

Isotopic Labelling Tools to Aid NMR  
Studies of Protein Structure and  
Interactions: Applications to the LIMD1  
Scaffold Protein

Benjamin Rowlinson

Doctor of Philosophy

University of York

Biology

December 2022

## Abstract

The scaffold protein LIM domain-containing protein 1 (LIMD1) has been identified as a tumour suppressing protein with roles in a range of cell signalling pathways. In microRNA (miRNA) mediated silencing, LIMD1 has been shown to be a regulator of the miRNA induced silencing complex (miRISC), able to bind to both Argonaute (AGO) and GW182 proteins and promote their co-localisation. In hypoxic signalling, LIMD1 is thought to promote degradation of hypoxia inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ) through scaffolding of prolyl hydroxylases (PHD) and von-Hippel Lindau (VHL) proteins. The regions of LIMD1 responsible for binding to Argonaute-2 (AGO2), Trinucleotide Repeat Containing Adaptor 6A (TNRC6A), PHD2 and VHL have been mapped through co-immunoprecipitation assays (Co-IPs), but precise binding sites and motifs have not been identified. This thesis investigates the structural biology and biophysics of the interactions of LIMD1 with its partner proteins in miRNA mediated silencing and hypoxic signalling pathways. For each system, this involved recapitulating the reported interactions of LIMD1 *in vitro* before more in depth analysis could be achieved. Investigations into the interactions of LIMD1 in the context of miRNA mediated silencing were largely unsuccessful but informed analysis of LIMD1 in hypoxic signalling. In that context, reconstitution of complexes of LIMD1 with PHD2 and VHL *in vitro* was demonstrated using NMR spectroscopy. Compelling evidence was obtained for identification of the binding interface between LIMD1 and PHD2 which was consistent with previously reported *in vivo* data. Finally, new methodologies were developed that combined specific isotopic unlabelling with filtered/edited NOESY experiments to aid in structural studies of challenging proteins by NMR spectroscopy. Overall, this thesis addresses the challenges of studying the structural biology and interaction mechanisms of multi-domain proteins with large unstructured regions, and presents case studies for the use of NMR spectroscopy together with new methodologies to support this important area of biology.

# Contents

Abstract .....	2
Contents.....	3
Table of Tables .....	7
Table of Figures .....	9
Acknowledgements .....	16
Declaration .....	17
1. Introduction .....	18
1.1. Scaffold Proteins.....	18
1.2. LIM Domains are Protein Binding Motifs with Diverse Modes of Binding.....	21
1.3. The LAW Family of LIM Domain Scaffold Proteins .....	32
1.4. The LIMD1 Scaffold Protein.....	35
1.5. miRNA Mediated Gene Silencing .....	36
1.6. The Role of LIMD1 in miRNA Mediated Gene Silencing.....	39
1.7. The Hypoxic Signalling Pathway.....	43
1.8. LIMD1 is a Key Regulator of Hypoxic Signalling.....	50
1.9. NMR Spectroscopy is a Powerful Tool to Study the Structural Biology and Biophysics of Scaffold Proteins .....	53
1.10. Aims.....	55
2. Materials and Methods .....	58
2.1. Protein production.....	58
2.1.1. Producing the LIMD1 166-260 Expression Vector .....	63
2.1.2. Mutagenesis.....	64
2.1.3. Expression of unlabelled LIMD1, PHD2 and TNRC6A proteins .....	67
2.1.4. Expression of uniformly isotope labelled protein .....	68
2.1.5. AGO2 Expression .....	69
2.1.6. Purification of LIMD1, PHD2 and TNRC6A proteins .....	70

2.1.7. Production of the VHL:ElonginB:ElonginC Complex .....	71
2.1.8. Production of Methyl-labelled PHD2 .....	72
2.1.9. Production of Specifically Isotopically Unlabelled Protein .....	72
2.2. Peptide Arrays .....	74
2.3. Pull Down Assays .....	75
2.4. Western Blotting .....	77
2.5. AlphaFold 2 Predictions .....	77
2.6. NMR Experiments.....	78
2.6.1. NMR Sample Preparation .....	78
2.6.2. NMR Experimental parameters .....	79
2.6.3. Analysis of Specific Isotope Unlabelled NMR Experiments.....	84
2.7. Mass Spectrometry Analysis of the VHL:ElonginB:ElonginC Complex.....	86
2.8. Analytical Size Exclusion and Co-elution.....	87
2.9. Succinate-Glo PHD2 Assay .....	87
2.10. HPEPDOCK Docking.....	87
2.11. Structure Analysis.....	88
2.12. NMR Structure Calculations .....	88
3. Results Chapter 1: Investigating the Protein:Protein Interactions of the LIMD1 Scaffold Protein in the miRNA Mediated Silencing Pathway .....	91
3.1. Introduction .....	91
3.2. Results and Discussion .....	92
3.2.1. Investigating the LIMD1:TNRC6a Interaction using Peptide Arrays .....	92
3.2.2. Investigating the LIMD1:TNRC6A Interaction using Pull-Down Assays	104
3.2.3. AlphaFold2 Guided Investigation of the LIMD1:TNRC6A Interaction ....	116
3.2.4. Production of Argonaute 2.....	121
3.3. Conclusion and Future Work .....	122
4. Results Chapter 2: Investigation of the LIMD1 Protein Interactions that Mediate HIF-1 $\alpha$ Degradation.....	124



4.1. Introduction .....	124
4.2. Results and Discussion .....	125
4.2.1. LIMD1 Binds to VHL though the Second LIM Domain of LIMD1 .....	125
4.2.2. Identification of the Binding Site of PHD2 on LIMD1 .....	133
4.2.3. Identifying the Binding Site of LIMD1 on PHD2.....	161
4.3. Conclusion and Future Work .....	174
5. Results Chapter 3: Using Specific Isotopic Unlabelling to Aid Structural Studies of Proteins by Solution NMR Spectroscopy.....	178
5.1. Introduction .....	178
5.1.1. Specific Isotope Unlabelling .....	179
5.1.2. Routes for Targeting Aliphatic Residues .....	180
5.1.3. Routes for Targeting Aromatic Residues .....	184
5.1.4. Routes for Targeting Polar Residues .....	186
5.1.5. Routes for Targeting Charged Residues .....	187
5.1.6. Special Cases .....	187
5.1.7. How Specific Isotope Unlabelling can Aid in Structural Studies of Proteins by Solution NMR Spectroscopy.....	190
5.2. Results and Discussion .....	195
5.2.1. Specific Isotope Unlabelling of SHIRT-R3.....	195
5.2.2. Specific Isotope Unlabelling of SasYr .....	202
5.2.3. Specific Isotope Unlabelling to Complement 3D Structure Calculations	208
5.2.4. Does the addition of filtered/edited NOESY data improve the rate of structure convergence?.....	215
5.2.5. Specific Isotope Unlabelling to Compliment Sparse Data .....	218
5.2.6. Validation of Automatic NOE Assignment by Specific Isotope Unlabelling .....	222
5.3. Conclusion and Future Work .....	223
6. Conclusion .....	227

7. Appendix .....	231
8. References .....	273

## Table of Tables

Table 1.1. Unique LIM domain interaction structures used for analysis of LIM domain interactions.....	25
Table 2.1. Table of proteins and associated vectors used for protein production.....	59
Table 2.2. Continuation of Table 2.1. ....	60
Table 2.3. Continuation of Table 2.1 and 2.2 .....	61
Table 2.4. Continuation of Table 2.1, 2.2 and 2.3 .....	62
Table 2.5. PCR mix components for the amplification of the LIMD1 166-260 gene .	63
Table 2.6. PCR steps for the amplification of the LIMD1 166-260 gene.....	64
Table 2.7. Temperatures and times for steps in the PCR program for TNRC6A mutagenesis.....	66
Table 2.8. PCR mix components for QuikChange Lighting mutagenesis .....	67
Table 2.9. Recipe for 1 L of M9 media as used to produce uniformly isotope labelled proteins .....	69
Table 2.10. Compounds and concentrations used for specific isotope unlabelling of SHIRT-R3.....	73
Table 2.11. Experimental parameters used for 2D ( <sup>1</sup> H, <sup>15</sup> N) correlation experiments. ....	80
Table 2.12. Continuation of Table 2.11 .....	81
Table 4.1. Mass spectrometry results for the VHL:EloB:EloC sample.....	127
Table 5.1. A summary of the scramble free isotope labelling strategies used in <i>E. coli</i> expression of recombinant proteins.....	189
Table 5.2. SHIRT-R3 sequence and amino acid composition used to inform unlabelling schemes.....	197
Table 5.3. SasYr sequence and amino acid composition used to inform unlabelling schemes.....	202
Table 5.4. Statistics from structure calculations of SasYr with different 2D filtered/edited NOESY spectra added.....	210
Table 5.5. Statistics from structure calculations of SasYr with data from different unlabelling schemes incorporated.....	214

Table 5.6. Structure statistics for SasYr structure calculations carried out with data from different unlabelling schemes used to compliment 25% of the peaks from the 3D ( $^1\text{H}$ , $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peak list.....	221
---	-----

## Table of Figures

Figure 1.1. Structures of p53 in complex with different partner proteins .....	20
Figure 1.2. Overview of the structure of a LIM domain .....	22
Figure 1.3. Structures of LIM domains bound to partner proteins . .....	26
Figure 1.4. Structure of LMO4 LIM1 bound to Idb1-LID and Secondary structure diagram showing the $\beta$ -zipper binding mode of LIM domains with linear motifs .....	28
Figure 1.5. Structure of LMO2 LIM12 in complex with LDB1, LMO4 LIM12 in complex with LDB1, LHX3 LIM12 in complex with ISL-1 and TES LIM23 in complex with ARP7A and MENA superposed by secondary structure matching on the first LIM domain. ....	30
Figure 1.6. Phylogenetic tree showing the evolutionary distance between members of the zyxin protein family with the LAW subfamily indicated.. .....	32
Figure 1.7. Multiple sequence alignment of human LIMD1, AJUBA and WTIP. ....	34
Figure 1.8. Disorder prediction for LIMD1 by IUPred3.....	36
Figure 1.9. Crystal structure of AGO1 bound to GW182 hook motif.....	38
Figure 1.10. Cartoon representation of the miRISC. ....	39
Figure 1.11. Schematic representation of LIMD1 regions essential for binding to AGO2 and TNRC6A .....	41
Figure 1.12. Overview of AGO2 domain composition and structure .....	42
Figure 1.13. Schematic representation of HIF-1 $\alpha$ regulation in response to oxygen availability .....	45
Figure 1.14. Cartoon representation of the domain composition of PHD1,2 and 3. .	46
Figure 1.15. Crystal structures of PHD2 in complex with HIF-1 $\alpha$ CODD and inhibitors showing the 'open' and 'closed' conformations of PHD2. and crystal structures of PHD2 bound to both HIF-1 $\alpha$ CODD and NODD.....	48
Figure 1.16. Crystal structure of the VHL, Elongin B, Elongin C complex bound to a peptide of the HIF-1 $\alpha$ CODD .....	49
Figure 1.17. Schematic representation of LIMD1 regions essential for binding to PHD2 and VHL.....	51
Figure 2.1. Example of a purge element pulse sequence with vector model description.....	85
Figure 2.2. Schematic description of the ARIA2 structure building process. ....	89

Figure 3.1. Chromatogram and SDS-PAGE analysis of LIM123 purification .....	93
Figure 3.2. Representation of the peptide array method and. Initial peptide array results without any partner protein, with LIM123-MBP or with just MBP.. .....	95
Figure 3.3. SDS-PAGE analysis of the purity of LIM123-His6 protein post MBP tag cleavage and being passed over an MBPTrap column.. .....	96
Figure 3.4. Peptide array assays with 20 $\mu$ M MBP, 20 $\mu$ M LIM123-His6, or 2 $\mu$ M LIM123-His6 using a more stringent protocol.....	98
Figure 3.5. Analysis of the SEC purification of MBP-LIM123. ....	99
Figure 3.6. Analysis of the purification of individual LIM domains. ....	101
Figure 3.7. TNRC6A peptide arrays with individual LIM domains of LIMD1 and SDS-PAGE analysis of the individual LIM domains and LIM123 probed with the ATTO 488 Ni-NTA conjugate. ....	103
Figure 3.8. A. SDS-PAGE analysis of MBP and TNRC6A 1-883 fractions from SEC. B. Western blot of MBP and TNRC6A SEC fractions C. SDS-PAGE analysis of the SEC purification of individual LIM domains .....	105
Figure 3.9. SDS-PAGE analysis of pull-down assays for individual LIM domains with MBP-TNRC6A 1-883.....	107
Figure 3.10. SDS-PAGE analysis of pull-down assay controls for individual LIM domains with MBP.....	108
Figure 3.11. SDS-PAGE analysis of pull-down assays for LIM2 with TNRC6A constructs.....	110
Figure 3.12. AGO2 pull-down assays with MBP-TNRC6A 1-202 and MBP-TNRC6A 720-883.....	113
Figure 3.13. SDS-PAGE analysis for the SEC purification of MBP tagged LIMD1 110-166 and AGO2 pull-down assay with LIMD1 110-166.....	115
Figure 3.14. AlphaFold2 predictions of LIM12 with TNRC6A 1-900 and AGO2 519-779 with TNRC6A 695-773 and LIM12.....	117
Figure 3.15. SDS-PAGE analysis of the purifications of LIM12 and a LIM12-TNRC6A715-739 fusion protein .....	118
Figure 3.16. 2D ( $^1$ H, $^{15}$ N) HMQC spectra of [U- $^{15}$ N]-LIM12 and [U- $^{15}$ N]-LIM12-TNRC6A715-739 (Pink) and an overlay of the two spectra.....	120
Figure 3.17. Western blot analysis of AGO2 production at different time points using different volumes of V1 viral stock .....	122
Figure 4.1. SDS-PAGE analysis of the VHL:EloB:EloC complex post purification	126

Figure 4.2. MS spectra of the VHL:EloB:EloC sample .....	128
Figure 4.3. Chromatograms from size exclusion chromatography of VHL with and without LIM2.....	129
Figure 4.4. 2D ( <sup>1</sup> H, <sup>15</sup> N) HMQC spectra of LIM2 (100 μM) without VHL:EloB:EloC, with VHL:EloB:EloC (250 μM), and an overlay of the spectra in with and without VHL:EloB:EloC .....	131
Figure 4.5. SDS-PAGE analysis for the purification of MBP tagged LIMD1 65-260 and untagged full length PHD2 .....	134
Figure 4.6 PHD2 pull downs using His6-MBP-tagged LIMD1 65-260 and just His6-MBP. ....	135
Figure 4.7. SEC chromatograms of PHD2, LIMD1 65-260 and a mixture of the two proteins followed by absorbance at 280 nm.....	137
Figure 4.8. SDS-PAGE analysis of the purification of unlabelled PHD2 181-426, pure unlabelled LIMD1 1-80, pure [U- <sup>15</sup> N] LIMD1 160-100, pure [U- <sup>15</sup> N] LIMD1 110-166, and Pure [U- <sup>15</sup> N] LIMD1 65-260.....	139
Figure 4.9. 2D ( <sup>1</sup> H, <sup>15</sup> N) HSQC or HMQC spectra for various regions of the pre-LIM of LIMD1.....	140
Figure 4.10. 2D ( <sup>1</sup> H, <sup>15</sup> N) HSQC spectra of [U- <sup>15</sup> N]-LIMD1 65-260, [U- <sup>15</sup> N]-LIMD1 160-300 and [U- <sup>15</sup> N]-LIMD1 110-166 with and without PHD2 181-426.....	141
Figure 4.11. <b>A.</b> Schematic representation showing how the LIMD1 166-260 related to the previously used pre-LIM regions and to full length LIMD1. <b>B.</b> Section of the chromatogram for the SEC purification of [U- <sup>15</sup> N]-LIMD1 166-260 showing absorbance at 280 nm. <b>C.</b> SDS-PAGE analysis of the purification of tag less [U- <sup>15</sup> N]-LIMD1 166-260. ....	143
Figure 4.12. Overlay of the 2D ( <sup>1</sup> H, <sup>15</sup> N) HSQC spectra for LIMD1 166-260, LIMD1 65-260, LIMD1 160-300 .....	145
Figure 4.13. 2D ( <sup>1</sup> H, <sup>15</sup> N) HMQC spectra for [U- <sup>15</sup> N]-LIMD1 166-260 with and without PHD2 181-426 .....	146
Figure 4.14. 2D ( <sup>1</sup> H, <sup>15</sup> N) HMQC spectra for [U- <sup>15</sup> N]-LIMD1 166-260 with and without PHD2 181-426 and HIF-1α CODD peptide. ....	147
Figure 4.15. SEC chromatograms of PHD2, LIMD1 166-260 and a mixture of the two proteins followed by absorbance at 280 nm .....	148
Figure 4.16. Schematic representation of the PHD2 activity assay used. ....	149
Figure 4.17. PHD2 activity assays using a succinate-glo luminescent assay.....	151

Figure 4.18. strips from the HNCACB and HNCOCACB spectra of LIMD1 166-260 for the assigned region of 168-193.....	153
Figure 4.19. Partially assigned 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HMQC spectrum of $[\text{U-}^{15}\text{N}]$ -LIMD1 166-260.....	154
Figure 4.20. The Z-score for the intensity change of peaks in the LIMD1 166-260 spectrum on addition of PHD2. ....	156
Figure 4.21. A heatmap showing the sequence conservation of human LIMD1 residues across all mammalian LIMD1 orthologs annotated in the NCBI database (214 sequences). ....	157
Figure 4.22. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HMQC spectra for $\text{U-}^{15}\text{N}$ -LIMD1 166-260 (grey), $\text{U-}^{15}\text{N}$ -LIMD1 166-260 with PHD2 <sub>181-426</sub> (red), and natural abundance LIMD1 168-191 (blue).....	159
Figure 4.23. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HMQC spectra of unlabelled LIMD1 168-191 in grey and of a mixture of unlabelled LIMD1 168-191 and PHD2 181-426 in red.....	160
Figure 4.24. The structure of PHD2 (PDB 2G1M) with <b>A.</b> assigned backbone nitrogen atoms shown as green spheres and <b>B.</b> unassigned backbone nitrogen atoms shown as red spheres. ....	162
Figure 4.25. 2D ( $^1\text{H}$ - $^{15}\text{N}$ ) TROSY spectra of $[\text{U-}^{15}\text{N}$ , $^2\text{H}]$ , ( $[\delta^1\text{L}^{\text{proSVproS}}]$ )- $^{13}\text{C}^1\text{H}_3$ -PHD2 <sub>181-426</sub> with (red) and without (black) LIMD1 168-191.....	164
Figure 4.26. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) TROSY spectra of $[\text{U-}^{15}\text{N}$ , $^2\text{H}]$ , ( $[\delta^1\text{L}^{\text{proSVproS}}]$ )- $^{13}\text{C}^1\text{H}_3$ -PHD2 <sub>181-426</sub> in black with unlabelled LIMD1 168-191 in red, with unlabelled HIF-1 $\alpha$ CODD in blue and with both unlabelled LIMD1 168-191 and unlabelled HIF-1 $\alpha$ CODD in yellow.....	166
Figure 4.27. 2D ( $^1\text{H}$ , $^{13}\text{C}$ ) HSQC spectra of $[\text{U-}^{15}\text{N}$ , $^2\text{H}]$ , ( $[\delta^1\text{L}^{\text{proSVproS}}]$ )- $^{13}\text{C}^1\text{H}_3$ -PHD2 <sub>181-426</sub> in black with LIMD1 168-191 in red, with HIF-1 $\alpha$ CODD in blue and with both LIMD1 168-191 and HIF-1 $\alpha$ CODD in yellow. ....	167
Figure 4.28. Chemical shift perturbations from the 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) TROSY spectra of PHD2 <sub>181-426</sub> on addition of LIMD1 168-191.....	169
Figure 4.29. The crystal structure of PHD2 (blue) with the backbone nitrogen of residues perturbed on addition of LIMD1 168-191 shown as red spheres. ....	171
Figure 4.30. <b>A.</b> AlphaFold2 models for the interaction of PHD2 with HIF CODD and LIMD1 168-191. <b>B.</b> The structure of PHD2 bound to a RaPID derived cyclic peptide. <b>C.</b> results of HPEPDOCK modelling of LIMD1 168-191 binding to PHD2. ....	173



Figure 5.1. Metabolic precursors that can be used for isotopic labelling and reverse labelling of the carbon sites in branched chain aliphatic amino acids..	183
Figure 5.2. Metabolic precursors used for isotopic labelling and reverse labelling of phenylalanine, tryptophan and tyrosine.	185
Figure 5.3. Histidine (un)labelling by the metabolic precursor Imidazolepyruvate with incorporated atoms shown in red.	186
Figure 5.4. Methionine (un)labelling by methylthio-2-oxobutanoate with isotopically labelled sites indicated in red.	188
Figure 5.5. Schematic representation of <b>A.</b> standard [U- <sup>13</sup> C, <sup>15</sup> N] labelling of a protein and collection of severely overlapping 2D ( <sup>1</sup> H, <sup>1</sup> H) NOESY spectrum <b>B.</b> The proposed unlabelling and 2D filtered NOESY approach to generate a simplified NOESY spectrum.	191
Figure 5.6. Descriptions of the transfer pathways of the various 2D filtered/edited NOESY experiments used with schematic representation of the anticipated patterns of peaks for each experiment.	194
Figure 5.7. <b>A.</b> <sup>15</sup> N-HSQC spectra of [U- <sup>13</sup> C, <sup>15</sup> N]-SHIRT-R3 in black with K unlabelled SHIRT-R3 in red. <b>B.</b> <sup>13</sup> C-HSQC spectra of [U- <sup>13</sup> C, <sup>15</sup> N]-SHIRT-R3 in black with LV unlabelled SHIRT-R3 in red.	198
Figure 5.8. <sup>15</sup> N incorporation for SHIRT-R3 HL, K, FK and W unlabelled samples.	199
Figure 5.9. F1fF2f and F1f NOESY spectra for FK unlabelled SHIRT-R3 in brown and pink respectively compared to a full 2D ( <sup>1</sup> H, <sup>1</sup> H) NOESY spectrum in grey.	201
Figure 5.10. <b>A.</b> <sup>15</sup> N and <b>B.</b> aliphatic <sup>13</sup> C isotope incorporation for SasYr unlabelled samples.	204
Figure 5.11. Examples of 2D filtered/edited NOESY spectra for the LVW unlabelled SasYr sample compared to a full 2D ( <sup>1</sup> H, <sup>1</sup> H) NOESY spectrum in grey.	206
Figure 5.12. Description of the transfer pathway for the different filtered/edited TOCSY and NOESY experiments with a focus on equivalent strips from each spectrum to demonstrate the experiments are working as intended.	208
Figure 5.13. A flowchart demonstrating the process for incorporating residue type information obtained from filtered/edited NOESY experiments into a single peak list.	212
Figure 5.14. Analysis of the convergence of SasYr structure calculation with and without LVW unlabelled data.	216

Figure 5.15. A graph of the average backbone RMSD across the ensemble for residues 27-109 for each iteration of five structure calculations for SasYr with a different 5% of the 3D ( $^1\text{H}$ , $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peaks randomly removed. The structure calculations were performed with LVW data (orange) and without (Blue).....	218
Figure 5.16. The results of structure calculations of SasYr with 75% of the 3D ( $^1\text{H}$ , $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peaks removed. Structure calculations were performed without any additional peaks added, with the LV unlabelled data added, and with the same number of peaks present in the LV data added from the previously removed 3D ( $^1\text{H}$ , $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peaks. ....	219
Figure 7.1. SEC chromatograms for molecular mass standards on a superdex increase 200 10/300 GL column. Used for FL-PHD2 with LIMD1 SEC experiments. ....	231
Figure 7.2. SEC chromatograms for molecular mass standards on a superdex increase 75 10/300 GL column. Used for PHD2 181-426 with LIMD1 166-260 and VHL with LIM2 SEC experiments. ....	232
Figure 7.3. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HSQC spectra of 100 $\mu\text{M}$ [ $\text{U}-^{15}\text{N}$ ]-LIMD1 65-260 with (grey) and without (red) PHD2 181-426.....	233
Figure 7.4. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HSQC spectra of 100 $\mu\text{M}$ [ $\text{U}-^{15}\text{N}$ ]-LIMD1 160-300 with (grey) and without (red) PHD2 181-426.....	234
Figure 7.5. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HSQC spectra of 100 $\mu\text{M}$ [ $\text{U}-^{15}\text{N}$ ]-LIMD1 110-166 with (red) and without (grey) PHD2 181-426.....	235
Figure 7.6. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HMQC spectra of [ $\text{U}-^{15}\text{N}$ ]-LIMD1 166-260 with (red) and without (grey) PHD2 181-426.....	236
Figure 7.7. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HMQC spectra for 100 $\mu\text{M}$ [ $\text{U}-^{15}\text{N}$ ]-LIMD1 166-260 with (red) and without (grey) 250 $\mu\text{M}$ PHD2 181-426 and 250 $\mu\text{M}$ HIF-1 $\alpha$ CODD peptide.....	237
Figure 7.8. strips from the HNCACB and HNCOCACB spectra of LIMD1 166-260 for the assigned region of 231-235.....	238
Figure 7.9. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) TROSY spectrum of [ $\text{U}-^{15}\text{N}$ , $^2\text{H}$ ],( $[\delta^{13}\text{C}^{\text{proSVproS}}]$ )-[ $^{13}\text{C}^1\text{H}_3$ ]-PHD2 <sub>181-426</sub> overlaid with the 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HMQC spectra of unlabelled LIMD1 168-191 with PHD2 181-426.....	239
Figure 7.10. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HSQC spectra for SHIRT-R3 samples .....	240
Figure 7.11. 2D ( $^1\text{H}$ , $^{13}\text{C}$ ) HSQC spectra of SHIRT-R3 samples .....	241

Figure 7.12 a 2D full ( $^1\text{H}$ , $^1\text{H}$ ) NOESY for the $[\text{U-}^{13}\text{C}$ , $^{15}\text{N}]$ -SHIRTr sample and 2D F1fF2f and F1f NOESY spectra for all SHIRT-R3 samples .....	243
Figure 7.13. 2D ( $^1\text{H}$ , $^{15}\text{N}$ ) HSQC spectra of SasYr samples .....	244
Figure 7.14. 2D ( $^1\text{H}$ , $^{13}\text{C}$ )-HSQC spectra of SasYr samples .....	245
Figure 7.15. 2D NOESY spectra of $[\text{U-}^{13}\text{C}$ , $^{15}\text{N}]$ -SasYr .....	246
Figure 7.16. 2D NOESY spectra of LV unlabelled SasYr .....	247
Figure 7.17. 2D NOESY spectra of K unlabelled SasYr .....	248
Figure 7.18. 2D NOESY spectra of LVK unlabelled SasYr .....	249
Figure 7.19. 2D NOESY spectra of LVW unlabelled SasYr .....	250
Figure 7.20. Examples of ambiguous restraints generated for LV unlabelled SasYr .....	251
Figure 7.21. A region of an ambiguously assigned peak list in NMRView format, as used for inputs for structure calculations with ARIA. ....	252
Figure 7.22. Graphs showing the intensity change for peaks in the spectra of PHD2 on addition of LIMD1 168-191. ....	253
Figure 7.23. SDS-PAGE analysis of the SHIRT-R3 samples used. ....	254
Figure 7.24. SDS-PAGE analysis of the purity of SasYr samples. ....	255
Figure 7.25. Statistics for the AlphaFold 2 prediction of TNRC6A 1-900 with LIM12 (LIMD1 467-594). ....	256
Figure 7.26. Statistics for the AlphaFold2 prediction of AGO2 519-779 (TNRC6A binding region) with TNRC6A 695-773 and LIM12 (LIMD1 467-594). ....	257
Figure 7.27. Statistics for the AlphaFold2 modelling of PHD2 181-426 with LIMD1 168-191 and HIF CODD. ....	258
Figure 7.28. The sequence of human VHL 54-213 produced in this study with peptides identified by MS in bold and red. ....	259

## Acknowledgements

I would like to thank Michael Plevin and Dimitris Lagos for giving me the opportunity to do this PhD and Michael for his supervision during the project. I would also like to thank Alex Heyam for his help with the NMR spectroscopy and Mike Hodgekinson for keeping the lab running.

I would like to thank my friends, Cameron McElrue, Mark Petchey, Amina Frese, Muhi Omar, Anibal Cuetos, Alex Snow, Rhianna Rowland, Sarah Barrass-Tanner, Beth Jenkins, Callum Silver, Frank, and Dougie for keeping me somewhat sane during this PhD. Thank you to Mahima Sharma for teaching me almost all I know about protein production and crystallography. Thank you to my parents for their endless love and support, I couldn't have made it this far without it.

Most importantly, this endeavour would not have been possible without the support of my wonderful fiancée Wendy Robinson. Your love and belief in me allowed me to overcome the lowest points during this PhD. Thank you for all you have done for me and for being the most amazing partner anyone could hope for.

## Declaration

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

Rachael Cooper generated the SasYr resonance assignments and 3D NOESY data, Lotte van Beek generated the SHIRT-R3 backbone resonance assignments, Alex Heyam modified the F1f and F1eF2f NOESY pulse sequences. Specific methyl labelling compounds were provided by NMR-Bio. Mass spectrometry experiments were carried out by the technology facility at the University of York. The AGO2 sample was supplied by Professor Ian MacRae (Scripps Institute, USA)

Some of the content in Chapter 5 was published in Biochemical Society Transactions [1]:

- Rowlinson B, Crublet E, Kerfah R, Plevin MJ. (2022) Specific isotopic labelling and reverse labelling for protein NMR spectroscopy: using metabolic precursors in sample preparation. *Biochem. Soc. Trans.* 13 <https://doi.org/10.1042/bst20210586>

# 1. Introduction

The ability of cells to respond to environmental pressures and signals is essential for the survival and proliferation of the cell and is common to all cellular life. This process is controlled by signal transduction pathways resulting in the repression or activation of various genes. In complex organisms the spatiotemporal organisation of these signal transduction pathways is essential to maintain the specificity of these pathways to ensure the correct response is achieved. This regulation is achieved through a range of mechanisms including the specificity of protein interactions, enzyme activity, protein abundance, cellular localisation and post-translational modifications. Dysregulation of these pathways often result in diseases such as cancer and diabetes. One of the key factors in the regulation and spatial organisation of signalling pathways is the behaviour of scaffold proteins.

## 1.1. Scaffold Proteins

Scaffold proteins bind to and co-localise multiple proteins with related functions and have been identified as key components in many signalling pathways [2]. One of the major mechanisms through which scaffold proteins facilitate cellular signalling pathways is the co-localisation of signalling components. Co-localisation of the signalling components increases their local concentration and thus the efficiency of the signalling processes. The effect of local concentration on signalling events has been investigated in depth using kinases tethered to substrates with different linker lengths [3]. This effect allows for regulation of signalling processes through modifying scaffold protein expression levels, post-translational modification or cellular localisation. Some scaffold proteins also improve efficiency of signalling processes by specifically orientating the partner proteins to facilitate the interaction. For example, this is thought to be the case for E3 ubiquitin ligase complexes scaffolded by Cullin 1 (Cul1) [4]. It was seen that introducing mutations into the Cul1 linker to increase flexibility of the scaffold inhibited substrate ubiquitination *in vitro*, suggesting the rigidity of the Cul1 scaffold allowed for specific orientation of the partner proteins. Scaffold proteins can also modulate signalling through allosteric effects such as the

activation of rapidly accelerated fibrosarcoma kinase (RAF) by binding of the kinase suppressor of RAS (KSR) scaffold protein, at a site distal to the active site, in the extracellular signal-regulated kinase (ERK) pathway [5].

Many scaffold proteins have been shown to be intrinsically disordered or feature intrinsically disordered regions (IDRs) [6]. IDRs have been predicted to occur in 33% of eukaryotic proteins, representing a significant proportion of the proteome [7]. Intrinsically disordered proteins (IDPs) or regions are characterised as having a low sequence complexity and are typically enriched in polar and charged residues whilst lacking a high proportion of bulky hydrophobic residues [8]. Rather than adopting a fixed 3D fold, associated with the classical model of protein activity, IDPs are highly flexible and exhibit a conformational ensemble [9]. Many IDPs adopt a more rigid structure upon binding to a partner protein, which may also be represented in the unbound structural ensemble. This allows for specificity to be conveyed by conformational selection in which a protein binds to a higher energy state in the conformational ensemble of an IDP rather than inducing a conformational change on binding (as in induced fit) [10]. Conformational selection was shown to determine the affinity of p53 binding to mouse double minute 2 homolog (MDM2) [11]. p53 was shown to adopt a helical conformation in the structure of p53 bound to MDM2 [12]. It was demonstrated that mutations that increased the helicity of the MDM2 binding region of p53 also increased the affinity of the interaction, which demonstrated the importance of conformational selection in this case. Some intrinsically disordered binding regions have been demonstrated to bind to multiple partner proteins by adopting different structures in the bound state. This can be seen in residue 374 to 388 of p53, which can interact with S100 $\beta$  with a  $\beta$ -helical structure, sirtuin with a  $\beta$ -sheet structure and CREB-binding protein (CBP) or Cyclin A2 with two different conformations (Figure 1.1) [13-17].

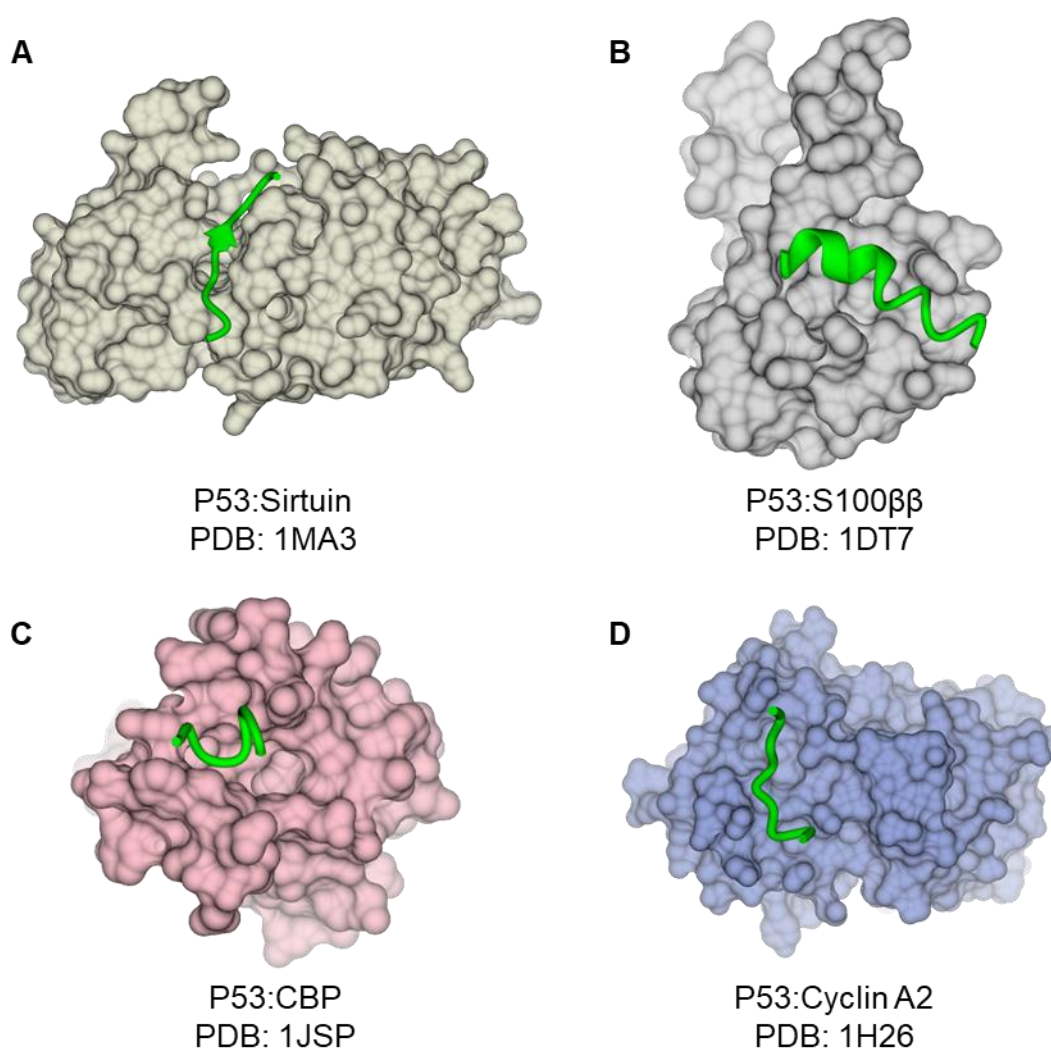


Figure 1.1. **A.** p53 372-389 (green) bound to Sir2-Af2 (beige) [15]. **B.** p53 375-388 (green) bound to S100β (grey) [14]. **C.** p53 380-386 (green) bound to a CBP bromodomain (pink) [16]. and **D.** p53 376-386 (green) bound to Cyclin A2 (blue) [17].

Although many IDPs adopt a single, specific, structure on binding to partner proteins it is thought that other interactions involve the IDP remaining disordered in the bound state [18]. The potential of IDPs to bind multiple partners at the same site, as well as the extended, exposed nature of IDPs, allows for a much higher concentration of binding sites than is possible with a persistent 3D fold. This potentially high concentration of binding sites makes IDRs highly efficient for scaffold proteins [6]. Additionally, the high flexibility of IDRs allows for interaction of partner proteins which can be fine-tuned by adjusting the distance between the protein binding sites [3].



More recently intrinsically disordered scaffold proteins have been heavily implicated in the formation and organisation of phase condensates [19-21]. This current study focusses on investigating the scaffold protein LIMD1. LIMD1 functions in a diverse range of signalling pathways and is believed to be comprised of three C-terminal LIM domains and a largely disordered N-terminal pre-LIM region [22, 23]. Study of the structure and biophysics of the protein-protein interactions formed by LIMD1, at both disordered regions and structured domains, will contribute to our understanding of scaffold protein biology.

## 1.2. LIM Domains are Protein Binding Motifs with Diverse Modes of Binding

LIMD1, the focus of this thesis, is predicted to contain three C-terminal tandem LIM domains which have been shown to be important for binding to various partner proteins [22, 24, 25]. LIM domains are protein binding motifs comprised of two zinc fingers separated by a short linker [26, 27]. LIM domains were originally identified in *lin-11*, insulin gene enhancer protein 1 (*isl-1*) and *mechanosensory protein 3* (*mec-3*), for which the LIM domain is named, and are typically 50-63 residues long with a  $C(X)_2C(X)_{16-23}(H/C)X_{2/4}(C/H/E)(X)_2C(X)_2C(X)_{14-21}(C/H)(X)_{2/1/3}(C/H/D/E)X$  sequence motif (Figure 1.2**A-B**) [23, 28-31].

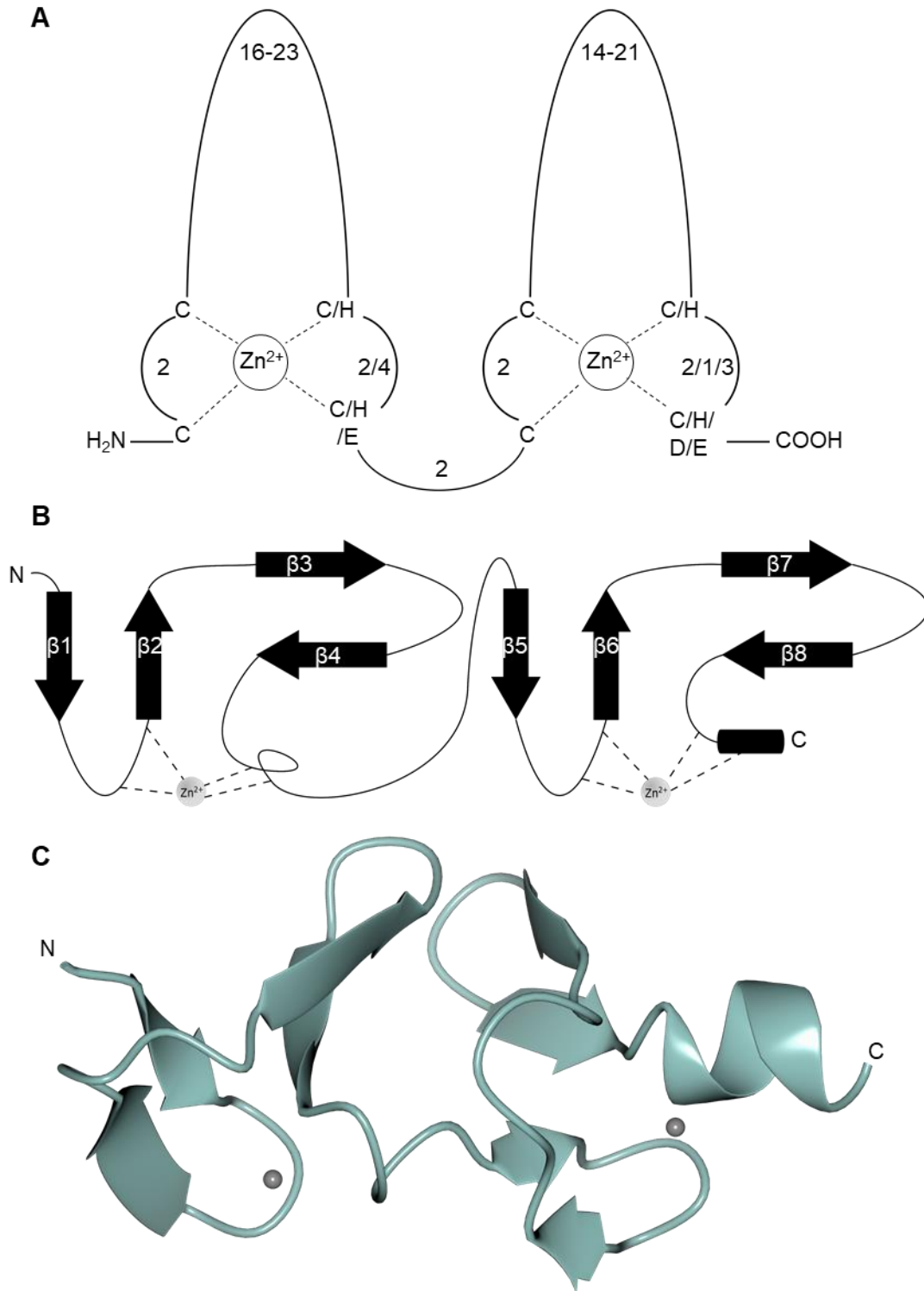


Figure 1.2. **A.** Schematic representation of a LIM domain with zinc binding residues and the number of residues between them indicated. **B.** Secondary structure diagram of a LIM domain. **C.** Solution NMR structure of CRP1. Model 1 of the ensemble is shown. PDB code: 1CTL

So far LIM domains have only been reported to function in protein-protein interactions, despite initial classification as DNA binding proteins, based on superposition of Cysteine-rich protein 1 (CRP1) LIM2 (Figure 1.2C) with the DNA binding domain of GATA-1 [32]. LIM domains are abundant in signalling proteins and are found in a number of different pathways including organ development, cytoskeleton organization and cell lineage specification [33]. Their importance in these pathways is highlighted by their relevance in a range of diseases including heart disease and various cancers [34, 35]. Evolutionary analysis reveals that LIM domain evolution likely played an important role in the development of animal multicellularity [36].

Classification of LIM domains has evolved over time, with five classes originally proposed by Dawid *et al.* (1995) [37]. Improved access to genomic-scale sequencing data allowed this to be expanded to 14 classes by Koch *et al.* (2012), based on phylogenetic groupings, domain architecture and motif signatures [36]. LIM domain proteins were thus classified as ABLIM, CRP, ENIGMA, EPLIN, LASP, LIMK, LHX, LMO, LMO7, MICAL, PXN, PINCH, TES or ZYX. Proteins in these classes are differentiated by their LIM, and non-LIM, domain composition as well as their evolutionary history. LIMD1 has been identified as a member of the zyxin (ZYX) class of proteins, all of which feature a variable N-terminal pre-LIM region followed by 3 closely linked C-terminal LIM domains [23]. This class of LIM domains is named after zyxin which is a LIM domain protein involved in actin polymerization at focal adhesions [38].

LIM domains are comprised of two type VII treble clef zinc fingers (Figure 1.2B-C) that have been compared to rubredoxin knuckles based on the similarity of the NOE patterns to those observed for rubredoxin from *P. furiosus* [32, 39]. The two zinc atoms appear to form an essential structural element that allows for the formation of a complex three-dimensional structure in a relatively short stretch of protein. The zinc atoms, bound by cystine residues located in the first  $\beta$ -hairpin of each zinc-finger and by cystine/histidine/aspartate residues located towards the C-terminal of each zinc-finger, do not appear to have any functional significance other than as

structural elements. The two zinc fingers stack on top of one another to form a small, conserved hydrophobic core. LIM domains are often present as tandem domains featuring a short flexible linker region [36].

An analysis of the available structures of LIM domains in complex with partner proteins (Table 1.1) reveals some interesting insight into the function of these domains (Figure 1.3). Interestingly, LIM domains do not appear to bind to a conserved recognition sequence but instead display groups of residues that facilitate recognition of their individual binding partners. LIM domains have been shown to bind both to proteins with a defined 3-dimensional fold and to proteins with more linear motifs. An analysis of protein complexes containing LIM domains (Table 1.1) reveals that the LIM domain employs distinct binding faces for linear and folded binding partners (Figure 1.3). This can be most clearly demonstrated through the structures of the third LIM domain of TES bound to the structured EVH1 (Ena/VASP Homology 1) domain of MENA (Mammalian-enabled protein) at the same time as the linear ARP7A (Actin-like protein 7A) motif, with both partner proteins bound to opposite faces of LIM domain (Figure 1.3B) [40]. The binding faces for LIM domains interacting with folded and linear partners was consistent for all of the available structures, irrespective of the number of LIM domains bound, the nature of the interaction with the folded partner, or the signalling pathways involved. Different modes of binding observed for linear and folded partners are explored below.

Table 1.1. Unique LIM domain interaction structures used for analysis of LIM domain interactions [40-47].

PDB code	LIM domain	Structured partner(s)	Linear Partner	Method
2YPA [41]	LIM domain only 2 (LMO2) LIM1 and 2	SCL, E47	LIM domain binding protein 1 (LDB1)	x-ray
2XQN [40]	Testin (TES) LIM1 and 2	MENA	ARP7A	x-ray
3IXE [42]	Particularly interesting new Cys-His protein 2 (PINCH2) LIM1	Integrin-linked protein kinase (ILK)	-	x-ray
3F6Q [48]	Particularly interesting new Cys-His protein 1 (PINCH1) LIM1	ILK	-	x-ray
1U5S [43]	PINCH1 LIM4	non-catalytic region of tyrosine kinase adaptor protein 2 (NCK-2) SRC homology 3 (SH3)	-	NMR
2RGT [44]	LIM homeobox protein 3 (LHX3) LIM1 and 2	-	ISL-1	x-ray
4JCJ [45]	ISL-1 LIM1 and 2	-	LDB1	x-ray
2MBV	LIM domain only 4 (LMO4) LIM2	-	Deformed epidermal autoregulatory factor 1 homolog (DEAF1)	NMR
1RUT [46]	LMO4 LIM1 and 2	-	LDB1	x-ray
3MMK [49]	LIM homeobox protein 4 (LHX4) LIM1 and 2	-	insulin gene enhancer protein 2 (ISL-2)	x-ray
2L4Z [47]	LMO4 LIM1	-	CtBP-interacting protein (CtIP)	NMR

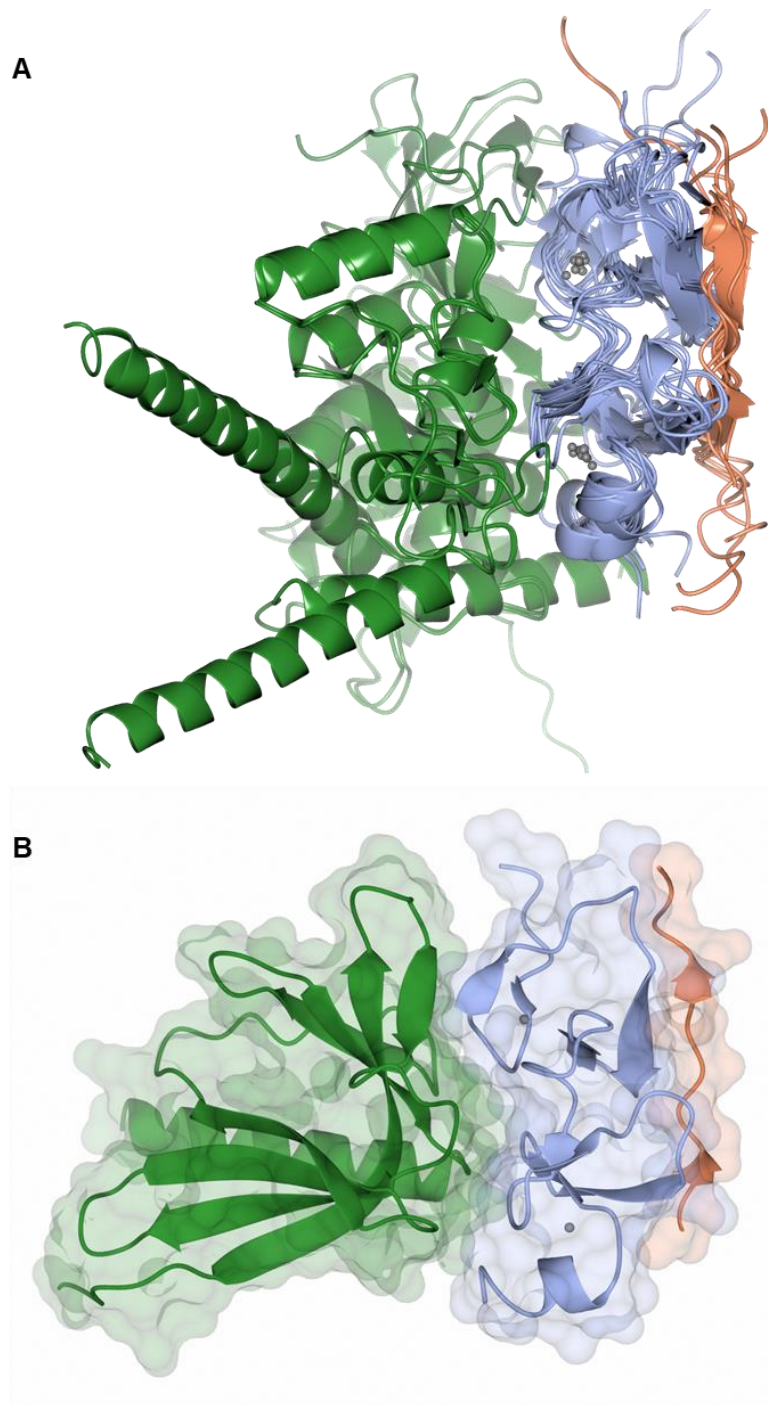


Figure 1.3. **A.** Superposed structures of LIM domains bound to partner proteins [40-47]. LIM domains are shown in blue, folded partners in green and linear motifs in orange. Only unique interactions are shown and only one model of ensemble structures is shown. PDB codes: 2YPA, 2XQN, 3IXE, 3F6Q, 1U5S, 2RGT, 4JCJ, 2MBV, 1RUT, 3MMK, 2L4Z. Superpositions were achieved by secondary structure matching in CCP4mg [50]. **B.** The structure of the third LIM domain of TES (blue) in complex with the EVH1 domain of MENA (green) and ARP7A (orange).

The linear motifs LDB1-LID, ARP7A and ISL-2 are seen bound across two tandem LIM domains, whereas the ISL1, DEAF1 and CtIP linear motifs are only seen to bound to a single LIM domain [40, 46, 47, 49, 51, 52].

The linear motifs, LDB1-LID, ARP7A N-terminal, ISL1, DEAF1 and CtIP were all found to bind in a similar fashion to the LIM domain scaffold. The linear motifs and LIM domains are always positioned anti-parallel to one another and interact *via* a tandem  $\beta$ -zipper formed between the linear motif and  $\beta$ -sheet 3 and 7 of the LIM domains (Figure 1.4) [53]. Specificity then appears to be conveyed through a small number of hydrophobic and electrostatic interactions between the LIM domain protein and the linear motif. It has been observed that binding to a linear motif *in vitro* can stabilise the LIM domain and improve yields from recombinant protein expression [49, 54, 55].

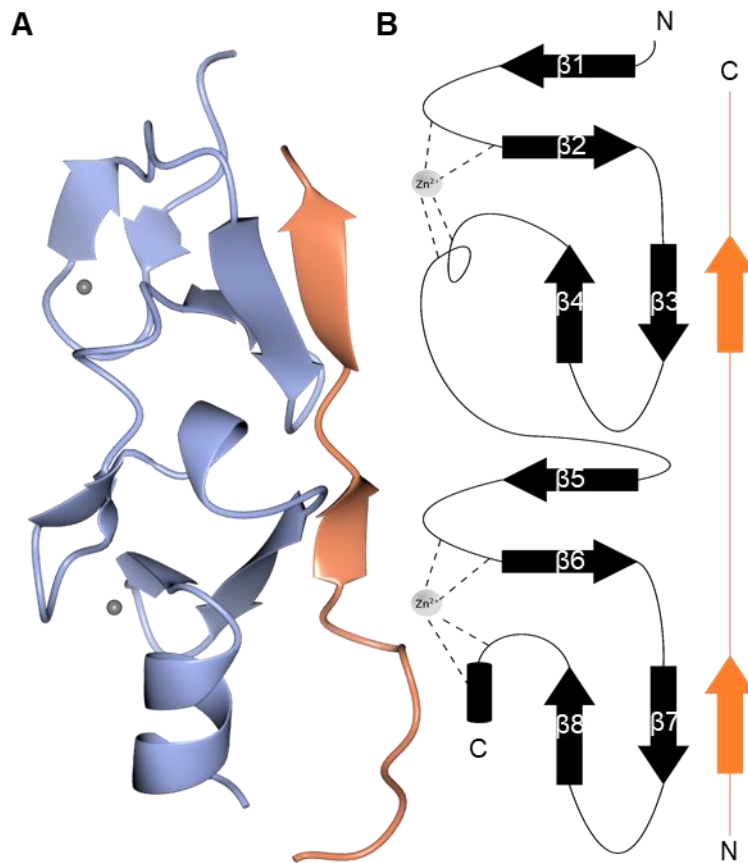


Figure 1.4. **A.** Structure of LMO4 LIM1 (blue) bound to ldb1-LID (orange) PDB code: 1RUT [46]. **B.** Secondary structure diagram showing the  $\beta$ -zipper binding mode of LIM domains (black) with linear motifs (orange).

Comparing the structures of tandem LIM domains in complex with linear motifs shows significant differences in the relative orientations of neighbouring LIM domains (Figure 1.5). A comparison of the structures of tandem LIM domains from LMO2, LMO4, TES and LHX3 shows the relative orientation of the two domains can vary by  $100^\circ$ . In the case of LMO2 and LMO4, which are bound to the same linear motif with a similar LIM domain orientation, it is possible that the relative position of tandem LIM domains is important for recognition of linear motifs, although greater study into the impact of the interdomain linker in tandem LIM domains would be required to confirm this [41, 46]. A R282G mutation in ISL-2 linear motif was seen to decrease affinity for the LHX4 LIM domains by increasing the flexibility of ISL-2 [51]. This suggests that shape complementarity of the linear motif for tandem LIM domains may be an important factor partner recognition by LIM domains. The relative domain position of



tandem LIM domains may be one of the aspects that conveys specificity for the linear motifs.

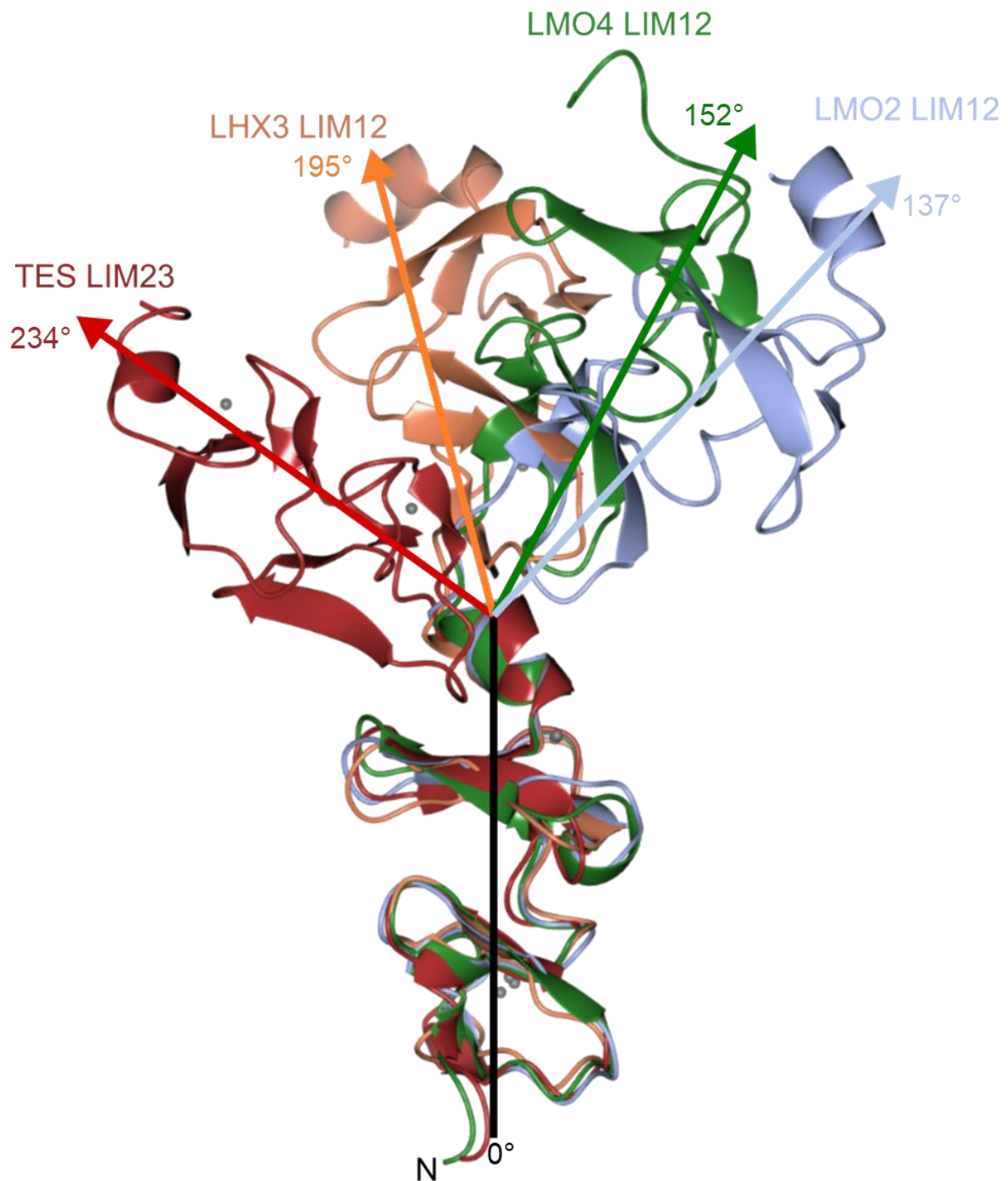


Figure 1.5. Structure of LMO2 LIM12 (blue) in complex with LDB1, LMO4 LIM12 (green) in complex with LDB1, LHX3 LIM12 (orange) in complex with ISL-1 and TES LIM23 (red) in complex with ARP7A and MENA superposed by secondary structure matching on the first LIM domain. The positions of the second LIM domain with respect to the first LIM domain is shown by the coloured arrows with the anticlockwise interdomain angle indicated.

LIM domain binding to folded protein partners occurs on the opposite side of the LIM domains to linear motif binding. The majority of interacting residues are localised on the first hairpin of the first zinc-finger, both hairpins in the second zinc-finger and the C-terminal helix of the LIM domain. Some partners, such as the SH3 domains of Nck-2, only interact with a small region of the LIM domain with only 480 Å<sup>2</sup> of total buried surface area. Other partners, such as the ankyrin repeats of ILK, interact across almost the full length of the LIM domain, resulting in a total buried surface area of 1893 Å<sup>2</sup> for the ILK:PINCH1 LIM1 interaction [42, 43]. Folded proteins are seen to interact with one or both zinc fingers and with multiple or single LIM domains. For some of these interacting partners, the LIM domain interaction site provides a novel mode of interaction. SH3 domains and EVH1 (Ena/Vasp homology domain 1) domains canonically bind to PxxP and FPPPP motifs respectively, whereas their interactions with LIM domains feature no such motifs [43, 56]. LIM domains appear to provide a short scaffold able to present clusters of amino acids on its surface to recognise specific partner proteins. The lack of a conserved mechanism of binding or conserved sites or motifs for folded partners across the available 3D structures demonstrates promiscuity of LIM domains as interaction scaffolds. Additional structures of LIM domains bound to a more diverse range of partner protein would hopefully reveal some common recognition themes that could be used to inform future studies of LIM domains.

The collation and analysis of the available LIM domain structures presented here provides a valuable resource to aid in the research of LIM domain structural biology. The analysis identified a previously unrecognised pattern of LIM domain partner binding (distinct interaction sites for linear and globular motifs) and highlighted the large variety in binding modes for LIM domains interacting with globular partners. This study aimed to provide additional insight into the structural biology and biophysics of LIM domain partner recognition by characterising the protein:protein interactions made by the LIM domains of LIMD1, which are known to bind to both globular and linear partners [24, 25]. It was hoped that in vitro binding assays could

reveal the stoichiometry and competitive or cooperative nature of LIMD1 LIM domains binding to linear and globular partners.

### 1.3. The LAW Family of LIM Domain Scaffold Proteins

The LAW (LIMD1, Ajuba and WTIP) family of proteins were identified as a family of LIM domain proteins with related function [57]. The LAW proteins are all zyxin-like proteins, which are characterised as having a variable N-terminal region followed by three C-terminal tandem LIM domains [23, 36]. The LAW proteins are more closely related to one another than they are to the other members of the zyxin proteins (Figure 1.6) and have related functions in a number of pathways [24, 57, 58].

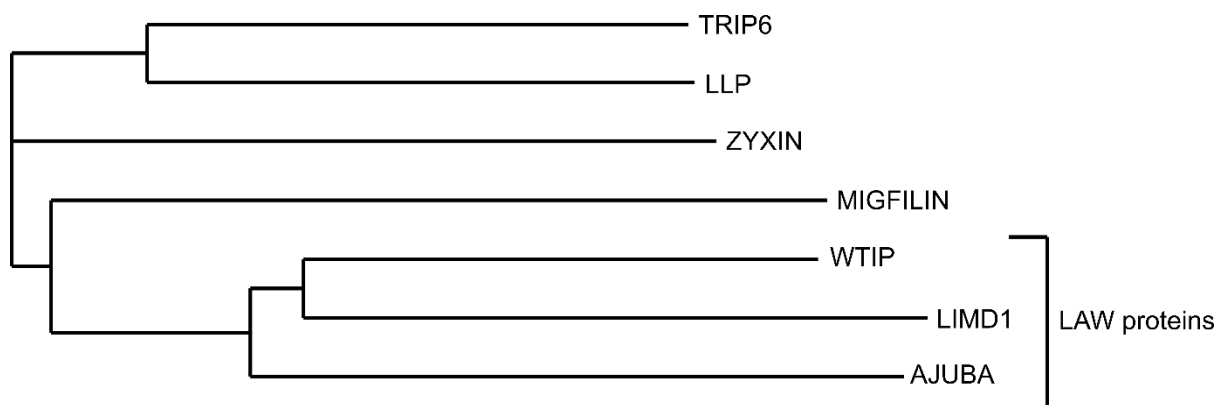


Figure 1.6. Phylogenetic tree showing the evolutionary distance between members of the zyxin protein family with the LAW subfamily indicated. Phylogenetic tree was generated by simple phylogeny using a neighbour-joining method from a multiple sequence alignment generated by Clustal Omega [59, 60].

Although other zyxin family members can contain additional folded domains in their N-terminal region, the N-terminal region of the LAW proteins, which is referred to as the pre-LIM region, is predicted to be disordered. The LAW proteins have high sequence similarity in their LIM domain region but very little similarity in either the sequence or the length of the pre-LIM region (Figure 1.7). The lack of sequence conservation in the pre-LIM raises questions about how recognition of common partner proteins that bind this region is achieved. This study aimed to characterise

the mode of partner protein recognition in the pre-LIM region of LIMD1, which may explain differences LAW protein function and utilisation [25]. The LAW proteins have been shown to function in a diverse range of signalling pathways including cell fate determination, cytoskeletal organisation, mitosis, focal adhesion, and migration [58, 61-64]. This study is focussed on investigating these proteins in the context of miRNA mediated silencing and hypoxic signalling with a focus on LIMD1 and the molecular basis of its interaction with its binding partners.



Figure 1.7. Multiple sequence alignment of human LIMD1, AJUBA and WTIP. Alignment was performed using Clustal Omega and visualised using jalview [60, 65]. Vertical bars represent the level of conservation. The location of LIM domains, predicted by Pfam, are also indicated [66].

#### 1.4. The LIMD1 Scaffold Protein

LIMD1 is a scaffold protein with tumour suppression properties, loss of which has been implicated in a range of cancers [67-72]. The importance of LIMD1 was demonstrated through LIMD1 knockout mice which were seen to develop more and larger tumours on exposure to a chemical carcinogen or crossbreeding with mice with an oncogenic mutation [68]. It has also been demonstrated that silencing of LIMD1 expression in colorectal carcinoma and non-Hodgkins lymphoma can reverse drug resistance, showing a potential case for LIMD1 as a therapeutic target [73, 74]. Given the tumour suppressing nature of LIMD1 it is surprising that silencing of LIMD1 leads to a reversal of drug resistance. This may be explained by a difference in the roles of LIMD1 in the different types of cancers being studied or an effect on the mechanisms of drug resistance in these cancers.

LIMD1 has been shown to bind to retinoblastoma protein (pRB) and influence cell proliferation by shuttling pRB to the nucleus to repress E2F-driven transcription [67]. LIMD1 has also been shown to: influence osteoclast development through an interaction with Traf6 [75, 76]; regulate Hippo signalling through interactions with SKI and LATS2 [58, 77, 78]; determine localisation of BRCA2 to the centrosome during cell proliferation [79]; and bind to snail proteins to downregulate E-cadherin expression in the snail/slugs pathway [80]. More recently a role for LIMD1 has emerged as a regulator of phase separation, particularly in focal adhesions under force, actin filament organisation, and processing body (P-body) formation [57, 63, 81].

Like other LIM proteins, LIMD1, is predicted to be comprised of a disordered N-terminal region called the pre-LIM followed by three C-terminal LIM domains (Figure 1.8). Proteins have been shown to bind to both the pre-LIM and LIM domain regions of LIMD1 [24, 25]. This project investigated the protein:protein interactions formed by LIMD1 in the context of miRNA mediated silencing and hypoxic signalling pathways. This included investigation of structured proteins, binding to the disordered pre-LIM

region of LIMD1 as well as both structured and disordered proteins binding to the LIM domain region of LIMD1.

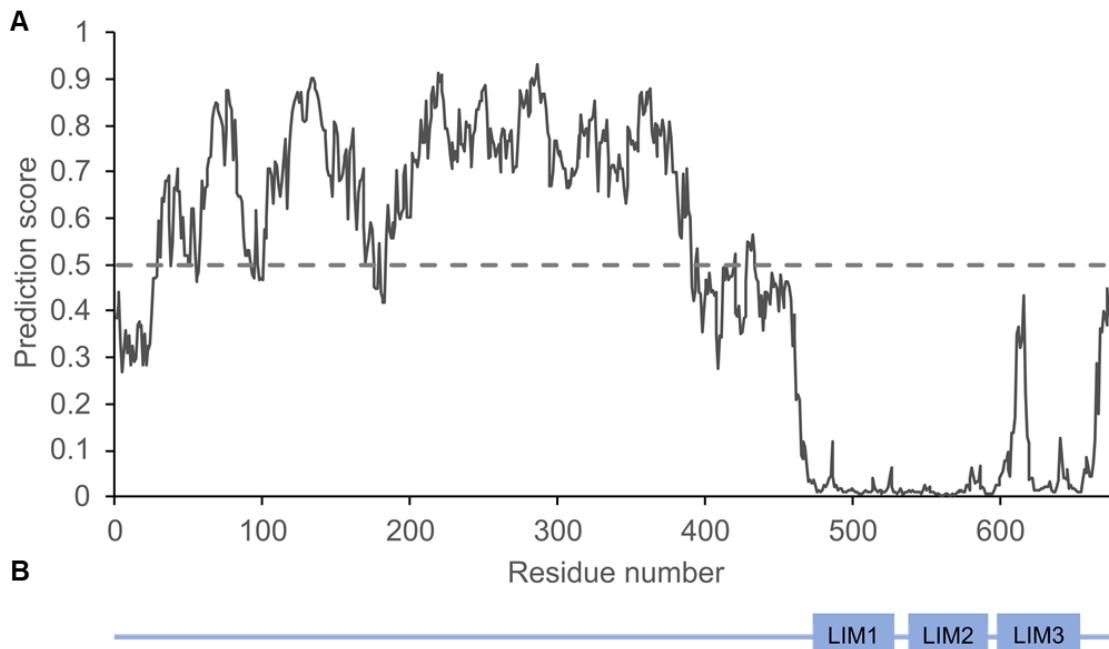


Figure 1.8. **A.** Disorder prediction for LIMD1 by IUPred3 [82]. The calculation was carried out using long disorder analysis mode. Scores  $>0.5$  indicate probable disorder and scores  $<0.5$  indicate potential structure. **B.** Domain composition of LIMD1 to demonstrate alignment of domains with the disorder prediction.

## 1.5. miRNA Mediated Gene Silencing

miRNAs are small non-coding RNAs, approximately 22 nucleotides in length, that act to silence the expression of certain genes by targeting mRNA for degradation and/or by repressing translation of the target mRNA [83]. In the canonical model of miRNA biogenesis, miRNAs are first transcribed and processed into pre-miRNA by DiGeorge Syndrome Critical Region 8 (DGCR8) and Drosha [83, 84]. Drosha, an RNase III enzyme, cleaves the primary miRNA to produce a pre-miRNA hairpin that is then cleaved by a second RNase III enzyme, DICER, to remove the terminal loop and produce a mature miRNA duplex. A single strand of mature miRNA can then be loaded into one of four Argonaute (AGO) proteins [85]. The miRNA directs AGO proteins to target mRNA by complimentary base pairing. AGO2 can directly cleave



perfectly complementary sequences though its slicing activity, although this appears to be an atypical mode of regulation [86]. The majority of miRNA recognition is only partially complementary, and three of the four human AGO proteins lack slicing activity, so regulation is most commonly achieved without direct slicing of the mRNA [86]. AGO proteins achieve this regulation through the recruitment of additional proteins to form the miRNA induced silencing complex (miRISC) (Figure 1.10) [87]. The miRISC regulates gene expression through repressing translation and/or promoting degradation of the target mRNA. The GW182 proteins TNRC6A, TNRC6B and TNRC6C have been shown to bind AGO proteins through their N-terminal AGO binding region and recruit accessory proteins to the target mRNA through a C-terminal silencing domain (Figure 1.10) [87]. GW182 proteins have been shown to be key components of the miRISC, essential for P-body formation and silencing activity [87-89]. The GW182 proteins are largely intrinsically disordered and are rich in glycine-tryptophan repeats. The interaction between GW182 proteins and AGO proteins has been well studied with several AGO binding sites identified on GW182 proteins by different groups using peptide arrays or pull-down assays with GW182 truncations/mutations [90-94]. Structures of AGO2 bound to tryptophans suggested a binding mode for GW182 proteins on AGO2 which was later confirmed by a structure of the hook motif of TNRC6A bound to AGO1 (Figure 1.9) [91, 95, 96].

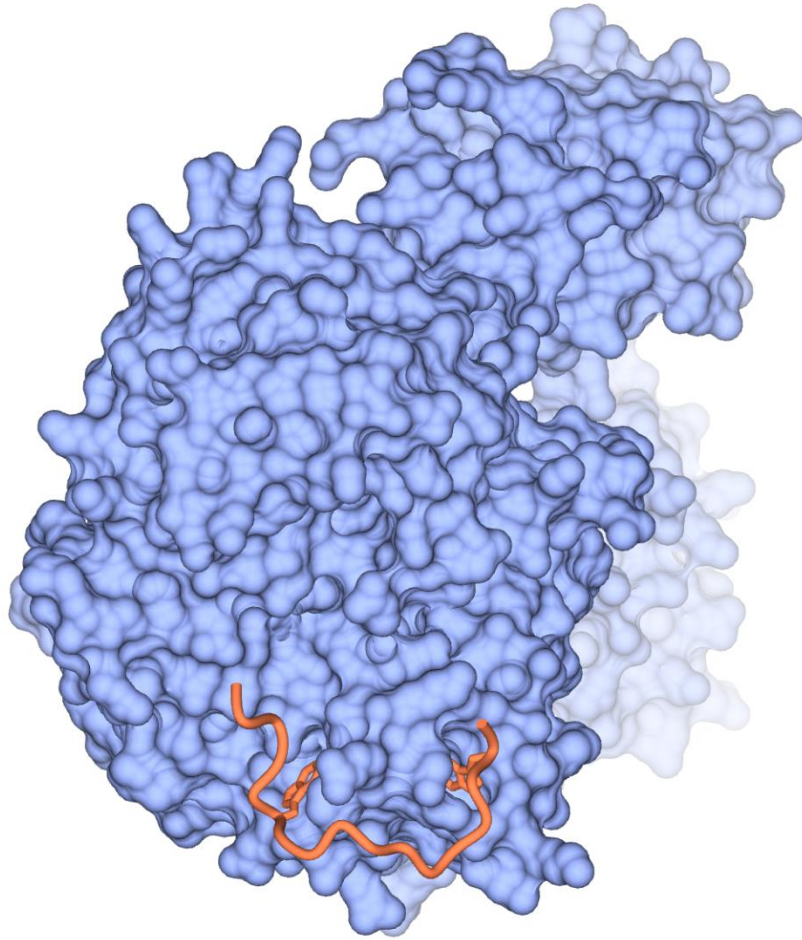


Figure 1.9. Crystal structure of AGO1 (blue) bound to GW182 hook motif (orange) [91]. W828 and W838 are shown inserted into pockets on AGO1. PDB code: 5W6V.

GW182 has been shown to recruit the PAN2/3 and CCR4-NOT deadenylating complexes which can then deadenylate the target mRNA (Figure 1.10) [97-99]. Decapping of the deadenylated mRNA by the mRNA decapping protein 1 (DCP1) and mRNA decapping protein 2 (DCP2) decapping complex (Figure 1.10) then leads to 5'-3' degradation of the mRNA by exoribonuclease 1 (XRN1) [99, 100]. GW182 proteins have also been shown to interact with poly(A) binding protein (PABP) (Figure 1.10) through a PAM2 motif in the GW182 silencing domain. This interaction may contribute to miRNA mediated silencing by directing miRISC components to the poly(A) tail, and/or by interfering with the translation promoting activity of PABP [101-103].

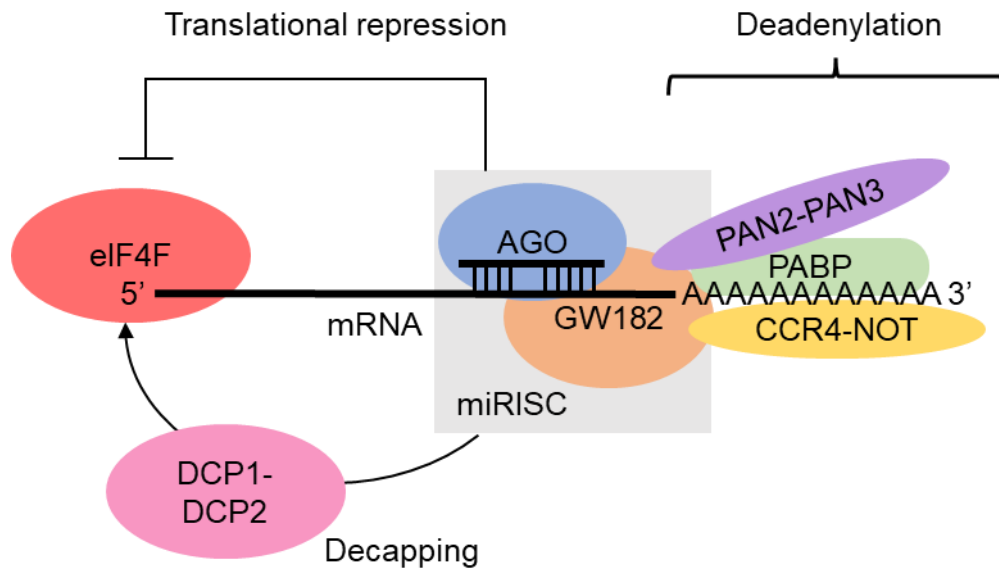


Figure 1.10. Cartoon representation of the miRISC. AGO and GW182 protein comprise the core miRISC complex. GW182 then interacts with PAN2-PAN3 and CCR4-NOT deadenylating complexes and PABP. The miRISC can also bring about decapping of the mRNA by DCP1/DCP2 which results in mRNA degradation. miRISC represses translation through preventing eukaryotic translation initiation factor 4F (eIF4F) assembly and/or activity [87].

## 1.6. The Role of LIMD1 in miRNA Mediated Gene Silencing

Although GW182 proteins have been shown to interact with AGO proteins directly, James *et al.* (2010) identified the LAW (LIMD1, Ajuba and WTIP) family of LIM domain proteins as core components of the miRISC, which were later shown to scaffold GW182 and AGO proteins [25, 57, 90, 91, 104]. LIMD1 was observed to co-localise with various components of P-bodies [57]. P-bodies are cytoplasmic ribonucleoprotein granules believed to form from accumulation of non-translated mRNAs and associated translational repressing proteins including miRISC components [88]. Co-immunoprecipitations of LIMD1 then revealed that LIMD1 formed complexes with the miRISC associated proteins AGO2, DDX6, DCP2 and eIF4E [57]. The importance of LIMD1 in miRNA mediated silencing was demonstrated though shRNA-directed knockdown of LIMD1, which led to a

derepression of miRNA-mediated silencing [57]. LIMD1 knockdown was also seen to lead to a reduction in the number of cytoplasmic P-bodies, suggesting a role for LIMD1 in P-body formation. Ajuba and WTIP knockdowns produced a similar effect to the LIMD1 knockdown, however depletion of other zyxin-like proteins did not have this effect. This suggested a specific role for the LAW proteins in miRNA silencing. LIMD1 was shown to pull down and co-purify with eIF4E and eIF4A, which are components of the eIF4E-m<sup>7</sup>GTP cap, as well as AGO proteins. This led to a mechanism of LIMD1 activity being proposed that involved LIMD1 simultaneously interacting with miRISC components and the eIF4E-m<sup>7</sup>GTP cap to form a closed loop complex to inhibit translation [57].

More insight into the role of LIMD1 in miRNA silencing was provided by Bridge *et al.* (2017) [25]. They proposed a new mechanism for LIMD1 activity, in which LIMD1 scaffolds AGO and GW182 proteins in order to facilitate the AGO:GW182 interaction. Genetic ablation of LIMD1 was seen to cause a shift from AGO2 mediated silencing to AGO3 mediated silencing. This demonstrated a key role for the LAW proteins in determining AGO utilisation for miRNA mediated silencing. As the majority of miRNAs are found across all AGO proteins the level of redundancy of the different AGO proteins remains unclear, although AGO utilisation may be determined by differences in their subcellular localisation [105]. Regardless, a number of examples of a specific activity of a particular AGO protein have been shown although little is still known about the distinct roles of individual AGO proteins [105]. Further investigation of effect of LIMD1 on miRISC formation, using proximity-ligation assays, demonstrated that LIMD1 promoted the association of AGO2 and TNRC6A, leading to the idea of LIMD1 as an AGO-TNRC6 scaffold [25]. Co-immunoprecipitation assays were then used in combination with truncations of LIMD1. It was seen that a construct of LIMD1 lacking residues 140-166 (LIMD1 $\Delta$ 140-166) was unable to bind to AGO2 whilst a construct lacking the 3 LIM domains (LIMD1 $\Delta$ LIM123) was unable to bind TNRC6A. Together, these results suggest that AGO2 binds to the pre-LIM region of LIMD1 whilst TNRC6A binds to the LIM domain region of LIMD1 (Figure 1.11).

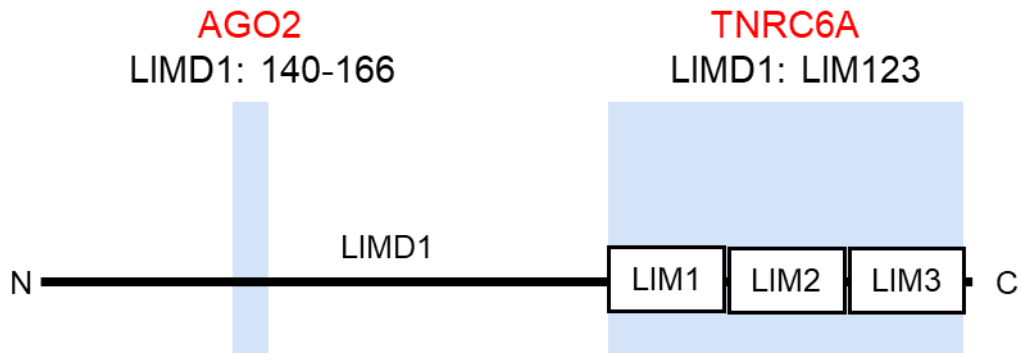


Figure 1.11. Schematic representation of LIMD1 regions essential for binding to AGO2 and TNRC6A [25]. A LIMD1 $\Delta$ 140-166 mutant was unable to bind to AGO2 and a LIMD1 $\Delta$ LIM123 mutant was unable to bind TNRC6A.

Mutation of AGO2 S387 to an alanine or a phospho-mimetic glutamate residue either reduced binding of AGO2 to LIMD1 or rescued binding of AGO2 to LIMD1 respectively [25, 106]. This residue had previously been identified as a target for Akt-3 phosphorylation by in vitro phosphorylation with purified protein and subsequent analysis by mass spectrometry [106]. siRNA-mediated depletion of Akt-3 led to a derepression of a reporter mRNA in LIMD1<sup>+/+</sup> cells but not in LIMD1<sup>-/-</sup> cells [25]. This confirmed that AGO2 binding by LIMD1 was Akt-3 phosphorylation dependent. AGO2 is comprised of six regions, the N-terminal domain (N), linker 1 (L1), the PAZ (PIWI/Argonaute/Zwille) domain, linker 2 (L2), the MID domain and the P-element-induced whimpy tested (PIWI) domain (Figure 1.12) [107]. The S387 residue, phosphorylated by Akt-3 to facilitate the interaction with LIMD1, is found in the linker 2 region of AGO2 (Figure 1.12).

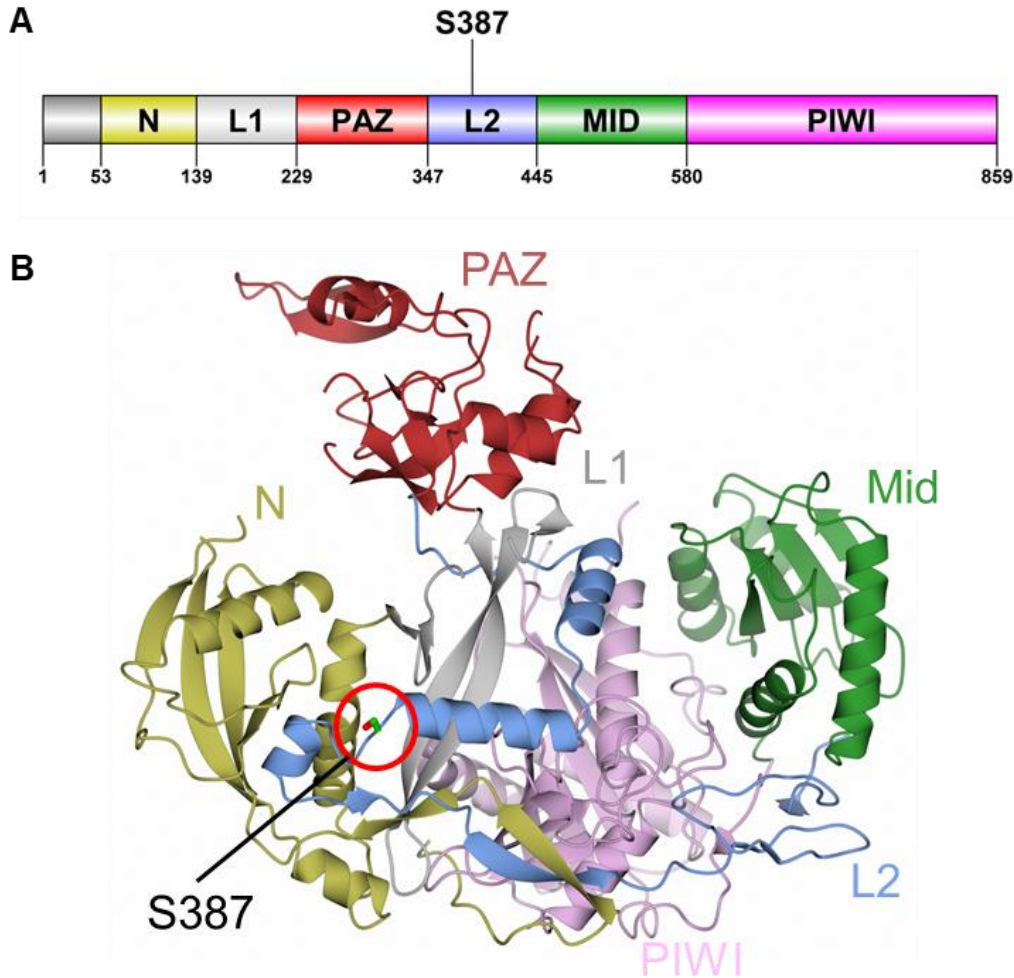


Figure 1.12. **A.** Schematic representation of the domain organisation of AGO2. **B.** The structure of AGO2 with domains indicated by different colours and the S387 residue key for LIMD1 binding indicated. PDB code: 4F3T [107].

AGO2 S387A was unable to rescue miRNA mediated silencing in a AGO2 knockdown however AGO2 S387E was able to rescue miRNA mediated silencing, providing additional evidence for the importance of AGO2 S387 phosphorylation. It was also found that loss of LIMD1 specifically inhibited the interaction of S387 phosphorylated AGO2 with TNRC6A. This all indicated that the phosphorylation event promoted the LIMD1:AGO2 interaction and that LIMD1 was essential in the formation of the AGO2 miRISC. The interactions of all three LAW proteins with AGO1, 2, and 4 were seen to be similarly phosphorylation dependent. However, the interaction with AGO3 was found to be independent of phosphorylation due to the

presence of a glutamate residue in place of the phosphorylated serine. This glutamate residue likely acts as a phospho-mimetic to facilitate interaction with the LAW proteins. This work provided strong evidence for the function of LIMD1 and the LAW proteins in scaffolding AGO and GW182 proteins to direct AGO utilisation in a Akt3 phosphorylation dependent manner. Part of the work presented in this thesis attempted to investigate the structural biology and biophysics of the interactions of LIMD1 with AGO2 and TNRC6A. Specifically, this study attempted to identify the structural determinants that underpin how AGO2 recognises the disordered re-LIM region of LIMD1. Work also focussed on determining the binding mode of TNRC6A to the LIM domain region of LIMD1. The overarching goal was therefore to explore general mechanisms of LIM domain partner recognition while also providing specific information on miRISC formation.

## 1.7. The Hypoxic Signalling Pathway

Sufficient oxygen availability is essential for maintaining cell function and survival. It is therefore important for cells to respond to low oxygen availability (hypoxia) to improve oxygen availability or adapt to survival in hypoxic conditions. The ability to survive in low oxygen environments is a key trait enabling tumour growth in many cancers, making the hypoxic signalling pathway a potential target for new cancer therapies [108, 109]. The importance of the hypoxic response was highlighted by the 2019 Nobel prize in physiology or medicine being awarded to researchers investigating this pathway [110]. A consensus has emerged that the hypoxic response is largely controlled by the hypoxia inducible factor 1 (HIF-1) transcription factor [111-113]. HIF-1 is comprised of two subunits, HIF-1 $\alpha$  and HIF-1 $\beta$  [114]. It is thought that HIF-1 activity is mediated through the degradation or stabilisation of the HIF-1 $\alpha$  subunit in response to normal or low oxygen levels respectively (Figure 1.13) [109, 113, 115, 116]. The degradation of HIF-1 $\alpha$  is believed to be driven by oxygen dependent hydroxylation by prolyl hydroxylase (PHD) enzymes, and subsequent ubiquitination of HIF-1 $\alpha$  by a von-Hippel-Lindau (VHL) E3 ubiquitin ligase complex, which signals for degradation of HIF-1 $\alpha$  by the proteasome (Figure 1.13) [115]. Under hypoxic conditions the lack of oxygen availability prevents hydroxylation of HIF-1 $\alpha$ , which stabilises HIF-1 $\alpha$  and allows for dimerization of HIF-1 $\alpha$  with HIF-1 $\beta$  to

form the active HIF-1 transcription factor (Figure 1.13). HIF-1 can then interact with its co-activators, including p300/CBP, and bind to the regulatory regions of its target genes to induce their expression. HIF-1 allows for adaptation to low oxygen by reducing oxygen uptake and increasing vascularisation. One way in which HIF-1 brings about this adaptation is through upregulating the expression of the C-Myc repressor MXI1 which in turn leads to decreased mitochondrial respiration and increased glycolysis [117]. In cancer cells the preferential metabolism of glucose to lactate even under non-hypoxic conditions is known as the Warburg effect [118]. HIF-1 is a key effector of the Warburg effect, allowing for greater proliferation by upregulating glucose transporter proteins, vascular endothelial growth factor (VEGF) and pyruvate dehydrogenase kinase (PDK) (which inhibits pyruvate dehydrogenase (PDH)) [118]. Two additional members of the HIF protein family, HIF-2 and HIF-3, have also been shown to be important in the regulation of the hypoxic response and are both regulated by the PHD-VHL degradation pathway [119-122]. HIF-3 function has yet to be fully elucidated but appears to act as both a negative regulator of HIF-1 and HIF-2 whilst also inducing expression of its own target genes [120]. HIF-1 and HIF-2 have been more extensively studied than HIF-3 and have been shown to have unique target genes [123]. HIF-1 has been found to be ubiquitously expressed whilst HIF-2 expression is more tissue specific. This study focussed on regulation of HIF-1 due to its ubiquitous expression and extensive prior research compared to HIF-2 or HIF-3.



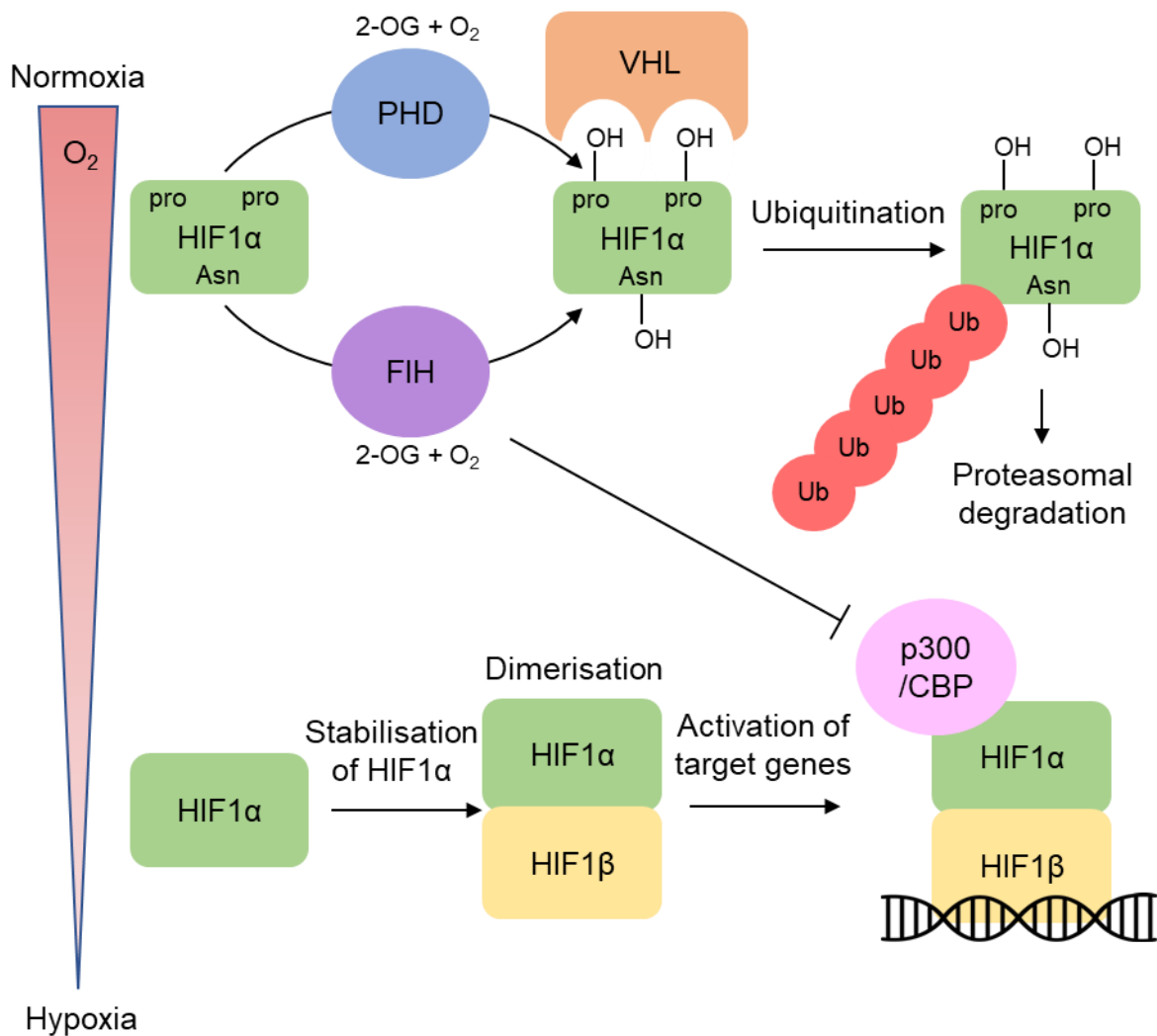


Figure 1.13. Schematic representation of HIF-1 $\alpha$  regulation in response to oxygen availability. Under normoxic (normal oxygen availability) conditions HIF-1 $\alpha$  is hydroxylated by FIH, which prevents binding to the p300 co-activator to inhibit HIF-1 activity, and PHD enzymes. Hydroxylation of conserved prolines by PHD enzymes promoted binding by VHL ubiquitin ligase complex. This leads to ubiquitination and subsequent proteosomal degradation of HIF-1 $\alpha$ . Under hypoxic conditions HIF-1 $\alpha$  is not hydroxylated and so can dimerise with HIF-1 $\beta$  activate its target genes.

Hydroxylation of HIF-1 $\alpha$  is catalysed by the non-heme Fe(II) and 2-oxoglutarate (2-OG) dependent oxygenases PHD 1,2 and 3 and Factor Inhibiting HIF-1 $\alpha$  (FIH). These enzymes use  $O_2$  and 2-OG to form a reactive Fe(IV)-oxo species, as well as

CO<sub>2</sub> and succinate as by-products [124]. The Fe(IV)-oxo species can then abstract a hydrogen from the substrate to form a substrate radical which can then form the hydroxylated product. FIH hydroxylates HIF-1 $\alpha$  N803 which is thought to inhibit HIF-1 by preventing the interaction of the HIF-1 $\alpha$  C-terminal transactivation domain with the p300 coactivator [125, 126]. The HIF-1 $\alpha$  N-terminal transactivation domain was found to be able to activate expression of target genes in splice variants lacking the C-terminal transactivation domain, suggesting that FIH hydroxylated HIF-1 $\alpha$  can still activate its target genes through its N-terminal region [127, 128]. The three PHD enzymes hydroxylate two conserved prolines, P402 and P564, on HIF-1 $\alpha$  [129, 130]. This hydroxylation promotes binding of HIF-1 $\alpha$  to the VHL E3 ubiquitin ligase complex [131-134]. HIF-1 $\alpha$  is then ubiquitinated and subsequently degraded by the proteasome. The three human PHD enzymes have all been shown to be capable of hydroxylating HIF-1 $\alpha$  and have a conserved C-terminal prolyl 4-hydroxylase domain (Figure 1.14) [129]. The N-terminal of the human PHD enzymes is more variable. PHD1 contains a large, predicted to be disordered, N-terminal region. PHD2 contains a MYND domain in an otherwise predicted to be disordered N-terminal region and PHD3 is mostly limited to the catalytic prolyl 4-hydroxylase domain (Figure 1.14).

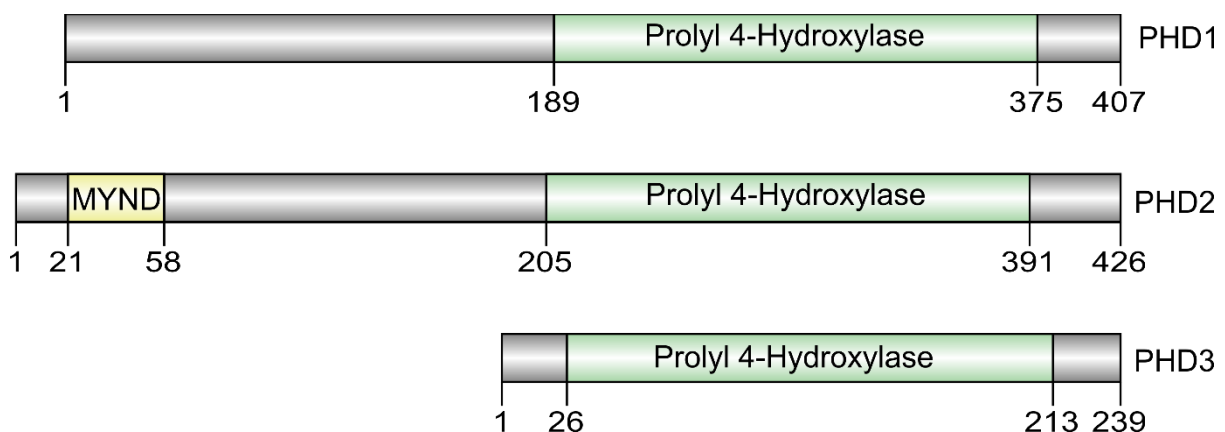


Figure 1.14. Cartoon representation of the domain composition of PHD1,2 and 3. Domain positions were taken from the Pfam database [66].

PHD1, 2 and 3 were seen to contribute differently towards the hydroxylation of different HIF subunits, with PHD1 preferentially hydroxylating HIF-2 $\alpha$  and PHD2

preferentially hydroxylating HIF-1 $\alpha$  under normoxic conditions [122]. PHD3 was found to contribute more to HIF-1 $\alpha$  and HIF-2 $\alpha$  hydroxylation under hypoxic conditions, where some level of HIF-1 $\alpha$  hydroxylation still occurs [122]. These results as well as the differential expression seen for the PHD enzymes indicate different roles for PHD1, 2 and 3. The current consensus points to PHD2 as the primary oxygen sensor of the cell [127]. Berra *et al.* (2003) found siRNA induced silencing of PHD1 and 3 had no effect on HIF-1 $\alpha$  stability whereas silencing of PHD2 led to activation of HIF-1 [127]. Germ-line deletion of PHD2 in mice lead to embryonic lethality whereas, PHD1<sup>-/-</sup> and PHD3<sup>-/-</sup> mice remained viable [135]. Hydroxylation of either of the HIF-1 $\alpha$  oxygen degradation domains (ODD) allows for binding of VHL [134]. VHL acts as the recognition component of a ubiquitin ligase complex containing Elongin B, Elongin C and Cul-2. VHL binding triggers ubiquitination and subsequent degradation of HIF-1 $\alpha$  and HIF-2 $\alpha$ . The interactions of both PHD2 and VHL with HIF-1 $\alpha$  have been extensively characterised both structurally and biophysically [136-139]. Crystal structures of PHD2 without substrate or in complex with HIF-1 $\alpha$  C-terminal ODD (CODD) or N-terminal ODD (NODD) are available and revealed a common binding site for the two ODDs (Figure 1.15B) [136, 140]. Crystal structures of PHD2 in complex with inhibitors also demonstrated the presence of open and closed conformations of PHD2 (Figure 1.15A) which are believed to be important in determining the substrate specificity of PHD enzymes [136]. Given that LIMD1 also binds to PHD2 it is important to consider the complexes of PHD2 and what these could mean for recognition of LIMD1. If LIMD1 binds to the substrate binding site of PHD2 it may be in competition with HIF-1 $\alpha$ . Alternatively, LIMD1 may bind preferentially to the open or closed conformation of PHD2 leading to allosteric regulation of LIMD1 binding by HIF-1 $\alpha$ . The effect of HIF-1 $\alpha$  binding on the interaction between LIMD1 and PHD2 is investigated in this study.

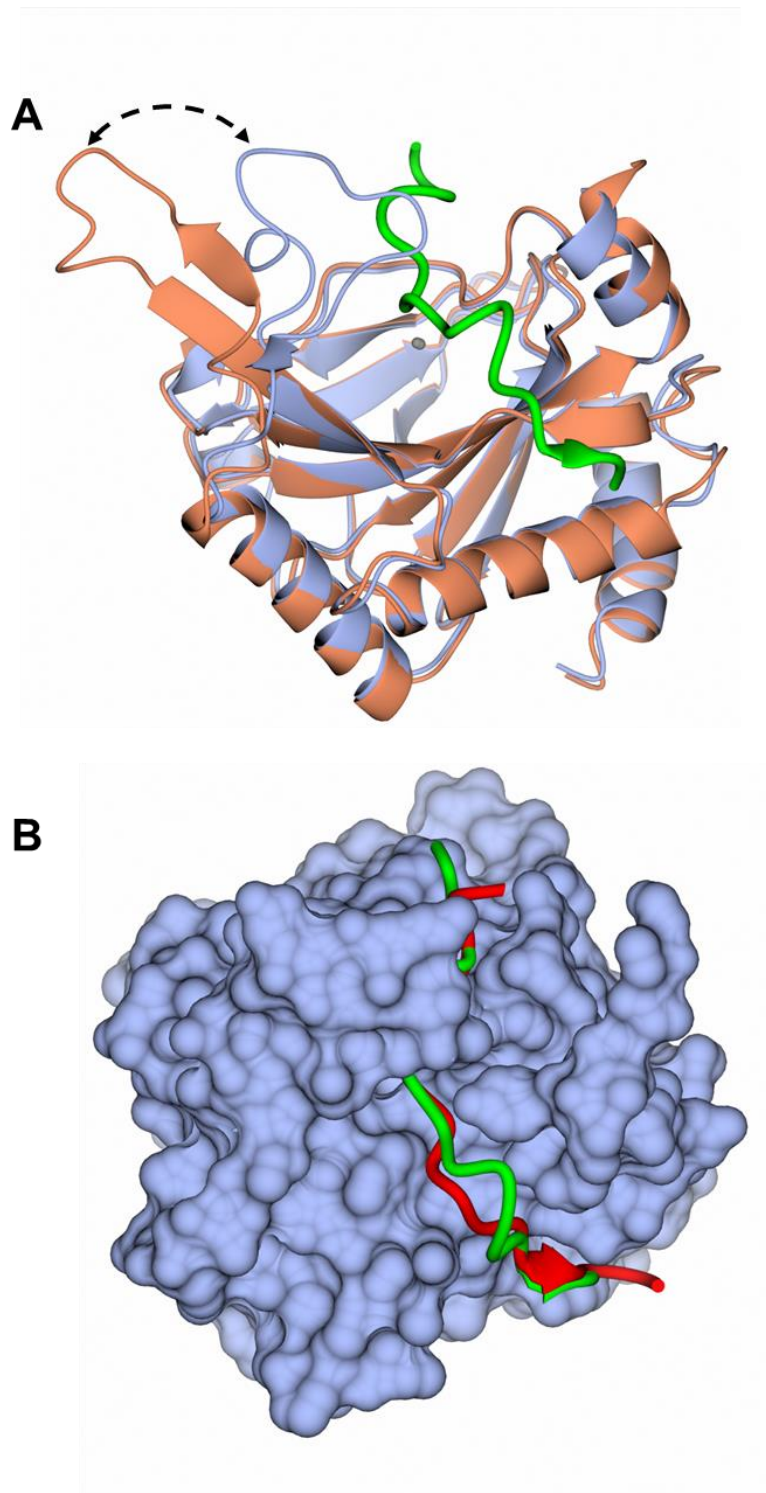


Figure 1.15. **A.** Crystal structures of PHD2 in complex with HIF-1 $\alpha$  CODD (Green) and inhibitors showing the 'open' (orange) and 'closed' (blue) conformations of PHD2. PDB codes: 3HQU and 3HQR [136]. **B.** crystal structures of PHD2 (blue) bound to both HIF-1 $\alpha$  CODD (green) and NODD (red). PDB codes: 5L9B and 5L9V [140].

Analysis of the effect of 2-OG on HIF-1 $\alpha$  binding by PHD2 revealed 2-OG hindered binding of hydroxylated HIF-1 $\alpha$  but not un-hydroxylated HIF-1 $\alpha$  [137]. This may be a mechanism to prevent product inhibition of PHD2 and could also lead to inhibition of HIF-1 $\alpha$  degradation by PHD2 binding when 2-OG is limiting. VHL was successfully crystallised with HIF-1 $\alpha$  CODD and revealed that the hydroxyproline was essential in the formation of the complex with 96% of its accessible surface area buried in the complex (Figure 1.16) [138]. Extensive work has focussed on targeting the interaction of VHL with HIF-1 $\alpha$  using small molecules which has provided a potential route for the development of new drugs to treat ischemic diseases [139].

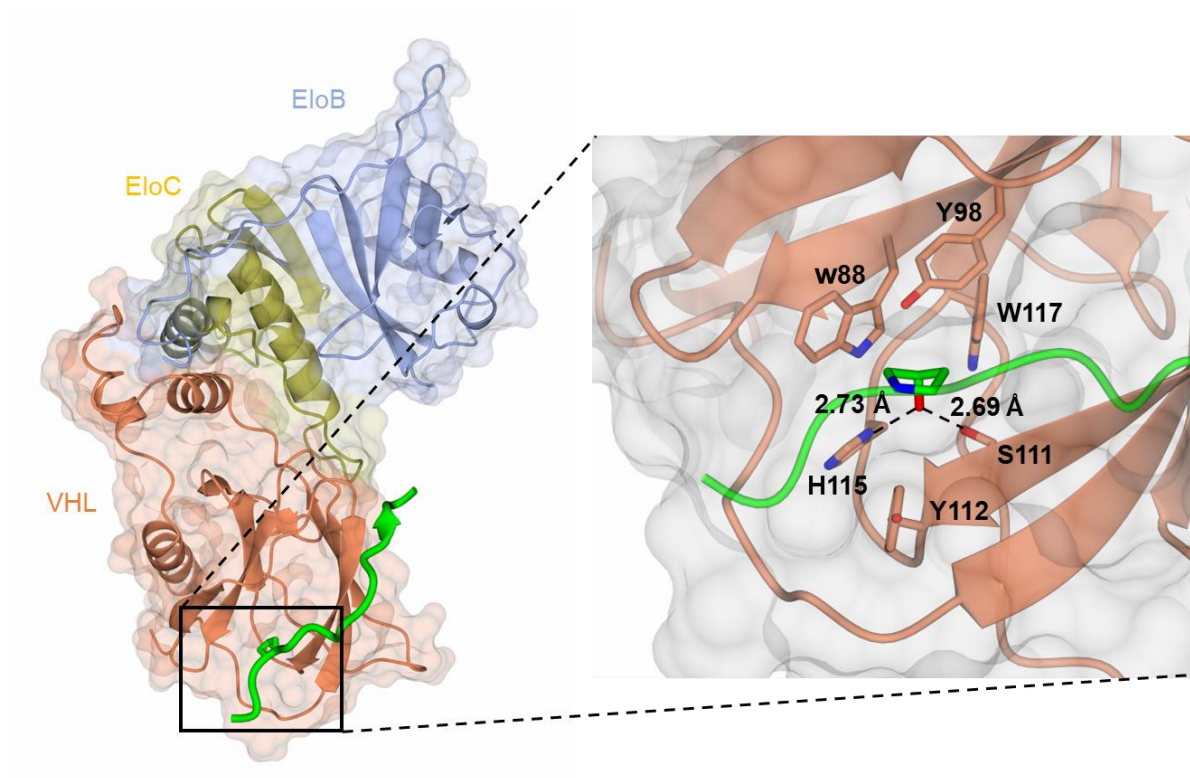


Figure 1.16. Crystal structure of the VHL (orange), Elongin B (blue), Elongin C (gold) complex bound to a peptide of the HIF-1 $\alpha$  CODD (green). A zoom in of the hydroxyproline binding site is given with VHL residues within 4 Å of the HIF-1 $\alpha$  CODD hydroxyproline shown. Close interactions of the hydroxyproline hydroxyl group with S111 and H115 on VHL are also indicated. PBD code: 1ML8 [138].

Several proteins have been implicated in scaffolding of PHD2 and VHL to enhance HIF-1 $\alpha$  degradation [141]. Phospholipase D1 (PDL1) and phospholipase D2 (PDL2) have been shown, through *in vitro* pull-down assays and co-immunoprecipitations, to interact directly with VHL, PHD2 and HIF-1 $\alpha$  [142, 143]. However, overexpression of PDL1 or 2 in cancer cells led to increased levels of HIF-1 $\alpha$  due to the production of phosphatidic acid by the enzymatic activity of PDL which lead to mammalian target of rapamycin (mTOR) inducing HIF-1 $\alpha$  expression [143, 144]. The A kinase anchor protein mAKAP was seen to co-precipitate HIF-1 $\alpha$ , PHD2, PHD3, and VHL and expression of mAKAP was consistent with lower HIF-1 $\alpha$  levels under normoxic conditions [145]. This suggested a likely scaffolding role for mAKAP in VHL mediated HIF-1 $\alpha$  regulation. Interestingly mAKAP appears to enhance HIF-1 $\alpha$  activity in hypoxic conditions, likely by localising HIF-1 $\alpha$  in the nucleus, close to its site of action [145]. Another A kinase anchor protein AKAP12 was also suggested to function as a hypoxic scaffold and was shown to increase the association of HIF-1 $\alpha$  with VHL and PHD2 [146]. Of the PHD2:VHL scaffolds mentioned above, none have been shown to consistently reduce HIF-1 $\alpha$  levels under both normoxic and hypoxic conditions. The LAW family of LIM domain scaffold proteins have been shown to fill this role [24].

### 1.8. LIMD1 is a Key Regulator of Hypoxic Signalling

Foxler *et al.* (2012) proposed a role for the LIMD1 scaffold protein as a PHD2, VHL scaffold protein [24]. It was demonstrated that the three LAW proteins were all able to bind to VHL but had different affinities for the different PHD proteins, with LIMD1 being the only LAW protein able to bind PHD1,2 and 3 whilst WTIP and AJUBA were only able to bind to PHD1 and 3. This difference in affinity may allow for LAW proteins to regulate PHD utilisation. As PHD2 is thought to be the primary oxygen sensor of the cell, LIMD1 was focused on as a more likely disease relevant protein than WTIP or AJUBA in this system [127]. LIMD1 co-immunoprecipitated with PHD2 and VHL and all three proteins were seen in the same fraction in a sucrose gradient of HEK293 cell extracts providing evidence for a PHD2-LIMD1-VHL complex [24]. Blots against HIF-1 $\alpha$  in the sucrose gradient were not shown and so it is not known whether HIF-1 $\alpha$  was also present in this complex. Overexpression of LIMD1 was

seen to result in increased HIF-1 $\alpha$  ubiquitination, demonstrating that LIMD1 is important for VHL mediated HIF-1 $\alpha$  regulation. LIMD1 scaffolding was seen to be important in both normoxic and hypoxic conditions, with LIMD1 depletions leading to increased HIF-1 $\alpha$  stabilisation in both conditions. It was also observed that LIMD1 depletion led to an inhibition of HIF-1 $\alpha$  degradation after reoxygenation. This suggested a potential role in LIMD1 scaffolding for chronic hypoxia and recovery from hypoxia, which could also explain the purpose of upregulation of LIMD1 expression by HIF-1 $\alpha$  [147]. Co-immunoprecipitations with LIMD1 truncations showed that a LIMD1 $\Delta$ 186-220 mutant was unable to bind PHD2 whilst a LIMD1 $\Delta$ 239-260 mutant retained the ability to bind PHD2 leading to an inferred PHD2 binding region of LIMD1 186-239. Additionally, a LIMD1 $\Delta$ LIM2 mutant was shown to be unable to bind VHL. These results indicated that PHD2 binds to the pre-LIM region of LIMD1 whilst VHL binds to LIM2 of LIMD1 (Figure 7.17).

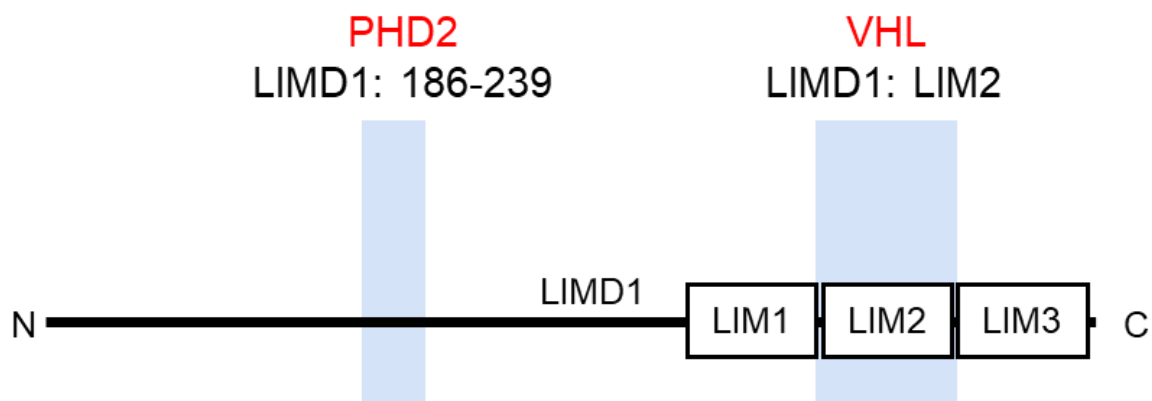


Figure 1.17. Schematic representation of LIMD1 regions essential for binding to PHD2 and VHL. A LIMD1 $\Delta$ 186-260 mutant was unable to bind to PHD2 whilst a LIMD1 $\Delta$ 239-260 retained the ability to bind PHD2 leading to an inferred PHD2 binding region of LIMD1 187-239. A LIMD1 $\Delta$ LIM2 mutant was shown to be unable to bind VHL.

Further investigation of the system revealed that LIMD1 expression is induced by HIF-1 in a negative feedback loop [147]. Under hypoxic conditions, HIF-1 induces expression of LIMD1 which then enables efficient degradation of HIF-1 $\alpha$  under hypoxic conditions. Female SCID-beige mice were subcutaneously injected with



either wild type A549 lung cancer epithelial cells or A549 cells with ablation of hypoxic response elements (HRE) in the LIMD1 promoter sequence. Ablation of HRE in the LIMD1 promoter was seen to increase tumour vascularity and growth providing further evidence for the importance of this system in chronic hypoxia and additional explanation for the tumour suppressing activity of LIMD1 [147]. Verification that loss of LIMD1 had no effect on HIF-1 $\alpha$  mRNA levels confirmed that this activity was not due to the role of LIMD1 in miRNA mediated silencing [147]. Evidence points to a role for LIMD1 in HIF-1 $\alpha$  regulation by scaffolding PHD2 and VHL in order to increase their local concentration, and thus signalling efficiency, to aid in HIF-1 $\alpha$  degradation [24]. It also appears that this system is of particular importance under chronic hypoxic conditions, such as is found in many cancers, in order to prevent pathological dysregulation of the HIF-1 system [147]. Rho-related BTB domain-containing protein 3 (RHOBTB3) has been suggested to act as a LIMD1-PHD2-VHL scaffold [148]. It was demonstrated that RHOBTB3 reduced HIF-1 $\alpha$  abundance, enhanced HIF-1 $\alpha$  hydroxylation and was seen to co-immunoprecipitate with VHL, PHD2 and LIMD1. Direct interaction of RHOBTB3 with the individual proteins was not confirmed and the association of RHOBTB3 with CUL3-ubiquitin ligase complexes suggests that RHOBTB3 may be more likely to facilitate HIF degradation through interactions with the VHL ubiquitin ligase complex rather than scaffolding of PHD, VHL and LIMD1 [148, 149]. It was also claimed, based on co-immunoprecipitations of RHOBTB3 and LIMD1, that RHOBTB3 formed a heterodimer with LIMD1 despite no evidence for the oligomeric state of the interaction or confirmation that the proteins interact directly [148]. Due to the lack of confidence in the role of RHOBTB3, this protein was not included in this thesis.

Part of the work presented in this thesis investigates the structural and biophysical nature of the protein:protein interactions of LIMD1 with PHD2 and VHL. Elucidation of the oligomeric states, binding regions, binding energies, and structures of the protein interactions was attempted. Investigation of the interaction between the second LIM domain of LIMD1 and VHL was carried out in order to inform on both the role of LIMD1 in hypoxic signalling and general mechanisms of LIM domain binding to structured partners. The interaction of the pre-LIM region of LIMD1 with PHD2 was also investigated in the presence and absence of HIF-1 $\alpha$  CODD in order to



confirm that LIMD1 acts *via* a scaffolding mechanism and to provide insight into the mode of recognition of the disordered pre-LIM region of LIMD1 despite the low sequence conservation in this region between the LAW proteins. Additionally, the effect of LIMD1 on the activity of PHD2 was investigated to evaluate the potential of LIMD1 to modulate the enzymatic activity of PHD2. Determining if LIMD1 is able to directly impact the enzymatic activity of PHD2 will help to support or question the proposed role of LIMD1 as a scaffold protein. It is hoped that this work can inform future hypoxia-targeted therapies and investigations into this pathway.

### 1.9. NMR Spectroscopy is a Powerful Tool to Study the Structural Biology and Biophysics of Scaffold Proteins

The activity of scaffold proteins is governed primarily through the protein:protein interactions they form with their partner proteins. Understanding the structural biology and biophysical nature of these interactions is therefore essential to understand, and thus target, scaffold protein biochemistry. The highly flexible, disordered regions commonly featured in scaffold proteins make structural studies of these regions in the unbound state impossible by x-ray crystallography or cryo-EM. Even where these regions fold upon binding to a partner protein, the dynamic nature of the interaction, and the individual proteins, often prevents crystal formation and can lead to poorly populated classes in cryo-EM. Additionally, scaffold proteins often feature low-affinity interactions with their partner proteins ( $K_d > \sim 1 \mu\text{M}$ ). These low-affinity interactions are more challenging to study *in vitro*, with approaches such as pull-down assays often failing to capture the interaction. The equilibrium between bound and unbound states for low-affinity complexes can also prevent crystal formation x-ray crystallography and can lead to a lack of observed complex for cryo-EM analysis. This difficulty has led to a dearth of structures of lower affinity interactions available on the PDB [43].

Solution NMR spectroscopy has proven to be a powerful tool for investigation of both IDPs [150-154] and low affinity interactions [43, 155, 156]. These factors make NMR spectroscopy a key technique for the study of scaffold protein structure and

biophysics. The highly dynamic internal motions in IDPs can result in an improved signal-to-noise ratio compared to folded proteins [157]. Despite the potential improvement in the signal-to-noise ratio associated with IDPs over folded proteins, several factors make analysis of IDPs by NMR more challenging. The position of peaks in NMR spectra, known as the chemical shift, is determined by the chemical environment of the observed nucleus. In disordered proteins the majority of residues will be solvent exposed, and any local structure is averaged by conformational dynamics. Together, these factors mean that nuclei in an IDP tend to experience similar chemical environments compared to the conditions of a folded protein. This similarity in the chemical environment across the IDP leads to a more narrow distribution of signals from backbone nuclei which can lead to a large number of overlapping peaks which can in turn be difficult to resolve. Additionally, NMR analysis of proteins often relies on observation of the backbone N-H atoms. As IDPs are often enriched in proline residues, which lack N-H groups, the number of N-H groups that can be observed can be severely reduced in IDPs compared to folded proteins, which do not have the same enrichment. The exquisite sensitivity of chemical shift to changes in local environment makes NMR spectroscopy well suited to the study of low affinity interactions [156].

This study utilised solution NMR spectroscopy as the primary technique to study the protein:protein interactions of the LIMD1 scaffold protein. Work was also carried out to develop methods to aid in the study of protein structure by NMR spectroscopy. This involved combining specific isotopic unlabelling approaches with filtered/edited NOESY (nuclear Overhauser effect spectroscopy) experiments to reduce the spectral complexity/overlap of NOESY spectra and to provide residue type assignments for structure calculations. More detail on this approach is presented in section 5.1

## 1.10. Aims

LIMD1 scaffolding has been shown to be essential for correct regulation of HIF-1 $\alpha$  in the hypoxic signalling pathway, and miRISC formation in miRNA mediated silencing [24, 25]. Current work on LIMD1 has largely focussed on determining its role in various signalling pathways and disease. This project focussed on uncovering the structural biology and biophysics of the interactions of the LIMD1 scaffold with its partner proteins in miRNA mediated silencing and hypoxic signalling pathways. LIMD1 has been identified as a target for cancer therapies, demonstrating the potential value in being able to develop additional LIMD1 targeted therapies [73, 74, 158]. More in depth analysis of the protein:protein interactions made by LIMD1 may aid in the development of these therapies. Specific targeting of the interaction between LIMD1 and PHD2 or VHL could also be valuable for the treatment of ischemic diseases. The investigation of the protein:protein interactions carried out here provides case studies to support the use of solution NMR in the study of challenging systems such as these.

The pre-LIM regions of LIMD1, WTIP and AJUBA are all predicted to be disordered and have low sequence similarity (Figure 1.7). Despite this, the similarity in the functions and protein binding partners of the LAW proteins in both miRNA mediated silencing and hypoxic signalling require specific recognition of protein partners by the pre-LIM regions [24, 25]. In addition to this, the differences in specificity seen for AGO and PHD enzymes by the pre-LIM regions require a high degree of specificity. Investigating the interactions of the LIMD1 pre-LIM region with both AGO2 and PHD2 was attempted in order to provide more information on how this specificity could be achieved in the absence of sequence or structural conservation.

Analysis of the available structures of LIM domains bound to partner proteins (Figure 1.3) revealed distinct binding faces on the LIM domains for linear and folded partner proteins. As the LIM domain region of LIMD1 is able to bind to both structured VHL and disordered TNRC6A, investigation of the structure of these interactions would contribute significantly to our understanding of LIM domain partner recognition [25].

For this reason, and to contribute to our understanding of the role of LIMD1 in miRNA mediated silencing and hypoxic signalling, work was carried out to attempt to identify the binding mode of the LIMD1 LIM domains to both VHL and TNRC6A.

LIMD1 has been proposed to function as a scaffold protein to facilitate the hydroxylation and ubiquitination of HIF-1 $\alpha$  through binding to and co-localising PHD and VHL proteins [24, 147]. It is however also possible that LIMD1 facilitates HIF-1 $\alpha$  degradation through direct modulation of the activity of its partner proteins, either instead of or in addition to its scaffolding activity. In order to investigate this the effect of LIMD1 on the enzymatic activity of PHD2 was investigated. This will help to better understand the role of LIMD1 in hypoxic signalling. Any scaffolding ability of LIMD1 is reliant on recruitment of its partner proteins. In order to better understand and model the scaffolding activity of LIMD1 it is necessary to determine the oligomeric state of the interaction formed between LIMD1 and its partner proteins. LIMD1 has also been suggested to homodimerize [148]. This study aimed to elucidate the oligomeric states of the complexes formed by LIMD1 in both hypoxic signalling and miRNA mediated silencing. This would provide valuable insights into the nature of these complexes and allow for better modelling of the systems of interest.

This project also involved the development of solution NMR methods to probe protein structure. It is hoped that this will aid similar studies of proteins by solution NMR spectroscopy. Specific isotope unlabelling was combined with filtered/edited NOESY experiments in order to simplify the resulting NOESY spectra and to provide residue type assignments to aid automatic resonance assignment pipelines. More information on this work is presented in chapter 5. It is hoped that this work provides an additional tool to aid protein structure analysis by NMR spectroscopy.

The aims of this project were to:

- Uncover the structural determinants of the disordered LIMD1 pre-LIM binding to AGO2
- Determine the mode of recognition of the LIMD1 LIM domains for binding to TNRC6A and VHL to improve understanding of LIM domain function
- Investigate the structural biology of the LIMD1:PHD2 interface in the absence and presence of HIF-1 $\alpha$
- Investigate the effect of LIMD1 on the enzymatic activity of PHD2
- Determine the oligomeric state of LIMD1 complexes
- Develop and test methods to aid in structural studies of proteins by NMR spectroscopy by combining specific isotope unlabelling with filtered/edited NOESY experiments

## 2. Materials and Methods

### 2.1. Protein production

An overview of the protein expression vectors used in this thesis is given in Tables 2.1-2.4.

Table 2.1. Table of proteins and associated vectors used for protein production. The protein tags and vector source are also indicated. ORF= Open Reading Frame, MBP = Maltose Binding Protein.

Protein	Vector backbone	Tags	Source
LIM1 (LIMD1 467-529)	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Purchased from York technology facility
LIM2 (LIMD1 532-594)	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Purchased from York technology facility
LIM3 (LIMD1 595-669)	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Purchased from York technology facility
LIM12 (LIMD1 467-594)	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Purchased from York technology facility
LIM123 (LIMD1 467-669)	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Purchased from York technology facility
LIM123-His <sub>6</sub>	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF-His <sub>6</sub>	Mutagenesis of LIM123 vector
TNRC6A 1-883	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Purchased from York technology facility

Table 2.2. Continuation of Table 2.1 showing proteins and associated vectors used for protein production. The protein tags and vector source are also indicted. ORF= Open Reading Frame, MBP = Maltose Binding Protein.

Protein	Vector backbone	Tags	Source
TNRC6A 157-883	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 1-883 vector by former Lab member
TNRC6A 301-883	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 1-883 vector by former Lab member
TNRC6A 544-883	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 1-883 vector by former Lab member
TNRC6A 1-202	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 1-883 vector
TNRC6A 157-408	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 157-883 vector
TNRC6A 301-573	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 301-883 vector
TNRC6A 544-787	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 544-883 vector
TNRC6A 720-883	pETFPP_2	His <sub>6</sub> -MBP- 3cProtease-ORF	Mutagenesis of TNRC6A 1-883 vector by former Lab member



Table 2.3. Continuation of Table 2.1 and 2.2 showing proteins and associated vectors used for protein production. The protein tags and vector source are also indicated. ORF= Open Reading Frame, MBP = Maltose Binding Protein

Protein	Vector backbone	Tags	Source
LIMD1 110-166	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Produced by York Biosciences Technology Facility (BTF)
LIM12-TNRC6A <sub>715-739</sub>	pET28a	His <sub>6</sub> -thrombin-MBP-3cProtease-ORF	Purchased from GenScript Biotech
AGO2 S387D, S824A, S828D, S831D, and S834A	pFastBacHT-A	His <sub>6</sub> -TEV-ORF	Gifted by Prof. Ian MacRae (Scripps Institute USA)
VHL 54-213 and Elongin B	pGEX-4T-3	GST-thrombin-VHL & Elongin B	Gifted by Sir Peter Radcliffe (Francis Crick Institute)
Elongin C 17-112	pBB75	-	Gifted by Sir Peter Radcliffe (Francis Crick Institute)
Full length PHD2	pETFPP_1	His <sub>6</sub> -3cProtease-ORF	Produced by York Biosciences Technology Facility (BTF)
PHD2 181-426	pETFPP_1	His <sub>6</sub> -3cProtease-ORF	Produced by York Biosciences Technology Facility (BTF)
LIMD1 65-260	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Produced by York Biosciences Technology Facility (BTF)
LIMD1 160-300	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Produced by York Biosciences Technology Facility (BTF)

Table 2.4. Continuation of Table 2.1, 2.2 and 2.3 showing proteins and associated vectors used for protein production. The protein tags and vector source are also indicated. ORF= Open Reading Frame, MBP = Maltose Binding Protein

Protein	Vector backbone	Tags	Source
LIMD1 166-260	pETFPP_2	His <sub>6</sub> -MBP-3cProtease-ORF	Gene purchased from GenScript Biotech and cloned into pETFPP_2
SHIRT-R3	pETFPP_1	His <sub>6</sub> -3cProtease-ORF	Generated by former Lab member
SasYr	pETFPP_1	His <sub>6</sub> -3cProtease-ORF	Generated by another lab member

### 2.1.1. Producing the LIMD1 166-260 Expression Vector

A codon optimised gene for expression of residues 166-260 of human LIMD1 (UniProt Code: Q9UGP4) in *E. coli* was purchased from Integrated DNA Technologies and cloned into the pETFPP\_2 vector using In-Fusion cloning (Takara Bio). Primers were designed to amplify the insert with the addition of overhangs complimentary to overhangs in the linearised vector. The primers were purchased from Sigma-Aldrich.

#### Primers

Forward: 5'-TCCAGGGACCAGCAATGCCCTGCGAAGATCCATC-3'

Reverse: 5'-TGAGGAGAAGGCGCGTTAGGTTCGGTTTCTCGGAG-3'

The insert was amplified by PCR using the PCR mix in table 2.5 and the PCR program in table 2.6

Table 2.5. PCR mix components for the amplification of the LIMD1 166-260 gene

Component	Volume
10X Buffer for KOD Hot Start DNA Polymerase	5 $\mu$ L
25 mM MgSO <sub>4</sub>	3 $\mu$ L
dNTPs (2 mM each)	5 $\mu$ L
Forward primer (10 $\mu$ M)	1.5 $\mu$ L
Reverse primer (10 $\mu$ M)	1.5 $\mu$ L
LIMD1 166-260 insert (10 ng/ $\mu$ L)	1 $\mu$ L
DMSO	1 $\mu$ L
KOD Hot Start DNA Polymerase (1 U/ $\mu$ l)	1 $\mu$ L
Pure water	31 $\mu$ L

Table 2.6. PCR steps for the amplification of the LIMD1 166-260 gene

Step	Temp and time
1.Initial denature	95°C for 2 min
2.Denature	95°C for 20 s
3.Annealing	57°C for 10 s
4.Extension	70°C for 4 s
Repeat steps 2-4	30 cycles
Cool	10°C

The PCR product was run on a 1% agarose gel containing sybr-safe. The band for the insert was excised and the DNA was extracted using a Monarch DNA Gel Extraction Kit. The In-Fusion reaction was carried out using 100 ng of both the PCR amplified insert and linearised vector with 2 µL 5x In-Fusion HD Enzyme Premix in a total of 10 µL. The reaction was incubated at 50°C for 15 min and then cooled on ice. 1 µL of the reaction was used to transform 50 µL DH5α competent cells.

The transformed cells were grown on an LB agar plate containing 50 µg/mL kanamycin. One colony was used to inoculate 10 mL LB media containing 50 µg/mL kanamycin. The culture was grown overnight at 37°C. DNA was extracted from the culture using a QIAprep spin Miniprep kit and the DNA was sent for sequencing by Eurofins Genomics. Sequencing using t7 forward and reverse primers confirmed the insert was successfully cloned into the vector.

### 2.1.2. Mutagenesis

In order to narrow down the binding site of LIMD1 on TNRC6A, and to allow for higher purity than could be achieved for TNRC6A 1-883, overlapping constructs spanning the AGO-binding domain of TNRC6A were produced. TNRC6A 1-202, 157-408, 301-573 and 544-787 constructs were generated by introducing stop codons into constructs of TNRC6A produced by previous lab members (TNRC6A 1-883, 1-

157, 1-301 or 544-883. Stop codons were introduced using a Quikchange Lightning kit with the following primers:

1-202

Forward: 5' GACTGGACCAAATAACTAGTAGACTACTAACTTTATGAC 3'

Reverse: 5' GTCATAAAGTTAGTAGTCTACTAGTTATTTGGTCCAGTC 3'

157-408

Forward: 5' CAAATGAGCAAAGCAGTTAATAAGTGTGGGCCAAAACAGG 3'

Reverse: 5' CCTGTTTTGGCCCACTTATTAAGTCTTTGCTCATTG 3'

301-573

Forward: 5' GCAAAGAGGAGAAGTAATAAGCTGCATGGAATGAC 3'

Reverse: 5' GTCATTCCATGCAGCTTACTTCTCCTCTTTGC 3'

544-787

Forward: 5' CAAGCACAGGTACATTAATAACAGCTGCTAACGCC 3'

Reverse: 5' GCGTTAGCAGCTGTTATTAATGTACCTGTGCTTG 3'

In order to remove the N-terminal His<sub>6</sub>-MBP tag on LIM123 whilst still retaining a His<sub>6</sub> tag for pull down experiments an additional C-terminal His<sub>6</sub> tag was inserted into the expression vector using a Quikchange Lightning mutagenesis kit. The newly generated construct was named LIM123His<sub>6</sub>. These primers allow for an 18 nucleotide insertion that code for a His<sub>6</sub> tag at the C-terminal of the protein. The resulting construct consists of an N-terminal His<sub>6</sub>-MBP tag with a 3C protease cleavage site, followed by the three LIM domains of LIMD1 and a C-terminal His<sub>6</sub> tag.

LIM123 His<sub>6</sub> inserttion mutagenesis primers:

Forward: 5' GACCCTCATCTACAGGTAGCCATCACCATCATTAACGCGCCTTCTC  
3'

Reverse: 5' GAGAAGGCGCGTTAATGATGGTGATGGTGATGGCTACCTGTAGAT-  
GAGGGTC 3'

PCR reactions were set up according to the Quikchange lightning protocol (Table 2.8) and run using the steps in Table 2.7.

Table 2.7. Temperatures and times for steps in the PCR program for TNRC6A mutagenesis

Step	Temp and time
1.Initial denature	95°C for 2 min
2.Denature	95°C for 20 s
3.Annealing	60°C for 10 s
4.Extension	68°C for 4.6 min for 1-202 and 151-408 4.15 min for 301-573 and 544-787 Or 4.3 min for LIM123His
Repeat steps 2-4	18 cycles
Final extension	68°C for 5 min

Table 2.8. PCR mix components for QuikChange Lighting mutagenesis

Component	Volume
10x reaction Buffer	5 $\mu$ L
dNTP mix	1 $\mu$ L
Forward primer (5 $\mu$ M)	0.75 $\mu$ L
Reverse primer (5 $\mu$ M)	0.75 $\mu$ L
Template (50 ng/ $\mu$ L)	1 $\mu$ L
Quiksolution reagent	1.5 $\mu$ L
QuikChange Lightning Enzyme	1 $\mu$ L
Pure water	39 $\mu$ L

Template DNA was then digested by Dpn1 digest, using 2  $\mu$ L of Dpn1 per reaction, for 5 min at 37°C before the product was used to transform stellar competent *E. coli*. The transformed cells were grown on an LB agar plate containing 50  $\mu$ g/mL kanamycin. One colony was used to inoculate 10 mL LB media containing 50  $\mu$ g/mL kanamycin. The culture was grown overnight at 37°C, 180 rpm. DNA was extracted from the culture using a QIAprep spin Miniprep kit and the DNA was sent for sequencing by Eurofins Genomics. Sequencing using t7 forward and reverse primers confirmed successful incorporation of the mutations.

### 2.1.3. Expression of unlabelled LIMD1, PHD2 and TNRC6A proteins

Expression plasmids (Table 2.1-2.4) were used to transform BL21 (DE3) *E. coli*. 1  $\mu$ L of plasmid DNA was added to 50  $\mu$ L of BL21 (DE3) and incubated on ice for 15 min. The cells were then heat shocked at 42°C for 10 s before incubation on ice for 5 min. 150  $\mu$ L LB media was added and cells were grown for 1 h before being spread onto an LB/agar plate containing 50  $\mu$ g/mL kanamycin. The plate was incubated at 37°C overnight. 5-10 colonies from the selection plate were used to inoculate 10 mL of LB media with 50  $\mu$ g/mL kanamycin. This was grown at 37°C 180 rpm overnight. The 10 mL culture was then used to inoculate 1 L LB media containing 50  $\mu$ g/mL kanamycin. The 1 L culture was then grown at 37°C, 180 rpm to an OD<sub>600</sub> of 0.8 before expression was induced by addition of 0.1 mM IPTG. Protein expression was carried

out at 20°C, 180 rpm for 18-20 h. Cells were then pelleted by centrifugation at 5000 × *g* for 20 min.

#### 2.1.4. Expression of uniformly isotope labelled protein

Uniform isotope labelling was achieved by expressing the protein of interest in isotopically enriched M9 media (Table 2.9). Expression plasmids (Table 2.1-2.4) were used to transform BL21 (DE3) *E. coli*. 1 µL of plasmid DNA was added to 50 µL of BL21 (DE3) and incubated on ice for 15 min. The cells were then heat shocked at 42°C for 10 s before incubation on ice for 5 min. 150 µL LB media was added and cells were grown for 1 h before being spread onto an LB/agar plate containing 50 µg/mL kanamycin. The plate was incubated at 37°C overnight. 10 mL LB containing 50 µg/mL kanamycin was inoculated with 5-10 colonies from the selection plate. This culture was grown at 37°C, 180 rpm for 5 h. The 10 mL culture was then used to inoculate a 50 mL M9 culture to a starting OD<sub>600</sub> of 0.05. The 50 mL culture was then grown overnight at 37°C, 180 rpm before being used to inoculate a 1 L culture at a starting OD<sub>600</sub> of 0.05. This culture was grown to OD<sub>600</sub> of 0.8 before expression was induced by addition IPTG to a final concentration of 0.1 mM. Expression was carried out at 20°C 180 rpm for 18-20 h. Cells were pelleted by centrifugation at 5000 × *g* for 20 min.



Table 2.9. Recipe for 1 L of M9 media as used to produce uniformly isotope labelled proteins.

Component	Quantity for 1 L M9
Na <sub>2</sub> HPO <sub>4</sub>	6 g
KH <sub>2</sub> PO <sub>4</sub>	3 g
NaCl	0.5 g
NH <sub>4</sub> Cl (U-[ <sup>15</sup> N] for <sup>15</sup> N labelling)	1 g
D-Glucose 20 % (w/v) (U-[ <sup>13</sup> C] for <sup>13</sup> C labelling)	10 mL
MgSO <sub>4</sub>	1 mL
Metal stock (FeCl <sub>3</sub> (50 mM), ZnSO <sub>4</sub> (50 mM), CaCl <sub>2</sub> (100 mM), CuSO <sub>4</sub> (10 mM), MnCl <sub>2</sub> (100 mM))	1 mL
Kanamycin 50 mg/mL	1 mL
MEM vitamin solution	10 mL

### 2.1.5. AGO2 Expression

Expression of full length AGO2 featuring S387D, S824A, S828D, S831D, and S834A mutations was achieved using a Bac-to-Bac system. The AGO2 gene in the pFastBac HT A vector was gifted by Prof. Ian MacRae (Scripps Institute, USA) and allowed for expression of AGO2 with an N-terminal His<sub>6</sub>-tag for purification. The plasmid was used to transform max efficiency DH10Bac competent cells which were streaked onto 10 µg/mL Gentamycin, 12 µg/mL Tetracyclin, 50 µg/mL Kanamycin, 1 mM IPTG, 100 µg/mL (in DMSO) X-gal LB-agar plates. IPTG is included to induce expression of *LacZ* which encodes for β-galactosidase which is required for blue-white screening. Plates were grown for 30 h at 37 °C. White colonies were re-streaked onto fresh plates and grown for an additional 30 h at 37°C. White colonies from the re-streaked plates were grown in 5 mL LB media with 50 µg/mL Kanamycin overnight at 37°C, 180 rpm. Cells were pelleted and lysed using buffers P1, P2 and N3 from a Qiagen miniprep kit. The supernatant was treated with 1:1 isopropanol to precipitate bacmid. The bacmid pellet was then washed twice with 70% ethanol before dissolving in 20 µL sterile water. 200 µL of insect cell media was then added to the bacmid. 100 µL of insect cell media was mixed with 10 µL FuGene and this

was added to the bacmid solution. In a six well plate 0.5-0.75 M SF9 cells in 2 mL and 150  $\mu$ L of the bacmid mixture was added to each well. The plate was stored at 28°C for 72 h. The media was taken and used as V0 viral stock. A 50 mL culture of SF9 cells at 0.5 Mcells/mL was infected by addition of 100  $\mu$ L V0 and left for 72 h before centrifugation at 100 rpm for 30 min. The supernatant was then taken and used as V1 viral stock.

Expression of AGO2 was optimised by infecting 20 mL cultures of SF9 cells by addition of 100  $\mu$ L 500  $\mu$ L or 1000  $\mu$ L of V1 and taking samples at 48, 92 and 76 h post infection. Protein production of each sample was then determined by western blotting.

#### 2.1.6. Purification of LIMD1, PHD2 and TNRC6A proteins

For purifications cells were first resuspended in lysis buffer (50 mM Tris, 300 mM NaCl, 25 mM imidazole, 0.5 mM TCEP, pH 7.5) and lysed by sonication or using a cell disruptor. The lysate was centrifuged at 18000 rpm for 20 min. The supernatant was loaded onto a 5 mL HisTrap column and eluted with an imidazole gradient of 25 mM to 300 mM over 20 column volumes using an AKTA FPLC system. For tag cleavage HRV-3C protease (purchased from York technology facility) was added in a 1:50 *w/w* protease:protein ratio and the protein was dialysed against low imidazole buffer (50 mM Tris, 300 mM NaCl, 25 mM imidazole, 0.5 mM TCEP, pH 7.5) overnight at 4°C. The sample was then passed through a HisTrap column to remove cleaved tag, uncleaved protein and protease. LIM123 samples were not subjected to additional purification steps. For all other proteins, the samples were then concentrated using amicon ultra centrifugal filters and purified by size exclusion chromatography using a HiLoad 26/600 or 16/600 Superdex 75 pg column or, for TNRC6A 1-883 and full length PHD2, a HiLoad 16/600 Superdex 200 pg column using an AKTA FPLC system.

### 2.1.7. Production of the VHL:ElonginB:ElonginC Complex

VHL 54-213 and Full length ElonginB were in the pGEX-4T-3 and ElonginC 17-112 in pBB75 were kindly gifted by Sir Peter Radcliffe [159]. The VHL:ElonginB and the ElonginC plasmid were used to co-transform BL21 (DE3) *E. coli* cells. 1  $\mu$ L of each plasmid was added to 50  $\mu$ L BL21 (DE3) and incubated on ice for 15 min before heat shock at 42°C for 10 s. The cells were then incubated on ice for 5 min before addition of 150  $\mu$ L LB media. The cells were then grown at 37°C 180 rpm for 1 h before being spread on LB agar containing 30  $\mu$ g/mL kanamycin and 100  $\mu$ g/mL ampicillin and incubated at 37°C overnight. 5-10 colonies from the selection plate were used to inoculate 10 mL LB media containing 30  $\mu$ g/mL kanamycin and 100  $\mu$ g/mL ampicillin. This was grown at 37°C and 180 rpm for 6 hours before being used to inoculate 50 mL LB media containing 30  $\mu$ g/mL kanamycin and 100  $\mu$ g/mL ampicillin to a starting OD<sub>600</sub> of 0.1. This culture was then grown at 37°C and 180 rpm overnight. The 50 mL culture was then used to inoculate 1L LB media containing 30  $\mu$ g/mL kanamycin and 100  $\mu$ g/mL ampicillin at a starting OD<sub>600</sub> of 0.1. The 1 L culture was grown at 37°C 180 rpm to OD<sub>600</sub> of 0.8 before expression was induced by the addition of 0.1 mM IPTG. After induction the culture was left at 20°C 180 rpm for 18-20 h before being pelleted by centrifugation at 5000 x g for 20 min.

Cells were resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 0.5 mM TCEP, pH 8) and lysed by sonication. The lysate was centrifuged at 18000 rpm for 20 min. The supernatant was loaded onto a 5 mL GSTrap HP column using an AKTA FPLC system. Protein was eluted with buffer containing 10 mM glutathione. The tag was cleaved by overnight incubation with human thrombin (CAS number: 9002-04-4) purchased from Sigma-Aldrich (1 U thrombin per 1 mg protein). Cleaved Tag and uncleaved protein were removed by passing the protein through a GSTrap column using an AKTA FPLC system. The unbound protein was then diluted 10-fold with a no salt buffer (50 mM Tris, 0.5 mM TCEP, pH 8) to (50 mM Tris, 50 mM NaCl, 0.5 mM TCEP, pH 8) before being loaded onto an 8 mL source 15Q column using an AKTA FPLC system. The protein was then eluted with gradient of 50 mM to 1 M NaCl. Fractions from the elution were analysed by SDS-PAGE and fractions containing VHL:ElonginB:ElonginC were pooled, concentrated and run on a HiLoad

26/600 Superdex 200 pg column using an AKTA FPLC system. The protein purity was then verified by SDS-PAGE.

#### 2.1.8. Production of Methyl-labelled PHD2

PHD2 181-426 was produced with Ile<sup>δ1</sup>, Leu<sup>proS</sup> and Val<sup>proS</sup> sites specifically <sup>13</sup>C<sup>1</sup>H<sub>3</sub> labelled in an otherwise [U-<sup>2</sup>H, <sup>12</sup>C, <sup>15</sup>N] labelled protein. The PHD2 181-426 expression vector was used to transform BL21 (DE3) cells which were spread onto a <sup>2</sup>H<sub>2</sub>O LB/agar plate containing 50 µg/mL kanamycin and grown at 37°C for 2 days. 5 colonies from the selection plate were used to inoculate 10 mL <sup>2</sup>H<sub>2</sub>O LB media containing 50 µg/mL kanamycin and this was grown at 37°C, 180 rpm for 6 h. The 10 mL culture was then used to inoculate a 25 mL culture of <sup>2</sup>H<sub>2</sub>O M9 media made with <sup>15</sup>NH<sub>4</sub>Cl and D-Glucose-1,2,3,4,5,6,6-d<sub>7</sub> at a starting OD<sub>600</sub> of 0.2. The 25 mL culture was grown overnight at 37°C, 180 rpm before being used to inoculate 250 mL <sup>2</sup>H<sub>2</sub>O M9 to a starting OD<sub>600</sub> of 0.2. The 250 mL culture was grown at 37°C, 180 rpm. 60 mg of 2-Hydroxy-2-[<sup>13</sup>C]methyl-3-oxo-4,4,4-tri-[<sup>2</sup>H]butanoate (pro-S acetolactate-13C) was added to label leucine and valine proS sites 1 h before induction [160]. 15 mg of 2-ketobutyrate-4-<sup>13</sup>C,3,3-d<sub>2</sub> was added 20 min before induction to label isoleucine δ1 methyl groups [161]. Once the culture had reached an OD<sub>600</sub> of 0.8 expression was induced by the addition of 0.1 mM IPTG. Expression was carried out at 20°C, 180 rpm for 18-20 h. Post expression, cells were harvested by centrifugation at 5000 x g for 20 min. The protein was then purified as previously described for the unlabelled PHD2 181-426.

#### 2.1.9. Production of Specifically Isotopically Unlabelled Protein

Isotope labelled proteins were expressed on a 500 mL scale following the protocol outlines in section 2.1.4 but with unlabelling compounds added 1 h before induction. For SHIRT-R3 unlabelling samples the unlabelling compounds used are outlined in table 2.10.

Table 2.10. Compounds and concentrations used for specific isotope unlabelling of SHIRT-R3

Unlabelling scheme	Compound 1	Compound 2
Leucine/Valine	$\alpha$ -ketoisovalerate 120 mg/L	
Leucine/Histidine	4-Methyl-2-oxopentanoate 120 mg/L	Histidine 400 mg/L
Lysine	Lysine 400 mg/L	
Lysine/Phenylalanine	Lysine 400 mg/L	Phenylpyruvate 400 mg/L
Tryptophan	Indole 100 mg/L	
Tyrosine	800 mg/L 4-hydroxyphenylpyruvic acid	
Phenylalanine	Phenylpyruvate 400 mg/L	

For SasYr unlabelling patterns, the same quantities were used apart from for Lysine which showed incomplete incorporation with the SHIRT-R3 samples. Lysine was instead used at 800 mg/L for SasYr unlabelling rather than the 400 mg/L used for SHIRT-R3 unlabelling.

SHIRT-R3 samples were resuspended in lysis buffer (20 mM Tris, 500 mM NaCl, 20 mM Imidazole pH 7.5) before lysis by sonication. The lysate was clarified by centrifugation at 18000 rpm for 20 min. The supernatant was loaded onto a 5 mL HisTrap HP column using an AKTA FPLC system and protein was eluted using a gradient from 20 mM to 500 mM imidazole over 20 column volumes. The samples were then concentrated and were all co-dialysed against the same NMR buffer (20 mM sodium phosphate, 150 mM NaCl, pH 6.5). Purity was confirmed by 15% SDS-PAGE (Appendix figure 7.23). SHIRT-R3 runs aberrantly on SDS-PAGE both in terms of its migration and the number of bands. The gels in appendix figure 7.23 are

consistent with other preparations of the same sample and 2D (1H, 15N) HSQC spectra confirmed a single species with the expected number and position of peaks.

SasYr samples were lysed by sonication in lysis buffer (50 mM Tris, 300 mM NaCl, 25 mM imidazole pH 8) before purification by IMAC using a 5 mL HisTrap HP column on an AKTA FPLC system and a gradient from 25 mM to 300 mM imidazole over 20 column volumes. The His<sub>6</sub> tag was cleaved overnight by incubation with HRV-3C protease 1:50 w/w protease:SasYr ratio and dialysed against 25 mM imidazole buffer. The His<sub>6</sub>, uncleaved protein and protease was removed by passing the sample through a 5 mL HisTrap HP column on an AKTA FPLC system. The unbound protein was concentrated and purified by SEC using a HiLoad Superdex 16/600 S75 column on an AKTA FPLC system. Purity was confirmed by 18% SDS-PAGE (appendix figure 7.24). As per SHIRT-R3 and as expected, SasYr runs apparently on SDS-PAGE. The protein has been confirmed to run at these masses on SDS-PAGE by previous work. Samples were co-dialysed into the same buffer for NMR spectroscopy.

## 2.2. Peptide Arrays

For initial peptide array experiments, the membranes were activated in MeOH for 10 min at RT, washed three times in TBS (50 mM Tris base, 27 mM KCl, 136 mM NaCl, pH 8) for 5 min. The membrane was then blocked for 1 h at room temperature with 5% sucrose, 4% skim milk and 0.05% tween-20 in TBS. Blocking buffer was washed off with TBS-T (50 mM Tris base, 27 mM KCl, 136 mM NaCl, 0.05% tween-20, pH 8). For no protein controls the membrane wash washed two more times with TBS-T before incubation with a mouse monoclonal anti-polyhistidine–peroxidase antibody (Sigma A7058) 1:120000 (v/v) for 2 h in blocking buffer. Protein sample was incubated at 20 µM for 4 h at room temperature in TBS-T before being washed three times with TBS-T for 5 min. The membrane was then incubated with a mouse monoclonal anti-polyhistidine–peroxidase antibody (Sigma A7058) 1:120000 (v/v) for 2 h in blocking buffer. Membranes were washed three times TBS-T for 10 min before detection with Amersham ECL prime western blotting detection reagent.

For optimised peptide arrays, membranes were activated in MeOH for 10 min at RT, washed three times in TBS (50 mM Tris base, 27 mM KCl, 136 mM NaCl, pH 8) for 5 min. The membrane was then blocked overnight at 4°C with 5% sucrose, 4% skim milk and 0.05% tween-20 in TBS. Blocking buffer was washed off with TBS-T (50 mM Tris base, 27 mM KCl, 136 mM NaCl, 0.05% tween-20, pH 8). For negative controls the membrane was washed two more times with TBS-T before incubation with a mouse monoclonal anti-polyhistidine-peroxidase antibody (Sigma A7058) 1:120000 for 2 h in blocking buffer or 1:1000 (v/v) ATTO 488 Ni-NTA conjugate for 1 h at room temperature. Protein sample was incubated at 2 µM for 4 h at room temperature in blocking buffer before being washed three times with TBS-T for 5 min. The membrane was then incubated with a mouse anti-polyhistidine-peroxidase antibody (Sigma A7058) 1:120000 for 2 h in blocking buffer or 1:1000 (v/v) ATTO 488 Ni-NTA conjugate for 1 h at room temperature. Membranes were washed three times TBS-T for 10 min before detection with Amersham ECL prime western blotting detection reagent or fluorescence detection.

### 2.3. Pull Down Assays

Initial pull downs and PHD2:LIMD1 pull downs were carried out by dispensing 45 µL of His mag sepharose excel bead slurry into a 1.5 mL Eppendorf. The supernatant was removed, and beads were resuspended in 113 µL equilibration buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, pH 7). The supernatant was removed, and 50 µL of the bait protein solution was added in equilibration buffer at a final concentration of 6 µM. This was incubated for 4 h at RT with end over end mixing. The supernatant was removed, and the beads were then washed with 3 x 113 µL wash buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, 30 mM imidazole, pH 7). 50 µL of the prey protein was added in equilibration buffer at a final concentration of 6 µM and incubated for 30 min at RT with end over end mixing. The supernatant was removed (and saved for SDS-PAGE analysis), and the beads were washed with 3 x 113 µL wash buffer, the last wash was saved for SDS-PAGE analysis, before elution with 40 µL elution buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, 500 mM imidazole, pH

7). The results were analysed by running 16  $\mu$ L of each sample on 18% SDS-PAGE gels

To increase sensitivity, pull-down assays were run by dispensing 45  $\mu$ L of His mag sepharose excel (Cytiva) bead slurry into a 1.5 mL Eppendorf. The supernatant was removed, and beads were resuspended in 113  $\mu$ L equilibration buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, pH 7). The supernatant was removed, and 50  $\mu$ L of the bait protein solution was added in equilibration buffer at a final concentration of 8  $\mu$ M. This was incubated for 4 h at RT with end over end mixing. The supernatant was removed, and the beads were then washed with 3 x 113  $\mu$ L wash buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, 30 mM imidazole, pH 7). 50  $\mu$ L of the prey protein was added in equilibration buffer at a final concentration of 20  $\mu$ M and incubated for 30 min at RT with end over end mixing. The supernatant was removed, and the beads were washed with 3 x 113  $\mu$ L wash buffer, the last wash was saved for SDS-PAGE analysis, before elution with 3 x 45  $\mu$ L elution buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, 500 mM imidazole, pH 7). 16  $\mu$ L of the first elution was taken for SDS-PAGE analysis and the rest of the elution fractions were precipitated by addition of 1 mL ice cold acetone, centrifuged for 5 min at 14000 rpm, the supernatant was removed, and the pellet was resuspended in 20  $\mu$ L equilibration buffer. The results were analysed by running 16  $\mu$ L of each sample on 18% SDS-PAGE gels

The AGO2 sample used for pull-down assays was gifted by Prof. Ian MacRae (Scripps Institute USA) and featured five mutations S387D, S824A, S828D, S831D, and S834A [95]. S387D should facilitate the interaction with LIMD1 by acting as a phosphomimetic. AGO2 pull downs were run by dispensing 20  $\mu$ L of His mag sepharose excel bead slurry into a 1.5 mL Eppendorf. The supernatant was removed, and beads were resuspended in 50  $\mu$ L equilibration buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, pH 7). The supernatant was removed, and 20  $\mu$ L of the bait protein solution was added in equilibration buffer at a final concentration of 8  $\mu$ M. This was incubated for 4 h at RT with end over end mixing. The supernatant was removed, and the beads were then washed with 3 x 50  $\mu$ L wash buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, 30 mM imidazole, pH 7). 20  $\mu$ L of the prey protein was



added in equilibration buffer at a final concentration of 2  $\mu$ M and incubated for 30 min at RT with end over end mixing. The supernatant was removed, and the beads were washed with 3 x 50  $\mu$ L wash buffer, the last wash was saved for SDS-PAGE analysis, before elution with 20  $\mu$ L elution buffer (50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, 500 mM imidazole, pH 7). The results were analysed by running 16  $\mu$ L of each sample on 12% SDS-PAGE gels

#### 2.4. Western Blotting

PVDF membrane was washed in 100% MeOH. A stack of filter paper/membrane/gel/filter paper was made and wetted with transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol, pH 8.3). protein was transferred at 25 V for 30 min using a semi-dry apparatus. The membrane was blocked overnight at 4°C in blocking buffer (50 mM Tris base, 27 mM KCl, 136 mM NaCl, 5% skimmed milk, 0.05% tween-20, pH 8). The membrane was washed three times with TBS-T (50 mM Tris base, 27 mM KCl, 136 mM NaCl, 0.05% tween-20, pH 8) before incubation with a mouse monoclonal anti-polyhistidine–peroxidase antibody (Sigma A7058) 1:120000 (v/v) in blocking buffer for 2 h at room temperature. The membrane wash then washed with TBS (50 mM Tris base, 27 mM KCl, 136 mM NaCl, pH 8) and imaged using Amersham ECL prime western blotting detection reagent.

#### 2.5. AlphaFold 2 Predictions

AlphaFold models were generated using the ColabFold notebook using MMseqs2 and AlphaFold multimer mode [162-164]. The statistical outputs from each run are shown in appendix figures 7.25-7.27.

## 2.6. NMR Experiments

### 2.6.1. NMR Sample Preparation

Individual LIM domains were concentrated to 100  $\mu$ M and dialysed into at in 20 mM Tris, 150 mM NaCl, 0.5 mM TCEP for 1D  $^1$ H experiments. 100  $\mu$ M samples of U- $^{15}$ N] LIM2 with and without 250  $\mu$ M unlabelled VHL:EloB:EloC were co-dialysed into 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O for 2D NMR experiments. [U- $^{15}$ N]-LIM12 and the [U- $^{15}$ N]-LIM12-TNRC6A<sub>715-739</sub> fusion protein were dialysed into 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O for NMR experiments. 100  $\mu$ M U- $^{15}$ N] LIMD1 166-260 with and without 400  $\mu$ M PHD2 were dialysed in to 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O.

100  $\mu$ M U- $^{15}$ N] LIMD1 65-260, 160-300 and 110-300 with and without 400  $\mu$ M unlabelled PHD2 181-426 were co-dialysed into 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O. 500  $\mu$ M LIMD1 1-80 at natural isotope abundance was dialysed in to 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O. 100  $\mu$ M U- $^{15}$ N] LIMD1 166-260 with and without 400  $\mu$ M PHD2 were dialysed in to 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O. 100  $\mu$ M U- $^{15}$ N] LIMD1 166-260 with and without unlabelled 250  $\mu$ M PHD2 and 250  $\mu$ M HIF CODD peptide were co-dialysed in to 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O. LIMD1 168-191, at natural isotope, abundance was diluted to 300  $\mu$ M in 20 mM Na phosphate, 150 mM NaCl, 0.5 mM TCEP, pH 6.6, 5% D<sub>2</sub>O both with and without 700  $\mu$ M unlabelled PHD2 181-426. For assignments, 250  $\mu$ M [U- $^{13}$ C,  $^{15}$ N] LIMD1 166-260 was dialysed into 20 mM Na phosphate, 150 mM NaCl, 1 mM TCEP, pH 6.6, 5% D<sub>2</sub>O.

100  $\mu$ M U- $^{15}$ N,  $^2$ H],( $^{\delta 1}$ L<sup>proSVproS</sup>)- $^{13}$ C $^1$ H $_3$ ]-PHD2<sub>181-426</sub> with and without unlabelled LIMD1 168-191 (300  $\mu$ M), HIF CODD (300  $\mu$ M) and LIM1 168-191 (300  $\mu$ M) and HIF CODD (300  $\mu$ M) together, were co-dialysed into 20 mM Na phosphate, 150 mM NaCl, 1 mM TCEP, pH 6.6, 5% D<sub>2</sub>O

SHIRT-R3 unlabelled samples were concentrated to 500  $\mu$ M and co-dialysed into in 20 mM sodium phosphate, 150 mM NaCl, pH 6.5 with 5% D<sub>2</sub>O. SasYr unlabelled samples were concentrated to 500  $\mu$ M and co-dialysed into 20 mM sodium phosphate, 150 mM NaCl, pH 6.5, 5% D<sub>2</sub>O.

### 2.6.2. NMR Experimental parameters

Initial 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectra were collected from LIMD1 65-260, 160-300, 110-166, and 166-260 with and without PHD2 using a Bruker Avance III 600 MHz NMR spectrometer. All other experiments were collected on a Bruker Avance Neo 700 MHz NMR spectrometer equipped with a 5 mm N<sub>2</sub>-cooled triple resonance TCI probe. All data were processed in TOPSPIN 4.0.6 and analysed in CCPN software [165].

A large proportion of the work on this thesis relied on the use of 2D (<sup>1</sup>H, <sup>15</sup>N) single bond correlation experiments such as 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC or HMQC experiments. SOFAST-HMQC experiments were used where additional sensitivity or shorter experiment time was beneficial, such as for unstable or low concentration samples. SOFAST experiments allow for more rapid scanning and so improves the sensitivity of the experiment [166]. 2D (<sup>1</sup>H, <sup>15</sup>N) TROSY experiments were used to improve the sensitivity of experiments observing larger proteins such as PHD2 [167]. An overview of the experimental parameters used for each sample for the 2D (<sup>1</sup>H, <sup>15</sup>N) single bond correlation experiments is given in Tables 2.11 and 2.12.

Table 2.11. Experimental parameters used for 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) correlation experiments. SFHMQCF3GPPH encodes for a SOFAST-HMQC [166]. HSQCETGPSI encodes for a 2D HSQC experiment.

Sample	Concentration ( $\mu\text{M}$ )	Temperature (K)	Pulse sequence	F2 FID (complex points)	F1 FID (complex points)	Scans per FID	F2 spectral width (ppm)	F1 spectral width (ppm)
[U- $^{15}\text{N}$ ]-LIM12	60	298	SFHMQCF3GPPH	2048	128	128	16.23	32
[U- $^{15}\text{N}$ ]-LIM12-TNRC6A <sub>715-739</sub>	30	298	SFHMQCF3GPPH	2048	128	1024	16.23	32
U-[ $^{15}\text{N}$ ]-LIM2	100	298	SFHMQCF3GPPH	2048	128	16	16.23	32
U-[ $^{15}\text{N}$ ]-LIMD1 65-260	100	288	HSQCETGPSI	1024	128	8	16.02	40
U-[ $^{15}\text{N}$ ]-LIMD1 110-166	100	288	HSQCETGPSI	1024	128	8	16.02	40
U-[ $^{15}\text{N}$ ]-LIMD1 160-300	100	288	HSQCETGPSI	1024	128	8	16.02	40
LIMD1 1-80	500	298	SFHMQCF3GPPH	2048	128	512	13.73	34
U-[ $^{15}\text{N}$ ]-LIMD1 166-260	100	298	HSQCETGPSI	2048	256	4	16.02	40

Table 2.12. Continuation of Table 2.11 showing experimental parameters used for 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) correlation experiments. SFHMQCF3GPPH encodes for a SOFAST-HMQC [166]. TROSYETF3GPSI encodes for a 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY experiment. H SQCETF3GPSI encodes for a 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC experiment [168-170].

Sample	Concentration ( $\mu\text{M}$ )	Temperature (K)	Pulse sequence	F2 FID (complex points)	F1 FID (complex points)	Scans per FID	F2 spectral width (ppm)	F1 spectral width (ppm)
U- $^{15}\text{N}$ -LIMD1 166-260	100	288	SFHMQCF3GPPH	2048	512	4	16.23	32
LIMD1 168-191	300	298	SFHMQCF3GPPH	2048	128	512	13.73	34
U- $^{15}\text{N}$ , $^2\text{H}$ ], ( $^{15}\text{N}$ - $^{13}\text{C}$ )- $^{13}\text{C}$ ]-PHD2 <sub>181-426</sub>	100	298	TROSYETF3GPSI	2048	256	8	13.73	40
SHIRT-R3 (all (un)labelling patterns)	500	298	HSQCETF3GPSI	2048	256	2	16.23	35
SasYr (all (un)labelling patterns)	400	298	HSQCETF3GPSI	2048	256	2	16.23	35

1D  $^1\text{H}$  spectra of the individual LIM domains were recorded using the pulse program ZGESGP (for solvent suppression by excitation sculpting) with 32768 complex points and 16 scans per FID. The spectral width was set to 15.87 ppm.

3D datasets to permit resonance assignment of  $[\text{U-}^{13}\text{C}, ^{15}\text{N}]$  LIMD1 166-260 were collected at 288 K. A 3D HNCACB experiment was collected using the pulse sequence HNCACBGPWG3D with 2048, 256 and 256 complex points in F3, F2 and F1 dimensions respectively and 16 scans per FID with 14% non-uniform sampling (NUS). Spectral widths were 11.7 ppm, 32 ppm and 80 ppm for F3, F2 and F1 respectively. A 3D HNCOCACB experiment was collected using the pulse sequence HNCOCACBGPWG3D with 2048, 256 and 256 complex points in F3, F2 and F1 dimensions respectively and 16 scans per FID with 14% non-uniform sampling (NUS). Spectral widths were 11.7 ppm, 32 ppm and 80 ppm for F3, F2 and F1 respectively. A 3D HNCO experiment was collected using the pulse sequence HNCOGP3D with 2048, 128 and 256 complex points in F3, F2 and F1 dimensions respectively and 16 scans per FID with 10% non-uniform sampling (NUS). Spectral widths were 11.7 ppm, 32 ppm and 12 ppm for F3, F2 and F1 respectively. Assignments were carried out using CcpNmr Analysis Assign software [171].

2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra of  $[\text{U-}^{15}\text{N}, ^2\text{H}]_2(\text{I}^{\delta 1}\text{L}^{\text{proSVproS}})-[^{13}\text{C}^1\text{H}_3]\text{-PHD2}_{181-426}$  were collected using the pulse sequence HSQCCTETGPSP with 1024 and 512 complex points in the F2 and F1 dimensions respectively. Spectral widths were 16.23 ppm centred on 4.7 ppm and 84 ppm centred on 43 ppm for F2 and F1 respectively.

NMR analysis of SHIRT-R3 samples was performed at 298 K using a protein concentration of 500  $\mu\text{M}$ . 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra were collected using the pulse sequence HSQCETGP with 2048 and 512 complex points in F2 and F1 dimensions respectively and 16 scans per FID [172]. Spectral widths were 16.2307 ppm and 100 ppm centred on 4.7 ppm and 50 ppm for F2 and F1 respectively. 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY spectra were collected using the pulse program NOESYFPGPPHWG [173, 174]. 2D F1fF2f NOESY were collected using the pulse sequence NOESYGPPHWGXF [173-

175]. 2D F1f NOESY was collected with the pulse sequence NOESYGPPHWGX1 [176-179]. All NOESY experiments were run with 2048 and 512 complex points in F2 and F1 dimensions respectively with 8 scans per FID and a spectral width of 16.2307 for both F2 and F1 dimensions. A mixing time of 120 ms was used for all 2D NOESY experiments.

SasYr experiments were collected at 500  $\mu$ M and 298 K. 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra were collected using the pulse sequence HSQCETGP. Aliphatic focussed 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra were collected with 1024 and 128 complex points in F2 and F1 dimensions respectively and 2 scans per FID and a spectral width of 16.2307 ppm and 100.0025 ppm in F2 and F1 centred on 4.7 ppm and 50 ppm. Aromatic focussed 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra were carried out with the same settings but a  $^{13}\text{C}$  spectral width of 50 ppm centred on 125 ppm. NOESY experiments were collected using the pulse sequences NOESYFPGPPHWG for the full 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY [173, 174], NOESYGPPHWGGXF for the 2D F1f2f NOESY [173-175], NOESYGPPHWGX1 for the 2D F1f NOESY [176-179], NOESYGPPHWGX2 for the 2D F2f NOESY and HSQCGPNOWGX33D for the 2D F1eF2f NOESY [176-179]. The 2D NOESY experiments were collected with 2048 and 512 complex points in F2 and F1 with 16 scans per FID and a spectral width of 16.22274 ppm for both F2 and F1. For the F1eF2f NOESY, F1 was set to 1 to record a 2D version of the experiment with the other parameters set to match the other NOESY experiments. A mixing time of 120 ms was used for all NOESY experiments. The 2D F1f TOCSY experiment was collected with a version of the pulse sequence DIPSI2GPPHWGX1 modified by Dr Alex Heyam to include  $^{15}\text{N}$  decoupling during the acquisition [176-179]. The F1f TOCSY experiments were collected with 2048 and 512 complex points in F2 and F1 dimensions respectively and with 16 scans per FID and a spectral width of 16.22274 ppm for both F2 and F1.

### 2.6.3. Analysis of Specific Isotope Unlabelled NMR Experiments

Isotope incorporation was assessed by comparing the mean normalised signal-to-noise ratio of peaks in unlabelled HSQC spectra to the equivalent peaks in the fully labelled HSQC spectra. Peak height was used as a measure of signal intensity. Severely overlapping peaks were excluded from the analysis. Peaks from unlabelled residues were not used to calculate the mean for mean normalisation.

The X-filtered/edited NOESY pulse sequences act to either select (edit) or remove (filter) signals resulting from  $^{13}\text{C}$  or  $^{15}\text{N}$  bound protons in either the F1 or F2 dimension [176]. This is achieved by applying a purge element before either or both of the  $t_1$  and  $t_2$  evolution periods. In a filtered experiment, protons bound to  $^{13}\text{C}$  or  $^{15}\text{N}$  will be filtered out whilst protons bound to  $^{12}\text{C}$  or  $^{14}\text{N}$  will be observed. In an edited experiment, protons bound to  $^{13}\text{C}$  or  $^{15}\text{N}$  are selectively observed meaning  $^{12}\text{C}$  and  $^{14}\text{N}$  bound protons are not observed. An example of a simple purge element is shown in figure 2.1A.



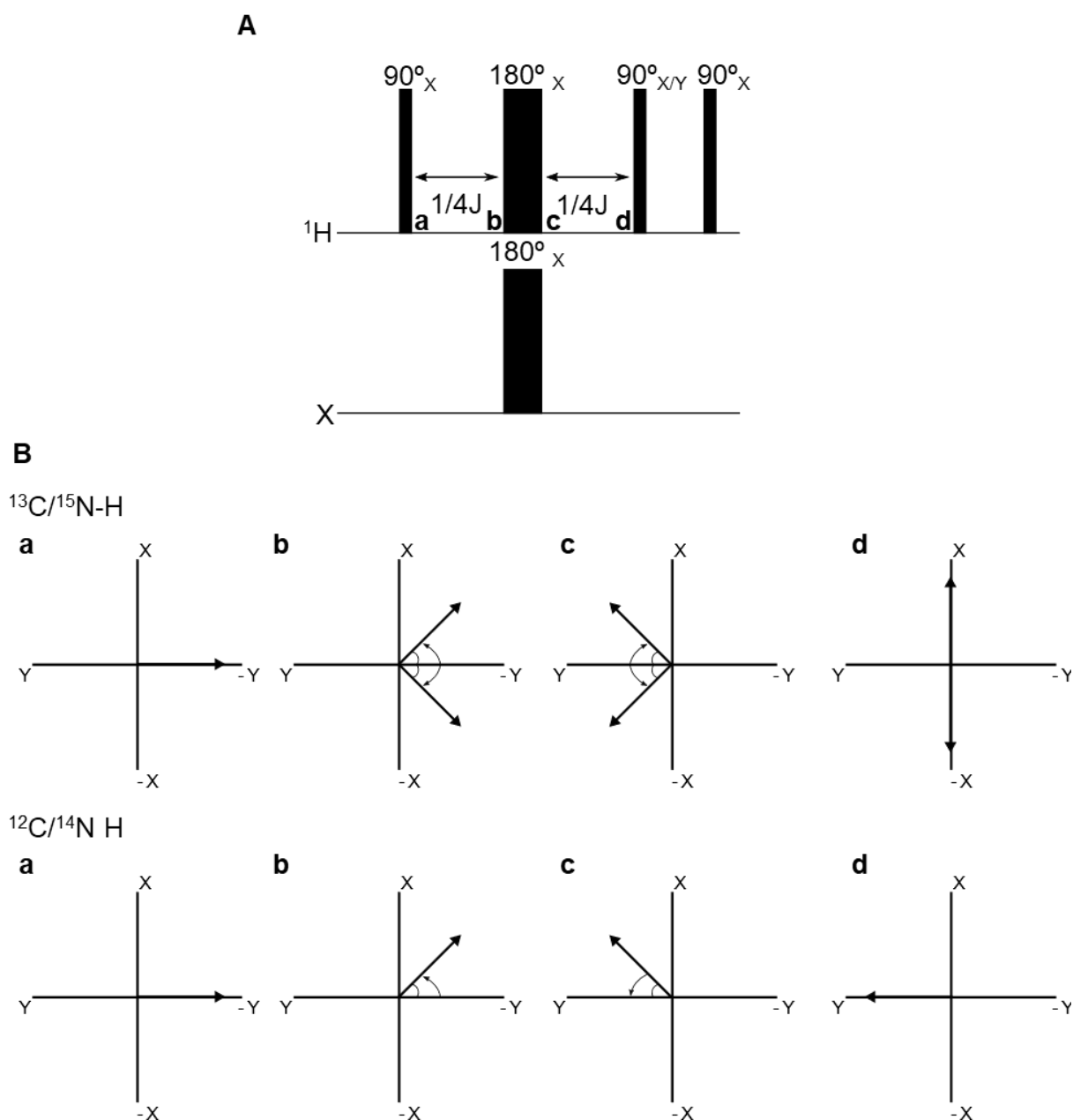


Figure 2.1. **A.** An example of a purge element sequence, specifically a BIRD filter. The top line corresponds to the  $^1\text{H}$  channel and the bottom line corresponds to the  $^{15}\text{N}$  or  $^{13}\text{C}$  channel.  $90^\circ$  pulses are shown as thinner lines and  $180^\circ$  pulses as thicker lines **B.** A vector model description of the purge element showing the spin states for both J-coupled  $^1\text{H}$  and non J-coupled  $^1\text{H}$  at different points during the pulse sequence. The phase of the  $90^\circ$  pulse at point **d** determines whether the purge acts as an edit or a filter.

The bulk magnetisation for both labelled and unlabelled protons during the purge element is shown as a vector representation in Figure 2.1B. At point **d** the magnetisation for X bound protons is antiphase along X and the magnetisation for unlabelled protons is along Y. This means that a filter can be applied *via* an X pulse to convert unlabelled-<sup>1</sup>H magnetisation to Z while labelled-<sup>1</sup>H stays in the transverse plane. The labelled-<sup>1</sup>H magnetisation can then be purged by applying a pulsed field gradient and unlabelled-<sup>1</sup>H magnetisation can be converted back along the transverse plane. Alternatively, an edit can be achieved by applying a Y pulse which converts the labelled-<sup>1</sup>H magnetisation to Z while the unlabelled-<sup>1</sup>H can be purged by gradient pulses. For editing, an additional purge element must be applied so that the magnetisation does not remain antiphase.

## 2.7. Mass Spectrometry Analysis of the VHL:ElonginB:ElonginC Complex

Mass spectrometry experiments were carried out by the bioscience technology facility at the University of York. Protein identification from gel samples was carried out using trypsin digest and MALDI-MS/MS. The spectra were then searched against the expected sequences.

For intact mass spectrometry the protein was diluted 1 in 10 and acidified with 0.1% TFA, then cleaned on a divinyl benzene column prior to elution into a Bruker maXis-HD qTOF mass spectrometer using electrospray ionisation. Positive ESI-MS spectra were averaged over the duration of the sample run. Multiply charged species were deconvoluted to average neutral masses using maximum entropy deconvolution, followed by Sum peak picking (S/N>10). Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control and DataAnalysis, Bruker Daltonics).

## 2.8. Analytical Size Exclusion and Co-elution

LIM2 and VHL were run on a Superdex 200 Increase 10/300 GL column using an AKTA Pure FPLC system. 100  $\mu$ L of VHL:EloB:EloC was run at 100  $\mu$ M and 100  $\mu$ L of LIM2 and VHL:EloB:EloC was run with both proteins at 100  $\mu$ M. Molecular mass standards from Sigma-Aldrich were also run. FL-PHD2 and LIMD1 65-260 were run on a Superdex 75 Increase 10/300 GL column. 500  $\mu$ L of each protein at 100  $\mu$ M was run individually and together. Molecular mass standards from Sigma-Aldrich were also run. PHD2 181-426 and LIMD1 65-260 were run on a Superdex 200 Increase 10/300 GL column using an AKTA Pure FPLC system. 500  $\mu$ L of each protein at 100  $\mu$ M was run individually and together. Molecular mass standards from Sigma-Aldrich were also run.

## 2.9. Succinate-Glo PHD2 Assay

PHD2 activity was monitored using a Succinate-Glo assay kit [180]. Assays were carried out in 384-well plates and analysed using a BMG Labtech Clariostar Plate Reader. Reactions were carried out on a 5  $\mu$ L scale in 25 mM HEPES pH 7.5, 1  $\mu$ M HIF CODD peptide, 10  $\mu$ M ascorbic acid, 1  $\mu$ M Fe(II), 1  $\mu$ M  $\alpha$ -ketoglutarate, 1% (v/v) DMSO plus PHD2 181-426 and/or LIMD2 166-260. The reaction was carried out for 30 min at 20°C without shaking. The reaction was stopped by addition of 5  $\mu$ L of succinate detection reagent 1. The plate was then incubated for 1 h before addition of 10  $\mu$ L succinate detection reagent 2. The plate was incubated for another 10 min then luminescence was recorded. Reactions were run in triplicate and normalised to a no enzyme control. The assay works by converting succinate produced by PHD2 into ATP which can then be used by a luciferase to produce light.

## 2.10. HPEPDOCK Docking

Peptide docking was achieved through the HPEPDOCK webserver using the 2G1M PHD2 structure with binding site residues set to 190-205 and LIMD1 168-191 for the peptide input [181, 182].

## 2.11. Structure Analysis

All structure analysis was performed using CCP4mg software [50]. Structure superpositions were performed by secondary structure matching. SasYr structure root mean square deviations (RMSDs) were calculated using MolMol between residues 27-109 [183]. G-factors (a measure of how normal or abnormal a protein structure is based the geometric parameters of the structure) were calculated using PROCHECK [184].

## 2.12. NMR Structure Calculations

Structure calculations were performed using ARIA2 (Ambiguous Restraints for Iterative Assignment) through the ARIAweb interface [185-187]. The core ability of ARIA to handle ambiguous assignments made it possible to incorporate known residue type assignments from the filtered/edited experiments by first ambiguously assigning these spectra in CCPN Analysis [165]. This ambiguous assignment was achieved by first producing distance restraints using the “make distance restraints” function in CCPN Analysis with the residue range limited to the potential residues specified by the labelled and filter/edit schemes. The peaks were assigned by matching chemical shifts to known chemical shifts from the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) and 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra with a tolerance of 0.025 ppm. Equivalent peaks could then be made from the restraints to produce an ambiguously assigned peak list which was assigned based on chemical shift and residue type. ARIA2 could readily handle these ambiguously assigned peak lists, whereas other structure building pipelines such as Cyana do not readily allow for ambiguously assigned peak lists to be used as in input [188]. The general ARIA pipeline is shown in figure 2.2.

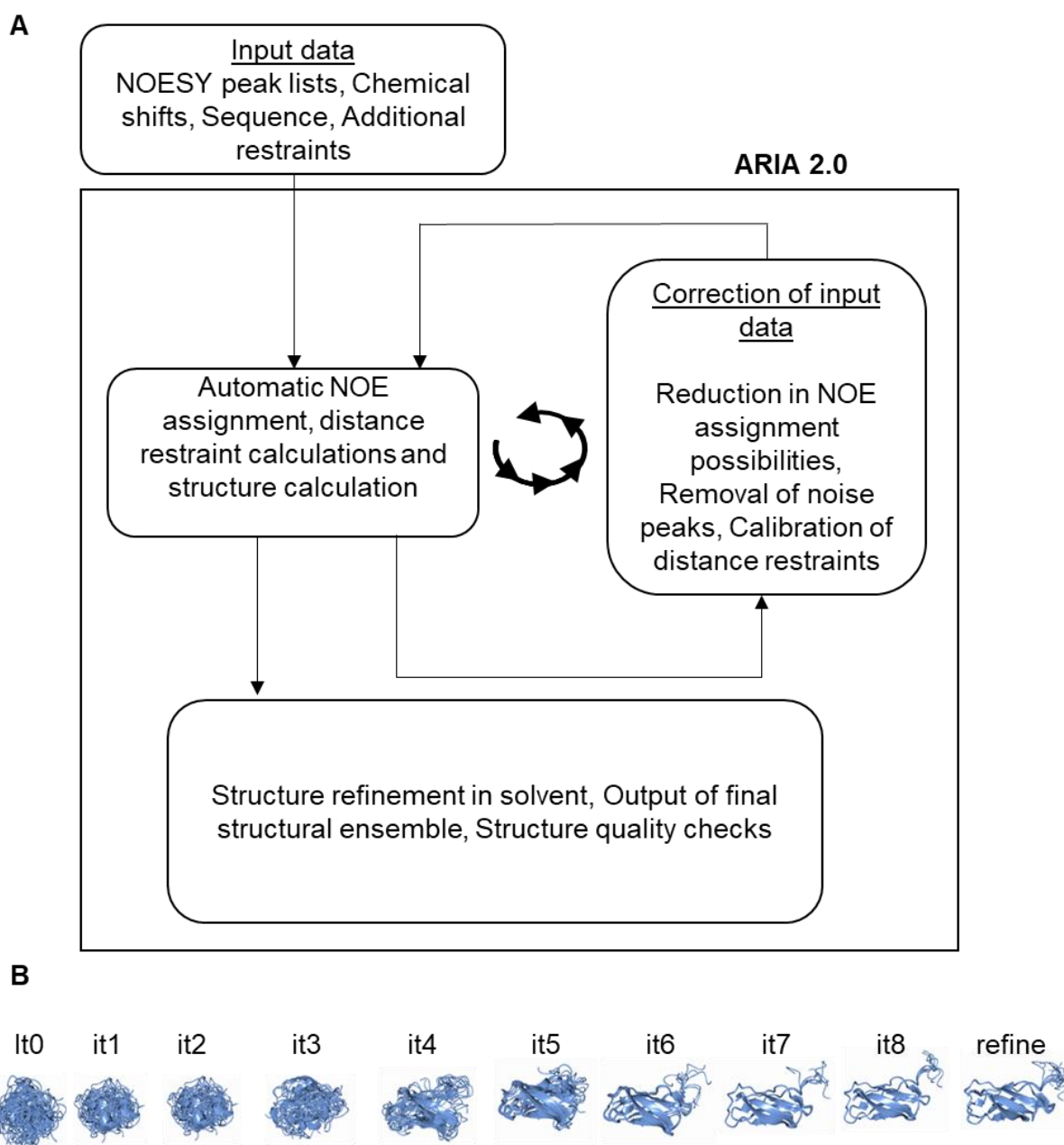


Figure 2.2. **A.** Schematic description of the ARIA2 structure building process showing the iterative structure building and data correction steps before eventual refinement of the final structure. **B.** Example of structures output at each iteration to show the improvement at each iteration. The structures are of SasYr with only NOEs from the 3D NOESY data included.

ARIA accepts assigned or unassigned NOESY peak lists and ambiguously assigns unassigned peaks based on chemical shift matching. Ambiguous distance restraints are then generated using the isolated spin pair approximation in which the distance

is equal to  $(\alpha^{-1}V_{ij})^{-1/6}$  where  $V_{ij}$  is the peak volume and  $\alpha$  is a correction factor [189]. The generated distance restraints are used, in combination with any additional restraints (e.g. dihedral angle restrains, residual dipolar couplings and/or disulphide bond restraints), for the initial structure calculation by molecular dynamics simulated annealing. The resulting structural ensemble is then used to improve the distance restraints by removing inconsistent cross peaks, reducing assignment possibilities, and calibrating  $\alpha$ . The improved distance restraints are then used for the next structure calculation iteration. There are typically nine iterations of structure calculation and restraint calibration before the structure ensemble is refined in explicit solvent.

All structure calculations were carried out using ARIAweb with proton1 shift error set to 0.04 ppm, proton2 shift error set to 0.02 ppm and hetero atom shift error set to 0.5 ppm and an upper bound correction of 6.0 Å. Dihedral angle restraints generated by Rachael Cooper using TALOS were included in all calculations [190, 191]. 20 structures were generated at each iteration with the 7 lowest energy structures used for restraint calibration. Molecular dynamics was performed using torsion angle dynamics [192]. The top 10 structures were refined in water. The different experiment tests for the LVW sample and the tests of the different labelling patterns with the full 3D data sets were run with 10000 high temperature steps, 5000 cool1 steps and 4000 cool2 steps in the simulated annealing. Calculations that only used 25% of the 3D data with and without LVW F2f and randomly added peaks were run with 30000 hot steps, 15000 cool1 steps and 12000 cool2 steps. Calculations that only used 25% of the 3D data to test different labelling schemes and the calculations with 5% of the 3D data randomly removed were run with 20000 high temperature steps, 10000 cool1 steps and 8000 cool2 steps. The random seed was fixed for all structure calculations.

### 3. Results Chapter 1: Investigating the Protein:Protein Interactions of the LIMD1 Scaffold Protein in the miRNA Mediated Silencing Pathway

#### 3.1. Introduction

As described in section 1.6, LIMD1 has been shown to be a component of the miRISC and is believed to scaffold AGO2 and TNRC6A proteins to facilitate miRISC formation [25]. Recent research into the role of LIMD1 in miRNA mediated silencing has focussed on investigating the function of LIMD1 *in vivo* [25, 57]. While previous investigations have shown that LIMD1 associates with miRISC proteins, it has not yet been demonstrated that LIMD1 interacts directly with either AGO2 or TNRC6A. Moreover, these interactions have not been characterised at an atomic level. As a general model for LIM domain partner protein recognition cannot be identified, the features allowing for recognition of TNRC6A by the LIM domains of LIMD1 remain unclear. It is also unclear how the disordered pre-LIM region of LIMD1 can recognise its partner proteins despite little sequence similarity to the related LAW proteins (Figure 1.7). The goal of this chapter is to add to the existing body of work on LIMD1 biology in miRNA mediated silencing by providing structural and biophysical characterisation of the LIMD1:AGO2:TNRC6A complex. It is hoped that greater understanding of molecular interactions formed by LIMD1 could provide insight into why LIMD1 is required for miRNA mediated silencing when AGO and GW182 proteins are able to interact directly. Characterisation of the structural nature of recognition in these protein:protein interactions could aid in the development of new tools to aid in the study of miRNA mediated silencing, such as inhibitors of miRISC assembly that target LIMD1 binding sites.

Investigation of the structural biology of the LIMD1 protein complexes first required the interactions to be recapitulated *in vitro*. Recapitulation of the interaction between LIMD1 and TNRC6A was attempted using peptide arrays, spanning the LIMD1

binding region of TNRC6A, and pull-down assays using various regions of LIMD1 and TNRC6A. Given that both the 180 kDa and 210 kDa isoforms of TNRC6A were seen to co-precipitate with LIMD1, this study utilised the smaller 180 kDa isoform of TNRC6A as the extended region in the larger construct is unlikely to participate in the interaction with LIMD1 [25, 193]. Although three TNRC6 proteins have been identified in humans, this study focussed on TNRC6A as the interaction with LIMD1 had already been characterised *in vivo* [25, 194]. To facilitate production of milligram quantities of TNRC6A and LIMD1 without excessive proteolytic degradation, small, overlapping regions of these proteins (spanning the previously identified binding regions) were used for recombinant expression in *E. coli*. Demonstration of an interaction between LIMD1 and AGO2 was attempted using pull-down assays but problems with production of AGO2 prevented additional study on this interaction.

## 3.2. Results and Discussion

### 3.2.1. Investigating the LIMD1:TNRC6a Interaction using Peptide Arrays

LIMD1 has been shown to interact with TNRC6A through the LIM domain region of LIMD1, with LIMD1 $\Delta$ LIM123 unable to bind to TNRC6A [25]. It was also demonstrated, through the loss of an interaction between LIMD1 with TNRC6A $\Delta$ 1-883, that LIMD1 binds within the Argonaute binding region (1-883) of TNRC6A [25]. Initial experiments were focussed on further narrowing down the binding region of LIMD1 on TNRC6A, in order to facilitate biophysical investigation of the interaction. Inspired by the work of Pfaff *et al.* (2013) on the AGO2:TNRC6 interaction, identification of the LIMD1 binding site was first attempted using peptide arrays [90]. The full-length LIM domain region (LIM123) was produced in *E. coli* as a maltose binding protein (MBP) fusion protein and purified by immobilised metal affinity chromatography (IMAC) (Figure 3.1).



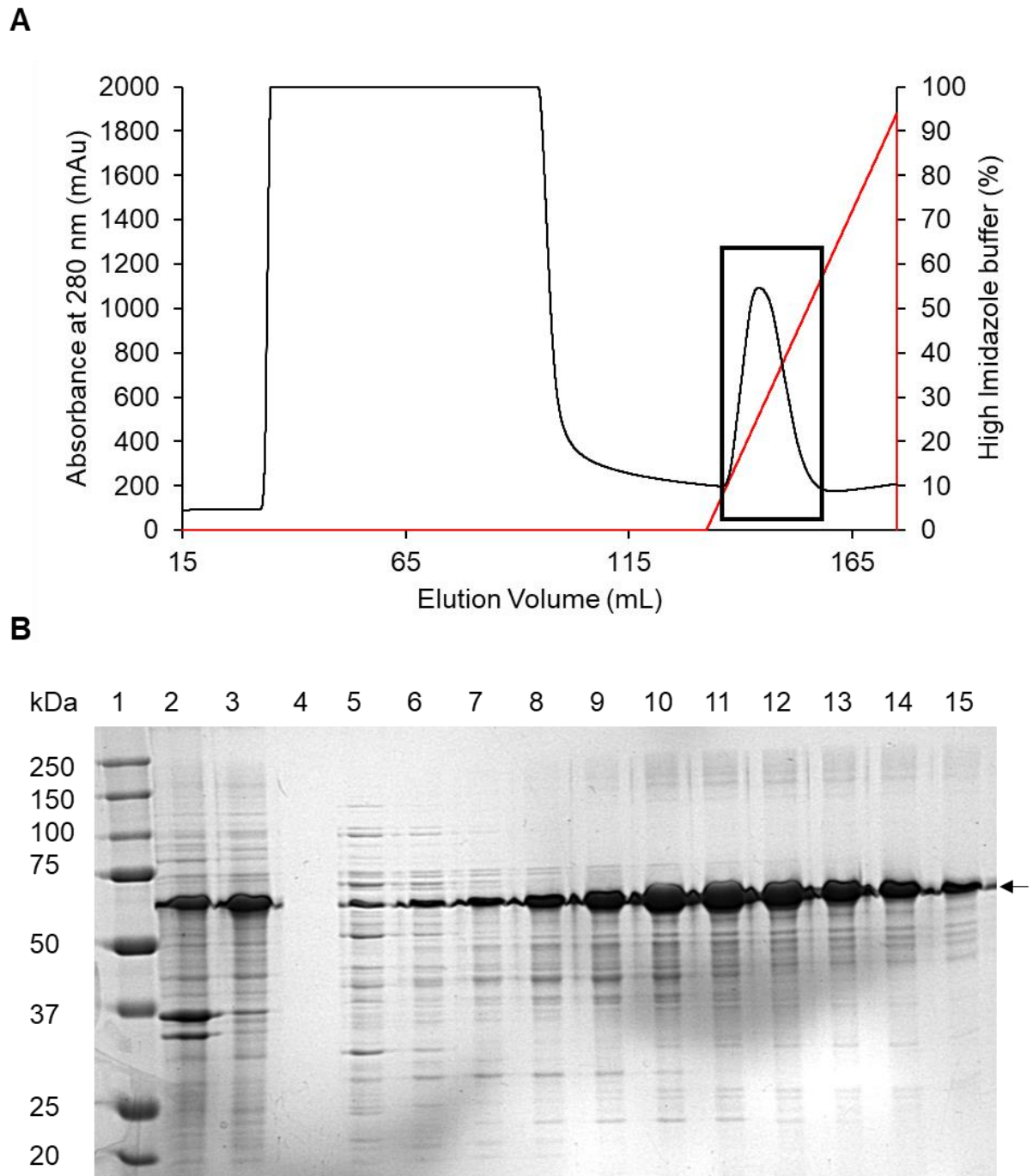


Figure 3.1. **A.** Chromatogram for the IMAC purification of MBP tagged LIM123. Fractions taken under the boxed region were analysed by SDS-PAGE. **B.** SDS-PAGE analysis of the single step IMAC purification of MBP tagged LIM123. Lane 1: molecular mass ladder, lane 2: total protein before lysis, lane 3: soluble protein after lysis, lanes 5-15: fractions from the elution peak shown in **A**. The expected location of MBP-tagged LIM123 is indicated by an arrow. MBP-LIM123 expected mass = ~67 kDa

Densitometry analysis using imageJ revealed that lanes 10-15 in Figure 3.1B LIM123 represented approximately 83% of the total protein [195]. This level of purity was deemed sufficient for use with peptide arrays where binding would be detected using a specific antibody, meaning high levels of purity were not required. A peptide array was used covering TNRC6A 1-900 that featured 15 residue long peptides with a 5-residue shift arrayed on a trioxatridecanediamine membrane (Figure 3.2). The length of peptides chosen was based on analysis of structures of LIM domains bound to linear partners (Figure 1.3), which showed 15 amino acids to be sufficient for binding to a single LIM domain. LIM123 was incubated with the array, unbound protein was washed off and bound protein was detected by probing with a horse radish peroxidase (HRP) coupled anti-His<sub>6</sub> antibody. His<sub>6</sub> control peptides were included in the corners of the array to test for antibody binding.

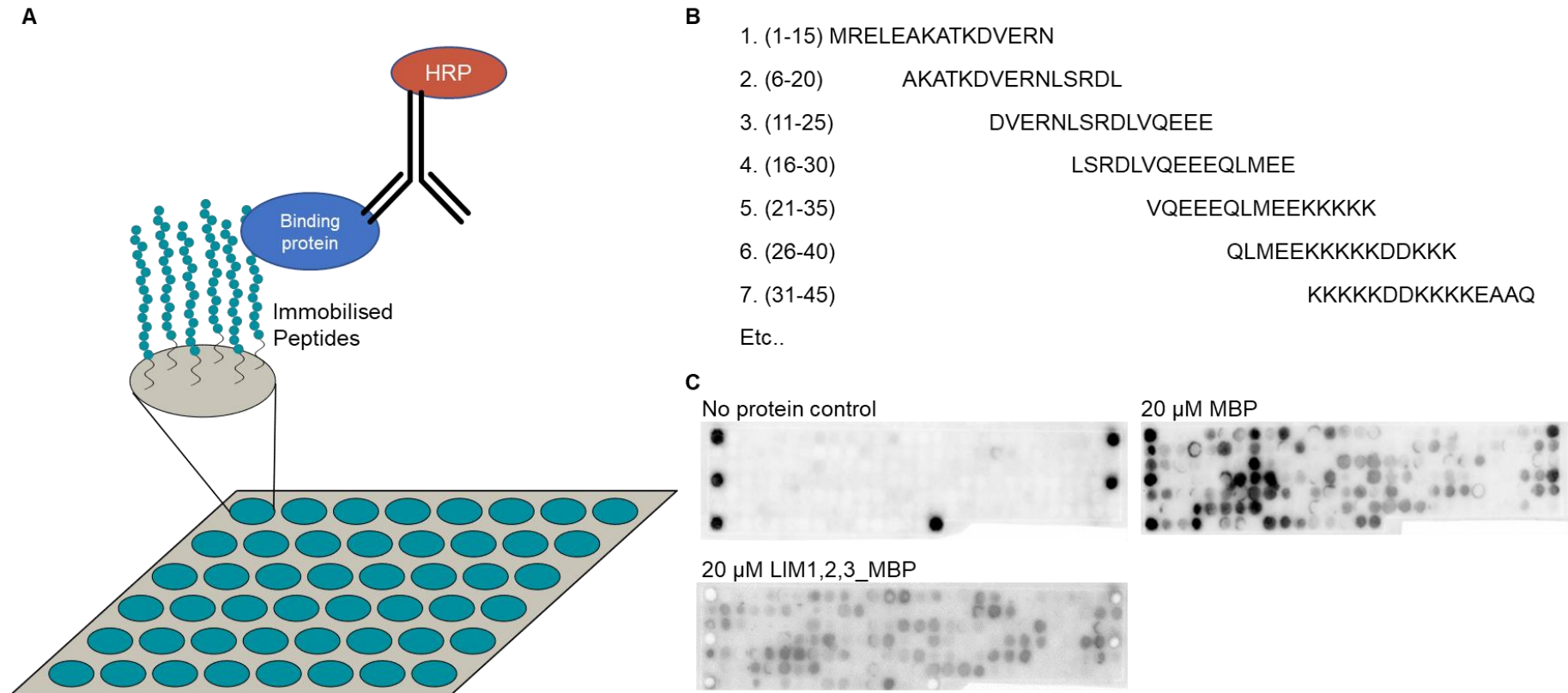


Figure 3.2. **A.** Schematic representation of the peptide array method where peptides are immobilised on a membrane and binding of partner proteins can be probed using a HRP conjugated anti-body. **B.** Examples of the first seven peptide sequences immobilised on the arrays. The 15 residue peptides are offset by 5 residues to prevent cleaving the binding site. **C.** Initial peptide array results without any partner protein, with LIM123-MBP or with just MBP (2 s exposures).

Initial results with the peptide arrays showed a strong response for the MBP control compared to MBP-LIM123 (Figure 3.2C). To remove the problem of potential non-specific binding by MBP a new LIM123 construct was produced incorporating a C-terminal His<sub>6</sub> tag, allowing for removal of the N-terminal MBP solubility tag by passing the cleaved protein through an MBPTrap column whilst still retaining a His<sub>6</sub> tag for detection (Figure 3.3).

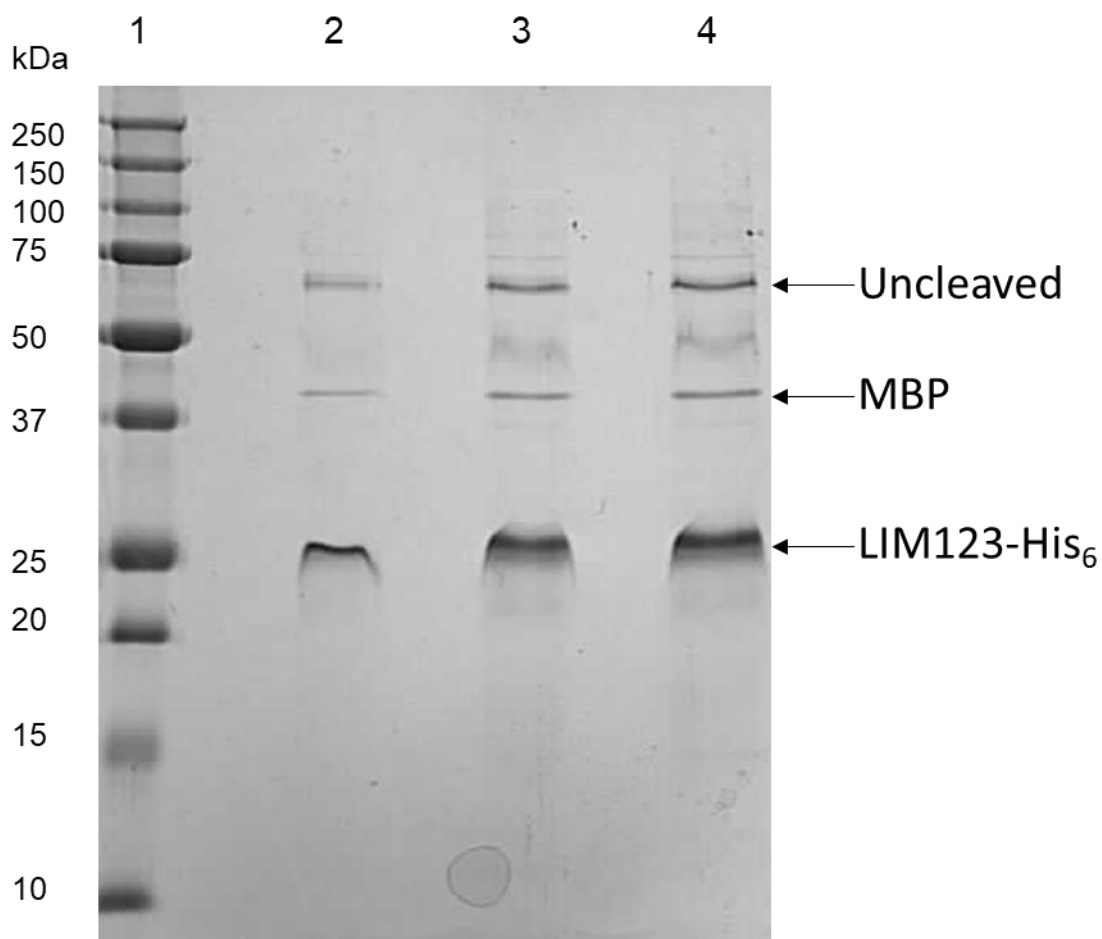
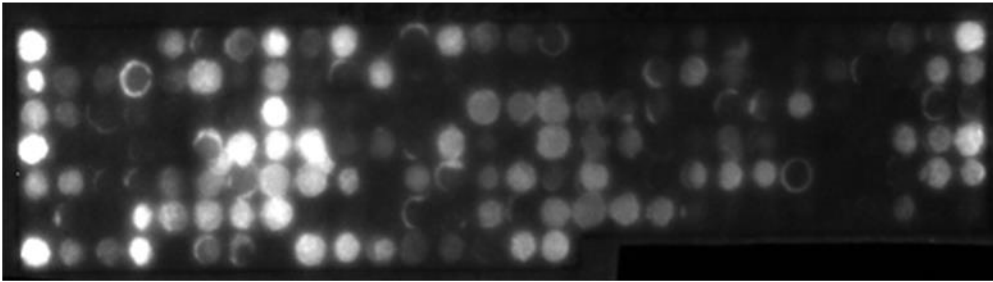


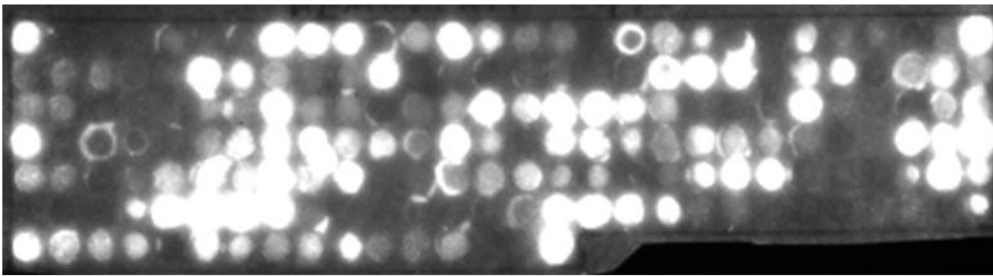
Figure 3.3. SDS-PAGE analysis of the purity of LIM123-His<sub>6</sub> protein post MBP tag cleavage and being passed over an MBPTrap column. Lane 1: molecular mass standards, lane 2-4: LIM123-His<sub>6</sub> sample serially diluted by a factor of 2 with the most concentrated sample in lane 4 and least concentrated in lane 2.. Expected mass of the cleaved protein = ~26 kDa, expected mass of tagged protein = 68 kDa, expected mass of MBP = 42 kDa.

Unfortunately, some residual MBP and uncleaved MBP-LIM123 remained in the sample after passing the sample through an MBPtrap column. The residual tag and uncleaved protein could not be removed by hydrophobic interaction chromatography (HIC) or anion exchange purification. It was hoped that the proportion of MBP in the sample had been sufficiently reduced to allow for use of the protein with peptide arrays. Initial experiments with the LIM123-His<sub>6</sub> showed that the protein has a stronger interaction with the array than MBP but that binding occurred across the array at the same locations as MBP, suggesting that the observed interactions were non-specific (Figure 3.4**A-B**).

**A.** 20 $\mu$ M MBP in TBS



**B.** 20 $\mu$ M LIM123-His<sub>6</sub> in TBS



**C.** 2  $\mu$ M LIM123his in blocking buffer

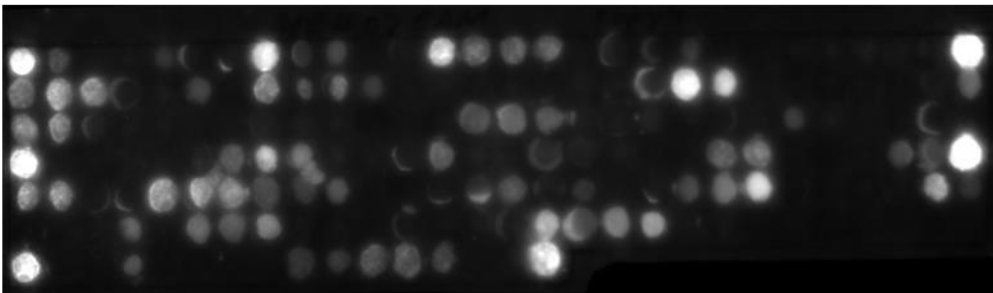


Figure 3.4. Peptide array assays with **A.** 20  $\mu$ M MBP (2 s exposure), **B.** 20  $\mu$ M LIM123-His<sub>6</sub> (2 s exposure), or **C.** 2  $\mu$ M LIM123-His<sub>6</sub> (30 s exposure) using a more stringent protocol.

In an attempt to reduce non-specific interactions of LIM123-His<sub>6</sub> with the array, the conditions of the incubation step were changed. The concentration of LIM123-His<sub>6</sub> was reduced from 20  $\mu$ M to 2  $\mu$ M and the incubation was carried out in blocking buffer instead of TBS. Although the amount of overall binding was reduced by these steps, no specific binding site could be identified (Figure 3.4**C**). The strongest response from the array corresponded to a false positive response seen in the no protein controls (Figure 3.2**C**). The lack of specific binding seen led to a reassessment of the quality of the LIM123 protein used. Size exclusion

chromatography of the protein revealed that the protein eluted in the void volume, showing that the protein was likely aggregated (Figure 3.5A and C). The protein was also seen to have a brown colour at higher concentration (Figure 3.5B).

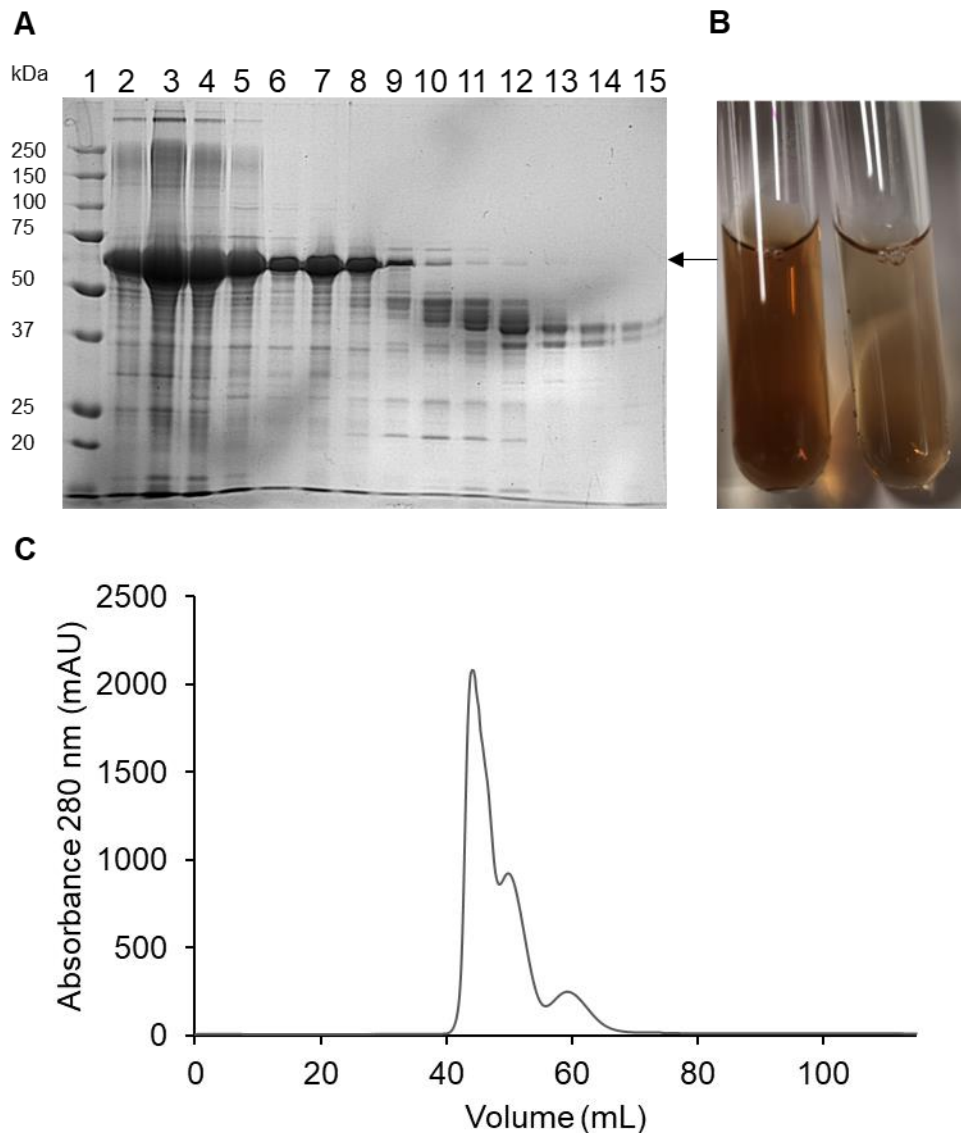


Figure 3.5. **A.** SDS-PAGE analysis of the SEC purification of MBP-LIM123. Lane 1: molecular mass standards, lanes 2-15 correspond to fractions of the peaks in the chromatogram in **C**. **B.** A picture of the MBP-LIM123 sample at two different concentrations showing the brown colour associated with the protein at higher concentrations. **C.** Chromatogram for the SEC purification of MBP-LIM123 showing the protein eluting much earlier than expected.

The brown colour seen was most likely the result of iron binding which is a known problem associated with the expression of LIM domains in *E. coli* [196, 197]. Performing the expression and purification in the presence of zinc did not have an effect on the colour of yields of protein and attempts to refold the protein in the presence of zinc were unsuccessful. If the LIM domains were loaded with iron rather than zinc any interactions, or lack of interactions, observed could be the result of the iron binding. For this reason, the LIM123 protein was deemed to be unusable. As an alternative, production of the individual LIM domains was attempted. After IMAC purification the individual LIM domains were seen to still have a brown colour suggesting the individual domains were also binding iron. As LIM domains are known to bind zinc *in vivo*, it was not desirable to have iron bound LIM domains which may behave differently than zinc bound LIM domains [198]. Iron binding also appears to drive aggregation of the LIM domains which is undesirable as the LIM domains are not thought to function as an aggregate. The SEC trace revealed the samples contained a mixture of aggregated and unaggregated protein (Figure 3.6A). It was seen that the aggregated protein retained the brown colour whilst the unaggregated protein, eluted at ~85 mL, was colourless. This suggested that the unaggregated protein was free of iron contamination. UV/Vis spectroscopy provided additional evidence for the removal of the iron contamination by SEC. Peaks in the UV/Vis spectra at 330 nm and 410 nm, reported to be caused by iron-sulfur clusters, were seen to be present in the sample prior to SEC purification but not after (Figure 3.6B). Peaks were seen above 9 ppm, between 5 and 6 ppm, and below 0 ppm in 1D <sup>1</sup>H NMR spectra of the individual LIM domains (Figure 3.6C). This dispersion of peaks confirmed the LIM domains were folded as unfolded proteins would have a much more narrow distribution of chemical shifts.



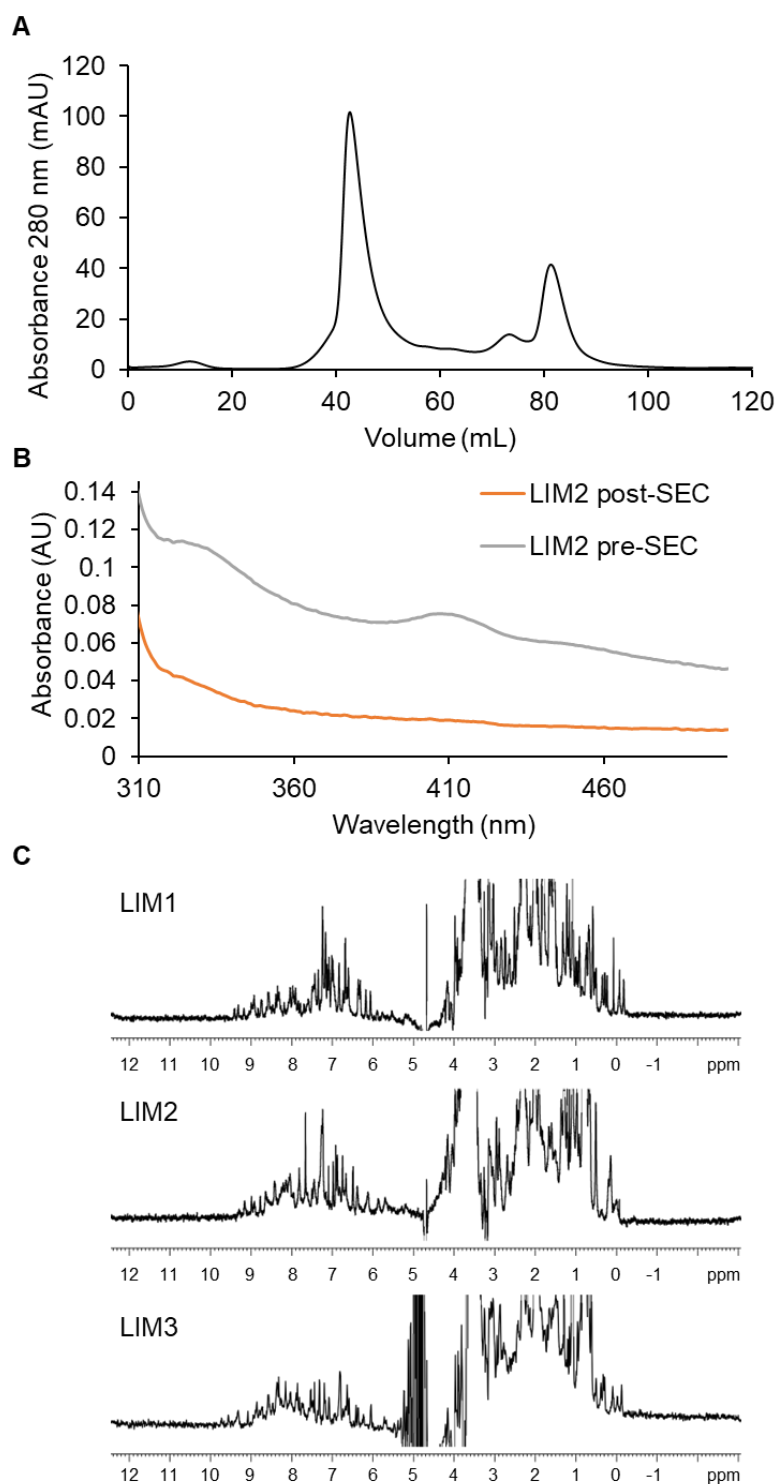


Figure 3.6. **A.** Chromatogram for the size exclusion purification of untagged LIM2 showing two major peaks. The peaks at 45 mL and 90 mL contained LIM2 with the early peak most likely to be a soluble aggregate. **B.** UV/Vis spectra of the LIM2 sample before and after SEC purification showing the loss of putative charge transfer bands at 330 nm and 410 nm. **C.** 1D  $^1\text{H}$  NMR spectra of the individual LIM domains showing a broad peak distribution consistent with the protein being folded.

It is possible that, as previously seen, some LIM domains are incorrectly loaded with iron during *E. coli* expression, and this then leads to mis-folding and aggregation of the iron containing LIM domains [197, 198]. With more confidence in the LIM domain proteins, TNRC6A peptide arrays were attempted again. It was hoped that the MBP tag of the LIM domains would bind less strongly than a specific interaction with the LIM domains and so the tag was not removed for these experiments. In an effort to improve the workflow and to develop new methods of analysing peptide arrays, the arrays were probed using a Ni-NTA-Atto 488 conjugate that can be detected by fluorescence, and so removes the need for chemiluminescent detection. Arrays were probed with each individual LIM domain using the Ni-NTA-Atto conjugate, but no sign of an interaction was seen (Figure 3.7A). Binding of the Ni-NTA-Atto conjugate was tested by probing an SDS-PAGE gel containing LIM1, LIM2, LIM3 and LIM123. The Ni-NTA-Atto conjugate was seen to give good signals for LIM1 and LIM123 but worse signals for LIM2 and LIM3 (Figure 3.7B). In order to test if the lack of an observed interaction was due to the use of the Ni-NTA-Atto conjugate or a lack of bound protein, the assay was repeated with LIM2 and probed with anti-His<sub>6</sub> antibody (Figure 3.7C). LIM2 was expected to interact with TNRC6A due to the previously observed binding mode of LIM domains binding to disordered partners through tandem LIM domains. This suggested that TNRC6A would likely bind across LIM1 and 2, LIM2 and 3 or all three LIM domains. TNRC6A binding to LIM2 was also supported by unpublished results from a collaborator.

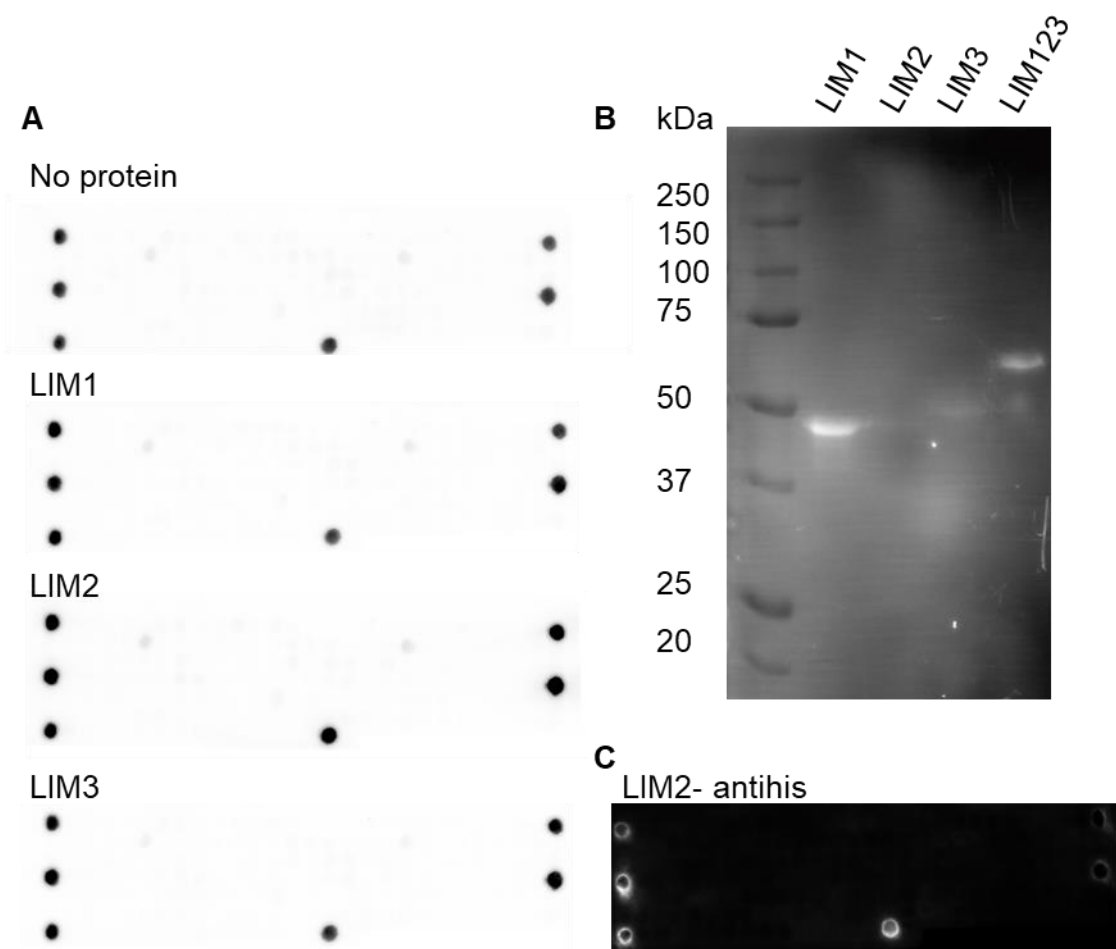


Figure 3.7. **A.** TNRC6A peptide arrays with no protein or the individual LIM domains of LIMD1 probed with an ATTO 488 Ni-NTA conjugate. **B.** SDS-PAGE of the individual LIM domains and LIM123 probed with the ATTO 488 Ni-NTA conjugate. Bands were seen for all LIM domain proteins although with different intensities. **C.** TNRC6A peptide array with LIM2 probed with anti-His antibody (Sigma A7058) with a 2 s exposure.

No sign of binding was observed with the peptide arrays under any of the conditions tested. The failure of this method to capture an interaction could be due to the length of the peptides being too short, the interaction being too weak to survive the multiple wash steps, the need for posttranslational modification or other factors affecting the interaction. At this stage, very few of the originally purchased arrays remained to be tested and regeneration of the arrays was not recommended or successful when attempted. Due to the high cost of peptide arrays, and the lack of success in

characterising the interaction between LIMD1 and TNRC6A, a new method was attempted to probe the interaction.

### 3.2.2. Investigating the LIMD1:TNRC6A Interaction using Pull-Down Assays

Pull-down assays were developed using magnetic Ni-NTA beads to attempt to capture a TNRC6A:LIMD1 interaction. Briefly, His-tagged bait protein is bound to the magnetic beads, untagged prey protein is then incubated with the beads, unbound protein is washed off and bound protein is eluted with an imidazole buffer. The proteins in each fraction can then be analysed by SDS-PAGE. His<sub>6</sub>-MBP-TNRC6A 1-883 and untagged LIM domains were produced for use as bait and prey proteins respectively (Figure 3.8). As the LIM domains are produced as MBP fusions, removal of the large His<sub>6</sub>-MBP tag is easily verified by SDS-PAGE analysis with the uncleaved proteins being ~43 kDa larger than the LIM domains (Figure 3.8C). The long, disordered, TNRC6A 1-883 did not purify to a high standard. It was thought that the poor purity of TNRC6A 1-883 was likely due to degradation of the protein. This was tested by performing a western blot of the TNRC6A 1-883 sample with anti-His<sub>6</sub> antibody to check if the lower molecular weight proteins contained a His<sub>6</sub>-tag (Figure 3.8B).

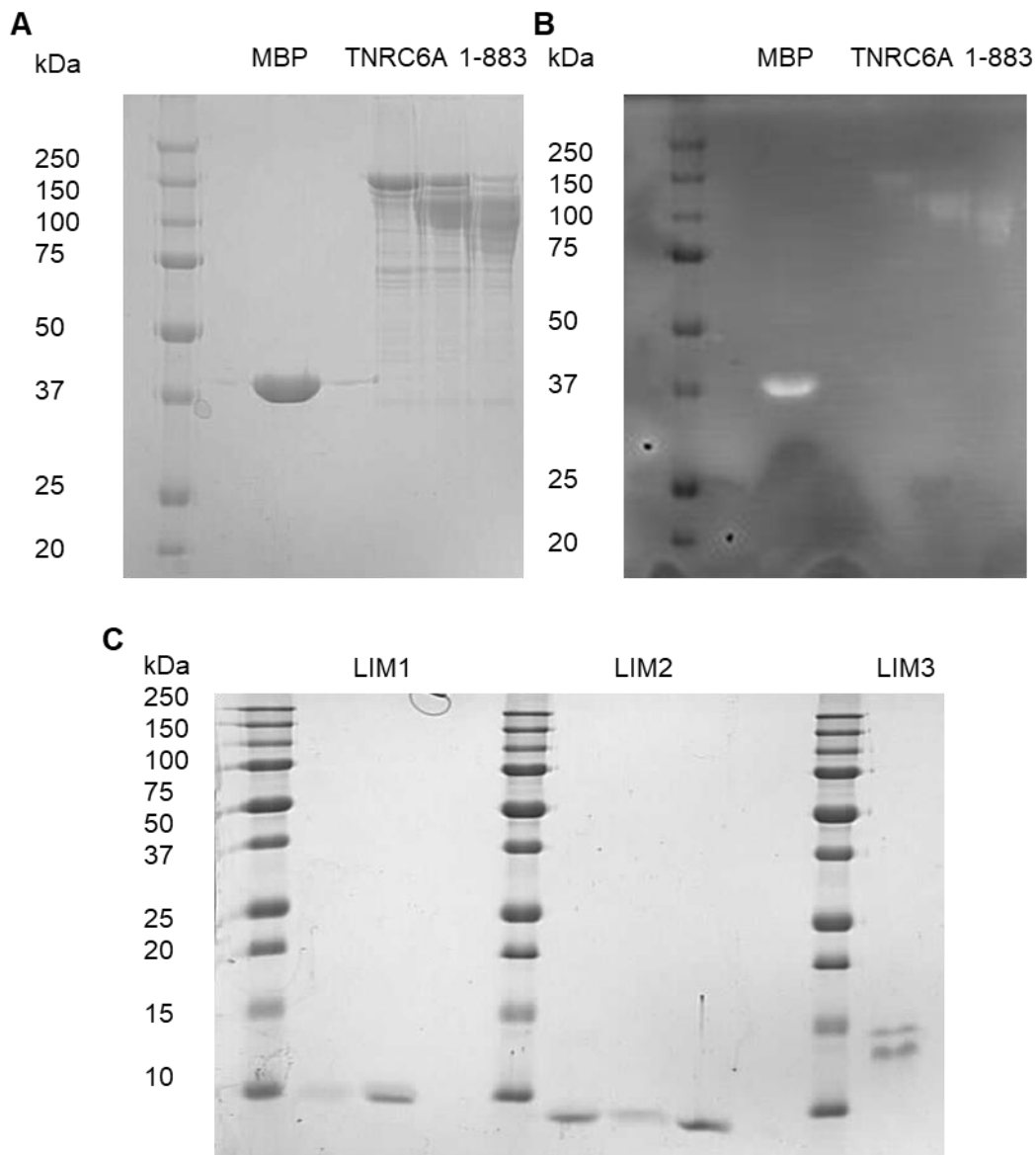


Figure 3.8. **A.** SDS-PAGE analysis of MBP and TNRC6A 1-883 fractions from SEC. Expected molecular weights: MBP = ~43 kDa, TNRC6A 1-883 = ~135 kDa. **B.** Western blot of MBP and TNRC6A SEC fractions probed with 1:120000 (v/v) anti-His antibody (Sigma A7058). **C.** SDS-PAGE analysis of the SEC purification of individual LIM domains. The individual domains run aberrantly on SDS-PAGE but positions are consistent with those previously observed in the Plevin group. Expected molecular weights: LIM1 = ~7 kDa, LIM2 = ~7 kDa, LIM3 = ~9 kDa.

The presence of a His<sub>6</sub> tag on the lower molecular weight proteins seen on the TNRC6A SDS-PAGE suggests that these proteins are likely to be products of degradation. The presence of degradation products was unlikely to cause a

significant impact on LIMD1 binding as it is likely that LIMD1 could also bind to some of these degradation products (unless LIMD1 binds to the C-terminal of TNRC6A). Pull down assays were carried out with the individual LIM domains and TNRC6A 1-883. Initial results from the pull-down assays appeared to show a very faint band for each of the LIM domains in the elution fractions with TNRC6A but not in the MBP control (Figure 3.9 and Figure 3.10). Although promising, these results remained unconvincing due to the low intensity of the LIM domain elution band and the poor purity of the TNRC6A 1-883 sample.

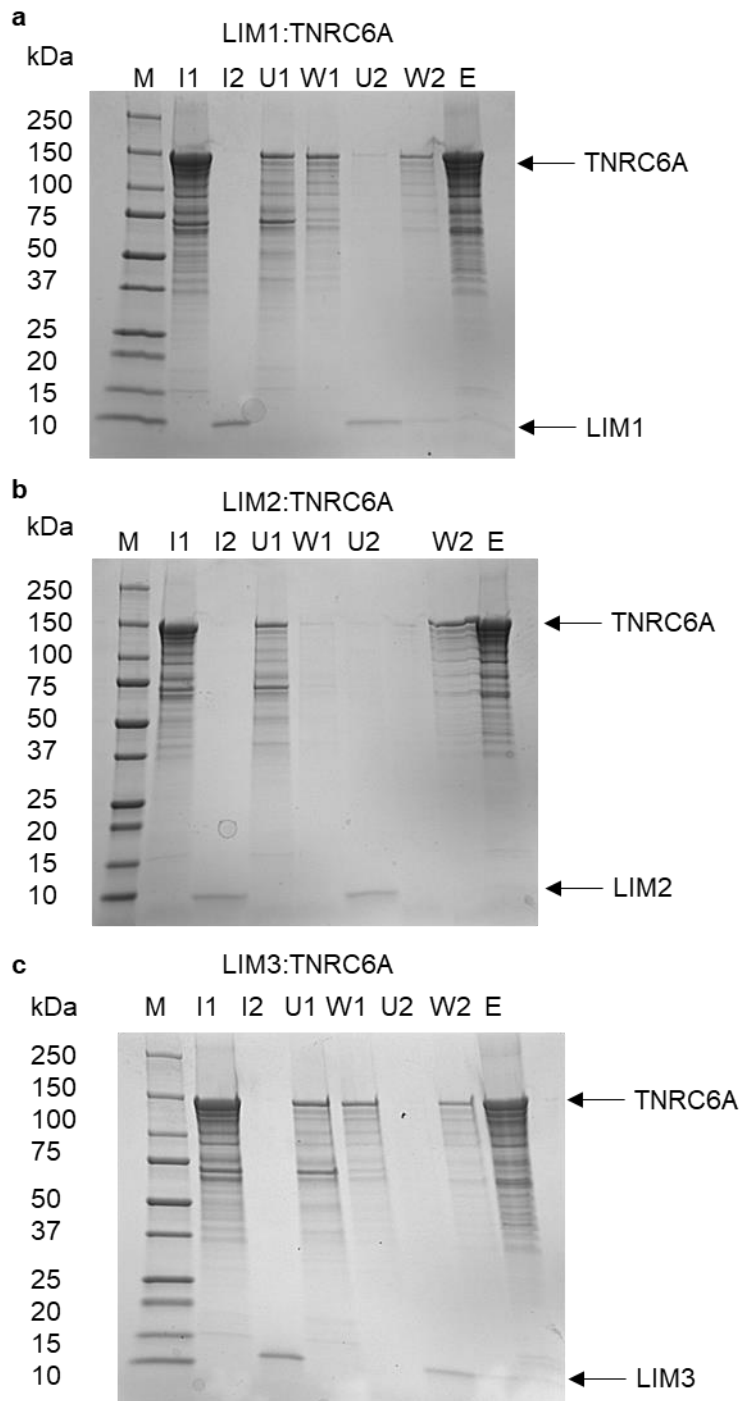


Figure 3.9. SDS-PAGE analysis of pull-down assays for **a.** LIM1 with MBP-TNRC6A 1-883, **b.** LIM2 with MBP-TNRC6A 1-883, **c.** LIM3 with MBP-TNRC6A 1-883. Lanes correspond to M = Molecular weight markers, I1 = bait protein input, I2 = prey protein input, U1 = unbound bait protein, W1 = first wash step, U2 = unbound prey protein, W2= second wash step, E = Elution. Expected molecular weights: LIM1 = ~7 kDa, LIM2 = ~7 kDa, LIM3 = ~9 kDa, MBP-TNRC6A 1-883 = 135 kDa.

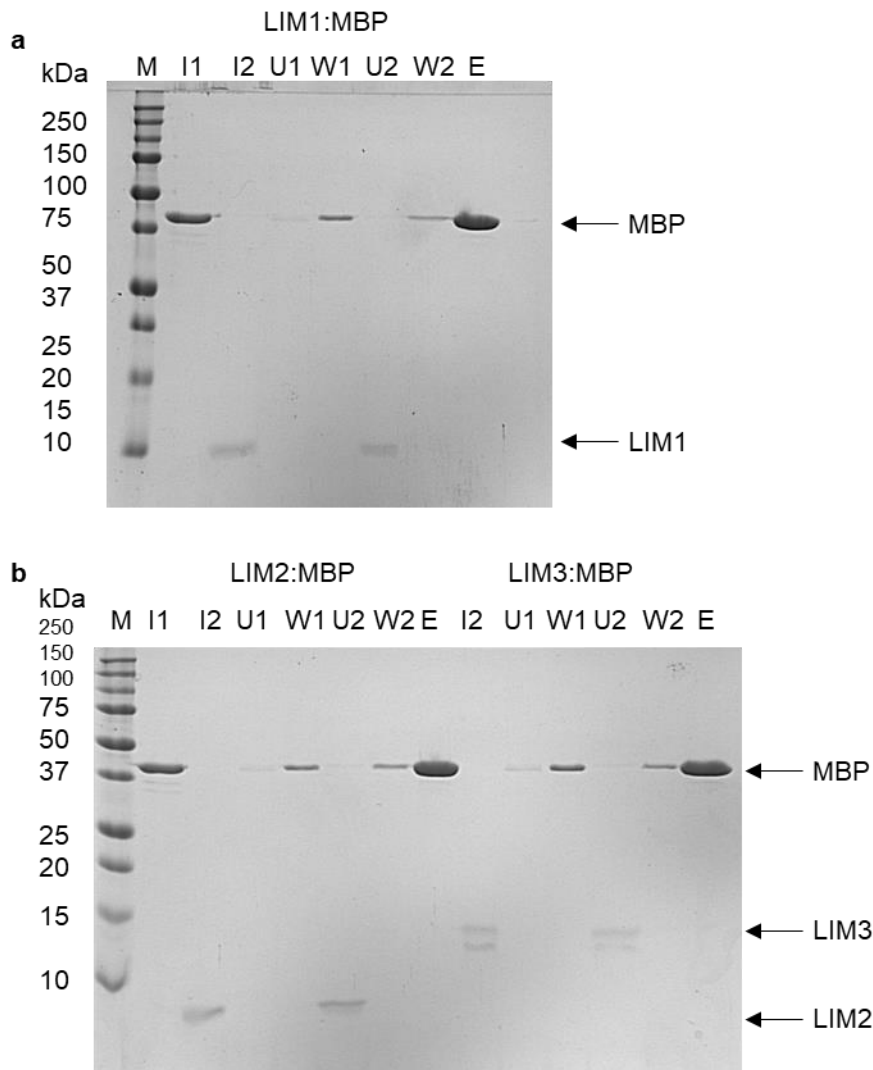


Figure 3.10. SDS-PAGE analysis of pull-down assay controls for **a**. LIM1 with MBP and **b**. LIM2 with MBP and LIM3 with MBP. Lanes correspond to M = Molecular weight markers, I1 = bait protein input, I2 = prey protein input, U1 = unbound bait protein, W1 = first wash step, U2 = unbound prey protein, W2= second wash step, E = Elution. Expected molecular weights: LIM1 = ~7 kDa, LIM2 = ~7 kDa, LIM3 = ~9 kDa, MBP = 43 kDa.

To allow for a higher purity TNRC6A sample, and to narrow down the LIMD1 binding site on TNRC6A, a range of smaller TNRC6A constructs were produced (Figure 3.11). The new TNRC6A constructs were designed to cover TNRC6A 1-883 with overlapping sequences and to avoid any known or predicted binding motifs. These



proteins were then used for pull-down assays with LIM2 (Figure 3.11). As discussed previously, the binding mode of LIM domains with linear motifs, and unpublished data from a collaborator suggested that LIM2 binds more strongly to TNRC6A than LIM1 or LIM3. For this reason, further experiments were mostly focussed on recapitulating an interaction between TNRC6A and LIM2.

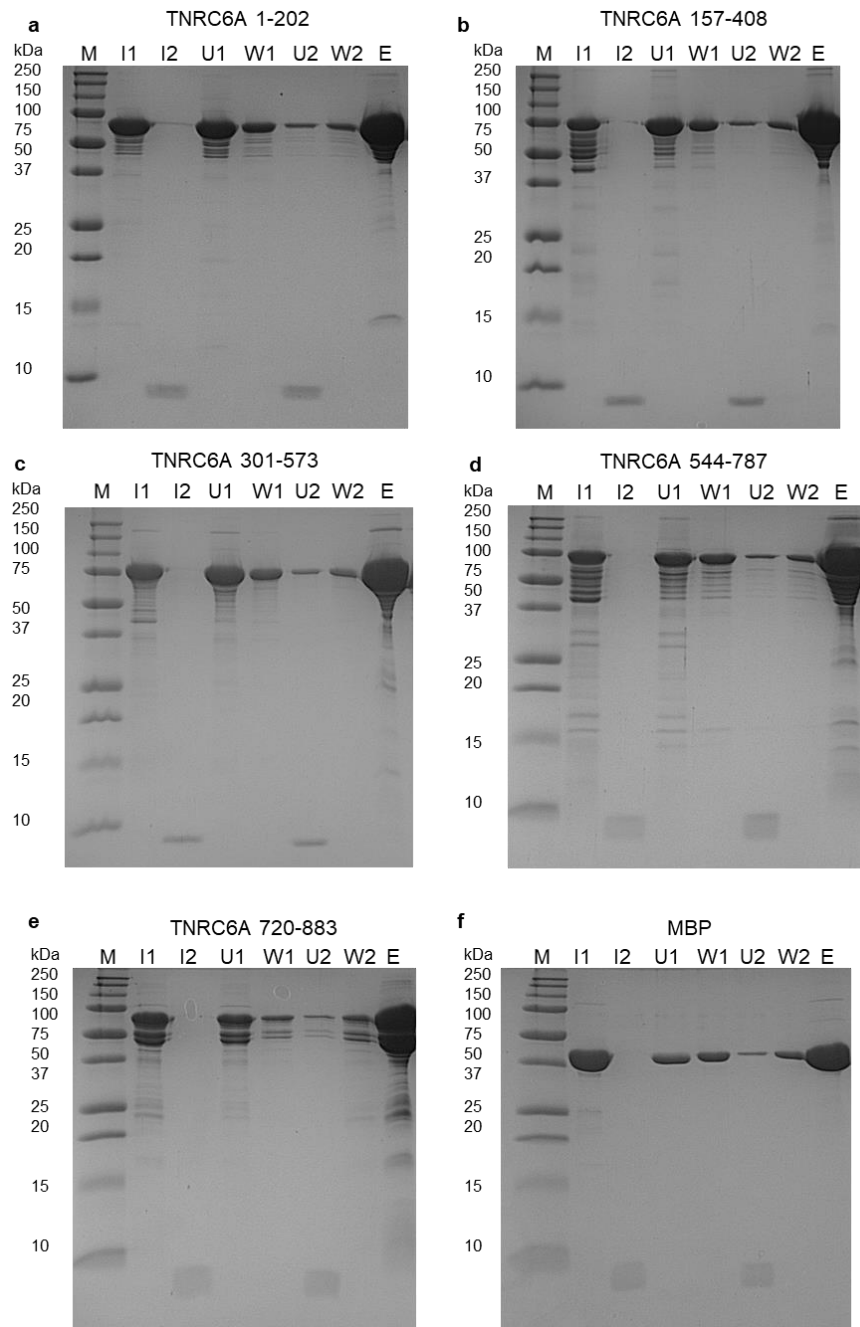


Figure 3.11. SDS-PAGE analysis of pull-down assays for LIM2 with **a.** MBP-TNRC6A 1-202, **b.** MBP-TNRC6A 157-408, **c.** MBP-TNRC6A 301-573, **d.** MBP-TNRC6A 544-787, **e.** MBP-TNRC6A 720-883 **f.** MBP. Lanes correspond to M = Molecular weight markers, I1 = bait protein input, I2 = prey protein input, U1 = unbound bait protein, W1 = first wash step, U2 = unbound prey protein, W2= second wash step, E = Elution. Expected molecular weights: LIM2 = ~7 kDa, MBP-TNRC6A 1-202 = 62 kDa, MBP-TNRC6A 157-408 = ~67 kDa, MBP-TNRC6A 301-573 = ~71 kDa, MBP-TNRC6A 544-787 = ~68 kDa, MBP-TNRC6A 720-883 = ~59 kDa, MBP = 43 kDa.

No prey proteins were seen in any of the elution fractions when using the new TNRC6A constructs. It was thought that the lack of observed binding may be due to a lower affinity with the smaller TNRC6A constructs. An alternative pull-down assay protocol was tested, in which the concentrations of both bait and prey proteins were increased and an acetone precipitation step was added to extract protein in the elution fraction for gel analysis (Figure 3.12).

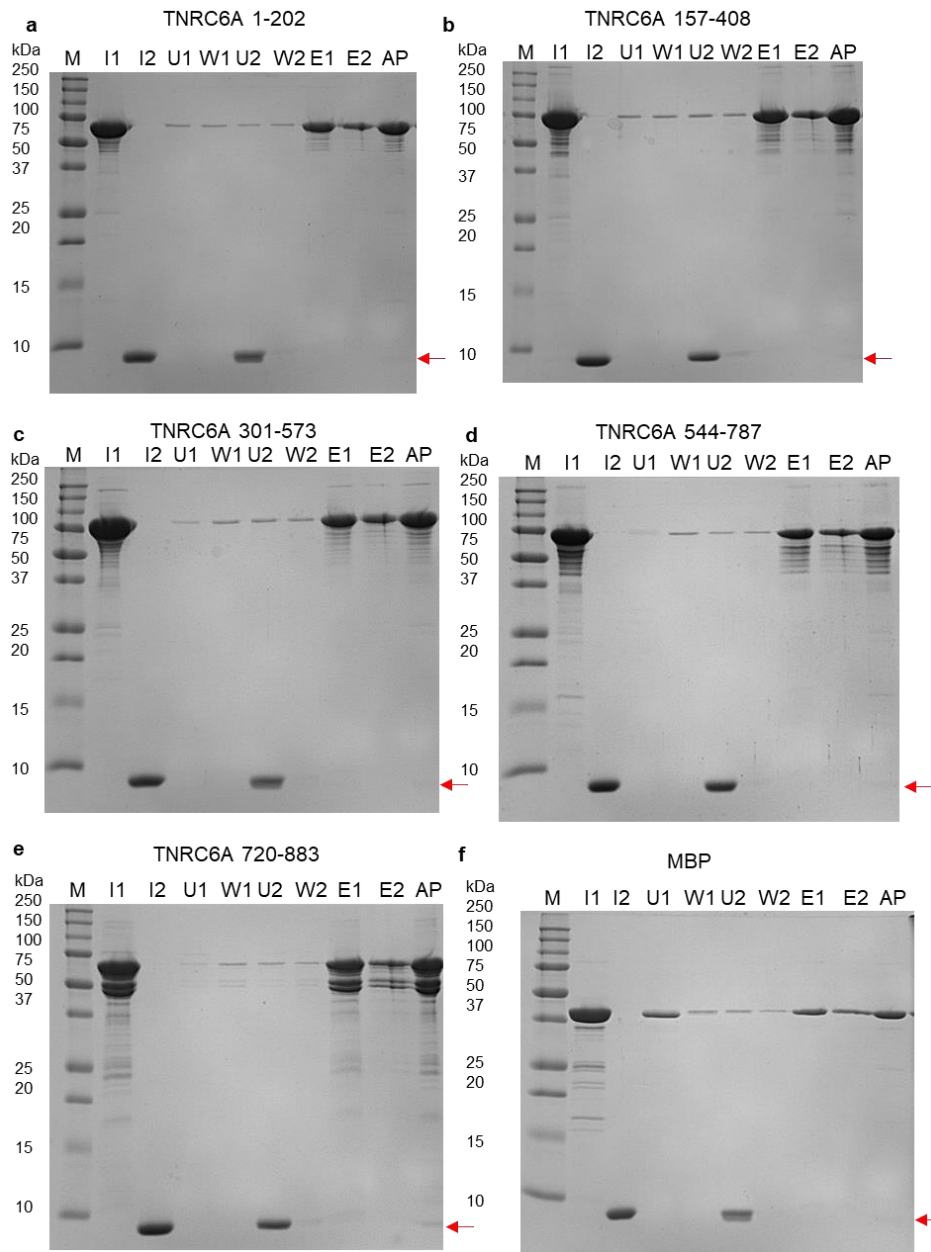


Figure 3.12 SDS-PAGE analysis of more sensitive pull-down assays for LIM2 with **a.** MBP-TNRC6A 1-202, **b.** MBP-TNRC6A 157-408, **c.** MBP-TNRC6A 301-573, **d.** MBP-TNRC6A 544-787, **e.** MBP-TNRC6A 720-883 **f.** MBP. Lanes correspond to M = Molecular weight markers, I1 = bait protein input, I2 = prey protein input, U1 = unbound bait protein, W1 = first wash step, U2 = unbound prey protein, W2= second wash step, E1 = first elution, E2 = second elution, AP = acetone precipitated elution fraction. Expected molecular weights: LIM2 = ~7 kDa (indicated by red arrow), MBP-TNRC6A 1-202 = 62 kDa, MBP-TNRC6A 157-408 = ~67 kDa, MBP-TNRC6A 301-573 = ~71 kDa, MBP-TNRC6A 544-787 = ~68 kDa, MBP-TNRC6A 720-883 = ~59 kDa, MBP = 43 kDa.

No indication of binding of prey proteins was seen in the untreated elution fractions. Conversely, following acetone precipitation, it was possible to see faint bands corresponding to LIM2 with TNRC6A 1-202, 157-408, 544-787, 720-883 and MBP (Figure 3.12a,b,d,e, and f). This is likely due to increasing the sensitivity of the assay to a point where residual protein, not removed by the wash steps, was observed. In order to confirm the assays were working as intended, a pull-down assay was used to demonstrate an interaction between TNRC6A and AGO2 (Figure 3.13a-b). Two constructs of TNRC6A were used; TNRC6A 720-883 that contains AGO2 binding motif II and hook, and TNRC6A 1-202 that is not believed to bind AGO2 [91, 194]. As expected, AGO2 was pulled down by TNRC6A 720-883 but not by TNRC6A 1-202. This result confirmed that the basic pull-down assay protocol works as intended.

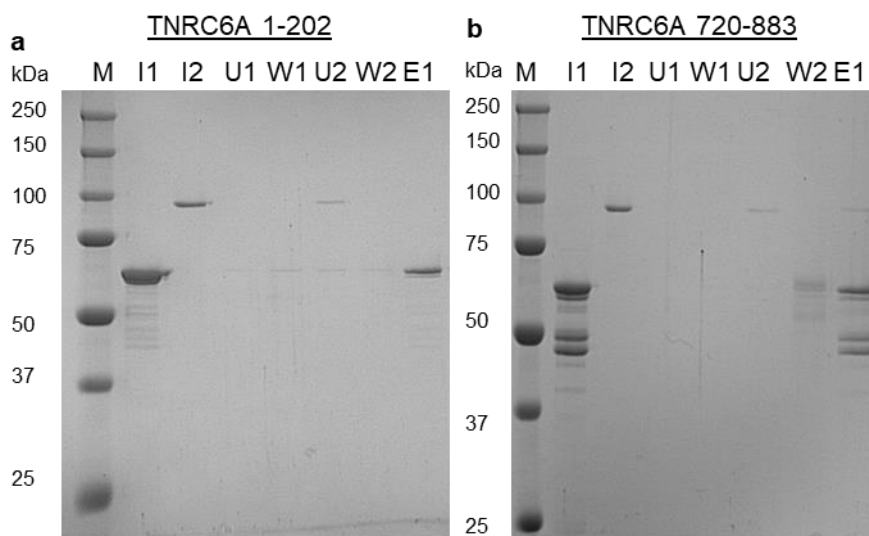


Figure 3.13. AGO2 pull-down assays with **a**. MBP-TNRC6A 1-202 and **b**. MBP-TNRC6A 720-883. Lanes correspond to M = Molecular weight markers, I1 = bait protein input, I2 = prey protein input, U1 = unbound bait protein, W1 = first wash step, U2 = unbound prey protein, W2= second wash step, E1 = Elution. Expected molecular weights: AGO2 = ~97 kDa, MBP-TNRC6A 1-202 = ~62 kDa, MBP-TNRC6A 720-883 = ~59 kDa.

In order to investigate the interaction between LIMD1 and AGO2 a construct of LIMD1 spanning 110-166 (the previously identified AGO2 binding region) was produced and used for a pull-down assay with AGO2 (Figure 3.14a). LIMD1 110-166 was not able to pull-down AGO2 (Figure 3.14b). The lack of LIMD1 binding observed could be caused by a lower affinity of the LIMD1:AGO2 interaction compared to the TNRC6A:AGO2 interaction or potentially be due to the 110-166 region not containing a complete AGO2 binding motif.

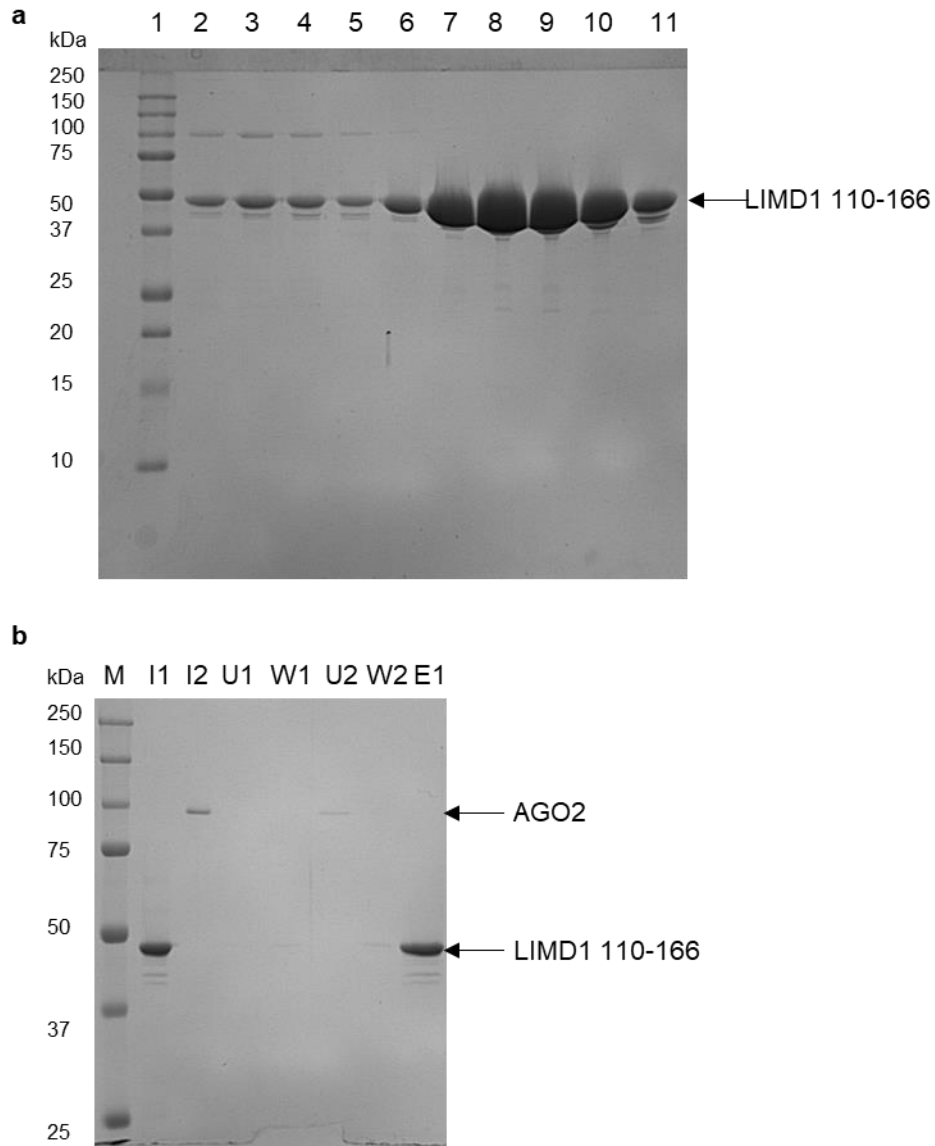


Figure 3.14. **a**. SDS-PAGE analysis for the SEC purification of MBP tagged LIMD1 110-166. Lane 1: molecular mass standards, lanes 2-11 fractions from the SEC purification. Fractions 7-10 were pooled and used for pull down assays.**b**. AGO2 pull-down assay with LIMD1 110-166. Lanes correspond to M = Molecular weight markers, I1 = bait protein input, I2 = prey protein input, U1 = unbound bait protein, W1 = first wash step, U2 = unbound prey protein, W2= second wash step, E1 = Elution. Expected molecular weights: AGO2 = ~97 kDa, LIMD1 110-166 = ~50 kDa.

Due to the lack of success in investigating the LIMD1:TNRC6A interaction, no further pull-down experiments were attempted. The difficulties in recapitulating the interaction *in vitro* could be caused by many factors, including the need for post-translational modifications, additional partner proteins or different buffer conditions. The lack of observed binding could also be due to the proteins existing in an equilibrium of bound and unbound states that would be driven towards the unbound state by the successive wash steps necessary with the techniques used. It may have been possible to enhance the affinity of the interaction by using tandem LIM domain constructs, although these are highly unstable. Future investigations could focus on using the smaller TNRC6A constructs that had been produced for NMR analysis of the interaction, similarly to the work presented in chapter 4. Alternatively, chemical cross linking could be used to covalently trap transient interactions.

### 3.2.3. AlphaFold2 Guided Investigation of the LIMD1:TNRC6A Interaction

During the course of this project AlphaFold 2, an artificial intelligence (AI) based predictor of protein structure, was released [163]. AlphaFold 2 has been shown to accurately and reliably predict the 3D structure of proteins from the amino acid sequence. This work was then later built upon with AlphaFold multimer, which is designed to predict the structure of protein-protein complexes [164]. AlphaFold multimer was shown to accurately predict the binding interface in heteromeric complexes in 70% of cases, however the quality of the structure predictions are often of poor quality [164]. AlphaFold multimer was employed here to investigate the interaction between LIMD1 and TNRC6A. Unpublished data from collaborators indicated that TNRC6A binds to the first two LIM domains of LIMD1 and so these were used for structure predictions. Running AlphaFold multimer predictions for the first two LIM domains of LIMD1 (LIM12) with TNRC6A 1-900 revealed a region of TNRC6A (712-741) predicted to interact with LIM1 and LIM2 (Figure 3.15A). The potential region identified corresponded to Motif II on TNRC6A, which is one of three regions on TNRC6A that have been previously identified as being responsible for interacting with AGO2 [194]. in order to provide more context for the structure



prediction, a second model was generated in which the region of AGO2 that interacts with TNRC6A was included in the prediction (Figure 3.15B).

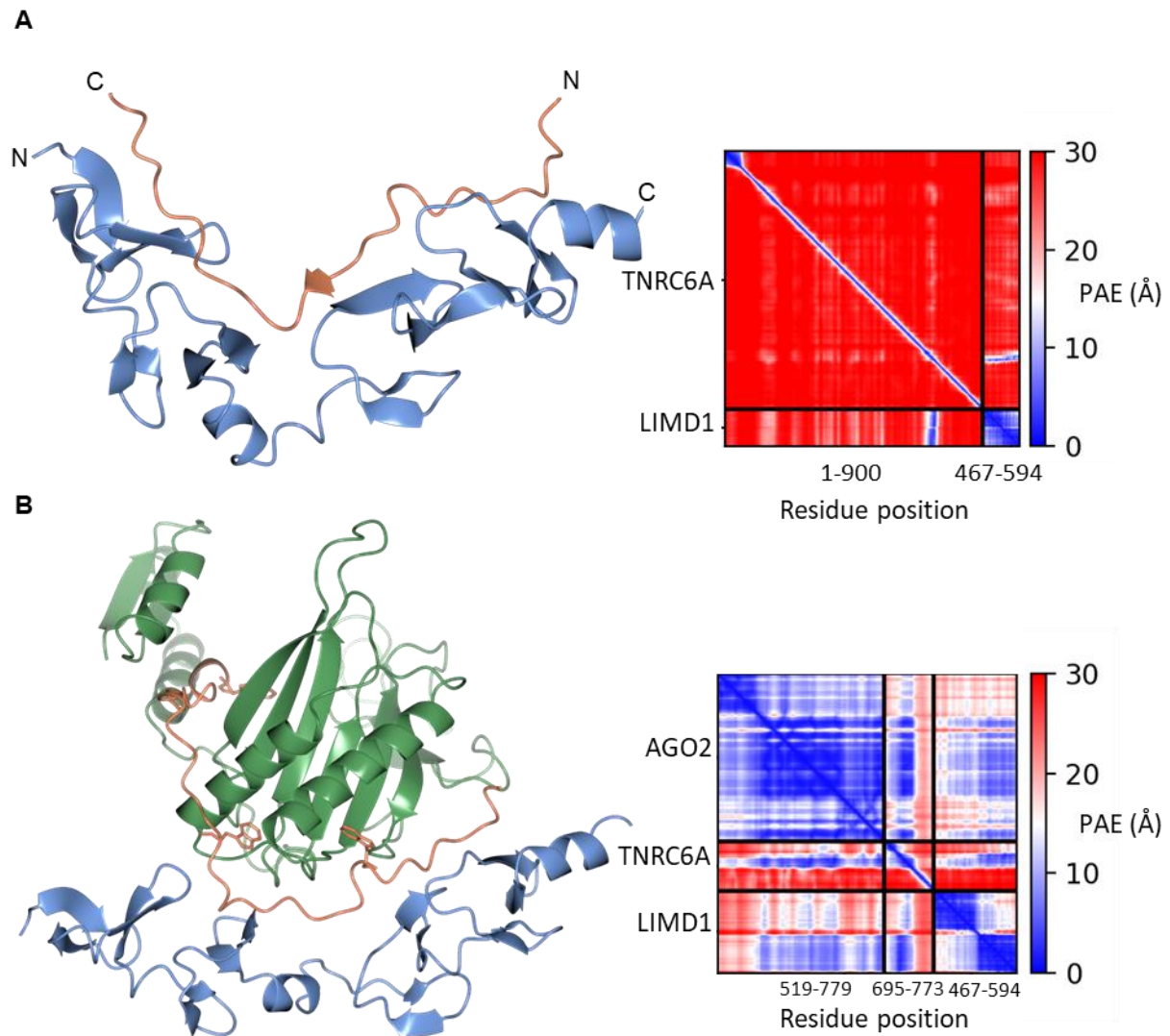


Figure 3.15. AlphaFold2 predictions of **A.** LIM12 with TNRC6A 1-900 showing LIM12 (blue) bound to TNRC6A 712-741 (orange) and the PAE plot showing high confidence of the relative positions of LIM12 and TNRC6A at this position. **B.** AGO2 519-779 with TNRC6A 695-773 and LIM12 showing AGO2 519-779 (green) and LIM12 (blue) bound to either side of TNRC6A 715-754.

The predicted structure of the LIMD1:TNRC6A:AGO2 interaction shows TNRC6A binding to AGO2 with tryptophan's inserted into the tryptophan binding pockets on

AGO2 [91, 194]. LIM12 is seen to bind on the opposite side of TNRC6A to AGO2 with the binding mode consistent with other LIM domains binding to disordered partners [40, 46, 47, 55, 199]. The structure shows a feasible mode by which the same region of TNRC6A could simultaneously bind to AGO2 and LIMD1. The PAE plot shows a high confidence for the interaction between LIM12 and TNRC6A 715-739, with a higher confidence for LIM2 than LIM1.

Inspired by work on the LMO2 and LMO4 LIM domains binding to Idb1, a fusion protein was designed with TNRC6A 715-739 connected to the C terminal of LIMD1 LIM12 by an 11 residue GGS linker. The fusion protein was expressed and purified (Figure 3.16B) but similarly to the tandem LIM domain proteins (Figure 3.16A), the fusion protein was unstable at higher concentrations.

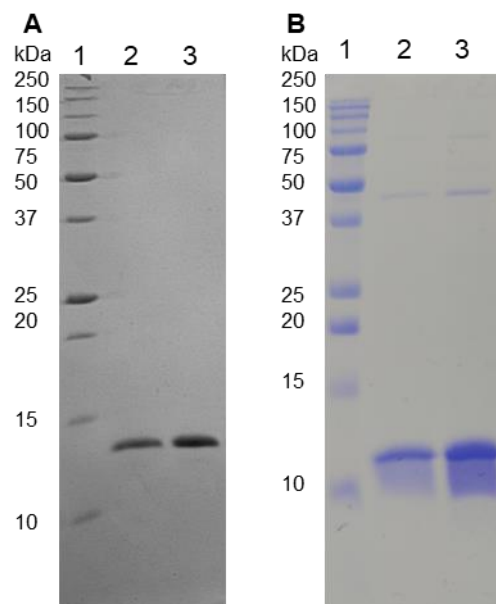


Figure 3.16. SDS-PAGE analysis of the purifications of **A**. LIM12 and **B**. a LIM12-TNRC6A<sub>715-739</sub> fusion protein. Lane 1: molecular mass standards, lane 2: protein post purification, lane 3: double the concentration of the sample in lane 2. Expected molecular weights = LIM12 = ~14 kDa, LIM12-TNRC6A<sub>715-739</sub> = ~17 kDa.

2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC spectra were collected for LIM12 and the LIM12-TNRC6A fusion protein (Figure 3.17). The 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC spectrum of the fusion protein was seen to have many peaks matching peaks in the LIM12 spectrum. This indicated that the LIM12 region was correctly folded in the fusion protein. There were also peaks in the LIM12 spectrum that had either shifted or disappeared in the LIM12-TNRC6A spectrum possibly due to an interaction with the TNRC6A region, although this could not be confirmed. The fusion protein spectrum also had some peaks of much higher intensity that likely correspond to the flexible linker or unbound regions of TNRC6A. The spectra were collected with different experimental parameters due to differences in the stabilities of the proteins. This makes more in depth comparison of the spectra difficult.

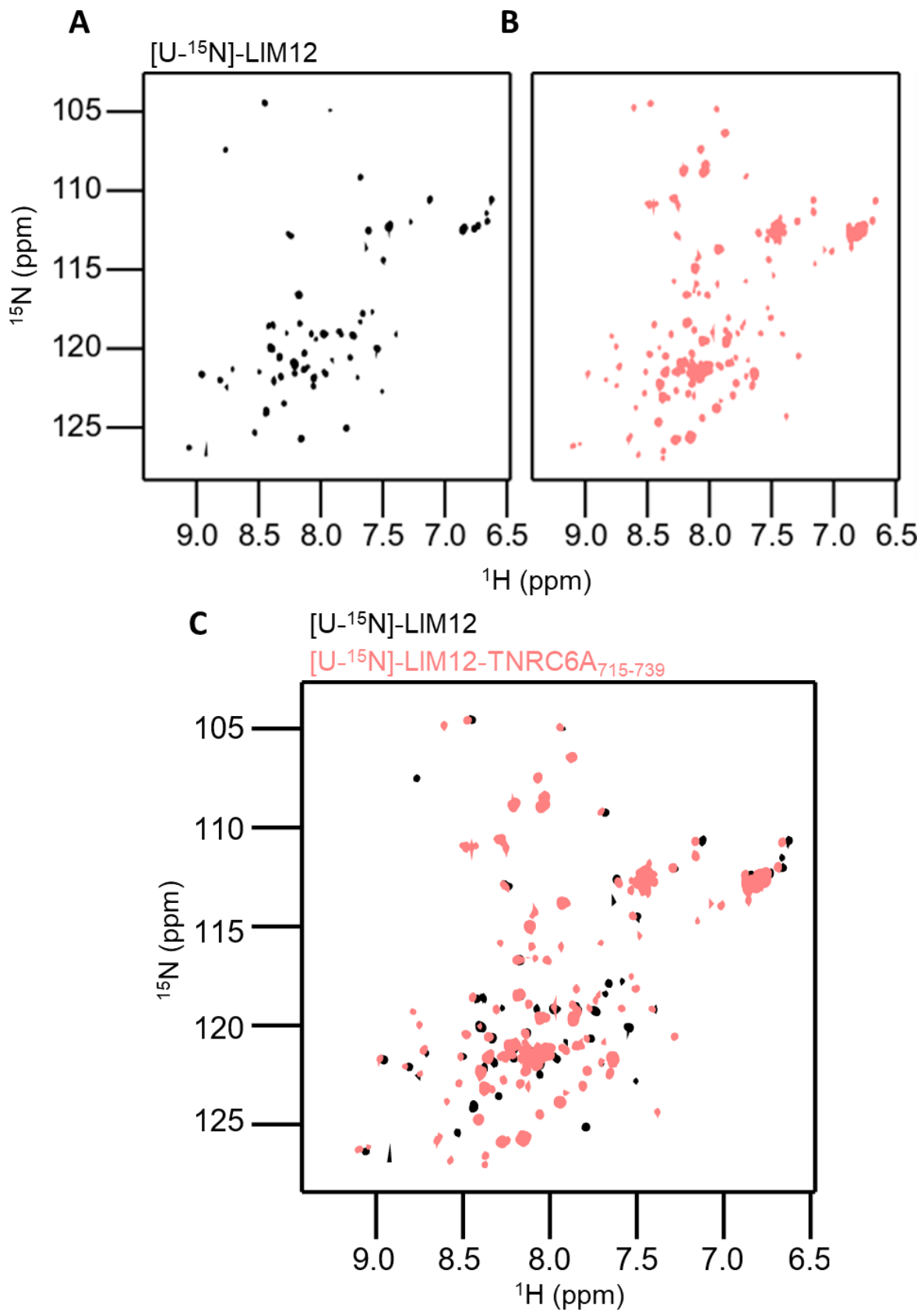


Figure 3.17. 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra of **A.** [U-<sup>15</sup>N]-LIM12 (black). **B.** [U-<sup>15</sup>N]-LIM12-TNRC6A<sub>715-739</sub> (Pink) and **C.** an overlay of the two spectra.

It was hoped that specific isotopic unlabelling of residues in the LIM12-TNRC6A fusion protein would allow for the observation of NOESY signals between LIM12 and the TNRC6A region (More detail on these types of experiments is presented in chapter 5). This would confirm the interaction between LIM12 and this region of TNRC6A. However, detection of NOEs was not possible at the concentrations of the fusion protein that could be achieved (30  $\mu$ M maximum concentration). It is possible that addition of AGO2 may have stabilised the complex by burying exposed tryptophan's in the TNRC6A region and allowed for higher concentrations to be achieved, however the increased size of the complex would likely have caused other issues.

#### 3.2.4. Production of Argonaute 2

To allow for study of the structural biology and biophysics of the interaction of LIMD1 with AGO2 it was necessary to produce milligram quantities of AGO2. AGO2 had previously been successfully expressed in insect cell expression systems and so this approach was used for this study. A mutant version of AGO2, as used in Schirle and MacRae *et al.* (2012), containing the mutations S387D, S824A, S828D, S831D, and S834A was used [95]. The S387D mutation should act as a phosphomimetic to facilitate the interaction with LIMD1 based on the similar property of AGO3 E390 [25]. The mutations were also thought to aid in crystallisation of AGO2. AGO2 was expressed in Sf9 cells using a bac-2-bac system [200]. Expression of AGO2 was optimised by testing different volumes of baculovirus and different expression times. Expression levels were assessed by western blot (Figure 3.18).

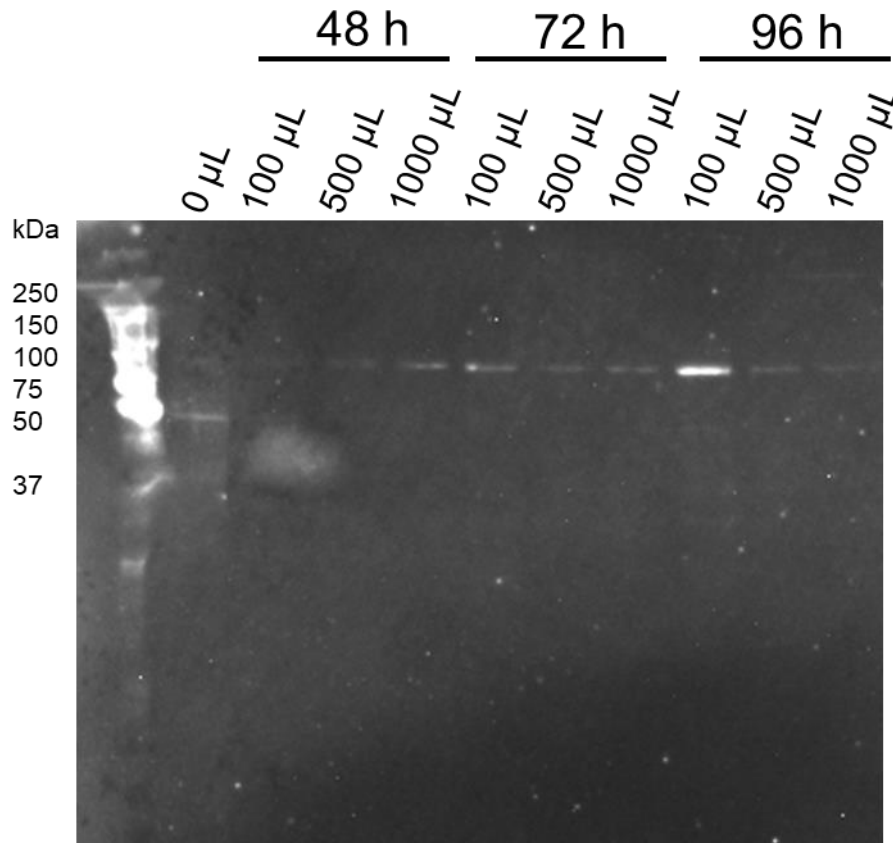


Figure 3.18. Western blot analysis of AGO2 production at different time points using different volumes of V1 viral stock. Probed with anti-His antibody (Sigma A7058) 1:120000 (v/v) with a 60 s exposure.

AGO2 was seen to be successfully expressing with the optimal conditions using 100 µL baculovirus for 96 h. With the expression conditions optimised, large scale expression could be carried out. Unfortunately, a shortage of insect cell media due to the COVID-19 pandemic made any further work on AGO2 impossible and so this work had to be abandoned.

### 3.3. Conclusion and Future Work

Identification of the LIMD1 binding region of TNRC6A was attempted using peptide arrays and pull-downs with neither method demonstrating clear evidence of an interaction. The pull-down method was validated by demonstrating an AGO2:TNRC6A interaction. The release of AlphaFold 2 allowed for prediction of the

binding site for LIMD1 on TNRC6A. Attempts to validate or disprove the prediction using a fusion protein, inspired by work on LIM domain only proteins, yielded a protein that could not be concentrated enough to confirm the interaction [54]. The lack of a demonstrated LIMD1:TNRC6A interaction may suggest that factors such as post-translational modification or accessory proteins may be needed for the interaction. The AlphaFold 2 prediction for the LIM12:TNRC6A:AGO2 complex suggest that TNRC6A binding to AGO2 could be required for LIMD1 binding [163, 164]. Unfortunately, AGO2 production was prevented by a lack of insect cell media availability and so this could not be tested. Additional *in vivo* investigation into the LIMD1:TNRC6A interaction may help to better inform future experiments. This could involve further reducing the potential LIMD1 binding site on TNRC6A by using more TNRC6A truncations and/or investigating potential post-translational modifications.

The lack of AGO2 production also prevented in-depth analysis of the AGO2:LIMD1 interaction. An AGO2 pull down using LIMD1 110-166 did not show any indication of an interaction. With expression of AGO2 having been established, production could be carried out as soon as media can be sourced. This would allow investigation of the LIMD1:AGO2 interface as well as testing to see if TNRC6A binding to AGO2 affects LIMD1 binding.

Despite binding not being demonstrated for LIMD1 with either TNRC6A or AGO2, the work presented in this chapter still informed the investigation of LIMD1 with its partner proteins in the hypoxic signalling pathway. The protocol to effectively produce LIM domains developed as part of the work in this chapter was essential to study of VHL binding to LIMD1. Additionally, the experimental methods used were informed by the limitations identified in this chapter. The AlphaFold 2 predictions for the interactions of LIMD1 with TNRC6A could inform future investigations into this system. Mutation of the proposed LIMD1 binding region on TNRC6A *in vivo* could be used to verify the AlphaFold 2 prediction.

## 4. Results Chapter 2: Investigation of the LIMD1 Protein Interactions that Mediate HIF-1 $\alpha$ Degradation

### 4.1. Introduction

In order to provide greater insight into the role of the LIMD1 scaffold protein in the hypoxic signalling pathway, the structural and biophysical determinants of partner protein recognition by LIMD1 will be investigated. As production of full-length LIMD1 was unsuccessful, and to limit the size of the complexes to allow for NMR analysis, the interactions of LIMD1 with PHD2 and VHL will be studied independently. The second LIM domain of LIMD1 had previously been shown to be responsible for the interaction with VHL [24]. As the mechanism of recognition of folded partner proteins by LIM domains remains poorly understood, investigation of the LIMD1 VHL interaction may provide additional insights into LIM domain biochemistry. Study of the interactions between LIM2 of LIMD1 and VHL will rely on work carried out on the production of LIM domains carried out in chapter 2. In this chapter, previously published protocols are used for the production of a recombinant VHL:Elongin B (EloB):Elongin C (EloC) complex [159]. It has been previously shown that VHL cannot be recombinantly produced independently and that it requires co-expression with EloB and EloC [159]. EloB and EloC have been shown to stabilise VHL through the formation of a heterotrimer [159]. Recapitulation of the interaction between LIMD1 LIM2 and VHL *in vitro* will be attempted using co-elution assays and solution NMR spectroscopy. The 186-239 region of LIMD1 had previously been identified as essential for the interaction with PHD2 [24]. A LIMD1 $\Delta$ 186-260 mutant was found to be unable to bind to PHD2 whilst a LIMD1 $\Delta$ 239-260 mutant retained the ability to bind to PHD2 [24]. This suggests that PHD2 binds to the disordered pre-LIM region of LIMD1 which raises the question of how specific recognition could be achieved in this region despite little sequence similarity in the pre-LIMs of the LAW proteins (Figure 1.7). In order to investigate the interaction between LIMD1 and PHD2, PHD2 and various regions of the pre-LIM region of LIMD1 will be recombinantly produced in *E. coli*. Recapitulation of the interaction between LIMD1 and PHD2 *in vitro* will then be attempted using co-elution, pull down assays and solution NMR



spectroscopy. Work will then utilise solution NMR spectroscopy to further characterise the interaction between LIMD1 and PHD2 on an atomic level. The results of this work could be used to inform the development of new LIMD1 targeted drugs/therapies to treat cancers and ischemic diseases.

## 4.2. Results and Discussion

### 4.2.1. LIMD1 Binds to VHL through the Second LIM Domain of LIMD1

LIMD1 is thought to interact with VHL through its second LIM domain (LIM2), due to the loss of binding seen with LIMD1 $\Delta$ LIM2 [24]. Because of the difficulty in producing milligram quantities of LIMD1 and the full VHL-E3-Ubiquitin ligase complex needed for structural analysis, work first focussed on producing a minimal LIM2:VHL complex. LIM2 had previously been produced to investigate the interactions with TNRC6A in chapter 3. VHL was co-expressed as a complex with elongin B and elongin C as has been used for previous structural studies of VHL complexes [159, 201, 202]. The VHL:EloB:EloC complex was expressed and purified and purity was analysed by SDS-PAGE (Figure 4.1).

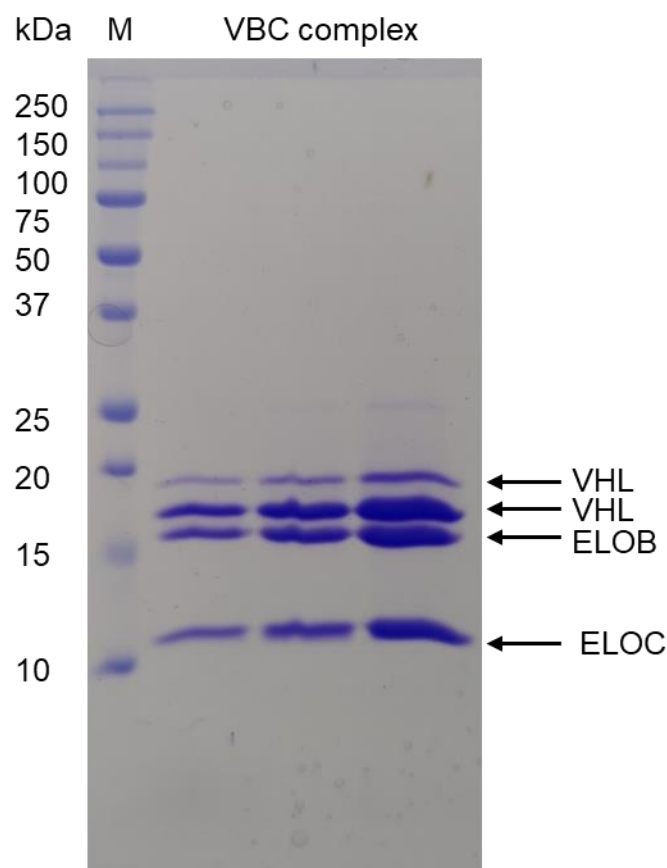


Figure 4.1. SDS-PAGE analysis of the VHL:EloB:EloC complex post purification. The protein was run at three concentrations and the protein corresponding to each band is indicated. Four bands were seen in place of the expected three however the top two bands were confirmed by mass spectrometry to contain VHL. Calculated molecular masses are VHL = ~18.7 kDa, ElonginB = ~13 kDa, ElonginC= ~10.8 kDa

The VHL:EloB:EloC complex was seen to run as four bands on SDS-PAGE rather than the expected three bands. Mass spectrometry (MS) was used to (a) identify which protein gave rise to the two highest bands and (b) confirm the intact mass of proteins responsible for these bands. The bands were excised, the proteins were digested with trypsin and analysed using MALDI-MS/MS. 5 VHL peptides were identified in both samples covering 40% of the VHL sequence (Appendix figure 7.28). This allowed us to determine that both bands corresponded to VHL. Although splice variants of VHL with different lengths have been described in the literature, the expression vectors used here for recombinant expression do not allow for alternative splicing [159, 203]. Intact mass spectrometry, without prior trypsin digestion, was

used to further confirm the species present in the sample. This was to investigate the possibility that VHL was truncated or longer than expected. This experiment was able to identify that full length VHL, ElonginB and ElonginC were present in the sample (Table 4.1), (Figure 4.2).

Table 4.1. Mass spectrometry results for the VHL:EloB:EloC sample. Expected  $m/z$  = Isotopically averaged molecular weight calculated from the protein sequence,  $m/z$  = mass to charge ratio, Res. = resolution, I = intensity, I% = relative intensity

Protein	Expected $m/z$	$m/z$	Res.	S/N	I	I %
Elongin C	10963.61	10962.9592	$1.82 \times 10^3$	$4.69 \times 10^6$	$7.23 \times 10^7$	100
Elongin B	13132.82	13132.3	$2.30 \times 10^3$	$3.19 \times 10^5$	$4.92 \times 10^6$	6.8
Methylated VHL	18676.32	18675.6	$3.08 \times 10^3$	$1.01 \times 10^4$	$1.5 \times 10^7$	28.5

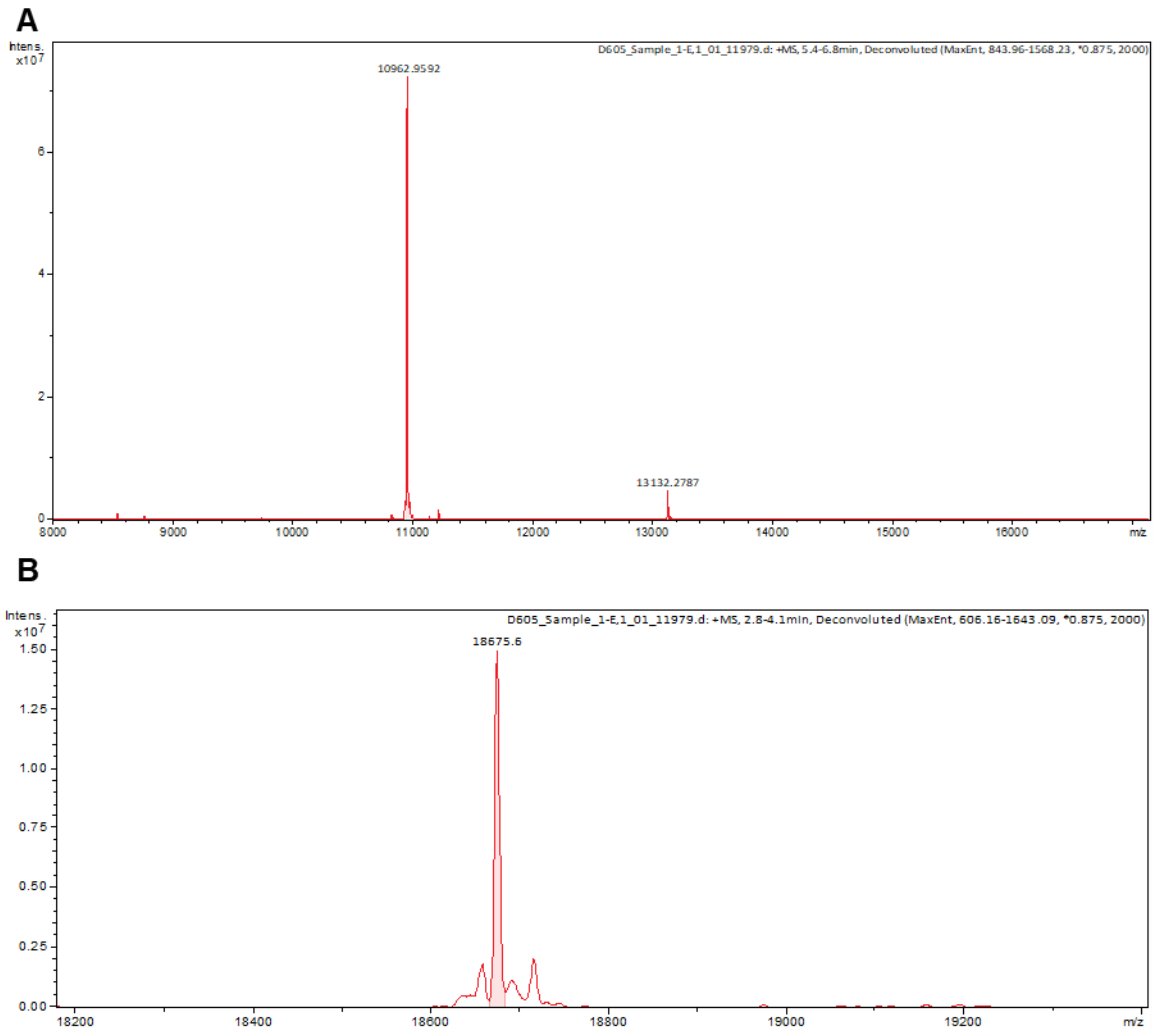


Figure 4.2. Deconvoluted MS spectra at **A.** 5.4-6.8 min elution from a divinyl benzene column showing peaks for ElonginB and ElonginC. **B.** 2.8-4.1 min elution from a divinyl benzene column showing a peak corresponding to VHL plus a methylation event.

All proteins were seen to have the expected mass apart from VHL where a methylation event was observed. The methylation event is not the cause of the two bands seen of SDS-PAGE as this mass difference would not lead to such a large shift on SDS-PAGE. The two bands are most likely the result of VHL not being fully denatured on the SDS-PAGE leading to multiple conformations of VHL being separated. The methylation of VHL is likely to be due to the action of *E. coli* proteins and so is unlikely to be biologically relevant although may impact the binding studies carried out in this study. The mass spectrometry results provided confidence in the

VHL:EloB:EloC sample quality, ensuring no truncations had occurred, allowing for the interaction with LIMD1 to be studied. In order to determine if an interaction could be formed *in vitro* between VHL and LIM2 it was first seen if the proteins co-eluted from size exclusion chromatography (Figure 4.3).

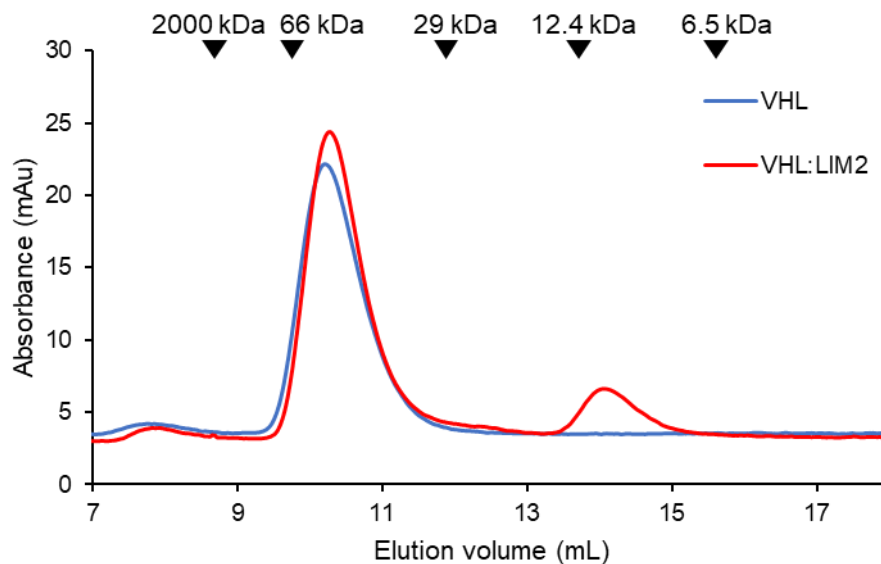


Figure 4.3. Chromatograms from size exclusion chromatography of VHL with (red) and without (Blue) LIM2 followed by absorbance at 280 nm. Peaks can be seen for the VHL:EloB:EloC complex (42 kDa) at ~10.5 mL and for LIM2 (9 kDa) at ~14 mL. there is no peak corresponding to a VHL:LIM2 complex. The elution volumes of molecular weight standards are indicated above the chromatogram. The VHL complex has a calculated MW of ~42.6 kDa and LIM2 has a calculated MW of ~7.1 kDa

VHL and LIM2 were seen to not co-elute from size exclusion chromatography, likely due to a low affinity of the interaction. Co-elution is a poor method for studying low affinity interactions as the presence of an equilibrium between bound and unbound states will result in unbound protein being separated. The technique therefore relies on the observation of long lived complexes such as the VHL:EloB:EloC complex. To better capture a potential interaction between VHL and LIM2, solution NMR

spectroscopy was used. [U-<sup>15</sup>N]-LIM2 was produced, and 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra were recorded in the absence and presence of unlabelled VHL (Figure 4.4).

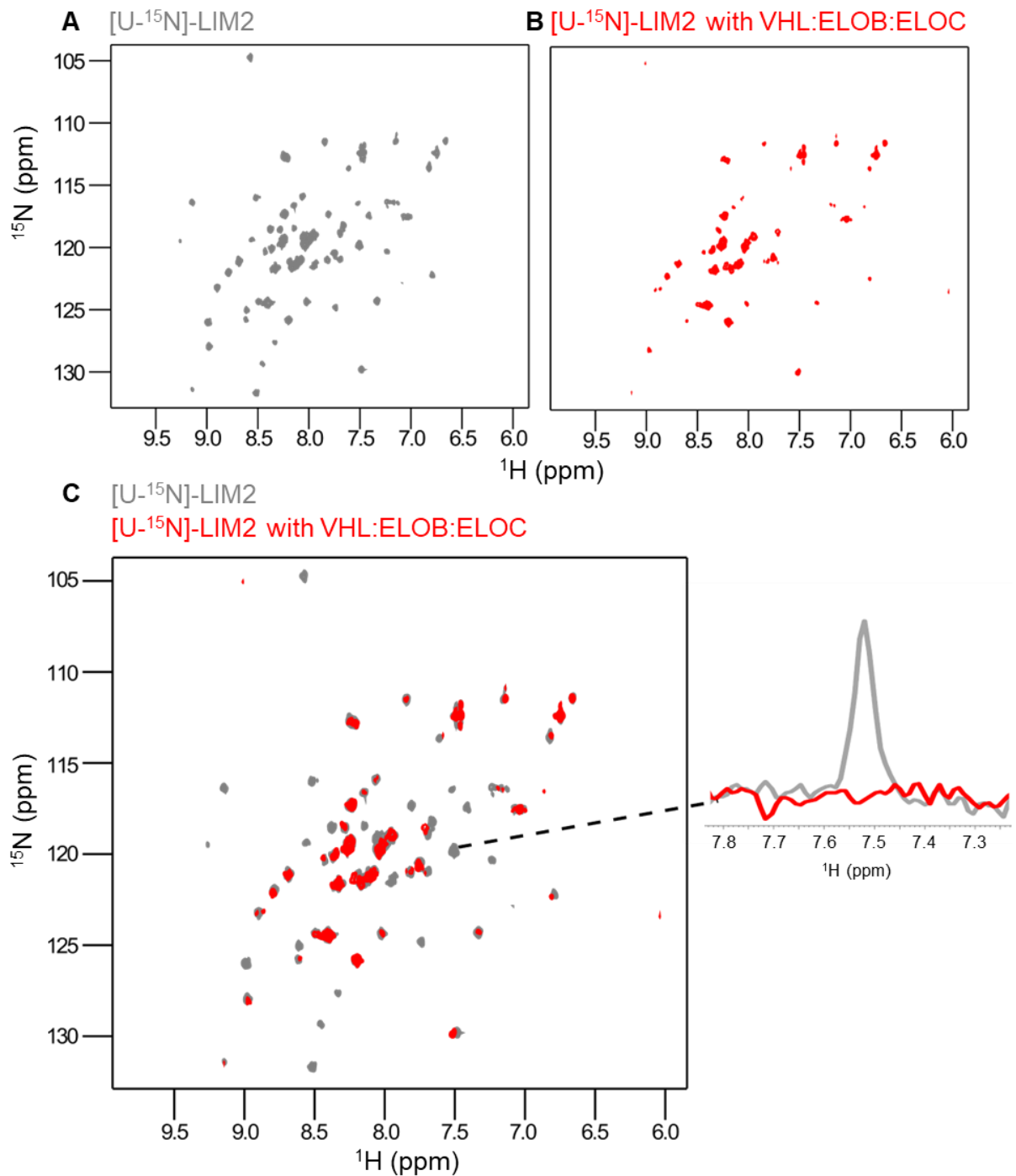


Figure 4.4. 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra of LIM2 (100 μM) **A**. without VHL:EloB:EloC (grey) **B**. with (red) VHL:EloB:EloC (250 μM) **C**. an overlay of the spectra in **A** and **B** with 1D <sup>1</sup>H traces are shown for one peak to demonstrate the loss in signal on addition of VHL:EloB:EloC.. A general loss in signal intensity and many peaks disappearing can be seen on addition of VHL:Elob:EloC.

The widespread reduction in peak intensity seen in the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC spectrum of LIM2 on addition of VHL:EloB:EloC is consistent with an interaction and the formation of a larger complex, which would be too large to observe by solution NMR spectroscopy without employing isotopic labelling approaches that have been developed for studying higher molecular weight systems. In order to extract more information on the nature of the interaction, backbone assignment of LIM2 was attempted. Triple resonance backbone assignment experiments were collected, however the poor stability of the protein meant that higher concentrations of LIM2 could not be achieved and the length of experiments that could be collected was limited. The low concentration used (150  $\mu\text{M}$ ) and the shorter experiment times meant that it was not possible to collect data of sufficient quality to allow for backbone resonance assignment of LIM2. In addition to this, a higher than expected number of peaks in the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC spectrum suggests that LIM2 may exist in multiple conformations in solution which would further complicate analysis. Optimisation of the conditions for assignment experiments was attempted by varying pH (pH 6-8), salt (50-300 mM NaCl), buffer (tris, sodium phosphate, HEPES, bis-tris), glycerol concentrations (5-20%) and temperature (10-30°C) but this did not alleviate the problems. The optimum conditions were 20 mM bis-tris pH 6.5, 150 mM NaCl, 1 mM TCEP, 5% glycerol at 20°C. Due to the combination of poor protein solubility and the higher than anticipated number of ( $^1\text{H}$ ,  $^{15}\text{N}$ ) peaks, backbone resonance assignment of LIM2 was not achieved. In addition, deuteration of the VHL complex was attempted in order to improve the relaxation properties of the large (42.5 kDa) VHL:EloB:EloC complex and allow observation of this by solution NMR spectroscopy. Adaptation of the *E. coli* expressing VHL:EloB:EloC to deuterated M9 media was not successful, possibly due to the added stress of dual anti-biotic selection. Extensive screening of 1728 crystallisation conditions for the LIM2:VHL:EloB:EloC complex were tested using commercial kits but no crystals were produced.

Additional work may benefit from expressing the full LIM domain region in insect cell systems which would allow for investigation of the complex by cryo-EM due to the size increase. The interaction with VHL may also benefit from the presence of the additional LIM domains. It would also be possible to extract a binding affinity for the



interaction by titrating [ $U\text{-}^{15}\text{N}$ ]-LIM2 with VHL and following the change in signal intensity. The interaction between LIM2 and VHL was not investigated by pull down assays due to the similarity in size between the proteins, although this could be done through the use of a VHL specific antibody.

#### 4.2.2. Identification of the Binding Site of PHD2 on LIMD1

PHD2 had previously been shown to bind within the disordered pre-LIM region of LIMD1, with LIMD1 $\Delta$ 1-467 and LIMD1 $\Delta$ 186-260 mutants unable to bind PHD2 *in vivo* [24]. To further investigate this interaction, a fragment of the pre-LIM region of LIMD1 spanning residues 65-260 (LIMD1 65-260) was produced along with full length PHD2 (Figure 4.5).

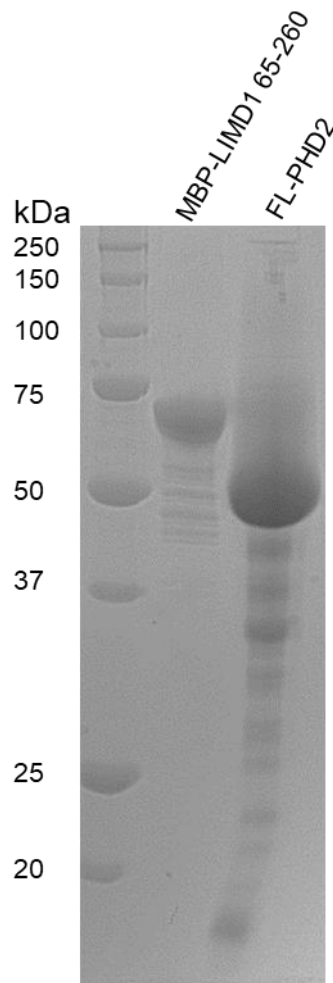


Figure 4.5. SDS-PAGE analysis for the purification of MBP tagged LIMD1 65-260 and untagged full length PHD2. MBP-LIMD1 65-260 = ~66 kDa, FL-PHD2 = ~46 kDa

A pull-down assay developed in chapter 3 was first attempted to demonstrate the formation of an interaction between these proteins in vitro (Figure 4.6). His<sub>6</sub>-MBP tagged LIM2 was immobilised on Ni-NTA magnetic beads and incubated with untagged PHD2. Unbound protein was washed off and bound protein was eluted with imidazole.

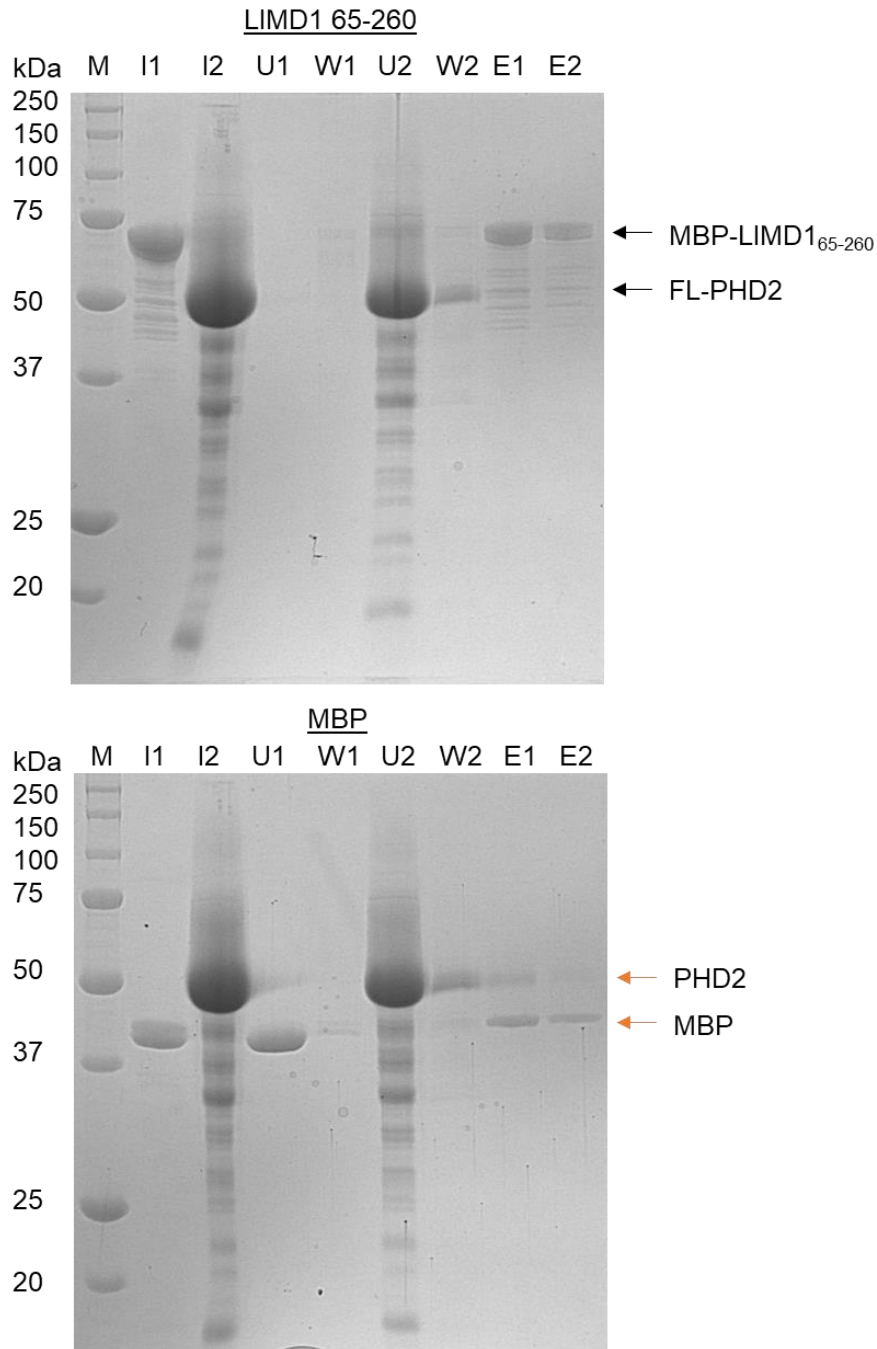


Figure 4.6 PHD2 pull downs using His<sub>6</sub>-MBP-tagged LIMD1 65-260 and just His<sub>6</sub>-MBP. I1 contains the bait protein input, LIMD1 65-260 or MBP. I2 contains the prey protein input, PHD2. U1 contains the unbound bait protein, W1 contains the protein from the first wash step, U2 contains the unbound prey protein, W2 contains the unbound protein from the second wash step, E1 and E2 contain the first and second elution fractions. The results show in interaction of PHD2 with MBP but not LIMD1 65-260. FL-PHD2 = ~46 kDa, MBP-LIMD1 = ~66 kDa, MBP = ~43 kDa

The presence of PHD2 in the second wash step means that no meaningful conclusions could be drawn from this experiment. Adding additional wash steps would likely remove the residual PHD2 in the wash step, however the lack of PHD2 in the LIMD1 65-260 elution fractions suggests that no interaction would be seen. As an alternative method of demonstrating the interaction, co-elution from SEC was attempted with FL-PHD2 and LIMD1 65-260 with the MBP tag removed (Figure 4.7).

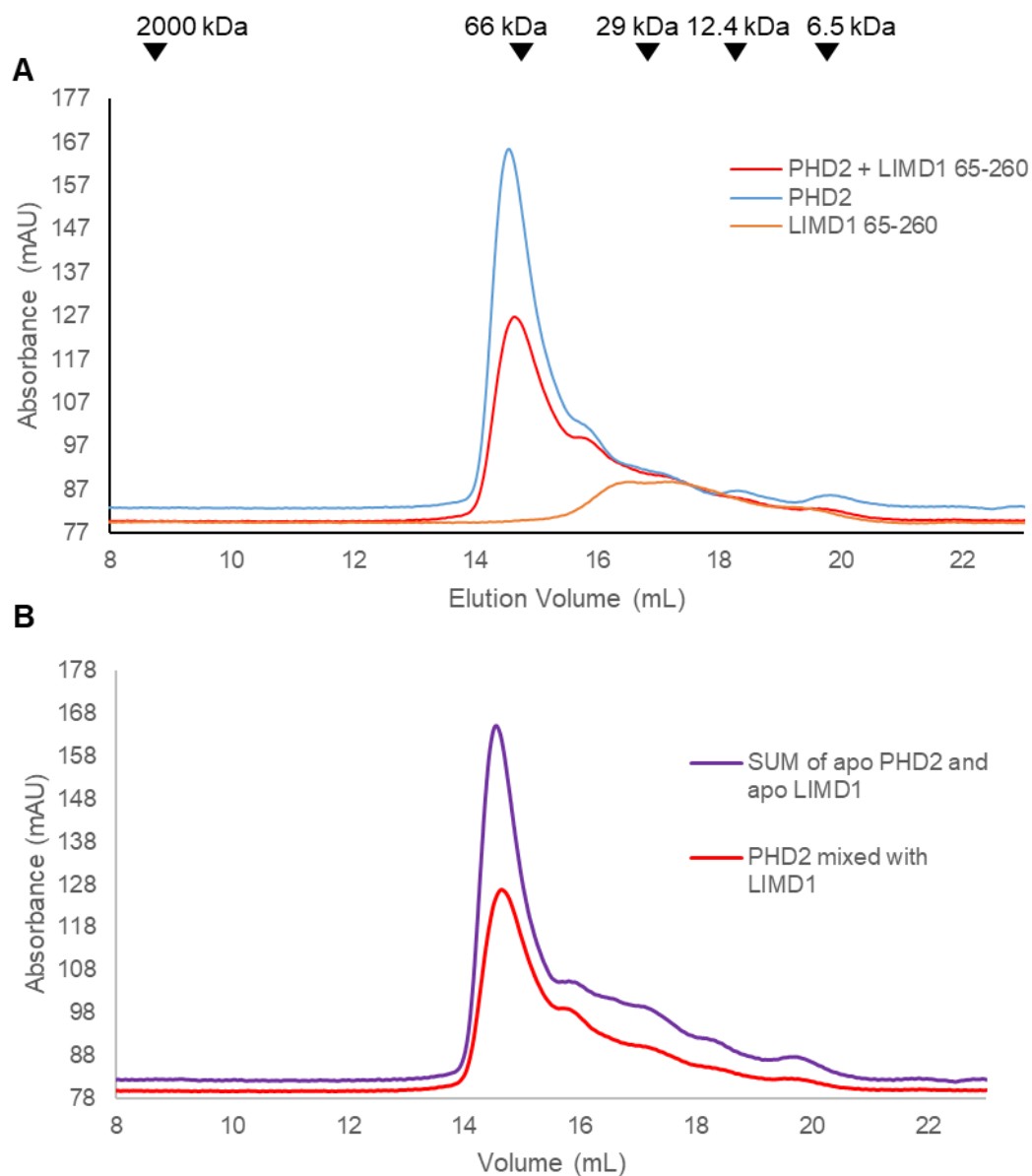


Figure 4.7. **A.** SEC chromatograms of PHD2 (blue), LIMD1 65-260 (orange) and a mixture of the two proteins (red) followed by absorbance at 280 nm. The elution volume of molecular weight standards are indicated above the chromatogram. **B.** The chromatogram for the PHD2, LIMD1 65-260 mixture in red and the sum of the individual PHD2 and LIMD1 65-260 chromatograms in purple. The sum of the individual chromatograms can be seen to closely follow the shape of the chromatogram of the mixed sample.

Similarly to the VHL:LIM2 results, the SEC results showed no indication of an interaction being formed between PHD2 and LIMD1 65-260. It was therefore decided to attempt to investigate the interaction by NMR which had worked for the VHL:LIM2 interaction.

FL-PHD2 was seen to elute from SEC earlier than expected and with a large tail to the peak. This suggested that the FL-PHD2 protein may be misfolded and possibly forming oligomers. Due to the reported ability of LIMD1 to bind to PHD1,2 and 3 it was assumed that these proteins must share a common binding site for LIMD1 [24]. The only region that appears in all three of these proteins is the core catalytic domain, meaning this is likely to be the site of interaction with LIMD1. Based on this assumption future work focussed on using only the catalytic domain of PHD2 (181-426). The smaller PHD2 construct has the added advantage of being easier to produce and, having been previously well studied, is known to crystallise readily [182]. In addition to demonstrating the formation of an interaction, it was also hoped to narrow down the binding region of PHD2 on LIMD1. For this reason, three constructs spanning overlapping regions of the pre-LIM region of LIMD1 were initially tested: 65-260, 110-166 and 160-300. LIMD1 65-260, 110-166 and 160-300 proteins as well as PHD2 181-426 were all successfully produced and purified (Figure 4.8A, C-D). Additionally, LIMD1 1-80, a region predicted by AlphaFold2 and disorder predictions to be ordered but previously annotated as a structured domain, was produced in order to verify the presence of any structured elements in this region (Figure 4.8B) [67, 204].

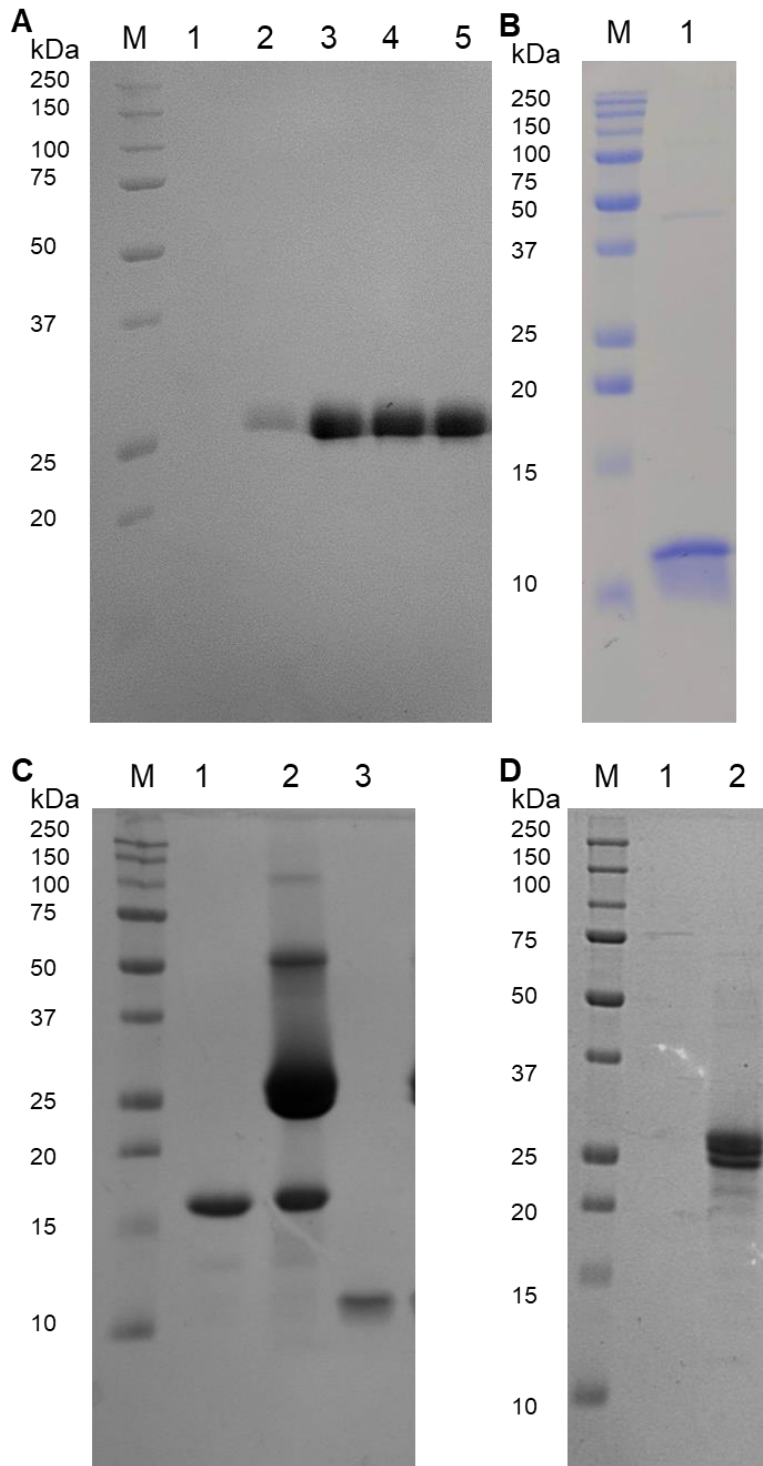


Figure 4.8. SDS-PAGE analysis of the purifications of **A.** unlabelled PHD2 181-426, with lanes 2-5 containing elution fractions from SEC. **B.** Pure unlabelled LIMD1 1-80. **C.** Lane 1: pure [U-<sup>15</sup>N] LIMD1 160-100. lane 3: pure [U-<sup>15</sup>N] LIMD1 110-166. **D.** Lane 2: Pure [U-<sup>15</sup>N] LIMD1 65-260. Estimated molecular weights based on amino acid sequence are: PHD2 181-426 = 27 kDa, LIMD1 160-300 = 14.3 kDa, LIMD1 110-166 = 5.85 kDa, LIMD1 65-260 = 27.6 kDa, LIMD1 1-80 = ~9.13 kDa

2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC or HMQC spectra were collected for the [ $\text{U-}^{15}\text{N}$ ] pre-LIM fragments (Figure 4.9). The spectra allowed for experimental characterisation of the pre-LIM as largely disordered due to the narrow  $^1\text{H}$  chemical shift distribution seen for all regions of the LIMD1 pre-LIM that were tested. The narrow  $^1\text{H}$  chemical shift distribution show that there is little change in the chemical environment across the protein and so the protein is likely to be disordered and lacking a persistent 3D fold. Verification that LIMD1 1-80 is disordered contradicts previous annotation of this region as a LEM (LAP2, emerin, MAN1) domain [67].

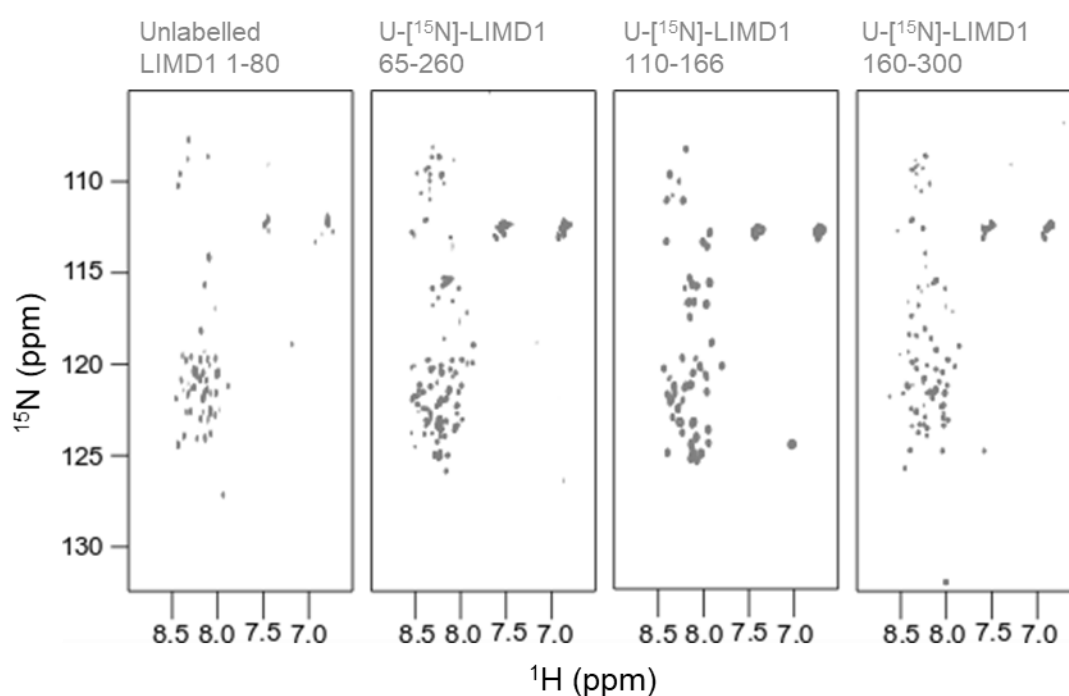


Figure 4.9. 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC or HMQC spectra for various regions of the pre-LIM of LIMD1. The narrow chemical shift distribution and sharp peaks suggest that all of these regions lack a persistent 3D fold.

In order to investigate PHD2 binding to LIMD1 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectra were also collected for [ $\text{U-}^{15}\text{N}$ ] LIMD1 65-260, 110-166 and 160-300 with PHD2 181-426 added (Figure 4.10).



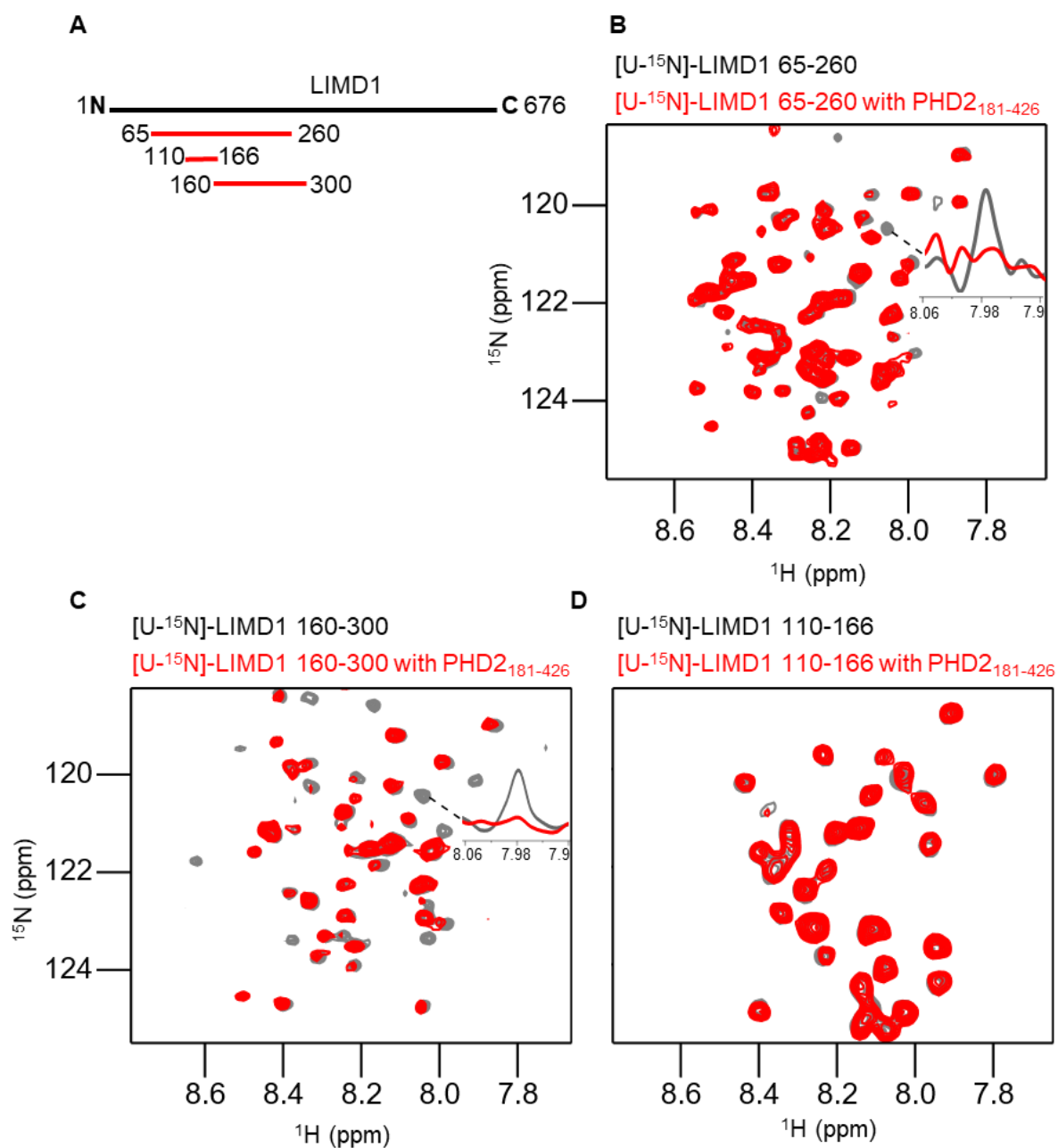


Figure 4.10. **A**. Schematic representation showing how the pre-LIM regions used relate to the sequence of full length LIMD1 and to each other **B-D**. 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectra of 100 μM [U-<sup>15</sup>N]-LIMD1 65-260 (**B**), [U-<sup>15</sup>N]-LIMD1 160-300 (**C**) and [U-<sup>15</sup>N]-LIMD1 110-166 (**D**) with (red) and without (grey) 400 μM PHD2 181-426. 1D <sup>1</sup>H traces are shown for one peak in the 65-260 and the 160-300 spectra to demonstrate the loss in signal on addition of PHD2<sub>181-426</sub>. Full spectra are shown in appendix figures 7.3-7.5

Peaks were seen to disappear in the spectra of LIMD1 65-260 (Figure 4.10B) and LIMD1 160-300 (Figure 4.10C) in the presence of an excess of PHD2 181-426 but not in the spectra of LIMD1 110-166 (Figure 4.10D). This suggested that the binding site for PHD2 on LIMD1 was located in the overlapping region between 65-260 and 160-300 but not in the 110-166 region (Figure 4.10A). Peaks disappearing in the presence of PHD2 is consistent with the formation of a LIMD1:PHD2 complex. These results narrowed the potential binding site of PHD2 to the region of 166-260 on LIMD1. This was in agreement with the LIMD1 186-239 region identified by in vivo studies [24]. To further test binding of PHD2 to the overlapping region of LIMD1 previously identified, a construct was made spanning LIMD1 166-260 (Figure 4.11). A smaller construct spanning the LIMD1 186-239 region previously implicated in PHD2 binding was not used in order to avoid truncating the binding site [24].

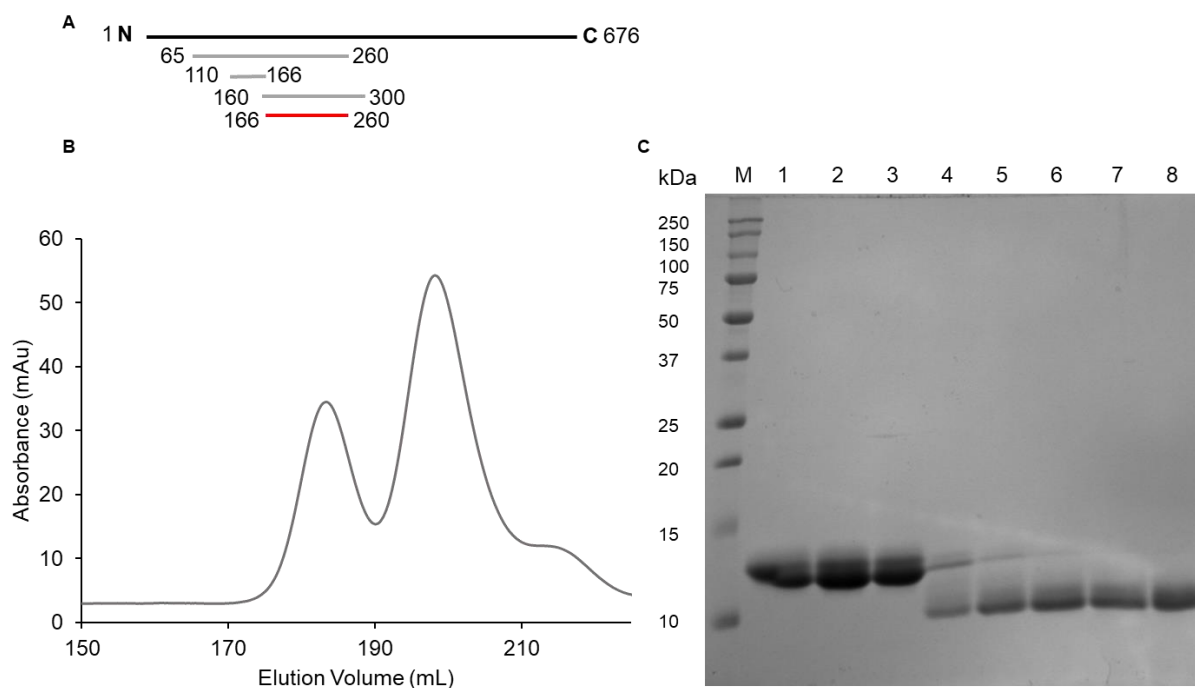


Figure 4.11. **A.** Schematic representation showing how the LIMD1 166-260 related to the previously used pre-LIM regions and to full length LIMD1. **B.** Section of the chromatogram for the SEC purification of [U-<sup>15</sup>N] LIMD1 166-260 showing absorbance at 280 nm. **C.** SDS-PAGE analysis of the purification of tag less [U-<sup>15</sup>N] LIMD1 166-260. Lanes contain fractions from the SEC purification of LIMD1 166-260 shown in **B**. Bands corresponding to LIMD1 166-260 were seen in lanes 4-8. Calculated molecular weight of LIMD1 166-260 = ~9.6 kDa.. Fractions from lanes 6-8 in **B** were pooled and used for future experiments.

As anticipated, comparison of the 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectrum of U-[<sup>15</sup>N]-LIMD1 166-260 with the 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectra of LIMD1 65-260 and 160-300 shows a mixture of matching and non-matching peaks (Figure 4.12A). The peaks in both the LIMD1 65-260 and 160-300 spectra most affected by PHD2 binding were all seen to contain matching or nearby peaks in the LIMD1 166-260 spectrum (Figure 4.12B). The spectra do not match perfectly potentially due to differences in the temperature and effects of truncating the proteins on chemical shift. The similarity in the regions of the spectra of LIMD1 166-260, 65-260 and 160-300 most effected by PHD2

supports the hypothesis that PHD2 is binding to the same region in these proteins, however the spectra do not overlay perfectly enough to confirm this.

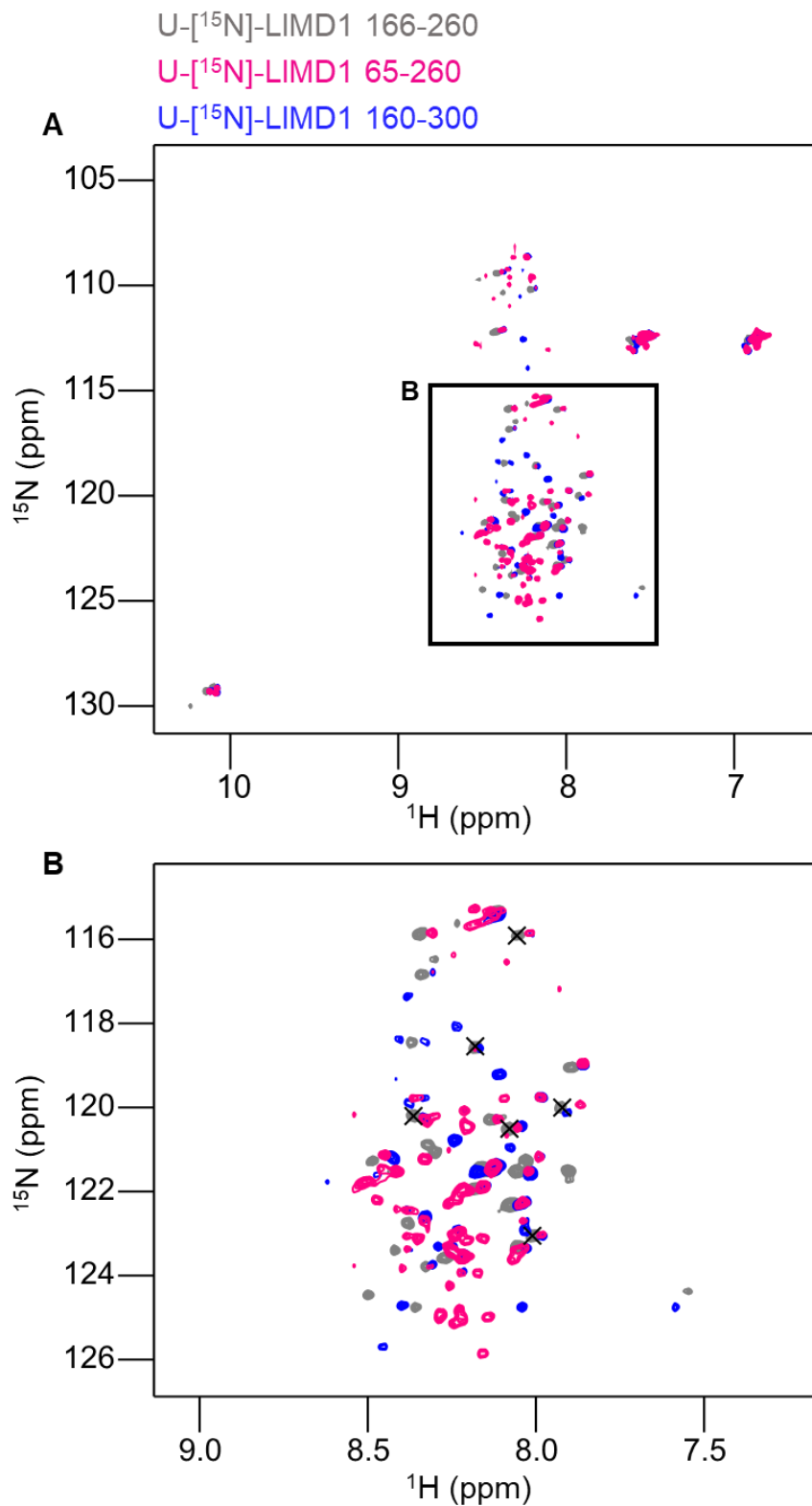


Figure 4.12. **A.** Overlay of the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectra for LIMD1 166-260 (grey), LIMD1 65-260 (pink), LIMD1 160-300 (blue). **B.** A zoomed in view of the region indicated in **A**. Peaks corresponding to the peaks most affected in the LIMD1 65-260 and 160-300 spectra on addition of PHD2 are indicated with a cross.

The ability of [U-<sup>15</sup>N]-LIMD1 166-260 to bind to PHD2 was tested by collecting 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra with and without unlabelled PHD2 181-426. (Figure 4.13)

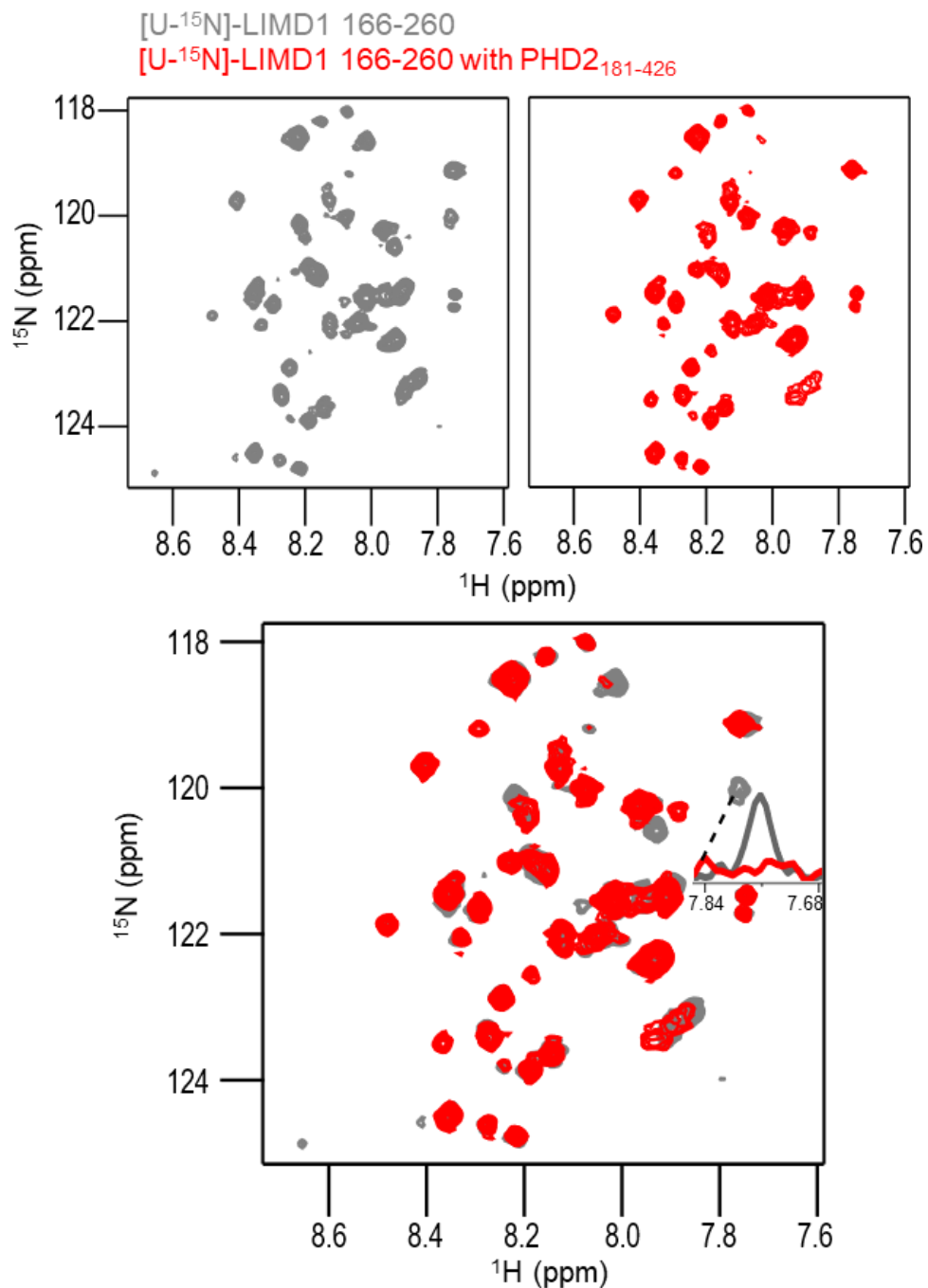


Figure 4.13. 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra for 100 μM [U-<sup>15</sup>N]-LIMD1 166-260 with (red) and without (grey) 400 μM PHD2 181-426. 1D <sup>1</sup>H traces are shown for one peak in the to demonstrate the loss in signal on addition of PHD2<sub>181-426</sub>. The full spectra are shown in appendix figure 7.6

The reduction of signal intensity seen for some peaks in the LIMD1 166-260 spectrum on the addition of PHD2 demonstrated that the LIMD1 166-260 construct retained the ability to bind to PHD2 181-426. As PHD2 binds to peptide substrates as part of its catalytic activity, it was important to verify that PHD2 was not binding LIMD1 non-specifically *via* its substrate binding pocket. To evaluate this, spectra of [U-<sup>15</sup>N] LIMD1 166-260 were collected in the presence of both PHD2 and an excess of HIF-1 $\alpha$  CODD peptide (Figure 4.14). As HIF-1 $\alpha$  CODD is one of the natural substrates of PHD2, adding an excess of the peptide to the sample would likely displace any LIMD1 bound in the active site.

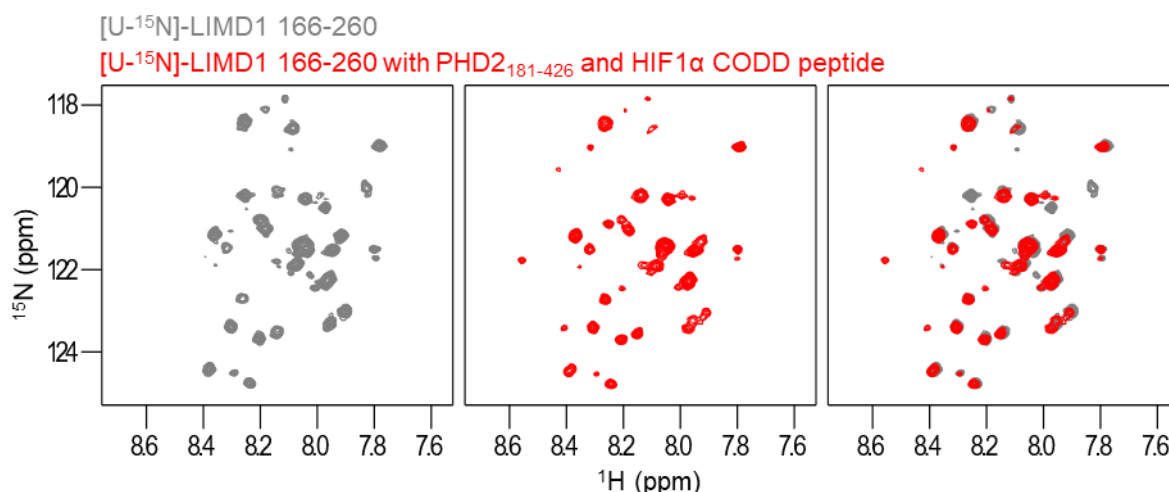


Figure 4.14. 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra for 100  $\mu$ M [U-<sup>15</sup>N]-LIMD1 166-260 with (red) and without (grey) 250  $\mu$ M PHD2 181-426 and 250  $\mu$ M HIF-1 $\alpha$  CODD peptide. Signals can be seen to disappear on addition of PHD2 and the HIF-1 $\alpha$  peptide indicating an interaction is formed.

The interaction between LIMD1 166-260 and PHD2 did not appear to be inhibited by the addition of the HIF-1 $\alpha$  CODD peptide. This suggested that LIMD1 166-260 binding does not compete with the CODD peptide which in turn implies that LIMD1 166-260 does not bind to the substrate binding pocket of PHD2. To further investigate the interaction between LIMD1 and PHD2, co-elution experiments were

carried out with the smaller PHD2 181-426 and LIMD1 166-260 constructs (Figure 4.15).

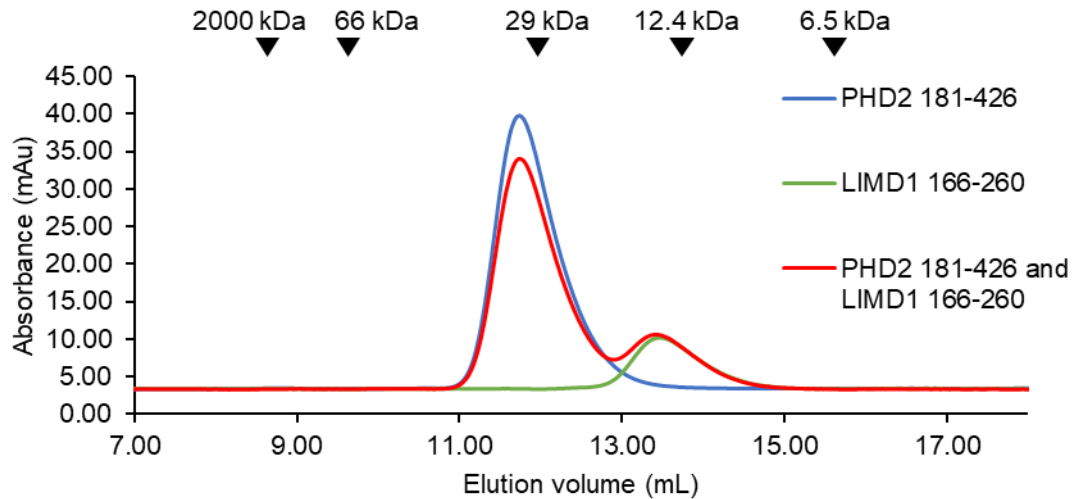


Figure 4.15. SEC chromatograms of PHD2 (blue), LIMD1 166-260 (green) and a mixture of the two proteins (red) followed by absorbance at 280 nm. The elution volumes of molecular weight standards are indicated above the chromatogram. PHD2 181-426 = ~ 27 kDa, LIMD1 166-260 = ~9.6 kDa.

Similarly to previous results, the co-elution experiments did not provide evidence of the interaction between PHD2 181-426 and LIMD1 166-260. The failure to see a complex of PHD2 and LIMD1 by SEC could suggest these constructs interact weakly and as such exist in an equilibrium between bound and unbound states. The separation of PHD2 and LIMD1 during SEC would serve to push the equilibria towards the unbound state. Co-elution would only be expected if the two components interacted with high enough affinity to form a stable complex during the SEC run.

The effect of LIMD1 to increase the efficiency of HIF-1 $\alpha$  degradation could be caused by a scaffolding effect that increases the local concentration of VHL and PHD2 and thus improves the efficiency of the signalling pathway, or LIMD1 could



directly modify the activity of its partner proteins. It is also possible that LIMD1 acts *via* a combination of these mechanisms. PHD2 activity assays were carried out to investigate any modulation of PHD2 activity by LIMD1. A succinate-glo assay was used to follow the production of succinate, a byproduct of proline hydroxylation, at various PHD2 concentrations (Figure 4.17A). Here, HIF-1 $\alpha$  CODD peptide was used as a substrate for PHD2. Hydroxylation of the peptide by PHD2 results in the conversion of 2-oxoglutarate to succinate (Figure 4.16). Components of the succinate-glo assay then use succinate to generate ATP which is used by a luciferase enzyme to produce light (Figure 4.16).

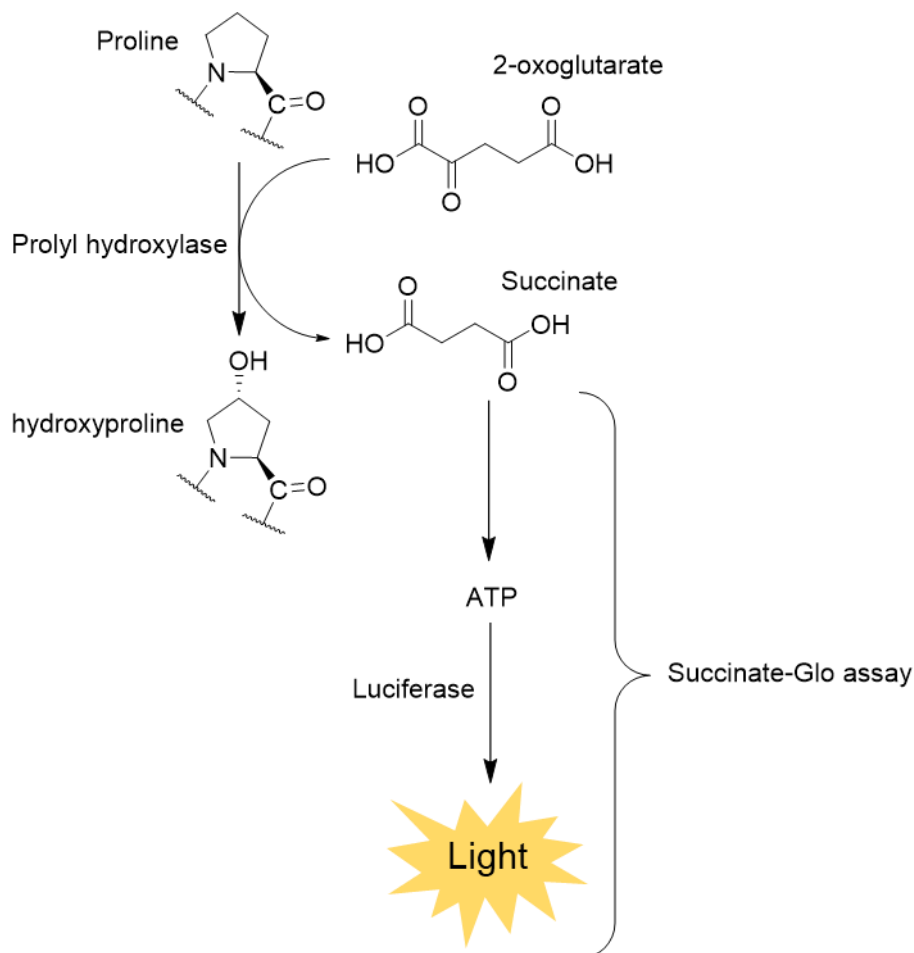


Figure 4.16. Schematic representation of the PHD2 activity assay used. PHD2 uses 2-oxoglutarate to hydroxylate a proline residue, resulting in the production of a succinate by-product. The succinate-Glo assay then converts succinate into ATP which can be used by a luciferase enzyme to produce light.

Initial assays with various PHD2 concentrations verified that the assay was working as intended and provided a working concentration to assay the effect of LIMD1 on PHD2 activity (Figure 4.17A). The signal at the highest PHD2 concentration was close to saturating the detector and so higher concentrations were not used. Further experiments used 1  $\mu$ M PHD2 in order to maximise the signal without saturating the detector. Carrying out the assay with a range of LIMD1 166-260 concentrations revealed that addition of LIMD1 166-260 had no apparent effect on the activity of PHD2 towards HIF-1 $\alpha$  CODD (Figure 4.17B). This lack of activity change suggests that LIMD1 functions through a purely scaffolding mechanism, by increasing the local concentrations of PHD2 and VHL, rather than through an allosteric effect modulating the activity of PHD2. Moreover, this result is consistent with LIMD1 interacting with PHD2 *via* a site that does not prevent binding of substrate peptides. It is however possible that LIMD1 could impact VHL activity as this was not tested.

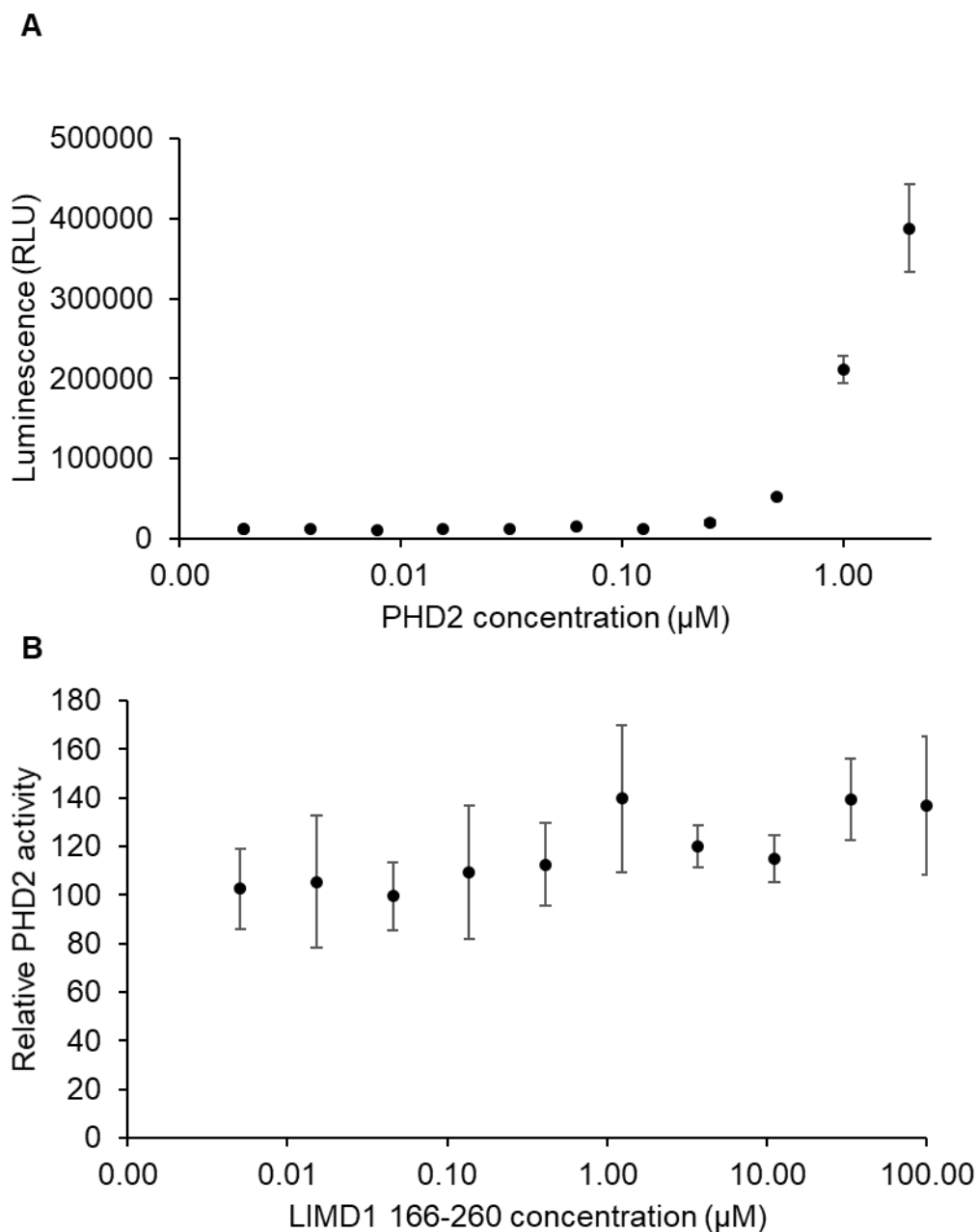


Figure 4.17. PHD2 activity assays using a succinate-glo luminescent assay. **A.** Luminescence at various PHD2 concentrations. **B.** relative activity of 1 μM PHD2 with and without LIMD1 166-260 at a range of concentrations. The results show no apparent effect of LIMD1 on PHD2 activity.

With the smaller binding region of LIMD1 identified it was then feasible to attempt to assign the backbone resonances for this region. Resonance assignment is the process by which signals in NMR spectra are linked to the specific atoms they result

from. This would allow identification of the peaks affected by PHD2 binding and thus further narrow the binding region on LIMD1. Triple resonance backbone assignment allows for sequential assignment of resonances resulting from atoms in the backbone of the protein. Typically, triple resonance backbone assignments rely on pairs of experiments where an experiment transfers magnetisation between the amide NH group of one residue to the C $\alpha$  and/or C $\beta$  protons of that residue and the preceding residue (in this case a HNCACB experiment) and another experiment where magnetisation is only transferred to the preceding residue (in this case a HN(CO)CACB experiment). The resulting signals can then be sequentially linked by matching chemical shifts of C $\alpha$  and C $\beta$  atoms to the i-1 signals from another amide. The chemical shift values of the C $\alpha$  and C $\beta$  atoms can be used to predict the amino acid type at a particular position which then allows for sequential signals in the spectra to be assigned to a stretch of the protein sequence. Assignment of LIMD1 166-260 was hindered by a high number of prolines (13), which break the sequential assignment due to the lack of amide protons, and low overall sequence complexity in this region, which caused poor dispersion of signals. Despite these difficulties, assignment of a stretch of residues from 168-193 (not including prolines) was possible. The HNCACB/HNCOACB strips for this region are shown in figure 4.18.

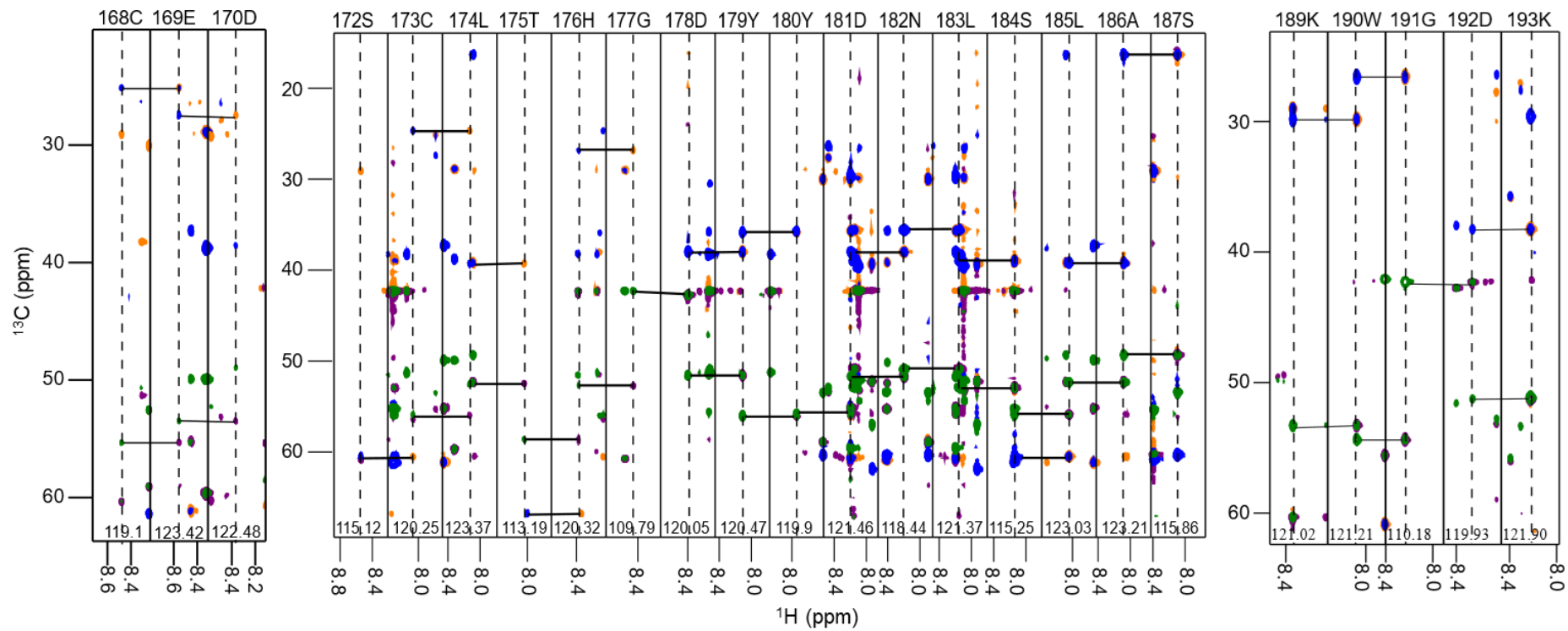


Figure 4.18. strips from the HNCACB and HNCOCACB spectra of LIMD1 166-260 for the assigned region of 168-193. The HNCACB spectra is shown with green positive and blue negative contours and the HNCOCACB spectra is shown with yellow negative and purple positive peaks.



the mean for the change in the signal-to-noise ratio of a peak on addition of PHD2. A higher Z-score means peaks are seen to disappear whilst a negative Z-score means peaks increase in intensity.





Figure 4.20 shows a region of peaks between 178-191 that significantly decreased in intensity on addition of PHD2. Intensity changes across the rest of the spectra were less significant with the exception of two peaks. It is possible that these two peaks arise from residues in the vicinity of the 178-193 region identified, although this cannot be confirmed. These residues may also represent an alternative PHD2 binding site or that these sites are affected by long range interactions with the PHD2 binding site. Closer analysis of the LIMD1 sequence revealed that the 168-191 region is one of the few regions of high sequence conservation within the pre-LIM region of mammalian LIMD1 orthologues (Figure 4.21).

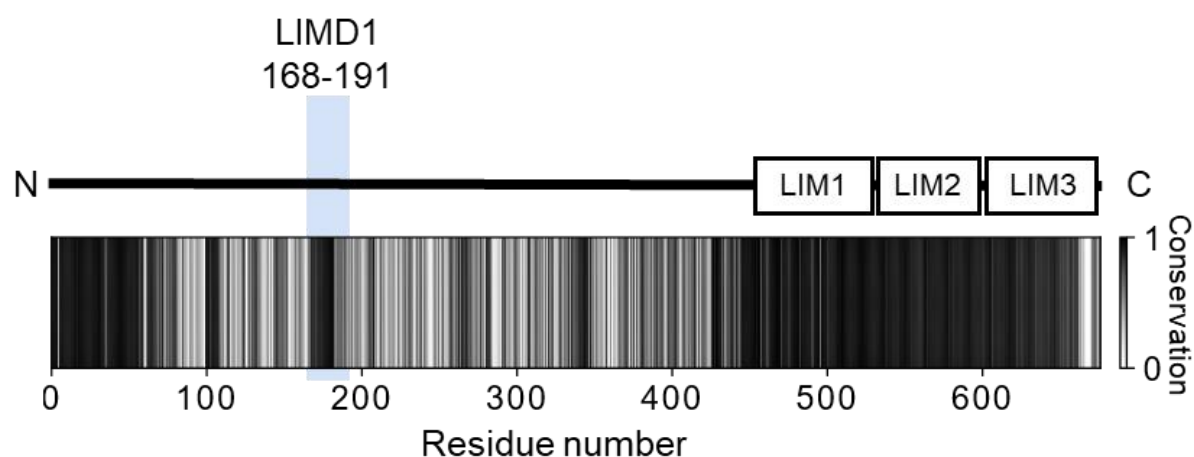


Figure 4.21. A heatmap showing the sequence conservation of human LIMD1 residues across all mammalian LIMD1 orthologs annotated in the NCBI database (214 sequences). Sequences were aligned using Clustal Omega and a conservation score was generated using the Jalview [60, 65, 205]. The LIMD1 domain positions are indicated above the heatmap with the 168-191 region of LIMD1 highlighted.

PHD2 had previously been shown *in vivo* to be unable to bind to LIMD1 $\Delta$ 186-260 but was still able to bind to a LIMD1 $\Delta$ 239-260 which, together with these results, would suggest the  $\Delta$ 186-260 mutation cut the PHD2 binding site in half. This consistency with the *in vivo* experimental data along with the sequence conservation and a decrease in the disorder prediction at this site (Figure 1.8) provide compelling

evidence that the 168-191 region of LIMD1 is an important PHD2 recognition site in LIMD1.

The ability of this small region of LIMD1 to bind to PHD2 was further explored using a peptide of LIMD1 168-191. It was initially hoped that co-crystallisation of PHD2 with this peptide could be achieved, however extensive testing of crystallisation conditions (1728 conditions tested) did not yield any viable crystals. Due to the small size of the peptide, it was possible to readily observe the peptide by NMR at natural isotope abundance. Comparison of the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC spectrum of LIMD1 168-191 with the spectra of LIMD1 166-260 with and without PHD2 revealed most of the peaks in the LIMD1 168-191 spectrum corresponded to peaks in the LIMD1 166-260 spectra that are impacted by addition of PHD2 (Figure 4.22). This both provides confidence in the assignments of LIMD1 166-260 and suggests that the LIMD1 168-191 peptide exists in a similar conformational ensemble as the longer LIMD1 construct.

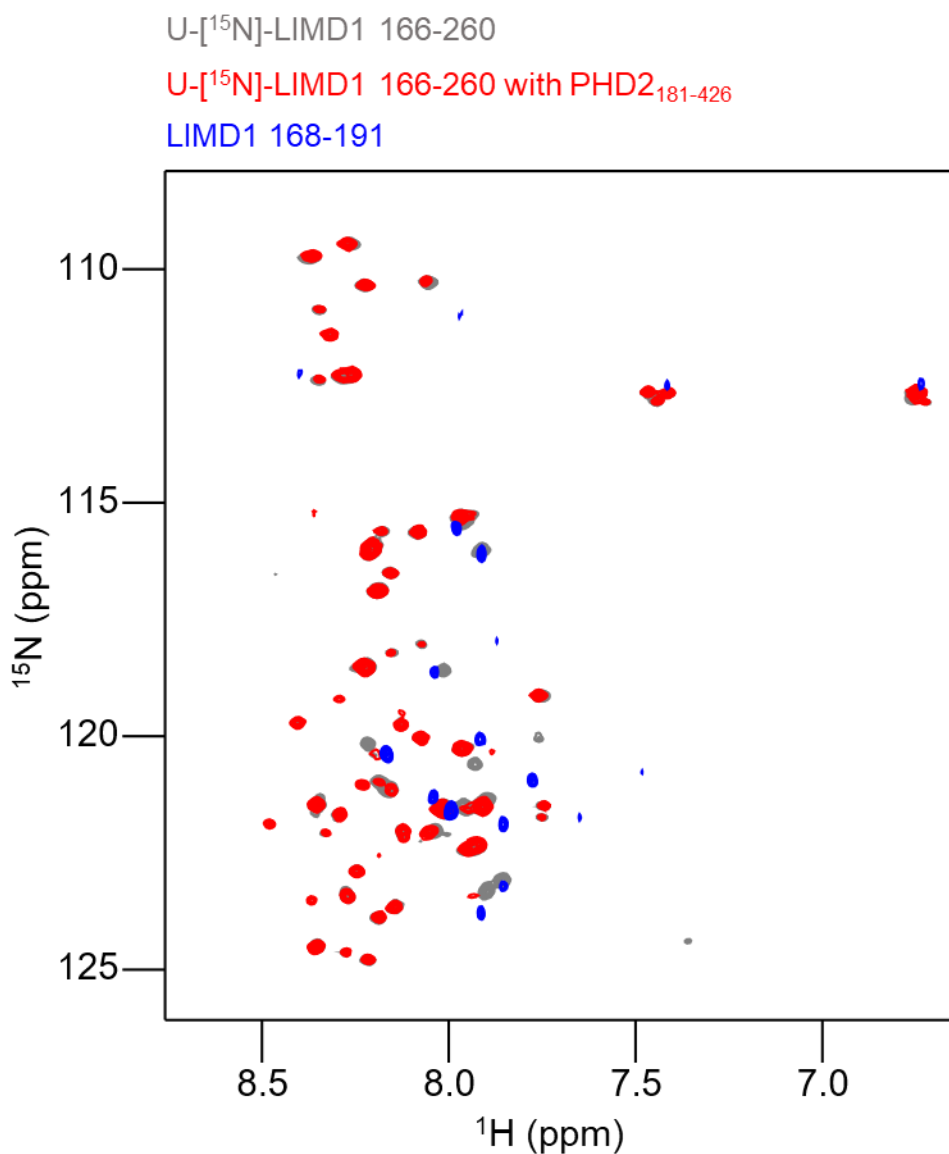


Figure 4.22. 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra for U-[<sup>15</sup>N]-LIMD1 166-260 (grey), U-[<sup>15</sup>N]-LIMD1 166-260 with PHD2<sub>181-426</sub> (red), and natural abundance LIMD1 168-191 (blue).

The PHD2 binding ability of the LIMD1 168-191 peptide could be demonstrated by following the loss of signal from all peaks in the LIMD1 168-191 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectrum on addition of PHD2 (Figure 4.23). Additional peaks are seen on addition of PHD2 which likely correspond to disordered residues or side chain NH groups from PHD2. This can be seen from the overlay of the 2D (<sup>1</sup>H, <sup>15</sup>N) spectra of LIMD1 168-191 with PHD2 spectrum with the spectrum of PHD2 on its own (appendix figure 7.9).

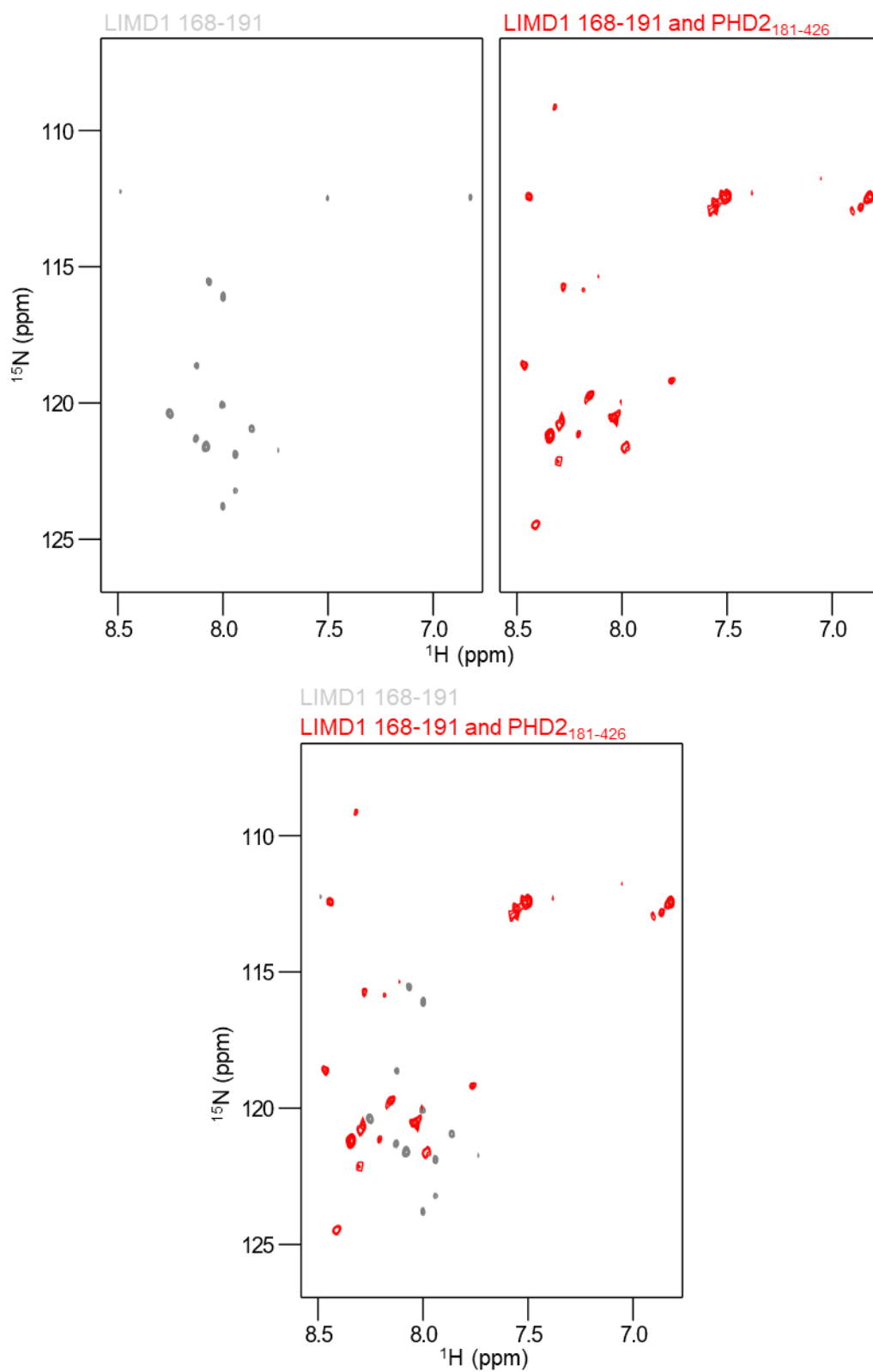


Figure 4.23. 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC spectra of unlabelled LIMD1 168-191 in grey and of a mixture of unlabelled LIMD1 168-191 and PHD2 181-426 in red.

### 4.2.3. Identifying the Binding Site of LIMD1 on PHD2

In order to better understand the interaction between LIMD1 and PHD2, we aimed to identify the LIMD1 binding site on PHD2. Due to the size of PHD2, it was necessary to use deuteration as a means of improving the signal intensity and signal overlap of PHD2 for observation by solution NMR spectroscopy. Deuteration involves producing the protein with protons replaced with deuterium but reintroducing protons at specific sites. This can be done by back-exchanging, exchangeable sites by simply putting the protein in  $^1\text{H}_2\text{O}$  based buffer or by specifically isotope labelling non-exchangeable sites. Deuterium is not observed by  $^1\text{H}$  NMR experiments, and so deuteration decreases the number of peaks observed in the 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra, alleviating problems with peak overlap. Deuteration also allows for greater signal intensity, even in very large proteins, by reducing the cross-relaxation due to dipolar interactions between  $^1\text{H}$  atoms [206]. Methyl groups are commonly selected for isotopic labelling as they offer high sensitivity due to the fast rotation of the three protons around a three-fold symmetry axis of the carbon-carbon bond. Even greater sensitivity can then be achieved through the use of TROSY experiments which use the cancellation between dipole-dipole relaxation and relaxation due to chemical shift anisotropy or different dipolar coupling to select for longer relaxation rates [207]. Perdeuterated PHD2 was produced with specific  $^{13}\text{C}$ ,  $^1\text{H}$  labelling of Ile $^{\delta 1}$ , Leu $^{\text{proS}}$  and Val $^{\text{proS}}$  methyl groups. A more detailed explanation of how this labelling, and others, can be achieved is presented in section 5.1.2. The protein was purified in  $^1\text{H}_2\text{O}$  based buffer resulting in back exchange of exchangeable protons. This back exchange allows for observation of the backbone NH group. The 2D ( $^1\text{H}$ - $^{15}\text{N}$ ) TROSY spectra of PHD2 with and without LIMD1 168-191 revealed a number of peaks shifting on addition of the peptide (Figure 4.24). Partial backbone assignments were available for PHD2 181-402 (BMRB entry: 26741) covering ~83% of the sequence [137]. It was possible to transfer 120 of 183 previously assigned peaks to the PHD2 181-426 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY spectra, allowing for ~49% coverage of PHD2 181-426. Mapping the assigned and unassigned residues onto the structure of PHD2 showed a good coverage of assigned residues across the surface of the protein (Figure 4.24).

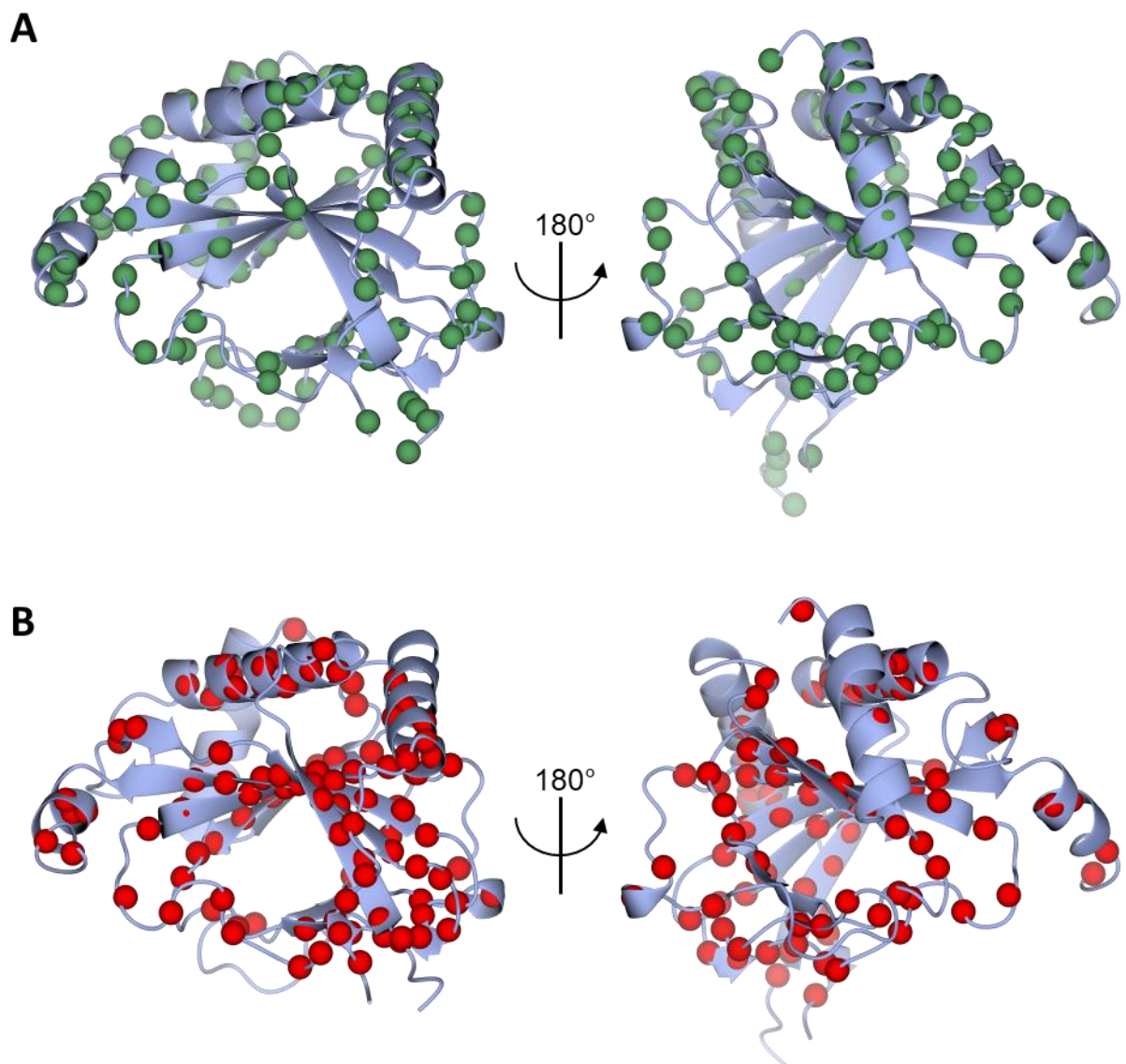


Figure 4.24. The structure of PHD2 (PDB 2G1M) with **A.** assigned backbone nitrogen atoms shown as green spheres and **B.** unassigned backbone nitrogen atoms shown as red spheres.

Differences in the spectra collected in this work and spectra used for resonance assignment were likely caused by differences in the buffers used as well as the

lengths of the PHD2 constructs. The assignments allowed identification of a majority of the peaks affected on addition of LIMD1 168-191 (Figure 4.25).

[U-<sup>15</sup>N, <sup>2</sup>H],(<sup>δ15</sup>L<sup>proSVproS</sup>)-[<sup>13</sup>C<sup>1</sup>H<sub>3</sub>]-PHD2<sub>181-426</sub>

[U-<sup>15</sup>N, <sup>2</sup>H],(<sup>δ15</sup>L<sup>proSVproS</sup>)-[<sup>13</sup>C<sup>1</sup>H<sub>3</sub>]-PHD2<sub>181-426</sub> With LIMD1<sub>168-191</sub>

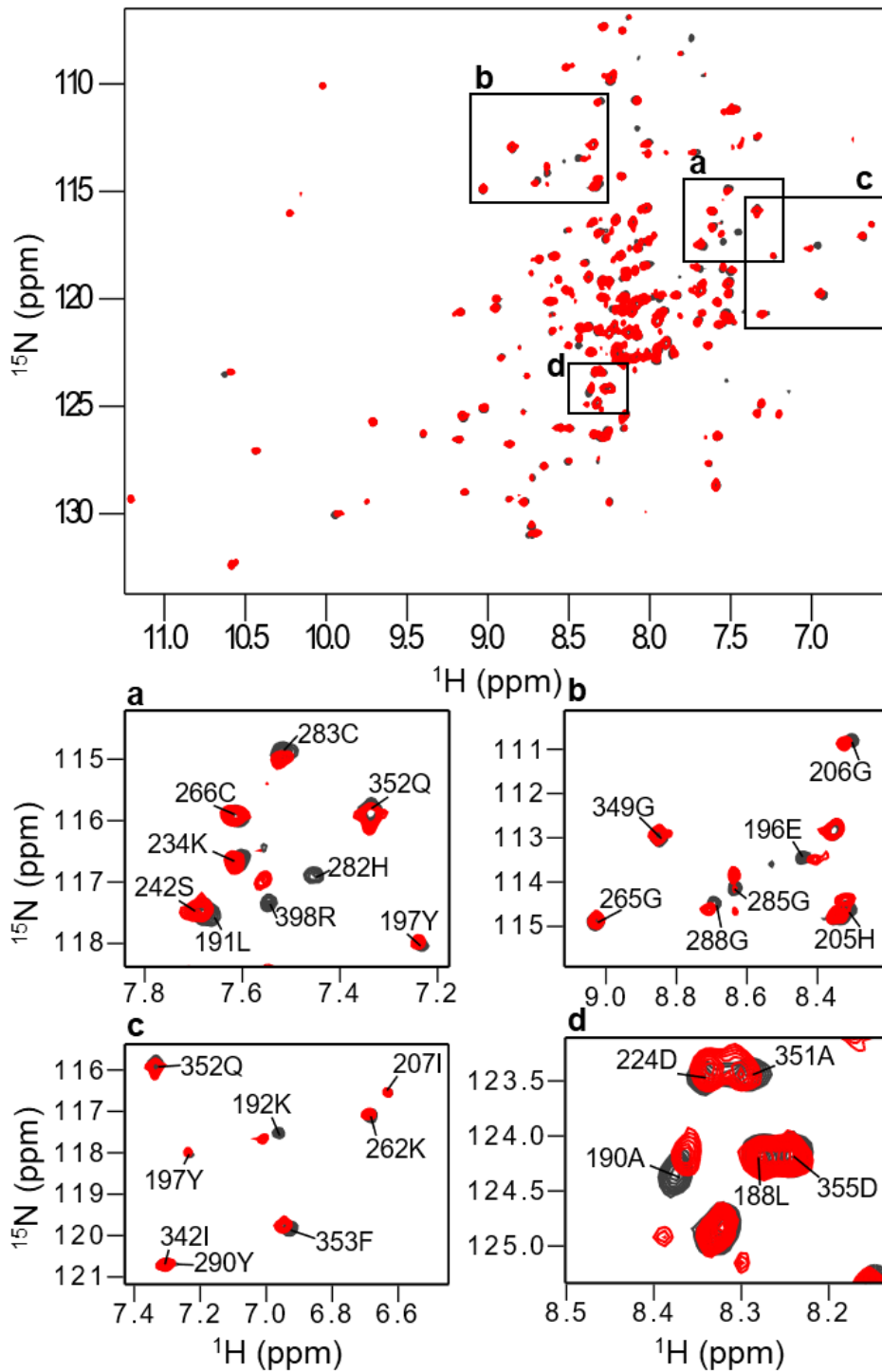


Figure 4.25. 2D (<sup>1</sup>H-<sup>15</sup>N) TROSY spectra of [U-<sup>15</sup>N, <sup>2</sup>H],(<sup>δ15</sup>L<sup>proSVproS</sup>)-[<sup>13</sup>C<sup>1</sup>H<sub>3</sub>]-PHD2<sub>181-426</sub> with (red) and without (black) LIMD1 168-191. Regions **a-d** of the spectra with relevant chemical shift perturbations are shown below the full spectra with residue assignments indicated.



The experiments were also collected in the presence of HIF-1 $\alpha$  CODD (Figure 4.26). These results show a much greater change in the PHD2 spectrum on addition of HIF-1 $\alpha$  CODD compared to the addition of LIMD1 168-191. The changes in the PHD2 spectra on addition of HIF1 $\alpha$  CODD are largely consistent with chemical shifts observed in previous work, although direct comparison is difficult due to differences in the buffers, protein constructs and the NMR experiments conducted [137]. The large chemical shift changes on addition of HIF-1 $\alpha$  CODD could be anticipated, as a large conformational change from the open to closed conformation is believed to occur on binding to HIF-1 $\alpha$  [136]. Interestingly there appears to be an additive effect on the chemical shift perturbation on addition of both LIMD1 168-191 and HIF-1 $\alpha$  CODD to PHD2 (Figure 4.26). This suggests that both peptides are able to bind to PHD2 simultaneously. Peaks can be seen to be affected by both LIMD1 and HIF-1 $\alpha$ , as well as by only one or the other providing additional evidence for distinct, binding sites for the two peptides.

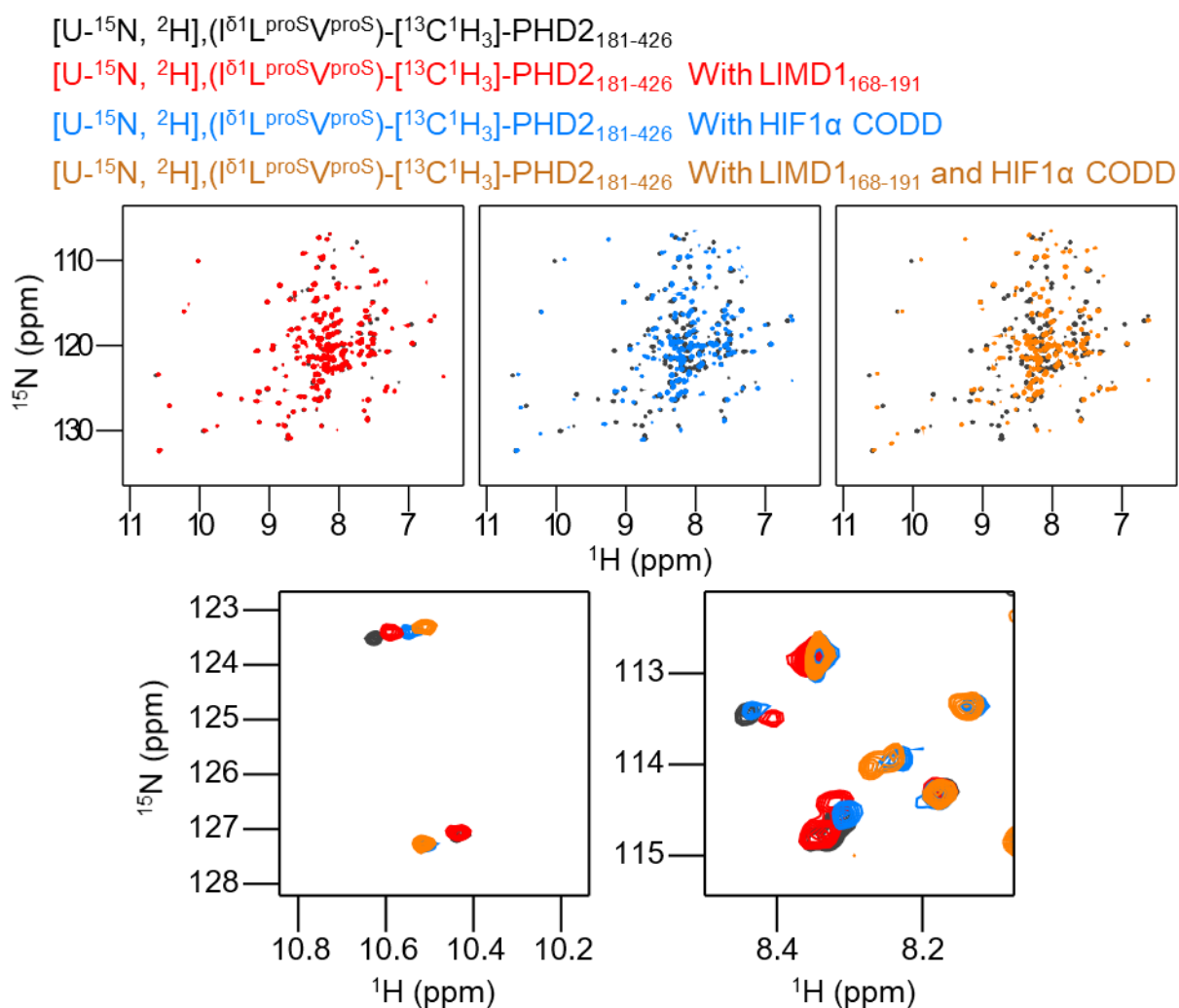


Figure 4.26. 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY spectra of  $[U-^{15}\text{N}, ^2\text{H}], (i^{15}\text{L}^{\text{proSVproS}})-[^{13}\text{C}^1\text{H}_3]\text{-PHD2}_{181-426}$  in black with unlabelled LIMD1 168-191 in red, with unlabelled HIF-1 $\alpha$  CODD in blue and with both unlabelled LIMD1 168-191 and unlabelled HIF-1 $\alpha$  CODD in yellow. Examples of shifted and unaffected peaks are shown in magnified regions of the spectra.

A similar trend was seen in the 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQCs to the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY experiments, with HIF-1 $\alpha$  CODD having a greater impact on the PHD2 spectra than LIMD1 168-191 but the combination of peptides producing a different effect on the spectra than either peptide individually (Figure 4.27). Two peaks were identified as being affected as a result of LIMD1 168-191 binding, with one of these peaks unaffected by HIF-1 $\alpha$  CODD binding (Figure 4.27).

$[U-^{15}N, ^2H], ([\delta^{13}C]^{proSVproS})-[^{13}C^1H_3]$ -PHD2<sub>181-426</sub>  
 $[U-^{15}N, ^2H], ([\delta^{13}C]^{proSVproS})-[^{13}C^1H_3]$ -PHD2<sub>181-426</sub> With LIMD1<sub>168-191</sub>  
 $[U-^{15}N, ^2H], ([\delta^{13}C]^{proSVproS})-[^{13}C^1H_3]$ -PHD2<sub>181-426</sub> With HIF1 $\alpha$  CODD  
 $[U-^{15}N, ^2H], ([\delta^{13}C]^{proSVproS})-[^{13}C^1H_3]$ -PHD2<sub>181-426</sub> With LIMD1<sub>168-191</sub> and HIF1 $\alpha$  CODD

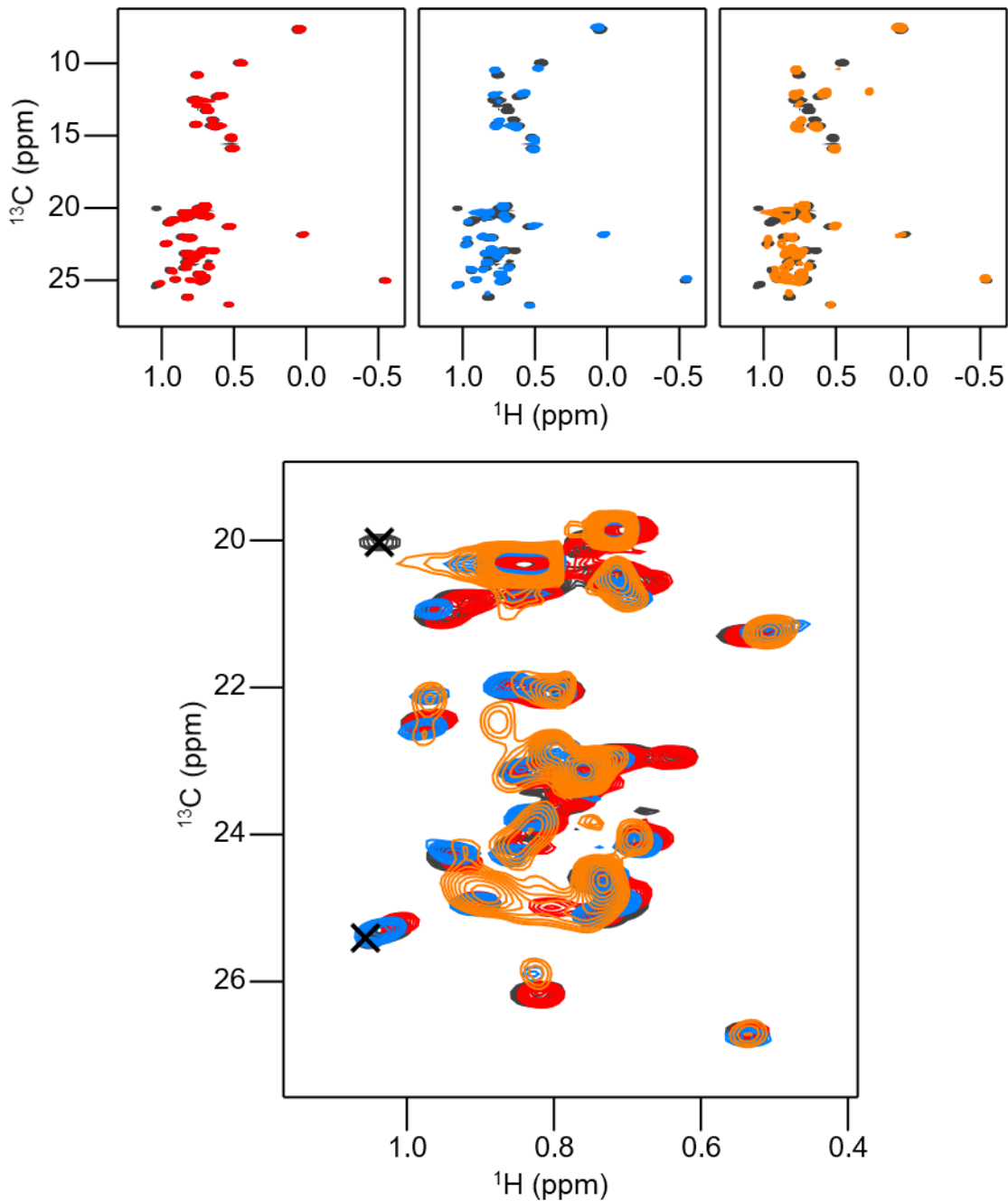


Figure 4.27. 2D ( $^1H$ ,  $^{13}C$ ) HSQC spectra of  $[U-^{15}N, ^2H], ([\delta^{13}C]^{proSVproS})-[^{13}C^1H_3]$ -PHD2<sub>181-426</sub> in black with LIMD1 168-191 in red, with HIF-1 $\alpha$  CODD in blue and with both LIMD1 168-191 and HIF-1 $\alpha$  CODD in yellow. A magnified version of the spectra is given with peaks perturbed by LIMD1 168-191 marked with a cross.

The 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra were less informative than the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY spectra due to the lack of available  $^{13}\text{C}$  assignments. Assignment of the 2D ( $^1\text{H}$ - $^{13}\text{C}$ ) HSQC peaks affected by LIMD1 168-191 would provide greater insight to help identification of the binding site of LIMD1 on PHD2. Assignments of these residues could be achieved through individually mutating the isoleucine, leucine, and valine residues in the protein [208]. Although this would be expensive and relies on none of these mutations severely impacting the structure of the protein. Alternatively, assignment could be achieved through the use of TOCSY experiments or NOE contacts [207]. These approaches would both be highly time consuming and would require the production of additional samples.

Chemical shift perturbations (CSPs) could be calculated for assigned, well resolved peaks in the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY spectra of PHD2 on addition of LIMD1 168-191 (Figure 4.28).

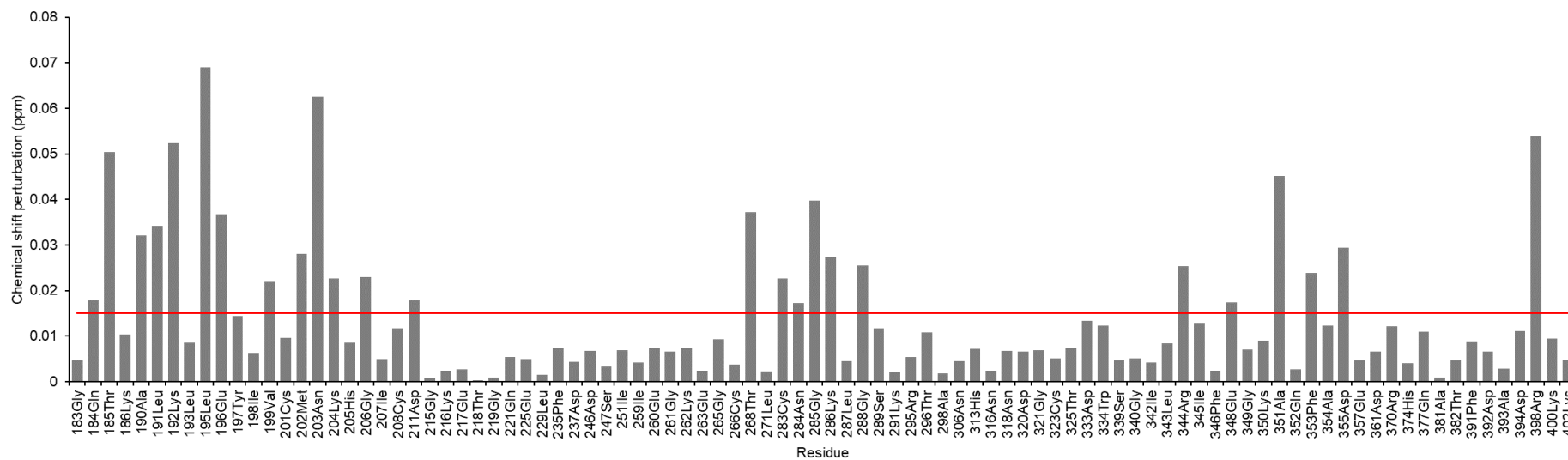


Figure 4.28. Chemical shift perturbations from the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY spectra of PHD2<sub>181-426</sub> on addition of LIMD1 168-191. The red line shows the cut off of 0.015 ppm that was used for subsequent analysis.

Peaks that were seen to shift by  $>0.015$  ppm on addition of LIMD1 168-191 made up 25% of the total CSPs. Mapping these shifted peaks and peaks that were seen to reduce in intensity by more than 75% on addition of LIMD1 168-191 (Appendix Figure 7.22), onto the previously determined structure of PHD2, revealed that these residues generally clustered around a single site on PHD2 (Figure 4.29A-B) [136]. The high density of residues affected in this region suggests that this is the location of LIMD1 binding to PHD2. A few additional residues were affected which may be due to long range effects of a conformational change. The proposed binding site is also distinct from the substrate binding site (Figure 4.29C).

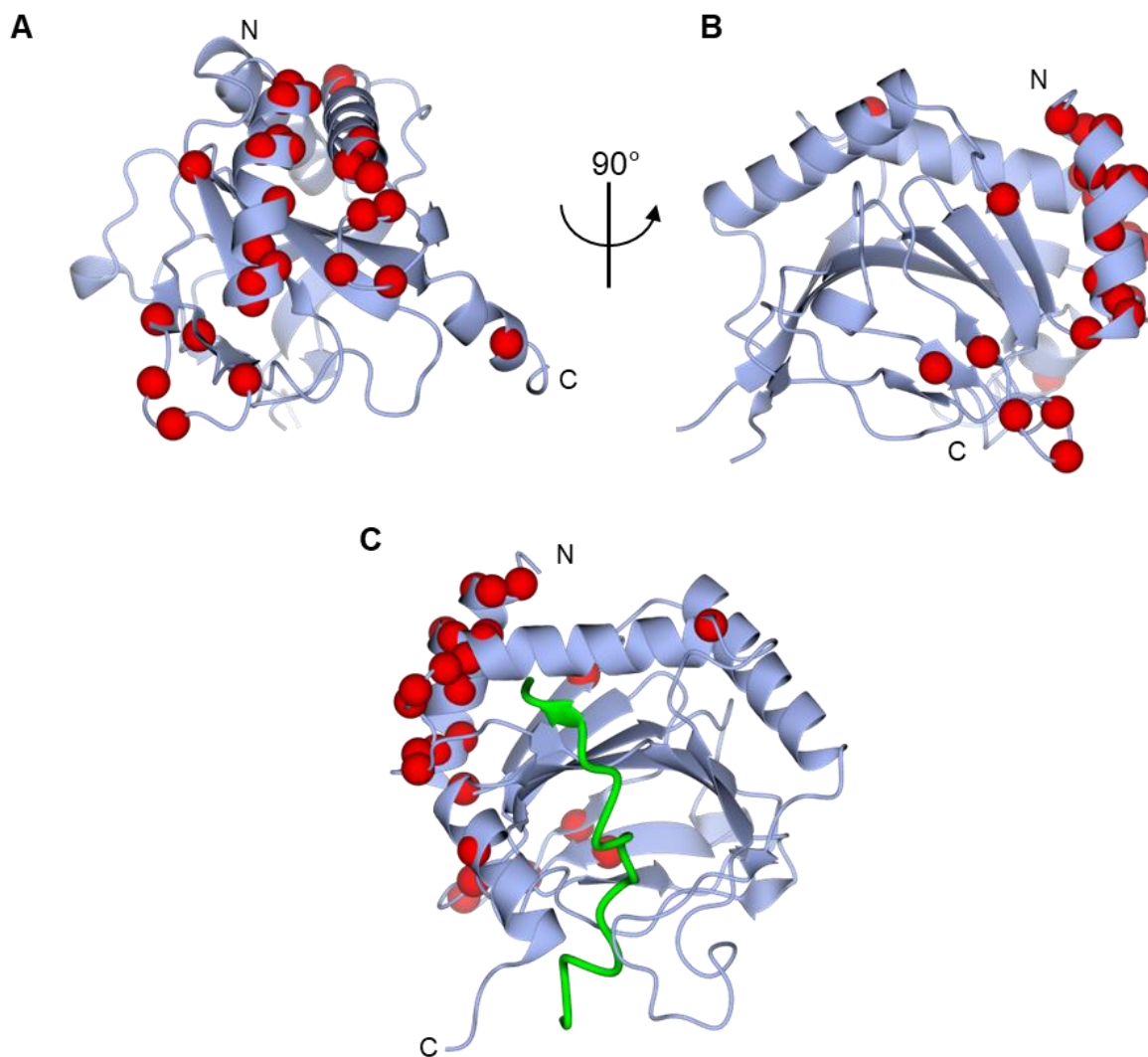


Figure 4.29. The crystal structure of PHD2 (blue) with the backbone nitrogen of residues perturbed on addition of LIMD1 168-191 shown as red spheres. **A.** PDB: 2G1M [182] **B.** 2G1M viewed from 90° compared to **A.** **C.** PHD2 bound to HIF1α CODD (green) PDB: 5L9B [136]. The N and C termini are indicated on the structures.

AlphaFold2 models of the HIF-1α CODD:PHD2 181-426 :LIMD1 168-191 interaction place the LIMD1 peptide alongside the α-helix at position 190 to 205 on PHD2 (Figure 4.30A), although with poor confidence in the conformation of the peptide [163]. The PAE plots for the models (Appendix figure 7.27B) show a high confidence in the position of the LIMD1 peptide (B) relative to the position of PHD2 (A) but a low confidence on the relative position of PHD2 (A) relative to the position of the LIMD1

peptide (B). This is likely due to the relatively small number of sequences in the multiple sequence alignment for LIMD1 168-191 (Appendix figure 7.27A) which results in a low confidence structure for the LIMD1 peptide (Appendix figure 7.27C) [163, 164]. The 190-205 region of PHD2 was also seen to have the largest degree of chemical shift perturbations on addition of LIMD1 168-191 (Figure 4.27) showing an agreement between AlphaFold2 and the experimental data. The predicted binding mode would also permit interactions with the 344-355 region which was also seen to experience chemical shift perturbations on addition of LIMD1 168-191. The reasonably high confidence in the position of the LIMD1 peptide relative to PHD2 and the agreement of this with the experimental results is consistent with LIMD1 interacting with the 190-205  $\alpha$ -helix on PHD2 (Figure 4.30A). The residues perturbed on the other face of this helix may not be in direct contact with the peptide but may be perturbed by changes to the position of the 190-205  $\alpha$ -helix on binding of LIMD1. Other residues in the vicinity of the proposed LIMD1 binding site may not have been identified due to the sparsity of the PHD2 backbone assignments (~49%).



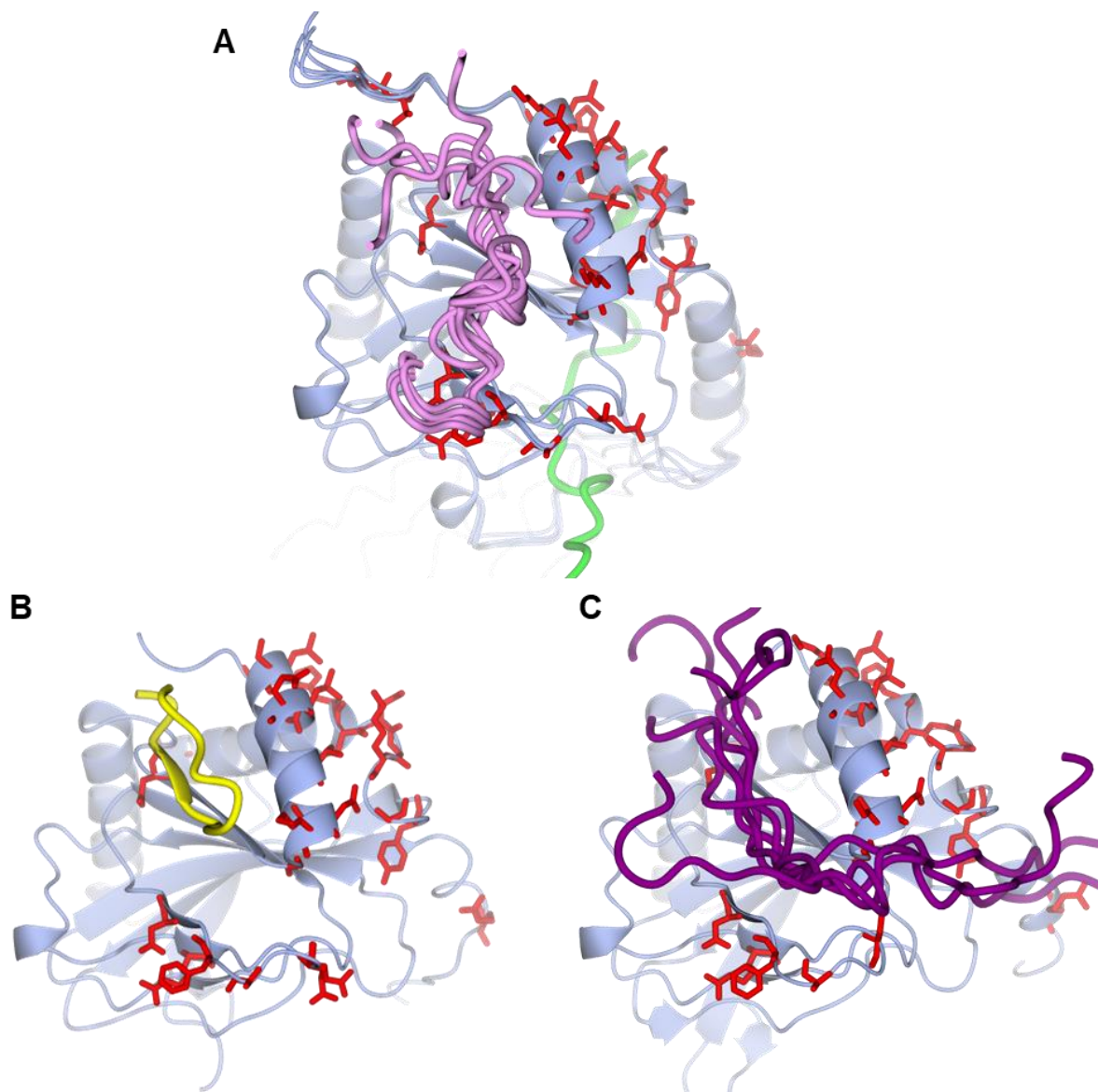


Figure 4.30. **A.** AlphaFold2 models for the interaction of PHD2 (blue) with HIF CODD (green) and LIMD1 168-191 (pink). 5 models are shown for LIMD1 168-191 and PHD2 with only the top ranked HIF CODD model shown. The 5 structures of PHD2 predicted by Alohafold2 superposed well with an experimental structure of PHD2 (PDB code: 2G1M) with RMSDs in secondary structure regions of 0.66 Å /200 C $\alpha$  atoms, 0.65 to 0.66 Å over 200 equivalent C $\alpha$  atoms [182]. The structure and position of HIF CODD was predicted with very high confidence (appendix figure 7.27). **B.** The structure of PHD2 (blue) bound to a RaPID derived cyclic peptide (yellow) PDB code:6YW1 [209]. **C.** results of HPEPDOCK modelling of LIMD1 168-191 (Purple) binding to PHD2 (blue) (PDB code: 2G1M) with binding site residues set to 190-205 [181, 182]. The top 5 results from the docking calculations are shown. Residues perturbed by LIMD1 168-191 binding are shown in red.

A structure of PHD2 in complex with a RaPID derived cyclic peptide binding to the same region of PHD2 as suggested for LIMD1 was reported in 2020 (Figure 4.30B) [209, 210]. McAllister T. E. *et. al.* (2018) had previously shown addition of the cyclic peptide to PHD2 produced similar changes to the PHD2 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectra as were seen with LIMD1 168-191, with the majority of CSPs between residues 190-215 [211]. The similarity of the results seen with the LIMD1 peptide and the RaPID derived cyclic peptide, suggests a potential shared binding mode for the two peptides. The mode of binding of this peptide may provide insight into the mode of binding of LIMD1. The 190-205  $\alpha$ -helix region on PHD2 (the region with the highest number of perturbed peaks) was used as an input for binding residues for computational docking of LIMD1 168-191 on PHD2 (Figure 4.30C) [181, 212]. The docking results are relatively consistent with the AlphaFold2 models. Future work could mutate residues within this region in order to see if binding of LIMD1 is diminished.. The region of PHD2 from 181 to 220 has been shown to be involved in nuclear import, although the exact mechanism for this remains unclear [213, 214]. This could suggest that either this region has multiple roles, in binding to LIMD1 and nuclear import, or that LIMD1 may be a regulator of PHD2 nuclear transport. Following PHD2 localisation by using a GFP-PHD2 fusion at different levels of LIMD1 expression would help to better understand any role of LIMD1 in regulating the nuclear transport of PHD2.

### 4.3. Conclusion and Future Work

The interactions of LIMD1 with both VHL and PHD2 have been validated and significantly further defined *in vitro* using recombinantly expressed proteins and demonstrated by NMR spectroscopy. This work highlights the power of NMR spectroscopy to investigate protein structure and interactions that are inaccessible to other techniques. Investigation of the interaction between LIMD1 and VHL was demonstrated although further investigation of this interaction was limited by issues with the stability and production of the individual protein components. Additional work may benefit by attempting to work with either the full LIM domain region of LIMD1 or full length LIMD1 as the other LIM domains may have a stabilising effect. Future work could also benefit by studying the effect of Rho-related BTB domain-containing

protein 3 (RHOBTB3) on the interaction of LIMD1 with both VHL and PHD2 [148]. RHOBTB3 has been shown to be an additional scaffold protein in the HIF-1 $\alpha$  degradation pathway with potential interactions demonstrated to VHL, PHD2 and LIMD1. In order to obtain a more complete picture of the function of the LIMD1 mediated HIF-1 $\alpha$  degradation complex, it may be important to incorporate this additional scaffold protein.

A potential PHD2 binding site was identified on the LIMD1 pre-LIM region. The observation of a narrow range of chemical shifts in the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectra for this region confirmed that this region is largely disordered. A series of overlapping constructs of this region of LIMD1 were used to narrow down the PHD2 binding site to between residues 166 and 260. Partial backbone assignment of this region revealed a potential binding site for PHD2 at position 168-191 of LIMD1. Further analysis using a LIMD1 168-191 peptide confirmed the ability of this peptide to bind to PHD2. High sequence conservation, and agreement of this result with *in vivo* results, provided compelling evidence for the 168-191 region of LIMD1 to be the PHD2 binding site. Binding to this region could be confirmed by generating a crystal structure of the complex or by using NMR to observe the effects of site directed mutagenesis on LIMD1 168-191 binding to PHD2. There is some sequence similarity seen for this region between LIMD1 and AJUBA (one of the only related sites in the pre-LIM region) but not for WTIP (Figure 1.7). The difference in the sequences of the LAW proteins at this site could explain differences in the affinities for PHD enzymes between the LAW proteins. The identified PHD2 binding site at 168-191 is close to the LIMD1 140-166 site believed to bind to AGO2 [25]. It may be that there is some shared binding site between the PHD2 and AGO2 or that the closeness of the binding sites acts to prevent simultaneous binding of AGO2 and PHD2, in order to limit crosstalk between the hypoxic and miRNA silencing pathways. Binding assays using a stretch of LIMD1 covering both binding sites could be used to investigate whether LIMD1 can bind to both PHD2 and AGO2 simultaneously or if there is a preference for binding to either PHD2 or AGO2.

The LIMD1 168-191 peptide was then used to investigate the binding site of LIMD1 on PHD2. Deuteration and specific methyl labelling were used to allow for observation of PHD2 by NMR spectroscopy. Available assignments for PHD2 were then used to map the binding site for LIMD1 onto the structure of PHD2. A potential binding site for LIMD1 on PHD2 was identified which agrees with AlphaFold 2 predictions and matches the binding site of a RaPID derived cyclic peptide [163, 209]. Further work could be carried out to test the effect of mutations in this region on the binding affinity of PHD2 for LIMD1. Single point mutations of all solvent accessible residues in the 190-205 region of PHD2 (L191, K192, L195, E196, Y197, P200, N203, K204, H205 ) would give valuable site specific information on the residues responsible for binding to LIMD1 [163, 181, 212]. The RaPID derived cyclic peptide could also be used for competition assays with LIMD1 [209]. The difficulty in demonstrating an interaction between LIMD1 and PHD2 by co-elution suggests that the interaction is relatively low affinity, conversely the cyclic peptide has been shown to have a very high affinity for PHD2 ( $K_D$  of 0.27 nM) [211]. The differences the affinities of LIMD1 and the cyclic peptide for PHD2 means that the cyclic peptide would be expected to outcompete LIMD1 for binding to PHD2 if they bind at the same site on PHD2. . It may also be worthwhile investigating any potential effect of LIMD1 on the nuclear import of PHD2 as the LIMD1 binding region identified has also been shown to be important for nuclear import of PHD2 [213]. PHD2 has been shown to hydroxylate HIF1 $\alpha$  in both the cytoplasm and the nucleus with nucleus-cytoplasmic shuttling of PHD2 being critical for interaction with HIF1 $\alpha$  [213]. The potential effect of LIMD1 on shuttling of PHD2 may be key to understanding the role of LIMD1 in hypoxic signalling. The characterisation of the LIMD1:PHD2 interaction may inform the development of drugs to target the interaction which could be used to treat ischemic diseases.

The findings from this study are in mixed agreement with a computational study investigating the interaction between LIMD1 and PHD2 [215]. The authors computationally predicted the effect of eight mutations at highly conserved residues on the structure of PHD2. The structural analysis identified a potential binding area on the surface of PHD2 formed by four polycythemia-related mutations (P200Q, N203K, K291I and R371H) that is not involved in any known protein-protein

interactions. As P200Q and R371H had previously been shown to have no effect on HIF-1 $\alpha$  hydroxylation *in vitro*, it was hypothesised that mutations in this region must impact partner protein binding [216]. As mutations impacting LIMD1 binding would not be expected to impact PHD2 activity *in vitro* the authors predicted that LIMD1 binds to the P200, N203, K291 and R371 region of PHD2 [215]. It is equally plausible that this site on PHD2 could be responsible for binding to other scaffold proteins known to interact with PHD2 such as PDL1/2 or mAKAP but these were not considered [142, 143, 145]. Residues P200 and N203 of PHD2, predicted in the computational study to interact with LIMD1, are within the 190-205 region identified in this study as a potential LIMD1 binding site with R371 close to this region in the structure. N203 was seen to have a large chemical shift change on addition of LIMD1 168-191 (Figure 4.28), whilst K291 is closer to the HIF-1 $\alpha$  binding site and was not affected by LIMD1 168-191 binding in the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC. The computational study assumed PHD2 binding occurred within the 186-260 region of LIMD1 identified by Foxler *et al.* (2012) which led to a predicted binding region of LIMD1 216-242 [24]. The work presented in this thesis identified LIMD1 168-191 as a potential PHD2 binding region which suggests the LIMD1 $\Delta$ 186-260 used in Foxler *et al.* (2012) cut the binding region in two [24]. The computational analysis may have better predicted the binding region of PHD2 on LIMD1 if a longer sequence was used to account for the potential of cleaving the binding site. The computational analysis performed by Minervini *et al.* (2016) could be repeated using the LIMD1 168-191 region to provide additional insights into the structural biology of the interaction between LIMD1 and PHD2 [215].

## 5. Results Chapter 3: Using Specific Isotopic Unlabelling to Aid Structural Studies of Proteins by Solution NMR Spectroscopy

### 5.1. Introduction

The work presented in this thesis focusses on the use of solution NMR spectroscopy to interrogate the structure and function of proteins. This chapter involves the development of new methods to aid in this type of study. Although the methods developed here were not successfully applied in the LIMD1 study, they remain applicable for similar investigations with more favourable characteristics (such as improved protein solubility).

NMR spectroscopy has the potential to observe a signal from every N, C and H atom in a protein. The technique typically relies on observation of spin active nuclei, particularly  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ .  $^1\text{H}$  is the most naturally abundant isotope of hydrogen (99.985% abundant) which allows for ready observation by NMR spectroscopy. However, the high abundance of  $^1\text{H}$  in proteins and the low signal dispersion observed for  $^1\text{H}$  in proteins makes interpretation of the spectra challenging due to the high degree of signal overlap. For the study of proteins, it is therefore common to use multidimensional, heteronuclear experiments to observe correlations between sets of  $^1\text{H}$ ,  $^{13}\text{C}$  and/or  $^{15}\text{N}$  nuclei [217]. These heteronuclear experiments provide improved signal dispersion compared to 1-dimensional  $^1\text{H}$  spectra and have allowed for detailed analysis of protein structure and dynamics. The observation of  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclei is however limited by the relatively low natural abundance of these isotopes, 1.1% and 0.05% respectively. For this reason, it is often necessary to enrich samples with  $^{13}\text{C}$  and/or  $^{15}\text{N}$  to increase the intensity of the observable signal. For proteins produced in *E. coli*, uniform  $^{13}\text{C}$  and/or  $^{15}\text{N}$  isotopic labelling can be achieved through the use of  $^{13}\text{C}$ -glucose and/or  $^{15}\text{NH}_4\text{Cl}$  as the sole carbon or nitrogen sources in the expression media [218]. Uniformly  $^{13}\text{C}$  and/or  $^{15}\text{N}$  labelled proteins contain a large number of observable nuclei, which can lead to a high

degree of signal overlap. This effect is more pronounced in larger proteins with more observable atoms and disordered proteins which have a narrower signal distribution due to the similarity in the chemical environment of each residue. In order to reduce signal overlap and aid in the interpretation of these spectra it can be useful to reduce the number of signals visible in NMR spectra (e.g., by selectively exciting only specific sets of atoms). Issues related to signal overlap can sometimes be reduced through the use of higher dimensionality experiments, although this is not always sufficient. Specific isotope labelling can be used to reduce the number of observable atoms in the protein and thus reduce the number of signals in spectra. Specific isotope labelling can be achieved in *E. coli* through the addition of an isotope labelled amino acid or biosynthetic precursor to an amino acid to the expression media. This results in a protein with only a select set of residues labelled with  $^1\text{H}$ ,  $^{13}\text{C}$  or  $^{15}\text{N}$  giving rise to greatly simplified NMR spectra. The use of specific isotope labelling is often limited by the high cost of isotope labelled amino acids or amino acid precursors. A more affordable alternative to specific isotope labelling is specific isotope unlabelling, also known as reverse labelling.

#### 5.1.1. Specific Isotope Unlabelling

Specific isotope unlabelling in *E. coli* can be achieved by adding a natural abundance amino acid or amino acid precursor to the expression media in addition to  $^{13}\text{C}$ -glucose and  $^{15}\text{NH}_4\text{Cl}$ . This media composition produces an otherwise uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$  labelled protein with a specific amino acid type or set of amino acids at natural isotope abundance [219, 220]. In a typical NMR spectrum, peaks resulting from unlabelled residues would disappear from the spectrum. This has the benefit of both simplifying the resulting spectra by reducing the number of peaks and thus reducing signal overlap as well as providing residue type information for the missing peaks. This residue type information and the low cost of the reagents required has allowed specific isotope unlabelling to be used as a tool to aid in resonance assignment of proteins [219-221].

The amino acids that can be targeted for either specific isotope labelling or unlabelling is limited by the metabolic pathways of the expression host. Some amino

acids can be metabolised by the host organism and the isotope (un)labelled sites will be incorporated into other residues. This effect is known as isotope scrambling. Considerable work has been carried out to understand these metabolic pathways and to determine which residues can be (un)labelled without scrambling [220, 222, 223]. Many amino acids can be effectively unlabelled through the use of metabolic precursors that lack the sites that would be scrambled if the amino acid itself was added. Amino acid precursors are also often cheaper to produce due to the reduced number of stereospecific sites. The type of amino acids that can be effectively unlabelled, free of scrambling was a key factor in deciding which amino acids to target in this study. An overview of specific isotope labelling and unlabelling approaches in *E. coli* is given.

#### 5.1.2. Routes for Targeting Aliphatic Residues

Aliphatic residues (Leu, Ile, Val and Ala) represent a highly desirable target for specific isotope labelling or reverse labelling due to their relatively high abundance (Leu: 9%, Ile: 5.2 %, Val: 6.6% and Ala: 8.3%) and broad distribution across proteins [224, 225]. Moreover, these residues contain methyl groups, which are excellent NMR probes for studying larger proteins [226].

Specific labelling and reverse labelling of the branched aliphatic amino acids has been used for generating backbone, sidechain and stereospecific assignments as well as for measuring NOEs [206, 220, 227-230].

The carbon atoms of leucine and valine can be labelled or reverse labelled with minimal scrambling through the use of the biosynthetic precursors  $\alpha$ -ketoisovalerate or acetolactate (Figure 5.1) [206, 220, 231-234]. Both precursors are chemically synthesised as a racemic mixture, which impacts how they label prochiral methyl groups. Only the 2*S* stereoisomer of acetolactate is a substrate of ketol-acid reductoisomerase, which means that it is possible to use acetolactate to stereospecifically target the prochiral methyl groups of leucine and valine [227, 233].  $\alpha$ -ketoisovalerate can be used for applications that require labelling of both prochiral



methyl groups. Leucine alone can be labelled by addition of 2-ketoisocaproate (Figure 15.1), a precursor of leucine that sits after the divergence of leucine and valine biosynthesis [206, 235]. Selectively deuterated versions of these precursors can be used to label larger proteins in combination with deuterated glucose and  $^2\text{H}_2\text{O}$  [233, 235].

The carbon atoms of isoleucine can be labelled with 2-ketobutyrate or 2-hydroxy-2-ethyl-3-ketobutyrate (Figure 5.1). These molecules also differentially target the methyl groups of Ile: 2-ketobutyrate is used to selectively label the Ile- $\delta_1$  methyl group while 2-(S)-hydroxy-2-ethyl-3-ketobutyrate can be used to target Ile- $\delta_1$  and/or Ile- $\gamma_2$  methyl groups [161, 236, 237].

Biosynthetic precursors that target isoleucine, leucine and valine can be used in combination with other metabolites or amino acids which suppress scrambling of carbon sites and off-target effects. For instance, prochiral methyl groups of valine can be selectively labelled by the addition of labelled pro-*R* acetolactate- $^{13}\text{C}_4$  or pro-*S* acetolactate- $^{13}\text{C}_3$  together with L-Leucine at natural abundance [160]. More details on methyl labelling of isoleucine, leucine and valine can be found in Kerfah *et al.* 2015b and Schutz and Sprangers 2020 [206, 207].

The methyl group of alanine provides a good probe for monitoring local structure and dynamics of the protein backbone [238]. Specific isotopic labelling or reverse labelling of alanine is hampered by the presence of alanine transaminases which convert alanine into the widely used metabolite pyruvate [239]. Pyruvate is an early precursor in isoleucine, valine and leucine biosynthesis, which means that the labelling pattern of alanine will be scrambled into other aliphatic residues in the target protein. Addition of 1 g/L natural abundance alanine (CAS: 56-41-7) will result in approximately 50% loss of signal from valine [221]. That said, scramble free  $^{13}\text{C}$  labelling of the carbonyl of alanine is possible via addition of 1- $^{13}\text{C}$ -alanine (CAS: 21764-56-7) [240], while labelling of the alanine methyl group can be achieved by adding other metabolites to suppress cross-talk between biosynthesis pathways

[241, 242]. In principle reverse labelling can be achieved using similar approaches, by adding alanine at natural abundance and other precursors with  $^{13}\text{C}$  labelling, though this would be expensive and impractical due to the high cost of  $^{13}\text{C}$  enriched amino acids and precursors.  $^{15}\text{N}$  labelling or reverse labelling of the backbone amine group of alanine, isoleucine, leucine and valine is problematic due to the action of various transaminases [243].

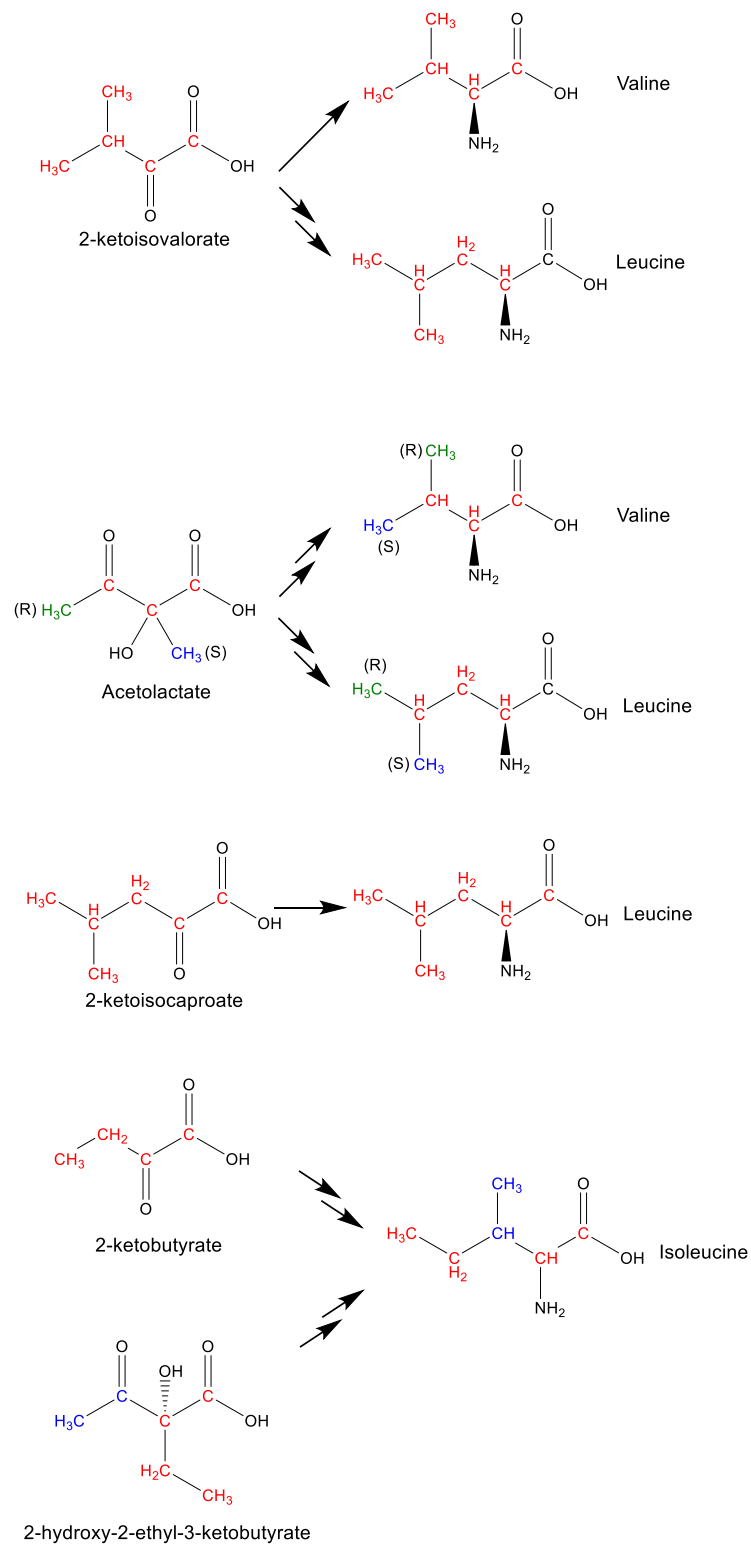


Figure 5.1. Metabolic precursors that can be used for isotopic labelling and reverse labelling of the carbon sites in branched chain aliphatic amino acids. Sites corresponding to labelled or unlabelled groups are coloured to show their starting and end positions.

### 5.1.3. Routes for Targeting Aromatic Residues

Aromatic residues are found at interaction interfaces and in the hydrophobic cores of proteins and hence can serve as excellent reporters of protein structure and interactions [244]. Tryptophan can be used as a sole carbon source by *E. coli* and so significant scrambling of carbon atoms occurs when tryptophan is added to the culture medium [245]. Additionally, tryptophanase can convert tryptophan to indole, pyruvate and ammonia which leads to nitrogen scrambling as ammonia is used in amino acid synthesis [221]. Aromatic amino acid transaminases cause significant nitrogen scrambling between tyrosine and phenylalanine when attempting to label or reverse label with either amino acid [221].

An early example of the selective labelling of aromatic residues through metabolic precursors was the use of shikimic acid to label aromatic protons of phenylalanine, tyrosine and tryptophan against a deuterated background (Figure 5.2) [246]. However, the synthesis of isotopically labelled shikimic acid is complicated, which has precluded widespread use. Phenylpyruvate and 4-hydroxy phenylpyruvate have been used to reverse label the carbon atoms phenylalanine and tyrosine, respectively (Figure 5.2) [220].  $^{13}\text{C}$  labelled versions of these precursors were later reported for isotopic labelling of phenylalanine and tyrosine [247]. Indole can be used for selective tryptophan labelling and reverse labelling of the tryptophan side chain (Figure 5.2) [248]. Isotopic labelling and reverse labelling of tryptophan can also be achieved through the use of indolepyruvate (Figure 5.2), which is part of the tryptophan degradation pathway rather than biosynthesis pathway [249]. Anthranilic acid (Figure 5.2) has also been reported as an alternative tryptophan labelling precursor, which allows both  $^{15}\text{N}$  and  $^{13}\text{C}$  labelling of side chain sites with minimal scrambling [250].

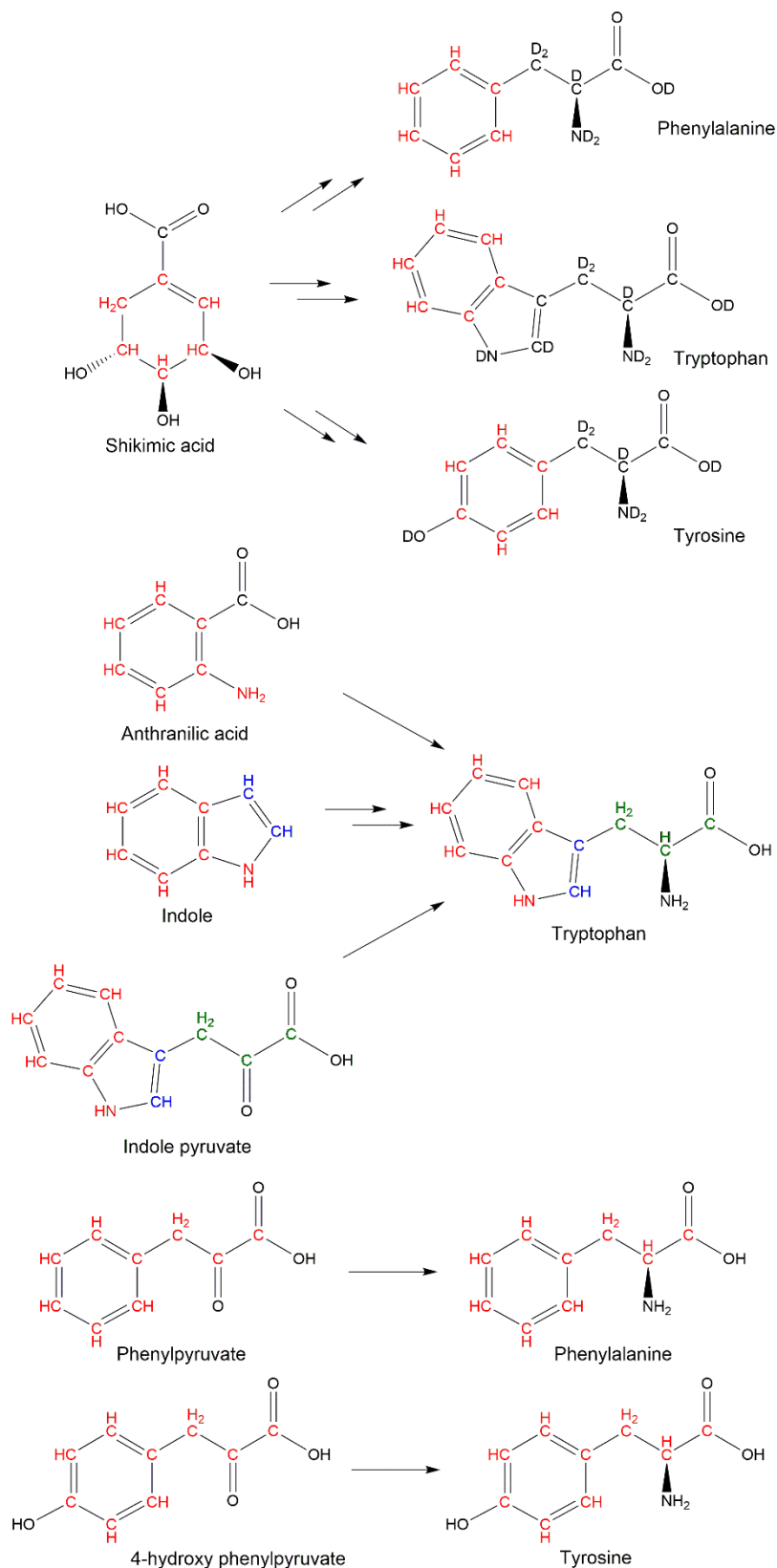


Figure 5.2. Metabolic precursors used for isotopic labelling and reverse labelling of phenylalanine, tryptophan, and tyrosine. Sites corresponding to labelled or unlabelled groups are coloured to show their starting and end positions.

Histidine can be labelled or reverse labelled by the addition of the amino acid itself to the expression media. Histidine can also be labelled, without scrambling, by the metabolic precursor imidazolepyruvate (Figure 5.3) [251].

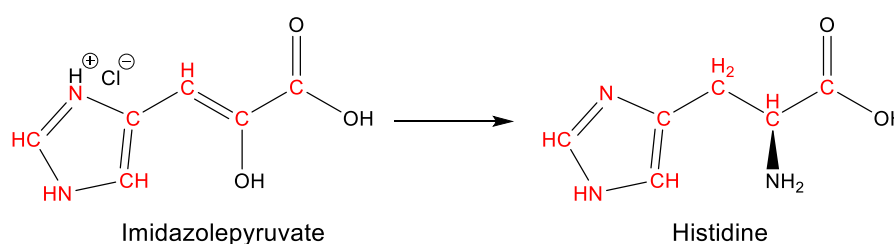


Figure 5.3. Histidine (un)labelling by the metabolic precursor Imidazolepyruvate with incorporated atoms shown in red.

#### 5.1.4. Routes for Targeting Polar Residues

Serine is connected to glycine *via* a glycine-hydroxymethyltransferase, which is in turn linked to threonine. Serine is also a precursor of tryptophan and cysteine biosynthesis and can be converted into pyruvate by serine dehydratase. Currently, there is no protocol for scramble-free specific labelling or reverse labelling of serine using traditional expression hosts. Similarly, cysteine is converted to pyruvate by cysteine desulfhydrases which leads to significant scrambling when cysteine is added to the culture medium.

Threonine is connected to glycine, serine, cysteine, tryptophan and isoleucine, which leads to significant scrambling for nitrogen labelling or reverse labelling [252]. For labelling or reverse labelling of carbon sites, threonine is connected to isoleucine and glycine biosynthesis and can cause significant scrambling when added to the expression media. This scrambling effect has been overcome by the addition of the isoleucine precursor 2-ketobutyrate (or isoleucine) and glycine to the expression media [252, 253].

Asparagine and glutamine are particularly difficult amino acids to specifically label or reverse label. Specific  $^{15}\text{N}$  labelling of these amino acids has been achieved through the use of media supplemented with all  $^{14}\text{N}$ -amino acids apart from asparagine or glutamine and  $^{15}\text{NH}_4\text{Cl}$  [254]. Specific labelling of side chain sites of these residues using metabolic precursors has not been possible due to their position in metabolic pathways. Asparagine and glutamine synthesis is closely linked to aspartate and glutamate synthesis, which are used in the synthesis of many amino acids. In addition, glutamate is the primary nitrogen donor in amino acid biosynthesis. This means that scramble-free specific labelling of these amino acids with the amino acids themselves or amino acid precursors is not possible without addition of a full amino acid complement to the media or the use of auxotrophic strains or cell free systems [245, 255].

#### 5.1.5. Routes for Targeting Charged Residues

The final steps of both lysine and arginine biosynthesis are irreversible, which means that both amino acids can be used directly for labelling and reverse labelling, thus negating the need for supplemental precursors to reduce isotopic scrambling [221].

For reasons discussed above, scrambling free specific labelling of aspartate and glutamate either with the amino acids themselves or with metabolic precursors has not been achieved.

#### 5.1.6. Special Cases

Methionine is commonly used to introduce methyl labelled probes for NMR analyses of proteins [256-258]. The relatively low abundance of methionine (2.4%) can reduce the chance of spectral overlap than with aliphatic residues [225]. Methionine can be both isotopically labelled and reverse labelled with minimal scrambling by addition of the amino acid itself to the media [221, 259]. An alternative approach uses the metabolic precursor methylthio-2-oxobutanoate (Figure 5.4) for labelling without nitrogen [260].

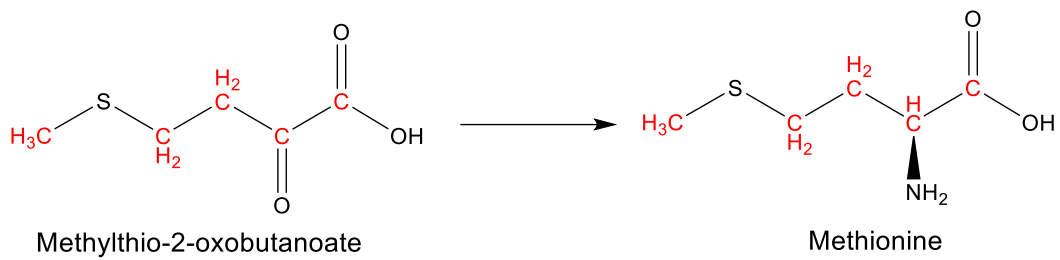


Figure 5.4. Methionine (un)labelling by methylthio-2-oxobutanoate with isotopically labelled sites indicated in red.

Glycine is linked to serine and threonine by glycine-hydroxymethyltransferase and threonine aldolase respectively, meaning extensive scrambling for both carbon and nitrogen sites occurs when attempting to label or reverse label with glycine.

Proline can be used as a sole carbon and nitrogen source in bacterial cell culture and so induces significant scrambling when added to the media [220, 221, 245, 261]. Metabolic precursors of proline have not been used for the production of proteins with proline labelling or reverse labelling. The interconnectivity of the proline biosynthetic pathway makes this unlikely.



Table 5.1. A summary of the scramble free isotope labelling strategies used in *E. coli* expression of recombinant proteins.

Amino Acid	Labelling compound
Alanine	1-[ <sup>13</sup> C]-alanine [240]
Arginine	Arginine [221, 262]
Aspartic acid	-
Asparagine	-
Cystine	-
Glutamic acid	-
Glutamine	-
Glycine	-
Histidine	Histidine [221, 262], imidazolepyruvate [251]
Isoleucine	2-ketobutyrate [161], 2-hydroxy-2-ethyl-3-ketobutyrate [236, 237]
Leucine	α-ketoisovalerate (L/V) [231, 232], acetolactate (L/V) [233], 2-ketoisocaproate [235]
Lysine	Lysine [221, 262]
Methionine	Methionine [221, 262], methylthio-2-oxobutanoate [260]
Phenylalanine	Phenylpyruvate [220, 247], , shikimic acid (F/Y/W) [246]
Proline	-
Serine	-
Threonine	Threonine with 2-ketobuterate/isoleucine and glycine [252, 253]
Tryptophan	Indole [248], indole pyruvate [249], anthranilic acid [250], shikimic acid (F/Y/W) [246]
Tyrosine	4-hydroxy phenylpyruvate [220, 247], shikimic acid (F/Y/W) [246]
Valine	α-ketoisovalerate (L/V) [231, 232], acetolactate (L/V) [233]

### 5.1.7. How Specific Isotope Unlabelling can Aid in Structural Studies of Proteins by Solution NMR Spectroscopy

Protein structure determination by solution NMR spectroscopy is generally achieved by applying experimentally derived structural restraints to the protein during folding by molecular dynamics simulations. Typically, the majority of these structural restraints are derived from the NOE (Nuclear Overhauser Effect) [263]. The NOE is caused by the relaxation of one nucleus by another nucleus due to dipole-dipole interactions. This effect can be measured by NOESY experiments, typically between  $^1\text{H}$  for proteins, and gives a through space correlation. The intensity of the NOE is proportional to the distance between the two interacting nuclei and so can be used to generate approximate distance restraints between hydrogen atoms in the protein, typically separated by  $<6 \text{ \AA}$ . These distance restraints make up the majority of restraints used in protein structure determination by NMR spectroscopy.

The H-H NOESY experiments used to generate distance restraints, give a signal for every interaction between one hydrogen and another hydrogen close in space (typically  $< 6 \text{ \AA}$ ). This leads to a large number of observable signals and thus a large degree of signal overlap which can cause difficulty in assigning the NOEs. The signal overlap can be somewhat alleviated by separating signals across extra dimensions (e.g. 3D or 4D datasets) through the use of X-edited pulse sequences. However even higher dimensionality doesn't always resolve issues with signal overlap.

Specific isotope unlabelling can be used in combination with filtered and/or edited NOESY experiments to both reduce the number of observed NOEs and to provide residue type information that inform NOE assignments (Figure 5.5). Filtered/edited experiments can select for (edit) or remove (filter) signals resulting from hydrogens directly bonded to  $^{13}\text{C}$  or  $^{15}\text{N}$  in either the F1 or F2 dimensions [176]. These experiments allow for specific observation of  $^1\text{H}$ - $^1\text{H}$  interactions in which one or both sites are unlabelled. Filtered/edited experiments have been used to great effect to identify intermolecular NOEs in protein:ligand or protein:protein complexes where the interacting partners are differently labelled [264-267]. Here, we combine

filtered/edited NOESY experiments with specific isotope unlabelling to look at intramolecular NOEs.

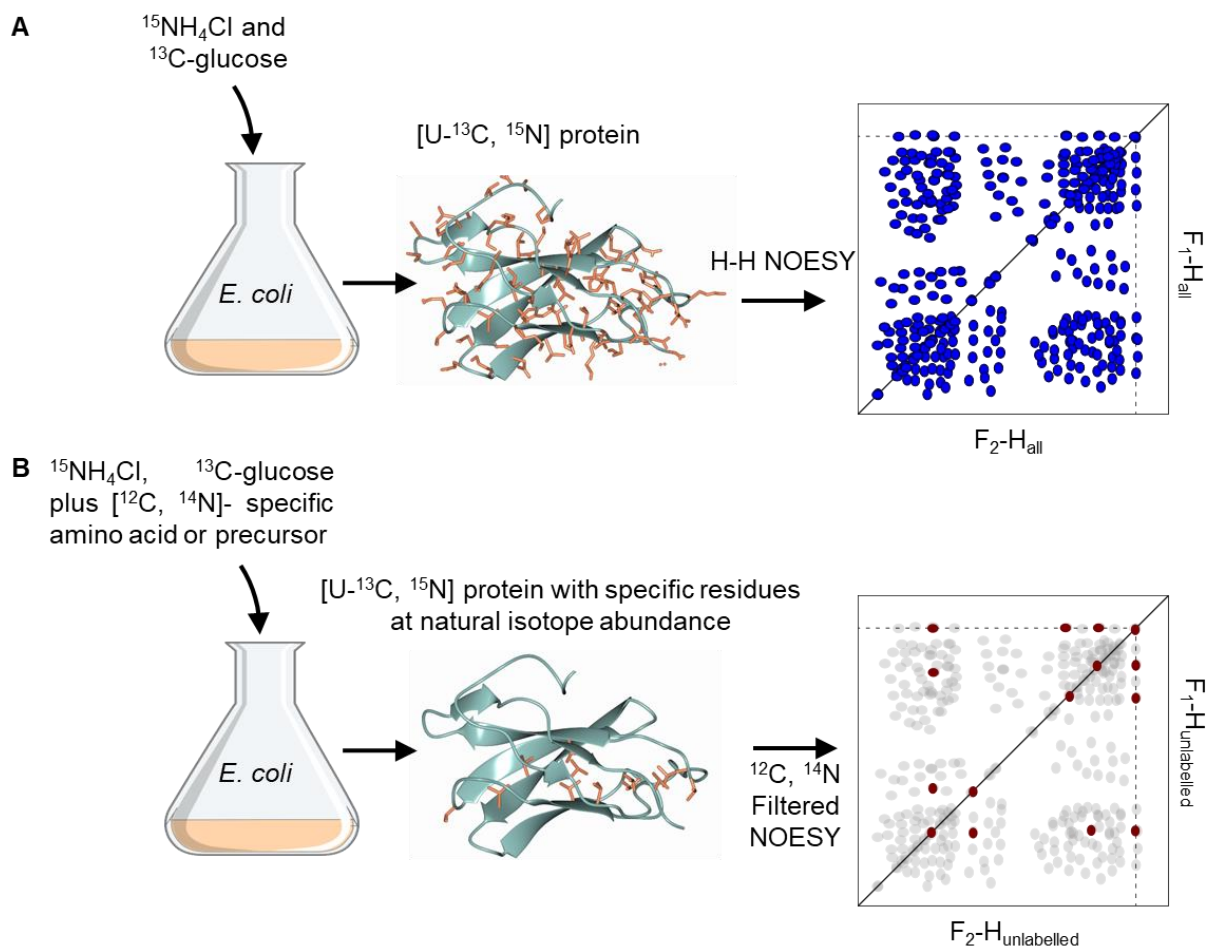


Figure 5.5. Schematic representation of **A.** standard  $[\text{U}-^{13}\text{C}, ^{15}\text{N}]$  labelling of a protein and collection of severely overlapping 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY spectrum **B.** The proposed unlabelling and 2D filtered NOESY approach to generate a simplified NOESY spectrum.

The combination of residue specific isotope unlabelling and 2D filtered/edited NOESY experiments to detect subsets of intramolecular NOEs was first employed by Vuister *et al.* (1994) to investigate phenylalanine residues in *Drosophila* heat shock factor [230]. Here, phenylalanine residues were unlabelled to alleviate problems associated with poor chemical shift dispersion and  $^{13}\text{C}$  induced line broadening. In this case phenylalanine was unlabelled by direct addition of the amino acid to the

culture with the authors reporting >90% isotope incorporation and minimal scrambling. However, later studies have demonstrated extensive isotope scrambling when attempting to label with phenylalanine [268]. This discrepancy may be due to differences in isotopic unlabelling compared to labelling or advancements in instrumentation and methodology for determining isotope incorporation between the two studies. Vuister *et al.* (1994) then combined the phenylalanine unlabelled samples with 2D filtered/edited NOESY spectra, which allowed identification of additional NOE restraints that had proved otherwise impossible to obtain due to the poor  $^{13}\text{C}$  signal dispersion in the aromatic region of the spectrum [230]. This approach yielded an additional 109 long-range NOEs and resulted in an improvement of RMSD from 1.4 to 0.87 Å. This study demonstrated the usefulness of this method for extracting additional structural restraints that are otherwise difficult to obtain. The Vuister *et al.* (1994) study is the only published use of this approach that could be found [172]. A number of important technical advances have been made since the study by Vuister *et al.* (1994) that warrant revisiting the use of specific isotope unlabelling with 2D filtered/edited NOESY experiments, these include: 1) The development of cryogenic probes for NMR spectrometers greatly reduced the concentration of protein required for sufficient signal strength, allowing for the use of smaller scale expressions which saves on the cost of specific (un)labelling reagents [269], and 2) the availability of new isotope (un)labelling methods/reagents expands the range of sites accessible to specific isotope unlabelling. Assessment of the value of different unlabelling schemes for use with filtered/edited NOESY experiments will aid in the design of future experiments of this type. Additionally, modern NMR structure building pipelines have been developed that are able to automatically assign NOEs and handle ambiguous data sets [185]. It was hypothesised that combining the residue type information obtained from these experiments with current iterative NOE assignment and structure building pipelines could prove useful for aiding correct early NOE assignments and thus improving structure convergence.

A similar approach was also applied to great effect by Peterson *et al.* (2004) to aid in the structural study of RNA and RNA-protein complexes [270]. Here RNA was produced, selectively labelled by nucleotide type, and filtered/edited NOESY spectra were used to both resolve overlapping peaks and provide base type assignment

information. The authors proposed filtered/edited NOESY experiments as a method of sequential assignment of RNA and found the improvement in data quality in using these experiments over 3D NOESYs was key to determining the conformation of their RNA of interest.

A range of different filtered/edited experiments can be collected to provide different residue type information (Figure 5.6), which would be useful in the assignment of NOEs. The objective here was to evaluate whether this data can be used to inform automatic NOE assignment and structure building pipelines such as ARIA in order to improve early NOE assignment and thus aid the structure calculation.

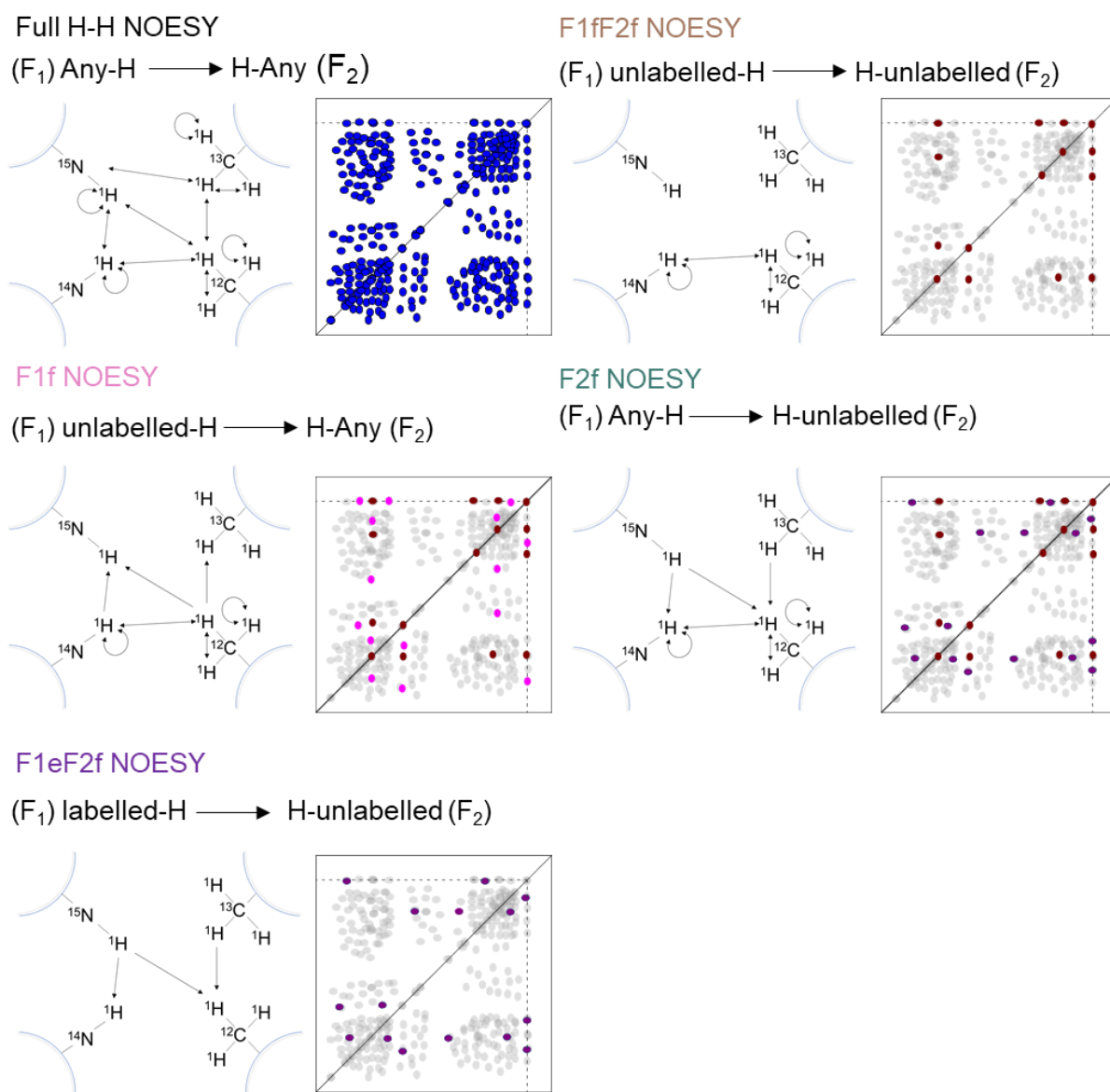


Figure 5.6. Descriptions of the transfer pathways of the various 2D filtered/edited NOESY experiments used with schematic representation of the anticipated patterns of peaks for each experiment. Peaks common across the different experiments are demonstrated by colour.

In order to investigate the potential benefits of this technique model proteins were used that were known to express well, have high stability and give good NMR spectra. This avoids potential protein specific problems that may give a false indication of the strengths or weaknesses of the technique. The biological role of the

proteins was not of importance for this study. A number of suitable candidates were available locally: 1) SHIRT-R3, a 9 kDa SHIRT domain from the bacterial protein Sgo\_0707 (residues 621-705), though a lack of sidechain assignments would limit the level of analysis that could be done [271], and 2) SasYr, a 9 kDa domain from the bacterial protein SasY (residues 23-111). Both backbone and side chain assignments were available for SasYr, which would allow for a better exploration into the potential of these methods and how the 2D filtered/edited NOESY data could best be implemented into structure calculation pipelines.

## 5.2. Results and Discussion

### 5.2.1. Specific Isotope Unlabelling of SHIRT-R3

In order to investigate the potential of specific isotope unlabelling and 2D filtered/edited NOESY experiments to aid protein structure elucidation, a model protein was selected that was known to express in high yields (~30 mg/L in LB), was highly stable at room temperature, and gave well dispersed NMR spectra with strong signals. The third SHIRT domain from Sgo\_0707 (SHIRT-R3) was first selected as it fit these parameters [271]. Using a well-behaved model protein limited sample specific problems that may have complicated the method development.

Unlabelling schemes to be used in combination with SHIRT-R3 were designed based on the distribution of residues within the sequence of the protein (Table 5.2) as well as the ease with which they could be unlabelled (as outlined in sections 5.1.2-5.1.6). Additionally, aromatic residues were targeted due to their general prevalence in the cores of proteins which could provide more relevant structural restraints. Seven unlabelling patterns were chosen, LV, K, F, Y, W, FK and HL. The inclusion single residue type and double residue type unlabeling patterns should provide information on the potential advantages of expanding the unlabelling scheme, although this will be highly protein dependent.





Table 5.2. SHIRT-R3 sequence and amino acid composition used to inform unlabelling schemes

SHIRT-R3		sequence:	
APTYKATHEFMSGTPGKELPQEVKDLLPADQTDLKDGSQATPTQPSKTEVKTAEGT			
WSFKSYDKTSETINGADAHFVGTWEFTPA			
Ala	8	Leu	4
Arg	0	Lys	8
Asn	1	Met	1
Asp	6	Phe	4
Cys	0	Pro	7
Gln	4	Ser	6
Glu	7	Thr	13
Gly	6	Trp	2
His	2	Tyr	2
Ile	1	Val	3

The unlabelled samples were expressed and purified, with the majority of samples expressing well (25 mg/L). The tryptophan unlabelled sample showed a decreased expression level of 15 mg/L. This seemed to be caused by an effect of the indole on the growth of the *E. coli* as the culture did not reach the same OD<sub>600</sub> as the other samples. Isotope incorporation could be seen by peaks disappearing from the 2D (<sup>1</sup>H, <sup>15</sup>N) and 2D (<sup>1</sup>H, <sup>13</sup>C) HSQC spectra when compared to the fully labelled sample (Figure 5.7).

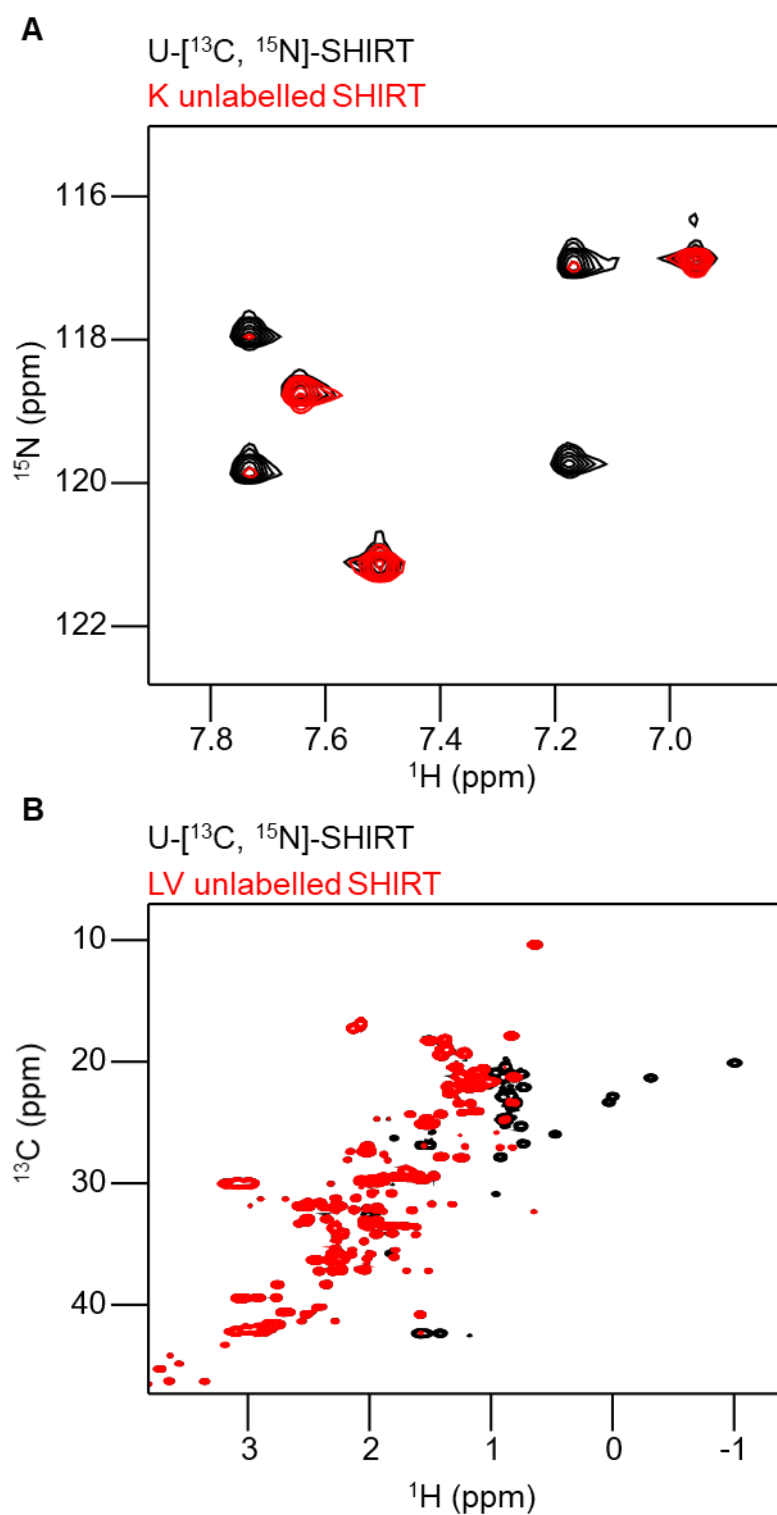


Figure 5.7. **A.**  $^{15}\text{N}$ -HSQC spectra of  $[\text{U-}^{13}\text{C}, ^{15}\text{N}]$ -SHIRT-R3 in black with K unlabelled SHIRT-R3 in red. **B.**  $^{13}\text{C}$ -HSQC spectra of  $[\text{U-}^{13}\text{C}, ^{15}\text{N}]$ -SHIRT-R3 in black with LV unlabelled SHIRT-R3 in red.

Isotope incorporation could be quantified for nitrogen unlabelling by comparing signal intensities from the unlabelled and fully labelled proteins (Figure 5.8). This was possible due to the available backbone assignments. Nitrogen unlabelling was seen to be >90% for all samples although some residual signal from lysine remained (~8 %).

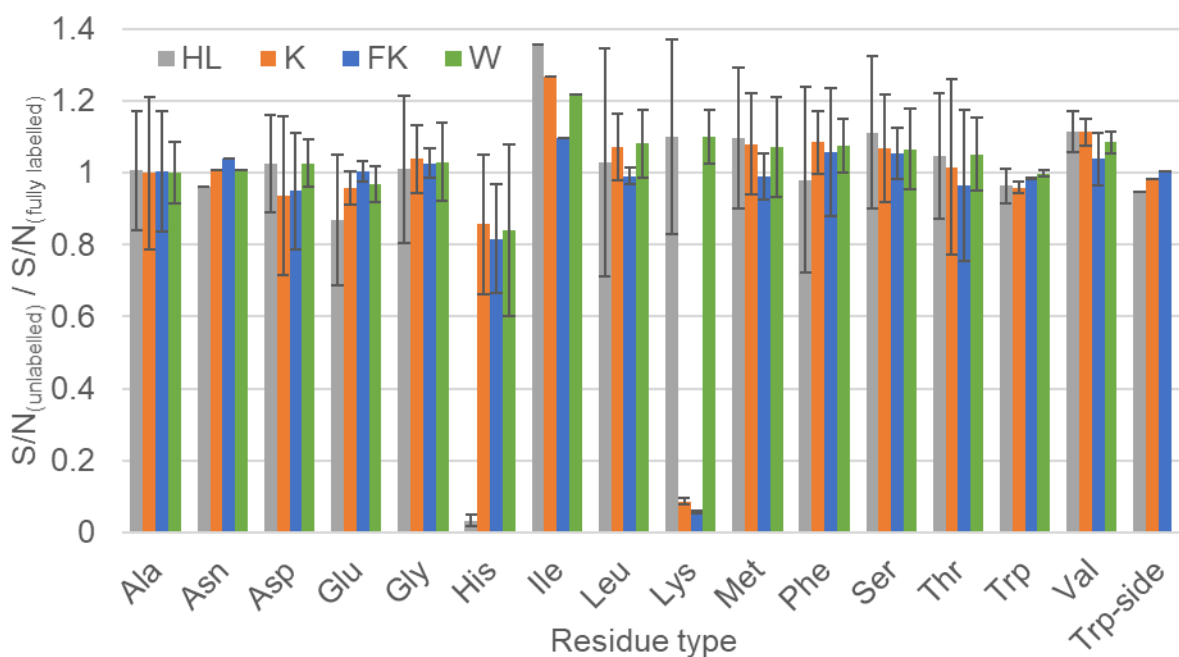


Figure 5.8.  $^{15}\text{N}$  incorporation for SHIRT-R3 HL, K, FK and W unlabelled samples. Incorporation is averaged over all residues of that type and error bars show the standard deviation. Other unlabelling patterns are not included as they do not unlabel nitrogen. Leucine and phenylalanine unlabelling do not unlabel the backbone amide and so are not observed here. Tryptophan is unlabelled on the sidechain amine but not the backbone amide.

Although peaks could be seen disappearing from the  $^{13}\text{C}$ -HSQC spectra (Figure 5.7B), the lack of available side chain assignments meant that it could not be confirmed which residues these peaks belonged to. This means that we would not be able to identify any isotope scrambling or differences in incorporation across the sequence. For this reason,  $^{12}\text{C}$  incorporation could not be effectively quantified for these samples without using an alternative technique such as mass spectrometry.

Incorporation of the backbone carbonyl carbon could be assessed using a HNCO experiment although not all of the labelling precursors unlabel the carbonyl.

2D F1f, 2D F1fF2f and full 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY spectra were collected for each labelling scheme. The resulting filtered spectra were greatly simplified compared to the full 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY, with much less signal overlap and peaks appearing in the expected locations based on the expected chemical shifts (Figure 5.9). Nitrogen bound  $^1\text{H}$  signals are only observed in the 2D F1fF2f spectrum for labelling schemes that incorporate  $^{14}\text{N}$  and combinatorial unlabelling of nitrogen unlabelled residues with aliphatic residues shows additional cross peaks to  $^1\text{H}$  bound to aliphatic carbons. This can be seen by comparing the spectra of lysine, leucine/valine and lysine/leucine/valine unlabelled samples (Appendix figure 7.12). Unlabelling of aromatic residues led to NOEs in the aromatic region of the 2D F1fF2f spectrum that are not observed in the 2D F1fF2f non-aromatic labelling schemes (Appendix figure 7.12). The 2D F1f spectrum was seen to contain all of the peaks present in the F1fF2f spectrum plus additional peaks (Figure 5.9), presumably resulting from unlabelled to labelled  $^1\text{H}$  interactions.

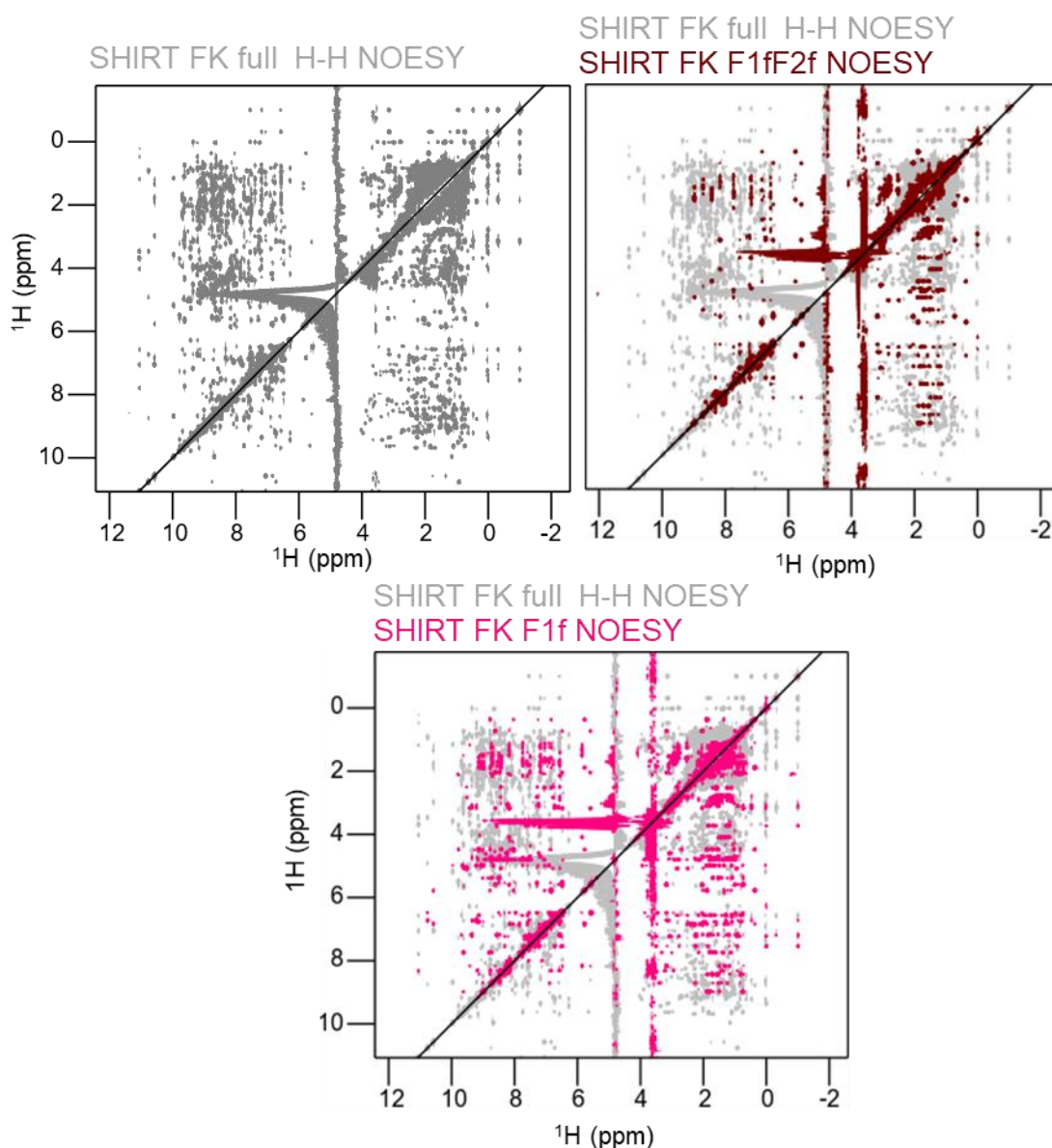


Figure 5.9. F1fF2f and F1f NOESY spectra for FK unlabelled SHIRT-R3 in brown and pink respectively compared to a full 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY spectrum in grey.

Initially it was expected that side chain assignments for SHIRT-R3 would have been generated by another researcher during the course of this investigation. However, assignment of the side chain residues proved more challenging than anticipated. Without side chain assignments, it was not feasible to use the NOESY spectra for structural analysis and so an alternative model protein, with sidechain assignments available, was chosen.

### 5.2.2. Specific Isotope Unlabelling of SasYr

SasYr has previously been fully assigned by Dr Rachael Cooper and so along with the other factors mentioned for SHIRT-R3 made a good model protein to use for the development of these methods.

As with SHIRT-R3, unlabelling patterns were planned based on the distribution of residues within the sequence (Table 5.3) as well as the ease of unlabelling. SasYr has fewer aromatic residues than SHIRT-R3 which reduced the number of unlabelling patterns that seemed sensible to test. Four unlabelling patterns were selected: K, LV, LVK and LVW.

Table 5.3. SasYr sequence and amino acid composition used to inform unlabelling schemes.

SasYr sequence: GPAMTTAPEAPSVNDTEVGSKKVSGKGHEVGNTVTVTFPDGKTATSKVDEKG NWTVDVPEGTELKVGNEITATETDMSGNKSESGKGKVTD			
Ala	5	Leu	1
Arg	0	Lys	10
Asn	5	Met	2
Asp	6	Phe	1
Cys	0	Pro	5
Gln	0	Ser	7
Glu	9	Thr	14
Gly	12	Trp	1
His	1	Tyr	0
Ile	1	Val	11

Isotope incorporation was quantified (Figure 5.10) and showed that all unlabelled residues were incorporated to >90%. Interestingly, despite doubling the amount of lysine added to the media there was still some residual <sup>15</sup>N, <sup>13</sup>C lysine present.

Tryptophan appeared to retain ~8%  $^{13}\text{C}$  labelling although almost no  $^{15}\text{N}$  labelling. This is likely an error in measuring the  $^{13}\text{C}$  incorporation due to the small number of peaks present in the aromatic focussed 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC.

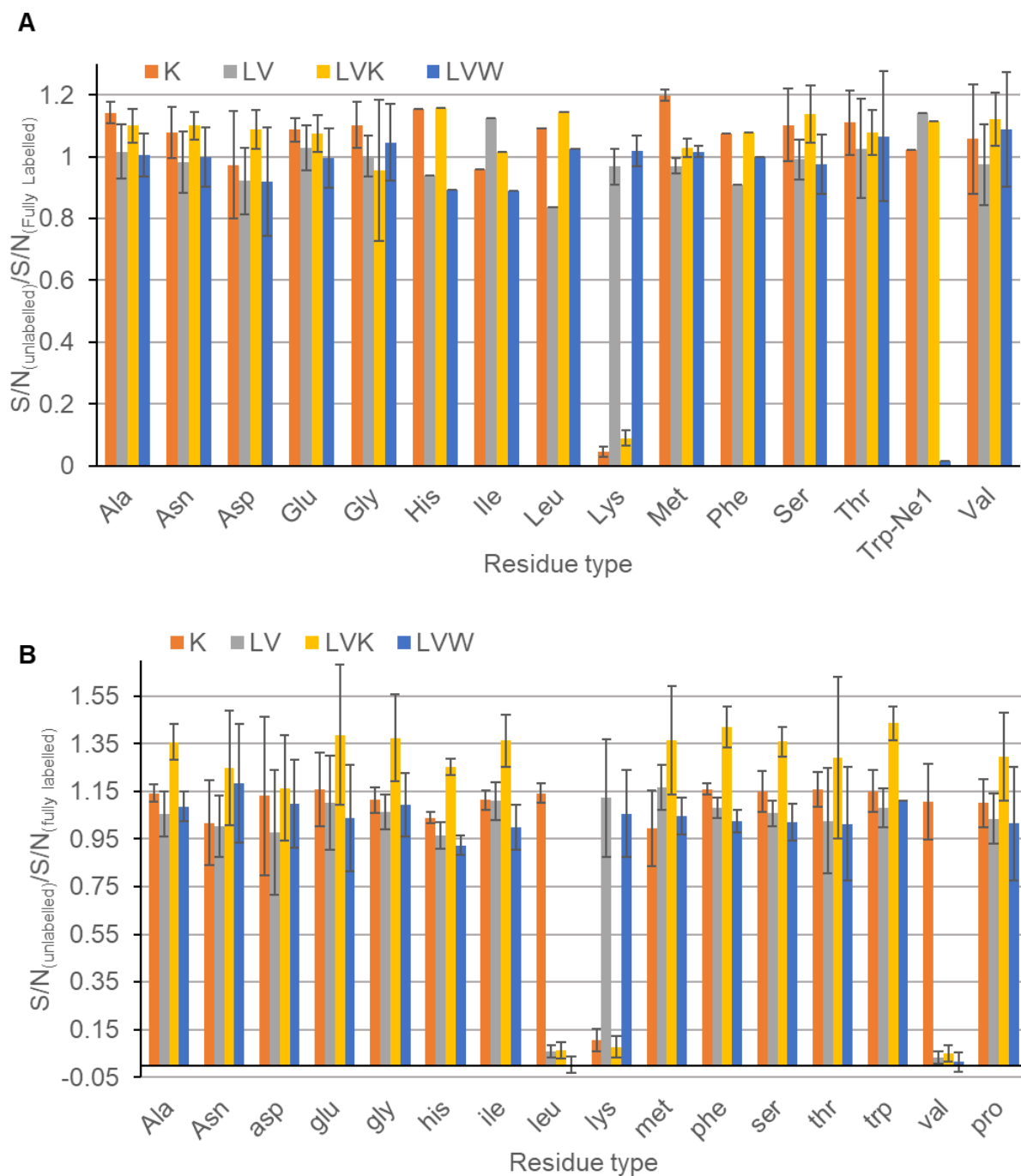


Figure 5.10. **A.**  $^{15}\text{N}$  and **B.** aliphatic  $^{13}\text{C}$  isotope incorporation for SasYr unlabelled samples. Tryptophan is only unlabelled at aromatic sites and so unlabelling is not shown in this analysis. The proportion of unlabelling is averaged over residue types and the standard deviation is shown as error bars. Leucine, valine and tryptophan unlabelling does not target backbone atoms and so leucine and valine unlabelling is not seen in the  $^{15}\text{N}$  analysis.



For each unlabelling pattern four different 2D filtered/edited NOESY experiments were collected: F1 filtered (F1f), F2 filtered (F2f), F1/F2 filtered (F1fF2f) and F1 edited F2 filtered (F1eF2f). As described in Figure 5.6, each of these experiments provides different residues type information to aid in assignment. The resulting spectra are greatly simplified compared to the full 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY and allow for identification of peaks from severely overlapping regions in the full spectrum (Figure 5.11).

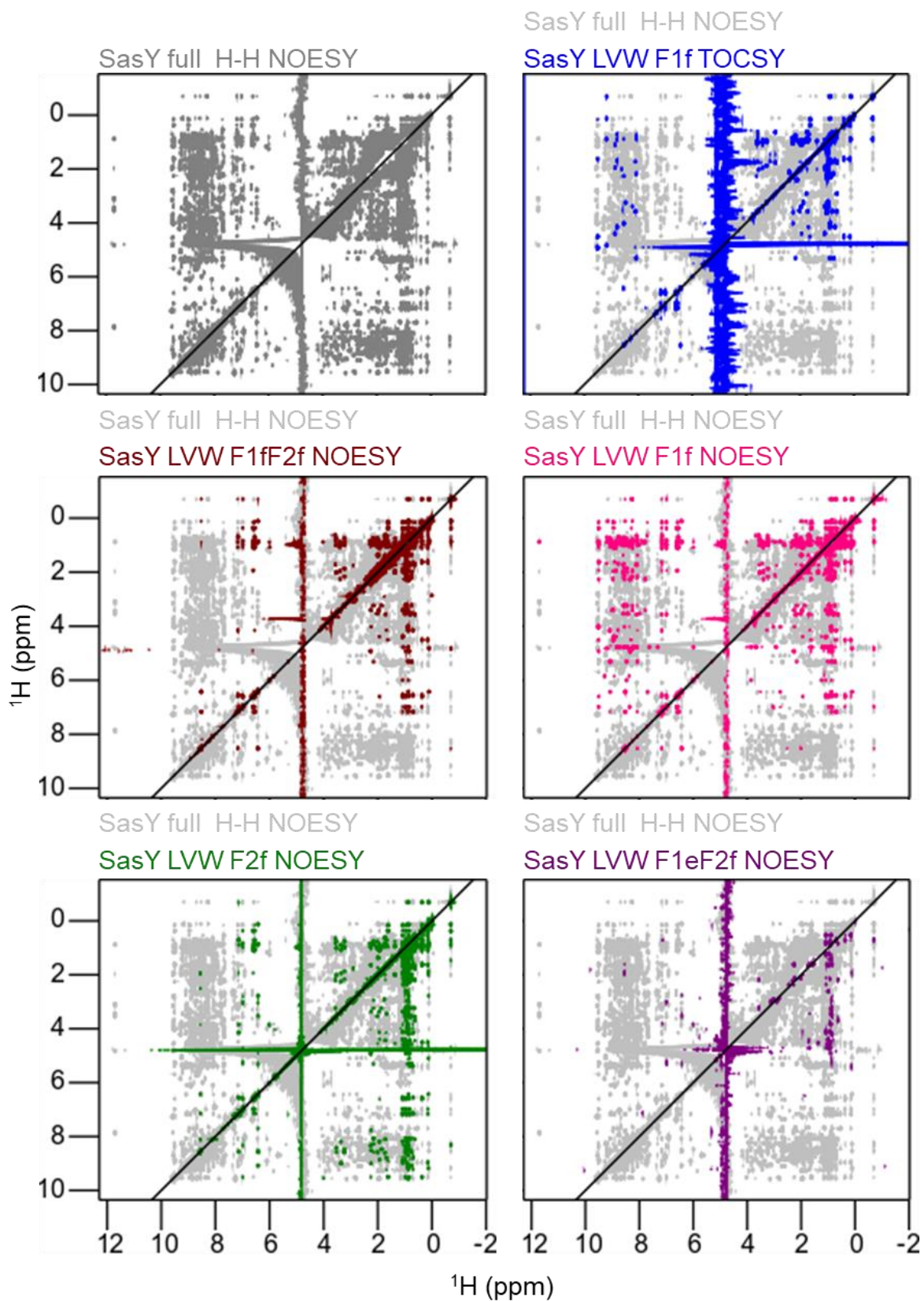
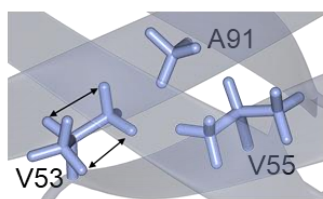


Figure 5.11. Examples of 2D filtered/edited NOESY spectra for the LVW unlabelled SasYr sample compared to a full 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY spectrum in grey.

Some peaks in the amide region of the LVW SasYr F1f TOCSY spectrum are not observed in the F1fF2f NOESY spectrum (Figure 5.11). This is due to the lack of backbone amide labelling in this region, meaning intra-residue sidechain to backbone interactions are filtered out in the F1fF2f NOESY but not in the F1f TOCSY experiment. A number of peaks in the spectra were easily assignable and demonstrated that the experiments were working as intended (Figure 5.12). This also demonstrates how the amino acid type information is encoded into the experiments.

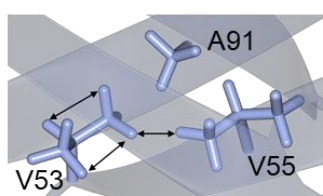
## F1f TOCSY

(F<sub>1</sub>) unlabelled-H → H-intraresidue (F<sub>2</sub>)



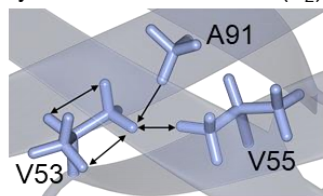
## F1fF2f NOESY

(F<sub>1</sub>) unlabelled-H → H-unlabelled (F<sub>2</sub>)



## F2f NOESY

(F<sub>1</sub>) Any-H → H-unlabelled (F<sub>2</sub>)



## F1eF2f NOESY

(F<sub>1</sub>) labelled-H → H-unlabelled (F<sub>2</sub>)

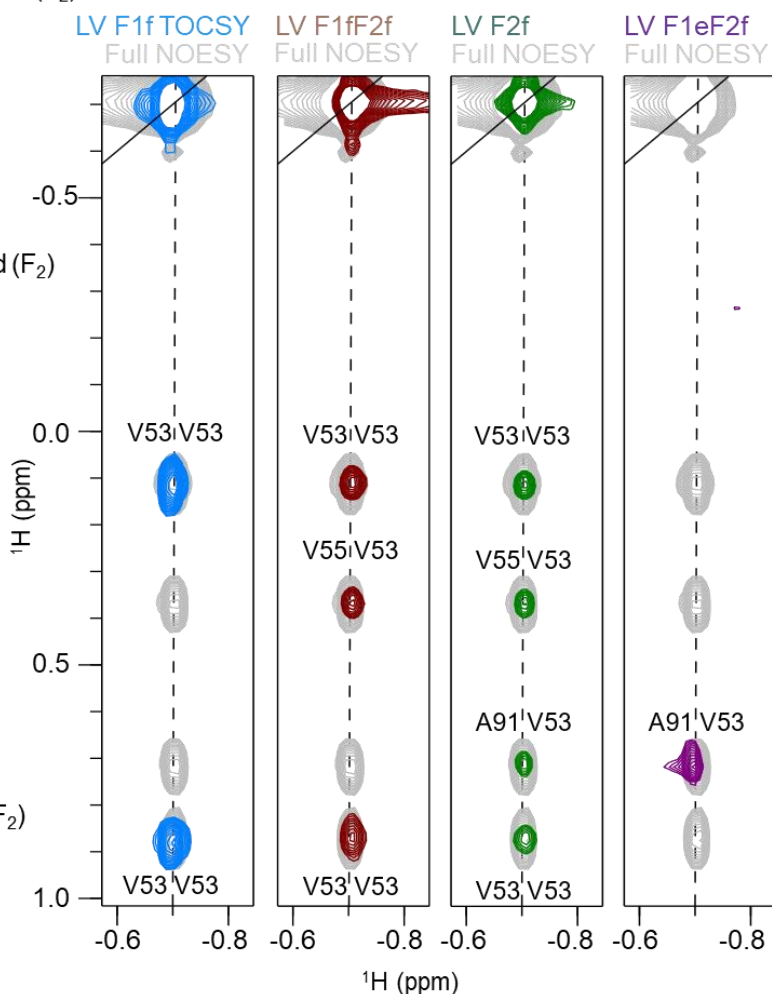
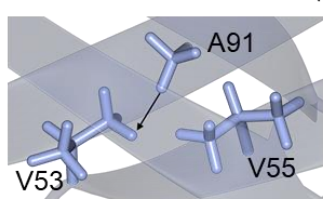


Figure 5.12. Description of the transfer pathway for the different filtered/edited TOCSY and NOESY experiments with a focus on equivalent strips from each spectrum to demonstrate the experiments are working as intended. The interactions are also shown on the structure of SasYr.

### 5.2.3. Specific Isotope Unlabelling to Complement 3D Structure Calculations

With all of the filtered/edited spectra obtained, the potential of using these spectra to aid in structure calculations could be explored. In order to incorporate the amino acid type information into the structure calculations, peaks were first ambiguously assigned automatically in CCPN, using the known potential residue types to limit the assignment possibilities (i.e. for LV unlabelled F1fF2f experiments possible

assignments were limited to only leucine or valine residues in both dimensions). The ambiguously assigned peak lists could then be used as inputs for structure calculations by ARIA. Examples of the ambiguous restraints generated are given in Appendix Figures 7.20-7.21.

In order to explore the potential of each type of filtered/edited NOESY experiment, the LVW unlabelled sample was used as a test case. Structure calculations were run with each filtered/edited spectrum added to complement data from a simultaneous 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC (which allows for simultaneous evolution of  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts in F2) (Table 5.4). It was observed that each individual experiment other than the F1eF2f experiment was able to both identify additional NOE restraints compared to standard calculation. It was also observed that these additional restraints aided structure calculation and resulted in a more compact structure ensemble (lower RMSDs). The 2D F1eF2f experiment appeared to impair the structure calculation, leading to fewer NOE restraints and higher RMSDs. This could occur if the poor quality of the F1eF2f experiments impaired the automated assignments of the 3D dataset, leading to more peaks in the 3D dataset being classified as violations. This could be caused by the lower sensitivity of the edited experiment which resulted in worse quality data than for the other experiments. The poor data quality may have led to mis-assigned peaks or could have negatively impacted the calibration of the NOE restraints and thus the rest of the calculation. It may be that running the 2D F1eF2f experiment with more scans would have resulted in a similar structure improvement to the other experiments. The other filtered/edited experiments all performed similarly to one another with the 2D F1f spectra performing slightly better than the 2D F2f spectra. This is likely due to the sensitivity improvement of having the more sensitive  $^{12}\text{C}$ - $^1\text{H}$  collected in the indirect dimension.

Table 5.4. Statistics from structure calculations of SasYr with different 2D filtered/edited NOESY spectra added. The filtered/edited data was used to complement data from a 3D (<sup>1</sup>H, <sup>1</sup>H) NOESY (<sup>13</sup>C/<sup>15</sup>N) HSQC experiment.

Filter/Edit spectra used	Number of restraints	Number of violations	Backbone RMSD	Heavy atom RMSD	Intra-residue restraints	Sequential restraints ( i-j =1)	Mid range restraints (1< i-j <5)	Long range restraints ( i-j ≥5)	G-factor
None	1834	100	0.52	1.22	592	380	129	597	-0.25
LVW F1fF2f	1870	86	0.36	0.97	595	372	125	641	-0.249
LVW F1f	1898	99	0.31	0.93	599	373	144	648	-0.203
LVW F2f	1896	81	0.42	1.03	605	374	135	630	-0.138
LVW F1eF2f	1816	129	1.25	1.95	594	380	131	568	-0.459
All LVW	1990	111	0.88	1.41	622	380	120	683	-0.258
Informed LVW F2f	1898	82	0.35	0.98	598	371	133	640	0.008

Adding all of the unlabelled experiments into the structure calculation resulted in a much higher number of restraints but also worse RMSD and G-factors. The G-factor reports on how the  $\Phi$ - $\Psi$ ,  $X_1$ - $X_2$  and  $X_1$  angles in a structure compare to the distribution of these angles for each amino acid type from a set of representative structures. This provides a measure of how probable a given conformation is, with higher values suggesting a more probable conformation. A G-factor  $>0$  suggests a highly probable conformation, G-factors between 0 and -0.5 are acceptable and a G-factor of  $<-0.5$  suggests an improbable conformation [272]. This suggests that many of the newly identified restraints are wrongly assigned or calibrated and are impairing the structure calculations. This could be due to increasing the error in the restraint calibration through addition of the poor quality F1eF2f data. An improved method for incorporating this data was therefore needed.

It was observed that the 2D F2f spectrum should contain all the same peaks as the combination of the 2D F1fF2f spectrum and the 2D F1eF2f spectrum (Figure 5.6). It was therefore possible to combine the information from all three of these experiments in the assignment of the 2D F2f spectrum. Ambiguous peak assignments could be generated by first producing ambiguous distance restraints (by chemical shift matching) (Appendix figure 7.20), limited to the residues defined by the unlabelling and filter/edited experiments. These distance restraints could then be used to generate ambiguous peak assignments (Figure 5.13). Ambiguous assignments were first generated for the 2D F1fF2f and 2D F1eF2f spectra and were then transferred to the 2D F2f spectrum. Ambiguous assignments were then generated for any peaks remaining in the 2D F2f spectrum (Appendix Figure 7.21). Through this approach, it was possible to combine all residue type information into a single peak list. The 2D F1f spectra were not included in this analysis as this should contain the same information as the F2f spectra. The ambiguously assigned peak list was used as an input for structure calculations. This allowed for restraints to be generated by ARIA rather than being pre-calculated [187]. This approach is referred to as “Informed LVW F2f” in Table 5.4.

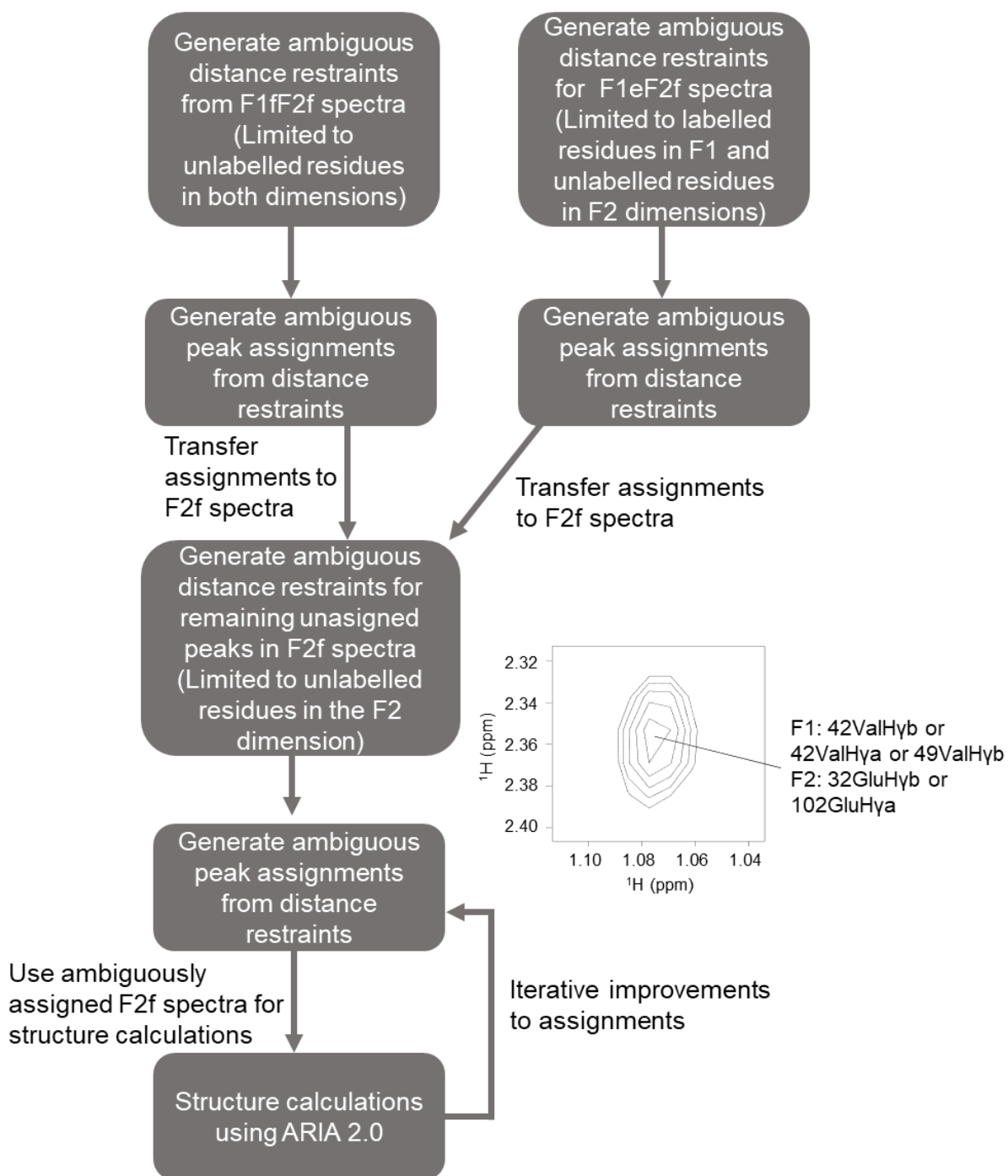


Figure 5.13. A flowchart demonstrating the process for incorporating residue type information obtained from filtered/edited NOESY experiments into a single peak list. An example of an ambiguously assigned peak from the LVW F2f spectra is also shown.

When using the informed LVW F2f approach, the resulting restraint numbers, violations and RMSD values are similar to those achieved with the half-filtered



experiments. However, the G-factor was significantly improved through this approach which indicated that the resulting structure are likely to be more accurate. This is most likely due to an improvement in the assignment of NOEs (leading to fewer misassigned NOEs) during structure building caused by the improved residue type assignment. Further experiments utilised this method of incorporating the filtered/edited data into structure calculations.

The effects of the different unlabelling patterns were then explored in order to inform any future work on the most effective unlabelling patterns to select. The results of structure calculations with the different unlabelling patterns using the informed assignment approach previously described are shown In Table 5.5. The structure calculations reported in Table 5.5 were run with a higher number of steps in the simulated annealing than the results in Table 5.4 to minimise the effects of variations between runs.

Table 5.5. Statistics from structure calculations of SasYr with data from different unlabelling schemes incorporated. The filtered/edited data was used to complement data from a 3D (<sup>1</sup>H, <sup>1</sup>H) NOESY (<sup>13</sup>C/<sup>15</sup>N) HSQC experiment.

Labelling scheme	Number of restraints	Number of violations	Backbone RMSD	Heavy atom RMSD	Intra-residue restraints	Sequential restraints ( i-j =1)	Mid range restraints (1< i-j <5)	Long range restraints ( i-j ≥5)	G-factor
Fully labelled	1834	100	0.52	1.22	592	380	129	597	-0.25
LV	1886	55	0.35	0.93	596	371	123	658	-0.159
K	1878	76	0.46	1.09	628	388	120	599	-0.223
LVK	1859	83	0.41	1.11	629	386	128	591	-0.281
LVW	1910	72	0.33	0.94	597	367	136	667	-0.103
All unlabelling	2008	109	0.35	1.00	663	379	126	672	-0.201

It was observed that each of the unlabelling patterns was able to identify additional restraints that led to an improved, more compact structural ensemble by all metrics. The most effective unlabelling pattern for SasYr was LVW, most likely due to tryptophan being buried in the core of the protein so that more long-range NOEs are observed, which are important for folding the protein correctly. Lysine appeared to be less useful than LV, which is predictable given that lysines are commonly surface exposed and flexible so will give rise to fewer structurally important NOEs. Interestingly the LVK sample performed worse than the LV sample. It is probable that the poorly distributed lysine shifts cause a higher degree of ambiguity in the initial assignments which then negatively impacts the structure calculation. For example, peaks in the LV unlabelled spectrum that had relatively few potential assignments may have many additional potential assignments to and from lysine residues in the LVK spectrum. For this reason, it is important take care to design a labelling scheme that targets residues predicted to be in the core of the protein as this would most likely have the greatest positive effect on structure calculations.

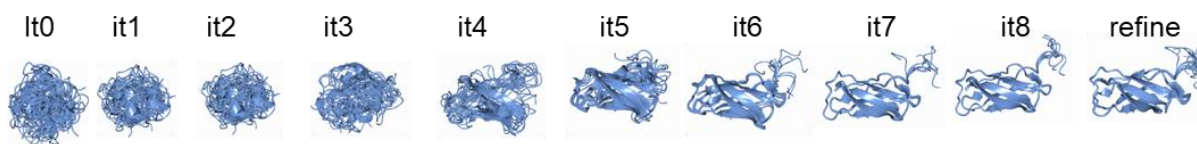
Adding all of the labelling patterns appeared to have an additive effect on the quality of the calculations, with more restraints than for the individual experiments. The limited improvement when compared to the LVW experiment alone however make it difficult to justify combining multiple unlabelling patterns given the effort required to produce multiple unlabelled samples. Using multiple unlabelling patterns may be more useful when fewer restraints can be generated from 3D or 4D data sets (e.g. disordered proteins suffering from signal overlap or unstable proteins that may not be usable with longer experiments).

#### 5.2.4. Does the addition of filtered/edited NOESY data improve the rate of structure convergence?

Initial results indicated that inclusion of filtered/edited data to structure calculations was able to improve how quickly the structure converged. This can be seen by

following the change in RMSD for structures calculated with and without filtered/edited experiments included (Figure 5.14).

#### No unlabelling data



#### With LVW unlabelled data

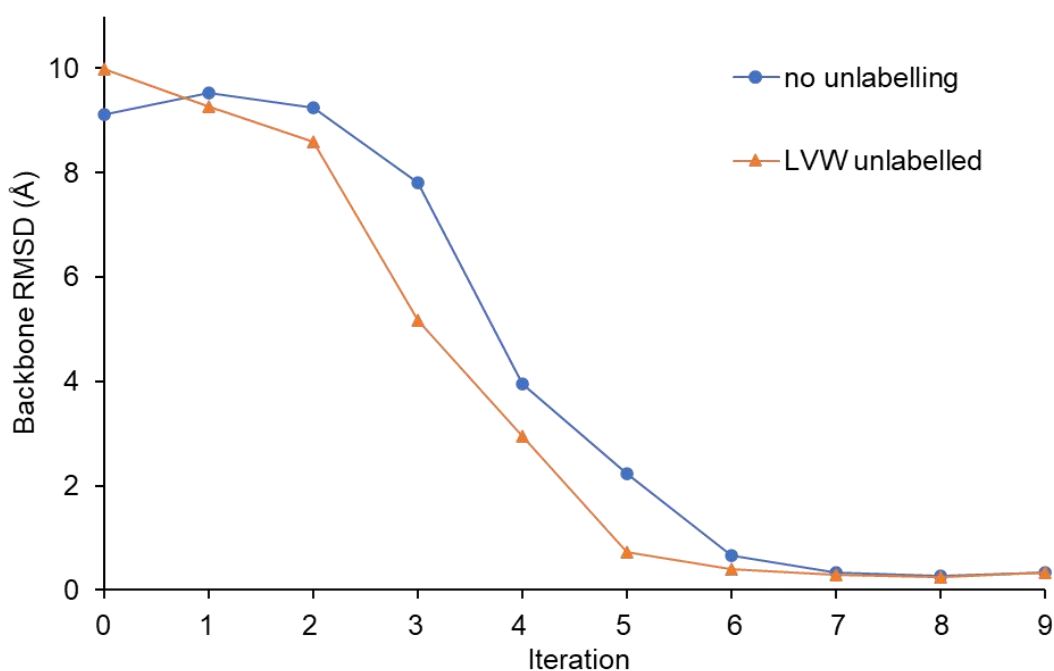


Figure 5.14. Analysis of the convergence of SasYr structure calculation with and without LVW unlabelled data. The structure ensemble from each iteration is shown and the backbone RMSD across the ensemble for residues 27-109 is plotted for each iteration. The standard calculation is shown in blue and the calculations with LVW data added are shown in orange.

It was hypothesised that this improvement in the rate of convergence may be caused by an improvement in the automated NOE assignments due to the addition of the

relatively small number of known assignments from the filtered/edited spectra. In order to investigate whether the differences in the rate of structure convergence were due to the unlabelling data, or due to inconsistencies in the structure calculations, the calculations were repeated five times with a different randomly selected 5% of the 3D NOESY data removed. This can account for variability in the structure calculations to better highlight differences caused by the addition of the filtered/edited data rather than the standard variation in the structure calculations or the effects of a small number of restraints. The results of this study (Figure 5.15) showed that the rate of convergence for these structures was not significantly affected by the addition of filtered/edited data. This suggests that the addition of the filtered/edited data does not significantly improve the rate of structure convergence in this case. Which implies that improvements to the structure are the result of the additional restraints identified, rather than the effect of the additional assignment information provided by the filtered/edited data. It also highlights the need to account for variability in the structure calculations. It is unclear whether addition of filtered/edited data would result in a similar behaviour for structure calculations of less ideal protein targets.

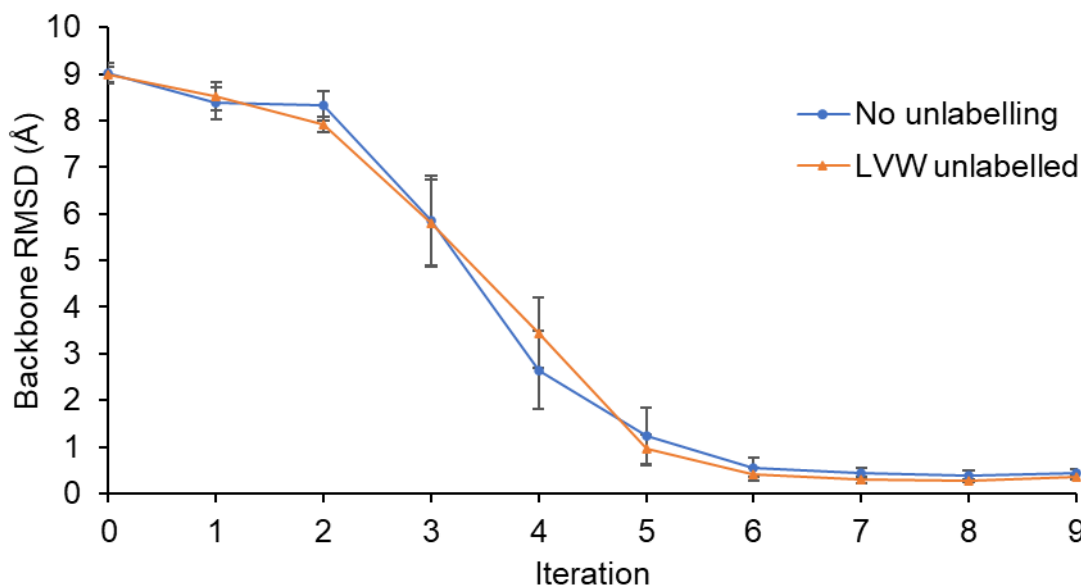


Figure 5.15. A graph of the average backbone RMSD across the ensemble for residues 27-109 for each iteration of five structure calculations for SasYr with a different 5% of the 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peaks randomly removed. The structure calculations were performed with LVW data (orange) and without (Blue). Error bars show the standard deviation for each point.

### 5.2.5. Specific Isotope Unlabelling to Compliment Sparse Data

The methods developed here are more likely to be of use in cases where the number of restraints from standard NOESY experiments is too sparse to produce a well-defined structural ensemble. In the absence of a less well behaved protein to test this, we attempted to artificially mimic a sparse dataset using the SasYr data. In order to achieve this, 75% of the peaks from the 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peak list were randomly removed. Structure calculations were then performed using this reduced dataset with and without data from unlabelling (Figure 5.16). The calculation was also performed using the reduced peak list with peaks randomly added back in to contain the same number of peaks as were used for the calculation containing the unlabelled data (Figure 5.16). This allows for analysis of the value of the residue type information encoded in the unlabelled data.

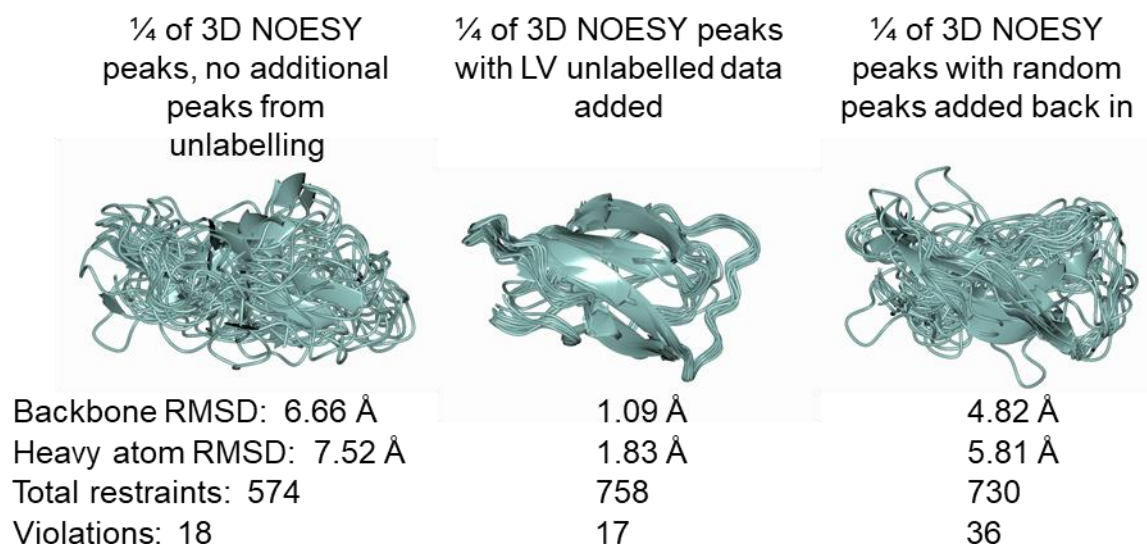


Figure 5.16. The results of structure calculations of SasYr with 75% of the 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peaks removed. Structure calculations were performed without any additional peaks added, with the LV unlabelled data added, and with the same number of peaks present in the LV data added from the previously removed 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peaks. Structural statistics for each structure is given below the structures.

As anticipated, removing 75% of the peaks from the 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peak list was sufficient to prevent folding of the structure. The structure could then be recovered effectively by adding in the LV unlabelled data but was not recovered to the same degree by adding in randomly selected peaks from the previously removed 3D NOESY peaks. This suggests that the additional residue-type information contained within the unlabelled data could have a significant impact on structure calculations with low quality or sparse data. However, the significance of this result to real case studies is highly questionable. Sparse datasets would rarely be uniformly sparse but would more likely have specific regions of the protein lacking restraints. In these cases, the unlabelled data may provide more or less value than demonstrated for the artificial sparse dataset. Additionally, for less ideal proteins that would give an initial sparse dataset, the filtered/edited experiments may also fail to identify NOEs in regions of sparse data.

The reduced data model was also used to test the effectiveness of the different isotope unlabelling patterns. The different isotope labelling patterns were tested individually and together with the reduced 3D data (Table 5.6).



Table 5.6. Structure statistics for SasYr structure calculations carried out with data from different unlabelling schemes used to compliment 25% of the peaks from the 3D (<sup>1</sup>H, <sup>1</sup>H) NOESY (<sup>13</sup>C/<sup>15</sup>N) HSQC peak list.

Labelling scheme	Number of restraints	Number of violations	Backbone RMSD	Heavy atom RMSD	Intra-residue restraints	Sequential restraints ( i-j =1)	Mid range restraints (1< i-j <5)	Long range restraints ( i-j ≥5)	G-factor
Fully labelled	571	29	6.47	7.43	196	154	43	136	-0.383
LV	738	27	3.68	4.45	250	162	43	225	-0.373
K	669	31	6.52	7.43	272	168	49	147	-0.382
LVK	752	17	6.10	7.00	299	169	54	172	-0.329
LVW	758	30	1.84	2.61	247	159	50	251	-0.256
LVW no residue type information	746	35	4.93	5.79	259	158	56	192	-0.418
All unlabelling schemes	1027	46	2.29	3.00	376	183	64	319	-0.261

The results for structure calculations using the individual labelling patterns and reduced 3D data (Table 5.6) follow the same pattern as the earlier results seen with the full 3D dataset (Table 5.5), although the effects are more pronounced. When including all unlabelled datasets, the resulting structure was less compact than for adding just LVW despite a higher number of restraints. This could be due to a larger number of mis-assigned restraints coming from the lysine unlabelled data. It may be that more iterations or more steps in the simulated annealing would help to improve the NOE assignments and lead to the combined unlabelling sets being more effective than LVW on its own.

The calculation was also performed with the LVW data without any residue type information included (Table 5.6). This led to a significantly worse structure than for the ambiguously assigned LVW data demonstrating the value of this information. The structure was still improved compared to the structure without any unlabelling data or with K or LVK data added. This further demonstrates the importance of the unlabelling scheme used.

The approach of artificially reducing the initial dataset is a poor approximation for true sparse datasets. This approach may represent scenarios where only poor quality 3D NOESY data could be collected, such as for unstable proteins, but challenging proteins will more likely be lacking assignments or feature specific regions of poor signal. For these reasons the investigation should be repeated with a less ideal model protein in order to better assess the usefulness of isotope unlabelling to aid in protein structure elucidation.

#### 5.2.6. Validation of Automatic NOE Assignment by Specific Isotope Unlabelling

The correct assignment of NOEs is key to NMR structure calculations. Due to the labour-intensive nature of NOE assignment and the power of iterative NOE assignment pipelines, automated NOE assignment has become standard practice in

NMR structure calculations. Incorrect assignment of NOEs can lead to the introduction of structural artifacts and a worse quality structure. The validation of automated NOE assignment pipelines is then an important factor for validation of the resulting structure. The possibility of using specific isotope unlabelling to validate automatic NOE assignments was explored. To do this, peaks in the 2D filtered/edited spectrum were first shift matched to peaks in the assigned 3D spectrum. The assignments of these peaks were then checked to see if they correspond to the potential residue types defined by the isotope unlabelling and filter/edit schemes. Peaks that did not match the possible residue types were likely to be mis-assigned. The flagged peaks were then manually inspected as they may be the result of overlapping peaks in the 2D spectrum that may or may not be resolved in the 3D spectrum. Validation of the automatically assigned SasYr 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}/^{15}\text{N}$ ) HSQC peaks was performed using the leucine/valine unlabelled F1fF2f peak list. This initially identified 27 potential mis-assigned peaks out of a total of 2126 peaks in the 3D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY ( $^{13}\text{C}$ ) HSQC peak list. This was then narrowed down to 11 mis-assigned NOEs after manual inspection. These 11 peaks all corresponded to intra-residue or sequential NOE peaks and so probably would not have significantly impacted the structure calculations. However, the ability to use specific isotope unlabelling to validate automatic NOE pipelines could be a valuable method of validation although requires more significant experimental investment than other methods of validation.

### 5.3. Conclusion and Future Work

This work demonstrates the potential of specific isotope unlabelling to help in the identification of NOEs that were missed when using standard NOESY experiments (due to peak overlap). It also demonstrates that the residue type information obtained can be used in combination with iterative NOE assignment and structure building pipelines to aid in protein structure elucidation. It was also shown that these techniques can be particularly useful when dealing with fewer restraints.

The results obtained with the different SasYr unlabelling patterns indicate that although all unlabelling patterns offered an improvement in the quality of the calculated structure, selection of residues more likely to be present in the core of the protein (bulky, hydrophobic and/or aromatic residues) are likely to be more effective than surface residues such as lysine. The results also suggest that combinatorial labelling can be highly effective when used with residues that give signals in distinct regions of the spectra as with leucine/valine and tryptophan. However, the results also suggest that combinatorial labelling with residues such as lysine, that give rise to a large number of unresolved peaks, can add ambiguity to the assignments and thus decrease the effectiveness of the unlabelling compared to single residue type unlabelling. This was seen in the worse results obtained for the LVK experiments compared to the LV experiments. These results can help to inform unlabelling patterns in any future work utilising these methods.

The method presented here has been shown to improve structure calculations through both allowing for the identification of new NOEs, due to the greatly simplified spectra, and by providing initial residue type information to inform early NOE assignments and calibrations. A method of effectively incorporating this residue type information was demonstrated. The specific experiments that should be applied to a specific protein will depend on the nature of the protein being studied. However, by combining the knowledge of transfer pathways in the different filtered/edited experiments, the number of experiments needed can be reduced to a minimal set of experiments. The information from the F1eF2f spectrum could be derived from the difference spectrum of F2f – F1fF2f making collecting all three potentially redundant. It would seem most efficient to collect only the double filtered and one half filtered NOESY as well as a filtered TOCSY. Collecting the additional experiments does however help to account for anomalies or leaky filtering. It would also be necessary to be sure of identifying every unlabelled-to-unlabelled NOE in the F1fF2f spectrum that would appear in the F2f spectrum. The labelling schemes used can also be limited to hydrophobic residues as these were shown to have the greatest impact on structure calculations.

A limiting factor in the use of NMR structures is the lack of good validation metrics. The use of specific isotope unlabelling and filtered/edited NOESY experiments provides a new method of validating automatic NOE assignment pipelines by using the potential residue type assignments based on the unlabelling and filter/edit schemes. The potential residue types can then be checked against the results from automatic NOE assignments of the fully labelled sample. This application is limited by the need for additional sample preparation and the need for manual inspection of peak matching, although this may be alleviated through the use of 3D filtered/edited experiments. Further investigation into the use of specific isotope unlabelling to aid in validation of automatic NOE assignments should be carried out.

In order to demonstrate the broad applicability of these techniques, a less ideal, larger protein should be used. A more poorly behaved, larger protein would be more likely to benefit more from these methods and would make a more realistic test case. An additional advantage would be that the specific isotope unlabelling would also benefit the backbone and sidechain assignment of the protein and so would serve an additional purpose. It was also shown that it is important to account for variability in structure calculations when comparing the results from different calculations. To do this, a similar approach to the one employed in Figure 5.15 could be used, where a random 5% of the standard NOESY peaks could be removed and multiple runs could be performed with slight variations in the starting data. Structure statistics could then be averaged across the repeats. However, when dealing with more sparse data, removal of NOESY peaks may have to great of an effect on structure calculations.

An additional potential application of the specific isotope unlabelling and filtered/edited NOESY experiments could be in the generation of accurate distance restraints from NOE build up curves [273, 274]. By following the change in signal intensity at different mixing times, accurate distance measurements can be extracted. 2D NOESY experiments are usually used for this due to the long acquisition times needed for 3D experiments. Using specific isotope unlabelled protein and filtered/edited NOESY experiments could allow measurement of a much larger number of distances to be extracted than with the full 2D ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY.

These accurate distances could then be used for structure calculation of validation. Recently AlphaFold2 has rendered experimental protein structure determination of small proteins largely redundant (except for cases such as novel folds and artificial sequences). Specific isotope unlabelling and filtered/edited NOESY experiments could potentially be used in the validation of AlphaFold2 models or to generate structures of alternate conformations/states, providing a valuable use case for these methods in a post AlphaFold2 world.

## 6. Conclusion

The scaffold protein LIMD1 has been characterised as a tumour suppressing protein with roles in miRNA mediated silencing [57], hypoxic signalling [24], HIPPO signalling [77], cell proliferation [79], cell differentiation [75], and focal adhesion [63]. A large amount of work has focussed on determining the role of LIMD1 in various signalling pathways and diseases. LIMD1 has been implicated in the development of a range of cancers and has been shown to play a key role in the non-small cell lung cancers [68, 70, 72, 275-277]. Therefore, an enhanced understanding of the molecular and cellular biology of LIMD1 may aid in the development of effective LIMD1-targeted therapies. It has been shown that stabilisation of LIMD1 mRNA by a long non-coding RNA was able to promote apoptosis and impede proliferation of non-small cell lung cancer cell lines [277]. Stabilisation of LIMD1, promoting LIMD1 activity or otherwise regulating targets of LIMD1 may then prove to be effective in the treatment of non-small cell lung cancer. As inhibition of PHD2 activity has been shown to be an effective method of treating renal anaemia, inhibiting the interaction of PHD2 with LIMD1 or between LIMD1 and VHL may also prove to be effective in the treatment of renal anaemia [278]. To aid in the potential development of LIMD1 targeted therapeutics, and to expand general understanding of LIM domain and disordered protein binding, a more in depth analysis of the protein:protein interactions made by LIMD1 was carried out.

Work initially focussed on attempting to investigate the interactions of LIMD1 with TNRC6A and AGO2 in miRNA mediated silencing. An attempt to analyse the interactions between LIMD1 and AGO2 was severely impacted by the pandemic due to a shortage of media for insect cell culture. Peptide arrays, pull-down assays and computational predictions were employed in an attempt to identify the LIMD1 binding site on TNRC6A, but no clear LIMD1:TNRC6A interaction could be demonstrated. Despite this, work on the production of recombinant LIM domain samples was valuable for analysis of LIMD1 in the hypoxic signalling pathway. Future work on LIMD1 in the miRNA mediated silencing pathway could benefit by focussing on the reconstruction of the potential complex identified by AlphaFold2 formed between the first 2 LIM domains of LIMD1, motif 2 of TNRC6A, and AGO2. NMR spectroscopy

could also provide an effective method to study these interactions, as was demonstrated for analysis of LIMD1 in the hypoxic signalling pathway. The lack of interaction demonstrated between LIMD1 and either TNRC6A or AGO2, despite the TNRC6A:AGO2 interaction forming under the same conditions, may also suggest that LIMD1 does not directly interact with AGO2 or TNRC6A or that the interaction requires additional proteins or post-translational modifications. Zhou et al. (2019) showed that phosphorylation of LIMD1 was important for the tumour suppressing activity of LIMD1, although the mechanism behind this remains unclear [279]. Future experiments into LIMD1 interacting with TNRC6A or AGO2 may benefit from using phosphorylated or phospho-mimetic LIMD1, in the case that these interactions are dependent on LIMD1 phosphorylation.

NMR spectroscopy allowed interactions to be captured for LIMD1 with both VHL and PHD2. The interaction with PHD2 was able to be studied in greater depth, with LIMD1 truncations revealing a smaller PHD2 binding region. Partial backbone resonance assignment of this region revealed a short, highly conserved, stretch of LIMD1 (168-191) responsible for binding to PHD2. This binding site agreed with *in vivo* studies and was verified by NMR spectroscopy using a peptide derived from the proposed PHD2 binding region of LIMD1 [24]. Perdeuteration of PHD2 allowed for identification of a LIMD1 binding site on PHD2, which agreed with models predicted by AlphaFold2 as well as being consistent with the structure of the complex between PHD2 and a synthetic cyclic peptide generated using RaPID [163, 209]. This work confirmed that LIMD1 is able to directly bind to both PHD2 and VHL, providing significant support for the scaffolding model of LIMD1 activity first proposed by Foxler *et al.* (2012) [24]. This work also provided compelling evidence that the interaction of LIMD1 and PHD2 occurs primarily at the site around residues 168-191 of LIMD1 and the  $\alpha$ -helix spanning residues 190-205 of PHD2. Future work should look to verify this mode of interaction by using mutational analysis and/or competition assays with the cyclic peptide. Comparison of the PHD2 binding site on LIMD1 to AJUBA and WTIP revealed only poor conservation with AJUBA and no conservation with WTIP. This could explain the differences in binding affinities between the LAW proteins for the different PHD enzymes [24]. A similar approach could be used to identify the PHD2 binding site on AJUBA and WTIP which would provide more



insight into the mode of recognition of these proteins and more general insight into how specificity is maintained in IDPs. Production of full-length LIMD1, as was achieved by Wang *et al.* (2021), would allow for investigation of the full LIMD1:PHD2:VHL complex which would allow for characterisation of the scaffolding behaviour of LIMD1 and the oligomeric states of the complex [63]. The identification of the potential interface between LIMD1 and PHD2 could be used to better inform computational modelling of the interaction than has previously been carried out [215]. The work presented in this thesis provides the first characterisation of a PHD2, VHL scaffold protein at a molecular level, which may provide valuable insights for the development of drugs to treat cancers or ischemic diseases by targeting the LIMD1:PHD2 interaction.

Work on LIMD1 relied heavily on the use of NMR spectroscopy to study the interactions of LIMD1 with its partner proteins. For certain applications, it was necessary to conduct specific isotopic labelling of PHD2 using metabolic precursors. An alternative and more affordable approach for obtaining site- or residue-specific labelling involves specific isotopic unlabelling. Work was carried out to allow for the use of specific isotope unlabelling to aid in structural studies of proteins. Specific isotope unlabelling was combined with 2D filtered/edited NOESY experiments to specifically observe  $^1\text{H}$ - $^1\text{H}$  interactions involving unlabelled residues. Methods for incorporating this data into structure calculations were presented and information was derived to inform on the selection of unlabelling schemes for the greatest positive effect on structure calculations. It was also demonstrated that specific isotope unlabelling could be used to validate automated NOE assignment pipelines. The work presented here showed that combining specific isotope unlabelling with filtered/edited NOESY experiments can significantly improve structure calculations, providing residues in the core of the protein are selected for unlabelling. The findings can be used to inform any work that applies this approach in the future. The exploration of these techniques may also encourage novel implementation of these methods such as investigating specific internal structural elements by careful selection of unlabelled sites or for validation of structures predicted by AlphaFold or other computational approaches. This type of experiment could be applied to future investigations of LIMD1 to probe the structure of LIM domains in the absence and

presence of partner proteins. This could report on potential domain movements related to partner protein binding and potentially aid in structural elucidation of the LIM domains of LIMD1 in complex with VHL or TNRC6A.

This thesis provides case studies for the use of solution NMR spectroscopy in the study of protein:protein interactions formed by intrinsically disordered proteins and provides new tools to aid in future studies of protein structure by solution NMR spectroscopy.

## 7. Appendix

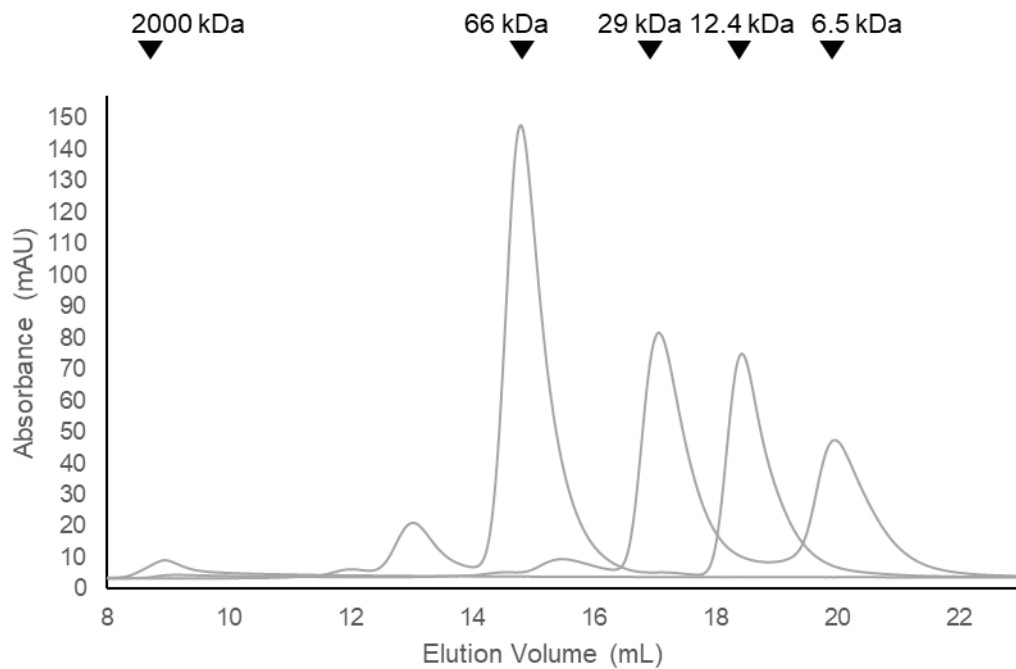


Figure 7.1. SEC chromatograms for molecular mass standards on a superdex increase 200 10/300 GL column. Used for FL-PHD2 with LIMD1 SEC experiments. Blue Dextran = 2000 kDa, albumin = 66 kDa, carbonic anhydrase = 29 kDa, cytochrome C = 12.4 kDa, aprotinin = 6.5 kDa. Sigma-Aldrich: MWFG70

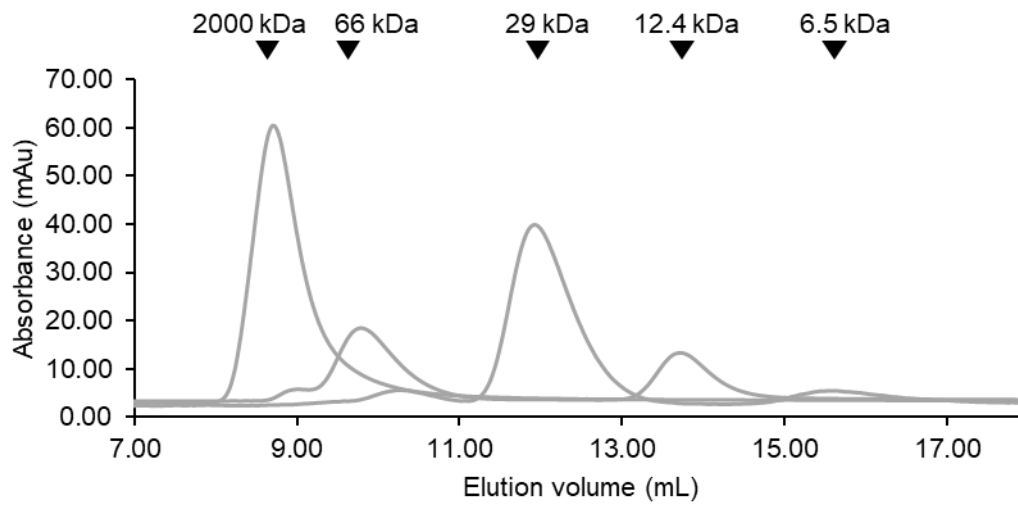


Figure 7.2. SEC chromatograms for molecular mass standards on a superdex increase 75 10/300 GL column. Used for PHD2 181-426 with LIMD1 166-260 and VHL with LIM2 SEC experiments. Blue Dextran = 2000 kDa, albumin = 66 kDa, carbonic anhydrase = 29 kDa, cytochrome C = 12.4 kDa, aprotinin = 6.5 kDa. Sigma-Aldrich: MWFG70.

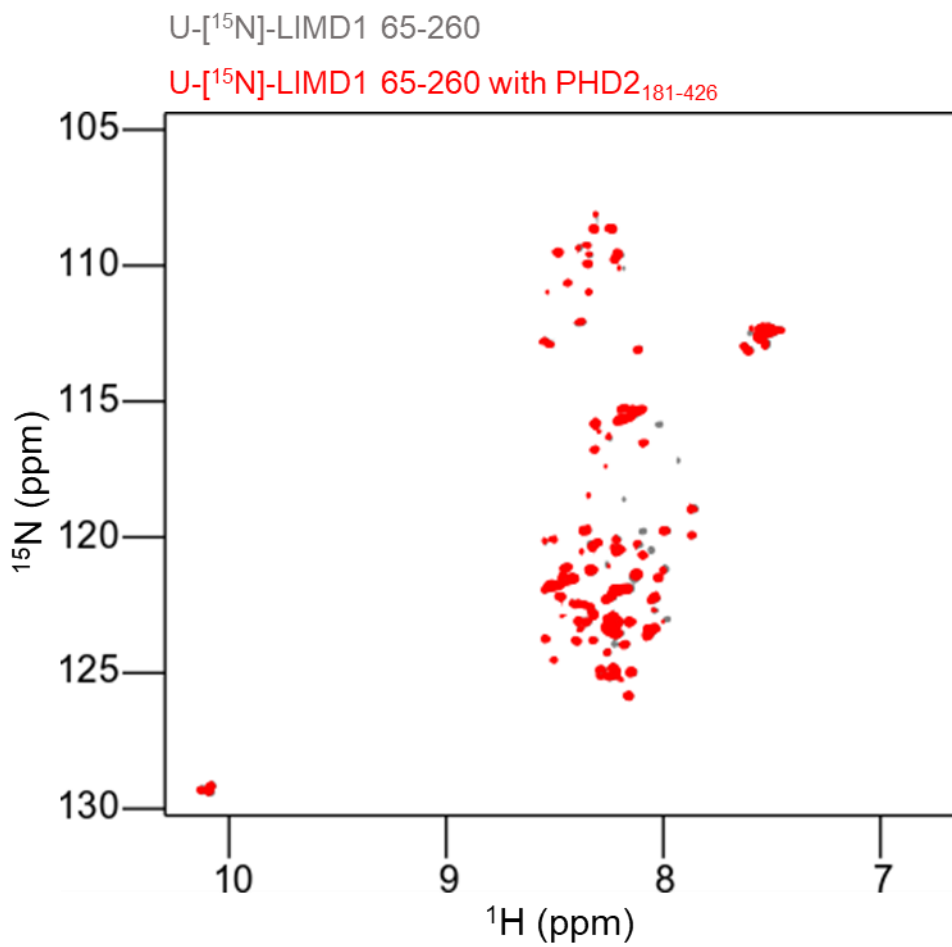


Figure 7.3. 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectra of 100 μM [U-<sup>15</sup>N]-LIMD1 65-260 with (grey) and without (red) PHD2 181-426

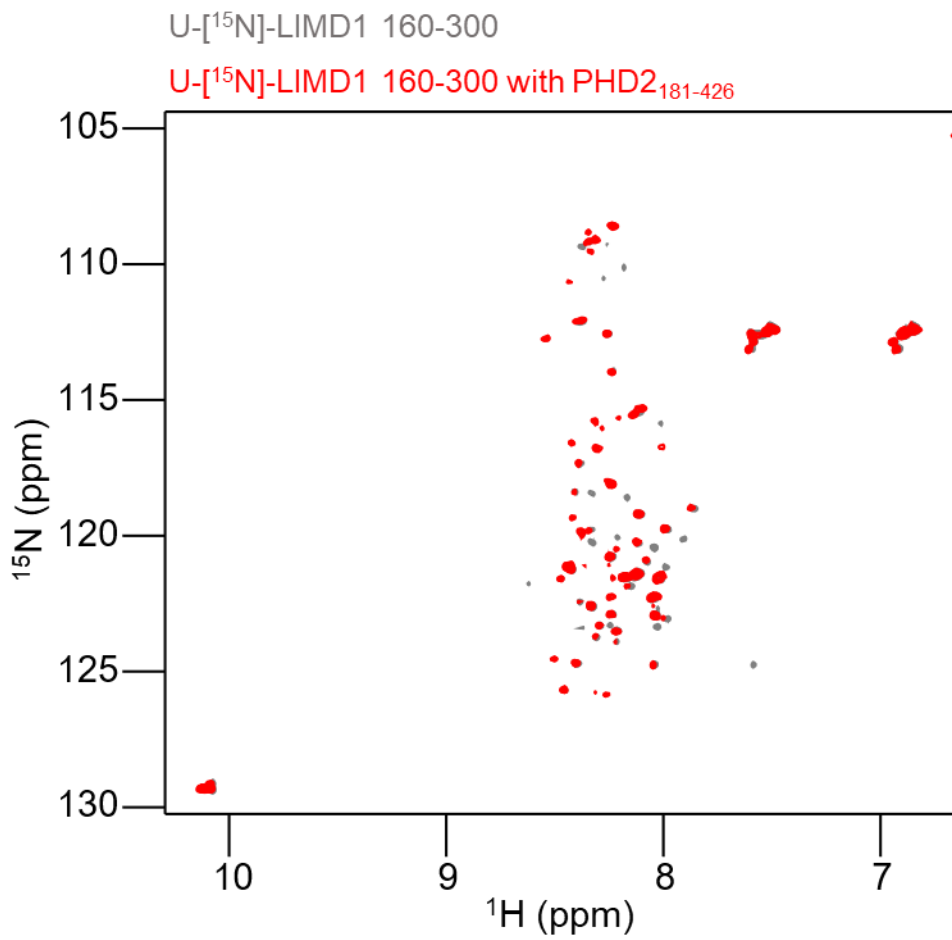


Figure 7.4. 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectra of 100 μM [U-<sup>15</sup>N]-LIMD1 160-300 with (grey) and without (red) PHD2 181-426

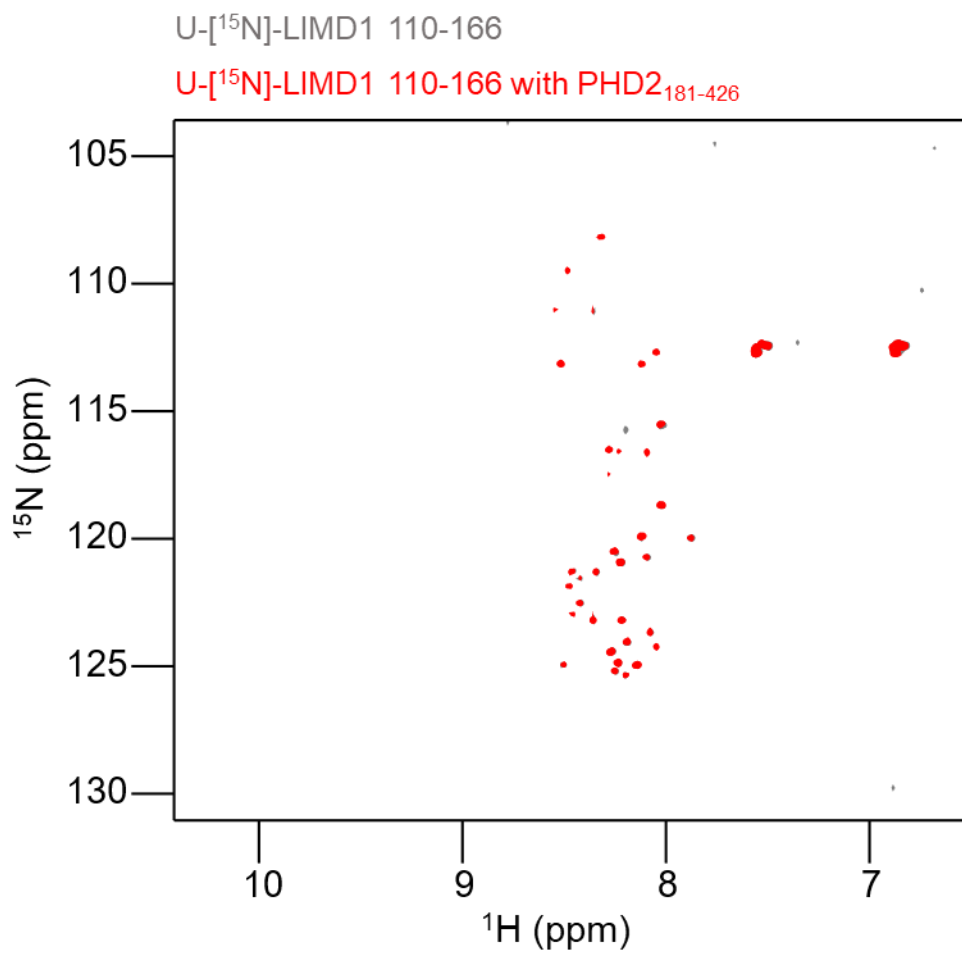


Figure 7.5. 2D (<sup>1</sup>H, <sup>15</sup>N) HSQC spectra of 100 μM [U-<sup>15</sup>N]-LIMD1 110-166 with (red) and without (grey) PHD2 181-426

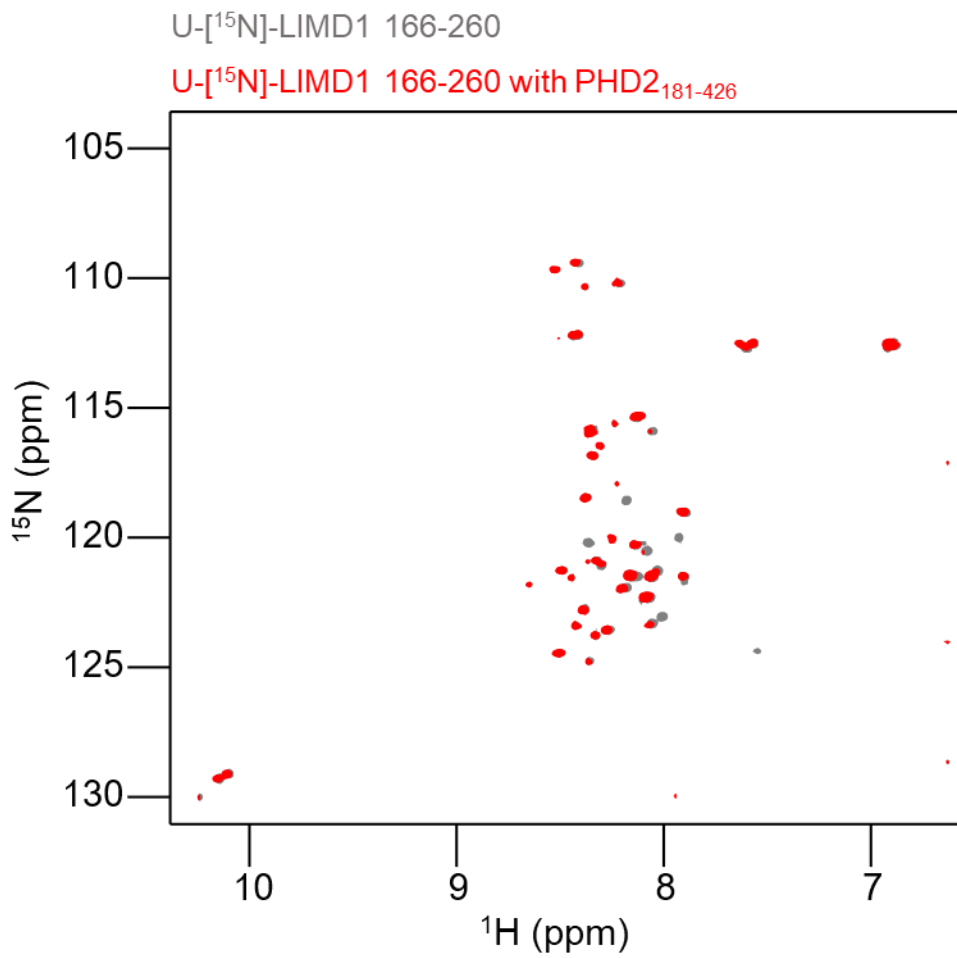


Figure 7.6. 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra of [U-<sup>15</sup>N]-LIMD1 166-260 with (red) and without (grey) PHD2 181-426



[U-<sup>15</sup>N]-LIMD1 166-260

[U-<sup>15</sup>N]-LIMD1 166-260 with PHD2<sub>181-426</sub> and  
HIF1α CODD peptide

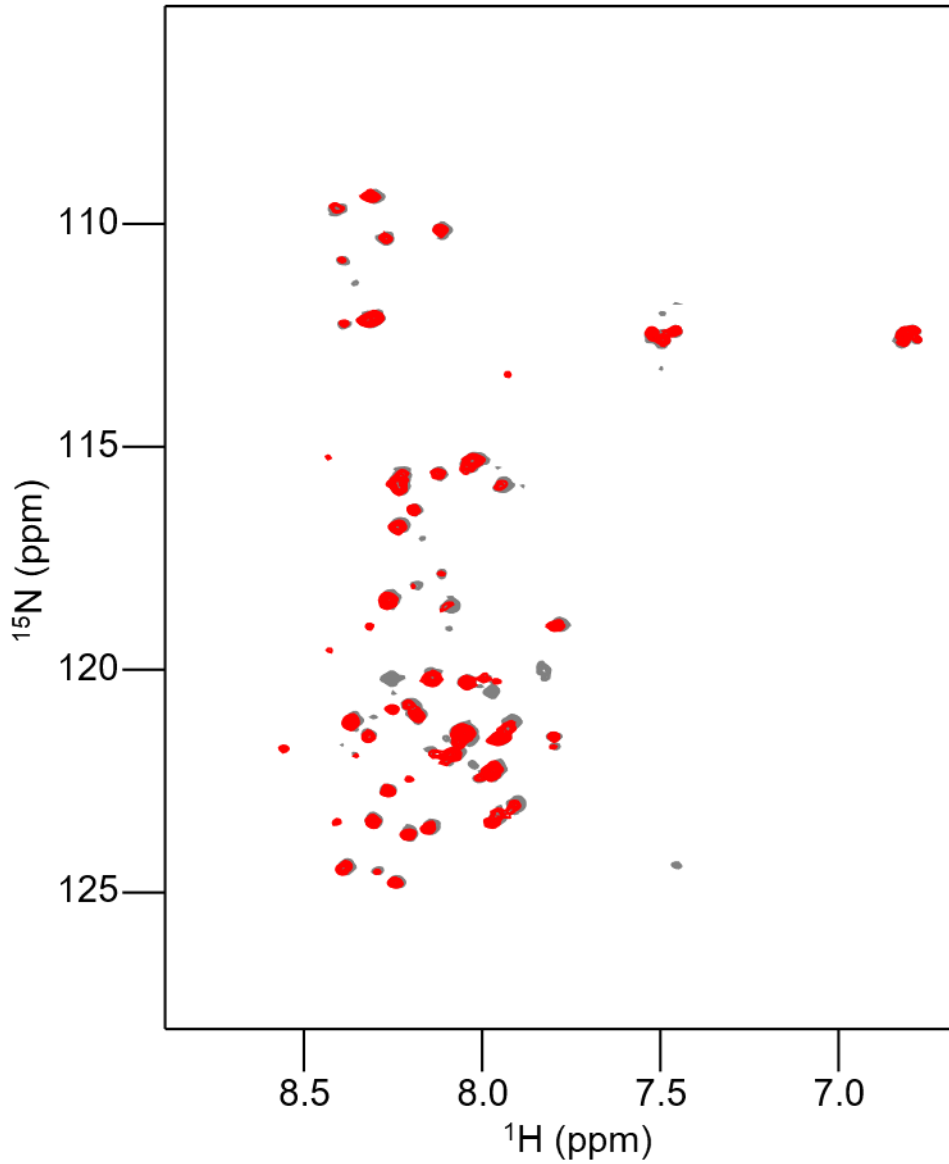


Figure 7.7. 2D (<sup>1</sup>H, <sup>15</sup>N) HMQC spectra for 100 μM [U-<sup>15</sup>N]-LIMD1 166-260 with (red) and without (grey) 250 μM PHD2 181-426 and 250 μM HIF-1α CODD peptide.

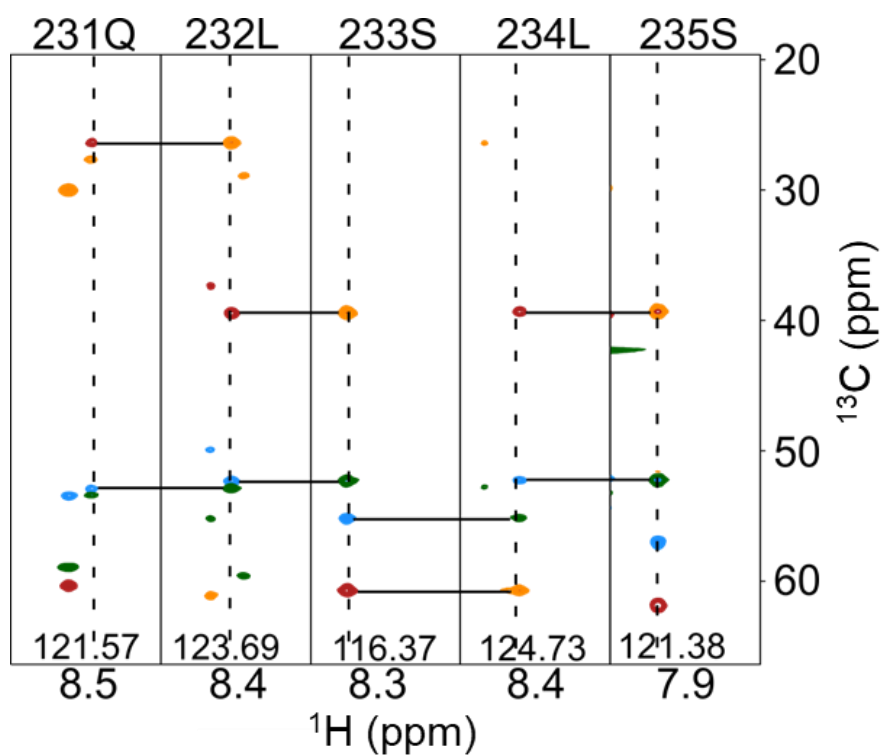


Figure 7.8. strips from the HNCACB and HNCOCACB spectra of LIMD1 166-260 for the assigned region of 231-235. The HNCACB spectra is shown with green positive and blue negative contours and the HNCOCACB spectra is shown with yellow negative and purple positive peaks.

[U- $^{15}\text{N}$ ,  $^2\text{H}$ ],( $^{15}\text{L}^{\text{proSVproS}}$ )-[ $^{13}\text{C}^1\text{H}_3$ ]-PHD2 $_{181-426}$

LIMD1 168-191 with PHD2

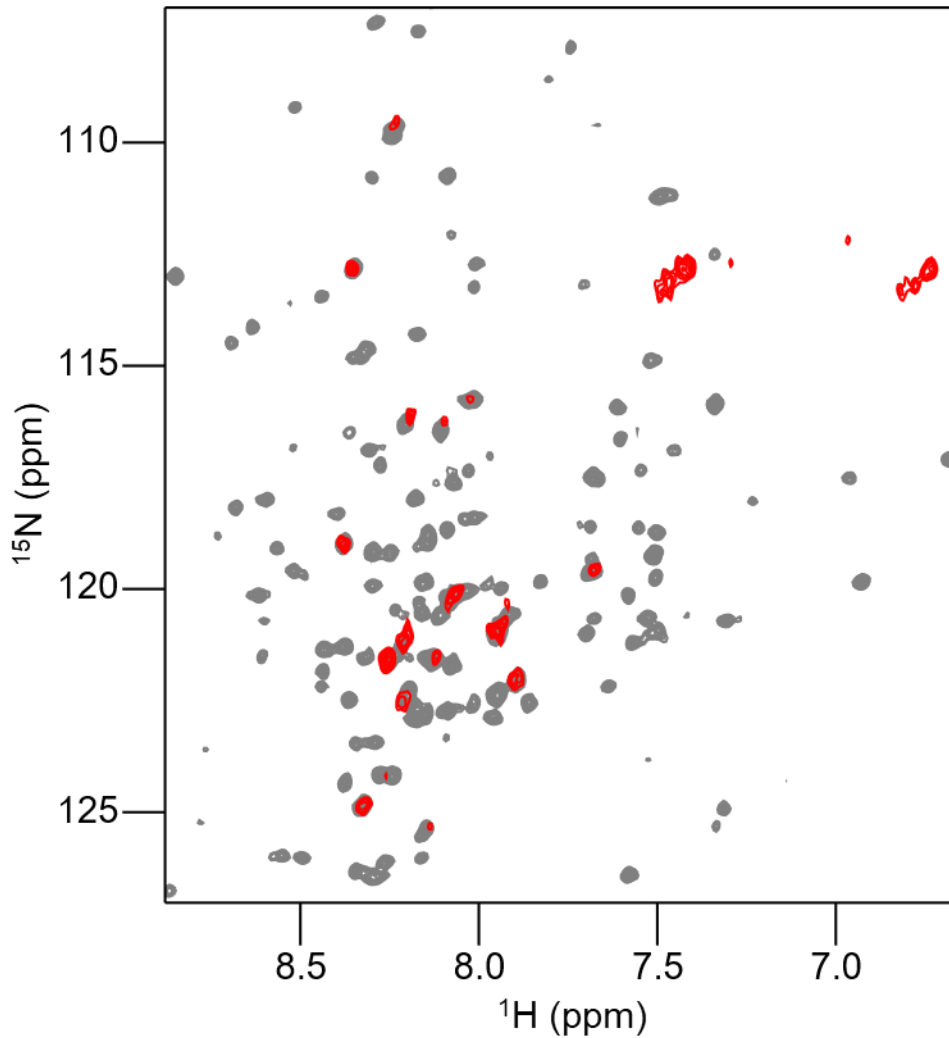


Figure 7.9. 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) TROSY spectrum of [U- $^{15}\text{N}$ ,  $^2\text{H}$ ],( $^{15}\text{L}^{\text{proSVproS}}$ )-[ $^{13}\text{C}^1\text{H}_3$ ]-PHD2 $_{181-426}$  overlaid with the 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HMQC spectra of unlabelled LIMD1 168-191 with PHD2 181-426. Peaks in the LIMD1 168-191 with PHD2 181-426 spectrum can be seen to match peaks in the PHD2 spectrum.

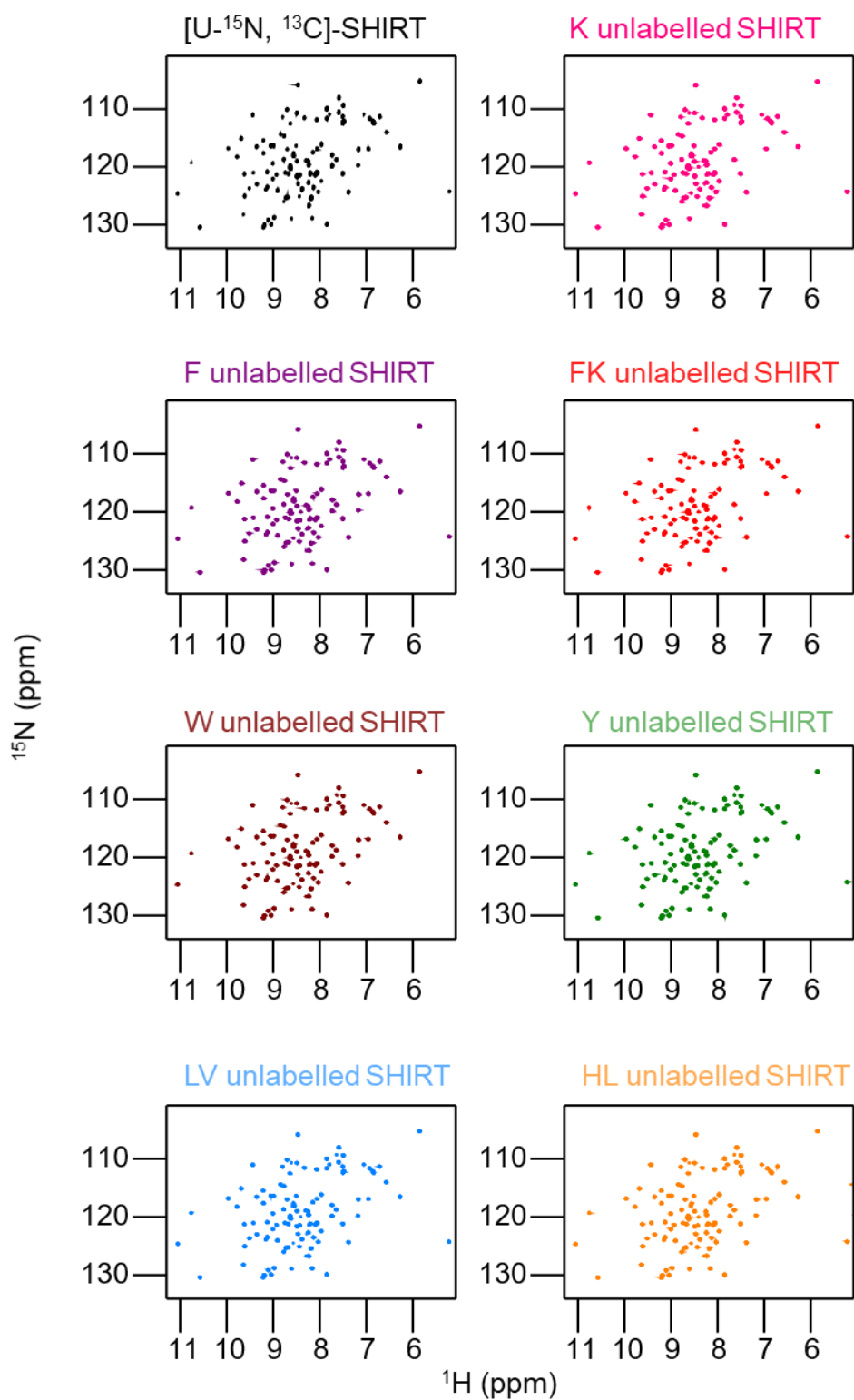


Figure 7.10. 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectra for SHIRT-R3 samples

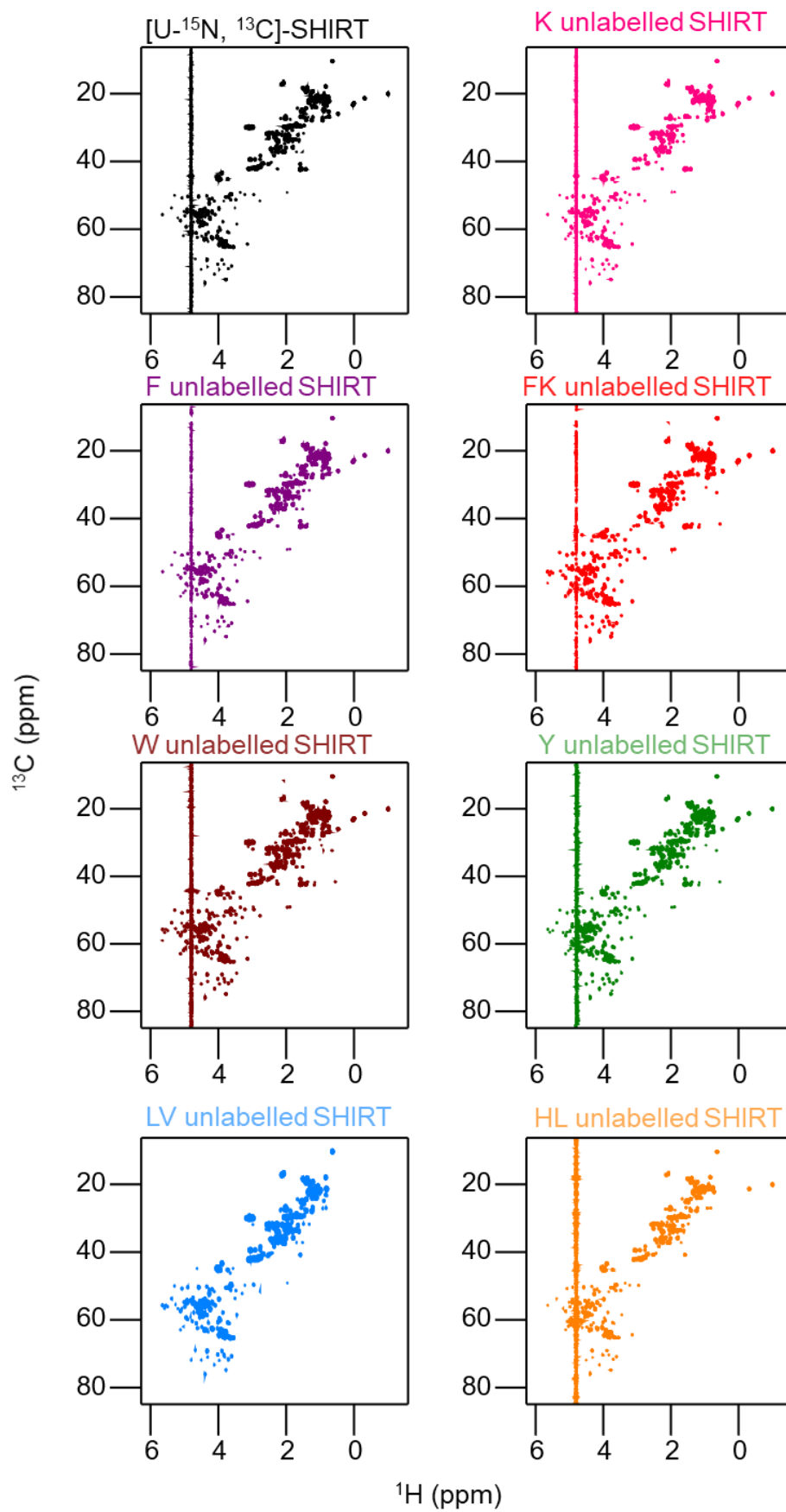
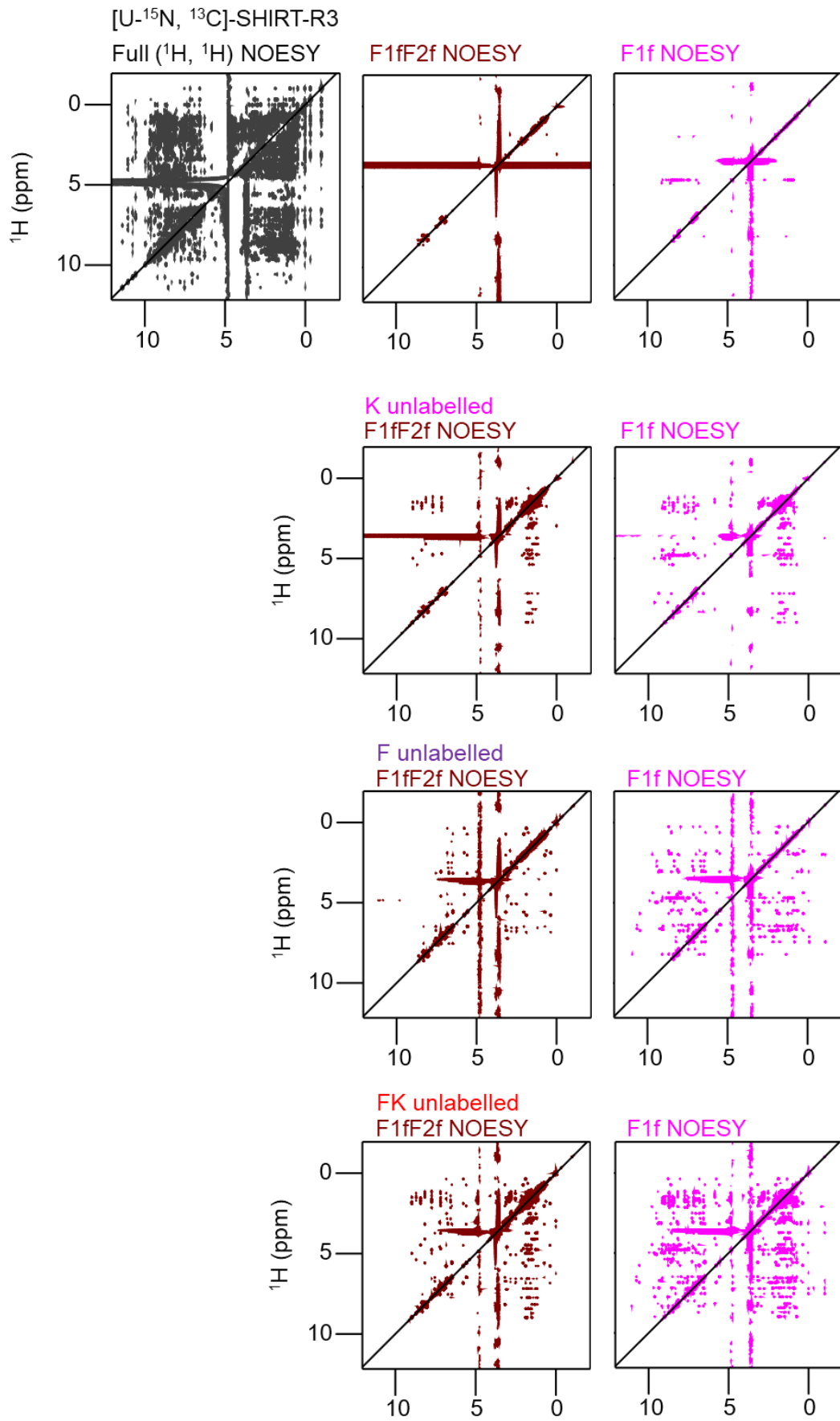


Figure 7.11. 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra of SHIRT-R3 samples



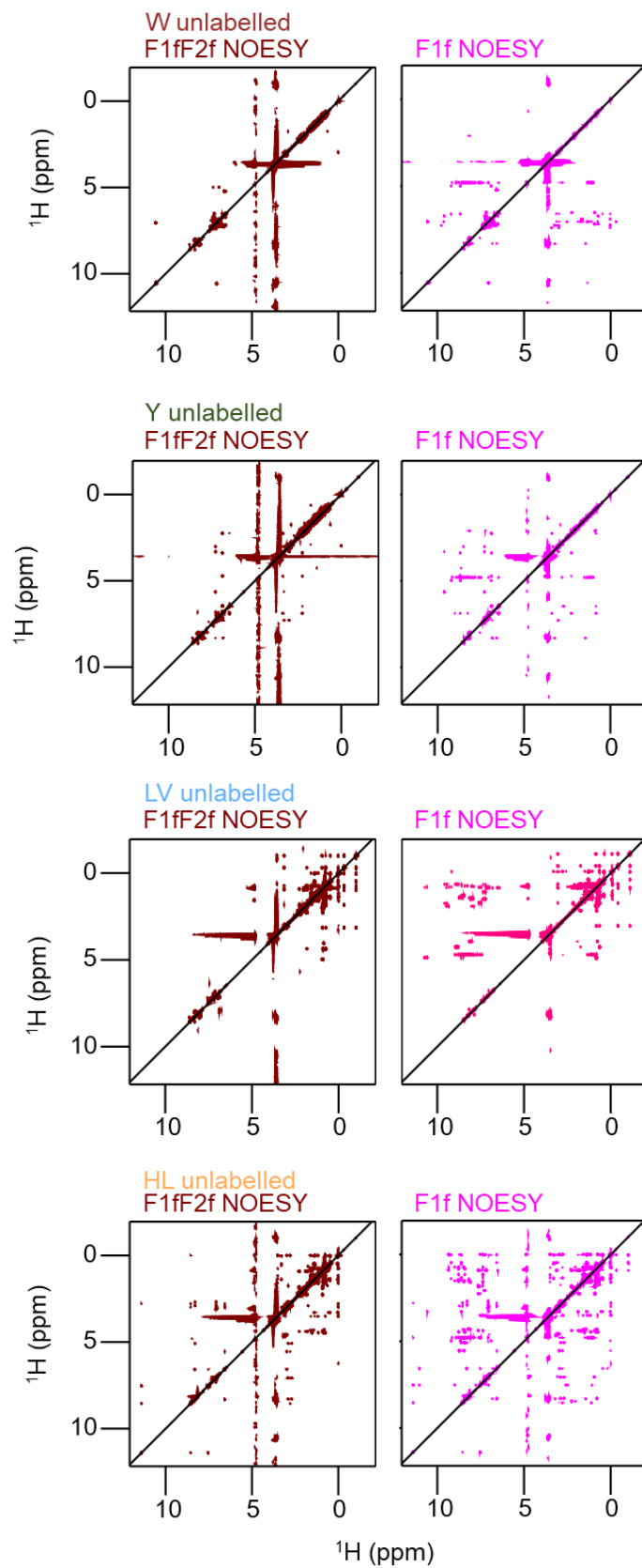


Figure 7.12 a 2D full ( $^1\text{H}$ ,  $^1\text{H}$ ) NOESY for the [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-SHIRT sample and 2D F1f2f and F1f NOESY spectra for all SHIRT-R3 samples

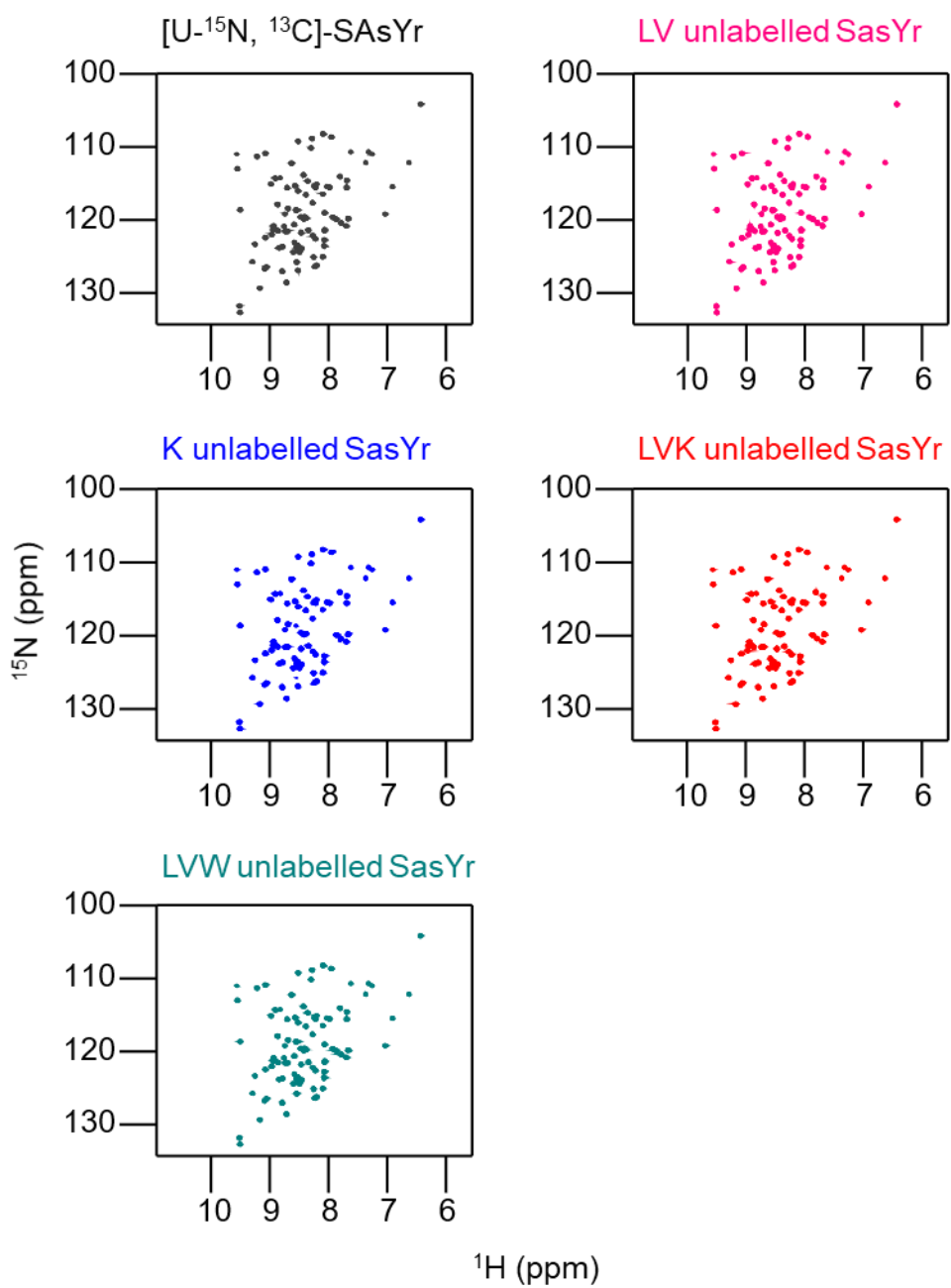


Figure 7.13. 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectra of SasYr samples



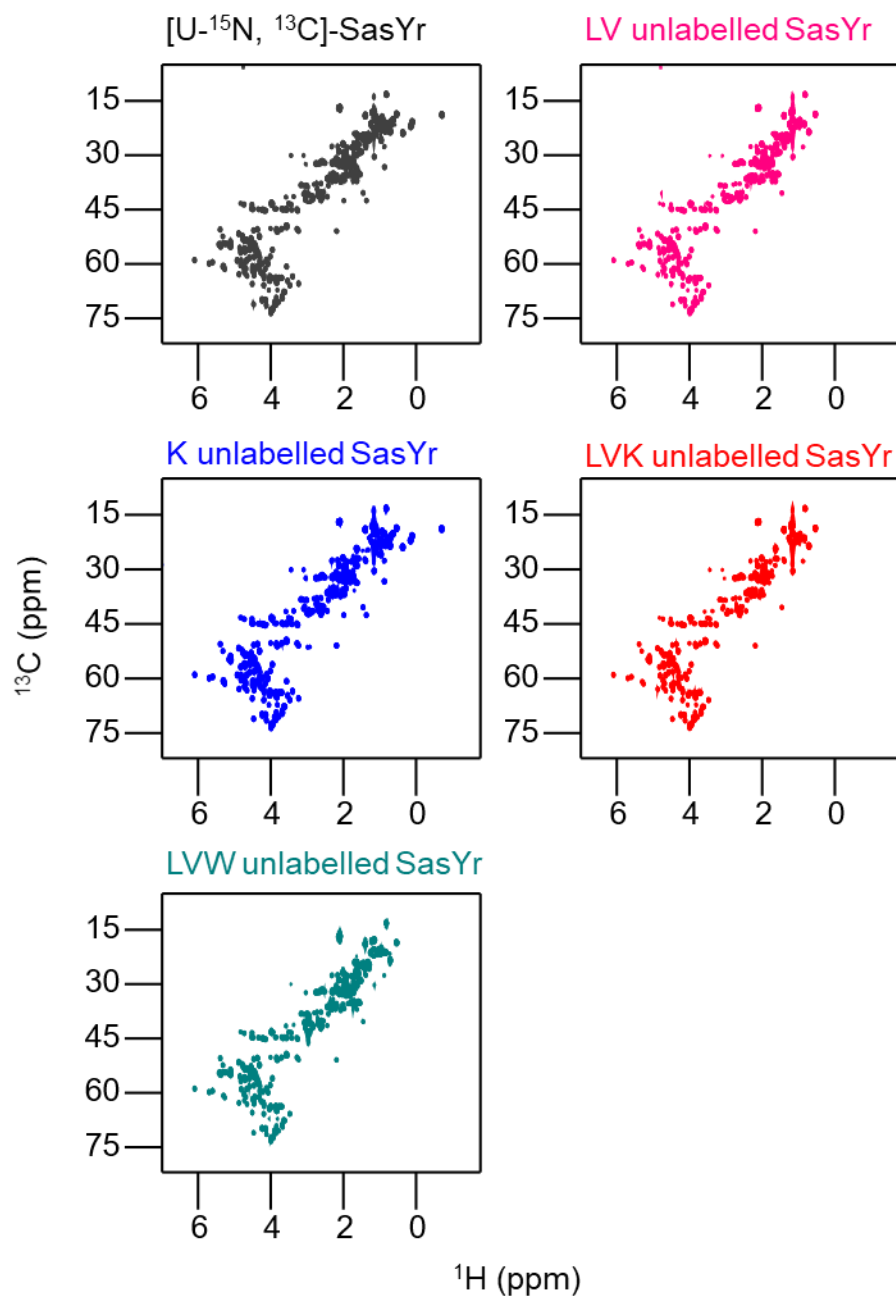


Figure 7.14. 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ )-HSQC spectra of SasYr samples

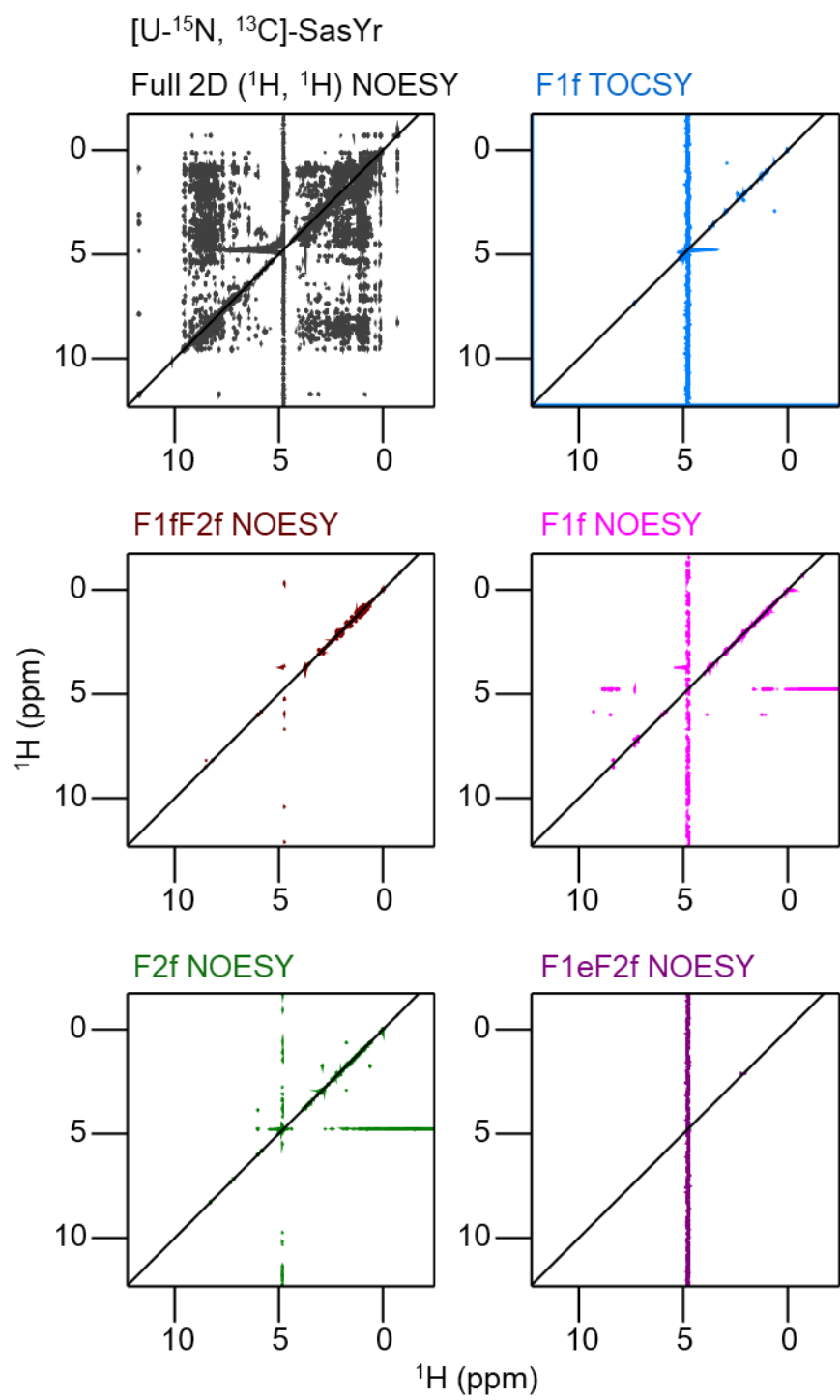


Figure 7.15. 2D NOESY spectra of [U-<sup>13</sup>C, <sup>15</sup>N]-SasYr

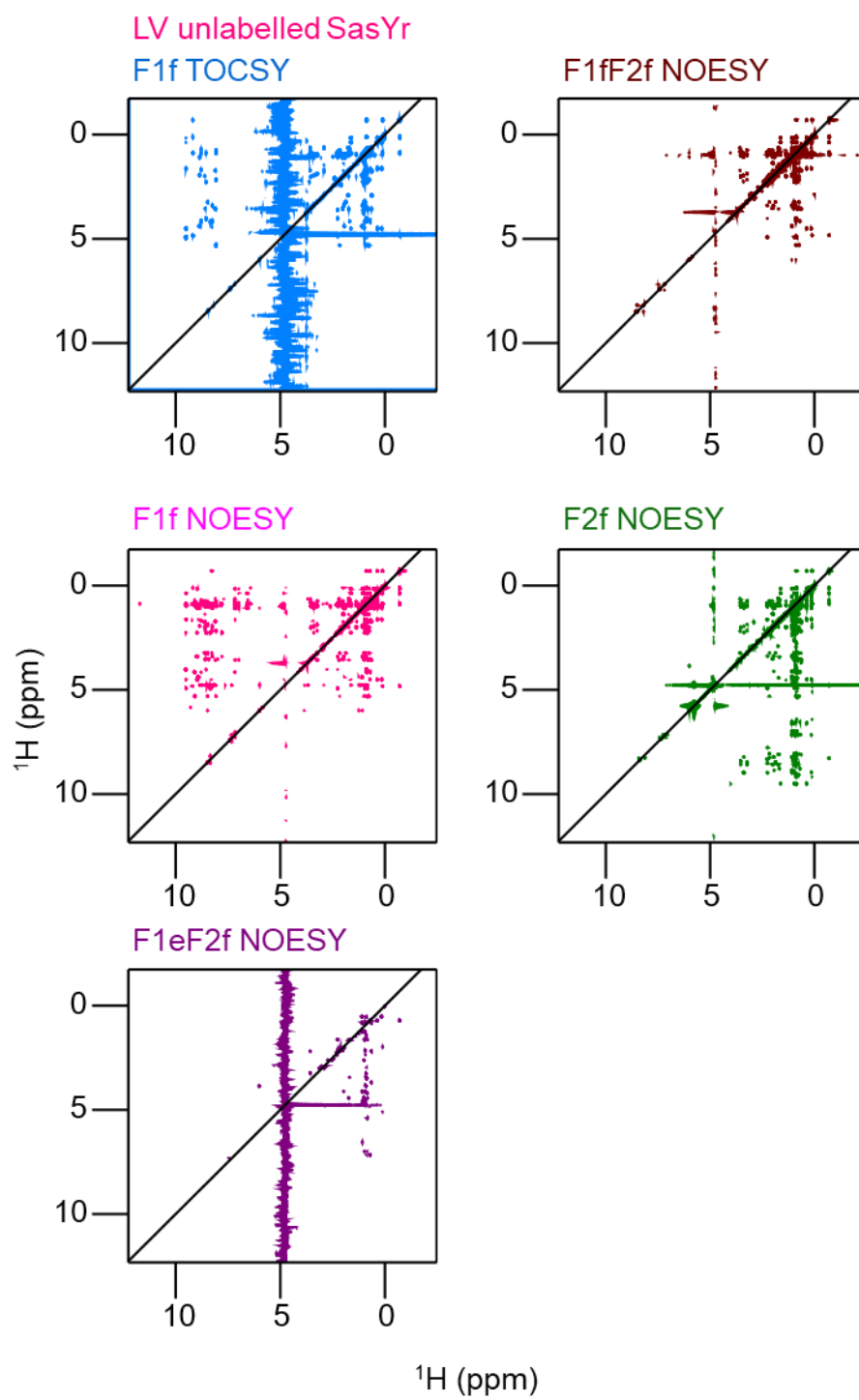


Figure 7.16. 2D NOESY spectra of LV unlabelled SasYr

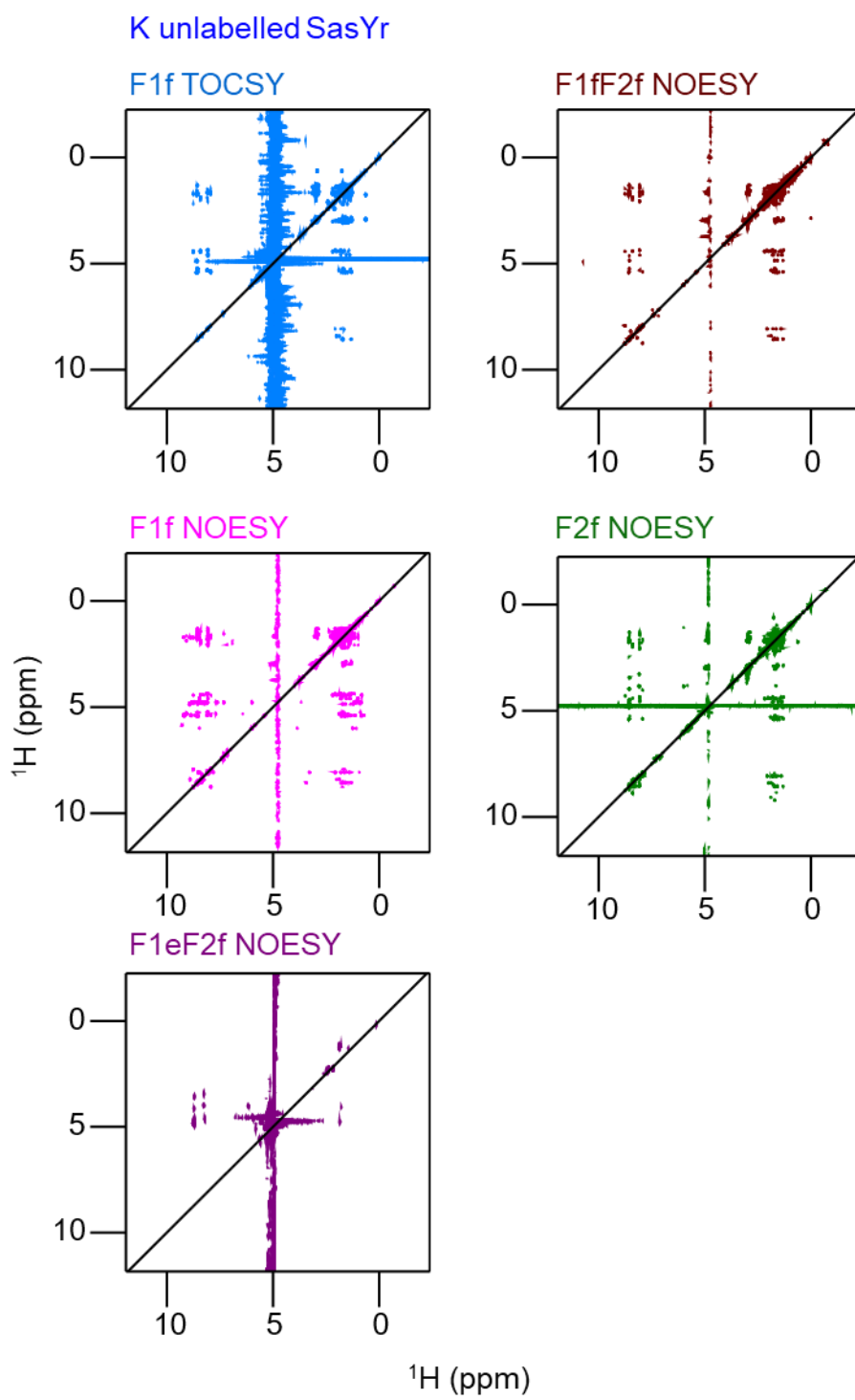


Figure 7.17. 2D NOESY spectra of K unlabelled SasYr

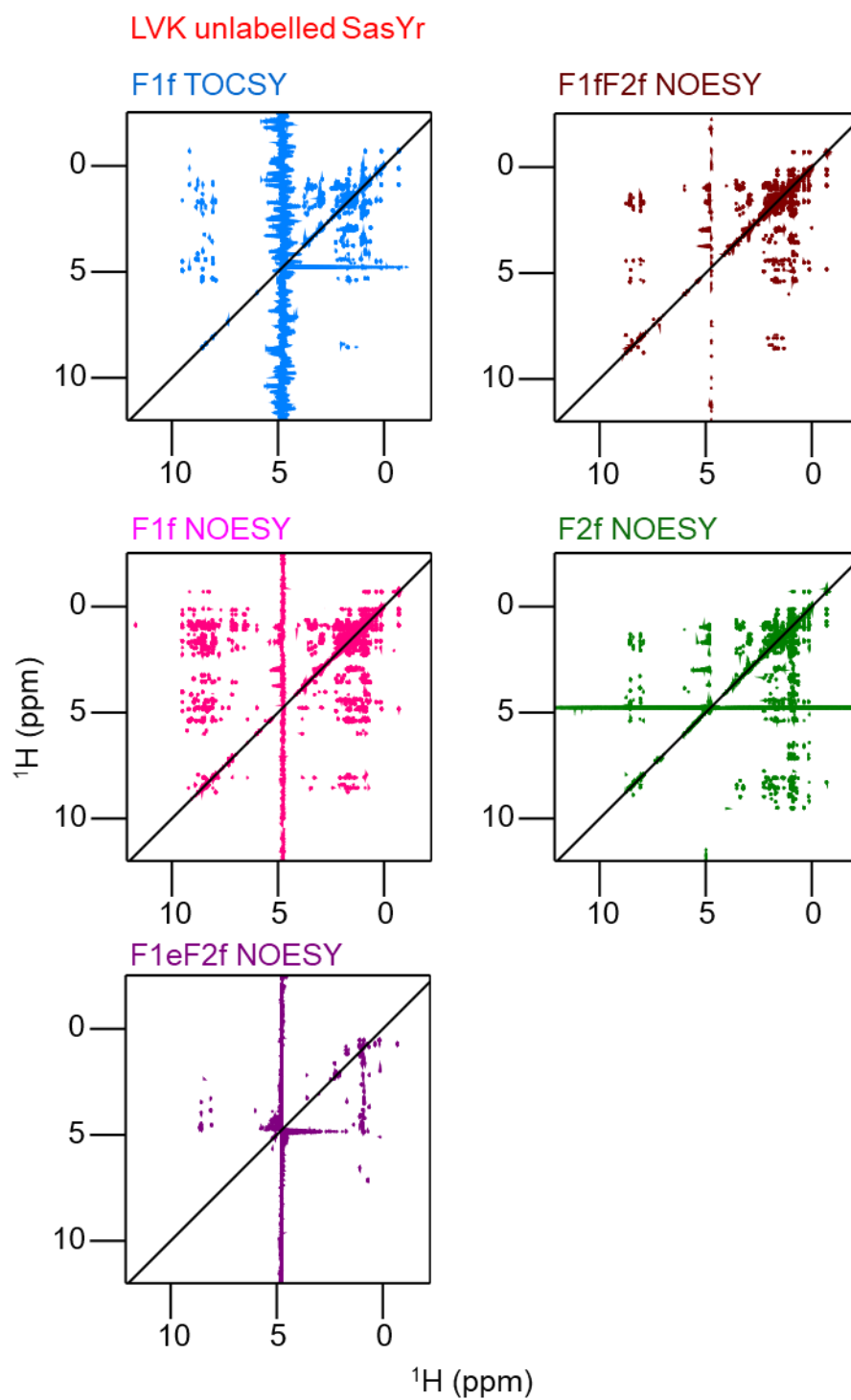


Figure 7.18. 2D NOESY spectra of LVK unlabelled SasYr

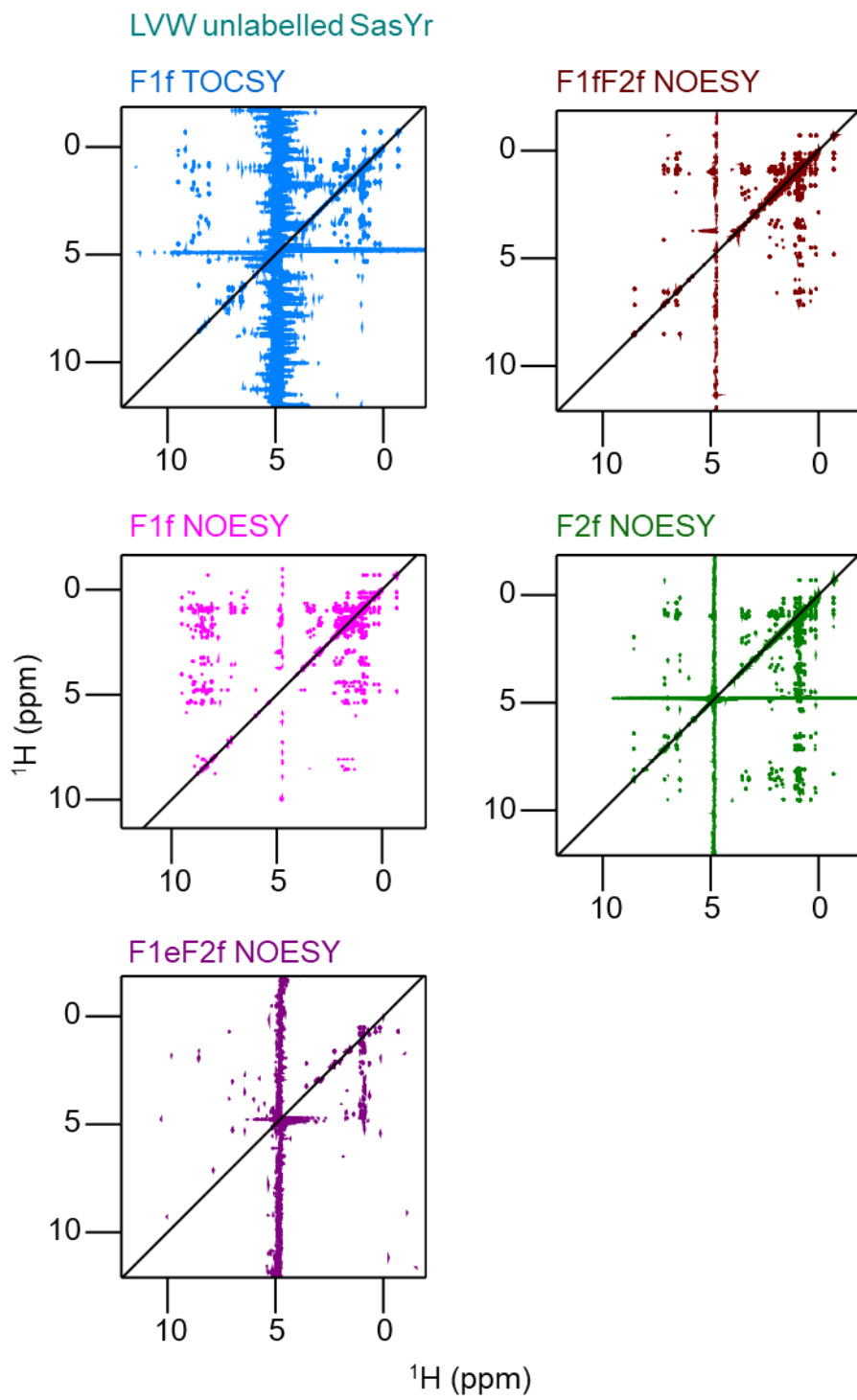


Figure 7.19. 2D NOESY spectra of LVW unlabelled SasYr

**A**

#	Resonances	Value	Upper Limit	Lower Limit	Error
1:0	75ValHa-67ValHb	3.009	3.611	2.407	1.204
1:01	75ValHa-75ValHb	3.009	3.611	2.407	1.204
2:0	75ValHa-67ValHga*	2.974	3.569	2.379	1.190
2:01	75ValHa-75ValHga*	2.974	3.569	2.379	1.190
3:0	53ValHa-108ValHga*	3.786	4.544	3.029	1.515
3:01	53ValHa-32ValHga*	3.786	4.544	3.029	1.515
3:02	53ValHa-53ValHb	3.786	4.544	3.029	1.515
3:03	53ValHa-67ValHga*	3.786	4.544	3.029	1.515
3:04	53ValHa-67ValHgb*	3.786	4.544	3.029	1.515
4:0	53ValHa-53ValHgb*	3.319	3.983	2.655	1.328
5:0	53ValHa-53ValHga*	3.305	3.966	2.644	1.322
6:0	55ValHa-37ValHga*	4.018	4.821	3.214	1.607
6:01	55ValHa-37ValHgb*	4.018	4.821	3.214	1.607
6:02	55ValHa-75ValHgb*	4.018	4.821	3.214	1.607
6:03	55ValHa-85ValHga*	4.018	4.821	3.214	1.607
7:0	32ValHa-32ValHgb*	3.276	3.931	2.620	1.310
8:0	32ValHa-108ValHga*	3.130	3.756	2.504	1.252

**B**

#	Resonances	Value	Upper Limit	Lower Limit	Error
1:0	73TrpHz2-42ValHga*	3.437	4.125	2.750	1.375
1:01	73TrpHz2-42ValHgb*	3.437	4.125	2.750	1.375
1:02	73TrpHz2-49ValHga*	3.437	4.125	2.750	1.375
2:0	73TrpHz2-37ValHga*	3.798	4.558	3.038	1.519
2:01	73TrpHz2-37ValHgb*	3.798	4.558	3.038	1.519
2:02	73TrpHz2-75ValHgb*	3.798	4.558	3.038	1.519
2:03	73TrpHz2-83LeuHdb*	3.798	4.558	3.038	1.519
2:04	73TrpHz2-85ValHga*	3.798	4.558	3.038	1.519
3:0	73TrpHz2-108ValHga*	3.634	4.360	2.907	1.453
3:01	73TrpHz2-32ValHga*	3.634	4.360	2.907	1.453
3:02	73TrpHz2-53ValHb	3.634	4.360	2.907	1.453
3:03	73TrpHz2-67ValHgb*	3.634	4.360	2.907	1.453
4:0	73TrpHz2-32ValHgb*	3.032	3.638	2.425	1.213
5:0	107LysHa-108ValHga*	3.425	4.111	2.740	1.370
5:01	107LysHa-32ValHga*	3.425	4.111	2.740	1.370
5:02	107LysHa-53ValHb	3.425	4.111	2.740	1.370
5:03	107LysHa-67ValHgb*	3.425	4.111	2.740	1.370

Figure 7.20. Examples of ambiguous restraints generated for LV unlabelled SasYr **A**. F1fF2f NOESY where assignment were limited to leucine and valine residues and **B**. F1eF2f NOESY experiments where F2 assignments were limited to leucine and valine and F1 assignments were limited to every other residue.

```

label dataset sw sf
1H 1H
None
{11363.64 } {11363.64 }
{700.2733 } {700.2733 }
1H.L 1H.P 1H.W 1H.B 1H.E 1H.J 1H.U 1H.L 1H.P 1H.W 1H.B 1H.E 1H.J 1H.U vol int stat comment flag0
1 {77.HG1#} 0.809 0.050 0.050 ? 0.000 {?} {{103.HN; 34.HN; 42.HN; 80.HN}} 8.843 0.050 0.050 ? 0.000 {?} 2778584064.00000 424573952.00000 1 {?} 0
2 {{49.HG2#; 83.HD2#; 85.HG1#}} 0.929 0.050 0.050 ? 0.000 {?} {{108.HN; 110.HN; 45.HN; 61.HN}} 8.063 0.050 0.050 ? 0.000 {?} 7130694656.00000 1026394112.00000 1 {?} 0
3 {{49.HG2#; 83.HD2#; 85.HG1#}} 0.926 0.050 0.050 ? 0.000 {?} {{28.HG1; 36.HG1; 48.HG2; 67.HB; 75.HB; 79.HG1}} 2.256 0.050 0.050 ? 0.000 {?} 7484303360.00000 912590848.00000 1 {?} 0
4 {55.HG2#} 0.149 0.050 0.050 ? 0.000 {?} {55.HG1#} 0.364 0.050 0.050 ? 0.000 {?} 26804942848.00000 4076949504.00000 1 {?} 0
5 {{49.HG2#; 83.HD2#; 85.HG1#}} 0.928 0.050 0.050 ? 0.000 {?} {73.HZ2} 7.151 0.050 0.050 ? 0.000 {?} 11491436544.00000 1735620608.00000 1 {?} 0
6 {37.HA} 3.530 0.050 0.050 ? 0.000 {?} {{37.HN; 66.HN; 83.HN; 86.HN}} 8.577 0.050 0.050 ? 0.000 {?} 5619737600.00000 871582720.00000 1 {?} 0
7 {32.HG2#} 0.699 0.050 0.050 ? 0.000 {?} {106.HN} 9.223 0.050 0.050 ? 0.000 {?} 2489009152.00000 359173120.00000 1 {?} 0
8 {{37.HA; 77.HA}} 3.554 0.050 0.050 ? 0.000 {?} {73.HZ2} 7.152 0.050 0.050 ? 0.000 {?} 1867474944.00000 299741184.00000 1 {?} 0
9 {108.HB} 2.157 0.050 0.050 ? 0.000 {?} {{100.HE2; 41.HE2; 61.HE1; 70.HE2; 87.HB2}} 2.955 0.050 0.050 ? 0.000 {?} 3808830464.00000 478594048.00000 1 {?} 0
10 {{49.HG2#; 83.HD2#; 85.HG1#}} 0.935 0.050 0.050 ? 0.000 {?} {{105.HN; 37.HN; 66.HN; 86.HN}} 8.569 0.050 0.050 ? 0.000 {?} 9698983936.00000 1385076736.00000 1 {?} 0

```

Figure 7.21. A region of an ambiguously assigned peak list in NMRView format, as used for inputs for structure calculations with ARIA [186, 280]. The peak list shows the LV F2f ‘informed’ peak assignments, showing the potential assignments in the F1 and F2 dimension for each peak.



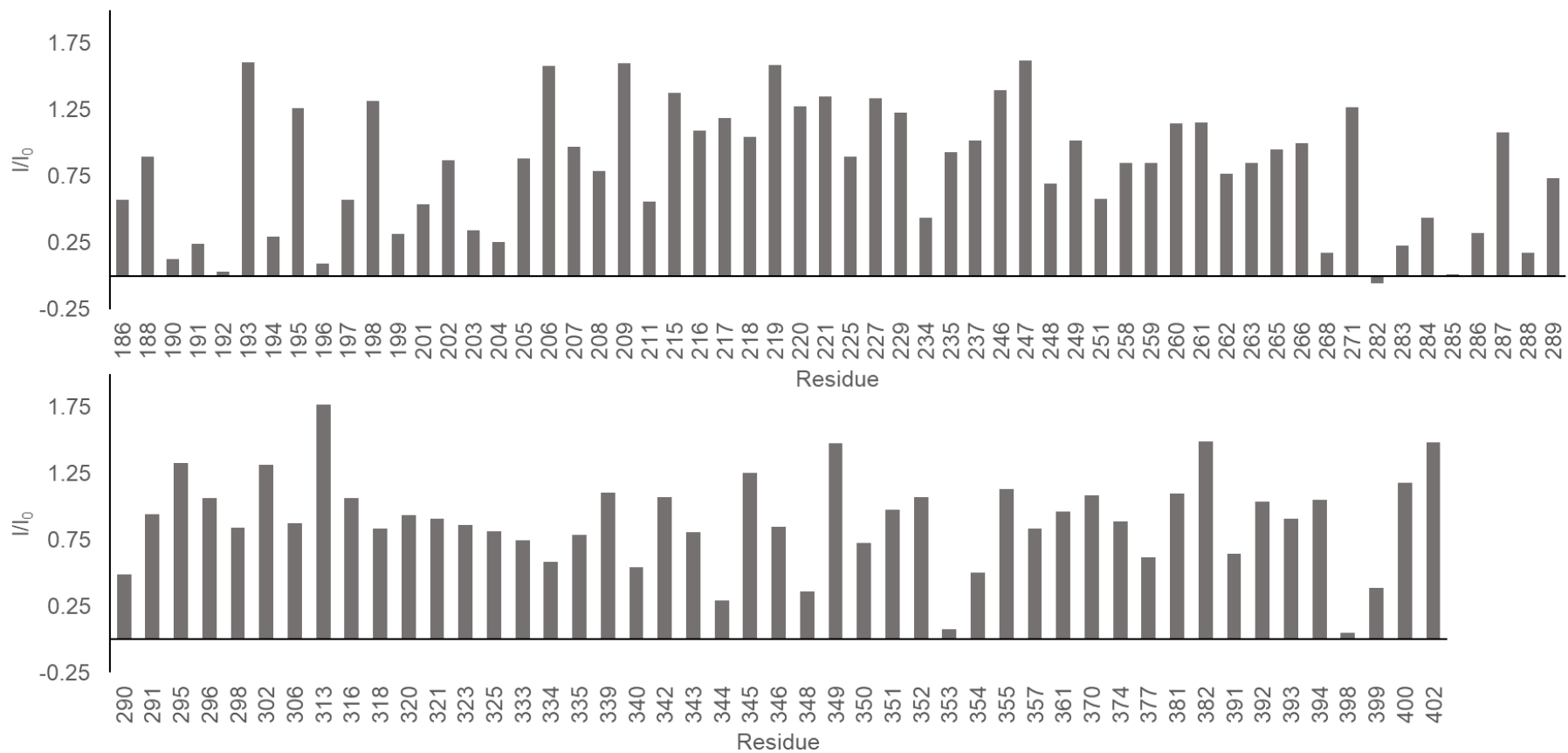


Figure 7.22. Graphs showing the intensity change for peaks in the spectra of PHD2 on addition of LIMD1 168-191. The Y axis shows the ratio of the signal-to-noise for peak intensities (peak heights) in the spectrum with LIMD1 168/191 divided by peak intensities in the apo spectrum.

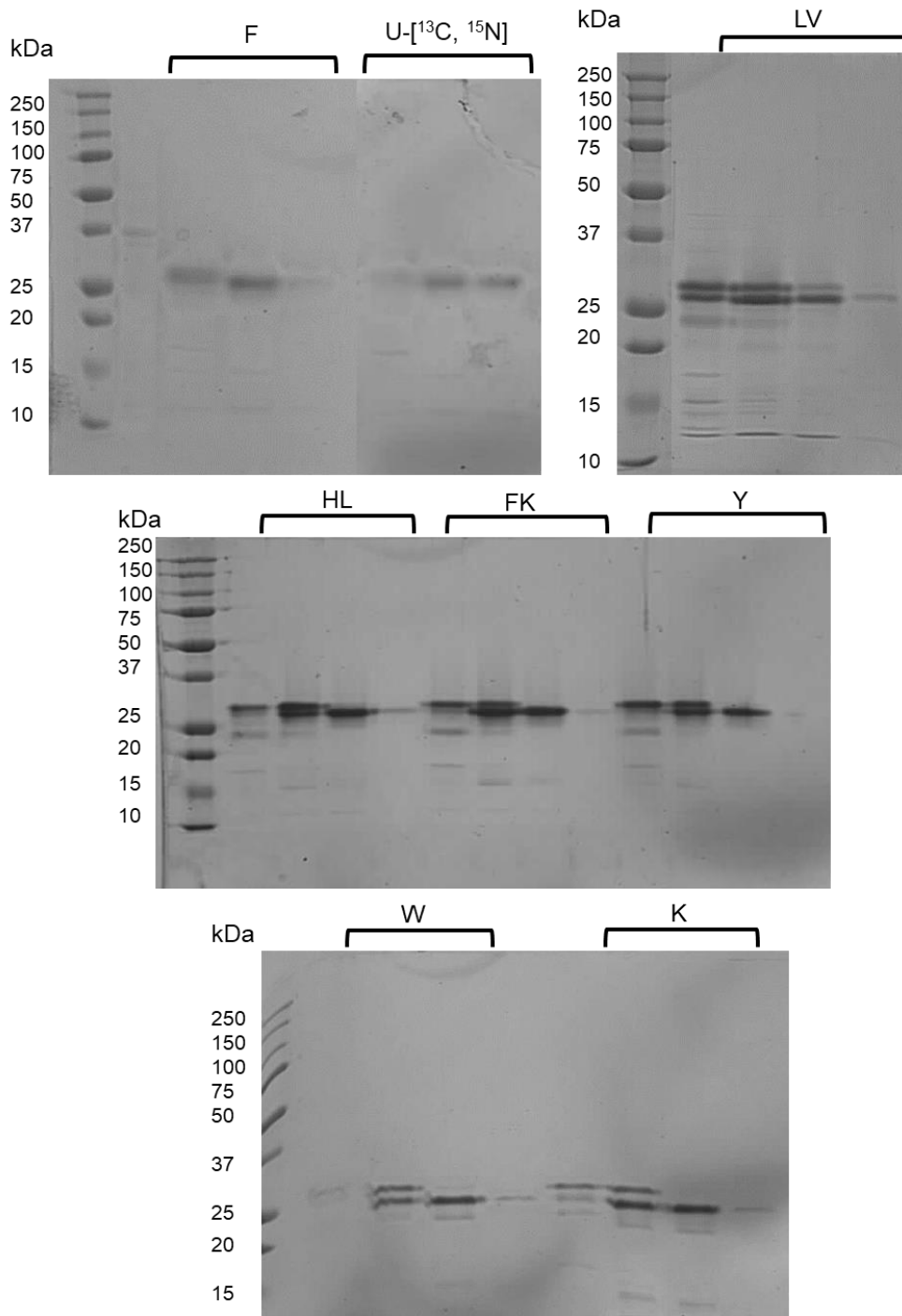


Figure 7.23. SDS-PAGE analysis of the SHIRT-R3 samples used. Each sample was run at 3 concentrations. 0.5 mg/mL, 0.25 mg/mL and 0.125 mg/mL on a 15% SDS-PAGE. The protein was calculated to be 11619.7 Da but is known to run aberrantly at ~26 kDa and often with multiple bands on SDS-PAGE.

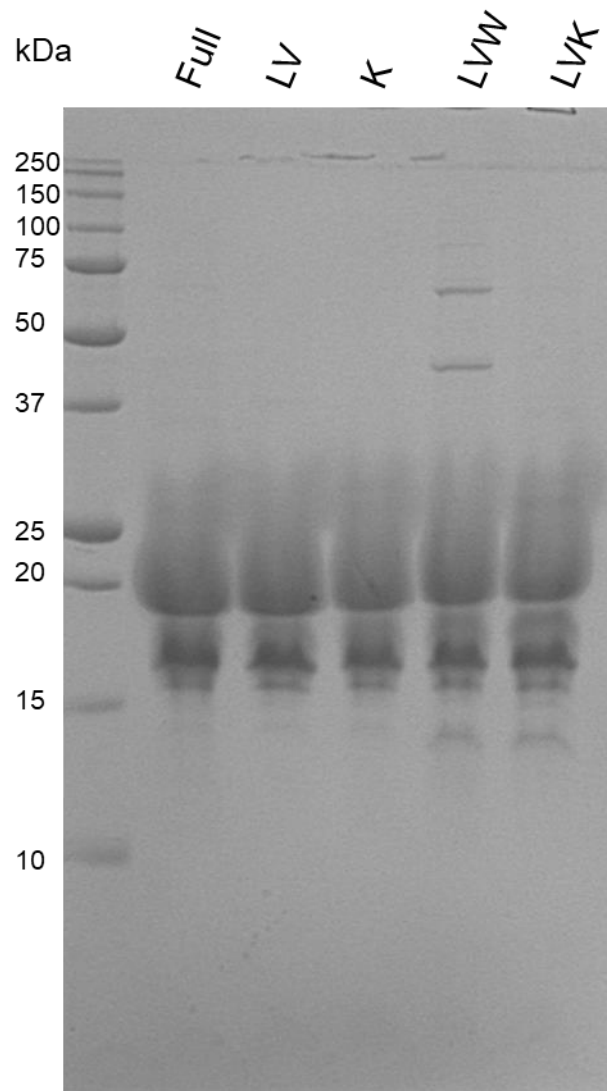


Figure 7.24. SDS-PAGE analysis of the purity of SasYr samples. Protein was used at 2 mg/mL and run on a 18% SDS-PAGE gel. SasYr was computed to be 9323.16 Da but is known to run aberrantly at ~20 kDa.

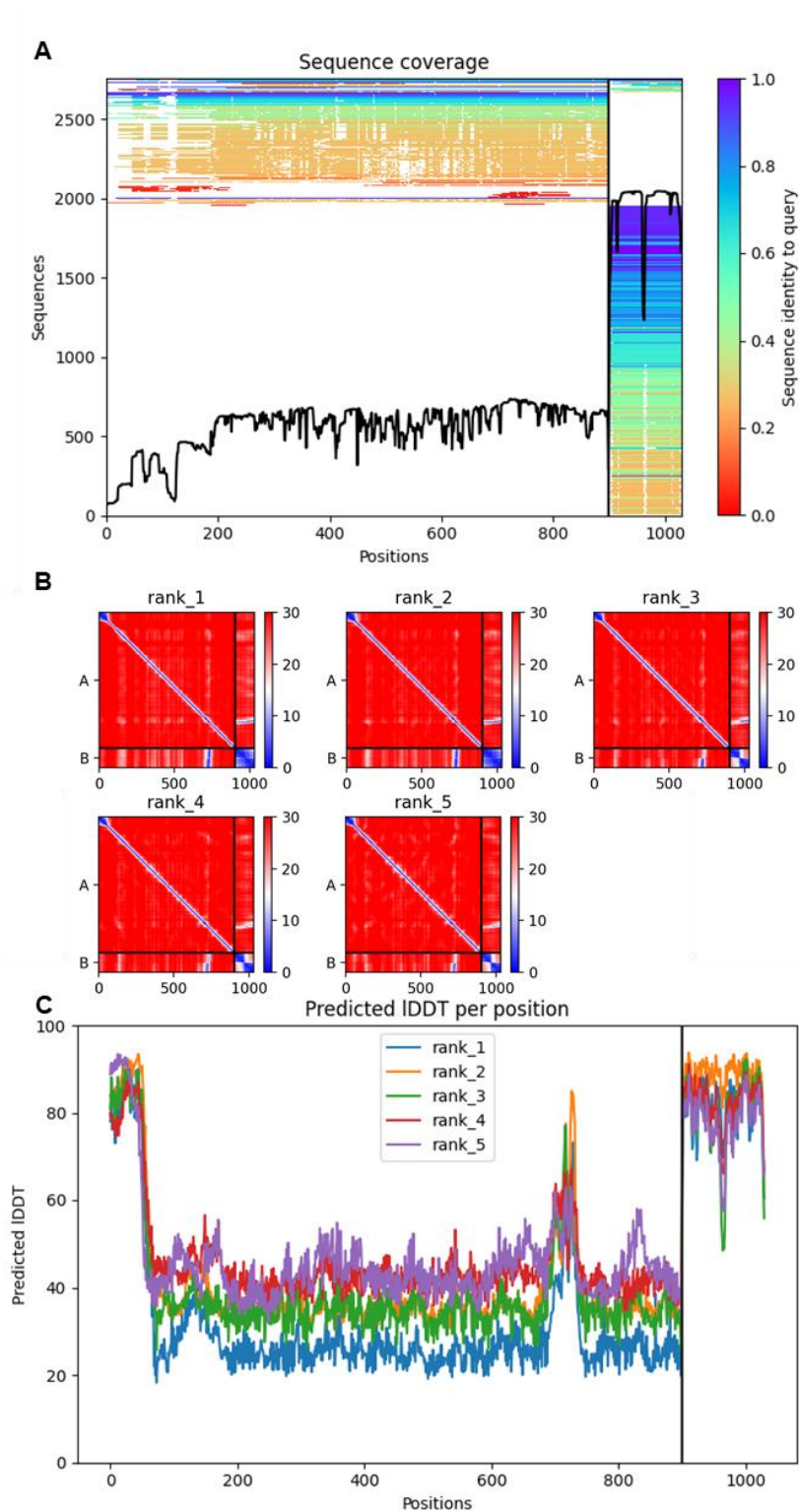


Figure 7.25. Statistic for the AlphaFold 2 prediction of TNRC6A 1-900 with LIM12 (LIMD1 467-594). **A.** sequence coverage in the MSA for TNRC6A 1-900 then LIM12. **B.** Predicted aligned error (PAE) plots for each model where A is TNRC6A 1-900 and B is LIM12. **C.** predicted local distance difference test (pLDDT) per residue for each model.

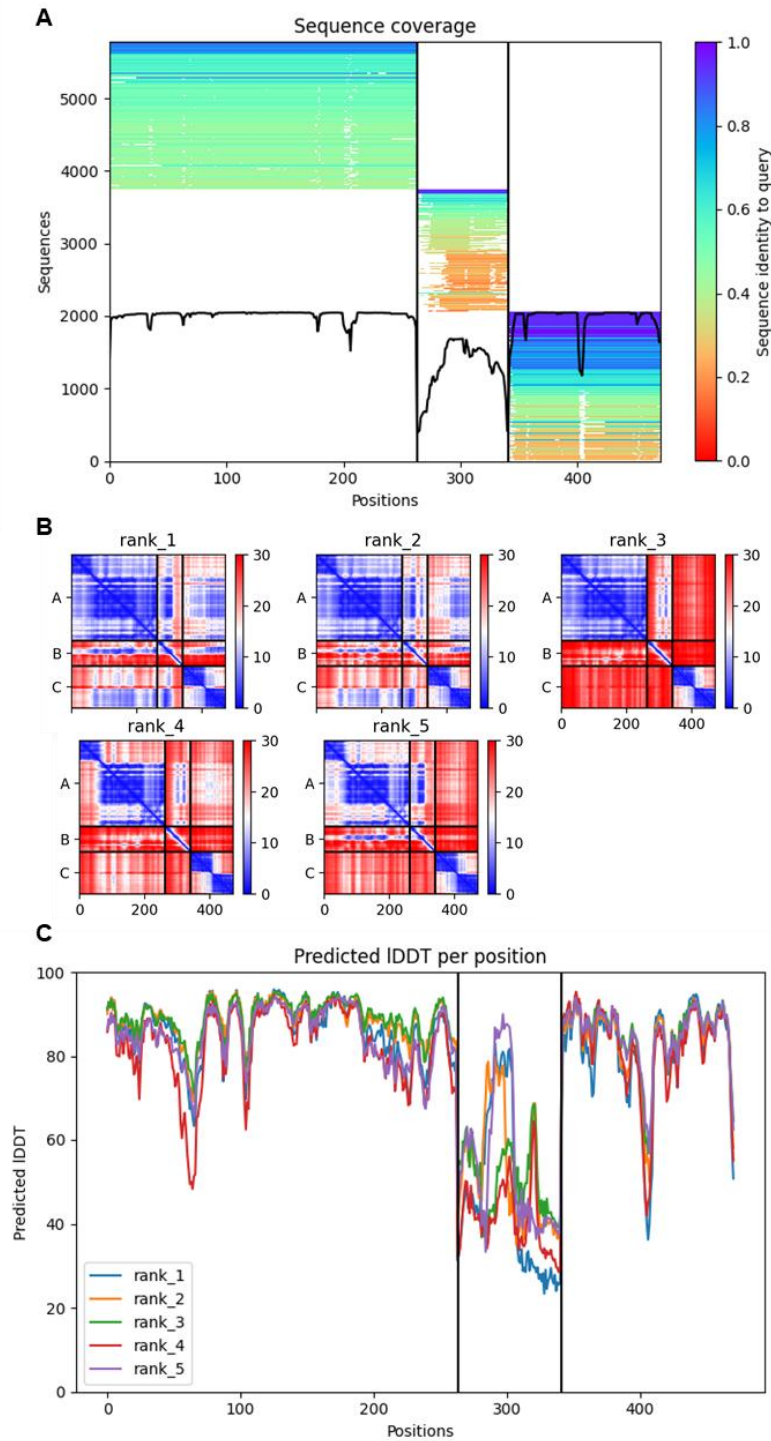


Figure 7.26. Statistic for the AlphaFold2 prediction of AGO2 519-779 (TNRC6A binding region) with TNRC6A 695-773 and LIM12 (LIMD1 467-594). **A.** sequence coverage in the MSA for AGO2 519-779, TNRC6A 695-773 then LIM12. **B.** Predicted aligned error (PAE) plots for each model where A is AGO2 519-779, B is TNRC6A 695-773 and C is LIM12. **C.** predicted local distance difference test (pLDDT) per residue for each model.

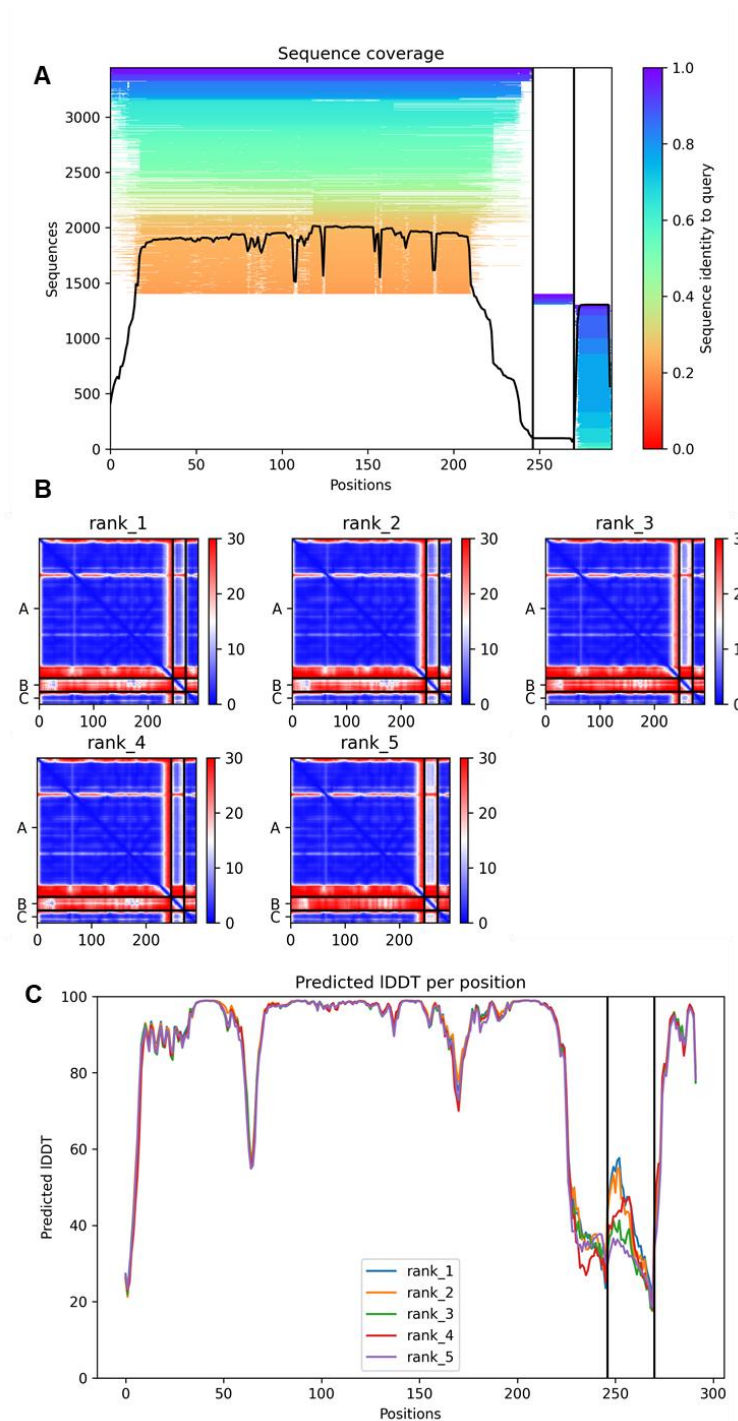


Figure 7.27. Statistics for the AlphaFold2 modelling of PHD2 181-426 with LIMD1 168-191 and HIF CODD. **A.** sequence coverage in the MSA for PHD2 181-426 then LIMD1 168-191 then HIF CODD. **B.** Predicted aligned error (PAE) plots for each model where A is PHD2 181-426, B is LIMD1 168-191 and C is HIF CODD. **C.** predicted local distance difference test (pLDDT) per residue for each model.

VHL 54-213


EMEAGRPRPVLRSVNS**REPSQVIFCNRS****PRVVL****PVWL**  
**NFDGEPQPYPTLPPGTG****RRIHSYRGHLWLF****RD**AGTHD  
GLLVNQTELFVPSLNVDGQPIFANITLPVYTLKERCLQVV  
**RSLVKPENYRR**LDIV**RS****LYEDLE****DHPNVQK**DLERLTQE  
RIAHQRMGD

Figure 7.28. The sequence of human VHL 54-213 produced in this study with peptides identified by MS in bold and red.



Review Article

# Specific isotopic labelling and reverse labelling for protein NMR spectroscopy: using metabolic precursors in sample preparation

Benjamin Rowlinson<sup>1</sup>, Elodie Crublet<sup>2</sup>, Rime Kerfah<sup>2</sup> and  Michael J. Plevin<sup>1</sup>

<sup>1</sup>York Structural Biology Laboratory, York Biomedical Research Institute, Department of Biology, University of York, York YO10 5DD, U.K.; <sup>2</sup>NMR-Bio, World Trade Center- 5 Place Robert Schuman, 38025 Grenoble Cedex 1, France

Correspondence: Michael J. Plevin (michael.plevin@york.ac.uk)



The study of protein structure, dynamics and function by NMR spectroscopy commonly requires samples that have been enriched ('labelled') with the stable isotopes <sup>13</sup>C and/or <sup>15</sup>N. The standard approach is to uniformly label a protein with one or both of these nuclei such that all C and/or N sites are in principle 'NMR-visible'. NMR spectra of uniformly labelled proteins can be highly complicated and suffer from signal overlap. Moreover, as molecular size increases the linewidths of NMR signals broaden, which decreases sensitivity and causes further spectral congestion. Both effects can limit the type and quality of information available from NMR data. Problems associated with signal overlap and signal broadening can often be alleviated though the use of alternative, non-uniform isotopic labelling patterns. Specific isotopic labelling 'turns on' signals at selected sites while the rest of the protein is NMR-invisible. Conversely, specific isotopic unlabelling (also called 'reverse' labelling) 'turns off' selected signals while the rest of the protein remains NMR-visible. Both approaches can simplify NMR spectra, improve sensitivity, facilitate resonance assignment and permit a range of different NMR strategies when combined with other labelling tools and NMR experiments. Here, we review methods for producing proteins with enrichment of stable NMR-visible isotopes, with particular focus on residue-specific labelling and reverse labelling using *Escherichia coli* expression systems. We also explore how these approaches can aid NMR studies of proteins.

## Introduction

Unlike many other spectroscopic techniques, NMR spectroscopy produces spectra in which it is possible to identify signals that correspond to specific atoms in the target molecule. NMR spectroscopy can, therefore, provide information about molecular structure, dynamics, interactions and biological function at atomic resolution even in molecules as large as proteins or nucleic acids. The full power of the technique is unlocked through the process of resonance assignment: That is, linking a signal in an NMR spectrum to a specific nucleus in the molecule. With assignments in hand, it is possible to determine information about local and global chemistry (i.e. structure, interactions, modifications, etc.) and how this changes with time (i.e. dynamics, kinetics, etc.), and to link this information to prior knowledge of the molecule (e.g. its basic chemical structure, experimental conditions, etc.). Resonance assignment requires being able to resolve individual signals in NMR spectra. Two major issues that complicate the assignment process are signal overlap (resolution) and signal-to-noise ratio (sensitivity). These two factors, which are often linked, can be addressed by combining specialised sample preparation techniques (i.e. isotopic labelling) with appropriate spectroscopic methods.

Received: 4 May 2022  
Revised: 18 October 2022  
Accepted: 24 October 2022

Version of Record published:  
16 November 2022



## Biomolecular NMR spectroscopy and low abundance stable isotopes

Spin- $\frac{1}{2}$  nuclei are particularly important for the NMR spectroscopy of biomacromolecules. The  $^1\text{H}$  isotope (natural abundance level: 99.985%) is spin- $\frac{1}{2}$ , which means that  $^1\text{H}$  NMR spectra can be recorded of proteins without the need for any special isotopic enrichment schemes. However, the dispersion of  $^1\text{H}$  chemical shifts typically found in spectra of proteins is relatively narrow and consequently it is common for  $^1\text{H}$  NMR spectra to suffer from signal overlap. In practice, it becomes challenging to resolve individual  $^1\text{H}$  NMR signals in NMR spectra of proteins larger than 10 kDa.

Spin- $\frac{1}{2}$  isotopes of carbon ( $^{13}\text{C}$ ) and nitrogen ( $^{15}\text{N}$ ) offer improved signal dispersion over  $^1\text{H}$ . However, both have low natural abundance ( $^{13}\text{C}$  at 1.1% and  $^{15}\text{N}$  at 0.05%) and it is, therefore, necessary to isotopically enrich proteins with these nuclei. Using a protein labelled with  $^{13}\text{C}$  and/or  $^{15}\text{N}$  affords better separation of signals by enabling the use of heteronuclear NMR experiments [1]. These experiments can correlate different sets of  $^1\text{H}$ ,  $^{13}\text{C}$  and/or  $^{15}\text{N}$  nuclei (spin systems) across different dimensions to generate multi-dimensional datasets, which can greatly reduce signal overlap seen in n-dimensional  $^1\text{H}$  spectra of proteins [1].

The combination of uniform isotopic labelling of proteins and multi-dimensional heteronuclear experiments is the foundation of modern biomolecular NMR spectroscopy. That said, the application of NMR spectroscopy to proteins is not always straightforward and it is often necessary to overcome significant obstacles.

### Too many signals

NMR spectroscopy can potentially detect signals from each NMR-visible nucleus in a protein, which can mean that even spectra of small proteins will have several hundred observable signals. This problem scales with the size of the protein: As molecular size increases so does the number of observable signals and the likelihood of nuclei having overlapping resonance frequencies.

Size is not the only issue that causes spectral congestion. Many proteins have low or reduced complexity sequences which can lead to signal overlap due to high numbers of residues of the same type or repeats of the same or similar sequence. Transmembrane proteins are often enriched in aliphatic and aromatic amino acids, and regions of NMR spectra where nuclei from these residues resonate can become highly congested [2]. Intrinsically disordered sequences can also present congested NMR spectra due to the lack of chemical shift dispersion that results from secondary and tertiary structure, and that fact that such sequences are often enriched in certain amino acid types and depleted of others [3]. Extreme examples are proteins carrying repeats of certain amino acid types, e.g. poly-glutamine stretches.

### Bigger is not always better

As well as issues relating to signal overlap, NMR spectra of larger proteins also suffer from reduced sensitivity. Larger molecules tumble more slowly in solution, which causes more rapid decay (relaxation) of the NMR signal, broadening of the signal linewidth and concomitant decrease in signal intensity (sensitivity). The principal problem here is that  $^1\text{H}$  nuclei efficiently relax other  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclei in their local vicinity. This effect increases with the size of the molecule and becomes highly detrimental for biomolecular NMR studies of proteins larger than 20–25 kDa.

The density of protons in a protein can be reduced by producing the sample in a deuterated expression medium. Deuteration replaces protons with deuterium atoms, which have a smaller impact on the relaxation rates of nearby nuclei [4]. Protein deuteration levels of up to 86% can be achieved using minimal media prepared in deuterium oxide ( $\text{D}_2\text{O}$ ) rather than water [5]. Should higher levels be required then all protons in the culture medium need to be replaced with deuterium, which can be achieved through the use of deuterated carbon sources (e.g. [ $^2\text{H}$ ,  $^{13}\text{C}$ ] glucose) [5]. While this approach successfully reduces the levels of  $^1\text{H}$  nuclei in proteins, it creates another problem: Protons are commonly used in NMR experiments to excite the system and to detect the final signal, and so retaining some  $^1\text{H}$  nuclei at select sites is beneficial. Protons can be selectively re-introduced into deuterated proteins by exposing the sample to water-based buffers (e.g. during purification) to facilitate the exchange of deuterium for protons in labile sites. Alternatively, protonated (i.e. natural abundance) molecules such as amino acids or metabolic precursors can be added to deuterated expression medium to introduce protons at specific sites in the protein (see below).

Combining deuteration with relaxation-optimised spectroscopic approaches such as TROSY can enable NMR characterisation of much larger proteins and protein assemblies, potentially exceeding MDa sizes (under favourable situations) [6,7].

## Isotopic labelling approaches for protein NMR spectroscopy

Many sample preparation approaches have been developed to alleviate issues of signal overlap and low sensitivity in the NMR spectra of proteins. At the base of most of these is the process of preparing a protein that is uniformly enriched in  $^{13}\text{C}$  and/or  $^{15}\text{N}$  nuclei. Protocols for uniform  $^{13}\text{C}$  and/or  $^{15}\text{N}$  enrichment of recombinant proteins produced in *Escherichia coli* are well established [8]. The isotopic labelling pattern of a recombinant protein over expressed from *E. coli* can be easily modified by adjusting the labelling pattern of the proton, nitrogen and carbon sources and by supplementing the medium with amino acids or metabolic precursors with the desired level and type of isotopic enrichment. These approaches can be applied to smaller proteins (<15 kDa) by over-expressing from media prepared in water, or to larger proteins (>20 kDa) by over-expression from media prepared with  $^2\text{H}_2\text{O}$ .

### Uniform or site-specific labelling with $^{15}\text{N}$ and/or $^{13}\text{C}$

An expression medium containing  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source would produce a  $[\text{U-}^{15}\text{N}]$ -labelled protein, i.e. a protein in which each nitrogen site is  $^{15}\text{N}$  labelled. A 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectrum of such a protein should show a single crosspeak for each residue (excluding proline) as well as crosspeaks for side-chain amine groups (Figure 1a). The enrichment pattern of the protein can be modified by adjusting the composition of the expression medium used to produce it. For example, supplementing unlabelled minimal medium with  $[\text{U-}^{15}\text{N}]$ -labelled lysine introduces  $^{15}\text{N}$  label at nitrogen sites in lysine residues [9]. A 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectrum of a protein with this labelling pattern would only show crosspeaks for each lysine NH group in the protein and hence be considerably simpler than that of a uniformly labelled sample (Figure 1a). Similar approaches can be considered for metabolic precursors of amino acids.

If a protein is produced using  $[\text{U-}^{13}\text{C}]$  glucose (CAS number of the unlabelled molecule: 50-99-7) as the sole carbon source, a 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectrum should show crosspeaks for each CH group in the protein (Figure 1b). Supplementing an unlabelled expression medium with 2- $^{13}\text{C}$ -methyl acetoacetate (CAS: 71698-08-3), which is a precursor in the biosynthesis of leucine and valine [10], produces a protein with  $[\text{H-}^{13}\text{C}]$  labelling of pro-S methyl groups. The resulting 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectrum is substantially simplified compared to the uniformly labelled protein as it only shows crosspeaks for pro-S methyl groups of leucine and valine (Figure 1b). All other CH sites are not  $^{13}\text{C}$  labelled and so are not observed in the spectrum.

### Selective reverse labelling of sites in $^{15}\text{N}$ and/or $^{13}\text{C}$ labelled proteins

An alternative approach is to reverse label specific residues or atoms in an otherwise uniformly labelled protein [11–13]. Supplementing a minimal medium containing  $^{15}\text{NH}_4\text{Cl}$  with unlabelled lysine produces a protein in which all nitrogen sites are  $^{15}\text{N}$  labelled except for those in lysine. The resulting 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectrum shows that crosspeaks corresponding to lysine residues have disappeared (Figure 1a). Likewise, supplementing  $^{13}\text{C}$  and/or  $^{15}\text{N}$  labelled media with unlabelled precursors can turn off signals of specific sets of atoms. For example, adding natural abundance phenylpyruvate (CAS: 114-76-1), a precursor of phenylalanine, to a medium containing  $[\text{U-}^{13}\text{C}]$ -glucose produces a  $[\text{H-}^{13}\text{C}]$ -labelled protein in which the side-chain groups of phenylalanine are unlabelled [12]. Comparing 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra of uniform and reverse labelled samples show select peaks disappear in the reverse labelled sample (Figure 1c).

### Isotopic labelling using amino acids and metabolites

Specific labelling and reverse labelling of residues and atom subsets is not uniformly applicable. Not every amino acid or precursor will be incorporated 'as is' into the target protein. Many amino acids and their precursors are metabolised by *E. coli*, which results in the scrambling of the isotopic labelling pattern of the molecule added to the culture medium. There has been considerable research over the last 30–40 years to identify amino acids and metabolic precursors that can be used to manipulate the isotopic enrichment pattern of recombinant proteins overexpressed in *E. coli* with no or minimal isotopic scrambling [4,14,15]. Only certain amino acids are compatible, typically those with an isolated biosynthesis pathway that includes one or more irreversible

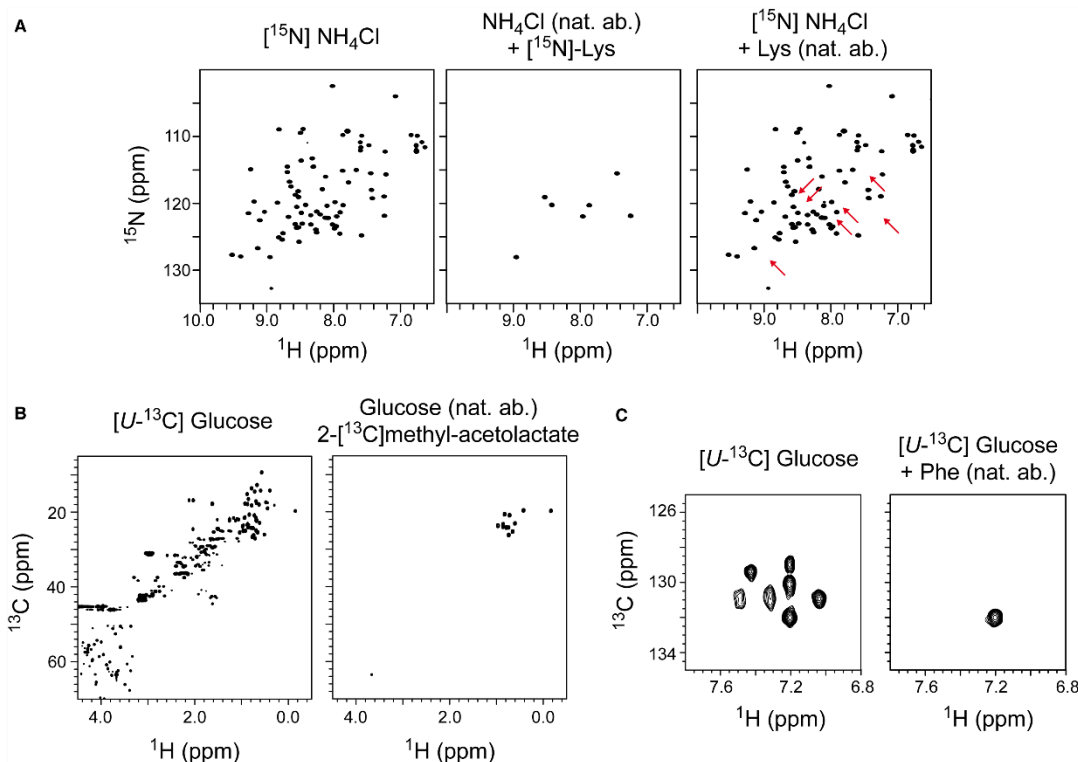


Figure 1. Turning NMR signals on and off using selective labelling or reverse labelling.

(a) 2D ( $^1\text{H}$ ,  $^{15}\text{N}$ ) HSQC spectra of ubiquitin prepared with [ $^{15}\text{N}$ ] labelled expression medium (uniform labelling; left), unlabelled medium supplemented with [ $^{15}\text{N}$ ] lysine (residue-specific labelling; centre), or [ $^{15}\text{N}$ ] labelled expression medium supplemented with unlabelled lysine (reverse labelling; right; red arrows indicate missing crosspeaks). (b) 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra of ubiquitin prepared with [ $^{13}\text{C}$ ] labelled expression medium (uniform labelling; left) or unlabelled medium supplemented with 2- [ $^{13}\text{C}$ ] methyl acetolactate (specific labelling of leucine and valine proS methyl groups; right). (c) 2D ( $^1\text{H}$ ,  $^{13}\text{C}$ ) HSQC spectra of ubiquitin prepared with [ $^{13}\text{C}$ ] labelled expression medium (uniform labelling; left) or [ $^{13}\text{C}$ ] labelled expression medium supplemented with unlabelled phenylpyruvate (reverse labelling of phenylalanine; right). Data taken from Rasia et al. 2012 (figure 1A,C) [12] and Gans et al. 2010 (figure 1B) [10].

step. Biosynthetic precursors have emerged as an alternative to using the full amino acid, as precursors often lack stereochemical sites that render the full amino acid expensive to synthesise. Below, we summarise current procedures for residue-specific isotopic labelling and reverse labelling of proteins.

#### Routes for targeting aliphatic residues

Aliphatic residues (Leu, Ile, Val and Ala) represent a highly desirable target for specific isotope labelling or reverse labelling due to their relatively high abundance (Leu: 9%, Ile: 5.2%, Val: 6.6% and Ala: 8.3%) and broad distribution across proteins [16,17]. Moreover, these residues contain methyl groups, which are excellent NMR probes for studying larger proteins [18].

Specific labelling and reverse labelling of branched-chain aliphatic amino acids has been used for generating backbone, side-chain and stereospecific assignments as well as for measuring NOEs [4,12,13,19–21].

The carbon atoms of leucine and valine can be labelled or reverse labelled with minimal scrambling through the use of the biosynthetic precursors  $\alpha$ -ketoisovalerate (CAS: 759-05-7) or acetolactate (CAS: 71698-08-3) (Figure 2) [4,10,12,22–24]. Both precursors are chemically synthesised as a racemic mixture, which impacts

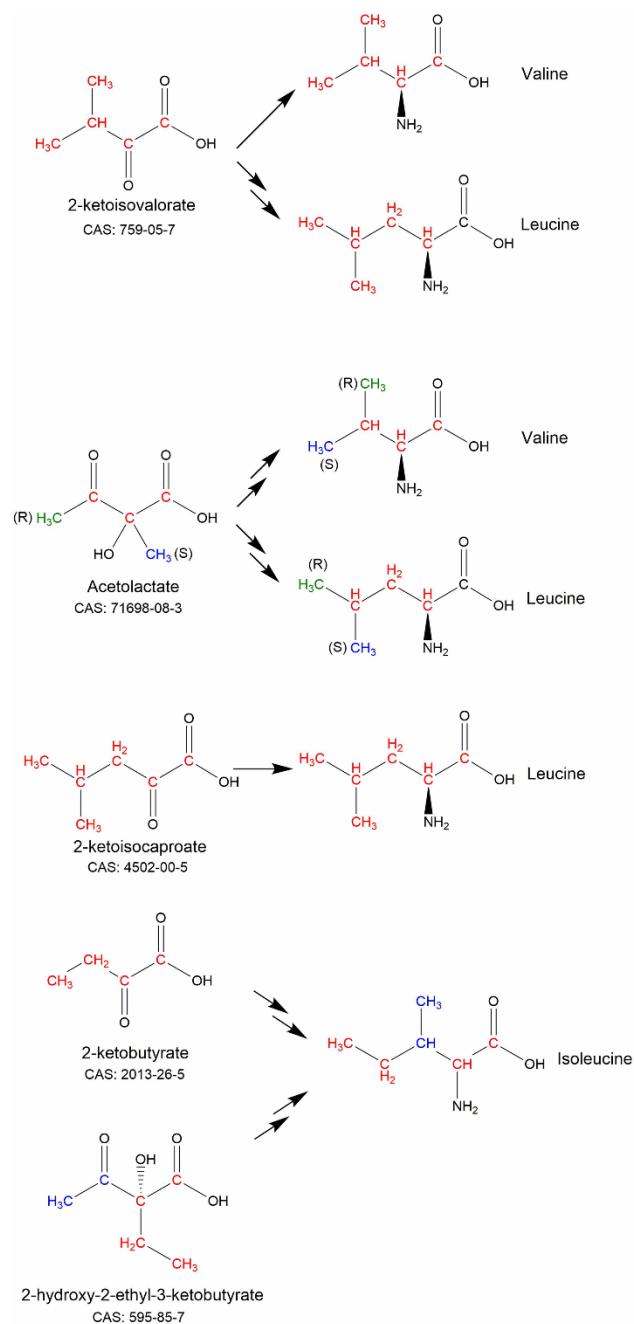


Figure 2. Metabolic precursors that can be used for isotopic labelling and reverse labelling of the carbon sites in branched-chain aliphatic amino acids. Sites corresponding to labelled or unlabelled groups are coloured to show their starting and end positions. The CAS numbers of the unlabelled precursors are given.



how they label prochiral methyl groups. Only the 2S stereoisomer of acetolactate is a substrate of ketol-acid reductoisomerase (EC: 1.1.1.86), which means that it is possible to use acetolactate to stereospecifically target the prochiral methyl groups of leucine and valine [10,19].  $\alpha$ -ketoisovalerate can be used for applications that require (or can tolerate) labelling of both prochiral methyl groups. Leucine alone can be labelled by the addition of 2-ketoisocaproate (CAS: 4502-00-5) (Figure 2), a precursor of leucine that sits after the divergence of leucine and valine biosynthesis [4,25]. Selectively deuterated versions of these precursors can be used to label larger proteins in combination with deuterated glucose and  $^2\text{H}_2\text{O}$  [10,25].

The carbon atoms of isoleucine can be labelled with 2-ketobutyrate (CAS: 2013-26-5) or 2-hydroxy-2-ethyl-3-ketobutyrate (CAS: 595-85-7) (Figure 2). These molecules also differentially target the methyl groups of Ile: 2-ketobutyrate is used to selectively label the Ile- $\delta_1$  methyl group, while 2-(S)-hydroxy-2-ethyl-3-ketobutyrate can be used to target Ile- $\delta_1$  and/or Ile- $\gamma_2$  methyl groups [26–28].

Biosynthetic precursors that target isoleucine, leucine and valine can be used in combination with other metabolites or amino acids which suppress scrambling of carbon sites and off-target effects. For instance, prochiral methyl groups of valine can be selectively labelled by the addition of labelled pro-R acetolactate- $^{13}\text{C}_4$  (CAS: 71698-08-3) or pro-S acetolactate- $^{13}\text{C}_3$  together with L-Leucine at natural abundance [29]. More details on methyl labelling of isoleucine, leucine and valine can be found in Kerfah et al. 2015 [4] and Schultz and Sprangers 2020 [6].

The methyl group of alanine provides a good probe for monitoring the local structure and dynamics of the protein backbone [30]. Specific isotopic labelling or reverse labelling of alanine is hampered by the presence of alanine transaminases which convert alanine into the widely used metabolite pyruvate [9]. Pyruvate is an early precursor in isoleucine, valine and leucine biosynthesis, which means that the labelling pattern of alanine will be scrambled into other aliphatic residues in the target protein. The addition of 1 g/L natural abundance alanine (CAS: 56-41-7) will result in approximately 50% loss of signal from valine [31]. That said, scramble-free  $^{13}\text{C}$  labelling of the carbonyl of alanine is possible via the addition of 1- $^{13}\text{C}$ -alanine [32], while labelling of the alanine methyl group can be achieved by adding other metabolites to suppress crosstalk between biosynthesis pathways [33,34]. In principle, reverse labelling can be achieved using similar approaches by adding alanine at the natural abundance and other precursors with  $^{13}\text{C}$  labelling, though this would be expensive and impractical. [ $^{15}\text{N}$ ] labelling or reverse labelling of the backbone amine group of alanine, isoleucine, leucine and valine is problematic due to the action of various transaminases [35].

### Routes for targeting aromatic residues

Aromatic residues are found at interaction interfaces and in the hydrophobic core of proteins and hence can serve as excellent reporters of protein structure, dynamics and interactions [36]. Tryptophan can be used as a sole carbon source by *E. coli* and so significant scrambling of carbon atoms occurs when tryptophan is added to the culture medium [37]. Additionally, tryptophanase (EC:4.1.99.1) can convert tryptophan to indole, pyruvate and ammonia which leads to nitrogen scrambling as ammonia is used in amino acid synthesis [31]. Aromatic amino acid transaminases cause significant nitrogen scrambling between tyrosine and phenylalanine when attempting to label or reverse label with either amino acid [31].

An early example of the selective labelling of aromatic residues through metabolic precursors was the use of shikimic acid (CAS: 138-59-0) to label aromatic protons of phenylalanine, tyrosine and tryptophan against a deuterated background (Figure 3) [38]. However, the synthesis of isotopically labelled shikimic acid is complicated, which has precluded widespread use. Phenylpyruvate (CAS: 156-06-9) and 4-hydroxy phenylpyruvate (CAS: 156-39-8) have been used to reverse label the carbon atoms phenylalanine and tyrosine, respectively (Figure 3) [12].  $^{13}\text{C}$  labelled versions of these precursors were later reported for isotopic labelling of phenylalanine and tyrosine [39]. Indole (CAS: 120-72-9) can be used for selective tryptophan labelling and reverse labelling of the tryptophan side chain (Figure 3) [40]. Isotopic labelling and reverse labelling of tryptophan can also be achieved through the use of indolepyruvate (CAS: 392-12-1) (Figure 3), which is part of the tryptophan degradation pathway rather than the biosynthesis pathway [41]. Anthranilic acid (CAS: 118-92-3) (Figure 3) has also been reported as an alternative tryptophan labelling precursor, which allows both  $^{15}\text{N}$  and  $^{13}\text{C}$  labelling of side-chain sites with minimal scrambling [42].

Histidine can be labelled or reverse labelled by the addition of the amino acid itself to the expression media. Histidine can also be labelled, without scrambling, by the metabolic precursor imidazolepyruvate (CAS: 2504-83-8; Figure 4) [43].

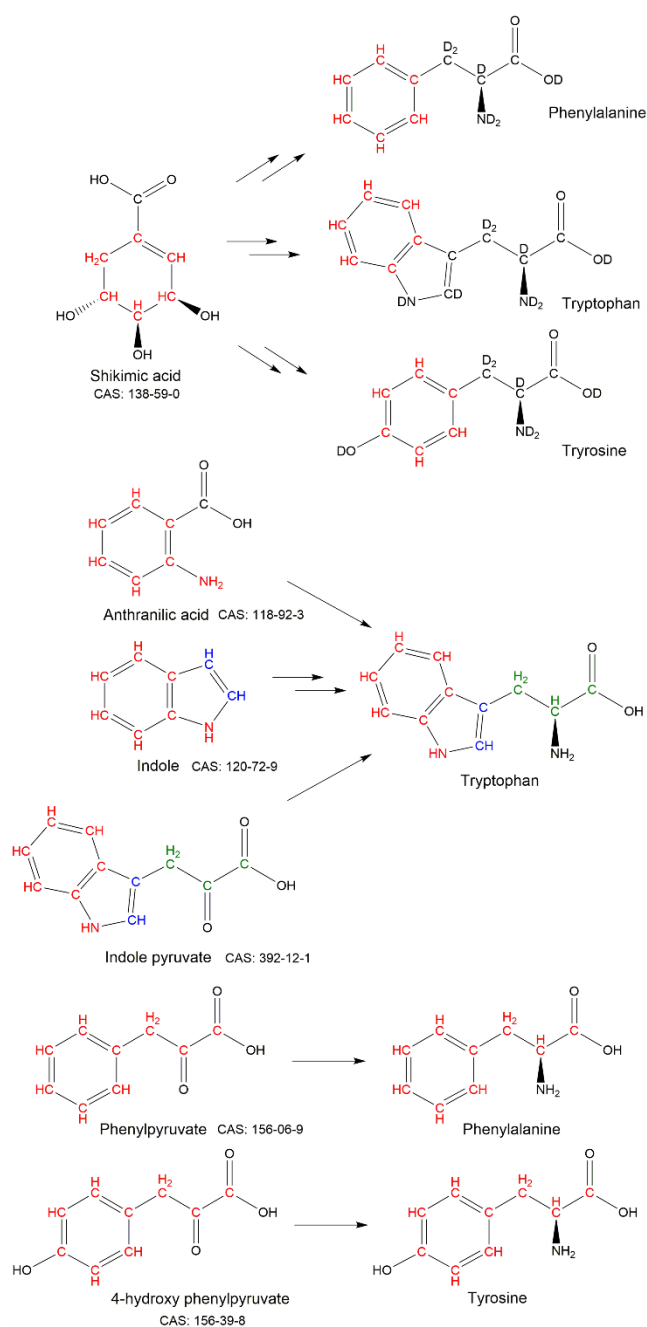


Figure 3. Metabolic precursors used for isotopic labelling and reverse labelling of phenylalanine, tryptophan and tyrosine. Sites corresponding to labelled or unlabelled groups are coloured to show their starting and end positions. The CAS numbers for the unlabelled precursors are given.

Downloaded from <http://portlandpress.com/biochemsoctrans/article-pdf/doi/10.1042/BST20210586/939564/bst-2021-0586c.pdf> by UK user on 12 December 2022

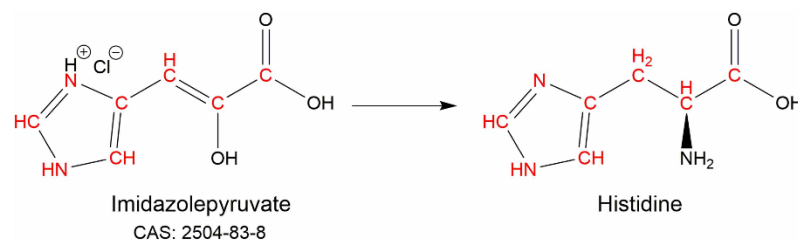


Figure 4. Histidine (un)labelling by the metabolic precursor imidazolepyruvate with incorporated atoms shown in red. The CAS number for the unlabelled precursor is given.

### Routes for targeting polar residues

Serine is connected to glycine via a glycine-hydroxymethyltransferase (EC: 2.1.2.1), which is in turn linked to threonine. Serine is also a precursor of tryptophan and cysteine biosynthesis and can be converted into pyruvate by serine dehydratase (EC: 4.3.1.17). Currently, there is no protocol for scramble-free specific labelling or reverse labelling of serine using traditional expression hosts. Similarly, cysteine is converted to pyruvate by cysteine desulfhydrases (EC: 4.4.1.28) which leads to significant scrambling when cysteine is added to the culture medium.

Threonine is connected to glycine, serine, cysteine, tryptophan and isoleucine, which leads to significant scrambling for nitrogen labelling or reverse labelling [44]. For labelling or reverse labelling of carbon sites, threonine is connected to isoleucine and glycine biosynthesis and can cause significant scrambling when added to the expression media. This scrambling effect has been overcome by the addition of the isoleucine precursor 2-ketobutyrate (CAS: 2013-26-5) (or isoleucine) and glycine to the expression media [44,45].

Asparagine and glutamine are particularly difficult amino acids to specifically label or reverse label. Specific  $^{15}\text{N}$  labelling of these amino acids has been achieved through the use of media supplemented with  $^{15}\text{NH}_4\text{Cl}$  and all amino acids at natural abundance apart from asparagine or glutamine [46]. Specific labelling of side-chain sites of these residues using metabolic precursors has not been possible due to their position in metabolic pathways. Asparagine and glutamine synthesis is closely linked to aspartate and glutamate synthesis, which are used in the synthesis of many amino acids. In addition, glutamate is the primary nitrogen donor in amino acid biosynthesis. This means that scramble-free specific labelling of these amino acids with the residues themselves or their precursors is not possible without the addition of a full amino acid complement to the media or the use of auxotrophic strains or cell-free systems [37,47].

### Routes for targeting charged residues

The final steps of both lysine and arginine biosynthesis are irreversible, which means that both amino acids can be used directly for labelling and reverse labelling (Figure 1a), thus negating the need for supplemental precursors to reduce isotopic scrambling.

For reasons discussed above, scrambling free specific labelling of aspartate and glutamate either with the amino acids themselves or with metabolic precursors has not been achieved.

### Special cases

Methionine is commonly used to introduce methyl labelled probes for NMR analyses of proteins [7,48,49]. The relatively low abundance of methionine in proteins (2.4%) can reduce the chance of spectral overlap than with aliphatic residues [17]. Methionine can be both isotopically labelled and reverse labelled with minimal scrambling by the addition of the amino acid itself to the media [31,50]. An alternative approach uses the metabolic precursor methylthio-2-oxobutanoate (CAS: 583-92-6) (Figure 5) for labelling without nitrogen [51].

Glycine is linked to serine and threonine by glycine-hydroxymethyltransferase (EC: 2.1.2.1) and threonine aldolase (EC: 4.1.2.5), respectively, meaning extensive scrambling for both carbon and nitrogen sites occurs when attempting to label or reverse label with glycine.

Proline can be used as a sole carbon and nitrogen source in bacterial cell culture and so induces significant scrambling when added to the media [12,31,37,52]. Metabolic precursors of proline have not been used for the

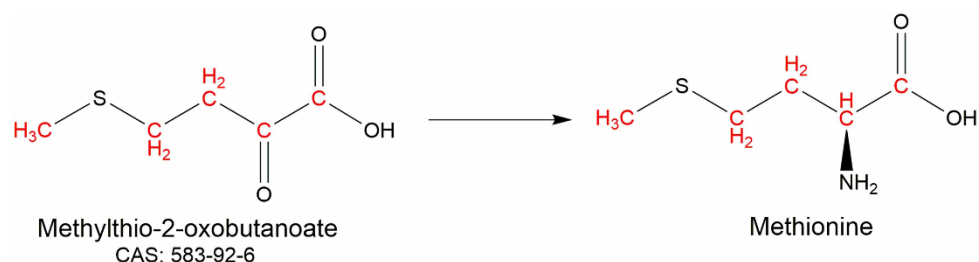


Figure 5. Methionine (un)labelling by methylthio-2-oxobutanoate with isotopically labelled sites indicated in red.

production of proteins with proline labelling or reverse labelling. The interconnectivity of the proline and other amino acid biosynthetic pathways makes this unlikely.

## Discussion

In this review, we have summarised approaches for manipulating the isotopic labelling patterns of recombinant proteins for NMR studies. We have focused on approaches that allow labelling or reverse labelling of specific sets of atoms or residues and discussed how this can be achieved through the addition of amino acids or their precursors to bacterial cell culture media. We have highlighted applications to solution NMR spectroscopy, but the labelling approaches described would also benefit solid-state NMR spectroscopy of proteins.

Extensive research into isotopic labelling protocols means that today's protein NMR scientist can make use of a wide range of labelling and reverse labelling schemes and enrichment patterns. The task now is to utilise these schemes to the greatest effect. In addition to helping with spectral crowding and line broadening, specific isotope labelling or reverse labelling can provide residue-type assignment. These approaches also place site-specific probes into protein that can report on the structure, dynamics and binding (often bypassing the need for full resonance assignment) [12]. For example, a study of the interaction of yeast ubiquitin hydrolase with ubiquitin used a sample prepared with specific  $^{13}\text{C}$  labelling of Met-C $\epsilon$ , Ala-C $\beta$ , His-C $\epsilon$ , Tyr-C $\epsilon$  and Trp-C $\delta$  and  $^{15}\text{N}$  labelling of Arg backbone amide groups [53]. Residues involved at the interface were predicted based on the number and type of amino acid chemical shifts that were perturbed on complex formation. These assignment predictions were then used as an input for computational docking using HADDOCK [54] and the resulting models were consistent with a crystal structure of a related complex. The potential of site-specific isotopic labelling or reverse labelling to provide useful information without full assignments can be particularly useful where full assignment may not be possible or be too time consuming to obtain. In the post-AlphaFold2 world [55], lower levels of resonance assignment will likely often be sufficient to confirm a predicted structure or to link protein structure to function [56].

This review has exclusively focussed on the over-expression of recombinant proteins using standard *E. coli*-based techniques, i.e. cytosolic protein production under the control of an IPTG-inducible T7-based expression system. However, not all proteins can be produced using *E. coli*-based approaches and eukaryotic hosts are often required. There has been considerable progress in isotopic labelling using eukaryotic cell types, including yeast, insect and mammalian cells [57–63]. Differences in the metabolic processing of amino acids and their precursors, and different requirements for cell culture media have meant that the production of labelled proteins using eukaryotic systems is still not widely reported in the literature. An overriding concern for isotopic labelling in insect or mammalian cells is the complexity of the cell culture medium compared to the minimal recipes that can be used for *E. coli*. Broadly speaking, modifying eukaryotic cell culture media composition to support isotopic labelling is considerably more expensive than bacterial cell culture. That said, a number of impressive studies have been reported that use eukaryotic expression systems to generate labelled proteins for NMR studies [60,64].

*In vitro* or 'cell free' protein synthesis is an alternative and highly adaptable approach for producing labelled protein, providing it is compatible with the protein of interest. Working with an S30 cell extract significantly reduces issues from metabolic scrambling compared to protein expression from live cells, and allows residue-specific labelling of each of the 20 proteinogenic amino acids [65,66]. Issues remain with the scrambling of Asp, Asn, Gln and Glu but these can be alleviated through the use of small molecule inhibitors or by using an auxotrophic strain of *E. coli* to produce the cell extract [67,68]. Moreover, the smaller scale of *in vitro* protein



synthesis reactions compared to cell culture reduces the costs of reagents. Consequently, the use of amino acids with isotopic enrichment patterns that are complicated or expensive to synthesise becomes feasible because only very low milligram-level quantities are required [69]. Lastly, in vitro protein synthesis using amber stop codons and pre-charged tRNA molecules can allow isotopic labelling of a single residue position in the protein of interest [66,70].

NMR spectra can be extremely rich in information. A drawback from selective labelling or reverse labelling of proteins is that it greatly reduces the number of NMR reporters in the target molecule, which can limit some applications and the types of question that can be addressed. Despite that, there are multiple examples of where specific labelling of proteins for NMR studies has considerable benefit over uniform labelling. Ultimately, the isotopic labelling pattern chosen for a particular protein target will depend on the parameters of the system being studied and the biological questions of interest. Many projects do not require a high number of NMR-visible sites or complete resonance assignment to provide relevant information. As NMR spectroscopy moves away from being a method for solving the 3D structures of proteins and embraces a role as a site-resolved spectroscopic technique, quick and efficient access to site-specific NMR probes will become more and more important. Specific isotopic labelling or reverse labelling can provide these important site-specific NMR-visible probes.

## Perspectives

- Uniform enrichment of recombinantly produced samples with  $^{13}\text{C}$  and/or  $^{15}\text{N}$  isotopes is a fundamental platform for NMR studies of proteins. Approaches that allow site-specific modulation of the enrichment pattern of a protein can simplify the process of data analysis, provide resonance assignment information and unlock experimental strategies for NMR analysis of structure, dynamics and function.
- Amino acids and/or their biosynthetic precursors can be added to bacterial culture media to achieve site- or residue-specific labelling or reverse labelling of proteins either alone or in combination with standard  $^{13}\text{C}$  and  $^{15}\text{N}$  labelling protocols. We review the approaches for achieving different isotopic labelling patterns, outline the motivations for employing these approaches and discuss the experimental benefits for their use.
- Combining experiments that exploit proteins with site-specific labelling patterns with the improved prediction of 3D structure now available via tools such as AlphaFold promises to greatly broaden the range of targets and biological questions that can be addressed by NMR spectroscopy.

### Competing Interests

NMR-Bio manufacture reagents for isotopic labelling of samples for NMR studies of proteins. The authors declare no other competing interests associated with this manuscript.

### Funding

B.R. receives PhD funding from White Rose BBSRC DTP (BB/M011151/1). M.J.P. acknowledges funding from EPSRC (EP/W024063/1).

### Open Access

Open access for this article was enabled by the participation of the University of York in an all-inclusive Read & Publish agreement with Portland Press and the Biochemical Society under a transformative agreement with JISC.

### Author Contributions

B.R. and M.J.P. outlined the review. B.R. wrote the first draft and made the figures. M.J.P., E.C. and R.K. revised the text.

## References

- 1 Sattler, M., Schleucher, J. and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93–158 [https://doi.org/10.1016/S0079-6565\(98\)00025-9](https://doi.org/10.1016/S0079-6565(98)00025-9)
- 2 Damraoui, G.I. and Hwang, P.M. (2020) Solution NMR spectroscopy of membrane proteins. *Biochim. Biophys. Acta-Biomembr.* 1862, 183356 <https://doi.org/10.1016/j.bbmem.2020.183356>
- 3 Konrat, R. (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* 241, 74–85 <https://doi.org/10.1016/j.jmr.2013.11.011>
- 4 Kerfah, R., Plevin, M.J., Soumier, R., Gans, P. and Boisbouvier, J. (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* 32, 113–122 <https://doi.org/10.1016/j.sbi.2015.03.009>
- 5 Leiting, B., Marsilio, F. and O'Connell, J.F. (1998) Predictable deuteration of recombinant proteins expressed in *Escherichia coli*. *Anal. Biochem.* 265, 351–355 <https://doi.org/10.1006/abio.1998.2904>
- 6 Schutz, S. and Sprangers, R. (2020) Methyl TROSY spectroscopy: a versatile NMR approach to study challenging biological systems. *Prog. Nucl. Magn. Reson. Spectrosc.* 116, 56–84 <https://doi.org/10.1016/j.pnmrs.2019.09.004>
- 7 Mas, G., Guan, J.Y., Grubler, E., Dabbed, E.C., Moriscot, C., Gans, P. et al. (2018) Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Sci. Adv.* 4, eaau4196 <https://doi.org/10.1126/sciadv.aau4196>
- 8 McIntosh, L.P. and Dahlquist, F.W. (1990) Biosynthetic incorporation of  $^{15}\text{N}$  and  $^{13}\text{C}$  for assignment and interpretation of nuclear-magnetic-resonance spectra of proteins. *Q. Rev. Biophys.* 23, 1–38 <https://doi.org/10.1017/s0033583500005400>
- 9 Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) Expression and  $^{15}\text{N}$  labeling of proteins for proton and  $^{15}\text{N}$  nuclear-magnetic-resonance. *Methods Enzymol.* 177, 44–73 [https://doi.org/10.1016/0076-6879\(89\)77005-1](https://doi.org/10.1016/0076-6879(89)77005-1)
- 10 Gans, P., Hamelin, O., Soumier, R., Ayala, I., Dura, M.A., Amaro, C.D. et al. (2010) Site-specific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew. Chem-Int. Ed.* 49, 1958–1962 <https://doi.org/10.1002/anie.200905660>
- 11 Krishnarajna, B., Jaipuria, G., Thakur, A., D'Silva, P. and Ahreja, H.S. (2011) Amino acid selective unlabeled for sequence specific resonance assignments in proteins. *J. Biomol. NMR* 49, 39–51 <https://doi.org/10.1007/s10858-010-9459-z>
- 12 Rasia, R.M., Brutscher, B. and Plevin, M.J. (2012) Selective isotopic unlabeled of proteins using metabolic precursors: application to NMR assignment of intrinsically disordered proteins. *ChemBioChem* 13, 732–739 <https://doi.org/10.1002/cbic.201100678>
- 13 Vuster, G.W., Kim, S.J., Wu, C. and Bax, A. (1994) 2D and 3D NMR-study of phenylalanine residues in proteins by reverse isotopic labeling. *J. Am. Chem. Soc.* 116, 9206–9210 <https://doi.org/10.1021/ja00099a041>
- 14 Schorghuber, J., Geist, L., Platzer, G., Feichtinger, M., Bisaccia, M., Scheibeleberger, L. et al. (2018) Late metabolic precursors for selective aromatic residue labeling. *J. Biomol. NMR* 71, 129–140 <https://doi.org/10.1007/s10858-018-0188-z>
- 15 Lacabanne, D., Meier, B.H. and Bockmann, A. (2018) Selective labeling and unlabeled strategies in protein solid-state NMR spectroscopy. *J. Biomol. NMR* 71, 141–150 <https://doi.org/10.1007/s10858-017-0156-z>
- 16 Janin, J. (1979) Surface and inside volumes in globular proteins. *Nature* 277, 491–492 <https://doi.org/10.1038/277491a0>
- 17 McCaldon, P. and Argos, P. (1988) Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide-sequences. *Proteins* 4, 99–122 <https://doi.org/10.1002/prot.340040204>
- 18 Ruschak, A.M. and Kay, L.E. (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. *J. Biomol. NMR* 46, 75–87 <https://doi.org/10.1007/s10858-009-9376-1>
- 19 Kerfah, R., Hamelin, O., Boisbouvier, J. and Marion, D. (2015)  $\text{CH}_3$ -specific NMR assignment of alanine, isoleucine, leucine and valine methyl groups in high molecular weight proteins using a single sample. *J. Biomol. NMR* 63, 389–402 <https://doi.org/10.1007/s10858-015-9998-4>
- 20 Tugarinov, V., Choy, W.Y., Orehkov, V.Y. and Kay, L.E. (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl Acad. Sci. U.S.A.* 102, 622–627 <https://doi.org/10.1073/pnas.0407792102>
- 21 Sprangers, R. and Kay, L.E. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445, 618–622 <https://doi.org/10.1038/nature05512>
- 22 Tugarinov, V. and Kay, L.E. (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR* 28, 165–172 <https://doi.org/10.1023/B:JNMR.0000013824.93994.1f>
- 23 Lichteneder, R., Ludwiczek, M.L., Schmid, W. and Konrat, R. (2004) Simplification of protein NCEST spectra using biorganic precursor synthesis and NMR spectral editing. *J. Am. Chem. Soc.* 126, 5348–5349 <https://doi.org/10.1021/ja049679n>
- 24 Tugarinov, V. and Kay, L.E. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc.* 125, 13868–13878 <https://doi.org/10.1021/ja030345s>
- 25 Lichteneder, R.J., Coudaville, N., Konrat, R. and Schmid, W. (2013) Selective isotope labelling of leucine residues by using alpha-ketoacid precursor compounds. *ChemBioChem* 14, 818–821 <https://doi.org/10.1002/cbic.201200737>
- 26 Ruschak, A.M., Velyvis, A. and Kay, L.E. (2010) A simple strategy for  $^{13}\text{C}$ ,  $^1\text{H}$  labeling at the Ile-gamma 2 methyl position in highly deuterated proteins. *J. Biomol. NMR* 48, 129–135 <https://doi.org/10.1007/s10858-010-9449-1>
- 27 Gardner, K.H. and Kay, L.E. (1997) Production and incorporation of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  ( $^1\text{H}$ - $\delta$ 1 methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc.* 119, 7599–7600 <https://doi.org/10.1021/ja9706514>
- 28 Ayala, I., Hamelin, O., Amaro, C., Pessey, O., Plevin, M.J., Gans, P. et al. (2012) An optimized isotopic labelling strategy of isoleucine- $\gamma$ (2) methyl groups for solution NMR studies of high molecular weight proteins. *Chem Commun.* 48, 1434–1436 <https://doi.org/10.1039/c1cc12932e>
- 29 Mas, G., Grubler, E., Hamelin, O., Gans, P. and Boisbouvier, J. (2013) Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J. Biomol. NMR* 57, 251–262 <https://doi.org/10.1007/s10858-013-9785-z>
- 30 Godoy-Ruiz, R., Guo, C.Y. and Tugarinov, V. (2010) Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. *J. Am. Chem. Soc.* 132, 18340–18350 <https://doi.org/10.1021/ja106365e>
- 31 Bellstedt, P., Seiboth, T., Häfner, S., Kutscha, H., Ramchandran, R. and Gorkach, M. (2013) Resonance assignment for a particularly challenging protein based on systematic unlabeled of amino acids to complement incomplete NMR data sets. *J. Biomol. NMR* 57, 65–72 <https://doi.org/10.1007/s10858-013-9768-0>

- 32 Takeuchi, K., Ng, E., Melia, T.J. and Wagner, G. (2007)  $1\text{-}^{13}\text{C}$  amino acid selective labeling in a  $(\text{H-N})\text{-}^2\text{H}\text{-}^{15}\text{N}$  background for NMR studies of large proteins. *J. Biomol. NMR* 38, 89–98 <https://doi.org/10.1007/s10858-007-9152-z>
- 33 Ayala, I., Sounier, R., Use, N., Gans, P. and Boisbouvier, J. (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J. Biomol. NMR* 43, 111–119 <https://doi.org/10.1007/s10858-008-9294-7>
- 34 Isaacson, R.L., Simpson, P.J., Liu, M., Cota, E., Zhang, X., Fremont, P. et al. (2007) A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J. Am. Chem. Soc.* 129, 15428–15429 <https://doi.org/10.1021/ja0761784>
- 35 Waugh, D.S. (1996) Genetic tools for selective labeling of proteins with  $\alpha\text{-}^{15}\text{N}$ -amino acids. *J. Biomol. NMR* 8, 184–192 <https://doi.org/10.1007/BF00211164>
- 36 Perez, L.M., Idasi, F.S., Bessa, L.M., Maurin, D., Kragelj, J., Blackledge, M. et al. (2022) Visualizing protein breathing motions associated with aromatic ring flipping. *Nature* 602, 695–700 <https://doi.org/10.1038/s41586-022-04417-6>
- 37 Reitzer, L. (2005) Catabolism of amino acids and related compounds 1, 1–56 <https://doi.org/10.1128/ecosolplus.3.4.7>
- 38 Rajesh, S., Nellispach, D., Nakayama, H., Takio, K., Laue, E.D., Shibata, T. et al. (2003) A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp. *J. Biomol. NMR* 27, 81–86 <https://doi.org/10.1023/a:1024710729352>
- 39 Lichteneder, R.J., Währhaup, K., Schmid, W. and Konrat, R. (2013)  $\alpha\text{-Ketoacids}$  as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *J. Biomol. NMR* 57, 327–331 <https://doi.org/10.1007/s10858-013-9796-9>
- 40 Rodriguez-Mias, R.A. and Pellecchia, M. (2003) Use of selective TRP side chain labeling to characterize protein-protein and protein-ligand interactions by NMR spectroscopy. *J. Am. Chem. Soc.* 125, 2892–2893 <https://doi.org/10.1021/ja029221q>
- 41 Schorghuber, J., Sara, T., Bisaccia, M., Schmid, W., Konrat, R. and Lichteneder, R.J. (2015) Novel approaches in selective tryptophan isotope labeling by using *Escherichia coli* overexpression media. *Chembiochem* 16, 746–751 <https://doi.org/10.1002/cbic.201402677>
- 42 Schorghuber, J., Geist, L., Bisaccia, M., Weber, F., Konrat, R. and Lichteneder, R. (2017) Anthranilic acid, the new player in the ensemble of aromatic residue labeling precursor compounds. *J. Biomol. NMR* 69, 13–22 <https://doi.org/10.1007/s10858-017-0129-2>
- 43 Schorghuber, J., Geist, L., Platzer, G., Konrat, R. and Lichteneder, R.J. (2017) Highly selective stable isotope labeling of histidine residues by using a novel precursor in *E. coli*-based overexpression systems. *Chembiochem* 18, 1487–1491 <https://doi.org/10.1002/cbic.201700192>
- 44 Ayala, I., Chari, L., Kerfah, R., Boisbouvier, J., Gans, P. and Harel, O. (2020) Asymmetric synthesis of methyl specifically labelled L-threonine and application to the NMR studies of high molecular weight proteins. *ChemistrySelect* 5, 5092–5098 <https://doi.org/10.1002/slct.202000827>
- 45 Velyis, A., Ruschak, A.M. and Kay, L.E. (2012) An economical method for production of  $^2\text{H}$  ( $\text{CH}_3$ )- $^{13}\text{C}$  threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One* 7, e43725 <https://doi.org/10.1371/journal.pone.0043725>
- 46 Cao, C., Chen, J.L., Yang, Y., Huang, F., Oting, G. and Su, X.C. (2014) Selective  $^{15}\text{N}$  labeling of the side-chain amide groups of asparagine and glutamine for applications in paramagnetic NMR spectroscopy. *J. Biomol. NMR* 59, 251–261 <https://doi.org/10.1007/s10858-014-9844-0>
- 47 Goux, W.J., Strong, A.A.D., Schneider, B.L., Lee, W.N.P. and Reitzer, L.J. (1995) Utilization of aspartate as a nitrogen source in *Escherichia coli* - analysis of nitrogen flow and characterization of the products of aspartate catabolism. *J. Biol. Chem.* 270, 638–646 <https://doi.org/10.1074/jbc.270.2.638>
- 48 Nygaard, R., Zou, Y.Z., Dor, R.O., Mildorf, T.J., Arlow, D.H., Manglik, A. et al. (2013) The dynamic process of  $\beta(2)$ -adrenergic receptor activation. *Cell* 152, 532–542 <https://doi.org/10.1016/j.cell.2013.01.008>
- 49 Stoffregen, M.C., Schwer, M.M., Renschler, F.A. and Wiesner, S. (2012) Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces. *Structure* 20, 573–581 <https://doi.org/10.1016/j.str.2012.02.012>
- 50 Gelis, I., Bonvin, A., Keranisano, D., Koukaki, M., Gouridis, G., Karananos, S. et al. (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* 131, 756–769 <https://doi.org/10.1016/j.cell.2007.09.039>
- 51 Fischer, M., Klöber, K., Hausler, J., Leddler, K., Konrat, R. and Schmid, W. (2007) Synthesis of a  $^{13}\text{C}$ -methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *Chembiochem* 8, 610–612 <https://doi.org/10.1002/cbic.200600551>
- 52 Frank, L. (1963) Proline metabolism in *Escherichia coli* II. Regulation of total growth of proline auxotroph by a proline-oxidizing system. *J. Bacteriol.* 86, 781–784 <https://doi.org/10.1128/jb.86.4.781-784.1963>
- 53 Kodama, Y., Reese, M.L., Shirata, N., Ono, K., Kanamori, E., Dotsch, V. et al. (2011) Rapid identification of protein-protein interfaces for the construction of a complex model based on multiple unassigned signals by using time-sharing NMR measurements. *J. Struct. Biol.* 174, 434–442 <https://doi.org/10.1016/j.jsb.2011.04.001>
- 54 Dominguez, C., Boelens, R. and Bonvin, A. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737 <https://doi.org/10.1021/ja026939x>
- 55 Junger, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 <https://doi.org/10.1038/s41586-021-03819-2>
- 56 Zvedstetter, M. (2021) NMR hawk-eyed view of AlphaFold2 structures. *Protein Sci.* 30, 2333–2337 <https://doi.org/10.1002/pro.4175>
- 57 Clark, L., Diky, I., Rosenbaum, D.M. and Gardner, K.H. (2018) On the use of *Pichia pastoris* for isotopic labeling of human GPCRs for NMR studies. *J. Biomol. NMR* 71, 203–211 <https://doi.org/10.1007/s10858-018-0204-3>
- 58 Franke, B., Qitz, C., Isogai, S., Graf, A., Delgado, L., Gossert, A.D. et al. (2018) Production of isotope-labeled proteins in insect cells for NMR. *J. Biomol. NMR* 71, 173–184 <https://doi.org/10.1007/s10858-018-0172-7>
- 59 Sugiki, T., Ichikawa, O., Miyazawa-Chami, M., Shimada, I. and Takahashi, H. (2012) Isotopic labeling of heterologous proteins in the yeast *Pichia pastoris* and *Kluyveromyces fragilis* 831, 19–36 [https://doi.org/10.1007/978-1-61779-480-3\\_2](https://doi.org/10.1007/978-1-61779-480-3_2)
- 60 Kofuku, Y., Yokomizo, T., Inai, S., Shiraishi, Y., Natsune, M., Itoh, H. et al. (2018) Deuteration and selective labeling of alanine methyl groups of  $\beta(2)$ -adrenergic receptor expressed in a baculovirus-insect cell expression system. *J. Biomol. NMR* 71, 185–192 <https://doi.org/10.1007/s10858-018-0174-5>
- 61 Sitariska, A., Skora, L., Klopp, J., Roest, S., Fernandez, C., Shrestha, B. et al. (2015) Affordable uniform isotope labeling with  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  in insect cells. *J. Biomol. NMR* 62, 191–197 <https://doi.org/10.1007/s10858-015-9935-6>
- 62 Yagi, H., Nakamura, M., Yokoyama, J., Zhang, Y., Yamaguchi, T., Kondo, S. et al. (2015) Stable isotope labeling of glycoprotein expressed in silkworms using immunoglobulin G as a test molecule. *J. Biomol. NMR* 62, 157–167 <https://doi.org/10.1007/s10858-015-9930-y>
- 63 Yanaka, S., Yagi, H., Yogo, R., Yagi-Utsuni, M. and Kato, K. (2018) Stable isotope labeling approaches for NMR characterization of glycoproteins using eukaryotic expression systems. *J. Biomol. NMR* 71, 193–202 <https://doi.org/10.1007/s10858-018-0169-2>

- 64 Werner, K., Richter, C., Klein-Seetharaman, J. and Schwalbe, H. (2008) Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J. Biomol. NMR* 40, 49–53 <https://doi.org/10.1007/s10858-007-9205-3>
- 65 Kigawa, T., Muto, Y. and Yokoyama, S. (1995) Cell-free synthesis and amino acid-selective stable-isotope labeling of proteins for NMR analysis. *J. Biomol. NMR* 6, 129–134 <https://doi.org/10.1007/bf00211776>
- 66 Yabuki, T., Kigawa, T., Dohmae, N., Takio, K., Terada, T., Ito, Y. et al. (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J. Biomol. NMR* 11, 295–306 <https://doi.org/10.1023/a:1008276001545>
- 67 Su, X.C., Loh, C.T., Qi, R.H. and Otting, G. (2011) Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enzymes for selective <sup>15</sup>N-labeling and production of perdeuterated proteins in H<sub>2</sub>O. *J. Biomol. NMR* 51, 409–409 <https://doi.org/10.1007/s10858-011-9562-9>
- 68 Calhoun, K.A. and Svartz, J.R. (2006) Total amino acid stabilization during cell-free protein synthesis reactions. *J. Biotechnol.* 123, 193–203 <https://doi.org/10.1016/j.jbiotec.2005.11.011>
- 69 Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Cho, A.M. and Guntert, P. (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440, 52–57 <https://doi.org/10.1038/nature04525>
- 70 Urbancik, A., Morato, A., Allemand, F., Delaforge, E., Fournet, A., Popovic, M. et al. (2018) A general strategy to access structural information at atomic resolution in polyglutamine homorepeats. *Angew. Chem. Int. Ed.* 57, 3598–3601 <https://doi.org/10.1002/anie.201711530>



## 8. References

1. Rowlinson B, Crublet E, Kerfah R, Plevin MJ. (2022) Specific isotopic labelling and reverse labelling for protein NMR spectroscopy: using metabolic precursors in sample preparation. *Biochem. Soc. Trans.* 13 <https://doi.org/10.1042/bst20210586>
2. Good MC, Zalatan JG, Lim WA. (2011) Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science.* **332**, 680-686 <https://doi.org/10.1126/science.1198701>
3. Dyla M, Kjaergaard M. (2020) Intrinsically disordered linkers control tethered kinases via effective concentration. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21413-21419 <https://doi.org/10.1073/pnas.2006382117>
4. Zheng N, Schulman BA, Song LZ, Miller JJ, Jeffrey PD, Wang P, *et al.* (2002) Structure of the Cul1-Rbx1-Skp1-F box(Skp2) SCF ubiquitin ligase complex. *Nature.* **416**, 703-709 <https://doi.org/10.1038/416703a>
5. Rajakulendran T, Sahmi M, Lefrancois M, Sicheri F, Therrien M. (2009) A dimerization-dependent mechanism drives RAF catalytic activation. *Nature.* **461**, 542-545 <https://doi.org/10.1038/nature08314>
6. Cortese MS, Uversky VN, Dunker AK. (2008) Intrinsic disorder in scaffold proteins: Getting more from less. *Prog. Biophys. Mol. Biol.* **98**, 85-106 <https://doi.org/10.1016/j.pbiomolbio.2008.05.007>
7. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635-645 <https://doi.org/10.1016/j.jmb.2004.02.002>
8. Wright PE, Dyson HJ. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18-29 <https://doi.org/10.1038/nrm3920>
9. Jensen MR, Blackledge M. (2014) Testing the validity of ensemble descriptions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1557-E1558 <https://doi.org/10.1073/pnas.1323876111>
10. Weikl TR, Paul F. (2014) Conformational selection in protein binding and function. *Protein Sci.* **23**, 1508-1518 <https://doi.org/10.1002/pro.2539>
11. Borchers W, Theillet FX, Katzer A, Finzel A, Mishall KM, Powell AT, *et al.* (2014) Disorder and residual helicity alter p53-Mdm2 binding affinity and

- signaling in cells. *Nat. Chem. Biol.* **10**, 1000-10002  
<https://doi.org/10.1038/nchembio.1668>
12. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, *et al.* (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*. **274**, 948-953  
<https://doi.org/10.1126/science.274.5289.948>
  13. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*. **9**, 20 <https://doi.org/10.1186/1471-2164-9-s1-s1>
  14. Rustandi RR, Baldisseri DM, Weber DJ. (2000) Structure of the negative regulatory domain of p53 bound to S100B( $\beta\beta$ ). *Nat. Struct. Biol.* **7**, 570-574
  15. Avalos JL, Celic I, Muhammad S, Cosgrove MS, Boeke JD, Wolberger C. (2002) Structure of a Sir2 enzyme bound to an acetylated p53 peptide. *Mol. Cell*. **10**, 523-535 [https://doi.org/10.1016/s1097-2765\(02\)00628-7](https://doi.org/10.1016/s1097-2765(02)00628-7)
  16. Mujtaba S, He Y, Zeng L, Yan S, Plotnikova O, Sachchidanand, *et al.* (2004) Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell*. **13**, 251-263  
[https://doi.org/10.1016/s1097-2765\(03\)00528-8](https://doi.org/10.1016/s1097-2765(03)00528-8)
  17. Lowe ED, Tews I, Cheng KY, Brown NR, Gul S, Noble MEM, *et al.* (2002) Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. *Biochemistry*. **41**, 15625-15634  
<https://doi.org/10.1021/bi0268910>
  18. Fuxreiter M, Tompa P. (2009) Fuzzy Complexes: Polymorphism And Structural Disorder In Protein-protein Interactions. *Biophys. J.* **96**, 319A-319A  
<https://doi.org/10.1016/j.bpj.2008.12.1600>
  19. Posey AE, Holehouse AS, Pappu RV. Phase Separation of Intrinsically Disordered Proteins. In: Rhoades E, editor. *Intrinsically Disordered Proteins. Methods in Enzymology*. 611. San Diego: Elsevier Academic Press Inc; 2018. p. 1-30.
  20. Choi JM, Holehouse AS, Pappu RV. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. In: Dill KA, editor. *Annual Review of Biophysics*, Vol 49, 2020. *Annual Review of Biophysics*. 49. Palo Alto: Annual Reviews; 2020. p. 107-133.

21. Jo Y, Jang J, Song D, Park H, Jung Y. (2022) Determinants for intrinsically disordered protein recruitment into phase-separated protein condensates. *Chem. Sci.* **13**, 522-530 <https://doi.org/10.1039/d1sc05672g>
22. Kiss H, Kedra D, Yang Y, Kost-Alimova M, Kiss C, O'Brien KP, *et al.* (1999) A novel gene containing LIM domains (LIMD1) is located within the common eliminated region 1 (C3CER1) in 3p21.3. *Hum. Genet.* **105**, 552-559 <https://doi.org/10.1007/s004390051144>
23. Kadrmas JL, Beckerle MC. (2004) The LIM domain: From the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* **5**, 920-931 <https://doi.org/10.1038/nrm1499>
24. Foxler DE, Bridge KS, James V, Webb TM, Mee M, Wong SCK, *et al.* (2012) The LIMD1 protein bridges an association between the prolyl hydroxylases and VHL to repress HIF-1 activity. *Nat. Cell Biol.* **14**, 201-208 <https://doi.org/10.1038/ncb2424>
25. Bridge KS, Shah KM, Li YG, Foxler DE, Wong SCK, Miller DC, *et al.* (2017) Argonaute Utilization for miRNA Silencing Is Determined by Phosphorylation-Dependent Recruitment of LIM-Domain-Containing Proteins. *Cell Rep.* **20**, 173-187 <https://doi.org/10.1016/j.celrep.2017.06.027>
26. Michelsen JW, Schmeichel KL, Beckerle MC, Winge DR. (1993) The LIM Motif Defines a Specific Zinc-Binding Protein Domain. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4404-4408 <https://doi.org/10.1073/pnas.90.10.4404>
27. Sanchezgarcia I, Rabbitts TH. (1994) The LIM Domain - a New Structural Motif Found in Zinc-Finger-Like Proteins. *Trends Genet.* **10**, 315-320 [https://doi.org/10.1016/0168-9525\(94\)90034-5](https://doi.org/10.1016/0168-9525(94)90034-5)
28. Freyd G, Kim SK, Horvitz HR. (1990) Novel Cysteine-Rich Motif and Homeodomain in the Product of the *Caenorhabditis-Elegans* Cell Lineage Gene LIN-II. *Nature.* **344**, 876-879 <https://doi.org/10.1038/344876a0>
29. Karlsson O, Thor S, Norberg T, Ohlsson H, Edlund T. (1990) Insulin Gene Enhancer Binding-Protein ISL-1 is a Member of a Novel Class of Proteins Containing Both a Homeodomain and a CYS-HIS Domain. *Nature.* **344**, 879-882 <https://doi.org/10.1038/344879a0>
30. Way JC, Chalfie M. (1988) MEC-3, a Homeobox-Containing Gene That Specifies Differentiation of the Touch Receptor Neurons in *C-Elegans*. *Cell.* **54**, 5-16 [https://doi.org/10.1016/0092-8674\(88\)90174-2](https://doi.org/10.1016/0092-8674(88)90174-2)

31. Gill GN. (1995) The enigma of LIM domains. *Structure*. **3**, 1285-1289  
[https://doi.org/10.1016/s0969-2126\(01\)00265-9](https://doi.org/10.1016/s0969-2126(01)00265-9)
32. Perezalvarado GC, Miles C, Michelsen JW, Louis HA, Winge DR, Beckerle MC, *et al.* (1994) Structure of the Carboxy-Terminal LIM Domain From the Cysteine-Rich Protein CRP. *Nat. Struct. Biol.* **1**, 388-398  
<https://doi.org/10.1038/nsb0694-388>
33. Bach I. (2000) The LIM domain: regulation by association. *Mech. Dev.* **91**, 5-17  
[https://doi.org/10.1016/s0925-4773\(99\)00314-7](https://doi.org/10.1016/s0925-4773(99)00314-7)
34. Li A, Ponten F, dos Remedios CG. (2012) The interactome of LIM domain proteins: The contributions of LIM domain proteins to heart failure and heart development. *Proteomics*. **12**, 203-225  
<https://doi.org/10.1002/pmic.201100492>
35. Matthews JM, Lester K, Joseph S, Curtis DJ. (2013) LIM-domain-only proteins in cancer. *Nat. Rev. Cancer*. **13**, 111-122  
<https://doi.org/10.1038/nrc3418>
36. Koch BJ, Ryan JF, Baxevanis AD. (2012) The Diversification of the LIM Superclass at the Base of the Metazoa Increased Subcellular Complexity and Promoted Multicellular Specialization. *PLoS One*. **7**, 14  
<https://doi.org/10.1371/journal.pone.0033261>
37. Dawid IB, Toyama R, Taira M. (1995) LIM Domain Proteins. *Comptes Rendus Acad. Sci. Ser. III-Sci. Vie-Life Sci.* **318**, 295-306
38. Hirata H, Tatsumi H, Sokabe M. (2008) Mechanical forces facilitate actin polymerization at focal adhesions in a zyxin-dependent manner. *J. Cell Sci.* **121**, 2795-2804  
<https://doi.org/10.1242/jcs.030320>
39. Velyvis A, Qin J. LIM domain and Its binding to target proteins. *Zinc Finger Proteins*: Springer; 2005. p. 99-105.
40. Boeda B, Knowles PP, Briggs DC, Murray-Rust J, Soriano E, Garvalov BK, *et al.* (2011) Molecular Recognition of the Tes LIM2-3 Domains by the Actin-related Protein Arp7A. *J. Biol. Chem.* **286**, 11543-11554  
<https://doi.org/10.1074/jbc.M110.171264>
41. El Omari K, Hoosdally SJ, Tuladhar K, Karia D, Hall-Ponsele E, Platonova O, *et al.* (2013) Structural Basis for LMO2-Driven Recruitment of the SCL:E47(bHLH) Heterodimer to Hematopoietic-Specific Transcriptional Targets. *Cell Rep.* **4**, 135-147  
<https://doi.org/10.1016/j.celrep.2013.06.008>



42. Chiswell BP, Stiegler AL, Razinia Z, Nalibotski E, Boggon TJ, Calderwood DA. (2010) Structural basis of competition between PINCH1 and PINCH2 for binding to the ankyrin repeat domain of integrin-linked kinase. *J. Struct. Biol.* **170**, 157-163 <https://doi.org/10.1016/j.jsb.2009.12.002>
43. Vaynberg J, Fukuda T, Chen K, Vinogradova O, Velyvis A, Tu YZ, *et al.* (2005) Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility. *Mol. Cell.* **17**, 513-523 <https://doi.org/10.1016/j.molcel.2004.12.031>
44. Bhati M, Lee C, Nancarrow AL, Lee M, Craig VJ, Bach I, *et al.* (2008) Implementing the LIM code: the structural basis for cell type-specific assembly of LIM-homeodomain complexes. *EMBO J.* **27**, 2018-2029 <https://doi.org/10.1038/emboj.2008.123>
45. Gadd MS, Jacques DA, Nisevic I, Craig VJ, Kwan AH, Guss JM, *et al.* (2013) A Structural Basis for the Regulation of the LIM-Homeodomain Protein Islet 1 (Isl1) by Intra- and Intermolecular Interactions. *J. Biol. Chem.* **288**, 12 <https://doi.org/10.1074/jbc.M113.478586>
46. Deane JE, Ryan DP, Sunde M, Maher MJ, Guss JM, Visvader JE, *et al.* (2004) Tandem LIM domains provide synergistic binding in the LMO4 : Ldb1 complex. *EMBO J.* **23**, 3589-3598 <https://doi.org/10.1038/sj.emboj.7600376>
47. Stokes PH, Liew CW, Kwan AH, Foo P, Barker HE, Djamirze A, *et al.* (2013) Structural Basis of the Interaction of the Breast Cancer Oncogene LMO4 with the Tumour Suppressor CtIP/RBBP8. *J. Mol. Biol.* **425**, 1101-1110 <https://doi.org/10.1016/j.jmb.2013.01.017>
48. Chiswell BP, Zhang R, Murphy JW, Boggon TJ, Calderwood DA. (2008) The structural basis of integrin-linked kinase-PINCH interactions. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20677-20682 <https://doi.org/10.1073/pnas.0811415106>
49. Gadd MS, Bhati M, Jeffries CM, Langley DB, Trewella J, Guss JM, *et al.* (2011) Structural Basis for Partial Redundancy in a Class of Transcription Factors, the LIM Homeodomain Proteins, in Neural Cell Type Specification. *J. Biol. Chem.* **286**, 42971-42980 <https://doi.org/10.1074/jbc.M111.248559>
50. McNicholas S, Potterton E, Wilson KS, Noble MEM. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D-Struct. Biol.* **67**, 386-394 <https://doi.org/10.1107/s0907444911007281>

51. Stokes PH, Robertson NO, Silva APG, Estephan T, Trehwella J, Guss JM, *et al.* (2019) Mutation in a flexible linker modulates binding affinity for modular complexes. *Proteins*. **87**, 425-429 <https://doi.org/10.1002/prot.25675>
52. Joseph S, Kwan AH, Stokes PH, Mackay JP, Cubeddu L, Matthews JM. (2014) The Structure of an LIM-Only Protein 4 (LMO4) and Deformed Epidermal Autoregulatory Factor-1 (DEAF1) Complex Reveals a Common Mode of Binding to LMO4. *PLoS One*. **9**, 13 <https://doi.org/10.1371/journal.pone.0109108>
53. Matthews JM, Potts JR. (2013) The tandem beta-zipper: Modular binding of tandem domains and linear motifs. *FEBS Lett*. **587**, 1164-1171 <https://doi.org/10.1016/j.febslet.2013.01.002>
54. Deane JE, Sum E, Mackay JP, Lindeman GJ, Visvader JE, Matthews JM. (2001) Design, production and characterization of FLIN2 and FLIN4: the engineering of intramolecular Idb1 : LMO complexes. *Protein Eng*. **14**, 493-499 <https://doi.org/10.1093/protein/14.7.493>
55. El Omari K, Hoosdally SJ, Tuladhar K, Karia D, Vyas P, Patient R, *et al.* (2011) Structure of the leukemia oncogene LMO2: implications for the assembly of a hematopoietic transcription factor complex. *Blood*. **117**, 2146-2156 <https://doi.org/10.1182/blood-2010-07-293357>
56. Boeda B, Briggs DC, Higgins T, Garvalov BK, Fadden AJ, McDonald NQ, *et al.* (2007) Tes, a specific Mena interacting partner, breaks the rules for EVH1 binding. *Mol. Cell*. **28**, 1071-1082 <https://doi.org/10.1016/j.molcel.2007.10.033>
57. James V, Zhang YN, Foxler DE, de Moor CH, Kong YW, Webb TM, *et al.* (2010) LIM-domain proteins, LIMD1, Ajuba, and WTIP are required for microRNA-mediated gene silencing. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12499-12504 <https://doi.org/10.1073/pnas.0914987107>
58. Jagannathan R, Schimizzi GV, Zhang K, Loza AJ, Yabuta N, Nojima H, *et al.* (2016) AJUBA LIM Proteins Limit Hippo Activity in Proliferating Cells by Sequestering the Hippo Core Kinase Complex in the Cytosol. *Mol. Cell. Biol*. **36**, 2526-2542 <https://doi.org/10.1128/mcb.00136-16>
59. Saitou N, Nei M. (1987) THE NEIGHBOR-JOINING METHOD - A NEW METHOD FOR RECONSTRUCTING PHYLOGENETIC TREES. *Mol. Biol. Evol*. **4**, 406-425 <https://doi.org/10.1093/oxfordjournals.molbev.a040454>

60. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 6  
<https://doi.org/10.1038/msb.2011.75>
61. Abe Y, Ohsugi M, Haraguchi K, Fujimoto J, Yamamoto T. (2006) LATS2-Ajuba complex regulates gamma-tubulin recruitment to centrosomes and spindle organization during mitosis. *FEBS Lett.* **580**, 782-788  
<https://doi.org/10.1016/j.febslet.2005.12.096>
62. Hirota T, Kunitoku N, Sasayama T, Marumoto T, Zhang DW, Nitta M, *et al.* (2003) Aurora-A and an interacting activator, the LIM protein Ajuba, are required for mitotic commitment in human cells. *Cell.* **114**, 585-598  
[https://doi.org/10.1016/s0092-8674\(03\)00642-1](https://doi.org/10.1016/s0092-8674(03)00642-1)
63. Wang Y, Zhang CL, Yang WZ, Shao SP, Xu XM, Sun YJ, *et al.* (2021) LIMD1 phase separation contributes to cellular mechanics and durotaxis by regulating focal adhesion dynamics in response to force. *Dev. Cell.* **56**, 1313-1325  
<https://doi.org/10.1016/j.devcel.2021.04.002>
64. Bai SW, Herrera-Abreu MT, Rohn JL, Racine V, Tajadura V, Suryavanshi N, *et al.* (2011) Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol.* **9**, 18  
<https://doi.org/10.1186/1741-7007-9-54>
65. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* **25**, 1189-1191  
<https://doi.org/10.1093/bioinformatics/btp033>
66. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427-D432  
<https://doi.org/10.1093/nar/gky995>
67. Sharp TV, Munoz F, Bourboulia D, Presneau N, Darai E, Wang HW, *et al.* (2004) LIM domains-containing protein 1 (LIMD1), a tumor suppressor encoded at chromosome 3p21.3, binds pRB and represses E2F-driven transcription. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16531-16536  
<https://doi.org/10.1073/pnas.0407123101>
68. Sharp TV, Al-Attar A, Foxler DE, Ding L, Vallim TQD, Zhang YN, *et al.* (2008) The chromosome 3p21.3-encoded gene, LIMD1, is a critical tumor suppressor

- involved in human lung cancer development. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19932-19937 <https://doi.org/10.1073/pnas.0805003105>
69. Spendlove I, Al-Attar A, Watherstone O, Webb TM, Ellis IO, Longmore GD, *et al.* (2008) Differential subcellular localisation of the tumour suppressor protein LIMD1 in breast cancer correlates with patient survival. *Int. J. Cancer.* **123**, 2247-2253 <https://doi.org/10.1002/ijc.23851>
70. Ghosh S, Ghosh A, Maiti GP, Mukherjee N, Dutta S, Roy A, *et al.* (2010) LIMD1 is more frequently altered than RB1 in head and neck squamous cell carcinoma: clinical and prognostic implications. *Mol. Cancer.* **9**, 13 <https://doi.org/10.1186/1476-4598-9-58>
71. Liao Z-L, Liu L, Wang Y, Fan H. (2015) Expression and clinical significance of LIMD-1 gene in adult patients with acute leukemia. **23**, 34-38 <https://doi.org/10.7534/j.issn.1009-2137.2015.01.007>
72. Basu M, Chatterjee A, Chakraborty B, Chatterjee E, Ghosh S, Samadder S, *et al.* (2021) High nuclear expression of HIF1 alpha, synergizing with inactivation of LIMD1 and VHL, portray worst prognosis among the bladder cancer patients: association with arsenic prevalence. *J. Cancer Res. Clin. Oncol.* **147**, 2309-2322 <https://doi.org/10.1007/s00432-021-03661-z>
73. Chen ZX, Zhu XS, Xie T, Xie JP, Quo K, Liu X. (2014) Drug resistance reversed by silencing LIM domain-containing protein 1 expression in colorectal carcinoma. *Oncol. Lett.* **8**, 795-798 <https://doi.org/10.3892/ol.2014.2155>
74. Tang J, Zhu LQ, Huang YJ, Shi B, Zhang SQ, Gu LL, *et al.* (2019) Silencing of LIMD1 promotes proliferation and reverses cell adhesion-mediated drug resistance in non-Hodgkin's lymphoma. *Oncol. Lett.* **17**, 2993-3000 <https://doi.org/10.3892/ol.2019.9921>
75. Luderer HF, Bai ST, Lonmore GD. (2008) The LIM protein LIMD1 influences osteoblast differentiation and function. *Exp. Cell Res.* **314**, 2884-2894 <https://doi.org/10.1016/j.yexcr.2008.06.003>
76. Feng YF, Zhao HB, Luderer HF, Epple H, Faccio R, Ross FP, *et al.* (2007) The LIM protein, LIMD1, regulates AP-1 activation through an interaction with TRAF6 to influence osteoclast development. *J. Biol. Chem.* **282**, 39-48 <https://doi.org/10.1074/jbc.M607399200>

77. Ibar C, Kirichenko E, Keepers B, Enners E, Fleisch K, Irvine KD. (2018) Tension-dependent regulation of mammalian Hippo signaling through LIMD1. *J. Cell Sci.* **131**, 16 <https://doi.org/10.1242/jcs.214700>
78. Landry NM, Rattan SG, Filomeno KL, Meier TW, Meier SC, Foran SJ, *et al.* (2021) SKI activates the Hippo pathway via LIMD1 to inhibit cardiac fibroblast activation. *Basic Res. Cardiol.* **116**, 27 <https://doi.org/10.1007/s00395-021-00865-9>
79. Hou XB, Li TH, Ren ZP, Liu Y. (2016) Novel BRCA2-Interacting Protein, LIMD1, Is Essential for the Centrosome Localization of BRCA2 in Esophageal Cancer *Cell. Oncol. Res.* **24**, 247-253 <https://doi.org/10.3727/096504016x14652175055765>
80. Langer EM, Feng YF, Zhaoyuan H, Rauscher FJ, Kroll KL, Longmore GD. (2008) Ajuba LIM proteins are Snail/Slug corepressors required for neural crest development in *Xenopus*. *Dev. Cell.* **14**, 424-436 <https://doi.org/10.1016/j.devcel.2008.01.005>
81. Graham K, Chandrasekaran A, Wang L, Ladak A, Lafer EM, Rangamani P, *et al.* (2022) Liquid-like VASP condensates drive actin polymerization and dynamic bundling. *bioRxiv.* <https://doi.org/10.1101/2022.05.09.491236>
82. Erdos G, Pajkos M, Dosztanyi Z. (2021) IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297-W303 <https://doi.org/10.1093/nar/gkab408>
83. O'Brien J, Hayder H, Zayed Y, Peng C. (2018) Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* **9**, 12 <https://doi.org/10.3389/fendo.2018.00402>
84. Han JJ, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016-3027 <https://doi.org/10.1101/gad.1262504>
85. Yoda M, Kawamata T, Paroo Z, Ye XC, Iwasaki S, Liu QH, *et al.* (2010) ATP-dependent human RISC assembly pathways. *Nat. Struct. Mol. Biol.* **17**, 17-29 <https://doi.org/10.1038/nsmb.1733>
86. Ameres SL, Zamore PD. (2013) Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14**, 475-488 <https://doi.org/10.1038/nrm3611>

87. Jonas S, Izaurralde E. (2015) NON-CODING RNA Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421-433 <https://doi.org/10.1038/nrg3965>
88. Liu JD, Rivas FV, Wohlschlegel J, Yates JR, Parker R, Hannon GJ. (2005) A role for the P-body component GW182 in microRNA function. *Nat. Cell Biol.* **7**, 1261-1266 <https://doi.org/10.1038/ncb1333>
89. Fabian MR, Sonenberg N. (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* **19**, 586-593 <https://doi.org/10.1038/nsmb.2296>
90. Pfaff J, Hennig J, Herzog F, Aebersold R, Sattler M, Niessing D, *et al.* (2013) Structural features of Argonaute-GW182 protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E3770-E3779 <https://doi.org/10.1073/pnas.1308510110>
91. Elkayam E, Faehnle CR, Morales M, Sun JC, Li HL, Joshua-Tor L. (2017) Multivalent Recruitment of Human Argonaute by GW182. *Mol. Cell.* **67**, 646-658 <https://doi.org/10.1016/j.molcel.2017.07.007>
92. Behm-Ansmant I, Rehwinkel J, Izaurralde E. (2006) MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. *Cold Spring Harbor Symp. Quant. Biol.* **71**, 523-530 <https://doi.org/10.1101/sqb.2006.71.013>
93. Till S, Lejeune E, Thermann R, Bortfeld M, Hothorn M, Enderle D, *et al.* (2007) A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nat. Struct. Mol. Biol.* **14**, 897-903 <https://doi.org/10.1038/nsmb1302>
94. Takimoto K, Wakiyama M, Yokoyama S. (2009) Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression. *RNA.* **15**, 1078-1089 <https://doi.org/10.1261/rna.1363109>
95. Schirle NT, MacRae IJ. (2012) The Crystal Structure of Human Argonaute2. *Science.* **336**, 1037-1040 <https://doi.org/10.1126/science.1221551>
96. Sheu-Gruttadauria J, MacRae IJ. (2018) Phase Transitions in the Assembly and Function of Human miRISC. *Cell.* **173**, 946-957 <https://doi.org/10.1016/j.cell.2018.02.051>



97. Chekulaeva M, Mathys H, Zipprich JT, Attig J, Colic M, Parker R, *et al.* (2011) miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat. Struct. Mol. Biol.* **18**, 1218-1262 <https://doi.org/10.1038/nsmb.2166>
98. Braun JE, Huntzinger E, Fauser M, Izaurralde E. (2011) GW182 Proteins Directly Recruit Cytoplasmic Deadenylation Complexes to miRNA Targets. *Mol. Cell.* **44**, 120-133 <https://doi.org/10.1016/j.molcel.2011.09.007>
99. Chen Y, Boland A, Kuzuoglu-Ozturk D, Bawankar P, Loh B, Chang CT, *et al.* (2014) A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing. *Mol. Cell.* **54**, 737-750 <https://doi.org/10.1016/j.molcel.2014.03.034>
100. Braun JE, Truffault V, Boland A, Huntzinger E, Chang CT, Haas G, *et al.* (2012) A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. *Nat. Struct. Mol. Biol.* **19**, 1324-1331 <https://doi.org/10.1038/nsmb.2413>
101. Huntzinger E, Braun JE, Heimstadt S, Zekri L, Izaurralde E. (2010) Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. *EMBO J.* **29**, 4146-4160 <https://doi.org/10.1038/emboj.2010.274>
102. Zekri L, Huntzinger E, Heimstadt S, Izaurralde E. (2009) The Silencing Domain of GW182 Interacts with PABPC1 To Promote Translational Repression and Degradation of MicroRNA Targets and Is Required for Target Release. *Mol. Cell. Biol.* **29**, 6220-6231 <https://doi.org/10.1128/mcb.01081-09>
103. Moretti F, Kaiser C, Zdanowicz-Specht A, Hentze MW. (2012) PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. *Nat. Struct. Mol. Biol.* **19**, 603-608 <https://doi.org/10.1038/nsmb.2309>
104. Pfaff J, Meister G. (2013) Argonaute and GW182 proteins: an effective alliance in gene silencing. *Biochem. Soc. Trans.* **41**, 855-860 <https://doi.org/10.1042/bst20130047>
105. Sala L, Chandrasekhar S, Vidigal JA. (2020) AGO unchained: Canonical and non-canonical roles of Argonaute proteins in mammals. *Front. Biosci.* **25**, 1-42 <https://doi.org/10.2741/4793>
106. Horman SR, Janas MM, Litterst C, Wang B, MacRae IJ, Sever MJ, *et al.* (2013) Akt-Mediated Phosphorylation of Argonaute 2 Downregulates

- Cleavage and Upregulates Translational Repression of MicroRNA Targets. *Mol. Cell.* **50**, 356-367 <https://doi.org/10.1016/j.molcel.2013.03.015>
107. Elkayam E, Kuhn CD, Tocilj A, Haase AD, Greene EM, Hannon GJ, *et al.* (2012) The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell.* **150**, 100-110 <https://doi.org/10.1016/j.cell.2012.05.017>
108. Harris AL. (2002) Hypoxia - A key regulatory factor in tumour growth. *Nat. Rev. Cancer.* **2**, 38-47 <https://doi.org/10.1038/nrc704>
109. Masoud GN, Li W. (2015) HIF-1 alpha pathway: role, regulation and intervention for cancer therapy. *Acta Pharm. Sin. B.* **5**, 378-389 <https://doi.org/10.1016/j.apsb.2015.05.007>
110. Lee CC, Wu CY, Yang HY. (2020) Discoveries of how cells sense oxygen win the 2019 Nobel Prize in Physiology or medicine. *Biomed. J.* **43**, 434-437 <https://doi.org/10.1016/j.bj.2020.05.019>
111. Semenza GL. (2000) HIF-1 and human disease: one highly involved factor. *Genes Dev.* **14**, 1983-1991
112. Semenza GL. (2001) HIF-1 and mechanisms of hypoxia sensing. *Curr. Opin. Cell Biol.* **13**, 167-171 [https://doi.org/10.1016/s0955-0674\(00\)00194-0](https://doi.org/10.1016/s0955-0674(00)00194-0)
113. Ke QD, Costa M. (2006) Hypoxia-inducible factor-1 (HIF-1). *Mol. Pharmacol.* **70**, 1469-1480 <https://doi.org/10.1124/mol.106.027029>
114. Wang GL, Jiang BH, Rue EA, Semenza GL. (1995) Hypoxia-Inducible Factor-1 is a Basic-Helix-Loop-Helix-PAS Heterodimer Regulated by Cellular O<sub>2</sub> Tension. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5510-5514 <https://doi.org/10.1073/pnas.92.12.5510>
115. Kaelin WG, Ratcliffe PJ. (2008) Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Mol. Cell.* **30**, 393-402 <https://doi.org/10.1016/j.molcel.2008.04.009>
116. Huang LE, Gu J, Schau M, Bunn HF. (1998) Regulation of hypoxia-inducible factor 1 alpha is mediated by an O<sub>2</sub>-dependent degradation domain via the ubiquitin-proteasome pathway. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7987-7992 <https://doi.org/10.1073/pnas.95.14.7987>
117. Zhang HF, Gao P, Fukuda R, Kumar G, Krishnamachary B, Zeller KI, *et al.* (2007) HIF-1 inhibits mitochondrial biogenesis and cellular respiration in VHL-deficient renal cell carcinoma by repression of C-MYC activity. *Cancer Cell.* **11**, 407-420 <https://doi.org/10.1016/j.ccr.2007.04.001>



118. Courtney R, Ngo DC, Malik N, Ververis K, Tortorella SM, Karagiannis TC. (2015) Cancer metabolism and the Warburg effect: the role of HIF-1 and PI3K. *Mol. Biol. Rep.* **42**, 841-851 <https://doi.org/10.1007/s11033-015-3858-x>
119. Maynard MA, Qi H, Chung J, Lee EHL, Kondo Y, Hara S, *et al.* (2003) Multiple splice variants of the human HIF-3 alpha locus are targets of the von Hippel-Lindau E3 ubiquitin ligase complex. *J. Biol. Chem.* **278**, 11032-11040 <https://doi.org/10.1074/jbc.M208681200>
120. Duan CM. (2016) Hypoxia-inducible factor 3 biology: complexities and emerging themes. *Am. J. Physiol.-Cell Physiol.* **310**, C260-C269 <https://doi.org/10.1152/ajpcell.00315.2015>
121. Mowat FM, Luhmann UFO, Smith AJ, Lange C, Duran Y, Harten S, *et al.* (2010) HIF-1 $\alpha$  and HIF-2 $\alpha$  Are Differentially Activated in Distinct Cell Populations in Retinal Ischaemia. *PLoS One.* **5**, 9 <https://doi.org/10.1371/journal.pone.0011103>
122. Appelhoff RJ, Tian YM, Raval RR, Turley H, Harris AL, Pugh CW, *et al.* (2004) Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor. *J. Biol. Chem.* **279**, 38458-38465 <https://doi.org/10.1074/jbc.M406026200>
123. Hu CJ, Wang LY, Chodosh LA, Keith B, Simon MC. (2003) Differential roles of hypoxia-inducible factor 1 alpha (HIF-1 alpha) and HIF-2 alpha in hypoxic gene regulation. *Mol. Cell. Biol.* **23**, 9361-9374 <https://doi.org/10.1128/mcb.23.24.9361-9374.2003>
124. Martinez S, Hausinger RP. (2015) Catalytic Mechanisms of Fe(II)-and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20702-20711 <https://doi.org/10.1074/jbc.R115.648691>
125. McNeill LA, Hewitson KS, Claridge TD, Seibel JF, Horsfall LE, Schofield CJ. (2002) Hypoxia-inducible factor asparaginyl hydroxylase (FIH-1) catalyses hydroxylation at the beta-carbon of asparagine-803. *Biochem. J.* **367**, 571-575 <https://doi.org/10.1042/bj20021162>
126. Lando D, Peet DJ, Whelan DA, Gorman JJ, Whitelaw ML. (2002) Asparagine hydroxylation of the HIF transactivation domain: A hypoxic switch. *Science.* **295**, 858-861 <https://doi.org/10.1126/science.1068592>
127. Berra E, Benizri E, Ginouves A, Volmat V, Roux D, Pouyssegur J. (2003) HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels

- of HIF-1 alpha in normoxia. *EMBO J.* **22**, 4082-4090  
<https://doi.org/10.1093/emboj/cdg392>
128. Gothie E, Richard DE, Berra E, Pages G, Pouyssegur J. (2000) Identification of alternative spliced variants of human hypoxia-inducible factor-1 alpha. *J. Biol. Chem.* **275**, 6922-6927 <https://doi.org/10.1074/jbc.275.10.6922>
129. Bruick RK, McKnight SL. (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. *Science.* **294**, 1337-1340  
<https://doi.org/10.1126/science.1066373>
130. Epstein ACR, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, *et al.* (2001) *C-elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell.* **107**, 43-54  
[https://doi.org/10.1016/s0092-8674\(01\)00507-4](https://doi.org/10.1016/s0092-8674(01)00507-4)
131. Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, Cockman ME, *et al.* (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature.* **399**, 271-275  
<https://doi.org/10.1038/20459>
132. Yu F, White SB, Zhao Q, Lee FS. (2001) HIF-1 alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9630-9635 <https://doi.org/10.1073/pnas.181341498>
133. Masson N, Willam C, Maxwell PH, Pugh CW, Ratcliffe PJ. (2001) Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation. *EMBO J.* **20**, 5197-5206  
<https://doi.org/10.1093/emboj/20.18.5197>
134. He WG, Batty-Stuart S, Lee JE, Ohh M. (2021) HIF-1 alpha Hydroxyprolines Modulate Oxygen-Dependent Protein Stability Via Single VHL Interface With Comparable Effect on Ubiquitination Rate. *J. Mol. Biol.* **433**, 17  
<https://doi.org/10.1016/j.jmb.2021.167244>
135. Minamishima YA, Moslehi J, Bardeesy N, Cullen D, Bronson RT, Kaelin WG. (2008) Somatic inactivation of the PHD2 prolyl hydroxylase causes polycythemia and congestive heart failure. *Blood.* **111**, 3236-3244  
<https://doi.org/10.1182/blood-2007-10-117812>
136. Chowdhury R, McDonough MA, Mecinovic J, Loenarz C, Flashman E, Hewitson KS, *et al.* (2009) Structural Basis for Binding of Hypoxia-Inducible

- Factor to the Oxygen-Sensing Prolyl Hydroxylases. *Structure*. **17**, 981-989  
<https://doi.org/10.1016/j.str.2009.06.002>
137. Abboud MI, McAllister TE, Leung IKH, Chowdhury R, Jorgensen C, Domene C, *et al.* (2018) 2-Oxoglutarate regulates binding of hydroxylated hypoxia-inducible factor to prolyl hydroxylase domain 2. *Chem. Commun.* **54**, 3130-3133 <https://doi.org/10.1039/c8cc00387d>
138. Min JH, Yang HF, Ivan M, Gertler F, Kaelin WG, Pavletich NP. (2002) Structure of an HIF-1 alpha-pVHL complex: Hydroxyproline recognition in signaling. *Science*. **296**, 1886-1889 <https://doi.org/10.1126/science.1073440>
139. Buckley DL, Van Molle I, Gareiss PC, Tae HS, Michel J, Noblin DJ, *et al.* (2012) Targeting the von Hippel-Lindau E3 Ubiquitin Ligase Using Small Molecules To Disrupt the VHL/HIF-1 alpha Interaction. *J. Am. Chem. Soc.* **134**, 4465-4468 <https://doi.org/10.1021/ja209924v>
140. Chowdhury R, Leung IKH, Tian YM, Abboud MI, Ge W, Domene C, *et al.* (2016) Structural basis for oxygen degradation domain selectivity of the HIF prolyl hydroxylases. *Nat. Commun.* **7**, 10 <https://doi.org/10.1038/ncomms12673>
141. Meneses AM, Wielockx B. (2016) PHD2: from hypoxia regulation to disease progression. *Hypoxia*. **4**, 53-67 <https://doi.org/10.2147/hp.s53576>
142. Park MH, Choi KY, Jung Y, Min DS. (2014) Phospholipase D1 protein coordinates dynamic assembly of HIF-1 alpha-PHD-VHL to regulate HIF-1 alpha stability. *Oncotarget*. **5**, 11857-11872 <https://doi.org/10.18632/oncotarget.2613>
143. Park MH, Bae SS, Choi KY, Min DS. (2015) Phospholipase D2 promotes degradation of hypoxia-inducible factor-1 alpha independent of lipase activity. *Exp. Mol. Med.* **47**, 7 <https://doi.org/10.1038/emm.2015.87>
144. Han S, Huh J, Kim W, Jeong S, Min DS, Jung Y. (2014) Phospholipase D activates HIF-1-VEGF pathway via phosphatidic acid. *Exp. Mol. Med.* **46**, 7 <https://doi.org/10.1038/emm.2014.86>
145. Wong W, Goehring AS, Kapiloff MS, Langeberg LK, Scott JD. (2008) mAKAP Compartmentalizes Oxygen-Dependent Control of HIF-1 alpha. *Sci. Signal.* **1**, 9 <https://doi.org/10.1126/scisignal.2000026>
146. Choi YK, Kim JH, Kim WJ, Lee HY, Park JA, Lee SW, *et al.* (2007) AKAP12 regulates human blood-retinal barrier formation by downregulation of hypoxia-

- inducible factor-1 alpha. *J. Neurosci.* **27**, 4472-4481  
<https://doi.org/10.1523/jneurosci.5368-06.2007>
147. Foxler DE, Bridge KS, Foster JG, Grevitt P, Curry S, Shah KM, *et al.* (2018) A HIF-LIMD1 negative feedback mechanism mitigates the pro-tumorigenic effects of hypoxia. *EMBO Mol. Med.* **10**, 18  
<https://doi.org/10.15252/emmm.201708304>
148. Zhang CS, Liu Q, Li MQ, Lin SY, Peng YY, Wu D, *et al.* (2015) RHOBTB3 promotes proteasomal degradation of HIF alpha through facilitating hydroxylation and suppresses the Warburg effect. *Cell Res.* **25**, 1025-1042  
<https://doi.org/10.1038/cr.2015.90>
149. Berthold J, Schenkova K, Ramos S, Miura Y, Furukawa M, Aspenstrom P, *et al.* (2008) Characterization of RhoBTB-dependent Cul3 ubiquitin ligase complexes - Evidence for an autoregulatory mechanism. *Exp. Cell Res.* **314**, 3453-3465  
<https://doi.org/10.1016/j.yexcr.2008.09.005>
150. Konrat R. (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* **241**, 74-85  
<https://doi.org/10.1016/j.jmr.2013.11.011>
151. Kurzbach D, Kontaxis G, Coudeville N, Konrat R. NMR Spectroscopic Studies of the Conformational Ensembles of Intrinsically Disordered Proteins. In: Felli IC, Pierattelli R, editors. *Intrinsically Disordered Proteins Studied by Nmr Spectroscopy. Advances in Experimental Medicine and Biology.* 870. Berlin: Springer-Verlag Berlin; 2015. p. 149-185.
152. Eliezer D. (2009) Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **19**, 23-30  
<https://doi.org/10.1016/j.sbi.2008.12.004>
153. Dyson HJ, Wright PE. (2019) Perspective: the essential role of NMR in the discovery and characterization of intrinsically disordered proteins. *J. Biomol. NMR.* **73**, 651-659  
<https://doi.org/10.1007/s10858-019-00280-2>
154. Gibbs EB, Cook EC, Showalter SA. (2017) Application of NMR to studies of intrinsically disordered proteins. *Arch. Biochem. Biophys.* **628**, 57-70  
<https://doi.org/10.1016/j.abb.2017.05.008>
155. Vaynberg J, Qin J. (2006) Weak protein-protein interactions as probed by NMR spectroscopy. *Trends Biotechnol.* **24**, 22-27  
<https://doi.org/10.1016/j.tibtech.2005.09.006>

156. Vinogradova O, Qin J. NMR as a Unique Tool in Assessment and Complex Determination of Weak Protein-Protein Interactions. In: Zhu G, editor. *Nmr of Proteins and Small Biomolecules. Topics in Current Chemistry-Series*. 326. Berlin: Springer-Verlag Berlin; 2012. p. 35-45.
157. Rezaei-Ghaleh N, Blackledge M, Zweckstetter M. (2012) Intrinsically Disordered Proteins: From Sequence and Conformational Properties toward Drug Discovery. *ChemBioChem*. **13**, 930-950 <https://doi.org/10.1002/cbic.201200093>
158. Davidson K, Grevitt P, Contreras-Gerenas MF, Bridge KS, Hermida M, Shah KM, *et al.* (2021) Targeted therapy for LIMD1-deficient non-small cell lung cancer subtypes. *Cell Death Dis.* **12**, 10 <https://doi.org/10.1038/s41419-021-04355-7>
159. Stebbins CE, Kaelin WG, Pavletich NP. (1999) Structure of the VHL-ElonginC-ElonginB complex: Implications for VHL tumor suppressor function. *Science*. **284**, 455-461 <https://doi.org/10.1126/science.284.5413.455>
160. Mas G, Crublet E, Hamelin O, Gans P, Boisbouvier J. (2013) Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J. Biomol. NMR*. **57**, 251-262 <https://doi.org/10.1007/s10858-013-9785-z>
161. Gardner KH, Kay LE. (1997) Production and incorporation of <sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H (<sup>1</sup>H- $\delta$  1 methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc.* **119**, 7599-7600 <https://doi.org/10.1021/ja9706514>
162. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*. **19**, 679-682 <https://doi.org/10.1038/s41592-022-01488-1>
163. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583-589 <https://doi.org/10.1038/s41586-021-03819-2>
164. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, *et al.* (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. <https://doi.org/10.1101/2021.10.04.463034>
165. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, *et al.* (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins*. **59**, 687-696 <https://doi.org/10.1002/prot.20449>

166. Schanda P, Kupce E, Brutscher B. (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR.* **33**, 199-211 <https://doi.org/10.1007/s10858-005-4425-x>
167. Xu YQ, Matthews S. TROSY NMR Spectroscopy of Large Soluble Proteins. In: Heise H, Matthews S, editors. *Modern Nmr Methodology. Topics in Current Chemistry-Series.* 335. Berlin: Springer-Verlag Berlin; 2013. p. 97-119.
168. Palmer AG, Cavanagh J, Wright PE, Rance M. (1991) Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Correlation NMR-Spectroscopy. *J. Magn. Reson.* **93**, 151-170 [https://doi.org/10.1016/0022-2364\(91\)90036-s](https://doi.org/10.1016/0022-2364(91)90036-s)
169. Kay LE, Keifer P, Saarinen T. (1992) Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy With Improved Sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665 <https://doi.org/10.1021/ja00052a088>
170. Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, Glaser SJ, *et al.* (1994) A General Enhancement Scheme in Heteronuclear Multidimensional NMR Employing Pulsed-Field Gradients. *J. Biomol. NMR.* **4**, 301-306
171. Skinner SP, Fogh RH, Boucher W, Ragan TJ, Mureddu LG, Vuister GW. (2017) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis (vol 66, pg 111, 2016). *J. Biomol. NMR.* **67**, 321-321 <https://doi.org/10.1007/s10858-017-0102-0>
172. Vuister GW, Bax A. (1992) Resolution Enhancement and Spectral Editing of Uniformly <sup>13</sup>C-Enriched Proteins by Homonuclear Broad-Band <sup>13</sup>C Decoupling. *J. Magn. Reson.* **98**, 428-435 [https://doi.org/10.1016/0022-2364\(92\)90144-v](https://doi.org/10.1016/0022-2364(92)90144-v)
173. Piotto M, Saudek V, Sklenar V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous-Solutions. *J. Biomol. NMR.* **2**, 661-665 <https://doi.org/10.1007/bf02192855>
174. Sklenar V, Piotto M, Leppik R, Saudek V. (1993) Gradient-Tailored Water Suppression for <sup>1</sup>H-<sup>15</sup>N HSQC Experiments Optimised to Retain Full Sensitivity. *J. Magn. Reson. Ser. A.* **102**, 241-245 <https://doi.org/10.1006/jmra.1993.1098>



175. Ikura M, Bax A. (1992) Isotope-Filtered 2D NMR of a Protein Peptide Complex - Study of a Skeletal-Muscle Myosin Light Chain Kinase Fragment Bound to Calmodulin. *J. Am. Chem. Soc.* **114**, 2433-2440 <https://doi.org/10.1021/ja00033a019>
176. Breeze AL. (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. Nucl. Magn. Reson. Spectrosc.* **36**, 323-372 [https://doi.org/10.1016/s0079-6565\(00\)00020-0](https://doi.org/10.1016/s0079-6565(00)00020-0)
177. Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119**, 6711-6721 <https://doi.org/10.1021/ja970224q>
178. Ogura K, Terasawa H, Inagaki F. (1996) An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wide-band inversion shaped pulse. *J. Biomol. NMR.* **8**, 492-498 <https://doi.org/10.1007/BF00228150>
179. Iwahara J, Wojciak JM, Clubb RT. (2001) Improved NMR spectra of a protein-DNA complex through rational mutagenesis and the application of a sensitivity optimized isotope-filtered NOESY experiment. *J. Biomol. NMR.* **19**, 231-241 <https://doi.org/10.1023/a:1011296112710>
180. Alves J, Vidugiris G, Goueli SA, Zegzouti H. (2018) Bioluminescent High-Throughput Succinate Detection Method for Monitoring the Activity of JMJC Histone Demethylases and Fe(II)/2-Oxoglutarate-Dependent Dioxygenases. *SLAS Discov.* **23**, 242-254 <https://doi.org/10.1177/2472555217745657>
181. Zhou P, Jin BW, Li H, Huang SY. (2018) HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* **46**, W443-W450 <https://doi.org/10.1093/nar/gky357>
182. McDonough MA, Li V, Flashman E, Chowdhury R, Mohr C, Lienard BMR, *et al.* (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9814-9819 <https://doi.org/10.1073/pnas.0601283103>
183. Koradi R, Billeter M, Wuthrich K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55 [https://doi.org/10.1016/0263-7855\(96\)00009-4](https://doi.org/10.1016/0263-7855(96)00009-4)

184. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. (1993) PROCHECK - a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **26**, 283-291 <https://doi.org/10.1107/s0021889892009944>
185. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics.* **23**, 381-382 <https://doi.org/10.1093/bioinformatics/btl589>
186. Allain F, Mareuil F, Menager H, Nilges M, Bardiaux B. (2020) ARIAweb: a server for automated NMR structure calculation. *Nucleic Acids Res.* **48**, W41-W47 <https://doi.org/10.1093/nar/gkaa362>
187. Linge JP, Habeck M, Rieping W, Nilges M. (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics.* **19**, 315-316 <https://doi.org/10.1093/bioinformatics/19.2.315>
188. Herrmann T, Guntert P, Wuthrich K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209-227 [https://doi.org/10.1016/s0022-2836\(02\)00241-3](https://doi.org/10.1016/s0022-2836(02)00241-3)
189. Habeck M, Rieping W, Linge JP, Nilges M. (2004) NOE assignment with ARIA 2.0: the nuts and bolts. **278**, 379-402 <https://doi.org/10.1385/1-59259-809-9:379>
190. Cooper R. Structural and biophysical characterisation of repetitive bacterial surface proteins [PhD]: University of York; 2021.
191. Cornilescu G, Delaglio F, Bax A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR.* **13**, 289-302 <https://doi.org/10.1023/a:1008392405740>
192. Guntert P, Mumenthaler C, Wuthrich K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283-298 <https://doi.org/10.1006/jmbi.1997.1284>
193. Li SQ, Lian SL, Moser JJ, Fritzler ML, Fritzler MJ, Satoh M, *et al.* (2008) Identification of GW182 and its novel isoform TNGW1 as translational repressors in Ago2-mediated silencing. *J. Cell Sci.* **121**, 4134-4144 <https://doi.org/10.1242/jcs.036905>
194. Lazzaretti D, Tournier I, Izaurralde E. (2009) The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts



- independently of Argonaute proteins. *RNA*. **15**, 1059-1066  
<https://doi.org/10.1261/rna.1606309>
195. Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. (2015) The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518-529 <https://doi.org/10.1002/mrd.22489>
196. Shimberg GD, Pritts JD, Michel SLJ. Iron-Sulfur Clusters in Zinc Finger Proteins. In: David SS, editor. Fe-S Cluster Enzymes, Pt B. Methods in Enzymology. 599. San Diego: Elsevier Academic Press Inc; 2018. p. 101-137.
197. El Mourabit H, Muller S, Tunggal L, Paulsson M, Aumailley M. (2003) Characterization of recombinant and natural forms of the human LIM domain-containing protein FHL2. *Protein Expr. Purif.* **32**, 95-103  
[https://doi.org/10.1016/s1046-5928\(03\)00211-0](https://doi.org/10.1016/s1046-5928(03)00211-0)
198. Archer VEV, Breton J, Sanchezgarcia I, Osada H, Forster A, Thomson AJ, *et al.* (1994) Cysteine-rich LIM Domains of LIM-Homeodomain and LIM-only Proteins Contain Zinc but not Iron. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 316-320  
<https://doi.org/10.1073/pnas.91.1.316>
199. Deane JE, Mackay JP, Kwan AHY, Sum EYM, Visvader JE, Matthews JM. (2003) Structural basis for the recognition of Idb1 by the N-terminal LIM domains of LMO2 and LMO4. *EMBO J.* **22**, 2224-2233  
<https://doi.org/10.1093/emboj/cdg196>
200. Anderson D, Harris R, Polayes D, Ciccarone V, Donahue R, Gerard G, *et al.* (1995) Rapid Generation of Recombinant Baculovirus and Expression of Foreign Genes Using the Bac-to-Bac Expression System. *FASEB J.* **9**, A1464-A1464
201. Gadd MS, Testa A, Lucas X, Chan KH, Chen WZ, Lamont DJ, *et al.* (2017) Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat. Chem. Biol.* **13**, 514-521  
<https://doi.org/10.1038/nchembio.2329>
202. Cardote TAF, Gadd MS, Ciulli A. (2017) Crystal Structure of the Cul2-Rbx1-EloBC-VHL Ubiquitin Ligase Complex. *Structure.* **25**, 901-911  
<https://doi.org/10.1016/j.str.2017.04.009>
203. Flores SK, Cheng ZM, Jasper AM, Natori K, Okamoto T, Tanabe A, *et al.* (2019) Synonymous but Not Silent: A Synonymous VHL Variant in Exon 2 Confers Susceptibility to Familial Pheochromocytoma and von Hippel-Lindau

- Disease. *J. Clin. Endocrinol. Metab.* **104**, 3826-3834  
<https://doi.org/10.1210/jc.2019-00235>
204. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439-D444 <https://doi.org/10.1093/nar/gkab1061>
205. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, *et al.* (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8-D13 <https://doi.org/10.1093/nar/gkx1095>
206. Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J. (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **32**, 113-122  
<https://doi.org/10.1016/j.sbi.2015.03.009>
207. Schutz S, Sprangers R. (2020) Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Prog. Nucl. Magn. Reson. Spectrosc.* **116**, 56-84 <https://doi.org/10.1016/j.pnmrs.2019.09.004>
208. Crublet E, Kerfah R, Mas G, Noirclerc-Savoie M, Lantez V, Vernet T, *et al.* (2014) A cost-effective protocol for the parallel production of libraries of <sup>13</sup>CH<sub>3</sub>-specifically labeled mutants for NMR studies of high molecular weight proteins. *Methods Mol. Biol.* **1091**, 229-244 [https://doi.org/10.1007/978-1-62703-691-7\\_17](https://doi.org/10.1007/978-1-62703-691-7_17)
209. Chowdhury R, Abboud MI, McAllister TE, Banerji B, Bhushan B, Sorensen JL, *et al.* (2020) Use of cyclic peptides to induce crystallization: case study with prolyl hydroxylase domain 2. *Sci. Rep.* **10**, 11 <https://doi.org/10.1038/s41598-020-76307-8>
210. Goto Y, Suga H. (2021) The RaPID Platform for the Discovery of Pseudo-Natural Macrocyclic Peptides. *Accounts Chem. Res.* **54**, 3604-3617  
<https://doi.org/10.1021/acs.accounts.1c00391>
211. McAllister TE, Yeh TL, Abboud MI, Leung IKH, Hookway ES, King ONF, *et al.* (2018) Non-competitive cyclic peptides for targeting enzyme-substrate complexes. *Chem. Sci.* **9**, 4569-4578 <https://doi.org/10.1039/c8sc00286j>
212. Dominguez C, Boelens R, Bonvin A. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731-1737 <https://doi.org/10.1021/ja026939x>

213. Pientka FK, Hu J, Schindler SG, Brix B, Thiel A, Jöhren O, *et al.* (2012) Oxygen sensing by the prolyl-4-hydroxylase PHD2 within the nuclear compartment and the influence of compartmentalisation on HIF-1 signalling. *J. Cell Sci.* **125**, 5168-5176 <https://doi.org/10.1242/jcs.109041>
214. Steinhoff A, Pientka FK, Mockel S, Kettelhake A, Hartmann E, Kohler M, *et al.* (2009) Cellular oxygen sensing: Importins and exportins are mediators of intracellular localisation of prolyl-4-hydroxylases PHD1 and PHD2. *Biochem. Biophys. Res. Commun.* **387**, 705-711 <https://doi.org/10.1016/j.bbrc.2009.07.090>
215. Minervini G, Quaglia F, Tosatto SCE. (2016) Computational analysis of prolyl hydroxylase domain-containing protein 2 (PHD2) mutations promoting polycythemia insurgence in humans. *Sci. Rep.* **6**, 9 <https://doi.org/10.1038/srep18716>
216. Ladroue C, Hoogewijs D, Gad S, Carcenac R, Storti F, Barrois M, *et al.* (2012) Distinct deregulation of the hypoxia inducible factor by *PHD2* mutants identified in germline DNA of patients with polycythemia. *Haematol-Hematol. J.* **97**, 9-14 <https://doi.org/10.3324/haematol.2011.044644>
217. Sattler M, Schleucher J, Griesinger C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93-158 [https://doi.org/10.1016/s0079-6565\(98\)00025-9](https://doi.org/10.1016/s0079-6565(98)00025-9)
218. McIntosh LP, Dahlquist FW. (1990) Biosynthetic Incorporation of <sup>15</sup>N and <sup>13</sup>C for Assignment and Interpretation of Nuclear-Magnetic-Resonance Spectra of Proteins. *Q. Rev. Biophys.* **23**, 1-38 <https://doi.org/10.1017/s0033583500005400>
219. Krishnarjuna B, Jaipuria G, Thakur A, D'Silva P, Atreya HS. (2011) Amino acid selective unlabeled for sequence specific resonance assignments in proteins. *J. Biomol. NMR.* **49**, 39-51 <https://doi.org/10.1007/s10858-010-9459-z>
220. Rasia RM, Brutscher B, Plevin MJ. (2012) Selective Isotopic Unlabeling of Proteins Using Metabolic Precursors: Application to NMR Assignment of Intrinsically Disordered Proteins. *ChemBioChem.* **13**, 732-739 <https://doi.org/10.1002/cbic.201100678>

221. Bellstedt P, Seiboth T, Hafner S, Kutscha H, Ramachandran R, Gorlach M. (2013) Resonance assignment for a particularly challenging protein based on systematic unlabeled amino acids to complement incomplete NMR data sets. *J. Biomol. NMR.* **57**, 65-72 <https://doi.org/10.1007/s10858-013-9768-0>
222. Lacabanne D, Meier BH, Bockmann A. (2018) Selective labeling and unlabeled strategies in protein solid-state NMR spectroscopy. *J. Biomol. NMR.* **71**, 141-150 <https://doi.org/10.1007/s10858-017-0156-z>
223. Schorghuber J, Geist L, Platzer G, Feichtinger M, Bisaccia M, Scheibelberger L, *et al.* (2018) Late metabolic precursors for selective aromatic residue labeling. *J. Biomol. NMR.* **71**, 129-140 <https://doi.org/10.1007/s10858-018-0188-z>
224. Janin J. (1979) Surface and Inside Volumes in Globular Proteins. *Nature.* **277**, 491-492 <https://doi.org/10.1038/277491a0>
225. McCaldon P, Argos P. (1988) Oligopeptide Biases in Protein Sequences and Their use in Predicting Protein Coding Regions in Nucleotide-Sequences. *Proteins.* **4**, 99-122 <https://doi.org/10.1002/prot.340040204>
226. Ruschak AM, Kay LE. (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. *J. Biomol. NMR.* **46**, 75-87 <https://doi.org/10.1007/s10858-009-9376-1>
227. Kerfah R, Hamelin O, Boisbouvier J, Marion D. (2015) CH<sub>3</sub>-specific NMR assignment of alanine, isoleucine, leucine and valine methyl groups in high molecular weight proteins using a single sample. *J. Biomol. NMR.* **63**, 389-402 <https://doi.org/10.1007/s10858-015-9998-4>
228. Tugarinov V, Choy WY, Orekhov VY, Kay LE. (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 622-627 <https://doi.org/10.1073/pnas.0407792102>
229. Sprangers R, Kay LE. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature.* **445**, 618-622 <https://doi.org/10.1038/nature05512>
230. Vuister GW, Kim SJ, Wu C, Bax A. (1994) 2D and 3D NMR-Study of Phenylalanine Residues in Proteins by Reverse Isotopic Labelling. *J. Am. Chem. Soc.* **116**, 9206-9210 <https://doi.org/10.1021/ja00099a041>

231. Tugarinov V, Kay LE. (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR.* **28**, 165-172  
<https://doi.org/10.1023/B:JNMR.0000013824.93994.1f>
232. Lichtenecker R, Ludwiczek ML, Schmid W, Konrat R. (2004) Simplification of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing. **126**, 5348-5349 <https://doi.org/10.1021/ja049679n>
233. Gans P, Hamelin O, Sounier R, Ayala I, Dura MA, Amero CD, *et al.* (2010) Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angew. Chem.-Int. Edit.* **49**, 1958-1962 <https://doi.org/10.1002/anie.200905660>
234. Tugarinov V, Kay LE. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc.* **125**, 13868-13878  
<https://doi.org/10.1021/ja030345s>
235. Lichtenecker RJ, Coudeville N, Konrat R, Schmid W. (2013) Selective Isotope Labelling of Leucine Residues by Using alpha-Ketoacid Precursor Compounds. *ChemBioChem.* **14**, 818-821  
<https://doi.org/10.1002/cbic.201200737>
236. Ruschak AM, Velyvis A, Kay LE. (2010) A simple strategy for <sup>13</sup>C, <sup>1</sup>H labeling at the Ile-gamma 2 methyl position in highly deuterated proteins. *J. Biomol. NMR.* **48**, 129-135 <https://doi.org/10.1007/s10858-010-9449-1>
237. Ayala I, Hamelin O, Amero C, Pessey O, Plevin MJ, Gans P, *et al.* (2012) An optimized isotopic labelling strategy of isoleucine-gamma(2) methyl groups for solution NMR studies of high molecular weight proteins. *Chem. Commun.* **48**, 1434-1436 <https://doi.org/10.1039/c1cc12932e>
238. Godoy-Ruiz R, Guo CY, Tugarinov V. (2010) Alanine Methyl Groups as NMR Probes of Molecular Structure and Dynamics in High-Molecular-Weight Proteins. *J. Am. Chem. Soc.* **132**, 18340-18350  
<https://doi.org/10.1021/ja1083656>
239. Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW. (1989) Expression and <sup>15</sup>N Labelling of Proteins for Proton and <sup>15</sup>N Nuclear-Magnetic-Resonance. *Methods Enzymol.* **177**, 44-73  
[https://doi.org/10.1016/0076-6879\(89\)77005-1](https://doi.org/10.1016/0076-6879(89)77005-1)

240. Takeuchi K, Ng E, Malia TJ, Wagner G. (2007)  $^{13}\text{C}$  amino acid selective labeling in a (HN)- $^2\text{H}$ - $^{15}\text{N}$  background for NMR studies of large proteins. *J. Biomol. NMR.* **38**, 89-98 <https://doi.org/10.1007/s10858-007-9152-z>
241. Ayala I, Sounier R, Use N, Gans P, Boisbouvier J. (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J. Biomol. NMR.* **43**, 111-119 <https://doi.org/10.1007/s10858-008-9294-7>
242. Isaacson RL, Simpson PJ, Liu M, Cota E, Zhang X, Freemont P, *et al.* (2007) A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J. Am. Chem. Soc.* **129**, 15428-15429 <https://doi.org/10.1021/ja0761784>
243. Waugh DS. (1996) Genetic tools for selective labeling of proteins with alpha- $^{15}\text{N}$ -amino acids. *J. Biomol. NMR.* **8**, 184-192
244. Perez LM, Ielasi FS, Bessa LM, Maurin D, Kragelj J, Blackledge M, *et al.* (2022) Visualizing protein breathing motions associated with aromatic ring flipping. *Nature.* **602**, 695-700 <https://doi.org/10.1038/s41586-022-04417-6>
245. Reitzer L. (2005) Catabolism of Amino Acids and Related Compounds. **1**, <https://doi.org/10.1128/ecosalplus.3.4.7>
246. Rajesh S, Nietlispach D, Nakayama H, Takio K, Laue ED, Shibata T, *et al.* (2003) A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp. *J. Biomol. NMR.* **27**, 81-86 <https://doi.org/10.1023/a:1024710729352>
247. Lichtenecker RJ, Weinhaupl K, Schmid W, Konrat R. (2013) alpha-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *J. Biomol. NMR.* **57**, 327-331 <https://doi.org/10.1007/s10858-013-9796-9>
248. Rodriguez-Mias RA, Pellecchia M. (2003) Use of selective TRP side chain labeling to characterize protein-protein and protein-ligand interactions by NMR spectroscopy. *J. Am. Chem. Soc.* **125**, 2892-2893 <https://doi.org/10.1021/ja029221q>
249. Schroghuber J, Sara T, Bisaccia M, Schmid W, Konrat R, Lichtenecker RJ. (2015) Novel Approaches in Selective Tryptophan Isotope Labeling by Using *Escherichia coli* Overexpression Media. *ChemBioChem.* **16**, 746-751 <https://doi.org/10.1002/cbic.201402677>



250. Schorghuber J, Geist L, Bisaccia M, Weber F, Konrat R, Lichtenecker R. (2017) Anthranilic acid, the new player in the ensemble of aromatic residue labeling precursor compounds. *J. Biomol. NMR.* **69**, 13-22 <https://doi.org/10.1007/s10858-017-0129-2>
251. Schorghuber J, Geist L, Platzter G, Konrat R, Lichtenecker RJ. (2017) Highly Selective Stable Isotope Labeling of Histidine Residues by Using a Novel Precursor in *E. coli*-Based Overexpression Systems. *ChemBioChem.* **18**, 1487-1491 <https://doi.org/10.1002/cbic.201700192>
252. Ayala I, Chiari L, Kerfah R, Boisbouvier J, Gans P, Hamelin O. (2020) Asymmetric Synthesis of Methyl Specifically Labelled L-Threonine and Application to the NMR Studies of High Molecular Weight Proteins. *ChemistrySelect.* **5**, 5092-5098 <https://doi.org/10.1002/slct.202000827>
253. Velyvis A, Ruschak AM, Kay LE. (2012) An Economical Method for Production of H-2, (CH3)-C-13-Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 kDa Proteasome. *PLoS One.* **7**, 8 <https://doi.org/10.1371/journal.pone.0043725>
254. Cao C, Chen JL, Yang Y, Huang F, Otting G, Su XC. (2014) Selective <sup>15</sup>N-labeling of the side-chain amide groups of asparagine and glutamine for applications in paramagnetic NMR spectroscopy. *J. Biomol. NMR.* **59**, 251-261 <https://doi.org/10.1007/s10858-014-9844-0>
255. Goux WJ, Strong AAD, Schneider BL, Lee WNP, Reitzer LJ. (1995) Utilisation of Aspartate as a Nitrogen-Source in *Escherichia-Coli* - Analysis of Nitrogen Flow and Characterization of the Products of Aspartate Catabolism. *J. Biol. Chem.* **270**, 638-646
256. Nygaard R, Zou YZ, Dror RO, Mildorf TJ, Arlow DH, Manglik A, *et al.* (2013) The Dynamic Process of beta(2)-Adrenergic Receptor Activation. *Cell.* **152**, 532-542 <https://doi.org/10.1016/j.cell.2013.01.008>
257. Stoffregen MC, Schwer MM, Renschler FA, Wiesner S. (2012) Methionine Scanning as an NMR Tool for Detecting and Analyzing Biomolecular Interaction Surfaces. *Structure.* **20**, 573-581 <https://doi.org/10.1016/j.str.2012.02.012>
258. Mas G, Guan JY, Crublet E, Debled EC, Moriscot C, Gans P, *et al.* (2018) Structural investigation of a chaperonin in action reveals how nucleotide

- binding regulates the functional cycle. *Sci. Adv.* **4**, 9  
<https://doi.org/10.1126/sciadv.aau4196>
259. Gelis I, Bonvin A, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, *et al.* (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell.* **131**, 756-769  
<https://doi.org/10.1016/j.cell.2007.09.039>
260. Fischer M, Kloiber K, Hausler J, Ledolter K, Konrat R, Schmid W. (2007) Synthesis of a <sup>13</sup>C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *ChemBioChem.* **8**, 610-612 <https://doi.org/10.1002/cbic.200600551>
261. Frank L. (1963) Proline Metabolism in *Escherichia Coli* .2. Regulation of Total Growth of Proline Auxotroph by a Proline-Oxidising System. *J. Bacteriol.* **86**, 781-784 <https://doi.org/10.1128/jb.86.4.781-784.1963>
262. Verardi R, Traaseth NJ, Masterson LR, Vostrikov VV, Veglia G. Isotope Labeling for Solution and Solid-State NMR Spectroscopy of Membrane Proteins. In: Atreya HS, editor. *Isotope Labeling in Biomolecular Nmr. Advances in Experimental Medicine and Biology.* 992. Berlin: Springer-Verlag Berlin; 2012. p. 35-62.
263. Driscoll PC. Structure Determination by NMR: Overview. In: Roberts GCK, editor. *Encyclopedia of Biophysics.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 2488-2495.
264. Kumar S, Akabayov SR, Kessler N, Cohen LS, Solanki J, Naider F, *et al.* (2020) The methyl <sup>13</sup>C-edited/<sup>13</sup>C-filtered transferred NOE for studying protein interactions with short linear motifs. **74**, 681-693  
<https://doi.org/10.1007/s10858-020-00340-y>
265. Tripsianes K, Schütz U, Emmanouilidis L, Gemmecker G, Sattler M. (2019) Selective isotope labeling for NMR structure determination of proteins in complex with unlabeled ligands. **73**, 183-189 <https://doi.org/10.1007/s10858-019-00241-9>
266. Anglister J, Srivastava G, Naider F. (2016) Detection of intermolecular NOE interactions in large protein complexes. *Prog. Nucl. Magn. Reson. Spectrosc.* **97**, 40-56 <https://doi.org/10.1016/j.pnmrs.2016.08.002>
267. Kang RS, Daniels CM, Francis SA, Shih SC, Salerno WJ, Hicke L, *et al.* (2003) Solution structure of a CUE-ubiquitin complex reveals a conserved



- mode of ubiquitin binding. *Cell*. **113**, 621-630 [https://doi.org/10.1016/s0092-8674\(03\)00362-3](https://doi.org/10.1016/s0092-8674(03)00362-3)
268. Ramaraju B, McFeeters H, Vogler B, McFeeters RL. (2017) Bacterial production of site specific <sup>13</sup>C labeled phenylalanine and methodology for high level incorporation into bacterially expressed recombinant proteins. *J. Biomol. NMR*. **67**, 23-34 <https://doi.org/10.1007/s10858-016-0081-6>
269. Russell DJ, Hadden CE, Martin CE, Gibson AA, Zens AP, Carolan JL. (2000) A comparison of inverse-detected heteronuclear NMR performance: Conventional vs cryogenic microprobe performance. *J. Nat. Prod*. **63**, 1047-1049 <https://doi.org/10.1021/np0003140>
270. Peterson RD, Theimer CA, Wu HH, Feigon J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR*. **28**, 59-67 <https://doi.org/10.1023/b:jnmr.0000012861.95939.05>
271. Whelan F, Lafita A, Gilbert J, Degut C, Griffiths SC, Jenkins HT, *et al.* (2021) Periscope Proteins are variable-length regulators of bacterial cell surface interactions. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 10 <https://doi.org/10.1073/pnas.2101349118>
272. Elengoe A, Abu Naser M, Hamdan S. (2014) Modeling and Docking Studies on Novel Mutants (K71L and T204V) of the ATPase Domain of Human Heat Shock 70 kDa Protein 1. *Int. J. Mol. Sci.* **15**, 6797-6814 <https://doi.org/10.3390/ijms15046797>
273. Wuthrich K. (2003) NMR studies of structure and function of biological macromolecules (Nobel Lecture). *J. Biomol. NMR*. **27**, 13-39 <https://doi.org/10.1023/a:1024733922459>
274. Vogeli B, Segawa TF, Leitz D, Sobol A, Choutko A, Trzesniak D, *et al.* (2009) Exact Distances and Internal Dynamics of Perdeuterated Ubiquitin from NOE Buildups. *J. Am. Chem. Soc.* **131**, 17215-17225 <https://doi.org/10.1021/ja905366h>
275. Chakraborty C, Mitra S, Roychowdhury A, Samadder S, Dutta S, Roy A, *et al.* (2018) Dereglulation of LIMD1-VHL-HIF-1 alpha-VEGF pathway is associated with different stages of cervical cancer. *Biochem. J.* **475**, 1793-1806 <https://doi.org/10.1042/bcj20170649>

276. Zhane D, Lil S, Yu WB, Chen C, Liu T, Sun YT, *et al.* (2018) LIMD1 is a survival prognostic marker of gastric cancer and hinders tumor progression by suppressing activation of YAP1. *Cancer Manag. Res.* **10**, 4349-4361 <https://doi.org/10.2147/cmar.s174856>
277. Pan JY, Tang YQ, Liu SM, Li LL, Yu B, Lu YY, *et al.* (2020) LIMD1-AS1 suppressed non-small cell lung cancer progression through stabilizing LIMD1 mRNA via hnRNP U. *Cancer Med.* **9**, 3829-3839 <https://doi.org/10.1002/cam4.2898>
278. Wu Y, Wang N, Lei YH, Hu TH, You QD, Zhang XJ. (2016) Small-molecule inhibitors of HIF-PHD2: a valid strategy to renal anemia treatment in clinical therapy. *MedChemComm.* **7**, 1271-1284 <https://doi.org/10.1039/c6md00240d>
279. Zhou JL, Zhang L, Zhou W, Chen YH, Cheng YF, Dong JX. (2019) LIMD1 phosphorylation in mitosis is required for mitotic progression and its tumor-suppressing activity. *FEBS J.* **286**, 963-974 <https://doi.org/10.1111/febs.14743>
280. Johnson BA, Blevins RA. (1994) NMR View - a Computer-Program for the Visualisation and Analysis of NMR Data. *J. Biomol. NMR.* **4**, 603-614 <https://doi.org/10.1007/bf00404272>