Characterisation of human peptidyl prolyl isomerase-like protein 1 (PPIL1) mutations

Alice Elizabeth Webb

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Leeds Institute of Biomedical and Clinical Sciences

June 2019

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Eamonn Sheridan for his guidance and support throughout my PhD project. I would also like to give particular thanks to Dr Iain Manfield for his help with training, useful scientific discussions and encouragement throughout my PhD. Thanks also go to Dr Arnout Kalverda and Dr Lars Kuhn who have given me a significant amount of help with the NMR work presented in this thesis, including with experiments, theory, data processing and many valuable scientific discussions. I am grateful to Prof. Sheena Radford and Prof. Alex Breeze for valuable discussions about this project. I would like to thank my other PhD supervisors Prof. David Bonthron and Dr Elizabeth Valleley for their advice on my PhD project.

During this PhD, I have been part of an ongoing collaboration with Dr Guoliang Chai and Prof. Joseph Gleeson, Laboratory of Paediatric Brain Disease, Howard Hughes Medical Institute, UC San Diego. I would like to thank all involved in this collaboration for the discussions about this project.

I am grateful to Dr Katarzyna Szymanska and Dr James Poulter for further help with lab work and to Dr James Ault for the mass spectrometry results. I would like to thank Rhys Thomas for further help with the NMR work. The next generation sequencing and homozygosity mapping for PPIL1 was conducted by Clare Logan, University of Leeds and the Laboratory of Paediatric Brain Disease, Howard Hughes Medical Institute, UC San Diego (unpublished data).

Special thanks are given to Peter Correy, whose generous donation supported the research presented here and whose interest and enthusiasm for the project were a source of great encouragement.

I would like to thank office 9.11 for their help and support throughout their PhD.

Finally, I would like to thank my Dad, sister Helen and partner Luke, for their continued love, help and support throughout the process. Special thanks go to my Mum, who although no longer with us, has been a particular inspiration to me, remaining positive through cancer treatment and always encouraging me throughout my studies.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

101_106 dup	Duplication of residues 101-106
3'SS	3’ Splice Site
5’SS	5’ Splice Site
A99T	Alanine 99 to Threonine substitution
AdML	“Adenovirus major late” splicing reporter gene
AS	Alternative Splicing
ATP	Adenosine Triphosphate
BPS	Branch Point Site
CAD	Coronary Artery Disease
CADD	Combined Annotation Dependent Depletion v.1.4
CD	Circular Dichroism
CLD	Cyclophilin-Like Domain
cryo-EM	Cryo-Electron Microscopy
CsA	Cyclosporin A
CyPA	Cyclophilin A
DNA	Deoxyribonucleic Acid
DNMR	Dynamic NMR
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDC	N-ethyl-N’-(dimethylaminopropyl) carbodiimide
EDTA	Ethylenediaminetetraacetic acid
eQTL	Expression Quantitative Trait Loci
G109C	Glycine 109 to Cysteine substitution
gnomAD	Genome Aggregation Database
GRCh38	Genome Reference Consortium Human Build 38
GST	Glutathione S-transferase tag

HSQC	Heteronuclear Single Quantum Coherence
I	Induced
ILS	Intron Lariat Spliceosome
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
k_{ex}	Rate of Exchange
LB	Luria Broth
MCPH	“Microcephaly Primary Hereditary” = autosomal recessive primary microcephaly
MRI	Magnetic Resonance Imaging
MW	Molecular Weight
NEB	New England Biolabs
NEP	Neuroepithelial progenitor cell
NHS	N-hydroxysuccinimide
NLS	Nuclear Localisation Signal
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NRMSD	Normalized Root-Mean-Square Deviation
NTC	Nineteen Complex = hPrp19/CDC5L Complex
NTR	PRP19 Related Proteins
OD_{600}	Optical Density at 600 nm
ORF	Open Reading Frame
PCH	Pontocerebellar Hypoplasia
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pNA	*p*-nitroanilide
Polyphen	Polymorphism Phenotyping v2
PPIase	Peptidyl Prolyl Isomerase
PPIL1	Peptidyl-Prolyl Isomerase-Like 1

pre-mRNA	Precursor Messenger RNA
PROVEAN	Protein Variation Effect Analyzer
PRP17	Pre-mRNA-Processing Factor 17
R_1	Longitudinal relaxation rate
R131Q	Arginine 131 to Glutamine substitution
r.f.	Radiofrequency
RGC	Radial Glial Cells
RNA	Ribonucleic Acid
ROESY	Rotating Frame Nuclear Overhauser Effect Spectroscopy
rpm	Revolutions Per Minute
RU	Response Units
S. cerevisiae	*Saccharomyces cerevisiae*
SD	Standard Deviation
SDM	Site-Directed Mutagenesis
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SIFT	Sorting Intolerant From Tolerant
siRNA	Small Interfering Ribonucleic Acid
SKIP	SKI-Interacting Protein
snRNAs	Small Nuclear RNAs
snRNPs	Small Nuclear Ribonucleoproteins
S. pombe	*Schizosaccharomyces pombe*
SPR	Surface Plasmon Resonance
suc	Succinyl
suc-AAPF-pNA	Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide
suc-AGPF-pNA	Succinyl-Ala-Gly-Pro-Phe-p-nitroanilide
T107A	Threonine 107 to Alanine substitution

TAE	Tris-Acetate-EDTA
TEMED	Tetramethylethylenediamine
TOCSY	TOtal Correlated SpectroscopY
Tri-snRNP	U4/U6.U5 triple Small Nuclear Ribonucleoprotein
TROSY	Transverse Relaxation Optimized Spectroscopy
U	Uninduced
UV	Ultraviolet
v/v	Volume/ Volume percentage
w/v	Weight/ Volume percentage

Abstract

A novel brain malformation syndrome was previously reported in six individuals from four families. Patients presented with profound microcephaly (OFC 7-9 SD < mean), severe cortical dysplasia, profound cerebellar hypoplasia of both vermis and hemispheres and brainstem hypoplasia. In previous work, autozygosity mapping and whole exome sequencing suggested variants in peptidyl-prolyl isomerase 1 (PPIL1) were responsible for the disease. All patient variants were homozygous and cause the following PPIL1 protein changes: p.A99T, p.T107A, p.R131Q and p.G109C;A101_D106dup. PPIL1 is expressed ubiquitously in humans and is part of the spliceosome, a complex RNA and protein enzyme responsible for splicing introns from pre-mRNA. To gain an insight into the mechanism of this disease, PPIL1 WT and mutant proteins have been expressed and purified from *E. coli*. This work has shown that two PPIL1 patient mutations significantly affect protein stability. Using SPR, a further patient mutation has been shown to abolish an interaction with SKIP, a protein which is thought to recruit PPIL1 to the spliceosome. PPIL1 is also a peptidyl prolyl-isomerase enzyme and using NMR it has been shown that a further patient mutation reduces this enzymatic activity on a model substrate. Currently the biological target of PPIL1 peptidyl prolyl-isomerase enzyme activity in the spliceosome is unknown. By studying published cryo-EM data and using NMR and ITC experiments it has been shown that PRP17, is a plausible biological target of PPIL1 in the spliceosome. Taken together these results confirm the deleterious effect of the identified patient variants and provide novel insights into the role of PPIL1.

Chapter 1 Introduction

1.1 The development of the brain and associated genetic disorders

1.1.1 Neurogenesis and brain development

Neural cell fate is determined early in development, in a population of cells of the ectoderm, by signals from the notochord[1]. Next, the neural plate folds inwards and fuses to form a hollow tube, the neural tube[2]. Around this time the patterning is established for specific regions of the nervous system. Before fusion of the neural tube is complete, the anterior region expands to form three primary vesicles from which the brain will develop: the prosencephalon (forebrain precursor), the mesencephalon (midbrain precursor) and the rhombencephalon (hindbrain precursor). These vesicles will further divide to form five secondary vesicles, which provide primary patterning for the central nervous system[3]. Formation of these structures is temporally controlled, with early migrating cells forming forebrain precursors and later migrating cells forming hindbrain precursors[3]. During formation of the cortex, the earliest migrating neurons form the deepest layer of the cortex and later migrating neurons form subsequent layers of the cortex, resulting in the characteristic 6-layered structure. (Each of the six layers of the cortex contain a different distribution of neurons and pattern of connections)[2,3]. The brain originally forms with a smooth surface and later acquires the characteristic folds (gyri and sulci), as a result of rapid neuronal proliferation and migration[3].

The neural plate is established by rapidly dividing neuroepithelial progenitor (NEP) cells. Mitosis initially takes place in a symmetrical manner, increasing the population of progenitor cells. At the onset of neurogenesis NEPs acquire the identity of radial glial cells (RGCs) and mitosis shifts from symmetrical to asymmetrical division. Asymmetric division of RGCs results in one progenitor cell (RGC) and a second differentiating cell with neural or glial cell fate (the two major classes of cells in the nervous system; Figure 1)[2,3]. The cells which undergo neurogenesis to form the brain, NEPs and NEP-derived radial

glial cells, are located in the ventricular and sub-ventricular zones of the neural tube (Figure 1). Shortly after they are formed, neurons migrate away from the ventricular zones and further differentiate into specialised cells with a range of different morphologies and functions. (Production of neurotransmitter and neurotrophic factors trigger dendrites, axons and synapses to form establishing neural circuits. Non-neuronal processes including gliogenesis and myelination also take place[4].)

Many factors are essential for correct regulation of neurogenesis. These include maintenance of apical polarity in RGCs, which is essential for maintaining the asymmetric division of RGCs; (apical polarity refers to the polarised organisation of RGCs which contain an apical surface that faces the ventricle and basal surface that contacts the pial basal membrane). Temporal and spatial regulation of the switch from symmetric to asymmetric division in RGCs is also important to ensure the correct number of neural cells are produced. This requires correct orientation of the mitotic spindle and segregation of cellular components and cell fate determinants[4]. The fate of differentiating cells is also tightly controlled by many factors such as: transcription factors, epigenetic modifications, post-transcriptional regulation of gene expression (e.g. pre-mRNA splicing), signalling pathways (Notch, Hedgehog, Wnt and Fgf signalling pathways) and the environment of the neural progenitor cells[1].

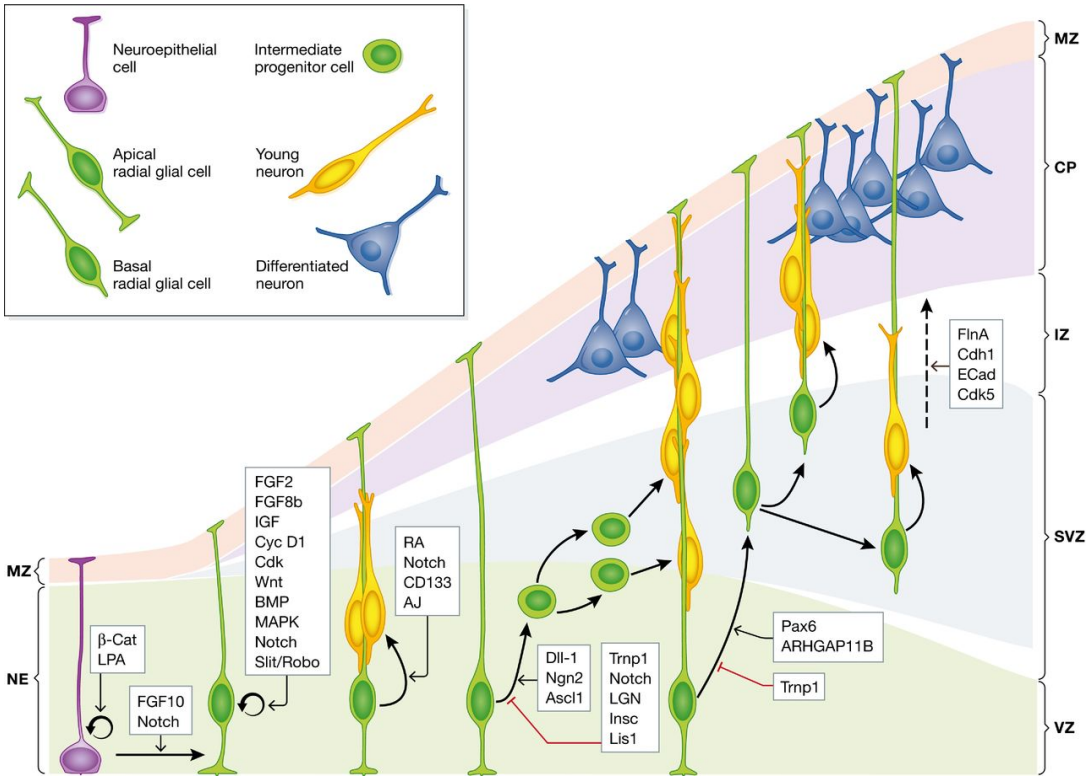


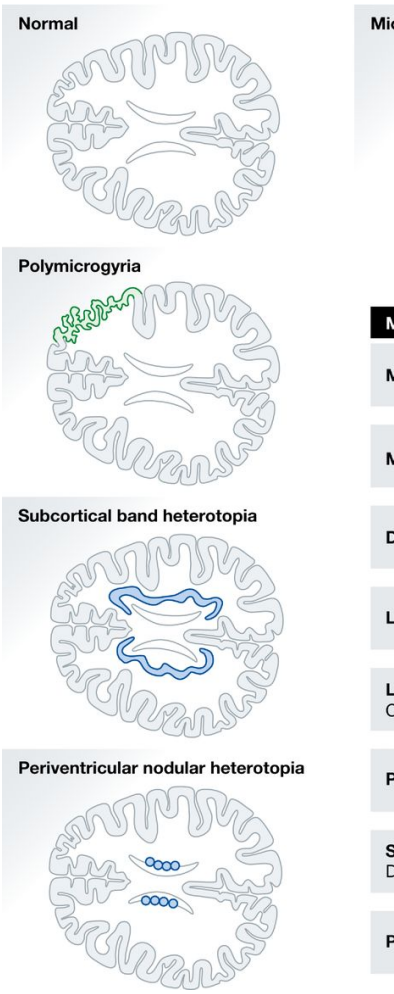
Figure 1 Neurogenesis during cerebral cortex development. Neuroepithelial progenitor cells (NEPs) undergo symmetrical divisions to produce radial glial cells (RGCs). RGCs subsequently divide asymmetrically, giving rise to cells with neural or glial fate (either directly or via intermediate steps, involving intermediate progenitor cells or basal radial glial cells). Progenitor (NEPs and RGCs) cell bodies are located at the ventricular zone (VZ). As cells differentiate, they migrate away from the VZ, through the subventricular zone (SVZ), intermediate zone (IZ), and cortical plate (CP) and marginal zone (MZ). Intermediate progenitor cells and basal radial glial cells populate the SVZ. Neurons migrate to the cortical plate, with newer neurons forming the outer layer, in an "inside-out" fashion. The MZ and VZ are primarily transient structures in the developing brain which are absent from the mature brain. Neurogenesis is a tightly controlled process and key pathways involved are indicated. The figure is taken from Fernández et al. 2016 (DOI 10.15252/embj.201593701)[5]. (*This is an open access article, which allows use and distribution of material, under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.*)

1.1.2 Microcephaly

Cortical neurogenesis is a tightly regulated process, however when this process is altered or disrupted cortex malformations may occur (Figure 2). Microcephaly describes a condition in which patients have a reduced head size as a result of impaired brain development (Figure 2) and is defined as an occipitofrontal head circumference three standard deviations below the

mean for a given age and sex[6]. The cerebral cortex makes up approximately 55% of the human brain and therefore is the structure most commonly affected in microcephaly; however other structures of the brain may also be reduced in size. Microcephaly is a heterogenous condition and can be caused by environmental factors, such as prenatal infection or exposure to toxins, or may have a genetic cause. Genetic cases of microcephaly may be primary (present at birth) or secondary (postnatal onset) and part of a syndrome or may be an isolated finding. In addition, non-syndromic patients may have either grossly normal or grossly abnormal brain architecture. Abnormal brain architecture associated with microcephaly may present as: holoprosencephaly, failure of the forebrain to develop in two hemispheres; lissencephaly a disorder characterised by loss of gyral folds and thickening of the cortex; or schizencephaly where abnormal clefts are present in the cerebral cortex[7]. Genetic cases of microcephaly, with no major changes in brain architecture or further syndromic features are referred to as autosomal recessive microcephaly syndrome (MCPH)[8].

Due to advances in next generation sequencing there has been a rapid increase in the number of known genes which contain mutations that cause microcephaly. The genes identified encode proteins with diverse functions, although many mutations have an autosomal recessive pattern of inheritance and influence crucial cellular processes[9]. (However mutations which cause microcephaly are not limited to an autosomal pattern of inheritance, for example, autosomal dominant mutations in *LIS1* lead to lissencephaly[10].) Interestingly the proteins encoded by many of the genes identified, have a role in the mitotic microtubule spindle or the centrosome, an organelle that contains the microtubule organising centre (mutations in these genes are mainly associated with MCPH)[11-13]. In addition, genes have also been identified which encode proteins that function in: DNA damage response, DNA replication, apoptosis pathways, cell cycle progression, cilia function and RNA processing[9,14-16]. (However, mutations in these genes may also cause additional symptoms in tissues beyond the brain.) Mutations in these genes lead to reduced proliferation or increased apoptosis of neural progenitor cells; therefore, reducing the final number of neurons. It remains unclear why the developing foetal brain is particularly sensitive to changes in processes which are critical in all cells[10,11,7].



MALFORMATION	SIZE	FOLDING	ECTOPIA
Microcephaly	Reduced	Mild – Normal Severe – Altered	No Yes
Megalencephaly	Increased	Mild – Normal Severe – Altered	No No
Dysplasia	Altered	Increased	Abnormal layers
Lissencephaly type I	Normal Reduced	Reduced	Abnormal layers
Lissencephaly type II Cobblestone	Normal	Increased	Cortical surface
Polymicrogyria	Increased Reduced	Increased	Abnormal layers
Subcortical band heterotopia Double cortex	Reduced	Normal	White matter
Periventricular heterotopia	Normal Reduced	Normal Normal	Periventricular

Figure 2 Types of cortex malformation. Diagrams show typical brain morphology of affected patients, viewed in the axial plane. The table lists features (size, folding and ectopia) associated with these disorders. The figure is taken from Fernández et al. 2016 (DOI 10.15252/embj.201593701)[5]. (*This is an open access article, which allows use and distribution of material, under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.*)

1.1.3 Pontocerebellar hypoplasia

Pontocerebellar hypoplasia (PCH) is a heterogenous group of inherited disorders characterised by reduced size of infratentorial brain regions; including the pons, cerebellum, and olivary nuclei. This may also be accompanied by developmental abnormalities of other brain regions including the cerebral cortex and basal ganglia. The onset of the disease is usually prenatal and results in global developmental delay. PCH is divided into 11 main sub-types based on: presentation of different symptoms, MRI findings and more recently the genetic cause. PCH was originally described

as a neurodegenerative condition, but has since been expanded to include nondegenerative sub-types. All cases of PCH currently identified follow an autosomal recessive pattern of inheritance. The genes implicated in PCH encode proteins which have a wide range of functions including: RNA processing, protein translation, mitochondria function, and golgi membrane trafficking[17,18].

1.1.4 A novel brain development disorder: molecular genetics and clinical details

Homozygous variants in PPIL1 have been identified as the cause of a novel form of microcephaly and pontocerebellar hypoplasia (unpublished data, University of Leeds and University of California, San Diego; Table 1). Two patients from a single family with: severe microcephaly (head circumference of 8-9 s.d. below the mean), severe cortical dysplasia with a smooth cortex surface, profound cerebellar hypoplasia of both vermis and hemispheres and brainstem hypoplasia, were suspected to have an autosomal recessive disease (Figure 3, Figure 4, Table 1). These patients also showed profound neonatal hypotonia, poor feeding but no seizures. No variants were detected in genes previously known to be responsible for congenital brain disorders [a]. Homozygosity mapping analysis indicated the mutation responsible to be Chr. 6:36855995T>C (GRCh38) homozygous single nucleotide transition in *PPIL1*, resulting in a p.Thr107Ala (p.T107A) amino acid substitution (unpublished data, University of Leeds). Sanger sequencing of the patients' and parents' DNA confirmed segregation of the PPIL1 variant with disease. More recently, a further patient also displaying the clinical features described above, has been found to be homozygous for Chr.6:36855996_36856013dup; 36855989C>A (GRCh38) resulting in Ala101_Asp106dup; Gly109Cys (p.A101_D106dup; G109C) (unpublished data, University of Leeds).

Independently, homozygosity mapping also mapped variants to *PPIL1* in a further four patients (two families) suspected for autosomal recessive

---

[a] Genes tested by Leeds Genetics Laboratories (NHS) include Brain Malformation Disorders Gene Panel and Pontocerebellar Hypoplasia Disorders Gene Panel (see https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/details/3836/ and https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/details/6597/ respectively; Appendix Table 19)

disease. These patients also displayed similar symptoms. Microcephaly was less severe (7-8 s.d. below the mean) and these patients did show seizures, (myoclonic, focal or generalised). They also had less severe cortical dysplasia and pachygyria only in one case. Otherwise symptoms were as described above. The variant observed in the first family (2 patients) was: homozygous Chr.6:.36856019C>T (GRCh38) resulting in p.Ala99Thr (p.A99T) and in the second family (2 patients): homozygous Chr.6:36855922C>T (GRCh38) resulting in p.Arg131Gln (p.R131Q) (unpublished data, University of California, San Diego; Figure 4, Table 1).

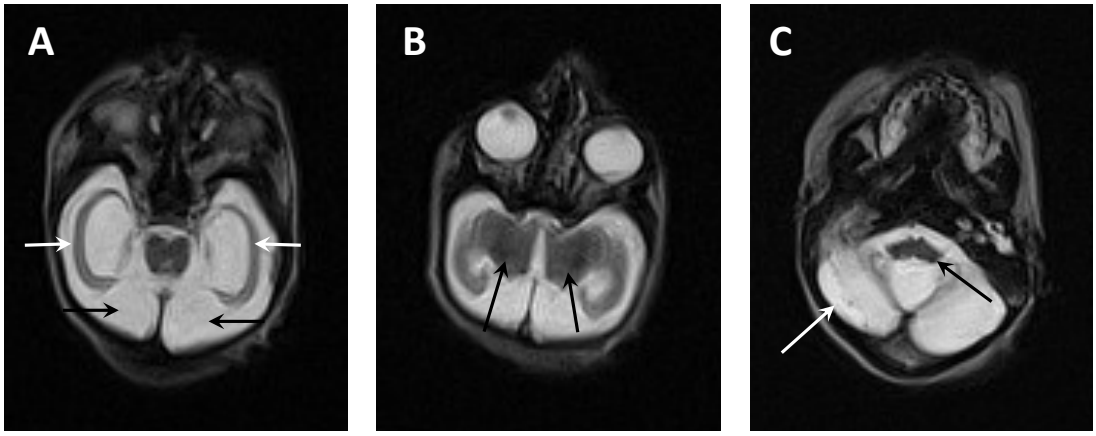


Figure 3 Axial MRI scans of a patient with severe microcephaly and pontocerebellar hypoplasia. The MRI images displayed, show the brain of a patient with homozygous PPIL1 p.T107A mutations. Scan A at the level of the cerebrum, shows large areas of fluid (white in appearance; black arrows), with only small regions of cerebral tissue (grey in appearance; white arrows). The region of the cerebellum is also filled with fluid, in the absence of brain tissue (black arrows). Scan B at the level of the midbrain shows that this region is largely unaffected (black arrows). Scan C at the level of the brainstem, shows a reduced brainstem size (black arrow) and absence of cerebellar tissue (white arrow).

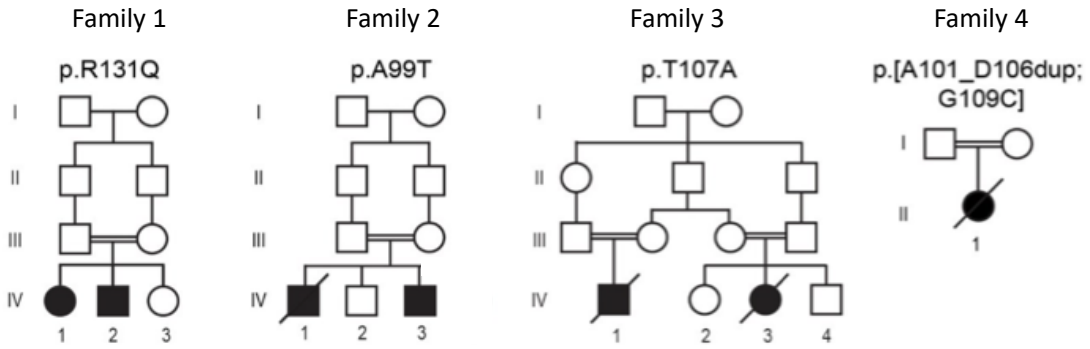


Figure 4 Pedigrees of the four families affected by microcephaly and pontocerebellar hypoplasia and discussed in this thesis.

Table 1 Patient details†.

Family	1		2		3		4
Mutation							
gDNA (GRCh38)	Chr.6: 36855922 C>T	Chr.6: 36855922 C>T	Chr.6: 36856019 C>T	Chr.6: 36856019 C>T	Chr. 6: 36855995 T>C	Chr. 6: 36855995 T>C	Chr.6:36855996_ 36856013dup; 36855989C>A
Protein	p.Arg131Gln	p.Arg131Gln	p.Ala99Thr	p.Ala99Thr	p.Thr107Ala	p.Thr107Ala	p.Ala101_Asp106 dup;p.Gly109Cys
cDNA	c.392G>A	c.392G>A	c.295G>A	c.295G>A	c.319A>G	c.319A>G	c.301_318dup; c.325G>T
Patient details							
Gender	Female	Male	Male	Male	Male	Female	Female
Ethnic origin	Egypt	Egypt	Egypt	Egypt	Pakistan	Pakistan	Pakistan
Pregnancy duration (weeks)	full term	full term	full term	full term	40 weeks	41 weeks	40 weeks
Weight at birth (kg)	N/A	N/A	1.75kg	2.5kg	1.8kg	2.0kg	1.7kg
Length at birth (cm)	N/A	N/A	N/A	50cm	45cm	46cm	48cm
HC at birth (SD)	N/A	N/A	26.4cm (-6.3SD)	26cm (-6.7SD)	25cm	24cm	26cm
HC at last examination (SD)	45cm (-7SD)	43.5 (-7SD)	29.5 (-8SD)	29.5cm (-8SD)	28.5cm (-8.5SD)	27cm (-9.0SD)	
Diagnosis age	6 months	2 months	Birth	Birth	Birth	Birth	Birth
Intellectual Disability	Severe	Severe	Severe	Severe	N/A	N/A	N/A
Development							
Gross motor (normal/delayed/ absent)	Delayed	Delayed	Absent	Absent	Absent	Absent	Absent
Fine motor (normal/delayed/ absent)	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Language (normal/delayed/ absent)	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Social (normal/delayed/ absent)	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Seizures							
Onset	6 months	2 months		3 weeks	None	None	None
Type	Generalized	Myoclonic	Focal	Myoclonic			N/A
Frequency	Infrequent	Monthly		Intractable			N/A
Neurological Findings							
Hypertonia	mild	mild	mild	mild	-	-	-
Hypotonia	mild	mild	-	-	severe	severe	severe

† The patient data refers to unpublished data from the University of Leeds and University of California, San Diego.

Deep tendon reflexes	brisk	brisk	brisk	brisk	brisk	brisk	brisk
Spastic tetraplegia	+	+	+	+	+	+	++
Ataxia	-	-	-	-	-	-	-
Investigations							
Metabolic	Neg	Neg	N/A	N/A	Neg	Neg	Neg
VEP/ERG	N/A	N/A	N/A	N/A	NO electrophysiology but normal neonatal eye exam	NO electrophysiology but normal neonatal eye exam	NO electrophysiology but normal neonatal eye exam
EEG	B frontotemp dc	B frontotemp dc	N/A	N/A	N/A	N/A	N/A
MRI							
Microcephaly (-SD)	-7SD	-7SD	-8SD	-8SD	-8.5SD	-9.0SD	-8SD
Cerebral cortex (Pachygyria)	-	-	+	+	+	+	++
Subcortical band heterotopia	-	-		-	-	-	-
Agenesis of corpus callosum	+	+	+	+	+	+	+
Cerebellar hypoplasia	+	+	+	+	+	+	+
Brainstem hypoplasia	-	-	+	+	+	+	+
Hydrocephalus	-	-	-	N/A	-		+
White matter abnormalities	-	-	-	-	-	-	-
Miscellaneous							
Polyhydramnios	-	-		-	-	-	-
Lung hypoplasia	-	-		-	-	-	-
Short stature	-	-		-	-	-	-
Optic atrophy	-	-		-	-	-	-
Autistic features	-	-		-	-	-	-
Dysmorphism	-	-		-	-	-	-
Other							
			died at 8 months		died aged 2months	died age 3 months	died age 2 months
			inguinal hernia		no thrombocytopenia	persistent thrombocytopenia	

1.1.5 Investigating pathogenicity of patient variants

Identifying the correct disease causing variant is critical and can have important implications for therapy or genetic counselling (e.g. prognostic and family planning advice). Recent evidence suggests that reporting of false positives remains a problem and it is therefore important that appropriate evidence is generated for variant pathogenicity[19]. Linking causality to a variant in rare diseases is often a challenge due to the small number of cases; this may include just a few affected families or even a single patient. In addition, researchers are often faced with several possible candidate gene variants and must examine the evidence to determine the causative variant. As recommended by the 2015 American College of Medical Genetics and Genomics and the Association for Molecular Pathology published guidelines for interpretation of variants, several lines of evidence should be used in order to implicate variants as pathogenic[20].

To be implicated in a rare genetic disease, variants are typically filtered to those with a minor allele frequency less that 1%[21]. If it is possible to obtain sequencing data from parents of the patient, this can also be used to confirm segregation of the variant with disease. The location of the variant is considered, for example whether the variant occurs in a previously implicated gene. Examination of the phenotype, is also used to help classify variants, for example whether the phenotype is consistent with the available published gene information. Recurrence of the variant in other individuals with similar symptoms is considered one of the ‘gold standard’ lines of evidence for variant pathogenicity. However, in many genetic diseases are so rare that this is not possible[19-21].

The likely effect on gene function is considered. For example, nonsense, frameshift mutations or deletion of essential splice sites may suggest disruption of gene function by creation of a null allele. However, this evidence must be examined with caution as these mutations may not always be pathogenic; for example, the affected exon may be alternatively spliced and therefore the protein is still produced[20]. For variants which do not result in a null allele, such as missense mutations, the effect on protein function can also be examined. Specifically, amino acid conservation between species and biochemical properties of the substituted amino acid, such as charge and hydrophobicity, are key lines of evidence for predicting variant pathogenicity. In addition, location of the mutation in the protein structure can

help to predict the consequence of the mutation. Various tools are available, including Polyphen and SIFT which make predictions based on these lines of evidence. However, these tools have relatively low prediction rates with an estimated specificity of approximately 68-69% and sensitivity of approximately 13-16%[22].

Animal models may also be used to demonstrate a genotype-phenotype link; where animals which contain an equivalent mutation in the gene and display a phenotype which resembles the human disease are given as evidence for a variant. However, this work is often slow, expensive is not always a relevant model. In addition, this would not be practicable in cases with many different variants in the same gene[21]. Biochemical assays can also provide important evidence of variant pathogenicity. Assays may be based on patient cells, cell line models or purified protein depending on the availability of tissue and the type of assay required. These assays are highly valuable, providing not just evidence of pathogenicity but also insights into the potential mechanism behind a disease.

In the case of autosomal recessive disease further criteria can be used to filter variants. Variants are removed which are homozygous in healthy control populations. In consanguineous families, autozygosity mapping, which identifies regions of homozygosity in affected patients, can also be used to narrow down the region containing the causative variant[23,24], as was used in the cases identified in Leeds.

1.2 PPIL1

1.2.1 Background

Peptidyl-prolyl isomerase-like 1 (PPIL1) encodes a single domain protein of the cyclophilin enzyme family. Like most other cyclophilins PPIL1 has peptidyl prolyl isomerase activity (PPIase), the ability to catalyse the transition between *cis* and *trans* isomers of a peptide bond preceding a proline residue (see section 1.2.2). PPIL1 is highly conserved across species, from human to *S. pombe* (Figure 5); although there is no homolog in S. cerevisiae. PPIL1 localises to both the cytoplasm and nucleus, and has also been detected in extracellular environments including plasma and urine[25,26]. Expression of PPIL1 was shown to be ubiquitous across all adult tissues (with highest expression in the adrenal gland, testis and heart)[27,28].

There is also recent evidence that PPIL1 expression may fluctuate with the circadian clock[29]. PPIL1 is also known to participate in the spliceosome; a complex and dynamic protein and RNA complex responsible for splicing introns from pre-mRNA[30].

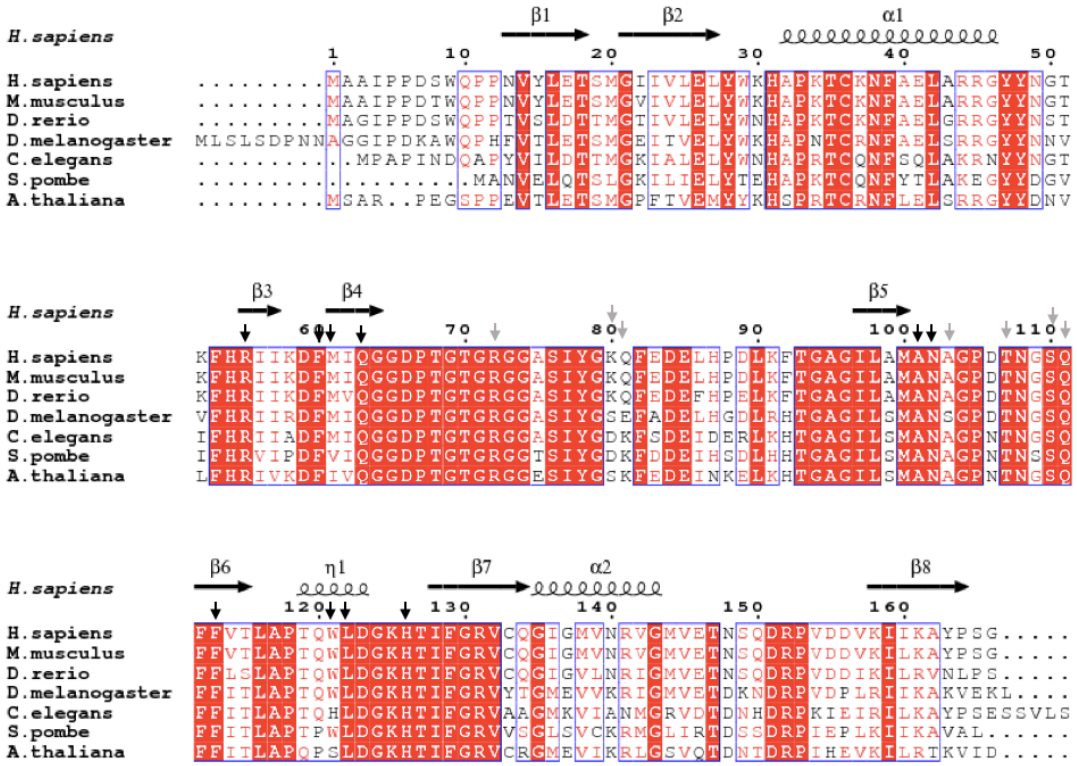


Figure 5 Alignment of PPIL1 amino acid sequence, showing conservation across species: *Homo sapiens* (NP_057143.1), *Mus musculus* (NP_081121.1), *Danio rerio* (NP_001029350.1), *Drosophila melanogaster* (NP_523874.1), *Caenorhabditis elegans* (NP_501118.1), *Schizosaccharomyces pombe* (NP_593308.1) and *Arabidopsis thaliana* (NP_181157.1). Conserved residues are shown with red letters and strictly conserved amino acids are highlighted in red with white letters. Sequences were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/)[31,32] and rendered for sequence similarities using ESPript (http://espript.ibcp.fr)[33]. Vertical black and grey arrows indicate PPIL1 active site and S2 pocket gatekeeper residues respectively.

1.2.2 PPIase activity

The transition between *cis* and *trans* isomers of a peptide bond preceding a proline residue (prolyl peptide bond; Figure 6) is a slow process and all

organisms contain PPIases which are able to accelerate this process, by stabilising the transition. Unlike most peptide bonds which are in the *trans* configuration, peptide bonds preceding proline are also found in the *cis* conformation; up to 30% of peptide bonds in peptides and 5-7 % of peptide bonds in folded proteins are present in the *cis* form[34]. This is a result of proline's unusual structure, which means that steric repulsion of the Cα atoms, either side of the imide bond, is almost equal in both the *cis* and *trans* conformations[35]. Due to the partial double bond character of peptide bonds, the energy required for rotation around the bond, between *cis* and *trans* states, is high, (approximately 80 kJ/mol). Therefore, in the absence of PPIase enzymes *cis-trans* isomerisation is slow (usually 10-100 s at 25 ^{o}C)[35].

In folded protein regions, local conformations and long range interactions have a significant effect on populations of the *cis* and *trans* isomer, whereas for disordered regions or peptides the preceding residue has a large impact on the ratio of *cis* and *trans* isomers. Aromatic and glycine amino acids which precede the proline peptide bond typically result in higher populations of the *cis* isomer compared to small, electron poor residues[36-38]. Due to the rigid backbone of proline, the peptide bond isomer can have a large effect on the secondary and tertiary structures of the peptide chain. PPIase activity may therefore be required for: correct protein folding [34,39,40], modulating protein function [35] or as a molecular switch in cellular processes such as signalling pathways [41,42].

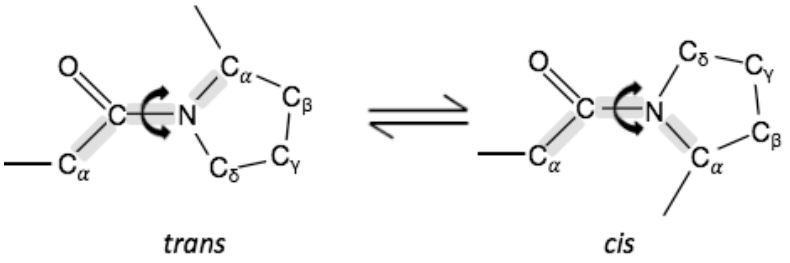


Figure 6 *Cis-trans* isomerization of a peptide bond preceding a proline residue. A peptide bond may occur in either a *cis* or *trans* state, where Cα atoms are found on the same or opposite sides of a peptide bond respectively. Interconversion between *cis* and *trans* states is an intrinsically slow process and is accelerated by a class of enzymes called PPIases.

1.2.3 Cyclophilins

Cyclophilins are one of three protein families with PPIase activity; (FK506 binding proteins and parvulins also possess PPIase activity although they are structurally unrelated)[43]. Cyclophilins are characterised by their ability to bind the immunosuppressant molecule cyclosporine A (CsA). All cyclophilins contain a conserved cyclophilin-like domain (CLD), which is a compact beta barrel structure capped by two alpha helices. In addition some cyclophilins contain other unique domains, which may allow further interactions with other proteins or targeting to subcellular locations[40]. Cyclophilins are ubiquitous proteins found in both prokaryotic and eukaryotic species and they have been reported to have a range of functions including in: mitochondrial apoptosis, mRNA splicing, interferon signalling pathway, and even stress tolerance and virulence of bacteria[39,44,45]. Of the 19 cyclophilins encoded by the human genome, 9 are localised in the nucleus and all 9 of these proteins function in the spliceosome (Table 2)[46].

Cyclophilin	**Alternative names**	**Domains**	**PPIase activity?** *[39]	**Details**
PPIL1 (peptidyl prolyl isomerase-like protein 1)	*CYPL1, hCyPX, CGI-124*	Cyclophilin only	Yes	Part of NTR spliceosome sub-complex (B^{act} -> ILS)[47,48,49,50].
PPIL2 (peptidyl prolyl isomerase-like protein 2)	*CYC4, Cyp60, UBOX7, Cyp58*	Cyclophilin, N-terminal U-box (E3 ligase), predicted coiled-coil region	No	Associates with tri-SNP (U4/U6.U5) in spliceosome before and during B^{act} [30,50,51].
PPIL3b (peptidyl prolyl isomerase-like protein 3)	*CyPJ*	Cyclophilin only	Not tested	Present in spliceosome B^{act} [(50)] to C* complexes[30,51,52].
PPIL4 (peptidyl prolyl isomerase-like protein 4)	*CyP57, HDCME13P*	Cyclophilin, C-terminal RRM, lysine rich domain	Not tested	Co-purifies with C, C* spliceosome complexes[53,51].
CWC27 (CWC27 spliceosome associated protein homolog)	*RPSKA, NY-CO-10, SDCCAG10*	Cyclophilin, predicted coiled-coil regions, extended C-terminus of unknown	No	Associates with spliceosome before B^{act} and present in C complex[30,52,49].

		function		
PPWD1 (peptidylprolyl isomerase domain and WD repeat containing 1)	*CyP73*	Cyclophilin, WD40 repeats	Yes	Spliceosome C complex[51].
PPIE (peptidylprolyl isomerase E)	*CYP33, CYP-33*	Cyclophilin, N-terminal RRM	Yes	NTR spliceosome sub-complex (B^{act} -> C*)[30,52,51,49,50].
PPIG (peptidylprolyl isomerase G)	*SR-Cyp, CARS-Cyp, CYPG, Matrin-CyP (rat)*	Cyclophilin, N-terminal RS domains	Yes	Present in spliceosome C complex[30].
PPIH (peptidylprolyl isomerase H)	*Snu-Cyp20, USA-Cyp, CyPH*	Cyclophilin	Yes	Associates with tri-SNP (U4/U6.U5) and is incorporated at B complex[30]. Dissociates with U4 snRNP before B^{act} complex[51].

Table 2 Nuclear cyclophilins.* PPIase activity was demonstrated on a model substrate (suc-AGPF-pNA peptide) by Davis et al. (PPIL3b and PPIL4 were insoluble / unstable and therefore not tested)[39]. Domains were annotated using the Uniprot database (https://www.uniprot.org/)[54].

1.2.4 PPIL1 structure

1.2.4.1 NMR and crystal structures

NMR spectroscopy and X-ray crystallography revealed that, like other cyclophilins, the structure of PPIL1 involves a compact eight stranded beta barrel enclosed by two alpha helices on either side and with a short 3_{10} helix between the β6 and β7 strands[55,56]. Although the majority of PPIL1 is relatively rigid, some of the loops linking beta strands show significant flexibility. The loop between β5 and β6 is dynamic and this is thought to be important for catalytic activity. In addition the loop linking β4 and β5 is also dynamic, although the role of this is less clear[55]. The structure of PPIL1 is highly similar to cyclophilin A, the first identified cyclophilin[55]. PPIL1 contains a single PPIase active site and, as with other cyclophilins, this is positioned on the β3 and β4 strands[55,56]. However, compared to CypA, PPIL1 also exhibits several differences in structure. PPIL1 β1-β2 loop is five residues

shorter than CypA (as shown for other cyclophilins including: PPIL2, PPIL3, PPIL4, PPWD1 and SDCCAG-10). In addition, PPIL1 also contains a shorter α1-β3 loop compared to CypA (by three residues), which results in an altered conformation of the C-terminal of α1 helix (this loop is also shorter in PPIL2, PPIL3, PPIL4, PPWD1 and SDCCAG-10). The α2-β8 loop also forms a different conformation compared to CypA, and instead shows greater similarities to other cyclophilins, such as NKTR (Figure 7)[39,55,56].

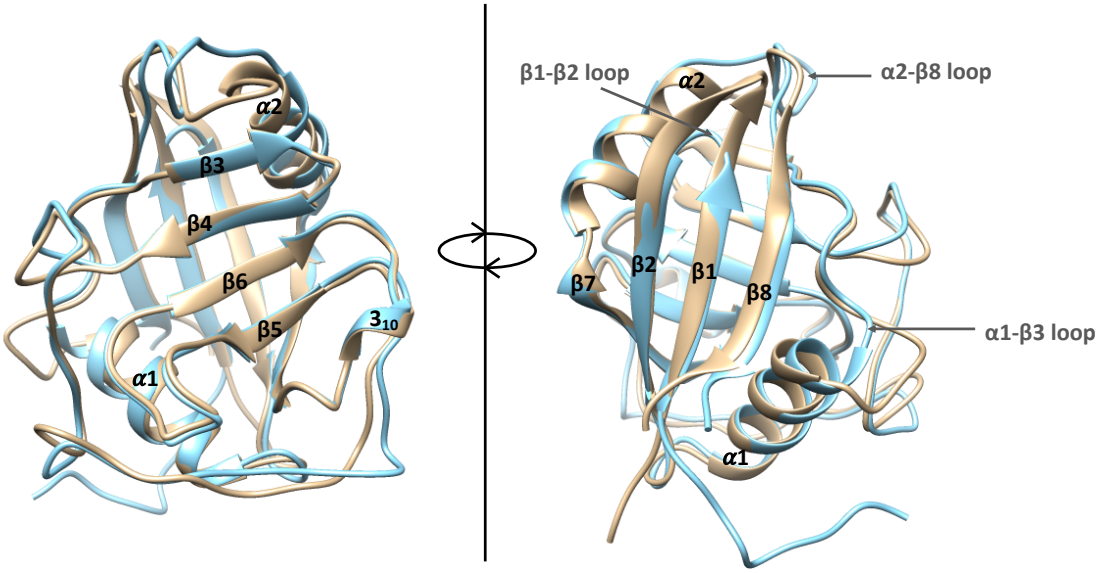


Figure 7 Overlay of PPIL1 (blue) and CypA (brown) crystal structures. Images were made using Chimera (http://www.rbvi.ucsf.edu/chimera)[57] and data from PDB (2X7K[56], 1ZKF). The β1-β2, α2-β8, and α1-β3 loops, which differ between CypA and PPIL1, are indicated with arrows

1.2.4.2 PPIL1 active site

The PPIase active site is defined by the following residues Arg55, Phe 60, Met 61, Gln 63, Ala 101, Asn 102, Phe 113, Leu 122 and His 126 (see Figure 8); these residues form the hydrophobic proline binding pocket or lie within 4Å of the Xaa-Pro peptide bond in model structures of the PPIL1 homolog CypA[39,58]. These residues in PPIL1 are strictly conserved compared to CypA and most residues are highly conserved across PPIL1 homologs and other cyclophilin PPIases (Figure 5; Appendix Figure 59). Residues Phe 113, Phe 60, Met 61, Leu 122 and His 126 line the hydrophobic pocket while Arg55,

Gln 63 and Asn 102 all form hydrogen bonds with the substrate[39,58]. Trp121 is not structurally part of the proline binding pocket, but is considered an active site residue as this residue forms a hydrogen bond with the substrate and mutating this residue significantly reduces catalytic activity of cyclophilins[58].

1.2.4.3 S2 pocket

A further hydrophobic pocket is located adjacent to the PPIase active site in cyclophilins (Figure 8A). This pocket has been named the S2 pocket, as it is thought to bind residues two and three N terminal to the prolyl peptide bond substrate (i.e. **P3-P2**-P1-Pro). The S2 pocket is defined by residues in the loop between β5 and β6 sheets. Further residues whose sidechains are positioned around the pocket have been named 'gatekeeper residues' as they are proposed to control access to the pocket. The gatekeeper residues were originally described in PPIA, and the corresponding residues in PPIL1 are: Arg72, Lys80, Gln81, Ala103, Thr107, Ser110, and Gln111[39]. The identity of these residues is variable across the cyclophilins and it is proposed this variation affects binding specificity of cyclophilin substrates[39]. In PPIL1 all of these residues have sidechains close to the S2 pocket, except Lys80 which points away from the pocket, due to the altered conformation of this loop compared to PPIA (see Figure 8)[56].

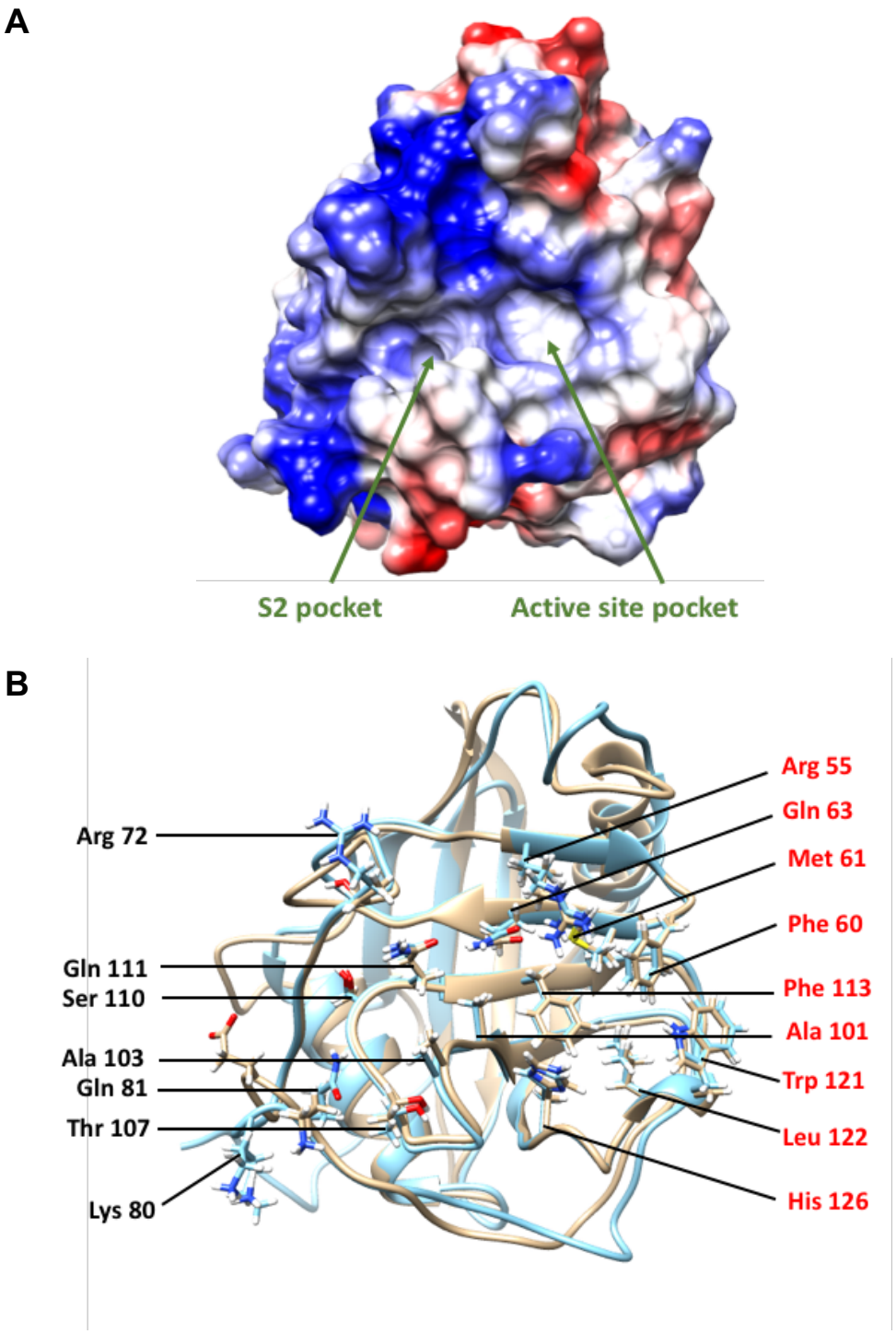


Figure 8 PPIL1 active site and S2 pocket. (A) Surface representation of PPIL1 crystal structure; electrostatic potential is shown using red (negative) and blue (positive). (B) Overlay of PPIL1 (light blue) and CypA (brown). PPIL1 gatekeeper residues (black labels) and active site residues (red labels) defined by Davis et al 2010 are shown[39]. (PPIL1 is shown in the same orientation in both images). Chimera software (http://www.rbvi.ucsf.edu/chimera)[57] and data from PDB were used to make these images (2X7K[56], 1ZKF).

1.2.5 Role of PPIL1 in disease

Several cyclophilins have been implicated in disease. For example, CypA has been shown to have a role in cardiovascular disease, virus infection and replication (HIV, HCV, influenza A), cancer and pro-inflammatory diseases (rheumatoid arthritis, sepsis)[59]. It is suggested that CypD has a role in collagen VI myopathies, reperfusion injury (of the heart and brain) and fulminant hepatitis. Increased CypB expression is also associated with breast cancer progression[60].

PPIL1 has also been linked to a wide range of diseases. Firstly, PPIL1 overexpression has been implicated in colon cancer; western blot and IHC staining show that that PPIL1 expression is increased in a significant number of colon cancer tissue samples. It was demonstrated that overexpression of PPIL1 in NIH3T3 cells and HEK293 cells also resulted in increased cellular proliferation. The authors also showed that reducing PPIL1 expression using siRNA reduces cell proliferation in SNUC4 and SNUC5 colon cancer cell lines[61].

PPIL1 is also implicated as a key regulatory gene or ‘driver gene’ in coronary artery disease (CAD); this is a central gene of a network which has been associated with significant changes in gene expression observed in CAD. Gene networks were assembled by integrating several types of published data. Genes which contain SNPs associated with CAD (from genome-wide association studies) were included into the model. Secondly, published CAD eQTLs (expression quantitative trait loci) were added to the model; these are genetic variants which explain variation in gene expression levels observed in CAD. Integration of the above data with known metabolic and signalling cascades and known gene-gene interaction networks lead to the creation of gene regulation networks involved in CAD. Genes at the centre of these networks were considered central regulatory genes in CAD. PPIL1 was found to be one of these central regulatory genes (driver genes) and was located in a network relating to antigen processing, including human leukocyte antigen complex genes. It was further shown that reducing PPIL1 expression in human aortic endothelial cells, using siRNA, results in changes of expression of 656 genes, including genes of the antigen processing network. These results suggest that PPIL1 has a direct or indirect role in the regulation of gene expression and may be associated with changes in gene expression observed in CAD[62].

PPIL1 was also shown to have elevated expression in T cells from patients with Psoriasis. Psoriasis is a chronic inflammatory skin condition, caused by unusually fast growth of the epidermal layer of skin. It was previously shown that patient T cells influence keratinocyte proliferation and the authors used differential gene expression of patient T cells to try to understand the mechanism for this. T-cells from patients with psoriasis, show elevated PPIL1 expression, compared to both unaffected twins and unaffected unrelated individuals. The authors suggest that PPIL1 may have a role in the abnormal keratinocyte proliferation observed in psoriasis[63].

Taken together these papers indicate PPIL1 may have an important role in regulation of cell proliferation across a range of tissues. In addition, these studies suggest a direct or indirect function for PPIL1 in gene expression. However, the exact mechanism involving PPIL1 in these diseases remains unclear.

1.3 The role of PPIL1 in the spliceosome

1.3.1 Assembly and mechanism of the spliceosome

PPIL1 participates in the spliceosome, a dynamic multi-protein and RNA enzyme, responsible for splicing introns from pre-mRNA (see Figure 9)[53,64]. In contrast to the majority of enzymes, the spliceosome does not have a preformed active site. Instead, the spliceosome is formed by sequential binding of spliceosomal components on the nascent pre-mRNA, followed by numerous conformational rearrangements. The spliceosome is composed of five stable small nuclear ribonucleoproteins (snRNPs; U1, U2, U4, U5 and U6), the PRP19-CDC5L complex, PRP19 related proteins (NTR) and numerous other proteins. SnRNPs contain specialized uridine rich small nuclear RNAs (snRNAs), which stably associate with specific proteins[30,65]. The PRP19-CDC5L (= NTC) complex is a multi-protein module and the NTR proteins are further proteins which associate with the PRP19/CDC5L complex during spliceosome assembly[66].

Introns contain consensus sequences which allow intron recognition by the spliceosome. The intron consensus sequences include the 5’ splice site (5’SS), branch point site (BPS) and 3’ splice site (3’SS); which in metazoans are GURAGU, YNYURAY and YAG respectively (where R is a purine, Y is a pyrimidine, and N is any nucleotide). In metazoans a polypyrimidine tract is also present which includes a 10-12 nucleotide sequence rich in pyrimidines,

particularly uracil[30]. The spliceosome catalyses two transesterification reactions which lead to intron excision and exon ligation. In the first step, known as the branching reaction, the 2' hydroxyl group of the conserved adenosine residue at the BPS attacks the 5'SS phosphodiester bond, producing a cleaved 5'exon and lariat intron. Secondly, the 3' hydroxyl group of the 5' exon attacks the 3'SS phosphodiester bond resulting in cleavage of the lariat intron and ligation of the neighbouring exons (Figure 9A).

The assembly of an active spliceosome starts with recognition of the 5'SS, BPS and 3'SS by U1 snRNP, SF1 and U2AF respectively, giving rise to the E complex. The U2 snRNP then binds to the BPS, which leads to release of SF1 and formation of the A complex[30]. Next the tri-snRNP, composed of U4, U5 and U6 snRNPs, binds to the pre-mRNA forming the catalytically inactive B complex[67]. U6 snRNA displaces U1 snRNA from the 5'SS, which then results in the release of the U1 snRNP. BRR2 helicase catalyses the separation of the U4/U6 snRNA duplex and the U4 snRNP is released. The PRP19-CDC5L and NTR proteins also bind to the spliceosome, leading to the formation of the B^{act} complex. PRP2 catalyses conformational changes of the spliceosome which results in the transition to the catalytically active B* complex. The spliceosome then catalyses the first transesterification reaction, yielding the C spliceosome complex which contains the lariat exon and cleaved 5' exon. Further conformational rearrangements catalysed by PRP16 helicase lead to the active C* complex[51]. The spliceosome then catalyses the second transesterification reaction, leading to cleavage of the lariat intron and exon ligation (P complex). PRP22 catalyses the dissociation of the spliced mRNA product from the P complex. This yields the ILS complex, which next dissociates releasing the lariat intron and the spliceosome components that are then recycled[51].

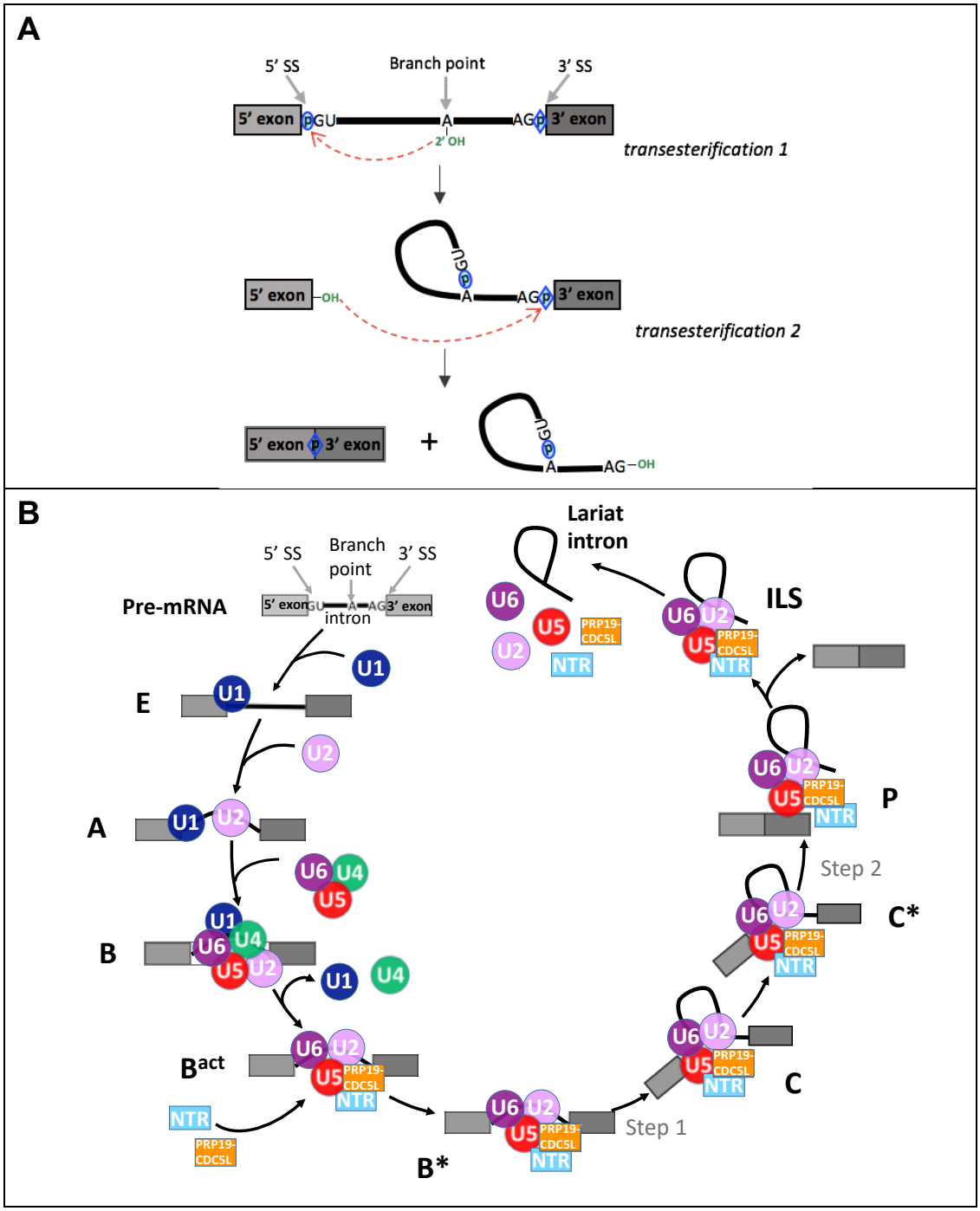


Figure 9 Pre-mRNA splicing by the major spliceosome. (A) The spliceosome catalyses two transesterification reactions to remove an intron. In the first transesterification reaction the 2' hydroxyl group of the conserved branch point adenosine attacks the 5' splice site phosphodiester bond, yielding the lariat intron and cleaved 5’ exon. In the second reaction, the 3' hydroxyl group of the 5' exon attacks the 3' splice site phosphodiester bond causing cleavage of the lariat intron and ligation of the adjacent exons. The phosphodiester linkages (blue oval / diamond) and key hydroxyl groups (green text) are indicated. (B) The spliceosome assembles in a stepwise manner on pre-mRNA substrates and performs two transesterification reactions (step 1 and step 2) to remove an intron. Components include small nuclear ribonucleoproteins (U1, U2, U4, U5 and U6) and large protein assemblies called the PRP19-CDC5L complex and PRP19 related proteins (NTR). Adapted from [30,67].

In addition to the major spliceosome (Figure 9), 700-800 genes in humans contain an intron which is spliced by the minor or U12 dependent spliceosome. U12 introns show key differences compared to introns spliced by the major spliceosome, including lack of a polypyrimidine tract and divergent 5'SS and 3'SS. The overall formation of the minor and major spliceosomes are related in terms of the components, assembly steps and mechanism. However, the minor spliceosome contains unique snRNPs (U11, U12, U4atac and U6atac), which are similar in function to snRNPs of the major spliceosome (U1, U2, U4 and U6 respectively). The U5 snRNP is common to both the minor and major spliceosomes. Unlike assembly of the major spliceosome, during assembly of the minor spliceosome the U11 and U12 snRNPs bind to the pre-mRNA together as a preformed complex. The splicing of U12 introns is also significantly slower than major introns. U12 introns are found in only a small group of eukaryotes and are absent in yeast and nematodes[68]. Minor introns are enriched in genes with the following functions: DNA replication, DNA repair, transcription, RNA processing, translation and cell cycle control[69,70].

1.3.2 Splicing regulation

Correct spliceosome assembly requires a large number of additional proteins, which are either essential for spliceosome function or play a modulatory role. In fact, over 170 proteins have been isolated from different stages of the spliceosome, during assembly and catalysis[71]. Recognition of splice sites is complex and *cis*-acting nucleotide sequences within both the exon and intron determine the overall strength of a splice site. In addition, trans acting splicing regulatory proteins bind to intronic or exonic sequences in the pre-mRNA further enhancing or silencing splicing at a particular site. The many proteins involved in recognising a splice site ensures accuracy, while the collective effect of the many weak interactions allows flexibility in splice site usage. It is this flexibility in splice site usage which makes both alternative splicing and spatial or temporal regulation of splicing possible[30].

Alternative splicing (AS) allows the inclusion or exclusion of certain regions in the mRNA transcript, which can lead to a range of protein or noncoding RNA products for a given gene (see Figure 10). More than 95% of human genes are subject to alternative splicing and each gene has on average seven alternative mRNA products[72]. In addition to *trans* acting splicing factors and *cis* acting pre-mRNA regulatory sequences, further processes within the cell

can also influence alternative splicing. Both the speed of pre-mRNA elongation by RNA polymerase and the epigenetic signature of the gene have been shown to affect splicing. Furthermore, even after spliceosome assembly has commenced, splicing may either proceed, pause or be discarded[73].

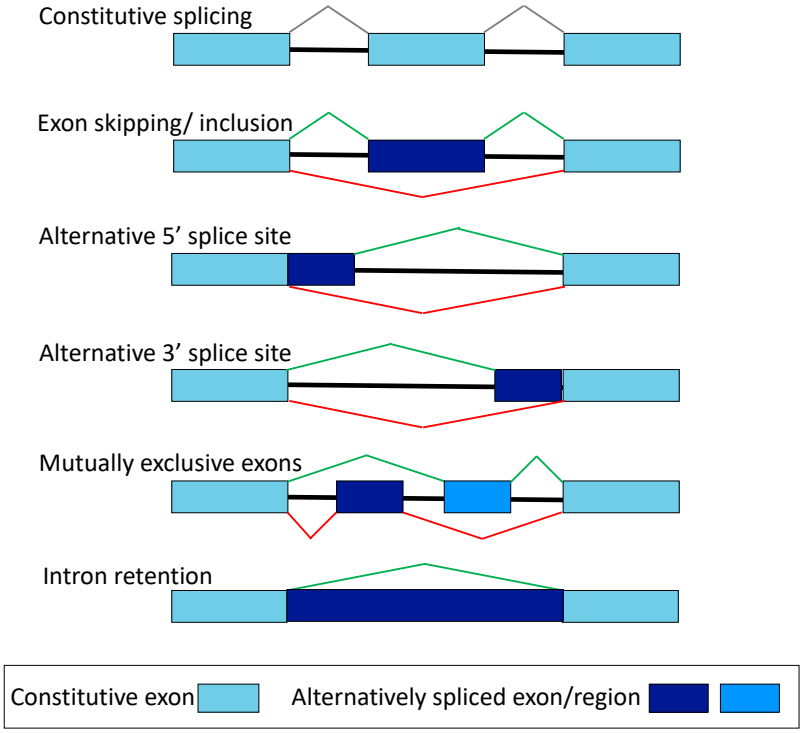


Figure 10 Categories of alternative splicing. Light blue boxes represent constitutively spliced exons and mid/ dark blue boxes represent alternative exons.

1.3.3 Splicing in humans vs yeast

Despite conservation of pre-mRNA splicing machinery from humans to yeast; there are many key differences between splicing in these organisms. In humans, the 5' SS, BP, polypyrimidine tract and 3' SS sequences are more degenerate compared to the highly conserved sequences in the yeast, S. cerevisiae; allowing for more flexibility in human splice site usage. Only 5% of S. cerevisiae genes contain introns and these are all constitutively spliced; whereas >95% of human genes contain introns which are alternative spliced. In addition, the average length of introns is approximately 3000 nt in humans compared to 270 nt in S. cerevisiae[72,74-76]. This means that the human spliceosome has to overcome many additional challenges, in order to ensure splicing fidelity and flexibility for alternative splicing. Many core spliceosome proteins are conserved from humans to yeast (S. cerevisiae); however, many

human spliceosome proteins include extra unstructured regions[47,49,77]. In addition, the composition of human spliceosomes involves many extra proteins compared to S. cerevisiae. These proteins include proteins of the hPrp19/CDC5L complex (NTC), specific Prp19- related proteins (NTR) and further splicing factors[49]. It is thought that these differences provide further flexibility of the spliceosome allowing spatial and temporal regulation of splicing and alternative splicing[30,47,78].

1.3.4 Current understanding of PPIL1 in the spliceosome

PPIL1 is present in purified human spliceosomes at stoichiometric levels; as analysis of spliceosome components using 2D electrophoresis ranked PPIL1 as highly abundant and at similar levels to other known core protein components[71]. It has also been shown that PPIL1 remains associated with the U5 snRNP in complex C even in 1M NaCl, suggesting that PPIL1 is stably bound to the spliceosome rather than engaging in a transient interaction with the spliceosome[79]. PPIL1 belongs to the NTR, a non snRNP sub-complex of the spliceosome, which is recruited to the spliceosome prior to activation (during formation of the B^{act} complex), remaining in the spliceosome until the intron lariat is released (after ILS complex; see Figure 9)[30,49,50]. (Compared to other subcomplexes within the spliceosome, proteins of the NTR make few direct contacts with each other in the spliceosome and instead interact indirectly through connections with PRP19/CDC5L proteins[48].)

PPIL1 was also detected in the 35S U5 complex, a spliceosome dissociation complex which contains PRP19/CDC5L and NTR proteins in addition to U5 snRNP components[80,81]. This shows that PPIL1 remains associated with spliceosome proteins even after initial stages of spliceosome disassembly commence. PPIL1 is then released from binding partners in the spliceosome before the U5 snRNP is recycled and enters a new spliceosome cycle. This is apparent from mass spectrometry analyses of the 20S U5 snRNP and tri-snRNP which both lack PPIL1[80]. It is unclear what leads to these subsequent dissociation steps, and causes proteins such as PPIL1 to be released from the U5 snRNP.

PPIL1 is one of nine cyclophilin proteins found in the spliceosome (Table 2). Due to the distribution of cyclophilins across different phases of spliceosome activation it is therefore likely that they have distinct roles in this process

(Table 2)[46]. Cyclophilins contain a PPIase active site and it is therefore proposed that these proteins may either catalyse proline isomerisation of spliceosome proteins or act as chaperones, binding to and stabilising proline rich regions of spliceosome components[43]. It is thought that the presence of cyclophilins across the spliceosome may assist with the many conformational changes of the spliceosome required for activation and splicing[30]. However, two spliceosome cyclophilins (PPIL2 and CWC27) do not display PPIase activity on the model PPIase substrate (suc-AGPF-pNA peptide)[39]. The function of spliceosomal cyclophilins is not limited to the PPIase active site and these proteins also interact with spliceosome proteins on surfaces distinct to their cyclophilin active site (for example cyclophilins play an important role linking various spliceosome subunits)[46,51].

1.3.5 Findings from splicing assays

Splicing assays can be a valuable tool for understanding the role of specific proteins in the spliceosome. Techniques used to study splicing include: *in vitro* splicing assays, cellular minigene assays or global splicing analysis. *In vitro* splicing reactions typically use radioactively labelled reporter gene constructs, incubated with HeLa nuclear extracts, which are enriched in spliceosome components. The products and intermediates of splicing can then be visualised by gel electrophoresis and autoradiography[82]. Transient transfections of minigene constructs can also be used to study splicing of reporter constructs in cell lines. The resulting mRNA is amplified using RT-PCR and analysed using agarose gel electrophoresis[83]. Alternatively, analysis of endogenous splicing events can be monitored using RT-PCR with primers flanking alternative splice sites[84]. Splicing can also be studied on a global scale in the relevant cell or tissue; this was previously carried out using microarrays with probes for splicing junctions but is now usually carried out using RNA sequencing. RNA sequencing offers several advantages including the ability to detect novel AS events and also provide quantitative data[85].

Adams et al. investigated the role of cyclophilins in the spliceosome, by adding excess of cyclophilin proteins to *in vitro* splicing assays. AdML and β-globin reporter gene constructs were used in the *in vitro* splicing assays. Addition of excess PPIL1 caused significant reduction of splicing in these assays with the β-globin reporter gene but no significant effect on splicing rates in the AdML (adenovirus major late) splicing reporter assay, suggesting

that PPIL1 may modulate splicing in a substrate dependent manner[46]. Experiments in which active site residues of spliceosome cylophilins were mutated to inactivate PPIase activity, show varying effects across the nuclear cyclophilins in the AdML splicing reporter assay. Inactivation of the PPIL1 active site caused only a small decrease in splicing for PPIL1.The authors concluded that cyclophilin PPIase sites may not be required for function of the spliceosome[46]. However, these conclusions are based on the assumption that if cyclophilins have a function in the spliceosome, adding excess cyclophilin protein will perturb spliceosome activity; therefore, these results may not reflect native conditions.

In a further study of splicing, genes encoding core spliceosome components and splicing factors, but also genes encoding proteins with roles in RNA processing or chromatin remodelling were knocked down using siRNA. Splicing of 38 alternative splicing events known to be important for cell proliferation or apoptosis were monitored in HeLa cells using RT-PCR and networks were created based on proteins which elicited similar effects. Interestingly PPIL1 is found in a region of the network which is enriched for helicases. Specifically, PPIL1 forms connections with DHX57 and DDX31 (or PPP1R25), both of which are probable ATP-dependent RNA helicases which form part of the mRNA bound proteome (DDX57 is found in spliceosome C complexes)[53,84]. In addition, other cyclophilins (PPIG, PPIL2, PPIE, PPIH) are dispersed in the network rather than clustering in a single region; which further supports the hypothesis that cyclophilins have specific roles across the spliceosome. Specifically, PPIL1 knockdown had a significant effect on either inclusion or skipping for 12 out of 40 studied splicing events (FAS, PAX6, RAC1, CASP2, MAP4K2, CASP9, DIABLO, H2AFY, BIM (EL), BIRC (D3), SMN2, and PHF19). Overall knockdown of splicing factors affects splice site selection instead of causing widespread inhibition of splicing, and PPIL1 knockdown causes both skipping and retention in different alternative splicing events[84].

1.3.6 SKIP- PPIL1 interaction

PPIL1 interacts with the spliceosomal binding partner, SNW domain-containing protein 1 (SNW1); also known as SKI-interacting protein (SKIP). SKIP is an essential spliceosome component which is also part of the NTR sub-complex[30]. GST-pulldown assays demonstrated that the PPIL1 binding region of SKIP is between residues 59-129[55]. PPIL1 binds to SKIP with

medium affinity, with a dissociation constant of 125 nM, as determined by SPR[55]. NMR studies show that a disordered region of SKIP (59-79) binds to a surface of PPIL1 distant to the active site, forming a hook like structure on PPIL1's surface[55,86]. The minimal PPIL1 binding fragment (PBF) of SKIP was narrowed down to residues 61 to 68, using peptide arrays[56]. The PBF region of SKIP is conserved from human to *S. pombe*; interestingly S. cerevisiae lacks both this region of the SKIP homolog and any PPIL1 homolog[86].

It was suggested that the disorder to order transition of SKIP which takes place upon binding to PPIL1 may play a role in the significant structural reorganisation required for spliceosome activation[86]. This interaction also leaves the PPIase site of PPIL1 available to potentially act as a PPIase or molecular chaperone which may further aid the spliceosome structural rearrangement[86]. SKIP is a predominantly disordered protein, in isolation and in the spliceosome it is able to adopt an extended conformation, interacting with many spliceosomal components (including PRP17, PPIL1, RBM22, SYF3, PRP8, SF3b155, U2 and U6 snRNA)[47-50,52,77]. Due to the high connectivity of SKIP across the spliceosome, it is thought that SKIP may be a key player in coordinating many dynamic transitions required for catalytic activation of the spliceosome[48,50]. The presence of SKIP in the spliceosome is essential for splicing and it has also been shown that SKIP has a role in transcription regulation and may act to couple these events[87]. PPIL1 does not have a clear nuclear localisation signal (NLS)[61] and in rice, it has also been demonstrated that SKIP enables translocation to the nucleus of the rice PPIL1 homolog Cyclophilin 18-2[88], although this has not been shown in mammalian cells.

1.3.7 Insights from cryo-EM structures

Recent advances in cryo-electron microscopy (cryo-EM) have made it possible to determine the structure of intact spliceosomes at various stages of assembly, at near atomic level resolution. This was made possible by improvements in the sensitivity of electron microscopes, due to the use of direct electron detectors in the microscopes, which allowed for increased signal to noise ratios in the data collected. In addition, improvements in software have improved resolution of structures by allowing for correction of beam-induced sample motions and for particles to be grouped according to similarities in conformation[65].

The spliceosome is a highly dynamic complex, with significant changes in composition and conformation, and this heterogeneity also meant it was previously difficult to obtain high resolution structures. Advances in experimental techniques to stall spliceosomes at specific stages of assembly have therefore been essential for solving spliceosome structures[78,89]. Most human spliceosome complexes are assembled using *in vitro* splicing reactions (as described above), with model pre-mRNA substrates (such as AdML or MINX model substrates) and nuclear cell extract from human cell lines (e.g. HeLa cells). A frequently used approach uses pre-mRNA substrates fused to an MS2 sequence, to allow affinity purification of the complex followed by sucrose gradient sedimentation[90]. Different methods have been used to stall spliceosomes at various stages of assembly, for example: reduction of $MgCl_2$ (from 2.5-3.5 mM to 0.3 mM) to obtain the B complex[91], addition of the splicing inhibitor BN82685 to assemble C complexes (by inhibiting exon ligation) and lowering of pH (7.9 to 6.4)[52] or 3'SS mutation (AG to GG)[47] to stall the complex at stage C*.

The first spliceosome cryo-EM structure was determined for *S. pombe* ILS complex[48] and since 2015 many further spliceosome structures have been deduced for both human and S. cerevisiae complexes. S. cerevisiae does not contain a PPIL1 homolog; however, PPIL1 is present in many human cryo-EM spliceosome structures[47,49,50,52,77] and the *S. pombe* structure[48], offering important insights into the role of PPIL1 in the spliceosome. Specifically, human spliceosome structures determined by cryo-EM include: the pre-B complex[92], B complex[91], a series of B^{act} complexes [49,50], the C complex[52,77], the C* complex[47,52] the P complex [93,94] and ILS complexes[94]. PPIL1 is stably integrated in all structures except for the pre-B complex, B complex and early B^{act} complexes. PPIL1 is also stably integrated with *S. pombe* ILS complex[48]. This data suggests that PPIL1 is recruited between B and B* stages, remaining in the spliceosome until after splicing is complete; and is consistent with data from previous mass spectrometry analyses [30,47,49,50,52,77,80,81,91].

PPIL1 is recruited during the transition to B^{act} spliceosome along with other NTR components[49,50]. Purification and cryo-EM of B^{act} complexes shows that this complex is dynamic and a series of conformational states have been identified[49,50]. During the data analysis, Zhang et al. divided B^{act} complex into three main conformational states and showed that PPIL1 is present in mature and late B^{act} complexes but not in the early complex[49]. Haselbach et al. also determined the structures of eight major sequential B^{act}

conformations. PPIL1 was shown to be stably integrated at step 2, at a point where the PRP19 helical bundle moves upwards in the complex[50]. PPIL1 is also shown stabilising the PRP19 bundle later in assembly at complex C*[52]. Interestingly, the region of the spliceosome into which PPIL1 is recruited has been shown to exhibit a large degree of conformational flexibility[50]. (Further insights into PPIL1 function in the spliceosome, gained from modelling of these structures is discussed in Chapter 3).

1.3.8 Spliceosome mutations and disease

Aberrations in splicing are well documented as the cause of many diseases, such as cancers and genetic disorders[95,96]. *Cis* acting splicing mutations currently account for approximately 10% of all known inherited disorders[95], a number which is expected to increase as more mutations are discovered[97,98]. *Trans* acting mutations of spliceosome components have also been shown to cause disease and cause a wide range of symptoms. For example, autosomal dominant retinitis pigmentosa is caused by mutations in numerous splicing related proteins (PRPF3, PRPF6, PRPF8, PRPF31, RP9 (PAP1), and SNRNP200); interestingly, all proteins are associated with the tri-snRNP [99]. Mutations of SNRPA, a component of the U1 snRNP, causes a syndrome of intellectual disability and craniofacial abnormalities [100]. Disease causing splicing mutations affect a range of components including also those of the minor spliceosome [70], snRNA[70] and in proteins that allow snRNP maturation [101,102].

More recently mutations in spliceosome components have been associated with genetic disorders which affect the brain; for example mutations in the essential spliceosome helicase (DHX16 and DHX38) cofactor, GPKOW, was shown to cause severe X linked microcephaly and intrauterine growth restriction[103]. In addition, haploinsufficiency of EFTUD2, a U5 snRNP GTPase, causes mandibulofacial dysostosis with microcephaly[104]. Heterozygous and homozygous mutations in RNU4ATAC, which encodes the minor spliceosome snRNA U4atac, have also been shown to cause microcephalic osteodysplastic primordial dwarfism type I[105]. In addition to spliceosome defects, mutations in components of further RNA processing and regulation complexes have also been implicated in brain developmental disorders. Mutations in the tRNA splicing complex (TSEN2, TSEN34, TSEN54) and a polynucleotide kinase (CLP1) also involved in tRNA splicing, have been shown to cause pontocerebellar hypoplasia and microcephaly[106-

[108]. The RNA exosome complex is required for RNA processing and degradation and mutations in the components (EXOSC3 and EXOSC8) also lead to a syndrome of pontocerebellar hypoplasia[109,110].

It is interesting that the pathological phenotype of mutations in ubiquitously expressed spliceosome components is often restricted to specific tissues. It is suggested that particular splicing defects lead to reduction of mature mRNAs for specific groups of genes that are particularly important for survival for specific cell types[111]. This is supported by the observation that knockdown of spliceosome components causes specific splicing defects rather than a general inhibition of splicing[84]. Alternatively, incomplete splicing may lead to accumulation of products which are toxic to specific cells; such as novel transcripts or protein isoforms. Another hypothesis is that certain tissues may be more susceptible to splicing defects due to an overall higher requirement of the splicing machinery[112]. For example mutations in spliceosome components often affect the retina[99], which has a high demand for splicing, as is reflected by the highest expression of spliceosome snRNAs compared to other tissues[112]. The brain, which is also affected in a large number of splicing disorders, has particularly high levels of alternative splicing and contains the highest number of tissue specific alternatively spliced transcripts[113].

1.4 Aims and objectives

The aim of this thesis is to investigate PPIL1 patient variants as a novel cause of microcephaly and pontocerebellar hypoplasia (section 1.1.4). The effect of the patient variants on PPIL1 protein function, will be studied in order to confirm the deleterious effect of the variants and gain further insights into the role of PPIL1.

The main objectives are:

1. To use *in silico* techniques to: assess the likely pathogenicity of the identified PPIL1 variants and gain a deeper understanding of the role of PPIL1 in the spliceosome.
2. To express and purify PPIL1 WT and mutant proteins to investigate the effect of the PPIL1 variants on: protein folding and stability, protein-protein interactions and enzyme activity of PPIL1.
3. To gain an insight into the role of the PPIL1 active site in the spliceosome. PRP17 was investigated as a putative PPIase substrate of PPIL1.

Due to the limited statistical power to confirm variant pathogenicity in rare genetic diseases, use of *in silico* and experimental approaches are important lines of evidence for confirming variant pathogenicity. Further investigating the role of PPIL1 will not only shed light on the mechanism of this disease, but also the mechanism of the spliceosome. Incorrect splicing is implicated in a large range of diseases, including many cancers and a wide range of genetic diseases, therefore knowledge of the spliceosome has wide applications to the understanding of many diseases[114,115].

Chapter 2 Methods

2.1 General molecular biology methods

2.1.1 Reagents

All general laboratory reagents were purchased from Sigma or Fisher, unless otherwise stated.

2.1.2 Gel electrophoresis

2.1.2.1 Agarose gel electrophoresis

Agarose gels were prepared with 1x TAE buffer (see Ogden *et al.*[116]), and agarose at a concentration of 0.7-1.5% w/v (depending on the DNA size to be resolved; Table 3). The solution was heated to dissolve the agarose and allowed to cool slightly before adding ethidium bromide (0.5 µg/ml). The solution was poured into a casting tray, with a gel comb and allowed to set. Samples were loaded with 6x loading buffer (Thermofisher) and Easyladder 1 (Bioline) was used to estimate size of separated DNA fragments/ PCR products. Gels were run at 80-100 V for 60 to 90 minutes, in TAE buffer, and visualised using a UV transilluminator.

Table 3 Percentage agarose gel required (w/v) for resolution of DNA fragments in agarose gel electrophoresis

Percentage agarose gel	DNA size resolution
0.7 %	800 bp to 12 kb
1.0 %	500 bp to 10 kb
1.2 %	400 bp to 7 kb
1.5 %	200 bp to 3 kb

*(*bp = base pairs, kb = kilo base pair)*

2.1.2.2 Polyacrylamide gel electrophoresis

12% SDS-PAGE gels, with a stacking layer, were made using Protogel reagents (National Diagnostics), in the quantities detailed below (Table 4). The ammonium persulfate and TEMED were added last to start the cross-

linking reaction, and the solution was immediately poured between the assembled glass gel casting plates. The separating layer was prepared first and allowed to set before pouring the stacking layer gel on top; this was left to set with the gel comb inserted. Samples were prepared using loading buffer (see Laemmli *et al*. [117]) and 50 mM dithiothreitol (DTT) and heated at 95 ^{o}C before use. Gels were run in glycine SDS-PAGE running buffer (see Laemmli *et al.* [117]) at 200 V for 50 minutes or until the Bromophenol blue dye front reached the bottom of the gel. Gels were stained with coomassie blue (Table 5) for 3 hours and placed in de-stain solution (Table 5) for a further three hours before imaging.

Table 4 SDS-PAGE gel reagents for 2x gels with 12% separating gel and stacking layer

Reagent	Stacking layer	12 % separating layer
dH_2O	4.20 ml	5.25 ml
30 % w/v acrylamide *(30 % acrylamide, 0.8 % bisacrylamide)*	0.65 ml	6.00 ml
4x resolving Tris solution *(1.5 M Tris.HCl pH 8.8, 0.4 % SDS)*	-	3.75 ml
4x stacking Tris solution *(0.5 M Tris.HCl pH 6.8, 0.4 % SDS)*	1.60 ml	-
10 % w/v ammonium persulphate	67 µl	150 µl
TEMED	6.7 µl	15 µl

Table 5 Reagents required for Coomassie blue or de-stain solution

Reagent	Coomassie blue solution	De-stain solution
Methanol	500 ml / l	100 ml / l
Acetic acid	100 ml / l	100 ml / l
Coomassie blue powder	1 g/ l	-

2.1.3 Sequencing

All sequencing reactions were prepared in 0.2 ml thin walled tubes using 10 µl volumes, 150 ng DNA, 0.2 µM primer, 0.5 µl big dye 3.1 (Applied Biosystems) and 2 µl 5x sequencing buffer (Applied Biosystems). Samples were incubated in the thermocycler (Dyad Peltier Thermal Cycler, Bio-Rad) for 1 minute 96 °C and 25 cycles of (96 °C 10 seconds, 50 °C 5 seconds and 60 °C 4 minutes), with a temperature ramp between steps of 1 °C/sec. Sequencing products were purified by ethanol precipitation. 10 µl of sequencing products was added to 5 µl of 125 mM EDTA and 60 µl 100% ethanol in a 96-well plate. Samples were spun at 2900 xg for 30 minutes, then inverted on tissue and spun at 10 xg for 15 seconds. 60 µl 70% (v/v) ethanol was added to the sample which was placed at 780 xg at 4 °C for 15 minutes. The plate was again inverted on tissue and spun at 10 xg for 15 seconds. Residual ethanol was removed by placing the plate on a heated block at 95 °C for 1 minute. 10 µl Hi Di Formamide (Applied Biosystems) was added to the samples prior to analysis. Samples were analysed using fluorescence-based capillary electrophoresis on the ABI 3130xl Genetic Analyzer (Applied Biosystems). Sequencing results were analysed using 4Peaks to visualise the electropherogram traces.

2.2 *E. coli* cell culture and cloning

2.2.1 Growth media

Transformed *E. coli* cells were grown on LB agar (Table 6) or in LB broth (Table 6); except for large scale expression of recombinant proteins where 2YT broth (Table 6) was used. Media was made as detailed below (Table 6) and sterilised by autoclaving. Ampicillin was added at a final concentration of 100 µg/ml to all media.

For expression of ^{15}N labelled protein, M9 media supplemented with ^{15}N ammonium chloride was prepared using the reagents in Table 7. 50 ml 10x M9-N salts (Table 8) and 2 ml 25% (w/v) ^{15}N NH_4Cl (Cambridge Isotope Laboratories) were added 430.55 ml dH_2O. The pH of the solution was adjusted to 7.4 and autoclaved. All other components were sterilised individually and subsequently added separately to the media in the quantities detailed in Table 7.

Table 6 Media details

Reagent	LB broth	LB agar	2YT broth*
NaCl	10 g/l	10 g/l	5 g/l
Yeast Extract	5 g/l	5 g/l	10 g/l
Tryptone	10 g/l	10 g/l	16 g/l
Agar	-	12 g/l	-

** The pH of the final solution was adjusted to 7.5 using NaOH.*

Table 7 Full list of reagents used for 500 ml M9 media

Reagent	*Details*	Autoclaved	Filter sterilised	Volume added
10x M9-N salts	See Table 8	✓		50 ml
25% (w/v) ^{15}N NH_4Cl	*Cambridge Isotope Laboratories*		✓	2 ml
1M $MgSO_4$		✓		1 ml
1M $CaCl_2$		✓		50 µl
20% w/v Glucose		✓		10 ml
5% (w/v) FeCl(III)			✓	400 µl
MEM vitamins (100x)	*Sigma #M6895*		✓	5 ml
Trace elements (1000x)**	See Table 9		✓	0.5 ml
Ampicillin (100 mg/ml)			✓	0.5 ml

Table 8 10x M9-N salts composition

Reagent	10x M9-N salts
Na_2HPO_4 (anhydrous)	6 g/l
KH_2PO_4	3 g/l
NaCl	0.5 g/l

Table 9 Reagents required for 50 ml 1000x trace elements solution

Reagent	
1 X 10^{-5} M $ZnSO_4$	3 X 10^{-5} M $CoCl_2$
8 X 10^{-5} M $MnCl_2$	4 X 10^{-4} M H_3BO_3
1 X 10^{-5} M $CuSO_4$	3 X 10^{-6} M $(NH_4)_6(Mo_7)$

*** 1000x trace elements solution was a gift from Gary Thomson, (previously University of Leeds).*

2.2.2 Plasmids

PPIL1 pET22b+ expression vectors were a gift from Guoliang Chai, UCSD. (Full length human PPIL1 cDNA was inserted into the vector, in frame with the C terminal His tag, using NdeI/XhoI restriction sites). Four vectors were received which contained WT and mutant PPIL1 (A99T, T107A and R131Q) open reading frames (ORF). In addition, a pDONR221 plasmid containing the full length SKIP ORF was also received.

2.2.3 In fusion cloning

A DNA fragment coding for amino acid residues 59-129 of SKIP (the PPIL1 binding region) was cloned into a pGEX-6P1 vector using the In-Fusion HD Cloning Kit (Clontech), as per manufacturer's instructions (Figure 11). For PCR amplification of SKIP 59-129 coding sequence, primers were designed, using the on-line In-Fusion Primer design tool [http://www.clontech.com/US/Products/Cloning_and_Competent_Cells/Selection_Guides/Online_In- Fusion_Tools] (forward: GAATTCCCGGGTCGAGGAGATGGAGGTGCTTTTC, reverse: AGTCACGATGCGGCCTTAATCATCTGCATTCATAACCTC). Primers

contain a gene specific 3' region (black) and a 5' region (blue) containing 15 bases with homology to the pGEX-6P1; a stop codon (underlined) was also included in the reverse primer. SKIP 59-129 was amplified from SKIP pDONR221 vector, using CloneAmp HiFi PCR Premix (Clontech) and the following conditions: 98 °C 30 seconds, 25 cycles: (98 °C 30 seconds, 55 °C 30 seconds, 72 °C 25 seconds) and 72 °C 7 minutes. The vector was digested with SalI and NotI restriction enzymes and the linearised vector was gel purified using a MinElute gel extraction kit (Qiagen). Insertion of the PCR product into the linearised vector was carried out as described in the in-fusion kit protocol. The resulting plasmids were then transformed into *E. coli* and validated as below (section 2.2.5).

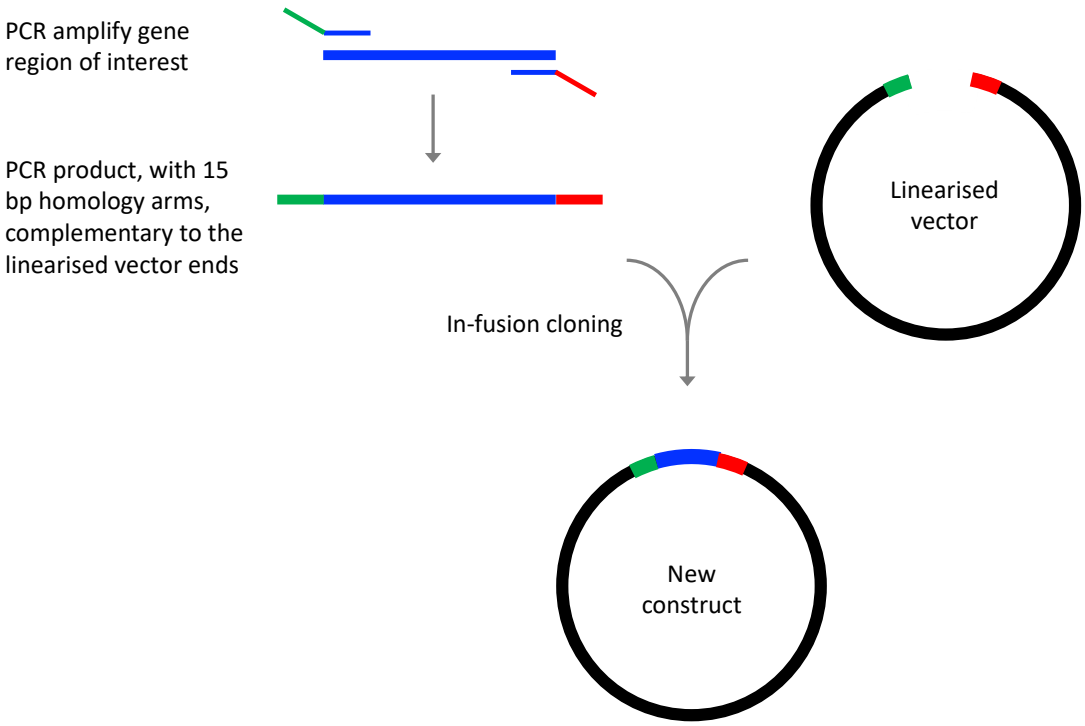


Figure 11 In-fusion cloning protocol, Clontech Laboratories. The gene region of interest is amplified using primers with a gene specific 3' region of 18-25 bp and a 5' 15 bp region complementary to the ends of the linearised vector (red/green). The (cloning enhancer treated) PCR product is inserted into the linearised vector using homologous recombination.

2.2.4 Site directed mutagenesis (SDM)

The Q5 Site-Directed Mutagenesis Kit (NEB) was used, as described in the kit protocol, to introduce the new patient mutation, c.301_318dup

p.(A101_D106dup); c.325G>T p.(G109C), into the pET22b+ PPIL1 expression vector. The first round of SDM was carried out using the primers AGATACCAATtGCAGCCAGTTCTTTGTGAC, GGCCCCGCATTGGCCATT to introduce c.325G>T p.(G109C): and the second round of SDM was carried out using the primers gggccagatACCAATTGCAGCCAGTTCTTTGTG, cgcattggcATCTGGCCCCGCATTGGC to introduce the second mutation c.301_318dup p.(A101_D106dup); (lowercase letters in primers represent the position of the mismatch or additional nucleotides required to insert the relevant mutation). (The primers were designed using NEBaseChanger online tool [https://nebasechanger.neb.com/], Figure 12) The exponential amplification step was carried out with the following conditions: initial denaturation 98°C 30 seconds, 25 cycles (98 °C 10 seconds, 69 °C 30 seconds, 72 °C 3 minutes), final extension 72 °C 2 minutes. All other steps were as described in the kit protocol. Site-directed mutagenesis plasmid products were transformed into *E. coli*, amplified and sequenced as below (sections 2.2.5, 2.1.3).

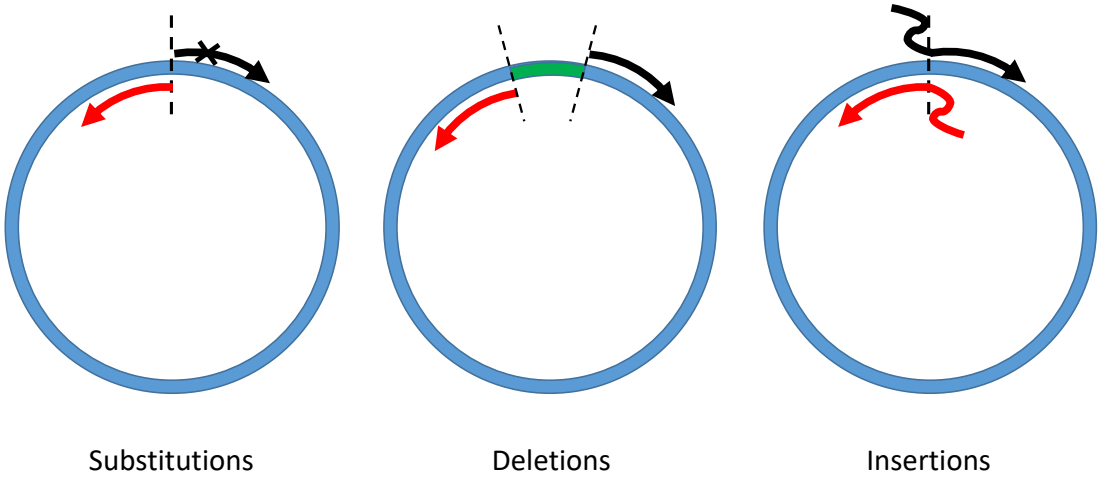


Figure 12 Site directed mutagenesis, using Q5 Site-Directed Mutagenesis Kit. Mutations are created in plasmid DNA, using exponential amplification, with forward (black) and reverse (red) primers. To create base substitutions, the desired nucleotide change (cross) is included in the forward primer. To create deletions, primers are designed which flank the target region to be deleted. For insertions, half of the desired insertion sequence is included at the 5' end of each primer.

2.2.5 Validation of constructs

For cloning, plasmids were transformed into chemically competent cells (Stellar competent cells- Clontech), using heat shock. 1 - 5 µl of DNA (10 pg - 100 ng) was added to 50 µl cells and incubated on ice for 30 minutes, at 42

°C for 45 seconds and on ice for a further 2 minutes. The cells were then incubated at 37 °C, 200 rpm, with 200-1000 µl SOC broth (New England Biolabs) for 1 hour, spread onto LB agar plates (100 µg/mL ampicillin) and incubated at 37 °C overnight. Bacteria from a single colony were inoculated into 5 ml LB cultures (100 µg/mL ampicillin) and grown at 37 °C overnight. Plasmid DNA was isolated using a Miniprep kit (Qiagen; using kit protocols) and constructs were verified by sequencing (see section 2.1.3), using primers in Table 10. If higher DNA yields were required, 200 ml LB cultures were grown at 37 °C overnight (100 µg/mL ampicillin) and DNA was extracted using a Maxiprep kit (Qiagen, using kit protocols); the vectors were again verified by sequencing. Samples were aligned to the predicted vector sequence using EMBOSS needle, pairwise sequence alignment (nucleotide) online algorithm (https://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html), to verify that genes were cloned in the correct orientation in frame with the relevant tag, mutations were in the correct location, where applicable, and that no additional mutations had been introduced.

Table 10 Primers used to sequence plasmids

Primer	Primer sequence	Plasmids sequenced
T7 forward	TAATACGACTCACTATAGGG	pET22b plasmids
T7 terminator	GCTAGTTATTGCTCAGCGG	pET22b plasmids
pGEX forward	GCAAGCTACCTGAAATGCTG	pGEX-6P1
pGEX reverse	GGCAGATCGTCAGTCAGTCA	pGEX-6P1

2.3 Protein methods

2.3.1 Protein expression

Expression vectors were transformed into chemically competent BL21(DE3) *E. coli* cells (Agilent/Stratagene) for protein expression, using the method described above (see section 2.2.5). For expression of unlabelled protein cells were grown in 400 ml 2YT broth (see 2.2.1) and for ^{15}N labelled protein cells were grown in 500 ml minimal media supplemented with ^{15}N ammonium chloride (see section 2.2.1); all media contained 100 µg/ml ampicillin.

Samples were inoculated with 5 ml overnight starter cultures and grown at 37°C, 200 rpm in 2 l flasks. Protein expression was induced at OD_{600} 0.5-0.8 by adding IPTG (Isopropyl β-D-1-thiogalactopyranoside) to a final concentration of 1 mM and incubation was continued at 18 °C overnight for PPIL1 proteins, or for 3 hours 37 °C for GST tagged proteins. Cells were harvested by centrifugation at 4000 xg for 30 minutes, at 5 °C.

2.3.2 Protein purification

Protein purification was carried out using the following steps; with specific buffers, columns and elution conditions appropriate for the protein (Table 11). Cell pellets were resuspended in 30 ml resuspension buffer (Table 11). PMSF, lysozyme (Sigma #62970) and triton X-100 were added to final concentrations of 1 mM, 0.4 mg/ml and 0.2 % (v/v) respectively and incubated at room temperature for 10 minutes. 2 ng/µl DNaseI (Sigma #D5025) and 3 mM $MgSO_4$ were added and the samples were incubated at room temperature for a further 10 minutes[3]. Insoluble material was pelleted by centrifugation at 7500 xg, 5 °C, for 30 minutes and the supernatant was loaded onto an affinity column using a peristaltic pump, at 4 °C. Columns were washed and protein was eluted using an ÄKTA explorer system (Amersham pharmacia biotech) (Table 11), at room temperature. The purity of protein solutions was assessed using SDS-PAGE and the molecular weight of purified proteins was confirmed using time-of-flight mass spectrometry (Astbury Centre Mass Spectrometry facility).

[3] Stock solutions were prepared as follows: PMSF, 100 mM solution in isopropanol; lysozyme, 20 mg/ml in deionised water; triton X-100, 10% v/v in deionised water; $MgSO_4$, 1M in deionised water; DNase I, 1 mg/ml in 10 mM Tris.HCl pH 7.9 150 mM NaCl. Lysozyme and DNase solutions were filtered (0.22 µm filter) and stored at -20 °C as 300 µl and 50 µl aliquots respectively.

Table 11 His-tagged PPIL1 and GST tagged protein purification conditions

	His-tagged PPIL1 proteins	GST-tagged proteins
Column	1 ml HisTrap HP columns (GE Healthcare)	Glutathione sepharose HP 1 ml columns (GE Healthcare)
Resuspension buffer	20 mM NaH_2PO_4, 0.3 M NaCl, 10 mM imidazole, pH 8.0	PBS (pH 7.4)
Elution conditions	imidazole gradient (20 mM NaH_2PO_4, 0.5 M NaCl, 10 mM to 500 mM imidazole, pH 8.0) over 25 column volumes	5-10 ml injection of 50 mM Tris-HCl, 10 mM L-glutathione reduced, pH 8.0
Column flow rate	1 ml/minute	0.5 ml/ minute
Elution fraction size	2 ml	1 ml

2.3.3 Calculation of protein concentration

Protein concentration was calculated by measuring absorbance at 280 nm (A_{280}) using a 1 cm pathlength quartz cuvette in a Genesys 6 spectrophotometer (Thermo Spectronic). A baseline reading was taken with the relevant buffer and samples were diluted so that the A_{280} reading was between 0.1 and 1. Protein concentration was then calculated using the Beer-Lambert law (Equation 1). The molar extinction coefficient (ε) was calculated using ExPASY ProtParam online tool[118], which uses Equation 2 to calculate the A_{280} molar extinction coefficient (M^{-1} cm^{-1}), assuming that all cysteine residues are reduced[118].

Equation 1

$$A = \varepsilon \, l \, c$$

(where A is absorbance, ε is the molar extinction coefficient, l is pathlength (cm) and c is concentration of solution (M).

Equation 2

ε = (number of tyrosine residues x 1490) + (number of tryptophan residues x 5500).

2.3.4 Buffer exchange and protein concentration

Protein solutions were exchanged into the relevant buffer using Spectra/Por 1 dialysis membrane, molecular weight cut off (MWCO) 6-8 kDa (Spectrum Laboratories). Experiments were carried out in 1x PBS (made from 10 x PBS stocks; Fisher), unless otherwise stated (for techniques where specific buffer components must be avoided or optimised to improve signal to noise ratios (e.g. CD, NMR)).

For experiments which required high concentrations of PPIL1 protein, the protein was concentrated using Vivaspin 500 concentrator columns (GE healthcare), with MWCO 3,000 Da.

2.3.5 Size exclusion chromatography (SEC)

Analytical SEC was carried out using a Superdex 75 10/300 gel filtration column (GE healthcare; which is suitable for analysis of 3–70 kDa globular proteins), connected to the ÄKTA pump system (as before). PBS buffer was used for both dialysis and running buffers; buffer was degassed and filtered using a vacuum pump and 0.45 µm filter prior to use on the ÄKTA. 0.25 mg/ml protein samples were loaded onto the column using a 250 µl sample loop and flow rate of 0.6 ml/minute.

A calibration curve was used to estimate protein molecular weight by injecting globular monomeric standards of known molecular weight, from the low molecular weight calibration kit (GE healthcare; #28-4038-41), at the recommended concentration. The curve was plotted using normalised K_{av} values calculated using the equation below (where V_0 is column void volume, V_e is elution volume and V_c is geometric column volume). The void volume was calculated using injections of blue dextran (1 mg/ml) from the low molecular weight calibration kit.

Equation 3

$$Kav = \frac{Ve - Vo}{Vc - Vo}$$

2.3.6 Circular dichroism (CD)

Protein secondary structure content was assessed using far-UV CD spectra (190 - 240 nm). Samples were dialysed extensively against 10 mM sodium phosphate buffer, 50 mM NaF, pH 7.4 and diluted to 0.2 mg/ml, 200 µl total volume for analysis. Spectra were recorded using a 1 mm quartz cuvette in a Chirascan CD spectrometer (Applied Photophysics) at 20 ^{o}C. The following parameters were used: 1 nm increments, scan speed = 1 nm/second and bandwidth = 2.5 nm. Results are the average of two consecutive spectra and were subtracted from the spectra of buffer alone. Data was collected in millidegrees (mdeg) and converted to mean residue ellipticity (θ MRW) for analysis using Equation 4 (where θ is the observed ellipticity [mdeg], MRW = mean residue weight, c = concentration [mg/ml], l is the pathlength [cm]. Secondary structure content was calculated using CDSSTR on the Dichroweb online tool using reference set 7, which is optimised for wavelengths of 190-240 nm. CDSSTR estimates secondary structure using a variable selection method, in which reference proteins which do not fit the data are removed and secondary structure content is determined by singular value decomposition. CDSSTR is recognised as a tool which gives better fits of globular proteins[119].

Equation 4

$$[\theta] MRW = \theta \times MRW^{*} / 10 \times c \times l$$

*(*MRW = molecular mass, Da / number of residues -1)*

2.3.7 Protein unfolding and aggregation screen

Protein aggregation and unfolding were measured simultaneously over a 16-90 °C temperature gradient using an Optim machine (Unchained Labs). Protein samples (9 µl) were loaded into 16-well Optim microcuvette arrays. Protein stability was measured using the barycentric mean wavelength of intrinsic protein fluorescence, to assess protein unfolding. Fluorescence was excited using a laser at 266 nm and emission monitored from 280 nm to 450 nm. Protein aggregation was detected by measuring static light scattering at 266 nm. During the assay, the temperature was increased from 16-90 °C in 1°C steps and the temperature was held constant for 30 seconds before recording fluorescence or static light scattering data. Analysis was performed with Optim Analysis Software (Avacta Analytical). Assays conducted to

compare stability and aggregation of WT and mutant PPIL1 proteins were carried out using protein at 1 mg/ml, in PBS.

Optim protein unfolding assays were also used as a screen, to determine optimal buffer compositions for future experiments. The buffer screen was designed and provided by Iain Manfield, University of Leeds. For the buffer screen PPIL1 WT protein (4 mg/ml), was diluted with the following reagents to give the range of buffer conditions. 1 M buffer stocks were added to stock PPIL1 protein (9 µl protein :1 µl buffer) to give 0.1 M buffer concentrations at a range of pH values (Table 12). For NaCl and glycerol concentration screens PPIL1 protein was diluted 1:1 in 2x stock reagents of NaCl (100-3000 mM) and glycerol (5-60 % v/v) solutions to give 50 -1500 mM NaCl and 2.5-30 % v/v conditions. An additive screen was also carried out using the following reagents: ammonium sulphate, potassium acetate, potassium glutamate, potassium chloride, sodium sulphate, sodium chloride, lithium chloride, magnesium chloride, magnesium nitrate, calcium chloride, lithium nitrate, sodium thiocyanate, lithium perchlorate, arginine. Protein was diluted (9:1) in 1 M additive solutions to give a final additive concentration of 0.1 M. As controls two further samples were monitored: an undiluted stock protein sample in PBS (pH 7.4) and protein samples diluted in water (1:1 or 9:1; to control for the effects of protein dilution).

Table 12 Optim pH screen for protein stability

Condition	Details
Water control	Water only
pH 3.0	0.1M citric acid: sodium citrate pH 3.0
pH 3.5	0.1M citric acid: sodium citrate pH 3.5
pH 4.0	0.1M acetic acid: sodium acetate pH 4.0
pH 4.5	0.1M acetic acid: sodium acetate pH 4.5
pH 5.0	0.1M acetic acid: sodium acetate pH 5.0
pH 5.5	0.1M MES.NaOH pH 5.5
pH 6.0	0.1M MES.NaOH pH 6.0
pH 6.5	0.1M MES.NaOH pH 6.
pH 7.0	0.1M HEPES.NaOH pH 7.0

pH 7.5	0.1M HEPES.NaOH pH 7.5
pH 8.0	0.1M HEPES.NaOH pH 8.0
pH 8.5	0.1M bicine.NaOH pH 8.5
pH 9.0	0.1M CHES.NaOH pH 9.0
pH 9.5	0.1M CHES.NaOH pH 9.5

2.4 Protein interaction assays

2.4.1 Surface plasmon resonance (SPR)

SPR binding experiments were performed on a Biacore 3000 system with a CM5 chip (GE healthcare) at 25 ^{o}C.

2.4.1.1 Amine coupling

Anti-GST antibodies were captured on the chip surface (flow cells two and three) using the amine coupling and GST Capture kits (GE Healthcare). The system was primed with 100 mM sodium acetate running buffer (pH 5.6) and the surface was activated using a 35 µl 1:1 injection of 0.1 M N-hydroxysuccinimide (NHS) and 0.4 M N-ethyl-N'-(dimethylaminopropyl) carbodiimide (EDC; both from GE Healthcare amine coupling kit) for 7 minutes. 30 µg/ml anti-GST antibody (immunosorbent-purified polyclonal goat antibody, GST capture kit) solution was injected over the chip surface at 5 µl/minute for 4 minutes. Unreacted material was eluted using 2x 20 µl, 1 minute injections of 1M NaCl, 75 mM sodium acetate (pH 5.6). Unreacted sites were capped using 35 µl, 7 minute injection of 1 M ethanolamine·HCl (pH 8.5; GE healthcare). Flow cell 1 was used as a reference surface and was treated as above, without protein injections.

2.4.1.2 Experiment design

GST-SKIP 59-129 and GST (control) proteins were immobilised on the anti-GST antibodies by injecting the protein solutions across the surface of the chip in separate flow cells. GST-SKIP 59-129 was immobilised at a density of 700 RU (response units) and GST at a density of 560 RU. PPIL1-SKIP binding was measured by injecting the three PPIL1 proteins across the chip

surface at varying concentrations (15.6 – 2000 nM), using a flow rate of 50 µl/min and injection time of 2.5 minutes. Buffer was washed across the chip for at least 20 minutes after the protein injection was complete and before any subsequent injections. All proteins used were dialysed extensively in PBS prior to use. Running buffer used for injections of GST tagged protein and PPIL1 proteins was PBS with 0.05 % v/v IGEPAL detergent. Data was analysed using Biacore BiaEvaluation software. All experimental data was overlaid, aligned and subsequently subtracted from reference flow cell data to account for any bulk refractive index changes between the running and sample buffer.

2.4.2 Isothermal titration calorimetry (ITC)

A peptide was designed which corresponds to residues 89-101 of PRP17 (Bio-FAPEFGPENPFRT-NH_2. Peptide modifications (N-terminal biotinylation and C-terminal amidation) were used to increase peptide stability and the biotin modification allowed use in other assays (such as SPR). The peptide was synthesised and purified by Peptide Synthetics to a purity of >98 % (as judged by HPLC analysis). ITC was used to measure binding between PRP17 peptide and PPIL1. ITC experiments were performed on Microcal ITC 200, at 25 ^{o}C. PPIL1 protein was extensively dialysed against PBS and the cell was filled with 40 µM protein (200 µl cell volume). Lyophilised PRP17 peptide was dissolved in PBS buffer from the final dialysis step. 2 mM PRP17 peptide was titrated into the cell, using 2 µl 4 second injections, spaced 2 minutes apart. The calorimetric signal was calculated by integration of the heat signal generated after each injection. The data point from the first injection, which is affected by diffusion of the syringe solution during equilibration, was removed (as is standard practice). The resulting binding isotherms were analysed using Microcal Origin 7 software and fit to a one-site binding model, with a fixed stoichiometry of 1 (as is necessary for low affinity interactions[120]).

2.5 NMR methods

2.5.1 NMR theory

Atomic nuclei with odd numbers of neutrons and protons have the property of spin which makes them possible to study by NMR (e.g. ^{1}H, ^{13}C, ^{15}N, ^{31}P,

^{19}F). When a magnetic field is introduced (B_0) these spins rotate around the field (around the z axis), in a process known as Larmor precession. The alignment of spins is either in the same direction of the magnetic field (the low energy state) or in the reverse direction to the field (the high energy state; Figure 13). Overall, the majority of spins will be in the low energy state (although the difference is small), and this creates bulk magnetisation along the Z axis. A radiofrequency (r.f.) pulse is used to rotate magnetisation by a defined angle (e.g. 90 ^{o}). The resulting magnetisation creates a current in the receiver coil which can be detected. This creates a signal which decays with time called the free induction decay (FID). An FID signal is a function of time and this can be converted to a function of frequency by Fourier Transformation, yielding an NMR spectrum.

Not all nuclei in a sample experience the same magnetic field due to a process known as shielding; this results in different chemical shifts for each nucleus in an NMR spectrum. Shielding is caused by electrons surrounding the nucleus. Therefore, larger atoms with more electrons typically exhibit larger chemical shifts. Nuclei of a specific type e.g. (^{1}H) can also exhibit different chemical shifts due to a process known as deshielding, where electrons are pulled away from the nucleus by bonding to electronegative atoms or by hydrogen bonding. This means that each nucleus has a specific position in the spectrum, as a result of the local electron distribution[121].

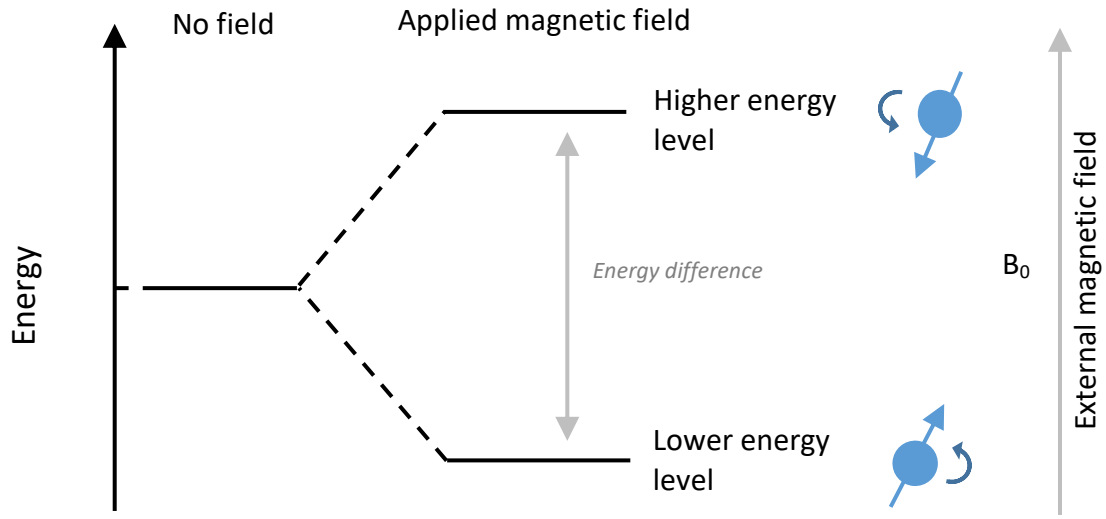


Figure 13 Energy levels for spin ½ nuclei (such as ^{1}H ^{13}C and ^{15}N). These nuclei are charged and have the property of spin which means they behave like a magnet. When exposed to an external magnetic field (B_0) nuclei have two possible orientations. Nuclei with higher energy create a magnetic field in the opposite direction to the external magnetic field, while nuclei with lower energy create a magnetic field in the same direction as the external magnetic field. The difference between the two energy levels depends on the strength of B_0 (the external magnetic field) and the gyromagnetic ratio of a nucleus (which is different for each isotope and is a ratio of its magnetic moment to its angular momentum). (The energy difference between higher and lower energy states also corresponds to the radio wave frequency required to rotate the magnetisation.)

In a simple one dimensional NMR experiment, during the preparation phase, r.f. pulses are applied to excite specific nuclei (e.g. ^{1}H); this is followed by a detection phase during which the resulting FID is measured for a specific time (t_1). The result is a spectrum with a peak for each nucleus occupying a different chemical environment, at different chemical shift frequencies. In peptides and proteins, amide and aromatic protons typically show the highest chemical shift, while aliphatic and methyl groups typically show lower chemical shifts. In two dimensional (2D) NMR experiments, specific pulse sequences are used to transfer magnetisation between nuclei; to detect correlations of nuclei that are connected through bonds (i.e. via J coupling) or through space (using the nuclear overhauser effect ;NOE). 2D experiments can be used to correlate nuclei of the same type (homonuclear experiments) or different types of nuclei (heteronuclear experiments). In 2D heteronuclear experiments magnetisation is typically transferred from more sensitive to less sensitive nuclei (e.g. ^{1}H -> ^{15}N). Nucleus sensitivity is

proportional to both the magnetic field strength (B_0) and the gyromagnetic ratio of the nucleus[121].

2.5.2 General NMR methods

NMR experiments were carried out in collaboration with Dr Arnout Kalverda, Dr Lars Kuhn and Dr Rhys Thomas in the NMR facility (University of Leeds). NMR spectra were acquired on 600 MHz, 750 MHz or 950 MHz spectrometers operating at 600.13, 747.98 and 950.13 MHz resonance frequencies respectively. NMR samples were loaded into standard Wilmad 5mm tubes, 3 mm tubes or susceptibility matched 5 mm Shigemi tubes depending on the experiment. Each spectrometer was equipped with a cryo-probe, Bruker Avance III HD console and pulsed field gradient along Z. Cryo probes used were as follows: triple resonance TXO cryo probe (TXO-CP), triple resonance TCI cryo-probe (TCI-CP) and quadruple resonance QCI-P cryo-probe (QCI-CP) for the 950 MHz, 750 MHz and 600 MHz spectrometers respectively. D_2O was added to all samples at 5% v/v, resulting in a 5% dilution of buffer components and peptide/protein samples. All experiments were conducted at 25 ^oC. Results were processed using NMRPipe[122] and analysed using CCPNMR Analysis[123]; except for 1D ^1H experiments which were analysed using Topspin 3.2 software (Bruker).

2.5.3 Binding experiments

2.5.3.1 PPIL1 and PRP17 peptide

The interaction of PPIL1 and PRP17 was monitored using ^{15}N uniformly labelled PPIL1 protein and unlabelled PRP17 peptide (Bio-FAPEFGPENPFRT-NH_2; as above). ^1H ^{15}N heteronuclear single quantum coherence (HSQC) spectra of ^{15}N PPIL1 were acquired with PRP17 peptide titrations. Samples contained 80 µM ^{15}N PPIL1 and 0 - 400 µM PRP17 peptide, in 10 mM sodium phosphate, 100 mM NaCl, 5% v/v D_2O pH 6.5. Sample volumes of 550 µl were used in a standard Wilmad tube and spectra were acquired using a 600 MHz spectrometer. Spectra were acquired using: *Hsqcfpf3gpphwg* pulse sequence (Topspin 3.2, Bruker), 2048 data points, 8 scans and spectral widths of 9578.544 and 2066.565 Hz for ^1H and ^{15}N respectively.

The published assignment of PPIL1[124] was used to identify the region of PPIL1 involved in binding to PRP17. Chemical shift perturbations were determined using Equation 5 (from Hewitt et al. 2017[125]), which calculates the average distance moved by each peak. Shifts in the proton dimension were multiplied by the gyromagnetic ratio of ^{15}N and ^{1}H nuclei (γ_H/γ_N) to account for the different magnitudes of chemical shifts in the proton and nitrogen dimensions. $\Delta\delta$ represents the total shift from the free state and $\Delta\delta_N$ and $\Delta\delta_H$ represent the shift in the nitrogen and proton dimensions respectively. Residues with significant chemical shift perturbations were defined as those which: a) broadened beyond detection or b) shifted more than one or two corrected standard deviations above the mean (as recommended and detailed by Schumann et al.[126]).

Equation 5

$$\Delta\delta = \sqrt{(\Delta\delta_N)^2 + (\Delta\delta_H)^2(\gamma_H/\gamma_N)^2}$$

2.5.3.2 PRP17 1-172 and PPIL1

^{15}N uniformly labelled protein corresponding to a larger region of PRP17, residues 1-172, was a gift from Guoliang Chai (UCSD). The interaction between PRP17 1-172 and full length PPIL1 was monitored by acquiring ^{1}H ^{15}N transverse relaxation optimized spectroscopy (TROSY) spectra of ^{15}N PRP17 1-172 in the presence and absence of PPIL1. Samples consisted of 30 µM PRP17 1-172 and 0-90 µM PPIL1 in 25 mM sodium phosphate buffer, 100 mM NaCl, 1 mM DTT, 5% v/v D_2O pH 7.0 in a Shigemi tube (300 µl total volume). Spectra were acquired using a 750 MHz spectrometer, 1272 data points, 32 scans and spectral widths of 8971.292 and 1895.006 Hz in ^{1}H and ^{15}N dimensions respectively.

2.5.4 Monitoring PPIase activity

2.5.4.1 Model substrate

The PPIase model substrate, *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide peptide (Sigma #S7388), was used to monitor PPIase activity of PPIL1 proteins. One-dimensional ^{1}H spectra were recorded to measure proline *cis-trans* isomerization of this peptide, using a gradient excitation sculpting sequence for water suppression. The peptide ^{1}H chemical shifts were previously assigned in the literature[127] and Ala2 methyl proton peaks were

used to monitor proline *cis-trans* isomerisation. Samples consisted of 550 µl, 500 µM peptide, dissolved in PBS (pH 7.4), 5% D_2O and where indicated sub-stochiometric concentrations of PPIL1 protein (0-10 µM). As a control, the active site inhibitor CsA (Cyclosporin A; #9973 Cell Signalling Technology), was also added to samples at equimolar ratios to PPIL1. Spectra were acquired using standard 5 mm Wilmad tubes and the 600 MHz spectrometer. The *zgesgp* pulse sequence (Topspin 3.2, Bruker) was used with 32768 data points, 128 scans and a ^{1}H spectra width of 9578.44 Hz. Exchange rates of proline isomerisation were determined using NMR signal lineshape analysis, on the Dynamic NMR (DNMR) module of the TopSpin 3.2 acquisition and processing software (Bruker). Exchange rates were determined iteratively, using fixed nuclei positions and coupling constants for Ala2 methyl *cis*- and *trans*- signals (*N*-succinyl-Ala-**Ala**-Pro-Phe-*p*-nitroanilide).

2.5.4.2 PRP17 substrate

PPIase activity of PPIL1 was monitored on a putative substrate of PPIL1, PRP17. For identification of *cis* and *trans* conformations of PRP17 Gly94-Pro95 peptide bond, a peptide corresponding to residues 89-101 of PRP17 (sequence as above; Ac-FAPEFG**P**ENPFRT-NH_2) was purchased with a ^{15}N and ^{13}C labelled proline, from Peptide Synthetics. The peptide had N-terminal acetylation and C-terminal amidation modifications (to mimic a peptide bond at the ends of the peptide, which increases stability to peptidases). PRP17 peptide was dissolved to 500 µM in 550 µl PBS (pH 7.4), 5 % v/v D_2O and placed in a standard 5 mm Wilmad tube. *Cis* and *trans* states were assigned using a constant time ^{1}H ^{13}C HSQC spectrum, recorded on a 750 MHz spectrometer. Spectra were collected using: *hsqcctetgpsisp* pulse sequence (Topspin 3.2, Bruker), 1024 data points, 4 scans and spectral widths of 11961.723 and 8839.628 Hz in ^{1}H and ^{13}C dimensions. *Cis* and *trans* peptide bond assignment was based on ^{13}C chemical shift for C_γ and C_β proton resonances, as shown by *Shen and Bax 2010*[37].

^{1}H,^{15}N - H(Cα)N ZZ exchange experiments (based on those used by Dujardin *et al.* 2015[128]) were conducted, to detect proline isomerisation of PRP17 Gly94-Pro95 peptide bond. Spectra were recorded using 550 µl 500 µM PRP17 peptide (with labelled proline, as described above) in the absence and presence of catalytic concentrations of PPIL1 (5 µM). The

sample buffer was PBS (pH 7.4), 5 % v/v D_2O, as above, and spectra were acquired using a standard 5 mm Wilmad tube and 950 MHz spectrometer. The spectra were acquired using 2048 data points, 1664 scans and spectral widths of 15243.902 and 192.577 Hz (2 ppm) for ^{1}H and ^{15}N dimensions respectively. The ^{15}N carrier was set to 134 ppm to centre on the proline ^{15}N resonance frequency and 10 complex increments were acquired for a total experiment time of 22.5 hours.

2.5.4.3 Identifying rates of proline isomerisation on PRP17

^{1}H-^{15}N ZZ exchange experiments were used to determine rates of proline isomerisation on PRP17. ^{15}N universally labelled PRP17 peptide was a gift from Guoliang Chai (UCSD). In brief this peptide was expressed in *E. coli* with a SUMO tag and purified firstly using affinity purification. After cleavage of the tag, the peptide was purified using HPLC and lyophilised for storage. PRP17 peptide corresponded to residues 84-101 of PRP17, TYETMFAPEFG**P**ENPFRT, (the peptide was five residues longer than PRP17 peptide used above, for optimal stability during expression and purification steps).

To assign the PRP17 peptide ^{1}H ^{15}N HSQC-TOCSY and ^{1}H ^{15}N HSQC-ROESY experiments were used. (^{1}H ^{15}N HSQC-TOCSY and ^{1}H ^{15}N HSQC-ROESY spectra contain HSQC pulse sequences at the start of the pulse sequence to help separate the large number of signals [which result from the multiple conformations of PRP17 peptide]). Spectra were recorded on the 600 MHz spectrometer using 300 µl sample volumes in a 5 mm Shigemi tube. Sample conditions were as follows: 400 µM peptide, 25 mM sodium phosphate buffer, 100 mM NaCl, 1mM DTT, 5% D_2O pH 7.0. ^{1}H ^{15}N HSQC-TOCSY spectrum (80 ms mixing time) was recorded using: 2048 data points, 80 scans and spectral width of 9578.544 and 1276.404 Hz for ^{1}H and ^{15}N respectively. ^{1}H ^{15}N HSQC-ROESY spectra (mixing times 200 and 300 ms) were acquired with: 2048 data points, 288 scans, and spectral widths of: 6893.382 and 1276.404 Hz for ^{1}H and ^{15}N respectively. 200 complex increments were acquired in ^{15}N for a total acquisition time of 65 hours.

^{1}H-^{15}N ZZ exchange experiments, to monitor proline isomerisation, were carried out using the uniformly ^{15}N labelled PRP17 peptide (details as above). Sample and buffer conditions were as above and spectra of PRP17 peptide (400 µM) were recorded in the absence and presence of 4 µM

PPIL1. Spectra were acquired on the 950 MHz spectrometer, using 3 mm NMR tubes (200 µl volume). The experiment was repeated using mixing times of 40, 80, 120, 200, 250, 350, 450, 600, 750 ms. The order of experiments was randomised and contained two technical repeats. The experiment was based on the *Hsqcexf3gpwgph* pulse sequence (Topspin 3.2, Bruker) and included: 4096 data points, 8 scans and spectral widths of 15243.902 and 2022.029 Hz for ^{1}H and ^{15}N dimensions.

In order to determine rates of proline isomerisation, auto and exchange peak intensities were measured for each experiment. Exchange peak intensities were normalised to the relevant exchange peak intensity at t=0. All peak intensities were then expressed as a fraction of total *cis* and *trans* auto peak intensity at t=0. The data was fit to the two state exchange equations, based on equations by Kleckner et al. 2011[129] (Equation 6, Equation 7, Equation 8), using GraphPad (Prism7). I refers to the peak intensity, at a given mixing time, for *cis* (AA), *trans* (BB), and exchange (AB and BA) peaks. T represents time in seconds, P is the population of *cis* or *trans* conformations, R_1 is the longitudinal relaxation rate constant and k_{ex} is the overall *cis* ↔ *trans* exchange rate per second.

Equation 6

$$I_{AA}(T) = P_A(P_A + P_B \exp(-k_{ex}T))\exp(-R_1T)$$

Equation 7

$$I_{BB}(T) = P_B(P_B + P_A \exp(-k_{ex}T))\exp(-R_1T)$$

Equation 8

$$I_{AB}(T) = I_{BA}(T) = P_AP_B(1 - \exp(-k_{ex}T))\exp(-R_1T)$$

2.6 Other methods

2.6.1 Variant pathogenicity predictions and protein modelling

Polyphen (Polymorphism Phenotyping v2)[130], SIFT (Sorting Intolerant From Tolerant)[131], PROVEAN (Protein Variation Effect Analyzer)[132] and CADD (Combined Annotation Dependent Depletion v.1.4)[133] online software tools

were used to predict the effect of patient PPIL1 variants. Software was available on the following links: http://genetics.bwh.harvard.edu/pph2/, http://sift.bii.a-star.edu.sg/, http://provean.jcvi.org and https://cadd.gs.washington.edu/score respectively. Prediction of variant pathogenicity was made using default parameters for all tools used.

Allele frequencies of patient variants were taken from gnomAD (Genome Aggregation Database; http://gnomad.broadinstitute.org/)[134].

Protein alignments were made using Clustal omega software (https://www.ebi.ac.uk/Tools/msa/clustalo/)[32] and annotated using ESPript ('Easy Sequencing in PostScript') 3.0 online software (http://espript.ibcp.fr/ESPript/ESPript/)[33]

UCSF Chimera software (downloaded from http://www.rbvi.ucsf.edu/chimera)[57] was used to model protein structures and produce images displayed in this thesis. The published data used (Table 13) was downloaded from the relevant entry in the protein data bank (PDB; www.rcsb.org)[135].

Table 13 PDB data used in this thesis

Structure	PDB ID	Reference
PPIL1 crystal structure	2X7K	[56]
PPIL1 NMR structure	1XWN	[55]
PPIL1 and SKIP NMR structure	2K7N	[86]
S. pombe ILS spliceosome	3JB9	[48]
Human C spliceosome	5YZG	[77]
Human C spliceosome	5O9Z	[52]
Human C* spliceosome	5XJC	[47]
Human Bact spliceosome (late)	5Z57	[49]
Human Bact spliceosome (mature)	5Z56	[49]

Human Bact spliceosome (core)	6FF7	[50]
Human P spliceosome	6QDV	[93]
Human P spliceosome	6ICZ	[94]
human ILS spliceosome [before prp43 loaded],	6ID0	[94]
Human ILS spliceosome [after prp43 loaded]	6ID1	[94]
Cyclophilin A crystal structure	1ZKF	[136]

2.6.2 Ethical approval

The families provided signed informed consent to participate in studies approved by the Leeds East Research Ethics Committee (07/H1306/113), see Appendix.

Chapter 3 *In silico* modelling

3.1 Introduction

Identification of causative gene variants in autosomal recessive disease has increased rapidly following the development of whole exome sequencing[137]. However, finding the correct disease causing variant in a background of many variants across the genome remains challenging. As a result, further lines of evidence are required to confirm if an identified variant causes the disease in question. *In silico* tools are frequently used to assess if an identified variant is likely to be pathogenic and may also be used to further investigate the mechanism behind a disease[138].

PPIL1 variants were previously detected as the cause of microcephaly and pontocerebellar hypoplasia, using whole exome sequencing and autozygosity mapping (unpublished data, University of Leeds and University of California San Diego; discussed in Chapter 1). Although many variants were detected in the patients' DNA, variants in PPIL1 were determined to be the most likely pathogenic variants using the following methods. Firstly common variants in dbSNP (2014 version), the Exome Sequencing Project (ESP) project, and in locally sequenced samples from individuals of Pakistani origin were excluded. Only variants that were present in biallelic state in both individuals and that were missense, splice site, frameshift or stop variants were called. After filtering variants as above, two missense variants remained: in RAB44 and PPIL1. The RAB44 variant had a very low CADD score (4.476), whereas the PPIL1 variant had a CADD score of 29.7 suggesting the PPIL1 variant is highly likely to be pathogenic and the RAB44 variant likely benign. Therefore mutations in PPIL1 were considered most likely to be the pathogenic variants (unpublished data, University of Leeds). Sanger sequencing of the patient's and parents DNA confirmed segregation of PPIL1 variant with disease. Independently, a laboratory at University of California San Diego also mapped the pathogenic variant in a case of microcephaly and pontocerebellar hypoplasia to be located in PPIL1 (unpublished data, UCSD).

Combined, the data from the University of Leeds and UCSD provided strong evidence that mutations in *PPIL1* cause microcephaly and pontocerebellar hypoplasia in the patients discussed. However, additional work was required

to provide further support for this hypothesis and *in silico* work was first carried out to investigate the consequences of the patient mutations in PPIL1. All *PPIL1* variants were homozygous and cause the following PPIL1 protein changes: A99T, T107A, R131Q and G109C;101_106dup. In this chapter, the consequences of the protein changes were assessed using pathogenicity prediction tools. In addition, the conservation of PPIL1 was assessed directly using a protein sequence species alignment. Structural analyses were also used to understand the functional consequences of the protein changes caused by the patient variants.

The pathogenicity prediction tools used to assess the likely pathogenicity of PPIL1 patient variants included: Polymorphism Phenotyping v2 (Polyphen), Sorting Intolerant From Tolerant (SIFT), CADD (Combined Annotation Dependent Depletion) and Protein Variation Effect Analyzer (PROVEAN). These tools are frequently used to help assess the likely pathogenicity of missense variants, (or in the case of PROVEAN missense or INDEL variants)[20]. The SIFT algorithm makes predictions based on amino acid conservation among species and also the properties of the amino acids in question[131]. Polyphen2 also uses sequence conservation data but also takes into account structural information (for example whether the mutation will perturb the hydrophobic core or further relevant features of the protein)[130]. SIFT and Polyphen2 are limited to predicting the effect of amino acid substitutions only. PROVEAN was developed to predict the effect of all types of protein sequence changes, including duplications. PROVEAN also uses amino acid conservation between species to predict pathogenicity, but also considers conservation of flanking sequences[132]. CADD is a tool which integrates many different methods of pathogenicity prediction including: conservation, sequence context of variant, gene annotations, epigenetic data and functional predictions from tools including PolyPhen and SIFT[139].

In addition, very little is currently known about the function of PPIL1. PPIL1 is a component of the spliceosome and many spliceosome cryo-EM spliceosome structures have recently been published. This was made possible, by both advances in cryo-EM and also in techniques to stall spliceosomes at specific stages of assembly (as discussed in Chapter 1). The advances in cryo-EM were made possible by increased sensitivity of microscopes and improvements in processing software. Various techniques have been employed to stall the spliceosome at specific phases of assembly (e.g. change of pH or addition of splicing inhibitors), allowing more

homogenous complexes to be analysed and thus increasing the resolution of spliceosome structures.

In this chapter, *in silico* modelling of published spliceosome cryo-EM data was used to gain novel insights into the role of PPIL1 in the spliceosome. Data was analysed for each stage of the spliceosome which contains PPIL1 and for which data was available (B^{act} to ILS spliceosome stages). This analysis suggests PPIL1 interacts with several proteins in the spliceosome and has led to the identification of a putative PPIL1 PPIase substrate.

3.2 Results

3.2.1 In silico prediction of variant pathogenicity

3.2.1.1 Pathogenicity prediction tools

The effect of the *PPIL1* variants identified in the patients was assessed using CADD[139], PolyPhen2[130], SIFT[131] and PROVEAN[132] (Table 14); this is standard practice after discovery of novel candidate variants. Overall, variant pathogenicity prediction tools suggest a likely pathogenic interpretation of PPIL1 variants. CADD scores for all variants are greater than 15 suggesting a pathogenic interpretation of the variants and the SIFT algorithm also ranks all PPIL1 variants as 'damaging'. For family 3 and 4 PPIL1 variants (p. R131Q, p.G109C;A101_D106dup) are predicted to negatively affect protein function across all 4 prediction tools used. Family 1 variant (p.T107A) is also predicted to affect protein function in SIFT, PROVEAN and CADD tools, and is classed as 'possibly damaging' by Polyphen2. However, predictions for the PPIL1 variant in family two (p.A99T) are most variable across the prediction tools, being classed as 'benign' or 'neutral' by PolyPhen2 and PROVEAN but 'damaging' or 'pathogenic' by SIFT and CADD. This indicates one of the drawbacks of using this approach.

Due to the rare occurrence of autosomal recessive genetic diseases it is predicted that disease causing variants must be relatively rare in the general population[138]. The occurrence of all patient single nucleotide variants in PPIL1 was therefore investigated using the Gnomad database (Table 14). For all patient single nucleotide variants, the PPIL1 variants are either not present in the Gnomad database or are found at less than 0.01 % in the general population (Table 14). The Gnomad database also provides allele frequencies for specific ethnic populations and shows that none of the single nucleotide variants in PPIL1 were found in the relevant ethnic population (Table 14). None of the patient variants were homozygous in any of the populations in the Gnomad database. The Gnomad database only reports two biallelic missense variants in PPIL1 p.Cys36Ser (79 individuals) and p.Gly46Asp (2 individuals; out of 141,456 individuals in the database), with allele frequencies of 0.01970 and 0.00007651 respectively. This indicates that *PPIL1* exhibits a low tolerance to biallelic missense variants across the gene. These results support the hypothesis that the identified PPIL1 variants are likely to be deleterious.

	Variant			Pathogenicity prediction				Allele frequencies (Gnomad)	
	Genomic variant (GRCh38)	Transcript variant	Amino acid change	CADD PHRED score	Polyphen	SIFT	PROVEAN	Total allele frequency	Population specific allele frequency
Family 1	Chr.6:36855995T>C	c.319A>G	**p.T107A**	26.8	Possibly damaging (0.779)	Deleterious (0)	Deleterious (-4.580)	Variant not present in Gnomad	
Family 2	Chr.6:.36856019C>T	c.295G>A	**p.A99T**	22.8	Benign (0.007)	Deleterious (0.04)	Neutral (-0.505)	0.0000142	0 (African population)
Family 3	Chr.6:36855922C>T	c.392G>A	**p.R131Q**	23.7	Possibly damaging (0.518)	Tolerated (0.06)	Deleterious (-3.884)	0.0000495	0 (African population)
Family 4*	*Chr.6:36855989C>A	c.325G>T	**p.G109C**	28.2	Probably damaging (0.989)	Deleterious (0)	Deleterious (-8.512)	Variant not present in Gnomad	
	*Chr.6:36855996_36856013dup	c.301_318dup	**p.A101_D106dup**	22.5	N.A.	N.A.	Deleterious (-14.714)	N.A.	

Table 14 Pathogenicity prediction of PPIL1 variants. All patients were homozygous for missense variants, except family 4 (*) which was homozygous for a complex allele consisting of a missense variant and duplication (*). Pathogenicity was predicted using CADD (https://cadd.gs.washington.edu/)[139], Polyphen2 (http://genetics.bwh.harvard.edu/pph2/)[130], SIFT (http://sift.bii.a-star.edu.sg/)[131] and PROVEAN (http://provean.jcvi.org)[132] and used the default cut off scores: CADD >15 predicted pathogenic; Polyphen >0.15 possibly damaging, >0.85 probably damaging; SIFT <0.05 predicted deleterious PROVEAN <-2.5 predicted deleterious. Allele frequencies in the general population were taken from Gnomad database and are shown for both the global population and specific populations.

3.2.1.2 Protein conservation

Conservation of specific residues across species may suggest an evolutionary constraint and functional importance of a particular residue in a protein[140]. Residue conservation is therefore a useful predictor of variant pathogenicity and was assessed directly using a protein sequence alignment for PPIL1 homologs (Figure 14). Both T107 and R131 residues are strictly conserved across species, suggesting the functional importance of these residues (Figure 14). A99 is also highly conserved across species, with serine being the only other residue found at this position (Figure 14); a residue which is equivalent to alanine except for a hydroxyl group in place of one methyl group protons. The patient variant affecting residue 99 of PPIL1 results in an alanine to threonine substitution. Although threonine is similar in structure to serine, a residue tolerated at this position in many species, threonine is larger than serine and contains a methyl group in place of a beta proton in serine. Threonine is not found at this position in any other species (Figure 14) (or in other human cyclophilins, Appendix Figure 59), suggesting that threonine may not be tolerated at this position. G109 is also highly conserved, with only *S. pombe* containing a different amino acid at this position (serine). The duplicated region (A101_D106dup) also contains four strictly conserved amino acids between species. The high degree of conservation of residues affected by patient variants in PPIL1 suggests that PPIL1 is intolerant of the variation observed at these residues.

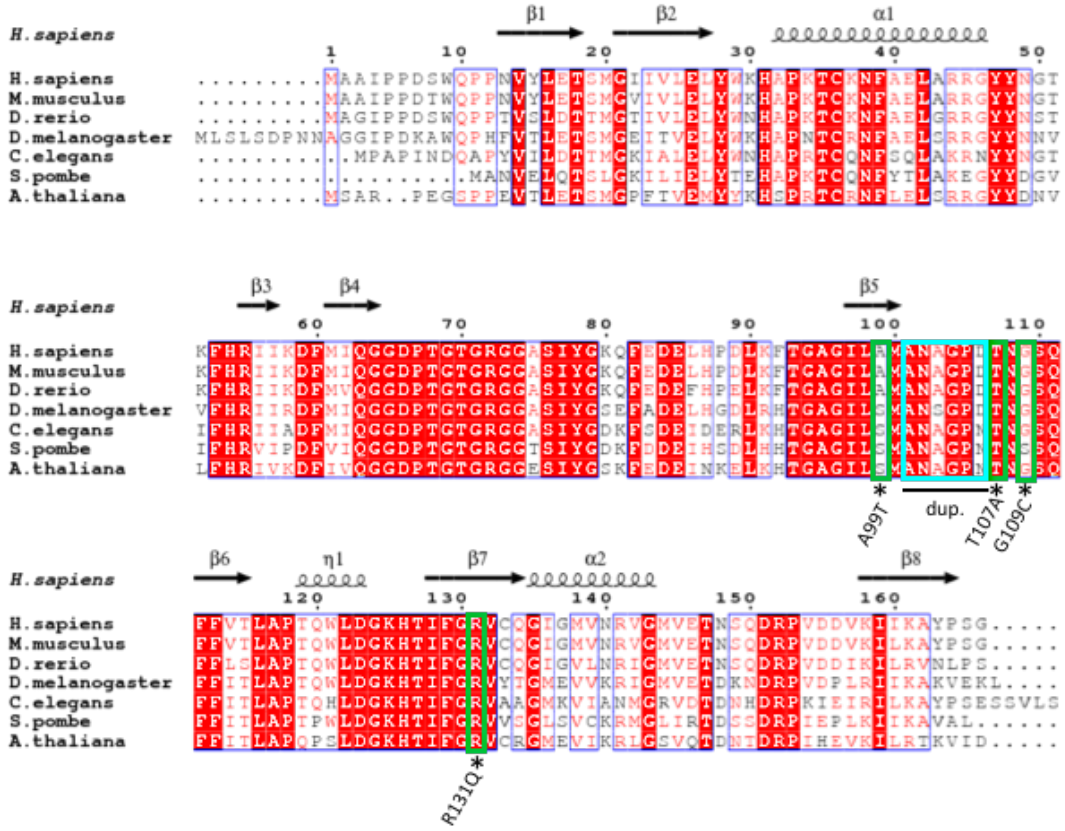


Figure 14 Alignment of PPIL1 amino acid sequence, showing conservation across species: *Homo sapiens* (NP_057143.1), *Mus musculus* (NP_081121.1), *Danio rerio* (NP_001029350.1), *Drosophila melanogaster* (NP_523874.1), *Caenorhabditis elegans* (NP_501118.1), *Schizosaccharomyces pombe* (NP_593308.1) and *Arabidopsis thaliana* (NP_181157.1). Conserved residues are shown with red letters and strictly conserved amino acids are highlighted in red with white letters. Positions of the patient mutations are indicated with green boxes for amino acid substitutions and a cyan box for the duplication region (A101_D106dup). Sequences were aligned using Clustal Omega[32] and rendered for sequence similarities using ESPript[33].

3.2.1.3 PPIL1 protein structural analysis

In order to gain an understanding of the region of PPIL1 affected by the patient variants, the relevant residues were highlighted on the published structure of PPIL1[56] (Figure 15). PPIL1 is a single domain globular protein with a PPIase active site and on the opposite side of the protein a SKIP interaction site (Figure 15)[55,56]. All mutations affect residues on the surface of the protein, except for A99T which is a buried residue in the hydrophobic core of the protein (Figure 15; Appendix Figure 56). G109, T107, A99 and the duplicated region (A101_D106dup) are found next to the PPIase active site (Figure 15). Specifically G109, T107 and A101_D106 are found in a loop between beta strands 5 and 6. This loop forms the S2 pocket, an adjacent pocket to the active site which is thought to confer PPIase substrate bind specificity[39]. It is also thought that flexibility of this loop is crucial for PPIase substrate turnover[55]. R131 is found on the SKIP binding surface and this residue is known specifically to be required for interaction with SKIP[56,86], other variants at this residue abolish the interaction with SKIP[86].

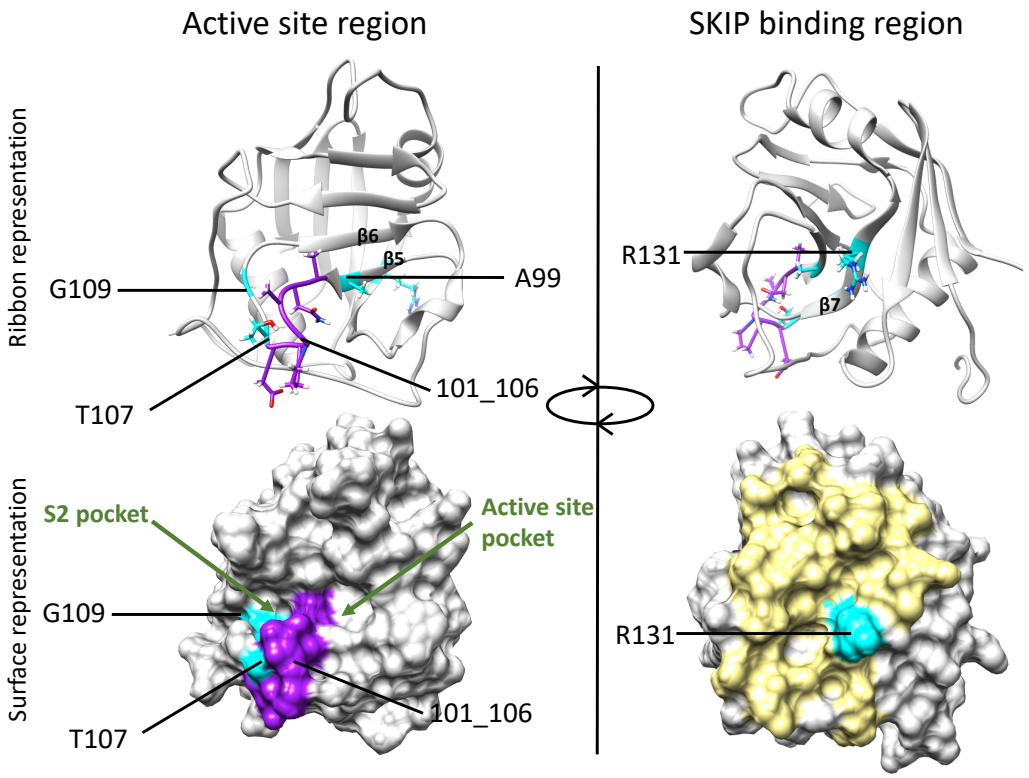


Figure 15 Location of residues affected by PPIL1 patient variants (A99T, T107A, R131Q, G109C; 101_106 dup). Light blue represents residues affected by missense mutations and purple represents residues affected by in-frame duplication mutations. All patients were homozygous for the PPIL1 variants, including G109C; 101_106 dup where patients were homozygous for both mutations. Most mutations cluster around the active site (left); however, the R131Q mutation affects a residue on the SKIP binding surface of the protein (right). The SKIP binding surface, as defined by Wang *et al.* 2010, is shown in yellow on the surface representation[86]. Images were created using UCSF Chimera[57] and the published crystal structure of PPIL1 (PDB: 2X7K)[56].

3.2.2 Insights from cryo-EM structures

3.2.2.1 PPIL1 interactions

Little is known about the role of PPIL1 in the spliceosome; therefore, spliceosome cryo-EM data was analysed to uncover novel insights about PPIL1. Data for spliceosome structures containing PPIL1 was available for: human B^{act} mature (5.1 Å)[49], human B^{act} late (6.5 Å)[49], human B^{act} [composite] (4.5 Å)[50], human C (4.1 Å)[77], human C (5.9 Å)[52], human C* (3.6 Å)[47], human P (3.3 Å)[93], human P (3.0 Å)[94], human ILS [before prp43 loaded] (2.9 Å)[94], human ILS [after prp43 loaded] (2.9 Å)[94] and *S. pombe* ILS (3.6 Å) spliceosome complexes (Table 15). The published data was imported from the protein data bank (www.rcsb.org) and modelled using Chimera software. Analysis of the spliceosome structures confirms that as shown previously, PPIL1 binds to SKIP (Figure 16)[55,56,86]. SKIP is a predominantly disordered protein in solution[86], but these structures show that SKIP adopts an ordered conformation in the spliceosome. Analysis also shows previously unknown interaction partners binding to PPIL1 in the spliceosome structures, including: PRP17 N terminus, RBM22 C terminus, CRNKL1, SYF2 and SPF27 of the PRP19 helical bundle (Figure 16).

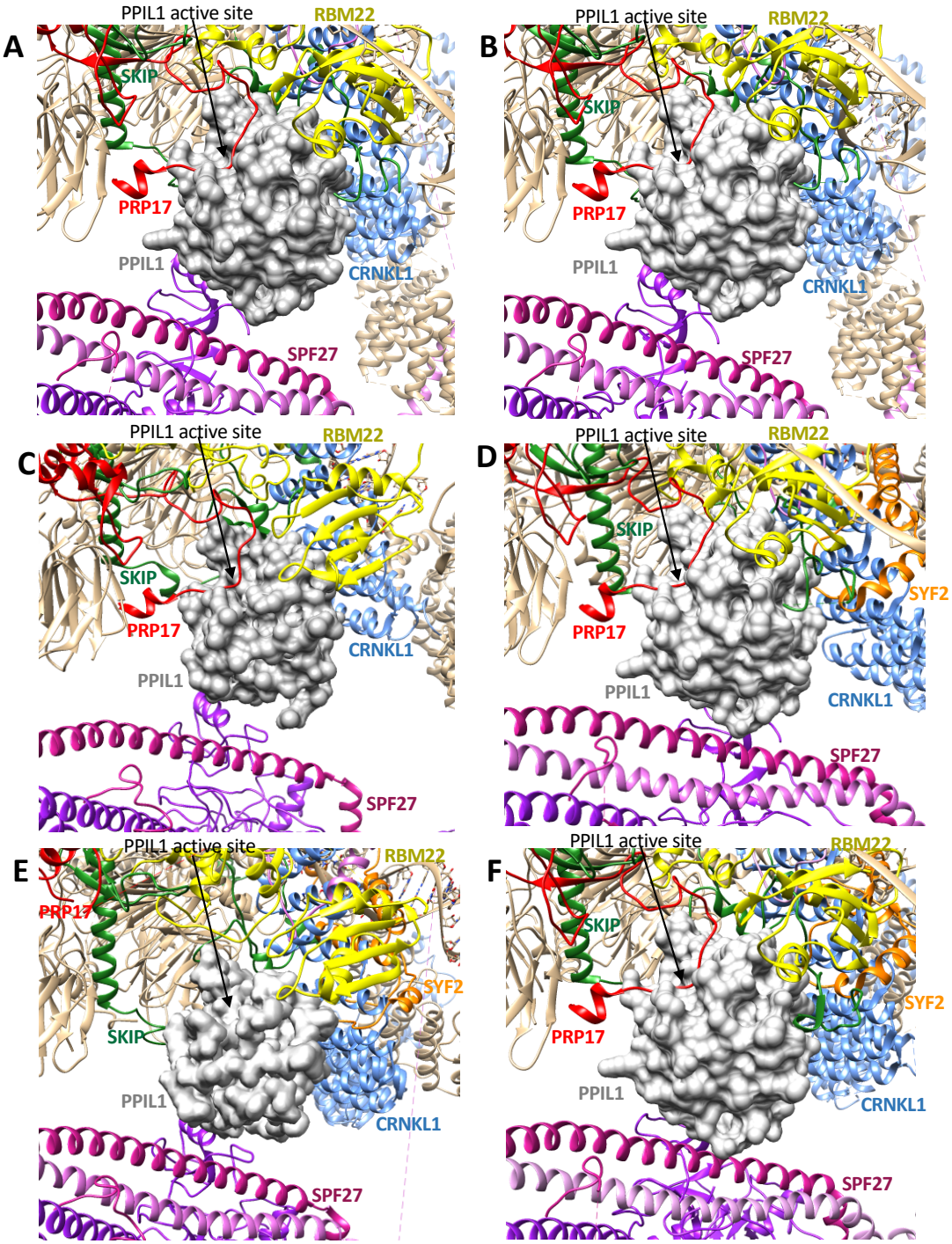
Table 15 Spliceosome cryo-EM structures, which include PPIL1 protein

Spliceosome complex	PDB code	Resolution
human B^{act} mature	5Z56	5.1 Å
human B^{act} late	5Z57	6.5 Å
human B^{act} [composite]	6FF7	4.5 Å
human C	5YZG	4.1 Å
human C	5O9Z	5.9 Å
human C*	5XJC	3.6 Å
human P	6QDV	3.3 Å
human P	6ICZ	3.0 Å
human ILS [before prp43 loaded]	6ID0	2.9 Å
human ILS [after prp43	6ID1	2.9 Å

loaded]

S. pombe ILS	3JB9	3.6 Å

All interactions observed, occur on surfaces of PPIL1 outside of the active site, except for the interaction with PRP17 (Figure 16). A loop of PRP17 protein interacts with PPIL1 active site. PPIL1 is a peptidyl prolyl isomerase enzyme, whose substrate is a prolyl peptide bond[55]. Interestingly, Gly94-Pro95 peptide bond of PRP17 occupies the centre of the active site (Figure 17). Gly94-Pro95 is strictly conserved between all species which contain this N-terminal region of PRP17 (Figure 18). S. cerevisiae PRP17 homolog does not contain an equivalent region corresponding to hPRP17 N-terminal region (1-135); however, S. cerevisiae also does not have a PPIL1 homolog (Figure 18; Appendix Figure 65). The PRP17-PPIL1 interaction is shown across all human spliceosome structures which contain PPIL1 and PRP17 (B^{act} to ILS complexes) and in the *S. pombe* ILS spliceosome structure (Figure 16). PPIL1 is a member of the cyclophilin family of PPIases, which contain an adjacent pocket to the active site called the S2 pocket; this region binds to substrate i-2 residues and this interaction is proposed to influence substrate binding specificity[39]. PRP17 Phe93 (i-2) sidechain binds to the S2 pocket in human spliceosome structures (Figure 17).


A
PPIL1 active site
RBM22
SKIP
PRP17
PPIL1
CRNKL1
SPF27
B
PPIL1 active site
RBM22
SKIP
PRP17
PPIL1
CRNKL1
SPF27
C
PPIL1 active site
RBM22
SKIP
PRP17
CRNKL1
PPIL1
SPF27
D
PPIL1 active site
RBM22
SKIP
SYF2
PRP17
PPIL1
CRNKL1
SPF27
E
PPIL1 active site
RBM22
PRP17
SYF2
SKIP
PPIL1
CRNKL1
SPF27
F
PPIL1 active site
RBM22
SKIP
SYF2
PRP17
PPIL1
CRNKL1
SPF27

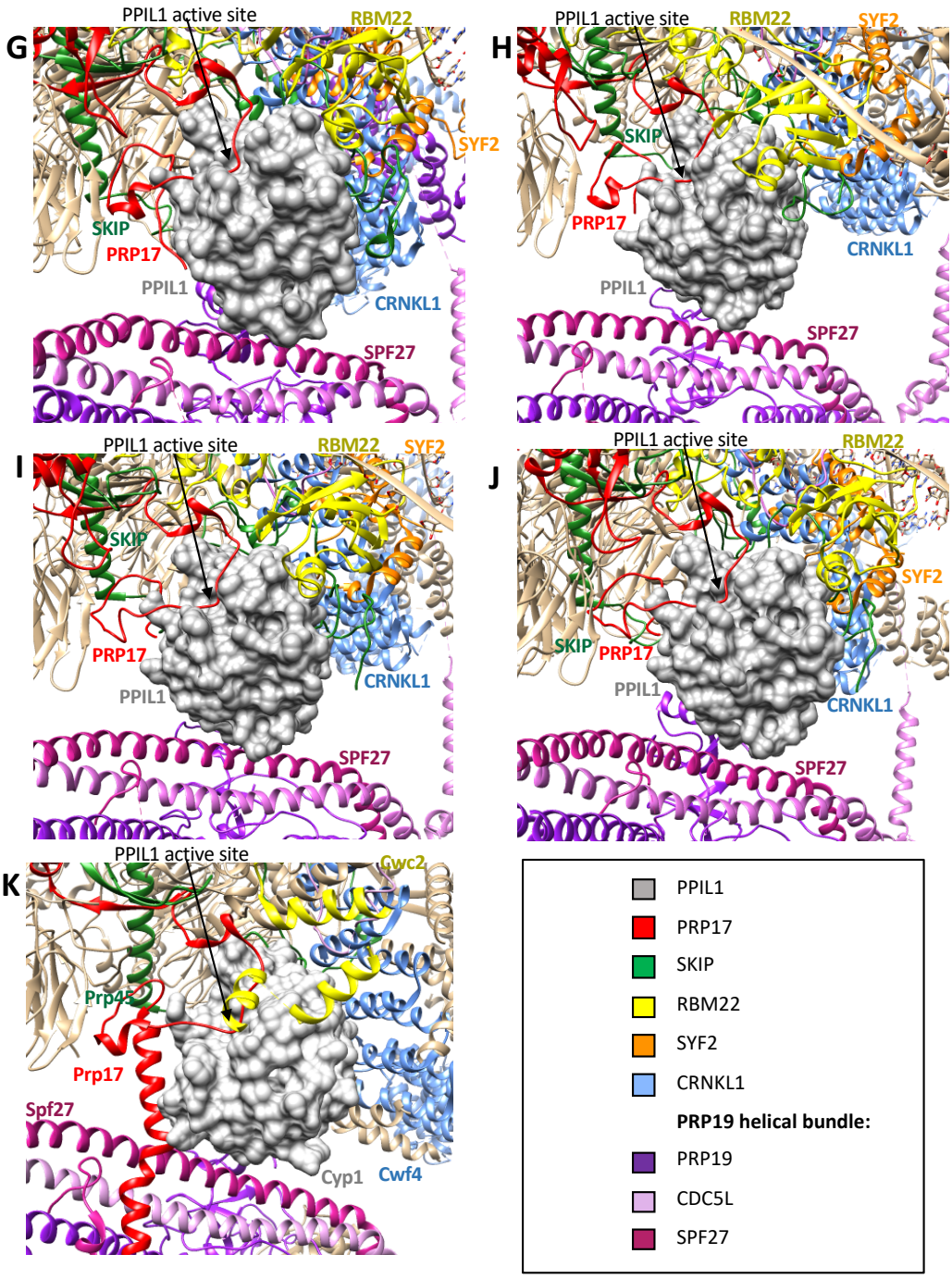


Figure 16 Cryo-electron microscopy structures of spliceosome complexes: human B^{act} mature (A)[49], human B^{act} late (B)[49], human B^{act} composite (C)[50], human C (D)[77], human C (E)[52], human C* (F)[47], human P (G)[93], human P (H)[94], human ILS [before prp43 loaded] (I)[94], human ILS [after prp43 loaded] (J)[94], *S. pombe* ILS (K). PPIL1 is shown as a grey surface fill model and interacting proteins are highlighted (PRP17 [red], SKIP [green], RBM22 [yellow], CRNKL1 [blue], SYF2 [orange] and the PRP19 helical bundle [purple/pink] in the human model and the corresponding homologs in *S. pombe*). Other components of the spliceosome are shown in brown. Images were created using published data from the protein data bank (www.rcsb.org[135]; 5Z56[49], 5Z57[49], 6FF7[50], 5YZG[77], 5O9Z[52], 5XJC[47], 6QDV[93], 6ICZ[94], 6ID0[94], 6ID1[94], 3JB9[48]) and Chimera software (http://www.rbvi.ucsf.edu/chimera)[57]. Regions of proteins without sufficient resolution were absent from the imported PDB data.

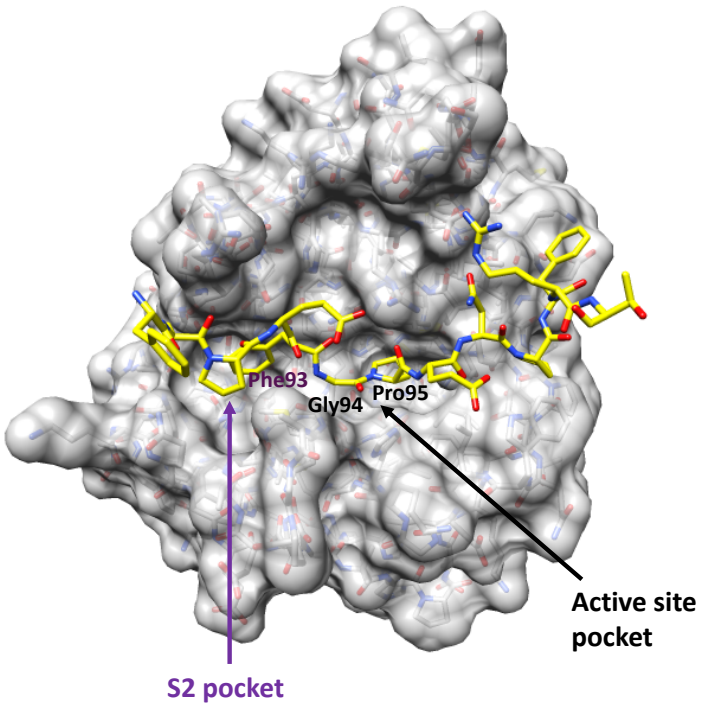


Figure 17 PRP17 (89-101) and PPIL1 interaction in the spliceosome. The published cryo-EM structure of the C* human spliceosome[47] shows a loop of PRP17 (residues 89-101; yellow) is positioned in the PPIase active site of PPIL1 (grey). This loop contains a conserved proline residue which sits at the centre of the PPIL1 proline isomerase active site. (Phe93 sidechain occupies the S2 pocket, a region thought to confer binding specificity for substrates of the PPIase active site[39]). This region of the model was rebuilt and refined by Chuangye Yan and Yigong Shi, Tsinghua University, to confirm conformation and interaction of PPIL1 and PRP17.

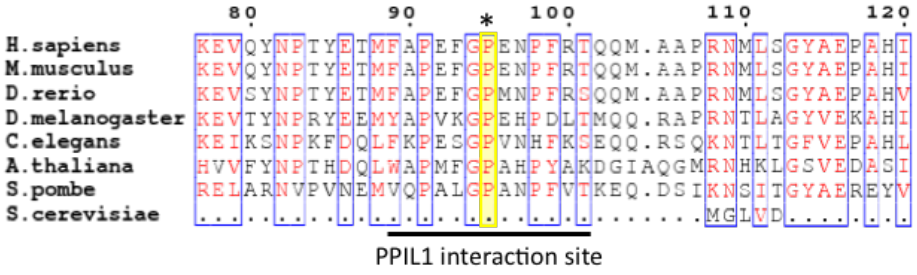


Figure 18 Amino acid alignment of PRP17 (77-120) showing conservation across species: *Homo sapiens* (NP_056975.1), *Mus musculus* (NP_082155.1), *Danio rerio* (NP_001009990.1), *Drosophila melanogaster* (NP_651005.1), *Caenorhabditis elegans* (NP_492851.1), Arabidopsis thaliana (NP_172528.2), *Schizosaccharomyces pombe* (NP_596089.1) and *Saccharomyces cerevisiae* (NP_010652.3). The loop of PRP17 (residues 89-101) which interacts with PPIL1 in the cryo-EM structures [47,48] is indicated. The critical proline residue which occupies the PPIL1 active site in cryo-EM data (highlighted in yellow; Pro95 human PRP17, Pro69 *S. pombe* PRP17) is conserved across species; from human to *S. pombe*. (See appendix for full length PRP17 sequence alignment.)

3.3 Discussion

Overall pathogenicity prediction tools, sequence conservation and structural analysis of PPIL1 support the hypothesis that the *PPIL1* variants are likely to be deleterious to function. However, despite most results supporting this conclusion some of the pathogenicity prediction tools gave different results for A99T depending on the tool used. This similarity in properties of serine, a residue tolerated at residue 99 and threonine (the substitution caused by a PPIL1 patient variant, A99T) might be the reason for this disparity. These prediction tools have a high error rate (e.g. SIFT and PolyPhen have an estimated sensitivity of 69% and 68% and specificity of 13% and 16% respectively)[22]. A significant proportion of human proteins do not have full domain annotations meaning algorithms using annotation data, such as Polyphen, may not provide the full picture. Analysis of sequence conservation and structural analysis coupled with the literature is also required in order to best predict effects of patient variants. The findings presented here also highlight the importance of directly investigating variant pathogenicity in the laboratory.

In addition to suggesting the identified PPIL1 variants are deleterious, the *in silico* results were also valuable for informing the design of further experiments. Structural analysis shows that one patient variant affects residue 99 in the hydrophobic core of PPIL1. In proteins, the arrangement and composition of the hydrophobic core is crucial for structure and stability. Substitutions at residue 99 may therefore affect PPIL1 stability and the folding and stability of PPIL1 proteins is investigated in Chapter 4. Structural analysis also shows that most of the PPIL1 patient variants affect residues around the PPIase active site and the effect of the variants on PPIase activity was monitored (Chapter 5). A further patient variant affects residue 131 at the SKIP binding interface; it was therefore also tested whether patient variants perturb this interaction (Chapter 5).

Previously only SKIP was a known binding partner of PPIL1[55], therefore the analysis of spliceosome structures provides important insights into the role of PPIL1 in the spliceosome. Interestingly the region of PPIL1 which binds to the PRP19 helical bundle, was suggested as a putative additional binding site in PPIL1 due to the presence of several positively charged residues[55]. The observation that a proline residue of PRP17 occupies the PPIL1 active site, in the spliceosome structures raises the questions of whether PRP17 may be a substrate of PPIL1 in the spliceosome. The high conservation of

PRP17 Pro95 across species suggests a functionally important role of this residue (see Chapter 6).

Many of the proteins PPIL1 binds to form long range interactions across the spliceosome, from peripheral regions to the spliceosome core. The PPIL1 interacting proteins include PRP17 which has an N-terminal region spanning over 100 Å and a WD40 domain adjacent to the spliceosome catalytic centre[47]. In addition, PPIL1 binds to the PRP19 helical bundle, of which CDC5L spans a large distance across the spliceosome. CDC5L Myb domains also contact the catalytic centre of the spliceosome, where they stabilise the catalytic U2/U6 structure and interact with PRP8; while the C-terminus of CDC5L forms part of the PRP19 helical bundle at the spliceosome periphery[47]. SKIP has an extended conformation interacting with more than 10 proteins across the spliceosome, contacting proteins such as PPIL1 at the periphery and also the catalytic RNA centre of the spliceosome[51]. These long range interactions may play a role in transferring information from regulatory factors to the catalytic core[47,77].

Overall this work suggests the identified PPIL1 variants are likely pathogenic and informs the design of laboratory experiments to test this hypothesis (Chapter 4-5). In addition, analysis of spliceosome structures has given important insights into the role of PPIL1, which will aid the understanding of the mechanism of this disease.

Chapter 4 Protein purification, folding and stability

4.1 Introduction

In silico work (Chapter 3), suggests that patient variants in PPIL1 are pathogenic and likely to have a significant impact on PPIL1 protein function. In order to confirm these findings and further understand the effect of these variants on PPIL1 protein function, PPIL1 proteins corresponding to WT and mutant PPIL1 protein were made. PPIL1 proteins were expressed and purified from BL21(DE3) strain of *E. coli*, with a C-terminal *His* tag for purification purposes, as this strategy has previously been shown to yield soluble folded PPIL1 WT protein[55]. The identity, purity and homogeneity of these proteins was first assessed using SDS-PAGE, mass spectrometry and size exclusion chromatography. In order to inform experimental design and buffer choice for future experiments a protein stability screen was also carried out using a range of different buffer conditions (pH range, salt concentrations, glycerol concentrations and additives).

PPIL1 interacts with SKIP in the spliceosome[47,55,56,77,86] and therefore SKIP was also expressed and purified to study this interaction. SKIP is predominantly disordered and interacts with PPIL1 via a linear epitope [56,86]. The PPIL1 binding region of SKIP (residues 59-129)[55,56,86], which contains this epitope, was expressed and purified from *E. coli*. This protein was expressed using a GST tag for purification purposes and also to allow immobilization during binding experiments with PPIL1. SKIP identity and purity were similarly assessed using SDS-PAGE and mass spectrometry as for PPIL1.

PPIL1 patient variants result in amino acid substitutions or amino acid duplications in the encoded PPIL1 protein (A99T, T107A, R131Q, 101_106dup; G109C; Figure 19). PPIL1 is a small globular protein (19 kDa) consisting of a compact eight-stranded beta barrel capped by two alpha helices[55,56]. A99T substitution affects a residue on the beta 5 strand of PPIL1 which is a buried residue in the hydrophobic core of PPIL1. The T107A substitution affects a solvent-exposed residue which is in a loop connecting beta strands 5 and 6. The R131Q substitution also affects a solvent-exposed residue, which is on beta 7 strand of PPIL1. The final patient was homozygous for variants that resulted in both a duplication and amino acid

substitution (101_106dup; G109C) and affects residues of the loop between beta 5 and 6 strands, which are also solvent exposed. (See Chapter 3 for an in depth analysis).

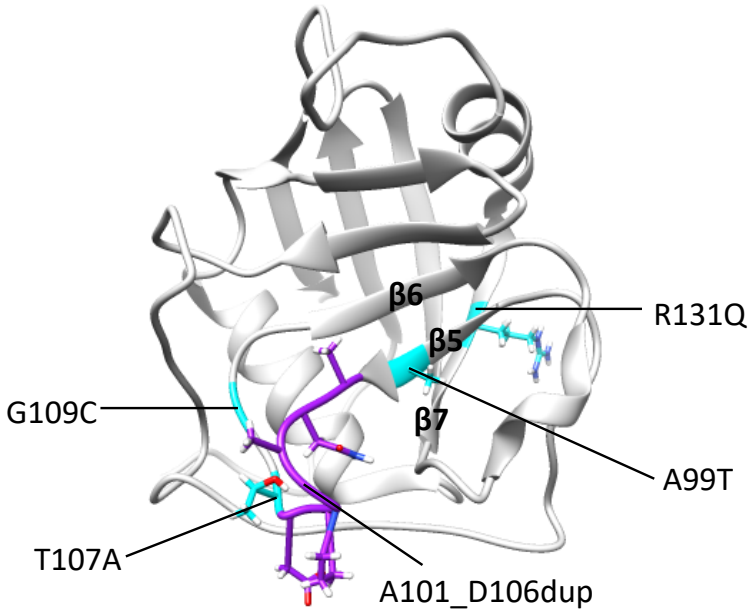


Figure 19 PPIL1 substitutions (light blue) and duplications (purple), caused by PPIL1 patient variants. Affected residues are highlighted on the published crystal structure of PPIL1[56].

Both amino acid substitutions and duplications can have a significant effect on protein folding and stability. Stability and folding of PPIL1 proteins were assessed using circular dichroism, aggregation and unfolding assays. Far UV circular dichroism was first used to determine whether any of the patient variants significantly perturbed PPIL1 protein secondary structure content. The overall stability of PPIL1 proteins was also assessed using a protein unfolding screen. Finally, the impact of patient variants on protein aggregation was also measured.

4.2 Results

4.2.1 Protein expression and purification

4.2.1.1 PPIL1 protein expression and purification

PPIL1 pET22b+ protein expression plasmids (WT, T107A, A99T and R131Q) were a gift from Guoliang Chai, UCSD. A further PPIL1 G109C 101_106dup pET22b+ plasmid was created using site directed mutagenesis. pET22b+ plasmids contain a C-terminal His tag to allow protein purification using affinity chromatography. All plasmids were sequenced to confirm the plasmids contained the correct coding sequence, inserted in the correct orientation, in-frame with the his tag and that no additional mutations were present (Appendix Figure 66).

Recombinant PPIL1 proteins were expressed in BL21(DE3) *E. coli* cells, transformed with the PPIL1 pET22b+ plasmids. Expression was induced with the addition of IPTG and expression trials confirm that this induced protein production of the correct size, 19 kDa (Figure 20A). PPIL1 protein was purified on Ni-NTA columns and eluted using an imidazole gradient. A significant proportion of each expressed protein was soluble, bound to and could be purified from the Ni-NTA columns (Figure 20B), with the exception of PPIL1 A99T and G109C;101_106dup. Both G109C;101_106dup and PPIL1 A99T proteins were insoluble when expressed at 37 ^{o}C (data not shown). Reducing the culture temperature after IPTG induction to 18 ^{o}C, resulted in a proportion of the expressed protein becoming soluble and purifiable. Elution chromatograms show a clear peak in absorbance at 280 nm, as the imidazole gradient is raised, showing PPIL1 proteins were eluted from the columns (Figure 20C). SDS-PAGE gels confirm PPIL1 proteins are of high purity (Figure 20D). Both A99T and 101_106dup; G109C PPIL1 proteins could only be purified at much lower yields compared to PPIL1 WT and exhibit very low solubility (Figure 20D).

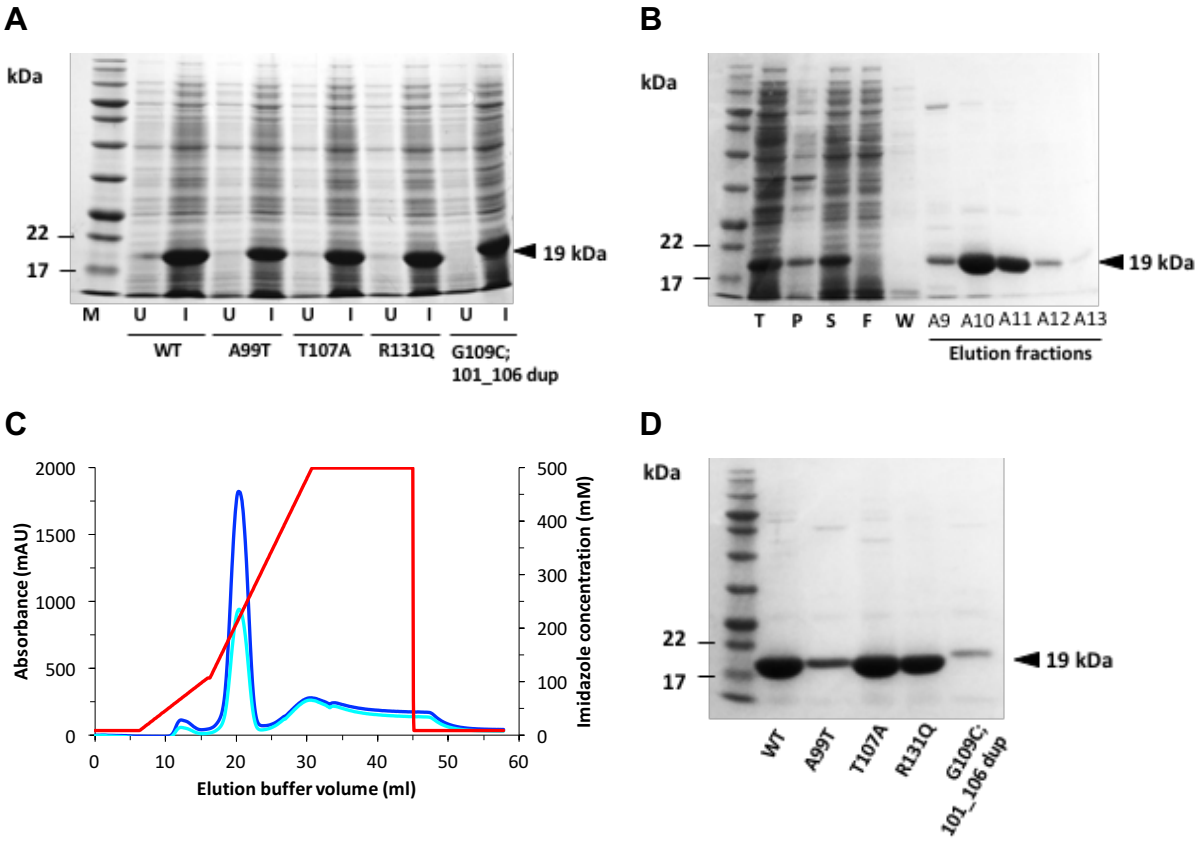


Figure 20 Protein expression and affinity purification of recombinant PPIL1 proteins (19 kDa). Representative Coomassie stained SDS-PAGE gels and purification chromatograms are shown.
A) Protein expression was induced with IPTG and samples were taken from uninduced (U) and induced (I) cultures. The expected 19 kDa band corresponding to PPIL1 is seen in all the induced cultures
B) Example protein purification gel for PPIL1 WT. Protein was purified using affinity chromatography and samples were taken during the procedure: total protein (T), insoluble protein in pellet (P), soluble protein (S), flow through from column (F), column wash (W) and elution fractions. The elution fractions which were analysed correspond to the peak in absorbance at 280 nm in the elution chromatogram, and fractions from either side of the peak.
C) Example chromatogram of PPIL1 WT protein purification, showing absorbance at 280 nm (blue) and 260 nm (cyan) and the imidazole gradient (red) used to elute the protein. A single sharp peak in absorbance at 280 nm indicates elution of PPIL1 protein.
D) SDS-PAGE gel of purified PPIL1 proteins, showing that all PPIL1 proteins are of high purity.

4.2.1.2 SKIP protein expression and purification

A sequence encoding SKIP residues 59-129 was cloned into pGEX-6P1 plasmid, to allow protein expression and purification (Appendix Figure 66). Recombinant SKIP 59-129 protein was expressed in BL21(DE3) *E. coli* cells, with an N-terminal GST tag for purification purposes and to allow immobilisation in binding experiments. GST tag alone was also expressed and purified alongside GST:SKIP 59-129 to act as a control reagent in binding experiments. Expression trials were carried out, as for PPIL1 proteins, and showed that induction with IPTG resulted in protein expression of approximately the expected protein size, 35 and 28 kDa for GST:SKIP 59-129 and GST tag respectively (Figure 21A). GST tagged proteins were purified on columns containing glutathione sepharose, which has a strong affinity for GST. A large fraction of expressed GST proteins were soluble, bound to and were subsequently purified from the columns (Figure 21B). GST proteins were eluted using a glutathione injection and the resulting chromatogram shows a single sharp peak in absorbance at 280 nm indicating GST-tagged proteins were eluted from the column (Figure 21C). GST-SKIP and GST purification yielded proteins that migrated at approximately the expected molecular weight in SDS-PAGE. However, an additional band is present on SDS-PAGE gels for GST-SKIP which corresponds to cleaved GST tag (as also shown by Stegmann et al.[56]; Figure 21D). Apart from the extra band representing a small proportion of cleaved GST in the GST:SKIP 59-129 sample, purified GST tagged proteins are also of high purity (Figure 21D). There is a small discrepancy in expected molecular weight of GST-tagged proteins and the apparent molecular weight from SDS-PAGE; likely caused by specific properties of the protein. For example, different residues vary in their charge and affinity for SDS which therefore affects migration through the gel.

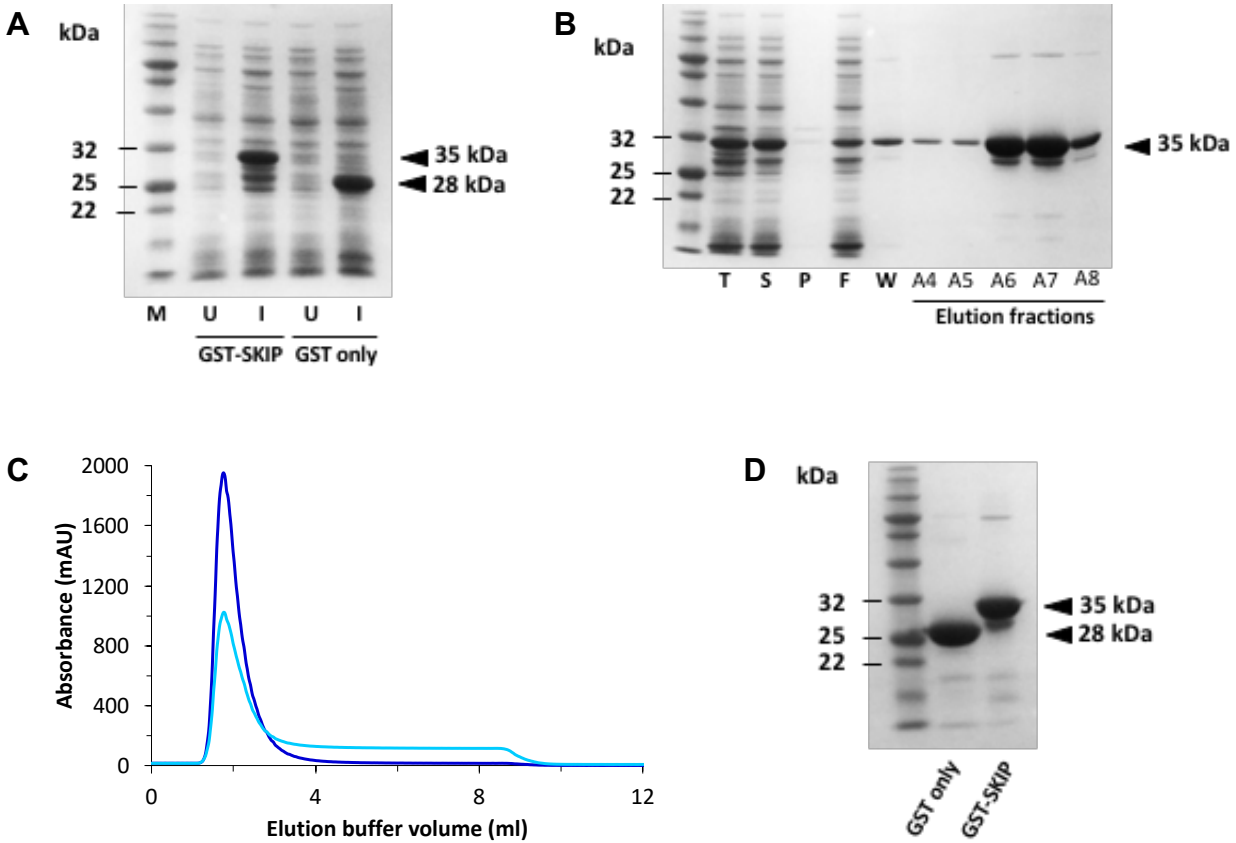


Figure 21 GST (28 kDa) and GST:SKIP 59-129 (35 kDa) protein expression and purification. Example SDS-PAGE gels (Coomassie stained) and purification chromatograms are displayed.
A) Samples were taken from uninduced (U) and induced (I) cultures to confirm protein expression had been induced by IPTG.
B) Representative protein purification gel for GST:SKIP 59-129. Protein was purified from *E. coli* lysates using a glutathione sepharose column. Total protein (T) represents all protein in the cell lysate. After centrifugation to remove insoluble protein, samples from the insoluble protein pellet (P) and remaining soluble protein (S) were collected. The soluble protein was loaded onto the column and the flow through (F), column wash (W) and elution fractions were collected.
C) Representative chromatogram of GST:SKIP 59-129 protein purification. GST tagged proteins were eluted using a single glutathione injection (10 mM) and a single sharp peak in absorbance can be seen at 280 nm (blue) and 260 nm (cyan) as the protein was eluted.
D) SDS-PAGE gel showing purity of proteins. (GST:SKIP 59-129 sample also contains a small proportion of cleaved GST.)

4.2.1.3 Validation of purified proteins

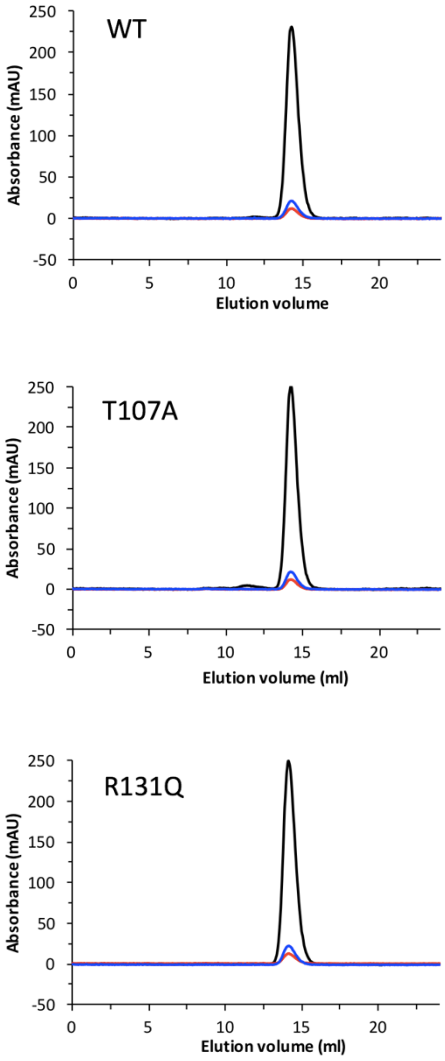
Soluble purified proteins were verified using time of flight mass spectrometry analysis (by Dr James Ault, Astbury Mass Spectrometry facility). All determined molecular weights were equal or within 1 Da of the expected molecular weight (Table 16), confirming the identity of the purified proteins. PPIL1 mutant proteins A99T and G109C; 101_106 dup could not be analysed due to a high propensity to aggregate (see section 4.2.2.1). These results also confirms that any small differences in expected and apparent molecular weight determined by SDS-PAGE is a result of altered migration rather than any other factor, such as degradation or incorrect product.

Sample	Predicted molecular weight (Da)	MS results (Da)
PPIL1 WT	19170.7	19170.58 $\pm$ 1.54
PPIL1 T107A	19140.7	19140.32 $\pm$ 1.13
PPIL1 R131Q	19142.6	19142.67 $\pm$ 0.24
GST-SKIP	35352.8	35352.76 $\pm$ 0.40
GST (control)	28429.9	28429.24 $\pm$ 0.11

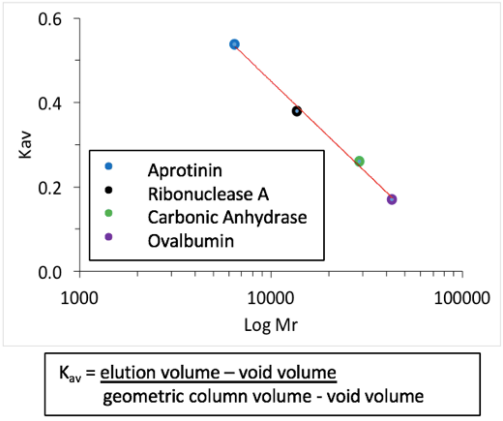
Table 16 Confirmation of the molecular masses of purified proteins. Protein molecular weight was determined by time of flight mass spectrometry and compared to the predicted molecular weight (calculated using ExPASy ProtParam tool). The predicted molecular weight for PPIL1 proteins is shown for removal of N-terminal methionine, as is often seen for recombinant proteins expressed in *E. coli*.

Size exclusion chromatography (SEC), using a Superdex 75 column, was used to confirm the purified PPIL1 proteins were homogenous, monomeric and of the expected size. All three soluble PPIL1 proteins (WT, T107A, R131Q) elute at a single elution volume and show a single sharp symmetrical peak in the elution chromatogram, confirming the purified PPIL1 protein solutions are homogenous and of high purity (Figure 22A). A calibration curve was plotted with protein standards, which are globular proteins of known molecular weight (Figure 22B), to calculate the apparent molecular weight of PPIL1 based on the elution volume. All three PPIL1 proteins elute at a volume consistent with a molecular weight of 15 kDa, which is close to the actual molecular weight of PPIL1 of 19 kDa (Figure 22C), suggesting that PPIL1 is monomeric. The small difference between the apparent and actual molecular weight of PPIL1 is likely due to the heterogenous shape of proteins, which make it difficult to predict molecular weights with high accuracy using this technique (see discussion)[141]. Increasing salt concentrations (from 137 mM to 300 or 500 mM NaCl) in the elution buffer to disrupt any interactions with the column that could lead to this difference, did not cause PPIL1 proteins to elute at a different volume (data not shown).

A



B



K_{av} = elution volume – void volume / geometric column volume - void volume

C

PPIL1 protein	Elution volume (ml)	K_{av} value	Apparent Molecular Weight
WT	14.26	0.375	14.7 kDa
T107A	14.24	0.373	14.8 kDa
R131Q	14.13	0.366	15.4 kDa

Figure 22 Analytical size exclusion chromatography (SEC) of PPIL1 proteins using a Superdex 75 column.

A) Representative SEC elution chromatograms of PPIL1 proteins (WT, T107A and R131Q), showing absorbance at 220 nm (black), 280 nm (blue) and 260 nm (red).

B) A calibration curve was made using globular proteins of known molecular weight (Aprotinin, 6.5 kDa; Ribonuclease A, 13.7 kDa; Carbonic Anhydrase, 29.0 kDa and Ovalbumin 43.0 kDa). To create the calibration curve, relative molecular mass (Mr) was plotted on a logarithmic scale against the partition coefficient, K_{av}. K_{av} was calculated, as shown above, using the elution volume (the volume between sample injection and elution from the column), void volume (the buffer volume that does not enter the pores; determined using blue dextran injections) and geometric column volume (the total volume of the column).

C) Using the calibration curve, the apparent molecular weights of PPIL1 proteins were calculated.

4.2.1.4 Buffer screen

In order to determine optimal buffer conditions for stability of PPIL1 protein, a protein stability screen was carried out using a range of: pH conditions, NaCl concentrations, glycerol concentrations and different additives. To measure protein stability, protein unfolding was monitored over a temperature gradient using the different buffer conditions. The high-throughput Optim system (Unchained Labs) was used, which can simultaneously measure protein unfolding in 48 samples. The degree of protein unfolding was measured using the barycentric mean wavelength of intrinsic protein fluorescence. Proteins fluoresce due to the presence of aromatic amino acids (tryptophan, phenylalanine and tyrosine) which are able to absorb and subsequently release light. Intrinsic fluorescence emission was analysed between 280 and 450 nm, which is primarily due to fluorescence emission of tryptophan. The emission wavelength of tryptophan is particularly sensitive to changes in environment and increases as a protein is unfolded due an increase in solvent exposure[142].

As the temperature was raised, PPIL1 protein transitions from folded to unfolded protein, which is represented by an increase in the barycentric mean wavelength of intrinsic protein fluorescence (Figure 23). PPIL1 WT shows a single transition from folded to unfolded protein as temperature is increased (Figure 23). This is consistent with the single domain structure of PPIL1[55]. Small, globular, single domain proteins typically show a single transition from folded to unfolded protein. (Multiple transitions may represent unfolding of multiple protein domains or several intermediate unfolding steps)[143].

Buffer pH can have a large effect on protein stability by altering charges on ionizable groups in amino acid sidechains, as a result of protonation or deprotonation[144]. A range of pH conditions (pH 3.0 - 9.5) were tested for impact on PPIL1 stability. The pH range 6.5-8.5 appears to be optimal for PPIL1 stability, as PPIL1 remains folded at higher temperatures, compared to other pH conditions tested (Figure 23A). A pH above or below this range is detrimental to PPIL1 stability.

Salt concentration may also have a significant effect on protein stability[145] and the effect of NaCl concentration (50-1500 mM) on PPIL1 stability was tested. All NaCl conditions tested, give rise to PPIL1 unfolding curves which

closely overlay (Figure 23B). This suggests that NaCl concentration, within this range, does not have a significant effect on protein stability.

Glycerol is commonly used to enhance protein stability due to the ability to promote compact protein conformations and also reduce aggregation[146]. Therefore, the presence of glycerol in sample buffers was also tested for impact on PPIL1 protein stability. Increasing glycerol concentrations, caused PPIL1 to unfold at higher temperatures, indicating increased PPIL1 protein stability (Figure 23C). However, the contribution towards PPIL1 protein stability is relatively small, compared to other factors such as pH.

It is widely known that protein stability may be improved with the inclusion of chemical additives in the sample buffer[145]. A range of chemical additives were tested for their effect on PPIL1 protein stability. However, none of the tested reagents caused any significant improvement in PPIL1 protein stability. All unfolding curves either overlay closely with the no additive control, or are shifted to lower temperatures, suggesting these the reagents tested may have a neutral or detrimental effect on PPIL1 protein stability, at the concentrations tested (Figure 23D).

Overall these results show that PBS (pH 7.4) is suitable as a general buffer for PPIL1 protein. Altering pH or salt concentration or presence of additives does not significantly improve PPIL1 stability. Although glycerol improves PPIL1 stability, the impact is only small. High concentrations of glycerol that are required to significantly improve PPIL1 stability may not be optimal for future experiments, such as SPR. In addition, specific types of experiments may also require different buffer conditions, such as NMR where a low pH is advantageous. These results show that, if required, the pH can be reduced to 6.5 without significantly perturbing protein stability.

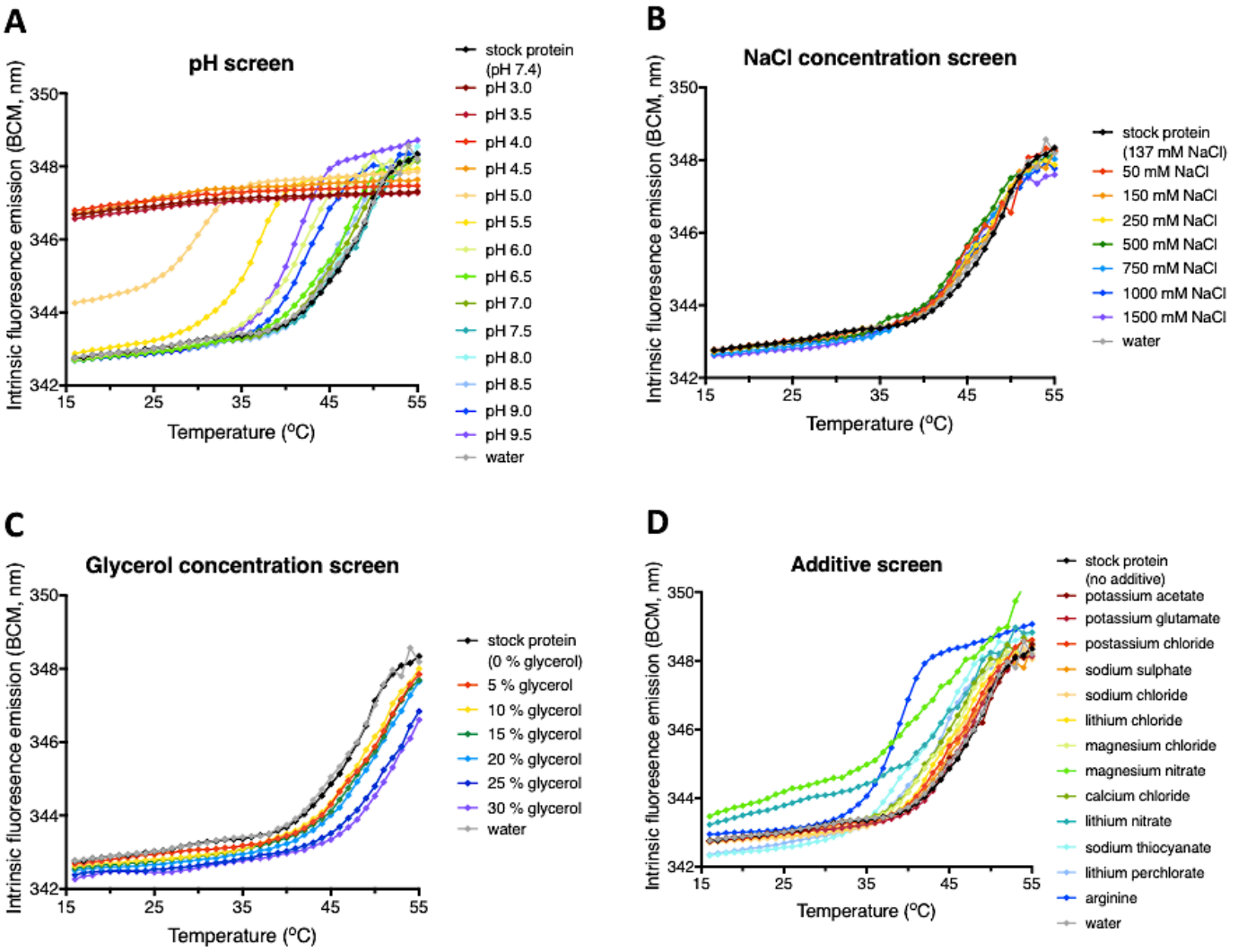


Figure 23 PPIL1 WT Optim buffer screen. Different buffer conditions were tested (pH range, glycerol concentrations, NaCl concentrations and use of additives) to determine optimal conditions for PPIL1 WT protein stability. Protein stability was measured over a temperature gradient using the barycentric mean (BCM) of intrinsic protein fluorescence as a measure of protein folding. Stock PPIL1 protein was in PBS, pH 7.4 (137 mM NaCl, 2.7 mM KCl, 11.9 mM phosphate buffer; no glycerol or additives).

4.2.2 Comparison of PPIL1 WT and mutant protein folding and stability

4.2.2.1 Protein solubility

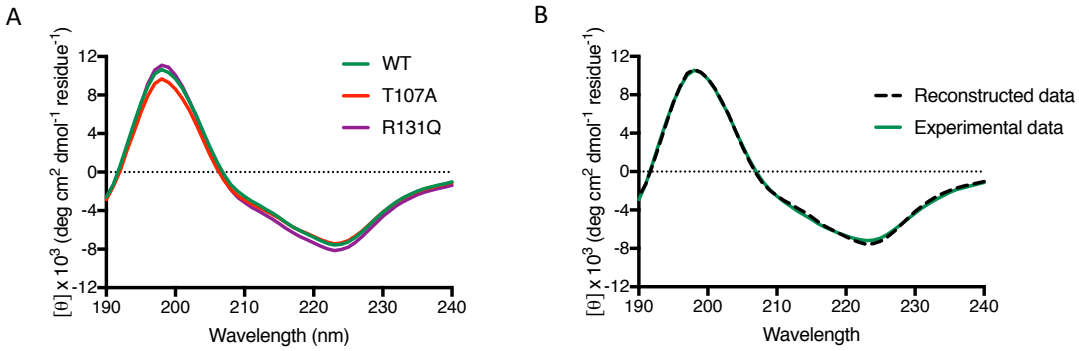
Both PPIL1 A99T and G109C; 101_106 dup proteins were highly unstable when expressed and purified from *E. coli*. Although small amounts could be purified, the protein rapidly aggregated and precipitated in solution. Due to these results, these mutant proteins could not be taken forward for further analysis. However, these results suggest that both PPIL1 A99T and G109C; 101_106 dup have a pathogenic effect on PPIL1 protein function by perturbing PPIL1 protein structure. All other proteins (WT, T107A and R131Q) were highly soluble and could be used for further analysis.

4.2.2.2 Circular dichroism (CD)

In order to determine whether any of the other patient variants had more subtle effects on PPIL1 protein structure, secondary structure content was compared between WT and mutant proteins. Circular dichroism spectroscopy (CD) at far UV wavelengths (190-240 nm) allows the rapid determination of protein secondary structure content and was therefore used to compare the secondary structure content of WT and mutant proteins. CD signals arise as a result of different absorption of right and left-handed circularly-polarised light. The peptide bond absorbs light below 240 nm and yields CD spectra with maxima and minima peaks at characteristic wavelengths, based on secondary structure content[119]. The spectra recorded for PPIL1 proteins show negative peaks at 223 nm and 208 nm consistent with alpha helical content, as expected for PPIL1 proteins (Figure 24A). Beta sheets have a less well-defined predicted maxima (from 190 nm to over 200 nm) due to the variation in possible conformations beta sheets can take. The CD spectra also contain a positive peak at 198 nm consistent with a significant beta sheet content, as expected (Figure 24A).

Spectra for WT and mutant proteins overlay, suggesting that the T107A and R131Q substitutions do not significantly perturb secondary structure in PPIL1 (Figure 24A). The secondary structure content was deconvoluted using Dichroweb online tool[147]. CDSSTR was the only analysis programme in Dichroweb which was able to fit the data (see Figure 24B for representative fit). This programme carries out a large number of calculations compared to

other programmes and often produces results where other analysis programmes fail[148]. NRMSD (normalized root-mean-square deviation) scores were below 0.05 for all proteins tested, suggesting a good fit to the data (Figure 24C). The predicted proportion of residues found in helices, beta strand or other conformation does not differ significantly between PPIL1 WT and mutant proteins and is in line with the published PPIL1 structures (Figure 24C). These results therefore suggest that the patient variants do not significantly perturb secondary structure content of PPIL1.



C

Protein	Helix	Stand	Other	NRMSD score
WT	0.18	0.28	0.55	0.04
T107A	0.18	0.26	0.56	0.04
R131Q	0.19	0.27	0.55	0.03
NMR structure	0.15	0.27	0.58	
Crystal structure	0.16	0.30	0.55	

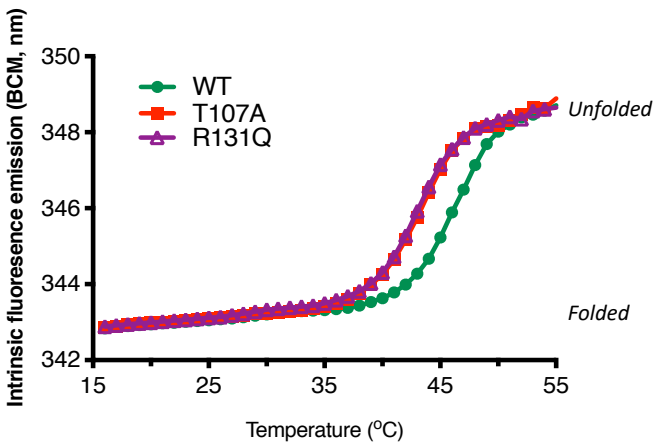
Figure 24 Far UV circular dichroism (CD) of PPIL1 wild type and mutant proteins. (A) Overlaid far UV CD spectra of PPIL1 proteins (WT, T107A and R131Q) at 20 °C. Spectra were acquired using wavelengths 190-240 nm and CD signal is shown as mean residue ellipticity [θ]. Samples consisted of 0.2 mg/ml protein in 10 mM sodium phosphate buffer, 50 mM NaF, pH 7.4. (B) Secondary structure content was estimated using CDSSTR on the Dichroweb online tool (http://www.cryst.bbk.ac.uk/cdweb)[147], using reference set 7, which is optimised for wavelengths of 190-240 nm. The reconstructed fit of the data calculated by CDSSTR is shown for PPIL1 WT CD spectrum. (C) Table of estimated secondary structure content of PPIL1 proteins, calculated by CDSSTR. Secondary structure content of PPIL1 in published crystal (PDB: 2X7K)[56] and NMR (PDB: 1XWN)[55] structures is shown for reference. (Helix includes α helices, 3_{10} helices and π-helices and sheet includes β strands).

4.2.2.3 Optim: protein unfolding and aggregation assays

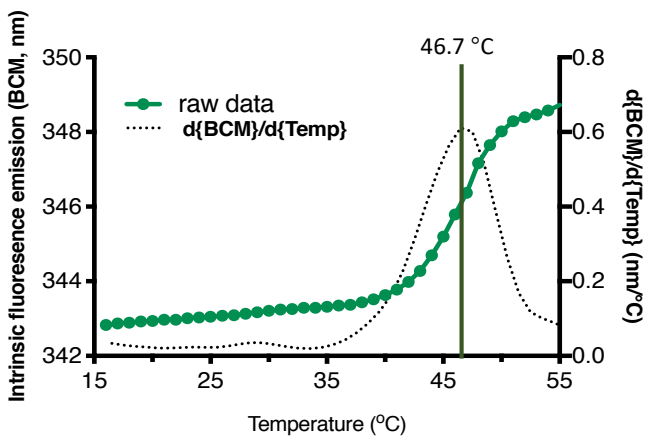
To establish whether the patient variants had any effect on the overall structure and stability of PPIL1, unfolding and aggregation screens were carried out, using an Optim machine. Protein unfolding was measured over a temperature gradient, as described previously (section 4.2.1.4). Intrinsic protein fluorescence was used to measure protein unfolding, as before. All PPIL1 proteins show a single transition from folded to unfolded proteins (Figure 25A, B), as is expected for small, globular, single domain proteins. The unfolding curves for PPIL1 T107A and R131Q proteins are shifted to the left, compared to PPIL1 WT, showing that PPIL1 T107A and R131Q begin to unfold at lower temperatures than PPIL1 WT (Figure 25A). The midpoint of the unfolding curve, called the transition temperature (Figure 25B), was determined using three separate preparations of protein. These results show that the unfolding transition temperature of PPIL1 WT is 46.8 ^{o}C, under the conditions tested. PPIL1 T107A and R131Q unfold at 3.0 and 3.4 ^{o}C lower temperatures, respectively, compared to PPIL1 WT.

In addition to the protein unfolding screen, protein aggregation was measured simultaneously across the same temperature gradient. Protein aggregation was measured as an increase in static light scattering at 266 nm. Comparison of the temperatures at which the static light scattering signals begin to increase, shows that PPIL1 T107A and R131Q are also more prone to aggregation than PPIL1 WT (Figure 25D). This trend was consistent across three experimental replicates with different preparations of protein.

A



B



C

Sample	Transition Temperature (°C)
WT	46.8 ± 0.46
T107A	43.8 ± 0.29
R131Q	43.4 ± 0.32

D

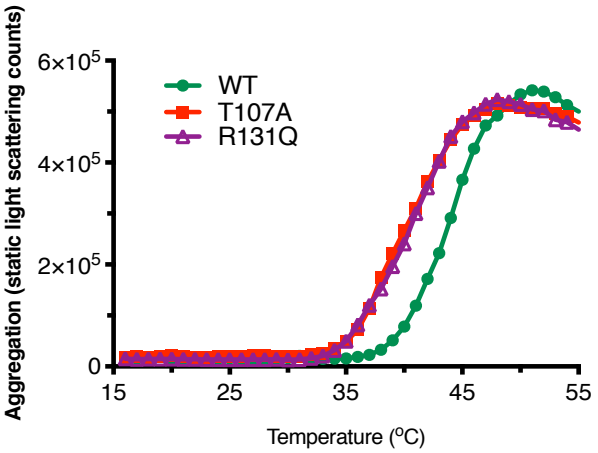


Figure 25 Protein unfolding and aggregation screen of PPIL1 wild type and mutant proteins. (A) Shift in the barycentric mean wavelength of intrinsic protein fluorescence was used to measure PPIL1 protein unfolding over a temperature gradient. (B) The transition temperature of folded to unfolded protein was calculated as the steepest point of the unfolding curve, which was determined by plotting the differential of the unfolding curve. (C) Mean transition temperatures of three experimental replicates, using different preparations of protein, are shown with standard deviation error. (D) Static light scattering was used to measure aggregation over a temperature gradient.

4.3 Discussion

PPIL1 WT protein and mutant proteins were successfully expressed and purified from *E. coli*. However, both PPIL1 A99T and G109C;101_106dup mutations disrupt the structure of PPIL1, causing significant aggregation and precipitation out of solution. The crystal structure of PPIL1 shows that A99 is a buried residue in PPIL1 protein core[56], so this may explain why substitution of this residue has such a profound effect on protein stability and aggregation. G109C;101_106dup causes both a six amino acid duplication and amino acid substitution in the loop between β5 and 6 strands of PPIL1. This loop shows no variation in size across human cyclophilins or between PPIL1 homologs in different species and may be critical to PPIL1 stability (Chapter 3).

SEC results suggest that all soluble PPIL1 proteins are monomeric. However, there was a small discrepancy in calculated apparent molecular weight of PPIL1 proteins (15 kDa), compared to the predicted molecular weight of PPIL1 proteins (19 kDa). Increasing the buffer salt concentration to reduce any non-specific binding to the column, which may cause this difference, did not cause PPIL1 proteins to elute at a different volume. Mass spectrometry results confirmed that the actual protein molecular weight is 19 kDa, as expected. The heterogenous shape of proteins mean that small deviations from theoretical protein molecular weight are often observed when molecular weight is deduced from SEC elution volumes[141]. PPIL1 WT protein has previously been characterized as monomeric using dynamic light scattering experiments[149]. Here it was shown that soluble WT and mutant PPIL1 proteins elute at similar volumes in SEC experiments. Taken together, the work by Xu et al. and our SEC experiments, indicate that all soluble PPIL1 proteins are monomeric (WT, T107A and R131Q)[56].

Characterisation of all soluble PPIL1 proteins, using SDS-PAGE, MS and SEC, has shown that these protein are of high purity and homogenous. These properties mean that further characterisation of these proteins was possible. PPIL1 T107A and R131Q mutant proteins exhibit reduced thermal stability and increased aggregation compared to PPIL1 WT, suggesting that these substitutions perturb the structure of PPIL1. Both T107A and R131Q substitutions affect solvent exposed residues, unlike A99T which is buried in the core of PPIL1. This likely explains why these substitutions have a more subtle impact on PPIL1 protein stability and aggregation compared to A99T.

However, it is clear that all patient PPIL1 variants have some impact on PPIL1 protein stability.

The availability of PPIL1 proteins will allow further investigation of the impact of PPIL1 patient variants on PPIL1 protein function. Little is understood about PPIL1 function, but it is known that PPIL1 interacts with another protein, SKIP, in the spliceosome[55]. Expression and purification of GST-SKIP 59-129 and GST was also successful and will therefore allow the interaction between SKIP 59-129 and PPIL1 to be studied (see Chapter 5). PPIL1 is also a PPIase enzyme[55] and the effect of the patient variants on this activity will be measured using the purified PPIL1 proteins (see Chapter 5).

Chapter 5 Effect of PPIL1 patient variants on protein function: interaction with SKIP and PPIase activity

5.1 Introduction

Biallelic variants at the PPIL1 locus cause a novel form of microcephaly and cortical dysplasia, together with cerebellar and brainstem hypoplasia (unpublished data: University of Leeds, UCSD). The variants result in the following PPIL1 protein changes: A99T, T107A, R131Q, G109C;101_106dup. In this chapter, the effect of PPIL1 patient variants on PPIL1 protein function was investigated. Little is known about the function of PPIL1; however, it is known that PPIL1 binds to SKIP and also exhibits peptidyl prolyl isomerase (PPIase) enzyme activity[55,56,86]. PPIL1 WT and mutant proteins that correspond to the PPIL1 patient variants have previously been expressed and purified (Chapter 4). PPIL1 A99T and G109C;101_106dup destabilise PPIL1 protein and cause significant aggregation and precipitation out of solution (Chapter 4). These results show a clear consequence of these variants; however, this also meant that these proteins were not suitable for further analysis. PPIL1 WT, T107A and R131Q proteins are soluble (Chapter 4) and were assessed for their interaction with SKIP and PPIase enzyme activity.

The PPIL1-SKIP interaction takes place in the spliceosome, a complex and dynamic ribonucleoprotein enzyme which splices introns from pre-mRNA[30,47,77]. SKIP is a 61.5 kDa, predominantly disordered protein, which spans a large region of the spliceosome[47,48,77]. SKIP is able to interact with many proteins throughout the spliceosome and it is thought that SKIP may play a key role in coordinating dynamics across the spliceosome[48,50]. SKIP binds to PPIL1 via a surface of PPIL1 distinct from the active site; which includes residue R131, a residue affected by a PPIL1 patient mutation (R131Q; see Chapter 3). The PPIL1 binding region of SKIP is found between residues 59-129 of SKIP, which is a disordered region[55]. It is known that a linear epitope of SKIP (residues 61-68) is essential for the interaction with PPIL1[56,86]. The PPIL1-SKIP interaction induces a disorder to order transition

of SKIP, and residues 59-79 of SKIP form a hook like structure on the surface of PPIL1 upon binding[56,86].

The impact of PPIL1 patient variants on the interaction with SKIP was investigated using surface plasmon resonance (SPR), a technique which has been successfully used previously, to study this interaction[55,86]. SPR monitors changes in the refractive index at a chip surface. Changes in mass near the chip surface, from processes such as binding, result in detectable shifts in the refractive index which are measured in arbitrary response units (RU)[150]. This technique offers many advantages over other protein-protein interaction methodologies, including the ability to use unlabeled protein, the potential to derive kinetic and affinity values from the data, and the requirement for relatively small quantities of protein interaction partners compared to other techniques[150]. GST:SKIP 59-129 and PPIL1 proteins were successfully expressed and purified from *E. coli* (see Chapter 4) and were used in SPR experiments to study this interaction.

PPIL1 also exhibits PPIase activity[55], the ability to catalyse the transition between a *cis* and *trans* peptide bond, before a proline residue. The type of peptide bond, *cis* or *trans*, may have important consequences for the local structure of the polypeptide chain. The peptide bond isomer may therefore be important for protein folding or protein conformational changes[34]. Proline isomerization is an intrinsically slow process and PPIase enzymes are ubiquitous in eukaryotes and prokaryotes in order to accelerate this interconversion (see Chapter 1 for more information)[35]. PPIL1 has previously been shown to exhibit PPIase activity on the model substrate *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (suc-AAPF-pNA)[55].

NMR was used to investigate the effect of PPIL1 patient variants on the PPIase enzyme activity of PPIL1. This analytical technique can be used to monitor proline isomerisation and is not subject to the same limitations of alternative assays[151]. The first published PPIase assay monitors proline isomerization on the tetrapeptide substrate suc-AAPF-pNA and uses chymotrypsin to cleave *p*-nitroanilide (pNA) from the peptide. Chymotrypsin can only cleave pNA from the peptide in the *trans* conformation of the Ala-Pro peptide bond and therefore cleavage of pNA is limited by the rate of proline isomerization. Accumulation of pNA can be used to monitor the rate of proline isomerization and is measured as increase in absorbance at 390 nm over time[152,153]. Although this has been a useful assay to identify PPIase enzymes[55,154], it has many limitations including susceptibility of some PPIase

enzymes to chymotrypsin cleavage[155-157]. This has previously led to some PPIase proteins incorrectly being labelled as inactive or having reduced activity[155-157]. A modified version of this assay, without chymotrypsin, has also been used[157]. This assay can also use the same substrate (suc-AAPF-pNA) and uses the difference in absorption coefficients of the peptide in the *cis* and *trans* isomers to monitor proline isomerization[157]. However, both of these assays are limited by low sensitivity and also use non-equilibrium conditions which may not be biologically relevant. NMR is sensitive to changes in the chemical environment of nuclei and can therefore be used to investigate proline isomerisation. This technique allows proline isomerisation to be monitored under equilibrium conditions[151]. In addition, there is greater flexibility in the buffers which can be used for NMR, allowing experiments to be performed in more physiologically relevant buffers.

In this chapter, PPIase enzyme activity of PPIL1 was monitored on the model substrate, suc-AAPF-pNA, using NMR. Specifically 1D ^{1}H NMR was used to monitor proline isomerization (as used by Hubner et al.[151]), which has high sensitivity to the changes in the chemical environment of nuclei that occur during proline isomerisation. In addition, 1D ^{1}H spectra can be acquired rapidly in minutes. Lineshape analysis of the spectra was used to measure rates of proline isomerisation, as the degree of line broadening can be used to measure chemical exchange such as proline *cis-trans* isomerisation. NMR lineshape is sensitive to slow-intermediate exchange (ms timescale), which is the timescale under which PPIase catalysed proline isomerisation typically takes place[151]. This technique allowed proline isomerization to be measured under equilibrium conditions without the use of chymotrypsin and enabled comparison of PPIL1 WT and mutant activity on suc-AAPF-pNA.

5.2 Results

5.2.1 PPIL1 interaction with SKIP

The impact of PPIL1 patient variants on the interaction between PPIL1 and SKIP was investigated using Biacore surface plasmon resonance (SPR). In order to measure the interaction, one component, the ligand, is immobilised on the chip surface and the interaction partner, or analyte, is then injected over the chip surface. GST:SKIP 59-129 (ligand) was immobilised on a sensor chip surface using an affinity capture system of anti-GST antibodies (Figure 26A). GST tag alone was also immobilised on anti-GST antibodies, in a separate flow cell, to confirm PPIL1 proteins did not bind to the GST tag. The anti-GST antibodies were first immobilised on the chip surface using amine coupling. This technique covalently links primary amine groups on the protein of interest (protein N-terminus or lysine sidechains) with the dextran matrix. This technique allowed immobilisation of approximately 10 000 RU of anti-GST protein on each flow cell and an example amine coupling sensorgram is shown (Figure 26B). After this procedure was complete, GST:SKIP 59-129 or GST (control) proteins were immobilised on the anti-GST antibodies using injections of protein into the relevant flow cell (e.g. Figure 26C). GST:SKIP 59-129 was immobilised at a density of approximately 700 RU and GST at a density of approximately 560 RU (e.g. Figure 26C).

All soluble PPIL1 proteins (WT, T107A, R131Q) were subsequently injected across the SPR chip surface to detect binding with SKIP 59-129 (Figure 26A, Figure 27A). Clear binding curves can be observed upon injection of PPIL1 WT protein across the flow cell containing SKIP 59-129, confirming binding between PPIL1 WT and SKIP 59-129 (Figure 27A), as shown previously[55,86]. PPIL1 T107A also interacts with SKIP 59-129, as shown by the binding curve generated upon PPIL1 T107A protein injection across SKIP 59-129 surface (Figure 27A). However, no response could be detected after injection of PPIL1 R131Q across SKIP 59-129 (Figure 27A), suggesting this amino acid substitution disrupts the interaction with SKIP. These results are consistent across three repeat experiments with separate preparations of PPIL1 and SKIP proteins. Control titrations against GST tag alone confirm PPIL1 proteins do not bind to GST tag (or anti-GST antibodies), (Figure 27B). The

Biacore software was unable to fit a 1:1 Langmuir binding model to any of the data which closely followed the binding curves obtained; therefore, kinetic or affinity values for the interaction were not determined (Appendix Figure 58). This was despite efforts to optimise and verify experimental conditions which might otherwise prevent 1:1 binding, including: ensuring purity and homogeneity of ligand and analyte protein, minimising mass transport effects, minimising non-specific binding, minimising bulk refractive index changes and use of appropriate reference and control flow cells (see discussion). Therefore, presented here are qualitative SPR data only (Figure 27A); however, these results demonstrate a clear pathogenic effect of the R131Q patient mutation on PPIL1 function.

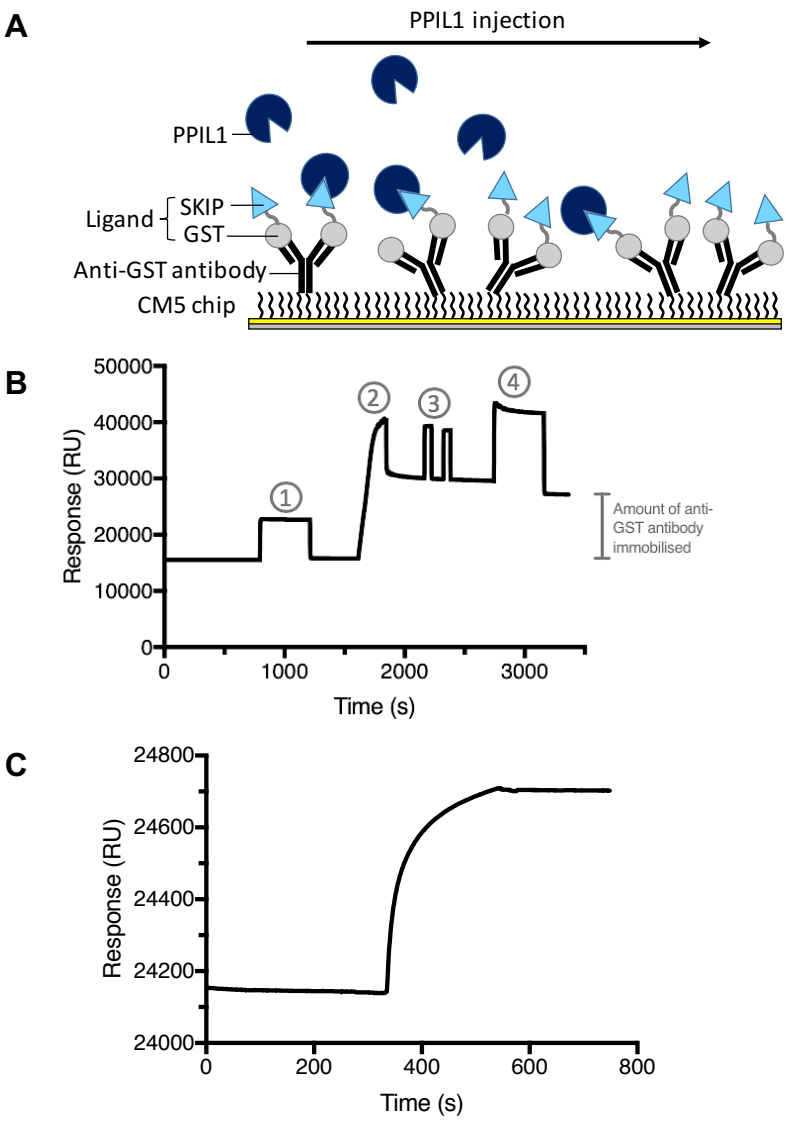


Figure 26 Experimental design and ligand immobilisation for PPIL1 and SKIP 59 -129 surface plasmon resonance (SPR) binding experiments. (A) SPR experimental design. Anti-GST antibodies were immobilised on a CM5 chip surface using amine coupling. GST:SKIP 59-129 protein or GST protein (control) was captured on the anti-GST antibodies and PPIL1 protein was injected across the surface to measure binding.
(B) Representative sensorgram for amine coupling of anti-GST antibodies to the CM5 chip. Amine coupling was carried out using 0.1 M sodium acetate running buffer, pH 5.6, at 5 µl/min. The surface was activated with a 7 minute injection of NHS/EDC (step 1). Anti-GST antibodies were injected across the activated matrix at 30 µg/ml for 4 minutes (step 2). Unreacted material was eluted with two, one-minute injections of 1 M NaCl at increased flow rate (20 µl/min; step 3). The surface was deactivated with a 7 minute injection of 1 M ethanolamine·HCl, pH 8.5 (step 4). More than 10 000 RU of anti-GST antibodies were immobilised on the chip surface, in each flow cell, using this method.
(C) Representative sensorgram of GST protein capture on anti-GST antibodies. GST and GST:SKIP 59-129 proteins (20 µg/ml) were injected across the chip surface at 5 µl/min in PBS (pH 7.4), 0.05 % v/v IGEPAL running buffer.

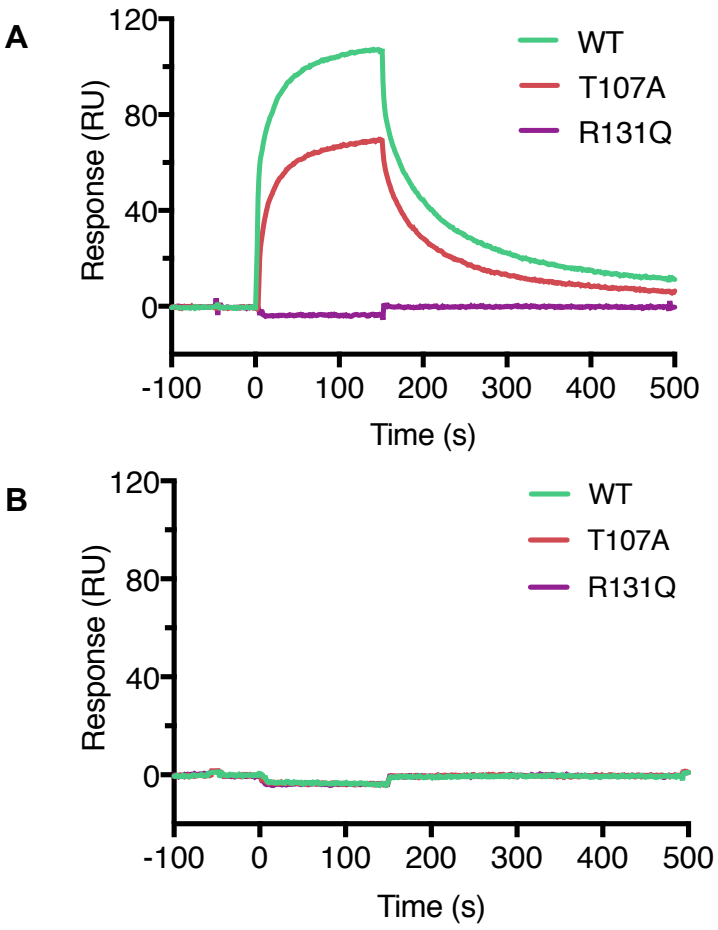


Figure 27 Surface plasmon resonance (SPR) binding and control experiments to analyse the interaction between PPIL1 proteins and SKIP 59-129. (A) PPIL1 protein injections (1 μM) across GST:SKIP 59-129. (B) GST was immobilised on the surface of a separate flow cell and PPIL1 proteins were also injected across the surface, to ensure PPIL1 proteins do not bind to the GST tag. The response is measured in arbitrary response units (RU), which is proportional to the refractive index change at the chip surface, as a result of binding.

5.2.2 PPIL1 PPIase activity

The effect of the PPIL1 patient variants on peptidyl prolyl *cis trans* isomerase (PPIase) enzyme activity of PPIL1 was determined on a model peptide substrate suc-AAPF-pNA. 1D ^{1}H NMR was used to measure proline isomerisation of the Ala-Pro peptide bond. NMR is highly sensitive to chemical exchange processes, such as proline isomerisation, and results in line broadening of signals for affected residues[151]. The degree of line broadening is proportional to the rate of proline isomerisation (relative to the chemical shift difference [Hz] between the *cis* and *trans* resonances)[151] and could be quantified using the mathematical technique, lineshape analysis.

1D ^{1}H spectra of the substrate were recorded, firstly in the absence of PPIL1 enzyme (Figure 28A). 1D ^{1}H experiments were suitable to monitor proline isomerisation in this instance, as the substrate suc-AAPF-pNA is a small molecule (four resides in length), meaning that there are relatively few signals in the spectrum and therefore signals were sufficiently dispersed for analysis (Figure 28A). Using the published assignment of this peptide[127], signals corresponding to the Ala2 methyl protons of suc-AAPF-pNA were identified (Figure 28B). Ala2 methyl proton signals are well-resolved in the spectrum and show distinct, well-separated signals for the *cis* and *trans* peptide bond conformations (Figure 28B) and were therefore suitable to monitor proline isomerisation. The relative area under the *cis* and *trans* signals, corresponds to the populations of the peptide found in each conformation. Integration of these signals shows that *cis* and *trans* conformations of the peptide are present at 10 % and 90 % respectively, under the conditions used. This is in line with previous studies of suc-AAPF-pNA[127].

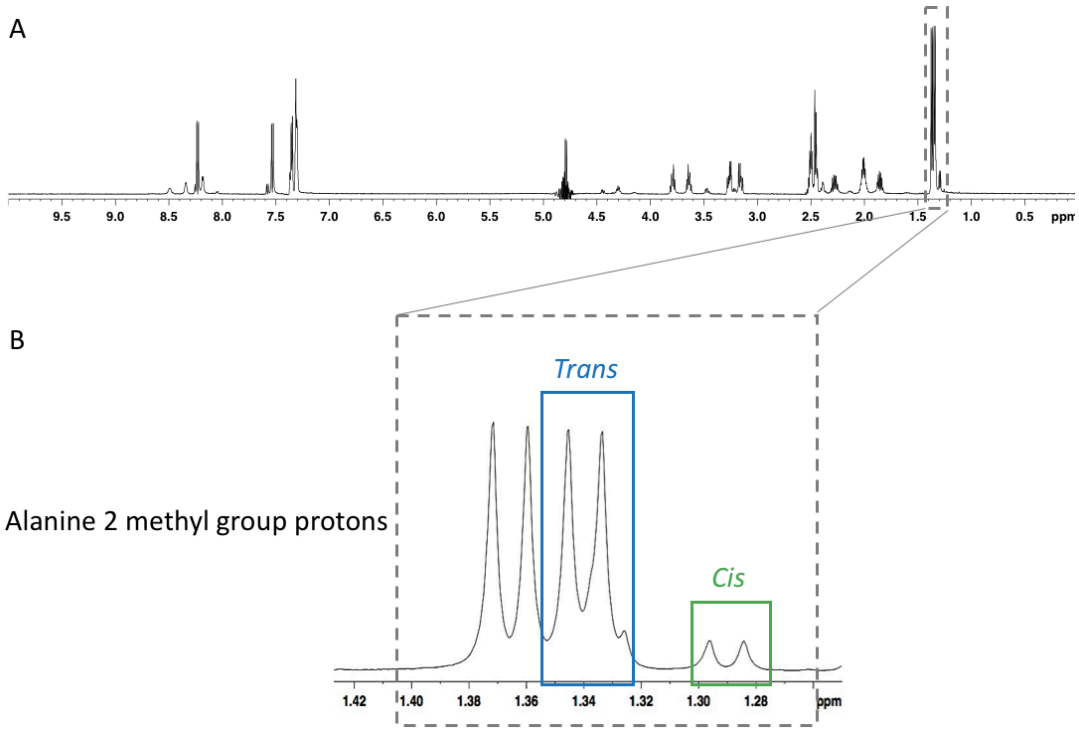


Figure 28 1D ^{1}H NMR spectrum of the model PPIase peptide substrate: *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide. Spectra were recorded in PBS at 25 °C. (A) Whole spectrum and (B) enlarged region of spectrum showing the region of the alanine 2 methyl proton resonances. The Ala-Pro peptide bond is found in both *cis* and *trans* conformations and gives rise to distinct signals in the spectrum. (Ala2 methyl proton signals appear as doublets due to coupling with Ala2 H_{α}.)

Spectra were recorded in the presence and absence of sub-stoichiometric concentrations of PPIL1 WT (Figure 29). Upon addition of PPIL1, *cis* and *trans* signals broaden, which is indicative of exchange (k_{ex}) between two states and thus here corresponds to increase in exchange between cis and *trans* conformations[151]. This confirms that as shown previously[55], PPIL1 is an active PPIase and can catalyse proline isomerisation on the model substrate. The PPIase activity is concentration dependent, as expected; demonstrated by increased line broadening upon addition of increasing concentrations of PPIL1 (Figure 29). It should be noted that addition of PPIL1 causes such a large degree of line broadening that the *cis* signals disappear into the noise; this is due to the low fraction of the *cis* conformation, (approximately 10 %). To further confirm that PPIase activity is the cause of line broadening, the PPIase competitive active site inhibitor CsA (Cyclosporin A), was added to the sample at equimolar ratios to PPIL1. (CsA has been reported to interact with PPIL1 at nanomolar affinity [9.8 nM][158] and equimolar ratios of CsA have

been shown to be sufficient to inhibit PPIase activity of cyclophilins[151]). Upon addition of CsA, line broadening is reduced and the spectra closely resembles that of peptide alone. This confirms that the observed activity is due to the action of the PPIL1 PPIase active site.

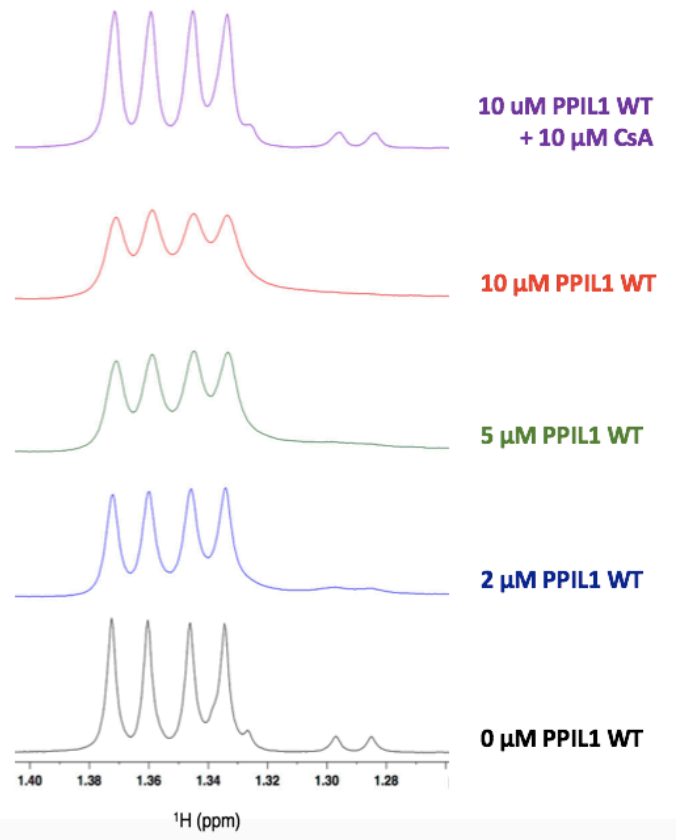


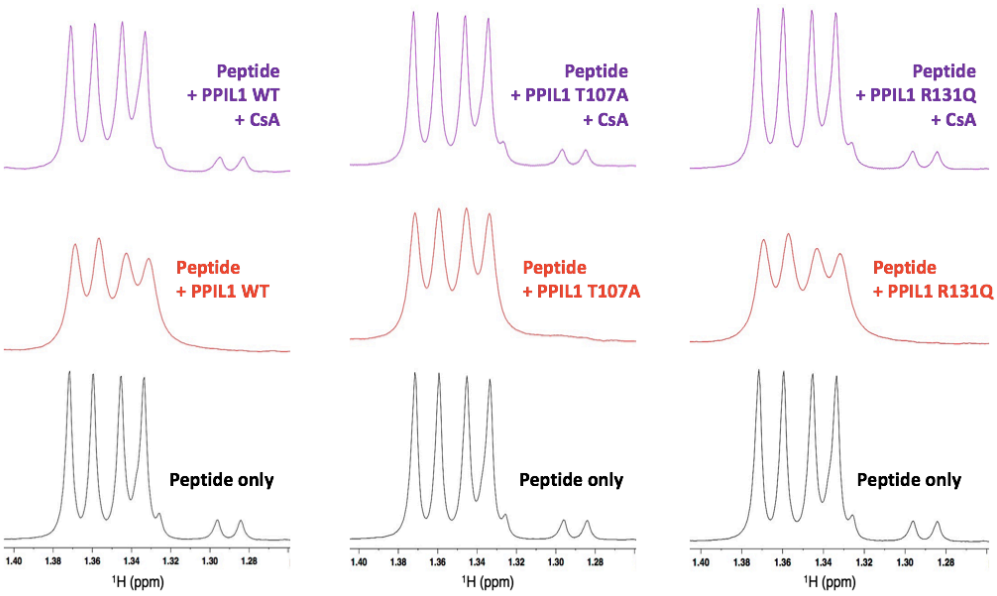
Figure 29 PPIase activity of PPIL1 WT. Spectra of the model peptide (*N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide) were recorded in the presence and absence of PPIL1. The region of the spectra containing Ala2 methyl proton signals that was used to monitor proline isomerization is shown. PPIL1 concentration was varied from 0-10 µM and the peptide concentration was held constant at 500 µM. Exchange between *cis* and *trans* Ala-Pro peptide bond conformations (i.e. proline isomerization) is indicated by line broadening. The PPIL1 competitive active site inhibitor CsA (cyclosporin A) was further added to samples, at an equimolar ratio to PPIL1, to confirm the observed activity is due to the active site of PPIL1 (purple spectrum).

The NMR assay was repeated to compare activity of PPIL1 WT and T107A and R131Q mutant proteins. Upon addition of catalytic concentrations of PPIL1 proteins (10 µM; WT, T107A and R131Q) to the peptide substrate, *cis* and *trans* Ala2 methyl proton signals broaden, indicating increased proline

isomerisation of the model substrate. This confirms that all three PPIL1 proteins tested (WT, T107A and R131Q) have PPIase activity. (Figure 30A). As before for PPIL1 WT (Figure 29), a control experiment was carried out in which the active site inhibitor (CsA) was added to the substrate and PPIL1 samples (Figure 30A). Addition of CsA reduces line broadening in all samples, confirming the activity observed is due to the PPIase active site (Figure 30A). Although all three proteins have PPIase activity on this substrate, T107A causes a smaller degree of line broadening compared to PPIL1 WT (Figure 30A). This suggests that PPIL1 T107A is a less active PPIase.

The rate of proline isomerisation was calculated from the spectra using the mathematical technique, lineshape analysis (Figure 30B). The analysis confirms that all PPIL1 proteins significantly accelerate proline isomerisation on the model substrate (Figure 30B). PPIL1 R131Q catalyses proline isomerisation to similar levels as PPIL1 WT (Figure 30B). However, addition of PPIL1 T107A results in a reduced rate of proline isomerisation, by twofold, compared to PPIL1 WT (Figure 30B).

A



B

Experimental condition	Rate of proline *cis-trans* isomerisation
Peptide alone	< 0.1 s^{-1}
Peptide + PPIL1 WT	9.6 ± 1.0 s^{-1}
Peptide + PPIL1 T107A	5.3 ± 0.2 s^{-1}
Peptide + PPIL1 R131Q	10.2 ± 1.5 s^{-1}

Figure 30 Comparison of PPIL1 WT and mutant PPIase activity on *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide peptide substrate. Proline isomerisation was measured by monitoring NMR lineshape of *cis* and *trans* signals (Ala2 methyl proton signals). (A) 1D ^{1}H NMR spectra were recorded of the peptide substrate alone (black) and with catalytic concentrations of PPIL1 protein (10 µM; red). Line broadening is indicative of exchange between *cis* and *trans* Ala-Pro peptide bond conformations (i.e. proline *cis-trans* isomerisation). As a control, CsA was subsequently added to the sample to block PPIase activity (purple). (B) Lineshape analysis was used to determine rates of *cis-trans* exchange in the substrate. The assay was repeated three times with independent preparations of protein and the mean and standard deviation of these results are shown. PPIL1 T107A shows reduced PPIase activity on this substrate compared to WT PPIL1.

5.3 Discussion

In this chapter, it has been shown that PPIL1 patient variants impact both binding to SKIP and PPIase activity of PPIL1. SPR results confirm that PPIL1 R131Q substitution completely abolishes binding with SKIP. This is consistent with previous reports which show that residue 131 is highly important for the interaction[56,86]. Specifically, R131 together with I97 of PPIL1 forms a hydrophobic groove of PPIL1 into which SKIP P65 binds. In addition, the sidechain of R131 also forms a hydrogen bond with the sidechain of Glu66 in SKIP[56]. SKIP is an essential splicing factor and is required for splicing to occur at the correct splice site[159]. Binding between PPIL1 and SKIP induces a disorder to order transition of SKIP and it is thought that this transition may contribute towards the large structural rearrangement of the spliceosome during assembly and catalytic activation[86]. In addition, it is proposed that this interaction is required to recruit PPIL1 PPIase active site to the correct location in the spliceosome, where it may bind to a target substrate[55,86].

It was not possible to achieve a good fit of the SPR data to a 1:1 Langmuir binding model and thus derive affinity and kinetic data for the interaction of PPIL1 proteins and SKIP. Previous SPR studies of this interaction[55,86] reported affinity values derived from 1:1 Langmuir binding model fits of the data. However, neither study showed the fits of the data[55,86]. Efforts were made to improve experimental design in order to rule out experimental conditions as the cause of deviations from a 1:1 interaction. PPIL1 analyte samples were confirmed to be of high purity and homogenous, as verified by SDS-PAGE, MS and SEC (Chapter 3). A low ligand immobilization density and high experimental flow rate were used to minimise the effects of mass transport. Any bulk refractive index changes were minimised by matching sample and running buffers, by extensive dialysis of PPIL1 proteins, and using a reference cell. Ligand homogeneity was ensured by confirming sample purity using SDS-PAGE and MS and also immobilising SKIP using an affinity capture system which results in a more uniform orientation of immobilised ligands. Non-specific binding was minimized by using a carboxymethyl dextran sodium salt in the sample buffer and was also controlled for with use of a reference cell. However, none of the above generated data that could be fitted to a 1:1 Langmuir binding model.

It is possible that the PPIL1-SKIP interaction is more complex than a 1:1 interaction. Cryo-EM data suggests that a second region of SKIP may also bind to PPIL1 (Appendix Figure 57)[47]. However, investigating this was beyond the scope of this project. Therefore, after the above efforts, further optimization was halted and it was decided that data would only be analysed qualitatively. Qualitative analysis, even in the absence of kinetic or affinity data was highly informative in this project and clearly demonstrated PPIL1 R131Q substitution abolished binding to SKIP.

NMR results also showed that T107A substitution reduced PPIase activity of PPIL1 on the model substrate suc-AAPF-pNA. This is unsurprising as residue T107 is positioned close to the active site (see Chapter 3). Specifically, T107 is part of the S2 pocket, a hydrophobic pocket which is thought to confer binding specificity for substrates of the neighbouring PPIase active site (see Chapter 1)[39]. T107 has been defined as one of the 'gatekeeper residues' of the S2 pocket, which may regulate binding of the substrate to the pocket as a result of the orientation and properties of the sidechain[39]. In addition, T107 is positioned in a flexible loop between beta 5 and beta 6 strands of PPIL1[55]. In the PPIL1 homolog, CypA, this loop becomes more flexible upon binding to a model substrate and it is thought that this flexibility may be crucial for substrate turnover, specifically by allowing Asn102 to form a hydrogen bond with the substrate[58]. It is clear that T107 and the surrounding loop have an important role in PPIase activity and therefore it is not surprising that T107A substitution affects PPIase activity of PPIL1. R131Q substitution does not significantly alter PPIase activity of PPIL1; however, this residue is found on a surface of PPIL1 distant to the active site[55,56].

All patient variants clearly perturb PPIL1 protein function either by affecting protein stability (Chapter 4), interaction with SKIP or PPIase enzyme activity. All characterized effects of the patient mutations on PPIL1 function, result in reduction or loss of PPIL1 function rather than any gain of function. Taken together, these results confirm the deleterious effect of PPIL1 patient variants on PPIL1 function.

Chapter 6 The role of PPIL1 active site in the spliceosome

6.1 Introduction

In this thesis it has been shown that PPIL1 patient variants, which cause a novel syndrome of pontocerebellar hypoplasia, clearly perturb known functions of PPIL1 protein (Chapter 3-5). However, the precise role of PPIL1 in the spliceosome remains unclear. In Chapter 5, NMR data demonstrated that a protein substitution caused by a patient variant (PPIL1 T107A), results in reduced PPIase activity of PPIL1 on a model substrate. However, it is not known what the role of the PPIL1 PPIase active site is in the spliceosome. The role of the PPIL1 active site in the spliceosome was investigated, to aid understanding of this disease. Advances in spliceosome biology also have implications for the understanding of many further diseases, as aberrant splicing activity is also implicated in a wide range of genetic diseases and cancers[96].

Cyclophilins, such as PPIL1, possess a PPIase active site, which can catalyse proline isomerisation - the transition between a *cis* and *trans* peptide bond preceding a proline residue (see Chapter 1). The type of peptide bond (*cis* or *trans*) affects the local conformation of the peptide chain, and the transition may be required for protein folding or protein conformational changes[34]. Several studies have shown that spliceosome cyclophilins, such as PPIL1, bind to other spliceosome components via surfaces outside the active site[55,160]; however the role of cyclophilin active sites in the spliceosome is unclear[43]. Cyclophilins that are associated with the spliceosome demonstrate a range of PPIase activity rates on a model substrate; and some spliceosome cyclophilins lack PPIase activity altogether[39,43]. Therefore, it is not clear whether PPIase activity is displayed by cyclophilins in the spliceosome.

Splicing reporter gene assays by Adams et al. suggests that PPIase activity may not be important in the spliceosome and instead the function of cyclophilins in the spliceosome may be to interact with other components on surfaces outside the active site[46]. However these conclusions are based on observations from *in vitro* splicing assays with a specific reporter gene, which also relies on addition of excess cyclophilin protein[46]. Therefore, these assays may not be biologically representative. It is now becoming clear that

even core components of the spliceosome, may have a modulatory role in the spliceosome, and the function of spliceosome components may be gene or intron specific; therefore, studies on reporter genes may not reveal the full function of spliceosome components[84].

Recent advances in both cryo-electron microscopy and spliceosome purification have meant that a large number of spliceosome structures have recently been released[65]. Due to the spliceosome's dynamic nature, it was previously difficult to determine spliceosome structures at high resolution. However, techniques to stall the spliceosome at various stages have now made it possible to determine the structure of a large part of the core spliceosome at near atomic level resolution[78,89]. Both the *S. pombe* ILS[48] and human C* spliceosome[47] structures have significant resolution for part of pre-mRNA-processing factor 17 (PRP17) N-terminal region. Detailed interrogation of the structures indicated that a loop of PRP17 (human PRP17 residues 89-101) interacts with PPIL1 in both structures (Chapter 3;Figure 17)[47,48]. Analysis of these structures showed that a highly conserved proline of PRP17 (human Pro95) is positioned directly in the PPIL1 active site (Figure 17; Figure 18). The sidechain of Phe93 also occupies the neighbouring S2 pocket of PPIL1, a region thought to influence binding specificity of PPIase substrates by interacting with sidechains of residues N-terminal to the prolyl peptide bond[39]. A phenylalanine residue at position i-2 (i.e. Phe93), has previously been shown to be favourable for PPIase substrate binding for a different cyclophilin, Cyclophilin A[161]. The N-terminal region of PRP17, including the putative PPIL1 interaction site (human PRP17 residues 89-101), is conserved in PRP17 from human to *S. pombe*; but is absent from S. cerevisiae which also lacks a PPIL1 homologue (Figure 18). (As this work was being completed further human spliceosome structures have been published which also indicate this interaction, [human B^{act}, C, P, ILS spliceosome complexes[13,35,36,49,50,77]] Chapter 3.)

PRP17 (also known as CDC40) is a core component of the spliceosome[30] and, like PPIL1, is recruited to the spliceosome during B^{act} stage of spliceosome assembly[49,50]. PRP17 consists of a C-terminal WD40 domain and N-terminal region of extended conformation[47,162]. The PRP17 WD40 domain is positioned close to the spliceosome catalytic centre and participates in the second catalytic step of the spliceosome[47,52,162]. The N-terminal region of PRP17 forms a large number of connections spanning over 100 Å of the spliceosome, including an interaction with PPIL1, which is positioned close to the periphery of the spliceosome[47]. The interactions of

spliceosome components with PRP17 N-terminus are proposed to be important for ensuring fidelity of the spliceosome catalytic centre[47,52]. It is also suggested that PRP17 coordinates conformational changes across the spliceosome, as many conformational changes coincide with stable integration of PRP17 into the spliceosome[50].

Given the evidence above, the hypothesis that PRP17 could act as the substrate of PPIL1 in the spliceosome, was tested. It was first confirmed that hPPIL1 and hPRP17 (residues 89-101) interact and that this interaction occurred at the PPIL1 active site. It was assessed whether the interaction had any impact on the conformation of PRP17. Patient variants in PPIL1 cause a novel syndrome of microcephaly and pontocerebellar hypoplasia (Chapters 1, 3, 4 and 5). The effect of the PPIL1 amino acid substitutions (T107A and R131Q) caused by the patient variants, on the PPIL1:PRP17 interaction was also investigated. Given that the interaction with PRP17 occurs at the PPIL1 active site and the cryo-EM data shows the PPIL1:PRP17 interaction involves a conserved proline residue of PRP17, it was also assessed whether PPIL1 has PPIase activity on PRP17. These results have given novel insights into the function of PPIL1 in the spliceosome.

6.2 Results

A peptide was synthesized (Peptide Synthetics) which corresponds to the putative PPIL1 binding region of PRP17 (residues 89-101 of PRP17; Figure 17 and Figure 18). This peptide was used in binding studies and an equivalent peptide with ^{13}C and ^{15}N labelled proline (Pro95) was used in initial proline isomerisation studies.

6.2.1 PPIL1 – PRP17 peptide interaction

6.2.1.1 NMR analysis

The interaction between PPIL1 and PRP17 was confirmed using ^{1}H ^{15}N heteronuclear single quantum coherence (HSQC) NMR. ^{1}H ^{15}N HSQC NMR is a highly valuable method to analyse protein interactions, as in addition to detecting an interaction, this technique can also provide information about the interaction site and affinity of the interaction[163]. ^{1}H ^{15}N HSQC spectra show H-N correlations of nuclei directly linked through covalent bonds, (by passing magnetisation between ^{1}H and ^{15}N nuclei; Figure 31). This spectrum yields a 'protein fingerprint' where specific signals for the amide group of each amino acid (except for proline and N-terminal residues) can be observed at specific chemical shifts. (Correlations in the following sidechain groups may also be seen in the spectrum: asparagine N_{δ}-$H_{\delta 2}$, glutamine N_{ε}-$H_{\varepsilon 2}$, Trp N_{ε}-H_{ε} and (at low pH) Arg N_{ε}-H_{ε}.)

PPIL1 was ^{15}N labelled using standard methods. A ^{1}H ^{15}N HSQC spectrum was first recorded of ^{15}N labelled PPIL1 alone, to evaluate suitability for NMR (Figure 32). The dispersion of signals in the ^{1}H dimension can be used to assess if a protein is folded, as ^{1}H nuclei are particularly sensitive to their environment[164,165]. In the ^{1}H ^{15}N HSQC spectrum of PPIL1, resonances are well dispersed in the ^{1}H dimension suggesting PPIL1 protein was folded (Figure 32). In addition, most signals are well resolved and of roughly uniform intensity, suggesting that PPIL1 protein is soluble and also suitable for further analysis (Figure 32).

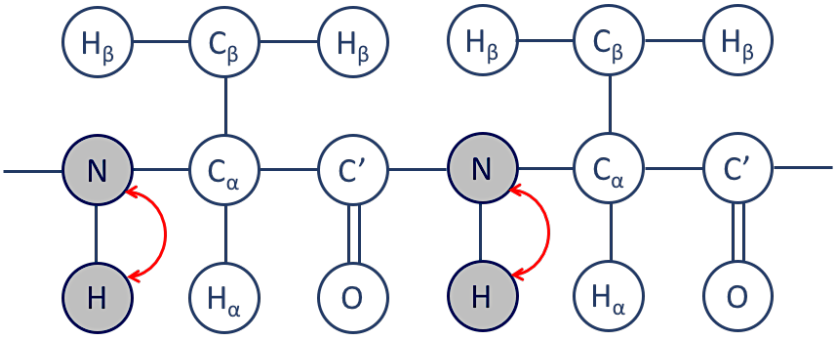


Figure 31 ^{1}H ^{15}N HSQC schematic, showing magnetisation pathway on two neighbouring amino acid residues. Magnetisation is first transferred from ^{1}H to ^{15}N (e.g. amide proton to nitrogen), chemical shift is then evolved on ^{15}N and magnetisation is finally passed back to ^{1}H for detection.

^{1}H ^{15}N HSQC spectra of PPIL1 were acquired with titrations of PRP17 peptide (0-5 molar equivalents), to measure binding. Upon addition of PRP17 peptide to PPIL1, resonances for certain residues of PPIL1 broaden beyond detection or shift, indicating PRP17 binds to PPIL1 (Figure 33). The chemical shift perturbations are indicative of specific binding rather than a non-specific effect; as only specific PPIL1 resonances broadened or shifted during the PRP17 titration and these shifts occur in different directions. Resonances which shift, move smoothly, from free to bound states suggesting a fast rate of exchange[iv]. (Exchange detected by these experiments may result from either exchange between bound and unbound states of PPIL1 or a conformational change in PPIL1 induced by PRP17 binding.) Several resonances also broaden beyond detection and this is likely indicative of exchange on the intermediate timescale (Figure 33).

The interaction site of PRP17 on PPIL1 was identified using the published assignment of PPIL1[124] and HSQC data described above. Weighted chemical shift changes were calculated for each residue (Figure 34). Resonances which shifted more than 1 or 2 corrected standard deviations above the mean or broadened beyond detection were considered to be significant (Figure 34) and were mapped onto the published crystal structure of PPIL1 (Figure 35)[56]. The significantly perturbed residues cluster around

[iv] Fast exchange is defined as k_{ex} (s^{-1}) >> Δδ, intermediate exchange as k_{ex} (s^{-1}) ≈ Δδ and slow exchange as k_{ex} (s^{-1}) << Δδ, where Δδ is the chemical shift difference in Hz between the interconverting states (e.g. free & bound forms) and k_{ex} is the overall exchange rate in seconds.

the active site surface of PPIL1 confirming that, as predicted from the cryo-EM structures, PRP17 binds to the PPIL1 active site (Figure 17; Figure 35). In addition, significantly perturbed residues also cluster around a neighbouring pocket to the active site, the S2 pocket (Figure 35). The S2 pocket binds residues N-terminal to the prolyl peptide bond and is thought to confer substrate binding specificity for the PPIase active site[39]. Significantly perturbed residues include those which have been specifically defined as either active site (R55, F60, N102 and W121) or S2 pocket 'gatekeeper' (T107, S110 and Q111) residues[39]. (Active site residues were defined as those which create the hydrophobic proline binding pocket or are within 4Å of the prolyl peptide bond in model structures of the PPIL1 homologue CypA[39,58]. S2 pocket gatekeeper residues include PPIL1 residues which have sidechains positioned around the entrance to the pocket and are proposed to control access to the pocket[39]). These results are therefore consistent with the published cryo-EM data, which suggests that PRP17 interacts with the active site and S2 pocket of PPIL1[47,48,77].

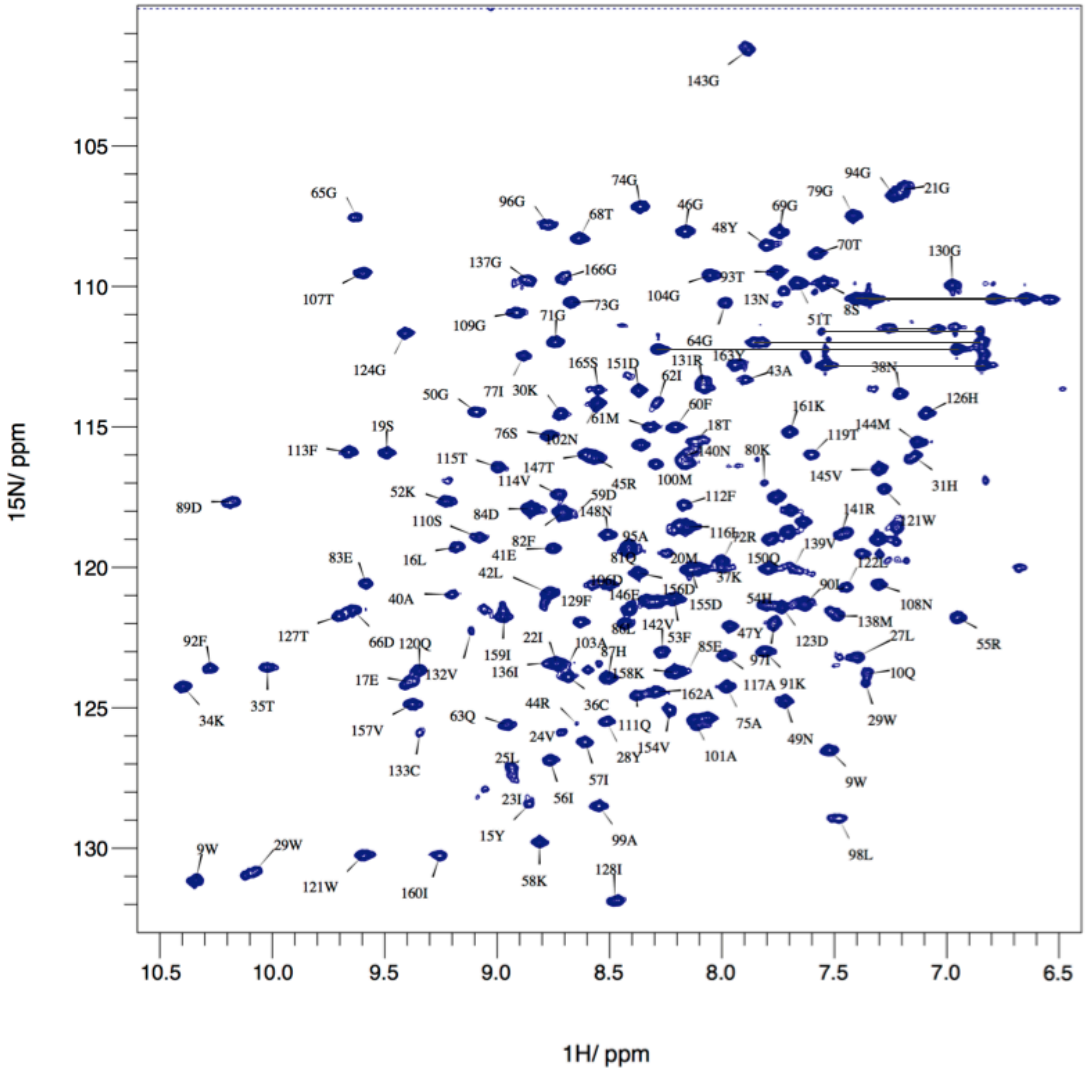


Figure 32 ^{1}H ^{15}N HSQC spectrum of ^{15}N labelled PPIL1 WT protein alone. Spectra were recorded using 80 µM PPIL1 protein in 10 mM sodium phosphate buffer, 100 mM NaCl, pH 6.5 in a 600 MHz NMR spectrometer at 25 °C Assignments were identified using the published assignments of PPIL1 by Xu et al.[124].

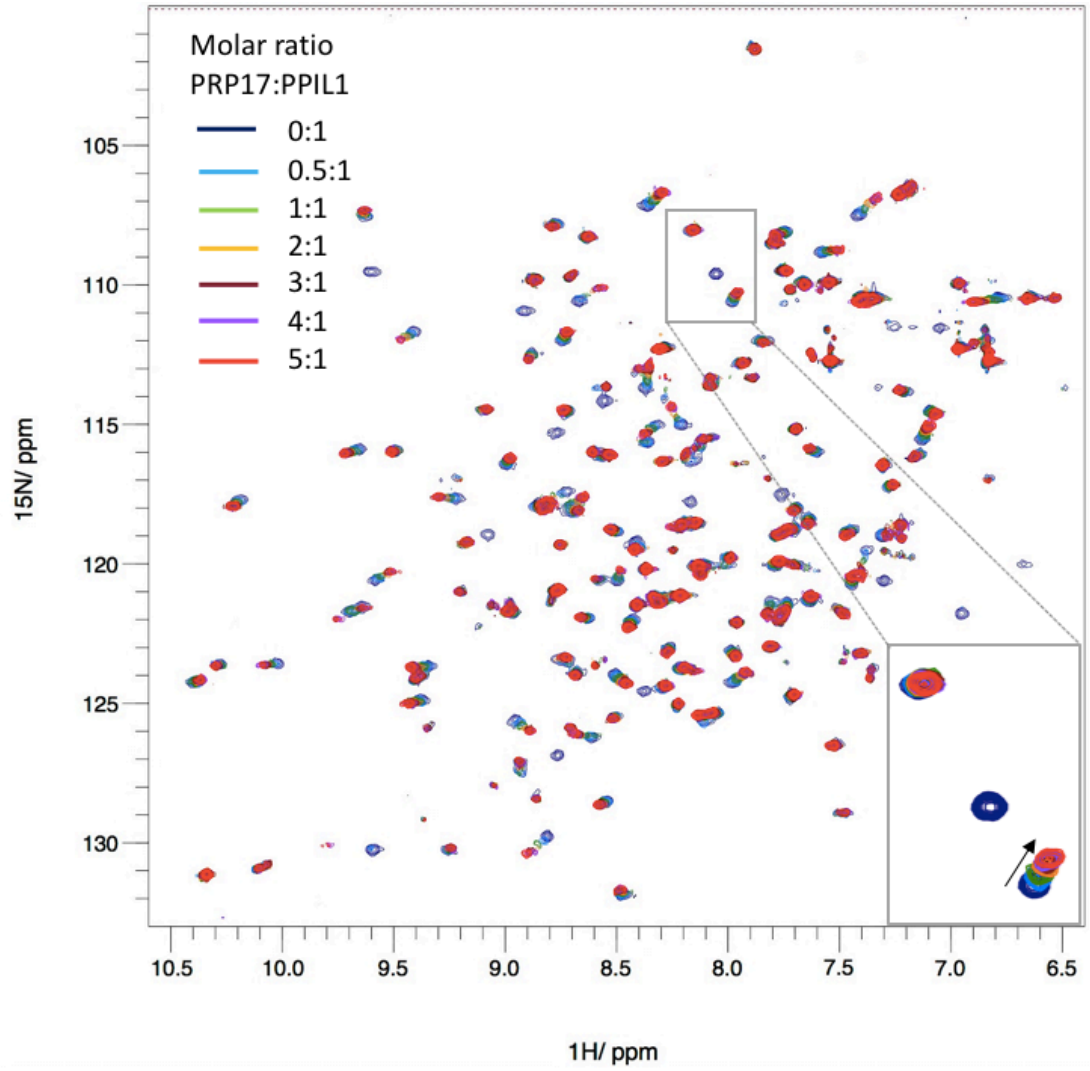


Figure 33 Overlaid ^{1}H ^{15}N HSQC spectra of PPIL1 WT protein before and after titrations with PRP17 peptide (0-5 molar equivalents). Concentration of PPIL1 protein used was 80 µM and PRP17 peptide was added at the relevant molar ratios. Specific resonances shift or broaden beyond detection indicating specific binding of PRP17 peptide to PPIL1. An example region of the spectra is enlarged to show resonances which shift (bottom), broaden beyond detection (centre) or are unaffected by the titration (top).

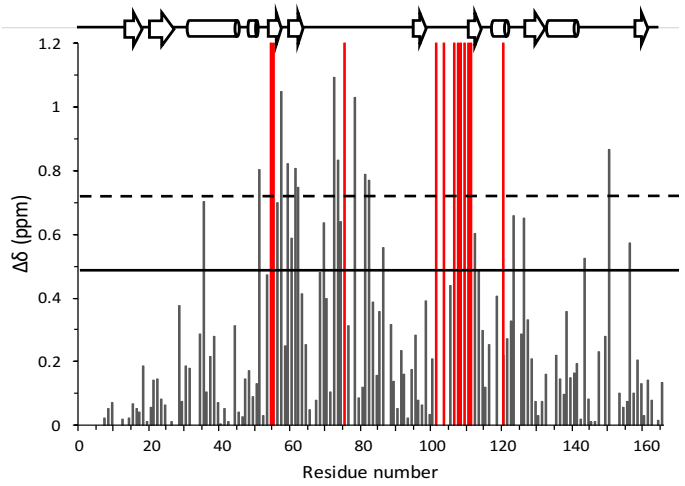


Figure 34 Histogram showing average chemical shift perturbations (Δδ) of PPIL1 residues upon addition of PRP17 peptide. PRP17 peptide was added to PPIL1 at a 5:1 molar ratio. Grey lines show the chemical shift differences (for residues which were visible in the NMR spectrum before and after addition of PRP17 peptide). Solid and dashed lines indicate the cut-off values for resonances which shift greater than 1 or 2 corrected standard deviations above the mean, respectively (as recommended by Schumann et al.[126]). Red lines show residues at which the signal disappears as a result of significant resonance broadening. Spaces are present at proline and unassigned residues. (Weighted chemical shift changes were calculated using the following equation: $\Delta\delta = \sqrt{(\Delta\delta_N)^2 + (\Delta\delta_H)^2(\gamma_H/\gamma_N)^2}$ from Hewitt et al. 2017[125], see Chapter 2)

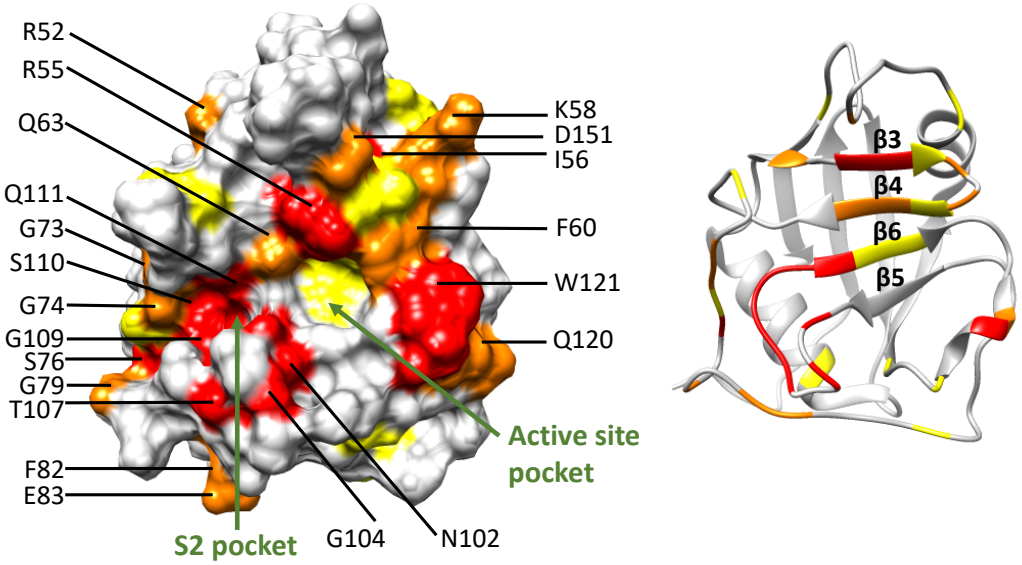


Figure 35 Identification of PRP17 interaction site on PPIL1, using chemical shift perturbation mapping. PPIL1 WT residues which showed significant chemical shift perturbations after titration with PRP17 peptide in ^{1}H ^{15}N HSQC spectra were mapped on the published structure of PPIL1. Residues at which the signal disappears as a result of significant resonance broadening are shown in red. Residues which shift > 2 or >1 corrected standard deviations above the mean are shown in orange and yellow respectively. Residues which broaden or shift significantly include residues which have been described previously[39] to define the active site (R55, F60, N102 and W121) and S2 pocket (T107, S110 and Q111). Images were created in Chimera using the published crystal structure of PPIL1[56].

PPIL1 PRP17 binding curves were plotted for all PPIL1 residues which shifted significantly in the HSQC spectra upon addition of PRP17 (Figure 36). The affinity of the PPIL1 PRP17 interaction was calculated using a global fit of the binding curves. The PPIL1 PRP17 interaction is low affinity (231.4 μM), which is consistent with the resonance shift patterns in the spectra, which move smoothly from an unbound to a bound state. However, this analysis does not consider residues which broaden beyond detection, as it was not possible to follow these resonances throughout the titration. (These resonances likely represent those for which the difference between free & bound chemical shifts [$\Delta\delta$] are much larger and therefore of the same order of magnitude as the exchange rate [k_{ex}].) ITC has also been used to confirm the interaction and determine the overall affinity of the interaction (see section 6.2.1.2).

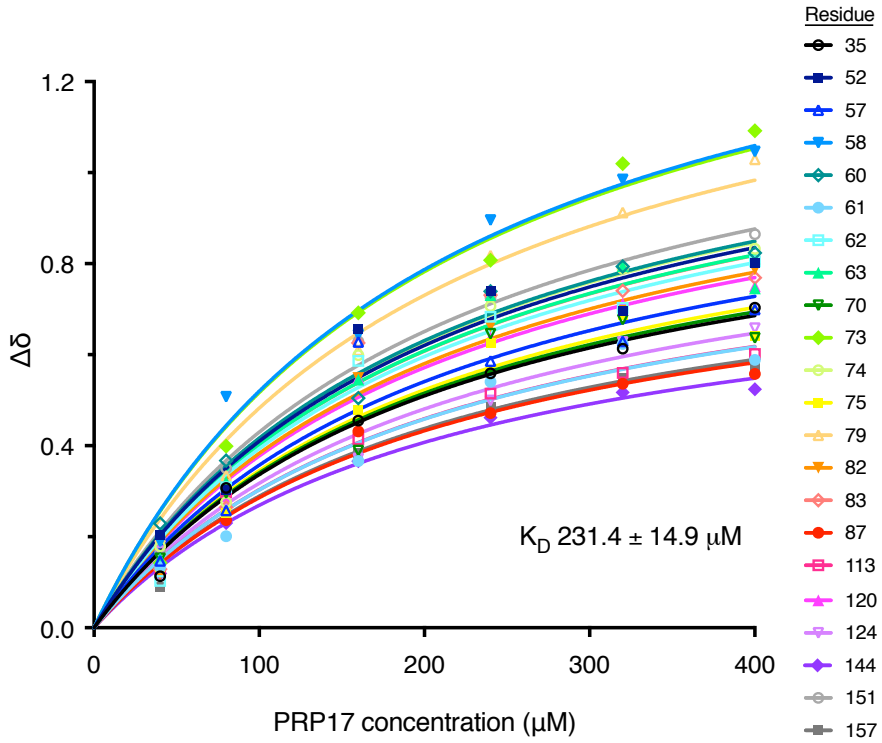


Figure 36 Global fitting of NMR ^{1}H ^{15}N HSQC titration curves. PRP17 peptide was titrated into ^{15}N PPIL1 protein (0-400 μM). For significantly shifting resonances (> 1 corrected standard deviation above the mean), chemical shift changes ($\Delta\delta$) are plotted against PRP17 concentration. The K_D was calculated with a global fit of the binding curves in Prism 7. (All datasets could be described by a single best fit K_D value; however, the B_{max} parameter was allowed to vary between curves due to the different magnitudes of $\Delta\delta$ between residues.) Any residues which were not possible to track throughout the titration were removed from the analysis (residues 114 and 127).

Given that variants in PPIL1 cause a novel syndrome of microcephaly and pontocerebellar hypoplasia (Chapters 1, 3, 4, 5), it was investigated whether the resulting PPIL1 residue substitutions (T107A and R131Q) affect the interaction with PPIL1. ^{1}H ^{15}N HSQC titration experiments of PPIL1 protein with PRP17 peptide were repeated for PPIL1 mutant proteins (T107A and R131Q), to determine if the amino acid substitutions had any impact on the PPIL1 PRP17 interaction. As shown for PPIL1 WT (Figure 33), specific resonances of mutant PPIL1 (T107A and R131Q) shift or broaden beyond detection upon addition of PRP17 peptide (Figure 37); indicating specific binding. Resonances which broaden or shift significantly in PPIL1 WT upon binding to PRP17 peptide, also broaden or shift in mutant PPIL1 (T107A and R131Q) upon binding to the PRP17 peptide (for resonances which could be assigned; Figure 38). This indicates that PRP17 peptide binds to the same site in all three PPIL1 proteins.

R131Q PPIL1 ^{1}H ^{15}N HSQC spectra closely resemble that of PPIL1 WT (Figure 33; Figure 37). However, for PPIL1 T107A, specific resonances show an unusual chemical shift pattern upon titration with PRP17 peptide. Unlike PPIL1 WT resonances which shift smoothly in a linear pattern, specific resonances for PPIL1 T107A residues move in a nonlinear fashion (Figure 37; Figure 39). This is only the case for resonances corresponding to specific residues of PPIL1 T107A and other resonances show smooth linear chemical shift patterns or do not shift (Figure 39). The affected residues are found in a loop connecting β strands 4 and 5, which lies below the S2 pocket in PPIL1 (Figure 40). These unusual shift changes are not simply representative of slow exchange, as resonances do not simply reduce in intensity and grow in intensity at a shifted location. Nonlinear chemical shift patterns indicate that something more complex than simple two state binding may be occurring (see discussion).

T107A

R131Q

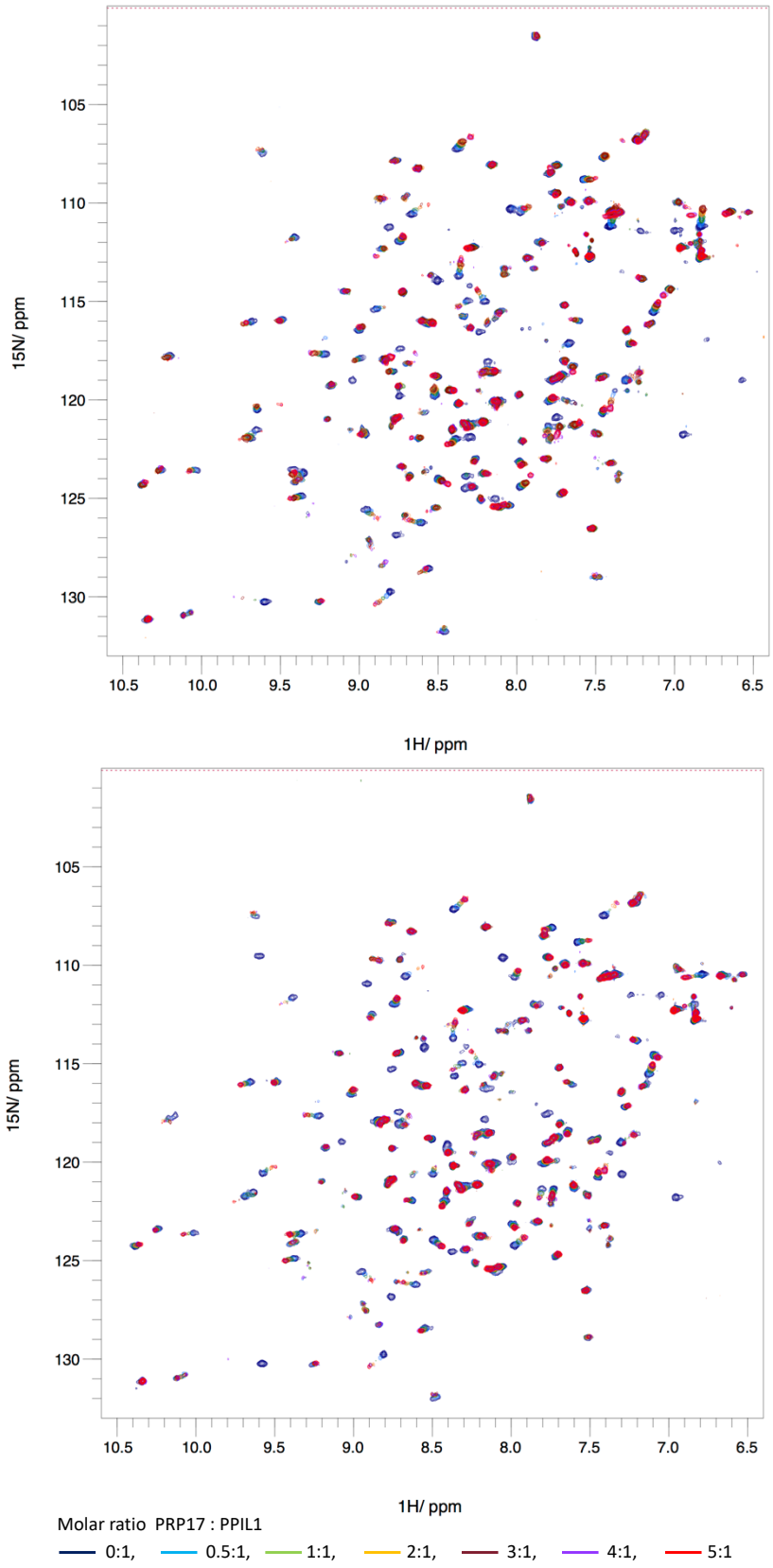


Figure 37 Overlaid ^{1}H ^{15}N HSQC spectra of PPIL1 T107A (top) and R131Q (bottom) protein before and after titrations with PRP17 peptide. Specific resonances shift or broaden beyond detection in the spectra, while other resonances are unaffected by the titration, indicating specific binding between PPIL1 proteins and PRP17 peptide.

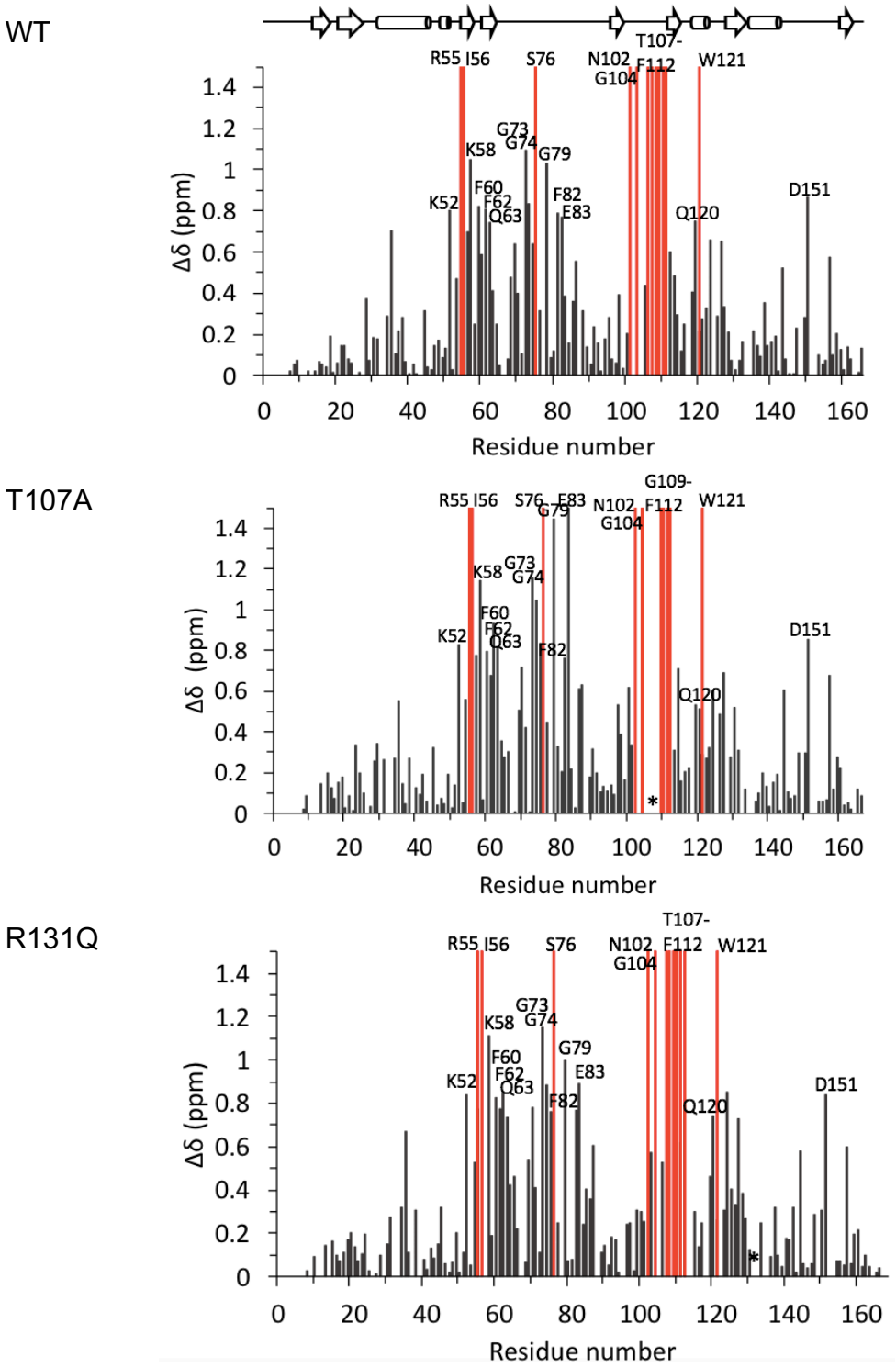


Figure 38 Chemical shift perturbation histograms of PPIL1 proteins (WT, T107A and R131Q) after addition of PRP17 peptide (5:1 molar ratio). PPIL1 residues which shift in the ^{1}H ^{15}N HSQC spectra, upon addition of PRP17 peptide, are represented by grey histogram bars; residues which broaden beyond detection are represented by red histogram bars. (* marks the position of the PPIL1 amino acid substitutions). Residues which broaden beyond detection or more than two corrected standard deviations above the mean for PPIL1 WT are labelled on the histograms.

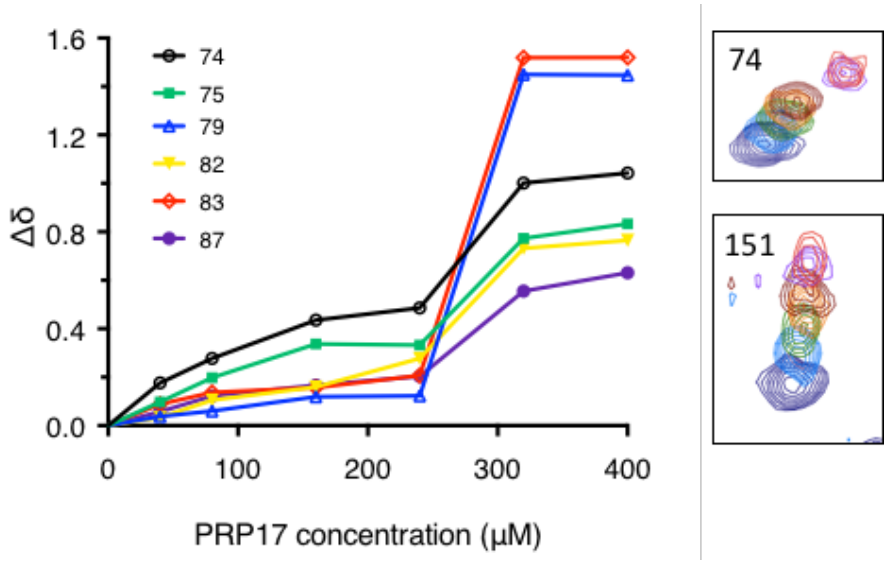


Figure 39 Addition of PRP17 peptide to PPIL1 T107A caused unusual chemical shift perturbation patterns for specific residues of PPIL1 T107A (74, 75, 79, 82, 83 and 87). ^{1}H ^{15}N HSQC spectra of ^{15}N PPIL1 T107A were recorded, with increasing titrations of PRP17 peptide, to measure binding. Specific resonances shifted significantly upon titrations with PRP17 peptide; however, some of these resonances shifted in a nonlinear manner (residues 74, 75, 79, 82, 83 and 87). Chemical shift perturbation (Δδ) is plotted against PRP17 peptide concentration (left). An example residue which shows an unusual nonlinear chemical shift perturbation (residue 74) and a residue which shifts in a smooth linear manner (residue 151) are shown (right).

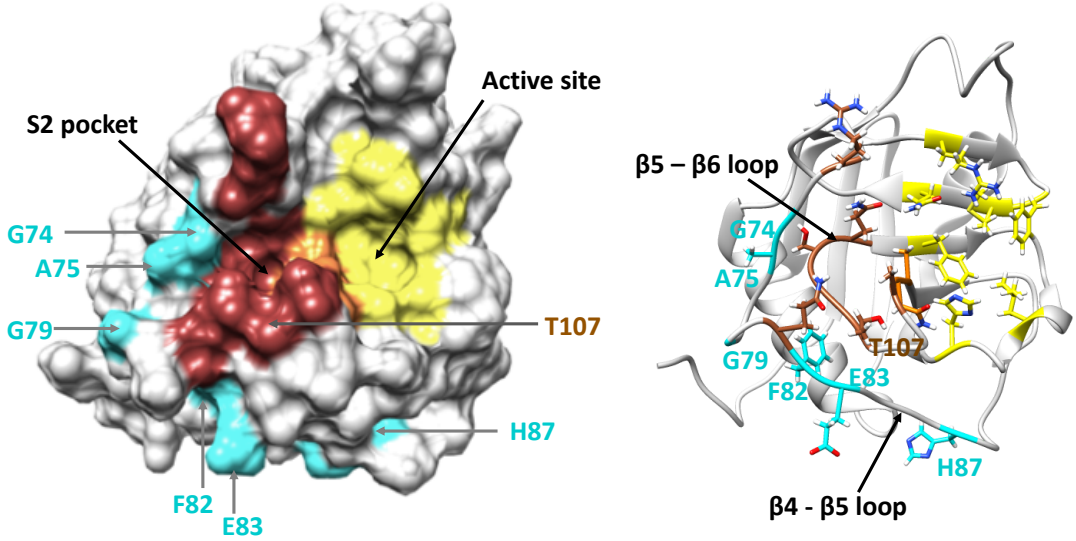


Figure 40 Surface and ribbon models of PPIL1 protein, showing residues of PPIL1 T107A which exhibit a nonlinear relationship between chemical shift and PRP17 concentration in HSQC spectra (cyan). Active site residues (yellow), S2 pocket residues (brown) and shared S2 pocket and active site residues (orange) are also highlighted. Residues with non-linear shift patterns in the HSQC are found in the β4-β5 loop. The β4- β5 loop is adjacent to the β5-β6 loop which forms the base of the S2 pocket.

6.2.1.2 ITC analysis

Isothermal titration calorimetry (ITC) was used to confirm binding between PPIL1 and PRP17 peptide. ITC is an accurate and sensitive method which can be used to measure molecular interactions in solution, without the need to label any interaction partner. This technique monitors binding by measuring the heat which is released or absorbed during an interaction. Titrations of binding partners are used to obtain binding curves, from which information such as binding affinity can be determined[150]. Weak affinity interactions are often difficult to characterise; however, Turnbull et al. showed that ITC produces reliable K_D values for such interactions given that sufficient ligand titrations are used and a fixed stoichiometry is applied[120]. (However, it is recommended not to derive further parameters such as $\Delta H°$ and $\Delta S°$, as these parameters can be less reliably determined using low affinity data[120].)

PRP17 peptide was titrated against PPIL1 protein to measure the interaction. Given the low affinity of the interaction, over 10 molar equivalents of ligand (PRP17) were titrated into the cell and K_D values were calculated with a fixed stoichiometry (n=1)[120]. Clear binding curves were observed between PPIL1 WT and PRP17 peptide with a calculated K_D of 111.9 µM (Figure 41). These results therefore confirm that PPIL1 interacts with PRP17 with weak affinity, on a similar order of magnitude to the binding detected by NMR (section 6.2.1.1). These results are therefore consistent with previous experimental observations. It was also investigated whether T107A and R131Q substitutions affect this interaction. ITC results also show that PPIL1 T107A and R131Q bind to PRP17 peptide with similar affinities to PPIL1 WT; 109.5 µM and 107.9 µM respectively, compared to 111.9 µM (Figure 41). This indicates that PPIL1 T107A and R131Q patient mutations do not significantly affect the overall binding affinity of the PPIL1 PRP17 interaction.

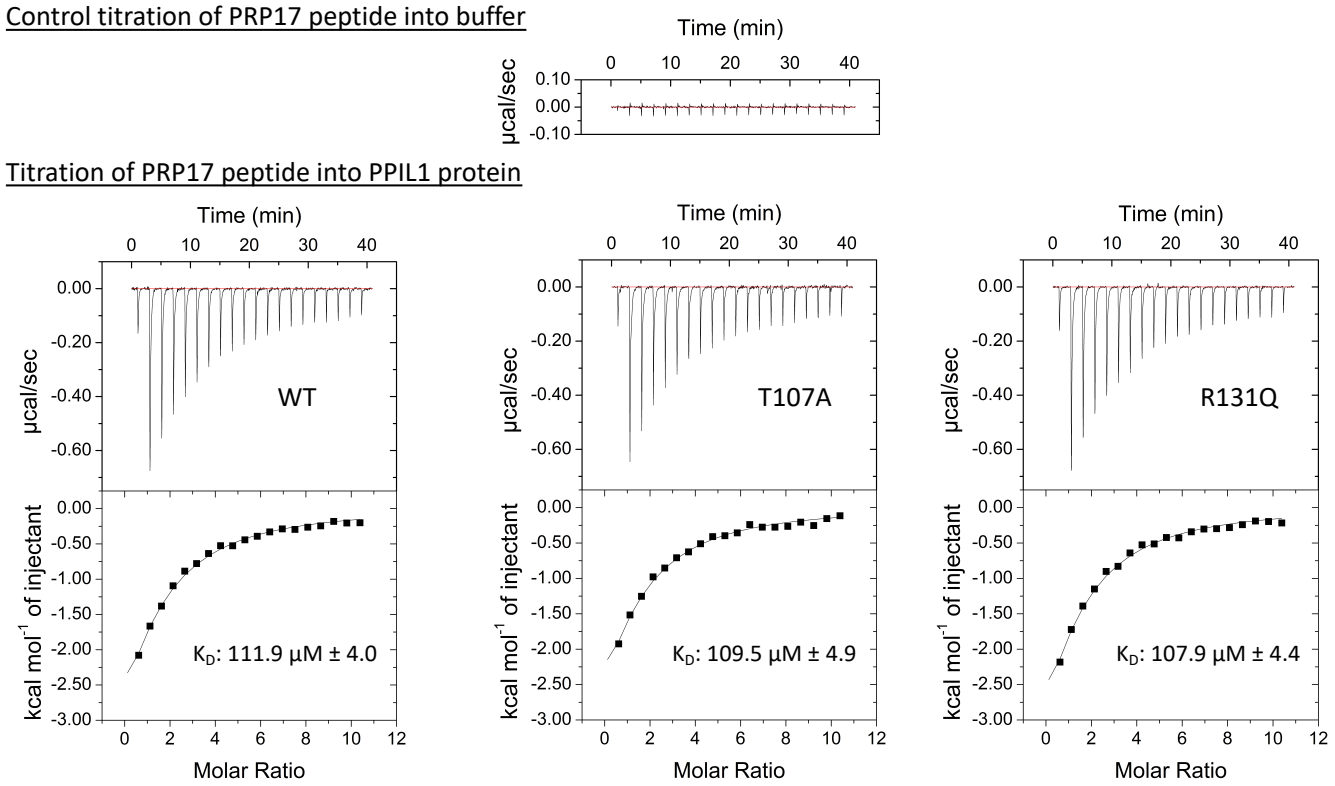


Figure 41 Isothermal titration calorimetry (ITC) analysis of the interaction between PPIL1 proteins (WT, T107A and R131Q) and PRP17 peptide (PRP17 residues 89-101). PRP17 peptide (2 mM) was titrated into PPIL1 WT (40 µM) at 25 °C in PBS (pH 7.4). A control titration of PRP17 peptide into buffer, was carried out to determine the heat of PRP17 peptide dilution (top). Sequential heat pulse data (µcal/sec) for the titration of PRP17 peptide into PPIL1 proteins (WT, T107A, R131Q) is shown (middle panels). The heat changes upon binding were integrated to give binding curves for the interaction of each PPIL1 protein (WT, T107A, R131Q) with PRP17 peptide (bottom panels). Curve fitting with a fixed stoichiometry of 1.0 was used to calculate the K_D of each interaction.

6.2.2 PRP17 1-172 and PPIL1 interaction

It is possible that the PPIL1:PRP17 interaction induces a conformational change in PRP17. To test this hypothesis ^{15}N labelled PRP17 1-172 was expressed and purified (this was a gift from Guoliang Chai, UCSD) and ^{1}H ^{15}N transverse relaxation optimized spectroscopy (TROSY) NMR was used to analyse PRP17 1-172. ^{1}H ^{15}N TROSY experiments are equivalent to ^{1}H ^{15}N HSQC (Figure 31) but show increased resolution and sensitivity for larger proteins[163]. The spectrum of the protein alone shows sharp well resolved resonances (Figure 42), suggesting the protein was suitable for further analysis. In addition, all resonances are found in a narrow ^{1}H spectral range, suggesting that this protein is predominantly disordered in solution. A few resonances are shifted slightly upfield in the ^{1}H dimension, suggesting that a small number of residues may have alpha helical tendencies (Figure 42).

PRP17 1-172 is mostly disordered in solution, it is possible that the interaction with PPIL1 stabilises a region of PRP17 1-172 and promotes a disorder to order transition. To investigate this further, unlabelled PPIL1 WT protein was titrated against ^{15}N labelled PRP17 1-172 protein and ^{1}H ^{15}N TROSY spectra were acquired (Figure 43). Upon titration of PPIL1 into PRP17, a small number of specific PRP17 resonances broaden significantly, in a concentration dependent manner. At molar ratios above 0.5:1 PPIL1:PRP17 these resonances broaden beyond detection (and do not reappear at another position in the spectrum). The observation that only specific resonances are affected by the PPIL1 titrations suggests that specific binding occurs between PRP17 1-172 and PPIL1; this is consistent with previous observations of the interactions between the PRP17 peptide and PPIL1 (sections 6.2.1.1 and 6.2.1.2). Resonance broadening is likely indicative of intermediate exchange, either between bound and unbound states of PRP17 or conformational exchange of PRP17 caused by binding. Resonance broadening here closely mirrors that of key active site PPIL1 residues during PRP17 peptide titration, discussed above (section 6.2.1.1; Figure 33; e.g. R55, W121). The small number of resonances affected suggests that the interaction involves a localised region of PRP17 1-172. The data suggests that there is no large disorder to order transition of PRP17 1-172 upon binding as there is absence of a significant number of resonances which shift in the spectra.

To confirm that the PPIL1 PRP17 1-172 interaction observed occurs via the PPIL1 active site, the PPIL1 active site competitive inhibitor, cyclosporin A (CsA) was used. ^{1}H ^{15}N TROSY experiments were repeated with PRP17 1-172 alone (30 μM), with addition of PPIL1 (15 μM) and upon subsequent addition of CsA to the same sample (30 μM; Figure 44). The spectra with PRP17 1-172 alone and with PPIL1 are identical to those in Figure 43 and show that specific resonances of PRP17 1-172 broaden beyond detection, upon addition of PPIL1 (Figure 43; Figure 44). Upon addition of CsA to the sample containing PRP17 1-172 and PPIL1, the resonances which broadened beyond detection upon addition of PPIL1, sharpen and can be resolved (Figure 44). These results show that CsA can block the PPIL1 PRP17 1-172 interaction and therefore confirm that the interaction occurs at the active site of PPIL1. In addition, these results further confirm that the observed resonance broadening is a result of specific binding rather than any other non-specific effect. The results of these experiments show that there is no large conformational change of PRP17 1-172 upon binding to PPIL1.

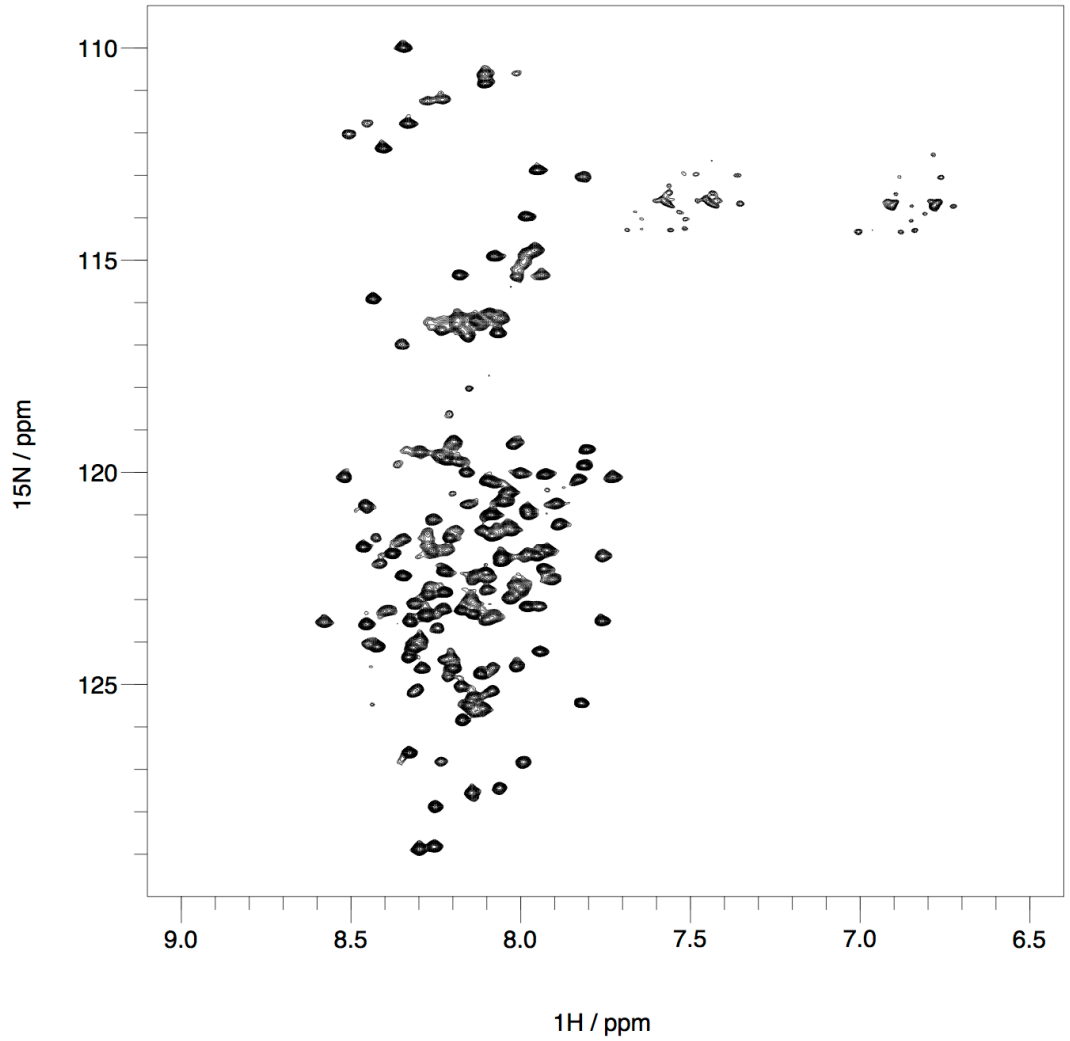


Figure 42 ^{1}H ^{15}N TROSY spectrum of PRP17 1-172 protein alone, in 25 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, 5% D_2O pH 7.0. Spectra were recorded using a 750 MHz NMR spectrometer at 25 ºC.

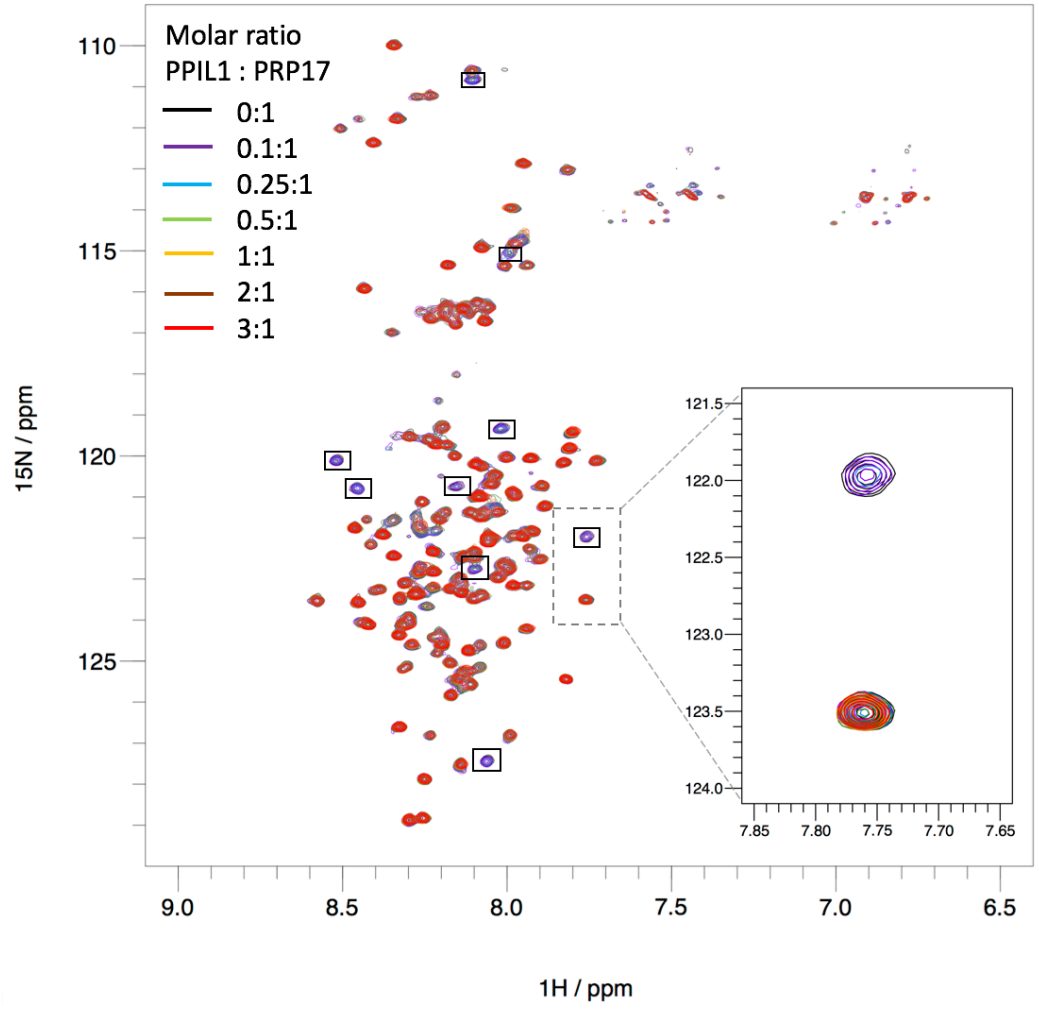


Figure 43 Overlay of PRP17 1-172 ^{1}H ^{15}N TROSY spectra before and after PPIL1 WT titrations. Spectra were recorded using 30 µM PRP17 and PPIL1 was titrated in from a 450 µM stock to give the following molar ratios. Resonances which broaden beyond detection are indicated by boxes. (There are currently no assignments for PRP17). The enlarged region of the spectra shows: an example resonance which broadens beyond the limit of detection upon titration of PPIL1 (top) and an example resonance which remains unaffected during the titration (bottom).

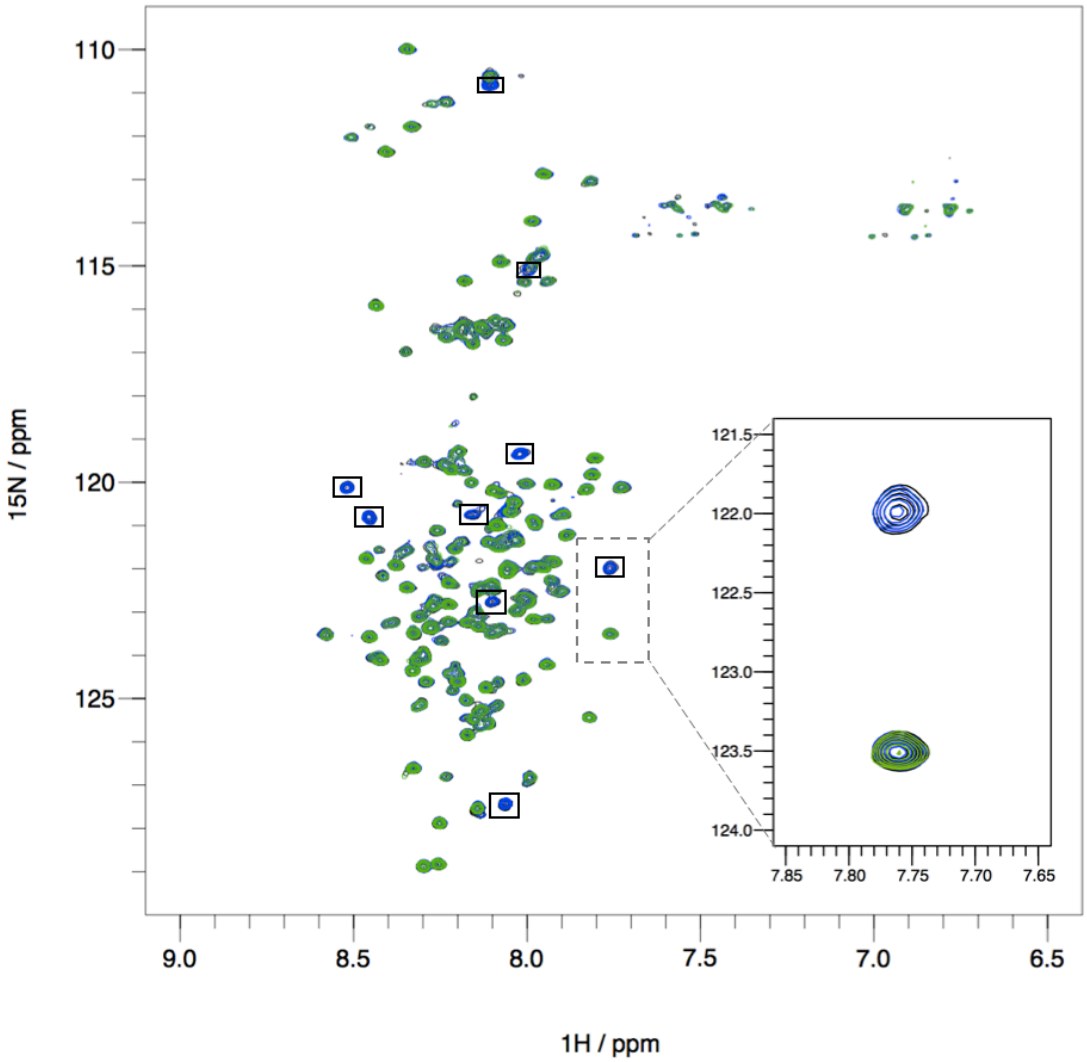


Figure 44 Overlay of PRP17 1-172 ^{1}H ^{15}N TROSY spectra: PRP17 1-172 alone (black), after addition of PPIL1 (green) and after addition of both PPIL1 and the PPIL1 active site inhibitor, CsA (blue). The resonances which broaden beyond detection upon addition of PPIL1, reappear upon addition of the PPIL1 active site inhibitor CsA (indicated with boxes; e.g. top resonance of inlay). All other resonances are unaffected by addition of CsA (e.g. bottom resonance of inlay). The concentrations of molecules used were: 30 µM PRP17, 15 µM PPIL1, 30 µM CsA.

6.2.3 PPIL1 PPIase activity on PRP17 (Gly94-Pro95)

6.2.3.1 NMR Gly94-Pro95 *cis* and *trans* isomer assignment

The data above suggests that binding between PRP17 and PPIL1 occurs at the PPIL1 PPIase active site. This raises the hypothesis that PRP17 is a substrate for the PPIase activity of PPIL1. Therefore, it was investigated whether addition of PPIL1 to PRP17 increases *cis/trans* proline isomerisation of the Gly94-Pro95 peptide bond; the prolyl peptide bond which occupies the PPIL1 active site in cryo-EM spliceosome structures (Figure 17; Figure 18). A peptide was synthesised (Peptide Synthetics) with an identical amino acid sequence to PRP17 peptide used for binding studies, but with the central proline (Pro95) uniformly ^{15}N and ^{13}C labelled.

^{13}C chemical shifts can be used to assign *cis* and *trans* peptide bond isomers[37]; therefore a ^{1}H ^{13}C HSQC of PRP17 peptide was acquired. ^{1}H ^{13}C HSQC spectra show all correlations between protons and ^{13}C labelled carbons in a sample which are one bond coupled (i.e. linked directly by a covalent bond; Figure 45). The ^{1}H ^{13}C HSQC spectra recorded for PRP17 peptide shows major resonances for each carbon of proline's pyrrolidine ring (α, β, γ, δ; Figure 46), as these are all linked to at least one proton. These signals show two ^{1}H shift values for β, γ, δ carbons as each carbon has covalent bonds with two protons; with the exception of C_{γ} *trans* resonances which are degenerate (i.e. the two resonances overlay and appear as a single resonance). There is a single major resonance for C_{α} as this atom has a covalent bond with a single ^{1}H. In addition to the major resonances, a set of adjacent minor resonances can also be observed, suggesting that significant populations of both *cis* and *trans* isomers of this peptide may exist under the conditions tested.

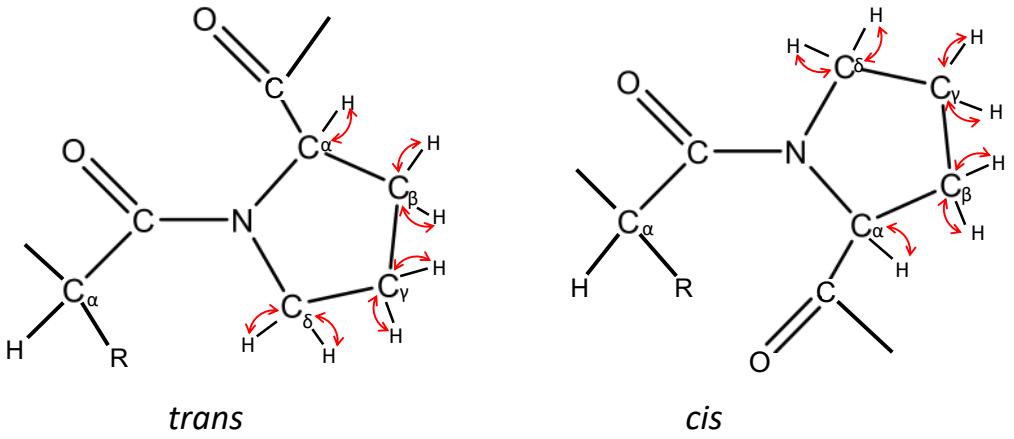


Figure 45 ^{1}H ^{13}C HSQC magnetisation pathway. ^{13}C labelled proline and i-1 residues are shown for a *trans* and *cis* peptide bond. Magnetisation is passed from ^{1}H to ^{13}C nuclei, where the chemical shift is evolved, and back to ^{1}H for detection (shown with red arrows). All ^{1}H-^{13}C correlations are detected in the spectrum. (Only proline carbon atoms are ^{13}C labelled).

The $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ chemical shift differences ($\Delta_{\beta\gamma}$) were used to determine whether the major and minor states correspond to *trans* and *cis* peptide bond isomers. Shen and Bax showed that $\Delta_{\beta\gamma}$ is higher when a proline peptide bond is in the *cis* conformation compared to the *trans* conformation and can be used to discriminate between *cis* and *trans* peptide bonds[37]. $\Delta_{\beta\gamma}$ is larger for the minor resonances (9.87 ppm) than major resonances (5.00 ppm) in the PRP17 peptide ^{1}H ^{13}C HSQC spectrum (Figure 46, right); consistent with minor resonances corresponding to the *cis* conformation and the major resonances corresponding to the *trans* conformation of the peptide bond. The major and minor resonances also fall within the expected $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ chemical shift ranges for *trans* and *cis* conformations (Figure 46), further confirming the assignment. Other possible causes of the minor resonances, such as low sample purity, were excluded as mass spectrometry and HPLC data shows that the peptide has a single molecular weight and is of >98 % purity (see appendix).

The ratios of *cis* and *trans* populations were determined by comparing the relative peak volumes of the *cis* and *trans* ^{1}H ^{13}C HSQC resonances. This analysis shows that the ratio of *cis* and *trans* populations are 12 % and 88 % respectively. This finding is consistent with data for unstructured peptides where the *cis* conformation is found at a lower population than the *trans* conformation[37]. The *cis* population is relatively high for an unstructured

peptide, which is likely a result of the preceding glycine residue, which is favourable for the *cis* conformation[37].

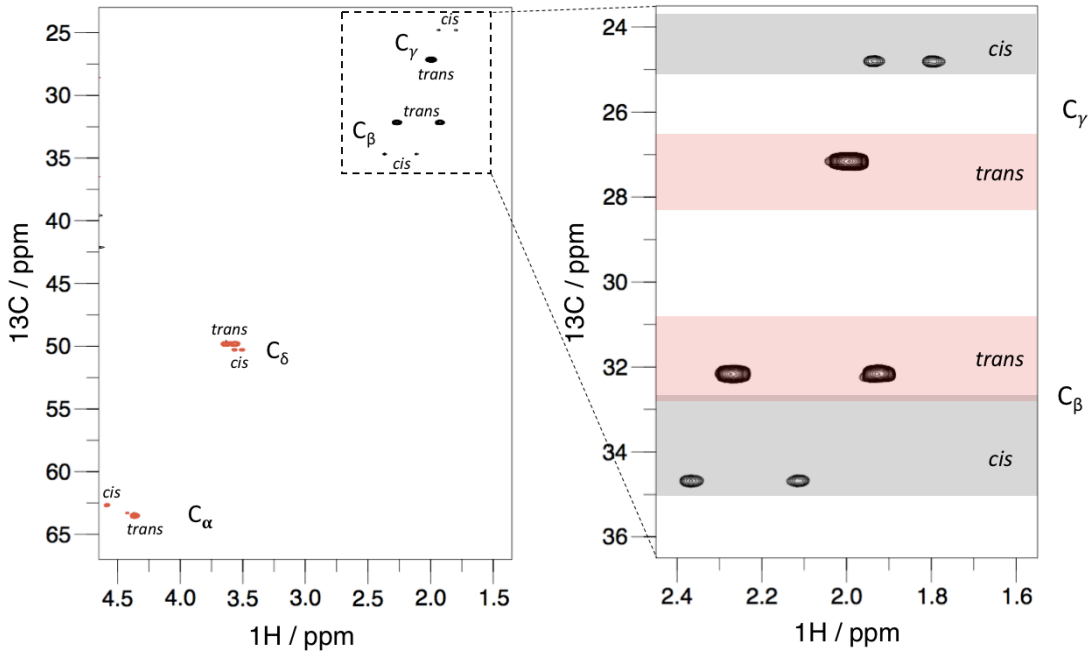


Figure 46 ^{1}H ^{13}C HSQC PRP17 peptide spectrum, showing resonances of the central proline residue (corresponding to Pro95 of PRP17). PRP17 peptide (corresponding to residues 89-101 of PRP17), was synthesised with a ^{13}C and ^{15}N labelled proline (corresponding to Pro95 of PRP17). The ^{1}H ^{13}C HSQC spectrum of PRP17 peptide was used to assign proline *cis* and *trans* peptide bond isomers based on $\Delta_{\beta\gamma}$ (the difference between $^{13}C_\beta$ and $^{13}C_\gamma$ shifts). Expected $^{13}C_\beta$ and $^{13}C_\gamma$ shifts for *trans* (pink) and *cis* (grey) isomers are shaded on the enlarged $^{13}C_\beta$ and $^{13}C_\gamma$ region of the spectrum (mean ± one standard deviation; Shen and Bax 2010[37]). Spectra were acquired at 25 °C, in PBS (pH 7.4), on a 750 MHz spectrometer.

6.2.3.2 ZZ exchange NMR experiments

To determine whether PPIL1 has PPIase activity on PRP17, NMR experiments were used to monitor exchange between the *cis* and *trans* states of the Gly94-Pro95 peptide bond. ZZ exchange experiments were chosen as these experiments can measure slow exchange (≈ 0.2 – 100 s^{-1}), which is the timescale proline isomerisation typically falls under[129]. During the ^{1}H ^{15}N, H(Cα)N ZZ exchange experiments, magnetisation is passed from the $^{1}H\alpha$ to ^{15}N amide via Cα. A mixing period follows where magnetisation is stored along the Z axis and during which exchange can occur (between proline peptide bond *cis* and *trans* conformations). The magnetisation is then transferred back to the Hα for detection, via Cα (Figure

47). Therefore, if there is significant interconversion between states (*cis* and *trans*) during the mixing time, exchange signals appear in the spectrum[128].

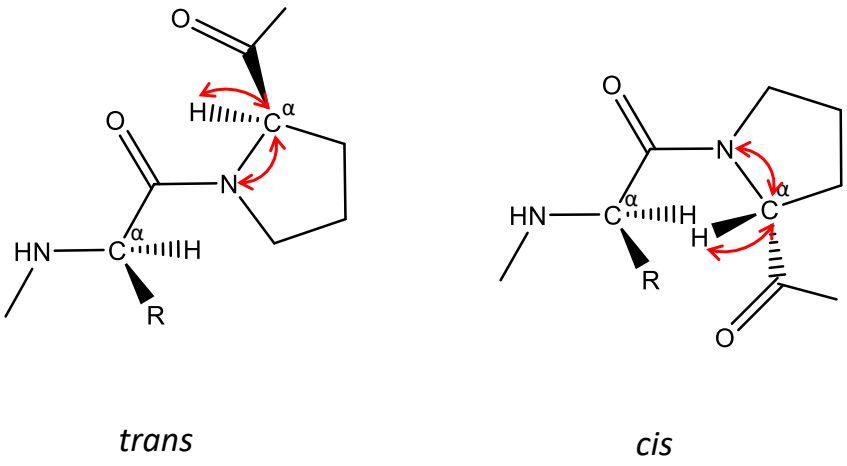


Figure 47 Magnetisation pathway of ^{1}H ^{15}N, H(Cα)N ZZ exchange experiments, showing proline and a preceding residue (*cis* and *trans* peptide bond conformations are shown). Magnetisation is transferred from $^{1}H\alpha$ to $^{13}C\alpha$ and then to ^{15}N amide, where it is labelled with ^{15}N chemical shift. The magnetisation is then stored as N_Z magnetisation (i.e. along the Z axis), for a defined period of time, during which exchange (e.g. proline isomerisation) may take place. Finally, magnetisation is passed back to $^{1}H\alpha$ for detection, via $^{13}C\alpha$.

^{1}H ^{15}N, H(Cα)N ZZ exchange spectra were first recorded in the absence of PPIL1. Distinct major and minor signals were observed for the labelled proline (corresponding to Pro95 of PRP17; Figure 48). The minor and major signals correspond to the *cis* and *trans* conformations of the Gly94-Pro95 peptide bond, as assigned in section 6.2.3.1. Levels of proline isomerisation were too low to detect in the absence of PPIL1, as no exchange crosspeaks are detected between the *cis* and *trans* conformations (Figure 48A). On addition of sub-stoichiometric concentrations of PPIL1, exchange cross peaks appear between the *cis* and *trans* resonances (Figure 48B). These results indicate that PPIL1 can significantly increase the rate of proline isomerisation on PRP17 (Gly94-Pro95 peptide bond) and therefore supports the conclusion that PRP17 is a substrate of PPIL1.

There is also a further minor signal adjacent to the *trans* signal (Figure 48). Since Pro95 is the only ^{15}N labelled residue the extra signal must also arise from this residue. (MS and HPLC analysis provided by the company confirm the peptide is of a single molecular weight and >98% purity; see Appendix

Figure 61, Figure 62) Since there are two further prolines in this peptide (89-FA**P**EFG**P**EN**P**FRT-101), it is likely that this reflects a minor cis conformation of another proline. *Cis* and *trans* peptide bond isomers also cause resonance doubling for neighbouring residues[166]. The extra resonance is only seen for Cα in the ^{1}H ^{13}C HSQC (Figure 46); suggesting that this resonance appears as a result of a minor conformation of the proline N-terminal to Pro95, Pro91.

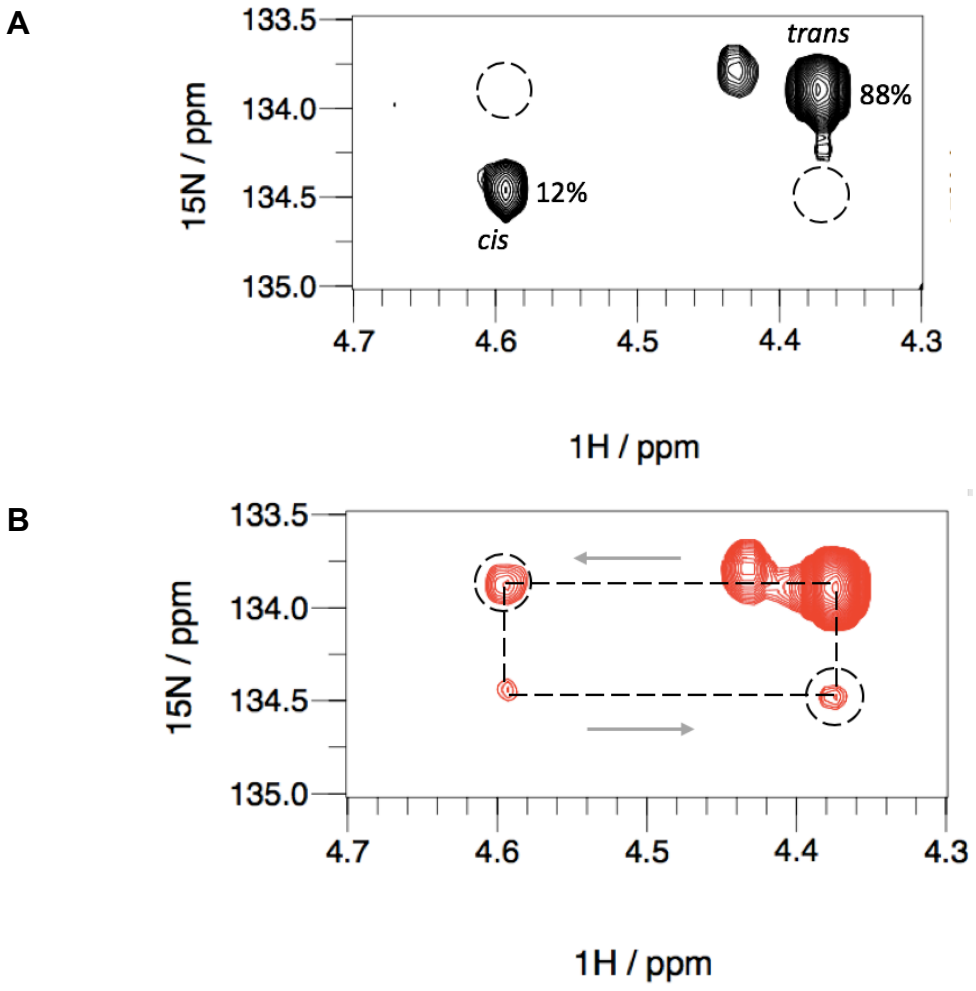


Figure 48 PPIL1 catalysed proline isomerisation of PRP17 peptide (Gly94-Pro95). PRP17 peptide was synthesised with ^{15}N (and ^{13}C) labelled proline (residues 89-101 and Pro95 of PRP17) to monitor proline isomerisation of PRP17 Gly94-Pro95 peptide bond. 2D ^{1}H ^{15}N, H(C**α**)N ZZ exchange NMR spectra were acquired of PRP17 peptide alone (A) and with sub-stoichiometric concentrations of PPIL1 (5 μM; B). Dashed circles represent the region of the spectra in which exchange peaks should appear if there is significant interconversion between *cis* and *trans* proline peptide bond isomers during the 150 ms delay. Spectra were acquired at 25 °C, in PBS (pH 7.4), on a 950 MHz spectrometer.

6.2.4 Further analysis of PPIL1 PPIase activity on PRP17

6.2.4.1 Assignment of PRP17 84 - 101 peptide

To measure rates of proline isomerisation in PRP17, alternative shorter ZZ exchange experiments were used, to allow acquisition of the many spectra required for such calculations (see section 6.2.4.2). This required use of a peptide with ^{15}N labelled amide groups, therefore a uniformly labelled PRP17 peptide was made (gift from Guoliang Chai, UCSD). This peptide corresponds to residues 84-101 of PRP17 (further residues of PRP17 sequence were included for expression and purification purposes) and all residues were uniformly ^{15}N labelled.

A ^{1}H ^{15}N HSQC spectrum was recorded of the peptide alone, which shows clear well resolved resonances (Figure 51). Resonances are found in a narrow ^{1}H spectral window, showing that the peptide is predominantly disordered under the conditions tested. As shown for PRP17 peptide with labelled Pro95 (section 6.2.3; Figure 46), the spectrum of uniformly labelled PRP17 peptide also contains major and minor resonances (Figure 51). In section 6.2.3 it was demonstrated that PRP17 peptide minor and major resonances correspond to *cis* and *trans* conformations respectively. The displayed assignment of major resonances (Figure 51; Table 17) were determined using ^{1}H ^{15}N HSQC, ^{1}H ^{15}N HSQC ROESY and ^{15}N HSQC-TOCSY spectra.

TOCSY (TOtal Correlated SpectroscopY) spectra were acquired to gain a ‘fingerprint’ for each amino acid, where the pattern and chemical shifts of resonances were used to help identify the type of amino acid residue. ROESY (Rotating frame nuclear Overhauser effect spectroscopy) spectra were also acquired to determine sequential connections between residues. Despite the small size of the peptide, the spectrum was complicated by the presence of many minor signals. Therefore, in order to assign the resonances ^{1}H ^{15}N HSQC-TOCSY and ^{1}H ^{15}N HSQC ROESY spectra were used which give better separations of signals, than standard ^{1}H-^{1}H TOCSY and ROESY spectra (Figure 50). ^{1}H ^{15}N HSQC-TOCSY results in a resonance for all ^{1}H-^{15}N direct couplings and corresponding ^{1}H chemical shifts for all protons linked by covalent bonds in that spin system (i.e. all protons in that residue); see Figure 49. ^{1}H ^{15}N HSQC-ROESY also shows resonances for all ^{1}H-^{15}N direct couplings but subsequent connections for all

protons linked in space (rather than through bonds). The precise correlations depend on factors such as the mixing time used and size of the molecule. The ^{1}H ^{15}N HSQC ROESY used here gave good correlations between $^{1}H\alpha$ of i and i-1 residues; which could be used for the assignment (Figure 50).

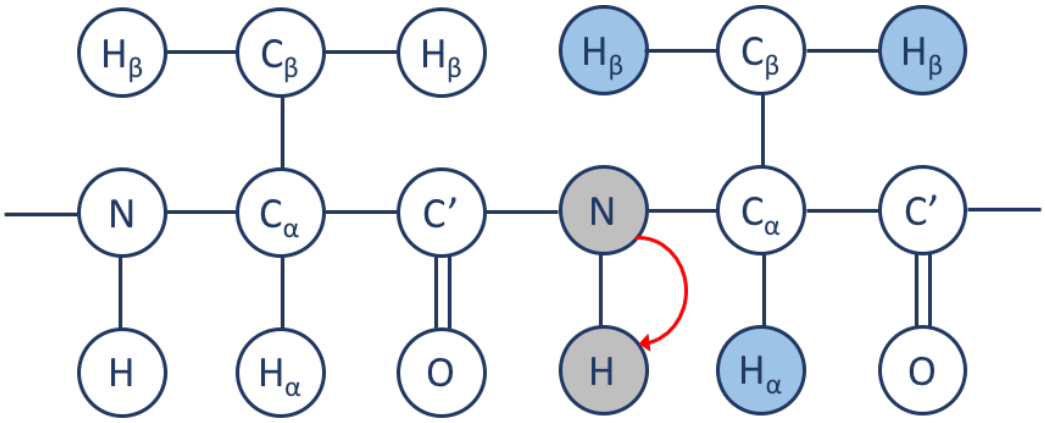


Figure 49 2D ^{1}H ^{15}N HSQC-TOCSY magnetisation pathway. An initial HSQC pulse sequence transfers magnetisation between ^{15}N and ^{1}H (shown in grey, with red arrow). The signal is then propagated via covalent bonds through the whole spin system and magnetisation is evolved on all ^{1}H nuclei (blue and grey). This results in correlations between ^{15}N amide and all protons in that residue.

Assignments were recorded for all major HSQC resonances and also all major proton resonances visible in the ^{1}H ^{15}N HSQC-TOCSY (Table 17). The N-terminal residue was absent in the spectrum, as is often the case, due to proton exchange with the solvent. Proline residues do not give a resonance in a ^{1}H ^{15}N HSQC, as they lack an amide proton; therefore, no assignment was determined for the three proline residues. The second residue of the peptide is also absent from the spectra, presumably due to high flexibility of the N-terminal region or accelerated amide proton exchange. The assignment was confirmed using sequential connections shown in the ^{1}H ^{15}N HSQC-ROESY spectra (Figure 50, Figure 52, Appendix Figure 63). The HSQC-ROESY spectrum showed near complete sequential HN-HN and HN-Hα crosspeaks for the *trans* state signals but could not detect the much weaker sequential crosspeaks from the *cis* state signals. HSQC-ROESY resonances allowed connections between all residues to be confirmed; except for interruptions by proline residues (which do not give any signals in these experiments due to lack of an amide proton). Figure 52 shows an

example sequential connection pathway for residues 3-7 (residues one and two were not visible and the sequential connection was interrupted at residue 8, as this residue is a proline residue).

The minor states could not be assigned using these methods, due to the low population of the minor states. This meant that any corresponding TOCSY or ROESY resonances were too low in intensity to detect. However, the data presented in section 6.2.3.1, indicates that the minor states correspond to the *cis* conformation and the major states correspond to the *trans* conformation.

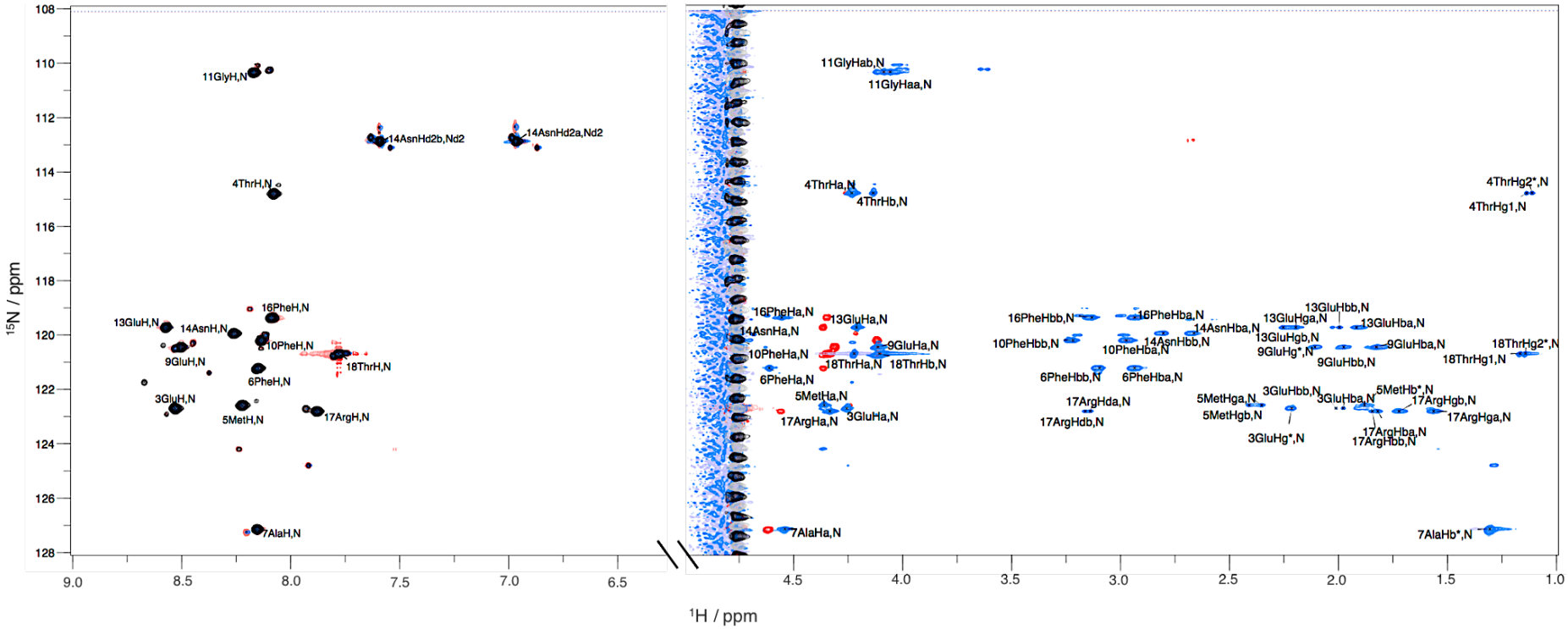


Figure 50 Overlay of ^{1}H ^{15}N HSQC (black), ^{1}H ^{15}N HSQC-TOCSY (blue) and ^{1}H ^{15}N HSQC-ROESY (red) spectra of PRP17 peptide (84-101) used for assignment. Assignments are shown for amide resonances (left), H_{α} resonances (centre) and sidechain resonances (right). 14Asn sidechain also contains ^{1}H-^{15}N connections, and the corresponding resonances are also found in the left region of the spectra. (The region of the spectra between 5.0 and 6.25 ppm (^{1}H) is omitted as no resonances were found here). Resonances are numbered according to their position in the peptide (i.e. residue 1, corresponds to residue 84 of full length PRP17).

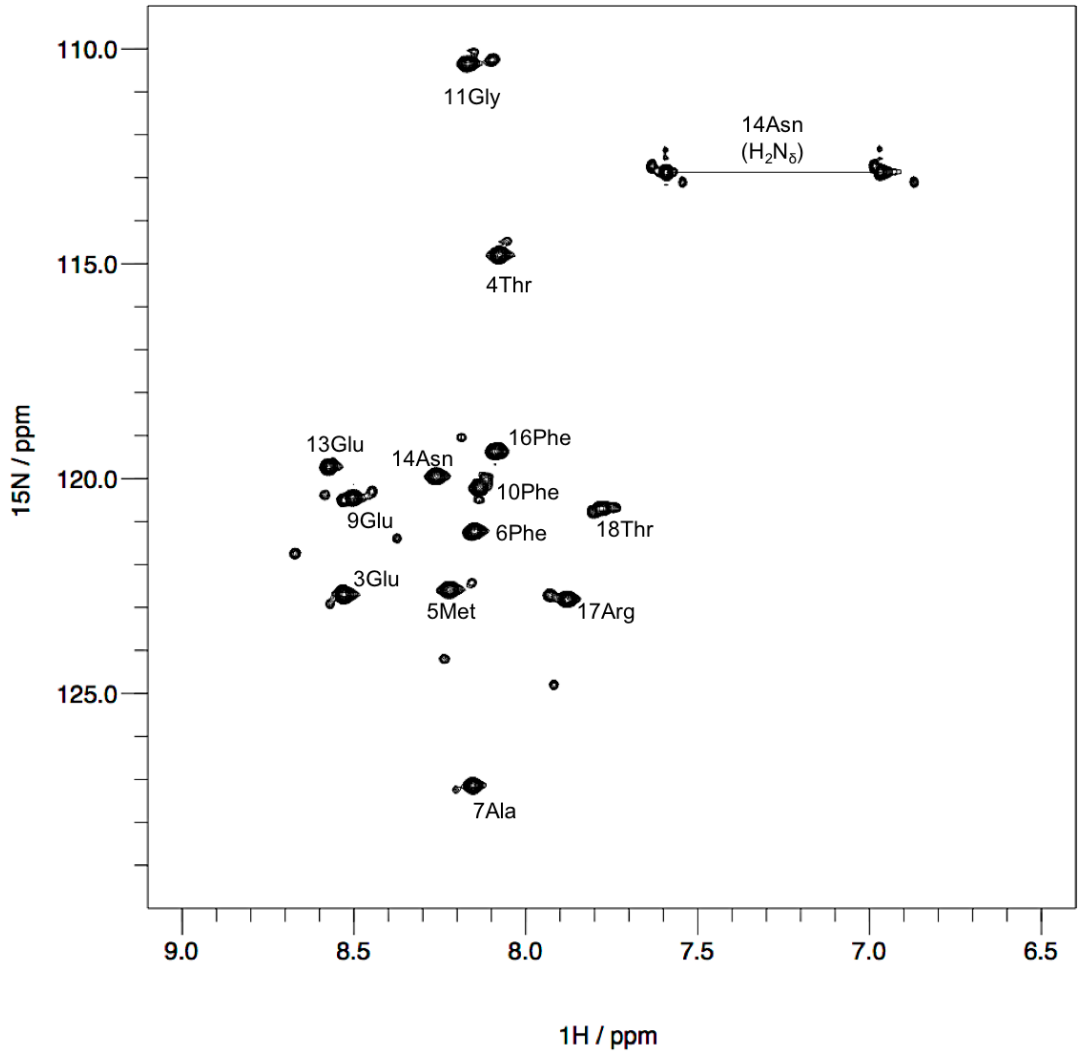


Figure 51 ^{1}H ^{15}N HSQC spectrum of PRP17 peptide (84-101), showing assignment of amide resonances and 14Asn sidechain (H_2N_δ) resonance. Resonances are numbered according to their position in the peptide (i.e. residue 1, corresponds to residue 84 of full length PRP17). Spectra were acquired at 25 ^{o}C, in 25 mM sodium phosphate buffer, 100 mM NaCl, 1mM DTT, 5% D_2O pH 7.0, using a 600 MHz spectrometer.

Residue no.	Amino Acid	N	H	H_α	H_β	H_γ	H_δ	Other
1	Thr	*						
2	Tyr	*						
3	Glu	122.70	8.52	4.26	2.01, 1.98	2.21, 2.21		
4	Thr	114.80	8.07	4.23	4.13	1.14, 1.11		
5	Met	122.59	8.22	4.36	1.88, 1.88	2.35, 2.41		
6	Phe	121.22	8.15	4.61	2.94, 3.10			
7	Ala	127.14	8.15	4.54	1.31			
8	Pro	*						
9	Glu	120.46	8.50	4.11	1.98, 1.83	2.11, 2.11		
10	Phe	120.21	8.13	4.74	3.22, 2.97			
11	Gly	110.36	8.17	4.09, 4.06				
12	Pro	*						
13	Glu	119.73	8.57	4.21	2.00, 1.92	2.25, 2.20		
14	Asn	119.95	8.26	4.68	2.80, 2.66		7.58, 6.96	N_δ 112.88
15	Pro	*						
16	Phe	119.37	8.09	4.56	3.13, 2.93			
17	Arg	122.82	7.87	4.33	1.84, 1.82	1.72, 1.56	3.16, 3.14	Hε 7.16, Nε 126.99
18	Thr	120.70	7.78	4.22	4.11	1.17, 1.14		

Table 17 ^{1}H and ^{15}N chemical shift assignment (ppm) of PRP17 peptide (residues 84-101). The following conditions were used: 400 µM peptide, 25 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, 5% D_2O, pH 7.0. (* Proline residues are absent from ^{1}H ^{15}N HSQC and ^{1}H ^{15}N HSQC-TOCSY spectra, due to lack of an amide proton. The two N terminal residues were also absent in these spectra.) Resonances are numbered according to their position in the peptide (i.e. residue 1, corresponds to residue 84 of full length PRP17).

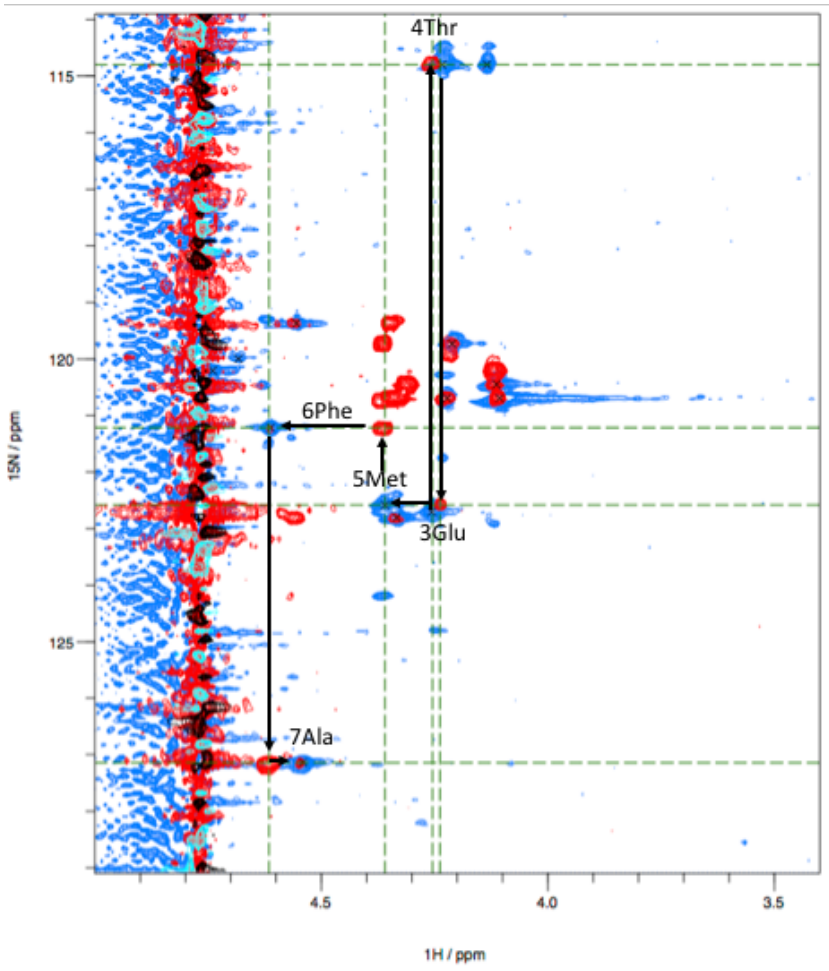


Figure 52 Example region of HSQC-TOCSY (blue) and HSQC-ROESY (red) spectra used for assignment of PRP17 peptide. The sequential connections of Cα protons for residues 3-7 of the peptide are shown.

6.2.4.2 ZZ exchange experiments

ZZ exchange experiments were again used to monitor proline isomerisation on PRP17 (84-101) peptide. However, it was now possible to monitor proline isomerisation on neighbouring amides, rather than directly on proline, due to the availability of a uniformly labelled peptide. Proline peptide bond isomers cause resonance doubling for surrounding residues; hence it is possible to monitor proline isomerisation on surrounding residues. ZZ exchange experiments are preferentially performed on residues with amide groups rather than proline residues, as it means that magnetisation can be passed directly from ^{1}H to ^{15}N rather than via $^{13}C\alpha$ (as proline lacks an amide proton). This means that the time taken to acquire a spectrum is significantly reduced and it was therefore possible to carry out the many experiments required for calculation of proline isomerisation rates.

^{1}H ^{15}N ZZ exchange spectra were acquired in the presence and absence of sub-stoichiometric concentrations of PPIL1 (4 μM). Upon addition of PPIL1, new exchange peaks appear in the spectrum (Figure 53); this is indicative of increased exchange between minor (*cis*) and major (*trans*) states. This therefore indicates that addition of PPIL1 causes increased proline isomerisation in PRP17, consistent with previous experiments (section 6.2.3). Residues 4Thr, 5Met, 7Ala, Glu9, Phe10, Gly11, Glu13, Asn14 and Arg17 all show major and minor resonances corresponding to *trans* and *cis* states. Proline isomerisation can clearly be seen for all residues with major and minor states. (Exchange peaks may be present for other residues, which cannot be observed due to overlapping chemical shifts with auto peaks).

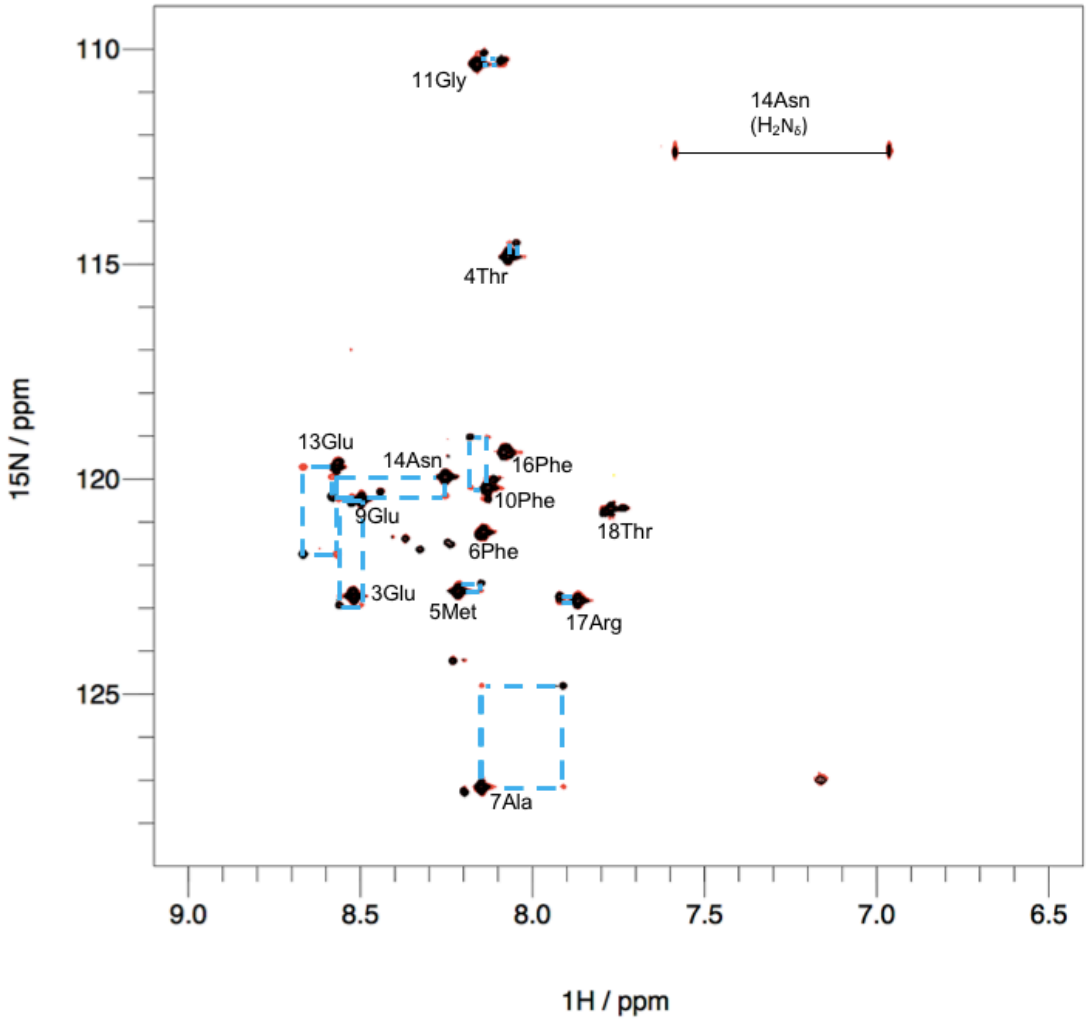


Figure 53 ^{1}H ^{15}N ZZ exchange spectra of PRP17 (84-101) with (red) and without (black) PPIL1. Exchange signals appear upon addition of PPIL1, due to exchange between *cis* and *trans* peptide bond conformations. Exchange peaks are connected to the corresponding auto peaks with dashed lines (blue). (4 µM PPIL1 was added to 400 µM PRP17 peptide).

To determine rates of exchange (i.e. of proline isomerisation), ^{1}H ^{15}N ZZ exchange experiments were repeated with different mixing times (Figure 54). This allowed auto and exchange peak intensities to be monitored as a function of mixing time (Figure 54; Figure 55). Example build up curves are shown for Glu13 (Figure 55), which has well separated and clearly resolved auto (*cis* and *trans*) and exchange signals (Figure 54). The auto (*cis* and *trans*) peaks decay biexponentially as a result of both relaxation and the exchange process; whereas the exchange (*cis*-*trans*, *trans*-*cis*) peaks have an initial build-up phase due to the exchange process but also show competing decay due to relaxation (Figure 54; Figure 55). Two state

exchange equations described by Kleckner and Foster[129] were fitted to the resulting build up curves, allowing the rate of exchange (k_{ex}) to be calculated (Figure 55; Table 18, Appendix Figure 64). Exchange rates were calculated for all residues with clear *cis* and *trans* resonances that were sufficiently separated and resolved for analysis (Table 18).

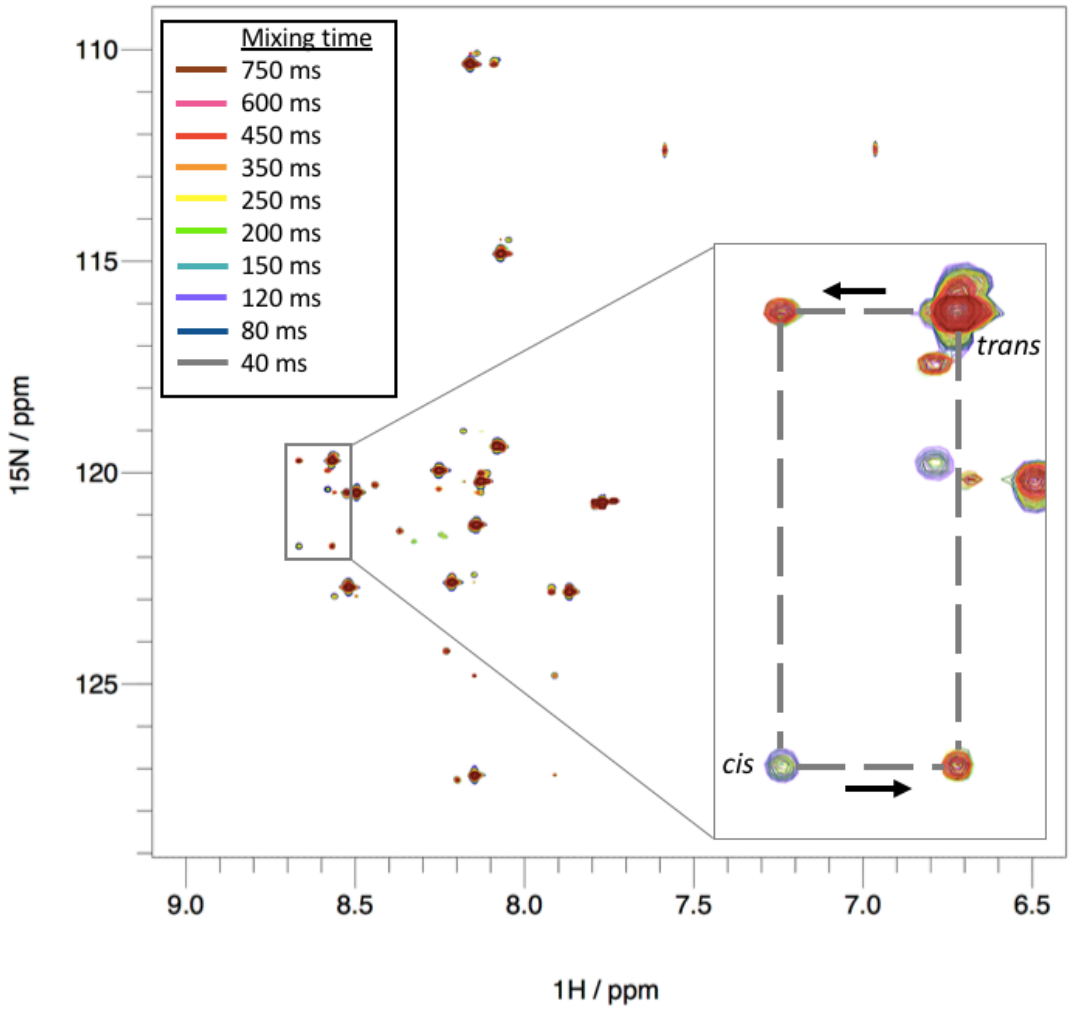


Figure 54 ^{1}H ^{15}N ZZ exchange spectra of PRP17 (84-101), recorded in the presence of PPIL1, using different mixing times. *Cis* and *trans* auto peak intensities reduce bi-exponentially with increased mixing time due to R_1 relaxation and the exchange process. Exchange peaks initially increase in intensity due to increased exchange events during the longer mixing time; but subsequently decrease in intensity due to R_1 relaxation. An example region of the spectra is enlarged to demonstrate this (the resonances correspond to Glu13 auto resonances [*cis* and *trans*] and exchange resonances [*cis*->*trans*, *trans*->*cis*]).

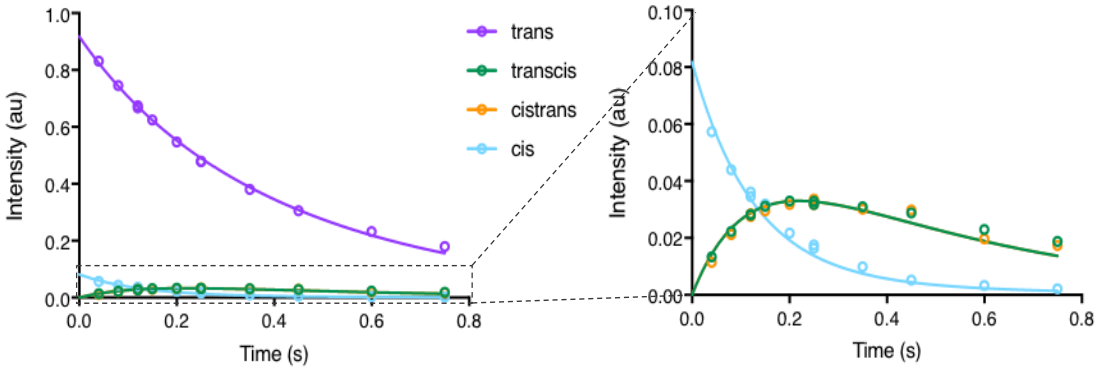


Figure 55 Proline isomerisation build up curves of PRP17 peptide, shown for Glu13. The normalised intensities of auto (*cis* and *trans*) and exchange peaks (*trans-cis*, *cis-trans*) from ^{1}H ^{15}N ZZ exchange spectra were plotted as a function of mixing time (seconds) and fit to two state exchange equations (see Chapter 2: materials and methods). Two experiments were collected in duplicate (experiments with 150 ms and 250 ms mixing times).

Residue	p_{cis} (%)	k_{ex} (s^{-1})
4Thr	6.8 ± 0.3	2.5 ± 0.4
7Ala	6.2 ± 0.2	2.3 ± 0.3
Glu9	5.5 ± 0.3	2.9 ± 0.4
Phe10	7.1 ± 0.3	2.7 ± 0.4
Gly11	11.1 ± 0.2	5.5 ± 0.3
Glu13	8.2 ± 0.2	5.8 ± 0.4
Asn14	9.8 ± 0.3	5.9 ± 0.6
Arg17	12.1 ± 0.3	5.3 ± 0.3

Table 18 The *cis* population (p_{cis}) and proline isomerisation rates (k_{ex}) of PRP17 peptide, measured across different PRP17 residues. (The error shown represents standard error for p_{cis} and k_{ex} values.)

The calculated rates of PPIL1 WT catalysed proline isomerisation on PRP17 fall between 2-6 s^{-1} (Table 18). Notably, proline isomerisation monitored on residues 4-10 shows a lower rate of exchange compared to residues 11-17. Residues 4-10 also show a lower *cis* population compared to residues 11-17 (Table 18). PRP17 peptide contains three proline residues and therefore the differences in exchange rates and *cis* populations are likely due to proline isomerisation of more than one proline in PRP17 (see discussion). The presence of more than one minor (*cis*) state is also shown in the ^{1}H ^{15}N HSQC spectrum of PRP17 peptide, where Gly11 has two minor resonances (Figure 51). This was also noted on PRP17 peptide with labelled Pro (section 6.2.3), where two minor states were observed (Figure 48). In this thesis it has already been shown that PRP17 Pro 95, which corresponds to Pro12 of this peptide is isomerised by PPIL1 (section 6.2.3). Due to the position in the sequence, it is likely that Pro8 and Pro12 contribute towards the observed resonance doublings and are isomerised by PPIL1; however further experiments would be required to confirm this (see discussion).

6.3 Discussion

Overall this work strongly supports the hypothesis that PPIL1 binds to PRP17, as suggested by the Cryo-EM spliceosome data (C* human spliceosomes[47] and *S. pombe* ILS spliceosome[48]). While working on this project further spliceosome structures (human B^{act}, C, P, ILS spliceosome complexes) have been published, which also suggests this interaction[13,35,36,49,50,77]. The interaction has been confirmed using both a peptide, corresponding to the proposed interaction region (PRP17 89-101), and also the N-terminal region of PRP17 (residues 1-172). These results suggest binding occurs at the active site of PPIL1 and is localized to a small region in PRP17, as also shown in the published Cryo-EM data[47,48,77]. Binding studies with the peptide confirm that binding occurs on the fast-intermediate timescale with weak affinity; as shown by the dissociation constants of 231.4 and 111.9 μM (using NMR and ITC respectively). Although dissociation constants could not be determined using the longer region of PRP17 (1-172), due to the high concentrations of protein required for such a low affinity interaction, the pattern of resonance shifts in this NMR data is consistent with binding on an intermediate time-scale (i.e. k_{ex} [s^{-1}] ≈ Δδ).

The K_D values for the PRP17: PPIL1 interaction, calculated using ITC and NMR, are of similar orders of magnitude (111.9 μM and 231.4 μM respectively). The difference between calculated K_D values for NMR and ITC experiments, is relatively small considering the different experimental techniques and conditions used. ITC results may be more accurate as a larger number of titrations were used meaning the complex was closer to saturation at the end of the titration. Overall the affinity of the PPIL1-PRP17 interaction, calculated using both methods, is in close agreement.

Comparison of binding between WT and mutant PPIL1 proteins with PRP17, yielded some unusual results for PPIL1 T107A. PPIL1 T107A shows clear evidence of binding to PRP17, also via the PPIL1 active site. However, the unusual chemical shift perturbation patterns caused by binding to PRP17, in the HSQC spectra, suggests something more complicated may be happening upon binding. Comparison of PPIL1 WT and mutant protein binding to PRP17, in ITC experiments, confirms that the PPIL1 T107A substitution does not significantly affect the overall binding affinity, compared to PPIL1 WT. It is unclear at present, what the cause of these differences may be and it is not possible to deduce the cause from this data alone. Possible causes of non-linear chemical shifts in HSQC spectra include: conformational change of PRP17 or PPIL1 upon binding or deviations from 1:1 binding stoichiometry. The residues with non-linear shifts, are found in the β4-β5 loop that lies beneath the PPIL1 S2 pocket, suggesting it is this region of PPIL1 which may be involved in either conformational change or deviations from 1:1 binding stoichiometry.
It is most likely that, in addition to bound and unbound states of PPIL1, there is a third state. The third state may correspond to a weaker secondary binding event or peptide binding induced dimerisation of PPIL1 and this could be investigated using analytical ultracentrifugation. To investigate conformational change of peptide or protein upon binding, further NMR experiments could be used. Specifically, Carr-Purcell-Meiboom-Gill sequence (CPMG) spin echo or Chemical Exchange Saturation Transfer (CEST) NMR experiments could be used to detect conformational changes upon binding, if they occur on timescales of milliseconds or less than a second respectively. To detect concentration dependent exchange, NMR Residual Dipolar Couplings (RDCs) could be used. In addition, transferred-NOEs or NOEs are NMR methods which could also be used determine conformation of the PPIL1:PRP17 complexes.

Many cyclophilins participate in the spliceosome, and until now the role of the cyclophilin active sites has been unclear. It has been proposed that cyclophilin active sites may be required to bind proline rich regions of the spliceosome, potentially acting as a chaperone or alternatively that PPIase activity itself is important[43,46]. Evidence here supports the latter: PPIL1 binds to PRP17 with weak affinity rather than engaging in a stable interaction with PRP17 which could stabilize a conformation of PRP17. Instead using NMR ZZ exchange experiments, here it has been shown for the first time that PPIL1 has PPIase activity on PRP17 and is therefore the likely substrate of PPIL1 in the spliceosome. In this chapter, it has been confirmed that PPIL1 has PPIase activity on the Gly94-Pro95 peptide bond, which is the peptide bond occupying the PPIL1 active site in the cryo-EM structures[47,48]. The work presented here is the first evidence that a cyclophilin has PPIase activity on a spliceosomal substrate.

The role of proline isomerisation in PRP17, in the context of the spliceosome is currently unclear. The spliceosome undergoes a large number of conformation changes during assembly, catalysis and disassembly and it has been suggested that proline isomerisation may contribute towards these conformational changes[30]. Work in this thesis showed that PRP17 is predominantly disordered, a property which means that PRP17 does not have a fixed structure and instead rapidly samples a wide range of conformations. It was therefore investigated whether interaction with PPIL1 may stabilise PRP17 and promote a particular conformation of PRP17. However, binding appears to be localised to a small number of residues and does not induce a large disorder to order transition of PRP17. It is possible the consequence of proline isomerisation cannot be observed outside of the spliceosome. Alternatively, it is possible that the small localised changes in PRP17 are sufficient to contribute to spliceosome assembly or catalysis.

Work presented here provides a methodology for monitoring proline isomerisation rates of PRP17. Proline isomerisation rates were calculated using PRP17 peptide (PRP17 84-101). The results presented strongly suggest that more than one proline in PRP17 undergoes PPIL1 catalysed proline isomerisation in PRP17 peptide. It will be necessary to confirm which prolines in PRP17 are isomerised, by substituting these residues with another residue such as alanine in PRP17 peptide. There is clear evidence that Pro95 is isomerised; however further work is required to confirm whether further proline residues (pro91 or pro98) also undergo proline isomerisation. Future experiments should be carried out to compare rates of PPIL1

catalysed proline isomerisation on PRP17 with WT vs mutant PPIL1 proteins, in order to determine whether patient PPIL1 variants affect this activity. This will offer further insights into the mechanism of the novel form of microcephaly and pontocerebellar hypoplasia, caused by mutations in PPIL1.

Despite potentially interacting with more than one proline in this loop region, PPIL1 clearly shows some binding specificity to particular proline peptide bonds in PRP17 1-172. Binding between PPIL1 and PRP17 is localised to a small number of residues in PRP17, despite PRP17 1-172 containing thirteen proline residues (which are all likely solvent accessible due to the disordered nature of PRP17 1-172). Piotukh et al. shows that for peptides binding to another cyclophilin's (CypA) active site, phenylalanine is favourable at position i-2 relative to proline (i), as it fits well into the S2 pocket[161]. PPIL1 is highly conserved compared to CypA and therefore PPIL1 may also exhibit a preference for phenylalanine at position i-2. Only Pro95 and Pro91 have a phenylalanine residue at position i-2, and therefore this may be the reason for PPIL1's specificity.

The role of other cyclophilin active sites in the spliceosome is still unclear. As this work was being completed, a study was published which showed the interaction of PPIH (spliceosome cyclophilin) and PRPF4 (spliceosome protein) involves two binding sites of PPIH, one of which involves PPIH PPIase active site[167]. Further cryo-EM spliceosome structures have since been published which also show interactions of spliceosomal proteins and cyclophilin active sites[77]. However some spliceosomal cyclophilins do not have PPIase activity on a model substrate[39], and it is therefore unlikely that all cyclophilins have PPIase activity in the spliceosome. This raises the intriguing question of the function of other cyclophilins which do not catalyse proline isomerization in the spliceosome.

Chapter 7 Conclusions and future perspectives

In this thesis it has been shown that the PPIL1 variants, identified in patients with microcephaly and pontocerebellar hypoplasia, clearly have deleterious effects on PPIL1 protein function (Chapters 3-5). All of the identified patient variants result in amino acid substitutions or duplications (A99T, T107A, R131Q and G109C;101_106dup). Following the expression and purification of PPIL1 WT and mutant proteins, it has been shown in an *in-vitro* system that PPIL1 A99T and G109C; 101_106dup protein changes significantly perturb PPIL1 protein structure and render these proteins insoluble (Chapter 4). T107A and R131Q amino acid substitutions also significantly affect PPIL1 protein folding and reduce protein stability, although to a lesser extent than A99T and G109C; 101_106dup (Chapter 4). PPIL1 R131Q abolishes binding to the PPIL1 spliceosome binding partner, SKIP (Chapter 5). It has also been shown that the PPIL1 T107A patient variant significantly reduces the PPIase enzyme activity of PPIL1 on a model substrate (Chapter 5). These findings strongly support the hypothesis that the novel syndrome of microcephaly and pontocerebellar hypoplasia is caused by the identified variants in PPIL1. It is interesting that the different patient variants affect different aspects of PPIL1 protein function: protein folding, interaction with SKIP, or PPIase activity of PPIL1. The severity of the phenotype varies between patients and interestingly patients with the T107A variant have the most severe phenotype (see Chapter 1: Introduction).

The discussed findings have important implications for the understanding of microcephaly and related brain development disorders. Together with findings by collaborators (Guoliang Chai and Joseph Gleeson, UCSD), there is strong evidence that variants in PPIL1 cause a novel syndrome of microcephaly and pontocerebellar hypoplasia (publication in progress “Biallelic mutations in the proline isomerase PPIL1 disrupt pre-mRNA splicing and cause microcephaly and pontocerebellar hypoplasia”). As

this thesis was being completed, further variants in PPIL1 were detected in patients with similar phenotypes to those reported here (data not yet published). These variants were: homozygous p.Y78C, compound heterozygous p.F82S; p.R131Q and compound heterozygous c.397A>G p.T127A; c.280+1G>A variants. These variants were identified independently in other laboratories, thus providing further independent evidence that variants in PPIL1 cause microcephaly and pontocerebellar hypoplasia.

This is the first time that variants in PPIL1 have been identified as the cause of a brain development disorder. Identifying the cause of genetic diseases has important implications for genetic counselling and therapy. Despite advances in identifying the underlying variant, a gene responsible can only be detected in approximately half of microcephaly cases[6]; therefore further work into the cause of microcephaly and associated disorders continues to be hugely important.

The experimental techniques used to investigate PPIL1 variant pathogenicity required use of purified protein and were therefore performed in an *in vitro* setting, which is a limitation of this study. However, the conclusions drawn from SPR experiments, to study the PPIL1-SKIP interaction, are in agreement with conclusions drawn from immunoprecipitation (IP) experiments (conducted by Guoliang Chai, UCSD) which are conducted in a more physiologically relevant situation. SPR results, however, offer advantages over IP experiments, as the direct interaction can be monitored and other indirect effects of the patient variants on protein interactions can be ruled out. It is currently not possible to study PPIase enzymatic activity in a cellular context (see below). The protein folding techniques used, also required the use of purified protein.

In order to gain a further understanding of the mechanism of the disease, the function of PPIL1 in the spliceosome was investigated. Previously it was only known that PPIL1 participates in the spliceosome, but little was known about its specific function in this complex. By modelling cryo-EM data PRP17 was identified as a putative substrate of PPIL1 in the

spliceosome (Chapter 3). It was subsequently shown, using NMR and ITC, that as suggested by the cryo-EM data[47-50,77], PRP17 binds to the PPIase active site of PPIL1. Further NMR experiments on PRP17 peptide, have demonstrated that PPIL1 exhibits PPIase activity on PRP17; providing strong evidence to support the hypothesis that PRP17 is the substrate of PPIL1 in the spliceosome (Chapter 5). This is the first time that a spliceosome cyclophilin has been shown to exhibit PPIase activity on a spliceosome substrate. It was previously unclear what the role of cyclophilin proteins, such as PPIL1 in the spliceosome might be[39,43]. These findings also have implications for the understanding of spliceosome biology in general. Aberrations in splicing cause a range of genetic diseases and cancers[96]; therefore knowledge of the spliceosome is important for the understanding many diseases.

Studies of proline isomerisation are currently limited to *in vitro* assays with purified protein, as NMR (and chromogenic peptide assays for model substrates[152]) are the only techniques capable of detecting proline isomerisation. A new *in vivo* method to detect proline isomerisation, using live cells and substrates labelled with a fluorescent probe is in development[168]. However, this method still requires further optimisation, such as increasing signal to noise, to be a reliable indicator of PPIase activity. In addition, this assay is only suitable for model substrates. The recent and ongoing development of “in cell NMR” means that there are now many examples of protein structural information derived from a cellular context. However this field is still in its infancy and has not been used to monitor proline isomerisation[169].

Further work will be required to understand the effect of the PPIL1 patient variants. It would be valuable to explore the effects of the newly identified PPIL1 variants on PPIL1 function, by expressing and purifying these proteins and repeating the assays used in this thesis (e.g. protein folding and stability, interaction with SKIP and PPIase activity assays). Furthermore, having established a protocol to determine rates of PPIL1 catalysed proline isomerisation of PRP17, it would be possible to compare the effect of the PPIL1 variants on this activity.

It still remains unclear how a mutation in a ubiquitous essential complex, the spliceosome, can cause symptoms which are limited to the brain. It was not possible to obtain patient cells to study splicing as the patients in the UK are deceased and cells obtained prior to this point did not survive. However collaborators (Joseph Gleeson and Guoliang Chai, UCSD) produced a mouse model (Ppil1$^{A99T/A99T}$) and performed RNA-seq analysis on E14.5 brains. Analysis using rMATS detected 3,797 significant differential splicing events, between mutant and WT brains, out of found 236,870 total splicing events. Significant differential splicing events affected all categories of splicing events (skipped exon, retained intron, alternative 5'SS, alternative 3'SS and mutually exclusive exons) and caused either increased inclusion or exclusion for different splicing events. GO-analysis showed that genes affected by differential splicing events in mutant brains are enriched for critical cellular functions including: RNA metabolism, protein translation/processing and cell cycle regulation.

The data above supports the conclusion that PPIL1 mutations affects the splicing of specific groups of genes, rather than a general inhibition of splicing (manuscript in progress as above). It may be that the additive effect of these changes accumulates over time. This model is further supported by previous work that shows that knockdown or mutation of spliceosome components, frequently affects the splicing of specific genes rather than a general inhibition of splicing[84,170,171]. High levels of alternative splicing takes place during brain development and AS is required at each stage of development, such as cell-fate determination, neuronal cell migration, axon growth and guidance and formation of synapses[172].

Many further genetic diseases, affecting a large range of tissues, are caused by mutations in spliceosome components. For example, mutations in several spliceosome components lead to autosomal dominant retinitis pigmentosa (RP), a genetic disorder which affects the retina. All spliceosome genes identified in RP (PRPF3, PRPF6, PRPF8, PRPF31, RP9 (PAP1), and SNRNP200) are components of the tri-

snRNP complex of the spliceosome. The tissue specificity has been attributed to the high demand of splicing in the retina; supported by the high expression of spliceosome snRNAs compared to other tissues[173]. However, mutations in genes encoding spliceosome components affect tissues beyond the retina and a range of craniofacial disorders (Nager syndrome, cerebrocostomandibular syndrome, Richieri-Costa–Pereira syndrome, Mandibulofacial dysostosis and Burn-McKeown syndrome) are also caused by spliceosome mutations (*SF3B4*, *SNRPB*, *EIF4A3, EFTUD2* and *TXNL4A* respectively)[174]. Spinal muscular atrophy is a neuromuscular disorder and is also caused by variants in *SMN1* which encodes a protein involved in assembly of spliceosome snRNPs[175]. Therefore, it is clear that tissue specificity in genetic disorders caused by spliceosome mutations is more complex, than originally thought.

Due to the complex combinatorial regulation of splicing by spliceosome components and regulatory factors[30,84], it is difficult to predict precisely how PPIL1 variants affect overall spliceosome conformation and function. Further spliceosome structures will be highly valuable to understand the mechanism of the spliceosome and function of individual components. This includes analysis of further states in the assembly and catalytic cycle of the spliceosome. In addition, analysis of spliceosome structures on alternatively spliced pre-mRNAs will provide valuable insights into regulation of the spliceosome and the mechanisms behind alternative splicing.

It has previously been shown that mutations in spliceosome components may affect stoichiometry of the spliceosome, which in turn leads to specific alternative splicing defects[173]. Tanackovic et al. showed that patient variants in PRPF31, PRPF3 or PRPF8 (which cause retinitis pigmentosa) caused alternations in composition of spliceosome subunits, using Northern and Western blotting[173]. Equivalent experiments in cell lines containing the patient PPIL1 variants would be valuable to further understand how PPIL1 variants lead to aberrant splicing.

It would also be valuable to investigate whether other cyclophilins have PPIase activity on the proteins which bind their active site in spliceosome structures. For example, Zhan et. al recently showed that CypE, and two further cyclophilin domains (tentatively assigned as PPWD1 and PPIG) have putative substrates in the C spliceosome (SmB, CCDC49 and Prp8)[77]. Further spliceosome structures may allow identification of further putative substrates of spliceosome cyclophilins. Two spliceosome cyclophilins are inactive on a model substrate (suc-AGPF-pNA), PPIL2 and CWC27[39]; therefore it would be interesting to understand the role of their PPIase active sites. These cyclophilins may bind to and stabilise proline rich regions of spliceosome components or it is possible that they have PPIase activity with high specificity; i.e. these cyclophilins may only be active on specific substrates and not the model substrate (suc-AGPF-pNA). This research will not only provide insights into the mechanism of the spliceosome, but also cyclophilin PPIase proteins. There are 19 cyclophilins in the human genome, many of which are implicated in human diseases, (e.g. PPIA[59] and PPID[176]).

Overall, this thesis provides strong evidence that variants in PPIL1 cause a syndrome of microcephaly and pontocerebellar hypoplasia and has provided novel insights into the role of PPIL1 in the spliceosome. Further research will be vital, not only for further understanding of this disease, but also to contribute towards understanding of the spliceosome and cyclophilin family of proteins (both of which are implicated in a wide range of human diseases).

Appendix

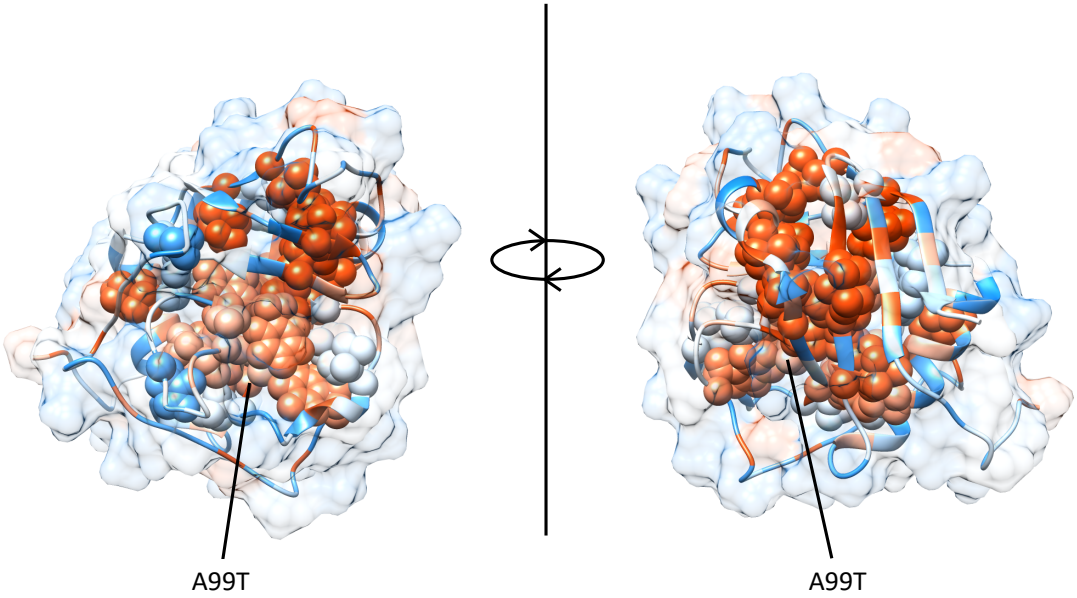


Figure 56 PPIL1 residue A99 is positioned in the hydrophobic core of PPIL1.

Hydrophobic residues are coloured in red and hydrophilic residues are coloured in blue (using Chimera kdHydrophobicity values) on the published crystal structure of PPIL1[56]). Buried residues are represented as spheres (and were calculated using normalised solvent accessibility data from STRIDE; <10 % solvent accessibility of residues[177-179]).

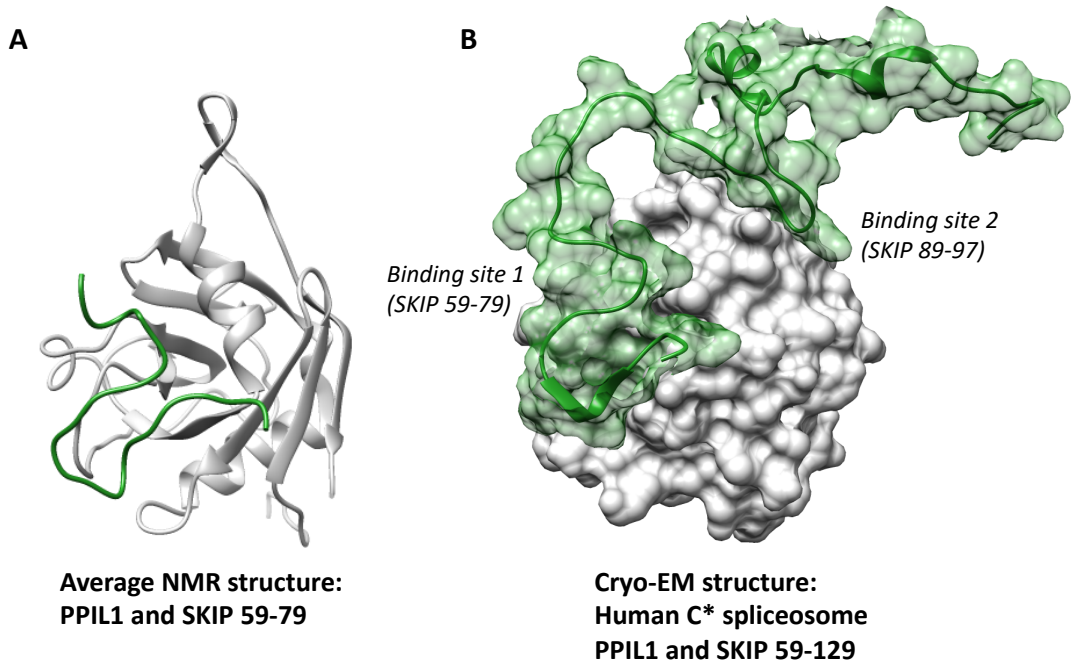


Figure 57 PPIL1 (grey) and SKIP (green) interaction, as shown by published NMR (A) and spliceosome cryo-EM (B) structures. NMR studies only modelled the interaction using a small fragment of SKIP (59-79)[86]; however, the cryo-EM model is based on full length SKIP and the PPIL1 interacting region (59-129) is shown here. The cryo-EM model shown is the human C* spliceosome complex[47]. The cryo-EM structure indicates that a second region of SKIP may bind to PPIL1 (SKIP 89-97). Images were created using data from PDB (2K7N[86]; 5XJC[47]).

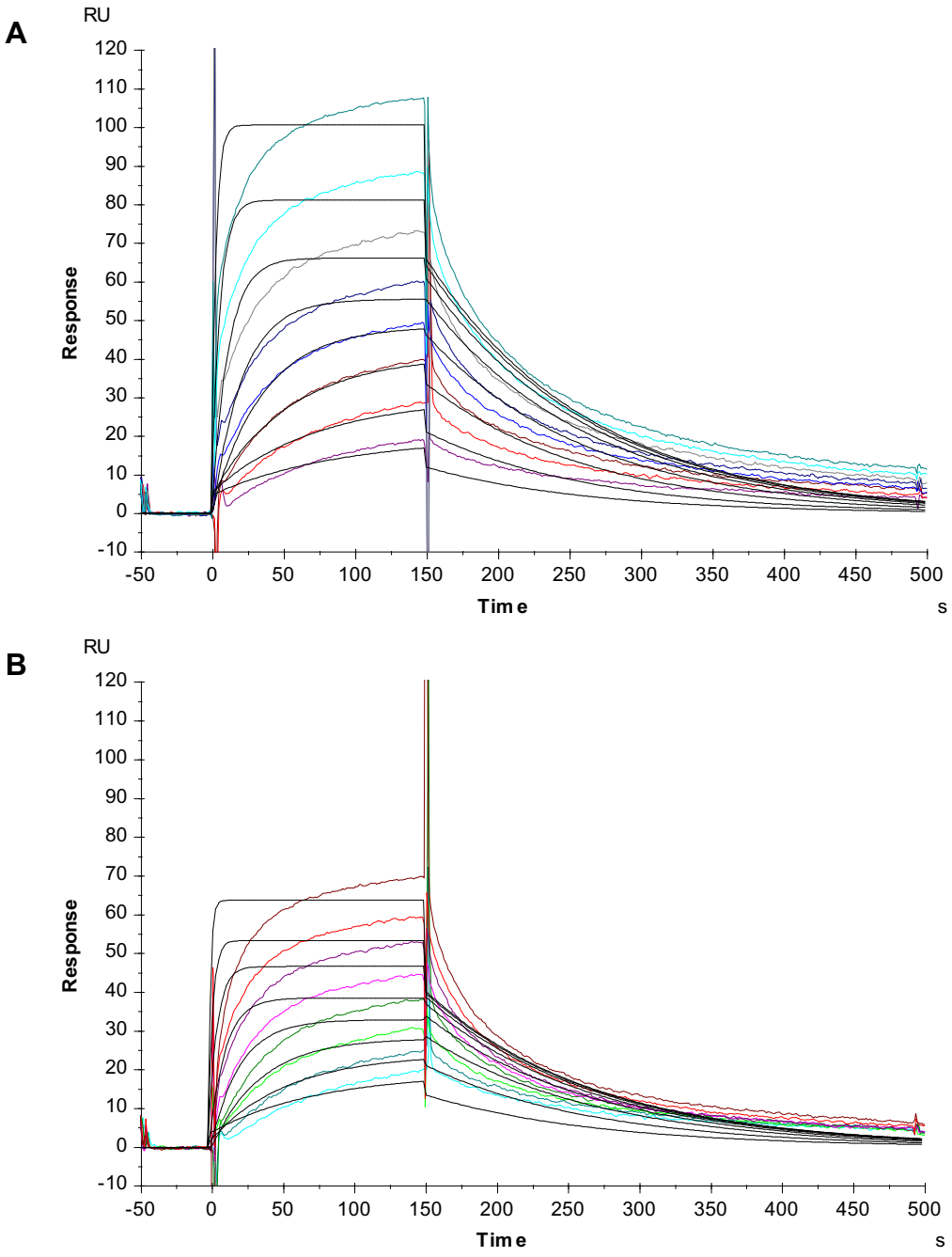


Figure 58 SPR data fitting for binding of PPIL1 proteins (WT [A], T107A [B]) to SKIP 59-129. SKIP 59-129 was immobilised on the chip surface and PPIL1 proteins were injected across the chip at a range of concentrations (15.6, 31.3, 62.5, 125, 250, 500 1000, 2000 nM; coloured lines). Binding curves were fit to a 1:1 Langmuir binding model, using BIAevaluation software (black lines). The software was unable to provide a fit which closely followed the data.

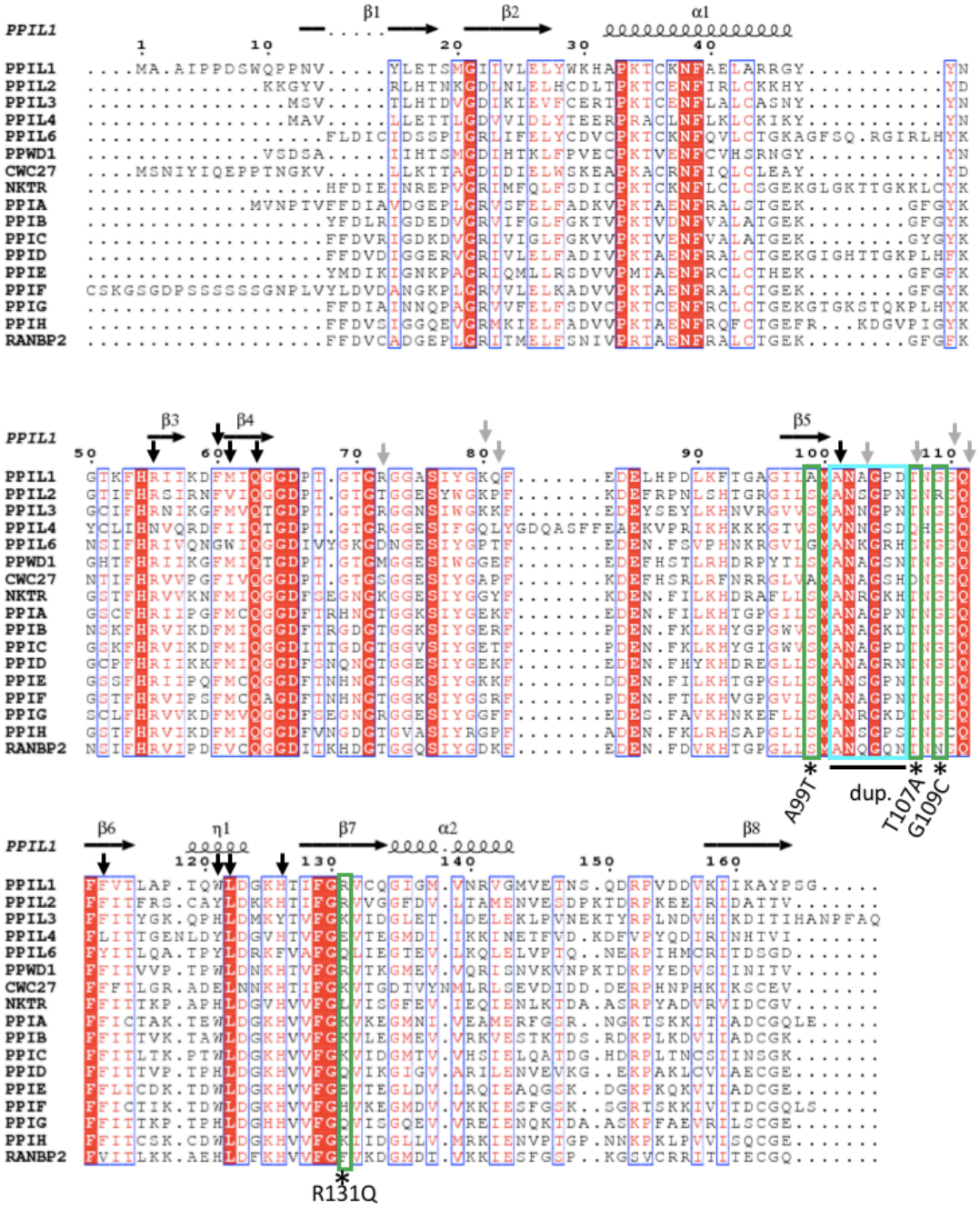


Figure 59 Comparison of amino acid conservation between the human cyclophilin protein family. Conserved residues are shown with red letters and strictly conserved amino acids are highlighted in red with white letters. Positions of the patient mutations are indicated with green boxes for amino acid substitutions and a cyan box for the duplication region. Active site and S2 pocket residues defined by Davis et al. are indicated with black and grey vertical arrows respectively[39]. Sequences were aligned using Clustal Omega and rendered for sequence similarities using ESPript.

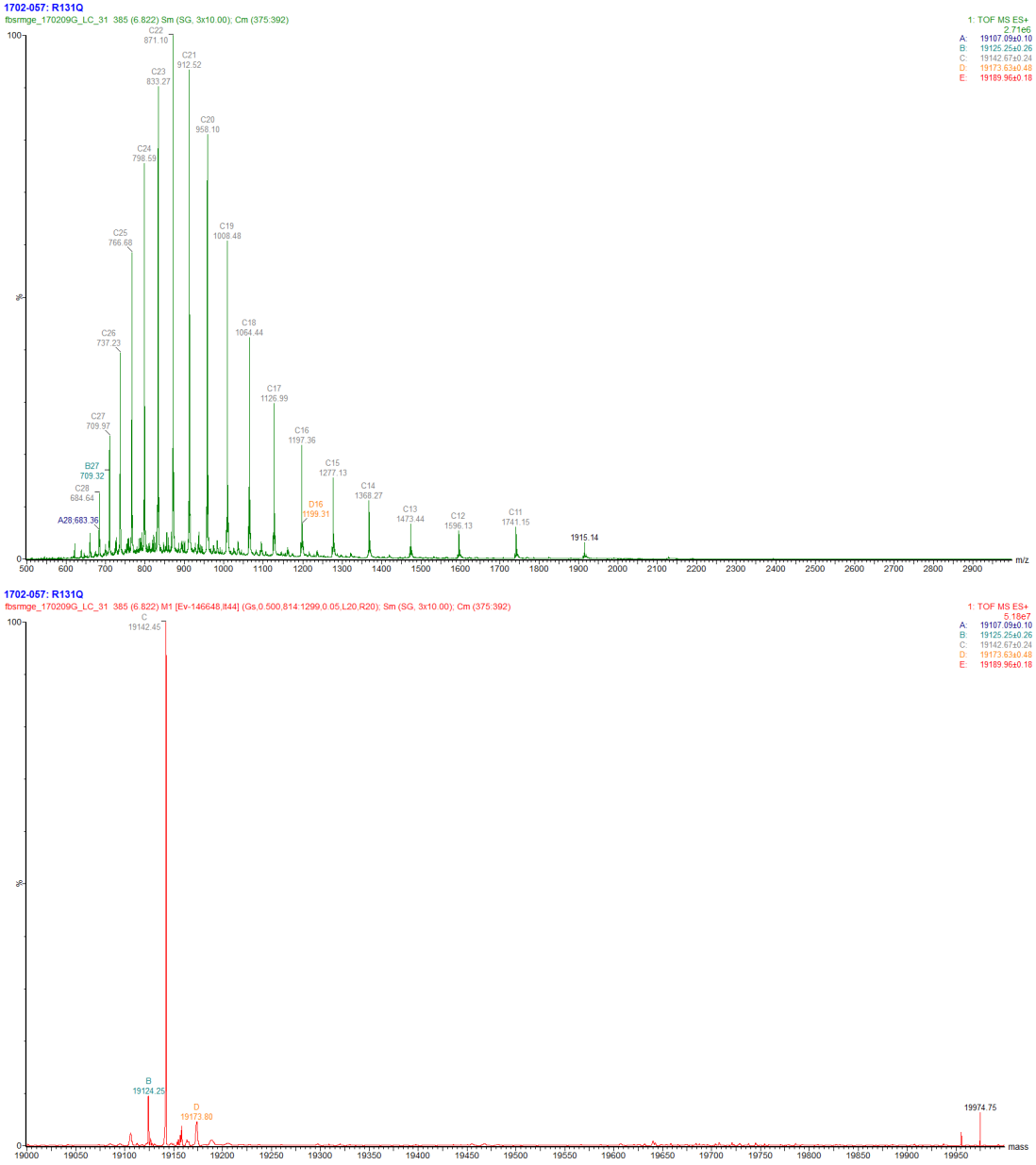
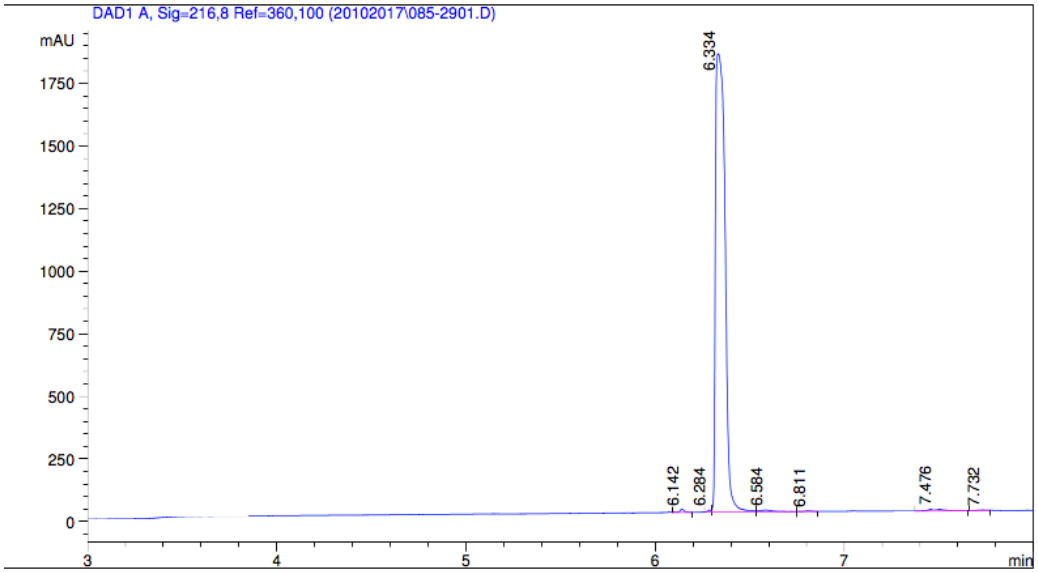


Figure 60 Representative time of flight mass spectrometry data, showing results for PPIL1 R131Q. Peak hight (%) is plotted against the mass-to-charge ratio (m/z) [top] or mass [bottom].

```
Sample No:                          45727RPR2PB
File:                 C:\CHEM32\2\DATA\20102017\085-2901.D


Column:               Kinetex 2.6u  XB-C18 100A
Method:               C1_FA0-80_60.M
Instrument:           Instrument  2
Flow Rate:            1.5mls/min
Injection Volume:     20ul
Method Info:          Analysis carried out using a 100A 4.6 x 50mm column, gradient from 0%
                      - 80% Acetonitrile, in 8 minutes. At 60C.
```



#	Meas. Ret.	Area	Area %
1	6.142	21.785	0.333
2	6.284	10.181	0.156
3	6.334	6409.684	98.094
4	6.584	32.000	0.490
5	6.811	5.468	0.084
6	7.476	45.170	0.691
7	7.732	9.956	0.152

Figure 61 Reverse phase HPLC data provided by Peptide Synthetics for PRP17 peptide (labelled proline). The area under the curve confirms that the peptide is of >98 % purity.

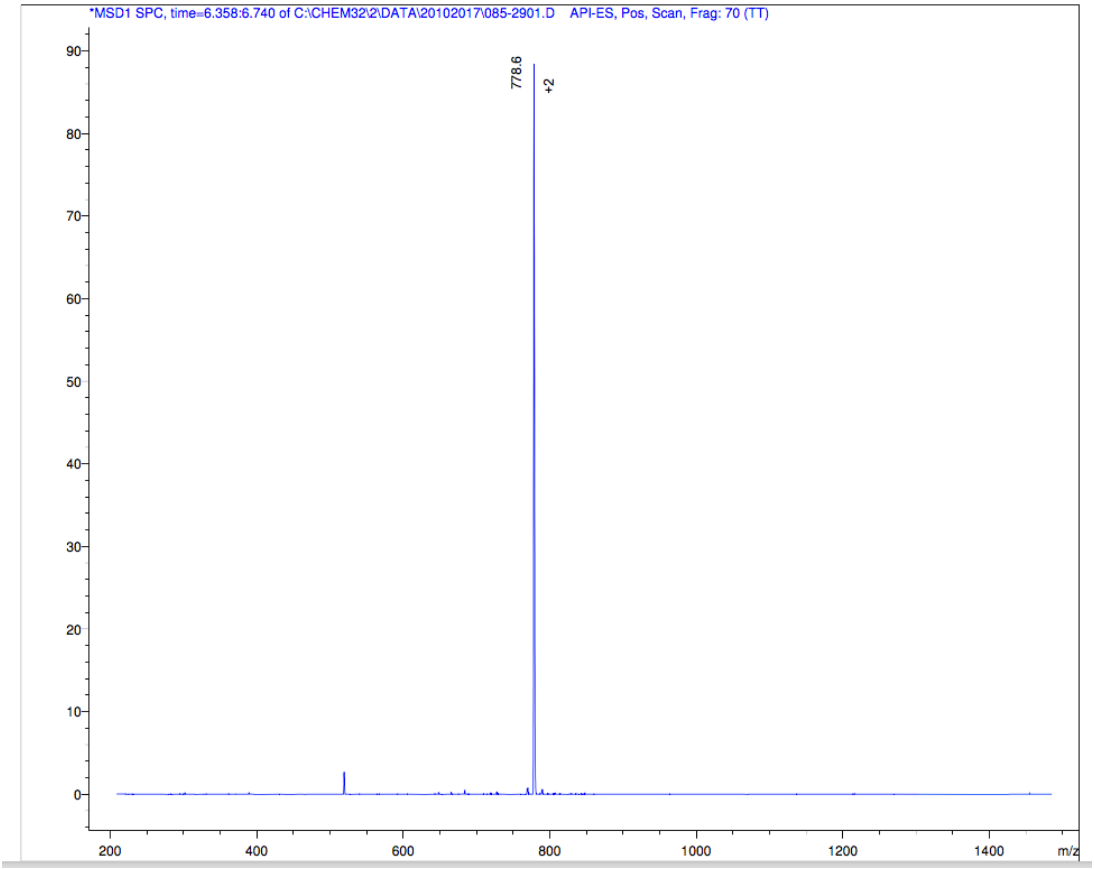


Figure 62 Mass spectrometry results provided by Peptide Synthetics. (Peaks of species with +2 and +3 charges are shown). Results therefore show that the peptide has a single mass of 15557 Da (within 1 kDa of the predicted molecular weight, 1556 kDa).

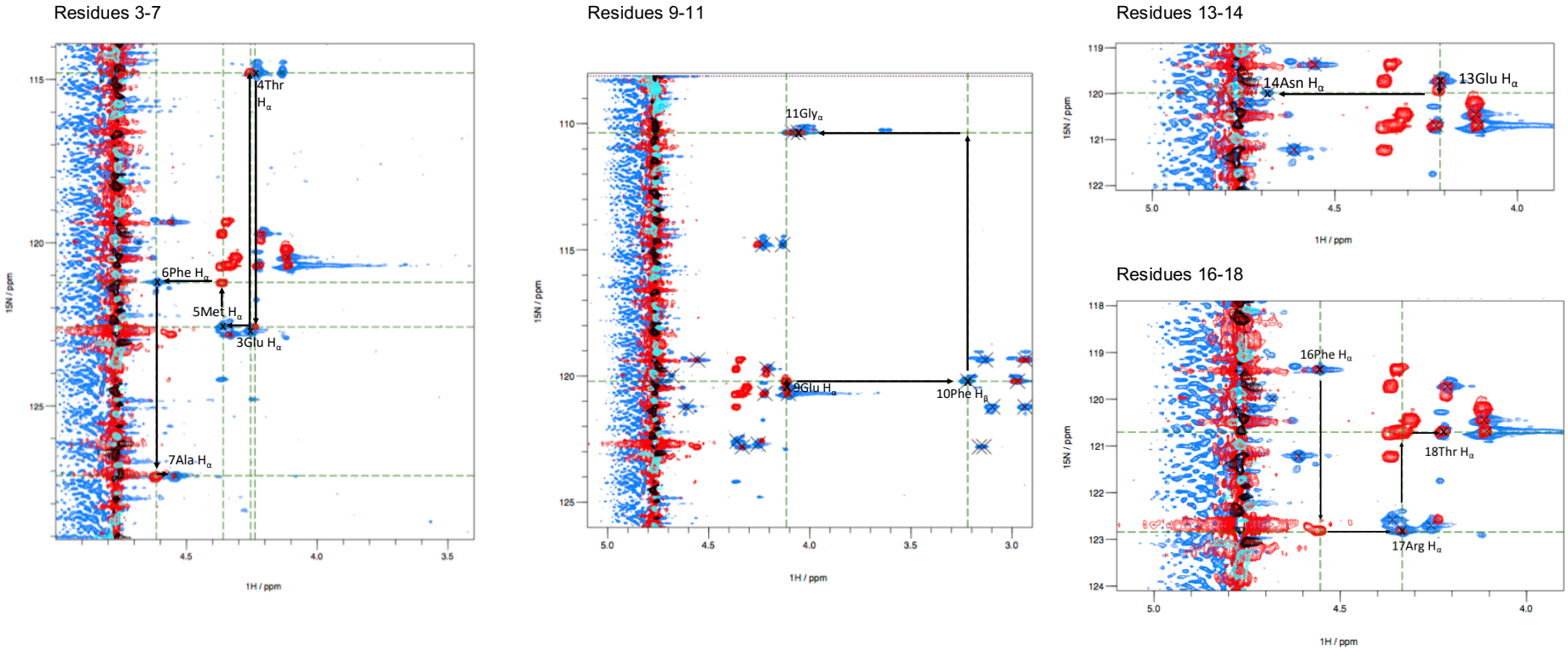
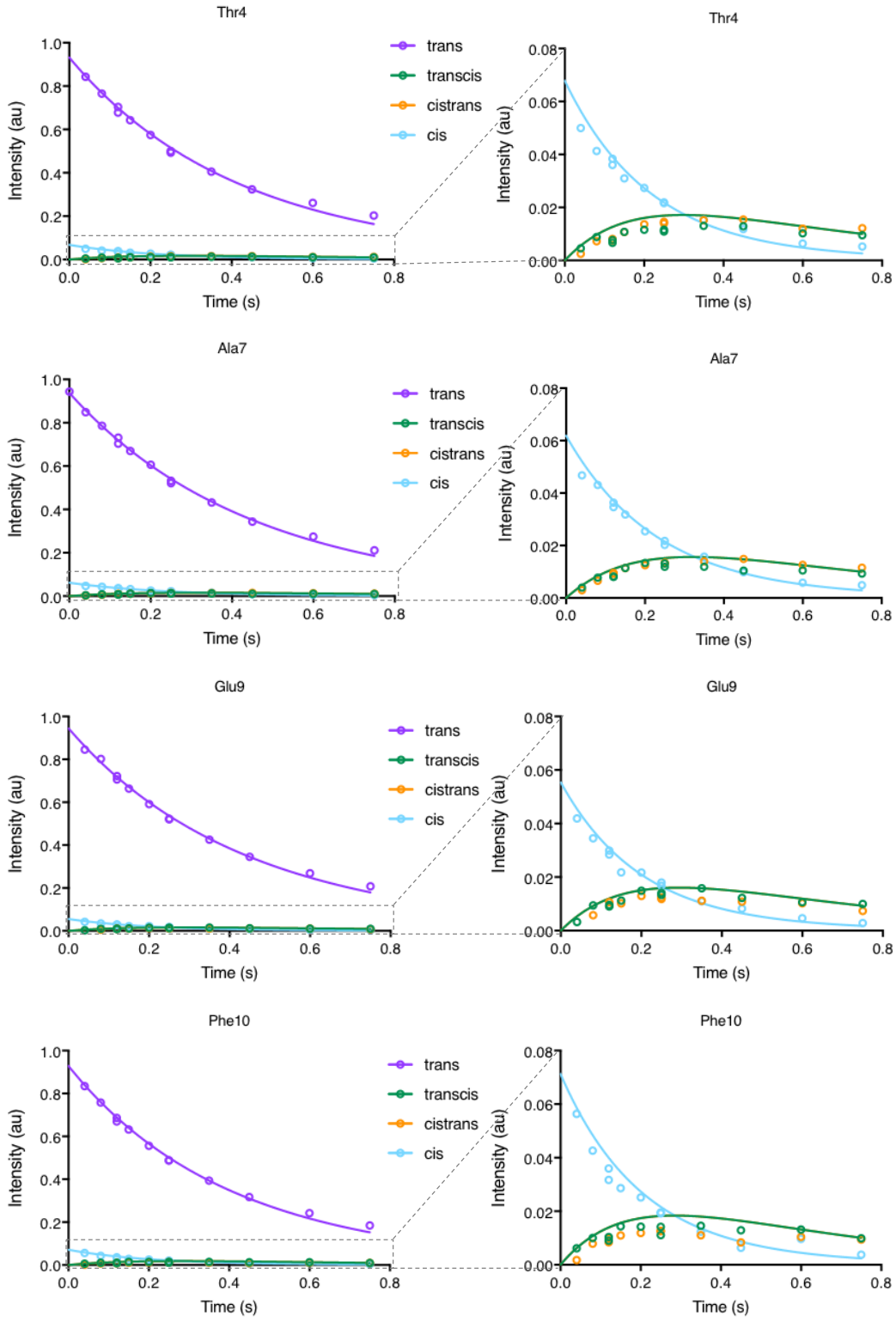


Figure 63 Enlarged regions of ^{1}H ^{15}N HSQC-TOCSY (blue) and ^{1}H ^{15}N HSQC-ROESY (red) spectra used for assignment of PRP17 peptide (84-101). Sequential connections could be observed between all consecutive residues, except where interrupted by proline residues (Pro8, Pro12, Pro15). (Residues 1 and 2 could not be observed in the spectra.)


Thr4
trans
transcis
cistrans
cis
Intensity (au)
Time (s)
Thr4
Intensity (au)
Time (s)
Ala7
trans
transcis
cistrans
cis
Intensity (au)
Time (s)
Ala7
Intensity (au)
Time (s)
Glu9
trans
transcis
cistrans
cis
Intensity (au)
Time (s)
Glu9
Intensity (au)
Time (s)
Phe10
trans
transcis
cistrans
cis
Intensity (au)
Time (s)
Phe10
Intensity (au)
Time (s)

continued…

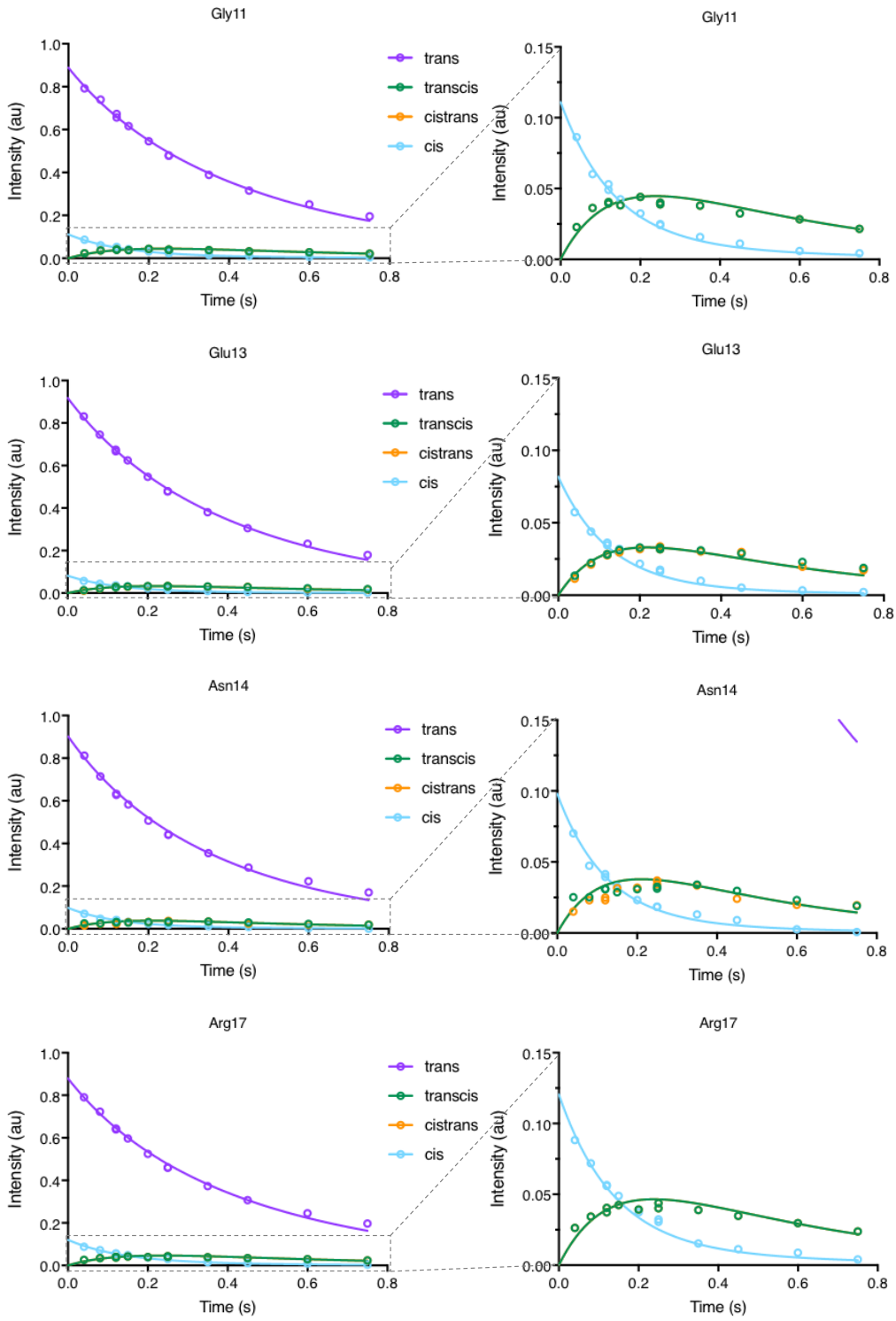
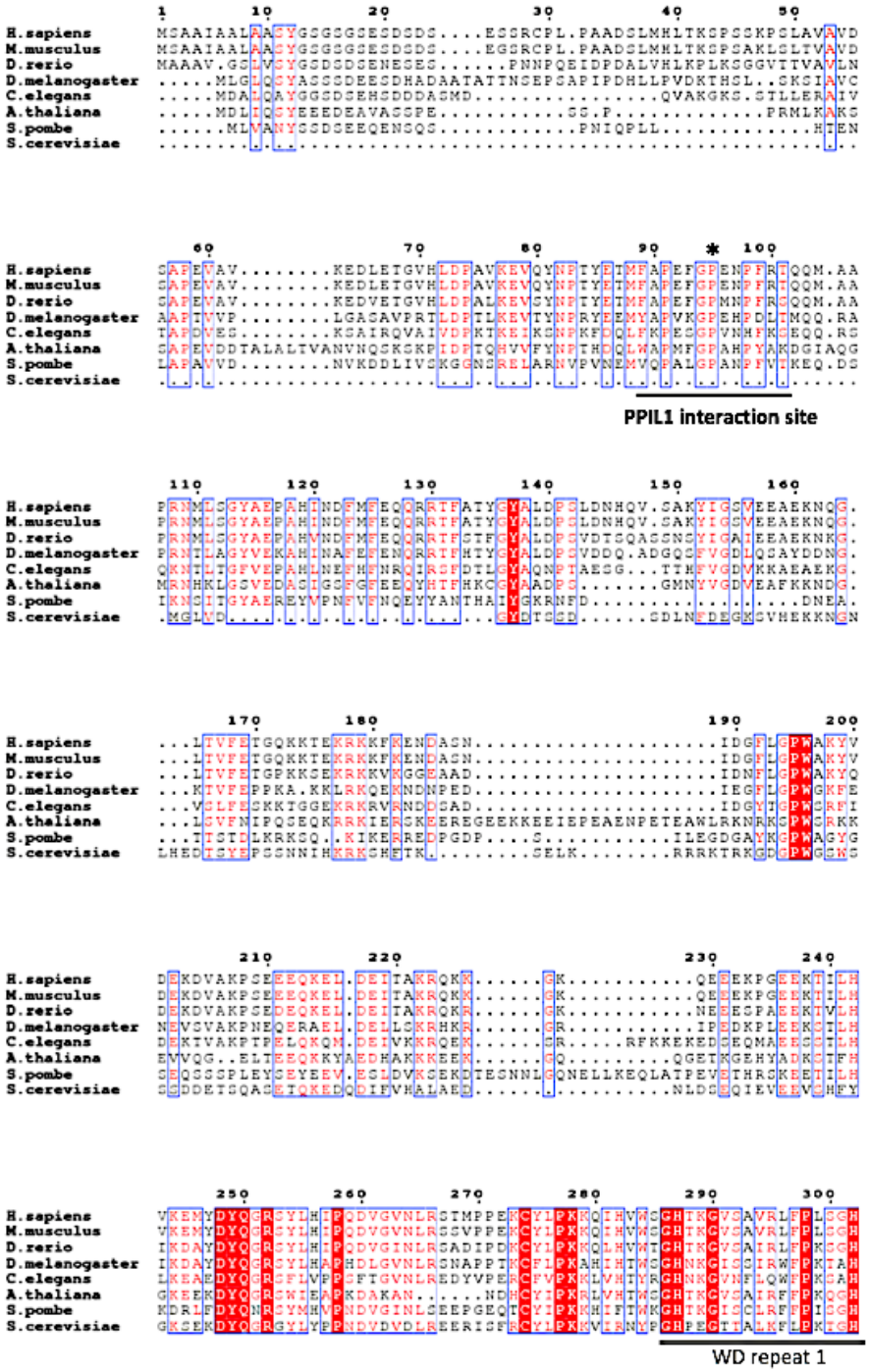


Figure 64 PRP17 peptide proline isomerisation build up curves. Build up curves were plotted for each residue of PRP17 with clearly resolved *cis*, *trans* and exchange peaks. The normalised intensities of auto (*cis* and *trans*) and exchange peaks (*trans-cis, cis-trans*) from ^{1}H ^{15}N ZZ exchange spectra were plotted as a function of mixing times and fit to two state exchange equations (see Chapter 2: materials and methods). Two experiments were collected in duplicate (experiments with 150 ms and 250 ms mixing times).


H.sapiens
M.musculus
D.rerio
D.melanogaster
C.elegans
A.thaliana
S.pombe
S.cerevisiae
PPIL1 interaction site
WD repeat 1

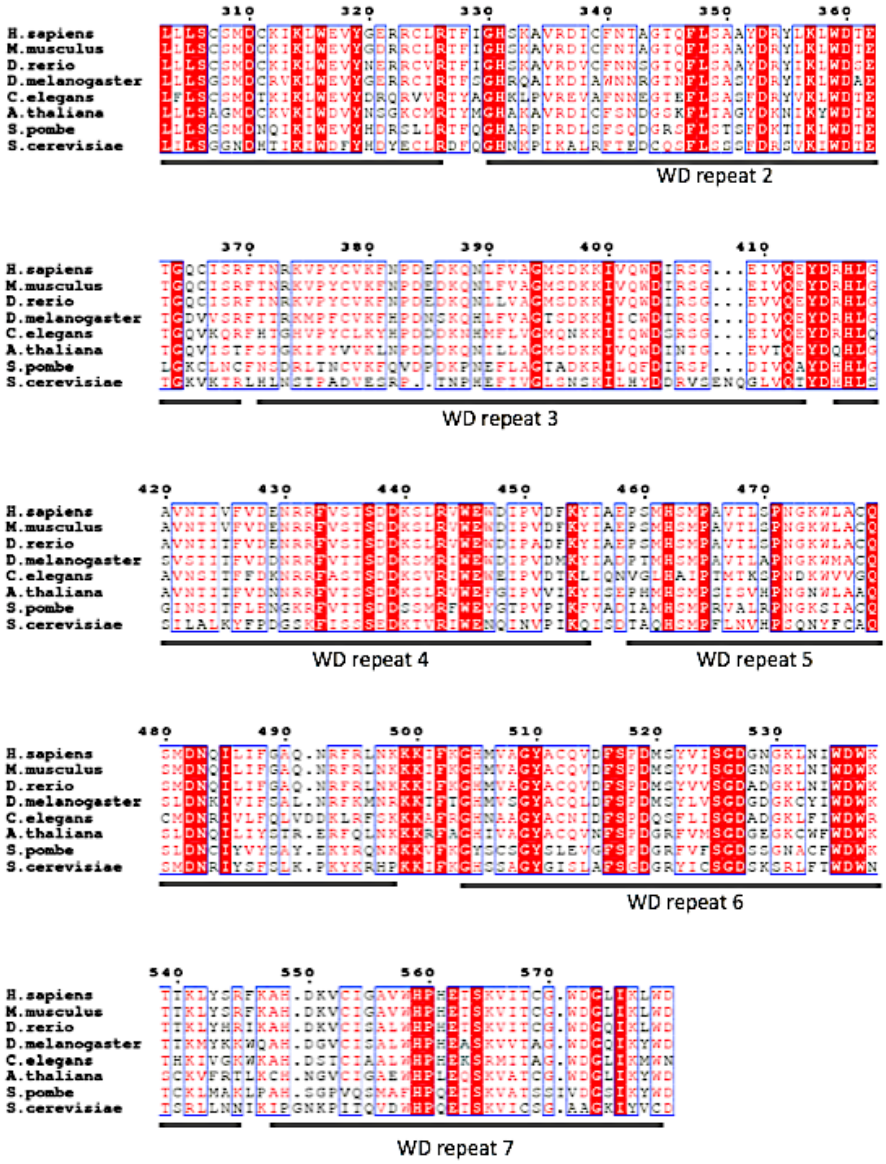


Figure 65 PRP17 conservation across species: *Homo sapiens* (NP_056975.1), *Mus musculus* (NP_082155.1), *Danio rerio* (NP_001009990.1), *Drosophila melanogaster* (NP_651005.1), *Caenorhabditis elegans* (NP_492851.1), *Arabidopsis thaliana* (NP_172528.2), *Schizosaccharomyces pombe* (NP_596089.1) and *Saccharomyces cerevisiae* (NP_010652.3). The PPIL1 interaction site, based on spliceosome cryo-EM structures[47,48] is indicated. The N terminal region of PRP17, including the PPIL1 binding site, is absent in S. cerevisiae, which also lacks a PPIL1 homologue. The WD40 C-terminal domain is conserved between species, from Human to S. cerevisiae. Sequences were aligned using Clustal Omega[32] and rendered for sequence similarities using ESPript[33].

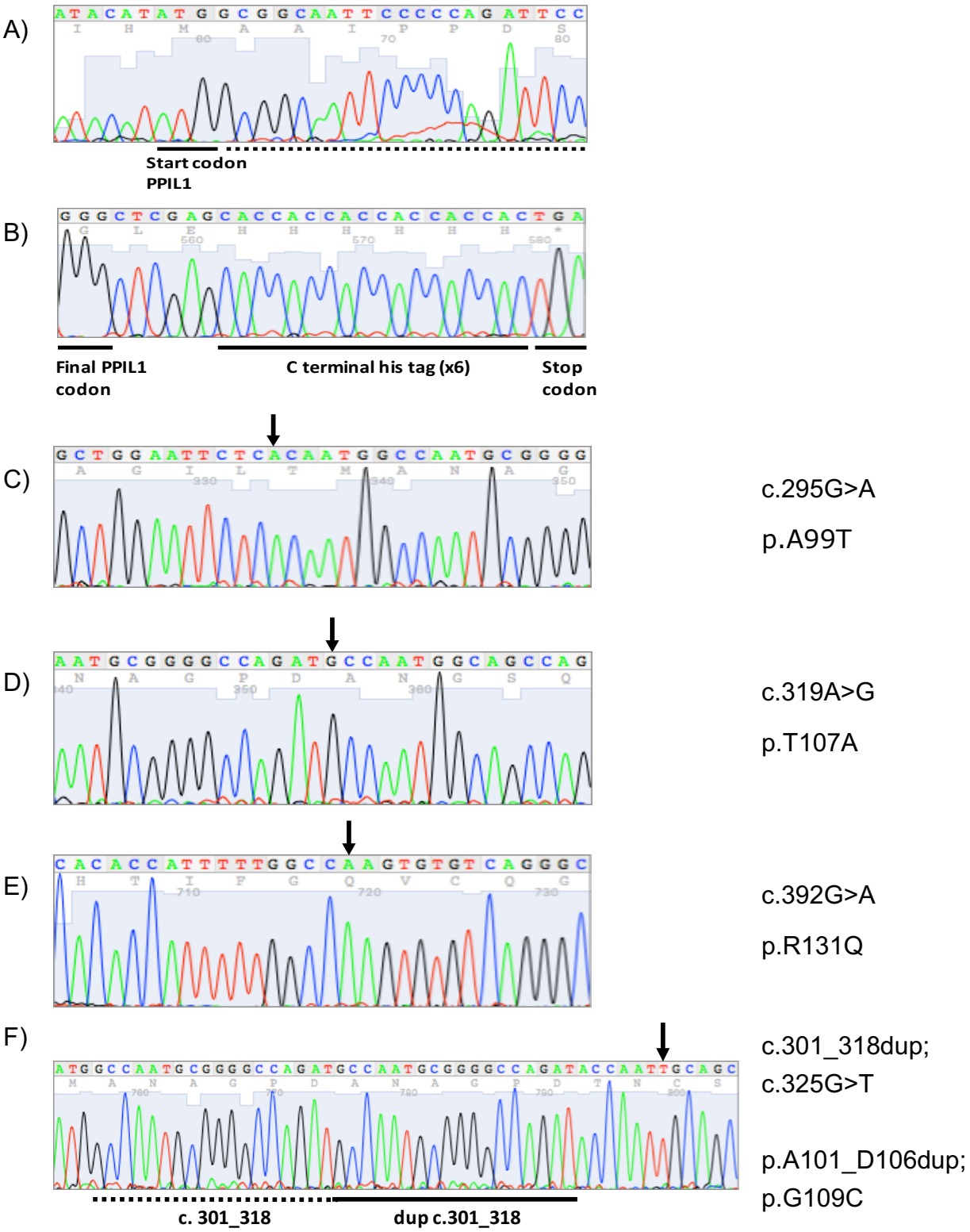


Figure 66 Example PPIL1 pET22b+ vector sequence chromatograms. Full length PPIL1 was cloned into the multiple cloning site; A) shows codon 1 and B) shows the C-terminal His-tag in frame with PPIL1 coding sequence. Arrows mark the positions of point mutations in the four mutant vectors C) A99T D) T107A E) R131Q and F) G109C and the duplication in the final vector is indicated by a horizontal line.

Brain Malformation Disorders Gene Panel

Genes	Symbol	MIM	HGNC
actin, beta	ACTB	102630	132
actin, gamma 1	ACTG1	102560	144
adhesion G protein-coupled receptor G1	ADGRG1	604110	4512
ADP-ribosylation factor guanine nucleotide-exchange factor 2	ARFGEF2	605371	15853
AKT serine/threonine kinase 3	AKT3	611223	393
aristaless related homeobox	ARX	300382	18060
calcium/calmodulin dependent serine protein kinase	CASK	300172	1497
cyclin D2	CCND2	123833	1583
doublecortin	DCX	300121	2714
dynein, cytoplasmic 1, heavy chain 1	DYNC1H1	600112	2961
ER membrane associated RNA degradation	ERMARD	615532	21056
filamin A	FLNA	300017	3754
fukutin	FKTN	607440	3622
fukutin related protein	FKRP	606596	17997
G protein signalling modulator 2	GPSM2	609245	29501
isoprenoid synthase domain containing	ISPD	614631	37276
KIF1 binding protein	KIF1BP	609367	23419
kinesin family member 2A	KIF2A	602591	6318
kinesin family member 5C	KIF5C	604593	6325
laminin subunit alpha 2	LAMA2	156225	6482

Disorders	Symbol	MIM
Baraitser-Winter Syndrome 1	BRWS1	243310
Baraitser-Winter Syndrome 2	BRWS2	614583
Cerebellar Ataxia, Mental Retardation, And Dysequilibrium Syndrome 1	CAMRQ1	224050

Chudley-Mccullough Syndrome	CMCS	604213
Cortical Dysplasia, Complex, With Other Brain Malformations 2	CDCBM2	615282
Cortical Dysplasia, Complex, With Other Brain Malformations 3	CDCBM3	615411
Cortical Dysplasia, Complex, With Other Brain Malformations 4	CDCBM4	615412
Cortical Dysplasia, Complex, With Other Brain Malformations 5	CDCBM5	615763
Cortical Dysplasia, Complex, With Other Brain Malformations 6	CDCBM6	615771
Cortical Dysplasia, Complex, With Other Brain Malformations 7	CDCBM7	610031
Cortical Dysplasia, Complex, With Other Brain Malformations 8	CDCBM8	613180
Cortical Malformations, Occipital	OCCM	614115
Goldberg-Shprintzen Syndrome	GOSHS	609460
Lissencephaly 1	LIS1	607432
Lissencephaly 2	LIS2	257320
Lissencephaly 3	LIS3	611603
Lissencephaly 4	LIS4	614019
Lissencephaly 5	LIS5	615191
Lissencephaly, X-Linked, 1	LISX	300067
Lissencephaly, X-Linked, 2	LISX2	300215

Pontocerebellar Hypoplasia Disorders Gene Panel

Genes	Symbol	MIM	HGNC
arginyl-tRNA synthetase 2, mitochondrial	RARS2	611524	21406
calcium/calmodulin dependent serine protein kinase	CASK	300172	1497
charged multivesicular body protein 1A	CHMP1A	164010	8740
exosome component 3	EXOSC3	606489	17944

Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase	SEPSECS	613009	30605
tRNA splicing endonuclease subunit 2	TSEN2	608753	28422
tRNA splicing endonuclease subunit 34	TSEN34	608754	15506
tRNA splicing endonuclease subunit 54	TSEN54	608755	27561
vaccinia related kinase 1	VRK1	602168	12718

Disorders	Symbol	MIM
Mental Retardation And Microcephaly With Pontine And Cerebellar Hypoplasia	MICPCH	300749
Pontocerebellar Hypoplasia, Type 1a	PCH1A	607596
Pontocerebellar Hypoplasia, Type 1b	PCH1B	614678
Pontocerebellar Hypoplasia, Type 2a	PCH2A	277470
Pontocerebellar Hypoplasia, Type 2b	PCH2B	612389
Pontocerebellar Hypoplasia, Type 2c	PCH2C	612390
Pontocerebellar Hypoplasia, Type 2d	PCH2D	613811
Pontocerebellar Hypoplasia, Type 4	PCH4	225753
Pontocerebellar Hypoplasia, Type 6	PCH6	611523
Pontocerebellar Hypoplasia, Type 8	PCH8	614961

Table 19 Brain Malformation Disorders Gene Panel and Pontocerebellar Hypoplasia Disorders Gene Panel - Yorkshire Regional Genetics Service (The Leeds Teaching Hospital NHS Trust). [https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/details/3836/][https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/details/6597/]

NHS
National Research Ethics Service

South Yorkshire Research Ethics Committee
Millside
Mill Pond Lane
Meanwood
Leeds
LS6 4RA

Telephone: 0113 305 0128

18 February 2011

Prof Colin A Johnson
Professor of Medical & Molecular Genetics
Section of Ophthalmology and Neuroscience
Leeds Inst of Molecular Medicine
St James University Hospital
Leeds LS9 7TF

Dear Prof Johnson

Study Title: **Molecular genetic investigations of autosomal recessive conditions**
REC reference number: **11/H1310/1**

Thank you for your letter of 8th February 2011, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where the only involvement of the NHS organisation is as a Participant Identification Centre (PIC), management permission for research is not required but the R&D office should be notified of the study and agree to the organisation's involvement. Guidance on procedures for PICs is available in IRAS. Further advice should be sought from the R&D office where necessary.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Investigator CV	1	14 December 2010
Covering Letter		08 February 2011
Letter from Sponsor		
REC application		14 December 2010
Response to Request for Further Information		
Participant Information Sheet: for children, aged 8 years or younger (Mirpuri Urdu translation)	1	14 December 2010
Evidence of insurance or indemnity		
Participant Information Sheet: Children aged 12 - 15 years	2	08 February 2011
Letter of invitation to participant	1	14 December 2010
Participant Information Sheet: for parents (Mirpuri Urdu translation)	1	14 December 2010
Participant Information Sheet: Relatives that are adults or young persons	2	08 February 2011
Participant Information Sheet: for children aged 8 year or younger	1	14 December 2010
Participant Information Sheet: Parents	2	08 February 2011
Participant Information Sheet: for children aged 8-12 years	1	14 December 2010
Participant Consent Form: Assent form for older children	1	14 December 2010
Protocol	1	14 December 2010

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

11/H1310/1	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Ms Jo Abbott
Chair

Email: Sinead.audsley@leedspft.nhs.uk

Enclosures: "After ethical review – guidance for researchers"

Copy to: *Mrs Rachel E de Souza, University of Leeds*
Mrs Anne Gowing, Leeds Teaching Hospitals NHS Trust

List of References

1 Paridaen, J. T. & Huttner, W. B. Neurogenesis during development of the vertebrate central nervous system. *EMBO Rep* **15**, 351-364, doi:10.1002/embr.201438447 (2014).

2 Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nat Rev Neurosci* **9**, 110-122, doi:10.1038/nrn2252 (2008).

3 Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol Rev* **20**, 327-348, doi:10.1007/s11065-010-9148-4 (2010).

4 Taverna, E., Götz, M. & Huttner, W. B. The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu Rev Cell Dev Biol* **30**, 465-502, doi:10.1146/annurev-cellbio-101011-155801 (2014).

5 Fernandez, V., Llinares-Benadero, C. & Borrell, V. Cerebral cortex expansion and folding: what have we learned? *EMBO J* **35**, 1021-1044, doi:10.15252/embj.201593701 (2016).

6 Nawathe, A., Doherty, J. & Pandya, P. Fetal microcephaly. *BMJ* **361**, k2232, doi:10.1136/bmj.k2232 (2018).

7 Duerinckx, S. & Abramowicz, M. The genetics of congenitally small brains. *Semin Cell Dev Biol* **76**, 76-85, doi:10.1016/j.semcdb.2017.09.015 (2018).

8 Naveed, M. *et al.* Comprehensive review on the molecular genetics of autosomal recessive primary microcephaly (MCPH). *Genet Res (Camb)* **100**, e7, doi:10.1017/S0016672318000046 (2018).

9 Jayaraman, D., Bae, B. I. & Walsh, C. A. The Genetics of Primary Microcephaly. *Annu Rev Genomics Hum Genet*, doi:10.1146/annurev-genom-083117-021441 (2018).

10 Mochida, G. H. Genetics and biology of microcephaly and lissencephaly. *Semin Pediatr Neurol* **16**, 120-126, doi:10.1016/j.spen.2009.07.001 (2009).

11 Dixon-Salazar, T. J. & Gleeson, J. G. Genetic regulation of human brain development: lessons from Mendelian diseases. *Ann N Y Acad Sci* **1214**, 156-167, doi:10.1111/j.1749-6632.2010.05819.x (2010).

12 Chavali, P. L., Pütz, M. & Gergely, F. Small organelle, big responsibility: the role of centrosomes in development and disease. *Philos Trans R Soc Lond B Biol Sci* **369**, doi:10.1098/rstb.2013.0468 (2014).

13 Faheem, M. *et al.* Molecular genetics of human primary microcephaly: an overview. *BMC Med Genomics* **8 Suppl 1**, S4, doi:10.1186/1755-8794-8-S1-S4 (2015).

14 Alcantara, D. & O'Driscoll, M. Congenital microcephaly. *Am J Med Genet C Semin Med Genet* **166C**, 124-139, doi:10.1002/ajmg.c.31397 (2014).

15 Gilmore, E. C. & Walsh, C. A. Genetic causes of microcephaly and lessons for neuronal development. *Wiley Interdiscip Rev Dev Biol* **2**, 461-478, doi:10.1002/wdev.89 (2013).

16 Barkovich, A. J., Guerrini, R., Kuzniecky, R. I., Jackson, G. D. & Dobyns, W. B. A developmental and genetic classification for malformations of cortical development: update 2012. *Brain* **135**, 1348-1369, doi:10.1093/brain/aws019 (2012).

17 van Dijk, T., Baas, F., Barth, P. G. & Poll-The, B. T. What's new in pontocerebellar hypoplasia? An update on genes and subtypes. *Orphanet J Rare Dis* **13**, 92, doi:10.1186/s13023-018-0826-2 (2018).

18 Rudnik-Schoneborn, S., Barth, P. G. & Zerres, K. Pontocerebellar hypoplasia. *Am J Med Genet C Semin Med Genet* **166C**, 173-183, doi:10.1002/ajmg.c.31403 (2014).

19 Wang, J. & Shen, Y. When a "disease-causing mutation" is not a pathogenic variant. *Clin Chem* **60**, 711-713, doi:10.1373/clinchem.2013.215947 (2014).

20 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).

21 MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-476, doi:10.1038/nature13127 (2014).

22 Flanagan, S. E., Patch, A. M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* **14**, 533-537, doi:10.1089/gtmb.2010.0036 (2010).

23 Carr, I. M. *et al.* Autozygosity mapping with exome sequence data. *Hum Mutat* **34**, 50-56, doi:10.1002/humu.22220 (2013).

24 Alkuraya, F. S. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. *Hum Genet* **132**, 1197-1211, doi:10.1007/s00439-013-1344-x (2013).

25 Schumann, M. *et al.* Identification of low abundance cyclophilins in human plasma. *Proteomics*, doi:10.1002/pmic.201600221 (2016).

26 Prunotto, M. *et al.* Proteomic analysis of podocyte exosome-enriched fraction from normal human urine. *J Proteomics* **82**, 193-229, doi:10.1016/j.jprot.2013.01.012 (2013).

27 Ozaki, K. *et al.* Cloning, expression and chromosomal mapping of a novel cyclophilin-related gene (PPIL1) from human fetal brain. *Cytogenet Cell Genet* **72**, 242-245 (1996).

28 *The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: [https://gtexportal.org/home/gene/PPIL1] the GTEx Portal on 12/05/19 and dbGaP accession number phs000424.v7.p2.*

29 Gustafson, C. L. *et al.* A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. *Mol Cell* **66**, 447-457 e447, doi:10.1016/j.molcel.2017.04.011 (2017).

30 Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718, doi:10.1016/j.cell.2009.02.009 (2009).

31 McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41**, W597-600, doi:10.1093/nar/gkt376 (2013).

32 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).

33 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324, doi:10.1093/nar/gku316 (2014).

34 Schmidpeter, P. A., Koch, J. R. & Schmid, F. X. Control of protein function by prolyl isomerization. *Biochim Biophys Acta* **1850**, 1973-1982, doi:10.1016/j.bbagen.2014.12.019 (2015).

35 Schmidpeter, P. A. & Schmid, F. X. Prolyl isomerization and its catalysis in protein folding and protein function. *J Mol Biol* **427**, 1609-1631, doi:10.1016/j.jmb.2015.01.023 (2015).

36 Reimer, U., el Mokdad, N., Schutkowski, M. & Fischer, G. Intramolecular assistance of cis/trans isomerization of the histidine-proline moiety. *Biochemistry* **36**, 13802-13808, doi:10.1021/bi9713916 (1997).

37 Shen, Y. & Bax, A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR* **46**, 199-204, doi:10.1007/s10858-009-9395-y (2010).

38 Zondlo, N. J. Aromatic-proline interactions: electronically tunable CH/pi interactions. *Acc Chem Res* **46**, 1039-1049, doi:10.1021/ar300087y (2013).

39 Davis, T. L. *et al.* Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. *PLoS Biol* **8**, e1000439, doi:10.1371/journal.pbio.1000439 (2010).

40 Kumari, S., Roy, S., Singh, P., Singla-Pareek, S. L. & Pareek, A. Cyclophilins: proteins in search of function. *Plant Signal Behav* **8**, e22734, doi:10.4161/psb.22734 (2013).

41 Theuerkorn, M., Fischer, G. & Schiene-Fischer, C. Prolyl cis/trans isomerase signalling pathways in cancer. *Curr Opin Pharmacol* **11**, 281-287, doi:10.1016/j.coph.2011.03.007 (2011).

42 Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* **3**, 619-629, doi:10.1038/nchembio.2007.35 (2007).

43 Thapar, R. Roles of Prolyl Isomerases in RNA-Mediated Gene Expression. *Biomolecules* **5**, 974-999, doi:10.3390/biom5020974 (2015).

44 Wang, P. & Heitman, J. The cyclophilins. *Genome Biol* **6**, 226, doi:10.1186/gb-2005-6-7-226 (2005).

45 Chou, I. T. & Gasser, C. S. Characterization of the cyclophilin gene family of Arabidopsis thaliana and phylogenetic analysis of known cyclophilin proteins. *Plant Mol Biol* **35**, 873-892 (1997).

46 Adams, B. M., Coates, M. N., Jackson, S. R., Jurica, M. S. & Davis, T. L. Nuclear cyclophilins affect spliceosome assembly and function in vitro. *Biochem J* **469**, 223-233, doi:10.1042/BJ20150396 (2015).

47 Zhang, X. *et al.* An Atomic Structure of the Human Spliceosome. *Cell* **169**, 918-929 e914, doi:10.1016/j.cell.2017.04.033 (2017).

48 Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182-1191, doi:10.1126/science.aac7629 (2015).
49 Zhang, X. *et al.* Structure of the human activated spliceosome in three conformational states. *Cell Res*, doi:10.1038/cr.2018.14 (2018).
50 Haselbach, D. *et al.* Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex. *Cell* **172**, 454-464 e411, doi:10.1016/j.cell.2018.01.010 (2018).
51 Kastner, B., Will, C. L., Stark, H. & Lührmann, R. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb Perspect Biol*, 345-364, doi:10.1101/cshperspect.a032417 (2019).
52 Bertram, K. *et al.* Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature*, doi:10.1038/nature21079 (2017).
53 Hegele, A. *et al.* Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* **45**, 567-580, doi:10.1016/j.molcel.2011.12.034 (2012).
54 UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699, doi:10.1093/nar/gky092 (2018).
55 Xu, C. *et al.* Solution structure of human peptidyl prolyl isomerase-like protein 1 and insights into its interaction with SKIP. *J Biol Chem* **281**, 15900-15908, doi:10.1074/jbc.M511155200 (2006).
56 Stegmann, C. M., Luhrmann, R. & Wahl, M. C. The crystal structure of PPIL1 bound to cyclosporine A suggests a binding mode for a linear epitope of the SKIP protein. *PLoS One* **5**, e10013, doi:10.1371/journal.pone.0010013 (2010).
57 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
58 McGowan, L. C. & Hamelberg, D. Conformational plasticity of an enzyme during catalysis: intricate coupling between cyclophilin A dynamics and substrate turnover. *Biophys J* **104**, 216-226, doi:10.1016/j.bpj.2012.11.3815 (2013).
59 Nigro, P., Pompilio, G. & Capogrossi, M. C. Cyclophilin A: a key player for human disease. *Cell Death Dis* **4**, e888, doi:10.1038/cddis.2013.410 (2013).
60 Fang, F., Flegler, A. J., Du, P., Lin, S. & Clevenger, C. V. Expression of cyclophilin B is associated with malignant progression and regulation of genes implicated in the pathogenesis of breast cancer. *Am J Pathol* **174**, 297-308, doi:10.2353/ajpath.2009.080753 (2009).
61 Obama, K. *et al.* Overexpression of peptidyl-prolyl isomerase-like 1 is associated with the growth of colon cancer cells. *Clin Cancer Res* **12**, 70-76, doi:10.1158/1078-0432.CCR-05-0588 (2006).
62 Mäkinen, V. P. *et al.* Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet* **10**, e1004502, doi:10.1371/journal.pgen.1004502 (2014).
63 Li, J. *et al.* Psoriatic T cells reduce epidermal turnover time and affect cell proliferation contributed from differential gene expression. *J Dermatol* **42**, 874-880, doi:10.1111/1346-8138.12961 (2015).

64 Mesa, A., Somarelli, J. A. & Herrera, R. J. Spliceosomal immunophilins. *FEBS Lett* **582**, 2345-2351, doi:10.1016/j.febslet.2008.06.006 (2008).
65 Scheres, S. H. & Nagai, K. CryoEM structures of spliceosomal complexes reveal the molecular mechanism of pre-mRNA splicing. *Curr Opin Struct Biol* **46**, 130-139, doi:10.1016/j.sbi.2017.08.001 (2017).
66 Shi, Y. The Spliceosome: A Protein-Directed Metalloribozyme. *J Mol Biol* **429**, 2640-2653, doi:10.1016/j.jmb.2017.07.010 (2017).
67 Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**, 108-121, doi:10.1038/nrm3742 (2014).
68 Will, C. L. & Luhrmann, R. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem* **386**, 713-724, doi:10.1515/BC.2005.084 (2005).
69 Turunen, J. J., Niemelä, E. H., Verma, B. & Frilander, M. J. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* **4**, 61-76, doi:10.1002/wrna.1141 (2013).
70 Verma, B., Akinyi, M. V., Norppa, A. J. & Frilander, M. J. Minor spliceosome and disease. *Semin Cell Dev Biol* **79**, 103-112, doi:10.1016/j.semcdb.2017.09.036 (2018).
71 Agafonov, D. E. *et al.* Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol* **31**, 2667-2682, doi:10.1128/MCB.05266-11 (2011).
72 Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim Biophys Acta* **1792**, 14-26, doi:10.1016/j.bbadis.2008.09.017 (2009).
73 Dvinge, H. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett*, doi:10.1002/1873-3468.13119 (2018).
74 Kaufer, N. F. & Potashkin, J. Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. *Nucleic Acids Res* **28**, 3003-3010 (2000).
75 van der Feltz, C., Anthony, K., Brilot, A. & Pomeranz Krummel, D. A. Architecture of the spliceosome. *Biochemistry* **51**, 3321-3333, doi:10.1021/bi201215r (2012).
76 Kuhn, A. N. & Kaufer, N. F. Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. *Curr Genet* **42**, 241-251, doi:10.1007/s00294-002-0355-2 (2003).
77 Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structure of a human catalytic step I spliceosome. *Science* **359**, 537-545, doi:10.1126/science.aar6401 (2018).
78 Fica, S. M. & Nagai, K. Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat Struct Mol Biol* **24**, 791-799, doi:10.1038/nsmb.3463 (2017).
79 Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. & Lührmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846-850, doi:10.1038/nature06842 (2008).

80 Makarov, E. M. *et al.* Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**, 2205-2208, doi:10.1126/science.1077783 (2002).
81 Makarova, O. V. & Makarov, E. M. The 35S U5 snRNP Is Generated from the Activated Spliceosome during In vitro Splicing. *PLoS One* **10**, e0128430, doi:10.1371/journal.pone.0128430 (2015).
82 Mayeda, A. & Krainer, A. R. Mammalian in vitro splicing assays. *Methods Mol Biol* **118**, 315-321, doi:10.1385/1-59259-676-2:315 (1999).
83 Baralle, D. & Baralle, M. Splicing in action: assessing disease causing sequence changes. *J Med Genet* **42**, 737-748, doi:10.1136/jmg.2004.029538 (2005).
84 Papasaikas, P., Tejedor, J. R., Vigevani, L. & Valcarcel, J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol Cell* **57**, 7-22, doi:10.1016/j.molcel.2014.10.030 (2015).
85 El Chehadeh, S. *et al.* Dominant variants in the splicing factor PUF60 cause a recognizable syndrome with intellectual disability, heart defects and short stature. *Eur J Hum Genet* **25**, 43-51, doi:10.1038/ejhg.2016.133 (2016).
86 Wang, X. *et al.* A large intrinsically disordered region in SKIP and its disorder-order transition induced by PPIL1 binding revealed by NMR. *J Biol Chem* **285**, 4951-4963, doi:10.1074/jbc.M109.087528 (2010).
87 Folk, P., Půta, F. & Skruzný, M. Transcriptional coregulator SNW/SKIP: the concealed tie of dissimilar pathways. *Cell Mol Life Sci* **61**, 629-640, doi:10.1007/s00018-003-3215-4 (2004).
88 Lee, S. S. *et al.* Rice cyclophilin OsCYP18-2 is translocated to the nucleus by an interaction with SKIP and enhances drought tolerance in rice and Arabidopsis. *Plant Cell Environ* **38**, 2071-2087, doi:10.1111/pce.12531 (2015).
89 Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* **18**, 655-670, doi:10.1038/nrm.2017.86 (2017).
90 Wilkinson, M. E., Lin, P. C., Plaschka, C. & Nagai, K. Cryo-EM Studies of Pre-mRNA Splicing: From Sample Preparation to Model Visualization. *Annu Rev Biophys*, doi:10.1146/annurev-biophys-070317-033410 (2018).
91 Bertram, K. *et al.* Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* **170**, 701-713 e711, doi:10.1016/j.cell.2017.07.011 (2017).
92 Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Res* **28**, 1129-1140, doi:10.1038/s41422-018-0094-7 (2018).
93 Fica, S. M., Oubridge, C., Wilkinson, M. E., Newman, A. J. & Nagai, K. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* **363**, 710-714, doi:10.1126/science.aaw5569 (2019).
94 Zhang, X. *et al.* Structures of the human spliceosomes before and after release of the ligated exon. *Cell Res*, doi:10.1038/s41422-019-0143-x (2019).

95 Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**, 749-761, doi:10.1038/nrg2164 (2007).
96 Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat Rev Genet* **17**, 19-32, doi:10.1038/nrg.2015.3 (2016).
97 Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777-793, doi:10.1016/j.cell.2009.02.011 (2009).
98 Anna, A. & Monika, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*, doi:10.1007/s13353-018-0444-7 (2018).
99 Liu, M. M. & Zack, D. J. Alternative splicing and retinal degeneration. *Clin Genet* **84**, 142-149, doi:10.1111/cge.12181 (2013).
100 Rangel-Sosa, M. M. *et al.* Exome sequencing reveals three homozygous missense variants in SNRPA in two sisters with syndromic intellectual disability. *Clin Genet*, doi:10.1111/cge.13235 (2018).
101 Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science* **332**, 240-243, doi:10.1126/science.1202205 (2011).
102 Neuenkirchen, N., Chari, A. & Fischer, U. Deciphering the assembly pathway of Sm-class U snRNPs. *FEBS Lett* **582**, 1997-2003, doi:10.1016/j.febslet.2008.03.009 (2008).
103 Carroll, R. *et al.* Variant in the X-chromosome spliceosomal gene GPKOW causes male-lethal microcephaly with intrauterine growth restriction. *Eur J Hum Genet*, doi:10.1038/ejhg.2017.97 (2017).
104 Huang, L. *et al.* Mandibulofacial Dysostosis with Microcephaly: Mutation and Database Update. *Hum Mutat* **37**, 148-154, doi:10.1002/humu.22924 (2016).
105 Jafarifar, F., Dietrich, R. C., Hiznay, J. M. & Padgett, R. A. Biochemical defects in minor spliceosome function in the developmental disorder MOPD I. *RNA* **20**, 1078-1089, doi:10.1261/rna.045187.114 (2014).
106 Budde, B. *et al.* tRNA splicing endonuclease mutations cause pontocerebellar hypoplasia. *Nature Genetics* **40**, 1113-1118, doi:10.1038/ng.204 (2008).
107 Karaca, E. *et al.* Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. *Cell* **157**, 636-650, doi:10.1016/j.cell.2014.02.058 (2014).
108 Schaffer, A. E. *et al.* CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. *Cell* **157**, 651-663, doi:10.1016/j.cell.2014.03.049 (2014).
109 Burns, D. T. *et al.* Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy. *Am J Hum Genet* **102**, 858-873, doi:10.1016/j.ajhg.2018.03.011 (2018).
110 Boczonadi, V. *et al.* EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nat Commun* **5**, 4287, doi:10.1038/ncomms5287 (2014).
111 Zhou, H., Mangelsdorf, M., Liu, J., Zhu, L. & Wu, J. Y. RNA-binding proteins in neurological diseases. *Sci China Life Sci* **57**, 432-444, doi:10.1007/s11427-014-4647-9 (2014).

112 Linder, B., Fischer, U. & Gehring, N. H. mRNA metabolism and neuronal disease. *FEBS Lett* **589**, 1598-1606, doi:10.1016/j.febslet.2015.04.052 (2015).
113 Xu, Q., Modrek, B. & Lee, C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res* **30**, 3754-3766 (2002).
114 Carey, K. T. & Wickramasinghe, V. O. Regulatory Potential of the RNA Processing Machinery: Implications for Human Disease. *Trends Genet* **34**, 279-290, doi:10.1016/j.tig.2017.12.012 (2018).
115 Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med* **18**, 472-482, doi:10.1016/j.molmed.2012.06.006 (2012).
116 Ogden, R. C. & Adams, D. A. Electrophoresis in agarose and acrylamide gels. *Methods Enzymol* **152**, 61-87 (1987).
117 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
118 Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. in *The Proteomics Protocols Handbook, Humana Press (2005).* (ed John M. Walker) 571-607 (Humana Press, 2005).
119 Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-2890, doi:10.1038/nprot.2006.202 (2006).
120 Turnbull, W. B. & Daranas, A. H. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc* **125**, 14859-14866, doi:10.1021/ja036166s (2003).
121 John Cavanagh, W. J. F., Arthur G. Palmer, III, Nicholas J. Skelton. *Protein NMR Spectroscopy: Principles and Practice*. Second edn, (Elsevier Academic Press, 2007).
122 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
123 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696, doi:10.1002/prot.20449 (2005).
124 Xu, C. *et al.* Backbone and side chain assignments of human Peptidylprolyl Isomerase Like 1 (hPPIL1). *J Biomol NMR* **31**, 179-180, doi:10.1007/s10858-004-8238-0 (2005).
125 Hewitt, S. H. *et al.* Protein Surface Mimetics: Understanding How Ruthenium Tris(Bipyridines) Interact with Proteins. *Chembiochem* **18**, 223-231, doi:10.1002/cbic.201600552 (2017).
126 Schumann, F. H. *et al.* Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J Biomol NMR* **39**, 275-289, doi:10.1007/s10858-007-9197-z (2007).
127 Justice, R. M., Jr. *et al.* The detection of proline isomerase activity in FK506-binding protein by two-dimensional 1H NMR exchange spectroscopy. *Biochem Biophys Res Commun* **171**, 445-450 (1990).
128 Dujardin, M. *et al.* A Proline-Tryptophan Turn in the Intrinsically Disordered Domain 2 of NS5A Protein Is Essential for Hepatitis C Virus RNA Replication. *J Biol Chem* **290**, 19104-19120, doi:10.1074/jbc.M115.644419 (2015).

129 Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**, 942-968, doi:10.1016/j.bbapap.2010.10.012 (2011).
130 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).
131 Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat Protoc* **11**, 1-9, doi:10.1038/nprot.2015.123 (2016).
132 Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688, doi:10.1371/journal.pone.0046688 (2012).
133 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315, doi:10.1038/ng.2892 (2014).
134 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
135 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).
136 Eisenmesser, E. Z., Thai, V., Pozharski, E., Kern, D. Cyrstal Structure of Human Cyclophilin-A in Complex with suc-AGPF-pNA. *PDB ID: 1ZKF* (2006).
137 Duncan, E., Brown, M. & Shore, E. M. The revolution in human monogenic disease mapping. *Genes (Basel)* **5**, 792-803, doi:10.3390/genes5030792 (2014).
138 Brunham, L. R. & Hayden, M. R. Hunting human disease genes: lessons from the past, challenges for the future. *Hum Genet* **132**, 603-617, doi:10.1007/s00439-013-1286-3 (2013).
139 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*, doi:10.1093/nar/gky1016 (2018).
140 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081, doi:10.1038/nprot.2009.86 (2009).
141 Burgess, R. R. A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting. *Protein Expr Purif* **150**, 81-85, doi:10.1016/j.pep.2018.05.007 (2018).
142 Royer, C. A. Probing protein folding and conformational transitions with fluorescence. *Chem Rev* **106**, 1769-1784, doi:10.1021/cr0404390 (2006).
143 Senisterra, G., Chau, I. & Vedadi, M. Thermal denaturation assays in chemical biology. *Assay Drug Dev Technol* **10**, 128-136, doi:10.1089/adt.2011.0390 (2012).
144 Pace, C. N., Grimsley, G. R. & Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J Biol Chem* **284**, 13285-13289, doi:10.1074/jbc.R800080200 (2009).
145 Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. Physical stability of proteins in aqueous solution: mechanism and driving

forces in nonnative protein aggregation. *Pharm Res* **20**, 1325-1336 (2003).
146 Vagenende, V., Yap, M. G. & Trout, B. L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry* **48**, 11084-11096, doi:10.1021/bi900649t (2009).
147 Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, W668-673, doi:10.1093/nar/gkh371 (2004).
148 *DichroWeb: On-line analysis for protein Circular Dichroism spectra*, <http://www.cryst.bbk.ac.uk/cdweb> (2001).
149 Xu, C. *et al.* Solution structure of human peptidyl prolyl isomerase-like protein 1 and insights into its interaction with SKIP. *Journal of Biological Chemistry* **281**, 15900-15908, doi:10.1074/jbc.M511155200 (2006).
150 Zhou, M., Li, Q. & Wang, R. Current Experimental Methods for Characterizing Protein-Protein Interactions. *ChemMedChem* **11**, 738-756, doi:10.1002/cmdc.201500495 (2016).
151 Hubner, D., Drakenberg, T., Forsen, S. & Fischer, G. Peptidyl-prolyl cis-trans isomerase activity as studied by dynamic proton NMR spectroscopy. *FEBS Lett* **284**, 79-81 (1991).
152 Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* **337**, 476-478, doi:10.1038/337476a0 (1989).
153 Kofron, J. L., Kuzmic, P., Kishore, V., Colón-Bonilla, E. & Rich, D. H. Determination of kinetic constants for peptidyl prolyl cis-trans isomerases by an improved spectrophotometric assay. *Biochemistry* **30**, 6127-6134 (1991).
154 Siekierka, J. J., Hung, S. H., Poe, M., Lin, C. S. & Sigal, N. H. A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature* **341**, 755-757, doi:10.1038/341755a0 (1989).
155 Scholz, C., Schindler, T., Dolinski, K., Heitman, J. & Schmid, F. X. Cyclophilin active site mutants have native prolyl isomerase activity with a protein substrate. *FEBS Lett* **414**, 69-73 (1997).
156 Ikolo, F. *et al.* Characterisation of SEQ0694 (PrsA/PrtM) of Streptococcus equi as a functional peptidyl-prolyl isomerase affecting multiple secreted protein substrates. *Mol Biosyst* **11**, 3279-3286, doi:10.1039/c5mb00543d (2015).
157 Janowski, B., Wöllner, S., Schutkowski, M. & Fischer, G. A protease-free assay for peptidyl prolyl cis/trans isomerases using standard peptide substrates. *Anal Biochem* **252**, 299-307, doi:10.1006/abio.1997.2330 (1997).
158 Daum, S. *et al.* Isoform-specific inhibition of cyclophilins. *Biochemistry* **48**, 6268-6277, doi:10.1021/bi9007287 (2009).
159 Nagai, K. *et al.* SKIP modifies gene expression by affecting both transcription and splicing. *Biochem Biophys Res Commun* **316**, 512-517, doi:10.1016/j.bbrc.2004.02.077 (2004).

160 Reidt, U. *et al.* Crystal structure of a complex between human spliceosomal cyclophilin H and a U4/U6 snRNP-60K peptide. *J Mol Biol* **331**, 45-56 (2003).
161 Piotukh, K. *et al.* Cyclophilin A binds to linear peptide motifs containing a consensus that is present in many human proteins. *J Biol Chem* **280**, 23668-23674, doi:10.1074/jbc.M503405200 (2005).
162 Ben Yehuda, S. *et al.* Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40 yeast gene involved in splicing and cell cycle control. *RNA* **4**, 1304-1312 (1998).
163 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **73**, 1-16, doi:10.1016/j.pnmrs.2013.02.001 (2013).
164 Yee, A. A., Semesi, A., Garcia, M. & Arrowsmith, C. H. Screening proteins for NMR suitability. *Methods Mol Biol* **1140**, 169-178, doi:10.1007/978-1-4939-0354-2_13 (2014).
165 Yao, J., Dyson, H. J. & Wright, P. E. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett* **419**, 285-289 (1997).
166 Deslauriers, R., Becker, J. M., Steinfeld, A. S. & Naider, F. STERIC EFFECTS OF CIS-TRANS ISOMERISM ON NEIGHBORING RESIDUES IN PROLINE OLIGOPEPTIDES - C-13-NMR STUDY OF CONFORMATIONAL HETEROGENEITY IN LINEAR TRIPEPTIDES. *Biopolymers* **18**, 523-538, doi:10.1002/bip.1979.360180305 (1979).
167 Rajiv, C., Jackson, S. R., Cocklin, S., Eisenmesser, E. Z. & Davis, T. L. The spliceosomal proteins PPIH and PRPF4 exhibit bi-partite binding. *Biochem J* **474**, 3689-3704, doi:10.1042/BCJ20170366 (2017).
168 Jiang, Q. *et al.* A fluorescent peptidyl substrate for visualizing peptidyl-prolyl cis/trans isomerase activity in live cells. *Chem Commun (Camb)*, doi:10.1039/c7cc09135d (2018).
169 Luchinat, E. & Banci, L. A Unique Tool for Cellular Structural Biology: In-cell NMR. *J Biol Chem* **291**, 3776-3784, doi:10.1074/jbc.R115.643247 (2016).
170 El Marabti, E. & Younis, I. The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Front Mol Biosci* **5**, 80, doi:10.3389/fmolb.2018.00080 (2018).
171 Vazquez-Arango, P. & O'Reilly, D. Variant snRNPs: New players within the spliceosome system. *RNA Biol* **15**, 17-25, doi:10.1080/15476286.2017.1373238 (2018).
172 Su, C. H., D, D. & Tarn, W. Y. Alternative Splicing in Neurogenesis and Brain Development. *Front Mol Biosci* **5**, 12, doi:10.3389/fmolb.2018.00012 (2018).
173 Tanackovic, G. *et al.* PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum Mol Genet* **20**, 2116-2130, doi:10.1093/hmg/ddr094 (2011).
174 Lehalle, D. *et al.* A review of craniofacial disorders caused by spliceosomal defects. *Clin Genet* **88**, 405-415, doi:10.1111/cge.12596 (2015).
175 Lanfranco, M., Vassallo, N. & Cauchi, R. J. Spinal Muscular Atrophy: From Defective Chaperoning of snRNP Assembly to Neuromuscular

Dysfunction. *Front Mol Biosci* **4**, 41, doi:10.3389/fmolb.2017.00041 (2017).

176 Giorgio, V. *et al.* Cyclophilin D in mitochondrial pathophysiology. *Biochim Biophys Acta* **1797**, 1113-1118, doi:10.1016/j.bbabio.2009.12.006 (2010).

177 Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-579, doi:10.1002/prot.340230412 (1995).

178 Eisenhaber, F. & Argos, P. Improved strategy in analytic surface calculation for molecular systems: Handling of singularities and computational efficiency. *Journal of Computational Chemistry* **14**, 1272-1280, doi:10.1002/jcc.540141103 (1993).

179 Bendell, C. J. *et al.* Transient protein-protein interface prediction: datasets, features, algorithms, and the RAD-T predictor. *BMC Bioinformatics* **15**, 82, doi:10.1186/1471-2105-15-82 (2014).