

Coevolution of expression of Prion-protein and Doppel genes

Tatiana Vassilieva

**A thesis submitted for the degree of Doctor of
Philosophy of the Australian National University**

June 2013

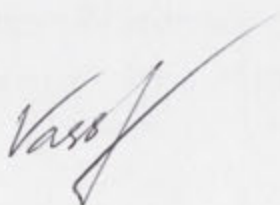
Statement of originality

The contents of this thesis are the result of my original research, which has been conducted under the principal supervision of Professor Jill Gready (John Curtin School of Medical Research, ANU).

The *X. tropicalis* genome database and the NCBI EST database screening for *PRNP* and *PRND* genes, resulting in identification and assembly of the *X. tropicalis* *PRNP/PRND* contig and of the *X. laevis* *PRNP* variants, was performed by Nagesh Chakka, the former PhD student in Prof Gready group. I used the obtained sequence information to design primers for the experimental work. Nagesh Chakka also conducted the phylogenetic footprinting analysis presented in Chapter 4.

Alignment of BPTF homologues from different species was performed by Dr Alex Zelinsky (Erasmus Medical Center, Netherlands).

I declare that the work presented in this thesis is to my believe original, except as acknowledged in the text and above, and that the material has not been submitted, either in whole or in part, for a degree at this or any other university.



Tatiana Vassilieva

June 2013

Acknowledgements

I am very grateful to my supervisor Prof Jill Gready for her patient and invaluable help and support during the course of my project. Most importantly, I would like to thank her for inspiring me to enrol into PhD study and giving me freedom to plan and perform this research using her critical guidance and thoughtful comments.

I would also like to express my special gratitude to Prof Jenny Graves for welcoming me in her laboratory at the beginning of this project. Her numerous suggestions and support are greatly appreciated.

I would like to acknowledge people who provided me with generous advises: Dr Kelly Jordan-Sciutto for sharing her knowledge in FAC1/BPTF subject and offering many valuable comments, Prof Tomas Preiss for discussion about my polysome binding assay, and Dr Alison Shield for advice on luciferase assay.

I would particularly like to thank Nagesh Chakka for being a great research partner and sharing his knowledge in bioinformatics with me.

I am also grateful to people who have put their energy to collect and share animal tissues, cell lines, and RNA essential for my research: Prof Kate L. Loveland (Monash Institute of Medical Research) and Prof Carolyn Geczy (University of New South Wales) for various cell lines, Dr Timothy C. Grammer (Harlands lab, University of California, Berkeley) for purified RNA from *X. tropicalis*, and Prof Norman R. Saunders (University of Melbourne) for opossum tissues.

Thanks to all past and present members of the Computational and Conceptual biology group: Mark Wallace, Babu Kannappan, Animesh Agrawal, Alex Zelensky, Sue Corley, Peter Cummins, Peter Mathews, and Aleu Mani George for their support, whenever I needed, which made the place so warm and friendly.

Finally, I would like to express my deepest love and gratitude to my wonderful daughters whose encouragement and support were very crucial in motivating me to finish this project.

Abstract

Gene duplication is a major mechanism for the origin of new genes. Subsequent function divergence and acquisition of new regulatory DNA elements is the necessary step for survival of a newly duplicated gene. However, despite of existence of several models, it remains unclear how paralogs evolve after the duplication and develop new functions. An example of evolution by gene duplication is *PRNP*, coding for prion protein (PrP), and its downstream duplicate, *PRND*, coding for doppel (Dpl). Both genes are well known in Eutherian mammals: *PRNP*, notoriously for its involvement in neurodegenerative disorders, and *PRND* as a testis-specific gene whose loss in knockout mice leads to male sterility. In addition, recent studies have shown that PrP possesses neuroprotective functions while Dpl is neurotoxic and induces apoptosis. Expression of PrP and Dpl is altered in various neurodegenerative disorders, such as Alzheimer's, as well as after brain injury and during development; this suggests their possible antagonistic coregulation by a so far unknown mechanism.

The aims of the work described in this thesis are to examine the evolution of the *PRNP* and *PRND* genes using three models: frog (*Xenopus laevis*), marsupial opossum (*Monodelphis domestica*) and mouse, and to define the molecular mechanisms underlying the antagonism in the functions of their proteins.

Previously unreported amphibian and marsupial *PRND* transcripts were cloned and characterised. Tissue expression analysis demonstrated that in frog *PRND* relies primarily on regulatory DNA elements of the *PRNP* gene and is expressed in brain as a chimeric transcript, sharing the first two 5' UTR exons of the *PRNP* gene. *PRND* acquires its own promoter in mammals and develops new tissue specificity. Conservation of the chimeric transcript in mouse and its translation (shown by polysome binding assay) in mice embryos suggests possible functional significance in development.

Transcription factors that might regulate the antagonistic relationship of PrP and Dpl by varying expression of the genes under specific cellular conditions were predicted by bioinformatics (phylogenetic footprinting analysis) and comparative genomics analysis

of the two genes within vertebrates. The predicted FAC1/BPTF binding site, situated next to the second non-coding exon of *PRNP*, was selected for experimental analysis based on the literature-deduced relationship between FAC1/BPTF, PrP, and Dpl. Reporter gene assays demonstrated upregulation of luciferase expression in constructs containing both a *Prnp* conserved non-coding DNA region with the predicted FAC1/BPTF binding site and a *Prnd* promoter region without FAC1/BPTF binding site; the latter might be explained by an indirect effect. Binding sites for the transcription factors E2F1, CREB, AP2, USF1, and NFY cluster were predicted in the *PRND* proximal promoter by computational analysis. Consequently, a model of PrP/Dpl coregulation is proposed that includes recruitment of BPTF by USF1, phosphorylation of retinoblastoma protein Rb induced by FAC1/BPTF followed by E2F1 derepression leading to activation of the *PRND* promoter. Discovery and characterisation of previously undetected mouse *Bptf* transcript variants suggests tissue-specific modulations of this mechanism, as well as having wide implications for the functions of “BPTF” and “FAC1” not previously suspected.

Overall, the results presented are important for understanding of evolution by gene duplication in general. For the first time, development of gene structure and expression are compared across three evolutionarily distant vertebrate species: amphibians, marsupials, and mammals. The proposed mechanism for PrP and Dpl coregulation by FAC1/BPTF provides novel insights into neuronal death.

Table of Contents

Statement of originality	i
Acknowledgement	ii
Abstract	iii
Table of contents	v
List of figures	ix
List of tables	xii
List of abbreviations	xiii
Chapter 1. Introduction.....	1
1.1 Evolution by gene duplication	1
1.1.1 Mechanisms of duplication and fate of duplicated genes	1
1.1.2 Chimeric transcription	4
1.2 <i>PRNP</i> and <i>PRND</i> and their functions.	5
1.2.1 Prion protein family genes	5
1.2.2 Expression and tissue distribution	8
1.2.3 Normal functions of PrP	9
1.2.4 Dpl and spermatogenesis	11
1.2.5 Role of PrP and Dpl in neurodegeneration	12
1.2.6 Regulation of PrP and Dpl gene expression	22
1.3 Motivation and aims of the thesis	24
Chapter 2. Materials and Methods.....	26
2.1 Homology modelling	26
2.2 RACE.....	26
2.2.1 5' GeneRacer™ kit (Invitrogen).....	26
2.2.2 BD SMART™ RACE cDNA Amplification Kit (Clontech)	27
2.3 Cloning	28
2.4 Site-Directed Mutagenesis.....	29
2.5 Luciferase assay.....	30
2.6 Polysome-binding assay	30
2.7 Reagents, Kits, Solutions, and Cell lines	31
2.8 General molecular biology protocols.....	31
2.8.1 Bacterial Cultures	31

2.8.2	DNA Preparations.....	31
2.8.3	Agarose Gel Electrophoresis	32
2.8.4	Oligonucleotide Primers	32
2.8.5	Automated Sequencing.....	32
2.8.6	Mammalian Cell Culture Methods	33
Chapter 3.	Evolution of <i>PRNP</i> and <i>PRND</i>.....	37
3.1	Introduction	37
3.2	Material and Methods.....	39
3.2.1	Total RNA extraction	39
3.2.2	Amplification of <i>X. laevis PRND</i> ORF fragment.....	40
3.2.3	5' and 3' Rapid amplification of cDNA ends (RACE)	40
3.2.4	Reverse transcription (RT)-PCR and tissue expression analysis.....	41
3.2.5	Database searches	41
3.2.6	Sequence analysis	42
3.2.7	Homology modelling	43
3.2.8	Transposable element content analysis.....	43
3.2.9	Polysome binding assay.....	43
3.3	Results	44
3.3.1	Cloning and characterisation of the <i>X. tropicalis</i> and <i>X. laevis PRNP</i> and <i>PRND</i> cDNAs.....	44
3.3.2	Characterisation of <i>M. domestica PRNP</i> and <i>PRND</i> cDNAs	53
3.3.3	<i>Prnd</i> transcription in mouse	58
3.3.4	Analysis of translation of mouse chimeric <i>Prnd</i> transcripts by polysome binding assay	62
3.4	Discussion.....	65
3.4.1	Expression of chimeric <i>PRND</i>	65
3.4.2	Mechanisms for alternative splicing of the bicistronic transcript.....	67
3.4.3	Conservation of chimeric transcripts throughout evolution	69
3.4.4	New exon creation	71
3.4.5	Evolution of <i>PRND</i> transcriptional regulation	73
3.4.6	3' UTR evolution: acquisition of CPE motif.....	74
3.4.7	Genomic context of <i>PRNP/PRND</i> locus	75
3.4.8	Duplicate of PrP in <i>X. Laevis</i>	76
3.4.9	Evolution of protein function	79
3.5	Conclusion	93

Chapter 4. Regulation of expression of <i>PRNP</i> and <i>PRND</i>	95
4.1 Introduction	95
4.1.1 Antagonistic functions of PrP and Dpl	95
4.1.2 Reported transcription factors for <i>PRNP</i>	96
4.1.3 Reported transcription factors for <i>PRND</i>	96
4.1.4 Chromatin and <i>PRNP/PRND</i> regulation	96
4.2 Materials and Methods	97
4.2.1 Cell lines	97
4.2.2 Total RNA extraction and cDNA synthesis	97
4.2.3 Oligonucleotide primers	97
4.2.4 Cloning of <i>Prnp</i> conserved non-coding DNA region and <i>Prnd</i> promoter	99
4.2.5 Site-directed mutagenesis of transcription factor binding sites	100
4.2.6 Transfection and luciferase reporter assays	100
4.2.7 Bioinformatics analysis	101
4.3 Results and Discussion	101
4.3.1 Bioinformatics prediction of <i>PRNP</i> transcriptional control sites	101
4.3.2 Biological significance of FAC1, MEF2, and MAZ and their relationship with PrP and Dpl	103
4.3.3 Effect of FAC1 on 630bp- <i>Prnp</i> conserved non-coding DNA region	104
4.3.4 <i>PRND</i> promoter contains conserved transcription factor binding sites	109
4.3.5 BPTF activates and E2F1 represses the 650bp- <i>Prnd</i> promoter	110
4.3.6 Tissue-specific correlation of <i>Fac1/Bptf</i> , <i>E2F1</i> , and <i>Prnd</i> expression	112
4.3.7 Expression of <i>FAC1/BPTF</i> and <i>PRND</i> changes throughout brain development	114
4.3.8 Mechanism of PrP and Dpl coregulation	115
4.4 Conclusion	120
Chapter 5. BPTF and its variants	123
5.1 Introduction	123
5.1.1 BPTF and its functions	123
5.2 Materials and methods	126
5.2.1 Total RNA extraction	126
5.2.2 RT-PCR analysis of tissue distribution	126
5.2.3 Amplification of large <i>Bptf</i> fragment	127
5.2.4 <i>Bptf</i> -specific library analysis	128
5.3 Results and Discussion	128

5.3.1	Discovery of new transcript variants and their expression profiles in mouse tissues.....	128
5.3.2	Existing computational and experimental methods cannot detect BPTF tissue-specific alternative splicing.....	132
5.3.3	Novel method for discovery of alternative splicing and tissue specificity of large transcripts.....	134
5.3.4	Analysis of FAC1/BPTF functional regions and their evolutionary conservation.....	139
5.4	Conclusions	148
Chapter 6. Summary and Conclusion		154
References.....		158
Appendices.....		185
1.	Sequences of oligonucleotide primers used in 5' and 3' RACE.....	184
2.	Sequences of oligonucleotide primers used in tissue expression analysis.....	185
3.	Original gel images for tissue and cell expression analysis.....	185
4.	<i>X. tropicalis</i> and <i>M. domestica</i> PrP and Dpl translated protein sequences.....	191
5.	Nucleotide and deduced amino acid sequences of cDNAs for chimeric and non-chimeric transcripts	192
6.	Nucleotide and amino acid sequences of cDNAs for transcripts of <i>M. domestica</i> PRNP, <i>M. domestica</i> PRND	197
7.	Multiple sequence alignment of PrP sequences.....	200
8.	Multiple sequence alignment of Dpl sequences.....	201
9.	Comparison of C-terminal domain of PrP and Dpl among different lineages.....	202
10.	Multiple sequence alignment of eukaryotic BPTF sequences	203

List of Figures

Figure 1-1. The most common mechanisms for origination of new gene copies through gene duplication	2
Figure 1-2. Evolutionary fate of duplicated genes	3
Figure 1-3. Figure 1-3. Phylogenetic tree of prion protein family genes.....	6
Figure 1-4. Protein structural regions of mammalian Shadoo (Sho), prion protein (PrP), and doppel (Dpl).	7
Figure 1-5. Three-dimentional structures of human PrP and Dpl.....	8
Figure 1-6. Malformation of spermatozoa from Dpl-deficient mice	12
Figure 1-7. Three-dimensional structure of human PrP with mapped amino acids associated with TSEs.....	13
Figure 1-8. Schematic representation of binding sites of PrP ligands	17
Figure 1-9. Hypothetical models of Dpl-PrP ^C interaction.....	21
Figure 2-1. Mechanism of BD SMART TM cDNA synthesis.	28
Figure 2-2. Site-directed mutagenesis by PCR amplification with mutagenic primers carrying desired mutation.....	29
Figure 2-3. Polysome-binding assay	31
Figure 3-1. The molecular evolutionary timescale	39
Figure 3-2. Transcripts coding for <i>PRNP</i> and <i>PRND</i> in <i>X. tropicalis</i> and <i>X. laevis</i>	47
Figure 3-3. The 3' ends of <i>PRNP</i> and <i>PRND</i> in <i>X. tropicalis</i> and <i>X. laevis</i>	50
Figure 3-4. Protein sequence comparison for <i>X. tropicalis</i> and <i>X. laevis</i> PrP and Dpl..	51
Figure 3-5. <i>X. laevis</i> tissue expression	53
Figure 3-6. Transcripts coding for <i>PRNP</i> and <i>PRND</i> in <i>M. domestica</i>	55
Figure 3-7. <i>M. domestica</i> <i>PRNP</i> transcripts with alternative coding regions amplified by 3' RACE and 3' nested RACE.....	56
Figure 3-8. Comparison of two <i>M. domestica</i> PrPs with variant ORF	56
Figure 3-9. <i>M. domestica</i> tissue expression for <i>PRNP</i> and <i>PRND</i>	58
Figure 3-10. Chimeric <i>Prnp</i> / <i>Prnd</i> transcripts discovered in <i>Prnp</i> ^{0/0} knockout mice	589
Figure 3-11. Transcripts coding for <i>Prnp</i> and <i>Prnd</i> in mouse	60
Figure 3-12. Mouse <i>Prnd</i> transcript expression	61
Figure 3-13. Polyribosome distribution after centrifugation on a sucrose gradient.....	63
Figure 3-14. Ribosomal distribution of <i>Prnd</i> and <i>Prnp</i> transcripts in adult and postnatal mouse brain is shown by RT-PCR analysis	64

Figure 3-15. Ribosomal distribution of <i>PRND</i> and <i>PRNP</i> transcripts in <i>X. laevis</i> is shown by RT-PCR analysis.	65
Figure 3-16. Schematic representation of bicistronic transcription leading to two different gene products, <i>PRNP</i> and chimeric <i>PRND</i>	68
Figure 3-17. Conserved sequence elements in <i>PRNP</i> 3' UTRs for mammals and marsupials.....	75
Figure 3-18. Comparative genomic context of <i>PRNP</i> and <i>PRND</i> genes in amphibian (<i>X. tropicalis</i>), marsupial (<i>M. domestica</i>), and mammals (mouse and human).....	76
Figure 3-19. Sequence and structural analysis of <i>X. laevis</i> PrP1 and PrP2	79
Figure 3-20. Overall structures of <i>X. laevis</i> , <i>M. domestica</i> and mouse PrP and Dpl.....	81
Figure 3-21. Evolution of Dpl protein sequence.....	83
Figure 3-22. Homology models for <i>X. tropicalis</i> and <i>M. domestica</i>	85
Figure 3-23. Distribution of evolutionarily conserved residues mapped onto the mouse Dpl structure.....	87
Figure 3-24. Evolution of PrP protein sequence	88
Figure 3-25. Homology models for <i>X. tropicalis</i> and <i>M. domestica</i>	89
Figure 3-26. Distribution of evolutionarily conserved residues mapped onto the mouse PrP structure	89
Figure 3-27 Sequence similarity between PrP and Dpl in <i>X. tropicalis</i> , <i>M. domestica</i> , and mouse.....	92
Figure 3-28. Structural conservation of PrP and Dpl among mouse, <i>M. domestica</i> and <i>X. tropicalis</i>	93
Figure 4-1. Transcription factors predicted in the <i>PRNP</i> conserved non-coding DNA region	102
Figure 4-2. Activity of the 630bp- <i>Prnp</i> conserved non-coding DNA region.....	105
Figure 4-3. Transcriptional activity of <i>Prnp</i> conserved non-coding DNA region constructs in GC2 cells	107
Figure 4-4. <i>Prnp</i> promoter activity in response to BPTF transfection has cell-specific response.....	109
Figure 4-5. Genomatix prediction of transcription factor binding sites on the proximal <i>PRND</i> promoter.....	110
Figure 4-6. Cotransfection with <i>Bptf</i> -pcDNA activates and with <i>E2F1</i> -pcDNA suppresses the proximal mouse <i>Prnd</i> promoter	112
Figure 4-7. Tissue-specific distribution of <i>Fac1/Bptf</i> , <i>E2F1</i> , and <i>Prnd</i> transcripts in mouse tissues by RT-PCR.....	113
Figure 4-8. Expression of <i>Fac1/Bptf</i> , <i>Prnd</i> , and <i>Prnp</i> in mouse and human brain and selected cell lines	115

Figure 4-9. The proposed model of how FAC1/BPTF might activate the <i>Prnd</i> promoter	117
Figure 5-1. FAC1 and BPTF protein structure	125
Figure 5-2. NCBI-annotated human <i>BPTF</i> variants (2011).....	130
Figure 5-3. Tissue-specific distribution of <i>Bptf</i> variants.....	131
Figure 5-4. Amplification of <i>Bptf</i> fragments from various mice tissues.....	135
Figure 5-5. Scheme for four PCR reactions performed on bacterial colonies harboring cloned <i>Bptf</i> fragments and examples of amplified products resolved on agarose gels	137
Figure 5-6. Relative abundance of transcripts in mouse tissues from colony analysis.	137
Figure 5-7. Functional domain of human BPTF variants	142
Figure 5-8. Alignment of human protein sequences for BPTF v1 and v2	153

List of Tables

Table 1-1. Tissue expression and distribution of Dpl as reported in the literature	9
Table 1-2. PrP-interacting ligands and their putative binding positions on PrP	15
Table 1-3. Functionally important domains and residues of PrP reported in the literature.....	17
Table 2-1. Public databases and software	33
Table 2-2. Commonly used reagents and their sources	34
Table 2-3. Kits and their suppliers.....	35
Table 2-4. Cell lines used and culture media	35
Table 2-5. Buffers and Solutions	36
Table 3-1. Summary of exon sizes for <i>X. tropicalis</i> and <i>X. laevis</i> <i>PRNP</i> and <i>PRND</i> transcripts.	48
Table 3-2. Exon sizes of <i>M. domestica</i> transcripts	55
Table 3-3. Exon sizes of mouse <i>Prnp</i> and <i>Prnd</i> transcripts.....	60
Table 3-4. Numbers of 5' UTR exons and total lengths of 5' UTRs for <i>M. domestica</i> and mouse <i>PRND</i> transcripts	72
Table 3-5. Transposable element and GC content in <i>X. tropicalis</i> , <i>M. domestica</i> , and mouse <i>PRNP/PRND</i> locus	72
Table 3-6. Matrix of sequence identity between Dpls from different species	83
Table 3-7. Matrix of sequence identity between PrPs from different species	83
Table 3-8. Sequence identity between the C-terminal region of PrP and Dpl in different species, excluding the signal sequences.....	91
Table 4-1. Summary of mouse-specific primers used for promoter amplification and site-directed mutagenesis	98
Table 4-2. Summary of mouse-specific primers used in RT-PCR analysis.....	98
Table 5-1. Summary of primers used in <i>Bptf</i> expression and sequencing analysis	127
Table 5-2. Summary of published literature on length of BPTF v1 and v2 (FAC1). ...	145

List of Abbreviations

AD	Alzheimer's disease
APP	Amyloid precursor protein
bp	base pairs
BPTF	Bromodomain PHD finger transcription factor
cDNA	complementary DNA
ChIP	Chromatin immunoprecipitation
CJD	Creutzfeldt–Jakob disease
C-terminal	Carboxy-terminal
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dpl	Doppel
EDTA	Ethylenediaminetetraacetic acid
EMBOSS	European Molecular Biology Open Software Suite
EST	Expressed sequence tag
FAC1	Fetal Alz-50 clone 1
FBS	Fetal bovine serum
GPI	Glycosylphosphatidylinositol
kb	kilobase pairs
kDa	kilodaltons
LB	Luria-Bertani
LINE	Long interspersed elements
LTR	Long terminal repeat
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
N-terminal	Amino-terminal
NURF	Nucleosome remodelling factor

ORF	Open Reading Frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEST	proline (P), glutamic acid (E), serine (S), and threonine (T)
PrP	Prion protein
PrP ^C	normal cellular form of prion protein
PrP ^{Sc}	scrapie-associated form of prion protein
RACE	Rapid amplification of cDNA ends
Rb	Retinoblastoma tumour suppressor protein
RNA	Ribonucleic acid
rRNA	ribosomal RNA
RT-PCR	Reverse transcriptase polymerase chain reaction
SDS	Sodium dodecyl sulfate
Sho	Shadoo
SINE	Short interspersed elements
SOD	Superoxide dismutase
TAE	Tris-acetate-EDTA
TF	Transcription factor
TFBS	Transcription factor binding site
TRANSFAC	TRANScriptioN FACtor database
TSA	Trichostatin A
TSS	Transcription start site
UTR	Untranslated region
VRC	Vanadyl ribonucleoside complexes

Chapter 1.

Introduction

The focus of this study is to characterise two paralogous genes, *PRNP* and *PRND*, and propose the mechanism for their coregulation. *PRNP* and *PRND*, together with their parental gene, *SPRN*, represent an example of a gene family created by gene duplication.

In the introduction I will briefly overview the mechanisms of gene duplication, fate of duplicated genes and chimeric transcription as one of the ways of survival of newly duplicated genes, and the aspects of prion protein (PrP) and doppel (Dpl) structure and function related to my research project.

1.1 Evolution by gene duplication

Gene duplication plays a key role in evolution by generating genomic material for new proteins to create new functions (Ohno, 1970). A duplicated copy of an existing gene is redundant and, therefore, would accumulate mutations, mostly deleterious. Mutations beneficial for species survival are retained by natural selection, resulting in appearance of new genes with new functions.

Gene duplication frequently creates clusters of paralogous genes referred to as gene families, the most studied cases being *Hox* genes (Pick and Heffer, 2012), globins (Storz et al., 2012), olfactory receptors (Zhang and Firestein, 2002), and major histocompatibility complex molecules (Yeager and Hughes, 1999).

1.1.1 Mechanisms of duplication and fate of duplicated genes

An existing locus may be duplicated by:

- tandem duplication or unequal crossing over occurring during meiosis if homologous chromosomes pair up incorrectly (Shoja and Zhang, 2006) (**Figure 1-1, a**);
- duplicative transposition or segmental duplication (Bailey et al., 2003, Samonte and Eichler, 2002);
- whole genome duplication (Skrabanek and Wolfe, 1998); or

- RNA-mediated retro-transposition (Brosius, 2003) (**Figure 1-1, b**).

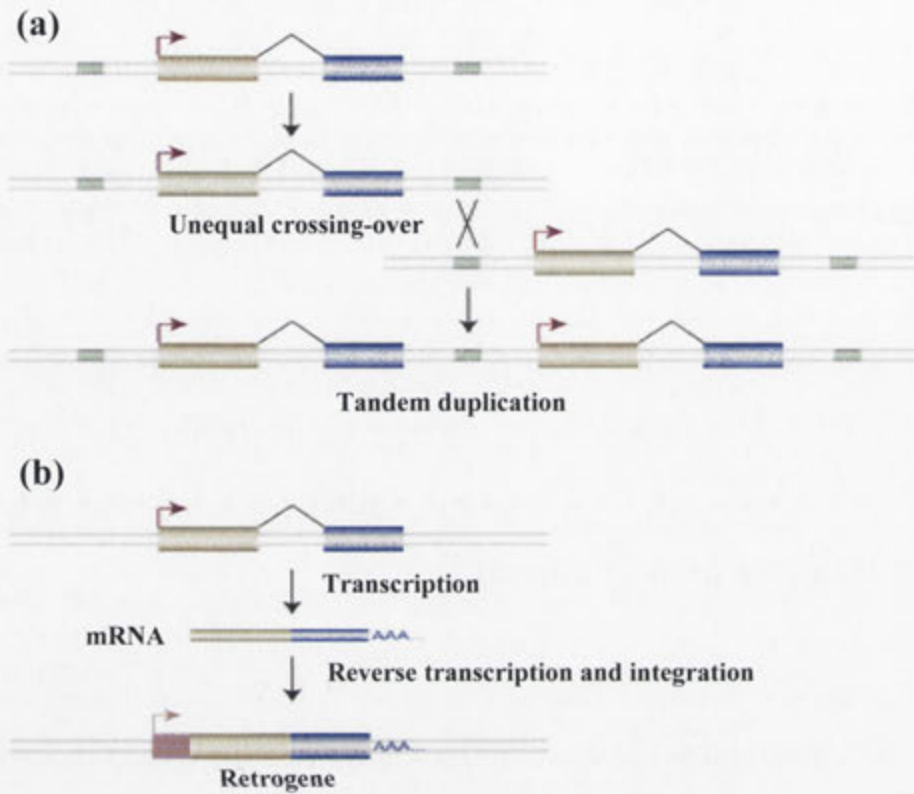


Figure 1-1. The most common mechanisms for origination of new gene copies through gene duplication. (a) DNA-based duplication. Tandem duplication may occur via unequal crossing-over that is mediated by transposable elements (light green). **(b)** RNA-based duplication or retroposition. Reverse transcription of mRNA and its integration into the genome results in retroposed gene copies. Exons are marked as gold and blue boxes; black connecting lines symbolise exon splicing; red right-angled arrows show transcriptional start sites; pink box shows additionally transcribed flanking sequence at the insertion site of the retrogene. Source: (Kaessmann, 2010)

Based on their ability to survive, the fate of duplicated genes can be (**Figure 1-2**):

- Pseudogenisation, the most common fate of duplicated genes, when the duplicated copy becomes nonfunctional by mutations (Lynch and Conery, 2000).
- Conservation of function, when paralogs retain very similar sequences and function under strong purifying selection (Nei and Rooney, 2005). This is the fate of strongly expressed genes, such as histones (Piontkivska et al., 2002), ubiquitins (Nei et al., 2000), and rRNAs (Rooney and Ward, 2005).

- Subfunctionalisation, when each duplicate performs part of the function of their parental gene (Hughes, 1994).
- Neofunctionalisation, when a new function emerges (Laukaitis et al., 2005, Deng et al., 2010, Lynch, 2007).

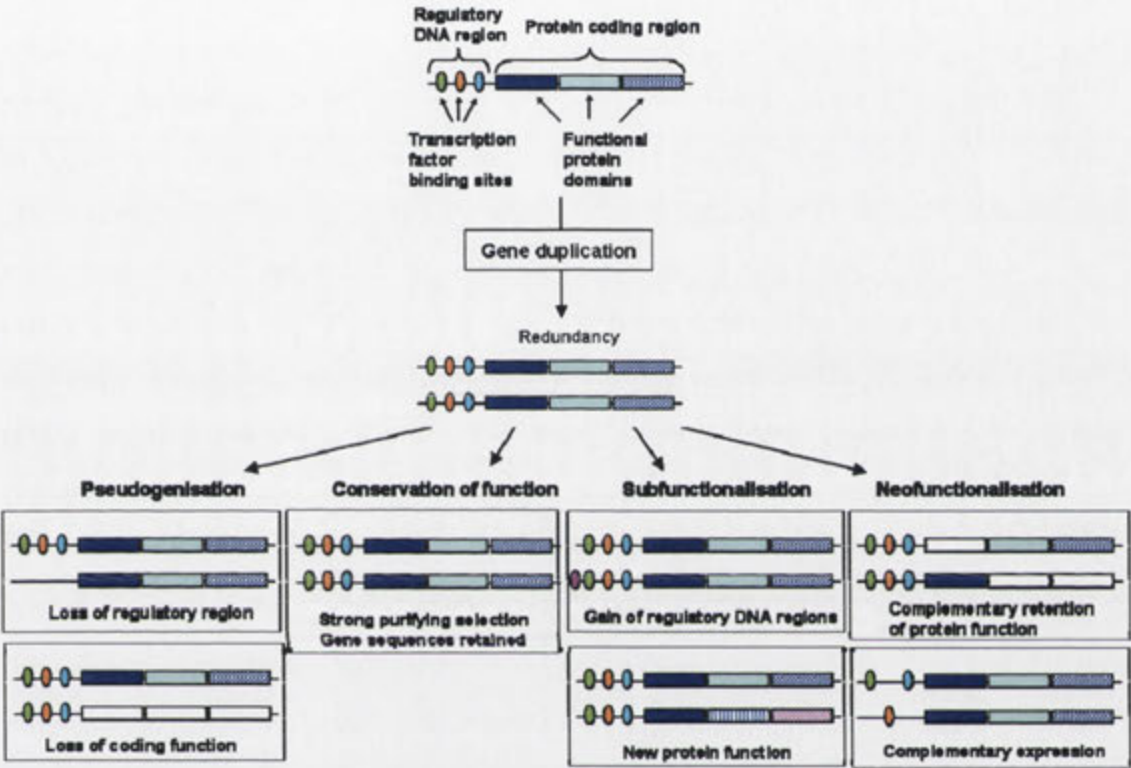


Figure 1-2. Evolutionary fate of duplicated genes.

Small-scale DNA-mediated gene duplications (tandem and segmental duplications) may include not only exon-intron regions of the gene, but the associated regulatory DNA sequences and transcription factor binding sites, resulting in functional duplicates.

Subfunctionalisation and neofunctionalisation result in changes in gene expression (Gu et al., 2002). Loss of some regulatory DNA elements and conservation of others will result in differential tissue expression of duplicates and could increase the probability of their conservation (Force et al., 1999). Therefore, evolution of regulatory DNA elements and establishment of regulatory networks is no less important for new function

development and survival of duplicated genes than differentiation of protein sequence (Ohta, 2003). Gene regulatory elements might evolve from inserted retroelements (long interspersed elements (LINEs), short interspersed elements (SINEs), and LTR retrotransposons (Ball et al., 2004, Muotri et al., 2005, Nigumann et al., 2002, Plant et al., 2001, Wheelan et al., 2005).

1.1.2 Chimeric transcription

Retroposed duplicates, unlike DNA-based events, lack the promoter and regulatory DNA sequences associated with the parental gene and thus immediately become pseudogenes (Mighell et al., 2000). However, there are multiple cases of expression of retroposed duplicates (Vinckenbosch et al., 2006, Sakai et al., 2007, Svensson et al., 2006). Duplicates without regulatory DNA elements can “hitchhike” the promoter from a neighbouring gene, and sometimes also pick its exons to create chimeric genes (Vinckenbosch et al., 2006). Examples of such chimeric ORFs made of exons from both the host and the retrogene are the *jingwei* gene in the fruit fly (Long and Langley, 1993) and the *TRIMCyp* gene in the owl monkey (Sayah et al., 2004). It is estimated that about 4-5% of the tandem gene pairs in the human genome can be transcribed as chimeric proteins (Lee et al., 2006b, Parra et al., 2006).

The outcome of polycistronic pre-mRNA expression is not necessarily chimeric protein. Consecutive alternative splicing might result in monocistronic mRNAs, as was first described in *C. elegans* by Blumenthal (Blumenthal, 1998). In *C. elegans*, ~25% of the genes are estimated to be contained in polycistronic units (Zorio et al., 1994). Such units are rare in mammals, one example being bicistronic *SNURF-SNRPN* transcript (Gray et al., 1999). Interestingly, genes expressed from the common promoter can share non-coding exons which remains present in both mRNAs after bicistronic transcription and alternative splicing, as in case of the *C. elegans unc-17/cha-1* genes (Alfonso et al., 1994).

In Chapter 3 I will describe how survival and evolution of the *PRNP* and *PRND* tandem gene pair relied on the ability of the downstream gene to share the promoter and resulted in chimeric transcription.

1.2 *PRNP* and *PRND* and their functions.

1.2.1 Prion protein family genes

The prion protein family includes three genes:

- *SPRN*, the gene coding for Shadoo (Sho) protein, is considered the oldest member of the family (Premzl et al., 2004).
- *PRNP*, the prion protein (PrP) gene (Oesch et al., 1985, Lee et al., 1998), is suggested as being the result of an ancient duplication of *SPRN* (Premzl et al., 2004).
- *PRND*, that encodes a protein called Doppel (Dpl), is a downstream duplication of *PRNP* (Moore et al., 1999). Curiously, it was discovered during attempts to elucidate PrP function as an explanation for why some strains of *Prnp* knockout mice develop ataxia and others do not.

In fish, PrP-like genes *PrP1* and *PrP2*, which are more related to Sho than to PrP, were first reported by Premzl et al. (2004) and further work was then published and reviewed by (Strumbo et al., 2006). It is unclear if the PrP duplication in fish/mammal ancestor evolved into PrP1 and PrP2 in fish, and PrP and Dpl in amphibians, reptiles, and mammals. The phylogenetic relationship between Sho, PrP, and Dpl, which includes 4 clusters (as the tetrapod Shos and fish PrP1,2s grouped separately) of then available sequences across vertebrate species, is shown in **Figure 1-3**.

All proteins are encoded by a single exon. Despite highly diverged sequence among the three proteins, many structural features are conserved (**Figure 1-4**): signal sequences, basic regions, and N-glycan attachment sites (CHO). Disulphide bridges (S-S) are present in PrP and Dpl, but not Sho. Repeats and the hydrophobic region are conserved between PrP and Sho, but Dpl lacks both of these structural features (**Figure 1-4**). The absence of the hydrophobic region in Dpl is very significant. It is suggested, that the hydrophobic domain determines selective polarized sorting between PrP (basolateral side of cells) and Dpl (apical surface) (Uelhoff et al., 2005) and is essential for the conformational change to the “scrapie” form of PrP associated with prion disease (Prusiner, 1998). PrP, Dpl and Sho are all extracellular, GPI-anchored proteins (Oesch et al., 1985, Premzl et al., 2003, Moore et al., 1999).

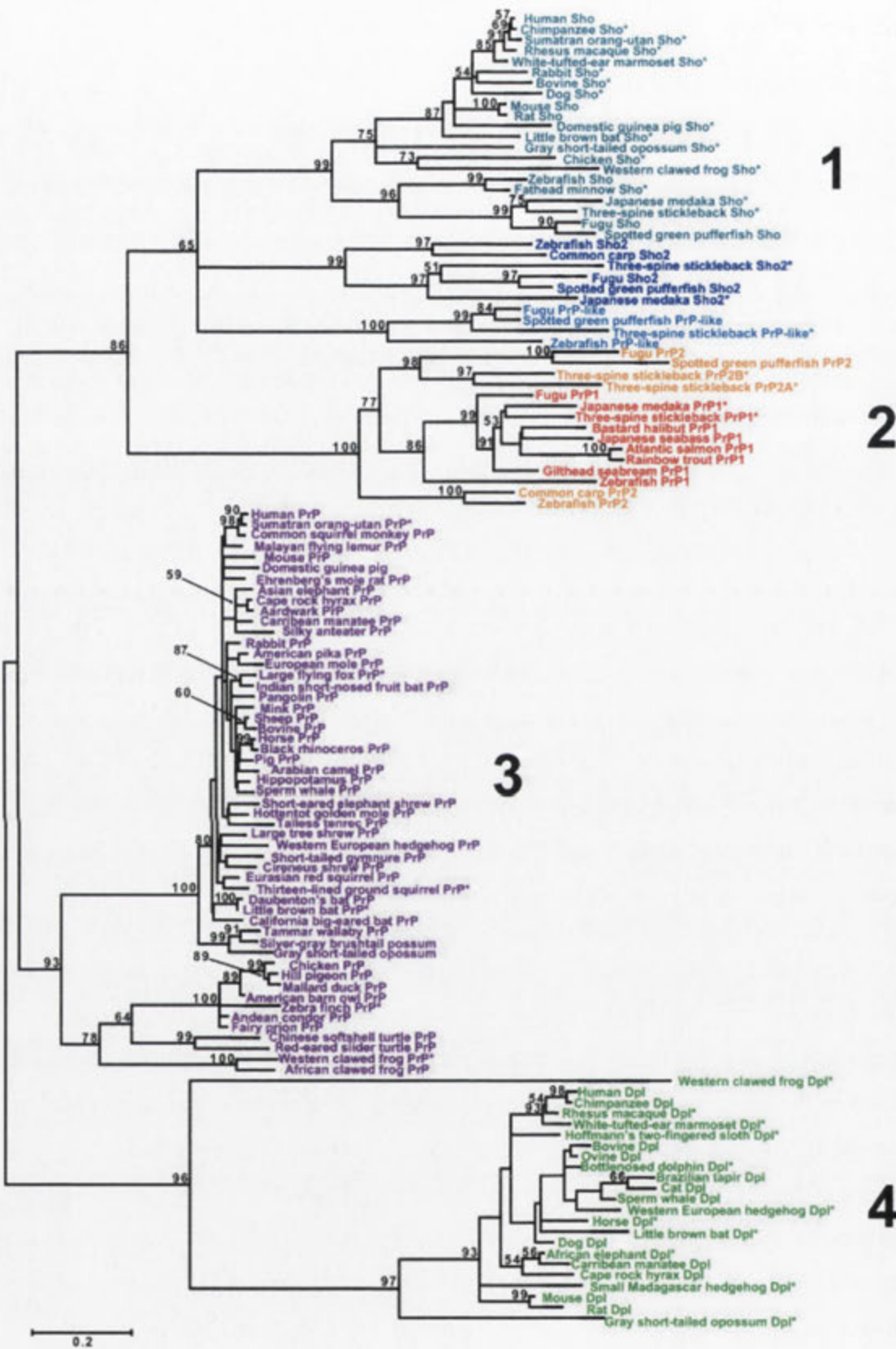


Figure 1-3. Phylogenetic tree of prion protein family genes. Numbers refer to four major clusters: (1) Shos, Sho2s and PrP-like; (2) fish PrP1s and PrP2s; (3) tetrapod PrPs; (4) Dpls. The bootstrap values >50 are depicted. The horizontal bar shows 0.2 amino acid substitutions per site. Source: Premzl & Gamulin, 2007.

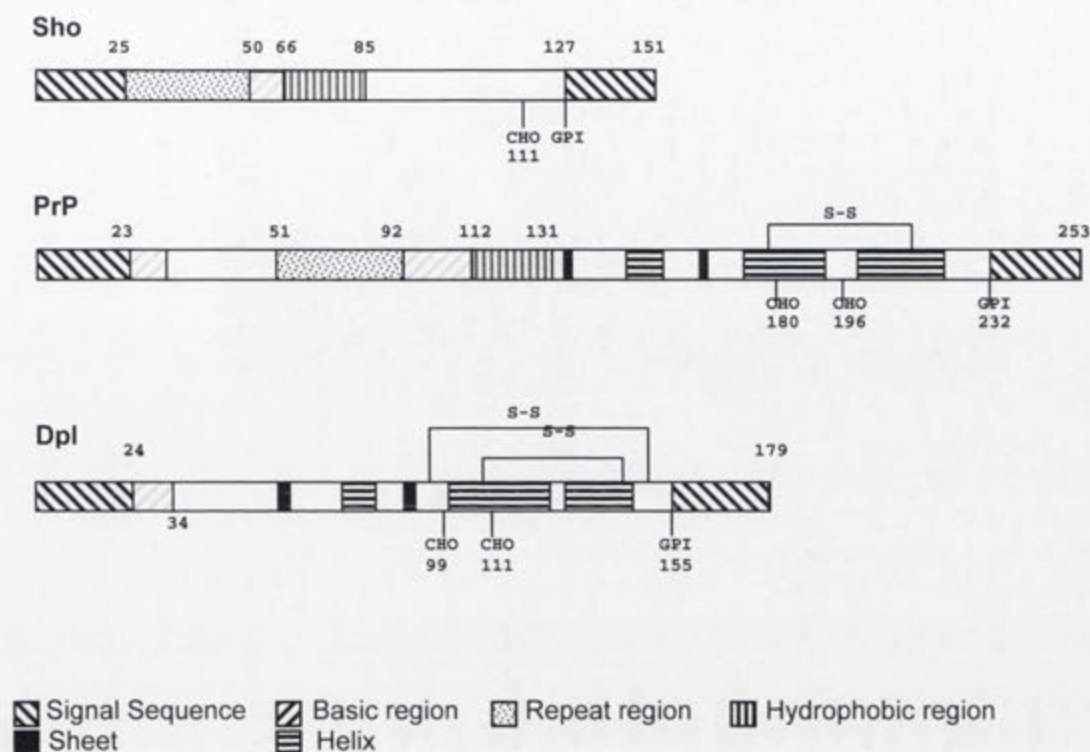


Figure 1-4. Protein structural regions of mammalian Shadoo (Sho), prion protein (PrP), and doppel (Dpl); numbers refer to human Sho and PrP (based on Premzl et al., 2003) and mouse Dpl (based on Moore et al., 1999). These structures are presented with more details based on their NMR analysis in Chapter 3 (**Figure 3-20**).

The resolved NMR structures revealed that Dpl folding is very similar to that of PrP, despite limited sequence homology, being characterised by three α -helices, and two β -strands (**Figure 1-5**). The Dpl structure differs by a kink in the second helix and a different orientation of the β -sheet. The structure is also stabilised by an additional disulfide bond (Luhrs et al., 2003).

The evolutionarily conserved positions involved in PrP and Dpl fold maintenance will be discussed in detail in Chapter 3.

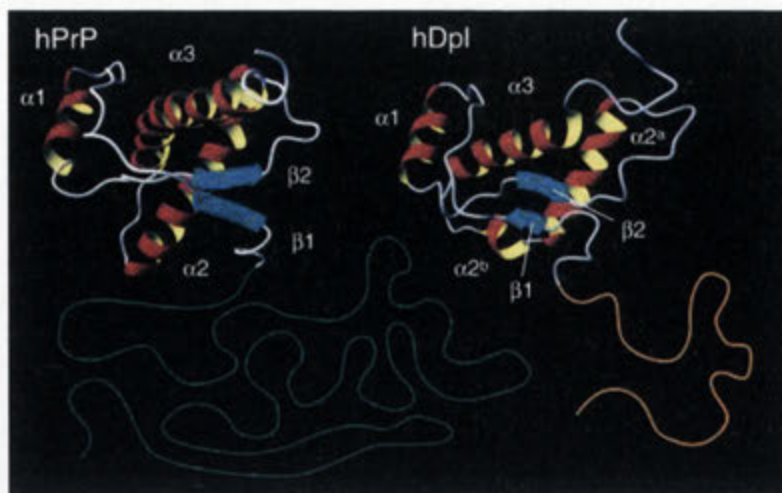


Figure 1-5. Three-dimensional structures of human PrP and Dpl. Helices $\alpha 1$, $\alpha 2$, and $\alpha 3$ are red and yellow, β -strands are cyan, nonregular secondary structures are grey, and the flexibly disordered segments of residues 23-124 of hPrP and 24-50 of hDpl are green and orange, respectively. Source: (Riek and Luhrs, 2003).

1.2.2 Expression and tissue distribution

PrP has wide tissue distribution with the highest amount in brain (Bendheim et al., 1992, Makrinou et al., 2002). In brain, PrP mRNA is detectable in neurones of the hippocampus, cortex, thalamus, cerebellum and medulla (McLennan et al., 2001). Considerable amounts of PrP are found in immune system cells (Dodelet and Cashman, 1998, Kubosaki et al., 2001). PrP is reported in heart, lung, pancreas, thymus, spleen, testis, kidney, gut, skin, stomach, ovary, skeletal muscle, and liver of adult mammals and in developing embryos (Manson et al., 1992, Horiuchi et al., 1995, Fournier et al., 1998, Moudjou et al., 2001, Ning et al., 2005). However, PrP's ubiquitous distribution can be explained by specialised expression at high levels only in discrete subpopulations of cells, especially those of the "neuroimmune network" (Ford et al., 2002b). Dpl expressed mainly in testes (Makrinou et al., 2002), although broader tissue distribution has been reported (**Table 1-1**). Similarly to PrP, presence of Dpl in other tissues can be attributed to its expression in endothelial cells, peripheral nerves and scattered lymphocytes (Rondena et al., 2005). Analysis of expression in mouse testes reveals that *Prnp* (Ford et al., 2002b) and *Prnd* (Peoc'h et al., 2000a) are coexpressed in Sertoli cells, spermatocytes, spermatids and epididymis. However, unlike Dpl knockout mice, PrP knockouts are fertile (Behrens et al., 2002).

During embryogenesis, *Prnp* is induced as early as E6.5 days in extra-embryonic tissues (placenta, umbilical cord and mesodermal layer of yolk sac) (Manson et al., 1992), between E8.5 and E9 days in embryonic tissues during transition from anaerobic to aerobic metabolism (Miele et al., 2003), and from E13.5 in developing brain and spinal cord (Manson et al., 1992). *Prnd* is widely expressed in foetal tissues (Makrinou et al., 2002).

Table 1-1. Tissue expression and distribution of Dpl as reported in the literature.

Tissue origin	Tissue (Technique)	Reference
Human	Testis (Northern blot)	(Makrinou et al., 2002)
Human fetal 11.5-16.5 weeks	Testis, intestine, limbs, kidney, brain (RT-PCR)	(Makrinou et al., 2002)
Mouse	Testis and heart (Northern blot); brain (RT-PCR); mammary gland, spleen (EST database)	(Moore et al., 1999)
Ovine, Bovine	Testis, epididymis (Northern blot); testis, spleen, ovary, heart, mammary gland, kidney, lung (RT-PCR)	(Tranulis et al., 2001)
Bovine	Testis (Western blot); endothelial cells, peripheral nerves and scattered lymphocytes in intestine, liver, spleen, lymph nodes, kidney, ovary, heart, lung, diaphragm (immunohistochemistry)	(Rondena et al., 2005)
Human	Testis, spermatozoa, seminal plasma (Western blot)	(Peoc'h et al., 2002)
Mouse, human	Testis, brain (RT-PCR)	(Peoc'h et al., 2003)
Mouse, 9 weeks	Testis, heart, spleen, muscle (Northern blot)	(Li et al., 2000b)
Neonatal mouse	Brain, spleen, gut (Northern blot)	(Li et al., 2000b)

1.2.3 Normal functions of PrP

The normal role of PrP remains unclear despite extensive research over the last 30 years and numerous functions proposed (Aguzzi and Polymenidou, 2004, Lasmezas, 2003). PrP gene knock out experiments on this widely expressed gene did not show obvious phenotype, although mice were resistant to scrapie infection (Bueler et al., 1992, Bueler et al., 1993). However, some subtle phenotypic changes have been reported, such as alterations to circadian rhythms and sleep disturbance (Tobler et al., 1996, Sanchez-

Alavez et al., 2007), to short- and long-term memory performance (Coitinho et al., 2003), altered fear-induced behaviour (Lobao-Soares et al., 2008) and sensitivity to oxidative stress (Brown et al., 2002). A recent microarray study revealed biological pathways altered in *Prnp*^{0/0} mice, such as cell signalling, cell-cell communication and transduction processes, calcium homeostasis, nervous system development, synaptic transmission, and cell adhesion (Benvegnu et al., 2011).

The proposed PrP functions are:

1. **Copper binding and antioxidant activity.** Mammalian PrP binds copper via histidine residues present in the proximal octarepeats (Brown et al., 1997). It was suggested this property of PrP is used for alteration of synaptic activity by modulating presynaptic copper concentration (Herms et al., 1999), for uptake of copper ions from the extracellular milieu, which also stimulates endocytosis of PrP from the cell surface (Pauly and Harris, 1998), and for activation of the copper-dependent antioxidant enzyme, superoxide dismutase (Brown and Besinger, 1998).
2. **Regulation of intracellular calcium concentration** (Herms et al., 1997, Herms et al., 2001, Krebs et al., 2007).
3. **Learning and memory.** The requirement of PrP for long-term memory formation was established from knockout mice experiments. *PrP*^{-/-} mice performed worse than wild-type mice in water and labyrinth learning tests (Nishida et al., 1997). Molecular evidence has been provided by PrP enrichment in the dentate gyrus of the hippocampus, which is involved in synaptic plasticity and memory formation (Bendheim et al., 1992). Expression on the surface of elongating axons suggests a PrP role in axon growth and polarisation (Sales et al., 2002); this is supported by experiments with cultured hippocampal neurons (Kanaani et al., 2005).
4. **Activation of lymphocytes.** PrP is expressed at high levels in T cells, B cells, monocytes and dendritic cells (Li et al., 2001). Activation of T lymphocytes increases the expression of PrP on the T-cell surface (Li et al., 2001, Bainbridge and Walker, 2005, Mabbott et al., 1997).
5. **Signal transduction.** PrP is involved in numerous signal transduction pathways, including cAMP/protein kinaseA, mitogen-activated protein (MAP) kinase,

phosphatidylinositol 3-kinase/Akt pathways, and non-receptor tyrosine kinases (reviewed in (Linden et al., 2008)).

6. **Cell-cell interaction.** PrP mediates binding of molecules in the extracellular matrix and on the cell membrane, such as the laminin receptor precursor (Rieger et al., 1997), laminin itself (Graner et al., 2000), and neural cell-adhesion molecule (N-CAM) (Schmitt-Ulms et al., 2001).
7. **Antiapoptotic properties,** discussed extensively in **section 1.2.5.4** below.
8. **Regulation of embryonic stem cell pluripotency and differentiation** (Miranda et al., 2011).

1.2.4 Dpl and spermatogenesis

One of the most studied functions of Dpl is its involvement in spermatogenesis; its main expression is in adult testes, as referenced in **Table 1-1**. Dpl is expressed in the seminiferous tubules, Sertoli cells, mature spermatozoa and their flagella, and seminal plasma (Peoc'h et al., 2002). Dpl is involved in early testes development, and later stages of spermatogenesis, in particular the final remodelling and maturation of elongated spermatids (Espenes et al., 2006).

Absence of Dpl in *Prnd*^{neo/neo} knockout mice results in male sterility due to malformed sperm, incapacity to perform the acrosome reaction in fertilisation, and poor sperm motility but otherwise does not affect embryonic or postnatal development or fertility of female mice (Behrens et al., 2002) (**Figure 1-6**). However, *Prnd* null mouse created by another research group, although also male-sterile, showed normal sperm concentration, motility, and morphology, and ability to fertilise *in vitro* albeit with unviable embryos (Paisley et al., 2004).

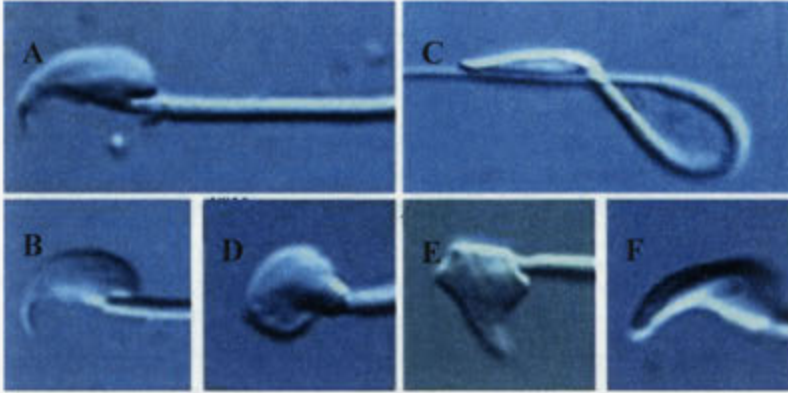


Figure 1-6. Malformation of spermatozoa from Dpl-deficient mice. Sperm morphology in wild-type (A, B) and *Prnd*^{neo/neo} knockout mice (C-F). Sperm in knockout mice has visible structural abnormalities of flagellum and malformation of sperm head (Behrens et al., 2002).

1.2.5 Role of PrP and Dpl in neurodegeneration

1.2.5.1 Prion disease

Prion diseases owe their name to proteinaceous infectious particle or “prion”, which seems to be the only pathological agent without any known organism or virus involved in infection (Prusiner, 1982). Prion diseases are a group of fatal neurodegenerative disorders affecting several species, for example, Creutzfeldt-Jakob disease (CJD), Gerstmann–Sträussler–Scheinker syndrome (GSS) and Fatal Familial Insomnia (FFI) in humans, bovine spongiform encephalopathy in cattle and scrapie in sheep. These diseases are associated with structural conversion of the normal alpha-helical cellular prion protein (PrP^{C}) into a beta-sheet-rich “scrapie” conformation (PrP^{Sc}) and are characterised by degeneration of brain tissue and amyloid deposit (Prusiner, 1998). Several PrP mutations have been linked to prion misfolding (**Figure 1-7**). The conversion is possibly assisted by an elusive ligand protein, “protein X” (Telling et al., 1995, Prusiner, 1998). It has been proposed that misfolded peptides form ion channels and as a result dysregulate Ca^{2+} signalling and depolarise cell membrane causing neuronal cell death (Lin et al., 1997a). In humans, approximately 10% of CJD has genetic causes (Prusiner and Scott, 1997) but 75% of all cases are spontaneous transformation of PrP^{C} to PrP^{Sc} (Kubler et al., 2003). There is no treatment for prion diseases and the diseases progresses inevitably to death (Knight and Will, 2004).

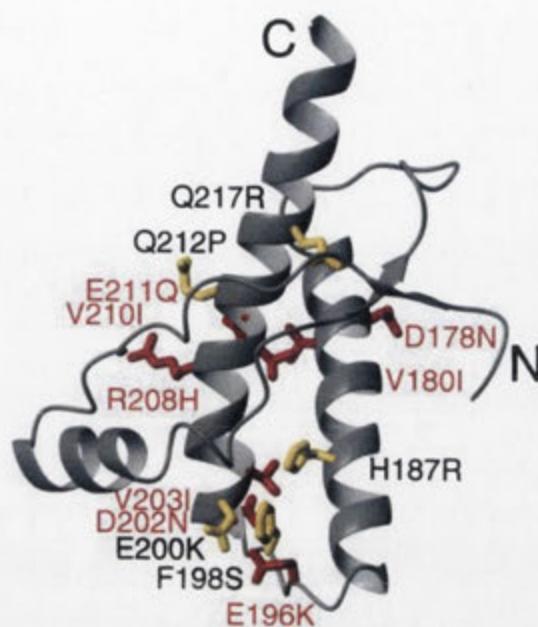


Figure 1-7. Three-dimensional structure of human PrP with mapped amino acids associated with TSEs. The mutated amino acids associated with CJD are labelled in red; with GSS are in black. Source: (Riek and Luhrs, 2003).

1.2.5.2 PrP and Dpl in Alzheimer's disease

Alzheimer's disease (AD) is the most common cause of mental disorders associated with ageing. It is associated with amyloid deposits, derived from cleavage of amyloid precursor protein (APP) into the amyloid-beta ($A\beta$) peptide of 40-42 amino acid residues, forming senile plaques with subsequent neuronal loss affecting memory, speech, and recognition (Hardy and Selkoe, 2002). There is an intriguing link between AD, PrP and Dpl, as expression of both proteins is altered in AD.

Abnormal PrP deposits were observed in AD brain as diffuse plaques, neuritic plaques, and amyloid cores of senile plaques, colocalising with β -amyloid (Ferrer et al., 2001). It was suggested that PrP expression is altered as a response to cellular stress (Kovacs et al., 2002, Voigtlander et al., 2001). However, overexpression of Syrian hamster PrP in transgenic mice carrying mutant human APP resulted in a 35% increase of $A\beta$ plaques suggesting that PrP participates in formation of $A\beta$ plaques (Schwarze-Eicker et al.,

2005). Nevertheless, alternative evidence, based on overexpression or depletion of PrP in neuronal cell lines and significant increase of A β in PrP^{-/-} mice, support a PrP-protective role during AD through PrP-mediated inhibition of the cleavage of APP by β -secretase and reduction of A β formation (Parkin et al., 2007). Surprisingly, although PrP inhibits A β formation, it might mediate A β -induced impairment of synaptic plasticity by direct interaction with soluble A β at synapses (Lauren et al., 2009). Pathology of this interaction is explained by Fyn kinase activation which leads to phosphorylation of NMDA (N-methyl-D-aspartate) glutamate receptors and alters their localisation (Um et al., 2012).

Dpl expression in AD was detected around amyloid deposits of neuritic plaques and, as was suggested, might trigger the astrocytic and microglial response (Ferrer et al., 2004). Surprisingly, considering increasing interest to PrP-Dpl antagonism and growing evidence of PrP importance in AD, no further research has been done to decipher the role of Dpl in AD.

1.2.5.3 Putative ligands and functional domains of the PrP

Attempts to identify a “protein X” which contributes to pathological conversion from normal PrP^C to PrP^{Sc} (section 1.2.5.1) have resulted in generation of vast amounts of data on putative ligands of the prion protein. Experimental screening techniques used for ligand detection (two-hybrid screen, crosslinking, coimmunoprecipitation, binding assays, ELISA (enzyme-linked immunosorbent assay), and mass-spectrometry) expose only approximate regions of binding and often do not reveal the amino acid residues most important for binding. Nevertheless, comparative analysis of binding sites might provide insights into PrP and Dpl evolution. Ligands of PrP with defined binding domains are summarised in **Table 1-2 and Figure 1-8**. The most important region for PrP function is formed by residues 106-126 (human). It plays the central role in prion disease neurotoxicity (Forloni et al., 1993). The isolated peptide forms ion-permeable channels in cell membranes causing dysregulation in ionic homeostasis, specifically, of calcium (Lin et al., 1997a). Deletions of this region (**Table 1-3**) cause neurodegenerative disorders, suggesting the existence of a putative ligand that mediates neuroprotection (Shmerling et al., 1998, Li et al., 2007, Baumann et al., 2007).

Dpl has structural similarities with PrP in its C-terminal end and truncation of the N-terminal tail of PrP results in Dpl-like neurotoxicity (Flechsigs et al., 2003, Radovanovic et al., 2005). The fusion protein of the N-terminal octarepeat region and half of the hydrophobic region of PrP to Dpl showed antiapoptotic function (Lee et al., 2006a). This indicates that common ligands might exist; however, almost no research has been done on putative Dpl ligands or common PrP-Dpl ligands. Two identified Dpl-binding partners, plasma metalloproteinase inhibitors alpha-1-inhibitor-3 (α_1I_3) and alpha-2-macroglobulin (α_2M), do not interact with PrP (Benvegnu et al., 2009). The only common interacting ligand discovered so far is laminin receptor (Yin et al., 2004).

Structural comparison of PrP and Dpl provided additional information to deduce functional properties of regions of secondary structure. The main aim of reported functional studies has been to solve the enigma of PrP^C misfolding to disease-associated conformation PrP^{Sc}. For example, substitution of residues at positions 167, 171, 214, or 218 prevented PrP^{Sc} formation (Kaneko et al., 1997). **Table 1-3** overviews findings in key functional domains and residues of PrP.

Table 1-2. PrP-interacting ligands and their putative binding positions on PrP. Amino acid residue numbers refer to mouse PrP^C. Data obtained from human or other animal PrP analysis are transposed to homologous mouse structure (based on review Linden et al., 2008)

Ligand	Description	Binding site on PrP	Reference
LRP1	Lipoprotein receptor-related protein 1, scavenger receptor. Endocytic trafficking of PrP ^C .	23-107	(Taylor and Hooper, 2007) (Parkyn et al., 2008)
LRP/LR	Laminin receptor precursor/Laminin receptor. Subcellular trafficking.	143-178	(Gauczynski et al., 2001, Hundt et al., 2001, Rieger et al., 1997)
LPR	Laminin receptor, residues 100-220. Interaction with Dpl, PrP-Dpl antagonism.	26-100	(Yin et al., 2004)
Laminin	Extracellular matrix component. Interaction promotes neuritogenesis.	173-182	(Coitinho et al., 2006, Graner et al., 2000)
Vitronectin	Extracellular matrix component.	105-119	(Hajj et al., 2007)
Heparin/heparin sulfate	Glucosaminoglycans.	23-52, 53-92, 109-127	(Horonchik et al., 2005, Pan et al., 2002, Warner et al., 2002)
14-3-3	Intracellular scaffolding protein.	23-137	(Satoh et al., 2005)
Cu ²⁺	Binds octapeptide repeat, histidine residues.	58-110	(Brown et al., 1997)

Hop/STI-1 Stress inducible protein1	Cell surface ligand, co-chaperone, roles in cell metabolism, secreted by tumor cell lines and primary astrocyte cultures. Interaction triggers endocytosis of the PrP and cellular signalling. Triggers neuroprotection. Enhances PrP properties in short-term memory formation and long-term memory consolidation.	113-128	(Martins et al., 1997, Zanata et al., 2002, Lopes et al., 2005, Coitinho et al., 2007)
N-CAM1	Cell adhesion molecule, mediates cell-cell interaction. Recruited by PrP to lipid rafts, activates Fyn kinase. Interaction important for neurite outgrowth.	143-153	(Santucci et al., 2005, Schmitt-Ulms et al., 2001)
Hsp60	Chaperone, apoptosis, heat shock.	179-209	(Edenhofer et al., 1996)
Pint1	Prion protein interactor 1, exonuclease motif.	90-231	(Spielhaupter and Schatzl, 2001)
CK2	Casein kinase 2. Induction of CK2 phosphotransferase activity.	93-228	(Meggio et al., 2000)
Grb2	Growth factor receptor-bound protein, adaptor protein for tyrosine kinase receptors. Interacts in intracellular vesicles, possibly signal transduction.	23-231	(Spielhaupter and Schatzl, 2001)
Synapsin 1b	Synaptic vesicle release regulator. Possibly regulation of cell-cell contact and extracellular signalling.	23-231	(Spielhaupter and Schatzl, 2001)
APLP1	Amyloid precursor-like protein.	23-231	(Yehiely et al., 1997)
Nrf2	Nuclear factor E2-related factor-2, transcription factor.	23-231	(Yehiely et al., 1997)
GASP	G protein-coupled receptor-associated sorting protein.	23-231	(Yehiely et al., 1997)
hnRNP A2/B1	Heterogeneous nuclear ribonucleoprotein, RNA-binding protein.	unknown	(Strom et al., 2006)
AldC	Aldolase C/zebrin.	unknown	(Strom et al., 2006)
Bcl-2	Antiapoptotic protein. PrP could sequester Bcl-2 from its intracellular organelle pools and contribute to apoptosis.	72-254	(Kurschner and Morgan, 1995, Kurschner and Morgan, 1996)
A β	Amyloid- β , accumulates in Alzheimer's disease.	95-105	(Lauren et al., 2009, Barry et al.)

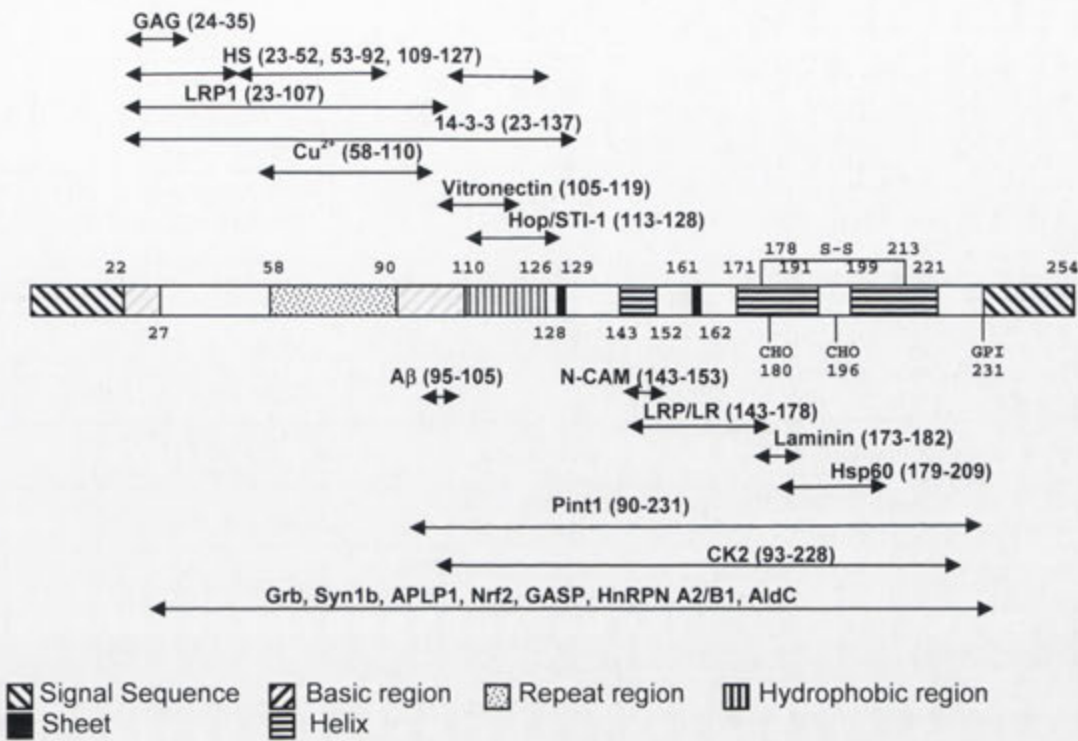


Figure 1-8. Schematic representation of binding sites of PrP ligands (modified from Linden et al., 2008). PrP structure with major domain shaded corresponds to mouse sequence. Ligand binding data based on research in other species (human, hamster, or bovine) were transposed to homologous mouse sequence. Numbers in parentheses next to each binding partner and arrows refer to PrP amino acid residues and PrP binding domain. GAG, glycosaminoglycans; HS, heparan sulfate; LRP1, low-density lipoprotein receptor-related protein; LRP, laminin receptor precursor protein; LR, laminin receptor; Pint1, prion protein interactor 1; CK2, casein kinase 2; Grb, growth factor receptor-bound protein; Syn1b, synapsin Ib; APLP1, amyloid precursor-like protein 1; Nrf2, nuclear factor E2-related factor-2; GASP, G protein-coupled receptor-associated sorting protein; Fbx, F-box only; HnRNP, heterogeneous nuclear ribonucleoprotein; AldC, aldolase C/zebrin; Aβ, amyloid-beta.

Table 1-3. Functionally important domains and residues of PrP reported in the literature. Amino acid residue numbers refer to mouse or human (when marked) PrP^C.

Description	Residue	Reference
Signalling peptide for reinternalisation of PrP.	23-42	(Nunziante et al., 2003)
N-terminal octapeptide repeat domain, contributes to extracellular copper ion binding.	65-95	(Viles et al., 1999, Jackson et al., 2001)
SH3-binding motif, binds C-terminal SH3 domain of Grb2.	100-110	(Spielhaupter and Schatzl, 2001, Lysek and Wuthrich, 2004)
Deletion results severe illness, mice died within 1 month after birth.	105-125	(Li et al., 2007)

Deletion results in ataxia and apoptosis.	32-121, 32-134	(Shmerling et al., 1998)
Unstructured central domain (charge cluster 95-110 and hydrophobic 112-134).	95-134	(Baumann et al., 2007, Li et al., 2007)
Deletion results in demyelination and axon loss in spinal cord and cerebellar white matter.	94-144	(Baumann et al., 2007, Radovanovic et al., 2005)
Globular C-terminal, GPI attachment.	134-231	(Riek et al., 1996)
Important role in intermolecular interactions, including potentially key determinant of prion protein aggregation.	beta2-alpha2 loop (166-175, human)	(Billeter et al., 1997)
Associated with species barrier differences.	beta2-alpha2 loop	(Bett et al., 2012)
Major role in PrP conversion.	127-157	(Cohen and Prusiner, 1998)
Surface regions that may be involved in initial binding of PrP ^C to PrP ^{Sc} .	156-175, 202-211, 246-264	(Horiuchi et al., 2000)
Predicted transmembrane helix.	113-128	(Hegde et al., 1998, Stewart et al., 2001)
Substitution of a basic residue at these positions prevented formation of PrP ^{Sc} .	167, 171, 214, 218 (human)	(Kaneko et al., 1997)
Protein X binding.	Q167, Q171, T214, Q218	(Kaneko et al., 1997)
Inhibition of prion replication in transgenic mice when substituted.	Q167R or Q218K	(Perrier et al., 2002)
Amino acid substitution associated with AD, important in stabilising conformation.	M129V (human)	(Berr et al., 2003)
Amino acid substitution associated with Creutzfeldt-Jakob disease.	D178N, V129, E196K, V203I, E211Q (human)	(Dossena et al., 2008) (Peoc'h et al., 2000b)
Cu ²⁺ coordination.	H96, H111 (human)	(Viles et al., 1999)
Constitutive endocytosis of PrP and cellular trafficking in neuronal cells.	N-terminal, positively charged, KKRPKP	(Sunnyach et al., 2003)
Amino acid substitution associated with neurodegeneration.	A117V	(Hegde et al., 1998)
Spontaneous neurodegeneration.	P101L	(Telling et al., 1996)
Variable susceptibility for prions	P102L, E200K	(Asante et al., 2009)
Results in high structural order of the beta2-alpha2 loop. Caused transmissible neurodegeneration.	S170N, N174T	(Sigurdson et al., 2009)
Progressive encephalopathy characterised by abundant PrP plaque formation, spongiform change and gliosis.	D167S	(Sigurdson et al., 2011)
Assists stabilisation of PrP fold.	179,180,184 (H2); 203,206,209,210,213,214 (H3); Val161 (S1); Tyr150 (H1); 134,137,139,141,157,158,198 (3 loop regions)	(Riek et al., 1996)
Residues important for species barrier.	96-167	(Telling et al., 1995)
Residues that form an epitope that is involved in the control of the species barrier and recognition of protein X.	184, 186, 203, 205	(Scott et al., 1997)

1.2.5.4 PrP is neuroprotective, Dpl is neurotoxic

Despite some similarity in their structure, PrP and Dpl seem to have antagonistic effects on neuronal cell survival (Behrens and Aguzzi, 2002, Qin et al., 2006a). Although overexpression of PrP is associated with neurodegeneration without conversion into PrP^{sc} (Westaway et al., 1994), there are multiple examples implicating PrP in protection against apoptotic cell death, demonstrated in both cell culture and animal models (Kuwahara et al., 1999, McLennan et al., 2004, Weise et al., 2004, Yamaguchi et al., 2004, Sakudo et al., 2005).

The first evidence of Dpl neurotoxicity was reported in 1999; and was the cause of Dpl discovery. Studies were prompted by the desire to find out why *Prnp*^{0/0} mice strains showed neurodegenerative features. Dpl overexpression was shown to be the reason why some of the strains develop late onset ataxia and Purkinje cell degeneration in the cerebellum that led to the discovery of Dpl (Moore et al., 1999, Moore et al., 2001). Similarly, transgenic mice selectively overexpressing Dpl in Purkinje cells develop ataxia and Purkinje cell loss despite of the presence of PrP (Anderson et al., 2004). Overexpression of an N-terminally truncated PrP, which makes it structurally more similar to Dpl, also causes ataxia (Shmerling et al., 1998). *Prnp*^{-/-} mice are rescued from Purkinje cell degeneration by introduction of transgenic PrP (Nishida et al., 1999).

The mechanism of Dpl neurotoxicity and PrP counteraction is still unclear. The proposed mechanisms include their opposite effects on apoptosis and calcium signalling pathways and oxidative stress. Experiments in primary neuronal cultures suggest Dpl-triggered apoptosis might be mediated by the proapoptotic protein BAX, caspase-3 and caspase-10 (Qin et al., 2006b, Didonna et al., 2012). By contrast, PrP demonstrates anti-BAX activity and protects neurons against BAX-mediated cell death (Park and Li, 2011, Bounhar et al., 2001). Dpl expression is correlated with the induction of heme-oxygenase-1 and nitric oxide synthases nNOS and iNOS in *Prnp*^{0/0} mice, which suggests that Dpl intensifies oxidative stress (Wong et al., 2001). The mechanism of Dpl toxicity mediated via nitric oxide production and its inhibition by PrP was confirmed in neuronal cells (Cui et al., 2003). PrP protects retinal explants against light damage and retinal degeneration; this serves as another example of PrP action against oxidative stress (Chiarini et al., 2002, Frigg et al., 2006). PrP and Dpl have opposite effects on Ca²⁺ homeostasis (Brini et al., 2005). In Chinese hamster ovary cells, transiently

transfected PrP regulates Ca^{2+} in an antiapoptotic way by limiting Ca^{2+} release from the endoplasmic reticulum and Ca^{2+} uptake by mitochondria. Alternatively, Dpl stimulates Ca^{2+} accumulation in mitochondria and its discharge from the endoplasmic reticulum. As expected, coexpression with PrP abolishes the Dpl effect on Ca^{2+} signalling (Brini et al., 2005). PrP antiapoptotic function might be explained by its ability to regulate copper homeostasis (Millhauser, 2007). Indeed, fusion of the PrP octarepeat to Dpl protects against apoptosis (Lee et al., 2006a). Intracellular copper might promote apoptosis by accelerating degradation of XIAP, an antiapoptotic copper-binding protein (Mufti et al., 2006). Coexpression of Dpl in neuronal cells might affect the copper-homeostasis functions of PrP by inhibition of its endoproteolytic cleavage (Sakudo and Onodera, 2011).

1.2.5.5 Interaction of PrP and Dpl

Biochemical and structural homology of PrP and Dpl and their antagonistic effects on neurodegeneration raises the question of their potential functional interaction.

The models of possible interaction are based on the “ligand” hypothesis first proposed by Shmerling et al. in 1998 to explain ataxia development in mice overexpressing PrP without octarepeats (Shmerling et al., 1998). According to these models, cell death is caused either by ligand binding to Dpl in the absence of PrP or ligand binding to PrP is necessary to prevent Dpl-induced cell death (**Figure 1-9, (a, b)**). Another possibility is the formation of toxic channels by Dpl, counteracted by PrP (**Figure 1-9, c**). A recently proposed model for antagonistic interaction of PrP and Dpl includes a still unidentified transmembrane molecule; it is proposed that PrP binds to that molecule via the N-terminal region, and that binding of Dpl that occurs when PrP is absent or truncated results in the toxic signal (Sakaguchi, 2008).

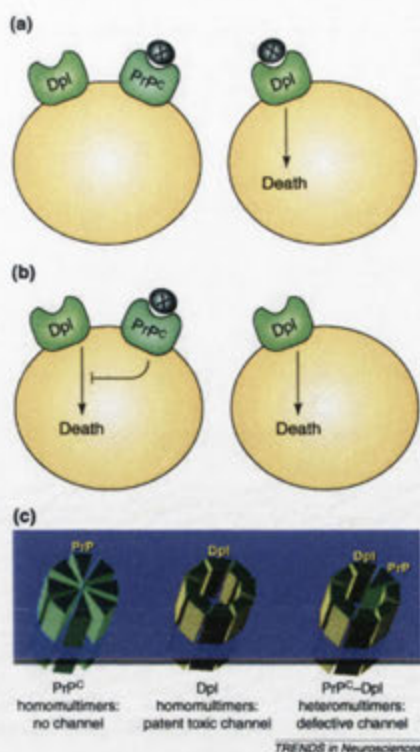


Figure 1-9. Hypothetical models of Dpl-PrP^C interaction. (a) PrP^C and Dpl compete for a common ligand (X). In the absence of PrP^C, binding of X to Dpl might cause cell death. (b) PrP^C and Dpl perform antagonistic functions through an as yet unidentified mechanism. Dpl might cause cell death when unopposed by PrP^C. (c) Dpl might form a homomultimeric toxic channel that is inactivated by participation of any full-length PrP^C. Source: (Behrens and Aguzzi, 2002)

The proposed direct interaction of PrP and Dpl has been tested experimentally in the human neuroblastoma cells. Exogenous expression of PrP and Dpl demonstrated their similar distribution when transfected separately and colocalisation at the plasma membrane after cotransfection (Massimino et al., 2004). Additional evidence of interaction was the internalisation of both proteins into the same vesicles when cells were stimulated by antibodies specific to either PrP or Dpl (Massimino et al., 2004). However, in similar experiments conducted in testicular cells by another group, PrP and Dpl failed to coimmunoprecipitate with the appropriate antibodies and so were not colocalised (Shaked et al., 2002). This inconsistency may be explained by cell-specific functions of Dpl that differ in neuronal and testis cells. Indeed, experiments in other neuronal cell lines support the interaction hypothesis (Qin et al., 2006a). Furthermore, recent evidence suggests that interaction occurs on the basolateral, detergent-resistant surface, of membranes (lipid rafts) and is cholesterol-dependent (Caputo et al., 2010).

1.2.6 Regulation of PrP and Dpl gene expression

A fundamental mechanism to control the function of a protein is to regulate its amount and localisation at the cellular, tissue and whole organism level. This can be achieved at several stages in the pathway from the gene DNA sequence to the protein at its site of action. PrP and Dpl are differentially expressed and localised in a species, developmental and tissue-specific manner (Tranulis et al., 2001, Ford et al., 2002b, Makrinou et al., 2002) suggesting that their functional evolution may be linked to the regulation of their expression pattern. There is evidence that both transcription factor-mediated expression and chromatin remodelling are important for PrP and Dpl expression.

Regulatory mechanisms described for PrP and Dpl are:

1. **Transcriptional control by transcription factors.** Transcription factors can control the rate of gene transcription by stimulating (activators) or inhibiting (repressors) basal transcription machinery carried out by RNA polymerase in a tissue-specific manner. The *PRNP* promoter region has been analysed in detail and is assumed to be analogous to a housekeeping gene based primarily on its constitutive expression in adult mammalian tissue (Oesch et al., 1985). The promoter was first characterised in hamster where it was shown to lack a TATA box for initiation and to contain three Sp-1 binding sites typically found in housekeeping genes (Oesch et al., 1985). Analysis of the human, bovine and rat *PRNP* promoter predicted only ubiquitous transcription factors, such as Sp1, Ap1 and Ap2 (Mahal et al., 2001, Inoue et al., 1997, Saeki et al., 1996). Comparisons in mammals showed some variation in promoter transcription factor binding sites: hamster, 3 Sp-1 and 1 AP-1; mouse, 2 Sp-1 and 1 AP-1; human 2 Sp-1 and 1 AP-1 (shared with rodents); but ovine, 1 AP2 and no Sp-1 or AP-1 (Lee et al., 1998, Premzl et al., 2005). However, phylogenetic footprinting of the marsupial, human, mouse, bovine and ovine *PRNP* sequences, conducted in our group, predicted both brain-specific factors (MEF-2), and other specific and ubiquitous transcription factor binding sites (MZF-1, MyT1, Oct-1, and NFAT) (Premzl et al., 2005). Most of the putative transcription factor binding sites for *PRNP* are found in the first intron suggesting this region might be involved in transcriptional regulation.

Evidence to date suggests that *PRND* is less ubiquitously expressed and its transcriptional regulation may have unique features in different cell types (Tranulis et al., 2001, Makrinou et al., 2002). Only two transcription factors, both ubiquitous, are reported so far to bind *Prnd* promoter. Nagyova *et al.* performed a mutational analysis of the mouse *Prnd* gene promoter using a luciferase reporter and identified NF-Y as a vital transcription factor and confirmed its binding to the promoter by gel shift analysis (Nagyova et al., 2004). Sepelakova *et al.* demonstrated USF-1 binding by a several-fold increase of the luciferase activity when coexpressed with the -185/+27 *Prnd*-promoter fragment inserted into luciferase reporter gene, by the gel shift assay, supershift with USF-1 antibody and UV crosslinking (Sepelakova et al., 2005).

2. Transcriptional control through the modification of chromatin structure. Gene activity can be altered through the recruitment of protein complexes which can modulate chromatin architecture and so accessibility to transcriptional machinery. Although no studies have been done to characterise the chromatin architecture of *PRNP* and *PRND* genes, some evidence suggests chromatin remodelling might be critical in regulating their expression. Firstly, despite the supposed regulation by only ubiquitous transcription factors, mouse *Prnp* has cell-specific expression pattern in multiple tissues (Ford et al., 2002b), during leukocyte maturation (Dodelet and Cashman, 1998), and embryogenesis (Manson et al., 1992, Miele et al., 2003). Also, *Prnd* shows high expression during spermatogenesis (Rondena et al., 2005), where unique chromatin remodelling events are known to regulate transcription (Sassone-Corsi, 2002) as well as wide distribution in foetal tissues but restricted in adults (Makrinou et al., 2002). Second, cells treated with a histone deacetylase inhibitor, trichostatin A, showed increased PrP expression in pheochromocytoma and glioma rat cell lines (Cabral et al., 2002). Interestingly, Sp-1 is known to interact with histone deacetylase 1 to initiate gene silencing (Doetzlhofer et al., 1999) suggesting that this sequence-specific DNA-binding protein may be required to initiate chromatin remodelling events over the promoter region. Finally, important roles for chromatin-dependent regulation have recently been proposed in synaptic plasticity and memory formation (Levenson and Sweatt, 2005), providing another link to PrP potential regulation.

3. Post-transcriptional control. Post-transcriptional regulatory mechanisms of PrP are also implicated, including presence of the conserved signal for nuclear-specific

polyadenylation and the cytoplasmic polyadenylation element (CPE) in the *PRNP* 3' UTR (Premzl et al., 2005), compartmentalisation of mRNA in vesicles (epididymosomes) (Ecroyd et al., 2004) for delayed translation, protein trafficking or degradation (Ford et al., 2002a), and cell-surface sorting (Uelhoff et al., 2005).

1.3 Motivation and aims of the thesis

Our early discovery of the “chimeric” expression of *PRND* from its parental gene *PRNP* promoter in the amphibian *X. tropicalis* provided an extraordinary opportunity to investigate the evolutionary mechanism of survival of duplicated genes. Although much research has been done to study gene evolution and gene families, nobody had a convenient model before to follow gene evolution from the “nascence” of the gene and to analyse systematically changes in gene structure and expression.

The *PRNP/PRND* genomic contig is a remarkable subject for evolutionary research due to its coexpression in vertebrates. Previous studies have not focused on evolutionary origin of this system. The objective of this thesis is to cover this gap by demonstrating the evolution of structure and function of *PRNP* and *PRND* genes and their proteins to provide new insight into their physiological activity, specifically to understanding of their roles in neurodegenerative diseases, such as prion diseases and Alzheimer's.

Throughout this thesis, the term “function” will refer to molecular processes in which gene products are involved. Similarly, the term “functional evolution” will refer to the distribution of parental gene function between two duplicates.

Ohno (1970) proposed that the first step in the functional divergence between duplicate genes is expression divergence. Thus, my first aim will be to characterise previously unknown amphibian and marsupial *PRND* transcripts and to study their tissue-expression patterns.

A more complex model of functional evolution suggests that changes in gene expression result from loss or acquisition of regulatory DNA elements (Ohta, 2003). The term “regulatory DNA region” will refer to DNA regions involved in transcription and transcriptional regulation. Transcriptional expression is achieved by proteins - transcription factors - and their target cis-regulatory DNA elements. Thus, my second

aim will be to investigate regulation of expression of *PRNP* and *PRND* as the molecular basis for expression divergence.

The characterisation of *Prnp*^{0/0} mice and mutant mice where various PrP/Dpl functional domains were deleted or swapped established an antagonistic functional relationship of PrP and Dpl, as referenced in **section 1.2.5.4**. These studies reveal the possibility that PrP and Dpl coevolved within vertebrates which led me to the proposition that the best way to understand their functional relationship is to study them simultaneously. I hypothesised that PrP and Dpl are coregulated to achieve balance between apoptosis and survival. Conceptual analysis of the existing data complemented with novel experimental results provides the necessary information for comparative analysis of *PRNP* and *PRND*. So, the third aim of this thesis is to analyse putative transcription factors that could regulate both genes simultaneously, and to develop a theory of how simultaneous regulation can be achieved.

Chapter 2.

Materials and Methods

In this chapter I describe the background of extensive range of methods I used. Interpretation of the results cannot be fully understood without understanding the basis of these methods. Detailed technical methodology can be found in relevant chapters. In addition, I summarise general molecular biology protocols (**Section 1.8**), software (**Table 2-1**), reagents (**Table 2-2**), kits (**Table 2-3**), buffers (**Table 2-4**), and cell lines (**Table 2-5**).

2.1 Homology modelling

Homology modelling is a computational method applied to generate a three-dimensional structure of a particular protein using its amino-acid sequence. This requires availability of experimentally determined structure and deposited X-ray or NMR data for an evolutionarily related homologous protein for use as a template. The model is constructed based on sequence similarity between the query (protein of interest) sequence and the template sequence comprising four main steps: selection of the evolutionarily related protein with known three-dimensional structure, sequence alignment, construction of the three-dimensional model, and model evaluation (Marti-Renom et al., 2000). As protein structures are conserved among homologues (Chothia and Lesk, 1986), the predicted structure for the target protein should be reliable. However, there are some limitations when homology is based on evolutionarily distant proteins, with higher sequence divergence. Proteins with sequence identity less than 20% can have significantly different structure (Chothia and Lesk, 1986). Therefore, the accuracy of the model should be carefully evaluated to assess the relevance of the model for protein function (Baker and Sali, 2001).

2.2 RACE

2.2.1 5' GeneRacer™ kit (Invitrogen)

This method is based on the assumption that intact mRNA has a special structure “cap” at its 5' end. The procedure starts with dephosphorylation. This step prevents ligation of the oligonucleotide adaptor, which requires 5'-phosphate group, to the 5' end of

degraded mRNA or other types of RNA. Therefore, they will not be converted to cDNA in later steps. The cap is then removed to make the 5' phosphorylated end available for the ligation reaction. The adaptor-linked mRNA is reverse-transcribed into full-length cDNA with oligo(dT) primer which contain another adaptor sequence at the 3'end. The cDNA ends are amplified in two separate reactions by using primers specific to the 5' or 3'adaptor sequences, respectively, together with the gene-specific primers listed in **Appendix 1**.

Main problem with this method is possible mRNA degradation after the cap, which protects the 5'end, is removed. Degraded mRNA will result in a range of cDNAs with variable, not fully amplified 5' UTRs. The detection step includes PCR amplification of the mixture of cDNAs and cloning into sequencing vector with subsequent bacterial colonies analysis leading to the reduced possibility of uncovering the full-sized mRNA among multiple products.

2.2.2 BD SMARTTM RACE cDNA Amplification Kit (Clontech)

In my experience, this method produces much more accurate results in terms of intact 5'UTR amplification. There is no need for adaptor ligation and other time-consuming enzymatic reactions. This method is based on the template switch from mRNA molecule to the adaptor RNA oligo, as shown in **Figure 2-1**. First-strand cDNA synthesis is primed using a modified oligo(dT) primer. BD PowerScript reverse transcriptase (RT) has terminal activity. It adds 3-5 dC residues to the 3'end of the first-strand cDNA. The BD SMARTTM oligo contains terminal G residues that anneal to the tail of the cDNA and serve as an extended template for reverse transcriptase (RT) (**Figure 2-1**).

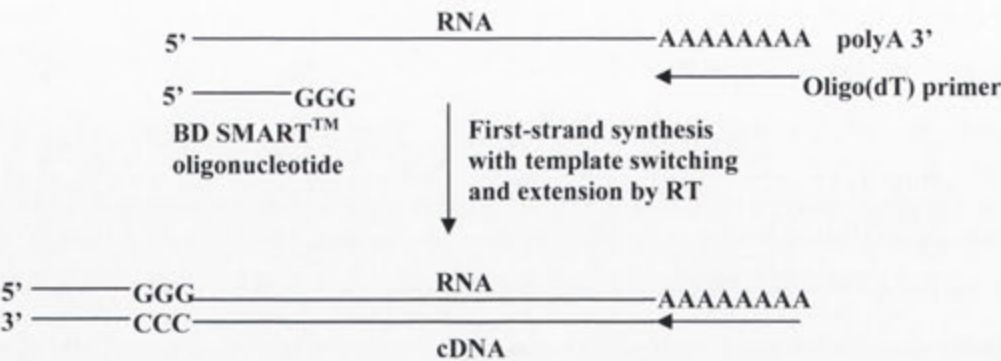


Figure 2-1. Mechanism of BD SMART™ cDNA synthesis.

2.3 Cloning

The choice of cloning method was based on the application.

Examination of transcripts requires fast and efficient cloning method for sequencing. The best approach is to use the TOPO-TA Cloning® Kit for Sequencing (Invitrogen). Gel-purified PCR product is directly inserted into topoisomerase-containing plasmid vector. The method is based on the properties of *Taq* polymerase to add 3' deoxyadenosine (A) overhangs to the PCR product and the ability of the topoisomerase to ligate the PCR product to the vector containing 3' thymidine (T) overhangs. Plasmid vector contains specific sites for forward and reverse sequencing primers. It allows sequencing previously unidentified transcripts by eliminating the need for sequence-specific primers.

More complex applications, such as protein expression or analysis of promoter activity, require a ligation into specialised vectors and, therefore, creation of a specific restriction site. I used three different approaches to address this problem. First, I analysed the sequences for presence of appropriate restriction sites. If the sites were present, the sequence was PCR-amplified over these sites, cleaved with appropriate restriction enzyme and cloned into corresponding restriction sites in the vector. Second, when the restriction sites were absent, I introduced them by using pCR®4-TOPO® cloning vector (Invitrogen) that contains multiple restriction sites. I used TOPO-TA cloning with consecutive restriction digestion to create the DNA fragment with appropriate ligation ends. Third, restriction sites were introduced by using PCR primers containing specific sequences.

2.4 Site-Directed Mutagenesis

Site-directed mutagenesis was performed using the Stratagene kit approach. The modification is introduced by using two synthetic complementary oligonucleotide primers containing the desired mutation and the plasmid with the DNA insert to be mutated (**Figure 2-2**). PCR using proof-reading Pfu Turbo DNA polymerase (Invitrogen) amplifies the whole plasmid incorporating the mutation from primers (**Figure 2-2**). Following the PCR, the product is treated with the endonuclease DpnI, which is specific for methylated DNA and, therefore, only digests the template DNA. The nicked vector mutant DNA is then transformed into *E.coli* (XL-1 blue strain) which repairs the nick and replicates the DNA. *E.coli* transformed with the plasmid is selected by plating on LB plates containing specific antibiotic. Successful nucleic acid mutagenesis is confirmed by sequencing over the site of the mutations.

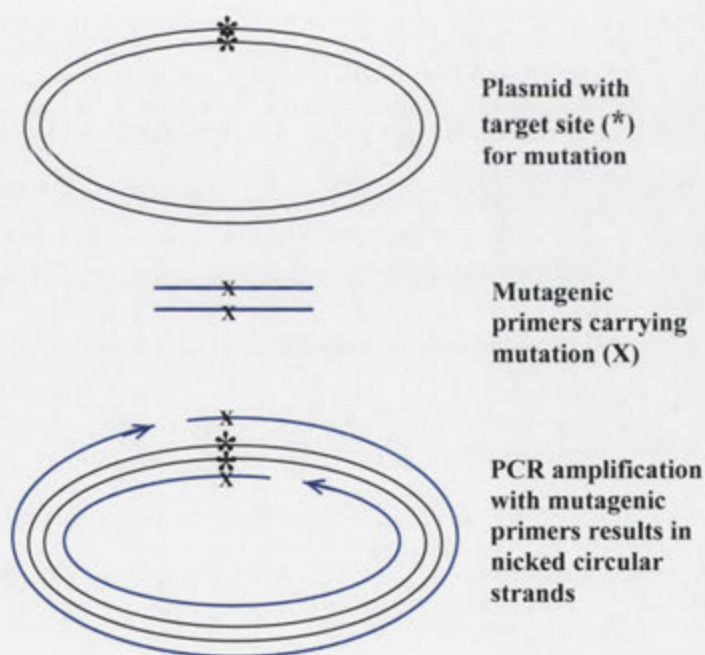


Figure 2-2. Site-directed mutagenesis by PCR amplification with mutagenic primers carrying desired mutation.

2.5 Luciferase assay

The luciferase assay is a powerful method for studies of transcriptional regulation designed to investigate the response of the gene promoter to specific transcription factors. The capacity of luciferase reaction to emit light allows monitoring the upregulation or downregulation of promoter elements. For this study I used Dual-GloTM Luciferase Assay System (Promega) that includes firefly luciferase vector and *Renilla* luciferase expression plasmid. Firefly luciferase vector is used for cloning the promoter of interest upstream of the luciferase gene originated from the firefly species *Photinus pyralis*. *Renilla* luciferase, originated from the sea pansy *Renilla reniformis*, is used as an internal control for DNA transfection efficiency and normalisation of expression of the firefly luciferase. Mammalian cells are transfected with plasmids containing specific gene promoters cloned into the luciferase vector, the *Renilla* luciferase plasmid, and with plasmids expressing the transcription factors being investigated. Activation or downregulation of the cloned promoter affects the base level of luciferase protein expression. This level can be easily detected using a luminometer that measures luminescence emitted in the presence of luciferase substrate (luciferin for firefly luciferase and coelenterazine for *Renilla* luciferase), with the consequent calculation of a normalised firefly luminescence / *Renilla* luminescence ratio for each sample. It is assumed that the transcription factor activates the promoter if the relative expression of luciferase cotransfected with the specific transcription factor is higher than without that factor.

2.6 Polysome-binding assay

The level of mRNA is not always reliable as an estimate of the corresponding protein abundances (Anderson and Seilhamer, 1997, Gygi et al., 1999). The polysome-binding assay is one of the most common methods to estimate transcript translation (Arava, 2003). It is based on the assumption that polysome-bound mRNAs are translated. This method takes into consideration translational control of mRNA expression by enabling the separation of free mRNA from ribosome-loaded mRNA and correlates closely with protein synthesis (Pradet-Balade et al., 2001, Smith et al., 1999).

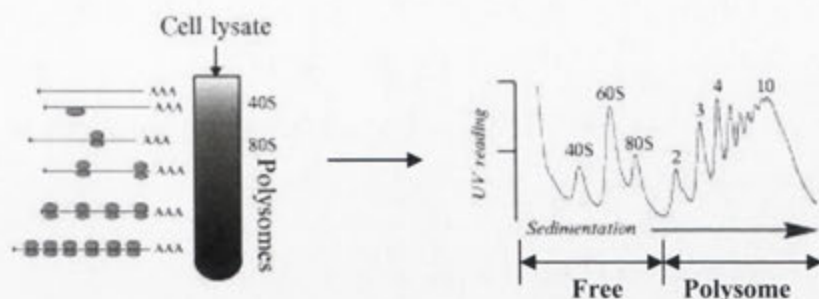


Figure 2-3. Polysome-binding assay. RNAs from the cell lysate are separated in a sucrose gradient based on their ribosomal load. The UV absorbance profile shows fractions that contain 40S, 60S, and 80S ribosomal RNA; and fractions that are separated according to the number of mRNA-associated ribosomes (polysomes); numbered are 2, 3, 4, and 10 ribosome fractions (modified from (Melamed and Arava, 2007)).

2.7 Reagents, Kits, Solutions, and Cell lines

The list of common reagents used for this study and their suppliers are listed in **Table 2-2**. Molecular biology kits are listed in **Table 2-3**. Cell lines and solutions used for experimental work are listed in **Tables 2-4 and 2-5** respectively. LB media and PBS buffer was provided by the John Curtin School of Medical Research (JCSMR), Australian National University media/wash-up facility.

2.8 General molecular biology protocols

2.8.1 Bacterial Cultures

Transformed bacterial strains were grown on LB plates supplemented with 100 µg/ml ampicillin overnight at 37 °C. Single colonies were selected for liquid culture mini or maxipreps (5 ml and 400 ml respectively) and grown overnight at 37 °C in LB media containing 100µg/ml ampicillin.

2.8.2 DNA Preparations

Plasmid constructs and their descriptions are listed in **Table 1.5**. Plasmids were purified from bacterial strains using Wizard® Plus SV Minipreps DNA Purification System (Promega) or GenElute™ Endotoxin-free Plasmid Midiprep kit (Sigma), according to the manufacturer's instructions.

2.8.3 Agarose Gel Electrophoresis

Electrophoretic separation of DNA was carried out using 1-2% agarose gels in 1xTAE buffer containing 0.5 µg/ml ethidium bromide. Purification of DNA excised from the gel was carried out using the QIAquick Gel extraction kit (Qiagen) or GenElute™ Agarose Spin Column (Sigma).

2.8.4 Oligonucleotide Primers

Oligonucleotide primers for PCR, sequencing, and mutagenesis were designed by using Primer3 software (<http://frodo.wi.mit.edu/primer3>) (Rozen and Skaletsky, 2000) and synthesised by Sigma-Aldrich (Australia). Primers were shipped as dry pellets and then resuspended in sterile ddH₂O to a 100 µM stock concentration. Later, primer stocks were further diluted to achieve the desirable working concentration, normally 10 µM for PCR reactions and 2 µM for sequencing reactions. Primer sequences are provided in the corresponding thesis chapters.

2.8.5 Automated Sequencing

DNA sequencing was performed based on the chain termination method using the ABI prism Big Dye v3.1 Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems), according to the manufacture's protocol. Briefly, 400 ng of purified plasmid DNA was combined with 1 µl of Big Dye, 3.5 µl of 5X Big Dye reaction buffer, 1.6 µl of 2 µM plasmid-specific primer (forward or reverse), and sterile ddH₂O to the final volume 15 µl. Sequencing PCR was carried out in an Eppendorf Thermal Cycler. The sequencing program consisted of 1 cycle of 96 °C for 1 min followed by 30 cycles of 96 °C for 20 sec, 50 °C for 10 sec, 60 °C for 4 min. Extension products were purified by adding 1.0 µl of 3 M NaOAc and 50 µl of 95 % EtOH. The mixture was vortexed briefly then incubated at room temperature for 15 min. The tubes were centrifuged for 15 min at 13200 rpm and the pellet washed with 1 ml 70 % EtOH, vortexed and centrifuged again for 3 min. The pellet was dried at room temperature for 10 min. Sequencing products were analysed at the Biomolecular Resource Facility (BRF), in the John Curtin School of Medical Research (JCSMR), Australian National University by using the ABI 3730 capillary DNA Analyzer (Applied Biosystems).

2.8.6 Mammalian Cell Culture Methods

Mammalian cell lines were maintained in the culture media, described in **Table 1.4**, at 37 °C in 5 % CO₂. Cells were grown in 10 ml dishes and 25 ml or 75 ml flasks (Nunc Brand products, Nalge Nunc International, Denmark). Frozen cell stocks were prepared in the culture media supplemented with 10 % FBS and 10 % DMSO. Cells were subcultured after they reached confluence at a ratio 1:10. For luciferase-assay experiments, cells were grown in 12-well plates to about 90 % confluency. Plasmids were transfected using Metafectene[®]Pro, a polycationic liposomal transfection reagent, obtained from Biontex Laboratories (Munich, Germany).

Table 2-1. Public databases and software

Name	Reference
BioEdit, version 5.0.9	http://www.mbio.ncsu.edu/BioEdit/bioedit.html
ClustalW	http://www.ebi.ac.uk/clustalw/
EMBOSS	http://www.ebi.ac.uk/emboss/
Ensembl	www.ensembl.org
FANTOM	www.gsc.riken.go.jp/e/FANTOM/
Genomatix	www.genomatix.de
GPI ancor	http://mendel.imp.univie.ac.at/sat/gpi/gpi_server.html
NCBI	www.ncbi.nlm.nih.gov
N-glycosylation	http://www.cbs.dtu.dk/services/NetNGlyc/
PolyA sites	http://www.imtech.res.in/raghava/polyapred/index.html
Primer3	http://frodo.wi.mit.edu/primer3
PyMol	http://pymol.org/educational/
RepeatMasker	http://ftp.genome.Washington.edu/RM/RepeatMasker.htm
Sanger EST database	http://www.sanger.ac.uk/cgi-bin/blast/submitblast/x_tropicalis
SignalP	http://www.cbs.dtu.dk/services/SignalP/
SWISS-MODEL	http://swissmodel.expasy.org
TRANSFAC	http://www.gene-regulation.com/pub/databases.html
XenBase	http://genome.jgi-psf.org/Xentr4/Xentr4.home.html

Table 2-2. Commonly used reagents and their sources

Reagent	Source
AccuTaq™ LA DNA Polymerase	Sigma-Aldrich
Ampicillin	Sigma-Aldrich
Cycloheximide	Sigma-Aldrich
Dithiothreitol	Astral Scientific
DMSO	Sigma-Aldrich
EDTA	Sigma-Aldrich
Fetal Bovine Serum (FBS)	Invitrogen
Glacial acetic acid	AJAX Chemicals
L-glutamine	Invitrogen
Heparin, sodium salt	Sigma-Aldrich
Horse Serum	Invitrogen
LiCl	Sigma-Aldrich
2-mercaptoethanol	Sigma-Aldrich
Metafectene® Pro	Biontex Laboratories, Germany
MgCl ₂	Sigma-Aldrich
NaCl	Sigma-Aldrich
Nonidet-P40	Roche Applied Science
Phenol-chloroform-isoamyl alcohol	Sigma-Aldrich
Platinum Taq polimerase	Invitrogen
Proteinase K	Sigma-Aldrich
Restriction enzymes	New England Biolab
RNAlater	Ambion
RNAasin	Promega
SDS	Sigma-Aldrich
Sucrose	Sigma-Aldrich
Tris-HCl	Merk
TRIzol® reagent	Invitrogen
Trypsin-EDTA	Invitrogen
VRC	New England Biolab

Table 2-3. Kits and their suppliers

Kit	Supplier
ABI prism Big Dye terminator cycle Sequencing Ready Reaction	Pelkin Elmer
BD SMART TM RACE cDNA Amplification	Clontech
Dual-Glo TM Luciferase Assay System	Promega
GenElute TM Agarose Spin Column	Sigma
GenElute TM Endotoxin-free Plasmid Midiprep kit	Sigma
GeneRacer TM	Invitrogen
QIAquick Gel extraction	Qiagen
RNAeasy Mini Kit	Qiagen
SuperscriptIII reverse transcriptase	Invitrogen
TOPO-TA Cloning [®] Kit for Sequencing	Invitrogen
Wizard [®] Plus SV Minipreps DNA Purification System	Promega

Table 2-4. Cell lines used and culture media

Cell line	Tissue origin	Media
bEND.3 (ATCC:CRL-2299)	Mouse brain endotelial	DMEM, 10% fetal bovine serum, 1% glutamine
GC1 (ATCC: CRL-2053)	Mouse testis, spermatogonia	DMEM, 10% fetal bovine serum, 1% glutamine
GC2 (ATCC: CRL-2196)	Mouse testis, spermatocyte	DMEM, 10% fetal bovine serum, 1% glutamine
TM4 (ATCC: CRL-1715)	Mouse Sertoli cells	DMEM/F12, 5% horse serum, 2.5% fetal bovine serum
SHSY5Y (ATCC: CRL-2266)	Human neuroblastoma	DMEM, 10% fetal bovine serum, 1% glutamine

Table 2-5. Buffers and Solutions

Name of media	Composition
Luria-Bertani (LB)	10 g/l bacto-tryptone; 5 g/l bacto-yeast extract; 10 g/l NaCl
LB plates	LB + 1.5% (w/v) agar + ampicillin antibiotic
PBS	137 mM NaCl, 7 mM Na ₂ HPO ₄ •2 H ₂ O, 2.5 mM NaH ₂ PO ₄ •H ₂ O, pH7.4
Polysome lysis buffer	10 mM Tris-HCl at pH 8.0, 140 mM NaCl, 1.5 mM MgCl ₂ , 0.5% Nonidet-P40, 20 mM dithiothreitol, 150 µg/ml cycloheximide, 500 U/ml RNAsin, 40 mM VRC
Sucrose gradient	15-40% sucrose, supplemented with 10 mM Tris-HCl at pH 7.5, 140 mM NaCl, 1.5 mM MgCl ₂ , 10 mM dithiothreitol, 100 µg/ml cycloheximide, 0.5 mg/ml heparin
TAE (50X)	2M Tris-HCl, 5.7% glacial acetic acid, 50 mM EDTA
TE	10 mM Tris-HCl, pH8.0; 1 mM EDTA, pH 8.0
Trypsin, 0.25%	0.25% (w/v) trypsin, 1 mM EDTA in PBS

Chapter 3.

Evolution of *PRNP* and *PRND*

Chapter 1 introduced gene duplication as a major mechanism for evolution of new proteins and functions. In order to survive, duplicates need to acquire new regulatory DNA elements for their expression and develop a new function or functions.

In Chapter 3 I will demonstrate “evolution in action” using *PRNP* gene duplication as an example.

3.1 Introduction

The *PRNP* gene has been reported in fish, amphibians, reptiles, birds, marsupials and eutherian mammals and characterised in different classes of these vertebrates (**Section 1.2**).

Currently, partially characterised *PRNP* transcripts are available for

- amphibian *X. laevis* with partial 5'UTR and unknown numbers of exons (Strumbo et al., 2001).
- marsupial brush-tailed possum genomic clone with unknown number of exons and non-defined 5' and 3' UTRs (Windl et al., 1995), *M. domestica* (unusual ORF, unknown number of exons), and tammar wallaby with one coding and one non-coding exon (Premzl et al., 2005).
- eutherian mammals with 2 non-coding and one coding exon (Inoue et al., 1997, Ford et al., 2002b, Van Rheede et al., 2003) except human and hamster transcripts that have only one non-coding exon (Oesch et al., 1985; Makrinou et al., 2002).

In contrast, the *PRND* gene is still poorly characterised. So far it has been reported only in eutherian mammals (**Table 1-1**). Bioinformatics prediction of *PRND* transcripts, including chimeric transcripts, is only available for mouse and can be found in the Ensembl database (www.ensembl.org).

Genome sequencing projects revealed that *PRND* is absent in ray-finned fish: examples are zebrafish (*Danio rerio*) and pufferfish (*Takifugu rubripes*) (Premzl et al., 2004).

From these results we can infer that the duplication of *PRNP*, the parental gene of *PRND*, occurred after the last common ancestor of fish and amphibians. Here, I investigate how the expression mechanism of the newly duplicated *PRND* gene developed. The aim of this chapter is to characterise *PRND* expression in early quadruped and intermediate vertebrates (amphibians and marsupials), that have not been reported and analysed before, and to compare it with *PRND* expression in mouse. For experimental transcript analysis I selected one model animal from three major evolutionary groups: amphibians, marsupials, and mammals (**Figure 3-1**). A model for the bird group was excluded as chicken was extensively studied by Nagesh Chakka, a former PhD student in our group. He reported that the *PRND* gene is absent in chicken. For the reptile group, turtle and lizard genomes were not assembled at the time of the experimental research for this thesis and tissues for the common model species are not available in Australia because of quarantine issues. Although a diploid amphibian, *Xenopus tropicalis* has been used as the model organism for genome sequencing, only the tetraploid *Xenopus laevis* is available for researchers in Australia due to the quarantine issues. Thus, although I was able to initially clone and characterise amphibian *PRNP* and *PRND* cDNAs from *X. tropicalis*, more detailed research was conducted using tissues from *X. laevis*. I found that whole genome duplication in *X. laevis* resulted in retention of both *PRNP* gene loci, but only one of the *PRND* genes. This provided an interesting opportunity to investigate gene duplication issues further.

In summary, I will follow and compare the development of the transcripts and expression of the *PRNP* and *PRND* gene duplicates using three vertebrate models: amphibian *Xenopus laevis*, marsupial *Monodelphis domestica* and mouse. This information provides essential knowledge for understanding coregulation of *PRNP* and *PRND*, as presented in Chapter 4.

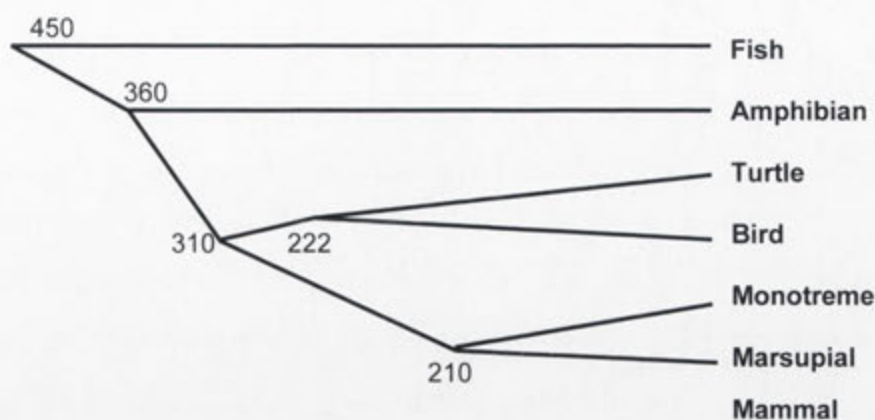


Figure 3-1. The molecular evolutionary timescale. The numbers indicate the years of separation in millions (Kumar and Hedges, 1998).

3.2 Material and Methods

3.2.1 Total RNA extraction

Purified RNA from *X. tropicalis* adult male brain and testis was obtained as a gift from Dr Timothy C. Grammer (Harlands lab, University of California, Berkeley). *X. laevis* total RNA was extracted from various adult male tissues (brain, testis, kidney, liver, lung, muscle, gut, spleen, skin, stomach, eye, heart, pancreas and gall bladder), using TRIZOL Reagent (Invitrogen) as recommended in the manufacturer's manual.

Adult opossum (*Monodelphis domestica*) tissues preserved in RNAlater solution (Ambion) were kindly provided by Prof Norman R. Saunders, University of Melbourne. Adult wild-type BALB/C mice (6, 12 and 18 weeks old) tissues were collected and preserved in RNAlater. RNA was isolated from brain, kidney, liver, ovaries, lung, muscle, gut, stomach, heart, and pancreas of female, and testis and eye of male opossum, and brain, testis, kidney, liver, lung, spleen, muscle, heart, gut, stomach, eye, and thymus of male mouse. All RNA extractions were conducted using the RNAeasy Mini Kit (Qiagen). DNase treatment (Ambion) was included to remove residual DNA contamination.

X. laevis and mouse tissues were salvaged from animals maintained locally in the John Curtin School of Medical Research (JCSMR), Australian National University.

3.2.2 Amplification of *X. laevis* PRND ORF fragment

A genomic fragment, corresponding to part of the *PRND* gene, was amplified from *X. laevis* DNA using primers specific to the *X. tropicalis* sequence in the genome database (<http://genome.jgi-psf.org/xenopus>). For the first PCR reaction, 100 ng of genomic DNA and 0.2 μ M F3 and R2 primers (**Appendix 2**) were used to amplify the *PRND* fragment. The PCR cycle was 94 °C 2 min, (94 °C for 30 sec, 51 °C for 30 sec, 72 °C for 30 sec) 35 cycles, 72 °C 10 min. 1 μ l of the first amplification reaction and 0.2 μ M F3 and R3 were used as a template for nested PCR with the same cycling parameters. The PCR products were cloned into pCR[®]4-TOPO vector (Invitrogen) and sequenced by the Australian Genome Research Facility (AGRF), Brisbane, Australia.

3.2.3 5' and 3' Rapid amplification of cDNA ends (RACE)

To isolate the cDNAs coding for the Dpl and PrP proteins in *X. tropicalis* and *X. laevis*, the GeneRacer[™] Kit (Invitrogen) was used. 1 μ g of total RNA from brain and testis was dephosphorylated, decapped, ligated to GeneRacer[™] RNA Oligo and reverse-transcribed into cDNA following the manufacturer's instructions. Cloned AMV RT enzyme and GeneRacer[™] Oligo(dT) Primer were used to synthesise full-length cDNA. To amplify the cDNA ends, GeneRacer[™] Primers homologous to sequence in the GeneRacer[™] RNA Oligo (5' end) or GeneRacer[™] Oligo(dT) Primer (3' end) were used together with the gene specific primers listed in **Appendix 1**. 50 ng of brain cDNA for *PRNP* or testis cDNA for *PRND* were amplified by PCR using Taq DNA polymerase (Promega). The PCR cycling conditions were 94 °C 2 min, (94 °C 30 sec, 72 °C 1 min) 5 cycles, (94 °C 30 sec, 70 °C 1 min) 5 cycles, (94 °C 30 sec, 63 °C 30 sec, 72 °C 1 min) 25 cycles, 72 °C 10 min. 1 μ l of the first amplification reaction was used as a template for nested PCR. Cycling conditions were 94 °C 2 min, (94 °C 30 sec, 63 °C 30 sec, 72 °C 1 min) 35 cycles, 72 °C 10 min. The PCR products were cloned into pCR[®]4-TOPO vector (Invitrogen) and sequenced by the Biomolecular Resource Facility (BRF), JCSMR, the Australian National University (ANU).

The 5' UTR and 3' UTR of *M. domestica* cDNAs coding for PrP and Dpl were amplified using the BD SMART[™] RACE cDNA Amplification Kit (Clontech) and 1 μ g of total testis RNA. Gene-specific primers (**Appendix 1**) were designed following the RACE kit manual instructions. PCR cycling conditions were (94 °C 30 sec, 72 °C 3min) 5 cycles, (94 °C 30 sec, 70 °C 30 sec, 72 °C 3 min) 5 cycles, (94 °C 30 sec, 68 °C

30 sec, 72 °C 3 min) 25 cycles, 72 °C 10 min. 1 µl of the first amplification reaction was used as a template for nested PCR. Cycling conditions were 94 °C 2 min, (94 °C 30 sec, 68 °C 30 sec, 72 °C 3 min) 35 cycles, 72 °C 10 min. The PCR products were cloned into pCR®4-TOPO vector (Invitrogen) and sequenced by the BRF, JCSMR, ANU.

3.2.4 Reverse transcription (RT)-PCR and tissue expression analysis

1 µg of total RNA from each *X. laevis*, opossum or mouse tissue was primed by an oligo(dT) primer and first strand cDNAs were synthesised by using SuperscriptIII reverse transcriptase (Invitrogen), following the manufacturer's instructions. 1 µl of the first-strand cDNA was used as a template for PCR with 0.2 µM primers (**Appendix 2**). Primer pairs for *X. laevis* tissue expression analysis were F4/R6 for *PRNP-1a*, F5/R4 for *PRNP-1b*, F7/R6 for *PRNP-2a*, F6/R6 for *PRNP-2b*, F4/R7 for *PRND-ch*, F9/R7 for *PRND-a*, F10/R7 for *PRND-b* and F11/R10 for *GAPDH* (**Figure 3-2**). Primer pairs for *M. domestica* tissue expression analysis were LmdPrP1/RmdPrP521 and LmdPrP1/RmdCPE167 for *PRNP*, LmdPrP1/RmdDpl350, Lmd2D/RmdDpl350 and Lmd5D/RmdDpl350 for *PRND* and LmdGAPDH/RmdGAPDH for *GAPDH* (**Figure 3-6**). Primer pairs for mouse tissue expression analysis were LmmPex1-94/RM587 for chimeric *Prnd*, LMex1Dpl/RM587 for non-chimeric *Prnd* and Lβ-actin/Rβ-actin for β-actin (**Figure 3-11**). To avoid genomic DNA amplification, all RNA samples were treated with DNase (Ambion) and primer pairs were designed as specific to different exons (in this case DNA amplification would result in a larger size PCR band) or overlapping exon-exon border (in this case DNA amplification was not possible). H₂O control was used to test for template contamination. PCR cycling conditions were 94 °C for 2 min initial denaturation followed by 25 cycles of the reaction (94 °C for 30 sec, 58 °C for 30 sec, 72 °C for 1 min) for *PRNP* variants and *GAPDH*, or 35 cycles for *PRND*, with final elongation at 72 °C for 10 min. The amplified DNA fragments were cloned into pCRII®TOPO vector (Invitrogen) and sequenced in the BRF, JCSMR, ANU, to confirm their identity.

3.2.5 Database searches

TBLASTN (Altschul et al., 1990) was used to search the *X. tropicalis* genome trace reads database (<http://genome.jgi-psf.org/xenopus>) using various mammalian Dpl

sequences as query. The query sequence from human Dpl containing the N-glycosylation sites and Cys involved in the disulfide bridge (HumDpl 57-140) picked up a hit from trace read TKS687278.g1, which corresponded to a stretch of the ORF of 159 residues but lacked the complete N-terminal region. The BLAST parameters used were a cut-off e-value of 1, default gap penalties, gapped blast program and low complexity regions permitted. A search with the nucleotide sequence from the 5' end of TKS687278.g1 (using BLASTN) gave another overlapping contig AAWU3059.g1 to complete the ORF for *X. tropicalis* Dpl. *X. tropicalis* PrP was identified using the *X. laevis* PrP sequence (Strumbo et al., 2001) as query. With the later release of assembled versions of the genome data, including the current version 4.1 (<http://genome.jgi-psf.org/Xentr4/Xentr4.home.html>), we were able confirm our findings obtained from trace reads and were also able to define the genes surrounding PrP and Dpl. Screening the EST library (NCBI EST database and Sanger EST database (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/x_tropicalis) for Dpl in *X. laevis* and *X. tropicalis* did not produce any positive hits but showed multiple hits for PrP for both species. Both BLASTN and TBLASTN programs with default Blast parameters were used for the searches.

M. domestica protein-coding sequences were revealed by construction of a local standalone BLAST database which was managed by Perl scripts (createBlastdb.pl). Initial searches were made on the assembled version 0.5, which was the latest at the time of study. Homology searches were performed using TBLASTN with the *Tricosurus vulpecular* amino acid query sequence for PrP and human sequence for Dpl. This database search work was carried out by Nagesh Chakka. I used the obtained sequence information to design primers for the experimental work. Amino acid sequences were predicted based on the genome data (**Appendix 4**).

3.2.6 Sequence analysis

N-terminal signal peptide cleavage sites for *X. tropicalis*, *X. laevis* and *M. domestica* PrPs and Dpls were predicted using the SignalP program (version 3.0) (Nielsen et al., 1997) (<http://www.cbs.dtu.dk/services/SignalP/>).

GPI modification sites were predicted by using the program Big-pi (Eisenhaber et al., 1999) (http://mendel.imp.univie.ac.at/sat/gpi/gpi_server.html); modification sites for PrPs were identified but clear sites for Dpls were not predicted.

N-glycosylation sites were predicted by using the NetNglyc server (Gupta and Brunak, 2002) (<http://www.cbs.dtu.dk/services/NetNGlyc/>)

Polyadenylation sites were predicted by using the PolyApred program (Ahmed et al., 2009) (<http://www.imtech.res.in/raghava/polyapred/index.html>).

Multiple sequence alignments were performed using web-based ClustalW (Higgins, 1994) (<http://www.ebi.ac.uk/clustalw/>) using default parameters, and then manually edited using BioEdit Version 5.0.9 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). The percentage identity in pairwise sequence alignments was calculated using the Needleman-Wunsch global alignment algorithm (Needleman and Wunsch, 1970) provided by EMBOSS (<http://www.ebi.ac.uk/emboss/align/>).

3.2.7 Homology modelling

The SWISS-MODEL Automated Protein Modelling Server (<http://swissmodel.expasy.org>) was used for homology modelling. Graphic images were produced using PyMol software (<http://pymol.org/educational/>). Representative structures from mouse Dpl (PDB ID 1I17), *X. laevis* PrP (PDB ID 1XU0), tammar wallaby PrP (PDB ID 2KFL), and mouse PrP (PDB ID 1AG2) were used as templates.

3.2.8 Transposable element content analysis

I used the Repeat Masker program to determine the relative content of repeats and transposable elements in the *X. tropicalis*, *M. domestica*, and mouse *PRNP/PRND* locus (<http://ftp.genome.Washington.edu/RM/RepeatMasker.htm>).

3.2.9 Polysome binding assay

Polysomes were prepared and analysed using published protocols (Del Prete et al., 2007, Mikulits et al., 2000). Briefly, a piece (~20 mg) of solid tissue was pulverised under liquid nitrogen and the obtained powdered tissue was lysed by pipetting up and down 10 times in 1 ml of lysis buffer (10 mM Tris-HCl at pH 8.0, 140 mM NaCl, 1.5 mM MgCl₂, 0.5% Nonidet-P40, 20 mM dithiothreitol, 150 µg/ml cycloheximide,

500 U/ml RNasin (Promega), 40 mM VRC (New England Biolabs). The nuclei were removed by centrifugation (12000 g, 10 sec, at 4 °C). The supernatant was supplemented with 30 µl heparin (50 mg/ml) and centrifuged (12000 g, 5 min, at 4 °C) to remove mitochondria and membrane particles. The supernatant was layered onto a 10 ml linear sucrose gradient (15-40% sucrose, supplemented with 10 mM Tris-HCl at pH 7.5, 140 mM NaCl, 1.5 mM MgCl₂, 10 mM dithiothreitol, 100 µg/ml cycloheximide, 0.5 mg/ml heparin) and centrifuged in a SW41Ti rotor (Beckman) for 120 min at 38000 rpm at 4°C, with brake off. Fractions (20 x 550 µl) were collected and treated with 100 µg proteinase K in 1% SDS and 10 mM EDTA for 30 min at 37°C. RNAs were recovered by phenol-chloroform-isoamyl alcohol extraction, followed by ethanol precipitation. For the EDTA release control experiments, lysis buffer was supplemented with 25 mM EDTA (pH 8.0). Heparin was neutralised by 2 M LiCl precipitation. RNAs from similar fractions (cytoplasmic, monosomal, polysomal) were pooled and reverse-transcribed using random hexamers and Superscript III polymerase (Invitrogen). Corresponding cDNAs were analysed by RT-PCR.

3.3 Results

3.3.1 Cloning and characterisation of the *X. tropicalis* and *X. laevis* PRNP and PRND cDNAs

3.3.1.1 Discovery of the most PRND in *X. tropicalis*

We searched the translated *X. tropicalis* genome database using a series of mammalian Dpl sequences as probes, as described in Methods, and identified 2 overlapping trace files containing the complete ORF. As the PRND gene had been reported so far only in mammals, and our group and others have established that it is not present in the available fish genomes (Premzl et al., 2004), this finding places amphibians as the earliest vertebrates known to contain the gene. We identified the *X. tropicalis* PRNP gene for PrP in the genome database without difficulty using the *X. laevis* PrP sequence (Strumbo et al., 2001) as query. No ESTs for *X. tropicalis* Dpl were found in databases, but there were several for PrP.

To confirm expression of PrP and Dpl in *X. tropicalis*, total RNAs from brain and testis were reverse-transcribed into cDNA and analysed by RT-PCR, using the F1/R1 and F2/R2 primer pairs (Appendix 2) specific to identified PRNP and PRND fragments, respectively. The resulting 400 bp fragments of the PRNP and PRND ORFs were

cloned and sequenced. Unexpectedly, Dpl expression was found not only in testis, but in brain as well; Dpl is not highly expressed in adult mammalian brain (Moore et al., 1999, Silverman et al., 2000).

To determine the structure of both genes, I amplified and sequenced cDNAs, and aligned them against genomic DNA. 5' and 3' RACE primers were constructed to amplify the complete *PRNP* and *PRND* cDNAs. To obtain better PCR products for cloning, I ran two consecutive PCRs with nested primers (**Appendix 1**). Amplification of the *PRNP* 5' end gave two PCR products of similar size, and a single product for the 3' end. The results for *PRND* indicated two variants of the 3' end and one variant of the 5' end. I aligned the sequences of these cDNAs against the *X. tropicalis* genomic DNA sequence. Both *PRNP* transcripts have three exons with total lengths 850 bp or 815 bp for the two 5' end variants. Both *PRND* transcripts have four exons with total lengths 1262 and 752 bp for the long and short 3' end variants, respectively (**Table 3-1, Figure 3-3**). As for mammalian (Lee et al., 1998) and marsupial (Premzl et al., 2005) *PRNP* and mammalian (Moore et al., 1999, Comincini et al., 2001) *PRND* genes, the entire ORF is contained within the last exon. Surprisingly, *PRNP* and *PRND* transcripts share the two first exons (**Figure 3-2(b), (c)**). As a result, the intron separating the second and third exons of *PRND* is 27,099 bp and contains exon 3 of *PRNP*. Because this sharing creates a chimera of two genes, I will refer to it as a chimeric transcript (*PRND-ch*). A non-chimeric *PRND* transcript, i.e. an analogue of mammalian *PRND* with its own transcriptional start site, was not found. Full nucleotide and deduced amino acid sequences for the *X. tropicalis* cDNAs are presented in **Appendix 5**.

3.3.1.2 Finding both chimeric and non-chimeric *PRNP* and *PRND* transcripts in *X. laevis*

RNAs from *X. tropicalis* tissues other than brain and testis were not available to conduct tissue expression studies. I continued amphibian *PRND* characterisation using the *X. laevis* frog model, as a colony of these is available in ANU. *X. laevis* cDNA coding for prion protein was known from earlier work (Strumbo et al., 2001), but as these authors had not been able to extend the 5' end, I decided to use 5' RACE to amplify the missing sequence. 5' RACE *PRNP* sequencing gave, as for *X. tropicalis*, two variants, one slightly shorter than the other (**Table 3-1**). The common *PRNP/PRND-ch* exon 1 was longer for *PRND-ch* than for both *PRNP* variants. I considered the possibility of RNA

degradation during the RACE procedure, and constructed a primer specific to the 5' end of the *PRND-ch* exon 1 (F8; **Appendix 1**). PCR amplification of exon 1 of *PRNP* with this primer and a primer specific to the ORF of *PRNP* (R4; **Appendix 1**) was performed. The 5' ends of *PRNP-1a* and *PRND-ch* were confirmed to be of equal length **Figure 3-2 (e), (g)**. However, this raises the question whether these transcripts (*PRNP-1a*, *PRNP-1a* (v1 and v2)) have alternatively spliced 5' UTRs or are just the result of RNA degradation.

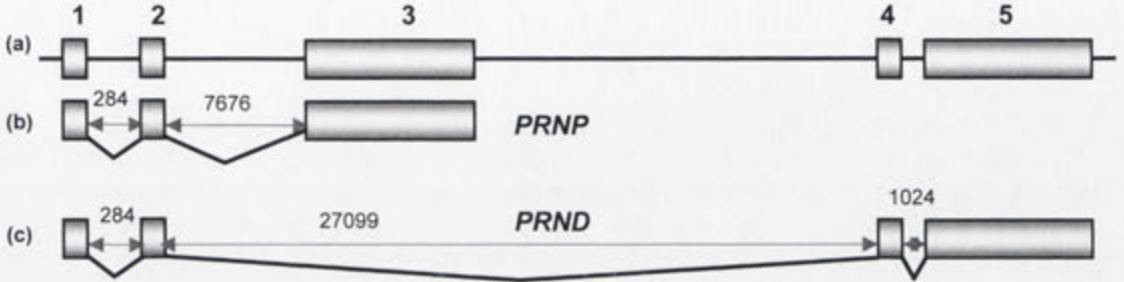
The *PRND* gene in *X. laevis* has not been characterised so far and genomic sequence was not available. Firstly, I amplified the *PRND* ORF fragment using primers specific to the *X. tropicalis* *PRND* sequence as described in Material and Methods (3.2.2). PCR was conducted with several combinations of primers specific to *X. tropicalis* *PRND* sequence with consecutive nested PCR reactions. Final PCR product was isolated, cloned into TOPO-TA vector and sequenced. Next, I constructed 5' and 3' RACE primers to amplify the complete *PRND* cDNA.

5' RACE for the *X. laevis* *PRND* gene produced more transcripts than for the *X. tropicalis* *PRND* gene. As for *X. tropicalis*, one was a chimeric transcript (*PRND-a* shown in **Figure 3-2 (g)**) sharing the first two exons with *PRNP* but there was no analogue to the third *X. tropicalis* exon (**Figure 3-2 (c)**). Two other transcripts were non-chimeric, one 328 bp longer than the other (**Figure 3-2 (h)** and **Table 3-1**). It is unclear if these transcripts (*PRND-a*, *PRND-b*) are alternatively spliced, have different promoters or are just the result of RNA degradation. Attempts to determine the intron size by amplifying genomic DNA between the non-chimeric exon and the *PRND* ORF were unsuccessful: hence, I suspect a long intron. Full nucleotide and deduced amino acid sequences for the *X. laevis* cDNAs are given in **Appendix 4**; in the absence of genomic sequence, the exon boundaries were predicted based on alignments of the *PRND* and *PRNP-1* sequences with those for *X. tropicalis*, and by aligning the *X. laevis* variants among themselves.

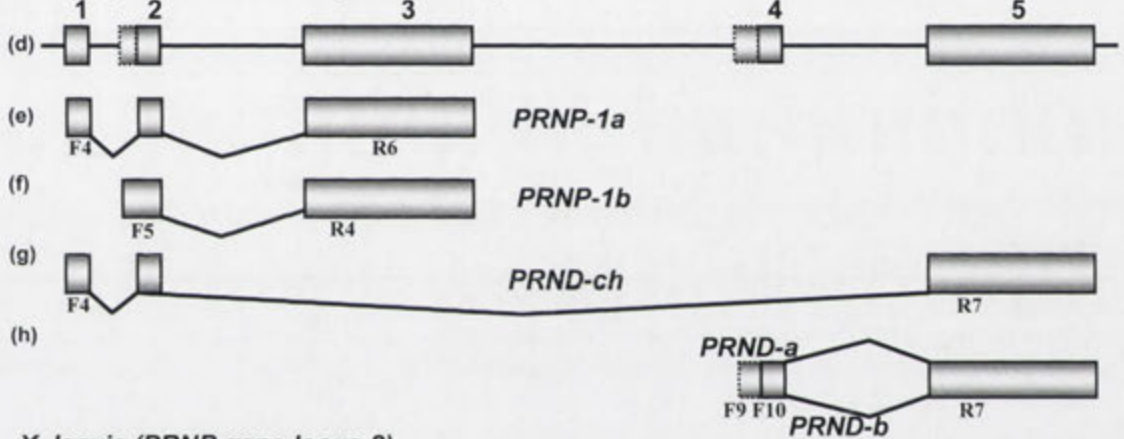
I used 3' RACE to evaluate the 3' UTR length of the *PRNP* and the *PRND*. I confirmed the unusually short, 80 bp, 3' UTR for *X. laevis* *PRNP* reported earlier (Strumbo et al., 2001). Similarly, the detected 3' UTR of *X. tropicalis* *PRNP* was only 30 bp, again much shorter than those of mammalian *PRNPs*. I detected two 3' UTR variants for both

X. tropicalis and *X. laevis* PRND (Table 3-1). My results do not confirm a very long *X. laevis* 3' UTR annotated by NCBI (BC094089) and cloned from oocyte cDNA library.

X. tropicalis (PRNP and PRND gene locus)



X. laevis (PRNP and PRND gene locus 1)



X. laevis (PRNP gene locus 2)

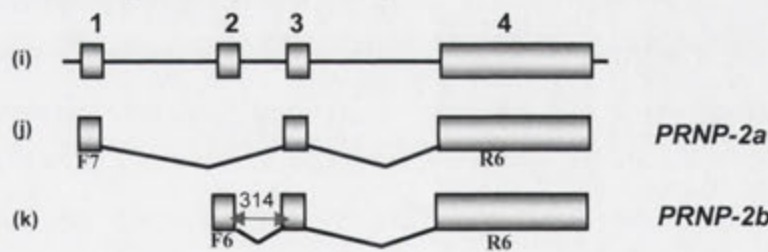


Figure 3-2. Transcripts coding for PRNP and PRND in *X. tropicalis* and *X. laevis*. Genomic contigs corresponding to PRNP/PRND gene loci are shown in (a), (d), and (i). Exons are numbered according to their relative positions in the corresponding locus. Exons 1, 2, and 4 from *X. tropicalis* locus and *X. laevis* locus 1 are non-coding; exon 5 is the coding exon. For *X. laevis* locus 2, exons 1, 2, and 3 are non-coding and exon 4 is the coding exon. Transcripts are shown for *X. tropicalis* PRNP (b), and *X. laevis* PRNP-1a and PRNP-1b with alternate start sites (e), (f). The first two non-coding exons of PRNP transcripts in both species are shared with PRND transcripts; these generate chimeric structure (c), (g). *X. laevis* PRND-a (dotted line) and PRND-b (solid line) have alternative non-chimeric start sites (h). Transcript structures for *X. laevis* PRNP at the second PRNP locus are shown in (j) and (k). Intron sizes for *X. tropicalis*

are shown in base pairs. Intron sizes for *X. laevis* are not available, except when sequenced (**k**), because of absence of genomic sequence information. Exon sizes are summarised in **Table 3-1**. Primers used for transcripts 5' and 3' RACE amplification are summarised in Appendix 1; approximate positions of primer pairs for *X. laevis* tissue expression analysis (F4/R6 for PRNP-1a, F5/R4 for PRNP-1b, F7/R6 for PRNP-2a, F6/R6 for PRNP-2b), F4/R7 for PRND-ch, F9/R7 for PRND-a, F10/R7 for PRND-b) are indicated.

Table 3-1. Summary of exon sizes for *X. tropicalis* and *X. laevis* PRNP and PRND transcripts. Exons are numbered according to their relative positions in the PRNP/PRND locus (**Figure 3-2**). Exon 1 and exon 2 are non-coding and shared between PRNP and chimeric PRND-ch in both *Xenopus* species. Coding exons are exon 3 for PRNP, PRNP-1a and PRNP-1b, exon 4 for PRNP-2a and PRNP-2b, exon 5 for PRND, PRND-ch and PRND-a, b. Alternative 5' and 3' UTRs are marked as v1 and v2.

	<i>X. tropicalis</i>		<i>X. laevis</i>					
	PRNP	PRND	PRNP-1a	PRND-ch	PRNP-1b	PRND-a, b	PRNP-2a	PRNP-2b
Exon1	79v1 44v2	64	142v1 115v2	208	-	-	44	-
Exon2	60	60	63	63	145	-	-	99
Exon3	712	-	741	-	649	399v1 77v2	56	56
Exon4	-	55	-	-	-		990	990
Exon5	-	1075v1 573v2		632v1 569v2		633v1 569v2		

Polyadenylation (polyA) signals were predicted by using recently developed PolyApred software (Ahmed et al., 2009). PolyApred is based on a support vector machine method and is expected to predict polyA signals with higher accuracy than other methods (Ahmed et al., 2009). A putative polyA signal is non-canonical for *X. laevis* PRNP-1a and PRNP-1b (TATAAA) and absent in *X. tropicalis* PRNP (**Figure 3-3 (a), (b)**). Nevertheless, a putative polyA signal is present in genomic DNA sequence downstream of the experimentally predicted *X. tropicalis* 3' UTR. Comparison of regions situated downstream of *X. tropicalis* and *X. laevis* PRNP ORFs (source: genomic DNA for *X. tropicalis* and BC094089 for *X. laevis* PRNP locus 1) reveals the conservation of the downstream region and indicates that the 3' UTR might be longer than detected (**Figure 3-3 (c)**). Unfortunately, attempts to extend the *X. tropicalis* PRNP 3' UTR to include the polyadenylation site were unsuccessful. The PolyApred software failed to predict polyA signals for both *X. tropicalis* and *X. laevis* PRND, although visual sequence analysis

identifies a canonical AATAAA signal for *X. tropicalis* PRND (**Figure 3-3 (d)**) and a non-canonical ACTAAA signal for *X. laevis* PRND (**Figure 3-3 (e)**) (complete only for the longer variant) (**Table 3-1**).

Amino acid sequences of *X. tropicalis* and *X. laevis* PrPs (**Figure 3-4**) show very strong conservation over the complete protein (88.9% sequence identity) apart from two short indels in the N-terminal region. Much greater sequence variation is observed between *X. tropicalis* and *X. laevis* Dpl sequences with only 69.2% sequence identity, but the lengths differ by only one amino acid (**Figure 3-4**).

(a) 3' end of *X. tropicalis* PRNP transcripts

TGTGTATCACTGAGTACAGGAGAGGATCGGGATTTAAGGTGCTCTCTAACCCTTGGCTGA 780
TCCTCACTATCACTCTTTTGTGTTACTTTGTGATAGAGTGAACGAGGGAAGGCCAAATG 840
TATGTATATAG 851

(b) 3' end of *X. laevis* PRNP-1a transcripts

ACAGGAGAGGATCGGGATTCAAAGTGTCTCTAAACCTTGGCTGATCCTTACTATCACTC 840
TCTTTGTTTACITTTGTGATAGAGTGA TCAAAGGAAATATTAATAAAAGGCCAAATGTAT 900
GTATATATAGAGAGAGTATAAACCGATTCTGAACTGTCCGTCTCA 946

(c) Downstream sequence conservation between *X. tropicalis* PRNP and *X. laevis* PRNP-1a

XtPRNP	CCTTGGCTGATCCTCACTATCACTCTPTTTGTTTACTTTGTGATAGAGTGA	827
XlPRNP	CCTTGGCTGATCCTTACTATCACTCTCTTTGTTTACTTTGTGATAGAGTGA	875

XtPRNP	-----GGAAGGCCAAATGTATGTATATAGAGATTTAAAGAGAA	878
XlPRNP	ATATTAATAAAAGGCCAAATGTATGTATATATAGAG-----AGAGAGTATAAA	931

XtPRNP	TGAAGTGTCTGTCTCAGCGCCAGCTTGACAGTCTCTGTTGTGATTCCCTTGC	938
XlPRNP	TGAAGTGTCTCGCTCTCAGCGCCAGCTTGACAGTCTCTATTGTGATTCCCT	991

XtPRNP	TGTTGGATCCAATGTGAATCCAGCCCAGCTTCTACCATCCGTAAGCAACTAT	998
XlPRNP	TGTTGGGTCCATGTAAATCCAGCCCAGCTTCTACTATCAGTTATCAACCTT	1051

XtPRNP	TACCTCTCGCTTCAATACTGACCGAAATCCCACTTCTGCATCAGTACGAT	1058
XlPRNP	TATATCTTGCTTCAATACCGAGCTGAATCCCACTTCTGCATCAACCACT	1111

XtPRNP	GACAACCTACATTTTGGACATCACAAATGGGCTTTATGTAACAGTGCATG	1117
XlPRNP	AACAACGCACCTTCTGACATTTGTAATGGGCTTTATGTAGCAGTGCATGT	1171

XtPRNP	AGATATCTTGAGGTTCCAGTACCGAAGAGCCACATTTTGTGAAGTTATGA	1176
XlPRNP	AGATACCTTGAGGTTCCCGCACCAAGAGCCCCGTTTCT-TTAATTTATGA	1230

XtPRNP	TGTAGAATCCGTCAACTGCAGATTTGTTTTCCATATCTTGGACACCC	1236
XlPRNP	TTTTTAATATCTTGTCACCTG-----TACCGATTGATACACAAC	1283

(d) 3' end of *X. tropicalis* PRND transcripts

AGTGACCTGCACTGGTGGATTGTGGGCTTTATAGGTGTCATGCATTTTTTTTTTTAAT 720
TAGGAAGGGGAGACTAAAGCCTAGGAATTCTGTattttatatgaactttaagaactaact 780
gtactagcccagaggttcagcagccctataacagcaatgatccaggccttcaaatttgtc 840
cacagcagctcttgatctcatcttgatcttttgagtgtcagtgacactgcacattctc 900
agtgtgcagggtctgtgttaagactacgccatctgtcatagaatgcacgtttctacaca 960
ggaatatactctaataataggaatataatgtgcacacccctactaatagttcagcctgc 1020
cagagatcaccaaggagttgcatagcctttttacaaacatatccttcattctaaaacttc 1080
atctgtgacttctaatagtcttataaaattacaacaggagcaatattatattctatattat 1140
atacatataagcagggaattatggctgtctgtgtgcacacagggaagtggaggtttggcc 1200
acggtttgttttcccttattataattgatgaataaaatcagggttaaactgg 1254

(e) 3' end of *X. laevis* PRND transcripts

TGTTATGAAAACATATGTAAAGCCTACCAATGGGGCAGTGACCTGCACAGGAGGACTGTGG 780
GTCTTCATAGGTGTCATTCACCTCCTTTTTTACTCAGAAAGGGAGTTAA TGGGAAC TA 840
AAGCCTGAATTCTGTATTCTTATATTGA ACTaaaatgtaccgccccagaggttcagcagc 900
tct 903

Figure 3-3. The 3' ends of PRNP and PRND in *X. tropicalis* and *X. laevis*. 3' UTRs for *X. tropicalis* and *X. laevis* PRNP and PRND were amplified by 3' RACE. Numbering starts from first nucleotide of cDNA sequence. Bases highlighted in grey correspond to the 3' end of protein coding sequences. Boxed bases were predicted as putative polyadenylation signals. (a) PolyA signal is absent in *X. tropicalis* PRNP. (b), (e) PolyA signals are non-canonical in *X. laevis* PRNP and PRND. (d), (e) PolyA sites in *X. tropicalis* and *X. laevis* PRND were predicted by visual search only, *X. laevis* PRND v2 does not have polyA site. The 3' end PRND variants are represented by lower case letters. (c) Conservation of sequence downstream ORF in *X. tropicalis* and *X. laevis* PRNP; ends of experimentally defined 3' UTRs are underlined; putative polyA signals are boxed.

(a)

XtPrP MLRSLWTSVLVLISLVCALT VSSKKSGSGKSKTGGWNSGSNRNPNYPGGYGWNTGGNTGGS 60
XlPrP MPQSLWTCVLVLISLICTLT VSSKKSGGGKSKTGGWNTGSNRNPNYPGGY----PGNTGGS 56
* :****:*****:*****:*****:*****:*****:*****:*****

XtPrP WG-QPYNPSSGGNNFNKNQWKPKPKSKTNMKAVAVGAAAGAIIGGYMLGNAMGRMSYHFSNPM 119
XlPrP WGQQPYNPSSG----YNKQWKPKPKSKTNMKSVAIIGAAAGAIIGGYMLGNAVGRMSYQFNNPM 112
** ***** :*****:*****:*****:*****:*****:*****:*****

XtPrP EARYYNDYYNQMPERVYRPMYRGEEHVSEDRFVTD CYNMSVTEYIIKPAEGKNTSEVNQL 179
XlPrP ESRYYNDYYNQMPNRVYRPMYRGEEYVSEDRFVRDCYNMSVTEYIIKPTEGKNNSELNQL 172
*:*****:*****:*****:*****:*****:*****:*****:*****:*****

XtPrP ETRVKSQIIREMCITEYRRSGGFKVLSNPWLILITITLFVYFVIE 223
XlPrP DTTVKSQIIREMCITEYRRSGGFKVLSNPWLILITITLFVYFVIE 216
:* *****

(b)

XtDpl1 MGRQNLFSCLILLLLILYCSLSSPRRAAS-SKKISKTTDLRSGAKRRPKVTNSPALGDLS 59
XlDpl1 MERQNVFSCLILLVLILYCSLSCPRRSGSGIKKYFKISDLRSGAKRKSVAHSPVLGHLF 60
* *:*****:*****:*****:*****:*****:*****:*****:*****:*****

XtDpl1 FRGRAINVNFNLTTEESELYTANLYSFPDGLYYPRPAHLSGAGGTDEFISGCLNTTIERNK 119
XlDpl1 FRSKELDVNLFNLTTEEELYTENLYRFPDGLYYPWRSQLNDAAGTEEFMNGCLNTTVERNK 120
*.: *:*****:*****:*****:*****:*****:*****:*****:*****:*****



Figure 3-4. Protein sequence comparison for *X. tropicalis* and *X. laevis* PrP and Dpl. (a) *X. tropicalis* (XtPrp) and *X. laevis* (XlPrp) PrP amino acid sequences show high level of conservation (identity 89%). (b) *X. tropicalis* (XtDpl) and *X. laevis* (XlDpl) Dpl amino acid sequence comparison shows relatively fast divergence after the duplication (identity 69%).

3.3.1.3 Alternatively spliced PRNP mRNA and PRNP expressed from the duplicated gene in tetraploid *X. laevis*

Several PRNP sequences were found in the *X. laevis* EST database (NCBI 31081059, 31094494, 31092138, 31092434, 43385448, and 46388120). This analysis was done in collaboration with another PhD student, Nagesh Chakka. In the following, I refer to a form described in the original report (Strumbo et al., 2001) as PRNP-1a/PrP1a and to the novel variant forms as PRNP-1b/PrP1b, PRNP-2a/PrP2a and PRNP-2b/PrP2b (**Figure 3-2**). We ran several PCRs to confirm expression of the transcripts and map their sequences. Initially, we supposed that the transcription start site for PRNP-1b is the same as for PRNP-1a. This implies that exon 1 could be missing in the sequence found in EST-database and that exon 2 could be spliced alternatively. To test this hypothesis we ran a PCR with primers F8/R9 (**Appendix 1**) complementary to exons 1 and 2 of PRNP-1a, respectively, using reverse transcribed RNA as a template. Two amplification products are expected if alternative splicing applies, but we found only one corresponding to the PRNP-1a variant. Next, we amplified the intron between the first and second exons of PRNP-1a with the F4/R9 primer pair (**Appendix 1**), using genomic DNA. It was found that sequence corresponding to exon 2 of PRNP-1a is common for both forms but that the PRNP-1b form has additional sequence preceding exon 2 (**Figure 3-2 (f)**). The PRNP-1b variant alternative start site was confirmed by PCR with the F5/R4 primer pair (**Appendix 1**). We found no corresponding PRND transcript with this extended exon 2 start site.

Expression of two PRNP-2 variants was confirmed by PCR using primer F6 for PRNP-2a, F7 for PRNP-2b, and the common primer R6 (**Appendix 1**). The EST sequences (NCBI 46388120 and 31081059 for PRNP-2a and PRNP-2b, respectively) indicated

they had different first exons but a shared second exon; the sequences were characterised by a series of PCR reactions. By amplification of genomic DNA between the first exon (exon 2) of *PRNP-2b* (primer F7) and the shared exon 3 (primer R5) the size of a small intron (314 bp) was determined but there was no PCR product from attempted amplification between the first exon (exon 1) of *PRNP-2a* (primer F6) and the shared exon 3 (primer R5), suggesting its intron is either large or rich in repeat sequence. The structures of the *PRNP-2* variants are shown in **Figure 3-2 (i), (j), (k)**. Significant differences in amino acid sequence for the ORF suggest *PRNP-2* is a retained duplicate, the result of the *X. laevis* allopoloidy; this is discussed later. We found no evidence for a retained duplicate of *PRND*.

3.3.1.4 *X. laevis* tissue expression

Novel genes are often expressed in tissue-specific patterns similar to those of the parental gene. In *X. laevis*, *PRNP-1a* and *PRNP-1b* are expressed in multiple tissues, but highest in brain (**Figure 3-5**). Although *PRND* expression does not follow this pattern closely, there is a high level of the chimeric transcript in brain. The results for *PRND-a* and *PRND-b* show that expression from the developing *PRND* promoter starts initially in testis, eye and pancreas. I will discuss the evolution of *PRND* tissue expression patterns in more detail, comparing the three vertebrate models, in later sections of this chapter.

The alternative *PRNP-1b* transcript shows a more restricted tissue distribution pattern than that of the ubiquitous expression of *PRNP-1a*. Similarly, the *PRNP-2a* transcript expressed from the duplicated gene is more specific than its ubiquitous alternative *PRNP-2b*. All *PRNP* variants are highly expressed in brain, eye, and heart. Differential distribution of *PRNP-1b* and *PRNP-2a* in other tissues and, interestingly, absence in testis, could indicate the importance of regulation of PrP functions through alternative promoters and thus provide an explanation for retention of the duplicate.

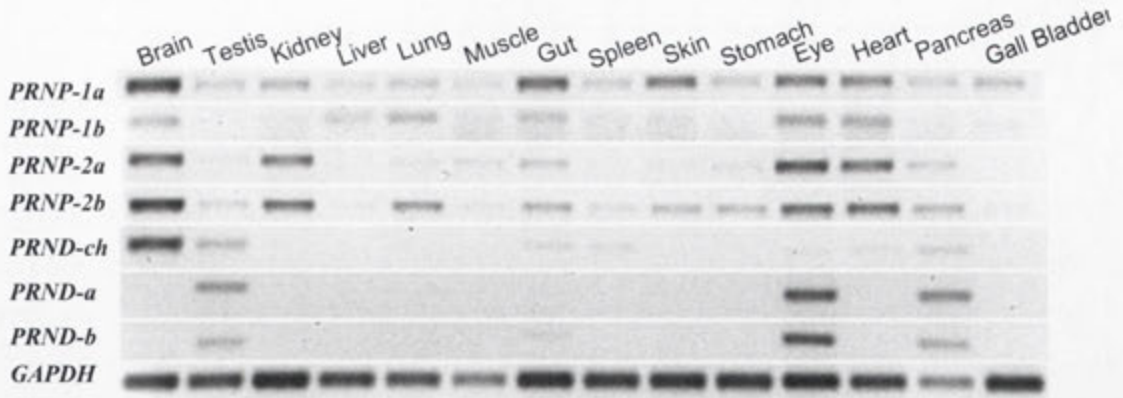


Figure 3-5. *X. laevis* tissue expression. RNA from adult male *X. laevis* tissues was analysed by RT-PCR for the presence of alternative *PRNP* and *PRND* transcripts. *PRNP-1a*, *PRNP-1b*, chimeric *PRND-ch*, and non-chimeric *PRND-a* and *PRND-b* are expressed from the first *PRNP/PRND* locus. *PRNP-2a* and *PRNP-2b* are expressed from the second, duplicated, locus. *GAPDH* was used as a positive control. Original gel images with a molecular marker and no template control can be found in **Appendix 3**.

3.3.2 Characterisation of *M. domestica* *PRNP* and *PRND* cDNAs

3.3.2.1 Cloning of *M. domestica* *PRNP* and *PRND* cDNAs

Isolation and characterisation of the *M. domestica* *PRNP* and *PRND* genes enables us to understand further their evolution after the duplication. Although marsupial *PRNP* sequence has been reported for brush-tailed possum (Windl et al., 1995) and tammar wallaby (Premzl et al., 2005), marsupial *PRND* sequence has not been characterised so far. Cloning of the 3' end of *M. domestica* *PRNP* was first reported by our group (Premzl et al., 2005). However, the 5' UTR was not characterised and the 3' UTR was found by screening a *M. domestica* testis cDNA library. Strangely, the last 11 amino acid residues of the reported ORF do not accord with the expected sequence pattern. Consequently, I decided to clone and characterise *M. domestica* *PRNP* as well as *PRND*.

Partial sequences of *PRNP* and *PRND* genes were found by searching the *M. domestica* genome database. Based on this initial information, I constructed primers (**Appendix 1**) to perform 5' and 3' RACE using total RNA isolated from *M. domestica* testis tissue.

3.3.2.2 Genomic organisation of PRNP/PRND locus

Mapping the *PRNP* and *PRND* cDNA sequences onto the genomic sequence located the intron/exon boundaries. Three variant transcripts for *PRNP* and five variants for *PRND* were identified (**Figure 3-6, Appendix 6**). The *PRNP/PRND* locus spans over 942126 bp and contains 8 exons (**Figure 3-6, Table 3-2**), two of which code for the PrP and Dpl proteins. The intergenic region between *M. domestica PRNP* and *PRND* genes is 68359 bp; this is 5.5 times larger than mouse *Prnp/Prnd* intergenic region (12433 bp).

3.3.2.3 PRNP structure and 3' end alternative splicing

M. domestica PRNP has two exons: a single non-coding 103 bp exon and an ORF-containing second exon. This is the same as the *PRNP* structure from another model marsupial, tammar wallaby (*Macropus eugenii*) (Premzl et al., 2005). Amplification of the 3' UTR by 3' RACE produced two transcripts (**Figures 3-6 (b), (d), 3-6**). I expected to find the cytoplasmic polyadenylation element (CPE) at the 3' end, which is characteristic of mammalian *PRNP* sequence (Premzl et al., 2005). However, one of the RACE-amplified *PRNP* transcripts (*PRNP-1*, **Figure 3-7**) did not contain the CPE motif (**Figure 3-6 (b)**). Analysis of the second RACE-amplified transcript (*PRNP-3*, **Figure 3-7**) revealed the presence of a 1607 bp intron interrupting an alternative 3' end of the ORF (**Figure 3-6 (c)**). The resulting protein-coding sequence has an alternative C-terminal end, not conserved with other marsupials (**Figure 3-8**). Nevertheless, as this C-terminal region is in the signal sequence for GPI attachment, no difference in the mature proteins is expected. This unusual protein was initially described as the unique marsupial PrP (Premzl et al., 2005) and annotated (NCBI accession number AY659989). The presence of the intron was not recognized due to lack of genomic sequence information at the time of their study.

I estimated the relative abundance of the *PRNP-1* and *-3* transcripts by RACE PCR band comparison (**Figure 3-7**). The ORF1 transcript without the CPE motif (*PRNP-1*) is the most abundant; the ORF2 transcript with the CPE (*PRNP-3*) is rare. It was amplified only by a second, nested, 3' RACE PCR. The *PRNP-2* transcript with the CPE motif was not amplified by RACE; from this result I assume this transcript is scarce. I confirmed the presence of the *PRNP* with the CPE motif by running RT-PCR with primers LmdPrP1/RmdCPE167 covering *PRNP* sequence from exon 1 to the CPE motif (**Appendix 2**).

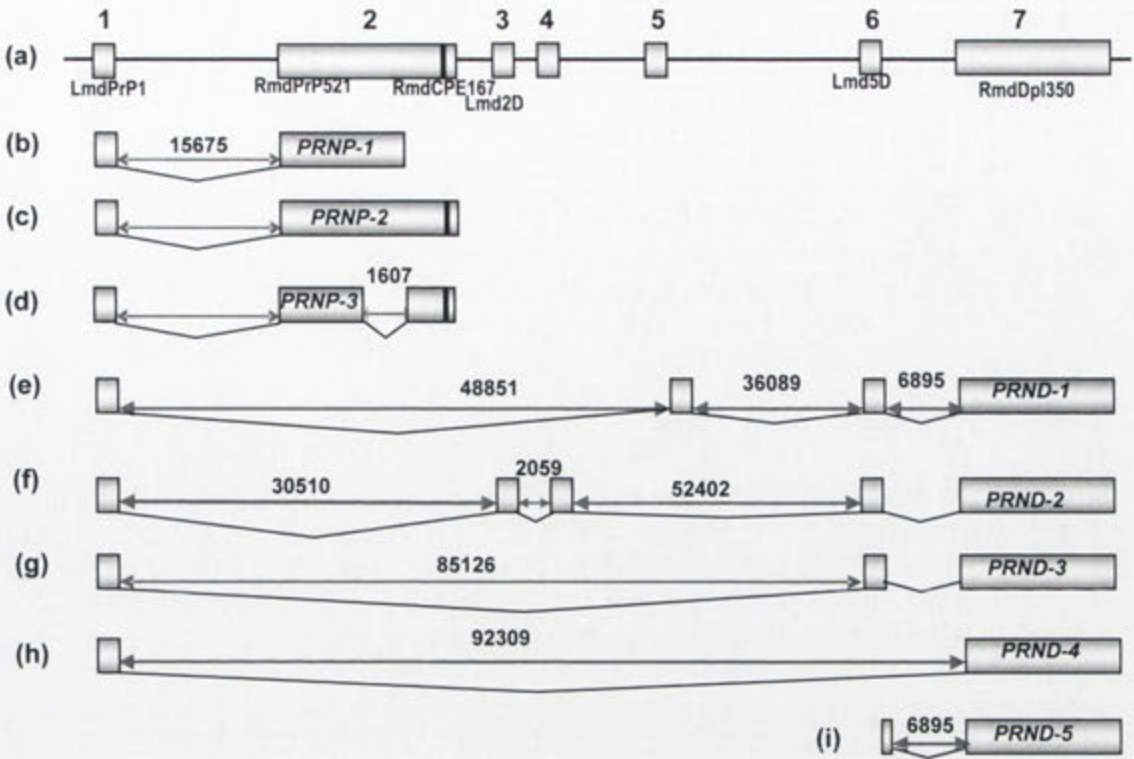


Figure 3-6. Transcripts coding for *PRNP* and *PRND* in *M. domestica*. (a) Genomic *PRNP/PRND* contig, exons are numbered. Exons 1, 3-6 are non-coding; exon 2 is coding for *PRNP*; exon 7 is coding for *PRND*. (b) *PRNP-1* variant without the CPE motif. (c) *PRNP-2* variant with the CPE motif. (d) *PRNP-3* variant with intron in the coding region and the CPE motif. (e) to (h) Chimeric *PRND* transcripts with alternative intergenic exons. (i) Non-chimeric *PRND*. The CPE motif is marked as a black line at the 3' end of *PRNP-2* and *PRNP-3* transcripts. The length of the introns is in base pairs. The length of exons is summarised in **Table 3-2**. Primers used for transcripts 5' and 3' RACE amplification are summarised in **Appendix 1**; approximate positions of primer pairs for *M. domestica* tissue expression analysis (LmdPrP1/RmdPrP521 and LmdPrP1/RmdCPE167 for *PRNP*, LmdPrP1/RmdDpl350, Lmd2D/RmdDpl350 and Lmd5D/RmdDpl350 for *PRND*) are indicated.

Table 3-2. Exon sizes of *M. domestica* transcripts. Exons are numbered according to their relative positions in the *PRNP/PRND* gene locus (**Figure 3-6 (a)**). Exon 1 is non-coding and shared between *PRNP* and chimeric *PRND-1*, *PRND-2*, *PRND-3*, and *PRND-4*. Coding exons are exon 2 for *PRNP* and exon 7 for *PRND*. Alternative 3' UTRs are marked as v1, v2 and v3.

Exon number	1	2	3	4	5	6	7
Exon size, bp	103	2269v1 2594v2 (793+203)v3	90	65	186	288	1125v1 1714v2

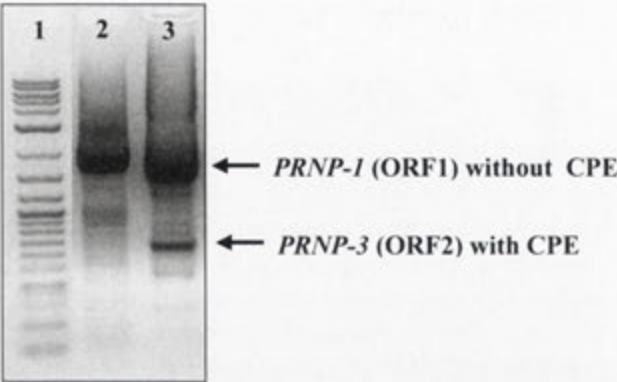


Figure 3-7. *M. domestica* PRNP transcripts with alternative coding regions (ORF1 and ORF2) amplified by 3' RACE and 3' nested RACE. Line 1, 2-log DNA Ladder (NEB); Line 2, ~1.7 kb product of 3' RACE; Line 3, ~1.5 kb product of 3' nested RACE corresponding to PRNP-1 without CPE motif (Figure 3-6 (b) and ~650 bp product corresponding to PRNP-3 with CPE motif and an intron inside the coding region (Figure 3-6 (c)).

MdORF1	MGKIH ^L GYWFLALFIMTWS ^{DL} TLCKKPKPRPGGGWNSGGNRYPGQSGGWGHPQGGGTNWG	60
MdORF2	MGKIH ^L GYWFLALFIMTWS ^{DL} TLCKKPKPRPGGGWNSGGNRYPGQSGGWGHPQGGGTNWG	60

MdORF1	QPHAGGSNWGQPRPGGSNWGQPHPGGSNWGQPHPGGSNWGQAGSSYNQKWKPDKPKTNMK	120
MdORF2	QPHAGGSNWGQPRPGGSNWGQPHPGGSNWGQPHPGGSNWGQAGSSYNQKWKPDKPKTNMK	120

MdORF1	HVAGAAAAGAVVGGLGGYMLGSAMSRPIMHFGNDYEDRYRENQYRYPNQVMYRPIDQYN	180
MdORF2	HVAGAAAAGAVVGGLGGYMLGSAMSRPIMHFGNDYEDRYRENQYRYPNQVMYRPIDQYN	180

MdORF1	NQNNEFVHDCVNITVKQHTTTTTTKGENFTETDIKIMERVVEQMCITQYQNEYRSAYSVAF	240
MdORF2	NQNNEFVHDCVNITVKQHTTTTTTKGENFTETDIKIMERVVEQMCITQYQNEYRSAYSVAF	240

MdORF1	FSAPPVTL ^{LLLLSFLIFL} VS-----	260
MdORF2	FSAPPVTL ^{LLLLSFLIFL} LPDAHSVEAT	269
	*****:	

Figure 3-8. Comparison of two *M. domestica* PrPs with variant ORF. The difference in the C-terminal sequence is boxed. The GPI-attachment site is indicated by the arrow.

3.3.2.4 Alternative *PRND* transcription in *M. domestica*

Five *PRND* transcripts were found in total: four chimeric (**(e)-(h)**), and one non-chimeric (**(i)**) (**Figure 3-6**). The non-chimeric *PRND* transcript shares its first, non-coding, exon with three of the alternative chimeric transcripts. Although exon 1 (exon 6 in the genomic contig diagram, **Figure 3-6**) of the non-chimeric transcript is shorter (66 bp) than the overlapping exon of the chimeric transcripts (288 bp) I am not able to exclude the possibility that this transcript is a product of degradation and that the full non-chimeric transcript (i.e. complete exon 6) was not found, or that it does not exist (i.e. it is a degradation product of a chimeric transcript). I assessed the relative abundance of the five *PRND* isoforms by RACE PCR band-size analysis (results not shown). The major chimeric transcript is *PRND-3* (**Figure 3-6 (g)**). The 3' RACE results showed a long and a short variant of the 3' UTR (**Table 3-2**). It is unclear whether all transcripts have both variants or whether the 3' UTR variations are transcript-specific. Nucleotide sequences and deduced amino acid sequences of *PRND* cDNA for the chimeric and non-chimeric transcripts are summarised in **Appendix 6**.

3.3.2.5 Tissue expression of *M. domestica PRNP* and *PRND* transcripts

To analyse tissue distribution of transcript variants identified by RACE, I performed RT-PCR analysis of total RNA derived from a range of tissue samples (brain, kidney, liver, ovary, lung, muscle, gut, stomach, heart, and pancreas of female and testis and eye of male).

PRND expression was not detected in liver, lung, muscle and stomach. *PRND-1* and *PRND-2* transcripts (**Figure 3-6 (e) and (f)**) are testis-specific (**Figure 3-9**); PCR amplification gave products of approximately the same size, not distinguishable on a 1% agarose gel. *PRND-3* (**Figure 3-6 (g)**) shows the highest level of expression of all transcripts. It is present in brain, testis, ovary, kidney, eye, and heart, with highest expression in testis, eye and heart (**Figure 3-9**). *PRND-4* (**Figure 3-6 (h)**) shows highest expression in brain and heart and is expressed in the same tissues as *PRND-3* plus also gut and pancreas (**Figure 3-9**). It is not possible to assess tissue specificity of the presumed non-chimeric transcript *PRND-5* (**Figure 3-6 (i)**) because it does not have unique sequence that is not shared with chimeric transcripts *PRND-1*, -2, and -3. Consequently, only four *PRND* variants will be amplified simultaneously when primers

specific to common exons 6 and exon 7 are used (*PRNP/PRND* contig numbering, **Figure 3-6 (a)**).

PRNP expression analysis showed that *PRNP-3* (**Figure 3-6 (d)**) is testis-specific. Both *PRNP-1* and *PRNP-2* are present in all tissues (**Figure 3-9**). Nevertheless, the transcript with the CPE motif (*PRNP-2*) is much less abundant than that without the CPE motif as assessed with different PCR cycling conditions. *PRNP-1* amplification is much stronger after 30 cycles than *PRNP-2* amplification. However, as the primers used to detect *PRNP-1* cannot discriminate between *PRNP-1* and *PRNP-2* the result shows amplification of both transcripts. Better evidence that *PRNP-2* is a minor transcript is its absence in 3' RACE amplification (**section 3.3.2.3**). It was detected by direct PCR amplification using primers specific to the ORF (RmdPrP521) and the CPE motif (RmdCPE167) (**Appendix 2**); this result is shown in **Figure 3-9**.



Figure 3-9. *M. domestica* tissue expression for *PRNP* and *PRND*. RNA from *M. domestica* tissues was analysed by RT-PCR. 40 cycles of the amplification was performed for *PRND* transcripts, except for *PRNP-1* and *PRNP-2* where 30 cycles of amplification was used. *GAPDH* was used as a positive control. Note that *PRNP-1* result includes *Prnp-2* due to common primers. *PRNP-2* result is from specific primer. Original gel images with a molecular marker and no template control can be found in **Appendix 3**.

3.3.3 *Prnd* transcription in mouse

3.3.3.1 Chimeric transcription in mouse

Chimeric *Prnd* in mouse was initially reported by two independent groups (Moore et al., 1999, Li et al., 2000a). Both groups discovered three alternative chimeric *Prnd* transcripts in *Prnp* knockout mice (**Figure 3-10**). One chimeric isoform was also detected in wild-type mouse brain by RT-PCR (Moore et al., 1999), but not by *in situ* hybridisation (Li et al., 2000a). In order to confirm the expression of chimeric *Prnd* in

wild-type mice, I first attempted to amplify the chimeric 5' UTR by 5' RACE. However, this resulted in non-chimeric transcripts only, probably because of the high relative proportion of the non-chimeric *Prnd* form in mouse. Despite unsuccessful RACE, I was able to amplify chimeric transcripts by RT-PCR using primers specific to *Prnp* exon 1 (LmmPex1-5 or LmmPex1-94, **Appendix 1**) and the *Prnd* ORF (RM587, **Appendix 1**). I found three alternative chimeric forms that share two non-coding exons with *Prnp* (**Figure 3-11**). Two of the chimeric *Prnd* transcripts (**Figure 3-11 (c), (d)**) have small exons in the intergenic region (**Table 3-3**).

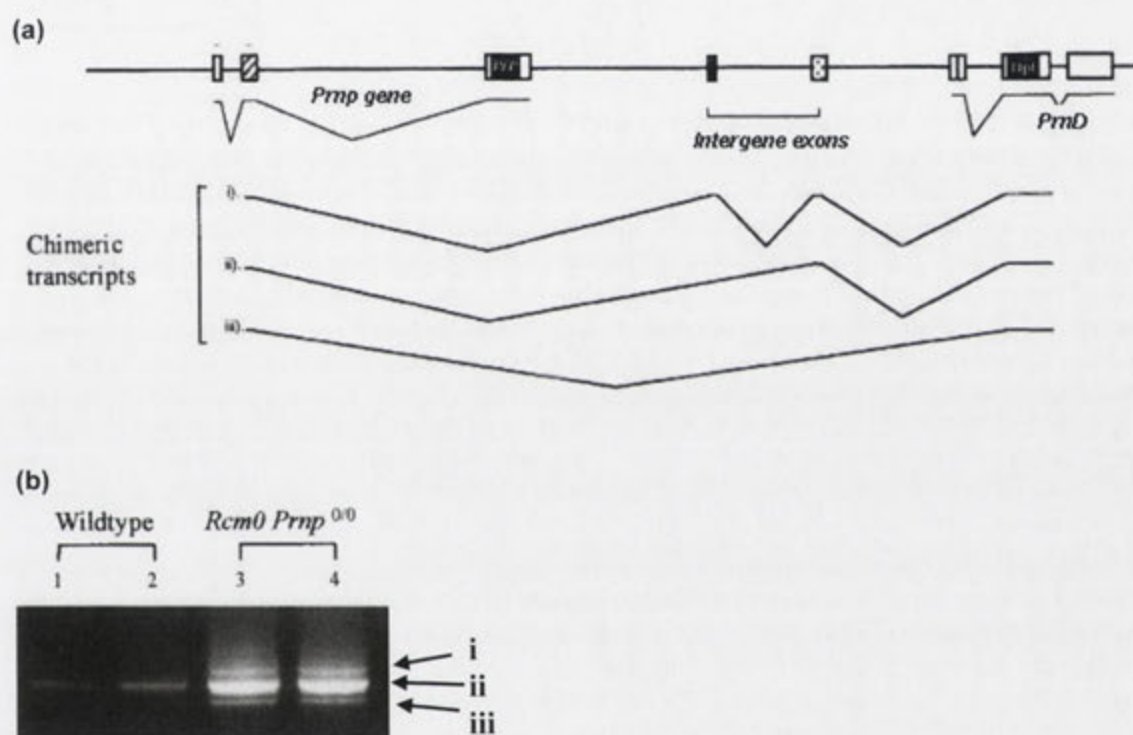


Figure 3-10. Chimeric *Prnp*/*Prnd* transcripts discovered in *Prnp*^{0/0} knockout mice (Moore et al., 1999). (a) Overview of *Prnp*/*Prnd* gene region with the most abundant forms *Prnp* and *Prnd* transcripts (shown at the top of the figure, non-chimeric *Prnd* is incorrectly presented as three-exon transcript) and chimeric transcripts (i-iii) shown underneath. (b) RT-PCR detection of chimeric *Prnd* in adult wildtype and knockout mouse brain, chimeric transcripts i-iii are indicated.

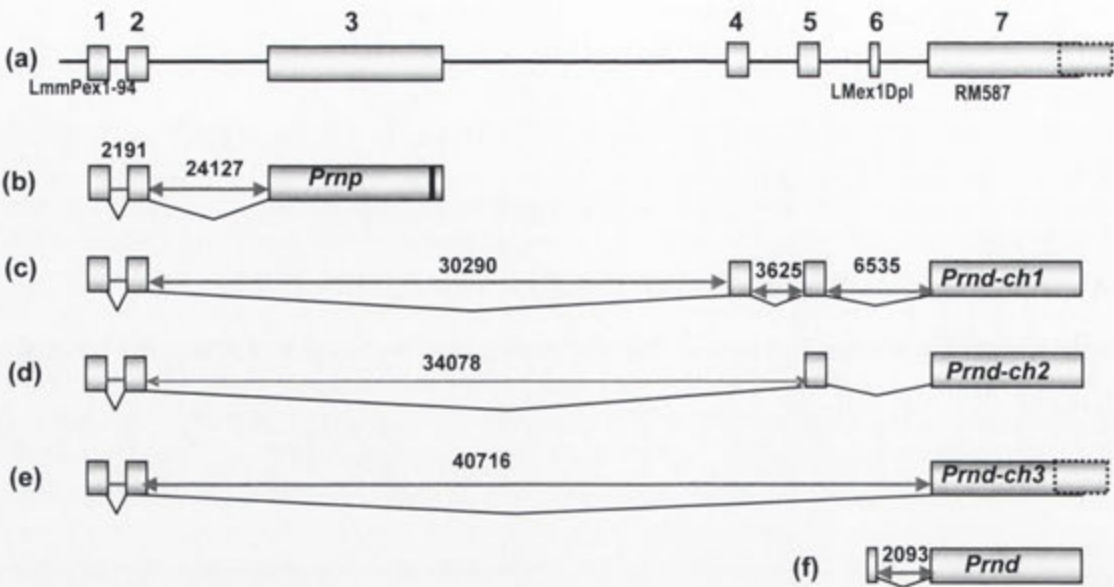


Figure 3-11. Transcripts coding for *Prnp* and *Prnd* in mouse. (a) Genomic *Prnp/Prnd* contig with exons numbered. Exons 1, 2, 4-6 are non-coding; exon 3 is coding for *Prnp*; exon 7 is coding for *Prnd*. (b) *Prnp* with the CPE motif, marked by a black line at the 3' end. (c), (d), (e) Chimeric *Prnd* transcripts with alternative intergenic exons. (f) Non-chimeric *Prnd*. The longer alternative *Prnd* 3' UTR is represented with dotted lines. The length of the introns is in base pairs. The length of exons is summarised in **Table 3-3**. Primers used for transcripts amplification are summarised in **Appendix 1**; approximate positions of primer pairs for mouse tissue expression analysis (LmmPex1-94/RM587 for chimeric *Prnd*, LMex1Dpl/RM587 for non-chimeric *Prnd*) are indicated.

Table 3-3. Exon sizes of mouse *Prnp* and *Prnd* transcripts. Exons are numbered according to their relative positions in the *Prnp/Prnd* gene locus (**Figure 3-11 (a)**). Exons 1 and 2 are non-coding and shared between *Prnp* and chimeric *Prnd-ch1*, *Prnd-ch2*, and *Prnd-ch3*. Coding exons are exon 3 for *Prnp* and exon 7 for *Prnds*. The source of the information for exons 1-3, 6, and 7 is the NCBI web site (<http://www.ncbi.nlm.nih.gov>) and for exons 4 and 5 is this sequencing of the chimeric *Prnd* PCR-amplified products.

Exon number	1	2	3	4	5	6	7
Exon size, bp	76	98	2012	163	103	55	3122v1 1884v2

3.3.3.2 Mouse *Prnd* tissue expression

I assessed *Prnd* tissue expression in BALB/C mice of different ages (6, 12 and 18 weeks old) using forward primer LmmPex1-94 and reverse primer RM587 for chimeric *Prnd*-ch (1,2,3) (**Figure 3-11 (c), (d), (e)**) and LMex1Dpl/RM587 for non-chimeric *Prnd* (**Figure 3-11 (f)**) (**Appendix 1**). *Prnd* tissue distribution was the same for all three stages. Expression analysis for 18-week-old mouse is presented in **Figure 3-12**. Although it is generally accepted that *Prnd* is predominantly expressed in testis and only weakly in other organs in adult animals (Behrens et al., 2002, Comincini et al., 2001, Paisley et al., 2004, Peoc'h et al., 2002), I found expression in all tissues (just traces in brain and liver). At first I thought that *Prnd*, which is known to be expressed in all mouse embryonic tissues (Li et al., 2000b), is still not switch off in 6-week-old mice but further experiments with older mice showed this theory is not correct and that expression is present in all tissues. This result confirms the more ubiquitous *Prnd* expression which was reported by the early workers in mouse (Moore et al., 1999, Li et al., 2000b), and for bovine and ovine tissues (Rondena et al., 2005, Tranulis et al., 2001).

Prnd-ch2 (**Figure 3-11 (d)**) is the most abundant of the three alternative chimeric transcripts. It is highly expressed in brain, testis, muscle, heart, and stomach and less so in kidney, liver, lung, eye, and thymus (**Figure 3-12**). The *Prnd*-ch3 transcript (**Figure 3-11 (e)**) is less abundant with highest expression in brain, testis, kidney, liver, and muscle. The *Prnd*-ch1 transcript (**Figure 3-11 (c)**) appears to be rare, with highest expression in testis (**Figure 3-12**).

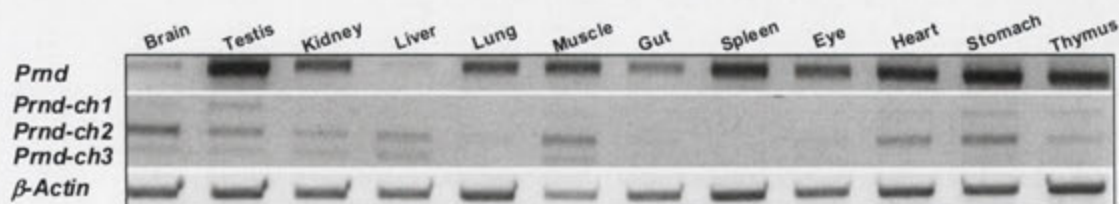


Figure 3-12. Mouse *Prnd* transcript expression. RNA from 18-week-old mouse tissues was analysed by RT-PCR. 40 cycles of amplification was performed for chimeric *Prnd* transcripts and 35 cycles for non-chimeric *Prnd*. β -Actin was used as a positive control. Original gel images with a molecular marker and no template control can be found in the **Appendix 3**.

3.3.4 Analysis of translation of mouse chimeric *Prnd* transcripts by polysome binding assay

Conservation of chimeric transcripts during evolution raises the question of their functional significance. We might assume that a transcript is functional if it is translated. In the case of *Prnd*, protein translated from both chimeric and non-chimeric transcripts are the same, so antibody detection cannot tell us from which transcript the protein comes. I used a polysome-binding assay to investigate whether chimeric transcripts are translated, as an initial means to assess functionality of the alternative *Prnd* transcripts. This assay is based on distribution of mRNA into different fractions on a sucrose gradient depending on the amount of ribosomes attached to it. The background to this method is given in Chapter 2.

3.3.4.1 Polyribosome distribution after centrifugation

Mouse brain tissue samples were homogenised and lysed under mild conditions and the resultant cytoplasmic extract loaded onto a sucrose gradient and centrifuged as described in Methods. mRNAs distribute among the different fractions of the gradient depending on the amount of ribosomes attached to them. Twenty fractions were collected and total RNA isolated and analysed for transcript presence. Fractions 1-9 contain cytoplasmic untranslated mRNA and free ribosomes. RNA in fractions 10-20 is associated with monosomes and polyribosomes and is thus translated. To exclude the possibility that mRNA is associated with other protein complexes, I conducted a control experiment where lysis buffer was supplemented with EDTA to dissociate polysomes. This causes polysome-bound, translated mRNA to be redistributed towards cytoplasmic fractions, as shown from a literature example in **Figure 3-13 (a)**.

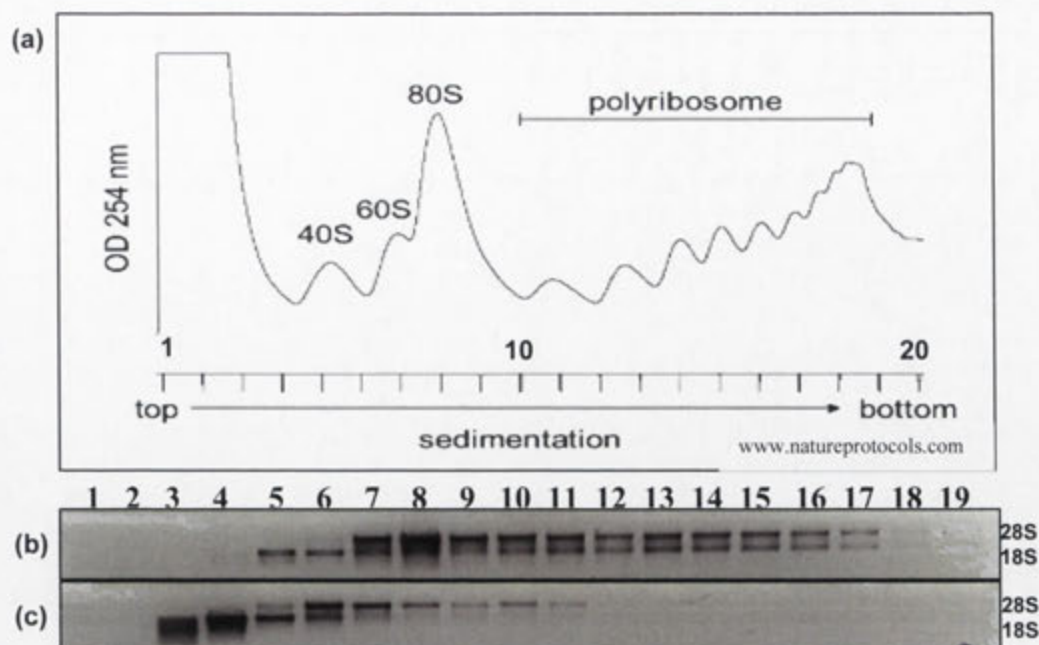


Figure 3-13. Polyribosome distribution after centrifugation on a sucrose gradient. (a) Literature example of the polysomal profile (www.natureprotocols.com). **(b)** Total RNA quality control (my work). Mouse brain total RNA was isolated from cytoplasmic fractions 1-9, monosomal fractions 10-11, and polysomal fractions 12-19, and resolved on a 1% agarose gel. Visible bands correspond to 28S (top band) and 18S (lower band) ribosomal RNA. **(c)** Same total RNA sample centrifuged after EDTA treatment. Polysomes are dissociated and RNA is absent in polysomic fractions.

3.3.4.2 Chimeric *Prnd* is translated in mouse brain

I compared mouse 12-week-old adult and 1-day-old postnatal brain tissues for *Prnd* translation (**Figure 3-14**). Total RNA from similar fractions were pooled, converted to cDNA and analysed by PCR for the transcript presence. Presence of chimeric transcripts in polysomal fractions implies they are translated. After polysome dissociation by EDTA, transcripts can be detected only in cytoplasmic fractions; this confirms their initial binding to ribosomes and not to other protein complexes. For 1-day-old brain all *Prnd* transcripts are present in higher polysomal fractions than in adult brain. This indicates they have more ribosomes attached and, thus, the translation rate for *Prnd* in developing mouse brain is higher than in adult brain. Three possible chimeric transcript variants (**Figure 3-11**) are detected in developing brain (although the *Prnd-ch1* band is very weak) but only one, *Prnd-ch2* (**Figure 3-11 (d)**), in adult brain (**Figure 3-14**). As a control for assay reliability I assessed *Prnp* and non-chimeric *Prnp* distribution in the gradient fractions. These two genes are known to be translated in brain at different levels (Li et al., 2000b, Moore et al., 1999). As expected, there is a high level of *Prnd*

translation in 1-day-old brain, much less in adult brain and a high level of *Prnp* translation in both cases.

These experiments confirm that chimeric transcripts are translated and possibly have some function, especially in development.



Figure 3-14. Ribosomal distribution of *Prnd* and *Prnp* transcripts in adult and postnatal mouse brain is shown by RT-PCR analysis. After separation on a sucrose gradient, RNA from similar fractions was pooled, converted to cDNA and analysed. Fractions 4-9 are cytoplasmic RNA; 10-11 are monosomes; 12-18 are polysomes. After EDTA treatment, a shift towards cytoplasmic fractions is observed. Strongly visible double PCR bands in the chimeric *Prnd*, 1-day-old brain panel, correspond to *Prnd-ch2* and *Prnd-ch3*. Only *Prnd-ch2* is visible in the adult brain panel.

3.3.4.3 Translation of alternative *PRNP* and *PRND* transcripts in *X. laevis*

Brains and testes of *X. laevis* were collected, processed, and analysed as described in Material and Methods (3.2.9). RT-PCR analysis of isolated cytoplasmic and polysomal fractions revealed that all four *PRNP* transcript variants (Figure 3-2) in frog are translated (Figure 3-15). However, the expression analysis of *PRND* (both chimeric and non-chimeric) was inconclusive; the RT-PCR results were not reproducible. Possible reasons are that the low level of *PRND* expression leads to barely detectable RT-PCR products or that incomplete removal of contaminants, such as heparin, during RNA purification from sucrose buffer might affect cDNA synthesis. Nevertheless, the preliminary results suggest that chimeric *PRND* is not translated in adult frog, as the transcript was detected in cytoplasmic fractions only. Further analysis is needed to

assess whether *PRND* transcripts are translated and, if so, define which tissues and development stage.

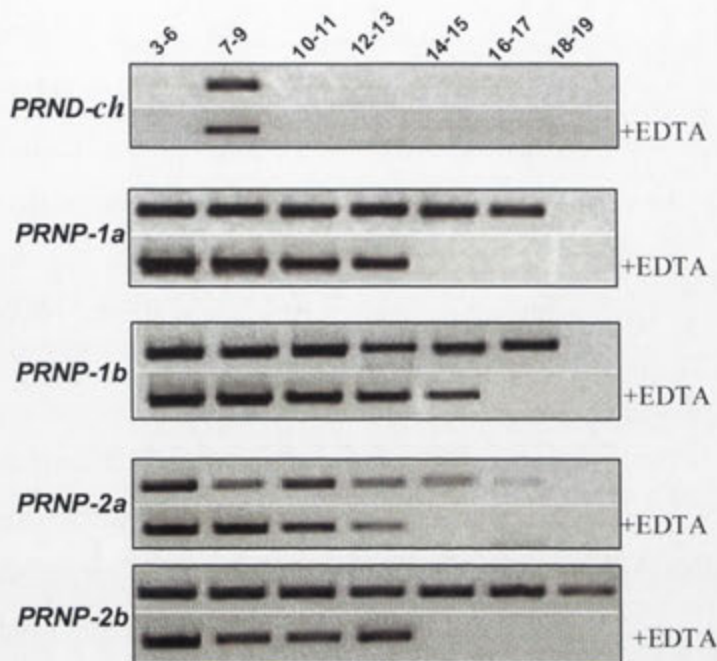


Figure 3-15. Ribosomal distribution of *PRND* and *PRNP* transcripts in *X. laevis* is shown by RT-PCR analysis. After separation on a sucrose gradient, RNA from similar fractions was pooled, converted to cDNA and analysed. Fractions 4-9 are cytoplasmic RNA; 10-11 are monosomes; 12-18 are polysomes. EDTA treatment results in shift towards cytoplasmic fractions. Chimeric *PRND* is in the cytoplasmic fraction. All *PRNP* transcripts are associated with ribosomes.

3.4 Discussion

3.4.1 Expression of chimeric *PRND*

Cloning and characterisation of previously unknown amphibian and marsupial *PRNP* and *PRND* transcripts and their comparison with analogous mouse sequences allows tracking of their evolution. A gene duplication event in an amphibian ancestor resulted in the birth of the *PRNP* paralogue gene, *PRND*. In both *Xenopus* species and in *M. domestica*, *PRND* is transcribed from its parental gene promoter and its transcripts share non-coding exons with *PRNP*. A transcript with chimeric *PRND* structure persists in mouse despite dominant transcription from its own promoter. It is noteworthy that although chimeric *Prnd* was initially discovered in *Prnp* knockout mice and reported as a low abundance transcript in wild-type mouse and rat (Moore et al., 1999), this interesting finding received no further attention.

The results for the recently duplicated *PRND* in amphibians show it is expressed mainly as a chimeric transcript sharing its 5' UTR with its parental gene. I take this as evidence of promoter sharing and *PRND* reliance on *PRNP* regulatory DNA elements. The highly brain-specific tissue expression pattern (**Figure 3-5**) for both *PRND-ch* and *PRNP* is consistent with this idea. However, another attribute of *PRNP* expression – ubiquity – is not present. Thus, although *PRND-ch* is expressed in several other tissues, such as testis, pancreas, intestine, spleen, eye and heart, there is no expression in kidney, liver, lung and muscle (**Figure 3-5**). This differences in tissue distribution despite a common promoter suggests the operation of other regulatory mechanisms for chimeric *PRND* expression, for example chromatin remodelling (Kornblihtt, 2006).

Surprisingly, the results show chimeric expression is more ubiquitous in mammals, *Prnd-ch* being expressed in multiple mouse tissues (**Figure 3-12**). It is possible, that functions requiring coregulated expression from shared *PRNP* promoter in amphibian and marsupial resisted in mammals. However, in interpreting this finding account needs to be taken of work of Ford et al. (2002), showing that, despite the *Prnp* wide tissue distribution, it is selectively expressed at high levels only in discrete subpopulations of cells, especially those of the “neuroimmune network”. Vertebrate evolution has led to the development of more complex neuronal and immune systems in mammals. It is possible that chimeric *Prnd* started in neural cells and then acquired a wider expression pattern with the evolution of the neuroimmune network.

It has been suggested that PrP and Dpl have antagonistic effects on neuronal cell survival, PrP being neuroprotective and Dpl neurotoxic (Moore et al., 2001, Qin et al., 2006b). Colocalisation of PrP and Dpl at the cell surface in neuronal cells points at their possible interaction (Massimino et al., 2004). This functional interaction can regulate apoptotic cell death (Sakudo et al., 2005). Simultaneous expression of *PRNP* and *PRND* from the common promoter, even though level of chimeric *PRND* expression is relatively low in mammals, might be necessary to provide an additional control over the balance between apoptosis and survival.

The most likely mechanism for formation of the chimeric *PRND* is by bicistronic transcription (**Figure 3-16, Section 3.4.2**). Polycistronic transcription of genes organized in operons occurs frequently (~25%) in the nematode *Caenorhabditis elegans* genome (Blumenthal, 1998). It is common in *Drosophila* as well (Andrews et al., 1996,

Nozawa et al., 2005, Brogna and Ashburner, 1997, Wang et al., 2000). The possible significance of this kind of coordinated transcription in mammals has received little attention but some cases of bicistronic transcription in mice and humans are reported (Sakuma-Takagi et al., 1999, Li et al., 2006b, Lee, 1991, Mcmillan and Singer, 1993, Szabo et al., 1994). Comprehensive bioinformatic analysis of sequenced genomes has provided new insights into chimeric transcription. Analysis of the human genome revealed that 4-5 % of tandem genes could be spliced into chimeric mRNAs (Parra et al., 2006). Transcripts of more than 20 gene pairs in the human genome share the 5' UTR (Yang et al., 2008).

Chimeric transcription might also occur as a consequence of “promoter hitchhiking” by retroposed genes (Vinckenbosch et al., 2006). Evolution of *PRND* might have some analogies with expression of retroposed genes. Thus, if *PRNP* duplication did not include the regulatory DNA region, *PRND* would have needed to rely on the regulatory DNA elements of its parental gene in order to survive.

The long intergenic intron between the duplicated genes, *PRNP* and *PRND*, poses additional problem for transcription. The distance between exon 1 of *PRNP* and the ORF exon of *PRND* is 28.6 kb in frog, 43 kb in mouse and 92.4 kb in opossum. The studies of Vinckenbosch et al. (2006) showed that retroposed genes might be able to use the promoter or regulatory DNA elements from a neighbour gene and be transcribed if the median distance is about 22.9 kb but that retrocopies are transcriptionally silenced at a median distance of 38.7 kb. It thus appears that *PRNP/PRND* gene pair chimerism challenges the transcription machinery and is truly remarkable.

3.4.2 Mechanisms for alternative splicing of the bicistronic transcript

My results indicate that after bicistronic transcription of the *PRNP* and *PRND* genes the initial pre-mRNA bicistronic transcript is processed to give alternative mature monocistronic transcripts. Possible mechanisms for an alternative splicing model in *C. elegans* have been proposed (Alfonso et al., 1994) for transcripts of the *unc-17* and *cha-1* proteins processed from a bicistronic pre-mRNA. Similarly to the *PRNP/PRND* pair, the two *C. elegans* genes share the first exon but in this case the *unc-17* coding region is entirely contained within the intron of the *cha-1* gene. The proposed model has “rapid

and slow” splicing steps together with exon skipping events. *PRNP/PRND* splicing might follow the same steps: a possible scheme is shown in **Figure 3-16**.

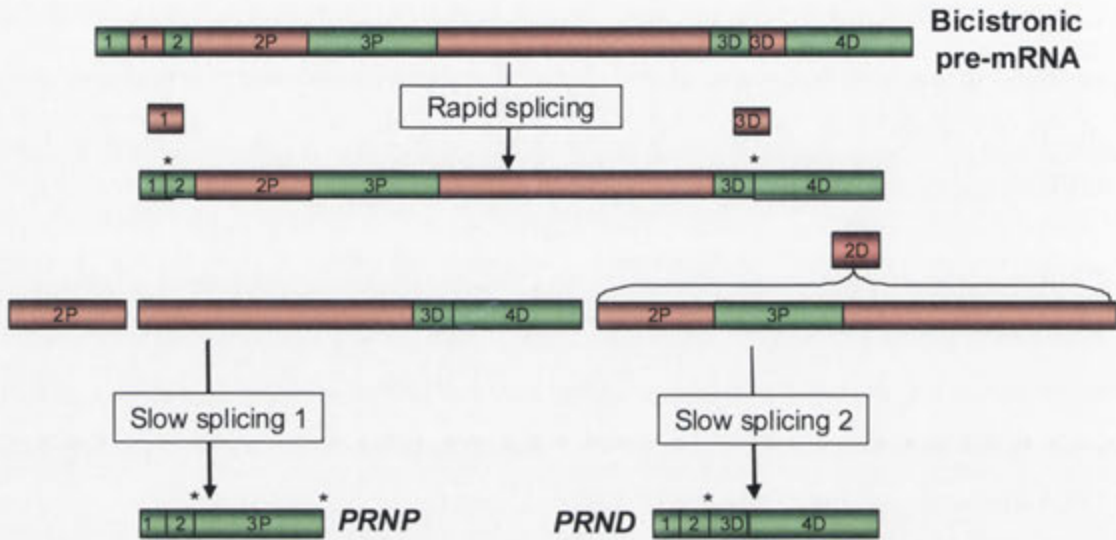


Figure 3-16. Schematic representation of bicistronic transcription leading to two different gene products, *PRNP* and chimeric *PRND*. The numbering corresponds to the exon/intron number; “P” and “D” represent PrP and Dpl, respectively. Exons are shown in green, introns in pink; splicing events are represented by “*”. Slow splicing 1 removes intron 2P, intergenic intron and *PRND* exons 3D and 4D resulting in *PRNP* mRNA. Slow splicing 2 removes intron 2D that harbours intron 2P, coding *PRNP* exon 3P, and the intergenic intron resulting in *PRND* mRNA. Figure adapted from thesis of N. Chakka.

Absence of a “canonical” AATAAA polyadenylation signal promotes chimeric transcription (Liu et al., 2001). The *PRNP* polyadenylation signal is canonical in eutherian mammals and marsupials (Premzl and Gamulin, 2007), but not in frogs. Although absence of the canonical polyadenylation signal may have promoted the initial chimeric transcription and pre-mRNA splicing soon after the duplication, conservation of chimeric transcription in mammals and marsupials suggests that factors other than the nature of the polyadenylation signal affect splicing.

A second possible reason for chimeric transcription is an inefficient splice acceptor. The removal of the splice acceptor from *Prnp* terminal intron in ataxic lines of *Prnp*^{0/0} mice resulted in chimeric *Prnd* (Moore et al., 1999, Li et al., 2000a). The idea that splicing elements in the *Prnp* terminal intron (intron 2P, **Figure 3-16**) are linked to bicistronic

transcription is supported by splice-site mutation research using the EGFP (enhanced green fluorescence protein) system (Yoshikawa et al., 2007). This *in vitro* transient transfection system with EGFP cloned downstream of *Prnp* mimics expression of a chimeric *Prnd*. Deletion of the splicing elements caused the expression of EGFP and fluorescence emission thus demonstrating functional linkage between chimeric *Prnd* expression and the *Prnp* exon 3 splice acceptor. The importance of the splice acceptor is underlined by another case of mammalian bicistronic transcription of genes sharing a promoter and the first two exons (Bailleul et al., 1997).

It is unclear if the *Prnd* transcript produced in this way is functional. It has been reported that only one ORF from an original bicistronic structure might be translated (Sakuma-Takagi et al., 1999). My investigation of the chimeric *Prnd* functionality using polysome binding experiments is discussed later.

3.4.3 Conservation of chimeric transcripts throughout evolution

Apart from the expected difficulties posed by the huge intron between the spliced genes, conservation of bicistronic transcription raises several questions:

- Why has chimeric *PRND* not disappeared during the evolution
- What is the reason for the parallel existence of chimeric and non-chimeric transcripts of *PRND* which, although having alternative 5' UTRs, share the same open reading frame, i.e. same confirmed or predicted protein?
- Does its retention indicate the necessity of additional 5' UTR regulatory elements for *PRND* expression?
- Do common promoter elements serve to coordinate transcription which might be desirable for correct function?

Alternative splicing increases the potential for protein complexity. Although most cases involve alternative splicing of a single gene, there are several examples reported in mammals of chimeric proteins containing domains (i.e. parts of ORFs) from two discrete proteins, translated from mRNA formed from bicistronic transcription of the two genes (Lee et al., 2006b, Parra et al., 2006, Akiva et al., 2006). However, in our case, only the 5' UTR is shared. Although such cases are reported in mammals (Yang et al., 2008), these genes tend to be expressed from a unique common promoter rather than

there being an option for the downstream gene to be transcribed efficiently from an “independent” alternative promoter.

Functionally related genes can be grouped in operons not only in prokaryotes but in eukaryotes as well (Blumenthal et al., 2002). To assess whether the chimeric transcripts were translatable, and thus functional, I performed polysome-binding assays. I could not find evidence of translation in adult *X. laevis*, the transcript being found in the cytoplasm only. It would be useful to repeat these experiments in embryonic (tadpole) tissues. However, our attempts of breeding frogs and raising tadpoles were confounded by irresolvable technical difficulties. But the results showed the mouse chimeric transcripts are polysome-bound in brain. I compared polysomal distribution in post-embryonic day 1 and adult mouse stages. The former is much higher; this might indicate the importance of chimeric transcription in development.

Coordinated bicistronic expression has been reported during development of chordate features in the ascidian tadpole larva (Swalla et al., 1999). It is interesting to note that bicistronic transcription of mammalian genes important for brain development and brain functions has also been reported (Sakuma-Takagi et al., 1999, Szabo et al., 1994, Lee, 1991, Eiden, 1998). Among reported genes, glutamic acid decarboxylase (GAD) is a well-known enzyme which catalyses the conversion of L-glutamate to γ -aminobutyric acid (GABA), the major inhibitory neurotransmitter in vertebrates (Krnjevic, 1970). Alternative splicing of a single exon during early stages of mouse development introduces overlapping stop-start codons and converts the original 67-kDa GAD into two tandemly arranged ORFs, coding for a 25-kDa leader peptide and a 44-kDa enzymatically active truncated GAD (Szabo et al., 1994). In a second case, expression of cholinergic neurotransmitters choline acetyltransferase (ChAT) and the vesicular acetylcholine transporter (VACHT), situated within the first intron of ChAT and sharing the first exon, is coordinated by common *cis*-acting elements (Eiden, 1998). Another bicistronic locus is *Lass1-Gdf1* (Longevity-assurance gene 1 homolog ceramide synthase 1 (initially discovered as Uog-1, upstream of Gdf1) and growth differentiation factor 1 (Szabo et al., 1994, Wang et al., 2007). Gdf-1 is a transforming growth factor- β (TGF- β) family member involved in brain development (Rankin et al., 2000). Mutations in the *Lass1* gene cause cerebellar ataxia and Purkinje cell degeneration (Zhao et al., 2011). Furthermore, disruption of *Gdf1* increases expression of *Lass1* (Bengtsson et al.,

2008). This is an interesting association with *Prnp/Prnd* genes, where removal of *Prnp* causes *Prnd* overexpression and late-onset ataxia (Moore et al., 1999).

Antagonistic effects of PrP and Dpl have been described (Brini et al., 2005). Overexpression of PrP is detrimental for cells and causes neurodegeneration (Westaway et al., 1994). It is possible, that coordinated expression of Dpl might offset neurodegeneration by antagonising PrP. Dpl protein might be expressed from a chimeric transcript because coordinated expression is necessary for regulation of the apoptosis/survival balance, especially under some circumstances, e.g. cell stress.

Another possible explanation of conservation of a chimeric *PRND* transcript is translational regulation from alternative 5' UTRs. A long 5' UTR might carry an internal ribosome entry site (IRES) to initiate cap-independent translation when eIF4G (5'cap) initiation machinery is inhibited. This mechanism of regulation has been proposed to play a crucial role during development (Van Der Velden and Thomas, 1999), apoptosis (Graber and Holcik, 2007, Riley et al.), Alzheimer's disease (Beaudoin et al., 2008), in stress responses (Thomas and Johannes, 2007), and cell cycle control (Pyronnet et al., 2000). Chimeric *PRND* might provide Dpl protein through IRES when normal Dpl expression is negatively affected.

Such additional mechanisms for regulation to maintain Dpl expression under as-yet-uncharacterised conditions might be important for functions well beyond Dpl's characterised major function in spermatogenesis.

3.4.4 New exon creation

Characterisation of *PRNP/PRND* transcripts from amphibian to mammals revealed the appearance of multiple intergenic exons in the marsupial, *M. domestica*. None of the intergenic exons was found in *X. laevis* and only one exists in the single chimeric transcript in *X. tropicalis* (**Figure 3-2**). *M. domestica* has four chimeric transcripts with none, one, two or three intergenic exons (**Figure 3-6**). Three chimeric transcripts in mouse have none, one or two intergenic exons (**Figure 3-11**).

The new exons in the *PRNP/PRND* intergenic area created variable 5' UTRs for chimeric *PRND* in marsupial and eutherian mammal (**Table 3-4**). They correspond to minor *PRND* transcripts, possibly because of inefficient splicing as noted. It is difficult

to assess if these new exons have been preserved by selection. The intergenic exons are not conserved between *M. domestica* and mouse, although the distance between the last intergene exon and the coding exon is similar for *M. domestica* and mouse (6895 bp and 6535 bp, respectively, **Figure 3-6** and **Figure 3-11**).

The presence of non-conserved exons in alternatively spliced genes indicates recent exon creation events (Modrek and Lee, 2003, Zhang and Chasin, 2006). It has been shown that new exons can be created from transposable elements (Sorek, 2007) and that nearly all such exons are alternatively spliced (Sorek et al., 2002). This correlates with high transposable element content in *M. domestica* (**Table 3-5**). Mapping the repeat elements to the intergenic region of the *PRNP/PRND* contig (**Figure 3-6 (a)**) revealed that *M. domestica* exon 3 overlaps with a SINE element, exon 4 overlaps with an LTR, and exons 5 and 6 are situated next to LINEs (downstream and upstream, respectively). Creation of such new 5' UTR exons may have increased opportunities for regulatory versatility of the developing *PRND* gene.

Table 3-4. Numbers of 5' UTR exons and total lengths of 5' UTRs for *M. domestica* and mouse *PRND* transcripts. *PRND*-1, 2, 3, and 4 are *M. domestica* transcripts (**Figure 3-6**). *Prnd*-ch1, 2, and 3 and *Prnd* are mouse transcripts (**Figure 3-11**).

Transcript	<i>PRND</i> -1	<i>PRND</i> -2	<i>PRND</i> -3	<i>PRND</i> -4	<i>Prnd</i> -ch1	<i>Prnd</i> -ch2	<i>Prnd</i> -ch3	<i>Prnd</i>
Number of 5' UTR exons	3	4	2	1	4	3	2	1
5' UTR length, bp	577	546	391	103	440	277	174	55

Table 3-5. Transposable element and GC content in *X. tropicalis*, *M. domestica*, and mouse *PRNP/PRND* locus.

Species	Locus length ^a (bp)	SINE ^b (%)	LINE ^b (%)	LTR ^b (%)	DNA ^b (%)	Total ^b (%)	GC (%)
<i>X. tropicalis</i>	29684	0	0	0	5	9	41
<i>M. domestica</i>	94126	10	37	14	2	66	39
Mouse	46174	6	4	33	2	46	46

^a Length from exon 1 to end of exons 5, 7 and 7 for *X. tropicalis*, *M. domestica* and mouse, respectively (see **Figures 3-2, 3-6** and **3-11** for locus architecture).

^b SINE, short interspersed element; LINE, long interspersed element; LTR, long terminal repeat (retrotransposons); DNA, DNA transposon.

3.4.5 Evolution of *PRND* transcriptional regulation

As discussed, survival of the newly duplicated *PRND* gene has been facilitated by borrowing regulatory DNA elements from its parental gene, *PRNP*. In order to evolve new functions, a newly duplicated gene needs to develop new regulatory DNA elements (Ohta, 2003). The next evolutionary step started with the acquisition of the alternative promoter and appearance of the non-chimeric *PRND* transcript. As introduced in the last section, potential sources of regulatory DNA sequences are transposable elements (Jordan et al., 2003, Van De Lagemaat et al., 2003, Bejerano et al., 2006, Santangelo et al., 2007). Considering the high transposable element content in *M. domestica* (Table 3-5) and their involvement in the intergenic exons in the chimeric transcripts, the alternative *PRND* promoter may have emerged from insertion of LTRs or LINEs which later evolved into an independent promoter that allowed introduction of novel tissue specificity and specialised functions.

I was unable to detect a non-chimeric transcript in *X. tropicalis*. However, it is present in testis, eye and pancreas in *X. laevis*. The tissue distribution of *M. domestica* non-chimeric *PRND* could not be assessed because it had common exons with the chimeric transcripts. Non-chimeric *Prnd* is expressed in all mouse tissues, with the highest level in testes.

It is interesting to note that initial transcription of non-chimeric *PRND* arose in the most recent common ancestor of frog and mammal initially in testis and that a dominant testis-specific function of Dpl has developed in mammals. Retrogenes, which can serve as an example of genes developing new promoters to survive, are mostly expressed in testis with function in spermatogenesis even though their parental genes show ubiquitous tissue distribution (Marques et al., 2005). This is consistent with the observation that genes involved in male reproduction have a high rate of evolution (Swanson and Vacquier, 2002). The fact that in mammals non-chimeric *PRND* gained broader expression pattern points to acquisition of more diverse regulatory DNA elements during *PRND* evolution.

A specific example of correlation of wider *PRND* expression with more complex functionality in mammals is the high *PRND* expression in the complex mammalian heart, which has four chambers and its absence in the three-chambered *Xenopus* heart. It

is tempting to speculate that *PRND* might have evolved as a part of the chamber formation system and/or cardiac conduction system and might be specifically expressed in Purkinje fibers that conduct impulses between two sets of chambers (Moorman and Christoffels, 2003).

3.4.6 3' UTR evolution: acquisition of CPE motif

3' UTRs of mRNA are involved in translational control, mRNA stability and degradation. The cytoplasmic polyadenylation element (CPE) motif resides in the 3' UTR, a few nucleotides upstream of the polyadenylation signal. It is important for translational control and modulates translational repression and mRNA localisation, which are involved in germ cell development (Charlesworth et al., 2004), synaptic plasticity (Wu et al., 1998) and long term memory formation (Darnell, 2003).

Our research group has previously reported conserved 3' UTR sequence motifs in mammalian and marsupial *PRNP* include the CPE motif and nuclear-specific polyadenylation signal site (Premzl et al., 2005) (**Figure 3-17**).

The CPE motif is absent in the reported frog *X. laevis* cDNA sequence (Strumbo et al., 2001). I confirmed this absence in *PRNP* transcripts of both *X. laevis* and *X. tropicalis*. Unexpectedly, I found three different *PRNP* 3' ends in *M. domestica*: long without the CPE motif, long with the CPE motif and short with the CPE motif. The short sequence with the CPE motif has an intron inside the ORF and is expressed in testis only.

The *PRNP* transcript with the CPE motif is present in all tissues, but is much less abundant than that without it, as shown by PCR amplification (**Figure 3-7**). My re-analysis of the previously described and annotated marsupial *PRNP* sequence (Premzl et al., 2005) revealed that the presence of the intron was missed at the time due to lack of genomic sequence information. This is the only finding of an intron in the ORF-coding exon of a PrP family gene and it is unclear if it has any functional significance. Although BAC sequencing of the marsupial tammar wallaby *PRNP* gene revealed the CPE motif, it was found in *PRNP* cDNA isolated from the pouch young cDNA library (Premzl et al., 2005).

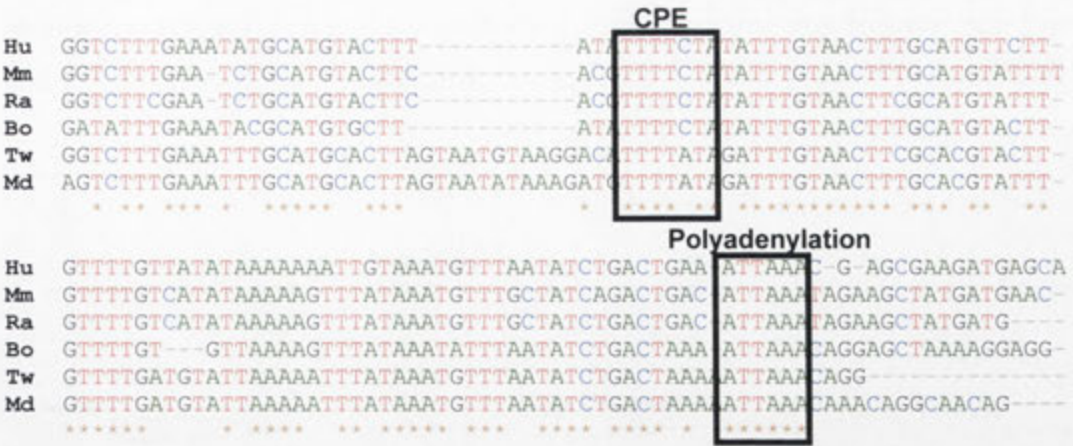


Figure 3-17. Conserved sequence elements in *PRNP* 3' UTRs for mammals and marsupials. Hu, human; Mm, mouse; Ra, rat; Bo, bovine; Tw, tammar wallaby; Md, opossum *M. domestica*.

The average length of 3' UTR sequences has increased during evolution, suggesting that their utilisation might contribute to organism complexity (Mazumder et al., 2003). The *PRNP* and *PRND* 3' UTRs in *X. tropicalis* are relatively short compared with those in mammals: 536 bp for *PRND* (opossum 1162 bp; mouse 1337 bp; human 2633 bp) and 48 bp for *PRNP* (opossum 1476 bp; mouse 1233 bp; human 1591bp).

In summary: combined results from the literature and my work indicate that *PRNP* with the CPE motif is absent in *Xenopus sp.*, appears in *M. domestica* as a minor transcript and is the only transcript in mammals. Phylogenetic analysis of *PRNP* paralogs, *SPRN* and *PRND*, suggests that the CPE motif is not an ancestral motif that has been lost in amphibians. As demonstrated by Premzl et al. (2004), *PRNP* arose after *SPRN* duplication. The CPE motif is absent in both *SPRN* and *PRND* and, thus, absent in the common ancestor before *SPRN* duplication (which resulted in the *PRNP* gene) and in the common ancestor before *PRNP* duplication (which resulted in the *PRND* gene). This strongly suggests that acquisition of the CPE motif allowed *PRNP* to develop additional post-transcriptional control that made it more effective for new specialised functions in the mammalian nervous system.

3.4.7 Genomic context of *PRNP/PRND* locus

I analysed the genomic environment of the *PRNP* and *PRND* locus using the annotated mouse genome and more recent *X. tropicalis* and *M. domestica* genomes. In mammals,

genes situated downstream of *PRND* are *RASSF2* and *SLC23A1*. This organisation, as well as direction of expression, has been conserved during evolution over 360 million years from amphibian to mammals (**Figure 3-18**).

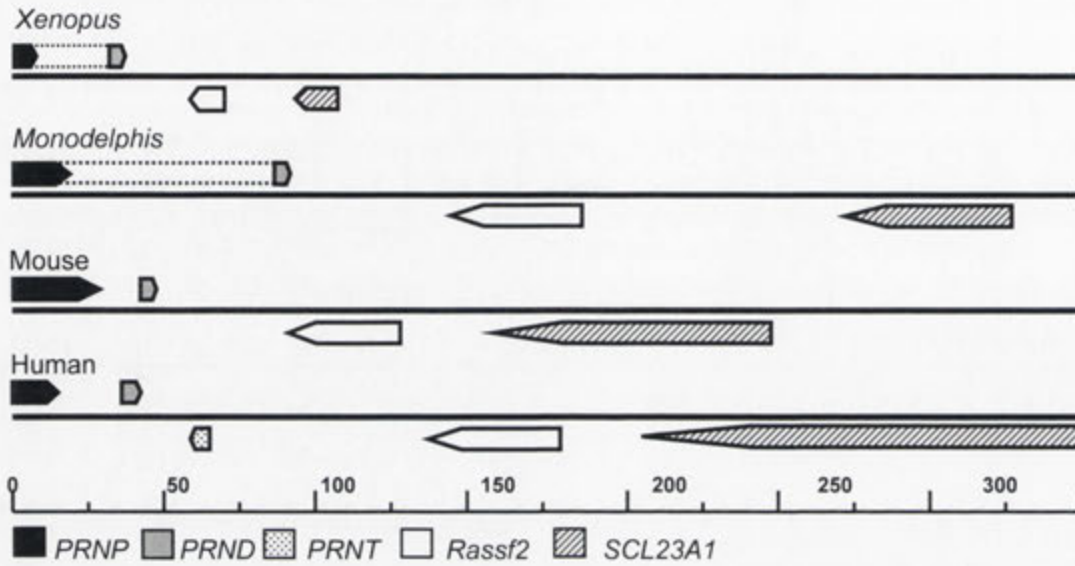


Figure 3-18. Comparative genomic context of *PRNP* and *PRND* genes in amphibian (*X. tropicalis*), marsupial (*M. domestica*), and mammals (mouse and human). The ruler shows approximative distance in kilobase pairs. Dotted lines indicate dominance of bicistronic gene expression.

3.4.8 Duplicate of PrP in *X. Laevis*

X. laevis is a pseudo-tetraploid species with 36 chromosomes, compared with the diploid number, 20, in *X. tropicalis*. Whole genome duplication in *X. laevis* occurred by allopolyploid combination of two different *Xenopus* species at least 30 million years ago (Hughes and Hughes, 1993) providing the possibility of existence of two copies for both *PRNP* and *PRND*. One copy of *X. laevis* *PRNP* has been reported earlier (Strumbo et al., 2001). We uncovered a different *PRNP* sequence by searching EST databases. Comparison of the translated EST sequence with that previously reported for PrP revealed the existence of two distinct proteins coded by two orthologous genes: *PRNP-1* and *PRNP-2*, coding for PrP1 and PrP2 respectively. A *PRND* duplicate was not found.

PrP1 and PrP2 share high sequence homology (**Figure 3-19 (a)**). Absence of repeats, specific for amphibian PrPs, characterises both proteins. N- and C-terminal signal

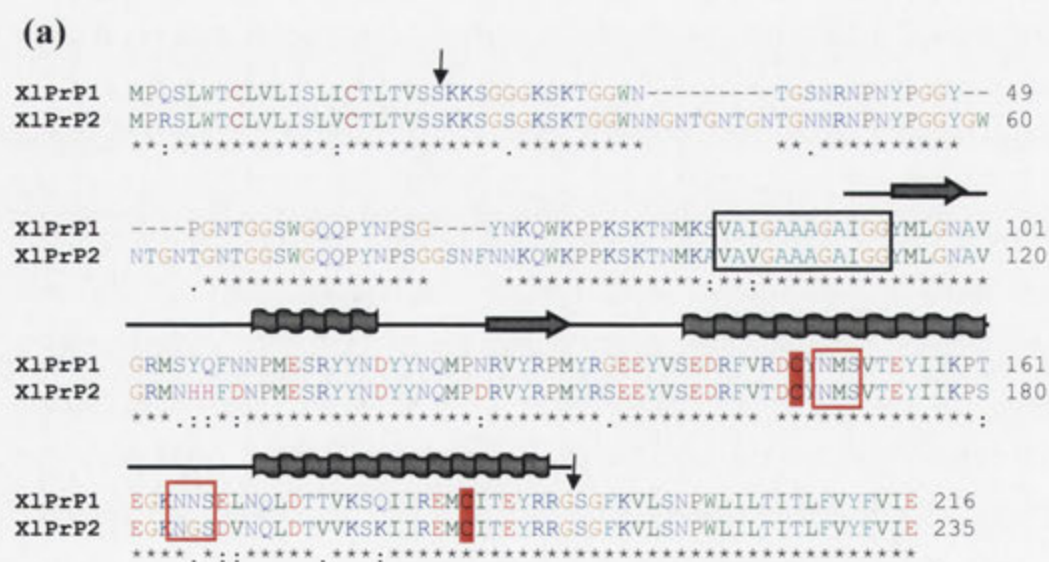
sequences, GPI anchor site, hydrophobic segment, and cysteines involved in the disulfide bridge formation are conserved.

PrP2 is 19 residues larger than PrP1 with extra residues positioned the N-terminal region. The maximum variation is observed in the unstructured N-terminal region and between residues 124-128 ("NHHFD") (**Figure 3-19 (a)**). There are very few variations in the sequences of the α -helices and β -strands; most replacements are conservative. It is possible that substitution of the Arg146 charged residue in the proximity of the Cys148 (PrP1) with neutral Thr165 residue in PrP2 might affect the disulphide bridge stability. Eutherian mammals and marsupials have a charged His residue in this position, but note that *X. tropicalis* PrP has the Thr residue, similar to PrP2. The first N-glycosylation site (PrP1 residues N150, M151, S152) is conserved between PrP1 and PrP2, but the second site (PrP1 residues N165, N166, S168) is not predicted for PrP2 (boxed in **Figure 3-19, a**). Absence of the large N-linked sugar residue might open this position for a ligand interaction. *X. laevis* PrP1 differs from other PrPs, including frog *X. tropicalis*, by the Gln107 substitution of a generally conserved histidine residue. However, the His residue is present in PrP2. This conservation might indicate subfunctionalisation between two PrPs after the polyploid event. In mammals, the loops between α -helix 1 and the first β -strand and between α -helices 2 and 3 are considered hot spots for pathogenic conversion of PrP, due to presence of ligand-binding sites (Kuwata et al., 2007). However, compared with mammals, residues in these positions in *X. laevis* are not conserved; thus, these two loops might bind ligands different from those in mammals.

Residue variation between known *X. laevis* PrP1 (PDB ID 1XUO) and a PrP2 homology model is mapped onto the PrP1 structure in **Figure 3-19 (b)**. N-terminal insertions of short GNT/GWNT repeats (3 in the first indel and 2 in the second) in PrP2 are probably just a consequence of unequal crossing-over, the same unrepeated GWNT and GNT motifs being present in PrP1. The N-terminal end has been shown to be important for intracellular trafficking (Nunziante et al., 2003). Despite low sequence similarity, the short N-terminal segment of *X. laevis* PrP (amino acids 23–69), i.e. PrP1, fused to truncated mouse PrP restored its secretory and endocytic function (Nunziante et al., 2003). Considering this exceptionally high level of evolutionary conservation of

trafficking function the insertion in this region of PrP2 is unlikely to disrupt this function.

Despite sequence similarity, the tissue distribution of PrP1 and PrP2 transcripts (**Figure 3-5**) are different. This signifies the importance of an alternate transcriptional start site and new regulatory DNA elements rather than new specific functions in retaining PrP2. Ohno (1970) has proposed that a redundant duplicate gene is free to accumulate mutations relatively rapidly. This copy might deteriorate to a pseudogene (nonfunctionalisation or pseudogenisation), specialise to perform a different subset of its original functions of the in the presence of its sibling copy (subfunctionalisation) or develop a new function (neofunctionalisation) and thus contribute to evolution. The preservation of *PRNP-2* indicates that purifying selection has led to a modified presumably beneficial second PrP copy. It is possible that the initial promiscuous functions of PrP provided the opportunity for subfunctionalisation or neofunctionalisation of the duplicate gene. In contrast, *PRND*, being a relatively young gene, was subject to negative selection and pseudogenisation.



(b)

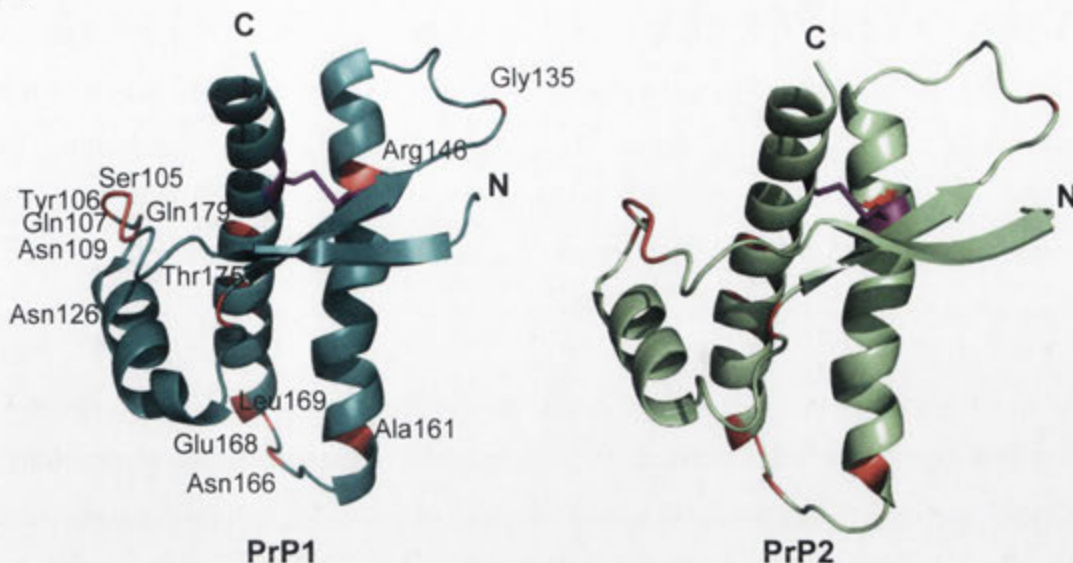


Figure 3-19. Sequence and structural analysis of *X. laevis* PrP1 and PrP2.

(a) Pairwise alignment of *X. laevis* PrP1 and PrP2. The secondary structural components are shown above the sequence (arrows indicate N- and C-terminal cleavage sites; cysteines involved in the disulphide bridge are highlighted in magenta. N-glycosylation sites and mutated second site in PrP2 are boxed in red; hydrophobic region is boxed in black. (b) *X. laevis* PrP1 (PDB 1XU0) and PrP2 (modelled using 1XU0 as a template); substituted residues between PrP1 and PrP2 are labelled and highlighted in red, numbering corresponds to *X. laevis* PrP1. Structured region starts at residue 95 in *X. laevis* PrP1.

3.4.9 Evolution of protein function

3.4.9.1 Overall structure comparison

We can observe from **Figure 3-4 (a)** that amino acid sequences of *X. tropicalis* and *X. laevis* PrP are strongly conserved, especially in the folded C-terminal domain from residue 95 onwards. A different picture is apparent in **Figure 3-4 (b)** for Dpl where multiple variations throughout the sequence have already accumulated between *X. tropicalis* and *X. laevis*. This observation is consistent with the fate of a recently duplicated gene for which the new copy is subject to rapid mutations leading to either acquisition of a new function and survival, or inactivation by pseudogenisation (Ohno, 1970).

Comparative sequence and structural analysis allows tracking of the main features of PrP and Dpl protein evolution. Despite differences in sequence length, the relative positions of the main structural elements (signal sequences, basic regions, hydrophobic

regions, α -helices and β -sheets, GPI-anchor site) are conserved among *Xenopus*, *Monodelphis* and mouse PrPs and Dpls (**Figure 3-20**). The N-terminal repeat region of PrP shows most difference among amphibians, marsupials and mammals. It is absent in amphibians, both in *X. laevis* (Strumbo et al., 2001) and *X. tropicalis* (our results). An important difference in Dpl (amphibian) is absence of the second disulphide bridge. Both PrP and Dpl have two N-glycosylation sites in folded C-terminal domain.

Protein N-glycosylation affects protein stability (Wormald and Dwek, 1999). Although the position of one N-glycosylation site is conserved between PrP and Dpl (**Figure 3-28**, residues N180 and N11, respectively in mouse), the second site is positioned differently and is also different between mouse and *Monodelphis* Dpl (at proximate end of helix 2) and *Xenopus* Dpl (at proximate end of helix 1). Divergent evolution of the N-glycosylation sites may have contributed to the functional divergence of PrP and Dpl through divergence of the fold structure and stability.

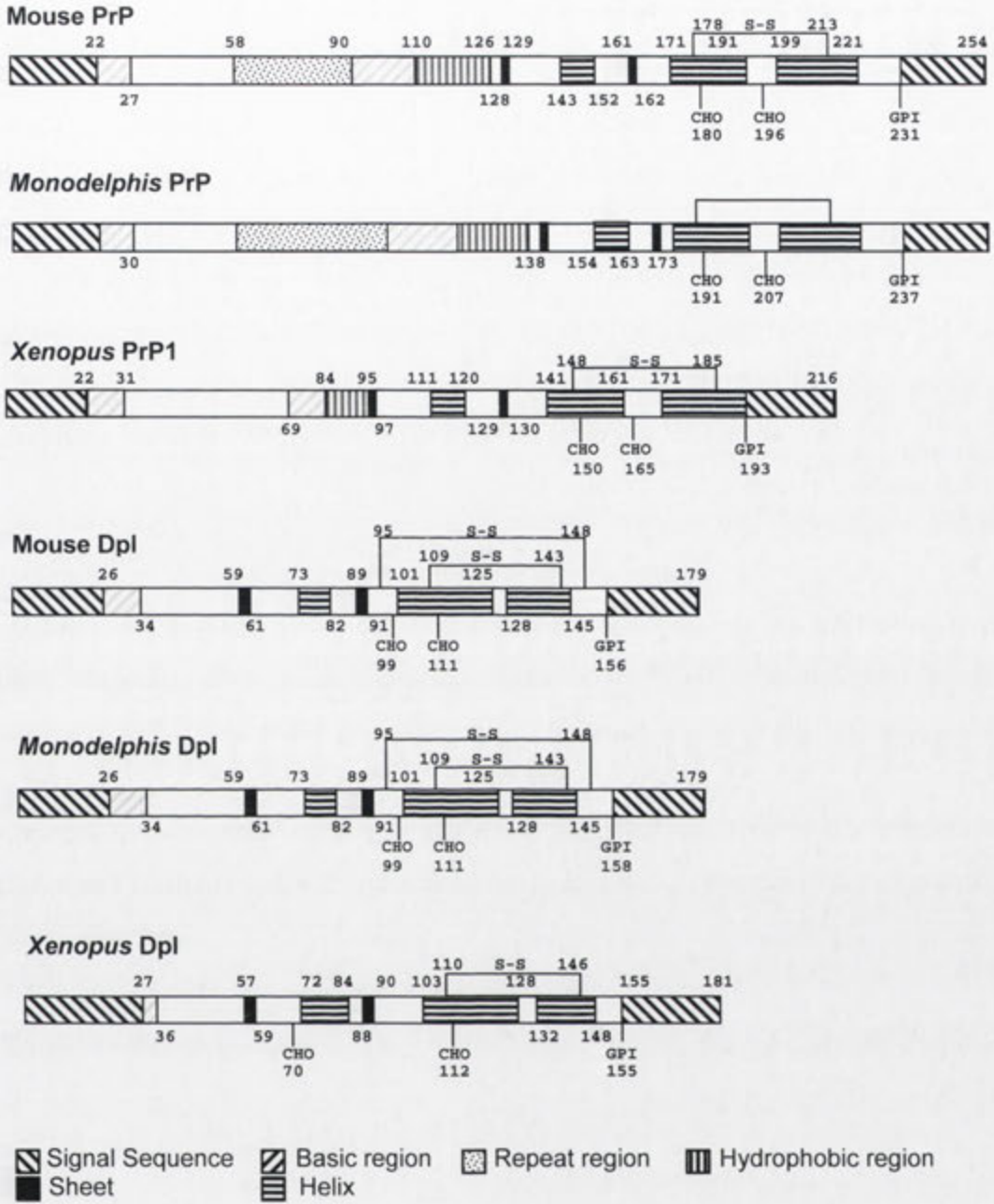


Figure 3-20. Overall structures of *X. laevis*, *M. domestica* and mouse PrP and Dpl. Positions of α -helices and β -strands were deduced from known PDB NMR structures and using SwissModel software for *Monodelphis* and *Xenopus* Dpl and *Monodelphis* PrP (PDB ID: 1AG2, mouse PrP; 2KLF, tammar wallaby PrP; 1XU0, *X.laevis* PrP1; 1I17, mouse Dpl). Numbers indicate the first and last residue of each motif and positions of post-translational modification sites. Shaded are signal sequence, basic region, repeats, hydrophobic region, β -sheets, and α -helices; CHO, N-glycosylation site; S-S, disulphide bond; GPI, glycoposphatidylinositol anchor.

3.4.9.2 Evolution of Dpl functions

Evolutionary history of Dpl starts after the *PRNP* gene duplication. A molecular estimation of divergence times places that event at about 360 million years ago (Kumar and Hedges, 1998). The most ancient elements of Dpl sequence, inherited from parental PrP, are the N-terminal signal sequence for extracellular export, the hydrophobic C-terminal end peptide for addition of a GPI-anchor, one N-glycosylation site, one disulphide bridge, and the general fold of the C-terminal domain. Their positions are well conserved, but both PrP and Dpl acquired a second N- glycosylation site in different positions and PrP acquired a second disulphide bridge between amphibians and marsupials (**Figure 3-20**).

The overall sequence conservation among amphibian, marsupial, and mammalian Dpls is low (**Table 3-6**). Taking *M. domestica* (diverged 173 myr ago) as a mid point of Dpl evolution, comparison of the two periods reveals fast initial Dpl divergence, with sequence identity between *M. domestica* and *X. tropicalis* being only 22%. This rapid initial evolution slowed down during the second period, with sequence identity between *M. domestica* and human Dpl being 45%. This is still much less than PrP sequence conservation of 70% between *M. domestica* and human for the same period (**Table 3-7**). Dpl sequence conservation among mammals is comparable with that of PrPs being, respectively for (human/mouse) 73% and 88%, for (human/bovine) 76% and 87%, and for (mouse/bovine) 76% and 83%. It reveals fast accumulation of mutations in Dpl sequence during the evolution.

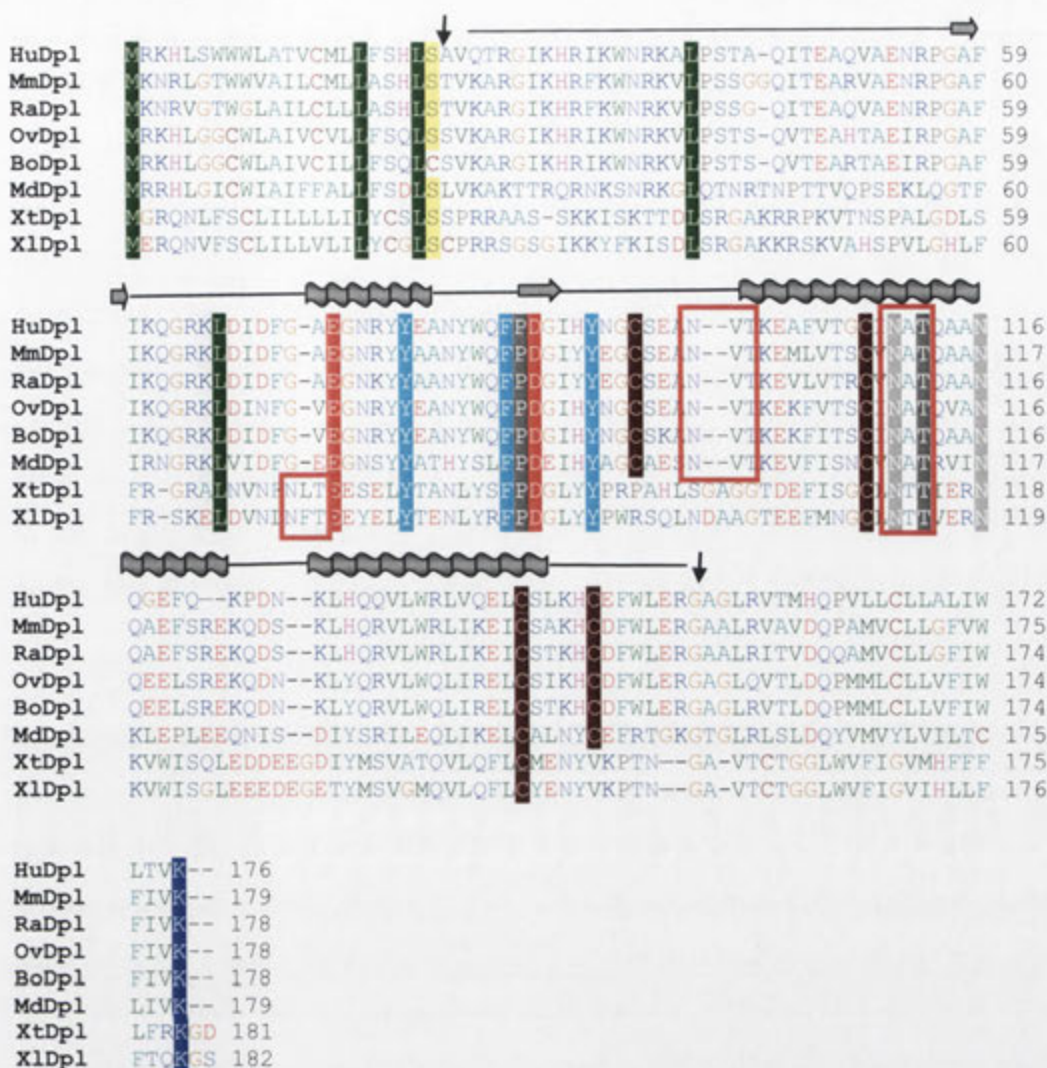


Figure 3-21. Evolution of Dpl protein sequence. Multiple alignment of human (HuDpl), mouse (MmDpl), rat (RaDpl), ovine (OvDpl), bovine (BoDpl), *X. tropicalis* (XtDpl), and *X. laevis* (XlDpl) sequences. α -helices and β -sheets are conserved in all species and indicated on top of the alignment based on mouse Dpl structure (PDB 1H17). The N-glycosylation sites are indicated in red boxes. Cysteine (C) residues involved in disulphide bridge formation are highlighted in brown. Residues conserved during evolution are highlighted in different colours; their significance is discussed in the text. Cleavage sites are marked by arrows (↓).

Table 3-6. Matrix of sequence identity between Dpls from different species. Matrix was created by ClustalW alignment of human (HuDpl), mouse (MmDpl), bovine (BoDpl), marsupial opossum (MdDpl), and amphibians *X. tropicalis* (XtDpl) and *X. laevis* (XlDpl) using BioEdit software.

	HuDpl	MmDpl	BoDpl	MdDpl	XtDpl	XlDpl
HuDpl		0.73	0.76	0.45	0.19	0.20
MmDpl	0.73		0.76	0.46	0.17	0.17
BoDpl	0.76	0.76		0.49	0.19	0.18
MdDpl	0.45	0.46	0.49		0.22	0.19
XtDpl	0.19	0.17	0.19	0.22		0.69
XlDpl	0.20	0.17	0.18	0.19	0.69	

Despite amino acid sequence divergence among mouse, *Monodelphis* and *Xenopus* Dpls, homology modelling of the latter using the mouse NMR structure (PDB ID 1I17) as template shows that protein folding of the 3 α -helices and 2 β -strands is predicted to be well conserved (**Figure 3-22**). The sequence identity between *M. domestica* and mouse Dpl proteins is high enough for reliable modelling, but the relatively low conservation for *Xenopus*, makes the model speculative. The models have been constructed with the Swiss-Model program (Bordoli et al., 2009) which automatically builds the homology model, using the mouse structure as template. As already discussed, *X. tropicalis* Dpl has only one disulfide bridge rather than two as in marsupials and mammals; its structural position is shown in **Figure 3-22**. Also, one of its N-glycosylation sites is displaced from the loop region between $\beta 2$ and $\alpha 2$ in marsupial and mammal to the loop region between $\beta 1$ to $\alpha 1$.

Multiple sequence alignment (**Figure 3-21, Appendix 8**) identifies totally conserved residues among species. Their spatial distribution mapped onto homology models can be seen in **Figure 3-23**. Apart from the conserved Cys109-Cys143 (mouse numbering) pair forming the disulphide bridge, the most highly conserved region is Phe86-Pro87-Asp88 situated just before and overlapping the first β -sheet. Asp88 is closely associated with Leu67 (**Figure 3-21**) forming a surface, as revealed in the 3-D model, that might be involved in interactions with other proteins or ligands. The other totally conserved residues include Leu-40, Glu-74, Tyr-79, Tyr-92, Asn-111, and Thr113. Another interesting conserved residue is Asn117 which is responsible for a kink in the second α -helix (Mo et al., 2001). Comparing conservation among different animal models allows some speculations about functional development of Dpl during evolution. For example, residues conserved between Dpls in higher and lower vertebrates are likely those necessary to preserve the fold and key posttranslational modifications of Dpl, whereas those conserved between mammals and marsupials will likely be important for more recently acquired functions.

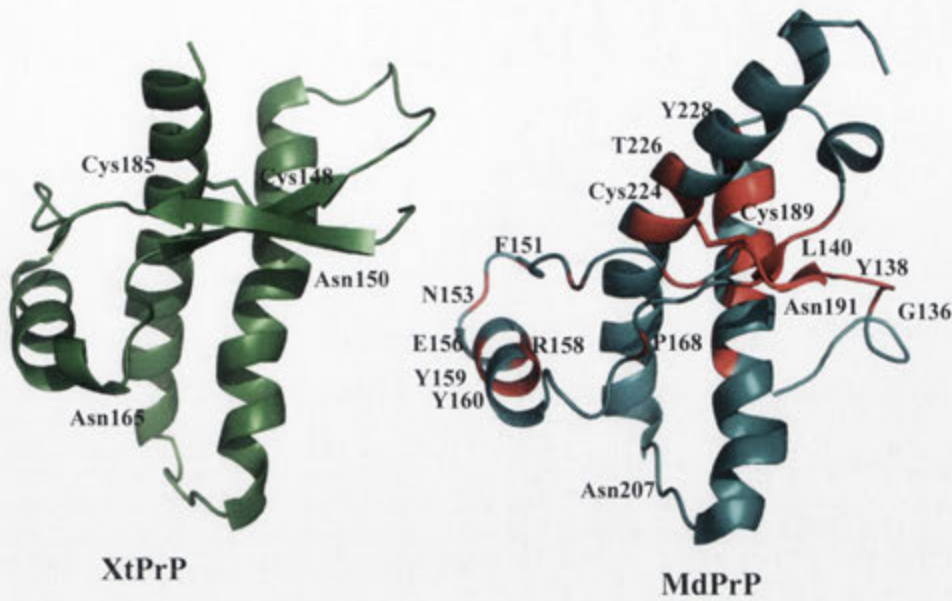


Figure 3-22. Homology models for *X. tropicalis* (XtDpl) and *M. domestica* (MdDpl). *X. tropicalis* and *M. domestica* models were created by using mouse Dpl template (PDB ID 1I17). Conserved residues (in red) between marsupial and amphibians were mapped onto the MdDpl surface using PyMOL graphics software. Positions of Cys involved in disulphide bridge formation, Asn involved in N-glycosylation, and some conserved residues are labelled. Numbers correspond to *Xenopus* or *Monodelphis* sequences.

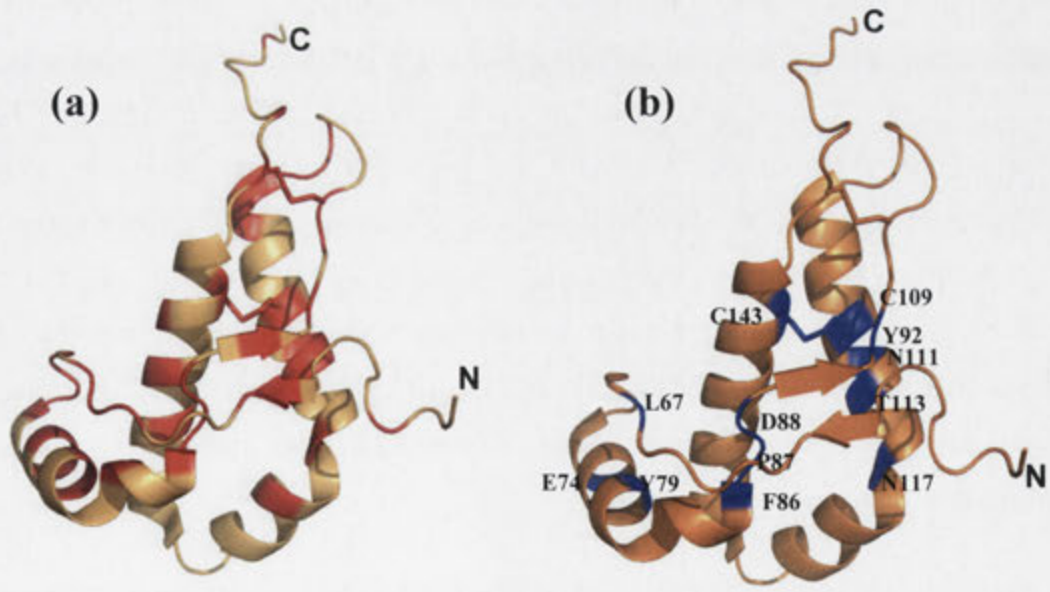


Figure 3-23. Distribution of evolutionarily conserved residues mapped onto the mouse Dpl structure. Residue numbers correspond to mouse sequence. (a) Residues conserved between marsupial (*M. domestica*) and mammal (mouse) are shown coloured red; (b) Residues conserved among mammal, marsupial, and amphibian are shown in blue.

3.4.9.3 Evolution of PrP functions

Gene duplication allows development of new functions of both duplicates (Ohno, 1970). According to this model, PrP either lost some of its functions to Dpl (subfunctionalisation) and/or acquired new ones (neofunctionalisation). PrP underwent fast initial divergence between amphibians and marsupials as shown by sequence identity between *Monodelphis* and *X. tropicalis* PrP being only 25% (**Table 3-7**). This rapid initial evolution might be attributed to the duplication event, which allowed PrP the opportunity to develop new functions. Amphibian PrP differs from that in marsupials and mammals by absence of the repeat region. Taking into account the overall structural conservation of PrP, it is clear that a significant part of evolution of PrP has been driven by repeat expansion. The number of repeats varies from octarepeats in mammals to decarepeats in marsupials. The P(H/R)(G/P)GG(--)WGQ motif is conserved. The process of PrP neofunctionalisation was probably mostly complete in the most recent common ancestor of marsupial and placental mammals as may be deduced from the fact that sequence identity between mammals and marsupials (70%) and within the mammalian group (88%) are fairly similar, compared with that between amphibian and marsupial, despite an evolutionary distance of 173 myr for marsupials and 20-40 myr for mammals (**Table 3-7**).

Evolution of PrP molecular domains may have an impact on its function. Conserved residues were identified from multiple sequence alignment (**Figure 3-24, Appendix 7**) and mapped onto homology models to reveal their spatial distribution (**Figures 3-25 and 3-26**). Various conserved residues are involved in fold integrity (Riek et al., 1996). Some of them are in similar positions to those of Dpl, as may be seen from **Figures 3-23 and 3-24** and will also be discussed later (**Figure 3-28**). However, some of the conserved residues are on the outer surface and these may be responsible for evolutionarily preserved functions of PrP: E145, Y148, Y149, E151, N152 (Calzolari et al., 2005). Substitutions in *Xenopus sp.* E119D and N120Y are conservative and thus could support the same function.

The most conserved region of PrP which is not involved in fold formation is the basic region (100-109). It is suggested to bind the C-terminal SH3 domain of Growth factor receptor-bound protein 2 (Grb2) and affect signal transduction (Spielhaupter and Schatzl, 2001, Lysek and Wuthrich, 2004). Another region is the hydrophobic region

(110-126). However, the frog hydrophobic region is shorter with GAAA deleted (85-96). Conserved Gly residues at the beginning of the N-terminal domain (30, 34, 35, 45) are probably important for intracellular trafficking and mediation of endocytosis; restoration of wild-type function by addition of the N-terminus of the *X. laevis* PrP to N-terminally truncated mouse PrP has been demonstrated (Nunziante et al., 2003).

Despite overall structural similarities, *Xenopus* PrP sequence is strongly diverged from those of mammalian PrPs. It has more stable structure as loop 131-140 appears to be stabilised by a 2-amino-acid insertion in XIPrP (Calzolari et al., 2005). Residues associated with PrP^{Sc} conversion in mammals (human M129, D178, V180, F198, E200, V210, Q217 (Prusiner, 1993)) are not conserved in amphibians, except for M129 and D178. The putative protein X binding residues (mouse Q167, Q171, Q218 (Kaneko et al., 1997)) are also not conserved in *Xenopus*.

Conservation of some PrP residues points at development of a stable C-terminal domain fold and appearance of some functions early in evolution. However, sequence divergence of mammalian PrP from that of amphibian suggests the development of novel functions later; additional experiments are needed to test this suggestion.

Table 3-7. Matrix of sequence identity between PrPs from different species. Matrix was created by ClustalW alignment of mammals (human (HuPrP), mouse (MmPrP), bovine (BoPrP)), marsupials (opossum (MdPrP), possum (PoPrP), tammar wallaby (TwPrP)), and amphibians (*X. tropicalis* (XtDpl) and *X. laevis* (XlDpl)) using BioEdit software.

	HuPrP	MmPrP	BoPrP	MdPrP	PoPrP	TwPrP	XtPrP	XlPrP
HuPrP		0.88	0.87	0.68	0.71	0.70	0.25	0.27
MmPrP	0.88		0.83	0.68	0.71	0.69	0.24	0.26
BoPrP	0.87	0.83		0.69	0.72	0.70	0.24	0.26
MdPrP	0.68	0.68	0.69		0.84	0.83	0.25	0.27
PoPrP	0.71	0.71	0.72	0.84		0.93	0.26	0.28
TwPrP	0.70	0.69	0.70	0.83	0.93		0.26	0.28
XtPrP	0.25	0.24	0.24	0.25	0.26	0.26		0.85
XlPrP	0.27	0.26	0.26	0.27	0.28	0.28	0.85	

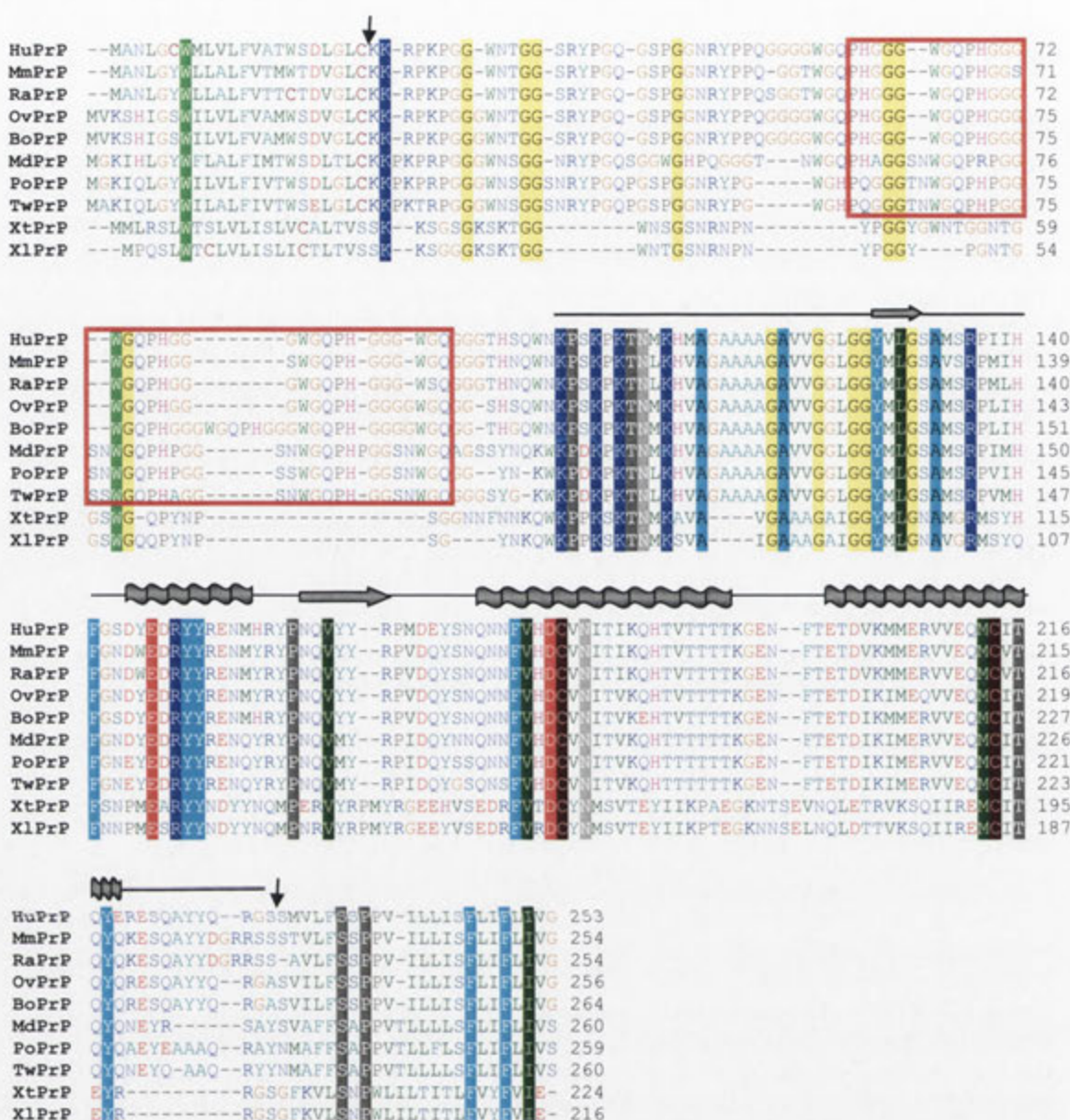


Figure 3-24. Evolution of PrP protein sequence. Multiple alignment of human (HuPrP), mouse (MmPrP), rat (RaPrP), ovine (OvPrP), bovine (BoPrP), *X. tropicalis* (XtPrP), and *X. laevis* (XlPrP) sequences; α -helices and β -sheets are conserved in all species and indicated on top of the alignment. The repeat region is indicated in red box. Cysteine (C) residues involved in disulphide bridge formation are highlighted brown. Residues conserved during evolution are highlighted in different colours; their significance is discussed in the text. Cleavage sites are marked by arrows (↓).

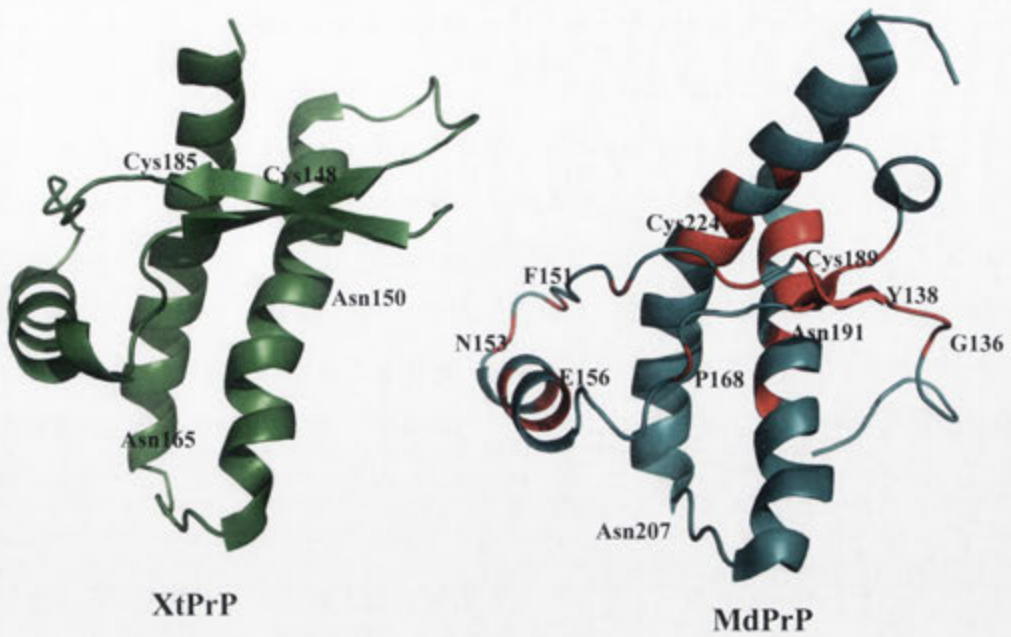


Figure 3-25. Homology models for *X. tropicalis* (XtPrP) and *M. domestica* (MdPrP). The *X. tropicalis* model was created by using *X. laevis* PrP template (PDB ID 1XU0). *M. domestica* model was created by using tammar wallaby PrP template (PDB ID 2KFL). Conserved residues between marsupial and amphibians are mapped onto the MdPrP surface in red using PyMOL graphics software. Positions of Cys involved in disulphide bridge formation and Asn involved in N-glycosylation, and some conserved residues are labelled. Numbers correspond to *Xenopus* or *Monodelphis* sequences.

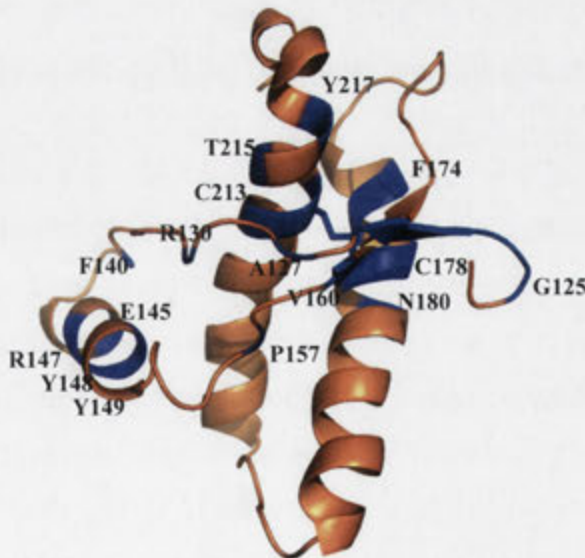


Figure 3-26. Distribution of evolutionarily conserved residues mapped onto the mouse PrP structure. Compared are NMR structures of *X. laevis* (PDB ID 1XU0), tammar wallaby (PDB ID 2KFL), and mouse (PDB ID 1AG2). Residue numbers correspond to mouse sequence. Residues conserved among mammal, marsupial, and amphibian are shown in blue.

3.4.9.4 Comparative analysis of PrP and Dpl structures

It is accepted that the longer the evolutionary distance between paralogues, the more different they become. We can expect that soon after the duplication, PrP and Dpl had similar amino acid and secondary structure that later diverged. To test this expectation I aligned PrP and Dpl sequences from different species (**Figure 3-27, Appendix 9**) and compared sequence identity of their C-terminal domains (**Table 3-8**). Sequence identity of the N-terminal region was not calculated because of large length differences and obvious divergence, and the particularly disordered PrP repeats and hydrophobic region. Sequence identity was calculated as the percentage of identical residues between the PrP and Dpl C-terminal ends, excluding the signal sequences. This analysis reveals that sequence identity of the C-terminal is similar (~20%) among amphibians, marsupials and eutherial mammals. This suggests that rapid evolution and sequence divergence of the C-terminal region was essentially complete in the most recent common ancestor of amphibians and mammals, with conserved residues restricted to those essential for fold maintenance. Later, functional divergence of PrP and Dpl was driven by large changes in the N-terminal disordered end and in C-terminal residues which are not important for folding.

Despite several hundred million years of divergence and low sequence similarity, the C-terminal domain structure characterised by 3 α -helices and 2 β -sheets is strongly retained between PrP and Dpl (**Table 3-8**). Both PrP and Dpl also retain the overall structural domain features (signal sequences, GPI anchor attachment, N-glycosylation site, folded C-terminal domain, disordered N-terminal domain and, as noted, share conserved residues (**Figure 3-28**). Conserved residues, stabilising fold formation in PrP are C178/C213 connected by a disulphide bond, Y162/T182 connected by a hydrogen bond, Y148/Y149 anchored to the protein core (Calzolari et al., 2005, Luhrs et al., 2003) (**Figure 3-28, PrP**, numbering correspond to mouse sequence). They correspond to residues C109/C148, Y92/T113, and Y78/Y79 mapped onto mouse Dpl structure (**Figure 3-28, Dpl**). PrP residues F140 (F71 in Dpl), N152 (N82), P157 (P87), Y159 (Y92), and F174 (F/L105) are also involved in structure maintenance (Luhrs et al., 2003). Leucine (L) residue at position 105 in mouse and rat Dpls is uniquely different from phenylalanine (F) in other species (**Appendices 8 and 9**). The N-glycosylation site (N180 for PrP and N111 for Dpl) also contributes to fold stability. The conserved

V179/V110 next to the disulphide bridge might contribute to fold stability and bridge formation.

It can be hypothesised that stabilisation of the C-terminal region fold was required for the function of the primordial ‘PrP’ before the duplication event. It is possible that this basic function is retained in both PrP and Dpl but the fact that PrP, in particular, has a reported plethora of binding partners involving both the C-terminal and N-terminal domains and putative functions makes it difficult to decipher what it might be. Functional studies have shown that N-terminally truncated PrP mimics Dpl (Shmerling et al., 1998, Flechsig et al., 2003), and that only N-terminally truncated PrP coexists with Dpl in mature spermatozoa (Peoc'h et al., 2002). Thus this hypothesis seems plausible.

Rapid evolution of both PrP and Dpl just after the duplication with a subsequent slow down is consistent with the normal fate of duplicates. Following duplication, genes usually experience a brief period of relaxed selection with subsequent strong purifying selection during which the majority of duplicates are silenced and only a few survive (Lynch and Conery, 2000); PrP and Dpl are such a pair.

Table 3-8. Sequence identity between the C-terminal region of PrP and Dpl in different species, excluding the signal sequences. % ID was calculated from ClustalW alignment of PrP and Dpl sequences from mouse, marsupial *M. domestica*, and amphibian *X. tropicalis* using BioEdit software. Starting points for C-terminal region shown by blue arrows in **Figure 3-27**.

Species	C-terminal, % ID
<i>X. tropicalis</i>	16.6
<i>M. domestica</i>	20.0
Mouse	18.5
Human	20.0

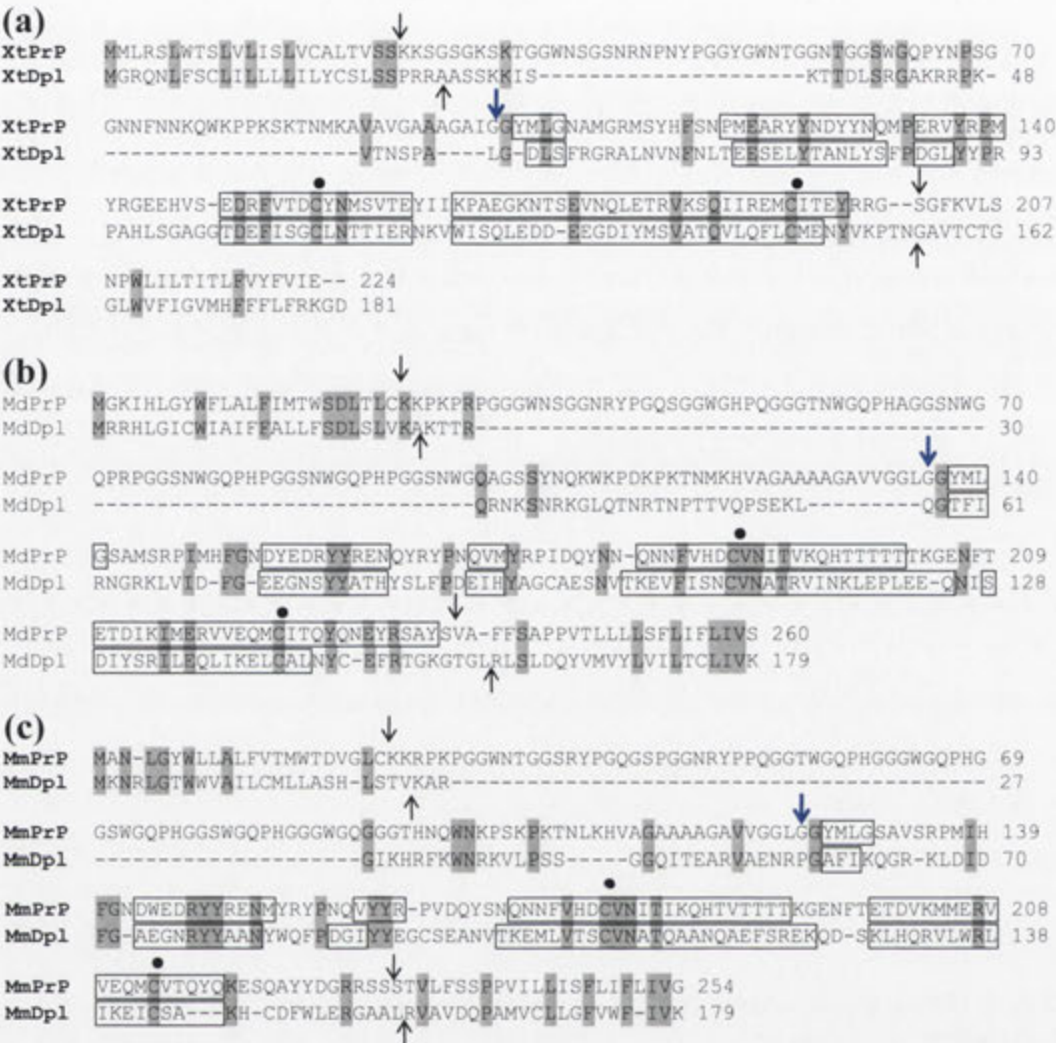


Figure 3-27 Sequence similarity between PrP and Dpl in (a) *X. tropicalis*; (b) *M. domestica*; (c) Mouse. Identical residues are highlighted in grey; α -helices and β -strands are boxed; their positions were deduced from known PDB NMR structures and using SwissModel software for *Monodelphis* and *Xenopus* Dpl and *Monodelphis* PrP (PDB ID: 1AG2, mouse PrP; 2KLF, tammar wallaby PrP; 1XU0, *X.laevis* PrP; 1I17, mouse Dpl). Signal sequence cleavage sites are indicated by arrows, and conserved Cys of disulphide bridge by dots. Starting point for C-terminal region for calculation of % identities in Table 3-8 are shown by blue arrows.

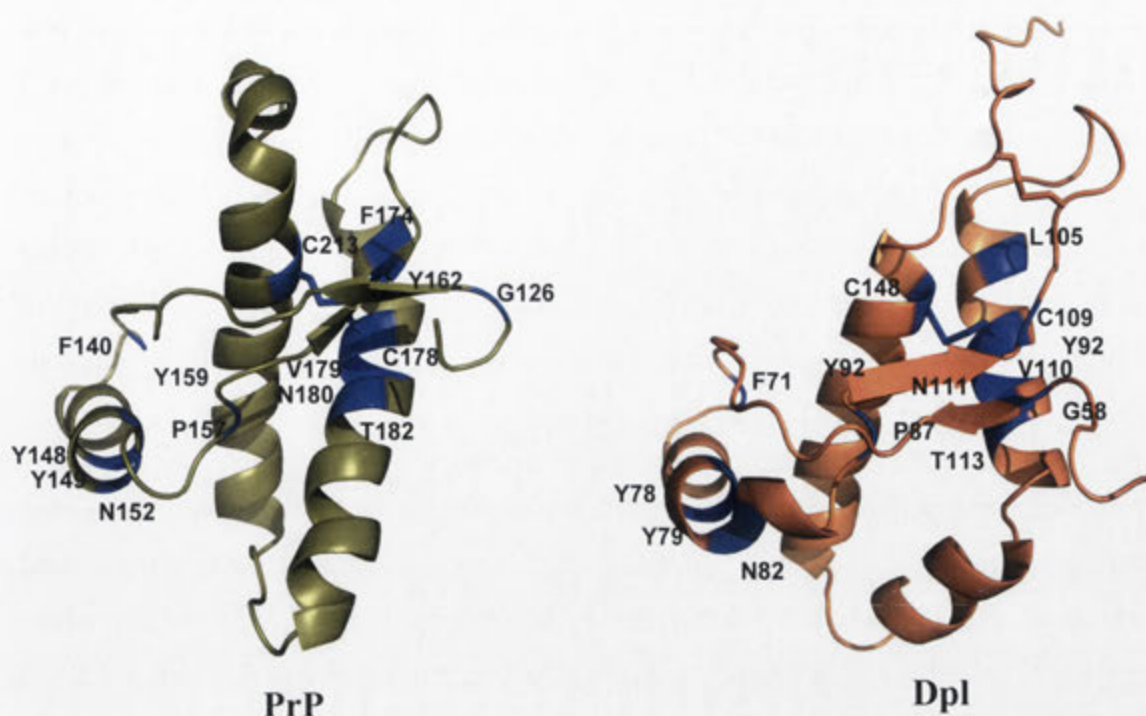


Figure 3-28. Structural conservation of PrP and Dpl among mouse, *M. domestica* and *X. tropicalis*. The conserved residues were identified from multiple sequence alignment of PrPs and Dpls from different species (**Appendix 9**) and are mapped in blue onto the mouse PrP structure (1AG2) and the mouse Dpl structure (1II7). Residue numbers correspond to mouse PrP or Dpl sequence, respectively. L105 residue in mouse Dpl corresponds to a conserved F Dpl residue in other species.

3.5 Conclusion

The evolution of functions of the *PRNP* and *PRND* genes illustrates a range of evolutionary processes after tandem and whole genome duplication. The novel finding of the *PRND* gene in *Xenopus* species allowed me to demonstrate mechanisms of gene retention, sequence divergence, and acquisition of new tissue distribution and apparent functions after the tandem gene duplication. Furthermore, the discovery of a variant *PRNP* gene, but not *PRND*, in the tetraploid frog *Xenopus laevis* provided additional insights into the divergence of duplicate genes or their loss after whole genome duplication from allopolyploidy. My results demonstrate that PrP and Dpl evolved concurrently and possibly cooperatively. Both genes acquired novel, more specialised functions, which contributed to physiological complexity of mammals. Evolution of PrP was given impetus by the acquisition of the CPE motif at the 3' UTR of the mRNA for posttranslational regulation and elaboration of N-terminal repeats in the protein. At the

same time, the *Dpl* gene evolved by gaining new regulatory DNA elements and new tissue specificity. From comparative sequence analysis and tissue distribution, I conclude that the functions of *PRNP* and *PRND* in mammals differ significantly from those in frogs. However, conservation of the main structural domains of PrP (α -helices, β -strands, N- and C-terminal signal sequences, N-glycosylation sites, GPI anchor attachment site, disulfide bonds, and basic and hydrophobic regions) suggests that PrP may have retained its original function and evolved additional promiscuous functions rather than gaining one completely new function.

The intrinsic functional flexibility of PrP is demonstrated by retention of its second copy after the whole genome duplication in *X. laevis*. Conversely, loss of the second variant of *PRND* might point to its functional immaturity at the time of the whole genome duplication and little selective advantage in retaining two copies with modified properties.

A key aspect of the early evolution of the *Dpl* gene is the existence of chimeric transcripts sharing 5' untranslated exons with the PrP gene in *Xenopus* sp., together with non-chimeric *PRND* transcripts expressed from a newly acquired promoter. Our findings suggest that divergence of PrP and *Dpl* started with subfunctionalisation followed by neofunctionalisation. This is consistent with a suggestion that functional evolution of duplicated genes cannot be explained by subfunctionalisation or neofunctionalisation alone, but that subfunctionalisation may be a transient mechanism before neofunctionalisation (He and Zhang, 2005, Rastogi and Liberles, 2005).

Neofunctionalisation of *Dpl* begins with the acquisition of a novel regulatory DNA region and, as a result, a novel tissue distribution: specifically higher expression in testis and repressed expression in brain. It seems that initial chimeric transcription allowed *PRND* to evolve very rapidly by mutation while its expression was preserved by coexpression with *PRNP*, thus allowing for selection for new functions and escape from pseudogenisation.

In summary, the discovery of the recently duplicated *PRND* gene in amphibians and marsupial, especially in its chimeric transcript form, allows us to better understand the process of gene evolution.

Chapter 4.

Regulation of expression of *PRNP* and *PRND*

As introduced in Chapter 1, expression of PrP and the PrP-homologue protein Dpl is altered in various neurodegenerative disorders, such as Alzheimer's, as well as after brain injury and during development. Recent studies suggest that PrP possesses neuroprotective functions while Dpl is neurotoxic and induces apoptosis. Although PrP/Dpl antagonism is currently explained by competitive ligand binding (Sakaguchi, 2008), there is no proposed mechanism for how this antagonism is regulated.

This chapter aims to define the regulatory mechanisms underlying this antagonism. I hypothesised that the antagonistic relationship of PrP and Dpl is transcriptionally coregulated to achieve balance between apoptosis and survival of neuronal cells. Using bioinformatics, experimental methods, and the conceptual biology approach, I propose a molecular mechanism for PrP and Dpl coregulation and novel insights into neuronal death.

4.1 Introduction

4.1.1 Antagonistic functions of PrP and Dpl

PrP's antiapoptotic activity is illustrated by protection against oxidative stress through SOD activity (Brown and Besinger, 1998), prevention of neurodegeneration in PrP-null mice which overexpress Dpl (Moore et al., 2001, Nishida et al., 1999), inhibition of proapoptotic protein Bax (Park and Li, Bounhar et al., 2001, Roucou et al., 2003), promotion of survival of cerebellar granule neurons in culture (Krebs et al., 2007, Chen et al., 2003), and antiapoptotic regulation of Ca^{2+} signalling by limiting Ca^{2+} accumulation in mitochondria (Brini et al., 2005). On the contrary, Dpl proapoptotic function is demonstrated by loss of Purkinje cells in PrP-null mice (Moore et al., 2001, Nishida et al., 1999), apoptosis in Dpl-expressing cerebellar neurons and astrocytes (in culture) mediated through caspase-3 and caspase-10 activation (Qin et al., 2006b), exacerbation of oxidative damage by Dpl expression (Wong et al., 2001), and

proapoptotic regulation of Ca^{2+} signalling when overexpressed Dpl increases mitochondrial Ca^{2+} (Brini et al., 2005).

4.1.2 Reported transcription factors for *PRNP*

PRNP is widely expressed in adult mammalian tissue and was originally considered a house-keeping gene (Oesch et al 1985). Only ubiquitous transcription factors (TFs), such as Sp1, Ap1 and Ap2 were predicted to bind the human, bovine and rat *PRNP* promoter (Mahal et al., 2001, Inoue et al., 1997, Saeki et al., 1996, Baybutt and Manson, 1997). However, cell-specific expression of PrP (Ford et al., 2002b) suggests some chromatin rearrangement and tissue-specific TFs might be involved to inhibit ubiquitous PrP expression. Indeed, early phylogenetic footprinting comparison of multiple mammalian and one marsupial sequences (Premzl et al., 2005) revealed conserved DNA regions with MEF2, Oct-1, MyT1 and NFAT transcription factor binding sites (TFBSs), and later work developed this approach (Chakka and Gready, 2006).

4.1.3 Reported transcription factors for *PRND*

Even though *PRND* is non-ubiquitously expressed and seems to have specialisation in spermatogenesis, only ubiquitous TFBSs, USF1 (Sepelakova et al., 2005) and NF-Y (Nagyova et al., 2004) are reported so far in the published literature in the mouse *Prnd* promoter region that spans -185/+27 relative to the transcription start site (Nagyova et al., 2004).

4.1.4 Chromatin and *PRNP/PRND* regulation

Important roles for chromatin-dependent regulation have recently been proposed in synaptic plasticity and memory formation (Levenson and Sweatt, 2005). Recent data indicate that TSA (histone deacetylase inhibitor) treatment affects PrP promoter activity and expression (Cabral et al., 2002), suggesting histone modification may be regulating accessibility of DNA to activating factors.

Dpl shows high expression during spermatogenesis (Rondena et al., 2005), where unique chromatin remodelling events are known to regulate transcription (Sassone-Corsi, 2002), as well as wide distribution in foetal tissues but restricted in adults.

4.2 Materials and Methods

4.2.1 Cell lines

The cell lines used are mouse spermatogonia GC-1 (ATCC: CRL-2053), mouse primary spermatocyte GC-2 (ATCC: CRL-2196), mouse sertoli cell TM4 (ATCC: CRL-1715) (provided by Prof Kate L. Loveland, Monash Institute of Medical Research, Melbourne), mouse cerebral cortex endothelial bEND.3 (ATCC:CRL-2299) (provided by Prof Carolyn Geczy, Inflammatory Disease Research Unit School of Medical Sciences, University of New South Wales), and human neuroblastoma SHSY5Y (ATCC: CRL-2266) (provided by Dr Alison Shield, Australian National University). All cell lines, except TM4, were maintained in DMEM media supplemented with 10% fetal bovine serum and 1% glutamine in a 37 °C incubator with 5% CO₂ supply. TM4 was maintained in DMEM/F12 media supplemented with 5% horse serum and 2.5% fetal bovine serum.

4.2.2 Total RNA extraction and cDNA synthesis

Wild-type BALB/C mice tissues at different stages (E14, embryonic day 14; P1, postnatal day 1; P15, postnatal day 15; and adult, 18 weeks old) were collected and preserved in RNAlater solution (Ambion). Total RNA was isolated from brain (E14, P1, P15, and adult), adult testis, heart, kidney, liver, lung, muscle, spleen, and thymus. GC1, bEND, and SHSY5Y cells were grown as described, trypsinised, washed in PBS solution and collected by centrifugation. All RNA extractions were conducted using the RNeasy Mini Kit (Qiagen). DNase treatment (Ambion) was included to remove residual DNA contamination. Human brain total RNA was purchased (Ambion® Human Brain Reference RNA, pooled several brain regions of multiple donors, catalog #6050).

1 µg of total RNA from each mouse tissue, cell lines and human brain was primed by an oligo(dT) primer and first strand cDNAs were synthesised using SuperscriptIII reverse transcriptase (Invitrogen), following the manufacturer's instructions.

4.2.3 Oligonucleotide primers

Oligonucleotide primers for mouse *Prnp* and *Prnd* promoter amplification and mutagenesis were synthesised by Sigma (Australia). Primer sequences and their descriptions are provided in **Table 4-1**.

Primers used for RT-PCR analysis of mouse *Prnp*, *Prnd*, *Bptf/Fac1*, *E2F1*, and β -actin were synthesised by Sigma (Australia); they are summarised in **Table 4-2**. To avoid genomic DNA amplification, primer pairs were designed as specific to different exons (in that case DNA amplification would result in a larger size PCR band). H₂O control was used to test for template contamination.

Table 4-1 Summary of mouse-specific primers used for promoter amplification and site-directed mutagenesis. Predicted transcription factor binding sites are highlighted in grey, mutated nucleotides are in bold.

LprFAC1	TACTGCTCCATTTTGC GTGA
RprPrP	CATACAGGGAGTTCCAGGTCA
LprDplBam	GAGGTTGGGTCTTGATGGTC
RprDplHind	CATACAGGGAGTTCCAGGTCA
RP3	ACAAGCACGGAGGTTTTCTG
LmutMAZ1	GATGGTCTTGGAGAGGGTTGGGAGGG
RmutMAZ1	CCCTCCCAACCCCTCTCCAAGACCATC
LmutMAZ2	GGGAGGGTTGGGAGAAGAACTGAAAAATCATC
RmutMAZ2	GATGATTTTTTCAGTTCTTCTCCCAACCCCTCCC
LmutFAC1	AGTTTTCTCTTTTTGGTGTTTTTAAAGGACTCC
RmutFAC1	GGAGTCCTTTAAACACCAAAAAGAGAAAACT
LmutFAC2	GGATTTTTCTTGTTTGGTGTTTGTTTTTAAATTGG
RmutFAC2	CCAAATTAAAAAACAAACACCAAAACAAGAAAAATCC
LmutE2F1	CTCAGTCCTGCTTCTCCACATACCCAG
RmutE2F1	CTGGGTATGTGGGAGAAAGCAGGACTGAG

Table 4-2. Summary of mouse-specific primers used in RT-PCR analysis. Superscript numbers indicate position of the first and the last nucleotides of the primer in cDNA relative to the ORF first nucleotide.

LmmFac1	¹⁷⁶² GACCCAGAAAAGGACAAGCA ¹⁷⁸¹
RmmFac1	²³⁶⁴ ACGAGTGGACCCAGATGTTCT ²³⁴⁵
LmmE2F1	³⁸³ ATCTGACCACCAACGCTTC ⁴⁰²
RmmE2F1	⁹⁰⁴ TAATCCCGTCTGCACTCTCC ⁸⁸⁵
LmmPex1-94	⁻⁶² GCATTCTGCCTTCCTAGTGG ⁻⁴³
Rpp3	⁵⁹² TGAAGTTCTCCCCCTTGGTG ⁵⁷³
LMex1Dpl	⁻³⁵ TCAGAGGCCACAGTAGCAGA ⁻¹⁶
RM587	⁵²⁶ ACCAAACGAAACCCAGCAG ⁵⁰⁸
LMex1Dpl	⁻³⁵ TCAGAGGCCACAGTAGCAGA ⁻¹⁶
L β -actin	³⁴⁹ AGATGACCCAGATCATGTTTG ³⁶⁹
R β -actin	⁹⁸⁹ AGCAATGATCTTGATCTTCATG ⁹⁶⁸
LhFAC1	¹⁷⁴⁵ GCCAAAAAGGGAGACATTGA ¹⁷⁶⁵
RhFAC1	²³⁸⁸ ATTCGAGTTGAGCCAGATGC ²³⁶⁹

LhE2F1	⁵⁴⁰ TGCCAAGAAGTCCAAGAACC ⁵⁵⁹
RhE2F1	¹⁰³⁵ ATCTGTGGTGAGGGATGAGG ¹⁰¹⁶
Lhex1Dpl	⁻⁴⁰ TCCAGAGGCACACTCCAGA ⁻²²
RH505	⁴³⁴ CAATGCTTGAGGGAGCAGAG ⁴¹⁵
LhPrPex	⁻³⁹ AGCTTCTCCTCTCCTCACGA ⁻²⁰
RhPrP	³¹³ GCTTACTCGGCTTGTTCCAC ²⁹⁴

4.2.4 Cloning of *Prnp* conserved non-coding DNA region and *Prnd* promoter

A 630 bp mouse *Prnp* fragment harbouring the predicted TFBSs for FAC1, MEF2, and MAZ transcription factors and a 650 bp mouse *Prnd* proximal promoter fragment were amplified by PCR (BD Advantage Polymerase mix, BD Biosciences; 95 °C 2 min, followed by 30 cycles (95 °C 30 sec, 60 °C 30 sec, 68 °C 3 min), 72 °C 10 min from mouse genomic DNA using primers LprFAC1/RprPrP for *Prnp* and LprDplBam/RprDplHind for *Prnd* (Table 4.1). The 630 bp *Prnp* PCR product was first cloned into TOPO-TA vector to create XhoI/HindIII restriction sites. The *Prnd* PCR product was directly digested with XhoI/HindIII restriction enzymes to obtain the 650 bp fragment. Both *Prnp* and *Prnd* fragments were then cloned into the corresponding sites of the pGL4 luciferase reporter vector (Promega). Ligation reactions (T4 DNA ligase, Invitrogen) were carried out by first incubating DNA at a vector to insert molar ratio of 1:6 at 65 °C for 5 minutes, followed by cooling on ice and adding 5X ligation buffer followed by incubation at room temperature for 1 hour.

Two truncations of the 630 bp *Prnp* fragment were cloned into the luciferase vector (Figure 4-3). A 320 bp *Prnp* fragment was amplified with the LprFAC1/RP3 primer pair (Table 4-1) to remove one of the two predicted FAC1 binding sites. A 190 bp *Prnp* fragment containing the second predicted FAC1 site, but with the predicted MAZ and MEF2 binding sites removed was obtained by KpnI digestion of the 630-*Prnp*-luciferase plasmid, isolation by gel purification and cloning into KpnI sites of the luciferase vector.

Plasmids for transfections were prepared using the GenElute™ Endotoxin-free Plasmid Midiprep kit (Sigma).

4.2.5 Site-directed mutagenesis of transcription factor binding sites

Site-directed mutagenesis was performed to introduce point mutations into the predicted FAC1 and MAZ binding sites on the 630 bp *Prnp* conserved non-coding DNA region and the predicted E2F1 binding site on the 650 bp *Prnd* promoter cloned into the pGL4 reporter vector (Promega). The method is based on using complementary forward and reverse primers containing mutations of corresponding TFBSs. The extension of primers annealed to templates occurred during PCR using high fidelity Platinum^(R) *Pfx* DNA polymerase (Invitrogen) under the following conditions: 94 °C for 15 sec, 55 °C for 30 sec, 68 °C for 5 min, 15 cycles. Following temperature cycling, the product was treated with the endonuclease DpnI (New England Biolab), which only digests the template DNA. The nicked vector mutant DNA is then transformed into the TOP10 *E.coli* strain (Invitrogen) which repairs the nick and replicates the DNA. Site-specific mutagenesis was confirmed by sequencing.

The primer pairs carrying point mutations were LmutMAZ1/RmutMAZ1 and LmutMAZ2/RmutMAZ2 for MAZ binding sites, LmutE2F1/RmutE2F1 for the E2F1 binding site, and LmutFAC1/RmutFAC1 and LmutFAC2/RmutFAC2 for FAC1 binding sites (Table 4-1).

4.2.6 Transfection and luciferase reporter assays

1×10^5 cells were seeded per well in serum-free and antibiotic-free media in 12-well plates. 1.5 µg of total plasmid DNA [800 ng of cloned promoter-pGL4-luciferase plasmid (described in Section 4.2.4), 300 ng of pRL-TK renilla (Promega), 400 ng of pcDNA3 expression vector containing transcription factor or empty for control] was transfected using Metafectene[®]Pro reagent (Biontex Laboratories, Germany), in a ratio 1.5 µg : 3 µl. Total DNA concentration was adjusted with pcDNA3 (a gift from Dr Alison Shield, ANU). Expression vectors used were human *FAC1*-M2flag in pcDNA3.1, mouse *Bptf* in pcDNA, human *E2F1* in pcDNA, and human *pRb* (all gifts from Dr Kelly Jordon-Sciutto, UPenn, USA). Transfected cells were incubated for 24 h in a 37 °C incubator with 5% CO₂ and then trypsinised, resuspended in 50 µl of DMEM medium and transferred into a 96-well plate for measurement of luciferase activity using the Dual-Luciferase Reporter Assay System (Promega). The luciferase activity was normalised against the renilla expression produced by the pRL-TK renilla vector

(Promega). Mean values with standard deviation were calculated from at least two independent experiments performed in triplicate.

4.2.7 Bioinformatics analysis

Phylogenetic footprinting and TFBS analysis of *PRNP* was carried out by Nagesh Chakka (Chakka and Gready, 2006). Phylogenetic footprinting is a computational method developed for identifying TFBS in a single gene by comparing orthologous sequences from several species. The basic assumption is that functional regions, including TFBS, are under greater selective pressure, and will be more conserved among species than nonfunctional regions. Motif-discovery and alignment-based phylogenetic footprinting methods with the integrated TRANSFAC professional database (<http://www.gene-regulation.com/pub/databases.html>) were applied to develop a high throughput screening pipeline to search for conserved motifs. Animal species were selected based on their evolutionary distances and availability of sequence information at the time of analysis. Representative sequences are from human, dog, rat, mouse, cow, and sheep for eutherian mammals, *Monodelphis domestica* and tammar wallaby for marsupials, frog *Xenopus tropicalis* for amphibians, and chicken for avians. Analysed genomic sequence covered a 2 kb region upstream of the transcription start site of *PRNP*, the whole exon-intron region, and 2 kb downstream from the transcription stop site.

I conducted additional analysis of the mouse and human *PRND* promoter region using Genomatix software (www.genomatix.de).

4.3 Results and Discussion

4.3.1 Bioinformatics prediction of *PRNP* transcriptional control sites

The phylogenetic footprinting technique was used to predict transcription factor binding sites for the PrP gene. This method is based on the assumption that sequence motifs that have some biological significance are likely to be conserved under selective pressure during evolution. First, conserved motifs were obtained with motif-discovery method analysis of multiple orthologous DNA sequences from mammals, marsupials, frog, and chicken. Then, the conserved sequence motifs obtained were screened against the TFBS database (<http://www.gene-regulation.com/pub/databases.html>). Total numbers of predictions were 30 TFs for mammals, 15 TFs for marsupials, 6 TFs for amphibian, and

3 TFs for chicken. These conserved binding sites for eutherian mammals (human, dog, mouse, rat, cow, and sheep), marsupials (*M. domestica* and tammar wallaby), amphibian *X. tropicalis*, and avian are mapped in **Figure 4-1** to illustrate overlapping predictions. Predicted TFBSs were mainly found in intron 1 and 2 (**Figure 4-1**). Interestingly, the most conserved TFBSs between species are for the FAC1 and MAZ TFrs (as can be seen from overlapping regions of the diagram), positioned as a cluster in the 630 bp region next to exon 2. The third most conserved TF, MEF2, is situated in intron 2, about 3 kb upstream of the *PRNP* ORF. There is another MEF2 transcription factor binding site in the 630 bp region that is conserved only in mammals. The FAC1 TF was selected for further experimental analysis. Its conservation among evolutionarily diverse species may reflect early specialisation of the *PRNP* gene. Absence of the FAC1 TFBS in chicken *PRNP* conserved non-coding DNA region is intriguing, considering that the *PRND* gene is absent in chicken (Nagesh Chakka, PhD thesis, 2008).

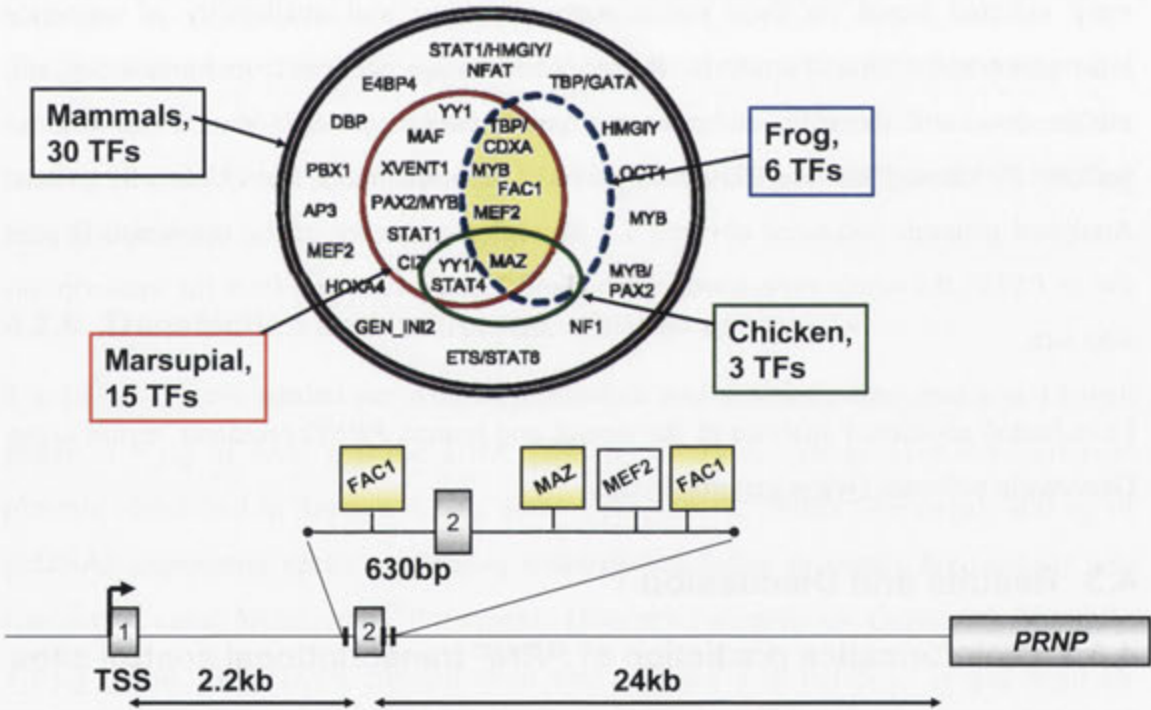


Figure 4-1. Transcription factors predicted in the *PRNP* conserved non-coding DNA region. The diagram illustrates overlapping predictions for 6 eutherian mammals (human, dog, mouse, rat, cow, and sheep), 2 marsupials (*M. domestica* and tammar wallaby), frog *X. tropicalis*, and chicken. Total numbers of predicted transcription factors (TFs) for each evolutionary animal group are indicated. The schematic of the mouse *Prnp* gene shows 2 non-coding exons and 1 coding exon, transcription start site (TSS), sizes of introns, the most conserved transcription factor binding sites among species (FAC1 and MAZ) and the MEF2 binding site, conserved in mammals only, positioned as a cluster in the 630 bp region next to exon 2.

4.3.2 Biological significance of FAC1, MEF2, and MAZ and their relationship with PrP and Dpl

The clustering of two of the most conserved TFs (FAC1 and MAZ) and mammalian-conserved MEF2 in the same 630 bp region of *PRNP* next to exon 2 makes them attractive candidates for experimental analysis. First, I searched key information about their function and possible relationship with PrP and Dpl expression.

The myocyte enhancer factor 2 (**MEF2**) family of TFs plays an important role in muscle differentiation (Martin et al., 1993). MEF2 is an essential regulator of cardiac myogenesis and right ventricular development (Lin et al., 1997b). Furthermore, MEF2s are also abundantly expressed in neurons and contribute to the development and function of the nervous system. Importance of MEF2 for development of the Purkinje cells in the cerebellum (Lyons et al., 1995), for vascular endothelium development (De Val et al., 2004), control of synapse number (Flavell et al., 2006), as well as expression of MEF2C in B cells and thymic medulla (Swanson et al., 1998), correlates well with possible roles for PrP activity, as both are involved in neuronal maturation and synaptic plasticity. MEF2 is able to recruit a histone methyltransferase to target gene promoters (Zhang et al., 2002) and is linked to both transcriptional repression through recruitment of histone deacetylase and activation by histone acetyltransferase interaction (Gruffat et al., 2002).

Fetal Alzheimer reactive Clone 1 (**FAC1**) was first identified in brain homogenates from patients with Alzheimer's disease (Bowser et al., 1995). Its alternatively spliced form was discovered later as BPTF (Jones et al., 2000). I describe FAC1 and BPTF proteins and their transcripts in detail in **Chapter 5** as BPTF variant 1 and variant 2.

There are several points of relationship between PrP, Dpl and FAC1.

- FAC1 and PrP are both involved in oxidative stress regulation (Strachan et al., 2004, Brown and Besinger, 1998). They might interact during common participation in the oxidative stress response Nrf2/Keap1 pathway. FAC1 interacts with Nrf2/Keap1 in the cytoplasm; in response to oxidative damage, Nrf2 is released (Strachan et al., 2005b). There is evidence of PrP interaction with Nrf2 (Yehiely et al., 1997).

- Both FAC1 and PrP are important for neurite outgrowth (Rhodes et al., 2003, Zawlik et al., 2006).
- PrP expression is upregulated in the dentate gyrus of the hippocampus after traumatic brain injury (Marciano et al., 2004, Weise et al., 2004). FAC1 demonstrates similar pattern of expression following entorhinal cortex lesion. (Styren et al., 1997).
- PrP, Dpl and FAC1 are all involved in Alzheimer's disease. PrP is abundant in synaptic terminals and aberrant neuritic of senile plaques in AD (Ferrer et al., 2001). Expression of Dpl increases in amyloid-containing diffuse and neuritic plaques (Ferrer et al., 2004). FAC1 is observed in diffuse and neuritic plaques (Schoonover et al., 1996) and activated microglia (Bowser and Reilly, 1998) in AD brain.
- FAC1 expression level increases rapidly following nerve growth factor (NGF) treatment of PC12 cells (Rhodes et al., 2003). PrP behaves similarly (Zawlik et al., 2006).
- FAC1 interacts with MAZ and reduces its transcriptional activity; both proteins show similar expression patterns in normal tissues and colocalise in AD brain (Jordan-Sciutto et al., 2000). The existence of a MAZ-binding site in the proximity of a FAC1-binding site is, thus, tantalizing evidence of a mechanism for such interaction and indicative of the power of the current bioinformatic predictions.

The link between FAC1, PrP, and Dpl points to FAC1 as a prospective candidate which simultaneously regulates PrP and Dpl.

4.3.3 Effect of FAC1 on 630bp-*Prnp* conserved non-coding DNA region

To analyse predicted FAC1-binding sites, I cloned mouse 190 bp, 320 bp, and 630 bp *Prnp* conserved non-coding DNA region fragments into the pGL4 luciferase reporter plasmid. The 630 bp fragment contained two FAC1, and one MAZ and MEF2 binding sites; in the 320 bp fragment the rightmost FAC1 and MEF2 sites were removed; the 190 bp fragment retained only the left FAC1 site (**Figure 4-2**). In addition, and taking account of the demonstrated interaction between FAC1 and MAZ (Jordan-Sciutto et al., 2000), point mutations were introduced into the MAZ binding site in two different

positions (Table 4-1). Unfortunately, attempts to mutate FAC1 binding sites were not successful.

Initially, mouse cerebral cortex endothelial bEND.3 cells were selected for transfections and luciferase assay experiments. Because the purpose of the experiment was to test for the coregulation of *PRNP* and *PRND* genes, the choice of cell line was made based on reported mouse *Prnd* promoter activity in bEND.3 cells (Nagyova et al., 2004). Comparison of luciferase activity between cells transfected with empty pGL4 vector and 630bp-*Prnp*-pGL4 construct demonstrated transcriptional activity of the 630 bp conserved non-coding DNA region with a 5-fold increase in luciferase activity. However, cotransfection of hFAC1-expressing plasmid with the 630bp-*Prnp*-pGL4 construct into the bEND.3 cell line showed no effect on luciferase activity (Figure 4-2).

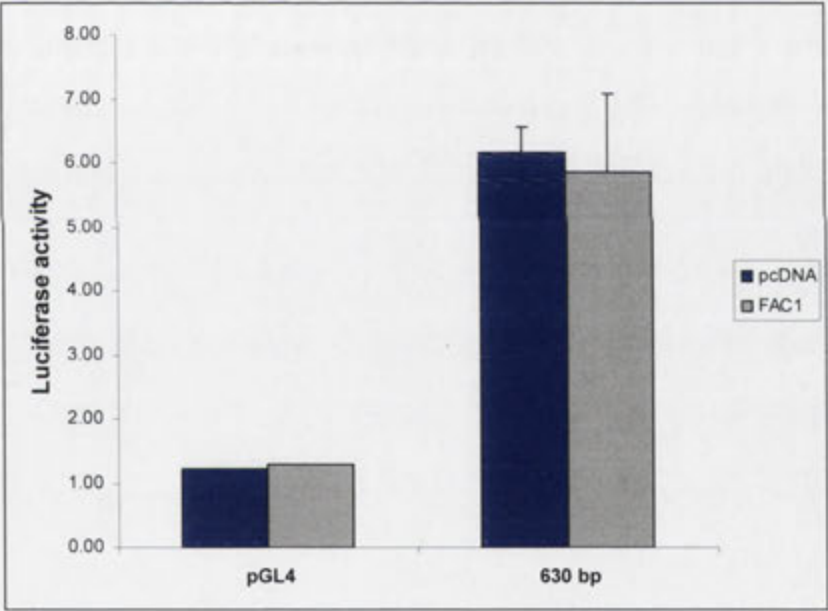


Figure 4-2. Activity of the 630bp-*Prnp* conserved non-coding DNA region. bEND.3 cells were transiently cotransfected with a pGL4 luciferase reporter vector or 630bp-*Prnp*-pGL4 construct and either pcDNA or hFAC1-expressing plasmid. Luciferase activity was assayed 24 h after transfection and normalised to pRL-TK renilla expression. The 630bp-*Prnp* region is transcriptionally active with a 5-fold increase in luciferase activity. Cotransfection with human FAC1-expressing plasmid has no further effect on promoter activation, possibly due to the 630 bp region being activated by endogenous factors. Standard deviation is indicated by the error bars; n (replicates) = 3, n (different biological samples) = 2.

The experiment was repeated in mouse primary spermatocyte GC-2 cells with additional mutated constructs. Both the 190 bp and 320 bp truncations of the 630 bp conserved non-coding DNA region and MAZ binding site mutated fragments were not affected by cotransfection with hFAC1 plasmid (**Figure 4-3**). However, comparison of luciferase activity of the original 630 bp *Prnp* conserved non-coding DNA region and its mutated variants demonstrated activity of the region in bEND.3 cells. Luciferase expression directed by the 320 bp construct was about 2.3 times lower than that induced by the 630 bp construct, whereas for the 190 bp construct carrying only one FAC1 binding site it was 2.4 times activated. Luciferase activity of the 630bp-*Prnp*-pGL4 construct was reduced 1.5 times by mutation of the MAZ-binding site (**Figure 4-3**).

The similarity of luciferase activity between control pcDNA plasmid and hFAC1-expressing pcDNA plasmid might be explained by presence of endogenous FAC1/BPTF in bEND.3 and GC2 cells or absence of some cofactors. It is interesting, that mutation of the MAZ binding site reduced promoter activity. This is consistent with studies of Jordan-Sciutto et al., (2000) which showed that FAC1 interacts with MAZ and represses its activity. So, it is possible that the MAZ binding site mutation affected activation of the promoter with endogenous FAC1/BPTF.

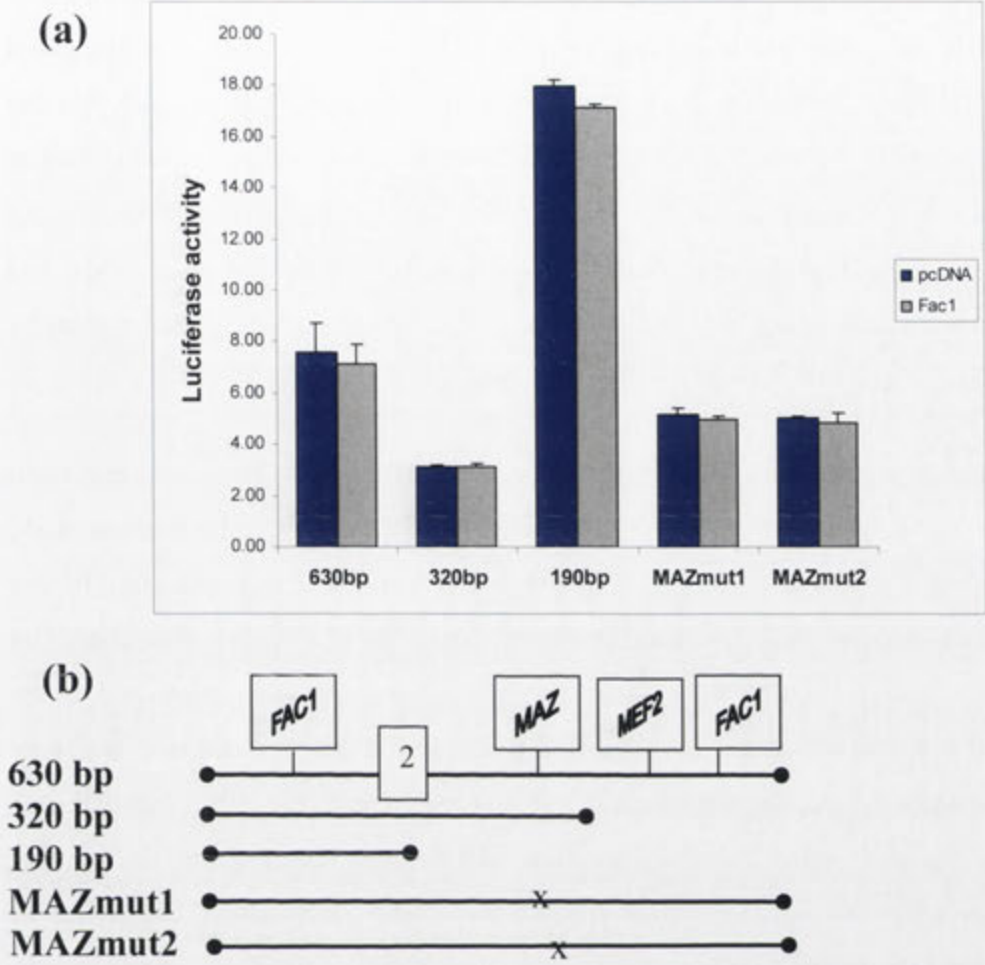


Figure 4-3. Transcriptional activity of *Prnp* conserved non-coding DNA region constructs in GC2 cells. (a) Cells were transiently cotransfected with *Prnp*-pGL4 constructs and pcDNA or hFAC1-expressing plasmid. Luciferase activity was assayed 24 h after transfection and normalised to pRL-TK renilla expression. Mutation or truncation of the 630bp-*Prnp* region results in variation in luciferase activity: 2.3-fold downregulation after truncation to 320 bp fragment, 2.4-fold upregulation after truncation to 190 bp fragment, and 1.5-fold downregulation after MAZ-binding site mutations. Standard deviation is indicated by the error bars; *n* (replicates) = 3, *n* (different biological samples) = 2. (b) *Prnp* conserved non-coding DNA region constructs created for luciferase assay. Schematic position of second *Prnp* exon and predicted transcription factors (FAC1, MAZ, and MEF2) are indicated on the 630 bp construct.

I hypothesised that cooperation of some cell-specific cofactors which are absent in BEND.3 and GC2 cells is needed to initiate FAC1 action. In order to test this, I compared 630bp-*Prnp* luciferase activity in various cell lines (mouse spermatogonia GC-1, mouse primary spermatocyte GC-2, mouse sertoli cell TM4, mouse cerebral

cortex endothelial bEND.3, and human neuroblastoma SHSY5Y) (**Figure 4-4**). Cells were cotransfected with the 630bp-*Prnp* luciferase construct and either with pcDNA3.1 or mouse *Bptf* in pcDNA3.1. Because *hFAC1*-M2flag-pcDNA expresses truncated protein (see **Chapter 5**), *Bptf*-pcDNA was used in this experiment to test whether the full length protein is required for promoter activation. The result showed a similar activation profile for cotransfection of bEND and GC2 cells with BPTF-expressing plasmid as for hFAC1-expressing plasmid (**Figure 4-4**). Cotransfection in TM4 and SHSY5Y cells resulted in activation of the *Prnp* conserved non-coding DNA region by approximately 2 and 3 times, respectively (**Figure 4-4**).

This latter outcome does not agree with hFAC1-mediated transcriptional repression reported previously in mouse embryonic fibroblast NIH3T3 cells (Jordan-Sciutto et al., 1999). This difference may be due to cell-specific variations in cofactors. In my experiments, luciferase activity was repressed only in the GC-1 cell line, by approximately 2-fold (**Figure 4-4**). However, the GC1 cells were suspected to be contaminated with mycoplasma (not analysed to confirm) and they did not look very healthy. This might explain the atypical result (compared with the other 4 cell lines I used). It is tempting to speculate that unhealthy cells have increased apoptosis, and that as a result, additional cofactors are expressed in these cells which contribute to downregulation of *Prnp* expression to shift the balance towards apoptosis. In conclusion, the variability in promoter activity in different cell lines shows that the effect of FAC1 is cell-specific and that the presence of some cofactors is required.

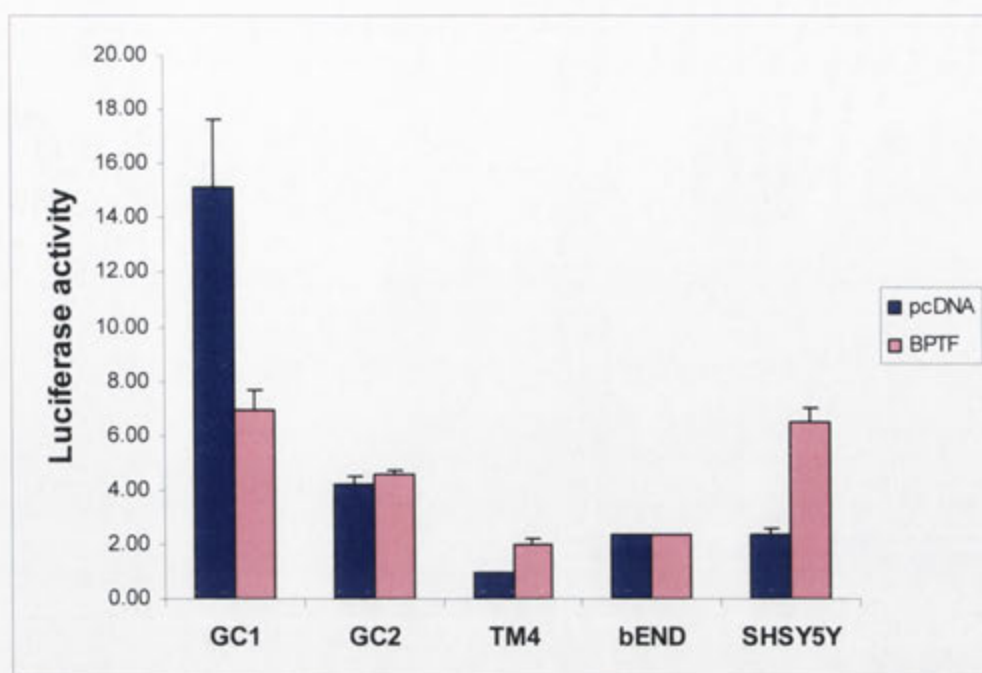


Figure 4-4. *Prnp* promoter activity in response to BPTF transfection has cell-specific response. Various cell lines were transiently cotransfected with the 630bp-*Prnp*-pGL4 construct and pcDNA or mouse *Bptf*-pcDNA plasmid. Luciferase activity was assayed 24 h after transfection and normalised to pRL-TK renilla expression. *Prnp*-pGL4 activity is 2.2-fold downregulated in GC1 cells, not affected in bEND and GC2 cells, 2.2-fold upregulated in TM4 cells, and 2.8-fold upregulated in SHSY5Y cells. GC-1, mouse spermatogonia; GC-2, mouse primary spermatocyte; TM4, mouse sertoli; bEND.3, mouse cerebral cortex endothelial; SHSY5Y, human neuroblastoma. Standard deviation is indicated by the error bars; **n** (replicates) = 3, **n** (different biological samples) = 2.

4.3.4 *PRND* promoter contains conserved transcription factor binding sites

The first step to develop a hypothesis about PrP and Dpl coregulation was to analyse the *PRND* promoter. Analysis of the proximal *PRND* promoter using Genomatix software resulted in prediction of E2F1, AP-2, CREB, USF1, and NFY TFBSs. As described in the literature, the E2F1 site can be organized in regulatory modules or clusters of TFs with AP-2 (Jin et al., 2006), CREB and NFY (Elkon et al., 2003). Thus, the presence of this cluster of binding sites in the 650bp *PRND* proximal promoter and its conservation between mouse and human (**Figure 4-5**) provide strong support for the reliability of this prediction.

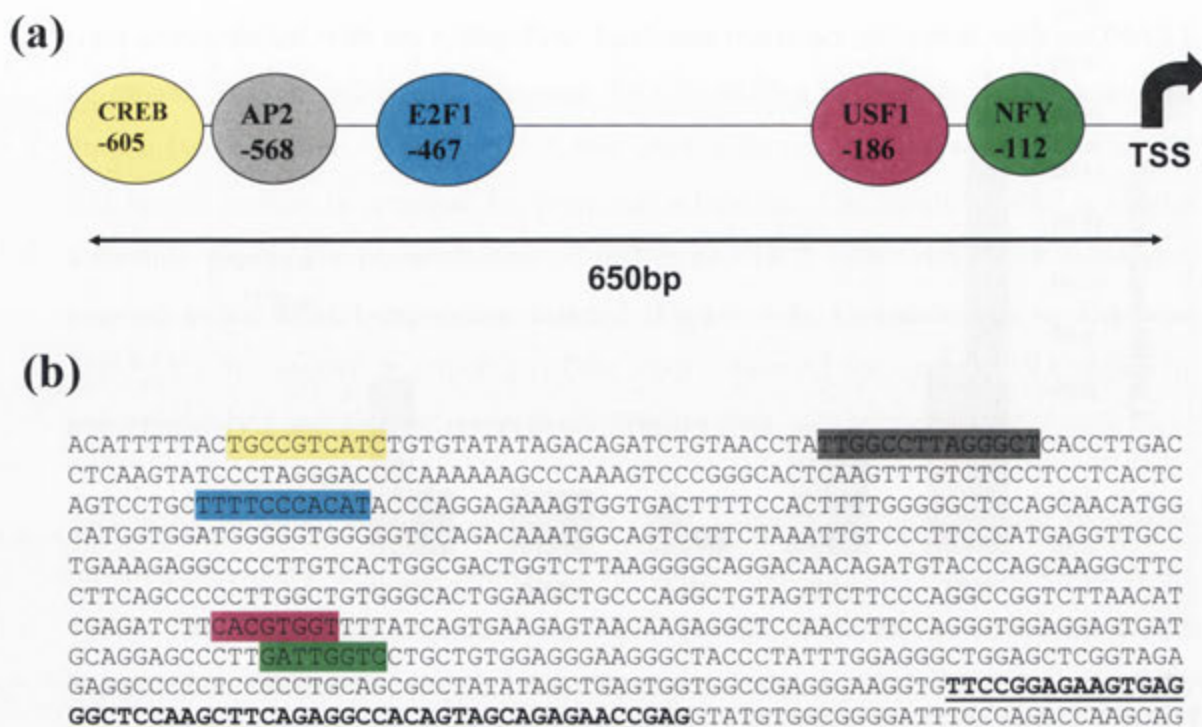


Figure 4-5. Genomatix prediction of transcription factor binding sites (TFBS) on the proximal *PRND* promoter. (a) Arrangement of predicted TFBS cluster in the 650 bp mouse *Prnd* proximal promoter. Numbers correspond to position of TFBS in the mouse *Prnd* proximal promoter relative to TSS. TSS, transcription start site. **(b)** The nucleotide sequence of the 650 bp mouse *Prnd* proximal promoter. Positions of TFBS are highlighted: CREB in yellow, AP-2 in grey, E2F1 in blue, USF1 in magenta, NFY in green. Exon 1 is in bold and underlined.

4.3.5 BPTF activates and E2F1 represses the 650bp-*Prnd* promoter

To verify if *Prnd* promoter activity can be modulated by FAC1/BPTF, the 650 bp mouse *Prnd* promoter fragment was cloned into the pGL4 luciferase reporter plasmid and cotransfected with mouse *Bptf*-pcDNA expressing plasmid into the different cell lines (GC-2, TM4, bEND.3, SHSY5Y) (**Figure 4-6**). Cotransfection resulted in increase of luciferase activity in GC-2 cells (1.4-fold), TM4 cells (2.6-fold), and SHSY5Y cells (2-fold), but had no effect in bEND cells. As in the case of the 630bp-*Prnp*-pGL4 experiments, this can be explained by strong promoter activation by endogenous factors.

I next examined the role of E2F1 in repression of the *Prnd* promoter. Luciferase assays were again conducted in the different cell lines (GC2, SHSY5Y, TM4, and bEND.3) (**Figure 4-6**). The human *E2F1*-pcDNA expressing plasmid or control pcDNA3 plasmid

was cotransfected with 650bp-*Prnd*-pGL4. There was a decrease of luciferase activity in GC2 cells (3.7-fold), in TM4 cells (2.9-fold), in bEND cells (3.1-fold), and in SHSY5Y cells (5.5-fold) compared with pcDNA3 control, supporting the prediction that E2F1 might be involved in *Prnd* regulation (**Figure 4-6**).

To provide further evidence for direct binding of E2F1 to the *Prnd* promoter, a point mutation T(-465)C was introduced into the predicted E2F1 binding site (**Figure 4-5b**, TTTTCCCACAT⁻⁴⁵⁷) with site-directed-mutagenesis primers LmutE2F1 and RmutE2F1 (**Table 4-1**). However, the mutation did not abolish E2F1-mediated downregulation of the *Prnd* promoter (**Figure 4-6**, tested in TM4 cell line only). It might be possible that E2F1 still binds the *Prnd* promoter because of ineffectiveness of the chosen point mutation. The E2F1 consensus binding motif is defined as TTTSSCGC (S stands for C or G) (Tao et al., 1997). However, whole genome analysis of E2F1-binding sites revealed that this canonical consensus binding motif is not common but rather it is very diverse in its nucleotide composition (Bieda et al., 2006, Rabinovich et al., 2008, Cao et al., 2011); this might cause problems with selecting the right point mutation to study E2F1 binding.

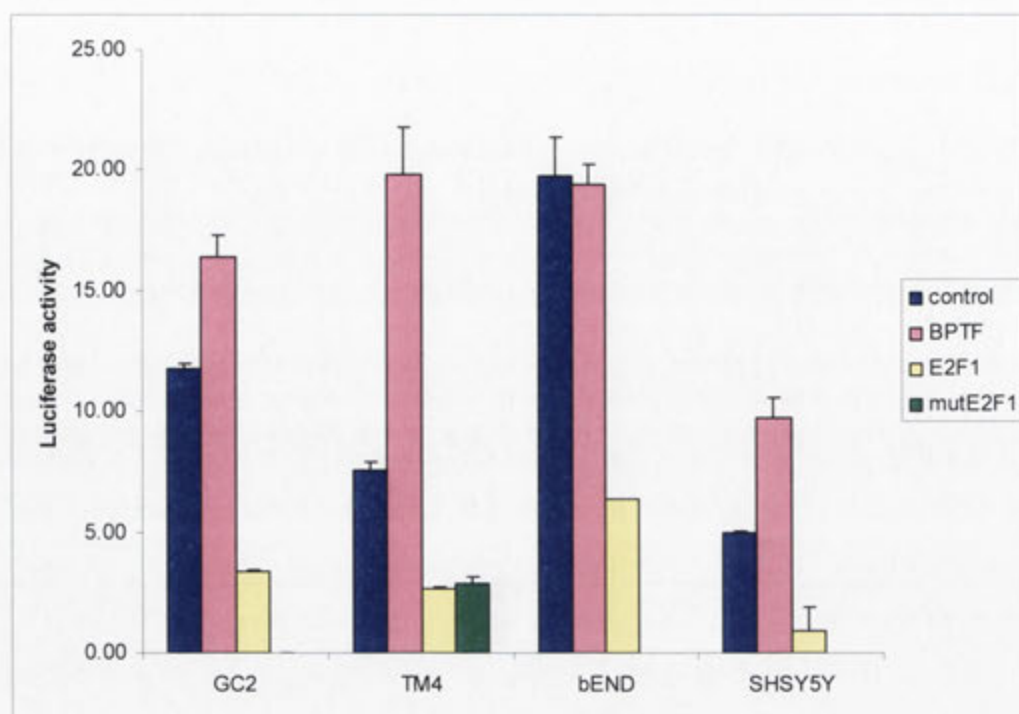


Figure 4-6. Cotransfection with *Bptf*-pcDNA activates and with *E2F1*-pcDNA suppresses the proximal mouse *Prnd* promoter. Comparison of 650bp-*Prnd*-pGL4 promoter activity in response to cotransfection with either control pcDNA plasmid or *Bptf*-pcDNA or *E2F1*-pcDNA plasmids. The 650bp-*Prnd*-pGL4 with mutated E2F1 binding site (mutE2F1) was cotransfected with *E2F1*-pcDNA into TM4 cell line. Luciferase activity was assayed 24 h after transfection and normalised to pRL-TK renilla expression. GC2, TM4, bEND3, and SHSY5Y cell lines were transfected and analysed. Standard deviation is indicated by the error bars; **n** (replicates) = 3, **n** (different biological samples) = 2.

4.3.6 Tissue-specific correlation of *Fac1/Bptf*, *E2F1*, and *Prnd* expression

To investigate the relationship between FAC1/BPTF, E2F1 and Dpl experimentally, I conducted tissue expression analysis by RT-PCR amplification of corresponding RNAs isolated from a range of mouse tissues (embryonic brain (E14), adult brain, testis, thymus, spleen, heart, muscle, kidney, liver, lung) (**Figure 4-7**). Interestingly, previously unreported splice variants of mouse *Bptf* are detected in different amounts in all these tissues and are major transcripts in muscle and heart, as will be discussed in **Chapter 5**. I will call them variants 3-6 to continue the NCBI-proposed numbering of transcript isoforms as variant 1 encoding for BPTF and variant 2 encoding for FAC1.

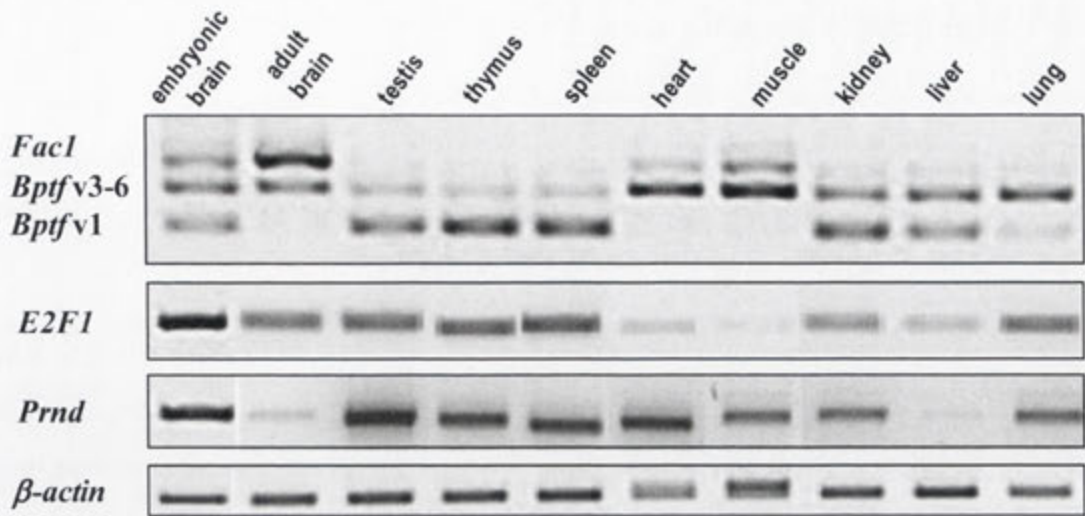


Figure 4-7. Tissue-specific distribution of *Fac1/Bptf*, *E2F1*, and *Prnd* transcripts in mouse tissues by RT-PCR. Analysed tissues are embryonic brain (stage E14), adult brain, testis, thymus, spleen, heart, muscle, kidney, liver, and lung. Top band on the *Fac1/Bptf* expression panel corresponds to *Fac1* (*Bptf* variant 2) signature; middle band reveals *Bptf* variants 3-6 with only one exon spliced, bottom band corresponds to *Bptf* (variant 1) signature (see Chapter 5 for detailed explanation). β -actin is shown as RT-PCR control. Original gel images with a molecular marker and no template control can be found in **Appendix 3**.

The observed patterns can be organized in three groups: brain, where *Fac1* is present and *Prnd* expression is suppressed in adults, possibly due to high level of *E2F1*; then E14 embryonic brain/testis/thymus/spleen, where *Bptf* v1 is present and both *Prnd* and *E2F1* are highly expressed and, so, *E2F1* repression might be alleviated by BPTF due to chromatin remodelling; muscle and heart, where *E2F1* is absent and no repression of *Prnd* is observed, and *Bptf* v3-6 are the most abundant transcripts; kidney and lung, where *Bptf* v3-6 seems to be equally or more abundant than *Bptf* v1 and both *E2F1* and *Prnd* are expressed; and liver that has a similar *Bptf* and *E2F1* expression profile to kidney and lung, but no *Prnd*, which indicates a different regulatory mechanism.

In conclusion, the tissue expression results support the proposed model where *E2F1* suppresses *Prnd* expression and this suppression can be removed by BPTF or its variants, coded by *Bptf* v3-6. *FAC1* expression does not remove *E2F1*-mediated repression of the *Prnd* promoter.

4.3.7 Expression of *FAC1/BPTF* and *PRND* changes throughout brain development

In order to emphasize the importance of various *BPTF* transcripts during development and verify the relevance of selected cell lines to study *BPTF*-mediated *PRND* induction, *FAC1/BPTF*, *PRNP*, and *PRND* expression was compared in brain tissues at different stages and in cell lines, both human and mouse (**Figure 4-8**).

The amount of mouse *Bptf* v1 in brain decreases from embryonic day 14 (E14) to adult stage, when it is no longer expressed. In contrast, expression of mouse *Fac1* increases becoming the strongest in adult stage; this is more apparent in the adult human brain sample. *Bptf* v3-6 is present in all stages. As expected, the increase of *Fac1* expression correlates with decrease of *Prnd* expression in brain. Both mouse spermatogonia GC-1 and mouse cerebral cortex endothelial bEND cell lines lack *Fac1* but express *Bptf* v1 and v3-6. Human neuroblastoma SHSY5Y cells show the most similarity to brain expression, resembling the mP15 pattern, however, *PRND* is absent in these cells, as well as in mouse bEND.3 cells. Based on these results, it would be best to use SHSY5Y cells for luciferase-assay analysis of *PRND* activation by *FAC1/BPTF*.

The comparison between tissue and cell-line expression profiles also emphasizes the problem of translating interpretations of results obtained from cultured cells to tissue samples and the importance of careful selection of the cell line model.

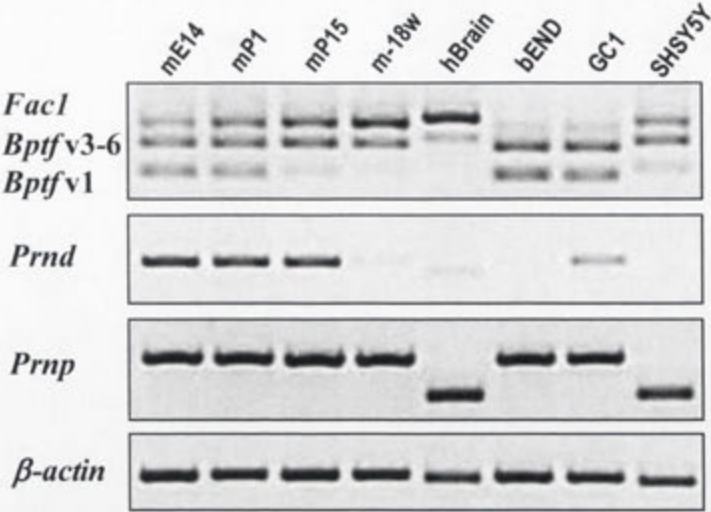


Figure 4-8. Expression of *Fac1/Bptf*, *Prnd*, and *Prnp* in mouse and human brain and selected cell lines. Comparison of expression between stages of brain development (mE14, mouse embryonic day 14; mP1, mouse postnatal day 1; mP15, mouse postnatal day 15; m-18w, mouse adult 18 weeks old; hBrain, human adult brain), two mouse cell lines, bEND and GC1, and human cell line, SHSY5Y. The top band on the *Fac1/Bptf* expression panel corresponds to *Fac1* (*Bptf* variant 2) signature; middle band reveals *Bptf* variants 3-6 with only one exon spliced, bottom band corresponds to *Bptf* (variant 1) signature (see Chapter 5 for detailed explanation). PCR products for human and mouse differ in sizes as a result of selected primers positions. β -actin is shown as RT-PCR control. Original gel images with a molecular marker and no template control can be found in **Appendix 3**.

4.3.8 Mechanism of PrP and Dpl coregulation

Experiments described in previous sections demonstrated that both *Prnp* and *Prnd* expression can be affected by FAC1/BPTF, although the FAC1-binding site is not predicted to be present in the *Prnd* promoter. This raises the question of the mechanism of this activation. My proposed hypothesis is based on my experimental results and literature-derived observations:

1. FAC1/BPTF has a predicted retinoblastoma protein (Rb)-binding site (Zhu and Bowser, 1996) (also see details in **Chapter 5, Figures 5.1 and 5.8**).
2. FAC1 overexpression appears to induce Rb phosphorylation and FAC1 binds phosphorylated isoforms of Rb as demonstrated by experimental analysis (Jordan-Sciutto, unpublished, personal communication).

3. The *Prnp* promoter has heat shock transcription factor (HSF) binding sites (Shyu et al., 2002); *Drosophila* NURF interacts with HSF (Xiao et al., 2001).

Therefore, I hypothesise key principles of PrP/Dpl coregulation, speculating that the FAC1 might activate transcription by binding to the predicted distant binding site.

(Figure 4-9):

1. In mature neurons FAC1 is bound to the *Prnp* conserved non-coding DNA region and might have repressing effect on the *Prnd* promoter. The *Prnp* promoter might not be affected.
2. During development, in Alzheimer's or under oxidative stress FAC1/BPTF directly or indirectly phosphorylates Rb; this leads to downstream effects on the E2F1-Rb gene regulation mechanism analogous to that of the basic cell cycle mechanism: when Rb is phosphorylated and no longer suppresses E2F1 activity, the *Prnd* promoter is activated which leads to apoptosis.
3. Recruitment of BPTF to USF1 is an initial step followed by Rb phosphorylation and E2F1 activation. BPTF-containing nucleosome remodelling complex NURF might compete with histone deacetylase complex HDAC1 for RbAp48 and facilitate E2F1 transcriptional activation of the the *Prnd* promoter involving CREB/CPB-mediated histone acetylation.
4. BPTF might be recruited to the *Prnp* promoter by HSF and increase *Prnp* expression resulting in neuronal growth or neuroprotection.
5. Additional unknown cell-specific factors are involved to direct FAC1/BPTF action towards *Prnp* or *Prnd* promoters. This effect might also be produced by cell-specific BPTF variants that are described in **Chapter 5**.

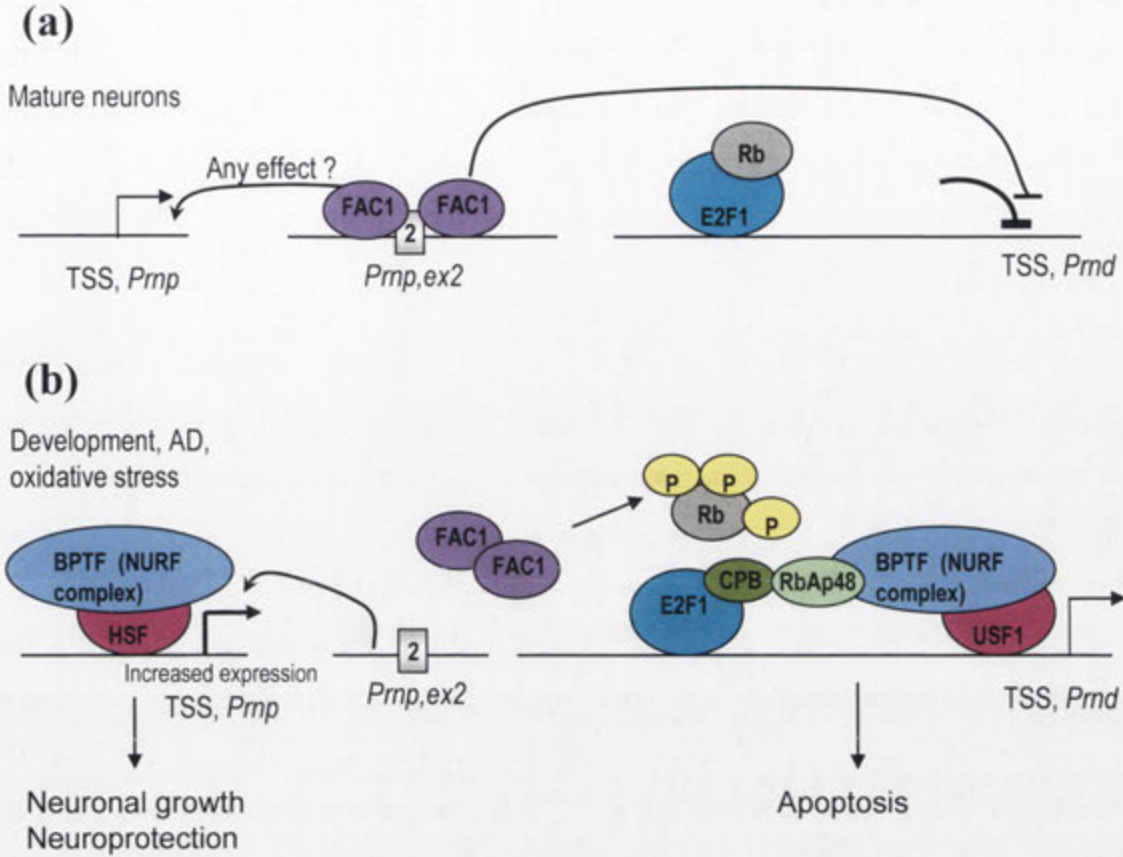


Figure 4-9. The proposed model of how FAC1/BPTF might activate the *Prnd* promoter. (a) In mature neurons FAC1 binds distant binding site next to the *Prnp* exon 2 and contributes to *Prnd* promoter repression mediated by the E2F1 assembly with Rb. The FAC1 might regulate *Prnd* expression by binding to the predicted distant binding site. **(b)** FAC1 does not bind DNA during development or Alzheimer's (AD), Rb is phosphorylated and removed from E2F1, and the *Prnd* promoter is activated. Recruitment of the BPTF/NURF chromatin remodelling complex by the USF1 transcription factor and its CBP/RbAp48-mediated interaction with E2F1 might be involved. TSS, transcription start site.

In vitro luciferase assay experiments demonstrated that FAC1 and BPTF have the same effect on the *Prnp* and *Prnd* promoters. The question remains how they act *in vivo* and how they could influence both the *Prnp* and *Prnd* genes. It is known that enhancers can modulate tissue-specific gene expression (Blackwood and Kadonaga, 1998). Simultaneous regulation from distant enhancers was reported for tandem paralogs Yp1/Yp2 (Chung et al., 1996) and Dlx6/Dlx4 (Zerucha et al., 2000) and even for unrelated neighbouring genes APOC3/APOA1 (Kan et al., 2000). So far, BPTF has been shown to be involved in chromatin remodelling and FAC1 is described as a

transcription factor. FAC1 differs from BPTF by the presence of the PEST domain and absence of the bromodomain. FAC1-induced apoptosis requires the presence of only the Rb-binding domain but doesn't require the PEST domain (Jordan-Sciutto, unpublished, personal communication). Therefore, I suggest that all BPTF isoforms might contribute to *Prnd* promoter activation by inducing both Rb phosphorylation and chromatin remodelling.

The FAC1/BPTF isoforms might create functionally distinct chromatin remodelling complexes to achieve cell-specific regulation. This kind of regulation was reported for *Drosophila* NURF (nucleosome remodelling factor) chromatin remodelling complex (Kwon et al., 2009). The largest subunit of *Drosophila* NURF, NURF301, is an ortholog of human BPTF with 3 full-length isoforms and one short isoform, C-terminally truncated. The short isoform is coexpressed with full-length NURF301 and can form a complex with all other NURF subunits, although it lacks the ability to target the H3K4me3 histone (Kwon et al., 2009). The formation of different FAC1/BPTF isoform NURF complexes might explain cell-type-specific altered expression of PrP and Dpl in AD with PrP activated in diffuse plaques, neuritic plaques, and amyloid cores of senile plaques and Dpl around amyloid deposits of neuritic plaques (Ferrer et al., 2001).

Additional information from the literature, presented below, can help with deeper understanding of the steps of this hypothesis and involved factors.

- E2F1 is a critical transcriptional regulator of the cell cycle and controls the G1/S transition (Johnson et al., 1993). Derepression of gene transcription is modulated through E2F1 interaction with retinoblastoma protein, Rb (Helin et al., 1993), and with CREB Binding Protein, CBP (Trouche et al., 1996). Rb protein contains multiple phosphorylation sites, and phosphorylation and dephosphorylation controls its activity. In particular, if Rb is hypophosphorylated it can form a complex with E2F1 at the promoter and repress transcription, whereas if Rb is hyperphosphorylated this complex is disrupted, and the repression is removed enabling free E2F1 to activate transcription of its target genes (Harbour and Dean, 2000). Rb is involved in tumor suppression and its inactivation is associated with the development of retinoblastoma and other types of human cancers (Weinberg, 1995, Burkhardt and

Sage, 2008). Thus, a balance between proliferation and apoptosis is controlled by E2F1 and Rb levels. It is reported that E2F1 modulates apoptosis in neurons and induces neuronal loss (Hou et al., 2000, O'hare et al., 2000, Strachan et al., 2005a, Dasgupta et al., 2006) and is involved in oxidative stress (Macleod, 2008). The lack of E2F1 in cortical neurons results in reduced proapoptotic protein BAX and caspase-3 activity and protects neurons from β -amyloid toxicity-mediated death (Giovanni et al., 2000). This hints that E2F1 might regulate *Prnd* as Dpl-induced apoptosis is linked to BAX and caspase-3 (Qin et al., 2006b, Didonna et al., 2012).

- BPTF-containing complex NURF can be recruited to the 5'HS4 insulator in the chicken β -globin and human α -spectrin loci in erythrocytes by USF1 protein (Li et al., 2011). USF1 mostly regulates chromatin structure accessibility and recruitment of transcription complexes (Crusselle-Davis et al., 2006) and, in case of *Prnd*, is described as a transcription factor capable of activating the proximal *Prnd* promoter (Sepelakova et al., 2005).
- RbAp48 (or RBBP4 retinoblastoma binding protein 4) binds at the C-terminal end of Rb protein (Qian et al., 1993). RbAp48 is suggested to target chromatin assembly factors, chromatin remodelling factors, histone acetyltransferases and histone deacetylases to their histone substrates, H4 in particular (Verreault et al., 1998). In addition, RbAp48 can mediate binding of several proteins, illustrated by distinct interaction sites for chromatin-modifying subunit, histone H4, and GATA-1 cofactor FOG-1 providing a link between transcriptional regulation and chromosome remodelling (Lejon et al., 2011). Therefore, it might be possible that RbAp48 brings together a kinase and Rb to promote Rb phosphorylation. RbAp48 is reported as a mediator of both transcriptional repression and activation. RbAp48 association with histone deacetylase HDAC1 in a Rb-containing complex mediates transcriptional repression of E2F-regulated genes (Nicolas et al., 2000). On the other hand, RbAp48 binding to phosphorylated CREB/CBP complex promotes histone acetylation during transcriptional activation (Zhang et al., 2000). In addition, NURF-mediated expression of genes determining cell fate during vulval development was described in *C. elegans* where NURF-like complex antagonises chromatin

remodelling activities of other complexes, repressors of proapoptotic genes (Andersen et al., 2006).

- FAC1 was shown to interact with CPB (Jordan-Sciutto, unpublished, personal communication). Both BPTF and CPB are required to activate Smad-dependent gene targets essential for cell proliferation and distal visceral endoderm development in early mouse embryos and embryonic stem cells (Landry et al., 2008). Similarly, CBP can bind E2F1 directly stimulating E2F1 transcriptional activation capacity (Trouche et al., 1996, Fry et al., 1999).
- Heat shock transcription factor (HSF1) is ubiquitously expressed and is implicated in cellular adaptation and survival (Page et al., 2006), aging (Ankar and Sistonen, 2011), and brain development (Walsh et al., 1997). Depletion of HSF1 promotes senescence, while its overexpression protects from senescence (Kim et al., 2012). A protective role of HSF1 against apoptosis under oxidative stress was reported in cardiac myocytes (Zhang et al., 2011) and under ischemic brain injury (Tucker et al., 2011). Binding sites for this transcription factor are found in both *Prnp* (Shyu et al., 2002) and *βAPP* promoters (Dewji and Do, 1996). Experiments with HSF1 knockout mice demonstrated its protective effect against prion-induced neurotoxicity (Steele et al., 2008). It is suggested that overexpression of HSF1 mediated by injection of lentivirus vector-HSF1 into cerebella of AD rats could rescue the Purkinje cells from degeneration (Jiang et al., 2013).

4.4 Conclusion

The results presented in this Chapter support the phylogenetic footprinting prediction of FAC1 binding sites in the *Prnp* conserved non-coding DNA region situated next to the second non-coding exon and point to a mechanism for PrP and Dpl coregulation by FAC1/BPTF mediated through the E2F1 binding site in the *Prnd* core promoter. This is confirmed by increase in the 630bp-*Prnp* and 650bp-*Prnd* promoter activity after cotransfection with FAC1/BPTF-expressing plasmids and 650bp-*Prnd* promoter suppression by E2F1-expressing plasmid. A conceptual approach was used to identify a relationship between expression of FAC1, PrP and Dpl in Alzheimer's, oxidative stress and neuronal development by analysing published data, and using it to design experiments to test it.

Coregulation of PrP and Dpl might be an important pathway regulating neuronal apoptosis. This chapter provides the first attempt to elucidate the mechanism of the PrP and Dpl antagonistic relationship by providing a hypothetical model based on the phylogenetically predicted FAC1/BPTF binding site, and experimental and conceptual analysis. However, the proposed model is exploratory and alternatives are possible. Unfortunately, no research has been published so far that would allow further speculation about the PrP/Dpl co-regulation mechanism. A more comprehensive study is needed to examine how PrP and Dpl are co-regulated.

Further confirmatory work might include:

- (1) Mutagenesis experiments of the predicted E2F1 binding site to demonstrate abolition of E2F1 recruitment to the *Prnd* promoter.
- (2) ChIP analysis to demonstrate direct binding of FAC1 to the *Prnp* conserved non-coding DNA region and E2F1 to *Prnd* promoter. The best approach would be comparison of expression in a healthy- and disease-state mouse model, or in embryonic brain, where Dpl is expressed, and adult brain, where Dpl is suppressed. In addition, based on previous results that both FAC1 and PrP are involved in neurite elongation, cell models might be used; appropriate models are rat adrenal PC2 cells or human neuroblastoma SHSY5Y cells. It has been shown that expression of both FAC1 and PrP is activated after PC2 cells are treated with NGF (Rhodes et al., 2003, Zawlik et al., 2006). Similarly, SHSY5Y cells differentiate into cells with characteristics of mature neurons after retinoic acid and BDNF treatment (Agholme et al., 2010). It is expected that FAC1 binding would be observed in untreated cells and BPTF binding in treated cells. Investigations by such cellular and animal models might help to develop better understanding of coregulation of PrP and Dpl during apoptosis.

Identification of all elements involved in the BPTF-chromatin remodelling mechanism in the *Prnd* promoter would contribute to knowledge of chromatin remodelling. The proximity of CREB/CBP, E2F1/Rb, and USF/NURF binding sites (all of which are reported to interact with RbAp48) in the *Prnd* promoter is intriguing. It would be interesting to check if other reported CREB/E2F1 clusters of binding sites have a USF1 binding site in proximity; this would provide evidence that the BPTF mechanism of activation is more universal.

Resolving issues regarding the PrP/Dpl coregulation pathway should allow a better understanding of the complexity of neuronal death and survival. These results underline the necessity for more study of the functions of BPTF and its involvement in gene regulation and chromatin remodelling. The next chapter (**Chapter 5**) is aimed to answer some questions about BPTF transcripts and their structure.

Chapter 5.

BPTF and its variants

In chapter 4 I proposed a model for coregulation of PrP and Dpl by the FAC1/BPTF transcription factor. Motifs for binding of FAC1/BPTF to the conserved non-coding DNA region of the *PRNP* gene – close to exon 2 - were predicted by phylogenetic footprinting analysis, based on their conservation during the evolution. I chose FAC1/BPTF for experimental study to test the predictions based on the similarity of its tissue distribution to that of PrP and Dpl, and because of predicted roles in neuronal development, oxidative stress, Alzheimer's disease, and spermatogenesis overlap with those predicted for PrP and Dpl. Little information is available on FAC1/BPTF in the literature. This motivated me to investigate its expression in mouse tissues. In this chapter I first describe my discovery of multiple FAC1/BPTF alternative transcripts and their tissue expression patterns, and then analyse the possible effects of alternative splicing on FAC1/BPTF functions. I will refer to proteins in accordance with their literature descriptions, BPTF for the full-length protein and FAC1 for its short, C-terminal truncated variant. As both proteins are expressed from the same gene, *BPTF*, I will refer to its alternative mRNA transcripts as *BPTF* transcripts using *BPTF/BPTF* as a generic name for all animals or *BPTF/Bptf* when presenting results for mouse.

5.1 Introduction

5.1.1 BPTF and its functions

The National Center for Biotechnology Information (NCBI) annotated two transcripts of the human *BPTF* gene and only one for mouse. Consequently, two proteins are described in the literature. FAC1 was identified first from screening senile plaques with monoclonal antibody Alz50 specific to Alzheimer's disease pathology (Bowser et al., 1995). It is described as a DNA-binding protein that represses transcription (Jordan-Sciutto et al., 1999). FAC1 is expressed at high levels during human brain and spinal cord development and at low levels in adult brain (Bowser et al., 1995, Mu et al., 1997). FAC1 expression increases in amyloid-containing diffuse and neuritic plaques in Alzheimer's disease (Schoonover et al., 1996). Some suggested roles of FAC1 are in

neuronal development, synaptogenesis, neurodegenerative diseases (Mu et al., 1997, Styren et al., 1997), oxidative stress response (Strachan et al., 2004), and apoptosis (Strachan et al., 2005b). Location of FAC1 in the cytoplasm during development and neuronal differentiation (Rhodes et al., 2003) might be explained by a possible role in activation of stored mRNA; this would be consistent with a proposed role in neuronal plasticity. As a member of the PHD/LAP zinc finger family, FAC1 might be involved in chromatin-mediated transcriptional regulation (Aasland et al., 1995).

The second protein, BPTF, is a more recently discovered transcription factor with extensive identity to the smaller FAC1 protein and alternatively transcribed from the same gene (Jones et al., 2000). In contrast with FAC1, BPTF is a transcriptional activator (Barak et al., 2003). Interestingly, recent microarray analysis of *Bptf* knockout mouse embryonic stem cells shows that BPTF can act as an activator or a repressor of some genes (Landry et al., 2008). *BPTF* expression analysis by Northern blot revealed ubiquitous expression, with highest levels of transcription in testis, spleen and thymus (Jones et al., 2000). In brain, *BPTF* transcript has been detected within the granule and Purkinje cell layer of the cerebellum and the pyramidal neurons throughout the hippocampus (Barak et al., 2003). BPTF is essential for early mouse embryogenesis (Goller et al., 2008, Landry et al., 2008) and regulates thymocyte maturation (Landry et al., 2011).

BPTF differs from FAC1 by containing the bromodomain at the C-terminal end. The bromodomain binds acetyl-lysine on histones and is found in many chromatin-associated proteins, transcription factors, and in nearly all nuclear histone acetyltransferases (Zeng and Zhou, 2002). Consequently, BPTF is part of a nucleosome-remodelling factor (hNURF) complex (Barak et al., 2003). Moreover, the second PHD finger of BPTF, situated upstream of the bromodomain, contains a binding site to histone H3 trimethylated at lysine 4 (H3K4me3) (Li et al., 2006a, Wysocka et al., 2006). H3K4me3 is associated with transcriptional activation (Santos-Rosa et al., 2002, Schneider et al., 2004). This correlates with BPTF-mediated activation of *Erg1*, a homeodomain protein that regulates neuronal development in the mid-hindbrain and thymocyte maturation (Barak et al., 2003, Landry et al., 2011). Similarly, the BPTF-containing NURF complex potentiates neurite outgrowth in cell culture (Barak et al., 2003).

FAC1 and BPTF proteins have similar N-terminal domains with a DNA-binding domain, PHD-zinc finger and Rb-binding motifs, but different C-terminal ends with glutamine-rich and bromodomain regions for BPTF and a PEST domain with a KEAP1-binding site for FAC1 (**Figure 5-1**). It is hypothesised that the PEST sequence acts as a signal peptide for protein degradation (Rogers et al., 1986). KEAP1 has a predicted role as a FAC1-interacting protein in neuronal response to oxidative stress (Strachan et al., 2004). The retinoblastoma protein, Rb, is an important regulator of the cell cycle (Weinberg, 1995), through E2F inhibition (Dyson, 1998), in particular, but also by interaction with chromatin remodelling proteins (Magnaghi-Jaulin et al., 1998, Dunaief et al., 1994). BPTF encodes a protein that is very similar to the largest subunit of the *Drosophila* NURF (nucleosome remodelling factor) complex (Barak et al., 2003). The NURF complex catalyzes nucleosome sliding on DNA and interacts with sequence-specific transcription factors (Xiao et al., 2001).

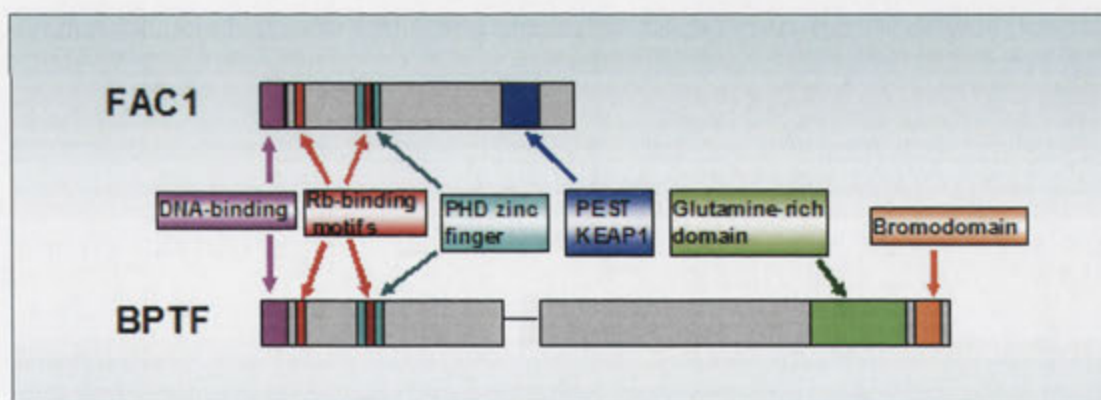


Figure 5-1. FAC1 and BPTF protein structure. Fac1, 810 amino acids. BPTF, 2781 amino acids, counting from TSS-2 (**Figure 5-8**). Common domains (DNA-binding motif, Rb-binding motifs, and PHD zinc finger) are located in the N-terminal end of both proteins. The FAC1 C-terminal end contains the PEST domain and KEAP1-binding site. The alternative BPTF C-terminal end contains the glutamine-rich domain and bromodomain. Figure is modified from a slide of Dr Kelly L. Jordan-Sciutto, University of Pennsylvania (personal communication).

It is necessary to note that the FAC1 protein has been observed only as a short, 125 kDa, protein, although only a large 10.5 kb mRNA has been detected by Northern blot analysis (Bowser et al., 1995). It has been assumed that the protein is truncated post-translationally, but no mechanism for this has been suggested.

The use of similar RNA probes and antibodies for detection of FAC1 and BPTF (despite their obvious structural differences) has created confusion about which protein or mRNA transcript is actually described. My goal in this chapter is to document the inconsistencies and fill in some gaps in BPTF-related research by characterising the alternative splicing of the mouse BPTF gene, and determine if different transcripts are preferentially expressed in different tissues and therefore have tissue specific functions.

5.2 Materials and methods

5.2.1 Total RNA extraction

Adult BALB/C mouse tissues were collected and preserved in RNAlater solution (Ambion). Total RNA was isolated from brain, testis, heart, eye, kidney, liver, lung, muscle, spleen, and thymus tissues. All RNA extractions were conducted using the RNeasy Mini Kit (Qiagen). DNase treatment (Ambion) was included to remove residual DNA contamination.

5.2.2 RT-PCR analysis of tissue distribution

First strand cDNA for use in subsequent amplification reactions was synthesised from 1 µg of total RNA extracted from each mouse tissue (brain, testis, kidney, liver, lung, muscle, spleen, eye, heart, thymus), using an oligo(dT) primer and SuperscriptIII reverse transcriptase (Invitrogen), following the manufacturer's instructions. 1 µl of the first strand cDNA reaction was used as a template in all subsequent amplification reactions that were performed with Platinum Taq polymerase (Invitrogen). Primer pairs LmmFac1/RmmFac1 and LmmBptf/RmmBptf were used to analyse expression profiles of *Bptf* transcripts (Table 5-1). To avoid genomic DNA amplification, primer pairs were designed as specific to different exons (in that case DNA amplification would result in a larger size PCR band). H₂O control was used to test for template contamination.

5.2.3 Amplification of large *Bptf* fragment

1 µg of total RNA from each mouse tissue (brain, testis, heart, eye, kidney, thymus) was primed by an oligo(dT) primer and first strand cDNAs were synthesised using SuperscriptIII reverse transcriptase (Invitrogen), following the manufacturer's instructions. 1 µl of the first-strand cDNA was used as a template for PCR with 0.2 µM *Bptf*-specific primers, Lm6Bptf and Rm6Bptf, to amplify 6kb sequence covering alternatively spliced regions (**Table 5-1**). The PCR was conducted using AccuTaqTM LA DNA Polymerase (Sigma) specially designed for long sequence amplification. After an initial 30 sec denaturation at 98 °C the long PCR amplification reaction carried out 5 cycles of (94 °C for 15 sec, 68 °C for 30 sec, 68 °C for 6 min), followed by 30 cycles of (94 °C for 15 sec, 65 °C for 30 sec, 68 °C for 6 min), with a final elongation step at 68 °C for 30 min.

Table 5-1. Summary of primers used in *Bptf* expression and sequencing analysis.

Superscript numbers position primer relative to ORF (exons 5 and 6 are present, exon 23 is unspliced).

Lm6Bptf	1629	GGATGTAACGGAAGACCTGACCAA	1652
Rm6Bptf	8184	AGAGCTTGCTGGCATTCTGTTTG	8206
LmmFac1	1762	GACCCAGAAAAGGACAAGCA	1781
RmmFac1	2345	ACGAGTGGACCCAGATGTTT	2364
LmmBptf	7515	TTCACAGCCGATTCCAATTC	7534
RmmBptf	8122	GCCGCCTGTTTTTCTTCTTT	8141
LhmmReg1-1	1908	GCCAACAGAAGTTGGGGATAAAGG	1931
RmmFac24	2348	TGATGATACGAGTGGACCCAGATG	2371
RmmReg1-2	2220	CTCCAATCTCAGGGTCATCC	2239
LmBptf1	2334	AGCTGGGAAAGGAACATCTG	2352
LmBptf2	2902	TGTGCAGTTAAGCCAGTTGTG	2921
LmBptf3	3542	GTGAGGGCTTCCATCTAAGG	3561
LmBptf4	4241	TCGATGGGAGACTAGGTGGT	4260
LmBptf5	5015	TGGAAACTGAAAGTCCAGAGG	5035
LmBptf6	5753	TGAGCCTGATGTTACGGTTG	5772
LmBptf7	6401	AGTCCAATTCAGGCATGGTC	6420
LmBptf8	7127	GCTCAGGGGAACAAAAACAG	7146
RmBptf1	7478	ACTTGGGATGGAGTTGATGG	7459
RmBptf2	6736	CAGAGATGGCAGTGGAACA	6717
RmBptf3	5982	TGTGGGTGTTTCTTTTGGTG	5963
RmBptf4	5263	TGGAAGAAACCACAGTTTGC	5244
RmBptf5	4500	GCCAGTTTCTTACCGTTTG	4481
RmBptf6	3748	CTTTTGTGGAGCCTGGAGAC	3729
RmBptf7	3165	AACGTGCGTCTTGCTAATCC	3146
RmBptf8	2483	ACCACCAGCACCTCTTCC	2465

5.2.4 *Bptf*-specific library analysis

PCR products were purified by isolation from agarose gels and cloned into the TOPO-TA sequencing vector (Invitrogen), transformed into TOP10 bacterial competent cells (Invitrogen), and grown overnight on LB-plates containing 100µg/ml ampicillin. 25 colonies were randomly picked from each plate and resuspended in 20 µl of water. 5 µl aliquots were used for PCR analyses with primer pairs specific to the different alternatively spliced regions: LmmFac1/RmmFac1 for Region 1, LmmBptf/RmmBptf for Region 2, LhmmReg1-1/RmmFac-24 for Region 1-1, and LmmFac1/RmmReg1-2 for Region 1-2 (**Table 5-1**). PCR reactions were performed with Platinum Taq polymerase (Invitrogen). Following an initial denaturation step of 94 °C for 5 min, the PCR amplification reaction consisted of 35 cycles of (94 °C 30 sec, 58 °C 30 sec, 72 °C 30 sec), followed by a final extension step of 72 °C 10 min for Regions 1 and 2; and 94 °C 5 min, (94 °C 30 sec, 64 °C 30 sec, 72 °C 30 sec) 35 cycles, 72 °C 10 min for Regions 1-1 and 1-2. To confirm identity of transcripts, each transcript variant was sequenced in both directions using the *Bptf*-specific primers LmBptf1-8 and RmBptf1-8 (**Table 5-1**), and the pCR®-TOPO vector-specific T3 and T7 sequencing primers).

5.3 Results and Discussion

5.3.1 Discovery of new transcript variants and their expression profiles in mouse tissues.

The tissue-specific distribution of mouse *Bptf* gene variants has not been characterised previously. Human BPTF is highly expressed in testis (Jones et al., 2000); its isoform, FAC1, is brain-specific (Bowser et al., 1995). I conducted initial mouse-tissue expression analysis based on the NCBI-annotated human transcript variants. Human *BPTF* is a large transcript about 11 kb long, characterised by two alternatively spliced regions (**Figure 5-2**). Region 1 includes two alternatively spliced exons at the 5' end of the open reading frame. BPTF protein variant 1 is encoded by the transcript with both exons in Region 1 absent. Region 2 is alternatively spliced from a cryptic site, creating a truncated exon. Absence of Region 2 is a signature of BPTF protein variant 2 for which the C-terminal truncated human version is named FAC1.

In order to determine whether each of the predicted *Bptf* transcript variants existed in the mouse, I constructed primers to amplify separately two variant-specific regions of

mouse *Bptf* (**Table 5-1, Figure 5-3 (a)**). I also carried out all amplifications on a panel of mouse tissues to collect information about the expression profile of each variant. I anticipated detecting 603 bp or 228 bp sized bands for PCRs over Region 1, indicating *Bptf* v2 and *Bptf* v1, respectively, and 640 bp or 230 bp size bands for PCRs over Region 2, indicating *Bptf* v2 and *Bptf* v1, respectively. RT-PCR analysis of Region 2 produced the expected results (**Figure 5-3 (c)**). However, Region 1 analysis gave the unexpected outcome of a three-band pattern (**Figure 5-3 (b)**). The NCBI BPTF structure predicted only two spliced variants for the two exons of Region 1, both present or both absent. This prediction was confirmed by PCR amplification (**Figure 5-3**, 603 bp band and 228 bp band). The third band (414 bp), for Region 1 indicates a third alternative with only one of the two exons present. Furthermore, as both exons of Region 1 are exactly the same size, four different isoforms of Region 1 can be predicted, i.e. with exon 5 or with exon 6. RT-PCR analysis of Region 2 produced bands of 640 bp and 230 bp as expected (**Figure 2c**). Therefore, taking into account the two possible splice variants of Region 2, a total of eight alternative splice variants are possible for the *Bptf* gene.

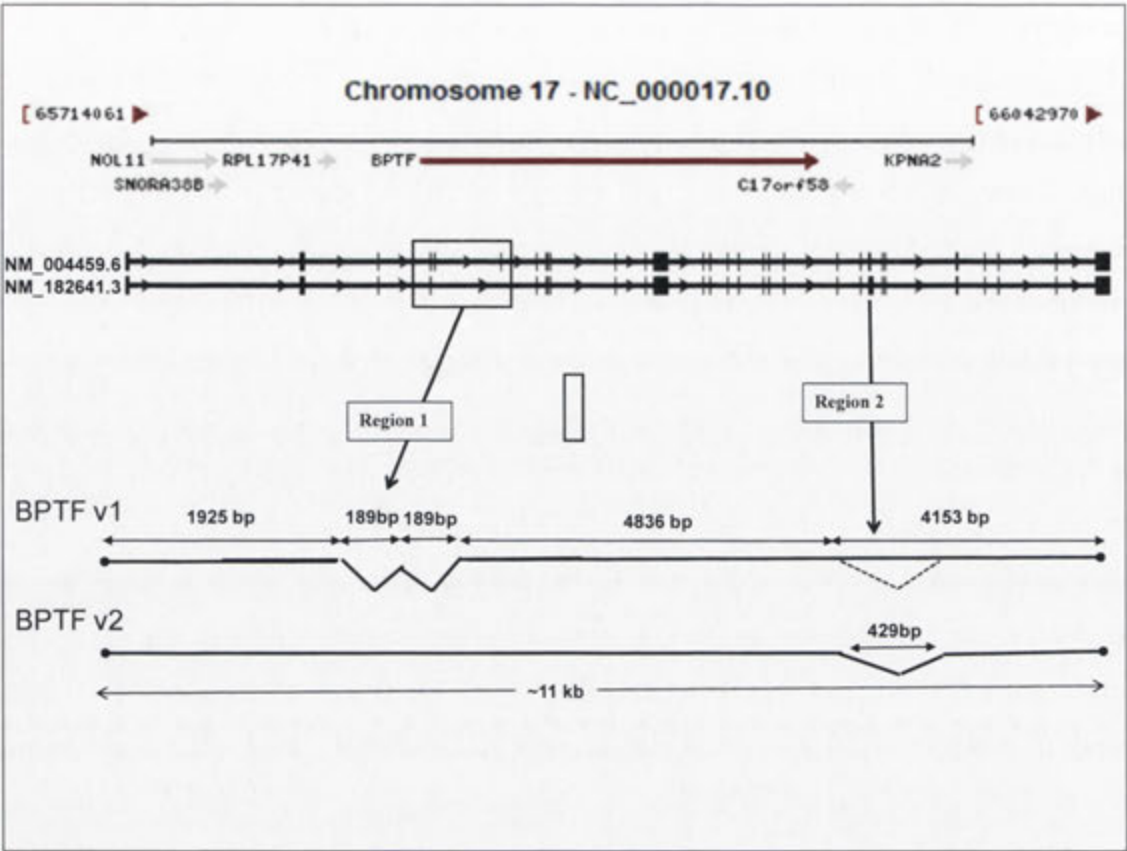


Figure 5-2. NCBI-annotated human *BPTF* variants (2011). *BPTF* v1 is distinguished by two alternatively spliced exons of equal size at the 5' end (exons 5-6), referred to as Region 1; *BPTF* v2 is distinguished by an alternatively spliced exon at the 3' end (exon 23), referred to as Region 2.

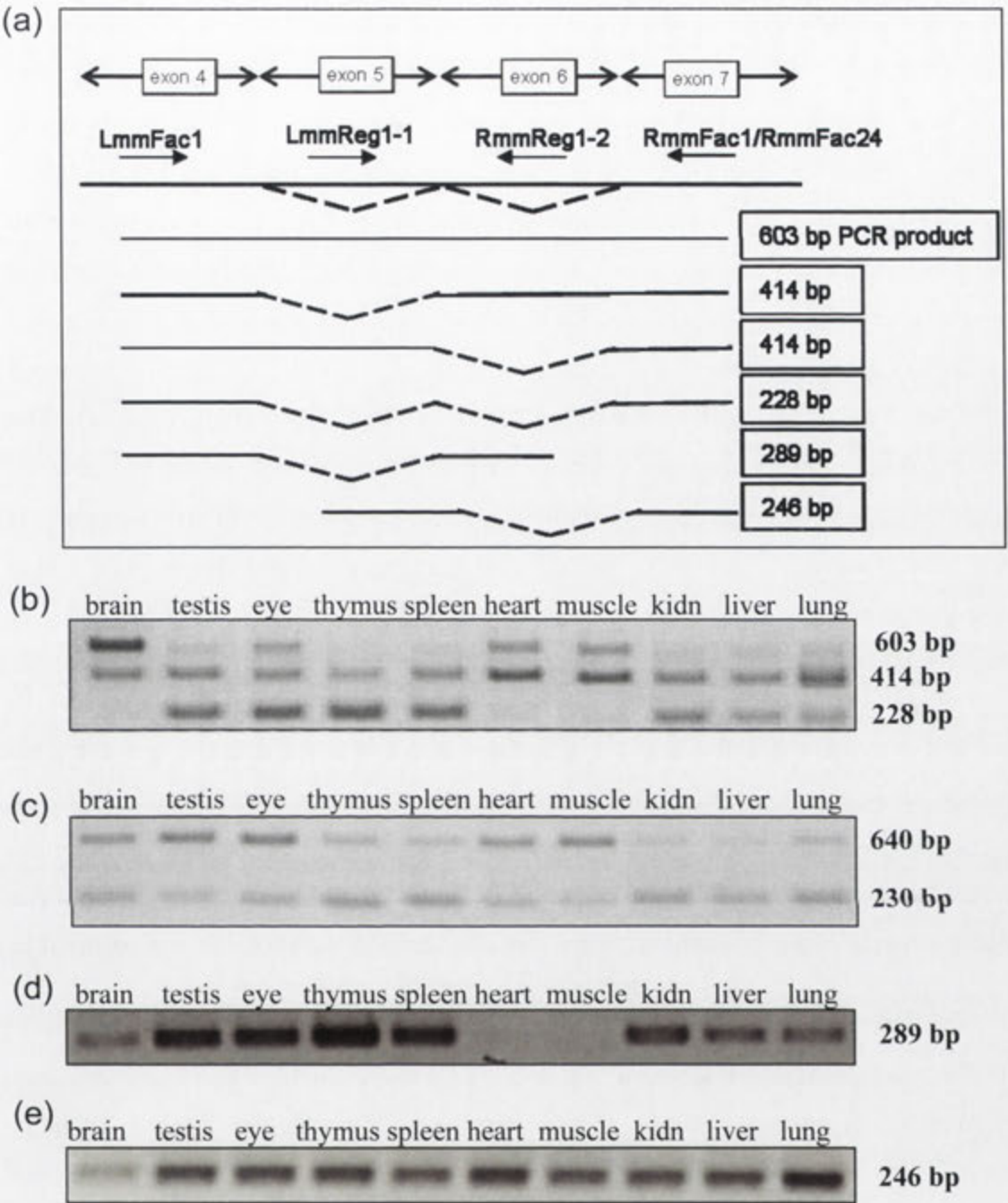


Figure 5-3. Tissue-specific distribution of *Bptf* variants. (a) Schematic map of the Region 1 exons, exon-specific PCR primers, and the resulting products (shown in gel images (b), (d), and (e)). (b) RT-PCR analysis of Region 1. Top band corresponds to *Bptf* v2 signature; bottom band corresponds to *Bptf* v1 signature; middle band reveals unknown *Bptf* variants with only one exon spliced. (c) RT-PCR analysis of Region 2. Top band corresponds to *Bptf* v1 signature; bottom band corresponds to *Bptf* v2 signature. (d) RT-PCR analysis of Region 1 exons showing that transcript variants without exon 5 are present in all tissues, except heart and muscle. (e) RT-PCR analysis of Region 1 exons showing that transcript variants without exon 6 are present in all tissues. All the PCRs were done at the same time; successful amplification from analysed tissues using other primer pairs serves as a positive control. Genomic DNA amplification was prevented by long intronic sequences. Original gel images with a molecular marker and no template control can be found in the **Appendix 3**.

I investigated further the existence of transcript variants of Region 1 by PCR with exon-specific primers. First, I analysed the presence of exons 5 and 6 using forward primer LmmFac1 specific to exon 4 and reverse primer RmmReg1-2 specific to exon 6 of Region 1 (**Figure 5-3 (a)**). The PCR product, resolved on an agarose gel, corresponds to a 289 bp fragment when exon 5 is absent but exon 6 is present. This is the case for all tissues tested except for heart and muscle. Absence of PCR products for heart and muscle tissues indicates the existence of transcript variants in which exon 6 of Region 1 is absent (**Figure 5-3 (d)**). Next, I used forward primer LmmReg1-1 specific to exon 5 of Region 1 and reverse primer RmmFac24 specific to the exon 7 (**Figure 5-3 (a)**). The appearance of a 246 bp fragment indicates transcripts with exon 6 absent. Existence of transcript variants of this type is confirmed for all analysed tissues (**Figure 5-3 (e)**). It is thus clear from these RT-PCR studies that there are multiple tissue-specific *Bptf* variants containing either of the two exons of Region 1.

Overall, RT-PCR analysis revealed considerable tissue specificity for a panel of 10 mouse tissues and the presence of multiple alternative transcript variants. However, this method for assessing tissue distribution of alternative transcripts cannot disclose the full picture. Thus, although ubiquitous expression of human *BPTF* was shown in 16 normal tissues by Northern blot (Bowser et al., 1995, Jones et al., 2000), this probe did not discriminate between possible transcript variants. A different approach was required to characterise *Bptf* variants.

In the next sections I describe current methods available for discovery and characterisation of alternative transcripts, and their deficiencies for analysis of large transcripts. I then present the method I developed to define tissue distribution of *Bptf* transcripts more accurately.

5.3.2 Existing computational and experimental methods cannot detect BPTF tissue-specific alternative splicing

The problem encountered arises from the large 11 kb size of *Bptf* transcripts and the relatively small size of the spliced exons situated 4.8 kb apart (**Figure 5-2**). An additional complexity is that the two exons of Region 1 are of equal size (**Figure 5-2**). In such circumstances the alternative transcripts are not resolvable on gels. To date, NCBI has annotated two *BPTF* transcript isoforms for human (NM_004459.6 and NM_182641.3) and only one for mouse (NM_176850.2). Protein-coding sequences of

both predicted human isoforms are very similar in size. There are two alternatively spliced regions: the first contains two exons, each of 189 bp, and the second of 429 bp alternatively spliced within one exon from a cryptic splice site (**Figure 5-2**). Either of the two exons of Region 1 can be spliced separately. The Region 1 diversity has remained undetected because previous experimental analysis has been done by Northern blot. The combination of three alternatively spliced exons over two regions produces eight possible mouse *Bptf* transcript variants.

The fact that all eight combinations of Region 1 and 2 configurations could be present and significant for BPTF function seems not to have been contemplated before. On average, 2.7 alternate transcripts have been found for human genes that are alternatively spliced (Takeda et al., 2006) and somewhat less for mouse and chicken (~2.3) (Tang et al., 2007). These averages determined from current mappings of ESTs to genome sequence are likely to underestimate alternate-splicing frequency due to incomplete and uneven EST coverage and other hidden biases. But nonetheless this instance of eight variants is well above average and might signify a particularly refined role of BPTF as a transcription and chromatin remodelling factor.

Missed computational predictions for *BPTF* variants are a good example of the insufficient coverage of rare alternative splice forms, despite the huge collections of mRNA-transcript data (ESTs, cDNA clones). The current method of prediction of alternative transcripts involves alignment of EST and cDNA sequences against genomic DNA. However, many putative alternative splice forms can be missed because EST sequences are biased towards the 5' and 3' ends of genes. Even when sufficient EST data for particular genes are available, there are problems of assembly and mapping to the genome (Tang et al., 2007). In the case of large or rare transcripts, as exemplified by *BPTF*, coverage of the internal sequence might be insufficient to correctly predict splice forms.

It is necessary to note that transcript finishing initiatives have been undertaken to bridge gaps between EST clusters mapped against the genomic sequence (Sogayar et al., 2004). However, although this laborious approach allows detection of low abundance transcripts in a restricted set of tissues, it validates EST clusters separated from each other by an average 2879 bp which is still inadequate for multiple alternatively spliced exons situated far apart, 4.8 kb in the case of *BPTF*. These issues are compounded if we

need to assess tissue distribution of the alternative transcripts because of the uneven coverage of EST tissue libraries (Gupta et al., 2004).

Standard approaches to solve these problems are RT-PCR and microarray analysis and *in situ* hybridisation. However, these methods are useful only if spliced exons are situated in one region of the transcript. Prediction of tissue specificity becomes more complicated in cases where transcripts have two or more spliced regions separated by common exons. RT-PCR amplification over multiple regions with consecutive agarose gel separation and band-size analysis is achievable only for relatively small transcripts. It is more difficult to discern variability in large transcripts, taking into account that about 80% of the exons in the human genome are <200 bp in length (Sakharkar et al., 2004).

In summary, application of standard computational or experimental methods of analysis cannot detect putative alternative transcripts of Regions 1 and 2 of *Bptf*. To address this difficulty, I designed a novel method to analyse tissue expression of the newly predicted *Bptf* transcripts.

5.3.3 Novel method for discovery of alternative splicing and tissue specificity of large transcripts.

5.3.3.1 Construction of a library of alternatively spliced *Bptf* transcripts.

To confirm the predicted alternative transcripts and to increase the possibility of finding all variants, I isolated RNA from six mouse tissues and coamplified all possible mRNA alternate splice forms in a single PCR reaction. The selected normal mouse tissues (brain, testis, heart, eye, kidney, and thymus) represent different patterns of *Bptf* expression (**Figure 5-3**). PCR primers (**Table 5-1**, Lm6Bptf/Rm6Bptf and **Figure 5-4 (a)**) were designed to amplify 6.2-6.6 kb sequences that include both alternatively spliced Regions 1 and 2 (**Figure 5-4 (a)**). Based on the sequence difference being only ~3% and spliced regions not being GC-rich, I have assumed that it is likely that all eight transcripts would be amplified similarly in a single PCR reaction and, thus, that the products of the long PCR reactions would correctly reflect the composition of the transcript pool. The resulting PCR products, depending on alternate transcription, have size differences of 189 bp, 378 bp, 429 bp, 618 bp, and 807 bp and are visible on an agarose gel as a single smeared band (**Figure 5-4 (b)**). Each smeared PCR band for each

of the tissue samples was isolated, cloned into a plasmid vector and transformed into an *E. coli* bacterial host to generate six tissue-specific libraries of *Bptf* alternative transcripts. Each library contained multiple single bacterial colonies on a Petri agar plate. Each single colony harbored only one transcript variant.

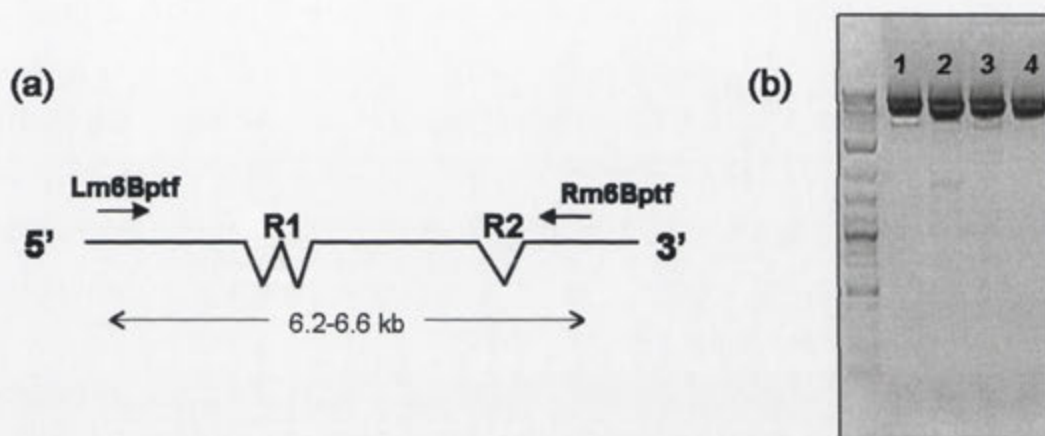


Figure 5-4. Amplification of *Bptf* fragments from various mice tissues. (a) Schematic representation of amplified *Bptf* fragment, Regions 1 (R1) and 2 (R2) are shown. (b) 6.2-6.6 kb *Bptf* fragments amplified from mouse brain (1), testis (2), heart (3), and kidney (4) run as single smeared bands on agarose gel.

5.3.3.2 Validation of *Bptf* alternative transcription and analysis of tissue distribution

To assess the tissue distribution of the alternative transcripts, 25 colonies of each tissue-specific library were randomly selected for PCR analysis, which used four pairs of primers to distinguish between each of the alternative splicing possibilities in Regions 1 and 2 (**Figure 5-5** and **Table 5-1**). Each of the picked bacterial colonies was resuspended in 20 μ l of double distilled water. Aliquots of 5 μ l were used for PCR over the alternatively spliced regions to identify transcripts. As the analysis in **Section 5.3.1** determined, up to four different PCR reactions are necessary (**Figure 5-3**). The first reaction, with primer pairs LmmFac1/RmmFac1, allows definition of the number of exons in Region 1; the top PCR band corresponds to transcript with both exons of Region 1 present, the bottom PCR band to transcripts with both exons spliced, and the middle PCR band for transcripts with only a single exon in Region 1 present (**Figure 5-**

3 (b), **Figure 5-5 (a)**). The second reaction, with primer pairs LmmBptf/RmmBptf, allows detection of alternative splicing of Region 2; the upper PCR band corresponds to unspliced Region 2, and the lower PCR band to the spliced variant (**Figure 5-3 (c)**, **Figure 5-5 (b)**). These two PCRs were conducted for all selected clones. Application of the third and fourth PCR reactions depends on the outcome: clones with a single exon in Region 1 require further analysis with primer pairs LmmReg1-1/RmmFac24 to detect whether the second exon of Region 1 is present or absent, and LmmFac1/RmmReg1-2 to confirm whether the first exon of Region 1 is present or absent. Thus, they analyse which of the two exons of equal size is present (**Figure 5-3 (a, d, e)**, **Figure 5-5 (c, d)**). PCR products were analysed by agarose gel electrophoresis. The nature of splice variants was deduced by matching consecutive PCR results for both regions to individual clones. Isolated plasmids harboring different predicted variants were sequenced to confirm their identity.

Overall, PCR analysis of individual clones confirmed all eight predicted alternate splice variants and gave an indication of their specific tissue distribution.

In addition to tissue-specific patterns of alternative *Bptf* transcription, clone analysis allowed a rough assessment of expression levels of different transcripts by calculating how many of the analysed bacterial colonies from each tissue carried transcripts of each particular variant (**Figure 5-6**). This result was very interesting as it suggested that each tissue has one major tissue-specific variant (or two in case of brain and heart). Other alternative splice variants were detected as well but at lower levels.

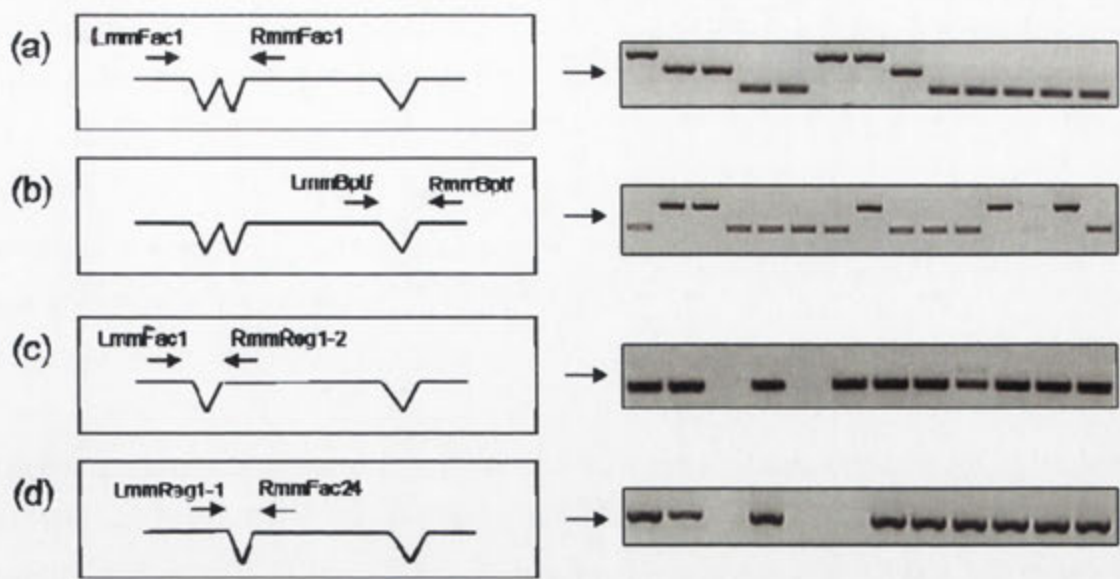


Figure 5-5. Scheme for four PCR reactions performed on bacterial colonies harboring cloned *Bptf* fragments and examples of amplified products resolved on agarose gels. (a) PCR over Region 1; three amplification products allow preliminary characterisation of exons in Region 1. The top, middle and bottom bands correspond to exon count two, one and zero, respectively. (b) PCR over Region 2; upper band corresponds to unspliced exon, lower band to spliced variant. (c) PCR conducted with the primer specific to the second exon of Region 1 (exon 6); absence of band signifies absence of exon 6. (d) PCR conducted with the primer specific to the first exon of Region 1 (exon 5); absence of band signifies absence of exon 5.

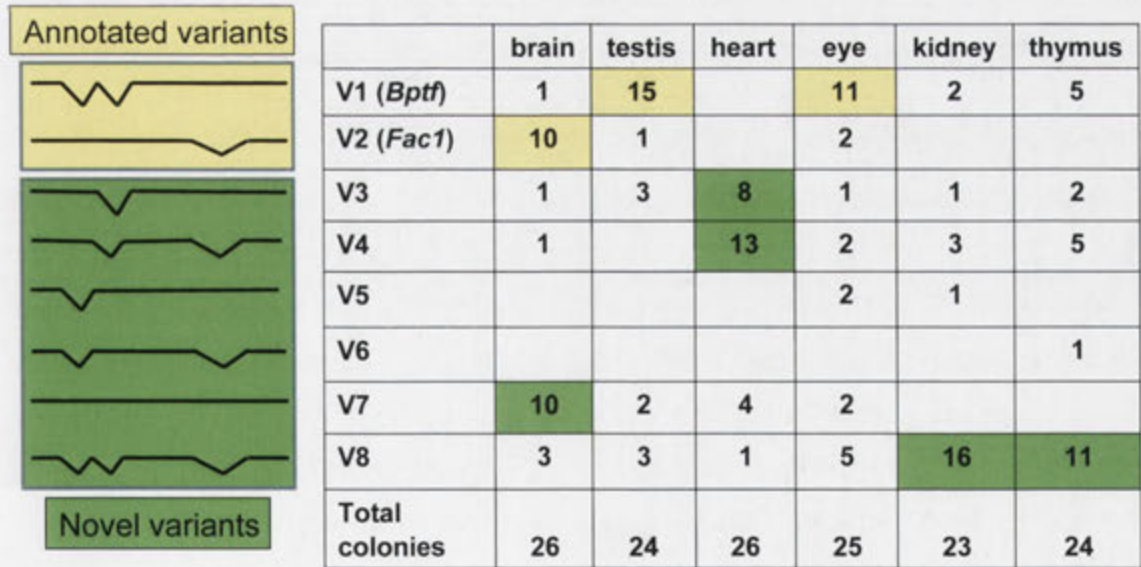


Figure 5-6. Relative abundance of transcripts in mouse tissues from colony analysis. Brain, testis, heart, eye, kidney, and thymus tissues were analysed. *Bptf* transcript variants (v1-v8), including previously annotated and novel variants are presented schematically. Total numbers of analysed colonies as well as the number of colonies found for each *Bptf* variant are shown in columns. Major tissue-specific variants are colored in yellow for annotated variants and in green for novel variants.

FAC1 has been described in the literature as a brain-specific transcription factor (Bowser et al., 1995). Indeed, a codominant variant for brain is FAC1 (10 colonies out of 26 analysed), but variant 7 has similar abundance to FAC1. As for FAC1, it is characterised by the presence of both exons of Region 1, but unlike FAC1 contains unspliced Region 2. Therefore, it appears that the full version of Region 1 is highly specific to brain. Exons of Region 1 code for the PEST domain and KEAP1-binding site (**Figure 5-1**). FAC1 transcript is present only in very low quantities in testis (1 out of 24 analysed colonies) and eye (2 out of 25 colonies), and variant 7 is similarly low in testis (2 out of 24 colonies), eye (2 out of 25 colonies), and heart (4 out of 26 colonies) (**Figure 5-6**). These results may reflect that these variants are expressed at a low level in these tissues, and further quantitative experiments would be required to determine if this was the case.

The other annotated transcript isoform, *BPTF*, has been described in the literature as being testis-specific and important for spermatogenesis in humans (Jones et al., 2000). The results confirm this is true for mice as well, with *Bptf* being the major transcript (15 colonies out of 24 analysed). It seems that the presence of Region 2 is very important for testis function (only 4 out of 24 analysed colonies lacking it). This pattern is somewhat similar for eye tissue, although a greater variety of transcript variants were found (**Figure 5-6**). It is interesting to note that there are only two tissues where the dominant *Bptf* variants have almost equal levels of transcripts with and without the Region 2 alternatively spliced; brain (10/10) and heart (8/13).

Transcript variants without the second exon (variant 3 and variant 4) in Region 1 are heart-specific. This could signify that this exon codes some function that is not needed for its function in heart or might even be deleterious. *Bptf* variant 8, although present in minor quantities in other tissues, is specifically expressed in kidney and thymus tissues with 16 out of 23, and 11 out of 24 colonies, respectively (**Figure 5-6**). It is worth mention that the NCBI annotation for mouse *Bptf* (NM_176850) has variant 1 signature, i.e. both exons of the Region 1 are absent. In addition, variant 5 has now been removed because insufficient data was initially annotated (NM_001080832). The Ensembl genome browser (<http://www.ensembl.org/index.html>) has annotated three full length *Bptf* transcripts corresponding to variant 1 (ENSMUST00000057892), variant 5 (ENSMUST00000106762), and variant 7 (ENSMUST00000106763).

Although *Bptf* variants 5 and 6 (exon 5 of Region 1 absent) were detected in eye, kidney, and thymus tissues, they are rare (**Figure 5-6**). It is possible that they are more abundant in tissues that were not analysed or might be more common in diseased or stressed tissues, for example cancer-associated. A BLAST search for EST data available in the NCBI database for human *BPTF* lacking exon 5 of Region 1 resulted in hits from a T84 carcinoma cell line (AA146707) and embryonal carcinoma (BG532502). However, there was no similar EST evidence for mouse.

In summary, construction of tissue-specific libraries and single-clone analysis can provide information about tissue-specificity and comparative level of expression of possible alternative large transcripts not obtainable by standard methods. It has provided intriguing novel insights into potentially different functional roles for *Bptf* variants in different tissues. For example, BPTF isoforms might result in functionally different chromatin remodelling complexes with altered histone-binding specificity because of alternatively sliced exons. I suggest this method of analysis can be applied to other problems involving large alternative transcripts and enable discovery of new splice isoforms and characterisation of their tissue expression.

5.3.4 Analysis of FAC1/BPTF functional regions and their evolutionary conservation

The discovery of multiple tissue-specific *BPTF* transcript isoforms raises the question of the physiological relevance of this expanded repertoire of transcripts and the functional roles of the alternatively spliced regions. Further bioinformatics analysis was undertaken to decipher the significance of these regions. I aligned the NCBI-annotated human BPTF variant 1 (NP_872579.2) and variant 2 (NP_004450.3) protein sequences and mapped onto it available information about functional motifs (<http://www.uniprot.org/uniprot/Q12830>) and antibodies used in BPTF research (Bowser et al., 1995, Jordan-Sciutto et al., 1999) (**Figure 5-8**). **Figure 5-8** shows which functional regions are affected by alternative splicing - such as the PEST domain, KEAP-interaction site, and glutamine-rich domain - as well as the antibody specificity to proteins coded by alternative transcripts. Alignment of BPTF homologues from different species and their comparative analysis allows tracking of protein domains that have remained conserved during evolution and, thus, provides an additional window on the regions which perform important BPTF functions. This alignment was performed by Dr Alex Zelinsky and is shown in **Appendix 10**. Comparison with *Drosophila*

NURF301 is especially valuable as properties of this protein have been studied extensively. Assuming that common conserved regions implies participation in common biological pathways, we can theorize about mammalian BPTF mechanisms of action.

5.3.4.1 Alternative transcriptional start site

Human BPTF sequence analysis reveals two alternative transcription start sites (TSS), with TSS-2 situated 139 amino acids downstream from the TSS-1 (**Figure 5-8**). The downstream site has been detected experimentally by 5' RACE using RNA isolated from human fetal brain for FAC1 (Zhu and Bowser, 1996) by screening a human testis cDNA library (Jones et al., 2000) and by protein sequencing (Barak et al., 2003) for BPTF. NCBI has annotated the alternative start codon 139 amino acids upstream of the reported FAC1/BPTF start site. This region, although conserved among mammals and less so for other vertebrates, does not have high homology with other species (insect, worm, fish, sea squirt), except for approximately 40-50 amino acids upstream of the TSS-2 (**Appendix 10**). This suggests that the alternative upstream TSS, conserved in vertebrates, has been recently acquired as an additional means of regulating BPTF activity in mammals.

Upstream open reading frames (ORF) are known to provide an additional mechanism for control by downregulation or inhibition of translation of the main ORF (Stoneley and Willis, 2003, Geballe and Morris, 1994, Tang et al., 2008). In addition to the upstream ORF, the BPTF sequence between the first and second ATG of TSS-1 and TSS-2 is 417 bp long and GC-rich, a feature generally associated with translational regulation (Kozak, 2005). I speculate that failure to inhibit BPTF translation could be a reason for its involvement in cancer (Grinberg-Rashi et al., 2009, Buganim et al., 2010).

5.3.4.2 Functional domains common to all transcripts

There are several functional domains common to all eight mouse alternative BPTF transcripts. These include the DDT domain, which determines the DNA-binding properties of BPTF, the glutamine-asparagine acidic domain, two nuclear localisation signals, the first PHD domain with Rb-binding motif, the second PHD domain with H3K4me3-binding motif, and the bromodomain (**Figures 5-7 and 5-8**). The MAZ-interaction site is common to all mouse transcripts, but as it is close to the end of FAC1 it could have different properties in cleaved and non-cleaved proteins (**Figures 5-7 and**

5-8). All these domains have a high degree of conservation among species from invertebrates to mammals revealing early specialisation of BPTF as a DNA-binding protein (**Appendix 10**). FAC1 is described as a transcription factor with a specific DNA-binding site (Jordan-Sciutto et al., 1999). Our phylogenetic footprinting analysis and luciferase assay experiments confirms this (**Chapter 4**). However, BPTF recruitment to promoters might be mediated by other factors (Wysocka et al., 2006, Landry et al., 2008). Similarly, the *Drosophila* ortholog of BPTF, NURF301, has been reported to be recruited to the promoter by the GAGA transcription factor (Xiao et al., 2001), and more recent studies demonstrated NURF301 sequence-specific binding to the *fzo* promoter in testis (Kwon et al., 2009). Thus, there is a possibility that both direct and indirect binding of BPTF to gene promoters are possible and alternative transcription might regulate the binding capacity.

It is necessary to note that it is unclear if proteins derived from both transcripts carrying both exons of Region 1 (v2 and v7) are truncated as apparently happens in case of FAC1 (v2), for which the full-length protein has not been detected. If so, then the transcripts v2 and v7 would both produce the same protein, i.e. without the bromodomain. Similarly, if the presence of the partial PEST-domain in Region 1 (i.e. derived from only one of the exons) is sufficient to drive protein cleavage, then BPTF-v3/BPTF-v4 and BPTF-v5/BPTF-v6 might also exist only as short proteins, with the Region 2 configuration again being irrelevant.

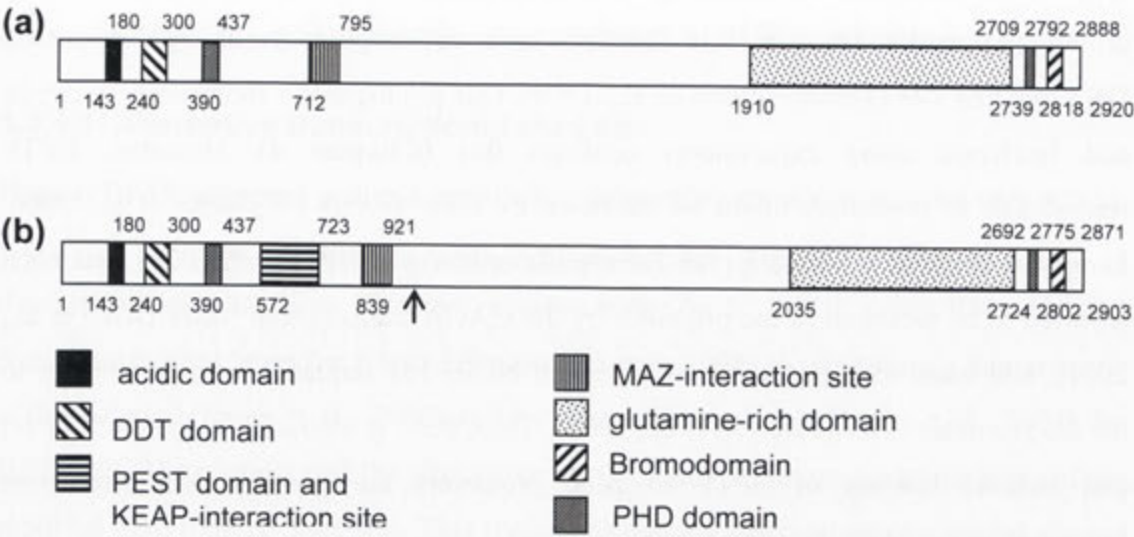


Figure 5-7. Functional domain of human BPTF variants. (a) BPTF with Region 1 absent. (b) BPTF with Region 2 absent. Arrow indicates possible FAC1 truncation site at approximate residue 1019 position. Numbers below specify the first residue of each domain; numbers above specify the last one.

5.3.4.3 Bromodomain

The bromodomain is the most conserved region of BPTF. It is required for spermatogenesis in *Drosophila* (Kwon et al., 2009). This is consistent with the reported highest level of human BPTF tissue expression being in testis (Jones et al., 2000) and my observations indicating that *Bptf* v1 is testis-specific. The high conservation of amino acids responsible for binding to histone H3K4me3 confirms that BPTF has maintained its function as a chromatin remodelling protein throughout evolution (Appendix 10). In effect, both *Drosophila* NURF301 and human BPTF can recognise H3K4me3 (Kwon et al., 2009, Wysocka et al., 2006). This region is not affected by alternative splicing, so potentially proteins coded by all transcript variants might be part of NURF complexes and participate in chromatin remodelling.

5.3.4.4 PEST domain

A PEST domain is defined as a hydrophilic region rich in proline (P), glutamic acid (E), serine (S), and threonine (T). Such regions are flanked by lysine (K), arginine (R) or histidine (H) residues. PEST sequences are proposed to act as signal peptides in

processes involving rapid degradation of proteins (Rogers et al., 1986, Rechsteiner and Rogers, 1996). Degradation may be mediated via the 26S proteasome pathway (Reverte et al., 2001, Spencer et al., 2004) or calpain proteolysis (Shumway et al., 1999, Sandoval et al., 2006, Marcilhac et al., 2006). PEST motifs can also contribute to various molecular interactions under diverse conditions due to their conformational flexibility (Singh et al., 2006, Sandhu and Dash, 2006). In the case of FAC1, the putative PEST domain carries the KEAP interaction site.

The human PEST region is 152 amino acids long and FAC1-specific as it extends across both exons of Region 1 (**Figures 5-7 and 5-8, amino acids 572-723**). In addition to proline, glutamic acid, serine, and threonine, it is enriched with aspartic acid (Zhu and Bowser, 1996). Additional analysis of FAC1 sequences from mouse, rat, dog, cow, human, and marsupial opossum was conducted for PEST sequences using the program PESTfind (<http://emboss.bioinformatics.nl/cgi-bin/emboss/pestfind>). PESTfind reported four PEST segments (**Figure 5-8, amino acids 296-310, 629-654, 673-724, and 2379-2433**) present across all species. This analysis confirmed that the PEST domain stretched across both exons of Region 1, although it is shorter than marked in human by Zhu & Bowser (1996) and is divided into two regions. The first of these (629-654) is situated in the first exon of Region 1 whereas the second (673-724) is positioned partly in the first exon but mostly in the second exon of Region 1. The PEST prediction for amino acids 2379-2433 may be an artefact because of its position within the glutamine-rich domain.

Similarly, the mouse BPTF gene encodes a PEST domain that is included in a region that undergoes alternative splicing. Alternative splicing of transcripts results in *Bptf* variants in which one or both Region 1 exons are removed. Clearly, BPTF proteins corresponding to transcript variants lacking Region 1 (**Figure 5-6, V1 and V8**) will not be subject to PEST-induced calpain cleavage. However, it is unclear whether proteins translated from transcripts with only one of the Region 1 exons (**Figure 5-6, V3-V6**) will have retained or modified PEST function. The PEST domain is not a discrete folded domain structure but rather a region with enhanced composition of Pro-Glu-Ser-Thr, and often also Asp, which is flexibly disordered. Disruption of PEST domains in reported splice variants of other genes appears not to affect susceptibility to calpain cleavage (Burgess and Reiner, 2002). Thus, the biological significance and level of

potential activity of PEST-segment variations for the *BPTF* alternative transcripts of Region 1 as well as the physiological conditions needed for proteolysis require further experimental study.

It is interesting to note that the PEST domain is apparently not present in *Drosophila* NURF301 (**Appendix 10**) even though it is regarded as the homologue of BPTF (Barak et al., 2003). The PEST sequence exons are a relatively recent addition likely acquired by insertion. It is known that PEST domains can be inserted into other folded domains, for example, the 35-residue unstructured PEST domain inserted in an SH2 domain of SOCS3 (Babon et al., 2006). The fact that the PEST domain in Region 1 is less conserved than preceding and following sequence may support this suggestion of recent insertion of PEST exons. Evolutionary conservation of the PEST region in vertebrates is low, but nonetheless the PESTfind program still identifies the PEST motifs in all vertebrates sampled - fish, chicken, frog *Xenopus tropicalis*, opossum *Monodelphis domestica*, cow, dog, mouse and human. However, the PEST domain is highly conserved in mammals, which seems to suggest development of specific functions later in BPTF evolution.

5.3.4.5 Possibility of FAC1 truncation due to PEST domain

The presence of the PEST domain in the FAC1 variant might explain its truncation. FAC1 was detected initially as a ~125 kDa protein translated from a large 10.5 kb mRNA (Bowser et al., 1995) (**Table 5-2**). Further analysis identified FAC1 as an 810-residue protein with a predicted molecular mass of 91.7 kDa (Zhu and Bowser, 1996) (**Table 5-2**). Unfortunately, that deduced protein sequence was incorrect because of sequencing errors (deletion of one adenosine nucleotide) in the 3' end of the cDNA which led to prediction of an incorrect C-terminal end for FAC1 (**Figure 5-8**) and use of plasmid constructs coding for this too-short protein in subsequent *in vitro* research. I discovered this error, which has not been corrected in the literature, while analysing available data on BPTF alternative transcripts. A FAC1 transcript that would correspond to a 125 kDa (~1125 residue) protein should be 3.4 kb of length but such a transcript has never been detected. More recently, a Western blot analysis suggested a ~100 kDa short BPTF isoform expressed exclusively in adult mouse brain, whereas a 311 kDa full-length BPTF isoforms was detected in testis, spleen, adult brain, and total embryo throughout development (Landry et al., 2008) (**Table 5-2**). The size of FAC1

transcript thus remains unclear. A full length *Bptf* transcript was detected in adult mouse brain by *in situ* hybridisation with a probe corresponding to a segment of the 3'-coding region and untranslated region (Barak et al., 2003) and in multiple human tissues with a probe specific to the region of the transcript coding for the DTT domain (Jones et al., 2000) (**Table 5-2**). Considering this critique of the literature and my own results as summarised in **Figure 5-6**, the most reasonable assumption at present is that FAC1 is translated as a full-length protein from full-length transcript and then cleaved. Although these results do not affect this assumption it should be noted that *Drosophila* NURF301 and *C. elegans* NURF-1A alternative transcripts that result in truncated proteins lacking the C-terminal PHD fingers and bromodomain have been reported (Andersen et al., 2006, Kwon et al., 2009).

Table 5-2. Summary of published literature on length of BPTF v1 and v2 (FAC1).

Reference	Detected BPTF sequence	Method
Bowser et al., 1995	1280 bp FAC1 cDNA with partial ORF, stop codon at position 1160.	Immunoscreening of expressed human fetal brain cDNA library with Alz50 antibody. Positive clone was isolated, subcloned and sequenced.
Bowser et al., 1995	10.5 kb mRNA in all analysed tissues.	Northern blot analysis of mRNA from human heart, brain, placenta, lung, liver, muscle, kidney, pancreas. Probed with partial (1280 bp) FAC1 cDNA.
Bowser et al., 1995	125 kDa FAC1 protein in fetal and adult brain.	Western blot with FA1 and FA2 antibodies (epitope is in Region 1, Figure 5-8)
Zhu and Bowser, 1996	1.4 kb FAC1 clone with TSS-2. Combined the novel sequence with that previously identified to deduce a 2433 bp ORF and 810 amino acid, 91.7 kDa FAC1. Mismatched protein sequence at the C-terminal end (Figure 5-8). Predicted acidic domain, PEST, nuclear localisation signal, and PHD zinc finger (Figure 5-8).	5'RACE using human fetal brain cDNA.

Jones et al., 2000	10.5 kb transcript expressed in all tissues examined with highest level in testis.	Northern blot on 16 normal human tissues with a 240 bp cDNA probe specific to the nucleotides of DTT domain.
Barak et al., 2003	Full-length transcript was detected in adult mouse brain.	<i>In situ</i> hybridisation using a 475 bp probe amplified from adult mouse brain (3'-coding region and untranslated region of BPTF).
Landry et al., 2008	Short ~100 kDa BPTF variant in adult mouse brain, and long 311 kDa BPTF variant in testis, spleen, adult brain, and total embryo E10.5, E12.5, and E16.5.	Western blot analysis of various adult and embryonic mouse tissues.

5.3.4.6 Calpain might cause FAC1 truncation

It is interesting to consider the possibility of a specific cleavage of FAC1 by calpain resulting in removal of the C-terminal region containing the bromodomain and the glutamine-rich region. Calpains are calcium-dependent cysteine proteases that modulate biological activities of proteins through limited proteolysis and contribute to posttranslational modification of cellular proteins (Carragher et al., 2001, Frame et al., 2002, Welm et al., 2002). Calpains are implicated in neuronal apoptosis and neurodegeneration (Chan and Mattson, 1999, Camins et al., 2006), Alzheimer’s disease (Raynaud and Marcilhac, 2006), brain injury (Huh et al., 2006), and neuronal differentiation (Grynspan et al., 1997), all processes in which FAC1 plays a role. Calpain cleaves Bid, Bax, p53, and AIF proteins into proapoptotic fragments which promote cytochrome C release and apoptosis (Gao and Dou, 2000, Chen et al., 2001, Sedarous et al., 2003, Polster et al., 2005). Nuclear FAC1 has been shown to repress transcription (Jordan-Sciutto et al., 1999). However, FAC1 relocation into the cytoplasm has been observed in development and disease (Schoonover et al., 1996, Mu et al., 1997). The altered subcellular localisation that promotes FAC1 proapoptotic activity (Strachan et al., 2005b) might be explained by cleavage of FAC1 by calpain.

5.3.4.7 Glutamine-rich domain

Another important region of FAC1/BPTF proteins is the glutamine-rich domain spanning amino acids 1910-2709 of BPTF v1 and 2035-2692 of BPTF v2. Our

bioinformatics analysis of eukaryotic BPTF sequences from fish to mammals demonstrates a high level of conservation of the glutamine-rich region, pointing to functional significance (**Appendix 10**). This domain is affected by alternative splicing over amino acids 2395-2538 in BPTF v1 (**Figures 5-7 and 5-8**).

Glutamine-rich domains have been associated with amyloid formation and neurodegenerative diseases (Michelitsch and Weissman, 2000), protein-protein interactions (Guo et al., 2007), and RNA polymerase II recruitment for transcriptional activation in a number of eukaryotic transcription factors (Courey and Tjian, 1988, Emili et al., 1994, Gerber et al., 1994). Activation can be effective even when the transcription factor is bound to the DNA at a long distance from the transcriptional start site (Courey et al., 1989). As shown in **Chapter 4**, the predicted FAC1/BPTF binding sites are situated in the proximity of the second *Prnp* exon, a long distance from the *Prnp* and, especially, *Prnd* transcription start sites. The potential of glutamine-rich domains to mediate a long-distance activation of gene expression should be kept in mind as a possible mechanism for how FAC1/BPTF could regulate both *Prnp* and *Prnd* genes.

Partial loss of a glutamine-rich domain has been reported to result in changes in functionality (Courey and Tjian, 1988, Davies et al., 2008, Mayr et al., 2005, Gerber et al., 1994). The spliced part of the domain in BPTF is situated 202 amino acids upstream of the PHD Zn-finger domain containing binding sites for histone H3-K4Me3 and 280 amino acids upstream of the bromodomain (**Figure 5-8**). The domain may assist the formation of the BPTF chromatin remodelling complex, and thus modification of it by splicing may change the protein complexes bound by BPTF and its activity in chromatin remodelling. However, as noted above, it is unclear if the presence of a full or partial Region 1 with the PEST domain always results in posttranslational truncation with removal of the glutamine-rich and bromodomain regions. In this case alternative splicing would be functionally irrelevant for the proteins from the v2-v7 transcripts as they would be short.

It is important to mention that glutamine-rich sequences have been found in transcriptional repressors also (Song et al., 2004). It is thus possible that partial deletion of the domain could change an activator to a repressor.

My transcript analysis shows that the majority (20 out of 24) of testis-specific clones contain the unspliced version of the glutamine-rich domain (**Figure 5-6**). Thus, this domain appears to be especially important for testis function as all other tissues either contain about equal amounts of Region 2 variants (**Figure 5-6**, brain, heart), lesser levels of variants with Region 2 present (**Figure 5-6**, eye), or variants with Region 2 absent (**Figure 5-6**, kidney, thymus).

5.4 Conclusions

The relatively recent discovery of BPTF involvement in chromatin remodelling as part of the remodelling complex - which has been most studied in the orthologous *Drosophila* nucleosome remodelling factor (NURF) complex - highlights the necessity for increased understanding of mammalian *BPTF* transcription and tissue-specificity. Despite a growing list of published studies addressing possible molecular mechanisms and targets of the BPTF complex, BPTF itself is poorly understood.

The fact that only a truncated-protein form of FAC1 has been described so far despite the presence of full length mRNA transcript adds an additional aspect to BPTF complexity. In this context, my discovery of multiple tissue-specific splice variants is especially important as it might provide an explanation for how mammalian BPTF acts as an activator in some circumstances and a repressor in others. I isolated eight alternative transcripts of mouse *Bptf* and showed that they had substantial tissue specificity. Brain-specific retention of the two exons containing the PEST domain of Region 1 at the 5' end of the *Bptf* ORF is of a particular interest as it might enable additional calpain-dependent regulation. I suggest that calpain can mediate FAC1 truncation through the PEST domain based on the capacity of PEST sequences to act as signals of protein degradation by calpain proteolysis.

Furthermore, the existence of alternative *Bptf* variants and their tissue specificity highlights problems with interpretation of previously reported experimental results. Great care needs to be taken when dealing with alternative transcripts, especially coding for proteins with complex functions. BPTF, with its known involvement in development, apoptosis, oxidative stress, transcriptional regulation, and chromatin remodelling, serves as a good illustration. Understanding which alternative transcripts are involved is essential, as chromatin remodelling complexes assembled with

alternative isoforms of BPTF may have altered histone-binding specificities or switch from an activator to a repressor of transcription.

My results make clear that use of antibodies or nucleic acid probes that do not discriminate multiple protein isoforms might lead to attribution of a particular function to an incorrect variant of the protein. Similarly, special attention is necessary when choosing tissues and cell lines because alternative transcripts are often tissue-specific. Although it seems obvious that the possibility of functionally relevant alternative transcripts should be recognized and investigated first, this was not done for BPTF. The growing literature is thus not securely anchored.

Bioinformatics analysis of FAC1/BPTF functional regions and evolutionary conserved motifs confirms that the oldest function of the *BPTF* gene is chromatin remodelling, and that the novel, brain-specific, functions which currently appear to be relevant only to truncated proteins are more recent. The analysis also illustrates which important functions of BPTF could be affected by alternative splicing: these have been identified as PEST-domain-associated function, in particular, possible regulation by calpain and interaction with KEAP1, and glutamine-rich-domain-associated functions.

Finally, I suggest that my approach to analyse previously undetected mouse *Bptf* transcript variants has general applicability to other problems involving detection of splice variants of large transcripts and their tissue distribution. Despite the accumulation of huge amounts of data relevant to mRNA transcripts many putative alternative splice forms are missed because EST sequences are biased towards the 5' and 3' ends of genes. In the case of large or rare transcripts, the coverage of the internal sequence may often be insufficient to correctly predict splice forms. The situation is even more unsatisfactory if we need to assess tissue distribution of alternative transcripts because of the uneven coverage of EST tissue libraries. The novel method I have designed is straightforward for identifying different transcripts and may be used to help to solve this problem.

hBPTFv1 hBPTFv2	TSS1 ↓ MRGRGRPPKQPAAPAAERCAPAPPPPPPTSGPIGGLSRHRGSSRGRWAAQAEVAP 60 MRGRGRPPKQPAAPAAERCAPAPPPPPPTSGPIGGLSRHRGSSRGRWAAQAEVAP 60 KTRLSSPRGSSSRKPPPPPPAPPSTSAFGRGGGGGGGTGGGGGGHGLARTTAARRA 120 KTRLSSPRGSSSRKPPPPPPAPPSTSAFGRGGGGGGGTGGGGGGHGLARTTAARRA 120 		
hBPTFv1 hBPTFv2	TSS2 ↓ FAC1-810aa start, AAA97522 VNKVYDDHESEEEEEEDMVSEEEEEEDGDAEETQDSEDEDEMEEDDDSDYPEEME 180 VNKVYDDHESEEEEEEDMVSEEEEEEDGDAEETQDSEDEDEMEEDDDSDYPEEME 180 DDDDASCYCTESSFRSHSTYSSTPCRRKPRVHRPRSPILEEKDIPPLEFKSSEDLMPV 240 DDDDASCYCTESSFRSHSTYSSTPCRRKPRVHRPRSPILEEKDIPPLEFKSSEDLMPV 240 	Glu/Asp-rich(acidic domain) (143-180, 38 aa) Nuclear localisation signal (Zhu,Bowser, 1996)	
hBPTFv1 hBPTFv2	SHIMVIAIYEVLNPGTVLRLSPRFDFCAALVSQEQTLAAEMHVLLKAVLREEDT 300 SHIMVIAIYEVLNPGTVLRLSPRFDFCAALVSQEQTLAAEMHVLLKAVLREEDT 300 SNTTFGPADLKDSVNSTLYFIDGMTWPEVLRVYCESDKYHHVL PYQEAEDYPYGPVENK 360 SNTTFGPADLKDSVNSTLYFIDGMTWPEVLRVYCESDKYHHVL PYQEAEDYPYGPVENK 360 	BDT domain (240-300, 61 aa) PEST-prediction, PESTfind (296-310,15 aa)	
hBPTFv1 hBPTFv2	IKVLQFLVDQFLTNIAREELMSEGVIOYDDHCRCVCHKLGDLCCETCSAVYHLECVKPP 420 IKVLQFLVDQFLTNIAREELMSEGVIOYDDHCRCVCHKLGDLCCETCSAVYHLECVKPP 420 LEEVPEDEMOCEVCAHKVPGVTDCAVEIQKNKPYIRHEPIGYDRSRKRYWFLNRLLIE 480 LEEVPEDEMOCEVCAHKVPGVTDCAVEIQKNKPYIRHEPIGYDRSRKRYWFLNRLLIE 480 	ZN FINGER PHD domain (390-437, 48 aa) Rb binding (L-X-C-X-E) (Zhu,Bowser, 1996)	
hBPTFv1 hBPTFv2	EDTENENKKIYYSTKVQLAELIDCLDKDYWEAELCKILEEMREEIHRHMDITEDLTNK 540 EDTENENKKIYYSTKVQLAELIDCLDKDYWEAELCKILEEMREEIHRHMDITEDLTNK 540 ARGSNKSFLLAAANEIILESIRAKKGDDINNVKSPEETEKDKNETENDSKDAKNREEFEDQ 600 ARGSNKSFLLAAANEIILESIRAKKGDDINNVKSPEETEKDKNETENDSKDAKNREEFEDQ 600 	PEST domain prediction (572-723, 152 aa) (Zhu,Bowser, 1996)	
hBPTFv1 hBPTFv2	SLEKSDDKTPDDPEQKSE-----621 SLEKSDDKTPDDPEQKSEETFEVGDGKNSVSNALGDGTTAAATGSGSPSGKSPAGC 660 SPSGRSPVGC	PEST-prediction, PESTfind (629-654, 26 aa) FA1,FA2 antibody epitop (650-695) (Bowser, 1995)	
hBPTFv1 hBPTFv2	LSPTDSSNNAEKVASELPQDVP EEPNKKICSS LSPTDSSNNAEKVASELPQDVP EEPNKKICSS ----- LSPTDSSNNAEKVASELPQDVP EEPNKKICSS	Interaction with KEAP1 (640-749) (Strachan et al., 2004)	720 Exon boundary

hBPTFv1	-----VGDFKSEKNGELSESPGAGKAGSGSTRIITRL 654
hBPTFv2	SSVPAKADDPENGSESHPTFVSTQEEVGV DFKSEKNGELSESPGAGKAGSGSTRIITRL 780

hBPTFv1	RNPDSKLSQKSOQVAAAHEANKLFKEGKEVLVNSQGEISRLSTKKEVIMKGNINNYF 714
hBPTFv2	RNPDSKLSQKSOQVAAAHEANKLFKEGKEVLVNSQGEISRLSTKKEVIMKGNINNN 840

hBPTFv1	KLGQEGKRVYHNOYSTNSFALNKHQHRHDHDKRRRLAHKFCLTPAGEFKWNGSVHGSKV 774
hBPTFv2	KLGQEGKRVYHNOYSTNSFALNKHQHRHDHDKRRRLAHKFCLTPAGEFKWNGSVHGSKV 900

hBPTFv1	LTISTLRITITOLENNIPSSFLHPNWASHRANWIKAVOMCSKPREFALALALECAVKPV 834
hBPTFv2	LTISTLRITITOLENNIPSSFLHPNWASHRANWIKAVOMCSKPREFALALALECAVKPV 960

hBPTFv1	LTISTLRITITOLENNIPSSFLHPNWASHRANWIKAVOMCSKPREFALALALECAVKPV
hBPTFv2	LTISTLRITITOLENNIPSSFLHPNWASHRANWIKAVOMCSKPREFALALALECAVKPV

hBPTFv1	VMLPIWRESLGHTRLHRMTSIEREEKEKVKKKKEKQEEETMQOATWVKYTFPVKHQVWK 894
hBPTFv2	VMLPIWRESLGHTRLHRMTSIEREEKEKVKKKKEKQEEETMQOATWVKYTFPVKHQVWK 1020

hBPTFv1	QKGEERYVTGYGWSWISKTHVYRVFVKPLPGNTNVNRYKSLGKTKNNNDENDESDDRKC 954
hBPTFv2	QKGEERYVTGYGWSWISKTHVYRVFVKPLPGNTNVNRYKSLGKTKNNNDENDESDDRKC 1080

hBPTFv1	SRSPKKIKIEPDSEKDEVKSDAAGADQNMDSKITEKKDDQVKLLDSDSDKPKCEE 1014
hBPTFv2	SRSPKKIKIEPDSEKDEVKSDAAGADQNMDSKITEKKDDQVKLLDSDSDKPKCEE 1140

hBPTFv1	PMEVDDDMKTESHVNCQESSQVDVNVVSEGFHLRTSYKKKTKSKLDGLLERRIKQFTLE 1074
hBPTFv2	PMEVDDDMKTESHVNCQESSQVDVNVVSEGFHLRTSYKKKTKSKLDGLLERRIKQFTLE 1200

hBPTFv1	EKQBLEKIKLEGGIKIGKSTSNSSKNLSSEPVITKAKEGQSDSMROFQSPNANDQPE 1134
hBPTFv2	EKQBLEKIKLEGGIKIGKSTSNSSKNLSSEPVITKAKEGQSDSMROFQSPNANDQPE 1260

hBPTFv1	DLIQGCSESDSSVLRMSPSHNTNKLYPKDRVLDDVSIIRSPETKCPKNSIENDIEEKVS 1194
hBPTFv2	DLIQGCSESDSSVLRMSPSHNTNKLYPKDRVLDDVSIIRSPETKCPKNSIENDIEEKVS 1320

hBPTFv1	DLASRQEPSSKTKGNDFEIDDSKLASADDIGTLICNKKKPLIOESDTIVSSSKSALH 1254
hBPTFv2	DLASRQEPSSKTKGNDFEIDDSKLASADDIGTLICNKKKPLIOESDTIVSSSKSALH 1380

hBPTFv1	SSVPKSTNDRDATPLSRAMD FEGLGCDSESNSTLENSSDVTVS IQDSEEDMIVONSNE 1314
hBPTFv2	SSVPKSTNDRDATPLSRAMD FEGLGCDSESNSTLENSSDVTVS IQDSEEDMIVONSNE 1440

hBPTFv1	ISEQFRTRQDVEVLEPLKCELVSGESTNCEDRLPVKGTANGKKPSQOKKLEERPYNK 1374
hBPTFv2	ISEQFRTRQDVEVLEPLKCELVSGESTNCEDRLPVKGTANGKKPSQOKKLEERPYNK 1500

hBPTFv1	CSDQIKLKNTTDKKNNRESEKKQGTSTFTQINGKDNKPKIYLKGECLKEISESRVVS 1434
hBPTFv2	CSDQIKLKNTTDKKNNRESEKKQGTSTFTQINGKDNKPKIYLKGECLKEISESRVVS 1560

PEST-prediction, PESTfind (673-724, 52 aa)

Nuclear localisation signal (Zhu, Bowser, 1996)
Interaction with MAZ (839-921) (Jordan-Sciutto et al., 2000)

Proposed FAC1 ending (Zhu, Bowser, 1996)

Bethyl lab (commercial) BPTF Ab (1350-1400)

hBPTFv1	NVEPKVNNINKIIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSSDAE	1494
hBPTFv2	NVEPKVNNINKIIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSSDAE	1620
hBPTFv1	GNYRDSLETLPSTKESDSTOTTTPSASCPSNSVNOVEDMEIETSEVKKVTSSPITSEEE	1554
hBPTFv2	GNYRDSLETLPSTKESDSTOTTTPSASCPSNSVNOVEDMEIETSEVKKVTSSPITSEEE	1680
hBPTFv1	SNLSNDFIDENGLPINKNENNGESKRKTWITEVTMTSTVATESKTIVIKVEKGDQTVV	1614
hBPTFv2	SNLSNDFIDENGLPINKNENNGESKRKTWITEVTMTSTVATESKTIVIKVEKGDQTVV	1740
hBPTFv1	SSTENCAKSTVTTTTTIVTKLSTPSTGGSDVIISVKEQSKTVTTTIVTDSLTGGLVT	1674
hBPTFv2	SSTENCAKSTVTTTTTIVTKLSTPSTGGSDVIISVKEQSKTVTTTIVTDSLTGGLVT	1800
hBPTFv1	SMTVSKEYSTRDKVLMKFSRPKPTRSGTALPSYRKFFVTKSSKKSIFVLPNDDLKKLARK	1734
hBPTFv2	SMTVSKEYSTRDKVLMKFSRPKPTRSGTALPSYRKFFVTKSSKKSIFVLPNDDLKKLARK	1860
hBPTFv1	GGIREVPYFNYNNAKALDIWPPSPRPTFGITWRYRLQTVKSLAGVSLMLRLMASLRWD	1794
hBPTFv2	GGIREVPYFNYNNAKALDIWPPSPRPTFGITWRYRLQTVKSLAGVSLMLRLMASLRWD	1920
hBPTFv1	DMAAKAPGGGTRTRETSETETITTEIIKRRDVGPGYIRSEYCIKIIICPIGVPTPKET	1854
hBPTFv2	DMAAKAPGGGTRTRETSETETITTEIIKRRDVGPGYIRSEYCIKIIICPIGVPTPKET	1980
hBPTFv1	PTPQRKGLRSSALRPKRPEPKQTGPVIIETWAEELWEIIRAFARVEKEKAQAVEQ	1914
hBPTFv2	PTPQRKGLRSSALRPKRPEPKQTGPVIIETWAEELWEIIRAFARVEKEKAQAVEQ	2040
hBPTFv1	QAKKRLEQQKPTVIATSTTSSTSTSTISPAQKVMVAPISGSVTTGTMVLTTKVGSPP	1974
hBPTFv2	QAKKRLEQQKPTVIATSTTSSTSTSTISPAQKVMVAPISGSVTTGTMVLTTKVGSPP	2100
hBPTFv1	IVTFQONKNFHOTFATWVKQGSNSGVVQVQVKVLGIIPSTGTSGOOTFTSFQPKTATVTI	2034
hBPTFv2	IVTFQONKNFHOTFATWVKQGSNSGVVQVQVKVLGIIPSTGTSGOOTFTSFQPKTATVTI	2160
hBPTFv1	IRPNTSGSGGTTNSOVIITGPQIRPGMTVIRTPLOOSTLIGKAIIRTPVMVOPGAPQQVMTI	2094
hBPTFv2	IRPNTSGSGGTTNSOVIITGPQIRPGMTVIRTPLOOSTLIGKAIIRTPVMVOPGAPQQVMTI	2220
hBPTFv1	QIIIRGQPVSTAVSAPNTVSTSTPGKSLTSATSTSNIOSSASQPPRPQGGQVVKLTMAQLTC	2154
hBPTFv2	QIIIRGQPVSTAVSAPNTVSTSTPGKSLTSATSTSNIOSSASQPPRPQGGQVVKLTMAQLTC	2280
hBPTFv1	LTQGHGNGOGLTVVIQGGGOTTGQLQIIPQGVTVLPGPGQOLMQAAMPNGTVQRFLFTPL	2214
hBPTFv2	LTQGHGNGOGLTVVIQGGGOTTGQLQIIPQGVTVLPGPGQOLMQAAMPNGTVQRFLFTPL	2340
hBPTFv1	ATTATTASTTTTTTIVSTTAAGTGEQRQSKLSQMQOVHQDKTLPQAQSSSVGPAAEQPQTAQ	2274
hBPTFv2	ATTATTASTTTTTTIVSTTAAGTGEQRQSKLSQMQOVHQDKTLPQAQSSSVGPAAEQPQTAQ	2400

Thr-rich domain (1709-11803, 95 aa)

Glutamine-rich domain (2035-2692) (Jones et al., 2000)

Thr-rich domain (2338-2361, 24 aa)

PEST-prediction, PESTfind (2379-2433, 55 aa)

Chapter 6.

Summary and Conclusion

The main goal of my PhD was to track the coevolution of the *PRNP* gene and its downstream duplicate *PRND* by investigation of their expression and transcripts in amphibian, marsupial and mouse models and by analysis of their coregulation using combined bioinformatics, comparative genomics and experimental techniques.

The importance of this research arises from the necessity to expand understanding of the fate of tandemly duplicated genes, as no suitable model was described before to examine gene evolution systematically from its “nascence” to acquisition of a new function. Evidence for coevolution of PrP and Dpl can also provide answers about how their antagonistic relationship in neurodegeneration might be regulated.

The first step towards developing the *PRNP/PRND* evolutionary model was a novel discovery of *PRND* in amphibians and marsupials and characterisation of *PRNP/PRND* tissue specificity, described in Chapter 3. A key aspect of the early evolution of the *PRND* gene is the existence of a chimeric transcript sharing 5' untranslated exons with the *PRNP* gene. This initial chimeric transcription allowed the *PRND* to evolve rapidly by mutation and escape pseudogenisation while its expression was preserved by coexpression from a parental gene promoter.

Results presented in this thesis demonstrate simultaneous evolution of *PRNP* and *PRND* by showing their expression divergence in three vertebrate models. Evolution of PrP was driven by the acquisition of the N-terminal repeats in the protein and the CPE motif at the 3'UTR by *PRNP*, which facilitates posttranslational regulation. Evolution of Dpl, on the other hand, was driven by acquisition of new regulatory DNA elements and increased tissue specificity in eutherian mammals. The emergence of monocistronic *PRND* in mammals together with a new tissue-expression pattern signifies the importance of an independent promoter for neofunctionalisation.

A significant discovery was persistence of chimeric *PRND* transcription in mouse. Moore et al. (1999) proposed that chimeric expression might be explained either as a result of inefficient transcriptional termination or as a means to synchronize synthesis of

the PrP and Dpl proteins. My analysis of translation of chimeric *PRND* transcripts by ribosome-binding assay reveals its association with polysomes during embryonic and early postnatal mouse stages and, thus, Dpl protein expression from the chimeric *PRND* transcript. This finding suggests the correctness of the idea of synchronized PrP/Dpl expression as well as demonstrates the functional importance of chimeric transcription of *PRND* during development.

Comparative sequence analysis suggests that the functions of PrP and Dpl in mammals differ from those in frogs. However, conservation of the main structural domains suggests changes in gene regulation rather than structural divergence was an important genetic basis for PrP and Dpl functional evolutionary development and that PrP may have retained its original function and evolved additional promiscuous functions rather than gaining a completely new function.

The discovery of a *PRNP* duplicate gene in the *X. laevis* allotetraploid genome provides an additional insight on gene fate after whole genome duplication. The intrinsic functional flexibility of PrP is demonstrated by retention of its second copy in the allotetraploid *X. laevis*. Conversely, loss of the second copy of *PRND* might be due to its functional immaturity at the time of the whole genome duplication.

Taken together, one of the most significant conclusions emerging from Chapter 3 is that PrP and Dpl coevolved within vertebrates and transcriptional and translational regulation of their expression plays an important role in their functional development.

The results presented in Chapter 4 address the issue of putative factors that might coregulate PrP and Dpl expression. It has been suggested that an antagonistic relationship between PrP and Dpl during neurodegeneration is based on competitive ligand binding. This study proposes a model for a regulatory mechanism behind this antagonism. Phylogenetic footprinting analysis predicted several transcription factor binding sites in the *PRNP* and *PRND* genes. Conceptual analysis of published data on the predicted transcription factors suggested the choice of FAC1/BPTF as a priority for further experimental analysis. This selection is based on absence of the FAC1 binding site in chicken, in which the *PRND* gene is absent, and on coexpression of FAC1, PrP and Dpl in Alzheimer's, oxidative stress and neuronal development. The results of *in vitro* experiments (luciferase assay, expression analysis) suggest a novel hypothesis that

FAC1/BPTF might coregulate expression of both *PRNP* and *PRND* genes by acting as a distant enhancer and engaging a BPTF-chromatin remodelling mechanism. A conceptual approach combined with bioinformatics analysis suggest that BPTF might be recruited to the *PRND* promoter by the USF1 transcription factor and activate expression through an E2F1/Rb mechanism. Alternatively, BPTF might be recruited to the *PRNP* promoter by the HSF transcription factor. This model provides an interesting potential mechanism for regulation of antagonistic antiapoptotic/proapoptotic functions of PrP and Dpl.

The research undertaken in Chapter 4 led to realisation that FAC1 and its isoform BPTF, two variants reported so far, are very poorly described in the literature. Thus, additional investigations, described in Chapter 5, were undertaken to provide a systematic characterisation of the structure and properties of BPTF using comparative genomics, expression, and transcript analysis. The results indicate the existence of eight *BPTF* variants in mouse tissues with specific tissue-expression patterns, and with alternative splicing affecting various functional protein domains. These findings imply that tissue-specific modulation of BPTF activity by alternative splicing should be taken into account when studying its role as a transcription factor or as a part of the chromatin remodelling complex. This information also can be used to develop BPTF-specific antibodies and *BPTF*-specific probes able to discriminate among isoforms.

The novel data presented in this thesis address the important issues of evolution by gene duplication, coregulation of antagonistic genes, and tissue-specific alternative transcription. However, this research has raised many questions that require further investigation.

First of all, translation of chimeric *PRND* in amphibians still has to be confirmed by polysome-binding assay. Development of novel methods of detection is needed to clarify the functional role of chimeric *PRND* in mammals.

Further experimental investigations are needed to deepen our understanding of PrP and Dpl coregulation. More detailed studies are required to confirm recruitment of FAC1 and E2F1 to their predicted binding sites. This can be achieved by mutagenesis of corresponding binding sites and ChIP analysis. Coimmunoprecipitation experiments may be applied to confirm recruitment of BPTF by USF1 in the *PRND* promoter.

Binding could be examined in cell or animal models at specific stages of development. Comparative analysis of healthy- and disease-state mouse models, or embryonic brain and adult brain, in which either *PRNP* or *PRND* is selectively activated, could provide more definitive evidence to the proposed mechanism of coregulation.

Finally, the existence of variable multiple *BPTF* variants in different tissue types requires analysis to determine whether all variants present are translated. This could be done by creating variant-specific antibodies followed by Western blot analysis and immunohistochemistry for detection of *BPTF* variants. .

In summary, this thesis demonstrated coevolution of PrP and Dpl and proposed a novel model for their coregulation. My research confirmed the concept that evolution of duplicated genes involves both sequence and expression divergence and presented a new way to look at the interaction between PrP and Dpl and regulation of their functions using a comparative evolutionary approach. The results also provided a novel basis for understanding how dysfunction of gene regulation of PrP-family genes might lead to outcomes in disease states such as prion disease or Alzheimer's neurodegeneration.

Finally, I would like to underline once more the advantages of applying a comparative approach using both computational and experimental analysis to successfully characterise the evolutionary relations between gene family members and define regulation of their functions.

References

- Aasland, R., Gibson, T. J. & Stewart, A. F. (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci*, 20, 56-9.
- Agholme, L., Lindstrom, T., Kagedal, K., Marcusson, J. & Hallbeck, M. (2010) An in vitro model for neuroscience: differentiation of SH-SY5Y cells into cells with morphological and biochemical characteristics of mature neurons. *J Alzheimers Dis*, 20, 1069-82.
- Aguzzi, A. & Polymenidou, M. (2004) Mammalian prion biology: one century of evolving concepts. *Cell*, 116, 313-27.
- Ahmed, F., Kumar, M. & Raghava, G. P. (2009) Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies. *In Silico Biol*, 9, 135-48.
- Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A. & Sorek, R. (2006) Transcription-mediated gene fusion in the human genome. *Genome Res*, 16, 30-6.
- Alfonso, A., Grundahl, K., McManus, J. R., Asbury, J. M. & Rand, J. B. (1994) Alternative splicing leads to two cholinergic proteins in *Caenorhabditis elegans*. *J Mol Biol*, 241, 627-30.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol*, 215, 403-10.
- Ankar, J. & Sistonen, L. (2011) Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu Rev Biochem*, 80, 1089-115.
- Andersen, E. C., Lu, X. & Horvitz, H. R. (2006) *C. elegans* ISWI and NURF301 antagonize an Rb-like pathway in the determination of multiple cell fates. *Development*, 133, 2695-704.
- Anderson, L., Rossi, D., Linehan, J., Brandner, S. & Weissmann, C. (2004) Transgene-driven expression of the Doppel protein in Purkinje cells causes Purkinje cell degeneration and motor impairment. *Proc Natl Acad Sci U S A*, 101, 3644-9.
- Anderson, L. & Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, 18, 533-7.
- Andrews, J., Smith, M., Merakovsky, J., Coulson, M., Hannan, F. & Kelly, L. E. (1996) The stoned locus of *Drosophila melanogaster* produces a dicistronic transcript and encodes two distinct polypeptides. *Genetics*, 143, 1699-711.
- Arava, Y. (2003) Isolation of polysomal RNA for microarray analysis. *Methods Mol Biol*, 224, 79-87.
- Asante, E. A., Gowland, I., Grimshaw, A., Linehan, J. M., Smidak, M., Houghton, R., Osiuguwa, O., Tomlinson, A., Joiner, S., Brandner, S., Wadsworth, J. D. & Collinge, J. (2009) Absence of spontaneous disease and comparative prion susceptibility of transgenic mice expressing mutant human prion proteins. *J Gen Virol*, 90, 546-58.
- Babon, J. J., McManus, E. J., Yao, S., DeSouza, D. P., Mielke, L. A., Sprigg, N. S., Willson, T. A., Hilton, D. J., Nicola, N. A., Baca, M., Nicholson, S. E. & Norton, R. S. (2006) The structure of SOCS3 reveals the basis of the extended SH2 domain function and identifies an unstructured insertion that regulates stability. *Mol Cell*, 22, 205-16.

- Bailey, J. A., Liu, G. & Eichler, E. E. (2003) An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet*, 73, 823-34.
- Bailleul, B., Akerblom, I. & Strosberg, A. D. (1997) The leptin receptor promoter controls expression of a second distinct protein. *Nucleic Acids Res*, 25, 2752-8.
- Bainbridge, J. & Walker, K. B. (2005) The normal cellular form of prion protein modulates T cell responses. *Immunol Lett*, 96, 147-50.
- Baker, D. & Sali, A. (2001) Protein structure prediction and structural genomics. *Science*, 294, 93-6.
- Ball, M., McLellan, A., Collins, B., Coadwell, J., Stewart, F. & Moore, T. (2004) An abundant placental transcript containing an IAP-LTR is allelic to mouse pregnancy-specific glycoprotein 23 (Psg23): cloning and genetic analysis. *Gene*, 325, 103-13.
- Barak, O., Lazzaro, M. A., Lane, W. S., Speicher, D. W., Picketts, D. J. & Shiekhattar, R. (2003) Isolation of human NURF: a regulator of Engrailed gene expression. *EMBO J*, 22, 6089-100.
- Barry, A. E., Klyubin, I., Mc Donald, J. M., Mably, A. J., Farrell, M. A., Scott, M., Walsh, D. M. & Rowan, M. J. (2011) Alzheimer's disease brain-derived amyloid-beta-mediated inhibition of LTP in vivo is prevented by immunotargeting cellular prion protein. *J Neurosci*, 31, 7259-63.
- Baumann, F., Tolnay, M., Brabeck, C., Pahnke, J., Klotz, U., Niemann, H. H., Heikenwalder, M., Rulicke, T., Burklee, A. & Aguzzi, A. (2007) Lethal recessive myelin toxicity of prion protein lacking its central domain. *EMBO J*, 26, 538-47.
- Baybutt, H. & Manson, J. (1997) Characterisation of two promoters for prion protein (PrP) gene expression in neuronal cells. *Gene*, 184, 125-31.
- Beaudoin, M. E., Poirel, V. J. & Krushel, L. A. (2008) Regulating amyloid precursor protein synthesis through an internal ribosomal entry site. *Nucleic Acids Res*, 36, 6835-47.
- Behrens, A. & Aguzzi, A. (2002) Small is not beautiful: antagonizing functions for the prion protein PrP^C and its homologue Dpl. *Trends Neurosci*, 25, 150-4.
- Behrens, A., Genoud, N., Naumann, H., Rulicke, T., Janett, F., Heppner, F. L., Ledermann, B. & Aguzzi, A. (2002) Absence of the prion protein homologue Doppel causes male sterility. *EMBO J*, 21, 3652-8.
- Bejerano, G., Lowe, C. B., Ahituv, N., King, B., Siepel, A., Salama, S. R., Rubin, E. M., Kent, W. J. & Haussler, D. (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, 441, 87-90.
- Bendheim, P. E., Brown, H. R., Rudelli, R. D., Scala, L. J., Goller, N. L., Wen, G. Y., Kascsak, R. J., Cashman, N. R. & Bolton, D. C. (1992) Nearly ubiquitous tissue distribution of the scrapie agent precursor protein. *Neurology*, 42, 149-56.
- Bengtsson, H., Epifantseva, I., Abrink, M., Kylberg, A., Kullander, K., Ebendal, T. & Usoskin, D. (2008) Generation and characterization of a Gdf1 conditional null allele. *Genesis*, 46, 368-72.
- Benvegnu, S., Franciotta, D., Sussman, J., Bachi, A., Zardini, E., Torreri, P., Govaerts, C., Pizzo, S. & Legname, G. (2009) Prion protein paralog doppel protein interacts with alpha-2-macroglobulin: a plausible mechanism for doppel-mediated neurodegeneration. *PLoS One*, 4, e5968.
- Benvegnu, S., Roncaglia, P., Agostini, F., Casalone, C., Corona, C., Gustincich, S. & Legname, G. (2011) Developmental influence of the cellular prion protein on the gene expression profile in mouse hippocampus. *Physiol Genomics*, 43, 711-25.
- Berr, C., Helbecque, N., Szadovitch, V., Mohr, M., Amant, C., Amouyel, P., Alperovitch, A. & Hauw, J. J. (2003) Polymorphism of the codon 129 of the

- prion protein (PrP) gene and neuropathology of cerebral ageing. *Acta Neuropathol (Berl)*, 106, 71-4.
- Bett, C., Fernandez-Borges, N., Kurt, T. D., Lucero, M., Nilsson, K. P., Castilla, J. & Sigurdson, C. J. (2012) Structure of the beta2-alpha2 loop and interspecies prion transmission. *Faseb J*, 26, 2868-76.
- Bieda, M., Xu, X., Singer, M. A., Green, R. & Farnham, P. J. (2006) Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res*, 16, 595-605.
- Billeter, M., Riek, R., Wider, G., Hornemann, S., Glockshuber, R. & Wuthrich, K. (1997) Prion protein NMR structure and species barrier for prion diseases. *Proc Natl Acad Sci U S A*, 94, 7281-5.
- Blackwood, E. M. & Kadonaga, J. T. (1998) Going the distance: a current view of enhancer action. *Science*, 281, 60-3.
- Blumenthal, T. (1998) Gene clusters and polycistronic transcription in eukaryotes. *Bioessays*, 20, 480-7.
- Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M. & Kim, S. K. (2002) A global analysis of *Caenorhabditis elegans* operons. *Nature*, 417, 851-4.
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc*, 4, 1-13.
- Bounhar, Y., Zhang, Y., Goodyer, C. G. & LeBlanc, A. (2001) Prion protein protects human neurons against Bax-mediated apoptosis. *J Biol Chem*, 276, 39145-9.
- Bowser, R., Giambrone, A. & Davies, P. (1995) FAC1, a novel gene identified with the monoclonal antibody Alz50, is developmentally regulated in human brain. *Dev Neurosci*, 17, 20-37.
- Bowser, R. & Reilly, S. (1998) Expression of FAC1 in activated microglia during Alzheimer's disease. *Neurosci Lett*, 253, 163-6.
- Brini, M., Miuzzo, M., Pierobon, N., Negro, A. & Sorgato, M. C. (2005) The prion protein and its paralogue Doppel affect calcium signaling in Chinese hamster ovary cells. *Mol Biol Cell*, 16, 2799-808.
- Brogna, S. & Ashburner, M. (1997) The Adh-related gene of *Drosophila melanogaster* is expressed as a functional dicistronic messenger RNA: multigenic transcription in higher organisms. *EMBO J*, 16, 2023-31.
- Brosius, J. (2003) The contribution of RNAs and retroposition to evolutionary novelties. *Genetica*, 118, 99-116.
- Brown, D. R. & Besinger, A. (1998) Prion protein expression and superoxide dismutase activity. *Biochem J*, 334 (Pt 2), 423-9.
- Brown, D. R., Nicholas, R. S. & Canevari, L. (2002) Lack of prion protein expression results in a neuronal phenotype sensitive to stress. *J Neurosci Res*, 67, 211-24.
- Brown, D. R., Qin, K., Herms, J. W., Madlung, A., Manson, J., Strome, R., Fraser, P. E., Kruck, T., von Bohlen, A., Schulz-Schaeffer, W., Giese, A., Westaway, D. & Kretzschmar, H. (1997) The cellular prion protein binds copper in vivo. *Nature*, 390, 684-7.
- Bueler, H., Aguzzi, A., Sailer, A., Greiner, R. A., Autenried, P., Aguet, M. & Weissmann, C. (1993) Mice devoid of PrP are resistant to scrapie. *Cell*, 73, 1339-47.
- Bueler, H., Fischer, M., Lang, Y., Bluethmann, H., Lipp, H. P., DeArmond, S. J., Prusiner, S. B., Aguet, M. & Weissmann, C. (1992) Normal development and

- behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature*, 356, 577-82.
- Buganim, Y., Goldstein, I., Lipson, D., Milyavsky, M., Polak-Charcon, S., Mardoukh, C., Solomon, H., Kalo, E., Madar, S., Brosh, R., Perelman, M., Navon, R., Goldfinger, N., Barshack, I., Yakhini, Z. & Rotter, V. (2010) A novel translocation breakpoint within the BPTF gene is associated with a pre-malignant phenotype. *PLoS One*, 5, e9657.
- Burgess, H. A. & Reiner, O. (2002) Alternative splice variants of doublecortin-like kinase are differentially expressed and have different kinase activities. *J Biol Chem*, 277, 17696-705.
- Burkhardt, D. L. & Sage, J. (2008) Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat Rev Cancer*, 8, 671-82.
- Cabral, A. L., Lee, K. S. & Martins, V. R. (2002) Regulation of the cellular prion protein gene expression depends on chromatin conformation. *J Biol Chem*, 277, 5675-82.
- Calzolari, L., Lysek, D. A., Perez, D. R., Guntert, P. & Wuthrich, K. (2005) Prion protein NMR structures of chickens, turtles, and frogs. *Proc Natl Acad Sci U S A*, 102, 651-5.
- Camins, A., Verdaguer, E., Folch, J. & Pallas, M. (2006) Involvement of calpain activation in neurodegenerative processes. *CNS Drug Rev*, 12, 135-48.
- Cao, A. R., Rabinovich, R., Xu, M., Xu, X., Jin, V. X. & Farnham, P. J. (2011) Genome-wide analysis of transcription factor E2F1 mutant proteins reveals that N- and C-terminal protein interaction domains do not participate in targeting E2F1 to the human genome. *J Biol Chem*, 286, 11985-96.
- Caputo, A., Sarnataro, D., Campana, V., Costanzo, M., Negro, A., Sorgato, M. C. & Zurzolo, C. (2010) Doppel and PrP^C co-immunoprecipitate in detergent-resistant membrane domains of epithelial FRT cells. *Biochem J*, 425, 341-51.
- Carragher, N. O., Fincham, V. J., Riley, D. & Frame, M. C. (2001) Cleavage of focal adhesion kinase by different proteases during SRC-regulated transformation and apoptosis. Distinct roles for calpain and caspases. *J Biol Chem*, 276, 4270-5.
- Chakka, N. & Gready, J. E. (2006) Pathway to Functional Studies: Pipeline Linking Phylogenetic Footprinting and Transcription-Factor Binding Analysis. *CRPIT*, 73, 15-21.
- Chan, S. L. & Mattson, M. P. (1999) Caspase and calpain substrates: roles in synaptic plasticity and cell death. *J Neurosci Res*, 58, 167-90.
- Charlesworth, A., Cox, L. L. & MacNicol, A. M. (2004) Cytoplasmic polyadenylation element (CPE)- and CPE-binding protein (CPEB)-independent mechanisms regulate early class maternal mRNA translational activation in *Xenopus* oocytes. *J Biol Chem*, 279, 17650-9.
- Chen, M., He, H., Zhan, S., Krajewski, S., Reed, J. C. & Gottlieb, R. A. (2001) Bid is cleaved by calpain to an active fragment in vitro and during myocardial ischemia/reperfusion. *J Biol Chem*, 276, 30724-8.
- Chen, S., Mange, A., Dong, L., Lehmann, S. & Schachner, M. (2003) Prion protein as trans-interacting partner for neurons is involved in neurite outgrowth and neuronal survival. *Mol Cell Neurosci*, 22, 227-33.
- Chiarini, L. B., Freitas, A. R., Zanata, S. M., Brentani, R. R., Martins, V. R. & Linden, R. (2002) Cellular prion protein transduces neuroprotective signals. *EMBO J*, 21, 3317-26.
- Chothia, C. & Lesk, A. M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J*, 5, 823-6.

- Chung, Y. D., Kwon, H. C., Chung, K. W., Kim, S. J., Kim, K. & Lee, C. C. (1996) Identification of ovarian enhancer-binding factors which bind to ovarian enhancer 1 of the *Drosophila* genes *yp1* and *yp2*. *Mol Gen Genet*, 251, 347-51.
- Cohen, F. E. & Prusiner, S. B. (1998) Pathologic conformations of prion proteins. *Annu Rev Biochem*, 67, 793-819.
- Coitinho, A. S., Freitas, A. R., Lopes, M. H., Hajj, G. N., Roesler, R., Walz, R., Rossato, J. I., Cammarota, M., Izquierdo, I., Martins, V. R. & Brentani, R. R. (2006) The interaction between prion protein and laminin modulates memory consolidation. *Eur J Neurosci*, 24, 3255-64.
- Coitinho, A. S., Lopes, M. H., Hajj, G. N., Rossato, J. I., Freitas, A. R., Castro, C. C., Cammarota, M., Brentani, R. R., Izquierdo, I. & Martins, V. R. (2007) Short-term memory formation and long-term memory consolidation are enhanced by cellular prion association to stress-inducible protein 1. *Neurobiol Dis*, 26, 282-90.
- Coitinho, A. S., Roesler, R., Martins, V. R., Brentani, R. R. & Izquierdo, I. (2003) Cellular prion protein ablation impairs behavior as a function of age. *Neuroreport*, 14, 1375-9.
- Comincini, S., Foti, M. G., Tranulis, M. A., Hills, D., Di Guardo, G., Vaccari, G., Williams, J. L., Harbitz, I. & Ferretti, L. (2001) Genomic organization, comparative analysis, and genetic polymorphisms of the bovine and ovine prion Doppel genes (PRND). *Mamm Genome*, 12, 729-33.
- Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. (1989) Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell*, 59, 827-36.
- Courey, A. J. & Tjian, R. (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell*, 55, 887-98.
- Crusselle-Davis, V. J., Vieira, K. F., Zhou, Z., Anantharaman, A. & Bungert, J. (2006) Antagonistic regulation of beta-globin gene expression by helix-loop-helix proteins USF and TFII-I. *Mol Cell Biol*, 26, 6832-43.
- Cui, T., Holme, A., Sassoon, J. & Brown, D. R. (2003) Analysis of doppel protein toxicity. *Mol Cell Neurosci*, 23, 144-55.
- Darnell, R. B. (2003) Memory, synaptic translation, and...prions? *Cell*, 115, 767-8.
- Dasgupta, P., Padmanabhan, J. & Chellappan, S. (2006) Rb function in the apoptosis and senescence of non-neuronal and neuronal cells: role in oncogenesis. *Curr Mol Med*, 6, 719-29.
- Davies, P., Watt, K., Kelly, S. M., Clark, C., Price, N. C. & McEwan, I. J. (2008) Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *J Mol Endocrinol*, 41, 301-14.
- De Val, S., Anderson, J. P., Heidt, A. B., Khiem, D., Xu, S. M. & Black, B. L. (2004) Mef2c is activated directly by Ets transcription factors through an evolutionarily conserved endothelial cell-specific enhancer. *Dev Biol*, 275, 424-34.
- del Prete, M. J., Vernal, R., Dolznig, H., Mullner, E. W. & Garcia-Sanz, J. A. (2007) Isolation of polysome-bound mRNA from solid tissues amenable for RT-PCR and profiling experiments. *Rna*, 13, 414-21.
- Deng, C., Cheng, C. H., Ye, H., He, X. & Chen, L. (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc Natl Acad Sci U S A*, 107, 21593-8.
- Dewji, N. N. & Do, C. (1996) Heat shock factor-1 mediates the transcriptional activation of Alzheimer's beta-amyloid precursor protein gene in response to stress. *Brain Res Mol Brain Res*, 35, 325-8.

- Didonna, A., Sussman, J., Benetti, F. & Legname, G. (2012) The role of Bax and caspase-3 in doppel-induced apoptosis of cerebellar granule cells. *Prion*, 6.
- Dodelet, V. C. & Cashman, N. R. (1998) Prion protein expression in human leukocyte differentiation. *Blood*, 91, 1556-61.
- Doetzlhofer, A., Rotheneder, H., Lagger, G., Koranda, M., Kurtev, V., Brosch, G., Wintersberger, E. & Seiser, C. (1999) Histone deacetylase 1 can repress transcription by binding to Sp1. *Mol Cell Biol*, 19, 5504-11.
- Dossena, S., Imeri, L., Mangieri, M., Garofoli, A., Ferrari, L., Senatore, A., Restelli, E., Balducci, C., Fiordaliso, F., Salio, M., Bianchi, S., Fioriti, L., Morbin, M., Pincherle, A., Marcon, G., Villani, F., Carli, M., Tagliavini, F., Forloni, G. & Chiesa, R. (2008) Mutant prion protein expression causes motor and memory deficits and abnormal sleep patterns in a transgenic mouse model. *Neuron*, 60, 598-609.
- Dunaief, J. L., Strober, B. E., Guha, S., Khavari, P. A., Alin, K., Luban, J., Begemann, M., Crabtree, G. R. & Goff, S. P. (1994) The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell*, 79, 119-30.
- Dyson, N. (1998) The regulation of E2F by pRB-family proteins. *Genes Dev*, 12, 2245-62.
- Ecroyd, H., Sarradin, P., Dacheux, J. L. & Gatti, J. L. (2004) Compartmentalization of prion isoforms within the reproductive tract of the ram. *Biol Reprod*, 71, 993-1001.
- Edenhofer, F., Rieger, R., Famulok, M., Wendler, W., Weiss, S. & Winnacker, E. L. (1996) Prion protein PrP^c interacts with molecular chaperones of the Hsp60 family. *J Virol*, 70, 4724-8.
- Eiden, L. E. (1998) The cholinergic gene locus. *J Neurochem*, 70, 2227-40.
- Eisenhaber, B., Bork, P. & Eisenhaber, F. (1999) Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol*, 292, 741-58.
- Elkon, R., Linhart, C., Sharan, R., Shamir, R. & Shiloh, Y. (2003) Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res*, 13, 773-80.
- Emili, A., Greenblatt, J. & Ingles, C. J. (1994) Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. *Mol Cell Biol*, 14, 1582-93.
- Espenes, A., Harbitz, I., Skogtvedt, S., Fuglestad, R., Berg, K. A., Dick, G., Krogenaes, A. & Tranulis, M. A. (2006) Dynamic expression of the prion-like protein Doppel in ovine testicular tissue. *Int J Androl*, 29, 400-8.
- Ferrer, I., Blanco, R., Carmona, M., Puig, B., Ribera, R., Rey, M. J. & Ribalta, T. (2001) Prion protein expression in senile plaques in Alzheimer's disease. *Acta Neuropathol (Berl)*, 101, 49-56.
- Ferrer, I., Freixas, M., Blanco, R., Carmona, M. & Puig, B. (2004) Selective PrP-like protein, doppel immunoreactivity in dystrophic neurites of senile plaques in Alzheimer's disease. *Neuropathol Appl Neurobiol*, 30, 329-37.
- Flavell, S. W., Cowan, C. W., Kim, T. K., Greer, P. L., Lin, Y., Paradis, S., Griffith, E. C., Hu, L. S., Chen, C. & Greenberg, M. E. (2006) Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science*, 311, 1008-12.
- Flechsigs, E., Hegyi, I., Leimeroth, R., Zuniga, A., Rossi, D., Cozzio, A., Schwarz, P., Rulicke, T., Gotz, J., Aguzzi, A. & Weissmann, C. (2003) Expression of truncated PrP targeted to Purkinje cells of PrP knockout mice causes Purkinje cell death and ataxia. *EMBO J*, 22, 3095-101.

- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. & Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151, 1531-45.
- Ford, M. J., Burton, L. J., Li, H., Graham, C. H., Frobert, Y., Grassi, J., Hall, S. M. & Morris, R. J. (2002a) A marked disparity between the expression of prion protein and its message by neurones of the CNS. *Neuroscience*, 111, 533-51.
- Ford, M. J., Burton, L. J., Morris, R. J. & Hall, S. M. (2002b) Selective expression of prion protein in peripheral tissues of the adult mouse. *Neuroscience*, 113, 177-92.
- Forloni, G., Angeretti, N., Chiesa, R., Monzani, E., Salmona, M., Bugiani, O. & Tagliavini, F. (1993) Neurotoxicity of a prion protein fragment. *Nature*, 362, 543-6.
- Fournier, J. G., Escaig-Haye, F., Billette de Villemeur, T., Robain, O., Lasmezas, C. I., Deslys, J. P., Dormont, D. & Brown, P. (1998) Distribution and submicroscopic immunogold localization of cellular prion protein (PrP^c) in extracerebral tissues. *Cell Tissue Res*, 292, 77-84.
- Frame, M. C., Fincham, V. J., Carragher, N. O. & Wyke, J. A. (2002) v-Src's hold over actin and cell adhesions. *Nat Rev Mol Cell Biol*, 3, 233-45.
- Frigg, R., Wenzel, A., Samardzija, M., Oesch, B., Wariwoda, H., Navarini, A. A., Seeliger, M. W., Tanimoto, N., Reme, C. & Grimm, C. (2006) The prion protein is neuroprotective against retinal degeneration *in vivo*. *Exp Eye Res*, 83, 1350-8.
- Fry, C. J., Pearson, A., Malinowski, E., Bartley, S. M., Greenblatt, J. & Farnham, P. J. (1999) Activation of the murine dihydrofolate reductase promoter by E2F1. A requirement for CBP recruitment. *J Biol Chem*, 274, 15883-91.
- Gao, G. & Dou, Q. P. (2000) N-terminal cleavage of bax by calpain generates a potent proapoptotic 18-kDa fragment that promotes bcl-2-independent cytochrome C release and apoptotic cell death. *J Cell Biochem*, 80, 53-72.
- Gauczynski, S., Peyrin, J. M., Haik, S., Leucht, C., Hundt, C., Rieger, R., Krasemann, S., Deslys, J. P., Dormont, D., Lasmezas, C. I. & Weiss, S. (2001) The 37-kDa/67-kDa laminin receptor acts as the cell-surface receptor for the cellular prion protein. *EMBO J*, 20, 5863-75.
- Geballe, A. P. & Morris, D. R. (1994) Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem Sci*, 19, 159-64.
- Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S. & Schaffner, W. (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science*, 263, 808-11.
- Giovanni, A., Keramaris, E., Morris, E. J., Hou, S. T., O'Hare, M., Dyson, N., Robertson, G. S., Slack, R. S. & Park, D. S. (2000) E2F1 mediates death of B-amyloid-treated cortical neurons in a manner independent of p53 and dependent on Bax and caspase 3. *J Biol Chem*, 275, 11553-60.
- Goller, T., Vauti, F., Ramasamy, S. & Arnold, H. H. (2008) Transcriptional regulator BPTF/FAC1 is essential for trophoblast differentiation during early mouse development. *Mol Cell Biol*, 28, 6819-27.
- Graber, T. E. & Holcik, M. (2007) Cap-independent regulation of gene expression in apoptosis. *Mol Biosyst*, 3, 825-34.
- Graner, E., Mercadante, A. F., Zanata, S. M., Forlenza, O. V., Cabral, A. L., Veiga, S. S., Juliano, M. A., Roesler, R., Walz, R., Minetti, A., Izquierdo, I., Martins, V. R. & Brentani, R. R. (2000) Cellular prion protein binds laminin and mediates neuritogenesis. *Brain Res Mol Brain Res*, 76, 85-92.

- Gray, T. A., Saitoh, S. & Nicholls, R. D. (1999) An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc Natl Acad Sci U S A*, 96, 5616-21.
- Grinberg-Rashi, H., Ofek, E., Perelman, M., Skarda, J., Yaron, P., Hajdich, M., Jacob-Hirsch, J., Amariglio, N., Krupsky, M., Simansky, D. A., Ram, Z., Pfeffer, R., Galanter, I., Steinberg, D. M., Ben-Dov, I., Rechavi, G. & Izraeli, S. (2009) The expression of three genes in primary non-small cell lung cancer is associated with metastatic spread to the brain. *Clin Cancer Res*, 15, 1755-61.
- Gruffat, H., Manet, E. & Sergeant, A. (2002) MEF2-mediated recruitment of class II HDAC at the EBV immediate early gene BZLF1 links latency and chromatin remodeling. *EMBO Rep*, 3, 141-6.
- Grynspan, F., Griffin, W. B., Mohan, P. S., Shea, T. B. & Nixon, R. A. (1997) Calpains and calpastatin in SH-SY5Y neuroblastoma cells during retinoic acid-induced differentiation and neurite outgrowth: comparison with the human brain calpain system. *J Neurosci Res*, 48, 181-91.
- Gu, Z., Nicolae, D., Lu, H. H. & Li, W. H. (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet*, 18, 609-13.
- Guo, L., Han, A., Bates, D. L., Cao, J. & Chen, L. (2007) Crystal structure of a conserved N-terminal domain of histone deacetylase 4 reveals functional insights into glutamine-rich domains. *Proc Natl Acad Sci U S A*, 104, 4297-302.
- Gupta, R. & Brunak, S. (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*, 310-22.
- Gupta, S., Zink, D., Korn, B., Vingron, M. & Haas, S. A. (2004) Strengths and weaknesses of EST-based prediction of tissue-specific alternative splicing. *BMC Genomics*, 5, 72.
- Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*, 19, 1720-30.
- Hajj, G. N., Lopes, M. H., Mercadante, A. F., Veiga, S. S., da Silveira, R. B., Santos, T. G., Ribeiro, K. C., Juliano, M. A., Jacchieri, S. G., Zanata, S. M. & Martins, V. R. (2007) Cellular prion protein interaction with vitronectin supports axonal growth and is compensated by integrins. *J Cell Sci*, 120, 1915-26.
- Harbour, J. W. & Dean, D. C. (2000) The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev*, 14, 2393-409.
- Hardy, J. & Selkoe, D. J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, 297, 353-6.
- He, X. & Zhang, J. (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics*, 169, 1157-64.
- Hegde, R. S., Mastrianni, J. A., Scott, M. R., DeFea, K. A., Tremblay, P., Torchia, M., DeArmond, S. J., Prusiner, S. B. & Lingappa, V. R. (1998) A transmembrane form of the prion protein in neurodegenerative disease. *Science*, 279, 827-34.
- Helin, K., Harlow, E. & Fattaey, A. (1993) Inhibition of E2F-1 transactivation by direct binding of the retinoblastoma protein. *Mol Cell Biol*, 13, 6501-8.
- Herms, J., Tings, T., Gall, S., Madlung, A., Giese, A., Siebert, H., Schurmann, P., Windl, O., Brose, N. & Kretschmar, H. (1999) Evidence of presynaptic location and function of the prion protein. *J Neurosci*, 19, 8866-75.
- Herms, J. W., Madlung, A., Brown, D. R. & Kretschmar, H. A. (1997) Increase of intracellular free Ca^{2+} in microglia activated by prion protein fragment. *Glia*, 21, 253-7.

- Herns, J. W., Tings, T., Dunker, S. & Kretschmar, H. A. (2001) Prion protein affects Ca^{2+} -activated K^+ currents in cerebellar purkinje cells. *Neurobiol Dis*, 8, 324-30.
- Higgins, D. G. (1994) CLUSTAL V: multiple alignment of DNA and protein sequences. *Methods Mol Biol*, 25, 307-18.
- Horiuchi, M., Priola, S. A., Chabry, J. & Caughey, B. (2000) Interactions between heterologous forms of prion protein: binding, inhibition of conversion, and species barriers. *Proc Natl Acad Sci U S A*, 97, 5836-41.
- Horiuchi, M., Yamazaki, N., Ikeda, T., Ishiguro, N. & Shinagawa, M. (1995) A cellular form of prion protein (PrP^{C}) exists in many non-neuronal tissues of sheep. *J Gen Virol*, 76 (Pt 10), 2583-7.
- Horonchik, L., Tzaban, S., Ben-Zaken, O., Yedidia, Y., Rouvinski, A., Papy-Garcia, D., Barritault, D., Vlodavsky, I. & Taraboulos, A. (2005) Heparan sulfate is a cellular receptor for purified infectious prions. *J Biol Chem*, 280, 17062-7.
- Hou, S. T., Callaghan, D., Fournier, M. C., Hill, I., Kang, L., Massie, B., Morley, P., Murray, C., Rasquinha, I., Slack, R. & MacManus, J. P. (2000) The transcription factor E2F1 modulates apoptosis of neurons. *J Neurochem*, 75, 91-100.
- Hughes, A. L. (1994) The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci*, 256, 119-24.
- Hughes, M. K. & Hughes, A. L. (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol Biol Evol*, 10, 1360-9.
- Huh, J. W., Franklin, M. A., Widing, A. G. & Raghupathi, R. (2006) Regionally distinct patterns of calpain activation and traumatic axonal injury following contusive brain injury in immature rats. *Dev Neurosci*, 28, 466-76.
- Hundt, C., Peyrin, J. M., Haik, S., Gauczynski, S., Leucht, C., Rieger, R., Riley, M. L., Deslys, J. P., Dormont, D., Lasmezas, C. I. & Weiss, S. (2001) Identification of interaction domains of the prion protein with its 37-kDa/67-kDa laminin receptor. *EMBO J*, 20, 5876-86.
- Inoue, S., Tanaka, M., Horiuchi, M., Ishiguro, N. & Shinagawa, M. (1997) Characterization of the bovine prion protein gene: the expression requires interaction between the promoter and intron. *J Vet Med Sci*, 59, 175-83.
- Jackson, G. S., Murray, I., Hosszu, L. L., Gibbs, N., Waltho, J. P., Clarke, A. R. & Collinge, J. (2001) Location and properties of metal-binding sites on the human prion protein. *Proc Natl Acad Sci U S A*, 98, 8531-5.
- Jiang, Y. Q., Wang, X. L., Cao, X. H., Ye, Z. Y., Li, L. & Cai, W. Q. (2013) Increased heat shock transcription factor 1 in the cerebellum reverses the deficiency of Purkinje cells in Alzheimer's disease. *Brain Res*.
- Jin, V. X., Rabinovich, A., Squazzo, S. L., Green, R. & Farnham, P. J. (2006) A computational genomics approach to identify cis-regulatory modules from chromatin immunoprecipitation microarray data--a case study using E2F1. *Genome Res*, 16, 1585-95.
- Johnson, D. G., Schwarz, J. K., Cress, W. D. & Nevins, J. R. (1993) Expression of transcription factor E2F1 induces quiescent cells to enter S phase. *Nature*, 365, 349-52.
- Jones, M. H., Hamana, N. & Shimane, M. (2000) Identification and characterization of BPTF, a novel bromodomain transcription factor. *Genomics*, 63, 35-9.
- Jordan-Sciutto, K. L., Dragich, J. M., Caltagarone, J., Hall, D. J. & Bowser, R. (2000) Fetal Alz-50 clone 1 (FAC1) protein interacts with the Myc-associated zinc finger protein (ZF87/MAZ) and alters its transcriptional activity. *Biochemistry*, 39, 3206-15.

- Jordan-Sciutto, K. L., Dragich, J. M., Rhodes, J. L. & Bowser, R. (1999) Fetal Alz-50 clone 1, a novel zinc finger protein, binds a specific DNA sequence and acts as a transcriptional regulator. *J Biol Chem*, 274, 35262-8.
- Jordan, I. K., Rogozin, I. B., Glazko, G. V. & Koonin, E. V. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet*, 19, 68-72.
- Kaessmann, H. (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Res*, 20, 1313-26.
- Kan, H. Y., Georgopoulos, S. & Zannis, V. (2000) A hormone response element in the human apolipoprotein CIII (ApoCIII) enhancer is essential for intestinal expression of the ApoA-I and ApoCIII genes and contributes to the hepatic expression of the two linked genes in transgenic mice. *J Biol Chem*, 275, 30423-31.
- Kanaani, J., Prusiner, S. B., Diacovo, J., Baekkeskov, S. & Legname, G. (2005) Recombinant prion protein induces rapid polarization and development of synapses in embryonic rat hippocampal neurons *in vitro*. *J Neurochem*, 95, 1373-86.
- Kaneko, K., Zulianello, L., Scott, M., Cooper, C. M., Wallace, A. C., James, T. L., Cohen, F. E. & Prusiner, S. B. (1997) Evidence for protein X binding to a discontinuous epitope on the cellular prion protein during scrapie prion propagation. *Proc Natl Acad Sci U S A*, 94, 10069-74.
- Kim, G., Meriin, A. B., Gabai, V. L., Christians, E., Benjamin, I., Wilson, A., Wolozin, B. & Sherman, M. Y. (2012) The heat shock transcription factor Hsf1 is downregulated in DNA damage-associated senescence, contributing to the maintenance of senescence phenotype. *Aging Cell*, 11, 617-27.
- Knight, R. S. & Will, R. G. (2004) Prion diseases. *J Neurol Neurosurg Psychiatry*, 75 Suppl 1, i36-42.
- Kornblihtt, A. R. (2006) Chromatin, transcript elongation and alternative splicing. *Nat Struct Mol Biol*, 13, 5-7.
- Kovacs, G. G., Zerbi, P., Voigtlander, T., Strohschneider, M., Trabattoni, G., Hainfellner, J. A. & Budka, H. (2002) The prion protein in human neurodegenerative disorders. *Neurosci Lett*, 329, 269-72.
- Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361, 13-37.
- Krebs, B., Wiebelitz, A., Balitzki-Korte, B., Vassallo, N., Paluch, S., Mitteregger, G., Onodera, T., Kretzschmar, H. A. & Herms, J. (2007) Cellular prion protein modulates the intracellular calcium response to hydrogen peroxide. *J Neurochem*, 100, 358-67.
- Krnjevic, K. (1970) Glutamate and gamma-aminobutyric acid in brain. *Nature*, 228, 119-24.
- Kubler, E., Oesch, B. & Raeber, A. J. (2003) Diagnosis of prion diseases. *Br Med Bull*, 66, 267-79.
- Kubosaki, A., Yusa, S., Nasu, Y., Nishimura, T., Nakamura, Y., Saeki, K., Matsumoto, Y., Itohara, S. & Onodera, T. (2001) Distribution of cellular isoform of prion protein in T lymphocytes and bone marrow, analyzed by wild-type and prion protein gene-deficient mice. *Biochem Biophys Res Commun*, 282, 103-7.
- Kumar, S. & Hedges, S. B. (1998) A molecular timescale for vertebrate evolution. *Nature*, 392, 917-20.
- Kurschner, C. & Morgan, J. I. (1995) The cellular prion protein (PrP) selectively binds to Bcl-2 in the yeast two-hybrid system. *Brain Res Mol Brain Res*, 30, 165-8.

- Kurschner, C. & Morgan, J. I. (1996) Analysis of interaction sites in homo- and heteromeric complexes containing Bcl-2 family members and the cellular prion protein. *Brain Res Mol Brain Res*, 37, 249-58.
- Kuwahara, C., Takeuchi, A. M., Nishimura, T., Haraguchi, K., Kubosaki, A., Matsumoto, Y., Saeki, K., Yokoyama, T., Itohara, S. & Onodera, T. (1999) Prions prevent neuronal cell-line death. *Nature*, 400, 225-6.
- Kuwata, K., Nishida, N., Matsumoto, T., Kamatari, Y. O., Hosokawa-Muto, J., Kodama, K., Nakamura, H. K., Kimura, K., Kawasaki, M., Takakura, Y., Shirabe, S., Takata, J., Kataoka, Y. & Katamine, S. (2007) Hot spots in prion protein for pathogenic conversion. *Proc Natl Acad Sci U S A*, 104, 11921-6.
- Kwon, S. Y., Xiao, H., Wu, C. & Badenhorst, P. (2009) Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities. *PLoS Genet*, 5, e1000574.
- Landry, J., Sharov, A. A., Piao, Y., Sharova, L. V., Xiao, H., Southon, E., Matta, J., Tessarollo, L., Zhang, Y. E., Ko, M. S., Kuehn, M. R., Yamaguchi, T. P. & Wu, C. (2008) Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genet*, 4, e1000241.
- Landry, J. W., Banerjee, S., Taylor, B., Aplan, P. D., Singer, A. & Wu, C. (2011) Chromatin remodeling complex NURF regulates thymocyte maturation. *Genes Dev*, 25, 275-86.
- Lasmezas, C. I. (2003) Putative functions of PrP^(C). *Br Med Bull*, 66, 61-70.
- Laukaitis, C. M., Dlouhy, S. R., Emes, R. D., Ponting, C. P. & Karn, R. C. (2005) Diverse spatial, temporal, and sexual expression of recently duplicated androgen-binding protein genes in *Mus musculus*. *BMC Evol Biol*, 5, 40.
- Lauren, J., Gimbel, D. A., Nygaard, H. B., Gilbert, J. W. & Strittmatter, S. M. (2009) Cellular prion protein mediates impairment of synaptic plasticity by amyloid-beta oligomers. *Nature*, 457, 1128-32.
- Lee, D. C., Sakudo, A., Kim, C. K., Nishimura, T., Saeki, K., Matsumoto, Y., Yokoyama, T., Chen, S. G., Itohara, S. & Onodera, T. (2006a) Fusion of Doppel to octapeptide repeat and N-terminal half of hydrophobic region of prion protein confers resistance to serum deprivation. *Microbiol Immunol*, 50, 203-9.
- Lee, I. Y., Westaway, D., Smit, A. F., Wang, K., Seto, J., Chen, L., Acharya, C., Ankener, M., Baskin, D., Cooper, C., Yao, H., Prusiner, S. B. & Hood, L. E. (1998) Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. *Genome Res*, 8, 1022-37.
- Lee, S. J. (1991) Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure. *Proc Natl Acad Sci U S A*, 88, 4250-4.
- Lee, Y., Ise, T., Ha, D., Saint Fleur, A., Hahn, Y., Liu, X. F., Nagata, S., Lee, B., Bera, T. K. & Pastan, I. (2006b) Evolution and expression of chimeric POTE-actin genes in the human genome. *Proc Natl Acad Sci U S A*, 103, 17885-90.
- Lejon, S., Thong, S. Y., Murthy, A., AlQarni, S., Murzina, N. V., Blobel, G. A., Laue, E. D. & Mackay, J. P. (2011) Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48.FOG-1 complex. *J Biol Chem*, 286, 1196-203.
- Levenson, J. M. & Sweatt, J. D. (2005) Epigenetic mechanisms in memory formation. *Nat Rev Neurosci*, 6, 108-18.
- Li, A., Christensen, H. M., Stewart, L. R., Roth, K. A., Chiesa, R. & Harris, D. A. (2007) Neonatal lethality in transgenic mice expressing prion protein with a deletion of residues 105-125. *EMBO J*, 26, 548-58.

- Li, A., Sakaguchi, S., Atarashi, R., Roy, B. C., Nakaoke, R., Arima, K., Okimura, N., Kopacek, J. & Shigematsu, K. (2000a) Identification of a novel gene encoding a PrP-like protein expressed as chimeric transcripts fused to PrP exon 1/2 in ataxic mouse line with a disrupted PrP gene. *Cell Mol Neurobiol*, 20, 553-67.
- Li, A., Sakaguchi, S., Shigematsu, K., Atarashi, R., Roy, B. C., Nakaoke, R., Arima, K., Okimura, N., Kopacek, J. & Katamine, S. (2000b) Physiological expression of the gene for PrP-like protein, PrPLP/Dpl, by brain endothelial cells and its ectopic expression in neurons of PrP-deficient mice ataxic due to Purkinje cell degeneration. *Am J Pathol*, 157, 1447-52.
- Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D. & Patel, D. J. (2006a) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, 442, 91-5.
- Li, P. W., Li, J., Timmerman, S. L., Krushel, L. A. & Martin, S. L. (2006b) The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition. *Nucleic Acids Res*, 34, 853-64.
- Li, R., Liu, D., Zanusso, G., Liu, T., Fayen, J. D., Huang, J. H., Petersen, R. B., Gambetti, P. & Sy, M. S. (2001) The expression and potential function of cellular prion protein in human lymphocytes. *Cell Immunol*, 207, 49-58.
- Li, X., Wang, S., Li, Y., Deng, C., Steiner, L. A., Xiao, H., Wu, C., Bungert, J., Gallagher, P. G., Felsenfeld, G., Qiu, Y. & Huang, S. (2011) Chromatin boundaries require functional collaboration between the hSET1 and NURF complexes. *Blood*, 118, 1386-94.
- Lin, M. C., Mirzabekov, T. & Kagan, B. L. (1997a) Channel formation by a neurotoxic prion protein fragment. *J Biol Chem*, 272, 44-7.
- Lin, Q., Schwarz, J., Bucana, C. & Olson, E. N. (1997b) Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science*, 276, 1404-7.
- Linden, R., Martins, V. R., Prado, M. A., Cammarota, M., Izquierdo, I. & Brentani, R. (2008) Physiology of the prion protein. *Physiol Rev*, 88, 673-728.
- Liu, Y., Huang, T., MacMorris, M. & Blumenthal, T. (2001) Interplay between AAUAAA and the trans-splice site in processing of a *Caenorhabditis elegans* operon pre-mRNA. *Rna*, 7, 176-81.
- Lobao-Soares, B., Walz, R., Prediger, R. D., Freitas, R. L., Calvo, F., Bianchin, M. M., Leite, J. P., Landemberger, M. C. & Coimbra, N. C. (2008) Cellular prion protein modulates defensive attention and innate fear-induced behaviour evoked in transgenic mice submitted to an agonistic encounter with the tropical coral snake *Oxyrhopus guibei*. *Behav Brain Res*, 194, 129-37.
- Long, M. & Langley, C. H. (1993) Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science*, 260, 91-5.
- Lopes, M. H., Hajj, G. N., Muras, A. G., Mancini, G. L., Castro, R. M., Ribeiro, K. C., Brentani, R. R., Linden, R. & Martins, V. R. (2005) Interaction of cellular prion and stress-inducible protein 1 promotes neuritogenesis and neuroprotection by distinct signaling pathways. *J Neurosci*, 25, 11330-9.
- Luhers, T., Riek, R., Guntert, P. & Wuthrich, K. (2003) NMR structure of the human doppel protein. *J Mol Biol*, 326, 1549-57.
- Lynch, M. & Conery, J. S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, 290, 1151-5.
- Lynch, V. J. (2007) Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol*, 7, 2.

- Lyons, G. E., Micales, B. K., Schwarz, J., Martin, J. F. & Olson, E. N. (1995) Expression of *mef2* genes in the mouse central nervous system suggests a role in neuronal maturation. *J Neurosci*, 15, 5727-38.
- Lysek, D. A. & Wuthrich, K. (2004) Prion protein interaction with the C-terminal SH3 domain of Grb2 studied using NMR and optical spectroscopy. *Biochemistry*, 43, 10393-9.
- Mabbott, N. A., Brown, K. L., Manson, J. & Bruce, M. E. (1997) T-lymphocyte activation and the cellular form of the prion protein. *Immunology*, 92, 161-5.
- Macleod, K. F. (2008) The role of the RB tumour suppressor pathway in oxidative stress responses in the haematopoietic system. *Nat Rev Cancer*, 8, 769-81.
- Magnaghi-Jaulin, L., Groisman, R., Naguibneva, I., Robin, P., Lorain, S., Le Villain, J. P., Troulen, F., Trouche, D. & Harel-Bellan, A. (1998) Retinoblastoma protein represses transcription by recruiting a histone deacetylase. *Nature*, 391, 601-5.
- Mahal, S. P., Asante, E. A., Antoniou, M. & Collinge, J. (2001) Isolation and functional characterisation of the promoter region of the human prion protein gene. *Gene*, 268, 105-14.
- Makrinou, E., Collinge, J. & Antoniou, M. (2002) Genomic characterization of the human prion protein (PrP) gene locus. *Mamm Genome*, 13, 696-703.
- Manson, J., West, J. D., Thomson, V., McBride, P., Kaufman, M. H. & Hope, J. (1992) The prion protein gene: a role in mouse embryogenesis? *Development*, 115, 117-22.
- Marciano, P. G., Brettschneider, J., Manduchi, E., Davis, J. E., Eastman, S., Raghupathi, R., Saatman, K. E., Speed, T. P., Stoeckert, C. J., Jr., Eberwine, J. H. & McIntosh, T. K. (2004) Neuron-specific mRNA complexity responses during hippocampal apoptosis after traumatic brain injury. *J Neurosci*, 24, 2866-76.
- Marcilhac, A., Raynaud, F., Clerc, I. & Benyamin, Y. (2006) Detection and localization of calpain 3-like protease in a neuronal cell line: possible regulation of apoptotic cell death through degradation of nuclear IkappaBalpha. *Int J Biochem Cell Biol*, 38, 2128-40.
- Marques, A. C., Dupanloup, I., Vinckenbosch, N., Reymond, A. & Kaessmann, H. (2005) Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol*, 3, e357.
- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, 29, 291-325.
- Martin, J. F., Schwarz, J. J. & Olson, E. N. (1993) Myocyte enhancer factor (MEF) 2C: a tissue-restricted member of the MEF-2 family of transcription factors. *Proc Natl Acad Sci U S A*, 90, 5282-6.
- Martins, V. R., Graner, E., Garcia-Abreu, J., de Souza, S. J., Mercadante, A. F., Veiga, S. S., Zanata, S. M., Neto, V. M. & Brentani, R. R. (1997) Complementary hydrophathy identifies a cellular prion protein receptor. *Nat Med*, 3, 1376-82.
- Massimino, M. L., Ballarin, C., Bertoli, A., Casonato, S., Genovesi, S., Negro, A. & Sorgato, M. C. (2004) Human Doppel and prion protein share common membrane microdomains and internalization pathways. *Int J Biochem Cell Biol*, 36, 2016-31.
- Mayr, B. M., Guzman, E. & Montminy, M. (2005) Glutamine rich and basic region/leucine zipper (bZIP) domains stabilize cAMP-response element-binding protein (CREB) binding to chromatin. *J Biol Chem*, 280, 15103-10.

- Mazumder, B., Seshadri, V. & Fox, P. L. (2003) Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci*, 28, 91-8.
- McLennan, N. F., Brennan, P. M., McNeill, A., Davies, I., Fotheringham, A., Rennison, K. A., Ritchie, D., Brannan, F., Head, M. W., Ironside, J. W., Williams, A. & Bell, J. E. (2004) Prion protein accumulation and neuroprotection in hypoxic brain damage. *Am J Pathol*, 165, 227-35.
- McLennan, N. F., Rennison, K. A., Bell, J. E. & Ironside, J. W. (2001) In situ hybridization analysis of PrP mRNA in human CNS tissues. *Neuropathol Appl Neurobiol*, 27, 373-83.
- McMillan, J. P. & Singer, M. F. (1993) Translation of the human LINE-1 element, L1Hs. *Proc Natl Acad Sci U S A*, 90, 11533-7.
- Meggio, F., Negro, A., Sarno, S., Ruzzene, M., Bertoli, A., Sorgato, M. C. & Pinna, L. A. (2000) Bovine prion protein as a modulator of protein kinase CK2. *Biochem J*, 352 Pt 1, 191-6.
- Melamed, D. & Arava, Y. (2007) Genome-wide analysis of mRNA polysomal profiles with spotted DNA microarrays. *Methods Enzymol*, 431, 177-201.
- Michelitsch, M. D. & Weissman, J. S. (2000) A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A*, 97, 11910-5.
- Miele, G., Alejo Blanco, A. R., Baybutt, H., Horvat, S., Manson, J. & Clinton, M. (2003) Embryonic activation and developmental expression of the murine prion protein gene. *Gene Expr*, 11, 1-12.
- Mighell, A. J., Smith, N. R., Robinson, P. A. & Markham, A. F. (2000) Vertebrate pseudogenes. *FEBS Lett*, 468, 109-14.
- Mikulits, W., Pradet-Balade, B., Habermann, B., Beug, H., Garcia-Sanz, J. A. & Mullner, E. W. (2000) Isolation of translationally controlled mRNAs by differential screening. *Faseb J*, 14, 1641-52.
- Millhauser, G. L. (2007) Copper and the prion protein: methods, structures, function, and disease. *Annu Rev Phys Chem*, 58, 299-320.
- Miranda, A., Pericuesta, E., Ramirez, M. A. & Gutierrez-Adan, A. (2011) Prion protein expression regulates embryonic stem cell pluripotency and differentiation. *PLoS One*, 6, e18422.
- Mo, H., Moore, R. C., Cohen, F. E., Westaway, D., Prusiner, S. B., Wright, P. E. & Dyson, H. J. (2001) Two different neurodegenerative diseases caused by proteins with similar structures. *Proc Natl Acad Sci U S A*, 98, 2352-7.
- Modrek, B. & Lee, C. J. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet*, 34, 177-80.
- Moore, R. C., Lee, I. Y., Silverman, G. L., Harrison, P. M., Strome, R., Heinrich, C., Karunaratne, A., Pasternak, S. H., Chishti, M. A., Liang, Y., Mastrangelo, P., Wang, K., Smit, A. F., Katamine, S., Carlson, G. A., Cohen, F. E., Prusiner, S. B., Melton, D. W., Tremblay, P., Hood, L. E. & Westaway, D. (1999) Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel. *J Mol Biol*, 292, 797-817.
- Moore, R. C., Mastrangelo, P., Bouzamondo, E., Heinrich, C., Legname, G., Prusiner, S. B., Hood, L., Westaway, D., DeArmond, S. J. & Tremblay, P. (2001) Doppel-induced cerebellar degeneration in transgenic mice. *Proc Natl Acad Sci U S A*, 98, 15288-93.
- Moorman, A. F. & Christoffels, V. M. (2003) Cardiac chamber formation: development, genes, and evolution. *Physiol Rev*, 83, 1223-67.

- Moudjou, M., Frobert, Y., Grassi, J. & La Bonnardiére, C. (2001) Cellular prion protein status in sheep: tissue-specific biochemical signatures. *J Gen Virol*, 82, 2017-24.
- Mu, X., Springer, J. E. & Bowser, R. (1997) FAC1 expression and localization in motor neurons of developing, adult, and amyotrophic lateral sclerosis spinal cord. *Exp Neurol*, 146, 17-24.
- Mufti, A. R., Burstein, E., Csomos, R. A., Graf, P. C., Wilkinson, J. C., Dick, R. D., Challa, M., Son, J. K., Bratton, S. B., Su, G. L., Brewer, G. J., Jakob, U. & Duckett, C. S. (2006) XIAP is a copper binding protein deregulated in Wilson's disease and other copper toxicosis disorders. *Mol Cell*, 21, 775-85.
- Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V. & Gage, F. H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, 435, 903-10.
- Nagyova, J., Pastorek, J. & Kopacek, J. (2004) Identification of the critical cis-acting elements in the promoter of the mouse *Prnd* gene coding for Doppel protein. *Biochim Biophys Acta*, 1679, 288-93.
- Needleman, S. B. & Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48, 443-53.
- Nei, M., Rogozin, I. B. & Piontkivska, H. (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci U S A*, 97, 10866-71.
- Nei, M. & Rooney, A. P. (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*, 39, 121-52.
- Nicolas, E., Morales, V., Magnaghi-Jaulin, L., Harel-Bellan, A., Richard-Foy, H. & Trouche, D. (2000) RbAp48 belongs to the histone deacetylase complex that associates with the retinoblastoma protein. *J Biol Chem*, 275, 9797-804.
- Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*, 10, 1-6.
- Nigumann, P., Redik, K., Matlik, K. & Speck, M. (2002) Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics*, 79, 628-34.
- Ning, Z. Y., Zhao, D. M., Yang, J. M., Cui, Y. L., Meng, L. P., Wu, C. D. & Liu, H. X. (2005) Quantification of prion gene expression in brain and peripheral organs of golden hamster by real-time RT-PCR. *Anim Biotechnol*, 16, 55-65.
- Nishida, N., Katamine, S., Shigematsu, K., Nakatani, A., Sakamoto, N., Hasegawa, S., Nakaoke, R., Atarashi, R., Kataoka, Y. & Miyamoto, T. (1997) Prion protein is necessary for latent learning and long-term memory retention. *Cell Mol Neurobiol*, 17, 537-45.
- Nishida, N., Tremblay, P., Sugimoto, T., Shigematsu, K., Shirabe, S., Petromilli, C., Erpel, S. P., Nakaoke, R., Atarashi, R., Houtani, T., Torchia, M., Sakaguchi, S., DeArmond, S. J., Prusiner, S. B. & Katamine, S. (1999) A mouse prion protein transgene rescues mice deficient for the prion protein gene from Purkinje cell degeneration and demyelination. *Lab Invest*, 79, 689-97.
- Nozawa, M., Aotsuka, T. & Tamura, K. (2005) A novel chimeric gene, siren, with retroposed promoter sequence in the *Drosophila bipectinata* complex. *Genetics*, 171, 1719-27.
- Nunziante, M., Gilch, S. & Schatzl, H. M. (2003) Essential role of the prion protein N terminus in subcellular trafficking and half-life of cellular prion protein. *J Biol Chem*, 278, 3726-34.

- O'Hare, M. J., Hou, S. T., Morris, E. J., Cregan, S. P., Xu, Q., Slack, R. S. & Park, D. S. (2000) Induction and modulation of cerebellar granule neuron death by E2F-1. *J Biol Chem*, 275, 25358-64.
- Oesch, B., Westaway, D., Walchli, M., McKinley, M. P., Kent, S. B., Aebersold, R., Barry, R. A., Tempst, P., Teplow, D. B., Hood, L. E. & et al. (1985) A cellular gene encodes scrapie PrP 27-30 protein. *Cell*, 40, 735-46.
- Ohno, S. (1970) *Evolution by gene duplication*, New York, Springer-Verlag.
- Ohta, T. (2003) Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. *Genetica*, 118, 209-16.
- Page, T. J., Sikder, D., Yang, L., Pluta, L., Wolfinger, R. D., Kodadek, T. & Thomas, R. S. (2006) Genome-wide analysis of human HSF1 signaling reveals a transcriptional program linked to cellular adaptation and survival. *Mol Biosyst*, 2, 627-39.
- Paisley, D., Banks, S., Selfridge, J., McLennan, N. F., Ritchie, A. M., McEwan, C., Irvine, D. S., Saunders, P. T., Manson, J. C. & Melton, D. W. (2004) Male infertility and DNA damage in Doppel knockout and prion protein/Doppel double-knockout mice. *Am J Pathol*, 164, 2279-88.
- Pan, T., Wong, B. S., Liu, T., Li, R., Petersen, R. B. & Sy, M. S. (2002) Cell-surface prion protein interacts with glycosaminoglycans. *Biochem J*, 368, 81-90.
- Park, K. W. & Li, L. (2011) Prion protein in *Caenorhabditis elegans*: Distinct models of anti-BAX and neuropathology. *Prion*, 5, 28-38.
- Parkin, E. T., Watt, N. T., Hussain, I., Eckman, E. A., Eckman, C. B., Manson, J. C., Baybutt, H. N., Turner, A. J. & Hooper, N. M. (2007) Cellular prion protein regulates beta-secretase cleavage of the Alzheimer's amyloid precursor protein. *Proc Natl Acad Sci U S A*, 104, 11062-7.
- Parkyn, C. J., Vermeulen, E. G., Mootoosamy, R. C., Sunyach, C., Jacobsen, C., Oxvig, C., Moestrup, S., Liu, Q., Bu, G., Jen, A. & Morris, R. J. (2008) LRP1 controls biosynthetic and endocytic trafficking of neuronal prion protein. *J Cell Sci*, 121, 773-83.
- Parra, G., Reymond, A., Dabbouseh, N., Dermitzakis, E. T., Castelo, R., Thomson, T. M., Antonarakis, S. E. & Guigo, R. (2006) Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res*, 16, 37-44.
- Pauly, P. C. & Harris, D. A. (1998) Copper stimulates endocytosis of the prion protein. *J Biol Chem*, 273, 33107-10.
- Peoc'h, K., Guerin, C., Brandel, J. P., Launay, J. M. & Laplanche, J. L. (2000a) First report of polymorphisms in the prion-like protein gene (*PRND*): implications for human prion diseases. *Neurosci Lett*, 286, 144-8.
- Peoc'h, K., Manivet, P., Beaudry, P., Attane, F., Besson, G., Hannequin, D., Delasnerie-Laupretre, N. & Laplanche, J. L. (2000b) Identification of three novel mutations (E196K, V203I, E211Q) in the prion protein gene (*PRNP*) in inherited prion diseases with Creutzfeldt-Jakob disease phenotype. *Hum Mutat*, 15, 482.
- Peoc'h, K., Serres, C., Frobert, Y., Martin, C., Lehmann, S., Chasseigneaux, S., Sazdovitch, V., Grassi, J., Jouannet, P., Launay, J. M. & Laplanche, J. L. (2002) The human "prion-like" protein Doppel is expressed in both Sertoli cells and spermatozoa. *J Biol Chem*, 277, 43071-8.
- Peoc'h, K., Volland, H., De Gassart, A., Beaudry, P., Sazdovitch, V., Sorgato, M. C., Creminon, C., Laplanche, J. L. & Lehmann, S. (2003) Prion-like protein Doppel expression is not modified in scrapie-infected cells and in the brains of patients with Creutzfeldt-Jakob disease. *FEBS Lett*, 536, 61-5.

- Perrier, V., Kaneko, K., Safar, J., Vergara, J., Tremblay, P., DeArmond, S. J., Cohen, F. E., Prusiner, S. B. & Wallace, A. C. (2002) Dominant-negative inhibition of prion replication in transgenic mice. *Proc Natl Acad Sci U S A*, 99, 13079-84.
- Pick, L. & Heffer, A. (2012) Hox gene evolution: multiple mechanisms contributing to evolutionary novelties. *Ann N Y Acad Sci*, 1256, 15-32.
- Piontkivska, H., Rooney, A. P. & Nei, M. (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol*, 19, 689-97.
- Plant, K. E., Routledge, S. J. & Proudfoot, N. J. (2001) Intergenic transcription in the human beta-globin gene cluster. *Mol Cell Biol*, 21, 6507-14.
- Polster, B. M., Basanez, G., Ettxebarria, A., Hardwick, J. M. & Nicholls, D. G. (2005) Calpain I induces cleavage and release of apoptosis-inducing factor from isolated mitochondria. *J Biol Chem*, 280, 6447-54.
- Pradet-Balade, B., Boulme, F., Beug, H., Mullner, E. W. & Garcia-Sanz, J. A. (2001) Translation control: bridging the gap between genomics and proteomics? *Trends Biochem Sci*, 26, 225-9.
- Premzl, M., Delbridge, M., Gready, J. E., Wilson, P., Johnson, M., Davis, J., Kuczek, E. & Marshall Graves, J. A. (2005) The prion protein gene: identifying regulatory signals using marsupial sequence. *Gene*, 349, 121-34.
- Premzl, M. & Gamulin, V. (2007) Comparative genomic analysis of prion genes. *BMC Genomics*, 8, 1.
- Premzl, M., Gready, J. E., Jermini, L. S., Simonic, T. & Marshall Graves, J. A. (2004) Evolution of vertebrate genes related to prion and Shadoo proteins--clues from comparative genomic analysis. *Mol Biol Evol*, 21, 2210-31.
- Premzl, M., Sangiorgio, L., Strumbo, B., Marshall Graves, J. A., Simonic, T. & Gready, J. E. (2003) Shadoo, a new protein highly conserved from fish to mammals and with similarity to prion protein. *Gene*, 314, 89-102.
- Prusiner, S. B. (1982) Novel proteinaceous infectious particles cause scrapie. *Science*, 216, 136-44.
- Prusiner, S. B. (1993) Genetic and infectious prion diseases. *Arch Neurol*, 50, 1129-53.
- Prusiner, S. B. (1998) Prions. *Proc Natl Acad Sci U S A*, 95, 13363-83.
- Prusiner, S. B. & Scott, M. R. (1997) Genetics of prions. *Annu Rev Genet*, 31, 139-75.
- Pyronnet, S., Pradayrol, L. & Sonenberg, N. (2000) A cell cycle-dependent internal ribosome entry site. *Mol Cell*, 5, 607-16.
- Qian, Y. W., Wang, Y. C., Hollingsworth, R. E., Jr., Jones, D., Ling, N. & Lee, E. Y. (1993) A retinoblastoma-binding protein related to a negative regulator of Ras in yeast. *Nature*, 364, 648-52.
- Qin, K., O'Donnell, M. & Zhao, R. Y. (2006a) Doppel: More rival than double to prion. *Neuroscience*.
- Qin, K., Zhao, L., Tang, Y., Bhatta, S., Simard, J. M. & Zhao, R. Y. (2006b) Doppel-induced apoptosis and counteraction by cellular prion protein in neuroblastoma and astrocytes. *Neuroscience*, 141, 1375-88.
- Rabinovich, A., Jin, V. X., Rabinovich, R., Xu, X. & Farnham, P. J. (2008) E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res*, 18, 1763-77.
- Radovanovic, I., Braun, N., Giger, O. T., Mertz, K., Miele, G., Prinz, M., Navarro, B. & Aguzzi, A. (2005) Truncated prion protein and Doppel are myelinotoxic in the absence of oligodendrocytic PrP^C. *J Neurosci*, 25, 4879-88.
- Rankin, C. T., Bunton, T., Lawler, A. M. & Lee, S. J. (2000) Regulation of left-right patterning in mice by growth/differentiation factor-1. *Nat Genet*, 24, 262-5.

- Rastogi, S. & Liberles, D. A. (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol*, 5, 28.
- Raynaud, F. & Marcilhac, A. (2006) Implication of calpain in neuronal apoptosis. A possible regulation of Alzheimer's disease. *Febs J*, 273, 3437-43.
- Rechsteiner, M. & Rogers, S. W. (1996) PEST sequences and regulation by proteolysis. *Trends Biochem Sci*, 21, 267-71.
- Reverte, C. G., Ahearn, M. D. & Hake, L. E. (2001) CPEB degradation during *Xenopus* oocyte maturation requires a PEST domain and the 26S proteasome. *Dev Biol*, 231, 447-58.
- Rhodes, J., Lutka, F. A., Jordan-Sciutto, K. L. & Bowser, R. (2003) Altered expression and distribution of FAC1 during NGF-induced neurite outgrowth of PC12 cells. *Neuroreport*, 14, 449-52.
- Rieger, R., Edenhofer, F., Lasmezas, C. I. & Weiss, S. (1997) The human 37-kDa laminin receptor precursor interacts with the prion protein in eukaryotic cells. *Nat Med*, 3, 1383-8.
- Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. & Wuthrich, K. (1996) NMR structure of the mouse prion protein domain PrP(121-321). *Nature*, 382, 180-2.
- Riek, R. & Luhrs, T. (2003) Three-dimensional structures of the prion protein and its doppel. *Clin Lab Med*, 23, 209-25.
- Riley, A., Jordan, L. E. & Holcik, M. (2010) Distinct 5' UTRs regulate XIAP expression under normal growth conditions and during cellular stress. *Nucleic Acids Res*, 38, 4665-74.
- Rogers, S., Wells, R. & Rechsteiner, M. (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*, 234, 364-8.
- Rondena, M., Cecilian, F., Comazzi, S., Pocacqua, V., Bazzocchi, C., Luvoni, C., Chigioni, S. & Paltrinieri, S. (2005) Identification of bovine doppel protein in testis, ovary and ejaculated spermatozoa. *Theriogenology*, 63, 1195-206.
- Rooney, A. P. & Ward, T. J. (2005) Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. *Proc Natl Acad Sci U S A*, 102, 5084-9.
- Roucous, X., Guo, Q., Zhang, Y., Goodyer, C. G. & LeBlanc, A. C. (2003) Cytosolic prion protein is not toxic and protects against Bax-mediated cell death in human primary neurons. *J Biol Chem*, 278, 40877-81.
- Rozen, S. & Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-86.
- Saeki, K., Matsumoto, Y. & Onodera, T. (1996) Identification of a promoter region in the rat prion protein gene. *Biochem Biophys Res Commun*, 219, 47-52.
- Sakaguchi, S. (2008) Antagonistic roles of the N-terminal domain of prion protein to doppel. *Prion*, 2, 107-11.
- Sakai, H., Koyanagi, K. O., Imanishi, T., Itoh, T. & Gojobori, T. (2007) Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes. *Gene*, 389, 196-203.
- Sakharkar, M. K., Chow, V. T. & Kanguane, P. (2004) Distributions of exons and introns in the human genome. *In Silico Biol*, 4, 387-93.
- Sakudo, A., Lee, D. C., Nakamura, I., Taniuchi, Y., Saeki, K., Matsumoto, Y., Itoharu, S., Ikuta, K. & Onodera, T. (2005) Cell-autonomous PrP-Doppel interaction regulates apoptosis in PrP gene-deficient neuronal cells. *Biochem Biophys Res Commun*, 333, 448-54.

- Sakudo, A. & Onodera, T. (2011) Tissue- and cell type-specific modification of prion protein (PrP)-like protein Doppel, which affects PrP endoproteolysis. *Biochem Biophys Res Commun*, 404, 523-7.
- Sakuma-Takagi, M., Tohyama, Y., Kasama-Yoshida, H., Sakagami, H., Kondo, H. & Kurihara, T. (1999) Novel related cDNAs (C184L, C184M, and C184S) from developing mouse brain encoding two apparently unrelated proteins. *Biochem Biophys Res Commun*, 263, 737-42.
- Sales, N., Hassig, R., Rodolfo, K., Di Giamberardino, L., Traiffort, E., Ruat, M., Fretier, P. & Moya, K. L. (2002) Developmental expression of the cellular prion protein in elongating axons. *Eur J Neurosci*, 15, 1163-77.
- Samonte, R. V. & Eichler, E. E. (2002) Segmental duplications and the evolution of the primate genome. *Nat Rev Genet*, 3, 65-72.
- Sanchez-Alavez, M., Conti, B., Moroncini, G. & Criado, J. R. (2007) Contributions of neuronal prion protein on sleep recovery and stress response following sleep deprivation. *Brain Res*, 1158, 71-80.
- Sandhu, K. S. & Dash, D. (2006) Conformational flexibility may explain multiple cellular roles of PEST motifs. *Proteins*, 63, 727-32.
- Sandoval, A., Oviedo, N., Tadmouri, A., Avila, T., De Waard, M. & Felix, R. (2006) Two PEST-like motifs regulate Ca²⁺/calpain-mediated cleavage of the CaVbeta3 subunit and provide important determinants for neuronal Ca²⁺ channel activity. *Eur J Neurosci*, 23, 2311-20.
- Santangelo, A. M., de Souza, F. S., Franchini, L. F., Bumashny, V. F., Low, M. J. & Rubinstein, M. (2007) Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. *PLoS Genet*, 3, 1813-26.
- Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J. & Kouzarides, T. (2002) Active genes are trimethylated at K4 of histone H3. *Nature*, 419, 407-11.
- Santuccione, A., Sytnyk, V., Leshchyn'ska, I. & Schachner, M. (2005) Prion protein recruits its neuronal receptor NCAM to lipid rafts to activate p59fyn and to enhance neurite outgrowth. *J Cell Biol*, 169, 341-54.
- Sassone-Corsi, P. (2002) Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science*, 296, 2176-8.
- Satoh, J., Onoue, H., Arima, K. & Yamamura, T. (2005) The 14-3-3 protein forms a molecular complex with heat shock protein Hsp60 and cellular prion protein. *J Neuropathol Exp Neurol*, 64, 858-68.
- Sayah, D. M., Sokolskaja, E., Berthoux, L. & Luban, J. (2004) Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature*, 430, 569-73.
- Schmitt-Ulms, G., Legname, G., Baldwin, M. A., Ball, H. L., Bradon, N., Bosque, P. J., Crossin, K. L., Edelman, G. M., DeArmond, S. J., Cohen, F. E. & Prusiner, S. B. (2001) Binding of neural cell adhesion molecules (N-CAMs) to the cellular prion protein. *J Mol Biol*, 314, 1209-25.
- Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C. & Kouzarides, T. (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol*, 6, 73-7.
- Schoonover, S., Davies, P. & Bowser, R. (1996) Immunolocalization and redistribution of the FAC1 protein in Alzheimer's disease. *J Neuropathol Exp Neurol*, 55, 444-55.

- Schwarze-Eicker, K., Keyvani, K., Gortz, N., Westaway, D., Sachser, N. & Paulus, W. (2005) Prion protein (PrP^C) promotes beta-amyloid plaque formation. *Neurobiol Aging*, 26, 1177-82.
- Scott, M. R., Safar, J., Telling, G., Nguyen, O., Groth, D., Torchia, M., Koehler, R., Tremblay, P., Walther, D., Cohen, F. E., DeArmond, S. J. & Prusiner, S. B. (1997) Identification of a prion protein epitope modulating transmission of bovine spongiform encephalopathy prions to transgenic mice. *Proc Natl Acad Sci U S A*, 94, 14279-84.
- Sedarous, M., Keramaris, E., O'Hare, M., Melloni, E., Slack, R. S., Elce, J. S., Greer, P. A. & Park, D. S. (2003) Calpains mediate p53 activation and neuronal death evoked by DNA damage. *J Biol Chem*, 278, 26031-8.
- Sepelakova, J., Takacova, M., Pastorekova, S. & Kopacek, J. (2005) Involvement of upstream stimulatory factor in regulation of the mouse Prnd gene coding for Doppel protein. *Biochim Biophys Acta*, 1731, 209-14.
- Shaked, Y., Hijazi, N. & Gabizon, R. (2002) Doppel and PrP^C do not share the same membrane microenvironment. *FEBS Lett*, 530, 85-8.
- Shmerling, D., Hegyi, I., Fischer, M., Blattler, T., Brandner, S., Gotz, J., Rulicke, T., Flechsig, E., Cozzio, A., von Mering, C., Hangartner, C., Aguzzi, A. & Weissmann, C. (1998) Expression of amino-terminally truncated PrP in the mouse leading to ataxia and specific cerebellar lesions. *Cell*, 93, 203-14.
- Shoja, V. & Zhang, L. (2006) A roadmap of tandemly arrayed genes in the genomes of human, mouse, and rat. *Mol Biol Evol*, 23, 2134-41.
- Shumway, S. D., Maki, M. & Miyamoto, S. (1999) The PEST domain of IkappaBalpha is necessary and sufficient for in vitro degradation by mu-calpain. *J Biol Chem*, 274, 30874-81.
- Shyu, W. C., Harn, H. J., Saeki, K., Kubosaki, A., Matsumoto, Y., Onodera, T., Chen, C. J., Hsu, Y. D. & Chiang, Y. H. (2002) Molecular modulation of expression of prion protein by heat shock. *Mol Neurobiol*, 26, 1-12.
- Sigurdson, C. J., Joshi-Barr, S., Bett, C., Winson, O., Manco, G., Schwarz, P., Rulicke, T., Nilsson, K. P., Margalith, I., Raeber, A., Peretz, D., Hornemann, S., Wuthrich, K. & Aguzzi, A. (2011) Spongiform encephalopathy in transgenic mice expressing a point mutation in the {beta}2-{alpha}2 loop of the prion protein. *J Neurosci*, 31, 13840-13847.
- Sigurdson, C. J., Nilsson, K. P., Hornemann, S., Heikenwalder, M., Manco, G., Schwarz, P., Ott, D., Rulicke, T., Liberski, P. P., Julius, C., Falsig, J., Stitz, L., Wuthrich, K. & Aguzzi, A. (2009) De novo generation of a transmissible spongiform encephalopathy by mouse transgenesis. *Proc Natl Acad Sci U S A*, 106, 304-9.
- Silverman, G. L., Qin, K., Moore, R. C., Yang, Y., Mastrangelo, P., Tremblay, P., Prusiner, S. B., Cohen, F. E. & Westaway, D. (2000) Doppel is an N-glycosylated, glycosylphosphatidylinositol-anchored protein. Expression in testis and ectopic production in the brains of Prnp(0/0) mice predisposed to Purkinje cell loss. *J Biol Chem*, 275, 26834-41.
- Singh, G. P., Ganapathi, M., Sandhu, K. S. & Dash, D. (2006) Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins*, 62, 309-15.
- Skrabanek, L. & Wolfe, K. H. (1998) Eukaryote genome duplication - where's the evidence? *Curr Opin Genet Dev*, 8, 694-700.
- Smith, C. W., Klaasmeyer, J. G., Edeal, J. B., Woods, T. L. & Jones, S. J. (1999) Effects of serum deprivation, insulin and dexamethasone on polysome

- percentages in C2C12 myoblasts and differentiating myoblasts. *Tissue Cell*, 31, 451-8.
- Sogayar, M. C., Camargo, A. A., Bettoni, F., Carraro, D. M., Pires, L. C., Parmigiani, R. B., Ferreira, E. N., de Sa Moreira, E., do Rosario, D. d. O. L. M., Simpson, A. J., Cruz, L. O., Degaki, T. L., Festa, F., Massirer, K. B., Sogayar, M. C., Filho, F. C., Camargo, L. P., Cunha, M. A., De Souza, S. J., Faria, M., Jr., Giuliani, S., Kopp, L., de Oliveira, P. S., Paiva, P. B., Pereira, A. A., Pinheiro, D. G., Puga, R. D., JE, S. d. S., Albuquerque, D. M., Andrade, L. E., Baia, G. S., Briones, M. R., Cavaleiro-Luna, A. M., Cerutti, J. M., Costa, F. F., Costanzi-Strauss, E., Espreafico, E. M., Ferrasi, A. C., Ferro, E. S., Fortes, M. A., Furchi, J. R., Giannella-Neto, D., Goldman, G. H., Goldman, M. H., Gruber, A., Guimaraes, G. S., Hackel, C., Henrique-Silva, F., Kimura, E. T., Leoni, S. G., Macedo, C., Malnic, B., Manzini, B. C., Marie, S. K., Martinez-Rossi, N. M., Menossi, M., Miracca, E. C., Nagai, M. A., Nobrega, F. G., Nobrega, M. P., Oba-Shinjo, S. M., Oliveira, M. K., Orabona, G. M., Otsuka, A. Y., Paco-Larson, M. L., Paixao, B. M., Pandolfi, J. R., Pardini, M. I., Passos Bueno, M. R., Passos, G. A., Pesquero, J. B., Pessoa, J. G., Rahal, P., Rainho, C. A., Reis, C. P., Ricca, T. I., Rodrigues, V., Rogatto, S. R., Romano, C. M., Romeiro, J. G., Rossi, A., Sa, R. G., Sales, M. M., Sant'Anna, S. C., Santarosa, P. L., Segato, F., Silva, W. A., Jr., Silva, I. D., Silva, N. P., Soares-Costa, A., Sonati, M. F., Strauss, B. E., Tajara, E. H., Valentini, S. R., Villanova, F. E., Ward, L. S. & Zanette, D. L. (2004) A transcript finishing initiative for closing gaps in the human transcriptome. *Genome Res*, 14, 1413-23.
- Song, H., Hasson, P., Paroush, Z. & Courey, A. J. (2004) Groucho oligomerization is required for repression in vivo. *Mol Cell Biol*, 24, 4341-50.
- Sorek, R. (2007) The birth of new exons: mechanisms and evolutionary consequences. *Rna*, 13, 1603-8.
- Sorek, R., Ast, G. & Gaur, D. (2002) Alu-containing exons are alternatively spliced. *Genome Res*, 12, 1060-7.
- Spencer, M. L., Theodosiou, M. & Noonan, D. J. (2004) NPDC-1, a novel regulator of neuronal proliferation, is degraded by the ubiquitin/proteasome system through a PEST degradation motif. *J Biol Chem*, 279, 37069-78.
- Spielhaupt, C. & Schatzl, H. M. (2001) PrP^C directly interacts with proteins involved in signaling pathways. *J Biol Chem*, 276, 44604-12.
- Steele, A. D., Hutter, G., Jackson, W. S., Heppner, F. L., Borkowski, A. W., King, O. D., Raymond, G. J., Aguzzi, A. & Lindquist, S. (2008) Heat shock factor 1 regulates lifespan as distinct from disease onset in prion disease. *Proc Natl Acad Sci U S A*, 105, 13626-31.
- Stewart, R. S., Drisaldi, B. & Harris, D. A. (2001) A transmembrane form of the prion protein contains an uncleaved signal peptide and is retained in the endoplasmic reticulum. *Mol Biol Cell*, 12, 881-9.
- Stoneley, M. & Willis, A. E. (2003) Aberrant regulation of translation initiation in tumorigenesis. *Curr Mol Med*, 3, 597-603.
- Storz, J. F., Opazo, J. C. & Hoffmann, F. G. (2012) Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Mol Phylogenet Evol*, In Press, Corrected Proof, Available online 27 July 2012.
- Strachan, G. D., Koike, M. A., Siman, R., Hall, D. J. & Jordan-Sciutto, K. L. (2005a) E2F1 induces cell death, calpain activation, and MDMX degradation in a transcription independent manner implicating a novel role for E2F1 in neuronal loss in SIV encephalitis. *J Cell Biochem*, 96, 728-40.

- Strachan, G. D., Morgan, K. L., Otis, L. L., Caltagarone, J., Gittis, A., Bowser, R. & Jordan-Sciutto, K. L. (2004) Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like Ech-associated protein. *Biochemistry*, 43, 12113-22.
- Strachan, G. D., Ostrow, L. A. & Jordan-Sciutto, K. L. (2005b) Expression of the fetal Alz-50 clone 1 protein induces apoptotic cell death. *Biochem Biophys Res Commun*, 336, 490-5.
- Strom, A., Diecke, S., Hunsmann, G. & Stuke, A. W. (2006) Identification of prion protein binding proteins by combined use of far-Western immunoblotting, two dimensional gel electrophoresis and mass spectrometry. *Proteomics*, 6, 26-34.
- Strumbo, B., Ronchi, S., Bolis, L. C. & Simonic, T. (2001) Molecular cloning of the cDNA coding for *Xenopus laevis* prion protein. *FEBS Lett*, 508, 170-4.
- Strumbo, B., Sangiorgio, L., Ronchi, S., Gready, J. E. & Simonic, T. (2006) Cloning and analysis of transcripts and genes encoding fish-specific proteins related to PrP. *Fish Physiol Biochem*, 32, 339-353.
- Styren, S. D., Bowser, R. & Dekosky, S. T. (1997) Expression of fetal ALZ-50 reactive clone 1 (FAC1) in dentate gyrus following entorhinal cortex lesion. *J Comp Neurol*, 386, 555-61.
- Sunyach, C., Jen, A., Deng, J., Fitzgerald, K. T., Frobert, Y., Grassi, J., McCaffrey, M. W. & Morris, R. (2003) The mechanism of internalization of glycosylphosphatidylinositol-anchored prion protein. *EMBO J*, 22, 3591-601.
- Svensson, O., Arvestad, L. & Lagergren, J. (2006) Genome-wide survey for biologically functional pseudogenes. *PLoS Comput Biol*, 2, e46.
- Swalla, B. J., Just, M. A., Pederson, E. L. & Jeffery, W. R. (1999) A multigene locus containing the *Manx* and *bobcat* genes is required for development of chordate features in the ascidian tadpole larva. *Development*, 126, 1643-53.
- Swanson, B. J., Jack, H. M. & Lyons, G. E. (1998) Characterization of myocyte enhancer factor 2 (MEF2) expression in B and T cells: MEF2C is a B cell-restricted transcription factor in lymphocytes. *Mol Immunol*, 35, 445-58.
- Swanson, W. J. & Vacquier, V. D. (2002) The rapid evolution of reproductive proteins. *Nat Rev Genet*, 3, 137-44.
- Szabo, G., Katarova, Z. & Greenspan, R. (1994) Distinct protein forms are produced from alternatively spliced bicistronic glutamic acid decarboxylase mRNAs during development. *Mol Cell Biol*, 14, 7535-45.
- Takeda, J., Suzuki, Y., Nakao, M., Barrero, R. A., Koyanagi, K. O., Jin, L., Motono, C., Hata, H., Isogai, T., Nagai, K., Otsuki, T., Kuryshev, V., Shionyu, M., Yura, K., Go, M., Thierry-Mieg, J., Thierry-Mieg, D., Wiemann, S., Nomura, N., Sugano, S., Gojobori, T. & Imanishi, T. (2006) Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res*, 34, 3917-28.
- Tang, H., Heeley, T., Morlec, R. & Hubbard, S. J. (2007) Characterising alternate splicing and tissue specific expression in the chicken from ESTs. *Cytogenet Genome Res*, 117, 268-77.
- Tang, T., Rector, K., Barnett, C. D. & Mao, C. D. (2008) Upstream open reading frames regulate the expression of the nuclear Wnt13 isoforms. *Biochem Biophys Res Commun*, 366, 1081-8.
- Tao, Y., Kassatly, R. F., Cress, W. D. & Horowitz, J. M. (1997) Subunit composition determines E2F DNA-binding site specificity. *Mol Cell Biol*, 17, 6994-7007.

- Taylor, D. R. & Hooper, N. M. (2007) The low-density lipoprotein receptor-related protein 1 (LRP1) mediates the endocytosis of the cellular prion protein. *Biochem J*, 402, 17-23.
- Telling, G. C., Haga, T., Torchia, M., Tremblay, P., DeArmond, S. J. & Prusiner, S. B. (1996) Interactions between wild-type and mutant prion proteins modulate neurodegeneration in transgenic mice. *Genes Dev*, 10, 1736-50.
- Telling, G. C., Scott, M., Mastrianni, J., Gabizon, R., Torchia, M., Cohen, F. E., DeArmond, S. J. & Prusiner, S. B. (1995) Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein. *Cell*, 83, 79-90.
- Thomas, J. D. & Johannes, G. J. (2007) Identification of mRNAs that continue to associate with polysomes during hypoxia. *Rna*, 13, 1116-31.
- Tobler, I., Gaus, S. E., Deboer, T., Achermann, P., Fischer, M., Rulicke, T., Moser, M., Oesch, B., McBride, P. A. & Manson, J. C. (1996) Altered circadian activity rhythms and sleep in mice devoid of prion protein. *Nature*, 380, 639-42.
- Tranulis, M. A., Espenes, A., Comincini, S., Skretting, G. & Harbitz, I. (2001) The PrP-like protein Doppel gene in sheep and cattle: cDNA sequence and expression. *Mamm Genome*, 12, 376-9.
- Trouche, D., Cook, A. & Kouzarides, T. (1996) The CBP co-activator stimulates E2F1/DP1 activity. *Nucleic Acids Res*, 24, 4139-45.
- Tucker, N. R., Middleton, R. C., Le, Q. P. & Shelden, E. A. (2011) HSF1 is essential for the resistance of zebrafish eye and brain tissues to hypoxia/reperfusion injury. *PLoS One*, 6, e22268.
- Uelhoff, A., Tatzelt, J., Aguzzi, A., Winklhofer, K. F. & Haass, C. (2005) A pathogenic PrP mutation and doppel interfere with polarized sorting of the prion protein. *J Biol Chem*, 280, 5137-40.
- Um, J. W., Nygaard, H. B., Heiss, J. K., Kostylev, M. A., Stagi, M., Vortmeyer, A., Wisniewski, T., Gunther, E. C. & Strittmatter, S. M. (2012) Alzheimer amyloid-beta oligomer bound to postsynaptic prion protein activates Fyn to impair neurons. *Nat Neurosci*.
- van de Lagemaat, L. N., Landry, J. R., Mager, D. L. & Medstrand, P. (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet*, 19, 530-6.
- van der Velden, A. W. & Thomas, A. A. (1999) The role of the 5' untranslated region of an mRNA in translation regulation during development. *Int J Biochem Cell Biol*, 31, 87-106.
- van Rheede, T., Smolenaars, M. M., Madsen, O. & de Jong, W. W. (2003) Molecular evolution of the mammalian prion protein. *Mol Biol Evol*, 20, 111-21.
- Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. (1998) Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr Biol*, 8, 96-108.
- Viles, J. H., Cohen, F. E., Prusiner, S. B., Goodin, D. B., Wright, P. E. & Dyson, H. J. (1999) Copper binding to the prion protein: structural implications of four identical cooperative binding sites. *Proc Natl Acad Sci U S A*, 96, 2042-7.
- Vinckenbosch, N., Dupanloup, I. & Kaessmann, H. (2006) Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A*, 103, 3220-5.
- Voigtlander, T., Kloppel, S., Birner, P., Jarius, C., Flicker, H., Verghese-Nikolaki, S., Sklaviadis, T., Guentchev, M. & Budka, H. (2001) Marked increase of neuronal

- prion protein immunoreactivity in Alzheimer's disease and human prion diseases. *Acta Neuropathol (Berl)*, 101, 417-23.
- Walsh, D., Li, Z., Wu, Y. & Nagata, K. (1997) Heat shock and the role of the HSPs during neural plate induction in early mammalian CNS and brain development. *Cell Mol Life Sci*, 53, 198-211.
- Wang, B., Shi, G., Fu, Y. & Xu, X. (2007) Cloning and characterization of a LASS1-GDF1 transcript in rat cerebral cortex: conservation of a bicistronic structure. *DNA Seq*, 18, 92-103.
- Wang, W., Zhang, J., Alvarez, C., Llopart, A. & Long, M. (2000) The origin of the Jingwei gene and the complex modular structure of its parental gene, yellow emperor, in *Drosophila melanogaster*. *Mol Biol Evol*, 17, 1294-301.
- Warner, R. G., Hundt, C., Weiss, S. & Turnbull, J. E. (2002) Identification of the heparan sulfate binding sites in the cellular prion protein. *J Biol Chem*, 277, 18421-30.
- Weinberg, R. A. (1995) The retinoblastoma protein and cell cycle control. *Cell*, 81, 323-30.
- Weise, J., Crome, O., Sandau, R., Schulz-Schaeffer, W., Bahr, M. & Zerr, I. (2004) Upregulation of cellular prion protein (PrP^C) after focal cerebral ischemia and influence of lesion severity. *Neurosci Lett*, 372, 146-50.
- Welm, A. L., Timchenko, N. A., Ono, Y., Sorimachi, H., Radomska, H. S., Tenen, D. G., Lekstrom-Himes, J. & Darlington, G. J. (2002) C/EBPalpha is required for proteolytic cleavage of cyclin A by calpain 3 in myeloid precursor cells. *J Biol Chem*, 277, 33848-56.
- Westaway, D., DeArmond, S. J., Cayetano-Canlas, J., Groth, D., Foster, D., Yang, S. L., Torchia, M., Carlson, G. A. & Prusiner, S. B. (1994) Degeneration of skeletal muscle, peripheral nerves, and the central nervous system in transgenic mice overexpressing wild-type prion proteins. *Cell*, 76, 117-29.
- Wheelan, S. J., Aizawa, Y., Han, J. S. & Boeke, J. D. (2005) Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res*, 15, 1073-8.
- Windl, O., Dempster, M., Estibeiro, P. & Lathe, R. (1995) A candidate marsupial PrP gene reveals two domains conserved in mammalian PrP proteins. *Gene*, 159, 181-6.
- Wong, B. S., Liu, T., Paisley, D., Li, R., Pan, T., Chen, S. G., Perry, G., Petersen, R. B., Smith, M. A., Melton, D. W., Gambetti, P., Brown, D. R. & Sy, M. S. (2001) Induction of HO-1 and NOS in doppel-expressing mice devoid of PrP: implications for doppel function. *Mol Cell Neurosci*, 17, 768-75.
- Wormald, M. R. & Dwek, R. A. (1999) Glycoproteins: glycan presentation and protein-fold stability. *Structure*, 7, R155-60.
- Wu, L., Wells, D., Tay, J., Mendis, D., Abbott, M. A., Barnitt, A., Quinlan, E., Heynen, A., Fallon, J. R. & Richter, J. D. (1998) CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. *Neuron*, 21, 1129-39.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C. & Allis, C. D. (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442, 86-90.
- Xiao, H., Sandaltzopoulos, R., Wang, H. M., Hamiche, A., Ranallo, R., Lee, K. M., Fu, D. & Wu, C. (2001) Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol Cell*, 8, 531-43.

- Yamaguchi, N., Sakaguchi, S., Shigematsu, K., Okimura, N. & Katamine, S. (2004) Doppel-induced Purkinje cell death is stoichiometrically abrogated by prion protein. *Biochem Biophys Res Commun*, 319, 1247-52.
- Yang, W., Ng, P., Zhao, M., Wong, T. K., Yiu, S. M. & Lau, Y. L. (2008) Promoter-sharing by different genes in human genome--CPNE1 and RBM12 gene pair as an example. *BMC Genomics*, 9, 456.
- Yeager, M. & Hughes, A. L. (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol Rev*, 167, 45-58.
- Yehiely, F., Bamborough, P., Da Costa, M., Perry, B. J., Thinakaran, G., Cohen, F. E., Carlson, G. A. & Prusiner, S. B. (1997) Identification of candidate proteins binding to prion protein. *Neurobiol Dis*, 3, 339-55.
- Yin, S. M., Sy, M. S., Yang, H. Y. & Tien, P. (2004) Interaction of Doppel with the full-length laminin receptor precursor protein. *Arch Biochem Biophys*, 428, 165-9.
- Yoshikawa, D., Kopacek, J., Yamaguchi, N., Ishibashi, D., Yamanaka, H., Yamaguchi, Y., Katamine, S. & Sakaguchi, S. (2007) Newly established in vitro system with fluorescent proteins shows that abnormal expression of downstream prion protein-like protein in mice is probably due to functional disconnection between splicing and 3' formation of prion protein pre-mRNA. *Gene*, 386, 139-46.
- Zanata, S. M., Lopes, M. H., Mercadante, A. F., Hajj, G. N., Chiarini, L. B., Nomizo, R., Freitas, A. R., Cabral, A. L., Lee, K. S., Juliano, M. A., de Oliveira, E., Jachieri, S. G., Burlingame, A., Huang, L., Linden, R., Brentani, R. R. & Martins, V. R. (2002) Stress-inducible protein 1 is a cell surface ligand for cellular prion that triggers neuroprotection. *EMBO J*, 21, 3307-16.
- Zawlik, I., Witusik, M., Hulas-Bigoszewska, K., Piaskowski, S., Szybka, M., Golanska, E., Liberski, P. P. & Rieske, P. (2006) Regulation of PrPC expression: nerve growth factor (NGF) activates the prion gene promoter through the MEK1 pathway in PC12 cells. *Neurosci Lett*, 400, 58-62.
- Zeng, L. & Zhou, M. M. (2002) Bromodomain: an acetyl-lysine binding domain. *FEBS Lett*, 513, 124-8.
- Zerucha, T., Stuhmer, T., Hatch, G., Park, B. K., Long, Q., Yu, G., Gambarotta, A., Schultz, J. R., Rubenstein, J. L. & Ekker, M. (2000) A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J Neurosci*, 20, 709-21.
- Zhang, C. L., McKinsey, T. A., Chang, S., Antos, C. L., Hill, J. A. & Olson, E. N. (2002) Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell*, 110, 479-88.
- Zhang, L., Jiang, H., Gao, X., Zou, Y., Liu, M., Liang, Y., Yu, Y., Zhu, W., Chen, H. & Ge, J. (2011) Heat shock transcription factor-1 inhibits H₂O₂-induced apoptosis via down-regulation of reactive oxygen species in cardiac myocytes. *Mol Cell Biochem*, 347, 21-8.
- Zhang, Q., Vo, N. & Goodman, R. H. (2000) Histone binding protein RbAp48 interacts with a complex of CREB binding protein and phosphorylated CREB. *Mol Cell Biol*, 20, 4970-8.
- Zhang, X. & Firestein, S. (2002) The olfactory receptor gene superfamily of the mouse. *Nat Neurosci*, 5, 124-33.
- Zhang, X. H. & Chasin, L. A. (2006) Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc Natl Acad Sci U S A*, 103, 13427-32.

- Zhao, L., Spassieva, S. D., Jucius, T. J., Shultz, L. D., Shick, H. E., Macklin, W. B., Hannun, Y. A., Obeid, L. M. & Ackerman, S. L. (2011) A deficiency of ceramide biosynthesis causes cerebellar purkinje cell neurodegeneration and lipofuscin accumulation. *PLoS Genet*, 7, e1002063.
- Zhu, P. & Bowser, R. (1996) Identification and analysis of the complete cDNA sequence for the human FAC1 gene. *Biochim Biophys Acta*, 1309, 5-8.
- Zorio, D. A., Cheng, N. N., Blumenthal, T. & Spieth, J. (1994) Operons as a common form of chromosomal organization in *C. elegans*. *Nature*, 372, 270-2.

Appendices

Appendix 1. Sequences of oligonucleotide primers used in 5' and 3' RACE. Numbering refers to the relevant cDNA sequence and positions primer relatively to ORF. Negative numbers correspond to 5'UTRs. GSP, gene specific primer, reverse (R) for 5' amplification and forward (F) for 3' amplification. *PRNP*, prion protein gene; *PRND*, doppel gene.

<i>PRNP, X. tropicalis</i>	
5'GSP-R	³¹⁵ ACCGAGCATGTAGCCCCCGATAG ²⁹³
5'nested GSP-R	²⁴⁴ TGGGAGGTTTCCATTGCTTGTTGTTG ²¹⁹
3'GSP-F	³⁸⁷ CCAGATGCCAGAGCGTGTTTACAG ⁴¹⁰
3'nested GSP-F	⁵⁰⁷ AGGGAAGAACACCAGCGAGGTAAAC ⁵³¹

<i>PRND, X. tropicalis</i>	
5'GSP-R	⁴⁷⁷ GGTCACTGCCCCATTGGTAGGCTT ⁴⁵⁴
5'nested GSP-R	²³⁰ AGCTCAGATTCCCTCGGTAAGGTTAAAG ²⁰⁴
3'GSP-F	¹⁷⁷ CTTCAGAGGCAGGGCACTCAATGTG ²⁰¹
	⁴¹¹ GGCCACGCAGGTCCTACAGTTTCTC ⁴³⁵
3'nested GSP-F	²⁹⁵ GGTGCTGGTGGGACTGACGAGTTT ³¹⁸
	⁴⁶⁸ GGCAGTGACCTGCACTGGTGGAT ⁴⁹⁰
F-exon	TGGACTGAAGGAGTAGAAA ⁻⁶⁵ GTCACCAT ⁻⁵⁸

<i>PRNP, X. laevis</i>	
5'GSP-R	³¹¹ CGACCCACTGCATTACCGAGCAT ²⁸⁹
5'nested GSP-R	⁹⁴ TACTTTTCCCACCACCGCTCTTCTTG ⁶⁹

<i>PRND, X. laevis</i>	
5'GSP-R	³¹⁸ CTCCTCCGTGCCGGCAGTATCATT ²⁹⁵
5'nested GSP-R	²⁹² GCTGGGAGCGCCATGGGTAGTAA ²⁶⁹
3'GSP-F	¹⁵⁹ TCCTGTCCTCGGACACCTATTCTTCA ¹⁸⁴
3'nested GSP-F	²⁴⁷ CTGTACAGATTCCCGGACGGACTTT ²⁷¹

<i>PRNP, M. domestica</i>	
5'mdPrP-GSP137	⁹² GGTCTTGGCTTCGGTTTCTTACAG ⁶⁹
5'mdPrP-nGSP83	³⁸ AGAGCCAAGAACCAGTATCCCAAG ¹⁵
3'mdPrP-GSP356	³¹¹ GCAGCTACAACCAGAAGTGGA ³³⁵
3'mdPrP-nGSP457	⁴¹¹ ATACATGCTGGGGAGTGCCATGA ⁴³³

<i>PRND, M. domestica</i>	
5'mdDpl-GSP387	³⁴² GCGGGTAGCATTACACAGTTGCT ³¹⁹
5'mdDpl-nGSP276	²³¹ ACTGTTGCCCTCCTCCCCAAAGT ²⁰⁹
3'mdDpl-GSP295	²⁵⁰ AGCCTGTTCCAGATGAAATCCA ²⁷²
3'mdDpl-nGSP351	³⁰⁶ GGAGGTCTTCATCAGCAACTGTGT ³²⁹

Appendix 2. Sequences of oligonucleotide primers used in tissue expression analysis. Numbering refers to the relevant cDNA sequence and positions primer relatively to ORF. Negative numbers correspond to 5'UTRs.

	<i>PRNP, X. tropicalis</i>	
F1	LxPrnp400	²⁴¹ CCTCCCAAGTCAAAAACCAA ²⁶⁰
R1	RxPrnp400	⁶³³ CAGCCAAGGGTTAGAGAGCA ⁶¹⁴

	<i>PRND, X. tropicalis</i>	
F2	LxDpl400	⁷² TAGAAGAGCAGCAAGCAGCA ⁹¹
R2	RxDpl400	⁴⁶⁹ CCCCATTGGTAGGCTTTACA ⁴⁵⁰
F3	LxDpl200	¹³³ AGAAGGCCAAAAGTGACCAA ¹⁵³
R3	RxDpl200	³⁴¹ GTGGTGTTAAGGCACCCACT ³²²

	<i>PRNP1, X. laevis</i>	
F4	LxlPp/D58	⁻¹⁵⁵ CTCACTGCTTTCCCGATCAC ⁻¹³⁶
F5	Lhit7ex1	⁻⁹⁶ GCTTCTCATTTGCCTTCCTG ⁻⁷⁷
R4	RxlPrnp	¹²³ GTTGCTCCCTGTGTTCCATC ¹⁰⁴

	<i>PRNP2, X. laevis</i>	
F6	Lhit6ex1	⁻⁹⁴ CGGATCACGAGCTTCTCATT ⁻⁷⁵
F7	Lhit1ex1	⁻¹¹⁶ TCCTGAACCTCCCCCTGTA ⁻⁹⁸
R5	Rhit6-1ex2	⁻¹⁶ TGAGAGGATAAGTAAGCCCAA ⁻³⁶
R6	RxlPp350	³⁶⁹ CCAAGCATGTAGCCTCCAAT ³⁵⁰

	<i>PRND, X. laevis</i>	
F8	LxlD6ex1	⁻²⁷⁷ CACACTCCCAGACAAGAGCA ⁻²⁵⁸
F4	LxlPp/D58	⁻¹⁵⁵ CTCACTGCTTTCCCGATCAC ⁻¹³⁶
F9	LxlDpl298	⁻¹¹³ TTTTCAGGGCAAAGCAGAGT ⁻⁹⁴
F10	LxlDpl373	⁻³⁸ GGAAGAAACAGCAGCAAAGG ⁻¹⁹
R7	RxlDpl	³¹¹ GTGCCGGCAGTATCATTC ²⁹³
R8	R180Dplex2	⁻¹⁹ CCTTTGCTGCTGTTTCTTCC ⁻³⁸
R9	RxlPex2	⁻¹⁷ TGAGGGTATAGAGTGTGCCAA ⁻³⁸

	<i>GAPDH, X. laevis</i>	
F11	LxlGAPDH	⁴⁵⁷ AACTGTCTGGCTCCTCTTGC ⁴⁷⁶
R10	RxlGAPDH	⁷⁵⁸ ACTTTGCAGGCTTCTTCAGG ⁷³⁸

	<i>PRND, X. tropicalis</i>	
F2	LxDpl400	⁷² TAGAAGAGCAGCAAGCAGCA ⁹¹
R2	RxDpl400	⁴⁶⁹ CCCCATTGGTAGGCTTTACA ⁴⁵⁰
F3	LxDpl200	¹³³ AGAAGGCCAAAAGTGACCAA ¹
R3	RxDpl200	³⁴¹ GTGGTGTTAAGGCACCCACT ³²²

	<i>PRNP, M. domestica</i>	
LmdPrP1		⁻¹¹³ AGTTAAGACCTGACAACGACCA ⁻⁹²
RmdPrP521		⁴⁰⁸ ACCAAGCCCTCCTACAACCTG ³⁸⁹
RmdCPE167		²⁵⁴⁸ CAAATACGTGCAAAGTTACAAATC ²⁵²⁵

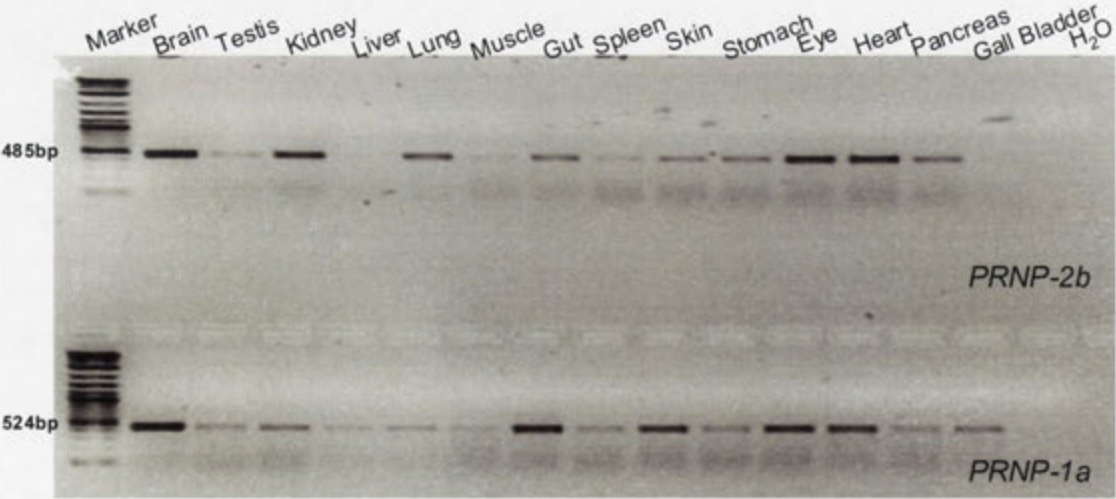
<i>PRND, M. domestica</i>	
RmdDpl350	³⁵⁰ TTGATTACGCGGGTAGCATT ³³¹
Lmd2D	⁻⁴³⁶ GACAAAGACCTCCTTGGATCAC ⁻⁴¹⁵
Lmd5D	⁻⁴⁰ AGCAACCTTTTCCAACAACG ⁻²¹

<i>GAPDH, M. domestica</i>	
LmdGAPDH	³⁵³ TTATTTCTGCCCCCTCTGCT ³⁷²
RmdGAPDH	⁸⁰⁴ ACGCTTGCTTCATCACCTTC ⁷⁸⁵

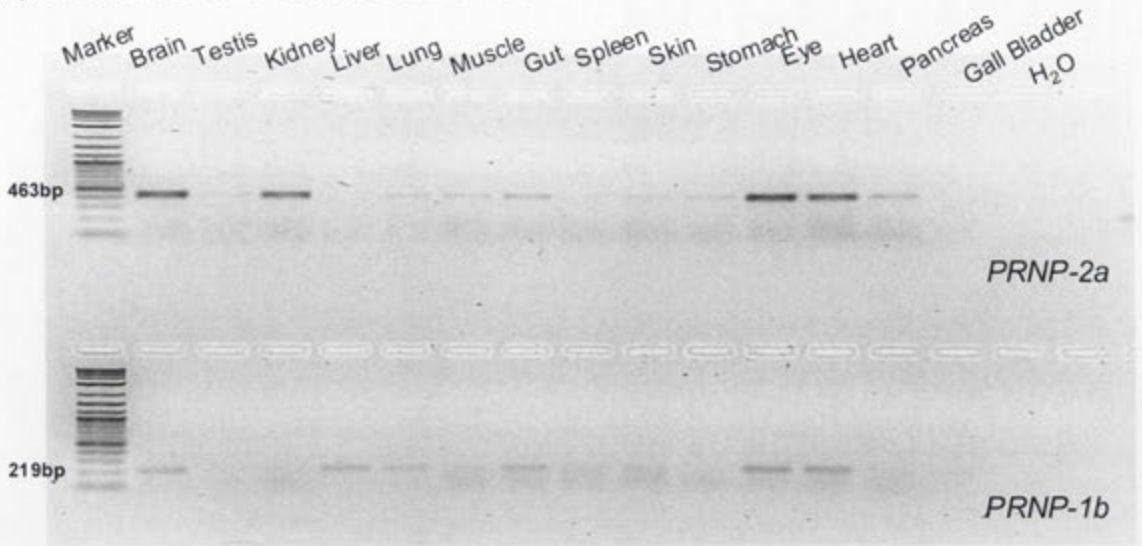
Mouse <i>Prnp</i>, <i>Prnd</i> and β-actin	
LmmPex1-5	⁻¹⁵¹ GATCAGCAGACCGATTCTGG ⁻¹³²
LmmPex1-94	⁻⁶² GCATTCTGCCTTCCTAGTGG ⁻⁴³
RM587	⁵²⁶ ACCAAACGAAACCCAGCAG ⁵⁰⁸
LMex1Dpl	⁻³⁵ TCAGAGGCCACAGTAGCAGA ⁻¹⁶
L β -actin	³⁴⁹ AGATGACCCAGATCATGTTTG ³⁶⁹
R β -actin	⁹⁸⁹ AGCAATGATCTTGATCTTCATG ⁹⁶⁸

Appendix 3. Original gel images for tissue and cell expression analysis. (a) *X. laevis* PRNP-2b and PRNP-1a, (b) *X. laevis* PRNP-2a and PRNP-1b, (c) *X. laevis* PRND-ch (chimeric), (d) *X. laevis* PRND-a (non-chimeric), (e) *X. laevis* PRND-b (non-chimeric), (f) *M. domestica* PRND-1, 2, 3, 4 (chimeric), (g) *M. domestica* PRNP-1 and PRNP-3, (h) mouse *Prnd* (non-chimeric), (i) mouse *Prnd-ch1*, 2, 3 (chimeric), (j) mouse *Bptf* Region 1, (k) mouse *Bptf* Region 2, (l) mouse *Bptf* Region 1-1 (exon5), (m) mouse *Bptf* Region 1-2 (exon6), (n) mouse and human *Bptf*, *Prnd*, and *Prnp*, (o) mouse *E2F1*. Analysed tissue samples are indicated on top of the images. H₂O was used as a negative no template control. 2-log ladder (Invitrogen) or HyperLadderTMIVbp ladder (Bioline) (in gels j and k) was used as a molecular marker. Numbering on the left side of the images refers to the size of the relevant PCR product.

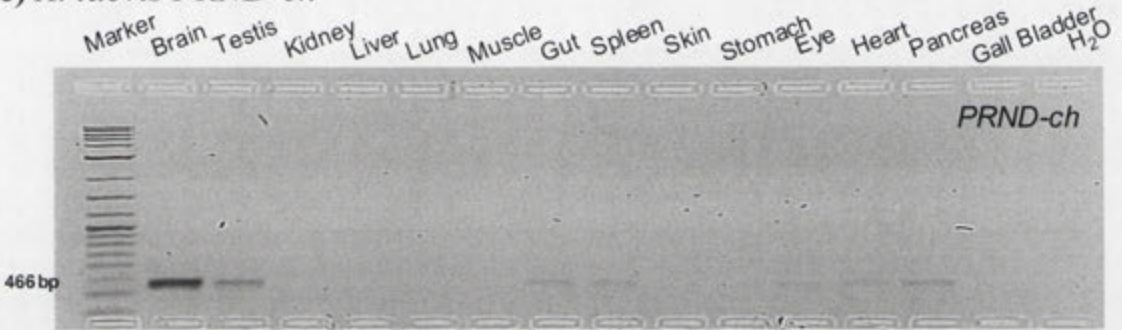
(a) *X. laevis* PRNP-2b and PRNP-1a



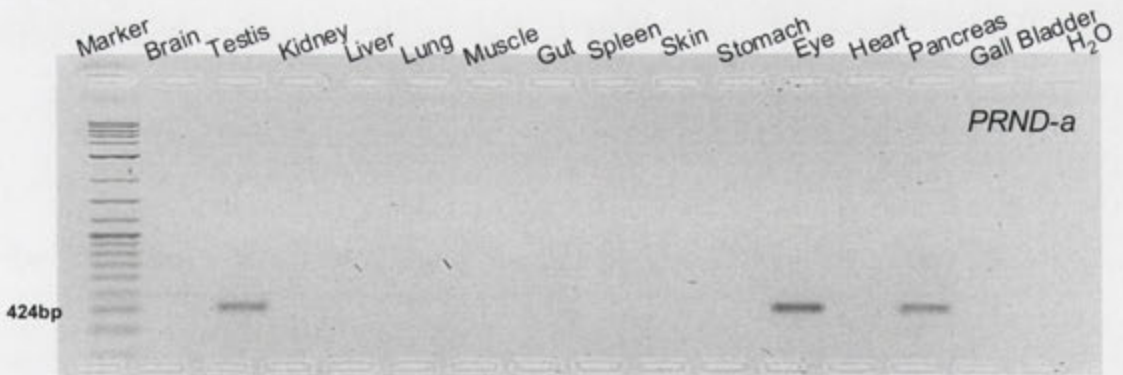
(b) *X. laevis* PRNP-2a and PRNP-1b



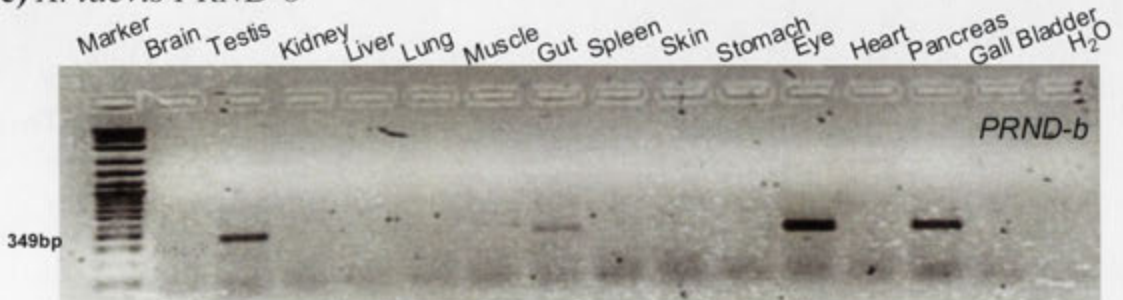
(c) *X. laevis* PRND-ch



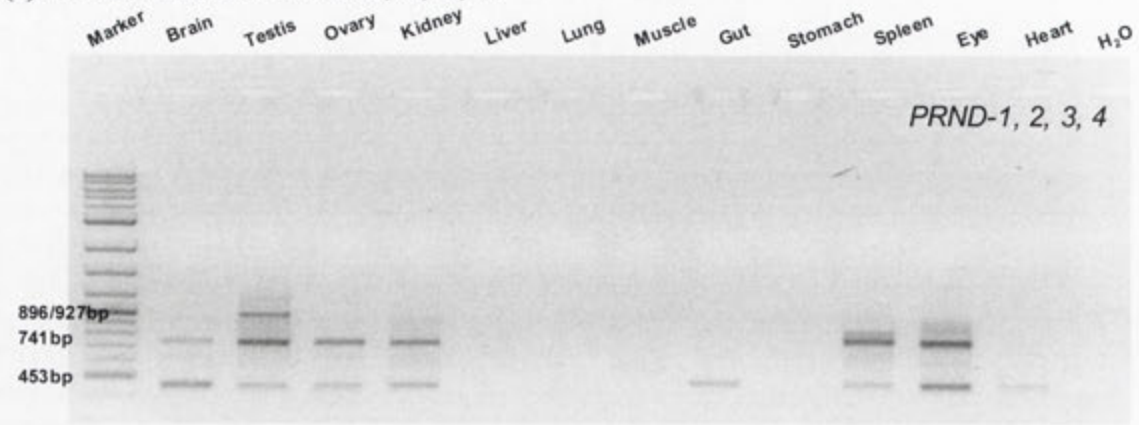
(d) *X. laevis* PRND-a



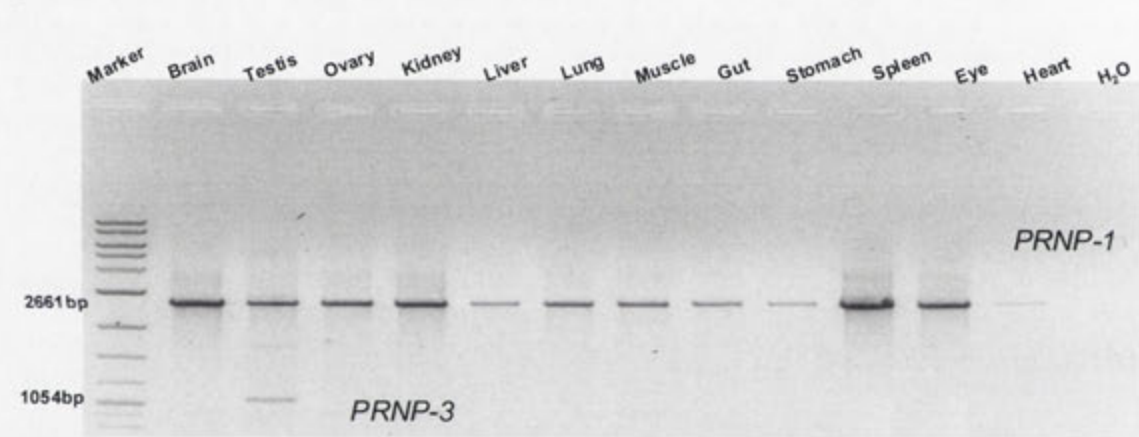
(e) *X. laevis* PRND-b



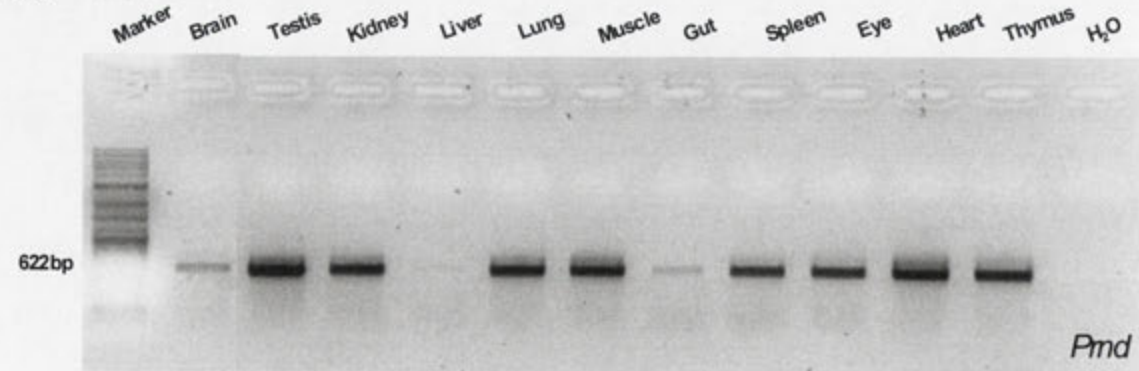
(f) *M. domestica* PRND-1, 2, 3, 4



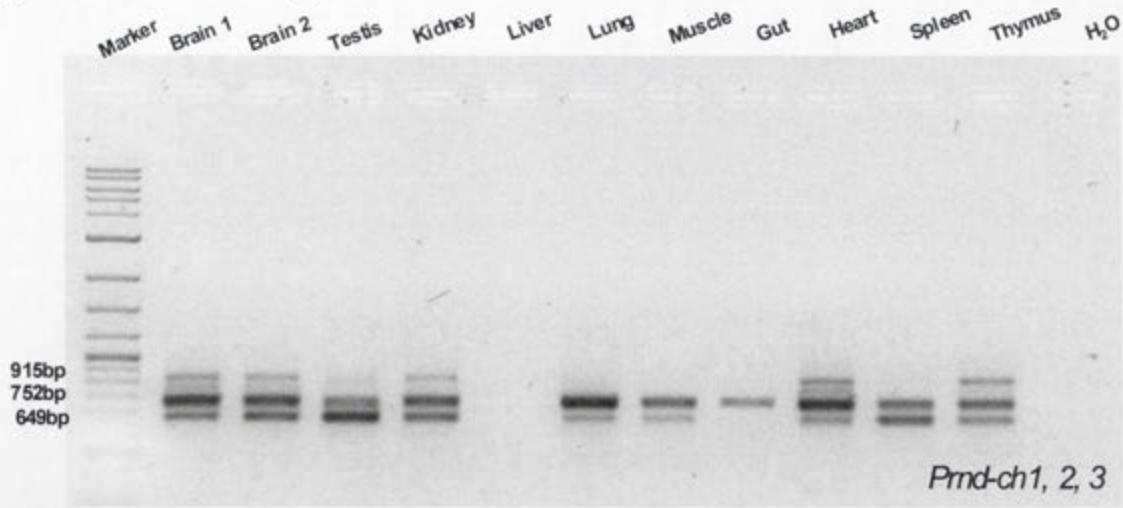
(g) *M. domestica* PRNP-1 and PRNP-3



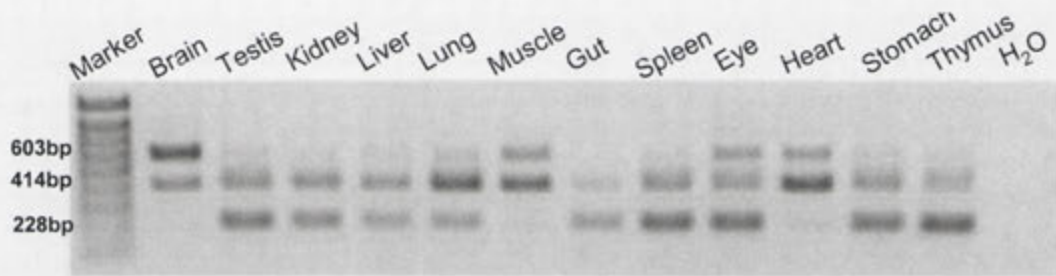
(h) Mouse *Prnd*



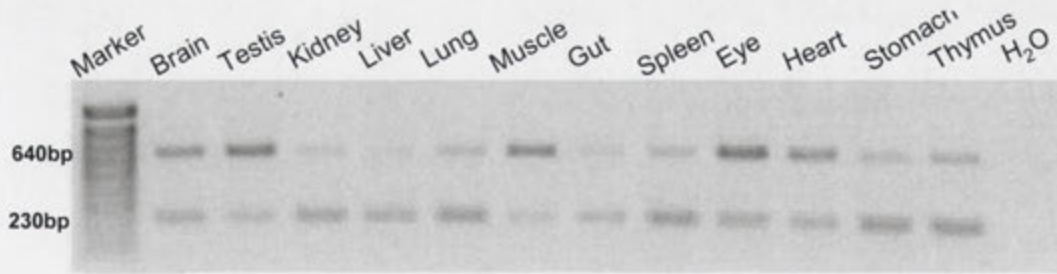
(i) Mouse *Prnd-ch1, 2, 3*



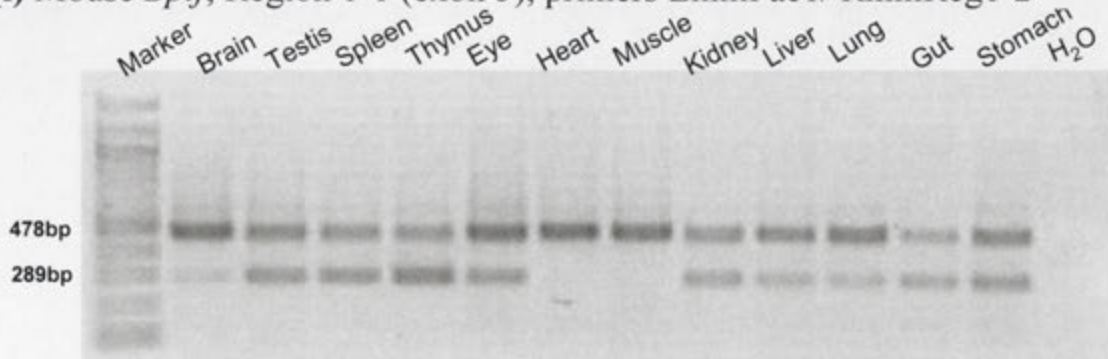
(j) Mouse *Bptf* Region 1, primers LmmFac1/RmmFac1



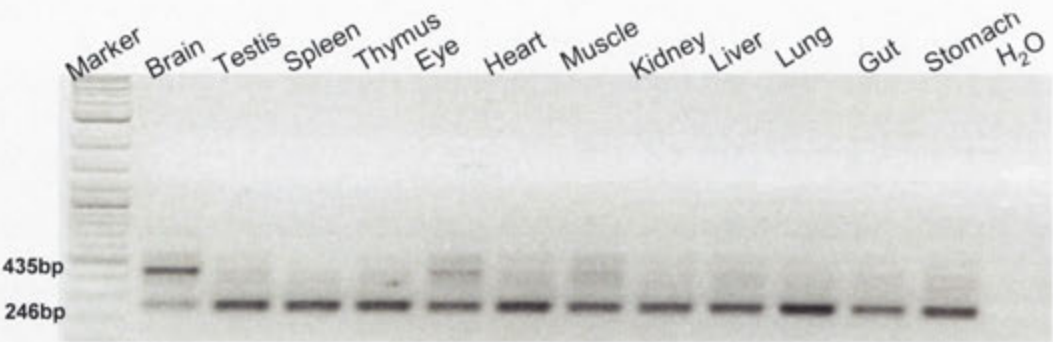
(k) Mouse *Bptf* Region 2, primers LmmBptf/RmmBptf



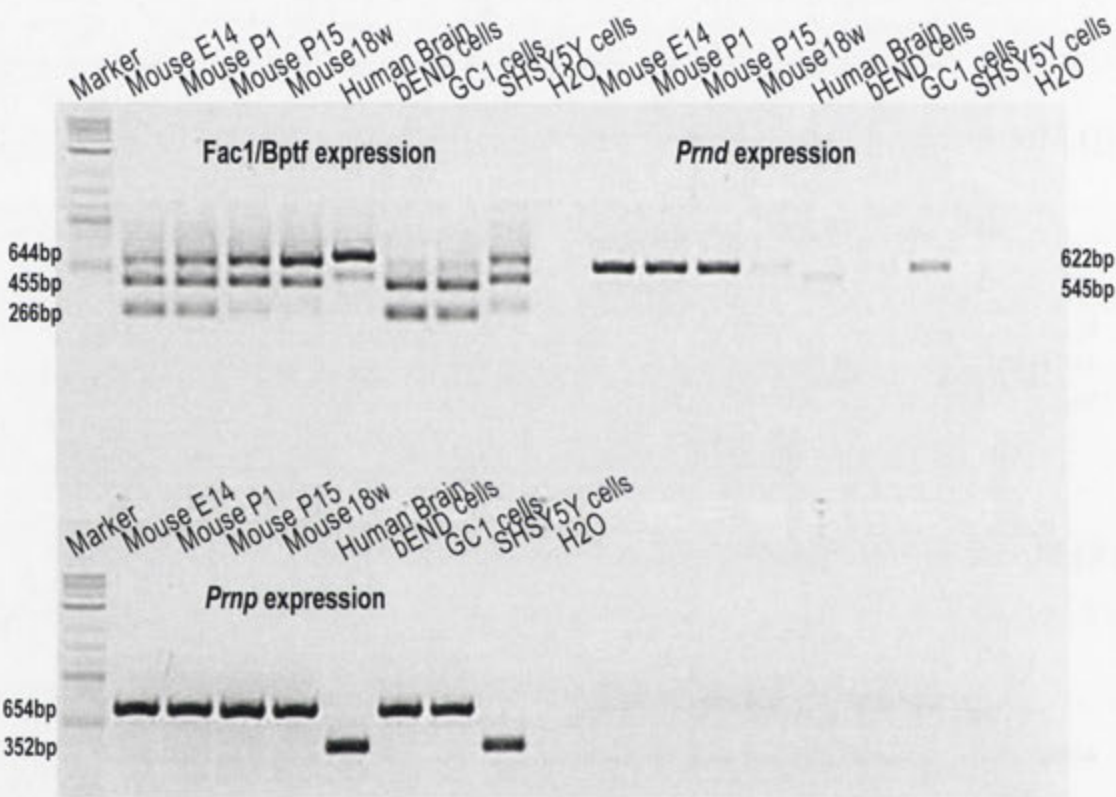
(l) Mouse *Bptf*, Region 1-1 (exon 5), primers LmmFac1/ RmmReg1-2



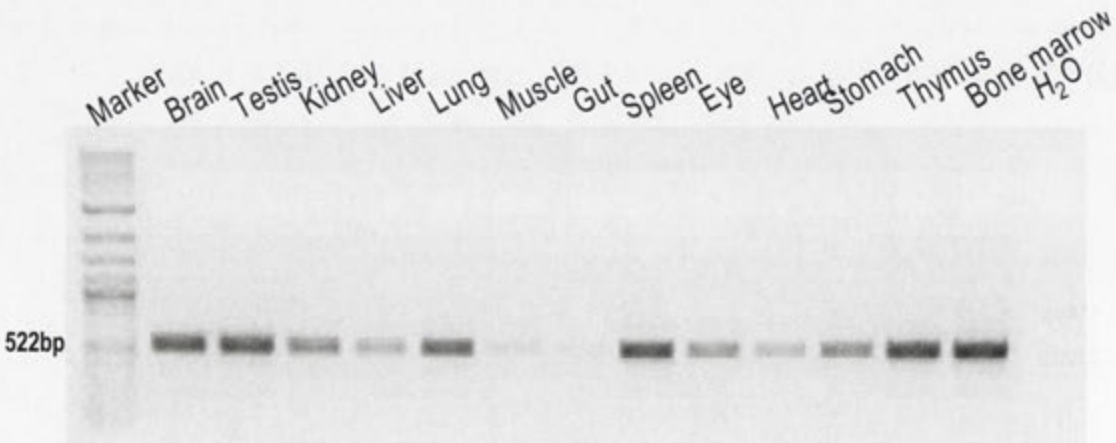
(m) Mouse *Bptf* Region 1-2 (exons 6), primers LmmReg1-1/ RmmFac24



(n) Mouse and human *Bptf*, *Prnd*, and *Prnp* in tissues and cell lines



(o) Mouse *E2F1*



MLRSLWTSVLVLISLVCALTVSSKKS GSGKSKTG GWSGSGNRNPYPGGY G WNTGGNTGGSWGQPYNPSSG
NNFNFKQWKPKPSKTNMKAVAVGAAAGAI GG YMLGNAMGRMSYHFSNPMEARYYNDYYNQMPERVYRPMY
RGEHVSEDRFVTD CY NMSVTEYIIKPAEGKNTSEVNQLETRVKSQIIREMCITEYRRSGFKVLSNPWL
ILTITLEVYEVIE

MGRQNLFSCILILLLILYCSLSSPRRAASSKKISKTTDLSRGAKRRPKVTNSPALGDLFRGRALNVNFN
LTEESELYTANLYSFPDGLYPRPAHLGAGGTDEFISGCLNTTIERNKVWISQLEDDEEGDIYMSVATQ
VLQFLCMENYVKPTGAVTCTGGWLWVFIGVMHFFFLERKGD

MGKIHLYWFLALFIMTWSDLTLCCKPKPRPGGGWNSGGNRYPGQSGGWGHPQGGGTNWGQPHAGGSNWG
QPRPGGSNWGQPHPGGSNWGQPHPGGSNWGQAGSSYNQKWKPKTNMKHVAGAAAAGAVVGGGLGGYML
GSAMSRPIMHFGNDYEDRYRENQYRPNQVMYRPIDQYNNQNNFVHDCVNIIVKQHTTTTTTCKGENFTE
TDIKIMERVVEQMCTITOYONEYRSAYVAFFSAPPVTLTLLLSFLIFLIVS

MRRHLGICWIAFFALLFSDLSLVKA KTTQRNKSNRKGLQTNRTNPPTTVQPSEKLQGT FIRNGRKLVID
FGEEGNSYYATHYSLFPDEIHYAGCAESNV KEVFISNCVNATRVINKLEPLEEQNIS DIYSRILEQLIK
ELCALNYCEFRGTGKGTSALFRPICYGLPGSDLLDSEIHKHRA

Cysteine involved in disulphide bridge formation

GPI anchor site

Appendix 5. Nucleotide and deduced amino acid sequences of cDNAs for chimeric and non-chimeric transcripts of (a) *X. tropicalis* PRNP, (b) *X. tropicalis* PRND (chimeric), (c) *X. laevis* PRNP-1a, (d) *X. laevis* PRNP-1b, (e) *X. laevis* PRND-ch (chimeric), (f) *X. laevis* PRND-a, PRND-b (non-chimeric), (g) *X. laevis* PRNP-2a, (h) *X. laevis* PRNP-2b. Exons are numbered and non-coding exons are enclosed by full, dashed or dotted boxes. In the absence of genomic sequence for *X. laevis* the exon boundaries have been predicted based on alignments of cDNA sequences with those for *X. tropicalis*. The cDNA sequence corresponding to the ORF is highlighted in grey. The 5' or 3' end variants are represented by lower case letters. Figure produced by Nagesh Chakka.

(a) *X. tropicalis* PRNP

Exon1	
1	actgctagtcccactgccctcctgatctgcctggcACATCCATCCTCAGTCCCTCTGTCTG 60
Exon2	
61	TATCCTGCCTGGCACAGAGGCACCCACCAGAGCCTGGCACCCAGCTAGCTTCTCTTTGGG 120
Exon3	
121	CATACTACCCACACACAGGTTTGCCATGATGCTAAGAAGCCTCTGGACTTCTTTAGTCC 180
1	M L R S L W T S L V L 11
181	TTATCTCACTTGTATGCGCACTGACTGTATCTTCCAAGAAGACTGGTAGTGGGAAAAGCA 240
12	I S L V C A L T V S S K K S G S G K S K 31
241	AAACCGGAGGATGGAACAGTGGGAGCAACCGGAACCCAACTACCCAGGAGGCTATGGCT 300
32	T G G W N S G S N R N P N Y P G G Y G W 51
301	GGAACACCGGAGGGAATACTGGAGGCGCTGGGGCCAACCTTATAATCCAGTGGCGGAA 360
52	N T G G N T G G S W G Q P Y N P S G G N 71
361	ACAATTTCAACAACAAGCAATGGAAACCTCCCAAGTCAAAAACCAATATGAAGGCTGTGG 420
72	N F N N K Q W K P P K S K T N M K A V A 91
421	CCGTAGGCGCTGCTGCAGGCGCTATCGGGGGCTACATGCTCGGTAATGCAATGGGTCTGA 480
92	V G A A A G A I G G Y M L G N A M G R M 111
481	TGAGCTATCATTTAGCAATCCCATGGAAGCACGTTATTATAACGACTACTACAACGAGA 540
112	S Y H F S N P M E A R Y Y N D Y Y N Q M 131
541	TGCCAGAGCGTGTTTACAGGCCAATGTACAGAGGCGAGGAGCACGTGTCAGAGGATAGGT 600
132	P E R V Y R P M Y R G E E H V S E D R F 151
601	TTGTACGGAAGTGTACAAATATGTCAAGTACAGAGTACATCAAGCCAGCTGAAGGGA 660
152	V T D C Y N M S V T E Y I I K P A E G K 171
661	AGAACACCAGCGAGGTAAACCAGTTGGAAACCAGGGTGAAGTCCCAAATTATTCGCGAGA 720
172	N T S E V N Q L E T R V K S Q I I R E M 191
721	TGTGTATCACTGAGTACAGGAGAGGATCGGGATTAAAGGTGCTCTCTAACCCCTTGGCTGA 780
192	C I T E Y R R G S G F K V L S N P W L I 211
781	TCCTCACTATCACTCTTTTGTGTTACTTTGTGATAGAGTGAACAGAGGGAAGGCCAAATG 840
212	L T I T L F V Y F V I E * 223
841	TATGTATATAG 851

(b) *X. tropicalis* PRND (chimeric)

Exon1

1 GCCCTCCTGATCTGCCTGGCACATCCATCCTCAGTCCCTCTGTGATCCTGCCTGGCAC 60

Exon2

61 AGAGGCACCCACCAGAGCCTGGCACCCAGCTAGCTTCTCTTTGGGCATACTCACCACAC 120

Exon3 **Exon4**

121 ACAGGGTCACCATAGATCACCATGGGGCCATTTAGCTCTTCCACACTTCTACCACAGG 180

-----|-----|-----|-----|-----|-----|

181 TGGTGACAGAAATGGGAAGGCAGAATCTATTCTCCTGTCTGATTCTTCTCCTGCTCATATT 240

1 M G R Q N L F S C L I L L L L I L 17

-----|-----|-----|-----|-----|-----|

241 ATATTGTAGTCTCTCTTCTCCTAGAAAGAGCAGCAAGCAGCAAAAAAATTAGCAAAACCAC 300

18 Y C S L S S P R R A A S S K K I S K T T 37

-----|-----|-----|-----|-----|-----|

301 AGATTTGAGCAGGGGAGCCAAAAGAAGGCCAAAAGTGACCAATTCTCCTGCCCTCGGAGA 360

38 D L S R G A K R R P K V T N S P A L G D 57

-----|-----|-----|-----|-----|-----|

361 TCTGTCTTTCAGAGGCAGGGCACTCAATGTGAACCTTTAACCTTACCGAGGAATCTGAGCT 420

58 L S F R G R A L N V N F N L T E E S E L 77

-----|-----|-----|-----|-----|-----|

421 TTATACAGCAAACTCTACAGCTTCCCGGATGGCCTGTACTACCCACGGCCTGCCACCT 480

78 Y T A N L Y S F P D G L Y Y P R P A H L 97

-----|-----|-----|-----|-----|-----|

481 CAGTGGTGCTGGTGGGACTGACGAGTTTATAAGTGGGTGCCTTAACACCACAATAGAAAG 540

98 S G A G G T D E F I S G C L N T T I E R 117

-----|-----|-----|-----|-----|-----|

541 AAACAAGGCTGGATCTCTCAACTGGAAGACGATGAAGAAGGGGATATTATATGAGCGT 600

118 N K V W I S Q L E D D E E G D I Y M S V 137

-----|-----|-----|-----|-----|-----|

601 GGCCACGCAGGTCTACAGTTTCTCTGTATGGAAAATTATGTAAGCCTACCAATGGGGC 660

138 A T Q V L Q F L C M E N Y V K P T N G A 157

-----|-----|-----|-----|-----|-----|

61 AGTGACCTGCACCTGGTGGATTGTGGGTCTTTATAGGTGTCATGCATTTTTTTTTTATT 720

158 V T C T G G L W V F I G V M H F F F L F 177

-----|-----|-----|-----|-----|-----|

721 TAGGAAGGGAGACTAAAGCCTAGGAATTCTGTattttatatgaacttttaagaactaact 780

178 R K G D * 181

781 gtactagcccagaggttcagcagccctataacagcaatgatccaggccttcaaatttgtc 840

841 cacagcagctcttgatctcatcttgatcttttgagtgtcagtgacactgcacattctc 900

901 agtgtgcagggtctgctgttaaagactacgccatctgtcatagaatgcagtttctacaca 960

961 ggaatatactctaataataggaatataatgtgcagtcacccactactaatagttcagcctgc 1020

1021 cagagatcaccaaggagttgcatagcctttttacaaacatatccttcatctaaaacttc 1080

1081 atctgtgacttctaataagtcctataaattacaacaggagcaatattatattctatattat 1140

1141 atacatataagcaggaattatggctgctgtgtgcacacaggggaagtggaggtttggcc 1200

1201 acggtttgttttcccttattataaattgatgaaataaaaaatcaggttaaactgg 1254

(c) *X. laevis* PRNP-1a

Exon1

1 agtccgccccaccctctcctgcatgaAGTCTCTTCCCCATCAGCTCATCCCTAGTCTC 60

61 ACTGCTTTCCCGATCACCCTGGAACAGCCATCCTGAATCCCCCCTGGCACATCCATTTC 120

Exon2

121 GTATTTTCCCTTGGCACAGAGGCACAGCACCCGGACCTGACACCCACATAGCTTCTCTT 180

-----|-----|-----|-----|-----|-----|

Exon3

181 TGGCACACTCTATACCTCACCAGTTGTTATGATGCCACAAAGTCTCTGGACTTGT 240

1 M P Q S L W T C L 9

241 TAGTCCTTATCTCCCTAGTATGCACATTGACTGTATCTTCCAAGAAGAGCGGTGGTGGGA 300
10 V L I S L V C T L T V S S K K S G G G K 29

301 AAAGTAAACTGGAGGATGGAACACAGGGAGCAACCGGAACCCCACTACCCAGGAGGCT 360
30 S K T G G W N T G S N R N P N Y P G G Y 49

361 ACCCAGGGAATACTGGAGGCAGCTGGGGGCAACAACCTTATAATCCTAGCGGTATAACA 420
50 P G N T G G S W G Q Q P Y N P S G Y N K 69

421 AGCAATGGAAACCTCCCAAGTCCAAAACCAACATGAAGTCGGTGGCCATAGGCGCTGCTG 480
70 Q W K P P K S K T N M K S V A I G A A A 89

481 CTGGTGCTATTGGAGGCTACATGCTCGGTAATGCAGTGGGTGGTATGAGTTATCAATTCA 540
90 G A I G G Y M L G N A V G R M S Y Q F N 109

541 ACAATCCCATGGAGTCCCGTTATTATAACGACTACTATAACCAGATGCCAAATCGCGTTT 600
110 N P M E S R Y Y N D Y Y N Q M P N R V Y 129

601 ACAGGCCTATGTACAGAGGAGAGGAGTACGTGTCAGAGGACAGGTTTCGTGAGGGACTGCT 660
130 R P M Y R G E E Y V S E D R F V R D C Y 149

661 ACAATATGTCTAGTGACAGAGTACATCATAAAGCCGACTGAAGGAAAGAACACAGCGAGC 720
150 N M S V T E Y I I K P T E G K N N S E L 169

721 TAAACCGAGTTGGATACCACGGTAAAGTCCCAAATTATTCGCGAGATGTGCATCACCGAGT 780
170 N Q L D T T V K S Q I I R E M C I T E Y 189

781 ACAGGAGAGGATCGGGATTCAAAGTGTCTCTAACCCTTGGCTGATCCTTACTATCACTC 840
190 R R G S G F K V L S N P W L I L T I T L 209

841 TCTTTGTTTACTTTGTGATAGAGTGA TCAAAGGAAATATTAATAAAAAGGCCAAATGTAT 900
210 F V Y F V I E * 216

901 GTATATATAGAGAGGTATAAACCATTCTGAACTGTTCCGTCTCA 946

(d) *X. laevis* PRNP-1b

Exon1

1 GACCCGGAATTCCCGGGATGATGGGAGCTCTCACTGCTGTAGTGTGTCAGCCTCACATGA 60

61 GCTTCTCATTTGCCTTCCTGTAGCACAGCACCCGGACCTGACACCCACATAGCTTCTCTT 120

-----Exon2-----

121 TGGCAGACTCTATACCCTACCCAGGTTGTTTATGATGCCACAAAGTCTCTGGACTTGT 180
1 M P Q S L W T C L 9

181 TAGTCCTTATCTCCCTAGTATGCACATTGACTGTATCTTCCAAGAAGAGCGGTGGTGGGA 240
10 V L I S L V C T L T V S S K K S G G G K 29

-----Sequence continues as in PRNP-1a exon3-----

(e) *X. laevis* PRND-ch (chimeric)

Exon1

1 ACTGTCACACTCCCAGACAAGAGCATTCCCCTCCTCCGACACGGACGTGAGGCCCAATTA 60

61 AGGGCAGTCCGCCCCCACCCTCTCCTGCATGAAGTCTCTTCCCCTTCAGCTCATCCCTA 120

121 GTCTCACTGCTTTCCCGATCACCTGGAACAGCTATCCTGAATCCCCCCTGGCACATC 180

Exon2

```

181 CATTTCGTATTTTCCCTTGGCAGAGGACACAGCACCCGGACCTGACACCCACATACCT 240
    -----|-----|-----|-----|-----|-----|-----|
    -Exon3-|-----|-----|-----|-----|-----|-----|
241 TCTCTTTGGCACACTCTATACCCCTACCCAGGTGGCGACAGAAATGGAAAGGCAGAACGTA 300
    1                                     M E R Q N V 6
    -----|-----|-----|-----|-----|-----|-----|
301 TTCTCCTGCCTGATTCTTCTTGTGCTGATATTATATTGTGGTCTCTCTTGTCCGAGAAGA 360
    7 F S C L I L L V L I L Y C G L S C P R R 26
    -----|-----|-----|-----|-----|-----|-----|
361 TCGGGAAGTGGCATTAAAAAATATTTCAAATCAGCGACTTGAGCAGGGGAGCCAAAAAA 420
    27 S G S G I K K Y F K I S D L S R G A K K 46
    -----|-----|-----|-----|-----|-----|-----|
421 AGGTCAAAAGTGGCCATTCTCTGTCTCGGACACCTATTCTTCAGAAGTAAGGAGCTC 480
    47 R S K V A H S P V L G H L F F R S K E L 66
    -----|-----|-----|-----|-----|-----|-----|
481 GATGTGAACCTTAACTTCACCGAGGAATATGAGCTTTATACAGAGAATCTGTACAGATTG 540
    67 D V N L N F T E E Y E L Y T E N L Y R F 86
    -----|-----|-----|-----|-----|-----|-----|
541 CCGGACGGACTTTACTACCCATGGCGCTCCAGCTGAATGATGCTGCCGGCAGCGAGGAG 600
    87 P D G L Y Y P W R S Q L N D A A G T E E 106
    -----|-----|-----|-----|-----|-----|-----|
601 TTTATGAACGGGTGCCTTAAACACCACCGTAGAGAGAAACAAGGTCTGGATCTCTGGACTG 660
    107 F M N G C L N T T V E R N K V W I S G L 126
    -----|-----|-----|-----|-----|-----|-----|
661 GAAGAAGAGGACGAAAGGGGAAACCTATATGAGTGTAGGCATGCAGGTCCTACAGTTCTG 720
    127 E E E D E G E T Y M S V G M Q V L Q F L 146
    -----|-----|-----|-----|-----|-----|-----|
721 TGTATGAAAACTATGTAAGCCTACCAATGGGGCAGTGACCTGCACAGGAGGACTGTGG 780
    147 C Y E N Y V K P T N G A V T C T G G L W 166
    -----|-----|-----|-----|-----|-----|-----|
781 GTCTTCATAGGTGTCATTACCTCCTTTTCTTACTCAGAAAGGGAGTTAATGGGAACATA 840
    167 V F I G V I H L L F F T Q K G S * 182
    -----|-----|-----|-----|-----|-----|-----|
841 AAGCCTGAATTCTGTATTCTTATATTGAACTaaatgtacccgccagaggttcagcagc 900
    901 tct 903

```

(f) *X. laevis* PRND-a, PRND-b (non-chimeric)

Exon1

1	aaaggcccagatttaggcaggactcattacatctctacatgtttggcactcaggcccgga	60
61	ctgacaatctgcccggttcggggcccgccgtctgcctgcatgacactgtcccgctctgcag	120
121	ccccctcccacctccacagagcagagtaatatcacagagcagaggtgagggcaggactgc	180
181	agaaaattttgaactagtgggggaggacgcccacctgacgccatcacacactggtctattt	240
241	acaaacagtggactggagacctgtattattactgggggaagctggagagaagtattttt	300
301	tcagggcaaagcagagttggtctGTAAAGTGGACTGGAGAGTGGGCCAGCAGTGTCTCAT	360

-----|-----|-----|-----**Exon2**-----|-----|

361	CAGTCTGAAGAAGGAAGAAACAGCAGCAAGGCTCTGAG	420
1		M E R Q 4

-----|-----|-----|-----|-----|-----|

421	AGAACGTATTCTCCTGCCTGATTCTTCTTGTGCTGATATTATATTGGTCTCTCTTGTG	480
5	N V F S C L I L L V L I L Y C G L S C P	24

-----Sequence continues as in *PRND-ch* exon3-----

(g) *X. laevis* PRNP-2a

Exon1

1 GGCTGGACCGGTCCGGATTCCCGGGATGATCAGCCTGGCACAGCCGCTCCTGAACCTCCCC 60

Exon2

61 CTGTACAGGCACAGTCTTTCTGAATCCCCCTGGCACAGAGGGCCTGACAGCCAGCTAGC 120

Exon3

121 TTCTCTTTGGGCTTACTTATCTCTCACACAGGTTGGCCAAGATGCCAAGAAGTCTCTGG 180
1 M P R S L W 6

181 ACTTGTITAGTCCTTATCTCCCTAGTGTGCACATTGACTGTATCTTCCAAGAAGAGTGGT 240
7 T C L V L I S L V C T L T V S S K K S G 26

241 AGTGGGAAAAGCAAAACCGGAGGCTGGAACAATGGGAACACTGGGAACACCGGGAACACT 300
27 S G K S K T G G W N N G N T G N T G N T 46

301 GGGAAACAACCGGAACCCCAACTATCCAGGAGGCTATGGCTGGAACACAGGGAACACAGGG 360
47 G N N R N P N Y P G G Y G W N T G N T G 66

361 AACACTGGAGGCAGTTGGGGGCAACAACCTTATAATCCTAGCGGAGGAAGCAATTTCAAC 420
67 N T G G S W G Q Q P Y N P S G G S N F N 86

421 AACAAAGCAATGGAACCTCCCAAGTCCAAAACCAATATGAAGGCCGTGGCCGTAGGCGCT 480
87 N K Q W K P P K S K T N M K A V A V G A 106

481 GCTGCTGGTGTCTATTGGAGGCTACATGCTTGGTAATGCAGTGGGTCTGATGAATCATCAT 540
107 A A G A I G G Y M L G N A V G R M N H H 126

541 TTCGACAATCCCATGGAATCCCGTTATTATAACGACTACTACAACCAGATGCCAGACCGG 600
127 F D N P M E S R Y Y N D Y Y N Q M P D R 146

601 GTTTACAGGCCAATGTACAAAACCGAGGAGTACGTGTCTGAAGATAGGTTCTGTCACGGAT 660
147 V Y R P M Y K T E E Y V S E D R F V T D 166

661 TGCTACAATATGTCAGTAACAGAGTACATCATCAAGCCATCCGAAGGGAAGAATGGCAGC 720
167 C Y N M S V T E Y I I K P S E G K N G S 186

721 GATGTAAACAGTTGGATACCGTGGTGAAATCCAAATTTATTCGCGAGATGTGCATCACT 780
187 D V N Q L D T V V K S K I I R E M C I T 206

781 GAATACAGGAGAGGATCAGGATTCAAAGTGCTTTCTAACCCTATGGCTGATCCTCACTATC 840
207 E Y R R G S G F K V L S N P W L I L T I 226

841 ACTCTCTTTGTTTACTTTGTGATAGAGTGAaccagaggggaatcgaaatccaaaggccaatt 900
227 T L F V Y F V I E * 235

901 gtatgtatatagagagtatcaaccaattctggactgtctcgtctcatgccaatatgaca 960
961 ctgttgggtgctatgttaatccagccagcttctaccatcagtaagcaaccatggatct 1020
1021 tgcctcaataccaacctgaattcccaacttctgcgtcaactacaacttctgcttgacaaca 1080
1081 catgtttctgacattgcaaatggcctttatgtatcagtgcattatataataagagatct 1140
1141 tcatt 1145

(h) *X. laevis* PRNP-2b

Exon1

1 gtcgggaattctccggatcacgagcttctcatttgccttcctgtagggcctgacagccag 60

Exon2

61 CTAGCTTCTCTTTGGGCTTACTTATCTCTCACACAGGTTGGCCAAGATGCCAAGAAGTC 120
1 M P R S L 5

Exon3

121 TCTGGACTTGTITAGTCCTTATCTCCCTAGTGTGCACATTGACTGTATCTTCCAAGAAGA 180
6 W T C L V L I S L V C T L T V S S K K S 25

-----Sequence continues as in PRNP-2a exon3-----

Appendix 6. Nucleotide and amino acid sequences of cDNAs for transcripts of (a) *M. domestica* PRNP, (b) *M. domestica* PRND. Exons are numbered; non-coding exons are enclosed by boxes; ORFs for PrP1 and PrP2 are highlighted in grey; 3'UTRs of PRNP variant 1 and variant 3 and PRND variant 1 are underlined. CPE motif at the 3'UTR of PRNP is in italic and boxed.

(a) *M. domestica* PRNP

Exon1
 AGTTAAGACCAGACAACGACCAGGCTGCTGCTGCCTTCTGCGCTCTACCAGAA
 Exon2
 TTCCTAGCTGCTCTGCTGAAAGACTCTACGTAGTTTGGATCCAAACGAAGATCAGCTACC

ATGGGGAAAATCCACTTGGGATACTGGTTCTTGGCTCTTTTTATTATGACCTGGAGTGT 60
 ATGGGGAAAATCCACTTGGGATACTGGTTCTTGGCTCTTTTTATTATGACCTGGAGTGT 60
 M G K I H L G Y W F L A L F I M T W S D

CTGACTCTCTGTAAGAAACCGAAGCCAAGACCTGGAGGGGGATGGAACAGTGGAGGCAAC 120
 CTGACTCTCTGTAAGAAACCGAAGCCAAGACCTGGAGGGGGATGGAACAGTGGAGGCAAC 120
 L T L C K K P K P R P G G G W N S G G N

CGCTATCCTGGCCAGTCTGGAGGCTGGGGACATCCTCAGGGAGGTGGCACC AATTGGGGT 180
 CGCTATCCTGGCCAGTCTGGAGGCTGGGGACATCCTCAGGGAGGTGGCACC AATTGGGGT 180
 R Y P G Q S G G W G H P Q G G G T N W G

CAACCCCATGCTGGGGGTTC AATTGGGGTCAGCCACGTCCTGGGGGTTC AATTGGGGC 240
 CAACCCCATGCTGGGGGTTC AATTGGGGTCAGCCACGTCCTGGGGGTTC AATTGGGGC 240
 Q P H A G G S N W G Q P R P G G S N W G

CAGCCCCATCCAGGGGGTTC AACTGGGGACAGCCCCATCCAGGGGGTTC AACTGGGGT 300
 CAGCCCCATCCAGGGGGTTC AACTGGGGACAGCCCCATCCAGGGGGTTC AACTGGGGT 300
 Q P H P G G S N W G Q P H P G G S N W G

CAAGCCGGTAGCAGCTACAACCAGAAGTGGAACACAGATAAACCCAAAACCAACATGAAA 360
 CAAGCCGGTAGCAGCTACAACCAGAAGTGGAACACAGATAAACCCAAAACCAACATGAAA 360
 Q A G S S Y N Q K W K P D K P K T N M K

CATGTGGCTGGGGCTGCTGCTGCTGGGGCAGTTGTAGGAGGGCTTGGTGGATACATGCTG 420
 CATGTGGCTGGGGCTGCTGCTGCTGGGGCAGTTGTAGGAGGGCTTGGTGGATACATGCTG 420
 H V A G A A A A G A V V G G L G G Y M L

GGGAGTGCCATGAGCAGGCCCATCATGCACTTTGGCAATGATTACGAAGACCGTTACTAT 480
 GGGAGTGCCATGAGCAGGCCCATCATGCACTTTGGCAGCGACTACGAAGACCGTTACTAT 480
 G S A M S R P I M H F G S D Y E D R Y Y

CGTGA AAAACCAATATCGCTATCCCAACCAAGTCATGTATCGGCCCATTGATCAGTACAAC 540
 CGTGA AAAACCAATATCGCTATCCCAACCAAGTCATGTATCGGCCCATTGATCAGTACAAC 540
 R E N Q Y R Y P N Q V M Y R P I D Q Y N

AACCAGAACAATTTTGTGCATGACTGTGTCAACATAACGGTGAAGCAACACACCACCACC 600
 AACCAGAACAATTTTGTGCATGACTGTGTCAACATAACGGTGAAGCAACACACCACCACC 600
 N Q N N F V H D C V N I T V K Q H T T T

ACCACTACCAAAGGGGAGAACTTCACTGAAACGGATATCAAGATCATGGAGCGAGTGGTG 660
 ACCACTACCAAAGGGGAGAACTTCACTGAAACGGATATCAAGATCATGGAGCGAGTGGTG 660
 T T T K G E N F T E T D I K I M E R V V

GAACAGATGTGCATCACGCAGTACCAGAACGAGTACCGCAGTGCTTACAGCGTGGCGTTC 720
 GAACAGATGTGCATCACGCAGTACCAGAACGAGTACCGCAGTGCTTACAGCGTGGCGTTC 720
 E Q M C I T Q Y Q N E Y R S A Y S V A F

TTCTCTGCCCCACCTGTGACCCTCCTCCTCCTCAGTTTCCTTATTTTCCTGATTGTGAGC 780
 F S A P P V T L L L L S F L I F L I V S

TTCTCTGCCCCACCTGTGACCCTCCTCCTCCTCAGTTTCCTTATTTTCCTGATTATCCCA 780
 F S A P P V T L L L L S F L I F L I I P

TAA GAAGCCTACCAATGTTTACTCTCTTCATGTTTCTTCTTAATCTTTGCAGAGAAGG PrP1

AGGTCCTTCTGTCTGCAAGGGCAGCCCAATAGCAGCAATTTCTCATTTCTATGTTTATC
TGTCCCCCATAGGTTAAGGCACATAAGTACTGGTGAATGTACAGTAGACCCCTAGATGC
CAGGCCACCACTCTTCCCCGAACCATTTTGATCATGCATCCATCAGGGCAATGCCATAC
TTGTCAGTATCCTTTACAAGAGAGGAGACCATTAAAGTAACTTCTGGTCCATCAAGACACT
TCTATAGTATAGCAGATTAAAGGCCAAAACAGAAATGATTTCAAACCTACATTTTCCAAATA
GACACAACCATGGGCCCTTTTGCTTCTGAAATGCCACCTAAATCTTTCTCCCTGCTTGT
ATAGTCAATTAATGAGTAGATAAAGAATTAGCTAATTTAGAGCCCCATCTCTTCTTTGGC
TTTACCAGCTGTGATATCATAGCCAGTTAAATATCTTTAGGAACTCATTATACATTTTC
AATACATCCTGATGCATTTTCTCTTCAAATAGAAATTTTCATCATTAGGAAAGAAAGAC
TAGAGACCATCTAGGCAACTATAGAAATCATCCGACCATGTGGGTCCAGGCTCGAGCTG
CTGAATCAAGCGTAGCTACACCACAGAGAAGTGGGACTACTGTCTGAATTTTATAAAT
GGACACACTCAGTAGAGGCTGCACCAAAATATATCCGGCCAGGGCTGGGTCTATTGT
TCAACAGTGGAAATGAGCATCGAGCATCATTGCGTGAGCTTTATCCTGTGCTACAGAGACT
GTTCTTCTGGGTAACCTACAATTTGGAAAGTAGTATTACTATGGGTATGATTTGTCTATCT
CAGACCATTTTGGTGTGTACATATTGTTCAATACCAGTGTAACCTATTTCTACCAGAGCA
TTCCACCCAGGAATATGAAATGCAAAGCAGATACTCCCGGATAACGAAGGAATCCCTCT
CCAATGTTAGCAAATAATCCAAGACTGGCAGGCAGCACTTGGCTAGGCCTCTTTGCCTGG
AAGCCACAAAGGTGCCAGCATTTAGAGTTACAGAGGAGGCTGGTTAGCTCAGAGGA
ATCAGCAACTGGTTAAATTGCAATTTTAGGAATTAATGAGATAGTTTGGTGAGTTTATAA
GCAAAAGAAGCACTTCTTTTTTTCATGTCTCAAGATAAAGCAGTACCTAACTATACTGGA
TGAGCCCTTAAAGTCAGATTAGAAGAAGATCATTCAGATGGAGGAAGCCAGCCTGTTGA
TTTTTGGAATTTAAATCCAATCTTCAAAACAATGTCTCAGAAAAGGAGGAGAGGGGAG
GAGAAGACGGGGAGATGATCTGTCTGTATAAGTGATTTTGTATTGAGTTGTGAAAAAGT
TGTTCTCATAGATAAATATTACTTTTTTCTTGTCACTATTTTGCATTGTGATGGGCTT
GCAGTCTTTGTAATCAGTAAATATCAGATATGATAAACTAGAAATCTGCTCAGAACAAAT

TCACCCAGAGTAGTATCCAATGTATTCTGTGTCTTTCCAAG GATGCCCATTCCTGGAAG
D A H S V E

CAATATCCTGA GGCCTACGGCACCACATGCTGAATGGACTCCTGCTTGAGTCCCCAGTGC PrP2

A I S *
ACCAGGGAAAGTAGTCTTTGAAATTTGCATGCACTTAGTAATATAAAGATG TTTTATA GA
TTTGTAACTTTGCACGTATTTGTTTGTATGATTAAAAATTTATAAATGTTTAATATC

(b) *M. domestica* PRND

Exon1 (common)
AGTTAAGACCAGACAACGACCAGGCTGCTGCTGCCTTCTGCGCTCTAC

Exon2

CAGAATTCCTAGCTGCTCTGCTGAAAGACTCTACGTAGTTTGGATCCAAACGAAG | GTTCATAGGATTTCT
ATTCTGACAAAGACCTCCTTGGATCAGATAGTCTGACATCTTCATTTACATATAAAGAAATGGAGACCC

Exon3

AGAGAG | AGCAAGACAGAAGCTGACCTCTATAATCAACAAGTAGACAAAGGGACAACAAACAACCTTTTTT

Exon4

TG | GATTTGTGGGAGGAGTGATGGATCAAGGTTGAGTGGTATTTCTGATCCCTGGATACTCCTTCTCTTT
CTTCCAAGACTTTTCATGGGCTACTATTGCCAGATAATGGTCATCAGAACTGATGGGATGTTCACTACCAT

Exon5

CCCCACCTTGCTCTCTGTCAGTTTAAAGGAGGAAGAGACAGAGAAACAG | GGATGCGTTTCTCCTCAACAC
ATTCTGAGGTCCTTACAAATGTACAGTCTTCAAGAGGCAGTTTTCATGAGAATGTCCATCTGGACAAA
TTGAGAGTGACCCGGAGACTTTTCAAGATTGGTACTCCAGGAAGTGAATAATATCATACCATATTTGGAGA
GAAGTATCAGTGCCACCTCCTTCTCTGCTATATAGTTGTTGAGTACAACCTGATAAAGAACCAGTTCTGT

Exon6

GGGGAGAATTTAACTTCTTCTGGCTTCCAAAGCAACCTTTTCCAAACAACGAATAAAAG GCTTTTAATACA

ATGAGGAGACATTTGGGAATTTGTTGGATAGCCATCTTCTTTGCCCTGCTGTTTAGTGATCTCTCCTTGG 70
M R R H L G I C W I A I F F A L L F S D L S L

TCAAGGCCAAAACCTACAAGACAGAGAAATAAGTCAAACAGAAAAGGTTTGCAAACCTAATAGAACCAACCC 140
V K A K T T R Q R N K S N R K G L Q T N R T N P

AACCACAGTGACGCCCTCAGAGAACTCCAGGGGACTTTTCATCAGGAATGGCAGAAAGCTTGTCTATTGAC 210
T T V Q P S E K L Q G T F I R N G R K L V I D

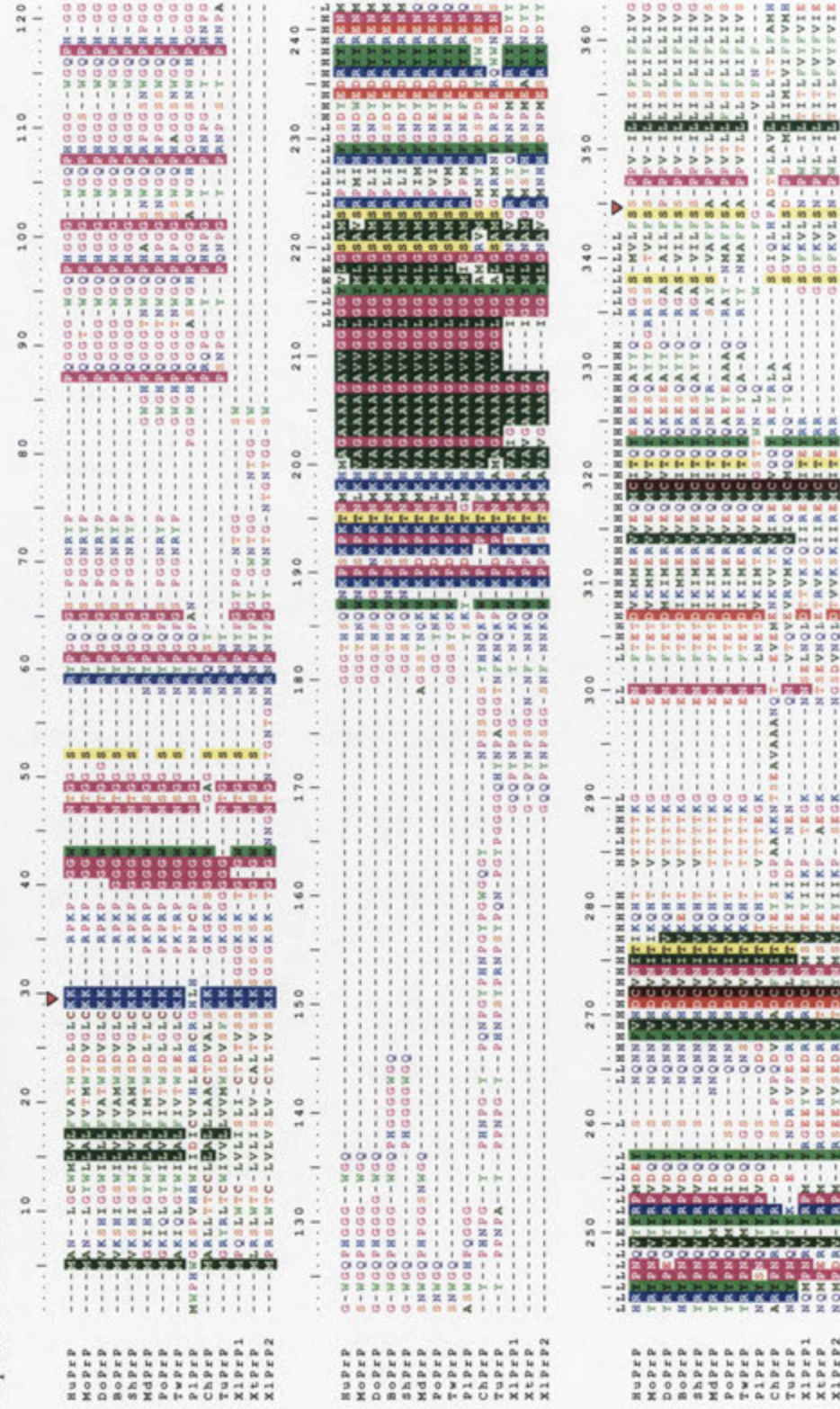
TTTGGGGAGGAGGGCAACAGTTACTATGCGACCCATTACAGCCTGTTCCAGATGAAATCCATTATGCTG 280
F G E E G N S Y Y A T H Y S L F P D E I H Y A

GGTGTGCCGAGTCCAATGTAACAAAGGAGGTCTTCATCAGCAACTGTGTGAATGCTACCCGCGTAATCAA 350
G C A E S N V T K E V F I S N C V N A T R V I N

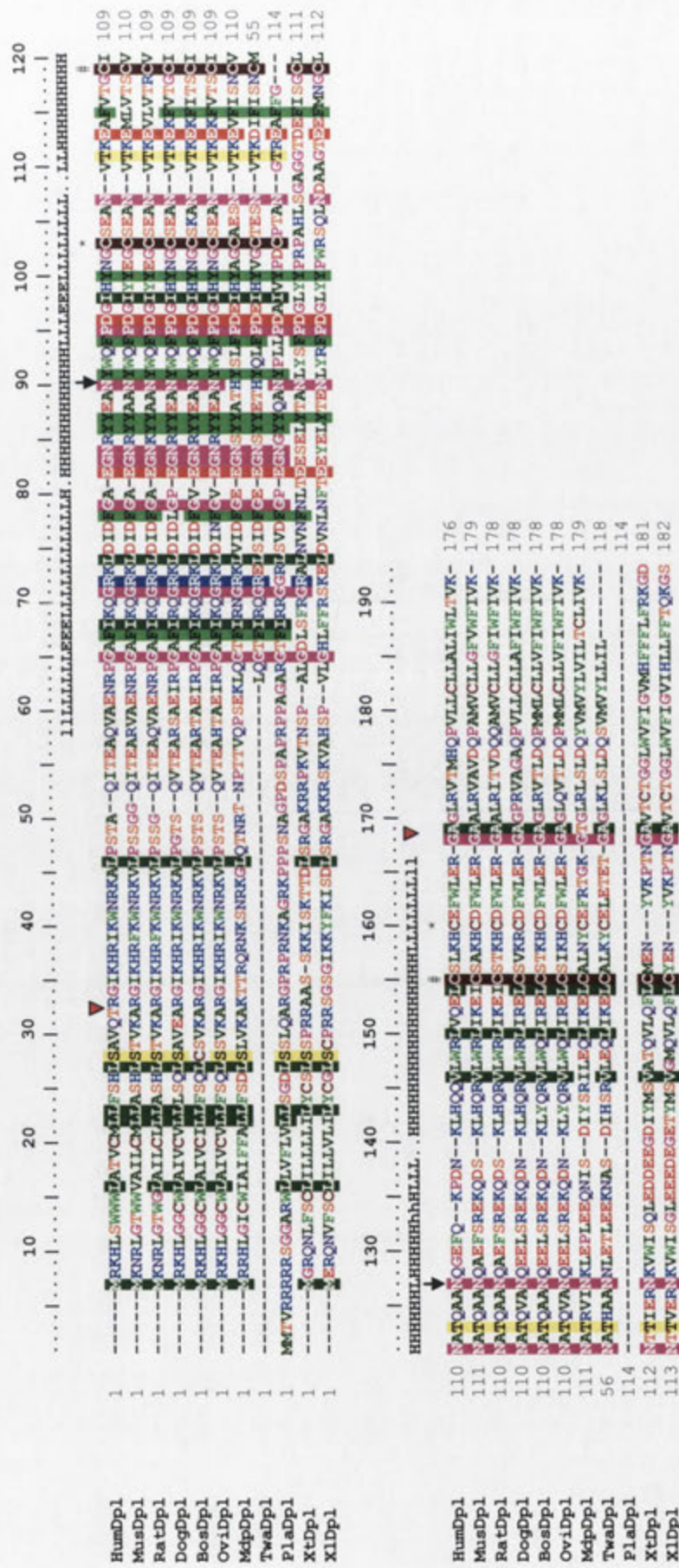
CAAGTTGGAGCCCCTAGAAGAACAAAACATAAGTGATATTTATTCTAGAATACTGGAACAATTAATAAAG 420
K L E P L E E Q N I S D I Y S R I L E Q L I K

GAATTATGTGCCCTTAACTATTGTGAATTTAGGACAGGGAAAGGAACTGGCCTTCGGCTCTCTTTAGACC 490
E L C A L N Y C E F R T G K G T G L R L S L D
AATATGTTATGGTCTACCTGGTGATTCTGACTTGCTTGATAGTGAAATAAATAAGCACAGAGCCTGAC
Q Y V M V Y L V I L T C L I V K *
AAGGAGAGATGGATAAGCCATCCAGCCACAGGCAAAGACCAAAGTTATCTCAGGCTCCCTTCC
TGCTGTAACAAAACCATTACCCCTTAAAAAAAACCTTCCAAAACCTCTCTGACAAAAGGTTTAAAT
TTCCCAAATATATAAGAACTAAGTCAAATTTATAAAAATTCAAGCCATTCCCCAATCAACTA
GTGGTCAAGGGATTTGAATAGGCAGTTCTCAGATGAAGAAATCAAACTATCTATAATCACAT
GAAAAAATGTTCTAAATCCCTCCTGATAAGAGAAATGCAAGTAAAAACAACCTCTGAGGAACCA
CCTCACACCTAGCAGATTGACTAATATGTGAACAAAAGAATGTGATAAATGTTGGAGGGGATG
TGGAAAAATTGGGACACTAATGCATTGCTGGTGGAGTTGTGAATTAATCCAACCATTCTGAAG
GGCAATTTGGAATTATGCCCAAAGGGCTTTAAAAGACTGTCTGCCCTTTGATCCAATCATACC
ACTTCTGGGTTTGTATCTAAAAGAGATAATAAAGAAAAAGACTTGTACAAAAATATTTATAGC
TGCACTCTTTGTGGTAGCAAAAAACTGGAAAATGAGGGAGTGCCATCGAATGGCTGAACAAA
TTGTGGTATCTGATGGTGATGGAATACTATTGTGCTGAAAGGAATGATGAATTGGAGGGATTCT
CATGTGAAGTGTAAAGACCTCCAGCAATTGATGCAGAGTGGAATGAGCAGAACCAGGAGAACA
TTATACACAGAGACTGATATACTGTGGTACAATTGAATGTAACAACTTTTCTACTGGCAGCA
ATGCAATGATCCAGGACAATTATGAGGGACTTATGGGAAAGAATGCTATCCACATGCAGAGAA
AGAACTATGGGAGTAGAAATGCAGAAGAAAAACATATGATCGATCACATGATTCAATGGGGAT
ATGATTAGGGTATTGGCTTTAAAAAATCATTCTACTACAAATATGAATAGCATGGAAATAGGC
TTTGAATGATGATACATGTATAACCCAATGGAATTGCTTGTCAACTCCAGGAGGTCGGGAGGG
AAGAGTGGTGGGAAAAATCATGAATCATGTAACCATAGAAAAATATTCTAAATAAATAGGTTT
TTTTAACTA

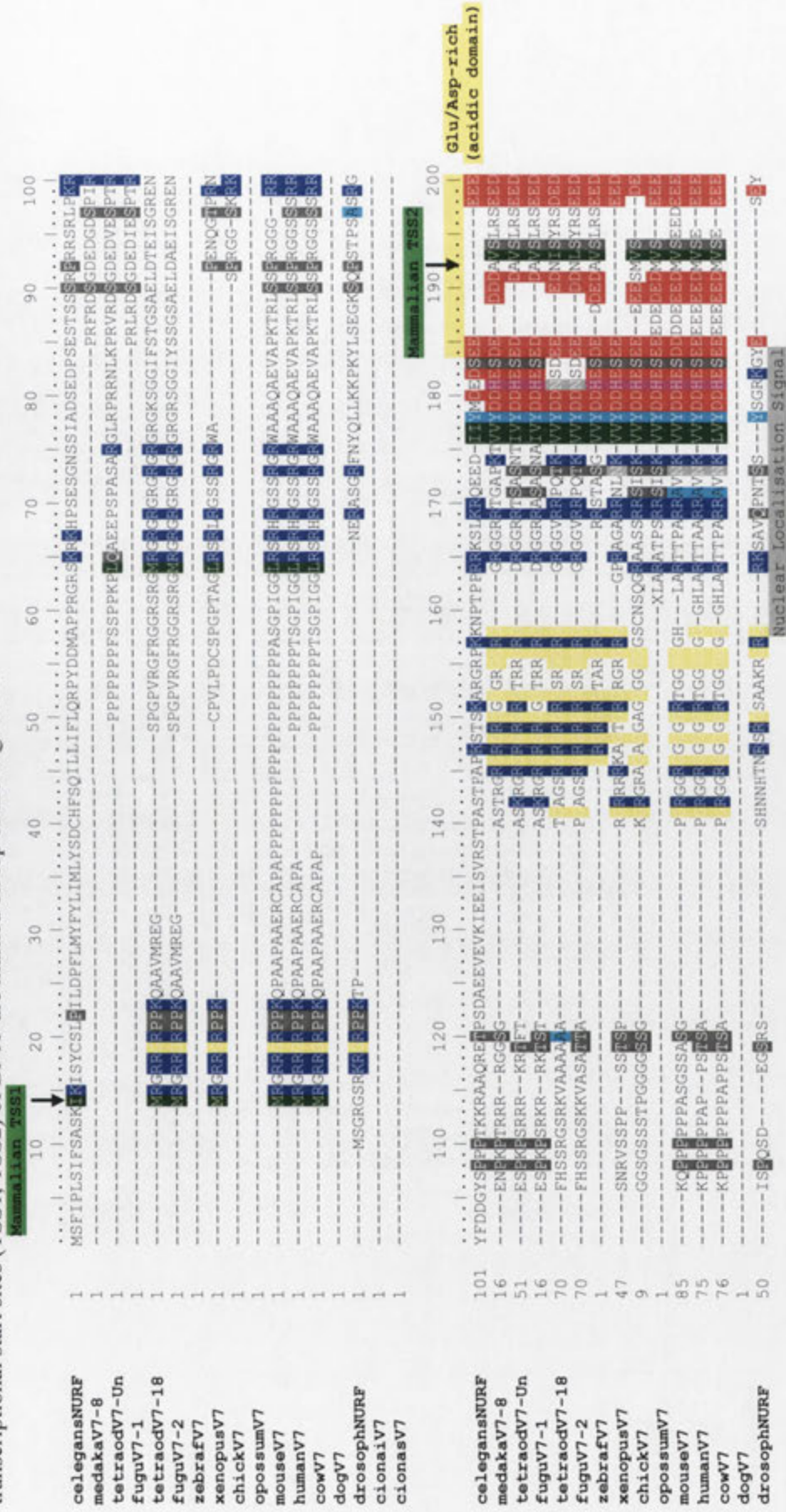
Appendix 7. Multiple sequence alignment of PrP sequences. The secondary structural units, predicted based on mouse PrP protein structure, are shown above the sequences (L- loop, E- sheet, H- helix). Cleavage site is represented by ▴. The C-terminal domain region shows a number of conserved residues among all the species compared. The most highly conserved region corresponds to the hydrophobic region numbered between 200 and 220. The N-terminal repeats show the maximum number of variations.



Appendix 8. Multiple sequence alignment of Dpl sequences. The secondary structural units, predicted based on mouse Dpl protein structure, are shown above the sequences (L-loop, E-sheet, H-helix). Cleavage site is represented by ▴. The C-terminal domain region shows fewer conserved residues compared with PrP. The most highly conserved regions correspond to those including the cysteines involved in disulphide bridge formation and the N-glycosylation sites. The Cys residues involved in disulphide bridge formation are indicated (*, #). Note the absence of the second disulphide bridge in *Xenopus* species. The Asn residues involved in Kink formation in helices 1 and 2 are indicated by arrows.



Appendix 10. Multiple sequence alignment of eukaryotic BPTF sequences. Analysed orthologous sequences correspond to *BPTF* v7 containing unspliced Regions 1 and 2. Selected species are: invertebrates *C.elegans* (worm), *Ciona intestinalis* and *Ciona savignyi* (sea squirts), and drosophila (insect), vertebrates Medaka, *Tetraodon*, fugu, and zebrafish (fish), *Xenopus tropicalis* (frog), chicken, opossum, cow, dog, mouse, and human. Numbering on top correspond to alignment position; numbering on the left corresponds to a first amino acid residue in line; (-) denotes gaps in alignment. The functional domains and both transcriptional start sites (TSS1, TSS2) of BPTF are marked on top of the alignment.



Appendices

188	colegansNURF	SSDPEFLNLEIQVQEEER	LNLTIKIEK	210	220	230	240	250	260	270	280	290	300
84	medakaV7-8	FEPEPSEEDGALKES	CLEDDVLDEE	210	220	230	240	250	260	270	280	290	300
117	tetraodV7-Un	FEPEPSEEDGALKES	CMDDRLNEERE	210	220	230	240	250	260	270	280	290	300
82	fuguV7-1	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
138	tetraodV7-18	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
138	fuguV7-2	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
44	zebrafV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
105	xenopusV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
80	chickV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
37	opossumV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
159	mouseV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
147	humanV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
150	cowV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
103	dogV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
103	drosophNURF	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
1	clonaiV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
1	clonaiV7	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300
1	Clustal Consensus	FEPEPSEEDGALKES	CMDDGLGDEE	210	220	230	240	250	260	270	280	290	300

247	colegansNURF	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
166	medakaV7-8	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
202	tetraodV7-Un	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
165	fuguV7-1	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
226	tetraodV7-18	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
223	fuguV7-2	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
125	zebrafV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
196	xenopusV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
169	chickV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
125	opossumV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
247	mouseV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
235	humanV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
230	cowV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
35	dogV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
183	drosophNURF	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
17	clonaiV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
13	clonaiV7	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400
6	Clustal Consensus	FEPEPSEEDGALKES	CMDDGLGDEE	310	320	330	340	350	360	370	380	390	400

	410	420	430	440	450	460	470	480	490	500
<i>celegans</i> NURF	MTHTLTATLQTH	YPTADASTDA	NVNYPVGV	NIDRDEFFNKHVRLV						
<i>medaka</i> V7-8	TLVILVTMPVLRV	ESDKY	HHVLPVLCVGC	ESDKY						
<i>tetraodon</i> V7-Un	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>fugu</i> V7-1	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>tetraodon</i> V7-18	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>fugu</i> V7-2	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>zebrafish</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>xenopus</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>chick</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>mouse</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>human</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>cow</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>dog</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>drosophila</i> NURF	SLTIDSTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>ciona</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>clonal</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
<i>clonal</i> V7	TLVIDVTMPVLRV	ESDRE	HHVLPVMDQDPA	ESDKY						
Clustal Consensus	46									

	510	520	530	540	550	560	570	580	590	600
<i>celegans</i> NURF	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>medaka</i> V7-8	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>tetraodon</i> V7-Un	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>fugu</i> V7-1	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>tetraodon</i> V7-18	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>fugu</i> V7-2	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>zebrafish</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>xenopus</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>chick</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>opossum</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>mouse</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>human</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>cow</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>dog</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>drosophila</i> NURF	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>ciona</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>clonal</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
<i>clonal</i> V7	SRVETCTCS	YHLVCKVPP	SEVPDEMQ	SEVCEV	LP--P	DAV	PEPRQPI	YDRHQRKWF	YDRHQRKWF	YDRHQRKWF
Clustal Consensus	81									

		610	620	630	640	650	660	670	680	690	700
520	colegansNURF	FOVGLKLR	FKK	CHLRI	DELEO	LVVTS	EE	ATVKECLIGY	FAP	TFPIYLH	SMRMASILR
428	medakaV7-8	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
452	tetraodV7-Un	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
427	fuguV7-1	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
488	tetraodV7-18	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
485	fuguV7-2	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
387	zabrafV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
457	chickV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
432	opossumV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
388	mouseV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
510	humanV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
498	humanV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
493	cowV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
298	dogV7	QVGLMEL	DKYET	HATPE	MEVOA	QITER	TKR	STFT	TPPG	GSAEN	ERT
443	drosophNURF	SKKLISL	SL	ABELE	TRHSQ	TRR	DEIERQ	ALTE	TLT	ERH	HEKRSVLEIQ
292	cionaV7	EDDY	RSK	CKHE	ED	VGH	GNH	CBRNH	CHRO	CH	NOEKKKEEEE
275	cionaV7	EDDY	RSK	CKHE	ED	VGH	GNH	CBRNH	CHRO	CH	NOEKKKEEEE
119	Clustal Consensus	..	..	..	..	..	..	..	..	..	..
		610	620	630	640	650	660	670	680	690	700
605	colegansNURF	CAE	---	---	---	---	---	---	---	---	---
517	medakaV7-8	PSL	---	---	---	---	---	---	---	---	---
541	tetraodV7-Un	TVB	---	---	---	---	---	---	---	---	---
521	fuguV7-1	LSLS	DR	LI	DS	GL	DES	GL	DES	GL	DES
579	tetraodV7-18	DV	---	---	---	---	---	---	---	---	---
582	fuguV7-2	DV	---	---	---	---	---	---	---	---	---
482	zabrafV7	PSG	---	---	---	---	---	---	---	---	---
539	xenopusV7	ATV	---	---	---	---	---	---	---	---	---
525	chickV7	DSH	---	---	---	---	---	---	---	---	---
483	opossumV7	DV	---	---	---	---	---	---	---	---	---
603	mouseV7	DV	---	---	---	---	---	---	---	---	---
589	humanV7	DV	---	---	---	---	---	---	---	---	---
585	cowV7	DV	---	---	---	---	---	---	---	---	---
390	dogV7	DV	---	---	---	---	---	---	---	---	---
514	drosophNURF	PKG	---	---	---	---	---	---	---	---	---
384	cionaV7	KT	---	---	---	---	---	---	---	---	---
367	cionaV7	PI	---	---	---	---	---	---	---	---	---
		710	720	730	740	750	760	770	780	790	800
605	colegansNURF	---	---	---	---	---	---	---	---	---	---
517	medakaV7-8	---	---	---	---	---	---	---	---	---	---
541	tetraodV7-Un	---	---	---	---	---	---	---	---	---	---
521	fuguV7-1	---	---	---	---	---	---	---	---	---	---
579	tetraodV7-18	---	---	---	---	---	---	---	---	---	---
582	fuguV7-2	---	---	---	---	---	---	---	---	---	---
482	zabrafV7	---	---	---	---	---	---	---	---	---	---
539	xenopusV7	---	---	---	---	---	---	---	---	---	---
525	chickV7	---	---	---	---	---	---	---	---	---	---
483	opossumV7	---	---	---	---	---	---	---	---	---	---
603	mouseV7	---	---	---	---	---	---	---	---	---	---
589	humanV7	---	---	---	---	---	---	---	---	---	---
585	cowV7	---	---	---	---	---	---	---	---	---	---
390	dogV7	---	---	---	---	---	---	---	---	---	---
514	drosophNURF	---	---	---	---	---	---	---	---	---	---
384	cionaV7	---	---	---	---	---	---	---	---	---	---
367	cionaV7	---	---	---	---	---	---	---	---	---	---
		710	720	730	740	750	760	770	780	790	800
605	colegansNURF	---	---	---	---	---	---	---	---	---	---
517	medakaV7-8	---	---	---	---	---	---	---	---	---	---
541	tetraodV7-Un	---	---	---	---	---	---	---	---	---	---
521	fuguV7-1	---	---	---	---	---	---	---	---	---	---
579	tetraodV7-18	---	---	---	---	---	---	---	---	---	---
582	fuguV7-2	---	---	---	---	---	---	---	---	---	---
482	zabrafV7	---	---	---	---	---	---	---	---	---	---
539	xenopusV7	---	---	---	---	---	---	---	---	---	---
525	chickV7	---	---	---	---	---	---	---	---	---	---
483	opossumV7	---	---	---	---	---	---	---	---	---	---
603	mouseV7	---	---	---	---	---	---	---	---	---	---
589	humanV7	---	---	---	---	---	---	---	---	---	---
585	cowV7	---	---	---	---	---	---	---	---	---	---
390	dogV7	---	---	---	---	---	---	---	---	---	---
514	drosophNURF	---	---	---	---	---	---	---	---	---	---
384	cionaV7	---	---	---	---	---	---	---	---	---	---
367	cionaV7	---	---	---	---	---	---	---	---	---	---

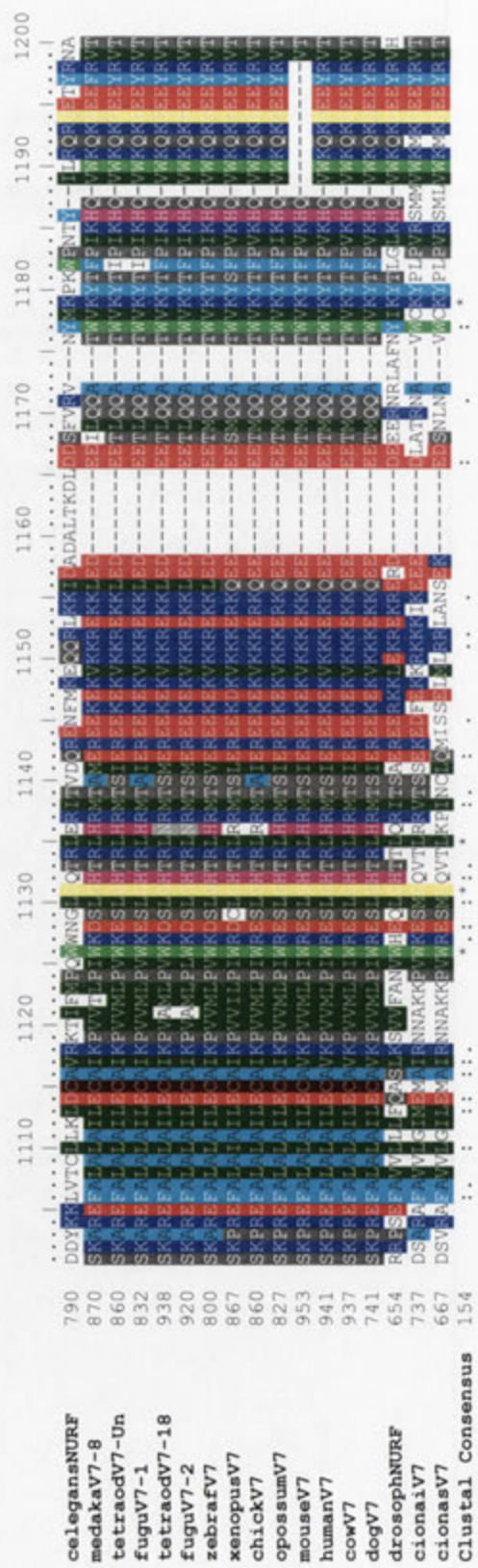
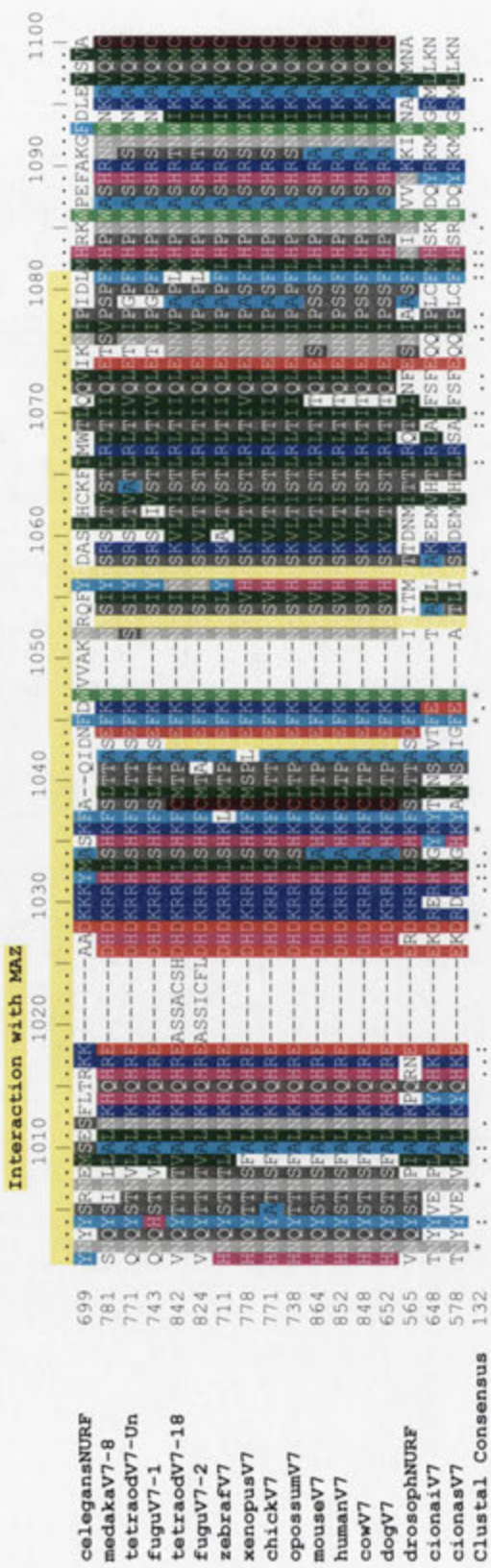
	810	820	830	840	850	860	870	880	890	900
celegansNURF	DVTERKIRRA	Q	EAKRAEAKQ	ESTSINAE	RVREP	SSPK	SSPK	SSPK	SSPK	SSPK
medakaV7-8	EHSDIVH	EN	N	EQORDE	SOPT	OKOLATIS	E	KNDRLHPE	DPVPI	SSQSF
tetraodV7-Un	NVSCQLSM	L	CKKUIY	YITDVTNN	EQQPDW	QLQT	OKPATIS	E	SSROLKE	DPVPI
fuguV7-1	FKDLKIC	YC	YS	SVVV	TAQEGD	GSLL	DGFMLOT	AGAPQS	TT	TSSSIAT
tetraodV7-18	ITDS	SPSLDA								
fuguV7-2	KGST	GV	EL	IS	SE	GA	HGEGD	SVLEALA	QAT	ES
zebrafV7	KSS	AC	GVSS	NI	PA	TA	RED	SS	CT	PODS
xenopusV7	KV	SE	EL	ST	GOMIPT	SS	SAAP	ADV	NSS	SEL
chickV7	KV	PE	POV	PS	TKLND	GN	TA	VNF	FF	Y
opossumV7	KV	SE	EL	PS	N	ET	SSNTS	TT	ASS	PNIT
mouseV7	KV	SE	EL	PS	N	ET	SSNTS	TT	ASS	PNIT
humanV7	KV	SE	EL	PS	N	ET	SSNTS	TT	ASS	PNIT
cowV7	KV	SE	EL	PS	N	ET	SSNTS	TT	ASS	PNIT
dogV7	KV	SE	EL	PS	N	ET	SSNTS	TT	ASS	PNIT
drosophNURF	TDGAPT	Q	VAP	VVPTDV	TDVAPT	GNVAP	SP	PTVV	PT	SV
clonaiV7	D	TS	LEN	KM	VP	DF	DT	SS	EL	K
cionasV7	D	TS	LEN	KM	VP	DF	DT	SS	EL	K

Region 1, Exon 5 Region 1, Exon 6

	910	920	930	940	950	960	970	980	990	1000
celegansNURF	FWS	GTQEE	VE	EF	Y	IS	NF	DAPS		
medakaV7-8	HE	QA	TPE	AP	EP	PP	VE	PE	PS	SG
tetraodV7-Un	RQCE	EQ	CS	PT	PS	CG	EL	OT		
fuguV7-1	VYH	C	EM	SR	CF	CF	CF	CF		
tetraodV7-18	ANG	QA	Q	GK	SS	TR	AV	TR	LN	PE
fuguV7-2	SNG	GANQ	GK	SA	TR	AV	TR	LN	PE	SL
zebrafV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
xenopusV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
chickV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
opossumV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
mouseV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
humanV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
cowV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
dogV7	ES	FT	K	TP	STR	IT	RL	NP	PS	KL
drosophNURF	TQ	NIT	PT	T	N	L	P	F	G	M
clonaiV7	EL	R	L	Q	F	A	P	M	K	P
cionasV7	EL	R	L	Q	F	A	P	M	K	P

Clustal Consensus 127

Appendices



885	celegansNURF	KGSM	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
959	medakaV7-8	---	...	...	...	...	...	...	...	...	...	...
949	tetraodV7-Un	---	...	...	...	...	...	...	...	...	...	...
921	fuguV7-1	---	...	...	...	...	...	...	...	...	...	...
1027	tetraodV7-18	---	...	...	...	...	...	...	...	...	...	...
1009	fuguV7-2	---	...	...	...	...	...	...	...	...	...	...
889	zebrafV7	---	...	...	...	...	...	...	...	...	...	...
956	xenopusV7	---	...	...	...	...	...	...	...	...	...	...
949	chickV7	---	...	...	...	...	...	...	...	...	...	...
1032	opossumV7	---	...	...	...	...	...	...	...	...	...	...
1030	mouseV7	---	...	...	...	...	...	...	...	...	...	...
1026	humanV7	---	...	...	...	...	...	...	...	...	...	...
830	cowV7	---	...	...	...	...	...	...	...	...	...	...
746	dogV7	---	...	...	...	...	...	...	...	...	...	...
827	drosophNURF	---	...	...	...	...	...	...	...	...	...	...
757	clonaiV7	---	...	...	...	...	...	...	...	...	...	...
172	clonasV7	---	...	...	...	...	...	...	...	...	...	...
	Clustal Consensus	---	...	...	...	...	...	...	...	...	...	...

920	celegansNURF	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
1044	medakaV7-8	---	...	...	...	...	...	...	...	...	...
1036	tetraodV7-Un	---	...	...	...	...	...	...	...	...	...
1006	fuguV7-1	---	...	...	...	...	...	...	...	...	...
1111	tetraodV7-18	---	...	...	...	...	...	...	...	...	...
1089	fuguV7-2	---	...	...	...	...	...	...	...	...	...
979	zebrafV7	---	...	...	...	...	...	...	...	...	...
1039	xenopusV7	---	...	...	...	...	...	...	...	...	...
1029	chickV7	---	...	...	...	...	...	...	...	...	...
1003	opossumV7	---	...	...	...	...	...	...	...	...	...
1118	mouseV7	---	...	...	...	...	...	...	...	...	...
1116	humanV7	---	...	...	...	...	...	...	...	...	...
1112	cowV7	---	...	...	...	...	...	...	...	...	...
916	dogV7	---	...	...	...	...	...	...	...	...	...
811	drosophNURF	---	...	...	...	...	...	...	...	...	...
837	clonaiV7	---	...	...	...	...	...	...	...	...	...
815	clonasV7	---	...	...	...	...	...	...	...	...	...

	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
celegansNURF	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
medakaV7-8	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
tetraodV7-Un	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
fuguV7-1	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
tetraodV7-18	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
fuguV7-2	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
zebrafV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
xenopusV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
chickV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
opossumV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
mouseV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
humanV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
cowV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
dogV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
drosophNURF	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
ciona1V7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
cionaV7	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700

	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
celegansNURF	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
medakaV7-8	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
tetraodV7-Un	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
fuguV7-1	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
tetraodV7-18	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
fuguV7-2	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
zebrafV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
xenopusV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
chickV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
opossumV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
mouseV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
humanV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
cowV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
dogV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
drosophNURF	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ciona1V7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
cionaV7	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800

Appendices

	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	
celegansNURF											
medakav7-8	PREHLSMAH	PREHLSMAH	PREHLSMAH	PREHLSMAH	PREHLSMAH	PREHLSMAH	PREHLSMAH	PREHLSMAH	PREHLSMAH	PREHLSMAH	
tetraodv7-Un	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	QATKEPVMAH	
fuguV7-1	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	CGNKEPAMAH	
tetraodv7-18	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	SPSPGILKDSVQCP	
fuguV7-2	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	VGEGGHLKDNLQCP	
zebrafV7	REPSMAH	REPSMAH	REPSMAH	REPSMAH	REPSMAH	REPSMAH	REPSMAH	REPSMAH	REPSMAH	REPSMAH	
xenopusV7	IVSHDILK	IVSHDILK	IVSHDILK	IVSHDILK	IVSHDILK	IVSHDILK	IVSHDILK	IVSHDILK	IVSHDILK	IVSHDILK	
chickV7	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	SPRHEMKTSSKSSK	
opossumV7	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	TIPEHEIKEDFKAPVKK	
mouseV7	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	VIPGHTKAGKSAK	
cowV7	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	
dogV7	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	IIPNDIKSATVKSATR	
drosophNURF											
cionaiV7	843										
cionasV7	1048	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	SLGLYTIFFQILQELK	
celegansNURF											
medakav7-8	1316	PPPLKPKTTHG	PPPLKPKTTHG	PPPLKPKTTHG	PPPLKPKTTHG	PPPLKPKTTHG	PPPLKPKTTHG	PPPLKPKTTHG	PPPLKPKTTHG	PPPLKPKTTHG	
tetraodv7-Un	1326	PPMKVLEH-YT	PPMKVLEH-YT	PPMKVLEH-YT	PPMKVLEH-YT	PPMKVLEH-YT	PPMKVLEH-YT	PPMKVLEH-YT	PPMKVLEH-YT	PPMKVLEH-YT	
fuguV7-1	1295	PPMKVLEHSHIT	PPMKVLEHSHIT	PPMKVLEHSHIT	PPMKVLEHSHIT	PPMKVLEHSHIT	PPMKVLEHSHIT	PPMKVLEHSHIT	PPMKVLEHSHIT	PPMKVLEHSHIT	
tetraodv7-18	1538	ESPTPPFENNA	ESPTPPFENNA	ESPTPPFENNA	ESPTPPFENNA	ESPTPPFENNA	ESPTPPFENNA	ESPTPPFENNA	ESPTPPFENNA	ESPTPPFENNA	
fuguV7-2	1522	RSTSPFSENNA	RSTSPFSENNA	RSTSPFSENNA	RSTSPFSENNA	RSTSPFSENNA	RSTSPFSENNA	RSTSPFSENNA	RSTSPFSENNA	RSTSPFSENNA	
zebrafV7	1346	PEFVASEP	PEFVASEP	PEFVASEP	PEFVASEP	PEFVASEP	PEFVASEP	PEFVASEP	PEFVASEP	PEFVASEP	
xenopusV7	1560	ROSLFA-TA-M	ROSLFA-TA-M	ROSLFA-TA-M	ROSLFA-TA-M	ROSLFA-TA-M	ROSLFA-TA-M	ROSLFA-TA-M	ROSLFA-TA-M	ROSLFA-TA-M	
chickV7	1601	KENLSSHES-M	KENLSSHES-M	KENLSSHES-M	KENLSSHES-M	KENLSSHES-M	KENLSSHES-M	KENLSSHES-M	KENLSSHES-M	KENLSSHES-M	
opossumV7	1550	KSSSNDEN-M	KSSSNDEN-M	KSSSNDEN-M	KSSSNDEN-M	KSSSNDEN-M	KSSSNDEN-M	KSSSNDEN-M	KSSSNDEN-M	KSSSNDEN-M	
mouseV7	1650	PRSSQVEL-M	PRSSQVEL-M	PRSSQVEL-M	PRSSQVEL-M	PRSSQVEL-M	PRSSQVEL-M	PRSSQVEL-M	PRSSQVEL-M	PRSSQVEL-M	
humanV7	1649	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	
cowV7	1645	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	
dogV7	1444	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	PESSNOVEL-M	
drosophNURF											
cionaiV7	843	1107	QDSKIDYES-M	QDSKIDYES-M	QDSKIDYES-M	QDSKIDYES-M	QDSKIDYES-M	QDSKIDYES-M	QDSKIDYES-M	QDSKIDYES-M	
cionasV7	1097	EASQVAFSPFRT	EASQVAFSPFRT	EASQVAFSPFRT	EASQVAFSPFRT	EASQVAFSPFRT	EASQVAFSPFRT	EASQVAFSPFRT	EASQVAFSPFRT	EASQVAFSPFRT	

Thr-rich

		2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
1011	celegansNURF	PDVLSRPNNOHSSGSSSLKQFGDAKVSSAAPA	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1384	medakaV7-8	KVKVVTGSGTGLTKN	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1394	tetraodV7-Un	KVKVVTGSGTGLTKN	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1363	fuguV7-1	SEALAPVSTSESSVSTLSTKTTKICGGLDSESD	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1621	fuguV7-2	DEFTFSTSESSVSTLSTKTTKICGGLDSESD	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1605	zebrafV7	DLASNGNLTATIDSKVSSSSVS	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1635	xenopusV7	QSS-NRAETIVSSITGSSIT	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1685	chickV7	ETLA-SNREKIVSSIECAISTV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1634	opossumV7	EVLASDQCSWSSIECAISTV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1733	mouseV7	KGKQIVSSIECAISTV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1732	humanV7	KGKQIVSSIECAISTV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1728	cowV7	KGKQIVSSIECAISTV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1527	dogV7	KGKQIVSSIECAISTV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1179	drosophNURF	LNQNERFPITRPVKRQCKKYSRETFENGNERVYSTSPR	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
843	clonaiV7	SGNVTVTKITVTAIV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK
1175	cionasV7	SGNVTVTKITVTAIV	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK	STRVTLKFSK

		2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
1011	celegansNURF	PEICITSTKRGKSTVLEOKKTRQIMG	COOVNE-GSAGISNIT	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1451	medakaV7-8	ARSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1437	tetraodV7-Un	ARSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1405	fuguV7-1	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1715	tetraodV7-18	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1700	fuguV7-2	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1528	zebrafV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1730	xenopusV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1780	chickV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1731	opossumV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1826	mouseV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1825	humanV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1821	cowV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1620	dogV7	TRSTLPSRRKVTSSKKSIFILPNDLKKTRRA	STREV-E-ESYGKPLI	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1253	drosophNURF	SR---YELISNLTGKRRSIFILPNDLKKTRRA	KSST-N-GHRAANN	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
843	clonaiV7	FDSTMF	PHC-HEVALPT-ELYREVECCBPIN	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC
1256	cionasV7	FDSTMF	PHC-HEVALPT-ELYREVECCBPIN	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC	MPVP-EPPTLTC

1279	celegansNURF	---Y---	SOQSRPN	N	RGFLE	KN	SSV	SPA	---	HGV	STPPP	YH	---	SPFN	ROAGY	QLPR	ET	IS	PF	NS	FR	AT	IT	---	2500
1700	medakaV7-8	ASTPTTLR	CAI	AKWL	SKL	ST	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2490
1686	tetraodV7-Un	TPVSA	LRL	QVIT	TKWL	SKLSSQ	---	VS	FE	KK	QSF	SWVK	Q	TS	SGA	---	---	---	---	---	---	---	---	---	2480
1654	fuguV7-1	SPST	LRM	QVIT	TKWL	SKL	SO	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2470
1956	tetraodV7-18	PKVIV	SIS	CGPPT	TKVL	STKV	TF	---	IT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2460
1939	fuguV7-2	PKVIV	SIS	CGPPT	TKVL	STKV	TF	---	IT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2450
1774	zebrafV7	PKVIV	SIS	QVSTK	VMTL	SP	---	VT	FO	KN	QSF	SWVK	Q	TS	SGA	---	---	---	---	---	---	---	---	---	2440
1976	xenopusV7	HKVW	---	SMPT	TKVL	TKV	SP	---	VT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2430
2030	chickV7	QKVW	P	PKPT	TKVL	TKV	SP	---	VT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2420
1981	opossumV7	QKVW	APIS	SVIT	TKVL	TKV	SP	---	VT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2410
2074	mouseV7	QKVW	APIS	SVIT	TKVL	TKV	SP	---	VT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2400
2073	humanV7	QKVW	APIS	SVIT	TKVL	TKV	SP	---	VT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2390
2069	cowV7	QKVW	APIS	SVIT	TKVL	TKV	SP	---	VT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2380
1868	dogV7	QKVW	APIS	SVIT	TKVL	TKV	SP	---	VT	FO	KN	QSF	TWVK	QS	TQ	---	---	---	---	---	---	---	---	---	2370
1493	drosohphNURF	SE	KEK	EQ	KL	ORA	HCQR	KIV	TGE	LR	SV	TP	VKG	VIG	RR	VIV	---	---	---	---	---	---	---	---	2360
844	clonaiV7	PHAL	LK	LN	NT	S	LE	K	---	FKC	IL	EV	TK	LT	H	C	---	---	---	---	---	---	---	---	2350
1546	cionasV7	PHAL	LK	LN	NT	S	LE	K	---	FKC	IL	EV	TK	LT	H	C	---	---	---	---	---	---	---	---	2340

1360	celegansNURF	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2600
1791	medakaV7-8	SSN	Q	CA	V	L	T	P	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2590
1779	tetraodV7-Un	AS	Q	Q	A	F	X	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2580
1747	fuguV7-1	ENTQ	Q	K	F	W	X	G	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2570
2045	tetraodV7-18	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2560
2031	fuguV7-2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2550
1866	zebrafV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2540
2068	xenopusV7	ATT	SC	Q	Q	A	F	X	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2530
2125	chickV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2520
2077	opossumV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2510
2170	mouseV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2500
2169	humanV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2490
2165	cowV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2480
1964	dogV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2470
1587	drosohphNURF	TS	PH	V	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2460
844	clonaiV7	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2450
1640	cionasV7	N	IGH	NP	T	N	L	S	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2440

1570	celegansNURF	TEGTVVRIENANGGR	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
1997	medakaV7-8	OLPQTIOLOSPISV	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2014	tetraodV7-Un	NHPOVTSLOSCXTPE	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
1982	fuguV7-1	NHPOVTSLOSCXTPE	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2264	tetraodV7-18	CADSASIKSEALE	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2257	fuguV7-2	TILTAPXS	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2076	zebrafV7	SPSPVRLAPQEPPLSLE	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2298	xenopusV7	PASGSPHPLTECPPLSLE	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2349	chickV7	AQNGS QPV--SPTOP	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2306	opossumV7	TVQPS QPQTITQPOSFNOTAQPC	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2402	mouseV7	EPQAPQ--EQEP--E	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2397	humanV7	TPQSP QPQ--EQTPQSP	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2395	cowV7	AQPQPS QPSAQCEQSP	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
2192	dogV7	TPQSP QPQ--EQTPQSP	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
1882	drosophNURF	IVCSSES CTKLKLVLVQ	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
844	clonaiV7		2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
1872	clonasV7		2810	2820	2830	2840	2850	2860	2870	2880	2890	2900

1636	celegansNURF	NVSPFRVSSGGHGGT	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2072	medakaV7-8	SHLEQVTCLOQVQTHSS	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2099	tetraodV7-Un	LCQITOTIHEQVQVCFHSS	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2067	fuguV7-1	LCQITOTIHEQVQVCFHSS	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2291	tetraodV7-18	PLTTAAPASASQSIQPHQ	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2318	fuguV7-2	TLTTVAPTSASPSQIPHQ	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2117	zebrafV7	TFQPTPPHPPVQLRECP	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2365	xenopusV7	PAVARQSSCSRPTCTALV	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2417	chickV7	VQSPQSNLEQSSQVREQV	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2388	opossumV7	QSPQSNLEQSSQVREQV	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2466	mouseV7	VQSPQSNLEQSSQVREQV	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2469	humanV7	VQSPQSNLEQSSQVREQV	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2472	cowV7	VQSPQSNLEQSSQVREQV	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
2264	dogV7	VQSPQSNLEQSSQVREQV	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
1978	drosophNURF	IVSNSATAVEQANPCTSVK	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
844	clonaiV7		2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
1886	clonasV7	RLITDQFAGGSSVYSP	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000

[illegible][illegible]

2161 **celegansNURF**
 2598 **medakav7-8**
 2654 **tetraodv7-Un**
 2617 **fuguv7-1**
 2752 **tetraodv7-18**
 2809 **fuguv7-2**
 2609 **zebrafv7**
 2860 **xenopusv7**
 2939 **chickv7**
 2917 **opossumv7**
 2976 **mousev7**
 2986 **humanv7**
 2985 **cowv7**
 2781 **dogv7**
 2615 **drosofphNURF**
 844 **cionaiv7**
 2355 **cionaiv7**

celeganaNURF	2261	MELYDF
medakav7-8	2658	
tetraodv7-Un	2714	
fuguV7-1	2677	
tetraodv7-18	2812	
fuguV7-2	2869	
zebrafV7	2669	
xenopusV7	2317	
chickV7	2999	
opossumV7	2977	
mouseV7	3036	
humanV7	3046	
cowV7	3045	
dogV7	2841	
drosophNURF	2669	
cionaiv7	844	
cionaasV7	2404	