

Biophysical Studies of Asymmetric
Homodimerisation of the microRNA
Biogenesis Cofactors PACT and TRBP

Alexander Peter Heyam

PhD

University of York

Biology

December 2016

Abstract

Processing of precursor microRNAs by Dicer is a key step in microRNA biogenesis. This process is assisted by the homologous proteins PACT and TRBP, which bind to the helicase domain of Dicer. The mechanism by which they assist microRNA biogenesis is poorly understood, but could include facilitating substrate positioning, assisting Argonaute loading, or discriminating between different classes of pre-miRNA.

PACT also regulates innate immune pathways that respond to viral double-stranded RNA, including via the kinase PKR. Mutations in PACT lead to early onset dystonia in humans, while depletion of PACT in mice results in both growth and fertility defects: both have been linked to inappropriate or altered activation of PKR. Homodimerisation of PACT via its C-terminal domain (PACT-D3) is thought to be necessary for it to induce PKR activation.

Homodimerisation of wild-type and mutant constructs of PACT-D3 was assayed using biophysical techniques. SEC-MALLS and analytical ultracentrifugation data demonstrate that PACT-D3 homodimerises via a different mechanism to a previously reported dsRBD homodimer, dsRBD-5 of Staufen1. Instead, NMR analyses show that PACT-D3 forms an asymmetric homodimer similar to that observed in the *Drosophila melanogaster* homologue Loquacious. Dimerisation could be abolished by the L273R mutation, while phospho-mimic mutations did not appear to significantly affect dimerisation. TRBP domain 3 also forms asymmetric dimers, but with weaker affinity due to sequence differences in its C-terminal α -helix. Asymmetry is caused by a register shift between intermolecular parallel α -strands, but the functional significance of asymmetric homodimerisation remains unclear.

The data presented in this thesis supports a model in which the homodimerisation interface of PACT-D3 overlaps with the surface that binds to Dicer, and suggests that PACT homodimerisation and the formation of a Dicer-PACT complex are incompatible.

Table of Contents

Abstract	2
Table of Contents	3
List of Tables	8
List of Figures	9
Acknowledgements	12
Author's Declaration	13
1. Roles of PACT and TRBP in double-stranded RNA recognition and processing of non-coding RNAs	14
1.1 Overview of the miRNA biogenesis pathway	15
1.2 Overview of antiviral responses mediated by RIG-I and PKR	17
1.2.1 RIG-I	17
1.2.2 PKR.....	17
1.3 Multi-domain structure of PACT and TRBP	20
1.3.1 TRBP and PACT domains 1 and 2 bind dsRNA.....	21
1.3.2 PACT and TRBP bind protein partners via their third domain.....	24
1.4 PACT and TRBP form part of the small RNA biogenesis pathway	26
1.4.1 How do PACT and TRBP affect processing by Dicer and the formation of active RISC?.....	26
1.4.2 Evidence from structural studies.....	28
1.4.3 Evidence from biochemical and biophysical studies.....	31
1.4.4 Evidence from molecular cell biology studies.....	33
1.5 TRBP and PACT mediate innate immune surveillance of dsRNA	35
1.5.1 PACT can stimulate RIG-I activation.....	36
1.5.2 PACT activates PKR, while TRBP inhibits it.....	37
1.5.3 Links between RNA interference and viral sensing in mammals	41
1.6 Other functions of PACT and TRBP	41
1.6.1 Fertility and development.....	41
1.6.2 Early onset dystonia-parkinsonism	42
1.7 General features and evolutionary history of PACT and TRBP homologues	43
1.7.1 Similarities between PACT homologues	45
1.7.2 Differences between PACT homologues	47
1.8 Summary	48
1.9 Aims of this study	49

2. Protein Production Methods.....	50
2.1 Protein expression in <i>E. coli</i>.....	50
2.1.1 Preparation of vectors	50
2.1.2 Transformation into <i>E. coli</i> strain BL21 (DE3).....	51
2.1.3 Site-directed mutagenesis.....	51
2.1.4 Culture and overexpression	52
2.2 Protein purification	52
2.2.1 Assessing solubility of constructs.....	52
2.2.2 Large-scale purification.....	53
2.3 SDS-PAGE	54
2.4 Media and buffer compositions	55
2.4.1 Media	55
2.4.2 Lysis and purification.....	57
2.4.3 SDS-PAGE.....	58
3. Biophysical Methods.....	60
3.1 SEC-MALLS.....	60
3.1.1 Theoretical basis for molecular weight measurement using MALLS.....	60
3.1.2 Practical aspects of data collection and analysis.....	63
3.2 Sedimentation equilibrium analytical ultracentrifugation (seAUC).....	65
3.2.1 Practical aspects of data collection and analysis.....	66
3.3 Differential Scanning Fluorimetry.....	67
4. NMR Methods	70
4.1 Introduction	70
4.2 Basic principles of NMR spectroscopy.....	70
4.2.1 Physical basis of the NMR signal	70
4.2.2 Intramolecular contributions to the magnetic field.....	71
4.2.3 Relaxation and the effects of molecular tumbling	72
4.2.4 The Nuclear Overhauser Effect	74
4.3 Overview of NMR experiments.....	75
4.3.1 (¹ H, ¹⁵ N) heteronuclear single quantum coherence (HSQC) experiment	75
4.3.2 ¹⁵ N longitudinal relaxation measurements and Exchange Spectroscopy (EXSY)	77
4.3.3 Isotope-filtered NOESY.....	78
4.4 Details of NMR data collection.....	79
4.4.1 Sample preparation.....	79
4.4.2 Spectrometers	80
4.4.3 (¹ H, ¹⁵ N) HSQC spectra	80
4.4.4 Assignment of PACT-D3 L273R	80

4.4.5	Assignment of wild-type PACT-D3	81
4.4.6	T ₁ relaxation and EXSY spectra	82
4.4.7	NOESY spectra	83
4.4.8	NMR experiments at high pressure.....	84
4.5	NMR data analysis	84
4.5.1	(¹ H, ¹⁵ N) HSQC processing.....	84
4.5.2	Fitting of longitudinal relaxation and exchange data	84
4.5.3	Fitting of pressure-induced chemical shift changes	86
5.	Cell and Molecular Biology Methods	87
5.1	Mammalian cell culture	87
5.2	Knockdown of PACT by siRNA	87
5.3	Construction of vectors for PACT overexpression	87
5.3.1	Cloning PACT constructs into pGEM-T	87
5.3.2	Construction of c-myc tag expression vectors.....	88
5.3.3	Subcloning of PACT constructs into mammalian expression vectors.....	89
5.4	PACT overexpression in mammalian cells.....	89
5.4.1	Optimisation of GeneJuice transfection.....	89
5.4.2	Overexpression in HEK293 and HeLa cells.....	89
5.5	Extraction of protein and Western blotting.....	89
5.6	Extraction of RNA and qPCR.....	91
6.	Production of recombinant PACT and TRBP domains	93
6.1	Introduction	93
6.2	Protein expression in <i>E. coli</i>	95
6.2.1	Solubility tests of protein constructs.....	95
6.2.2	Scale up and purification	96
6.3	Optimisation of size exclusion buffer for protein stability	98
6.3.1	Differential scanning fluorimetry results	98
6.4	NMR spectroscopy indicates folding state of PACT-D3 and TRBP-D3 constructs	100
7.	Biophysical Characterisation of PACT Domain 3	103
7.1	Introduction	103
7.2	Both PACT-D3 and PACT-Ext-D3 homodimerise.....	103
7.3	Sedimentation equilibrium AUC confirms that PACT-D3 is dimeric.....	105
7.4	Phospho-mimic mutations do not appear to alter homodimerisation behaviour	106
7.5	The L273R mutation disrupts PACT-D3 dimerisation	108
7.6	Discussion	111
7.6.1	The N-terminal extension is neither necessary nor sufficient for dimerisation	111

7.6.2	Implications for NMR studies	113
8.	Characterisation of PACT Domain 3 by NMR Spectroscopy.....	114
8.1	Introduction	114
8.2	Characterisation of PACT-D3 L273R by NMR.....	114
8.3	Characterisation of PACT-D3 dimers by NMR.....	118
8.3.1	Backbone assignment of PACT-D3	118
8.3.2	Investigation of phospho-mimic mutants of PACT-D3	122
8.3.3	(¹ H, ¹⁵ N) HSQC of PACT-Ext-D3.....	123
8.4	Investigations into the two states of PACT-D3.....	125
8.4.1	Two-state behaviour is maintained across a range of buffers and conditions	125
8.4.2	Chemical shift analysis suggests differences between the two states are focused on the third β -strand.....	127
8.4.3	Proline isomerisation does not explain the two states.....	128
8.4.4	The two-state behaviour is not due to disulphide bond formation	129
8.4.5	The two states do not represent different protonation states	129
8.4.6	The two states exchange at a rate inconsistent with domain swapping.....	131
8.4.7	NOESY experiments show that PACT-D3 forms asymmetric dimers.....	134
8.4.8	The third β -strands form an intermolecular parallel β -sheet.....	137
8.4.9	Loqs-D3 also forms asymmetric dimers in solution	139
8.5	Evidence for the appearance of a monomeric state at low concentrations	141
8.6	NMR spectroscopy of PACT-D3 at high pressure	143
8.7	Discussion	146
8.7.1	State-specific assignment of PACT-D3	146
8.7.2	Phosphorylation of PACT-D3	147
9.	Biophysical Characterisation of TRBP Domain 3	150
9.1	Introduction	150
9.2	NMR spectra of TRBP-Ext _{DD} -D3 show fewer peaks than expected.....	150
9.3	SEC-MALLS shows that TRBP-Ext-D3 is in equilibrium between monomer and dimer	151
9.4	Mutations in TRBP-Ext-D3 can inhibit or enhance dimerisation	153
9.5	NMR suggests that TRBP-Ext _{DD} -D3 RR353,354HN forms asymmetric dimers.....	155
9.6	Discussion	157
9.6.1	Factors affecting the stability of TRBP domain 3	157
9.6.2	Possible causes for intermediate exchange in TRBP-Ext _{DD} -D3 NMR spectra	157
10.	Development of Tools for Characterisation of PACT Dimerisation in Mammalian Cells.....	159

10.1	Introduction	159
10.2	Knockdown of PACT by siRNA	159
10.3	Overexpression of PACT	161
10.4	Future directions	166
11.	Conclusions and Discussion	168
11.1	Summary of results from this thesis	168
11.1.1	PACT-D3 forms homodimers	168
11.1.2	PACT-D3 homodimers are asymmetric, and similar in structure to Loqs-D3 dimers	168
11.1.3	TRBP domain 3 homodimerises with lower affinity than PACT domain 3	169
11.2	Discussion	169
11.2.1	Incompatibility of PACT homodimerisation and Dicer/PACT complex formation	169
11.2.2	Homo- and heterodimerisation of dsRBDs	173
11.2.3	Asymmetric dimerisation of PACT	175
11.2.4	Suggestions for future research	182
Appendix 1: Sequences of PACT, TRBP and Loqs constructs		184
Appendix 2: Analytical Ultracentrifugation raw data		187
Appendix 3: Resonance assignments of PACT-D3 and PACT-D3 L273R		189
Appendix 4: R code for simultaneous fitting of longitudinal relaxation and exchange rates		205
Appendix 5: Dissecting the roles of TRBP and PACT in double-stranded RNA recognition and processing of noncoding RNAs		215
Abbreviations		235
References		236

List of Tables

Table 1.1 Summary of viruses linked to TRBP or PACT.....	40
Table 1.2. PACT and TRBP homologue sequences	46
Table 2.1. UV extinction coefficients of PACT and TRBP constructs	53
Table 4.1. NMR spectrometer details.....	80
Table 4.2. Experimental parameters for assignment of PACT-D3 L273R.....	81
Table 4.3. Experimental parameters for assignment of PACT-D3.....	82
Table 4.4. Experimental parameters for NOESY experiments on PACT-D3.....	83
Table 5.1. Primers used for cloning PACT into pGEM-T vector.....	88
Table 5.2. Oligonucleotides used in construction of c-myc expression vector	88
Table 5.3. Antibodies used for Western blotting	91
Table 5.4. Primers used for qRT-PCR.....	92
Table 6.1. PACT and TRBP Domain 3 Constructs.....	93
Table 6.2. Overview of protein expression systems.....	94

List of Figures

Figure 1.1 Overview of miRNA biogenesis	16
Figure 1.2. Overview of RIG-I and PKR antiviral pathways.....	19
Figure 1.3. Domain structure and interactions of PACT and TRBP	21
Figure 1.4 dsRBD sequence and structure in PACT and TRBP	22
Figure 1.5 Potential roles of PACT and TRBP in the Dicer complex.....	27
Figure 1.6. Structure of the Dicer complex.....	30
Figure 1.7 PACT-mediated activation of RIG-I.....	36
Figure 1.8. Regulation of PKR by PACT and TRBP.....	39
Figure 1.9. Phylogenetic analysis of PACT, TRBP and Loqs homologues.....	45
Figure 2.1. Schematic of the pETFPP expression vector.....	51
Figure 3.1. Schematic of the Wyatt Dawn HELEOS-II light scattering detector.	64
Figure 3.2. Example differential scanning fluorimetry curve.....	69
Figure 4.1 Relaxation rates depend on rotational correlation time	74
Figure 4.2. Overview of NMR pulse sequences	76
Figure 6.1. Solubility tests of PACT and TRBP domain 3 constructs.....	95
Figure 6.2. Large scale purification of PACT and TRBP domains	97
Figure 6.3. Differential Scanning Fluorimetry (DSF) to optimise buffer for protein stability.....	98
Figure 6.4. 1D ¹ H spectra of [¹⁵ N]-labelled purified protein domains.....	101
Figure 7.1. SEC-MALLS of wild-type PACT-D3 and PACT-Ext-D3.....	104
Figure 7.2. Diagnostic plots of sedimentation equilibrium AUC of wild-type PACT-D3.....	105
Figure 7.3. SEC-MALLS of PACT-D3 with phospho-mimic mutations	107
Figure 7.4. SEC-MALLS of PACT-D3 and PACT-Ext-D3 mutants	110

Figure 7.5. SEC-MALLS of Loqs-D3a and L426R mutants	111
Figure 7.6. Positions of mutated residues with reference to Loqs-D3	112
Figure 8.1. Assigned (¹ H, ¹⁵ N) HSQC of PACT-D3 L273R.....	115
Figure 8.2. TALOS-N analysis of PACT-D3 L273R	116
Figure 8.3. (¹ H, ¹⁵ N) HSQC of PACT-Ext-D3 L273R	117
Figure 8.4. Assigned (¹ H, ¹⁵ N) SOFAST-HMQC of PACT-D3	119
Figure 8.5. TALOS-N analysis of PACT-D3	121
Figure 8.6. (¹ H, ¹⁵ N) HSQC spectra of PACT-D3 phosphomimic mutants.....	123
Figure 8.7. (¹ H, ¹⁵ N) HSQC of PACT-Ext-D3	124
Figure 8.8. Distribution of HSQC peak heights under different conditions	126
Figure 8.9. Chemical shift differences between the two states of PACT-D3	128
Figure 8.10. NMR spectra of PACT-D3 proline, cysteine and histidine residues ..	130
Figure 8.11. Chemical exchange in PACT-D3	132
Figure 8.12. ¹³ C -filtered NOESY spectra of PACT-D3	135
Figure 8.13. Intermolecular NOEs identified by ¹³ C-filtered NOESY.....	136
Figure 8.14. NOESY spectra provide evidence for asymmetry in strand β-3	139
Figure 8.15. (¹ H, ¹⁵ N) HSQC spectra of Loqs-D3 and Loqs-D3 L426R.....	140
Figure 8.16. (¹ H, ¹⁵ N) SOFAST-HMQC of low concentration PACT-D3	142
Figure 8.17. NMR of PACT-D3 and PACT-D3 L273R at high pressure.....	144
Figure 9.1. (¹ H, ¹⁵ N) HSQC of TRBP-Ext _{DD} -D3.....	151
Figure 9.2. SEC-MALLS of TRBP-Ext-D3 and TRBP-Ext _{DD} -D3	152
Figure 9.3. SEC-MALLS of TRBP-Ext _{DD} -D3 L326R and RR353,354HN	155
Figure 9.4. (¹ H, ¹⁵ N) HSQC spectra of TRBP-Ext _{DD} -D3 L326R and RR353,354HN	156
Figure 10.1. qRT-PCR of miRNA-related genes after treatment with siPACT	160
Figure 10.2. Test of siRNA targeting PACT	161
Figure 10.3. Major elements of the pSF-CMV-Puro-NH2-FLAG plasmid	162
Figure 10.4. Optimisation of transfection in HEK293 cells	163

Figure 10.5. Stable overexpression of FLAG- and myc-tagged PACT constructs in HEK293 cells.....	165
Figure 11.1. Overlap of the PACT-D3 homodimerisation and Dicer-binding interfaces.....	170
Figure 11.2. Comparison of the Dicer/TRBP and Loqs homodimer interfaces	172
Figure 11.3. Asymmetry in Loqs-D3 homodimers	176

Acknowledgements

First of all, I'd like to thank my supervisors Dr Michael Plevin and Dr Dimitris Lagos for all their discussions, support and advice over the past 4 years, especially at the start of my PhD and during my period of part-time study. I'm also grateful to the members of my Training Advisory Panel, Prof Fred Antson and Prof Peter McGlynn, for their suggestions and for keeping me on track, and to our collaborators Nicky Baxter, Mike Williamson, Leonhard Jakob and Gunter Meister for productive discussions, and for sharing their expertise and samples.

I would not have been able to complete this PhD without the generous support and friendship of many members of the Plevin, Lagos and Potts groups, including Banushan Balansethupathy, Mike Hodgkinson, Claire Coupland, Dr James Hewitson and Daniel Yee. I'd especially like to thank Dr Fiona Whelan, Dr Andrew Brentnall and Judith Hawkhead for teaching me many of the experimental techniques I relied on in my research, and for finding time to answer my many questions. I'm indebted to Dr Pedro Aguire, Dr Geoff Kelly and Dr Alain Oregioni for assistance with NMR data collection and interpretation, and to Dr Andrew Leech and Shirley Williams for assistance with various biophysical techniques. I'd also like to thank Julie Knox for her help in accessing the University's support mechanisms.

I'd like to thank the Wellcome Trust for providing funding for this project through the CIDCATS doctoral training programme, the MRC Biomedical NMR Facility for providing access to their spectrometers, and the Instruct programme for funding several visits to the Institut de Biologie Structurale in Grenoble.

Finally, I want to express my enormous gratitude to my husband Kit for his patience, understanding and support throughout my PhD, and especially during the writing of this thesis.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author (except as detailed below). This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1 (“Roles of TRBP and PACT in double-stranded RNA recognition and processing of non-coding RNAs”) is modified from a review article co-authored by myself, Dr Dimitris Lagos and Dr Michael Plevin, and published in Wiley Interdisciplinary Reviews RNA (Heyam, Lagos and Plevin, 2015). I wrote the initial text, while my supervisors suggested modifications and rewrote some sections. The version published as Chapter 1 has been updated to include recently published studies, and to include additional material. The published review article is included as Appendix 5.

Roles of PACT and TRBP in double-stranded RNA recognition and processing of non-coding RNAs

Double-stranded RNA (dsRNA) is a vital part of normal cellular function. Any structured RNA, such as the ribosome, contains regions of dsRNA, which are often part of higher order tertiary structure. In addition, dsRNA is an intermediate in the biogenesis of short regulatory RNA such as microRNAs (miRNAs), endogenous short interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs) (Kim, Han and Siomi, 2009).

However, dsRNA is also an important replication intermediate for RNA viruses, such as SARS coronavirus, poliovirus and hepatitis C. It is crucial for cells to distinguish cellular dsRNA from viral dsRNA and to respond appropriately. In mammals, this is achieved by proteins that have evolved to recognise chemical features specific to viral dsRNA, such as terminal 5' triphosphate groups. These proteins are termed pattern recognition receptors, and include TLRs (Toll-like receptors) 3, 7 and 8, RIG-I (Retinoic acid-inducible gene I), and MDA5 (Melanoma differentiation associated protein 5) (Gantier and Williams, 2007).

The closely related mammalian proteins PACT (Protein Activator of PKR; PRKRA) and TRBP (HIV Trans-activation responsive RNA binding protein; HGNC symbol, TARBP2) bridge several of these pathways. TRBP was initially identified through its interaction with the HIV TAR element, and both proteins were found to regulate the response to viral dsRNA through the protein PKR (double-stranded RNA-activated protein kinase; EIF2AK2) (Park *et al.*, 1994; Patel and Sen, 1998). Subsequently, PACT and TRBP were shown to interact with Dicer (DICER1), the ribonuclease responsible for processing the precursors of miRNAs and siRNAs (Chendrimada *et al.*, 2005; Haase *et al.*, 2005). More recently, PACT has been shown to activate RIG-I, another innate immune sensor of viral RNA (Kok *et al.*, 2011). However, these different roles are often studied independently, and the links between them remain largely unexplored.

I will begin by providing brief overviews of miRNA biogenesis and the innate immune pathways which detect and respond to dsRNA in mammalian somatic

cells. I will then discuss PACT and TRBP in more detail, including their structure, their roles in miRNA and antiviral pathways, and their evolutionary history.

1.1 Overview of the miRNA biogenesis pathway

RNA interference (RNAi) is a highly conserved process in which small RNA molecules cause post-transcriptional silencing of complementary messenger RNAs (Figure 1.1) (Malone and Hannon, 2009; Liu and Paroo, 2010). Somatic cells contain 2 major classes of small regulatory RNA: miRNA and siRNA (Kim, Han and Siomi, 2009). Both are approximately 22 nt long, but differ in their biogenesis and target recognition. miRNAs are derived from RNA hairpins which usually contain several mismatched bases, while siRNAs are processed from longer dsRNA molecules. siRNAs are typically highly complementary to their targets, whereas miRNAs only require a 6-8 nt 'seed' region to be perfectly matched, with other positions contributing only weakly to target specificity (Brennecke *et al.*, 2005; Grimson *et al.*, 2007). Recent crystal structures of Argonaute-RNA complexes show that the seed region of the miRNA or siRNA is initially presented for binding, and that supplemental binding at other positions (as in siRNAs) favours cleavage of the target RNA (Schirle, Sheu-Gruttaduria and MacRae, 2014). Although there are relatively few reports of endogenous mammalian siRNAs, dsRNA can be introduced exogenously either as RNA hairpins (shRNA) or short siRNA duplexes (Silva *et al.*, 2005).

In the canonical miRNA biogenesis pathway, miRNAs are transcribed as long primary (pri-) miRNAs that contain a 60-80 nucleotide hairpin structure (Kim, 2005). The pri-miRNA is processed in the nucleus by Drosha/DGCR8 (the 'microprocessor' complex) to a shorter precursor (pre-) miRNA (Lee *et al.*, 2003; Han, 2004). The pre-miRNA is exported from the nucleus by Exportin 5, and further cleaved by a protein complex containing the ribonuclease Dicer, an Argonaute protein, and either TRBP or PACT (Bernstein *et al.*, 2001; Hutvagner, 2001; Chendrimada *et al.*, 2005; Haase *et al.*, 2005; Lee *et al.*, 2006). This second processing step removes the terminal hairpin loop, leaving an RNA duplex ~22 bp in length.

One strand of the dsRNA Dicer product is removed, while the other (the guide strand) is loaded into one of the four Argonaute proteins (Ago1-4) (Martinez *et al.*, 2002; Liu, 2004). This Ago:miRNA complex, together with its protein partners, is

referred to as the RNA induced silencing complex (RISC). Despite recent progress, the mechanism of strand-selective RISC loading remains to be determined (Gredell *et al.*, 2010; Noland and Doudna, 2013); however, many miRNAs can contribute either strand to RISC (Yang *et al.*, 2011).

Once loaded, the guide strand targets RISC to complementary mRNAs, which results in either suppression of translation, enhanced degradation (via decapping and polyA tail removal), and/or site-specific cleavage of the mRNA (Jonas and Izaurralde, 2015). In cultured mammalian cells at steady state, degradation via the 5'-to-3' mRNA decay pathway is most significant (Guo *et al.*, 2010; Eichhorn *et al.*, 2014). Target mRNAs are first deadenylated by the PAN2-PAN3 or CCR4-NOT complexes, decapped by DCP2, and finally degraded by XRN1 (Behm-Ansmant, 2006; Wahle and Winkler, 2013). These processes are coordinated by the scaffold proteins TNRC6A-C, which link the deadenylation complexes to Argonaute (Liu *et al.*, 2005; Eulalio, Tritschler and Izaurralde, 2009; Pfaff and Meister, 2013). Translational repression is less well understood, but was found to account for at most a quarter of miRNA-mediated silencing (Eichhorn *et al.*, 2014). Although only Ago2 has the ability to cleave mRNAs, miRNAs and siRNAs appear to distribute among all 4 Argonautes with little specificity (Meister *et al.*, 2004; Burroughs *et al.*, 2011).

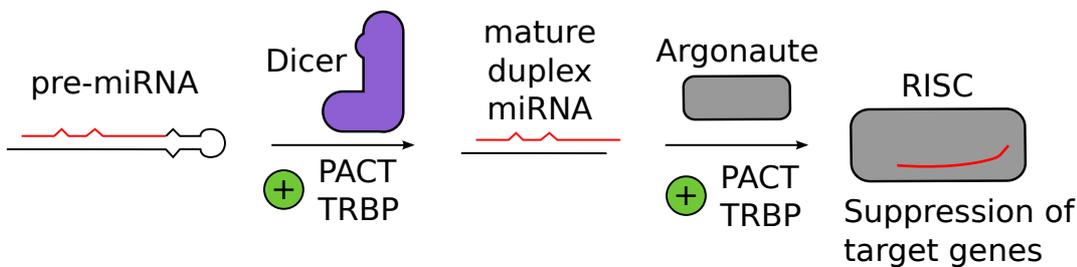


Figure 1.1 Overview of miRNA biogenesis

Precursor (pre-)miRNAs are RNA hairpins that are produced in the nucleus, and exported to the cytoplasm. They contain the ~22 nt sequence of the mature miRNA, indicated in red. The endonuclease Dicer removes the terminal loop to give an RNA duplex, one strand of which is loaded into an ArgonAUT (Ago) protein to form RISC. TRBP and PACT are implicated in both Dicing and RISC loading, as discussed in section 1.4.

1.2 Overview of antiviral responses mediated by RIG-I and PKR

1.2.1 RIG-I

RIG-I is an early response pattern recognition receptor that recognises dsRNA with a 5' triphosphate group (5'ppp-dsRNA) (Yoneyama *et al.*, 2005; Takeuchi and Akira, 2008; Schlee and Hartmann, 2010). RIG-I is ubiquitously expressed, and is one of the main viral RNA sensors in non-immune cells (Gantier and Williams, 2007). It is part of a family of related cytoplasmic viral sensors alongside MDA-5 and LGP2, which act cooperatively to recognise longer dsRNA molecules (>200 bp) (Peisley *et al.*, 2011; Bruns *et al.*, 2014).

The mechanism of RIG-I activation has been characterised at the structural level (Figure 1.2B) (Kowalinski *et al.*, 2011; Luo *et al.*, 2011; Kolakofsky, Kowalinski and Cusack, 2012). RIG-I consists of two caspase recruitment domains (CARDs), a DExD/H helicase domain and a C-terminal domain (CTD). In the absence of 5'ppp-dsRNA, the tandem CARDs are bound to the helicase domain, and are inactive. If 5'ppp-dsRNA is present, RIG-I binds using both its helicase and C-terminal domain, displacing the CARDs in an ATP-dependent manner. The CARDs are then poly-ubiquitinated by TRIM25, which encourages RIG-I multimerisation and interaction with the Mitochondrial Antiviral Signalling protein (MAVS) at the mitochondrial membrane (Meylan *et al.*, 2005; Seth *et al.*, 2005; Gack *et al.*, 2007; Jiang *et al.*, 2012). This stimulates the production of type I interferons and pro-inflammatory cytokines via IRF3/7 and NF- κ B, alerting nearby cells and the adaptive immune system of the viral threat (Figure 1.2A). Interferon can also act on the original cell, upregulating the production of antiviral effectors such as PKR (Gantier and Williams, 2007).

1.2.2 PKR

PKR is a kinase that acts to block translation in response to viral dsRNA, or to other cellular stresses such as oxidative stress, accumulation of misfolded proteins in the ER, or external signals from cytokines and growth factors (Donnelly *et al.*, 2013). It has low basal expression, but is strongly induced by interferon, making it a 'late' responder to viruses (Figure 1.2A) (Meurs *et al.*, 1990).

PKR has two N-terminal type A double-stranded RNA-binding domains (dsRBDs, discussed further in Section 1.3), and a kinase domain (Figure 1.2C). Related

kinase domains are found in other proteins that regulate translation in response to various cellular stresses, including PERK, GCN2 and HRI (Donnelly *et al.*, 2013). Initial studies suggested that the dsRBDs exerted an autoinhibitory effect on the kinase in the absence of dsRNA (Wu and Kaufman, 1997). However, this model was unable to account for inhibition of PKR upon transfection of very high dsRNA loads, and it is now accepted that dimerisation of the kinase domain is the trigger for activation (Ung *et al.*, 2001; Vatter, Staschke and Wek, 2001). The first two domains bind dsRNA and act as a scaffold to bring PKR molecules close together, and increase the likelihood of dimerization of the kinase domain (Cole, 2007).

After dimerization, the kinase domain autophosphorylates, dissociates from dsRNA and phosphorylates eIF2 α (eukaryotic translation initiation factor 2 α) (Vatter, Staschke and Wek, 2001). This prevents eIF2 from passing Met-tRNA to the ribosome, thereby inhibiting translation (Kimball, 1999). PKR also directly or indirectly phosphorylates several other targets, including pro-apoptotic p53, and the STAT1 transcription factor (Cuddihy *et al.*, 1999; Ramana *et al.*, 2000).

Although eIF2 α phosphorylation leads to general translational repression, a number of transcription factors are specifically upregulated, including ATF4 (Harding *et al.*, 2000). ATF4 controls amino acid metabolism and redox homeostasis, and promotes expression of the pro-apoptotic factors BIM and DR5, while downregulating anti-apoptotic BCL-2 (McCullough *et al.*, 2001; Yamaguchi, 2004; Puthalakath *et al.*, 2007). Through this mechanism, prolonged activation of PKR leads to apoptosis.

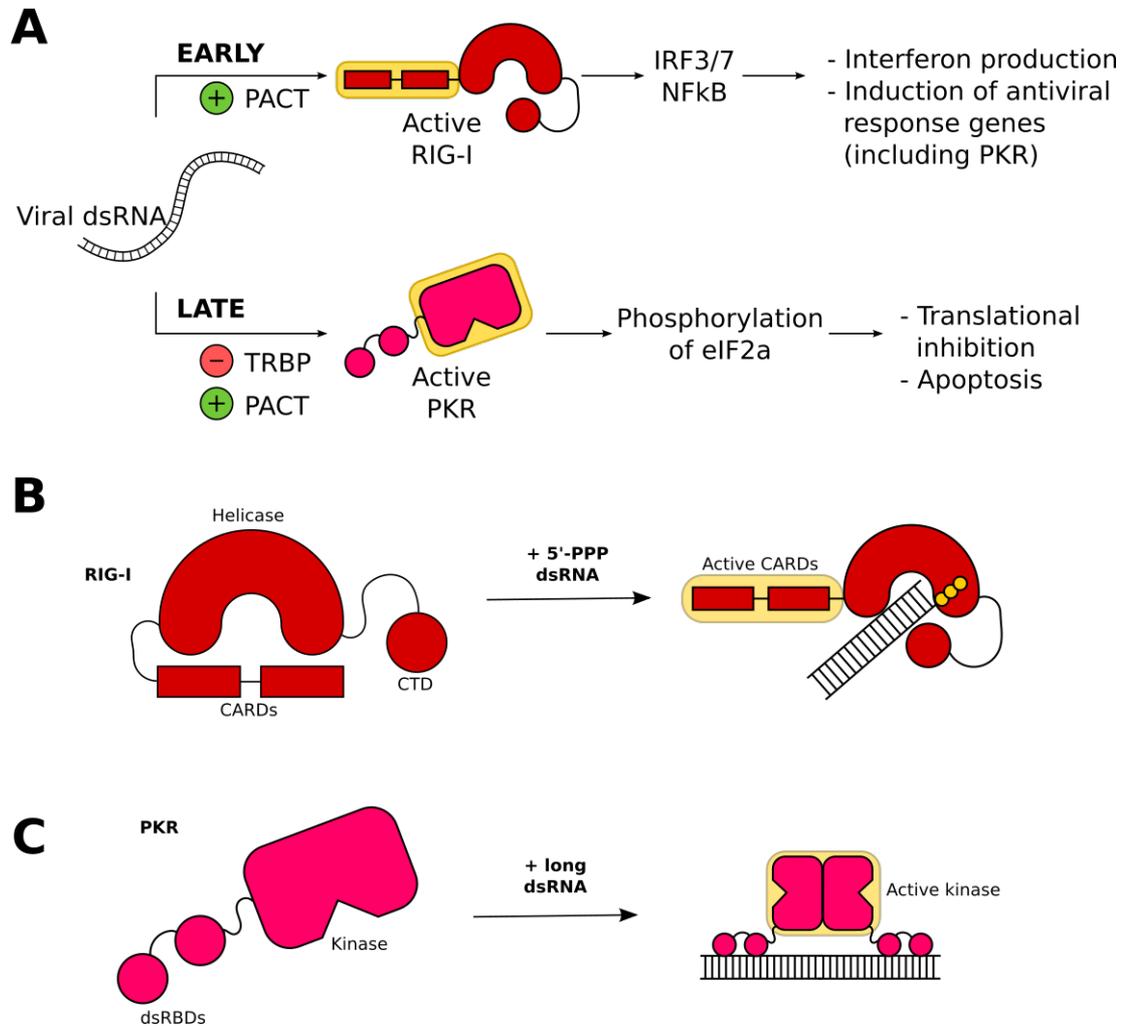


Figure 1.2. Overview of RIG-I and PKR antiviral pathways

A) PACT and TRBP have roles in at least two viral response pathways. Firstly, PACT can facilitate activation of RIG-I by viral dsRNA (distinguished from cellular dsRNA by distinct molecular features, discussed in section 1.5). This begins a signalling cascade that results in the production of interferon and other antiviral genes. PACT and TRBP also regulate PKR, a kinase which targets the translation initiation factor eIF2 α to inhibit protein production and promote apoptosis. B) In the absence of appropriate ligands, RIG-I has an inactive conformation, in which the helicase domain binds to the CARDs. 5'-triphosphate dsRNA binds to the CTD and helicase domain, which displaces the CARDs and results in signalling. C) PKR can bind to long dsRNA through two N-terminal dsRBDs. This brings the kinase domains together to form dimers, which can then autophosphorylate and become active.

1.3 Multi-domain structure of PACT and TRBP

PACT and TRBP each contain three double-stranded RNA binding domains (dsRBDs) that are separated by linker regions which are predicted to be unstructured (St Johnston *et al.*, 1992; Peters *et al.*, 2001) (Figure 1.3). Many RNA binding proteins contain multiple dsRBDs: PKR, ADAR2 and DGCR8 all have tandem dsRBDs, while human Staufen has five (Nanduri *et al.*, 1998; Marión *et al.*, 1999; Stefl *et al.*, 2006; Sohn *et al.*, 2007). The presence of multiple dsRBDs in the same protein is thought to allow greater affinity and specificity, and to allow functional divergence of individual domains (Chang and Ramos, 2005; Lunde, Moore and Varani, 2007).

Although all dsRBDs share the core α - β - β - β - α fold (Figure 1.4B), they can be divided into two subgroups depending on sequence conservation. The type A dsRBD is the canonical form, which shows amino acid conservation in three regions involved in dsRNA binding (Figure 1.4A). Type B dsRBDs only show conservation at the C-terminal end of the domain, and are generally unable to bind dsRNA, despite retaining the same overall three-dimensional (3D) structure (St Johnston *et al.*, 1992; Masliah, Barraud and Allain, 2013). PACT and TRBP contain both classes of dsRBD (Figure 1.4A).

Some dsRBDs have additional structural elements that contribute to their function: for example, Rnt1p dsRBD from *Saccharomyces cerevisiae* has an additional α -helix that contributes to domain stability and RNA binding (Wu *et al.*, 2004), while dsRBDs from the *Caenorhabditis elegans* TRBP homologue, RDE-4, have recently been shown to contain numerous additional helices and extended loops regions (Chiliveri and Deshmukh, 2014).

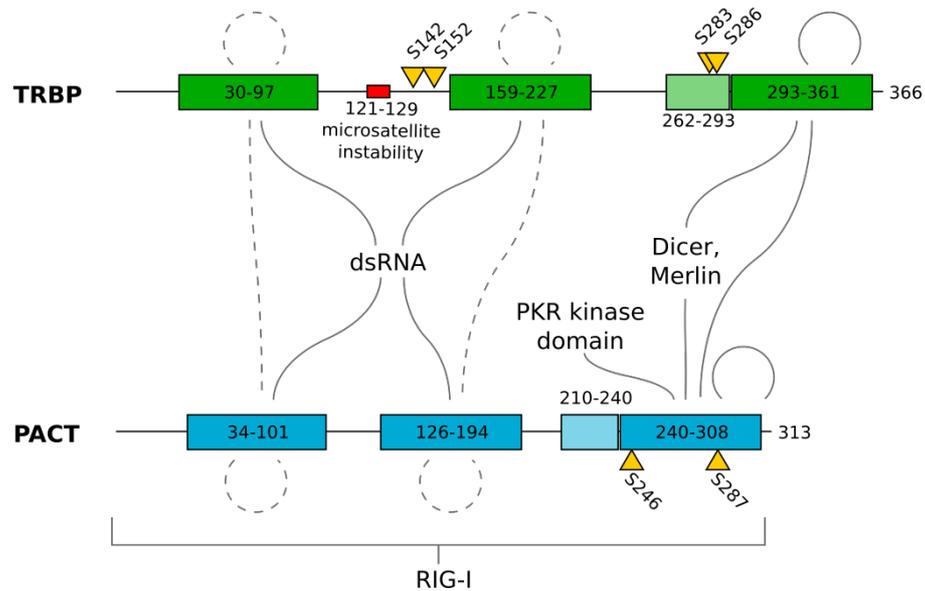


Figure 1.3. Domain structure and interactions of PACT and TRBP

Both PACT and TRBP contain 3 double-stranded RNA binding domains (dsRBDs). The first two domains can bind dsRNA, while the third cannot. The interdomain linkers are longer in TRBP than in PACT. Interactions with many other proteins have been documented, particularly for the third domain. Solid lines indicate direct protein/protein interactions, while dashed lines indicate interactions that may be mediated via dsRNA. It is not known which domain of PACT interacts with RIG-I, shown as a solid bracket. The third dsRBD potentially has an N-terminal extension, based on sequence conservation. Each protein has a number of phosphorylation sites (yellow triangles) that regulate function under certain conditions. A region of TRBP implicated in cancers exhibiting microsatellite instability is indicated in red.

1.3.1 TRBP and PACT domains 1 and 2 bind dsRNA

Domains 1 and 2 of both PACT and TRBP are type A dsRBDs, and all bind dsRNA (Benoit *et al.*, 2013; Takahashi *et al.*, 2013). In type A dsRBDs, three distinct regions participate in dsRNA binding (Figure 1.4A,B). Unlike double-stranded DNA, dsRNA adopts an A form helix in which the major groove is deep and narrow, limiting access to the bases and therefore to sequence specific information. dsRBDs bind across two adjacent minor grooves and the intervening phosphate backbone, burying $\sim 780\text{\AA}^2$ of the domain surface (Yang *et al.*, 2010). Generally dsRBDs bind non-specifically to dsRNA. On discovery, TRBP was thought to specifically recognise certain structured RNAs, such as HIV-1 trans-activation responsive (TAR) RNA (Gatignol *et al.*, 1991; Dorin *et al.*, 2003). However, later

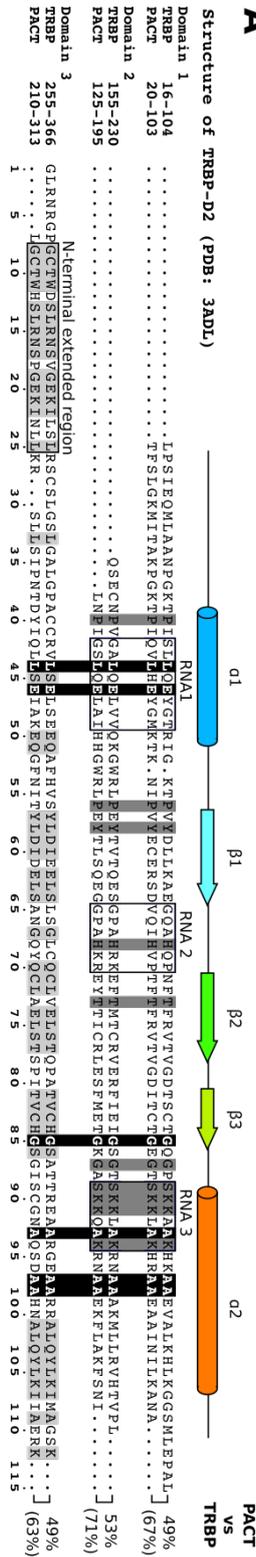
evidence has shown that TRBP and PACT interact with a broad range of targets (Yamashita *et al.*, 2011; Benoit *et al.*, 2013).

Several well-studied dsRBD-containing proteins do act on specific targets. For example, Staufen can regulate translation and decay of certain mRNAs, while ADAR proteins can convert adenosine to inosine at specific positions in a variety of dsRNAs (Bass, 2002; Ricci *et al.*, 2014). In both cases, RNA recognition is believed to occur through tertiary structures, such as bulges and loops (Ramos *et al.*, 2000; Wu *et al.*, 2004). Some specificity may also arise through contact with the edges of bases in the minor groove, and through the combination of dsRBDs with weak sequence or secondary structure preferences (Lunde, Moore and Varani, 2007; Masliah, Barraud and Allain, 2013). Current evidence does not rule out PACT and TRBP having a degree of substrate specificity.

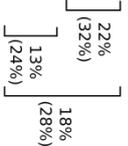
Figure 1.4 dsRBD sequence and structure in PACT and TRBP

A) Sequence alignment of the dsRBDs of human TRBP and PACT. The top line shows the secondary structure of a 'typical' dsRBD, taken from the 3D structure of TRBP domain 2 (PDB accession: 3ADL). Residues conserved between all domains are highlighted in black; those conserved between domains 1 and 2 are highlighted in dark grey; while those conserved in the third C-terminal domain are shown in light grey. The regions of domain 1 and 2 which bind RNA (located in helix $\alpha 1$, the loop between β strands 1 and 2 (loop- $\beta 12$), and helix $\alpha 2$) are boxed, as is a conserved region upstream of domain 3. On the right, the % identity (% similarity) shows that equivalent dsRBDs between TRBP and PACT are more similar than dsRBDs within the same protein. The sequences were aligned using Multalin (Corpet, 1988), and rendered using ESPript (Gouet, 2003) (<http://espript.ibcp.fr>). B) Two views of TRBP-D2 bound to two molecules of 10 bp dsRNA (PDB accession 3ADL). The RNA-interacting regions shown in part A are highlighted. Several parts of the dsRNA binding interface are flexible, most notably loop- $\beta 12$ in the second RNA-interacting region, which contains a highly conserved histidine residue. C) Structure of a Staufen-D5 dimer (PDB accession 4DKK). The N-terminal extension (consisting of two α -helices) is thought to interact with the dsRBD core of a second molecule. The linker between the dsRBD core and the extended region is not visible in the crystal structure, and is indicated here with dashed lines. D) Structure of a Loqs-D3 dimer (PDB accession 4X8W).

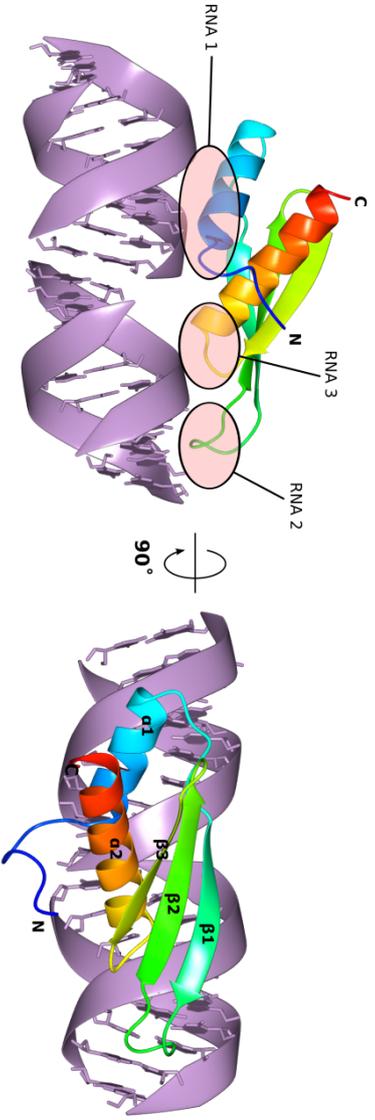
A



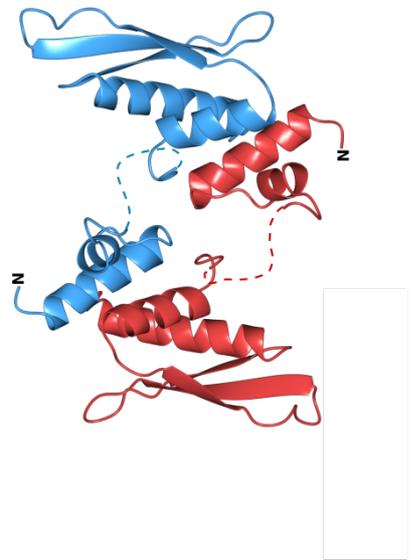
PACT domain comparison



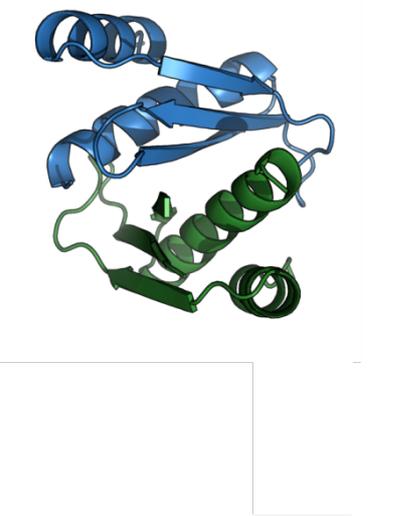
B



C



D



1.3.2 PACT and TRBP bind protein partners via their third domain

The C-terminal dsRBDs of both PACT and TRBP (hereafter referred to as PACT-D3 or TRBP-D3) are type B dsRBDs. They do not bind RNA, but are instead required for interaction with other proteins. It is thought that many type B dsRBDs mediate protein-protein interactions: for example, *Drosophila* Staufen domain 5 binds to Miranda to enable mRNA localisation (Gleghorn *et al.*, 2013; Gleghorn and Maquat, 2014). Structural information about how dsRBDs interact with other proteins has only recently become available, and is discussed further below and in section 1.4.2 (Wilson *et al.*, 2015; Jakob *et al.*, 2016).

Both PACT-D3 and TRBP-D3 bind to Dicer (Chendrimada *et al.*, 2005; Haase *et al.*, 2005; MacRae *et al.*, 2008; Daniels *et al.*, 2009; Wilson *et al.*, 2015), while PACT-D3 has also been shown to interact with PKR (Peters *et al.*, 2001; Huang, Hutchins and Patel, 2002). TRBP-D3 is not thought to interact with PKR (Gupta, Huang and Patel, 2003), suggesting that binding to Dicer and PKR are mediated by different regions of PACT-D3. These interactions are discussed further in sections 1.4 and 1.5 below. Other binding partners have been suggested for TRBP and PACT, such as the tumour suppressor Merlin (Lee *et al.*, 2004), but the nature of their interactions has not yet been studied in detail.

Although several dsRBDs have been reported to dimerise (for example, the dsRBDs from PKR (Patel *et al.*, 1995) and ADAR2 (Poulsen *et al.*, 2006)), it has been difficult to distinguish between direct binding and indirect association via dsRNA (Zhang *et al.*, 2001). There is substantial biochemical evidence that PACT-D3 can homodimerise, or heterodimerise with TRBP-D3 (Laraki *et al.*, 2008; Singh *et al.*, 2011; Singh and Patel, 2012). While there is no direct biophysical evidence that TRBP-D3 homodimerises, full length TRBP can form dimers (Yamashita *et al.*, 2011), but a construct lacking TRBP-D3 does not dimerise (Benoit *et al.*, 2013). It has been suggested that dimerisation of PACT and TRBP could modulate activation of PKR (Singh and Patel, 2012), but it could also increase their avidity for dsRNA, by bringing together four type A dsRBDs (Takahashi *et al.*, 2013).

The 3D structures of domain 5 of Staufen (Staufen-D5) and domain 3 of Loquacious (Loqs-D3) have given the first high-resolution information about type-B dsRBD dimers (Gleghorn *et al.*, 2013; Jakob *et al.*, 2016). These reveal that the overall fold is not substantially different from type A dsRBDs, but also that the

charge on the canonical dsRNA binding surface is different, with type-B dsRBDs being more negatively charged. The two structures demonstrate different dimerisation mechanisms. Staufen-D5 includes a region of extra-dsRBD structure, which interacts with the two α -helices of the dsRBD core, and is required for dimerisation. As the linker between the core domain and the extended region is flexible, dimerization can occur via a domain-swapping mechanism (Figure 1.4C) (Gleghorn *et al.*, 2013). In contrast, Loqs-D3 dimerises via an interface on the dsRBD itself, including the β -sheet and second α -helix (Figure 1.4D) (Jakob *et al.*, 2016). Loqs-D3 dimerisation will be discussed in more detail in Chapters 8 and 11. PACT-D3 and TRBP-D3 both lack the key histidine residue and have negatively charged residues at sites that would correspond to the dsRNA binding surface. In addition, both domains are preceded by a highly conserved 20 amino acid sequence, which could be consistent with a similar dsRBD/dsRBD interaction as that seen in Staufen-D5. On the other hand, neither the N-terminal extension nor helix α -1 are well conserved with Staufen-D5, while the β -sheet and helix α -2 are highly conserved with Loqs-D3 (discussed further in section 1.7).

There is some evidence that the type A dsRBDs of PACT and TRBP can also mediate protein-protein interactions. Domains 1 and 2 of PACT are reported to directly interact with the dsRBDs of PKR (Daher *et al.*, 2001; Peters *et al.*, 2001). PKR also dimerises partly through its dsRBDs (Patel *et al.*, 1995), and PACT can still activate PKR if its first two domains are replaced with those from PKR (Peters *et al.*, 2001). The first two dsRBDs of PACT and TRBP have also been reported to homo- and heterodimerise (Laraki *et al.*, 2008; Singh and Patel, 2012), although we have seen no evidence of this using recombinant proteins in the absence of dsRNA (Benoit *et al.*, 2013; Heyam *et al.*, (in preparation)). Overall, current evidence is consistent with these reported interactions between type A dsRBDs being indirect and mediated by dsRNA.

1.4 PACT and TRBP form part of the small RNA biogenesis pathway

1.4.1 How do PACT and TRBP affect processing by Dicer and the formation of active RISC?

There is now clear evidence that PACT and TRBP can associate with Dicer, both from immunoprecipitation experiments (Chendrimada *et al.*, 2005; Gregory *et al.*, 2005; Haase *et al.*, 2005; Lee *et al.*, 2006; Daniels *et al.*, 2009) and from *in vitro* reconstitution of the complex from the individually purified components (Macrae *et al.*, 2006; Chakravarthy *et al.*, 2010). Furthermore, there is strong support for the idea that PACT and TRBP are involved in the processing of pre-miRNAs and RISC-loading (Chendrimada *et al.*, 2005; Lee *et al.*, 2006). However, the molecular mechanism by which PACT and TRBP facilitate miRNA and siRNA processing remains unclear. It has long been established that Dicer alone is able to cleave pre-miRNA to mature miRNA *in vitro* (Provost, 2002), and there is evidence that some substrates can be loaded into Ago2 *in vitro* in the absence of PACT and TRBP (Kawamata, Seitz and Tomari, 2009; Yoda *et al.*, 2010; Noland and Doudna, 2013).

The importance of tackling this question is underscored by the evidence linking miRNA biogenesis in general, and TRBP in particular, to a number of different cancers (Gregory and Shiekhattar, 2005; Melo *et al.*, 2009; Fu *et al.*, 2010; Caramuta *et al.*, 2013; Lin *et al.*, 2014). In particular, tumours exhibiting microsatellite instability are prone to frameshift mutations in TRBP (see Figure 1.3), leading to impaired miRNA production (Melo *et al.*, 2009). The drug enoxacin has recently been investigated as a cancer therapeutic due to its effect on miRNA biogenesis, and is believed to act through TRBP (Melo *et al.*, 2011; Sousa *et al.*, 2013). Although several studies have linked PACT expression to cancer, the effect is not as well documented as for TRBP (Chiosea *et al.*, 2008; Sand *et al.*, 2012).

There are several points along the miRNA biogenesis and RISC loading pathway at which TRBP and PACT could act (Figure 1.5):

1) Recruitment of substrates to Dicer

- a) PACT and/or TRBP may increase the affinity of the Dicer complex for dsRNA substrates, or

b) more specifically increase Dicing rate through favouring binding of substrates over products

2) Facilitating efficient substrate cleavage by Dicer

PACT and/or TRBP could ensure that the substrate dsRNA or pre-miRNA has the optimum orientation for Dicer cleavage.

3) Removing the Dicer product

a) *PACT and/or TRBP may aid unloading of Dicer or loading of Dicer products into Argonaute proteins,*

b) *in particular affecting which strand is loaded.*

4) Controlling which type of dsRNA is loaded into Argonaute

PACT and/or TRBP may preferentially recognise different subsets of miRNA or siRNA (or their precursors), resulting in differences in processing or RISC loading.

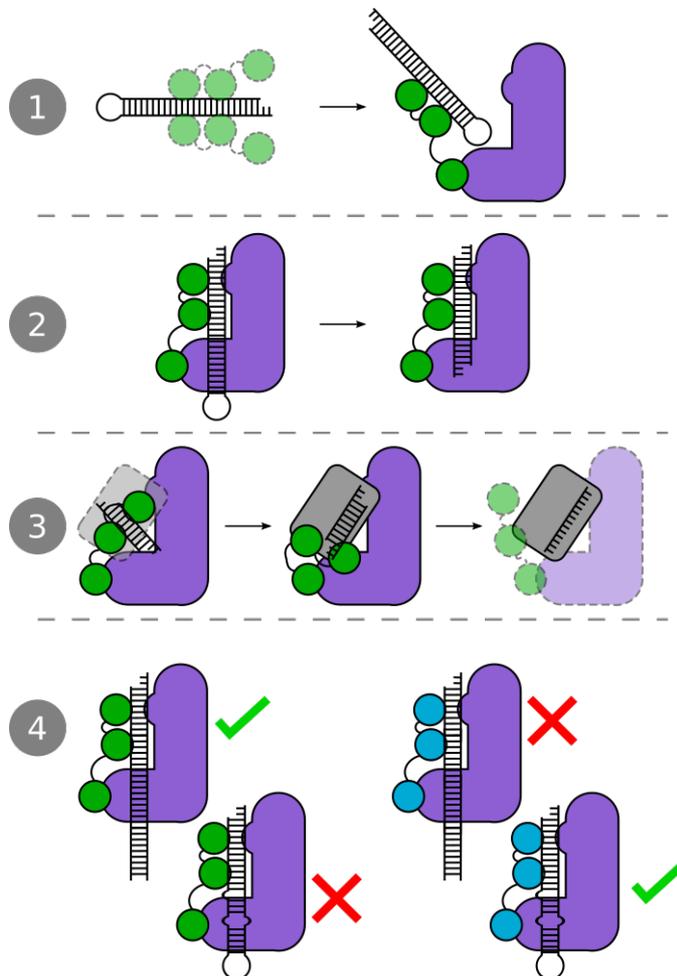


Figure 1.5 Potential roles of PACT and TRBP in the Dicer complex

At least 4 possible roles for PACT/TRBP can be envisaged (see section 1.4.1). The RNA substrates shown for role 4 are illustrative only.

I will now discuss evidence from structural, biochemical, biophysical and molecular cell biology studies in the context of these possible mechanisms.

1.4.2 Evidence from structural studies

As described above, PACT and TRBP have been shown to interact with Dicer, a 1922 amino acid type-III ribonuclease (Figure 1.6). The mechanism by which Dicer controls the length of small RNAs was elegantly resolved in an early crystallographic study (Macrae *et al.*, 2006; MacRae, Zhou and Doudna, 2007). The Dicer PAZ domain binds one end of the pre-miRNA, while a 'ruler' domain positions the PAZ domain at a distance from the RNase III site that corresponds to roughly 22 base pairs of dsRNA. This simple structural mechanism seems to function independently of binding partners as isolated Dicer can still cleave dsRNA targets.

While a considerable amount is known about the structure-function relationship of the RNase III/PAZ region of Dicer, much less is known about the N-terminal DExD/H helicase domain. The helicase domain consists of Hel1 and Hel2 regions separated by a 100 amino acid insert region (Figure 1.6A), which has no close sequence homologues outside of the Dicer family. This domain does not appear necessary for dsRNA cleavage *in vitro* (Ma *et al.*, 2008), nor does processing of pre-miRNAs require ATP (Zhang, 2002). The helicase domain has however been shown to restrict processing of long dsRNA substrates (Kennedy *et al.*, 2015), and the insert region was first identified as the site of TRBP (and presumably PACT) binding through yeast-2-hybrid screening (Daniels *et al.*, 2009).

20-30 Å resolution cryo-electron microscopy studies of a Dicer alone (Lau *et al.*, 2012), in complex with TRBP (Lau *et al.*, 2009), or TRBP and Ago2 (a minimal RISC loading complex) (Wang *et al.*, 2009) have been reported. They reveal an L-shaped overall structure with the helicase located in the base, and the RNase III and PAZ domains located in the long arm (Figure 1.6C). More recently, the crystal structure of TRBP domain 3 in complex with the Dicer helicase insert region was solved, and reveals several interesting features (Wilson *et al.*, 2015).

Firstly, the helicase insert has a broadly similar structure to the equivalent regions of RIG-I and MDA5 helicases, despite having only low sequence similarity (<10%) (Figure 1.6B). Secondly, the N-terminal extension of TRBP was found to form a helix that interacted with the core dsRBD, fitting into a cleft between the two α -

helices and the loop β -2,3. This differs from the location seen in Staufen-D5, where the additional N-terminal helices contacted only the α -helices of the dsRBD core (Gleghorn *et al.*, 2013). Unfortunately, the electron density in this region was poor, so the amino acid sequence of the helix could not be identified.

The structure shows that TRBP-D3 interacts with Dicer via its β -sheet and the C-terminal end of helix α -2. These regions are highly conserved with PACT-D3, which is expected to bind to Dicer using the same interface, but are not conserved in domains 1 and 2 of either protein. No direct contact was observed between Dicer and the additional helix of TRBP. The Dicer-binding surface is also highly conserved with TRBP-D3 and Loqs-D3 (discussed in section 1.7).

By combining crystallographic and electron microscopy data, Doudna and coworkers were able to construct a model of the Dicer/TRBP complex, which suggests that domains 1 and 2 of TRBP have considerable freedom of movement, consistent with roles 1-3 outlined above (Figure 1.6D) (Wilson *et al.*, 2015). Presumably the first two domains of PACT would have less freedom of movement, due to the shorter inter-domain linkers. This is potentially an important structural difference between PACT and TRBP, given the high conservation of the RNA- and Dicer-binding surfaces (see section 1.7), and could plausibly lead to differences in function. It is worth noting that several papers have suggested that Dicer and/or associated proteins undergo conformational shifts (Wang *et al.*, 2009; Taylor *et al.*, 2013) and that these complexes contain multiple dsRNA binding sites (Noland, Ma and Doudna, 2011). This raises the possibility that the role of TRBP may be more complex than is currently believed.

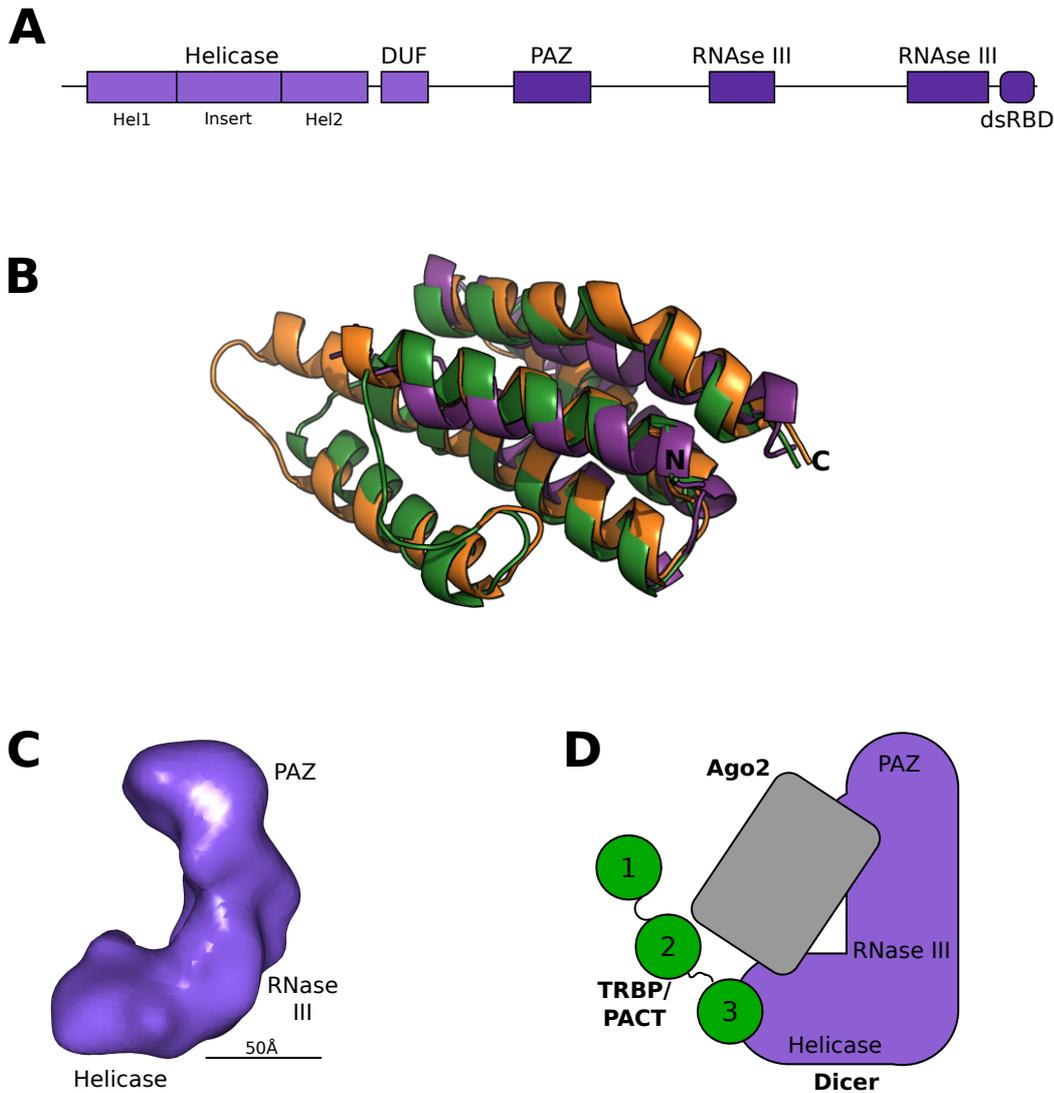


Figure 1.6. Structure of the Dicer complex

A) Layout of domains within Dicer. B) Structural alignment of the helicase insert regions of Dicer (purple, 4WYQ), RIG-I (orange, 4A2W) and MDA-5 (green, 4I1S). Alignments were performed using the DALI server (Holm and Rosenström, 2010). C) Reconstruction of the Dicer-TRBP complex from cryo-electron microscopy (EM) data, with $\sim 15\text{\AA}$ resolution (EM data bank accession EMD-1646) (Lau et al., 2009). The locations of the RNase III and helicase domains are inferred from epitope tagged Dicer (Lau et al., 2012). The position of TRBP is not resolved. D) A schematic of a minimal RISC-loading complex of Dicer, Ago2 and TRBP/PACT based on cryo-EM data (Wang et al., 2009). All components are approximately to scale. It is unknown whether all three components assemble prior to dsRNA binding, or if the complex is more dynamic. *In vivo*, it is likely that other proteins associate with the RISC-loading complex (Robb and Rana, 2007; Pare et al., 2009).

1.4.3 Evidence from biochemical and biophysical studies

A number of *in vitro* studies have shown that the complex of Dicer with PACT or TRBP has a higher affinity for dsRNA than Dicer alone by several orders of magnitude (Chakravarthy *et al.*, 2010; Lee and Doudna, 2012; Lee *et al.*, 2013; Fareh *et al.*, 2015). However, these studies report a more modest effect on small RNA processing rate, ranging from a 5-fold increase in rate, to a 5-fold reduction. Furthermore, TRBP has been reported to bind a pre-miRNA and the duplex produced by Dicer with similar affinities (Chakravarthy *et al.*, 2010; Benoit *et al.*, 2013), which is inconsistent with a role in modulating Dicer processing rate due to preferential binding of substrate over product. With the caveat that the reaction conditions may differ considerably from those present *in vivo*, this suggests that PACT and TRBP fulfil role 1a above, but not 1b.

In an attempt to more closely replicate crowded, heterogeneous cellular conditions, Fareh and coworkers recently studied pre-miRNA processing by Dicer complexes in the presence of excess tRNA molecules (Fareh *et al.*, 2016). They found that Dicer/TRBP cleaved the pre-miRNA more efficiently than Dicer alone, suggesting TRBP and PACT may play a role in discriminating between proper and improper substrates (role 1a/4).

The measured affinity of TRBP (without Dicer) for short dsRNA has varied considerably between different studies (Chakravarthy *et al.*, 2010; Yamashita *et al.*, 2011; Benoit *et al.*, 2013; Acevedo *et al.*, 2015). While this may simply represent variation in sample preparation and technique, it could also be due to the different RNAs studied, which would be consistent with TRBP and PACT showing a degree of specificity for different small RNAs (role 4). Additional evidence comes from studies of reconstituted complexes of Dicer. Dicer/TRBP complexes show different affinities for pre-miRNAs and siRNAs, although Dicer is at least partially responsible for this specificity (Chakravarthy *et al.*, 2010; Lee and Doudna, 2012). A complex containing Dicer and PACT showed considerably slower processing of pre-siRNA substrates than a Dicer/TRBP complex, even though they seemed to have similar processing rates for pre-miRNAs (Lee *et al.*, 2013). This potentially conflicts with an earlier study, which concluded that both PACT and TRBP increase processing of long dsRNA by Dicer (Kok *et al.*, 2007). This discrepancy could reflect a genuine difference in how Dicer complexes

process dsRNAs of different lengths: further experiments using matched reaction conditions will be required to resolve this issue.

Both *in vitro* and *in vivo*, cleavage of pre-miRNAs by Dicer may give a range of products with different lengths, termed iso-miRs (Koscianska, Starega-Roslan and Krzyzosiak, 2011; Neilsen, Goodall and Bracken, 2012). The size distribution of iso-miRs varies depending on the pre-miRNA. One observation that supports the idea that TRBP helps position RNA for cleavage by Dicer (role 2) is that the Dicer/TRBP complex produces different length products compared to Dicer alone (Koscianska, Starega-Roslan and Krzyzosiak, 2011; Fukunaga *et al.*, 2012; Lee and Doudna, 2012; Lee *et al.*, 2013; Kim *et al.*, 2014; Wilson *et al.*, 2015), although PACT has not been observed to affect iso-miR distribution. It is not yet clear what causes the change in cleavage site: possibilities include altering the length of the pre-miRNA (for example, by inducing a change in helical pitch), or changing the position in which the pre-miRNA binds to Dicer. Moving the cleavage site can affect strand-specific loading of Argonaute proteins (Kim *et al.*, 2014), and can alter the seed sequence of the 5' strand.

In *Drosophila*, current models suggest that the TRBP homologue R2D2 plays a significant role in selecting which strand of the miRNA/siRNA is loaded into Argonaute. There is some evidence that this is also the case in humans. The thermodynamic stability of the miRNA- or siRNA-duplex produced by Dicer is dependent on the nucleotide sequence. Experiments with photo-crosslinking dsRNA concluded that TRBP has some propensity to bind to the most thermodynamically stable end of the siRNA (Gredell *et al.*, 2010). Furthermore, strand specific crosslinking to the helicase domain of Dicer required TRBP or PACT and depended on thermodynamic stability, terminal nucleotides and mismatched bases (Noland, Ma and Doudna, 2011). However, strand selectivity is not solely determined at this step: a study that measured which strand was loaded into RISC concluded that Ago2 alone had some strand selectivity (dependent on substrate), though TRBP and PACT could enhance this (Noland and Doudna, 2013).

Lastly, it has recently been shown that TRBP can diffuse along dsRNA (Koh *et al.*, 2013). This behaviour has been observed in a number of other proteins which contain multiple dsRBDs, including PACT, and is dependent upon inter-domain

linker length (Wang *et al.*, 2015; Koh *et al.*, 2016). Although it is not yet clear what effect this would have on the Dicer complex, one-dimensional diffusion may allow the dsRNA to explore a greater range of positions and orientations, potentially facilitating RNA positioning for Dicer processing or RISC loading.

1.4.4 Evidence from molecular cell biology studies

The initial evidence that TRBP and PACT bound Dicer came from immunoprecipitation experiments using HEK293 cells (Chendrimada *et al.*, 2005; Haase *et al.*, 2005; Lee *et al.*, 2006). In one case, epitope-tagged Dicer was used to recover a complex containing TRBP and Ago2 that was able to process pre-let-7 to mature let-7, and cleave target mRNAs (Gregory *et al.*, 2005). The same studies used RNA interference to investigate the function of TRBP and PACT, but gave somewhat contradictory results. However, one common feature is that knocking down any of the 4 components (Dicer, Ago2, TRBP or PACT) resulted in a decrease in levels of mature miRNA (Haase *et al.*, 2005; Lee *et al.*, 2006; Koscianska, Starega-Roslan and Krzyzosiak, 2011).

There are some hints that PACT and TRBP may act on different sets of miRNAs and/or siRNAs (role 4). In one study, knocking down TRBP dramatically reduced the effectiveness of exogenous siRNA while having a small effect on levels of an inducible miRNA, whereas targeting PACT had the opposite effect (Lee *et al.*, 2006). It is currently unclear whether knockdown of PACT or TRBP differentially affects a certain subset of miRNAs.

To determine whether PACT and TRBP aid Dicer processing (roles 1 and 2) or RISC loading (role 3), several studies have examined how depletion of these proteins affects silencing mediated by shRNA (which requires processing by Dicer) or duplex siRNA (which can in principle be incorporated directly into RISC). The consensus is that both are affected, implying an important role in RISC loading, while not ruling out a supporting contribution to Dicer processing (Chendrimada *et al.*, 2005; Lee *et al.*, 2006; Kok *et al.*, 2007). Although one of these studies reports siRNA-mediated silencing to be unaffected by depletion of TRBP or PACT (Kok *et al.*, 2007) the effectiveness of PACT/TRBP knockdown was not measured. It is therefore possible that PACT and TRBP were not effectively depleted due to saturation of the RNAi machinery with siRNA.

It is worth noting that it can be challenging to use RNAi to knock down proteins involved in small RNA processing, as the effectiveness of RNAi is itself dependent on the presence and activity of these proteins. To avoid this problem, a recent study used TALENs (Transcription activator-like effector nucleases) to delete PACT and TRBP in HeLa cells (Kim *et al.*, 2014). Surprisingly, no difference in overall miRNA levels was observed in single or double knockout lines, not even in a subset of miRNAs that had previously been observed to be TRBP dependent (Paroo *et al.*, 2009; Wilson *et al.*, 2015), although changes in strand-selectivity and iso-miR distribution were seen for some miRNAs. Although this casts doubt on a critical role for PACT and TRBP in miRNA biogenesis, a separate study confirmed that TRBP is necessary for production of specific miRNAs in cardiac tissue (Ding *et al.*, 2015). Cardiac-specific deletion of TRBP (via Cre-Lox recombination) was found to cause a reduction of miR-208a levels in the heart, resulting in misregulated gene expression due to overexpression of the miR-208a target SOX6. Taken together, these studies suggest that TRBP and PACT play important roles in the production of particular miRNAs in some cells, but are unlikely to be necessary for miRNA biogenesis in general.

TRBP can be phosphorylated at a number of sites by ERK and downstream kinases (Figure 1.3) (Paroo *et al.*, 2009; Warner *et al.*, 2016). When TRBP with phospho-mimic mutations at S142, S152, S283 and S286 was transfected into human cell lines, miRNA production was shown to increase relative to controls (Paroo *et al.*, 2009). This effect was due to increased stability of the phospho-mimic TRBP protein, since *in vitro* tests showed negligible differences in Dicer processing between complexes containing wild-type and phospho-mimic TRBP (Chakravarthy *et al.*, 2010). A separate set of JNK-mediated phosphorylation events were observed to occur specifically in M phase of the cell cycle (discussed further in section 1.5.2) (Kim *et al.*, 2014). Although PACT also undergoes phosphorylation at several sites, the effects have never been studied in the context of miRNA biogenesis. TRBP domain 1 was recently suggested to undergo SUMOylation: however, the somewhat artificial conditions used in this study leave questions about whether this process occurs naturally (Chen *et al.*, 2015). Post-translational modification of TRBP and PACT is consistent with the idea that their functions are context dependent, and are not constitutively required.

The interpretation of experimental data is complicated by the possibility that TRBP and Dicer may stabilise one another. Several studies have reported that knocking down TRBP also reduces Dicer levels, confounding interpretation in terms of TRBP alone (Chendrimada *et al.*, 2005; Melo *et al.*, 2009; Paroo *et al.*, 2009; Matsui *et al.*, 2015), especially since this effect has not been unanimously reported (Haase *et al.*, 2005). A similar stabilising effect on Dicer was seen for PACT – but not for TRBP – in a separate study (Lee *et al.*, 2006).

The majority of both TRBP and PACT is found in the cytoplasm, particularly in the perinuclear space, though a smaller amount is present in the nucleus (Kok *et al.*, 2007; Laraki *et al.*, 2008; Ohrt *et al.*, 2012). Förster resonance energy transfer (FRET) experiments using fluorescently-tagged TRBP and PACT provide evidence they interact *in vivo* (Kok *et al.*, 2007). The distribution of TRBP and PACT mirrors that of Dicer and Ago2, which are also predominantly cytoplasmic with a small nuclear fraction. Nuclear Dicer has been proposed to regulate gene transcription via Argonaute1, and to prevent accumulation of endogenous dsRNA (White *et al.*, 2014).

Dicer, Ago2, TRBP and PACT have lower diffusion rates in the cytoplasm than in the nucleus (Ohrt *et al.*, 2008, 2012). This observation has been interpreted either as evidence for a large cytoplasmic complex which is disassembled in the nucleus; or as suggesting that these proteins are anchored to a larger cellular feature, such as P-bodies or the rough endoplasmic reticulum. This latter possibility is consistent with reports of co-localisation of RISC components with ribosomes or the endoplasmic reticulum (Chendrimada *et al.*, 2007; Stalder *et al.*, 2013; Kim, Maizel and Chen, 2014; Suzuki and Yokota, 2014).

1.5 TRBP and PACT mediate innate immune surveillance of dsRNA

RNA viruses produce dsRNA in the course of their lifecycle, either as genomic material or as a replication intermediate. Viral dsRNA has features that distinguish it from cellular dsRNA, which may include its longer length, lack of the 5' cap characteristic of eukaryotic mRNAs, and absence of mismatches. The innate immune system can recognise these features through pattern recognition receptors, including RIG-I, MDA5 and TLR3.

The PRRs used by mammalian cells can be divided into ‘early’ responders, which stimulate production of interferons and pro-inflammatory cytokines, and ‘late’ responders, which directly block viral proliferation through translation inhibition and RNA degradation (Figure 1.2A). These pathways have been reviewed in more detail elsewhere (Gantier and Williams, 2007). PACT and TRBP have roles in several of these pathways, and have been implicated in the response to a number of viruses (summarised in Table 1.1).

1.5.1 PACT can stimulate RIG-I activation

Evidence has emerged that PACT can activate RIG-I in response to a number of viruses (Kok *et al.*, 2011; Luthra *et al.*, 2013; Siu *et al.*, 2014) (Table 1.1). Although the helicase domains of RIG-I and Dicer are part of the same subfamily (Fairman-Williams, Guenther and Jankowsky, 2010), it is unlikely that PACT binds these two proteins in the same manner as there is low sequence homology in the helicase insert domain (Daniels *et al.*, 2009; Wilson *et al.*, 2015). This is consistent with a report that PACT binds to the C-terminal domain of RIG-I (Kok *et al.*, 2011). TRBP has not been reported to interact with RIG-I, but has recently been shown to interact with the RIG-I homologue LGP2, and to play a role in the interferon response to several cardioviruses (Komuro *et al.*, 2016).

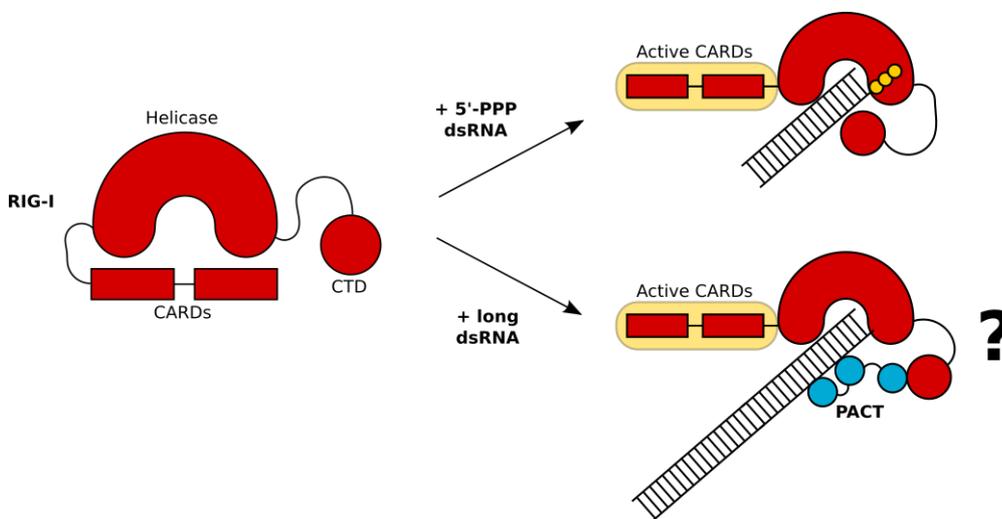


Figure 1.7 PACT-mediated activation of RIG-I

As described in section 1.2.1, the mechanism of RIG-I activation by 5'ppp-dsRNA is well understood. It is less clear how PACT enables RIG-I activation: one possibility is that it increases RIG-I binding to additional ligands such as long dsRNA which lacks a 5'-triphosphate.

Many unanswered questions surround the role of PACT in RIG-I activation. It is unclear under which conditions PACT activates RIG-I, and the mechanism is not known. To date, the interaction has been most studied in the context of Sendai virus (Kok *et al.*, 2011; Luthra *et al.*, 2013), a negative sense single-stranded RNA (ssRNA) virus that infects rodents. In this case, PACT can act to enhance activation of IRF3 and transcription of genes under the control of IFN- β . A similar effect was seen for Ebola virus (also a negative sense ssRNA virus), and overexpressing PACT was able to slow viral replication (Luthra *et al.*, 2013). To counter this effect, Ebola virus encodes a protein that is reported to disrupt the interaction between PACT and both Dicer and RIG-I (Fabozzi *et al.*, 2011; Luthra *et al.*, 2013). An earlier study found that PACT increased type 1 interferon production in response to Newcastle disease virus, which also has a negative sense ssRNA genome (Iwamura *et al.*, 2001). It now seems plausible that this effect was mediated by RIG-I.

One hypothesis is that PACT can extend the range of ligands that can activate RIG-I. Overexpression of PACT did not enhance RIG-I activation by 5'-triphosphate dsRNA, but did enhance the response to the dsRNA analogue poly-I:C (Kok *et al.*, 2011). However, further work will be needed to confirm this, and to disentangle the effects of the other RIG-I (and Dicer) related helicases, MDA5 and LGP2. MDA5 functions similarly to RIG-I, but is activated by longer dsRNA, while LGP2 lacks the CARD signalling domains, and is thought to have an inhibitory effect (Yoneyama and Fujita, 2009). The C-terminal domains of RIG-I, MDA5 and LGP2 are well conserved and so it is plausible that PACT could interact with all of them.

1.5.2 PACT activates PKR, while TRBP inhibits it

TRBP and PACT can regulate PKR: TRBP has an inhibitory effect, while PACT is a conditional activator (Park *et al.*, 1994; Patel and Sen, 1998; Peters, Li and Sen, 2006). The current model is summarised in Figure 1.8. Domain swap and mutational analyses indicate that the first two dsRBDs of PACT and TRBP are functionally interchangeable, and interact with the dsRBDs of PKR (Benkirane *et al.*, 1997; Peters *et al.*, 2001; Gupta, Huang and Patel, 2003). There is some evidence that this interaction is independent of dsRNA, indicating the dsRBDs bind one another directly (Benkirane *et al.*, 1997). However, there is some disagreement on this issue (Cosentino *et al.*, 1995), and it has never been explicitly examined biophysically.

The differences between PACT and TRBP stem from their C-terminal dsRBD. In response to cellular stresses, PACT-D3 is phosphorylated at S246 and S287 by an unknown kinase (or kinases), leading to activation of PKR (Peters, Li and Sen, 2006). The crystal structures of TRBP-D3 reveals that these residues are on opposite faces of the domain, and that S287 is on the edge of the Dicer-binding interface (Wilson *et al.*, 2015). Two models have been proposed to explain PACT-induced PKR activation. Firstly, binding of phosphorylated PACT-D3 to PKR could cause a conformational change that promotes activation (Peters *et al.*, 2001). Supporting this model, PACT-D3 alone has been reported to interact weakly with PKR (Peters *et al.*, 2001; Li *et al.*, 2006). Alternatively, PACT dimers might bind two molecules of PKR, enhancing PKR dimerization (Singh and Patel, 2012). This is consistent with evidence that PACT-D3 phosphorylation promotes homodimerisation and disfavours heterodimerisation with TRBP (Singh *et al.*, 2011; Singh and Patel, 2012).

In contrast, TRBP-D3 has an inhibitory effect on PKR, which appears to be important for preventing inappropriate activation (Gupta, Huang and Patel, 2003; Daher *et al.*, 2009). A similar inhibitory effect is also observed for truncated PKR or PACT constructs containing only the first two dsRBDs (Peters *et al.*, 2001; Huang, Hutchins and Patel, 2002), which suggests that rather than a specific effect of TRBP-D3, anything that binds to the dsRBDs of PKR and disrupts PKR dimerization will cause inhibition. Hyper-phosphorylation of TRBP by JNK increases binding of PKR by TRBP, leading to a greater inhibitory effect (Kim *et al.*, 2014). This occurs during M phase of the cell cycle, and is thought to prevent PKR activation by endogenous dsRNAs during cell division.

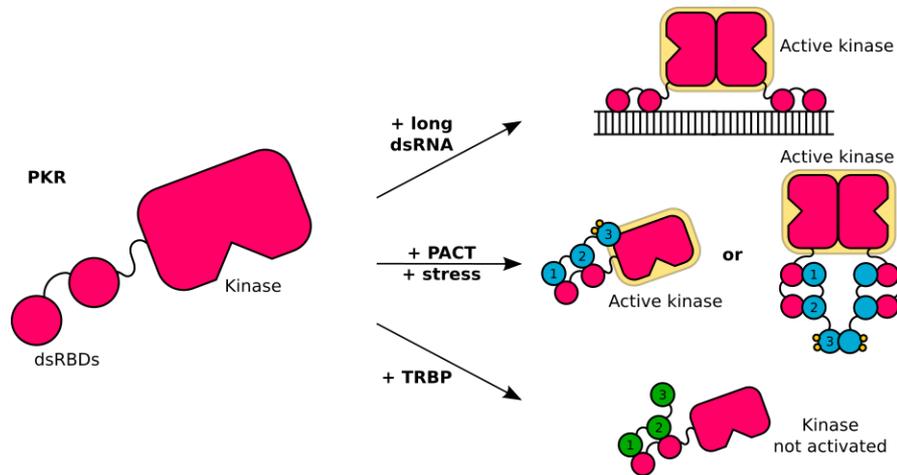


Figure 1.8. Regulation of PKR by PACT and TRBP

As described in section 1.2.2, dimerisation of the kinase domain on dsRNA triggers PKR activation. PACT (when phosphorylated during cellular stress) can also activate PKR, although the mechanism is unclear. The two main hypotheses are: PACT-D3 contacts the kinase domain, somehow favouring activation; or, PACT dimers can bind 2 molecules of PKR, promoting their dimerisation and activation. TRBP acts as an inhibitor of PKR, either because its third domain cannot interact with PKR's kinase domain, or because its third domain exhibits weaker dimerisation.

It is worth noting that many studies prior to 2009 used PACT constructs containing a frameshift mutation that replaces the last 13 amino acids (including part of the C-terminal helix) with 5 unrelated amino acids (Patel and Sen, 1998; Peters *et al.*, 2001; Huang, Hutchins and Patel, 2002; Gupta, Huang and Patel, 2003; Li and Sen, 2003; Li *et al.*, 2006; Laraki *et al.*, 2008; Peters, Dickerman and Sen, 2009). This mutant constitutively activated PKR and disrupted PACT-D3 dimerisation (Daher *et al.*, 2009). The experiments which originally identified PACT phosphorylation were performed on the frameshifted PACT (Peters, Li and Sen, 2006), and have not been repeated on the native sequence.

Table 1.1 Summary of viruses linked to TRBP or PACT

Virus	Reference	Details
<i>Viruses linked to TRBP and PACT</i>		
Newcastle disease virus (NDV)	(Iwamura <i>et al.</i> , 2001)	Overexpressing PACT increased type 1 interferon production from NDV infection, potentially due to interaction with RIG-I.
Sendai virus	(Kok <i>et al.</i> , 2011)	RIG-I activation was enhanced by PACT (note conflict with (Marques <i>et al.</i> , 2008)).
Herpes simplex virus 1 (HSV-1)	(Peters <i>et al.</i> , 2002; Kew <i>et al.</i> , 2013)	HSV-1 protein Us11 reduces type 1 interferon production due to PACT. Blocks PACT interaction with PKR and/or RIG-I
Ebola virus	(Fabozzi <i>et al.</i> , 2011; Luthra <i>et al.</i> , 2013)	Viral proteins VP30 and VP35 interfere with RNAi and interact with TRBP, PACT and/or Dicer. PACT overexpression inhibits viral replication, while VP35 prevents PACT from activating RIG-I.
Influenza	(Tafforeau <i>et al.</i> , 2011)	Viral proteins interact with PACT, resulting in increased viral replication.
MERS-CoV	(Siu <i>et al.</i> , 2014)	Viral protein 4a inhibited PACT activation of RIG-I, resulting in reduced interferon production.
HIV-1	(Bennasser, Yeung and Jeang, 2006; Sanghvi and Steel, 2011a, 2011b)	TRBP blocks PKR activation by HIV transcripts. HIV was suggested to disrupt RNAi by sequestering TRBP, but this is disputed.
Encephalomyocarditis virus (EMCV); Theiler's murine encephalitis virus (TMEV)	(Komuro <i>et al.</i> , 2016)	Knockdown of TRBP reduced interferon β production in infected cells.
<i>Viruses found not to depend on PACT</i>		
Sindbis virus; EMCV	(Iwamura <i>et al.</i> , 2001)	Overexpressing PACT had no effect on type I interferon response.
EMCV; vesicular stomatitis virus (VSV); Sendai virus	(Marques <i>et al.</i> , 2008)	Found no difference between PACT knockout cells vs wild-type.

1.5.3 Links between RNA interference and viral sensing in mammals

In response to infection with RNA viruses, invertebrates and plants are capable of producing antiviral siRNA targeting the viral genome or RNA products (Voinnet, 2005; Ding, 2010). In contrast, mammals do not generally produce detectable levels of siRNA during viral infection, and siRNA is not thought to form a major part of the antiviral response in most cell types (Cullen, 2006; Backes *et al.*, 2014). Instead of processing viral dsRNA into siRNA, it triggers the interferon system through RIG-I or other pattern recognition receptors, as described above.

Recently, several studies have observed siRNA production in mammalian cells with artificially inactivated interferon systems (Benitez *et al.*, 2015; Maillard *et al.*, 2016), or in undifferentiated cells with naturally attenuated interferon responses (Flemr *et al.*, 2013; Maillard *et al.*, 2013). This suggests that the interferon system is incompatible with siRNA biogenesis, and may have supplanted it in differentiated cells (tenOever, 2016).

1.6 Other functions of PACT and TRBP

1.6.1 Fertility and development

Several mouse models of PACT and TRBP depletion have exhibited fertility and developmental defects, but interpretation of their phenotypes has proven challenging to date. Complete PACT deletion resulted in embryonic lethality for homozygous null mice (Bennett *et al.*, 2008), while deletion of TRBP caused a reduction in size, and a severe male fertility defect (Zhong *et al.*, 1999). Two further mouse models of PACT disruption (a truncation after domain 2 that produced no detectable protein expression, and a mutation in domain 2) resulted in a similar phenotype to TRBP deletion, exhibiting slow growth, reduced fertility, and defects in ear and face development (Rowe *et al.*, 2006; Dickerman *et al.*, 2011). These could be attributed (at least in part) to problems in hormone production caused by reduced growth of the anterior pituitary gland (Peters *et al.*, 2009).

Surprisingly, the development and fertility defects of these PACT depletion mouse models could be rescued by additionally knocking out PKR function, suggesting that PACT disruption leads to inappropriate activation of PKR, at least in the anterior pituitary gland (Dickerman *et al.*, 2015). This implies that PACT plays

only a minor role in miRNA biogenesis, consistent with results from PACT knockout cell lines (Kim *et al.*, 2014), but conflicting with many other cell biology studies (see section 1.3.4). The complete rescue of the fertility defect is perhaps unexpected, as a number of studies have found TRBP and PACT to be highly expressed in the testis (Lee, Fajardo and Braun, 1996; Siffroi *et al.*, 2001; Suzuki and Yokota, 2014). Additionally, the *Drosophila* homologue Loqs has been shown to be important for fertility and the maintenance of germline stem cells, despite the lack of PKR in insects (Bennett *et al.*, 2008).

1.6.2 Early onset dystonia-parkinsonism

In humans, point mutations in PACT have been implicated in numerous cases of heritable dystonia-parkinsonism (Camargos *et al.*, 2008, 2012; Zech *et al.*, 2014). Presentation varies, but generally patients experience progressively worsening dystonia starting in childhood or early adulthood, sometimes together with the features of Parkinson's disease. Other than this, patients appear neurologically normal.

The original study designated the locus responsible as DYT16, and identified PACT P222L as the causative mutation (Camargos *et al.*, 2008). This falls in the centre of the highly conserved region in the N-terminal extension of PACT domain 3 (Figure 1.4A). Although the residues surrounding it are highly conserved, P222 is not well conserved among PACT homologues in other species, and is replaced by valine in TRBP. Other causative mutations in PACT domains 1 and 2 have since been identified (Zech *et al.*, 2014; de Carvalho Aguiar *et al.*, 2015).

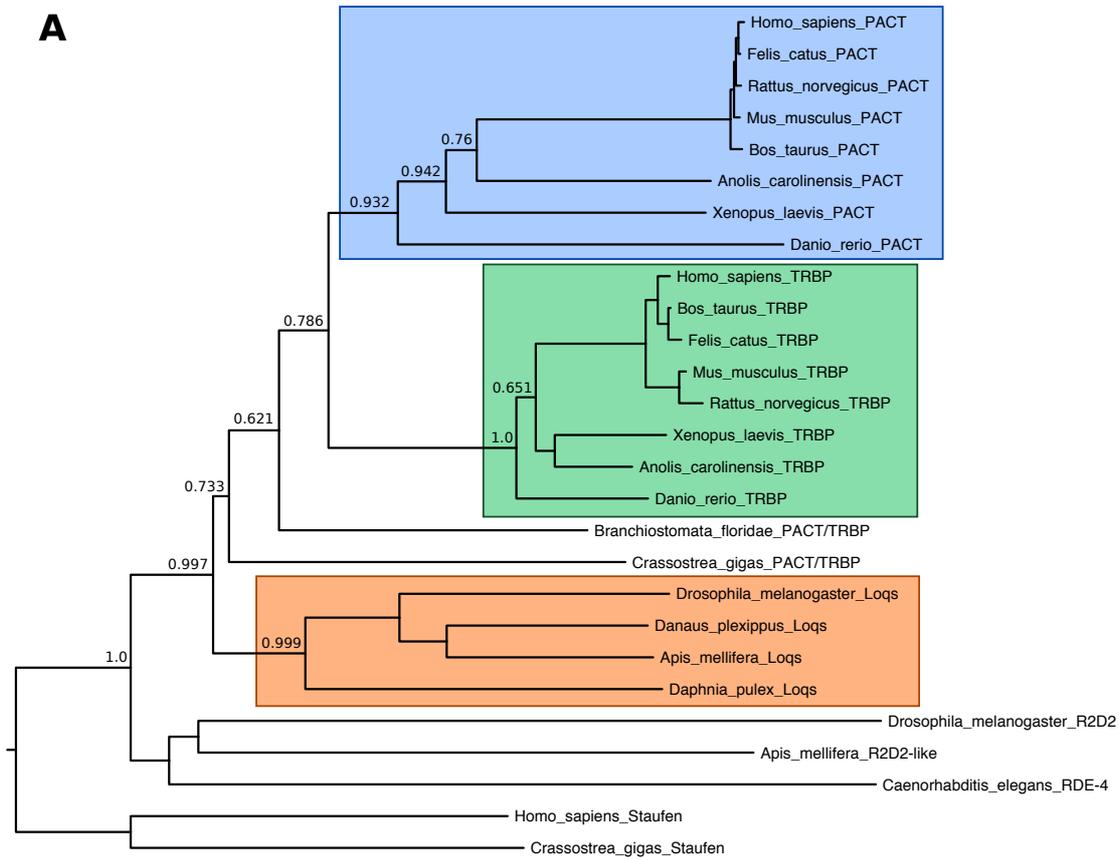
A recent study has reported that PACT P222L mutation alters the kinetics of PKR activation in patient cells, and potentially alters PACT's affinity for PKR and TRBP (Vaughn *et al.*, 2015). This suggests that PACT-associated dystonia may be caused by improper activation of PKR, similar to the developmental defects seen for PACT depletion (section 1.6.1). The symptoms of these two disruptions of PACT are quite different, perhaps due to the stronger PKR/PACT P222L interaction in one case, and the absence of PACT/PKR interactions in the other. Alternatively, PACT function may be impaired in different cell types in these two conditions.

1.7 General features and evolutionary history of PACT and TRBP homologues

Recent advances in next-generation sequencing have vastly reduced the amount of time and effort to sequence an entire genome (Goodwin, McPherson and McCombie, 2016). As a result, there has been an explosion in the number of organisms for which genomes are available, albeit often in draft form. This allows the evolutionary relationships, conservation and divergence of PACT and TRBP to be examined across a broad range of species.

In the phylogenetic tree, distinct clades can be identified for PACT, TRBP and Loqs (shown by coloured boxes in Figure 1.9). Interestingly, the PACT and TRBP clades are more closely related to one another than to the Loqs clade. This implies that PACT and TRBP diverged from one another after the split from Loqs, but before the divergence of ray-finned fishes (represented by *Danio rerio*) and vertebrates. This is supported by the presence of only a single PACT homologue in the model chordate *Branchiostoma floridae*. Note however that the bootstrap confidence values for the nodes between PACT, TRBP and Loqs are all substantially below 1, indicating that the tree architecture here is slightly ambiguous.

R2D2 and RDE-4 are dsRBD-containing proteins from *D. melanogaster* and *C. elegans* respectively which have been shown to interact with Dicer proteins to assist small RNA-mediated silencing (Liu *et al.*, 2003; Tomari *et al.*, 2004; Parker, Eckert and Bass, 2006). Examination of the sequence alignment shows that while these proteins show weak sequence similarity in the first two domains, conservation of domain 3 is extremely poor, and lacks residues conserved in almost all other known dsRBDs (Masliah, Barraud and Allain, 2013).

A**B**

1	Homo sapiens PACT	201IS	TNVVGHSLGCT	WHSLRNP	GEKINL	KRSL	..	SIPNTDYIQL	LS						
	Bos taurus PACT	201IS	TNVVGHSLGCT	WHSLRNP	GEKINL	KRSL	..	SIPNTDYIQL	LS						
	Mus musculus PACT	201IS	TNVVGHSLGCT	WHSLRNP	GEKINL	KRSL	..	SLPNTDYIQL	LS						
	Rattus norvegicus PACT	201IS	TNVVGHSLGCT	WHSLRNP	GEKINL	KRSL	..	SLPNTDYIQL	LS						
	Felis catus PACT	176IS	TNVVGHSLGCT	WHSLRNP	GEKINL	KRSL	..	SIPNTDYIQL	LS						
	Anolis carolinensis PACT	392IS	NLSRLNGCT	WHLRNP	REKTL	KMSF	..	SIPNTDYVHL	LR						
	Xenopus laevis PACT	197IS	NKLTGNKMGCT	WHSLRNP	GEKISM	KRSP	..	SIPNTDYVKM	PK						
	Danio rerio PACT	180ET	WSPSRVYV	WHLRNP	GEKIVL	KRTP	..	SLPNTDYIQML	LS						
2	Homo sapiens TRBP	247	.IGVG.....SR	DGLRNRGPGCT	WDSLRNV	VEKILS	RSCS	GSLGALGPACCRV	LS							
	Bos taurus TRBP	247	.IGVG.....SR	DGLRNRGPGCT	WDSLRNV	VEKILS	RSCS	GSLGALGPACCRV	LS							
	Felis catus TRBP	247	.IGVG.....SR	DGLRNRGPGCT	WDSLRNV	VEKILS	RSCS	GALGALGPACCSV	LS							
	Mus musculus TRBP	246	.IGVS.....SR	DGLRNRGPGCT	WDSLRNV	VEKILS	RSCS	GSLGALGSAACCSV	LS							
	Rattus norvegicus TRBP	246	.IGVS.....SR	DGLRNRGPGCT	WDSLRNV	VEKILS	RSCS	GSLGALGSAACCSV	LS							
	Xenopus laevis TRBP	238	.V.....GK	DGSRSRGTACT	WDSLRNV	CEKILH	RSNP	..	TILSSGFCSL	LO						
	Anolis carolinensis TRBP	245	.ITAG.....NK	DGVKSRGSGCT	WDSLRNV	AGEKILH	KSNP	..	GVLNAGFCSL	LE						
	Danio rerio TRBP	227	.MQIG.....GR	EGGKSKGLGCT	WDSLRNV	AGEKILQ	RCNP	GQSDSIDSNFCSL	LR							
3	Branchiostomata floridae PACT/TRBP	207TG	KPMHAARIGT	WATLRNV	AGEKITR	KSTS	..	SPNSNYCQL	LO						
	Crassostrea gigas PACT/TRBP	197	LNAS.....E	IRASYTALKEGKVLPT	SPQNKELQ	QYQKKN	..	HKIKNSK	..	TAPATNYQM	LO					
4	Drosophila melanogaster Loqs	340	SSE.....NY	YGEIKDI	SVPT	TQHSNKVSP	ERTK	KVAT	TKLLK	QKCL	..	KNKIDYFK	LC			
	Danaus plexippus Loqs	277	AR.....Y	ADLK	.KITT	TTSHSHKVS	FFH	KHLKQ	VGNL	VK	QVTP	..	NNKDFNFVQ	FO		
	Apis mellifera Loqs	259	AR.....Y	ADLK	.GSKIST	TTIHSKVS	FFH	KSLKS	TGVK	LFE	QNTC	N	.DGDVNLVQ	FO		
	Daphnia pulex Loqs	167LED	ED	EVA	..	KDPCLDYQ	LO

1	Homo sapiens PACT	247	SI	AK	OG	NIT	DI	EL	S	ANG	QV	OC	IA	EL	SL	SP	V	GS	G	IS	C	GN	OS	DA	AL	V	K	I	A	E	R	K										
	Bos taurus PACT	247	SI	AK	OG	NIT	DI	EL	S	ANG	QV	OC	IA	EL	SL	SP	V	GS	G	IS	C	GN	OS	DA	AL	V	K	I	A	E	R	K										
	Mus musculus PACT	247	SI	AK	OG	NIT	DI	EL	S	ANG	QV	OC	IA	EL	SL	SP	V	GS	G	IS	C	GN	OS	DA	AL	V	K	I	A	E	R	K										
	Rattus norvegicus PACT	247	SI	AK	OG	NIT	DI	EL	S	ANG	QV	OC	IA	EL	SL	SP	V	GS	G	IS	C	GN	OS	DA	AL	V	K	I	A	E	R	K										
	Felis catus PACT	222	SI	AK	OG	NIT	DI	EL	S	ANG	QV	OC	IA	EL	SL	SP	V	GS	G	IS	C	GN	OS	DA	AL	V	K	I	A	E	R	K										
	Anolis carolinensis PACT	436	SI	AK	OG	HAT	DI	EL	S	VNG	QV	OC	IA	EL	SL	HP	V	GT	G	IS	W	GN	HN	DA	AL	V	K	I	A	E	R	K										
	Xenopus laevis PACT	243	DVAE	LD	NLT	DI	EL	S	LVN	QV	OC	IA	EL	SL	NP	V	GT	G	IS	C	GN	HN	DA	AL	V	K	I	A	E	R	K											
	Danio rerio PACT	223	IS	LG	QV	DI	EL	S	LVN	QV	OC	IA	EL	SL	RP	V	GT	G	IS	C	GN	HN	DA	AL	V	K	I	A	E	R	K											
2	Homo sapiens TRBP	300	EL	SE	QA	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	A	T	RE	A	R	GE	A	R	AL	V	K	I	A	E	R	K								
	Bos taurus TRBP	300	EL	SE	QA	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	A	T	RE	A	R	GE	A	R	AL	V	K	I	A	E	R	K								
	Felis catus TRBP	300	EL	SE	QA	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	A	T	RE	A	R	GE	A	R	AL	V	K	I	A	E	R	K								
	Mus musculus TRBP	299	EL	SE	QA	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	A	T	RE	A	R	GE	A	R	AL	V	K	I	A	E	R	K								
	Rattus norvegicus TRBP	299	EL	SE	QA	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	A	T	RE	A	R	GE	A	R	AL	V	K	I	A	E	R	K								
	Xenopus laevis TRBP	285	EL	SE	QA	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	A	T	RE	A	R	GE	A	R	AL	V	K	I	A	E	R	K								
	Anolis carolinensis TRBP	295	EL	SE	QA	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	A	T	RE	A	R	GE	A	R	AL	V	K	I	A	E	R	K								
	Danio rerio TRBP	280	EL	SE	QR	HVS	DI	EL	S	LS	SL	QC	IV	EL	SL	OP	V	GS	F	A	S	S	L	D	A	R	A	S	A	A	AL	V	K	I	A	E	R	K				
3	Branchiostomata floridae PACT/TRBP	253	EL	AE	ON	RE	V	DI	EL	S	ASS	L	OC	IV	EL	SL	OP	V	GS	G	H	T	R	D	E	A	H	A	A	AL	V	K	I	A	E	R	K					
	Crassostrea gigas PACT/TRBP	258	EI	SE	VR	RE	V	DI	EL	S	TR	QR	OC	IV	EL	SL	LP	V	GS	G	T	V	D	E	A	H	A	A	AL	V	K	I	A	E	R	K						
4	Drosophila melanogaster Loqs	400	EI	AT	NO	RE	V	DI	EL	S	TR	QR	OC	IV	EL	SL	LP	V	GS	G	P	T	A	A	D	Q	R	H	A	A	AL	V	K	I	A	E	R	K				
	Danaus plexippus Loqs	334	EI	AS	CO	RE	V	DI	EL	S	TR	QR	OC	IV	EL	SL	LP	V	GS	G	L	T	S	K	D	Q	S	A	A	AL	V	K	I	A	E	R	K					
	Apis mellifera Loqs	316	EI	AS	CO	RE	V	DI	EL	S	TR	QR	OC	IV	EL	SL	LP	V	GS	G	V	T	S	K	D	Q	S	A	A	AL	V	K	I	A	E	R	K					
	Daphnia pulex Loqs	189	EI	SA	OK	D	A	T	V	DI	EL	S	TR	QR	OC	IV	EL	SL	LP	V	GS	G	S	S	Y	E	B	O	K	E	A	F	S	A	L	V	K	I	A	E	R	K

Figure 1.9. Phylogenetic analysis of PACT, TRBP and Loqs homologues

The amino acid sequences of PACT, TRBP and Loqs homologues were retrieved from NCBI, supplemented in some cases with organism-specific databases (Table 1.2). These were aligned using ClustalX with default settings, and a phylogenetic tree constructed from ungapped positions. A) Phylogenetic tree of PACT, TRBP, Loqs and R2D sequences. Staufen sequences were used as an outgroup, as 3 of their 5 dsRBDs can be aligned to PACT, but their different domain structure implies they diverged prior to all PACT, TRBP and Loqs homologues examined here. Selected nodes are labelled with the bootstrap value from 1000 resamplings. B) A subset of the main alignment showing domain 3 of PACT, TRBP and Loqs sequences.

The third domain of RDE-4 has been shown to be required for its effect on miRNA biogenesis, so it is possible that this domain is still responsible for interacting with Dicer, perhaps through a different interface (Parker, Eckert and Bass, 2006). As the third domains of R2D2 and RDE-4 are so divergent from PACT, TRBP and Loqs, they are excluded from the following discussion.

1.7.1 Similarities between PACT homologues

There are a number of sequence features common to all the PACT, TRBP and Loqs sequences examined. The presence of these features in both mammalian and insect sequences implies that they have been conserved for at least 500-600 million years, and are therefore crucial to the common function of these proteins (Wray, 2015).

Firstly, the majority of the motifs previously identified as required for the dsRBD fold are conserved in all three domains (Masliah, Barraud and Allain, 2013). Similarly, the residues required for dsRNA binding (in particular, the histidine in loop β -12 and the lysines in helix α -2) are conserved in domains 1 and 2, but not in domain 3. Together, these two considerations account for almost all the perfectly conserved residues in domains 1 and 2.

The third domain contains a number of highly conserved residues that cannot be explained by dsRBD structure alone. Examples include the **YLDIEE** motif in strand β -1, the **QCL** motif in strand β -2 and the **VC[H/Y]** in strand β -3. The majority of

these conserved residues fall within the β -sheet region, which was recently shown to form the Dicer-binding interface (Wilson *et al.*, 2015).

Table 1.2. PACT and TRBP homologue sequences

Species	Protein	Accession
<i>Homo sapiens</i>	PACT	O75569
<i>Felis catus</i>	PACT	XP_006935445
<i>Rattus norvegicus</i>	PACT	EDL79230
<i>Mus musculus</i>	PACT	EDL27212
<i>Bos taurus</i>	PACT	NP_001039335
<i>Anolis carolinensis</i>	PACT	XP_008116415
<i>Xenopus laevis</i>	PACT	Q7ZYA5
<i>Danio rerio</i>	PACT	XP_005172565
<i>Homo sapiens</i>	TRBP	NP_599150
<i>Bos taurus</i>	TRBP	NP_001069146
<i>Felis catus</i>	TRBP	XP_003988820
<i>Rattus norvegicus</i>	TRBP	NP_001030113
<i>Mus musculus</i>	TRBP	NP_033345
<i>Xenopus laevis</i>	TRBP	XP_018100523
<i>Anolis carolinensis</i>	TRBP	XP_008101858
<i>Danio rerio</i>	TRBP	NP_956291
<i>Branchiostoma floridae</i>	PACT/TRBP	XP_002587993
<i>Crassostrea gigas</i>	PACT/TRBP	XP_011456094
<i>Drosophila melanogaster</i>	Loqs	AAF53295
<i>Danaus plexippus</i>	Loqs	EHJ63609
<i>Apis mellifera</i> *	Loqs	gnl Amel_4.5 GB47214-PA
<i>Daphnia pulex</i>	Loqs	EFX79291
<i>Drosophila melanogaster</i>	R2D2	AAF52561
<i>Apis mellifera</i>	R2D2-like	XP_006560091
<i>Caenorhabditis elegans</i>	RDE-4	AAL61544
<i>Homo sapiens</i>	Staufen	CAB40082
<i>Crassostrea gigas</i>	Staufen	XP_011432604

* The accession is for the Amel_4.5_OGSv3.2_pep database of predicted protein sequences based on the *Apis mellifera* genome assembly 4.5, not for NCBI.

1.7.2 Differences between PACT homologues

The maintenance of both PACT and TRBP in vertebrates suggests they have distinct functions, while the differences in small RNA biogenesis and function between vertebrates and arthropods suggests that the function of Loqs will be somewhat different again. Sequence comparison can shed light on which parts of the protein are differentially selected in these different lineages.

A particularly striking difference is the length of the inter-domain linkers. The linker between the first two domains of PACT is approximately 25 residues long, while the equivalent region of TRBP is approximately 67 residues long. The specific sequence of the linker is not well conserved, but in both cases contains a high proportion of proline residues. These features are consistent with this region being unstructured as was previously found for human TRBP (Benoit *et al.*, 2013), and suggest that the main function of this linker is to determine the spatial separation between domains 1 and 2. Linker length has been found to tune RNA-binding affinity of other multidomain proteins, and could plausibly perform a similar role in PACT and TRBP (Shamoo, Abdul-Manan and Williams, 1995). The length of the linker between TRBP domains 1 and 2 has recently been shown to influence 1 dimensional diffusion of the protein along dsRNA (Koh *et al.*, 2016). While the function of this diffusion is unknown, it could be speculated that differences in diffusion between TRBP and PACT would affect the mobility of miRNAs and pre-miRNAs within the Dicer complex.

The linker between the second and third domains also differs in length, being approximately 47 residues long in PACT and 67 residues long in TRBP. This linker does contain a 15 amino acid region that is extremely highly conserved among vertebrate PACT and TRBP sequences, and which is thought to form an α -helix that docks onto the third dsRBD (Wilson *et al.*, 2015). This region is not conserved with Loqs, suggesting it may function in a vertebrate-specific role of PACT and TRBP, such as in interferon-mediated immune responses. The remainder of the linker is not well conserved and contains a high proportion of charged and hydrophilic residues, so likely acts mainly to determine the distance between the second and third domains. This distance will affect the positioning of the dsRNA-binding domains, and any associated RNA molecules, within the Dicer complex.

Neither linker sequence nor length is well conserved among the Loqs sequences examined here, so no conclusions can be drawn about their function. There are also numerous sequence differences within the dsRBDs of all three proteins, but these are very difficult to interpret without experimental data.

1.8 Summary

PACT and TRBP each have two canonical dsRBDs and one non-canonical dsRBD, separated by linker regions that are predicted to be flexible. There is about 50% sequence identity between equivalent domains, but much poorer conservation in the linkers, which are substantially longer in TRBP. The first two domains interact non-specifically with RNA, while the third is thought to mainly mediate protein-protein interactions.

PACT and TRBP have been conclusively shown to associate with the Dicer complex, and almost certainly increase its affinity for dsRNA. TRBP can alter iso-miR distribution, and this can affect which strand is loaded into Argonaute. Beyond this, the functions of PACT and TRBP in miRNA biogenesis are unclear, due to conflicts between the results of different studies.

PACT can likely increase activation of RIG-I in response to viral RNA, but the mechanism is unclear. Both PACT and TRBP can inhibit PKR, while phosphorylated PACT can also activate it, with the caveat that these experiments were conducted with a truncated version of PACT. The mechanism by which PACT and TRBP regulate PKR has not been resolved, but is dependent on the third domain.

PACT is important for the development of the anterior pituitary gland, potentially by preventing inappropriate PKR activation. Mutations in PACT can cause early onset dystonia/Parkinsonism, again potentially through regulation of PKR. Both TRBP and PACT are important for fertility and are expressed in germ cells, but this remains poorly understood. Overall, there is a disconnect in our understanding of PACT and TRBP between the molecular and whole organism levels, especially their role in miRNA biogenesis.

1.9 Aims of this study

It is clear that the third domains of PACT and TRBP are important both for binding to Dicer in the miRNA biogenesis pathway, and for their different effects on PKR in response to viral infection and cellular stress. However, no structural information about these domains was available at the outset of this project, and very little was known about the structure of class B dsRBDs more generally, let alone how they interact with protein binding partners. Biochemical data had shown that PACT domain 3 could homodimerise, and that this may be regulated by phosphorylation, but this dimerisation had not been studied in the context of miRNA biogenesis. The stoichiometry and organisation of the Dicer complex depends on whether PACT homodimerisation and Dicer-binding can occur simultaneously.

Therefore, this study has three major aims:

- 1) Investigate the structure of PACT-D3, to determine whether it deviates from the canonical dsRBD fold.
- 2) Characterise the dimerisation mechanism of PACT-D3 and TRBP-D3, to establish whether it is compatible with Dicer-binding.
- 3) Explore whether phosphorylation alters dimerisation behaviour.

Protein Production Methods

2.1 Protein expression in *E. coli*

2.1.1 Preparation of vectors

Codon-optimised DNA sequences were ordered from Eurofins Genomics (see Appendix 1). These were then cloned into pETFPP expression vectors developed by the University of York Technology Facility, which are based on the commonly used pET-28a plasmid (Figure 2.1). These vectors contain a gene conferring resistance to kanamycin, and an open reading frame (ORF) containing a 6xHis tag (MGSSHHHHHSS), a solubility tag, and a human rhinovirus (HRV) 3C protease cleavage site (GLEVLFQ[^]GPAM). Four solubility tags were tested: Maltose Binding Protein (MBP), Glutathione-S-transferase (GST), *Escherichia coli* immunity protein 9 (Im9) (James, Kleanthous and Moore, 1996) and Green Fluorescent Protein (GFP). In some cases, a construct was tested with a 6xHis tag and HRV 3C cleavage site, but no solubility tag (His).

The protein of interest was cloned directly after the 3C cleavage site using an InFusion cloning strategy, eliminating the need for specific restriction sites (cloning was performed by the University of York Technology Facility). Upstream of the ORF is a T7 promoter, and a lac operator element which allows protein expression only when induced by isopropyl β -D-1-thiogalactopyranoside (IPTG). All constructs were sequenced before use to ensure no errors had been introduced during cloning. Note that because HRV 3C protease cleaves between glutamine and glycine, the non-native sequence GPAM remain at the N-terminus of the protein of interest.

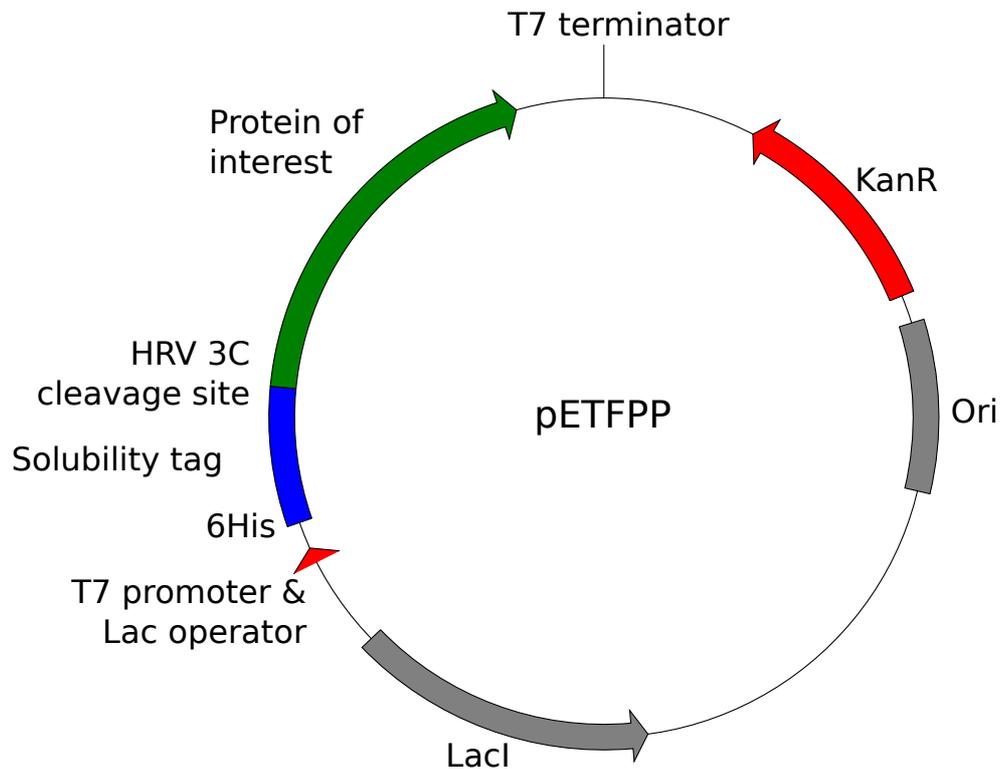


Figure 2.1. Schematic of the pETFPP expression vector

2.1.2 Transformation into *E. coli* strain BL21 (DE3)

The BL21 (DE3) strain of *E. coli* was used for protein expression for two reasons. Firstly, many of the genes encoding proteases have been deleted, which helps prevent degradation of the overexpressed protein. Secondly, it expresses the T7 polymerase under the control of a lac operator, which is necessary for transcription of the mRNA encoding the protein of interest.

For the transformation, 50 μL of chemically competent *E. coli* was incubated on ice with 2 μL of plasmid ($>100 \text{ ng}/\mu\text{L}$) for 30 minutes in a 1.5 mL tube. The mixture was then incubated at 42 $^{\circ}\text{C}$ for 45 seconds followed by incubation on ice for 2 minutes. 450 μL of LB media was added, and the mixture was incubated at 37 $^{\circ}\text{C}$ for 30-40 minutes, with agitation. The cells were then grown overnight at 37 $^{\circ}\text{C}$ on LB/agar plates supplemented with 50 $\mu\text{g}/\text{mL}$ kanamycin.

2.1.3 Site-directed mutagenesis

PrimerX (<http://www.bioinformatics.org/primerx/>) was used to design mutagenesis primers with a length of 25-45 nt, with a melting temperature of at least 78 $^{\circ}\text{C}$, and terminating in at least one C or G. Where possible, the CG content was kept between 40% and 60%. The mutagenesis reaction was carried out on 100 ng of freshly prepared plasmid DNA using the QuikChange Lightning

mutagenesis kit (Agilent), according to the manufacturer's instructions. An extension period of 3 m 30 s was used in the thermocycle. The original, methylated plasmid was digested with DpnI, and the product transformed into XL10-Gold supercompetent cells. All mutants were verified by sequencing.

2.1.4 Culture and overexpression

10 mL Luria Broth (LB)/kanamycin was inoculated with a single colony, and grown for 6 hours at 37 °C with shaking at 120-180 rpm (media and buffer recipes are given in section 4 below). The optical density at 600 nm (OD_{600}) was measured, then a volume V_{in} was centrifuged and resuspended in 50 mL M9/kanamycin minimal media, such that the new OD_{600} was 0.05.

$$V_{in} = \frac{(OD_{600} \text{ of preculture}) \cdot (\text{final volume})}{(\text{desired final } OD_{600})} = \frac{OD_{600} \cdot 50 \text{ ml}}{0.05}$$

This M9 preculture was grown for 16 hours at 37 °C. Then the OD_{600} was measured, and a volume resuspended in 1 L M9/kanamycin to give a starting OD_{600} of 0.05. This culture was grown at 37 °C until it reached an OD_{600} of 0.6-0.8, at which point it was supplemented with IPTG to a final concentration of 1 mM. It was then grown for 14-16 hours at 20 °C, and then centrifuged for 20 mins at 5000 RCF to pellet the cells. The cell pellet was then lysed immediately (as described below), or stored at -20 °C until ready for use.

2.2 Protein purification

2.2.1 Assessing solubility of constructs

To determine whether the constructs were expressed and soluble, cultures were grown as above, but were induced and harvested after the 50 mL M9 culture. Cell pellets were resuspended in a volume of lysis buffer supplemented with 1 mg/mL hen egg lysozyme (Sigma) and 0.5% (v/v) Triton X-100, to give a final OD_{600} of 40.

$$V_{lysis} = \frac{V_{culture} \cdot OD_{600}}{40}$$

The cell suspension was then agitated at 30°C for 40 minutes, then centrifuged for 5 minutes at 20000 relative centrifugal force (RCF). The insoluble pellet was then resuspended in the same volume of lysis buffer, and samples of total protein (before centrifugation), supernatant and pellet were examined by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis).

2.2.2 Large-scale purification

1 litre cultures of the MBP-tagged constructs were grown as described above, and lysed using either continuous flow French press or sonication. These methods were chosen over chemical lysis for two reasons: firstly, chemical lysis efficiency was quite variable when the reaction was scaled up to larger volumes, possibly due to less effective sample mixing; secondly, the high concentration of lysozyme used proved difficult to separate using nickel affinity chromatography. Lysis by French press or sonication was more reproducible, and avoided the need to separate exogenous lysozyme.

Cell pellets were resuspended in 35 mL of lysis buffer, and lysed by continuous flow French press cooled to 4 °C (Constant Systems) or by sonication. In the latter case, a Sonicator 3000 (Misonix) was used to sonicate samples on ice in two stages, with 3 minutes total on time each (3 s on, 3 s off, power level 6), and separated by a 10 minute cool-down period. The lysate was then centrifuged at 40000 RCF for 45 minutes, and the supernatant passed over a nickel affinity column equilibrated in Ni-IMAC binding buffer at 2 mL/min. Bound protein was eluted with Ni-IMAC elution buffer, and concentration assessed by UV spectrophotometry (using hypothetical extinction coefficients calculated by the ProtParam server (web.expasy.org/protparam/)).

Table 2.1. UV extinction coefficients of PACT and TRBP constructs

Protein construct	Molecular weight (kDa)	ϵ_{280} (cm⁻¹ M⁻¹)
PACT-D3	8.1	6000
PACT-Ext-D3	11.6	11600
TRBP-D3	8.5	3100
TRBP-Ext-D3	11.9	8650
Loqs-D3	8.1	4500

To remove the solubility tag from the eluted protein, His-tagged HRV 3C protease (produced by the University of York Technology Facility) was added in a ratio of 1 mg protease per 50 mg eluted protein. This mixture was dialysed at 4 °C for 14 hours against a >50-fold larger volume of 20 mM Tris, 0.5 M NaCl, 0.5 mM dithiothreitol (DTT), pH 7.5, using a dialysis membrane with a molecular weight cut-off (MWCO) of 3.5 kDa (Spectrum Labs). DTT was necessary for efficient cleavage, possibly by preventing oxidation of a catalytic cysteine residue (Matthews *et al.*, 1994). A second nickel affinity purification was then carried out to separate the cleaved protein of interest (found in the column flow through) from the His-solubility tag and 3C protease.

To remove remaining impurities (mostly MBP and uncleaved protein), the protein was first concentrated using spin columns with a 5 kDa MWCO (VivaSpin) to a volume of approximately 500 µL. This was then passed over a S75 Superdex 16/60 column (GE Healthcare) equilibrated in size exclusion buffer (see Chapter 6, section 6.3), at a flow rate of 1 mL/min. Fractions were visualised by SDS-PAGE, and fractions containing the protein of interest were pooled and concentrated. For samples requiring TCEP, an additional dialysis step against 50 mL of the TCEP-containing buffer was performed.

2.3 SDS-PAGE

15% SDS-PAGE were hand-cast using the recipes in section 2.4.3 below. Protein samples were prepared by mixing 30 µL protein, 20 µL SDS-PAGE sample buffer and 10 µL 1 M DTT and boiling for 5 minutes. 10 µL of sample was loaded per lane, and gels run at 180-200 V for approximately 50 minutes using a mini-PROTEAN tetra cell system (BioRad). Gels were then stained for approximately 30 minutes in Coomassie stain, then destained until the bands were visible. Gels were imaged using a GeneGenius Bio Imaging System (Syngene).

2.4 Media and buffer compositions

2.4.1 Media

LB media

Component	Concentration (g/L)
Tryptone	10
Yeast Extract	5
NaCl	10
Agar (only if making plates)	15

Autoclave. Optionally add 50 µg/mL kanamycin before use.

M9 base

Component	Concentration (g/L)
Na ₂ HPO ₄ (anhydrous)	6
KH ₂ PO ₄ (anhydrous)	3
NaCl	0.5
NH ₄ Cl *	1

Adjust pH to 6.8-7.2, then autoclave.

* For 50 mL cultures, ¹⁴NH₄Cl was used. For 1 L cultures, ¹⁵NH₄Cl was used (Cambridge Isotope Laboratories, Inc.)

1 L M9

Component	Volume
M9 base	1 L
20% D-Glucose *	10 mL
1 M MgSO ₄	1 mL
100 mM CaCl ₂	1 mL
100 mM MnCl ₂	1 mL

50 mM ZnSO ₄	1 mL
100 mM FeCl ₃	0.5 mL
500x Vitamin Cocktail	2 mL
50 mg/mL Kanamycin	1 mL

Nb. All components except the M9 base are passed through a 0.22 µm filter before use.

* When producing samples for 3D NMR experiments, 99% [U-¹³C]-labelled D-glucose was used (Cambridge Isotope Laboratories, Ltd.)

50 mL M9 500x vitamin cocktail

Component	Mass (mg)
Pyridoxine	25
Biotin	25
D-pantothenic acid hemicalcium salt	25
Folic acid	25
Choline Chloride	25
Niacinamide	25
Riboflavin	2.5
Thiamine	125

Adjust pH to 7 to solubilise, then pass through 0.22 µm filter. Adjust to pH 5 for storage.

2.4.2 Lysis and purification

Lysis buffer

Component	Concentration
Tris	20 mM
NaCl	150 mM
CaCl	10 mM
MgCl	10 mM
DNase I	20 µg/µL
RNase A	20 µg/µL
Leupeptin	1 µg/µL

Nickel affinity purification buffers

Component	Ni-IMAC binding buffer	Ni-IMAC elution buffer
Tris	20 mM	20 mM
NaCl	0.5 M	0.5 M
Imidazole	20 mM	0.5 M

Adjust pH to 7.5, then pass through 0.22 µm filter

SEC-MALLS buffer

Component	Concentration
MES	20 mM
NaCl	150 mM

Adjust pH to 6.5, then pass through 0.22 µm filter

2.4.3 SDS-PAGE

Acrylamide gel recipes

Component	15% resolving gel (10 mL)	4% Stacking gel (5mL)
30% Acrylamide	4.9 mL	650 µL
1.5 M Tris, pH8.8	2.5 mL	-
0.5 M Tris pH 6.8	-	1.25 mL
10% SDS	100 µL	50 µL
Distilled H ₂ O	2.5 mL	3.05 mL
10% APS*	100 µL	50 µL
TEMED*	10 µL	5µL

* Add just before pouring

SDS-PAGE Running buffer

Component	Concentration
Tris	3 g/L
Glycine	14 g/L
SDS	1 g/L

SDS-PAGE sample buffer

Component	
1 M Tris pH 7.2	1 mL
Distilled H ₂ O	3 mL
10% SDS	10 mL
Bromophenol blue	0.06 g
Glycerol	12 g

1 L Coomassie stain

Component

Ethanol	450 mL
Acetic acid	100 mL
Distilled H ₂ O	450 mL
Brilliant Blue R	2.5 g

1 L Destain

Component

Ethanol	100 mL
Acetic acid	100 mL
Distilled H ₂ O	800 mL

Biophysical Methods

This chapter describes the theoretical basis and experimental details for the biophysical techniques used in this thesis: SEC-MALLS (Size Exclusion Chromatography coupled to Multi-Angle Laser Light Scattering), sedimentation equilibrium AUC (Analytical Ultracentrifugation), and differential scanning fluorimetry.

3.1 SEC-MALLS

SEC-MALLS is a technique for measuring the masses of polymers in solution, using light scattering and refractive index measurements. This allows protein oligomeric state to be determined, by comparing the measured mass to the monomer molecular weight predicted from sequence. In this section, I will describe the theoretical basis of MALLS, the principles of the MALLS detectors, and practical aspects of the technique as applied in this thesis.

3.1.1 Theoretical basis for molecular weight measurement using MALLS

MALLS exploits the fact that both the degree of light scattering and the refractive index of a polymer solution vary with concentration and molecular weight, but with different dependencies. The refractive index depends only on the mass concentration of the polymer, and is given by the equation

$$n(c) = n_{solvent} + \left(\frac{dn}{dc}\right) c \quad [3.1]$$

where n is the refractive index, c is the protein mass concentration, and dn/dc is the difference in refractive index per concentration unit of protein.

For proteins, $\left(\frac{dn}{dc}\right)$ lies in a fairly narrow range between 0.18 and 0.20 mL g⁻¹, and is generally not measured directly (Zhao, Brown and Schuck, 2011). $\left(\frac{dn}{dc}\right)$ does also depend on the refractive index of the solvent, but this is usually neglected, despite the strong dependence of n on salt concentration (Tan and Huang, 2015).

The light scattered by a sample of (possibly oligomeric) protein is proportional to both the mass concentration and the molecular weight. I will first give a physical explanation, then briefly describe a more mathematical derivation.

Consider two solutions of a particular protein with the same concentration, but different oligomeric states: one is dimeric, while the other is monomeric (perhaps because it has a mutation on its dimerisation interface). As a light wave passes through the solution of monomers, it will induce charge polarisation in each protein molecule, causing it to act as a driven oscillating dipole.

In the monomeric protein solution, each dipole will give rise to a wave with amplitude $aM \left(\frac{dn}{dc} \right)$, where M is the protein molecular weight, and a is a constant. The waves from each dipole add together, but because the proteins are randomly distributed through space, the waves are not in phase with one another and the overall amplitude is proportional to $\sqrt{c}aM \left(\frac{dn}{dc} \right)$. The intensity is the square of the amplitude, so the overall intensity of scattered light is proportional to $ca^2M^2 \left(\frac{dn}{dc} \right)^2$

In the dimeric sample, each protomer will again generate a wave with amplitude $aM \left(\frac{dn}{dc} \right)$. But this time, the waves from the two halves of each dimer will interfere constructively, so each dimer gives rise to a wave of amplitude $2aM \left(\frac{dn}{dc} \right)$. The molar concentration of dimer is half the monomer concentration, so these sum to a total amplitude of $\sqrt{\frac{c}{2}} 2aM \left(\frac{dn}{dc} \right)$, giving a final intensity of $2ca^2M^2 \left(\frac{dn}{dc} \right)^2$. So, the dimeric solution scatters twice as much light as the monomeric solution, despite having the same protein concentration.

A full mathematical treatment of light scattering from a polymer is given by Zimm (Zimm, 1948), and summarised by Wyatt (Wyatt, 1993). Briefly, we start by assuming a polymer is made up of n identical scattering segments, which are distributed relative to one another according to a function $\rho(r)$. This distribution function accounts both for the probability of finding a segment within the same polymer chain at a distance r, and the probability of finding a segment from a different polymer chain. The scattering from a solution with N polymer molecules in a volume V is then given by the Raleigh-Gans-Debye approximation:

$$I(\theta) = \frac{K^* n^2 N^2}{V^2} \int \rho(r) \exp\left(\frac{2\pi i s \cdot r}{\lambda}\right) dr \quad [3.2]$$

where $I(\theta)$ is the excess Raleigh scattering at an angle theta between the incident and scattered rays, K^* is a constant determined by the refractive indexes of the

solvent and polymer (Wyatt, 1993), \mathbf{s} is the vector difference between the incident and scattered ray direction vectors, and λ is the wavelength of incident light.

By separately considering the parts of the distribution function representing intra- and intermolecular distances, and assuming the radial distribution function for a random coil, Zimm derives

$$I(\theta) = \frac{K^*n^2N}{v} \left[P(\theta) + \frac{Nn^2X}{v} P^2(\theta) \right] \quad [3.3]$$

where $P(\theta)$ is an integral of the random coil radial distribution function, and X is a constant integral (Zimm, 1948). This is more commonly expressed as

$$I(\theta) = K^*McP(\theta)[1 - 2A_2McP(\theta)] \quad [3.4]$$

where A_2 is the second Virial coefficient (Wyatt, 1993). This makes clear that the scattered intensity is proportional to both the molecular weight and the mass concentration, provided that neither is high enough for the second term to become significant.

A number of approximations and assumptions are made in this derivation.

The Raleigh-Gans-Debye approximation only holds when

$$\left| \frac{n_{protein}}{n_{solvent}} - 1 \right| \ll 1 \quad [3.5]$$

and

$$2a\pi\lambda \left| \frac{n_{protein}}{n_{solvent}} - 1 \right| \ll 1 \quad [3.6]$$

where $n_{protein}$ is the refractive index of the protein molecules, $n_{solvent}$ is the refractive index of the solvent, and $2a$ is the characteristic diameter of the protein. Respectively, these mean the protein must not significantly slow the light wave as it passes through the solution, and the protein must not significantly affect the phase of the propagating light wave. These conditions are both met because the refractive index of the protein is similar to that of the solvent ($(\frac{dn}{dc}) \sim 0.18 \text{ mL g}^{-1}$), and the protein is much smaller than the wavelength of light used (658 nm).

The protein is assumed to be made up of identical segments with a random coil distribution. This is clearly not the case for folded proteins, but is a reasonable approximation considering that the scattered light is an average over all orientations of the protein, and that the scattering depends only on the locations

of scattering segments, not their connectivity. It is possible that this assumption could break down for rod-like or other non-globular proteins: however, this does not apply to the domains studied in this thesis, which are expected to adopt a globular dsRBD fold.

In order to account for interactions between nearby molecules, the derivation assumes that they touch only at a single point. This appears to be a reasonable model for transient non-specific interactions between proteins in close proximity. In any case, this assumption only affects the second term in the formula for scattering. Although protein oligomers interact via a surface and not a single point, that is unimportant here because the entire oligomer acts as a single 'molecule' over the timescale of each scattering measurement.

3.1.2 Practical aspects of data collection and analysis

100 μL of protein sample at 3 mg/mL was passed over a Superdex 75 10/30 analytical gel filtration column (GE Healthcare) pre-equilibrated in 20 mM MES, 200 mM NaCl, 1 mM DTT, pH 6.5, at a flow rate of 0.5 mL/min. Light scattering and refractive index measurements were recorded using in-line Wyatt Dawn HELEOS-II and Wyatt rEX Optilab detectors, and analysed using ASTRA software version 5.3.4.14 (Wyatt Technology).

Refractive index detectors work by passing a beam through a flow cell at an angle, so that refraction causes a change in direction. The displacement of the beam relative to a reference beam can then be used to calculate the refractive index. The Wyatt rEX Optilab elaborates on this by passing the beam through a diffraction grating, and inferring the displacement from the resulting diffraction pattern as measured by a photodiode array. This is claimed to result in higher accuracy through combining the signals from multiple photodiodes, as well as an increased range of refractive indexes (Wyatt Technology, 2005).

The Wyatt Dawn HELEOS-II instrument consists of 18 detectors arranged around a cylindrical flow cell, as shown in Figure 3.1. Each detects light scattered in a particular direction from a volume of the flow cell. Depending on the exact positioning, the volume 'seen' by each detector can vary, which would lead to systematic errors in the measured scattering intensity per unit volume. This was corrected for by performing SEC-MALLS on a sample of bovine serum albumin (BSA, Fisher Scientific), and normalising the signals of the different detectors to

counteract any differences. BSA is sufficiently small that the scattering should have no angular dependence.

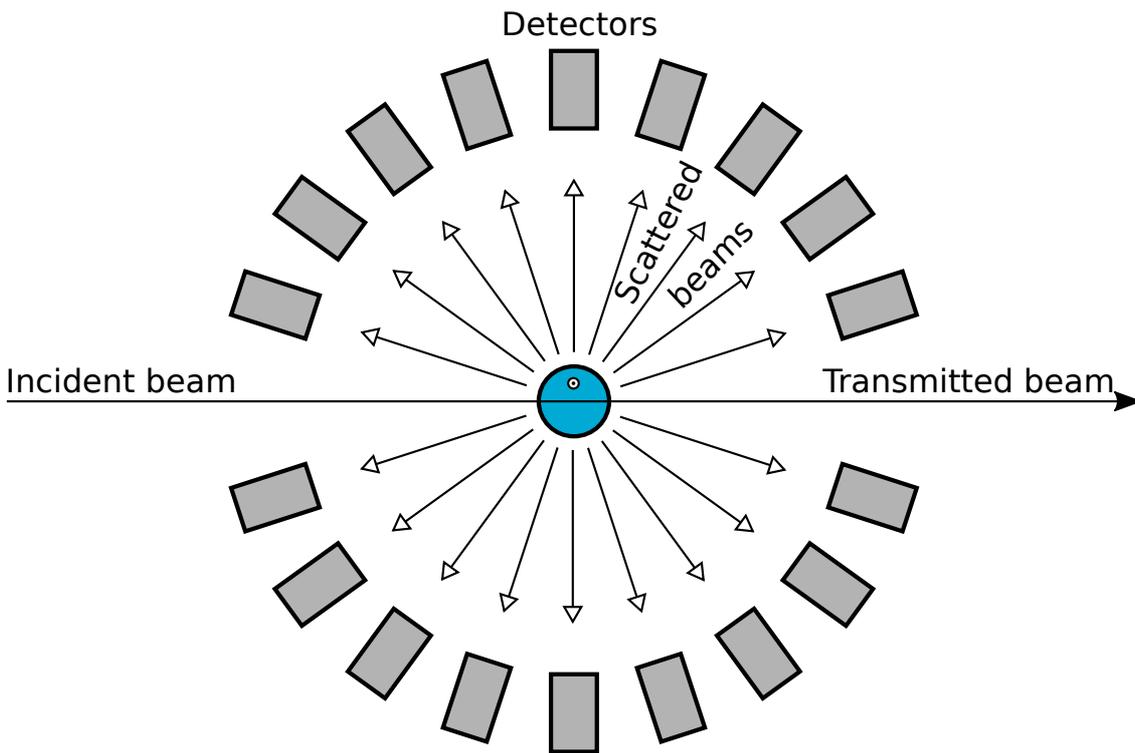


Figure 3.1. Schematic of the Wyatt Dawn HELEOS-II light scattering detector.

The sample flows perpendicular to the page through a cylindrical flow cell (shown in blue). A laser beam (polarised perpendicular to the page) is incident on the cuvette: the intensity of scattered light is measured by a series of photodiode detectors (grey) arranged around the flow cell.

As previously mentioned, fitting the molecular mass requires knowledge of $\left(\frac{dn}{dc}\right)$. I did not make direct measurements of $\left(\frac{dn}{dc}\right)$: instead, I found the value of $\left(\frac{dn}{dc}\right)$ that gave the expected molecular weight of BSA (66 kDa), after normalisation of the light scattering detectors as described above. This value and normalisation was then used for all subsequent samples run on that day. Normalisation and $\left(\frac{dn}{dc}\right)$ estimation was performed every time the instrument was used. $\left(\frac{dn}{dc}\right)$ was found to be between 0.176 and 0.184.

As each peak elutes from the column and passes through the detectors, multiple measurements of the refractive index and light scattering are made. The calculated mass varied continuously over the width of the peak, but generally plateaued near the centre. The deviations at the edges of the peak are likely due to systematic

errors that become more significant when the signal is low, although in some cases deviations can be attributed to sample impurities. Therefore, only the central region of the peak was used to estimate the molar mass.

Both the light scattering and refractive index detectors have high precision, so random measurement errors do not contribute significantly to the error in calculated molecular weight. Instead, errors are mostly caused by violations of the assumptions used to derive the expression for the intensity of scattered light. Of these, the assumption that $\left(\frac{dn}{dc}\right)$ is similar across all proteins and the assumption of a single species contributing to scattering seem most likely to be violated. The overall error in calculated molecular weight has been estimated empirically to be 5-10% for the SEC-MALLS system used in this thesis (Andrew Leech, personal communication).

3.2 Sedimentation equilibrium analytical ultracentrifugation (seAUC)

Sedimentation equilibrium analytical ultracentrifugation is a method of determining molecular weight and equilibrium association properties of proteins based on their sedimentation properties. During centrifugation, molecules experience a force

$$F = M \left(1 - \frac{\bar{v}}{\rho}\right) \omega^2 r \quad [3.7]$$

where M is the protein molar mass, \bar{v} is the partial specific volume of the protein, ρ is the solvent density, ω is the angular speed and r is the radial distance from the rotor axis. This force causes them to move towards the outer edge of the rotor until the centrifugal force is balanced by diffusion, forming a concentration gradient. The analytical ultracentrifuge is equipped with a spectrophotometer that allows the absorbance to be measured at different radii, allowing the radial distribution of protein to be calculated. This data can then be used to calculate molecular weight and association constants, and additionally detect the presence of aggregates or non-ideality.

When a protein sample is centrifuged, it distributes itself in accordance with the Boltzmann distribution, with higher concentrations in regions with lower (more

negative) potential energy (Brautigam, 2011). It can be shown that the equilibrium distribution of protein is given by:

$$\frac{\ln(c(r))}{r^2} = M \left(1 - \frac{\bar{v}}{\rho}\right) \frac{\omega^2}{2RT} \quad [3.8]$$

where $c(r)$ is the mass concentration at r , R is the gas constant and T the temperature. Of the variable parameters, ω and T are controlled by the experimenter, while \bar{v} and ρ can be estimated accurately with knowledge of the protein sequence and buffer composition respectively. This allows the molar mass to be calculated for a single species in solution simply by plotting $\ln(c(r))$ against r^2 .

For solutions containing a mixture of oligomeric states, or with a mixture of different proteins, more complicated data analysis is required. For analysing a mixture of monomer and dimer, the full equation for the radial concentration distribution is given by:

$$c(r) = c_{monomer}(r_0)e^{M\left(1-\frac{\bar{v}}{\rho}\right)\frac{\omega^2(r^2-r_0^2)}{2RT}} + c_{dimer}(r_0)e^{2M\left(1-\frac{\bar{v}}{\rho}\right)\frac{\omega^2(r^2-r_0^2)}{2RT}} + B \quad [3.9]$$

where B corrects for any baseline offset in the data. This equation is fitted to the data using non-linear least-squares fitting methods, in this case using a modified version of the Gauss-Newton method (Johnson *et al.*, 1981).

Accurately fitting a sum of exponentials is in general an ill conditioned problem, meaning that small changes in the input data can lead to large changes in the optimal fit. To counteract this problem, the experiment is repeated at different rotor speeds and with different protein concentrations. The latter of these is particularly important when estimating dissociation constants.

3.2.1 Practical aspects of data collection and analysis

A dilution series of PACT-D3 was prepared in 20 mM MES, 200 mM NaCl, pH 6.5, giving concentrations of 3, 1.5, 0.75, 0.38, 0.19 and 0.09 mg/mL. 118 μ L of each sample was loaded into each well of a seAUC cell, with 120 μ L of buffer in a matched reference well. Unfortunately, the buffer used for dilution and reference was from a different batch from the buffer in the protein sample, leading to baseline offsets in the data. This was accounted for during data analysis by including a baseline offset term in the fitted model, but the extra degrees of freedom from this term would be expected to degrade the accuracy of the fit.

A Beckman Optima XL/I ultracentrifuge was used to spin samples at velocities of 22,000, 24,000, 26,000, 33,000, 38,000 and 42000 rpm, while recording absorption profiles at 250 and 280 nm. Measurements were taken every 4 hours, and each velocity was maintained for 20 hours to ensure time to reach equilibrium. The temperature was maintained at 20°C.

Data pre-processing was carried out using custom R scripts, to determine whether equilibrium had been reached, and to remove regions where the absorbance was too high to measure accurately. Non-linear least-squares fitting to a single species model (refer to previous equation number) was then performed using Origin software. The program SEDNTERP was used to calculate the partial specific volume and buffer density, giving values of 0.7285 mL g⁻¹ and 1.00742 g mL⁻¹ respectively.

3.3 Differential Scanning Fluorimetry

Differential scanning fluorimetry measures the stability of a protein sample, by monitoring the fluorescence of SYPRO orange dye as the sample is heated (Reinhard et al., 2013). As the protein denatures, newly exposed hydrophobic regions interact with the dye, generating a fluorescent signal with maximum emission intensity at a wavelength of 570 nm (Simpson, 2010). The melting temperature (T_m) is calculated by fitting a sigmoidal curve to the fluorescence vs temperature data, and taking the temperature of half-maximum intensity.

Melting temperature is used here as a proxy for the less well-defined “protein stability”, under the assumption that a higher melting temperature indicates a structure that is less prone to deviations from the native fold. The melting temperature clearly depends upon both protein sequence and structure, but the relationship is insufficiently well understood to make accurate melting temperature predictions (Franzosa, Lynagh and Xia, 2010). The difficulty in predicting melting temperature may stem from the need to account for both enthalpic and entropic contributions from the solvent, and from the ensemble of unfolded states.

Protein oligomerisation is also known to increase protein stability (Goodsell and Olson, 2000; Tanakai *et al.*, 2004; Hashimoto and Panchenko, 2010), but the magnitude of the effect on melting temperature is again unpredictable (Pollegioni

et al., 2003; Shallom *et al.*, 2004). That said, it is clear than the effect depends on the dimer dissociation constant, and that no stabilisation will be seen at concentrations significantly below this.

Proteins were purified as described in Chapter 2, using 20 mM Tris, 200 mM NaCl, pH 7.5 as the size exclusion buffer, and 1:400 SYPRO orange dye (ThermoFisher Scientific) added. Stock solutions of the trial buffers were prepared, each containing 100 mM buffering agent and sufficient NaCl so that the final concentration (once mixed with protein) would be 50 mM, 100 mM, 200 mM or 500 mM. Then, 6 μ L of 1 mg/mL protein stock was mixed with 24 μ L buffer stock in a 96 well PCR plate, with three replicates of each condition. Using a Stratagene Mx3005P thermocycler (Agilent), the plate was then heated from 25°C to 95°C at a rate of 2°C/min, with fluorescence measurements taken every 30 s.

The fluorescence data was then exported and analysed with a custom R script, based on the previously reported MTSA program (Schulz, Landström and Hubbard, 2013). Briefly, the script extracts the region of the fluorescence curve which represents the protein unfolding transition, and fits a function of the form:

$$F(T) = F_{min} + \frac{\Delta}{\left(1 + e^{\frac{T^* - T}{b}}\right)^c} \quad [3.10]$$

where F is fluorescence intensity, T is the temperature and F_{min} , b, c and T^* are constants. F_{min} and $(F_{min} + \Delta)$ are constrained to lie within 2% of the maximum and minimum values of the raw data. The melting temperature is calculated as the temperature at which the fluorescence is at half maximum (see Figure 3.2).

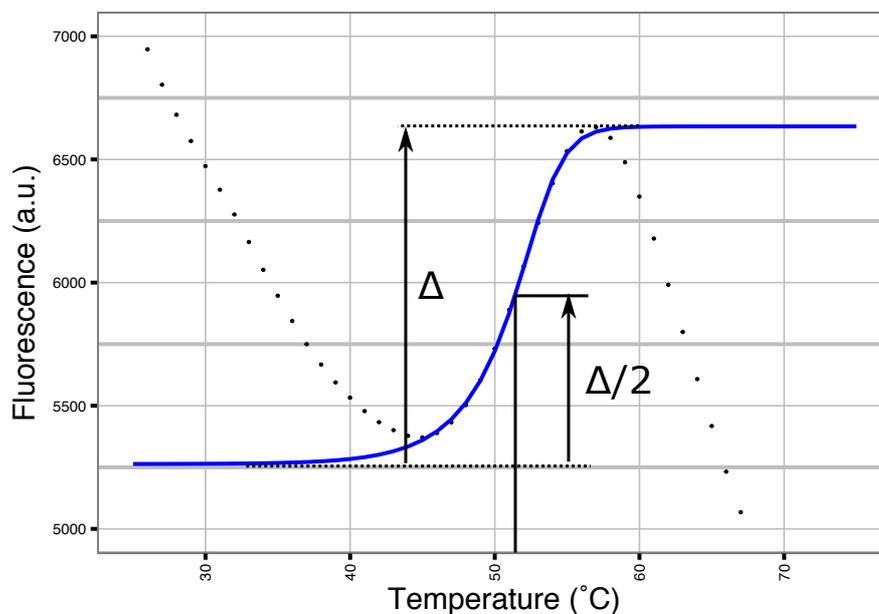


Figure 3.2. Example differential scanning fluorimetry curve

Fluorescence measurements are shown as points, while the fitted curve is shown in blue. Δ is the difference between the fitted maxima and minima. The melting temperature is the point at which the fluorescence is halfway between the fitted minimum and maximum.

There are several confounding factors that affect the shape of the fluorescence curve. Firstly, in some cases there is a high level of background fluorescence from the SYPRO orange, which reduces as the temperature is increased. This can occur if the protein has hydrophobic surface patches or if it is already partially unfolded (Reinhard et al., 2013). Secondly, after reaching a maximum, the fluorescence intensity decreases again at higher temperatures. It has been suggested that this is due to aggregation of the unfolded protein, or perhaps because of increased thermal motion of the dye (Reinhard *et al.*, 2013). These factors make it difficult to accurately determine the start and end points of the transition, preventing a reliable determination of melting temperature. However, it is sufficient for determining whether there is a difference in melting temperature between two conditions, providing the size of the confounders is not too great.

NMR Methods

4.1 Introduction

This chapter describes the basic principles of NMR spectroscopy, then gives a brief summary of the experiments used in this thesis. So far as possible, these are described in physical terms, without giving detailed formulae: a more mathematical description can be found elsewhere (Cavanagh *et al.*, 2006; Levitt, 2008). A more detailed description is then given of the experimental setup and parameters, and of subsequent data processing and analysis.

4.2 Basic principles of NMR spectroscopy

4.2.1 Physical basis of the NMR signal

All nuclei have a quantum-mechanical property called spin, which is determined by the arrangement of protons and neutrons within the nucleus. Nuclei of a given isotope have either zero (^{12}C , ^{16}O), $\frac{1}{2}$ (^1H , ^{15}N , ^{13}C , ^{19}F), or higher order spin quantum numbers. Nuclei with spin $\frac{1}{2}$ act as magnetic dipoles, so can interact with electromagnetic fields. In particular, a nuclear dipole oriented perpendicular to an external magnetic field will precess at its Larmor frequency:

$$\omega = -\gamma B$$

where γ is the gyromagnetic ratio for that nucleus, and B is the magnetic field strength.

In the absence of an external magnetic field, the energy of each nuclear dipole is independent of its orientation. Therefore, the dipoles in a sample will be distributed isotropically, leading to zero net magnetisation. When an external field is applied, the energy becomes dependent on the angle between the dipole and the field: spins oriented parallel to the magnetic field have a lower energy than those oriented antiparallel, resulting in a net magnetisation parallel to the field at thermal equilibrium.

In the simplest form of Fourier-transform NMR spectroscopy, a radio-frequency (r.f.) pulse at the Larmor frequency causes all spins to rotate by 90° , leading also to a rotation of net magnetisation into a plane perpendicular to the external field. As the spins precess, the net magnetisation precesses with them, and can be

detected by a receiver coil within the probe. Other experiments utilise a more complex initial series of pulses and delays, but detect the final signal in the same way. Over time, the precessing spins lose coherence (transverse relaxation) and return to thermal equilibrium (longitudinal relaxation). These processes are discussed further in the sections below.

4.2.2 Intramolecular contributions to the magnetic field

The above description neglects several crucial factors that affect the overall magnetic field experienced by each nucleus, which allow different sites within a molecule to be distinguished, and enable the system to return to thermal equilibrium after a perturbation. I have chosen to discuss the effects of these factors in terms of magnetic fields; for a more complete description in terms of the spin Hamiltonian, see (Levitt, 2008).

The first of these factors is the chemical shift effect: the external magnetic field sets up currents in the electrons of the molecule, which then create an additional induced magnetic field. The magnitude (and direction) of the induced field depends on the local electronic configuration, and is proportional to the external field. The induced field is therefore different for each chemically distinct nucleus in a molecule. A tensor is required to fully describe the chemical shift, but often only the isotropic chemical shift is reported. Chemical shift effects are generally larger for nuclei surrounded by more electrons, such as heavier or more electronegative atoms, or chemical groups with delocalised electrons.

The second factor is direct dipole-dipole coupling, whereby each nucleus is affected by the dipole fields of nearby spin $\frac{1}{2}$ nuclei. These fields depend on the orientation of the internuclear vector with respect to the external field, and the magnitude of the interaction between spins i and j is given by the dipole-dipole coupling constant:

$$b_{jk} = -\frac{\mu_0 \hbar \gamma_j \gamma_k}{4\pi r^3} \quad [4.1]$$

where γ_j and γ_k are the gyromagnetic ratios of the two spins, r is the separation between them, and μ_0 and \hbar are physical constants.

A third factor is indirect dipole-dipole coupling, also known as scalar coupling or J coupling. This occurs due to the interaction of nuclear spins with the spins of the electrons in chemical bonds. Consider a pair of identical nuclei separated by a

single bond. In the absence of the nuclear spins, the antiparallel bonding electrons will have identical distributions. If we introduce a nuclear spin at one site, one of the electrons will be aligned so as to interact favourably with the nucleus, while the other will interact unfavourably. As a result, the electron distributions change so that the favourably aligned electron is closer to the nuclear spin, while the unfavourably aligned one is further away. The second site therefore feels uneven effects from the two electrons, resulting in a net magnetic field whose direction depends on the orientation of the nuclear spin at the first site. The strength of the scalar coupling is independent of the external field, and is usually measured in terms of the difference in Larmor frequency between the favourably and unfavourably aligned states. Scalar couplings are exploited to transfer magnetisation between bonded nuclei, as in the Heteronuclear Single Quantum Coherence (HSQC) experiment described in Section 4.3.1.

4.2.3 Relaxation and the effects of molecular tumbling

All three of these contributions to the overall magnetic field vary as the molecule tumbles in solution. In isotropic liquids, the rotational correlation time is generally much faster than the sampling timescale of the NMR experiment (on the orders of 10^{-9} s and 10^{-4} s respectively). Therefore, in many cases we can treat the nuclear spins as experiencing a constant field that is the average over all possible molecular orientations. It can be shown that the rotational average direct dipole-dipole interaction is zero, but that chemical shift and J-coupling do have isotropic components (Levitt, 2008). This averaged field alters the Larmor frequency for each nucleus, allowing different sites in the molecule to be distinguished.

Although the effects of chemical shift and dipole-dipole coupling are averaged on long timescales, at shorter timescales they result in each nucleus experiencing a fluctuating magnetic field. This fluctuating field is responsible for both transverse and longitudinal relaxation. The timescale, or correlation time, of the fluctuations is determined by the rotational diffusion coefficient of the molecule, and thus increases with molecular weight. More precisely, the fluctuations can be decomposed into components with different frequencies, described by a spectral density function. Intramolecular motions can introduce fluctuations on additional timescales: these can be described using the model-free approach (Lipari and Szabo, 1982). In this method, the correlation function is divided into two

independent parts, representing overall molecular motion and internal motions. The internal part is characterised by the timescale of motions τ_e , and by the S^2 order parameter. S^2 measures the long-term correlation of the magnetic field fluctuations due to local motion, and represents how restricted that motion is (with a completely rigid region having $S^2 = 1$, and a completely flexible region having $S^2 = 0$).

Transverse relaxation (also known as T_2 or spin-spin relaxation) is due to fluctuations parallel to the external field. These cause the total field at each nucleus to vary, resulting in slightly different precession rates. Therefore the nuclei do not precess synchronously, but instead gradually lose coherence (dephase), resulting in a reduction in net magnetisation. The rate of transverse relaxation is strongly dependent on the correlation time of the fluctuations: rapid changes in field tend to cancel each other out, while slower ones lead to a greater degree of dephasing. Therefore, the transverse relaxation rate (T_2) increases as molecules become larger and tumble more slowly (Figure 4.1).

In contrast, longitudinal relaxation (T_1 or spin-lattice relaxation) occurs due to fluctuations in any direction. These fluctuations can act in a similar way to the applied r.f. field, rotating the nuclear spins by small amounts. These random perturbations tend to return the spins to thermal equilibrium, restoring net magnetisation parallel to the external field. Only fluctuations with timescales similar to the Larmor frequency are effective at rotating the spins, while fluctuations on faster or slower timescales do not effectively contribute to longitudinal relaxation. Therefore, the longitudinal relaxation rate (T_1) is low for small molecules, increases for molecules around the size of a small protein, and decreases for larger macromolecules (Figure 4.1).

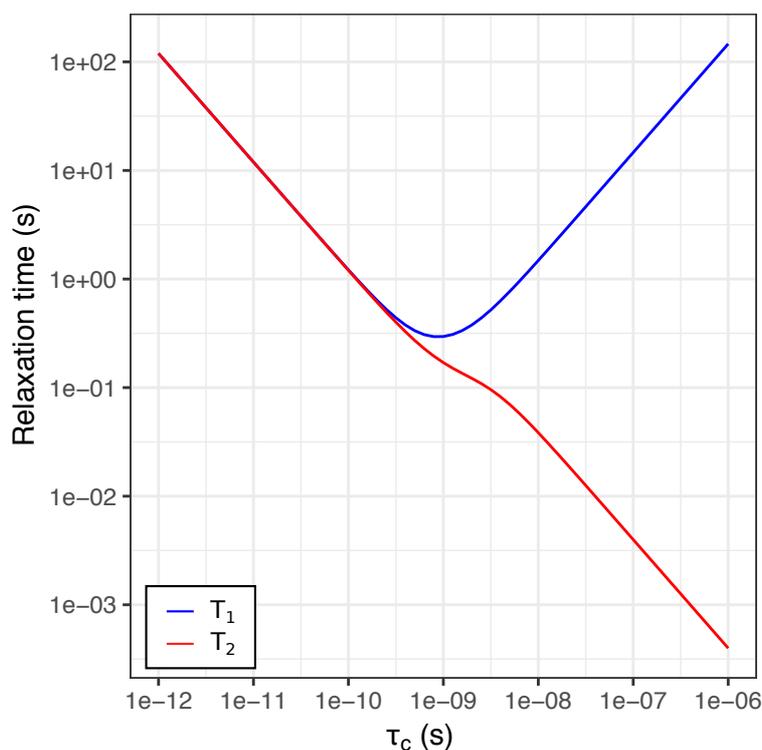


Figure 4.1 Relaxation rates depend on rotational correlation time

As molecules increase in size, their rotational correlation time increases. This leads to a steady decrease in T_2 , while T_1 initially decreases before rising again. The curve is plotted for amide N-H bonds, assuming that dipolar coupling is the only contribution to relaxation (Levitt, 2008, p. 564). Fortuitously, the minimum in T_1 is similar to the rotational correlation time of small proteins (5-20 ns) (García de la Torre, Huertas and Carrasco, 2000).

4.2.4 The Nuclear Overhauser Effect

The Nuclear Overhauser Effect (NOE) is a manifestation of direct dipole-dipole coupling, and affects how the net magnetisation parallel to the external field evolves following perturbations. The details are challenging to explain in terms of the vector model used so far, and are perhaps better understood in the more abstract terminology of energy levels and spin flips (Levitt, 2008, p. 566).

Consider a pair of nuclei, I and S, coupled only by the direct dipole-dipole interaction. This system has 4 states (both spin down, both spin up, and two mixed states), with equilibrium populations determined by the energy differences between them. Now suppose that a selective pulse is used to invert spin I (this is the basis of the transient NOE experiment). The populations are now out of equilibrium, and will relax via a number of possible transitions.

If the spins were not coupled, the only pathway would be via single-quantum transitions corresponding to flipping of spin I – the net magnetisation of spin S would remain unaffected. However, the direct dipole-dipole coupling means that the magnetic field fluctuations experienced by the two nuclei are correlated, allowing the possibility of zero-order transitions (where antiparallel I and S spins flip simultaneously) and second-order transitions (where parallel I and S spins flip simultaneously). These transitions do affect the net magnetisation of spin S, leading to an increase of magnetisation for large molecules, and a decrease for small ones. Importantly, the probability of these cross-relaxation transitions is proportional to r^{-6} , so the NOE is restricted to nuclei in close proximity to one another. This can be exploited to generate short-range (<5 Å) distance restraints by systematically measuring the NOE between pairs of nuclei using NOESY experiments.

4.3 Overview of NMR experiments

4.3.1 (^1H , ^{15}N) heteronuclear single quantum coherence (HSQC) experiment

The (^1H , ^{15}N)-HSQC is one of the most important experiments for the study of proteins by solution state NMR spectroscopy, because it allows the proton and nitrogen chemical shifts of the backbone amide groups to be correlated (Bodenhausen and Ruben, 1980; Cavanagh *et al.*, 1991; Kay, Keifer and Saarinen, 1992). Provided the amino acid sequence does not have a high proline content, these backbone amides are evenly distributed throughout the entire protein, and can be used as a basis for backbone assignment, characterisation of relaxation and exchange, and determination of secondary structure. A wide variety of important NMR experiments are based on the (^1H , ^{15}N)-HSQC, including the 3D HNC(O), CBCANH, CBCA(CO)NH, H(CCO)NH-TOCSY, ^{15}N -NOESY-HSQC experiments used in this thesis.

A simplified schematic of the overall (^1H , ^{15}N)-HSQC pulse sequence is shown in Figure 4.2A. The sequence begins with transfer of magnetisation from the amide proton to the nitrogen using an INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer) pulse element (Morris and Freeman, 1979). The one-bond scalar couplings of backbone amide groups fall in a narrow range around -90 Hz, so it is possible to achieve near optimal magnetisation transfer to almost all backbone nitrogen spins (Rule and Hitchens, 2006, p. 137). Following a period of

nitrogen chemical shift evolution, polarisation is transferred back to the amide proton through a reverse INEPT element, and the proton free induction decay is directly detected.

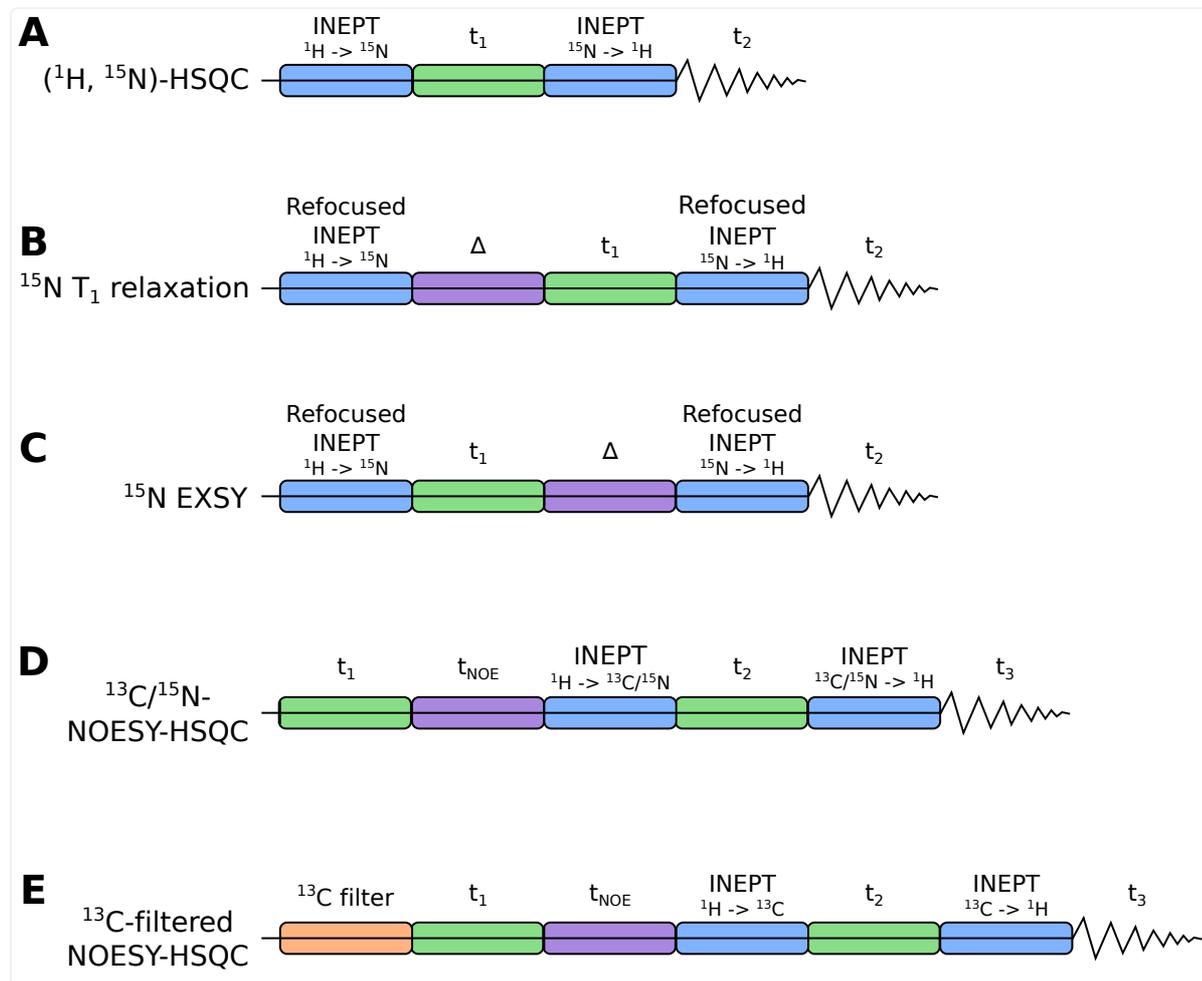


Figure 4.2. Overview of NMR pulse sequences

The main functional blocks are shown for each pulse sequence. INEPT polarisation transfers are depicted in blue, variable delays for chemical shift evolution of the indirect dimension in green, periods of pure z magnetisation in purple, and other elements in orange. Acquisition of the free induction decay is displayed as a zigzag.

A) $(^1\text{H}, ^{15}\text{N})$ -HSQC. B) ^{15}N longitudinal relaxation measurement. C) ^{15}N EXSY. D) $^{13}\text{C}/^{15}\text{N}$ -NOESY-HSQC. E) ^{13}C -filtered NOESY-HSQC.

If magnetisation was completely transferred without losses, the volume of each peak in the resulting (^1H , ^{15}N)-HSQC spectrum would be directly proportional to the abundance of that amide group. However, each amide group has different relaxation properties, disrupting the proportionality. Firstly, transverse relaxation causes magnetisation to be lost during the INEPT transfers, resulting in an approximately 7% reduction in signal for a typical 10 kDa protein (Rule and Hitchens, 2006, p. 210). This reduction will be larger in residues with fast transverse relaxation rates.

Secondly, incomplete longitudinal relaxation reduces the initial polarisation of each spin. Generally the delay between scans is 1-1.5 s, which is too short for equilibrium to be reached. This reduces the initial magnetization available for transfer and hence the peak intensity in the spectrum. Globally, this loss of intensity is more than offset by the increased signal to noise ratio from the higher scan rate, but differences in longitudinal relaxation rate still cause relative reductions in the intensity of peaks from slowly relaxing amides.

4.3.2 ^{15}N longitudinal relaxation measurements and Exchange Spectroscopy (EXSY)

The (^1H , ^{15}N)-HSQC experiment can be extended to measure the ^{15}N longitudinal relaxation rate by adding an additional delay (Δ) before the nitrogen evolution period t_2 (Figure 4.2B,C). To remove the influence of the amide proton, refocused INEPT elements are used to generate pure ^{15}N polarisation along the z-axis during Δ . The ^{15}N nucleus is used because it does not exchange with the solvent, and because Nuclear Overhauser Effects (NOEs) between ^{15}N nuclei are negligible (due to its small gyromagnetic ratio).

In the absence of slow chemical exchange, the only factor affecting the ^{15}N spins will be longitudinal relaxation. To determine the longitudinal relaxation rate, a series of spectra are recorded with different relaxation delays (typically up to about 1 second), and an exponential model fitted to the peak height. To ensure an accurate measure of the relaxation rate, it is critical that the delay between scans is sufficiently long that all spins can return to equilibrium (generally 3-5 times T_1). If this is not done, the initial magnetisation will be different for each delay, causing systematic errors in the peak height and hence the estimate of T_1 . Note that both

the longitudinal relaxation and ^{15}N EXSY experiments should give identical spectra in this case.

If the protein is in slow exchange ($k_{ex} \ll \sim 50 \text{ s}^{-1}$ (Mittermaier and Kay, 2009)), transitions between states can occur during the relaxation delay. This means that a more complicated model is required to fit the data (described in section 4.4.2 below), although simplifications are possible if the two states can be assumed to have equal relaxation rates. Attempting to fit exchange rates from the ^{15}N longitudinal relaxation experiment alone is an ill-conditioned problem, as was encountered for the seAUC data fitting in Chapter 3. The advantage of the EXSY experiment is that the nitrogen evolution period is before Δ , so signals from exchanged and non-exchanged spins appear at different positions in the spectrum. This helps separate the effects of relaxation and exchange, allowing their rates to be fitted more accurately.

4.3.3 Isotope-filtered NOESY

NOESY experiments exploit dipolar coupling to transfer polarisation between spins separated by less than approximately 5 \AA via the Nuclear Overhauser Effect, described above (Wüthrich, 1990). The variants most commonly used for protein structural characterisation are the ^{13}C - and ^{15}N -NOESY-HSQC experiments (Figure 4.2 D,E). Again, these are extensions of the HSQC experiment, but include an additional evolution and transfer period to measure the chemical shifts of nearby protons.

The NOE transfer occurs during the period t_{NOE} , in which the proton magnetization is aligned with the z-axis. This arrangement is similar to that of the EXSY experiment, and as a result, chemical exchange can also be detected in the spectrum (Rule and Hitchens, 2006, p. 411). This is important to consider when analysing NOESY data of proteins in slow exchange between two or more states, as misinterpretation of exchange peaks as NOE peaks will lead to erroneous distance restraints.

^{13}C -filtered NOESY is an extension of standard NOESY experiments that enables identification of intermolecular contacts (Otting and Wüthrich, 1989). In its standard form, a [^{15}N , ^{13}C]-labelled protein and its unlabelled binding partner are mixed. In this system, intramolecular NOEs always occur between protons bound to the same isotope (that is, $^1\text{H}^{13}\text{C}$ to $^1\text{H}^{13}\text{C}$ or $^1\text{H}^{12}\text{C}$ to $^1\text{H}^{12}\text{C}$). In contrast, proton

pairs linked by intermolecular NOEs are always bound to different isotopes (ie. $^1\text{H}^{12}\text{C}$ to $^1\text{H}^{13}\text{C}$ or $^1\text{H}^{13}\text{C}$ to $^1\text{H}^{12}\text{C}$). These intermolecular NOEs can be picked out by pulse sequence elements that filter out all NOE donors bound to ^{13}C , and select (edit) only for NOE acceptors bound to ^{13}C .

The editing step is achieved effectively by the ^{13}C evolution period t_2 , but the initial ^{13}C filtering is more difficult, due to the wide range of chemical shifts and scalar coupling constants of aliphatic groups. The approach used here selectively purges protons bonded to ^{13}C using an adiabatic wideband inversion pulse (Zwahlen *et al.*, 1997), while leaving protons bonded to ^{12}C unaffected. This is not completely effective, as it relies on an empirical correlation between the carbon chemical shift and one-bond C-H scalar coupling: aliphatic groups which deviate from this relationship are not fully suppressed. Fortunately, these residual NOEs originating from $^1\text{H}^{13}\text{C}$ pairs can be identified due to the evolution of scalar coupling during t_1 , which causes them to appear as doublets with a splitting of approximately 140 Hz.

4.4 Details of NMR data collection

4.4.1 Sample preparation

Unless otherwise stated, NMR samples were prepared as follows. Purified protein samples were dialysed overnight against 20 mM MES, 50 mM NaCl, 10 mM TCEP, pH 6.5. A 600 μL NMR sample was prepared by mixing 540 μL protein solution, 60 μL D_2O and 0.6 μL 50 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), and loaded into a 5 mm thin wall NMR tube (Wilmad). This gives a final buffer composition of 18 mM MES, 45 mM NaCl, 9 mM TCEP, 10% D_2O , 50 μM DSS, pH 6.5, and is referred to as “standard NMR buffer” throughout this thesis.

Spectrometers

Five spectrometers were used for data collection, and are summarised in Table 4.1:

Table 4.1. NMR spectrometer details

Spectrometer	Field (MHz)	Probe	Abbreviation
Bruker Avance II	700	triple-resonance room temperature probe	York_700
Bruker Avance II+	600	triple-resonance cryoprobe	MRC_600
Bruker Avance III HD	700	quadruple-resonance cryoprobe	MRC_700
Bruker Avance III HD	800	triple-resonance cryoprobe	MRC_800
Bruker Avance I	800	triple-resonance room temperature probe	Shef_800

Unless otherwise stated, all NMR experiments were performed at 25°C, with a recycle delay of 1 second.

4.4.2 (^1H , ^{15}N) HSQC spectra

All (^1H , ^{15}N) HSQC spectra (with the exception of those recorded at high pressure) were recorded on the York_700 spectrometer with a triple-resonance room temperature probe, using pulse sequence hsqcetf3gpsi. The nitrogen and proton offsets were 118 ppm and approximately 4.7 ppm respectively; spectral widths of either 28 ppm or 32 ppm were used for nitrogen, and 16 ppm for proton. 2048 real and imaginary points were collected in the proton dimension, and 128 real and imaginary points in the nitrogen dimension.

4.4.3 Assignment of PACT-D3 L273R

A 1.1 mM sample of [^{15}N , ^{13}C]-labelled PACT-D3 L273R in 18 mM MES, 45 mM NaCl, 5 mM TCEP, 10% D₂O, 50 μM DSS, pH 6.5 was used for backbone assignment. CBCANH (cbcanhgpwg3d), CBCA(CO)NH (cbcaconhgpwg3d) and HNCO (hncogpwg3d) spectra were recorded using the York_700 spectrometer.

Table 4.2. Experimental parameters for assignment of PACT-D3 L273R

Pulse sequence		Offset (ppm)	Spectral width (ppm)	# points (real + imaginary)
cbcanhgpwg3d	¹ H	4.68	14	2048
	¹⁵ N	118	28	64
	¹³ C	39	75	128
cbcaconhgpwg3d	¹ H	4.68	14	2048
	¹⁵ N	118	28	64
	¹³ C	39	75	128
hncogpwg3d	¹ H	4.68	14	2048
	¹⁵ N	118	28	64
	¹³ C	173	12	32

4.4.4 Assignment of wild-type PACT-D3

Experimental data for backbone assignment of wild-type PACT-D3 was collected using the MRC_600 spectrometer, on a 1.3 mM [¹⁵N, ¹³C]-labelled protein sample in standard NMR buffer. HNCACB (hncacbgpwg3d) and CBCA(CO)NH (cbcaconhgpwg3d) spectra were recorded with the parameters given in Table 4.3 below.

For side chain assignment, a [¹⁵N, ¹³C]-labelled sample at 2.4 mM was used. The York_700 spectrometer was used to record HNCO (b_hncogp3d), H(CCO)NH (hccconhgpwg3d2) and C(CO)NH (ccconhgp3d) spectra, while the MRC_800 spectrometer was used to record constant-time ¹³C-HSQC (hsqcctetgppsp) and HC(C)H-TOCSY (hcchdigp3d) spectra, using the parameters given in Table 4.3 below. For the b_hncogp3d, an offset of 8.3 ppm was used for the band-selective pulse, with a bandwidth of 3.5 ppm. Non-uniform sampling (NUS) was used for the HC(C)H-TOCSY: 25% of the points were collected, distributed so that the earlier times were sampled more frequently (exponentially weighted, with time constants of 10 ms for the indirect proton dimension and 20 ms for the indirect carbon dimension).

Table 4.3. Experimental parameters for assignment of PACT-D3

Pulse sequence		Offset (ppm)	Spectral width (ppm)	# points (real + imaginary)
hncacbgpwg3d, cbcaconhgpwg3d	¹ H	4.7	14	2048
	¹⁵ N	118	28	80
	¹³ C	39	75	200
b_hncogp3d	¹ H	4.70	14	2048
	¹⁵ N	118	28	56
	¹³ C	176	16	32
hccconhgpwg3d2	¹ H	4.69	14	2048
	¹⁵ N	118	28	56
	¹ H	4.69	14	128
ccconhgp3d	¹ H	4.69	14	2048
	¹⁵ N	118	28	56
	¹³ C	40	70	128
hsqcctetgppsp	¹ H	4.73	14	2048
	¹³ C	39	70	400
hcchdigp3d	¹ H	4.75	14	2048
	¹³ C	39	70	256
	¹ H	2.9	7.5	256

4.4.5 T₁ relaxation and EXSY spectra

T₁ relaxation data was recorded using the York_700 spectrometer, on a 2.2 mM [¹⁵N]-labelled sample of PACT-D3 in standard NMR buffer. The pulse sequence hsqct1etf3gpsi3d was used, with randomly-ordered relaxation delays of 0.01, 0.05, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.51, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5 and 2.0 seconds. A series of EXSY spectra were recorded of a 1.4 mM [¹⁵N]-labelled sample of PACT-

D3, with 10 equally-spaced, randomly-ordered exchange delays between 0.1 and 1.0 seconds. The York_700 spectrometer was used with pulse sequence hsqcetexf3gp. All other parameters were kept the same as for (^1H , ^{15}N) HSQC spectra, except that a recycle delay of 3 seconds was used.

4.4.6 NOESY spectra

The MRC_800 spectrometer was used to record ^{15}N -NOESY-HSQC (noesyhsqcf3gpwg3d) and ^{13}C -NOESY-HSQC (noesyhsqcetgp3d) spectra of PACT-D3. The ^{15}N -NOESY-HSQC sample had a concentration of 1.1 mM, while the ^{13}C -NOESY-HSQC sample was at 2.4 mM; both samples were [^{15}N , ^{13}C]-labelled, and in standard NMR buffer. The NOESY mixing time was set to 120 ms. 21% NUS was used for the ^{15}N -NOESY-HSQC, and 25% NUS for the ^{13}C -NOESY-HSQC.

The MRC_700 spectrometer was used to record a ^{13}C -filtered NOESY-HSQC spectrum (Zwahlen *et al.*, 1997) of a sample containing 700 μM [^{15}N , ^{13}C]-labelled and 700 μM ^{15}N -labelled PACT-D3 in standard NMR buffer.

Table 4.4. Experimental parameters for NOESY experiments on PACT-D3

Pulse sequence		Offset (ppm)	Spectral width (ppm)	# points (real + imaginary)
noesyhsqcf3gpwg3d	^1H	4.73	14	2048
	^{15}N	118	32	128
	^1H	4.73	14	512
noesyhsqcetgp3d	^1H	4.69	14	2048
	^{13}C	39	39	128
	^1H	4.69	14	460
^{13}C -filtered NOESY-HSQC	^1H	4.7	16	2048
	^{13}C	39	39	52
	^1H	47	13	192

4.4.7 NMR experiments at high pressure

300 μ L samples of 0.5 mM wild-type PACT-D3 or 0.75 mM PACT-D3 L273R were loaded into a 5 mm ceramic pressure cell, overlaid with paraffin oil and pressurised using an Xtreme-60 syringe pump (Daedalus Innovations). (^1H , ^{15}N)-HSQC spectra were then recorded at pressures of 1 bar, 0.5 kbar, 1.0 kbar, 1.5 kbar, 2.0 kbar and 2.5 kbar on the Shef_800 spectrometer. Offsets were (4.70 ppm, 118 ppm), spectral widths were (12.5 ppm, 28 ppm) and the number of real and imaginary points collected were (2048, 128) for the proton and nitrogen dimensions respectively.

4.5 NMR data analysis

4.5.1 (^1H , ^{15}N) HSQC processing

NMR data was processed either using NMRpipe (Delaglio *et al.*, 1995), or in some cases using TopSpin version 3 (Bruker Biospin). For (^1H , ^{15}N)-HSQC spectra, the following processing scheme was used:

- 1) Direct dimension
 - a. High pass filtering to remove residual solvent signal.
 - b. Apodisation with a quadratic sine function to reduce influence of noise at later timepoints.
 - c. Fourier transformation.
 - d. Polynomial baseline correction.
- 2) Indirect dimension
 - a. Apodisation with a sine bell function.
 - b. Zero-filling to 512 points, to enhance the digital resolution.
 - c. Fourier transformation.
 - d. Polynomial baseline correction.

CCPN Analysis version 2 was used for all peak picking and assignment (Vranken *et al.*, 2005).

4.5.2 Fitting of longitudinal relaxation and exchange data

Peaks were picked and assigned manually using CCPN Analysis version 2, and peak heights estimated using a parabolic model. Only residues with 4 well-resolved, non-overlapping cross peaks in the EXSY spectra were analysed. Exchange rates for individual residues were determined by nonlinear least-

squares fitting, using a Gauss-Newton algorithm. The equations describing the exchange model are given below (Farrow, Zhang, *et al.*, 1994; 3rd, Kroenke and Loria, 2001):

$$I_{AA}(t) = I_0 C_A \frac{(\lambda_2 - R_A - k_{ex})e^{\lambda_1 t_{mix}} + (\lambda_1 - R_A - k_{ex})e^{\lambda_2 t_{mix}}}{\lambda_2 - \lambda_1} \quad [4.2]$$

$$I_{BB}(t) = I_0 C_B \frac{(\lambda_2 - R_B - k_{ex})e^{\lambda_1 t_{mix}} + (\lambda_1 - R_B - k_{ex})e^{\lambda_2 t_{mix}}}{\lambda_2 - \lambda_1} \quad [4.3]$$

$$I_{AB}(t) = I_0 C_A k_{ex} \frac{e^{-\lambda_1 t_{mix}} - e^{-\lambda_2 t_{mix}}}{\lambda_2 - \lambda_1} \quad [4.4]$$

$$I_{BA}(t) = I_0 C_B k_{ex} \frac{e^{-\lambda_1 t_{mix}} - e^{-\lambda_2 t_{mix}}}{\lambda_2 - \lambda_1} \quad [4.5]$$

$$I_A(t) = \frac{I_1}{I_0} (I_{AA} + I_{BA}) \quad [4.6]$$

$$I_B(t) = \frac{I_1}{I_0} (I_{BB} + I_{AB}) \quad [4.7]$$

$$\lambda_1 = \frac{1}{2} (R_A + R_B + 2k_{ex} + \sqrt{(R_A - R_B)^2 + 4k_{ex}^2}) \quad [4.8]$$

$$\lambda_2 = \frac{1}{2} (R_A + R_B + 2k_{ex} - \sqrt{(R_A - R_B)^2 + 4k_{ex}^2}) \quad [4.9]$$

Where:

I_{AA} and I_{BB} are the heights of the auto peaks for states A and B in the EXSY spectra;

I_{AB} and I_{BA} are the heights of the exchange peaks in the EXSY spectra;

I_A and I_B are the peak heights for states A and B in the T_1 relaxation spectra;

I_0 and I_1 are scaling factors for the EXSY and T_1 relaxation spectra respectively;

C_A and C_B are account for initial differences in intensity of states A and B;

R_A and R_B are the longitudinal relaxation rates of states A and B;

k_{ex} is the exchange rate between states A and B;

t_{mix} is the mixing time allowed for relaxation and exchange.

Note that the model as given above assumes that the rate of conversion from A to B and from B to A are equal, and only accounts for relaxation that occurs during the mixing time, not during other parts of the pulse sequence. This would be expected to give equal populations to each, but in practice deviations occur,

presumably due to relaxation in other parts of the pulse sequence. The parameters C_A and C_B account for this difference in initial population.

The model is also intended to describe peak volumes, not peak heights. In practice, the estimates of peak height and peak volume produced by CCPN Analysis are highly correlated for non-overlapping peaks; therefore, either can be used. The high correlation suggests that peak volumes are not being accurately calculated within CCPN Analysis, as height to volume ratios should vary with peak width.

To calculate the global exchange rate, peak heights from the EXSY spectra were fit to the following model (Miloushev *et al.*, 2008) using least-squares in R:

$$\mathcal{E}(t) = \frac{I_{AB}(t)I_{BA}(t)}{I_{AA}(t)I_{BB}(t) - I_{AB}(t)I_{BA}(t)} \cong k_{ex}^2 t^2 \quad [4.10]$$

The calculation was restricted to the data used for the per-residue fitting above. The 95% confidence interval was calculated using bootstrapping. Note that the quadratic relationship between $\mathcal{E}(t)$ and t breaks down at longer time points, as higher order terms become significant.

4.5.3 Fitting of pressure-induced chemical shift changes

The reference frequencies of the (^1H , ^{15}N) HSQC spectra were first corrected using the chemical shift of the internal DSS standard in a proton spectrum recorded at each pressure. Peaks were manually assigned using CCPN analysis using a spectrum recorded at ambient pressure as a reference. Peaks which could not be unambiguously assigned, due to overlap with other peaks, were excluded from further analysis. Chemical shift values were exported, and fit to either a linear or quadratic model using least-squares in R. The two fitting methods did not result in significantly different linear coefficients.

Cell and Molecular Biology Methods

5.1 Mammalian cell culture

All cells were grown in Dulbecco's modified Eagle medium (Gibco) supplemented with 10% Foetal Bovine Serum (Hyclone), 2mM L-Glutamine, 50 U/ml penicillin and 50 µg/ml streptomycin (henceforth referred to as "full media"), at 37°C and 5% CO₂. Cells were split every 3-4 days, and reseeded at an approximate density of 2x10⁶ (for a 10cm plate).

5.2 Knockdown of PACT by siRNA

ON-TARGET plus SMART pool against human PRKRA (Dharmacon) was used to target PACT, while ON-TARGET plus control siRNA pool (Dharmacon) was used as a control. Cells were seeded in 6 well plates at a density of 5-8 x10⁴ cells per well 16 hours prior to transfection. 2 µL siQuest reagent (Mirus Bio) and siRNA (Dharmacon) were separately mixed with Opti-MEM+Gluta-MAX medium; the two solutions were then mixed so as to give a total volume of 160 µL for each well to be transfected (containing 1.25 % (v/v) transfection reagent and 250 nM siRNA), and incubated at room temperature for 20 minutes. The media in each well was replaced with 640 µL Opti-MEM, then 160 µL of siRNA transfection mix was added dropwise. After 6 hours, the transfection mixture was replaced with full media, and cells were harvested 48 hours later.

5.3 Construction of vectors for PACT overexpression

5.3.1 Cloning PACT constructs into pGEM-T

Full-length PACT, PACT-Ext-D3 and PACT-D3 were amplified from pOTB7 lentiviral expression vectors using the GoTaq G2 Flexi kit (Promega) according to the manufacturer's instructions, with 2 mM MgCl₂ included in the reaction buffer. The thermal cycle was 2 min at 95°C; 29 cycles of 30 s 57°C, 1 min 72°C, 30 s 95°C; and finishing with 30 s at 57°C and 5 mins at 72°C. Primers are shown in table 1, and were designed to introduce a Bsp119I site at the 5' end, and an XbaI site at the 3' of the PCR product. Regions matching/complementary to PACT are shown as uppercase, restriction sites and extensions in lowercase. The PCR products were

then ligated into pGEM-T vectors (Promega) using T4 DNA ligase, and verified by sequencing.

Table 5.1. Primers used for cloning PACT into pGEM-T vector

Name	Sequence	T _m /°C
PACT forward	aatgaattcgaacATGTCCCAGAGCAGGCACC	64.8
PACT-D3 forward	aatgaattcgaacACAGATTACATCCAGCTGC TTAGTGAAATTG	63.1
PACT-Ext-D3 forward	aatgaattcgaacGGACATTCTTTAGGATGTA CTTGGCATTTCCTTG	64.6
PACT reverse	acaacttctagaTTACTTTCTTTCTGCTATTA TCTTTAAATACTGCAAAGCA	62.1

5.3.2 Construction of c-myc tag expression vectors

Oligonucleotides encoding c-myc and flanking sticky-ended NotI and BSP199I restriction sites (see table two) were annealed by heating to 95°C for 5 minutes, followed by a slow cooling to room temperature. pSF-CMV-PURO-NH2-FLAG vector (OG3213, Oxford Genetics) was then cleaved using NotI and Bsp199I, and purified using a Wizard SV Gel and PCR Cleanup Kit (Promega). The annealed oligos and linearized vector were then ligated using T4 DNA ligase, and verified by sequencing.

Table 5.2. Oligonucleotides used in construction of c-myc expression vector

Name	Sequence	T _m /°C
c-myc forward	GGCCGCAGGAGGTACTCACGATGGCGGAACAG AAACTGATTAGCGAAGAAGACCTGATGAATT	64.8
c-myc reverse	CGAATTCATCAGGTCTTCTTCGCTAATCAGTT TCTGTTCCGCCATCGTGAGTACCTCCTGC	63.1

5.3.3 Subcloning of PACT constructs into mammalian expression vectors

The pGEM-T/PACT plasmids (from section 5.3.1) and the FLAG and c-myc expression vectors (from section 5.3.2) were both digested with Bsp119I and XbaI, and purified from a 1% agarose gel. The linear vectors and PACT constructs were then mixed, ligated with T4 DNA ligase and transformed into XL10-Gold cells (Agilent). Colony PCR was used to screen colonies for the insert, using the same primers and thermocycle as in section 5.3.1. Plasmids from positive colonies were then verified by sequencing.

5.4 PACT overexpression in mammalian cells

5.4.1 Optimisation of GeneJuice transfection

HEK293 cells were seeded in a black-walled, clear-bottomed 96 well plate, with densities of either 1.0×10^4 , 1.5×10^4 , 2.0×10^4 or 2.5×10^4 cells per well, and grown in full media for 24 hours.

Transfection mix was prepared using GeneJuice (Novagen), Opti-MEM serum-free medium with Gluta-MAX (Gibco) and mKate2 plasmid (Evrogen) according to the manufacturer's instructions, in various ratios (either 60 ng or 120 ng plasmid per 30 μ L, with between 0 and 6 μ L GeneJuice per μ L DNA). 30 μ L of transfection mix was added dropwise to each well (with three replicates per condition), and incubated for 48 hours. Fluorescence was then measured using a BMG Labtech POLARstar OPTIMA plate reader, with excitation at 584 nm and detection at 620 nm.

5.4.2 Overexpression in HEK293 and HeLa cells

1.5×10^5 HEK293 cells per well were seeded into a 12 well plate, and grown for 24 hours in full media. GeneJuice transfection mix was prepared for each FLAG- or c-myc-tagged PACT construct according to the manufacturer's instructions, to give 0.5 μ g plasmid and 1.5 μ L GeneJuice in a total volume of 50 μ L. This was added dropwise to each well. Cells were grown for 72 hours, after which cells were cultured for 14 days in full media supplemented with 3 μ g/ml puromycin.

5.5 Extraction of protein and Western blotting

For a 6 well plate, each well was washed with 1 ml ice-cold PBS, then cells were lysed with 35 μ L RIPA buffer (25 mM Tris pH 7.5, 150 mM NaCl, 1% sodium

deoxycholate, 0.5% Triton X-100, 0.1 % SDS, 0.1% EDTA. The lysate was incubated on ice for 15 minutes, centrifuged at 10,000 RCF for 15 minutes at 4°C, and the supernatant extracted.

The protein concentration was measured using a Pierce BCA Protein Assay Kit (Thermo Scientific). A series of BSA solutions with concentrations equally spaced between 0 and 2 mg/ml was prepared as a reference. 5 µL of the BSA standards and protein extracts were loaded in duplicate into a 96 well plate: 200 µL of BCA reagent was added to each well, and the plate incubated at 37°C for 1 hour. The absorbance at 562 nm was then measured, and the protein concentration calculated according to a linear regression on the BSA standards.

Protein samples were diluted to 1 mg/ml, mixed with SDS-PAGE loading buffer, and run on a 15% SDS-PAGE gel as described previously in Chapter 2. Proteins were then transferred to a PVDF membrane (Millipore): the membrane was first washed in methanol for 1 minute, distilled water for 1 minute, and transfer buffer (20 mM Tris, 150 mM glycine, 20% methanol) for 3 minutes, then a Trans-Blot SD semi-dry transfer cell (BioRad) was used, with a transfer time of 90 minutes at 25 volts. The PVDF membrane was then blocked with 5% Marvel milk powder in TBS/T (10 mM Tris, 150 mM NaCl, 0.1% Tween-20, pH 8.0) for 1 hour, rinsed with TBS/T three times, then incubated for 16 hours at 4°C in 10 ml primary antibody solution. Following this, the membrane was washed three times for 5 minutes in TBS/T, then incubated with 6 ml secondary antibody solution for 1 hour, then again washed three times for 5 minutes in TBS/T. All wash/incubation steps were performed on a tube rotator/rocker.

To visualise the stained protein, the PVDF membrane was soaked in 4 ml ECL Western Blotting reagent (GE Healthcare) for 5 minutes, then visualised with Amersham Hyperfilm ECL (GE Healthcare). The film was then developed using a Compact X4 film processor (Xograph).

For re-probing with α -beta-actin, the membrane was stripped using Restore™ Western Blot Stripping Buffer (Thermo Scientific), then re-blocked and probed as above, except the incubation with primary antibody was 1 hour at room temperature.

Table 5.3. Antibodies used for Western blotting

Antibody	Description	Dilution	Manufacturer, catalogue number
α -hsPACT	Rabbit polyclonal	1:1000*	Abcam, ab31967
α -c-myc	Mouse monoclonal (9E10) IgG1	1:500*	Sigma-Aldrich, M4439
α -FLAG	Mouse monoclonal (M2) IgG1	1:5000*	Sigma-Aldrich, F3165
α -beta Actin	Mouse monoclonal (AC-15) IgG1	1:50000*	Abcam, ab6276
α -rabbit IgG	Goat polyclonal coupled to HRP	1:5000**	Dakocytomation, P0448
α -mouse IgG	Goat polyclonal coupled to HRP	1:5000**	Dakocytomation, P0447

* diluted with 4% BSA in TBS/T. ** diluted with 5% Marvel milk powder in TBS/T.

5.6 Extraction of RNA and qPCR

RNA was extracted using miRNeasy kits (Qiagen) according to the manufacturer's instructions. The concentration of the eluted RNA, as well as its 260/280 and 260/230 absorbance ratios, was quantified using a Nanodrop ND 1000 (Thermo Scientific).

cDNA of messenger RNAs was synthesised using a SuperScript II reverse transcriptase kit (Invitrogen). 1 μ L of RNA (100 ng/ μ L), 0.5 μ L oligo-dT (5 μ M), 1 μ L dNTP (10 μ M) and 9.5 μ L distilled water were mixed, heated to 70°C for 6 minutes, then cooled to 4°C. To each sample, 4 μ L 5x reaction buffer, 2 μ L DTT (0.1 M), 1 μ L RNase Out and 1 μ L SuperScript II reverse transcriptase was added. Samples were then heated to 42°C for 1 hour, followed by 70°C for 10 minutes. To synthesise miRNA cDNAs, 25 ng of RNA was reverse transcribed using TaqMan® Small RNA Assay kits (Life Technologies) for human miR-21 and U6, according to the manufacturer's instructions.

qPCR was performed on 1 μ L of cDNA in a total volume of 20 μ L. For all genes except GAPDH, primers and quenched fluorescent reporters from the appropriate TaqMan Small RNA Assay kit were used. For GAPDH, the primer pair 5'-GGAGTCAACGGATTTGGTCGTA-3' and 5'-GGCAACAATATCCACTTTACCAGAGT-3'

were used at a concentration of 300 nM, and SYBR Green (Applied Biosystems) was used for quantification. All samples were loaded in triplicate into a MicroAmp Fast Optical 96-well reaction plate (Applied Biosystems). The amplification reaction was performed on a StepOnePlus Real-Time PCR System (Applied Biosystems), with a thermocycle consisting of 50°C for 2 minutes, 95°C for 10 minutes, then 40 cycles of 95°C for 15 seconds, 60°C for 60 seconds. A common fluorescence threshold was set for all genes in the exponential part of the amplification curve, and used to calculate a C_t value for each well. The concentration was calculated as e^{2C_t} , and normalised to the value of U6 (for miR-21) or GAPDH (for PACT, TRBP and Dicer).

Table 5.4. Primers used for qRT-PCR

Gene	Type	Catalogue number
PACT	TaqMan	Hs00269379_m1
TRBP	TaqMan	Hs00998379_m1
Dicer	TaqMan	Hs00229023_m1
Hsa-miR-21	TaqMan	000397
U6 snRNA	TaqMan	001973

Production of recombinant PACT and TRBP domains

6.1 Introduction

There are a number of common methods for expressing proteins for use in structural and biophysical studies, including in *E. coli*, *Pichia pastoris*, insect cells, mammalian cells or using cell free expression systems. Each has its own strengths and weaknesses (see Table 6.2): for example, mammalian cells provide the most 'native' environment for protein expression and folding, but are expensive due to the specific media requirements, and typically produce a lower yield of protein.

Structural and biophysical studies typically require milligram quantities of highly purified protein, and may require isotope labelling. As both PACT and TRBP domain 3 are relatively small (approximately 75 residues for the core dsRBD, or 110 residues including the N-terminal extension), initial expression and solubility tests were carried out in *E. coli*. Two different construct lengths were tested: D3 covers only the predicted dsRBD, while Ext-D3 includes the conserved N-terminal region (see Chapter 1 section 1.3.2). Two phosphorylation sites have been identified in the N-terminal region of TRBP domain 3 (see Chapter 1 section 1.4.4), so a phospho-mimic construct was also tested.

Table 6.1. PACT and TRBP Domain 3 Constructs

	Residues	Length	Predicted molecular weight (kDa)
PACT-D3	239-313	75	8.1
PACT-Ext-D3	208-313	106	11.6
TRBP-D3	293-366	74	8.0
TRBP-Ext-D3	258-366	109	11.6
TRBP-Ext _{DD} -D3 *	258-366	109	11.6

* This construct contains S283D and S286D mutations.

6.2 Protein expression in *E. coli*

6.2.1 Solubility tests of protein constructs

To determine whether the constructs were expressed and soluble, 50 mL cultures of *E. coli* BL21 (DE3) were grown as described in Chapter 2. The cultures were then centrifuged, and the cell pellets lysed using a buffer containing lysozyme and Triton X-100 as previously described (Benoit and Plevin, 2013). This allowed the lysis of all samples to be performed in parallel, minimising differences between samples, and reducing the amount of time cell pellets were stored on ice.

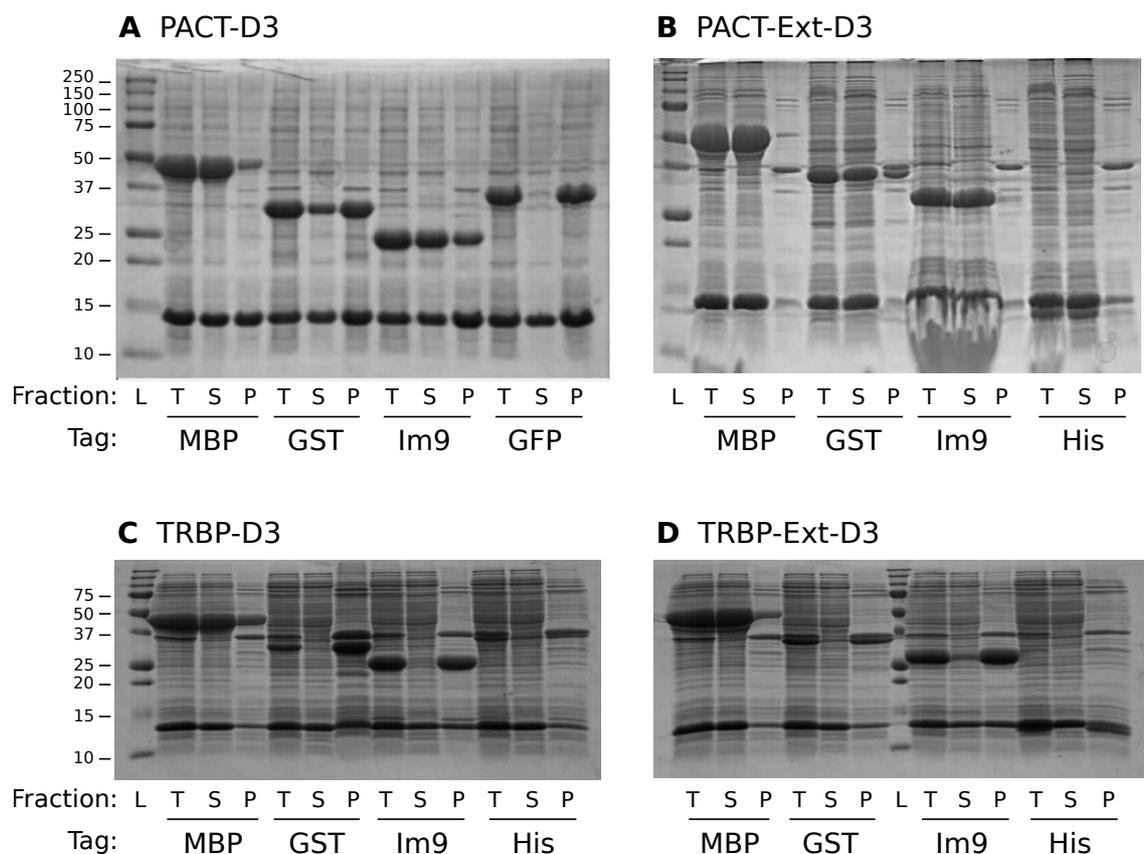


Figure 6.1. Solubility tests of PACT and TRBP domain 3 constructs

15 % SDS-PAGE gels showing the results of solubility tests for A) PACT-D3, B) PACT-Ext-D3, C) TRBP-D3, and D) TRBP-Ext-D3. For each tag, 3 lanes were run: total protein before centrifugation (T), supernatant after centrifugation (S), and pellet after centrifugation (P). Lane L contains the protein ladder, with masses given in kDa. The tags are Maltose Binding Protein (MBP), Glutathione-S-transferase (GST), *E. coli* Immunity Protein 9 (Im9) (James, Kleanthous and Moore, 1996), Green Fluorescent Protein (GFP) or His-tag only (His). The band just below 15 kDa in all lanes is lysozyme.

The solubility tests shown in Figure 6.1A,B demonstrate that both PACT-D3 and PACT-Ext-D3 constructs are soluble and highly expressed when tagged with either MBP or Im9. To a lesser extent, they are also soluble when GST-tagged. PACT-Ext-D3 appears to be soluble with only a His tag, but PACT-D3 is insoluble when GFP-tagged. Figure 6.1C,D show that TRBP-D3 and TRBP-Ext-D3 are soluble when fused to an MBP tag, but not to any of the other tags. TRBP-Ext_{DD}-D3 was not tested, but would be expected to behave similarly to wild-type TRBP-Ext-D3.

6.2.2 Scale up and purification

As described in Chapter 2, 1 L cultures of the MBP-tagged constructs were then grown in M9 minimal media containing [¹⁵N]-labelled ammonium chloride. Cell pellets were lysed using either continuous flow French press or sonication. Two nickel immobilised metal affinity purification steps were used to separate the tagged protein from endogenous *E. coli* proteins (Figure 6.2A), and to separate the tag from the domain of interest after 3C cleavage (Figure 6.2B). Finally, size exclusion chromatography (SEC) was used to remove any remaining contaminants (Figure 6.2C,D).

Following SEC, only very minor contaminants can be seen by SDS-PAGE (Figure 6.2D,E). Per litre of M9 media, typical yields were 15 mg PACT-D3, 10 mg PACT-Ext-D3, 1 mg TRBP-Ext-D3 and 5 mg TRBP-Ext_{DD}-D3. It was not possible to accurately measure the concentration of TRBP-D3 due to its small number of aromatic residues (Pace *et al.*, 1995). However, the yield appeared to be lower than for the other constructs, based on both Coomassie-stained SDS-PAGE gels (data not shown) and the NMR signal strength (see section 6.4). This is consistent with attempts to produce TRBP-D3 by Doudna and coworkers, who found that the N-terminal extension was necessary for TRBP domain 3 stability (Wilson *et al.*, 2015). Both TRBP-D3 and TRBP-Ext-D3 showed signs of aggregation during purification, such as white precipitate after centrifugation, and multiple peaks on the size exclusion chromatogram (data not shown).

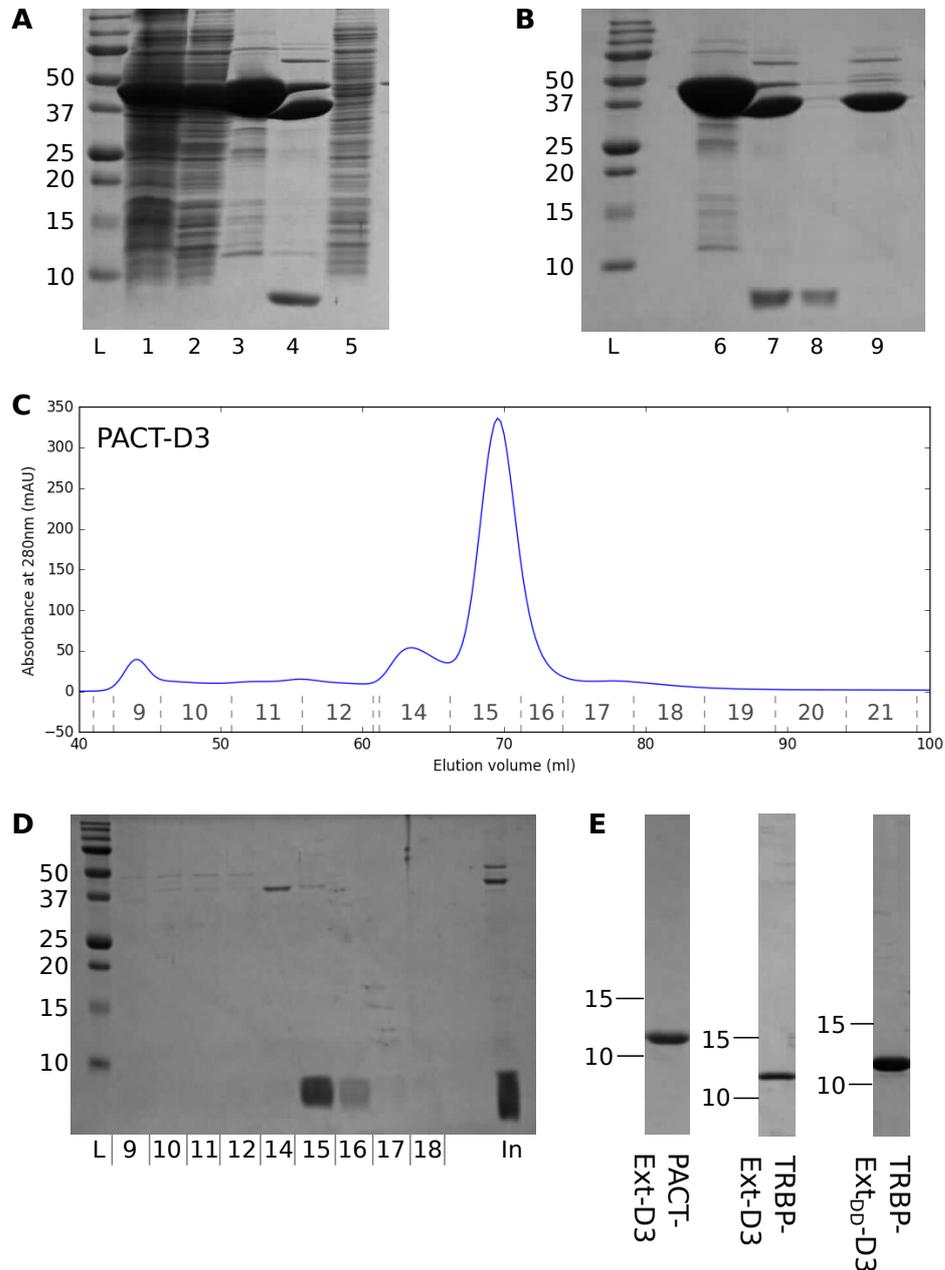


Figure 6.2. Large scale purification of PACT and TRBP domains

A) SDS-PAGE analysis of the lysis and first nickel affinity purification of PACT-D3. 1: Total cell lysate; 2: Soluble fraction; 3: Ni-IMAC eluate; 4: Eluate after dialysis with 3C protease; 5: Ni-IMAC flow through. L is the protein ladder, with masses in kDa. B) Second nickel affinity purification of PACT-D3. 6: First Ni-IMAC eluate; 7: Eluate after dialysis with 3C; 8: Second Ni-IMAC flow through; 9: Ni-IMAC eluate. C) Size exclusion profile of PACT-D3 passed over Superdex S75 16/60 column at a flow rate of 1 ml/min. D) SDS-PAGE analysis of size exclusion fractions of PACT-D3. L is the protein ladder; In is the input, diluted 1:5 E) SDS-PAGE analysis of concentrated samples of PACT-Ext-D3, TRBP-Ext-D3 and TRBP-Ext_{DD}-D3 following size exclusion chromatography.

6.3 Optimisation of size exclusion buffer for protein stability

Because there were signs of aggregation for several of the domains, differential scanning fluorimetry (DSF) was used to examine the stability of each protein in different buffers. Two main variables were tested: pH (between 5.5 and 8.5) and NaCl concentration (between 50 mM and 500 mM). In addition, various additives were examined (5% glycerol, 1 mM DTT, 0.5% (v/v) beta-mercaptoethanol and 1 mM tris(2-carboxyethyl)phosphine (TCEP)). Glycerol is frequently used to prevent protein aggregation, while DTT, beta-mercaptoethanol and TCEP are all reducing agents which prevent the formation of disulphide bonds. As PACT and TRBP are intracellular proteins, they are not expected to form intramolecular disulphide bonds *in vivo*, despite the extended third domains containing 4 and 7 cysteines respectively.

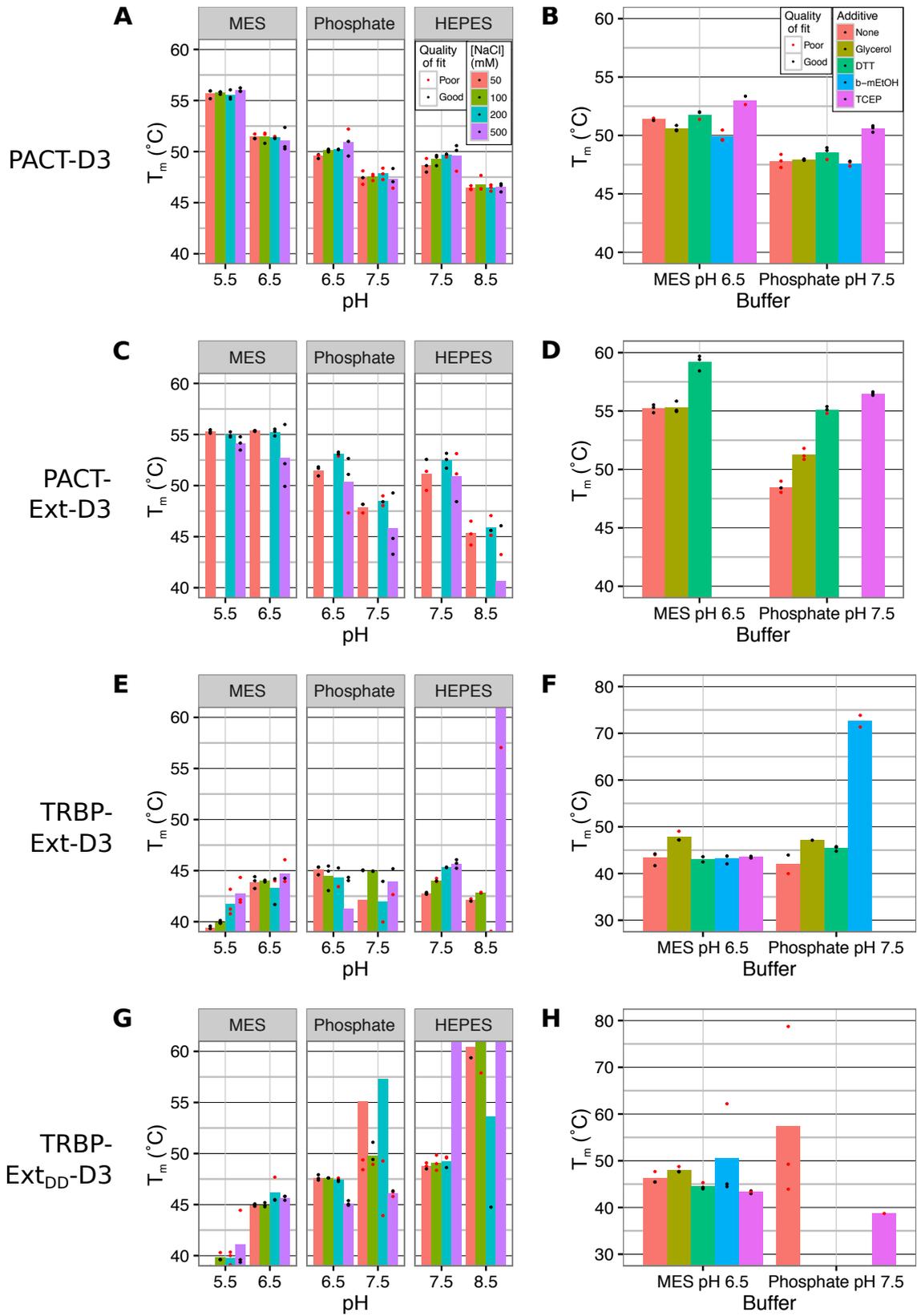
6.3.1 Differential scanning fluorimetry results

For PACT-D3 (Figure 6.3A,B), pH was found to have a large influence on stability, with the melting temperature at pH 5.5 almost 10°C higher than at pH 8.5. In contrast, salt concentration made very little difference to melting temperature. Both DTT and TCEP increased the melting temperature slightly, while glycerol either had no effect or slightly reduced it.

For PACT-Ext-D3 (Figure 6.3C,D), low pH was again found to increase melting temperature, while salt concentration had very little effect. DTT and TCEP were also found to increase melting temperature. This suggested that PACT-Ext-D3 would be stable in the same buffer as PACT-D3.

Figure 6.3. Differential Scanning Fluorimetry (DSF) to optimise buffer for protein stability.

A,C,E,G) Melting temperatures for each domain when pH and salt concentration are varied. B,D,F,H I) Melting temperatures when various cosolvents are added to either MES pH 6.5 or Na/K PO₄ pH 7.5, each with 200 mM NaCl. Points are the individual replicates, while bars show the average. Points are coloured red if the R² correlation between the data and fitted model is less than 0.999, indicating that the fit may be unreliable, and that the calculated melting temperature should be treated with caution.



It proved too difficult to produce sufficient TRBP-D3 to perform DSF, due to aggregation during the purification. When DSF was performed for TRBP-Ext-D3 and TRBP-Ext_{DD}-D3 (Figure 6.3E-H), several differences were apparent. Firstly, melting temperatures were in the range 40-45 °C, below the 45-55 °C range found for PACT-D3 and PACT-Ext-D3, indicating that the structure of TRBP-Ext-D3 is less stable. Secondly, many of the plots of fluorescence against temperature did not have the expected sigmoidal shape, instead showing additional 'shoulders'. This is reflected in the larger number of results marked as having a poor-quality fit. Therefore, a single-step unfolding transition is not a good fit to the data, suggesting that a more complex process is occurring. This could be interpreted as a two-step unfolding process, or could be due to the presence of multiple folded or partially-folded states. In either case, the melting temperatures calculated from this experiment should be treated with caution. Bearing this in mind, the melting temperature of TRBP-Ext_{DD}-D3 appears to be slightly higher than that of wild-type TRBP-Ext-D3.

Based on the data for PACT-D3 and PACT-Ext-D3, 20 mM MES, 200 mM NaCl, 10 mM TCEP, pH 6.5 was chosen as the working buffer for biophysical experiments. pH 6.5 was used rather than pH 5.5 because it is closer to the physiological pH range, and 200 mM NaCl was chosen to remain close to the physiological salt concentration of 150 mM. For many NMR experiments, the NaCl concentration was reduced to 50 mM to reduce sample conductivity, thereby increasing sensitivity (Kelly *et al.*, 2002). 10 mM TCEP was chosen to ensure that there was always a molar excess over the 3 cysteines of PACT-D3, even for highly concentrated samples. Because TCEP is relatively expensive, it was omitted from the size exclusion buffer, and the concentrated size exclusion fractions were dialysed against a smaller volume of buffer containing TCEP. While the DSF data suggests that TRBP-Ext-D3 is less stable than PACT-D3 in all the buffers examined, the buffer described above does not appear to be significantly worse than the others. Therefore, the same buffer was used for TRBP-Ext-D3 and TRBP-Ext_{DD}-D3.

6.4 NMR spectroscopy indicates folding state of PACT-D3 and TRBP-D3 constructs

To assess whether the domains were folded, 1D proton spectra were recorded of [¹⁵N]-labelled samples. The amide and methyl regions are particularly useful for

assessing folding state. In an unfolded protein, amide protons rapidly switch between different environments with a range of chemical shifts: the rapid exchange leads to observation only of the average frequency, resulting in peaks tightly clustered around 8 ppm. In a folded protein, the chemical environments of the amide protons are relatively fixed, and vary substantially between different sites, leading to peaks dispersed between approximately 6 and 10 ppm. Methyl protons in an unfolded protein typically have chemical shifts close to 1 ppm (Sahakyan *et al.*, 2011), but methyl groups in folded proteins can resonate at lower chemical shifts if in close proximity to an aromatic side chain.

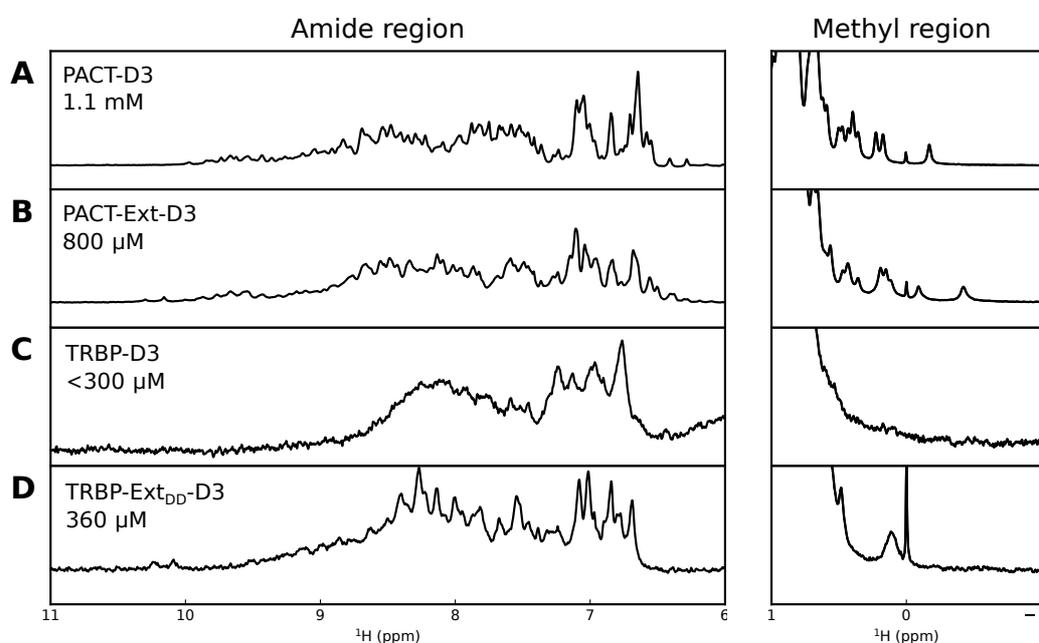


Figure 6.4. 1D ^1H spectra of [^{15}N]-labelled purified protein domains.

Data was recorded using Bruker pulse sequence *zgesgp* at 25°C on a Bruker AVANCE II 700 MHz spectrometer equipped with a triple-resonance room temperature probe. The spectra were then processed using TopSpin 3.2, using 5 Hz exponential line broadening. Protein concentration, determined by UV absorbance spectroscopy, is given at the top left of each spectrum. The concentration of TRBP-D3 was determined to be 300 μM but the intensity of the NMR spectrum suggests the true concentration is much lower. Note that ^{15}N decoupling was not used, so each amide resonates as a doublet.

For both PACT-D3 and PACT-Ext-D3 (Figure 6.4A,B), signals in the amide region are dispersed between approximately 6.5 and 10 ppm, indicating that both constructs are folded. In addition, both spectra contain signals with chemical shifts below 0 ppm, which are likely to represent methyl groups experiencing ring-current effects from aromatic groups. Taken together, these spectra provide strong evidence that both PACT-D3 and PACT-Ext-D3 are folded under these conditions. However, the spectra do not prove that the proteins have adopted the native fold, or that the sample is homogenous.

The 1D spectrum of TRBP-D3 (Figure 6.4C) has a lower signal to noise ratio than the other spectra, despite having significantly more scans (80, compared to 16 for the others). This is due to the low protein concentration, which is a consequence of the aggregation that occurred during purification. Signals in the amide region are weak, but appear to be clustered between 7.5 and 8.5 ppm, and there are no methyl signals below 0.5 ppm. The linewidths are broad, suggesting that the domain is forming aggregates with a long rotational correlation time.

The amide region of the TRBP-Ext_{DD}-D3 1D spectrum (Figure 6.5D) shows a broad peak between approximately 7 and 9 ppm, with a number of sharper peaks visible in the central region (the doublet at approximately 10 ppm most likely corresponds to the NεH amide group of the tryptophan side chain, and does not provide information about folding state). There is a methyl peak just above 0 ppm, but it is considerably broader than the methyl peaks of PACT-Ext-D3. Overall, these spectra suggest that TRBP-Ext_{DD}-D3 is folded, but experiences either aggregation or an intermediate exchange process that leads to considerable line broadening. The narrow, high intensity peaks in the centre of the amide region may suggest that part of the domain is flexible.

Biophysical Characterisation of PACT Domain 3

7.1 Introduction

PACT-D3 has previously been reported to homodimerise, and to heterodimerise with TRBP-D3, in Yeast-2-Hybrid and pull-down experiments (Laraki *et al.*, 2008; Singh *et al.*, 2011). Heterodimerisation between full-length PACT and TRBP has also been demonstrated in cells using FRET microscopy (Kok *et al.*, 2007). However, the interaction has never been demonstrated using purified protein, leaving open the possibility that this interaction is not direct, but is mediated via another protein or via RNA. The exact region required for PACT-D3 dimerisation has not been defined – the non-canonical dsRBD Staufen domain 5 dimerises via a pair of N-terminal helices, but this region is not conserved in PACT and TRBP (Gleghorn *et al.*, 2013). By contrast, the recently solved structure of Loquacious domain 3 (Loqs-D3) shows dimerisation mediated by the β -sheet region of the dsRBD, a region which shows considerable conservation with PACT and TRBP (Jakob *et al.*, 2016). In this chapter, wild-type and mutant forms of PACT domain 3 is investigated using SEC-MALLS and analytical ultracentrifugation in order to determine their oligomeric state and identify important residues for dimerisation.

7.2 Both PACT-D3 and PACT-Ext-D3 homodimerise

PACT-D3 and PACT-Ext-D3 were purified as described in Chapter 2, and examined by SEC-MALLS. In each case, 100 μ L of protein at 3 mg/mL was injected onto the column. Chromatograms are shown in Figure 7.1, and show a single dominant peak in each case. The calculated molecular weights are 16.6 kDa for PACT-D3, and 23.7 kDa for PACT-Ext-D3. These are very close to twice the expected molecular weights based on their amino acid sequence (8.5 kDa and 11.9 kDa, respectively). This shows that both constructs are almost entirely dimeric at this concentration. Both PACT-D3 and PACT-Ext-D3 eluted at a volume of 12.3 mL, suggesting that the N-terminal extension does not significantly increase the hydrodynamic radius of the dimer.

For PACT-Ext-D3, a small but wide peak elutes prior to the main peak, with an estimated mass of 50 kDa (Figure 7.1, lower). This is likely to represent a small amount of aggregated protein, as the molar mass estimate continues to increase towards the start of the peak. For both proteins, a number of peaks elute between 15 ml and 20 ml, but do not produce a light scattering signal. These are due to a mismatch between the running and sample buffers, and do not provide information about protein oligomeric state.

That PACT-D3 is able to dimerise in the absence of the N-terminal extension demonstrates that it must dimerise via a different mechanism to that shown for Staufen-D5. This is consistent with the lack of sequence conservation between the N-terminal regions of these two domains.

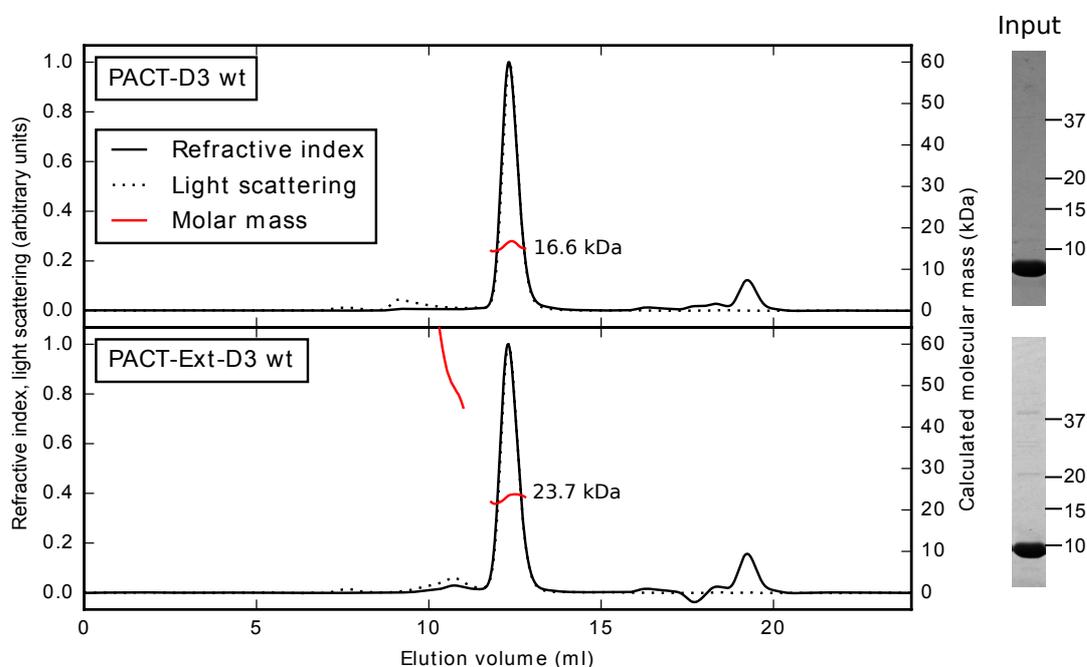


Figure 7.1. SEC-MALLS of wild-type PACT-D3 and PACT-Ext-D3

Differential refractive index is displayed as a solid line; the Rayleigh ratio (a measure of light scattering) is shown as a dotted line. The weight-averaged molar mass calculated over the centre of the peak is displayed next to each peak. SDS-PAGE analysis of the input is shown on the right.

7.3 Sedimentation equilibrium AUC confirms that PACT-D3 is dimeric

To further validate the oligomeric state, I examined PACT-D3 using sedimentation equilibrium analytical ultracentrifugation (seAUC). As described in Chapter 3, different concentrations of PACT-D3 were loaded into an An-50 Ti rotor, and centrifuged at speeds between 22,000 and 42,000 rpm until equilibrium was achieved. If only a single molecular species is present, a plot of $\ln(\text{absorbance})$ against r^2 should show a straight line. Therefore, this type of plot is commonly used to assess the overall data quality, and check for aggregation and non-ideal behaviour, such as that resulting from non-specific electrostatic interactions. Figure 7.2 shows a selection of these plots for PACT-D3.

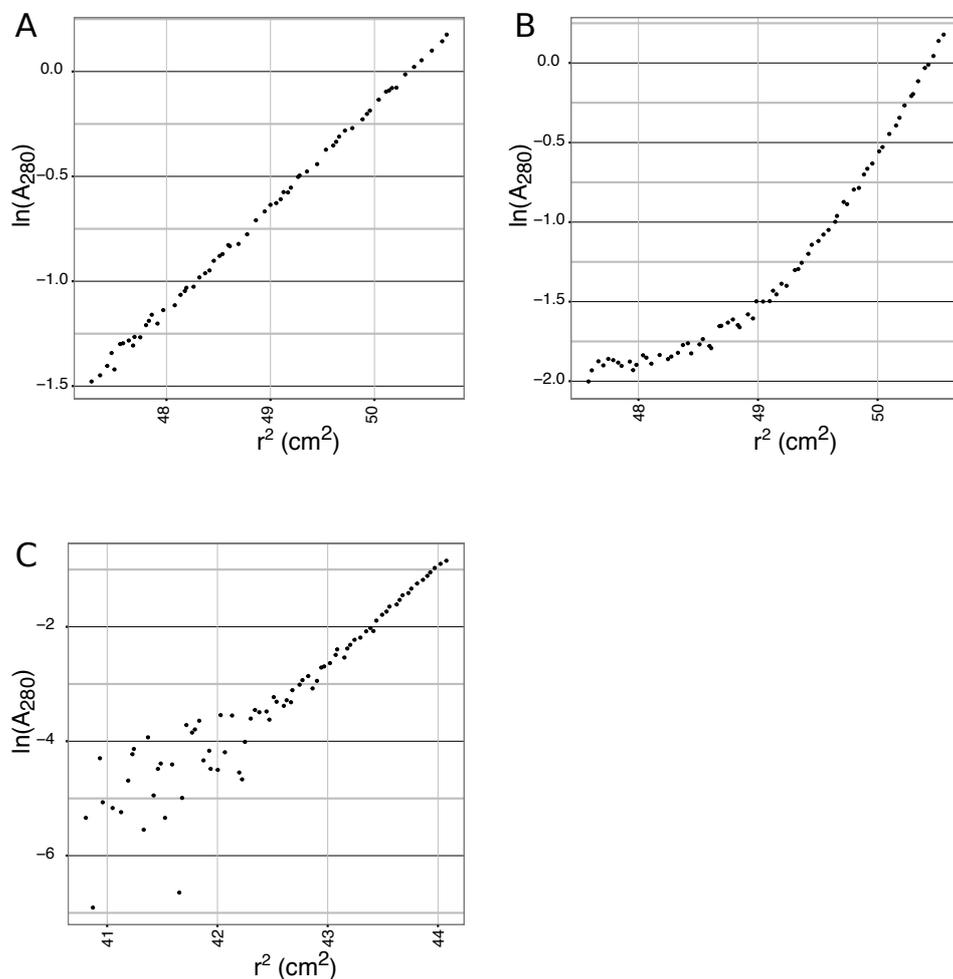


Figure 7.2. Diagnostic plots of sedimentation equilibrium AUC of wild-type PACT-D3

A) 0.75 mg/ml (88 μM), 22,000 rpm. B) 0.75 mg/ml (88 μM), 42,000 rpm. C) 0.19 mg/ml (22 μM), 42,000 rpm.

The majority of plots for lower speeds (below approximately 38,000 rpm) appear similar to Figure 7.2A, with an approximately linear relationship between $\ln(A_{280})$ and r^2 (see Appendix 2). This shows that a single species predominates, and that there are no serious problems with aggregation or non-ideality. However, at higher speeds many of the plots show a distinct curve at smaller radii, as in Figure 7.2B. The gradient is proportional to buoyant molecular mass, so this indicates that a mixture of species with different molecular weights is present, perhaps a monomer and a dimer. If this is the case, the monomer would be expected to be relatively more abundant in the samples with a lower protein concentration. Unfortunately, these samples also necessarily have lower absorbance, which means the signal from the smaller component becomes lost in the noise, as in Figure 7.2C. PACT-D3 contains no tryptophan residues, and its extinction coefficient at 280 nm is predicted to be only $6000 \text{ M}^{-1} \text{ cm}^{-1}$.

Although the buoyant mass can be determined approximately from the gradient of the diagnostic plots above, this does not account for the fact that the error is not constant across all data points. A more accurate method is to use non-linear least-squares methods to fit the raw data directly to a model. Initially, a model of a single species was fit to all well-equilibrated data, giving a mass of 16.9 kDa (16.7, 17.1) (fits and residuals are provided in Appendix 2). This is consistent with the predicted mass of a PACT-D3 dimer of 17.1 kDa.

An attempt was made to fit a two-species model of an 8.5 kDa monomer and a 17 kDa dimer in equilibrium, but this did not consistently converge. This is likely due to the low concentration of monomer, resulting in its signal being obscured by noise, and by the more intense dimer signal. To increase the relative population of monomer would require a lower overall protein concentration; however, the absorbance would then be too weak to detect. Therefore, this seAUC data shows that PACT-D3 is predominantly dimeric at concentrations above approximately $10 \mu\text{M}$ (the lowest concentration used here), but is not sufficient to determine the dissociation constant.

7.4 Phospho-mimic mutations do not appear to alter homodimerisation behaviour

Phosphorylation of PACT-D3 at positions S246 and S287 has been reported in response to cellular stress, although the kinase(s) responsible is unknown (Peters,

Li and Sen, 2006). It has been proposed that these phosphorylation events promote homodimerisation, while disrupting heterodimerisation with TRBP (Singh *et al.*, 2011; Singh and Patel, 2012). To test the first part of this hypothesis, I introduced S246D and S287D phospho-mimic mutations into PACT-D3, both individually and in combination, and examined the mutant domains by SEC-MALLS (Figure 7.3).

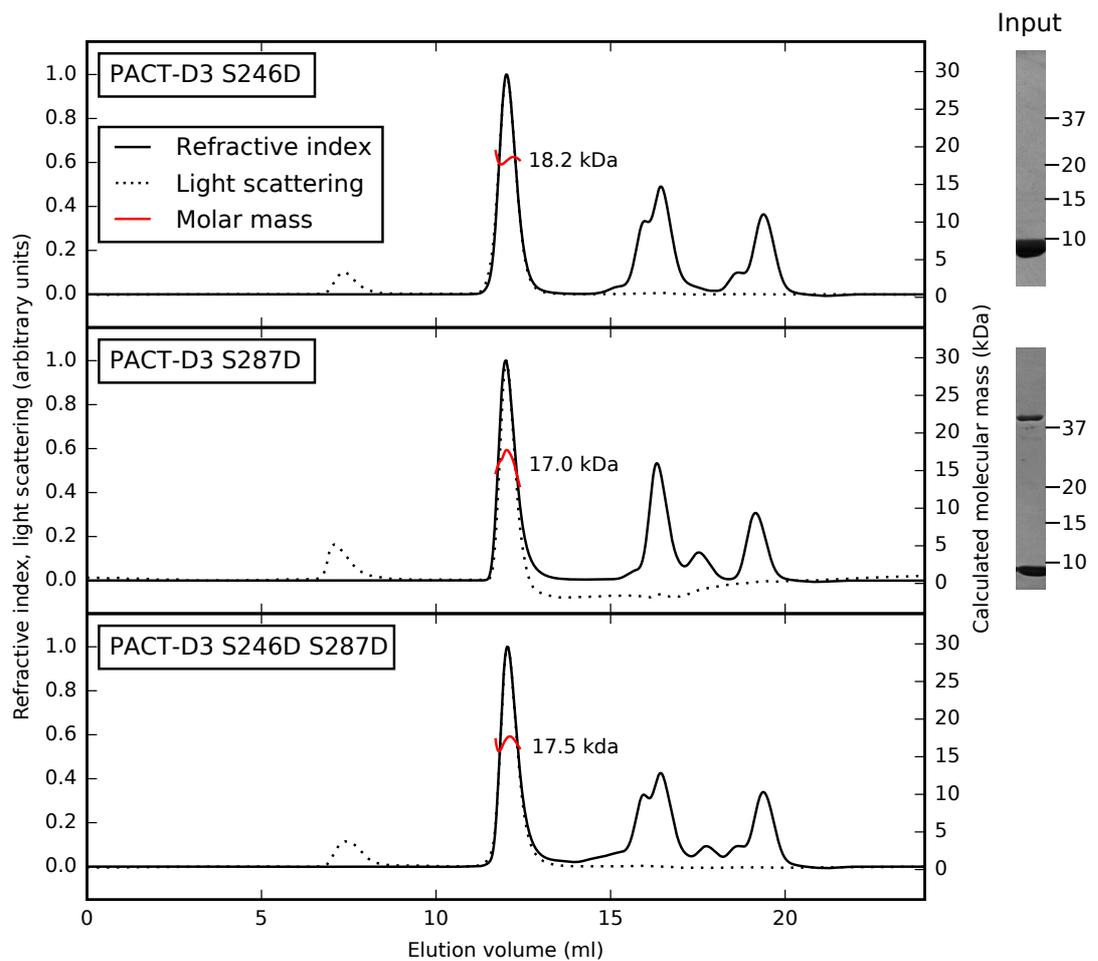


Figure 7.3. SEC-MALLS of PACT-D3 with phospho-mimic mutations

Differential refractive index is displayed as a solid line; the Rayleigh ratio (a measure of light scattering) is shown as a dotted line. The weight-averaged molar mass calculated over the centre of the peak is displayed next to each peak. SDS-PAGE analysis of the input is shown on the right (where available).

All three variants eluted at a similar volume to wild-type PACT-D3 (at approximately 12.0 mL), and were calculated to have similar masses. The fitted masses were slightly higher than for the wild-type (17-18.2 kDa compared to 16.6 kDa), but still more consistent with a dimer than with any other oligomeric state. Therefore, phospho-mimic mutations do not affect dimerisation at this concentration (3 mg/mL, 350 μ M). It remains possible that some of the phospho-mimic forms dimerise with higher affinity, but that this can only be observed at lower, more physiologically relevant concentrations. It is also possible that the phospho-mimic mutations studied here behave differently to true phosphorylation: if this is the case, it would also affect a number of previous studies (Peters 2006, Singh 2011, Singh 2012).

7.5 The L273R mutation disrupts PACT-D3 dimerisation

Conserved surface residues of PACT-D3 were identified using a structural model of the domain generated using the I-TASSER server (Roy, Kucukural and Zhang, 2010), and the sequence alignment of PACT and TRBP homologues (Figure 1.9). Several of these residues were then mutated to test whether they were required for dimerisation:

- E275 is conserved between TRBP and PACT, and is surrounded by other conserved residues (Figure 1.9). It was mutated to lysine to reverse the charge.
- H285, C284 and Y305 are highly conserved, not just between PACT and TRBP but also in the more distantly related Loqs (Figure 1.9). H285 and Y305 were mutated to alanine to effectively remove the sidechains, while C284 was mutated to serine, which is of similar size, but cannot form disulphide bonds.
- L273 is completely conserved between PACT, TRBP and Loqs (Figure 1.9), and it has recently been shown that mutating it to arginine disrupts dimerisation in Loqs-D3 (Jakob *et al.*, 2016). The L273R mutation was introduced into both PACT-D3 and PACT-Ext-D3.

All constructs were expressed and purified as described previously, and examined by SEC-MALLS. 1 and 2 dimensional NMR spectroscopy was used to check protein misfolding (data not shown): these spectra suggested that Y305A might be partially unfolded/misfolded, but that all other mutants were correctly folded.

E275K had no discernible effect on dimerisation, having an elution volume and calculated mass very similar to that of the wild-type. PACT-D3 C284S also eluted at the expected volume, but has a mass slightly below that expected for a dimer. PACT-D3 H285A eluted slightly earlier than wild-type PACT-D3, and with a higher calculated mass. This suggests it may have a tendency to aggregate, although it is still predominantly dimeric. The slight shoulder is likely due to a small amount of contaminating MBP (see associated gel in Figure 7.4). The masses of the C284S and H285A mutants deviate from the expected dimer mass by approximately 10%, but are still closer to the dimer mass than to a monomer or trimer.

PACT-D3 Y305A elutes later, and has a calculated mass of 9.7 kDa. This is more than 10% higher than the monomer mass of 8.5 kDa, which suggests the peak contains a small proportion of higher molecular weight particles, potentially a dimer. The small early-eluting peak has a mass of approximately 40 kDa, and is likely to be contaminating MBP (see inset gel in Figure 7.4).

The L273R mutation disrupts dimerisation in both PACT-D3 and PACT-Ext-D3 constructs. This is apparent both in their calculated masses (8.8 kDa and 12.3 kDa respectively), and from their significantly later elution time. No dimer appears to be present for either state. This suggests that PACT-D3 dimerises using a similar interface to that reported for Loqs-D3.

Jakob and coworkers demonstrated dimerisation of Loqs-D3 by crystallisation and cross-linking assays (Jakob *et al.*, 2016). To validate their results by a different method, I purified both wild-type and L426R forms of Loqs-D3, and examined them using SEC-MALLS (Figure 7.5). Wild type Loqs had a mass of 15.3 kDa, while the mutant domain had mass 8.0 kDa, consistent with dimer and monomer respectively (the expected monomer mass based on sequence is 8.1 kDa). An early eluting peak with a mass of 15.3 kDa was seen for Loqs-D3 L426R, which may correspond to a small population of homodimer. Alternatively, this peak may represent non-specific aggregation: both Loqs-D3 samples were stored for approximately 4 weeks at 4°C between purification and SEC-MALLS, considerably longer than for the other proteins shown.

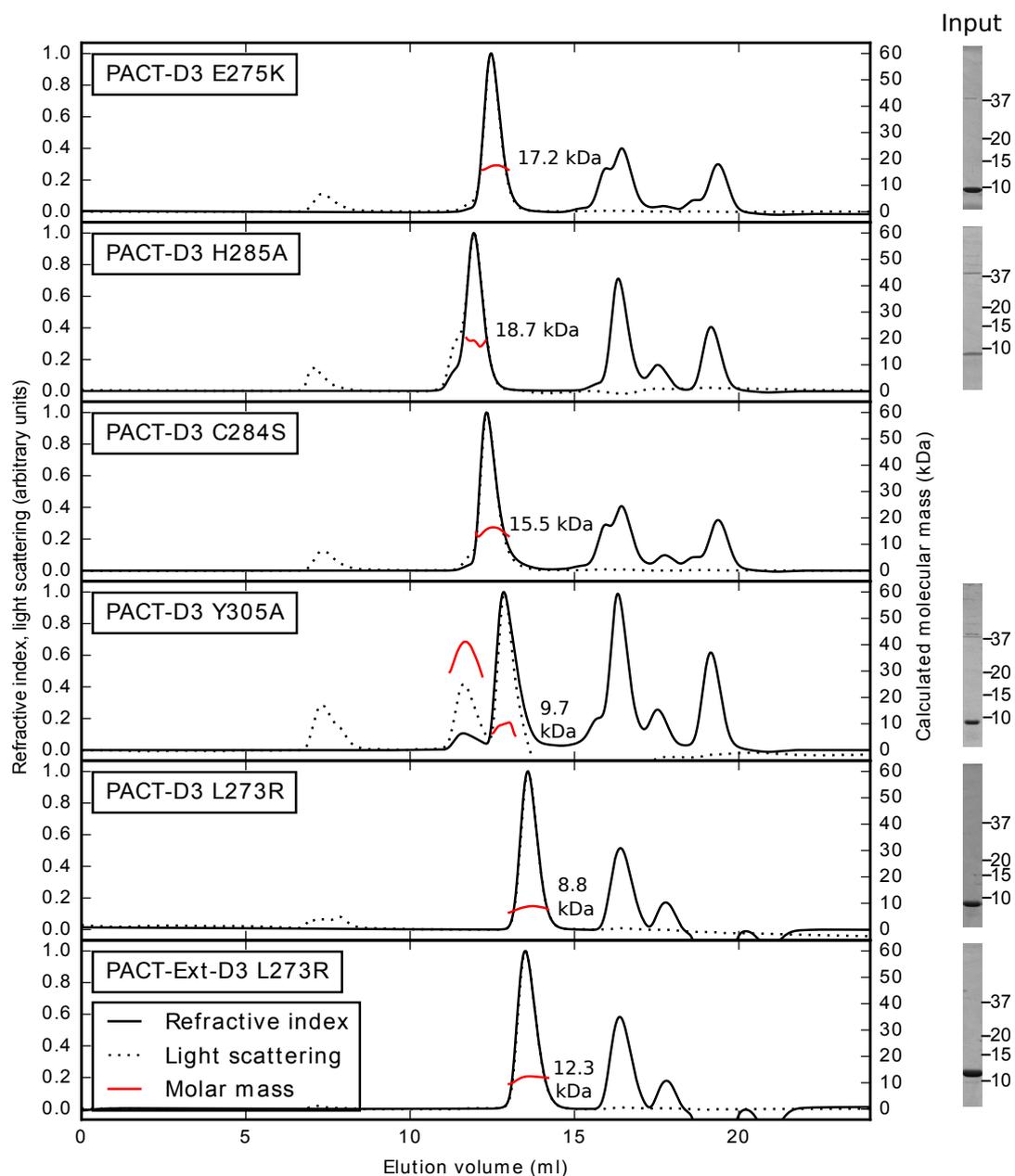


Figure 7.4. SEC-MALLS of PACT-D3 and PACT-Ext-D3 mutants

Differential refractive index is displayed as a solid line; the Rayleigh ratio (a measure of light scattering) is shown as a dotted line. The weight-averaged molar mass calculated over the centre of the peak is displayed next to each peak. SDS-PAGE analysis of the input is shown on the right (where available).

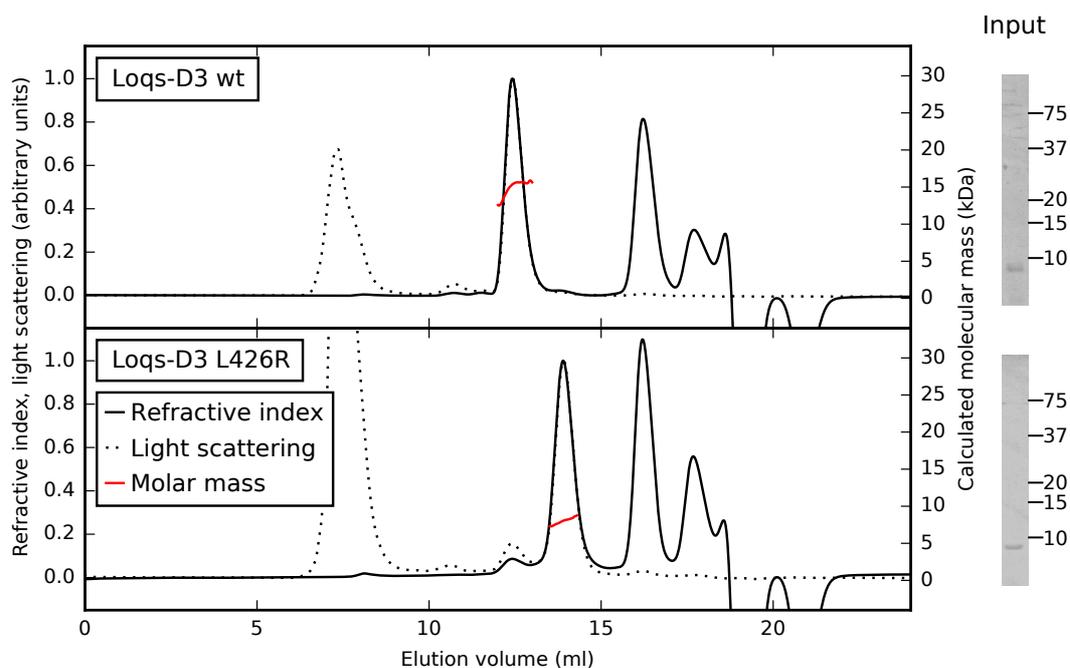


Figure 7.5. SEC-MALLS of Loqs-D3a and L426R mutants

Differential refractive index is displayed as a solid line; the Rayleigh ratio (a measure of light scattering) is shown as a dotted line. The weight-averaged molar mass calculated over the centre of the peak is displayed next to each peak. SDS-PAGE analysis of the input is shown on the right.

7.6 Discussion

7.6.1 The N-terminal extension is neither necessary nor sufficient for dimerisation

The structures of two dsRBD homodimers have been previously reported, for Stauf domain 5 and for Loqs domain 3 (Gleghorn *et al.*, 2013; Jakob *et al.*, 2016). Stauf domain 5 was found to dimerise via domain swapping of an N-terminal helical region, while Loqs domain 3 dimerised via its beta sheet region. The data presented in this chapter is sufficient to conclusively rule out the former mechanism, as removal of the N-terminal region does not prevent dimerisation, as measured either by SEC-MALLS or by equilibrium AUC (Figure 7.1). Furthermore, inclusion of the N-terminal region does not compensate for mutations on the dsRBD core that block dimerisation, since PACT-Ext-D3 L273R remains monomeric (Figure 7.4). Therefore, the N-terminal region of PACT-D3 is neither necessary nor sufficient for homodimerisation. The fact that mutations in this region (such as P222L) can cause heritable disease suggests that it plays an

important role unrelated to dimerisation, perhaps through increased PKR activation (Camargos *et al.*, 2008; Vaughn *et al.*, 2015).

It is possible to rationalise the effects of the different mutations on dimerisation with reference to the structure of Loqs domain 3 (Figure 7.6) (Jakob *et al.*, 2016). Q428 in Loqs (E275 in PACT) is completely exposed in one protomer, and is at the very edge of the interface in the other protomer (Figure 7.6A), so mutating it to lysine does not affect the interface. The side chain of C437 (C284) points into the domain and does not form a disulphide bond, so is relatively unaffected by the mutation to similarly-sized serine (Figure 7.6B). H438 (H285) is directly at the interface of both protomers, but does not appear to make any cross-interface interactions beyond standard van der Waals forces (Figure 7.6B): therefore, replacement with alanine does not drastically weaken the dimer.

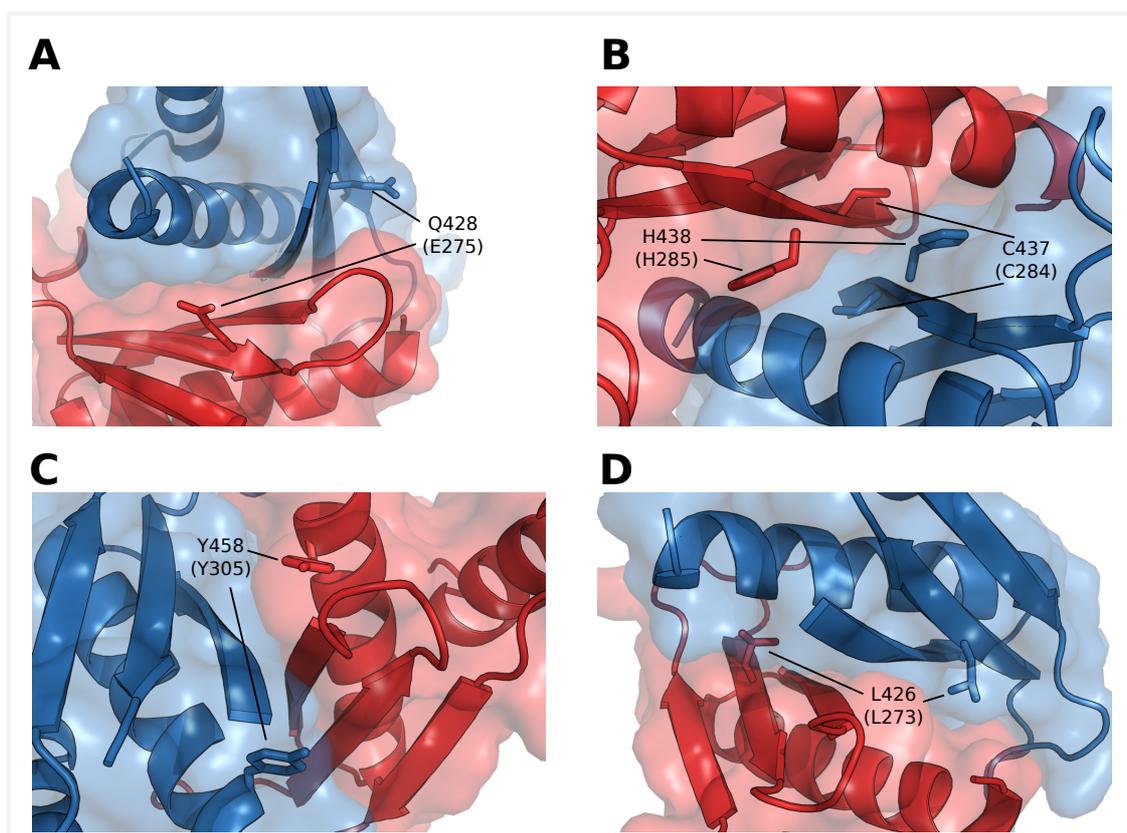


Figure 7.6. Positions of mutated residues with reference to Loqs-D3

Loqs-D3 (4X8W) chain A is shown in red, chain D/F in blue. The labels give the residue number in *Loqs*, followed by the equivalent *PACT* residue number in brackets. A) Q428 (E275); B) C437 (C284) and H438 (H285); C) Y458 (Y305); D) L426 (L273).

Y458 (Y305) is at the interface on both protomers, but also fills a gap between the third β -strand and the C-terminal α -helix (Figure 7.6C) – it seems plausible that losing both these functions would destabilise the interface through loss of direct contact, and perhaps through destabilisation of the domain structure. There is some evidence for this in the NMR spectrum of this domain, which shows increased signals in the central region of the spectrum, although this may be partly attributable to sample degradation (data not shown).

L426 (L273) is buried well within the binding surface of one protomer, and is partially exposed in the other (Figure 7.6D). Replacing the first of these two side chains with arginine causes a steric clash with the backbone of the C-terminal alpha helix of the second protomer, preventing dimerisation. Interestingly, modelling suggests that the leucine on the second protomer could be replaced with arginine with only minor steric clashes, possibly allowing the formation of heterodimers between wild-type and L273R forms of PACT-D3.

7.6.2 Implications for NMR studies

For both wild-type PACT-D3 and PACT-D3 L273R, a single oligomeric state was found to be present at the concentrations studied here. This is ideal for NMR studies, which operate in a similar or slightly higher concentration range to SEC-MALLS. The presence of multiple states complicates interpretation of NMR spectra, through the appearance of additional peaks or through exchange broadening of some or all signals. Oligomeric state (or more specifically, overall complex mass) also affects the rotational correlation time, which heavily influences the relaxation rate. Larger oligomeric states result in slower rotational averaging and faster relaxation, reducing the intensity of all NMR signals, especially for longer pulse sequences. The monomer and dimer states have masses of 8.5 kDa and 17 kDa respectively, putting them within the range of masses that can usually be studied with standard NMR methods. Finally, oligomeric state can affect the structure calculation process: a symmetric dimer will have intermolecular NOEs that are not easily distinguishable from intramolecular NOEs. If a structure is calculated assuming a monomer, when the true oligomeric state is a dimer or higher, these restraints based on the intermolecular NOEs will be applied between the wrong atoms, leading to a distorted structure (Nabuurs *et al.*, 2006).

Characterisation of PACT Domain 3 by NMR Spectroscopy

8.1 Introduction

The first section of this chapter presents the backbone assignment of PACT-D3 L273R, together with some preliminary data on PACT-Ext-D3 L273R. The later sections detail work on wild-type PACT-D3, and present clear evidence that PACT-D3 homodimers are asymmetric, and that the asymmetry is focused on the third β -strand. The dimer interface involves the third β -strand and second α -helix, and is very similar to that recently observed in Loqs-D3 (Jakob *et al.*, 2016).

8.2 Characterisation of PACT-D3 L273R by NMR

Following initial tests, [^{15}N , ^{13}C]-labelled PACT-D3 L273R was prepared at a concentration of 1 mM and a standard set of triple-resonance backbone assignment experiments was recorded, namely CBCANH, CBCA(CO)NH and HNC(O). Sequence-specific assignment was carried out using the $C\alpha$ and $C\beta$ chemical shifts to link neighbouring residues.

Figure 8.1 shows an assigned (^1H , ^{15}N) HSQC of PACT-D3 L273R. It contains 103 peaks (including sidechain amides), close to the 98 peaks predicted from its sequence (several low intensity peaks were observed that could not be assigned to the PACT-D3 L273R sequence). 98% of backbone amide resonances were assigned, together with 96% of C_0 spins, 99% of $C\alpha$ spins and 100% of $C\beta$ spins. It was also possible to assign all 4 asparagine side chains, and 4 out of 6 glutamine side chains, using the CBCA(CO)NH to link the NHD signals from the side chain amides to known $C\alpha$ and $C\beta$ chemical shifts. Full assignment data is provided in Appendix 3.

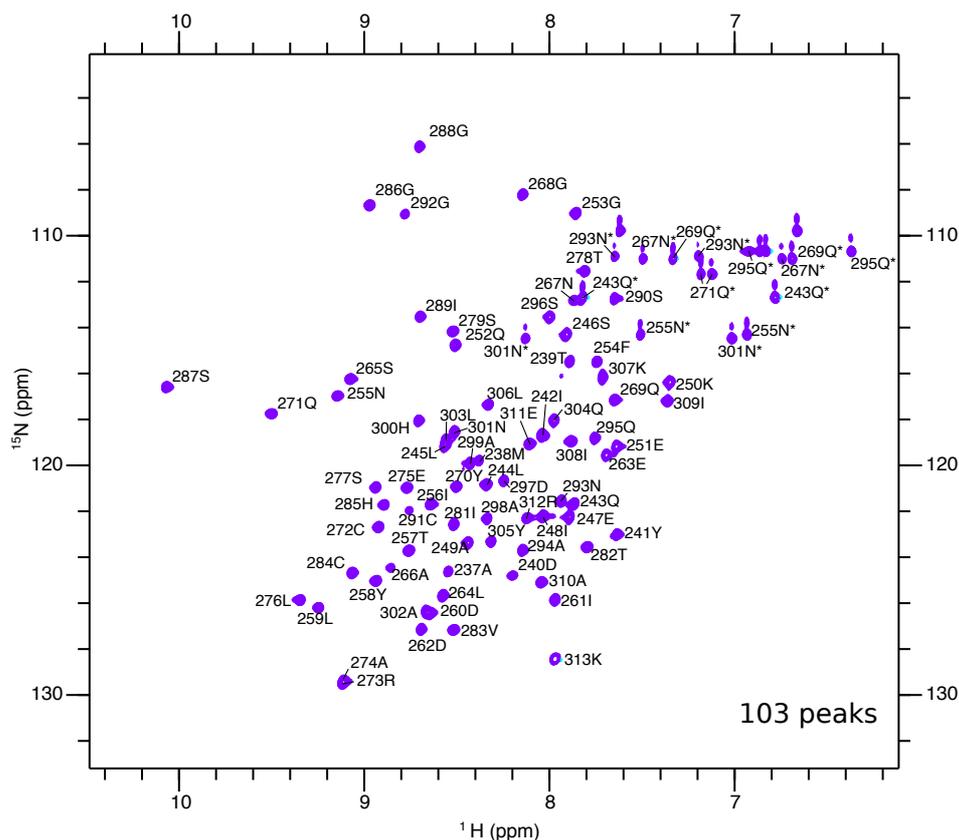


Figure 8.1. Assigned (^1H , ^{15}N) HSQC of PACT-D3 L273R

Experiments were performed on a sample in standard NMR buffer, except that the TCEP concentration was only 4.5 mM. The unassigned peaks in the top left of the spectrum correspond to Q252 and Q304, which have similar $\text{C}\beta$ chemical shifts. The nearby peak labelled 295Q appears broader than the other side chain peaks and lacks the NHD signal. However, the CBCA(CO)NH spectrum strongly supports this assignment (data not shown), suggesting that this amide proton is likely broadened by a dynamic process.

The secondary structure can be predicted from the backbone chemical shifts using TALOS-N (Shen and Bax, 2013). This is an updated version of the popular TALOS program, which predicts dihedral angles from chemical shifts using a database of X-ray structures for which chemical shifts have been measured or predicted, together with a trained neural network.

Figure 8.2 shows the TALOS-N secondary structure prediction, together with the secondary structures of TRBP-D3 and Loqs-D3 based on crystallographic data (Wilson *et al.*, 2015; Jakob *et al.*, 2016). The TALOS-N predictions for PACT-D3 L273R are in very good agreement with the TRBP-D3 and Loqs-D3 secondary structures, strongly suggesting that it adopts the same fold.

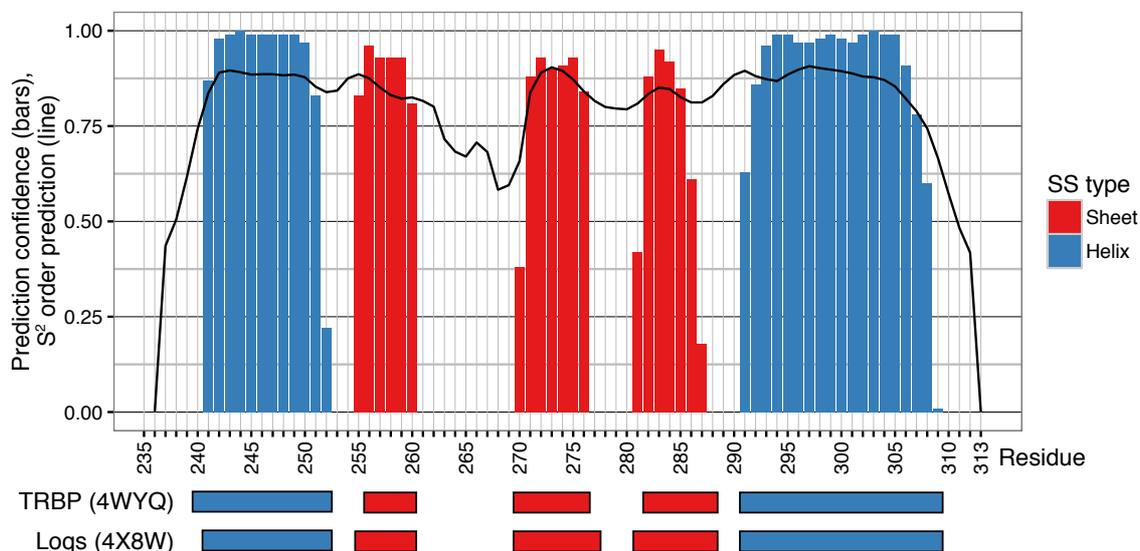


Figure 8.2. TALOS-N analysis of PACT-D3 L273R

H_N , N , $C\alpha$, $C\beta$ and C_0 chemical shifts were used for the prediction. Coloured bars show the prediction confidence for each secondary structure type; the solid line shows the predicted S^2 order parameter. Below, the secondary structures of TRBP domain 3 (from PDB accession 4WYQ) and Loqs domain 3 (PDB accession 4X8W) are displayed.

TALOS-N also predicts the S^2 order parameter for each residue, which describes how restricted internal molecular motions are at each site (see Chapter 4, Section 4.2.3). It can be estimated by combining the deviations of the backbone chemical shifts from their random coil values into a single Random Coil Index (RCI) (Berjanskii and Wishart, 2005). The predicted S^2 parameters from this method are in good agreement with known features of dsRBD structure, such as the increased flexibility of loop β -1,2 (Nanduri et al., 2000; Benoit and Plevin, 2013).

There is some crystallographic evidence that the conserved N-terminal extension of TRBP contains an α -helix that docks onto the core domain (Wilson *et al.*, 2015). However, the poor quality of the electron density did not allow sequence specific assignment of this helix, and the length of the helix did not account for the entire N-terminal region, suggesting other residues are disordered. To investigate the equivalent region of PACT, preliminary NMR data was recorded for PACT-Ext-D3 L273R.

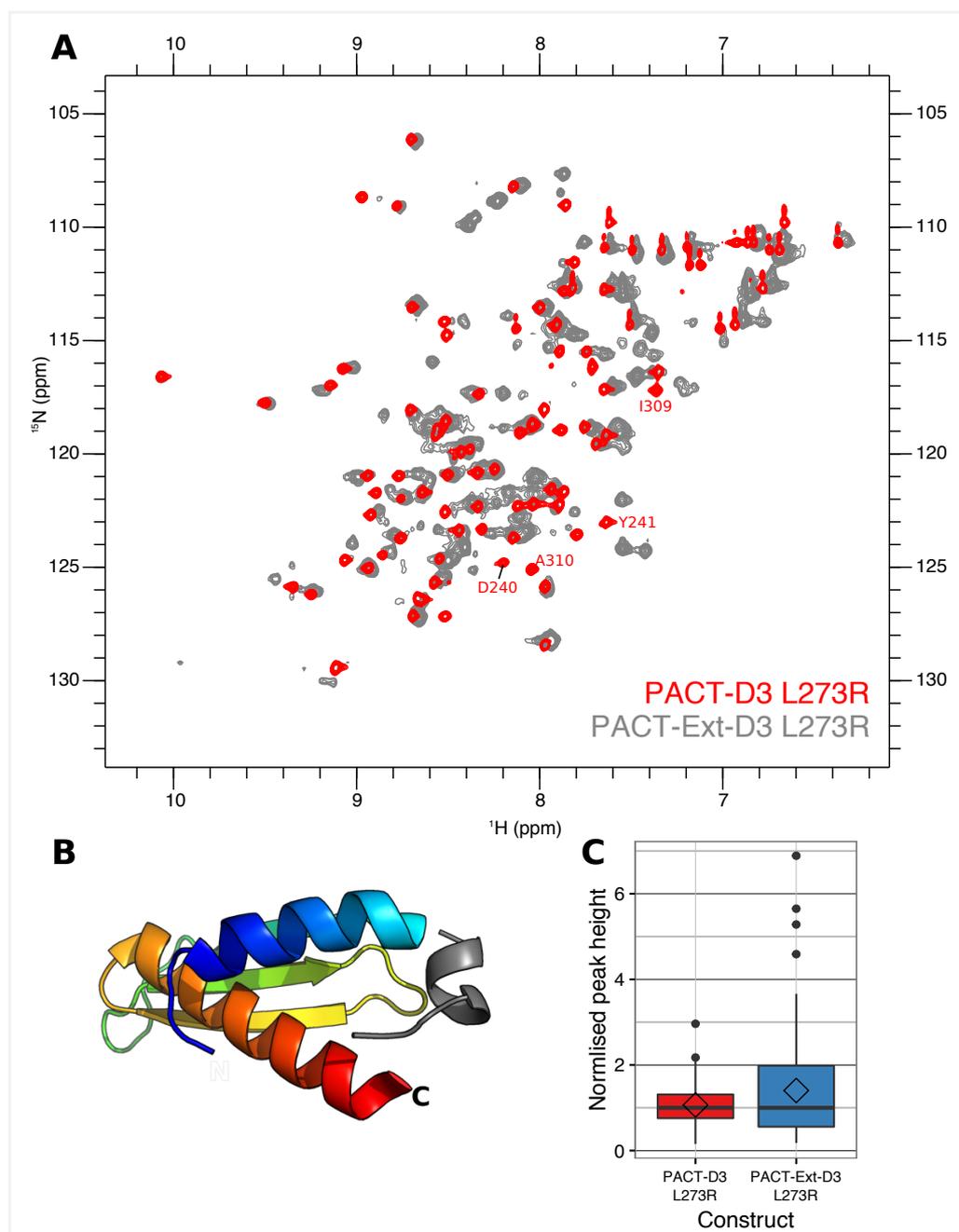


Figure 8.3. (^1H , ^{15}N) HSQC of PACT-Ext-D3 L273R

A) Overlaid (^1H , ^{15}N) HSQC spectra of PACT-D3 L273R (red) and PACT-Ext-D3 L273R (grey). The PACT-D3 L273R sample was in standard NMR buffer, while the buffer for PACT-Ext-D3 L273R was 9 mM sodium phosphate, 45 mM NaCl, 4.5 mM DTT, 10% D_2O , 50 μM DSS, pH 6.5. B) Structure of TRBP domain 3 (4WYQ), showing the location of the N-terminal extension α -helix (Wilson et al., 2015). C) Boxplot showing the distribution of peak heights in the (^1H , ^{15}N) HSQC spectra shown in part A. Peaks heights have been normalised to the median height for each protein. Diamonds show the mean peak height.

Figure 8.3A shows superimposed HSQCs of PACT domain 3 with and without the N-terminal extension. There is reasonable overlay of many peaks, which is expected given that they should share the same core structure. A number of new peaks have appeared, mostly in the central part of the spectrum, suggesting that the N-terminal extension is either unfolded or α -helical. A relatively small number of peaks seem to have moved significantly – these include N-terminal residues such as D240 and Y241, but also C-terminal residues such as I309 and A310. This is consistent with the crystal structure of TRBP domain 3 in complex with Dicer, which suggests that the N-terminal extension docks into a space between the two α -helices and loop β -1,2 (Figure 8.3B). Overall, there are around 159 peaks, slightly more than the 139 expected from the sequence. However, it is difficult to accurately determine the number of peaks due to the wide range of peak intensities present (Figure 8.3C). This suggests either that different parts of the protein have different correlation times (for example, if one part is unfolded), or that some peaks are being broadened by exchange processes, such as transient interactions between the N-terminus and the core domain.

8.3 Characterisation of PACT-D3 dimers by NMR

8.3.1 Backbone assignment of PACT-D3

Backbone assignment of PACT-D3 was performed using the same strategy as for PACT-D3 L273R, on a 1.3 mM protein sample in standard NMR buffer. For side chain assignment, H(CCCO)NH-TOCSY, CC(CO)NH-TOCSY HC(C)H-TOCSY and ^{13}C -NOESY-HSQC spectra were recorded of a 2.5 mM sample, and ^{15}N -NOESY-HSQC data was collected for a 1.1 mM sample, both in standard NMR buffer. An assigned (^1H , ^{15}N) SOFAST-HMQC spectrum is shown in Figure 8.4.

The most striking feature of Figure 8.4 is the presence of approximately twice the expected number of peaks (186, compared to the 97 predicted from the sequence). This complicated the assignment process, but it was still possible in most cases to link peaks unambiguously through the CBCANH and CBCA(CO)NH spectra. There were however a number of spin systems which had identical $\text{C}\alpha$ and $\text{C}\beta$ shifts, which were initially left unlinked. This led to a series of fragments that could be assigned to the PACT-D3 sequence based on the characteristic $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts of glycine, alanine, serine and threonine. At this stage, it

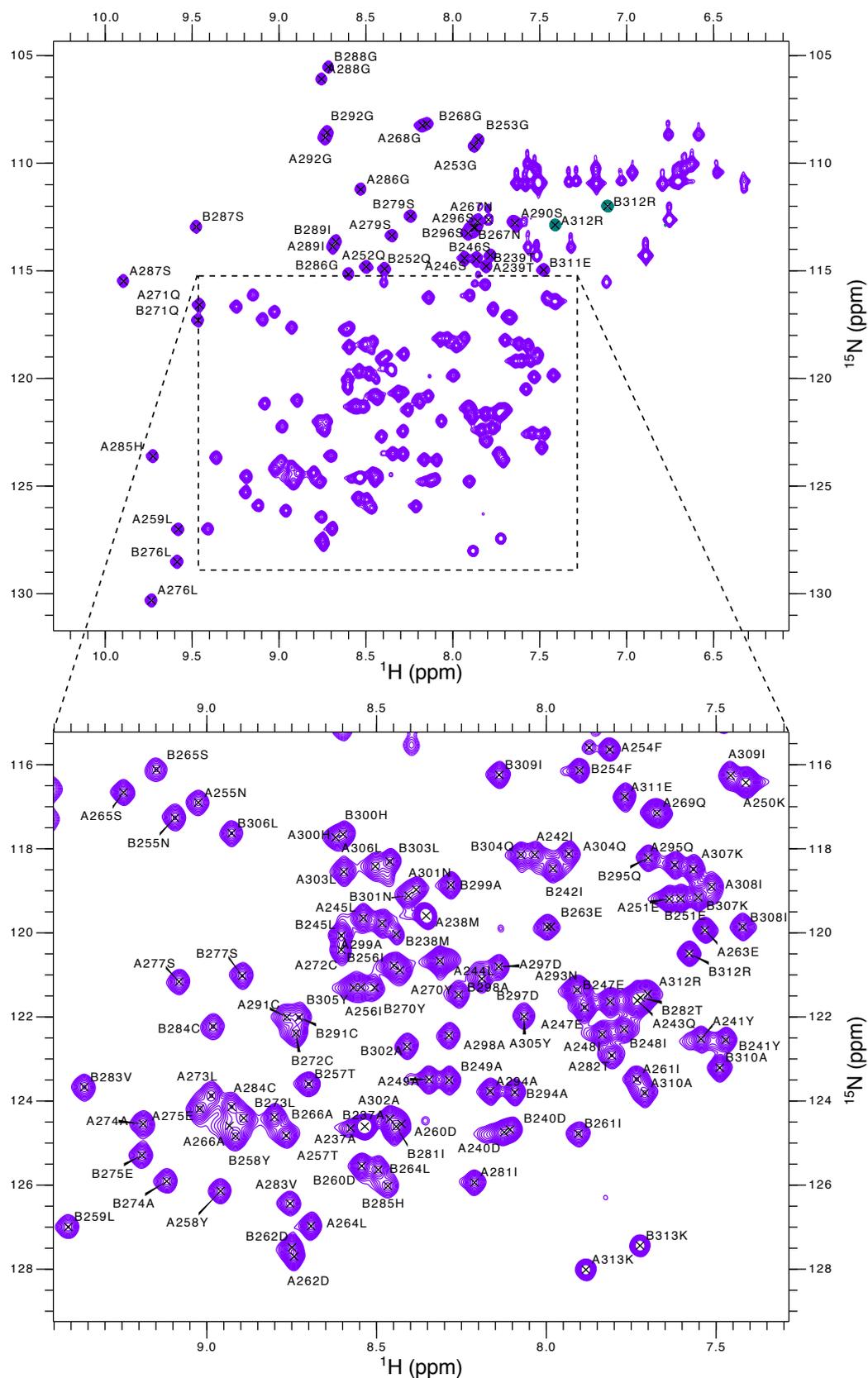


Figure 8.4. Assigned (^1H , ^{15}N) SOFAST-HMQC of PACT-D3

Peaks are labelled with the protein state they originate from (A/B), the residue number, then the amino acid type. See Chapter 4, section 4.3.5 for experimental details, and Appendix 3 for full assignments.

became clear that there are two resonances associated with each nucleus, indicating that two distinct states are present. (The nature of the two states is investigated in section 8.4 below.)

With this in mind, it was possible to assign the remainder of the peaks. 95% of backbone amides could be assigned, together with 99% of C α spins. In addition, 93% of side chain protons were assigned, together with 80% of side chain carbons (C β outwards). Ambiguous C α and C β chemical shifts were typically found towards the N-terminus, in loops, or for glycine residues (due to the lack of C β). In many cases, spin systems with similar C α and C β chemical shifts also had similar N and H_N chemical shifts, such as D262, I289.

In order to correctly link fragments belonging to the same state, backbone and side chain NOEs from the ¹⁵N- and ¹³C-NOESY-HSQC experiments were used. This approach allowed states to be consistently assigned to residues 255-313, but was unable to unambiguously allocate states to 239-254, due to the similar backbone and side chain chemical shifts of these residues. For most practical purposes, the exact assignment of states in this region is unimportant, since the NMR parameters are so similar (see section 8.7.1 for further discussion).

Secondary structure predictions for each state were made using TALOS-N, using N, H_N, H α , C α , C β and C_O chemical shifts (Figure 8.5). The two states of PACT-D3 have almost identical secondary structures, which are also very similar to the secondary structures of TRBP-D3 and Loqs-D3. The secondary structures of wild-type PACT-D3 and PACT-D3 L273R appear identical across most of the sequence, but differ slightly at the C-terminus, which has a slightly shorter α -helix in the L273R mutant (cf Figure 8.2). The predicted S² order parameter is again very similar across most of the two states of wild-type PACT-D3, but differs slightly in loop β -1,2, which is predicted to be more flexible in state B. Examining the predicted dihedral angles (Figure 8.5B), it can be seen that there are slightly larger differences in the β -strands than in the α -helices, but also that the β -strand predictions have greater associated uncertainties.

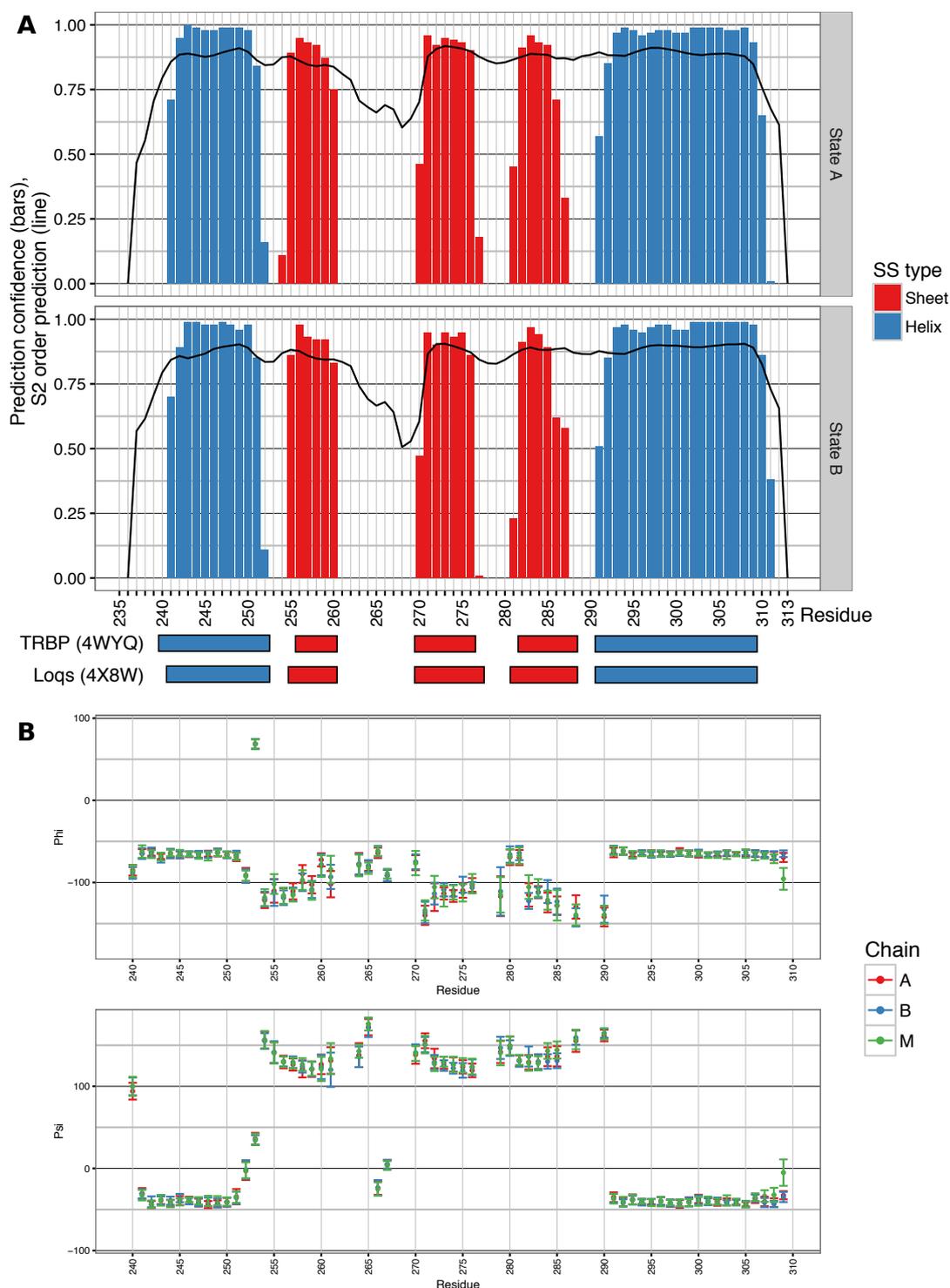


Figure 8.5. TALOS-N analysis of PACT-D3

A) Secondary structure was predicted by TALOS-N using predictions H_N , N , $C\alpha$, $C\beta$ and $C\gamma$ chemical shifts. Coloured bars show the prediction confidence for each secondary structure type; the solid line shows the predicted S^2 order parameter.

Below, the secondary structures of TRBP domain 3 (from PDB accession 4WYQ) and Loqs domain 3 (PDB accession 4X8W) are displayed. B) Dihedral angle predictions for each dimer state, and for monomeric PACT-D3 L273R.

8.3.2 Investigation of phospho-mimic mutants of PACT-D3

As discussed in Chapter 1, phosphorylation of PACT-D3 at serines 246 and 287 has been suggested to control its ability to form homo- and heterodimers, and to activate PKR. Much previous work has used serine-to-aspartate mutations to mimic the effects of phosphorylation: this approach has also been adopted here. In Chapter 7, SEC-MALLS data demonstrated that introduction of phospho-mimic mutations into PACT-D3 does not alter the oligomeric state at concentrations on the order of 100 μ M. Here, NMR is used to examine the effects of the mutations in more detail.

Figure 8.6 shows (^1H , ^{15}N) HSQCs of PACT-D3 S246D, PACT-D3 S287D and PACT-D3 S246D S287D. The spectrum of the S246D mutant (Figure 8.6A) suggests that the majority of residues are largely unaffected by the mutation, since most peaks overlay well with the spectrum of wild-type PACT-D3. A few residues in the neighbourhood of the mutation show larger shifts: for example, T239, Y241, L245, A249 and F254 all show moderate shifts. The two states behaviour is still present, and peaks from both states seem to be affected by the mutation in the same way (for example, the peaks from Y241_A and Y241_B are both shifted by the same amount).

The spectrum of PACT-D3 S287D (Figure 8.6B) shows slightly larger changes, with chemical shift changes for a larger number of residues. Some of these are from residues near the mutation site (for example H285, G288), but others are expected to be quite distant, based on the structural model of the domain (for example A310_A, S265). This could suggest a larger change in structure. Again, the two states behaviour is unaffected by the phospho-mimic mutation.

The (^1H , ^{15}N) HSQC of PACT-D3 S246D S287D shows numerous changes from the wild-type domain. However, when superimposed on the spectra of the two single mutants (as in Figure 8.6C), it appears that almost all peaks overlay well on at least one of the two single mutants. This suggests that the effects of the two mutations are purely additive, and that there are no cooperative effects on protein structure. Additionally, the PACT-D3 S246D S287D spectrum contains a number of weaker peaks that appear characteristic of degraded protein, suggesting this mutant may be less stable than the others.

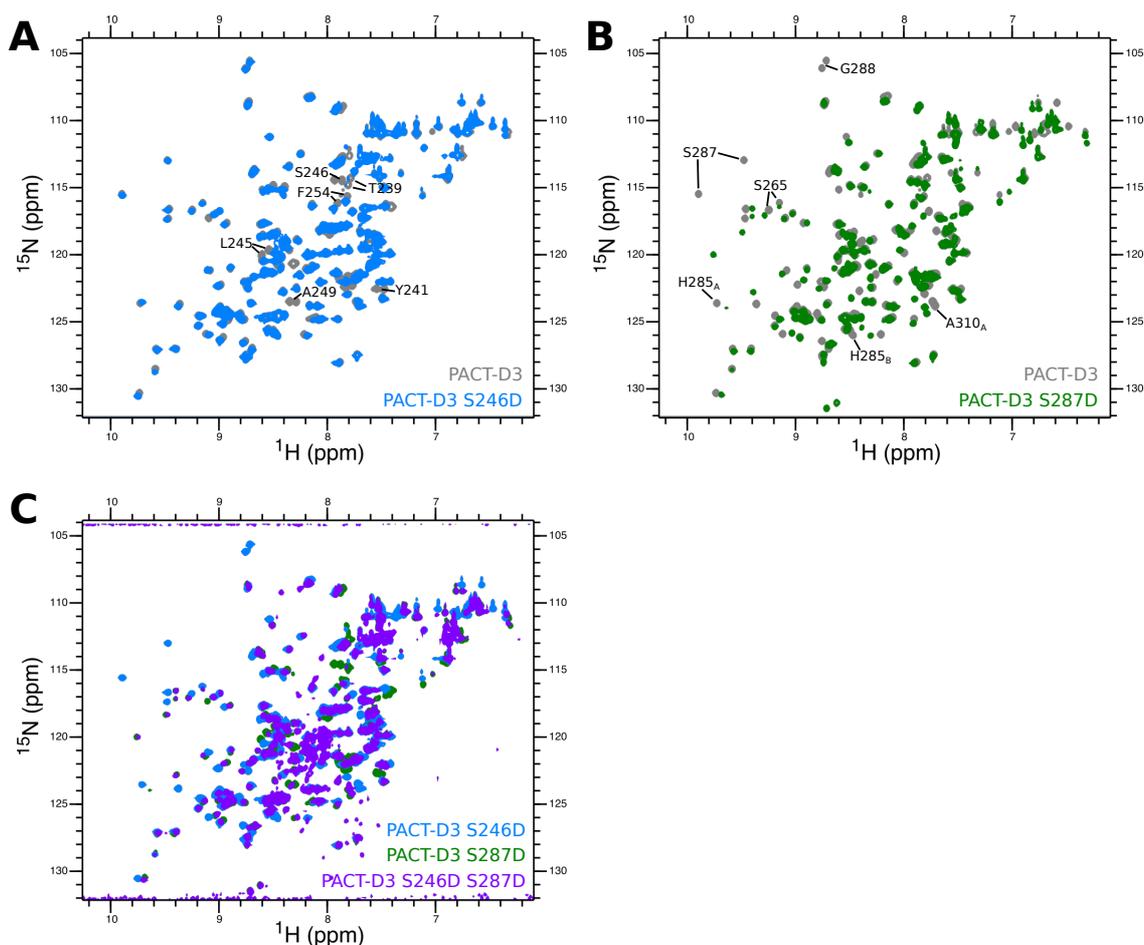


Figure 8.6. (^1H , ^{15}N) HSQC spectra of PACT-D3 phosphomimic mutants

Overlaid (^1H , ^{15}N) HSQC spectra of PACT-D3: A) Wild-type (grey) and S246D (blue); B) Wild-type (grey) and S287D (green); C) S246D (blue), S287D (green) and S246D S287D (purple). Note that the two peaks of G288 are aliased in the S287D and S246D S287D spectra.

8.3.3 (^1H , ^{15}N) HSQC of PACT-Ext-D3

A (^1H , ^{15}N) HSQC was also recorded of wild-type PACT-Ext-D3 (Figure 8.7). This spectrum contains approximately 200 peaks, about 50% more than the 139 expected from its sequence. This suggests peak doubling is occurring, but perhaps that some peaks are being lost due to exchange broadening or signal overlap. There are substantial shifts in the positions of many peaks, whether the spectrum is compared to wild-type PACT-D3 (Figure 8.7A) or PACT-Ext-D3 L273R (Figure 8.7B). As with PACT-Ext-D3 L273R, a wide range of peak intensities are present, also consistent with the presence of an intermediate exchange process (data not shown). Given the difficulties encountered for the monomeric mutant form of this

domain, and the additional complication of peak doubling, it was decided that wild-type PACT-Ext-D3 is too complicated to analyse further at this time.

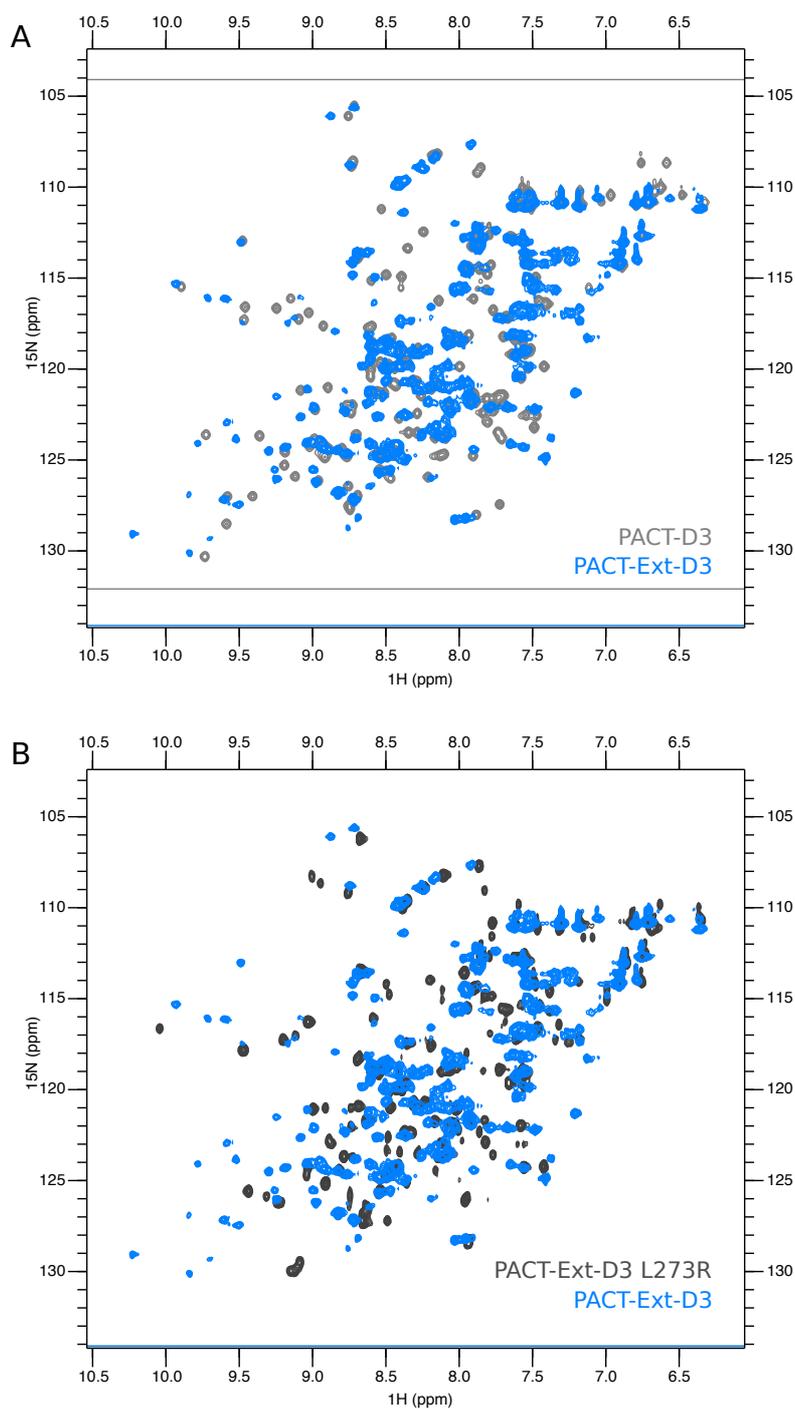


Figure 8.7. (^1H , ^{15}N) HSQC of PACT-Ext-D3

Overlaid (^1H , ^{15}N) HSQC spectra of PACT-Ext-D3 (blue) with A) PACT-D3 (grey) or B) PACT-Ext-D3 L273R (dark grey). All samples were in standard NMR buffer except for PACT-Ext-D3 L273R, which contained 4.5 mM DTT instead of TCEP.

8.4 Investigations into the two states of PACT-D3

The data presented in the previous section shows that two states of PACT-D3 are present in solution. It was important to work out the reason for the presence of two states, to determine whether they are a genuine feature of the protein, or an artefact of the conditions it is studied under.

Several possibilities can be immediately ruled out. Firstly, the two states cannot represent monomeric and dimeric PACT-D3. This is because the SEC-MALLS data demonstrates that the protein is entirely dimeric at the concentration studied, while the two states have approximately equal intensities (as discussed in section 8.4.1. below).

Secondly, the two states cannot represent folded and misfolded protein, as TALOS-N shows both states to have the same secondary structure. While there is a possibility that the tertiary structures could be different despite the identical secondary structure, this would require differences in the loops between secondary structure elements, which is not supported by the TALOS-N predictions (Figure 8.5B).

The following sections investigate other potential explanations for the two states behaviour.

8.4.1 Two-state behaviour is maintained across a range of buffers and conditions

The first question is whether the two-state behaviour is intrinsic to the protein, or is instead caused by particular experimental conditions. To test this, NMR spectra of PACT-D3 were recorded in a range of buffers (varying buffering compound, pH and denaturant concentration) and conditions (varying temperature and pressure). The intensity of each well-resolved peak was extracted, and the relative population of state A calculated. Figure 8.8A shows the data for denaturant concentration, temperature and pressure.

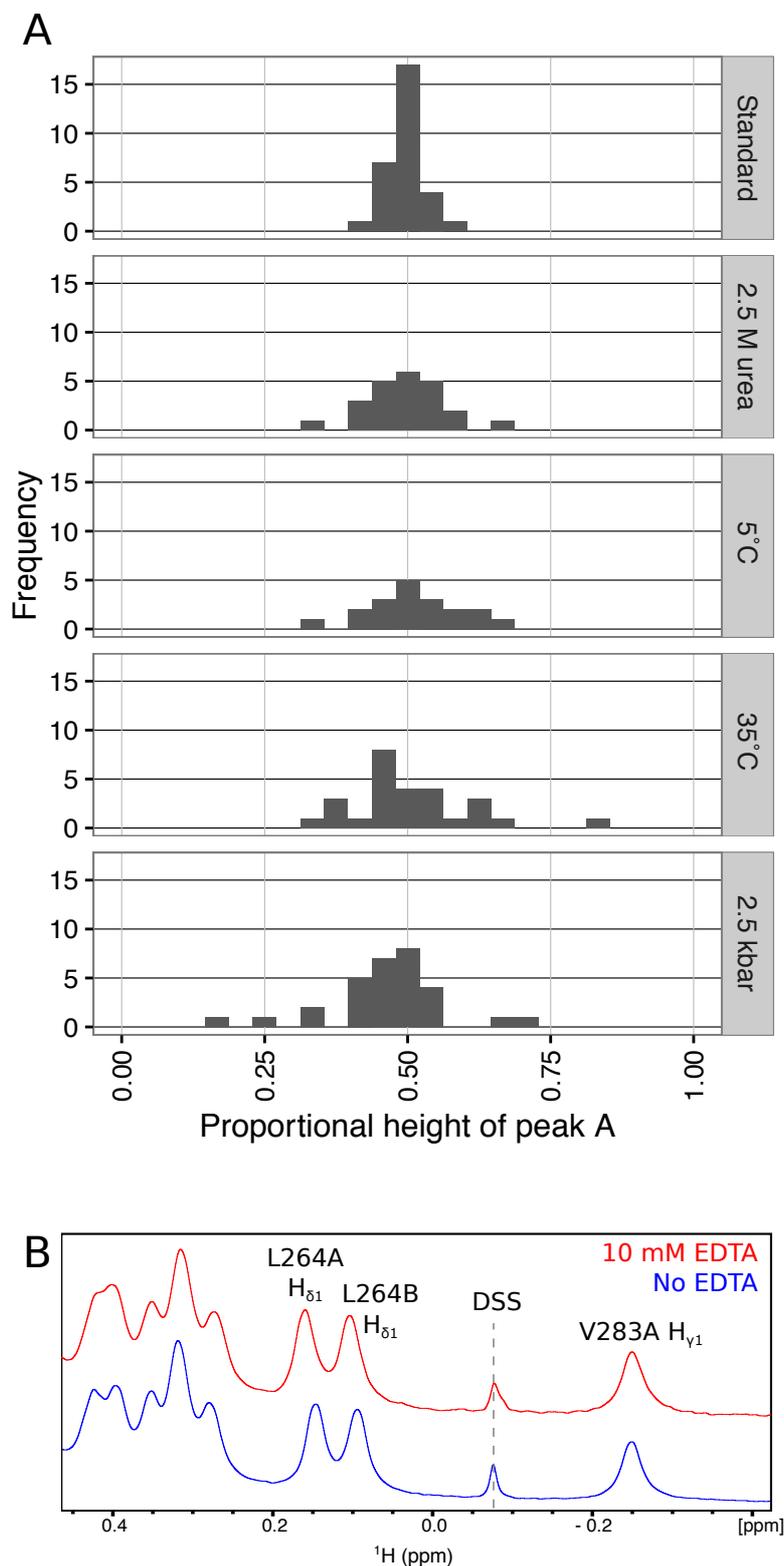


Figure 8.8. Distribution of HSQC peak heights under different conditions

A) Population distribution of the two states of PACT-D3 under various conditions. The x axis is the proportion of each residue in state A (calculated as $Height_A / (Height_A + Height_B)$). B) 1D proton spectra showing the upfield methyl region of PACT-D3 in standard NMR buffer, with or without the inclusion of 10 mM EDTA.

The first observation is that two states were still present under all the conditions studied. More specifically, the relative intensities of the two states remained approximately equal, suggesting that the two states are equally populated. This strongly argues that the two-state behaviour is an intrinsic property of PACT-D3. (More details of the high pressure NMR are given in section 8.6.)

The above experiments do not explicitly test for possible sub-stoichiometric binding of contaminating nickel ions carried through from the nickel affinity purification. To rule out this as an explanation for the two states, proton NMR spectra of PACT-D3 were recorded in the presence or absence of 10 mM EDTA (Figure 8.8B). The peak at -0.2 ppm corresponds to one of the methyl groups of V283_A, and is diagnostic of two-state behaviour. Additionally, two peaks are visible for L264 H δ 1, again showing that two-state behaviour is maintained in the presence of EDTA, and indicating that contaminating ions are unlikely to be the cause of the two-state behaviour.

8.4.2 Chemical shift analysis suggests differences between the two states are focused on the third β -strand.

To examine the regions of the protein with the greatest differences between the two states, the compound chemical shift between the amides of the two states was plotted (Figure 8.9A). This shows that the two states differ most in the third β -strand, and least in the N-terminal helix. When this data is plotted on a structural model of PACT-D3, it can be seen that the residues with the largest compound chemical shifts are clustered on one face of the protein. This strongly suggests that this region, focused around the third β -strand, differs significantly between the two states.

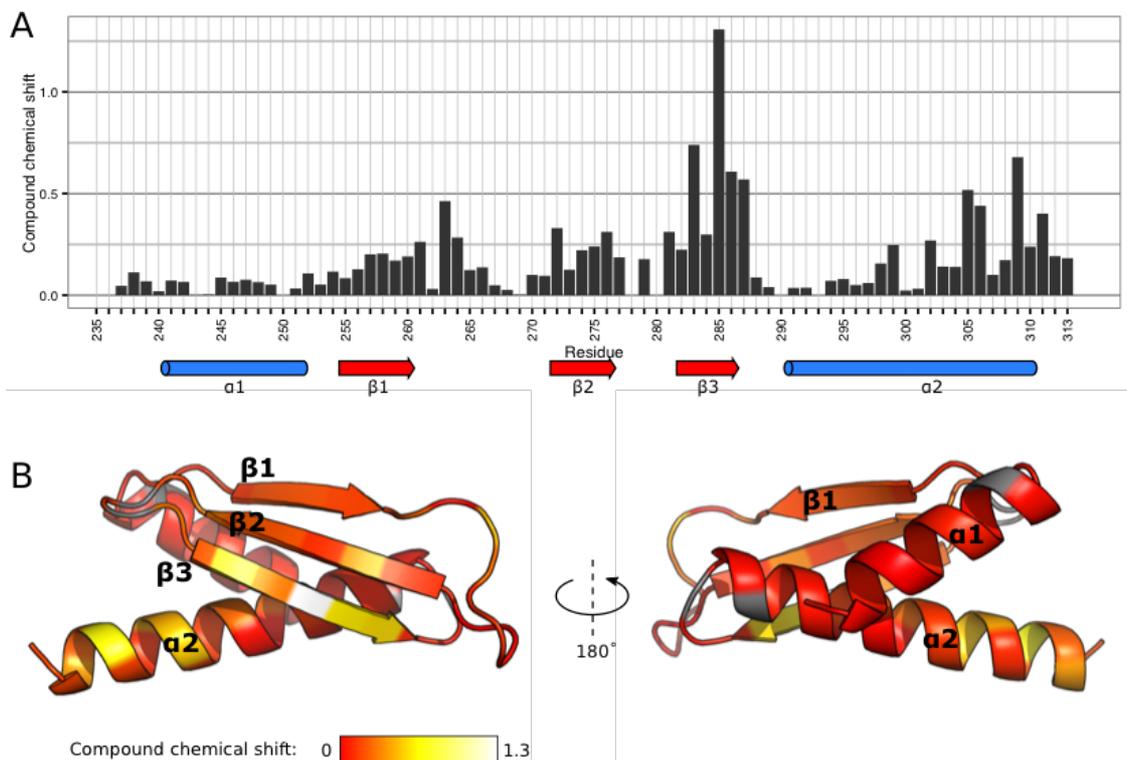


Figure 8.9. Chemical shift differences between the two states of PACT-D3

A) Compound chemical shifts calculated between the two states of PACT-D3,

according to the formula $\delta_{compound} = \sqrt{\delta_H^2 + \left(\frac{\delta_N}{6.5}\right)^2}$. Below, the secondary structure of PACT-D3 is shown, as calculated by TALOS-N above. B) Structural model of PACT-D3 (generated using the I-TASSER server (Roy, Kucukural and Zhang, 2010)) coloured according to compound chemical shift.

8.4.3 Proline isomerisation does not explain the two states

The differences between the two states might be due to different proline isomerisation states. PACT-D3 contains a proline at position 280, just before the start of the third β -strand, and thus in the same region as the largest compound chemical shifts. (There is also the non-native proline P236, but this is far from the region which differs between the two states.) It has been shown previously that carbon chemical shifts can be used to determine which proline isomer is present, based on the difference between $C\beta$ and $C\gamma$ chemical shifts (Schubert *et al.*, 2002). A difference in chemical shifts of less than 4.8 ppm indicates the trans isomer, while a difference of greater than 9.15 ppm indicates the cis isomer.

Figure 8.10A shows strips from the CBCA(CO)NH and CC(CO)NH-TOCSY, focusing on P280 in each state. The carbon chemical shifts are similar in both states A and

B: in particular, the difference between the C β and C γ chemical shifts is approximately 10 ppm in each case, which is characteristic of the cis isomer. This is also consistent with the crystal structure of Loqs-D3. Therefore, proline isomerisation is not responsible for the two-state behaviour.

8.4.4 The two-state behaviour is not due to disulphide bond formation

PACT-D3 contains 3 cysteine residues, so it was possible that the two states are caused by alternative disulphide bond configurations. The likelihood of this is reduced due to the inclusion of TCEP in the NMR buffer, but there is a possibility that disulphide bonds could form in a location that is inaccessible to TCEP. This can be tested by NMR, because the cysteine C β chemical shift has been shown to be predictive of redox state (Sharma and Rajarathnam, 2000; Mielke and Krishnan, 2009). As shown in Figure 8.10B, all three cysteines in both states have C β chemical shifts between 26 and 32 ppm, implying that they are all in the reduced state. The only cysteine to show significant C β chemical shift differences between the two states is C284. This residue is in the third β -strand, and all neighbouring residues also have large chemical shift differences between the two states. Therefore, the C β chemical shift difference can likely be attributed to a more general difference between the two states, rather than being related to redox state.

8.4.5 The two states do not represent different protonation states

The histidine at position 285 has the largest compound chemical shift between the two states (Figure 8.9A), raising the possibility that the two states differ in histidine protonation. The pK_a of histidine varies considerably depending on local structure and nearby residues, but is generally close to pH 6. This is close to the buffer pH, so it is plausible that a mix of protonation states could exist. To test this, the C α and C β chemical shifts of each histidine were examined (Figure 8.10C). For H285, there is a 1-2 ppm chemical shift difference between the two states, whereas for H300, there is no difference in chemical shift. While this is potentially consistent with a difference in protonation state at H285, a separate experiment shows that it is not responsible for the two-state behaviour: peak doubling is still present in the (¹H, ¹⁵N) HSQC of PACT-D3 H285A (Figure 8.10D), showing that protonation of H285 is not responsible for the presence of the two states.

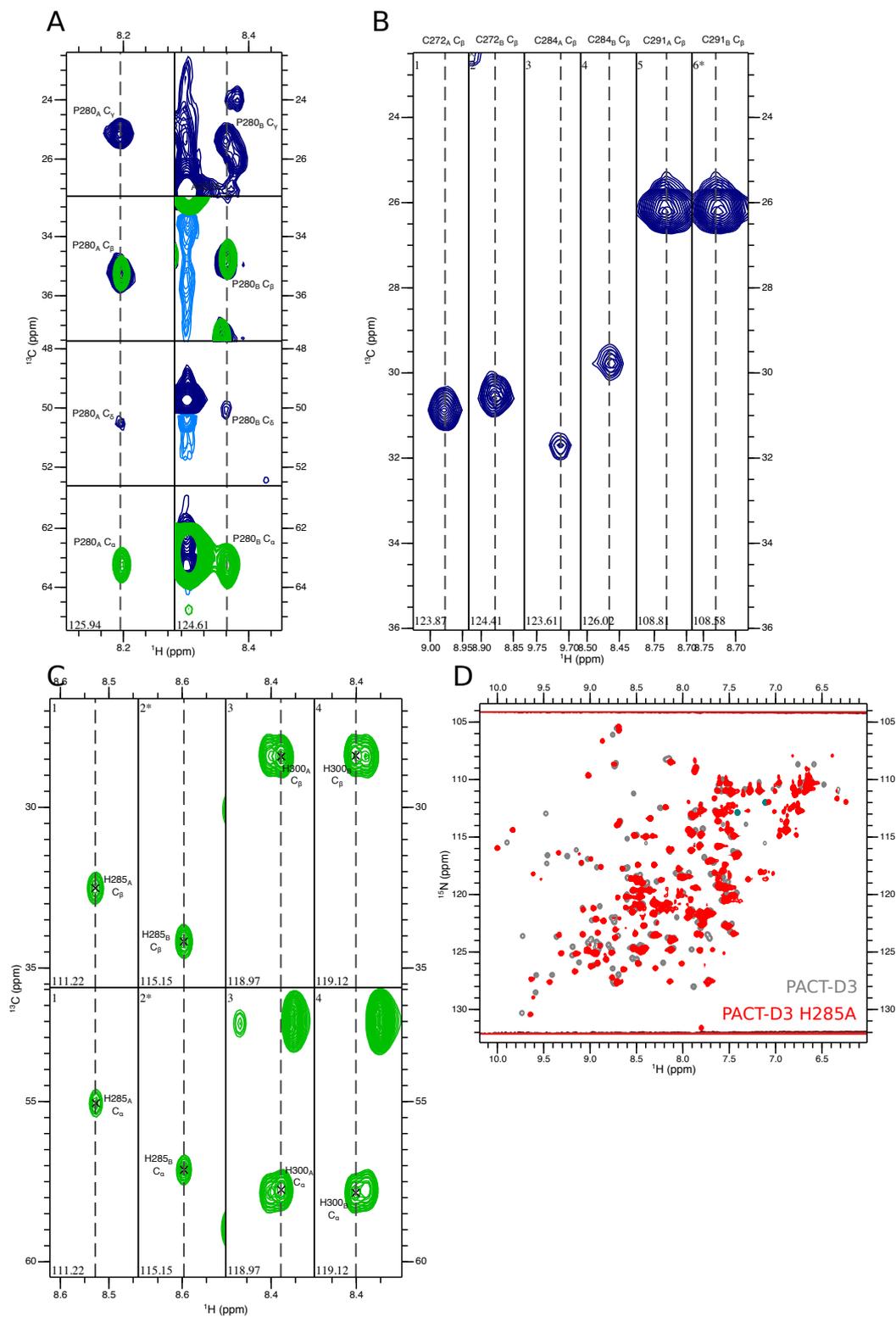


Figure 8.10. NMR spectra of PACT-D3 proline, cysteine and histidine residues

A) Strips from the CBCA(CO)NH spectrum showing P280 in each state. B) Strips from the CC(CO)NH-TOCSY spectrum showing all cysteine C_β peaks. C) Strips from the CBCA(CO)NH spectrum showing all histidine C_α and C_β peaks. D) Overlay of HSQC spectra of PACT-D3 H285A (red) on the wild-type domain (grey). See Chapter 4 for sample and data acquisition details.

8.4.6 The two states exchange at a rate inconsistent with domain swapping

The fact that two distinct peaks are observed for each residue in the (^1H , ^{15}N) HSQC shows that the two states of PACT exchange slowly with respect to the NMR timescale, if at all. This is potentially consistent with a phenomenon called domain swapping, where secondary structure elements become interchanged between the two dimer subunits (Bennet, Schlunegger and Eisenberg, 1995). The crystal structure of Loqs-D3 contains a domain swap of the third β -strand and C-terminal helix, so it is possible that PACT-D3 could also be swapped at this position. One feature of domain swapping is that it is a very slow process, taking on the order of days to weeks (Bennet, Schlunegger and Eisenberg, 1995; Rousseau *et al.*, 2001; Ghasriani *et al.*, 2014). If the two states of PACT-D3 also exchange on this timescale, it would suggest domain swapping is occurring.

The stability of domain swapped dimers arises because there is a high energy barrier between the native and domain swapped states (Liu and Eisenberg, 2002; Rousseau, Schymkowitz and Itzhaki, 2003). The domain swapped dimer has similar energy to the native state, because the domain swapped part forms native-like interactions with the other subunit. A significant degree of unfolding must occur to convert between the native and misfolded states, including the breaking of multiple native or native-like interactions. It is this unfolding that is primarily responsible for the slow rate of domain swapping, although a slow folding pathway can also increase the likelihood of domain swapping, by increasing the chance of contact between partially folded proteins (Rousseau, Schymkowitz and Itzhaki, 2003).

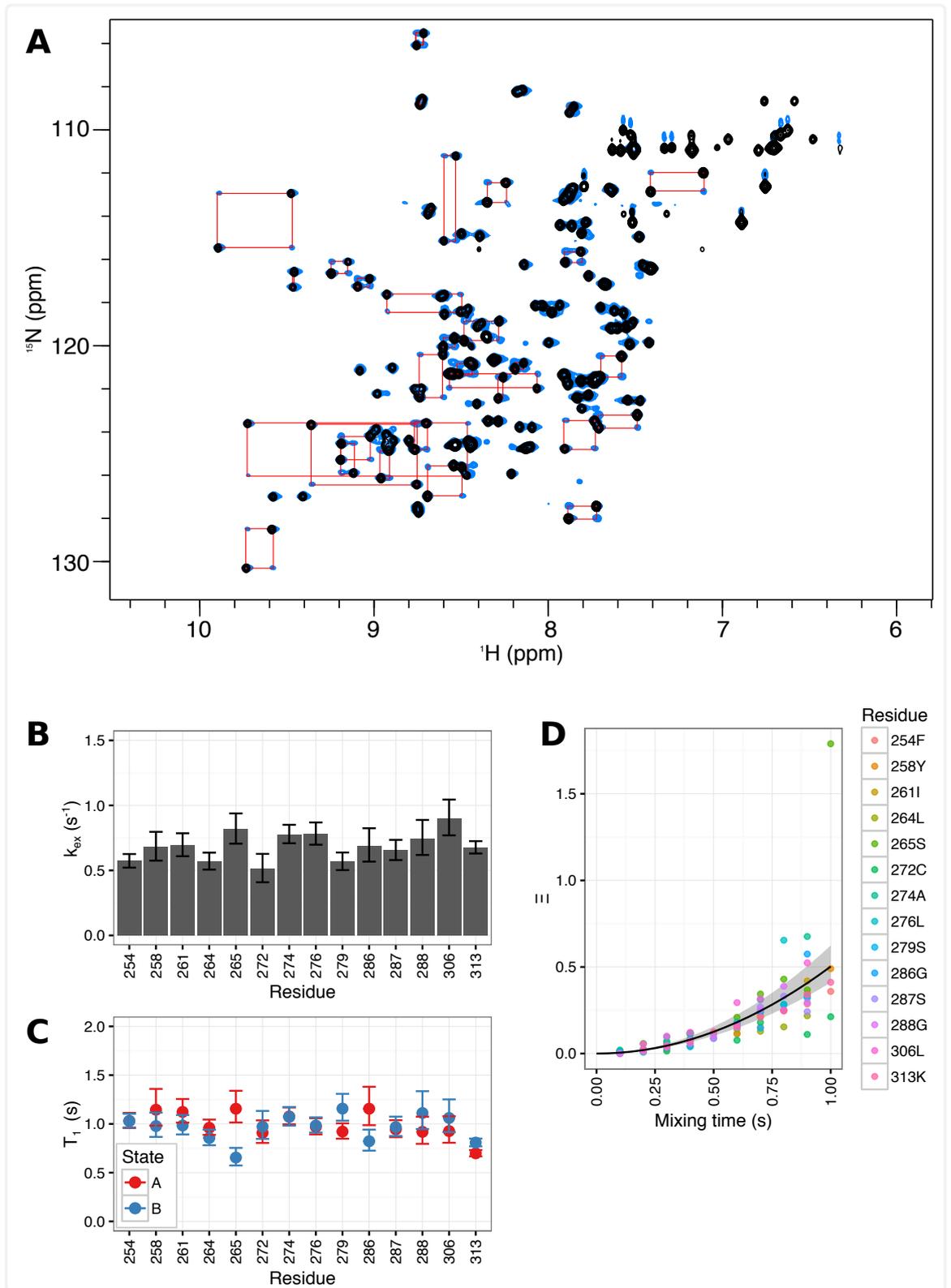
To test whether the two states of PACT-D3 are in exchange, EXSY spectra were recorded with mixing times between 0.1 and 1 second (Farrow, Zhang, *et al.*, 1994). The appearance and positions of new exchange peaks in Figure 8.11A shows that substantial exchange takes place within 0.5 seconds. Exchange cross peaks can be observed for all residues for which the two states have distinct proton and nitrogen chemical shifts. If either nucleus has the same chemical shift in both states, the exchange peaks overlap with the auto peaks (i.e. the peaks representing non-exchanged protein, which appear at the same positions as in the (^1H , ^{15}N) HSQC), and therefore cannot be resolved.

To calculate the exchange rate for individual residues, intensities of well-resolved cross peaks were fit to a two-state exchange model (Farrow, Zhang, *et al.*, 1994). To increase the data available for fitting, T_1 relaxation experiments were performed with delay times up to 2 seconds, and included in the fit (Farrow, Muhandiram, *et al.*, 1994). This is possible because the T_1 and EXSY pulse sequences are almost identical: the only difference is that the positions of t_1 and the relaxation/exchange delay are swapped. The peak intensities in the T_1 experiment are the sum of the auto and exchange peak intensities in the EXSY. In principle, this approach allows exchange peak intensity to be inferred from the difference between the T_1 and EXSY auto peaks, even if the exchange peaks are not resolvable in the EXSY spectra due to overlap. In practice, exchange rates calculated for residues with overlapped cross peaks tended to fall outside the range observed for residues with well-resolved cross peaks, suggesting the fit is less accurate. Therefore, the analysis below was restricted to residues with four well-resolved EXSY peaks.

The fit results, along with their 95% confidence intervals, are shown in Figure 8.11B,C. All residues shown have exchange rate constants (k_{ex}) between 0.5 and 1 s^{-1} . The similar exchange rates support the idea that PACT-D3 undergoes a global transition between the two states. Most residues have similar exchange rates, but S265, S279 and G286 show significant differences.

Figure 8.11. Chemical exchange in PACT-D3

A) Overlay of the (1H , ^{15}N) HSQC (black) with a (1H , ^{15}N) EXSY spectrum (blue) with a mixing time of 0.5 s. The auto- and exchange peaks from individual residues are linked in red, where resolvable. B) Exchange rate constants (k_{ex}) and C) longitudinal relaxation rates, calculated by fitting a two-state exchange model (allowing for differential relaxation) to EXSY and T_1 relaxation data for PACT-D3. Error bars denote 95% confidence intervals. D) Plot of the parameter $\bar{\Xi}$ calculated for each residue with different EXSY mixing times. The quadratic fit is shown as a solid line ($k_{ex} = 0.71 s^{-1}$), while the grey area shows the region bounded by the 95% confidence interval ($0.65 s^{-1}$, $0.86 s^{-1}$). The script used to perform the fitting and plotting is available as Appendix 4.



To capture the global exchange rates, well-resolved EXSY crosspeaks were analysed using a method which extracts the exchange rate while accounting for the possibility of different relaxation rates in the two states (Miloushev *et al.*, 2008). Least-squares fitting of the data gave a global exchange rate of 0.71 s^{-1} , with a 95% confidence interval of $(0.65 \text{ s}^{-1}, 0.86 \text{ s}^{-1})$ (Figure 8.11D). At later time points, there are large discrepancies between the data and the fit: this probably occurs because the quadratic relationship between \overline{E} and time breaks down as cubic terms in the series expansion of \overline{E} become significant. However, restricting the fit to time points less than or equal to 0.5 s did not substantially change the fitted exchange rate ($k_{\text{ex}} = 0.70 \text{ s}^{-1}$).

The global exchange rate is consistent with the exchange rates calculated for individual residues. By either method, the exchange rate is much faster than has been reported for domain swapping, and does not provide sufficient time for the degree of unfolding and refolding that would need to occur (Barrientos *et al.*, 2002; Rousseau, Schymkowitz and Itzhaki, 2003; Ghasriani *et al.*, 2014).

Additionally, if PACT-D3 unfolding was fast enough to for domain swapping to occur on this timescale, we would expect to see some evidence for this in PACT-D3 L273R in the form of exchange broadening: this is not the case. Therefore, the second state of PACT-D3 is unlikely to represent a domain-swapped dimer.

8.4.7 NOESY experiments show that PACT-D3 forms asymmetric dimers

Having ruled out the possibilities above, only two potential explanations remained for the two-state behaviour. Firstly, the two states could represent two different symmetric conformations of the PACT-D3 dimer, with similar free energy and with a low enough energy barrier to allow exchange. Alternatively, PACT-D3 could form asymmetric homodimers, with each half of the dimer giving rise to a separate set of peaks in NMR spectra.

To distinguish between these two possibilities, a ^{13}C -filtered NOESY-HSQC was recorded on a mixed sample of isotope-labelled PACT-D3, with 50% [^{13}C , ^{15}N]-labelled protein and 50% [^{15}N]-labelled protein (Figure 8.12). In this experiment, NOE cross peaks are only detected if the NOE donor is [^{12}C]-labelled, and the NOE acceptor is [^{13}C]-labelled. Since half of the protein molecules are uniformly

labelled with ^{13}C , the spectrum will contain only NOE cross peaks from dimers containing one $[^{13}\text{C},^{15}\text{N}]$ -labelled and one $[^{15}\text{N}]$ -labelled subunit.

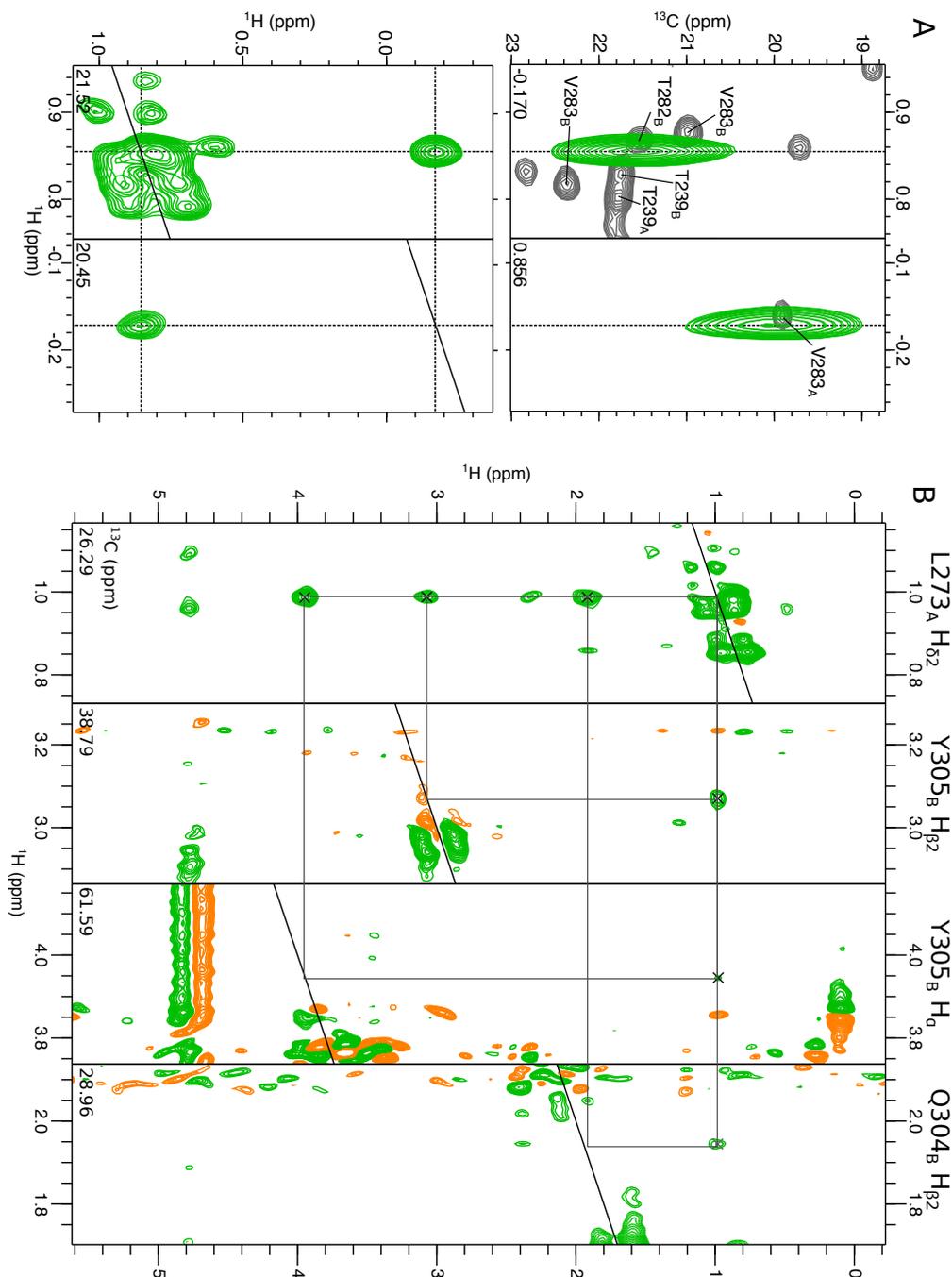


Figure 8.12. ^{13}C -filtered NOESY spectra of PACT-D3

A) Strips from the ^{13}C -filtered NOESY-HSQC (green) of PACT-D3, showing the NOE between methyl groups of V283_A and T282_B. The upper panel shows (^1H , ^{13}C) planes overlaid with the assigned (^1H , ^{13}C) constant-time HSQC (grey). The lower panel shows the (^1H , ^1H) planes of the same peaks. B) (^1H , ^1H) strips from the ^{13}C -filtered NOESY-HSQC showing NOEs between L273_A, Q304_B and Y305_B.

If PACT-D3 forms symmetric homodimers, then NOE cross peaks should appear only between protons from the same state. Alternatively, if PACT forms asymmetric homodimers, then NOE cross peaks should only be observed between protons from different states. As shown in Figure 8.12, the ^{13}C -filtered NOESY-HSQC demonstrates that the latter possibility is correct. Figure 8.12A shows the NOE between one of the methyl groups of V283_A, and the methyl group of T282_B. In Figure 8.12B, NOE cross peaks can be seen between L273_A and the H $_{\alpha}$ and H $_{\beta}$ protons of Q304_B and Y305_B. These prove that PACT-D3 forms asymmetric dimers.

These interactions are also consistent with PACT-D3 dimers adopting the same structure as Loqs-D3 dimers (Figure 8.13). They also allow the two NMR states to be assigned to specific chains of the dimer structure: state A observed by NMR correspond to chain A of the Loqs-D3 dimer, while NMR state B corresponds to the domain-swapped protomer formed from chains D and F. To maintain a consistent nomenclature between PACT and Loqs, the two Loqs domains will also be referred to as protomers A and B.

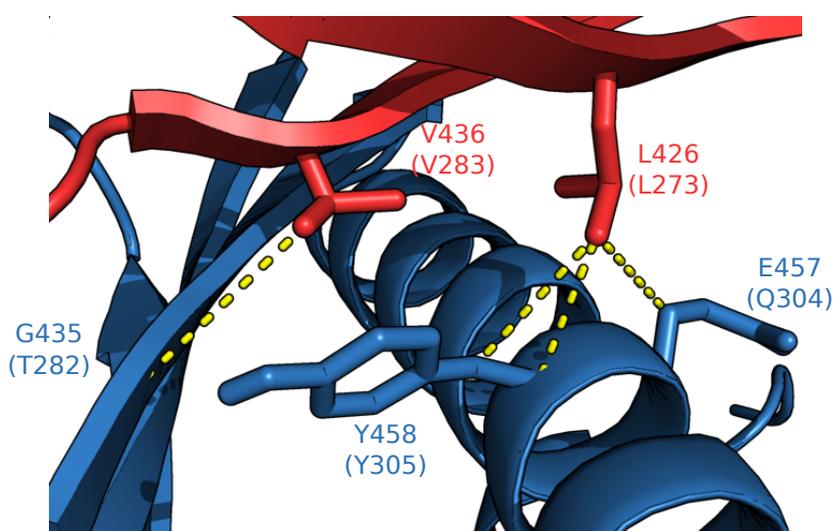


Figure 8.13. Intermolecular NOEs identified by ^{13}C -filtered NOESY

The NOEs from the ^{13}C -filtered NOESY-HSQC displayed on the equivalent residues of Loqs-D3 (4WYQ). Loqs-D3 chain A is coloured red, while chain D/F is coloured blue.

8.4.8 The third β -strands form an intermolecular parallel β -sheet.

As shown in Figure 8.9, the largest difference between the two states occur in the third β -strand. In the crystal structure of Loqs-D3, it can be seen that the third β -strands form an intermolecular parallel β -sheet (Figure 8.14A). Notably, this arrangement is not possible in a symmetric homodimer, as the symmetry would prevent alignment of hydrogen bond donors and acceptors (Figure 8.14B,C). To test whether the same geometry occurs in PACT-D3, NOESY data of residues in the third β -strand was examined. Figure 8.14D shows strips of V283 and H285 from the ^{15}N -NOESY-HSQC spectrum. An NOE is observed between V283_A and H285_B, while no NOE occurs between V283_B and H285_A. This is entirely consistent with the Loqs-D3 structure, where V436_A H_N is 4 Å from H438_B H_N, but V436_B H_N is greater than 10 Å from H438_A H_N.

Note that during the NOESY mixing time, approximately 10% of molecules will exchange between states. This results in the appearance of exchange cross peaks linking different states of the same spin (marked with asterisks in Figure 8.14D) regardless of spatial proximity. NOE cross peaks may also appear in these positions, but they cannot be distinguished from the exchange contributions.

The same pattern can be seen in the ^{13}C - and ^{15}N -NOESY-HSQC for the C284 H _{α} protons (Figure 8.14E,F). NOEs are observed between C284_A H _{α} and H285_B H_N, and also between C284_B H _{α} and V283_A H_N. This is equivalent to the Loqs-D3 structure, where there C437_A H _{α} is 3 Å from H438_{D/F} H_N, and C437_{D/F} H _{α} is 2.8 Å from V436_A H_N. The ^{15}N - and ^{13}C - NOESY-HSQC were recorded on different samples, which explains the slight differences in the H _{α} chemical shifts between these two spectra.

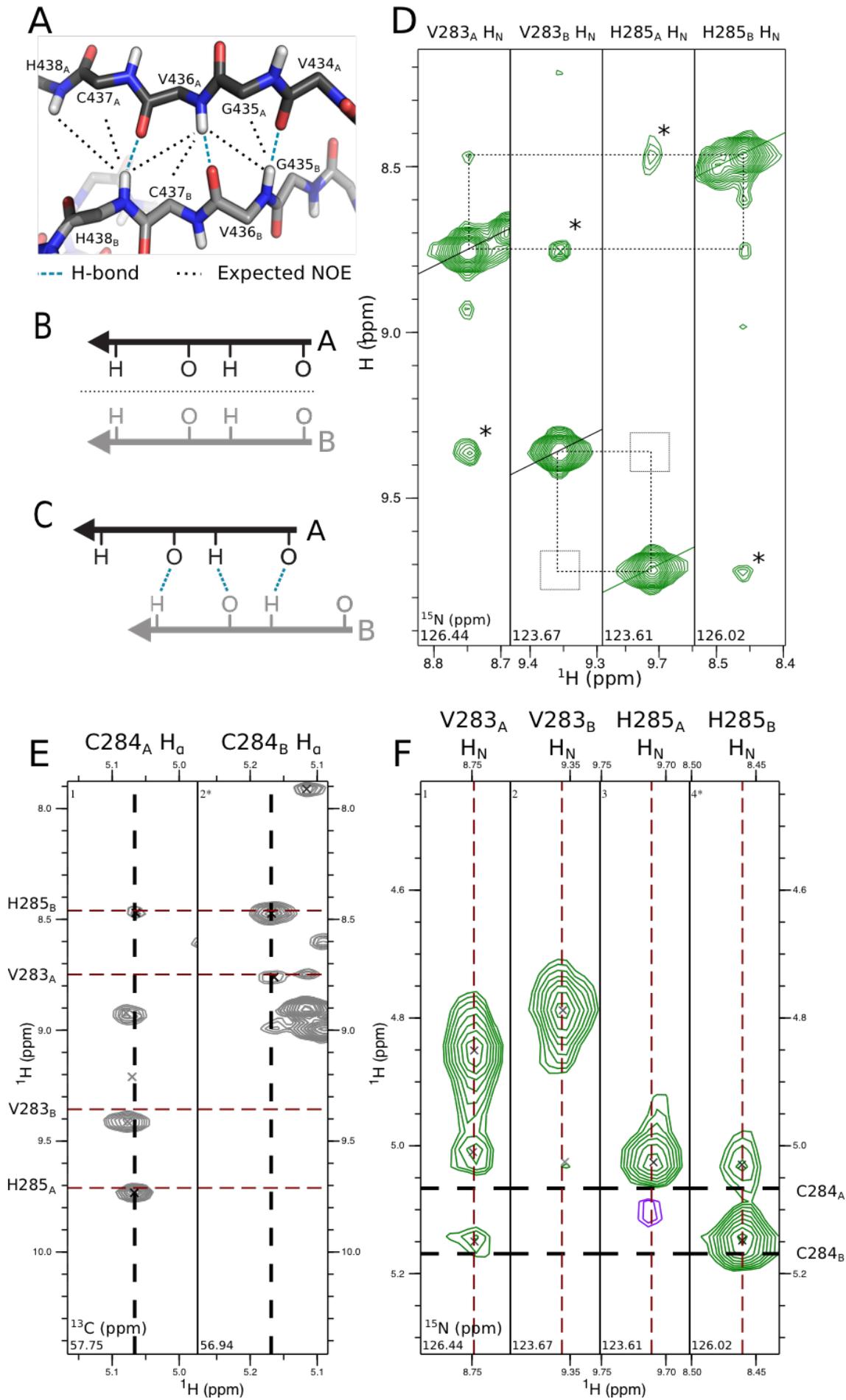


Figure 8.14. NOESY spectra provide evidence for asymmetry in strand β -3

A) Structure of the β -3 strands of the Loqs-D3 dimer, taken from chains A and D/F of PDB accession 4X8W. Intermolecular hydrogen bonds are displayed as blue dashed lines, while short intermolecular proton-proton distances are shown as black dotted lines. B) In a symmetric dimer, intermolecular parallel β -strands are forced to adopt a symmetric arrangement, preventing formation of hydrogen bonds. The dotted line shows the symmetry axis. C) If a register shift is introduced between the strands, the backbone amide and carboxyl groups are correctly positioned for hydrogen bond formation. However, no symmetry axis is possible. D) Strips from the ^{15}N NOESY-HSQC showing the amide protons of V283 and H285. Cross peaks caused by chemical exchange between states A and B are marked with asterisks. E) Strips from the ^{13}C HSQC-NOESY showing NOEs between H_α of C284 and the amide region. F) Strips from the ^{15}N HSQC-NOESY showing NOEs between the amide protons of V283 and H285, and the H_α region.

8.4.9 Loqs-D3 also forms asymmetric dimers in solution

If PACT-D3 forms asymmetric dimers and has the same structure as Loqs-D3, that implies that Loqs-D3 should also form asymmetric dimers in solution. To test this, (^1H , ^{15}N) HSQCs were recorded of wild-type Loqs-D3 (residues 392-463) and Loqs-D3 L426R (Figure 8.15). Expression vectors were kindly provided by Leonhard Jakob and Gunter Meister. The Loqs-D3 L426R spectrum has 100 peaks, close to the 90 expected based on sequence. The wild-type Loqs-D3 spectrum contains around 207 peaks, more than twice the expected number. As with PACT-D3, the peaks in the L426R spectrum do not overlay well with the wild-type spectrum. Although the spectra have not been assigned, and all alternative explanations have not been ruled out, it seems probable that this peak doubling is due to asymmetric dimerisation.

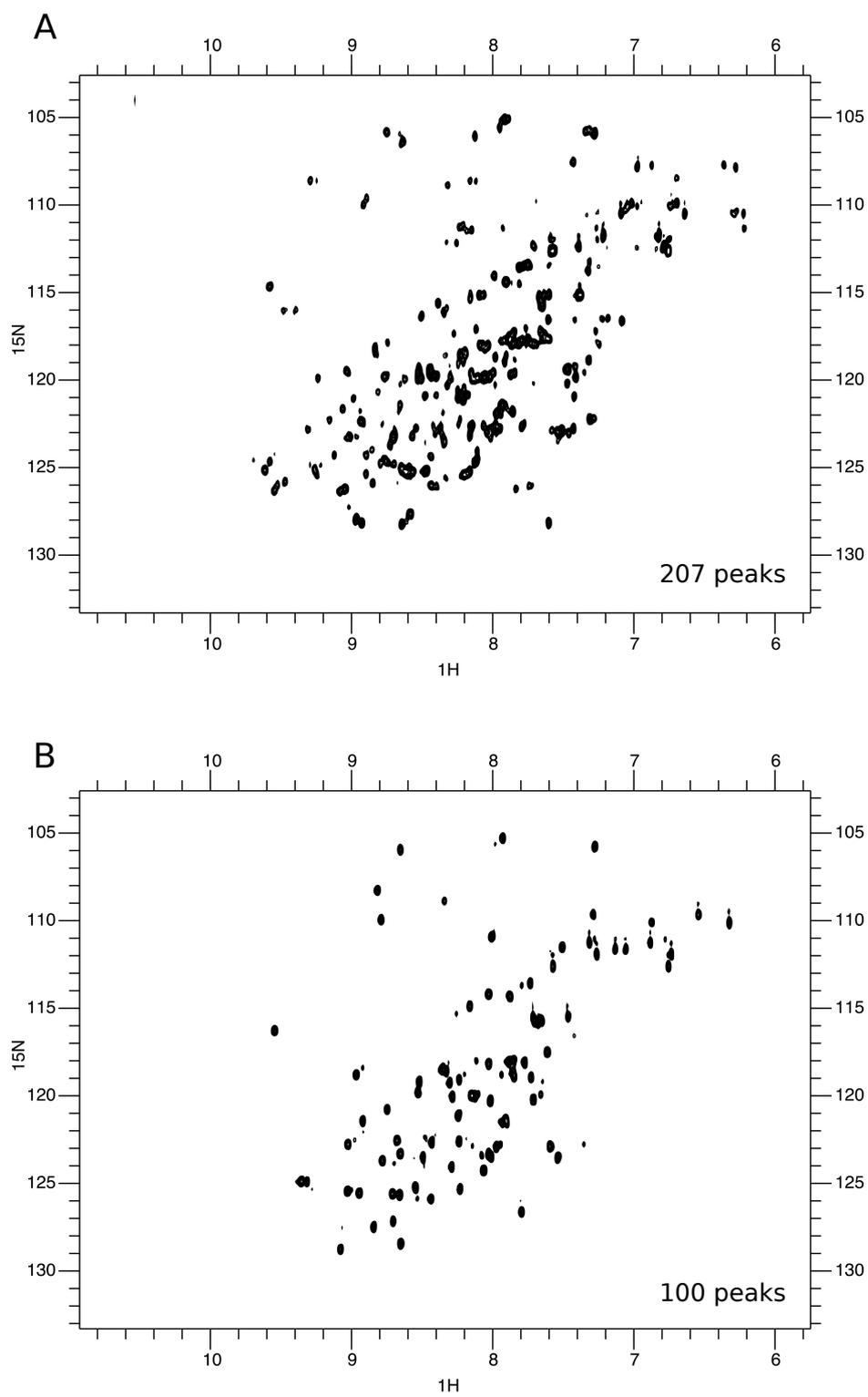


Figure 8.15. (^1H , ^{15}N) HSQC spectra of Loqs-D3 and Loqs-D3 L426R

(^1H , ^{15}N) HSQC spectra of A) 120 μM wild-type Loqs-D3 and B) 110 μM Loqs-D3 L426R. The buffer was standard NMR buffer, except with only 4.5 mM TCEP.

8.5 Evidence for the appearance of a monomeric state at low concentrations

As discussed in Chapter 7, sedimentation equilibrium AUC showed that PACT-D3 was predominantly dimeric at concentrations above 10 μM . To test whether it remained dimeric below this concentration, a SOFAST-HMQC spectrum of a 5 μM sample of PACT-D3 was recorded. To allow detection of such a low protein concentration, data was acquired on a 700 MHz spectrometer equipped with a cryoprobe, and a large number of scans (2400) were recorded. If the protein remains dimeric, the spectrum should be the same as for high concentration PACT-D3. If monomeric protein is present, peaks should appear in positions similar to the L273R mutant form of PACT-D3.

The NMR spectrum of 5 μM PACT-D3 is shown in Figure 8.16A, together with spectra of wild type and L273R mutant at high concentration. The majority of peaks in the 5 μM spectrum overlay well with high concentration wild-type PACT-D3, showing that the protein remains predominantly dimeric even at very low concentrations. A number of additional peaks appear in the 5 μM spectrum, mostly on or close to the positions of peaks in PACT-D3 L273R (Figure 8.16B-D). Because the SOFAST-HMQC uses a very short recycle delay of 0.2 s, peak intensity is not directly proportional to population (Schanda, Kupçe and Brutscher, 2005). Despite this, a number of well-resolved peaks were quantified, and were found to have fairly consistent intensities relative to the dimer state peaks (Figure 8.16E). This suggests the dimeric state of PACT-D3 is still more prevalent than the monomer at 5 μM , implying that the dissociation constant is on the order of 1 μM or lower.

It is not entirely clear why monomer peaks are observed for some residues but not others. The peaks that are present originate mostly from the α -helices; only C272 and G288 are visible from the β -sheet. This raises the possibility that peaks at the interface are being broadened, perhaps due to an exchange process with the dimer state.

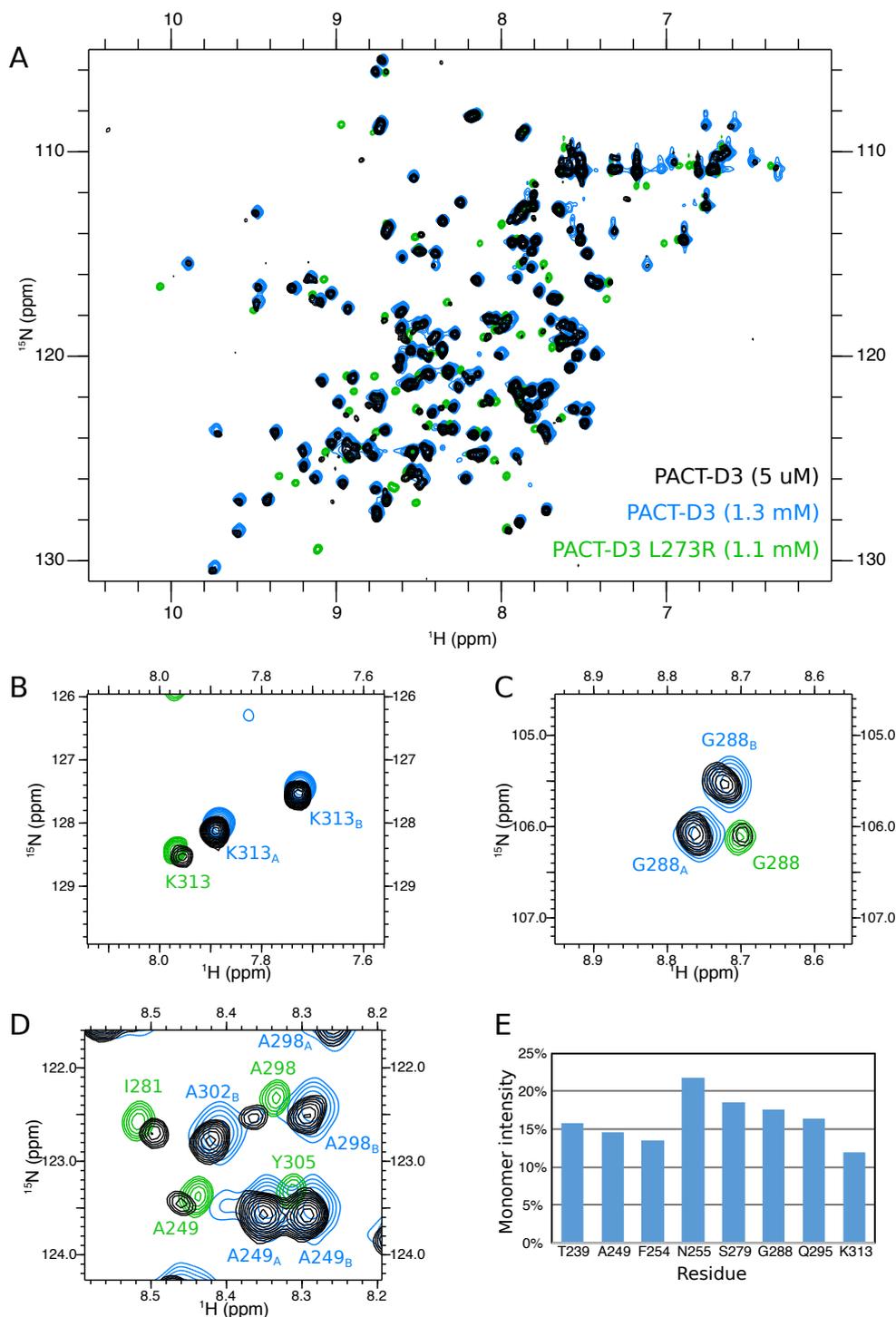


Figure 8.16. (^1H , ^{15}N) SOFAST-HMQC of low concentration PACT-D3

A) Overlay of (^1H , ^{15}N) SOFAST-HMQCs of $5\ \mu\text{M}$ PACT-D3 (black) and $1.3\ \text{mM}$ PACT-D3 (blue), and a (^1H , ^{15}N) HSQC of $1.1\ \text{mM}$ PACT-D3 L273R (green). All spectra were recorded at 25°C in standard NMR buffer. The $1.3\ \text{mM}$ PACT-D3 spectrum was recorded on a $600\ \text{MHz}$ instrument, the other two at $700\ \text{MHz}$. B-D) Magnified regions from the spectra in part (A). E) Quantification of the relative peak heights for various residues in the $5\ \mu\text{M}$ PACT-D3 SOFAST-HMQC. The Y axis shows the height of the monomer peak as a percentage of the sum of all three peaks from that residue.

8.6 NMR spectroscopy of PACT-D3 at high pressure

High pressure NMR was performed on PACT-D3 and PACT-D3 L273R with two aims. Firstly, high pressure has been reported to increase the dissociation constants of oligomers, allowing monomer to be observed at higher concentrations (Ingr 2016). Secondly, pressure-dependent chemical shifts indirectly encode information about solvent exposure, providing complementary evidence for the dimer interface found using NOESY experiments in section 8.4.7.

As discussed by Kitahara and coworkers, the amide proton chemical shift depends mostly on the length and angles of hydrogen bonds: these will differ depending on whether the hydrogen bond is to the protein backbone, to a side chain, or to the solvent (Wagner, Pardi and Wuethrich, 1983; Asakura *et al.*, 1995). Applying high pressure can cause compression of protein structure, so will alter intra-protein hydrogen bond lengths and angles (Weber and Drickamer, 1983; Nisius and Grzesiek, 2012). High pressure can also cause changes to the water packing at the protein surface, and favour the formation of unfolded states (which may have a lower volume due to imperfect side chain packing in the core of the folded protein). For solvent exposed amides, proton chemical shift tends to increase consistently with pressure, while amides forming hydrogen bonds within the protein generally show smaller or negative pressure-induced chemical shift changes.

Amide nitrogen chemical shifts also vary with pressure. However, the relationship is complex, and depends on hydrogen bonding, backbone dihedral angles and side chain conformation in roughly equal measure (Xu and Case, 2002). It is hard to predict confidently how these factors will vary with pressure, so less information can be extracted from the pressure-induced chemical shift changes of nitrogen nuclei.

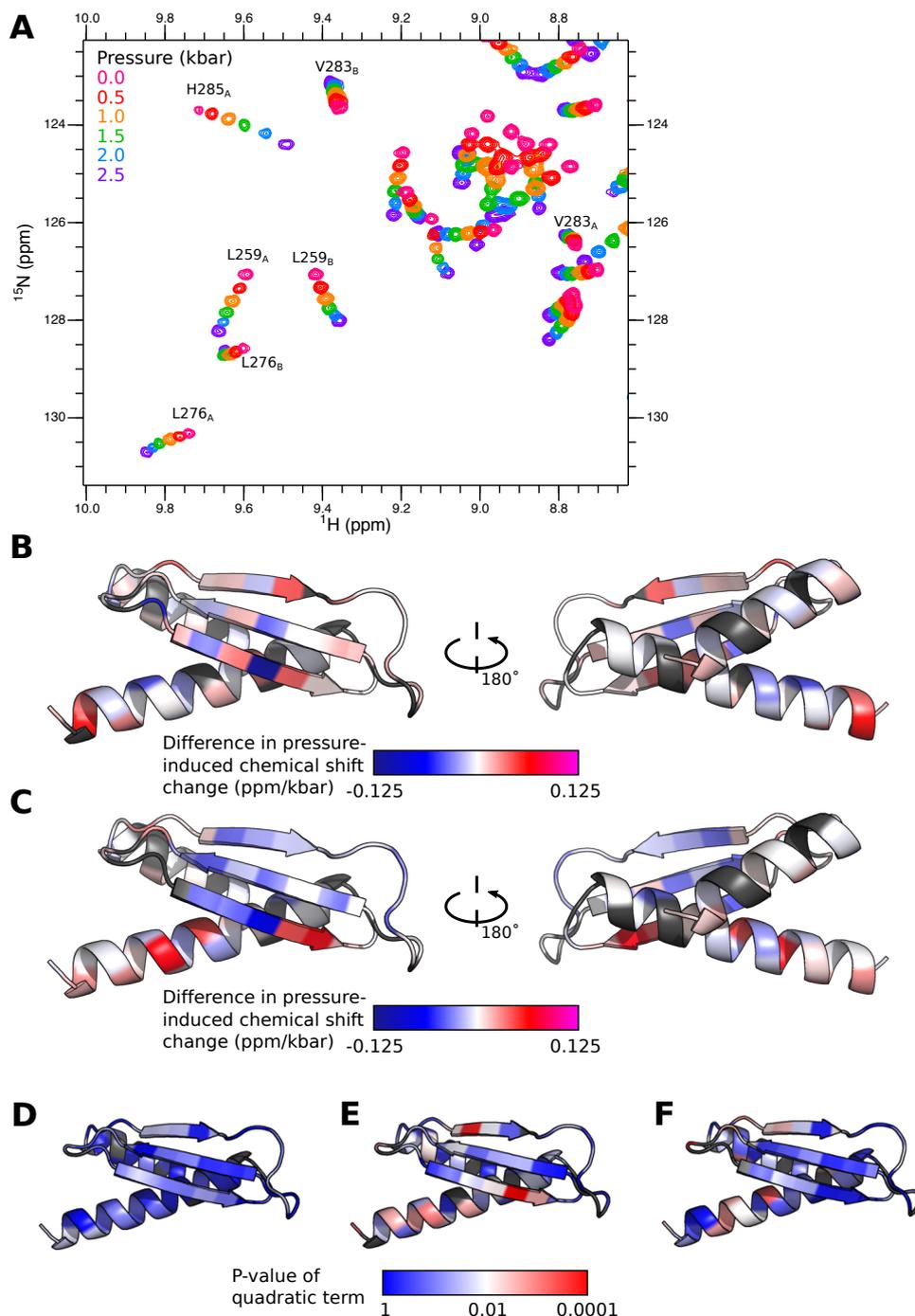


Figure 8.17. NMR of PACT-D3 and PACT-D3 L273R at high pressure

A) Overlay of (^1H , ^{15}N) HSQCs of wild-type PACT-D3 at different pressures. B) Difference in pressure-induced proton chemical shift change for PACT-D3 state A relative to PACT-D3 L273R, displayed on a model of PACT-D3. Dark grey indicates missing data. C) Difference in pressure-induced proton chemical shift change for PACT-D3 state B relative to PACT-D3 L273R. D-F) Statistical significance of quadratic coefficient of pressure-induced chemical shift change for (D) PACT-D3 L273R, (E) wild-type PACT-D3 state A, and (F) wild-type PACT-D3 state B. Low statistical significance is shown in blue, high significance is shown in red.

Away from the dimer interface, there is no difference in solvent exposure between the monomeric and dimeric forms of PACT-D3, so pressure-induced proton chemical shift changes should be similar between the wild-type and L273R mutant of PACT-D3. At the interface, dimerization will cause some protons to switch from being hydrogen bonded to water, to being bonded to protein or to nothing. This will cause a change in pressure-induced chemical shift change between the wild-type and L273R mutant, and so provides an NOE-independent method of evaluating the dimer interface.

In Figure 8.17A, (^1H , ^{15}N) HSQCs of PACT-D3 at different pressures are overlaid in order to show the changes in chemical shift that occur. For all residues, peak doubling remains throughout the entire pressure series, showing that PACT-D3 dimers are not disrupted over this pressure range. A range of behaviours can be seen: for example, both peaks of L276 show the expected increase of chemical shift with pressure. The proton chemical shift of H285_A has a negative pressure dependence; in contrast, the proton pressure-induced chemical shift of V283 is small but positive in both states, but the nitrogen chemical shift decreases as the pressure is increased.

To quantify this behaviour, peaks were assigned in all spectra, and their chemical shifts (referenced to DSS) were extracted. Linear models were then fitted to the proton chemical shifts for each state of wild-type PACT-D3, and for PACT-D3 L273R: the linear coefficient will be referred to as c_p^X , where X can be L273R, or dimer states A or B. The differences between the dimer states and the monomer are referred to as $\Delta c_p^{A/B} = c_p^{A/B} - c_p^{L273R}$. Figures 8.17B and 8.17C displays these differences in pressure-induced chemical shift change on a structural model of the domain.

In both states, the β -sheet and C-terminus of the protein show large changes in pressure-induced chemical shift coefficient, while the opposite face of the domain has only minor changes. In dimer state B (Figure 8.17C), most residues in the β -sheet have negative Δc_p^B values, consistent with this region being buried in the dimer. In dimer state A, some residues again have negative Δc_p^A values, particularly V283 and H285: V283 likely forms hydrogen bonds as part of the intermolecular β -sheet, while H285 is close enough to the dimer interface to

restrict its interaction with solvent (see Figure 8.14A). Other residues have more positive Δc_p^A values, which would at first appear to contradict the idea that this area is more buried in the dimer than the monomer. However, it can be reconciled if dimerization also induces a bending in the β -sheet, altering the degree of hydrogen bonding of these residues.

As mentioned above, high pressure acts to compress protein structure. If multiple states are present in fast exchange ($k_{ex} \gg \Delta\omega$), high pressure favours states with smaller volumes, leading to a change in the relative populations. This can be detected in high pressure datasets as a non-linear dependence of chemical shift on pressure. For each peak, a quadratic model was fitted using R, and the statistical significance of the quadratic term calculated (Figure 8.17D-F).

For PACT-D3 L273R (Figure 8.17D), the proton chemical shifts of most residues do not have a significant quadratic pressure dependence, indicating that there is minimal fast exchange between states, or at least that any exchanging states have similar chemical shifts or volumes. For both states of wild-type PACT-D3, a number of residues in helix α -2 and the β -sheet have statistically significant quadratic terms ($p < 0.01$) (Figure 8.17E,F). It is plausible that this is due to fast conformational exchange at the C-terminus of helix α -2, as TALOS-N predicts that the length of this helix differs between the monomeric and dimeric states (Figures 8.2 and 8.5). This also explains the quadratic pressure dependence of the start of strand β -1, as dimerisation brings this strand in close proximity to the C-terminus of the other protomer.

8.7 Discussion

8.7.1 State-specific assignment of PACT-D3

In order to prove that asymmetric dimerization was occurring, it was necessary to show NOE cross peaks between residues in different states. This is only possible if NMR signals have been correctly assigned to each state. As discussed earlier, state-specific assignment was performed in two stages: firstly by linking residues into fragments using $C\alpha$ and $C\beta$ chemical shifts, and then by linking these fragments using NOESY data.

After the first stage, there were 5 fragments of 7 residues or more with internally consistent states, and a number of smaller fragments of 1-4 residues. Each

fragment was separated from its neighbours by a residue with very similar $C\alpha$ and $C\beta$ chemical shifts. Three of the large fragments corresponded to the three β -strands, while the other two covered helix α -2. These could be fairly simply linked using NOEs between residues in the centre of the fragments, where the chemical shift differences between the two states are greater. This resulted in an internally consistent state assignment for residues 256-288 and 294-313, which includes all the residues used in Figures 8.12 and 8.14 to show asymmetric dimerisation (273, 282-285, 304, 305).

The shorter fragments corresponding to helix α -1 and the loop between β -3 and α -2 could not be linked using this method. This was because chemical shifts were similar between states throughout the whole fragment, and the short length meant there were fewer NOEs between fragments. Therefore, the state assignments of these residues should be considered arbitrary. The small chemical shift differences between the states suggests these regions have similar structures. Indeed, the crystal structure of Loqs-D3 shows that these areas are distant from the dimer interface. The lack of accurate state assignments does not affect calculation of the compound chemical shift.

8.7.2 Phosphorylation of PACT-D3

Several phospho-mimic mutants of PACT-D3 (S246D, S287D and S246D S287D) were studied both by SEC-MALLS (Chapter 7, section 7.4) and by NMR (section 8.3.2). The results of these investigations have been somewhat inconclusive. The SEC-MALLS data shows that they do not affect oligomeric state at high concentrations (350 μ M): however, the literature suggests phosphorylation should increase dimerisation affinity, an effect that cannot be detected when dimerisation is already saturated. The cellular concentration of PACT is unknown and likely context dependent. However, it is almost certainly lower than that used for SEC-MALLS, so it is possible that a change in affinity could affect the formation of dimers *in vivo*.

The simplest approach to investigate this further would be to measure the dimer dissociation constants of the different constructs. This is somewhat challenging, because homodimers will already be present at any concentration above at least 5 μ M. Microscale thermophoresis (MST) appears to be a promising approach, as this can detect fluorescently-labelled protein at nanomolar concentrations. Given the

evidence that PACT-D3 is partially monomeric at 5 μM (Section 8.5), a nanomolar solution of PACT-D3 should be entirely monomeric.

NMR data shows that the S246D mutation only changes the chemical environment of local residues, while S287D has an effect on distant residues. This can be explained with reference to the structure of Loqs-D3: the serine equivalent to S287 (S440) is close to the dimer interface, allowing it to affect helix α -2 on the other protomer. By contrast, the residue equivalent to S246 (G399) is on the opposite face of the protein from the dimer interface. This large separation also explains why the effects of the two mutations do not appear to be cooperative.

It has been reported that S246 phosphorylation is required before S287 can be phosphorylated (Peters, Li and Sen, 2006). It is not clear how this is compatible with their distant positioning, and the lack of communication between the two sites. There are a number of possible explanations for this discrepancy.

Firstly, the requirement for a specific ordering of phosphorylation could be imposed by the binding of other proteins. For example, phosphorylation of S246 may cause PACT to dissociate from one complex, and expose S287 for phosphorylation, perhaps as part of a separate complex.

Secondly, aspartate may not be a good model for phosphorylation in this case: the greater bulk or charge of a phosphate group might lead to a significant difference in behaviour. The two S287 serines are positioned relatively close to the dimer interface and to one another: it may be that phosphate groups are large enough to cause steric clashes, while aspartate is too small. This could perhaps be tested by using glutamate instead of aspartate to mimic phosphorylation, since it is slightly larger, although still only has a single negative charge. If there is a significant difference in behaviour between aspartate and phosphoserine at this position, it would cast doubt on the conclusions of other papers which use aspartate phospho-mimic mutations to investigate PACT phosphorylation (Singh et al., 2011; Singh and Patel, 2012).

Thirdly, the original evidence for phosphorylation may be misleading. The experiments in the paper by Peters and coworkers use a truncated version of PACT in which residues 301-313 are replaced with 5 unrelated amino acids due to a frameshift mutation (Peters, Li and Sen, 2006; Daher *et al.*, 2009), severely altering helix α -2. All subsequent studies have used aspartate and alanine

mutations to mimic or block phosphorylation, and there is no published direct detection of phosphorylation of PACT with the correct sequence. In addition, both S246 and to a lesser extent S287 are in relatively rigid regions of the protein, while phosphorylation sites are usually in flexible loops. These considerations suggest that phosphorylation has not yet been adequately demonstrated *in vivo*, and that further biochemical and biophysical investigations may be premature until the experiments of Peters and coworkers are replicated with the correct PACT sequence.

Biophysical Characterisation of TRBP Domain 3

9.1 Introduction

TRBP domain 3 has approximately 50% sequence identity with PACT domain 3, and the residues which form the dimer interface in PACT and Loqs are very well conserved in TRBP (Figure 1.9). However, the evidence for TRBP dimerisation in the literature is somewhat ambiguous: full-length TRBP was found to form dimers with a K_d of approximately 54 μM using analytical ultracentrifugation (Yamashita 2011). In contrast, no homo-interaction was detected for TRBP domain 3 by yeast two-hybrid, despite the same assay readily detecting PACT homodimerisation (Laraki 2008). This chapter attempts to answer two questions: whether TRBP domain 3 homodimerises, and whether its behaviour differs from that of PACT domain 3, given their sequence similarity.

9.2 NMR spectra of TRBP-Ext_{DD}-D3 show fewer peaks than expected

As described in Chapter 6, it was not possible to produce a construct of TRBP domain 3 containing only the dsRBD in sufficient quantities for biophysical characterisation, and appeared to be aggregated by 1D NMR (Figure 6.4). TRBP-Ext-D3, containing an additional N-terminal region conserved between TRBP and PACT, was somewhat more soluble, but significant improvements in yield and stability were achieved by adding the phospho-mimic mutation S283D and S286D (dubbed TRBP-Ext_{DD}-D3). These mutations were previously shown to enhance protein expression/stability *in vivo* (Paroo 2009), although the mechanism for this stabilisation is unknown.

The (^1H , ^{15}N) HSQC spectrum of TRBP-Ext_{DD}-D3 suggests that the domain is folded, as judged by the distribution of peaks between 7 and 9 ppm. The sidechain amides also give rise to a number of discrete signals centred around approximately (7.2, 113) ppm, whereas an unfolded protein would be expected to have poorly dispersed overlapping peaks in this region. However, there are considerably fewer NH and NH₂ signals in the spectrum than expected: only 66, compared to the 129

predicted based on the sequence. This suggests that some peaks are broadened beyond detection by fast T_2 relaxation or by intermediate exchange (i.e. where $k_{ex} \sim \Delta\omega$, the exchange rate is similar to the frequency difference between the two chemical shifts). Several different exchange processes could potentially be responsible, including a change in conformation, a change in oligomeric state, or binding of a buffer component.

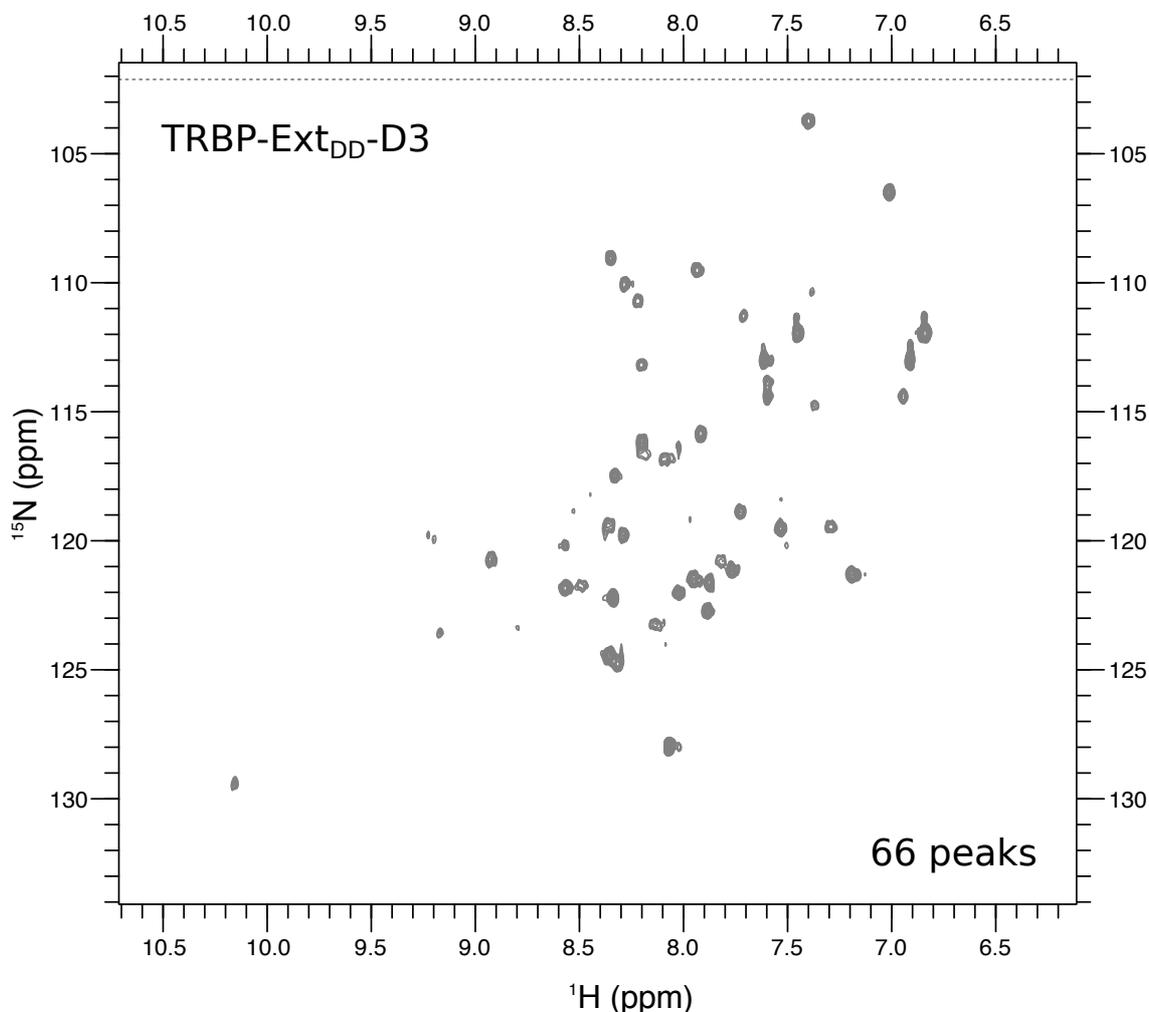


Figure 9.1. (^1H , ^{15}N) HSQC of TRBP-Ext_{DD}-D3

The sample concentration was 350 μM , in 18 mM Tris, 180 mM NaCl, 1 mM DTT, 10% D₂O, 50 μM DSS, pH 7.5. Peaks were picked and curated manually.

9.3 SEC-MALLS shows that TRBP-Ext-D3 is in equilibrium between monomer and dimer

Both TRBP-Ext-D3 and TRBP-Ext_{DD}-D3 were examined by SEC-MALLS, using the method described previously (Figure 9.2). Both constructs eluted with similar volumes (13.0 ml and 12.8 ml respectively) and were calculated to have similar

masses (16.9 kDa, 16.7 kDa). This is approximately 1.4 times the predicted monomer molecular mass of 12.0 kDa: since SDS-PAGE shows a single species of the correct molecular weight, this means the peak contains a mixture of oligomeric states. In principle this could include trimers, tetramers and other higher order oligomers, but the simplest assumption is that it contains a mixture of monomers and dimers.

Dimeric PACT-Ext-D3 eluted at 12.3 ml, while monomeric PACT-Ext-D3 L273R eluted at 13.5 ml (Figures 7.1 and 7.4). If we assume that TRBP-Ext-D3 behaves similarly, this difference in elution volume should be sufficient to resolve monomer and dimer peaks, provided the lifetime of each oligomeric state is longer than the elution time. The fact that we see a single intermediate peak indicates that the monomer and dimer states must exchange over the course of the experiment. This does not put a very strong bound on the lifetime of the dimer state, only telling us that it is less than approximately 30 minutes.

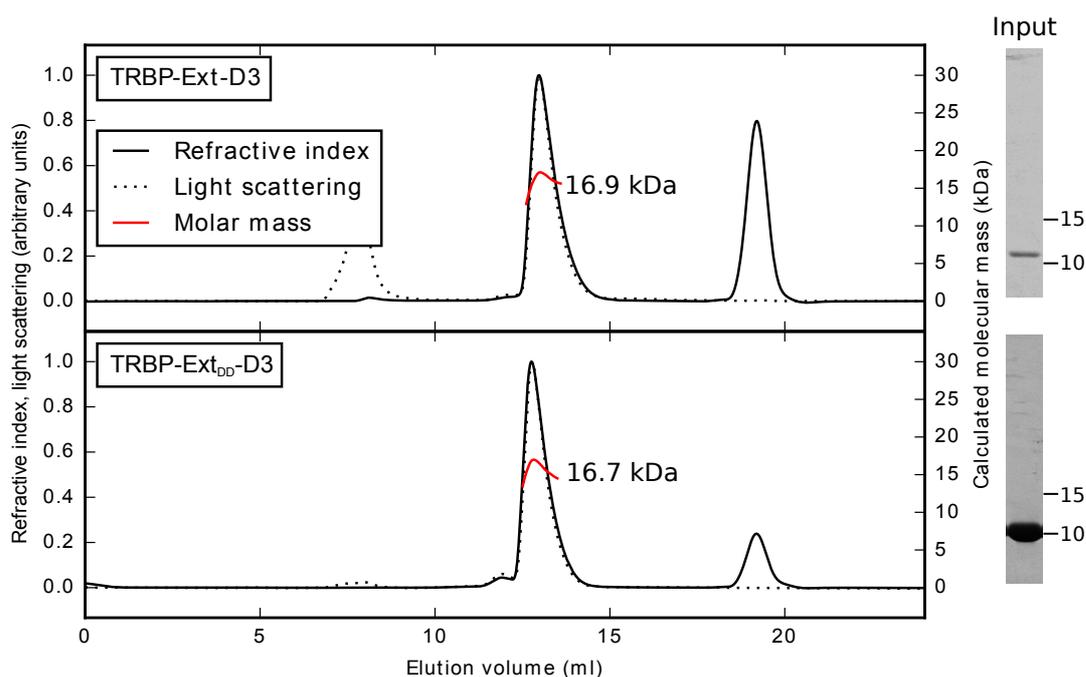


Figure 9.2. SEC-MALLS of TRBP-Ext-D3 and TRBP-Ext_{DD}-D3

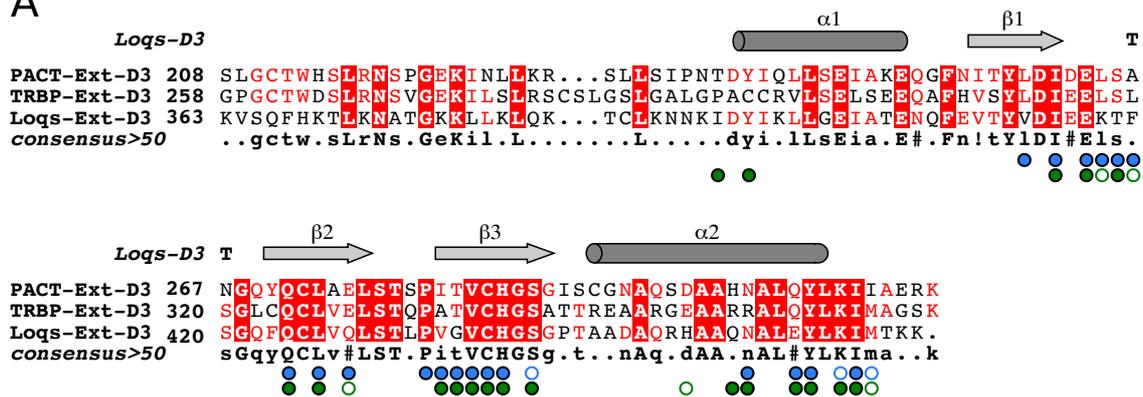
Differential refractive index is displayed as a solid black line; Rayleigh ratio (a measure of light scattering) is shown as a dotted line. The weight-averaged molar mass calculated over the centre of the peak is displayed next to each peak. SDS-PAGE analysis of the input is shown to the right.

9.4 Mutations in TRBP-Ext-D3 can inhibit or enhance dimerisation

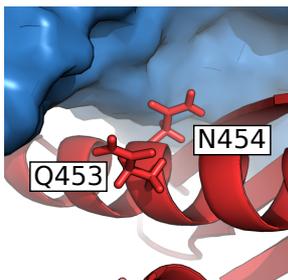
The SEC-MALLS data presented above demonstrates that TRBP domain 3 self-association is weaker than in PACT and Loqs domain 3, despite the high sequence similarity. The change in affinity is presumably due to differences in amino acids at the dimerisation interface. A sequence alignment of the three domains is shown in Figure 9.3A; residues that are buried by dimerisation in the Loqs-D3 dimer (as calculated by PISA (Krissinel and Henrick, 2007)) are indicated below the consensus sequence. The majority of these residues (19/26) are conserved between TRBP and at least one of the dimeric proteins, so are unlikely to be responsible for weakening dimerisation. Of the 7 residues which do not match, A334 and E350 have similar chemical properties to the equivalent residues in PACT, and A292, C294 and L319 are part of an N-terminus/loop β -1,2 interaction that may be a crystal artefact. In contrast, R353 and R354 have bulkier, more charged side chains than the residues they replace, and could cause steric clashes upon dimer formation. Intriguingly, the equivalent residues in Loqs-D3 (Q453, N454) experience significantly different environments in the two halves of the dimer (Figures 9.3B,C), being partially exposed in one protomer and buried in the other.

To test whether these residues are responsible for the weaker dimerisation affinity of TRBP, the mutations R353H and R354HN were introduced, replacing both arginines with the equivalent residues from PACT. Separately, L326 was mutated to arginine, to test whether this blocked dimerisation as in PACT and Loqs. During purification, the L326R mutant appeared to be less stable than the RR353,354HN mutant and more prone to aggregation, resulting in a significantly lower overall yield (L326R: 2 mg; RR353,354HN: 20.6 mg). When examined by SEC-MALLS (Figure 9.3D), TRBP-Ext_{DD}-D3 L326R eluted at 13.5 ml and had a mass of 12.5 kDa, consistent with a monomer and comparable to PACT-Ext-D3 L273R. TRBP-Ext_{DD}-D3 RR353,354HN eluted at 12.2 ml, with an estimated mass of 22.4 kDa, consistent with a dimer and similar to wild-type PACT-Ext-D3. This demonstrates that the TRBP domain 3 dimer interface is the same as that in PACT and Loqs, and that R353 and R354 are the main residues responsible for the weaker dimerisation of TRBP domain 3.

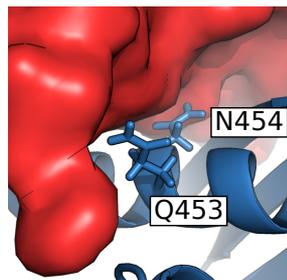
A



B



C



D

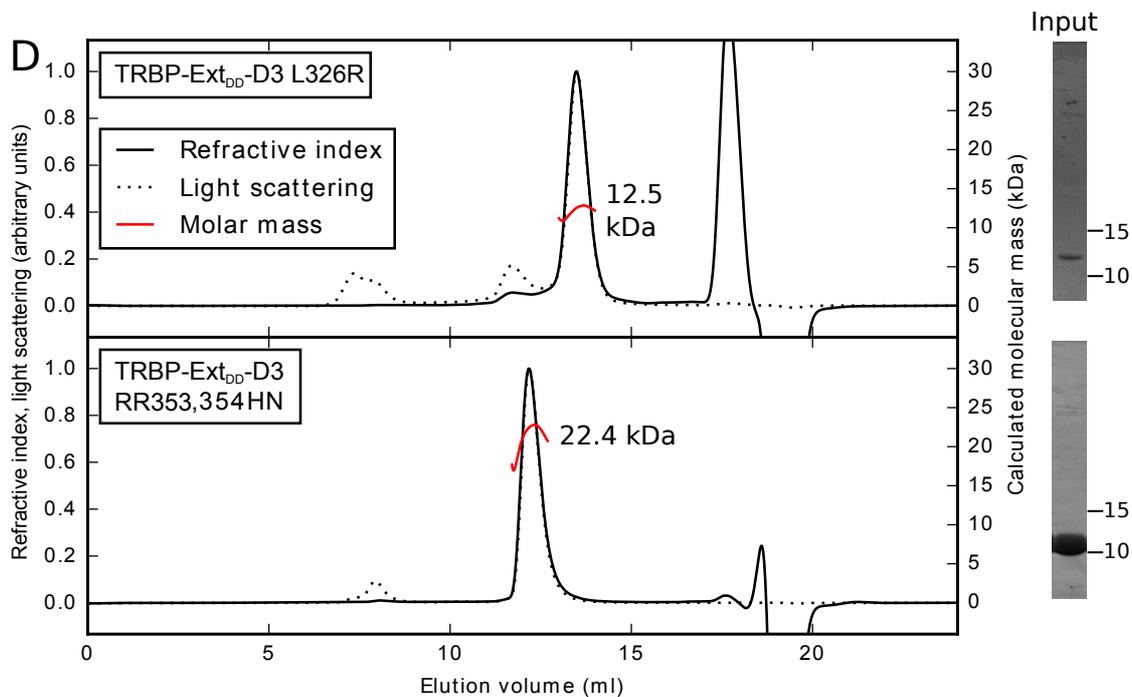


Figure 9.3. SEC-MALLS of TRBP-Ext_{DD}-D3 L326R and RR353,354HN

A) Sequence alignment of the third domains of PACT, TRBP and Loqs, with Loqs-D3 secondary structure shown above. The circles below the alignment show the residues which are buried by dimerisation in protomer A (blue) or protomer B (green) of the Loqs-D3 crystal structure (4X8W) (Jakob et al., 2016). Open circles are < 30% buried, filled circles are > 30% buried. Calculations were performed using the PDBePISA server (Krissinel and Henrick, 2007). B,C) Residues Q453 and N454 in the Loqs-D3 crystal structure, in B) protomer A, or C) protomer B. D) SEC-MALLS traces of TRBP-Ext_{DD}-D3 with either L326R or RR453,354HN mutations. Differential refractive index is displayed as a solid black line; Rayleigh ratio is shown as a dotted line. The weight-averaged molar mass calculated over the centre of the peak is displayed next to each peak. SDS-PAGE analysis of the input is shown on the right.

9.5 NMR suggests that TRBP-Ext_{DD}-D3 RR353,354HN forms asymmetric dimers

(¹H, ¹⁵N)-HSQC spectra were recorded of the L326R and RR353,354HN mutants of TRBP-Ext_{DD}-D3. Both spectra are well dispersed with generally well-resolved peaks, indicating that both constructs are folded. The largest difference between them is in the number of observable signals: 123 peaks are visible for TRBP-Ext_{DD}-D3 L326R, while the RR353,354HN spectrum contains 211 peaks. Since the number of signals based on sequence should be 129, the L326R mutant has the expected number of peaks, while the RR353,354HN mutant gives rise to close to twice the expected number of peaks. This is exactly the same behaviour as was seen for monomeric and dimeric forms of PACT and Loqs domain 3, suggesting that TRBP-Ext_{DD}-D3 RR353,354HN also forms asymmetric homodimers. The weak dimers formed by TRBP-Ext_{DD}-D3 are presumably also asymmetric, since they utilise the same interface as the RR353,354HN mutant. However, peak doubling is not observed in that case because too many of the peaks are broadened beyond detection by intermediate exchange.

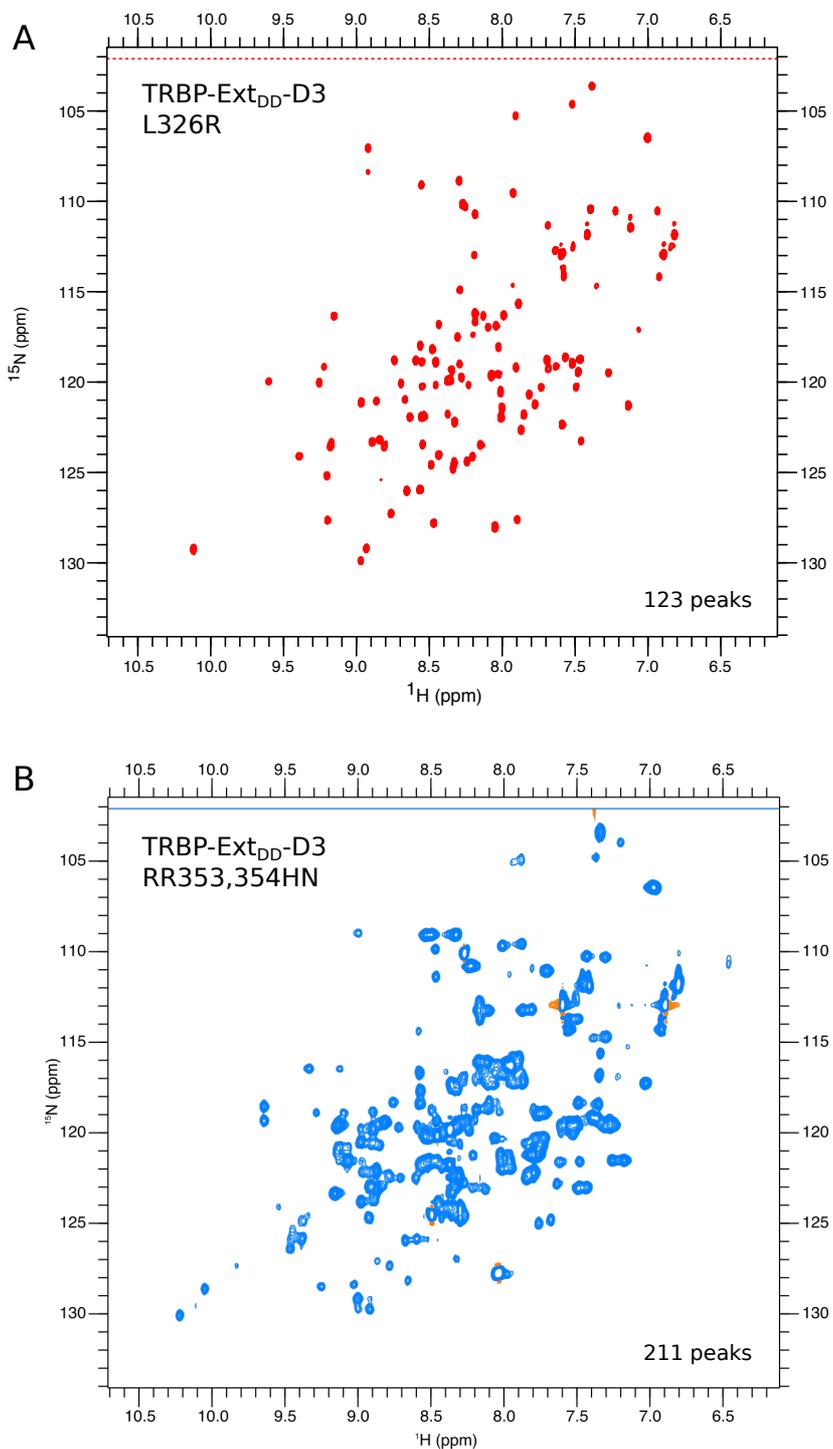


Figure 9.4. (^1H , ^{15}N) HSQC spectra of TRBP-Ext_{DD}-D3 L326R and RR353,354HN

A) 50 μM TRBP-Ext_{DD}-D3 L326R, or B) 1 mM TRBP-Ext_{DD}-D3 RR353,354HN, in 18 mM MES, 45 mM NaCl, 5 mM DTT, 10% D₂O, 50 μM DSS, pH 6.5. Peaks were picked and curated manually.

9.6 Discussion

9.6.1 Factors affecting the stability of TRBP domain 3

Qualitatively, TRBP domain 3 appears to be less stable *in vitro* than PACT domain 3. The stability appears to be affected by both the presence of the N-terminal extended region and by dimerisation. It was not possible to produce the dsRBD alone in sufficient amounts for study; adding the N-terminal extension increased solubility somewhat, while adding phospho-mimic mutations to the extension increased stability further. These additions enhanced the yield of protein that could be purified from a given culture volume, and also the melting temperature calculated by differential scanning fluorimetry (Figure 6.3). This effect on stability can be explained with reference to the structure of TRBP in complex with Dicer, which shows part of the N-terminal region forming a helix against the dsRBD (Figure 8.3B) (Wilson *et al.*, 2015). The N-terminal extension presumably forms additional attractive interactions with the core dsRBD, stabilising the overall structure. It is unclear whether this is a genuine part of its function *in vivo*, although the report that S283D and S286D mutations stabilise TRBP in cells suggests that it might be (Paroo *et al.*, 2009). The exact residues responsible for the stabilisation effect remain to be defined.

Dimeric TRBP-Ext_{DD}-D3 RR353,354HN also appeared to be more stable than the monomeric or partially dimeric versions of the domain. This can again be explained through the formation of attractive interactions upon dimerisation, forming a greater free energy barrier to unfolding and subsequent misfolding or degradation. It is unclear whether the difference in stability is functionally relevant.

9.6.2 Possible causes for intermediate exchange in TRBP-Ext_{DD}-D3 NMR spectra

The clearest difference between the NMR spectra of TRBP-Ext_{DD}-D3 and the equivalent region of PACT is in the number of observable signals: instead of the peak doubling seen for PACT, TRBP-Ext_{DD}-D3 has fewer signals than expected. This phenomena was also observed in a previous study of TRBP-D3 (Wilson *et al.*, 2015). The reduced number of signals is most likely due to intermediate exchange between monomer and dimer states. SEC-MALLS shows that a mix of oligomeric states are present at this concentration, and mutagenesis shows that blocking

dimerisation completely (by introducing L326R) results in the expected number of peaks in the (^1H , ^{15}N) HSQC spectrum. The other possible causes of exchange (conformational changes in the N-terminal extension, binding of buffer components) can be ruled out, as they would be expected to equally affect the wild-type and L326R mutants. Further evidence in support of this hypothesis is that the majority of peaks in the TRBP-Ext_{DD}-D3 spectrum overlay well on peaks in both L326R and RR353,354HN spectra (data not shown). To be the same in both of these spectra, the residue must be relatively unaffected by dimerisation, so would also not be affected by exchange between oligomeric states. Given that the line broadening is due to exchange, the exchange rate must be similar to the difference in chemical shifts between monomeric and dimeric states. For PACT-D3, the mean frequency difference between monomer and dimer was approximately 100 Hz for amide protons on a 700 MHz spectrometer, implying that TRBP-Ext_{DD}-D3 exchange rate is also on the order of 100 s⁻¹.

Development of Tools for Characterisation of PACT Dimerisation in Mammalian Cells

10.1 Introduction

All the experiments described so far were carried out on purified recombinant protein samples. While this approach is useful for studying molecular mechanisms, one must be careful when extrapolating conclusions found *in vitro* to the more complex crowded environment inside living cells. Functional studies in cells are therefore necessary to validate and further characterize the behaviours observed in biophysical experiments. This chapter does not present the results of such experiments, but does present initial steps towards setting up the tools and systems that would be needed to perform them.

10.2 Knockdown of PACT by siRNA

One strategy to validate hypotheses about protein function is to deplete the protein in cells, and observe how gene expression and cell behaviour change. A common approach is to transfect cells with siRNA targeting that gene – this generally results in partial downregulation at the protein level, although the magnitude of the effect size can vary significantly between different targets. Knockdown can be less effective on stable proteins with low turnover rates, where the experiment is too short for substantial degradation of protein to occur. In other cases, other transcriptional or post-transcriptional mechanisms may counteract the siRNA-mediated knockdown.

Three cell types were transfected with siRNA targeting PACT according to the protocol described in Chapter 5. HeLa and HEK293 cells were chosen due to their ease of growth and transfection. Primary human foreskin fibroblasts (Hff cells) were used because they are a better model of ‘real’ cells in their native context, compared to transformed cells containing multiple mutations and genome rearrangements.

qRT-PCR was performed on RNA extracted from Hff cells (Figure 10.1). This shows a large reduction in PACT at the mRNA level, without affecting the mRNA

levels of other genes involved in miRNA processing, or affecting the levels of miR-21, a miRNA highly expressed in Hff cells. A drop of this magnitude would be expected to result in reduced protein expression, but this would need to be confirmed by Western blot.

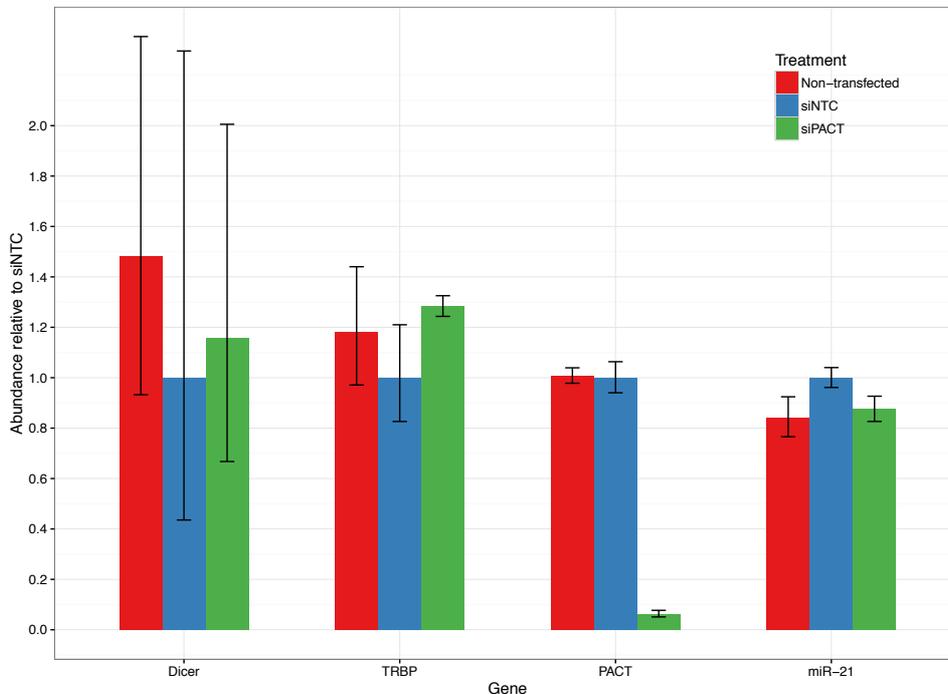


Figure 10.1. qRT-PCR of miRNA-related genes after treatment with siPACT

All results were first normalized to a housekeeping gene (HPRT for Dicer, TRBP and PACT; U6 for miR-21), then normalized to the non-targeting siRNA condition. Error bars show the standard deviation of three technical replicates.

The levels of PACT following siRNA transfection of HeLa and HEK293 cells were examined by Western blotting (Figure 10.2). In both cell lines, the levels of the PACT band (at approximately 34 kDa) are decreased in the siPACT condition relative to the controls, showing that the knockdown was successful.

Unfortunately, no staining for a loading control was performed for this blot.

Multiple bands are observed in addition to the expected band at 34 kDa. The lower molecular weight band probably represents either an isoform or a degradation product of PACT, consistent with the drop in intensity in the siPACT condition. The higher molecular weight band is presumably due to non-specific staining (as no

larger isoforms of PACT are known), although this does not explain the apparent drop in intensity of this band in HeLa cells treated with siPACT.

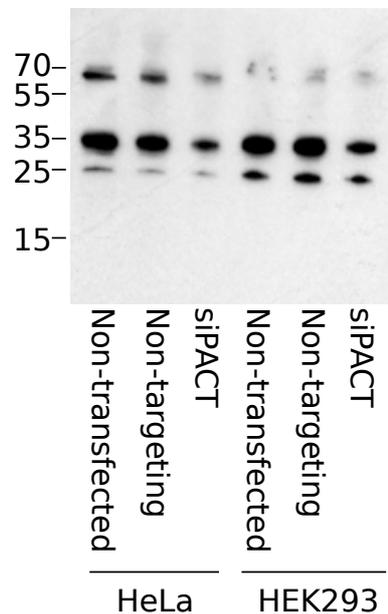


Figure 10.2. Test of siRNA targeting PACT

HeLa and HEK293 cells were transfected with siRNA as described in Chapter 5, Section 5.2. 10 µg of cell lysate per lane was then run on a 15% SDS-PAGE gel, and blotted with an α-PACT antibody as described in Section 5.5.

10.3 Overexpression of PACT

While siRNA knockdown would allow us to see how the function of the cell is perturbed when PACT levels are reduced, overexpression of PACT would allow hypotheses about dimerisation and the role of individual domains to be tested more explicitly. The approach taken here was to clone various PACT constructs into a plasmid with a cytomegalovirus (CMV) promoter, and encoding for puromycin resistance (pSF-CMV-Puro-NH2-FLAG/Cmyc, Oxford Genetics; Figure 10.3). This method offers a number of favourable properties:

- 1) Overexpression can be transient, or stable cell lines can be generated (if integrated plasmids are selected for with puromycin).
- 2) Plasmids can be transfected into stable cell lines, allowing the expression of multiple constructs in the same cell.

- 3) The plasmid can be transfected directly, and does not require a separate virus production stage (as in lentiviral overexpression).
- 4) Plasmid construction and cloning can be carried out in *E. coli*.
- 5) The CMV promoter allows expression in a broad range of cell types.

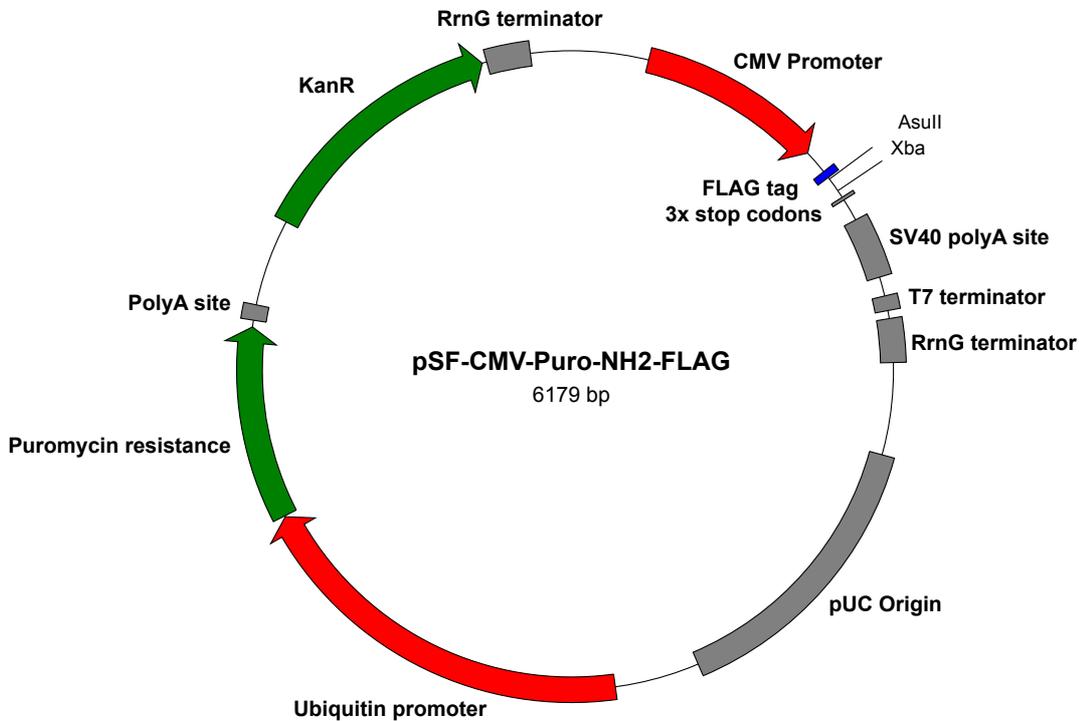


Figure 10.3. Major elements of the pSF-CMV-Puro-NH2-FLAG plasmid

Promoters are shown in red, resistance markers in green, and all other features in grey.

This system would allow several different types of experiment to be performed. Firstly, the effect of PACT overexpression on the levels of mature miRNA, or on sensitivity to cellular stress, could be measured, allowing probing of the biological function of PACT. Secondly, the interactions of PACT with endogenous proteins could be examined by immunoprecipitation. Thirdly, FLAG and C-myc tagged PACT constructs could be overexpressed in the same cell, and homodimerisation assessed using immunoprecipitation or proximity ligation assay. Finally, any of the above experiments could be repeated with mutant forms of PACT, to test the contributions of specific residues to binding or function.

Unfortunately, developments in the wider research field led me to focus on biophysical studies at the expense of the cell biology component of this project, and meant that I was unable to perform the above experiments. However, I have

prepared the expression plasmids, and carried out initial tests of the expression system.

Full-length PACT, PACT-Ext-D3 and PACT-D3 were cloned into pSF-CMV-Puro-NH2-FLAG and pSF-CMV-Puro-NH2-Cmyc vectors as described in Chapter 5. This results in the fusion of an N-terminal epitope tag (FLAG or C-myc respectively) to each construct.

To optimise transfection conditions, a reporter plasmid containing the red-fluorescent protein mKate was transfected into HEK293 cells, varying the cell density, the volume of transfection reagent, and the amount of DNA used (Figure 10.4) (Shcherbo *et al.*, 2007). Lower cell densities were found to result in higher overall expression of mKate, while the volume of transfection reagent made no difference in the range studied. The transfection was slightly more effective with the lower concentration of DNA. These values agree well with the ‘typical’ values provided by the manufacturer, so were used in the subsequent experiment.

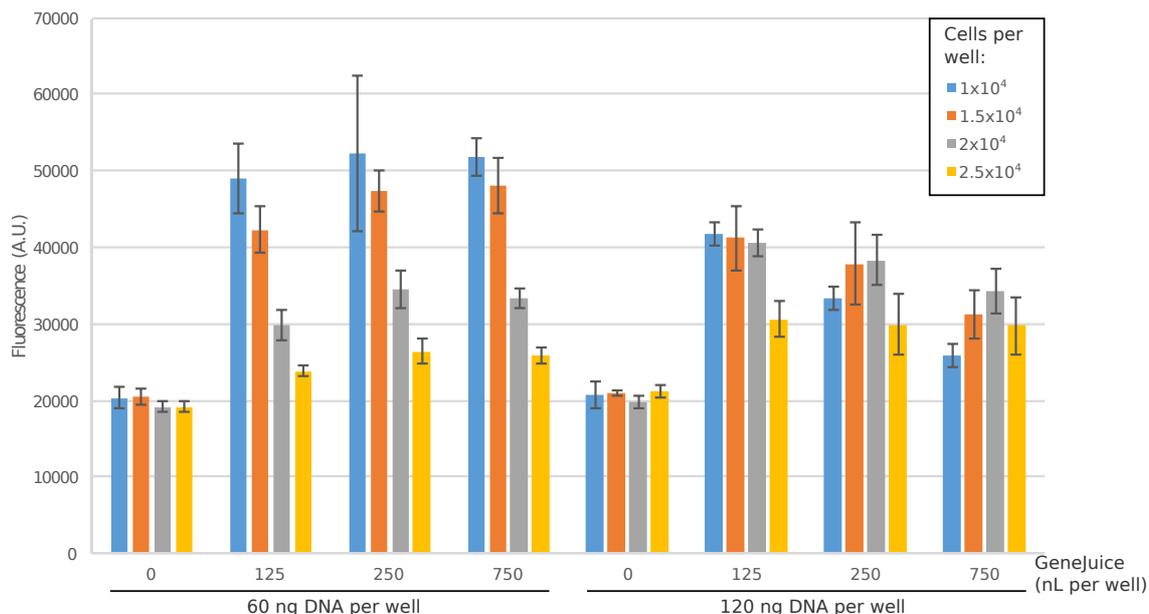


Figure 10.4. Optimisation of transfection in HEK293 cells

Results of transfection optimization in HEK293 cells, using GeneJuice transfection reagent (Novagen) and a plasmid coding for the fluorescent protein mKate. The experiment was carried out in a 96 well plate format, with three replicates for each condition. Error bars show the standard deviation.

Each of the PACT constructs, together with corresponding empty vectors, was transfected into HeLa and HEK293 cells. Cells with integrated plasmids were then

selected using media supplemented with puromycin. No HeLa cells survived the selection (perhaps because the transfection was ineffective), but some cells survived for all HEK293 transfections. Lysates from these new mixed cell lines were then examined by Western blot (Figure 10.5).

In all cases, a high degree of background signal was seen, showing that further optimisation of the antibody concentrations and blocking procedures is required. However, some specific staining is visible. In Figure 10.5A, a band can be seen in the FLAG-PACT lane, just above 35 kDa. This is the expected weight for full-length PACT with the FLAG epitope tag, showing that the overexpression has been successful. This is confirmed in Figure 10.5C, as a second, slightly larger band appears above endogenous PACT in the FLAG-PACT lane. Among the myc-tagged constructs, only myc-PACT-Ext-D3 appears to be expressed (Figure 10.5B). It is unfortunately not possible to confirm this in the same way as for FLAG-PACT, because the α -PACT antibody recognises a region in domain 2 that is not included in this construct.

Four of the six constructs did not appear to be expressed. To test whether this was due to lack of plasmid integration, or a problem with expression from an integrated plasmid, each line was grown in media with puromycin concentrations up to 3 $\mu\text{g}/\mu\text{L}$ (Figure 10.5D). If cells survive and grow, they alter the pH of the media, causing a colour change from pink to yellow. In this way, it can be seen that all cell lines were resistant to puromycin, while HeLa and HEK293 cells which were not transfected were not resistant. This shows that the plasmid integrated successfully, but does not always express the desired construct. This may be due to integration into sites where expression is disfavoured (for example, in regions of heterochromatin), but could also be due to disruption of the PACT sequence when the plasmid is linearised during integration.

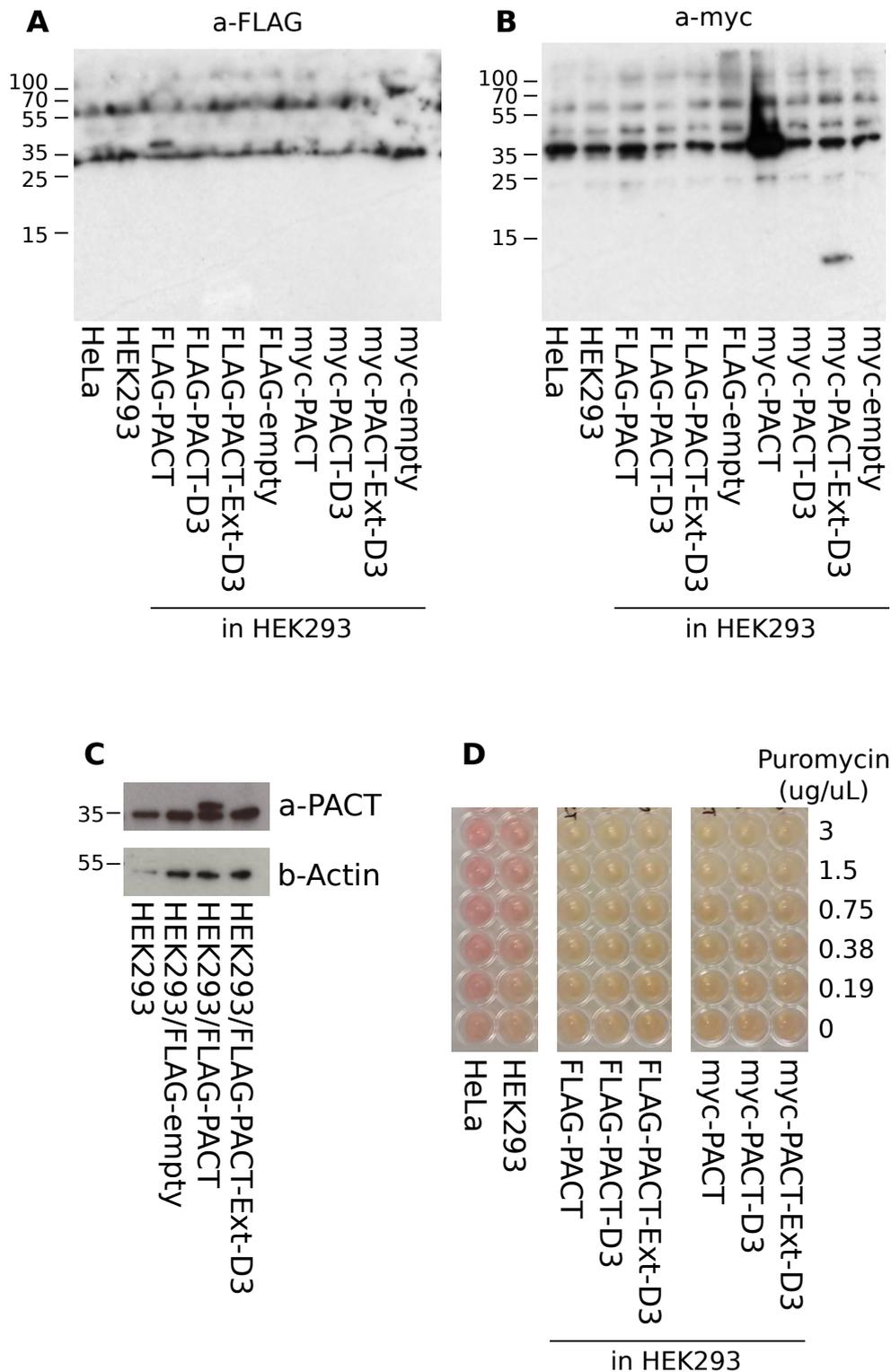


Figure 10.5. Stable overexpression of FLAG- and myc-tagged PACT constructs in HEK293 cells.

Cells were transfected with PACT constructs, and cells with stably integrated plasmids were selected for using puromycin as described in Section 5.4.2. Cell lysates were then examined by Western blotting, with primary antibodies against (A) FLAG or (B) myc. C) For selected cell lines, Western blots were stained with primary antibodies against PACT. D) Puromycin resistance tests of all cell lines.

10.4 Future directions

The preliminary results presented above show that it is possible to both knockdown and overexpress PACT, at least in HEK293 cells. There are several clear ways in which this work could be extended to allow insight into PACT function within cells.

Firstly, transfection of the PACT overexpression vectors into HEK293 cells should be repeated to generate the remaining overexpression lines. The fact that some constructs expressed successfully shows that this expression strategy can work, but that the probability of successful integration and expression is less than 100%. It would also be beneficial to generate monoclonal lines from the mixed populations, so that the site of integration can be checked to ensure important cellular functions have not been disrupted. In addition, the effectiveness of transient transfection of PACT constructs into Hff cells could be tested, to examine the effects in a more physiologically relevant system. As mentioned previously, optimisation of the Western blot blocking and staining protocol is required in order to reduce the non-specific signal.

Once the cell lines have been generated, homodimerisation of PACT within cells can be tested by transfecting a second PACT construct into the cell lines, with a different epitope tag. Having two separate tags allows one to be used for immunoprecipitation, and the other to be used for detection, and would also permit the use of proximity ligation assays to detect homodimers in cells. For the full-length PACT constructs, immunoprecipitations would need to be performed with and without the presence of an RNase that can degrade dsRNA (such as RNase III or RNase V1), to distinguish direct dimerisation via domain 3 from indirect dimerisation via dsRNA. Alternatively, mutations could be introduced into domains 1 and 2 to abolish dsRNA-binding. The necessity of PACT domain 3 for dimerisation could be tested by introducing the L273R mutation, or by truncating PACT after domain 2.

Further work is required to optimise siRNA-mediated knockdown of PACT in Hff cells to ensure that expression is reduced at the protein level. Once this is achieved, the levels of individual miRNAs, and of Dicer and TRBP, can be examined by qRT-PCR and Western blotting in both PACT knockdown and overexpression

conditions. The same system could be used to probe the effects of PACT on PKR activation in response to cellular stresses.

The availability of CRISPR tools offers the possibility of creating cell lines where PACT expression is completely abolished. This provides an alternative strategy if problems are found with the optimisation of siRNA-mediated knockdown of PACT, as might be expected if PACT plays a significant role in the RNA interference pathway. CRISPR also offers two approaches for studying mutant forms in the absence of wild type PACT: either PACT mutants can be overexpressed in PACT^{-/-} cells, or the endogenous copy of PACT can be edited directly.

Conclusions and Discussion

11.1 Summary of results from this thesis

Prior to this study, the homodimerisation of PACT-D3 and TRBP-D3 had been established using yeast-2-hybrid assays (Laraki *et al.*, 2008), but the structural basis of this dimerisation was unknown. Indeed, no structures of type B dsRBDs were publicly available at that time, and it was unclear to what degree their structure diverged from that of canonical dsRBDs. Since then, the field has developed considerably, with recent structures of Loqs-D3 homodimers and of the TRBP/Dicer complex, together with the application of genome editing techniques to study PACT and TRBP function (Kim *et al.*, 2014; Wilson *et al.*, 2015; Jakob *et al.*, 2016). The main contributions of this thesis are summarised below.

11.1.1 PACT-D3 forms homodimers

Recombinant PACT-D3 was shown to dimerise *in vitro* using both SEC-MALLS (Figure 7.1A) and sedimentation equilibrium AUC (Figure 7.2). The conserved N-terminal region was found to be unnecessary for dimerisation (Figure 7.1B), in marked contrast to Staufen-D5 (Gleghorn *et al.*, 2013). Phosphorylation (modelled by phospho-mimic mutations) was found to be unnecessary for dimerisation at concentrations above approximately 100 μM (Figure 7.3), but might still play a role at lower concentrations. Dimerisation of PACT-D3 can be disrupted by mutation of leucine 273 to arginine, without disrupting the structure of the domain (Figure 7.4). NMR data collected at 5 μM appears to show the presence of monomeric PACT-D3, implying the dimer dissociation constant on the order of 1 μM (Figure 8.15).

11.1.2 PACT-D3 homodimers are asymmetric, and similar in structure to Loqs-D3 dimers

NMR spectroscopy was used to show that PACT-D3 adopts two states in solution (Figure 8.5); that these states undergo chemical exchange (Figure 8.11); and that they represent the two protomers of an asymmetric dimer (Figure 8.12). A number of intermolecular NOEs were identified (Figures 8.12, 8.13 and 8.14), all of which are consistent with PACT-D3 homodimers adopting the same structure as Loqs-D3 homodimers (Jakob *et al.*, 2016). In particular, the NOE data supports the

β -3 strands of the two subunits lying parallel to one another, but with a register shift to allow for intermolecular hydrogen bond formation (Figure 8.14).

11.1.3 TRBP domain 3 homodimerises with lower affinity than PACT domain

3

TRBP-D3 was not stable *in vitro*, but a longer construct containing two phosphomimic mutations (TRBP-Ext_{DD}-D3) could be produced (Figures 6.2 and 6.4). Despite extensive sequence conservation with PACT and Loqs at the dimer interface (Figure 9.3A), TRBP-Ext_{DD}-D3 was found to only weakly dimerise in the concentration range studied (Figure 9.2). The weak dimerisation is primarily due to R353 and R354: mutating these to match the equivalent residues in PACT restored asymmetric homodimerisation of TRBP-Ext_{DD}-D3 (Figures 9.3 and 9.4). By contrast, dimerisation was completely abolished by the mutation L326R, further confirming that TRBP domain 3 homodimerises, albeit weakly, using the same interface as PACT and Loqs (Figure 9.3).

11.2 Discussion

11.2.1 Incompatibility of PACT homodimerisation and Dicer/PACT complex formation

The crystal structures of TRBP-D3 in complex with the Dicer helicase insert and of Loqs-D3 homodimers suggested that Dicer binding and dimerisation use the same interface (Wilson *et al.*, 2015; Jakob *et al.*, 2016). PACT-D3 homodimerisation via the same interface is supported both by mutagenesis (Figure 7.4) and NMR data (Figures 8.12, 8.13 and 8.16). Given the strong conservation with TRBP at the Dicer-binding surface, it is reasonable to conclude that PACT also binds Dicer using the same interface, and therefore that PACT homodimerisation is incompatible with formation of PACT/Dicer complexes (Figure 11.1).

It has previously been suggested that the specialisation required to form a specific protein-protein interface means that interactions with different proteins will utilise separate protein surfaces (Kundrotas and Vakser, 2013). This is clearly not the case for PACT-D3, whose Dicer binding and homodimerisation interfaces are almost completely coincident despite the lack of any structural homology between the Dicer helicase insert and PACT-D3. Close examination of the Loqs-D3 homodimer and TRBP/Dicer complex reveals that these two interactions are

possible partly because conserved TRBP/Loqs residues mediate similar interactions in the two complexes, and partly through rearrangement of side chains at the interface.

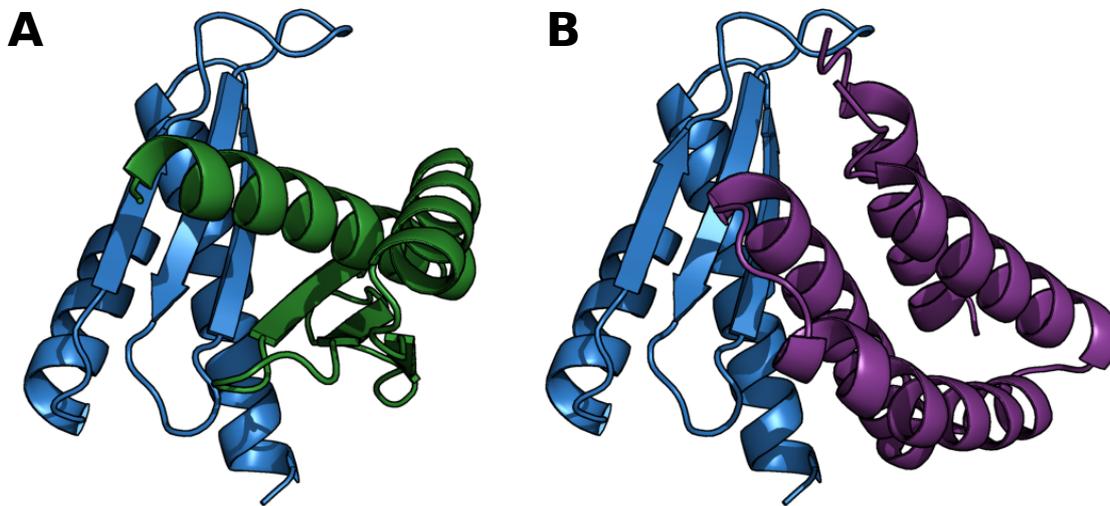


Figure 11.1. Overlap of the PACT-D3 homodimerisation and Dicer-binding interfaces

A) Model of a PACT-D3 homodimer created by aligning the I-TASSER model of PACT-D3 to each protomer of the Loqs-D3 dimer (4X8W). B) Model of the Dicer/PACT-D3 complex created by aligning the I-TASSER model of PACT-D3 to the Dicer/TRBP-D3 structure (4WYQ). PACT-D3 is in blue, Dicer is in purple.

To identify the Dicer residues which contribute most to the interaction with TRBP, it is helpful to examine sequence conservation at the interface, especially between the distantly related mammalian and insect homologues (Figure 11.2A). There is significant variability at many positions, especially in helix α -4, in contrast to the much higher conservation between PACT, TRBP and Loqs. The best conserved interface residues are F280 and D283 in helix α -1, and T350 and L358 in helix α -2.

In the TRBP/Dicer complex, Dicer T350 forms a hydrogen bond to the backbone amide of TRBP V336 (Figure 11.2B). The methyl group of T350 is also positioned so that it can interact weakly with the π electrons of TRBP Y358. In the Loqs-D3 homodimer, similar interactions are formed, but to the other Loqs-D3 protomer: the backbone amide of V436_B hydrogen bonds to the backbone oxygen of V434_A, while the ring of Y458_B interacts with the methyl of V436_A (Figure 11.2C). All these residues are highly conserved between all the vertebrate and insect PACT homologues examined in Figure 1.9. There is direct evidence that these

interactions also occur in PACT-D3 homodimers: NOESY data supports a similar β -sheet arrangement around V283_B (Figure 8.14), and the upfield shift of V283_A H γ 1 (-0.17 ppm; see Figure 8.12) is consistent with ring current effects due to a location above the plane of an aromatic ring (Haigh and Mallion, 1979).

Figure 11.2D shows that Dicer F280 inserts into a shallow pocket formed by TRBP I314, Q324, L326 and H338. All five of these residues are extremely well conserved among vertebrates and arthropods, suggesting that this is a key part of the interface (with the only common variant being replacement of the histidine with tyrosine). In both subunits of the Loqs-D3 homodimer (Figure 11.2E), the conserved residues are rearranged, with the H438 side chain rotating to fill the pocket. This also prevents the histidine sidechain from clashing with the β -3 strand of the other Loqs-D3 subunit.

The incompatibility of PACT homodimerisation and PACT/Dicer interaction implies that these two complexes compete inside the cell. The extent to which this is significant depends both on the relative affinities of the interactions, and the relative concentrations of PACT, TRBP and Dicer. The affinity of TRBP for Dicer is reportedly in the low nanomolar range (Wilson *et al.*, 2015), while NMR of low concentration PACT-D3 suggests that the homodimerisation affinity of PACT is on the order of 1 μ M (Chapter 8 section 8.5). Unfortunately, information on the relative abundance of PACT, TRBP and Dicer is not currently available. If the affinity of PACT for Dicer is similar to that of TRBP, then homodimers will only form if PACT is in excess of Dicer. However, dimerisation could only have been selected for if functional dimers are present. Ultimately, this question is unlikely to be resolved except through affinity measurements of PACT homodimerisation and PACT/Dicer complex formation, and in-cell tests such as Proximity Ligation Assays to detect which complexes are present.

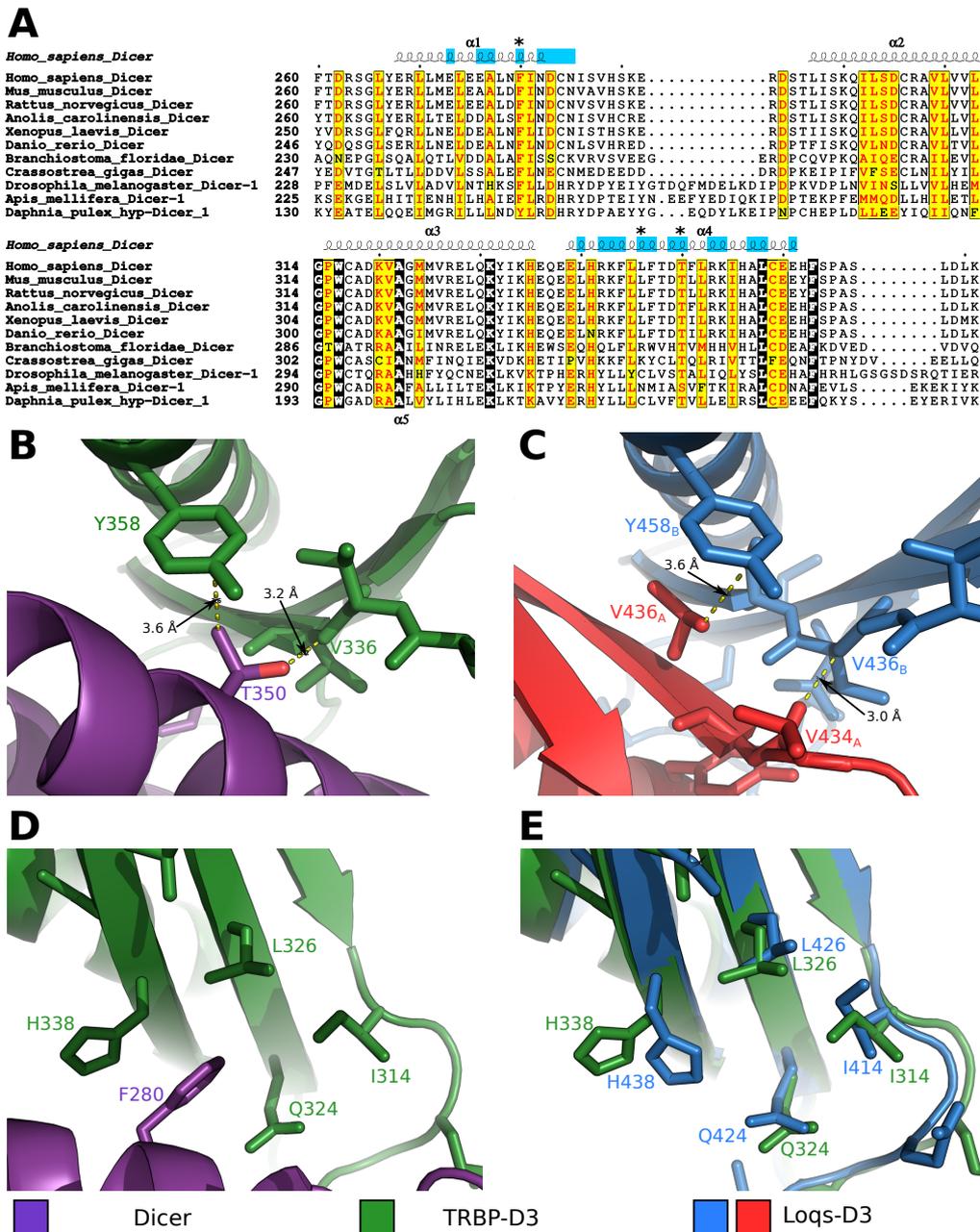


Figure 11.2. Comparison of the Dicer/TRBP and Loqs homodimer interfaces

A) Alignment of Dicer-1 helicase insert sequences from a range of animal species. The secondary structure of human Dicer insert (4WYQ) is shown above; cyan highlights indicate residues that interact with TRBP; asterisks indicate residues which can be mutated to disrupt the interaction with TRBP and PACT (Wilson et al., 2015). B) Interactions of Dicer T350. Interactions described in the text are marked in yellow. C) Similar interactions are formed by Loqs V434_A and V436_A. Protomer A here corresponds to chain A in PDB accession 4X8W, while protomer B corresponds to chains D/F. D) Interactions of Dicer F280 with a surface pocket of TRBP. E) Overlay of the surface pocket in TRBP and Loqs, showing rotation of the histidine side chain. (Colours are consistent throughout.)

11.2.2 Homo- and heterodimerisation of dsRBDs

11.2.2.1 Potential functions of PACT homodimerisation

Because both PACT-D3 and Loqs-D3 homodimerise, and the sequence of other distant homologues appears compatible with dimerisation, it is likely that the ancestral version of PACT-D3 could dimerise. This conservation suggests that dimerisation is functional. There are several possible ways in which PACT dimerisation might be beneficial to the cell, but current evidence does not conclusively prove which, if any, are driving conservation of this property.

Dimerisation via domain 3 may alter the RNA-binding behaviour of full-length PACT by bringing 4 type A dsRBDs into a single complex. This would be expected to lead to increased avidity for dsRNA (Shamoo, Abdul-Manan and Williams, 1995), and indeed full-length PACT has up to 10-fold higher apparent affinity for siRNA than a PACT construct which lacks the third domain (Takahashi *et al.*, 2013). However, the same study found that full-length TRBP had a higher affinity still, despite its weaker dimerisation, demonstrating that differences in other regions of the protein can exert a greater effect. 4 dsRBDs is also the maximum number that can bind to miRNA-length substrates, suggesting a PACT dimer bound to miRNA would cover most of the RNA surface (Benoit *et al.*, 2013; Acevedo *et al.*, 2015, 2016). However, it is unclear what the effects of these properties would be *in vivo*. It is also unknown whether PACT dimers assemble on (pre-)miRNAs, or are pre-formed prior to RNA binding.

PACT dimerisation is reportedly required for PKR activation in response to cellular stress. While this may be true in mammals (subject to the caveats given in Chapter 8, section 8.7.2 regarding phosphorylation), PACT homologues can dimerise even in organisms that lack an interferon pathway. For example, *Drosophila* Loqs-D3 has been shown to dimerise using SEC-MALLS (Figure 7.5), and other insects and molluscs have highly conserved dimer interfaces (Figure 1.9). Therefore, dimerisation of the ancestral PACT/Loqs protein must have served a different function, presumably related to miRNA or siRNA biogenesis.

PACT dimerisation may function to increase protein stability. dsRBDs are fairly small domains, and the Dicer-binding interface takes 20% of their surface (calculated from 4WYQ, excluding the poorly resolved N-terminal helix). There is a trade-off between being optimised for protein-interaction and optimisation for

solvent exposure. Dimerisation may provide a way to 'hide' the exposed interaction surface from the solvent, until a Dicer molecule can be found. This is consistent with trends in *in vitro* stability of monomeric and dimeric PACT and TRBP constructs observed in this thesis. However, further work is clearly needed to quantify the difference in stability, and determine whether it is significant *in vivo*.

11.2.2.2 Lack of homodimerisation in TRBP

TRBP-D3 homodimerises with lower affinity than PACT-D3, largely because of two arginines at positions 353 and 354 (Figure 9.3), which replace an [H/Q]N motif that is conserved in almost all PACT and Loqs sequences. They are relatively recent additions to TRBP: they appear in human, rat and bovine TRBP, but not in *Xenopus laevis* or *Danio rerio* (Figure 1.9). The evidence presented here clearly shows that R353 and R354 are the cause of weaker TRBP homodimerisation, but does not directly address the reason these substitutions have evolved.

As discussed above and in Chapter 9, weakening homodimerisation seems to decrease the stability of the domain, but it is not clear why this would be desirable. An alternative possibility is that weaker dimerisation of TRBP-D3 might liberate it for interactions with other binding partners, notably Dicer. The Dicer binding interface identified by Wilson and coworkers is incompatible with the homodimerisation interface (Wilson *et al.*, 2015), so these two interactions will compete – reducing the affinity of one complex will increase formation of the other. In further support of the hypothesis that the function of these substitutions favours Dicer binding, the crystal structure of the complex shows that TRBP R354 lies at the interface with Dicer, forming a salt bridge to Dicer E278.

11.2.2.3 Possibility of PACT/TRBP heterodimerisation

The structure of Loqs-D3 shows that R353 and R354 are buried in one half of the dimer but exposed on the other (Figure 9.3). Therefore, dimers may be able to tolerate substitution of bulkier amino acids in one half of the dimer, but clashes would occur if both halves of the dimer had such substitutions. This is consistent with Yeast-2-hybrid experiments showing that TRBP domain 3 can heterodimerise with PACT domain 3, but not form homodimers (Laraki *et al.*, 2008). PACT/TRBP heterodimers would presumably have a different dsRNA

binding mode to either PACT homodimers or TRBP monomers, and have been suggested to prevent activation of PKR (Singh *et al.*, 2011).

However, heterodimerisation remains to be validated *in vitro* using purified proteins. Several approaches have been tried in our lab, including pull-down assays, analytical size exclusion and chemical cross-linking. However, the results of these studies are inconclusive (data not shown). Part of the problem may be the strength of PACT-D3 homodimerisation, which will always compete with PACT-D3/TRBP-D3 heterodimerisation. A promising approach may be to use microscale thermophoresis (MST), as this technique is sufficiently sensitive to work at nanomolar concentrations of PACT-D3, where the domain would be expected to be monomeric.

Only one of the two L426 sidechains is buried in the Loqs-D3 homodimer, while the other remains partially solvent exposed. The structure suggests that a heterodimer may be able to form between Loqs-D3 L426R, and Loqs-D3 QN453,454RR. If this is true, it implies that PACT-D3 L273R may retain the ability to heterodimerise with TRBP-D3. This would provide an alternative method to test heterodimerisation between PACT and TRBP, while removing the competing effects of PACT homodimerisation.

The possibility of post-translational modification in PACT adds an additional layer. Based on the Loqs-D3 crystal structure, TRBP R353 and R354 would be positioned near PACT S287D in a heterodimer. If S287 is phosphorylated, there would likely be a steric clash between the arginines and the phosphate group, which could plausibly prevent dimerisation as previously reported (Singh *et al.*, 2011). It would be possible to test this idea using an assay for heterodimerisation (such as MST), and replacement of S287 with bulky amino acids.

11.2.3 Asymmetric dimerisation of PACT

11.2.3.1 Asymmetry in Loqs-D3 homodimers

In Chapter 8, NMR data was used to show that PACT-D3 homodimers are asymmetric, and adopt a structure similar to that of Loqs-D3 homodimers. It was also shown that Loqs-D3 forms asymmetric homodimers in solution, and that asymmetric dimerisation is not merely an artefact of crystal packing. With this established, we can examine the Loqs-D3 structure (4X8W) for clues as to how

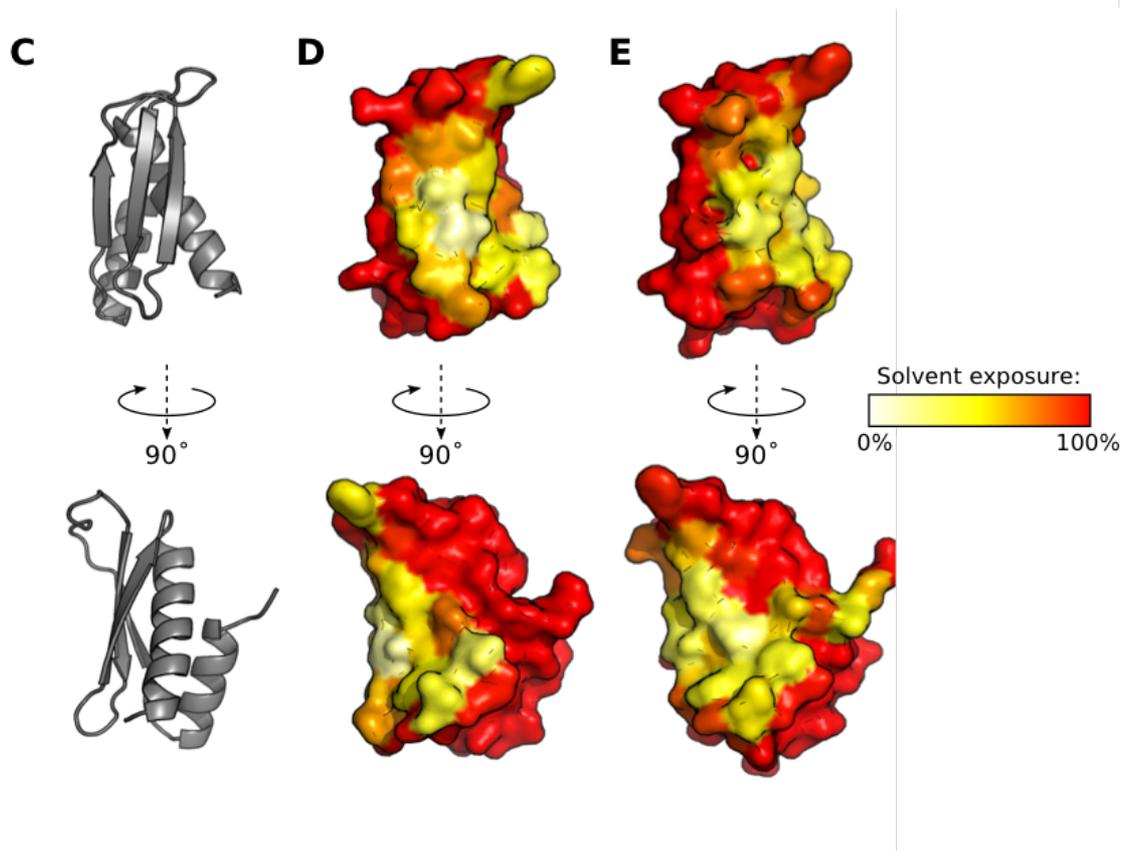
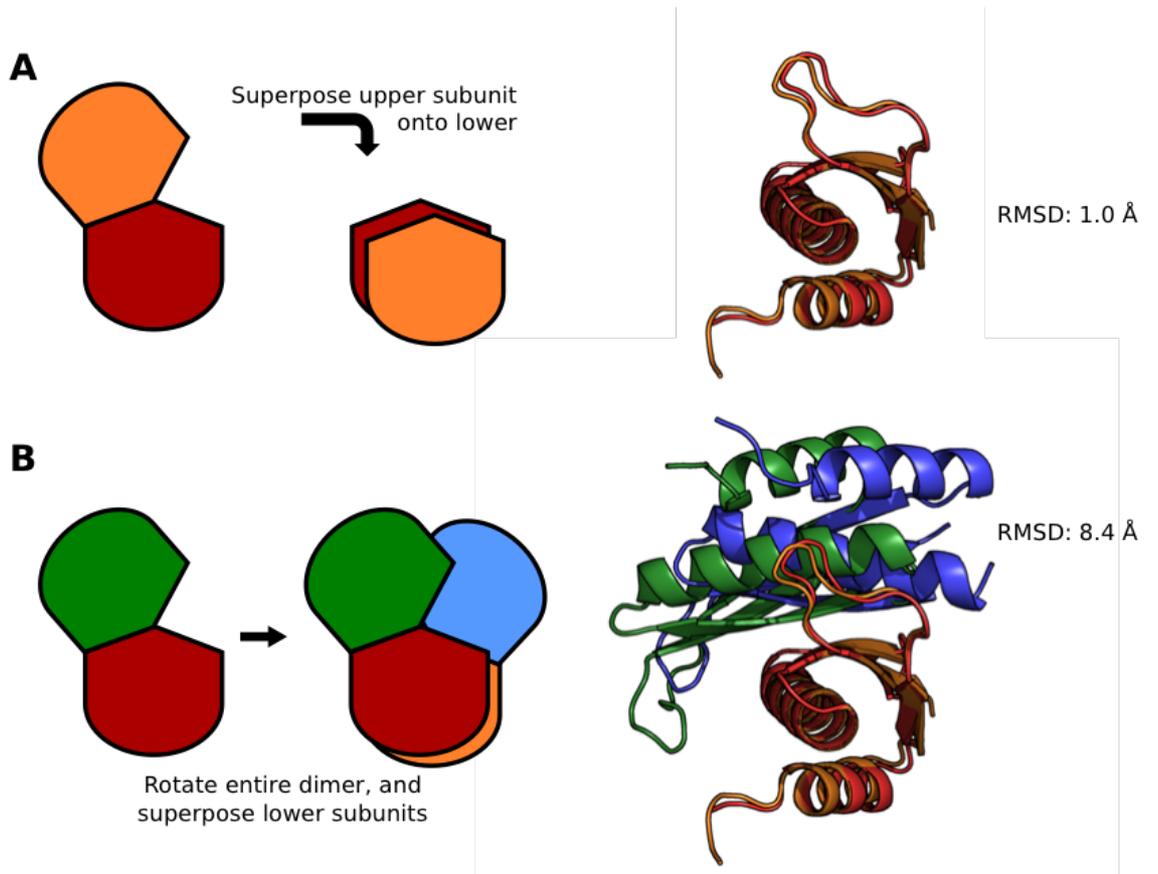
and why asymmetric dimerisation occurs (Jakob *et al.*, 2016). The unit cell for this crystal contained 6 well-resolved molecules, some of which exhibited domain swapping of strand β -3 and helix α -2. The discussion below is based on the dimer formed by chain A (protomer A), and β -3 and α -2 of chain D together with α -1, β -1 and β -2 of chain F (protomer B). Similar results hold for the other two dimers in the unit cell.

A protein complex is symmetric if the positions of its subunits are related by a symmetry operator. Since reflection is not allowed, the possible operators are translation and rotation. Translational symmetry occurs in protein crystals and fibres: since translation can be repeatedly applied, complexes with translational symmetry operators are not naturally bounded, and can form infinite arrays or fibrils. In contrast, repeatedly applying a rotational symmetry operator to a protein will eventually return it to its original position, resulting in a complex of a defined size. Subunits of a protein complex can be related by multiple symmetry operators: in these cases, all axes of rotation must intersect at a common point, and the complex is said to exhibit point group symmetry. All symmetric homodimers fall into the C2 symmetry group, meaning that the two subunits are related by a single 180° rotation.

Given this definition of symmetry, two types of asymmetric dimer can be distinguished. Firstly, a dimer could be asymmetric because the two protomers have different structures, meaning there is no possible rigid transformation that relates them. Alternatively, the two protomers could have the same structure, but their positions may not be related by a rotational symmetry operator. In this case, superposition of the two subunits requires either multiple rotations along non-intersecting axes, or a combination of rotation and translation.

Figure 11.3. Asymmetry in Loqs-D3 homodimers

A) Overlay of the two dimer subunits reveals only small changes in conformation. B) Rotating the dimer such that opposite protomers are superposed demonstrates the asymmetry. The C α RMSD is calculated only between the blue and green protomers. C-E) Location of the dimer interface in Loqs-D3 protomer A (D) and protomer B (E). The surface is coloured according to solvent exposure in the dimer (calculated using the POPSCOMP server (Kleijnung and Fraternali, 2005)).



Loqs-D3 can be seen to exhibit the latter form of asymmetry. As shown in Figure 11.3A, the two halves of the homodimer have very similar structures, with a C α root mean square displacement (RMSD) of 1.0 Å. However, there does not exist a 180° rotation that maps one subunit onto the other. If instead the homodimer is partially aligned with itself so that protomer A of one dimer superposes onto protomer B of the other dimer (Figure 11.3B), the nonaligned subunits are found to have a C α RMSD of 8.4 Å (for a symmetric dimer, this procedure would result in perfect superposition).

In an asymmetric dimer, the two subunits use different interfaces to bind their partner. Figures 11.3D,E show that these interfaces overlap significantly. Examining the interface closely, the main difference between the two binding modes is a register shift between the parallel β -3 strands (Figure 8.14). As discussed in Chapter 8, section 8.4.8, identical β -strands require a register shift to interact in parallel, because if they were aligned symmetrically, the amide and carbonyl groups would not be correctly positioned for hydrogen bond formation. Therefore, the presence of an intermolecular parallel β -sheet between equivalent β -strands in the two halves of a dimer always causes asymmetry.

A quantitative measure of asymmetry was proposed by Andre and coworkers (Andre *et al.*, 2008):

$$S_{dev} = \frac{\sum_i \sum_j |A_i B_j - B_i A_j|}{N \times N}$$

where $A_i B_j$ is the distance between the C α atoms of residue i in protomer A and residue j in protomer B, and N is the protein length. For a symmetric dimer, S_{dev} equals 0, as $A_i B_j = B_i A_j$ for all i and j . There is no maximum possible value for S_{dev} , nor intrinsic thresholds to distinguish degrees of asymmetry, but Swapna *et al.* suggest that S_{dev} values greater than 3 represent clear asymmetry, while S_{dev} values between 1 and 3 represent limited asymmetry (Swapna, Srikeerthana and Srinivasan, 2012). For Loqs-D3, S_{dev} is equal to 2.3, consistent with the fact that the structures of the individual protomers are similar, and that the two possible dimer interfaces overlap.

11.2.3.2 Evidence for similar asymmetry in PACT-D3 homodimers

Do PACT-D3 homodimers also exhibit this form of asymmetry, and have the same structure as Loqs-D3 dimers? Several pieces of NMR data suggest that they do. Firstly, secondary structure predictions from the TALOS-N program (based on comparison of backbone chemical shifts to a database of matched shifts and structures) match well to Loqs-D3, and suggest that there are no significant differences in secondary structure between the two halves of the PACT-D3 homodimer (Figure 8.5). Comparison of the predicted dihedral angles of PACT-D3 also reveals no significant differences between the two states. There are some small differences in predicted dihedral angles for residues in the β -sheet, but these could be due to the larger prediction error in this region. Alternatively, the differences might be consistent with a slight bending of the β -sheet, as suggested by the proton pressure-induced chemical shift changes Chapter 8, section 8.6. Secondly, both NOESY and high pressure experiments suggest that PACT-D3 uses the same dimer interface as Loqs-D3, including strands β -2 and β -3 and the C-terminal half of helix α -2 (Figures 8.12 and 8.17). Compound chemical shift differences between the two states are largest on the third β -strand and second α -helix, showing that the residues in these regions experience the most pronounced difference in chemical environment (Figure 8.9). NOESY data shows that the β -3 strands of the two subunits form a parallel β -sheet, and that there is a register shift between the two strands (Figure 8.14). Taken together, this confirms that all the main features of asymmetric Loqs-D3 dimerisation also occur in PACT-D3 homodimers.

11.2.3.3 General arguments for the prevalence of symmetry

Asymmetric homodimers are relatively rare, with a recent study finding that only 5-10% of homodimers in the PDB show significant asymmetry (Swapna, Srikeerthana and Srinivasan, 2012). To explain why PACT-D3 homodimers might be asymmetric, it is necessary to understand why symmetry is so prevalent in protein complexes.

One argument is that symmetry ensures that the protein will assemble into a finite complex. In general, because asymmetric dimerisation uses different interfaces on each protomer, a single molecule could bind to two other molecules

simultaneously, and each of those could bind two molecules, leading to formation of long fibrils. Fibril formation is useful for forming large scale structures such as microtubules and actin filaments, but would not be useful for most proteins. In practice however, fibril formation of this kind requires that the two possible interfaces are well separated. In PACT-D3 and Loqs-D3, the two interfaces overlap, and so steric clashes prevent fibril formation.

A second argument is that strong symmetric dimers can evolve more easily than strong asymmetric dimers. A symmetric dimer interface contains two copies of each possible interaction, so mutations at the interface cause twice the change in interaction free energy (Monod, Wyman and Changeux, 1965). This is partially offset by the smaller number of residues involved in a symmetric interface, reducing the mutation rate compared to an asymmetric interface of equivalent area, resulting overall in a $\sqrt{2}$ -fold greater variation in interaction free energy (Andre *et al.*, 2008). The greater variation in interaction free energy leads to faster emergence of stable complexes, which can then be selected for. The same result can be arrived at by considering the energies of randomly chosen interfaces (Andre *et al.*, 2008).

Several computational studies into simplified models of dimer evolution have been published, with results consistent with the argument above. Schulz found that in the case of a defined interface with fixed orientation, symmetric dimers have on average approximately equal interaction energies to asymmetric dimers, but a wider distribution of energies. This led to the formation of strong symmetric dimer interfaces with significantly fewer mutations than for asymmetric dimers (Schulz, 2010). Lukatsky and coworkers studied a dimer model which allowed rotation of the two subunits, finding that randomly generated symmetric dimer interfaces had both larger variance and lower average energy than asymmetric interfaces (Lukatsky *et al.*, 2007). The lower energy can be explained by the freedom of the subunits to rotate and sample a range of possible interfaces: the greater variance of symmetric interface energy means the best orientation will on average have a lower energy than for an asymmetric interface. Andre and coworkers performed random docking of protein structures, and found that symmetric complexes had a greater variance in interaction energy, and were heavily over-represented among low energy dimers (Andre *et al.*, 2008). Taken

together, these studies provide compelling theoretical reasons for the predominance of symmetric dimers.

11.2.3.4 Cause of asymmetry in PACT-D3 and Loqs-D3

Several explanations have been put forward to explain symmetry breaking within homodimers. Local asymmetry can allow optimisation of sidechain packing, especially near a symmetry axis (Goodsell and Olson, 2000). Slightly larger scale asymmetry can occur due to register slips between antiparallel intermolecular β -strands or coiled coils (Brown, 2006). Reported cases of gross asymmetry are usually caused by symmetry mismatch between the dimer and a single ligand, which can be either a small molecule, another protein or an asymmetric nucleic acid sequence (Swapna, Srikeerthana and Srinivasan, 2012). As stated above, Loqs-D3 and PACT-D3 break symmetry in a different and (to our knowledge) novel manner, via interaction of offset parallel β -strands.

However, the functional significance of asymmetry of PACT-D3 and Loqs-D3 remains unclear. Asymmetry has been reported to play a role in transporter function (Morrison *et al.*, 2011), to introduce bends into large fibril forming proteins (Brown *et al.*, 2001; Madrazo *et al.*, 2001), to occur in enzymes with half-of-sites reactivity (Renatus *et al.*, 2001) and to allow a dimer to interact with a monomeric partner. However, none of these apply to PACT-D3 and Loqs-D3 homodimers. Although PACT and Loqs do bind to asymmetric RNA molecules, binding is mediated through the first two domains. Since these are believed to be flexibly linked to the third domain, the asymmetry of the RNA should not impose asymmetry on the dimerisation domain.

I believe two hypotheses are consistent with the data. Firstly, the interaction between the parallel β -strands contributes at least three hydrogen bonds to the dimer interface. The stabilising effect of these bonds may provide a sufficient advantage to overcome the disadvantages associated with asymmetry. It is perhaps notable that the simulation studies cited above only include side chain interactions, and do not account for the possibility of intermolecular contacts with the protein backbone (Lukatsky *et al.*, 2007; Schulz, 2010).

Alternatively, the evolutionary advantages of symmetry may be partially negated by selection for Dicer binding. This places a competing evolutionary pressure on the interface, preventing it from freely exploring some mutations that might lead

to symmetric dimers. It may also disfavour formation of a strong homodimer, as it is necessary for the dimer to dissociate for PACT and Loqs to bind Dicer.

The conservation of asymmetry in PACT and Loqs suggests that asymmetry was present before the divergence of PACT and TRBP. Therefore, asymmetry cannot have evolved to allow selective blocking of TRBP homodimerisation while still allowing heterodimer formation. In any case, this can be achieved without asymmetry: even pseudosymmetric heterodimers can minimise homodimerisation by adding complementary charged residues to each heterodimer subunit (Lukatsky *et al.*, 2007)

11.2.4 Suggestions for future research

The new structural data available on PACT-D3, TRBP-D3 and Loqs-D3 provide a strong base for further investigations of the protein-protein interactions of these domains. The peak doubling and chemical exchange in PACT-D3 poses challenges for directly determining the dimer structure using traditional NMR structural methods. However, the availability of monomeric PACT-D3 L273R will allow the monomer structure to be solved using standard NOESY methods, and restraints from wild-type PACT-D3 used to reconstruct the dimer by computational docking.

The conserved N-terminal region of PACT-D3 has been implicated in improper PKR activation, but little is known about its structure. Further NMR studies of PACT-Ext-D3 L273R are likely the only way to investigate this, as there appears to be a degree of flexibility in this region (Wilson *et al.*, 2015). Backbone assignment would allow the secondary structure to be defined using TALOS-N, and the effects of the P222L mutation studied. It may also be possible to determine the structure of the entire domain with extension, provided the effects of intermediate chemical exchange are not too severe (Figure 8.3).

The line broadening in TRBP-Ext_{DD}-D3 means that little additional progress can be made by NMR, unless conditions can be found which favour a single oligomeric state, or significantly speed or slow the exchange rate. The monomeric mutant form produces the cleanest spectra, but the low yield may be a challenge: if conditions can be found which confer greater stability, backbone assignment may be possible, especially with a sensitive probe. The dimeric mutant is stable, but its spectra are complicated by peak doubling, and attempts to study it by NMR are likely to run into the same problems as for PACT-Ext-D3. If TRBP-D3

RR353,354HN is found to be stable and dimeric, it may be a promising target for further NMR studies.

The dimerisation affinity for PACT-D3 is currently unknown, and has proved challenging to measure using traditional biophysical methods. As noted above, MST may provide a solution, as the fluorescently-labelled protein can be detected even at nanomolar concentrations, well below the expected homodimerisation affinity. MST would also allow interaction measurements with other partner proteins (such as Dicer, TRBP or even PKR) without the confounding factor of homodimerisation.

More broadly, the conflicting reports of PACT and TRBP function in the literature indicate the need for the functions of PACT and TRBP to be more clearly defined *in vivo*. The preliminary work presented in Chapter 10 provides a set of tools for investigating PACT-D3 function, allowing for overexpression of various PACT constructs. Mutations can be easily introduced into this system, facilitating testing of hypotheses based on structural data. The system could be further enhanced by the generation of PACT knock-out cell lines using CRISPR technology, so that endogenous PACT could be completely replaced.

The relatively minor effects of PACT and TRBP knockout in HeLa cells have cast doubt on the results of many earlier studies which used siRNA-mediated knockdown (Kim *et al.*, 2014). However, HeLa cells are in many ways non-physiological, and it is possible that other alterations in miRNA biogenesis in this cell line mask the effects of PACT and TRBP. Validation in a different cell type would greatly increase the confidence in this result.

The interaction mechanisms of PACT with PKR and RIG-I remain to be defined. Yeast-2-hybrid methods have proved successful in the past for locating interaction sites in these proteins, provided the domains do not have intrinsic dsRNA-binding activity. This means it may be preferable to work with versions of domains 1 and 2 with mutations in the RNA binding regions. The method of interaction with PKR is particularly important, as it is becoming clear that PACT and TRBP mediated inhibition of PKR is crucial for proper development.

Appendix 1: Sequences of PACT, TRBP and Loqs constructs

6x His with MBP tag and 3C cleavage site:

```

ATGGGCAGCA GCCATCATCA TCATCATCAC AGCAGCATGA AAATCGAAGA
AGGTAAACTG GTAATCTGGA TTAACGGCGA TAAAGGCTAT AACGGTCTCG
CTGAAGTCGG TAAGAAATTC GAGAAAGATA CCGGAATTAA AGTCACCGTT
GAGCATCCGG ATAAACTGGA AGAGAAATTC CCACAGGTTG CGGCAACTGG
CGATGGCCCT GACATTATCT TCTGGGCACA CGACCGCTTT GGTGGCTACG
CTCAATCTGG CCTGTTGGCT GAAATCACCC CGGACAAAGC GTTCCAGGAC
AAGCTGTATC CGTTTACCTG GGATGCCGTA CGTTACAACG GCAAGCTGAT
TGCTTACCCG ATCGCTGTTG AAGCGTTATC GCTGATTTAT AACAAAGATC
TGCTGCCGAA CCCGCCAAAA ACCTGGGAAG AGATCCCGGC GCTGGATAAA
GAACTGAAAG CGAAAGGTAA GAGCGCGCTG ATGTTCAACC TGCAAGAACC
GTACTTCACC TGGCCGCTGA TTGCTGCTGA CGGGGGTTAT GCGTTCAAGT
ATGAAAACGG CAAGTACGAC ATTAAAGACG TGGGCGTGGA TAACGCTGGC
GCGAAAGCGG GTCTGACCTT CCTGGTTGAC CTGATTAAAA ACAAACACAT
GAATGCAGAC ACCGATTACT CCATCGCAGA AGCTGCCTTT AATAAAGGCG
AAACAGCGAT GACCATCAAC GGCCCGTGGG CATGGTCCAA CATCGACACC
AGCAAAGTGA ATTATGGTGT AACGGTACTG CCGACCTTCA AGGGTCAACC
ATCCAAACCG TTCGTTGGCG TGCTGAGCGC AGGTATTAAC GCCGCCAGTC
CGAACAAAGA GCTGGCGAAA GAGTTCCTCG AAAACTATCT GCTGACTGAT
GAAGGTCTGG AAGCGGTAA TAAAGACAAA CCGCTGGGTG CCGTAGCGCT
GAAGTCTTAC GAGGAAGAGT TGGCGAAAGA TCCACGTATT GCCGCCACCA
TGAAAACGC CCAGAAAGGT GAAATCATGC CGAACATCCC GCAGATGTCC
GCTTTCTGGT ATGCCGTGCG TACTGCGGTG ATCAACGCCG CCAGCGGTGCG
TCAGACTGTC GATGAAGCCC TGAAAGACGC GCAGACTCGT ATCACCAAGG
GCCTGGAAGT TCTGTTCCAG GGACCAGCAA TG

```

```

MGSSHHHHHH SSMKIEEGKL VIWINGDKGY NGLAEVGKKF EKDTGIKVTV
EHPDKLEEFK PQVAATGDGP DIIFWAHDRF GGYAQSGLLA EITPDKAFQD
KLYPFTWDAV RYNGKLIAYP IAVEALSLIY NKDLLPNPPK TWEEIPALDK
ELKAKGKSAL MFNLQEPYFT WPLIAADGGY AFKYENKDYD IKDVGVDNAG
AKAGLTFLLVD LIKNKHMNAD TDYSIAEAAF NKGETAMTIN GPWAWSNIDT
SKVNYGVTVL PTFKGQPSKP FVGVLGAGIN AASPNKELAK EFLNYLLTD
EGLEAVNKDK PLGAVALKSY EEELAKDPRI AATMENAQKG EIMPNIPOMS
AFWYAVRTAV INAASGRQTV DEALKDAQTR ITKGLEVLFQ^GPAM

```

The cleavage site is marked by ^.

PACT-D3 (239-313):

```

ACCGATTATA TTCAGCTGCT GTCTGAAATT GCGAAAGAAC AGGGCTTTAA
CATCACCTAT CTGGATATTG ATGAACTGAG CGCGAACGGC CAGTATCAGT

```

GCCTGGCCGA ACTGAGCACC AGCCCGATTA CCGTGTGCCA TGGCAGCGGC
ATTAGCTGCG GCAACGCGCA GTCTGATGCG GCGCATAACG CGCTGCAGTA
TCTGAAAATT ATCGCGGAAC GCAA

TDYIQLLSEI AKEQGFNITY LDIDELSA NG QYQCLAELST SPITVCHGSG
ISCGNAQSDA AHNALQYLKI IAERK

PACT-Ext-D3 (208-313):

GGCCATAGCC TGGGCTGCAC CTGGCATAGC CTGCGTAACA GCCCGGGTGA
AAAAATTAAC CTGCTGAAAC GTAGCCTGCT GTCTATTCCG AACACCGATT
ATATTCAGCT GCTGTCTGAA ATTGCGAAAG AACAGGGCTT TAACATCACC
TATCTGGATA TTGATGAACT GAGCGCGAAC GGCCAGTATC AGTGCCTGGC
CGAACTGAGC ACCAGCCCGA TTACCGTGTG CCATGGCAGC GGCATTAGCT
GCGGCAACGC GCAGTCTGAT GCGGCGCATA ACGCGCTGCA GTATCTGAAA
ATTATCGCGG AACGCAA

GHSLGCTWHS LRNSPGEKIN LLKRSLLSIP NTDYIQLLSE IAKEQGFNIT
YLDIDELSAN GQYQCLAELS TSPITVCHGS GISCGNAQSD AHNALQYLK
IIAERK

TRBP-D3 (293-366):

GCGTGTGACC GTGTTCTGTC TGAAGTGAAG GAAGAACAGG CGTTTCATGT
GAGCTATCTG GATATTGAAG AACTGAGCCT GAGCGGCCTG TGCCAGTGCC
TGGTGGAACT GAGCACCCAG CCGGCGACCG TGTGTCATGG TAGCGCGACC
ACCCGTGAAG CGGCGCGTGG TGAAGCGGCC CGTCGTGCGC TGCAGTATCT
GAAAATTATG GCGGGCAGCA AA

ACCRVLSELS EEQAFHVSYL DIEELSLSGL CQCLVELSTQ PATVCHGSAT
TREAARGEAA RRALQYLKIM AGSK

TRBP-Ext-D3 (258-366):

AATCGTGGCC CTGGTTGCAC CTGGGATAGC TTACGCAACA GCGTGGGTGA
GAAAATTCTG AGTCTGCGCA GCTGCAGCCT GGGTAGCTTA GGTGCATTAG
GCCCGGCATG TTGTCTGTGT CTGAGCGAGC TGAGCGAAGA ACAGGCCTTC
CACGTGAGTT ACCTGGACAT CGAAGAGCTG AGTCTGAGCG GCCTGTGCCA
GTGCCTGGTG GAGCTGAGCA CACAACCGGC AACAGTGTGT CACGGCAGTG
CAACCACCCG CGAAGCAGCA CGCGGTGAAG CAGCACGTCTG CCGACTGCAG
TACCTGAAGA TTATGGCCCG CAGCAAG

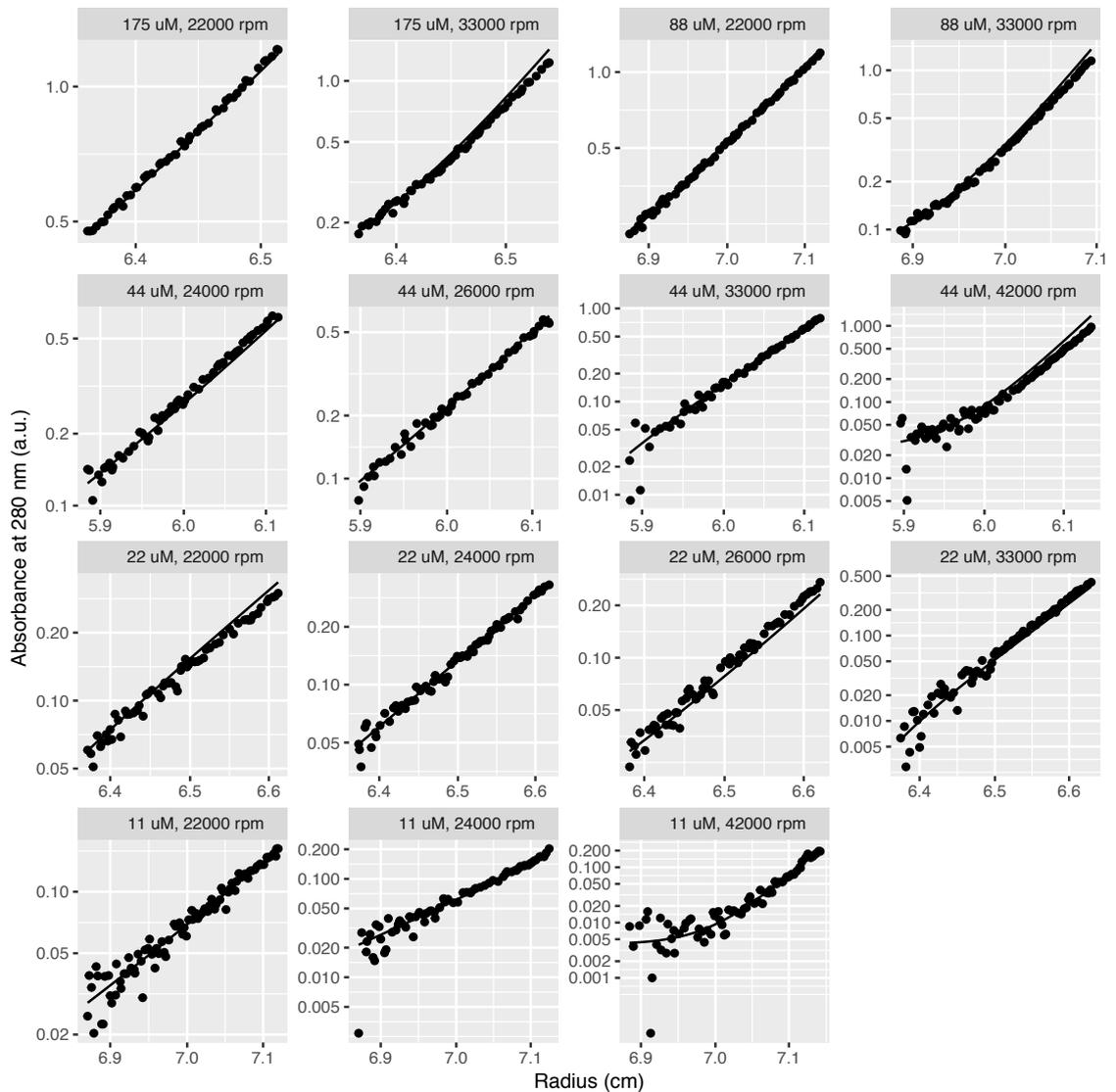
NRPGCTWDS LRNSVGEKIL SLRSCSLGSL GALGPACCRV LSELSEEQAF
HVSYL DIEEL SLSGLCQCLV ELSTQPATVC HGSATTREAA RGEAARRALO
YLKIMAGSK

Loqs-D3 (392-463):

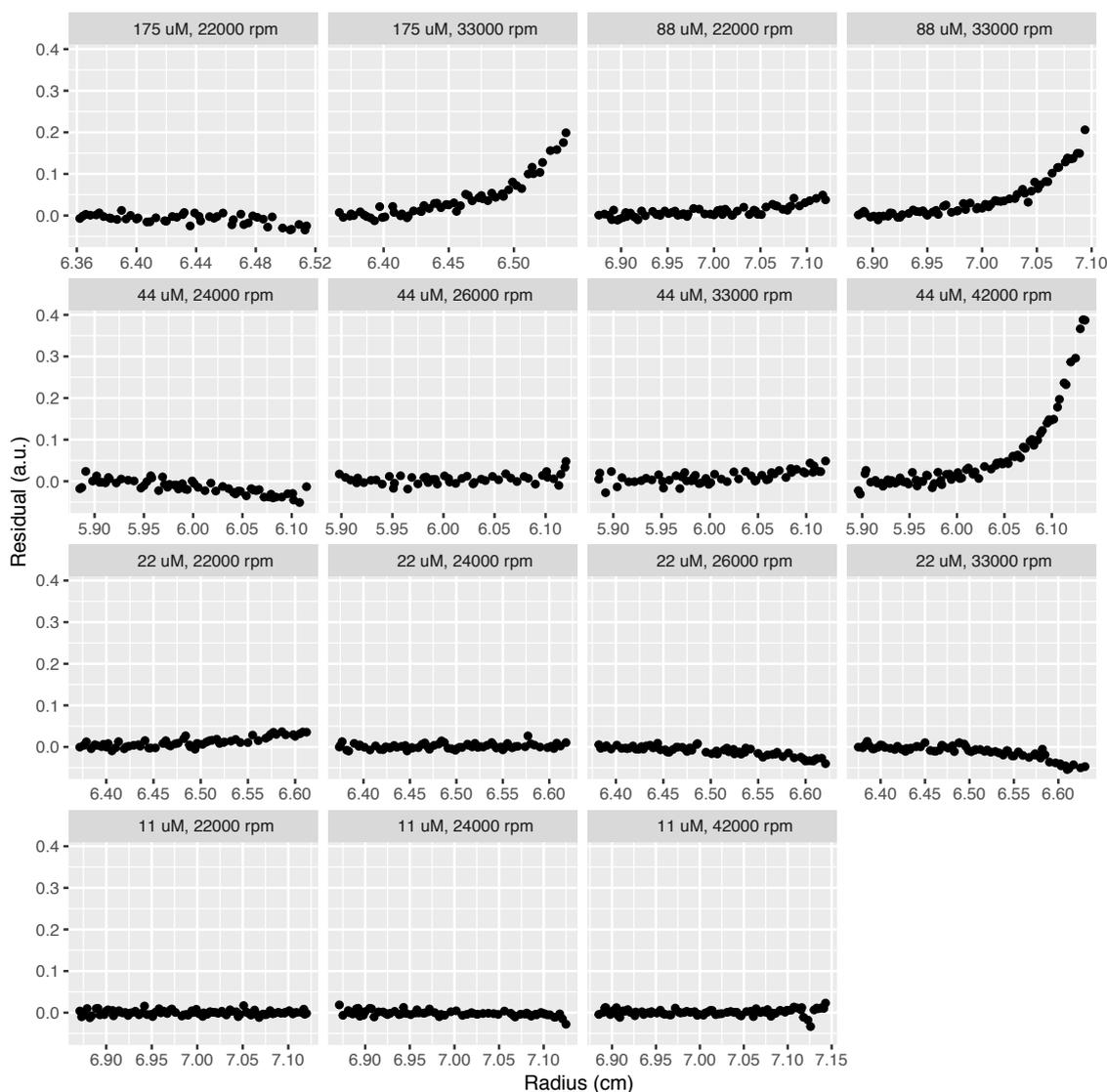
IDYIKLLGEI ATENQFEVTY VDIEEKTFSG QFQCLVQLST LPVGVCHGSG
PTAADAQRHA AQNALEYLKI MT

Note that after TEV cleavage, 3 non-native glycines remain at the N-terminus.

Appendix 2: Analytical Ultracentrifugation fitted data



Raw data and fitted curves for PACT-D3 sedimentation equilibrium AUC data (see Section 7.3)



Residuals from fitting the sedimentation equilibrium AUC data of PACT-D3.

Systematic deviations at larger radii are apparent in most measurements, with several cases (mostly at higher speeds) showing extremely poor fits to the data. This is most likely due to problems fitting the baseline offset term, as other possible sources of error (failure to reach equilibrium, presence of a small fraction of monomer, protein aggregation) were either checked or would give rise to a different pattern of errors. Due to limitations of the fitting software, offset terms could not be fit at the same time as the overall mass and concentrations, and had to be set separately on a more restricted dataset. Although these systematic errors cast doubt on the accuracy of the mass estimate from AUC, this data is sufficient to support the dimeric PACT-D3 being the predominant oligomeric state in solution.

Appendix 3: Resonance assignments of PACT-D3 and PACT-D3 L273R

PACT-D3 L273R backbone resonance assignments

All assignments are given in ppm, referenced to DSS.

Residue	Type	H	N	C	Cα	Cβ	Cγ	Cδ	Hδ	Hϵ	Nδ	Nϵ
236	Pro	-	-	176.44	62.84	32.3	-	-	-	-	-	-
237	Ala	8.55	124.64	177.76	52.4	19.16	-	-	-	-	-	-
238	Met	8.38	119.82	175.9	55.49	32.9	-	-	-	-	-	-
239	Thr	7.88	115.45	173.04	62.05	69.47	-	-	-	-	-	-
240	Asp	8.2	124.8	176.11	52.46	40.15	-	-	-	-	-	-
241	Tyr	7.64	123	177.24	61.15	38.57	-	-	-	-	-	-
242	Ile	8.04	118.69	179.49	65.17	37.93	-	-	-	-	-	-
243	Gln	7.86	121.66	179.03	58.58	28.16	33.46	179.78	-	6.78,7.82	-	112.69
244	Leu	8.34	120.86	179.42	58.37	42	-	-	-	-	-	-
245	Leu	8.57	119.16	177.97	58.12	40.95	-	-	-	-	-	-
246	Ser	7.9	114.31	177.1	61.8	62.65	-	-	-	-	-	-
247	Glu	7.89	122.25	179.52	59.3	29.34	-	-	-	-	-	-
248	Ile	8.03	122.23	177.57	63.49	37.09	-	-	-	-	-	-
249	Ala	8.44	123.38	179.92	54.85	18.9	-	-	-	-	-	-
250	Lys	7.35	116.38	179.21	58.8	32.32	-	-	-	-	-	-
251	Glu	7.64	119.18	178.62	58.91	30.15	-	-	-	-	-	-
252	Gln	8.51	114.79	177.02	56.69	29.11	-	-	-	-	-	-
253	Gly	7.85	109.01	174.2	46.93	-	-	-	-	-	-	-
254	Phe	7.74	115.51	172.36	54.87	41.23	-	-	-	-	-	-
255	Asn	9.15	117.01	174.52	51.74	41.49	176.88	-	6.93,7.51	-	114.31	-
256	Ile	8.64	121.71	175.59	60.32	41.48	-	-	-	-	-	-
257	Thr	8.76	123.72	172.81	61.35	71.1	-	-	-	-	-	-
258	Tyr	8.93	125.05	175.12	57.89	40.14	-	-	-	-	-	-
259	Leu	9.24	126.18	174.83	53.89	43.97	-	-	-	-	-	-

PACT-D3 wild-type resonance assignments (carbons and backbone amides)

State	Residue	Type	H	N	C	C α	C β	C γ	C δ	C ϵ
A	236	<i>Pro</i>	-	-	176.38	62.79	32.21	27	49.62	-
A	237	<i>Ala</i>	8.53	124.6	177.8	52.4	19.11	-	-	-
A	238	<i>Met</i>	8.35	119.59	175.86	55.61	32.82	31.87	-	-
A	239	<i>Thr</i>	7.8	114.72	173	61.87	69.37	21.78	-	-
A	240	<i>Asp</i>	8.12	124.73	175.93	52.33	40.05	-	-	-
A	241	<i>Tyr</i>	7.54	122.55	177.24	61.24	38.51	-	-	-
A	242	<i>Ile</i>	8.03	118.16	175.07	65.14	37.94	16.25,29.19	13.24	-
A	243	<i>Gln</i>	7.73	121.61	179.03	58.62	28.07	33.47	179.76	-
A	244	<i>Leu</i>	8.31	120.67	179.77	58.36	42.21	-	-	-
A	245	<i>Leu</i>	8.54	119.66	177.94	57.91	40.98	26.92	23.00,26.59	-
A	246	<i>Ser</i>	7.93	114.41	177.02	61.59	62.73	-	-	-
A	247	<i>Glu</i>	7.88	121.78	179.41	59.3	29.31	36.16	-	-
A	248	<i>Ile</i>	7.83	122.42	177.4	64.3	37.73	17.55,29.04	12.72	-
A	249	<i>Ala</i>	8.34	123.49	179.8	54.83	18.66	-	-	-
A	250	<i>Lys</i>	7.41	116.42	179.34	58.78	32.25	24.8	29.18	42.02
A	251	<i>Glu</i>	7.64	119.19	178.55	58.94	30.03	35.95	-	-
A	252	<i>Gln</i>	8.49	114.81	176.98	56.77	29.13	34.43	-	-
A	253	<i>Gly</i>	7.88	109.22	174.13	46.91	-	-	-	-
A	254	<i>Phe</i>	7.81	115.65	172.1	54.91	41.3	-	-	-
A	255	<i>Asn</i>	9.04	116.92	174.2	51.78	41.49	177	-	-
A	256	<i>Ile</i>	8.55	121.29	175.51	60.21	41.35	18.27,28.07	16.24	-
A	257	<i>Thr</i>	8.77	124.83	172.61	61.7	70.83	21.43	-	-
A	258	<i>Tyr</i>	8.97	126.14	175.28	58.53	38.94	-	-	-
A	259	<i>Leu</i>	9.58	127.02	174.62	52.93	41.96	26.17	23.79,25.74	-

A 260	Asp	8.43	124.55	175.75	55.5	41.2	-	-	-
A 261	Ile	7.73	123.49	-	60.28	39.22	17.52,27.47	13.12	-
A 262	Asp	8.75	127.6	176.28	55.71	40.74	-	-	-
A 263	Glu	7.53	119.93	177.07	56.32	31.3	35.95	-	-
A 264	Leu	8.7	126.99	178.27	54.64	42.43	26.91	21.87,24.99	-
A 265	Ser	9.26	116.65	-	58.01	64.78	-	-	-
A 266	Ala	8.93	124.57	178.01	54.71	18.25	-	-	-
A 267	Asn	7.85	112.7	175.79	52.16	37.88	-	-	-
A 268	Gly	8.17	108.25	173.76	46.12	-	-	-	-
A 269	Gln	7.67	117.15	175.84	55.52	29.85	34.49	-	-
A 270	Tyr	8.42	120.91	175.99	58.92	39.1	-	-	-
A 271	Gln	9.47	116.58	174.44	54.07	35.02	35.08	-	-
A 272	Cys	8.6	120.43	172.24	56.94	30.75	-	-	-
A 273	Leu	8.99	123.89	175.05	52.95	44.3	26.64	23.39	-
A 274	Ala	9.19	124.56	174.18	50.27	19.68	-	-	-
A 275	Glu	9.02	124.2	175	54.51	32.77	36.55	-	-
A 276	Leu	9.74	130.32	178.19	53.07	42.52	28.42	24.01,25.28	-
A 277	Ser	9.09	121.15	177.1	57.99	62.5	-	-	-
A 278	Thr	7.51	110.89	172.9	61.61	70.42	20.87	-	-
A 279	Ser	8.35	113.36	-	53.99	63.96	-	-	-
A 280	Pro	-	-	176.27	63.1	35.25	25.06	50.37	-
A 281	Ile	8.21	125.94	175.13	62.56	38.06	16.90,28.54	12.45	-
A 282	Thr	7.8	122.93	172.91	62.25	70.8	20.42	-	-
A 283	Val	8.75	126.42	176.3	60.78	33.96	19.91, 22.3	-	-
A 284	Cys	8.92	124.15	172.08	57.76	31.63	-	-	-
A 285	His	9.72	123.67	172.63	54.98	32.41	-	-	-
A 286	Gly	8.52	111.23	171.77	44.52	-	-	-	-

A 287	Ser	9.89	115.44	173.87	56.86	67.25	-	-	-	-
A 288	Gly	8.76	106.06	172.51	46.34	-	-	-	-	-
A 289	Ile	8.69	113.77	174.5	62.81	38.51	18.34,26.78	14.2	-	-
A 290	Ser	7.64	112.79	-	55.65	66.82	-	-	-	-
A 291	Cys	8.77	122.02	-	62.35	26.03	-	-	-	-
A 292	Gly	8.74	108.79	176.56	46.93	-	-	-	-	-
A 293	Asn	7.9	121.38	177.6	56	38.48	-	-	-	-
A 294	Ala	8.16	123.76	178.43	55.41	17.62	-	-	-	-
A 295	Gln	7.62	118.38	-	59.13	27.44	33.05	179.29	-	-
A 296	Ser	7.88	112.99	176.11	61.88	63.02	-	-	-	-
A 297	Asp	8.14	120.82	177.68	57.68	42.92	-	-	-	-
A 298	Ala	8.28	122.44	179.57	55.34	17.71	-	-	-	-
A 299	Ala	8.48	119.78	179.43	55.06	17.95	-	-	-	-
A 300	His	8.62	117.75	177.51	57.66	28.39	-	-	-	-
A 301	Asn	8.38	118.97	177.25	56.16	37.4	175.71	-	-	-
A 302	Ala	8.46	124.42	179.08	55.65	18.6	-	-	-	-
A 303	Leu	8.59	118.55	180.23	58.34	42.31	27.28	24.47, 27.41	-	-
A 304	Gln	7.93	118.12	178.54	59.04	28.72	34.55	179.99	-	-
A 305	Tyr	8.06	121.99	176.57	61.32	37.67	-	-	-	-
A 306	Leu	8.5	118.42	179.26	57.33	41.38	27.01	21.70,26.47	-	-
A 307	Lys	7.56	118.51	177.88	58.8	32.18	24.8	29.18	42.03	-
A 308	Ile	7.51	118.9	178.76	63.79	37.74	17.13,28.34	13.43	-	-
A 309	Ile	7.46	116.24	177.72	62.63	37.49	17.71, 26.86	12.45	-	-
A 310	Ala	7.71	123.81	178.75	53.81	18.91	-	-	-	-
A 311	Glu	7.76	116.76	176.53	56.53	29.96	36.23	-	-	-
A 312	Arg	7.7	121.44	175.43	56.56	30.57	27.1	43.65	-	-
A 313	Lys	7.88	128.01	-	57.63	33.71	24.76	29.09	42.21	-

B 236	<i>Pro</i>	-	-	-	62.23	34.61	24.76	50.23	-
B 236	<i>Pro</i>	-	-	-	-	-	-	-	-
B 237	<i>Ala</i>	8.58	124.69	177.72	52.55	19.13	-	-	-
B 238	<i>Met</i>	8.44	120.05	175.85	55.6	32.81	-	-	-
B 239	<i>Thr</i>	7.78	114.29	173.04	61.82	69.35	21.77	-	-
B 240	<i>Asp</i>	8.11	124.68	175.97	52.33	40.04	-	-	-
B 241	<i>Tyr</i>	7.47	122.56	177.1	61.26	38.51	-	-	-
B 242	<i>Ile</i>	7.98	118.44	-	65.11	37.9	16.27,29.25	13.29	-
B 244	<i>Leu</i>	-	-	179.82	58.2	42.2	26.89	23.75,24.62	-
B 245	<i>Leu</i>	8.6	120.05	177.89	57.98	40.94	-	-	-
B 246	<i>Ser</i>	7.86	114.45	177.05	61.61	62.73	-	-	-
B 247	<i>Glu</i>	7.81	121.65	179.47	59.41	29.33	-	-	-
B 248	<i>Ile</i>	7.77	122.3	177.36	64.43	37.79	-	-	-
B 249	<i>Ala</i>	8.28	123.5	-	54.71	18.84	-	-	-
B 250	<i>Lys</i>	-	-	179.21	58.8	32.26	24.81	29.12	42.06
B 251	<i>Glu</i>	7.6	119.19	178.54	58.94	30.05	35.95	-	-
B 252	<i>Gln</i>	8.39	114.91	176.83	56.7	29.17	-	-	-
B 253	<i>Gly</i>	7.85	108.93	174.16	46.87	-	-	-	-
B 254	<i>Phe</i>	7.9	116.13	172.22	54.94	41.26	-	-	-
B 255	<i>Asn</i>	9.1	117.25	174.01	51.83	41.51	177	-	-
B 256	<i>Ile</i>	8.45	120.77	175.79	59.57	41.13	18.31,28.00	15.65	-
B 257	<i>Thr</i>	8.7	123.6	172.5	61.25	71.22	21.42	-	-
B 258	<i>Tyr</i>	8.92	124.86	175.29	58.24	39.65	-	-	-
B 259	<i>Leu</i>	9.41	127.02	174.74	53.62	43.28	26.93	24.58,24.66	-
B 260	<i>Asp</i>	8.54	125.54	175.89	55	41.15	-	-	-
B 261	<i>Ile</i>	7.9	124.8	176.29	59.81	38.12	17.50,27.33	11.74	-
B 262	<i>Asp</i>	8.75	127.54	176.63	56.22	41	-	-	-

B 263	<i>Glu</i>	7.99	119.86	177.22	57.01	30.31	36	-	-
B 264	<i>Leu</i>	8.49	125.65	178.17	54.61	42.51	26.96	21.79,24.99	-
B 265	<i>Ser</i>	9.17	116.14	175.73	57.96	64.67	-	-	-
B 266	<i>Ala</i>	8.81	124.4	177.77	54.55	18.32	-	-	-
B 267	<i>Asn</i>	7.87	112.99	175.78	52.14	37.88	-	-	-
B 268	<i>Gly</i>	8.15	108.18	-	46.16	-	-	-	-
B 269	<i>Gln</i>	-	-	175.77	55.46	29.95	-	-	-
B 270	<i>Tyr</i>	8.51	121.3	175.98	58.87	39.07	-	-	-
B 271	<i>Gln</i>	9.48	117.27	174.25	54.18	34.18	35.37	-	-
B 272	<i>Cys</i>	8.74	122.38	171.73	56.83	30.42	-	-	-
B 273	<i>Leu</i>	8.89	124.43	175.23	52.86	43.97	26.64	22.81,25.26	-
B 274	<i>Ala</i>	9.12	125.92	174.84	50.28	18.87	-	-	-
B 275	<i>Glu</i>	9.19	125.34	174.95	55.23	31.11	36.45	-	-
B 276	<i>Leu</i>	9.59	128.54	178.1	52.98	43.07	27.98	24.65	-
B 277	<i>Ser</i>	8.9	121.01	-	57.86	62.47	-	-	-
B 278	<i>Thr</i>	-	-	172.71	61.63	70.53	20.96	-	-
B 279	<i>Ser</i>	8.24	112.44	-	54.28	64.09	-	-	-
B 280	<i>Pro</i>	-	-	176.65	63.17	34.71	25.21	50.08	-
B 280	<i>Pro</i>	-	-	-	-	-	-	-	-
B 281	<i>Ile</i>	8.44	124.64	-	63.34	38.61	17.27,29.00	13.04	-
B 282	<i>Thr</i>	7.72	121.56	172.21	62.01	72.56	21.46	-	-
B 283	<i>Val</i>	9.36	123.68	174.2	59.93	33.7	20.95,22.35	-	-
B 284	<i>Cys</i>	8.98	122.22	173.61	56.96	29.65	-	-	-
B 285	<i>His</i>	8.46	126	174.74	57.15	34.14	-	-	-
B 286	<i>Gly</i>	8.6	115.15	171	44.61	-	-	-	-
B 287	<i>Ser</i>	9.48	112.91	173.78	56.79	67.17	-	-	-
B 288	<i>Gly</i>	8.72	105.51	172.75	46.34	-	-	-	-

B 289	Ile	8.69	113.7	-	-	62.9	38.55	-	-	-	-
B 290	Ser	-	-	-	-	55.67	66.76	-	-	-	-
B 291	Cys	8.73	122.01	-	-	62.33	26.03	-	-	-	-
B 292	Gly	8.73	108.66	-	-	46.9	-	-	-	-	-
B 293	Asn	-	-	177.66	-	55.95	38.45	-	-	-	-
B 294	Ala	8.09	123.79	178.72	-	55.49	17.47	-	-	-	-
B 295	Gln	7.69	118.22	177.74	-	59.27	27.24	33.12	179.29	-	-
B 296	Ser	7.91	113.25	176	-	61.78	63	-	-	-	-
B 297	Asp	8.18	121.07	177.49	-	58.01	43.61	-	-	-	-
B 298	Ala	8.25	121.45	179.28	-	55.82	17.59	-	-	-	-
B 299	Ala	8.28	118.86	179.27	-	54.98	18.21	-	-	-	-
B 300	His	8.6	117.66	177.51	-	57.64	28.38	-	-	-	-
B 301	Asn	8.41	119.12	176.71	-	55.84	37.2	174.1	-	-	-
B 302	Ala	8.41	122.7	179.3	-	55.61	19.32	-	-	-	-
B 303	Leu	8.46	118.33	180.71	-	58.38	42.34	27.28	24.47,27.41	-	-
B 304	Gln	8.07	118.14	178.42	-	59.01	28.85	33.84	-	-	-
B 305	Tyr	8.57	121.31	176.57	-	61.6	38.77	-	-	-	-
B 306	Leu	8.93	117.65	179.96	-	57.57	41.39	27.04	21.67,26.45	-	-
B 307	Lys	7.55	119.15	178.49	-	59.31	32.48	24.9	29.59	42.2	-
B 308	Ile	7.42	119.85	179.46	-	64.48	38.15	17.84,29.46	14.43	-	-
B 309	Ile	8.13	116.24	178.61	-	63.8	36.93	17.63,26.44	12.45	-	-
B 310	Ala	7.49	123.21	179.18	-	54.35	18.54	-	-	-	-
B 311	Glu	7.47	114.95	176.87	-	56.48	29.94	36.15	-	-	-
B 312	Arg	7.57	120.49	175.44	-	56.88	30.64	27.11	44.1	-	-
B 313	Lys	7.72	127.44	-	-	-	33.71	-	-	-	-

PACT-D3 wild-type resonance assignments (proton and nitrogen, except backbone amide)

State	Residue	Type	H α	H β	H δ	H ϵ	H γ	N δ	N ϵ
A	236	Pro	4.45	1.93,2.29	3.56	-	1.98	-	-
A	237	Ala	4.28	1.37	-	-	-	-	-
A	238	Met	4.41	2.00,2.02	-	-	2.53,2.57	-	-
A	239	Thr	3.87	3.7	-	-	0.81	-	-
A	240	Asp	4.56	2.42,2.92	-	-	-	-	-
A	241	Tyr	4.11	2.93,3.08	-	-	-	-	-
A	242	Ile	3.65	1.94	0.69	-	0.90,1.04,1.59	-	-
A	243	Gln	4.06	2.03	-	6.76,7.80	2.37	-	112.62
A	244	Leu	-	-	-	-	-	-	-
A	245	Leu	4.25	1.54,2.18	1.03,1.11	-	1.79	-	-
A	246	Ser	4.05	4.04,4.09	-	-	-	-	-
A	247	Glu	4.08	2.13,2.24	-	-	2.33,2.49	-	-
A	248	Ile	3.85	1.97	0.87	-	1.03,1.34,1.68	-	-
A	249	Ala	2.24	1.07	-	-	-	-	-
A	250	Lys	3.98	1.91	1.69	2.97	1.42,1.57	-	-
A	251	Glu	4.04	2.16,2.18	-	-	2.27,2.42	-	-
A	252	Gln	4.3	1.95,2.44	-	-	2.46,2.61	-	-
A	253	Gly	3.88,4.08	-	-	-	-	-	-
A	254	Phe	5.03	2.82,3.05	-	-	-	-	-
A	255	Asn	5.04	2.74,2.76	6.89,7.52	-	-	114.31	-
A	256	Ile	4.78	1.51	0.87	-	0.36,0.98,1.36	-	-
A	257	Thr	4.47	4.03	-	-	1.16	-	-
A	258	Tyr	4.9	2.63,2.77	-	-	-	-	-
A	259	Leu	4.64	1.40,2.05	0.84,0.97	-	1.58	-	-
A	260	Asp	4.67	2.55,2.79	-	-	-	-	-

A 261	Ile	4.26	1.79		0.73	-		1.07,1.15,1.50	-	-
A 262	Asp	4.58	2.73,2.75	-	-	-	-	-	-	-
A 263	Glu	4.32	1.93	-	-	-	-	2.22,2.27	-	-
A 264	Leu	4.52	1.24,1.65	0.27, 0.71	-	-	-	1.51	-	-
A 265	Ser	4.38	3.80,4.27	-	-	-	-	-	-	-
A 266	Ala	4.2	1.47	-	-	-	-	-	-	-
A 267	Asn	4.82	2.88,2.90	-	-	-	-	-	-	-
A 268	Gly	3.62,4.09	-	-	-	-	-	-	-	-
A 269	Gln	4.45	1.89,2.24	-	-	-	-	2.00,2.23	-	-
A 270	Tyr	4.51	2.43,3.02	-	-	-	-	-	-	-
A 271	Gln	5.44	1.97, 2.06	-	-	-	-	2.32,2.55	-	-
A 272	Cys	5.09	2.41,3.20	-	-	-	-	-	-	-
A 273	Leu	5.43	1.34,1.91	0.62,0.67	1	-	-	1.81	-	-
A 274	Ala	4.72	0.86	-	-	-	-	-	-	-
A 275	Glu	4.83	2.34	-	-	-	-	1.71, 1.95	-	-
A 276	Leu	5.44	1.12,1.73	0.62,0.67	-	-	-	1.66	-	-
A 277	Ser	4.72	3.89, 4.44	-	-	-	-	-	-	-
A 278	Thr	3.98	4.14	-	-	-	-	0.43	-	-
A 279	Ser	4.84	3.67,3.83	-	-	-	-	-	-	-
A 280	Pro	4.84	1.82,2.11	3.42, 3.55	-	-	-	1.67,1.80	-	-
A 281	Ile	3.91	1.98	0.88	-	-	-	1.32,1.70	-	-
A 282	Thr	4.86	3.6	-	-	-	-	0.59	-	-
A 283	Val	5.04	1.85	-	-	-	-	-0.16,0.59	-	-
A 284	Cys	5.07	2.48, 2.97	-	-	-	-	-	-	-
A 285	His	5.53	3.08,3.22	-	-	-	-	-	-	-
A 286	Gly	3.85,4.75	-	-	-	-	-	-	-	-
A 287	Ser	5.82	3.85,3.85	-	-	-	-	-	-	-

A 288	<i>Gly</i>	3.88, 4.42	-	-	-	-	-	-	-	-	-	-	-
A 289	<i>Ile</i>	4.58	2.31	0.94	-	-	1.09,1.31,1.41	-	-	-	-	-	-
A 290	<i>Ser</i>	4.32	3.82,3.98	-	-	-	-	-	-	-	-	-	-
A 291	<i>Cys</i>	3.84	2.66,2.73	-	-	-	-	-	-	-	-	-	-
A 292	<i>Gly</i>	3.70,3.78	-	-	-	-	-	-	-	-	-	-	-
A 293	<i>Asn</i>	4.53	2.73,2.97	-	-	-	-	-	-	-	-	-	-
A 294	<i>Ala</i>	3.71	1.25	-	-	-	-	-	-	-	-	-	-
A 295	<i>Gln</i>	4.2	1.68,2.01	-	-	6.33,7.05	2.10,2.35	-	-	-	-	-	110.82
A 296	<i>Ser</i>	4.16	4.13,4.17	-	-	-	-	-	-	-	-	-	-
A 297	<i>Asp</i>	4.6	2.83,2.95	-	-	-	-	-	-	-	-	-	-
A 298	<i>Ala</i>	3.97	1.38	-	-	-	-	-	-	-	-	-	-
A 299	<i>Ala</i>	4.08	1.73	-	-	-	-	-	-	-	-	-	-
A 300	<i>His</i>	3.8	2.42,3.23	-	-	-	-	-	-	-	-	-	-
A 301	<i>Asn</i>	4.37	2.64,2.93	7.12,8.41	-	-	-	-	115.52	-	-	-	-
A 302	<i>Ala</i>	4.42	1.45	-	-	-	-	-	-	-	-	-	-
A 303	<i>Leu</i>	4	1.58,2.17	0.95,1.25	-	-	-	-	-	-	-	-	-
A 304	<i>Gln</i>	4.04	2.18	-	-	6.47,6.97	2.22,2.47	-	-	-	-	-	110.44
A 305	<i>Tyr</i>	4.11	2.58,2.89	-	-	-	-	-	-	-	-	-	-
A 306	<i>Leu</i>	3.67	1.44,1.86	0.74,0.87	-	-	-	1.91	-	-	-	-	-
A 307	<i>Lys</i>	3.98	1.91	1.69	-	2.97	1.42,1.57	-	-	-	-	-	-
A 308	<i>Ile</i>	3.81	1.99	0.86	-	-	0.82,1.20,1.68	-	-	-	-	-	-
A 309	<i>Ile</i>	3.85	1.77	0.66	-	-	0.97,1.13	-	-	-	-	-	-
A 310	<i>Ala</i>	4.19	1.47	-	-	-	-	-	-	-	-	-	-
A 311	<i>Glu</i>	4.26	1.99,2.14	-	-	-	2.27,2.45	-	-	-	-	-	-
A 312	<i>Arg</i>	4.32	1.80,1.93	3.16	-	7.41	1.68,1.78	-	-	-	-	-	112.86
A 313	<i>Lys</i>	4.16	1.74,1.85	1.69	-	3.01	1.43	-	-	-	-	-	-
B 236	<i>Pro</i>	4.51	2.14,2.38	3.58	-	-	1.82,1.93	-	-	-	-	-	-

Appendix 4: R code for simultaneous fitting of longitudinal relaxation and exchange rates

The following code was executed using R version 3.3.3, and used to generate the individual parts of Figure 8.11.

```
# Script to fit T1 and EXSY data of wild-type PACT-D3

setwd("~/Dropbox/Phd/Data/Analysis/P3a") # For Mac

multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  # Multiplot function written by Winston Chang, and available from
  www.cookbook-r.com
  require(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                      ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout),
ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this
      subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
layout.pos.col = matchidx$col))
    }
  }
}
```

```

options = theme_bw() + theme(axis.text.x = element_text(angle=90,
vjust=0.5))
colours = scale_colour_brewer(palette="Set1")
fills = scale_fill_brewer(palette="Set1")

#### First, import, tidy and filter data ####
# First import the raw T1 data
t1_peaks.raw = read.delim("Peak lists/P3a_T1_relaxation_raw_data.txt",
as.is=TRUE,
col.names=c("Dim.1", "Dim.2", "10", "600", "250", "700", "100", "510", "300", "
500", "200", "1500", "1000", "10.001", "400", "2000", "800", "900", "50", "10.00
2", "Fit.Error", "Fit.A", "Fit.B"))

t1_peaks = cbind(t1_peaks.raw[,1], stack(t1_peaks.raw[,3:20])) #
Convert to long format
names(t1_peaks)=c("Assign", "Height", "t_mix")
t1_peaks$t_mix = as.numeric(substring(t1_peaks$t_mix,2))/1000
# Convert text labels into times in s
t1_peaks = cbind(parse_analysis_assignments(t1_peaks$Assign)[,-4],
t1_peaks[,2:3])
t1_peaks$Height = t1_peaks$Height*100 # This just scales the heights
from the t1 experiment to make them comparable to the EXSY
t1_peaks = t1_peaks[!is.na(t1_peaks$Res.N),]
names(t1_peaks)[1]="type"

# Now import EXSY data
exsy = rbind(cbind(read.delim("EXSY peak
lists/EXSY_11.txt",na.strings="None", as.is=TRUE), t_mix=0.9),
cbind(read.delim("EXSY peak
lists/EXSY_12.txt",na.strings="None", as.is=TRUE), t_mix=0.4),
cbind(read.delim("EXSY peak
lists/EXSY_13.txt",na.strings="None", as.is=TRUE), t_mix=0.1),
cbind(read.delim("EXSY peak
lists/EXSY_14.txt",na.strings="None", as.is=TRUE), t_mix=0.5),
cbind(read.delim("EXSY peak
lists/EXSY_15.txt",na.strings="None", as.is=TRUE), t_mix=0.2),
cbind(read.delim("EXSY peak
lists/EXSY_16.txt",na.strings="None", as.is=TRUE), t_mix=1.0),
cbind(read.delim("EXSY peak
lists/EXSY_17.txt",na.strings="None", as.is=TRUE), t_mix=0.8),
cbind(read.delim("EXSY peak
lists/EXSY_18.txt",na.strings="None", as.is=TRUE), t_mix=0.3),
cbind(read.delim("EXSY peak
lists/EXSY_19.txt",na.strings="None", as.is=TRUE), t_mix=0.7),
cbind(read.delim("EXSY peak
lists/EXSY_20.txt",na.strings="None", as.is=TRUE), t_mix=0.6))

exsy = cbind(parse_analysis_assignments(exsy$Assign.F1)[,c(1,2,5)],
Chain.2=parse_analysis_assignments(exsy$Assign.F2)[,1], exsy[,c(-3,-
4)])
exsy$type = paste0(exsy$Chain.2, exsy$Chain) # Determine whether
it's an auto peak (AA, BB) or an exchange peak (AB, BA)
# Note that Chain.2 (the nitrogen assignment) is first, because the N
frequency is measured before the mixing time

#Make a dataframe containing the T1 and EXSY data
names(t1_peaks)[1]="type"
comb_data = rbind(exsy[,c(2,3,7,11,12)],t1_peaks[, c(2,4,5,6,1)])
comb_data = comb_data[!is.na(comb_data$Res.N),] # Remove NAs

# Remove data which is affected by peak overlap
exclude.1 = c(237, 240, 243, 244, 250, 256, 262,
266, 267, 269, 278, 290,
293, 296, 300) # Peaks where neither auto- or cross-peaks are distinct

```

```

exclude.2 = c(239, 241, 242, 245, 246, 247, 248, 249,
              251, 252, 253, 255, 259,
              263, 268, 271,
              277, 289, 291, 292, 294,
              295, 297, 298,
              303, 304, 307, 309) # Peaks
where cross-peaks overlap auto-peaks
comb_data = comb_data[!comb_data$Res.N %in% c(exclude.1, exclude.2),]

exclude.3 = data.frame(Res.N=c(257, 265,275,283,284,286,299,308),
                       type.ex=
c('BA','AB','AB','BA','AB','BA','BA','BA')) # Individual cross-peaks
with overlaps
tmp = merge(comb_data, exclude.3, by="Res.N", all.x=TRUE)
comb_data = comb_data[tmp$type!=tmp$type.ex | is.na(tmp$type.ex),]

exclude.4a = data.frame(Res.N=c(261,272,305,312),
                       type.ex=c('A','A','B','A'))
exclude.4b = data.frame(Res.N=c(261,272,305,312),
                       type.ex=c('AA','AA','BB','AA')) # Individual auto-peaks with overlaps
tmpa = merge(comb_data, exclude.4a, by="Res.N", all.x=TRUE)
tmpb = merge(comb_data, exclude.4b, by="Res.N", all.x=TRUE)
comb_data = comb_data[(tmpa$type!=tmpa$type.ex &
tmpb$type!=tmpb$type.ex) | is.na(tmpa$type.ex),]

exclude.5 = c(238, 270, 273, 301, 311) # Residues with no clear cross-
peaks
comb_data=comb_data[!(comb_data$Res.N %in% exclude.5 & comb_data$type
%in% c('AB','BA')),]

exclude.6a = data.frame(Res.N=c(260,281,282,302,310),
                       type.ex=c('A','B','B','A','A'))
exclude.6b = data.frame(Res.N=c(260,281,282,302,310),
                       type.ex=c('AB','AB','BA','AB','BA')) # Individual auto-peaks with
overlaps
tmpa = merge(comb_data, exclude.6a, by="Res.N", all.x=TRUE)
tmpb = merge(comb_data, exclude.6b, by="Res.N", all.x=TRUE)
comb_data = comb_data[(tmpa$type!=tmpa$type.ex &
tmpb$type!=tmpb$type.ex) | is.na(tmpa$type.ex),]

#### Test the fitting on a single residue ####
active_residue = 313

red_data = comb_data[comb_data$Res.N==active_residue,]
ggplot(data=red_data[red_data$Res.N==active_residue,]) +
  geom_point(aes(x=t_mix, y=Height, colour=type))

# First fit the case where R1_A = R1_B
# According to Rule & Hitchens, peak intensities are given by:
# I_AA(t) = p_A * (p_A + p_B*exp(-k_ex*t)) * exp(-R1*t)
# I_BB(t) = p_B * (p_B + p_A*exp(-k_ex*t)) * exp(-R1*t)
# I_AB(t) = I_BA(t) = p_A * p_B * (1 - exp(-k_ex*t)) * exp(-R1*t)
modell <- nls(Height ~ (type=="AA")*I0*pA*(pA+(1-pA)*exp(-
k_ex*t_mix))*exp(-R1*t_mix) +
                                                    (type=="BB")*I0*(1-
pA)*((1-pA)+pA*exp(-k_ex*t_mix))*exp(-R1*t_mix) +
                                                    (type=="AB")*I0*pA*(1-
pA)*(1-exp(-k_ex*t_mix))*exp(-R1*t_mix) +
                                                    (type=="BA")*I0*pA*(1-
pA)*(1-exp(-k_ex*t_mix))*exp(-R1*t_mix) +
                                                    (type=="A")*I1*(pA*(pA+(1-pA)*exp(-k_ex*t_mix))*exp(-R1*t_mix) +
pA*(1-pA)*(1-exp(-k_ex*t_mix))*exp(-R1*t_mix)) +

```

```

                                (type=="B")*I1*((1-
pA)*((1-pA)+pA*exp(-k_ex*t_mix))*exp(-R1*t_mix) +
                                pA*(1-pA)*(1-exp(-k_ex*t_mix))*exp(-R1*t_mix)),
                                data =
red_data[red_data$Res.N==active_residue,],
                                start = list(I0=7e8, I1=5e8
,pA=0.5, k_ex=0.6, R1=1.25),
                                algorithm="port", # This
algorithm allows setting upper and lower bounds
                                lower=c(1e8, 5e7, 0.3, 0.01,
0.1),
                                upper=c(12e8, 5e9, 0.7, 5,
2),

                                control=nls.control(warnOnly=TRUE), # setting warnOnly means
that if a good fit can't be found, a fit will be returned anyway.
                                trace=FALSE)
summary(modell)

ggplot(data=cbind(red_data[red_data$Res.N==active_residue,],
fit=fitted(modell))) +
  geom_point(aes(x=t_mix, y=Height, colour=type)) +
  geom_line(aes(x=t_mix, y=fit, group=type, colour=type))

# And now the case where R1_A != R1_B
# The below equations are from Farrow, Zhang, Forman-Kay, Kay (1994) J
Bio NMR
L1 = function(pA, rA, rB, k_ex) {
  return( 0.5*(rA+rB+2*k_ex + sqrt((rA-rB)^2 + 4*k_ex^2)) )
}
L2 = function(pA, rA, rB, k_ex) {
  return( 0.5*(rA+rB+2*k_ex - sqrt((rA-rB)^2 + 4*k_ex^2)) )
}
denom = function(pA, rA, rB, k_ex) {
  return( L1(pA, rA, rB, k_ex) - L2(pA, rA, rB, k_ex) )
}
I_AA = function(I0, pA, rA, rB, k_ex, t_mix) {
  return(
    I0*pA*(-(L2(pA, rA, rB, k_ex)-rA-k_ex)*exp(-L1(pA, rA, rB,
k_ex)*t_mix) +
    (L1(pA, rA, rB, k_ex)-rA-k_ex)*exp(-L2(pA, rA, rB,
k_ex)*t_mix))/denom(pA, rA, rB, k_ex)
  )
}
I_BB = function(I0, pA, rA, rB, k_ex, t_mix) {
  return(
    I0*(1-pA)*(-(L2(pA, rA, rB, k_ex)-rB-k_ex)*exp(-L1(pA, rA, rB,
k_ex)*t_mix) +
    (L1(pA, rA, rB, k_ex)-rB-k_ex)*exp(-L2(pA, rA, rB,
k_ex)*t_mix))/denom(pA, rA, rB, k_ex)
  )
}
I_AB = function(I0, pA, rA, rB, k_ex, t_mix) {
  return (
    I0*pA*-k_ex*(exp(-L1(pA, rA, rB, k_ex)*t_mix) - exp(-L2(pA, rA,
rB, k_ex)*t_mix))/denom(pA, rA, rB, k_ex)
  )
}
I_BA = function(I0, pA, rA, rB, k_ex, t_mix) {
  return (
    I0*(1-pA)*-k_ex*(exp(-L1(pA, rA, rB, k_ex)*t_mix) - exp(-L2(pA,
rA, rB, k_ex)*t_mix))/denom(pA, rA, rB, k_ex)
  )
}

```

```

}
# Guess starting parameters
I0_start = sum(red_data[red_data$Res.N==active_residue & red_data$type
%in% c("AA","BB") & red_data$t_mix==0.1,"Height"])
I1_start = sum(red_data[red_data$Res.N==active_residue & red_data$type
%in% c("A","B") & red_data$t_mix==0.01,"Height"])
pA_start = red_data[red_data$Res.N==active_residue &
red_data$type=="A" & red_data$t_mix==0.01,"Height"]/I1_start
#pA_start=0.6
model2 <- nls(Height ~ (type=="AA")*I_AA(I0, pA, rA, rB, k_ex, t_mix)
+
                                                                    (type=="BB")*I_BB(I0,
pA, rA, rB, k_ex, t_mix) +
                                                                    (type=="AB")*I_AB(I0,
pA, rA, rB, k_ex, t_mix) +
                                                                    (type=="BA")*I_BA(I0,
pA, rA, rB, k_ex, t_mix) +
                                                                    (type=="A")*(I_AA(I1,
pA, rA, rB, k_ex, t_mix) + I_BA(I1, pA, rA, rB, k_ex, t_mix))+
                                                                    (type=="B")*(I_BB(I1,
pA, rA, rB, k_ex, t_mix) + I_AB(I1, pA, rA, rB, k_ex, t_mix)),
data =
red_data[red_data$Res.N==active_residue,],
start = list(I0=I0_start,
I1=I1_start, pA=pA_start, k_ex=1.0, rA=0.8, rB=0.8),
#algorithm="port", # This
algorithm allows setting upper and lower bounds
#lower=c(1e6, 1e6, 0, 0.1,
0.1, 0.1),
#upper=c(1e10, 1e10, 1, 10,
10,10),

control=nls.control(warnOnly=TRUE), # setting warnOnly means
that if a good fit can't be found, a fit will be returned anyway.
trace=FALSE)
summary(model2)

# Guess starting parameters
I0_start = sum(red_data[red_data$Res.N==active_residue & red_data$type
%in% c("AA","BB") & red_data$t_mix==0.1,"Height"])
I1_start = sum(red_data[red_data$Res.N==active_residue & red_data$type
%in% c("A","B") & red_data$t_mix==0.01,"Height"])
pA_start = red_data[red_data$Res.N==active_residue &
red_data$type=="A" & red_data$t_mix==0.01,"Height"]/I1_start
#pA_start=0.6
model3 <- nls(Height ~ (type=="AA")*I_AA(I0, pA, rA, rB, k_ex, t_mix)
+
                                                                    (type=="BB")*I_BB(I0,
pA, rA, rB, k_ex, t_mix) +
                                                                    (type=="AB")*I_AB(I0,
pA, rA, rB, k_ex, t_mix) +
                                                                    (type=="BA")*I_BA(I0,
pA, rA, rB, k_ex, t_mix), # +
#
                                                                    (type=="A")*(I_AA(I1,
pA, rA, rB, k_ex, t_mix) + I_BA(I1, pA, rA, rB, k_ex, t_mix))+
#
                                                                    (type=="B")*(I_BB(I1,
pA, rA, rB, k_ex, t_mix) + I_AB(I1, pA, rA, rB, k_ex, t_mix)),
data =
red_data[red_data$Res.N==active_residue,],
start = list(I0=I0_start,
pA=pA_start, k_ex=1, rA=1, rB=1),
#
algorithm="port", # This
algorithm allows setting upper and lower bounds

```

```

#                               lower=c(1e6, 0, 0.1, 0.1,
0.1),
#                               upper=c(1e10, 1, 10, 10,10),

      control=nls.control(warnOnly=TRUE), # setting warnOnly means
that if a good fit can't be found, a fit will be returned anyway.
      trace=FALSE)
summary(model3)

ggplot(data=cbind(red_data[red_data$Res.N==active_residue,],
fit=fitted(model3))) +
  geom_point(aes(x=t_mix, y=Height*(type %in%
c("AA","BB","AB","BA"))+Height*(type %in% c("A","B")), colour=type)) +
  geom_line(aes(x=t_mix, y=fit*(type %in% c("AA","BB","AB","BA")),
group=type, colour=type))

#### Fit all residues where data is available ####
models = list()
fit_results = data.frame()

# Exclude residues which have less than three EXSY cross peaks (on
average)
tmp = table(comb_data[comb_data$type %in%
c('AA','BB','AB','BA'),c(1,4)])
tmp_list = unique(comb_data$Res.N)
exclusions = tmp_list[rowMeans(tmp)<3]

exclusions = c(exclusions, 285, 305, 312) # Also exclude residues 305
and 312 due to peak overlap

residue_list = setdiff(comb_data$Res.N, exclusions)

starting_values = data.frame(Res.N=residue_list, I0=NA, I1=NA, pA=NA,
rA=1, rB=1, k_ex=1)

for (i in residue_list){      # For each residue
  print(paste0("Fitting residue ",i,"..."))
  # fit the model, catching any errors
  # Guess starting parameters
  I0_start = sum(comb_data[comb_data$Res.N==i & comb_data$type
%in% c("AA","BB") & comb_data$t_mix==0.1,"Height"])
  I1_start = sum(comb_data[comb_data$Res.N==i & comb_data$type
%in% c("A","B") & comb_data$t_mix==0.01,"Height"])
  pA_start = comb_data[comb_data$Res.N==i & comb_data$type=="A" &
comb_data$t_mix==0.01,"Height"]/I1_start

  if(length(c(I0_start, I1_start, pA_start))==3) {
    starting_values[starting_values$Res.N==i, c("I0","I1","pA")] =
c(I0_start, I1_start, pA_start)
  }

  m <- try(nls(Height ~ (type=="AA")*I_AA(I0, pA, rA, rB, k_ex,
t_mix) +
                                (type=="BB")*I_BB(I0, pA,
rA, rB, k_ex, t_mix) +
                                (type=="AB")*I_AB(I0, pA,
rA, rB, k_ex, t_mix) +
                                (type=="BA")*I_BA(I0, pA,
rA, rB, k_ex, t_mix) +
                                (type=="A")*(I_AA(I1, pA,
rA, rB, k_ex, t_mix) + I_BA(I1, pA, rA, rB, k_ex, t_mix))+
                                (type=="B")*(I_BB(I1, pA,
rA, rB, k_ex, t_mix) + I_AB(I1, pA, rA, rB, k_ex, t_mix)),

```

```

                                data =
comb_data[comb_data$Res.N==i,],
                                start = list(I0=I0_start,
I1=I1_start, pA=pA_start, k_ex=1, rA=1, rB=1),
                                #algorithm="port", # This
algorithm allows setting upper and lower bounds
                                lower=c(1e6, 1e6, 0, 0.1, 0.1,
0.1),
                                upper=c(1e10, 1e10, 1, 10,
10,10),

control=nls.control(warnOnly=TRUE), # setting warnOnly means that if a
good fit can't be found, a fit will be returned anyway.
                                trace=FALSE), silent=TRUE)
    if (class(m)=="try-error"){
        # If there was an error, set all stored values to NA and
move onto next residue
        models=c(models,NA)
        result = data.frame(Res.N=i, I0=NA, I1=NA, pA=NA,
                                rA=NA, rB=NA,
k_ex=NA,
                                I0.lo=NA,
I0.hi=NA,I1.lo=NA, I1.hi=NA,
                                pA.lo=NA,
pA.hi=NA,rA.lo=NA, rA.hi=NA,
                                rB.lo=NA,
rB.hi=NA,k_ex.lo=NA, k_ex.hi=NA,

message1=geterrmessage(), message2=NA)
        fit_results=rbind(fit_results, result)
        next
    }
    # Calculate 95% confidence intervals, catching any errors
    ci = try(confint(m), silent=TRUE)
    if(class(ci)=="try-error") {
        result = data.frame(Res.N=i, I0=coef(m)[["I0"]],
I1=coef(m)[["I1"]], pA=coef(m)[["pA"]],
                                rA=coef(m)[["rA"]], rB=coef(m)[["rB"]], k_ex=coef(m)[["k_ex"]],
                                I0.lo=NA,
I0.hi=NA,I1.lo=NA, I1.hi=NA,
                                pA.lo=NA,
pA.hi=NA,rA.lo=NA, rA.hi=NA,
                                rB.lo=NA,
rB.hi=NA,k_ex.lo=NA, k_ex.hi=NA,

message1=m$convInfo$stopMessage, message2=geterrmessage())
    }
    else {
        result = data.frame(Res.N=i, I0=coef(m)[["I0"]],
I1=coef(m)[["I1"]], pA=coef(m)[["pA"]],
                                rA=coef(m)[["rA"]], rB=coef(m)[["rB"]], k_ex=coef(m)[["k_ex"]],

                                I0.lo=ci["I0",1], I0.hi=ci["I0",2],I1.lo=ci["I1",1],
I1.hi=ci["I1",2],

                                pA.lo=ci["pA",1], pA.hi=ci["pA",2],rA.lo=ci["rA",1],
rA.hi=ci["rA",2],

                                rB.lo=ci["rB",1], rB.hi=ci["rB",2],k_ex.lo=ci["k_ex",1],
k_ex.hi=ci["k_ex",2],

        message1=m$convInfo$stopMessage, message2=NA)

```

```

    }
    # Store the model results
    models=c(models,list(m))
    fit_results=rbind(fit_results, result)
}

# Export the fitted data
write.table(fit_results, "Relaxation and EXSY/exported fits/full
fit.txt", sep="\t", row.names=FALSE)

#### Plot the results ####
# Exclude residues where the model doesn't appear to fit the data well
fit_results.filtered = fit_results[fit_results$message1%in%c("relative
convergence (4)","converged"),]
excluded_residues = c(238, 270, 273, 301, 311, 283, 284, 285)
fit_results.filtered =
fit_results.filtered[!fit_results.filtered$Res.N %in%
excluded_residues,]

#### Try fitting a single exchange constant to all residues
simultaneously ####
# Based on a method from Miloushev et. al. (2008), Structure
# This uses an approximation to remove the effects of differential
relaxation rates for 1st and 2nd order in t.
# Equation is  $\xi = (I_{AB} \cdot I_{BA}) / (I_{AA} \cdot I_{BB} - I_{AB} \cdot I_{BA})$ 

# Make an appropriate data frame
global_data = comb_data[comb_data$type %in% c('AA','AB','BA','BB') &
comb_data$Res.N %in% residue_list,]
library(reshape2)
tmp = melt(global_data, id.vars=c('Res.N','Res.name','t_mix','type'))
# Puts the data into long format
global_data = dcast(tmp, Res.N+Res.name+t_mix ~ type,
value.var="value") # Put it back into wide format, with separate
columns for AA,AB,BA and BB heights

# Calculate  $\xi$  for each residue
global_data$Xi = NA
attach(global_data)
global_data$Xi = (AB*BA)/(AA*BB-AB*BA)
detach(global_data)
global_data = global_data[!is.na(global_data$Xi),] # Get rid of NA
values
global_data = global_data[global_data$Res.N %in% residue_list,] #
Restrict to the same residues used in the residue-by-residue analysis
#global_data = global_data[global_data$t_mix<=0.5,] # Test restricting
to early timepoints. This only marginally affects the exchange rate.

# Fit  $\xi$  against  $t_{mix}^2$ , with no intercept or linear component
global_fit = lm(Xi ~ 0 + I(t_mix^2), data=global_data)
#[global_data$Res.N==313,])
k_ex.global = sqrt(coef(global_fit))

# Perform bootstrapping to find the error
library(boot)
stat = function(data, indices) {
  d = data[indices,]
  model = lm(Xi ~ 0 + I(t_mix^2), data=d)
  return(sqrt(coef(model)))
}

boot.obj = boot(data=global_data, statistic=stat, R=1000)
print(boot.obj)

```

```

plot(boot.obj)
boot.ci(boot.obj, type='bca')
x=seq(0,1,0.05)
Xi.model = data.frame(t_mix=x,
Xi.pred=boot.obj$t0^2*x^2,Xi.lo=0.64^2*x^2, Xi.hi=0.79^2*x^2)

#### Output pdfs of plots ####
# Plot the EXSY/T1 fits for each residue
for (i in 1:length(models)) {
  if (!is.na(models[[i]])) {
    pdf(paste0("Relaxation and EXSY/plots/exchange
fits/",comb_data[comb_data$Res.N==residue_list[i],]$Res.name[1],".pdf"
), width=11.69/2, height=8.27/2, useDingbats=FALSE)
    print(
      ggplot(data=cbind(comb_data[comb_data$Res.N==residue_list[i],],
fit=fitted(models[[i]])) +
      geom_point(aes(x=t_mix,
y=Height/coef(models[[i]])[["I0"]]*(type %in% c("AA","BB","AB","BA"))+
      Height/coef(models[[i]])[["I1"]]*(type %in%
c("A","B"))),
      colour=substr(type,1,1),
shape=factor(nchar(type))) +
      geom_line(aes(x=t_mix, y= fit/coef(models[[i]])[["I0"]]*(type
%in% c("AA","BB","AB","BA"))+
      fit/coef(models[[i]])[["I1"]]*(type %in%
c("A","B"))),
      group=type, colour=substr(type,1,1),
linetype=factor(nchar(type))) +
      xlab("Mixing time (s)") + ylab("Relative peak height") +
      ggtitle(paste("Raw data and fitted values for
residue",residue_list[i])) +
      scale_colour_brewer(name="Chain", palette="Set1") +
      scale_linetype_discrete(name="Experiment", labels=c("T1",
"EXSY")) +
      scale_shape_discrete(name="Experiment", labels=c("T1",
"EXSY")) +
      theme_bw()
    )
    dev.off()
  }
}

# Plot various parameters from the T1/EXSY fits, with poorly fitting
residues excluded
pdf(paste0("Relaxation and EXSY/plots/Exchange rate plot (combined
EXSY,T1 fit).pdf"), height=11.69/4, width=8.27/2, useDingbats=FALSE)
print(
  ggplot(data=fit_results.filtered, aes(x=factor(Res.N))) +
  geom_bar(aes(y=k_ex), stat="identity") +
  geom_errorbar(aes(ymin=pmax(k_ex.lo, 0.1, na.rm=TRUE),
ymax=pmin(k_ex.hi,10,na.rm=TRUE)), width=.5) +
  xlab("Residue") + ylab("Exchange rate (/s)") +
  ggtitle("Fitted exchange rate with 95% confidence intervals") +
  coord_cartesian(ylim=c(0, 1.5)) + options
)
dev.off()

pdf(paste0("Relaxation and EXSY/plots/T1 plot (combined EXSY,T1
fit).pdf"), height=11.69/4, width=8.27/2, useDingbats=FALSE)
print(
  ggplot(data=fit_results.filtered, aes(x=factor(Res.N))) +
  geom_point(aes(y=1/rA, colour="A"), size=3) +
  geom_point(aes(y=1/rB, colour="B"), size=3) +

```

```

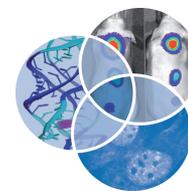
    geom_errorbar(aes(ymin=1/pmax(rA.lo, 0.1, na.rm=TRUE),
ymax=1/pmin(rA.hi,10,na.rm=TRUE), colour="A", width=.5)) +
    geom_errorbar(aes(ymin=1/pmax(rB.lo, 0.1, na.rm=TRUE),
ymax=1/pmin(rB.hi,10,na.rm=TRUE), colour="B", width=.5)) +
    xlab("Residue") + ylab("T1 (s)") +
    ggtitle("Fitted T1 relaxation time with 95% confidence intervals")
+
    coord_cartesian(ylim=c(0, 2)) + scale_colour_brewer(name="Chain",
palette="Set1") + options + theme(legend.position=c(0.2, 0.2))
)
dev.off()

pdf(paste0("Relaxation and EXSY/plots/Population plot (combined
EXSY,T1 fit).pdf"), width=11.69, height=8.27/2, useDingbats=FALSE)
print(
    ggplot(data=fit_results.filtered, aes(x=factor(Res.N))) +
    geom_bar(aes(y=pA), stat="identity") +
    geom_errorbar(aes(ymin=pmax(pA.lo, 0, na.rm=TRUE),
ymax=pmin(pA.hi,1,na.rm=TRUE), width=.5)) +
    xlab("Residue") + ylab("Fractional population of state A") +
    ggtitle("Fractional population of state A with 95% confidence
intervals") +
    coord_cartesian(ylim=c(0, 1)) + scale_colour_brewer(name="Chain",
palette="Set1") + options
)
dev.off()

# Plot results from the global fit
pdf("Relaxation and EXSY/plots/Global fit of EXSY data.pdf",
width=8.27/2, height=8.27/2, useDingbats=FALSE)
print(
    ggplot() + geom_ribbon(data=Xi.model, aes(x=t_mix, ymax=Xi.hi,
ymin=Xi.lo), alpha=0.25) +
    geom_point(data=global_data, aes(x=t_mix, y=Xi, colour=Res.name),
alpha=0.75) +
    geom_line(data=Xi.model, aes(x=t_mix, y=Xi.pred)) + xlab("Mixing
time (s)") + ylab("Composite parameter Xi") +
    ggtitle(paste0("Global fit of exchange rate = ",
format(k_ex.global, digits=3), " /s")) +
    options
)
dev.off()

```

Appendix 5: Dissecting the roles of
TRBP and PACT in double-stranded
RNA recognition and processing of
noncoding RNAs



Dissecting the roles of TRBP and PACT in double-stranded RNA recognition and processing of noncoding RNAs

Alex Heyam,¹ Dimitris Lagos² and Michael Plevin^{1*}

HIV TAR RNA-binding protein (TRBP) and Protein Activator of PKR (PACT) are double-stranded (ds) RNA-binding proteins that participate in both small regulatory RNA biogenesis and the response to viral dsRNA. Despite considerable progress toward understanding the structure–function relationship of TRBP and PACT, their specific roles in these seemingly distinct cellular pathways remain unclear. Both proteins are composed of three copies of the double-stranded RNA-binding domain, two of which interact with dsRNA, while the C-terminal copy mediates protein–protein interactions. PACT and TRBP are found in a complex with the endonuclease Dicer and facilitate processing of immature microRNAs. Their precise contribution to the Dicing step has not yet been defined: possibilities include precursor recruitment, rearrangement of dsRNA within the complex, loading the processed microRNA into the RNA-induced silencing complex, and distinguishing different classes of small dsRNA. TRBP and PACT also interact with the viral dsRNA sensors retinoic acid-inducible gene I (RIG-I) and double-stranded RNA-activated protein kinase (PKR). Current models suggest that PACT enables RIG-I to detect a wider range of viral dsRNAs, while TRBP and PACT exert opposing regulatory effects on PKR. Here, the evidence that implicates TRBP and PACT in regulatory RNA processing and viral dsRNA sensing is reviewed and discussed in the context of their molecular structure. The broader implications of a link between microRNA biogenesis and the innate antiviral response pathway are also considered. © 2015 John Wiley & Sons, Ltd.

How to cite this article:

WIREs RNA 2015. doi: 10.1002/wrna.1272

INTRODUCTION

Double-stranded (ds) RNA has a vital role in normal cellular function. Any structured RNA, such as the ribosome, contains regions of dsRNA, which are often part of higher order tertiary structure. In addition, dsRNA is an intermediate in the biogenesis of short regulatory RNAs such as micro

RNAs (miRNAs), endogenous short interfering RNAs (siRNAs), and Piwi-interacting RNAs (piRNAs).¹

However, dsRNA is also an important replication intermediate for RNA viruses, for instance SARS coronavirus, poliovirus, and hepatitis C virus. It is crucial for cells to distinguish cellular dsRNA from viral dsRNA, and respond appropriately. Differentiation of self versus non-self dsRNA is achieved by proteins that have evolved to recognize chemical features specific to viral dsRNA, such as terminal 5' triphosphate groups. These proteins are termed pattern recognition receptors (PRRs), and include TLRs (Toll-like receptors) 3, 7, and 8; RIG-I (retinoic acid-inducible gene I);

*Correspondence to: michael.plevin@york.ac.uk

¹Department of Biology, University of York, York, UK

²Centre for Immunology and Infection, University of York, York, UK

Conflict of interest: The authors have declared no conflicts of interest for this article.

and MDA5 (melanoma differentiation associated protein 5).²

The closely related mammalian proteins TRBP [HIV trans-activation responsive (TAR) RNA-binding protein; HGNC symbol, TARBP2; UniProt, Q15633] and PACT (Protein Activator of PKR; PRKRA; O75569) bridge several of these pathways (Figure 1(a) and (b)). TRBP was initially identified through its interaction with the HIV TAR RNA element, and both proteins were found to regulate the response to viral dsRNA through the protein PKR (double-stranded RNA-activated protein kinase; EIF2AK2).^{3,4} Subsequently, TRBP and PACT were shown to interact with Dicer (DICER1), the ribonuclease responsible for processing the precursors of miRNAs and siRNAs.^{5,6} More recently, PACT has been shown to activate RIG-I (DDX58), another innate immune sensor of viral RNA.⁷ However, these different roles are often studied independently, and the links between them remain largely unexplored.

This review will summarize current knowledge about TRBP and PACT, and how they influence small RNA biogenesis and viral sensing.

TRBP AND PACT EACH CONTAIN THREE DOUBLE-STRANDED RNA-BINDING DOMAINS

TRBP and PACT each contain three double-stranded RNA-binding domains (dsRBDs) that are separated by unstructured linker regions^{8,9} (Figure 1(c)). Many RNA-binding proteins contain multiple dsRBDs: PKR, ADAR2, and DGCR8 all have tandem dsRBDs, while human Staufen has five.^{10–13} The presence of multiple dsRBDs in the same protein is thought to allow greater affinity and specificity, and to allow functional divergence of individual domains.^{14,15}

Although all dsRBDs share the core α - β - β - α fold (Figure 2), they can be divided into two subgroups depending on sequence conservation. The type A dsRBD is the canonical form, which shows amino acid conservation in three regions involved in dsRNA binding (Figure 2(a)). Type B dsRBDs show conservation only at the C-terminal end of the domain, and are generally unable to bind dsRNA, despite retaining the same overall three-dimensional (3D) structure.^{8,18} PACT and TRBP contain both classes of dsRBD (Figure 2(a)).

Some dsRBDs have additional structural elements that contribute to their function: for example, Rnt1p dsRBD from *Saccharomyces cerevisiae* has an additional α -helix that contributes to domain stability and RNA binding,¹⁹ while dsRBDs from the *Caenorhabditis elegans* TRBP homolog, RDE-4, have

recently been shown to contain numerous additional helices and extended loop regions²⁰ (see Box 1).

BOX 1

HOMOLOGS OF TRBP AND PACT:

C. elegans RDE-4

C. elegans RDE-4 forms a complex with homologs of Dicer and Argonaute, and is required for dsRNA-mediated RNA interference (RNAi), but not miRNA-mediated silencing.²¹ RDE-4 contains three dsRBDs and binds nonspecifically to dsRNA, with an affinity for siRNA similar to that reported previously for TRBP.^{22–24} The first two dsRBDs appear to function independently, with the second having a higher affinity for dsRNA, as was shown for TRBP.^{20,22} RDE-4 has a higher affinity for longer dsRNAs, which may be due to cooperative binding, or an effect related to the overlap in potential binding sites.

As in TRBP and PACT, the C-terminal dsRBD of RDE-4 is required for interaction with Dicer, and also mediates dimerization.²⁴ In addition, the linker between domains 1 and 2 is required for dsRNA processing by Dicer.²⁵ The first two dsRBDs of RDE-4 have additional structural elements that extend into this linker region.²⁰ It will be interesting to see how these additional elements affect the function of the domains—it has already been shown that domain 1 is required for binding siRNAs, while having only a small influence on binding to longer dsRNA.²⁵

TRBP and PACT Domains 1 and 2 Bind dsRNA

Domains 1 and 2 of both TRBP and PACT are type A dsRBDs, and all bind dsRNA.^{22,26} In type A dsRBDs, three distinct regions participate in dsRNA binding (Figure 2(a) and (b)). Unlike dsDNA, dsRNA adopts an A form helix in which the major groove is deep and narrow, limiting access to the bases and therefore to sequence-specific information. dsRBDs bind across two adjacent minor grooves and the intervening phosphate backbone, burying $\sim 780 \text{ \AA}^2$ of the domain surface. In general, it is thought that dsRBDs bind nonspecifically to dsRNA. On discovery, TRBP was thought to specifically recognize certain structured RNAs, such as HIV-1 TAR RNA^{27,28}; however, later evidence has shown that TRBP and PACT interact with a broad range of targets.^{22,23}

However, several well-studied dsRBD-containing proteins do act on specific targets. For example, Staufen can regulate translation and decay

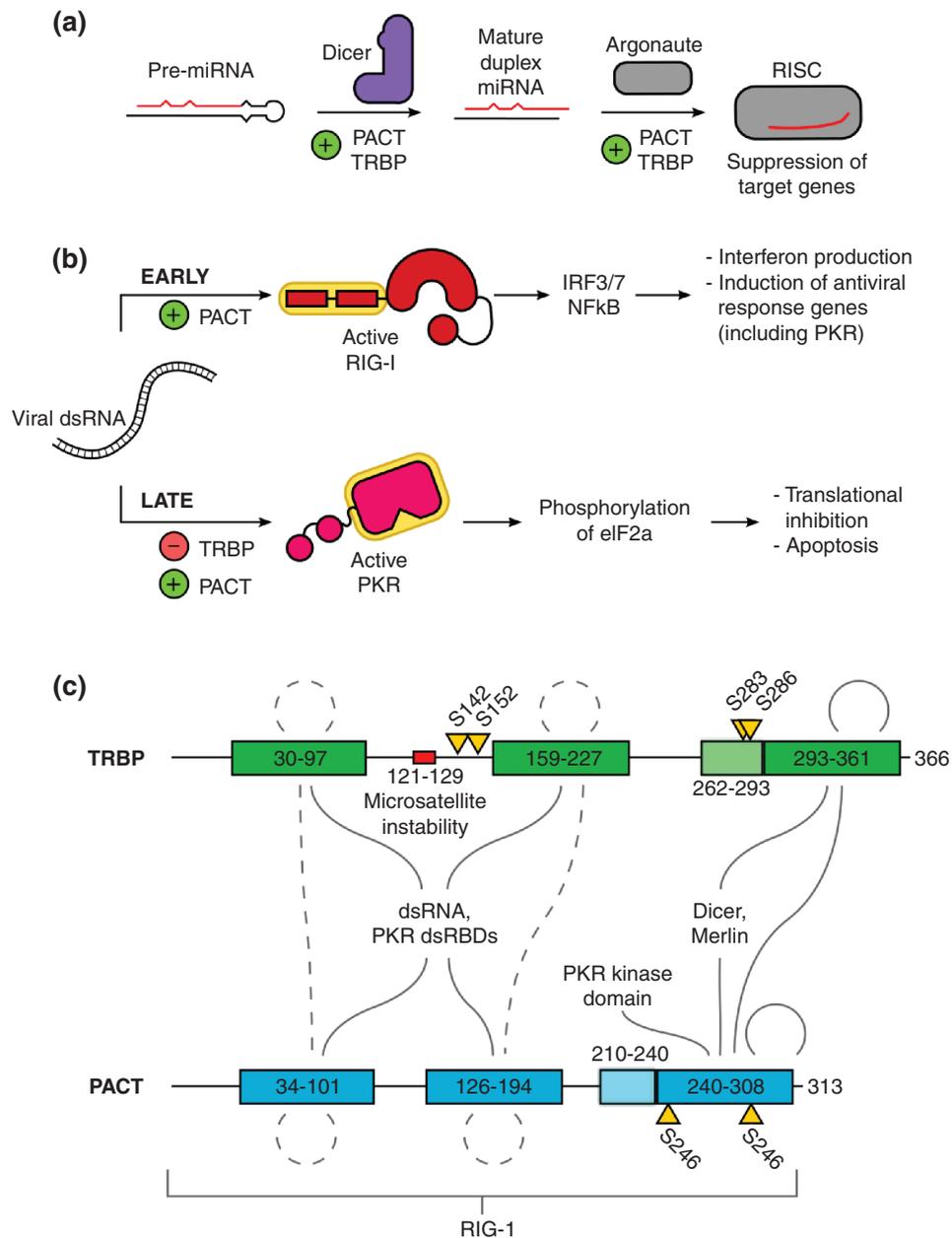


FIGURE 1 | Functions and domain composition of TAR RNA-binding protein (TRBP) and Protein Activator of PKR (PACT). (a) Precursor (pre-)microRNAs (miRNAs) are RNA hairpins that are produced in the nucleus, and exported to the cytoplasm. They contain the ~22 nt sequence of the mature miRNA, indicated in red. The endonuclease Dicer removes the terminal loop to give an RNA duplex, one strand of which is loaded into an Argonaute (Ago) protein to form RNA-induced silencing complex (RISC). TRBP and PACT are implicated in both Dicing and RISC loading. (b) PACT and TRBP have roles in at least two viral response pathways. First, PACT can facilitate activation of retinoic acid-inducible gene I (RIG-I) by viral double-stranded RNA (dsRNA) (distinguished from cellular dsRNA by distinct molecular features, discussed in section *TRBP and PACT mediate innate immune surveillance*). This begins a signaling cascade that results in the production of interferon and other antiviral genes. PACT and TRBP also regulate PKR, a kinase that targets the translation initiation factor eIF2 α to inhibit protein production and promote apoptosis. TRBP inhibits PKR, while PACT can activate it in response to cellular stress. (c) Both PACT and TRBP contain three double-stranded RNA-binding domains (dsRBDs). The first two domains can bind dsRNA, while the third cannot. Interactions with many other proteins have been documented, particularly for the third domain. Solid lines indicate direct protein–protein interactions, while dashed lines indicate interactions that may be mediated via dsRNA. The third dsRBD potentially has an N-terminal extension, based on sequence conservation. Each protein has a number of phosphorylation sites (marked by yellow triangles) that regulate function under certain conditions. A region of TRBP implicated in cancers exhibiting microsatellite instability is indicated in red.

of specific mRNAs, while ADAR proteins can convert adenosine to inosine at precise positions in a variety of dsRNAs.^{29,30} In both cases, RNA recognition is believed to occur through tertiary structures, such as bulges and loops.^{19,31} Some specificity may also arise through contact with the edges of bases in the minor groove, and through the combination of dsRBDs with weak sequence or secondary structure preferences.^{15,18} Current evidence does not rule out PACT and TRBP having a degree of substrate specificity.

TRBP and PACT Bind Protein Partners Primarily via Their Third Domain

The C-terminal dsRBDs of both TRBP and PACT (hereafter referred to as TRBP-D3 or PACT-D3) are type B dsRBDs. They do not bind RNA, but are instead required for interaction with other proteins. It is thought that many type B dsRBDs mediate protein–protein interactions: for example, *Drosophila* Stauf domain 5 binds to Miranda to enable mRNA localization.^{32,33} However, currently, there is no structural information available about how dsRBDs interact with other proteins.

Both PACT-D3 and TRBP-D3 bind to Dicer,^{5,6,34,35} while PACT-D3 has also been shown to interact with the virus-sensing protein PKR.^{7,9} TRBP-D3 is not thought to interact with PKR,³⁶ suggesting that binding to Dicer and PKR is mediated by different regions of PACT-D3. These interactions are discussed further in sections *TRBP and PACT Form Part of the Small RNA Biogenesis Pathway* and *TRBP and PACT Mediate Innate Immune Surveillance of dsRNA* below. Other binding partners have been suggested for TRBP and PACT, such as the tumor suppressor Merlin,³⁷ but the nature of their interactions has not yet been studied in detail.

Although several dsRBDs have been reported to dimerize (e.g., the dsRBDs from PKR³⁸ and ADAR2³⁹), it has been difficult to distinguish between direct binding and indirect association via dsRNA.⁴⁰ There is substantial biochemical evidence that PACT-D3 can homodimerize or heterodimerize with TRBP-D3.^{41–43} While there is no direct biophysical evidence that TRBP-D3 homodimerizes, full-length TRBP can form dimers,²³ though a construct lacking TRBP-D3 does not.²² Dimerization of PACT and TRBP has been suggested to modulate PACT activation of PKR,⁴³ but could also increase their affinity for dsRNA, by bringing together four type A dsRBDs.²⁶

The recent 3D structure of dsRBD 5 of Stauf (Stauf-D5) gave the first high-resolution information about a dsRBD dimer³² and revealed several features that distinguish type B dsRBDs, including

the absence of the conserved histidine residue in loop- β 1,2, the presence of large negatively charged patches on the canonical dsRNA-binding surface, and a region of extra-dsRBD structure, which is required for dimerization. As the linker between the core domain and the extended region is flexible, it was proposed that dimerization occurs via a domain-swapping mechanism (Figure 2(c)).

TRBP-D3 and PACT-D3 both lack the key histidine residue and have negatively charged residues at sites that would correspond to the dsRNA-binding surface. Furthermore, both domains are preceded by a highly conserved 20 amino acid sequence, which may suggest a similar dsRBD/dsRBD interaction to that seen in Stauf-D5. Interestingly, the conserved residues between TRBP-D3 and PACT-D3 are predicted to cluster on the opposite surface to the typical dsRNA-binding face, which potentially suggests the location of a conserved protein interaction site.

There is also evidence that the type A dsRBDs of TRBP and PACT mediate protein–protein interactions. Domains 1 and 2 of PACT are reported to directly interact with the dsRBDs of PKR.^{9,44} PKR also dimerizes partly through its dsRBDs,³⁸ and PACT can still activate PKR if its first two domains are replaced with those from PKR.⁹ The first two dsRBDs of PACT and TRBP have also been reported to homodimerize and heterodimerize,^{41,43} which suggests there may be a common dimerization mechanism between these domains. However, current evidence does not exclude the possibility that these interactions are indirect or mediated by dsRNA.

TRBP AND PACT FORM PART OF THE SMALL RNA BIOGENESIS PATHWAY

RNAi is a highly conserved process (see Figure 1(a)) in which short RNA molecules cause translational (or transcriptional) silencing of complementary messenger RNAs (mRNAs). The cellular pathways that underpin RNAi have been discussed in detail by several recent reviews.^{1,45,46} Here, we will provide only a brief summary of RNAi in mammals.

Somatic cells contain two major classes of small regulatory RNA: miRNA and siRNA. Both are approximately 22 nt long, but differ in their biogenesis and target recognition. miRNAs are derived from RNA hairpins that usually contain several mismatched bases, while siRNAs are processed from longer dsRNA molecules. siRNAs are typically highly complementary to their targets, whereas miRNAs require a 6–8 nt ‘seed’ region to be perfectly matched, with other positions contributing only weakly to target specificity. Although there are relatively few

reports of endogenous mammalian siRNAs, dsRNA can be introduced exogenously either as RNA hairpins (shRNA) or as short dsRNA duplexes.

In the canonical miRNA biogenesis pathway, miRNAs are transcribed as long primary (pri-) miRNAs that contain a 60–80 nucleotide hairpin structure. The pri-miRNA is processed in the nucleus by Drosha/DGCR8 (the ‘microprocessor’ complex) to a shorter precursor (pre-) miRNA. The pre-miRNA is exported from the nucleus by Exportin 5, and further cleaved by a protein complex containing the ribonuclease Dicer, one of four Argonaute proteins (Ago1–4), and either TRBP or PACT (Figure 3(a)–(c)). This second processing step removes the terminal hairpin loop, leaving an RNA duplex of ~22 bp in length.

One strand of the dsRNA Dicer product is removed, while the other (the guide strand) is loaded into Argonaute. This Ago:miRNA complex, together with its protein partners, is referred to as the RNA-induced silencing complex (RISC). Despite recent progress, the mechanism of strand-selective RISC loading remains to be determined^{47,48}; however, many miRNAs can contribute either strand to RISC.⁴⁹ Once loaded, the guide strand targets RISC to complementary mRNAs, which results in either suppression of translation, enhanced degradation and/or cleavage of the mRNA. Although only Ago2 has the ability to cleave mRNAs, miRNAs and siRNAs appear to distribute among all four Argonautes with little specificity.^{50,51}

How Do TRBP and PACT Affect Processing by Dicer and the Formation of Active RISC?

There is now clear evidence that TRBP and PACT can associate with Dicer, both from immunoprecipitation experiments^{5,6,35,52,53} and from *in vitro* reconstitution of the complex from the individually purified components.^{54,55} Furthermore, there is strong support for the idea that PACT and TRBP are involved in processing of pre-miRNAs and RISC loading.^{5,58} However, the molecular mechanism by which TRBP and PACT facilitate miRNA and siRNA processing remains unclear. It has long been established that Dicer alone is able to cleave pre-miRNA to mature miRNA *in vitro*,⁵⁶ and there is evidence that some substrates can be loaded into Ago2 *in vitro* in the absence of TRBP or PACT.^{48,57,58}

The importance of tackling this question is underscored by the evidence linking miRNA biogenesis in general, and TRBP in particular, to a number of different cancers.^{64–68} In particular, tumors exhibiting microsatellite instability are prone to frameshift mutations in TRBP (see Figure 1(c)), leading to impaired

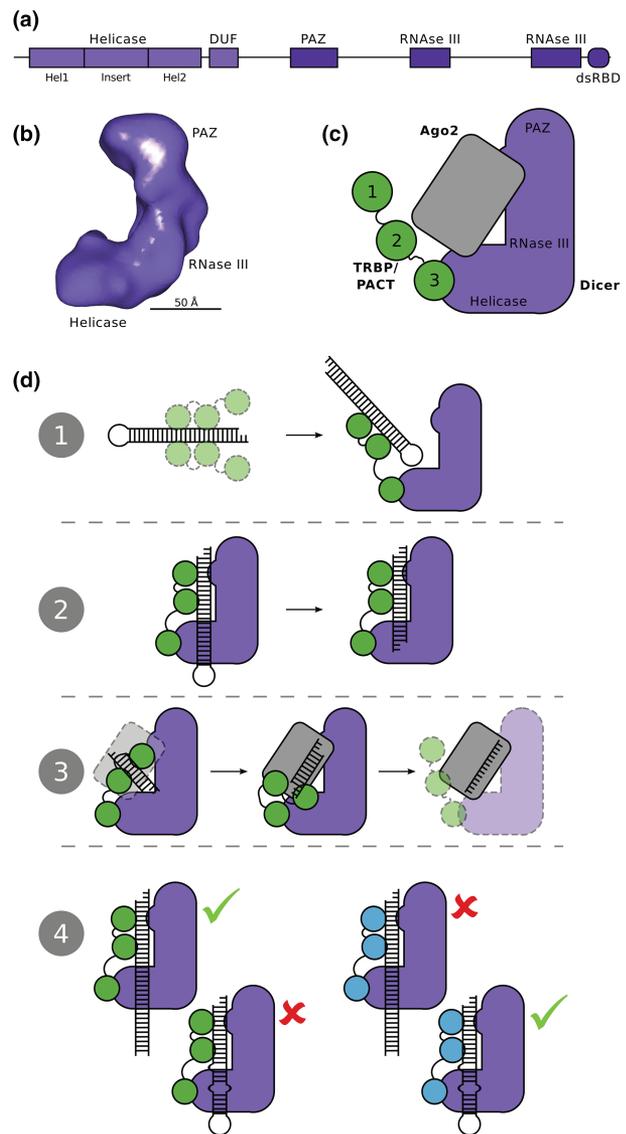


FIGURE 3 | (a) Layout of domains within Dicer. (b) Reconstruction of the Dicer–TRBP RNA-binding protein (TRBP) complex from cryo-electron microscopy (EM) data, with ~15 Å resolution (EM data bank accession EMD-1646).⁵⁹ The locations of the RNase III and helicase domains are inferred from epitope-tagged Dicer.⁶⁰ The position of TRBP is not resolved. (c) A schematic of a minimal RNA-induced silencing complex (RISC)-loading complex of Dicer, Ago2, and TRBP/PACT (Protein Activator of PKR) based on cryo-EM data.⁶¹ All components are approximately to scale. It is unknown whether all three components assemble prior to double-stranded RNA (dsRNA) binding, or if the complex is more dynamic. *In vivo*, it is likely that other proteins associate with the RISC-loading complex.^{62,63} (d) At least four possible roles for TRBP/PACT can be envisaged (see section *How Do TRBP and PACT Affect Processing by Dicer and the Formation of Active RISC?*): (1) TRBP/PACT may help recruit dsRNA to Dicer; (2) TRBP/PACT may aid alignment of dsRNA for cleavage by Dicer; (3) TRBP/PACT may help dsRNA unwinding and/or loading into Argonaute proteins; or (4) TRBP/PACT may favor processing and loading of different substrates into RISC (the substrates shown are illustrative only).

miRNA production.⁶⁵ The drug enoxacin has recently been investigated as a cancer therapeutic owing to its effect on miRNA biogenesis, and is believed to act through TRBP.^{69,70} Although several studies have linked PACT expression to cancer, the effect is not as well documented as for TRBP.^{71,72}

There are several points along the miRNA biogenesis and RISC loading pathway at which TRBP and PACT could act (see also Figure 3(d)):

- (1) Recruitment of substrates to Dicer
 - (a) TRBP and/or PACT may increase the affinity of the Dicer complex for dsRNA, or
 - (b) more specifically increase Dicing rate through favoring binding of substrates over products.

- (2) Facilitating efficient substrate cleavage by Dicer

TRBP and/or PACT could ensure that the substrate dsRNA or pre-miRNA has the optimum orientation for Dicer cleavage.

- (3) Removing the Dicer product
 - (a) TRBP and/or PACT may aid unloading of Dicer or loading of Dicer products into Argonaute proteins,
 - (b) in particular affecting which strand is loaded.

- (4) Controlling which type of dsRNA is loaded into Argonaute

TRBP and/or PACT may preferentially recognize different subsets of miRNA or siRNA (or their precursors), resulting in differences in processing or RISC loading.

We will now discuss evidence from structural, biochemical, biophysical, and molecular cell biology studies in the context of these possible mechanisms.

Evidence from Structural Studies

As described above, TRBP and PACT have been shown to interact with Dicer, a 1922-amino acid type III ribonuclease (Figure 3). The mechanism by which Dicer controls the length of small RNAs was elegantly resolved in an early crystallographic study.^{59,73} The Dicer PAZ domain binds one end of the pre-miRNA, while a 'ruler' domain positions the PAZ domain at a distance from the RNase III site that corresponds to

roughly 22 base pairs of dsRNA. This simple structural mechanism appears to function independently of binding partners as isolated Dicer can still cleave dsRNA targets.

While a considerable amount is known about the structure–function relationship of the RNase III/PAZ region of Dicer, much less is known about the N-terminal DExD/H helicase domain. The helicase domain consists of Hel1 and Hel2 regions separated by a 100-amino acid insert region (Figure 4(a)), which has no sequence homologs outside of the Dicer family. Yeast-2-hybrid screens have identified the insert region as the site of TRBP (and presumably PACT) binding.³⁵ This domain does not appear necessary for dsRNA cleavage *in vitro*,⁷⁴ nor does processing of pre-miRNAs require ATP.⁷⁵

Atomic resolution structures of larger constructs of human Dicer or complexes containing TRBP or PACT are not currently available. The 20–30 Å resolution cryo-electron microscopy studies of a Dicer alone,⁶⁰ in complex with TRBP,⁵⁹ or TRBP and Ago2 (a minimal RISC-loading complex)⁶¹ have been reported. They reveal an L-shaped overall structure with the helicase located in the base, and the RNase III and PAZ domains located in the long arm. However, these studies have been unable to identify precisely where and how TRBP interacts with Dicer.

Because only the C-terminal domains (CTDs) of TRBP or PACT bind to Dicer, it appears likely that domains 1 and 2 have considerable freedom of movement, consistent with roles 1–3 outlined above. To exclude proposed roles for TRBP, it will be necessary to determine more precisely where it interacts with Dicer. Similarly, a greater understanding of how flexible TRBP remains when bound to Dicer and dsRNA would allow possible roles to be considered or discounted. It is worth noting that several papers have suggested that Dicer and/or associated proteins undergo conformational shifts^{61,76} and that these complexes contain multiple dsRNA-binding sites.⁷⁷ This raises the possibility that the role of TRBP may be more complex than currently believed.

Evidence from Biochemical and Biophysical Studies

A number of *in vitro* studies have shown that the complex of Dicer with TRBP or PACT has a higher affinity for dsRNA than Dicer alone by several orders of magnitude.^{55,78,79} However, these studies report a more modest effect on small RNA processing rate, ranging from a fivefold increase in processing rate to a fivefold reduction. With the caveat that the reaction conditions may differ considerably from those present *in vivo*, this suggests TRBP and PACT fulfill role

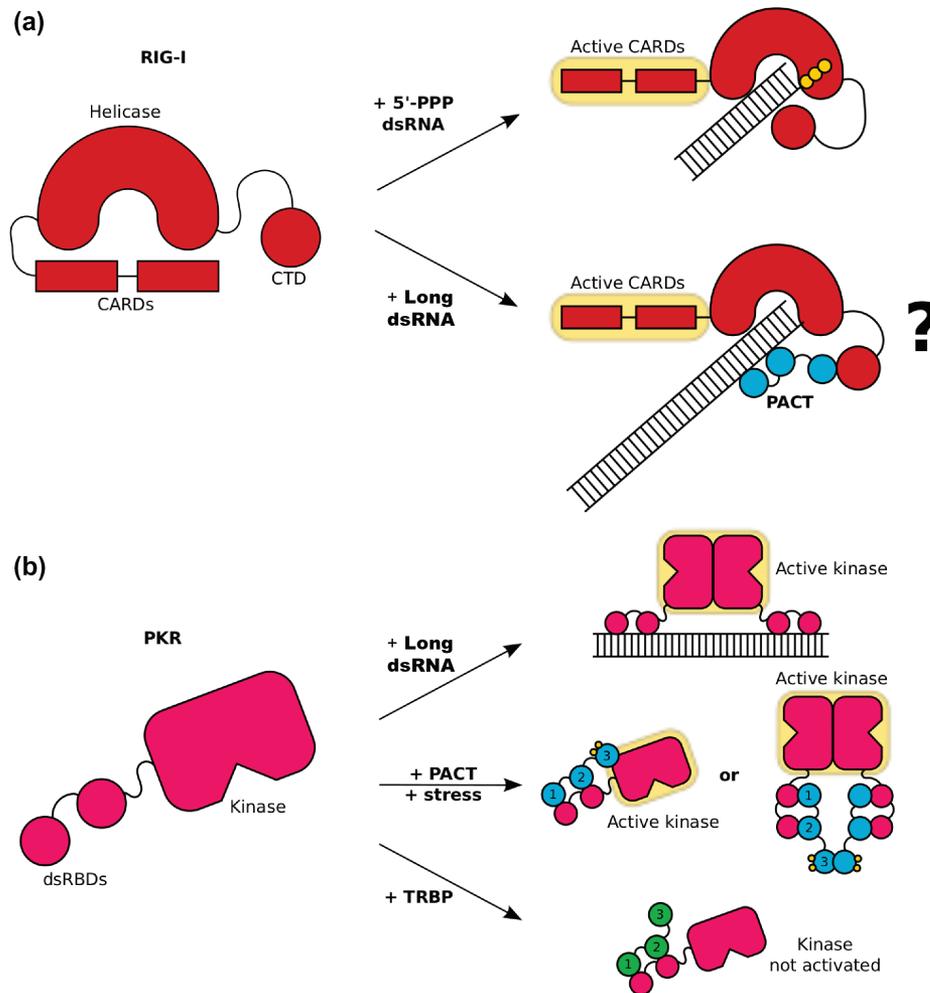


FIGURE 4 | (a) In the absence of appropriate ligands, retinoic acid-inducible gene I (RIG-I) has an inactive conformation, in which the helicase domain binds to the caspase recruitment domains (CARDs). 5'-Triphosphate double-stranded RNA (dsRNA) binds to the C-terminal domain (CTD) and helicase domain, which displaces the CARDs and results in signaling. It is less clear how Protein Activator of PKR (PACT) enables RIG-I activation: one possibility is that it increases RIG-I binding to additional ligands such as long dsRNA that lacks a 5'-triphosphate. (b) PKR can bind to long dsRNA through two N-terminal double-stranded RNA-binding domains (dsRBDs). This brings PKR molecules together to form dimers, which can then autophosphorylate and become active. PACT (when phosphorylated during cellular stress) can also activate PKR, although the mechanism is unclear. The two main hypotheses are: PACT-D3 contacts the kinase domain, somehow favoring activation, or PACT dimers can bind two molecules of PKR, promoting their dimerization and activation. TRBP acts as an inhibitor of PKR, either because its third domain cannot interact with PKR's kinase domain or because its third domain exhibits weaker dimerization.

1a above, but not 1b. Furthermore, TRBP has been reported to bind a pre-miRNA and the duplex Dicer product with similar affinities,^{22,55} which is inconsistent with a role in modulating Dicer processing rate due to preferential binding of substrate over product.

The measured affinity of TRBP (without Dicer) for short dsRNA has varied considerably between different studies.^{22,23,55} While this may simply represent variation in sample preparation and technique, it could also be due to the different RNAs studied, which would be consistent with TRBP and PACT showing a degree of specificity for different small RNAs (role 4).

Additional evidence comes from studies of reconstituted complexes of Dicer. Dicer/TRBP complexes show different affinities for different pre-miRNAs and siRNAs, although Dicer itself is at least partially responsible for this specificity.^{55,78} A complex containing Dicer and PACT showed considerably slower processing of short dsRNA substrates than a Dicer/TRBP complex, even though they seemed to have similar processing rates for pre-miRNAs.⁷⁹ These data potentially conflict with an earlier study, which concluded that both PACT and TRBP increase processing of long dsRNAs by Dicer.⁸⁰ This discrepancy could reflect

a genuine difference in how Dicer complexes process dsRNAs of different lengths; further experiments using matched reaction conditions will be required to resolve this issue.

Both *in vitro* and *in vivo*, cleavage of pre-miRNAs by Dicer can give a range of products with different lengths, termed iso-miRs.^{81,82} The size distribution of iso-miRs varies depending on the pre-miRNA. The Dicer/TRBP complex produces different length products compared with Dicer alone,^{82,83,78,79} supporting the idea that TRBP helps position RNA for cleavage by Dicer (role 2). PACT has not yet been observed to affect iso-miR distribution.

In *Drosophila*, current models suggest that the TRBP homolog R2D2 (see Box 2) plays a significant role in selecting which strand of the miRNA/siRNA is loaded into Argonaute. There is some evidence that this is also the case in humans. The thermodynamic stability of the miRNA or siRNA duplex produced

by Dicer is dependent on the nucleotide sequence. Experiments with photo-cross-linking dsRNA concluded that TRBP has some propensity to bind to the most thermodynamically stable end of the siRNA.⁴⁷ Furthermore, strand-specific crosslinking to the helicase domain of Dicer required TRBP or PACT and depended on thermodynamic stability, terminal nucleotides, and mismatched bases.⁷⁷ However, strand selectivity is not solely determined at this step: a study that measured which strand was loaded into RISC concluded that Ago2 alone had some strand selectivity (dependent on substrate), though TRBP and PACT could enhance this.⁴⁸

Lastly, it has recently been shown that TRBP can diffuse along dsRNA *in vitro*.⁸⁴ Although it is not yet clear what effect this would have on the Dicer complex, one-dimensional diffusion may allow the dsRNA to explore a greater range of positions and orientations within the complex. Together with the flexibility of TRBP, this could potentially facilitate RNA positioning for Dicer processing or RISC loading.

BOX 2

RNA INTERFERENCE IN *D. melanogaster*

Drosophila melanogaster has emerged as an important model system for research into RNAi, often revealing interactions and mechanisms that are later shown to occur in humans. Unlike mammals, *Drosophila* uses siRNA as an antiviral defense, termed exogenous (exo-) siRNA.⁸⁵ In addition, an endogenous (endo-) siRNA pathway exists to target mobile genetic elements.⁸⁶ In general, the miRNA pathway utilizes the proteins Dicer-1 and Ago1, while siRNA pathways rely on Dicer-2 and Ago2.⁸⁷

Several TRBP homologs have been identified in *Drosophila*, most notably the proteins R2D2 and Loquacious (Loqs). R2D2 forms a complex with Dicer-2 and assists strand-specific loading of siRNA into Ago2,^{88,89} while Loqs plays a role in processing of pre-miRNAs by Dicer-1.^{90,91} At least four isoforms of Loqs have been identified, which differentially process specific miRNAs, lead to altered iso-miR distributions, and even interact with different Dicer proteins.^{83,92,93} At the whole organism level, reduction of Loqs expression causes infertility owing to terminal differentiation of germ cells, while complete knockout is lethal.^{83,94,95} While TRBP and PACT share several properties with *Drosophila* dsRBPs,^{78,96} currently, there is no clear evidence for a similar delineation of function on the protein level in the mammalian system.

Evidence from Molecular Cell Biology Studies

The initial evidence that TRBP and PACT bound Dicer came from immunoprecipitation experiments using HEK293 cells.^{5,6,53} In one case, epitope-tagged Dicer was used to recover a complex containing TRBP and Ago2 that was able to process pre-let-7 to mature let-7, and cleave target mRNAs.⁵² The same studies used RNAi to investigate the function of TRBP and PACT, but gave somewhat contradictory results. However, one common feature is that knocking down any of the four components (Dicer, Ago2, TRBP, or PACT) resulted in a decrease in levels of mature miRNA.^{6,53,82}

There are some hints that PACT and TRBP may act on different sets of miRNAs and/or siRNAs (role 4). In one study, knocking down TRBP dramatically reduced the effectiveness of exogenous siRNA while having a small effect on levels of an inducible miRNA, whereas targeting PACT had the opposite effect.⁵³ It is currently unclear whether knockdown of PACT or TRBP differentially affects a certain subset of miRNAs.

It has been reported that TRBP contains multiple phosphorylation sites (Figure 1(c)), which modulate its stability. When TRBP with phospho-mimic mutations at these sites was transfected into human cells, miRNA production was shown to increase relative to controls.⁹⁷ The mechanism by which this occurs is unclear, as *in vitro* tests showed negligible differences in Dicer processing between complexes containing wild-type and phospho-mimic TRBP.⁵⁵ Transfection with phospho-mimic TRBP also caused a relative

downregulation of the let-7 miRNA family, hinting that TRBP may differentially affect at least some miRNAs (role 4).⁹⁷ Although PACT also undergoes phosphorylation at several sites, the effects have never been studied in the context of miRNA biogenesis. Post-translational modification of TRBP and PACT may indicate that their functions are context dependent, and are not constitutively required.

To determine whether TRBP and PACT aid Dicer processing (roles 1 and 2) or RISC loading (role 3), several studies have examined how depletion of these proteins affects silencing mediated by shRNA (which requires processing by Dicer) or duplex siRNA (which can in principle be incorporated directly into RISC). The consensus is that both are affected, implying an important role in RISC loading, while not ruling out a supporting contribution to Dicer processing.^{5,53,80} Although one of these studies reports siRNA-mediated silencing to be unaffected by depletion of TRBP or PACT, the effectiveness of TRBP/PACT knockdown was not measured.⁸⁰ It is therefore possible that TRBP and PACT were not effectively depleted, owing to saturation of the RNAi machinery with luciferase siRNA.

It is worth noting that it can be challenging to use RNAi to knock down proteins involved in small RNA processing, as the effectiveness of RNAi is itself dependent on the presence and activity of these proteins. The recent development of CRISPR/Cas genome editing may allow future studies to avoid this problem.⁹⁸ The interpretation of experimental data is also complicated by the possibility that these proteins may stabilize one another. Several studies have reported that knocking down TRBP also reduces Dicer levels, confounding interpretation in terms of TRBP alone,^{5,65,97} especially because this effect has not been unanimously reported.⁶ A similar stabilizing effect on Dicer was seen for PACT—but not for TRBP—in a separate study.⁵³

The majority of both TRBP and PACT is found in the cytoplasm, particularly in the perinuclear space, though a smaller amount is present in the nucleus.^{41,80,99} Förster resonance energy transfer experiments using fluorescently tagged TRBP and PACT provide evidence that they interact *in vivo*.⁸⁰ The distribution of TRBP and PACT mirrors that of Dicer and Ago2, which are also predominantly cytoplasmic with a small nuclear fraction. Dicer, Ago2, TRBP, and PACT have lower diffusion rates in the cytoplasm than in the nucleus.^{99,100} This observation has been interpreted as evidence for a large cytoplasmic complex that is disassembled in the nucleus. Alternatively, it is consistent with the idea that these proteins are anchored to a larger cytoplasmic feature,

such as P-bodies or the rough endoplasmic reticulum (ER). This latter possibility is consistent with reports of co-localization of RISC components with ribosomes or the ER.^{101–103}

TRBP AND PACT MEDIATE INNATE IMMUNE SURVEILLANCE OF dsRNA

RNA viruses produce dsRNA in the course of their lifecycle, either as genomic material or as a replication intermediate. Viral dsRNA has features that distinguish it from cellular dsRNA, which may include its longer length, lack of the 5' cap characteristic of eukaryotic mRNAs, and absence of mismatches. The innate immune system can recognize these features through PRRs, including RIG-I, MDA5, and TLR3.

The PRRs used by mammalian cells can be divided into 'early' responders, which stimulate production of interferons and proinflammatory cytokines, and 'late' responders, which directly block viral proliferation through translation inhibition and RNA degradation (Figure 1(b)). These pathways have been reviewed in more detail elsewhere.² TRBP and PACT have roles in several of these pathways, and have been implicated in the response to a number of viruses (summarized in Table 1).

PACT Can Stimulate RIG-I Activation

RIG-I is an early response PRR that detects dsRNA with a 5' triphosphate group (5'ppp-dsRNA).¹⁰⁴ RIG-I is ubiquitously expressed, and is one of the main viral RNA sensors in nonimmune cells.² The mechanism of RIG-I activation has been characterized at the structural level^{105–107} (Figure 4(a)). In the absence of 5'ppp-dsRNA, the tandem caspase recruitment domains (CARDs) responsible for signaling are bound to a DExD/H helicase domain, and are inactive. 5'ppp-dsRNA binds to the helicase and CTDs of RIG-I, displacing the CARDs. The CARDs then initiate a signaling pathway that leads to the activation of interferon regulatory factor 3 (IRF3), increasing transcription of type 1 interferon and other antiviral genes (Figure 1(c)). Several recent reviews have described RIG-I function in more detail.^{107–109}

Evidence has emerged that PACT can activate RIG-I in response to a number of viruses^{7,110,111} (Table 1). Although the helicase domains of RIG-I and Dicer are part of the same subfamily,¹¹² it is unlikely that PACT binds these two proteins in the same manner: there is no evident sequence homology in the helicase insert domain, which forms the PACT-binding site in Dicer. This is consistent with a report that

TABLE 1 | Evidence Linking TRBP and PACT to Viral Sensing.

Virus	Details	Reference
<i>Viruses linked to TRBP and PACT</i>		
Newcastle disease virus (NDV)	Overexpressing PACT increased type 1 interferon production from NDV infection	113
Sendai virus	RIG-I activation during infection was enhanced by PACT	7
Herpes simplex virus 1 (HSV-1)	HSV-1 protein Us11 reduces type 1 interferon production due to PACT. Viral proteins block PACT interaction with PKR and/or RIG-I	114,115
Ebolavirus	Viral proteins VP30 and VP35 interfere with RNAi and interact with TRBP, PACT, and/or Dicer. PACT overexpression inhibits viral replication, while VP35 prevents PACT from activating RIG-I	110,116
Influenza	Viral proteins interact with PACT, resulting in increased viral replication	117
MERS-CoV	Viral protein 4a inhibited PACT activation of RIG-I, resulting in reduced interferon production	111
HIV-1	TRBP blocks PKR activation by HIV transcripts. HIV was suggested to disrupt RNAi by sequestering TRBP, but this is disputed	118–120
<i>Viruses found not to depend on TRBP or PACT</i>		
Sindbis virus; encephalomyocarditis virus (EMCV)	Overexpressing PACT had no effect on type I interferon response	113
EMCV; vesicular stomatitis virus (VSV); Sendai virus	Found no difference between PACT knockout and wild-type cells	121

TRBP, TAR RNA-binding protein; PACT, Protein Activator of PKR; RIG-I, retinoic acid-inducible gene I; RNAi, RNA interference.

PACT binds to the CTD of RIG-I.⁷ There have been no reports of TRBP interacting with RIG-I.

Many unanswered questions surround the role of PACT in RIG-I activation. It is unclear under which conditions PACT activates RIG-I, and the mechanism is not known. To date, the interaction has been most studied in the context of Sendai virus,^{7,110} a negative sense single-stranded RNA (ssRNA) virus that infects rodents. In this case, PACT can act to enhance activation of IRF3 and transcription of genes under the control of IFN- β . A similar effect was seen for Ebola virus (also a negative sense ssRNA virus), and overexpressing PACT was able to slow viral replication.¹¹⁰ To counter this effect, Ebola virus encodes a protein that is reported to disrupt the interaction between PACT and both Dicer and RIG-I.^{110,116} An earlier study found that PACT increased type 1 interferon production in response to Newcastle disease virus, which also has a negative sense ssRNA genome.¹¹³ It now appears plausible that this effect was mediated by RIG-I, although it had not been identified at the time.

One hypothesis is that PACT can extend the range of ligands that can activate RIG-I. Overexpression of PACT did not enhance RIG-I activation by

5'-ppp dsRNA, but did enhance the response to the dsRNA analog poly-I:C.⁷ However, further work will be needed to confirm this, and to disentangle the effects of the other RIG-I (and Dicer)-related helicases, MDA5 and LGP2. MDA5 functions similarly to RIG-I, but is activated by longer dsRNA, while LGP2 lacks the CARD signaling domains, and is thought to inhibit RNA sensing.¹⁰⁸ The CTDs of RIG-I, MDA5, and LGP2 are well conserved and so it is plausible that PACT could interact with all of them.

PACT Activates PKR, While TRBP Inhibits It

PKR is a kinase that acts to block translation in response to viral dsRNA, or to other cellular stresses such as oxidative stress, accumulation of misfolded proteins in the ER, or external signals from cytokines and growth factors.¹²² It has low basal expression, but is strongly induced by interferon,¹²³ making it a 'late' responder to viruses.

PKR has two N-terminal type A dsRBDs and a kinase domain (Figure 4(b)). The first two domains bind dsRNA and act as a scaffold to bring PKR

molecules close together, and increase the likelihood of dimerization of the kinase domain.¹²⁴ After dimerization the kinase domain autophosphorylates,¹²⁵ then phosphorylates eukaryotic translation initiation factor 2 α (eIF2 α), in turn blocking translation. The kinase domain is present in a number of other proteins that phosphorylate eIF2 α in response to various stresses.¹²² PKR function has been reviewed in more detail elsewhere.^{122,124,126}

TRBP and PACT can regulate PKR: TRBP has an inhibitory effect, while PACT is a conditional activator^{3,4,127} (Figure 4(b)). Domain swap and mutational analyses indicate that the first two dsRBDs of PACT and TRBP are functionally interchangeable, and interact with the dsRBDs of PKR.^{9,36,128} There is some evidence that this interaction is independent of dsRNA, indicating that dsRBDs bind one another directly.¹²⁸ However, there is some disagreement on this issue,¹²⁹ and it has never been explicitly examined biophysically.

The differences between PACT and TRBP stem from their C-terminal dsRBD. In response to cellular stresses, PACT-D3 is phosphorylated at S246 and S287 by an unknown kinase (or kinases), leading to activation of PKR.¹²⁷ Two models have been proposed to explain how PACT enables PKR activation. First, binding of phosphorylated PACT-D3 to PKR could cause a conformational change that promotes activation.⁹ Supporting this model, PACT-D3 alone has been reported to interact weakly with PKR.^{9,130} Alternatively, PACT dimers might bind two molecules of PKR, enhancing PKR dimerization.⁴³ This is consistent with evidence that PACT-D3 phosphorylation promotes homodimerization and disfavors heterodimerization with TRBP.^{42,43}

In contrast, TRBP-D3 has an inhibitory effect on PKR, which appears to be important for preventing inappropriate activation of PKR.^{36,131} A similar inhibitory effect is also observed for truncated PKR or PACT constructs containing only their first two dsRBDs,^{9,132} which suggests that rather than a specific effect of TRBP-D3, anything that binds to the dsRBDs of PKR and disrupts PKR dimerization will cause inhibition. To our knowledge, TRBP phosphorylation (discussed in section *Evidence from Biochemical and Biophysical Studies*) has not been studied in the context of PKR inhibition.

It is worth noting that many studies prior to 2009 used PACT constructs containing a frameshift mutation that replaces the last 13 amino acids (including part of a predicted helix) with 5 unrelated amino acids.^{4,9,36,41,130,132–134} This mutant constitutively activated PKR and disrupted PACT-D3 dimerization.¹³¹

CONCLUSION

It is more than 15 years since TRBP and PACT were found to interact with PKR, and almost 10 years since their interaction with Dicer was uncovered. It is therefore somewhat surprising that (with a few exceptions^{116,120}) there has been so little crosstalk between these two areas of study. RNAi is a key defense against viruses in plants and invertebrates, and while this is no longer the case in mammals,¹³⁵ it is intriguing that small RNA biogenesis and viral sensing are still linked through PACT and TRBP. It remains to be seen whether this is an evolutionary accident, or whether these two proteins play a genuine linking role that has yet to be discovered. Interestingly, the helicase domains of Dicer and RIG-I are from the same family, termed RIG-I-like helicases, which also hints at an evolutionary or functional link between the two processes.^{112,136} A more mundane consequence is that results from *in vivo* experiments must always be interpreted with all the roles of TRBP and PACT in mind, even if the investigation is intended to focus on only one.

Although much has been learnt about the function of PACT and TRBP in miRNA biogenesis, two important questions remain unanswered: why does miRNA biogenesis utilize dsRBD-containing proteins, and why do vertebrates have two dsRBD-containing proteins which appear to be partially redundant? dsRBD-containing proteins interact with Dicer proteins in plants, insects, nematodes, and mammals, suggesting that there is some evolutionary pressure to conserve them. Results from *Drosophila* demonstrate that dsRBD-containing proteins may function at multiple steps in small RNA biogenesis pathways, rather than having a single role. While it is possible that this is also true for PACT and TRBP, it currently appears more likely that they function at the same step, potentially on different substrates. To work out which sequence or structural features lead to differential processing, it will likely be necessary to measure the affinities and processing rate of Dicer/TRBP and Dicer/PACT on a wider range of small RNAs. A complementary approach would be to examine on a genome-wide scale which miRNAs are differentially affected when TRBP or PACT are depleted from the cell.

Filling in the details of the role(s) of TRBP and PACT in miRNA biogenesis will help to unravel the diverse phenotypes associated with defective expression of these proteins and their homologs. These range from infertility and growth defects in mice and flies,^{94,96,137} to early onset dystonia/Parkinson's disease and cancer in humans.^{65,138} Biochemical and biophysical experiments must be complemented by cell and whole organism work to tease apart

these complex conditions. Although misregulation of particular microRNAs may play a large role, it is plausible that additional functions for TRBP and PACT will emerge.

Note Added in Press

We would like to alert the reader to two important studies into TRBP and PACT that were published after submission of this review: Kim and colleagues produced knock-outs of TRBP and/or PACT in HeLa

cells and found altered Dicer cleavage in a subset of pre-miRNAs, but no effect on steady-state miRNA levels or Dicer stability. They also showed hyperphosphorylation of TRBP by JNK during M phase. Wilson et al. reported the 3D structure of a fragment of Dicer in complex with TRBP domain 3. A Dicer mutant deficient for TRBP and PACT binding was shown to alter strand selection and iso-miR distribution for certain miRNAs. Please see Further Reading.

ACKNOWLEDGMENTS

The authors would like to acknowledge Herman Fung and Dr. Matthew Warner for critical reading of this manuscript. APH is a recipient of a PhD studentship from the Wellcome Trust-funded CIDCATS programme at the University of York (WT095024MA).

REFERENCES

1. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 2009, 10:126–139. doi: 10.1038/nrm2632.
2. Gantier MP, Williams BRG. The response of mammalian cells to double-stranded RNA. *Cytokine Growth Factor Rev* 2007, 18:363–371. doi: 10.1016/j.cytogfr.2007.06.016.
3. Park H, Davies MV, Langland JO, Chang HW, Nam YS, Tartaglia J, Paoletti E, Jacobs BL, Kaufman RJ, Venkatesan S. TAR RNA-binding protein is an inhibitor of the interferon-induced protein kinase PKR. *Proc Natl Acad Sci U S A* 1994, 91:4713–4717.
4. Patel RC, Sen GC. PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J* 1998, 17:4379–4390. doi: 10.1093/emboj/17.15.4379.
5. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 2005, 436:740–744. doi: 10.1038/nature03868.
6. Haase AD, Jaskiewicz L, Zhang H, Lainé S, Sack R, Gagnol A, Filipowicz W. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* 2005, 6:961–967. doi: 10.1038/sj.embor.7400509.
7. Kok K-H, Lui P-Y, Ng M-HJ, Siu K-L, Au SWN, Jin D-Y. The double-stranded RNA-binding protein PACT functions as a cellular activator of RIG-I to facilitate innate antiviral response. *Cell Host Microbe* 2011, 9:299–309. doi: 10.1016/j.chom.2011.03.007.
8. St Johnston D, Brown NH, Gall JG, Jantsch M. A conserved double-stranded RNA-binding domain. *Proc Natl Acad Sci U S A* 1992, 89:10979–10983.
9. Peters GA, Hartmann R, Qin J, Sen GC. Modular structure of PACT: distinct domains for binding and activating PKR. *Mol Cell Biol* 2001, 21:1908–1920. doi: 10.1128/MCB.21.6.1908-1920.2001.
10. Nanduri S, Carpick BW, Yang Y, Williams BR, Qin J. Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J* 1998, 17:5458–5465. doi: 10.1093/emboj/17.18.5458.
11. Stefl R, Xu M, Skrisovska L, Emeson RB, Allain FH-T. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* 2006, 14:345–355. doi: 10.1016/j.str.2005.11.013.
12. Sohn SY, Bae WJ, Kim JJ, Yeom K-H, Kim VN, Cho Y. Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* 2007, 14:847–853. doi: 10.1038/nsmb1294.
13. Marión RM, Fortes P, Beloso A, Dotti C, Ortín J. A human sequence homologue of Staufen is an RNA-binding protein that is associated with polysomes and localizes to the rough endoplasmic reticulum. *Mol Cell Biol* 1999, 19:2212–2219.
14. Chang K-Y, Ramos A. The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J* 2005, 272:2109–2117. doi: 10.1111/j.1742-4658.2005.04652.x.
15. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 2007, 8:479–490. doi: 10.1038/nrm2178.
16. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 1988, 16:10881–10890. doi: 10.1093/nar/16.22.10881.

17. Gouet P. ESPrpt/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 2003, 31:3320–3323. doi: 10.1093/nar/gkg556.
18. Masliah G, Barraud P, Allain FH-T. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci* 2013, 70:1875–1895. doi: 10.1007/s00018-012-1119-x.
19. Wu H, Henras A, Chanfreau G, Feigon J. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc Natl Acad Sci U S A* 2004, 101:8307–8312. doi: 10.1073/pnas.0402627101.
20. Chiliveri SC, Deshmukh MV. Structure of RDE-4 dsRBDs and mutational studies provide insights into dsRNA recognition in the *Caenorhabditis elegans* RNAi pathway. *Biochem J* 2014, 458:119–130. doi: 10.1042/BJ20131347.
21. Tabara H, Yigit E, Siomi H, Mello CC. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. *Cell* 2002, 109:861–871. doi: 10.1016/S0092-8674(02)00793-6.
22. Benoît MP, Imbert L, Palencia A, Pérard J, Ebel C, Boisbouvier J, Plevin MJ. The RNA-binding region of human TRBP interacts with microRNA precursors through two independent domains. *Nucleic Acids Res* 2013, 41:4241–4252. doi: 10.1093/nar/gkt086.
23. Yamashita S, Nagata T, Kawazoe M, Takemoto C, Kigawa T, Güntert P, Kobayashi N, Terada T, Shirouzu M, Wakiyama M, et al. Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci* 2011, 20:118–130. doi: 10.1002/pro.543.
24. Parker GS, Eckert DM, Bass BL. RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA. *RNA* 2006, 12:807–818. doi: 10.1261/rna.2338706.
25. Parker GS, Maity TS, Bass BL. dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *J Mol Biol* 2008, 384:967–979. doi: 10.1016/j.jmb.2008.10.002.
26. Takahashi T, Miyakawa T, Zenno S, Nishi K, Tanokura M, Ui-Tei K. Distinguishable in vitro binding mode of monomeric TRBP and dimeric PACT with siRNA. *PLoS One* 2013, 8:e63434. doi: 10.1371/journal.pone.0063434.
27. Gatignol A, Buckler-White A, Berkhout B, Jeang K. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science* 1991, 251:1597–1600. doi: 10.1126/science.2011739.
28. Dorin D, Bonnet MC, Bannwarth S, Gatignol A, Meurs EF, Vaquero C. The TAR RNA-binding protein, TRBP, stimulates the expression of TAR-containing RNAs in vitro and in vivo independently of its ability to inhibit the dsRNA-dependent kinase PKR. *J Biol Chem* 2003, 278:4440–4448. doi: 10.1074/jbc.M208954200.
29. Ricci EP, Kucukural A, Cenik C, Mercier BC, Singh G, Heyer EE, Ashar-Patel A, Peng L, Moore MJ. Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol* 2014, 21:26–35. doi: 10.1038/nsmb.2739.
30. Bass BL. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 2002, 71:817–846. doi: 10.1146/annurev.biochem.71.110601.135501.
31. Ramos A, Grünert S, Adams J, Micklem DR, Proctor MR, Freund S, Bycroft M, St Johnston D, Varani G. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J* 2000, 19:997–1009. doi: 10.1093/emboj/19.5.997.
32. Gleghorn ML, Gong C, Kielkopf CL, Maquat LE. Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat Struct Mol Biol* 2013, 20:515–524. doi: 10.1038/nsmb.2528.
33. Gleghorn ML, Maquat LE. “Black sheep” that don’t leave the double-stranded RNA-binding domain fold. *Trends Biochem Sci* 2014, 39:328–340. doi: 10.1016/j.tibs.2014.05.003.
34. MacRae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. In vitro reconstitution of the human RISC-loading complex. *Proc Natl Acad Sci U S A* 2008, 105:512–517. doi: 10.1073/pnas.0710869105.
35. Daniels SM, Melendez-Peña CE, Scarborough RJ, Daher A, Christensen HS, El Far M, Purcell DFJ, Lainé S, Gatignol A. Characterization of the TRBP domain required for dicer interaction and function in RNA interference. *BMC Mol Biol* 2009, 10:38. doi: 10.1186/1471-2199-10-38.
36. Gupta V, Huang X, Patel RC. The carboxy-terminal, M3 motifs of PACT and TRBP have opposite effects on PKR activity. *Virology* 2003, 315:283–291. doi: 10.1016/S0042-6822(03)00589-0.
37. Lee JY, Kim H, Ryu CH, Kim JY, Choi BH, Lim Y, Huh P-W, Kim Y-H, Lee K-H, Jun T-Y, et al. Merlin, a tumor suppressor, interacts with transactivation-responsive RNA-binding protein and inhibits its oncogenic activity. *J Biol Chem* 2004, 279:30265–30273. doi: 10.1074/jbc.M312083200.
38. Patel RC, Stanton P, McMillan NM, Williams BR, Sen GC. The interferon-inducible double-stranded RNA-activated protein kinase self-associates in vitro and in vivo. *Proc Natl Acad Sci* 1995, 92:8283–8287. doi: 10.1073/pnas.92.18.8283.
39. Poulsen H, Jorgensen R, Heding A, Nielsen FC, Bonven B, Egebjerg J. Dimerization of ADAR2 is mediated by the double-stranded RNA binding domain. *RNA* 2006, 12:1350–1360. doi: 10.1261/rna.2314406.
40. Zhang F, Romano PR, Nagamura-Inoue T, Tian B, Dever TE, Mathews MB, Ozato K, Hinnebusch AG. Binding of double-stranded RNA to protein

- kinase PKR is required for dimerization and promotes critical autophosphorylation events in the activation loop. *J Biol Chem* 2001, 276:24946–24958. doi: 10.1074/jbc.M102108200.
41. Laraki G, Clerzius G, Daher A, Melendez-Peña C, Daniels S, Gatignol A. Interactions between the double-stranded RNA-binding proteins TRBP and PACT define the Medipal domain that mediates protein-protein interactions. *RNA Biol* 2008, 5:92–103. doi: 10.4161/rna.5.2.6069.
 42. Singh M, Castillo D, Patel CV, Patel RC. Stress-induced phosphorylation of PACT reduces its interaction with TRBP and leads to PKR activation. *Biochemistry* 2011, 50:4550–4560. doi: 10.1021/bi200104h.
 43. Singh M, Patel RC. Increased interaction between PACT molecules in response to stress signals is required for PKR activation. *J Cell Biochem* 2012, 113:2754–2764. doi: 10.1002/jcb.24152.
 44. Daher A, Longuet M, Dorin D, Bois F, Segéral E, Bannwarth S, Battisti PL, Purcell DF, Benarous R, Vaquero C, et al. Two dimerization domains in the trans-activation response RNA-binding protein (TRBP) individually reverse the protein kinase R inhibition of HIV-1 long terminal repeat expression. *J Biol Chem* 2001, 276:33899–33905. doi: 10.1074/jbc.M103584200.
 45. Liu Q, Paroo Z. Biochemical principles of small RNA pathways. *Annu Rev Biochem* 2010, 79:295–319. doi: 10.1146/annurev.biochem.052208.151733.
 46. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell* 2009, 136:656–668. doi: 10.1016/j.cell.2009.01.045.
 47. Gredell JA, Dittmer MJ, Wu M, Chan C, Walton SP. Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry* 2010, 49:3148–3155. doi: 10.1021/bi902189s.
 48. Noland CL, Doudna JA. Multiple sensors ensure guide strand selection in human RNAi pathways. *RNA* 2013, 19:639–648. doi: 10.1261/rna.037424.112.
 49. Yang J-S, Phillips MD, Betel D, Mu P, Ventura A, Siepel AC, Chen KC, Lai EC. Widespread regulatory activity of vertebrate microRNA* species. *RNA* 2011, 17:312–326. doi: 10.1261/rna.2537911.
 50. Burroughs AM, Ando Y, de Hoon ML, Tomaru Y, Suzuki H, Hayashizaki Y, Daub CO. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol* 2011, 8:158–177. doi: 10.4161/rna.8.1.14300.
 51. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* 2004, 15:185–197. doi: 10.1016/j.molcel.2004.07.007.
 52. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 2005, 123:631–640. doi: 10.1016/j.cell.2005.10.022.
 53. Lee Y, Hur I, Park S-Y, Kim Y-K, Suh MR, Kim VN. The role of PACT in the RNA silencing pathway. *EMBO J* 2006, 25:522–532. doi: 10.1038/sj.emboj.7600942.
 54. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. Structural basis for double-stranded RNA processing by Dicer. *Science* 2006, 311:195–198. doi: 10.1126/science.1121638.
 55. Chakravarthy S, Sternberg SH, Kellenberger CA, Doudna JA. Substrate-specific kinetics of dicer-catalyzed RNA processing. *J Mol Biol* 2010, 404:392–402. doi: 10.1016/j.jmb.2010.09.030.
 56. Provost P. Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J* 2002, 21:5864–5874. doi: 10.1093/emboj/cdf578.
 57. Yoda M, Kawamata T, Paroo Z, Ye X, Iwasaki S, Liu Q, Tomari Y. ATP-dependent human RISC assembly pathways. *Nat Struct Mol Biol* 2010, 17:17–23. doi: 10.1038/nsmb.1733.
 58. Kawamata T, Seitz H, Tomari Y. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* 2009, 16:953–960. doi: 10.1038/nsmb.1630.
 59. Lau P-W, Potter CS, Carragher B, MacRae IJ. Structure of the human Dicer-TRBP complex by electron microscopy. *Structure* 2009, 17:1326–1332. doi: 10.1016/j.str.2009.08.013.
 60. Lau P-W, Guiley KZ, De N, Potter CS, Carragher B, MacRae IJ. The molecular architecture of human Dicer. *Nat Struct Mol Biol* 2012, 19:436–440. doi: 10.1038/nsmb.2268.
 61. Wang H-W, Noland C, Siridechadilok B, Taylor DW, Ma E, Felderer K, Doudna JA, Nogales E. Structural insights into RNA processing by the human RISC-loading complex. *Nat Struct Mol Biol* 2009, 16:1148–1153. doi: 10.1038/nsmb.1673.
 62. Robb GB, Rana TM. RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol Cell* 2007, 26:523–537. doi: 10.1016/j.molcel.2007.04.016.
 63. Pare JM, Tahbaz N, López-Orozco J, LaPointe P, Lasko P, Hobman TC. Hsp90 regulates the function of argonaute 2 and its recruitment to stress granules and P-bodies. *Mol Biol Cell* 2009, 20:3273–3284. doi: 10.1091/mbc.E09-01-0082.
 64. Gregory RI, Shiekhattar R. MicroRNA biogenesis and cancer. *Cancer Res* 2005, 65:3509–3512. doi: 10.1158/0008-5472.CAN-05-0298.
 65. Melo SA, Ropero S, Moutinho C, Aaltonen LA, Yamamoto H, Calin GA, Rossi S, Fernandez AF, Carneiro F, Oliveira C, et al. A TARBP2 mutation in human cancer impairs microRNA processing and

- DICER1 function. *Nat Genet* 2009, 41:365–370. doi: 10.1038/ng.317.
66. Fu X, Xue C, Huang Y, Xie Y, Li Y. The activity and expression of microRNAs in prostate cancers. *Mol Biosyst* 2010, 6:2561–2572. doi: 10.1039/c0mb00100g.
67. Caramuta S, Lee L, Ozata DM, Akçakaya P, Xie H, Höög A, Zedenius J, Bäckdahl M, Larsson C, Lui W-O. Clinical and functional impact of TARBP2 over-expression in adrenocortical carcinoma. *Endocr Relat Cancer* 2013, 20:551–564. doi: 10.1530/ERC-13-0098.
68. Lin X, Wu M, Liu P, Wei F, Li L, Tang H, Xie X, Liu X, Yang L, Xie X. Up-regulation and worse prognostic marker of cytoplasmic TARBP2 expression in obstinate breast cancer. *Med Oncol* 2014, 31:868. doi: 10.1007/s12032-014-0868-9.
69. Melo S, Villanueva A, Moutinho C, Davalos V, Spizzo R, Ivan C, Rossi S, Setien F, Casanovas O, Simo-Riudalbas L, et al. Small molecule enoxacin is a cancer-specific growth inhibitor that acts by enhancing TAR RNA-binding protein 2-mediated microRNA processing. *Proc Natl Acad Sci U S A* 2011, 108:4394–4399. doi: 10.1073/pnas.1014720108.
70. Sousa E, Graça I, Baptista T, Vieira FQ, Palmeira C, Henrique R, Jerónimo C. Enoxacin inhibits growth of prostate cancer cells and effectively restores microRNA processing. *Epigenetics* 2013, 8:548–558. doi: 10.4161/epi.24519.
71. Sand M, Skrygan M, Georgas D, Arenz C, Gambichler T, Sand D, Altmeyer P, Bechara FG. Expression levels of the microRNA maturing microprocessor complex component DGCR8 and the RNA-induced silencing complex (RISC) components argonaute-1, argonaute-2, PACT, TARBP1, and TARBP2 in epithelial skin cancer. *Mol Carcinog* 2012, 51:916–922. doi: 10.1002/mc.20861.
72. Chiosea S, Acquafondata M, Luo J, Kuan S, Seethala R. DICER1 and PRKRA in colon adenocarcinoma. *Biomark Insights* 2008, 3:253–258.
73. MacRae IJ, Zhou K, Doudna JA. Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* 2007, 14:934–940. doi: 10.1038/nsmb1293.
74. Ma E, MacRae IJ, Kirsch JF, Doudna JA. Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol* 2008, 380:237–243. doi: 10.1016/j.jmb.2008.05.005.
75. Zhang H. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J* 2002, 21:5875–5885. doi: 10.1093/emboj/cdf582.
76. Taylor DW, Ma E, Shigematsu H, Cianfrocco MA, Noland CL, Nagayama K, Nogales E, Doudna JA, Wang H-W. Substrate-specific structural rearrangements of human Dicer. *Nat Struct Mol Biol* 2013, 20:9–11. doi: 10.1038/nsmb.2564.
77. Noland CL, Ma E, Doudna JA. siRNA repositioning for guide strand selection by human Dicer complexes. *Mol Cell* 2011, 43:110–121. doi: 10.1016/j.molcel.2011.05.028.
78. Lee HY, Doudna JA. TRBP alters human precursor microRNA processing in vitro. *RNA* 2012, 18:2012–2019. doi: 10.1261/rna.035501.112.
79. Lee HY, Zhou K, Smith AM, Noland CL, Doudna JA. Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic Acids Res* 2013, 41:6568–6576. doi: 10.1093/nar/gkt361.
80. Kok K-H, Ng M-HJ, Ching Y-P, Jin D-Y. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J Biol Chem* 2007, 282:17649–17657. doi: 10.1074/jbc.M611768200.
81. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs—the overlooked repertoire in the dynamic microRNAome. *Trends Genet* 2012, 28:544–549. doi: 10.1016/j.tig.2012.07.005.
82. Koscińska E, Starega-Roslan J, Krzyzosiak WJ. The role of Dicer protein partners in the processing of microRNA precursors. *PLoS One* 2011, 6:e28548. doi: 10.1371/journal.pone.0028548.
83. Fukunaga R, Han BW, Hung J-H, Xu J, Weng Z, Zamore PD. Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell* 2012, 151:533–546. doi: 10.1016/j.cell.2012.09.027.
84. Koh HR, Kidwell MA, Ragunathan K, Doudna JA, Myong S. ATP-independent diffusion of double-stranded RNA binding proteins. *Proc Natl Acad Sci U S A* 2013, 110:151–156. doi: 10.1073/pnas.1212917110.
85. Wang X-H, Aliyari R, Li W-X, Li H-W, Kim K, Carthew R, Atkinson P, Ding S-W. RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* 2006, 312:452–454. doi: 10.1126/science.1125694.
86. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler ELW, Zapp ML, Weng Z, et al. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 2008, 320:1077–1081. doi: 10.1126/science.1157396.
87. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 2004, 117:69–81.
88. Liu Q, Rand TA, Kalidas S, Du F, Kim H-E, Smith DP, Wang X. R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* 2003, 301:1921–1925. doi: 10.1126/science.1088710.
89. Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD. A protein sensor for siRNA asymmetry.

- Science* 2004, 306:1377–1380. doi: 10.1126/science.1102755.
90. Saito K, Ishizuka A, Siomi H, Siomi MC. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol* 2005, 3:e235. doi: 10.1371/journal.pbio.0030235.
 91. Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev* 2005, 19:1674–1679. doi: 10.1101/gad.1334005.
 92. Hartig JV, Esslinger S, Böttcher R, Saito K, Förstemann K. Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J* 2009, 28:2932–2944. doi: 10.1038/emboj.2009.220.
 93. Marques JT, Kim K, Wu P-H, Alleyne TM, Jafari N, Carthew RW. Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nat Struct Mol Biol* 2010, 17:24–30. doi: 10.1038/nsmb.1735.
 94. Förstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* 2005, 3:e236. doi: 10.1371/journal.pbio.0030236.
 95. Park JK, Liu X, Strauss TJ, McKearin DM, Liu Q. The miRNA pathway intrinsically controls self-renewal of *Drosophila* germline stem cells. *Curr Biol* 2007, 17:533–538. doi: 10.1016/j.cub.2007.01.060.
 96. Bennett RL, Blalock WL, Choi E-J, Lee YJ, Zhang Y, Zhou L, Oh SP, May WS. RAX is required for fly neuronal development and mouse embryogenesis. *Mech Dev* 2008, 125:777–785. doi: 10.1016/j.mod.2008.06.009.
 97. Paroo Z, Ye X, Chen S, Liu Q. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell* 2009, 139:112–122. doi: 10.1016/j.cell.2009.06.044.
 98. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013, 339:819–823. doi: 10.1126/science.1231143.
 99. Ohrt T, Muetze J, Svoboda P, Schwille P. Intracellular localization and routing of miRNA and RNAi pathway components. *Curr Top Med Chem* 2012, 12:79–88. doi: 10.2174/156802612798919132.
 100. Ohrt T, Mütze J, Staroske W, Weinmann L, Höck J, Crell K, Meister G, Schwille P. Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res* 2008, 36:6439–6449. doi: 10.1093/nar/gkn693.
 101. Chendrimada TP, Finn KJ, Ji X, Baillat D, Gregory RI, Liebhaber SA, Pasquinelli AE, Shiekhattar R. MicroRNA silencing through RISC recruitment of eIF6. *Nature* 2007, 447:823–828. doi: 10.1038/nature05841.
 102. Stalder L, Heusermann W, Sokol L, Trojer D, Wirz J, Hean J, Fritzsche A, Aeschmann F, Pfanzagl V, Basselet P, et al. The rough endoplasmic reticulum is a central nucleation site of siRNA-mediated RNA silencing. *EMBO J* 2013, 32:1115–1127. doi: 10.1038/emboj.2013.52.
 103. Kim YJ, Maizel A, Chen X. Traffic into silence: endomembranes and post-transcriptional RNA silencing. *EMBO J* 2014, 33:968–980. doi: 10.1002/emboj.201387262.
 104. Schlee M, Hartmann G. The chase for the RIG-I ligand—recent advances. *Mol Ther* 2010, 18:1254–1262. doi: 10.1038/mt.2010.90.
 105. Kowalinski E, Lunardi T, McCarthy AA, Loubser J, Brunel J, Grigorov B, Gerlier D, Cusack S. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 2011, 147:423–435. doi: 10.1016/j.cell.2011.09.039.
 106. Luo D, Ding SC, Vela A, Kohlway A, Lindenbach BD, Pyle AM. Structural insights into RNA recognition by RIG-I. *Cell* 2011, 147:409–422. doi: 10.1016/j.cell.2011.09.023.
 107. Kolakofsky D, Kowalinski E, Cusack S. A structure-based model of RIG-I activation. *RNA* 2012, 18:2118–2127. doi: 10.1261/rna.035949.112.
 108. Yoneyama M, Fujita T. RNA recognition and signal transduction by RIG-I-like receptors. *Immunol Rev* 2009, 227:54–65. doi: 10.1111/j.1600-065X.2008.00727.x.
 109. Takeuchi O, Akira S. MDA5/RIG-I and virus recognition. *Curr Opin Immunol* 2008, 20:17–22. doi: 10.1016/j.coi.2008.01.002.
 110. Luthra P, Ramanan P, Mire CE, Weisend C, Tsuda Y, Yen B, Liu G, Leung DW, Geisbert TW, Ebihara H, et al. Mutual antagonism between the Ebola virus VP35 protein and the RIG-I activator PACT determines infection outcome. *Cell Host Microbe* 2013, 14:74–84. doi: 10.1016/j.chom.2013.06.010.
 111. Siu K-L, Yeung ML, Kok K-H, Yuen K-S, Kew C, Lui P-Y, Chan C-P, Tse H, Woo PCY, Yuen K-Y, et al. Middle east respiratory syndrome coronavirus 4a protein is a double-stranded RNA-binding protein that suppresses PACT-induced activation of RIG-I and MDA5 in the innate antiviral response. *J Virol* 2014, 88:4866–4876. doi: 10.1128/JVI.03649-13.
 112. Fairman-Williams ME, Guenther U-P, Jankowsky E. SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* 2010, 20:313–324. doi: 10.1016/j.sbi.2010.03.011.
 113. Iwamura T, Yoneyama M, Koizumi N, Okabe Y, Namiki H, Samuel CE, Fujita T. PACT, a double-stranded RNA binding protein acts as a positive regulator for type I interferon gene induced by

- Newcastle disease virus. *Biochem Biophys Res Commun* 2001, 282:515–523. doi: 10.1006/bbrc.2001.4606.
114. Peters GA, Khoo D, Mohr I, Sen GC. Inhibition of PACT-mediated activation of PKR by the herpes simplex virus type 1 Us11 protein. *J Virol* 2002, 76:11054–11064. doi: 10.1128/JVI.76.21.11054-11064.2002.
115. Kew C, Lui P-Y, Chan C-P, Liu X, Au SWN, Mohr I, Jin D-Y, Kok K-H. Suppression of PACT-induced type I interferon production by herpes simplex virus 1 Us11 protein. *J Virol* 2013, 87:13141–13149. doi: 10.1128/JVI.02564-13.
116. Fabozzi G, Nabel CS, Dolan MA, Sullivan NJ. Ebola virus proteins suppress the effects of small interfering RNA by direct interaction with the mammalian RNA interference pathway. *J Virol* 2011, 85:2512–2523. doi: 10.1128/JVI.01160-10.
117. Tafforeau L, Chantier T, Pradezynski F, Pellet J, Mangot PE, Vidalain P-O, Andre P, Rabourdin-Combe C, Lotteau V. Generation and comprehensive analysis of an influenza virus polymerase cellular interaction network. *J Virol* 2011, 85:13010–13018. doi: 10.1128/JVI.02651-10.
118. Bennasser Y, Yeung ML, Jeang K-T. HIV-1 TAR RNA subverts RNA interference in transfected cells through sequestration of TAR RNA-binding protein, TRBP. *J Biol Chem* 2006, 281:27674–27678. doi: 10.1074/jbc.C600072200.
119. Sanghvi VR, Steel LF. The cellular TAR RNA binding protein, TRBP, promotes HIV-1 replication primarily by inhibiting the activation of double-stranded RNA-dependent kinase PKR. *J Virol* 2011, 85:12614–12621. doi: 10.1128/JVI.05240-11.
120. Sanghvi VR, Steel LF. A re-examination of global suppression of RNA interference by HIV-1. *PLoS One* 2011, 6:e17246. doi: 10.1371/journal.pone.0017246.
121. Marques JT, White CL, Peters GA, Williams BRG, Sen GC. The role of PACT in mediating gene induction, PKR activation, and apoptosis in response to diverse stimuli. *J Interferon Cytokine Res* 2008, 28:469–476. doi: 10.1089/jir.2007.0006.
122. Donnelly N, Gorman AM, Gupta S, Samali A. The eIF2 α kinases: their structures and functions. *Cell Mol Life Sci* 2013, 70:3493–3511. doi: 10.1007/s00018-012-1252-6.
123. Meurs E, Chong K, Galabru J, Thomas NS, Kerr IM, Williams BR, Hovanessian AG. Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell* 1990, 62:379–390.
124. Cole JL. Activation of PKR: an open and shut case? *Trends Biochem Sci* 2007, 32:57–62. doi: 10.1016/j.tibs.2006.12.003.
125. Vattem KM, Staschke KA, Wek RC. Mechanism of activation of the double-stranded-RNA-dependent protein kinase, PKR. *Eur J Biochem* 2001, 268:3674–3684. doi: 10.1046/j.1432-1327.2001.02273.x.
126. García MA, Meurs EF, Esteban M. The dsRNA protein kinase PKR: virus and cell control. *Biochimie* 2007, 89:799–811. doi: 10.1016/j.biochi.2007.03.001.
127. Peters GA, Li S, Sen GC. Phosphorylation of specific serine residues in the PKR activation domain of PACT is essential for its ability to mediate apoptosis. *J Biol Chem* 2006, 281:35129–35136. doi: 10.1074/jbc.M607714200.
128. Benkirane M, Neuveut C, Chun RF, Smith SM, Samuel CE, Gatignol A, Jeang KT. Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR. *EMBO J* 1997, 16:611–624. doi: 10.1093/emboj/16.3.611.
129. Cosentino GP, Venkatesan S, Serluca FC, Green SR, Mathews MB, Sonenberg N. Double-stranded-RNA-dependent protein kinase and TAR RNA-binding protein form homo- and heterodimers in vivo. *Proc Natl Acad Sci U S A* 1995, 92:9445–9449.
130. Li S, Peters GA, Ding K, Zhang X, Qin J, Sen GC. Molecular basis for PKR activation by PACT or dsRNA. *Proc Natl Acad Sci U S A* 2006, 103:10005–10010. doi: 10.1073/pnas.0602317103.
131. Daher A, Laraki G, Singh M, Melendez-Peña CE, Bannwarth S, Peters AH, Meurs EF, Braun RE, Patel RC, Gatignol A. TRBP control of PACT-induced phosphorylation of protein kinase R is reversed by stress. *Mol Cell Biol* 2009, 29:254–265. doi: 10.1128/MCB.01030-08.
132. Huang X, Hutchins B, Patel RC. The C-terminal, third conserved motif of the protein activator PACT plays an essential role in the activation of double-stranded-RNA-dependent protein kinase (PKR). *Biochem J* 2002, 366:175–186. doi: 10.1042/BJ20020204.
133. Li S, Sen GC. PACT-mediated enhancement of reporter gene expression at the translational level. *J Interferon Cytokine Res* 2003, 23:689–697. doi: 10.1089/107999003772084806.
134. Peters GA, Dickerman B, Sen GC. Biochemical analysis of PKR activation by PACT. *Biochemistry* 2009, 48:7441–7447. doi: 10.1021/bi900433y.
135. Cullen BR. Is RNA interference involved in intrinsic antiviral immunity in mammals? *Nat Immunol* 2006, 7:563–567. doi: 10.1038/ni1352.
136. Ulvila J, Hultmark D, Rämét M. RNA silencing in the antiviral innate immune defence—role of DEAD-box RNA helicases. *Scand J Immunol* 2010, 71:146–158. doi: 10.1111/j.1365-3083.2009.02362.x.
137. Rowe TM, Rizzi M, Hirose K, Peters GA, Sen GC. A role of the double-stranded RNA-binding protein

- PACT in mouse ear development and hearing. *Proc Natl Acad Sci U S A* 2006, 103:5823–5828. doi: 10.1073/pnas.0601287103.
138. Camargos S, Scholz S, Simón-Sánchez J, Paisán-Ruiz C, Lewis P, Hernandez D, Ding J, Gibbs JR, Cookson MR, Bras J, et al. DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA. *Lancet Neurol* 2008, 7:207–215. doi: 10.1016/S1474-4422(08)70022-X.

FURTHER READING

Kim Y, Yeo J, Lee JH, Cho J, Seo D, Kim J-S, Kim VN. Deletion of human tarbp2 reveals cellular microRNA targets and cell-cycle function of TRBP. *Cell Rep* 2014, 9:1061–1074.

Wilson RC, Tambe A, Kidwell MA, Noland CL, Schneider CP, Doudna JA. Dicer-TRBP complex formation ensures accurate mammalian MicroRNA biogenesis. *Mol Cell* 2015. doi: 10.1016/j.molcel.2014.11.030.

Abbreviations

5-ppp	5' triphosphate
Ago	Argonaute
AUC	Analytical Ultracentrifugation
dsRNA	Double-stranded RNA
DSF	Differential Scanning Fluorimetry
DSS	4,4-dimethyl-4-silapentane-1-sulphonic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EXSY	Exchange Spectroscopy
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear Single Quantum Coherence
MES	2-(N-morpholino)ethanesulfonic acid
miRNA	microRNA
mRNA	Messenger RNA
MST	Microscale Thermophoresis
NCBI	National Centre for Biotechnology Information
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
pre-miRNA	Precursor microRNA
PRR	Pattern Recognition Receptor
RNA	Ribonucleic Acid
SEC	Size Exclusion Chromatography
SEC-MALLS	Size Exclusion Chromatography coupled to Multi-Angle Laser Light Scattering
SOFAST-HMQC	Band-Selective Flip-Angle-Optimised Short-Transient Heteronuclear Multiple Quantum Coherence
TBS/T	Tris/Tween wash buffer – see page 88 for recipe
TCEP	Tris(2-carboxyethyl)phosphine

References

- 3rd, A. P., Kroenke, C. and Loria, J. (2001) 'Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules', *Methods Enzymol.*, 339(1997), pp. 204–238.
- Acevedo, R. *et al.* (2015) 'Helical Defects in MicroRNA Influence Protein Binding by TAR RNA Binding Protein', *PLoS One*. Edited by F. Kashanchi, 10(1), p. e0116749.
- Acevedo, R. *et al.* (2016) 'Binding by TRBP-dsRBD2 Does Not Induce Bending of Double-Stranded RNA', *Biophys. J.*, 110(12), pp. 2610–2617.
- Andre, I. *et al.* (2008) 'Emergence of symmetry in homooligomeric biological assemblies', *Proc. Natl. Acad. Sci.*, 105(42), pp. 16148–16152.
- Asakura, T. *et al.* (1995) 'The relationship between amide proton chemical shifts and secondary structure in proteins.', *J. Biomol. NMR*, 6(3), pp. 227–36.
- Backes, S. *et al.* (2014) 'The Mammalian Response to Virus Infection Is Independent of Small RNA Silencing', *Cell Rep.*, 8(1), pp. 114–125.
- Barrientos, L. G. *et al.* (2002) 'The domain-swapped dimer of cyanovirin-N is in a metastable folded state: Reconciliation of X-ray and NMR structures', *Structure*, 10(5), pp. 673–686.
- Bass, B. L. (2002) 'RNA editing by adenosine deaminases that act on RNA.', *Annu. Rev. Biochem.*, 71(1), pp. 817–46.
- Behm-Ansmant, I. (2006) 'mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes', *Genes Dev.*, 20(14), pp. 1885–1898.
- Benitez, A. A. *et al.* (2015) 'Engineered Mammalian RNAi Can Elicit Antiviral Protection that Negates the Requirement for the Interferon Response', *Cell Rep.*, 13(7), pp. 1456–1466.
- Benkirane, M. *et al.* (1997) 'Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR.', *EMBO J.*, 16(3), pp. 611–24.
- Bennasser, Y., Yeung, M. L. and Jeang, K.-T. (2006) 'HIV-1 TAR RNA subverts RNA interference in transfected cells through sequestration of TAR RNA-binding protein, TRBP.', *J. Biol. Chem.*, 281(38), pp. 27674–8.
- Bennet, M. J., Schlunegger, M. P. and Eisenberg, D. (1995) '3D Domain swapping: A mechanism for oligomeric assembly', *Protein Sci.*, 4, pp. 2455–2468.
- Bennett, R. L. *et al.* (2008) 'RAX is required for fly neuronal development and mouse embryogenesis.', *Mech. Dev.*, 125(9–10), pp. 777–85.
- Benoit, M. P. M. H. *et al.* (2013) 'The RNA-binding region of human TRBP interacts with microRNA precursors through two independent domains.', *Nucleic Acids Res.*, 41(7), pp. 4241–52.
- Benoit, M. P. M. H. and Plevin, M. J. (2013) 'Backbone resonance assignments of the micro-RNA precursor binding region of human TRBP.', *Biomol. NMR Assign.*, 7(2), pp. 229–33.

- Berjanskii, M. V. and Wishart, D. S. (2005) 'A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts', *J. Am. Chem. Soc.*, 127(43), pp. 14970–14971.
- Bernstein, E. *et al.* (2001) 'Role for a bidentate ribonuclease in the initiation step of RNA interference.', *Nature*, 409(6818), pp. 363–6.
- Bodenhausen, G. and Ruben, D. J. (1980) 'Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy', *Chem. Phys. Lett.*, 69(1), pp. 185–189.
- Brautigam, C. A. (2011) 'Using Lamm-Equation modeling of sedimentation velocity data to determine the kinetic and thermodynamic properties of macromolecular interactions', *Methods*, 54(1), pp. 4–15.
- Brennecke, J. *et al.* (2005) 'Principles of MicroRNA–Target Recognition', *PLoS Biol.* Edited by James C. Carrington, 3(3), p. e85.
- Brown, J. H. *et al.* (2001) 'Deciphering the design of the tropomyosin molecule', *Proc. Natl. Acad. Sci.*, 98(15), pp. 8496–8501.
- Brown, J. H. (2006) 'Breaking symmetry in protein dimers: designs and functions.', *Protein Sci.*, 15(1), pp. 1–13.
- Bruns, A. M. *et al.* (2014) 'The innate immune sensor LGP2 activates antiviral signaling by regulating MDA5-RNA interaction and filament assembly.', *Mol. Cell*, 55(5), pp. 771–81.
- Burroughs, A. M. *et al.* (2011) 'Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin', *RNA Biol.*, 8(1), pp. 158–177.
- Camargos, S. *et al.* (2008) 'DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA.', *Lancet Neurol.*, 7(3), pp. 207–15.
- Camargos, S. *et al.* (2012) 'DYT16: the original cases.', *J. Neurol. Neurosurg. Psychiatry*, 83(10), pp. 1012–4.
- Caramuta, S. *et al.* (2013) 'Clinical and functional impact of TARBP2 over-expression in adrenocortical carcinoma.', *Endocr. Relat. Cancer*, 20(4), pp. 551–64.
- de Carvalho Aguiar, P. *et al.* (2015) 'Novel compound heterozygous mutations in PRKRA cause pure dystonia', *Mov. Disord.*, 30(6), pp. 877–878.
- Cavanagh, J. *et al.* (1991) 'Sensitivity improvement in proton-detected 2-dimensional heteronuclear relay spectroscopy', *J. Magn. Reson.*, 91, pp. 429–436.
- Cavanagh, J. *et al.* (2006) *Protein NMR Spectroscopy: Principles and Practice*. 2nd edn. Elsevier Academic Press.
- Chakravarthy, S. *et al.* (2010) 'Substrate-Specific Kinetics of Dicer-Catalyzed RNA Processing', *J. Mol. Biol.*, 404(3), pp. 392–402.
- Chang, K.-Y. and Ramos, A. (2005) 'The double-stranded RNA-binding motif, a versatile macromolecular docking platform.', *FEBS J.*, 272(9), pp. 2109–17.
- Chen, C. *et al.* (2015) 'SUMOylation of TARBP2 regulates miRNA/siRNA efficiency', *Nat. Commun.*, 6, p. 8899.
- Chendrimada, T. P. *et al.* (2005) 'TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing', *Nature*, 436(7051), pp. 740–4.

- Chendrimada, T. P. *et al.* (2007) 'MicroRNA silencing through RISC recruitment of eIF6.', *Nature*, 447(7146), pp. 823–8.
- Chiliveri, S. C. and Deshmukh, M. V (2014) 'Structure of RDE-4 dsRBDs and mutational studies provide insights into dsRNA recognition in the *Caenorhabditis elegans* RNAi pathway.', *Biochem. J.*, 458(1), pp. 119–30.
- Chiosea, S. *et al.* (2008) 'DICER1 and PRKRA in Colon Adenocarcinoma.', *Biomark. Insights*, 3, pp. 253–258.
- Cole, J. L. (2007) 'Activation of PKR: an open and shut case?', *Trends Biochem. Sci.*, 32(2), pp. 57–62.
- Corpet, F. (1988) 'Multiple sequence alignment with hierarchical clustering', *Nucleic Acids Res.*, 16(22), pp. 10881–10890.
- Cosentino, G. P. *et al.* (1995) 'Double-stranded-RNA-dependent protein kinase and TAR RNA-binding protein form homo- and heterodimers in vivo.', *Proc. Natl. Acad. Sci. U. S. A.*, 92(21), pp. 9445–9.
- Cuddihy, A. R. *et al.* (1999) 'The double-stranded RNA activated protein kinase PKR physically associates with the tumor suppressor p53 protein and phosphorylates human p53 on serine 392 in vitro', *Oncogene*, 18(17), pp. 2690–2702.
- Cullen, B. R. (2006) 'Is RNA interference involved in intrinsic antiviral immunity in mammals?', *Nat. Immunol.*, 7(6), pp. 563–7.
- Daher, A. *et al.* (2009) 'TRBP control of PACT-induced phosphorylation of protein kinase R is reversed by stress.', *Mol. Cell. Biol.*, 29(1), pp. 254–65.
- Daher, a *et al.* (2001) 'Two dimerization domains in the trans-activation response RNA-binding protein (TRBP) individually reverse the protein kinase R inhibition of HIV-1 long terminal repeat expression.', *J. Biol. Chem.*, 276(36), pp. 33899–905.
- Daniels, S. M. *et al.* (2009) 'Characterization of the TRBP domain required for dicer interaction and function in RNA interference.', *BMC Mol. Biol.*, 10, p. 38.
- Delaglio, F. *et al.* (1995) 'NMRPipe: a multidimensional spectral processing system based on UNIX pipes.', *J. Biomol. NMR*, 6(3), pp. 277–93.
- Dickerman, B. K. *et al.* (2011) 'Missense mutation in the second RNA binding domain reveals a role for Prkra (PACT/RAX) during skull development.', *PLoS One*, 6(12), p. e28537.
- Dickerman, B. K. *et al.* (2015) 'The protein activator of protein kinase R, PACT/RAX, negatively regulates protein kinase R during mouse anterior pituitary development.', *FEBS J.*
- Ding, J. *et al.* (2015) 'Trbp regulates heart function through microRNA-mediated Sox6 repression', *Nat. Genet.*, 47(7), pp. 776–783.
- Ding, S.-W. (2010) 'RNA-based antiviral immunity.', *Nat. Rev. Immunol.*, 10(9), pp. 632–44.
- Donnelly, N. *et al.* (2013) 'The eIF2 α kinases: their structures and functions.', *Cell. Mol. Life Sci.*, 70(19), pp. 3493–511.
- Dorin, D. *et al.* (2003) 'The TAR RNA-binding protein, TRBP, stimulates the expression of TAR-containing RNAs in vitro and in vivo independently of its ability to inhibit the dsRNA-dependent kinase PKR.', *J. Biol. Chem.*, 278(7), pp. 4440–8.

- Eichhorn, S. W. *et al.* (2014) 'mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues', *Mol. Cell*, 56(1), pp. 104–115.
- Eulalio, A., Tritschler, F. and Izaurralde, E. (2009) 'The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing.', *RNA*, 15(8), pp. 1433–42.
- Fabozzi, G. *et al.* (2011) 'Ebola virus proteins suppress the effects of small interfering RNA by direct interaction with the mammalian RNA interference pathway.', *J. Virol.*, 85(6), pp. 2512–23.
- Fairman-Williams, M. E., Guenther, U.-P. and Jankowsky, E. (2010) 'SF1 and SF2 helicases: family matters.', *Curr. Opin. Struct. Biol.*, 20(3), pp. 313–24.
- Fareh, M. *et al.* (2015) 'Single-molecule pull-down for investigating protein-nucleic acid interactions', *Methods*.
- Fareh, M. *et al.* (2016) 'TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments', *Nat. Commun.*, 7, p. 13694.
- Farrow, N. A., Zhang, O., *et al.* (1994) 'A heteronuclear correlation experiment for simultaneous determination of ¹⁵N longitudinal decay and chemical exchange rates of systems in slow equilibrium.', *J. Biomol. NMR*, 4(5), pp. 727–34.
- Farrow, N. A., Muhandiram, R., *et al.* (1994) 'Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation.', *Biochemistry*, 33(19), pp. 5984–6003.
- Flemer, M. *et al.* (2013) 'A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes.', *Cell*, 155(4), pp. 807–16.
- Franzosa, E. A., Lynagh, K. J. and Xia, Y. (2010) 'Structural Correlates of Protein Melting Temperature', *Proc. 4th Int. Beilstein Symp. Exp. Stand. Cond. Enzym. Charact.*, pp. 99–106.
- Fu, X. *et al.* (2010) 'The activity and expression of microRNAs in prostate cancers.', *Mol. Biosyst.*, 6(12), pp. 2561–72.
- Fukunaga, R. *et al.* (2012) 'Dicer partner proteins tune the length of mature miRNAs in flies and mammals.', *Cell*, 151(3), pp. 533–46.
- Gack, M. U. *et al.* (2007) 'TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity', *Nature*, 446(7138), pp. 916–920.
- Gantier, M. P. and Williams, B. R. G. (2007) 'The response of mammalian cells to double-stranded RNA.', *Cytokine Growth Factor Rev.*, 18(5–6), pp. 363–71.
- García de la Torre, J., Huertas, M. . and Carrasco, B. (2000) 'HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations', *J. Magn. Reson.*, 147(1), pp. 138–146.
- Gatignol, A. *et al.* (1991) 'Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR', *Science (80-.)*, 251(5001), pp. 1597–1600.
- Ghasriani, H. *et al.* (2014) 'Micelle-Catalyzed Domain Swapping in the GlpG Rhomboid Protease Cytoplasmic Domain.', *Biochemistry*, 53(37), pp. 5907–15.
- Gleghorn, M. L. *et al.* (2013) 'Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay.', *Nat. Struct. Mol.*

Biol., 20(4), pp. 515–24.

Gleghorn, M. L. and Maquat, L. E. (2014) “Black sheep” that don’t leave the double-stranded RNA-binding domain fold’, *Trends Biochem. Sci.*, 39(7), pp. 328–340.

Goodsell, D. S. and Olson, A. J. (2000) ‘Structural Symmetry and Protein Function’, *Annu. Rev. Biophys. Biomol. Struct.*, 29(1), pp. 105–153.

Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016) ‘Coming of age: ten years of next-generation sequencing technologies’, *Nat. Rev. Genet.*, 17(6), pp. 333–351.

Gouet, P. (2003) ‘ESPrict/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins’, *Nucleic Acids Res.*, 31(13), pp. 3320–3323.

Gredell, J. a *et al.* (2010) ‘Recognition of siRNA asymmetry by TAR RNA binding protein.’, *Biochemistry*, 49(14), pp. 3148–55.

Gregory, R. I. *et al.* (2005) ‘Human RISC couples microRNA biogenesis and posttranscriptional gene silencing.’, *Cell*, 123(4), pp. 631–40.

Gregory, R. I. and Shiekhattar, R. (2005) ‘MicroRNA biogenesis and cancer.’, *Cancer Res.*, 65(9), pp. 3509–12.

Grimson, A. *et al.* (2007) ‘MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing’, *Mol. Cell*, 27(1), pp. 91–105.

Guo, H. *et al.* (2010) ‘Mammalian microRNAs predominantly act to decrease target mRNA levels.’, *Nature*, 466(7308), pp. 835–40.

Gupta, V., Huang, X. and Patel, R. C. (2003) ‘The carboxy-terminal, M3 motifs of PACT and TRBP have opposite effects on PKR activity’, *Virology*, 315(2), pp. 283–291.

Haase, A. D. *et al.* (2005) ‘TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing.’, *EMBO Rep.*, 6(10), pp. 961–7.

Haigh, C. W. and Mallion, R. B. (1979) ‘Ring current theories in nuclear magnetic resonance’, *Prog. Nucl. Magn. Reson. Spectrosc.*, 13(4), pp. 303–344.

Han, J. (2004) ‘The Drosha-DGCR8 complex in primary microRNA processing’, *Genes Dev.*, 18(24), pp. 3016–3027.

Harding, H. P. *et al.* (2000) ‘Regulated Translation Initiation Controls Stress-Induced Gene Expression in Mammalian Cells’, *Mol. Cell*, 6(5), pp. 1099–1108.

Hashimoto, K. and Panchenko, A. R. (2010) ‘Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states’, *Proc. Natl. Acad. Sci.*, 107(47), pp. 20352–20357.

Heyam, A., Lagos, D. and Plevin, M. (2015) ‘Dissecting the roles of TRBP and PACT in double-stranded RNA recognition and processing of noncoding RNAs’, *Wiley Interdiscip. Rev. RNA*, 6(3), pp. 271–289.

Holm, L. and Rosenström, P. (2010) ‘Dali server: Conservation mapping in 3D’, *Nucleic Acids Res.*, 38(SUPPL. 2), pp. 545–549.

Huang, X., Hutchins, B. and Patel, R. C. (2002) ‘The C-terminal, third conserved motif of the protein activator PACT plays an essential role in the activation of

- double-stranded-RNA-dependent protein kinase (PKR).', *Biochem. J.*, 366(Pt 1), pp. 175–86.
- Hutvagner, G. (2001) 'A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA', *Science (80-.)*, 293(5531), pp. 834–838.
- Iwamura, T. *et al.* (2001) 'PACT, a double-stranded RNA binding protein acts as a positive regulator for type I interferon gene induced by Newcastle disease virus.', *Biochem. Biophys. Res. Commun.*, 282(2), pp. 515–23.
- Jakob, L. *et al.* (2016) 'Structural and functional insights into the fly microRNA biogenesis factor Loquacious', *RNA*, 22(3), pp. 383–396.
- James, R., Kleanthous, C. and Moore, G. R. (1996) 'The biology of E colicins: Paradigms and paradoxes', *Microbiology*, 142(7), pp. 1569–1580.
- Jiang, X. *et al.* (2012) 'Ubiquitin-Induced Oligomerization of the RNA Sensors RIG-I and MDA5 Activates Antiviral Innate Immune Response', *Immunity*, 36(6), pp. 959–973.
- Johnson, M. L. *et al.* (1981) 'Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques.', *Biophys. J.*, 36(3), pp. 575–588.
- Jonas, S. and Izaurralde, E. (2015) 'Towards a molecular understanding of microRNA-mediated gene silencing.', *Nat. Rev. Genet.*, 16(7), pp. 421–433.
- Kawamata, T., Seitz, H. and Tomari, Y. (2009) 'Structural determinants of miRNAs for RISC loading and slicer-independent unwinding.', *Nat. Struct. Mol. Biol.*, 16(9), pp. 953–60.
- Kay, L., Keifer, P. and Saarinen, T. (1992) 'Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity', *J. Am. Chem. Soc.*, 114(26), pp. 10663–10665.
- Kelly, A. E. *et al.* (2002) 'Low-Conductivity Buffers for High-Sensitivity NMR Measurements', *J. Am. Chem. Soc.*, 124(40), pp. 12013–12019.
- Kennedy, E. M. *et al.* (2015) 'Production of functional small interfering RNAs by an amino-terminal deletion mutant of human Dicer', *Proc. Natl. Acad. Sci.*, 112(50), pp. E6945–E6954.
- Kew, C. *et al.* (2013) 'Suppression of PACT-induced type I interferon production by herpes simplex virus 1 Us11 protein.', *J. Virol.*, 87(24), pp. 13141–9.
- Kim, V. N. (2005) 'MicroRNA biogenesis: coordinated cropping and dicing', *Nat. Rev. Mol. Cell Biol.*, 6(5), pp. 376–385.
- Kim, V. N., Han, J. and Siomi, M. C. (2009) 'Biogenesis of small RNAs in animals.', *Nat. Rev. Mol. Cell Biol.*, 10(2), pp. 126–39.
- Kim, Y. *et al.* (2014) 'Deletion of Human tarbp2 Reveals Cellular MicroRNA Targets and Cell-Cycle Function of TRBP.', *Cell Rep.*, 9(3), pp. 1061–74.
- Kim, Y. J., Maizel, A. and Chen, X. (2014) 'Traffic into silence: endomembranes and post-transcriptional RNA silencing.', *EMBO J.*, 33(9), pp. 968–980.
- Kimball, S. R. (1999) 'Eukaryotic initiation factor eIF2', *Int. J. Biochem. Cell Biol.*, 31(1), pp. 25–29.
- Kleinjung, J. and Fraternali, F. (2005) 'POPSCOMP: an automated interaction

- analysis of biomolecular complexes', *Nucleic Acids Res.*, 33(Web Server), pp. W342–W346.
- Koh, H. R. *et al.* (2013) 'ATP-independent diffusion of double-stranded RNA binding proteins.', *Proc. Natl. Acad. Sci. U. S. A.*, 110(1), pp. 151–6.
- Koh, H. R. *et al.* (2016) 'RNA scanning of a molecular machine with a built-in ruler', *J. Am. Chem. Soc.*, p. jacs.6b10387.
- Kok, K.-H. *et al.* (2007) 'Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA.', *J. Biol. Chem.*, 282(24), pp. 17649–57.
- Kok, K.-H. *et al.* (2011) 'The double-stranded RNA-binding protein PACT functions as a cellular activator of RIG-I to facilitate innate antiviral response.', *Cell Host Microbe*, 9(4), pp. 299–309.
- Kolakofsky, D., Kowalinski, E. and Cusack, S. (2012) 'A structure-based model of RIG-I activation.', *RNA*, 18(12), pp. 2118–27.
- Komuro, A. *et al.* (2016) 'The TAR-RNA binding protein is required for immunoresponses triggered by Cardiovirus infection', *Biochem. Biophys. Res. Commun.*, 480(2), pp. 187–193.
- Koscianska, E., Starega-Roslan, J. and Krzyzosiak, W. J. (2011) 'The role of Dicer protein partners in the processing of microRNA precursors.', *PLoS One*, 6(12), p. e28548.
- Kowalinski, E. *et al.* (2011) 'Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA.', *Cell*, 147(2), pp. 423–35.
- Krissinel, E. and Henrick, K. (2007) 'Inference of Macromolecular Assemblies from Crystalline State', *J. Mol. Biol.*, 372(3), pp. 774–797.
- Kundrotas, P. J. and Vakser, I. A. (2013) 'Protein-protein alternative binding modes do not overlap', *Protein Sci.*, 22(8), pp. 1141–1145.
- Laraki, G. *et al.* (2008) 'Interactions between the double-stranded RNA-binding proteins TRBP and PACT define the Medipal domain that mediates protein-protein interactions', *RNA Biol.*, 5(2), pp. 92–103.
- Lau, P.-W. *et al.* (2009) 'Structure of the human Dicer-TRBP complex by electron microscopy.', *Structure*, 17(10), pp. 1326–32.
- Lau, P.-W. *et al.* (2012) 'The molecular architecture of human Dicer.', *Nat. Struct. Mol. Biol.*, 19(4), pp. 436–40.
- Lee, H. Y. *et al.* (2013) 'Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing.', *Nucleic Acids Res.*, pp. 1–9.
- Lee, H. Y. and Doudna, J. a (2012) 'TRBP alters human precursor microRNA processing in vitro.', *RNA*, 18(11), pp. 2012–9.
- Lee, J. Y. *et al.* (2004) 'Merlin, a tumor suppressor, interacts with transactivation-responsive RNA-binding protein and inhibits its oncogenic activity.', *J. Biol. Chem.*, 279(29), pp. 30265–73.
- Lee, K., Fajardo, M. and Braun, R. (1996) 'A testis cytoplasmic RNA-binding protein that has the properties of a translational repressor.', *Mol. Cell. Biol.*, 16(6), pp. 3023–3024.

- Lee, Y. *et al.* (2003) 'The nuclear RNase III Droscha initiates microRNA processing', *Nature*, 425(6956), pp. 415–419.
- Lee, Y. *et al.* (2006) 'The role of PACT in the RNA silencing pathway.', *EMBO J.*, 25(3), pp. 522–32.
- Levitt, M. H. (2008) *Spin dynamics: Basics of nuclear magnetic resonance, 2nd edition*. John Wiley & Sons, Ltd.
- Li, S. *et al.* (2006) 'Molecular basis for PKR activation by PACT or dsRNA.', *Proc. Natl. Acad. Sci. U. S. A.*, 103(26), pp. 10005–10.
- Li, S. and Sen, G. C. (2003) 'PACT-mediated enhancement of reporter gene expression at the translational level.', *J. Interferon Cytokine Res.*, 23(12), pp. 689–97.
- Lin, X. *et al.* (2014) 'Up-regulation and worse prognostic marker of cytoplasmic TARBP2 expression in obstinate breast cancer.', *Med. Oncol.*, 31(4), p. 868.
- Lipari, G. and Szabo, A. (1982) 'Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity', *J. Am. Chem. Soc.*, 104(17), pp. 4546–4559.
- Liu, J. (2004) 'Argonaute2 Is the Catalytic Engine of Mammalian RNAi', *Science (80-.)*, 305(5689), pp. 1437–1441.
- Liu, J. *et al.* (2005) 'A role for the P-body component GW182 in microRNA function', *Nat. Cell Biol.*, 7(12), pp. 1261–1266.
- Liu, Q. *et al.* (2003) 'R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway.', *Science*, 301(5641), pp. 1921–5.
- Liu, Q. and Paroo, Z. (2010) 'Biochemical principles of small RNA pathways.', *Annu. Rev. Biochem.*, 79, pp. 295–319.
- Liu, Y. and Eisenberg, D. (2002) '3D domain swapping: As domains continue to swap', *Protein Sci.*, 11(6), pp. 1285–1299.
- Lukatsky, D. B. *et al.* (2007) 'Structural Similarity Enhances Interaction Propensity of Proteins', *J. Mol. Biol.*, 365(5), pp. 1596–1606.
- Lunde, B. M., Moore, C. and Varani, G. (2007) 'RNA-binding proteins: modular design for efficient function.', *Nat. Rev. Mol. Cell Biol.*, 8(6), pp. 479–90.
- Luo, D. *et al.* (2011) 'Structural insights into RNA recognition by RIG-I.', *Cell*, 147(2), pp. 409–22.
- Luthra, P. *et al.* (2013) 'Mutual Antagonism between the Ebola Virus VP35 Protein and the RIG-I Activator PACT Determines Infection Outcome.', *Cell Host Microbe*, 14(1), pp. 74–84.
- Ma, E. *et al.* (2008) 'Autoinhibition of human dicer by its internal helicase domain.', *J. Mol. Biol.*, 380(1), pp. 237–43.
- Macrae, I. J. *et al.* (2006) 'Structural basis for double-stranded RNA processing by Dicer.', *Science*, 311(5758), pp. 195–8.
- MacRae, I. J. *et al.* (2008) 'In vitro reconstitution of the human RISC-loading complex.', *Proc. Natl. Acad. Sci. U. S. A.*, 105(2), pp. 512–7.
- MacRae, I. J., Zhou, K. and Doudna, J. A. (2007) 'Structural determinants of RNA recognition and cleavage by Dicer.', *Nat. Struct. Mol. Biol.*, 14(10), pp. 934–40.

- Madrado, J. *et al.* (2001) 'Crystal structure of the central region of bovine fibrinogen (E5 fragment) at 1.4-Å resolution.', *Proc. Natl. Acad. Sci. U. S. A.*, 98(21), pp. 11967–72.
- Maillard, P. V *et al.* (2013) 'Antiviral RNA interference in mammalian cells.', *Science*, 342(6155), pp. 235–8.
- Maillard, P. V *et al.* (2016) 'Inactivation of the type I interferon pathway reveals long double-stranded RNA-mediated RNA interference in mammalian cells', *EMBO J.*, p. e201695086.
- Malone, C. D. and Hannon, G. J. (2009) 'Small RNAs as guardians of the genome.', *Cell*, 136(4), pp. 656–68.
- Marión, R. M. *et al.* (1999) 'A human sequence homologue of Staufen is an RNA-binding protein that is associated with polysomes and localizes to the rough endoplasmic reticulum.', *Mol. Cell. Biol.*, 19(3), pp. 2212–9.
- Marques, J. T. *et al.* (2008) 'The role of PACT in mediating gene induction, PKR activation, and apoptosis in response to diverse stimuli.', *J. Interferon Cytokine Res.*, 28(8), pp. 469–76.
- Martinez, J. *et al.* (2002) 'Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi', *Cell*, 110(5), pp. 563–574.
- Masliyah, G., Barraud, P. and Allain, F. H.-T. (2013) 'RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence.', *Cell. Mol. Life Sci.*, 70(11), pp. 1875–95.
- Matsui, M. *et al.* (2015) 'Reduced Expression of Argonaute 1, Argonaute 2, and TRBP Changes Levels and Intracellular Distribution of RNAi Factors', *Sci. Rep.*, 5(August), p. 12855.
- Matthews, D. A. *et al.* (1994) 'Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein', *Cell*, 77(5), pp. 761–771.
- McCullough, K. D. *et al.* (2001) 'Gadd153 Sensitizes Cells to Endoplasmic Reticulum Stress by Down-Regulating Bcl2 and Perturbing the Cellular Redox State', *Mol. Cell. Biol.*, 21(4), pp. 1249–1259.
- Meister, G. *et al.* (2004) 'Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs.', *Mol. Cell*, 15(2), pp. 185–97.
- Melo, S. *et al.* (2011) 'Small molecule enoxacin is a cancer-specific growth inhibitor that acts by enhancing TAR RNA-binding protein 2-mediated microRNA processing.', *Proc. Natl. Acad. Sci. U. S. A.*, 108(11), pp. 4394–9.
- Melo, S. a *et al.* (2009) 'A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function.', *Nat. Genet.*, 41(3), pp. 365–70.
- Meurs, E. *et al.* (1990) 'Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon.', *Cell*, 62(2), pp. 379–90.
- Meylan, E. *et al.* (2005) 'Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus', *Nature*, 437(7062), pp. 1167–1172.
- Mielke, S. P. and Krishnan, V. V (2009) 'Characterization of protein secondary structure from NMR chemical shifts.', *Prog. Nucl. Magn. Reson. Spectrosc.*, 54(3–4),

pp. 141–165.

Miloushev, V. Z. *et al.* (2008) 'Dynamic Properties of a Type II Cadherin Adhesive Domain: Implications for the Mechanism of Strand-Swapping of Classical Cadherins', *Structure*, 16(8), pp. 1195–1205.

Mittermaier, A. K. and Kay, L. E. (2009) 'Observing biological dynamics at atomic resolution using NMR.', *Trends Biochem. Sci.*, 34(12), pp. 601–11.

Monod, J., Wyman, J. and Changeux, J. (1965) 'On the nature of allosteric transitions: A plausible model', *J. Mol. Biol.*, 12(1), pp. 88–118.

Morris, G. A. and Freeman, R. (1979) 'Enhancement of nuclear magnetic resonance signals by polarization transfer', *J. Am. Chem. Soc.*, 101(3), pp. 760–762.

Morrison, E. A. *et al.* (2011) 'Antiparallel EmrE exports drugs by exchanging between asymmetric structures', *Nature*, 481(7379), pp. 45–50.

Nabuurs, S. B. *et al.* (2006) 'Traditional biomolecular structure determination by NMR spectroscopy allows for major errors', *PLoS Comput. Biol.*, 2(2), pp. 71–79.

Nanduri, S. *et al.* (1998) 'Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation.', *EMBO J.*, 17(18), pp. 5458–65.

Nanduri, S. *et al.* (2000) 'A dynamically tuned double-stranded RNA binding mechanism for the activation of antiviral kinase PKR.', *EMBO J.*, 19(20), pp. 5567–74.

Neilsen, C. T., Goodall, G. J. and Bracken, C. P. (2012) 'IsomiRs--the overlooked repertoire in the dynamic microRNAome.', *Trends Genet.*, 28(11), pp. 544–9.

Nisius, L. and Grzesiek, S. (2012) 'Key stabilizing elements of protein structure identified through pressure and temperature perturbation of its hydrogen bond network', *Nat. Chem.*, 4(9), pp. 711–717.

Noland, C. L. and Doudna, J. a. (2013) 'Multiple sensors ensure guide strand selection in human RNAi pathways.', *RNA*, 19(5), pp. 639–48.

Noland, C. L., Ma, E. and Doudna, J. a (2011) 'siRNA repositioning for guide strand selection by human Dicer complexes.', *Mol. Cell*, 43(1), pp. 110–21.

Ohrt, T. *et al.* (2008) 'Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells.', *Nucleic Acids Res.*, 36(20), pp. 6439–49.

Ohrt, T. *et al.* (2012) 'Intracellular Localization and Routing of miRNA and RNAi Pathway Components', *Curr. Top. Med. Chem.*, 12(2), pp. 79–88.

Otting, G. and Wüthrich, K. (1989) 'Extended heteronuclear editing of 2D 1H NMR spectra of isotope-labeled proteins, using the X(ω_1 , ω_2) double half filter', *J. Magn. Reson.*, 85(3), pp. 586–594.

Pace, C. N. *et al.* (1995) 'How to measure and predict the molar absorption coefficient of a protein', *Protein Sci.*, 4(11), pp. 2411–2423.

Pare, J. M. *et al.* (2009) 'Hsp90 regulates the function of argonaute 2 and its recruitment to stress granules and P-bodies.', *Mol. Biol. Cell*, 20(14), pp. 3273–84.

Park, H. *et al.* (1994) 'TAR RNA-binding protein is an inhibitor of the interferon-induced protein kinase PKR.', *Proc. Natl. Acad. Sci. U. S. A.*, 91(11), pp. 4713–7.

- Parker, G. S., Eckert, D. M. and Bass, B. L. (2006) 'RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA.', *RNA*, 12(5), pp. 807–18.
- Paroo, Z. *et al.* (2009) 'Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling.', *Cell*, 139(1), pp. 112–22.
- Patel, R. C. *et al.* (1995) 'The interferon-inducible double-stranded RNA-activated protein kinase self-associates in vitro and in vivo.', *Proc. Natl. Acad. Sci.*, 92(18), pp. 8283–8287.
- Patel, R. C. and Sen, G. C. (1998) 'PACT, a protein activator of the interferon-induced protein kinase, PKR.', *EMBO J.*, 17(15), pp. 4379–90.
- Peisley, A. *et al.* (2011) 'Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition', *Proc. Natl. Acad. Sci.*, 108(52), pp. 21010–21015.
- Peters, G. A. *et al.* (2001) 'Modular structure of PACT: distinct domains for binding and activating PKR.', *Mol. Cell. Biol.*, 21(6), pp. 1908–20.
- Peters, G. A. *et al.* (2002) 'Inhibition of PACT-Mediated Activation of PKR by the Herpes Simplex Virus Type 1 Us11 Protein', *J. Virol.*, 76(21), pp. 11054–11064.
- Peters, G. A. *et al.* (2009) 'The double-stranded RNA-binding protein, PACT, is required for postnatal anterior pituitary proliferation.', *Proc. Natl. Acad. Sci. U. S. A.*, 106(26), pp. 10696–701.
- Peters, G. A., Li, S. and Sen, G. C. (2006) 'Phosphorylation of specific serine residues in the PKR activation domain of PACT is essential for its ability to mediate apoptosis.', *J. Biol. Chem.*, 281(46), pp. 35129–36.
- Peters, G. a, Dickerman, B. and Sen, G. C. (2009) 'Biochemical analysis of PKR activation by PACT.', *Biochemistry*, 48(31), pp. 7441–7.
- Pfaff, J. and Meister, G. (2013) 'Argonaute and GW182 proteins: an effective alliance in gene silencing', *Biochem. Soc. Trans.*, 41(4), pp. 855–860.
- Pollegioni, L. *et al.* (2003) 'Contribution of the dimeric state to the thermal stability of the flavoprotein D-amino acid oxidase.', *Protein Sci.*, 12(5), pp. 1018–1029.
- Poulsen, H. *et al.* (2006) 'Dimerization of ADAR2 is mediated by the double-stranded RNA binding domain.', *RNA*, 12(7), pp. 1350–60.
- Provost, P. (2002) 'Ribonuclease activity and RNA binding of recombinant human Dicer', *EMBO J.*, 21(21), pp. 5864–5874.
- Puthalakath, H. *et al.* (2007) 'ER Stress Triggers Apoptosis by Activating BH3-Only Protein Bim', *Cell*, 129(7), pp. 1337–1349.
- Ramana, C. V *et al.* (2000) 'Regulation of c-myc expression by IFN- γ through Stat1-dependent and -independent pathways', *EMBO J.*, 19(2), pp. 263–272.
- Ramos, A. *et al.* (2000) 'RNA recognition by a Staufen double-stranded RNA-binding domain.', *EMBO J.*, 19(5), pp. 997–1009.
- Reinhard, L. *et al.* (2013) 'Optimization of protein buffer cocktails using Thermofluor.', *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.*, 69(Pt 2), pp. 209–14.

- Renatus, M. *et al.* (2001) 'Dimer formation drives the activation of the cell death protease caspase 9.', *Proc. Natl. Acad. Sci. U. S. A.*, 98(25), pp. 14250–5.
- Ricci, E. P. *et al.* (2014) 'Staufen1 senses overall transcript secondary structure to regulate translation.', *Nat. Struct. Mol. Biol.*, 21(1), pp. 26–35.
- Robb, G. B. and Rana, T. M. (2007) 'RNA helicase A interacts with RISC in human cells and functions in RISC loading.', *Mol. Cell*, 26(4), pp. 523–37.
- Rousseau, F. *et al.* (2001) 'Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues', *Proc. Natl. Acad. Sci.*, 98(10), pp. 5596–5601.
- Rousseau, F., Schymkowitz, J. W. H. and Itzhaki, L. S. (2003) 'The unfolding story of three-dimensional domain swapping', *Structure*, 11(3), pp. 243–251.
- Rowe, T. M. *et al.* (2006) 'A role of the double-stranded RNA-binding protein PACT in mouse ear development and hearing.', *Proc. Natl. Acad. Sci. U. S. A.*, 103(15), pp. 5823–8.
- Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: a unified platform for automated protein structure and function prediction.', *Nat. Protoc.*, 5(4), pp. 725–38.
- Rule, G. S. and Hitchens, T. K. (2006) *Fundamentals of Protein NMR Spectroscopy*. Springer-Verlag.
- Sahakyan, A. B. *et al.* (2011) 'Structure-based prediction of methyl chemical shifts in proteins', *J. Biomol. NMR*, 50(4), pp. 331–346.
- Sand, M. *et al.* (2012) 'Expression levels of the microRNA maturing microprocessor complex component DGCR8 and the RNA-induced silencing complex (RISC) components argonaute-1, argonaute-2, PACT, TARBP1, and TARBP2 in epithelial skin cancer.', *Mol. Carcinog.*, 51(11), pp. 916–22.
- Sanghvi, V. R. and Steel, L. F. (2011a) 'A re-examination of global suppression of RNA interference by HIV-1.', *PLoS One*, 6(2), p. e17246.
- Sanghvi, V. R. and Steel, L. F. (2011b) 'The cellular TAR RNA binding protein, TRBP, promotes HIV-1 replication primarily by inhibiting the activation of double-stranded RNA-dependent kinase PKR.', *J. Virol.*, 85(23), pp. 12614–21.
- Schanda, P., Kupçê, E. and Brutscher, B. (2005) 'SOFAS-T-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds', *J. Biomol. NMR*, 33(4), pp. 199–211.
- Schirle, N. T., Sheu-Gruttadauria, J. and MacRae, I. J. (2014) 'Structural basis for microRNA targeting', *Science (80-.)*, 346(6209), pp. 608–613.
- Schlee, M. and Hartmann, G. (2010) 'The chase for the RIG-I ligand--recent advances.', *Mol. Ther.*, 18(7), pp. 1254–62.
- Schubert, M. *et al.* (2002) 'A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ¹³C chemical shift statistics.', *J. Biomol. NMR*, 24(2), pp. 149–54.
- Schulz, G. E. (2010) 'The Dominance of Symmetry in the Evolution of Homooligomeric Proteins', *J. Mol. Biol.*, 395(4), pp. 834–843.
- Schulz, M. N., Landström, J. and Hubbard, R. E. (2013) 'MTSA - A Matlab program to fit thermal shift data', *Anal. Biochem.*, 433(1), pp. 43–47.

- Seth, R. B. *et al.* (2005) 'Identification and Characterization of MAVS, a Mitochondrial Antiviral Signaling Protein that Activates NF- κ B and IRF3', *Cell*, 122(5), pp. 669–682.
- Shallom, D. *et al.* (2004) 'Effect of dimer dissociation on activity and thermostability of the β -glucuronidase from *Geobacillus stearothermophilus*: Dissecting the different oligomeric forms of family 67 glycoside hydrolases', *J. Bacteriol.*, 186(20), pp. 6928–6937.
- Shamoo, Y., Abdul-Manan, N. and Williams, K. R. (1995) 'Multiple RNA binding domains (RBDs) just don't add up.', *Nucleic Acids Res.*, 23(5), pp. 725–728.
- Sharma, D. and Rajarathnam, K. (2000) '¹³C NMR chemical shifts can predict disulfide bond formation', *J. Biomol. NMR*, 18, pp. 165–171.
- Shcherbo, D. *et al.* (2007) 'Bright far-red fluorescent protein for whole-body imaging', *Nat. Methods*, 4(9), pp. 741–746.
- Shen, Y. and Bax, A. (2013) 'Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks.', *J. Biomol. NMR*, 56(3), pp. 227–41.
- Siffroi, J. P. *et al.* (2001) 'Expression of the TAR RNA binding protein in human testis.', *Mol. Hum. Reprod.*, 7(3), pp. 219–25.
- Silva, J. M. *et al.* (2005) 'Second-generation shRNA libraries covering the mouse and human genomes', *Nat. Genet.*, 37(11), p. 1281.
- Simpson, R. J. (2010) 'SYPRO Orange Fluorescent Staining of Protein Gels', *Cold Spring Harb. Protoc.*, 2010(4), p. pdb.prot5414-prot5414.
- Singh, M. *et al.* (2011) 'Stress-induced phosphorylation of PACT reduces its interaction with TRBP and leads to PKR activation.', *Biochemistry*, 50(21), pp. 4550–60.
- Singh, M. and Patel, R. C. (2012) 'Increased interaction between PACT molecules in response to stress signals is required for PKR activation.', *J. Cell. Biochem.*, 113(8), pp. 2754–64.
- Siu, K.-L. *et al.* (2014) 'Middle east respiratory syndrome coronavirus 4a protein is a double-stranded RNA-binding protein that suppresses PACT-induced activation of RIG-I and MDA5 in the innate antiviral response.', *J. Virol.*, 88(9), pp. 4866–76.
- Sohn, S. Y. *et al.* (2007) 'Crystal structure of human DGCR8 core.', *Nat. Struct. Mol. Biol.*, 14(9), pp. 847–53.
- Sousa, E. *et al.* (2013) 'Enoxacin inhibits growth of prostate cancer cells and effectively restores microRNA processing.', *Epigenetics*, 8(5), pp. 548–58.
- St Johnston, D. *et al.* (1992) 'A conserved double-stranded RNA-binding domain.', *Proc. Natl. Acad. Sci. U. S. A.*, 89(22), pp. 10979–83.
- Stalder, L. *et al.* (2013) 'The rough endoplasmic reticulum is a central nucleation site of siRNA-mediated RNA silencing', *EMBO J.*, 32(8), pp. 1115–1127.
- Stefl, R. *et al.* (2006) 'Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs.', *Structure*, 14(2), pp. 345–55.
- Suzuki, J. and Yokota, S. (2014) 'PRKRA Localizes to Nuage Structures and the Ectoplasmic Specialization and Tubulobulbar Complexes in Rat and Mouse Testis', *J. Histol.*, 2014, pp. 1–9.

- Swapna, L. S., Srikeerthana, K. and Srinivasan, N. (2012) 'Extent of Structural Asymmetry in Homodimeric Proteins: Prevalence and Relevance', *PLoS One*. Edited by H. W. van Veen, 7(5), p. e36688.
- Tafforeau, L. *et al.* (2011) 'Generation and comprehensive analysis of an influenza virus polymerase cellular interaction network.', *J. Virol.*, 85(24), pp. 13010–8.
- Takahashi, T. *et al.* (2013) 'Distinguishable In Vitro Binding Mode of Monomeric TRBP and Dimeric PACT with siRNA.', *PLoS One*, 8(5), p. e63434.
- Takeuchi, O. and Akira, S. (2008) 'MDA5/RIG-I and virus recognition.', *Curr. Opin. Immunol.*, 20(1), pp. 17–22.
- Tan, C. Y. and Huang, Y. X. (2015) 'Dependence of Refractive Index on Concentration and Temperature in Electrolyte Solution, Polar Solution, Nonpolar Solution, and Protein Solution', *J. Chem. Eng. Data*, 60(10), pp. 2827–2833.
- Tanakai, Y. *et al.* (2004) 'How oligomerization contributes to the thermostability of an archaeon protein: Protein L-isoaspartyl-O-methyltransferase from *Sulfolobus tokodaii*', *J. Biol. Chem.*, 279(31), pp. 32957–32967.
- Taylor, D. W. *et al.* (2013) 'Substrate-specific structural rearrangements of human Dicer.', *Nat. Struct. Mol. Biol.*, 20(6), pp. 9–11.
- tenOever, B. R. (2016) 'The Evolution of Antiviral Defense Systems', *Cell Host Microbe*, 19(2), pp. 142–149.
- Tomari, Y. *et al.* (2004) 'A protein sensor for siRNA asymmetry.', *Science*, 306(5700), pp. 1377–80.
- Ung, T. L. *et al.* (2001) 'Heterologous dimerization domains functionally substitute for the double-stranded RNA binding domains of the kinase PKR.', *EMBO J.*, 20(14), pp. 3728–37.
- Vattem, K. M., Staschke, K. A. and Wek, R. C. (2001) 'Mechanism of activation of the double-stranded-RNA-dependent protein kinase, PKR', *Eur. J. Biochem.*, 268(13), pp. 3674–3684.
- Vaughn, L. S. *et al.* (2015) 'Altered activation of protein kinase PKR and enhanced apoptosis in dystonia cells carrying a mutation in PKR activator protein PACT', *J. Biol. Chem.*, p. jbc.M115.669408.
- Voinnet, O. (2005) 'Induction and suppression of RNA silencing: insights from viral infections', *Nat. Rev. Genet.*, 6(3), pp. 206–220.
- Vranken, W. F. *et al.* (2005) 'The CCPN data model for NMR spectroscopy: Development of a software pipeline', *Proteins Struct. Funct. Genet.*, 59(4), pp. 687–696.
- Wagner, G., Pardi, A. and Wuethrich, K. (1983) 'Hydrogen bond length and proton NMR chemical shifts in proteins', *J. Am. Chem. Soc.*, 105(18), pp. 5948–5949.
- Wahle, E. and Winkler, G. S. (2013) 'RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes', *Biochim. Biophys. Acta - Gene Regul. Mech.*, 1829(6–7), pp. 561–570.
- Wang, H.-W. *et al.* (2009) 'Structural insights into RNA processing by the human RISC-loading complex.', *Nat. Struct. Mol. Biol.*, 16(11), pp. 1148–53.
- Wang, X. *et al.* (2015) 'Dynamic profiling of double-stranded RNA binding proteins', *Nucleic Acids Res.*, 43(15), pp. 7566–7576.

- Warner, M. J. *et al.* (2016) 'S6K2-mediated regulation of TRBP as a determinant of miRNA expression in human primary lymphatic endothelial cells', *Nucleic Acids Res.*, p. gkw631.
- Weber, G. and Drickamer, H. G. (1983) 'The effect of high pressure upon proteins and other biomolecules', *Q. Rev. Biophys.*, 16(1), p. 89.
- White, E. *et al.* (2014) 'Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA.', *Nat. Struct. Mol. Biol.*, (April).
- Wilson, R. C. *et al.* (2015) 'Dicer-TRBP Complex Formation Ensures Accurate Mammalian MicroRNA Biogenesis', *Mol. Cell*, 57(3), pp. 397–407.
- Wray, G. A. (2015) 'Molecular clocks and the early evolution of metazoan nervous systems', *Philos. Trans. R. Soc. B Biol. Sci.*, 370(1684), pp. 20150046–20150046.
- Wu, H. *et al.* (2004) 'Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III.', *Proc. Natl. Acad. Sci. U. S. A.*, 101(22), pp. 8307–12.
- Wu, S. and Kaufman, R. J. (1997) 'A Model for the Double-stranded RNA (dsRNA)-dependent Dimerization and Activation of the dsRNA-activated Protein Kinase PKR', *J. Biol. Chem.*, 272(2), pp. 1291–1296.
- Wüthrich, K. (1990) 'Protein structure determination in solution by NMR spectroscopy', *J. Biol. Chem.*, 265(36), pp. 22059–22062.
- Wyatt, P. J. (1993) 'Light scattering and the absolute characterization of macromolecules', *Anal. Chim. Acta*, 272(1), pp. 1–40.
- Wyatt Technology (2005) *Optilab rEX User's Guide (M1500 Rev. D)*.
- Xu, X.-P. and Case, D. A. (2002) 'Probing multiple effects on ¹⁵N, ¹³C alpha, ¹³C beta, and ¹³C' chemical shifts in peptides using density functional theory.', *Biopolymers*, 65(6), pp. 408–23.
- Yamaguchi, H. (2004) 'CHOP Is Involved in Endoplasmic Reticulum Stress-induced Apoptosis by Enhancing DR5 Expression in Human Carcinoma Cells', *J. Biol. Chem.*, 279(44), pp. 45495–45502.
- Yamashita, S. *et al.* (2011) 'Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein.', *Protein Sci.*, 20(1), pp. 118–30.
- Yang, J.-S. *et al.* (2011) 'Widespread regulatory activity of vertebrate microRNA* species.', *RNA*, 17(2), pp. 312–26.
- Yang, S. W. *et al.* (2010) 'Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing.', *Structure*, 18(5), pp. 594–605.
- Yoda, M. *et al.* (2010) 'ATP-dependent human RISC assembly pathways.', *Nat. Struct. Mol. Biol.*, 17(1), pp. 17–23.
- Yoneyama, M. *et al.* (2005) 'Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity.', *J. Immunol.*, 175(5), pp. 2851–8.
- Yoneyama, M. and Fujita, T. (2009) 'RNA recognition and signal transduction by RIG-I-like receptors.', *Immunol. Rev.*, 227(1), pp. 54–65.
- Zech, M. *et al.* (2014) 'DYT16 revisited: Exome sequencing identifies PRKRA

- mutations in a European dystonia family', *Mov. Disord.*, 29(12), pp. 1504–1510.
- Zhang, F. *et al.* (2001) 'Binding of double-stranded RNA to protein kinase PKR is required for dimerization and promotes critical autophosphorylation events in the activation loop.', *J. Biol. Chem.*, 276(27), pp. 24946–58.
- Zhang, H. (2002) 'Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP', *EMBO J.*, 21(21), pp. 5875–5885.
- Zhao, H., Brown, P. H. and Schuck, P. (2011) 'On the distribution of protein refractive index increments', *Biophys. J.*, 100(9), pp. 2309–2317.
- Zhong, J. *et al.* (1999) 'A double-stranded RNA binding protein required for activation of repressed messages in mammalian germ cells.', *Nat. Genet.*, 22(2), pp. 171–4.
- Zimm, B. H. (1948) 'The Scattering of Light and the Radial Distribution Function of High Polymer Solutions', *J. Chem. Phys.*, 16(12), p. 1093.
- Zwahlen, C. *et al.* (1997) 'Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage λ N-Peptide/boxB RNA Complex', *J. Am. Chem. Soc.*, 119(29), pp. 6711–6721.