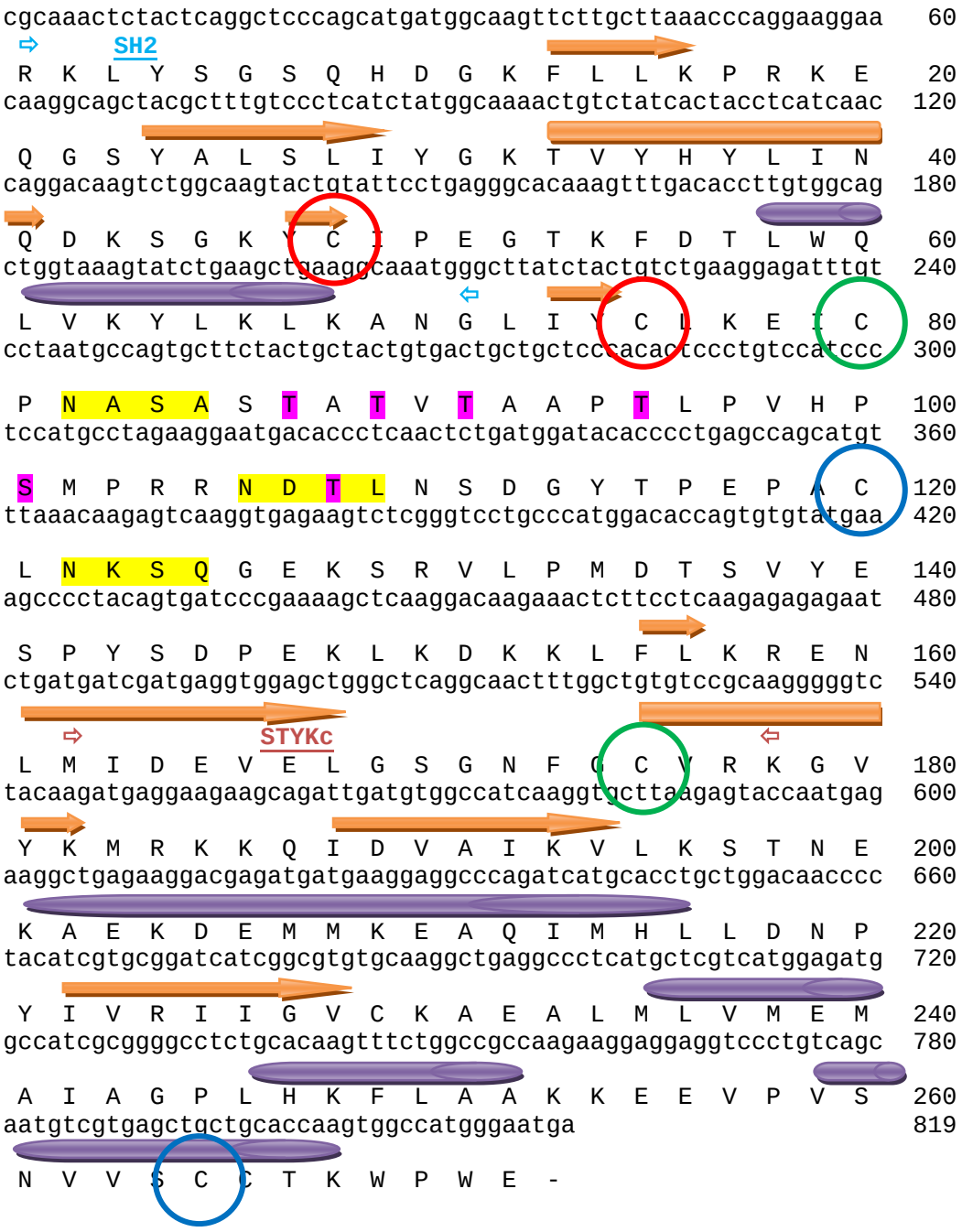


Figure 5.14. *M. domestica* partial ZAP-70 primary sequence (partial) and secondary structure prediction. K = glycation sites. Y = tyrosine phosphorylation sites. S = serine phosphorylation sites. T = threonine phosphorylation sites. Highlighted in yellow = N-linked glycosylation sites. W = C- mannosylation site. T = threonines are O-linked glycosylation sites. Different colours = disulphide bridges. C = ligand binding cysteine.▼ = PKC binding sites. = Transmembrane helix. = Strands.

5.3.3.7 ZAP-70 - Phylogenetic analysis

The evolutionary history of ZAP-70 was inferred by the Maximum Likelihood method based on the Poisson correction model. The highest logarithmic likelihood in the phylogenetic tree was -9,093.93. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 26 amino acid sequences. There were a total of 745 positions in the final dataset (Fig. 5.15).

The marsupial clade is sister to a clade containing eutherian mammals. The amphibian clade is basal to both groups. The branch lengths of *M. eugenii* and *O. fraenata* are short indicating a close relationship. By comparison, *M. domestica* had an extended branch length indicating a more distant relationship. Although only a partial sequence was used for *O. fraenata* it was confirmed in the phylogenetic analysis as being part of the marsupial clade due to the high conservation of this molecule in both *M. eugenii* and *O. fraenata*.

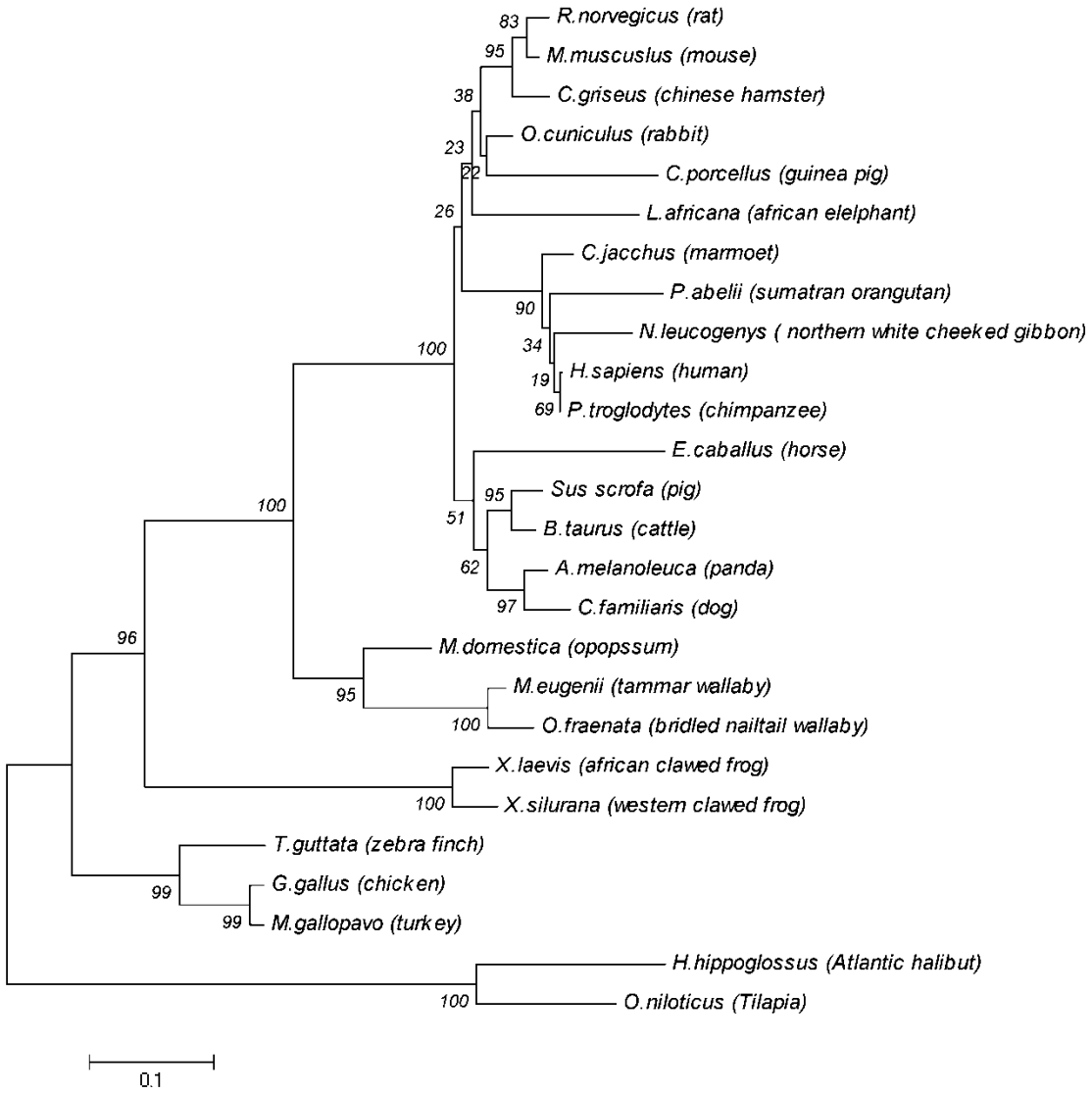


Figure 5.15. Phylogenetic tree for ZAP-70 compiled using the Maximum Likelihood method.

5.3.3.8 ZAP-70 - Structure modelling

Homology modelling carried out with a number of programs indicated that the I-TASSER (Fig. 5.16 c) and Modweb (Fig. 5.16 b) programs were in agreement with the template 1m61. RaptorX could not model the full open reading frame of the *M. eugenii* ZAP-70 sequence but modelled the domains. The template 1m61 is annotated as a monomer. The Phyre2 program built the model on the template c2ozoA with a confidence level of 100% over 87% of the sequence (Figure 5.16a). A model constructed using the Modeller 9.10 program had a customized Python shell (Appendix 5B) and the model obtained is shown in Figure 5.16 (b).

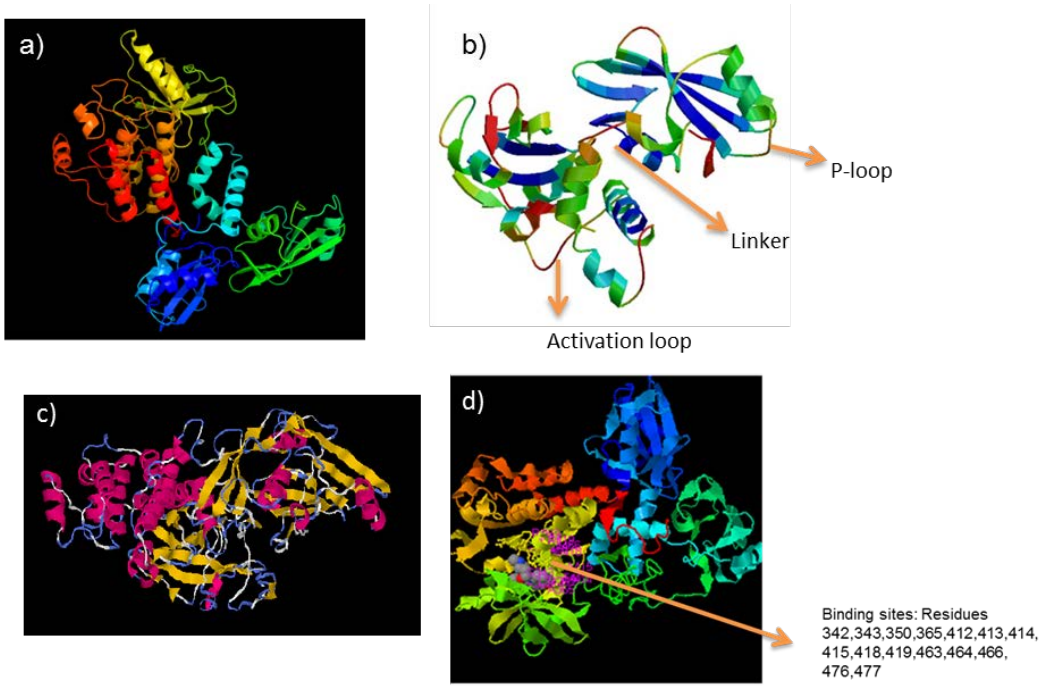


Figure 5.16. Homology models and binding site predictions of ZAP-70 molecule in *M. eugenii* (a) Phyre2 model, (b) Modweb model, (c) I-TASSER model, (d) Binding site prediction by 3D ligand prediction server (buried binding sites).

The 3D ligand prediction server was used to predict exposed binding sites. 20 exposed residues were revealed in contrast to the 15 predicted buried residues (Fig. 5.17). A binding site prediction carried out with I-TASSER revealed a slightly different number of residues involved in the ligand binding to that predicted by the 3D ligand binding

prediction server (diagram not shown). Residues involved in ligand binding according to the I-TASSER prediction were amino acid residues 342, 344, 345, 347, 350, 365, 367, 397, 412, 413, 414, 415, 416, 419, 463, 464, 466, 476, 477 and 480.

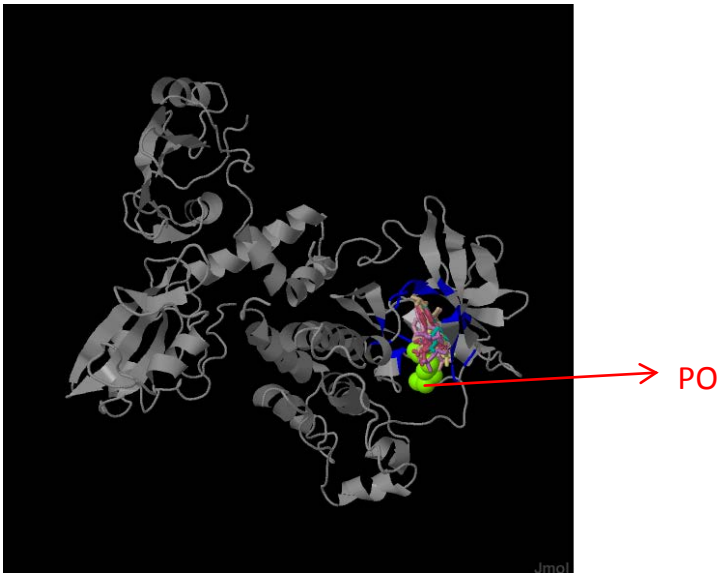


Figure 5.17. Exposed ligand binding sites in *M. eugenii* ZAP-70 predicted by 3D ligand prediction server.

5.3.3.9 ZAP-70 - Western Blot analysis

The human anti-ZAP-70 antibody was unsuccessful in detecting the marsupial ZAP-70 protein in *M. eugenii* thymus and in *O. fraenata* spleen, thymus, liver and lung tissues (Fig. 5.18).

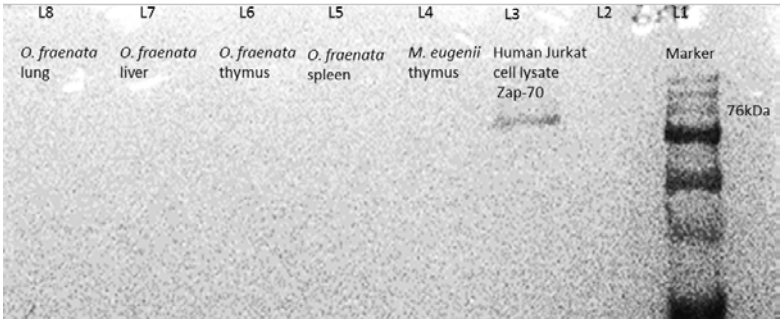


Figure 5.18. Western Blot analysis of human anti-ZAP-70 antibody in *O. fraenata* tissues and *M. eugenii* thymus tissue.

5.3.4 Lymphocyte specific kinase (Lck)

5.3.4.1 Lck - Homology

A BLAST homology search showed that the *O. fraenata* sequence had a 90% identity to the predicted *M. domestica* sequence at the nucleotide level and a 95% identity at the putative amino acid level. The *M. eugenii* partial Lck sequence had a 90% identity to *M. domestica* at the nucleotide level and 95% identity at the amino acid level. These and other homology percentages and their respective e-values are shown in Table 5.18. The high identity levels indicated a conserved function of this molecule across mammalian taxa.

Table 5.18. Homology search results for the *O. fraenata* and *M. eugenii* Lck nucleotide and amino acid sequences, their identities and their respective e-values.The values are calculated over the 1528bp/503aa length of the partial *O. fraenata* sequence.

Homology search results for *M. eugenii* and *O. fraenata* Lck								
	O. fraenata				*M. eugenii*			
Species	nt	e-value	aa	e-value	nt	e-value	aa	e-value
M.domestica	91%	0.0	96%	0.0	90%	0.0	95%	0.0
N. leucogenys	----	----			85%	8e.138	----	----
E. caballus	81%	0.0	91%	0.0	83%	0.0	92%	0.0
P. toglodytes	80%	0.0			83%	0.0	----	----
Hylobates	82%	0.0	92%	0.0	82%	0.0	92%	0.0
A. nancymaae	82%	0.0	----	----	----	----	92%	0.0
C .familiaris	81%	0.0	93%	0.0	82%	0.0	92%	0.0
S.scrofa	81%	0.0	92%	0.0	81%	0.0	91%	0.0
H.sapiens	82%	0.0	91%	0.0	81%	0.0	92%	0.0
S. sciureus	81%	0.0	90%	0.0	81%	0.0	----	----
A.melanoleuca	82%	0.0	93%	0.0	81%	0.0	92%	0.0
C. griseus	81%	0.0	----	----	81%	0.0	91%	0.0
L. africana	81%	0.0	93%	0.0	80%	0.0	92%	0.0
C. porcellus	81%	0.0	91%	0.0	80%	0.0	91%	0.0
R. norvegicus	81%	0.0	91%	0.0	82%	0.0	91%	0.0
B. taurus	81%	0.0	91%	0.0	81%	0.0	90%	0.0
O. aries	81%	0.0	92%	0.0	81%	0.0	91%	0.0
O. cuniculus	81%	0.0	92%	0.0	81%	0.0	91%	0.0
M. musculus	81%	0.0	92%	0.0	80%	0.0	91%	0.0
G. gallus	77%	0.0	91%	0.0	79%	0.0	84%	0.0
O. mykiss	74%	0.0	81%	0.0	75%	0.0	74%	0.0
S. salar	74%	0.0	70%	0.0	75%	0.0	74%	0.0
A.carolinensis	75%	0.0	71%	0.0	74%	2e.133	----	----
X. silurana	72%	0.0	76%	0.0	73%	1e.173	79%	0.0
C. auratus	71%	0.0	68%	0.0	72%	9e.150	74%	0.0
H. hippoglussus	----	----	72%	0.0	72%	2e-152	73%	0.0
S. maximus	71%	0.0	69%	0.0	72%	1e-153	74%	0.0

nt = nucleotide, aa = amino acid, ---- = not recognized by BLAST.

5.3.4.2 Lck - Domain structure

Three principal domains (SH_3, SH_2 and tyrosine kinase) were identified in the *O. fraenata* and *M. eugenii* Lck sequences. The identified domain structures showed high e- values ranging from 1.39e-35 to 2.03e-17 indicating a high probability that these domains occur within the sequences. The locations and respective e-values are shown in Table 5.19.

Table 5.19. Predicted domain structures, locations, and e-values for the *O. fraenata* and *M. eugenii* Lck sequences.

Domain structures in *O.fraenata* and *M. eugenii* Lck sequences				
	O. fraenata		*M. eugenii*	
Domain	Location	e-value	Location	e-value
SH_3	43-99	2.03e-17	64-120	1.56e-18
SH_2	104-194	1.39e-35	125-215	1.39e-35
TyrKc	224-473	2.20e-134	245-494	1.42e-128

An SH_4 domain (unique domain) was identified in *O. fraenata* by comparing its domains with the published Lck sequences for *C. auratus langsdorfii (Araki et al., 2007), O. aries* (Yu *et al.,* 2010) and *H. sapiens* (Perlmutter *et al.,* 1988).

An N-terminal signal peptide was not identified in the coding domain of the *O. fraenata* Lck molecule because the program identified the molecule as being soluble.

5.3.4.3 Lck - Glycosylation and glycation sites

One predicted O-linked glycosylation site was found in the *O. fraenata* Lck coding domain at position 226 with a confidence value of 55%. The *M. eugenii* partial sequence also contained a predicted O-linked glycosylation site at position 205 with a 55% probability. Two putative N-linked glycosylation sites were found in both the *O. fraenata* and the *M. eugenii* Lck sequences. In *O. fraenata,* one predicted N-linked glycosylation site was located at position 40 with a probability of 72.3%, and the second site was at position 321 and had a 62% probability. In *M. eugenii*, one N-linked glycosylation site was located at position 19 with a probability of 76.36% and the second site was at position 300 and had a probability of 64%. The sequences and locations are shown in Table 5.20.

Table 5.20. Predicted N-linked glycosylation sites their sequence, location, probability, and signal strength in the *O. fraenata* and *M. eugenii* Lck sequences.

Putative N-linked glycosylation sites in *O. fraenata* and *M. eugenii* Lck sequences.					
O. fraenata			*M. eugenii*		
Sequence/location	Probability	Signal	Sequence/location	Probability	Signal
NGSD (40)	72.32%	++	NGSD (19)	76.36%	+++
NGSL (321)	62%	+	NGSL (300)	64%	+

Predicted glycation sites were found in both *O. fraenata* and *M. eugenii* indicating the presence of covalent bonding sites to sugar moieties. The location of the putative glycated lysines are shown with their respective confidence levels in Table 5.21. Validation of this result requires sophisticated mass spectrometry or Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOFF) analysis which was outside the scope of this study.

Table 5.21. Predicted glycated lysine residues, their locations, and confidence levels in *O. fraenata* and *M.eugenii* Lck sequences.

Locations and confidence levels of glycated lysine residues			
O. fraenata		*M. eugenii*	
Location	Confidence	Location	Confidence
73	82.7%	52	82.5%
84	89%	63	88.9%
118	93.8%	97	94%
182	93.8%	161	94%
269	79%	248	84%
276	92.5%	255	92.7%
340	91%	319	91.5%

5.3.4.4 Lck - Phosphorylation sites

Predicted phosphorylated amino acid residues were found in the *O. fraenata* and *M. eugenii* Lck sequences. Thirteen serine, five threonine and seven tyrosine phosphorylation sites were predicted in the *O. fraenata* Lck sequence by the EXPASY prediction program. A similar distribution was found in the partial *M. eugenii* Lck sequence with the exception that there were six predicted threonine and eight predicted tyrosine phosphorylation sites. The confidence levels of the predictions ranged from 50% to 98% in the *O. fraenata* Lck sequence and 54% to 99% in the *M. eugenii* Lck sequence. Their locations and the predicted confidence levels are shown in Table 5.22.

Table 5.22. Predicted phosphorylation sites and their confidence levels for Lck sequences in *O. fraenata* and *M. eugenii.*

Predicted phosphorylation sites in Lck of *O. fraenata* and *M. eugenii*						
Species	Serine Positions	Confidence levels	Threonine Positions	Confidence levels	Tyrosine Positions	Confidence levels
O. fraenata	7	96.1%	159	66.5%	25	50.1%
	42	97.4%	226	95.9%	110	68.4%
	121	91.2%	244	98.4%	192	54.1%
	133	99.8%	395	94.4%	313	97.3%
	156	86.3%	418	93%	318	56.4%
	164	58.8%			360	75.6%
	166	99.7%			394	85.4%
	194	83.2%				
	221	63.9%				
	274	98.4%				
	279	89.1%				
	281	90.8%				
	323	61.7%				
M. eugenii	21	89.7%	138	66.5%	89	68.4%
	50	90.9%	205	95.9%	171	54.1%
	100	91.2%	223	98.4%	292	97.3%
	112	99.8%	374	94.4%	297	56.4%
	135	86.3%	397	93%	339	75.6%
	143	58.8%	424	69.6%	373	85.4%
	145	99.7%			449	95%
	173	83.2%			484	68%
	200	63.9%				
	253	98.4%				
	258	89.1%				
	260	90.8%				
	302	61.7%				

5.3.4.5 Lck - Disulphide bonds

The *O. fraenata* Lck sequence contained nine cysteines of which four can build predicted disulphide bridges. In the *O. fraenata* Lck sequence, four predicted disulphide bridges were detected compared to three in the partial Lck sequence of *M. eugenii*. The sequences and locations of these predicted bridges are shown in Table 5.23. The software predicted a disulphide bond located at position 217-224, however the construction of the sequence suggested that this might be an intrachain disulphide bond.

Table 5.23. Predicted disulphide bonds and their locations and sequences in the *O. fraenata* and *M. eugenii* Lck sequences.

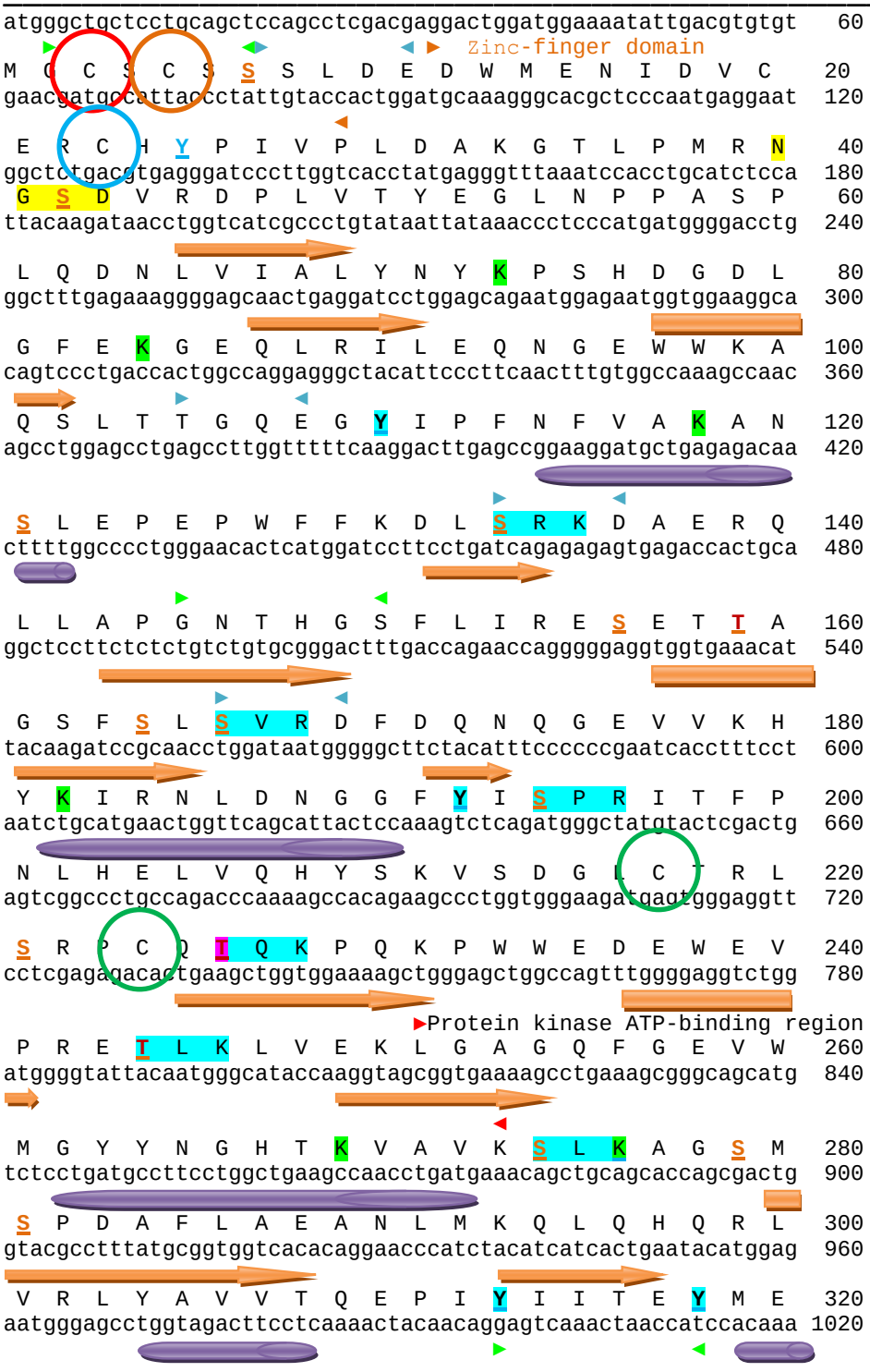
Predicted disulphide bridges in *O. fraenata* and *M. eugenii* Lck sequences			
O. fraenata		*M. eugenii*	
Sequence	Location	Sequence	Location
XXXMGCSCSSS - VRPDNCPEELY	3-465	XXXXRCHYPIV - KLMMLCWKERP	2-455
XMGCSCSSSLD - KPMMLCWKERP	5-476	VSDGLCTRLSR - SDTLNCKIADF	196-357
DVCERCHYPIV - SDTLNCKIADF	23-378	RLSRPCQTQKP - VRPDNCPEELY	203-444
VSDGLCTRLSR - RLSRPCQTQKP	217-224		

5.3.4.6 Lck - Primary sequence and secondary structure prediction

The *O. fraenata* Lck coding domain consisted of 1,528bp which translated into 509 amino acids compared to the human Lck sequence that consisted of 1,530 bp and translated into 510 amino acids (Fig. 5.19). For *M. eugenii,* a 1,467bp long partial nucleotide sequence was amplified. This translated into 488 amino acids to the stop codon and a further 426bp in the 3' untranslated region to a polyA tail (raw data in Appendix 5C) (Fig. 5.20).

The distribution of the alpha helices and beta strands differ between the two species over the same areas. The *O. fraenata* secondary structure prediction indicated that it had two more beta strands than *M. eugenii*. This may indicate a different fold pattern since the beta strands are more rigid than the alpha helices thus rendering the molecule less flexible.

In the *M. eugenii* Lck a portion of the 3' end which was amplified appeared to have a polyA tail. However upon closer inspection no polyadenylation sequence was observed, hence this may not have actually been the polyA tail.



Cont. page 242

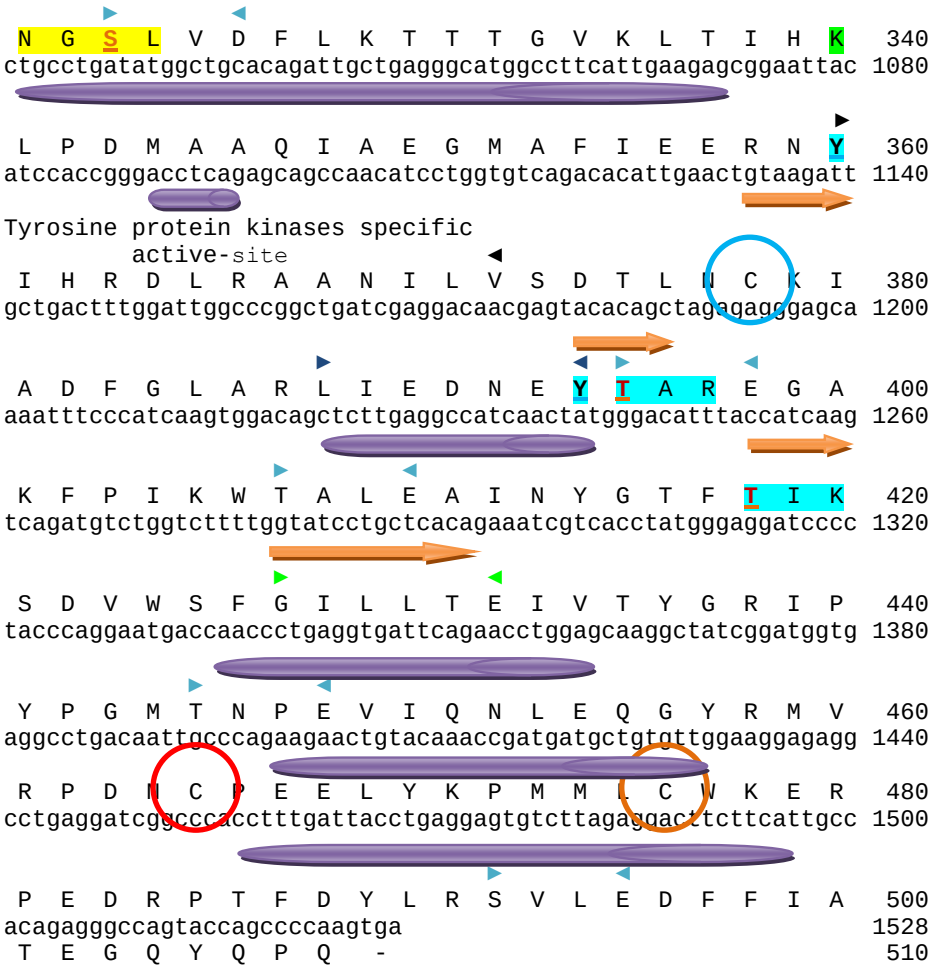
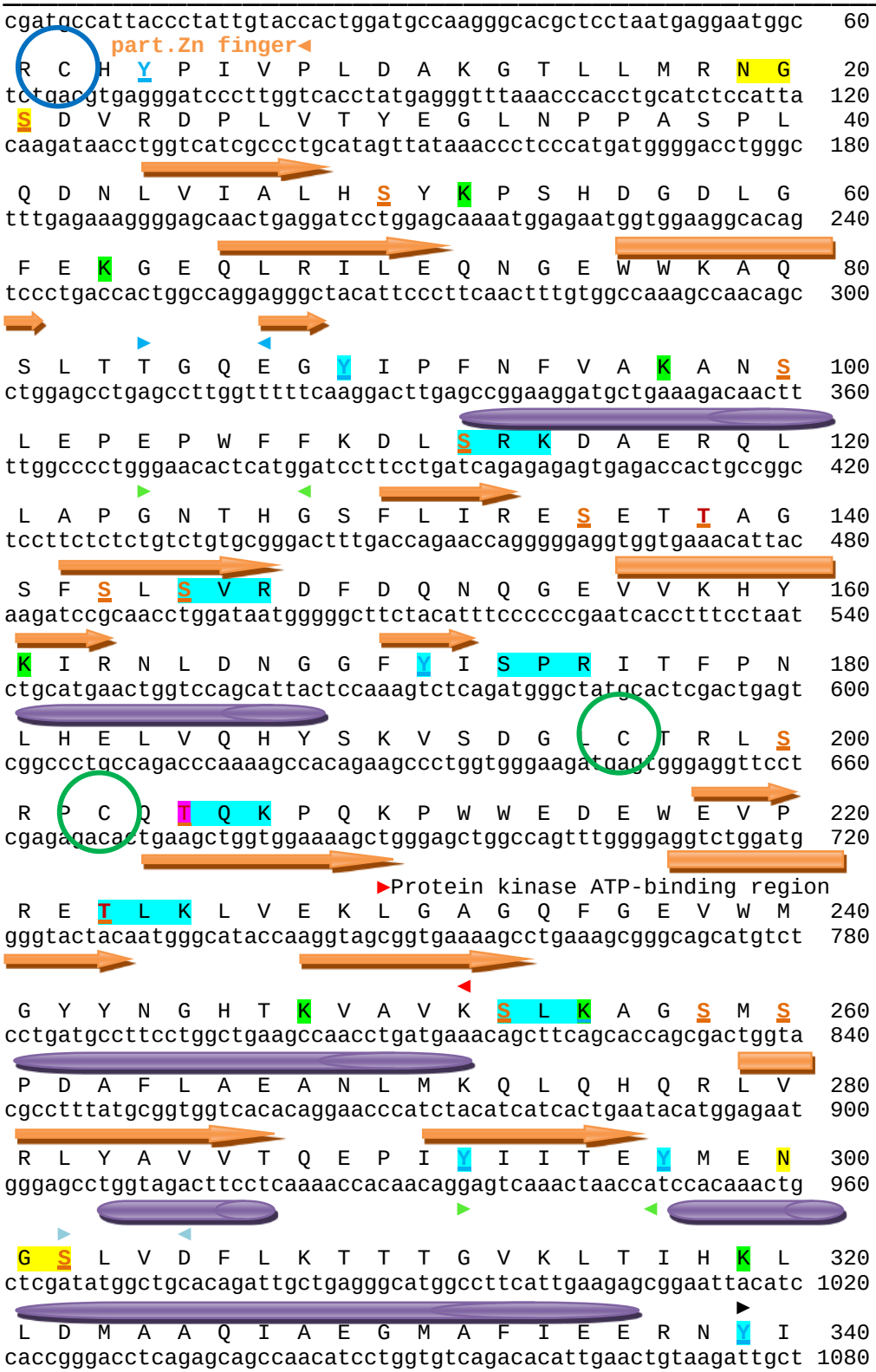


Figure 5.19. *O. fraenata* Lck primary sequence and secondary structure prediction. ▶◀ = Myristoylation sites. ▶◀ = Protein kinase ATP- binding region. ▶◀= Tyrosine protein kinases specific active sites. ▶◀ = Zinc finger domain. ▶◀ = Tyrosine kinase phosphorylation site. ▶◀ = Casein kinase II phosphorylation site. Highlighted in yellow = N-linked glycosylation sites. K = glycation sites. Protein kinase C phosphorylation sites. T = threonine phosphorylation sites. Y = tyrosines phosphorylation sites. Different colours = disulphide bonds. = Transmembrane helix, = Strand.



Cont. page 244

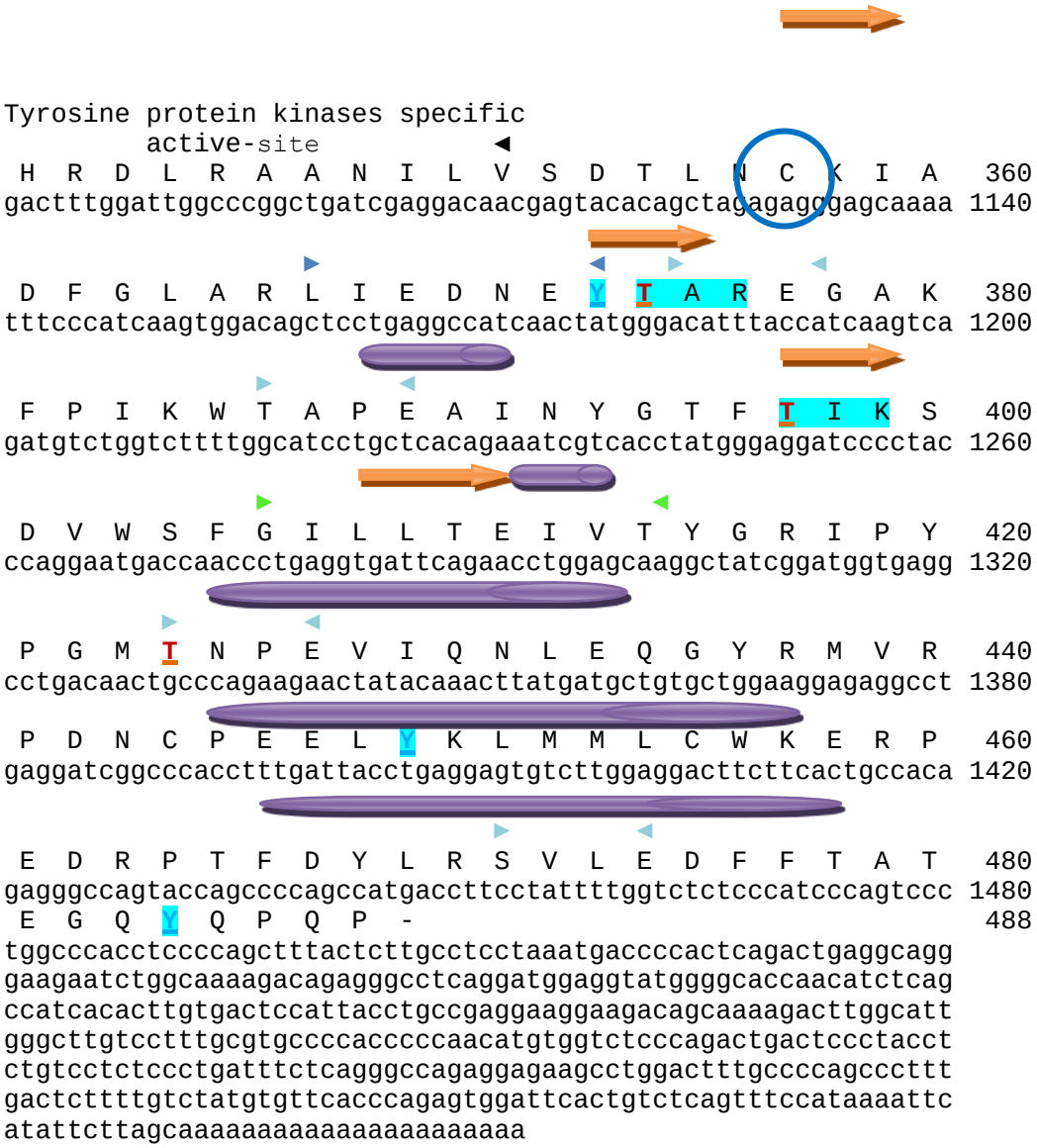


Figure 5.20. *M. eugenii* partial primary Lck sequence and secondary structrure prediction. ▶◀ = Myristoylation sites. ▶◀ = Protein kinase ATP- binding region. ▶◀= Tyrosine protein kinases specific active sites. ▶◀ = Zinc finger domain. ▶◀ = Tyrosine kinase phosphorylation site. ▶◀ = Casein kinase II phosphorylation site. Highlighted in yellow = N-linked glycosylation sites. K = glycation sites. Protein kinase C phosphorylation sites. T = threonine phosphorylation sites. Y = tyrosines phosphorylation sites. Different colours = disulphide bonds. = Transmembrane helix, = Strand.

5.3.4.7 Lck - Phylogenetic analysis

The Neighbor-Joining phylogenetic tree inferred the evolutionary history of Lck and indicated that Lck is highly conserved in mammals. The optimal tree showed the sum of the branch length was 2.01 (Fig. 5.21). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. The tree is drawn to scale with the branch length showing the same units as those of the evolutionary distances used to infer the tree. The evolutionary distances are in the units of the number of amino acid substitutions per site. This analysis involved 31 amino acid sequences and all positions containing gaps and missing data were eliminated. There were a total of 57 positions in the final dataset. A clear clustering of the marsupial sequences with those of other mammals was observed. There is a distinct separation between the macropods within the marsupial clade. The clustering of the taxa was intact despite the fact that only a partial sequence of the *M. eugenii* Lck molecule was included in the phylogenetic analysis.

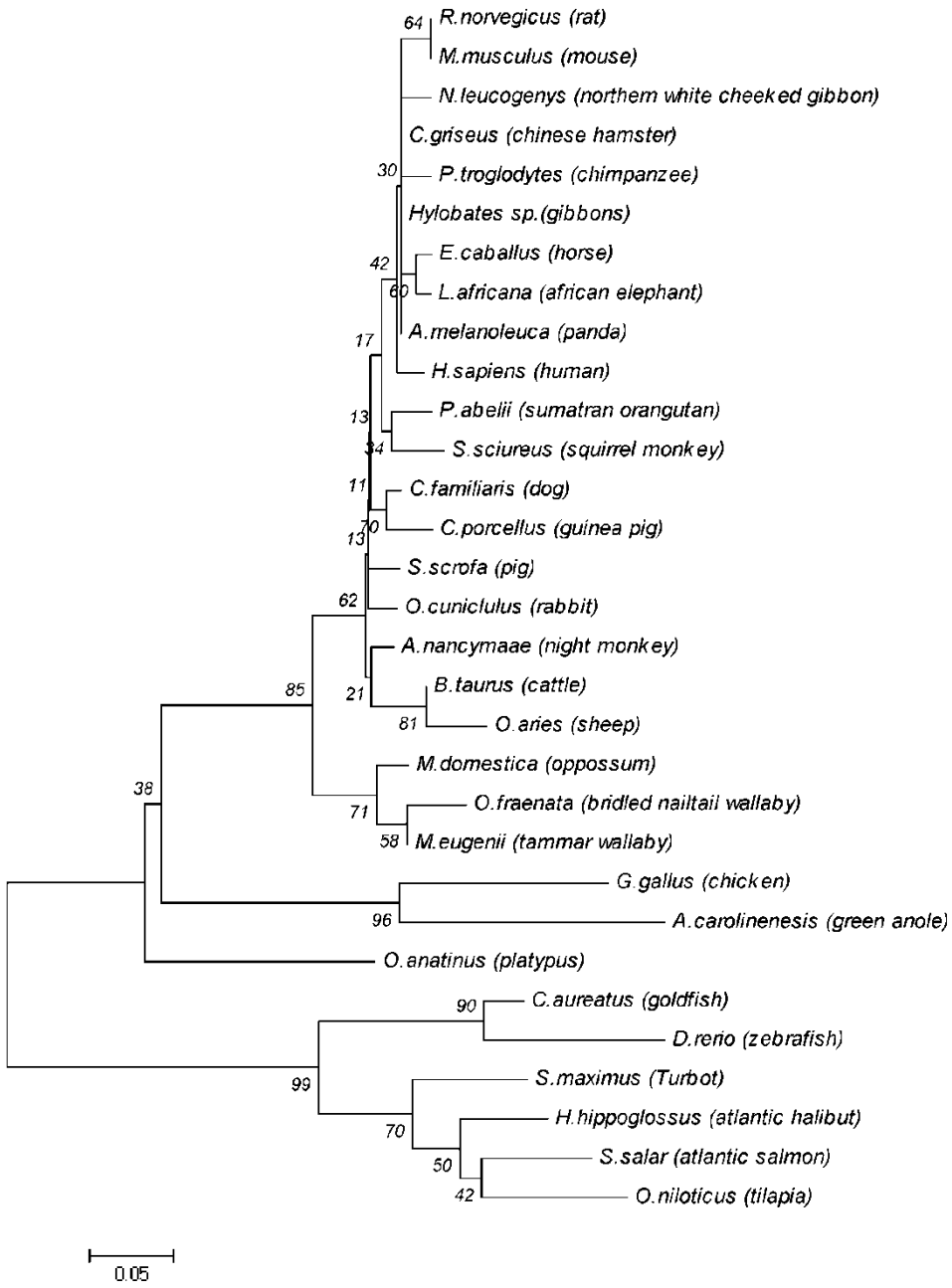


Figure 5.21. Phylogenetic tree for Lck compiled using the Neighbor-Joining method based on the Poisson correction model.

5.3.4.8 Lck - Structure modelling

Homology modelling of Lck carried out with different programs showed that the closest template was the pdb structure 2ofu_1. RaptorX did not model the full open reading frame of the *O. fraenata* Lck sequence but modelled the individual domains (data not shown). The Phyre2 program modelled 85% of the *O. fraenata* putative protein sequence with more than a 90% confidence value. The interactive program Modeller 9.10 used the template 1qpc to generate a model for the *O. fraenata* Lck sequence. The different structural models are shown in Fig. 5.22.

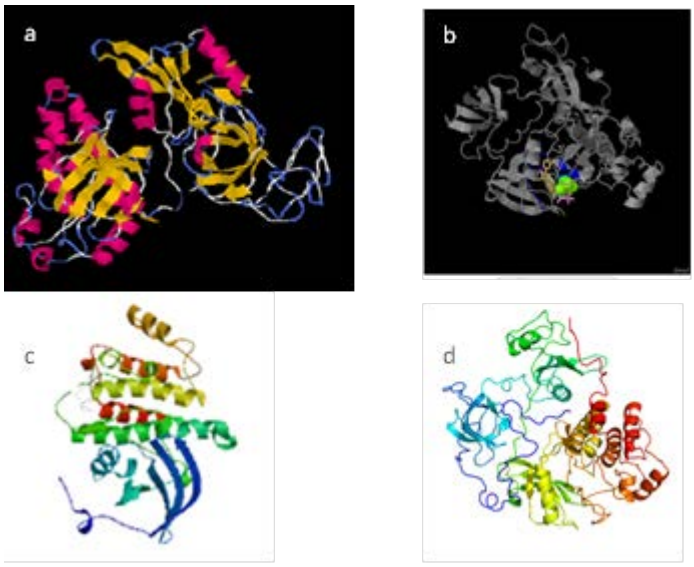


Figure 5.22. Structure models of Lck in *O. fraenta* (a) I-Tasser model (b) ligand binding capacity model (c), Swiss Model model (d) Modeller 9.10 interactive model.

Sixteen predicted amino acid residues involved in ligand binding and their locations were identified in the *O. fraenata* Lck sequence with the 3DLigand binding site prediction server (http.//www.sbg.bio.ic.ac.uk/3dligandsite/3dligand). These are shown in Table 5.24.

Table 5.24. Predicted ligand binding sites and their locations within the *O. fraenata* Lck sequence.

Predicted ligand binding residues of *O. fraenata* Lck			
Residue	Amino Acid	Residue	Amino Acid
251	Leucine	319	Methionine
259	Valine	322	Glycine
271	Alanine	323	Serine
273	Lysine	368	Alanine
301	Valine	369	Asparagine
316	Threonine	371	Leucine
317	Glutamine	382	Aspartic acid
318	Tyrosine	385	Leucine

It can be seen from the statistics in Fig. 5.23a that the structure prediction indicated a close relationship to the template selected. The Z-score slider indicated that the model predicted for the *O. fraenata* Lck is close to Lck structures that were solved by others using X-ray crystallography (Fig. 5.23a,b). A density plot based on the QMEAN score of all reference models indicated the location of the *O. fraenata* Lck model (Fig. 5.23a). The number of reference models used in the calculation is given at the bottom of the plot (Fig. 5.23c).

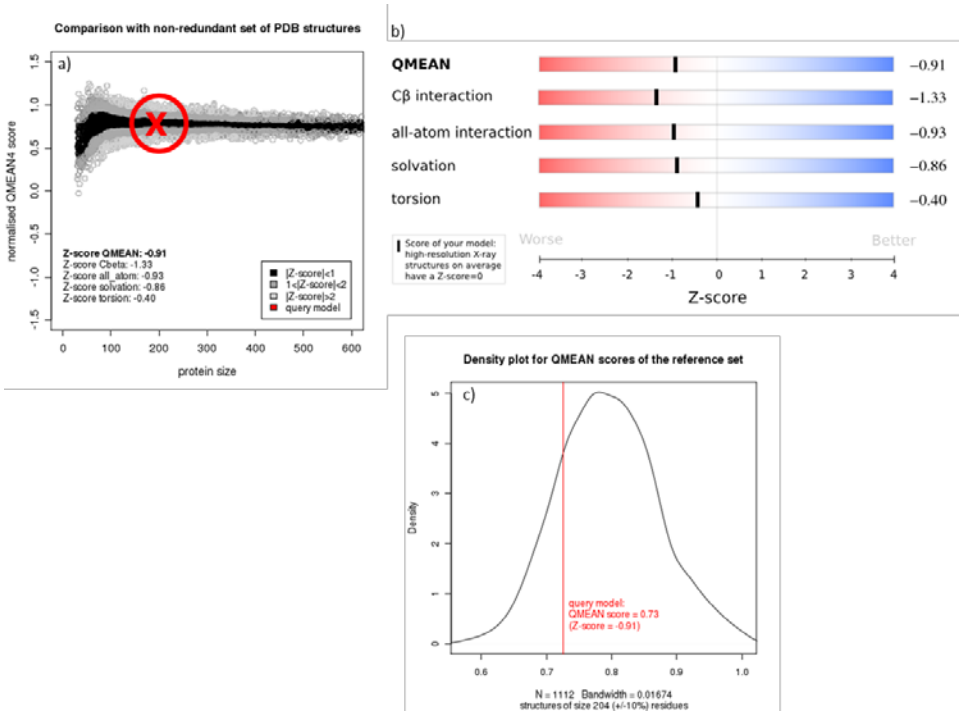


Figure 5.23. Statistical evaluation of *O. fraenata* Lck model prediction (a) location of *O. fraenata* homology model compared with structures in pdb, (b) Z-score slider indicating quality of the model, (c) Bell curve for QMEAN.

5.3.4.9 Lck - Expression studies

The gene expression measured against the housekeeping gene GAPDH indicated that the Lck molecule is highly expressed in all tissues which were investigated. The image of the 2% agarose gel revealed a high intensity of expression of the Lck gene in spleen and thymus tissue. This may indicate that the Lck gene is upregulated in all tissues investigated (Fig. 5.24).

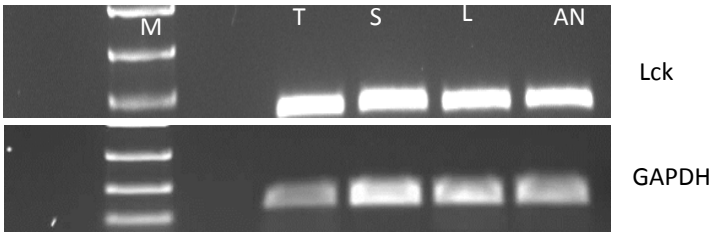


Figure 5.24. Gel image showing expression of *O. fraenata* Lck across four tissue types. T = thymus, S = spleen, L = lung, AN = axial node, M = marker.

5.3.4.10 Lck - Western Blot analysis

The selected N-terminal anti–human Lck antibody did not detect the Lck protein in the spleen, axial node, gut node, thymus, liver and lung tissues from *O. fraenata* and thymus tissue from *M. eugenii* (Fig. 5.25).

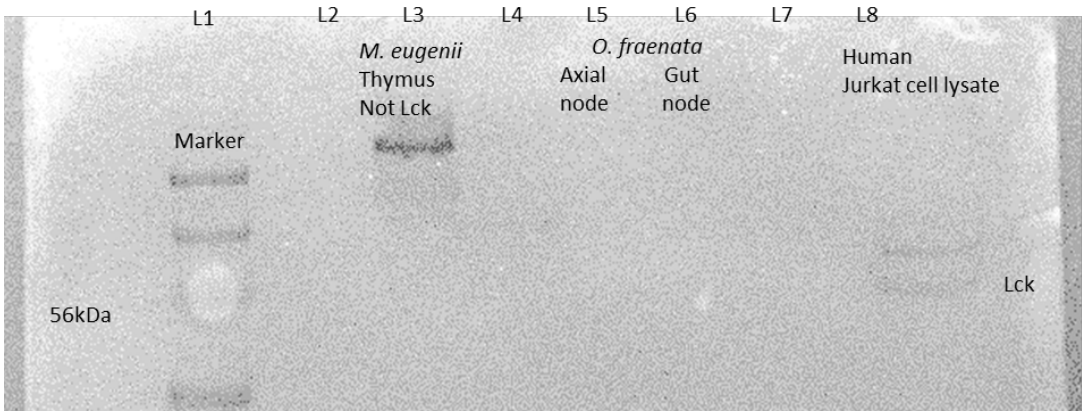


Figure 5.25. Western Blot image for Lck protein in spleen, axial node, gut node, thymus, liver and lung tissues from *O. fraenata* and thymus tissue from *M. eugenii.*

5.4 Discussion

The TCRζ chain is one of the main signalling units in the T cell signalling cascade. The *M. eugenii* molecule is not annotated in the ensembl database. This is the first time that the TCRζ sequences of *M. eugenii* and *O. fraenata* have been characterized. While the identity percentage of the putative amino acid residues is high among mammals (Table 5.2) there were differences in the structurally important motifs of the TCRζ chains in the two marsupial species when compared to humans and mice. In this study, a glycine rich region similar to that found in humans was found in the TCRζ in the two marsupials. It was also found that the marsupial glycine region had a different topology when compared to humans and mice (Fig. 5.26). Two additional amino acid residues, asparagine (N) and glutamine (Q), were found in the marsupial sequences. While the human and rodent sequences have the consensus sequence G(129)XXXXGKGXXGX XXG(143) (Peter *et al.,*

1992), the corresponding marsupial sequence was found to be G(130)XXXXXXGKGXXXXXXXG(146).

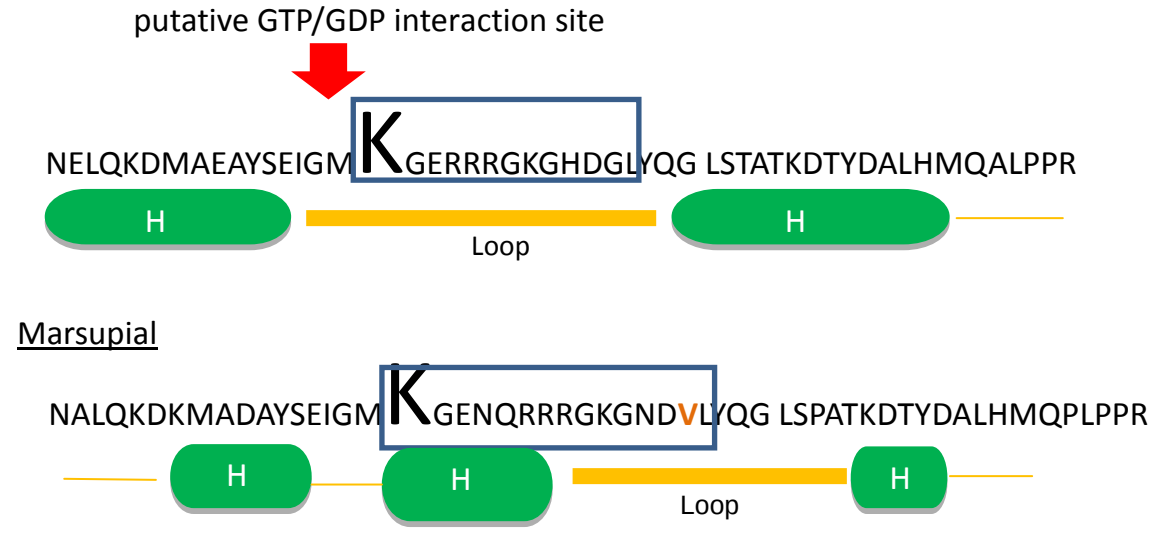


Figure 5.26. Structural differences in the cytoplasmic tail of the marsupial TCRζ chain compared with human and rodents.

The conserved arginine motif in the glycine region is responsible for the maintenance of the tertiary structure of the molecule (Borders *et al.,* 1994). The marsupial arginine motif is preceded by two additional amino acid residues similar to that found in other mammals. However, the human and rodent sequences were devoid of the asparagine and glutamine residues in the glycine region. Another interesting finding was the substitution of one of the glycine residues (G^{128}) with valine in the marsupial sequence. There are reports of mutation of G^{134} to valine which in humans reduces antigen induced IL-2 production and phosphorylation of the zeta chain (Frank *et al.,* 1990). It would therefore appear that this glycine is functionally important. However, no studies have been undertaken to determine whether or not the substitution of glycine to valine in marsupials has similar effects.

This study found that the *O. fraenata* TCRζ amino terminus resembles a leader sequence with the signal peptidase cleavage site located between residues 21 and 22, yielding a

mature peptide of 143 amino acid residues which is consistent with other known TCRζ sequences (Appendix 5A). However, in this study it was found that the putative amino acid sequence for the *M. eugenii* TCRζ did not show the presence of a signal peptide. The TCRζ chain is a single-pass membrane protein with an internal signal sequence into the endoplasmic reticulum membrane. The hydrophobic residues in the domains that traverse the plasma membrane and the low occurrence of these types of transmembrane proteins define the ability of the cell to transduce a signal or take part in cell to cell contact (Bledi *et al.,* 2003). The leader sequence is cleaved off. The motifs of the mature protein are conserved in marsupials.

A single cysteine at position 32 of the putative *O. fraenata* amino acid sequence was present on the external face of the transmembrane region. This indicated that all disulphide interactions between the zeta chain and other CD3 polypeptide chains occur through this residue. The *M. eugenii* sequence has a single cysteine residue at position 19 due to the absence of the leader sequence and all disulphide interactions therefore occur through this residue. It was found in this study that the transmembrane region DPR$\underline{L^{1}CYLLDGILF^{10}IYGVIITALF^{20}LRA}$ KFSKIAK of the marsupial TCRζ chain contained two aspartic acid residues. These play an important part in TCR assembly and in the dimerization of the molecule as well as in the formation of a disulphide bridge. This was established by others in experiments carried out in the human TCR ζ chain (Call *et al.,* 2006, Rutledge *et al.,* 1992). Due to the conservation of these residues in marsupials it was concluded that the aspartic acid residues had the same function as in humans. It is the negative charge of these asparagine residues that are of importance as well as their location in the transmembrane domain (Weissman *et al.,* 1988). The residues involved in the dimerization of the marsupial TCRζ chain were found to be C^{2}, D^{6}, L^{9}, Y^{12}, L^{16}, T^{17} and F^{20} which are the same residues responsible for the dimerization of the human molecule. Residue V^{23} (numbering from motif only) in humans is also involved in this process but the marsupial TCRζ chain has a substitution to alanine in that position. Three of the residues are polar and mutational analyses in humans have identified D^{6}, Y^{12} and T^{17} as the most important residues in the dimerization process (Call *et al.,* 2006, Rutledge *et al.,* 1992).

The glycine residue in the second part of the motif (GVXXT) drives the homodimerization of this molecule (Bolliger and Johansson, 1999). Residue D^6 contacts the TCRα chain at the arginine residue and stabilizes the TCR/CD3 complex (Bolliger and Johansson, 1999). It appeared that these residues are conserved in *O. fraenata* and *M. eugenii* (Appendix 5A) and therefore may have the same function. The ability of the residues, especially the aspartic acid residues, to assist in the dimerization of the molecule is dependent on their positions within the helix. It was found in the present study that the marsupial models show the locations of the two aspartic residues at similar positions to those in the human molecule.

The extracellular domain of TCRζ consists of nine amino acid residues across all mammalian sequences. Mutational studies in humans have highlighted the importance of this short region by reducing the extracellular domain to two residues with the result that the receptor assembly was inhibited (Johansson *et al.,* 1999). The conservation of this domain across marsupials therefore indicated a similar function. It was further reported that architectural changes occurred in human T lymphocytes which had been stimulated for an unstated but lengthy period of time (La Gruta *et al.,* 2004, Osono *et al.,* 1997). The characterization of TCRζ in *M. eugenii* was carried out on cDNA from stimulated lymphocytes and the sequences appeared to have conventional domain structures as shown in Table 5.15. The stimulation time point for the *M. eugenii* lymphocytes was 48 h. This was not regarded as 'lengthy', and a comparison with the work of La Gruta *et al.* (2004) and Osono *et al.* (1997) may not therefore be valid. Further stimulation assays may have to be conducted to determine if the changes in the domain structure of the marsupial TCRζ chain molecules do in fact occur.

In humans, the binding of Lck to the TCRζ via the phosphorylation event on the ITAM recruits additional signalling molecules such as ZAP-70 via the Src homology 2 (SH_2) domains (Straus and Weiss, 1993, Wange *et al.,* 1992, Wange *et al.,* 1993). The activation of ZAP-70 is possible through the interaction between ZAP-70, Lck and the TCRζ-CD4-CD8 complex (Chan *et al.,* 1995, Neumeister *et al.,* 1995). Due to the signalling function of Lck

it was not surprising to find that the marsupial Lck had a greater than 90% identity to other mammalian Lck molecules, thus indicating that this gene is highly conserved and serves an important function.

A domain structure search conducted in this study indicated that the marsupial Lck molecule, like its human counterpart (Pawson *et al.,* 2001) contained SH_2, SH_3, and SH_4 domains and a tyrosine kinase domain (Table 5.15) (supporting data in Appendix 5C). In *O. fraenata* and *M. eugenii* the structure of the SH_2 domain encompasses a central anti-parallel β-sheet sandwiched between two alpha helices. This is a module of approximately 100 amino acids that bind specific phospho-tyrosine motifs. This study also found that the hydrophobic core and the invariant arginine residue that coordinates the phosphate oxygen of the phosphorylated tyrosine were conserved in the *O. fraenata* Lck molecule. These findings support the work of Pawson *et al.* (2001) in the human SH_2 domain. It was also reported by Pawson *et al.* (2001) that the SH_3 domains in humans contain five-antiparallel beta strands which are packed to form two perpendicular beta sheets. In this study both the SH_3 and the SH_2 domains were found to be conserved in the marsupial sequences.

In the human SH_2 domain a hydrophobic patch consisting of a cluster of conserved aromatic residues resides in the SH_3 domain (Nguyen *et al.,* 1998). This feature was also identified in the *O. fraenata* sequence in the present study. SH_3 domains recognize RKXXPXXP or PXXPXR motifs and these motifs together with the two proline residues that bind within the two hydrophobic pockets of the SH_3 domain were also identified in the *O. fraenata* Lck molecule. SH_4 has a consensus sequence of MGC (methionine, glycine, cysteine) which is located at the N-terminal end, and is responsible for promoting membrane binding and plasma membrane targeting (Liang *et al.,* 2004). In the case of Lck, the SH_4 domain is essential for membrane localization and function which is dependent on the myristoylation of the glycine residue in position two of the Lck N- terminal sequence. Human studies have shown that the serine at position six is responsible for the localization of the Lck molecule (Yasuda *et al.,* 2000). In the present study this residue was

identified in the *O. fraenata* and *M. eugenii* sequences and therefore it is proposed that the function of this motif is conserved in those species.

Lck not only interacts with ZAP-70 but also interacts with CD4. This interaction occurs between the short cytoplasmic domain of CD4 and the NH_2-terminal domain of Lck through the sequence motifs CXC and CXXC. These motifs were not only identified in the *O. fraenata* and *M. eugenii* sequences in Lck, but were also found to be located in their CD8 sequences indicating the interaction between these two molecules also occurs in marsupials. The NH_2 terminal is located in the unique domain of Lck and holds the CXXC motif representing the site of signal transduction (Isakov and Biesinger, 2000). The dissociation of Lck from the cytoplasmic tail of CD4 occurs during T cell activation. This induces serine phosphorylation of CD4, internalization and degradation of the Lck molecule (Pelchen-Matthews *et al.,* 1992). It was found in the present study that these binding sites were conserved in all vertebrates with the exception of the teleost fish (supporting data in Appendix 5C). It was found in this study that the Lck sequence of *O. fraenata* also contained this binding site and it is concluded that the interaction of the marsupial CD4 and CD8 molecule with Lck is conserved.

It was also found in this study that in *O. fraenata,* and in the partial sequence from *M. eugenii*, the catalytic domain HRDLRAAN was conserved. In human studies it has been determined that RDL accommodates the binding of protein phosphatase 2 (PP2-an Src family selective tyrosine kinase inhibitor) (Zhu *et al.,* 1999). A portion of this catalytic domain (DLRAAN) was also found in the *M. eugenii* ZAP-70 sequence indicating the interaction of the two molecules was conserved. The composition of the hydrophobic pocket of the marsupial Lck appears to be unique to the Src kinase family as is the case in other mammals (supporting data in Appendix 5C). Two lysine (K) residues which interact with adenosinetriphosphate (ATP) in the Lck molecule were found to be conserved in both *O. fraenata* and *M. eugenii* (supporting data in Appendix 5C).

A major event in the T cell signalling cascade is the activation of ZAP-70 which occurs after the Lck has bound the CD4 or CD8 TCR co-receptors. In order to draw any conclusions as

to the sophistication of the marsupial adaptive immune system it was therefore important to investigate the main motifs for any differences in sequence and secondary structure. Prior to this study, the sequence of ZAP-70 in marsupials was unknown although a partial annotation exists in the ensembl database. A comparison of the ensembl sequence and the *M. eugenii* sequence expressed in this study revealed that there were significant differences between the two (supporting data are in Appendix 5B). The expressed *M. eugenii* sequence differed in important sequence motifs such as the loop motifs and in the interdomain B region to those in the annotated sequence.

Like other characterized mammalian ZAP-70 molecules, ZAP-70 in *M. eugenii* interacts with the CD3 chains and is comprised of two SH_2 domains (SH_2-N and SH_2-C) that bind the phosphorylated ITAMs on TCRζ. This phosphorylation event allows for the recruitment of ZAP-70 to the TCR via the SH_2 domains consequently phosphorylating the tyrosine residues in ZAP-70 which affects the catalytic activation of the molecule. This was deduced by reference to experiments carried out in humans where it was found that the cytoplasmic ZAP-70 molecule rapidly translocated the TCR-CD3 complex upon ligand binding in preparation for signal transduction (Wange *et al.,* 1993, Chan *et al.,* 1992).

In this study a comparison of the *M. eugenii* ZAP-70 with the human sequence of ZAP-70 revealed similarities in the P-loop and the functional motifs contained therein. The glycine loop or P-loop was identified in the *M. eugenii* sequence. This lies between the β1 and β2 strands that are important for localizing the phosphates of ATP, substrate binding, catalysis and regulation of kinase activity. The P-loop has a consensus sequence of GxGxϕG (G = glycine, x represents any amino acid, and ϕ is either a phenylalanine or a tyrosine) and is conserved in the *O. fraenata*, *M. eugenii, L. hirsutus* and *M. domestica* ZAP-70 sequences. In humans, this loop forms the upper side of the ATP-binding cleft and because this loop is located in the same area in the marsupial sequence and with the same topology a similar function is indicated. It was found in this study that the F^{349} residue in the four marsupial sequences is conserved. Human studies have shown that this residue has a side chain that is folded back into the binding cleft (Jin *et al.,* 2004a).

This appears to be same for the marsupial ZAP-70 sequence due to conservation of the topography. The glycine residues increases the flexibility of this structure which enables it to adopt different conformations to accommodate different inhibitors (Jin *et al.,* 2004a). The sequence VELGAGNF (V = valine, E = glutamic acid, L = leucine, G = glycine, A = alanine, N = asparagine, F = phenylalanine) of the marsupial P-loop is conserved except for the first amino acid. In marsupials, amphibians, fish and birds the first amino acid of this motif is a valine whereas in all other vertebrates, except for *L. Africana,* this is substituted with isoleucine. However, no reports have indicated any functional impairment due to this substitution and with the biochemical activity being similar between the two residues it is highly unlikely that the function is compromised in marsupials.

The negative regulator in the linker region (Y^{288}) is located between the carboxyl terminal SH_2 and the catalytic domains and is conserved across species including the marsupials. The phosphorylation of residue Y^{292} in humans down-regulates antigen receptor function through mechanisms independent of ZAP-70 activity (Ahmed *et al.,* 2005). Phosphorylation of Y^{292} may induce a conformational change in ZAP-70 to stabilize its enzymatic activity following antigen receptor cross-linking while phosphorylation on Y^{319} is important for the interferon-stimulated ERK activation (Ahmed *et al.,* 2005). A species alignment of ZAP-70 shows that the four marsupial sequences all have the same biochemical environment around the negative regulatory site, and it was therefore concluded that the function of this molecule is conserved.

The lysine (K^{369}, human numbering) is responsible for transferring phosphate groups and was found to be conserved in marsupial ZAP-70. The methionine (M^{411}, human numbering) which is the gatekeeper in the gatekeeper pocket were shown to be conserved across all species, including marsupials, which indicated that both the lysine and the methionine are functionally important residues.

Human functional studies have shown that the function of ZAP-70 is regulated by phosphorylation on multiple tyrosine residues (Di Bartolo *et al.,* 1999, Chan *et al.,* 1992) including the conserved tyrosine (Y^{493} human numbering) located in the catalytic loop.

The phosphorylation of ZAP-70 on Y^{493} is essential for antigen receptor mediated activation of calcium and the ras pathways. ZAP-70 is first phosphorylated by Lck at Y^{493} which up-regulates the catalytic activity of ZAP-70. This residue was shown to be conserved in the four marsupial sequences thus indicating a conservation of function.

The motif S^{520} DVWS^{524} (human numbering) is present in all phosphotyrosine kinases and especially the amino acid residue S^{520} and is important in the enzymatic activity and function of ZAP-70 in the signalling cascade (Yang *et al.,* 2003b, Qian and Weiss, 1997). The extended sequence for this motif in humans is FRKFSSRS^{520}DVW<u>S</u>^{524}YGV. It was shown in this study that the *M. eugenii* sequence YRKFSCQS^{520}DVWTYGF and *the O. fraenata* sequence YRKFSSRS^{520}DVW<u>S</u>YGV differ in their amino acid compositions and are both different to the human sequence. The *O. fraenata* sequence is closer to the human sequence, by differing only in one amino acid residue, than to that of the other macropod *M. eugenii*. Human studies have shown that D^{521} stabilizes the catalytic loop by hydrogen bonding to the arginine residue in the catalytic loop (H<u>R</u>DLAARN) (Taylor *et al.,* 1995). In humans, the catalytic loop determines the substrate specificity of the protein kinase (Taylor *et al.,* 1995). This appeared to be the same for the marsupials since the catalytic loop sequence, including the residue S^{520}, is 100% conserved.

In humans, the first two residues of the <u>SD</u>VWS motif are responsible for the activation of NFκB and the IL-2 promoter (Yang *et al.,* 2003b). The present study did not investigate the functionality of this motif in marsupials, however the sequence is conserved thus indicating it has a similar function.

The DLAARN motif is 100% conserved among vertebrates. The amino acid residues in this motif are responsible for generating a signal that induces thymocyte development (Wiest *et al.,* 1997). Human studies have shown that mutations in this motif halt thymocyte development at the $CD4^+CD8^+$ stage (Wiest *et al.,* 1997). In the present study it was found that the DLAARN motif is 100% conserved in the four marsupials thus indicating that the function of this motif is also conserved.

The N-terminal human ZAP-70 antibody did not detect the marsupial ZAP-70 molecule in the *O. fraenata* spleen, axial node, gut node, thymus, liver and lung tissues, because the sequence conservation in this area of the putative marsupial amino acid sequence was only 60%.

5.5 Conclusion

Marsupial genomes have been published only for *M. eugenii* and *M. domestica*. The TCRζ chain, one of the main signalling units in the T cell signalling cascade, has not been annotated for those two species in ensembl. In this study, TCRζ was characterized in *O. fraenata, M. eugenii* and *M. domestica*, ZAP-70 was characterized in *O. fraenata, M. eugenii, L. hirsutus* and *M. domestica*, and Lck was characterized in *O. fraenata* and *M. eugenii*. This is the first time these signalling molecules have been characterized in those marsupial species.

The predicted marsupial TCRζ, ZAP-70 and Lck structures determined in this study were compared to the molecular structures of other mammals resolved by others using X-ray crystallography. It was found that the amino acid sequences were highly conserved in different areas of the molecules and the structures appeared to be similar to those of other mammals. It was demonstrated that the signalling molecules TCRζ, ZAP-70 and Lck have homologies well above 70% to the same signalling molecules in other mammals. This study showed that the SH_2, SH_3 and tyrosine kinase domains were present in the TCRζ chain of *O. fraenata* and *M. eugenii*. These domains were also found in the ZAP-70 sequences of *O. fraenata, M. eugenii, L. hirsutus* (in part) and *M. domestica* (in part). The SH_4 domain in Lck was detected in *O. fraenata* and *M. eugenii* (in part). The triple ITAM motif found in other mammalian TCRζ chains was found in the *O. fraenata* and *M. eugenii* TCRζ molecules. The catalytic domain in ZAP-70, which is important in the enzymatic function of the molecule, was shown to be conserved in the *O. fraenata* and *M. eugenii* ZAP-70 sequences. The classic CxxC motif was identified in the *O. fraenata* and *M. eugenii* Lck molecules and it was therefore concluded that the attachment of Lck to the co-receptors CD4 and CD8 also occurs in those species.

The functions of many of the motifs discussed in this chapter are known in humans and other mammals. Because those motifs have now been shown to be conserved in the *O. fraenata* and *M. eugenii* immune molecules it is concluded that their functionality is also conserved. Functional studies were outside the scope of this study. Future work should therefore include functional studies to confirm that these marsupial immune molecules have the same functional capabilities as in other mammals. It would also be useful to investigate whether or not the calcium ion influx upon ZAP-70 activation also occurs in marsupials. This event in humans stimulates the Linker of activated T lymphocytes and therefore boosts the T cell signal.

Chapter 6

Cytokines – Interleukin-2, Interleukin-17 and the regulatory T cell surface marker

Chapter 6 - Cytokines – Interleukin-2 (IL-2), Interleukin-17 (IL-17) and the Foxp3 transcription factor

6.0 Abstract

The cytokine IL-2 is the principal T cell growth factor and, prior to this study, had been identified in various vertebrates but not in marsupials. The full open reading frame of the *T. vulpecula* IL-2 sequence was characterized in this study and was published by Young *et al.* (2011). The full *M. eugenii* IL-2 sequence was also identified in this study for the first time but has yet to be published.

The mRNA derived IL-2 sequences, including the 5' and 3' ends of the molecule, were identified in both *T. vulpecula* and *M. eugenii* using RACE-PCR. The IL-2 sequences were investigated for polymorphisms, using RT-PCR, in both *T. vulpecula* and *M. eugenii*. The promoter region of the *M. eugenii* IL-2 gene was identified by data mining the published wallaby genome in the ensembl database. A marsupial-specific antibody was designed by predicting the secondary structure and finding a hydrophobic stretch using the Kyte Doolittle and Hopp- Woods plots. A Dot Blot was used to determine the binding of the *M. eugenii* IL-2 antibody to its peptide while a Western Blot was used to determine that the *M. eugenii* IL-2 antibody binds the IL-2 protein in a cell lysate made from thymus. IL-17 and Foxp3 were also identified in *M. eugenii* using RT-PCR. Structure prediction tools were used to elucidate a probable tertiary structure and the ligand binding properties of the two cytokines were investigated using bioinformatics tools.

It was found that the *T. vulpecula* IL-2 gene sequence was devoid of both synonymous and non-synonymous substitutions. In contrast, the M*. eugenii* IL-2 sequence contained both types of substitutions. It appeared that the TATA box, the NFκB, AP-1 and Oct-1 binding sites were present in the *M. eugenii* IL-2 gene. The marsupial-specific antibody for *M. eugenii* was successfully tested in a tissue lysate. The cytokine IL-17 was characterized, and it identified the T_h17 cell sub-population and the specific knot motif characteristic for

the IL-17 cytokine family in *M. eugenii*. The T_{reg} sub-population was also identified in *M. eugenii* and *O. fraenata* by characterizing the Foxp3 molecule.

6.1 Introduction

Interleukin-2 promotes T cell proliferation, activation, and differentiation, and is produced mainly by an activated T_h1 cell population (Weaver *et al.,* 2007, Smith, 1988a, Morgan *et al.,* 1976). The activation of IL-2 producing $CD4^+$ T lymphocytes occurs in response to TCR stimulation (Bendiksen and Rekvig, 2004, Nelson *et al.,* 1996). IL-2 can be produced by a number of events such as activation of dendritic cells by gram negative bacteria (Granucci *et al.,* 2003), activation of natural killer (NK) cells (Yu *et al.,* 2000) and immunoglobulin synthesis by B lymphocytes (Gold and DeFranco, 1994). IL-2 has been identified in a number of vertebrates (Table 6.1, Appendix 6A) and activity analogous to IL-2 has been detected in lower vertebrates such as *Spalerosophis diadema* (Egyptian or diadem rat snake) (El-Ridi *et al.,* 1986). Once the cytokine is produced, the magnitude of the response is regulated at the genetic level, however IL-2 is difficult to detect due to the low level of expression in resting T lymphocytes (Margeta-Mitrovic, 2002). A review of the literature revealed that this cytokine can only be detected in cells that were either stimulated with a mitogen, such as ConcanavalinA (ConA) or Phytohaemagglutinin (PHA) (Bird *et al*., 2005b), or in cells where the TCR complex interacted with an antigen/MHC complex (Smith, 1988b).

IL-2 is implicated in a multitude of functions in the immune system and it also plays an important role in providing homeostasis for regulatory T lymphocytes (T_{reg}) (Fehérvari and Sakaguchi, 2004). Regulatory T lymphocytes are characterized by a high expression of the forkhead box protein 3 (Foxp3) gene which can up- or down-regulate around 700 genes (Campbell and Ziegler, 2007). Foxp3 is a multifunctional protein acting either as a transcriptional repressor by inactivating the transcription of the cytokine IL-2, or as a transcriptional activator (Moon *et al.,* 2009, Chen *et al.,* 2006, Wu *et al.,* 2006). It is thought that Foxp3 is essential for the development of the T_{reg} population as immune

cells, and is involved in the molecular mechanism which controls immune tolerance (Moon *et al.,* 2009, Mantel *et al.,* 2006).

The differentiation of T_{regs} is tightly linked to the development of IL-17 producing T_h17 cells. The T_h17 cell population was not identified in humans until 2007 (Beriou *et al.,* 2009, Stockinger and Veldhoen, 2007). The *M. eugenii* sequence for IL-17 is annotated in ensembl. In this study, the mRNA derived expressed sequence of the *M. eugenii* IL-17 was obtained and characterized for important functional motifs.

$CD4^+IL\text{-}17^+$ cells have been implicated in allograft rejection and several autoimmune diseases (Kappel *et al.,* 2009). Given the fact that marsupials present with a difference in skin graft rejection response time, it was considered that an investigation of the *M. eugenii* sequence would reveal important differences when compared to other mammalian IL-17 sequences. IL-17 is produced in response to IL-23 which is induced, for example, by *M. tuberculosis* infected cells (Lockhart *et al.,* 2006). The evidence that marsupials suffer from mycobacterial infections meant that the structural components of IL-17 in *M. eugenii* warranted investigation.

This study investigated the Foxp3 gene in *M. eugenii* and *O. fraenata*, the cytokine IL-17 in *M. eugenii,* and identified the protein of IL-2 in *M. eugenii*.

6.2 Aims and Objectives.

The aims of the study were to determine whether or not the marsupial immune system overall is as sophisticated as that of other mammals by identifying important regulatory molecules and, in particular:

- To characterize the mRNA derived IL-2 molecule and to design a custom made marsupial IL-2 antibody to verify that the mRNA product was translated into a protein.

- To characterize the expressed sequence of IL-17 and identify the cysteine knot family motif of the IL-17 cytokine family.
- To demonstrate that T_{reg} cells were also present in the repertoire of the marsupial immune system by characterizing the cell surface marker Foxp3 that identifies a T cell as belonging to the T_{reg} cell sub-population.

6.3 Specific Materials and Methods

The following materials and methods were unique to this chapter. General materials and methods are detailed in Chapter 2.

6.3.1 RNA, mRNA and cDNA

RNA and RACE DNA from *M. eugenii* PHA stimulated lymphocytes and RACE DNA from *T. vulpecula* stimulated lymphocytes were donated by Dr. L. J. Young. cDNA was synthesized from the RNA of *M. eugenii* PHA stimulated lymphocytes as described in Chapter 2.

RNA and mRNA were isolated from spleen, gut node, thymus and liver, and Foxp3 was isolated from spleen and thymus from *O. fraenata*.

6.3.2 Primer design for Interleukin-2 (IL-2), Interleukin-17 (IL-17) and Forkhead box P3 (Foxp3).

Primers were designed as detailed in Chapter 2, section 2.1.2. The primer sequences used are presented in Table 6.1.

Table 6.1. Primer sequences employed to deduce the open reading frames of IL-2, IL-17, the partial open reading frame of Foxp3 in *M. eugenii,* and primers for the deduction of potential polymorphisms in *M. eugenii* and *T. vulpecula* IL-2.

Species/ Gene	Forward primer 3' – 5'	Reverse primer 3' – 5'	T_m	Size	Primer ID
IL-2					
M. eugenii	ctcctgtybtgcrtbgcactaactcttg (Exon-1) 3'RACE	gctgtcaacgatacgctacgtaacg (3'RACE primer) cgctacgtaacggcatgacagtg (3' Nested RACE primer)	61.6°C/76°C 72°C	~700bp	*2FW$_c$-5*
	cgactggagcacgaggacactga (5'RACE primer) ggacactgacatggactgaaggagta (5'Nested RACE primer)	catgttagtgcatgaatctttggcag ac (Exon-4) 5'RACE	74°C/58°C 78°C	550bp	*2TR-R*
	ccactctctaatcatctacccagag (Exon-1) 3'UTR	cagttaggatcataagatctattac (Exon-4) 5' UTR	55°C/61°C	607bp	*IL2-UTF/* IL2-UTR*
	atgaacaaggtcccgctcttgtcctg (Exon-1)	ttaagatccttcaatcctcatcttgg (Exon-4)	57°C/51°C	420bp	*IL2START/ IL2STOP*
	cccgtctcttgtcctgtattg (Exon-1)	ctacgtgacttaatggaggtcc (Exon-2)	54.6°C/54.6°C	150bp	*IL-2fexp/** IL-2rexp*
T. vulpecula	attccactctctaatcactactcag (5' UTR)	actgctgttggtcttagtcgtc (3'UTR))	53.4°C/56.8°C	650bp	*P2UTRF/ P2UTR**
O. fraenata	ccaccactgtgctgcagtacttactacg	cgatgactgctggtattacttgggat gtag	63°C/61°C	122bp	*HRM2TF/ HRM2TR****
IL-17					
M.eugenii	atgtcttctctgggcaacttgccaggg (Exon-1)	tcaggacactgtgcgtggggtcacac (Exon-3)	64°C/64°C	460bp	*T17F/T17R*
Foxp3					
M. eugenii	ggccygghtgkgaraaggtcttc	gatctcrttgagkgtccgctgyttctc	60°C/62°C	~500bp	*FoxF/FoxR*
	gagaaacagcggacactcaatgagatc 3' RACE	gctgtcaacgatacgctacgtaacg	60°C/76°C	~1200bp	*TFORKF*

* designed by Dr. L.J. Young and used to deduce IL-2 polymorphisms in *M. eugenii* and *T. vulpecula.* ** used for expression studies. ***used in qPCR.

6.3.3 *Trichosurus vulpecula* PHA stimulated lymphocytes.

IL-2 was identified in PHA stimulated lymphocytes of *T. vulpecula*. The animal was taken from the wild and euthanized with appropriate approval from the Animal Ethics Committee of Landcare Research, Lincoln in accordance with the 1987 Animals Protection Regulation of New Zealand (Approval No. 10/02/01). Blood and spleen suspensions were prepared from homogenized tissues of the retropharyngeal lymph nodes and the mesenteric lymph node, and were used in a stimulation assay carried out by New Zealand collaborators. Control and stimulated cells were harvested for molecular studies at the 24 h and 72 h time points, shipped to Australia and processed in the Young laboratory.

6.3.4 *Macropus eugenii* stimulated lymphocytes.

The stimulation assay was performed by Dr. L. J. Young and the donated cell pellet was used in this study. RNA and mRNA were isolated from the cell pellet, as described in Chapter 2, sections 2.1.3 – 2.1.3.3, from which cDNA was subsequently synthesized as outlined in Chapter 2, section 2.1.5.

6.3.5 Polymerase chain reaction, cloning and sequencing.

The PCR conditions used, and the concentration of the PCR mixes employed, are stated in Chapter 2, section 2.1.8. Table 6.2 lists the PCR templates used. A detailed description of the templates can be found in Chapter 2, section 2.1.8.1.

Table 6.2. RT-PCR and RACE-PCR templates used for amplification of IL-2, IL-17 and Foxp3.

Species	Gene of interest	PCR template	RACE PCR template
M. eugenii	IL-2	Nos. 2 and3	No. 2
T. vulpecula	IL-2	No. 2	---
M. eugenii	IL-17	No. 3	---
M. eugenii	Foxp3	Nos. 1, 2 and 3	No. 2

RACE primers were designed from initial sequences derived from the RT-PCR products, and RACE PCRs were performed to obtain the 5’ and 3’ ends of the molecule as described in Chapter 2, section 2.1.9.

6.3.6 Investigation of polymorphisms in *T. vulpecula* and *M. eugenii* IL-2 sequences.

Primers were designed in the untranslated (UTR) regions of the IL-2 gene in *M. eugenii* and *T. vulpecula*. Initial PCR’s were run for 35 cycles but to ensure minimal mispriming the annealing cycles were reduced to the point of clear visibility of the products on a 0.8% agarose gel matrix. This meant that 29 cycles for the *M. eugenii* and 30 cycles for *T. vulpecula* were necessary. Three PCR’s (Table 6.3) were performed for the IL-2 molecules from each species. The resultant products were visualized on a 0.8% agarose gel and subsequently purified as outlined in Chapter 2, section 2.1.8.3. The purified cDNA was

inserted into the pCR™ 4TOPO® vector (Invitrogen, Carlsbad, USA) for cloning as described in Chapter 2, section 2.1.10.

Table 6.3. RT-PCR and RACE-PCR conditions for polymorphism investigation.

Species	Primers	Cycling conditions	Product size
M. eugenii	IL-2UTF/IL-2UTR	94°C– 2min 94°C– 30s 55°C– 50s } 29x; 30x; 35x 68°C– 55s 68°C– 5min	600 bp
T. vulpecula	P2UTF/P2UTR	94°C– 2min 94°C– 30s 50°C– 55s } 30x; 35x 68°C– 50s 68°C– 5min	650 bp

The products were excised under minimal exposure to UV light in order to reduce any UV induced alteration of DNA. Ten clones of *M. eugenii* IL-2 and *T. vulpecula* IL-2 were sent for sequencing in both directions to AGRF as outlined in Chapter 2, section 2.1.10.6. These were subsequently investigated for possible polymorphisms (alignments of IL-2 sequences and highlighted polymorphic sequences are shown in Appendix 6A).

6.3.7 Genomic DNA isolation and amplification

A genomic DNA isolation was carried out as detailed in Chapter 2, section 2.4. An amplification of the IL-2 gene was carried out using primers positioned in exon-3 and in the 3' UTR end of the IL-2 gene. Polymerase chain reaction conditions were 94°C denaturation for 30 s, 94°C for 10 s, 55°C for 30 s, 65° for 4 min 15 s (repeated for 35 cycles), and an extension at 65°C for 10 min. A special enzyme that is able to amplify longer products (Long-Amp Taq, Biolabs New England, Australia) was used for the amplification of IL-2 in genomic DNA. This was carried out to confirm the shorter *M. eugenii* IL-2 sequence by investigating the sequence downstream from exon-3. The obtained gDNA was investigated with Genscan as described in Chapter 2, section 2.4.2 to determine the exon/intron boundaries.

6.3.7.1 Gel electrophoresis for gDNA

Gel electrophoresis was carried out on a 0.8% agarose gel containing 0.3 µg/mL of ethidium bromide and run for 1.5 h at 100 V in a Bio-rad gel electrophoresis tank. A Lambda DNA/HindIII marker (Fermentas, Thermofisher, Australia), which was vortexed prior to use, was employed. The marker was prepared by adding 1 µL of DNA marker (0.5 µg concentration) together with 1 µL of 6 X DNA Loading dye and 4 µL of deionized water. The mixture was heated for 5 min at 65°C and then cooled on ice for 3 min. This was then used in a 0.3 µg/mL ethidium bromide gel for visualization of the genomic DNA.

6.3.8 Real Time PCR (qPCR) for IL-2

Real Tim PCR was performed as described in Chapter 2, section 2.7 - 2.7.1.1. The primer pair used for the qPCR experiment was HRM2TF and HRM2TR. These primers spanned a 122 bp product and were verified in a RT-PCR where they yielded the correct product for *M. eugenii* prior to use in qPCR. The product of the one step qPCR was visualized on a 1% gel matrix excised and processed as outlined in Chapter 2, section 2.1.8.2 and sent for sequencing to identify the amplicon as outlined in Chapter 2, section 2.1.10.6.

A High Resolution Melt analysis approach was used and a HRM kit from Qiagen (Doncaster, Vic, Australia) was employed (Chapter 2, section 2.7.2). A Melt analysis was carried out using the Rotorgene Gene Q HRM Software which characterized the double-stranded PCR products based on their dissociation (melt) behavior.

6.3.9 Annotation of the *M. eugenii* IL-2 promoter region

The IL-2 promoter region sequence was extracted from the ensembl database. The 5' region amplified from the RACE-DNA by RACE-PCR was investigated for the presence of any functional motifs such as the start site (atg), and then used in the ensembl BLAST function to determine the scaffolds on which the *M. eugenii* IL-2 gene was located. It was noted that the IL-2 gene was fractionated and different segments of the gene were

located on different scaffolds. Once the IL-2 sequence was located, the TATA box was localized by searching approximately 52 bp upstream from the start site. When the TATA box was located, the search for the promoter region was extended by searching 1,000 bp at a time upstream from the TATA box. This included a search for important motifs such as the NFAT, Oct1, NFκB, and AP-1 binding sites. The identified sequences were investigated with Promoter Scan to identify the threshold percentages for valid promoter motifs.

6.3.10 Peptide design

A custom designed antibody was used to show that the expressed gene sequence translated into a protein. The peptide to produce the antibody was designed using the PSIpred bioinformatics tool. This was used to predict the secondary structure as described in Chapter 2, section 2.2.6.1. The Kyte Doolittle and Hopp- Woods plots were used to identify a region suitable for an antigenic peptide. Areas that indicated a buried helix were avoided and hydrophilic residues were preferred. Cysteine, methionine, multiple serine, and proline residues were avoided, where possible, because of their negative effect on peptide purification.

The electrochemical property of the chosen amino acid sequence was determined using Prot Param from the EXPASY suite of programs. A Kyte Doolittle plot was employed to visualize the hydrophilicity over the length of the chosen peptide sequence. A moving 'window' determined the summed hydropathy at each point in the sequence (this represented the Y coordinate), and these were plotted against their respective positions (X coordinate). The Kyte-Doolittle scale indicated hydrophobic amino acids while the Hopp-Woods scale measured hydrophilic residues. While the 'window' used in the Kyte-Doolittle scale is between 19 and 21 amino acid residues in length, the Hopp-Woods scale used a 'window' of between 5 and 7 amino acid residues in length. The optimal length of the peptide is between 10 and 15 amino acid residues. Although longer peptides are preferable, the number of possible epitopes increases and the synthesis becomes more

difficult. Problems can arise in the purification and coupling to carrier proteins if the peptide is too long.

The theoretical isoelectric point was calculated at 4.43 and a hydrophilicity plot was established using the Innovagen website (http.//www.innovagen.se/custom-peptide-synthesis/peptide-property-calculator/peptide-property-calculator.asp).

The area chosen was a unique stretch of sequence not affected by any post-translational modifications and was not important for the functionality of the molecule. The aim was to find a peptide that generated an antibody that would react specifically with its target protein. The sequence of the designed peptide was sent to Sapphire Bioscience (http.//www.sapphirebioscience.com) for synthesis. The peptide was synthesized in conjunction with ProSci Incorporated (Poway, CA, USA - www.prosci-inc.com) with an N-terminal cysteine.

The antibody produced was a polyclonal antibody, although preference is normally given to a monoclonal antibody to avoid indiscreet binding to non-target proteins. In this case the polyclonal antibody was preferred since it had a better chance of recognizing the target protein in a crude cell lysate. An antibody titer was carried out by the manufacturer of the antibody. A lower percentage of sodium azide (0.05%) was added by the manufacturer to the antibody primarily for preservation during the journey and to stop any interference with target protein detection. The target animal was rabbit, and two rabbits were used for the antigen production. The antibody was affinity purified by the manufacturer and diluted by the user to 1.0mg/mL using 1 X PBS.

6.3.11 Western Blot

Western Blot analyses were carried out as described in Chapter 2, section 2.8.4. The isotype control was carried out beforehand in a Dot Blot (Chapter 2, section 2.8.4.2) as well as by Western Blot to prove that the antibodies did not recognize each other. The *M. eugenii* customized IL-2 antibody (mpIL-2) lyophilized powder was mixed with PBS to

make a stock solution with a concentration of 0.2 mg/mL. Decimal dilutions of the stock solution were prepared and then tested against the peptide.

6.3.12 Phylogeny

A phylogenetic analysis using Mega5 was undertaken for all genes after the full coding domains were determined. The method is detailed in Chapter 2, section 2.3.

6.3.13 Bioinformatics

All of the bioinformatics tools detailed in Chapter 2 were used to elucidate putative domain structures, glycosylation sites, disulphide bonds, amino acid sequences, as well as to predict secondary and tertiary structures, and to conduct homology searches.

6.4 Results

6.4.1 Interleukin-2

The *M. eugenii* IL-2 sequence was characterized in this study for the first time, while the gene sequence for IL-2 in *T. vulpecula* was elucidated by a collaborative effort between the Young laboratory and a New Zealand research team (Young *et al.,* 2011). Both molecules were investigated for post-translational modifications and important structural motifs as well as for gene polymorphisms. Structural comparisons were also made between the *M. eugenii* and the human IL-2 molecules.

6.4.1.1 IL-2 - Homology

A homology search conducted with the BLAST algorithm found that the *M. eugenii* IL-2 had an 87% identity to *T. vulpecula* and a 76% identity to the predicted sequence of *M. domestica* at the nucleotide level. At the amino acid level the identity percentage was 78% to *T. vulpecula* and 63% to *M. domestica*. The identity percentages fell sharply to 38% and 37% when compared with *M. monax* and *S. scrofa* respectively. The lowest identity percentage was detected in *M. musculus* with 28%. The nucleotide and amino

acid identity percentages are listed in Table 6.4. Many of the species only had identities to the *M. eugenii* IL-2 at the amino acid level while they were not recognized at the nucleotide level. The identity percentages at the amino acid levels were, on average, around 35% which was expected since IL-2 is species and tissue specific. A similar outcome was observed for *T. vulpecula.* The IL-2 sequence of this marsupial had an identity of 73% to the predicted sequence of *M. domestica* with an e-value of 8e-63 at the nucleotide level, while at the amino acid level the sequence identity was 63% with an e-value of 8e-50.

Table 6.4. Homology search results for the *M. eugenii* and *T. vulpecula* IL-2 nucleotide and amino acid sequences, their identities and their respective e-values.

Homology search result for IL-2 in *M. eugenii* and *T. vulpecula*								
.	*M. eugenii*				*T. vulpecula*			
Species	Nucleotide	e-value	Amino acid	e-value	Nucleotide	e-value	Amino acid	e-value
T. vulpecula	87%	4e-142	78%	9e-69	100%	0.0	100%	4e-87
M. domestica	76%	3e-80	63%	1e-58	73%	8e-63	63%	8e-50
S.scrofa	------	------	37%	1e-18	------	------	36%	4e-15
S. caffer	79%	3e-80	32%	3e-11	79%	0.049	34%	3e-10
O. aries	79%	8e-06	33%	3e-10	79%	0.049	34%	6e-10
B. bison	79%	8e-06	31%	3e-10	79%	0.049	------	------
B. bubalis	79%	8e-06	31%	6e-10	79%	0.049	34%	4e-10
B. taurus	79%	8e-06	33%	3e-10	79%	0.049	33%	9e-10
C. falconeri	79%	8e-06	32%	2e-10	80%	0.04	34%	1e-09
O. cuniculus	-------	-------	35%	1e-18	------	------	34%	4e-15
M. monax	-------	-------	38%	1e-18	------	------	37%	3e-16
D. leuca	-------	-------	35%	5e-16	------	------	34%	1e-11
A.melanoleuca	-------	-------	36%	1e-15	------	------	34%	2e-11
F. catus	-------	-------	35%	1e-14	------	------	34%	4e-12
H. sapiens	-------	-------	36%	1e-14	------	------	36%	7e-12

For Genbank Accession number refer to Appendix 6A. ------ = not recognized by BLAST algorithms.

6.4.1.2 IL-2 - Domain structure

A domain search for *M. eugenii* indicated that 138 of the 139 amino acid residues belong to the interleukin-2 family. This result carried an e-value of 3.20e-08. Also identified was a predicted signal peptide of 20 amino acids in length leaving a mature protein of 109 amino acid residues in length. The putative signal peptide cleavage site was predicted

with a 98% probability to lie between residues 20 and 21. The *T. vulpecula* IL-2 sequence was also mapped to the IL-2 family from amino acid residue 1 through to 141. A putative signal peptide was predicted in this sequence with the cleavage site being predicted between residues 20 and 21 with a probability of 80.5%.

6.4.1.3 IL-2 - Glycosylation and glycation sites

Two putative N-linked glycosylation sites were predicted in the *M.eugenii* IL-2 at positions 62 and 101 respectively. The sequence NTSS had a probability of 73.6% and the sequence NVTV had a probability of 59.9%, both being predicted phosphorylation sites. Both sites were above the threshold of 50% and were therefore highly probable. The *T. vulpecula* sequence contained two predicted N-linked glycosylation sites at positions 47 (NVSE) and 62 (NTSS). The latter location was also conserved in the *M. eugenii* sequence.

A single putative O-linked glycosylation site was predicted at position 23 with a probability of 67.9% in *M. eugenii* and 65.6% in *T. vulpecula*. Further inspection of the sequence established that the threonine in position 23 was the only one surrounded by a number of prolines (at positions -1 and +3) and threonines making it a prime site for an O-linked glycosylation event. Three glycation sites were also predicted in the *M. eugenii* IL-2 sequence at positions 3, 53 and 88 with probabilities of 94%, 58% and 79% respectively. The *T. vulpecula* sequence had four predicted glycation sites at positions 46, 77, 84 and 122 with probabilities of 82%, 78%, 72% and 83% respectively.

Validation of the predicted glycation sites requires more sophisticated mass spectrometry or Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOFF) analysis which was outside the scope of this study.

6.4.1.4 IL-2 - Phosphorylation sites

In total, six predicted phosphorylation sites were found in the *M. eugenii* IL-2 amino acid sequence. These included three serines, one threonine and two tyrosines residues. In

total, ten putative phosphorylation sites were predicted in the *T. vulpecula* IL-2 amino acid sequence. These included five serines, four threonines, and one tyrosine residues. The respective locations and probability percentages of those predicted phosphorylation sites are listed in Table 6.5.

Table 6.5. Predicted phosphorylation sites of IL-2 in *M. eugenii* and *T. vulpecula*.

Predicted phosphorylation sites of *M. eugenii* and *T. vulpecula* IL-2						
	M. eugenii			*T. vulpecula*		
Amino Acid	Sequence	Location	Probability	Sequence	Location	Probability
Serine	LKIV**S**ERMK	49	90.1%	LKN**S**ERMK	49	90.8%
	PSNT**S**SIEN	64	98.7%	PSNT**S**SIEA	64	98.8%
	SNTSSIENL	66	81.2%	**S**NTSSIEAL	65	62.9%
				LKYE**S**EDAQ	87	92.5%
				CHYA**S**KKKI	119	86.6%
Threonine	GPEI**T**QCHY	113	58.5%	NGAP**T**SRPP	23	96.4%
				RPPT**T**VLQF	29	59.8%
				VNRLTGPET	108	89.3%
				GPET**T**QCHY	113	92%
Tyrosine	K**Y**ELYIPSN	58	56.1%	GALK**Y**ESED	85%	65%
	GALK**Y**ESKD	85	81.2%			

6.4.1.5 IL-2 - Disulphide bonds

Two disulphide bonds were predicted in the *M. eugenii* IL-2 sequence. The cysteines involved in the predicted disulphide bridges were located at position 9, 71, 115 and 134. The connectivity patterns were 9 to 71 and 115 to 134 (Fig. 6.1). The cysteines involved in the predicted disulphide bridges for *T. vulpecula* were located at positions 9, 71, 115 and 134. The connectivity patterns for those predicted disulphide bridges were 9 to 115 and 71 to 134 (Fig. 6.2).

6.4.1.6 IL-2 - Primary sequence and secondary structure prediction

The *M. eugenii* IL-2 molecule consisted of 420bp which translated into 139 putative amino acids. The *T. vulpecula* IL-2 molecule consisted of 429bp which translated into 142 putative amino acids. Eight alpha helices were predicted in the IL-2 sequences of both species. The sequence of *T. vulpecula* had a single beta strand while *M. eugenii* had no beta strands in its sequence.

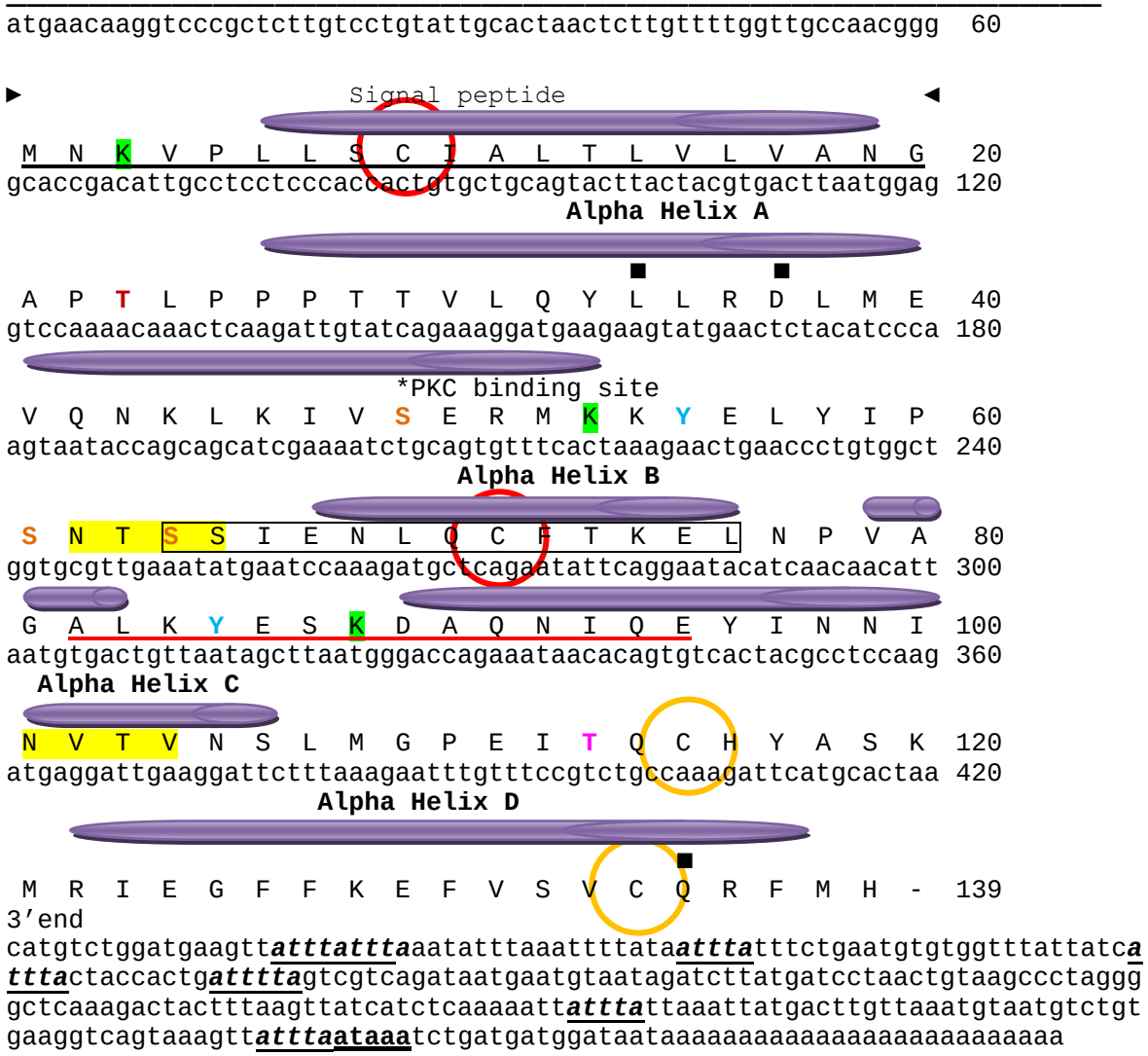


Figure 6.1. *M. eugenii* IL-2 primary sequence and secondary structure prediction. ►Signal peptide is underlined◄. K = glycation sites. Yellow highlight = N-linked glycosylation sites. T = O-linked glycosylation sites. S = serine phosphorylation sites. T = threonine phosphorylation sites. Y = tyrosine phosphorylation sites. Different colours = disulphide bridges . ■ = biologically important residues. Red underlined = custom made *M. eugenii* antibody sequence. **ataaa** = polyadenylation signal. ***attta*** = mRNA instability motifs. = Helices.

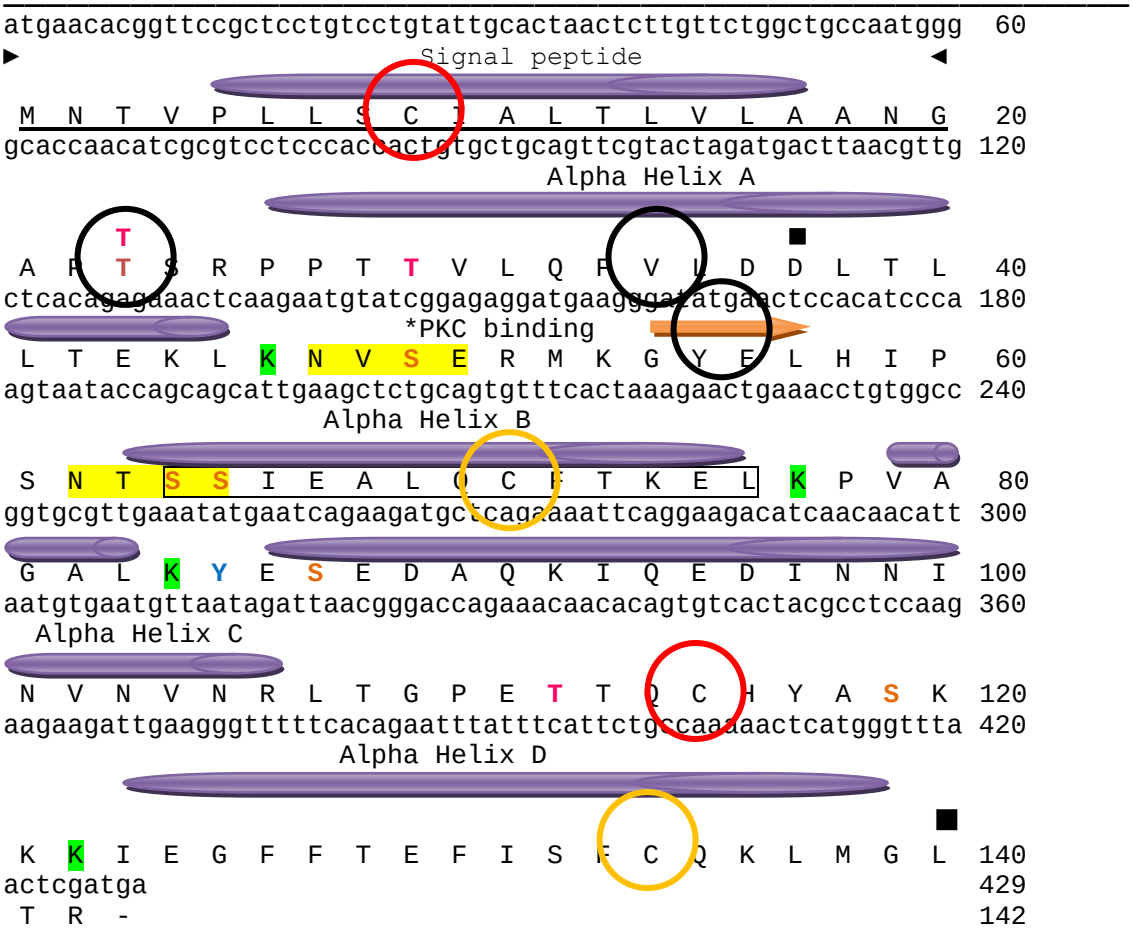


Figure 6.2. *T. vulpecula* IL-2 primary sequence and secondary structure prediction.
►Signal peptide is underlined◄. K = glycation sites. Yellow highlight = N-linked glycosylation sites. T = O-linked glycosylation sites. S = serine phosphorylation sites. T = threonine phosphorylation sites. Y = tyrosine phosphorylation sites. Different colours = disulphide bridges . ■ = biologically important residues. Red underlined = custom made *M. eugenii* antibody sequence. **ataaa** = polyadenylation signal. ***attta*** = mRNA instability motifs. = different to *M. eugeneii* IL-2.

T T = O-linked glycosylation, threonine phosphorylation sites.
= Helices. = Strands

6.4.1.7 IL-2 amplification in genomic DNA

Genomic DNA was used to amplify and confirm the 3' end of the *M. eugenii* IL-2 exon-4 sequence (gel picture in Appendix 6A). When compared to other IL-2 sequences, including the IL-2 sequence of *T. vulpecula,* the *M. eugenii* sequence was found to be three amino acid residues shorter. This may indicate that the *M. eugenii* IL-2 sequence is truncated. The implications of a truncated IL-2 gene are not known. Exon/intron boundaries were also identified in *M. eugenii* and are detailed in Appendix 6A.

6.4.1.8 IL-2 gene polymorphisms

Initially, sequence discrepancies were found in the *M. eugenii* IL-2 sequence in different clones, which led to a search for polymorphisms within the IL-2 gene for both *M. eugeni*i and *T. vulpecula*. Polymorphisms were suspected at the 3' end of exon-4 because of sequence variations. Although polymorphisms were found in the *M. eugenii* IL-2 N-terminal region they were not confirmed at the suspected position in exon-4. A single non-synonymous substitution of an asparagine (N) for a lysine (K) was found in the expressed IL-2 gene of *M. eugenii*. In contrast, no substitutions were detected in the *T. vulpecula* IL-2 sequence.

6.4.1.9 IL-2 promoter annotation using the ensembl database

The ensembl database was used to ascertain the promoter region of the *M. eugenii* IL-2 gene. On scaffolds where the TATA box could not be identified, a search for the characteristic sequence of the individual transcription factors was carried out. A number of transcription elements such as IgNF-A (ATTGCAT), NFAT (GGAGGA), AP-1(TTCAGTCAGT), NFκB (GGGATTTCAC), Oct-1 (AATTGCAT), and NF-IL2-D (ATGCAATTAA) were identified. The TATA box (TATAAAT) was located 85bp upstream from the START site. Promoter sites with a threshold of >53% were considered to be valid. Scaffolds 846545 and 332747 were found to contain several important transcription factors (Appendix 6A).

6.4.1.10 IL-2 - Real Time Polymerase Chain Reaction (qPCR)

The Melt curve shows the appropriate melt temperature for a correct product in *M. eugenii* (Fig. 6.3). However, the gel did not resolve the second product.

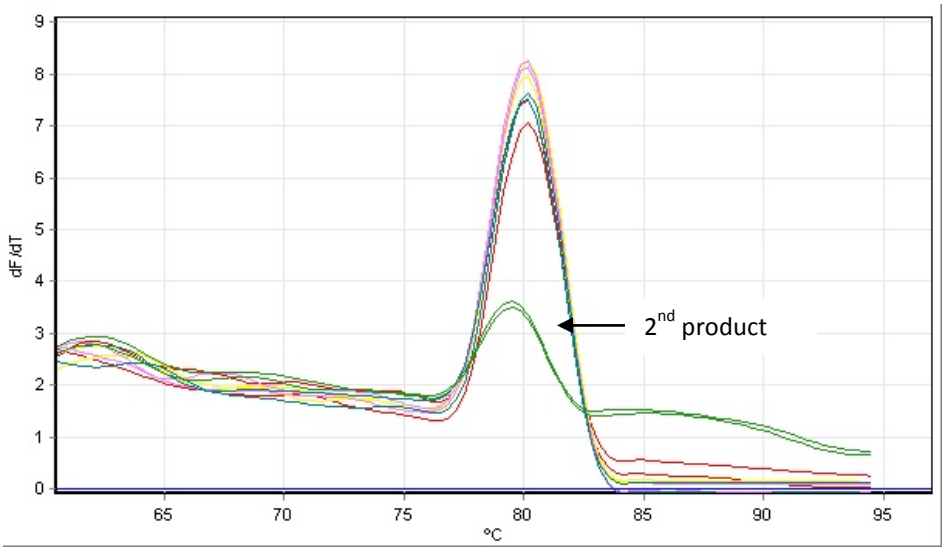


Figure 6.3. Melt curve for *M. eugenii* IL-2. The lines represent the duplicates from each tissue (thymus, spleen, liver, gut node and mesenteric lymph node). The lower curves (green) were observed from liver tissue.

Because the IL-2 sequence in *O. fraenata* has not been characterized or annotated, it was not possible to design a primer specific for that species. *M. eugenii* specific primers were therefore trialled in *O. fraenata* tissue with an HRM approach, and initially appeared to show a good result. However after data analyses and gel electrophoresis it was found that the gene amplified was not interleukin-2 (Fig. 6.4).

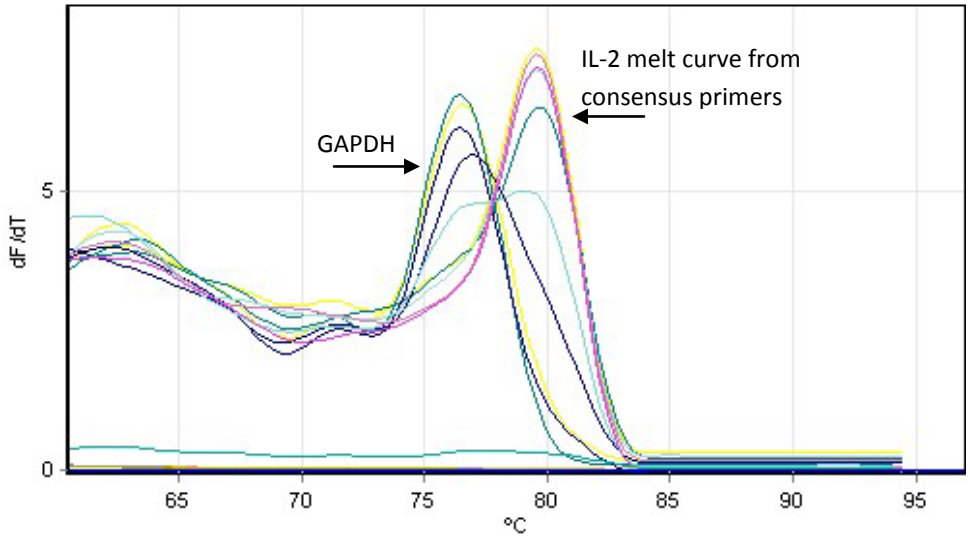


Figure 6.4. *O. fraenata* IL-2 melt curve (Melt curves depicted to the right), against GAPDH (curves to the left). Although the peaks are in the correct region of the melt temperature the IL-2 curve did not sequence as IL-2. The GAPDH curves are in fact the GAPDH product.

This indicated that a consensus approach was not appropriate for a qPCR method. Expression primers used for *M. eugenii* were used for *O. fraenata* cDNA in a real time approach but, due to the specificity of the primer to the *M. eugenii* sequence, an incorrect product was found. While the the qPCR for *M. eugenii* yielded the correct result, the *O. fraenata* qPCR was unable to detect the IL-2 gene in the cDNA from unstimulated tissue.

6.4.1.11 IL-2 - Phylogenetic analysis

The Maximum Likelihood phylogenetic tree inferred the evolutionary history of IL-2 through the Poisson correction model with the highest log likelihood of -7709.06. The percentage of replicate trees in which the associated taxa were clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter=3.13)). The tree is drawn to scale with branch lengths measured in the number of substitutions per site. The analysis involved 62 amino acid sequences with a

total of 193 positions in the final dataset. The marsupial sequences cluster with the mammals and are located between the rodents and the amphibian clade (Fig. 6.5).

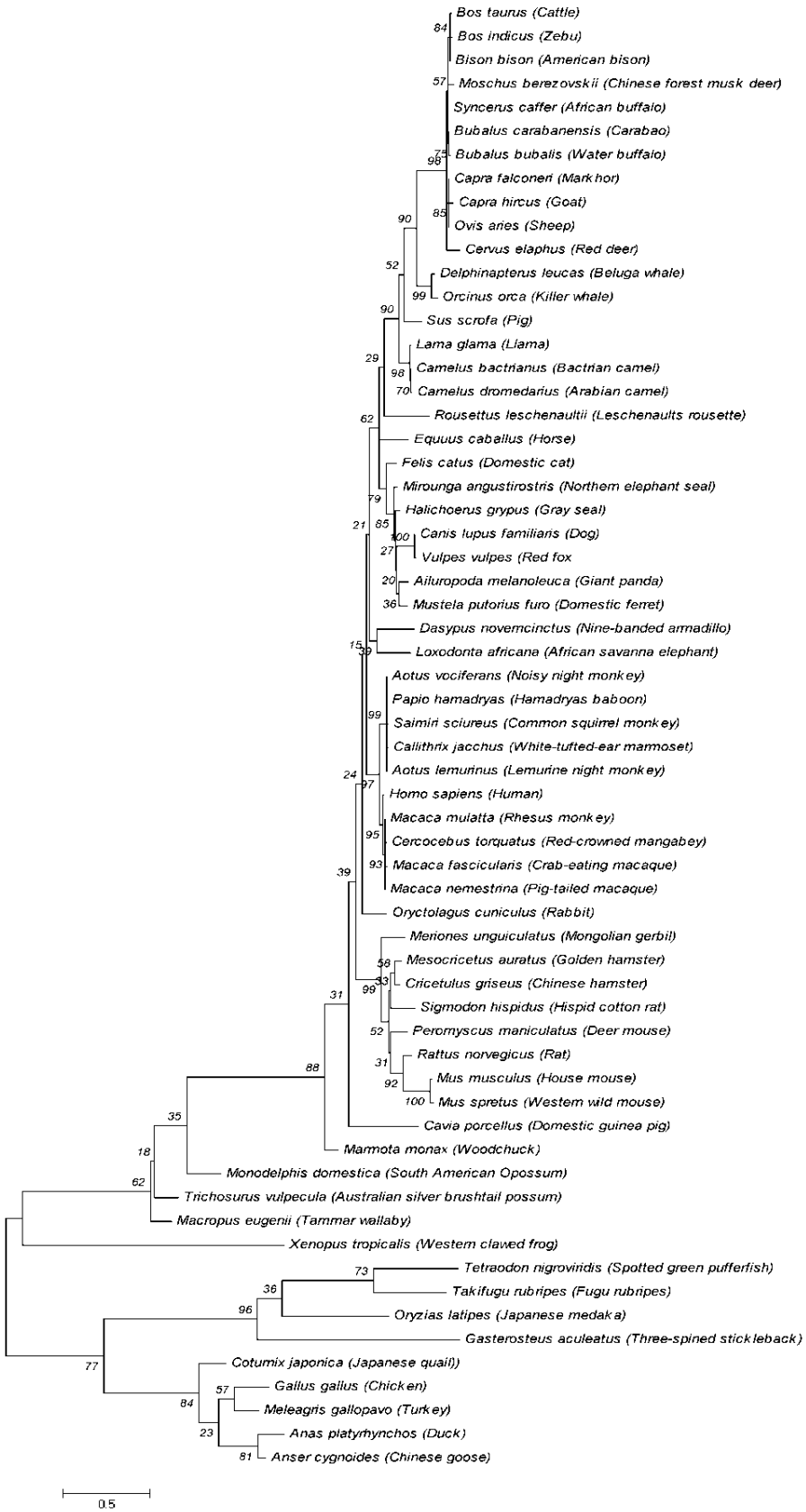
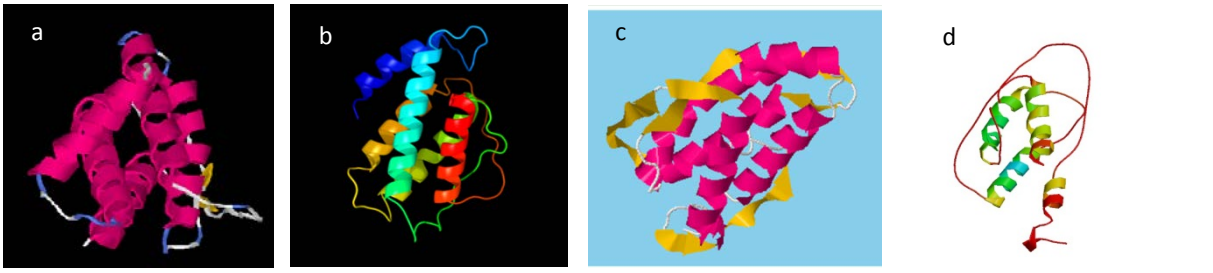


Figure 6.5. Maximum likelihood phylogenetic tree of known interleukin-2 sequences. Short branch lengths for the marsupial IL-2 molecules indicate a close relationship. The branch length between *T. vulpecula* and *M. eugenii* are short signifying a close evolutionary tie, while the branch length to *M. domestica* is longer indicating a greater evolutionary distance.

6.4.1.12 IL-2 - Structure modelling

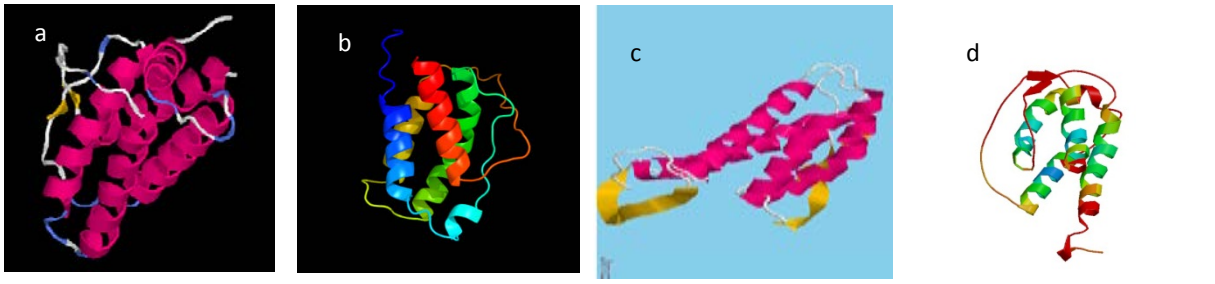
Homology modelling indicated that both the *M. eugenii* and *T. vulpecula* IL-2 molecules had a four helix bundle structure. The models for *M. eugenii* and *T. vulpecula* IL-2 molecules as predicted by the I-TASSER program are illustrated in Fig. 6.6 (a). The models show a different orientation of the helices, and a different distribution of the loops and strands between the two species. The IL-2 models produced by Phyre illustrate the four helices, loops, and beta strands for each species, and the differences in their distribution (Fig. 6.6 b). Those models also show the closest homology to the structure d1m47a. In order to obtain a Phyre model, 30 residues were modelled *ab initio* and are therefore unreliable with regard to their positions within the structure. Seventy eight percent (78%) of the residues were modelled with >90% confidence for *M. eugenii*, while 77% of the sequence from *T. vulpecula* was modelled with 100% confidence. This suggested that the folding pattern of the protein for the two species may be different and this possibility was confirmed by the visual representation of the molecules in Fig. 6.6 (b). The RaptorX program produced a very different model for each of the two marsupial IL-2 molecules (Fig. 6.6 c). Both molecules showed the typical four helical structure, however the *T. vulpecula* model is clearly elongated and the loops and beta strands are further apart. This may suggest that the torsion angles in the *T. vulpecula* IL-2 molecule are different to that of *M. eugenii*. The RaptorX program showed that the closest homology to the pdb was 1irlA, and the second closest template was 1m47a. Similarly, the Modeller 9.10 models for the two species confirms the four helical structure but again shows clear differences between the structures of the two molecules (Fig. 6.6d).

M. eugenii IL-2



I-TASSER model | Phyre model | RaptorX model | Modeller 9.10 model

T. vulpecula IL-2



I-TASSER model | Phyre model | RaptorX model | Modeller 9.10 model

Figure 6.6. *M. eugenii* and *T. vulpecula* IL-2 models.. (a) the I-TASSER models showing the four helix bundle structure, (b) the Phyre models showing small variations between the two models. (c) RaptorX models showing considerable variation between the torsion angles of the molecules.(d) Modeller 9.10 models showing the difference in the loop lengths of the molecules.

The *M. eugenii* and *T. vulpecula* IL-2 structures are indicated by **X** in Figs. 6.7 (a) and (c) and in both cases were clearly outside the non-redundant set of pdb structures. The QMEAN for both species is shown in Figs. 6.7 (b) and (d) and was well below zero in both cases. Both models were poor when compared to the structures resolved by X-ray crystallography deposited in the pdb database.

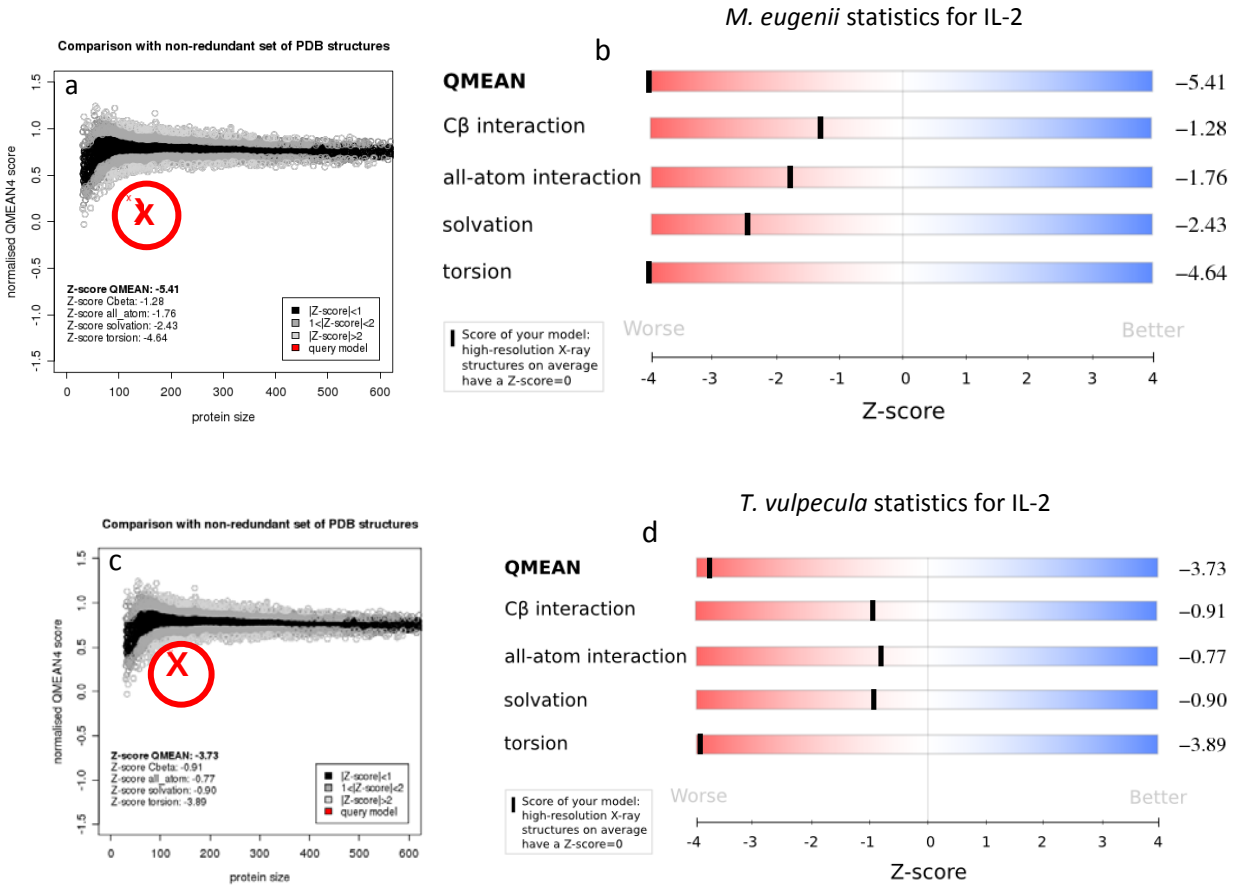
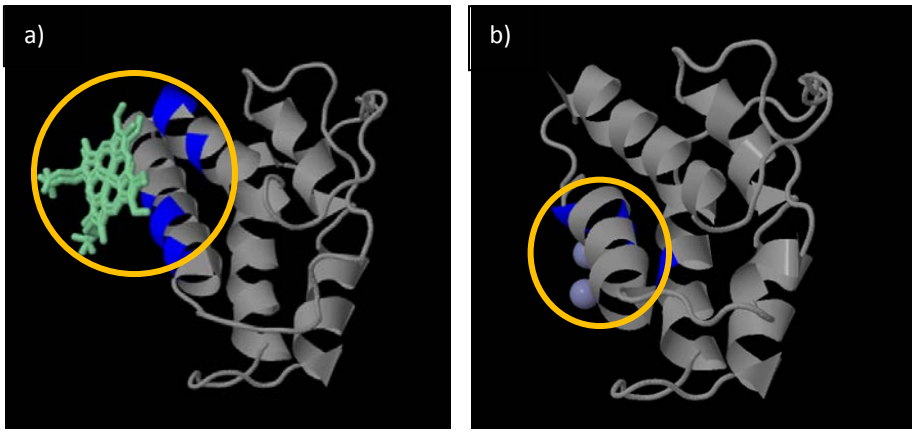


Figure 6.7. Comparison of *O. fraenata* and *T. vulpecula* IL-2 structures with the pdb structure.
a) ⊗ = *M. eugenii* IL-2 sequence compared to homologous structures in the pdb database.
b) Z- score slider indicating the quality of the *M. eugenii* IL-2 model showing a very low QMEAN.
c) ⊗ = *T. vulpecula* IL-2 sequence compared to homologous structures in the pdb database.
(d) Z-score slider indicating the quality of the *T. vulpecula* IL-2 model showing a low QMEAN but is higher than the QMEAN for *M. eugenii*.
.

6.4.1.12.1 IL-2 - Ligand binding

The IL-2 molecules from both *M. eugenii* and *T. vulpecula* were found to bind a number of ligands, however differences were found in the residues involved in this process. Eight amino acid residues at positions 35, 38, 39, 42, 131, 135, 136 and 138 were recognized as being the sites involved in ligand binding for *M. eugenii*. *T. vulpecula* engaged the

residues 32, 33, 37 and 99. The binding site areas are illustrated in Fig. 6.8. These sites were quite different to each other in the structures. An examination of the IL-2 sequences indicated a different biochemical environment surrounding these amino acid residues in each of the two species. This may explain why different residues in the two species are available for ligand contact.



M.eugenii IL-2 *T. vulpecula* IL-2

Figure 6.8. Ligand binding sites of a) *M. eugenii* and b) *T. vulpecula*. The *M. eugenii* model shows a much larger ligand binding area than does *T. vulpecula*. Circled areas are the ligand binding sites. The positions of the amino acid residues involved are listed in Chapter 6, section 6.4.1.12.1.

6.4.1.13 IL-2 - Peptide design

The *M. eugenii* IL-2 sequence ALKYESKDAQNIQE was the most promising in terms of hydrophilicity, isoelectric properties and solubility, and this was confirmed by the Kyte-Doolittle plot shown in Fig. 6.9.

Because of the properties illustrated in Figs. 6.9 and 6.10, the *M. eugenii* sequence ALKYESKDAQNIQE was selected as the peptide most suitable for an antibody. Analysis revealed an overall acidic profile of this peptide at -1 with a pI of 4.4. This 14 amino acid long peptide presented with a predicted MW of 1,636.8 with an average hydrophilicity of 0.6. A Hopp and Woods plot is illustrated in Fig. 6.10.

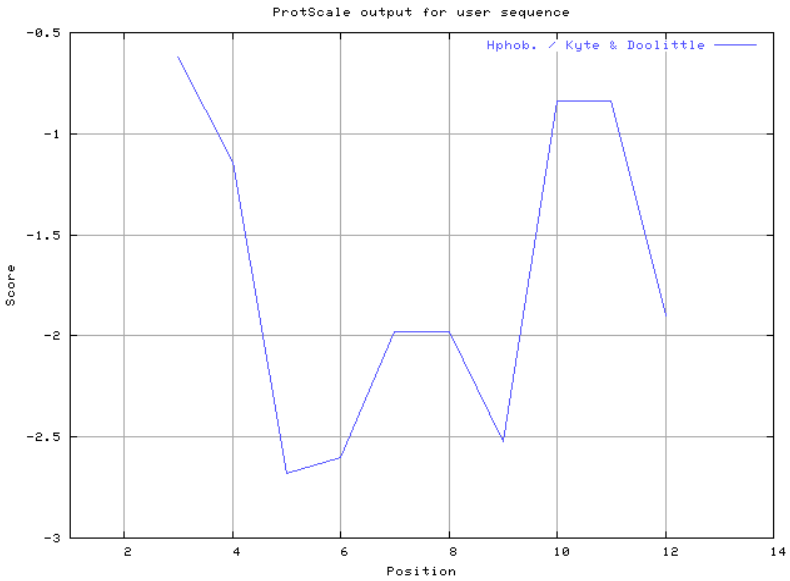


Figure 6.9. Kyte-Doolittle plot obtained through EXPASY for *M. eugenii* IL-2 (Kyte and Doolittle, 1982). This plot shows the hydrophobicity of the selected peptide sequence and allows for the selection of suitable amino acid residues.

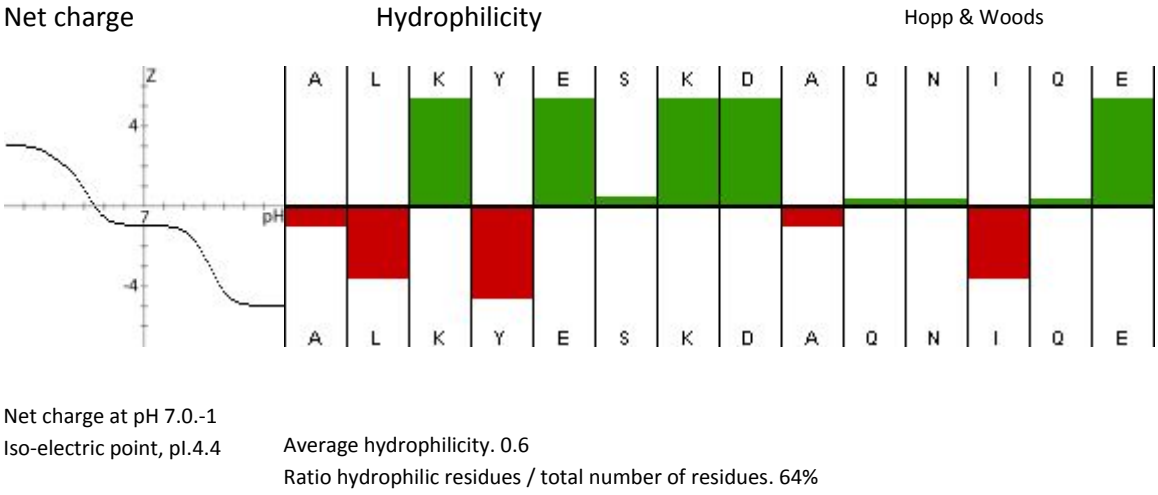


Figure 6.10. Hydrophilicity depicted in a Hopp and Woods diagram (Hopp and Woods, 1983). The plot shows the hydrophilicity of the amino acid residues. It also calculates the net charge and the pI of the selected peptide sequence and indicates the ratio between the hydrophilic and other residues within the peptide sequence.

6.4.1.14 IL-2 - Dot Blot and Western Blot

The Dot Blot carried out to test the custom designed antibody against the peptide with the necessary isotype control is illustrated in Fig. 6.11. The isotype control for the anti-Rabbit IgG was negative while the peptide was recognized by the antibody.

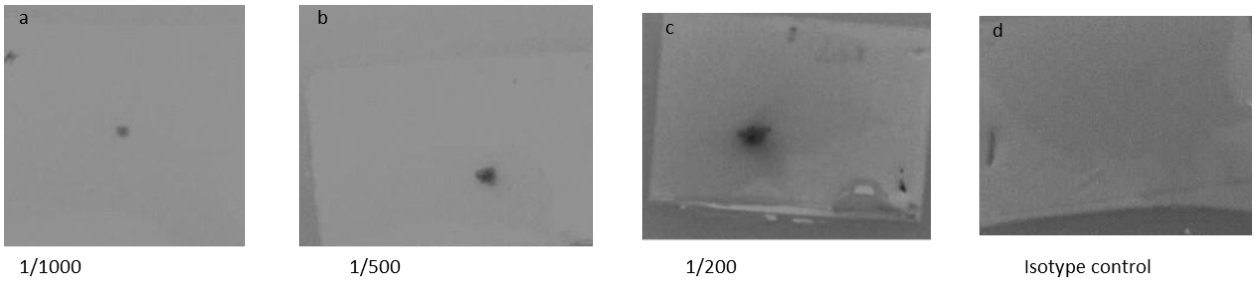


Figure 6.11. Dot Blot of marsupial IL-2 peptide including isotype control for secondary antibody (anti-Rabbit IgG). a) Peptide dot incubated with mpIL-2 antibody (0.1µg/µL diluted to 1/1000). b) Peptide dot incubated with mpIL-2 antibody (0.1µg/µL diluted to 1/500). c) Peptide dot incubated with mpIL-2 antibody (0.1µg/µL diluted to 1/200).

A 12% SDS-PAGE gel confirmed the *M. eugenii* IL-2 protein at ~21kDa which is consistent with an IL-2/IL-2 receptor complex. IL-2 is often found complexed to its receptor rather than as a free molecule in a crude cell lysate (Fig. 6.12).

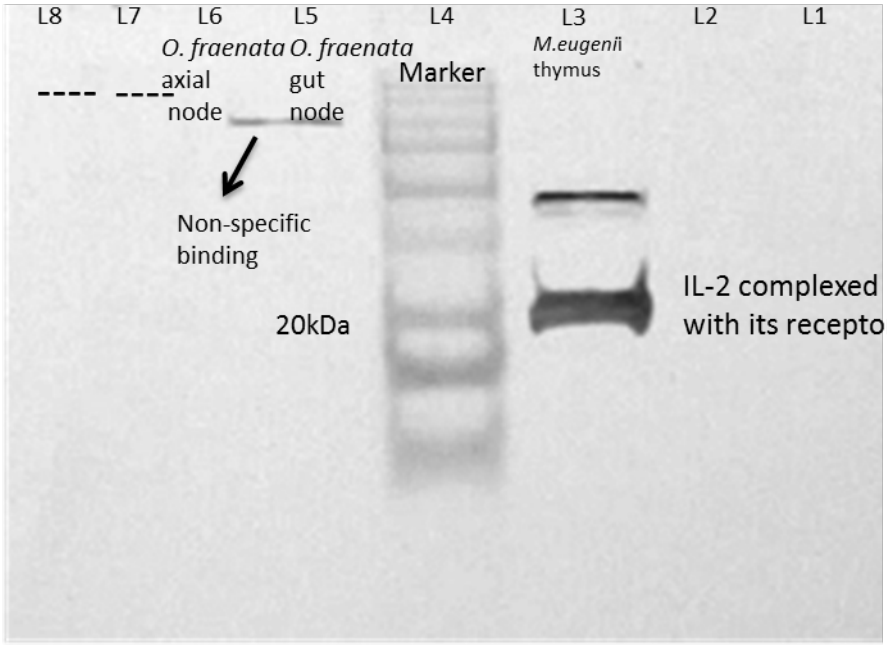


Figure 6.12. Western Blot for IL-2 in *M. eugenii* thymus tissue. *O. fraenata* gut node and *O.fraenata* axial node (L1-L8 are the line numbers).

6.4.2 Interleukin-17

6.4.2.1 IL-17 - Homology

The homology search revealed that the *M. eugenii* IL-17A had a sequence identity of 84% to the predicted *M. domestica* IL-17A at the nucleotide level, and 81% at the amino acid level. The results of the homology search of a number of mammalian species are listed in Table 6.6. The identity percentage to human IL-17A at the nucleotide level was 73% (Table 6.6).

Table 6.6. Homology search results for the *M. eugenii* IL-17 nucleotide and amino acid sequences their identities and respective e- values.

Homology search result for *M. eugenii* for IL-17A				
Species	Nucleotide	e-value	Amino acid	e-value
M. domestica	84%	7e-127	81%	2e-74
H. sapiens	73%	4e-40	61%	4e-51
N. leucogenys	72%	1e-55	61%	5e-52
B. taurus	73%	1e-54	63%	4e-53
P. abelii	72%	5e-53	60%	6e-51
C. hircus	72%	3e-50	63%	6e-53
M. mulatta	72%	7e-51	59%	3e-49
C. Jacchus	71%	3e-50	60%	4e-51
E. caballus	71%	1e-48	60%	1e-50
S. scrofa	71%	6e-46	61%	4e-51
C. familiaris	70%	1e-42	59%	2e-50

For Accession number refer to Appendix 6B

The members of this particular cytokine family are the only molecules with a characteristic knot structure. The homology search indicated that there was a high sequence homology of IL-17A to IL-17F, another member of the IL-17 family. The sequence conservation of IL-17A to IL-17F is $\geq$50% at the amino acid level for a number of mammalian species including the marsupial *M. eugenii* (Table 6.7).

Table 6.7. Conservation of nucleotide and amino acid residues of *M. eugenii* IL-17A to IL-17F of other vertebrate species.

Conservation of *M. eugenii* IL-17A compared to other vertebrate IL-17F molecules		
Species	Nucleotide conservation to IL-17F	Amino acid conservation to IL-17F
N. leucogenys	74%	55%
S. scrofa	74%	59%
H. sapiens	73%	54%
L. africana	73%	52%
M. domestica	73%	52%
M. mulatta	73%	50%
M. gallopavo	72%	52%

6.4.2.2 IL-17 - Domain structure

In *M. eugenii*, a putative protein family domain (PFAM) was identified at positions 8 to 148 with an e-value of 2.20e-51. A possible PAN domain, which is important in protein-protein or protein-carbohydrate interaction, was identified at positions 10 to 83. PAN domains contain a hair-pin loop-like structure, similar to knottins, but have a different disulphide bond pattern (Tordai *et al.,* 1999). The PAN domain had a high probability since IL-17A belongs to a knot structure protein family. This was substantiated by a putative C-terminal cystine knot-like domain (CTCK) at positions 82 to 152. McDonald and Hendrickson (1993) reported that this domain has been found in growth factors but with little sequence homology. These growth factors have an unusual arrangement of six cysteines linked to form a 'cystine knot' formation (Isaacs, 1995).

A predicted nuclear receptor box motif (LxxLL) was found in IL-17A of *M. eugenii* thus conferring binding properties to its nuclear receptors. This structure was highly hydrophobic. It should be noted that it is often identified incorrectly when buried in globular structures (Heery *et al.,* 1997).

The *M. eugenii* IL-17 sequence had a long signal peptide of 28 amino acids. The potential signal peptide cleavage site was located between positions 28 and 29.

6.4.2.3 IL-17 - Glycosylation sites

The program employed to elucidate any putative N-linked glycosylation sites in the *M. eugenii* IL-17A sequence indicated three possible positions. However, upon inspection of the sequence and the probability values, two sites were rejected since they were below the threshold of 50%. The site with the highest probability of 57.49% was found at position 41 and had the sequence NDSS.

6.4.2.4 IL-17 - Phosphorylation sites

The predicted amino acid sequence of *M. eugenii* IL-17A contained seven phosphorylated serines, three threonines and one tyrosine. The positions of these predicted residues, and the probabilities of phosphorylation occurring at these sites, are listed in Table 6.8.

Table 6.8. Predicted phosphorylation sites their location, sequences and probabilities in *M. eugenii* IL-17.

M. eugenii IL-17 phosphorylation sites			
Amino Acid	Sequence	Location	Probability
Serine	KNDS**S**QRVS	44	92.1%
	SQRV**S**INMN	48	96.4%
	YKNR**S**TSPW	70	97.0%
	NRST**S**PWDM	72	99.8%
	KCRH**S**GCIN	97	96.7%
	NCSTSFRLE	132	74.3%
	PRTV**S**----	153	77.1%
Threonine	KNRS**T**SPWD	71	52.4%
	RLPR**T**IWEA	88	54.5%
	CTCV**T**PRTV	148	77.4%
Tyrosine	ISPD**Y**KNRS	66	89.8%

6.4.2.5 IL-17 - Disulphide bonds

Three disulphide bonds were predicted in the *M. eugenii* IL-17A sequence. The six cysteines needed to form the cystine knot in IL-17A were located at positions 35, 129, 94, 144, 99 and 146. The connectivity patterns of the three resultant disulphide bridges were 35 to 144, 99 to 129, and 94 to 146.

6.4.2.6 IL-17 - Primary sequence and secondary structure prediction

The *M. eugenii* IL-17A consisted of 462bp which translated into 153 amino acids. The secondary structure prediction revealed a single transmembrane helix located in the N-terminal end and therefore situated in the leader sequence. Five beta sheets of varying length were also identified in the *M. eugenii* IL-17A sequence (Fig.6.13).

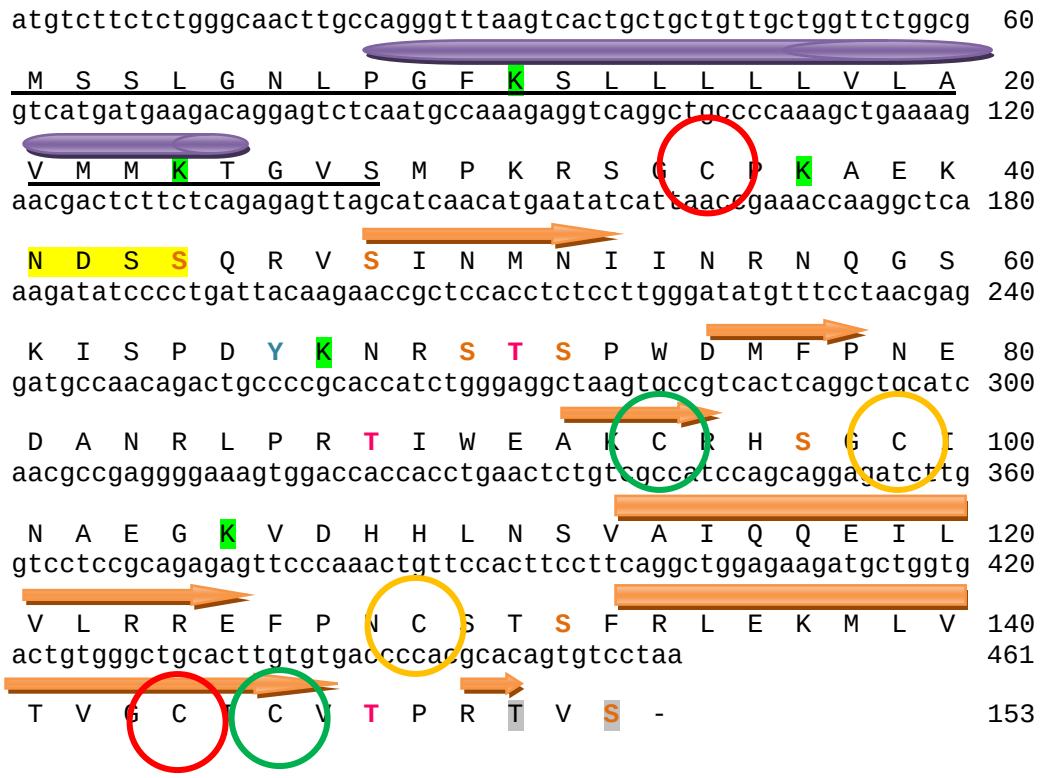


Figure 6.13. *M. eugenii* IL-17 primary sequence and secondary structure prediction. Signal peptide = underlined. K = glycation sites. S = serine phosphorylation. T = threonine phosphorylation sites. Y = tyrosine phosphorylation sites. T = O-linked glycosylation sites. Yellow highlighted = N-linked glycosylation sites = Transmembrane helices. , = Strands .

6.4.2.7 IL-17 - Phylogenetic analysis

The optimal phylogenetic tree inferred using a Neighbor-Joining method and with the sum of branch length of 4.63 is illustrated in Fig. 6.14. The percentages of replicate trees in which the associated taxa are clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 25 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 114 positions in the final dataset. A clear distinction within the marsupial clade was observed with the branch length of *M. eugenii* being shorter than the branch length of *M. domestica* thus indicating a distant relationship.

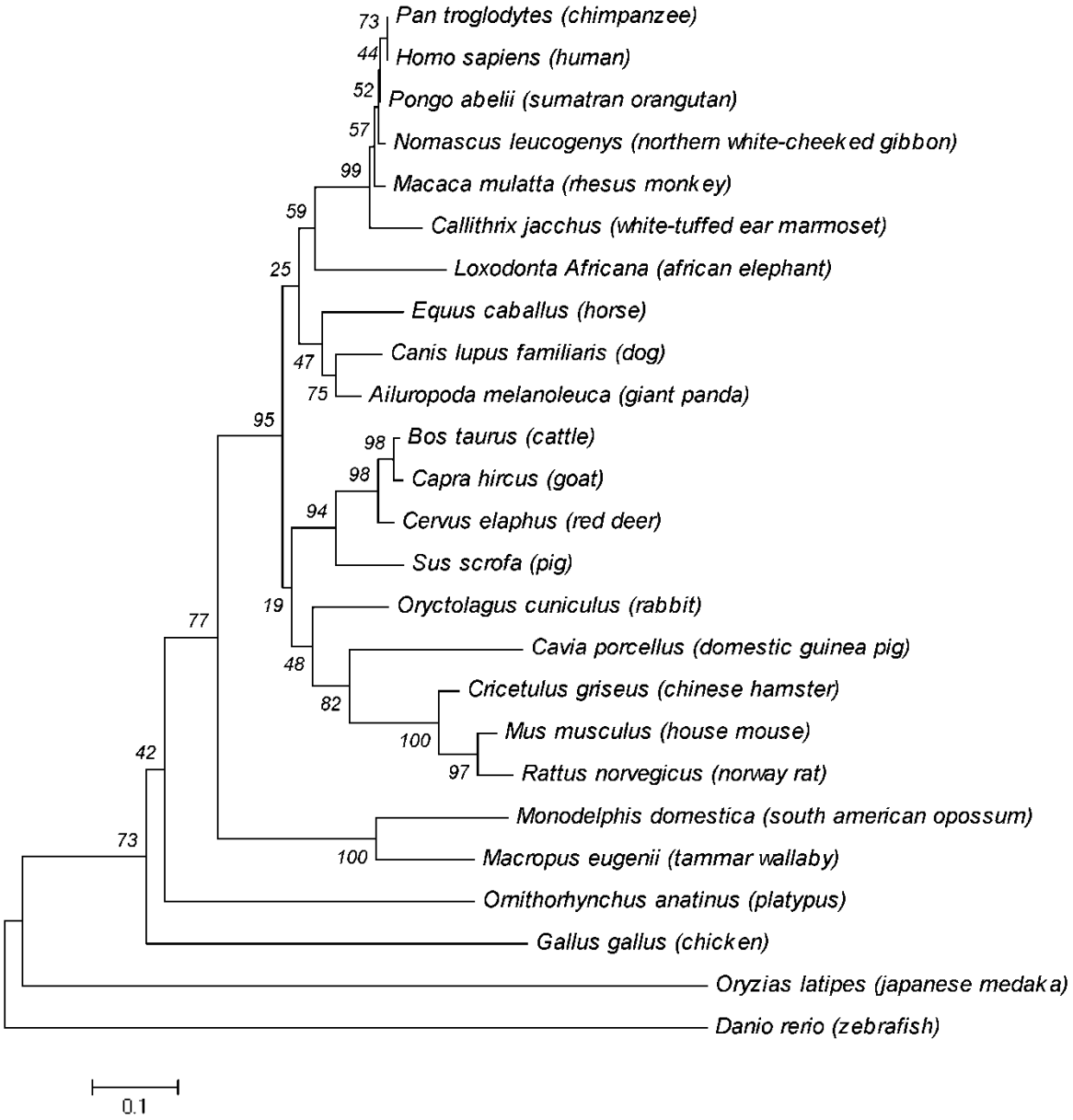


Figure 6.14. Neighbor-Joining phylogenetic tree for IL-17 showing the *M. eugenii* IL-17 evolutionary relationship to other taxa.

6.4.2.8 IL-17 - Structure modelling

The models produced by the different structure prediction programs are shown in Fig. 6.15. All structures in Fig. 6.15 (a-c) show a knot configuration indicating that the *M. eugenii* IL-17 molecule, like other mammalian IL-17 molecules, belongs to the cysteine knot family of cytokines. Each prediction program produced a significantly different model. Although all three models have the classic cysteine knot structure, there are significant differences in the loop and strand positions. The I-TASSER model indicated the possible existence of a number of loops which were only modelled with a 40% probability (Fig. 6.15 a). The RaptorX model (Fig. 6.15 b) had a high homology to the Modweb model structure (Fig. 6.15 c). The Modweb program produced two models of high probability using the templates 1jpyA and 2vxsA. The model 1ijpyA predicted the amino acid residues at positions 29 to 149 successfully, and is annotated in the pdb library as the crystal structure of IL-17F. The model 2vxsA predicted amino acid residues at positions 58 to 151 with a high probability, and this concurred with the structure of human IL-17A in the pdb database. 1ijpyA had a sequence identity of 50% to *M. eugenii* IL-17A, and the cysteine knot domain 1jpyA00 was 97% conserved. The sequence identity of 2vxsA was 70% conserved to the *M. eugenii* IL-17 in the region encompassing amino acid residues 58 to 151. The predicted folded protein is illustrated in Fig. 6.15 (c).

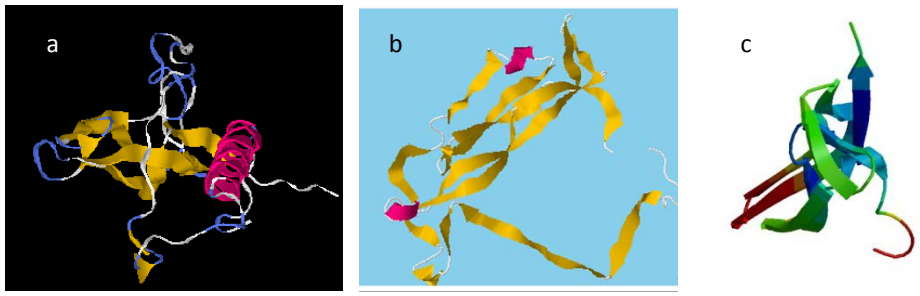


Figure 6.15. Structure predictions for *M. eugenii* IL-17A using different prediction software programs. (a) I-TASSER predicted structure of IL-17, (b) RaptorX predicted structure of IL-17, (c) Modweb prediction of cystine-knot in IL-17.

The QMEAN for the *M. eugenii* IL-17A SWISS-Model was negative. It appeared at the edge of the homologous structures found in the pdb database (Fig. 6.16 a), and the Z- score slider indicated that the model was not a good fit (Fig. 6.16 b). This indicated that, although the different prediction software programs selected the same homology model, the sequence conservation and the resultant structural comparisons could only be mapped to the model selected by each program and no other comparison was possible.

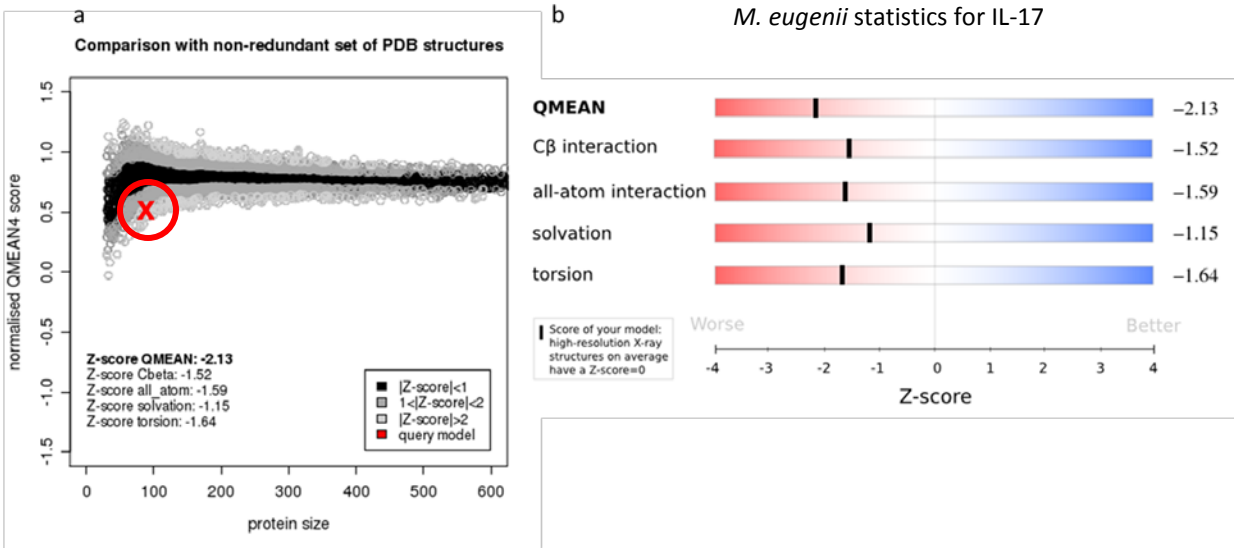


Figure 6.16. Comparison of the *M. eugenii* IL-17structure with the pdb structures. a) X = M. eugenii IL-17 sequence compared to homologous structures in the pdb database. b) Z-score slider indicating the quality of the *M. eugenii* IL-17 model showing a low QMEAN.

6.4.3 Forkhead box protein 3 (FOXP3)

Foxp3 is a regulator in the development of natural regulatory T lymphocytes (nT_{regs}) and induced regulatory T lymphocytes (iT_{regs}). The homology and phosphorylation status is reported in the following sections.

6.4.3.1 Foxp3 - Homology

The BLAST homology search of the partial 426bp long *M. eugenii* nucleotide sequence revealed a sequence identity of 89% with an e-value of 2e-146 to the annotated sequence

of *M. domestica*. At the amino acid level the identity percentage was 88% with an e-value of 2e-74. The *O. fraenata* 421bp long partial nucleotide sequence had an identity percentage of 89% with an e- value of 9e-151 to the annotated *M. domestica* sequence. At the amino acid level the identity percentage was 91% with an e-value of 1e-75. The identity percentages to other mammalian nucleotide and amino acid sequences are listed in Table 6.9.

Table 6.9. Homology search results for the O. fraenata and M. eugenii partial nucleotide and amino acid residues, their identities and respective e- values for the *M. eugenii* and *O. fraenata* partial Foxp3 sequences. The values were calculated over the 426bp/133aa and 421bp/140aa respectively.

Homology search *for M. eugenii and O. fraenata* partial Foxp3 sequences								
	M. eugenii Foxp3 426bp/139aa				*O. fraenata* Foxp3 421bp/140 aa			
Species	Nucleotide	e-value	Amino acid	e-value	Nucleotide	e-value	Amino acid	e-value
M. domestica	89%	2e-146	88%	2e-74	89%	9e-151	91%	1e-75
E. caballus	69%	1e-40	59%	8e-39	69%	6e-71	61%	1e-38
A.melanoleuca	69%	6e-39	59%	5e-39	68%	2e-38	61%	7e-41
F. catus	69%	2e-38	58%	8e-38	69%	7e-38	61%	1e-38
H. sapiens	69%	9e-37	59%	2e-38	69%	1e-35	60%	1e-39
O. aries	68%	1e-34	------	------	68%	4e-35	62%	1e-39
B. taurus	68%	2e-32	60%	3e-39	69%	9e-37	62%	1e-39
L. africana	68%	2e-32	59%	3e-39	67%	2e-31	61%	3e-39
N. leucogenys	68%	2e-32	------	------	68%	7e-32	------	------
C. familiaris	68%	8e-31	57%	3e-37	68%	7e-32	59%	7e-39
P. abelii	68%	1e-29	58%	3e-37	67%	4e-29	60%	1e-37
S. scrofa	68%	1e-29	59%	1e-37	69%	4e-35	60%	7e-38
M. fascicularis	67%	1e-29	58%	2e-37	67%	4e-29	60%	1e-37
C. jacchus	67%	4e-29	58%	7e-37	67%	4e-29	58%	9e-36
M. musculus	68%	2e-27	61%	3e-39	68%	1e-27	61%	1e-38
O. cuniculus	67%	2e-27	57%	2e-36	67%	1e-27	58%	7e-36
C. porcellus	67%	5e-27	61%	4e-35	67%	5e-27	61%	1e-36
M. auratus	67%	2e-26	59%	1e-39	67%	4e-29	62%	8e-41
C. elaphus	67%	3e-24	58%	4e-33	68%	1e-27	59%	5e-34
R. norvegicus	81%	3e-23	58%	3e-37	80%	8e-25	59%	1e-36
C. griseus	78%	6e-20	59%	1e-37	66%	1e-22	61%	4e-38
M. agrestis	67%	9e-18	58%	1e-33	------	------	59%	4e-34
P. maniculatus	77%	1e-16	60%	1e-39	66%	1e-22	63%	4e-41
M. mulatta	80%	6e-07	------	------	67%	5e-27	59%	1e-37
O. anatinus	------	------	60%	1e-37	67%	2e-26	61%	7e-38
H. glaber	------	------	59%	5e-34	------	------	60%	4e-36
M. furo	------	------	56%	3e-28	------	------	58%	2e-30

For Accession numbers refer to Appendix 6D. ------ = not recognized by BLAST algorithms.

It was found that the *M. eugenii* partial sequence had 12 amino acids that were low complexity sequences and none were detected in the *O. fraenata* partial sequence. A sequence variation was detected in the 3' end of the *M. eugenii* forkhead domain which may imply an isoform of the Foxp3 gene (Appendix 6C).

6.4.3.2 Foxp3 - Domain structure

Two putative domains were identified in both *M. eugenii* and *O. fraenata* partial Foxp3 sequences. The domain from positions 116 to 140 was predicted as a Transcription DNA-Binding factor (nucleus regulation Full Forkhead Box factor) with an e-value of 2e-06. The other domain from positions 2 to 38 was predicted as a metal ion binding domain with an e-value of 1e-08. In the *M. eugenii* sequence, the metal binding domain was identified as the isoform Foxp2.

6.4.3.3 Foxp3 - Glycosylation and glycation sites

Six putative O-linked glycosylation sites were identified in the partial sequence of *M. eugenii* and their confidence values ranged from 57% to 64%. There were seven putative O-linked glycosylation sites found in *O. fraenata* with confidence values ranging from 55% to 64% (Table 6.10). Glycation sites were also predicted in both species and, together with their positions and confidence levels, are listed in Table 6.10. However validation of the predicted glycation sites requires more sophisticated mass spectrometry or Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOFF) analysis which was outside the scope of this study. No putative N-linked glycosylation sites were predicted in the partial Foxp3 sequence of either of the two species.

Table 6.10. Predicted O-linked glycosylation and glycation sites in *M. eugenii* and *O. fraenata* Foxp3 and their probabilities.

Predicted O-linked glycosylation sites in *M. eugenii* and *O. fraenata*				
Species	O-linked glycosylation site positions	Probability	Glycation site positions	Probability
M. eugenii	71	57%	1	77.7%
	75	64.6%	11	64.1%
	79	59.8%	31	79.1%
	81	56.8%	40	92.2%
	85	61.2%	58	82.8%
	92	58.8%	63	79.7%
			73	57.9%
O. fraenata	54	55.5%	5	95.8%
	58	63.6%	7	81%
	62	59.4%	1	81.5%
	64	56.3%	23	92.8%
	68	61.7%	41	68.7%
	75	59.8%	46	94.4%
	140	54.7%		

6.4.3.4 Foxp3 - Phosphorylation sites

There were four serine, two threonine and one tyrosine phosphorylation sites predicted for both the *M. eugenii* and *O. fraenata* Foxp3 partial sequences. The positions of the phosphorylation sites and their respective confidence levels are shown in Table 6.11.

Table 6.11. Predicted phosphorylation sites, their positions and probabilities of Foxp3 in *O.fraenata* and *M. eugenii*.

Predicted phosphorylation sites of Foxp3 in *O. fraenata* and *M. eugenii*					
O. fraenata			*M. eugenii*		
Amino acid	Position	Probability	Amino acid	Position	Probability
Serine	39	75.7%	Serine	56	75.7%
	51	98.0%		68	97.4%
	53	55.6%		70	86.5%
	102	59.8%		119	59.8%
Threonine	54	92.2%	Threonine	71	91.5%
	75	91.9%		92	91.9%
Tyrosine	59	73.1%	Tyrosine	76	80.3%

6.4.3.5 Foxp3 - Disulphide bond prediction

There were two predicted cysteines in both the *M. eugenii* and *O. fraenata* Foxp3 sequences. These formed single disulphide bonds and therefore formed a cystine in both species.

6.4.3.6 Foxp3 – Primary sequence and secondary structure prediction

The *O. fraenata* partial Foxp3 sequence contained 426bp which translated into 139 amino acids, while the *M. eugenii* sequence contained 421bp which translated into 140 amino acids. There were differences in the secondary structures of the sequence fragments. The *O. fraenata* sequence contained four helices compared to *M. eugenii* which contained six helices. No beta strands were detected in either of the partial sequences.

Figure 6.17. *O. fraenata* partial Foxp3 primary sequence and secondary structure prediction. S = serine phosphorylation sites. T =, threonine phosphorylation sites. Y = tyrosine phosphorylation sites. = disulphide bridge. = Helices.

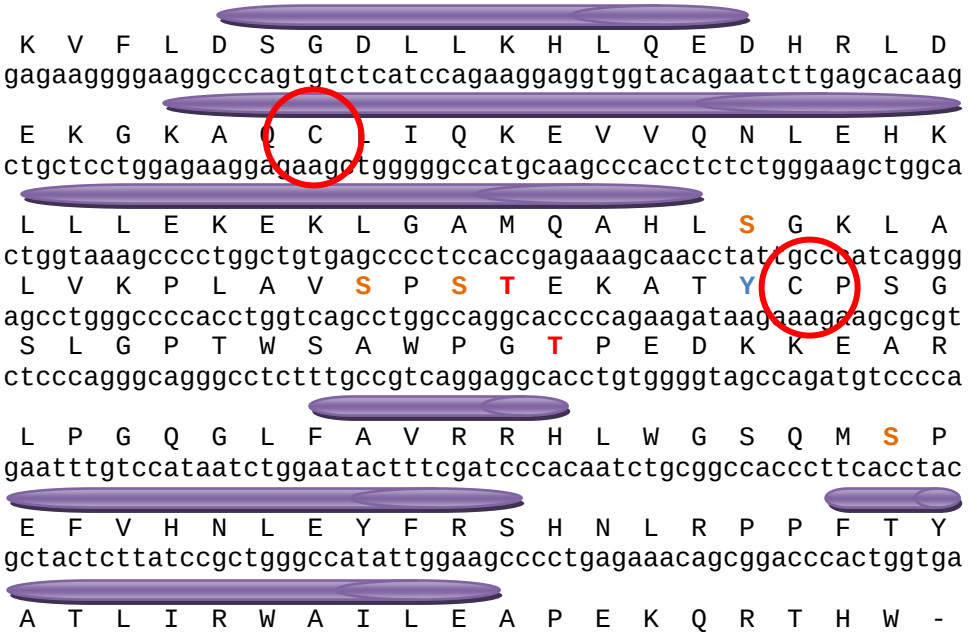


Figure 6.18. *M. eugenii* partial Foxp3 primary sequence and secondary structure prediction. S = serine phosphorylation sites. T =, threonine phosphorylation sites. Y = tyrosine phosphorylation sites. = disulphide bridge. = Helices.

6.5 Discussion

This investigation provided the first conclusive evidence of the existence of IL-2 in marsupials. IL-2 was initially described as the T cell growth factor and was purified from mitogen-stimulated lymphocyte cultures (Morgan *et al.,* 1976). This cytokine controls the amplification of naïve T lymphocytes by initially stimulating growth following antigen activation but later it also promotes an activation-induced cell death (Bird *et al.*, 2005a, Waldmann *et al.*, 2001, Smith, 1988a). The *T. vulpecula* IL-2 sequence was characterized in this study and was published by Young *et al.* (2011). The *M. eugenii* sequence was also characterized in this study and is yet to be published.

It has been reported that IL-2 is tissue-specific and the DNA sequence can vary when isolated from different tissue types. It has also been reported that human placental IL-2 cDNA is 247 nucleotides longer in the 5' untranslated region than cDNA derived from T

lymphocytes (Chernicky *et al.,* 1996). In the present study, the *M. eugenii* IL-2 molecule was identified in PHA stimulated lymphocytes only, therefore no conclusion can be drawn as to the length of the sequence in other tissue types and how much the sequence may vary. Two splice variants of this gene have been reported in humans. One of the splice variants is missing exon-2 (21bp) and the other is missing exon-3 (22bp) (Tsytsikov *et al.,* 1996). The *M. eugenii* interleukin-2 was amplified with primers located in the 5' and 3' untranslated regions in order to examine the sequence for possible splice variants. No splice variants could be detected in either *M. eugenii* or *T. vulpecula*.

Two putative N-linked glycosylation sites were identified in the *M. eugenii* and the *T. vulpecula* sequences but the locations of those sites differed in the two species. The consequence of this is unknown, however it has been reported that location changes of N-linked glycosylated sites can have an effect on transcription activity (Chan *et al.,* 2010). The IL-2 molecule in humans and other mammals has a distinct family signature (TELxxLxCLxxE). This family signature sequence was shown to be only partially conserved in *M. eugenii* and *T. vulpecula* and is located in the alpha helix B region. In position five of the family signature sequence a substitution occurred from asparagine (N) in *M. eugenii* to alanine (A) in *T. vulpecula*. Structure modelling indicated that this substitution does not change the four helical structure of the molecule even though alanine has a non-polar side chain compared to a polar side chain in asparagine. The consensus sequence in mammals showed a histidine in this position which also has a polar side chain, while the bird sequence has a leucine residue which has a non-polar side chain. Within this motif in both the *M. eugenii* and *T. vulpecula* sequences is one cysteine at position 71 which builds a disulphide bond with the cysteine located at position 131. Strict conservation of these cysteines is essential for bioactivity of many mammalian IL-2 molecules (Gaffen *et al.,* 1998). This may indicate that the bioactivity of the marsupial interleukin-2 gene is preserved.

When examining the species alignment of IL-2 (Appendix 6A) it was noted that the *M.eugenii* sequence was three amino acid residues shorter than that of all other mammals including *T. vulpecula*. The *M. eugenii* sequence is similar in length to the sequence of birds. Kaiser and Mariani (1999) reported that human and mouse IL-2 molecules have a four helical structure. It was shown in the present study that the *M. eugenii* and *T. vulpecula* molecules also have a four alpha helical structure. Neither of the two marsupial sequences have the glutamine repeat that has been found in the mouse IL-2 sequence. The phylogenetic analysis showed the marsupials are basal to the eutherian mammals while the amphibians are basal to both groups. This indicates an evolutionary trend that was found in many of the molecules investigated in marsupials in this study.

The stability of the mRNA is regulated by AU-rich elements (AREs) which lead to a rapid degradation of the mRNA. AREs differ in length and sequence but contain several copies of the AUUUA pentamers which are the RNA instability motifs (Chen and Shyu, 1995). These sequence elements serve as binding sites for proteins that regulate mRNA stability and have a dual function. The mRNA can either be stabilized or marked for degradation (Graham *et al.,* 2010). Expression of cytokines is dependent on the highly conserved instability motifs. Transcriptional and translational regulation and modulation of mRNA is an important part of gene expression by altering the amount of translatable mRNA (Hargrove and Schmidt, 1989). Studies carried out on human cDNA have shown that the number of AUUUA cassettes plays a role in mRNA degradation (Akashi *et al.,* 1994) and can also affect mRNA translational efficiency (Kruys and Huez, 1994). It was shown in this study that the *M. eugenii* 3' untranslated region has seven mRNA instability motifs (ATTTA) in contrast to the *T. vulpecula* 3' untranslated region which had only one. These compare to the *Fugu rubripes* (tiger blowfish) which has eight (Bird *et al.*, 2005a) and humans which, like *M. eugenii,* has seven mRNA instability motifs. *T. vulpecula* had the lowest number of instability motifs in the 3' untranslated region compared to all other vertebrates. This may indicate that the half-life of the *T. vulpecula* IL-2 molecule is longer than the IL-2 molecules in other mammals including *M. eugenii.* The *T. vulpecula* IL-2

molecule may therefore be more stable than that of *M. eugenii*. This would mean that the *T. vulpecula* IL-2 molecule would not disappear as rapidly from the cytoplasm after a decrease in transcription which occurs in humans as reported by Ohme-Takagi *et al.* (1993).

In this study it was found that the *M. eugenii* IL-2 sequence contained a non-synonymous substitution where a lysine (K) was substituted with asparagine (N). Both amino acids have polar side chains and positive charges. Due to the polar nature of asparagine it generally prefers to be exposed to an aqueous environment and can therefore be located on the surface of proteins. Asparagine is often found in the active sites of proteins and if found in an N-linked glycosylated site it cannot be substituted (Betts and Russell, 2003). This is certainly not the case in the *M. eugenii* IL-2 sequence. However a substitution to lysine is quite possible since this amino acid is also polar and carries the same charge as asparagine. Similarly, lysine is also found in the active sites of proteins and fulfills the same function as asparagine (Betts and Russell, 2003). There appeared to be no change in function of the *M. eugenii* IL-2 molecule.

IL-2 belongs to the four helical cytokine family. In contrast, IL-17A belongs to the cysteine knot family and represents a link between innate and adaptive immunity. Prior to this study, the expressed sequence of IL-17A had not been identified in marsupials although an annotation of the gene sequence is in the ensembl database. This cytokine has only recently been discovered in human tissues and has been reported as belonging to the T_h17 T cell sub-population (Korn *et al.,* 2009). While IL-2 was thought to be the principal T cell growth factor it appears to restrain IL-17 production (Adler, 2007). IL-17 has been found to play a role in inducing and mediating pro-inflammatory responses (Gaffen, 2009). IL-17A has a unique sequence and is one of six members of the cysteine knot family of cytokines. Human IL-17(A) has a ~60% sequence homology to IL-17(F). A BLAST homology search showed a similar homology in *M. eugenii*.

The *M. eugenii* IL-17A sequence was two amino acids shorter than the IL-17A sequence found in other mammals and one amino acid shorter than the IL-17A of teleost fish. The human IL-17 gene consists of a 19 amino acid residue signal sequence and a 136 amino acid residue mature segment (Fossiez *et al.,* 1996). Six cysteine residues and one potential N-linked glycosylation site are contained within the sequence (Yao *et al.,* 1995a). In this study, it was found that the *M. eugenii* IL-17 putative leader sequence had a length of 30 amino acids which was almost double the length of the human sequence. The purpose of such a long leader sequence is unknown. It has been reported that, in humans, two sets of paired β-strands connect the disulphide linkages between strands 2 and 3, and a third disulphide bridge connects strands 1 and 3. These characterize the cysteine knot fold of IL-17 (Hymowitz *et al.,* 2001). This typical fold configuration was also observed in the *M. eugenii* IL-17 molecule (Fig. 6.15 c).

The predicted IL-17A sequence of *M. domestica* was quite different to the same sequences in all other vertebrates. No comparison could be made between the expressed gene sequence of *M. eugenii* and the predicted sequence of *M. domestica*. The *M. domestica* sequence was 206 amino acid residues long compared to the 153 amino acid residues of *M. eugenii*. The *M. domestica* IL-17A sequence is the longest ever reported in any vertebrate. A multi-valine motif was found in the *M. domestica* sequence and this was absent in all other mammalian and non-mammalian IL-17 sequences. Such a large number of valines following each other have not been previously reported. The closest sequence motif that was found to be conserved in the *M. domestica* IL-17 sequence was the DART motif which is contained in receptors where this motif features as an alpha helical structure. In this study, this motif was also identified in the M*. eugenii* IL-17 molecule. This is often associated with an inhibition of receptor homomultimerization. However IL-17 is a cytokine, not a cytokine receptor, and no reference can be found as to the high level of valines in such a circumstance (del Rio *et al.,* 2007).

The multi-leucine motif identified in the *M. eugenii* IL-17A sequence is a potential leucine zipper and was present in all vertebrate IL-17 sequences with the exception of *Cavia porcellus* (guinea pig).

The expressed *M. eugenii* IL-17A sequence found in the present study differs from the annotated sequence. The expressed sequence had an extra two amino acids in the leader sequence. No other substitutions were detected between the two sequences. By comparing the 3' end of IL-17A with the 3' end of IL-2 it was found that IL-17A does not contain the mRNA instability motif. This was assumed by using the ensembl database and may not be correct since the full 3' end has not been annotated.

The two cytokines IL-17(A) and IL-17(F) have pro-inflammatory functions similar to the regulatory T lymphocytes, another new addition to the T cell lineage. T_{reg} cells are identified by the expression of the transcription factor Foxp3 and are reliant on IL-2 for maintaining homeostasis (Fontenot *et al.,* 2005a). It was reported that circulating memory $Foxp3^+$ T_{reg} cells also secrete IL-17 (Ayyoub *et al.,* 2009). In order to show that marsupials have the full range of T cell sub-populations compared with humans and mice, it was important to show that the cell surface marker of the T_{reg} cells was expressed. An annotation of a small fragment of the Foxp3 molecule is found in ensemble, however that fragment is too small to find important structural motifs. *O. fraenata* and *M. eugenii* Foxp3 partial sequences were therefore amplified to identify structural motifs in Foxp3.

The human full length protein of Foxp3 contains a forkhead DNA-binding domain at the C terminal end which, directly or by forming a repressor complex with nuclear factor of activated T lymphocytes (NFAT), can bind to the IL-2 promoter and repress IL-2 mRNA transcription. That protein encodes a zinc-finger (C_2H_2) (aa 200-223) and a leucine zipper (aa 240 -261) domain that permit homodimerization or heterodimerization with other forkhead family members or other DNA-binding co-factors (Li *et al.,* 2007, Lopes *et al.,* 2006). Little is known about the molecular mechanism by which Foxp3 functions although it has been established that it plays an important role in the maintenance of self-

tolerance within the immune system (Floess *et al.,* 2007). This molecule acts as a transcriptional repressor when expressed in either non-lymphoid cells or T cell lines where it inhibits activation induced cytokine expression (Schubert *et al.,* 2001).

The cloned fragments of the expressed Foxp3 genes characterized in this study for *M. eugenii* and *O. fraenata* revealed a zinc finger domain in both species. In mammals, the main motif in the zinc finger has either the sequence DFLKH (D = aspartic acid, F = phenylalanine, L = leucine, K = lysine, H = histidine) or EFLKH (E = glutamic acid, F = phenylalanine, L = leucine, K = lysine, H = histidine). The *M. eugenii* sequence is DLLKH (D = aspartic acid, L = leucine, K = lysine, H = histidine) and in *M. domestica* the sequence is ELLKH (E = glutamic acid, L = leucine, K = lysine, H = histidine) indicating a substitution in the second position of this motif. In this study, it was found that the substitution of the second amino acid phenylalanine to leucine is unique in *M. eugenii*. Since both amino acids are non-polar and neutral, this substitution does not appear to change the functionality of the zinc finger domain. Unfortunately, the *O. fraenata* sequence did not extend far enough to be able to investigate this particular sequence motif in that species.

The Foxp3 gene also contains a DNA-binding domain which independently folds and recognizes double or single-stranded DNA. The domain sequence for most mammals is GAMQ (G = glycine, A = alanine, M = methionine, Q = glutamine). It was found in this study that the *M. eugenii* amino acid sequence in the motif changes to SAMQ (S = serine, A = alanine, M = methionine, Q = glutamine). The substitution in position one of the *M. eugenii* motif does not appear to change the functionality of the domain.

The forkhead domain of Foxp3 is highly conserved among all mammalian species. Since it was also shown to be conserved in *M. eugenii*, the functionality of this important transcription factor must also be conserved in that species.

6.6 Conclusion

Marsupials are an ancient mammalian lineage, and consequently, it would be expected that the sequences in their immune modulatory molecules may be different to those of more recent lineages. Some, but not all, of the marsupial cytokines have been characterized and reported by others. However, prior to this study, the cytokines IL-2 and IL-17, together with the transcription factor Foxp3, have been characterized in many mammals but not in marsupials. IL-2 and Foxp3 immune molecules are vital for the effective functioning of the T lymphocyte responses in mammals, and have now been characterized in marsupials for the first time. It has been reported that marsupials have diminished T lymphocyte responses compared to other mammals (Infante *et al*., 1991, Stone *et al*., 1996). However, the presence of these vital immune molecules in *M. eugenii*, as demonstrated in this study, suggests that their T lymphocyte responses should not be notably different to those of other mammals. The apparent contradiction therefore suggests that the reported impaired T lymphocyte responses in marsupials may be linked to factors other than the IL-2 and Foxp3 molecules. IL-17 is the main cytokine in the T_h17 cell sub-population, and Foxp3 is the cell surface marker for the T_{reg} sub-population. Since IL-17 and Foxp3 were shown to be expressed in *M. eugenii,* it follows that the T_h17 and T_{reg} cell sub-populations were also present. It is therefore concluded that the marsupial adaptive immune system is not significantly different to that of other mammals.

Future work should include functional studies of IL-2 and IL-17 to confirm the bioactivity of those cytokines in marsupials.

Chapter 7

Research summary, concluding discussion and future work

7.0 Research summary, concluding discussion and future work

7.1 Research summary and discussion

It has been reported in the literature that marsupials have diminished T cell responses such as an absent mixed lymphocyte reaction (Stone *et al*., 1998, 1997a, 1997b, Infante *et al*., 1991). Anecdotal evidence from zoo keeper observations also indicates that there is considerable variation in disease resistance among marsupial species. For example, *O. fraenata* appears to be more resistant to mycobacterial infections, while *L. hirsutus* is reported to be highly susceptible to such infections. This study was undertaken to characterize key molecules in the marsupial immune system to help explain those different immune responses, and to form the direction of further immunological studies. Specifically, the key receptors and signalling molecules of the T cell signalling cascade and the resultant expression of the main cytokines were investigated. The expressed sequences were compared with the sequences of the recently published wallaby genome (Renfree *et al*., 2011) since any differences between the predicted and expressed sequences may indicate splice variants or polymorphisms. Such differences were encountered in the molecules CD28 and ZAP-70. CD28 is only partially annotated in ensembl for *M. eugenii* since the start and stop sites are missing. The CD28 5' annotated sequence is missing 51bp, while the 3' end is missing 163bp. On the other hand, ZAP-70 was fully annotated for *M. eugenii*, however significant differences were found between the annotated sequence and the expressed sequence reported in this study. It is possible that the annotated sequence is incorrect and, if so, the expressed sequence will provide a better template for further studies.

These findings were substantiated by the identification of important structural motifs contained within the putative amino acid sequences of the molecules investigated in this study. It was also found that, in most cases, these motifs were highly conserved between marsupial species and between marsupials and eutherian mammals. This high sequence homology did not always translate into structural homologies. Tertiary structures were determined by using various bioinformatics tools. Although there were similarities in the

structures, there were often differences in the torsion angles and only a few of the structures, such as the CD8α and CD28 receptors and the signalling molecules TCRζ and CD3ε, could be modelled with a high confidence level. It was also found that the transmembrane region in the TCRα chain appeared to differ between *O. fraenata* and *M. eugenii*. This region is an important functional motif of this molecule. A structure prediction indicated that the *O. fraenata* sequence had two phenylalanines where the ring structures faced each other while in *M. eugenii,* which also had two phenylalanines, the ring structures faced away from each other. The positions of the phenylalanines in the marsupial TCRα are unique to marsupials and monotremes. In other mammals, leucine and valine take the place of phenylalanines in those positions. Human studies have shown that the leucines in those positions are important for the function of the receptor. Whether or not this substitution in marsupials has any effect on the biological function of the TCRα receptor is not known.

It was also found that the β-chain of the TCR receptor is highly conserved in marsupials. However in this study, one important difference was identified in the TCR-C beta-beta strand E where eight amino acids were found to be missing in the *O. fraenata* sequence. This may have implications for the T cell responses in that species. This motif represents the site of interaction between the TCRα and TCRβ chains and, since this motif determines the regulatory function of the TCRβ chain, its absence may indicate a down-regulation of the receptor.

The molecules CD3ε, CD4, CD8αβ, CD28, CD86, and CTLA-4 belong to the immunoglobulin superfamily. The prominent Ig-fold, which characterizes this superfamily, consists of two beta sandwiches, highly conserved disulphide bonds, and the complementarity determining region. These features were all found in the marsupial molecules investigated in this study but some differences were identified. CD8αβ and CD28 were modelled with high QMEAN scores indicating a high probability of the determined structures. However CTLA-4, although highly conserved between *M. eugenii* and *O. fraenata*, was an exception. The prediction markers for that molecule in both species

were highly negative and the software, although able to make a comparison with other pdb structures, was unable to produce the Z-score slider and Bell curve for this molecule in those species. While the *M. eugenii* CTLA-4 was modelled as a dimer, the *O. fraenata* molecule could only be produced as a monomer, a structure that has been shown in humans to be linked to the inhibition of the mixed lymphocyte response (Oaks *et al*., 2000). This monomer represents the soluble form of the receptor which has been shown to function in the up- and down-regulation of CTLA-4 in different disease states (Toussirot *et al*., 2009). It has been suggested that CTLA-4 has the ability to confer either resistance or susceptibility to disease depending on its expression kinetics (Walker and Sansom, 2011). The difference in the *O. fraenata* CTLA-4 structure may therefore explain the reported disease resistance of that species.

The characterization of CD4 in *O. fraenata* showed that almost an entire domain (D4) was missing in this species. A literature search revealed that there are CD4 molecules with unusual transcripts in various species (Moore *et al.,* 1992, Lonberg *et al.,* 1988). The missing D4 domain in *O. fraenata* may indicate that this CD4 is an isoform and, if so, this would be the first time an isoform of a main receptor has been identified in a marsupial. This missing D4 domain may also be a factor in the reported disease resistance of *O. fraenata*.

Historically, the immune regulatory molecule IL-2 has been difficult to find in marsupials, but in this study it was successfully characterized in PHA stimulated *M. eugenii* lymphocytes for the first time. The wallaby genome was published by Renfree *et al*. (2011) but the IL-2 gene was not annotated. When mining the genome for the promoter region of the IL-2 gene in this study, it was found that IL-2 was located on multiple scaffolds. The characteristic transcription factor binding sites contained within the promoter region were identified for the *M. eugenii* IL-2 promoter sequence. IL-2 was also characterized in *T. vulpecula* in conjunction with New Zealand collaborators thus allowing a comparison of the molecule in those two species. The IL-2 molecule in both species was investigated for SNPs, and a non-synonymous substitution was identified in the *M. eugenii* sequence but

not in the *T. vulpecula* sequence. A custom designed marsupial IL-2 antibody positively identified the IL-2 protein in a cell lysate of *M. eugenii* thymus tissue thus showing that the mRNA translated into a protein. In *O. fraenata,* attempts were made to identify IL-2 in various unstimulated tissues with qPCR but without success.

The expression of IL-2 is dependent on the calcium ion influx which is regulated by the expression of the ZAP-70 molecule. Conserved structural motifs were found within the ZAP-70 molecule in *O. fraenata*, *M. eugenii*, *L. hirsutus* and *M. domestica*. However, when the *M. eugenii* sequence was compared to the annotated sequence in ensembl it was found that 19% of the annotated sequence differed from the sequence characterized in this study over the same number of base pairs. A human anti-ZAP-70 antibody did not recognize the 60% conserved epitope present in *M. eugenii* and *O. fraenata* spleen cell lysate.

The cytokine IL-17 is the link between the adaptive and innate immune systems. This molecule was characterized in *M. eugenii* for the first time in this study. This characterization confirmed that the T_h17 sub-population is present in marsupials. A comparison of the IL-17 expressed sequence confirmed the annotated *M. eugenii* sequence in ensembl.

The T_{reg} cell population was also identified for the first time in *M. eugenii* and *O. fraenata* by characterizing a partial sequence of the T_{reg} cell surface marker Foxp3. The winged helix formation and the forkhead domain were identified thus confirming that this transcription factor is expressed in marsupials.

Phylogenetic analyses conducted for all molecules where the open reading frame was available, and in some cases for important functional motifs alone, showed consistent results. The branch lengths of *O. fraenata* and *M. eugenii* were short indicating their close relationship, while the branches of *M. domestica* were always longer indicating its evolutionary distance from the other two species. When *L. hirsutus* was included in the

analyses it clustered with the two macropods *O. fraenata* and *M. eugenii* indicating that it also belonged to the macropod family.

This study has advanced the understanding of the marsupial adaptive immune system and specifically the molecular aspects of the marsupial T cell biology. A graphical illustration of the overall advances made by this study is shown in Fig. 7.1.

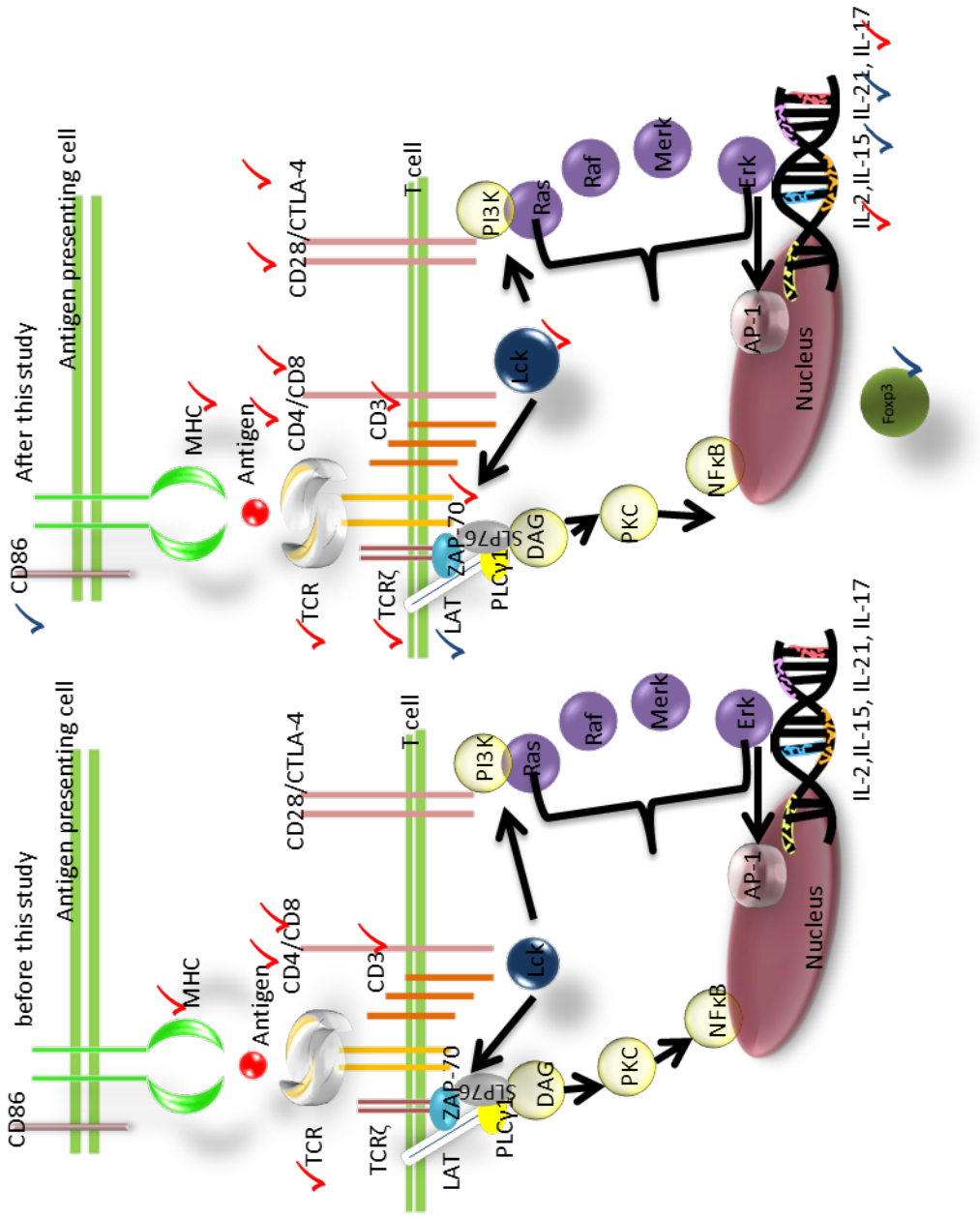


Figure 7.1. Comparison of the marsupial T cell signalling cascade prior and subsequent to this study
√= open reading frame. √ = partial sequence.

7.2 Future work

IL-2 was characterized in *M. eugenii* and *T. vulpecula* in this study but it could not be determined in *O. fraenata* due to its low expression in unstimulated tissues. Since detailed functional studies were outside the scope of this project, stimulation assays, together with qPCR, should be undertaken to determine whether or not this molecule is expressed in *O. fraenata*. Once the *O. fraenata* IL-2 sequence is determined, a comparison with the sequences in *M. eugenii* and *T. vulpecula* may reveal any differences in the *O. fraenata* IL-2 molecule which could explain the reported disease resistance of this species.

Further work should also be undertaken to determine the biological activity of IL-2 in *M. eugenii*. In order to do so, a purified IL-2 recombinant protein should be used as a T cell growth factor to ascertain if marsupial lymphocytes have similar responses to those of eutherian mammals. Real Time PCR experiments could be undertaken to determine whether or not the marsupial IL-2 also contains isoforms. If isoforms were detected, sequence comparisons could then be made between marsupial species with different immune responses to infections, and the prevalence of isoforms could be ascertained. This would help to determine whether or not isoforms of IL-2 are responsible for the observed resistance to infection in *O. fraenata*.

Very few immunological tools are available for the investigation of the marsupial immune system. The only antibodies that have been proven to be cross reactive in marsupial tissues are the human anti-CD3, anti-CD5 and anti-human CD79 antibodies. As a precursor to further protein work in marsupial immunology, additional work on trialling commercially available antibodies, from either human or other species, to identify any which may be cross reactive in marsupial tissue would be beneficial. Such cross reactive antibodies would permit more detailed analyses of protein expression and signalling pathways in marsupials using flowcytometry, T cell depletion, and other techniques.

The two phenylalanine residues with different ring configurations in the *O. fraenata* TCRα chain should be investigated to determine whether or not they are competing for space

and, if so, what implication this has for the immune competency of *O. fraenata*. Site directed mutagenesis could be used to identify any structural hindrance which will indicate whether or not the molecule functions differently to that of other mammals.

Because of its potential impact on the immune status of *O. fraenata,* the expression kinetics of the predicted monomeric CTLA-4 structure in that species should be determined and compared with that of the dimeric structure of *M. eugenii*. In particular, it should be determined if the *O. fraenata* CTLA-4 sequence characterized in this study is the only form of this molecule in that species. Real Time experiments will determine if there are other isoforms present which may affect its immune response.

The published CD4 sequence of *M. eugenii* did not show any anomalies when compared with other mammalian CD4 sequences. However, in the *O. fraenata* CD4 sequence the D4 domain was missing altogether suggesting that the *O. fraenata* CD4 may be an isoform. Further determination of what type of transcript is inherent in the *O. fraenata* CD4 molecule together with characterization of the full 5' and 3' regions of this particular form of the CD4 molecule is needed. A search for other isoforms in marsupials may determine if other transcripts, similar to those in other mammals, exist in marsupials. Additional research on the transcript presented in this thesis may confirm the significance of this variation and whether or not it translates into a functional protein.

Some *M. eugenii* molecules characterized in this study, such as CD28 and ZAP-70, showed marked differences to the annotation found in ensembl. Additional work is needed to determine if these differences are due to non-recognition of exon/intron boundaries or if there are variations in the expressed genes.

Since the marsupial species investigated in this study are endangered, the availability of fresh tissue samples was limited. Consequently, future work would ideally be conducted in association with an endangered marsupial breeding facility where fresh tissue samples may be more readily and quickly available.

References

ABE, Y., URAKAMI, H., OSTANIN, D., ZIBARI, G., HAYASHIDA, T., KITAGAWA, Y. & GRISHAM, M. B. 2009. Induction of Foxp3-expressing regulatory T-cells by donor blood transfusion is required for tolerance to rat liver allografts. *PLoS One,* 4**,** e7840.

ABRAM, C. L. & LOWELL, C. A. 2007. The expanding role for ITAM-based signaling pathways in immune cells. *Sci STKE,* 2007**,** re2.

ACUTO, O., HUSSEY, R. E., FITZGERALD, K. A., PROTENTIS, J. P., MEUER, S. C., SCHLOSSMAN, S. F. & REINHERZ, E. L. 1983. The human T cell receptor: appearance in ontogeny and biochemical relationship of alpha and beta subunits on IL-2 dependent clones and T cell tumors. *Cell,* 34**,** 717-26.

ADAMS, J. E., PENA, M. T., GILLIS, T. P., WILLIAMS, D. L., ADAMS, L. B. & TRUMAN, R. W. 2005. Expression of nine-banded armadillo (Dasypus novemcinctus) interleukin-2 in E. coli. *Cytokine,* 32**,** 219-25.

ADLER, E. M. 2007. IL-2 Antagonizes Th17 Differentiation. 2007**,** 103.

AHMED, Z., BEETON, C. A., WILLIAMS, M. A., CLEMENTS, D., BALDARI, C. T. & LADBURY, J. E. 2005. Distinct spatial and temporal distribution of ZAP70 and Lck following stimulation of interferon and T-cell receptors. *J Mol Biol,* 353**,** 1001-10.

AITIO, O., HELLMAN, M., KESTI, T., KLEINO, I., SAMUILOVA, O., PAAKKONEN, K., TOSSAVAINEN, H., SAKSELA, K. & PERMI, P. 2008. Structural basis of PxxDY motif recognition in SH3 binding. *J Mol Biol,* 382**,** 167-78.

AKASHI, M., SHAW, G., HACHIYA, M., ELSTNER, E., SUZUKI, G. & KOEFFLER, P. 1994. Number and location of AUUUA motifs: role in regulating transiently expressed RNAs. *Blood,* 83**,** 3182-7.

ALCOVER, A., MARIUZZA, R. A., ERMONVAL, M. & ACUTO, O. 1990. Lysine 271 in the transmembrane domain of the T-cell antigen receptor beta chain is necessary for its assembly with the CD3 complex but not for alpha/beta dimerization. *J Biol Chem,* 265**,** 4131-5.

ALEGRE, M. L., FRAUWIRTH, K. A. & THOMPSON, C. B. 2001. T-cell regulation by CD28 and CTLA-4. *Nat Rev Immunol,* 1**,** 220-8.

ALI, M., SALAM, N. K., AMON, M., BENDER, V., HIBBS, D. E. & MANOLIOS, N. 2006. T-Cell Antigen Receptor-alpha Chain Transmembrane Peptides: Correlation between Structure and Function. *Intl. J. of Pep. Res. and Therap.,* 12**,** 261-267.

ALONSO, A., RAHMOUNI, S., WILLIAMS, S., VAN STIPDONK, M., JAROSZEWSKI, L., GODZIK, A., ABRAHAM, R. T., SCHOENBERGER, S. P. & MUSTELIN, T. 2003. Tyrosine phosphorylation of VHR phosphatase by ZAP-70. *Nat Immunol,* 4**,** 44-8.

ALTSCHUL, S. F. & GISH, W. 1996. Local alignment statistics. *Methods in Enzymology,,* 266**,** 460 – 480.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J of Mol Biol,* 215**,** 403 – 410.

ANDERTON, S. M. 2006. Avoiding autoimmune disease--T cells know their limits. *Trends Immunol,* 27**,** 208-14.

APPAY, V., VAN LIER, R. A., SALLUSTO, F. & ROEDERER, M. 2008. Phenotype and function of human T lymphocyte subsets: consensus and issues. *Cytometry A,* 73**,** 975-83.

ARAKI, K., HIRANO, K., TAKIZAWA, F., MORITIOMO, T., OTOTAKE, M. & NAKANISHI, T. 2007. Identification and characterization of multiple *lck* genes in ginbuna crucian carp *Carassius auratus langsdorfii*. *Fisheries Science,* 73**,** 1017 - 1024.

ARAKI, K., SUETAKE, H., KIKUCHI, K. & SUZUKI, Y. 2005. Characterization and expression analysis of CD3var epsilon and CD3gamma/delta in fugu, Takifugu rubripes. *Immunogenetics,* 57**,** 158-63.

ARCARO, A., GREGOIRE, C. & BAKKER, T. 2001. CD8β endows CD8 with efficient coreceptor function by coupling TCR/CD3 to raft-associated CD8/p56lck complexes *J Exp Med,* 194**,** 1485–1495.

ARNAUD, J., CHENU, C., HUCHENQ, A., GOUAILLARD, C., KUHLMANN, J. & RUBIN, B. 1996. Defective Interactions Between TCR Chjains and CD3 Heterodimers Prevent Membrane Exprssion of TCR-αβ in Human T Cells. *J. Immunol.,* 156**,** 2155-2162.

ARNAUD, J., HUCHENQ, A., VERNHES, M. C., CASPAR-BAUGUIL, S., LENFANT, F., SANCHO, J., TERHORST, C. & RUBIN, B. 1997. The interchain disulfide bond between TCR alpha beta heterodimers on human T cells is not required for TCR-CD3 membrane expression and signal transduction. *Int Immunol,* 9**,** 615-26.

ARPAIA, E., SHAHAR, M., DADI, H., COHEN, A. & ROIFMAN, C. M. 1994. Defective T cell receptor signaling and CD8+ thymic selection in humans lacking zap-70 kinase. *Cell,* 76**,** 947-958.

ARUFFO, A. & SEED, B. 1987. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl. Acad. Sci. USA Immunology,* 84**,** 8573-8577.

ASHE, J. M., WIEST, D. L., ABE, R. & SINGER, A. 1999. ZAP-70 protein promotes tyrosine phosphorylation of T cell receptor signaling motifs (ITAMs) in immature CD4(+)8(+) thymocytes with limiting p56(lck). *J Exp Med,* 189**,** 1163-8.

ASHMAN, R., WARING, H. & STANLEY, N. F. 1975. Adult mortality after neonatal thymectomy in a marsupial, *Setonix brachyurus* (quokka). *Proc Soc Exp Biol Med Sci,* 144**,** 819-822.

ASHMAN, R. B., HOLMES, R. M. & KEAST, D. 1976. The *In Vitro* Response to Phytomitogens of Marsupial Leukocytes. *Lab. Anim. Sci.,* 26**,** 777-780.

ASHMAN, R. B., HOLMES, R. M. & KEAST, D. 1977. The effect of neonatal thymectomy on the ontogeny of mitogen responses in the quokka (Setonix brachyurus). *Dev Comp Immunol,* 1**,** 47-57.

AYYOUB, M., DEKNUYDT, F., RAIMBAUD, I., DOUSSET, C., LEVEQUE, L., BIOLEY, G. & VALMORI, D. 2009. Human memory FOXP3+ Tregs secrete IL-17 ex vivo and constitutively express the T(H)17 lineage-specific transcription factor RORgamma t. *Proc Natl Acad Sci U S A,* 106**,** 8635-40.

AZUMA, M., ITO, D., YAGITA, H., OKUMURA, K., PHILLIPS, J. H., LANIER, L. L. & SOMOZA, C. 1993. B70 antigen is a second ligand for CTLA-4 and CD28. *Nature,* 366**,** 76-9.

AZUMA, T., TAKAHASHI, T., KUNISATO, A., KITAMURA, T. & HIRAI, H. 2003. Human CD4+ CD25+ regulatory T cells suppress NKT cell functions. *Cancer Res,* 63**,** 4516-20.

BABIK, W., DURKA, W. & RADWAN, J. 2005. Sequence diversity of the MHC DRB gene in the Eurasian beaver (Castor fiber). *Mol Ecol,* 14**,** 4249-57.

BACKSTROM, B. T., HAUSMANN, B. T. & PALMER, E. 1997. Signaling efficiency of the T cell receptor controlled by a single amino acid in the beta chain constant region. *J Exp Med,* 186**,** 1933-8.

BACKSTROM, B. T., MILIA, E., PETER, A., JAUREGUIBERRY, B., BALDARI, C. T. & PALMER, E. 1996. A motif within the T cell receptor alpha chain constant region connecting peptide domain controls antigen responsiveness. *Immunity,* 5**,** 437-47.

BAKER, M. L., GEMMELL, E. & GEMMELL, R. T. 1999. Ontogeny of the Immune System of the Brushtail Possum, *Trichosurus vulpecula*. *The Anatomical Record,* 256**,** 354 - 356.

BAKER, M. L. & MILLER, R. D. 2007. Evolution of mammalian CD1: marsupial CD1 is not orthologous to the eutherian isoforms and is a pseudogene in the opossum Monodelphis domestica. *Immunology,* 121**,** 113-21.

BAKER, M. L., OSTERMAN, A. K. & BRUMBURGH, S. 2005. Divergent T-cell receptor delta chains from marsupials. *Immunogenetics,* 57**,** 665-673.

BAKER, M. L., ROSENBERG, G. H., ZUCCOLOTTO, P., HARRISON, G. A., DEANE, E. M. & MILLER, R. D. 2001. Further characterization of T cell receptor chains of marsupials. *Developmental and Comparative Immunology,* 25**,** 495 - 507.

BAKKER, T. R. & VAN DER MERWE, P. A. 2002. Self-help in T cell recognition. *Nature Immunology* 3**,** 11-12.

BALZANO, C., BUONAVISTA, N., ROUVIER, E. & GOLSTEIN, P. 1992. CTLA-4 and CD28: similar proteins, neighbouring genes. *Int J Cancer Suppl,* 7**,** 28-32.

BARNES, T. S., HINDS, L. A., JENKINS, D. J. & COLEMAN, G. T. 2007. Precocious development of hydatid cysts in a macropodid host. *Int J Parasitol,* 37**,** 1379-89.

BAROJA, M. L., LUXENBERG, D., CHAU, T., LING, V., STRATHDEE, C. A., CARRENO, B. M. & MADRENAS, J. 2000. The inhibitory function of CTLA-4 does not require its tyrosine phosphorylation. *J Immunol,* 164**,** 49-55.

BATES, P. A., KELLEY, L. A., MACCALLUM, R. M. & STERNBERG, M. J. 2001. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins,* Suppl 5**,** 39-46.

BAXTER, A. G. & HODGKIN, P. D. 2002. Activation rules: the two-signal theories of immune activation. *Nat Rev Immunol,* 2**,** 439-46.

BAZAN, J. F. 1992. Unraveling the structure of IL-2. *Science,* 257**,** 410-413.

BECKER, J. C., BRABLETZ, T., KIRCHNER, T., CONRAD, C. T., BROCKER, E. B. & REISFELD, R. A. 1995. Negative transcriptional regulation in anergic T cells. *Proc Natl Acad Sci U S A,* 92**,** 2375-8.

BELLO, R., FEITO, M. J., OJEDA, G., PORTOLES, P. & ROJO, J. M. 2009. N-terminal negatively charged residues in CD3varepsilon chains as a phylogenetically conserved trait potentially yielding isoforms with different isoelectric points: analysis of human CD3varepsilon chains. *Immunol Lett,* 126**,** 8-15.

BELOV, K., DEAKIN, J. E., PAPENFUSS, A. T., BAKER, M. L., MELMAN, S. D., SIDDLE, H. V., GOUIN, N., GOODE, D. L., SARGEANT, T. J., ROBINSON, M. D., WAKEFIELD, M. J., MAHONEY, S., CROSS, J. G. R., BENOS, P. V., SAMOLLOW, P. B., SPEED, T. P., MARSHALL GRAVES, J. A. & MILLER, R. D. 2006. Reconstructing an Ancestral Mammalian Immune Supercomplex from a Marsupial Major Histocompatibility Complex. *PLOS Biology,* 4**,** 0317 - 0328.

BELOV, K., MILLER, R. D., ILIJESKI, A., HELLMAN, L. & HARRISON, G. A. 2004. Isolation of monotreme T-cell receptor α and β chains. *Immunogenetics,* 56**,** 164 - 169.

BELOV, K., NGUYEN, M. A., ZENGER, K. R. & COOPER, D. W. 2002. Ontogeny of immunoglobulin expression in the brushtail possum (Trichosurus vulpecula). *Dev Comp Immunol,* 26**,** 599-602.

BELOV, K., SANDERSON, C. E., DEAKIN, J. E., WONG, E. S., ASSANGE, D., MCCOLL, K. A., GOUT, A., DE BONO, B., BARROW, A. D., SPEED, T. P., TROWSDALE, J. & PAPENFUSS, A. T. 2007. Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res,* 17**,** 982-91.

BENDIKSEN, S. & REKVIG, O. P. 2004. Interleukin-2, but not Interleukin-15, is required to terminate experimentally induced clonal T-Cell Anergy. *Scandinavian Journal of Immunology,* 60**,** 64–73.

BENDTSEN, J. D., NIELSEN, H., VON HEIJNE, G. & BRUNAK, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol,* 340**,** 783-95.

BERIOU, G., COSTANTINO, C. M., ASHLEY, C. W., YANG, L., KUCHROO, V. K., BAECHER-ALLAN, C. & HAFLER, D. A. 2009. IL-17-producing human peripheral regulatory T cells retain suppressive function. *Blood,* 113**,** 4240-9.

BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The Protein Data Bank. *Nucleic Acids Res,* 28**,** 235-42.

BERNARD, D., HANSEN, J. D., DU PASQUIER, L., LEFRANC, M.-P., BENMANSOUR, A. & BOUDINOT, P. 2006. Costimulatory receptors in jawed vertebrates: Conserved CD28, odd CTLA4 and multiple BTLAs.

BERRY, A., TECTOR, M. & OAKS, M. K. 2008. Lack of association between sCTLA-4 levels in human plasma and common CTLA-4 polymorphisms. *J Negat Results Biomed,* 7**,** 8.

BETTS, M. J. & RUSSELL, R. B. 2003. Amino Acid properties and consequences of substitutions. *Bioinformatics for Geneticists,* ISBN 0-470-84393-4.

BEVERLY, B., KANG, S. M., LENARDO, M. J. & SCHWARTZ, R. H. 1992. Reversal of in vitro T cell clonal anergy by IL-2 stimulation. *Int Immunol,* 4**,** 661-71.

BEYERSMANN, D. & HAASE, H. 2001. Functions of zinc in signaling, proliferation and differentiation of mammalian cells. *Biometals,* 14**,** 331-41.

BHATNAGAR, A., GULL;AND, S., BASCAND, M., PALMER, E., GARDNER, T. G., KEARSE, K. P. & BACKSTROM, B. T. 2003. Mutational analysis of conserved amino acids in the T cell receptor α–chain transmembrane region:a critical role of leucine 112 and phenylalanine 127 for assembly and surface expression. *Molecular Immunology,* 39**,** 953-963.

BINGULAC-POPOVIC, J., FIGUEROA, F., SATO, A., TALBOT, W. S., JOHNSON, S. L., GATES, M., POSTLETHWAIT, J. H. & KLEIN, J. 1997. Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, Danio rerio. *Immunogenetics,* 46**,** 129-34.

BIRD, S., ZOU, J., KONO, T., SAKAI, M., DIJKSTRA, J. M. & SECOMBES, C. 2005a. Characterisation and expression analysis of interleukin 2 (IL-2) and IL-21 homologues in the Japanese pufferfish, Fugu rubripes, following their discovery by synteny. *Immunogenetics,* 56**,** 909–923.

BIRD, S., ZOU, J., KONO, T., SAKAI, M., DIJKSTRA, J. M. & SECOMBES, C. 2005b. Characterisation and expression analysis of interleukin 2 (IL-2) and IL-21 homologues in the Japanese pufferfish, Fugu rubripes, following their discovery by synteny. *Immunogenetics,* 56**,** 909-23.

BJELLQVIST, B., HUGHES, G. J., PASQUALI, C., PAQUET, N., RAVIER, F., SANCHEZ, J. C., FRUTIGER, S. & HOCHSTRASSER, D. 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis,* 14**,** 1023-31.

BLANCHARD, N., DI BARTOLO, V. & HIVROZ, C. 2002. In the immune synapse, ZAP-70 controls T cell polarization and recruitment of signaling proteins but not formation of the synaptic pattern. *Immunity,* 17**,** 389-99.

BLEDI, Y., INBERG, A. & LINIAL, M. 2003. PROCEED: A proteomic method for analysing plasma membrane proteins in living mammalian cells. *Brief Funct Genomic Proteomic,* 2**,** 254-65.

BLOM, N., GAMMELTOFT, S. & BRUNAK, S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol,* 294**,** 1351-62.

BLUESTONE, J. A. 1997. Is CTLA-4 a master switch for peripheral T cell tolerance? *J Immunol,* 158**,** 1989-93.

BOISE, L. H., MINN, A. J., NOEL, P. J., JUNE, C. H., ACCAVITTI, M. A., LINDSTEN, T. & THOMPSON, C. B. 1995. CD28 costimulation can promote T cell survival by enhancing the expression of Bcl-XL. *Immunity,* 3**,** 87-98.

BOLLIGER, L. & JOHANSSON, B. 1999. Identification and functional characterization of the zeta-chain dimerization motif for TCR surface expression. *J Immunol,* 163**,** 3867-76.

BORDERS, C. L., JR., BROADWATER, J. A., BEKENY, P. A., SALMON, J. E., LEE, A. S., ELDRIDGE, A. M. & PETT, V. B. 1994. A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci,* 3**,** 541-8.

BORN, W. K., REARDON, C. L. & O'BRIEN, R. L. 2006. The function of γδ T cells in innate immunity. Curr Opin Immunol. *Curr Opin Immunol,* 18**,** 21-38.

BORNBERG-BAUER, E., RIVALS, E. & VINGRON, M. 1998. Computational approaches to identify leucine zippers. *Nucleic Acids Res,* 26**,** 2740-6.

BORRIELLO, F., OLIVEROS, J., FREEMAN, G. J., NADLER, L. M. & SHARPE, A. H. 1995. Differential expression of alternate mB7-2 transcripts. *J Immunol,* 155**,** 5490-7.

BORROTO, A., LAMA, J., NIEDERGANG, F., DAUTRY-VARSAT, A., ALARCON, B. & ALCOVER, A. 1999. The CD3 epsilon subunit of the TCR contains endocytosis signals. *J Immunol,* 163**,** 25-31.

BOSCARIOL, R., PLEASANCE, J., PIEDRAFITA, D. M., RAADSMA, H. W. & SPITHILL, T. W. 2006. Identification of two allelic forms of ovine CD4 exhibiting a Ser183/Pro183 polymorphism in the coding sequence of domain 3. *Vet Immunol Immunopathol,* 113**,** 305-12.

BOSSELUT, R. 2004. CD4/CD8 - Lineage differentiation in the thymus: from nuclear effectors to membrane signals. *Nature Reviews Immunology,* 4**,** 529 - 540.

BOSSELUT, R., KUBO, S., GUINTER, T., KOPACZ, J. L., ALTMAN, J. D., FEIGENBAUM, L. & SINGER, A. 2000. Role of CD8beta domains in CD8 coreceptor function: importance for MHC I binding, signaling, and positive selection of $CD8^+$ T cells in the thymus. *Immunity,* 12**,** 409-418.

BOSSELUT, R., ZHANG, W., ASHE, J. M., KOPACZ, J. L., SAMELSON, L. E. & SINGER, A. 1999. Association of the adaptor molecule LAT with CD4 and CD8 coreceptors identifies a new coreceptor function in T cell receptor signal transduction. *J Exp Med,* 190**,** 1517-26.

BOYD, A. M., HALL, R. A., GEMMELL, R. T. & KAY, B. H. 2001. Experimental infection of Australian brushtail possums, Trichosurus vulpecula (Phalangeridae: Marsupialia), with Ross River and Barmah Forest viruses by use of a natural mosquito vector system. *Am J Trop Med Hyg,* 65**,** 777-82.

BRADSHAW, J. D., LU, P., LEYTZE, G., RODGERS, J., SCHIEVEN, G. L., BENNETT, K. L., LINSLEY, P. S. & KURTZ, S. E. 1997. Interaction of the cytoplasmic tail of CTLA-4 (CD152) with a clathrin-associated protein is negatively regulated by tyrosine phosphorylation. *Biochemistry,* 36**,** 15975-82.

BRAND, S. 2009. Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. *Gut,* 58**,** 1152-67.

BRDICKA, T., KADLECEK, T. A., ROOSE, J. P., PASTUSZAK, A. W. & WEISS, A. 2005. Intramolecular regulatory switch in ZAP-70: analogy with receptor tyrosine kinases. *Mol Cell Biol,* 25**,** 4924-33.

BRIESE, L. & WILLBOLD, D. 2003. Structure determination of human Lck unique and SH3 domains by nuclear magnetic resonance spectroscopy. *BMC,* 3.

BRODEUR, J. F., LI, S., MARTINS MDA, S., LAROSE, L. & DAVE, V. P. 2009. Critical and multiple roles for the CD3epsilon intracytoplasmic tail in double negative to double positive thymocyte differentiation. *J Immunol,* 182**,** 4844-53.

BROZEK, C. M., KALETA, E. W., KUSEWITT, D. F. & LEY, R. D. 1992. Proliferative responses of lymphocytes to mitogens in the gray, short-tailed opossum, Monodelphis domestica. *Vet Immunol Immunopathol,* 31**,** 11-9.

BROZEK, C. M. & LEY, R. D. 1991. Production of interleukin-1 in a South American opossum (Monodelphis domestica). *Dev Comp Immunol,* 15**,** 401-12.

BRYSON, K., MCGUFFIN, L. J., MARSDEN, R. L., WARD, J. J., SODHI, J. S. & JONES, D. T. 2005. Protein structure prediction servers at University College London. *Nucleic Acids Res,* 33**,** W36-8.

BUDDLE, B. M. & YOUNG, L. J. 2000. Immunobiology of mycobacterial infections in marsupials. *Dev Comp Immunol,* 24**,** 517-529.

BUGEON, L. & DALLMAN, M. J. 2000. Costimulation of T cells. *Am J Respir Crit Care Med,* 162**,** S164-8.

BUONOCORE, F., RANDELLI, E., BIRD, S., SECOMBES, C. J., COSTANTINI, S., FACCHIANO, A., MAZZINI, M. & SCAPIGLIATI, G. 2006. The CD8alpha from sea bass (Dicentrarchus labrax L.): Cloning, expression and 3D modelling. *Fish Shellfish Immunol,* 20**,** 637-46.

BURBACH, B. J., MEDEIROS, R. B., MUELLER, K. L. & SHIMIZU, Y. 2007. T-cell receptor signaling to integrins. *Immunol Rev,* 218**,** 65-81.

CALL, M. E., PYRDOL, J., WIEDMANN, M. & WUCHERPFENNIG, K. W. 2002. The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell,* 111**,** 967-79.

CALL, M. E., PYRDOL, J. & WUCHERPFENNIG, K. W. 2004. Stoichiometry of the T-cell receptor-CD3 complex and key intermediates assembled in the endoplasmic reticulum. *EMBO,* 23**,** 2348 - 2357.

CALL, M. E., SCHNELL, J. R., XU, C., LUTZ, R. A., CHOU, J. J. & WUCHERPFENNIG, K. W. 2006. The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor. *Cell,* 127**,** 355-68.

CAMPBELL, D. J. & ZIEGLER, S. F. 2007. FOXP3 modifies the phenotypic and functional properties of regulatory T cells. *Nat Rev Immunol,* 7**,** 305-10.

CANADAY, D. H., ZIEBOLD, C., NOSS, E. H., CHERVENAK, K. A., HARDING, C. V. & BOOM, W. H. 1999. Activation of human CD8+ alpha beta TCR+ cells by Mycobacterium tuberculosis via an alternate class I MHC antigen-processing pathway. *J Immunol,* 162**,** 372-9.

CANFIELD, P., HEMSLEY, S. & CONNOLLY, J. 1996. Histological and immunohistological study of the developing and involuting superficial cervical thymus in the koala (Phascolarctos cinereus). *J Anat,* 189 (Pt 1)**,** 159-69.

CARRERA, A. C., PARADIS, H., BORLADO, L. R., ROBERTS, T. M. & MARTINEZ, C. 1995. Lck unique domain influences Lck specificity and biological function. *J Biol Chem,* 270**,** 3385 - 3391.

CASTRO-PRIETO, A., WACHTER, B. & SOMMER, S. 2011. Cheetah paradigm revisited: MHC diversity in the world's largest free-ranging population. *Mol Biol Evol,* 28**,** 1455-68.

CHAMBERS, C. A. 2001. The expanding world of co-stimulation: the two-signal model revisited. *Trends Immunol,* 22**,** 217-23.

CHAMIZO, C., RUBIO, J. M., MORENO, J. & ALVAR, J. 2001. Semi-quantitative analysis of multiple cytokines in canine peripheral blood mononuclear cells by [correction of zby] a single tube RT-PCR. *Vet. Immunol. Immunopathol.,* 83**,** 191-202.

CHAN, A. C., DALTON, M., JOHNSON, R., KONG, G. H., WANG, T., THOMA, R. & KUROSAKI, T. 1995. Activation of ZAP-70 kinase activity by phosphorylation of tyrosine 493 is required for lymphocyte antigen receptor function. *EMBO J,* 14**,** 2499-508.

CHAN, A. C., IWASHIMA, M., TURCK, C. W. & WEISS, A. 1992. ZAP-70: a 70 kd protein-tyrosine kinase that associates with the TCR zeta chain. *Cell,* 71**,** 649-662.

CHAN, C. P., MAK, T. Y., CHIN, K. T., NG, I. O. & JIN, D. Y. 2010. N-linked glycosylation is required for optimal proteolytic activation of membrane-bound transcription factor CREB-H. *J Cell Sci,* 123**,** 1438-48.

CHAN, J., ROBINSON, E. S., YEH, I. T. & MCCARREY, J. R. 2002. Absence of ras gene mutations in UV-induced malignant melanomas correlates with a dermal origin of melanocytes in Monodelphis domestica. *Cancer Lett,* 184**,** 73-80.

CHAPLIN, P. J., PIETRALA, L. N. & SCHEERLINCK, J.-P. Y. 1999. Cloning and sequence comparison of sheep CD28 and CTLA-4. *Immunogenetics,* 49**,** 583-584.

CHARLES, N., HARDWICK, D., DAUGAS, E., ILLEI, G. G. & RIVERA, J. 2010. Basophils and the T helper 2 environment can promote the development of lupus nephritis. *Nat Med,* 16**,** 701-7.

CHATILA, T. A., BLAESER, F., HO, N., LEDERMAN, H. M., VOULGAROPOULOS, C., HELMS, C. & BOWCOCK, A. M. 2000. JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation syndrome. *J Clin Invest,* 106**,** R75-81.

CHEN, C., GAULT, A., SHEN, L. & NABAVI, N. 1994. Molecular cloning and expression of early T cell costimulatory molecule-1 and its characterization as B7-2 molecule. *J Immunol,* 152**,** 4929-36.

CHEN, C., ROWELL, E. A., THOMAS, R. M., HANCOCK, W. W. & WELLS, A. D. 2006. Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. *J Biol Chem,* 281**,** 36828-34.

CHEN, C. Y. & SHYU, A. B. 1995. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci,* 20**,** 465-70.

CHEN, F., LEE, Y., JIANG, Y., WANG, S., PEATMAN, E., ABERNATHY, J., LIU, H., LIU, S., KUCUKTAS, H., KE, C. & LIU, Z. 2010. Identification and characterization of full-length cDNAs in channel catfish (Ictalurus punctatus) and blue catfish (Ictalurus furcatus). *PLoS One,* 5**,** e11546.

CHERNICKY, C. L., TAN, H., BURFEIND, P. & ILAN, J. 1996. Sequence of interleukin-2 isolated from human placental poly A+ RNA: possible role in maintenance of fetal allograft. *Mol Reprod Dev,* 43**,** 180-6.

CHIDA, A. S., GOYOS, A. & ROBERT, J. 2011. Phylogenetic and developmental study of CD4, CD8 alpha and beta T cell co-receptor homologs in two amphibian species, Xenopus tropicalis and Xenopus laevis. *Dev Comp Immunol,* 35**,** 366-77.

CHOI, I., CHO, B., KIM, S. D., PARK, D., KIM, J. Y., PARK, C. G., CHUNG, D. H., HWANG, W. S., LEE, J. S. & AHN, C. 2006. Molecular cloning, expression and functional characterization of miniature swine CD86. *Mol Immunol,* 43**,** 480-6.

CHOI, I. S., HASH, S. M. & COLLISSON, E. W. 2000a. Molecular cloning and expression of feline CD28 and CTLA-4 cDNA. *Vet Immunol Immunopathol,* 76**,** 45-59.

CHOI, I. S., HASH, S. M., WINSLOW, B. J. & COLLISSON, E. W. 2000b. Sequence analyses of feline B7 costimulatory molecules. *Vet Immunol Immunopathol,* 73**,** 219-31.

CHOTHIA, C., NOVOTNY, J., BRUCCOLERI, R. & KARPLUS, M. 1985. Domain association in immunoglobulin molecules. The packing of variable domains. *J Mol Biol,* 186**,** 651-63.

CHOWDHURY, B., TSOKOS, C. G., KRISHNAN, S., ROBERTSON, J., FISHER, C. U., WARKE, R. G., WARKE, V. G., NAMBIAR, M. P. & TSOKOS, G. C. 2005. Decreased stability and translation of T cell receptor zeta mRNA with an alternatively spliced 3'-untranslated region contribute to zeta chain down-regulation in patients with systemic lupus erythematosus. *J Biol Chem,* 280**,** 18959-66.

CHRETIEN, I., MARCUZ, A., FELLAH, J., CHARLEMAGNE, J. & DU PASQUIER, L. 1997. The T cell receptor beta genes of Xenopus. *Eur J Immunol,* 27**,** 763-71.

CHUANG, E., ALEGRE, M. L., DUCKETT, C. S., NOEL, P. J., VANDER HEIDEN, M. G. & THOMPSON, C. B. 1997. Interaction of CTLA-4 with the clathrin-associated protein AP50 results in ligand-independent endocytosis that limits cell surface expression. *J Immunol,* 159**,** 144-51.

CISTERNAS, P. A. & ARMATI, P. J. 2000. Immune system cell markers in the northern brown bandicoot, *Isoodon macrourus*. *Dev Comp Immunol,* 24**,** 771-82.

CLARK, G. J. & DALLMAN, M. J. 1992. Identification of a cDNA encoding the rat CD28 homologue. *Immunogenetics,* 35**,** 54-7.

CLARK, S. J., JEFFREIES, W. A., BARCLAY, A. N., GAGNON, J. & WILLIAMS, A. F. 1987. Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for derivation from a structure with four immunoglobulin-related domains. *Proc Natl Acad Sci USA,* 84.

CLASSON, B. J., TSAGARATOS, J., MCKENZIE, I. F. C. & WALKER, I. D. 1986. Partial primary structure of the T4 antigens of mouse and sheep: Assignment of intrachain disulfide bonds. *Proc Natl Acad Sci USA,* 83**,** 4499 -4503.

CLAYTON, L. K., D'ADAMIO, L., HOWARD, F. D., SIEH, M., HUSSEY, R. E., KOYASU, S. & REINHERZ, E. L. 1991. CD3 eta and CD3 zeta are alternatively spliced products of a common genetic locus and are transcriptionally and/or post-transcriptionally regulated during T-cell development. *Proc Natl Acad Sci U S A,* 88**,** 5202-6.

CLEMENTS, C. S., DUNSTONE, M. A., MACDONALD, W. A., MCCLUSKEY, J. & ROSSJOHN, J. 2006. Specificity on a knife-edge: the alphabeta T cell receptor. *Curr Opin Struct Biol,* 16**,** 787-95.

CLEVERS, H., ALARCON, B., WILEMAN, T. & TERHORST, C. 1988a. The T cell receptor/CD3 complex: a dynamic protein ensemble. *Ann. Rev. Immunol.,* 6**,** 629.

CLEVERS, H. C., DUNLAP, S., WILEMAN, T. & TERHORST, C. 1988b. Human CD3-ε gene contains three miniexons and is transcribed from a non-TATA promoter. *Proc Natl. Acad Sci USA,* 85**,** 8156 - 8160.

COHEN, G. B., REN, R. & BALTIMORE, D. 1995. Modular binding domains in signal transduction proteins. *Cell,* 80**,** 237-48.

COLE, C., BARBER, J. D. & BARTON, G. J. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res,* 36**,** W197-201.

COLE, D. K., PUMPHREY, N. J., BOULTER, J. M., SAMI, M., BELL, J. I., GOSTICK, E., PRICE, D. A., GAO, G. F., SEWELL, A. K. & JAKOBSEN, B. K. 2007. Human TCR-binding affinity is governed by MHC class restriction. *J Immunol,* 178**,** 5727-34.

CONNOLLY, J. H., CANFIELD, P. J., MCCLURE, S. J. & WHITTINGTON, R. J. 1999. Histological and immunohistological investigation of lymphoid tissue in the platypus (Ornithorhynchus anatinus). *J Anat,* 195 (Pt 2)**,** 161-71.

COUTINHO, H. B., SEWELL, H. F., TIGHE, P., KING, G., NOGUEIRA, J. C., ROBALINHO, T. I., COUTINHO, V. B. & CAVALCANTI, V. M. 1995. Immunocytochemical study of the ontogeny of the marsupial Didelphis albiventris immune system. *J Anat,* 187 (Pt 1)**,** 37-46.

COZZI, P. J., PADRID, P. A., TAKEDA, J., ALEGRE, M. L., YUHKI, N. & LEFF, A. R. 1993. Sequence and functional characterization of feline interleukin 2. *Biochem. Biophys. Res. Commun. ,* 194**,** 1038-1043.

CRESPO,M., VILLAMOR, N., GINE, E., MUNTANOLA, A., COLOMER, D., MARAFIOTI, T.JONES, M., CAMOS, M., CAMPO, E., MONTSERRAT, E., BOSCH,F. 2006. ZAP-70 expression in normal pro/pre B lymphocytes, mature B lymphocytes, and in B-cell acute lymphoblastic leukemia. *Clin Cancer Res,* 12, 3, 726-734.

CRISCITIELLO, M. F., OHTA, Y., SALTIS, M., MCKINNEY, E. C. & FLAJNIK, M. F. 2010. Evolutionarily conserved TCR binding sites, identification of T cells in primary lymphoid tissues, and surprising trans-rearrangements in nurse shark. *J. Immunol.,* 184**,** 6950–6960.

CRISCITIELLO, M. F., SALTIS, M. & FLAJNIK, M. F. 2006. An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. *Proc Natl Acad Sci U S A,* 103**,** 5036-5041.

CUI, S., HOPE, R. M., RATHJEN, J., VOYLE, R. B. & RATHJEN, P. D. 2001. Structure, sequence and function of a marsupial LIF gene: conservation of IL-6 family cytokines. *Cytogenet Cell Genet,* 92**,** 271-8.

CUI, S. & SELWOOD, L. 2000. cDNA cloning characterization, expression and recombinant protein production of leukemia inhibitory factor (LIF) from the marsupial the brushtail possum (Trichosurus vulpecula). *Gene,* 243**,** 167 - 178.

CUROTTO DE LAFAILLE, M. A. & LAFAILLE, J. J. 2009. Natural and adaptive foxp3+ regulatory T cells: more of the same or a division of labor? *Immunity,* 30**,** 626-35.

CURRIER, J. R., DESOUZA, M., CHANBANCHERD, P., BERNSTEIN, W., BIRX, D. L. & COX, J. H. 2002. Comprehensive screening for human immunodeficiency virus type 1 subtype-specific CD8 cytotoxic T lymphocytes and definition of degenerate epitopes restricted by HLA-A0207 and -C(W)0304 alleles. *J Virol,* 76**,** 4971-86.

CUTHBERTSON, J. M., DOYLE, D. A. & SANSOM, M. S. 2005. Transmembrane helix prediction: a comparative evaluation and analysis. *Protein Eng Des Sel,* 18**,** 295-308.

DAVE, V. P. 2009. Hierarchical role of CD3 chains in thymocyte development. *Immunol Rev*, 232, 1 22-33.

DAVIS, M. M. 2002. A new trigger for T cells. *Cell,* 110**,** 285-7.

DAYHOFF, M. O., SCHWARTZ, R. M. & ORCUTT, B. C. 1978. A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure* 5**,** 345-352.

DE LA CRUZ, J., KRUGER, T., PARKS, C. A., SILGE, R. L., VAN OERS, N. S., LUESCHER, I. F., SCHRUM, A. G. & GIL, D. 2011. Basal and antigen-induced exposure of the proline-rich sequence in CD3epsilon. *J Immunol,* 186**,** 2282-90.

DEAN, E. M. & COOPER, D. W. 1998. Immunolgical development of pouch young marsupials. In: Tyndale-Briscoe CH., Jansens, PA., editors. The developing marsupial. Models for biomedical research. *Berlin: Springer***,** 190-199.

DEEN, K. C., MCDOUGAL, J. S., INACKER, R., FOLENA-WASSERMAN, G., ARTHOS, J., ROSENBERG, J., MADDON, P. J., AXEL, R. & SWEET, R. W. 1988. A soluble form of CD4 (T4) protein inhibits AIDS virus infection. *Nature,* 331**,** 82-4.

DEGERMANN, S., SOLLAMI, G. & KARJALAINEN, K. 1999a. Impaired NK1.1 T cell development in mice transgenic for a T cell receptor beta chain lacking the large, solvent-exposed cbeta FG loop. *J Exp Med,* 190**,** 1357-62.

DEGERMANN, S., SOLLAMI, G. & KARJALAINEN, K. 1999b. T cell receptor beta chain lacking the large solvent-exposed Cbeta FG loop supports normal alpha/beta T cell development and function in transgenic mice. *J Exp Med,* 189**,** 1679-84.

DEINDL, S., KADLECEK, T., BRDLICKA, T., CAO, X. & WEIS, A. 2007. Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. *Cell,* 129**,** 735-746.

DEL RIO, G., KANE, D. J., BALL, K. D. & BREDESEN, D. E. 2007. A Novel Motif Identified in Dependence Receptors. *PLoS ONE,* 2.

DEVINE, L., KIEFFER, L. J., AITKEN, V. & KAVATHAS, P. B. 2000. Human CD8 beta, but not mouse CD8 beta, can be expressed in the absence of CD8 alpha as a beta beta homodimer. *J Immunol,* 164**,** 833-8.

DEVINE, L., SUN, J., BARR, M. & KAVATHAS, P. 1999. Orientation of the Ig domains of CD8 alpha beta relative to MHC class I. *J. Immunol.,* 162**,** 846-851.

DI BARTOLO, V., MEGE, D., GERMIAN, V., PELOSI, M., DUFOUR, E., MICHEL, F., MAGISTRELLI, G., ISAACCHI, A. & ACUTO, O. 1999. Tyrosine 319, a newly identified phosphorylationsite of ZAP-70, plays a critical role in T cell antigen receptor signaling. *J Biol Chem,* 274**,** 6285-6294.

DI TOMMASO, S., ANTONACCI, R., CICCARESE, S. & MASSARI, S. 2010. Extensive analysis of D-J-C arrangements allows the identification of different mechanisms enhancing the diversity in sheep T cell receptor beta-chain repertoire. *BMC Genomics,* 11**,** 3.

DIDELOT, X. 2010. Sequence-Based Analysis of Bacterial Population Structures. In D. Ashley Robinson, Daniel Falush, Edward J. Feil. Bacterial Population Genetics in Infectious Disease. . *John Wiley and Sons,* ISBN 978-0-470-42474-2**,** p. 46–47.

DONG, S., CORRE, B., NIKA, K., PELLEGRINI, S. & MICHEL, F. 2010. T cell receptor signal initiation induced by low-grade stimulation requires the cooperation of LAT in human T cells. *PLoS One,* 5**,** e15114.

DRANOFF, G. 2004. Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer,* 4**,** 11-22.

DRESZER, T. R., KAROLCHIK, D., ZWEIG, A. S., HINRICHS, A. S., RANEY, B. J., KUHN, R. M., MEYER, L. R., WONG, M., SLOAN, C. A., ROSENBLOOM, K. R., ROE, G., RHEAD, B., POHL, A., MALLADI, V. S., LI, C. H., LEARNED, K., KIRKUP, V., HSU, F., HARTE, R. A., GURUVADOO, L., GOLDMAN, M., GIARDINE, B. M., FUJITA, P. A., DIEKHANS, M., CLINE, M. S., CLAWSON, H., BARBER, G. P., HAUSSLER, D. & JAMES KENT, W. 2012. The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res,* 40**,** D918-23.

DUCHARDT, E., SIGALOV, A. B., AIVAZIAN, D., STERN, L. J. & SCHWALBE, H. 2007. Structure induction of the T-cell receptor zeta-chain upon lipid binding investigated by NMR spectroscopy. *Chembiochem,* 8**,** 820-7.

DUNCAN, L. G., NAIR, S. V. & DEANE, E. M. 2007. Molecular characterisation and expression of CD4 in two distantly related marsupials: The gray short-tailed opossum (*Monodelphis domestica*) and tammar wallaby (*Macropus eugenii)*. *Mol Immunol,* 44**,** 3641-3652.

DUNCAN, L. G., NAIR, S. V. & DEANE, E. M. 2008. The marsupial CD8 gene locus: Molecular cloning and expression analysis of the alpha and beta sequences in the gray short-tailed opossum (*Monodelphis domestica*) and the tammar wallaby *(Macropus eugenii*). *Vet Immunol Immunopathol*.

ECK, M. J., ATWELL, J. K., SHOELSON, S. E. & HARRISON, S. H. 1994. Structure of the regulatory domains of the Src-family tyrosine kinase Lck. *Nature,* 368**,** 764-769.

EDHOLM, E. S., STAFFORD, J. L., QUINIOU, S. M., WALDBIESER, G., MILLER, N. W., BENGTEN, E. & WILSON, M. 2006. Channel catfish, *Ictalurus punctatus*, CD4-like molecules. *Developmental and Comparative Immunology*.

EL-RIDI, R., WAHBY, A. F. & SAAD, A.-H. 1986. Characterization of snake Interleukin-2. *Dev Comp Immunol,* 10**,** 128.

EMAMAULLEE, J. A., DAVIS, J., MERANI, S., TOSO, C., ELLIOTT, J. F., THIESEN, A. & SHAPIRO, A. M. 2009. Inhibition of Th17 cells regulates autoimmune diabetes in NOD mice. *Diabetes,* 58**,** 1302-11.

ESWAR, N., JOHN, B., MIRKOVIC, N., FISER, A., ILYIN, V. A., PIEPER, U., STUART, A. C., MARTI-RENOM, M. A., MADHUSUDHAN, M. S., YERKOVICH, B. & SALI, A. 2003. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res,* 31**,** 3375-80.

ESWAR, N., WEBB, B., MARTI-RENOM, M. A., MADHUSUDHAN, M. S., ERAMIAN, D., SHEN, M. Y., PIEPER, U. & SALI, A. 2006. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics,* Chapter 5**,** Unit 5 6.

EVANS, E. J., ESNOUF, R. M., MANSO-SANCHO, R., GILBERT, R. J., JAMES, J. R., YU, C., FENNELLY, J. A., VOWLES, C., HANKE, T., WALSE, B., HUNIG, T., SORENSEN, P., STUART, D. I. & DAVIS, S. J. 2005. Crystal structure of a soluble CD28-Fab complex. *Nat Immunol,* 6**,** 271-9.

FALLAH-ARANI, F., SCHWEIGHOFFER, E., VANES, L. & TYBULEWICZ, V. L. 2008. Redundant role for Zap70 in B cell development and activation. *Eur J Immunol,* 38**,** 1721-33.

FALLARINO, F., GROHMANN, U., HWANG, K. W., ORABONA, C., VACCA, C., BIANCHI, R., BELLADONNA, M. L., FIORETTI, M. C., ALEGRE, M. L. & PUCCETTI, P. 2003. Modulation of tryptophan catabolism by regulatory T cells. *Nat Immunol,* 4**,** 1206-12.

FALLARINO, F., GROHMANN, U., YOU, S., MCGRATH, B. C., CAVENER, D. R., VACCA, C., ORABONA, C., BIANCHI, R., BELLADONNA, M. L., VOLPI, C., SANTAMARIA, P., FIORETTI, M. C. & PUCCETTI, P. 2006. The combined effects of tryptophan starvation and tryptophan catabolites down-regulate T cell receptor zeta-chain and induce a regulatory phenotype in naive T cells. *J Immunol,* 176**,** 6752-61.

FARES, F. 2006. The role of O-linked and N-linked oligosaccharides on the structure-function of glycoprotein hormones: development of agonists and antagonists. *Biochim Biophys Acta,* 1760**,** 560-7.

FEHÉRVARI, Z. & SAKAGUCHI, S. 2004. CD4+ Tregs and immune control. *The Journal of Clinical Investigation,* 114

FELLAH, J. S., KERFOURN, F., GUILLET, F. & CHARLEMAGNE, J. 1993. Conserved structure of amphibian T cell antigen receptor β chain. *Proc Natl Acad Sci USA,* 90**,** 6811 -6814.

FELLAH, J. S., TUFFERY, P., ETCHEBEST, C., GUILLET, F., BLEUX, C. & CHARLEMAGNE, J. 2002. Cloning and modeling of CD8 beta in the amphibian ambystoma Mexicanum. Evolutionary conserved structures for interactions with major histocompatibility complex (MHC) class I molecules. *Gene,* 288**,** 95-102.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. . *Evolution,* 39.

FERRE, F. & CLOTE, P. 2005. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res,* 33**,** W230-2.

FISHER, R. A., BERTONIS, J. M., MEIER, W., JOHNSON, V. A., COSTOPOULOS, D. S., LIU, T., TIZARD, R., WALKER, B. D., HIRSCH, M. S., SCHOOLEY, R. T. & ET AL. 1988. HIV infection is blocked in vitro by recombinant soluble CD4. *Nature,* 331**,** 76-8.

FLOESS, S., FREYER, J., SIEWERT, C., BARON, U., OLEK, S., POLANSKY, J., SCHLAWE, K., CHANG, H.-D., BOPP, T., SCHMITT, W., KLEIN-HESSLING, S., SERFLING, E. & HARMANN, A. 2007. Epigenetic Control of the foxp3 Locus in Regulatory T Cells. *PLoS Biol,* 5.

FLOREA, L., DI FRANCESCO, V., MILLER, J., TURNER, R., YAO, A., HARRIS, M., WALENZ, B., MOBARRY, C., MERKULOV, G. V., CHARLAB, R., DEW, I., DENG, Z., ISTRAIL, S., LI, P. & SUTTON, G. 2005. Gene and alternative splicing annotation with AIR. *Genome Res,* 15**,** 54-66.

FOMSGAARD, A., HIRSCH, V. M. & R., J. P. 1992. Cloning and sequences of primate CD4 molecules: Diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus. *European Journal of Immunology,* 22**,** 2973 - 2981.

FONTENOT, J. D., GAVIN, M. A. & RUDENSKY, A. Y. 2003. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol,* 4**,** 330-6.

FONTENOT, J. D., RASMUSSEN, J. P., GAVIN, M. A. & RUDENSKY, A. Y. 2005a. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nat Immunol,* 6**,** 1142-51.

FONTENOT, J. D., RASMUSSEN, J. P., WILLIAMS, L. M., DOOLEY, J. L., FARR, A. G. & RUDENSKY, A. Y. 2005b. Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity,* 22**,** 329-41.

FOSSIEZ, F., BANCHEREAU, J., MURRAY, R., VAN KOOTEN, C., GARRONE, P. & LEBECQUE, S. 1998. Interleukin -17. *Int. Rev. Immunol.,* 16**,** 541 - 551.

FOSSIEZ, F., DJOSSOU, O., CHOMARAT, P., FLORES-ROMO, L., AIT-YAHIA, S., MAAT, C., PIN, J. J., GARRONE, P., GARCIA, E., SAELAND, S., BLANCHARD, D., GAILLARD, C., DAS MAHAPATRA, B., ROUVIER, E., GOLSTEIN, P., BANCHEREAU, J. & LEBECQUE, S. 1996. T cell interleukin-17 induces stromal cells to produce proinflammatory and hematopoietic cytokines. *J Exp Med,* 183**,** 2593-603.

FRANK, S. J., NIKLINSKA, B. B., ORLOFF, D. G., MERCEP, M., ASHWELL, J. D. & KLAUSNER, R. D. 1990. Structural mutations of the T cell receptor zeta chain and its role in T cell activation. *Science,* 249**,** 174-7.

FREEMAN, A. R., LYNN, D. J., MURRAY, C. & BRADLEY, D. G. 2008. Detecting the effects of selection at the population level in six bovine immune genes. . *BMC Genet,* 9.

FREEMAN, G. J., LOMBARD, D. B., GIMMI, C. D., BROD, S. A., LEE, K., LANING, J. C., HAFLER, D. A., DORF, M. E., GRAY, G. S., REISER, H. & ET AL. 1992. CTLA-4 and CD28 mRNA are coexpressed in most T cells after activation. Expression of CTLA-4 and CD28 mRNA does not correlate with the pattern of lymphokine production. *J Immunol,* 149**,** 3795-801.

FREEMAN, J. D., WARREN, R. L., WEBB, J. R., NELSON, B. H. & HOLT, R. A. 2009. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res,* 19**,** 1817-24.

FUJII, Y., MATSUTANI, T., KITAURA, K., SUZUKI, S., ITOH, T., TAKASAKI, T., SUZUKI, R. & KURANE, I. 2010. Comprehensive analysis and characterization of the TCR alpha chain sequences in the common marmoset. *Immunogenetics,* 62**,** 383-95.

FUJITA, P. A., RHEAD, B., ZWEIG, A. S., HINRICHS, A. S., KAROLCHIK, D., CLINE, M. S., GOLDMAN, M., BARBER, G. P., CLAWSON, H., COELHO, A., DIEKHANS, M., DRESZER, T. R., GIARDINE, B. M., HARTE, R. A., HILLMAN-JACKSON, J., HSU, F., KIRKUP, V., KUHN, R. M., LEARNED, K., LI, C. H., MEYER, L. R., POHL, A., RANEY, B. J., ROSENBLOOM, K. R., SMITH, K. E., HAUSSLER, D. & KENT, W. J. 2011. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res,* 39**,** D876-82.

GAFFEN, S. L. 2009. Structure and signalling in the IL-17 receptor family. *Nat Rev Immunol,* 9**,** 556-67.

GAFFEN, S. L., GOLDSMITH, M. A. & GREENE, W. C. 1998. Interleukin-2 and the interleukin-2 receptor. In: Thomson A (ed) The cytokine handbook 3rd edn. . *Academic, London***,** 73-103.

GAO, G. F. & JAKOBSEN, B. K. 2000. Molecular interactions of coreceptor CD8 and MHC class I: the molecular basis for functional coordination with the T-cell receptor. *Immunol Today,* 21**,** 630-6.

GAYNOR, W. T., COUSINS, D. V. & FRIEND, J. A. 1990. Mycobacterial infections in numbats (Myrmecobius fasciatus). *J Zoo Wildl Med,* 21**,** 476-479.

GEER, L. Y., DOMRACHEV, M., LIPMAN, D. J. & BRYANT, S. H. 2002. CDART: protein homology by domain architecture. *Genome Res,* 12**,** 1619-23.

GEISLER, C., KUHLMANN, J. & RUBIN, B. 1989. Assembly, intracellular processing, and expression at the cell surface of the human alpha beta T cell receptor/CD3 complex. Function of the CD3-zeta chain. *J Immunol,* 143**,** 4069-77.

GERLONI, M., RIZZI, M., CASTIGLIONI, P. & ZANETTI, M. 2004. T cell immunity using transgenic B lymphocytes. *Proc Natl Acad Sci U S A,* 101**,** 3892-7.

GHENDLER, Y., SMOLYAR, A., CHANG, H. C. & REINHERZ, E. L. 1998. One of the CD3epsilon subunits within a T cell receptor complex lies in close proximity to the Cbeta FG loop. *J Exp Med,* 187**,** 1529-36.

GLAICHENHAUS, N., SHASTRI, N., LITTMAN, D. R. & TURNER, J. M. 1991. Requirement for association of p56lck with CD4 in antigen-specific signal transduction in T cells. *Cell,* 64**,** 511-20.

GOBEL, T. W. & BOLLIGER, L. 1998. The chicken TCR zeta-chain restores the function of a mouse T cell hybridoma. *J Immunol,* 160**,** 1552-4.

GOBEL, T. W., CHEN, C. L., LAHTI, J., KUBOTA, T., KUO, C. L., AEBERSOLD, R., HOOD, L. & COOPER, M. D. 1994. Identification of T-cell receptor alpha-chain genes in the chicken. *Proc Natl Acad Sci U S A,* 91**,** 1094-8.

GOBEL, T. W. & FLURI, M. 1997. Identification and analysis of the chicken CD3epsilon gene. *Eur J Immunol,* 27**,** 194-8.

GOEBEL, T. W. F. & DANGY, J. P. 2000. Evidence for a stepwise evolution of the CD3 family. *Immunology,* 164**,** 879 - 883.

GOLD, D. P., CLEVERS, J., ALARCON, B., DUNLAP, S., NOVOTNY, J., WILLIAMS, A. F. & TERHORST, C. 1987. Evolutionary relationship between the T3 chains of the T cell receptor complex and the immunoglobulin supergene family. *Proc Natl. Acad Sci USA,* 84**,** 7649 - 7633.

GOLD, D. P., PUCK, J. M., PETTEY, C. L., CHO, M., COLIGAN, J., WOODY, J. N. & TERHORST, C. 1986. Isolation of cDNA clones encoding the 20K non-glycosylated polypeptide chain of the human T-cell receptor/T3 complex. *Nature,* 321**,** 431-4.

GOLD, M. R. & DEFRANCO, A. I. 1994. Biochemistry of B lymphocyte activation. *Adv Immunol,* 55**,** 221-295.

GOLDSTEIN, D. R. 2011. T Cell Costimulation Blockade and Organ TrnsplantationL A Change of Philosophy for Transplant Immunologists? *J Immunol***,** 2691-2692.

GONG, Q., CHENG, A. M., AKK, A. M., ALBEROLA-ILA, J., GONG, G., PAWSON, T. & CHAN, A. C. 2001. Disruption of T cell signaling networks and development by Grb2 haploid insufficiency. *Nat Immunol,* 2**,** 29-36.

GOODING, L. R. 1992. Virus proteins that counteract host immune defenses. *Cell,* 71**,** 5-7.

GORMAN, S. D., FREWIN, M. R., COBBOLD, S. P. & WALDMANN, H. 1994. Isolation and expression of cDNA encoding the canine CD4 and CD8 alpha antigens. *Tissue Antigens,* 43**,** 184 - 188.

GOUGH, S. C., WALKER, L. S. & SANSOM, D. M. 2005. CTLA4 gene polymorphism and autoimmunity. *Immunol Rev,* 204**,** 102-15.

GRAHAM, J. R., HENDERSHOTT, M. C., TERRAGNI, J. & COOPER, G. M. 2010. mRNA degradation plays a significant role in the program of gene expression regulated by phosphatidylinositol 3-kinase signaling. *Mol Cell Biol,* 30**,** 5295-305.

GRANUCCI, F., ZANONI, I., FEAU, S. & RICCIARDI-CASTAGNOLI, P. 2003. Dendritic cell regulation of immune responses: a new role for interleukin 2 at the intersection of innate and adaptive immunity. *EMBO J,* 22**,** 2546-51.

GRIFFITHS, A. L., RENFREE, M. B., SHAW, G., WATTS, L. M. & HUTSON, J. M. 1993. The tammar wallaby (Macropus eugenii) and the Sprague-Dawley rat: comparative anatomy and physiology of inguinoscrotal testicular descent. *J Anat,* 183 (Pt 3)**,** 441-50.

GROSS, J. A., CALLAS, E. & ALLISON, J. P. 1992. Identification and distribution of the costimulatory receptor CD28 in the mouse. *J Immunol,* 149**,** 380-8.

GROSSBERGER, D., MARCUZ, A., FICHTEL, A., DUDLER, L. & HEIN, W. R. 1993. Sequence analysis of sheep T-cell receptor beta chains. *Immunogenetics,* 37**,** 222-6.

GUJAR, S. A. & MICHALAK, T. I. 2006. Characterization of bioactive recombinant woodchuck interleukin-2 amplified by RLM-RACE and produced in eukaryotic expression system. *Vet. Immunol. Immunopathol.,* 112**,** 183-198.

GULLBERG, J. 1997. Mathematics from the birth of numbers. . New York: W. W. Norton**,** pp. 963–965.

GUNIMALADEVI, I., SAVAN, R. & SAKAI, M. 2006. Identification, cloning and characterization of interleukin-17 and its family from zebrafish. *Fish Shellfish Immunol,* 21**,** 393-403.

GUPTA, R., JUNG, E. & BRUNAK, S. 2004. Prediction of N-glycosylation sites in human proteins. *In preparation*.

GUSTAFSSON, K., GERMANA, S., SUNDT, T. M., 3RD, SACHS, D. H. & LEGUERN, C. 1993. Extensive allelic polymorphism in the CDR2-like region of the miniature swine CD4 molecule. *J Immunol,* 151**,** 1365-70.

GUY, C. S., MULROONEY-COUSINS, P. M., CHURCHILL, N. D. & MICHALAK, T. I. 2008. Intrahepatic expression of genes affiliated with innate and adaptive immune responses immediately after invasion and during acute infection with woodchuck hepadnavirus. *J Virol,* 82**,** 8579-91.

HABICHT, A., NAJAFIAN, N., YAGITA, H., SAYEGH, M. H. & CLARKSON, M. R. 2007. New insights in CD28-independent allograft rejection. *Am J Transplant,* 7**,** 1917-26.

HAGENS, G., GALLEY, Y., GLASER, I., DAVIS, W. C., BALDWIN, C. L., CLEVERS, H. & DOBBELAERE, D. A. 1996. Cloning, sequencing and expression of the bovine CD3 epsilon and TCR-zeta chains, two invariant components of the T-cell receptor complex. *Gene,* 169**,** 165-71.

HAGUE, B. F., SAWASDIKOSOL, S., BROWN, T. J., LEE, K., RECKER, D. P. & KINDT, T. J. 1992. CD4 and its role in infection of rabbit cell lines by human immunodeficiency virus type 1. *Proc Natl Acad Sci USA,* 89**,** 7963 - 7967.

HANGER, J. J., BROMHAM, L. D., MCKEE, J. J., O'BRIEN, T. M. & ROBINSON, W. F. 2000. The nucleotide sequence of koala (Phascolarctos cinereus) retrovirus: a novel type C endogenous virus related to Gibbon ape leukemia virus. *J Virol,* 74**,** 4264-72.

HANKS, S. K. & QUINN, A. M. 1991. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol,* 200**,** 38-62.

HANNA, Z., SIMARD, C., LAPERRIERE, A. & JOLICOEUR, P. 1994. Specific expression of the human CD4 gene in mature CD4+ CD8- and immature CD4+ CD8+ T cells and in macrophages of transgenic mice. *Mol Cell Biol,* 14**,** 1084-94.

HANSEN, J. A., MARTIN, P. J. & NOWINSKI, R. C. 1980. Monoclonal antibodies identifying a novel T-cell antigen and Ia antigens of human T lymphocytes. *Immunogenetics,* 10**,** 247-260.

HANSEN, J. D. & STRASSBURGER, P. 2000. Description of an Ectothermic TCR Coreceptor, CD8α, in Rainbow Trout. *Journal of Immunology,* 164**,** 3132 - 3139.

HANSON, K., ROSKOP, L., DJUROVICH, P. I., ZAHARIEV, F., GORDON, M. S. & THOMPSON, M. E. 2010. A Paradigm for Blue- or Red-Shifted Absorption of Small Molecules Depending on the Site of pi-Extension. *J Am Chem Soc,* 132**,** 16247-55.

HARADA, Y., OHGAI, D., WATANABE, R., OKANO, K., KOIWAI, O., TANABE, K., TOMA, H., ALTMAN, A. & ABE, R. 2003. A Single Amino Acid Alteration in Cytoplasmic Domain Determines IL-2 Promoter Activation by Ligation of CD28 but Not Inducible costimulator (ICOS). *J. Exp. Med.,* 197**,** 257–262.

HARGROVE, J. L. & SCHMIDT, F. H. 1989. The role of mRNA and protein stability in gene expression. *FASEB J,* 3**,** 2360-70.

HARHAY, G. P., SONSTEGARD, T. S., KEELE, J. W., HEATON, M. P., CLAWSON, M. L., SNELLING, W. M., WIEDMANN, R. T., VAN TASSELL, C. P. & SMITH, T. P. 2005. Characterization of 954 bovine full-CDS cDNA sequences. *BMC Genomics,* 6**,** 166.

HARRISON, G. A., BROUGHTON, M. J., YOUNG, L. J., COOPER, D. W. & DEANE, E. M. 1999. Conservation of 3' untranslated region elements in tammar wallaby (*Macropus eugenii*) TNF-α mRNA. *Immunogenetics,* 49**,** 464 - 467.

HARRISON, G. A. & DEANE, E. M. 1999. cDNA sequence of the lymphotoxin beta chain from a marsupial, Macropus eugenii (Tammar wallaby). *J Interferon Cytokine Res,* 19**,** 1099-102.

HARRISON, G. A. & DEANE, E. M. 2000. cDNA cloning of lymphotoxin alpha (LT-alpha) from a marsupial, Macropus eugenii. *DNA Seq,* 10**,** 399-403.

HARRISON, G. A., TAYLOR, C. L., MILLER, R. D. & DEANE, E. M. 2003a. Primary structure and variation of the T-cell receptor delta-chain from a marsupial, Macropus eugenii. *Immunol Lett,* 88**,** 117-25.

HARRISON, G. A., TAYLOR, C. L., MILLER, R. D. & DEANE, E. M. 2003b. Primary structure and variation of the T cell recptor δ-chain from a marsupial, *Macropus eugenii*. . *Immunol Lett,* 88**,** 117-125.

HARRISON, G. A. & WEDLOCK, D. N. 2000. Marsupial cytokines strucutre, function and evolution. *Developmental and Comparative Immunology,* 24**,** 473 -484.

HARRISON, G. A., YOUNG, L. J., WATSON, C. M., MISKA, K. B., MILLER, R. D. & DEANE, E. M. 2003c. A survey of type I interferons from a marsupial and monotreme: implications for the evolution of the type I interferon gene family in mammals. *Cytokine,* 21**,** 105-19.

HAWKEN, R. J., MACCARONE, P., TODER, R., MARSHALL GRAVES, J. A. & MADDOX, J. F. 1999. Isolation and characterization of marsupial IL5 genes. *Immunogenetics,* 49**,** 942 - 948.

HAYDAY, A. C. 2000. γδ cells: a right time and a right place for a conserved third way of protection. *Annu Rev Immuno,* 18**,** 975-1026.

HAYNES, M. R. & WU, G. E. 2007. Gene discovery at the human T-cell receptor alpha/delta locus. *Immunogenetics,* 59**,** 109-121.

HEERY, D. M., KALKHOVEN, E., HOARE, S. & PARKER, M. G. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature,* 387**,** 733-6.

HEHNER, S. P., HOFMANN, T. G., DIENZ, O., DROGE, W. & SCHMITZ, M. L. 2000. Tyrosine-phosphorylated Vav1 as a point of integration for T-cell receptor- and CD28-mediated activation of JNK, p38, and interleukin-2 transcription. *J Biol Chem,* 275**,** 18160-71.

HEIN, W. R., MARCUZ, A., FICHTEL, A., DUDLER, L. & GROSSBERGER, D. 1991. Primary structure of the sheep T-cell receptor alpha chain. *Immunogenetics,* 34**,** 39-41.

HEIN, W. R. & TUNNACLIFFE, A. 1993. Invariant components of the sheep T-cell antigen receptor: cloning of the CD3 epsilon and Tcr zeta chains. *Immunogenetics,* 37**,** 279-84.

HEMSLEY, S. W., CANFIELD, P. J. & HUSBAND, A. J. 1995. Immunohistological staining of lymphoid tissue in four Australian marsupial species using species cross-reactive antibodies. *Immunology and Cell Biology* 73**,** 321-325.

HENIKOFF, S. & HENIKOFF, G. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A,* 89**,** 10915-10919.

HENNECKE, J., CARFI, A. & WILEY, D. C. 2000. Structure of a covalently stabilized complex of a human αβ T cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *EMBO J.,* 19**,** 5611-5624.

HERAUD, J. M., LAVERGNE, A. & KAZANJI, M. 2002. Molecular cloning, characterization, and quantification of squirrel monkey (Saimiri sciureus) Th1 and Th2 cytokines. *Immunogenetics,* 54**,** 20-29

HICKFORD, D., FRANKENBERG, S. & RENFREE, M. B. 2009. The tammar wallaby, Macropus eugenii: a model kangaroo for the study of developmental and reproductive biology. *Cold Spring Harb Protoc,* 2009**,** pdb emo137.

HIGGINS, D., HEMSLEY, S. & CANFIELD, P. J. 2004. Assessment of anti-bovine IL4 and IFN gamma antibodies to label IL4 and IFN gamma in lymphocytes of the koala and brushtail possum. *Veterinary Immunology and Immunopathology* 101.

HOEK, A., RUTTEN, V. P., KOOL, J., ARKESTEIJN, G. J., BOUWSTRA, R. J., VAN RHIJN, I. & KOETS, A. P. 2009. Subpopulations of bovine WC1(+) gammadelta T cells rather than CD4(+)CD25(high) Foxp3(+) T cells act as immune regulatory cells ex vivo. *Vet Res,* 40**,** 6.

HOFSTETTER, H. H., SMOLIANOV, V. & HARTUNG, H. P. 2009. Untired regulatory T cells: untying fatigue from T cell regulation in multiple sclerosis. *Eur Neurol,* 62**,** 327-9.

HOLST, J., WANG, H., EDER, K. D., WORKMAN, C. J., BOYD, K. L., BAQUET, Z., SINGH, H., FORBES, K., CHRUSCINSKI, A., SMEYNE, R., VAN OERS, N. S., UTZ, P. J. & VIGNALI, D. A. 2008. Scalable

signaling mediated by T cell antigen receptor-CD3 ITAMs ensures effective negative selection and prevents autoimmunity. *Nat Immunol,* 9**,** 658-66.

HOLTMEIER, W. & KABELITZ, D. 2005. γδ T cells link innate and adaptive immune responses. Chem Immunol Allergy. *Chem Immunol Allergy,* 86**,** 151-183.

HOPP, T. P. & WOODS, K. R. 1983. A computer program for predicting protein antigenic determinants. *Mol Immunol,* 20**,** 483-9.

HORI, S., NEMURA, T. & SAKAGUCHI, S. 2003. Control of regulatory T cell development by the transcription factor Foxp3. *Science,* 299**,** 1057-1061.

HOROHOV, D. W., BEADLE, R. E., MOUCH, S. & POURCIAU, S. S. 2005. Temporal regulation of cytokine mRNA expression in equine recurrent airway obstruction. *Vet Immunol Immunopathol,* 108**,** 237-45.

HORTON, R., WILMING, L., RAND, V., LOVERING, R. C., BRUFORD, E. A., KHODIYAR, V. K., LUSH, M. J., POVEY, S., TALBOT, C. C., JR., WRIGHT, M. W., WAIN, H. M., TROWSDALE, J., ZIEGLER, A. & BECK, S. 2004. Gene map of the extended human MHC. *Nat Rev Genet,* 5**,** 889-99.

HURLEY, J. H., MASON, D. A. & MATTHEWS, B. W. 1992. Flexible-geometry conformational energy maps for the amino acid residue preceding a proline. *Biopolymers,* 32**,** 1443-6.

HUSE, M., ECK, M. J. & HARRISON, S. C. 1998. A Zn2+ ion links the cytoplasmic tail of CD4 and the N-terminal region of Lck. *J Biol Chem,* 273**,** 18729-33.

HYMOWITZ, S. G., FILVAROFF, E. H., YIN, J. P., LEE, J., CUI, I. & RISSER, P. 2001. IL-17s adopt a cystine knot fold: structure and activity of a novel cytokine, IL17F and implications for receptor binding. *EMBO J,* 20**,** 5332-5341.

IHA, K., OMATSU, T., WATANABE, S., UEDA, N., TANIGUCHI, S., FUJII, H., ISHII, Y., KYUWA, S., AKASHI, H. & YOSHIKAWA, Y. 2009. Molecular cloning and sequencing of the cDNAs encoding the bat interleukin (IL)-2, IL-4, IL-6, IL-10, IL-12p40, and tumor necrosis factor-alpha. . *J. Vet. Med. Sci.,* 71**,** 1691-1695.

IRVING, B. A., CHAN, A. C. & WEISS, A. 1993. Functional characterization of a signal transducing motif present in the T cell antigen receptor zeta chain. *J Exp Med,* 177**,** 1093-103.

ISAACS, N. W. 1995. Cystine knots. *Curr Opin Struct Biol,* 5**,** 391-5.

ISAKOV, N. & BIESINGER, B. 2000. Lck protein tyrosine kinase is a key regulator of T-cell activation and a target for signal intervention by Herpesvirus saimiri and other viral gene products. *Eur J Biochem,* 267**,** 3413-21.

ISHIGURO, N., TANAKA, A. & SHINAGAWA, M. 1990. Sequence analysis of bovine T-cell receptor alpha chain. *Immunogenetics,* 31**,** 57-60.

ISONO, T., KIM, C. J. & SETO, A. 1995. Sequence and diversity of rabbit T-cell receptor gamma chain genes. *Immunogenetics,* 41**,** 295-300.

ISONO, T. & SETO, A. 1995. Cloning and sequencing of the rabbit gene encoding T-cell costimulatory molecules. *Immunogenetics,* 42**,** 217-20.

ITOH, Y., MATSUURA, A., KINEBUCHI, M., HONDA, R., TAKAYAMA, S., ICHIMIYA, S., KON, S. & KIKUCHI, K. 1993. Structural analysis of the CD3 zeta/eta locus of the rat. Expression of zeta but not eta transcripts by rat T cells. *J Immunol,* 151**,** 4705-17.

IWASHIMA, M., IRVING, B. A., VAN OERS, N. S., CHAN, A. C. & WEISS, A. 1994. Sequential interactions of the TCR with two distinct cytoplasmic tyrosine kinases. *Science,* 263**,** 1136-9.

JANEWAY, C. A. J. 2001. How the immune system works to protect the host from infection: A personal view. *Proc. Natl. Acad. Sci. USA - Immunology,* 98**,** 7461 - 7468.

JENKINS, M. K. 1992. The role of cell division in the induction of clonal anergy. *Immunol Today,* 13**,** 69-73.

JENSEN, J. P., HOU, D., RAMSBURG, M., TAYLOR, A., DEAN, M. & WEISSMAN, A. M. 1992. Organization of the human T cell receptor zeta/eta gene and its genetic linkage to the Fc gamma RII-Fc gamma RIII gene cluster. *J Immunol,* 148**,** 2563-71.

JIANG, H. & CHESS, L. 2009. How the immune system achieves self-nonself discrimination during adaptive immunity. *Adv Immunol,* 102**,** 95-133.

JIN, L., PLUSKEY, S., PETRELLA, E. C., CANTIN, S. M., GORGA, J. C., RYNKIEWICZ, M. J., PANDEY, P., STRICKLER, J. E., BABINE, R. E., WEAVER, D. T. & SEIDL, K. J. 2004a. The Three-dimensional Structure of the ZAP-70 Kinase Domain in Complex with Staurosporine. *THE JOURNAL OF BIOLOGICAL CHEMISTRY,* Vol. 279**,** 42818–42825.

JIN, L., PLUSKEY, S., PETRELLA, E. C., CANTIN, S. M., GORGA, J. C., RYNKIEWICZ, M. J., PANDEY, P., STRICKLER, J. E., BABINE, R. E., WEAVER, D. T. & SEIDL, K. J. 2004b. The three-dimensional structure of the ZAP-70 kinase domain in complex with staurosporine: implications for the design of selective inhibitors. *J Biol Chem,* 279**,** 42818-25.

JOHANSEN, M. B., KIEMER, L. & BRUNAK, S. 2006. Analysis and prediction of mammalian protein glycation. *Glycobiology,* 16**,** 844-53.

JOHANSSON, B., PALMER, E. & BOLLIGER, L. 1999. The extracellular domain of the zeta chain is essential for TCR function. *J. Immunol.,* 162**,** 878-885.

JOLY, E., LE ROLLE, A. F., GONZALEZ, A. L., MEHLING, B., STEVENS, J., COADWELL, W. J., HUNIG, T., HOWARD, J. C. & BUTCHER, G. W. 1998. Co-evolution of rat TAP transporters and MHC class I RT1-A molecules. *Curr Biol,* 8**,** 169-72.

JONES, M., CORDELL, J. L., BEYERS, A. D., TSE, A. G. & MASON, D. Y. 1993. Detection of T and B lymphocytes in many animal species using cross-reactive anti-peptide antibodies. *J Immunol,* 150**,** 5429-35.

JOSE, E. S., SAHUQUILLO, A. G., BRAGADO, R. & ALARCON, B. 1998. Assembly of the TCR/CD3 complex: CD3ε/δ and CD3ε/γ dimers associate indistinctly with both TCRα and TCRβ chains. Evidence for a double TCR heterodimer model. *European Journal of Immunology,* 28**,** 12 - 21.

JOSEFOWICZ, S. Z., NIEC, R. E., KIM, H. Y., TREUTING, P., CHINEN, T., ZHENG, Y., UMETSU, D. T. & RUDENSKY, A. Y. 2012. Extrathymically generated regulatory T cells control mucosal TH2 inflammation. *Nature,* 482**,** 395-9.

JULENIUS, K. 2007. NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. *Glycobiology,* 17**,** 868-76.

JUNE, C. H., LEDBETTER, J. A., LINSLEY, P. S. & THOMPSON, C. B. 1990. Role of the CD28 receptor in T cell activation. *Immunol Today,* 11**,** 211-216.

JURD, R. D. 1994. "Not proper mammasl": immunity in monotremes and marsupials. Comp *Immunol Microbiol Infect Dis.* 17, 1, 41-52.

KAISER, P., MARIANI, P. 1999. Promoter sequence, exon:intron structure, and synteny of genetic location show that a chicken cytokine with T-cell proliferative activity is IL2 and not IL15. *Immunogenetics*, 49, 1, 26-35.

KAMPER, S. M. & MCKINNEY, C. E. 2002. Polymorphism and evolution in the constant region of the T-cell receptor beta chain in an advanced teleost fish. *Immunogenetics,* 53**,** 1047-54.

KAPPEL, L. W., GOLDBERG, G. L., KING, C. G., SUH, D. Y., SMITH, O. M., LIGH, C., HOLLAND, A. M., GRUBIN, J., MARK, N. M., LIU, C., IWAKURA, Y., HELLER, G. & VAN DEN BRINK, M. R. 2009. IL-17 contributes to CD4-mediated graft-versus-host disease. *Blood,* 113**,** 945-52.

KASHIWADA, M., PHAM, N. L., PEWE, L. L., HARTY, J. T. & ROTHMAN, P. B. 2011. NFIL3/E4BP4 is a key transcription factor for CD8alpha dendritic cell development. *Blood,* 117**,** 6193-7.

KATOH, S., KITAZAWA, H., SHIMOSATO, T., TOHNO, M., KAWAI, Y. & SAITO, T. 2004. Cloning and characterization of Swine interleukin-17, preferentially expressed in the intestines. *J Interferon Cytokine Res,* 24**,** 553-9.

KAUFMAN, J., MILNE, S., GOBEL, T. W., WALKER, B. A., JACOB, J. P., AUFFRAY, C., ZOOROB, R. & BECK, S. 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature,* 401**,** 923-5.

KELLEY, J., WALTER, L. & TROWSDALE, J. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics,* 56**,** 683-95.

KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4**,** 363-71.

KENT, W. J. 2002. BLAT - The BLAST-Like Alignment Tool. *Genome Res.,* 12**,** 656-664.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KESTI, T., RUPPELT, A., WANG, J. H., LISS, M., WAGNER, R., TASKEN, K. & SAKSELA, K. 2007. Reciprocal regulation of SH3 and SH2 domain binding via tyrosine phosphorylation of a common site in CD3epsilon. *J Immunol,* 179**,** 878-85.

KHATLANI, T. S., OHNO, K., MA, Z., INOKUMA, H. & ONISHI, T. 2000. Cloning and sequencing of dog cDNA encoding the T-cell costimulatory molecule (CTLA-4). *Immunogenetics,* 51**,** 79-81.

KIM, P. W., SUN, Z. Y., BLACKLOW, S. C., WAGNER, G. & ECK, M. J. 2003. A zinc clasp structure tethers Lck to T cell coreceptors CD4 and CD8. *Science,* 301**,** 1725 - 1728.

KIM, S. T., TOUMA, M., TAKEUCHI, K., SUN, Z. Y., DAVE, V. P., KAPPES, D. J., WAGNER, G. & REINHERZ, E. L. 2010. Distinctive CD3 heterodimeric ectodomain topologies maximize antigen-triggered activation of alpha beta T cell receptors. *J Immunol,* 185**,** 2951-9.

KIRKBRIDGE, K. C., RAY, B. N. & BLOBE, G. C. 2005. Cell-surface co-receptors: emerging roles in signaling and human disease. *Trends Biochem. Sci. ,* 30 611–621.

KIRKHAM, P. A., TAKAMATSU, H., YANG, H. & PARKHOUSE, R. M. E. 1996. Porcine CD3ε its characterization, expression and involvement in activation of procine T lymphocytes. *Immunology,* 87**,** 616 - 623.

KLAUSNER, R. D. & SAMELSON, L. E. 1991. T cell antigen receptor activation pathways: the tyrosine kinase connection. *Cell,* 64**,** 875-8.

KLEIN, S. L., STRAUSBERG, R. L., WAGNER, L., PONTIUS, J., CLIFTON, S. W. & RICHARDSON, P. 2002. Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. *Dev. Dyn.,* 225**,** 384-391.

KOLLS, J. K. & LINDEN, A. 2004. Interleukin-17 family members and inflammation. *Immunity,* 21**,** 467-76.

KONIG, R., SHEN, X., MAROTO, R. & DENNING, T. L. 2002. The role of CD4 in regulating homeostasis of T helper cells. *Immunol Res,* 25**,** 115-30.

KONO, T., KORENAGA, H. & SAKAI, M. 2011. Genomics of fish IL-17 ligand and receptors: a review. *Fish Shellfish Immunol,* 31**,** 635-43.

KORN, T., BETTELLI, E., OUKKA, M. & KUCHROO, V. K. 2009. IL-17 and Th17 Cells. *Annu Rev Immunol,* 27**,** 485-517.

KOSKINEN, R., LAMMINMAEKI, U., TREGASKES, C. A., SALOMONSEN, J., YOUNG, J. R. & VAINIO, O. 1999. Cloning and Modeling of the First Nonmammalian CD4. *Immunology,* 162**,** 4115-4121.

KOTHLOW, S., MANNES, N. K., SCHAERER, B., REBESKI, D. E., KASPERS, B. & SCHULTZ, U. 2005. Characterization of duck leucocytes by monoclonal antibodies. *Dev Comp Immunol,* 29**,** 733-48.

KRANGEL, M. S., HERNANDEZ-MUNAIN, C., LAUZURICA, P., MCMURRY, M., ROBERTS, J. L. & ZHONG, X. P. 1998. Developmental regulation of V(D)J recombination at the TCR alpha/delta locus. *Immunol Rev,* 165**,** 131-47.

KREIJVELD, E., KOENEN, H. J., VAN CRANENBROEK, B., VAN RIJSSEN, E., JOOSTEN, I. & HILBRANDS, L. B. 2008. Immunological monitoring of renal transplant recipients to predict acute allograft rejection following the discontinuation of tacrolimus. *PLoS One,* 3**,** e2711.

KRUYS, V. & HUEZ, G. 1994. Translational control of cytokine expression by 3' UA-rich sequences. *Biochimie,* 76**,** 862-6.

KU, G., MALISSEN, B. & MATTEI, M. G. 1994. Chromosomal location of the Syk and ZAP-70 tyrosine kinase genes in mice and humans. *Immunogenetics,* 40**,** 300-2.

KUNJIBETTU, S., FULLER-ESPIE, S., CAREY, G. B. & SPAIN, L. M. 2001. Conserved transmembrane tyrosine residues of the TCR beta chain are required for TCR expression and function in primary T cells and hybridomas. *Int Immunol,* 13**,** 211-22.

KYTE, J. & DOOLITTLE, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol,* 157**,** 105-32.

LA GRUTA, N. L., LIU, H., DILIOGLOU, S., RHODES, M., WIEST, D. L. & VIGNALI, D. A. A. 2004. Architecutral Changes in the TCR:CD3 Complex induced by MHC:Peptide ligation. *Immunology,* 172**,** 3662 - 3669.

LAFAGE-POCHITALOFF, M., COSTELLO, R., COUEZ, D., SIMONETTI, J., MANNONI, P., MAWAS, C. & OLIVE, D. 1990. Human CD28 and CTLA-4 Ig superfamily genes are located on chromosome 2 at bands q33-q34. *Immunogenetics,* 31**,** 198-201.

LAFFERTY, K. J., WARREN, H. S., WOOLNOUGH, J. A. & TALMAGE, D. W. 1978. Immunological induction of T lymphocytes: role of antigen and the lymphocyte costimulator. *Blood Cells,* 4**,** 395-406.

LAING, K. J., DUTTON, S. & HANSEN, J. D. 2007. Molecular and biochemical analysis of rainbow trout LCK suggests a conserved mechanism for T-cell signaling in gnathostomes. *Mol. Immunol. ,* 44**,** 2737-2748.

LALOR, P., BUCCI, C., FORNARO, M., RATTAZZI, M. C., NAKAUCHI, H., HERZENBERG, L. A. & ALBERTI, S. 1992. Molecular cloning, reconstruction and expression of the gene encoding the alpha-chain of the bovine CD8--definition of three peptide regions conserved across species. *Immunology,* 76**,** 95-102.

LANGER, P. 1984. Comparative anatomy of the stomach in mammalian herbivores. *Q J Exp Physiol,* 69**,** 615-25.

LANIER, L. L., YU, G. & PHILLIPS, J. H. 1989. Co-association of CD3 with a receptor (CD16) for IgG Fc on human natural killer cells. . *Nature* 342**,** 803-805.

LANKFORD, S., PETTY, C., LAVOY, A., RECKLING, S., TOMPKINS, W. & DEAN, G. A. 2008. Cloning of feline FOXP3 and detection of expression in CD4+CD25+ regulatory T cells. *Vet Immunol Immunopathol,* 122**,** 159-66.

LAWLOR, D. A. & PARHAM, P. 1992. Structure of CD8 alpha and beta chains of the orangutan: novel patterns of mRNA splicing encoding hingeless polypeptides. *Immunogenetics,* 36**,** 121-5.

LAWSON, S., ROTHWELL, L. & KAISER, P. 2000. Turkey and chicken interleukin-2 cross-react in in vitro proliferation assays despite limited amino acid sequence identity. . *J. Interferon Cytokine Res. ,* 20**,** 161-170.

LAZAR-MOLNAR, E., ALMO, S. C. & NATHENSON, S. G. 2006. The interchain disulfide linkage is not a prerequisite but enhances CD28 costimulatory function. *Cell Immunol,* 244**,** 125-9.

LENSCHOW, D. J., WALUNAS, T. L. & BLUESTONE, J. A. 1996. CD28/B7 system of T cell costimulation. *Annu Rev Immunol,* 14**,** 233-58.

LEONG, J. S., JANTZEN, S. G., VON SCHALBURG, K. R., COOPER, G. A., MESSMER, A. M., LIAO, N. Y., MUNRO, S., MOORE, R., HOLT, R. A., JONES, S. J., DAVIDSON, W. S. & KOOP, B. F. 2010. Salmo salar and Esox lucius full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. *BMC Genomics,* 11**,** 279.

LETUNIC, I., DOERKS, T. & BORK, P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res,* 40**,** D302-5.

LEWINSOHN, D. A., HEINZEL, A. S., GARDNER, J. M., ZHU, L., ALDERSON, M. R. & LEWINSOHN, D. M. 2003. Mycobacterium tuberculosis-specific CD8+ T cells preferentially recognize heavily infected cells. *Am J Respir Crit Care Med,* 168**,** 1346-52.

LEWKOWICZ, P., LEWKOWICZ, N., SASIAK, A. & TCHORZEWSKI, H. 2006. Lipopolysaccharide-activated CD4+CD25+ T regulatory cells inhibit neutrophil function and promote their apoptosis and death. *J Immunol,* 177**,** 7155-63.

LI, B., SAMANTA, A., SONG, X., IACONO, K. T., BRENNAN, P., CHATILA, T. A., RONCADOR, G., BANHAM, A. H., RILEY, J. L., WANG, Q., SHEN, Y., SAOUAF, S. J. & GREENE, M. I. 2007. FOXP3 is a homo-oligomer and a component of a supramolecular regulatory complex disabled in the human XLAAD/IPEX autoimmune disease. *Int Immunol,* 19**,** 825-35.

LI, Q. R., MA, J., WANG, H. & LI, J. S. 2005. Interleukin-2α Receptor in Membrane Lipid Rafts. *Transplantation Proceedings,* 37**,** 2395–2397.

LI, R., FAN, W., TIAN, G., ZHU, H., HE, L., CAI, J., HUANG, Q., CAI, Q., LI, B., BAI, Y., ZHANG, Z., ZHANG, Y., WANG, W., LI, J., WEI, F., LI, H., JIAN, M., NIELSEN, R., LI, D., GU, W., YANG, Z., XUAN, Z., RYDER, O. A., LEUNG, F. C., ZHOU, Y., CAO, J., SUN, X., FU, Y., FANG, X., GUO, X., WANG, B., HOU, R., SHEN, F., MU, B., NI, P., LIN, R., QIAN, W., WANG, G., YU, C., NIE, W., WANG, J., WU, Z., LIANG, H., MIN, J., WU, Q., CHENG, S., RUAN, J., WANG, M., SHI, Z., WEN, M., LIU, B., REN, X., ZHENG, H., DONG, D., COOK, K., SHAN, G., ZHANG, H., KOSIOL, C., XIE, X., LU, Z., LI, Y., STEINER, C. C., LAM, T. T., LIN, S., ZHANG, Q., LI, G., TIAN, J., GONG, T., LIU, H., ZHANG, D., FANG, L., YE, C., ZHANG, J., HU, W., XU, A., REN, Y., ZHANG, G., BRUFORD, M. W., LI, Q., MA, L., GUO, Y., AN, N., HU, Y., ZHENG, Y., SHI, Y., LI, Z., LIU, Q., CHEN, Y., ZHAO, J., QU, N., ZHAO, S., TIAN, F., WANG, X., WANG, H., XU, L., LIU, X., VINAR, T., WANG, Y., LAM, T. W., YIU, S. M., et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature,* 463**,** 311-7.

LI, S., CHOKSI, S., SHAN, S., HU, X., GAO, J., KORNGOLD, R. & HUANG, Z. 1998. Identification of the CD8 DE loop as a surface functional epitope. Implications for major histocompatibility complex class I binding and CD8 inhibitor design. *J Biol Chem,* 273**,** 16442-5.

LIANG, X., LU, Y., WILKES, M., NEUBERT, T. A. & RESH, M. D. 2004. The N-terminal SH4 region of the Src family kinase Fyn is modified by methylation and heterogeneous fatty acylation: role in membrane targeting, cell adhesion, and spreading. *J Biol Chem,* 279**,** 8133-9.

LIN, A. F., XIANG, L. X., WANG, Q. L., DONG, W. R., GONG, Y. F. & SHAO, J. Z. 2009. The DC-SIGN of zebrafish: insights into the existence of a CD209 homologue in a lower vertebrate and its involvement in adaptive immunity. *J. Immunol.,* 183**,** 7398-7410.

LIN, H., RATHMELL, J. C., GRAY, G. S., THOMPSON, C. B., LEIDEN, J. M. & ALEGRE, M. L. 1998. Cytotoxic T lymphocyte antigen 4 (CTLA4) blockade accelerates the acute rejection of cardiac allografts in CD28-deficient mice: CTLA4 can function independently of CD28. *J Exp Med,* 188**,** 199-204.

LINSLEY, P. S., NADLER, S. G., BAJORATH, J., PEACH, R., LEUNG, H. T., ROGERS, J., BRADSHAW, J., STEBBINS, M., LEYTZE, G., BRADY, W. & ET AL. 1995. Binding stoichiometry of the cytotoxic T lymphocyte-associated molecule-4 (CTLA-4). A disulfide-linked homodimer binds two CD86 molecules. *J Biol Chem,* 270**,** 15417-24.

LIS, H. & SHARON, N. 1993. Protein glycosylation. Structural and functional aspects. *Eur J Biochem,* 218**,** 1-27.

LITTMAN, D. R., THOMAS, Y., MADDON, P. J., CHESS, L. & AXEL, R. 1985. The isolation and sequence of the gene encoding T8: a molecule defining functional classes of T lymphocytes. *Cell,* 40**,** 237-46.

LIU, Y., MOORE, L., KOPPANG, E. O. & HORDVIK, I. 2008. Characterization of the CD3zeta, CD3gammadelta and CD3epsilon subunits of the T cell receptor complex in Atlantic salmon. *Dev Comp Immunol,* 32**,** 26-35.

LIVOLSI, A., BUSUTTIL, V., IMBERT, V., ABRAHAM, R. T. & PEYRON, J. F. 2001. Tyrosine phosphorylation-dependent activation of NF-kappa B. Requirement for p56 LCK and ZAP-70 protein tyrosine kinases. *Eur J Biochem,* 268**,** 1508-15.

LOCKHART, E., GREEN, A. M. & FLYNN, J. L. 2006. IL-17 production is dominated by gammadelta T cells rather than CD4 T cells during Mycobacterium tuberculosis infection. *J Immunol,* 177**,** 4662-9.

LONBERG, N., GETTNER, S. N., LACY, E. & LITTMAN, D. R. 1988. Mouse brain CD4 transcripts encode only the COOH-terminal half of the protein. *Mol Cell Biol,* 8**,** 2224-8.

LOPES, J. E., TORGERSON, T. R., SCHUBERT, L. A., ANOVER, S. D., OCHELTREE, E. L., OCHS, H. D. & ZIEGLER, S. F. 2006. Analysis of FOXP3 reveals multiple domains required for its function as a transcriptional repressor. *J Immunol,* 177**,** 3133-42.

MACIAS, M. J., WIESNER, S. & SUDOL, M. 2002. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett,* 513**,** 30-7.

MADDON, P. J., LITTMAN, D. R., GODFREY, M., MADDON, D. E., CHESS, L. & AXEL, R. 1985. The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: A new member of the immunoglobulin gene family. *Cell,* 42**,** 93-104.

MAHER, S. E., KARMANN, K., MIN, W., HUGHES, C. C., POBER, J. S. & BOTHWELL, A. L. 1996. Porcine endothelial CD86 is a major costimulator of xenogeneic human T cells: cloning, sequencing, and functional expression in human endothelial cells. *J Immunol,* 157**,** 3838-44.

MAI, Z., KOUSOULAS, K. G., HOROHOV, D. W. & KLEI, T. R. 1994. Cross-species PCR cloning of gerbil (Meriones unguiculatus) interleukin-2 cDNA and its expression in COS-7 cells. *Vet. Immunol. Immunopathol. ,* 40**,** 63-71.

MALLAUN, M., NAEHER, D., DANIELS, M. A., YACHI, P. P., HAUSMANN, B., LUESCHER, I. F., GASCOIGNE, N. R. & PALMER, E. 2008. The T cell receptor's alpha-chain connecting peptide motif promotes close approximation of the CD8 coreceptor allowing efficient signal initiation. *J Immunol,* 180**,** 8211-21.

MANEL, N., UNUTMAZ, D. & LITTMAN, D. R. 2008. The differentiation of human T(H)-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. *Nat Immunol,* 9**,** 641-9.

MANNERING, S. I. & CHEERS, C. 2002. Interleukin-2 and loss of immunity in experimental Mycobacterium avium infection. *Infect Immun,* 70**,** 27-35.

MANOLIOS, N. 1995. Hierarchy of T cell antigen receptor assembly. *Immunol Cell Biol,* 73**,** 544-8.

MANOLIOS, N., COLLIER, S., TAYLOR, J., POLLARD, J., HARRISON, L. C. & BENDER, V. 1997. T-cell antigen receptor transmembrane peptides modulate T-cell function and T cell-mediated disease. *Nat Med,* 3**,** 84-8.

MANOLIOS, N., LETOURNEUR, F., BONIFACINO, J. S. & KLAUSNER, R. D. 1991. Pairwise, cooperative and inhibitory interactions describe the assembly and probable structure of the T-cell antigen receptor. *EMBO J,* 10**,** 1643-51.

MANTEL, P. Y., OUAKED, N., RUCKERT, B., KARAGIANNIDIS, C., WELZ, R., BLASER, K. & SCHMIDT-WEBER, C. B. 2006. Molecular mechanisms underlying FOXP3 induction in human T cells. *J Immunol,* 176**,** 3593-602.

MARCHALONIS, J. J., BERNSTEIN, R. M., SHEN, S. X. & SCHLUTER, S. F. 1996. Emergence of the immunoglobulin family: conservation in protein sequence and plasticity in gene organization. *Glycobiology,* 6**,** 657-63.

MARCHE, P. N. & KINDT, T. J. 1986. Two distinct T-cell receptor alpha-chain transcripts in a rabbit T-cell line: implications for allelic exclusion in T cells. *Proc Natl Acad Sci U S A,* 83**,** 2190-4.

MARENGERE, L. E., OKKENHAUG, K., CLAVREUL, A., COUEZ, D., GIBSON, S., MILLS, G. B., MAK, T. W. & ROTTAPEL, R. 1997. The SH3 domain of Itk/Emt binds to proline-rich sequences in the cytoplasmic domain of the T cell costimulatory receptor CD28. *J Immunol,* 159**,** 3220-9.

MARGETA-MITROVIC, M. 2002. Assembly-dependent trafficking assays in the detection of receptor–receptor interactions. *Methods* 27**,** 311–317.

MARGULIES, D. H. 2003. CD28, costimulator or agonist receptor? *J Exp Med,* 197**,** 949-53.

MARRACK, P., ENDRES, R., SHIMONKEVITZ, R., ZLOTNIK, A., DIALYNAS, D., FITCH, F. & KAPPLER, J. 1983. The major histocompatibility complex-restricted antigen receptor on T cells. II. Role of the L3T4 product. *J Exp Med,* 158**,** 1077-91.

MARSHALL GRAVES, J. A. & WESTERMAN, M. 2002. Marsupial genetics and genomics. *Trends in Genetics,* 18**,** 517 - 522.

MARTIN-OROZCO, N., CHUNG, Y., CHANG, S. H., WANG, Y. H. & DONG, C. 2009. Th17 cells promote pancreatic inflammation but only induce diabetes efficiently in lymphopenic hosts after conversion into Th1 cells. *Eur J Immunol,* 39**,** 216-24.

MATESANZ, F., ALCINA, A. & PELLICER, A. 1992. A new cDNA sequence for the murine interleukin-2 gene. *Biochim. Biophys. Acta,* 1132**,** 335-336.

MATESANZ, F., ALCINA, A. & PELLICER, A. 1993. Existence of at least five interleukin-2 molecules in different mouse strains. *Immunogenetics,* 38**,** 300-303

MATSUZAKI, G. & UMEMURA, M. 2007. Interleukin-17 as an effector molecule of innate and acquired immunity against infections. *Microbiol Immunol,* 51**,** 1139-47.

MATZINGER, P. 2002. The danger model: a renewed sense of self. *Science,* 296**,** 301-5.

MCCOY, K. D. & LE GROS, G. 1999. The role of CTLA-4 in the regulation of T cell immune responses. *Immunol Cell Biol,* 77**,** 1-10.

MCDONALD, N. Q. & HENDRICKSON, W. A. 1993. A structural superfamily of growth factors containing a cystine knot motif. *Cell,* 73**,** 421-4.

MCGUFFIN, L. J., BRYSON, K. & JONES, D. T. 2000. The PSIPRED protein structure prediction server. *Bioinformatics,* 16**,** 404-5.

MCKENNA, K., BEIGNON, A. S. & BHARDWAJ, N. 2005. Plasmacytoid dendritic cells: linking innate and adaptive immunity. *J Virol,* 79**,** 17-27.

MCKNIGHT, A. J., MASON, D. W. & BARCLAY, A. N. 1989. Sequence of rat interleukin 2 and anomalous binding of a mouse interleukin 2 cDNA probe to rat MHC class II-associated invariant chain mRNA. *Immunogenetics* 30**,** 145-147.

MEDZHITOV, R. & JANEWAY, C. A., JR. 1997. Innate immunity: impact on the adaptive immune response. *Curr Opin Immunol,* 9**,** 4-9.

MELTRETTER, J. & PISCHETSRIEDER, M. 2008. Application of mass spectrometry for the detection of glycation and oxidation products in milk proteins. *Ann N Y Acad Sci,* 1126**,** 134-40.

METHI, T., NGAI, J., VANG, T., TORGERSEN, K. M. & TASKEN, K. 2007. Hypophosphorylated TCR/CD3ζ signals through a Grb2-SOS1-Ras pathway in Lck knockdown cells. *Eur. J. Immunol,* 37**,** 2539–2548.

MEUER, S. C., HUSSEY, R. E., CANTRELL, D. A., HODGDON, J. C., SCHLOSSMAN, S. F., SMITH, K. A. & REINHERZ, E. L. 1984. Triggering of the T3-Ti antigen-receptor complex results in clonal T-cell proliferation through an interleukin 2-dependent autocrine pathway. *Proc. Natl. Acad. Sci. USA Immunology,* 81**,** 1509-1513.

MEYER-LUCHT, Y., OTTEN, C., PUTTKER, T. & SOMMER, S. 2008. Selection, diversity and evolutionary patterns of the MHC class II DAB in free-ranging Neotropical marsupials. *BMC Genet,* 9**,** 39.

MICHEL, F., ATTAL-BONNEFOY, G., MANGINO, G., MISE-OMATA, S. & ACUTO, O. 2001. CD28 as a molecular amplifier extending TCR ligation and signaling capabilities. *Immunity,* 15**,** 935-45.

MIKKELSEN, T. S., WAKEFIELD, M. J., AKEN, B., AMEMIYA, C. T., CHANG, J. L., DUKE, S., GARBER, M., GENTLES, A. J., GOODSTADT, L., HEGER, A., JURKA, J., KAMAL, M., MAUCELI, E., SEARLE, S. M., SHARPE, T., BAKER, M. L., BATZER, M. A., BENOS, P. V., BELOV, K., CLAMP, M., COOK, A., CUFF, J., DAS, R., DAVIDOW, L., DEAKIN, J. E., FAZZARI, M. J., GLASS, J. L., GRABHERR, M., GREALLY, J. M., GU, W., HORE, T. A., HUTTLEY, G. A., KLEBER, M., JIRTLE, R. L., KOINA, E., LEE, J. T., MAHONY, S., MARRA, M. A., MILLER, R. D., NICHOLLS, R. D., ODA, M., PAPENFUSS, A. T., PARRA, Z. E., POLLOCK, D. D., RAY, D. A., SCHEIN, J. E., SPEED, T. P., THOMPSON, K., VANDEBERG, J. L., WADE, C. M., WALKER, J. A., WATERS, P. D., WEBBER, C., WEIDMAN, J. R., XIE, X., ZODY, M. C., GRAVES, J. A., PONTING, C. P., BREEN, M., SAMOLLOW, P. B., LANDER, E. S. & LINDBLAD-TOH, K. 2007. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature,* 447**,** 167-77.

MILDE, K. F., CONNER, G. W., MINTZ, D. H. & ALEJANDRO, R. 1993. Primary structue of the canine CD4 antigen. *Biochimica et Biophysica Acta,* 1172**,** 315 -318.

MILLER, R. D. & ROSENBERG, G. H. 1997. Recombination activating gene-1 of the opossum *Monodelphis domestica*. *Immunogenetics,* 45**,** 341 - 342.

MINGALA, C. N., KONNAI, S., IKEBUCHI, R. & OHASHI, K. 2011. Characterization of CTLA-4, PD-1 and PDL-1 of swamp and riverine type water buffaloes. *Comp Immunol Microbiol Infect Dis,* 34**,** 55-63.

MINGALA, C. N., ODBILEG, R., KONNAI, S., OHASHI, K. & ONUMA, M. 2006. Comparative assessment of Th1 and Th2 cytokines of swamp type buffalo and other bubaline breeds by molecular cloning, sequencing and phylogenetics. *Vet. Immunol. Immunopathol.,* 113**,** 348-356.

MINGUENEAU, M., SANSONI, A., GREGOIRE, C., RONCAGALLI, R., AGUADO, E., WEISS, A., MALISSEN, M. & MALISSEN, B. 2008. The proline-rich sequence of CD3ε controls T cell antigen receptor expression on and singaling potency in preselection $CD4^{+}CD8^{+}$ thymocytes. *Nat Immunol,* 9**,** 522-532.

MINGUET, S., SWAMY, M., DOPFER, E. P., DENGLER, E., ALARCON, B. & SCHAMEL, W. W. 2008. The extracellular part of zeta is buried in the T cell antigen receptor complex. *Immunol Lett,* 116**,** 203-10.

MITTRUCKER, H. W. & KAUFMANN, S. H. 2004. Mini-review: regulatory T cells and infection: suppression revisited. *Eur J Immunol,* 34**,** 306-12.

MIZUNO, T., SUZUKI, R., UMEKI, S. & OKUDA, M. 2009. Crossreactivity of antibodies to canine CD25 and Foxp3 and identification of canine CD4+CD25 +Foxp3+ cells in canine peripheral blood. *J Vet Med Sci,* 71**,** 1561-8.

MLADINIC, M., LEFEVRE, C., DEL BEL, E., NICHOLLS, J. & DIGBY, M. 2010. Developmental changes of gene expression after spinal cord injury in neonatal opossums. *Brain Res,* 1363**,** 20-39.

MOLINA, I. J., KENNEY, D. M., ROSEN, F. S. & REMOLD-O'DONNELL, E. 1992. T cell lines characterize events in the pathogenesis of the Wiskott-Aldrich syndrome. *J Exp Med,* 176**,** 867-74.

MOLLER, S., LESER, U., FLEISCHMANN, W. & APWEILER, R. 1999. EDIT to TrEMBL: a distributed approach to high-quality automated protein sequence annotation. *Bioinformatics,* 15**,** 219-27.

MONTALI, R. J., BUSH, M., CROMIE, R., HOLLAND, S. M., MASLOW, J. N., WORLEY, M., WITEBSKY, F. G. & PHILLIPS, T. M. 1998. Primary *Mycobacterium avium* Complex infections correlate with lowered cellular immune reactivity in Matschie's Tree Kangaroos (*Dendrolagus matschiei*). *Infectious Diseases,* 178.

MOON, C., KIM, S. H., PARK, K. S., CHOI, B. K., LEE, H. S., PARK, J. B., CHOI, G. S., KWAN, J. H., JOH, J. W. & KIM, S. J. 2009. Use of epigenetic modification to induce FOXP3 expression in naive T cells. *Transplant Proc,* 41**,** 1848-54.

MOORE, L. J., SOMAMOTO, T., LIE, K. K., DIJKSTRA, J. M. & HORDVIK, I. 2005. Characterisation of salmon and trout CD8α and CD8β. *Molecular Immunology,* 42**,** 1225-1234.

MOORE, S. C., ANDERSON, S. J. & WALKER, W. S. 1992. Mouse macrophages contain a truncated CD4 transcript. *Leukocyte Biology,* 52**,** 128 - 129.

MORGAN, D. A., RUSCETTI, F. W. & GALLO, R. 1976. Selective in vitro growth of T lymphocytes from normal human bone marrows. *Science,* 193**,** 1007-8.

MORITA, C. T., MARIUZZA, R. A. & BRENNER, M. B. 2000. Antigen recognition by human γδ T cells: pattern recognition by the adaptive immune system. *springer Semin Immunopathol,* 22**,** 191-217.

MORRISSEY, D. V. & COLLINS, M. L. 1989. Nucleic acid hybridization assays employing dA-tailed capture probes. Single capture methods. *Mol Cell Probes,* 3**,** 189-207.

MURANSKI, P., BONI, A., ANTONY, P. A., CASSARD, L., IRVINE, K. R., KAISER, A., PAULOS, C. M., PALMER, D. C., TOULOUKIAN, C. E., PTAK, K., GATTINONI, L., WRZESINSKI, C., HINRICHS, C. S., KERSTANN, K. W., FEIGENBAUM, L., CHAN, C. C. & RESTIFO, N. P. 2008. Tumor-specific Th17-polarized cells eradicate large established melanoma. *Blood,* 112**,** 362-73.

MURCHISON, E. P., SCHULZ-TRIEGLAFF, O. B., NING, Z., ALEXANDROV, L. B., BAUER, M. J., FU, B., HIMS, M., DING, Z., IVAKHNO, S., STEWART, C., NG, B. L., WONG, W., AKEN, B., WHITE, S., ALSOP, A., BECQ, J., BIGNELL, G. R., CHEETHAM, R. K., CHENG, W., CONNOR, T. R., COX, A. J., FENG, Z. P., GU, Y., GROCOCK, R. J., HARRIS, S. R., KHREBTUKOVA, I., KINGSBURY, Z., KOWARSKY, M., KREISS, A., LUO, S., MARSHALL, J., MCBRIDE, D. J., MURRAY, L., PEARSE, A. M., RAINE, K., RASOLONJATOVO, I., SHAW, R., TEDDER, P., TREGIDGO, C., VILELLA, A. J., WEDGE, D. C., WOODS, G. M., GORMLEY, N., HUMPHRAY, S., SCHROTH, G., SMITH, G., HALL, K., SEARLE, S. M., CARTER, N. P., PAPENFUSS, A. T., FUTREAL, P. A., CAMPBELL, P. J., YANG, F., BENTLEY, D. R., EVERS, D. J. & STRATTON, M. R. 2012. Genome sequencing and analysis of the tasmanian devil and its transmissible cancer. *Cell,* 148**,** 780-91.

MURZIN, A. G., BRENNER, S. E., HUBBARD, T. & CHOTHIA, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol,* 247**,** 536-40.

NAGARAJAN, U. M., O'CONNELL, C. & RANK, R. G. 2004. Molecular characterization of guinea - pig (Cavia porcellus) CD8alpha and CD8beta cDNA. *Tissue Antigens,* 63**,** 184-9.

NAKAJIMA, A., WATANABE, N., YOSHINO, S., YAGITA, H., OKUMURA, K. & AZUMA, M. 1997. Requirement of CD28-CD86 co-stimulation in the interaction between antigen-primed T helper type 2 and B lymphocytes. *Int Immunol,* 9**,** 637-44.

NAKAUCHI, H., SHINKAI, Y. & OKUMURA, K. 1987. Molecular cloning of Lyt-3, a membrane glycoprotein marking a subset of mouse T lymphocytes: molecular homology to

immunoglobulin and T-cell receptor variable and joining regions. *Proc Natl Acad Sci U S A,* 84**,** 4210-4.

NASH, R. A., SCHERF, U. & STORB, R. 1991. Molecular cloning of the CD3 epsilon subunit of the T-cell receptor/CD3 complex in dog. *Immunogenetics,* 33**,** 396-8.

NEL, A. E. & SLAUGHTER, N. 2002. T-cell activation through the antigen receptor. Part 2: role of signaling cascades in T-cell differentiation, anergy, immune senescence, and development of immunotherapy. *J Allergy Clin Immunol,* 109**,** 901-15.

NELSON, B. H., LORD, J. D. & GREENBERG, P. D. 1996. A membrane-proximal region of the interleukin-2 receptor gamma c chain sufficient for Jak kinase activation and induction of proliferation in T cells. *Mol Cell Biol,* 16**,** 309-17.

NESS, T. L., BRADLEY, W. G., ROESS, W. B. & REYNOLDS, J. E. I. 1997. Isolation and Expression of the Interleukin-2 Gene from the Killer Whale, Orcinus orca. . *Mar. Mamm. Sci.*

NEUMEISTER, E. N., ZHU, Y., RICHARD, S., TERHORST, C., CHAN, A. C. & SHAW, A. S. 1995. Binding of ZAP-70 to phosphorylated T-cell receptor zeta and eta enhances its autophosphorylation and generates specific binding sites for SH2 domain-containing proteins. *Mol Cell Biol,* 15**,** 3171-8.

NGUYEN, J. T., TURCK, C. W., COHEN, F. E., ZUCKERMANN, R. N. & LIM, W. A. 1998. Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science,* 282**,** 2088-92.

NISHIMURA, Y., MIYAZAWA, T., IKEDA, Y., IZUMIYA, Y., NAKAMURA, K., CAI, J. S., SATO, E., KHOMOTO, M. & MIKAMI, T. 1998. Molecular cloning and expression of feline CD3ε.

NOLAN, A., KOBAYASHI, H., NAVEED, B., KELLY, A., HOSHINO, Y., HOSHINO, S., KARULF, M. R., ROM, W. N., WEIDEN, M. D. & GOLD, J. A. 2009. Differential role for CD80 and CD86 in the regulation of the innate immune response in murine polymicrobial sepsis. *PLoS One,* 4**,** e6600.

NONANAKA, S., SOMAMAOTO, T., KATO-UNOKI, Y., OTOTTAKE, M., NAKANISHI, T. & NAKAO, M. 2008. Molecular cloning of CD4 from ginbuna crucian carp *Carassius auratus langsdorfii*. *Fisheries Science,* 72**,** 341-346.

NORIMINE, J., MIYAZAWA, T., KAWAGUCHI, Y., TOHYA, Y., KAI, C. & MIKAMI, T. 1992. A cDNA encoding feline CD4 has a unique repeat sequence downstream of the V-like region. *Immunology,* 75**,** 74-9.

NORMENT, A. M. & LITTMAN, D. R. 1988. A second subunit of CD8 is expressed in human T cells. *EMBO J,* 7**,** 3433-9.

OAKS, M. K. & HALLETT, K. M. 2000. Cutting Edge: A Soluble Form of CTLA-4 in Patients with Autoimmune Thyroid disease. *J. Immunol.,* 164**,** 50115-5018.

ODBILEG, R., LEE, S. I., YOSHIDA, R., CHANG, K. S., OHASHI, K., SUGIMOTO, C. & ONUMA, M. 2004. Cloning and sequence analysis of llama cytokines related to cell-mediated immunity. . *Vet. Immunol. Immunopathol. ,* 99**,** 1-10.

ODBILEG, R., PUREVTSEREN, B., BATSUKH, Z., KONNAI, S., OHASHI, K. & ONUMA, M. 2006. Complete cDNA sequences and phylogenetic analyses of the Th1 and Th2 cytokines of the Bactrian Camel (Camelus bactrianus). . *J. Vet. Med. Sci. ,* 68**,** 941-946.

OFFMAN, M. N., FITZJOHN, P. W. & BATES, P. A. 2006. Developing a move-set for protein model refinement. *Bioinformatics,* 22**,** 1838-45.

OHME-TAKAGI, M., TAYLOR, C. B., NEWMAN, T. C. & GREEN, P. J. 1993. The effect of sequences with high AU content on mRNA stability in tobacco. *Proc Natl Acad Sci U S A,* 90**,** 11811-5.

OHTA, Y., GOETZ, W., HOSSAIN, M. Z., NONAKA, M. & FLAJNIK, M. F. 2006. Ancestral organization of the MHC revealed in the amphibian Xenopus. *J Immunol,* 176**,** 3674-85.

OKAMOTO, R., SOUMA, S. & KAJIHARA, Y. 2009. Efficient substitution reaction from cysteine to the serine residue of glycosylated polypeptide: repetitive peptide segment ligation strategy and the synthesis of glycosylated tetracontapeptide having acid labile sialyl-T(N) antigens. *J Org Chem,* 74**,** 2494-501.

OKKENHAUG, K., BILANCIO, A., EMERY, J. L. & VANHAESEBROECK, B. 2004. Phosphoinositide 3-kinase in T cell activation and survival. *Biochem Soc Trans,* 32**,** 332-5.

OKOYE, F. I., KRISHNAN, S., CHANDOK, M. R., TSOKOS, G. C. & FARBER, D. L. 2007. Proximal signaling control of human effector CD4 T cell function. *Clin Immunol,* 125**,** 5-15.

OLD, J. M., CARMAN, R. L., FRY, G. & DEANE, E. M. 2006. The immune tissues of the endangered red-tailed phascogale (*Phascogale calura)*. *Anatomy,* 208**,** 381 - 387.

OLD, J. M. & DEANE, E. M. 2000. Development of the immune system and immunological protection in marsupial pouch young. *Dev Comp Immunol,* 24**,** 445-54.

OLD, J. M. & DEANE, E. M. 2001. Histology and immunohistochemistry of the gut-associated lymphoid tissue of the eastern grey kangaroo, Macropus giganteus. *J Anat,* 199**,** 657-62.

OLD, J. M. & DEANE, E. M. 2002a. The gut-associated lymphoid tissues of the northern brown bandicoot (Isoodon macrourus). *Dev Comp Immunol,* 26**,** 841-8.

OLD, J. M. & DEANE, E. M. 2002b. Immunohistochemistry of the lymphoid tissues of the tammar wallaby, *Macropus eugenii*. *Anatomy,* 201**,** 257 - 266.

OLD, J. M. & DEANE, E. M. 2003. The detection of mature T-and B-cells during development of the lymphoid tissues of the tammar wallaby (Macropus eugenii). *Anatomy,* 203**,** 123 - 131.

OLD, J. M. & DEANE, E. M. 2005. Antibodies to the Ross River virus in captive marsupials in urban areas of eastern New South Wales, Australia. *J Wildl Dis,* 41**,** 611-4.

OLD, J. M., DEANE, E. M. & HARRISON, G. A. 2001. Molecular characterisation of the tammar wallaby CD3 epsilon chain cDNA. *Molecular Immunology,* 38**,** 359-364.

OLD, J. M., SELWOOD, L. & DEANE, E. M. 2003. Development of lymphoid tissues of the stripe-faced dunnart (Sminthopsis macroura). *Cells Tissues Organs,* 175**,** 192-201.

OMATSU, T., NISHIMURA, Y., BAK, E. J., ISHII, Y., TOHYA, Y., KYUWA, S., AKASHI, H. & YOSHIKAWA, Y. 2006. Molecular cloning and sequencing of the cDNA encoding the bat CD4. *Veterinary Immunology and Immunopathology***,** 8.

OSONO, E., SATO, N., YOKOMURO, K. & SAIZAWA, K. 1997. Changes in arrangement and in conformation of molecular components of peripheral T cell antigen receptor complex after ligand binding: analyses by co-precipitation profiles. *Scand J Immunol,* 45**,** 487-493.

OSTROV, D. A., SHI, W., SCHWARTZ, J. C., ALMO, S. C. & NATHENSON, S. G. 2000. Structure of murine CTLA-4 and its role in modulating T cell responsiveness. *Science,* 290**,** 816-9.

OUYANG, W., KOLLS, J. K. & ZHENG, Y. 2008. The biological functions of T helper 17 cell effector cytokines in inflammation. *Immunity,* 28**,** 454-67.

OVERGARD, A. C., HORDVIK, I., NERLAND, A. H., EIKELAND, G. & PATEL, S. 2009. Cloning and expression analysis of Atlantic halibut (Hippoglossus hippoglossus) CD3 genes. *Fish Shellfish Immunol,* 27**,** 707-13.

OVERGARD, A. C., NERLAND, A. H. & PATEL, S. 2010a. Cloning, characterization, and expression pattern of Atlantic halibut (Hippoglossus hippoglossus L.) CD4-2, Lck, and ZAP-70. *Fish Shellfish Immunol. ,* 29**,** 987-997.

OVERGARD, A. C., NERLAND, A. H. & PATEL, S. 2010b. Cloning, characterization, and expression pattern of Atlantic halibut (Hippoglossus hippoglossus L.) CD4-2, Lck, and ZAP-70. . *Fish Shellfish Immunol.,* 29**,** 987-997.

OWEN, D. J. & EVANS, P. R. 1998. A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science,* 282**,** 1327-32.

PAGES, F., RAGUENEAU, M., ROTTAPEL, R., TRUNEH, A., NUNES, J., IMBERT, J. & OLIVE, D. 1994. Binding of phosphatidylinositol-3-OH kinase to CD28 is required for T-cell signalling. *Nature,* 369**,** 327-9.

PAN, Z., BARRY, R., LIPKIN, A. & SOLOVIEV, M. 2007. Selection strategy and the design of hybrid oligonucleotide primers for RACE-PCR: cloning a family of toxin-like sequences from Agelena orientalis. *BMC Mol Biol,* 8**,** 32.

PANACCIO, M., GILLESPIE, M. T., WALKER, I. D., KIRSZBAUM, L., SHARPE, J. A., TOBIAS, G. H., MCKENZIE, I. F. & DEACON, N. J. 1987. Molecular characterization of the murine cytotoxic T-cell membrane glycoprotein Ly-3 (CD8). *Proc Natl Acad Sci U S A,* 84**,** 6874-8.

PARNES, J. R., SIZER, K. C., SUKHATME, V. P. & HUNKAPILLER, T. 1985. Structure of Leu-2/T8 as deduced from the sequence of a cDNA clone. *Behring Inst. Mitt.,* 77**,** 48-55.

PARRA, Z. E., BAKER, M. L., LOPEZ, A. M., TRUJILLO, J., VOLPE, J. M. & MILLER, R. D. 2009. TCRmu recombination and transcription relative to the conventional TCR during postnatal development in opossums. *J Immunol,* 182**,** 154-63.

PARRA, Z. E., BAKER, M. L., SCHWARZ, R. S., DEAKIN, J. E., LINDBALD-TOH, K. & D, M. R. 2007. A unique T cell receptor discovered in marsupials. *Proc Natl Acad Sci USA,* 104**,** 9776 – 9781.

PARSONS, K. R., YOUNG, J. R., COLLINS, B. A. & HOWARD, C. J. 1996. Cattle CTLA-4, CD28 and chicken CD28 bind CD86: MYPPPY is not conserved in cattle CD28. *Immunogenetics,* 43**,** 388-91.

PARTULA, S., DE GUERRA, A., FELLAH, J. S. & CHARLEMAGNE, J. 1996. Structure and diversity of the TCR alpha-chain in a teleost fish. *J Immunol,* 157**,** 207-12.

PASTORI, R. L., MILDE, K. F. & ALEJANDRO, R. 1994. Molecular cloning of the dog homologue of the lymphocyte antigen CD28. *Immunogenetics,* 39**,** 373.

PATEL, S., OVERGARD, A. C. & NERLAND, A. H. 2009. A CD4 homologue in Atlantic halibut (Hippoglossus hippoglossus): molecular cloning and characterisation. *Fish Shellfish Immunol,* 26**,** 377-84.

PAWSON, T. 1995. Protein modules and signalling networks. *Nature,* 373**,** 573-80.

PAWSON, T., GISH, G. & NASH, P. 2001. SH2 domains, interaction modules and cellular wiring. *Trends in Cell Biology,* 11.

PECORARO, M. R., KAWAGUCHI, Y., MIYAZAWA, T., NORIMINE, J., MAEDA, K., TOYOSAKI, T., TOHYA, Y., KAI, C. & MIKAMI, T. 1994. Isolation, sequence and expression of a cDNA encoding the alpha-chain of the feline CD8. *Immunology,* 81**,** 127-31.

PECORARO, M. R., SHIMOJIMA, M., MAEDA, K., INOSHIMA, Y., KAWAGUCHI, Y., KAI, C. & MIKAMI, T. 1996. Molecular cloning of the feline CD8 beta-chain. *Immunology,* 89**,** 84-8.

PELCHEN-MATTHEWS, A., BOULET, I., LITTMAN, D. R., FAGARD, R. & MARSH, M. 1992. The protein tyrosine kinase p56lck inhibits CD4 endocytosis by preventing entry of CD4 into coated pits. *J Cell Biol,* 117**,** 279-90.

PENG, J. & XU, J. 2011. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins,* 79 Suppl 10**,** 161-71.

PERKINS, H. D., VAN LEEUWEN, B. H., HARDY, C. M. & KERR, P. J. 2000. The complete cDNA sequences of IL-2, IL-4, IL-6 AND IL-10 from the European rabbit (Oryctolagus cuniculus). *Cytokine,* 12**,** 555-565.

PERLMUTTER, R. M., MARTH, J. D., LEWIS, D. B., PEET, R., ZIEGLER, S. F. & WILSON, C. B. 1988. Structure and expression of lck transcripts in human lymphoid cells *J. Cell. Biochem.,* 38**,** 111-126.

PETER, M. E., HALL, C., RUHLMAN, A., SANCHO, J. & TERHORST, C. 1992. The T-cell receptor ζ chain contains a GTP/GDP binding site. *EMBO J,* 11**,** 933.

PETERSEN, B., PETERSEN, T. N., ANDERSEN, P., NIELSEN, M. & LUNDEGAARD, C. 2009. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol,* 9**,** 51.

PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods,* 8**,** 785-786.

POLLASTRI, G., PRZYBYLSKI, D., ROST, B. & BALDI, P. 2002. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins,* 47**,** 228-35.

POLLOCK, J. M., WELSH, M. D. & MCNAIR, J. 2005. Immune responses in bovine tuberculosis: towards new strategies for the diagnosis and control of disease. *Vet Immunol Immunopathol,* 108**,** 37-43.

POWELL, F., LAWSON, M., ROTHWELL, L. & KAISER, P. 2009. Development of reagents to study the turkey's immune response: Identification and molecular cloning of turkey CD4, CD8alpha and CD28. *Dev Comp Immunol,* 33**,** 540-6.

PRAKASH, V., BHATTACHARYA, T. K., JYOTSANA, B. & PANDEY, O. P. 2011. Molecular Cloning, Characterization, Polymorphism, and Association Study of the Interleukin-2 Gene in Indian Crossbred Cattle. *Biochem Genet,* In press.

PRASAD, K. V., CAI, Y. C., RAAB, M., DUCKWORTH, B., CANTLEY, L., SHOELSON, S. E. & RUDD, C. E. 1994. T-cell antigen CD28 interacts with the lipid kinase phosphatidylinositol 3-kinase by a cytoplasmic Tyr(P)-Met-Xaa-Met motif. *Proc Natl Acad Sci U S A,* 91**,** 2834-8.

PRESTRIDGE, D. S. 1995. Predicting Pol II promoter sequences using transcription factor binding sites. *J Mol Biol,* 249**,** 923-32.

PYZ, E., NAIDENKO, O., MIYAKE, S., YAMAMURA, T., BERBERICH, I., CARDELL, S., KRONENBERG, M. & HERRMANN, T. 2006. The complementarity determining region 2 of BV8S2 (V beta 8.2) contributes to antigen recognition by rat invariant NKT cell TCR. *J Immunol,* 176**,** 7447-55.

QI, Q. & AUGUST, A. 2007. Keeping the (kinase) party going: SLP-76 and ITK dance to the beat. *Sci STKE,* 2007**,** pe39.

QIAN, D. & WEISS, A. 1997. T cell antigen receptor signal transduction. *Curr Opin Cell Biol,* 9**,** 205-12.

QUINIOU, S. M., SAHOO, M., EDHOLM, E. S., BENGTEN, E. & WILSON, M. 2011. Channel catfish CD8alpha and CD8beta co-receptors: characterization, expression and polymorphism. *Fish Shellfish Immunol,* 30**,** 894-901.

QURESHI, O. S., KAUR, S., HOU, T. Z., JEFFERY, L. E., POULTER, N.S., BRIGGS, Z., KENEFECK, R., WILLOX, A.K., ROYLE, S. J., RAPPOPORT, J.Z. SANSOM, D. M. 2012. Constitutive clathrin-mediated endocytosis of CTLA-4 persists during T cell activation. *J Biol Chem*, 287, 12, 9429-9440.

RABBITTS, T. H., LEFRANC, M. P., STINSON, M. A., SIMS, J. E., SCHRODER, J., STEINMETZ, M., SPURR, N. L., SOLOMON, E. & GOODFELLOW, P. N. 1985. The chromosomal location of T-cell receptor genes and a T cell rearranging gene: possible correlation with specific translocations in human T cell leukaemia. *EMBO J,* 4**,** 1461-5.

RANDELLI, E., FOGLIETTAA, A., MAZZINIA, M., SCAPIGLIATIA, G. & BUONOCORE, F. 2006. Cloning and expression analysis of the co-receptor CD8α in sea bream (Sparus aurata L.). *Aquaculture,* 256**,** 631-637.

RAST, J. P. & LITMAN, G. W. 1994. T-cell receptor gene homologs are present in the most primitive jawed vertebrates. *Proc Natl Acad Sci U S A,* 91**,** 9248-52.

RAVICHANDRAN, K. S. & BURAKOFF, S. J. 1994. The adapter protein Shc interacts with the interleukin-2 (IL-2) receptor upon IL-2 stimulation. *J Biol Chem,* 269**,** 1599-602.

REEVES, R., SPIES, A. G., NISSEN, M. S., BUCK, C. D., WEINBERG, A. D., BARR, P. J., MAGNUSON, N. S. & MAGNUSON, J. A. 2008. Molecular cloning of a functional bovine interleukin 2 cDNA. *Proc. Natl. Acad. Sci. U.S.A.,* 83**,** 3228-3232.

REINHERZ, E. L., ROYER, H. D., CAMPEN, T. J., RAMARLI, D., FABBI, M. & ACUTO, O. 1986. The ontogeny, structure and function of the human T lymphocyte receptor for antigen and major histocompatibility complex. *Biochem Soc Symp,* 51**,** 211-32.

REINHERZ, E. L. & SCHLOSSMAN, S. F. 1980. The differentiation and function of human T lymphocytes: a review. *Cell,* 19**,** 821-827.

REN, R., MAYER, B. J., CICCHETTI, P. & BALTIMORE, D. 1993. Identification of a ten-amino acid proline-rich SH3 binding site. *Science,* 259**,** 1157-61.

RENFREE, M. B., PAPENFUSS, A. T., DEAKIN, J. E., LINDSAY, J., HEIDER, T., BELOV, K., RENS, W., WATERS, P. D., PHARO, E. A., SHAW, G., WONG, E. S., LEFEVRE, C. M., NICHOLAS, K. R., KUROKI, Y., WAKEFIELD, M. J., ZENGER, K. R., WANG, C., FERGUSON-SMITH, M., NICHOLAS, F. W., HICKFORD, D., YU, H., SHORT, K. R., SIDDLE, H. V., FRANKENBERG, S. R., CHEW, K. Y., MENZIES, B. R., STRINGER, J. M., SUZUKI, S., HORE, T. A., DELBRIDGE, M. L., PATEL, H., MOHAMMADI, A., SCHNEIDER, N. Y., HU, Y., O'HARA, W., AL NADAF, S., WU, C., FENG, Z. P., COCKS, B. G., WANG, J., FLICEK, P., SEARLE, S. M., FAIRLEY, S., BEAL, K., HERRERO, J., CARONE, D. M., SUZUKI, Y., SUGANO, S., TOYODA, A., SAKAKI, Y., KONDO, S., NISHIDA, Y., TATSUMOTO, S., MANDIOU, I., HSU, A., MCCOLL, K. A., LANSDELL, B., WEINSTOCK, G., KUCZEK, E., MCGRATH, A., WILSON, P., MEN, A., HAZAR-RETHINAM, M., HALL, A., DAVIS, J., WOOD, D., WILLIAMS, S., SUNDARAVADANAM, Y., MUZNY, D. M., JHANGIANI, S. N., LEWIS, L. R., MORGAN, M. B., OKWUONU, G. O., RUIZ, S. J., SANTIBANEZ, J., NAZARETH, L., CREE, A., FOWLER, G., KOVAR, C. L., DINH, H. H., JOSHI, V., JING, C., LARA, F., THORNTON, R., CHEN, L., DENG, J., LIU, Y., SHEN, J. Y., SONG, X. Z., EDSON, J., TROON, C., THOMAS, D., STEPHENS, A., YAPA, L., LEVCHENKO, T., GIBBS, R. A., COOPER, D. W., SPEED, T. P., FUJIYAMA, A., JA, M. G., et al. 2011. Genome sequence of an Australian kangaroo, Macropus eugenii, provides insight into the evolution of mammalian reproduction and development. *Genome Biol,* 12**,** 414.

RETTIG, L., MCNEILL, L., SARNER, N., GUILLAUME, P., LUESCHER, I., TOLAINI, M., KIOUSSIS, D. & ZAMOYSKA, R. 2009. An essential role for the stalk region of CD8 beta in the coreceptor function of CD8. *J Immunol,* 182**,** 121-9.

RICH, S. S. & RICH, R. R. 1974. Regulatory mechanisms in cell-mediated immune responses. I. Regulation of mixed lymphocyte reactions by alloantigen-activated thymus-derived lymphocytes. *J Exp Med,* 140**,** 1588-603.

ROBBIN, M. G., WAGNER, B., NORONHA, L. E., ANTCZAK, D. F. & DE MESTRE, A. M. 2011. Subpopulations of equine blood lymphocytes expressing regulatory T cell markers. *Vet Immunol Immunopathol,* 140**,** 90-101.

ROGERS, K. A., SCINICARIELLO, F. & ATTANASIO, R. 2006. IgG Fc receptor III homologues in nonhuman primate species: genetic characterization and ligand interactions. *J. Immunol. ,* 177**,** 3848-3856.

ROLLAND-TURNER, M., FARRE, G. & BOUE , F. 2006. Cloning of fox (Vulpes vulpes) Il2, Il6, Il10 and IFNgamma and analysis of their expression by quantitative RT-PCR in fox PBMC after in vitro stimulation by Concanavalin A. *Vet. Immunol. Immunopathol.,* 110**,** 369-375.

ROLLAND, J. & O'HEHIR, R. 1999. "Turning off the T-cells: Peptides for treatment of allergic Diseases" . *Today's life science publishing***,** 32.

ROMANO, T. A., RIDGWAY, S. H., FELTEN, D. L. & QUARANTA, V. 1999. Molecular cloning and characterization of CD4 in an aquatic mammal, the white whale *Delphinapterus leucas*. *Immunogenetics,* 49**,** 376 - 383.

ROMIR, J., LILIE, H., EGERER-SIEBER, C., BAUER, F., STICHT, H. & MULLER, Y. A. 2007. Crystal structure analysis and solution studies of human Lck-SH3; zinc-induced homodimerization competes with the binding of proline-rich motifs. *J Mol Biol,* 365**,** 1417-28.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc,* 5**,** 725-38.

RUDD, C. F., TREVILLYAN, J. M., DASGUPTA, J. D., WONG, L. L. & SCHLOSSMAN, S. F. 1988. The CD4 receptor is complexed in detergent lysates to a protein-tyrosine kinase (pp58) from human T lymphocytes. *Proc Natl Acad Sci USA,* 85**,** 5190-5194.

RULIFSON, I. C., SPERLING, A. I., FIELDS, P. E., FITCH, F. W. & BLUESTONE, J. A. 1997. CD28 costimulation promotes the production of Th2 cytokines. *J Immunol,* 158**,** 658-65.

RUTLEDGE, T., COSSON, P., MANOLIOS, N., BONIFACINO, J. S. & KLAUSNER, R. D. 1992. Transmembrane helical interactions: zeta chain dimerization and functional association with the T cell antigen receptor. *EMBO J,* 11**,** 3245-54.

RYU, S. E., KWONG, P. D., TRUNEH, A., PORTER, T. G., ARTHOS, J., ROSENBERG, M., DAI, X. P., XUONG, N. H., AXEL, R., SWEET, R. W. & ET AL. 1990. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature,* 348**,** 419-26.

SAITOU, N. & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol,* 4**,** 406-25.

SAKAGUCHI, S., SAKAGUCHI, N., ASANO, M., ITOH, M. & TODA, M. 1995. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J Immunol,* 155**,** 1151-64.

SAKAGUCHI, S., YAMAGUCHI, T., NOMURA, T. & ONO, M. 2008. Regulatory T cells and immune tolerance. *Cell,* 133**,** 775-87.

SALI, A. & BLUNDELL, T. L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol,* 234**,** 779-815.

SAMOLLOW, P. B. 2006. Status and applications of genomic resources for the gray, short-tailed opossum, *Monodelphis domestica*, an American marsupial model for comparative biology. *Aust J of Zool,* 54**,** 173-196.

SAMOLLOW, P. B. 2008. The opossum genome: insights and opportunities from an alternative mammal. *Genome Res,* 18**,** 1199-215.

SAN JOSÉ, E., SAHUQUILLO, A. G., BRAGADO, R. & ALARCÓN, B. 1998. Assembly of the TCR/CD3 complex: CD3 epsilon/delta and CD3 epsilon/gamma dimers associate indistinctly with both TCR alpha and TCR beta chains. Evidence for a double TCR heterodimer model. *Eur J Immunol,* 28**,** 12-21.

SASADA, T., TOUMA, M., CHANG, H. C., CLAYTON, L. K., WANG, J. H. & REINHERZ, E. L. 2002. Involvement of the TCR Cbeta FG loop in thymic selection and T cell function. *J Exp Med,* 195**,** 1419-31.

SCHNEIDER, H., PRASAD, K. V., SHOELSON, S. E. & RUDD, C. E. 1995. CTLA-4 binding to the lipid kinase phosphatidylinositol 3-kinase in T cells. *J Exp Med,* 181**,** 351-5.

SCHOUNTZ, T., GREEN, R., DAVENPORT, B., BUNIGER, A., RICHENS, T., ROOT, J. J., DAVIDSON, F., CALISHER, C. H. & BEATY, B. J. 2004. Cloning and characterization of deer mouse (Peromyscus maniculatus) cytokine and chemokine cDNAs. *BMC Immunol.,* 5.

SCHOUNTZ, T., PRESCOTT, J., COGSWELL, A. C., OKO, L., MIROWSKY-GARCIA, K., GALVEZ, A. P. & HJELLE, B. 2007. Regulatory T cell-like responses in deer mice persistently infected with Sin Nombre virus. *Proc Natl Acad Sci U S A,* 104**,** 15496-501.

SCHUBERT, L. A., JEFFERY, E. W., ZHANG, C., RAMSDELL, F. & ZIEGLER, S. F. 2001. Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. *J Biol Chem,* 276**,** 37672-37679.

SCHWARTZ, J. C., ZHANG, X., FEDOROV, A. A., NATHENSON, S. G. & ALMO, S. C. 2001. Structural basis for co-stimulation by the human CTLA-4/B7-2 complex. *Nature,* 410**,** 604-8.

SCHWARTZ, R. H. 2001. Immunology. It takes more than two to tango. *Nature,* 409**,** 31-2.

SCHWARTZ, R. H. 2003. T cell Anergy. *Annu Rev Immunol,* 21**,** 305-334.

SEHGAL, A. & BERGER, M. S. 2000. Basic concepts of immunology and neuroimmunology. *Neurosurg Focus,* 9**,** e1.

SHARMA, R., FU, S. M. & JU, S. T. 2011. IL-2: a two-faced master regulator of autoimmunity. *J Autoimmun,* 36**,** 91-7.

SHEVACH, E. M. 2000. Regulatory T cells in autoimmmunity. *Annu Rev Immunol,* 18**,** 423-49.

SHIN, J., LEE, S. & STROMINGER, J. L. 1993. Translocation of TCR alpha chains into the lumen of the endoplasmic reticulum and their degradation. *Science,* 259**,** 1901-4.

SHIN, S. & STEFFEN, D. L. 1993. Frequent activation of the lck gene by promoter insertion and aberrant splicing in murine leukemia virus-induced rat lymphomas. *Oncogene,* 8**,** 141-149.

SHIRATORI, T., MIYATAKE, S., OHNO, H., NAKASEKO, C., ISONO, K., BONIFACINO, J. S. & SAITO, T. 1997. Tyrosine phosphorylation controls internalization of CTLA-4 by regulating its interaction with clathrin-associated adaptor complex AP-2. *Immunity,* 6**,** 583-9.

SHODA, L. K., BROWN, W. C. & RICE-FICHT, A. C. 1998. Sequence and characterization of phocine interleukin 2. *J. Wildl. Dis. ,* 34**,** 81-90.

SIDDLE, H. V., DEAKIN, J. E., COGGILL, P., WHILMING, L., HARROW, J., KAUFMAN, J., BECK, S. & BELOV, K. 2011. The tammar wallaby major histocompatibility complex shows evidence of past genomic instability. *BMC Genomics,* 12**,** 421.

SIDDLE, H. V., KREISS, A., ELDRIDGE, M. D., NOONAN, E., CLARKE, C. J., PYECROFT, S., WOODS, G. M. & BELOV, K. 2007. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proc Natl Acad Sci U S A,* 104**,** 16221-6.

SMITH, K. A. 1988a. Interleukin-2: inception, impact and implications. *Science,* 240**,** 1169-1176.

SMITH, K. A. 1988b. The interleukin 2 receptor. *Adv Immunol,* 42**,** 165-79.

SMITH, K. A., BAKER, P. E., GILLIS, S. & RUSCETTI, F. W. 1980a. Functional and molecular characteristics of T-cell growth factor. *Mol Immunol,* 17**,** 579-89.

SMITH, K. A., LACHMAN, L. B., OPPENHEIM, J. J. & FAVATA, M. F. 1980b. The functional relationship of the interleukins. *J Exp Med,* 151**,** 1551-6.

SMITH, S. M. & DOCKRELL, H. M. 2000. Role of $CD8^+$ in mycobacterial infections. *Immunol. and Cell Biol. ,* 78**,** 325-333.

SODING, J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics,* 21**,** 951-60.

SOMAMOTO, T., YOSHIURA, Y., NAKANISHI, T. & OTOTAKE, M. 2005. Molecular cloning and characterization of two types of CD8alpha from ginbuna crucian carp, Carassius auratus langsdorfii. *Dev Comp Immunol,* 29**,** 693-702.

SPELLBERG, B. & EDWARDS, J. E., JR. 2001. Type 1/Type 2 immunity in infectious diseases. *Clin Infect Dis,* 32**,** 76-102.

SREEKUMAR, E., PREMRAJ, A. & RASOOL, T. J. 2005. Duck (Anas platyrhynchos), Japanese quail (Coturnix coturnix japonica) and other avian interleukin-2 reveals significant conservation of gene organization, promoter elements and functional residues. *Int J Immunogenet,* 32**,** 355-65.

SRINIVASAN, M., WARDROP, R. M., GIENAPP, I. E., STUCKMAN, S. S., WHITACRE, C. C. & KAUMAYA, P. T. 2001. A retro-inverso peptide mimic of CD28 encompassing the MYPPPY motif adopts a polyproline type II helix and inhibits encephalitogenic T cells in vitro. *J Immunol,* 167**,** 578-85.

ST-LAURENT, G., BELIVEAU, C. & ARCHAMBAULT, D. 1999. Molecular cloning and phylogenetic analysis of beluga whale (Delphinapterus leucas) and grey seal (Halichoerus grypus) interleukin 2. *Vet. Immunol. Immunopathol.,* 67**,** 385-394.

STARR, T. K., JAMESON, S. C. & HOGQUIST, K. A. 2003. Positive and negative selection of T cells. *Annu Rev Immunol,* 21**,** 139-76.

STEFANOWICZ, P., KIJEWSKA, M., KLUCZYK, A. & SZEWCZUK, Z. 2010. Detection of glycation sites in proteins by high-resolution mass spectrometry combined with isotopic labeling. *Anal Biochem,* 400**,** 237-43.

STOCKINGER, B. & VELDHOEN, M. 2007. Differentiation and function of Th17 T cells. *Curr Opin in Immunol,* 19 281–286.

STONE, W. H., BRUUN, D. A., FOSTER, E. B., MANIS, G. H., HOFFMAN, E. S., SAPHIRE, D. G., VANDEBERG, J. L. & INFANTE, A. J. 1998. Absence of a significant mixed lymphocyte reaction in a marsupial (*Monodelphis domestica*). *Lab Anim Sci,* 48**,** 184-189.

STONE, W. H., BRUUN, D. A., MANIS, G. S., HOLSTE, S. B., HOFFMAN, E. S., SPONG, K. D. & WALUNAS, T. 1996. The Immunobiology of the Marsupial *Monodelphis domestica*. In: Stolen J S, Fletcher, T C, Bayne C J, et al. (eds) Modulators of immune responses: the evolutionary trail. *SOS Publications, Fair Haven, NJ, pp 149-165.*

STONE, W. H., MANIS, G. H., HOFFMAN, E. S., SAPHIRE, D. G., HUBBARD, G. B., VANDEBERG, J. L. & INFANTE, A. J. 1997a. Fate of allogeneic skin tranplantations in a marsupial *Monodelphis domestica*. *Lab Anim Sci,* 47**,** 283-287.

STONE, W. H., MANIS, G. S. & HOFFMAN 1997b. Allogeneic skin transplantation in a marsupial. *Lab. Anim. Sci.,* 47**,** 283-287.

STRAUS, D. B. & WEISS, A. 1993. The CD3 chains of the T cell antigen receptorassociates with the ZAP-70 tyrosine kinase and are tyrosine phosphorylated after receptor stimulation. *J. Exp. Med.,* 178**,** 1523-1530.

STREBHARDT, K., MULLINS, J. I., BRUCK, C. & RUBSAMEN-WAIGMANN, H. 1987. Additional member of the protein-tyrosine kinase family: the src- and lck-related protooncogene c-tkl. *Proc. Natl. Acad. Sci. U.S.A.,* 84**,** 8778-8782

SUETAKE, H., ARAKI, K. & SUZUKI, Y. 2004. Cloning, expression, and characterization of fugu CD4, the first ectothermic animal CD4. *Immunogenetics,* 56**,** 368 - 374.

SUKHATME, V. P., SIZER, K. C., VOLLMER, A. C., HUNKAPILLER, T. & PARNES, J. R. 1985. The T cell differentiation antigen Leu-2/T8 is homologous to immunoglobulin and T cell receptor variable regions. *Cell,* 40**,** 591-7.

SUNDICK, R. S. & GILL-DIXON, C. 1997. A cloned chicken lymphokine homologous to both mammalian IL-2 and IL-15. *J Immunol,* 159**,** 720-5.

SUVAS, S. & ROUSE, B. T. 2006. Treg control of antimicrobial T cell responses. *Curr Opin Immunol,* 18**,** 344-8.

TAKAHASHI, T., KUNIYASU, Y., TODA, M., SAKAGUCHI, N., ITOH, M., IWATA, M., SHIMIZU, J. & SAKAGUCHI, S. 1998. Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int Immunol,* 10**,** 1969-80.

TAKANO, M., HAYASHI, N., GOITSUKA, R., OKUDA, M., MOMOI, Y., YOUN, H. Y., WATARI, T., TSUJIMOTO, H. & HASEGAWA, A. 1994. Identification of dog T-cell receptor beta chain genes. *Immunogenetics,* 40**,** 246.

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M. & KUMAR, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol*.

TAVERNOR, A. S., ALLEN, W. R. & BUTCHER, G. W. 1993. cDNA cloning of equine interleukin-2 by polymerase chain reaction. *Equine Vet. J.,* 25**,** 242-243.

TAYLOR, S. S., RADZIO-ANDZELM, E. & HUNTER, T. 1995. How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor protein-tyrosine kinase. *FASEB J,* 9**,** 1255-66.

TERZO, E. A., DE VILLARREAL, M. P., MICK, V., MUNOZ, F., AMORENA, B., DE ANDRES, D. & PEREZ DE LA LASTRA, J. M. 2006. Molecular cloning of multiple forms of the ovine B7-2 (CD86) costimulatory molecule. *Vet Immunol Immunopathol,* 114**,** 149-58.

THE MHC SEQUENCING CONSORTIUM 1999. Complete sequence and gene map of a human major histocompatibility complex. . *Nature,* 401**,** 921-3.

THOME, A., SAALMULLER, A. & PFAFF, E. 1993. Molecular cloning of porcine T cell receptor alpha, beta, gamma and delta chains using polymerase chain reaction fragments of the constant regions. *Eur J Immunol,* 23**,** 1005-10.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res,* 22**,** 4673-80.

TJOELKER, L. W., CARLSON, L. M., LEE, K., LAHTI, J., MCCORMACK, W. T., LEIDEN, J. M., CHEN, C. L., COOPER, M. D. & THOMPSON, C. B. 1990. Evolutionary conservation of antigen recognition: the chicken T-cell receptor beta chain. *Proc Natl Acad Sci U S A,* 87**,** 7856-60.

TOMPKINS, D., HUDGENS, E., HOROHOV, D. & BALDWIN, C. L. 2010. Expressed gene sequences of the equine cytokines interleukin-17 and interleukin-23. *Vet Immunol Immunopathol,* 133**,** 309-13.

TORDAI, H., BANYAI, L. & PATTHY, L. 1999. The PAN module: the N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. *FEBS Lett,* 461**,** 63-7.

TOUMA, M., CHANG, H. C., SASADA, T., HANDLEY, M., CLAYTON, L. K. & REINHERZ, E. L. 2006. The TCR C beta FG loop regulates alpha beta T cell development. *J Immunol,* 176**,** 6812-23.

TOUSSIROT, E., SAAS, P., DESCHAMPS, M., POUTHIER, F., PERROT, L., PERRUCHE, S., CHABOD, J., TIBERGHIEN, P. & WENDLING, D. 2009. Increased production of soluble CTLA-4 in patients with spondylarthropathies correlates with disease activity. *Arthritis Res Ther,* 11**,** R101.

TREGASKES, C. A., KONG, F. K., PARAMITHIOTIS, E., CHEN, C. L., RATCLIFFE, M. J., DAVISON, T. F. & YOUNG, J. R. 1995. Identification and analysis of the expression of CD8 alpha beta and CD8 alpha alpha isoforms in chickens reveals a major TCR-gamma delta CD8 alpha beta subset of intestinal intraepithelial lymphocytes. *J Immunol,* 154**,** 4485-94.

TSYTSIKOV, V. N., YUROVSKY, V. V., ATAMAS, S. P., ALMS, W. J. & WHITE, B. 1996. Identification and characterization of two alternative splice variants of human interleukin-2. *J Biol Chem,* 271**,** 23055-60.

TURNER, J. M., BRODSKY, M. H., IRVING, B. A., LEVIN, S. D., PERLMUTTER, R. M. & LITTMAN, D. R. 1990. Interaction of the unique N-terminal region of tyrosine kinase p56lck with cytoplasmic domains of CD4 and CD8 is mediated by cysteine motifs. *Cell,* 60**,** 755 - 765.

TURNI, C. & SMALES, L. R. 2001. Parasites of the bridled nailtail wallaby (Onychogalea fraenata) (Marsupialia:Macropodidae). *Wildlife Research,* 28**,** 403-411.

TYNDALE-BISCOE, C. H. 1973. Life of Marsupials. *Edward Arnold Ltd, London***,** 147-169.

UDA, A., TANABAYASHI, K., MUKAI, R., YACHI, M., NAM, K. H. & YAMADA, A. 2001. CD3 polymorphism in cynomolgus monkeys (Macaca fascicularis). *J Med Primatol,* 30**,** 141-7.

UENISHI, H., EGUCHI, T., SUZUKI, K., SAWAZAKI, T., TOKI, D., SHINKAI, H., OKUMURA, N., HAMASIMA, N. & AWATA, T. 2004. PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucleic Acids Res,* 32**,** D484-8.

URETA-VIDAL, A., GARCIA, Z., LEMONNIER, F. A. & KAZANJI, M. 1999. Molecular characterization of cDNAs encoding squirrel monkey (Saimiri sciureus) CD8 alpha and beta chains. *Immunogenetics,* 49**,** 718-21.

VAN BEELEN, A. J., ZELINKOVA, Z., TAANMAN-KUETER, E. W., MULLER, F. J., HOMMES, D. W., ZAAT, S. A., KAPSENBERG, M. L. & DE JONG, E. C. 2007. Stimulation of the intracellular bacterial sensor NOD2 programs dendritic cells to promote interleukin-17 production in human memory T cells. *Immunity,* 27**,** 660-9.

VAN DEN STEEN, P., RUDD, P. M., DWEK, R. A. & OPDENAKKER, G. 1998. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol,* 33**,** 151-208.

VAN RIJT, L. S., VOS, N., WILLART, M., KLEINJAN, A., COYLE, A. J., HOOGSTEDEN, H. C. & LAMBRECHT, B. N. 2004. Essential role of dendritic cell CD80/CD86 costimulation in the induction, but not reactivation, of TH2 effector responses in a mouse model of asthma. *J Allergy Clin Immunol,* 114**,** 166-73.

VILLINGER, F., BOSTIK, P., MAYNE, A., KING, C. L., GENAIN, C. P., WEISS, W. R. & ANSARI, A. A. 2001. Cloning, sequencing, and homology analysis of nonhuman primate Fas/Fas-ligand and co-stimulatory molecules. *Immunogenetics,* 53**,** 315-28.

VILLINGER, F., BRAR, S. S., MAYNE, A., CHIKKALA, N. & ANSARI, A. A. 1995. Comparative sequence analysis of cytokine genes from human and nonhuman primates. *J. Immunol.,* 155**,** 3946-3954.

VOLPE, E., SERVANT, N., ZOLLINGER, R., BOGIATZI, S. I., HUPE, P., BARILLOT, E. & SOUMELIS, V. 2008. A critical function for transforming growth factor-beta, interleukin 23 and proinflammatory cytokines in driving and modulating human T(H)-17 responses. *Nat Immunol,* 9**,** 650-7.

VORONOVA, A. F. & SEFTON, B. M. 1986. Expression of a new tyrosine protein kinase is stimulated by retrovirus promoter insertion. *Nature,* 319**,** 682-685.

WAKSMAN, G., KOMINOS, D., ROBERTSON, S. C., PANT, N., BALTIMORE, D., BIRGE, R. B., COWBURN, D., HANAFUSA, H., MAYER, B. J., OVERDUIN, M., RESH, M. D., RIOS, C. B., SILVERMAN, L. & KURIYAN, J. 1992. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature,* 358**,** 646-53.

WALDMANN, T. A. 2006. Effective cancer therapy through immunomodulation. *Annu Rev Med,* 57**,** 65-81.

WALDMANN, T. A., DUBOIS, S. & TAGAYA, Y. 2001. Contrasting roles of IL-2 and IL-15 in the life and death of lymphocytes: implications for immunotherapy. *Immunity,* 14**,** 105-10.

WALKER, L. S. & SANSOM, D. M. 2011. The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. *Nat Rev Immunol,* 11**,** 852-63.

WALSH, P. T., TAYLOR, D. K. & TURKA, L. A. 2004. Tregs and transplantation tolerance. *J Clin Invest,* 114**,** 1398-403.

WANDLESS, T. J. 1996. SH2 domains: a question of independence. *Curr Biol,* 6**,** 125-7.

WANG, J., LIM, K., SMOLYAR, A., TENG, M., LIU, J., TSE, A. G., HUSSEY, R. E., CHISHTI, Y., THOMSON, C. T., SWEET, R. M., NATHENSON, S. G., CHANG, H. C., SACCHETTINI, J. C. & REINHERZ, E. L.

1998a. Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR fab fragment derived from a mitogenic antibody. *EMBO J,* 17**,** 10-26.

WANG, J. H., LIM, K., SMOLYAR, A., TENG, M. K., LIU, J. H., TSE, A. G. D., LIU, J., HUSSEY, R. E., CHISHTI, Y., THOMSON, C. T., SWEET, R. M., NATHESNSON, S. G., C., C. H., SACCHETTINI, J. C. & REINHERZ, E. L. 1998b. Atomic structure of an αβ T cell receptor (TCR) heterodimer in comples with an anti-TCR Fab fragment derived from a mitogenic antibody. *EMBO,* 17**,** 10 - 26.

WANG, L. & BOSSELUT, R. 2009. CD4-CD8 lineage differentiation: Thpok-ing into the nucleus. *J Immunol,* 183**,** 2903-10.

WANG, L., FAN, J., YU, M., ZHENG, S. & ZHAO, Y. 2011a. Association of goat (Capra hircus) CD4 gene exon 6 polymorphisms with ability of sperm internalizing exogenous DNA. *Mol Biol Rep,* 38**,** 1621-8.

WANG, Q., SMITH, J. B., HARRISON, M. L. & GEAHLEN, R. L. 1991. Identification of tyrosine 67 in bovine brain myelin basic protein as a specific phosphorylation site for thymus p56lck. *Biochem. Biophys. Res. Commun.,* 178**,** 1393.

WANG, Q. J., HANADA, K. & YANG, J. C. 2008. Characterization of a novel nonclassical T cell clone with broad reactivity against human renal cell carcinomas. *J Immunol,* 181**,** 3769-76.

WANG, X., PARRA, Z. E. & MILLER, R. D. 2011b. Platypus TCRmu provides insight into the origins and evolution of a uniquely mammalian TCR locus. *J Immunol,* 187**,** 5246-54.

WANG, Z., HUBBARD, G. B., PATHAK, S. & VANDEBERG, J. L. 2003. In Vivo Opossum Xenograft Model for Cancer Research. *Cancer Research,* 63**,** 6121 - 6124.

WANGE, R. L., KONG, A. N. & SAMELSON, L. E. 1992. A tyrosine-phosphorylated 70kDa protein binds a photoaffinity analogue of ATP and associates with both the zeta chain and CD3 components of the activated T cell antigen receptor. *J Biol Chem,* 267**,** 11685-11688.

WANGE, R. L., MALEK, S. N., DESIDERIO, S. & SAMELSON, L. E. 1993. Tandem SH2 domains of ZAP-70 bind to T cell antigen receptor zeta and CD3 epsilon from activated Jurkat T cells. *J Biol Chem,* 268**,** 19797-801.

WARING, H., HOLMES, R., COCKSON, A., ASHMAN, R. B. & STANLEY, N. V. 1978. Induction of long-term tolerance in the quokka (*Setonix brachyurus*) by thymus and skin allografts into early pouch young. *Aust. J. Exp. Bio. Med. Sci.,* 56**,** 597-604.

WASS, M. N., KELLEY, L. A. & STERNBERG, M. J. 2010. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res,* 38**,** W469-73.

WATANABE, M., IWASAKI, Y., MITA, Y., OTA, S., YAMADA, S., SHIMIZU, M. & TAKAGAKI, Y. 2007. Porcine T-cell receptor beta-chain: a genomic sequence covering Dbeta1.1 to Cbeta2 gene segments and the diversity of cDNA expressed in piglets including novel alternative splicing products. *Mol Immunol,* 44**,** 2332-43.

WATERHOUSE, P., PENNINGER, J. M., TIMMS, E., WAKEHAM, A., SHAHINIAN, A., LEE, K. P., THOMPSON, C. B., GRIESSER, H. & MAK, T. W. 1995. Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science,* 270**,** 985-8.

WEAVER, J. R., GOOD, K., WALTERS, R. D., KUGEL, J. F. & GOODRICH, J. A. 2007. Characterization of the sequence and architectural constraints of the regulatory and core regions of the human interleukin-2 promoter. *Mol Immunol,* 44**,** 2813–2819.

WEDLOCK, D. N., ALDWELL, F. E. & BUDDLE, B. M. 1996. Molecular cloning and characterization of tumour necrosis factor alpha (TNF-a) from the Australian common brushtail possum, *Trichosurus vulpecula*.

WEDLOCK, D. N., ALDWELL, F. E. & BUDDLE, B. M. 1998. Nucleotide sequence of a marsupial interleukin-10 cDNA from the Australian brushtail possum (Trichosurus vulpecula). *DNA Seq,* 9**,** 239-44.

WEDLOCK, D. N., GOH, L. P., PARLANE, N. A. & BUDDLE, B. M. 1999. Molecular cloning and physiological effects of brushtail possum interleukin-1beta. *Vet Immunol Immunopathol,* 67**,** 359-72.

WEISS, A. & LITTMAN, D. R. 1994. Signal transduction by lymphocyte antigen receptors. *Cell,* 76**,** 263-74.

WEISSMAN, A. M., HOU, D., ORLOFF, D. G., MODI, S. W., SEUANEZ, H., O'BRIEN, S. J. & KLAUSNER, R. D. 1988. Molecular cloning and chromosomal localizaiton of the human T-cell receptor ζ chain: Distinction from the molecular CD3 complex. *Proc Natl. Acad Sci USA,* 85**,** 9709 - 9713.

WEISSMAN, A. M., HOU, D., ORLOFF, D. G., MODI, W. S., SEUANEZ, H., O'BRIEN, S. J. & KLAUSNER, R. D. 1998. Molecular cloning of the ζ chain of the T cell antigen receptor. . *Science* 239**,** 1018-1021.

WELLS, A. D., WALSH, M. C., BLUESTONE, J. A. & TURKA, L. A. 2001. Signaling through CD28 and CTLA-4 controls two distinct forms of T cell anergy. *J Clin Invest,* 108**,** 895-903.

WIEST, D. L., ASHE, J. M., HOWCROFT, T. K., LEE, H. M., KEMPER, D. M., NEGISHI, I., SINGER, D. S., SINGER, A. & ABE, R. 1997. A spontaneously arising mutation in the DLAARN motif of murine ZAP-70 abrogates kinase activity and arrests thymocyte development. *Immunity,* 6**,** 663-71.

WILKINSON, R., KOTLARSKI, I. & BARTON, M. 1992. Koala lymphoid cell: analysis of antigen-specific responses. *Vet. Immunol. Immunopathol,* 33**,** 237-247.

WILKINSON, R., KOTLARSKI, I. & BARTON, M. 1994. Further characterisation of the immune response of the koala. *Vet Immunol Immunopathol,* 40**,** 325-39.

WILLIAMSON, M. P. 1994. The structure and function of proline-rich regions in proteins. *Biochem. J.,* 297**,** 249 - 260.

WILSON, M. R., ZHOU, H., BENGTEN, E., CLEM, L. W., STUGE, T. B., WARR, G. W. & MILLER, N. W. 1998. T-cell receptors in channel catfish: structure and expression of TCR alpha and beta genes. *Mol Immunol,* 35**,** 545-57.

WINEMAN, J. P., GILMORE, G. L., GRITZMACHER, C., TORBETT, B. E. & MULLER-SIEBURG, C. E. 1992. CD4 is expressed on murine pluripotent hematopoietic stem cells. *Blood,* 80**,** 1717-24.

WONG, E. S. W., YOUNG, L. J., PAPENFUSS, A. T. & BELOV, K. 2006. In *silico* identification of oposum cytokine genes suggests the complexity of the marsupial immune system rivals that of eutherian mammals. *Immunome Research,* 24.

WOODBURNE, M. O., RICH, T. H. & SPRINGER, M. S. 2003. The evolution of tribospheny and the antiquity of mammalian clades. *Molecular Phylogenetics and Evolution,* 28**,** 360-385.

WU, H., KWONG, P. D. & HENDRICKSON, W. A. 1997. Dimeric association and segmental variability in the structure of human CD4. *Nature,* 387**,** 527-30.

WU, S. & ZHANG, Y. 2010. Recognizing protein substructure similarity using segmental threading. *Structure,* 18**,** 858-67.

WU, Y., BORDE, M., HEISSMEYER, V., FEUERER, M., LAPAN, A. D., STROUD, J. C., BATES, D. L., GUO, L., HAN, A., ZIEGLER, S. F., MATHIS, D., BENOIST, C., CHEN, L. & RAO, A. 2006. FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell,* 126**,** 375-87.

WU, Y., WU, W., WONG, W. M., WARD, E., THRASHER, A. J., GOLDBLATT, D., OSMAN, M., DIGARD, P., CANADAY, D. H. & GUSTAFSSON, K. 2009. Human gamma delta T cells: a lymphoid lineage cell capable of professional phagocytosis. *J Immunol,* 183**,** 5622-9.

XU, C., GAGNON, E., CALL, M. E., SCHNELL, J. R., SCHWIETERS, C. D., CARMAN, C. V., CHOU, J. J. & WUCHERPFENNIG, K. W. 2008. Regulation of T cell receptor activation by dynamic membrane binding of the CD3ε cytoplasmic tyrosine-based motif. *Cell,* 135**,** 702 - 713.

XU, S. & CAO, X. 2010. Interleukin-17 and its expanding biological functions. *Cell Mol Immunol,* 7**,** 164-74.

YABE, M., MATSUURA, Y. & TATSUMI, M. 1997. Molecular cloning and expression of cynomolgus monkey interleukin-2 cDNA by the recombinant baculovirus system. *Int. Arch. Allergy Immunol. ,* 113.

YANDELL, M., MUNGALL, C. J., SMITH, C., PROCHNIK, S., KAMINKER, J., HARTZELL, G., LEWIS, S. & RUBIN, G. M. 2006. Large-scale trends in the evolution of gene structures within 11 animal genomes. *PLoS Comput Biol,* 2**,** e15.

YANG, D., ROGGENDORF, M. & LU, M. 2003a. Molecular characterization of CD28 and cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) of woodchuck (Marmota monax). *Tissue Antigens,* 62**,** 225-32.

YANG, S. & SIM, G. K. 1999. New forms of dog CD80 and CD86 transcripts that encode secreted B7 molecules. *Immunogenetics,* 50**,** 349-53.

YANG, Y., VILLAIN, P., MUSTELIN, T. & COUTURE, C. 2003b. Critical role of Ser-520 phosphorylation for membrane recruitment and activation of the ZAP-70 tyrosine kinase in T cells. *Mol Cell Biol,* 23**,** 7667-77.

YAO, Q., FISCHER, K. P., MOTYKA, B., FERLAND, S., LI, L., TYRRELL, D. L. & GUTFREUND, K. S. 2010. Identification of cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4) isoforms in the Pekin duck. *Dev Comp Immunol,* 34**,** 749-58.

YAO, Z., FANSLOW, W. C., SELDIN, M. F., ROUSSEAU, A. M., PAINTER, S. L., CONNEAU, M. R., COHEN, J. I. & SPRIGGS, M. K. 1995a. Herpesvirus samiri encodes a new cytokine, IL-17, which binds to a novel cytokine receptor. *Immunity* 3**,** 811-821.

YAO, Z., PAINTER, S. L., FANSLOW, W. C., ULRICH, D., MACDUFF, B. M., SPRIGGS, M. K. & ARMITAGE, R. J. 1995b. Human IL-17: a novel cytokine derived from T cells. *J. Immunol.,* 155**,** 5483-5486.

YAO, Z., TIMOUR, M., PAINTER, S., FANSLOW, W. & SPRIGGS, M. 1996. Complete nucleotide sequence of the mouse CTLA8 gene. *Gene,* 168**,** 223-5.

YASUDA, K., KOSUGI, A., HAYASHI, F., SAITOH, S., NAGAFUKU, M., MORI, Y., OGATA, M. & HAMAOKA, T. 2000. Serine 6 of Lck tyrosine kinase: a critical site for Lck myristoylation, membrane localization, and function in T lymphocytes. *J Immunol,* 165**,** 3226-31.

YONGGANG, L. & XUESHAN, X. 2011. Molecular characterization, tissue expression, polymorphism and association of porcine LCK gene. *Mol. Biol. Rep.,* In press.

YOSHIKAI, Y., CLARK, S. P., TAYLOR, S., SOHN, U., WILSON, B. I., MINDEN, M. D. & MAK, T. W. 1985. Organization and sequences of the variable, joining and constant region genes of the human T-cell receptor alpha-chain. *Nature,* 316**,** 837-40.

YOUNG, J. R., DAVISON, T. F., TREGASKES, C. A., RENNIE, M. C. & VAINIO, O. 1994. Monomeric homologue of mammalian CD28 is expressed on chicken T cells. *J Immunol,* 152**,** 3848-51.

YOUNG, L. J. 2010. Expressed sequence identification and characterisation of the cDNA for Interleukin-4 from the mitogen-stimulated lymphoid tissue of a marsupial, *Macropus eugenii*. *Vet. Immunol. Immunopathol.*

YOUNG, L. J., CROSS, M. L., DUCKWORTH, J. A., FLENADY, S. & BELOV, K. 2011. Molecular identification of interleukin-2 in the lymphoid tissues of the common brushtail possum, *Trichosurus vulpecula*. *Dev & Comp Immunol,* Manuscript accepted.

YOUNG, L. J. & DEANE, E. M. 2003. Immunocytochemical identifiaction of CD5 positive peripheral blood lymphocytes in four marsupial species. *Comparative Clinical Pathology,* 12**,** 16 - 32.

YOUNG, L. J. & DEANE, E. M. 2005. Morphology and ultrastructure of blood cellls of the tammar wallaby Macropus eugenii. *Comparative Clinical Pathology,* 14**,** 36 - 47.

YOUNG, L. J., MCFARLANE, R., SLENDER, A. L. & DEANE, E. M. 2003. Histological and immunohistological investigation of the lymphoid tissue in normal and mycobacteria-affected specimens of the Rufous Hare-wallaby (*Lagorchestes hirsutus)*. *Anatomy,* 202**,** 315 - 325.

YU, H., CHEN, S., XI, D., HE, Y., LIU, Q., MAO, H. & DENG, W. 2010. Molecular cloning, sequence characterization and tissue transcription profile analyses of two novel genes: LCK and CDK2 from the Black-boned sheep (Ovis aries). *Mol. Biol. Rep.,* 37**,** 39-45.

YU, H. & SCHREIBER, S. L. 1994. Signalling an interest. *Nature Structural Biology,* 1**,** 417-420.

YU, T. K., CAUDELL, E. G., SMID, C. & GRIMM, E. A. 2000. IL-2 activation of NK cells: involvement of MKK1/2/ERK but not p38 kinase pathway. *J Immunol,* 164**,** 6244-6251.

ZAMOYSKA, R. 1994. The CD8 coreceptor revisted: one chain good, two chains better. *Immunity,* 1**,** 243-246.

ZHANG, Y. A., HIKIMA, J., LI, J., LAPATRA, S. E., LUO, Y. P. & SUNYER, J. O. 2009. Conservation of structural and functional features in a primordial CD80/86 molecule from rainbow trout (Oncorhynchus mykiss), a primitive teleost fish. *J Immunol,* 183**,** 83-96.

ZHANG, Z. D., WEINSTOCK, G. & GERSTEIN, M. 2008. Rapid evolution by positive Darwinian selection in T-cell antigen CD4 in primates. *J Mol Evol,* 66**,** 446-56.

ZHOU, J. Y., WANG, J. Y., CHEN, J. G., WU, J. X., GONG, H., TENG, Q. Y., GUO, J. Q. & SHEN, H. G. 2005. Cloning, in vitro expression and bioactivity of duck interleukin-2. *Mol Immunol,* 42**,** 589-98.

ZHU, L., RUAN, X. D., GE, Y. F., WAN, Q. H. & FANG, S. G. 2007. Low major histocompatibility complex class II DQA diversity in the Giant Panda (Ailuropoda melanoleuca). *BMC Genet,* 8**,** 29.

ZHU, X., KIM, J. L., NEWCOMB, J. R., ROSE, P. E., STOVER, D. R., TOLEDO, L. M., ZHAO, H. & MORGENSTERN, K. A. 1999. Structural analysis of the lymhocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure,* 7**,** 651-661.

ZIMIN, A. V., DELCHER, A. L., FLOREA, L., KELLEY, D. R., SCHATZ, M. C., PUIU, D., HANRAHAN, F., PERTEA, G., VAN TASSELL, C. P., SONSTEGARD, T. S., MARCAIS, G., ROBERTS, M., SUBRAMANIAN, P., YORKE, J. A. & SALZBERG, S. L. 2009. A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol,* 10**,** R42.

ZOLLER, K. E., MACNEIL, I. A. & BRUGGE, J. S. 1997. Protein tyrosine kinases Syk and ZAP-70 display distinct requirements for Src family kinases in immune response receptor signal transduction. *J Immunol,* 158**,** 1650-9.

ZUCCOLOTTO, P., E., HARRISON, G. A. & DEANE, E. M. 2000. Cloning of marsupial T cell receptor α and β constant region cDNAs. *Immunology and Cell Biology,* 78**,** 103 - 109.

Appendix 3A

Monodelphis domestica (South American grey short tailed opossum) CD3ε nucleotide sequence – open reading frame

atgcagttgggatctctctggaccgtgcttggattctttcttctctcagcttgtgtttgggggggaa
gacttagaagaagatcctcagaaatacaaatttggagtctccatctcaggaacccaagtgacactt
acctgccctgagaaatcagaagatcttataatatggaagaaaaataatgtactaataaatggtgtg
gaaagttatcaactcactctagatgactcagagactgaatacagcggccacttccattgtaaaaaa
aaatccagtccatcagatgaaggctattttctttacctgaaagcaagagtgtgtcatggttgcttg
gagatgggtgtgctgactgtggctgggattattattgctgatgtcttcatcactctgggagtactg
attttggtatatcactggagcaagaagcaaaaggccaagagcaagccggttcgaggggggaggtgct
ggtggcaagactagaggagtaaacaaagagaggcctccacctgttcccaaccctgactatgagccc
atccgtaaaggtcaacgggaactgtatgctggcctgaatcagagagccatatga

M. domestica putative protein sequence

MQLGSLWTVLGFFLLSACVWGEDLEEDPQKYKFGVSISGTQVTLTCPEKSEDLIIWKKNNLINGVE
SYQLTLDDSETEYSGHFHCKKKSSPSDEGYFLYLKARVCHGCLEMGVLTVAGIIIADVFITLGVLI
LVYHWSKKQKAKSKPVRGGGAGGKTRGVNKERPPPVPNPDYEPIRKGQRELYAGLNQRAI

Appendix 3A

Secondary structure prediction for *M. domestica* CD3ε using PSIpred

Figure 3A.1. Secondary structure prediction for *M. domestica* CD3ε.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 3A

Aligment between predicted *M. domestica* putative protein sequence and expressed sequence.

```
M. domestica          MQLGSLWTVLGFFLLSACVWGEDLEEDPQKYKFGVSISGTQVTLTCPEKSEDLIIWKKN- 59
pred. M. domestica    MQLGSLWTVLGFFLLSACVWGEDLEEEFG-----VSISGTQVTLTCPEKSEDLIIWKKNN 55
                      *************************.       ************************

M. domestica          NLINGVESYQLTLDDSETEYSGHFHCKKKSSPSDEGYFLYLKARVCHGCLEMGVLTVAGI 119
pred. M. domestica    VLINGVESYQLTLDDSETEYSGHFHCKKKSSPSDEGYFLYLKARVCHGCLEMGVLTVAGI 115
                      ************************************************************

M. domestica          IIADVFITLGVLILVYHWSKKQKAKSKPVRGGGAGGKTRGVNKERPPPVPNPDYEPIRKG 179
pred. M. domestica    IIADVFITLGVLILVYHWSKKQKAKSKPVRGGGAGGKTRGVNKESPPPVPNPDYEPIRKG 175
                      ****************************************** ***************

M. domestica          QRELYAGLNQRAI                                               192
pred. M. domestica    QRELYAGLNQRAM                                               188
                      ************.
```

Figure 3A.2. Alignment of the predicted *M. domestica* sequence with the expressed sequence. Blue highlight = the insert in the expressed sequence. Magenta highlight = differences between the expressed and predicted sequences.

Appendix 3A

Onychogalea fraenata (Bridled nailtail wallaby) - CD3ε nucleotide sequence – open reading frame

5'end

agcattcaacaaagatgttccttggaggcctcaccttctagcaggagctttgaactaagctcaagattacttt
ttcagaagaaataaacccaccaaaccctgcctgctggatcaaggatccaagcagagaatg
atgcatttggaagctctctggactgtggtaggattctgtctgctctcagcctgtgtctgggggcaaagcccgg
aaggcgaatttgacgtctacatctcaggaactgaagtaatactcacctgccccgataaaactagtgaggacat
agaatggaagaaaaatgatgaaaccgtaaacggtgtggacggcagtacactcaccctaacaaactccgagatt
cagtatggctacttcctttgtaaaaagaaaggatcaaaagatcacgaaggccattatctctacctgaaagcaa
gagtatgtgaaggttgtgtggaaatggacgtgctgacggtggctgggattgtcattgctgacgtcttcatcac
tctgggagtgctgctcttggtgtattactggagcaaggcgcgaaaggccaaggccaagcctgttggtcgaggg
ggaggtggcggtggcaggacaagaggagcaaacaaggagaggcctccacctgttcccaaccctgactatgagc
ccatccgcaaaggccaacgggacctgtatgctggcctgaatcagagagccatctga
3'end

gaatccctgaagagctcctccttccactgtggctcctattctgggacttctacttttaccctccacttccagc
ctggggcatcccagaactataggatacaaaaagaaaaaatggttttctcttctaaatgttgattctacttcta
gaaaaactaacttctctgccccccacagtcttgttctcagcaaccttcttgctctccactctaattccccaacc
atagggcataactattttcctactgttgaaattatatttacaaggtatcatggacaccaaccttaaaggataa
acccaaagctctggaacctttgaaccccttgctcgttccctttgtgatttggtgtttttatggtttgctctttc
ttgtttccttttttcttgctttctttttttttatttcagattgtcccagtacccctttttccacccctttttcc
gtgcctagcttgtcactttttttttcatttttttggaacctggcatgtataatgttttcaagtgttgtgtccccc
acatggagtttggtgccagagtctttccctctagcttaattttgcctgccttctttgtgaaatatttggtttt
catgtagactatggtaactcattaaagtacatttgagctgaaaaaaaaaaaaaaaaaaaaaaaaa

O. fraenata putative protein sequence

MHLEALWTVVGFCLLSACVWGQSPEGEFDVYISGTEVILTCPDKTSEEIEWKKNDETVKGVDGSTLTLTNSEI
QYGYFLGKKKGSKDHEGHYLYLKAKVCEGCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKAKPVGRG
GGGGGRTRGANKERPPPVPNPDYEPIRKGQRDLYAGLNQRAI

Secondary structure prediction for *O. fraenata* CD3ε using PSIpred

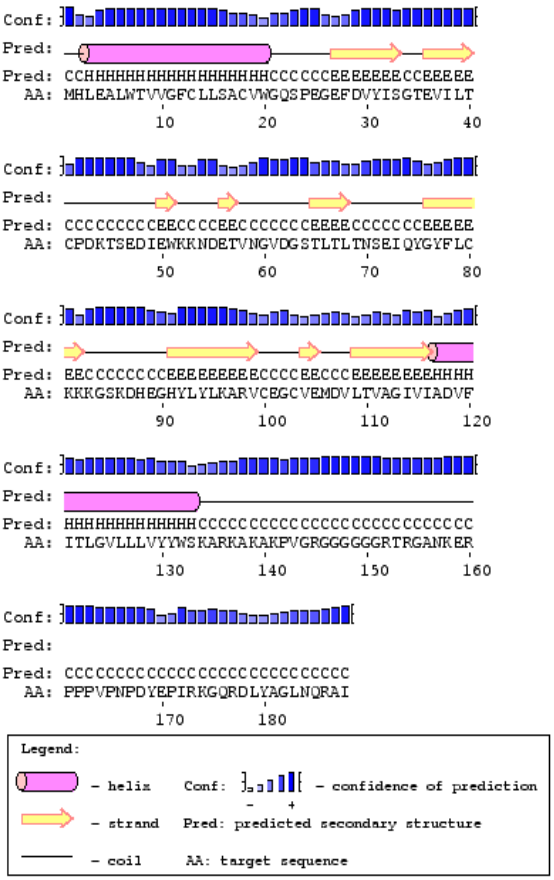


Figure 3A.3. Secondary structure prediction for *O. fraenata* CD3ε.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Lagorchestes hirsutus (Rufous hare wallaby or mala) CD3ε – nucleotide sequence - open reading frame

```
atgcatttggaagctctctggactgtggtaggattctgtcagctctcagcctgtgtctgggggcaaagcctgg
aaaccgataagaactatgaatttgaagtctccatctcaggaactgaagtaacactcacctgccccgaaaaagc
taatgaggacatagaatggaagaaaaatgatgtaaccgtaaacggtgtggacagcagtttattcaccctatca
gaccccgagactgagtataatggtcacttcttttgtaaaaagaaaggatcagatggcgaaggctattatctct
acctgaaagcaagagtatgtgaaggttgtgtggaaatggacgtgctgacggtggctggcattgtcattgctga
cgtcttcatcactctgggagtactgctgttggtgtattactggagcaaggcgcgaaaggccaaggccaagcct
gttggtcgaggggaggtggcggtggcaggacaagaggagcaaacaaggagaggcctccacctgttcccaacc
ctgactatgaccccatccgcaaaggccaacaggacctgtatgctggcctgaatcacagagccatctga
```

L. hirsutus putative protein sequence

```
MHLEALWTVVGFCQLSACVWGQSLETDKNYEFEVSISGTEVTLTCPEKANEDIEWKKNDVTVNGVDSSLFTLS
DPETEYNGHFFCKKKGSDGEGYYLYLKARVCEGCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKAKP
VGRGGGGGGRTRGANKERPPPVPNPDYDPIRKGQQDLYAGLNHRAI
```

Secondary structure prediction for *L. hirsutus* CD3ε using PSIpred

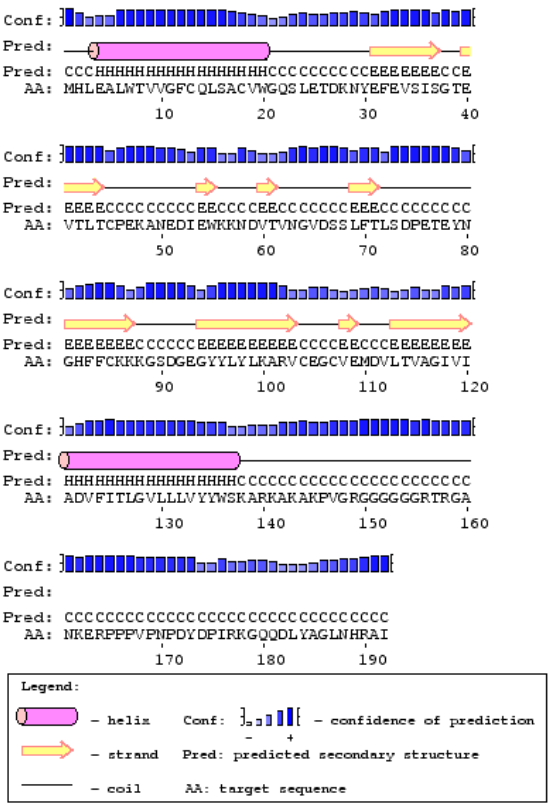


Figure 3A.4. Secondary structure prediction for *L. hirsutus* CD3ε.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 3A

Alignment between marsupial species and human

```
M. eugenii      MHLEALWTVVGFCLLSACVWGQSLES--------EFGVSISGTKVTLTCPEKSGEEIEWK  52
L. hirsutus     MHLEALWTVVGFCQLSACVWGQSLETD----KNYEFEVSISGTEVTLTCPEKANEDIEWK  56
O. fraenata     MHLEALWTVVGFCLLSACVWGQSPEG--------EFDVYISGTEVILTCPDKTSEEIEWK  52
M. domestica    MQLGSLWTVLGFFLLSACVWGEDLEEDP---QKYKEGVSISGTQVTLTCPEKSEDLIIWK  57
H. sapiens      MQSGTHWRVLGLCLLSVGVWGQDGNEEMGGITQTPYKVSISGTTVILTCPQYPGSEILWQ  60
                *.  . * *.*.  **. ***.. .          . * **** * ****. . . * *.

M. eugenii      KNDVTINGVNTNSLTLSD---------LETEYNGHFFCKKKGPEDGEG-YYLYLKAKVCE 103
L. hirsutus     KNDVTVNGVDSSLFTLSD---------PETEYNGHFFCKKKG-SDGEG-YYLYLKARVCE 107
O. fraenata     KNDETVKGVDGSTLTLTN---------SEIQYG-YFLGKKKGSKDHEG-HYLYLKAKVCE 103
M. domestica    KNN-LINGVESYQLTLDD---------SETEYSGHFHCKKKSSPSDEG-YFLYLKARVCH 107
H. sapiens      HNDKNIGGDEDDKNIGSDEDHLSLKEFSELEQSGYYVCYPRGSKPEDANFYLYLRARVCE 120
                .*.  . * .       .          * . . ..    ..    .. ..***.*.**.

M. eugenii      GCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKAKPVGRGGGGGGRTRGANKERP 163
L. hirsutus     GCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKAKPVGRGGGGGGRTRGANKERP 167
O. fraenata     GCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKAKPVGRGGGGGGRTRGANKERP 163
M. domestica    GCLEMGVLTVAGIIIADVFITLGVLILVYHWSKKQKAKSKPVR-GGGAGGKTRGVNKERP 167
H. sapiens      NCMEMDVMSVATIVIVDICITGGLLLLVYYWSKNRKAKAKPVTRGAGAGGRQRGQNKERP 180
                .*.**.*..** *.*.*. ** *.*.***.*** .***.***  *.*.**. ** *****

M. eugenii      PPVPNPDYEPIRKGQRDLYAGLNQRAI 190
L. hirsutus     PPVPNPDYDPIRKGQQDLYAGLNHRAI 194
O. fraenata     PPVPNPDYEPIRKGQRDLYAGLNQRAI 190
M. domestica    PPVPNPDYEPIRKGQRELYAGLNQRAI 194
H. sapiens      PPVPNPDYEPIRKGQRDLYSGLNQRRI 207
                ********.******..**.***.* *
```

Figure 3A.5. Alignment of marsupial and CD3ε and human.
Grey area = exon boundary does not confirm prediction in ensembl. Bold Underlined = the exon boundaries as predicted in ensembl. Green highlight shows the gap in the marsupial sequence compared to the human sequence.

M. domestica CD3ε in Spleen and Thymus cDNA library

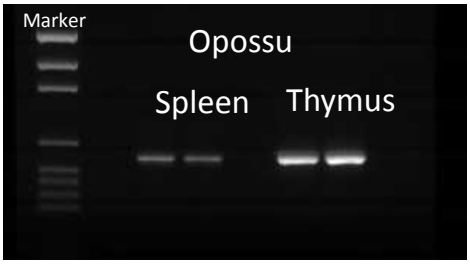


Figure 3A.6. Gel image of CD3ε in *M. domestica.*
Spleen and thymus cDNA library using UTR primers.

L. hirsutus CD3ε

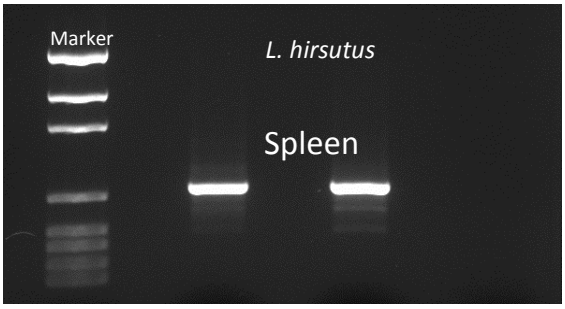


Figure 3A.7. Gel Image of *L. hirsutus* CD3ε in spleen
using UTR primers. Product in 1.5% agarose Gel matrix.

Appendix 3A

Alignment for CD3ε all species using CLUSTALW2 (blosum 62 matrix) (putative protein)

```
H.hippoglossus  -----MKIHSMGVRAVIAMTLLLSLAASEEETAK------------GSVKFKGGNFTMR 42
P.olivaceus     -----MKINTMDVRAVIAMTLLLYVAADDSE----------------PVTFEGEYFTMR 38
T.rubripes      ----------MIVRALLCFAVLFLVAGQEEEADK------------PGGVIFWRTSVTMT 38
S.salar         -------MNRDGVYGGLVFLLLIMTSVEGGG-----------------DVSFWRTTVTLT 36
O.mykiss        -------MDRDGVYGGLVFLLLIMTTVEGNG-----------------YVSFWRTKVTMT 36
A.ruthenus      MKPTMKKAASLLALFIAVCSDLVSSEGGDEDSTA------------QGSVDVSGTSVTLT 48
A.mexicanum     ------MARDAQVVATLLSVVLYGVSTLANTPGN------------EIAVRISGTSVFLT 42
A.platyrhynchos -------MRFELSLPFLGLLLCVGGTAAQDVADEELT-----ASPEELQVEISGTTVKVR 48
G.gallus        -------MRCEVPLPLLGLLLCVVGAAAQ-------------GGQEEFAVEISGTTVTIT 40
P.troglodytes   -------MQSGTHWRVLGLCLLSVGVWGQDGEINFPS-----FSP-AYKVSISGTTVILT 47
M.mulatta       -------MQSGTHWRVLGLCLLSVGVWGQDGEINFPS-----FSP-AYKVSISGTTVILT 47
H.sapiens       -------MQSGTHWRVLGLCLLSVGVWGQDGNEEMGG-----ITQTPYKVSISGTTVILT 48
M.fascicularis  -------MQSGTRWRVLGLCLLSIGVWGQDGNEEMGS-----ITQTPYQVSISGTTVILT 48
C.jacchus       -------MQSGTRWRVLGLCLLSVGIWGQDGNEEMGD-----TTQNPYKVSISGTTVTLT 48
B.taurus        -------MQSGNLWRALGLCLLLVGAWAQDADE-----------QKPYEVSISGNTVELT 42
O.aries         -------MQTGNLWQVLGLCLLLVGAWAQDDTE-----------QNPYEVSISGNSVELT 42
S.scrofa        -------MPSGNLWKVLGLCLLSVGAWGQEDIERPDE-----DTQKTFKVSISGDKVELT 48
C.familiaris    -------MQSRNLWRILGLCLLSVGAWGQDEDFKASDDLTSISPEKRFKVSISGTEVVVT 53
M.furo          -------MQPGNLWRILGLCLLLVGAWGQEE--------------GYKVSISGTMVVLT 38
F.catus         -------MPSGSLWRVLGLCLLSVGAWGQEDNEDPLEPSPQTSASARYKVSISGTTVVLT 53
M.musculus      -------MRWNTFWGILCLSLLAVGTCQDDAEN-----------IEYKVSISGTSVELT 41
R.norvegicus    -------MQWNAFWSILGLSLLAVGTCQE----------------EYEVSISGTSVELT 36
O.cuniculus     ----MSRMRAGTLWRVLALWLLSVAAWGQEDDDHADD-----YTQKLFTVSISGTRVVLT 51
M.eugenii       -------MHLEALWTVVGFCLLSACVWGQSLES-----------EFGVSISGTKVTLT 40
L.hirsutus      -------MHLEALWTVVGFCQLSACVWGQSLETD-KN--------YEFEVSISGTEVTLT 44
O.fraenata      -------MHLEALWTVVGFCLLSACVWGQSPEG-----------EFDVYISGTEVILT 40
M.domestica     -------MQLGSLWTVLGFFLLSACVWGEDLEEDPQK--------YKFGVSISGTQVTLT 45
T.guttata       -------MPGGKALSAWALLASLAMASLGVRG-------------QIYVKEFSGKVFLE 39
                                                             *      . .

H.hippoglossus  CPVYGKWYR-----------RSAQVGDGE---------DSLELNYNSDTKGLYKCVYKDN 82
P.olivaceus     CPGEQKLLK-----------DNSDANN-------------TVQYHDQTKGLYRC-----  68
T.rubripes      CPQDGKWFK-----------DKTQVSQGS---------E-YAFEYDNK-KGRYHCTY---  73
S.salar         CPDKGDWYD-----------NTIKMNEEE--------SKEIKMDYDESKKNVYQCKYLYD 77
O.mykiss        CPGKGEWYE-----------GTNNLDKNSR-------SEQIEENYDESKKRVYHCEYQYD 78
A.ruthenus      CPLTGTVSDPGSTTWQYKEEEKKIPDTDGK-------TQITLQTYNSTNNGLYKCSN---  98
A.mexicanum     CPSRDSPSRDTN-----STLSGPNVLSTNT-------MDHHLKNYDEGMNGEYHCQVLHG 90
A.platyrhynchos CSLS-ET----------VQWKPYSPDTED-----------TTFIKTN----HDSSPLNLT 82
G.gallus        CPSSGDD----------IKWKPDPALGDN----------NKYIIQN----HDSSPLTVS 75
P.troglodytes   CPQYPGS-E--------ILWQHNDKNIGSDEDDKNIGSDEDHLSLKEFSELEQSGYYVCY 98
M.mulatta       CPQYPGS-E--------ILWQHNDKNIGSDEDDKNIGSDEDHLSLKEFSELEQSGYYVCY 98
H.sapiens       CPQYPGS-E--------ILWQHNDKNIGGDEDDKNIGSDEDHLSLKEFSELEQSGYYVCY 99
M.fascicularis  CSQHLGS-E--------AQWQHNGKNKG---------DSGDQLFLPEFSEMEQSGYYVCY 90
C.jacchus       CPRYDGH-E--------IKWLVNSQNKE---------GHEDHLLLEDFSEMEQSGYYACL 90
B.taurus        CPREFEG-E--------IHWKQNDEQMKG--------YTGKQLLLENFSEMDNSGYYQCY 85
O.aries         CPKDFEN-G--------IQWKRNNEQMKG--------HNEKYLLLDQFSEMESSGYYQCL 85
S.scrofa        CPEDPESEK--------MTWKRNDMQIYE--------SYDNYMLLESFSEVENSGYYTCT 92
C.familiaris    CPDVFGYDN--------IKWEKNDNLVEG--------ASNRELSQKEFSEVDDSGYYACY 97
M.furo          CPEEVDTNS--------ITWEKNGEAIKG--------ANERQYTMHTFSEVEDSGSYTCF 82
F.catus         CPEDLGSES--------IKWERNGDLLPN--------EYGEQLFLDDFSEMENSGYYACY 97
M.musculus      CPLDSDE-N--------LKWEKNGQELPQ--------KHDKHLVLQDFSEVEDSGYYVCY 84
R.norvegicus    CPLENED-N--------LKWEKNDKVLPD--------KNEKHLVLEDFSEVKDSGYYVCY 79
O.cuniculus     CPVEAEGGD--------IHWERDEKSLPN---------TKKELDLTDFSEMEHSGYYSCY 94
M.eugenii       CPEKSGE-E--------IEWKKNDVTING--------VNTNSLTLSD-LETEYNGHFFCK 82
L.hirsutus      CPEKANE-D--------IEWKKNDVTVNG--------VDSSLFTLSD-PETEYNGHFFCK 86
O.fraenata      CPDKTSE-D--------IEWKKNDETVNG--------VDGSTLTLTN-SEIQY-GYFLCK 81
M.domestica     CPEKSED-L--------IIWKKNN-LING--------VESYQLTLDD-SETEYSGHFHCK 86
T.guttata       CVRDQGSKN--------ITWWRDGSTVGH----------EAQLDLNRVYD-DPRGLFVCE 80
                *

H.hippoglossus  -----EEDKEYYFYVKGKACANCIELEAAPLALAITVDMAGTIILMMIIYKCTKKKSSA- 136
P.olivaceus     ------EKGKYFFYVKANLCLNCFNLDGALFAQVIAVDMTGTIILMIVIYRCTKKRS-A- 120
T.rubripes      ------ENTRYDFYVQGKVCENCFELDGGFWGMIIAADMFLTVVVMVMVYKCAKKRS--- 124
S.salar         --QYDTEKTTYQFYFKGKVCKDCYELNPTVVAGAIIGDLLVTGGVILIVYLRARKKS--- 132
O.mykiss        PQDYPEKTAIYQFYFKGKVCKDCYELNPTLVAGAIIGDLLVTGGVILIVYLRARKKS--- 135
A.ruthenus      ------GNDYYHFYVKVKVCESCVEVEVVPMIGIIFADLLVTAGVAILVYYWAQNRK-G- 150
A.mexicanum     -----TKPNVSRLFLKVKVCHHCVDLGFWTVAGILVTDILVTVGVSILVYYWSKGRKRIP 145
A.platyrhynchos CTADNK---NIHMYLKARVCTNCMELDTLTVTGIIIADLLITFGLLILVYYFSKDKKG-- 137
```

```
G.gallus          CTAGDQ---EHTMYLNAKVCANCEELDTFTVVGIIAADLLITLGVLILVYYFSKNKKG-- 130
P.troglodytes     PRGSKPEDANFYLYLRARVCENCMEMDVMSVATIVIVDICITGGLLLLVYYWSKNRKAKA 158
M.mulatta         PRGSKPEDANFYLYLRARVCENCMEMDVMSVATIVIVDICITGGLLLLVYYWSKNRKAKA 158
H.sapiens         PRGSKPEDANFYLYLRARVCENCMEMDVMSVATIVIVDICITGGLLLLVYYWSKNRKAKA 159
M.fascicularis    PRGSNPEDASHHLYLKARVCENCMEMDVMAVATIVIVDICITLGLLLLVYYWSKNRKAKA 150
C.jacchus         SKETPAEEASHYLYLKARVCENCVEVDVMAVATIVIVDICITLGLLLLVYYWSKNRKAKA 150
B.taurus          MTEGNKE-AAHTLYLKARVCQNCMEVNLMEVATIIVVDICVTLGLLLLVYYWSKSRKAKA 144
O.aries           ATEGNTE-AAHTLYLKARVCKNCMEVNLLEVATIIVVDICVTLGLLLLVYYWSKSRKAKA 144
S.scrofa          VGEKTS----HRLYLKARVCENCVEVDLMAVVTIIVVDICITLGLLMVVYYYSKSRKAKA 148
C.familiaris      ADSIKE---KSYLYLRARVCANCIEVNLMAVVTIIVADICLTLGLLLMVYYWSKTRKANA 154
M.furo            SDKLK----KNSLYLKARVCKNCIEVSPMAVAAILVTDICITLGLLLLVYYWSKNRKANA 138
F.catus           TSNSLE---KNYLYLKARVCQNCVEVDTMTAVAIVVADVCITLGLLLLVYYWSKNKKASS 154
M.musculus        TPASNK---NTYLYLKARVCEYCVEVDLTAVAIIIIVDICITLGLLMVIYYWSKNRKAKA 141
R.norvegicus      TESSRK---NTYLYLKARVCENCMEVDLTAVSIIIIVDICITLGLLMVVYYWSKKRKAKA 136
O.cuniculus       VGTKNKE-NEHILYLKARVCEACMEVDLTTVASIVVADVCVTLGLLLLVYYWSKNRKAKC 153
M.eugenii         -KKGPEDGEGYYLYLKAKVCEGCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKA 141
L.hirsutus        -KKGS-DGEGYYLYLKARVCEGCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKA 144
O.fraenata        -KKGSKDHEGHYLYLKARVCEGCVEMDVLTVAGIVIADVFITLGVLLLVYYWSKARKAKA 140
M.domestica       -KKSSPSDEGYFLYLKARVCHGCLEMGVLTVAGIIIADVFITLGVLILVYHWSKKQKAKS 145
T.guttata         TGSK-----RSSLQVHYRMCQNCIEVDAPTVSGIVIADVVATLFLAVAVYCITGHNR--- 132
                             . .. . *  * ..        .  *.  *  . . .*  .  .

H.hippoglossus    ---GSTQASKAPARA-----GGRAPP-VPSPDYEALNPHTRSQDPYAIVS----RTG 180
P.olivaceus       ---GSTNTSKAPARA-----VGRAPP-VPSPDYEPLNPHTRAQDPYSIVN----RTG 164
T.rubripes        ----SAALPRVP-KA-----GGRAPP-LPSPDYEPLNPHTRSQGTYSEVHP--KRMG 168
S.salar           ---GPAAPQKPTSRS-----AGRGPPVVPSPDYEPLSVATRSSDIYATTQTSTQRTG 181
O.mykiss          ---GPAAPQKPTSRS-----AGRGPPVVPSPDYEPLSLATRSRDIYAT-----HRTG 179
A.ruthenus        ---ASAMAPAARPGR-----QNRAPP-VPNPDYEPIRTGNR--EVYSGLN---KRT- 192
A.mexicanum       APGGASAGGRRPRDY-----NKERPPPVPNPDYEPIRKGQR--EVYDGLKP---QY- 191
A.platyrhynchos   ---RPSAGAGSRPRG----QKTQRPPPVPNPDYEPIRKGQR--EVYAGLES---RGY 182
G.gallus          ---QSRAAAGSRPRA----QKMQRPPPVPNPDYEPIRKGQR--DVYAGLEH---RGF 175
P.troglodytes     KPVTRGAGAGGRQRG----QNKEKPPPVPNPDYEPIRKGQR--DLYSGLNQ---RRI 206
M.mulatta         KPVTRGAGAGGRQRG----QNKEKPPPVPNPDYEPIRKGQR--DLYSGLNQ---RRI 206
H.sapiens         KPVTRGAGAGGRQRG----QNKERPPPVPNPDYEPIRKGQR--DLYSGLNQ---RRI 207
M.fascicularis    KPVTRGAGAGGRQRG----QNKERPPPVPNPDYEPIRKGQQ--DLYSGLNQ---RRI 198
C.jacchus         KPVTRGVGAGGRQRG----QNKERPPPVPNPDYEPIRKGQR--DLYSGLNQ---RGI 198
B.taurus          SPMTRGAGAGGRPRG----QNKGRPPPVPNPDYEPIRKGQR--DLYAGLNQ---RGV 192
O.aries           TPMTRGAGAGGRPRG----QNRERPPPVPNPDYEPIRKGQR--DLYSGLNQ---RGV 192
S.scrofa          MPVTRGAGAGGRPRG----QNRERPPPVPNPDYEPIRKGQR--DLYSGLNQ---RGR 196
C.familiaris      KPVMRGTGAGSRPRG----QNKEKPPPVPNPDYEPIRKGQQ--DLYSGLNQ---RGI 202
M.furo            TTVMRAKGAGGRTRG----QNKEKPPPVPNPDYEPIRKGQQ--DLYSGLNQ---RGI 186
F.catus           VTMMRGPGAGGRPRG----QNKEKPPPVPNPDYEPIRKGQQ--DLYSGLNQ---RGI 202
M.musculus        KPVTRGTGAGSRPRG----QNKERPPPVPNPDYEPIRKGQR--DLYSGLNQ---RAV 189
R.norvegicus      KPVTRGTGTGGRPRGKAQGQNKERPPPVPNPDYEPIRKGQR--DLYSGLNQ---RAV 188
O.cuniculus       KPVTRGAGAGGRPRG----QNKERPPPVPNPDYEPIRKGQR--DLYSGLNQ---RGI 201
M.eugenii         KPVGRGGGGGGRTRG----ANKERPPPVPNPDYEPIRKGQR--DLYAGLNQ---RAI 189
L.hirsutus        KPVGRGGGGGGRTRG----ANKERPPPVPNPDYDPIRKGQQ--DLYAGLNH---RAI 192
O.fraenata        KPVGRGGGGGGRTRG----ANKERPPPVPNPDYEPIRKGQR--DLYAGLNQ---RAI 188
M.domestica       KPV-RGGGAGGKTRG----VNKERPPPVPNPDYEPIRKGQR--ELYAGLNQ---RAI 192
T.guttata         -------GHTSRASD--------RQNLIANELYQPLGERDD--EQYSRLAPARARK- 171
                                        ...  *...         *        .
```

Figure 3A.8. Alignment of all CD3ε sequences found in Genbank. Underlined is the ARAM site (antigen recognition activation motif) and the endoplasmic retention signal is marked in red.

Appendix 3A

Genbank Accession Numbers for CD3 epsilon

Table 3A.1. Accession numbers for CD3ε sequences found in Genbank and the relevant references

Species	Common Name	Accession Number	References
Acipenser ruthenus	Sterlet	AJ242941	unpublished
Ambystoma mexicanum	Axolotl	AY212509	unpublished
Anas platyrhynchos	Duck	AF378704	Direct submission
Aotus nancymaae	Ma's night monkey	EF547186	unpublished
Bos taurus	Cattle	U25687	(Hagens *et al.,* 1996)
Cairina moschata	Muscovy duck	AY738734	(Kothlow *et al.,* 2005)
Callithrix jacchus	Common marmoset	DQ189218	unpublished
Canis lupus familiaris	Dog	M55410	(Nash *et al.,* 1991)
Felis catus	Cat	AB195839	(Nishimura *et al.,* 1998)
Gallus Gallus	Chicken	NM_206904	(Gobel and Fluri, 1997)
Hippoglossus hippoglossus	Atlantic halibut	FJ769816	(Overgard *et al.,* 2009)
Homo sapiens	Human	NM_000733	(Gold *et al.,* 1986)
Macaca fascicularis	Crab eating macaque	AB073994	(Uda *et al.,* 2001)
Macaca mulatta	Rhesus monkey	XM_001097204	Annotated
Macropus eugenii	Tammar wallaby	AY028923	(Old *et al.,* 2001)
Marmota monax	Woodchuck	AF232727	Direct submission
Monodelphis domestica	Grey short tailed opossum	XM_001380690	Annotated
Mus musculus	Mouse	NM_007648	(de la Cruz *et al.,* 2011)
Mustela putorius furo	Domestic ferret	EF492054.1	unpublished
Oncorhynchus mykiss	Rainbow trout	GU074379	Direct submission
Oryctolagus cuniculus	Rabbit	NM_001082001	Direct submission
Ovis aries	Sheep	NM_001009418	(Hein and Tunnacliffe, 1993)
Pan trogldytes	Chimpanzee	XM_001160645	Annotated
Paralichthys olivaceus	Japanese flounder	AB081751	Direct submission
Rattus norvegicus	Rat	NM_001108140	Direct submission
Salmo salar	Atlantic salmon	EF421421	(Liu *et al.,* 2008)
Sus scrofa	Pig	AY323829	Direct submision
Taeniopygia guttata	Zebra finch	XM_002189574	Annotated
Takifugu rubripes	Pufferfish	AB166799	(Araki *et al.,* 2005)

Appendix 3B

Macropus eugenii T cell receptor α-chain (TCRα) partial nucleotide sequence

```
tgcctcttcaccgactttgactcttccattacaaatacaagtggtaccaacccaacggtactggaaatgatgt
cgatggactctaagagctatggatcactgcactggggtcacaaagaaaacttcgattgctcaaaggcattcaa
accagacatcaataatttgaagaccagtataaaggtgccacatgcaaagtccaagatgtacagcaaagcttt
gaaacagacaaagatttgaacttgatgaacatatctctgatttttcttcgtgttatcttcttgaagactgtgg
gat
```

Partial putative amino acid sequence of *M. eugenii* TCRα chain

```
CLFTDFDSSITNTSGTNPTVLEMMSMDSKSYGSLHWGHKENFDCSKAFKPDINNFEDQYKGATCKVQDVQQSF
ETDKDLNLMNISLIFLRVIFLKTVG
```

Secondary structure prediction using PSIpred for M. eugenii TCRα

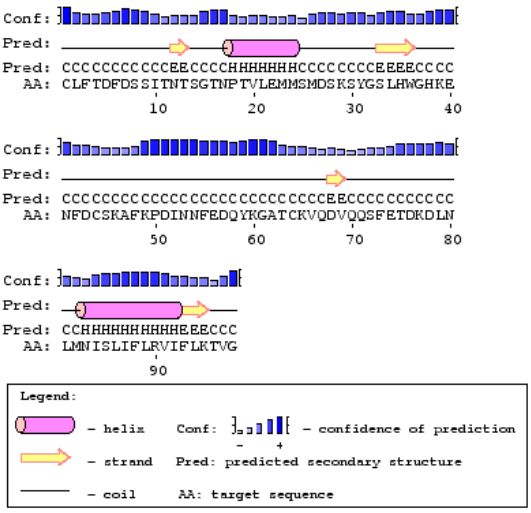


Figure 3B.a. Secondary structure prediction for *M. eugenii* TCRα.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 3B

Onychogalea fraenata partial TCRα chain partial nucleotide sequence

```
agcaacaccctggtgtgcctcttcacagattttgactcttccattacaaatacaaatggtaccaacccaacag
tactggaaatgatatcgatggaatctaagagctatggatcagtgtactggggtcacaaagaaaactccagttg
cacagatgcattcagtccaaacatcatcggtcctttggctgacccctcagatgccacatgcaaagtccaagat
gtacagcaaagctttgaaacagacaaagatttgaacttgatgaacatatctctgatttttcttcgtgtcatct
tcttgaagactgtgga
```

Partial putative amino acid sequence of *O. fraenata* TCRα chain

```
SNTLVCLFTDFDSSITNTNGTNPTVLEMISMESKSYGSVYWGHKENSSCTDAFSPNIIGPLADPSDATCKVQD
VQQSFETDKDLNLMNISLIFLRVIFLKTV
```

Secondary structure prediction using PSIpred for *O. fraenata* TCRα

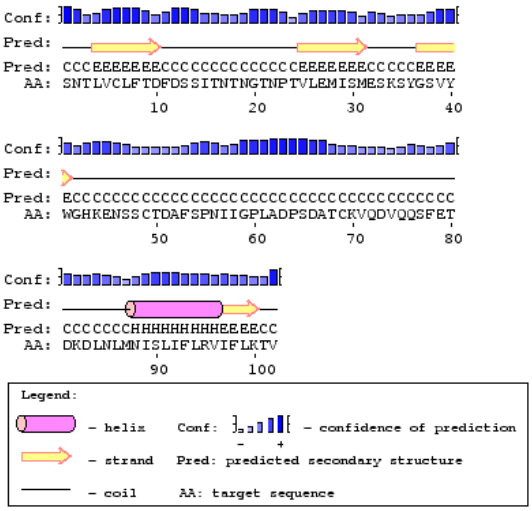


Figure 3B.b. Secondary structure prediction for *O. fraenata* TCRα.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Amino acid alignment for TCRα using CLUSTALW2 (blosum62 matrix)

```
M.eugenii         ------------------------------------------------------------
O.fraenata        ------------------------------------------------------------
T.vulpecula       ----------------------------------------CTYT-SSVN-SLQWYRH 15
M.domestica       --MNSALTLMIWILLLFGDTYGDSVTQTEGRIILTEGASLTLNCSYQ-TSGSPFLSWYIQ 57
H.sapiens         -MSLSSLLKVVTASLWLGPGIAQKITQTQPGMFVQEKEAVTLDCTYDTSDQSYGLFWYKQ 59
B.taurus          ----MRLVTVVTVFLTLGTVVDAKTTQPNS-MDCAEGENVNLPCNHSIIRGEDYIHWYRQ 55
O.anatinus        --------------------MP------YSRVGTS-GGGS-----------------YNL 16
T.aculeatus       ----MARGVTATLLLLLSPVLPNQVTQLTSVEATA-GHAVNLQCKHTSLTSS-PIFWYQQ 54
M.musculus        MKKRLSACWVVLWLHYQWVAGKTQVEQSPQSLVVRQGENCVLQCNYSVTPDN-HLRWFKQ 59
R.norvegicus      MKAPIHTVFLFSWLWLDWESHGEKVEQVLSTLSVQEGDTAVTNCTYTDSASS-YFPWYKQ 59

M.eugenii         ------------------------------------------------------------
O.fraenata        ------------------------------------------------------------
T.vulpecula       HPGTGPTFLF-AMFSDGDEKQQGRFKATLNTKSRHSSLSISATQLSDSATYFCAV--NTG 72
M.domestica       HPNEGLKLLVNEAKRKDQEKDNNGFWTKKIKEKSFFSLEKTSVQVKDSAVYYCVL--SKG 115
H.sapiens         PSSGEMIFLIYQGSYDEQNATEGRYSLNFQKARKSANLVISASQLGDSAMYFCAMR-EYP 118
B.taurus          NPSQSPQYVIHGLRGTVNNSMA---SLHIASDRKSSTLVLPQVTLRDAAVYYCTP--SSS 110
O.anatinus        LFGKG------------------TKVTVVP---------------------------- 28
T.aculeatus       MFNQAPELVLSGYSST----RSAKAKLTIQEDRKSNILTLHNVQDRDSAVYYCALD-GYD 109
M.musculus        DTGKGLVSLTVLVDQK-DKTSNGRYSATLDKDAKHSTLHITATLLDDTATYICVVADRGS 118
R.norvegicus      GAGKGLHFVID-IRSNVDRKETQKFTVLMDKKAKKFSLHITATQAEDSAIYFCAKT---- 114

M.eugenii         -----------------NIKNPE-PAVYQLKSPKSSNTSVCLFTDFNSTI---------- 32
O.fraenata        ----------------------------------SNTLVCLFTDFDSSI---------- 15
T.vulpecula       AGNKLIFGIGSSLKIKPNIKNPE-PALYQLTSPKSSDTSVCLFTDFDS------------ 119
M.domestica       ETSQHIFGKGTQVAVLPNIQNPQ-PALYQLRSPKSSNTSVCLLTDFGFYNGSIKN----- 169
H.sapiens         SYDKVIFGPGTSLSVIPNIQNPD-PAVYQLRDSKSSDKSVCLFTDFDS-QTNVSQ---SK 173
B.taurus          SGWQLTFGSGTQLTVVPEVKDPN-PTVYQLRSPQSSDTSVCLFTDFDSNQVNMEKIMGSE 169
O.anatinus        SLNKQIFGTGTKLTVQPNVTNPQ-PRMYHLKKPRVNDLSVCLFTDFGNEEVNMMG----- 82
T.aculeatus       SGNKVIFGKGTSLTVTPNVTNPQ-PRMYRLKKPQVNDLSICLFTDFGNDEVNMTG----- 163
M.musculus        ALGRLHFGAGTQLIVIPDIQNPE-PAVYQLKDPRSQDSTLCLFTDFDSQINVPKT----M 173
R.norvegicus      --LKMDSSPGFVAVILLLLGRTHGDSVTQTEGQVTISENGFLRINCTY------------ 160
                                                     *  .

M.eugenii         -TNTSGTNPTVLEMMSMDSKSYGSLHWGHKENFDCSKAFKPDINNF--EDQYKGATCKVQ 89
O.fraenata        -TNTNGTNPTVLEMISMESKSYGSVYWGHKENSSCTDAFSPNIIGP--LADPSDATCKVQ 72
T.vulpecula       -SDPSGSNATVLEMMTMESKSYGAVTWGSKSSFNCTNSFQQSDVT---LTLPSGSTCKVK 175
M.domestica       -ETVTGSEATVLEMMTMESKSYGAVTWGSKSNFTCTDAFRKDMFD---FNQFSGSKCNSS 225
H.sapiens         DSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPEDTFFPSPESSCDVK 233
B.taurus          GSTVHKTNSTVLNMEILGSKSNGIVTWGNTSDAGCEYTFNETIP----FASSLEISCNAK 225
O.anatinus        -INNIKRTPSMVAEKRLASKSLGIVAWNNNLDWKCQAKISNITYS---LSNSSGKVCNTT 138
T.aculeatus       -IRNIMRAPSVVDVKRLESKSLGIVAWDNSLDWDCQAQASEAVYS---LSNSSGKVCNAK 219
M.musculus        ESGTFITDKTVLDMKAMDSKSNGAIAWSNQTSFTCQDIFKETNAT----YPSSDVPCDAT 229
R.norvegicus      -SATSIAYPTLFWYVQYPGEGLQLLLKVFTAGQKGSSRGFEATYN-------KETTSFHL 212
                        ...       ..   .       .                        .

M.eugenii         DVQQSFETDKDLNLMNISLIFLRVIFLKTVEFNVLMTLRLWSN- 132
O.fraenata        DVQQSFETDKDLNLMNISLIFLRVIFLKTV-------------- 102
T.vulpecula       DVQQSFETDKDLNLMNMSLVFLRIIFLKTVGFNLFMTLRLWSN- 218
M.domestica       HAEQGFETDRDLNLMNLSLIVLRIIFLKTVGFNLLMTLRLWSN- 268
H.sapiens         LVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWSS- 276
B.taurus          LVEKSFETDINLNSQNLSVIVFRILLLKVVGFNLLMTLRLWSS- 268
O.anatinus        AVTENFSSDPYLNSINLAMIFLRVIFVKTVGFNLLMTLKLWSS- 181
T.aculeatus       VVNENFSSDPYLNSRNLAMIFLRVVFVKTMGFNLLMTLKLWSS- 262
M.musculus        LTEKSFETDMNLNFQNLSVMGLRILLLKVAGFNLLMTLRLWSS- 272
R.norvegicus      QKASVQESDSAVYYCALG----DTVVETTGGAEHKPRGNRWSGC 252
```

Fig.3Bc. Alignment of TCRα sequences.
Underlined are putative glycosylation sites. Boxed the core-pepitde. Green the connecting peptide.

Appendix 3B

Genbank Accession Numbers

Table 3B.1. Genbank Accession Numbers for TCRα found in Genbank and the relevnat references.

Species Name	Common Name	Accession Number	References
Bos taurus	Cattle	BC102771,(D10394 D90030	(Ishiguro *et al.,* 1990)
Callithrix jacchus	White-tuffed-ear marmoset	AB504389 (partial sequence)	(Fujii *et al.,* 2010)
Homo sapiens	Human	EF101779, M12959, X02883	(Wang *et al.,* 2008, Rabbitts *et al.,* 1985, Yoshikai *et al.,* 1985)
Macaca mulatta	Rhesus monkey	HQ622180	Unpublished
Monodelphis domestica	South American gray-short tailed opossum	AY014504	(Baker *et al.,* 2001)
Mus msuculus	Mouse	DQ340292, U46581	(Pyz *et al.,* 2006)
Ornithorhynchus anatinus	Platypus	XM_001507749	Annotation
Oryctolagus cuniculus	Rabbit	M12885	(Marche and Kindt, 1986)
Rattus norvegicus	Rat	M18853	Annotated
Tachyglossus aculeatus	Australian Echidna	AY423736	(Belov et al., 2004)
Trichosurus vulpecula	Australian silver-grey brushtail possum	AF133097	(Zuccolotto *et al.,* 2000)
Ictalurus punctatus	Catfish	U39194	(Wilson *et al.,* 1998)
Gallus gallus	Chicken	U04611	(Goebel and Dangy, 2000)
Ovis aries	Sheep	M55622	(Hein *et al.,* 1991)

Appendix 3C

M. eugenii TCR beta (TCRβ) partial nucleotide sequence

cggaccgggcaccaggctcactgtaacagatgacctgaccagggtgactcccccaaaggtgactgtgtttcag
ccatctgaggaagagatggcgaataagggaaaggccacattggtctgtctggccacaggcttctaccctgacc
ttgtggagctgaagtggtgggtgaatgggcaggagacccaaattggggtcagcacagaccctcagccctccaa
agagcagccccataacaacttctccacatactccctgagcagtcgtcttcgggtgtctgctcccttctggcgc
aatcccaagaacagcttccggtgccaagtattgttccatgggattggagagaatgagacctggacaagtaacc
tgaagaaacccatcacccagaatgtcagtgaccagatctggggaaaagcagattgtggggtttcctctgaatc
ctatcaacatagtatccagtctgccaccttcttgtatgagatcctgctggggaaagccgtgctctatggcctg
ctggtcagtgctctggtgtggagaaccatggccaagaaaaaacattcctga
3′end
aggcatctgtggaggtgagagcagggaagacagagcaagagaaccctatcccctgttttgcctactgtgttct
ctattcctgccctttgttccctaaaaac**tataat**atttgctcttcctttctttggttcctcatacctcctaa
atttcaactctgaaaatcatgtaggaccaggaggatggcttaatcctattaatcttaaaacatcttgcccgtc
ctggtgtggtctcactcttaaaaaaaaaaaaaaaaaa

Partial putative protein sequence TCRβ *M. eugenii*

GPGTRLTVTDDLTRVTPPKVTVFQPSEEEMANKGKATLVCLATGFYPDLVELKWWVNGQETQIGVSTDPQPSK
EQPHNNFSTYSLSSRLRVSAPFWRNPKNSFRCQVLFHGIGENETWTSNLKKPITQNVSDQIWGKADCGVSSES
YQHSIQSATFLYEILLGKAVLYGLLVSALVWRTMAKKKHS

Secondary structure prediction for *M. eugenii* TCRβ chain

Figure 3C.a. *M. eugenii* TCRβ secondary structure prediction.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 3C

O. fraenata TCRβ partial nucleotide sequence

atgtataagactgtaactaactatagtgagcttcattttggacctggcacaaggctaagtgtcgtagatgacc
tgaccagggcgactccccaaaggtgactgtgtttcagccatctgaggaagagatggcgaataagggaaaggc
cacactggtctgtctggccacaggcttctaccctgacctcgtggagctgaagtggtgggtgaatgggcaggaa
acccaagttggggtcagcacagaccctcagccctccaaggagcagccccataaaaatttctccagatactccc
tgagcagtcgtcttcgggtgtctgctcccttctggcgcaatcccaagaacagcttccggtgccaagtattgtt
ccatgggattggagagaatgagacctggacaagtaacctgaccaaacccatcacccggaatgtcagtgaccag
atctgggaaaaggcagattgtggaagtatccagtttgccaccttgttttatgagatttcctggggaaagcca
tgctgtatggcctgctggtcagtgctttggtgtggagaaccatggccaagagaaaacattcctga
3'end
ggccacctatggaggtaacaacagagagtacagagcaggagaatccagcaccctgttttggctactgctttct
ctattcctatcctatgttccagaagaactattatttttcagttcccatcttcatgtttcatagatctcctta
aaaagaactttgaagctcacagaggacaaggatggtggcttaatcctgaaaaaaaaaaaaaaaaaaaaaaa

Partial putative protein sequence for *O. fraenata* TCRβ

MYKTVTNYSELHFGPGTRLSVVDDLTRATPPKVTVFQPSEEEMANKGKATLVCLATGFYPDLVELKWWVNGQE
TQVGVSTDPQPSKEQPHKNFSRYSLSSRLRVSAPFWRNPKNSFRCQVLFHGIGENETWTSNLTKPITRNVSDQ
IWEKADCGSIQFATLFYEILLGKAMLYGLLVSALVWRTMAKRKHS

Secondary structure prediction using PSIpred for O. fraenata TCRβ

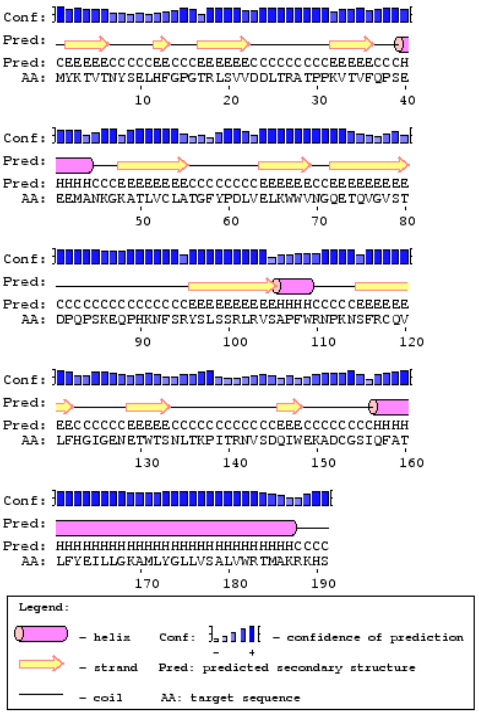


Figure 3C.b. *O. fraenata* TCRβ secondary structure prediction.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Elucidation of TCRβ chain

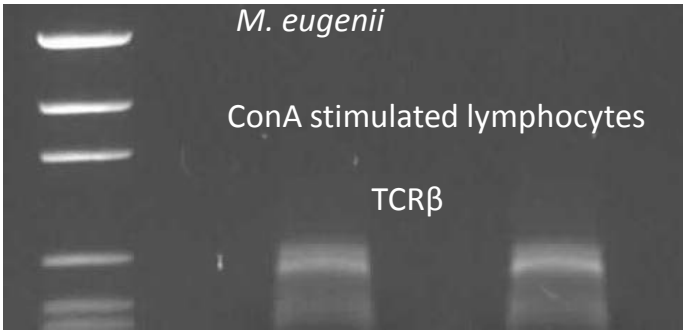


Figure 3C.c. Gel image (1.5%) of TCRβ in *M.eugenii* in ConA stimulated lymphocytes.

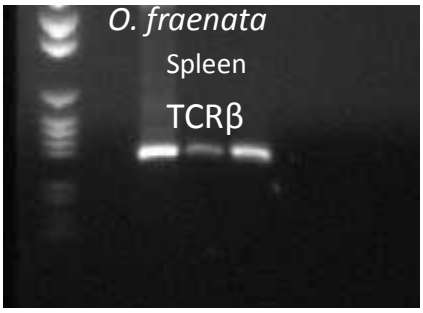


Figure 3C.d. Gel image (1.5%) of TCRβ in *O. fraenata* spleen.

Appendix 3C

Alignment of TCRβ chain sequences using CLUSTALW2 (blosum62 matrix)

```
A.platyrhynchos ------------MGMWTAWCVATFFFGARAKITQTSSLVLKEDGEATLKCSQNDNHN-Y 46
G.gallus        ---------------MWTIWCMVLYFFGARAEINQPSILVLKEDENATLSCSQNDDHN-Y 44
M.musculus      ---------MGS-RLFFVLSSLLCSKHMEAAVTQSPRNKVAVTGGKVTLSCNQTNNHNNM 50
R.norvegicus    ---------MGS-RFLLVVLSFLCAKHMEAAVTQSPRNKVTLKGGKVTLSCKQNNNHNNM 50
O.aries         ------------------------------------------------------------
S.scrofa        MRFFSGAENIPR-LFSCVALCLLWTGHAEAWITQSPRYEITVTRETVALQCYQTYNHDCV 59
C.familiaris    ---------MGSRLLCCVALFSWEPAPVESEVIQTPRHMIKVKRTDSDLRC-PYLWTLSV 50
H.sapiens       ---------MDSWTFCCVSLCILVAKHTDAGVIQSPRHEVTEMGQEVTLRCKPISGHNSL 51
T.vulpecula     ------------------------------------------------------------
O.fraenata      ------------------------------------------------------------
M.eugenii       ------------------------------------------------------------
M.domestica     ------------------------------------------------------------
A.mexicanum     ------------------------------------------------------------
O.mykiss        ----MIRILISITMGYRA---WAAGSSPSNQVHQGPADLYKNQGELAKMECSHSISTYNV 53
S.partitus      ----MKHVLIITGLCFTFNIILVSGSSLSDKVDQAPTDIYGKQGETAEITCSHKIDNYNR 56
P.olivaceus     ----MIPSLNTLTFFVLR----AAGVSHSVLITQWPHDISRFPSGSAEMHCYQNDTDYNH 52
X.laevis        ---------MGGYLTVLLLLSLLVGPNYGVKVTQVQKLLIIKSGEAAELYCEHDDSSYYN 51
E.caballus      ---------MGSRLLCCVALCLLGTGPVDSGVTQTPRHLIKARGQQVTLRCSPISGHNRV 51

A.platyrhynchos MSWYLQQPGKGLQLLYYSIGADQ-EAVGDTHPGYKATRLNLSDFHLVIKPVKMNHSADYF 105
G.gallus        MYWYLQQPGKGLQLIYYSYGTNL-DYKGDIHTGYEAKRLSQEVFRLNIVSVKKNHSAIYF 103
M.musculus      YWYRQDTGHGLRLIH-YSYGAGS-TEKGDIPDGYKASRPSQENFSLILELATPSQTSVYF 108
R.norvegicus    YWYRQDMGHGLRLIH-YSYDVNS-TEKGDVPNGYKVSRPSQGDFFLTLESASPSQTSVYF 108
O.aries         ----------------YLCAS----SKD------------------LAGGVSS------ 15
S.scrofa        YWYQQDQGHGLRLIF-YTCDVGI-LNKEEVPNGYNVSRPSMEDFSLILESVVPSRTSVYL 117
C.familiaris    YWYQQALM--VRLPVSHSVIIVK-KETSGQDSQCSSSVTTA--SQLEMNSLEPGDSALYL 105
H.sapiens       FWYRQTMMRGLELLIYFNNNVPI-DDSGMPEDRFSAKMPNASFSTLKIQPSEPRDSAVYF 110
T.vulpecula     ----------------------------------------------GVTLREELFMYK 12
O.fraenata      ------------------------------------------------------MYK 3
M.eugenii       ------------------------------------------------------------
M.domestica     ------------------------------------------------------------
A.mexicanum     ------------------------------------------------------------
O.mykiss        ILWYKQSNYRELVFLGYMQLKTG-FPEVGFD--IEGDANAGGTSTLTIKQLTPNSSAVYY 110
S.partitus      ILWYKQLN-RNLQFLGYLNINKG-YPEDGVDVTIDGDANKGRNCTLTINSLSVSSSAVYF 114
P.olivaceus     MYWYRQQRGKEPQLVVYLVGSSA-NLEEGFKSGFEAEIVQKKKWSLKIPSIQEKDEAVYL 111
X.laevis        MFWYQQKPDQGLKLMLHSLNVGSEDVESDYKDNWGTDRKFVLNSTLILKKGNVEDSATYF 111
E.caballus      FWYQQPLGQGPQFLFYYYNGKEN---------------------------------- 74

A.platyrhynchos CASSP---NRGSNTQYFGEGTKITVLEKNDVIKPPA-VAIFSPSKQEIQEK----SKATL 157
G.gallus        CASTR---DRVSGNMIFGDGTKLTVIGKNSEIIEPD-VVIFSPSKQEIQGK----KKATL 155
M.musculus      CASGE--GGLGGPTQYFGPGTRLLVLEDLRNVTPPK-VSLFEPSKAEIANK----QKATL 161
R.norvegicus    CASSD--S---GNVLYFGEGSRLLVVEDLKTVTPPK-VSLFEPSEAEIADK----QKATL 158
O.aries         ------------ETQYFGPGTRLLVLDDLRQVHPPK-VAVFEPSEAEISRT----QKATL 58
S.scrofa        CASSR--Q---GNTQHFGPGTWLTVLEDLQQVRPPK-VAVFEPSEAEISRT----QKATL 167
C.familiaris    CASSG---YSESYERYFGAGTRLTVLEDLQKVTPPT-VTVFEPSEAEISRT----QKATL 157
                •                    ◆  ◆
H.sapiens       CASS----FNGAGEAFFGQGTRLTVVEDLNKVFPPE-VAVFEPSEAEISHT----QKATL 161
T.vulpecula     TVTN-------YSELHFGPGTRLSVVDDLTKVTPPK-VTVFQPSEEEMEEK----GKATL 60
O.fraenata      TVTN-------YSELHFGPGTRLSVVDDLTRATPPK-VTVFQPSEEEMANK----GKATL 51
M.eugenii       -----------------GPGTRLTVTDDLTRVTPPK-VTVFQPSEEEMANK----GKATL 38
M.domestica     -------------------------DDLERVTPPK-VTVFQPSEEEIGEK----GKATL 29
A.mexicanum     ---------------RFGQGTKLTVLEEGLSVTQPS-VVLFDPSPQEIKKK----GKATL 40
O.mykiss        CAA---TGTKNYNPAFFGAGTKLTVLDPNIKVTEPT-VKVLAPSAKECEDR-NKKKKKTL 165
S.partitus      CAASYGTGGPQTEPAYFGKGTKLTVLETDRTVTPPTKVKIFPPSAKECRNKKDDIRKKTL 174
P.olivaceus     CAA---SGTRILYEAYFGQGTKPTVLEPGQAVKSPK-VKVFRPSSKECRNPIDNEREKTL 167
X.laevis        CAARS---GQGADRLYFGDGTILTVLEKGQEEKNVANVAIFRPNPKEQTDH----GYLSI 164
E.caballus      -----------------EKGDVPARFSGQQFSDYR--SELNVSVLELKDS------ALY 108
                                                        .  .  *

A.platyrhynchos VCLASGFYPDTLNLVWKVNGAERTEGVGTDETSTS------------YENTYSLTSRLRI 205
G.gallus        VCLASGFFPDHLNLVWKVNGVKRTEGVGTDEISTS------------NGSTYSLTSRLRI 203
M.musculus      VCLARGFFPDHVELSWWVNGKEVHSGVSTDPQAYK-------E---SN-YSYCLSSRLRV 210
R.norvegicus    VCLARGFFPDHVELSWWVNGKEIRNGVSTDPQAYK-------E---SNNITYCLSSRLRV 208
O.aries         VCLATGFYPDHVELTWWVNRKQVTTGVSTDPEPYK-------EDLTQNDSRYCLSSRLRV 111
S.scrofa        VCLATGFYPDHVELSWWVNGKQVQSGVSTDLQPYR-------EDPSRNDSCYCLSSRLRV 220
C.familiaris    VCLARGFYPDHVELSWWN----------------------------DSSYCLSSRLRV 187
```

Appendix 3C

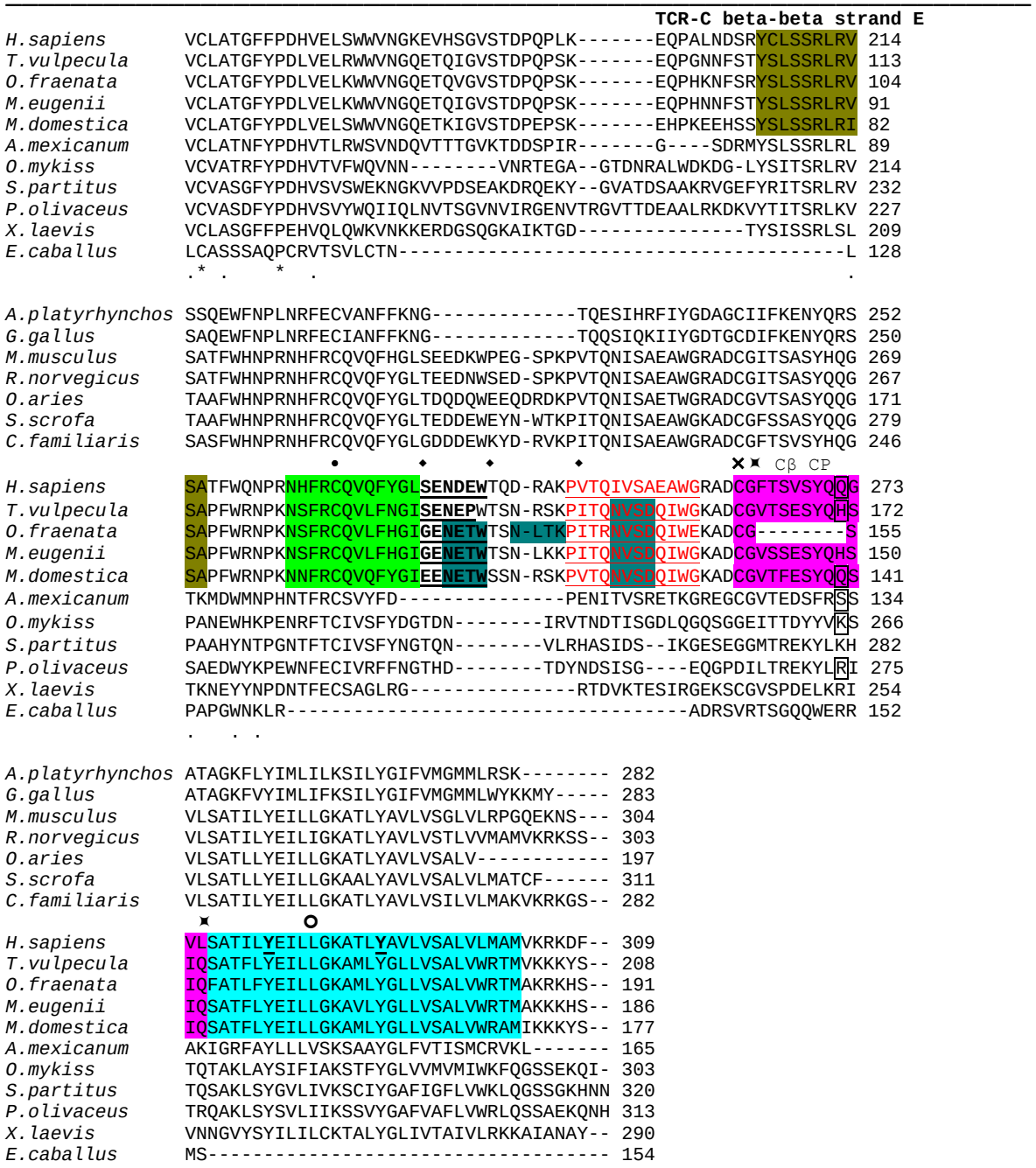


Fig. 3Cc. Sequence alignment for TCRβ.
IG domain = grey. **Y** = evolutionary conserved tyrosine residues. Blue = the transmembrane domain. G-loop = red. Cβ elbow loop = black bold. F-loop = green. ◆ = base of FG loop. ✖ = intradomain disulphide bridge. Cβ CP = magenta. Boxed residue important in signal transduction. ✕ = Interchain disulphide bond to TCRα. • = intradomain disufide with C147 of Cβ. O = conserved leucine motif involved in signal transduction. N-linked glycosylation sites.

Appendix 3C

Genbank Accession Numbers

Table3C.1. Accession Numbers for TCR beta chain found in Genbank and the relevant references.

Species name	Common name	Accession Numbers	Reference
Ambystoma mexicanum	Axolotl	L08498	(Fellah *et al.,* 1993) (partial sequence)
Anas platyrhynchos	Duck	AY039002	Unpublished
Canis lupus familiaris	Dog	D16410	(Takano *et al.,* 1994)
Equus caballus	Horse	XM_003364894	Annotated
Gallus gallus	Chicken	EF554759, M37803	Unpublished (Tjoelker *et al.,* 1990)
Heterodontus francisci	Horn shark	U07624	(Rast and Litman, 1994)
Homo sapiens	Human	AY232284	Direct Submission
Monodelphis domestica	American short tailed opossum	AY014506	(Baker *et al.,* 2001) (partial sequence)
Mus musculus	Mouse	DQ340294	Direct Submission
Oncorhynchus mykiss	Rainbow trout	AF329700	Direct Submission
Ovis aries	Sheep	FM993981, M94182	(Di Tommaso *et al.,* 2010) (partial sequence) (Grossberger *et al.,* 1993)
Paralichthys olivaceus	Japanese flounder	AB053228	Direct Submission
R.norvegicustus norvegicus	Norway rat	AY228549	Direct Submission
Stegastes partitus	Bicolor Damselfish	AF324823	(Kamper and McKinney, 2002)
Sus scrofa	Pig	AB079530	(Watanabe *et al.,* 2007)
Trichosurus vulpecula	Australian silver brushtail possum	AF133098	(Zuccolotto *et al.,* 2000)
Xenopus laevis	Frog	U60424	(Chretien *et al.,* 1997)

Appendix 4A

O. fraenata partial CD4 nucleotide sequence

```
ttcaaggggacagccaccccaagtgactatgtgacctctggaaccaatgtgactttaactttgcacagctctt
ccaaccttcttgcattcaaggtggaatggaggggtccaggagataaaagtaaacagatcatgaatcaagacaa
gaagactttgaacttggtgaaaatggggccaaatgaaacaggtctctgggactgtactgtctctgtgagtgag
aaaaccctgaaactgggcatcaaagtcacagcatttggtttcacaaaatcttctcagaccttctatacgatgg
tgggcaaagctgtcaaattctccttccctctgaatttaaatgaccaagagctgaacagggaacagccaaatgg
agaactgaggtggaaggtggaagaccctgcttcttctctacaggtggccaagttttcatggaagagtgactcc
ttgactctaaaaacaacgactccacgtttcagtcgtgatcccaagttcccactcacgatcactctttcctccg
tcttgccttctgatgctggctcaggagtcttcttactaaagttttcttcggggactgtggaacagaaggtcaa
ccttgtggtaatgaaagctatgtccagggaatcaccacatcacgaactgtgctgtgaagtgctgggtcccata
atcctgggttggttctgacatggattcgagaaaaccagcacagaaaaagaacaactgctattgcggaaaaaga
aacaactaaaatcagattccttcaaactgacagcaaaacaggagtggaacctacccttcatctgggcatcgg
tctgggggccggagcaagtctgttgcttctctctggactctgtatattcttttgtgccagacgaaggcacagg
```

O. fraenata partial putative protein sequence

```
FKGTATPSDYVTSGTNVTLTLHSSSNLLAFKVEWRGPGDKSKQIMNQDKKTLNLVKMGPNETGLWDCTVSVSE
KTLKLGIKVTAFGFTKSSQTFYTMVGKAVKFSFPLNLNDQELNREQPNGELRWKVEDPASSLQVAKFSWKSDS
LTLKTTTPRFSRDPKFPLTITLSSVLPSDAGSGVFLLKFSSGTVEQKVNLVVMKAMSRESPHHELCCEVLGPI
ILGWF-HGFEKTSTEKEQLLLRKKKQLKSDSFKLTAKQEWNLPFHLGIGLGAGASLLLLSGLCIFFCARRRHR
```

Appendix 4A

Secondary structure prediction for *O. fraenata* CD4 using PSIpred

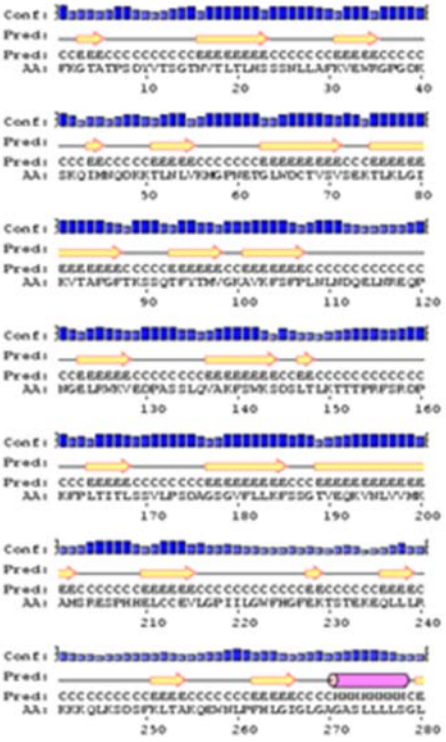


Figure 4A.1. *O. fraenata* secondary structure prediction for partial CD4
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 4A

Secondary structure prediction for *M. eugenii* CD4 using PSIpred

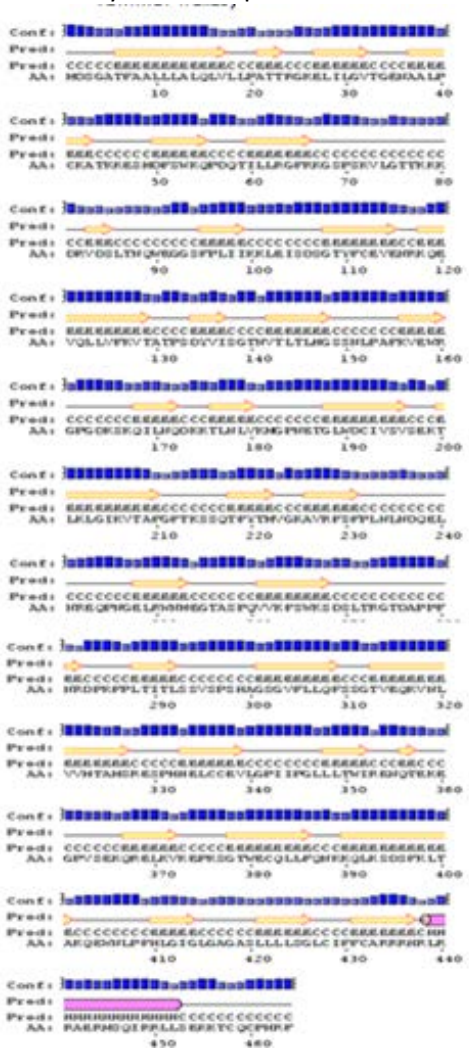


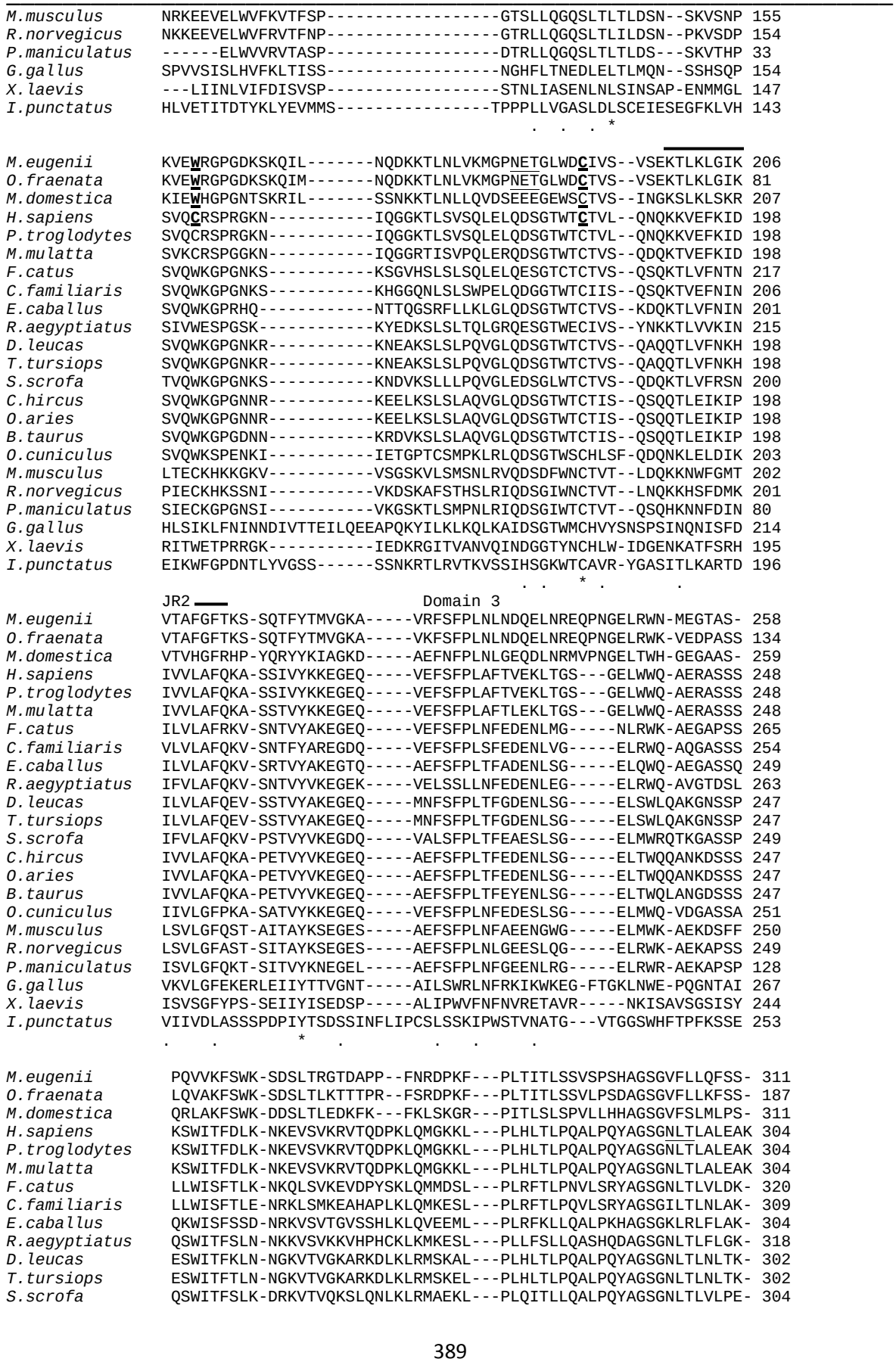
Figure 4A.2. *M. eugenii* CD4 secondary structure prediction.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 4A

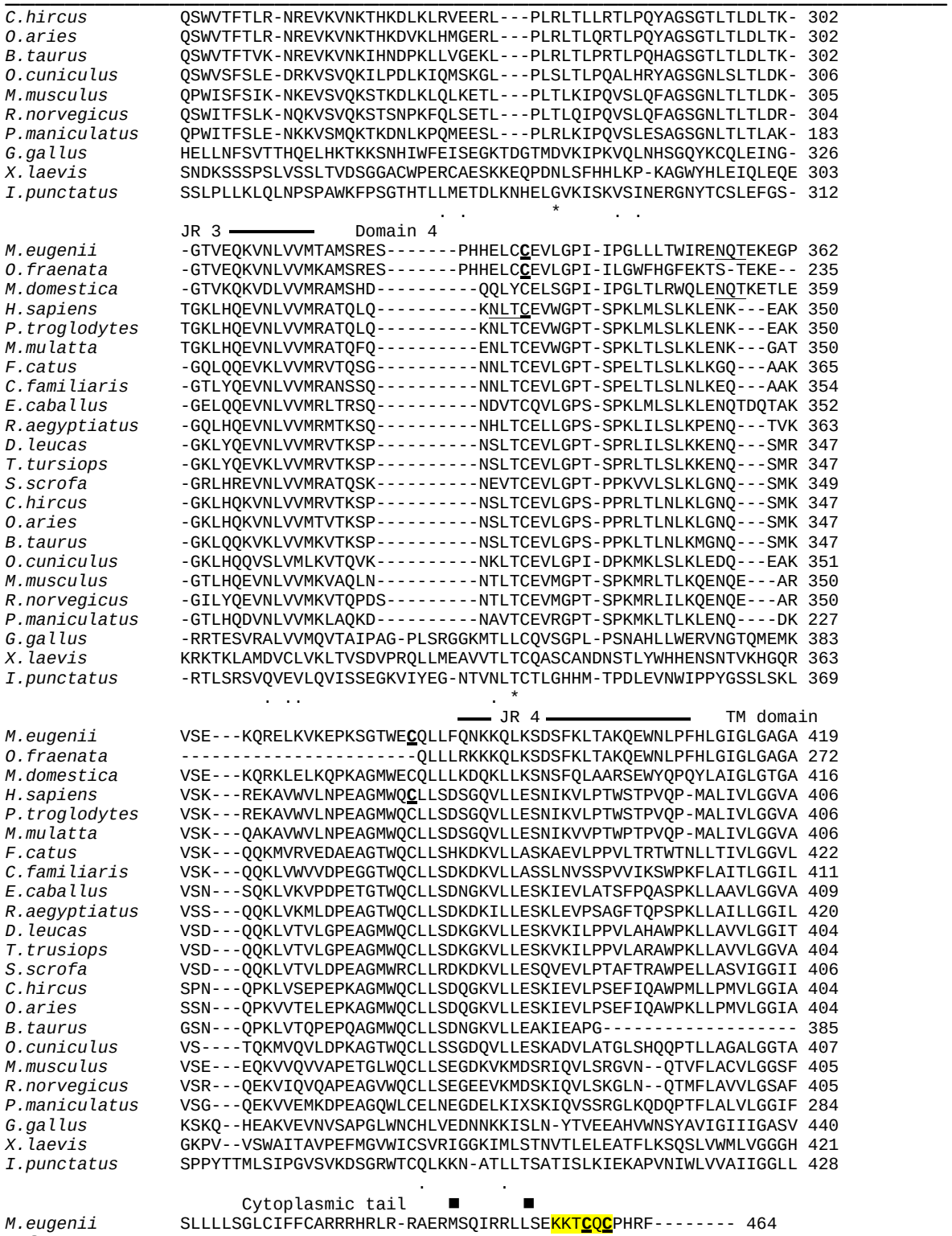
Amino Acid Alignment for CD4

```
M.eugenii       -MDSGATFAALLLALQLVLLPATT--RGKELILGVTGENAALPCKATK-KESMDFSWKQP 56
O.fraenata      ------------------------------------------------------------
M.domestica     -MSRGAALAMLLLALQLVLLPAMT--RGKESVLGQVGGTVELPCKASR-KERMDFAWKQQ 56
H.sapiens       -MNRGVPFRHLLLVLQLALLPAAT--QGKKVVLGKKGDTVELTCTASQ-KKSIQFHWKNS 56
P.troglodytes   -MNRGVPFRHLLLVLQLALLPAAT--QGKKVVLGKKGDTVELTCTASQ-KKSIQFHWKNS 56
M.mulatta       -MNRGIPFRHLLLVLQLALLPAVT--QGKKVVLGKKGDTVELTCTASQ-KKNTQFHWKNS 56
F.catus         -MNQGAVFRHLLLVLQLVMLKAAVPQ-GKEVVLGKAGGTAELPCQASQ-KKYMTFTWRLS 57
C.familiaris    -MNQEAAFRHLLLMLQLVMLPAVTP--VREVVLGKAGDAVELPCQTSQ-KKNIHFNWRDS 56
E.caballus      -MDRGTSFGHLLLLLQLALLPAVTQ--GREVVLGKAGETVELPCQGSQ-KKNVFFNWKYP 56
R.aegyptiatus   -MNLGSSFRHLLLLLQLALLPATTQ--GKEVVLGKAGDKAELPCQASQ-KKNMKFSWKY- 55
D.leucas        -MDPRTSLRHLFLVLQLVMLPAG--TQGKKVVLGKAGELAELPCKASQ-NKSLFFSWKNS 56
T.trusiops      -MDPRTSLRHLFLVLQLVMLPAG--TQGKKVVLGKAGELAELPCKASQ-NKSLFFSWKNS 56
S.scrofa        -MDPGTSLRHLFLVLQLAMLPAASGTQEKYLVLGKAGDLAELPCHSSQ-KKNLPFNWKNS 58
C.hircus        -MGPGTSLRHLFLVLQLVMLPAG--TQGKAVVLGKAGGQAELPCQASQ-KKNIVFSWKDS 56
O.aries         -MGPGTSLRHLFLVLQLVMLPAG--TQGKAVVLGKAGGQAELPCQASQ-KKNIVFSWKDS 56
B.taurus        -MGPGTSLRHLFLVLQLAMLPAG--TQGKTVVLGEAGDKAELPCQASQ-KKNMVFSWKDS 56
O.cuniculus     -MNRRIYFQCLLLVLPLALLPAAT--WGKTVVRGKAGAIVELPCQSSQ-KRNSVFNWKHA 56
M.musculus      -MCRAISLRRLL-LLLLQLSQLLAVTQGKTLVLGKEGESAELPCESSQ-KKITVFTWKFS 57
R.norvegicus    -MCRGFSFRHLLPLLLLQLSKLLVVTQGKTVVLGKEGGSAELPCESTS-RRSASFAWKSS 58
P.maniculatus   ------------------------------------------------------------
G.gallus        MERCGAVVSCVFAVILVLQLGLTPIMAQQEQQIGIAGKEVILSCKAINNQKDGTCTWKYK 60
X.laevis        --MEIQVITFISWLLFLQMGPSLTSPQIQMQMWTTVGASVIMPCNINT---NGDFTWKKN 55
I.punctatus     --------MSFLLGLLLLLAPCHSAADEPKGIFAQFGNSVTLPRRIWGIEGKIHVNWYFQ 52

M.eugenii       DQTILLRGFRKGSPSKVLGTT-KKKDRVDSLTNQWEGGSFPLIIKKLEISDSGTYFCEVE 115
O.fraenata      ------------------------------------------------------------
M.domestica     NQDLILKSFQKGLSRMMWGTSNILRNRADSSSNQWDMGSFPLTIQYLETSDSGMYFCEVE 116
H.sapiens       NQIKILGNQGSF---LTKG-PSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVE 112
P.troglodytes   NQTKILGNQGSF---LTKG-PSKLNDRVDSRRSLWDQGNFTLIIKNLKIEDSDTYICEVG 112
M.mulatta       NQIKILGIQGLF---LTKG-PSKLSDRADSRKSLWDQGCFSMIIKNLKIEDSDTYICEVE 112
F.catus         SQVKILESQHSS---LCLTGSSKLKTRFESKKILWDQGSFPLVIKSLQVADSGIYTCEVE 114
C.familiaris    SMVQILGNQGS----FWTVGSSRLKHRVESKKNLWDQGSFPLVIKDLEVADSGIYFCDTD 112
E.caballus      SEVKILGGQR-T---WWVKGNTKLKDRIESKTTLWDQGSFPLIIKYLEITDSGTYICEVE 112
R.aegyptiatus   SGIKVLENFPGS---YKMLGTAVQSTRSDSSRNLWEQGSFPLIIRNLDIKDSGIYTCDVE 112
D.leucas        YQTKILGRHGYF---WHKG-ASNLHSRVESKINLWEQGSFPLVIKDLEVPDSGTYICEVE 112
T.trusiops      YQTKILGRHGYF---WHKG-ASNLNSRVESKINLWDQGSFPLVIKDLEVPDSGTYICEVE 112
S.scrofa        NQTKILGGHGSF---WHTASVTELTSRLDSKKNMWDHGSFPLIIKNLEVTDSGIYICEVE 115
C.hircus        SQSKILGSHNSF---LHKG-NTELSRRVESKRNLWDQGSFPLIIKNLQVTDSGTYTCEVD 112
O.aries         SQSKILGSHNSF---LHKG-NTELSHRVESKKNLWDQGSFPLIIKNLQVTDSGTYTCEVD 112
B.taurus        SQSNILGKRGLF---FYKG-TTELSHRVESKKNLWDQGSFPLIIKNLQVTDSGTYTCEVD 112
O.cuniculus     NQVKILGNQGSSSSSFWLKGNSPLSNRVESKKNMWDQGSFPLVIKDLRMDDSGTYICEVG 116
M.musculus      DQRKILGQHGKG--VLIRGGSPSQFDRFDSKKGAWEKGSFPLIINKLKMEDSQTYICELE 115
R.norvegicus    DQKTILGYKNK---LLIKG-SLELYSRFDSRKNAWERGSFPLIINKLRMEDSQTYVCELE 114
P.maniculatus   ------------------------------------------------------------
G.gallus        YKEVSSTIISFS------KAQVFKGKAPMTHRSELNSNSKKLKVSDLSLDDAGIYTCACY 114
X.laevis        GVTYARRLNSQIN------YGSLAESSRISFPHETKLKNYSMQLVNIKLNDFGKYYCDEK 109
I.punctatus     DNLLISRNPTLS-----------ASKTVHNRFSLSSDSS--LIISNVEKSDFGIFKCEQH 99

                        JR 1                          Domain 2
M.eugenii       NRKQEVQLLVFKVTATP-----------------SDYVISGTNVTLTLHGS--SNLPAF 155
O.fraenata      ----------FKGTATP-----------------SDYVTSGTNVTLTLHSS--SNLLAF 30
M.domestica     DKKQQVQLLVFKVTANP-----------------SESVLSGNNVTLTLHSP--PNLPEL 156
H.sapiens       DQKEEVQLLVFGLTANS-----------------DTHLLQGQSLTLTLES---PPGSSP 151
P.troglodytes   DQKEEVQLLVFGLTANS-----------------DTHLLQGQSLTLTLES---PPGSSP 151
M.mulatta       NKKEEVELLVFGLTANS-----------------DTHLLEGQSLTLTLES---PPGSSP 151
F.catus         NKKREVELLVFGLTAKVDPSGSGGS-SSSSTSTSTSIYLLQGQSLTLTLES---PSSSNP 170
C.familiaris    -KRQEVELLVFNLTAKWD---------SGSSSGSSNIRLLQGQQLTLTLEN---PSGSSP 159
E.caballus      DKKTEVELLVFRLTAN-------------SHTHSSIRLLLGQNLTLTLES---PSGSNP 155
R.aegyptiatus   DKKREVELLVFSLNADLDTGSSSSSGSSSGSSSGSSIRLRPGERLTLSLES---PPGVNP 169
D.leucas        DKKIEVELQVFRLTASS-----------------DTRLLLGQSLTLTLEG---PSGSNP 151
T.tursiops      DKKIEVELQVFRLTASS-----------------DTRLLLGQSLTLTLEG---PSGSNP 151
S.scrofa        DKRIEVQLLVFRLTAS------------------VTRVLLGQSLTLTLEG---PSGSHP 153
C.hircus        SKKLEVELKVFGLTASS-----------------DTRVLLGQSLTLTLES---PSGSNP 151
O.aries         SKKLEVELKVFGLTASS-----------------DTRVLLGQSLTLTLES---PSGSNP 151
B.taurus        KKTLEVELQVFRLTASS-----------------DTHVLLGQSLTLTLES---PSGSNP 151
O.cuniculus     DKKMEVELLVFRLTANP-----------------NTRLLHGQSLTLTLEG---PSVGSP 155
```

Appendix 4A


M.musculus NRKEEVELWVFKVTFSP----------GTSLLQGQSLTLTLDSN--SKVSNP 155
R.norvegicus NKKEEVELWVFRVTFNP----------GTRLLQGQSLTLILDSN--PKVSDP 154
P.maniculatus ------ELWVVRVTASP----------DTRLLQGQSLTLTLDS---SKVTHP 33
G.gallus SPVVSISLHVFKLTISS----------NGHFLTNEDLELTLMQN--SSHSQP 154
X.laevis ---LIINLVIFDISVSP----------STNLIASENLNLSINSAP-ENMMGL 147
I.punctatus HLVETITDTYKLYEVMMS----------TPPPLLVGASLDLSCEIESEGFKLVH 143
M.eugenii KVEWRGPGDKSKQIL-------NQDKKTLNLVKMGPNETGLWDCIVS--VSEKTLKLGIK 206
O.fraenata KVEWRGPGDKSKQIM-------NQDKKTLNLVKMGPNETGLWDCTVS--VSEKTLKLGIK 81
M.domestica KIEWHGPGNTSKRIL-------SSNKKTLNLLQVDSEEEGEWSCTVS--INGKSLKLSKR 207
H.sapiens SVQCRSPRGKN----------IQGGKTLSVSQLELQDSGTWTCTVL--QNQKKVEFKID 198
P.troglodytes SVQCRSPRGKN----------IQGGKTLSVSQLELQDSGTWTCTVL--QNQKKVEFKID 198
M.mulatta SVKCRSPGGKN----------IQGGRTISVPQLERQDSGTWTCTVS--QDQKTVEFKID 198
F.catus SVQWKGPGNKS----------KSGVHSLSLSQLELQESGTCTCTVS--QSQKTLVFNTN 217
C.familiaris SVQWKGPGNKS----------KHGGQNLSLSWPELQDGGTWTCIIS--QSQKTVEFNIN 206
E.caballus SVQWKGPRHQ----------NTTQGSRFLLKLGLQDSGTWTCTVS--KDQKTLVFNIN 201
R.aegyptiatus SIVWESPGSK----------KYEDKSLSLTQLGRQESGTWECIVS--YNKKTLVVKIN 215
D.leucas SVQWKGPGNKR----------KNEAKSLSLPQVGLQDSGTWTCTVS--QAQQTLVFNKH 198
T.tursiops SVQWKGPGNKR----------KNEAKSLSLPQVGLQDSGTWTCTVS--QAQQTLVFNKH 198
S.scrofa TVQWKGPGNKS----------KNDVKSLLLPQVGLEDSGLWTCTVS--QDQKTLVFRSN 200
C.hircus SVQWKGPGNNR----------KEELKSLSLAQVGLQDSGTWTCTIS--QSQQTLEIKIP 198
O.aries SVQWKGPGNNR----------KEELKSLSLAQVGLQDSGTWTCTIS--QSQQTLEIKIP 198
B.taurus SVQWKGPGDNN----------KRDVKSLSLAQVGLQDSGTWTCTIS--QSQQTLEIKIP 198
O.cuniculus SVQWKSPENKI----------IETGPTCSMPKLRLQDSGTWSCHLSF-QDQNKLELDIK 203
M.musculus LTECKHKKGKV----------VSGSKVLSMSNLRVQDSDFWNCTVT--LDQKKNWFGMT 202
R.norvegicus PIECKHKSSNI----------VKDSKAFSTHSLRIQDSGIWNCTVT--LNQKKHSFDMK 201
P.maniculatus SIECKGPGNSI----------VKGSKTLSMPNLRIQDSGIWTCTVT--QSQHKNNFDIN 80
G.gallus HLSIKLFNINNDIVTTEILQEEAPQKYILKLKQLKAIDSGTWMCHVYSNSPSINQNISFD 214
X.laevis RITWETPRRGK----------IEDKRGITVANVQINDGGTYNCHLW-IDGENKATFSRH 195
I.punctatus EIKWFGPDNTLYVGSS------SSNKRTLRVTKVSSIHSGKWTCAVR-YGASITLKARTD 196
JR2
Domain 3
M.eugenii VTAFGFTKS-SQTFYTMVGKA-----VRFSFPLNLNDQELNREQPNGELRWN-MEGTAS- 258
O.fraenata VTAFGFTKS-SQTFYTMVGKA-----VKFSFPLNLNDQELNREQPNGELRWK-VEDPASS 134
M.domestica VTVHGFRHP-YQRYYKIAGKD-----AEFNFPLNLGEQDLNRMVPNGELTWH-GEGAAS- 259
H.sapiens IVVLAFQKA-SSIVYKKEGEQ-----VEFSFPLAFTVEKLTGS---GELWWQ-AERASSS 248
P.troglodytes IVVLAFQKA-SSIVYKKEGEQ-----VEFSFPLAFTVEKLTGS---GELWWQ-AERASSS 248
M.mulatta IVVLAFQKA-SSTVYKKEGEQ-----VEFSFPLAFTLEKLTGS---GELWWQ-AERASSS 248
F.catus ILVLAFRKV-SNTVYAKEGEQ-----VEFSFPLNFEDENLMG-----NLRWK-AEGAPSS 265
C.familiaris VLVLAFQKV-SNTFYAREGDQ-----VEFSFPLSFEDENLVG-----ELRWQ-AQGASSS 254
E.caballus ILVLAFQKV-SRTVYAKEGTQ-----AEFSFPLTFADENLSG-----ELQWQ-AEGASSQ 249
R.aegyptiatus IFVLAFQKV-SNTVYVKEGEK-----VELSSLLNFEDENLEG-----ELRWQ-AVGTDSL 263
D.leucas ILVLAFQEV-SSTVYAKEGEQ-----MNFSFPLTFGDENLSG-----ELSWLQAKGNSSP 247
T.tursiops ILVLAFQEV-SSTVYAKEGEQ-----MNFSFPLTFGDENLSG-----ELSWLQAKGNSSP 247
S.scrofa IFVLAFQKV-PSTVYVKEGDQ-----VALSFPLTFEAESLSG-----ELMWRQTKGASSP 249
C.hircus IVVLAFQKA-PETVYVKEGEQ-----AEFSFPLTFEDENLSG-----ELTWQQANKDSSS 247
O.aries IVVLAFQKA-PETVYVKEGEQ-----AEFSFPLTFEDENLSG-----ELTWQQANKDSSS 247
B.taurus IVVLAFQKA-PETVYVKEGEQ-----AEFSFPLTFEYENLSG-----ELTWQLANGDSSS 247
O.cuniculus IIVLGFPKA-SATVYKKEGEQ-----VEFSFPLNFEDESLSG-----ELMWQ-VDGASSA 251
M.musculus LSVLGFQST-AITAYKSEGES-----AEFSFPLNFAEENGWG-----ELMWK-AEKDSFF 250
R.norvegicus LSVLGFAST-SITAYKSEGES-----AEFSFPLNLGEESLQG-----ELRWK-AEKAPSS 249
P.maniculatus ISVLGFQKT-SITVYKNEGEL-----AEFSFPLNFGEENLRG-----ELRWR-AEKAPSP 128
G.gallus VKVLGFEKERLEIIYTTVGNT-----AILSWRLNFRKIKWKEG-FTGKLNWE-PQGNTAI 267
X.laevis ISVSGFYPS-SEIIYISEDSP-----ALIPWVFNFNVRETAVR-----NKISAVSGSISY 244
I.punctatus VIIVDLASSSPDPIYTSDSSINFLIPCSLSSKIPWSTVNATG---VTGGSWHFTPFKSSE 253
M.eugenii PQVVKFSWK-SDSLTRGTDAPP--FNRDPKF---PLTITLSSVSPSHAGSGVFLLQFSS- 311
O.fraenata LQVAKFSWK-SDSLTLKTTTPR--FSRDPKF---PLTITLSSVLPSDAGSGVFLLKFSS- 187
M.domestica QRLAKFSWK-DDSLTLEDKFK---FKLSKGR---PITLSLSPVLLHHAGSGVFSLMLPS- 311
H.sapiens KSWITFDLK-NKEVSVKRVTQDPKLQMGKKL---PLHLTLPQALPQYAGSGNLTLALEAK 304
P.troglodytes KSWITFDLK-NKEVSVKRVTQDPKLQMGKKL---PLHLTLPQALPQYAGSGNLTLALEAK 304
M.mulatta KSWITFDLK-NKEVSVKRVTQDPKLQMGKKL---PLHLTLPQALPQYAGSGNLTLALEAK 304
F.catus LLWISFTLK-NKQLSVKEVDPYSKLQMMDSL---PLRFTLPNVLSRYAGSGNLTLVLDK- 320
C.familiaris LLWISFTLE-NRKLSMKEAHAPLKLQMKESL---PLRFTLPQVLSRYAGSGILTLNLAK- 309
E.caballus QKWISFSSD-NRKVSVTGVSSHLKLQVEEML---PLRFKLLQALPKHAGSGKLRLFLAK- 304
R.aegyptiatus QSWITFSLN-NKKVSVKKVHPHCKLKMKESL---PLLFSLLQASHQDAGSGNLTLFLGK- 318
D.leucas ESWITFKLN-NGKVTVGKARKDLKLRMSKAL---PLHLTLPQALPQYAGSGNLTLNLTK- 302
T.tursiops ESWITFTLN-NGKVTVGKARKDLKLRMSKEL---PLHLTLPQALPQYAGSGNLTLNLTK- 302
S.scrofa QSWITFSLK-DRKVTVQKSLQNLKLRMAEKL---PLQITLLQALPQYAGSGNLTLVLPE- 304

Appendix 4A


C.hircus QSWVTFTLR-NREVKVNKTHKDLKLRVEERL---PLRLTLLRTLPQYAGSGTLTLDLTK- 302
O.aries QSWVTFTLR-NREVKVNKTHKDVKLHMGERL---PLRLTLQRTLPQYAGSGTLTLDLTK- 302
B.taurus QSWVTFTVK-NREVKVNKIHNDPKLLVGEKL---PLRLTLPRTLPQHAGSGTLTLDLTK- 302
O.cuniculus QSWVSFSLE-DRKVSVQKILPDLKIQMSKGL---PLSLTLPQALHRYAGSGNLSLTLDK- 306
M.musculus QPWISFSIK-NKEVSVQKSTKDLKLQLKETL---PLTLKIPQVSLQFAGSGNLTLTLDK- 305
R.norvegicus QSWITFSLK-NQKVSVQKSTSNPKFQLSETL---PLTLQIPQVSLQFAGSGNLTLTLDR- 304
P.maniculatus QPWITFSLE-NKKVSMQKTKDNLKPQMEESL---PLRLKIPQVSLESAGSGNLTLTLAK- 183
G.gallus HELLNFSVTTHQELHKTKKSNHIWFEISEGKTDGTMDVKIPKVQLNHSGQYKCQLEING- 326
X.laevis SNDKSSSPSLVSSLTVDSGGACWPERCAESKKEQPDNLSFHHLKP-KAGWYHLEIQLEQE 303
I.punctatus SSLPLLKLQLNPSPAWKFPSGTHTLLMETDLKNHELGVKISKVSINERGNYTCSLEFGS- 312
JR 3
Domain 4
M.eugenii -GTVEQKVNLVVMTAMSRES-------PHHELCCEVLGPI-IPGLLLTWIRENQTEKEGP 362
O.fraenata -GTVEQKVNLVVMKAMSRES-------PHHELCCEVLGPI-ILGWFHGFEKTS-TEKE-- 235
M.domestica -GTVKQKVDLVVMRAMSHD----------QQLYCELSGPI-IPGLTLRWQLENQTKETLE 359
H.sapiens TGKLHQEVNLVVMRATQLQ----------KNLTCEVWGPT-SPKLMLSLKLENK---EAK 350
P.troglodytes TGKLHQEVNLVVMRATQLQ----------KNLTCEVWGPT-SPKLMLSLKLENK---EAK 350
M.mulatta TGKLHQEVNLVVMRATQFQ----------ENLTCEVWGPT-SPKLTLSLKLENK---GAT 350
F.catus -GQLQQEVKLVVMRVTQSG----------NNLTCEVLGPT-SPELTLSLKLKGQ---AAK 365
C.familiaris -GTLYQEVNLVVMRANSSQ----------NNLTCEVLGPT-SPELTLSLNLKEQ---AAK 354
E.caballus -GELQQEVNLVVMRLTRSQ----------NDVTCQVLGPS-SPKLMLSLKLENQTDQTAK 352
R.aegyptiatus -GQLHQEVNLVVMRMTKSQ----------NHLTCELLGPS-SPKLILSLKPENQ---TVK 363
D.leucas -GKLYQEVNLVVMRVTKSP----------NSLTCEVLGPT-SPRLILSLKKENQ---SMR 347
T.tursiops -GKLYQEVKLVVMRVTKSP----------NSLTCEVLGPT-SPRLTLSLKKENQ---SMR 347
S.scrofa -GRLHREVNLVVMRATQSK----------NEVTCEVLGPT-PPKVVLSLKLGNQ---SMK 349
C.hircus -GKLHQKVNLVVMRVTKSP----------NSLTCEVLGPS-PPRLTLNLKLGNQ---SMK 347
O.aries -GKLHQKVNLVVMTVTKSP----------NSLTCEVLGPS-PPRLTLNLKLGNQ---SMK 347
B.taurus -GKLQQKVKLVVMKVTKSP----------NSLTCEVLGPS-PPKLTLNLKMGNQ---SMK 347
O.cuniculus -GKLHQQVSLVMLKVTQVK----------NKLTCEVLGPI-DPKMKLSLKLEDQ---EAK 351
M.musculus -GTLHQEVNLVVMKVAQLN----------NTLTCEVMGPT-SPKMRLTLKQENQE---AR 350
R.norvegicus -GILYQEVNLVVMKVTQPDS---------NTLTCEVMGPT-SPKMRLILKQENQE---AR 350
P.maniculatus -GTLHQDVNLVVMKLAQKD----------NAVTCEVRGPT-SPKMKLTLKLENQ----DK 227
G.gallus -RRTESVRALVVMQVTAIPAG-PLSRGGKMTLLCQVSGPL-PSNAHLLWERVNGTQMEMK 383
X.laevis KRKTKLAMDVCLVKLTVSDVPRQLLMEAVVTLTCQASCANDNSTLYWHHENSNTVKHGQR 363
I.punctatus -RTLSRSVQVEVLQVISSEGKVIYEG-NTVNLTCTLGHHM-TPDLEVNWIPPYGSSLSKL 369
JR 4
TM domain
M.eugenii VSE---KQRELKVKEPKSGTWECQLLFQNKKQLKSDSFKLTAKQEWNLPFHLGIGLGAGA 419
O.fraenata ----------QLLLRKKKQLKSDSFKLTAKQEWNLPFHLGIGLGAGA 272
M.domestica VSE---KQRKLELKQPKAGMWECQLLLKDQKLLKSNSFQLAARSEWYQPQYLAIGLGTGA 416
H.sapiens VSK---REKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP-MALIVLGGVA 406
P.troglodytes VSK---REKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP-MALIVLGGVA 406
M.mulatta VSK---QAKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVVPTWPTPVQP-MALIVLGGVA 406
F.catus VSK---QQKMVRVEDAEAGTWQCLLSHKDKVLLASKAEVLPPVLTRTWTNLLTIVLGGVL 422
C.familiaris VSK---QQKLVWVVDPEGGTWQCLLSDKDKVLLASSLNVSSPVVIKSWPKFLAITLGGIL 411
E.caballus VSN---SQKLVKVPDPETGTWQCLLSDNGKVLLESKIEVLATSFPQASPKLLAAVLGGVA 409
R.aegyptiatus VSS---QQKLVKMLDPEAGTWQCLLSDKDKILLESKLEVPSAGFTQPSPKLLAILLGGIL 420
D.leucas VSD---QQKLVTVLGPEAGMWQCLLSDKGKVLLESKVKILPPVLAHAWPKLLAVVLGGIT 404
T.trusiops VSD---QQKLVTVLGPEAGMWQCLLSDKGKVLLESKVKILPPVLARAWPKLLAVVLGGVA 404
S.scrofa VSD---QQKLVTVLDPEAGMWRCLLRDKDKVLLESQVEVLPTAFTRAWPELLASVIGGII 406
C.hircus SPN---QPKLVSEPEPKAGMWQCLLSDQGKVLLESKIEVLPSEFIQAWPMLLPMVLGGIA 404
O.aries SSN---QPKVVTELEPKAGMWQCLLSDQGKVLLESKIEVLPSEFIQAWPKLLPMVLGGIA 404
B.taurus GSN---QPKLVTQPEPQAGMWQCLLSDNGKVLLEAKIEAPG------------------ 385
O.cuniculus VS----TQKMVQVLDPKAGTWQCLLSSGDQVLLESKADVLATGLSHQQPTLLAGALGGTA 407
M.musculus VSE---EQKVVQVVAPETGLWQCLLSEGDKVKMDSRIQVLSRGVN--QTVFLACVLGGSF 405
R.norvegicus VSR---QEKVIQVQAPEAGVWQCLLSEGEEVKMDSKIQVLSKGLN--QTMFLAVVLGSAF 405
P.maniculatus VSG---QEKVVEMKDPEAGQWLCELNEGDELKIXSKIQVSSRGLKQDQPTFLALVLGGIF 284
G.gallus KSKQ--HEAKVEVNVSAPGLWNCHLVEDNNKKISLN-YTVEEAHVWNSYAVIGIIIGASV 440
X.laevis GKPV--VSWAITAVPEFMGVWICSVRIGGKIMLSTNVTLELEATFLKSQSLVWMLVGGGH 421
I.punctatus SPPYTTMLSIPGVSVKDSGRWTCQLKKN-ATLLTSATISLKIEKAPVNIWLVVAIIGGLL 428
Cytoplasmic tail
M.eugenii SLLLLSGLCIFFCARRRHRLR-RAERMSQIRRLLSEKKTCQCPHRF-------- 464
O.fraenata SLLLLSGLCIFFCARRRHR---------- 291
M.domestica SLLLLFGFIMFCYARRRHRLR-RAERMSQIRRLLSEKKTCQCPHRF-------- 461
H.sapiens GLLLFIGLGIFFCVRCRHRRR-QAERMSQIKRLLSEKKTCQCPHRFQKTCSPI- 458
P.troglodytes GLLLFIGLGIFFCVRCRHRRR-QAQRMSQIKRLLSEKKTCQCPHRFQKTCSPI- 458
M.mulatta GLLLFTGLGIFFCVRCRHRRR-QAERMSQIKRLLSEKKTCQCPHRFQKTCSPI- 458
F.catus GLVLYIGLWVYCCVKCWHRRR-QAARMSHIKRLLSEKKTCQCSHRLQKTCNPI- 474
C.familiaris GLLLLIGLCVFCCVKCWRRRR-QAERMSQIKRLLSEKKTCQCSHRIQKTCSLI- 463
E.caballus GLLLFTGFFIFCCVKCWHRRR-QAERMSQIKRLLSEKKTCQCHQ---------- 452

```
R.aegyptiatus     GFLTFTVICIFCCVKCWHRRRRQAERMSQIKRLLSEKKTCQCPHRYQKTGLI-- 472
D.leucas          SLLLLAGFCIFS-AKCWHRRR-RAERTSQIKRLLSEKKTCHCSHRLQKTCSLT- 455
T.tursiops        SLLLLTGFCIFS-AKYWHRRR-RAQRTSQIKRLLSEKKTCHCSHRLQKTCSLT- 455
S.scrofa          GLLFLAGFCIAC-VKCWHRRR-RAERMSQIKRLLSEKKTCQCAHRQQKNYSLT- 457
C.hircus          GLALLTGSCIFC-VKCWHRRR-QAERMSQIKRLLSEKKTCQCPHRLQKTHSLT- 455
O.aries           GLALLTGSCIFC-VKCWHRRR-QAERMSQIKRLLSEKKTCQCPHRLQK------ 450
B.taurus          --------------------------------RTDVSNQEAP----------- 395
O.cuniculus       GLVLFAGLCIYCCVKCRHRRH-QAQRMSQIKKLLSEKKTCQCPHRLQKTYNLL- 459
M.musculus        GFLGFLGLCILCCVRCRHQQR-QAARMSQIKRLLSEKKTCQCPHRMQKSHNLI- 457
R.norvegicus      SFLVFTGLCILFCVRCRHQQR-QAARMSQIKRLLSEKKTCQCSHRMQKSHNLI- 457
P.maniculatus     SFLTFIGLCILCCVRCRHQQR-QAER--------------------------- 309
G.gallus          LVIGLACMCIITGMRWQRRRK-RARRMAQAKQYLLEKKTCQCQRRMYK------ 487
X.laevis          PALVGMVTIVILAARCRRKRR--ARRGAWILMNLDQQRRCQCKGFAPMRLREKD 473
I.punctatus       VFILIAVITVFIIRRHRQMMR----------YRCRKGRVCCCKNPKPKGFYKT- 471
```

Fig. 4A.3. Sequence alignment of all CD4 sequences including partial CD4 sequence of *O. fraenata* Glycosylation sites are underlined. Conserved cysteines are bold and underlined. Yellow highlight shows the $p56^{lck}$ binding site. ■ = serine phosphorylation site.

Appendix 4A

Table 4A.1. Accession Numbers for CD4 found in Genbank and the relevant references.

Species	Common name	Accession Numbers	References
Bos taurus	Cattle	NM_001103225	(Hoek *et al.,* 2009)
Canis lupus familiaris	Dog	NM_001003252	(Milde *et al.,* 1993)
Capra hircus	Goat	EU913093	(Wang *et al.,* 2011a)
Delphinapterus leucas	Beluga whale	AF071799	(Romano *et al.,* 1999)
Equus caballus	Horse	XM_001497051	Annotated
Felis catus	Cat	NM_001009250	(Norimine *et al.,* 1992)
Gallus gallus	Chicken	NM_204649	(Koskinen *et al.,* 1999)
Homo sapiens	Human	NM_000616	(Maddon *et al.,* 1985)
Ictalurus punctatus	Channel Catfish	DQ435305	(Edholm *et al.,* 2006)
Macaca mulatta	Rhesus monkey	NM_001042662	(Fomsgaard *et al., 1*992)
Macropus eugenii	Tammar wallaby	EF490599	(Duncan *et al.,* 2007)
Monodelphis domestica	South American grey short tailed opossum	NM_001099290, DQ665840	(Duncan *et al.,* 2007)
Mus musculus	Mouse	NM_013488	(Wineman *et al.,* 1992)
Nomascus leucogenys	Northern white-cheeked gibbon	XM_003273751	Annotated
Oryctolagus cuniculus	Rabbit	NM_001082313	(Hague *et al.,* 1992)
Ovis aries	Sheep	NM_001129902	(Boscariol *et al.,* 2006, Classon *et al.,* 1986)
Pan troglodytes	Chimpanzee	NM_001009043	Annotated
Peromyscus maniculatus	Deer mouse	DQ836358	Direct submission
Rattus norvegicus	Rat	NM_012705	(Clark *et al.,* 1987)
Rousettus aegyptiacus	Egyptian rousette	AB210837	(Omatsu *et al.,* 2006)
Sus scrofa	Pig	NM_001001908	(Gustafsson *et al.,* 1993)
Takifugu rubripes	Fugu rubripes	NM_001078623	(Suetake *et al.,* 2004)
Tursiops truncatus	Bottlenosed dolphin	AF408402	Romano unpublished
Xenopus laevis	Frog	HQ116782	(Chida *et al.,* 2011)

Appendix 4B

O. fraenata CD8α nucleotide sequence

atgggctccctcttggctgtacgatccctgctcctgccgctggccctgctgctccagtccgtcgggtcc
caggcacaggtgaaattccggatgaatcccctggagaagcgggacgtccggccatccgacaaggtgcag
ctgcagtgcgagactctgagcagttcgcccacgggatgttcgtggctgcgcttggtccctggaaaagtg
gtccctacctttctcctcttcatctcgagcactagcttaaacgcgaagctggccgaaggcctggacccc
aagcgattccggggcgaaaggatctcgtcttccacataccgcctgaccctgcagaacttcagagaggag
gaccagggctactactactgcgtggtcacccggaactcggcgctgttcttcagccccttcgtgccggtc
ttcctgccagtaaagactaccactacccctgcacccaaacccaagaccaccattcttccagccaccacc
agctcttcaacccagatttctgaaaactgcaagctcaccgtcaagaagcaaggaaagaagggattggat
ttctcctgtgacctgtacatctgggtgcccctcgctggcgtctgtgttatcttgtttctggccctgatc
actaccattaccatctgccagaggtcacgaaaacgagtctgccgatgtccgaggcccctgatcagacca
ggaggaaaggctggcccatcagagagatatgcttaa

O. fraenata CD8α putative amino acid sequence

MGSLLAVRSLLLPLALLLQSVGSQAQVKFRMNPLEKRDVRPSDKVQLQCETLSSSPTGCSWLRLVPGKV
VPTFLLFISSTSLNAKLAEGLDPKRFRGERISSSTYRLTLQNFREEDQGYYYCVVTRNSALFFSPFVPV
FLPVKTTTTPAPKPKTTILPATTSSSTQISENCKLTVKKQGKKGLDFSCDLYIWVPLAGVCVILFLALI
TTITICQRSRKRVCRCPRPLIRPGGKAGPSERYA

Secondary structure prediction CD8α (*M. eugenii*)

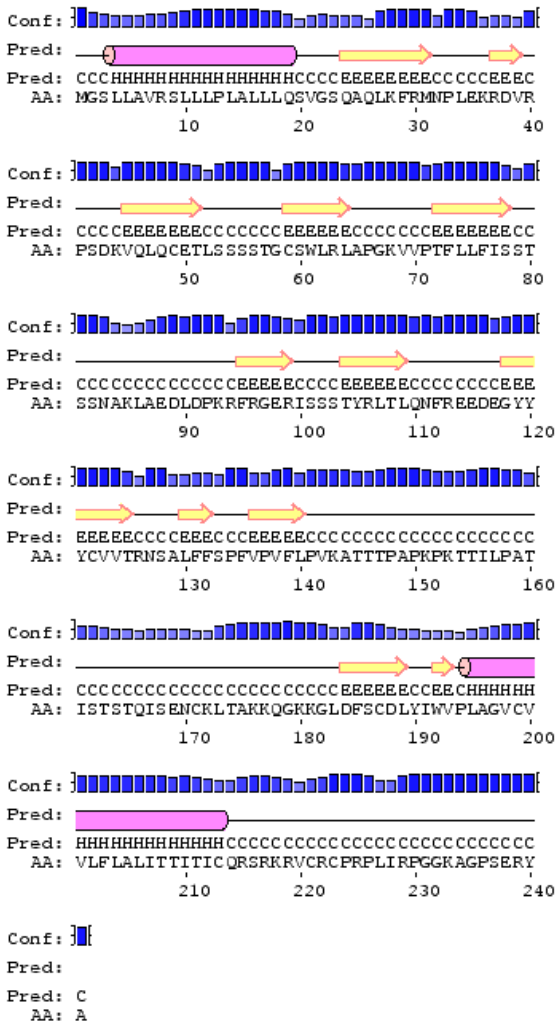


Figure 4B.1. *M. eugenii* CD8α secondary structure prediction.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 4B

L. hirsutus CD8α chain nucleotide sequence

atgggctccctcttggctgtacgatccctgctcctgccgctggccctgctgctccagtccgtcgggtcc
caggcacagctgaaattccggatgaatcccctggagaagcgggacgtccggccatccgacaaggtgcag
ctgcagtgcgatacgctgaacagttcgcccacgggatgttcgtggctgcgcttggcccctggaaaagtg
gtccctacctttctgctcttcatctcgagcagtagctcaaacgcgaagctggccgaagacctggacccc
aagcgattccggggcgaaaggatctcgtcttccacataccgcctgaccctgcagaacttcagagaggag
gacgagggctactactactgcgtggtcacccggaactcggcgctgttcttcagccccttcgtgccggtt
ttcctgccagtaaaggctaccaccacccctgcgcccaaacccaagaccaccattcttccagccaccacc
agctcttccacccagatttctgaaaactgcaagctcaccgtcaagaaacaaggaaagaagggattggat
ttctcctgtgacctgtacatctgggtgcccctcgctggcgtctgtgttatcttgtttctggccctgatc
actaccattaccatctgccagaggtcacgaaaacgagtctgccgatgtccgaggcccctgatcagacca
ggaggaaaggctggcccatcagagagatatgcttaa

L. hirsutus putative CD8α amino acid sequence

MGSLLAVRSLLLPLALLLQSVGSQAQLKFRMNPLEKRDVRPSDKVQLQCDTLSSSPTGCSWLRLAPGKV
VPTFLLFISSSSSNAKLAEDLDPKRFRGERISSSTYRLTLQNFREEDEGYYYCVVTRNSALFFSPFVPV
FLPVKATTTPAPKPKTTILPATTSSSTQISENCKLTVKKQGKKGLDFSCDLYIWVPLAGVCVILFLALI
TTITICQRSRKRVCRCPRPLIRPGGKAGPSERYA

Secondary structure prediction for CD8α (*L. hirsutus*)

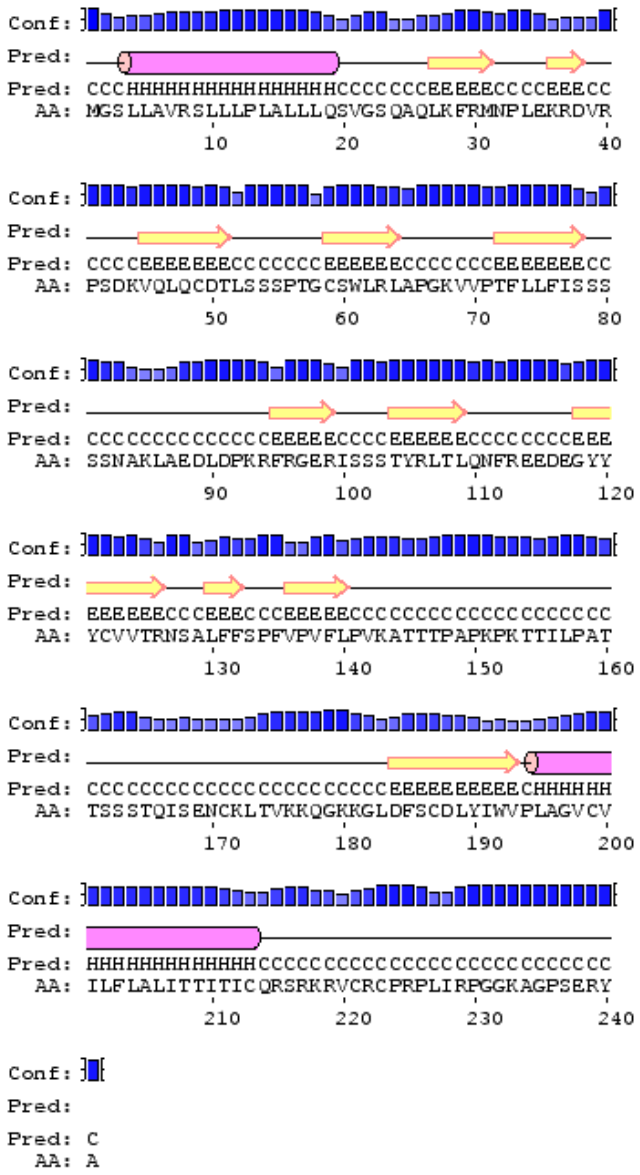


Figure 4B.2. *L. hirsutus* CD8α secondary structure prediction.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 4B

Alignment of marsupial CD8α chain putative amino acid sequences

```
M. eugenii      MGSLLAVRSLLLPLALLLQSVGSQAQLKFRMNPLEKRDVRPSDKVQLQCETLSSSSTGCS
L. hirsutus     MGSLLAVRSLLLPLALLLQSVGSQAQLKFRMNPLEKRDVRPSDKVQLQCDTLSSSPTGCS
O. fraenata     MGSLLAVRSLLLPLALLLQSVGSQAQVKFRMNPLEKRDVRPSDKVQLQCETLSSSPTGCS
M. domestica    MGWPSAVRSLLLPLALLLEPLGAQEHVTFRMNPVEKRDAGQGRELKLQCETVNSSPTGCS
                **   *************...*.* ...*****.****.  . ...***.*..**.****

M. eugenii      WLRLAPGKVVPTFLLFISSTSSNAKLAEDLDPKRFRGERISSSTYRLTLQNFREEDEGYY
L. hirsutus     WLRLAPGKVVPTFLLFISSSSSNAKLAEDLDPKRFRGERISSSTYRLTLQNFREEDEGYY
O. fraenata     WLRLVPGKVVPTFLLFISSTSLNAKLAEGLDPKRFRGERISSSTYRLTLQNFREEDQGYY
M. domestica    WLRLTPGAVVPTFLLYLSGSSQSVKVAPELDSRRFGGSRSSS-SYFLTLKDFQKDDQGYY
                ****.** *******..*..* ..*.*  **..** *.* ** .* ***..*...*.***

M. eugenii      YCVVTRNSALFFSPFVPVFLPVKATTTPAPKPKTTILPATISTSTQIS--ENCKLTAKKQ
L. hirsutus     YCVVTRNSALFFSPFVPVFLPVKATTTPAPKPKTTILPATTSSSTQIS--ENCKLTVKKQ
O. fraenata     YCVVTRNSALFFSPFVPVFLPVKTTTTPAPKPKTTILPATTSSSTQIS--ENCKLTVKKQ
M. domestica    YCVVARNSRLFFSPFVPVFMPVKATTAPAPRPKPTTPAATNSSIQNAAGSEKCKSFSNTS
                ****.*** **********.***.**.***.**.*  .** *.  . .  *.**   ...

M. eugenii      GKKGLDFSCDLYIWVPLAGVCVVLFLALITTITICQRSRKRVCRCPRPLIRPGGKAGPSE
L. hirsutus     GKKGLDFSCDLYIWVPLAGVCVILFLALITTITICQRSRKRVCRCPRPLIRPGGKAGPSE
O. fraenata     GKKGLDFSCDLYIWVPLAGVCVILFLALITTITICQRSRKRVCRCPRPLIRPGGKAGPSE
M. domestica    EKNGLDFSCDLYIWMPLTGGCVVLLLALIITITIFRRSRRRVCQCPRPLIRPGGKTGR--
                 *.**********.**.* **.*.**** **** .***.***.**********.*

M. eugenii      RYA
L. hirsutus     RYA
O. fraenata     RYA
M. domestica    ---
```

Fig. 4B.3. Alignment of all to date known marsupial CD8α putative amino acid sequences. Line 5(*) indicates the consensus between the macropods (*M. eugenii, L. hirsutus, O. fraenata*) and the didelphid *M. domestica*. In red CXC motif for binding Lck.

Appendix 4B

4B.1. Genbank Accession numbers for CD8α chain and relevant references.

Species	Common Name	Accession Number	References
Anas platyrhynchos	Duck	AY738733	(Kothlow *et al.,* 2005)
Bos taurus	Cattle	NM_174015	(Lalor *et al.,* 1992)
Cairina moschata	Muscovy duck	AY738735	(Kothlow *et al.,* 2005)
Callithrix jacchus	White-tuffed ear marmoset	DQ189217	Direct submission
Canis lupus familiaris	Dog	NM_001002935	(Gorman *et al.,* 1994)
Carassius auratus langsdorfii	Japanese silver crucian carp	AB186397	(Somamoto *et al.,* 2005)
Cyprinus carpio	Common carp	EU251078	Direct submission
Danio rerio	Zebrafish	NM_001040049	(Somamoto *et al.,* 2005)
Cavia porcellus	Domestic Guinea Pig	NM_001172876	(Nagarajan *et al.,* 2004)
Ctenopharyngodon idella	Grass carp	GQ355586	Direct submission
Dicentrarchus labrax	European Seabass	AJ846849	(Buonocore *et al.,* 2006)
Equus caballus	Horse	XM_001496953	Annotated
Felis catus	Cat	D16536	(Pecoraro *et al.,* 1994)
Gallus gallus	Chicken	NM_001048080	Provisional entry
Homo sapiens	Human	M12828	(Littman *et al.,* 1985)
Ictalurus punctatus	Channel catfish	NM_001200331	(Quiniou *et al.,* 2011)
Loxodonta africana	African savanna elephant	XM_003420420	Annotated
Macaca mulatta	Rhesus monkey	XM_001092778	Annotated
Macropus eugenii	Tammar wallaby	EU152103	(Duncan *et al.,* 2008)
Marmota monax	Woodchuck (partial cds)	EF621766	(Guy *et al.,* 2008)
Monodelphis domestica	South American grey short tailed opossum	EU152102	(Duncan *et al.,* 2008)
Mus musculus	Mouse	NM_001081110	(Kashiwada *et al.,* 2011)
Mustela putorius furo	Domestic ferret	EF492056	Direct submission
Nomascus leucogenys	Northern white-cheeked gibbon	XM_003268782	Annotated
Oncorhynchus mykiss	Rainbow trout	AF178053	(Hansen and Strassburger, 2000)
Oryctolagus cuniculus	Rabbit	XM_002709594	Annotated
Pan troglodytes	Chimpanzee	XM_001138871	Annotated
Paralichthys olivaceus	Japanese flounder	AB082958	Direct submission
Peromyscus maniculatus	North American deer mouse (partial cds)	EF648004	(Schountz *et al.,* 2007)
Pongo abelii	Sumatran orangutan	XR_092850	Annotated
Pongo pygmaeus	Bornean orangutan	X60223	(Lawlor and Parham, 1992)
Rattus norvegicus	Norway rat	NM_031538	(Parnes *et al.,* 1985)
Saimiri sciureus	Common squirrel monkey	AJ130818	(Ureta-Vidal *et al.,* 1999)
Salmo salar	Atlantic salmon	AY693393	(Moore *et al.,* 2005)
Salmo trutta	Brown trout	AY701523	(Moore *et al.,* 2005)
Sigmodon hispidus	Hispid cotton rat	AY065643	Direct submission
Siniperca chuatsi	Chinese perch	GQ863494	Direct submission
Sparus aurata	Gilthead Seabream	AJ878605	(Randelli *et al.,* 2006)
Sus scrofa	Pig	NM_001001907	(Uenishi *et al.,* 2004)
Tursiops truncatus	Bottlenosed dolphin	EU081776	Direct submission
Xenopus laevis	African clawed frog	HQ116783	(Chida *et al.,* 2011)

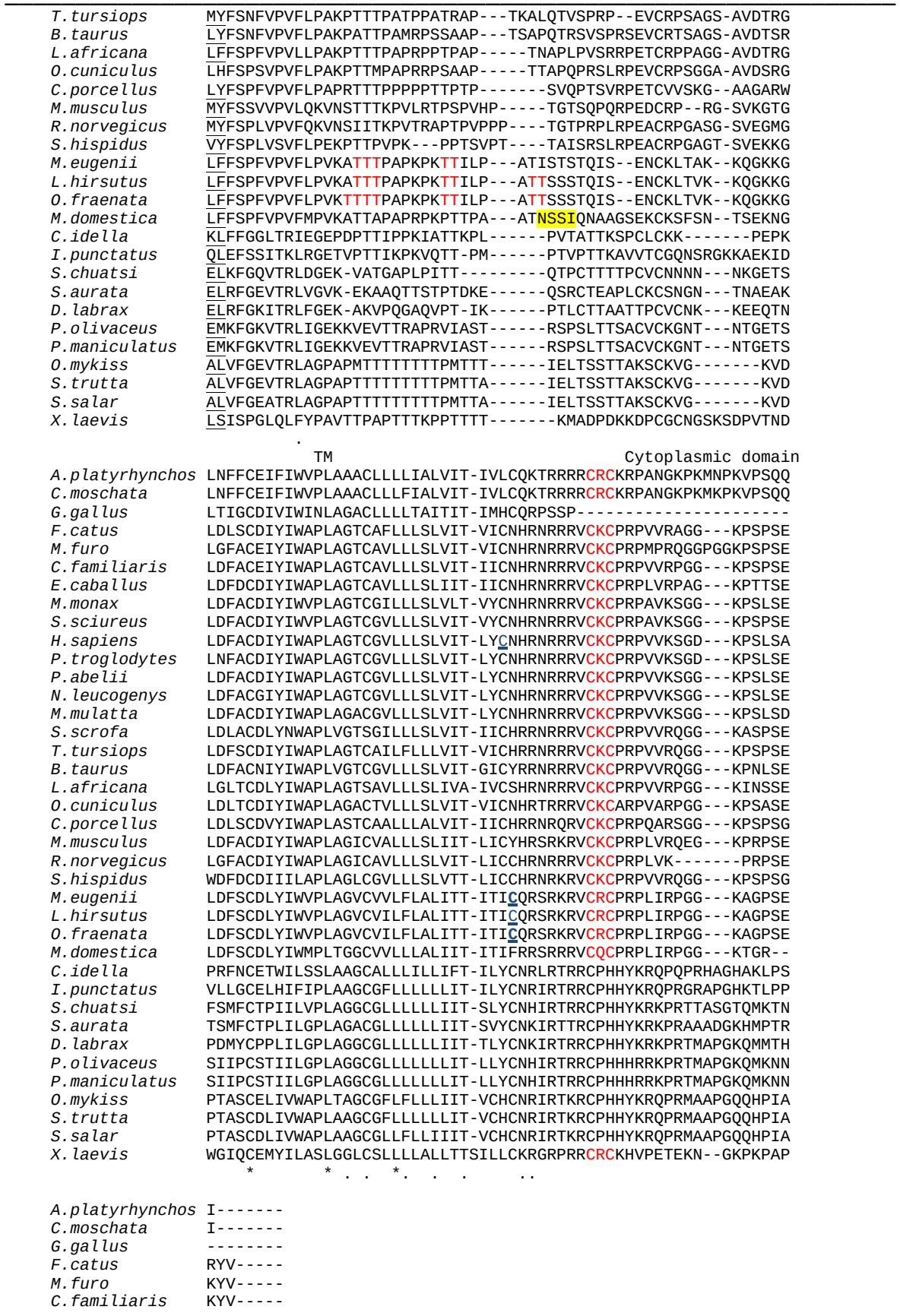
Appendix 4B

Alignment for CD8α chain including marsupial species *O. fraenata* and *L. hirsutus* and the previously by Duncan *et al*. (2008), identified *M. eugenii* and *M. domestica* sequences.

```
                                                                    Signal peptide
A.platyrhynchos -----------------------------------------MTDASPALLLLLSLGLCCP
C.moschata      -----------------------------------------MTDASPALLLLLSLGLCCP
G.gallus        ------------------------------------------MAASPALLLLLSLGLCCT
F.catus         -----------------------------------------MASPVTAQLLPLALLLH-A
M.furo          -----------------------------------------MASRVTPLLLPLALLLH-A
C.familiaris    -----------------------------------------MASRVTALLLPLALLLR-A
E.caballus      -----------------------------------------MAEPGTFLLLPLALLLH-A
M.monax         -----------------------------------------MASPVTALLLPLAMLLH-A
S.sciureus      -----------------------------------------MASPVTALLLPLALLLH-A
H.sapiens       -----------------------------------------MALPVTALLLPLALLLH-A
P.troglodytes   MRNQAPGRPKGATSPPRRPTGSRAPPLAPELRAKQRPGERVMALPVTALLLPLALLLH-A
P.abelii        -----------------------------------------MALPVTALLLPLALLLH-A
N.leucogenys    -----------------------------------------MSLPVTALLLPLALMLH-A
M.mulatta       MRNQAPGRPKGATSPPPLPTGSRAPPVAPELRAEPRPGERVMAPPVTALLLPLVLLLH-A
S.scrofa        -----------------------------------------MASLVTALLLPLVLQLH-P
T.tursiops      -----------------------------------------MASPVTALLLPLGLLLH-A
B.taurus        -----------------------------------------MASLLTALILPLALLLLDA
L.africana      -----------------------------------------MALRVTALLLPLALLLH-A
O.cuniculus     -----------------------------------------MASSATALLLPLTLLLH-F
C.porcellus     ---------------------------------------MAPRGSAWLLLLPVALLLD-A
M.musculus      -----------------------------------------MASPLTRFLSLNLLLLGES
R.norvegicus    -----------------------------------------MASRVICFLSLNLLLLDVI
S.hispidus      -----------------------------------------MAPRVTRFLCLTLLLEFIA
M.eugenii       ---------------------------------------MGSLLAVRSLLLPLALLLQSV
L.hirsutus      ---------------------------------------MGSLLAVRSLLLPLALLLQSV
O.fraenata      ---------------------------------------MGSLLAVRSLLLPLALLLQSV
M.domestica     ---------------------------------------MGWPSAVRSLLLPLALLLEPL
C.idella        -----------------------------------------MYQIYIGFCVSLSLFYGVS
I.punctatus     -----------------------------------------MALILSVLCVVLLGFHGVE
S.chuatsi       -----------------------------------------MDQKWI--LVILVFYQKMT
S.aurata        -----------------------------------------MEQKWIHFLLILAFYQKIT
D.labrax        ----------------------------------------------MIHLIQQRRRLQIT
P.olivaceus     -----------------------------------------MDQKWIQMLVILVFYQIMA
P.maniculatus   -----------------------------------------MDQKWIQMLVILVFYQIMA
O.mykiss        ---------------------------------------MKMVQKWMQTLVLLFFCQETL
S.trutta        ---------------------------------------MKMVRKWMETLVLLFFCQETL
S.salar         ---------------------------------------MKMVQKWMETLVLLFFCQETL
X.laevis        ---------------------------------------------MELFLMALFLCNWIT

                                  IGSf V domain        CDR1
A.platyrhynchos GAQGQKS----MVTAKFYNSKTTHPQSGKPLELECES---SVDSGVSWIRQEKS--GTLH
C.moschata      GAQGQRI----TVTAKFYNSKSTHPQSGKSLELECTN---SQDSGVSWIRQDKA--GTLH
G.gallus        SAQGQRD----AEVVWFVPRNMKNPQEGQRLEMECSPK--NSDSGASWIRQDKD--GKLH
F.catus         AAAAGP------SPFRLSPVR-VEGRLGQRVELQCEVLLSSAAPGCTWLFQKNEPAARPI
M.furo          VAAAGP------SQFRLSPAK-VVGQLGEKVELQCEVLLPSAAPGCSWLLQKNEPAARPV
C.familiaris    AAASGP------SRFRMTPPK-VVGQLHAQVELQCQVLLSTAAPGCSWLYQRNEPAARPV
E.caballus      VLALGS------NAFRMTQPE-GPPQPGKTLNLRCQVLLSNQAPGCSWLYQPPGPAARPV
M.monax         ARP---------SRFRVSPLD-RTWNLGDTVELKCEVLLSNPASGCSWLFQPRGAAASPN
S.sciureus      ARP---------SRFRVSPLD-RTWNLGDKVELKCEVLLSNPSSGCSWLFQKRGAAASPT
H.sapiens       ARP---------SQFRVSPLD-RTWNLGETVELKCQVLLSNPTSGCSWLFQPRGAAASPT
P.troglodytes   ARP---------SQFRVSPLD-RTWNLGETVELKCQVLLSNPTSGCSWLFQPRGAAASPT
P.abelii        ARP---------SQFRVSPLD-RTWNLGETVELKCQVLLSNPTSGCSWLFQPRGAAASPT
N.leucogenys    ARP---------SQFRVSPLD-RTWNLGETVELKCQVLLSNPTSGCSWLFQPRGAAASPT
M.mulatta       ARP---------NQFRVSPLG-RTWNLGETVELKCQVLLSNPTSGCSWLFQPRGTAARPT
S.scrofa        AKVLGS------SLFRTSPEM-VQASLGETVKLRCEVMHSNTLTSCSWLYQKPGAASKPI
T.tursiops      ATVLES------FSFRMSPVR-VQARLGEKVKLHCEVLQSSMTSSCSWLYQKPGDASRPI
B.taurus        AKVLGS------LSFRMSPTQ-KETRLGEKVELQCELLQSGMATGCSWLRHIPGDDPRPT
L.africana      AEAQGS------NQFRMSPKE-VKATLGKPVTLQCEVLLSNAGSGCSWLFQRLDAAASPI
O.cuniculus     AAAFGQ------SQIRMKPKE-VTAILDKPVTLICEVLLSDLGDSCSWVFQRS-AAASPT
C.porcellus     ATAQGA------SQFRMSPRE-LVAQVGTKVTLRCEVLVPNAPAGCSWLFQPRHDAKGPT
M.musculus      IILGSGEAKPQAPELRIFPKK-MDAELGQKVDLVCEVLGSVSQGCSWLFQNSSSKLPQPT
R.norvegicus    TRLQVS------GQLQLSPKK-VDAEIGQEVKLTCEVLRDTSQGCSWLFRNSSSELLQPT
S.hispidus      ELGGSK-------DFEMSPKK-VVAHLGKEVRLTCEVWVSTSQGCSWLFLEHGS-GVKPT
M.eugenii       GSQAQLK-------FRMNPLEKRDVRPSDKVQLQCETLSSSS-TGCSWLRLAPG-KVVPT
L.hirsutus      GSQAQLK-------FRMNPLEKRDVRPSDKVQLQCDTLSSSP-TGCSWLRLAPG-KVVPT
```

```
O.fraenata        GSQAQVK-------FRMNPLEKRDVRPSDKVQLQCETLSSSP-TGCSWLRLVPG-KVVPT
M.domestica       GAQEHVT-------FRMNPVEKRDAGQGRELKLQCETVNSSP-TGCSWLRLTPG-AVVPT
C.idella          ---------------------ARIYKEGEKVQVDCDLKQ--SGTLTFWFRINSK-GADYL
I.punctatus       GNVK-----------------ELVIKEKQPATITCDKNY--P-STVFWLRLKEN-GQGFE
S.chuatsi         PGAGV----------------NVTAKEGAQVEISCQPRE--TGSMVVWFRVLDQSGMEFI
S.aurata          SGTDE----------------VKAVTEGDNAEIKCHPSD--PGSMIIWFRVRDKSGMEFI
D.labrax          SGAGE----------------DKATTEGQLVEIHCQSG---TGTMIIWFRVLDKTGMEFI
P.olivaceus       SGAG-----------------ELVVKEGAKVDIECKPAE--MFNTVIWFRVLDNSGMEFI
P.maniculatus     SGAG-----------------ELVVKEGAKVDIECKPAE--MFNTVIWFRVLDNSGMEFI
O.mykiss          QLSS-----------------LTEKTDGERVEITCAPVSKTKSNMVIWFRVQDNAGMEFI
S.trutta          QLSS-----------------LTEKTDGERVEITCAPVSKTKSNMVIWFRVQDNAGMEFI
S.salar           QLSS-----------------LTEKTDGKRVEITCAPVS-IKSNMVMWFRVQDNAGMEFI
X.laevis          GTQQLG---------LTQTSGSLTKEKSETVKLECKPGPKESMDNAVFWFRQRKDTKTPE
                                                    . *
                  ◀DE-loop
                          ▼ CDR2-CD8α binding site to MHC I                  CDR3
A.platyrhynchos   FIVYISTLSKPTFKENQMMPSHFG----TFKSGKSYRLTVKSFKAEDEGTYFCTVNYNQQ
C.moschata        FIVYISALSKPTFEGNQMTSSRYG----VSKSGRSYRLTVTSFKAQDEGIYFCTVNYNQV
G.gallus          FIVYISVLSRTTYSGNENTSPNFE----ASQKGNSYRLVVKNFRAQDQGTYFCIANINQV
F.catus           FLAYLSRSR--TKLAEELDPKQIS---GQRIQDTLYSLTLHRFRKEEEGYYFCSVVSNSV
M.furo            FLMYLSQTR--TKLADGLDSEQIS---GKKIRDTLYSLTLRRFRKEDEGYYFCSVLSNSV
C.familiaris      FLMYISQSR--AKPAEGLDTKHIS---GQKKTDSTYSLTLSRFRKEDEGYYFCSVLSNSI
E.caballus        FLMYITKGR--IKTAEDLDTKKFS---GERIQDAVFGLTLHHFSEKDQGYYFCSVLSNSI
M.monax           FLLYISQTK--SKVADGLDTQRYS---GKK-MGDSFILTLSDFREENQGYYFCSALRNSI
S.sciureus        FLLYISQTK--PKVADGLDAQRFS---GKK-MGDSFILTLRDFREEDQGFYFCSALSNSI
H.sapiens         FLLYLSQNK--PKAAEGLDTQRFS---GKR-LGDTFVLTLSDFRRENEGYYFCSALSNSI
P.troglodytes     FLLYLSQNK--PKAAEGLDTQRFS---GKR-LGDTFVLTLSDFRRENEGYYFCSALSNSI
P.abelii          FLLYLSQNK--PKAAEGLDTQRFS---GKR-LGDTFVLTLSDFRRENEGYYFCSALSNSI
N.leucogenys      FLLYLSQNK--PKAAEGLDTQRFS---GKR-LGDTFVLTLSDFRRENEGYYFCSALSNSI
M.mulatta         FLLYLSQNK--PKAAEGLDTQRFS---GKR-LGDTFVLTLRDFRQENEGYYFCSALSNSI
S.scrofa          FLMYLSKTR--NKTAEGLDTRYIS---GYKAN-DNFYLILHRFREEDQGYYFCSFLSNSV
T.tursiops        FLMYLSSTR--SKPAEGLDTSYIS---GTKAEGANFHLTLHRFHEEHQGYYFCSFMSNSV
B.taurus          FLMYLSAQR--VKLAEGLDPRHIS---GAKVSGTKFQLTLSSFLQEDQGYYFCSVVSNSI
L.africana        FLLYISGTG--IKRVQGQDSTRF----SGAKTSSGFQLTLNHFQEKDQGYYFCSVLSNSA
O.cuniculus       FLLYLSKT-----RNQVDSPLNS----GKRVQEKVFSLTLHHFREEDQGYYFCVVVGSLS
C.porcellus       FLLYHSASG--TKLAPGLEQKRFS---PSK-SSNTYTLTVNSFQKRDEGYYFCSVSGNMM
M.musculus        FVVYMASSH--NKITWDEKLNSSKLFSAMRDTNNKYVLTLNKFSKENEGYYFCSVISNSV
R.norvegicus      FIIYVSSSR--SKLNDILDPN----LFSARKENNKYILTLSKFSTKNQGYYFCSITSNSV
S.hispidus        FLIYLSGSR--NERNNKIPST----KLSGKKEDKKYTLTLNNFAKEDEGYYFCSVTSNSV
M.eugenii         FLLFISSTSSNAKLAEDLDPKRFR---GERISSSTYRLTLQNFREEDEGYYYCVVTRNSA
L.hirsutus        FLLFISSSSSNAKLAEDLDPKRFR---GERISSSTYRLTLQNFREEDEGYYYCVVTRNSA
O.fraenata        FLLFISSTSLNAKLAEGLDPKRFR---GERISSSTYRLTLQNFREEDQGYYYCVVTRNSA
M.domestica       FLLYLSGSSQSVKVAPELDSRRFG---GSRSSS-SYFLTLKDFQKDDQGYYYCVVARNSR
C.idella          FSVKSADIKENDLNSEHYTV-NTN-------SGKVQLDIKSFKKKTDSGVYVCAAMN-SN
I.punctatus       YIASYSKTKKSGKVAEGG---NYK-------VAEKTFDINSFETQKDSGTYSCVFIN-NN
S.chuatsi         ASFSNNGIRKSPTTSLSLIFSEAK-------IGQHILTLQSFDKARDSGVYGCASLYKGI
S.aurata          ASFSSNGMPKPNTKSPSSTFIDSK-------IGQNILILQSFKEAVDSGVYSCATLYKGT
D.labrax          GSFSNNGVLKS--TSLSNIYRQTK-------INQNILILQSFNKSRDSGIYSCASLYKGN
P.olivaceus       ASFGRDGKMKSNPSPLSPYIDSSK-------VDKHILTLKSFSKARDSGTYSCTIIQ-SN
P.maniculatus     ASFGRDGKMKSNPSPLSPYIDSSK-------VDKHILTLKSFSKARDSGTYSCTIIQ-SN
O.mykiss          ASFSTKDGMKK-TDFNNEVFSEEQ-------INKNILILKAFKKARDSGVYSCASIN-GN
S.trutta          ASFSTKDGTKK-TDFNNEVFSEEQ-------INKNILILKAFKKARDSGVYSCASIN-GN
S.salar           ASFSTKDGMKK-TDFNNEVFSEEQ-------INKNILILKAFKKARDSGVYSCASIN-GN
X.laevis          SILYLSSVSKQKLSDDNVFKHFTAN-----KGAFTFTLNIFPFQEKDEGTYYCMININSV
                                                                  ..* * *
                      Hinge region
A.platyrhynchos   LYFSSGLPAFFPVTTTAAPIKTVPPTTQ--------GSQVTKRDVCLQSHEAE-TSKEKE
C.moschata        LYFSSGLPAFFPVTTTAAPVTSVSPTTQ--------GSRVTKRDVCLQSHEAE-TSKEKK
G.gallus          LHFSSGQSAFFPVTTEAP--TTPAATTQ--------SHQVTKKDTSHQSLHPG-TSSENT
F.catus           LYFSAFVPVFLPVKPTTTPAPRPPTQAP-----ITTSQRVSLRPGTCQPSAGS-TVEASG
M.furo            LYFSSFVPVFLPVKPTPTPAPRLPTRAP-----TNTSQPVSQRPGICRPPAGK-PVEKGV
C.familiaris      LYFSPFVPVFLPVKPPTTPAPRPPTRAP-----TNASKPVSPRGETCRPAAGS-AVKTSG
E.caballus        IYFSPFVPVFLPAKPTTTPAPRPPTPRPPMRAATNASQPVTRRPETCRPAKGS-PVGKKW
M.monax           MYFSSFVPVFLPAKPTTTPAPRPPTPEP-----TTASQPLSLRPQACRPAAGV-AGDTRG
S.sciureus        MYFSPFVPVFLPAKPTTTPAPRPPTPEP-----TTASQPLSLRPQACRPPAGG-AVDTRG
H.sapiens         MYFSHFVPVFLPAKPTTTPAPRPPTPAP-----TIASQPLSLRPEACRPAAGG-AVHTRG
P.troglodytes     MYFSHFVPVFLPAKPTTTPAPRPPTPAP-----TIASQPLSLRPEACRPAAGG-AVHTRG
P.abelii          MYFSHFVPVFLP----------------------------------------VHTRG
N.leucogenys      MYFSHFVPVFLPEKPTTTPAPRPPTPAA-----TTASQPLSLRPEACRPAAGG-AVHTRG
M.mulatta         MYFSHFVPVFLPAKPTTTPAPRSPTPAP-----TTASQPLSLRPEACRPAAGG-SVNTRG
S.scrofa          LYFSNFMSVFLPAKPTKTPTTPPPKRTP---TKASHAVSVAP--EVCRPSG---NADPRK
```

Appendix 4B


T.tursiops MYFSNFVPVFLPAKPTTTPATPPATRAP---TKALQTVSPRP--EVCRPSAGS-AVDTRG
B.taurus LYFSNFVPVFLPAKPATTPAMRPSSAAP---TSAPQTRSVSPRSEVCRTSAGS-AVDTSR
L.africana LFFSPFVPVLLPAKPTTTPAPRPPTPAP-----TNAPLPVSRRPETCRPPAGG-AVDTRG
O.cuniculus LHFSPSVPVFLPAKPTTMPAPRRPSAAP-----TTAPQPRSLRPEVCRPSGGA-AVDSRG
C.porcellus LYFSPFVPVFLPAPRTTTPPPPPTTPTP-------SVQPTSVRPETCVVSKG--AAGARW
M.musculus MYFSSVVPVLQKVNSTTTKPVLRTPSPVHP-----TGTSQPQRPEDCRP--RG-SVKGTG
R.norvegicus MYFSPLVPVFQKVNSIITKPVTRAPTPVPPP----TGTPRPLRPEACRPGASG-SVEGMG
S.hispidus VYFSPLVSVFLPEKPTTPVPK---PPTSVPT----TAISRSLRPEACRPGAGT-SVEKKG
M.eugenii LFFSPFVPVFLPVKATTTPAPKPKTTILP---ATISTSTQIS--ENCKLTAK--KQGKKG
L.hirsutus LFFSPFVPVFLPVKATTTPAPKPKTTILP---ATTSSSTQIS--ENCKLTVK--KQGKKG
O.fraenata LFFSPFVPVFLPVKTTTTPAPKPKTTILP---ATTSSSTQIS--ENCKLTVK--KQGKKG
M.domestica LFFSPFVPVFMPVKATTAPAPRPKPTTPA---ATNSSIQNAAGSEKCKSFSN--TSEKNG
C.idella KLFFGGLTRIEGEPDPTTIPPKIATTKPL------PVTATTKSPCLCKK-------PEPK
I.punctatus QLEFSSITKLRGETVPTTIKPKVQTT-PM------PTVPTTKAVVTCGQNSRGKKAEKID
S.chuatsi ELKFGQVTRLDGEK-VATGAPLPITT---------QTPCTTTTPCVCNNNN---NKGETS
S.aurata ELRFGEVTRLVGVK-EKAAQTTSTPTDKE------QSRCTEAPLCKCSNGN---TNAEAK
D.labrax ELRFGKITRLFGEK-AKVPQGAQVPT-IK------PTLCTTAATTPCVCNK---KEEQTN
P.olivaceus EMKFGKVTRLIGEKKVEVTTRAPRVIAST------RSPSLTTSACVCKGNT---NTGETS
P.maniculatus EMKFGKVTRLIGEKKVEVTTRAPRVIAST------RSPSLTTSACVCKGNT---NTGETS
O.mykiss ALVFGEVTRLAGPAPMTTTTTTTTPMTTT------IELTSSTTAKSCKVG-------KVD
S.trutta ALVFGEVTRLAGPAPTTTTTTTTTPMTTA------IELTSSTTAKSCKVG-------KVD
S.salar ALVFGEATRLAGPAPTTTTTTTTTPMTTA------IELTSSTTAKSCKVG-------KVD
X.laevis LSISPGLQLFYPAVTTPAPTTTKPPTTTT-------KMADPDKKDPCGCNGSKSDPVTND
TM
Cytoplasmic domain
A.platyrhynchos LNFFCEIFIWVPLAAACLLLLIALVIT-IVLCQKTRRRRCRCKRPANGKPKMNPKVPSQQ
C.moschata LNFFCEIFIWVPLAAACLLLFIALVIT-IVLCQKTRRRRCRCKRPANGKPKMKPKVPSQQ
G.gallus LTIGCDIVIWINLAGACLLLLTAITIT-IMHCQRPSSP----------
F.catus LDLSCDIYIWAPLAGTCAFLLLSLVIT-VICNHRNRRRVCKCPRPVVRAGG---KPSPSE
M.furo LGFACEIYIWAPLAGTCAVLLLSLVIT-VICNHRNRRRVCKCPRPMPRQGGPGGKPSPSE
C.familiaris LDFACEIYIWAPLAGTCAVLLLSLVIT-IICNHRNRRRVCKCPRPVVRPGG---KPSPSE
E.caballus LDFDCDIYIWAPLAGTCAVLLLSLIIT-IICNHRNRRRVCKCPRPLVRPAG---KPTTSE
M.monax LDFACDIYIWVPLAGTCGILLLSLVLT-VYCNHRNRRRVCKCPRPAVKSGG---KPSLSE
S.sciureus LDFACDIYIWVPLAGTCGVLLLSLVIT-VYCNHRNRRRVCKCPRPAVKSGG---KPSPSE
H.sapiens LDFACDIYIWAPLAGTCGVLLLSLVIT-LYCNHRNRRRVCKCPRPVVKSGD---KPSLSA
P.troglodytes LNFACDIYIWAPLAGTCGVLLLSLVIT-LYCNHRNRRRVCKCPRPVVKSGD---KPSLSE
P.abelii LDFACDIYIWAPLAGTCGVLLLSLVIT-LYCNHRNRRRVCKCPRPVVKSGG---KPSLSE
N.leucogenys LDFACGIYIWAPLAGTCGVLLLSLVIT-LYCNHRNRRRVCKCPRPVVKSGG---KPSLSE
M.mulatta LDFACDIYIWAPLAGACGVLLLSLVIT-LYCNHRNRRRVCKCPRPVVKSGG---KPSLSD
S.scrofa LDLACDLYNWAPLVGTSGILLLSLVIT-IICHRRNRRRVCKCPRPVVRQGG---KASPSE
T.tursiops LDFSCDIYIWAPLAGTCAILFLLLVIT-VICHRRNRRRVCKCPRPVVRQGG---KPSPSE
B.taurus LDFACNIYIWAPLVGTCGVLLLSLVIT-GICYRRNRRRVCKCPRPVVRQGG---KPNLSE
L.africana LGLTCDLYIWAPLAGTSAVLLLSLIVA-IVCSHRNRRRVCKCPRPVVRPGG---KINSSE
O.cuniculus LDLTCDIYIWAPLAGACTVLLLSLVIT-VICNHRTRRRVCKCARPVARPGG---KPSASE
C.porcellus LDLSCDVYIWAPLASTCAALLLALVIT-IICHRRNRQRVCKCPRPQARSGG---KPSPSG
M.musculus LDFACDIYIWAPLAGICVALLLSLIIT-LICYHRSRKRVCKCPRPLVRQEG---KPRPSE
R.norvegicus LGFACDIYIWAPLAGICAVLLLSLVIT-LICCHRNRRRVCKCPRPLVK-------PRPSE
S.hispidus WDFDCDIIILAPLAGLCGVLLLSLVTT-LICCHRNRKRVCKCPRPVVRQGG---KPSPSG
M.eugenii LDFSCDLYIWVPLAGVCVVLFLALITT-ITICQRSRKRVCRCPRPLIRPGG---KAGPSE
L.hirsutus LDFSCDLYIWVPLAGVCVILFLALITT-ITICQRSRKRVCRCPRPLIRPGG---KAGPSE
O.fraenata LDFSCDLYIWVPLAGVCVILFLALITT-ITICQRSRKRVCRCPRPLIRPGG---KAGPSE
M.domestica LDFSCDLYIWMPLTGGCVVLLLALIIT-ITIFRRSRRRVCQCPRPLIRPGG---KTGR--
C.idella PRFNCETWILSSLAAGCALLLILLIFT-ILYCNRLRTRRCPHHYKRQPQPRHAGHAKLPS
I.punctatus VLLGCELHIFIPLAAGCGFLLLLLLIT-ILYCNRIRTRRCPHHYKRQPRGRAPGHKTLPP
S.chuatsi FSMFCTPIILVPLAGGCGLLLLLLIIT-SLYCNHIRTRRCPHHYKRKPRTTASGTQMKTN
S.aurata TSMFCTPLILGPLAGACGLLLLLLIIT-SVYCNKIRTTRCPHHYKRKPRAAADGKHMPTR
D.labrax PDMYCPPLILGPLAGGCGLLLLLLIIT-TLYCNKIRTRRCPHHYKRKPRTMAPGKQMMTH
P.olivaceus SIIPCSTIILGPLAGGCGLLLLLLLIT-LLYCNHIRTRRCPHHHRRKPRTMAPGKQMKNN
P.maniculatus SIIPCSTIILGPLAGGCGLLLLLLLIT-LLYCNHIRTRRCPHHHRRKPRTMAPGKQMKNN
O.mykiss PTASCELIVWAPLTAGCGFLFLLLIIT-VCHCNRIRTKRCPHHYKRQPRMAAPGQQHPIA
S.trutta PTASCDLIVWAPLAAGCGFLLLLLLIT-VCHCNRIRTKRCPHHYKRQPRMAAPGQQHPIA
S.salar PTASCDLIVWAPLAAGCGLLFLLIIIT-VCHCNRIRTKRCPHHYKRQPRMAAPGQQHPIA
X.laevis WGIQCEMYILASLGGLCSLLLLALLTTSILLCKRGRPRRCRCKHVPETEKN--GKPKPAP
I-------
I-------

RYV-----
KYV-----
KYV-----

```
E.caballus      RYV-----
M.monax         RCV-----
S.sciureus      RYV-----
H.sapiens       RYV-----
P.troglodytes   RYV-----
P.abelii        RYV-----
N.leucogenys    RYV-----
M.mulatta       RYV-----
S.scrofa        RFI-----
T.tursiops      RCT-----
B.taurus        KYV-----
L.africana      GYI-----
O.cuniculus     RTI-----
C.porcellus     KLV-----
M.musculus      KIV-----
R.norvegicus    KFV-----
S.hispidus      KLV-----
M.eugenii       RYA-----
L.hirsutus      RYA-----
O.fraenata      RYA-----
M.domestica     --------
C.idella        NHF-----
I.punctatus     PDY-----
S.chuatsi       RHI-----
S.aurata        Q-------
D.labrax        RHV-----
P.olivaceus     THV-----
P.maniculatus   THV-----
O.mykiss        NNRLF---
S.trutta        NNRLF---
S.salar         NNRLF---
X.laevis        RYV-----
```

Fig. 4B.4. Sequence alignment of all to date known CD8α sequences. Alignment used for phylogenetic analysis of the CD8α and CD8β chains. **T** = O-linked glycosylation sites. Yellow highlight = N-linked glycosylation site in *M. domestica*. Double underlined is a site that indicates a short helix. Underlined are the areas that contact the MHC class I molecule. Bold underlined is the signal peptide from position 1 to 21.

O. fraenata nucleotide sequence for mature protein of CD8β

```
aggtattctatagggcgattgatttagcggccgcgaattcgcccttatggctcagcctcttgcccattc
agatctcaggcatctctggagtcccttctgtcattcagagcccagaaaccctcctggttcagactgaca
aagaggcaaagttcctctgtgacatgaggtcatcctcaagtacttacagaatctactggtttcgtcagg
tagcgcctcccagccctgacagtcactatcagtttctggtccatttggactccaagaccacttatgggg
aaggcattgatccaaaacatgtgcttacgtccagagagtcatacagatccacactcaggatgctgaatg
tgaaaccttcagacagtggcatctatatctgcggcatttttgaaagctatcaactggtctttggaagcg
gaacccggctgaatgtggttgatgttttgcctacctcccctataccaccaaaaagaccacccccagga
agagaccatgcaatacgaaacgctcagaagtcactcaacagaatggtttcttctgcagtgccctccccc
taagcctgctggtgggatgtgctgtggtgctgctcattcccctaattgtgatcatccgaatgaattact
tgtggaatgtagctcggcaccacgtcgtgaaatag
```

O. fraenata putative mature protein sequence

```
LWLSLLPIQISGISGVPSVIQSPETLLVQTDKEAKFLCDMRSSSSTYRIYWFRQVAPPSPDSHYQFLVH
LDSKTTYGEGIDPKHVLTSRESYRSTLRMLNVKPSDSGIYICGIFESYQLVFGSGTRLNVVDVLPTSPI
PTKKTTPRKRPCNTKRSEVTQQNGFFCSALPLSLLVGCAVVLLIPLIVIIRMNYLWNVARHHVVK
```

Secondary structure prediction CD8β (*O. fraenata*)

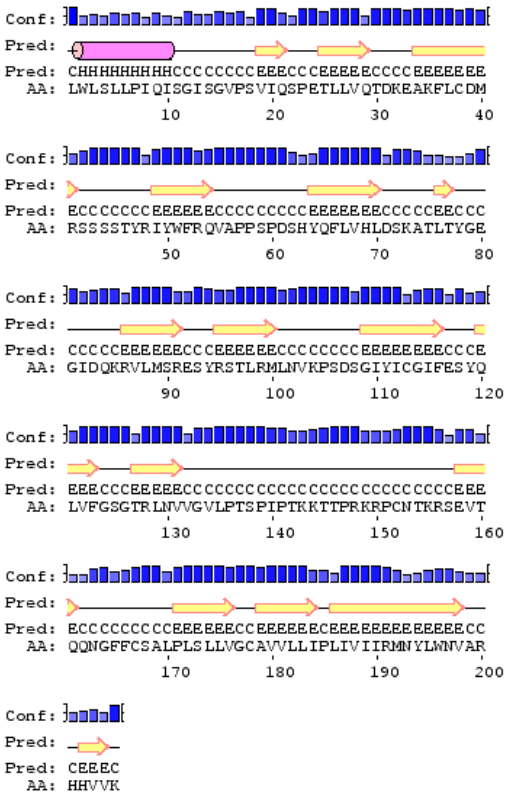


Figure 4B.5. Secondary structure prediction for *O. fraenata* CD8β chain.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

L. hirsutus nucleotide sequence for mature protein of CD8β

```
ttatggctcagcctcttgcccattcagatctcagagatctctggagtcccctctgtcatgcagagccca
gaaaccctcctggttcagactgaccaagaggcaaagttcctctgtgacctgaggtcatcctcaagtact
tacagaatctactggtttcgtcaggtagcgcctcctagccctgacagtcactatcagtttctggtccat
ttggactccaaggccactctcacttatggggaaggcattgatcaaaaacgtgtccttatgtccagagag
tcatacagatccacactcaggatgctgaatgtgaagccttcagacagtggcatctatatctgtggcatt
tttgaaagctatcaactggtctttggaagcggaacccggctgaatgtggttgatgttttgcctacctcc
cctatacccaccaaaaagaccacccccaggaagagaccatgcagtacgaaacgctcagaagtcactcaa
cagaatggtttcttctgcagtgccctcccccctaagcctgctggtgggatgtgctgtggttctgctcatt
cccctaattgtgatcatccgaatgaattacttgtggaatgtagctcggcaccacgtcgtgaaatag
```

L. hirsutus putative protein sequence

```
LWLSLLPIQISGISGVPSVIQSPETLLVQTDKEAKFLCDMRSSSSTYRIYWFRQVAPPSPDSHYQFLVH
LDSKATLTYGEGIDQKRVLMSRESYRSTLRMLNVKPSDSGIYICGIFESYQLVFGSGTRLNVVGVLPTS
PIPTKKTTPRKRPCNTKRSEVTQQNGFFCSALPLSLLVGCAVVLLIPLIVIIRMNYLWNVARHHVVK
```

Secondary structure prediction for *L. hirsutus* CD8β

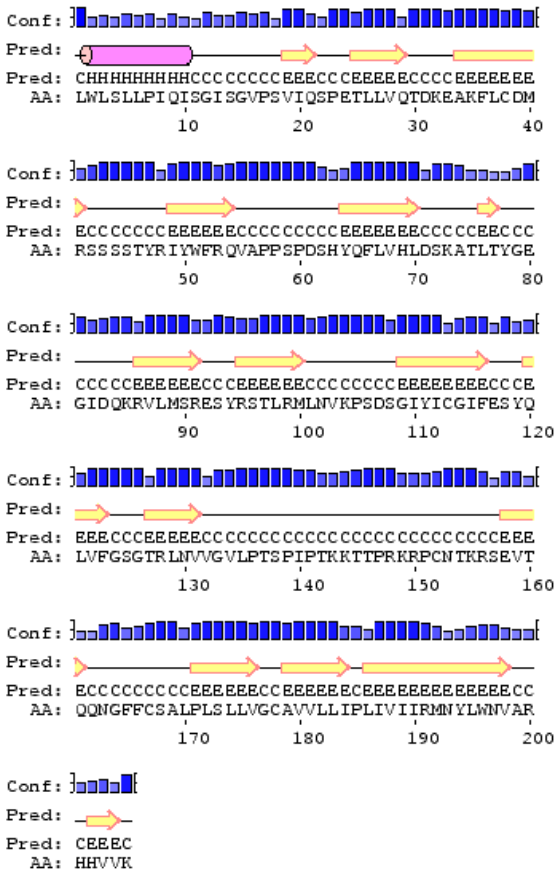


Figure 4B.6. Secondary structure prediction CD8β for *L. hirsutus*.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

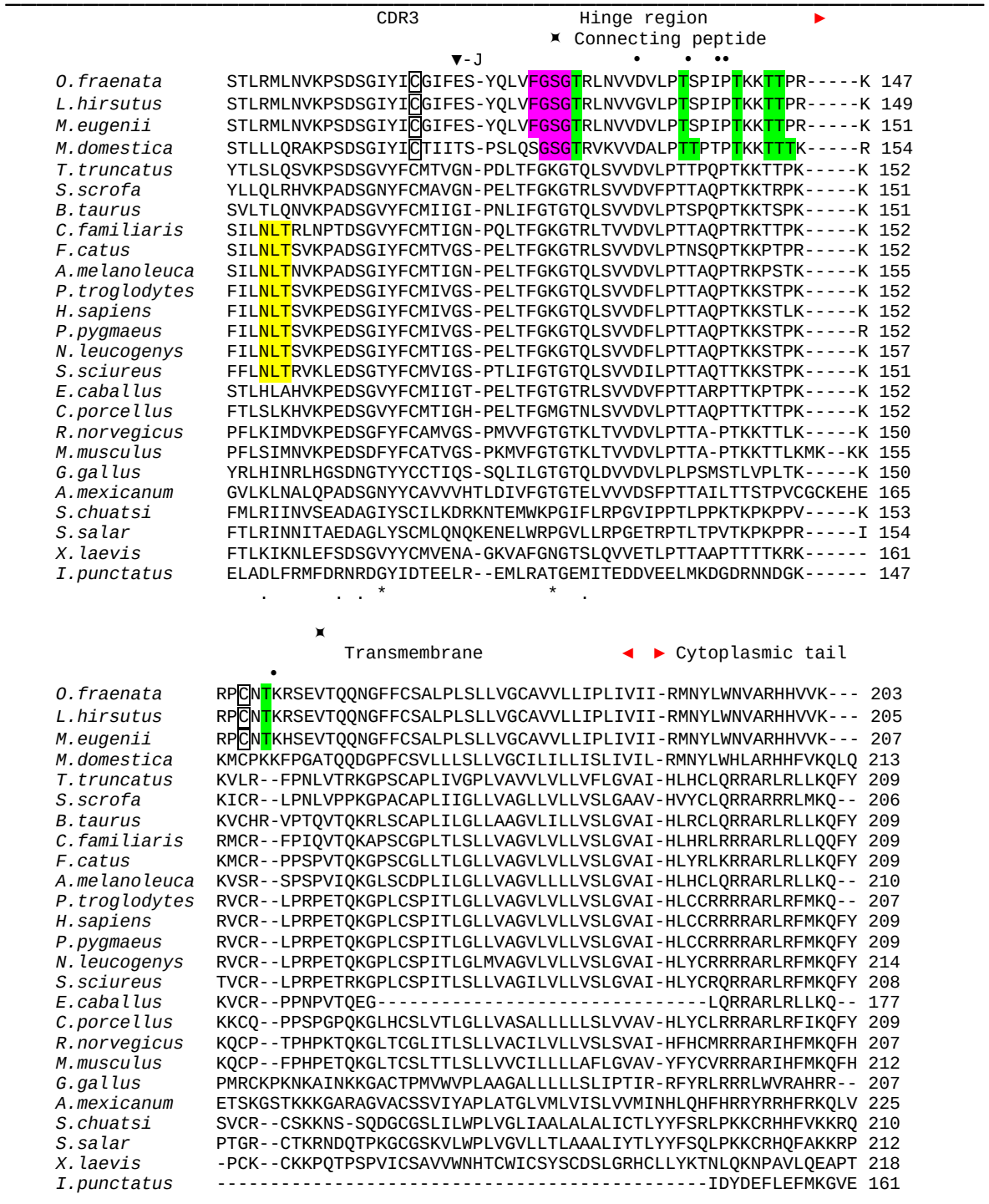
.

Appendix 4B

Alignment of known CD8β amino acid sequences

```
                                                  Mature peptide
                                                     ▼ variable region   CDR1
O.fraenata      ---------L----WLSL-LPIQIS----GISGVPSVIQSPETLLVQTDKEAKFLCDMRS 42
L.hirsutus      ---------L----WLSL-LPIQIS----GISGVPSVIQSPETLLVQTDKEAKFLCDMRS 42
M.eugenii       ------MQLL----WLSLFLPIQIS----GISGVPSVMQSPETLLVQTDQEAKFLCDMRS 46
M.domestica     ------MELL----WLSVFLSGQIS----GIFGVPALSQSPETLLVQTDKQARLICDIKL 46
T.truncatus     ------MQPG-----LWLLVAAQLA----ALRGSSVLLQTPASTTAQTNQTVTLSCEAKT 45
S.scrofa        ------MQPR-----LWLLIAAQLS----TLHGGSALLQTPSSVMAQTNQTVKLSCEART 45
B.taurus        ------MQAR-----LWLLLAAQLA----ALHGSLALVQTPAFLMVQTNQMVTLSCKTQT 45
C.familiaris    ------MQPG-----LWLLLAAQLA----ALRGSSVLQQVPVSIMVQTNQMVTMSCEVRT 45
F.catus         ------MQPG-----LWLLLATQLA----ALRGSSVLQQAPGSVMVQTNGMVIMSCEAKT 45
A.melanoleuca   ------MTLE-----LKDSSGWRFGNR-LALRGSSPLQQVPASIVIQTKQMVTMSCEART 48
P.troglodytes   ------MRPR-----LWLLLAAQLT----VLHGSSVLQQTPAYIKVQTNKMVMLSCEAKI 45
H.sapiens       ------MRPR-----LWLLLAAQLT----VLHGNSVLQQTPAYIKVQTNKMVMLSCEAKI 45
P.pygmaeus      ------MRPR-----LWLLLAAQLA----VLHGSSVLQQTPAYIKVQTNKMVMLSCEAKI 45
N.leucogenys    ------MGPRSLSHECRQDFKHEVL----ILHGNSVLQQTPAYIKVQTNKAVMLSCEAKI 50
S.sciureus      ------MRPR-----MWLLLSAQLA----ALHGNSVLQQTPAYIMVQTNQMVMLSCKA-I 44
E.caballus      ------MQPR-----LWLLLAAQLA----ALHGSSVLQQNPEYQMVQTNQMVVLTCEART 45
C.porcellus     ------MQPR-----LWLLVAAKLA----ALRGICGLQQSPHFLMLQTNQSATMTCESKT 45
R.norvegicus    ------MQPW-----LWLVFSVKLS----ALWGSSALLQTPSSLLVQTNQTAKMSCEAKT 45
M.musculus      ------MQPW-----LWLVFSMKLA----ALWSSSALIQTPSSLLVQTNHTAKMSCEVKS 45
G.gallus        ------MARP----WLWLWLCLQLP----GFCTNLLSSQTPGYILTKTNNSTEIVCPMKG 46
A.mexicanum     ------MQESQRGRIPDPPTPQDLVKQDIGVRSSSLLPQSPTSIIALTNNKTEIQCAVRE 54
S.chuatsi       ------MIPLP---LVWTLLTVSLWT---SGSSQILQQETARVLYPKILSTEVIECDCVN 48
S.salar         ------MTPLA---LVWT--TVCLWK---TVYS-LTPTESHFVRYPPINDTEVVTCECSS 45
X.laevis        MKHPSDMSPHS---CCFLLLIIIIS---FWGKQVSRTRTNVKVTTCTAQRNTEMFCTMKN 54
I.punctatus     --MDD----------IYKAAVEQLT------EEQKNEFKAAFEVFVQDAEDGCISTKELG 42
                                          .                          .
                          IgG variable-like region    CDR2
O.fraenata      SSSTYR-IYWFRQVAPPSPDS-HYQFLVHLDSET--TYGEGIDPKHVLTS-----RESYR 93
L.hirsutus      SSSTYR-IYWFRQVAPPSPDS-HYQFLVHLDSKATLTYGEGIDQKRVLMS-----RESYR 95
M.eugenii       SSSIYR-IYWFRQVAPPSPDS-HYQFLVHLDSKA--TYGEGIDQKHVLIS-----KESHR 97
M.domestica     PSINYG-IYWFRLIGSPHSDSLSYEFLLQSEARGSCTYGKGVNSTHVTAS-----RESSK 100
T.truncatus     SPTNSR-IYWLRQRLAPSANS-HFEFLAFWDLTRGTVYGEEVGQERLTVL-----RDSSR 98
S.scrofa        FPTNTR-IYWLRLRQALSANS-HYEFLALWIPNGNPVYGKET-EKQLTVL-----QDSSR 97
B.taurus        SPSNTR-IYWLRLRQALSANS-HYEFLAYWE-SKKTVYGKEVDSEKLTVH-----GNPPQ 97
C.familiaris    SPTSTR-IYWLRQRQAPSPDS-HHEFLAFWDPKKGIVYDQMAEQEKLTVF-----PGTTR 98
F.catus         SPTSTR-IYWLRHRQAPSPDS-HYECLAYWDPIKGIVYGQEVEPEKLTVF-----PDATR 98
A.melanoleuca   SPTSTR-IYWLRQRQAPSSDS-HHEFLAFWDSIKGIVYGQEVDQEKLTVF-----PEATR 101
P.troglodytes   SLSNMR-IYWLRQRQAPSSDS-HHEFLALWDSAKGTIHGEEVEQEKIAVF-----RDASR 98
H.sapiens       SLSNMR-IYWLRQRQAPSSDS-HHEFLALWDSAKGTIHGEEVEQEKIAVF-----RDASR 98
P.pygmaeus      SLSNMR-IYWLRQRQAPSSDS-HHEFLALWDSAKGTIHSEEVEQEKVAVF-----RDASR 98
N.leucogenys    SLSNMR-IYWLRQRQAPSSDS-HHEFLALWDSAKGTIHGEEVEQEKIAVF-----RDGSR 103
S.sciureus      SSSTTR-IYWLRQLHAPSSNS-HHEILAFWDSSKGTIHSEGVEQKKITVF-----RDGSL 97
E.caballus      TYPNKR-IYWLRQRQAPSADS-HQEFLAAWDSIKGIVYGEKLDQEKLTLS-----QDASR 98
C.porcellus     SSINGR-LYWLRQRGAPSPDS-HHEFLASGDFSNNIVYGRGMTSEMLTLS-----RLSTR 98
R.norvegicus    FPKGTT-IYWLRELQDSNKNK-HFEFLASRTSTKGIKYGERV--KKNMTL----SFNSTL 97
M.musculus      ISKLTS-IYWLRERQDP-KDK-YFEFLASWSSSKGVLYGESVDKKRNIILE---SSDSRR 99
G.gallus        EHTG---VYWYR----WNQGRQHFEFLLFSSPLGKATYGTNISQEKFSIRG---TSSYHS 96
A.mexicanum     QHMDRLGIYWYR------HHRGDFQFILYLSILDKSTYGSSISKESFTASR---DSFRRM 105
S.chuatsi       ISCDTV--YWFR----SISNHSKVQFLGKCNNADRATYGEG-VETARFKLSRR---SSMS 98
R.norvegicus    FPKGTT-IYWLRELQDSNKNK-HFEFLASRTSTKGIKYGERV--KKNMTL----SFNSTL 97
S.salar         RSCQTV--FWFR----THLNNSGFQFLLSLNNADRTYYGPGLMDEHRFKASKRDTGSKVA 99
X.laevis        QNIDQIGVYWYKQ---TSTTADKLEFVVFSHILNKIRYGALFSSERVSVAR---ENFRNT 108
I.punctatus     KVMRMLGQNPTPQELQEMIDEVDEDSSGTVDFDEFLVMMVRCMKDDSRAK-------SEE 95
                                          .
```

Appendix 4B


CDR3
Hinge region
Connecting peptide
-J
Transmembrane
Cytoplasmic tail
O.fraenata STLRMLNVKPSDSGIYICGIFES-YQLVFGSGTRLNVVDVLPTSPIPTKKTTPR-----K 147
L.hirsutus STLRMLNVKPSDSGIYICGIFES-YQLVFGSGTRLNVVGVLPTSPIPTKKTTPR-----K 149
M.eugenii STLRMLNVKPSDSGIYICGIFES-YQLVFGSGTRLNVVDVLPTSPIPTKKTTPR-----K 151
M.domestica STLLLQRAKPSDSGIYICTIITS-PSLQSGSGTRVKVVDALPTTPTPTKKTTTK-----R 154
T.truncatus YTLSLQSVKPSDSGVYFCMTVGN-PDLTFGKGTQLSVVDVLPTTPQPTKKTTPK-----K 152
S.scrofa YLLQLRHVKPADSGNYFCMAVGN-PELTFGKGTRLSVVDVFPTTAQPTKKTRPK-----K 151
B.taurus SVLTLQNVKPADSGVYFCMIIGI-PNLIFGTGTQLSVVDVLPTSPQPTKKTSPK-----K 151
C.familiaris SILNLTRLNPTDSGVYFCMTIGN-PQLTFGKGTRLTVVDVLPTTAQPTRKTTPK-----K 152
F.catus SILNLTSVKPADSGIYFCMTVGS-PELTFGKGTRLSVVDVLPTNSQPTKKPTPR-----K 152
A.melanoleuca SILNLTNVKPADSGIYFCMTIGN-PELTFGKGTQLSVVDVLPTTAQPTRKPSTK-----K 155
P.troglodytes FILNLTSVKPEDSGIYFCMIVGS-PELTFGKGTQLSVVDFLPTTAQPTKKSTPK-----K 152
H.sapiens FILNLTSVKPEDSGIYFCMIVGS-PELTFGKGTQLSVVDFLPTTAQPTKKSTLK-----K 152
P.pygmaeus FILNLTSVKPEDSGIYFCMIVGS-PELTFGKGTQLSVVDFLPTTAQPTKKSTPK-----R 152
N.leucogenys FILNLTSVKPEDSGIYFCMTIGS-PELTFGKGTQLSVVDFLPTTAQPTKKSTPK-----K 157
S.sciureus FFLNLTRVKLEDSGTYFCMVIGS-PTLIFGTGTQLSVVDILPTTAQTTKKSTPK-----K 151
E.caballus STLHLAHVKPEDSGVYFCMIIGT-PELTFGTGTRLSVVDVFPTTARPTTKPTPK-----K 152
C.porcellus FTLSLKHVKPEDSGVYFCMTIGH-PELTFGMGTNLSVVDVLPTTAQPTTKTTPK-----K 152
R.norvegicus PFLKIMDVKPEDSGFYFCAMVGS-PMVVFGTGTKLTVVDVLPTTA-PTKKTTLK-----K 150
M.musculus PFLSIMNVKPEDSDFYFCATVGS-PKMVFGTGTKLTVVDVLPTTA-PTKKTTLKMK--KK 155
G.gallus YRLHINRLHGSDNGTYYCCTIQS-SQLILGTGTQLDVVDVLPLPSMSTLVPLTK-----K 150
A.mexicanum GVLKLNALQPADSGNYYCAVVVHTLDIVFGTGTELVVVDSFPTTAILTTSTPVCGCKEHE 165
S.chuatsi FMLRIINVSEADAGIYSCILKDRKNTEMWKPGIFLRPGVIPPTLPPKTKPKPPV-----K 153
S.salar FTLRINNITAEDAGLYSCMLQNQKENELWRPGVLLRPGETRPTLTPVTKPKPPR-----I 154
X.laevis FTLKIKNLEFSDSGVYYCMVENA-GKVAFGNGTSLQVVETLPTTAAPTTTTKRK------ 161
I.punctatus ELADLFRMFDRNRDGYIDTEELR--EMLRATGEMITEDDVEELMKDGDRNNDGK------ 147
O.fraenata RPCNTKRSEVTQQNGFFCSALPLSLLVGCAVVLLIPLIVII-RMNYLWNVARHHVVK--- 203
L.hirsutus RPCNTKRSEVTQQNGFFCSALPLSLLVGCAVVLLIPLIVII-RMNYLWNVARHHVVK--- 205
M.eugenii RPCNTKHSEVTQQNGFFCSALPLSLLVGCAVVLLIPLIVII-RMNYLWNVARHHVVK--- 207
M.domestica KMCPKKFPGATQQDGPFCSVLLLSLLVGCILILLISLIVIL-RMNYLWHLARHHFVKQLQ 213
T.truncatus KVLR--FPNLVTRKGPSCAPLIVGPLVAVVLVLLVFLGVAI-HLHCLQRRARLRLLKQFY 209
S.scrofa KICR--LPNLVPPKGPACAPLIIGLLVAGLLVLLVSLGAAV-HVYCLQRRARRRLMKQ-- 206
B.taurus KVCHR-VPTQVTQKRLSCAPLILGLLAAGVLILLVSLGVAI-HLRCLQRRARLRLLKQFY 209
C.familiaris RMCR--FPIQVTQKAPSCGPLTLSLLVAGVLVLLVSLGVAI-HLHRLRRRARLRLLQQFY 209
F.catus KMCR--PPSPVTQKGPSCGLLTLGLLVAGVLVLLVSLGVAI-HLYRLKRRARLRLLKQFY 209
A.melanoleuca KVSR--SPSPVIQKGLSCDPLILGLLVAGVLLLLVSLGVAI-HLHCLQRRARLRLLKQ-- 210
P.troglodytes RVCR--LPRPETQKGPLCSPITLGLLVAGVLVLLVSLGVAI-HLCCRRRRARLRFMKQ-- 207
H.sapiens RVCR--LPRPETQKGPLCSPITLGLLVAGVLVLLVSLGVAI-HLCCRRRRARLRFMKQFY 209
P.pygmaeus RVCR--LPRPETQKGPLCSPITLGLLVAGVLVLLVSLGVAI-HLCCRRRRARLRFMKQFY 209
N.leucogenys RVCR--LPRPETQKGPLCSPITLGLMVAGVLVLLVSLGVAI-HLYCRRRRARLRFMKQFY 214
S.sciureus TVCR--LPRPETRKGPLCSPITLSLLVAGILVLLVSLGVAI-HLYCRQRRARLRFMKQFY 208
E.caballus KVCR--PPNPVTQEG------------------------------LQRRARLRLLKQ-- 177
C.porcellus KKCQ--PPSPGPQKGLHCSLVTLGLLVASALLLLLSLVVAV-HLYCLRRRARLRFIKQFY 209
R.norvegicus KQCP--TPHPKTQKGLTCGLITLSLLVACILVLLVSLSVAI-HFHCMRRRARIHFMKQFH 207
M.musculus KQCP--FPHPETQKGLTCSLTTLSLLVVCILLLLAFLGVAV-YFYCVRRRARIHFMKQFH 212
G.gallus PMRCKPKNKAINKKGACTPMVWVPLAAGALLLLLSLIPTIR-RFYRLRRRLWVRAHRR-- 207
A.mexicanum ETSKGSTKKKGARAGVACSSVIYAPLATGLVMLVISLVVMINHLQHFHRRYRRHFRKQLV 225
S.chuatsi SVCR--CSKKNS-SQDGCGSLILWPLVGLIAALALALICTLYYFSRLPKKCRHHFVKKRQ 210
S.salar PTGR--CTKRNDQTPKGCGSKVLWPLVGVLLTLAAALIYTLYYFSQLPKKCRHQFAKKRP 212
X.laevis -PCK--CKKPQTPSPVICSAVVWNHTCWICSYSCDSLGRHCLLYKTNLQKNPAVLQEAPT 218
I.punctatus --IDYDEFLEFMKGVE 161

```
O.fraenata      --
L.hirsutus      --
M.eugenii       --
M.domestica     R- 214
T.truncatus     K- 210
S.scrofa        --
B.taurus        K- 210
C.familiaris    K- 210
F.catus         K- 210
A.melanoleuca   --
P.troglodytes   --
H.sapiens       K- 210
P.pygmaeus      K- 210
N.leucogenys    K- 215
S.sciureus      K- 209
E.caballus      --
C.porcellus     K- 210
R.norvegicus    K- 208
M.musculus      K- 213
G.gallus        --
A.mexicanum     K- 226
S.chuatsi       MT 212
S.salar         LK 214
X.laevis        EL 220
I.punctatus     --
```

Figure 4B.7. Sequence alignment of all to date known CD8β sequences
✖ CP = connecting peptide. FGXG=β-g strand. Yellow highlight = N-linked glycosylation sites. T = O-linked glycosylation. Boxed cysteines conserved across mammals. ▶ = Domains.

Appendix 4B

Table 4B.2. Genbank Accession numbers for CD8β chain and the relevant references.

Species	Common Name	Accession Numbers	References
Ailuropoda melanoleuca	Giant Panda	XM_002922344	Annotated
Ambystoma mexicanum	*Axolotl*	AF242416	(Fellah *et al.,* 2002)
Bos taurus	Cattle	NM_001105344	(Zimin *et al.,* 2009)
Canis lupus familiaris	Dog	XM_859954	Annotated
Cavia porcellus	Domestic guinea pig	NM_001172877 AY303774	(Nagarajan *et al.,* 2004)
Equus caballus	Horse	NM_003362954	Annotated
Felis catus	Cat	NM_001009867 AB000484	(Pecoraro *et al.,* 1996)
Gallus gallus	Chicken	NM_205247 Z26484	(Tregaskes *et al.,* 1995)
Homo sapiens	Human	NM_004931, Y00805 M36712	(Parnes *et al.,* 1985) (Norment and Littman, 1988)
Ictalurus punctatus	Channel catfish	NM_001201049	(Chen *et al.,* 2010)
Macaca mulatta	Rhesus monkey	XR_010580	Annotated
Macropus eugenii	Tammar wallaby	EU152105	(Duncan *et al.,* 2008)
Monodelphis domestica	South American grey short tailed opossum	NM_001146331	(Duncan *et al.,* 2008)
Mus musculus	Mouse	NM_009858, M17534, M16799	(Rettig *et al.,* 2009) (Panaccio *et al.,* 1987) (Nakauchi *et al.,* 1987)
Nomascus leucogenys	Northern white-cheeked gibbon	NM_003282487	Annotated
Pan troglodytes	Chimpanzee	XM_001139023	Annotated
Pongo pygmaeus	Bornean Orangutan	X60222	(Lawlor and Parham, 1992)
Rattus norvegicus	Norway Rat	NM_031539	(Parnes *et al.,* 1985)
Salmo salar	Atlantic salmon	AY693392 AY693394 AY701522	(Moore *et al.,* 2005)
Samiri sciureus	Squirrel monkey	AJ130819	(Ureta-Vidal *et al.,* 1999)
Siniperca chuatsi	Mandarin fish	GU550707	Direct submission
Sus scrofa	Pig	NM_213762	(Uenishi *et al.,* 2004)
Tursios truncatus	Bottle nosed dolphin	AY744139	Direct submission
Xenopus laevis	African clawed frog	HQ116784	(Chida *et al.,* 2011)

Appendix 4C

M. eugenii nucleotide sequence CD28

5′end
gcaaactggcctgtctcacagctctttgtgaaaagaggaggccacagctttcagctcaac
gctacc
atgatccgcaaggtgcttctggtcctcagtttcttcccttcagttcaagtaacagataagattttggtcaaac
agccaccttggctgttggtggataatcacgaagtggcaacccctttcctgtaactacatctgtgataaaactcc
aacggaatttcgagcttcactccaaaagggaacgaacagtgcccttgaagtctgttttgtgtatgtaaatgga
acccataagcccttgctttcctcaatggaggacttcaactgctccgtgaattttgataataaaacagtgaagt
ttctactccagaatatgagtatcaaccaaacggatatttacttctgtaaaattgaattcatgtatcctcctcc
atatctctccaatgaaaaaagtaatggaaccatcatttttgtgaaagagaaggaggtctttccgactcctgga
acttctgagtcccccaaaccctttgggcggttgttgcggctctgtgtgtccttgctttctatgttttgttaa
tgacagtgacttttttaactgctggttgaaaatcaaaaaaagcacaattcttcagtgggactacatgaacat
gactcctcgaaaaccgagaccaactaataagctttgtctgccttatgccccactttgggactatgccgcctat
cgatcctga
3′end
ctcctcgaaaaccgagaccaactaataagctttgtctgccttatgccccactttgggactatgccgcctatcg
atcctgaaactaactcctatccttaatccagccggttaaaacccctctacctcctcaaccctgtttgtctgga
taggaaatgaccatctcccatctctagccggttatattctgccttttacaggtcactcctgtctttttcatg
aagagaaaaaaaaa

M. eugenii putative protein sequence for CD28

MIRKVLLVLSFFPSVQVTDKILVKQPPWLLVDNHEVATLSCNYICDKTPTEFRASLQKGTNSALEVCFVYVNG
THKPLLSSMEDFNCSVNFDNKTVKFLLQNMSINQTDIYFCKIEFMYPPPYLSNEKSNGTIIFVKEKEVFPTPG
TSESPKPFWAVVAALCVLAFYVLLMTVTFFNCWLKIKKSTILQWDYMNMTPRKPRPTNKLCLPYAPLWDYAAY
RS

Secondary structure prediction for *M. eugenii* CD28

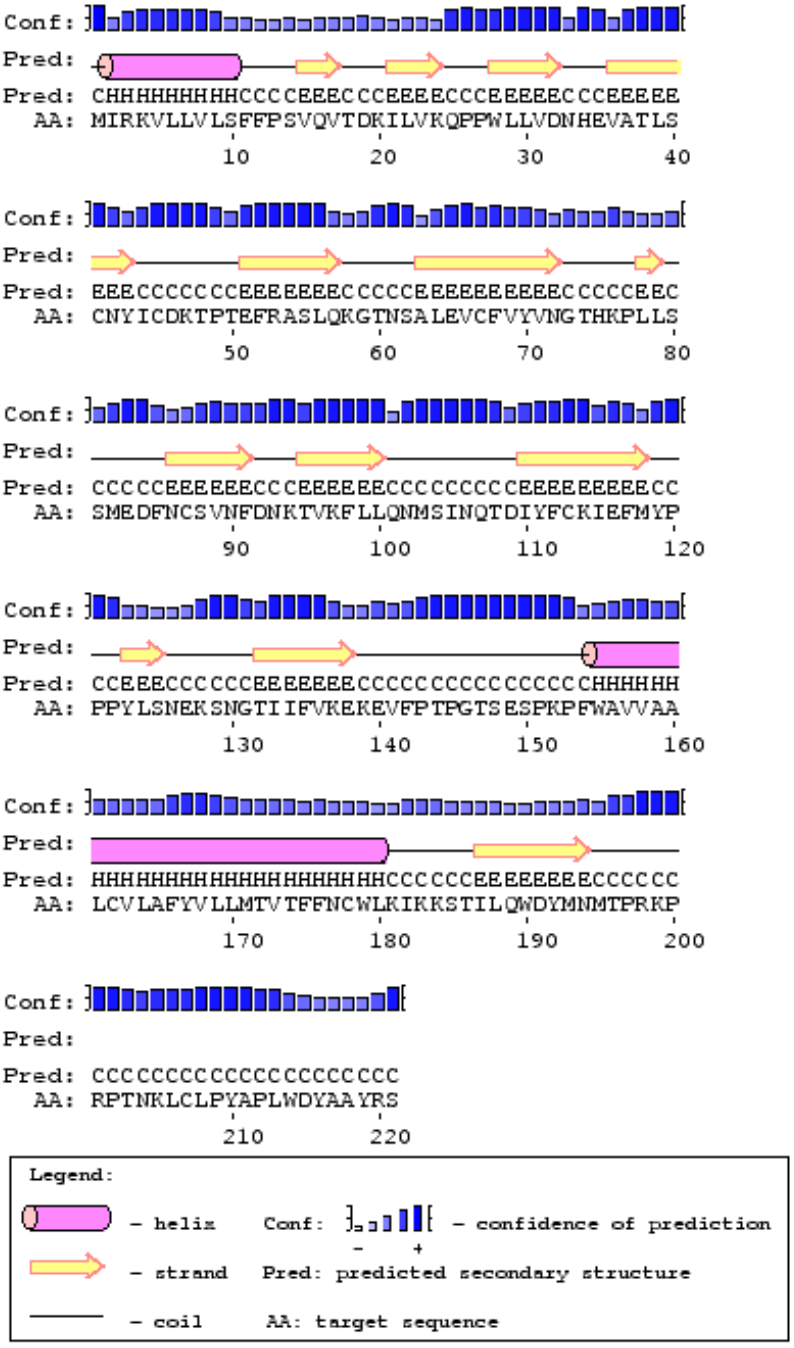


Figure 4C.1. PSIpred secondary structure prediction *M. eugenii* CD28.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Alignment of annotated *M. eugenii* CD28 sequence (ensembl) and actual *M. eugenii* sequence.

```
Annotated       ----------------------------------------------------CAGGATAAG 9
M.eugenii       ATGATCCGCAAGGTGCTTCTGGTCCTCAGTTTCTTCCCTTCAGTTCAAGTAACAGATAAG 60
                                                                    ******

Annotated       ATTTTGGTCAAACAGCCACCTTGGCTGTTGGTGGATAATCACGAAGTGGCAACCCTTTCC 69
M.eugenii       ATTTTGGTCAAACAGCCACCTTGGCTGTTGGTGGATAATCACGAAGTGGCAACCCTTTCC 120
                ************************************************************

Annotated       TGTAACTACATCTGTGATAAAACTCCAACGGAATTTCGAGCTTCACTCCAAAAGGGAACG 129
M.eugenii       TGTAACTACATCTGTGATAAAACTCCAACGGAATTTCGAGCTTCACTCCAAAAGGGAACG 180
                ************************************************************

Annotated       AACAGTGCCCTTGAAGTCTGTTTTGTGTATGTAAATGGAACCCATAAGCCCTTGCTTTCC 189
M.eugenii       AACAGTGCCCTTGAAGTCTGTTTTGTGTATGTAAATGGAACCCATAAGCCCTTGCTTTCC 240
                ************************************************************

Annotated       TCAATGGAGGACTTCAACTGCTCCGTGAATTTTGATAATAAAACAGTGAAGTTTCTACTC 249
M.eugenii       TCAATGGAGGACTTCAACTGCTCCGTGAATTTTGATAATAAAACAGTGAAGTTTCTACTC 300
                ************************************************************

Annotated       CAGAATATGAGTATCAACCAAACGGATATTTACTTCTGTAAAATTGAATTCATGTATCCT 309
M.eugenii       CAGAATATGAGTATCAACCAAACGGATATTTACTTCTGTAAAATTGAATTCATGTATCCT 360
                ************************************************************

Annotated       CCTCCATATCTCTCCAATGAAAAAAGTAATGGAACCATCATTTTTGTGAAAGAGAAGGAG 369
M.eugenii       CCTCCATATCTCTCCAATGAAAAAAGTAATGGAACCATCATTTTTGTGAAAGAGAAGGAG 420
                ************************************************************

Annotated       GTCTTTCCGACTCCTGGAACTTCTGAGTCCCCCAAACCCTTTTGGGGCGGTTGTTGCGGC 429
M.eugenii       GTCTTTCCGACTCCTGGAACTTCTGAGTCCCCCAAACCCTTTTGGGGCGGTTGTTGCGGC 479
                ************************************************************

Annotated       TCTGTGTGTCCTTGCTTTCTATAG---------------------------------- 453
M.eugenii       TCTGTGTGTCCTTGCTTTCTATGTTTTGTTAATGACAGTGACTTTTTTTAACTGCTGGTT 539
                **********************

Annotated       ------------------------------------------------------------
M.eugenii       GAAAATCAAAAAAAAGCACAATTCTTCAGTGGGACTACATGAACATGACTCCTCGAAAACC 599

Annotated       ------------------------------------------------------------
M.eugenii       GAGACCAACTAATAAGCTTTGTCTGCCTTATGCCCCACTTTGGGACTATGCCGCCTATCG 659

Annotated       -------
M.eugenii       ATCCTGA 666
```

Fig. 4C.2. *M. eugenii* annotated CD28 sequence and expressed sequence showing that the annotated sequence is incomplete.

Appendix 4C

M. domestica partial CD28 nucleotide sequence

```
attggtaaatgtttcctgtgactacatctatgataaaactccaactgaatttcgagcttcactccag
aagggaaaaaatggtgcccatgaagtctgttctgtgtatgtaaatggaacccataagcccttgatta
ccacaactgaggacttcagatgccatgtgaattttgataacaaaacggtgatgttttgtcttttgaa
tatgagtatcttccaaacagatatttacttctgtaaaattgaattcatgtatcctcctccatatttc
tccagtgaagtgaacaacggaaccttcatctatgtgaaagaaaaggacgtctgtacaactctgggat
cttctgagcctcccaaacccttctggccagttgtggctgctctgtgtgttttgctttctatagcat
gctaataacagtggcttttgtaactgctggttgaagagcaaaaagaatagaattctccatagt
```

M. domestica putative CD28 protein sequence

```
LVNVSCDYIYDKTPTEFRASLQKGKNGAHEVCSVYVNGTHKPLITTTEDFRCHVNFDNKTVMFCLLN
MSIFQTDIYFCKIEFMYPPPYFSSEVNNGTFIYVKEKDVCTTLGSSEPPKPFWPVVAALCVFAFYSM
LITVAFCNCWLKSKKNRILHS
```

Appendix 4C

Alignment of all species including the *M.eugenii* for CD28

```
M.musculus            ---MTLRLLFLALN-FFS----------VQVTENKILVKQSPLLVVDSNEVS-LSCRYSY 45
R.norvegicus          ---MTLRLLFLALS-FFS----------VQVTENKILVKQSPLLVVDNNEVS-LSCRYSY 45
C.griseus             ---MTPMLLFLALN-FFSS---------VQVTENKILVKQSPMLVVDNNDVT-LSCRYSH 46
M.mulatta             -----MLRLLLALN-LLPS--------IQVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
M.fascicularis        -----MLRLLLALN-LLPS--------IQVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
M.nemestrina          -----MLRLLLALN-LLPS--------IQVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
C.torquatus           -----MLRLLLALN-LLPS--------IRVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
P.anubis              -----MLRLLLALN-LLPS--------IQVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
H.sapiens             -----MLRLLLALN-LFPS--------IQVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
P.troglodytes         -----MLRLLLALN-LFPS--------IQVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
N.leucogenys          -----MLRLLLALN-LFPS--------IQVT-GNKILVKQSPMLVAYDNAVN-LSCKYSY 44
C.jacchus             -----MLRLLLVLN-LFPS--------IQAT-GIKILVKQSPMLEAYDNTVN-LTCKYSC 44
E.grevyi              -----MLRLLLALN-FFPS--------IQVT-ENKILVKQSPMLVVHNNAVN-LSCKYTY 44
E.zebra               -----MLRLLLALN-FFPS--------IQVT-ENKILVKQSPMLVVHNNAVN-LSCKYTY 44
E.caballus            -----MLRLLLALN-FFPS--------IQVT-ENKILVKQSPMLVVHNNAVN-LSCKYTY 44
E.asinus              -----MLRLLLALN-FFPS--------IQVT-ENKILVKQSPMLVVHNNAVN-LSCKYTY 44
R.unicornis           -----MLRLLLALN-FFPS--------IQVT-ENKILVKQSPMLVVHNNAVN-LSCKYTY 44
E.maximus             -----MLRLLLALH-LFPS--------IQATGENKIFVKQAPMLVAYNNAVN-LSCEYTN 45
L.africana            ----MLLRLFLALH-LFPS--------IQATAENKIFVKQAPMLVAYNNAVN-LSCEYTN 46
F.catus               ----MILRLLLALN-FFPS--------IQVT-ENKILVKQLPRLVVYNNEVN-LSCKYTH 45
A.melanoleuca         ----MILRLLLALN-FFPS--------IQVT-ENKILVKQLPRLVVYDNEVN-LSCKYTH 45
C.familiaris          ----MILRLLLALN-FFPS--------IQVT-ENKILVKQLPRLVVYNNEVN-LSCKYTY 45
B.taurus              -----MLRLLLALN-FFPS---------IQVAENKILVKQSPMLVVNDNEVN-LSCKYTY 44
S.caffer              -----MLRLLLALN-FFPS---------IQVAENKILVKQSPMLVVNDNEVN-LSCKYTY 44
B.bonasus             -----MLRLLLALN-FFPS---------IQVAENKILVKQSPMLVVNDNEVN-LSCKYTY 44
G.camelopardalis      -----MLRLLLALN-FFPS---------IQVAENKILVKQSPMLVVNDNEVN-LSCKYTY 44
O.aries               -----MLRLLLALN-FFPS---------IQVAENKILVKQSPMLVVNDNEVN-LSCKYTY 44
S.scrofa              MYCKHHVDLLLPGSTFLPCGESLLENEADPERGNKILVKQSPILVVNDNEVN-LSCKYTY 59
O.cuniculus           ----MILRLLLAFN-FFPS---------IQGTENKILVKQSPMLVVNNNEVN-LSCKYTY 45
M.monax               ----MLRLLLLALN-FFPS--------IQVT-EDKILVKQSPRLEVYNNEVN-LSCKYTY 45
M.domestica           ---------------MLYG----------TATYNKILVKQPHLLVTYNHELVNVSCDYIY 35
M.eugenii             ----MIRKVLLVLS-FFPS----------VQVTDKILVKQPPWLLVDNHEVATLSCNYIC 45
G.gallus              -----MLGILVVLC-LIPA---------ADVTENKILVAQRPLLIVANRTAT-LVCNYTY 44
M.gallopavo           -----MLGILVVLC-LIPA---------ADVTENKILVAQSPLLIVANRTAT-LVCNYTY 44
T.guttata             ----MLLGILVVLF-FIPT---------ADVTENKILVAQHPLLIVANQTAT-LVCNYTY 45
                                    ..               **.* *   * . ..    . * *

M.musculus            NLLAKEFRASLYKGVNSDVEVCVGNGNFTYQPQFRSNAEFNCDGDFDNETVTFRLWNLHV 105
R.norvegicus          NLLAKEFRASLYKGVNSDVEVCVGNGNFTYQPQFRPNVGFNCDGNFDNETVTFRLWNLDV 105
C.griseus             NLLSKEFRASLYKGVNSDVEVCVVNVNFSYQLQFHSNVGFNCDGNLDNETVTFYLRNLHV 106
M.mulatta             NLFSREFRASLHKGLDSAVEVCVVYGNYSQQLQVYSKTGFNCDGKLGNESVTFYLQNLYV 104
M.fascicularis        NLFSREFRASLHKGLDSAVEVCVVYGNYSQQLQVYSKTGFNCDGKLGNESVTFYLQNLYV 104
M.nemestrina          NLFSREFRASLHKGLDSAVEVCVVYGNYSQQLQVYSKTGFNCDGKLGNESVTFYLQNLYV 104
C.torquatus           NLFSREFRASLHKGLDSAVEVCVVYGNYSQQLQVYPKTGFNCDGKLGNESVTFYLQNLYV 104
P.anubis              NLFSREFRASLHKGLDSAVEVCVVYGNYSQQLQVYSKTGFNCGGKLGDESVTFYLQNMYV 104
H.sapiens             NLFSREFRASLHKGLDSAVEVCVVYGNYSQQLQVYSKTGFNCDGKLGNESVTFYLQNLYV 104
P.troglodytes         NLFSREFRASLHKGLDSAVEVCVVYGNYSQQLQVYSKTGFNCDGKLGNESVTFYLQNLYV 104
N.leucogenys          NLFSREFRASLHKGLDSAVEVCVVYGNYSQHPQVYSKTGFNCDGKLGNESVTFYLQNLYV 104
C.jacchus             NLFSRQFQASLHKGVDSAVEVCAVHGNYSQLLQVHSATGFNCDGKLGNESVTFYLQNLYV 104
E.grevyi              NLFSKEFRASLYKGADSAVEVCVVNGNHSHQLQFHSNTGFNCDGKLGNETVTFYLWNLYV 104
E.zebra               NLFSKEFRASLYKGADSAVEVCVVNGNHSHQLQFHSNTGFNCDGKLGNETVTFYLWNLYV 104
E.caballus            NLFSKEFRASLYKGADSAVEVCVVNGNHSHQLQFHSNTGFNCDGKLGNETVTFYLWNLYV 104
E.asinus              NLFSKEFRASLYKGADSAVEVCVVNGNHSHQLQFHSNTGFNCDGKLGNETVTFYLWNLYV 104
R.unicornis           NLFSKEFRASLYKGADSAVEVCVVNGNYSHQLQFHSNTGFNCDGKLGNETVTFYLRNLYV 104
E.maximus             NLFSKEFQAALYKGVDSDVEVCVVNGNYSHQLQFHSNTGFNCDGKLGNETVTFYLWNLYV 105
L.africana            NLFSKEFQAALYKGVDSDVEVCVVNGNYSHQLQFHSNTGFNCDGKLGNETVTFYLWNLYV 106
F.catus               NFFSKEFRASLYKGVDSAVEVCVVNGNYSHQPQFYSSTGFDCDGKLGNETVTFYLRNLFV 105
A.melanoleuca         NLFSKEFRASLYKGVDSAVEVCVVNGNYSHQPQFYSSTGFDCDGKMGNETVTFYLRKLFV 105
C.familiaris          NLFSKEFRASLYKGVDSAVEVCVVNGNYSHQPQFYSSTGFDCDGKLGNETVTFYLRNLFV 105
B.taurus              NLFSKEFRASLYKGADSAVEVCAVNGNHSHPLQST-NKEFNCTVKVGNETVTFYLQDLYV 103
S.caffer              NLFSKEFRASLYKGADSAVEVCAVNGNHSHPLQST-NKEFNCTVKVGNETVTFYLQDLYV 103
B.bonasus             NLFSKEFRASLYKGADSAVEVCAVNGNHSHPLQST-NKEFNCTVKVGNETVTFYLQDLYV 103
G.camelopardalis      NLFSKEFRASLYKGADSAVEVCAVNGNHSHPLQST-NKEFNCTVKVGNETVTFYLQDLYV 103
O.aries               NLFSKEFRASLYKGADSAVEVCAVNGNHSHPLQST-NKEFNCTVKVGNETVTFYLQDLYV 103
S.scrofa              NLFSKEFRASLYKGADSAVEVCVVNVNYSRLLQFKPNTGFNCDVKYGNETVTFYLRNLHV 119
O.cuniculus           NLFSKEFRASLYKGADSAVEVCVVNGNFSHPHQFHSTTGFNCDGKLGNETVTFYLKNLYV 105
M.monax               NLFSKEFRASLYKGVDSAVEVCVVNGNFSHQLQFYSHTGFNCDGKLGNETVTFYLRNLYV 105
```

Appendix 4C

```
M.domestica          DKTPTEFRASLQKGKNGAHEVCSVYVNGTHKPLITTTEDFRCHVNFDNKTVMFCLLNMSI 95
M.eugenii            DKTPTEFRASLQKGTNSALEVCFVYVNGTHKPLLSSMEDFNCSVNFDNKTVKFLLQNMSI 105
G.gallus             NGTGKEFRASLHKGTDSAVEVCFISWN-MTKINSNSNKEFNCRGIHDKDKVIFNLWNMSA 103
M.gallopavo          NGTGKEFRASLHKGTDSSVEVCFISWN-MTKSNSNSNKEFNCWGNHDKDKVIFNLRNMTA 103
T.guttata            NGTGKEFRASLQKGTDSSVEVCVISWN-TTKISSNSNKGFNCQGSYDKDKVIFNLWNMNT 104
                     .   .*.*.* ** ..  ***    *           * *    ....* * * ..

M.musculus           NHTDIYFCKIEFMYPPPYLDNERSNGTIIHIKEKHLCHTQSSP---KLFWALVVVAGVLF 162
R.norvegicus         NHTDIYFCKIEVMYPPPYLDNEKSNGTIIHIKEKHLCHAQTSP---KLFWPLVVVAGVLL 162
C.griseus            NQTDIYFCRIEVMYPPPYLDNEKSNGTIIHVKEKHLCPDQETP---KLLWVLIGGAGVLF 163
M.mulatta            NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWALVVVGGVLA 164
M.fascicularis       NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWALVVVGGVLA 164
M.nemestrina         NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWALVVVGGVLA 164
C.torquatus          NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWALVVVGGVLA 164
P.anubis             NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWALVVVGGVLA 164
H.sapiens            NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVVVGGVLA 164
P.troglodytes        NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVVVGGVLA 164
N.leucogenys         NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWALVVVGGVLA 164
C.jacchus            NQTDIYFCKIEIMYPPPYLDSEKSNGTIIHVKGKHLCPGPSFSGPSQPFWALAVVGGVLA 164
E.grevyi             NQTDIYFCKIEVMYPPPYIDNEKSNGTIIHVKEKHLCPVHAFTESSTPFWALAVTGGVLA 164
E.zebra              NQTDIYFCKIEVMYPPPYIDNEKSNGTIIHVKEKHLCPVHAFTESSTPFWALVVTGGVLA 164
E.caballus           NQTDIYFCKIEVMYPPPYIDNEKSNGTIIHVKEKHLCPVHPFTESSTPFWALAVTGGVLA 164
E.asinus             NQTDIYFCKIEVMYPPPYIDNEKSNGTIIHVKEKHLCPVHLFTESSTPFWALAVTGGVLA 164
R.unicornis          NQTDIYFCKIEVMYPPPYIDNEKSNGTIIHVKEKHLCPDRPSPVSSKPFWALVVVG-VLA 163
E.maximus            NQTDIYFCKIEVMYPPPYIHNEKNNGTIIHVKEKHICPAPPSTESSKPFWALVVVNGVLA 165
L.africana           NQTDIYFCKIEVMYPPPYIHNEKNNGTIIHVKEKHICPAPPSTESSKPFWALVVVNGVLA 166
F.catus              NQTDIYFCKIEVMYPPPYIDNEKSNGTIIHVKEKHLCPAQLSPESSKPFWALVVVGGILG 165
A.melanoleuca        NQTDIYFCKIEVMYPPPYIDNEKSNGTIIHVKEKHHCPAQPSPESSKPFWALVVVGGVLV 165
C.familiaris         NQTDIYFCKIEVMYPPPYIGNEKSNGTIIHVKEKHLCPDELFPDSSKPFWALVVVGAVLV 165
B.taurus             NQTDIYFCKLEVLYPPPYIDNEKSNGTIIHVKEKHLCPSPRSPESSKPFWALVVVNGVLV 163
S.caffer             NQTDIYFCKLEVLYPPPYIDNEKSNGTIIHVKEKHLCPSPRSPESSKPFWALVVVNGVLV 163
B.bonasus            NQTDIYFCKLEVLYPPPYIDNEKSNGTIIHVKEKHLCPSPRSPESSKPFWALVVVNGVLV 163
G.camelopardalis     NQTDIYFCKLEVLYPPPYIDNEKSNGTIIHVKEKHLCPSPRSPESSKPFWALVVVNGVLV 163
O.aries              NQTDIYFCKLEVLYPPPYIDNEKSNGTIIHVKEKHLCPSPQSPESSKPFWALVVVNGVLV 163
S.scrofa             NQTDIYFCKIEVLYPPPYIDNEKSNGTIIHVKEKH-CPAPRPPESSKIFWVLVVVNGVVA 178
O.cuniculus          NQTDIYFCKIEVMYPPPYLDNEKSNGTIIHVKEQHFCPAHPSPKSSTLFWVLVVVGAVLA 165
M.monax              NQTDIYFCKIEVMYPPPYLDNEKSNGTVIHVKENNICPGVPSPEPPKPFWTLVVFSGVLG 165
M.domestica          FQTDIYFCKIEFMYPPPYFSSEVNNGTFIYVKEKDVCTTLGSSEPPKPFWPVVAALCVFA 155
M.eugenii            NQTDIYFCKIEFMYPPPYLSNEKSNGTIIFVKEKEVFPTPGTSESPKPFWAVVAALCVLA 165
G.gallus             SQTDIYFCKIEAMYPPPYVYNEKSNGTVIHVRETP--IQTQEPESATSYWVMVAVTGLLG 161
M.gallopavo          SQTDIYFCKIEAMYPPPYVYNEKSNGTVIHVRETP--IQTQEPESATSYWVMVALTGLLG 161
T.guttata            NQTDIYFCKIEVMYPPPYVYNEKSNGTVIHVKETP--TQIQEPQSAIPLWILATVTGILA 162
                      .******..* .*****. .* .***.*...          .      * .     ..

M.musculus           CYGLLVTVALCVIWTNSRRNRLLQVTTMNMTPRRPGLTR-KPYQPYAPARDFAAYRP--- 218
R.norvegicus         CYGLLVTVTLCIIWTNSRRNRLLQSDYMNMTPRRLGPTR-KHYQPYAPARDFAAYRP--- 218
C.griseus            LYGLVVTVALCIIWKNSKRNRLLQSDYMNMTPRRLGPTR-KHYQPYAPARDFAAYSP--- 219
M.mulatta            CYSLLVTVAFCIFWMRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDFAAYRS--- 220
M.fascicularis       CYSLLVTVAFCIFWMRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDFAAYRS--- 220
M.nemestrina         CYSLLVTVAFCIFWMRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDSAAYRS--- 220
C.torquatus          CYSLLVTVAFRIFWMRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDFAAYRS--- 220
P.anubis             CYSLLVTVAFSIFCMRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDFAAYRS--- 220
H.sapiens            CYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDFAAYRS--- 220
P.troglodytes        CYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDFAAYRS--- 220
N.leucogenys         CYSLLVTVAFSIFWMRSKRSRLLHSDYMNMTPRRPGPTR-KHYQPYAPPRDFAAYRS--- 220
C.jacchus            SYSLLVTVALSVFWMRSRRSRLLHSDYMNMTPRCPGPTR-RHYQPYAPPRDFAAYRS--- 220
E.grevyi             FYSLLVTVALCTCWMRNRRSRTLQSDYMNMTPRRPGPTR-KHYQPYAPARDFAAYRS--- 220
E.zebra              FYSLLVTVALCTCWMRNRRSRTLQSDYMNMTPRRPGPTR-KHYQPYAPARS--------- 214
E.caballus           FYSLLVTVALCTCWMRNRRSRTLQSDYMNMTPRRPGPTR-KHYQPYAPARDFAAYRS--- 220
E.asinus             FYSLLVTVALCTCWMRNRRSRTLQSDYMNMTPRRPGPTR-KHYQPYAPARDFAAYRS--- 220
R.unicornis          FYSLLVTVALCTCWMRSKRSRTLQSDYMNMTPRRPGPTR-KHYQPYAPTRDFAAYRS--- 219
E.maximus            FYSLVITVALCICWMKNKRSRILQSDYMNMTPRRPGPTR-KHYQPYAPARDFAAYRS--- 221
L.africana           FYSLVITVALCICWMKNKRSRILQSDYMNMTPRRPGPTR-KHYQPYAPARDFAAYRS--- 222
F.catus              FYSLLATVALGACWMKTKRSRILQSDYMNMTPRRPGPTR-RHYQPYAPARDFAAYRS--- 221
A.melanoleuca        FYSLLVTVALCACWMKNKRSRILQSDYMNMTPRRPGPTR-RHYQPYAPTRDFAAYRS--- 221
C.familiaris         FYSLLVTVALCAYWIKSKSSRILQSDYMNMTPRRPGPTR-RHYQPYAPARDFAAYRS--- 221
B.taurus             FYSLLVTVALSNCWMKNKRNRMLQSDYMNMTPRRPGPTR-RHYQPYAPARDFAAYRS--- 219
S.caffer             FYSLLVTVALSNCWMKNKRNRMLQSDYMNMTPRRPGPTR-RHYQPYAPARDFAAYRS--- 219
B.bonasus            FYSLLVTVALSNCWMKNKRNRMLQSDYMNMTPRRPGPTR-RHYQPYAPARDFAAYRS--- 219
G.camelopardalis     FYSLLVTVALSNCWMKSKRNRMLQSDYMNMTPRRPGPTR-RHYQPYAPARDFAAYRS--- 219
O.aries              FYSLLVTVALCNCWMKSKRNRMHQSDYMNMTPRRPGPTR-RHYQPYAPTRDFAAYRS--- 219
```

Appendix 4C

```
S.scrofa          FYSLVVTLALFFYWMKSKRTRMLQSDYMNMTPRRLGPTR-KHYQPYAPARDFAAYRS--- 234
O.cuniculus       FYSMLVTVALFSCWMKSKKNRLLQSDYMNMTPRRPGPTR-KHYQPYAPARDFAAYRS--- 221
M.monax           IYSLLSTMLLCYLWTKRQRTRLLQSDYMNMTPRRPGPSR-KHYQPYAP------YRP--- 215
M.domestica       FYSMLITVAFCNCWLKSKKNRILHSDYMNMTPQRPGPTK-KHYQPYAPSRDYAAYRS--- 211
M.eugenii         FYVLLMTVTFFNCWLKIKKSTILQWDYMNMTPRKPRPTN-KLCLPYAPLWDYAAYRS--- 221
G.gallus          FYSMLITAVFIIYRQKSKRNRYRQSDYMNMTPRHPPHQKNKGYPSYAPTRDYTAYRSWQP 221
M.gallopavo       FYSMLVTAVFIIYRQKSKRNRYRQSDYMNMTPRHPPHQKNKGYPPYAPTRDYTAYRSWQP 221
T.guttata         LYSTLITAVSINYWQKSKKYMYRQSDYMNMTPRHPPYQKNKGYPSYAPTRDYTAYRSWQP 222
                   *  . *       . .     .   *****.     . .   .***
```

Figure 4C.3. Alignment of putative protein sequences including the predicted protein sequence of the *M. domestic.* Cysteine = red. NMT = glycosylation site below threshold. Grey = potential glycosylation sites. Magenta = MYPPPY motif.

Table 4C.1. Genbank accession numbers for CD28 and the relevant references.

Species Name	Common name	Accession Number	References
Bison bonasus	European bison	EU000418	Direct submission
Bos taurus	Cattle	NM_181004	(Parsons *et al.,* 1996)
Canis lupus familiaris	Dog	NM_001003087	(Pastori *et al.,* 1994)
Cercocebus torquatus atys	Sooty mangabey	AF344842	(Villinger *et al.,* 2001)
Elephas maximus	Asian elephant	EU000417	Direct submission
Equus asinus somalicus	Somali Wild Ass	EU000413	Direct submission
Equus caballus	Horse	NM_001100179	Direct submission
Equus grevyi	Grevy's Zebra	EU000415	Direct submission
Equus zebra hartmannae	Zebra	EU000414	Direct submission
Felis catus	Cat	AB025316	Direct submission
Fugu rubripes	Pufferfish	BK005769	(Bernard *et al.,* 2006)
Gallus Gallus	Chicken	NM_205311	(Young *et al.,* 1994)
Giraffa camelopardalis	Giraffe	EU000420	Direct submission
Homo sapiens	Human	NM_006319	(Lafage-Pochitaloff *et al.,* 1990)
Macaca fascicularis	Crab-eating macaque	NM_001042641	Annotated
Macaca mulatta	Rhesus monkey	DQ872187	(Villinger *et al.,* 2001)
Marmota monax	Woodchuck	EF534209	Direct submission
Meleagris gallopavo	Turkey	AM884252	(Powell *et al.,* 2009)
Monodelphis domestica	Opossum	XM_001371298	Annotated
Mus musculus	Mouse	NM_007642	(Gross *et al.,* 1992)
Oryctolagus cuniculus	Rabbit	NM_001082207	(Isono *et al.,* 1995)
Ovis aries	Sheep	AF092739	(Chaplin *et al.,* 1999)
Rattus norvegicus	Norway Rat	NM_013121	(Clark and Dallman, 1992)
Rhinocerus unicornus	Greater Indian rhinoceros	EU000416	Direct submission
Syncerus caffer	African buffalo	EU000419	Direct submission
Taeniopygia guttata	Zebrafinch	XM_002187839	Annotated

Appendix 4D

M. eugenii CTLA-4 nucleotide sequence

atggttctcctgggatccaggagacaaatggaaaaggttcaccctcccaagaactggccctgcacggcaatgc
tctctctgctctttatcccaagtatctccaaaggggtacatgtgactcaaccagcagtggtcgtggccagcgg
caaagggattgcaagctttgtgtgtaactttgagttgacaaacaaaaccacagagatccgagtgggccttctt
cgacagatggacaaccaaatggttgaagtctgtgcctcaacatacctggtacagaatcaaccagtgttcatgg
atgacatgcttgagtgcacgggaaatgccagtgggaataaattgatgctcactctcacaggactgaaggcctc
agacagtggactgtacatctgcaaagtggagctcatgtatcctccccctactacatgggtttgggcaatgga
acacagatatatgccattgatcctgaaccttgcccagactttgaagttatgctctggatcctagctatagtga
gctccgcattatttttttacagctccctcattacagctgtctccttgaataaaatgctaaagaaagggagtct
tctcactacaggagtctatgtgaaaatgcctccaacagaacctgaacatgagaagcaatttcagccctacttt
attcccatcaactga

M. eugenii CTLA-4 putative amino acid sequence

MVLLGSRRQMEKVHPPKNWPCTAMLSLLFIPSISKGVHVTQPAVVVASGKGIASFVCNFELTNKTTEIRVGLL
RQMDNQMVEVCASTYLVQNQPVFMDDMLECTGNASGNKLMLTLTGLKASDSGLYICKVELMYPPPYYMGLGNG
TQIYAIDPEPCPDFEVMLWILAIVSSALFFYSSLITAVSLNKMLKKGSLLTTGVYVKMPPTEPEHEKQFQPYF
IPIN

Alignment of annotated *M. eugenii* CTLA-4 and the expressed gene obtained from mRNA

```
ensembl         MVLLGSRRQMEKVHPPKNWPCTAMLSLLFIPSISKGVHVTQPAVVVASGRGIASFVCNFE 60
M.eugenii       MVLLGSRRQMEKVHPPKNWPCTAMLSLLFIPSISKGVHVTQPAVVVASGKGIASFVCNFE 60
                ************************************************.********** 

ensembl         LTNKTTEIRVGLLRQMDNQMVEVCASTYLVQNQPVFMDDMLECTGNASGNKLMLTLTGLK 120
M.eugenii       LTNKTTEIRVGLLRQMDNQMVEVCASTYLVQNQPVFMDDMLECTGNASGNKLMLTLTGLK 120
                ************************************************************ 

ensembl         ASDSGLYICKVELMYPPPYYMGLGNGTQIYAIDPEPCPDFEVMLWILAIVSSALFFYSFL 180
M.eugenii       ASDSGLYICKVELMYPPPYYMGLGNGTQIYAIDPEPCPDFEVMLWILAIVSSALFFYSSL 180
                ************************************************************ * 

ensembl         ITAVSLNKMLKKRSLLTTGVYVKMPPTEPEHEKQFQPYFIPIN 223
M.eugenii       ITAVSLNKMLKKGSLLTTGVYVKMPPTEPEHEKQFQPYFIPIN 223
                ************ ***************************** 
```

Figure 4D.1. *M. eugenii* CTLA-4. Comparison between ensembl deposited sequence and expressed gene.

Secondary structure prediction *M. eugenii* CTLA-4 using PSIpred

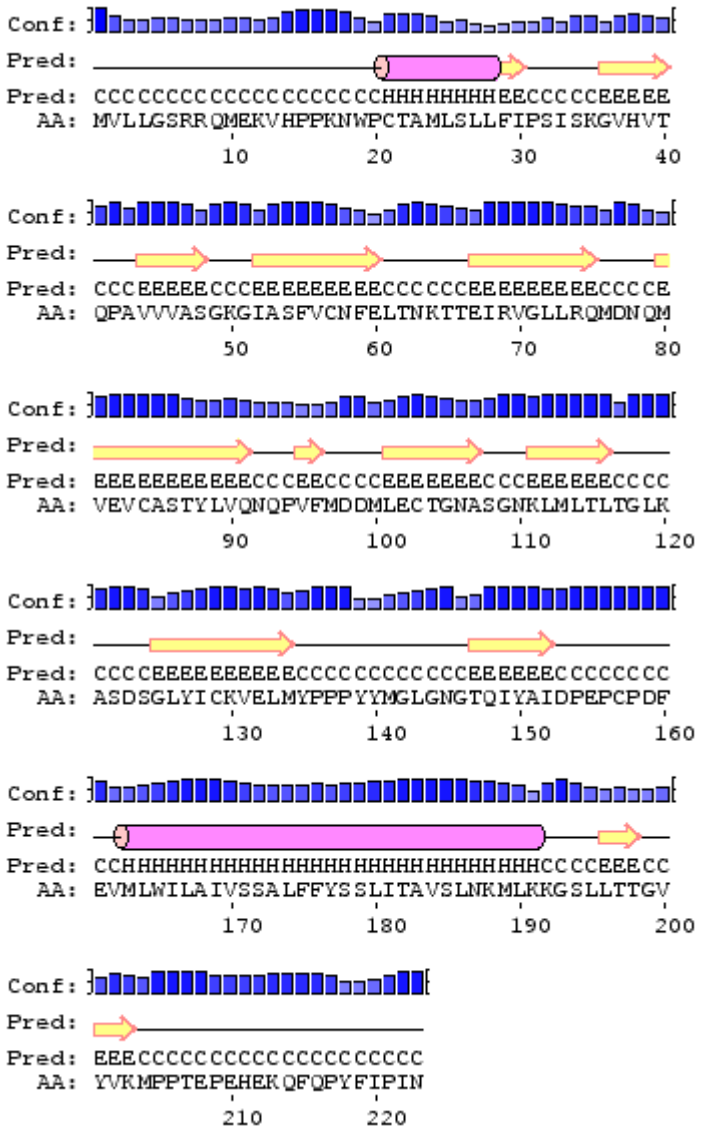


Figure 4D.2. *M. eugenii* CTLA-4 secondary structure prediction.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

O. fraenata CTLA-4 nucleotide sequence

atggttctcgtgggatacaggagacaaatggaaaaggttcaccctcccaagaactggccgtgcacggcaatgc
tctgtctgctctttatcccaagtatctccaaaggggtacatgtgactcaaccagcagtggtagtggccagcgg
cagagggattgcaagctttgtgtgtaactttgagttgtcaaacaaaaccacggagatccgagtgggccttttt
ggacagatggacaaccaaatggttgaagtgtgtgcctcaacatacttggtacagaatcaaccagtgttcatgg
atgacatgcttgagtgcacgggaaatgccagtgggaataaattgatgctcactctcacaggatcgaaggcctc
agacagtggactgtacatatgcaaagtggagctcatgtatcctcccccctactacatgggtttgggcaatgga
acacagatatacgccattgatcctgaaccttgcccagactttgaagttatgctctggatcctagctatagtga
gctccgcattatttttttacagcttcctcattacagctgtctccttgaataaaatgctaaagaaaaggagtct
tctcactacaggagtgtatgtgaaaatgcctccaacagaacctgaacatgagaagcaatttcagccatacttt
attcccataaactga

O. fraenata CTLA-4 putative protein sequence

MVLVGYRRQMEKVHPPKNWPCTAMLCLLFIPSISKGVHVTQPAVVVASGRGIASFVCNFELSNKTTEIRVGLF
GQMDNQMVEVCASTYLVQNQPVFMDDMLECTGNASGNKLMLTLTGSKASDSGLYICKVELMYPPPYYMGLGNG
TQIYAIDPEPCPDFEVMLWILAIVSSALFFYSFLITAVSLNKMLKKRSLLTTGVYVKMPPTEPEHEKQFQPYF
IPIN

Secondary structure prediction *O. fraenata* CTLA-4 using PSIpred

Figure 4D.3. Secondary structure prediction of *O. fraenata* CTLA-4.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 4D

M. domestica CTLA-4 partial nucleotide sequence

```
gcatggacacagatatatgtcattgattctgaaccttgtccagactctgatgtctcactctggatattggcta
tagtcagctctggactatttttttacagcctcctcatcacagctgtttccttgaataaaatgctaaagaaaag
gagtcttctgactacaggggtctacgtgaaaatgcccacaccattgcacttaattacttga
```

M. domestica putative partial amino acid

```
AWTQIYVIDSEPCPDSDVSLWILAIVSSGLFFYSLLITAVSLNKMLKKRSLLTTGVYVKMPTPLHLIT
```

M. domestica secondary structure prediction of partial CTLA-4 sequence using PSIpred.

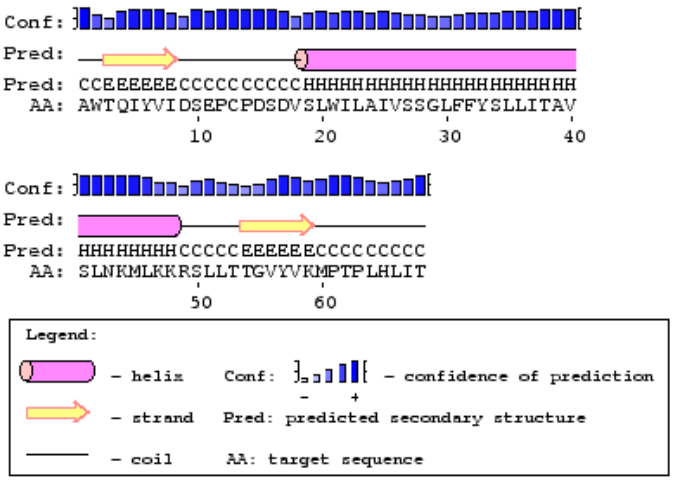


Figure 4D.4. Secondary structure prediction of partial *M. domestica* CTLA-4 sequence.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 4D

Alignment of expressed partial *M. domestica* CTLA-4 sequence and predicted *M. domestica* sequence

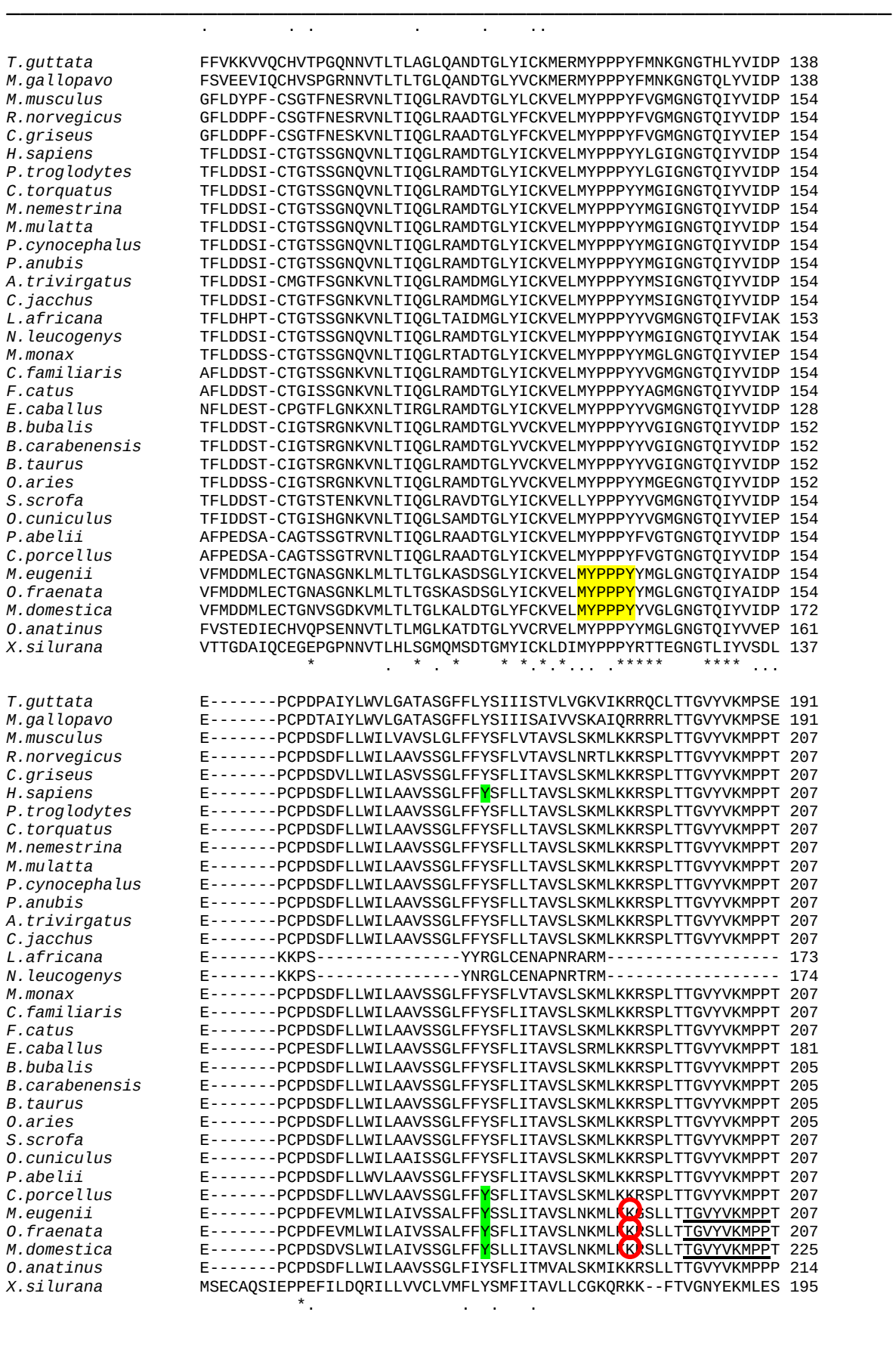
```
M.domestica       ------------------------------------------------------------
pred.opossum      MVSLNFCYLGDFIGCNQAMLLLGSRREMGKLHHPMNWPCTAMLSLLFIPSISKGIHVTQP

M.domestica       ------------------------------------------------------------
pred.opossum      AVILANSRGVASFVCEYELTAKTKEIRVSLLRQMDDELVEVCASTYLVQNQPVFMDDMLE

M.domestica       ----------------------------------------AWTQIYVIDSEPCPDSDV
pred.opossum      CTGNVSGDKVMLTLTGLKALDTGLYFCKVELMYPPPYYVGLGNGTQIYVIDPEPCPDSDV
                                                         *******.********

M.domestica       SLWILAIVSSGLFFYSLLITAVSLNKMLKKRSLLTTGVYVKMPTPLHLIT----------
pred.opossum      SLWILAIVSSGLFFYSLLITAVSLNKMLKKRSLLTTGVYVKMPPTEPEHEKQFQPYFITI
                  ****************************************..

M.domestica       -
pred.opossum      H
```

Figure D4.5. Alignment of expressed partial *M. domestica* CTLA-4 and the predicted *M. domestica* sequence (Accession No. XM_001371277.1).

Appendix 4D

Sequence alignment of known CTLA-4 sequences (for accession number see Table 4D.1).

```
T.guttata           -----------------------------------MLSILVTMGFLCTAT------AIA 18
M.gallopavo         -----------------------------------MLSAWVIVSFLCAAT------ATA 18
M.musculus          -----------------MACLGLRRYKAQLQLPSRTWPFVALLTLLFIP-------VFS 35
R.norvegicus        -----------------MACLGLQRYKTHLQLPSRTWPFGVLLSLLFIP-------IFS 35
C.griseus           -----------------MAGLGVQRCRAQLQLASRTWPFEALLAFLFIP-------TFS 35
H.sapiens           -----------------MACLGFQRHKAQLNLAARTWPCTLLFFLLFIP-------VFC 35
P.troglodytes       -----------------MACLGFQRHKAQLNLATRTWPCTLLFFLLFIP-------VFC 35
C.torquatus         -----------------MACLGFQRHKARLNLATRTRPYTLLFSLLFIP-------VFS 35
M.nemestrina        -----------------MACLGFQRHKARLNLATRTRPYTLLFSLLFIP-------VFS 35
M.mulatta           -----------------MACLGFQRHKARLNLATRTRPYTLLFSLLFIP-------VFS 35
P.cynocephalus      -----------------MACLGFQRHKAQLNLATRTRPYTLLFSLLFIP-------VFS 35
P.anubis            -----------------MACLGFQRHKAQLNLATRTRPYTLLFSLLFIP-------VFS 35
A.trivirgatus       -----------------MACLGFQRHKAQLDLATRTWPCTFLFSLLFIL-------VFS 35
C.jacchus           -----------------MACLGFRRHKAQLDLATRTWPCTLLFSLLFIP-------VFS 35
L.africana          -----------------MAYIVFHRCGAQLHQP-RTWPCTAMLSLLFIP-------ICS 34
N.leucogenys        -----------------MACLGFQRHTAQLNLATRTWPCTLLFFLLFIP-------VFC 35
M.monax             -----------------MACLGLQRPEAGLEPAPRAWSCTALLSLLFIP-------IFS 35
C.familiaris        -----------------MAGFGFRRHGAQPDLASRTWPCTALFSLLFIP-------VFS 35
F.catus             -----------------MACFGFRRHGAQLDLASRTWPCTALFSLLFIP-------VFS 35
E.caballus          -----------------MAGFGFRRRGARLDPAPRTWPCTALFSLLFMP-------VFS 35
B.bubalis           -----------------MACSGFQSHGTWR--TSRTWPCTALFFLLFIP-------VFS 33
B.carabenensis      -----------------MACSGFQSHGTWR--TSRTWPCTALFFLLFIP-------VFS 33
B.taurus            -----------------MACSGFQSHGTWW--TSRTWPCTALFFLVFIP-------VFS 33
O.aries             -----------------MACSGFQSHGTWR--TSRTWPCTALFFLLFIP-------VFS 33
S.scrofa            -----------------MACSGFQSHGAWLELTSRTWPCTALFSLLFIP-------VFS 35
O.cuniculus         -----------------MARLGFQRQGTQLDLASRTWSCAALFSLLFLP-------VFS 35
P.abelii            -----------------MARLELRRHQVQLWLLPKTWPCSALLSLLVMP-------VFS 35
C.porcellus         -----------------MARLELRRHQVQLWLLPKTWPCSALLSLLVMP-------VFS 35
M.eugenii           -----------------MVLLGSRRQMEKVHPP-KNWPCTAMLSLLFIP-------SIS 34
O.fraenata          -----------------MVLVGYRRQMEKVHPP-KNWPCTAMLCLLFIP-------SIS 34
M.domestica         MVSLNFCYLGDFIGCNQAMLLLGSRREMGKLHHP-MNWPCTAMLSLLFIP-------SIS 52
O.anatinus          -----------------MIRLGSKR-QIRERHHYRNRSWSVMVSFLFVAGFGGPAAGLA 41
X.silurana          ----------------------------------MMRIPFITGIFCLITN---------A 17
                                                       .    .  .            .

T.guttata           EVMEVTQPAIVLANRQGVASLVCKYKNIGNAKEIRVTLLKQTGDQVTEICASSYTTEFKT 78
M.gallopavo         KVMEVTQPAIVLANRQGVASLVCNYKHIGNAKEIRVTLLKQTGDKFTEICASTYTMEFEM 78
M.musculus          EAIQVTQPSVVLASSHGVASFPCEYSPSHNTDEVRVTVLRQTNDQMTEVCATTFTEKNTV 95
R.norvegicus        EAIQVTQPSVVLASSHGVASFPCEYASSHNTDEVRVTVLRQTNDQVTEVCATTFTVKNTL 95
C.griseus           KAIHVAQPSVVLASSHGVASFSCEYTSSHNTDEVRVTVLRQTNSQMTEVCATTFTMKNKL 95
H.sapiens           KAMHVAQPAVVLASSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMTGNEL 95
P.troglodytes       KAMHVAQPAVVLASSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNEL 95
C.torquatus         KAMHVAQPAVVLANSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNEL 95
M.nemestrina        KAMHVAQPAVVLANSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNEL 95
M.mulatta           KAMHVAQPAVVLANSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNEL 95
P.cynocephalus      KAMHVAQPAVVLANSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNEL 95
P.anubis            KAMHVAQPAVVLANSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNEL 95
A.trivirgatus       NAMHVAQPAVVLASSRGIASFVCEYASPGNTTEIRVTVLRQTDSQVTEVCAGTYIMGNEL 95
C.jacchus           NAMHVAQPAVVLASSRGIASFACEYASPGKATEIRVTVLRQTDSQVTEVCAGTYIMGNEL 95
L.africana          KALDVSQPAVVLASNRGVASFVCEYESLHKVKEVRVTVLRQANSQMTEVCASTFEVENEL 94
N.leucogenys        KAMHVAQPAVVLASSRGIASFVCEYASPGKATEVRVTVLRQADSQETEVCAATYMMGNEL 95
M.monax             KAMHVAQPAVVLASSRGVASFVCEYAAFHKATEVRVTVLRQIASQVTEVCATTYTVENEL 95
C.familiaris        KGMHVAQPAVVLASSRGVASFVCEYGSSGNAAEVRVTVLRQAGSQMTEVCAATYTVEDEL 95
F.catus             KGMHVAQPAVVLASSRGVASFVCEYGSSGNAAEVRVTVLRQTGSQMTEVCAATYTVENEL 95
E.caballus          K--------VWIFRQAQVRSDIATQGAPKNEGWPHV-----------------MVENEV 69
B.bubalis           KGMNVTQPPVVLASSRGVASFSCEYESSGKADEVRVTVLREAGSQVTEVCAGTYMVEDEL 93
B.carabenensis      KGMNVTQPPVVLASSRGVASFSCEYESSGKADEVRVTVLREAGSQVTEVCAGTYMVEDEL 93
B.taurus            KGMNVTQPPVVLASSRGVASFSCEYESSGKADEVRVTVLREAGSQVTEVCAGTYMVEDEL 93
O.aries             KGMNVTQPPVVLASSRGVASFTCEYESSGKADEVRVTVLRKAGIQVTEVCAGTYMVEDEL 93
S.scrofa            KGMHVAQPAVVLANSRGVASFVCEYGSAGKAAEVRVTVLRRAGSQMTEVCAATYTVEDEL 95
O.cuniculus         KALHVSQPAVVLASSRGVASFVCEYASSHKATEVRVTVLRQANSQMTEVCAMTYTVENEL 95
P.abelii            KAMHVAQPAVVLASSRGVASFECEYASSHNANEVRVTVLQQVASRTTEICAATYTVEREL 95
C.porcellus         KAMHVAQPAVVLASSRGVASFECEYASSHNANEVRVTVLQQVASRTTEICAATYTVEREL 95
M.eugenii           KGVHVTQPAVVVASGKGIASFVCNFELTNKTTEIRVGLLRQMDNQMVEVCASTYLVQNQP 94
O.fraenata          KGVHVTQPAVVVASGRGIASFVCNFELSNKTTEIRVGLFGQMDNQMVEVCASTYLVQNQP 94
M.domestica         KGIHVTQPAVILANSRGVASFVCEYELTAKTKEIRVSLLRQMDDELVEVCASTYLVQNQP 112
O.anatinus          EALQVTQPRVVLASMKGVASLACEYEFTGKAKEIRVTLIRQTGNEFHEVCASSFTTEYEP 101
X.silurana          SGLKVTQPDIIVANRHGKAMLVCDYRIHAKVEEMRFRLLRKMGNQVKEICAFSYSTNYES 77
```



Appendix 4D

```
                    .        . .          .      .    ..

T.guttata           FFVKKVVQCHVTPGQNNVTLTLAGLQANDTGLYICKMERMYPPPYFMNKGNGTHLYVIDP 138
M.gallopavo         FSVEEVIQCHVSPGRNNVTLTLTGLQANDTGLYVCKMERMYPPPYFMNKGNGTQLYVIDP 138
M.musculus          GFLDYPF-CSGTFNESRVNLTIQGLRAVDTGLYLCKVELMYPPPYFVGMGNGTQIYVIDP 154
R.norvegicus        GFLDDPF-CSGTFNESRVNLTIQGLRAADTGLYFCKVELMYPPPYFVGMGNGTQIYVIDP 154
C.griseus           GFLDDPF-CSGTFNESKVNLTIQGLRAADTGLYFCKVELMYPPPYFVGMGNGTQIYVIEP 154
H.sapiens           TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYLGIGNGTQIYVIDP 154
P.troglodytes       TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYLGIGNGTQIYVIDP 154
C.torquatus         TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYMGIGNGTQIYVIDP 154
M.nemestrina        TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYMGIGNGTQIYVIDP 154
M.mulatta           TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYMGIGNGTQIYVIDP 154
P.cynocephalus      TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYMGIGNGTQIYVIDP 154
P.anubis            TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYMGIGNGTQIYVIDP 154
A.trivirgatus       TFLDDSI-CMGTFSGNKVNLTIQGLRAMDMGLYICKVELMYPPPYYMSIGNGTQIYVIDP 154
C.jacchus           TFLDDSI-CTGTFSGNKVNLTIQGLRAMDMGLYICKVELMYPPPYYMSIGNGTQIYVIDP 154
L.africana          TFLDHPT-CTGTSSGNKVNLTIQGLTAIDMGLYICKVELMYPPPYYVGMGNGTQIFVIAK 153
N.leucogenys        TFLDDSI-CTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYMGIGNGTQIYVIAK 154
M.monax             TFLDDSS-CTGTSSGNQVNLTIQGLRTADTGLYICKVELMYPPPYYMGLGNGTQIYVIEP 154
C.familiaris        AFLDDST-CTGTSSGNKVNLTIQGLRAMDTGLYICKVELMYPPPYYVGMGNGTQIYVIDP 154
F.catus             AFLDDST-CTGISSGNKVNLTIQGLRAMDTGLYICKVELMYPPPYYAGMGNGTQIYVIDP 154
E.caballus          NFLDEST-CPGTFLGNKXNLTIRGLRAMDTGLYICKVELMYPPPYYVGMGNGTQIYVIDP 128
B.bubalis           TFLDDST-CIGTSRGNKVNLTIQGLRAMDTGLYVCKVELMYPPPYYVGIGNGTQIYVIDP 152
B.carabenensis      TFLDDST-CIGTSRGNKVNLTIQGLRAMDTGLYVCKVELMYPPPYYVGIGNGTQIYVIDP 152
B.taurus            TFLDDST-CIGTSRGNKVNLTIQGLRAMDTGLYVCKVELMYPPPYYVGIGNGTQIYVIDP 152
O.aries             TFLDDSS-CIGTSRGNKVNLTIQGLRAMDTGLYVCKVELMYPPPYYMGEGNGTQIYVIDP 152
S.scrofa            TFLDDST-CTGTSTENKVNLTIQGLRAVDTGLYICKVELLYPPPYYVGMGNGTQIYVIDP 154
O.cuniculus         TFIDDST-CTGISHGNKVNLTIQGLSAMDTGLYICKVELMYPPPYYVGMGNGTQIYVIEP 154
P.abelii            AFPEDSA-CAGTSSGTRVNLTIQGLRAADTGLYICKVELMYPPPYFVGTGNGTQIYVIDP 154
C.porcellus         AFPEDSA-CAGTSSGTRVNLTIQGLRAADTGLYICKVELMYPPPYFVGTGNGTQIYVIDP 154
M.eugenii           VFMDDMLECTGNASGNKLMLTLTGLKASDSGLYICKVELMYPPPYYMGLGNGTQIYAIDP 154
O.fraenata          VFMDDMLECTGNASGNKLMLTLTGSKASDSGLYICKVELMYPPPYYMGLGNGTQIYAIDP 154
M.domestica         VFMDDMLECTGNVSGDKVMLTLTGLKALDTGLYFCKVELMYPPPYYVGLGNGTQIYVIDP 172
O.anatinus          FVSTEDIECHVQPSENNVTLTLMGLKATDTGLYVCRVELMYPPPYYMGLGNGTQIYVVEP 161
X.silurana          VTTGDAIQCEGEPGPNNVTLHLSGMQMSDTGMYICKLDIMYPPPYRTTEGNGTLIYVSDL 137
                               *       .  * . *    * *.*.*... .*****    **** ...

T.guttata           E-------PCPDPAIYLWVLGATASGFFLYSIIISTVLVGKVIKRRQCLTTGVYVKMPSE 191
M.gallopavo         E-------PCPDTAIYLWVLGATASGFFLYSIIISAIVVSKAIQRRRRLTTGVYVKMPSE 191
M.musculus          E-------PCPDSDFLLWILVAVSLGLFFYSFLVTAVSLSKMLKKRSPLTTGVYVKMPPT 207
R.norvegicus        E-------PCPDSDFLLWILAAVSSGLFFYSFLVTAVSLNRTLKKRSPLTTGVYVKMPPT 207
C.griseus           E-------PCPDSDVLLWILASVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 207
H.sapiens           E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
P.troglodytes       E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
C.torquatus         E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
M.nemestrina        E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
M.mulatta           E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
P.cynocephalus      E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
P.anubis            E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
A.trivirgatus       E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
C.jacchus           E-------PCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPT 207
L.africana          E-------KKPS---------------YYRGLCENAPNRARM------------------ 173
N.leucogenys        E-------KKPS---------------YNRGLCENAPNRTRM------------------ 174
M.monax             E-------PCPDSDFLLWILAAVSSGLFFYSFLVTAVSLSKMLKKRSPLTTGVYVKMPPT 207
C.familiaris        E-------PCPDSDFLLWILAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 207
F.catus             E-------PCPDSDFLLWILAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 207
E.caballus          E-------PCPESDFLLWILAAVSSGLFFYSFLITAVSLSRMLKKRSPLTTGVYVKMPPT 181
B.bubalis           E-------PCPDSDFLLWILAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 205
B.carabenensis      E-------PCPDSDFLLWILAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 205
B.taurus            E-------PCPDSDFLLWILAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 205
O.aries             E-------PCPDSDFLLWILAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 205
S.scrofa            E-------PCPDSDFLLWILAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 207
O.cuniculus         E-------PCPDSDFLLWILAAISSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 207
P.abelii            E-------PCPDSDFLLWVLAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 207
C.porcellus         E-------PCPDSDFLLWVLAAVSSGLFFYSFLITAVSLSKMLKKRSPLTTGVYVKMPPT 207
M.eugenii           E-------PCPDFEVMLWILAIVSSALFFYSSLITAVSLNKMLKKGSLLTTGVYVKMPPT 207
O.fraenata          E-------PCPDFEVMLWILAIVSSALFFYSFLITAVSLNKMLKKRSLLTTGVYVKMPPT 207
M.domestica         E-------PCPDSDVSLWILAIVSSGLFFYSLLITAVSLNKMLKKRSLLTTGVYVKMPPT 225
O.anatinus          E-------PCPDSDFLLWILAAVSSGLFIYSFLITMVALSKMIKKRSLLTTGVYVKMPPP 214
X.silurana          MSECAQSIEPPEFILDQRILLVVCLVMFLYSMFITAVLLCGKQRKK--FTVGNYEKMLES 195
                              *.               .  .   .
```

```
T.guttata           --KLEKKVIPFHITVDCNKEREKPFPSWDGESDNSFQLKK 229
M.gallopavo         --KLEKKVIPFHITVN------------------------ 205
M.musculus          EPECEKQFQPYFIPIN------------------------ 223
R.norvegicus        EPECEKQFQPYFIPIN------------------------ 223
C.griseus           EPECEKQFQPYFIPIN------------------------ 223
H.sapiens           EPECEKQFQPYFIPIN------------------------ 223
P.troglodytes       EPECEKQFQPYFIPIN------------------------ 223
C.torquatus         EPECEKQFQPYFIPIN------------------------ 223
M.nemestrina        EPECEKQFQPYFIPIN------------------------ 223
M.mulatta           EPECEKQFQPYFIPIN------------------------ 223
P.cynocephalus      EPECEKQFQPYFIPIN------------------------ 223
P.anubis            EPECEKQFQPYFIPIN------------------------ 223
A.trivirgatus       EPECEKQFQPYFIPIN------------------------ 223
C.jacchus           EPECEKQFQPYFIPIN------------------------ 223
L.africana          ---------------------------------------
N.leucogenys        ---------------------------------------
M.monax             EPECEKQFQPYFIPIN------------------------ 223
C.familiaris        EPECEKQFQPYFIPIN------------------------ 223
F.catus             EPECEKQFQPYFIPIN------------------------ 223
E.caballus          EPECEKQFQPYFIPIN------------------------ 197
B.bubalis           EPECEKQFQPYFIPIN------------------------ 221
B.carabenensis      EPECEKQFQPYFIPIN------------------------ 221
B.taurus            EPECEKQFQPYFIPIN------------------------ 221
O.aries             EPECEKQFQPYFIPIN------------------------ 221
S.scrofa            EPECEKQFQPYFIPIN------------------------ 223
O.cuniculus         EPECEKQFQPYFIPIN------------------------ 223
P.abelii            EPECEKQFQPYFIPIN------------------------ 223
C.porcellus         EPECEKQFQPYFIPIN------------------------ 223
M.eugenii           EPEHEKQFQPYFIPIN------------------------ 223
O.fraenata          EPEHEKQFQPYFIPIN------------------------ 223
M.domestica         EPEHEKQFQPYFITIH------------------------ 241
O.anatinus          EPEHEKQFQPYFIPIN------------------------ 230
X.silurana          --DQGNGFSPYYIRVN------------------------ 209
```

Figure 4D.5. Sequence alignment of known CTLA-4 .

TGVYVMPPT = the intracellular localization motif, which restricts CTLA-4 expresssion to intracellular membranes such as the perinuclear Golgi or post- Golgi compartments (does not exist in CD28)

Appendix 4D

Table 4D.1. Genbank Accession numbers for CTLA-4 and the relevant references.

Species	Common name	Accession Numbers	References
Ailuropoda melanoleuca (PREDICTED)	Giant Panda	XM_002919948	Annotated
Anas platyrhyncos	Duck	GQ995931	(Yao *et al.,* 2010)
Aotus trivirgatus	Three-striped monkey	AF344834	(Villinger *et al.,* 2001)
Bos Taurus	Cattle	NM_174297.1	(Parsons *et al.,* 1996)
Bubalus bubalis	Water buffalo	FJ827143.1	(Mingala *et al.,* 2011)
Bubalus carabanensis	Carabao	FJ827142.1	(Mingala *et al.,* 2011)
Callithrix jacchus	White-tuffed ear marmoset	GQ284838.1	Direct submission
Canis lupus familiaris	Dog	NM_001003106.1 AF154842 AF215893	(Khatlani *et al.,* 2000) Unpublished Unpublished
Cavia porcellus (PREDICTED)	Guinea Pig	XM_003474180	Annotated
Cercocebus torquatus atys	Sootey mangabey	AF344848	(Villinger *et al.,* 2001)
Cricetulus griseus	Chinese hamster	AF307318	Direct submission
Equus caballus (PREDICTED)	Horse	XM_001497853.2	Annotated
Felis catus	Cat	AF170725.1	(Choi *et al.,* 2000a)
Gallus gallus	Chicken	NM_001040091.1	Annotated
Homo sapiens	Human	AF414120.1 AY209009	Direct submission
Loxodonta africana (PREDICTED)	African elephant	XM_003406118	Annotated
Macaca mulatta	Rhesus monkey	NM_001044739.1	(Villinger *et al.,* 2001)
Macaca nemestrina	Pig tailed macaque	AF344854	(Villinger *et al.,* 2001)
Marmota monax	Woodchuck	AF130428.1	Unpublished
Meleagris gallopavo (PREDICTED)	Turkey	XM_003207503	Annotated
*Monodelphis domestica (*PREDICTED)	Opossum	XM_001371277.1	Annotated
Mus musculus	Mouse	NM_009843.3	(Freeman *et al.,* 1992)
Nomascus leucogenys (PREDICTED)	White cheeked Gibbon	XM_003253971	Annotated
Ornithorhynchus anatinus (PREDICTED)	Platypus	XM_001514865.1	Annotated
Oryctolagus cuniculus	Rabbit	NM_001082685.1	(Isono *et al.,* 1995)
Ovis aries	Sheep	NM_001009214.1	(Chaplin *et al.,* 1999)
Papio Anubis	Olive baboon	NM_001112634.1	(Villinger *et al.,* 2001)
Pongo abelii (PREDICTED)	Orangutan	XM_002812770	Annotated
Rattus norvegicus	Norway Rat	NM_031674.1 U90271	(Waterhouse *et al.,* 1995)
Sus scrofa	Pig	AF281633.1	Direct Submission
Taeniopygia guttata (PREDICTED)	Zebra finch	XM_002197453	Annotated

Appendix 4E

M. domestica CD86 partial nucleotide sequence

cgggactgtagacctgtcttgtaattttaagaatcctgaaggaatcagcctggaagaactactgatattttgg
caagatgctaatgatcttgttctgtatgagctatatcaaggaagagagaagcaagatcacatccatgagaagt
accttaaccgaaccgagtacaaccaaaccacgtggactttacaactccggaatatccagattgaggatcagag
ggaatataaatgtttagtccaacaccgtagccccagaggcttagttcttgtccatcggttttcttttcagctg
tttgtctttgctcctttcagtcaacctgaaataacacgacttgataacatgacagtaaaaattggggacgtgt
tgaatttttcgaaataacacgacttgataacatgacagtaaaaattggggacgtgttgaatttttc

M. domestica CD86 putative amino acid sequence

GTVDLSCNFKNPEGISLEELLIFWQDANDLVLYELYQGREKQDHIHEKYLNRTEYNQTTWTLQLRNIQIEDQR
EYKCLVQHRSPRGLVLVHRFSFQLFVFAPFSQPEITRLDNMTVKIGDVLNFSK

Secondary structure prediction for *M. domestica* CD86 using PSIpred

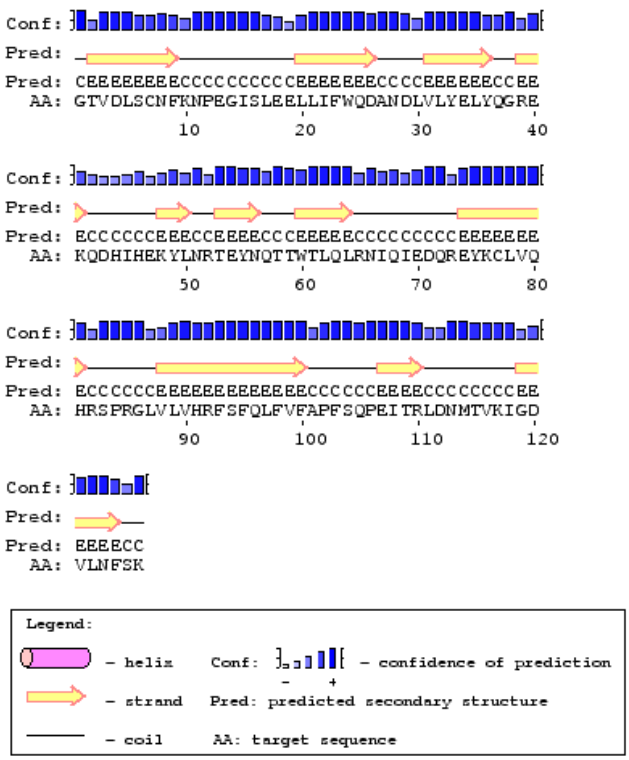


Figure 4E.1. Secondary structure prediction of *M. domestica* CD86.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 4E

Amino acid alignment of CD86 (eCD86= ensembl sequence), *M. domestica* is the obtained sequence from amplification of the molecule in *M. domestica* thymus cDNA library

```
eCD86           ATKPKVEALFNGTVDLSCNFKNPEGISLEELLIFWQDANDLVLYELYQGREKQDHIHEKY 60
M.domestica     -----------GTVDLSCNFKNPEGISLEELLIFWQDANDLVLYELYQGREKQDHIHEKY 49
                           *******************************************

eCD86           LNRTEYNQTTWTLQLRNIQIEDQREYKCLVQHRSPRGLVLVHRFSFQLFVFAPFSQPEIT 120
M.domestica     LNRTEYNQTTWTLQLRNIQIEDQREYKCLVQHRSPRGLVLVHRFSFQLFVFAPFSQPEIT 109
                ************************************************************

eCD86           RLDNMTVKIGDVLNFSCSSEQGYPEPEEMYWMITTENSTKIPGIMDLSQDKTTQLYNVRS 180
M.domestica     RLDNMTVKIGDVLNFSK---------------------------------------- 126
                ***************

eCD86           TLTLTFNETTRTNISCYLQTVRQKEP 206
M.domestica     -------------------------
```

Figure 4E.2. Alignment of the predicted *M. domestica* CD86 (eCD86) from ensembl and the expressed sequence.

Appendix 4E

Alignment of known CD86 sequences

```
                                                                IGV-domain
R.norvegicus          MYVVKTCVTCTMYLGILFSVLAYLL--------------SDAVPVKRQAYFNSTAYLPCP 46
M.musculus            -----MDPRCTMGLAILIFVTVLLI--------------SDAVSVETQAYFNGTAYLPCP 41
M.unguiculatus        -----MDPRRIMELGILLLVTAFTL--------------SGAASMKIQAYFNSTACLPCP 41
M.monax               -----------MGLRITLFVTAFLL--------------SDVASLKSQASFNETADLPCQ 35
C.griseus             -----MDPQCTMGLSNILFVMAFLL--------------SGAAPLKIQAYFNETADLPCQ 41
P.troglodytes         -----------MGLSNTLFVMAFLL--------------SGAAPLKIQAYFNETADLPCQ 35
P.anubis              -----------MGLSNILFVMAFLL--------------SGAAPLKIQAYFNETADLPCQ 35
C.troquatus           -----------MGLSNILFVMAFLL--------------SGAAPLKIQAYFNETADLPCQ 35
M.mulatta             -----------MGLSNILFVMAFLL--------------SGAAPLKIQAYFNETADLPCQ 35
M.nemestrina          -----------MGLSNILFVMAFLL--------------SGAAPLKIQAYFNETADLPCQ 35
C.aethiops            -----------MGLINILFVMAFLL--------------SGAAPLKIQAYFNETADLPCQ 35
C.jacchus             -----MDLQCAMGRITILFMVGFLI--------------SGVDSLKIHAYFNETADLPCQ 41
O.cuniculus           -----MDAGCTMGLSVTVFVMALLL--------------SGAASLRIQAYFNKTADLPCQ 41
S.scrofa              -----------MGLSNILFVMVLLL--------------SGAASLKSQAYFNETGELPCH 35
B.taurus              ----MRFK-CTMGLRNILMGMALRLSVSK-------VPFSGAASLKSHAFFNETGELPCH 48
C.familiaris          ----MYLR-CTMELNNILFVMTLLL--------------YGAASMKSQAYFNKTGELPCH 41
A.melanoleuca         ----MGICDSTMGLSNTLLGMALLI--------------SGAASMMSQGYFNKTGELPCH 42
L.africana            ---MYAQFQSKMAMSSGPLLIRQLQ---------------GAAAAIKTYFNETAYLPCQ 41
pred.M.domestica      ---MEAHLPRELGETEALGIIEDTGDGRGSSRLRRWCAVAAATKPKVEALFNGTVDLSCN 57
M.domestica           ------------------------------------------------------GTVDLSCN 8
G.gallus              -----------MEVCIFFLYAIILLPG-----------IAAN-VHHVKSFLNHTAYLSCY 37
T.guttata             -----------MEVCIFFLCAMIFLPG-----------IAAASVHPVQAFLHHTVYLSCY 38
F.catus               -MGHAAKWKTPLLKHPYPKLFPLLMLAS------LFYFCSGII--QVNKTVEEVAVLSCD 51
O.aries               -MGHTMKWGTLLPKRPCLWLSQLLVLAG------LFHFCSGITPKSVTKRVKETVVLSCD 53
O.mykiss              -----MTVDYGMCARRLLRRCVILFLVS-----------VQFIQSKDVTMVIGEVGGSVT 44
                                                                              .

R.norvegicus          FTKAQNISPSELVVFWQDRKKSVLYEHYLGAEKLDNVNAKYLGR-TSFDRDNQALRLHNV 105
M.musculus            FTKAQNISLSELVVFWQDQQKLVLYEHYLGTEKLDSVNAKYLGR-TSFDRNNWTLRLHNV 100
M.unguiculatus        HSKVQNMSLSELVVFWQDQKKLVLYEHYLGREKSNSVNAKYLGR-TSFDEHNWALRLPNV 100
M.monax               FINSQNISLGELIIFWQDQQKLVLYELYLGNEKPDNVAAKYLGR-TSFDQDNWTMQLHNV 94
C.griseus             FANSQNQSLSELVVFWQDQENLVLNEVYLGKEKFDSVHSKYMGR-TSFDSDSWTLRLHNL 100
P.troglodytes         FANSQNQRLSELVVFWQDQENLVLNEVYLGKEKFDSVHSKYMGR-TSFDSDSWTLRLHNL 94
P.anubis              FANSQNRSLSELVVFWQNQENLVLNEVYLGREKFDSVHSKYMGR-TSFDPESWTLRLHNL 94
C.troquatus           FANSQNRSLSELVVFWQNQENLVLNEVYLGREKFDSVHSKYMGR-TSFDPDSWTLRLHNL 94
M.mulatta             FANSQNRSLSELVVFWQNQENLVLNEVYLGKEKFDSVHSKYMGR-TSFDPESWTLRLHNL 94
M.nemestrina          FANSQNRSLSELVVFWQNQENLVLNEVYLGKEKFDSVHSKYMGR-TRFDPESWTLRLRNL 94
C.aethiops            FANSQNRSLSELVVFWQNQENLVLNEVYLGQEKFDSVHSKYMGR-TSFDPESWTLRLHNL 94
C.jacchus             FANSQNLSLSELVAFWQNQENLVLNEVYLGKEKFDSVHSRYMGR-TSFDPDSWTLRLHNL 100
O.cuniculus           FTNSQSRSLSELVVFWQDQERLVLYELFLGREKPDNVDPKYIGR-TSFDQESWNLQLHNV 100
S.scrofa              FTNSQNLSLDELVIFWQDQDNLVLYELYRGQEKPHNVNSKYMGR-TSFDQATWTLRLHNV 94
B.taurus              FPNTQNLSLDELVIFWQDQNKLVLYELFKGQEKPNNVNPKYIGR-TSFDQDSWTLRLHNV 107
C.familiaris          FTNSQNISLDELVVFWQDQDKLVLYELYRGKENPQNVHRKYKGR-TSFDKDNWTLRLHNI 100
A.melanoleuca         FTNSQNISLDELVVFWQDQDKLVLYELYRGKENPQNVHPKYKGR-TSFDKDNWTLKLHNI 101
L.africana            FINSQNISLDELVVFWQNQEKLVVYELYQGKEKFDNVDPKYKNRNISFDEENWTLRLHNV 101
pred.M.domestica      FKNPEGISLEELLIFWQDANDLVLYELYQGREKQDHIHEKYLNR-TEYNQTTWTLQLRNI 116
M.domestica           FKNPEGISLEELLIFWQDANDLVLYELYQGREKQDHIHEKYLNR-TEYNQTTWTLQLRNI 67
G.gallus              FPNSQKTDINNVIVFWQKGTGEVVHEVYLGQEKHDHLNSKYINR-TKMDMDKWTLQLLNV 96
T.guttata             FPNSQKFDVKDLIIFWQKESKKVLHEVYHGQEKHDNLSPEYINR-TKVDMGKWTLQLLNA 97
F.catus               Y-NISTKELTEIRIYWQKDDEMVLAVMSG----KVQVWPKYKNRTFTDVTDNHSIVIMAL 106
O.aries               Y-NTTTEELASLRIYWQKDSKMVLAILLG----KVQVWSEYENRTITDINNNPRIVILAL 108
O.mykiss              LPCTSDIQRLPNHLYVQRPDPNKFINGYHKTRDLPSPHPEYSNR-TQVDHTQGTMRLWSI 103
                                 . *      .              .* .*          . .
                                                  ▼PKA binding site
R.norvegicus          QIKDTGLYDCFIQQKTPTGSIILQQWETELSVIAN-FSEPEIEEAQNETRNT--GINLTC 162
M.musculus            QIKDMGSYDCFIQKKPPTGSIILQQTLTELSVIAN-FSEPEIKLAQNVTGNS--GINLTC 157
M.unguiculatus        QITDMGSYDCYIQQKRPTGSVILQQTNMELSVVAN-FSEPEIELAQNVTRNS--GINLTC 157
M.monax               QIKDKGSYQCIVHHKGPQGIVHLYQMTSELSVFAN-FSEPEIRPLSNITGDS--GINLTC 151
C.griseus             QIKDKGLYQCIIHHKKPTGMIRIHQMNSELSVLAN-FSQPEIVPISNITENV--YINLTC 157
P.troglodytes         QIKDKGLYQCIIHHKKPTGMIRIHQMNSELSVLAN-FSQPEIVPISNITENV--YINLTC 151
P.anubis              QIKDKGLYQCIIHHKRPTGMIRIHQMNSELSVLAS-FSQPEIVPISNITENM--YINLTC 151
C.troquatus           QIKDKGLYQCIIHHKRPTGMIRIHQMNSELSVLAN-FSQPEIVPISNITENM--YINLTC 151
M.mulatta             QIKDKGLYQCIIHHKRPTGMIRIHQMNSELSVLAN-FSQPEIVPISNITENM--YINLTC 151
M.nemestrina          QIKDKGLYQCIIHHKRPTGMIRIHQMNSELSVLAN-FSQPEIVPISNITENM--YINLTC 151
C.aethiops            QIKDKGLYQCIIHHKRPTGMIRIHQMNSELSVLAN-FSQPEIVPISNITENM--YINLTC 151
C.jacchus             QITDKGLYRCIIHHKKPTGMIRIHQMNSDLLVLAN-FSQPEIVPISNITENL--YINLTC 157
```

Appendix 4E

```
O.cuniculus          QIKDKGVYQCFVHHRGAKGLVPIYQMNSELSVLAN-FTQPEITLISNITRNS--AINLTC 157
S.scrofa             QIKDKGSYQCFIHHKGPHGLVPIHQMSSDLSLLAN-FSQPEINLLTNHTENS--VINLTC 151
B.taurus             QIKDTGSYQCFIHHRRSQGLVSIHQMSSDLIVLAN-FSQPEIRLIANQTEKSN-IINLTC 165
C.familiaris         QIKDKGLYQCFVHHKGPKGLVPMHQMNSDLSVLAN-FSQPEIMVTSNRTENSG-IINLTC 158
A.melanoleuca        QIKDKGSYQCFIHHKGPKGLVPMYHMGSELSVLAN-FTQPEIMVTSNRTENSG-IINLTC 159
L.africana           EIKDQGDYQCFIHHKGPKGLVPSHKMTRELKVLSN-FSQPEIMEGSNSSSKS--YRNLTC 158
pred.M.domestica     QIEDQREYKCLVQHRSPRGLVLVHRFSFQLFVFAP-FSQPEITRLDNMTVKIGDVLNFSC 175
M.domestica          QIEDQREYKCLVQHRSPRGLVLVHRFSFQLFVFAP-FSQPEITRLDNMTVKIGDVLNFSK 126
G.gallus             GIVDEGQYKCIIMHVDKGPKKLIHESECLLNITAN-YSQPVIAQLHTGEPKPNENLNLSC 155
T.guttata            EIEDEGLYQCIIQKIVEQSKEVVHQSECSLRIVAN-YSQPEIAELHSGELKPNGYLNLSC 156
F.catus              RLSDNGKYTCIIQKIE-KGSYKVKHLTSVMLLVRG------------------------ 140
O.aries              RLSDSGTYTCVIQKPDLKGTYKVEHLTSVKLMIRADFPVPTINDLGNPSPNIR---RVIC 165
O.mykiss             RLSDEGLYECHIGYPT-----KNNQKNIQLSVTAN-YSIPNVTVACDNGSCL-----VTC 152
                      . *   * * .           .      .

R.norvegicus         SSKQGYPKPTKMYFLIT--NSTNEYGDNMQISQDNVTKLFSVSISLSLPFPDGVYNMTIV 220
M.musculus           TSKQGHPKPKKMYFLIT--NSTNEYGDNMQISQDNVTELFSISNSLSLSFPDGVWHMTVV 215
M.unguiculatus       TSEHGFPKPMKMYFLII--NSTNKQGDDMEISQDNVTELFSVSTSLSLPFPEDAYNVTFW 215
M.monax              S------------------------------------------------------------ 152
C.griseus            SSIHGYPEPKKMSVLLRTKNSTIEYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 217
P.troglodytes        SSIHGYPEPKKMSVLLRTKNSTIEYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 211
P.anubis             SSIHGYPEPEKMSVLLRTKNSTIEYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 211
C.troquatus          SSIHGYPEPEKMSVLLRTKNPTIEYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 211
M.mulatta            SSIHGYPEPEKMSVLLRTKNSTIEYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 211
M.nemestrina         SSIHGYPEPEKMSVLLRTKNSTIEYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 211
C.aethiops           SSIHGYPEPEKMSVLLRTKNSTIEYDGVMQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 211
C.jacchus            SSIHGYPEPEKMSFLLITKNSTTEYDGVIQKSQDNVTELYDVSISLSVSFPDVTSNMTIF 217
O.cuniculus          SSVQGYPEPKKMFFVLKTENATTEYDGVIEKSQDNVTGLYNISISGSITFSDDIRNATIY 217
S.scrofa             SSTQGYPEPQRMYMLLNTKNSTTEHDADMKKSQNNITELYNVSIRVSLPIPPET-NVSIV 210
B.taurus             SSIQGYPEPQRMYVSLNTTNSSSTYDAVMKKSQSNITELYNVSISVSFPIPPET-NVTIF 224
C.familiaris         SSIQGYPEPKEMYFLVKTENSSTKYDTVMKKSQNNVTELYNVSISLSFSVPEAS-NVSIF 217
A.melanoleuca        SSVQGYPEPKEMYFKLKTENSTTKYDTAMKKSQNNVTELYNVSISLSFSVPEAS-NVSIT 218
L.africana           SSIQGYPQPEEMYFLLTTENSTIKYSTTMQKSQDNITELYNVSISWSYSGFHDTTNRSII 218
pred.M.domestica     SSEQGYPEPEEMYWMITTENSTKIPG-IMDLSQDKTTQLYNVRSTLTLTFNETT-RTNIS 233
M.domestica          ------------------------------------------------------------
G.gallus             SSSGGYPEPKQMIWLISSENITDRLIRHMDVLQDAVTKLYNVTSKLNIPVPTNT-LTNIS 214
T.guttata            SSSGGYPEPKEMTWLISHENITHSSTAHMDVSQDAVTKLYNVTSKLNIPVPTKS-RTNIS 215
F.catus              ----------------------------VTPSTEP------------------------ 147
O.aries              STSGGFPRP----YLSWLENGEELNATNTTLSQDPETKLYTISSELDFNMTSDHNFLCLV 221
O.mykiss             SSDNGYPRRDVEWSLNPPLNQSHWGVVNSSGWTDPVSMLFSVFSSISINCSSGP-RLNLS 211


R.norvegicus         CILETESMN--ISSKPHNMVFSQPQF---DRKTWIQIAGPSSLLCCLFLLVVYKA----- 270
M.musculus           CVLETESMK--ISSKPLNFTQEFP-S---PQTYWKEITAS----VTVALLLVMLL----- 260
M.unguiculatus       CVLETKSMN--ISSRPFSVVLPEPRP---VQENWRVTVVVA---VVVAVLGAVLP----- 262
M.monax              ------------------------------------------------------------
C.griseus            CILETDKTR--LLSSPFSIELEDPQP---PPDHIPWITAVLPTVIICVM-VFCLI-LWKW 270
P.troglodytes        CILETDKTR--LLSSPFSIELEDPQP---PPDHIPWITAVLPTVIICVM-VFCLI-LWKW 264
P.anubis             CVLETDKTQ--LLSSPFSI---------------------------------------- 228
C.troquatus          CVLETDKTQ--LLSSPFSIELEDPQP---PPDHIPWITAVLPTVIICVM-AFCLI-LWKW 264
M.mulatta            CVLETDKTQ--LLSSPFSIELEDPQP---PPDHIPWITAVLPTVIICVM-AFCLI-LWKW 264
M.nemestrina         CVLETDKTQ--LLSSPFSIELEDPQP---PPDHIPWITAVLPSVVICVM-AFCLI-LWKW 264
C.aethiops           CVLETDKTQ--LLSSPFSIELEDPQP---PPDHIPWITAVLPTVIICVM-AFCLI-LWKC 264
C.jacchus            CVLQTKKTQ--LLSSPFSIEIEDLQP---PPDRIPWIAAVLLAIIICVMMVFCLLNLWKW 272
O.cuniculus          CVLQTESTE--TYSQHFPIVPADPVP---VEKPRLWIAAVALTLIVVCGIVLFLT--LWK 270
S.scrofa             CVLQLEPSKTLLFSLPCNIDAKPPVQ-PPVPDHILWIAALLVTVVVVCG--MVSFVTLRK 267
B.taurus             CALQLEPTK-IILSQPYNIDAKSPVPSPPVPDHILWIAALLVTVVVS-G--MV-FLTLKK 279
C.familiaris         CVLQLESMK--LPSLPYNIDAHTKPT--PDGDHILWIAALLVMLVILCG--MVFFLTLRK 271
A.melanoleuca        CVLQLESME--LHSLPYNIDAHTKPP--PARDPILWIAALLVMLVILCGMVLVFFLTLRK 274
L.africana           CVLCVSEMC--LFSKPYDIVPPKTHP--PPKDDILWITALSLVIVGVTVFFLVRW----- 269
pred.M.domestica     CYLQTNYWR--PPAQPAVHRQDKQQK----TLKMKAEAGAKLDLLRVPRKAATLG----- 282
M.domestica          ------------------------------------------------------------
G.gallus             CLLHLGEQQGSLVSVPLVIEIPAEEM----EPVKVNFFGPLVAVILLVT--LLLG----- 263
T.guttata            CLLHLREQLGSLVSVPLGIEIQEKEM----EQAKINFFGPLIAVVVLITSALLLG----- 266
F.catus              ----------------------------NAHAELEIMT--------------------- 157
O.aries              KYGDLTVSQ--------TFYWQESKPTPSANQHLPWTIIIPVSACGISVIIAVIL----- 268
O.mykiss             CAVGGALSQEHTVCRPPDISVVSVIC------AVSVIAAVLLCFLVLVS----------- 254


R.norvegicus         ----------VK-KCLKMQNQPGRPSRKTCESKQDSGVD-ESINLEEVEPQLHQQ----- 313
M.musculus           ----------II-VCHKKPNQPSRPSNTASKLERDSNADRETINLKELEPQIASAKPNAE 309
```

Appendix 4E

```
M.unguiculatus      ----------LI-IYFLCSCWSRRKDARRSRSRVASVS---------------------- 289
M.monax             ------------------------------------------------------------
C.griseus           KKKKRPRNSYKC-GTNTMEREESEQTKKREKIHIPERSDEAQRVFKSSKTSSCDKSDTCF 329
P.troglodytes       KKKKRPRNSYKC-GTNTMEREESEQTKKREKIHIPERSDEAQRVFKSSKTSSCDKSDTCF 323
P.anubis            ------------GTNTMEREESEQTKKREKINVPERSDEAQCVFKSLKTPSCDKSDTHF 275
C.troquatus         KKKKQPRNSYNC-GTNTMEREESEQTKKREKINVPERSDEAQCVFKSLKTPSCDKSDTHF 323
M.mulatta           KKKKQPRNSYKC-GTNTMEREESEQTKKREKINVPERSDEAQCVFKSLKTPSCDKSDTRF 323
M.nemestrina        KKKKQPRNSYKC-GTNTMEREESEQTKKREKINVPERSDEAQCVFKSLKTPSCDKSDTRF 323
C.aethiops          KKKKQPRNSYKR-GTNTMEREESEQTKKREKINVPERSDETQCVFKSLKTPSCDKSDTRF 323
C.jacchus           KKKQQPCISCEC-EHINMERGESEQTQERDSSRWRGRSSLG---LQQLRIAALSEGSEV- 327
O.cuniculus         RKKEQQPGVCEC-ETIKMDKAENEHVEERVKIHEPEKIPAKAAKCEHRLKTPSSDKSAAH 329
S.scrofa            RKKKQPGPSNECGETIKMNRKASEQTKNRAEVHE--RSDDAQCDVNILKTASDDNSTTDF 325
B.taurus            RKKKQPGPSNEC-EIIKVEEKESEQTAKRVELQEPERSDEVQCDVNISKTASDNKSATNL 338
C.familiaris        RKKKQPGPSHEC-ETNKVERKESEQTKERVRYHETERSDEAQC-VNISKTASGDNSTTQF 329
A.melanoleuca       RKKKQPGPSHEC-ETNKMERKESEQTMERVQPHVPERADEAQC-VNISKTTSGNKSTTH- 331
L.africana          ------------KRKKKQPDLSRECQESETIKVEEALTL--------------------- 296
pred.M.domestica    --------------FSHEDPAPNPKPQKKGEGNPNRETLALQRVQYRC------------ 316
M.domestica         ------------------------------------------------------------
G.gallus            ------------FLILK-NRNISSTSQSVSLAV-------------------------- 283
T.guttata           ------------FVILKKNSKILSTSQSVSLAV-------------------------- 287
F.catus             --------------LRSRPELRSRVGRLID----------------------------- 173
O.aries             ------------ACLTCRNAARCRRRRRNENMEMERWSVSIKSVEG------------- 302
O.mykiss            -------------SRKKKKAQTARSNQGGKGAASSEENVQLT----------------- 283


R.norvegicus        -
M.musculus          -
M.unguiculatus      -
M.monax             -
C.griseus           -
P.troglodytes       -
P.anubis            -
C.troquatus         -
M.mulatta           -
M.nemestrina        -
C.aethiops          -
C.jacchus           -
O.cuniculus         F 330
S.scrofa            -
B.taurus            -
C.familiaris        -
A.melanoleuca       -
L.africana          -
pred.M.domestica    -
M.domestica         -
G.gallus            -
T.guttata           -
F.catus             -
O.aries             -
O.mykiss            -
```

Fig. 4E.3 Multiple sequence alignment for CD86 including the predicted and actual sequence of *M. domestica* CD86. Underlined = Ig-V domain. C = structurally important cysteines. **N** = putative N-linked glycosylation sites. S = predicted serine phosphorylation sites. Y = tyrosine phosphorylation sites. ▼ = putative PKA binding site.

Appendix 4E

Table 4E.1 Genbank accession numbers for CD86 and the relevant references.

Species	Common name	Accession number	References
Ailuropoda melanoleuca	Giant Panda	XM_002927223	Annotated
Bos taurus	Cattle	NM_001038017	(Harhay *et al.,* 2005)
Callithrix jacchus	White tuffed ear marmoset	XM_002758743	Annotated
Canis lupus familiaris	Dog	NM_001003146	(Yang and Sim, 1999)
Cercocebus torquatus atys	Sooty mangabey	AF344840	(Villinger *et al.,* 2001)
Cercopithecus aethiops	Green monkey	AF344861	(Villinger *et al.,* 2001)
Felis catus	Cat	NM_001009229	(Choi *et al.,* 2000b)
Gallus gallus	Chicken	NM_001037839	Direct submission
Homo sapiens	Human	NM_001206925, U04343 NM_175862	(Chen *et al.,* 1994) (Azuma *et al.,* 1993) (Chen *et al.,* 1994)
Loxodonta africana	African Elephant	XM_003412829	Annotated
Macaca mulatta	Rhesus monkey	NM_001042644	(Villinger *et al.,* 2001)
Macaca nemestrina	Pig-tailed macaque	AF344851	(Villinger *et al.,* 2001)
Marmota monax	Woodchuck	EU586564	Direct submission
Meriones unguiculatus	Mongolian gerbril	AY095931	Direct submission
Monodelphis domestica	South American grey short tailed opossum	XM_001371482	Annotated
Mus musculus	Mouse	NM_019388	(Borriello *et al.,* 1995)
Oncorhynchus mykiss	Rainbow trout	NM_001160477	(Zhang *et al.,* 2009)
Oryctolagus cuniculus	Rabbit	NM_001082208	(Isono and Seto, 1995)
Ovis aries	Sheep	DQ304077	(Terzo *et al.,* 2006)
Pan troglodytes	Chimpanzee	XM_001166230	Annotated
Papio anubis	Olive babboon	NM_001112636, AF344836	(Villinger *et al.,* 2001)
Pongo abelii	Orangutan	XM_002813228	Annotated
Rattus norvegicus	Norwegian Rat	NM_020081	
Sus scrofa	Pig	NM_214222 AY834754	(Maher *et al.,* 1996) (Choi *et al.,* 2006)
Taeniopygia guttata	Zebra Finch	XM_002197578	Annotated

Appendix 5A

Macropus eugenii TCRζ chain nucleotide sequence

atgcaattcctttccacagaggcccagagttttggactggcagacccaagactgtgctaccttctagatggca
tcctcttcatttatggagtcatcatcacggccctgttcctgagagcaaagttcagtaagactgccaaaatctc
cagctaccaacaggaccagaatcaactctacaatgagctctccccaggacgaagagaagaatatgacatttta
gataagagaagaggccgtgacccagagatgggaggaaaacagagaaggaagaatcctacagaaagcgtctaca
atgcactgcagaaagacaagatggcggatgcatacagtgagattggaatgaaaggagagaaccagcggagacg
aggcaaaggaaatgatgtcctgtaccagggcctcagtccagccaccaaggacacctatgatgctctccacatg
cagcccctgcctccccgttaa
3′end
aaagaacaaccatcactgctccactgaccaggtttgccaggacgcagc**attta**aaaccgacccccaccaaaaa
aaaaaaaaaaaaaaaggggggggggttttgtggggttttttaaaaaaaaaaaaaaaaaaaatgggtttatc
cggacgtaaatcaccttgtctgaacggaacatttcttcagtactttgtccccgtaaatacacatgacgagtat
ccggatcttggaaaatgccctcgccttctcgtaaatgcacaatattaaatgaataatccaaaccttccctgca
tgctgttttgacggtgccatccccatcgccgccgtcgtataacgaacccactattctggattgctcagctcat
attgattcgcgccatatttcgctttattttcttctataatcggcaataaccattgagataactgttcttgatc
tgcttgcataatgtcggcattaaaagcaagcctatcttgttgtgcgagctaccttgccagtccttgtaaagaa
cgagtggatttaaagttgcaagaatatcctgattccttttataccccgtgaaaacccacaatcgcactctgc
agcttcatttctggcttgtgattattgtctttataagtacgatggcctttaggaagcagagaaactttctcat
agctcacgccgtcagtatctctatttcagtgtaattcga

M. eugenii putative protein of TCRζ chain

MQFLSTEAQSFGLADPRLCYLLDGILFIYGVIITALFLRAKFSKTAKISSYQQDQNQLYNELSPGRREEYDIL
DKRRGRDPEMGGKQRRKNPTESVYNALQKDKMADAYSEIGMKGENQRRRGKGNDVLYQGLSPATKDTYDALHM
QPLPPR

Appendix 5A

Onychogalea fraenata TCRζ chain nucleotide sequence

atgaagtggaaggggattgttatcacagccgtcctgcaggcacgggtcccaattacagaggcccagagttttg
gactggcagacccaagactgtgctaccttctagatggcatcctcttcatttatggagtcatcatcacggccct
gttcctgagagcaaagttcagtaagattgccaaaatctccagctaccaacaggaccaaaatcaactctacaat
gagctttccccaggacgcagagaagaatatgacattttagataagagaagaggccgtgacccagagatgggag
gaaaacagagaaggaagaatcctacagaaaccgtctacaatgcactgcagaaagacaagatggcggacgcata
cagtgagattggaatgaaaggagagaaccagcggagacgaggcaaaggaaatgatgtcctgtaccagggcctc
agtccagccaccaaggacacctatgatgctctccacatgcagcccctgcctccccgt**taa**

O. fraenata TCRζ putative Protein

MKWKGIVITAVLQARVPITEAQSFGLADPRLCYLLDGILFIYGVIITALFLRAKFSKIAKISSYQQDQNQLYN
ELSPGRREEYDILDKRRGRDPEMGGKQRRKNPTETVYNALQKDKMADAYSEIGMKGENQRRRGKGNDVLYQGL
SPATKDTYDALHMQPLPPR

Appendix 5A

M. domestica TCRζ partial nucleotide sequence

```
tggctgaccccagactgtgttattttctagatggcatcctcttcatatatggagtcatcatcacggccctatt
cctaagagcaaagttctccaagactgccagagtttctgcctaccaacgagatcagaaccaagtctacaatgag
ctctctatgggacgaagagaagaatatgacattttagataagagaagaggaggccatgacccagagattggag
gaaaacagagaaggaagaatcctcaagaaaccgtgtacaattcactgcaaaaagacaagatggcagaagcata
cagtgagattggaatgaaaggcgagaaacagcggagacgtggcaaaggaaatgatgtcctgtaccagggcctc
agcccagccaccaaggacacctatgacgccctccacatgcagc
```

M. domestica putative amino acid sequence

```
ADPRLCYFLDGILFIYGVIITALFLRAKFSKTARVSAYQRDQNQVYNELSMGRREEYDILDKRRGGHDPEIGG
KQRRKNPQETVYNSLQKDKMAEAYSEIGMKGEKQRRRGKGNDVLYQGLSPATKDTYDALHMQ
```

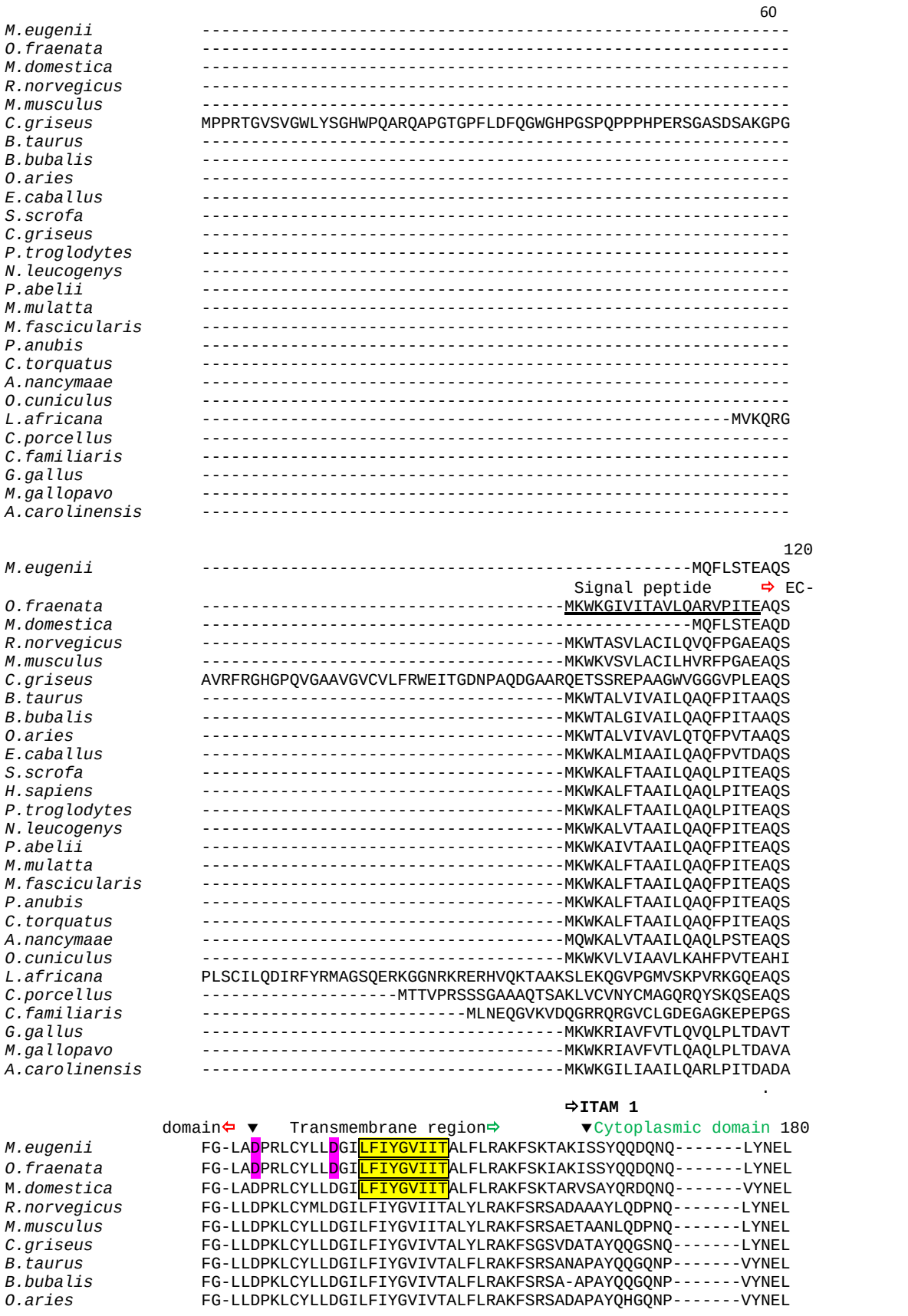
BLAT result in USCS for TCRζ *O. fraenata*

Input sequence. (*O. fraenata*)

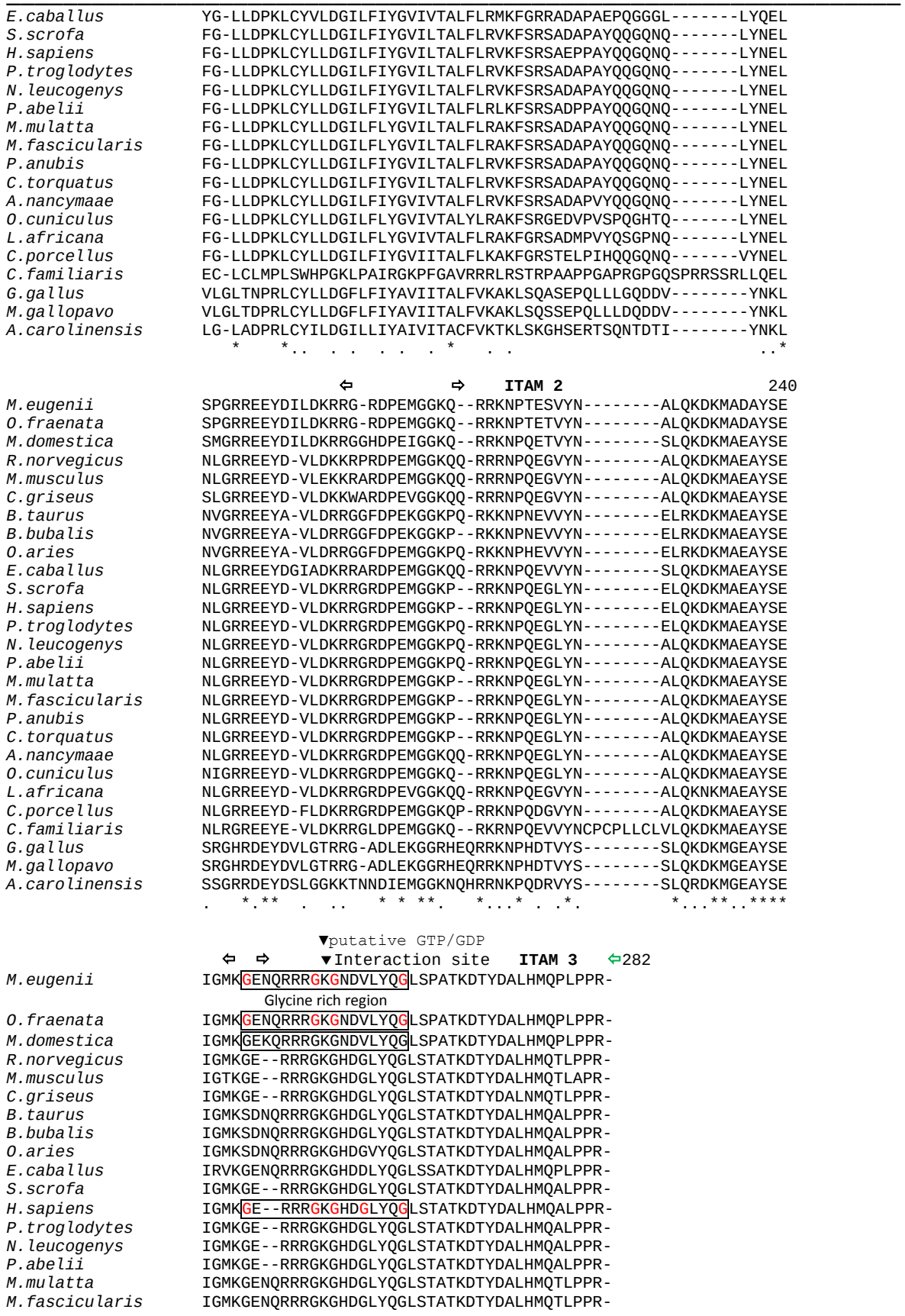
```
ATGAAGTGGAAGGGGATTGTTATCACAGCCGTCCTGCAGGCCCGGGTCCCAATTACAGAGGCCCAGAGTTTTG
GACTGGCAGACCCAAGACTGTGCTACCTTCTAGATGGCA/TCCTCTTCATTTATGGAGTCATCATCACGGCCC
TGTTCCTGAGAGCAAAGTTCAGTAAGATTGCCAAAAT/CTCCAGCTACCAACAGGACCAGAATCAACTCTACA
ATGAGCTTTCCCCAGGA/CGCAGAGAAGAATATGACATTTTAGATAAGAGAAGAGGCCGT/GACCCAGAGATG
GGAGGAAAACAGAGAAGGAAGAATCCTACAGAAACCGTCTACAATGCACTGCAGAAAGACAAGATGGCGGACG
CATACAGTGAGATTGGAATGAAAGGAGAGAACCAGCGGAGACGAGGCAAAGGAAATGATGTCCTGTACCAGGG
CCTCAGTCCAGCCACCAAGGACACCTATGATGCTCTCCACATGCAGCCCCTGCCTCCCCGTTAA
```

Figure 5A.1. **/** Exon boundaries according to UCSC.

Appendix 5A


60
M.eugenii
O.fraenata
M.domestica
R.norvegicus
M.musculus
C.griseus MPPRTGVSVGWLYSGHWPQARQAPGTGPFLDFQGWGHPGSPQPPPHPERSGASDSAKGPG
B.taurus
B.bubalis
O.aries
E.caballus
S.scrofa
C.griseus
P.troglodytes
N.leucogenys
P.abelii
M.mulatta
M.fascicularis
P.anubis
C.torquatus
A.nancymaae
O.cuniculus
L.africana -------MVKQRG
C.porcellus
C.familiaris
G.gallus
M.gallopavo
A.carolinensis
120
M.eugenii -------MQFLSTEAQS
Signal peptide
EC-
O.fraenata -------MKWKGIVITAVLQARVPITEAQS
M.domestica -------MQFLSTEAQD
R.norvegicus -------MKWTASVLACILQVQFPGAEAQS
M.musculus -------MKWKVSVLACILHVRFPGAEAQS
C.griseus AVRFRGHGPQVGAAVGVCVLFRWEITGDNPAQDGAARQETSSREPAAGWVGGGVPLEAQS
B.taurus -------MKWTALVIVAILQAQFPITAAQS
B.bubalis -------MKWTALGIVAILQAQFPITAAQS
O.aries -------MKWTALVIVAVLQTQFPVTAAQS
E.caballus -------MKWKALMIAAILQAQFPVTDAQS
S.scrofa -------MKWKALFTAAILQAQLPITEAQS
H.sapiens -------MKWKALFTAAILQAQLPITEAQS
P.troglodytes -------MKWKALFTAAILQAQLPITEAQS
N.leucogenys -------MKWKALVTAAILQAQFPITEAQS
P.abelii -------MKWKAIVTAAILQAQFPITEAQS
M.mulatta -------MKWKALFTAAILQAQFPITEAQS
M.fascicularis -------MKWKALFTAAILQAQFPITEAQS
P.anubis -------MKWKALFTAAILQAQFPITEAQS
C.torquatus -------MKWKALFTAAILQAQFPITEAQS
A.nancymaae -------MQWKALVTAAILQAQLPSTEAQS
O.cuniculus -------MKWKVLVIAAVLKAHFPVTEAHI
L.africana PLSCILQDIRFYRMAGSQERKGGNRKRERHVQKTAAKSLEKQGVPGMVSKPVRKGQEAQS
C.porcellus -------MTTVPRSSSGAAAQTSAKLVCVNYCMAGQRQYSKQSEAQS
C.familiaris -------MLNEQGVKVDQGRRQRGVCLGDEGAGKEPEPGS
G.gallus -------MKWKRIAVFVTLQVQLPLTDAVT
M.gallopavo -------MKWKRIAVFVTLQAQLPLTDAVA
A.carolinensis -------MKWKGILIAAILQARLPITDADA
⇨ITAM 1
domain⇦ ▼ Transmembrane region⇨ ▼Cytoplasmic domain 180
M.eugenii FG-LADPRLCYLLDGILFIYGVIITALFLRAKFSKTAKISSYQQDQNQ-------LYNEL
O.fraenata FG-LADPRLCYLLDGILFIYGVIITALFLRAKFSKIAKISSYQQDQNQ-------LYNEL
M.domestica FG-LADPRLCYLLDGILFIYGVIITALFLRAKFSKTARVSAYQRDQNQ-------VYNEL
R.norvegicus FG-LLDPKLCYMLDGILFIYGVIITALYLRAKFSRSADAAAYLQDPNQ-------LYNEL
M.musculus FG-LLDPKLCYLLDGILFIYGVIITALYLRAKFSRSAETAANLQDPNQ-------LYNEL
C.griseus FG-LLDPKLCYLLDGILFIYGVIVTALYLRAKFSGSVDATAYQQGSNQ-------LYNEL
B.taurus FG-LLDPKLCYLLDGILFIYGVIVTALFLRAKFSRSANAPAYQQGQNP-------VYNEL
B.bubalis FG-LLDPKLCYLLDGILFIYGVIVTALFLRAKFSRSA-APAYQQGQNP-------VYNEL
O.aries FG-LLDPKLCYLLDGILFIYGVIVTALFLRAKFSRSADAPAYQHGQNP-------VYNEL

Appendix 5A


E.caballus YG-LLDPKLCYVLDGILFIYGVIVTALFLRMKFGRRADAPAEPQGGGL-------LYQEL
S.scrofa FG-LLDPKLCYLLDGILFIYGVILTALFLRVKFSRSADAPAYQQGQNQ-------LYNEL
H.sapiens FG-LLDPKLCYLLDGILFIYGVILTALFLRVKFSRSAEPPAYQQGQNQ-------LYNEL
P.troglodytes FG-LLDPKLCYLLDGILFIYGVILTALFLRVKFSRSADAPAYQQGQNQ-------LYNEL
N.leucogenys FG-LLDPKLCYLLDGILFIYGVILTALFLRVKFSRSADAPAYQQGQNQ-------LYNEL
P.abelii FG-LLDPKLCYLLDGILFIYGVILTALFLRLKFSRSADPPAYQQGQNQ-------LYNEL
M.mulatta FG-LLDPKLCYLLDGILFLYGVILTALFLRAKFSRSADAPAYQQGQNQ-------LYNEL
M.fascicularis FG-LLDPKLCYLLDGILFLYGVILTALFLRAKFSRSADAPAYQQGQNQ-------LYNEL
P.anubis FG-LLDPKLCYLLDGILFIYGVILTALFLRVKFSRSADAPAYQQGQNQ-------LYNEL
C.torquatus FG-LLDPKLCYLLDGILFIYGVILTALFLRVKFSRSADAPAYQQGQNQ-------LYNEL
A.nancymaae FG-LLDPKLCYLLDGILFIYGVIVTALFLRVKFSRSADAPVYQQGQNQ-------LYNEL
O.cuniculus FG-LLDPKLCYLLDGILFLYGVIVTALYLRAKFSRGEDVPVSPQGHTQ-------LYNEL
L.africana FG-LLDPKLCYLLDGILFLYGVIVTALFLRAKFGRSADMPVYQSGPNQ-------LYNEL
C.porcellus FG-LLDPKLCYLLDGILFIYGVIITALFLKAKFGRSTELPIHQQGQNQ-------VYNEL
C.familiaris EC-LCLMPLSWHPGKLPAIRGKPFGAVRRRLRSTRPAAPPGAPRGPGQSPRRSSRLLQEL
G.gallus VLGLTNPRLCYLLDGFLFIYAVIITALFVKAKLSQASEPQLLLGQDDV--------YNKL
M.gallopavo VLGLTDPRLCYLLDGFLFIYAVIITALFVKAKLSQSSEPQLLLDQDDV--------YNKL
A.carolinensis LG-LADPRLCYILDGILLIYAIVITACFVKTKLSKGHSERTSQNTDTI--------YNKL
ITAM 2
240
M.eugenii SPGRREEYDILDKRRG-RDPEMGGKQ--RRKNPTESVYN--------ALQKDKMADAYSE
O.fraenata SPGRREEYDILDKRRG-RDPEMGGKQ--RRKNPTETVYN--------ALQKDKMADAYSE
M.domestica SMGRREEYDILDKRRGGHDPEIGGKQ--RRKNPQETVYN--------SLQKDKMAEAYSE
R.norvegicus NLGRREEYD-VLDKKRPRDPEMGGKQQ-RRRNPQEGVYN--------ALQKDKMAEAYSE
M.musculus NLGRREEYD-VLEKKRARDPEMGGKQQ-RRRNPQEGVYN--------ALQKDKMAEAYSE
C.griseus SLGRREEYD-VLDKKWARDPEVGGKQQ-RRRNPQEGVYN--------ALQKDKMAEAYSE
B.taurus NVGRREEYA-VLDRRGGFDPEKGGKPQ-RKKNPNEVVYN--------ELRKDKMAEAYSE
B.bubalis NVGRREEYA-VLDRRGGFDPEKGGKP--RKKNPNEVVYN--------ELRKDKMAEAYSE
O.aries NVGRREEYA-VLDRRGGFDPEMGGKPQ-RKKNPHEVVYN--------ELRKDKMAEAYSE
E.caballus NLGRREEYDGIADKRRARDPEMGGKQQ-RRKNPQEVVYN--------SLQKDKMAEAYSE
S.scrofa NLGRREEYD-VLDKRRGRDPEMGGKP--RRKNPQEGLYN--------ELQKDKMAEAYSE
H.sapiens NLGRREEYD-VLDKRRGRDPEMGGKP--RRKNPQEGLYN--------ELQKDKMAEAYSE
P.troglodytes NLGRREEYD-VLDKRRGRDPEMGGKPQ-RRKNPQEGLYN--------ELQKDKMAEAYSE
N.leucogenys NLGRREEYD-VLDKRRGRDPEMGGKPQ-RRKNPQEGLYN--------ALQKDKMAEAYSE
P.abelii NLGRREEYD-VLDKRRGRDPEMGGKPQ-RRKNPQEGLYN--------ALQKDKMAEAYSE
M.mulatta NLGRREEYD-VLDKRRGRDPEMGGKP--RRKNPQEGLYN--------ALQKDKMAEAYSE
M.fascicularis NLGRREEYD-VLDKRRGRDPEMGGKP--RRKNPQEGLYN--------ALQKDKMAEAYSE
P.anubis NLGRREEYD-VLDKRRGRDPEMGGKP--RRKNPQEGLYN--------ALQKDKMAEAYSE
C.torquatus NLGRREEYD-VLDKRRGRDPEMGGKP--RRKNPQEGLYN--------ALQKDKMAEAYSE
A.nancymaae NLGRREEYD-VLDKRRGRDPEMGGKQQ-RRKNPQEGLYN--------ALQKDKMAEAYSE
O.cuniculus NIGRREEYD-VLDKRRGRDPEMGGKQ--RRKNPQEGLYN--------ALQKDKMAEAYSE
L.africana NLGRREEYD-VLDKRRGRDPEVGGKQQ-RRKNPQEGVYN--------ALQKNKMAEAYSE
C.porcellus NLGRREEYD-FLDKRRGRDPEMGGKQP-RRKNPQDGVYN--------ALQKDKMAEAYSE
C.familiaris NLRGREEYE-VLDKRRGLDPEMGGKQ--RKRNPQEVVYNCPCPLLCLVLQKDKMAEAYSE
G.gallus SRGHRDEYDVLGTRRG-ADLEKGGRHEQRRKNPHDTVYS--------SLQKDKMGEAYSE
M.gallopavo SRGHRDEYDVLGTRRG-ADLEKGGRHEQRRKNPHDTVYS--------SLQKDKMGEAYSE
A.carolinensis SSGRRDEYDSLGGKKTNNDIEMGGKNQHRRNKPQDRVYS--------SLQRDKMGEAYSE
putative GTP/GDP
Interaction site
ITAM 3
282
M.eugenii IGMKGENQRRRGKGNDVLYQGLSPATKDTYDALHMQPLPPR-
Glycine rich region
O.fraenata IGMKGENQRRRGKGNDVLYQGLSPATKDTYDALHMQPLPPR-
M.domestica IGMKGEKQRRRGKGNDVLYQGLSPATKDTYDALHMQPLPPR-
R.norvegicus IGMKGE--RRRGKGHDGLYQGLSTATKDTYDALHMQTLPPR-
M.musculus IGTKGE--RRRGKGHDGLYQGLSTATKDTYDALHMQTLAPR-
C.griseus IGMKGE--RRRGKGHDGLYQGLSTATKDTYDALNMQTLPPR-
B.taurus IGMKSDNQRRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
B.bubalis IGMKSDNQRRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
O.aries IGMKSDNQRRRGKGHDGVYQGLSTATKDTYDALHMQALPPR-
E.caballus IRVKGENQRRRGKGHDDLYQGLSSATKDTYDALHMQPLPPR-
S.scrofa IGMKGE--RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
H.sapiens IGMKGE--RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
P.troglodytes IGMKGE--RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
N.leucogenys IGMKGE--RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
P.abelii IGMKGE--RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
M.mulatta IGMKGENQRRRGKGHDGLYQGLSTATKDTYDALHMQTLPPR-
M.fascicularis IGMKGENQRRRGKGHDGLYQGLSTATKDTYDALHMQTLPPR-

```
P.anubis            IGMKGENQRRRGKGHDGLYQGLSTATKDTYDALHMQTLPPR-
C.torquatus         IGMKGENQRRRGKGHDGLYQGLSTATKDTYDALHMQTLPPR-
A.nancymaae         IGMKGENQRRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
O.cuniculus         IGMKGENQRRRGKGHDGLYQGLSAATKDTYDALHMQTLPPR-
L.africana          IGMKGENQRRRGKGQDGLYQGLSTPTKDTYDALHMQALPPR-
C.porcellus         IGMKGE--RRRAKGQDGLYQGLSTATKDTYDALHMQTLPPR-
C.familiaris        IGIKSE--RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR-
G.gallus            IGKKGEQ-RRRGKGNDAVYQGLSAATRDTYDALHMQPLPPR-
M.gallopavo         IGKKGEQQRRRGKGNEAVYQVGNTSTAAPFIYLCS-------
A.carolinensis      IGKKGER--RRGKGNDAVYQGLSAATKDTYDVLQMQPMQTPY
                    *  *..   **.**.. .**  ...*  ..  *
```

Figure 5A.2. Multiple sequence alignment for identification of structural and functional elements. Highlighted and boxed in yellow is the dimerization motif. Boxed is the glycine rich region that represents the GTP binding site. Domains are signified with⇨⇦.

Appendix 5A

Table 5A.1. Genbank Accession numbers for TCRζ and the relevant references.

Species name	Common name	Accession Number	References
Anolis carolinensis	Green anole	XM_003219127.1	Annotated
Aotus nancymaae	Ma's night monkey	EF656481.1	Direct submission
Bos taurus	Cattle	NM_174012.2	(Hagens *et al.,* 1996)
Bubalus bubalis	Water buffolo	DQ057984.1	Direct submission
Canis lupus familiaris	Dog	XM_849255.2	Annotated
Cavia porcellus	Domestic Guinea Pig	XM_003468676.1	Annotated
Cercocebus torquatus	Red crowend managaby (variant 2)	DQ437666.1	(Rogers *et al.,* 2006)
Cercocebus torquatus	Red crowned mangaby (variant 1)	DQ437665.1	(Rogers *et al.,* 2006)
Cricetulus griseus	Chinese hamster	XM_003502297.1	Annotated
Equus caballus	Horse	XM_003364910.1 DQ885232.1	Annotated (Horohov *et al.,* 2005)
Gallus gallus	Chicken	NM_206879.1	(Gobel and Bolliger, 1998)
Homo sapiens	Human (variant 2)	NM_000734.3	(Jensen *et al.,* 1992)
Homo sapiens	Human (variant 1)	J04132.1	(Weissman *et al.,* 1998)
Loxodonta africana	African elephant	XM_003415037.1	Annotated
Macaca fascicularis	Crab eating macaque (variant 2)	DQ437668.1	(Rogers *et al.,* 2006)
Macaca fascicularis	Crab eating macaque (variant 1)	DQ437667.1	(Rogers *et al.,* 2006
Macaca mulatta	Rhesus monkey (variant 2)	NM_001077423.1	(Rogers *et al.,* 2006)
Macaca mulatta	Rhesus monkey (variant 1)	DQ437669.1	(Rogers *et al.,* 2006)
Meleagris gallopavo	Turkey	XM_003202846.1	Annotated
Monodelphis domestica	Opossum	XM_001371336.1	Annotated
Mus musculus	Mouse	NM_001113391.1 M19729	(Clayton *et al.,* 1991)
Nomascus leucogenys	Northern white-cheeked gibbon	XM_003258814	Annotated
Oryctolagus cuniculus	Rabbit	NM_001082002.1 AB035152.1	 Direct submission
Ovis aries	Sheep	NM_001009417.1	(Hein and Tunnacliffe, 1993)
Pan troglodytes	Chimpanzee (variant 2)	XM_001174745.1	Annotated
Pan troglodytes	Chimpanzee (variant 1)	XM_001174731.1	Annotated
Pan troglodytes	Chimpanzee	XM_001174731.2	Annotated
Papio Anubis	Olive Baboon (variant 2)	DQ437664.1	(Rogers *et al.,* 2006)
Papio Anubis	Olive Baboon (variant 1)	NM_001112652.1	(Rogers *et al.,* 2006)
Pongo abelii	Sumatran Orangautan	XM_002809835.1	Annotated
Rattus norvegicus	Rat	NM_170789.1 L08447	 (Itoh *et al.,* 1993)
Sus scrofa	Pig	AF153830.1	Unpublished

Appendix 5B

Macropus eugenii ZAP-70 nucleotide sequence

5’ end

aatttgtctctcgtgggggggccgggacgttatttcctctttttggccccctagctacactccttttagat
cgacagtaaacagattcgttgacagagggtggtagctagagccagtggtggctgggcctagaccctgtccgct
aactgccttccccaccctgagaccctggaacccagggtagcctttccctccatggaacttggccctccaagag
actgaagaagactgagggagccaagaaagcctggttccttgggcctagcgcctccggaacctccacaggacca
tttctccttgtgggaaggaactggaaagcctctcacccttgtcagttttgctccagcctagacaccatcaag
tggtgccctcagggagtccacg
atgccagatgcagctgcccatttgccctttttttacgggagcatctcgagggcggaggccgaggagcacctga
agctggcaggcatggcggatgggcttttcctgctccgccagtgcctccgtagtttgggggggtatgtgctctc
attggtatacgacctgcacatccatcattaccccatcgagcgtcagctgaacggcacctatgccattgctggg
ggcaagcctcattgcggcccggctgagctctgtgagttttactccaaggatgctgatggcctcccctgtgctt
tacgcaagccttgtaacaggcccagtgggatggagccccagccaggtgtctttgacagttttcgggacagcat
ggttcgagactatgtgcgccagacctggaaactagagggtgatgcccttgagcaggccatcatcagccaggcc
ccccaggtagagaagctcattgccaccacagcccatgagcggatgccttggtaccacagctccatctccagag
aggaagcaaaacgcaaactctactcaggctcccagcatgatggcaagttcttgcttaaacccaggaaggaaca
aggcagctacgctttgtccctcatcaatggcaaaactgtctaccactacctcatcaaccaggacaagtctggc
aagtactgtattcctgagggcacaaagtttgacaccttgtggcagctggtaaagtatctgaagctgaaggcaa
atgggcttatctactgtctgaaggagatttgtcctaatgccagtgcttctactgctactgtgactgctgctcc
cacactccctgtccatccctccatgcctagaaggaatgacaccctcaactctgatggatacacccctgagcca
gcatgtttaaacaagagtcaaggtgagaagtctcgggtcctgcccatggacaccagtgtgtatgagagcccct
acagtgatcccgaagagctcaaggacaagaaactcttcctcaagagagagaatctgatgatcgatgaggtgga
gctgggctcaggcaactttggctgtgtccgcaaggggtctacaagatgaggaagaagcagattgatgtggcc
atcaaggtgcttaagagtaccaatgagaaggctgagaaggacgagatgatgaaggaggcccagatcatgcacc
tgctggacaacccctacatcgtgcggatcatcggcgtgtgcaaggctgaggccctcatgctcgtcatggagat
ggccatcgcggggcctctgcacaagtttctggccgccaagaaggaggaggtccctgtcagcaatgtcgtggag
ctgctgcaccaggtggccatgggaatgaaatacctagaagaaaaaaattttgtgcaccgtgacctggctgccc
gaaatgttcttctggtcaaccagcactatgccaagattagtgactttggtttatccaaggcactgggggctga
tgacagctactacaccgcccgctctgcagggaaatggccactcaaatggtatgctccagagtgcatcaactac
cggaaattctcctgccaaagcgacgtgtggacctatggattcaccatgtggaaacctttcacctatggccaga
acccttataagaaaatgaaaggccctgaggtcctcaaattcattgaaaagggtaagcggatggatccccctcc
tgagtgcccaccaaacatgtacacactcatgaaaaaatgctggatatacaaatgggaacatcgtccaaacttc
ccatatgtggaacagcccattaaaacctactattaccgcctggccagtaaggcggaaaaggtcttatatgccc
ctcaatcagaggggctactcctgcctga
3’ end

cgaccatctccctttctctggatcttcacaattcttcccagaacctggctttgctctttgtttttcccagaa
aagtgttgtccccccagcctctcccatgccagcaccaccaaaccctcctcccttgcctacatgcctctacttt
cctccccagttgaacatgctccctccacgcaaaggacccctgactcccaactttgaggtggatgatgcacata
aagataggcctattccagcctccattgtcttaaataccccggtttcaaaaacctgttactcctatctgaaaa
ttaccgtcggccgtcattaat

Appendix 5B

M. eugenii ZAP-70 putative protein

MPDAAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVYDLHIHHYPIERQLNGTYAIAG
GKPHCGPAELCEFYSKDADGLPCALRKPCNRPSGMEPQPGVFDSFRDSMVRDYVRQTWKLEGDALEQAIISQA
PQVEKLIATTAHERMPWYHSSISREEAKRKLYSGSQHDGKFLLKPRKEQGSYALSLINGKTVYHYLINQDKSG
KYCIPEGTKFDTLWQLVKYLKLKANGLIYCLKEICPNASASTATVTAAPTLPVHPSMPRRNDTLNSDGYTPEP
ACLNKSQGEKSRVLPMDTSVYESPYSDPEELKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYKMRKKQIDVA
IKVLKSTNEKAEKDEMMKEAQIMHLLDNPYIVRIIGVCKAEALMLVMEMAIAGPLHKFLAAKKEEVPVSNVVE
LLHQVAMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYTARSAGKWPLKWYAPECINY
RKFSCQSDVWTYGFTMWKPFTYGQNPYKKMKGPEVLKFIEKGKRMDPPPECPPNMYTLMKKCWIYKWEHRPNF
PYVEQPIKTYYYRLASKAEKVLYAPQSEGATPA

Onychogalea fraenata partial ZAP-70 nucleotide sequence

```
tctactcaggctcccagcatgatggcaagttcttgcttaaacccaggaaggaacaaggcacctacgctttgtc
cctcatctatggcaaaactgtctaccactacctcatcaaccaggacaagtctggcaagtactgtattcctgag
ggcacaaagtttgacaccttgtggcagctggtaaagtatctgaagctgaaggcaaatgggcttatctactgtc
tgaaggagatttgtcctaatgccagtgcttctactgctactgtgactgctgctcccacactccctgtccatcc
ctccatgcctagaaggaatgacaccctcaactctgatggatacacccctgacccagcatgtttaaacaaaagt
caaggtgaaaagtctcgggtcctgcccatggacaccagtgtgtatgaaagcccttacagtgaccccaaaaagc
tcaaggacaaaaaactcttcctcaaaagagagaatctgatgattgatgaggtggagctgggctcaggcaactt
tggctgtgtccgcaaggggtctacaagatgaggaaaaagcaaattgatgtacccatcaaggtgcttaaaagt
accaatgaaaaggctgaaaaggacaagatgatgaaggaggcccaaatcatgcaccagctggacaaccccctaca
tcgtgcgtatcatcggcgtgtgcaaggctgaggccctcatgctcgtcatgaagatggccatcgcggggcctct
gcacaagttcctggccgccaagaaggaggaggtccctgtaagcaatgttgtggagctactgcaccaggtggcc
atgggaatgaaatacctggaagaaaaaaattttgtgcaccgtgacctggctgcccgaaatgttcttctggtca
accagcactatgccaagattagtgactttggtttatccaaggcactgggggctgatgacagctactacaccgc
ccgctctgcagggaagtggccactcaaatggtatgccccagagtgcatcaactatcggaaattctccagccga
agcgatgtgtggagctatggagtcaccatgtgggaagctttcacctatggccagaagccttataagaaaatga
aaggccctgaggtcatcaaattcattgaagagggtaagcggatggatcgccctcctgagtgcccaccagacat
gtacacgctcatgaaagactgctggatatacaagtgggaagatcgtccaaacttctcagatgtggaacagcgc
attagaacctactattacagcctggccagtaaggcggaagcggttttagatgcccctcaagcagaggggctca
ctagtgcctga
3'end
ggatcatctccctttctctggatcttcccaattcttcccaaaacctggccttgctctttgctttttcccagag
aggcttgtccccctgcctctcccatgccagcaagaccactccttcccccttccctagatgcctgaaactcctt
cccatttgcatctgcttcctcccagcagaggactgaaaactccagactttgaggtgaatgttctgcaggaaag
ttaggcccctttctgcatctttttccttgatactcagggcttggagatcttggaaggcacaactggctgtatc
cttctgccttcttctttgatggatttggtcctcacccctgggctcctccctactgtggttggatgggatctgg
tgggg
```

O. fraenata putative protein sequence

```
YSGSQHDGKFLLKPRKEQGTYALSLIYGKTVYHYLINQDKSGKYCIPEGTKFDTLWQLVKYLKLKANGLIYCL
KEICPNASASTATVTAAPTLPVHPSMPRRNDTLNSDGYTPDPACLNKSQGEKSRVLPMDTSVYESPYSDPKKL
KDKKLFLKRENLMIDEVELGSGNFGCVRKGVYKMRKKQIDVPIKVLKSTNEKAEKDKMMKEAQIMHQLDNPYI
VRIIGVCKAEALMLVMKMAIAGPLHKFLAAKKEEVPVSNVVELLHQVAMGMKYLEEKNFVHRDLAARNVLLVN
QHYAKISDFGLSKALGADDSYYTARSAGKWPLKWYAPECINYRKFSSRSDVWSYGVTMWEAFTYGQKPYKKMK
GPEVIKFIEEGKRMDRPPECPPDMYTLMKDCWIYKWEDRPNFSDVEQRIRTYYYSLASKAEAVLDAPQAEGAT
SA
```

Appendix 5B

M. domestica ZAP-70 (partial nucleotide sequence)

```
cgcaaactctactcaggctcccagcatgatggcaagttcttgcttaaacccaggaaggaacaaggcagctacg
ctttgtccctcatctatggcaaaactgtctatcactacctcatcaaccaggacaagtctggcaagtactgtat
tcctgagggcacaaagtttgacaccttgtggcagctggtaaagtatctgaagctgaaggcaaatgggcttatc
tactgtctgaaggagatttgtcctaatgccagtgcttctactgctactgtgactgctgctcccacactccctg
tccatccctccatgcctagaaggaatgacaccctcaactctgatggatacacccctgagccagcatgtttaaa
caagagtcaaggtgagaagtctcgggtcctgcccatggacaccagtgtgtatgaaagcccctacagtgatccc
gaaaagctcaaggacaagaaactcttcctcaagagagagaatctgatgatcgatgaggtggagctgggctcag
gcaactttggctgtgtccgcaaggggtctacaagatgaggaagaagcagattgatgtggccatcaaggtgct
taagagtaccaatgagaaggctgagaaggacgagatgatgaaggaggcccagatcatgcacctgctggacaac
ccctacatcgtgcggatcatcggcgtgtgcaaggctgaggccctcatgctcgtcatggagatggccatcgcgg
ggcctctgcacaagtttctggccgccaagaaggaggaggtccctgtcagcaatgtcgtgagctgctgcaccaa
gtggccatgggaatgaaatacctagaagaaaaaaattttgtgcaccgtgacctggctgcccgaaatgttcttc
tggtcaaccagcactatgccaagattagtgactttggtttatccaaggcactgggggctgatgacagctacta
caccgcccgctctgcagggaagtggccactcaaatggtatgccccagagtgcatcaactacaggaaatctcca
gcc
```

Putative protein *M. domestica* ZAP-70 (partial sequence)

```
RKLYSGSQHDGKFLLKPRKEQGSYALSLIYGKTVYHYLINQDKSGKYCIPEGTKFDTLWQLVKYLKLKANGLI
YCLKEICPNASASTATVTAAPTLPVHPSMPRRNDTLNSDGYTPEPACLNKSQGEKSRVLPMDTSVYESPYSDP
EKLKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYKMRKKQIDVAIKVLKSTNEKAEKDEMMKEAQIMHLLDN
PYIVRIIGVCKAEALMLVMEMAIAGPLHKFLAAKKEEVPVSNVVSCCTKWPWE
```

Appendix 5B

L. hirsutus ZAP-70 (partial nucleotide sequence)

```
gacaagtctggcaagtactgtattcctgagggcacaaagtttgacaccttgtggcagctggtggagtatctga
agctgaaggcagatgggcttatctactgtctgaaggagatttgtcctaatgccagtgcttctactgctactgt
gactgctgctcccacactccctgtccatccctctatgcctagaaggaatgacaccctcaactctgatggatat
acccctgagccagcatgtttaaacaagagtcaaggtgagaagtctcgggtcctgcccatggacaccagtgtgt
atgagagcccctacagtgaccccgaagagctcaaggacaagaaactcttcctcaagagagagaatctgatgat
cgatgaggtggagctgggctcaggcaactttggctgtgtccgcaaggggtctacaggatgaggaagaagcag
attgatgtagccatcaaggtgcttaagagtaccaatgagaaggctgagaaggacgagatgatgaaggaggccc
agatcatgcaccagctggacaacccctacatagtgcgtatcatcggcgtgtgcaaggctgaggccctcatgct
cgtcatggagatggccatcgcggggcctctgcacaagttcctggctgccaagaaaggaggaggtccctgtaag
caatgtcgtggagctgctgcaccaagtggccatgggaatgaaatacctagaagaaaaaaattttgtgcaccgt
gacctggctgcccgaaatgttcttctggtcaaccagcactatgccaagattagtgactttgggtttatccaag
gcactgggggctgatgacagctactacaccgcccgctctgcagggaagtggccactcaaatggtatgccccag
agtgcatc
```

L. hirsutus putative partial amino acid sequence

```
DKSGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEICPNASASTATVTAAPTLPVHPSMPRRNDTLNSDGY
TPEPACLNKSQGEKSRVLPMDTSVYESPYSDPEELKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYRMRKKQ
IDVAIKVLKSTNEKAEKDEMMKEAQIMHQLDNPYIVRIIGVCKAEALMLVMEMAIAGPLHKFLAAKKGGGPCK
QCRGAAAPSGHGNEIPRRKKFCAP
```

Appendix 5B

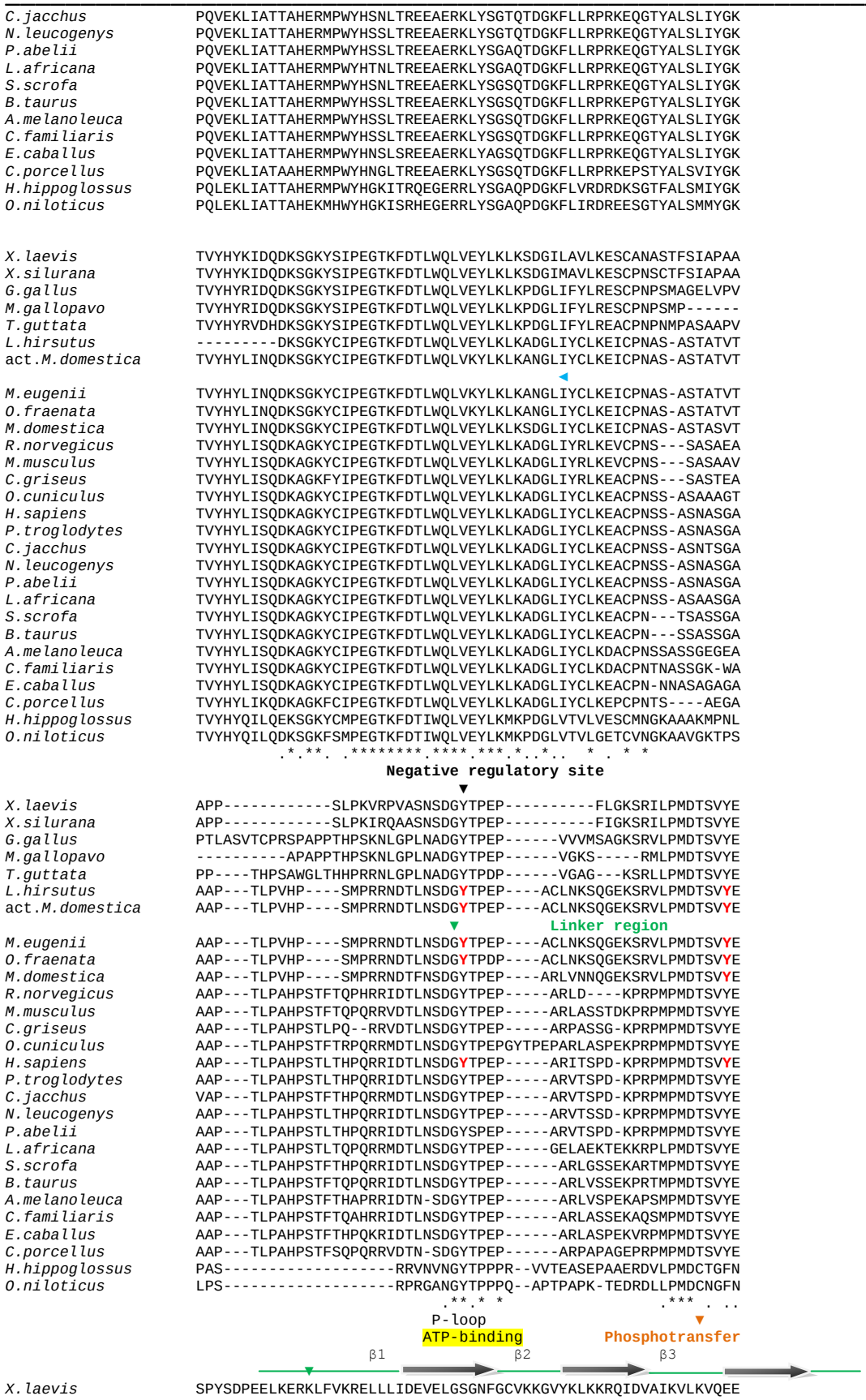
Sequence alignment of known ZAP-70 sequences

```
X.laevis            ------------------------------------------------------------
X.silurana          ------------------------------------------------------------
G.gallus            ------------------------------------------------------------
M.gallopavo         ------------------------------------------------------------
T.guttata           ------------------------------------------------------------
L.hirsutus          ------------------------------------------------------------
act.M.domestica     ------------------------------------------------------------
M.eugenii           ------------------------------------------------------------
O.fraenata          ------------------------------------------------------------
M.domestica         ------------------------------------------------------------
R.norvegicus        ------------------------------------------------------------
M.musculus          ------------------------------------------------------------
C.griseus           ------------------------------------------------------------
O.cuniculus         ------------------------------------------------------------
H.sapiens           ------------------------------------------------------------
P.troglodytes       ------------------------------------------------------------
C.jacchus           -------------------------------------------------------------M
N.leucogenys        ------------------------------------------------------------
P.abelii            MSPGTPLLPTTAAYPPVPGNRLVSDKPACRSSADTGPSGQAWPTVGLRAAAGAFRTGSPL
L.africana          ------------------------------------------------------------
S.scrofa            ------------------------------------------------------------
B.taurus            ------------------------------------------------------------
A.melanoleuca       ------------------------------------------------------------
C.familiaris        ------------------------------------------------------------
E.caballus          MRPKQQMGVTGVHPSGDPASRRALSACLVPTGQRYQCSRQLESRLVGEQIVRLDERVGAQ
C.porcellus         ------------------------------------------------------------
H.hippoglossus      ------------------------------------------------------------
O.niloticus         ------------------------------------------------------------


X.laevis            ----------------------------MPDVAGHLPFYYGSISRADAEEYLKLGG-MMD
X.silurana          ----------------------------MPDAAGHLPFYYGSISRADAEEYLKLGG-MMD
G.gallus            ----------------------------MPDAAAHLPFYYGSIARSEAEEYLKLAG-MSD
M.gallopavo         ----------------------------MPDAAAHLPFYYGSIARSEAEEYLKLAG-MSD
T.guttata           ----------------------------MPDAAAHLPFYYGSIARSEAEEHLKLGG-MAD
L.hirsutus          ------------------------------------------------------------
act.M.domestica     ------------------------------------------------------------
                                                    ▶SH2 domain-N TCRζ and CD3
                                                                 interaction
M.eugenii           ----------------------------MPDAAAHLPFFYGSISRAEAEEHLKLAGMADG
O.fraenata          ------------------------------------------------------------
M.domestica         ----------------------------MPDAAAHLPFFYGSISRAEAEDHLKLAGMADG
R.norvegicus        ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
M.musculus          ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
C.griseus           ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
O.cuniculus         ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
H.sapiens           ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
P.troglodytes       ------------------------------------------------------------
C.jacchus           ALGPETRQVACLLGHPPVGPQVWGGAGAMPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
N.leucogenys        ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
P.abelii            ALGPETPQVACLPGHPPVRPQVWGGPGAMPDPAAHLPFFYGSISRAEAEEHLKLAG-MGD
L.africana          ----------------------------MPDPAAHLPFFYGSISRAEAEDHLKLAG-MSD
S.scrofa            ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
B.taurus            ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
A.melanoleuca       ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
C.familiaris        ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAG-MAD
E.caballus          EALATTDVYSLGTGEAQELAPCTGGGPDIPETARTCPSSTAGISRAEAEEAPGSWRGMAD
C.porcellus         ----------------------------MPDPAAHLPFFYGSISRAEAEEHLKLAACFNG
H.hippoglossus      ---------------------------MSTDPAAELPFYYGSISRTDAEQHLKLAG-MVD
O.niloticus         ---------------------------MSIDPAAELPFFYGSISRSEAEQHLKLAG-MSD
```

Appendix 5B

```
X.laevis            GLFLLRQCLR-----TLGGYVLSMVYNVHFHHYPVERQLNGTYAIAGGKAHCGPAELCEY
X.silurana          GLFLLRQCLR-----TLGGYVLSMVYNLHFHHYPVERQLNGTYAIAGGKAHCGPAELCEY
G.gallus            GLFLLRQCLR-----SLGGYVLSMVCNLQFYHYAIERQMNGTYAIAGGKAHCGPEELCEF
M.gallopavo         GLFLLRQCLR-----SLGGYVLSMVCNLQFYHYAIERQMNGTYAIAGGKAHCGPEELCEF
T.guttata           GLFLLRQCLR-----SLGGYVLSMVCDLQFYHYPIERQLNGTYAILGGKAHCGPEELCEF
L.hirsutus          ------------------------------------------------------------
act.M.domestica     ------------------------------------------------------------
M.eugenii           LFLLRQCLRSLGG------YVLSLVYDLHIHHYPIERQLNGTYAIAGGKPHCGPAELCEF
O.fraenata          ------------------------------------------------------------
M.domestica         LFLLRQCLRSLGG------YVLSLVYNLTFHHYPIERQLNGTYAIAGGKPHCGPAELCEF
R.norvegicus        GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCQF
M.musculus          GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCQF
C.griseus           GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCQF
O.cuniculus         GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
H.sapiens           GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
P.troglodytes       ------------------------------------------------------------
C.jacchus           GLFLLRQC----LRSLG-GYVLSLVHDVSFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
N.leucogenys        GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
P.abelii            GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
L.africana          GLFLLRQC----LRSLGGLRALAGARRGPFHHFPVERQG--------------------
S.scrofa            GLFLLRQC----LRSLG-GYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
B.taurus            GLFLLRQC----LRSLG-GYVLSLVHEVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
A.melanoleuca       GLFLLRQC----LRSLG-GYVLSLVHEVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEF
C.familiaris        GLFLLRQC----LRSLG-GYMLSLVHNVRFHHFX---------------------------
E.caballus          GLFLLRQW----LRSLG-GYVLSLVHNVRFQHFPIERQLNGTYAIAGGKAHCGPAELCEF
C.porcellus         TYGIVGGP----RGIAG-----------------------------------AADLCEF
H.hippoglossus      GLFLLRQCLR-----SLGGYVLSIACNVEFNHYTIEKLLNGTYCIVGGKPHCGPAELCEF
O.niloticus         GLFLLRQCLR-----SLGGYVLSIVWNLDFHHYSVEKQLNGTYCITGGKPHCGPAELCEF


X.laevis            YSKDADGLSCTLRRPCNRPVGVECQAGVFDNMRDNMMREYVRQTWKLEGDALEQAIISQA
X.silurana          YSKDADGLCCTLRRPCNRPAGVECQAGVFDNMRDNMMREYVRQTWKLEGDALEQAIISQA
G.gallus            YSKDADGLCCTLRKPCNRPSGVEPQPGVFDSMRDNMVREYVRQTWRLEGDALEQAIISQA
M.gallopavo         YSKDADGLCCTLRKPCNRPSGVEPQPGVFDSMRDNMVREYVRQTWRLEGDALEQAIISQA
T.guttata           YSKDADGLCCPLRKPCNRPSGVEPQPGVFDSMRDNMVREYVRQTWKLEGDALEQAIISQA
L.hirsutus          ------------------------------------------------------------
act.M.domestica     ------------------------------------------------------------
                          ◄
M.eugenii           YSKDADGLPCALRKPCNRPSGMEPQPGVFDSFRDSMVRDYVRQTWKLEGDALEQAIISQA
O.fraenata          ------------------------------------------------------------
M.domestica         YSKDADGLPCPLRKPCNRPSGLEPQPGVFDSLRDSMVRDYVRQTWKLEGEALEQAIISQA
R.norvegicus        YSQDPDGLPCNLRKPCNRPPGLEPQPGVFDCLRDAMVRDYVRQTWKLEGDALEQAIISQA
M.musculus          YSQDPDGLPCNLRKPCNRPPGLEPQPGVFDCLRDAMVRDYVRQTWKLEGDALEQAIISQA
C.griseus           YSQDPDGLPCNLRKPCNRPPGLEPQPGVFDCLRDAMVRDYVRQTWKLEGDALEQAIISQA
O.cuniculus         YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRDAMVRDYVRQTWKLEGEALEQAIISQA
H.sapiens           YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRDAMVRDYVRQTWKLEGEALEQAIISQA
P.troglodytes       ------------------------------------------------------------
C.jacchus           YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDGLRDAMVRDYVRQTWKLEGEALEQAIISQA
N.leucogenys        YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRDAMVRDYVRQTWKLEGEALEQAIISQA
P.abelii            YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRDAMVRDYVRQTWKLEGEALEQAIISQA
L.africana          -------------------------------------------------EALEQAIISQA
S.scrofa            YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDSLRDGMVRDYVRQTWKLEGEALEQAIISQA
B.taurus            YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDNLRDAMVRDYVRQTWKLEGEALEQAIISQA
A.melanoleuca       YSRDPDGLPCNLRKPCNRPSGLXX--GVFDCLRDAMVRDYVRQTWKLEGEALEQAIISQA
C.familiaris        -------------------------------------------------GEALEQAIISQA
E.caballus          YSRDPDGLPCNLRKPCNRPSGLEPQPGVFDSLRDAMVRDYVRQTWKLEGEALEQAIISQA
C.porcellus         CSPRSDGLPCALRRPCNRPTGLEPQPGVFDSMRDAMVRDYVRQSWKLEGEALEQAILSQA
H.hippoglossus      YSKDADGLVSNLRKPCLRAPDTPIQPGVFDSLRENMLREYVKQTWALEGEAMEQAIISQA
O.niloticus         YSKDSDGLVCLLKKPCLRSPDTPIRQGVFDSLRENMLREYVKQTWNLEGEAMEQAIISQA


X.laevis            PQVEKLIATTAHERMAWYHGSISRDEAERKLYSGAQPDGKFLMRERKENGTYALSVMYGK
X.silurana          PQVEKLIATTAHERMAWYHGTISRDEAERKLYSGAQPDGKFLMRERKENGTYALSVMYGK
G.gallus            PQVEKLIATTAHERMPWYHGNIARDEAERRLYSGAQPDGKFLLRDKKENGAYALSLVYGK
M.gallopavo         PQVEKLIATTAHERMPWYHGNIARDEAERRLYSGAQPDGKFLLRDKKESGTYALSLVYGK
T.guttata           PQVEKLVATTAHERMPWYHGNISREEAERRLYAGAQPDGKFLLREKKENRAYALSLVYGK
L.hirsutus          ------------------------------------------------------------
act.M.domestica     ----------------------------RKLYSGSQHDGKFLLKPRKEQGSYALSLIYGK


                                  ►      SH2 domain-C
M.eugenii           PQVEKLIATTAHERMPWYHSSISREEAKRKLYSGSQHDGKFLLKPRKEQGSYALSLINGK
O.fraenata          ----------------------------YSGSQHDGKFLLKPRKEQGTYALSLIYGK
M.domestica         PQVEKLIATTAHERMPWYHSSISREEAERKLYSGTQHDGKFLLRPRKEHGSYALSLIYGK
R.norvegicus        PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGQQTDGKFLLRPRKEQGTYALSLVYGK
M.musculus          PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGQQTDGKFLLRPRKEQGTYALSLVYGK
C.griseus           PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGQQTDGKFLLRPRKEQGTYALSLIYGK
O.cuniculus         PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGAQTDGKFLLRPRKEQGTYALSLVYGK
H.sapiens           PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGAQTDGKFLLRPRKEQGTYALSLIYGK
P.troglodytes       --------------MPWYHSSLTREEAERKLYSGAQTDGKFLLRPRKEQGTYALSLIYGK
```

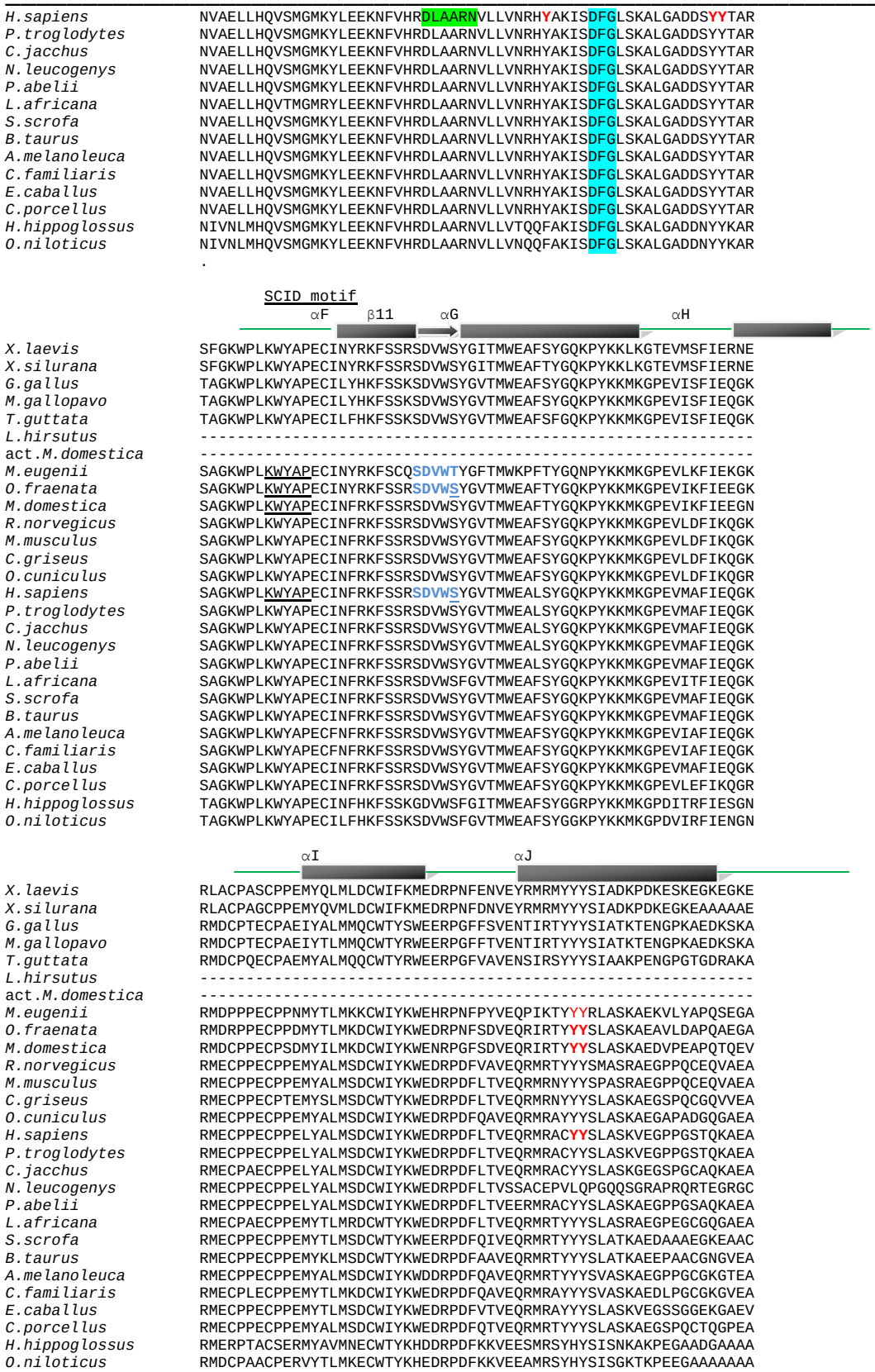
Appendix 5B


C.jacchus PQVEKLIATTAHERMPWYHSNLTREEAERKLYSGTQTDGKFLLRPRKEQGTYALSLIYGK
N.leucogenys PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGTQTDGKFLLRPRKEQGTYALSLIYGK
P.abelii PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGAQTDGKFLLRPRKEQGTYALSLIYGK
L.africana PQVEKLIATTAHERMPWYHTNLTREEAERKLYSGAQTDGKFLLRPRKEQGTYALSLIYGK
S.scrofa PQVEKLIATTAHERMPWYHSNLTREEAERKLYSGSQTDGKFLLRPRKEQGTYALSLIYGK
B.taurus PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGSQTDGKFLLRPRKEPGTYALSLIYGK
A.melanoleuca PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGSQTDGKFLLRPRKEQGTYALSLIYGK
C.familiaris PQVEKLIATTAHERMPWYHSSLTREEAERKLYSGSQTDGKFLLRPRKEQGTYALSLIYGK
E.caballus PQVEKLIATTAHERMPWYHNSLSREEAERKLYAGSQTDGKFLLRPRKEQGTYALSLIYGK
C.porcellus PQVEKLIATAAHERMPWYHNGLTREEAERKLYSGSQTDGKFLLRPRKEPSTYALSVIYGK
H.hippoglossus PQLEKLIATTAHERMPWYHGKITRQEGERRLYSGAQPDGKFLVRDRDKSGTFALSMIYGK
O.niloticus PQLEKLIATTAHEKMHWYHGKISRHEGERRLYSGAQPDGKFLIRDREESGTYALSMMYGK
X.laevis TVYHYKIDQDKSGKYSIPEGTKFDTLWQLVEYLKLKSDGILAVLKESCANASTFSIAPAA
X.silurana TVYHYKIDQDKSGKYSIPEGTKFDTLWQLVEYLKLKSDGIMAVLKESCPNSCTFSIAPAA
G.gallus TVYHYRIDQDKSGKYSIPEGTKFDTLWQLVEYLKLKPDGLIFYLRESCPNPSMAGELVPV
M.gallopavo TVYHYRIDQDKSGKYSIPEGTKFDTLWQLVEYLKLKPDGLIFYLRESCPNPSMP------
T.guttata TVYHYRVDHDKSGKYSIPEGTKFDTLWQLVEYLKLKPDGLIFYLREACPNPNMPASAAPV
L.hirsutus ---------DKSGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEICPNAS-ASTATVT
act.M.domestica TVYHYLINQDKSGKYCIPEGTKFDTLWQLVKYLKLKANGLIYCLKEICPNAS-ASTATVT
M.eugenii TVYHYLINQDKSGKYCIPEGTKFDTLWQLVKYLKLKANGLIYCLKEICPNAS-ASTATVT
O.fraenata TVYHYLINQDKSGKYCIPEGTKFDTLWQLVKYLKLKANGLIYCLKEICPNAS-ASTATVT
M.domestica TVYHYLINQDKSGKYCIPEGTKFDTLWQLVEYLKLKSDGLIYCLKEICPNAS-ASTASVT
R.norvegicus TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYRLKEVCPNS---SASAEA
M.musculus TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYRLKEVCPNS---SASAAV
C.griseus TVYHYLISQDKAGKFYIPEGTKFDTLWQLVEYLKLKADGLIYRLKEACPNS---SASTEA
O.cuniculus TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSS-ASAAAGT
H.sapiens TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSS-ASNASGA
P.troglodytes TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSS-ASNASGA
C.jacchus TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSS-ASNTSGA
N.leucogenys TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSS-ASNASGA
P.abelii TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSS-ASNASGA
L.africana TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSS-ASAASGA
S.scrofa TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPN---TSASSGA
B.taurus TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPN---SSASSGA
A.melanoleuca TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKDACPNSSASSGEGEA
C.familiaris TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKDACPNTNASSGK-WA
E.caballus TVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPN-NNASAGAGA
C.porcellus TVYHYLIKQDKAGKFCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEPCPNTS----AEGA
H.hippoglossus TVYHYQILQEKSGKYCMPEGTKFDTIWQLVEYLKMKPDGLVTVLVESCMNGKAAAKMPNL
O.niloticus TVYHYQILQDKSGKFSMPEGTKFDTIWQLVEYLKMKPDGLVTVLGETCVNGKAAVGKTPS
Negative regulatory site
X.laevis APP------------SLPKVRPVASNSDGYTPEP----------FLGKSRILPMDTSVYE
X.silurana APP------------SLPKIRQAASNSDGYTPEP----------FIGKSRILPMDTSVYE
G.gallus PTLASVTCPRSPAPPTHPSKNLGPLNADGYTPEP------VVVMSAGKSRVLPMDTSVYE
M.gallopavo ----------APAPPTHPSKNLGPLNADGYTPEP------VGKS-----RMLPMDTSVYE
T.guttata PP----THPSAWGLTHHPRRNLGPLNADGYTPDP------VGAG---KSRLLPMDTSVYE
L.hirsutus AAP---TLPVHP----SMPRRNDTLNSDGYTPEP----ACLNKSQGEKSRVLPMDTSVYE
act.M.domestica AAP---TLPVHP----SMPRRNDTLNSDGYTPEP----ACLNKSQGEKSRVLPMDTSVYE
Linker region
M.eugenii AAP---TLPVHP----SMPRRNDTLNSDGYTPEP----ACLNKSQGEKSRVLPMDTSVYE
O.fraenata AAP---TLPVHP----SMPRRNDTLNSDGYTPDP----ACLNKSQGEKSRVLPMDTSVYE
M.domestica AAP---TLPVHP----SMPRRNDTFNSDGYTPEP----ARLVNNQGEKSRVLPMDTSVYE
R.norvegicus AAP---TLPAHPSTFTQPHRRIDTLNSDGYTPEP-----ARLD----KPRPMPMDTSVYE
M.musculus AAP---TLPAHPSTFTQPQRRVDTLNSDGYTPEP-----ARLASSTDKPRPMPMDTSVYE
C.griseus AAP---TLPAHPSTLPQ--RRVDTLNSDGYTPEP-----ARPASSG-KPRPMPMDTSVYE
O.cuniculus AAP---TLPAHPSTFTRPQRRMDTLNSDGYTPEPGYTPEPARLASPEKPRPMPMDTSVYE
H.sapiens AAP---TLPAHPSTLTHPQRRIDTLNSDGYTPEP-----ARITSPD-KPRPMPMDTSVYE
P.troglodytes AAP---TLPAHPSTLTHPQRRIDTLNSDGYTPEP-----ARVTSPD-KPRPMPMDTSVYE
C.jacchus VAP---TLPAHPSTFTHPQRRMDTLNSDGYTPEP-----ARVTSPD-KPRPMPMDTSVYE
N.leucogenys AAP---TLPAHPSTLTHPQRRIDTLNSDGYTPEP-----ARVTSSD-KPRPMPMDTSVYE
P.abelii AAP---TLPAHPSTLTHPQRRIDTLNSDGYSPEP-----ARVTSPD-KPRPMPMDTSVYE
L.africana AAP---TLPAHPSTLTQPQRRMDTLNSDGYTPEP-----GELAEKTEKKRPLPMDTSVYE
S.scrofa AAP---TLPAHPSTFTHPQRRIDTLNSDGYTPEP------ARLGSSEKARTMPMDTSVYE
B.taurus AAP---TLPAHPSTFTQPQRRIDTLNSDGYTPEP------ARLVSSEKPRTMPMDTSVYE
A.melanoleuca AAP---TLPAHPSTFTHAPRRIDTN-SDGYTPEP------ARLVSPEKAPSMPMDTSVYE
C.familiaris AAP---TLPAHPSTFTQAHRRIDTLNSDGYTPEP------ARLASSEKAQSMPMDTSVYE
E.caballus AAP---TLPAHPSTFTHPQKRIDTLNSDGYTPEP------ARLASPEKVRPMPMDTSVYE
C.porcellus AAP---TLPAHPSTFSQPQRRVDTN-SDGYTPEP------ARPAPAGEPRPMPMDTSVYE
H.hippoglossus PAS-------------------RRVNVNGYTPPPR--VVTEASEPAAERDVLPMDCTGFN
O.niloticus LPS-------------------RPRGANGYTPPPQ--APTPAPK-TEDRDLLPMDCNGFN
P-loop
ATP-binding
Phosphotransfer
β1
β2
β3
X.laevis SPYSDPEELKERKLFVKRELLLIDEVELGSGNFGCVKKGVYKLKKRQIDVAIKVLKVQEE

Appendix 5B

```
X.silurana          SPYSDPEELKERKLFVKRELLLIDEVELGSGNFGCVKKGVYKLKKRQIDVAIKVLKVQDE
G.gallus            SPYSDPEELKDKKLFLKRDHLMIDEVELGAGNFGCVKKGVYKMRKKQIDVAIKVLKSNNE
M.gallopavo         SPYSDPEELKDKKLFLKRDHLMIDEVELGAGNFGCVKKGVYKMRKKQIDVAIKVLKSNNE
T.guttata           SPYSDPEELKDKKLFLKRDHLMIDEVELGAGNFGCVKKGVYKMRKKQIDVAIKVLKSNNE
L.hirsutus          SPYSDPEELKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYRMRKKQIDVAIKVLKSTNE
act.M.domestica     SPYSDPEKLKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYKMRKKQIDVAIKVLKSTNE
M.eugenii           SPYSDPEELKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYKMRKKQIDVAIKVLKSTNE
O.fraenata          SPYSDPKKLKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYKMRKKQIDVPIKVLKSTNE
M.domestica         SPYSDPEELKDKKLFLKRENLIIDEVELGSGNFGCVRKGVYKMRKKQIDVAIKVLKSTNE
R.norvegicus        SPYSDPEELKDKKLFLKRENLLVADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQSTE
M.musculus          SPYSDPEELKDKKLFLKRENLLVADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQGTE
C.griseus           SPYSDPEELKDKKLFLKRENLIVADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKHSTE
O.cuniculus         SPYSDPEELKDKKLFLKRENLLVADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKHSTE
H.sapiens           SPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQGTE
P.troglodytes       SPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQGTE
C.jacchus           SPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQGTE
N.leucogenys        SPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQGTE
P.abelii            SPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKKEEL---
L.africana          SPYSDPEELKDKKLFLKRENLLVADLELGCGNFGSVRQGVYRMRKYGGPARVG-------
S.scrofa            SPYSDPEELKNKKLFLKRENLLMADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQSTE
B.taurus            SPYSDPEELKNKKLFLKRENLLMADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQSTE
A.melanoleuca       SPYSDPEELKDKKLFLKRENLLMADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKHSTE
C.familiaris        SPYSDPEELKDKKLFLKRENLLMADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKHSTE
E.caballus          SPYSDPEELKDKKLFLKRENLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKHSTE
C.porcellus         SPYSDPEELKDKKLFLKRENLLVADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQSTE
H.hippoglossus      PYHNPNE---VKRFNIKRTQLLIDEVELGCGNFGSVKKGVLKTDAGQIDVAIKVLKNENE
O.niloticus         PYHNPND---IKRFNIQRSDLLMDEVELGSGNFGCVKKGVLKTESGQIDVAVKVLKNENE
                    . ..  .    ... ..*  *.. ..***.****.*..** .       . .
                                                Gatekeeper pocket
                                                ▼       ▼
                          αC          β4          β5      αD          αE
X.laevis            KNVRDEMMKEAEFMHQLDNPYIVRMIGVCEAENLMLVMEMASGGPLNKFLGAKKDTITVS
X.silurana          KNVRDEMMKEAEIMHQLDNPYIVRMIGVCEAENLMLVMEMASGGPLNKFLGAKKDVIPMS
G.gallus            KTVRDEMMTEAQIMHQLDNPYIVRMIGVCEAESLMLVMEMASGGPLNKFLSSKKDEITVS
M.gallopavo         KTVRDEMMTEAQIMHQLDNPYIVRMIGVCEAESLMLVMEMASGGPLNKFLSSKKDEITVS
T.guttata           KTVKDEMMKEAQIMHQLDNPYIVRMIGVCEAESLMLVMEMASGGPLNRFLASKKDEITTS
L.hirsutus          KAEKDEMMKEAQIMHQLDNPYIVRIIGVCKAEALMLVMEMAIAGPLHKFLAAKKGGGPCK
act.M.domestica     KAEKDEMMKEAQIMHLLDNPYIVRIIGVCKAEALMLVMEMAIAGPLHKFLAAKKEEVPVS
M.eugenii           KAEKDEMMKEAQIMHLLDNPYIVRIIGVCKAEALMLVMEMAIAGPLHKFLAAKKEEVPVS
O.fraenata          KAEKDKMMKEAQIMHQLDNPYIVRIIGVCKAEALMLVMKMAIAGPLHKFLAAKKEEVPVS
M.domestica         KAEKEEMMKEAHIMHQLDNPYIVRIIGVCKAEALMLVMEMATGGPLNKFLSSKRGEIPVS
R.norvegicus        KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLIGKKEEIPVS
M.musculus          KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLLGKKEEIPVS
C.griseus           KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLLGKKEEIPVS
O.cuniculus         KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLLGKKEEIPVS
H.sapiens           KADTEEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLVGKREEIPVS
P.troglodytes       KADTEEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLVGKREEIPVS
C.jacchus           KADTEEMMREAQIMHQLDNPYIVRLIGVCRAEALMLVMEMAGGGPLHKFLIGKKEEIPVS
N.leucogenys        KADTEEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLVGKKEEIPVS
P.abelii            ----------------------------LERTVEEPYVCAHGDGDNQATLSLRCLEEIPVS
L.africana          -------------------------------------------------ALPGEEIPVS
S.scrofa            KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGAGPLHKFLVGKKEEIPVS
B.taurus            KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLVGKKEEIPVS
A.melanoleuca       KTDKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLLSKKEEIPIS
C.familiaris        KTDKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLLGKKEEIPIS
E.caballus          KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLIMEMAGGGPLHKFLLGKKEEIPVS
C.porcellus         KADKDEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLVGRKEEIPVS
H.hippoglossus      KLVKEEMMREAEIMHQLNNPFIVRMLGLCNAESLMLVMEMAAAGPLNKFLSSNKDTVTAE
O.niloticus         KLVKEEMMREAEIMHQLSNPFIVRMLGLCNADCLMLVMEMASAGPLNKFLSSKKDTVSVE
                                                                  . .
                                                                -reg./+reg
                                                      DFG-loop    ▼▼
                                                      ▼ ▼
                                        Catalytic   activation      loop
                          αE                  β7      β8      β9      β10
X.laevis            NVVELMHQVSMGMKYLEGKNFVHRDLAARNVLMVNQHYAKISDFGLSKALAADDSYYKAK
X.silurana          NIVELMHQVSIGMKYLEGKNFVHRDLAARNVLMVNQHYAKISDFGLSKALAADDSYYKAK
G.gallus            NIVELMHQVSMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYKAR
M.gallopavo         NIVELMHQVSMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYKAR
T.guttata           NIVELMHQVSMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYKAR
L.hirsutus          QCRGAAAPSGHGNEIPRRKKF-------------------------------------------
act.M.domestica     NVVSCCTKWPWE-----------------------------------------------------
M.eugenii           NVVELLHQVAMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYTAR
O.fraenata          NVVELLHQVAMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYTAR
M.domestica         NVVELLHQVAMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYTAR
R.norvegicus        NVAELLHQVAMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
M.musculus          NVAELLHQVAMGMKYLEEKNFVHRDLAACNVLLVNRHYAKISDFGLSKALGADDSYYTAR
C.griseus           NVAELLHQVAMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
O.cuniculus         NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
```

Appendix 5B




```
H.sapiens             NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
P.troglodytes         NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
C.jacchus             NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
N.leucogenys          NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
P.abelii              NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
L.africana            NVAELLHQVTMGMRYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
S.scrofa              NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
B.taurus              NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
A.melanoleuca         NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
C.familiaris          NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
E.caballus            NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
C.porcellus           NVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTAR
H.hippoglossus        NIVNLMHQVSMGMKYLEEKNFVHRDLAARNVLLVTQQFAKISDFGLSKALGADDNYYKAR
O.niloticus           NIVNLMHQVSMGMKYLEEKNFVHRDLAARNVLLVNQQFAKISDFGLSKALGADDNYYKAR
                      .

X.laevis              SFGKWPLKWYAPECINYRKFSSRSDVWSYGITMWEAFSYGQKPYKKLKGTEVMSFIERNE
X.silurana            SFGKWPLKWYAPECINYRKFSSRSDVWSYGITMWEAFTYGQKPYKKLKGTEVMSFIERNE
G.gallus              TAGKWPLKWYAPECILYHKFSSKSDVWSYGVTMWEAFSYGQKPYKKMKGPEVISFIEQGK
M.gallopavo           TAGKWPLKWYAPECILYHKFSSKSDVWSYGVTMWEAFSYGQKPYKKMKGPEVISFIEQGK
T.guttata             TAGKWPLKWYAPECILFHKFSSKSDVWSYGVTMWEAFSFGQKPYKKMKGPEVISFIEQGK
L.hirsutus            ------------------------------------------------------------
act.M.domestica       ------------------------------------------------------------
M.eugenii             SAGKWPLKWYAPECINYRKFSCQSDVWTYGFTMWKPFTYGQNPYKKMKGPEVLKFIEKGK
O.fraenata            SAGKWPLKWYAPECINYRKFSSRSDVWSYGVTMWEAFTYGQKPYKKMKGPEVIKFIEEGK
M.domestica           SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFTYGQKPYKKMKGPEVIKFIEEGN
R.norvegicus          SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVLDFIKQGK
M.musculus            SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVLDFIKQGK
C.griseus             SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVLDFIKQGK
O.cuniculus           SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVLDFIKQGR
H.sapiens             SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQKPYKKMKGPEVMAFIEQGK
P.troglodytes         SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQKPYKKMKGPEVMAFIEQGK
C.jacchus             SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQKPYKKMKGPEVMAFIEQGK
N.leucogenys          SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQKPYKKMKGPEVMAFIEQGK
P.abelii              SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQKPYKKMKGPEVMAFIEQGK
L.africana            SAGKWPLKWYAPECINFRKFSSRSDVWSFGVTMWEAFSYGQKPYKKMKGPEVITFIEQGK
S.scrofa              SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVMAFIEQGK
B.taurus              SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVMAFIEQGK
A.melanoleuca         SAGKWPLKWYAPECFNFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVIAFIEQGK
C.familiaris          SAGKWPLKWYAPECFNFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVIAFIEQGK
E.caballus            SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVMAFIEQGK
C.porcellus           SAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEAFSYGQKPYKKMKGPEVLEFIKQGR
H.hippoglossus        TAGKWPLKWYAPECINFHKFSSKGDVWSFGITMWEAFSYGGRPYKKMKGPDITRFIESGN
O.niloticus           TAGKWPLKWYAPECILFHKFSSKSDVWSFGVTMWEAFSYGGKPYKKMKGPDVIRFIENGN

X.laevis              RLACPASCPPEMYQLMLDCWIFKMEDRPNFENVEYRMRMYYYSIADKPDKESKEGKEGKE
X.silurana            RLACPAGCPPEMYQVMLDCWIFKMEDRPNFDNVEYRMRMYYYSIADKPDKEGKEAAAAAE
G.gallus              RMDCPTECPAEIYALMMQCWTYSWEERPGFFSVENTIRTYYYSIATKTENGPKAEDKSKA
M.gallopavo           RMDCPTECPAEIYTLMMQCWTYRWEERPGFFTVENTIRTYYYSIATKTENGPKAEDKSKA
T.guttata             RMDCPQECPAEMYALMQQCWTYRWEERPGFVAVENSIRSYYYSIAAKPENGPGTGDRAKA
L.hirsutus            ------------------------------------------------------------
act.M.domestica       ------------------------------------------------------------
M.eugenii             RMDPPPECPPNMYTLMKKCWIYKWEHRPNFPYVEQPIKTYYYRLASKAEKVLYAPQSEGA
O.fraenata            RMDRPPECPPDMYTLMKDCWIYKWEDRPNFSDVEQRIRTYYYSLASKAEAVLDAPQAEGA
M.domestica           RMDCPPECPSDMYILMKDCWIYKWENRPGFSDVEQRIRTYYYSLASKAEDVPEAPQTQEV
R.norvegicus          RMECPPECPPEMYALMSDCWIYKWEDRPDFVAVEQRMRTYYYSMASRAEGPPQCEQVAEA
M.musculus            RMECPPECPPEMYALMSDCWIYKWEDRPDFLTVEQRMRNYYYSPASRAEGPPQCEQVAEA
C.griseus             RMECPPECPTEMYSLMSDCWTYKWEDRPDFLTVEQRMRNYYYSLASKAEGSPQCGQVVEA
O.cuniculus           RMECPPECPPEMYALMSDCWIYKWEDRPDFQAVEQRMRAYYYSLASKAEGAPADGQGAEA
H.sapiens             RMECPPECPPELYALMSDCWIYKWEDRPDFLTVEQRMRACYYSLASKVEGPPGSTQKAEA
P.troglodytes         RMECPPECPPELYALMSDCWIYKWEDRPDFLTVEQRMRACYYSLASKVEGPPGSTQKAEA
C.jacchus             RMECPAECPPELYALMSDCWIYKWEDRPDFLTVEQRMRACYYSLASKGEGSPGCAQKAEA
N.leucogenys          RMECPPECPPELYALMSDCWIYKWEDRPDFLTVSSACEPVLQPGQQSGRAPRQRTEGRGC
P.abelii              RMECPPECPPELYALMSDCWIYKWEDRPDFLTVEERMRACYYSLASKAEGPPGSAQKAEA
L.africana            RMECPAECPPEMYTLMRDCWTYKWEDRPDFLTVEQRMRTYYYSLASRAEGPEGCGQGAEA
S.scrofa              RMECPPECPPEMYTLMSDCWTYKWEERPDFQIVEQRMRTYYYSLATKAEDAAAEGKEAAC
B.taurus              RMECPPECPPEMYKLMSDCWTYKWEDRPDFAAVEQRMRTYYYSLATKAEEPAACGNGVEA
A.melanoleuca         RMECPPECPPEMYALMSDCWIYKWDDRPDFQAVEQRMRTYYYSVASKAEGPPGCGKGTEA
C.familiaris          RMECPLECPPEMYTLMKDCWIYKWEDRPDFQAVEQRMRAYYYSVASKAEDLPGCGKGVEA
E.caballus            RMECPPECPPEMYTLMSDCWIYKWEDRPDFVTVEQRMRAYYYSLASKVEGSSGGEKGAEV
C.porcellus           RMECPPECPPEMYALMSDCWIYKWEDRPDFQTVEQRMRTYYYSLASKAEGSPQCTQGPEA
H.hippoglossus        RMERPTACSERMYAVMNECWTYKHDDRPDFKKVEESMRSYHYSISNKAKPEGAADGAAAA
O.niloticus           RMDCPAACPERVYTLMKECWTYKHEDRPDFKKVEEAMRSYHYSISGKTKPEEGAAAAAAA
```

Appendix 5B

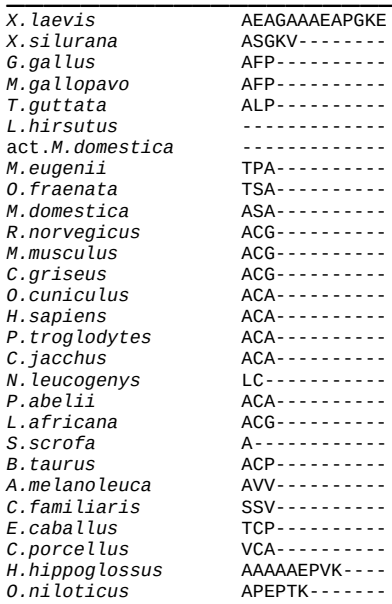


Figure 5B.1. Amino acid alignment for known ZAP-70 sequences.
Domains are marked with solid arrows ◄►. Important tyrosines in red, yellow marked M in Gatekeeper pocket is the gatekeeper. Black bold underlined is SCID motif. –reg/+reg, are the negative and positive regulator in the PTK domain. **K** is responsible for phosphotransfer. Green lines repesent coils and turns, while boxes indicate α-helices and arrows indicate β-strands.

Sequence alignment of ensembl annotated *M. eugenii* ZAP-70 (eZAP) sequence and the expressed sequence from the cloned *M. eugenii* ZAP-70 gene (ZAP).

```
eZAP            MPDPAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVHDVRFHHFP 60
ZAP             MPDAAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVYDLHIHHYP 60
                ***.**********************************************.*...**.*

eZAP            IERQLNGTYAIAGGKAHCGPAELCEFYSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRD 120
ZAP             IERQLNGTYAIAGGKPHCGPAELCEFYSKDADGLPCALRKPCNRPSGMEPQPGVFDSFRD 120
                ***************.************.*.***** **********.********..**

eZAP            AMVRDYVRQTWKLEGEALEQAIISQAPQVEKLIATTAHERMPWYHSSLTREEAERKLYSG 180
ZAP             SMVRDYVRQTWKLEGDALEQAIISQAPQVEKLIATTAHERMPWYHSSISREEAKRKLYSG 180
                .**************.*******************************..****.******

eZAP            AQTDGKFLLRPRKEQGTYALSLIYGKTVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLK 240
ZAP             SQHDGKFLLKPRKEQGSYALSLINGKTVYHYLINQDKSGKYCIPEGTKFDTLWQLVKYLK 240
                .* ******.******.****** *********.***.******************.***

eZAP            LKADGLIYCLKEACPNSSASNASGAAAPTLPAHPSTLTHPQRRIDTLNSDGYTPEPARIT 300
ZAP             LKANGLIYCLKEICPNASASTATVTAAPTLPVHPSMP----RRNDTLNSDGYTPEPACLN 296
                ***.******** ***.***.*. .******.***      ** ************ ..

eZAP            SP--DKPRPMPMDTSVYESPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYR 358
ZAP             KSQGEKSRVLPMDTSVYESPYSDPEELKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYK 356
                ..  .*.* .*************************.**.* ..***.****.**.***.

eZAP            MRKKQIDVAIKVLKQGTEKADTEEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAG 418
ZAP             MRKKQIDVAIKVLKSTNEKAEKDEMMKEAQIMHLLDNPYIVRIIGVCKAEALMLVMEMAI 416
                **************. .***...***.****** ********.****.***********

eZAP            GGPLHKFLVGKREEIPVSNVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKIS 478
ZAP             AGPLHKFLAAKKEEVPVSNVVELLHQVAMGMKYLEEKNFVHRDLAARNVLLVNQHYAKIS 476
                .*******..*.**.*****.******.************************.******

eZAP            DFGLSKALGADDSYYTARSAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQK 538
ZAP             DFGLSKALGADDSYYTARSAGKWPLKWYAPECINYRKFSCQSDVWTYGFTMWKPFTYGQN 536
                *********************************.****..****.**.***....***.

eZAP            PYKKMKGPEVMAFIEQGKRMECPPECPPELYALMSDCWIYKWEDRPDFLTVEQRMRACYY 598
ZAP             PYKKMKGPEVLKFIEKGKRMDPPPECPPNMYTLMKKCWIYKWEHRPNFPYVEQPIKTYYY 596
                **********. ***.****. ******..*.**..*******.**.*  *** ... **

eZAP            SLASKVEGPPGSTQKAEAACA- 619
ZAP             RLASKAEKVLYAPQSEGATPA- 617
                 ****.*    ..*.  *. *
```

Figure 5B.2. Alignment between the ensembl ZAP-70 sequence for *M. eugenii* and the cloned sequence of *M. eugenii* ZAP-70 obtained from mRNA.

Appendix 5B

Alignment of *M. eugenii* and *H. sapiens* ZAP-70 amino acid sequences for determination of possible antibody cross reactivity. The coloured sequences indicate antibodies available for *H. sapiens*.

```
M.eugenii          MPDAAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVYDLHIHHYP 60
H.sapiens          MPDPAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVHDVRFHHFP 60
                   ***.**************************************************.*...**.*

M.eugenii          IERQLNGTYAIAGGKPHCGPAELCEFYSKDADGLPCALRKPCNRPSGMEPQPGVFDSFRD 120
H.sapiens          IERQLNGTYAIAGGKAHCGPAELCEFYSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRD 120
                   ***************.************.*.***** **********.********..**

M.eugenii          SMVRDYVRQTWKLEGDALEQAIISQAPQVEKLIATTAHERMPWYHSSISREEAKRKLYSG 180
H.sapiens          AMVRDYVRQTWKLEGEALEQAIISQAPQVEKLIATTAHERMPWYHSSLTREEAERKLYSG 180
                   .**************.*******************************..****.******

M.eugenii          SQHDGKFLLKPRKEQGSYALSLINGKTVYHYLINQDKSGKYCIPEGTKFDTLWQLVKYLK 240
H.sapiens          AQTDGKFLLRPRKEQGTYALSLIYGKTVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLK 240
                   .* ******.******.****** *********.***.*****************.***

M.eugenii          LKANGLIYCLKEICPNASASTATVTAAPTLPVHPSMP----RRNDTLNSDGYTPEPACLN 296
                                      This Antibody was used
H.sapiens          LKADGLIYCLKEACPNSSASNASGAAAPTLPAHPSTLTHPQRRIDTLNSDGYTPEPARIT 300
                   ***.******** ***.***.*. .******.***       ** ************ ..

M.eugenii          KSQGEKSRVLPMDTSVYESPYSDPEELKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYK 356
H.sapiens          SP--DKPRPMPMDTSVYESPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYR 358
                   ..  .*.* .**************************.**.* ..***.****.**.***.

M.eugenii          MRKKQIDVAIKVLKSTNEKAEKDEMMKEAQIMHLLDNPYIVRIIGVCKAEALMLVMEMAI 416
H.sapiens          MRKKQIDVAIKVLKQGTEKADTEEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAG 418
                   **************. .***...***.****** ********.****.***********

M.eugenii          AGPLHKFLAAKKEEVPVSNVVELLHQVAMGMKYLEEKNFVHRDLAARNVLLVNQHYAKIS 476
H.sapiens          GGPLHKFLVGKREEIPVSNVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKIS 478
                   .*******..*.**.*****.******.************************.******

M.eugenii          DFGLSKALGADDSYYTARSAGKWPLKWYAPECINYRKFSCQSDVWTYGFTMWKPFTYGQN 536
H.sapiens          DFGLSKALGADDSYYTARSAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQK 538
                   **********************************.****..****.**.***....***.

M.eugenii          PYKKMKGPEVLKFIEKGKRMDPPPECPPNMYTLMKKCWIYKWEHRPNFPYVEQPIKTYYY 596
H.sapiens          PYKKMKGPEVMAFIEQGKRMECPPECPPELYALMSDCWIYKWEDRPDFLTVEQRMRACYY 598
                   **********. ***.****. ******..*.**..*******.**.*  *** ... **

M.eugenii          RLASKAEKVLYAPQSEGATPA 617
H.sapiens          SLASKVEGPPGSTQKAEAACA 619
                   ****.*    ..*.  *. *
```

Figure 5B.3. Alignment of M. eugenii ZAP-70 and H. sapiens ZAP-70 showing the commercially available ZAP-70 Antibodies.

Appendix 5B

Python shell for Tammar ZAP-70 model.

```
mod9v1 build_profile.py
>P1;TZAP
Sequence. TZAP.....0.00. 0.00
MPDAAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVYDLHIHHYPIERQLNGTYAIAG
GKPHCGPAELCEFYSKDADGLPCALRKPCNRPSGMEPQPGVFDSFRDSMVRDYVRQTWKLEGDALEQAIISQA
PQVEKLIATTAHERMPWYHSSISREEAKRKLYSGSQHDGKFLLKPRKEQGSYALSLINGKTVYHYLINQDKSG
KYCIPEGTKFDTLWQLVKYLKLKANGLIYCLKEICPNASASTATVTAAPTLPVHPSMPRRNDTLNSDGYTPEP
ACLNKSQGEKSRVLPMDTSVYESPYSDPEELKDKKLFLKRENLMIDEVELGSGNFGCVRKGVYKMRKKQIDVA
IKVLKSTNEKAEKDEMMKEAQIMHLLDNPYIVRIIGVCKAEALMLVMEMAIAGPLHKFLAAKKEEVPVSNVVE
LLHQVAMGMKYLEEKNFVHRDLAARNVLLVNQHYAKISDFGLSKALGADDSYYTARSAGKWPLKWYAPECINY
RKFSCQSDVWTYGFTMWKPFTYGQNPYKKMKGPEVLKFIEKGKRMDPPPECPPNMYTLMKKCWIYKWEHRPNF
PYVEQPIKTYYYRLASKAEKVLYAPQSEGATPA*
from modeller import*
log.verbose ()
env = environ ()
#--Prepare the input files
#-- Read in the sequence database
sdb= sequence_db(env)
sdb.read (seq_database_file= 'pdb_95.pir',
seq_database_format='PIR'
            chains_list='ALL', minmax_db_seq_len= (30,4000),
clean_sequences=True)

#--Write the sequence database in binary form
sdb.write(seq_database_file='pdb_95.bin',,
seq_database_format='BINARY',
             chains_list='ALL')

#--Now, read in the binary dataset
sdb.read(seq_datase_file='pdb_95.bin',
seq_database_forma='BINARY",
           chains_list='ALL'

#--Read in the target sequence/alignment
aln= alignment (env)
aln.append (file='c2ozo.ali', alignment_format='PIR',
align_codes= 'ALL')

#--Convert the input sequence/alignment into
#   profile format
pr f=  aln.to_profile ()

#-- Scan sequence database to pick up homologous sequences
prf.build (sdb, matrix_offset=-450 ,
rr_files='${LIB}/blosum62.sim.mat ' ,
           gap_penalties=False , max_aln_evalue=0.01)

#-- Write out the profile in text format
prf.write (file= 'build_profile.prf' , profile_format= 'TEXT'  )

#-- Convert the profile back to alignment format
```

```
aln = prf.to_alignment ()

#-- Write out the aligment file
aln.write (file= 'build_profile.ali' , alignment_format= 'PIR' )

from modeller import *

env = environ ()
aln  = alignment (env)
mdl + model (env, file= 'c2ozo' ,
model_segment= ('FIRST. A ', 'LAST. A' ) )
aln.append_model (mdl, align_codes=' c2ozo ' ,
atom_filex=' c2ozo.pdb ' )
aln.append (file= 'TZAP.ali' , align_codes= 'TZAP' )
aln.align2d ()
aln.write (file=' TZAP-c2ozo.ali' , alignment_format=' PIR ' )
aln.write (file=' TZAP-c2ozo.pap' , alignment_format='PAP' )

from modeller import *
from modeller.automodel import *

env = environ ()
a = automodel (env, alnfile= 'TZAP-c2ozo.ali',
                    knowns=' TZAP' , sequence='TZAP' ,
                    assess_methods=(assess.DOPE, assess.GA341) )
a.starting_model = 1
a.ending_model = 5
a.make ()
```

Appendix 5B

Table 5B.1. Genbank Accession numbers for ZAP-70 and the relevant references.

Species Name	Common Name	Accession Numbers	References
Ailuropoda melnaoleuca	Giant Panda	XM_002912427.1	Annotated
Bos taurus	Cattle	XM_865562	(Zimin *et al.,* 2009)
Callithrix jacchus	Common marmoset	XM_002757393.1	Annotated
Canis lupus familiaris	Dog	XM_003431497.1	Annotated
Cricetulus griseus	Chinese hamster	XM_003498289.1	Annotated
Danio rerio	Zebrafish	NM_001020589	(Lin *et al.,* 2009)
Equus caballus	Horse	XM_001488128	Annotated
Gallus gallus	Chicken	XM_418206.2	Annotated
Hippoglossus hippoglossus	Atlantic halibut	GU985452.1	(Overgard *et al.,* 2010b)
Homo sapiens	Human	NM_001079	(Chan *et al.,* 1992)
Loxodonta africana	African elephant	XM_003422028.1	Annotated
Macaca mulatta	Rhesus monkey	XM_001101293	Annotated
Meleagris gallopavo	Turkey	XM_003213184.1	Annotated
Monodelphis domestica	Gray short tailed opossum	XM_001376170	Annotated
Mus musculus	Mouse	NM_009539 U77667.1 U4379	(Ku *et al.,* 1994) Direct submission Direct submission
Nomascus leucogenys	White-cheeked gibbon	XM_003279222.1	Annotated
Oreochromis niloticus	Nile tilapia	XM_003445128.1	Annotated
Ornithorhynchus anatinus	Platypus	XM_001515221	Annotated
Oryctolagus cuniculus	Rabbit	XM_002710002.1	Annotated
Pan troglodytes	Chimpanzee	XM_515637.3	Annotated
Pongo abelii	Sumatran Orangutan	XM_002811671.1	Annotated
Rattus norvegicus	Norway Rat	NM_001012002	(Arpaia *et al.,* 1994)
Sus scrofa	Pig	XM_003481152.1	Annotated
Taeniopygia guttata	Zebra finch	XM_002191899	Annotated
Xenopus laevis	African clawed frog	NM_001093536	(Klein *et al.,* 2002)
Xenopus tropicalis	Silurana tropiclais	NM_001006824	(Klein *et al.,* 2002)

Appendix 5C

Onychogalea fraenata Lymphocyte specific kinase (Lck) nucleotide sequence

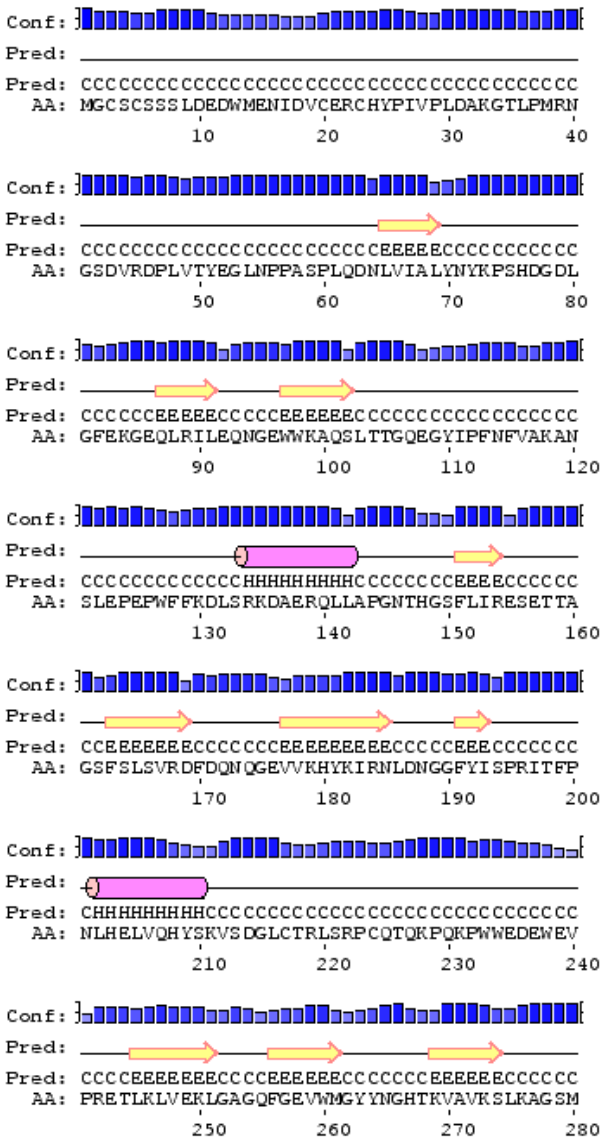
atgggctgctcctgcagctccagcctcgacgaggactggatggaaaatattgacgtgtgtgaacgatgccatt
accctattgtaccactggatgcaaagggcacgctcccaatgaggaatggctctgacgtgagggatcccttggt
cacctatgagggtttaaatccacctgcatctccattacaagataacctggtcatcgccctgtataattataaa
ccctcccatgatggggacctgggctttgagaaaggggagcaactgaggatcctggagcagaatggagaatggt
ggaaggcacagtccctgaccactggccaggagggctacattcccttcaactttgtggccaaagccaacagcct
ggagcctgagccttggtttttcaaggacttgagccggaaggatgctgagagacaacttttggcccctgggaac
actcatggatccttcctgatcagagagagtgagaccactgcaggctccttctctctgtctgtgcgggactttg
accagaaccaggggggaggtggtgaaacattacaagatccgcaacctggataatgggggcttctacatttcccc
ccgaatcacctttcctaatctgcatgaactggttcagcattactccaaagtctcagatgggctatgtactcga
ctgagtcggccctgccagacccaaaagccacagaagccctggtgggaagatgagtgggaggttcctcgagaga
cactgaagctggtggaaaagctgggagctggccagtttggggaggtctggatggggtattacaatgggcatac
caaggtagcggtgaaaagcctgaaagcgggcagcatgtctcctgatgccttcctggctgaagccaacctgatg
aaacagctgcagcaccagcgactggtacgcctttatgcggtggtcacacaggaacccatctacatcatcactg
aatacatggagaatgggagcctggtagacttcctcaaaactacaacaggagtcaaactaaccatccacaaact
gcctgatatggctgcacagattgctgagggcatggccttcattgaagagcggaattacatccaccgggacctc
agagcagccaacatcctggtgtcagacacattgaactgtaagattgctgactttggattggcccggctgatcg
aggacaacgagtacacagctagagagggagcaaaatttcccatcaagtggacagctcttgaggccatcaacta
tgggacatttaccatcaagtcagatgtctggtcttttggtatcctgctcacagaaatcgtcacctatgggagg
atcccctacccaggaatgaccaaccctgaggtgattcagaacctggagcaaggctatcggatggtgaggcctg
acaattgcccagaagaactgtacaaaccgatgatgctgtgttggaaggagaggcctgaggatcggcccacctt
tgattacctgaggagtgtcttagaggacttcttcattgccacagagggccagtaccagccccaagtga

O. fraenata Lck putative protein sequence

MGCSCSSSLDEDWMENIDVCERCHYPIVPLDAKGTLPMRNGSDVRDPLVTYEGLNPPASPLQDNLVIALYNYK
PSHDGDLGFEKGEQLRILEQNGEWWKAQSLTTGQEGYIPFNFVAKANSLEPEPWFFKDLSRKDAERQLLAPGN
THGSFLIRESETTAGSFSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPNLHELVQHYSKVSDGLCTR
LSRPCQTQKPQKPWWEDEWEVPRETLKLVEKLGAGQFGEVWMGYYNGHTKVAVKSLKAGSMSPDAFLAEANLM
KQLQHQRLVRLYAVVTQEPIYIITEYMENGSLVDFLKTTTGVKLTIHKLPDMAAQIAEGMAFIEERNYIHRDL
RAANILVSDTLNCKIADFGLARLIEDNEYTAREGAKFPIKWTALEAINYGTFTIKSDVWSFGILLTEIVTYGR
IPYPGMTNPEVIQNLEQGYRMVRPDNCPEELYKPMMLCWKERPEDRPTFDYLRSVLEDFFIATEGQYQPQG

Appendix 5C

Secondary structure prediction for *O. fraenata* Lck using PSIpred



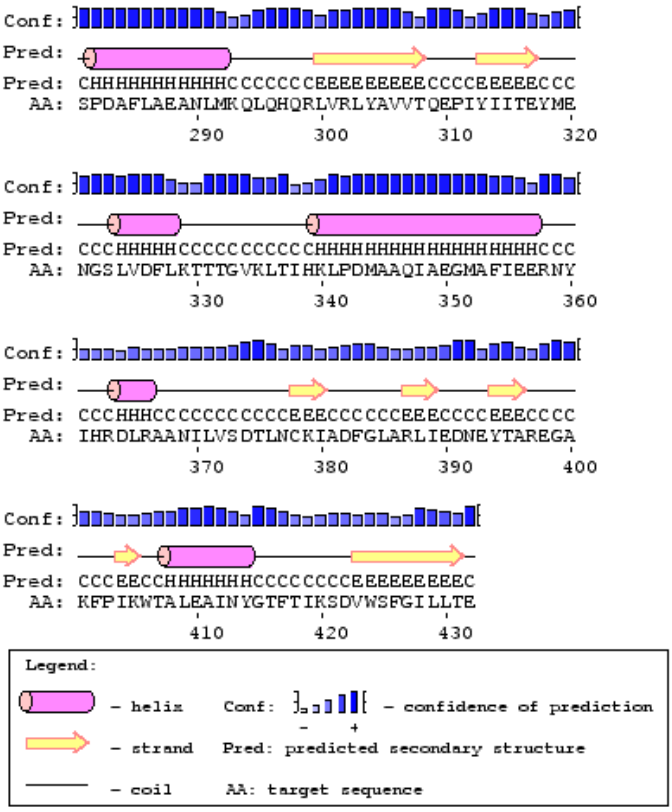


Figure 5C.1. Secondary structure prediction *O. fraenata* Lck.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 5C

Macropus eugenii lymphocyte specific kinase (Lck) nucleotide sequence

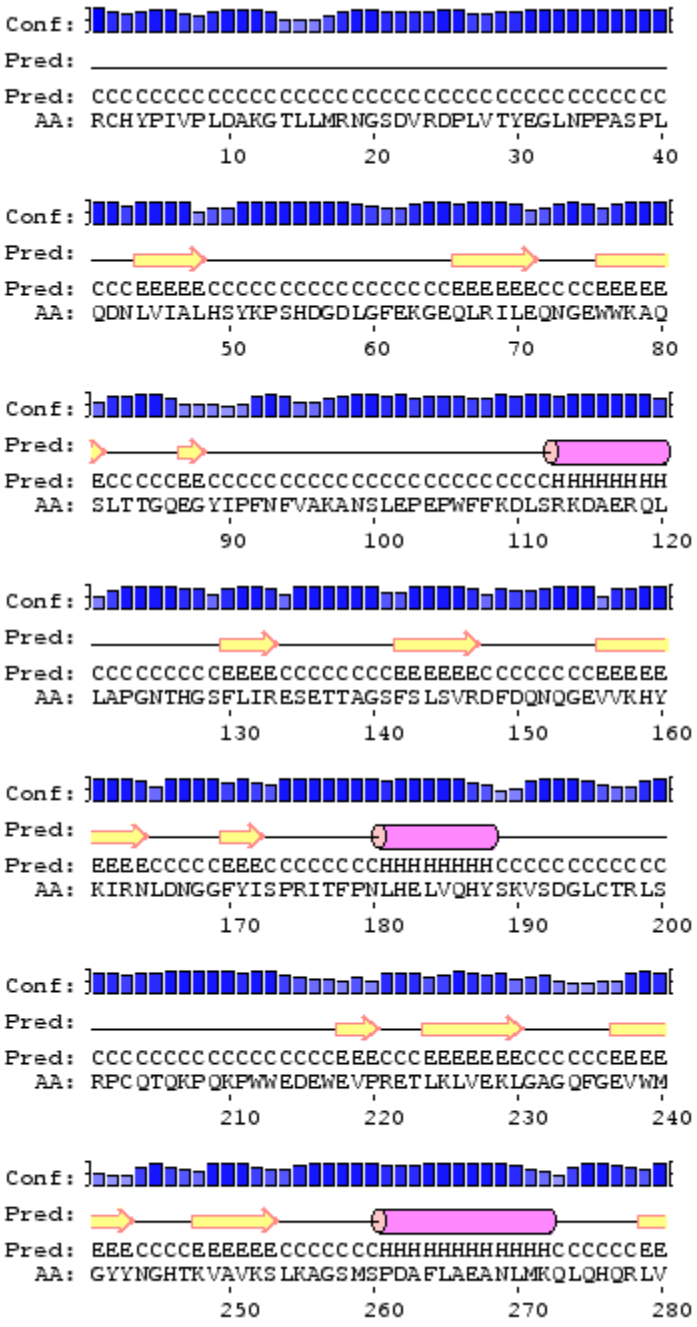
cgatgccattaccctattgtaccactggatgccaagggcacgctcctaatgaggaatggctctgacgtgaggg
atcccttggtcacctatgagggtttaaacccacctgcatctccattacaagataacctggtcatcgccctgca
tagttataaaccctcccatgatggggacctgggctttgagaaaggggagcaactgaggatcctggagcaaaat
ggagaatggtggaaggcacagtccctgaccactggccaggagggctacattcccttcaactttgtggccaaag
ccaacagcctggagcctgagccttggtttttcaaggacttgagccggaaggatgctgaaagacaacttttggc
ccctgggaacactcatggatccttcctgatcagagagagtgagaccactgccggctccttctctctgtctgtg
cgggactttgaccagaaccaggggaggtggtgaaacattacaagatccgcaacctggataatgggggcttct
acatttccccccgaatcacctttcctaatctgcatgaactggtccagcattactccaaagtctcagatgggct
atgcactcgactgagtcggccctgccagacccaaaagccacagaagccctggtgggaagatgagtgggaggtt
cctcgagagacactgaagctggtggaaaagctgggagctggccagtttggggaggtctggatggggtactaca
atgggcataccaaggtagcggtgaaaagcctgaaagcgggcagcatgtct
cctgatgccttcctggctgaagccaacctgatgaaacagcttcagcaccagcgactggtacgcctttatgcgg
tggtcacacaggaacccatctacatcatcactgaatacatggagaatgggagcctggtagacttcctcaaaac
cacaacaggagtcaaactaaccatccacaaactgctcgatatggctgcacagattgctgagggcatggccttc
attgaagagcggaattacatccaccgggacctcagagcagccaacatcctggtgtcagacacattgaactgta
agattgctgactttggattggcccggctgatcgaggacaacgagtacacagctagagagggagcaaaatttcc
catcaagtggacagctcctgaggccatcaactatgggacatttaccatcaagtcagatgtctggtcttttggc
atcctgctcacagaaatcgtcacctatgggaggatcccctacccaggaatgaccaaccctgaggtgattcaga
acctggagcaaggctatcggatggtgaggcctgacaactgcccagaagaactatacaaacttatgatgctgtg
ctggaaggagaggcctgaggatcggcccacctttgattacctgaggagtgtcttggaggacttcttcactgcc
acagagggccagtaccagccccagccatga
3′end
ccttcctattttggtctctcccatcccagtccctggcccacctccccagctttactcttgcctcctaaatgac
cccactcagactgaggcagggaagaatctggcaaaagacagagggcctcaggatggaggtatggggcaccaac
atctcagccatcacacttgtgactccattacctgccgaggaaggaagacagcaaaagacttggcattgggctt
gtcctttgcgtgccccacccccaacatgtggtctcccagactgactccctacctctgtcctctccctgatttc
tcagggccagaggagaagcctggactttgccccagccctttgactctttgtctatgtgttcacccagagtgg
attcactgtctcagtttccataaaattcatattcttagcaaaaaaaaaaaaaaaaaaaaa

Putative protein sequence *M. eugenii*

RCHYPIVPLDAKGTLLMRNGSDVRDPLVTYEGLNPPASPLQDNLVIALHSYKPSHDGDLGFEKGEQLRILEQN
GEWWKAQSLTTGQEGYIPFNFVAKANSLEPEPWFFKDLSRKDAERQLLAPGNTHGSFLIRESETTAGSFSLSV
RDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPNLHELVQHYSKVSDGLCTRLSRPCQTQKPQKPWWEDEWEV
PRETLKLVEKLGAGQFGEVWMGYYNGHTKVAVKSLKAGSMSPDAFLAEANLMKQLQHQRLVRLYAVVTQEPIY
IITEYMENGSLVDFLKTTTGVKLTIHKLLDMAAQIAEGMAFIEERNYIHRDLRAANILVSDTLNCKIADFGLA
RLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTEIVTYGRIPYPGMTNPEVIQNLEQGYRM
VRPDNCPEELYKLMMLCWKERPEDRPTFDYLRSVLEDFFTATEGQYQPQP

Appendix 5C

Secondary structure prediction for *M. eugenii* Lck using PSIpred



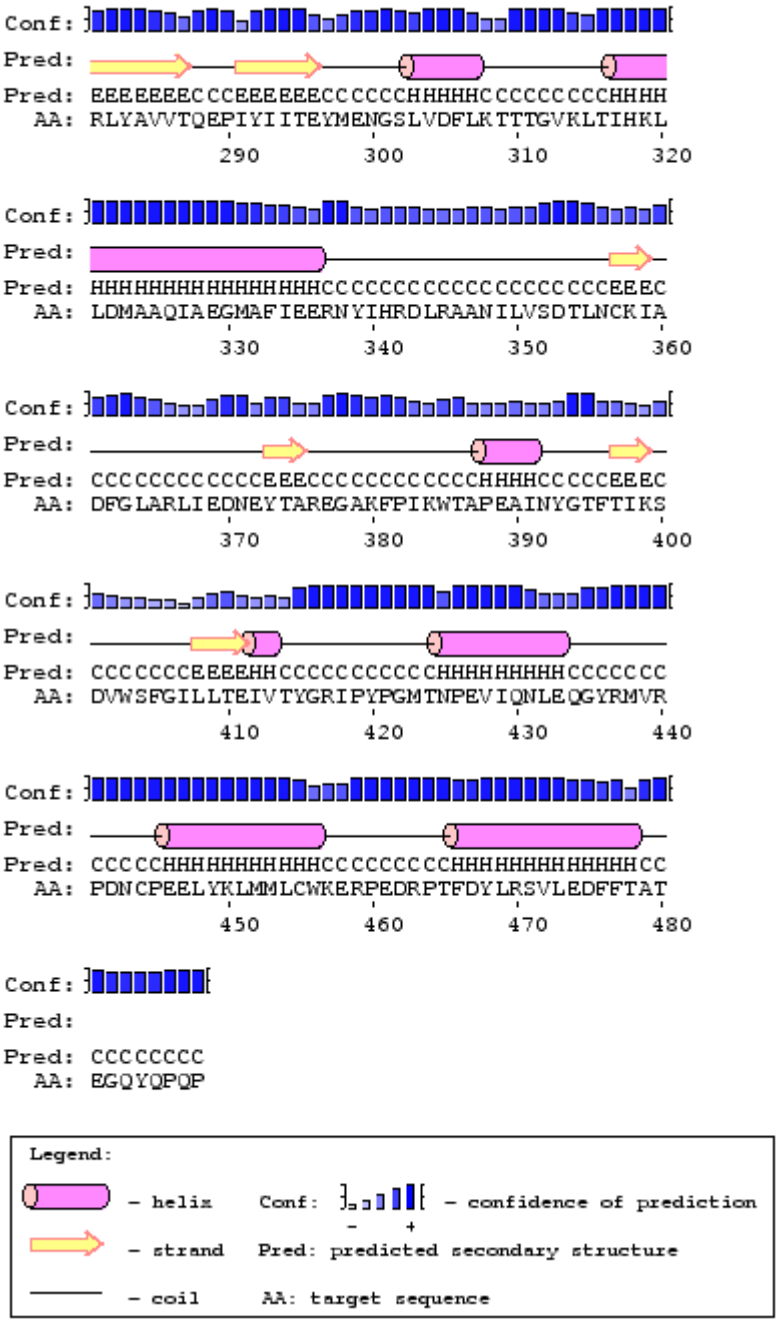
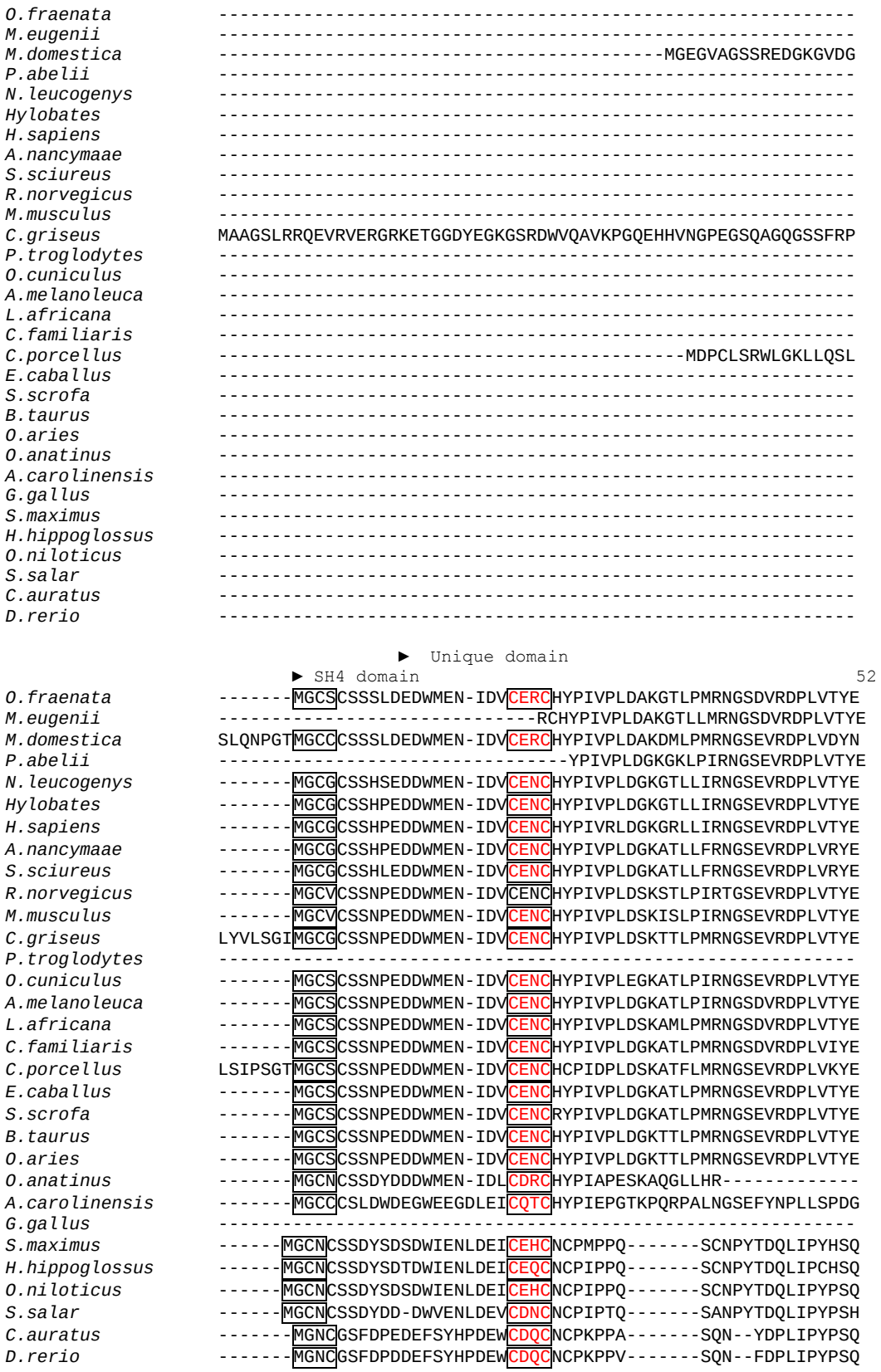


Figure 5C.2. Secondary structure prediction *M.eugenii* Lck.
Yellow arrows indicate beta strands. Purple cylinders represent α-helices. Confidence level of prediction is represented by the blue bars. Numbering indicates the number of amino acids within the sequence.

Appendix 5C

Sequence Alignment for known Lck sequences

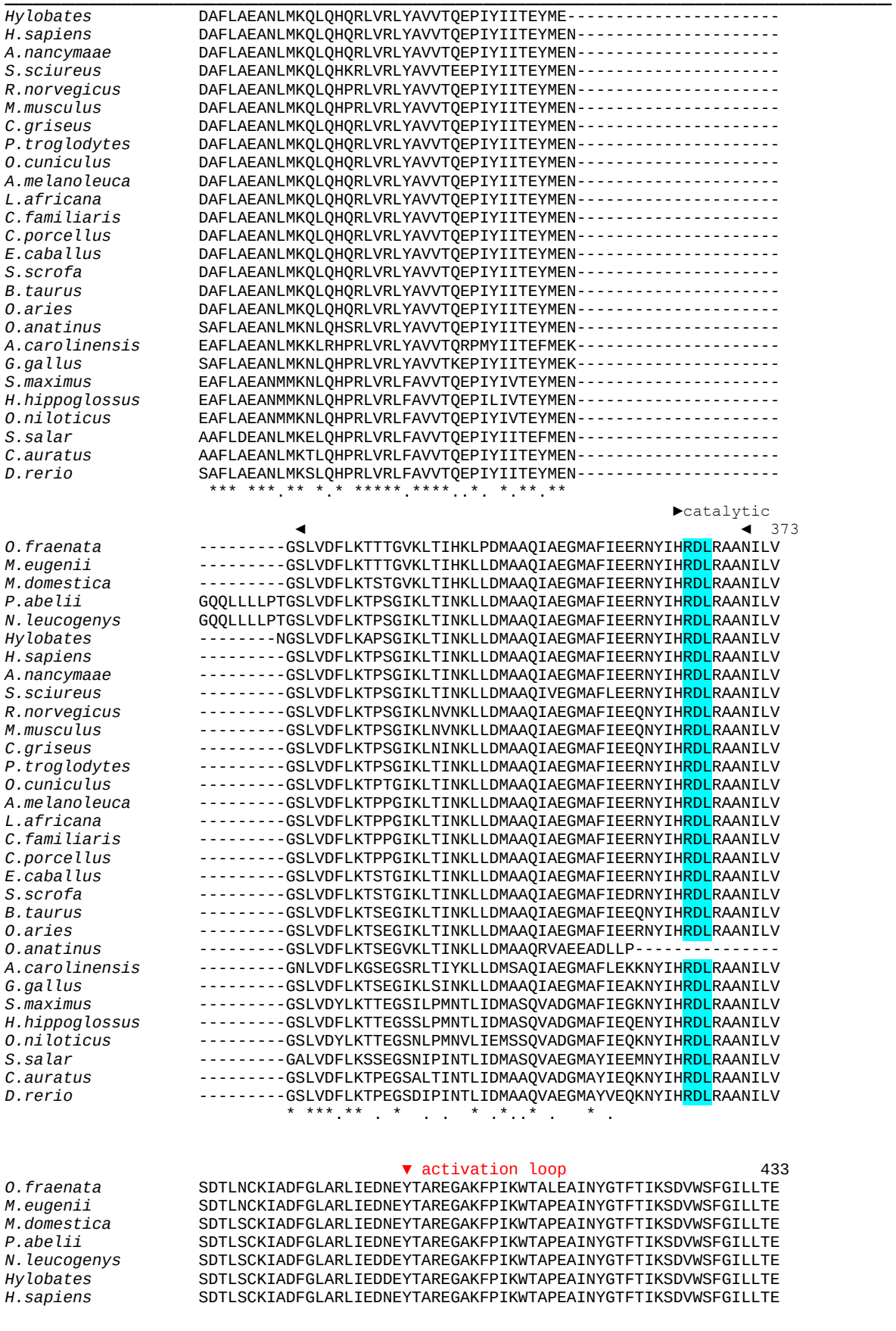


```
                                                                    ►SH3 domain  111
O.fraenata          GLNPPASPLQDNLVIALYNYKPSHDGDLGFEKGEQLRILE-QNGEWWKAQSLTTGQEGYI
M.eugenii           GLNPPASPLQDNLVIALHSYKPSHDGDLGFEKGEQLRILE-QNGEWWKAQSLTTGQEGYI
M.domestica         GLDPPVSPLQDNLVIALHSYKPSHDGDLGFEKGEQLKILE-QNGEWWKAQSLTTGQEGFI
P.abelii            GCNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-ESGEWWKAQSLTTGQEGFI
N.leucogenys        GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QSG----------------
Hylobates           GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
H.sapiens           GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEPLRILE-QSGEWWKAQSLTTGQEGFI
A.nancymaae         GSNPPASPLQDNLVIALHSYKPSHDGDLGFEKGEQLRILE-QNGEWWKAQSLTTGQEGFI
S.sciureus          GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEHLRILE-QNGEWWKAQSLTTGQEGFV
R.norvegicus        GSLPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
M.musculus          GSLPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
C.griseus           GSIPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
P.troglodytes       ----MPLVFTDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
O.cuniculus         GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QNGEWWKAQSLTTGQEGFI
A.melanoleuca       GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
L.africana          GSIPPASPLQDNLVIALHNYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
C.familiaris        GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-RNGEWWKAQSLTTGQEGFI
C.porcellus         DINPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-HNGEWWKAQSLTTGQEGFI
E.caballus          GSNPPASPLQDNLVIALHRYEPSHDGDLGFEKGEQLRILE-QSGEWWKAQSLTTGQEGFI
S.scrofa            GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QNGEWWKAQSLTTGLEGFI
B.taurus            GSNPPASPLQDNLVIALHSYEPSHDGDLGFEKGEQLRILE-QNGEWWKAQSLTTGQEGFI
O.aries             GSNPPASPLQDNLVIALHSYVPSHDGDLGFEKDEQLRILE-QNGEWWKAQSLTTGQEGFI
O.anatinus          ---PVPPFSPDNLVVALHNYEPVHDGDLGFQKSEQLRILE-QSGEWWKAQSLTTGQEGFI
A.carolinensis      MTPPPSSPLQDKLVVALYDYDSTHEGDLVLRTGEQLRVLE-ESGEWWKAQSLTTGRVGYI
G.gallus            -MSPPCSPLQDKLVVALYDYEPTHDGDLGLKQGEKLRVLE-ESGEWWRAQSLTTGQEGLI
S.maximus           HSPP-TSPLPDKLVVAIYSYEPNHDGDLGFEKGEKLKIINKDDPEWYLAESLTTGQQGYI
H.hippoglossus      HSPP-MSPLPDNLVEAIYSYEPNHDGDLGFEKGDKLKIINKDDPEWYLAESLTTGQQGFI
O.niloticus         MTPP-TSPLPVNVVVAIYSYEPTHDGDLGFDKGDKLKILNKDDPEWYLAESLTTGQQGYI
S.salar             LTPP-SSPLPDSLVIAIYSYEPNHNDDLGFEKGDKLKILNKDDPEWFMAESLITGQKGFI
C.auratus           YTPPPSSPLPENLVVALYKYEPCHSDDLGFEKGEKLKILNIDDPEWFMAESLFTGQKGYI
D.rerio             YTPPPSSPLPENLVVAIYKYDPAHSDDLGFEKGEKMKILDCDDPEWYMAESLFTGQRGYI
                                  ..* *.. * . *..** .  .. .....  .
                                                                    ► SH2 domain 163
O.fraenata          PFNFVAKANSLEPEP--------WFFKDLSRKDAERQLLAPGNTHGSFL-IRESETTAGS
M.eugenii           PFNFVAKANSLEPEP--------WFFKDLSRKDAERQLLAPGNTHGSFL-IRESETTAGS
M.domestica         PFNFVAKVNSLEPEP--------WFFKDLSRKDAERQLLAPGNTHGSFL-IRESETTAGS
P.abelii            PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
N.leucogenys        ------------------------------------------------------------
Hylobates           PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
H.sapiens           PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
A.nancymaae         PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
S.sciureus          PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
R.norvegicus        PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
M.musculus          PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
C.griseus           PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
P.troglodytes       PFNFVAKANSLEPEP--------WFFKNLSRXXXXRRFVHP-------------SIHSGS
O.cuniculus         PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFLDSGKSESTAGS
A.melanoleuca       PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
L.africana          PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTVGS
C.familiaris        PFNFVAKANSLEPEP--------WFFKNLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
C.porcellus         PFNFVAKANSLEPEP--------WFFKSLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
E.caballus          PFNFVAKANSLEPEPXVGTRRGGWMGADLGDSGAERQLLAPGNTHGSFL-IRESESTAGS
S.scrofa            PFNFVAKANSLEPEP--------WFFKNLSRRDAERQLLAPGNTHGSFL-IRESESTAGS
B.taurus            PFNFVAKANSLEPEP--------WFFKTLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
O.aries             PFNFVAKANSLEPEP--------WFFKTLSRKDAERQLLAPGNTHGSFL-IRESESTAGS
O.anatinus          PFNFVARANSLEPEP--------WFFKDLSRKDAERQLLAPGNMHGSFL-VRESETTKGS
A.carolinensis      PSNFVAKLNSLEQEP--------WFFKALSRKDAERQLLTAGNTHGAFL-IRESESTKGS
G.gallus            PHNFVAMVNSLEPEP--------WFFKNLSRKNAEARLLASGNTHGSFL-IRESETSKGS
S.maximus           PHNFVGMT-RVETEP--------WFIKNISRNEAMRLLLAPGNTQGSYL-IRESETTPGS
H.hippoglossus      PYNFIAMT-TVETEP--------WFFKSISRNEAMRLLLAPGNTQGSFL-IRESETIQGS
O.niloticus         PHNFVALS-TVETEP--------WFFRNISRNEAMRLLLAPGNTQGSFL-IRESETAKGS
S.salar             PYNFVAPLNSMEMET--------WFFKNLSRNDAMRLLLAPGNTQGSFM-VRESETTQGS
C.auratus           PQNFVAKLNSMETEP--------WFFKNLSRNDAMRQLLAPGNTQGSFL-IRESETTPGS
D.rerio             PKNFVAKLNSMETEP--------WFYKNLSRNDAMRQLLAPGNTQGSFL-IRESETQPGS

                                                                           223
O.fraenata          FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPNLHELVQHYSKVSDGLCTRLSR
M.eugenii           FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPNLHELVQHYSKVSDGLCTRLSR
M.domestica         FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVQHYSKVSDGLCTRLSR
P.abelii            FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFSGLHELVRHYTNASDGLCTRLSR
```

Appendix 5C

```
N.leucogenys          -------------------------DGFYISPRITFPGLHELVRHYTNASDGLCTRLSR
Hylobates             FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTNASDGLCTRLSR
H.sapiens             FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTNASDGLCTRLSR
A.nancymaae           FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTNASDGLCTRLSR
S.sciureus            FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFSGLHELVRHYTNASDGLCTRLSR
R.norvegicus          FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHDLVRHYTNASDGLCTKLSR
M.musculus            FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHDLVRHYTNASDGLCTKLSR
C.griseus             FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTNASDGLCTKLSR
P.troglodytes         FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTNASDGLCTRLSR
O.cuniculus           FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTSAPDGLCTRLSR
A.melanoleuca         FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTNASDGLCTRLSR
L.africana            FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHELVRHYTNAPDGLCTRLSR
C.familiaris          FSLSVRDFDQNQGEVVKHYKIRNLDKGGFYISPRITFPGLHELVRHYTNSSDGLCTRLSR
C.porcellus           FSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLQELVRHYTNAPDGLCTRLSR
E.caballus            FSLSVRDFDQNQGEVVKHYKIRNLDKGGFYISPRITFPGLHELVRHYTNASDGLCTRLSR
S.scrofa              FSLSVRDFDQNQGEVMKHYKIRNLDKGGFYISPRITFPGLHELVHHYTNSSDGLCTRLNR
B.taurus              FSLSVRDFDQTQGEVVKHYKIRNLDKGGFYISPRVTFPGLHELVRHYMNTSDGLCTRLSR
O.aries               FSLSVRDFDQTQGEVVKHYKIRNLDKGGFYISPRVTFPGLHELVRHYMNTSDGLCTRLSR
O.anatinus            YSLSVRDFDQTQGEVVKHYKIRNMDNGGYYISPRITFRSLQDLVKHYSR-----------
A.carolinensis        FSLSVRDFDQDQGEVVKHYKIRNMDNGGFYISPRITFDSLHNLVEHYMRNTDGLCTRLGK
G.gallus              YSLSVRDFDQNQGETVKHYKIRNMDNGGYYISPRVTFSSLHELVEYYSSSSDGLCTRLGK
S.maximus             YSLSIRDLDHNTGEGVKHYRIRNMDNGGFYITTRISFNSLKELIQHHSRDADGLCTKLVK
H.hippoglossus        YSLSIRDLDHNTGEGVKHYRIRNMDNGGFYITAKISFSSLKELVQHHVRETDGLCTKLGK
O.niloticus           YSLSVRDLDHNTGEGVKHYRIRNMDNGGFYITAKISFNSLKELVQHHSRDADGLCTKLVK
S.salar               FSLSVRDLDPDTGDTVKHYRIRNLDTGGFYITAKISFNSLKELVQHHSREADGLCTRLMK
C.auratus             FSLSVRDLDHTMGDIIKHYRIRNLDEGGFYITTKISFSSLSELVKHYSREADGLCTRLVK
D.rerio               FSISVRDLDPMQGDIIKHYRIRNMDAGGFYITNKISFNSLSELVKHYSREADGLCTRLVK
                                             .*.**. ...* .* .*....

                                                    ▼ATP binding site  ▼
                             ►Tyrosine kinase domain                  ▼        283
O.fraenata            PCQTQKPQKPWWEDEWEVPRETLKLVEKLGAGQFGEVWMGYYNGHTKVAVKSLKAGSMSP
M.eugenii             PCQTQKPQKPWWEDEWEVPRETLKLVEKLGAGQFGEVWMGYYNGHTKVAVKSLKAGSMSP
M.domestica           PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKAGSMSP
P.abelii              PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
N.leucogenys          PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
Hylobates             PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
H.sapiens             PCQTQKPQKPWWEDEWEVPRETLKLVERLGAAQFGEVWMGYYNGHTKVAVKSLKQGSMSP
A.nancymaae           PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNDHTKVAVKSLKQGSMSP
S.sciureus            PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNEHTKVAVKSLKQGSMSP
R.norvegicus          PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
M.musculus            PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
C.griseus             PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
P.troglodytes         PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
O.cuniculus           PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
A.melanoleuca         PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
L.africana            PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
C.familiaris          PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
C.porcellus           PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
E.caballus            PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
S.scrofa              PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
B.taurus              PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
O.aries               PCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKVAVKSLKQGSMSP
O.anatinus            ----------------------------------------YYNGHTKVAVKNLKPGSMSP
A.carolinensis        PCQTQKPQKPWWQDEWEVPRENLKLEEKLGAGQFGEVWKGMYNGHTKVAIKSLKQGSMSP
G.gallus              PCRTQKPQKPWWQDEWEVPRESLKLVEKLGAGQFGEVWMGFYNGHTKVAIKNLKQGSMSP
S.maximus             PCQSRAPQKPWWQDEWEIPRESLKLEQKLGAGQFGEVWMGVYNNDRRVAIKNLKIGTMSV
H.hippoglossus        PCQSRAPQKPWWQDEWEIPRESLKLLRRLGAGQFGEVWMGEYNNDREVAIKKLKMGTMSV
O.niloticus           PCQSRAPQKPWWQDEWEIPRESLKLERRLGAGQFGEVWMGVYNNDRKVAIKNLKMGTMSV
S.salar               PCQSRVPQKPWWQDEWEIPRESLKMERRLGAGQFGEVWMGLYNNHRRVAIKNLKVGTMSM
C.auratus             PCQTRAPQKPWWQDEWEVPRESLKLERRLGQGQFGEVWMGLYNNNRRVAIKSLKTGTMSI
D.rerio               PCQTRAPQKPWWQDEWEVPRESLKLERRLGQGQFGEVWMGLYNNNRQVAIKSLKPGTMSI
                                                        ** . .**.*.** *.**

                      Hydrophobic pocket                 ▼  ►    linker           322
O.fraenata             DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------------------
M.eugenii              DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------------------
M.domestica            DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------------------
P.abelii               DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMENDTILDSQLEEKGLGASPWGNL
N.leucogenys           DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMENDTLLDSQLEEKGLGASPWGNL
```

Appendix 5C


Hylobates DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYME----------
H.sapiens DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
A.nancymaae DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
S.sciureus DAFLAEANLMKQLQHKRLVRLYAVVTEEPIYIITEYMEN----------
R.norvegicus DAFLAEANLMKQLQHPRLVRLYAVVTQEPIYIITEYMEN----------
M.musculus DAFLAEANLMKQLQHPRLVRLYAVVTQEPIYIITEYMEN----------
C.griseus DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
P.troglodytes DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
O.cuniculus DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
A.melanoleuca DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
L.africana DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
C.familiaris DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
C.porcellus DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
E.caballus DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
S.scrofa DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
B.taurus DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
O.aries DAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMEN----------
O.anatinus SAFLAEANLMKNLQHSRLVRLYAVVTQEPIYIITEYMEN----------
A.carolinensis EAFLAEANLMKKLRHPRLVRLYAVVTQRPMYIITEFMEK----------
G.gallus SAFLAEANLMKNLQHPRLVRLYAVVTKEPIYIITEYMEK----------
S.maximus EAFLAEANMMKNLQHPRLVRLFAVVTQEPIYIVTEYMEN----------
H.hippoglossus EAFLAEANMMKNLQHPRLVRLFAVVTQEPILIVTEYMEN----------
O.niloticus EAFLAEANMMKNLQHPRLVRLFAVVTQEPIYIVTEYMEN----------
S.salar AAFLDEANLMKELQHPRLVRLFAVVTQEPIYIITEFMEN----------
C.auratus AAFLAEANLMKTLQHPRLVRLFAVVTQEPIYIITEYMEN----------
D.rerio SAFLAEANLMKSLQHPRLVRLFAVVTQEPIYIITEYMEN----------
►catalytic
◄ ◄ 373
O.fraenata ---------GSLVDFLKTTTGVKLTIHKLPDMAAQIAEGMAFIEERNYIHRDLRAANILV
M.eugenii ---------GSLVDFLKTTTGVKLTIHKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
M.domestica ---------GSLVDFLKTSTGVKLTIHKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
P.abelii GQQLLLLPTGSLVDFLKTPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
N.leucogenys GQQLLLLPTGSLVDFLKTPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
Hylobates --------NGSLVDFLKAPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
H.sapiens ---------GSLVDFLKTPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
A.nancymaae ---------GSLVDFLKTPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
S.sciureus ---------GSLVDFLKTPSGIKLTINKLLDMAAQIVEGMAFLEERNYIHRDLRAANILV
R.norvegicus ---------GSLVDFLKTPSGIKLNVNKLLDMAAQIAEGMAFIEEQNYIHRDLRAANILV
M.musculus ---------GSLVDFLKTPSGIKLNVNKLLDMAAQIAEGMAFIEEQNYIHRDLRAANILV
C.griseus ---------GSLVDFLKTPSGIKLNINKLLDMAAQIAEGMAFIEEQNYIHRDLRAANILV
P.troglodytes ---------GSLVDFLKTPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
O.cuniculus ---------GSLVDFLKTPTGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
A.melanoleuca ---------GSLVDFLKTPPGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
L.africana ---------GSLVDFLKTPPGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
C.familiaris ---------GSLVDFLKTPPGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
C.porcellus ---------GSLVDFLKTPPGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
E.caballus ---------GSLVDFLKTSTGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
S.scrofa ---------GSLVDFLKTSTGIKLTINKLLDMAAQIAEGMAFIEDRNYIHRDLRAANILV
B.taurus ---------GSLVDFLKTSEGIKLTINKLLDMAAQIAEGMAFIEEQNYIHRDLRAANILV
O.aries ---------GSLVDFLKTSEGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILV
O.anatinus ---------GSLVDFLKTSEGVKLTINKLLDMAAQRVAEEADLLP----------
A.carolinensis ---------GNLVDFLKGSEGSRLTIYKLLDMSAQIAEGMAFLEKKNYIHRDLRAANILV
G.gallus ---------GSLVDFLKTSEGIKLSINKLLDMAAQIAEGMAFIEAKNYIHRDLRAANILV
S.maximus ---------GSLVDYLKTTEGSILPMNTLIDMASQVADGMAFIEGKNYIHRDLRAANILV
H.hippoglossus ---------GSLVDFLKTTEGSSLPMNTLIDMASQVADGMAFIEQENYIHRDLRAANILV
O.niloticus ---------GSLVDYLKTTEGSNLPMNVLIEMSSQVADGMAFIEQKNYIHRDLRAANILV
S.salar ---------GALVDFLKSSEGSNIPINTLIDMASQVAEGMAYIEEMNYIHRDLRAANILV
C.auratus ---------GSLVDFLKTPEGSALTINTLIDMAAQVADGMAYIEQKNYIHRDLRAANILV
D.rerio ---------GSLVDFLKTPEGSDIPINTLIDMAAQVAEGMAYVEQKNYIHRDLRAANILV
▼ activation loop 433
O.fraenata SDTLNCKIADFGLARLIEDNEYTAREGAKFPIKWTALEAINYGTFTIKSDVWSFGILLTE
M.eugenii SDTLNCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
M.domestica SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
P.abelii SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
N.leucogenys SDTLSCKIADFGLARLIEDDEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
Hylobates SDTLSCKIADFGLARLIEDDEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
H.sapiens SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE

Appendix 5C

```
A.nancymaae         SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
S.sciureus          SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILMTE
R.norvegicus        SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
M.musculus          SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
C.griseus           SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
P.troglodytes       SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
O.cuniculus         SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
A.melanoleuca       SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
L.africana          SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
C.familiaris        SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
C.porcellus         SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
E.caballus          SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
S.scrofa            SDTLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
B.taurus            SHSLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
O.aries             SHSLSCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
O.anatinus          ------------------------------------------------------------
A.carolinensis      SEAICCKIADFGLARLIEDDEYTAQEGGSAL----------------------------
G.gallus            SEALCCKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTE
S.maximus           SQELICKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFSIKSDVWSFGILLTE
H.hippoglossus      SHEHICKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFSIKSDVWSFGILLTE
O.niloticus         SHELICKVADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFSIKSDVWSFGILLTE
S.salar             SDELICKIADFGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFSIKSDVWSFGILLTE
C.auratus           SQELICKIADFGLARLIENNEYTAREGAKFPIKWTAPEAINYGTFSIKSDVWSFGVLLTE
D.rerio             SHELTCKIADFGLARLIKNNEYTAREGAKFPIKWTAPEAINYGTFSIKSDVWSFGVLLTE

                                                                          493
O.fraenata          IVTYGRIPYPGMTNPEVIQNLEQGYRMVRPDNCPEELYKPMMLCWKERPEDRPTFDYLRS
M.eugenii           IVTYGRIPYPGMTNPEVIQNLEQGYRMVRPDNCPEELYKLMMLCWKERPEDRPTFDYLRS
M.domestica         IVTYGRIPYPGMTNPEVIQNLEQGYRMVRPDNCPEELYKLMMLCWKERPEERPTFDYLRS
P.abelii            IVTHGRIPYPGMTNPEVIQNLERGYRMVCPDNCPEELYQLMRLCWKERPEDRPTFDYLRS
N.leucogenys        IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMMLCWKERPEDRPTFDYLRS
Hylobates           IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMMLCWKERPEDRPTFDYLRS
H.sapiens           IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMRLCWKERPEDRPTFDYLRS
A.nancymaae         IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYHLMMLCWKERPEDRPTFDYLRS
S.sciureus          IVTHGRIPYPGMTNPEVIQNLERGYRMPRPDNCPEELYKLMMQCWRERPDDRPTFDYLRS
R.norvegicus        IVTHGRIPYPGMTNPEVIQNLEKGYRMVRPDNCPEELYHLMMLCWKERPEDRPTFDYLRS
M.musculus          IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYHLMMLCWKERPEDRPTFDYLRS
C.griseus           IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQIMMLCWKERPEERPTFDYLRS
P.troglodytes       IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMRLCWKERPEDRPTFDYLRS
O.cuniculus         IVTHGRIPYPGMTNPEVIQNLERGYRMVRPENCPEELYHLMKLCWKERPEDRPTFDYLRS
A.melanoleuca       IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYHLMMLCWKERPEERPTFDYLRS
L.africana          IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYDLMLLCWKERPEDRPTFDYLRS
C.familiaris        IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMMLCWKERPEDRPTFDYLRS
C.porcellus         IVTHGRIPYPGMTNPEVIQNLERGYRMVQPDNCPAELYQLMMQCWKERPEDRPTFDYLRS
E.caballus          IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMMLCWKERPEDRPTFDYLRS
S.scrofa            IVTHGRIPYPGMTNTEVIQNLERGYRMVRPDNCPEELYHLMMLCWKERPEERPTFDYMRS
B.taurus            IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMMLCWKERPEERPTFDYLRS
O.aries             IVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMMLCWKERREDRPTFDYLRS
O.anatinus          --FFIGPGGAGMTNPEVIQNLERGYRMPQPENCPEELYDLMQLCWKEKPENRPTFEYMKS
A.carolinensis      ------------------------------------------------------------
G.gallus            IVTYGRIPYPGMTNPEVIQNLERGYRMPQPDNCPQELYELMMQCWKEQPEERPTFEYMKS
S.maximus           IVTYGRIPYPGMSNPEVIQNLERGYRMPKPENCPEGLYQVMGMCWKENQDDRPTFEYLRG
H.hippoglossus      IVTYGRIPYPGMSNPEVIQNLERAYRMPKPDNCPEGLYNVMGMCWRETPDDRPTFEYLRS
O.niloticus         IVTYGRIPYPGMSNPEVIQNLERGYRMPQPDNCSDALYSIMCHCWKESPEERPTFEYLRN
S.salar             IVTYGRIPYPGMSNPEVIQNLEKGYRMPRPENCPEDLYNIMNLCWKESPENRPTFEYLRS
C.auratus           IVTHGRIPYPGMTNPEVIANLERGYRMPCPDNCPEDLYDVMKRCWTENPDSRPTFEYLRS
D.rerio             IVTYGRIPYPGMTNPEVIANLERGYRMPCPDNCPEALYNVMKHCWTENPDNRPTFEFLRS
```

Appendix 5C

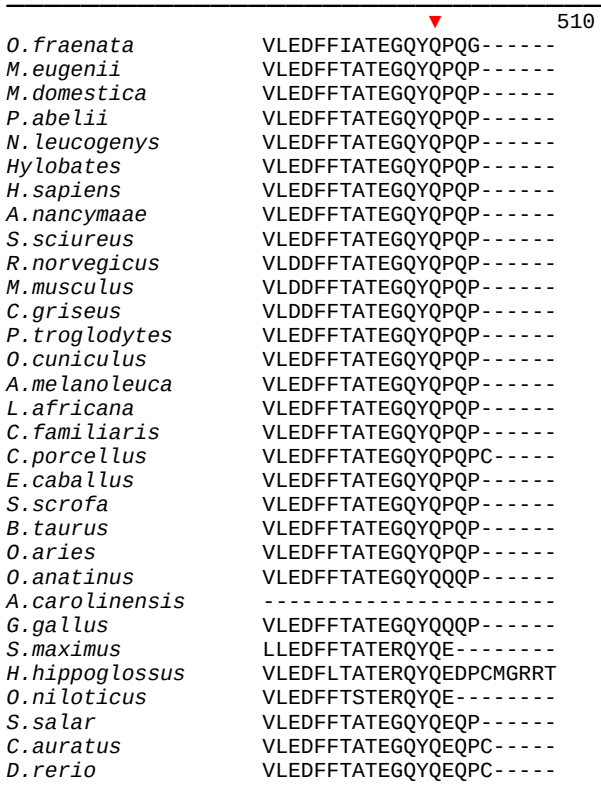


Figure 5C.1. Multiple sequence alignment for Lck.
Boxed sequences are the myristoylation (G^2) and palmitylation (C^3) in the SH4 domain. CXXC = motif in the unique domain. ▼ = tyrosine residues related to activation and inhibition of the PKC activities. * = fully conserved residues. RDL = PP2 binding site. Residues marked in brown and bold mark the hydrophobic pocket. Lysines (K) in green and bold signify the ATP binding site.

Appendix 5C

Table 5C.1. Genbank accession numbers and the relevant references.

Species Name	Common Name	Accession Numbers	References
A.carolinensis	Green anole	XM_003228719.1	Annotated
A.melanoleuca	Giant Panda	XM_002921866	Annotated
A.nancymaae	Ma's night monkey	AY821852	Direct submission
B. taurus	Cattle	NM_001034334.1	(Wang *et al.,* 1991)
C. auratus langsdorfii	Japanese silver crucian carp	AB279595	(Araki *et al.,* 2007)
C. griseus	Chinese hamster	XM_003500655	Annotated
C. jacchus	White tuffed ear marmoset	XR_088449.1	Annotated
C. l. familiaris	Dog	XM_846879.2	Annotated
C. porcellus	Domestic Guinea Pig	XM_003471081	Annotated
D. rerio	Zebrafish	AY390224.1	Direct submission
E. caballus	Horse	XM_001917285	Annotated
G. gallus	Chicken	XM_427615 J30579	Annotated (Strebhardt *et al.,* 1987)
H. hippoglossus	Halibut	FJ769822.1	(Overgard *et al.,* 2010a)
H. sapiens	Human	M36881.1	(Perlmutter *et al.,* 1988)
Hylobate	Gibbon	AJ320182.1	Direct submission
L. africana	African elephant	XM_003415462	Annotated
M. domestica	Opossum	XM_001366178.2	Annotated
M. mulatta	Rhesus monkey	XM_001109718	Annotated
M. musculus	Mouse	X03533	(Voronova and Sefton, 1986)
N. leucogenys	Northern white-cheeked gibbon	XM_003276431.1	Annotated
O. anatinus	Platypus	XM_001509225.2	Annotated
O. aries	Sheep	NM_001142515.1	(Yu *et al.,* 2010)
O. cuniculus	Rabbit	XM_002720629	Annotated
O. mykiss	Rainbow trout	NM_001124542.1	(Laing *et al.,* 2007)
O. niloticus	Nile tilapia	XM_003446183.1	Annotated
P. abelii	Sumatran orangutan	XM_002811121	Annotated
P. troglodytes	Chimpanzee	XM_003307949	Annotated
R. norvegicus	Rat	NM_001100709.1	(Shin and Steffen, 1993)
S. salar	Atlantic Salmon	NM_001139907.1	(Leong *et al.,* 2010)
S. sciureus	Common squirrel monkey	AJ277921	Direct submission
S. scrofa	Pig	NM_001143713.1	(Yonggang and Xueshan, 2011)
S.maximus	Turbot	DQ848967	Direct submission
X. Silurana tropicalis	Western clawed frog	XM_002939295.1	Annotated

Appendix 6A

Macropus eugenii interleukin-2 nucleotide sequence

5′
ccactctctaatcactacccagagtaacctgaagttactgagctcttgacacg
atgaacaaggtcccgctcttgtcctgtattgcactaactcttgttttggttgccaacggggcaccgacattgc
ctcctcccaccactgtgctgcagtacttactacgtgacttaatggaggtccaaaacaaactcaagattgtatc
agaaaggatgaagaagtatgaactctacatcccaagtaataccagcagcatcgaaaatctgcagtgtttcact
aaagaactgaaccctgtggctggtgcgttgaaatatgaatccaaagatgctcagaatattcaggaatacatca
acaacattaatgtgactgttaatagcttaatgggaccagaaataacacagtgtcactacgcctccaagatgag
gattgaaggattctttaaagaatttgtttccgtctgccaaagattcatgcactaa
3′
catgtctggatgaagtt**atttattta**aat**attta**aattttata**attta**tttctgaatgtgtggtttattatc**a**
tttactaccactgattttagtcgtcagataatgaatgtaatagatcttatgatcctaactgtaagccctaggg
gctcaaagactactttaagttatcatctcaaaaatt**attta**ttaaattatgacttgttaaatgtaatgtctgt
gaaggtcagtaaagtt**attt**ataaatctgatgatggataataaaaaaaaaaaaaaaaaaaaaaaaaaa

Macropus eugenii putative protein sequence

MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIPSNTSSIENLQCFT
KELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASKMRIEGFFKEFVSVCQRFMH

Exon predictions for *M. eugenii* IL-2 (using ensembl, BLAT and Genscan)

Exon-1
ATGAACAAGGTCCCGCTCTTGTCCTGTATTGCACTAACTCTTGTTTTGGTTGCCAACGGGGCACCGACATTGC
CTCCTCCCACCACTGTGCTGCAGTACTTACTACGTGACTTAATGGAGGTCCAAAACAAACTCAAG
Exon-2
ATTGTATCAGAAAGGATGAAGAAGTATGAACTCTACATCCCAAGTAAT
Exon-3
ACCAGCAGCATCGAAAATCTGCAGTGTTTCACTAAAGAACTGAACCCTGTGGCTGGTGCGTTGAAATATGAAT
CCAAAGATGCTCAGAATATTCAGGAATACATCAACAACATTAATGTGACTGTTAATAGCTTAATGG
Exon-4
GACCAGAAATAACACAGTGTCACTACGCCTCCAAGATGAGGATTGAAGGATTCTTTAAAGAATTTGTTTCCGT
CTGCCAAAGATTCATGCACTAA

Amplification of exon-4 in genomic DNA

```
gctacaaccaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgaaggcattcatagaaa
attcacacgtacagaatgagagaggtaaagtacaagtttgagaacagtacaaaatgaagc
aagatttgtctagtaccatgatttttaagttagtttatttcaaaaatgttagaagacttc
atacttaatgtaacaatctctagcatgtcaacttattgacatgttgcagacctaatgttg
gtctgtctgaacgtccattgtctactgtagttggaatagattcatcaccaggtcagatac
aaagggatgaattgggggtaggagcggtagttttaactttcaaagactgtatactccatt
gtaaggggttatgatacacttcagagaaagcatgtccacaccgcaaaagtcatagaggct
caaggtcctgaagtaagaatgcttaatgtttgaaaagaaacagttctgaactgggtaggg
tctgatctccctaaaaaggagcccatttaaaagtcacttgagaggcatcacataacattt
ccttgttctgtttttgtgtttctttagggaccagaaataacacagtgtcactacgcctcc
                          G  P  E  I  T  Q  C  H  Y  A  S
aagatgaggattgaaggattctttaaagaatttgtttccgtctgccaaagattcatgcac
 K  M  R  I  E  G  F  F  K  E  F  V  S  V  C  Q  R  F  M  H
taa
 -
```

gDNA

IL-2

Gel picture of *M. eugenii* IL-2 genomic DNA

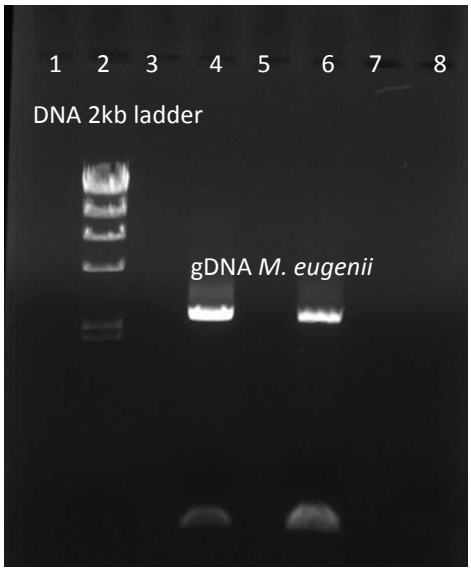


Figure 6A.1. gDNA *M. eugenii* on 1.5%gel, runtime 1.5h at 100V

Polymorphisms in *M. eugenii* IL-2

```
C1F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C1R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C2F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C2R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C3F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C3R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C4F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C4R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C5F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C5R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C6F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C6R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C7F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C7R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C8F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C8R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C9F             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C9R             MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
C10F            MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLNIVSERMKKYELYIP 60
C10R            MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLNIVSERMKKYELYIP 60
                ********************************************.**************

C1F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C1R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C2F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C2R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C3F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C3R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C4F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C4R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C5F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C5R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C6F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C6R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C7F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C7R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C8F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C8R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C9F             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C9R             SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C10F            SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
C10R            SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
                ************************************************************

C1F             MRIEGFFKEFVSVCQRFMH 139
C1R             MRIEGFFKEFVSVCQRFMH 139
C2F             MRIEGFFKEFVSVCQRFMH 139
C2R             MRIEGFFKEFVSVCQRFMH 139
C3F             MRIEGFFKEFVSVCQRFMH 139
C3R             MRIEGFFKEFVSVCQRFMH 139
C4F             MRIEGFFKEFVSVCQRFMH 139
C4R             MRIEGFFKEFVSVCQRFMH 139
C5F             MRIEGFFKEFVSVCQRFMH 139
C5R             MRIEGFFKEFVSVCQRFMH 139
C6F             MRIEGFFKEFVSVCQRFMH 139
C6R             MRIEGFFKEFVSVCQRFMH 139
C7F             MRIEGFFKEFVSVCQRFMH 139
C7R             MRIEGFFKEFVSVCQRFMH 139
C8F             MRIEGFFKEFVSVCQRFMH 139
C8R             MRIEGFFKEFVSVCQRFMH 139
C9F             MRIEGFFKEFVSVCQRFMH 139
C9R             MRIEGFFKEFVSVCQRFMH 139
C10F            MRIEGFFKEFVSVCQRFMH 139
C10R            MRIEGFFKEFVSVCQRFMH 139
                ******************
```

Figure 6A.2. Non-synonymous substitution at position 46 lysine to asparagine

Appendix 6A

Identification of the promoter region of the *M. eugenii* Interleukin-2 gene.

>**Scaffold84654 dna.scaffold scaffold.Meug_1.0.Scaffold84654.1.9143.1**

```
CACACACACACATATATATATGAGAACATCCCTGGCATCT
GTCCTAACTCATCTCTTGTGATCAACTTAACTTGCCTTTTAGATACCTCAAACTGGATGT
CCCATAGGCATCTCAAATTCAATATGTATATAACACTTTTGCCCAAACCCACCCTTCTTT
CCAGCTACCTTCTTACTGTCAAGGGTAATACCATCCTCCCAGTCACCCAGCCTTACAAAC
CAGGTGTCATCTTCACCTCCTCCCTCTCATGACCCCCCAAATCCAATCCATCACCAAACA
TTTTCTCTCCCCTTACATCCTTTCTCAGCACTCATCCAGCCATGGCTGTAGTTACCTCTC
TTACCTGAACTACTGTCATAGCCTTCCAATAACTAGCCTTCCTGCTTCAAATCTCCCCCC
ACTCCAATCCACCCTCCTCTCAGCAGCCAGAGTAACTTTTTATAAAGTAAGGCTTTCTAT TATA
                          Transcription
AAATGTTACCTCCGTCCTTGCCCCCAACTCGATGAACTCCAGGGGCTCCCCATGACCTAC
AGGAGCAAACCTCGGCTCCTATTTGGAACTGAAATGTCTTCACAGCCTCCAGTCTTTTCA 5'end
                                    Translation
CACTTAGTATAGCCATCTCCTGGCAGCTCCTCAAATGGCCATTCTCCCCTTTCTGCAGC
TATGCACTGGCTGACCCCCCTGCCTGGAACACTTTCTCCCCTCACTGCCACCTCACAGTT
TCCCCGACTTCCTTCCTGACTCAGTTCAGACCCCACCTTTTGTAGGTGGTCTTTCCTGGC
TCCCACCCAGCAACTGCAATATCTTCTCTGACACTGCCTTCTCTTCAGTCCACTTCATAG
CTTCTAGTTATTCACATGGTACCTCCCTTTAGAAAGTAAGTTCCCTGAGAGCAAGGTCCA
TTTTTGTCTTTATAACTCCGGTGTCCAACCCAATGTCCGGCACATAGAAAGTACTGAAGA
AATGTTTGCTGAGCAGATTAATGACGGATGAATACAGAAGTGTGTAAGAACAAGGGGAAC
TAAACACTCAAAAGGAGCGGTACTTGGCAAGGAATTCTAAGAATAGATTTGGCAACTTGT
ACTGCAGAAAAAGAAGATTGTATGAGGGATGTGTCCGTGCTGTTCATTCACATTGTGCAT
CTCCTACCTTTCACTCTCACTGGGAAACTAACCCATCTCCTCTTAAACCTTGTTTTTCGT
TTGATACATATATACATACCTGGTG
T
CGGGAGGGGATGAGGGAGGGTTTCATGGAGGTGGTGCTTGACCTGTGTCCTGAAGGAAAA
GACTCTACACGCCAATGTCCATGCCATCCTATAAGCTTTCCCCAAGGGTTTTGCCCAGTG
TGCCAGGGGAGCCCTTGCTACTGCATACCTGCCAATCAAACTCATGAGCTCTCTGCTGGT
TTTCCCTCATTCTCACTCTATGACTGATTCATTTCTTTCTGCAAATGTACAATTCTCTAA
CCTTCTTGACATGTATGTATGTGATATATGCAGGGATGTGCTGGATCAGAAGAGCTAACT
AGAGTTAAATTTTTAGGATGAGCATTTACACCTTAGAAATCAGCAAATCACTGGGCAAAT
CAACCTTGGTTTCTCCGTCTGTAAAATGGGGTTAATAATCGCACCTGCCTTGCAGGGTTC
TTGTGAGACTCAAGTGAGACAATATTTGTAAAGGGCTTAGCAGAGTGCCTGGCCCATAGA
AGGTGAGATATAAATATTTATTATTATTGATACCCAAAGTATGAAAGCATATCCCTTGAA
GATACAACAGTCCAGACTTCTATTATCTTGTCAGATCAGCAACATGGATTAGTGATAGAG
AATTCTGGCCCTGGAGTCAGGAAGACCTGGGTTTAAATGCCGTTTCTGACAAATAATAGG
TGATTTCACCTGTTTAAGTCTCTGGCCAGTCTCTAAGACTTATCTAATAAGTCCCAGACT
GGTTGCACCCTCCCTCCTAGAAATCCCGGATAGCTAATGTGTACATGATATATCCTTCCT
TCCCACTTACCACCCTGAAGTGACTTCAGGGCAATTACCTCATTCATTCTTGTCTCCCTA
TCCCAAGGGACTGGCATAATAAGACAAACTTATGGCTCTTAAGAAATACTTGCTATATTG
AGTTAAATGAAATACTTTTCCTTTCCTCACAAAACAATGGGATTGTTATAAATCAAAATA
AGAACATATTAAAAGGTTGTTTTTAGGAAGTGACAATGTCAACCAAAGAGGGGGCTAGAG
AGAACACCCCAACATATCTCCCCTCCCACCAAGCGCTTACTGAGCACTGTGCTAAATGGA
AATTCACTGGTCACTAGAGGCTTTGGATGAATGATTAGTAAAGTCACCACATAGTTAGGA
AAGGCTACTCAAAGTCCTACTCCAGGTCAACTTTCATTGGGTGATACTCATTCCATCTTC
AAAGAGAAAGTCAGAGGCATGTCACTCAAGTCGTGGTTAATAGAGTGTACTGACAACATA
ACCACTTCAGCCAGAAGTCAGCCTATGTGTCAGTCAAGGTAAAAAGCGAAAGTATGTCAG
AAGGAAGAGCATCTCCTTCAGCGTTTCAAAAACTAAGCATTCAGCTGCCTCCTGCACAGG
TAGGCAGTGCTGGCTGCCCAGCAGGGGAGACCAGTGGGCACCTGGGGTTCAGTTTTCACC
AGGGTCCCAGAAACGGATCTGAACGCTTTGGGGCCTTCCCCTGTCTCGAATGGAAAGAGT
TATGGAATCTTGAAGACTAAGATTCTGTGAGAACATCAAAGGTCTCATTGGAAAGGCTCC
CACACAGAAAATTGTGTGGGCCTTCAATTTTCCACTCTGGAAGGATGATTGTCTGAGCTA
AAAATGGCAGGATGAGCCCATAAGCTCCTCACCCAGCTATCAGATCGATGCTAGGAGGAT
TAAGCAGCCCTTCCCCTCCACAAAAGACACACCTTGTGGTTTAGACTTGGAAGGTCCAAG
ATGAGTTTGTATTTTCTGCTAGACAGAAGAGTAGTCATGGAATCTGTGCTTTGGGATTCC
TTCATTCTCCATAAGTCTTTTTTCATCACCCCCTCCATCCATATCTCCATCTGAGGAACT
```

ATGATTTACAATGGTGATTTTATATTCGTTATAAAAATAATAGCTCACCTTGTATAGGAT
TTTACAAAGTCCTTTCCTTACAATAACCTCTAAAGATACGTAGTGCATGGATTAGGAATC
CCTTTCCCAAGGCCCTTGAGCCTCAGAGGTTACCTGGCAAAATAATAGTAGAACTAAGAC
TTGGAAATAGTTTCTCTGACTCTAAGTTCTATACGCTTCCCACTAAACTGTATGAGAAAT
AGAGAGAGAAGATAGAGACAGACAGGCAGAGAGATGGATAGAGTTCAGTAAAATGAAAGG
GCTCAATAGTATTATCTATTAGACAACACGTAAGACGACCAGACTTGAAATCAGGACCTG
AATTCTAATCCTACCTCAGACATTTTCTAGCTGTGTGATCCTGGGCTAGTCATTCAACCT
ATGCCAGCCTCAGTTTCCTTATCTGTAAAATGGGGGGGTTAAAAATCACACCTATCTCAA
AGGATTGTCTGGGAGGATCAAGTGAGATAACACATGTAAAGCGTTTTGCAAATCTCAAAG
AGTTAGCTTTTATTGCCCCTCTCCCCAACCTAAATTACTCTCTGGTCATGTGGCTGAGGC
TCCTAGAGGCATTTTCTTTTCTTTTGTGCCTACGTAAGCCATTGTTTTAGCTAAAAGACG
CTTCCACGAGTATCTCCTGAGAATACTGTCCTTGGCCCTCGCTGAGTTTTAGTGAAAAAA
GGCTTTGTATAATGGAAAGTAGTTATAAAGAATGTAAAAGTGAATGGTGAGCATAGTAAT
GAATCCTATCAAGAAGAAAGAAAAAAAACTTGATAAGATCATACAGTATGGAAAGGACCC
TAACTCCTGTGGAGGCACAGACATTAGGGAGGACTCCCCAGTGAAGAGCCCCTTTTATCA
ACACTGCTTGGTCTGAGCCTTTTAACTTAATCATCTTAAAGAGTCTAGAACAAAGGGTTC
AAACAGAACATTCCTGTGTAGCTCAAACAAATTTAAAATGTGATTGGGAAATATTTAACA
AAATAAAAGAAATACAATAAAACCTGGATAACATATTTTAAAATTAACTCCATATGGGGC
CTGCAGATCTCCTTACATGTGCAGTTGTTCAGTCATGTCCGACTTCATGCCCCCTTTTGA
GGTTTTCTTGGCAGAGATCCTGGAGTGGTTTGCCATTTCCTCTTCTTTTACAGACGAGGA
AACTGAGACCAACAAGGTGAAGTGACTTGCCCAAGGTCACACAGCTAGCAGGTATCTGAG
GCTGTATTTCAACTCAGGTCTTCCTGACTCCAGGCCCAGCATTCTGTCCACTTCGCCACC
TAACTGCCCAGGTGGATTAATGACTTCCATTTCTATTTGAGTTTGAGGCTTTGCGCCTAT
AACCCAAGGATACACAGCGAACATGCATCCACTCATCTATGTCCAGGTGGGAGCTGGACT
AGATGATCTCTGGGGCACTTTTTGGACATGAGTCTATTTTATCCCAGTGGAGTTTCTATA
TGTTTACAGTGAGTAGAATTCAAGGAAGCATGCTCTTCTATTAACTGAATCACATTTACA
CCATGCTTAGCTTATAGACTGAAATGTCACTAGTTCCCCGTTTGTGTAGAACAAGTAATT
TCTGCTCACATCACTTTGTAGGCATTAGCTCAGTCTTTGTAAAAGAAAGGCAAAACCACC
CTAAAACAGGAAACCAATATTCTTCCTGTTTAATCAACAAATCTAAAATTCTATTTTGTT
TATCTGTTTATATTTCTCCCCCTGCACATGTCCACTGAAA
TTGAT
TGTCCCAGTTCTGTGAGGTAGGTGTTATTGTCATTCTCATTTTACAAATGATATTAATGA
GGCTGGGAGAGATAAAGGAACCTGCCCAGAGTCACACAGCTAGAAAGAATCAGAGATGAG
GTTTGAACATGTTGGAGTCCTCCTGTCTCCAAAGCCTGCCCTGAAGCCTCTAGGAGACCT
AGCTACCTCTGATGGGACTTGAGTCTCACCTGTAATGAAGCTAAATAGGCCAGACCCAGG
TGAAGACGGGATGTGCTCTGGGCAGGAAGGCTGCCTCTGCTAACGCACAGAGATGGGAAA
AGAGGGGAAAGCCATGTTCAGGGAACTAACTGCTGGGAGGCCAGTTTGACTTGAACGTAG
GCTGCATAAAGGGGTGTAATGTGGCCTTGGTCTGAAAAGGTCAATTGCAATCAAACTACA
GAGAGCTTAGAAAGCTCCTCTGAGCAAAAGGTTTTAGAAAGGCTTCAAAGTATTTCATTA
AAAAGTGAAGTAATTTTAACAACCTAAAACAGAGACCACTGGCCTGTCAGCAAGATGCAG
ATTCCAATTCTGCTTCTTACTCTAACAGCCCAGCGGGTCCCTGGGCAGGTCATTCAGCCT
CTTTCAGTTTGTTTTCCCATCTTTAAAAAGGAAATAATCATTGCACCAACCTCACAGGTT
GTGAGGCTCAAGTGCAACAATTTATAAATATTATCATGGTTACTCTCCCATAAAAGAAAG
AACCCATCTTGTGGTCCAAAGACTGGGGCTCCAGTCTCATATTTAGGACCTTTATTTAAT
CAGTCACTTAAGCTCTCTAAATTAATTTTCTCATTTGTAAAAAAAAAAAAAAAAAGAAAG
AGATCAACCCTTGTCCAATTCACAGTTGCTGTCTGGAACTAGACAGAGATATAGATGCAT
AGACATGTAGACAAATGAAGTGTTAGTGTAAATGTGTGTAGATGTGCGTAGGCTAGCATT
TACATAATGTTTTAAGCTTTGCAAGACACTTTACAAAGATCATCACAGTTCTTCTCCTCC
CAACAACCATGGAAGGCGGATGCTAGTATTATCCTCACTCTACAGATGAGTAAACTGAGG
CAGACACAGCTTAAGAGATTTGCCCAAGGTGATACAATCAATAAGGGTCTGAGGTTGGAT
TTTAATCCAGATCTTCCTGACTCTGACACTCTAATCGCTGTGCCACCTAAAGGCCCCCTA
CATATGAAAGTGCTTGGTAAACGACAAGGCACTATGTAAATCTCAGATTAGTAGTATTAA
CAGAAAACTGAAAGTAAAGTTTAAAGCAAGTTCTTTCAGAATCTTCCACATATAATGTGT
TTTGAGGATGCAGCTGGAGAAAAATGTAGCAATAAAAACAAAGAAATGTTTAACACTTGG
TTATATGTGACAGGACCACAGCTATATGTAAA**ATTGCAT**GAACAAAACCTGGCCTCATTC **IgNF-A**
CTATACTACAATCCATCTTCAGTACCTGGAAGGGAGAAAACCCACCTAGCCTGCTCCTCT
GGAGAGGGGATCAGGATAATTCTCCATACCCAGCTTTTCCCTCCCTGGGCTTACAGAATC
TCTCACTCTTTTGACTCCAG**ACCAGCAGCATCGAAAATCTGCAGTGTTTCACTAAAGAAC**
TGAACCCTGTGGCTGGTGCGTTGAAATATGAATCCAAAGATGCTCAGAATATTCAGGAAT **IL-2 (SSIEN..)**

Appendix 6A

ACATCAACAACATTAATGTGACTGTTAATAGCTTAATGGTAAGGAGGATATTATTTTTTT
TTATTTCTTTCAAACTTAATACAAGAGAAAAGAATTATATCCTAAAGACTGCAAGATTAA
ACAATAGTCTAGACTAGTGTTGTGACATAGGATTAAGAGGTATTCACCCTTTTGGTGTTT
GGGGACACACCACCTGTATATTTGACTCTCACAATCATGATTA

>scaffold.Meug_1.0.Scaffold84654.8521.9259.1
TTAAAGCAAGTTCTTTCAGAATCTTCCACATATAATGTGTTTTGAGGATGCAGCTGGAGA
AAAATGTAGCAATAAAAACAAAGAAATGTTTAACACTTGGTTATATGTGACAGGACCACA
GCTATATGTAAAATTGCATGAACAAAACCTGGCCTCATTCCTATACTACAATCCATCTTC Oct1
AGTACCTGGAAGGGAGAAAACCCACCTAGCCTGCTCCTCTGGAGAGGGGATCAGGATAAT
TCTCCATACCCAGCTTTTCCCTCCCTGGGCTTACAGAATCTCTCACTCTTTTGACTCCAG
ACCAGCAGCATCGAAAATCTGCAGTGTTTCACTAAAGAACTGAACCCTGTGGCTGGTGCG
TTGAAATATGAATCCAAAGATGCTCAGAATATTCAGGAATACATCAACAACATTAATGTG BLASTx IL-2
ACTGTTAATAGCTTAATGGTAAGGAGGATATTATTTTTTTTTATTTCTTTCAAACTTAAT
ACAAGAGAAAAGAATTATATCCTAAAGACTGCAAGATTAAACAATAGTCTAGACTAGTGT
TGTGACATAGGATTAAGAGGTATTCACCCTTTTGGTGTTTGGGGACACACCACCTGTATA
TTTGACTCTCACAATCATGATTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNN

2000bp 5’ and 3’ added to above

>scaffold.Meug_1.0.Scaffold84654.6821.10959.1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTGAT
TGTCCCAGTTCTGTGAGGTAGGTGTTATTGTCATTCTCATTTTACAAATGATATTAATGA
GGCTGGGAGAGATAAAGGAACCTGCCCA**GAGTCA**CACAGCTAGAAAGAATCAGAGATGAG AP-1
GTTTGAACATGTTGGAGTCCTCCTGTCTCCAAAGCCTGCCCTGAAGCCTCTAGGAGACCT
AGCTACCTCTGATGGGACTTGAGTCTCACCTGTAATGAAGCTAAATAGGCCAGACCCAGG
TGAAGACGGGATGTGCTCTGGGCAGGAAGGCTGCCTCTGCTAACGCACAGAGATGGGAAA
AGAGGGGAAAGCCATGTTCAGGGAACTAACTGCTGGGAGGCCAGTTTGACTTGAACGTAG
GCTGCATAAAGGGGTGTAATGTGGCCTTGGTCTGAAAAGGTCAATTGCAATCAAACTACA
GAGAGCTTAGAAAGCTCCTCTGAGCAAAAGGTTTTAGAAAGGCTTCAAAGTATTTCATTA
AAAAGTGAAGTAATTTTAACAACCTAAAACAGAGACCACTGGCCTGTCAGCAAGATGCAG
ATTCCAATTCTGCTTCTTACTCTAACAGCCCAGCGGGTCCCTGGGCAGGTCATTCAGCCT
CTTTCAGTTTGTTTTCCCATCTTTAAAAAGGAAATAATCATTGCACCAACCTCACAGGTT
GTGAGGCTCAAGTGCAACAATTTATAAATATTATCATGGTTACTCTCCCATAAAAGAAAG
AACCCATCTTGTGGTCCAAAGACTGGGGCTCCAGTCTCATATTTAGGACCTTTATTTAAT
CAGTCACTTAAGCTCTCTAAATTAATTTTCTCATTTGTAAAAAAAAAAAAAAAAAAGAAAG AP-1
AGATCAACCCTTGTCCAATTCACAGTTGCTGTCTGGAACTAGACAGAGATATAGATGCAT
AGACATGTAGACAAATGAAGTGTTAGTGTAAATGTGTGTAGATGTGCGTAGGCTAGCATT
TACATAATGTTTTAAGCTTTGCAAGACACTTTACAAAGATCATCACAGTTCTTCTCCTCC
CAACAACCATGGAAGGCGGATGCTAGTATTATCCTCACTCTACAGATGAGTAAACTGAGG
CAGACACAGCTTAAGAGATTTGCCCAAGGTGATACAATCAATAAGGGTCTGAGGTTGGAT
TTTAATCCAGATCTTCCTGACTCTGACACTCTAATCGCTGTGCCACCTAAAGGCCCCCTA
CATATGAAAGTGCTTGGTAAACGACAAGGCACTATGTAAATCTCAGATTAGTAGTATTAA
CAGAAAACTGAAAGTAAAGTTTAAAGCAAGTTCTTTCAGAATCTTCCACATATAATGTGT
TTTGAGGATGCAGCTGGAGAAAAATGTAGCAATAAAAACAAAGAAATGTTTAACACTTGG
TTATATGTGACAGGACCACAGCTATATGTAA**AATTGCAT**GAACAAAACCTGGCCTCATTC Oct1
CTATACTACAATCCATCTTCAGTACCTGGAAGGGAGAAAACCCACCTAGCCTGCTCCTCT
GGAGAGGGGATCAGGATAATTCTCCATACCCAGCTTTTCCCTCCCTGGGCTTACAGAATC

BLASTx IL-2

Appendix 6A

```
TCTCACTCTTTTGACTCCAGACCAGCAGCATCGAAAATCTGCAGTGTTTCACTAAAGAAC
TGAACCCTGTGGCTGGTGCGTTGAAATATGAATCCAAAGATGCTCAGAATATTCAGGAAT
ACATCAACAACATTAATGTGACTGTTAATAGCTTAATGGTAAGGAGGATATTATTTTTTT
TTATTTCTTTCAAACTTAATACAAGAGAAAAGAATTATATCCTAAAGACTGCAAGATTAA
ACAATAGTCTAGACTAGTGTTGTGACATAGGATTAAGAGGTATTCACCCTTTTGGTGTTT
GGGGACACACCACCTGTATATTTGACTCTCACAATCATGATTANNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

+3 frame reverse transcription

```
tctcactcttttgactccagaccagcagcatcgaaaatctgcagtgtttcactaaagaactg
   S  L  F  -  L  Q  T  S  S  I  E  N  L  Q  C  F  T  K  E  L
aaccctgtggctggtgcgttgaaatatgaatccaaagatgctcagaatattcaggaatac
 N  P  V  A  G  A  L  K  Y  E  S  K  D  A  Q  N  I  Q  E  Y
atcaacaacattaatgtgactgttaatagcttaatgg
 I  N  N  I  N  V  T  V  N  S  L  M
```

Genomic

>scaffold.Meug_1.0.Scaffold332747.-144.593.1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNAAGGCCTTTGAAAATGTGTATATGTAAAATTCTTT
TATA box → Transcription
AACACCCCCAGGATACTTTTCCAGAATTAAGAA**TATAAAT**TGTCCCTCTGATTTAGAGAC
TTCCATTCCACTCTCTAATCACTACCCAGAGTAACCTGAAGTTACTGAGCTCTTGACACG
Translation
ATGAACAAGGTCCCGCTCTTGTCCTGTATTGCACTAACTCTTGTTCTGGTTGCCAACGGG
GCACCGACATTGCCTCCTCCCACCACTGTGCTGCAGTACTTACTACGTGACTTAATGGAG
GTCCAAAACAAACTCAAGGTAAGTATCTCATTGTTTTTTAAAGACTAAAAATATTGTACT
GAATTCATTCTCATGGGGTGTTCATTTTATTAATGACATTATTAACTTAATTGACTAATT
ACTTTTACTTCCTTAG**ATTGTATCAGAAAGGATGAAGAAGTATGAACTCTACATCCCAAG**
TAATGTAAGTAACATTTTGCATATTGAATTTTGGAGATTTTCAAATTTAAGATCATTAGG
AAGTAGAATTTTTTTGCAGTCTGATGACTTTAGTATCTAAGATTTAAGTCAGGTCTAGGG
GAATGTGTCAAAAAATTT

BLASTx IL-2

Genomic

scaffold.Meug_1.0.Scaffold332747.112.759.1
AATCACTACCCAGAGTAACCTGAAGTTACTGAGCTCTTGACACG**ATGAACAAGGTCCCGC**
TCTTGTCCTGTATTGCACTAACTCTTGTTCTGGTTGCCAACGGGGCACCGACATTGCCTC
CTCCCACCACTGTGCTGCAGTACTTACTACGTGACTTAATGGAGGTCCAAAACAAACTCA
AGGTAAGTATCTCATTGTTTTTTAAAGACTAAAAATATTGTACTGAATTCATTCTCATGG
GGTGTTCATTTTATTAATGACATTATTAACTTAATTGACTAATTACTTTTACTTCCTTAG
ATTGTATCAGAAAGGATGAAGAAGTATGAACTCTACATCCCAAGTAATGTAAGTAACATT
TTGCATATTGAATTTTGGAGATTTTCAAATTTAAGATCATTAGGAAGTAGAATTTTTTTG
CAGTCTGATGACTTTAGTATCTAAGATTTAAGTCAGGTCTAGGGGAATGTGTCAAAAAAT
TTTTCTGTACTTTTTAAACGCAAAACAATCAAATTAGGCTTATATATAGTTGGGTAGATA
GATAACACTTCTCTATACCTTCTCCCCACCCCACATAAATTGAAAAAGAAAATGAGGTAG
AACATCTGCTTTTCTTTTGCAGCAATGTTTTATTTTCTAATAGATAAC

IL-2 BLASTx

1000 bases 5’ and 3’ added

>scaffold.Meug_1.0.Scaffold332747.-588.1459.1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Scaffold332747 dna.scaffold
scaffold.Meug_1.0.Scaffold332747.1.1278.1
AAGGCCTTTGAAAATGTGTATATGTAAAATTCTTTAACACCCCCAGGATACTTTTCCAGA
ATTAAGAATATAAATTGTCCCTCTGATTTAGAGACTTCCATTCCACTCTCTAATCACTAC
CCAGAGTAACCTGAAGTTACTGAGCTCTTGACACGATGAACAAGGTCCCGCTCTTGTCCT
GTATTGCACTAACTCTTGTTCTGGTTGCCAACGGGGCACCGACATTGCCTCCTCCCACCA
CTGTGCTGCAGTACTTACTACGTGACTTAATGGAGGTCCAAAACAAACTCAAGGTAAGTA
TCTCATTGTTTTTTAAAGACTAAAAATATTGTACTGAATTCATTCTCATGGGGTGTTCAT
TTTATTAATGACATTATTAACTTAATTGACTAATTACTTTTACTTCCTTAGATTGTATCA
GAAAGGATGAAGAAGTATGAACTCTACATCCCAAGTAATGTAAGTAACATTTTGCATATT
GAATTTTGGAGATTTTCAAATTTAAGATCATTAGGAAGTAGAATTTTTTTGCAGTCTGAT
GACTTTAGTATCTAAGATTTAAGTCAGGTCTAGGGGAATGTGTCAAAAAATTTTTCTGTA
CTTTTTAAACGCAAAACAATCAAATTAGGCTTATATATAGTTGGGTAGATAGATAACACT
TCTCTATACCTTCTCCCCACCCCACATAAATTGAAAAAGAAAATGAGGTAGAACATCTGC
TTTTCTTTTGCAGCAATGTTTTATTTTCTAATAGATAACTCTGTGTTTTAGATTATATAA
ATGCTTAAAAATGTGTTTTCAAAACTGGAAATCCATATTGATTATAGTACTGTATTGGAC
CCAAGGGTCAAAGGAGATCAGATGTATTAAGTGGGTGGCAAAGTTTAAAGTCCTGGATAA
GCATCAGCTGTTATCATTATTACTCAAGGGTAGTGAGGTGTACTCCACACCCCCTACTGC
TCCCATCCTCCCAACAAGAGTCAGAAACACTGAATTGAATCAGCCTCTGATACACACAAC
TTGAGTAACCCTGGGCCAGTCATAGGTTAGAGAAGTTTGCTGATCTGTATTAGAGGAAGG
GGTTTCCTTCAGTGCAGTGTTACCCACACCAATGAAATGACAGACAGATCTCTACAGATC
TATCTGTGGACATACATATTCATAGATAATCATATCAAATGATGGCAAAGCTTCACAGAG
AACACAGAAACCAGCCAGGTTCTCATTCTGTGTGGAGACAGCAAATAGGGTAGAGACTGG
GCCTGTGATCTCATCCAG

This is IL-2 in red exon/intron boundaries

Proscan. Version 1.7
Processed Sequence. 6754 Base Pairs

```
Promoter region predicted on forward strand in 5978 to 6228
Promoter Score. 58.03 (Promoter Cutoff = 53.000000)
TATA found at 6201, Est.TSS = 6231
Significant Signals.
 Name                   TFD #  Strand  Location  Weight
HNF1                   S01619  -       5978    1.012000
ISGF1                  S01978  -       6125    17.211000


Promoter region predicted on reverse strand in 659 to 409
Promoter Score. 54.47 (Promoter Cutoff = 53.000000)
TATA found at 446, Est.TSS = 414
Significant Signals.
 Name                 Strand  Location  Weight
AP-1                    +       657     1.091000
AP-2                    +       655     1.672000
NF-kB                   +       502     1.008000
NF-kB                   +       501     1.434000
T-Ag                    -       484     1.086000
TFIID                   -       444     2.618000
TFIID                   -       444     1.971000
TFIID                   -       444     2.920000
CTF                     -       409     1.704000
```

Genscan output

```
GENSCAN 1.0     Date run. 29-Mar-111   Time. 22.23.02

Sequence /tmp/03_29_11-22.23.02.fasta . 6754 bp . 42.23% C+G . Isochore 1
( 0 - 43 C+G%)

Parameter matrix. HumanIso.smat

Predicted genes/exons.


Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..

----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------
 1.01 Init +   4181   4282  102  1  0   63   67    89 0.455   4.60

 1.02 Intr +   6432   6569  138  2  0   58   91   172 0.992  14.24
```

Predicted peptide sequence(s).

>/tmp/03_29_11-22.23.02.fasta|GENSCAN_predicted_peptide_1|80_aa

MSDFMPPFEVFLAEILEWFAISSSFTDEETETNKTSSIENLQCFTKELNPVAGALKYESKDAQNIQ
EYINNINVTVNSLM = **IL-2**

>/tmp/03_29_11-22.39.16.fasta|GENSCAN_predicted_CDS_1|240_bp
atgtccgacttcatgcccccttttgaggttttcttggcagagatcctggagtggtttgccatttcc
tcttctttacagacgaggaaactgagaccaacaagaccagcagcatcgaaaatctgcagtgtttc
actaaagaactgaaccctgtggctggtgcgttgaaatatgaatccaaagatgctcagaatattcag
gaatacatcaacaacattaatgtgactgttaatagcttaatg

EXPASY translation of nt sequence

```
atgtccgacttcatgcccccttttgaggttttcttggcagagatcctggagtggtttgcc
 M  S  D  F  M  P  P  F  E  V  F  L  A  E  I  L  E  W  F  A
atttcctcttctttacagacgaggaaactgagaccaacaagaccagcagcatcgaaaat
 I  S  S  S  F  T  D  E  E  T  E  T  N  K  T  S  S  I  E  N
ctgcagtgtttcactaaagaactgaaccctgtggctggtgcgttgaaatatgaatccaaa
 L  Q  C  F  T  K  E  L  N  P  V  A  G  A  L  K  Y  E  S  K
gatgctcagaatattcaggaatacatcaacaacattaatgtgactgttaatagcttaatg
 D  A  Q  N  I  Q  E  Y  I  N  N  I  N  V  T  V  N  S  L  M
```

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAGGCCTTTGAAAATGTGTATATGTAAAATT
CTTTAACACCCCCAGGATACTTTTCCAGAATTAAGAATATAAATTGTCCCTCTGATTTAG
5′ UTR
AGACTTCCATTCCACTCTCTAATCACTACCCAGAGTAACCTGAAGTTACTGAGCTCTTGA
CACGATGAACAAGGTCCCGCTCTTGTCCTGTATTGCACTAACTCTTGTTCTGGTTGCCAA
CGGGGCACCGACATTGCCTCCTCCCACCACTGTGCTGCAGTACTTACTACGTGACTTAAT
GGAGGTCCAAAACAAACTCAAGGTAAGTATCTCATTGTTTTTTAAAGACTAAAAATATTG
TACTGAATTCATTCTCATGGGGTGTTCATTTTATTAATGACATTATTAACTTAATTGACT
AATTACTTTTACTTCCTTAGATTGTATCAGAAAGGATGAAGAAGTATGAACTCTACATCC
CAAGTAATGTAAGTAACATTTTGCATATTGAATTTTGGAGATTTTCAAATTTAAGATCAT
TAGGAAGTAGAATTTTTTTGCAGTCTGATGACTTTAGTATCTAAGATTTAAGTCAGGTCT
AGGGGAATGTGTCAAAAAATTTTTCTGTACTTTTTAAACGCAAAACAATCAAATTAGGCT
TATATATAGTTGGGTAGATAGATAACACTTCTCTATACCTTCTCCCCACCCCACATAAAT
TGAAAAAGAAAATGAGGTAGAACATCTGCTTTTCTTTTGCAGCAATGTTTTATTTTCTAA
TAGATAACTCTGTGTTTTAGATTATATAAATGCTTAAAAATGTGTTTTCAAAACTGGAAA
TCCATATTGATTATAGTACTGTATTGGACCCAAGGGTCAAAGGAGATCAGATGTATTAAG
TGGGTGGCAAAGTTTAAAGTCCTGGATAAGCATCAGCTGTTATCATTATTACTCAAGGGT
AGTGAGGTGTACTCCACACCCCCTACTGCTCCCATCCTCCCAACAAGAGTCAGAAACACT
GAATTGAATCAGCCTCTGATACACACAACTTGAGTAACCCTGGGCCAGTCATAGGTTAGA
GAAGTTTGCTGATCTGTATTAGAGGAAGGGGTTTCCTTCAGTGCAGTGTTACCCACACCA
ATGAAATGACAGACAGATCTCTACAGATCTATCTGTGGACATACATATTCATAGATAATC
ATATCAAATGATGGCAAAGCTTCACAGAGAACACAGAAACCAGCCAGGTTCTCATTCTGT
GTGGAGACAGCAAATAGGGTAGAGACTGGGCCTGTGATCTCATCCAGNNNNNNNNNNNN

GENSCAN 1.0 Date run. 29-Mar-111 Time. 22.45.02

Sequence /tmp/03_29_11-22.45.02.fasta . 1200 bp . 36.38% C+G . Isochore 1 (0 - 43 C+G%)

Parameter matrix. HumanIso.smat

Predicted genes/exons.

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.01	Init	+	125	262	138	1	0	84	115	78	0.991	9.87
1.02	Intr	+	381	428	48	2	0	105	101	46	0.969	5.56

>/tmp/03_29_11-22.45.02.fasta|GENSCAN_predicted_peptide_1|62_aa

MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIPSN = **IL-2**

>/tmp/03_29_11-22.45.02.fasta|GENSCAN_predicted_CDS_1|186_bp

atgaacaaggtcccgctcttgtcctgtattgcactaactcttgttctggttgccaacggggcaccgacattgc
ctcctcccaccactgtgctgcagtacttactacgtgacttaatggaggtccaaaacaaactcaagattgtatc
agaaaggatgaagaagtatgaactctacatcccaagtaat
Promoter region predicted on forward strand in 10183 to 10433
Promoter Score. 55.71 (Promoter Cutoff = 53.000000)

TATA found at 10416, Est.TSS = 10446
Significant Signals.

Name	TFD #	Strand	Location	Weight
Oct-factors	S01029	+	10198	2.754000
IgHC.2	S00814	-	10205	2.754000
INF.1	S01152	+	10240	1.044000
IgNF-A	S00830	+	10307	1.434000
NF-IL2-D	S01626	-	10313	8.606000

Promoter region predicted on forward strand in 8729 to 8979
Promoter Score. 54.65 (Promoter Cutoff = 53.000000)
TATA found at 8930, Est.TSS = 8960
Significant Signals.

Name	TFD #	Strand	Location	Weight
T-Ag	S00974	+	8925	1.086000
TFIID	S00087	+	8931	2.618000
Y	S01848	+	8974	9.680000
CTF	S00780	+	8976	1.704000

Promoter region predicted on reverse strand in 6287 to 6037
Promoter Score. 55.70 (Promoter Cutoff = 53.000000)
TATA found at 6065, Est.TSS = 6033
Significant Signals.

Name	Strand	Location	Weight
TFIID	-	6287	2.618000
TFIID	-	6287	1.971000
TFIID	-	6287	2.920000
NF-kB	-	6094	1.080000
TFIID	-	6063	1.971000
TFIID	-	6063	2.920000
TFIID	-	6063	2.618000

Promoter region predicted on forward strand in 313 to 563
Promoter Score. 90.42 (Promoter Cutoff = 53.000000)

Significant Signals.

Name	TFD #	Strand	Location	Weight
UCE.2	S00437	+	350	1.278000
AP-2	S01936	-	375	1.091000
UCE.1	S00436	-	399	1.700000
PuF	S02016	+	430	1.082000
JCV_repeated_sequenc	S01193	+	430	1.427000
AP-2	S00346	-	439	1.672000
UCE.2	S00437	+	448	1.278000
JCV_repeated_sequenc	S01193	-	491	1.658000
UCE.2	S00437	-	506	1.216000
GCF	S01964	-	519	2.284000
TTR_inverted_repeat	S01112	+	548	2.151000
EIIF	S00659	+	552	50.000000
element_II_rs-4	S01507	-	559	25.816999
E2F	S00147	-	559	50.000000

T. vulpecula IL-2 nucleotide sequence

atgaacacggttccgctcctgtcctgtattgcactaactcttgttctggctgccaatggggcaccaacatcgc
gtcctcccaccactgtgctgcagttcgtactagatgacttaacgttgctcacagagaaactcaagaatgtatc
ggagaggatgaagggatatgaactccacatcccaagtaataccagcagcattgaagctctgcagtgtttcact
aaagaactgaaacctgtggccggtgcgttgaaatatgaatcagaagatgctcagaaaattcaggaagacatca
acaacattaatgtgaatgttaatagattaacgggaccagaaacaacacagtgtcactacgcctccaagaagaa
gattgaagggtttttcacagaatttatttcattctgccaaaaactcatgggtttaactcga**tga**

Putative protein of *T. vulpecula* IL-2

MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIPSNTSSIEALQCFT
KELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASKKKIEGFFTEFISFCQKLMGLTR

Alignment of *M. eugenii* and *T. vulpecula* IL-2 using CLUSTALW (blosum 62 matrix)

```
T.vulpecula     MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
M.eugenii       MNKVPLLSCIALTLVLVANGAPTLPPPTTVLQYLLRDLMEVQNKLKIVSERMKKYELYIP 60
                **.*************.******  *******..* **  . .*** ****** ***.**

T.vulpecula     SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
M.eugenii       SNTSSIENLQCFTKELNPVAGALKYESKDAQNIQEYINNINVTVNSLMGPEITQCHYASK 120
                ******* ********.**********.***.*** ******.** * *** ********

T.vulpecula     KKIEGFFTEFISFCQKLMGLTR 142
M.eugenii       MRIEGFFKEFVSVCQRFMH--- 139
                 .*****.**.*.**..*
```

Figure 6A.3 . Sequence comparison of *T. vulpecula* and *M. eugenii* IL-2

Alignment of *M. eugenii*, *T. vulpecula* and *H. sapiens* using CLUSTALW (blosum 62 matrix)

```
M.eugenii       MNKVPLLSCIALTLVLVANGAPTL---PPPTTVLQYLLRDLMEVQNKLKIVS----ERMK 53
T.vulpecula     MNTVPLLSCIALTLVLAANGAPTS---RPPTTVLQFVLDDLTLLTEKLKNVS----ERMK 53
H.sapiens       MYRMQLLSCIALSLALVTNSAPTSSSTKKTQLQLEHLLLDLQMILNGINNYKNPKLTRML 60
                *  . *******.*.*..*.***      .   *...* **  . . ..  .     **

M.eugenii       KYELYIPSNTSSIENLQCFTKELNPVAGALKYESKD--AQNIQEYINNINVTVNSLMGPE 111
T.vulpecula     GYELHIPSNTSSIEALQCFTKELKPVAGALKYESED--AQKIQEDINNINVNVNRLTGPE 111
H.sapiens       TFKFYMPKKATELKHLQCLEEELKPLEEVLNLAQSKNFHLRPRDLISNINVIVLELKGSE 120
                 .....*....... ***. .**.*.  .*.  ...    . .. *.**** *  * *.*

M.eugenii       IT-QCHYASK-MRIEGFFKEFVSVCQRFMH--- 139
T.vulpecula     TT-QCHYASK-KKIEGFFTEFISFCQKLMGLTR 142
H.sapiens       TTFMCEYADETATIVEFLNRWITFCQSIISTLT 153
                 *  *.**..   *  *.......** ..
```

Figure 6A.4. Sequence comparison of marsupial and human IL-2 sequences

Appendix 6A

Polymorphism in *T. vulpecula* IL-2

```
1CP7F           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP8R            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP4F            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP5R            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP9F            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP10R           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
2CP1F           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
2CP7F           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP3F            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP6R            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP5F            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP8F            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP10F           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP11R           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
2CP1R           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP11F           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP9R            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP7R            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP4R            MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
1CP1F           MNTVPLLSCIALTLVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
CP3R            MNTVPLLSCIALALVLAANGAPTSRPPTTVLQFVLDDLTLLTEKLKNVSERMKGYELHIP 60
1CP1R           MNTVPLLSCIGLALVLAANGAPASRPPTTVLQFVLDDLTLLAEKLKNVSERMKGYELRIP 60
                **********.*.*********.******************.***************.**

1CP7F           SNTSSIEAVQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCQYASK 120
CP8R            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP4F            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP5R            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP9F            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP10R           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
2CP1F           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
2CP7F           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP3F            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP6R            SNTSGIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP5F            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP8F            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP10F           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP11R           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
2CP1R           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP11F           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP9R            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP7R            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP4R            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
1CP1F           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
CP3R            SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
1CP1R           SNTSSIEALQCFTKELKPVAGALKYESEDAQKIQEDINNINVNVNRLTGPETTQCHYASK 120
                ****.***.*********************************************.****

1CP7F           KKIEGFFTEFISLCQKLMGLTR 142
CP8R            KKIEGFFTEFISLCQKLMGLTR 142
CP4F            KKIEGFFTEFISFCQKLMGLTR 142
CP5R            KKIEGFFTEFISFCQKLMGLTR 142
CP9F            KKIEGFFTEFISFCQKLMGLTR 142
CP10R           KKIEGFFTEFISFCQKLMGLTR 142
2CP1F           KKIEGFFTEFISFCQKLMGLTR 142
2CP7F           KKIEGFFTEFISFCQKLMGLTR 142
CP3F            KKIEGFFTEFISFCQKLMGLTR 142
CP6R            KKIEGFFTEFISFCQKLMGLTR 142
CP5F            KKIEGFFTEFISFCQKLMGLTR 142
CP8F            KKIEGFFTEFISFCQKLMGLTR 142
CP10F           KKIEGFFTEFISFCQKLMGLTR 142
CP11R           KKIEGFFTEFISFCQKLMGLTR 142
2CP1R           KKIEGFFTEFISFCQKLMGLTR 142
CP11F           KKIEGFFTEFISFCQKLMGLTR 142
CP9R            KKIEGFFTEFISFCQKLMGLTR 142
CP7R            KKIEGFFTEFISFCQKLMGLTR 142
CP4R            KKIEGFFTEFISFCQKLMGLTR 142
1CP1F           KKIEGFFTEFISFCQKLMGLTR 142
```

```
CP3R            KKIEGFFTEFISFCQKLMGLTR 142
1CP1R           KKIEGFFTEFISFCQKLMGLTR 142
                ************.*********
```

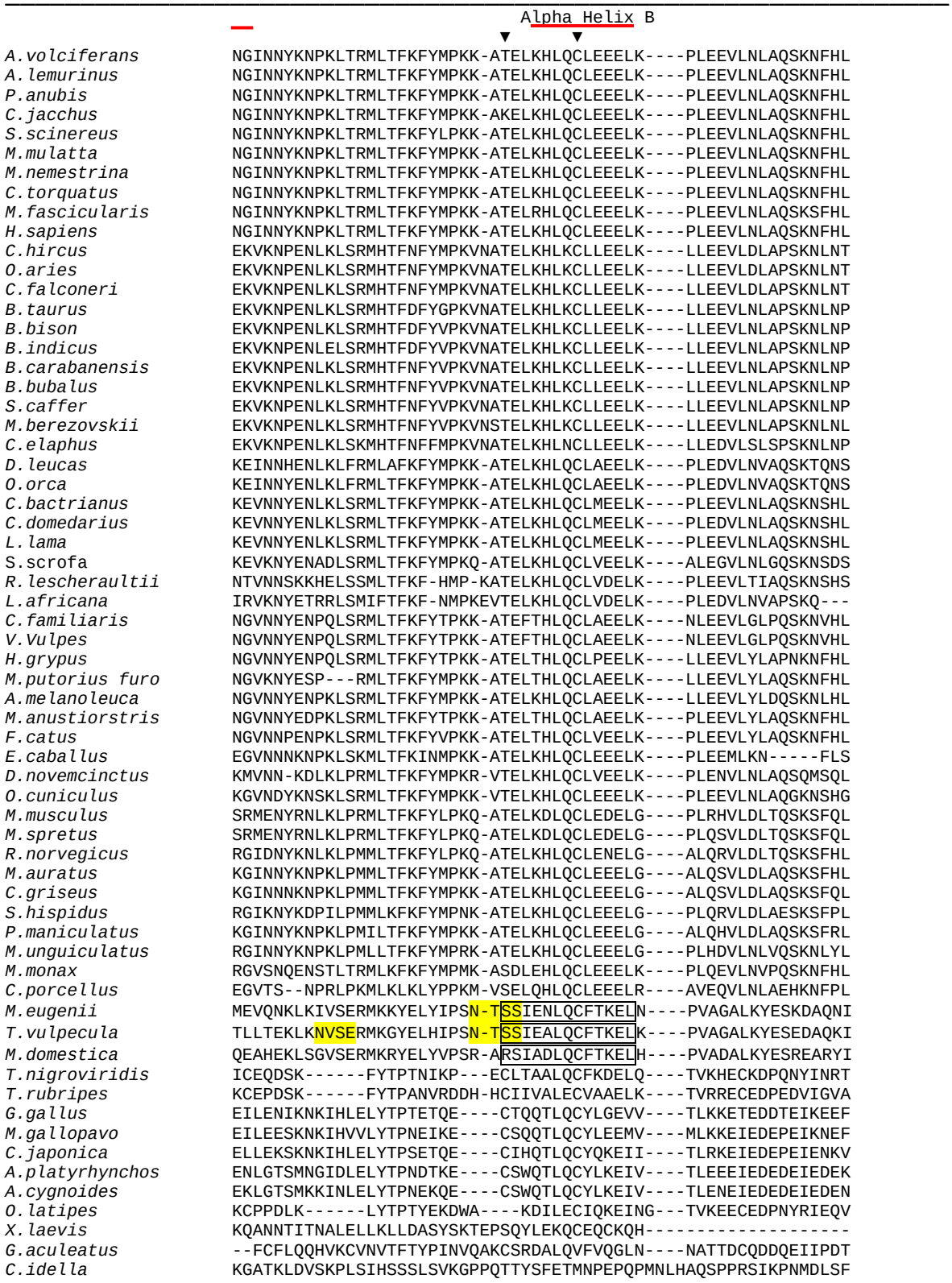
Figure 6A.5. Sequence of 10 clones in forward and reverse direction using CLUSTALW. The differences in sequence are sequence variations and not substitutions.

Appendix 6A

Multiple amino acid alignment using CLUSTALW (blosum 62 matrix)

```
                                                                                                  Alpha Helix A
                                                                                                    ■   ■
A.volciferans          -MYRMQLLSCIALSLALITNSAP-TSSSTK-------------KTQLQLEHLLLDLQMLL
A.lemurinus            -MYRMQLLSCIALSLALITNSAP-TSSSTK-------------KTQLQLEHLLLDLQMLL
P.anubis               -MYRMQLLSCIALSLALITNSAP-TSSSTK-------------KTQLQLEHLLLDLQMLL
C.jacchus              -MYRMQLLSCIALSLALITNSAP-TSSSTK-------------KTQLQLEHLLLDLQMLL
S.scinereus            -MYRMQLLSCIALSLALITNSAP-TSSSTK-------------KTQLQLEHLLLDLQMLL
M.mulatta              -MYRMQLLSCIALSLALVTNSAP-TSSSTK-------------KTQLQLEHLLLDLQMIL
M.nemestrina           -MYRMQLLSCIALSLALVTNSAP-TSSSTK-------------KTQLQLEHLLLDLQMIL
C.torquatus            -MYRMQLLSCIALSLALVTNSAP-TSRSTK-------------KTQLQLEHLLLDLQMIL
M.fascicularis         -MYRMQLLSCIALSLALVTNSAP-TSSSTK-------------KTQLQLEHLLLDLQMIL
H.sapiens              -MYRMQLLSCIALSLALVTNSAP-TSSSTK-------------KTQLQLEHLLLDLQMIL
C.hircus               -MYQIPLLSCIALTLALVANGAT-TSSSTG-------------NPMKEVKSLLLDLQLLL
O.aries                -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
C.falconeri            -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
B.taurus               -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
B.bison                -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
B.indicus              -MYRIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
B.carabanensis         -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
B.bubalus              -MYKIQLLSCIVLTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
S.caffer               -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
M.berezovskii          -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
C.elaphus              -MYKIQLLSCIALTLALVANGAP-TSSSTG-------------NTMKEVKSLLLDLQLLL
D.leucas               -MYKMQLLSCIALTLALVANGAP-TSSSTE-------------NTKKQVQSLLQDLHLLL
O.orca                 -MYRMQLLSCIALTLALVANGAP-TSSSTE-------------NTKKQVQSLLQDLQLLL
C.bactrianus           -MYKLQFLSCIALTLALVANSAP-TLSSTK-------------DTKKQLEPLLLDLQFLL
C.domedarius           -MYKLQFLSCIALTLALVANSAP-TLSSTK-------------DTKKQLEPLLLDLQFLL
L.lama                 -MYKLQLLSCIALTLALVANSAP-TLSSTK-------------DTKKQLEPLLLDLQFLL
S.scrofa               -MYKMQLLCCIALTLALMANGAP-TSSSTK-------------NTKKQLEPLLLDLQLLL
R.lescheraultii        -MHKMYFLSCIALTLALVADGAP-TSSSRK-------------ETQQQLEHLLKDLQLLL
L.africana             -MFKMQLLSCIALTLALVANSAP-TSSSTK-------------ETQQQLEQLLLDLQMLL
C.familiaris           -MYKMQLLSCIALTLVLVANSAPITSSSTK-------------ETEQQMEQLLLDLQLLL
V.vulpes               -MYKMQLLSCIALMLVLVANSAPITSSSTK-------------ETEQQMEQLLLDLQLLL
H.grypus               -MYKMQLLSCIALTLVLVANSAPTTSSSTK-------------ETQQQLEQLLLDLRLLL
M.putorius furo        -MYKMQLLSCIVLTLALFANSAPTTSSSTK-------------EAQQQLEQLLLDLQLLL
A.melanoleuca          -MYKMQLLYCIALTLVLVANSAPTPSSPTK-------------ETQQQLEQLMLDLRLLL
M.anustiorstris        -MCKMQLLSCIALSLVLVANSAPTTSS-TK-------------ETQQQLEQLLLDLRLLL
F.catus                -MYKIQLLSCIALTLILVTNSAP-ASSSTK-------------ETQQQLEQLLLDLRLLL
R.norvegicus           -MYSMQLASCVALTLVLLVNSAPTSSPAK--------------ETQQHLEQLLLDLQVLL
M.auratus              -MYSMQLASCLALTLALLVSSAPTSSSKK--------------ETQQHLEQLLLDLQELL
C.griseus              -MYTMQLASCLALTLALLVNSAPTSSSKK--------------ETQQHLEQLLLDLQVLL
S.hispidus             -MYNMQLASCVALMLALLVNSAPTSSSTK--------------ETQQHLEQLLLDLQVLL
P.maniculatus          -MYSRQLASCVALTLVLLVNSAPTSSSTK--------------ETQQHLEQLLMDLEVLL
M.unguiculatus         -MYSRQLASCVALALVLLANSAPTSSPAK--------------EAQQYLEQLLLDLQQLL
M.monax                -MHTMPLLSCLALTLALVAHGAPTSGSAEET--------------RQQLEQLLLDLQMLS
C.porcellus            -MYKTLLLSCLALTLALLTSSAPTSSSPKQ--------------TQDRLELLLRDLQTLL
M.eugenii              -MNKVPLLSCIALTLVLVANGAPTLPPPTT---------------------VLQYLLRDL
T.vulpecula            -MNTVPLLSCIALTLVLAANGAPTSRPPTT---------------------VLQFVLDDL
M.domestica            -MSKVPLLLCVALTLAVLAGGAPTSPPPTS---------------------LLEYLLLDL
T.nigroviridis         MENFIRINVWLGILCLCFPANPFPLHLED----------------------SNIDVIREDV
E.caballus             -MYKMQLLACIALTLAVLANSAP-TSSSKR-------------ETQQQLKQLQMDLKLLL
D.novemcinctus         -MYKMQLVACIALSLVLITNSAP-TSSSTK-------------ETQQQLEQLLLDLKMLS
O.cuniculus            -MYKVQLLSCIALTLALLTSSAP-TSSSTK-------------ETQEQLDQLLLDLQVLL
M.musculus             -MYSMQLASCVTLTLVLLVNSAPTSSSTSSSTAEAQQQQQQQQQQQQHLEQLLMDLQELL
M.spretus              -------------------MAPTSSSTSSSTAEAQ-------QQQQHLEQLLMDLQELL
R.norvegicus           METFNRIYFGMVIVCVCLPANSNPMPLLDD---------------------SDIGDMKKNV
T.rubripes             MENFIRINVWLGILCLCFPANPFPLHLED----------------------SNIDVIREDV
G.gallus               MMCKVLIFGCISVAMLMTTAYGASLSSEKW---------------------KTLQTLIKDL
M.gallopavo            MMCKVLIFSCISVALLMTTAYGASLSPEKL---------------------EILPALIKDL
C.japonica             -MCKVLIFACISVAMLMTTAYGATLPPKEQ---------------------DILPALISDL
A.platyrhynchos        -MCKVLIFSCLSVLMLMTTAYGAPLS-EKD---------------------NTLKTLIKDL
A.cygnoides            -MCKVLIFSCLSVVMLMTTAYGAPLS-EKN---------------------DTLTTLIKDL
O.latipes              MEHLFKIAIWIFVLSGCHLTSSKCIPTDDD---------------------WVLDALQEEV
X.laevis               MKCRISALCCFVHELYTVEKYLGTLNEKAG-----------------------------
G.aculeatus            MEHSLRTALWVFCLFGFLQATPPCYGQGDLG-----------------------------
C.idella               -MYSMQLQKQEAKKANGTRNGSKGADWTIE---------------PNSVTDIEKLDIGSG
```

Appendix 6A


Alpha Helix B
A.volciferans NGINNYKNPKLTRMLTFKFYMPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
A.lemurinus NGINNYKNPKLTRMLTFKFYMPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
P.anubis NGINNYKNPKLTRMLTFKFYMPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
C.jacchus NGINNYKNPKLTRMLTFKFYMPKK-AKELKHLQCLEEELK----PLEEVLNLAQSKNFHL
S.scinereus NGINNYKNPKLTRMLTFKFYLPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
M.mulatta NGINNYKNPKLTRMLTFKFYMPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
M.nemestrina NGINNYKNPKLTRMLTFKFYMPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
C.torquatus NGINNYKNPKLTRMLTFKFYMPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
M.fascicularis NGINNYKNPKLTRMLTFKFYMPKK-ATELRHLQCLEEELK----PLEEVLNLAQSKSFHL
H.sapiens NGINNYKNPKLTRMLTFKFYMPKK-ATELKHLQCLEEELK----PLEEVLNLAQSKNFHL
C.hircus EKVKNPENLKLSRMHTFNFYMPKVNATELKHLKCLLEELK----LLEEVLDLAPSKNLNT
O.aries EKVKNPENLKLSRMHTFNFYMPKVNATELKHLKCLLEELK----LLEEVLDLAPSKNLNT
C.falconeri EKVKNPENLKLSRMHTFNFYMPKVNATELKHLKCLLEELK----LLEEVLDLAPSKNLNT
B.taurus EKVKNPENLKLSRMHTFDFYGPKVNATELKHLKCLLEELK----LLEEVLNLAPSKNLNP
B.bison EKVKNPENLKLSRMHTFDFYVPKVNATELKHLKCLLEELK----LLEEVLNLAPSKNLNP
B.indicus EKVKNPENLELSRMHTFDFYVPKVNATELKHLKCLLEELK----LLEEVLNLAPSKNLNP
B.carabanensis EKVKNPENLKLSRMHTFNFYVPKVNATELKHLKCLLEELK----LLEEVLNLAPSKNLNP
B.bubalus EKVKNPENLKLSRMHTFNFYVPKVNATELKHLKCLLEELK----LLEEVLNLAPSKNLNP
S.caffer EKVKNPENLKLSRMHTFNFYVPKVNATELKHLKCLLEELK----LLEEVLNLAPSKNLNP
M.berezovskii EKVKNPENLKLSRMHTFNFYVPKVNSTELKHLKCLLEELK----LLEEVLNLAPSKNLNL
C.elaphus EKVKNPENLKLSKMHTFNFFMPKVNATELKHLNCLLEELK----LLEDVLSLSPSKNLNP
D.leucas KEINNHENLKLFRMLAFKFYMPKK-ATELKHLQCLAEELK----PLEDVLNVAQSKTQNS
O.orca KEINNYENLKLFRMLTFKFYMPKK-ATELKHLQCLAEELK----PLEDVLNVAQSKTQNS
C.bactrianus KEVNNYENLKLSRMLTFKFYMPKK-ATELKHLQCLMEELK----PLEEVLNLAQSKNSHL
C.domedarius KEVNNYENLKLSRMLTFKFYMPKK-ATELKHLQCLMEELK----PLEDVLNLAQSKNSHL
L.lama KEVNNYENLKLSRMLTFKFYMPKK-ATELKHLQCLMEELK----PLEEVLNLAQSKNSHL
S.scrofa KEVKNYENADLSRMLTFKFYMPKQ-ATELKHLQCLVEELK----ALEGVLNLGQSKNSDS
R.lescheraultii NTVNNSKKHELSSMLTFKF-HMP-KATELKHLQCLVDELK----PLEEVLTIAQSKNSHS
L.africana IRVKNYETRRLSMIFTFKF-NMPKEVTELKHLQCLVDELK----PLEDVLNVAPSKQ---
C.familiaris NGVNNYENPQLSRMLTFKFYTPKK-ATEFTHLQCLAEELK----NLEEVLGLPQSKNVHL
V.Vulpes NGVNNYENPQLSRMLTFKFYTPKK-ATEFTHLQCLAEELK----NLEEVLGLPQSKNVHL
H.grypus NGVNNYENPQLSRMLTFKFYTPKK-ATELTHLQCLPEELK----LLEEVLYLAPNKNFHL
M.putorius furo NGVKNYESP---RMLTFKFYMPKK-ATELTHLQCLAEELK----LLEEVLYLAQSKNFHL
A.melanoleuca NGVNNYENPKLSRMLTFKFYMPKK-ATELKHLQCLAEELK----LLEEVLYLDQSKNLHL
M.anustiorstris NGVNNYEDPKLSRMLTFKFYTPKK-ATELTHLQCLAEELK----PLEEVLYLAQSKNFHL
F.catus NGVNNPENPKLSRMLTFKFYVPKK-ATELTHLQCLVEELK----PLEEVLYLAQSKNFHL
E.caballus EGVNNNKNPKLSKMLTFKINMPKK-ATELKHLQCLEEELK----PLEEMLKN-----FLS
D.novemcinctus KMVNN-KDLKLPRMLTFKFYMPKR-VTELKHLQCLVEELK----PLENVLNLAQSQMSQL
O.cuniculus KGVNDYKNSKLSRMLTFKFYMPKK-VTELKHLQCLEEELK----PLEEVLNLAQGKNSHG
M.musculus SRMENYRNLKLPRMLTFKFYLPKQ-ATELKDLQCLEDELG----PLRHVLDLTQSKSFQL
M.spretus SRMENYRNLKLPRMLTFKFYLPKQ-ATELKDLQCLEDELG----PLQSVLDLTQSKSFQL
R.norvegicus RGIDNYKNLKLPMMLTFKFYLPKQ-ATELKHLQCLENELG----ALQRVLDLTQSKSFHL
M.auratus KGINNYKNPKLPMMLTFKFYMPKK-ATELKHLQCLEEELG----ALQSVLDLAQSKSFHL
C.griseus KGINNNKNPKLPMMLTFKFYMPKK-ATELKHLQCLEEELG----ALQSVLDLAQSKSFQL
S.hispidus RGIKNYKDPILPMMLKFKFYMPNK-ATELKHLQCLEEELG----PLQRVLDLAESKSFPL
P.maniculatus KGINNYKNPKLPMILTFKFYMPKK-ATELKHLQCLEEELG----ALQHVLDLAQSKSFRL
M.unguiculatus RGINNYKNPKLPMLLTFKFYMPRK-ATELKHLQCLEEELG----PLHDVLNLVQSKNLYL
M.monax RGVSNQENSTLTRMLKFKFYMPMK-ASDLEHLQCLEEELK----PLQEVLNVPQSKNFHL
C.porcellus EGVTS--NPRLPKMLKLKLYPPKM-VSELQHLQCLEEELR----AVEQVLNLAEHKNFPL
M.eugenii MEVQNKLKIVSERMKKYELYIPSN-TSSIENLQCFTKELN----PVAGALKYESKDAQNI
T.vulpecula TLLTEKLKNVSERMKGYELHIPSN-TSSIEALQCFTKELK----PVAGALKYESEDAQKI
M.domestica QEAHEKLSGVSERMKRYELYVPSR-ARSIADLQCFTKELH----PVADALKYESREARYI
T.nigroviridis ICEQDSK------FYTPTNIKP---ECLTAALQCFKDELQ----TVKHECKDPQNYINRT
T.rubripes KCEPDSK------FYTPANVRDDH-HCIIVALECVAAELK----TVRRECEDPEDVIGVA
G.gallus EILENIKNKIHLELYTPTETQE----CTQQTLQCYLGEVV----TLKKETEDDTEIKEEF
M.gallopavo EILEESKNKIHVVLYTPNEIKE----CSQQTLQCYLEEMV----MLKKEIEDEPEIKNEF
C.japonica ELLEKSKNKIHLELYTPSETQE----CIHQTLQCYQKEII----TLRKEIEDEPEIENKV
A.platyrhynchos ENLGTSMNGIDLELYTPNDTKE----CSWQTLQCYLKEIV----TLEEEIEDEDEIEDEK
A.cygnoides EKLGTSMKKINLELYTPNEKQE----CSWQTLQCYLKEIV----TLENEIEDEDEIEDEN
O.latipes KCPPDLK------LYTPTYEKDWA----KDILECIQKEING---TVKEECEDPNYRIEQV
X.laevis KQANNTITNALELLKLLDASYSKTEPSQYLEKQCEQCKQH----------
G.aculeatus --FCFLQQHVKCVNVTFTYPINVQAKCSRDALQVFVQGLN----NATTDCQDDQEIIPDT
C.idella KGATKLDVSKPLSIHSSSLSVKGPPQTTYSFETMNPEPQPMNLHAQSPPRSIKPNMDLSF

Appendix 6A

```
                          Alpha Helix C                                 Alpha Helix D
                                        ▼ ▼                                   ■
A.volciferans        RDT-------RDIISNINVLVLELKGSETTFTCEYDDDTATIIEFLNGWITFCQSIISTL
A.lemurinus          RDT-------RDIISNINVLVLELKGSETTFTCEYDDDTATIIEFLNGWITFCQSIISTL
P.anubis             RDT-------RDIISNINVLVLELKGSETTFTCEYDDDTATIIEFLNGWITFCQSIISTL
C.jacchus            RDT-------RDIISNINVLVLELKGSETTFTCEYDDDTATIIEFLNGWITFCQSIISTL
S.scinereus          RDT-------RDIISNINVLVLELKGSETTFTCEYDDDTATIIEFLNGWITFCQSIISTL
M.mulatta            RDT-------KDLISNINVIVLELKGSETTLMCEYADETATIVEFLNRWITFCQSIISTL
M.nemestrina         RDT-------KDLISNINVIVLELKGSETTLMCEYADETATIVEFLNRWITFCQSIISTL
C.torquatus          RDT-------KDLISNINVIVLELKGSETTLMCEYADETATIVEFLNRWITFCQSIISTL
M.fascicularis       RDT-------KDLISNINVIVLELKGSETTLMCEYADETATIVEFLNRWITFCQSIISTL
H.sapiens            RP--------RDLISNINVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTL
C.hircus             R-------EIKDSMDNIKRIVLELQGSETRFTCEYDDATVKAVEFLNKWITFCQSIYSTM
O.aries              R-------EIKDSMDNIKRIVLELQGSETRFTCEYDDATVKAVEFLNKWITFCQSIYSTM
C.falconeri          --------RIKDSMDNIKRIVLELQGSETRFTCEYDDATVKAVEFLNKWITFCQSIYSTM
B.taurus             R-------EIKDSMDNIKRIVLELQGSETRFTCEYDDATVNAVEFLNKWITFCQSIYSTM
B.bison              R-------EIKDSMDNIKRIVLELQGSETRFTCEYDDATVNAVEFLNKWITFCQSIYSTM
B.indicus            R-------EIKDSMDNIKRIVLELQGSETRFTCEYDDATVNAVEFLNKWITFCQSIYSTM
B.carabanensis       R-------EIKDSMDNIKRIVLELQGSETGFTCEYDDATVKAVEFLNKWITFCQSIYSTM
B.bubalus            R-------EIKDSMDNIKRIVLELQGSETGFTCEYDDATVKAVEFLNKWITFCQSIYSTM
S.caffer             R-------EIKDSMDNIKRIVLELQGSETRFTCEYDDATVKAVEFLNKWITFCQSIYSTM
M.berezovskii        R-------EIKDSMDNIKRIVLELQGSETTFTCEYDNATVKAAEFLNKWITFCQSIYSTM
C.elaphus            KEIKDSMDEIKDLMDNIKRIVLELQGSETSFKCEYDAATVKAVEFLNKWITFCQRIYSTM
D.leucas             ID-------IKDLMDNINRIVLTLKGSETRFTCEYDDETVTAVEFLNKWITFCQSIYSTM
O.orca               ID-------IKDLMDNINRIVLTLKGSETRFTCEYDEKTVTAVELLNKWITFCQSIIST-
C.bactrianus         TN-------IKDSMNNINLTVSELKGSETGFTCEYDDETVTVVEFLNKWITFCQSIYSTL
C.domedarius         TN-------IKDSMNNINLTVSELKGSETGFTCEYDDETVTVVEFLNKWITFCQSIYSTL
L.lama               TN-------IKDSMNNINLTVSELKGSETGFTCEYDDETVTVVEFLNKWITFCQSIYSTM
S.scrofa             AN-------IKESMNNINVTVLELKGSETSFKCEYDDETVTAVEFLNKWITFCQSIYSTL
R.lescheraultii      --------KIKELVSNINVTAQKLKGPETKSTCDYDDDRVTVREFLNNWITFCQSIFSTL
L.africana           --------NTRELISNINVTALELQGSETTFMCEYDDHAATIEEFLNKWIVFCQSIISTL
C.familiaris         TD-------TKELISNMNVTLLKLKGSETSYNCEYDDETATITEFLNKWITFCQSIFSTL
V.Vulpes             TD-------TKELISNMNVTLLKLKGSETSYNCEYDDETATITEFLNKWITFCQSIFSTL
H.grypus             TD-------IKELMSNINVTLLKLKGSETRFKCEYDDETATITEFLNKWITFCQSIFSTL
M.putorius furo      TD-------IKELMSNINVTLLKLKGSETSFKCEYDDETVTITEFLNKWITFCQSIFSTL
A.melanoleuca        TD-------IKELMSNINVTLLKLKGSEASFKCEYDDETVTITEFLNKWITFCQSIFSTL
M.anustiorstris      TD-------IKELMSNINVTLLKLKGSETRFKCEYDDETATITEFLNKWITFCQSIFSTL
F.catus              NH-------IKELMSNINVTVLKLKGSETRFTCNYDDETATIVEFLNKWITFCQSIFSTL
E.caballus           KD-------IKELMSNINVTVLGLKGSETRFTCEYDDETGTIVEFLNKWITFCQSIFSTM
D.novemcinctus       EHN------GDLISNINITVLELKGSETTFMCDYDDEAATIVEFLNKWIIFCQSIISKR
O.cuniculus          GN-------TRESISNINVTVLKLKGSET-FMCEYD-ETVTIVEFLNRWITFCQSIISAS
M.musculus           EDA-------ENFISNIRVTVVKLKGSDNTFECQFDDESATVVDFLRRWIAFCQSIISTS
M.spretus            EDA-------ENFISNIRVTVVKLKGSDNTIECQFDDESATVVDFLRRWIAFCQSIISTS
R.norvegicus         EDA-------GNFISNIRVTVVKLKGSENKFECQFDDEPATVVEFLRRWIAICQSIISTM
M.auratus            EDA-------ENFISNIRVTVVKLKGSESTFRCEFDDETVTVVEFLNRWITFCQSSIATM
C.griseus            EDT-------ENFISNIRVTVVKLKGSENTFTCEFDDETVTVVEFLSRWITFCQSIIATM
S.hispidus           EDA-------ENSISNIRVIVVKLKGSENPSMCEFDDERVTVVEFLSRWITFCQSTIAAM
P.maniculatus        EDA-------ENSISNIRVTVAKLKGSENTSTCEFHDETVTVVEFLNTWIAFCQSAIST-
M.unguiculatus       EDA-------GNFISNIRVTVMKLKGSENTLNCEFDDETVTVVEFLSRWITFCQSAISTM
M.monax              KDT-------RNFISNINVTVLKLKGSATTFTCEYAPETANIVEFLNTWITFCQSIISKL
C.porcellus          IHT-------KDFISNINVTVLSLKGSETAFVCDLEDESVNIVEFLKRWTAFCQKIMSRL
M.eugenii            QEY----------INNINVTVNSLMGPEITQCHYAS--KMRIEGFFKEFVSVCQRFMH--
T.vulpecula          QED----------INNINVNVNRLTGPETTQCHYAS--KKKIEGFFTEFISFCQKLMGLT
M.domestica          QDH----------IRNINVTVNRLMGTATTHCQYAV--KIKIRGFFGEWITFCQRLIHLT
T.nigroviridis       KGF------LEHVISTMKN----EEVNSNACS-CESYSEEPFPEFLNAMETLVQRFNSKA
T.rubripes           EEF------LTHTIQKLKNGVKIEKSNSTECSTCESWPEKPLTNFLDATESLLQQVQSGA
G.gallus             VTA------IQNIEKNLKSLTGLNHTG-SECKICEANNKKKFPDFLHELTNFVRYLQK--
M.gallopavo          KNA------LQNIKKNLHRLKDLSPTG-GECKICEANDKKNFPDFLQQLTNFVRYLQK--
C.japonica           RTA------LQHIENNLDTLMEQTPIG-GECKICEATTKKNFPDFHQKLTDFLRSMHK--
A.platyrhynchos      VSS------VRNIKMNLQKLMDLIPPG-TGCNICEAN-ANNFPEFRQELTNFLRSMLK--
A.cygnoides          VFS------VRNIEKNLQKLTDLIPPG-TGCKICEANDKKEFPEFHRELTNFLRSMLK--
O.latipes            ISMLKN--VSPDNGTGQNSTNSTCEGSPVKSFQEFVTSVKVILQKIRSGKCLTQNEEKQK
X.laevis             ---------------------------------KAHTTTKNLEGFIEDFEALLKNLATNL
G.aculeatus          LES-----------------LAWKFPTTDSTNCKLQTKESQFEDFVKDLERLVQLINASG
C.idella             PNLSNHTIAPKPRPGKAGAAAAASVVTGPQRFLESLKTPFKLQLSDEPGDPMLRQVIIST
                                                                            . .

A.volciferans        T------
A.lemurinus          T------
P.anubis             T------
C.jacchus            T------
```

```
S.scinereus             T------
M.mulatta               T------
M.nemestrina            T------
C.torquatus             T------
M.fascicularis          T------
H.sapiens               T------
C.hircus                T------
O.aries                 T------
C.falconeri             T------
B.taurus                T------
B.bison                 T------
B.indicus               T------
B.carabanensis          T------
B.bubalus               T------
S.caffer                T------
M.berezovskii           T------
C.elaphus               T------
D.leucas                T------
O.orca                  -------
C.bactrianus            T------
C.domedarius            T------
L.lama                  T------
S.scrofa                T------
R.lescheraultii         P------
L.africana              I------
C.familiaris            T------
V.Vulpes                T------
H.grypus                T------
M.putorius furo         T------
A.melanoleuca           T------
M.anustiorstris         T------
F.catus                 T------
E.caballus              T------
D.novemcinctus          LDN----
O.cuniculus             SS-----
M.musculus              PQ-----
M.spretus               PQ-----
R.norvegicus            TQ-----
M.auratus               TQ-----
C.griseus               TQ-----
S.hispidus              TQ-----
P.maniculatus           -------
M.unguiculatus          TQ-----
M.monax                 T------
C.porcellus             T------
M.eugenii               -------
T.vulpecula             R------
M.domestica             R------
T.nigroviridis          RQNQQR-
T.rubripes              IPSAEGS
G.gallus                -------
M.gallopavo             -------
C.japonica              -------
A.platyrhynchos         -------
A.cygnoides             -------
O.latipes               STIRNK-
X.laevis                -------
G.aculeatus             DK-----
C.idella                -------
```

Figure 6A.6. Alignment of known IL-2 molecules. Boxed the family signature of IL-2. Yellw highlight = Putative N-linked glycosylation sites marked in yellow. ► = disulphide bridge.■ = biologically Important amino acids.

Appendix 6A

Table 6A.1. Genbank Accession numbers for IL-2 and the relevant references.

Species name	Common name	Accession number	Reference
Ailuropoda melanoleuca	Giant Panda	DQ852339	(Li *et al.,* 2010)
Anas platyrhynchos	Duck	AY707747	(Sreekumar *et al.,* 2005)
Anser cygnoides	Chinese goose	AY392557	(Sreekumar *et al.,* 2005)
Aotus lemurinus	Northern grey-necked night monkey	U88364	Direct submission
Aotus vociferans	Spix's owl monkey	U88362	Direct submission
Bison bison	Bison	EF118563.1	(Freeman *et al.,* 2008)
Bos taurus	Cattle	EF056472.1, NM_180997 AF348423	(Reeves *et al.,* 2008)
Bos indicus	Zebu	EF569671	(Prakash *et al.,* 2011)
Bubalus bubalis	Water buffalo	EF118562, EF407852	(Freeman *et al.,* 2008)
Bubalus carabanensis	Carabao	AB246271	(Mingala *et al.,* 2006)
Callithrix jacchus	White-tuffed-ear marmoset	XM_002745432	Annotated
Camelus bactrianus	Bactrian camel	AB246671	(Odbileg *et al.,* 2006)
Camelus dromedarius	Arabian camel	HM051105	Direct submission
Canis lupus familiaris	Dog	NM_001003305	(Chamizo *et al.,* 2001)
Capra falconeri	Markhor	EF056471	Direct submission
Capra hircus	Goat	EF375707, U34274	Direct submission
Cavia porcellus	Domestic Guinea Pig	AB010093	Direct submission
Cercocebus torquatus	Red-crowned mangaby	U19846	(Villinger *et al.,* 1995)
Cervus elaphus	Red Deer	CEU14682	Direct submision
Cricetulus griseus	Chinese Hamster	XM_003506764	Annotated
Coturnix japonica	Japanese quail	AY707748	(Sreekumar *et al.,* 2005)
Ctenopharyngodon idella	Grass carp	AF486820	Direct submission
Dasypus novemcinctus	Nine-banded armadillo	DQ092925	Direct submission
Delphinapterus leucas	Beluga whale	AF072870	(St-Laurent *et al.,* 1999)
Equus caballus	Horse	EU438768, NM_001085433	(Tavernor *et al.,* 1993)
Felis catus	Cat	L19402	(Cozzi *et al.,* 1993)
Gallus gallus	Chicken	NM_204153	(Sundick and Gill-Dixon, 1997)
Gasterosteus aculeatus	Three-spined stickleback	EF513157	Direct submission
Halichoerus grypus	Gray seal	AF072871	(St-Laurent *et al.,* 1999)
Homo sapiens	Human	NM_000586	(Bazan, 1992)
Llama glama	Lama	AB107651	(Odbileg *et al.,* 2004)
Loxodonta africana	African elephant	XM_003410509	Direct submission
Macaca fascicularis	Crab-eating macaque	D63352	(Yabe *et al.,* 1997)
Macaca mulatta	Rhesus monkey	NM_001047130	(Villinger *et al.,* 1995)
Macaca nemestrina	Pig-tailed macaque	U19852	(Villinger *et al.,* 1995)
Marmota monax	Woodchuck	DQ272238	(Gujar and Michalak, 2006)
Meleagris gallopavo	Turkey	AJ007463	(Lawson *et al.,* 2000)
Meriones unguiculatus	Mongolian gerbil	X68779	(Mai *et al.,* 1994)
Mesocricetus auratus	Golden hamster	EU729351	Direct submission
Mirounga anustiorstris	Northern elephant seal	U79187	(Shoda *et al.,* 1998)

Monodelphis domestica	South-American opossum	XM_0033413731	Annotated
Moschus berezovskii	Chinese forest musk deer	AY840980	Direct submission
Mus musculus	House mouse	NM_008366	(Matesanz *et al.,* 1992)
Mus spretus	Western wild mouse	U41495	(Matesanz *et al.,* 1993)
Mustela putorius furo	Domestic ferret	EF368206	Direct submission
Orcinus orca	Ocra	AF009570	(Ness *et al.,* 1997)
Oryzias latipes	Japanese medaka	NM_001128514	Direct submission
Oryctolagus cuniculus	Rabbit	AF169168, AF068057.1 DQ852342	(Perkins *et al.,* 2000)
Ovis aries	Sheep	EF118564.1, AF287479	(Freeman *et al.,* 2008)
Papio anubis	Olive baboon	NM_001112654, AY234220	Direct submission
Papio hamadryas	Hamadras baboon	U88365	Direct submission
Peromyscus maniculatus	Deer mouse	AY247760	(Schountz *et al.,* 2004)
Rattus norvegicus	Norway Rat	NM_053836	(McKnight *et al.,* 1989)
Rousettus lescheraultii	Leschenault's rousette	AB472358	(Iha *et al.,* 2009)
Saimiri sciureus	Common squirrel monkey	AF294755	(Heraud *et al.,* 2002)
Sigmodon hispidus	Cotton rat	AF398549	Direct submission
Sus scrofa	Pig	NM_213861	(Li *et al.,* 2005)
Syncerus caffer	African buffalo	AB571123.1	Direct submission
Takifugu rubripes	Tiger pufferfish	NM_001037994	(Bird et al., 2005a)
Tetraodon nigroviridis	Spotted green pufferfish	EF513163	Direct submission
Trichosurus vulpecula	Silver brushtail posssum	HQ717721.1	(Young *et al.,* 2011)
Vulpes vulpes	Red fox	AJ621188	(Rolland-Turner *et al.,* 2006)
Xenous tropicalis	Western clawed frog	EF513165	Direct submission

Appendix 6B

M. eugenii Interluekin-17 (IL_17) nucleotide sequence

atgtcttctctgggcaacttgccagggtttaagtcactgctgctgttgctggttctggcggtcatgatgaaga
caggagtctcaatgccaaagaggtcaggctgccccaaagctgaaaagaacgactcttctcagagagttagcat
caacatgaatatcattaaccgaaaccaaggctcaaagatatcccctgattacaagaaccgctccacctctcct
tgggatatgtttcctaacgaggatgccaacagactgccccgcaccatctgggaggctaagtgccgtcactcag
gctgcatcaacgccgaggggaaagtggaccaccacctgaactctgtcgccatccagcaggagatcttggtcct
ccgcagagagttcccaaactgttccacttccttcaggctggagaagatgctggtgactgtgggctgcacttgt
gtgaccccacgcacagtgtcctaa

M. eugenii putative protein sequence

MSSLGNLPGFKSLLLLLVLAVMMKTGVSMPKRSGCPKAEKNDSSQRVSINMNIINRNQGSKISPDYKNRSTSP
WDMFPNEDANRLPRTIWEAKCRHSGCINAEGKVDHHLNSVAIQQEILVLRREFPNCSTSFRLEKMLVTVGCTC
VTPRTVS

CLUSTALW alignment of putative IL-17 *M. eugenii* amino acid sequence and the ensembl sequence.

```
M.eugenii       MSSLGNLPGFKSLLLLLVLAVMMKTGVSMPKRSGCPKAEKNDSSQRVSINMNIINRNQGS 60
ensembl         MSSLGNLPG--SLLLLLVLAVMMKTGVSMPKRSGCPKAEKNDSSQRVSINMNIINRNQGS 58
                *********  **********************************************

M.eugenii       KISPDYKNRSTSPWDMFPNEDANRLPRTIWEAKCRHSGCINAEGKVDHHLNSVAIQQEIL 120
ensembl         KISPDYKNRSTSPWDMFPNEDANRLPRTIWEAKCRHSGCINAEGKVDHHLNSVAIQQEIL 118
                ************************************************************

M.eugenii       VLRREFPNCSTSFRLEKMLVTVGCTCVTPRTVS 153
ensembl         VLRREFPNCSTSFRLEKMLVTVGCTCVTPRTVS 151
                ********************************
```

Figure 6B.1. Alignment of expressed sequence with annotated sequence found in ensembl

```
M.domestica          ----------------MAFDGQNRPTKSNRIRSRNWRSVVVVVVVVVVKFKSLLLLLILA
M.eugenii            ----------------------------------------MSSLGNLPGFKSLLLLLVLA
E.caballus           -------------------------------------------MAPLRTSSVSLLLLLSLV
C.familiaris         ----------------------------------------MTLVTTSSMFQSLLLLLSLV
O.cuniculus          -----------------------------------------MSLGRISSVSLLLLLCLV
P.troglodytes        -----------------------------------------MTPGKTSLVSLLLLLSLE
H.sapiens            -----------------------------------------MTPGKTSLVSLLLLLSLE
P.abelii             -----------------------------------------MTPGKTSLVSLLLLLSLE
M.mulatta            -----------------------------------------MTPGKTSLVLLLLLLSLE
N.leucogenys         -----------------------------------------MTPGKTSLVSLLLLLSLE
C.jacchus            -----------------------------------------MTPGKTSLVSLLLLLILE
L.africana           -----------------------------------------MSSVRISSLSLLLLLSML
B.taurus             -----------------------------------------MASMRTSSMSLLLLLSLV
C.hircus             -----------------------------------------MASMRTASMSLLLLLSLV
C.elaphus            -----------------------------------------MASLRTSSMSLLLLLSLV
S.scrofa             -----------------------------------------MTPVRSSSLSLLLLLSLV
M.musculus           ----------------------------------------MSPGRASSVSLMLLLLLSLA
R.norvegicus         ----------------------------------------MSPRRIPSMCLMLLLLLNLE
C.griseus            ----------------------------------------MSPGRTSSVSLLLLLLLSLE
C.porcellus          MVIKVDQEKSKSIVGVRAEEKIIGSLQSLQRAPSWKGTRSPYAFPPTARTFRSLLLLSLM
O.anatinus           ----------------------------MTPGENVQFLFPTLVLMATLQESVLGKAIAA
G.gallus             -------------------------------------MSPIPYSPLFRPLLLVLLAMLS
O.latipes            ----------------------------------------MELPTHSICILMVICCSLR
D.rerio              ---------------------------------------------MSSALNLRFLMVACMM
                                                                              .

M.domestica          VMMKMGVSMPKRSGCP-----------KIEGNDSLQSIRVNMNMINRN---QGSKISPD
M.eugenii            VMMKTGVSMPKRSGCP-----------KAEKNDSSQRVSINMNIINRN---QGSKISPD
E.caballus           AIVKAGIVIPQNPECP-----------NTGDKNFPQNVKINLNVLNR---KTNSRRASD
C.familiaris         AIIKAGIAFPQNPGCR-----------NTEDKNFPQHVKVNLNILNR---NTNSRRPSD
O.cuniculus          ATVKNGIAMPRNPGCP-----------NAEDKNFPQNVKVSLNILNK---SVNSRRPSD
P.troglodytes        AIVKAGIAIPRNPGCP-----------NSEDKNFPRTVMVNLNIHNRN-TNTNPKRSSD
H.sapiens            AIVKAGITIPRNPGCP-----------NSEDKNFPRTVMVNLNIHNRN-TNTNPKRSSD
P.abelii             AIVKAGIAIPRNPGCP-----------NSEDKNFPRTVMVNLNIHNRN-TNTNPKRSSD
M.mulatta            AIVKAGIAIPRNPGCP-----------NSEDKTFPRTVMVNLNIHNRN-TNTNPKRSSD
N.leucogenys         AIVKAGIAIPQNPGCP-----------NSEDKNFPRTVMVNLNIHNRN-TNTNPKRSSD
C.jacchus            AIVKAGIASPQNPGCP-----------NAEDKNFPRTVMVNLNIRNRN-TNS--KRASD
L.africana           AMVIAGPAAPRNPGCP-----------NAEDKNFPHTVRLNLNITNRN-PNTNSRRSSD
B.taurus             ALVKAGVIIPQSPGCP-----------PTEDKNFPQHVRVNLNIVNR---STNSRRPTD
C.hircus             ALVKAGVIIPQSPGCP-----------PTEDKNFPQHVRVNLNIVNR---NTNSRRPTN
C.elaphus            ALVKAGVIIPQSPGCP-----------PTEDKNFPQHVRVNLNIVNR---NTNSRRPTD
S.scrofa             ALVKAGIMIPQSPGCP-----------KTEDKNFPQHVRVNLNILNR---STPARRPSD
M.musculus           ATVKAAAIIPQSSACP-----------NTEAKDFLQNVKVNLKVFNSLGAKVSSRRPSD
R.norvegicus         ATVKAAVLIPQSSVCP-----------NAEANNFLQNVKVNLKVLNSLSSKASSRRPSD
C.griseus            AVVKAGLPIPQSSECP-----------NTEAKNFLQNVKVNLKVLNSLSPKVNSRRPSD
C.porcellus          ATVKAGIPIPRNPGCPT----------ATEGKNFLQNVKLNLSIFNPLTQNVNSRRSSD
O.anatinus           KQPQNSLHLKKEKGCP-----------SSESDDFPHSVTVNLSITNGN---GTSKKFPS
G.gallus             ASISAHGKVIRPGLEPESLFKKADAGCLTQKDGKFPQTVRVNISISNMN---QDTKVTLD
O.latipes            FSSCSDEGVLHPPDCN-------------------VTLQFSSEIFSLS------RGNGN
D.rerio              GLVLISFGAEGASVRS-----------DQKNKNSHPEADHSYRLVLDAEFKASTNPIHP
                                                       .  .          .

M.domestica          YKNRSTSPWDLVQNVDENRQPRVIWEARCRYSGCINVEG--KVDYHRNSVPIQQEIMVLR
M.eugenii            YKNRSTSPWDMFPNEDANRLPRTIWEAKCRHSGCINAEG--KVDHHLNSVAIQQEILVLR
E.caballus           YHNRSTSPWNLHRNEDPERYPSVIWEAKCRHLGCVNAEG--KVDFHMNSVPIQQEILVLR
C.familiaris         YYNRSTSPWNLHRNEDPERYPSVIWEAKCRHLGCVNNEG--NINYHMNSVPIQQEILVLR
O.cuniculus          YYNRSTSPWTLHRNEDRERYPSVIWEAKCRHLGCVNAEG--NEDHHMNSVPIQQEILVLR
P.troglodytes        YYNRSTSPWNLHRNEDPERYPSVIWEAKCRHLGCINADG--NVDYHMNSVPIQQEILVLR
H.sapiens            YYNRSTSPWNLHRNEDPERYPSVIWEAKCRHLGCINADG--NVDYHMNSVPIQQEILVLR
P.abelii             YYNRSTSPWNLHRNEDPERYPSVIWEAKCRHLGCVNADG--NVDYHMNSVPIQQEILVLR
M.mulatta            YYNRSTSPWNLHRNEDPERYPSVIWEAKCRHLGCVNADG--NVDYHMNSVPIQQEILVLR
N.leucogenys         YYNRSTSPWNLHRNEDPERYPSVIWEAKCRHLGCVNADG--KVDYHMNSVPIQQEILVLR
C.jacchus            YYNRSSSPWNLHRNEDPERYPSVIWEAKCRHLGCVDADG--NVDYHMNSVPIQQEILVLR
L.africana           YYKRSTSPWSLHRNEDPERYPSVIWEAKCLHLGCVNADG--QVNHHMNSVPVKQEILVLR
B.taurus             YHKRSTSPWTLHRNEDPERYPSVIWEAKCSHSGCINAEG--KVDHHMNSVTIQQEILVLR
C.hircus             YHKRSTSPWTLHRNEDPERYPSVIWEAKCSHSGCINAEG--KVDHHMNSVTIQQEILVLR
C.elaphus            YHKRSTSPWTLHRNEDPERYPSVIWEAKCSHSGCINAEG--KVDHHMNSVTIQQEILVLR
S.scrofa             YSKRFTSPWTLQRNEDPERYSSVIWEAKCSHSGCINAEG--KEDHHMNSVPIQQEILVLR
```

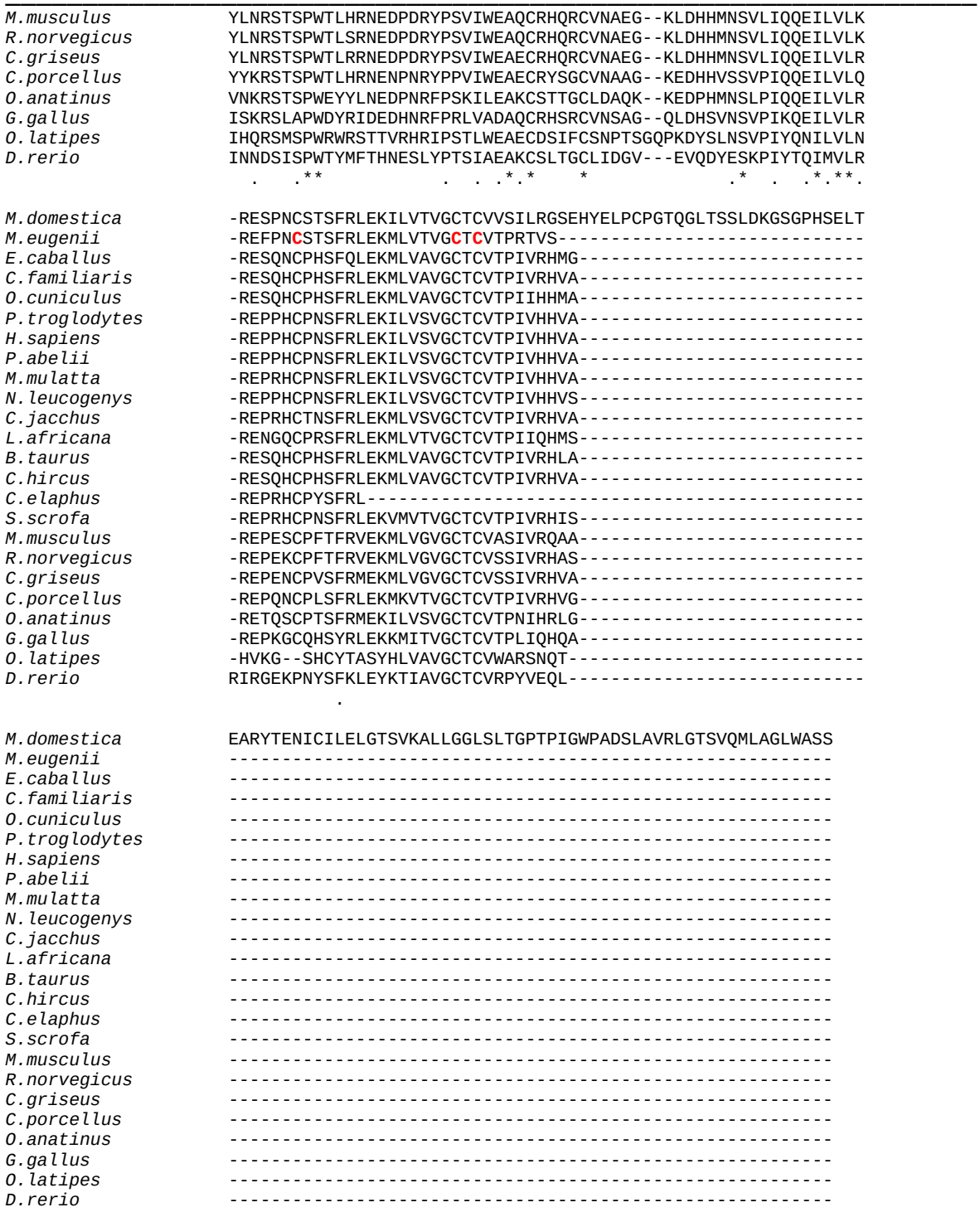


Figure 6B.2. Sequence alignment of IL-17.

C = red and bold. Signal peptide is underlined. Green highlight = polyleucine motif. Yellow highlight = polyvaline motif of *M.domestica*.

Appendix 6B

Table 6B.1. Genbank Accession numbers for IL-17 and the relevant references.

Species	Common Name	Genbank Accession Number	Reference
Ailuropoda melanoleuca	Giant Panda	XM_002915419	Annotated
Bos taurus	Cattle	NM_001008412	(Zimin *et al.,* 2009)
Callithrix jacchus	White-tuffed ear marmoset	EF534212	Direct submission
Canis lupus familiaris	Dog	AB514445 NM_001165878 XM_538958	Direct submission
Capra hircus	Goat	GU269912	Direct submission
Cavia porcellus	Domestic guinea pig	XM_003474763	Annotated
Cervus elaphus	Red Deer	EU860095	Direct submission
Cricetulus griseus	Chinese hamster	XM_003509128	Annotated
Danio rerio	Zebrafish	NM_001020787	(Gunimaladevi *et al.,* 2006)
Equus caballus	Horse	NM_001143792 XM_001498896, EU744563	(Tompkins *et al.,* 2010)
Gallus gallus	Chicken	AM773756	Direct submission
Homo sapiens	Human	NM_002190 Z58820	(Yao *et al.,* 1995b) (Fossiez *et al.,* 1998)
Loxodonta africana	African elephant	XM_003404140	Annotated
Macaca mulatta	Rhesus monkey	XM_001106391	Annotated
Monodelphis domestica	South American opossum	XM_001370118	Annotated
Mus musculus	Mouse	NM_010552	(Yao *et al.,* 1996)
Nomascus leucogenys	White cheeked gibbon	XM_003254163	Annotated
Ornithorhynchus anatinus	Platypus	XM_001509837	Annotated
Oryza latipes	Japanese medaka	NM_001204786	(Kono *et al.,* 2011)
Pan troglodytes	Chimpanzee	XM_527408	Annotated
Pongo abelii	Sumatran orangutan	XM_002816996	Annotated
Rattus norvegicus	Rat	NM_001106897 XM_001070601 XM_236985	(Florea *et al.,* 2005)
Sus scrofa	Pig	NM_001005729	(Katoh *et al.,* 2004)

Appendix 6C

Partial nucleotide sequence for *M. eugenii* Foxp3

aaggtcttcctggactcaggggatctcttaaaacacctccaagaagaccaccgcctggatgagaaggggaagg
cccagtgtctcatccagaaggaggtggtacagaatcttgagcacaagctgctcctggagaaggagaagctggg
ggccatgcaagcccacctctctgggaagctggcactggtaaagcccctggctgtgagcccctccaccgagaaa
gcaacctattgcccatcagggagcctgggccccacctggtcagcctggccaggcaccccagaagataagaaag
aagcgcgtctcccagggcagggcctctttgccgtcaggaggcacctgtggggtagccagatgtccccagaatt
tgtccataatctggaatactttcgatcccacaatctgcggccacccttcacctacgctactcttatccgctgg
gccatattggaagcccctgagaaacagcggacccactggtgagatcaa

Putative protein sequence for *M. eugenii* Foxp3

KVFLDSGDLLKHLQEDHRLDEKGKAQCLIQKEVVQNLEHKLLLEKEKLGAMQAHLSGKLALVKPLAVSPSTEK
ATYCPSGSLGPTWSAWPGTPEDKKEARLPGQGLFAVRRHLWGSQMSPEFVHNLEYFR
SHNLRPPFTYATLIRWAILEAPEKQRTHW

Appendix 6C

Partial nucleotide sequence for *O. fraenata* Foxp3

```
ccgcttggatgagaaggggaaggcccagtgtctcatccagaaggaggtggtacagaatcttgaacacaagctg
ctcctggagaaggagaagctgggggccatgcaagcccacctctccgggaagctggcactggtgaagcccctgg
ctatgagcccctccaccgagaaaggaacctattgcccatcagagagcctgggccccacctggtcagcctggcc
aggcaccccagaagataagaaagaagcgcgactcccagggcagggcctctttgccgtcaggaggcacctgtgg
ggtagccagatgtccccagaatttgtccataatctggaatactttcgatcccacaatctgcggccacccttca
cctatgctactcttatccgctgggccatattggaagcccctgagaaacagcggacc
```

Putative protein sequence for *O. fraenata* Foxp3

```
RLDEKGKAQCLIQKEVVQNLEHKLLLEKEKLGAMQAHLSGKLALVKPLAMSPSTEKGTYCPSESLGPTWSAWP
GTPEDKKEARLPGQGLFAVRRHLWGSQMSPEFVHNLEYFRSHNLRPPFTYATLIRWAILEAPEKQRT
```

Proposed isoform of Foxp3 yet to be confirmed

```
taggcctgcttggggacaacccagccgcatggaagagcgccattcgcgcaacgggagtctg
  R   P   A   W   G   Q   P   S   R   M   E   E   R   H   S   R   N   G   S   L
cataaatgctttgtgggaggggagaatgagaaaggggagggggagtgtggatgagttt
 H   K   C   F   V   G   G   E   N   E   K   G   G   G   G   S   V   D   E   F
gaattccgaaagaagagaagtcgggggcccagaagacgccaagacctgagacgcctacta
 E   F   R   K   K   R   S   R   G   P   R   R   R   Q   D   L   R   R   L   L
gcctgcgccgcacgggatgagacagaggccgggctccctctgccccccagctgaccatcc
 A   C   A   A   R   D   E   T   E   A   G   L   P   L   P   P   S   -   3'end
acacctgaacctagctgctgaaggaaccatttttattggttgtctttctcccgcaccccc
ccccccctcccagggaaggagggccttggtccccttccaccaataaaggcccgggggctg
aaccacaaaaaaaaaaaaaaaaaaaaaaacgttcgggggggggg
```

```
RPAWGQPSRMEERHSRNGSLHKCFVGGENEKGGGGSVDEFEFRKKRSRGPRRRQDLRRLLACAARDETEAGLP
LPPS
```

Appendix 6C

Amino Acid alignment for Foxp3

```
T.nigroviridis    ----------------------------------MMTPPAEAPQQLQQTPQQQIPSPQQ 25
X.laevis          -----MARLSGDRWLCPKWMGSLNCRMPNPQNPKATSAPSKESETQPDGKGKEQINPWSR 55
O.aries           -------------------------MPNPRPAKPLAPSLVLSPSPGASPSWR-AAPKAS 33
B.taurus          -------------------------MPNPRPAKPLAPSLVLSPSPGASPSWR-AAPKAS 33
S.scrofa          -------------------------MPNPRPAKPLAPSSVLSPSPGASPSWR-AVPKTS 33
C.familiaris      -------------------------MPNPRPAKPSAPSLAPGPSPGALPSWR-AAPKAS 33
A.melanoleuca     -------------------------MPNPRPAKPSAASLALGPSPGASPSWR-AAPKAS 33
F.catus           -------------------------MPNPRPAKPSAPSLALGPSPGASPSWR-AGPKTS 33
E.caballus        -------------------------MPNSRPAKPSAPSLALGPSPGTSPSWR-AAPKAS 33
H.sapiens         -------------------------MPNPRPGKPSAPSLALGPSPGASPSWR-AAPKAS 33
P.abelii          -------------------------MPNPRPGKPSAPSLALGPSPGASPSWR-AAPKAS 33
M.fascicularis    -------------------------MPNPRPGKPSAPSLALGPSPGASPSWR-AAPKAS 33
M.mulatta         -------------------------MPNPRPGKPSAPSLALGPSPGASPSWR-AAPKAS 33
C.jacchus         MMVDIWNPKSLETRALTRGPLSEKDSMPNPRPVKPSAPSLALGPSPGASPSWR-AAPKAS 59
R.norvegicus      -------------------------MPNPRPAKPMAPSLAPGPSPGGLPSWK-TAPKGS 33
M.musculus        -------------------------MPNPRPAKPMAPSLALGPSPGVLPSWK-TAPKGS 33
O.anatinus        ----MGKRPASRKWRQKVAINQSEDLMPSPKLNKSSSSSLLPNAKSAAGPVSKGAEPVAR 56
M.domestica       -------------------MYFFINPLNLGLSSTQPDFLQTEIDRCEHPRGPNCISLQK 40
O.fraenata        ------------------------------------------------------------
M.eugenii         ------------------------------------------------------------
O.mykiss          ----------------------------------MLQTESERLKSGRLNSGRHQQQREQR 26
D.rerio           ----------------------------------MLLNATGTHRGDDNRSSHQHLYQDED 26

T.nigroviridis    ----LQALLQQQKALMLH--------QQIQEVFKNQQEQLSMQLLQ-QKNAGIVSQELT- 71
X.laevis          RGAGVISQLQQNHGVMVTPSAGFSPPSQLQALLEDKKQTVVFHPNPSLMQVGVLNTELLS 115
O.aries           DQLGTKSPGTIFQGRDLRSGAHTSSSS--LNPMPPSQLQMPTVPLVMVAPSGARLGPSPH 91
B.taurus          DQLGTKSPGTTFQGRDLRSGAHTSSSS--LNPMPPSQLQMPTVPLVMVAPSGARLGPSPH 91
S.scrofa          DQQGAKGPGAAFQGRELRGGAHASSSS--LNPMPPSQLQLPTVPLVMVAPSGARLGPSPH 91
C.familiaris      DLLGAKGPGVTFQGRDLRGGTHASSS---LNPMPPSQLQLPTVPLVMVAPSGARLGPSPH 90
A.melanoleuca     DLLGAKGPGANFQSRDLRGGAHASSSSSSLNPMPPSQLQLPTVPLVMVAPSGARLGPSPH 93
F.catus           DPLGAKGPGATFQGRDLRGGTHASSS---LNPMPPSQLQLPTVPLVMVAPSGTRLGPSPH 90
E.caballus        DLLGAKGPGAAFQGRDLRGGAHASSS---LNPVPPSQLQLPTVPLVMVAPSGARLGPSPH 90
H.sapiens         DLLGARGPGGTFQGRDLRGGAHASSSS--LNPMPPSQLQL-------------------- 71
P.abelii          DLLGARGPGGAFQGRDLRGGAHASSSS--LNPMPPSQLQLPTLPLVMVAPSGARLGPLPH 91
M.fascicularis    DLLGARGPGGIFQGRDLRGGAHASSSS--LNPMPPSQLQLPTLPLVMVAPSGARLGPLPH 91
M.mulatta         DLLGARGPGGIFQGRDLRGGAHASSSS--LNPMPPSQLQLPTLPLVMVAPSGARLGPLPH 91
C.jacchus         DLLGARGPGGALQGRDLRGGAHASSSSS-LNPMPPSQLQLPTLPLVMVAPSGARLGPLPH 118
R.norvegicus      ELLGTRGPGGPFQGRDLRSGAHTSSSS--LNPLPPSQLQLPTVPLVMVAPSGARLGPSPH 91
M.musculus        ELLGTRGSGGPFQGRDLRSGAHTSSS---LNPLPPSQLQLPTVPLVMVAPSGARLGPSPH 90
O.anatinus        GVSGFHSQEALLNPR--YATATASS--------PQPQPSLPAFPIMMVTPPPGRLSTSPH 106
M.domestica       DYLKTSSMPNHSKAHQIPAASSSVASSS-LVAHMKDTELLPTGPMVMVAPPGGQLSTLPH 99
O.fraenata        ------------------------------------------------------------
M.eugenii         ------------------------------------------------------------
O.mykiss          EEPDACTQPKDSASPQLCARTSVTQMGFPLMIRPGVSLMASSQLQSILLRQCSSEEEGRS 86
D.rerio           CATFSIIQMKSRISNSLLTSPKPMATKISVALESDLRGLGSSRNQNFTLQKQSASG---- 82

T.nigroviridis    ------AQQIAIQQQLLQVQQHLLNLQRQGLLSVLPASPITAPGCENGSLLSAGGDARES 125
X.laevis          KLSRDTASQSPIIHLSPTSSTSILNLQPARLYPVMAKHKNQHPITQGFNLTSLGWAQQEA 175
O.aries           LQALLQDRPHFVHQLSTVDAHARTPVLQVRPLDSPAMISLPPPTAATGLFSLKARPG--- 148
B.taurus          LQALLQDRPHFVHQLSTVDAHARTPVLQVRPLDSPAMISLPPPTAATGLFSLKARPG--- 148
S.scrofa          LQALLQDRPHFVHQLSTVDAHARTPVLQVRPLDSPAMISLPPPTAATGVFSLKARPG--- 148
C.familiaris      LQALLQDRPHFMHQLSTVGTHTRTPVLQVRPLDSPAMISLPPPTAATSVFSLKARPG--- 147
A.melanoleuca     LQALLQDRPHFMHQLSMVDTHARTPVLQVRPLDSPAMISLPPPTAATSVFSLKARPG--- 150
F.catus           LQALLQDRPHFMHQLSTVDTHARTPVLQVRPLDSPAMISLPPPTAATGVFSLKARPG--- 147
E.caballus        LQALLQDRPHFMHQLSTVDTHARTPVLQVRPLDSPAMISLPPPTAATGVFSLKARPG--- 147
H.sapiens         --------------STVDAHARTPVLQVHPLESPAMISLTPPTTATGVFSLKARPG--- 113
P.abelii          LQALLQDRPHFMHQLSTVDAHARTPVLQVHPLESPAMISLPPPTTTTGVFSLKARPG--- 148
M.fascicularis    LQALLQDRPHFMHQLSTVDAHARTPVLQVHPLESPAMISLPPPTTATGVFSLKARPG--- 148
M.mulatta         LQALLQDRPHFMHQLSTVDAHARTPVLQVHPLESPAMISLPPPTTATGVFSLKARPG--- 148
C.jacchus         LQALLQDRPHFMHQLSTVDAHARTPVLQVHPLESPAMISLPPPTTTTGVFSLKARPG--- 175
R.norvegicus      LQALLQDRPHFMHQLSTVDAHGHTPVLQVRPLDNPAMIGLPPPTAATGVFSLKARPG--- 148
M.musculus        LQALLQDRPHFMHQLSTVDAHAQTPVLQVRPLDNPAMISLPPPSAATGVFSLKARPG--- 147
O.anatinus        LQALLQDKQQFVQQLSIENR-GRTPFLHVTPLSSPSLLNVPPP---TGVFSLKARPAQLH 162
M.domestica       PQALLQDKQHFVHQLTSTEVLGRPSLVHMTPLSTPALINLPSPPDIIAYKTRTSQLHS-- 157
O.fraenata        ------------------------------------------------------------
M.eugenii         ------------------------------------------------------------
```

```
O.mykiss          LLQRVSLCPQLSQHRPSVLRQGVQAAHVRPQAPCTLTEASSAQPISLCKVEVDTGSR--- 143
D.rerio           -----STTKYFKQHRPSVLRKGNQP---FPQA-CSAHDWVVDT---VCKTEPDSEPS--- 127


T.nigroviridis    ------------SSQQCT-----VNGHQPL---LRKKDS------------------- 144
X.laevis          RLGKPEEVILGKNSSTFSS-----LSHSQPLNTVKTQKKL------------------- 210
O.aries           -LPPGINVASLEWVSREPA-----LLCTFPSP--GMPRKD------------------- 180
B.taurus          -LPPGINVASLEWVSREPA-----LLCTFPSP--GMPRKD------------------- 180
S.scrofa          -LPPGINVASLEWVSREPA-----LLCTFPSP--GVPRKD------------------- 180
C.familiaris      -LPPGINVASLEWVSREPA-----LLCTFPSP--STPRKD------------------- 179
A.melanoleuca     -LPPGINVASLEWVSREPA-----LLCTFPSP--STPRKD------------------- 182
F.catus           -LPPGINVASLEWVSREPA-----LLCTFPSP--STPRKD------------------- 179
E.caballus        -LPPGINVASVEWVSREPA-----LLCTFPSP--SAPRKD------------------- 179
H.sapiens         -LPPGINVASLEWVSREPA-----LLCTFPNP--SAPRKD------------------- 145
P.abelii          -LPPGINVASLEWVSREPA-----LLCTFPNP--GALRKDRSVDRAGKDLRPPIPSPDTL 200
M.fascicularis    -LPPGINVASLEWVSREPA-----LLCTFPNP--GAPRKD------------------- 180
M.mulatta         -LPPGINVASPEWVSRELA-----LLCTFPNP--GAPRKD------------------- 180
C.jacchus         -LPPGINVASVEWLSREPT-----LLCTFPNP--GAPRKD------------------- 207
R.norvegicus      -LPPGINVASLEWVSREPA-----LLCTFPRS--GTPRKD------------------- 180
M.musculus        -LPPGINVASLEWVSREPA-----LLCTFPRS--GTPRKD------------------- 179
O.anatinus        SLSHGINLASLEWVPKEPTNTLTNYLCAVPSPGAGEGRKE------------------- 202
M.domestica       -LPPGINLANFEWLPKEPANMLTTYLCTFPSGPSTGDATAFR----------------- 198
O.fraenata        ------------------------------------------------------------
M.eugenii         ------------------------------------------------------------
O.mykiss          ------GQSSPPHSEHSPG-----PTRHPSPPRTASPKQSSIITR-------------- 177
D.rerio           ------DAIPLYTGQSESR-----IGAYASSPQPGSP----------------------- 153
```

Transcirption DNA binding nuclear box regulation

Forkhead coiled coil P2 zinc

↓

C2H2 Zinc finger

```
T.nigroviridis    --GCPD-ENTQNSHPLYGNGMCKWPGC---ETVFGDLQAFLKHLNSEHILDDKSTAQCRV 198
X.laevis          --GVQNKESPEPICPVYYRGACTFPGC---GKAFEDHRHFLRHLHSDHHLDDKSTVQCLI 265
O.aries           ---STLLTVPQGSYSLLANGVCKWPGC---EKVFKEPEDFLKHCQADHLLDEKGRAQCLL 234
B.taurus          ---STLSTVPQGSYSLLANGVCKWPGC---EKVFKEPEDFLKHCQADHLLDEKGRAQCLL 234
S.scrofa          ---STLSTVPQGSYSLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 234
C.familiaris      ---STLPTVPQGSYSLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 233
A.melanoleuca     ---STLSTVPQGSYSLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 236
F.catus           ---STLSTXPQGSYSLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 233
E.caballus        ---STLSTMPQGSYSLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 233
H.sapiens         ---STLSAVPQSSYPLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 199
P.abelii          CPPSTLSAVPQSSYPLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 257
M.fascicularis    ---STLSAMPQSSYPLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 234
M.mulatta         ---STLSAMPQSSYPLLANGVCKWPGC---EKVFEEPEDFLKHCQADHLLDEKGRAQCLL 234
C.jacchus         ---STLSATPQSSYPLLANGICKWPGC---EKVFEEPEDFLKHCQEDHLLDEKGRAQCLL 261
R.norvegicus      ---SNLLAAPQGSYPLLANGVCKWPGC---EKAFEEPGEFLKHCQADHLLDEKGKAQCLL 234
M.musculus        ---SNLLAAPQGSYPLLANGVCKWPGC---EKVFEEPEEFLKHCQADHLLDEKGKAQCLL 233
O.anatinus        ---SAAPSSPDGSHPLLANGACRWPGC---EKVFEESKEFLKHFHTDHRMDEKGRAQCLV 256
M.domestica       --LGPKKESILQTCPLDSSQSCWWPGC---EKVFLEPGELLKHLQEDHRLDEKGKAQCLI 253
O.fraenata        ------------------------------------------------RLDEKGKAQCLI 12
M.eugenii         ------------------------------KVFLDSGDLLKHLQEDHRLDEKGKAQCLI 29
O.mykiss          -QHEAGHPTLEGSSALFLNGLCCWPGC---DAVFEEFPSFLKHLHSDHGHGDRSIAQWKV 233
D.rerio           -EYTGKHPYSLSGDYLCVKGQCRWPGCSKSEDVFTEYGHFLRHLSTDHAPGDRSIGQLRM 212
                                                                 ....  *  .
```

Transcription DNA-binding Box Regulation
nuclear Metal-Binding Zinc-Finger

```
T.nigroviridis    QMQVVQQLELQLKKDKERLQAMMAHLKSSEPKPAAQPINLASNVSLSQATLPKGPAPMSV 258
X.laevis          QTEVVHKLEEQLAVEKERLHHMQSQMSG---KLNTQALHLSKQRECGLILHPTHPS--IS 320
O.aries           QREVVQSLEQQLVLEKEKLGAMQAHLAGKMAQTKAPSA-ASSDKGSCCIVATGTPGTTVP 293
B.taurus          QREVVQSLEQQLVLEKEKLGAMQAHLAGKMAQTKAPSA-ASSDKGSCCIVATGTPGTTVP 293
S.scrofa          QREVVQSLEQQLVLEKEKLGAMQAHLAGKMPSPKAPSA-ASSDKGSCCIVATGTPGTAVP 293
C.familiaris      QREVVQSLEQQLVLEKEKLGAMQAHLAGKMTLTKAPST-ASSDKGSCCIVAAGTPATTGP 292
A.melanoleuca     QREVVQSLEQQLVLEKEKLGAMQAHLAGKMALTKAPST-ASSDKGSCC-----TPVATGP 290
F.catus           QREVVQSLEQQLVLEKEKLGAMQAHLAGKMALTKAPST-ASSDKGSCCIVATGTPAATGP 292
E.caballus        QREVVQSLEQQLVLEKEKLGAMQAHLAGKMALTKAPSA-ASSDKGSCRLAATGTPGTAVP 292
H.sapiens         QREMVQSLEQQLVLEKEKLSAMQAHLAGKMALTKASSV-ASSDKGSCCIVAAGSQGPVVP 258
P.abelii          QREMVQSLEQQLVLEKEKLSAMQAHLAGKMALTKASSV-ASSNKGSCCIVAAGSQGSVIP 316
M.fascicularis    QREMVQSLEQQLVLEKEKLSAMQAHLAGKMALTKASSV-ASSDKGSCCIVAAGSQGSAVP 293
M.mulatta         QREMVQSLKQQLVLEKEKLSAMQAHLAGKMALTKASSV-ASSDKGSCCIVAAGSQGSAVP 293
C.jacchus         QREMVQSLEQQLVLEKEKLSAMQAHLAGKMAVPKAPSV-ASSDKASCCIVAAGSQGSMVP 320
R.norvegicus      QREVVQSLEQQLELEKEKLGAMQAHLAGKMALTKAPPV-ASVDKSSCCLVATSTQGSVLP 293
M.musculus        QREVVQSLEQQLELEKEKLGAMQAHLAGKMALAKAPSV-ASMDKSSCCIVATSTQGSVLP 292
O.anatinus        QKEVVQSLEQQLVLEKEKLSAMQAHLTGKLSLPKLPSS-ISTEKPNGCLPGTLSPSLAT- 314
M.domestica       QKEVVQNLEQKLLLEKEKLGAMQAHLSGKLALVKPLAN-PSTEKVTYCP--SRSLGPTWS 310
O.fraenata        QKEVVQNLEHKLLLEKEKLGAMQAHLSGKLALVKPLAMSPSTEKGTYCP--SESLGPTWS 70
M.eugenii         QKEVVQNLEHKLLLEKEKLGAMQAHLSGKLALVKPLAVSPSTEKATYCP--SGSLGPTWS 87
O.mykiss          QQDMVQYMETQLTVEKQKLFAMQLHLHLSGHKSTVLKAASDWPYRHSLSLGLPQNRGGVS 293
D.rerio           QKDRVQHMENQLTAERQKLQAMQLHLLD--VKSTSEGGNIVEKPAHLSGLLQPASSNDHY 270
                  * . *. .. .*  ....*  *  ..

T.nigroviridis    SQSATAPTTPLTPHSESPSVLTPSSMFTGTPVRRRYSRSVSQDISDNKEFYLSTEVRPPF 318
X.laevis          AWSGLDLSSPLQKEFSDTILALRRQLWEGS------SLNIFQNMANCIEYYKTNNVRPPF 374
O.aries           AWPGPQE------APDG-LFAVRRHLWGSHGNS------TFPEFFHNMDYFKFHNMRPPF 340
B.taurus          AWPGPQE------APDG-LFAVRRHLWGSHGNS------TFPEFFHNMDYFKFHNMRPPF 340
S.scrofa          AWPGPQE------APDG-LFAVRRHLWGSHGNS------TFPDFFHNMEYFKFHNMRPPF 340
C.familiaris      AWSSPQE------APDG-LFAVRRHLWGSHGNS------TFPEFFHNMDYFKFHNMRPPF 339
A.melanoleuca     AWPSPQE------APDG-LFAVRRHLWGSHGNS------TFPEFFHNMDYFKFHNMRPPF 337
F.catus           AWPSPQE------APDG-LFAVRRHLWGSHGNS------TFPEFFHNMDYFKFHDMRPPF 339
E.caballus        AWPSPQE------APDG-LFAVRRHLWGSHGNS------TFPEFFHNMDYFKFHNMRPPF 339
H.sapiens         AWSGPRE------APDS-LFAVRRHLWGSHGNS------TFPEFLHNMDYFKFHNMRPPF 305
P.abelii          AWSGPRE------APDS-LFAVRRHLWGSHGNS------TFPEFLHNMDYFKFHNMRPPF 363
M.fascicularis    AWSGPRE------APDS-LFAVRRHLWGSHGNS------TFPEFLHNMDYFKFHNMRPPF 340
M.mulatta         AWSGPRE------APDS-LFAVRRHLWGSHGNS------TFPEFLHNMDYFKFHNMRPPF 340
C.jacchus         AWSGPRE------APDS-LFAVRRHLWGSHGNS------TFPEFLHNMDYFKFHNMRPPF 367
R.norvegicus      AWSSPRE------ASDS-LFAVRRHLWGSHGNS------TFPEFFHNMDYFKYHNMRPPF 340
M.musculus        AWSAPRE------APDGGLFAVRRHLWGSHGNS------SFPEFFHNMDYFKYHNMRPPF 340
O.anatinus        AWPSSKE------SPDS-LFAMRRHLWSSHGVS------MCPDILHNMEYFKFNNMRPPF 361
M.domestica       SWPGSLEDKDKAQLPGQGLFAVRRHLWGSHMS---------PDFVHNLEYFRSHNLRPPF 361
O.fraenata        AWPGTPEDKKEARLPGQGLFAVRRHLWGSQMS---------PEFVHNLEYFRSHNLRPPF 121
M.eugenii         AWPGTPEDKKEARLPGQGLFAVRRHLWGSQMS---------PEFVHNLEYFRSHNLRPPF 138
O.mykiss          RCTTKEP-----------EELEQHEYWPSAAP----HHLRPDLIPSVECYKYNNIRPLY 337
D.rerio           DCERAAT-----------EALTQG-YWQIST-----SQVIPGIIPSFEYYKFTNMRPPF 312
                                              .               .    . .   ..** .

                         Forkhead domain
T.nigroviridis    TYASLIRQAIFESPRSQLTLNEIYNWFTRNFAYFRSNAATWKNAVRHNLSLHKCFVRLEN 378
X.laevis          TYASLIRWAILESPQKQLALNEIYHWFTRMFAFFRYNTATWKNAVRHNLSLHKCFVRVEN 434
O.aries           TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 400
B.taurus          TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 400
S.scrofa          TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 400
C.familiaris      TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 399
A.melanoleuca     TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 397
F.catus           TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 399
E.caballus        TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 399
H.sapiens         TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 365
P.abelii          TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 423
M.fascicularis    TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 400
M.mulatta         TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 400
C.jacchus         TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAFFRNHPATWKNAIRHNLSLHKCFVRVES 427
```

```
R.norvegicus      TYATLIRWAILEAPERQRTLNEIYHWFTRMFAYFRNHPATWKNAIRHNLSLHKCFVRVES 400
M.musculus        TYATLIRWAILEAPERQRTLNEIYHWFTRMFAYFRNHPATWKNAIRHNLSLHKCFVRVES 400
O.anatinus        TYATLIRWAILEAPEKQRTLNEIYHWFTRMFAYFRNQPATWKNAIRHNLSLHKCFVRVEN 421
M.domestica       TYATLIRWAILEAPEKQRTLNEIYHWFTHTFAFFRTHPATWKNAIRHNLSLHKCFVRVEN 421
O.fraenata        TYATLIRWAILEAPEKQRT---------------------------------------- 140
M.eugenii         TYATLIRWAILEAPEKQRT-----HW-------------------------------- 159
O.mykiss          TYACMIRWSILESPDKQRSLNDIYNWFTTMFFYFRHNTATWKNAVRHNLSLHKCFVRVEG 397
D.rerio           TYASMIRWAILKSPEKQLTLKEIYQWFTSMFFYFRHNTATWKNAVRHNLSLHKCFVRVEG 372
                  *** .** .*...*  * .

T.nigroviridis    VKGAVWTVDEIEFHRRRPQKPAGAG---------------------------------- 403
X.laevis          IKGAVWMVDELEFQRKRGVRNSR------------------------------------- 457
O.aries           EKGAVWTVDEFEFRKKRSQRPSRCSNPT-------------PGP--------------- 431
B.taurus          EKGVVWTVDEFEFRKKRSQRPSRCSNPT-------------PGP--------------- 431
S.scrofa          EKGAVWTVDEFEFRKKRSQRPSRCSNPT-------------PGP--------------- 431
C.familiaris      EKGAVWTVDEFEFRKKRSQRPSRSSNPT-------------PGP--------------- 430
A.melanoleuca     EKGAVWTVDEFEFRKKRSQRPSRSANPT-------------PGP--------------- 428
F.catus           EKGAVWTVDEFEFRKKRSQRPSRCSNPT-------------PGP--------------- 430
E.caballus        EKGAVWTVDEFEFRKKRSQRPSRCSNPT-------------PGP--------------- 430
H.sapiens         EKGAVWTVDELEFRKKRSQRPSRCSNPT-------------PGP--------------- 396
P.abelii          EKGAVWTVDELEFRKKRSQRPSRCSNPT-------------PGP--------------- 454
M.fascicularis    EKGAVWTVDELEFRKKRSQRPSRCSNPT-------------PGP--------------- 431
M.mulatta         EKGAVWTVDELEFRKKRSQRPSRCSNPT-------------PGP--------------- 431
C.jacchus         EKGAVWTVDELEFRKKRSQRPSRCSNPT-------------PGP--------------- 458
R.norvegicus      EKGAVWTVDEFEFRKKRSQRPSKCSNPC-------------P----------------- 429
M.musculus        EKGAVWTVDEFEFRKKRSQRPNKCSNPC-------------P----------------- 429
O.anatinus        EKGAVWTVDEVEYRRKRSQRPSSPADSSGVVSRRGDGEEERHPSQRVGNRVGQTKLTAAA 481
M.domestica       EKGAVWTVDEFEFRKKRRPRPSRDQDLKR-----------LLACTPMAVTEAWLPLPHRN 470
O.fraenata        ------------------------------------------------------------
M.eugenii         ------------------------------------------------------------
O.mykiss          GKGAVWTVDEMEYQRRKGQKYHRDHHVKWLAPYSLFRPEEP------------------- 438
D.rerio           RKGSVWTVDEEEFLRRKGQKLHRDHDMDWMAPFQLFPLTPQGESYQM------------ 419


T.nigroviridis    ----------
X.laevis          ----------
O.aries           ----------
B.taurus          ----------
S.scrofa          ----------
C.familiaris      ----------
A.melanoleuca     ----------
F.catus           ----------
E.caballus        ----------
H.sapiens         ----------
P.abelii          ----------
M.fascicularis    ----------
M.mulatta         ----------
C.jacchus         ----------
R.norvegicus      ----------
M.musculus        ----------
O.anatinus        SSLASNLPLTSSRSELEKTTL 502
M.domestica       SHI------------------ 473
O.fraenata        ----------
M.eugenii         ----------
O.mykiss          ----------
D.rerio           ----------
```

Figure 6C.1. Alignment of Foxp3 sequences from sequences deposited in Genbank.
Forkhead domain, Zinc Finger domain and Forkhead coiled coil Zinc ligand are marked in the sequence.

Appendix 6C

Table 6C.1. Genbank Accession Numbers for Foxp3 and the relevant references.

Species	Common Name	Genbank Accession Number	Reference
Sus scrofa	Pig	AY669812	Direct submission
Tetraodon nigroviridis	Spotted green pufferfish	FJ358692	Direct submission
Homo sapiens	Human	NM_001114377	(Chatila *et al.,* 2000)
Callithrix jacchus	Common marmoset	GQ284839	Direct submission
Equus caballus	Horse	NM_001163272	(Robbin *et al.,* 2011)
Xenopus laevis	African clawed frog	AB359948	Direct submission
Ovis aries	Sheep	FJ491732	Direct submission
Felis catus	Cat	EF419427	(Lankford *et al.,* 2008)
Ornithorhynchus anatinus	Platypus	XM_001507231	Annotated
Macaca fascicularis	Crab eating macaque	AY376065.1	Direct submission
Rattus norvegicus	Rat	NM_001108250	(Abe *et al.,* 2009)
Bos taurus	Cattle	NM_001045933	(Zimin *et al.,* 2009)
Canis lupus familiaris	Dog	XM_548996, NM_001168461	(Mizuno *et al.,* 2009)
Pongo abelii	Sumatran orangutan	XM_002831634	Annotated
Macaca mulatta	Rhesus monkey	NM_001032918	Annotated
Ailuropoda melanoleuca	Giant Panda	XM_002917744	Annotated
Mus musculus	Mouse	NM_001199347	(Josefowicz *et al.,* 2012)
Monodelphis domestica	South American grey short tailed opossum	XM_001372363	Annotated
Oncorhynchus mykiss	Rainbow trout	FM883710	Direct submission
Danio rerio	Zebrafish	FN435333	Direct submission

Appendix 7

Recipes of reagents used in this body of work

1.0 Kit reagents

1.1 Promega PCR Clean up kit.

1.1.1 Membrane bind solution.
4.5M guanidine isothiocyanate
0.5M potassium acetate at pH 5.0
1x Tris EDTA buffer (TE)(10 mM Tris-HCl at pH 7.5, 1mM EDTA at pH 8.0)
1x TBE buffer containing 89mM Tris base
89mM boric acid and 2mM EDTA at ph8.0
1x Tris acetate EDTA (TAE) buffer (40mM Tris base, 5mM sodium acetate
1mM EDTA at pH 8.0)

1.1.2 Wash solution.
10mM potassium acetate at pH 5.0
80% ethanol
16.7µM EDTA at pH8.0

1.2 Promega DNA clean up kit.

1.2.1 Cell Resuspension Solution
50mM Tris-HCl (pH7.5)
10mM EDTA,
100µg/mL RNase A

1.2.2 Cell Lysis Solution
0.2 M NaOH
1% SDS

1.2.3 Neutralization Solution
1.32M potassium acetate at pH 4.8

1.2.4 Wash solution
80mM potassium acetate
8.3mM Tris-HCl at pH 7.5
40µM EDTA

1.3 RIPA Buffer (Pierce)

20mM Tris-HCl at pH 7.5,
150mM NaCl, 1mM Na_2EDTA,
1mM EDTA, 1% NP-40,
1% sodium deoxycholate,

2.5mM sodium pyrophosphate,
1mM β-glycerophosphate,
1µg/mL leupeptin

1.4 Lambda DNA Marker.
10mM Tris-HCl (pH 7.6), 1mM EDTA
6X DNA Loading Dye
10mM Tris-HCl (Ph 7.6), 0.03% bromophenol blue
0.03% xylene-cyanol FF, 60% glycerol and 60nM EDTA

Reagents prepared in laboratory

Tris Borate EDTA Buffer

Component	Amount	10X Stock Concentration	Final 1X Concentration
Tris Base	108g	890mM	89mM
Boric Acid	55g	890mM	89mM
EDTA (pH 8.0)	40mL	20mM	2mM

4X Resolving Gel Buffer (100mL)

Component	Amount
Tris HCl (pH 8.8,2.0M)	75mL (1.5M)
10% SDS	4mL (0.4%)
MilliQ water	21mL

10% APS working solution
0.1g APS in 1mL water

10mL Resolving Gel

Component	Amount
30% Acrylamide/Bisacrylamide 29.1	4mL
4X resolving buffer	2.5mL
Milli Q water	3.5mL
10% Ammonium Persulfide (APS)	50µL
TEMED	5µL

4x Stacking Gel Buffer (100mL)

Component	Amount
Tris HCl (pH 6.8, 1M)	50mL (0.5M)
10% SDS	4mL
Milli Q water	46mL

4% Stacking Gel (10mL)

Component	Amount
30% Acrylamide/Bisacrylamide 29.1	1.4mL
4X stacking buffer	2.5mL
Milli Q water	6.15mL
10% Ammonium Persulfide (APS)	50µL
TEMED	5µL

Laemmli Buffer 2X

4% SDS
10% 2-mercaptoethanol
20% glycerol
0.004% bromophenol blue
0.125M Tris HCl
pH to 6.8

Coomassie brilliant blue solution

500mL Methanol
100mL Acetic Acid
1g Coomassie brilliant blue-R250 (Bio-Rad)
Dissolve in 1L of MilliQ water

De-staining solution

450mL Methanol
450mL water
100mL Acetic acid

Blocking Buffers

5% non-fat milk powder in TBST
3% non-fat milk powder in TBST

Tris Buffered Saline (TBS) 10X

Component	Amount
Tris HCl	24.23g
Sodium Chloride (NaCl)	80.06g
MilliQ water	800mL
pH to with pure HCl	7.6

Tris Buffered Saline Tween (TBS-T)

For 1 liter. 100mL of TBS 10X + 900mL of MilliQ water + 1mL of Tween 20 or 10ml of a 10% solution (easier to dispense then the very viscous undiluted Tween20).

10% solution. 2mL of Tween 20 + 18mL of MilliQ water

Transfer Buffer-Tris-glycine 1X

Component	Amount
Tris Base	6.06g
Glycine	28.83g
Methanol	50mL
MilliQ to	2.0 L

Or

Component	Amount
Tris Base	2.9g
Glycine	5.8g
Methanol	200mL
SDS	0.37g
MilliQ to	1.0 L

(Both buffer systems were used)

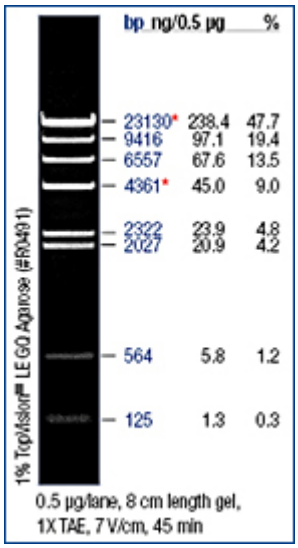
SYPRO Ruby Fixing solution (200mL)

Component	Amount
50% Methanol	100mL
7% Acetic Acid	14mL
MilliQ water	86mL

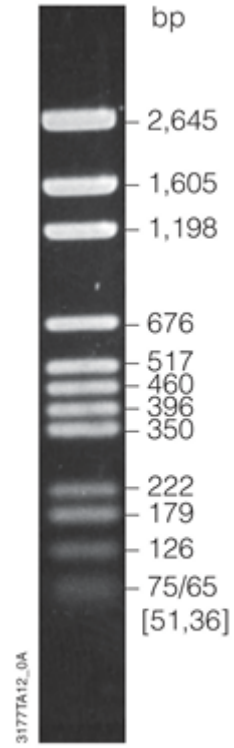
SYPRO Ruby Wash solution (100mL)

Component	Amount
10% Methanol	10mL
7% Acetic Acid	7mL
MilliQ water	83mL

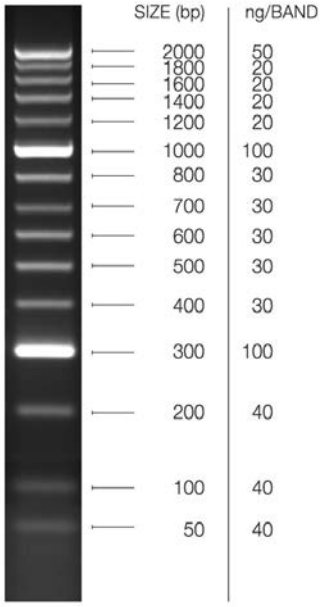
Marker used in this study.



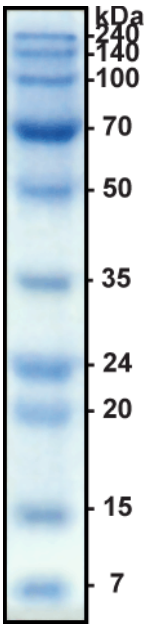
DNA marker (benchtop pgem, promega, Australia)



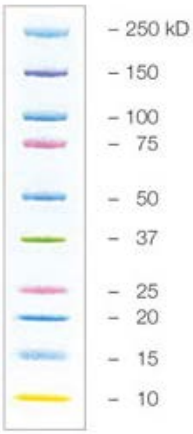
Hyperladder II (Bioline, Australia)



Protein markers (Bio-rad, Australia)



All Blue
Bio-rad



Kaleidoscope marker
Bio-rad

Appendix 7