

---

# The Interaction of $\alpha$ -Actinin-2 with ZASP and Titin

---

By

Yunghan Au

January 2004

Thesis for the degree of Doctor of Philosophy  
from University College London

Supervisor: Dr. Annalisa Pastore

Division of Molecular Structure  
National Institute for Medical Research  
The Ridgeway, Mill Hill  
London, NW7 1AA

UMI Number: U602696

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U602696

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

*To my parents*

*Perhaps the most valuable result of all education is the ability to make yourself do the thing you have to do, when it ought to be done, whether you like it or not; it is the first lesson that ought to be learned; and however early a man's training begins, it is probably the last lesson that he learns thoroughly.*

Thomas Henry Huxley (1825-1895)



---

# Abstract

---

Z-band Alternately Spliced PDZ-containing protein (ZASP) is a sarcomeric Z-disk protein expressed in human cardiac and skeletal muscle that is involved in a dominant familial dilated cardiomyopathy. ZASP interacts with the last 150 amino acids of  $\alpha$ -actinin-2, the major component of the Z-disk, via an N-terminal PDZ domain. ZASP and  $\alpha$ -actinin-2 are thought to play an important structural role, probably by forming a ternary complex with titin Z-repeats. We have determined the structure of ZASP PDZ and characterised its interaction with  $\alpha$ -actinin-2. We show that ZASP PDZ is a classical class 1 PDZ domain that recognises with micromolar affinity the carboxy-terminal sequence of an  $\alpha$ -actinin-2 calmodulin-like domain. We also characterised the ternary complex ZASP/ $\alpha$ -actinin-2/titin and the role of each component, showing that the  $\alpha$ -actinin-2/ZASP PDZ interaction involves a binding surface distinct from that involved in the recognition of the titin Z repeats. Finally, we used the ZASP PDZ structure to model other members of the enigma family by homology and to predict their abilities to bind  $\alpha$ -actinin-2.

---

# Table Of Contents

---

<b>Abstract .....</b>	<b>4</b>
<b>Table Of Contents .....</b>	<b>5</b>
<b>List Of Figures .....</b>	<b>8</b>
<b>List Of Tables.....</b>	<b>16</b>
<b>Acknowledgements .....</b>	<b>17</b>
<b>Abbreviations.....</b>	<b>18</b>
 <b>Chapter 1.....</b>	 <b>20</b>
<b>Introduction .....</b>	<b>20</b>
1.1 Muscle .....	20
1.2 $\alpha$ -Actinin-2 .....	27
1.3 Titin .....	31
1.4 ZASP .....	33
1.5 Nuclear Magnetic Resonance .....	37
1.6 Aims .....	39
 <b>Chapter 2.....</b>	 <b>40</b>
<b>Materials and Methods .....</b>	<b>40</b>
2.1 Bioinformatics .....	40
2.2 Plasmid Production.....	40
2.3 Protein Expression.....	41
2.3.1 Transformation of Competent Cells .....	41
2.3.2 Fermentation.....	41
2.4 Protein Purification.....	43
2.4.1 Cell Lysis.....	43
2.4.2 Nickel NTA Agarose Protein Purification From Lysate .....	44
2.4.3 Tobacco-Etch Virus Cleavage.....	44
2.4.4 Gel Filtration (FPLC) Protein Purification.....	44
2.4.5 Nickel NTA Protein Purification After TEV Cleavage.....	45
2.4.6 Buffer Exchange Using PD-10 Columns .....	45
2.4.7 Concentration Determination Using Absorption Spectroscopy .....	45
2.5 NMR.....	47
2.5.1 Spectroscopy .....	47
2.5.2 Spectral Processing.....	48
2.6 NMR Structure Determination .....	49
2.6.1 NOESY spectra .....	50
2.6.2 Distance Calibration .....	50
2.6.3 Hydrogen Bonds, Dihedral Angles, and Residual Dipolar Couplings .....	51
2.6.4 CYANA.....	52
2.6.5 CANDID .....	52
2.6.6 XPLOR.....	52
2.6.7 CHARM22 Water Refinement.....	53
2.7 $^{15}\text{N}$ -HSQC Titration .....	54
2.8 Fluorometric Titration .....	56
2.9 Peptide synthesis and purification of the Minimal Titin ZR7 peptide .....	57

<b>Chapter 3.....</b>	<b>58</b>
<b>Protein Expression and Purification.....</b>	<b>58</b>
3.1 Protein Purification of ZASP-PDZ.....	58
3.2 Protein Purification of Act-EF34 .....	61
3.3 Purification of Act-EF1234.....	64
3.4 Conclusion.....	66
<b>Chapter 4.....</b>	<b>67</b>
<b>Resonance Assignment: ZASP-PDZ with and without Act-EF1234 .....</b>	<b>67</b>
4.1 Assignment of ZASP-PDZ.....	67
4.2 Assignment of <sup>15</sup> N-ZASP-PDZ bound to Act-EF1234 .....	70
4.3 Assignment of Act-EF34.....	72
4.4 Assignment of Act-EF1234.....	77
4.5 Conclusion.....	79
<b>Chapter 5.....</b>	<b>80</b>
<b>NMR Structure of ZASP-PDZ .....</b>	<b>80</b>
5.1 Secondary Structure Prediction of ZASP-PDZ .....	80
5.2 Manual Structure Determination .....	81
5.3 Automatic Structure Determination .....	86
5.4 Structure Description.....	87
5.5 Restraint Comparison: Manual vs. Automatic .....	89
5.6 Physical Comparison of ZASP-PDZ structure calculations.....	92
5.7 Precision, Quality, and Accuracy Comparison .....	99
5.7.1 Precision: RMSD Calculations.....	99
5.7.2 Quality: WHATIF scores .....	101
5.7.3 Quality: Ramachandran plots .....	103
5.7.4 Accuracy: Comparison Of Known PDZ Domains .....	105
5.8 Conclusion.....	108
<b>Chapter 6.....</b>	<b>110</b>
<b>Interaction of <math>\alpha</math>-Actinin-2 with ZASP and Titin .....</b>	<b>110</b>
6.1 NMR Titrations .....	111
6.1.1 <sup>15</sup> N-Act-EF1234 and <sup>15</sup> N-Act-EF34 with ZASP-PDZ .....	111
6.1.2 <sup>15</sup> N-ZASP-PDZ with Act-EF1234 and Act-EF34.....	113
6.1.3 <sup>15</sup> N-Act-EF34 in complex with Minimal Titin ZR7 and ZASP-PDZ .....	115
6.1.4 NMR Kd calculations between ZASP, $\alpha$ -actinin-2, and Titin.....	118
6.2 Fluorometric Titration .....	120
6.3 Modelling of $\alpha$ -Actinin, Titin and ZASP Tertiary Complex.....	121
6.4 Conclusion.....	124
<b>Chapter 7 .....</b>	<b>128</b>
<b>Homology Modelling of the Enigma Family PDZ Domains .....</b>	<b>128</b>
7.1 Homology modelling of the enigma family of PDZ .....	129
7.2 Conclusion.....	133
<b>Chapter 8.....</b>	<b>134</b>
<b>Discussion .....</b>	<b>134</b>
<b>Chapter 9.....</b>	<b>138</b>
<b>Bibliography.....</b>	<b>138</b>

<b>Chapter 10</b>	<b>150</b>
<b>Appendix</b>	<b>150</b>
10.1 Protein Sequences	150
10.1.1 ZASP [1-85] (ZASP-PDZ)	150
10.1.2 $\alpha$ -Actinin-2 [745-894] (Act-EF1234)	150
10.1.3 $\alpha$ -Actinin-2 [823-894] (Act-EF34)	150
10.1.4 Titin ZR7	150
10.1.5 Titin ZR7 minimum sequence	150
10.2 Media and Buffers	151
10.2.1 Luria-Bertani Media (LB)	151
10.2.2 Preparation of M9 Minimal Media	151
10.2.3 Nickel-NTA Column Buffers	152
10.2.4 FPLC Column Buffer	152
10.2.5 NMR Sample Buffer	152
10.2.6 Broad Molecular Weight Marker	153
10.3 NMR Acquisition Parameters	154
10.3.1 ZASP-PDZ Spectra	154
10.3.2 Act-EF34 Spectra	161
10.4 NMR Spectral Assignment	164
10.4.1 Assignment of ZASP-PDZ	164
10.4.2 Assignment of ZASP-PDZ bound to Act-EF1234	170
10.4.3 Assignment of Act-EF34	172
10.5 CYANA Manual Calculation	174
10.5.1 Init.cya	174
10.5.2 Calibration.cya	175
10.5.3 CYCLE.cya	175
10.5.4 REDAC Annealing Protocol (ANNEAL_STR.cya)	176
10.6 CYANA Automatic Calculation (CANDID)	177
10.6.1 Init.cya	177
10.6.2 CANDID.CYA	177
10.7 Homology Modelling	178
10.8 Titration Data	179
10.8.1 Titration of $^{15}\text{N}$ -ZASP-PDZ with Unlabelled Act-EF1234	179
10.8.2 Titration of $^{15}\text{N}$ -ZASP-PDZ with Unlabelled Act-EF34	179
10.8.3 Titration of $^{15}\text{N}$ -Act-EF1234 with Unlabelled ZASP-PDZ	180
10.8.4 Titration of $^{15}\text{N}$ -Act-EF34 with Unlabelled ZASP-PDZ	180
10.8.5 Titration of $^{15}\text{N}$ -Act-EF34 with Unlabelled tZR7, followed by ZASP-PDZ	181
10.9 Perl Scripts	182
10.9.1 D2X.PL: DYANA To XPLOR Distance Restraint Conversion Script	182
10.9.2 TALOS2DYANA.PL: TALOS To DYANA Dihedral Angle Converter	189
10.9.3 ACO2TBL: DYANA to XPLOR Dihedral Angle Conversion	190
10.9.4 COMPARE.PL: Dyana Restraint Comparison Script	191
10.9.5 DIST_SEARCH.PL: MOLMOL Distance To XEASY Possible Assignment Comparison	192
10.9.6 COMBINE.PL: XEASY Atom List Combiner	193
10.9.7 CALIBRATE.PL: Volume/Intensity To Distance Calibration	194

---

# List Of Figures

---

- Figure 1.1:** a) Schematic representation of the sarcomere. The position of the Z-disk and M-line is indicated. A single molecule titin filament (indicated in green) connects the Z-disk to the M-line. In the Z-disk  $\alpha$ -actinin-2 (in yellow) forms transversal connections between actin filaments (in red) and is bound to ZASP (in blue). Myosin and F-actin are coloured purple and red respectively. b) Domain architecture of the titin region localised in the Z-disk,  $\alpha$ -actinin-2 and ZASP (variant 3).....21
- Figure 1.2:** Transmission Electron Microscope picture of striated mammalian muscle (<http://www.udel.edu/Biology/Wags/>), labelled to indicate the position of the sarcomere, I and A bands, Z and M lines, and the sarcoplasmic reticulum. ....21
- Figure 1.3:** The myosin II molecule, indicating the production of different subfragments by two types proteolytic cleavage: trypsin cleavage resulting in HMM and LMM, and papain cleavage, producing S1 and S2.....23
- Figure 1.4:** The lever arm hypothesis of muscle contraction, adapted from (Rayment et al., 1993a). The diagram shows the process between myosin, actin, and ATP. The cycle begins from actin and myosin bound together in the 'rigor complex', after which ATP hydrolysis leads to a series of step-wise conformational changes. The sequence of events ends with the power-stroke, hence, re-starting the cycle. ....25
- Figure 1.5:** The Telegraph 28 August 2003 .....29
- Figure 1.6:** Diagram of the domain organisation of  $\alpha$ -actinin-2 and its binding partners. The arrows indicate the region of  $\alpha$ -actinin-2 that the binding partners have been found to interact with - CapZ has only been localised to all of the spectrin-like motifs. The program SMART (Schultz et al., 2000) was used to predict the different domains: calponin-homology domain (CH); spectrin-like domains (SPEC), and EF-hand domains (EFh). Interacting proteins include actin (Panassenko and Gusev, 2001), titin/connectin (Ohtsuka et al., 1997b) ZASP (Faulkner et al., 1999), ALP (Xia et al., 1997), CapZ (Papa et al., 1999), Myotilin (Salmikangas et al., 1999), Myopalladin (Bang et al., 2001), FATZ/Calsarcin-2/Myozenin (Faulkner et al., 2000; Frey et al., 2000; Takada et al., 2001) .....29
- Figure 1.7:** Solved structures from regions of  $\alpha$ -actinin-2; a) the second calponin homology domain of  $\beta$ -spectrin (1AA2), 57% sequence identity with  $\alpha$ -actinin-2's CH2 domain, coloured by secondary structure succession (from blue to red); b) the

last two EF-hands in complex with titin's seventh Z-repeat (1H8B), coloured yellow and green respectively; c) four spectrin-like repeats in an anti-parallel homodimer (1HCI), coloured green and orange to demarcate each monomer. B and C have blue and red spheres to distinguish the N and C termini respectively. ....	30
<b>Figure 1.8:</b> Predicted domain representation of the primary sequence of the cardiac titin filament cDNA (From (Labeit and Kolmerer, 1995)). The immunoglobulin-like domains are represented in red, while the fibronectin type 3-like domains are in white. Irregular areas of sequences are shown in blue, PEVK region in yellow, and titin kinase in black. Titin domains have been labelled according to the main sarcomeric region (Z-line, A-band, I-band, and M-line). Note that the Z-repeats are found between Ig domains 3 and 4 in the Z-disk region of titin.....	32
<b>Figure 2.1:</b> Schematic of the steps of protein purification for 6HIS (6H), GST, and 'TEV cleavage site' (red circles) tagged proteins. Cells are lysed after protein expression, and the cell lysate is passed down a nickel-agarose column. The fusion protein is recovered and subsequently cleaved using the TEV protease. The target protein is then purified from the 6HIS-GST by size exclusion or another nickel-agarose purification. ....	43
<b>Figure 2.2:</b> NMR structure determination pipeline. First, the backbone atoms of the protein 'chemical shift' are assigned, allowing for sidechain atoms assignment. Structural restraints are then analysed (distance, angular, and orientational), which are then used for structure calculation.....	49
<b>Figure 3.1:</b> Polyacrylamide SDS gel of the purification of ZASP-PDZ; from cell lysis up till TEV protease cleavage. 1) Broad molecular weight marker (Appendix 10.2.6), 2) cell lysate supernatant, 3 to 7) nickel agarose purification (in order, unbound fraction, column buffer wash, '1M NaCl' wash, 30 mM imidazole wash, and 300 mM imidazole wash, 8) TEV protease cleavage product containing 6HIS-GST and ZASP-PDZ. ....	59
<b>Figure 3.2:</b> FPLC elution profile of ZASP-PDZ monitoring fluorescence at 260 and 280 nm. ....	59
<b>Figure 3.3:</b> Polyacrylamide SDS gel of a ZASP-PDZ FPLC result. 1) Marker, 2) un-cleaved 6HIS/GST/ZASP-PDZ, 3) GST (eluent at 55 ml from Figure 3.2), 4-7) Pure ZASP-PDZ elution (elution volume of 95 ml). ....	60
<b>Figure 3.4:</b> UV/Visible light absorbance spectra of ZASP-PDZ (at 0.57 mM), used to measure protein concentration at 280 nm.....	60

<b>Figure 3.5:</b> Polyacrylamide SDS gel of Act-EF34 purification, from cell lysis until TEV protease cleavage. 1) Molecular weight marker, 2) cell lysate, 3 to 7) nickel agarose purification (in order of, unbound fraction, column buffer wash, '1M NaCl' wash, 30 mM imidazole wash, and 300 mM imidazole wash, 8) TEV protease cleavage product. ....	62
<b>Figure 3.6:</b> UV/Visible light absorbance spectra of Act-EF34 (at 0.27 mM), used to measure protein concentration at 280 nm. ....	62
<b>Figure 3.7:</b> FPLC elution profile of Act-EF34 monitoring fluorescence at 260 and 280 nm. Act-EF34 is shown here to elute at a volume of 90 ml. ....	63
<b>Figure 3.8:</b> Polyacrylamide SDS Gel of an Act-EF34 FPLC result. 1) Marker, 2) 6HIS-GST-Act-EF34 fusion protein (elution volume of 55 ml) 3) GST (elution volume of 70 ml), 4-7) Pure Act-EF34 elution (elution volume of 90 ml). ....	63
<b>Figure 3.9:</b> Polyacrylamide SDS gel of the purification of Act-EF1234 from cell lysis until TEV protease cleavage. 1) Broad molecular weight marker, 2) cell lysate, 3 to 7) nickel agarose purification (in order, unbound fraction, column buffer wash, '1M NaCl' wash, 30 mM imidazole wash, and 300 mM imidazole wash). ....	64
<b>Figure 3.10:</b> Polyacrylamide SDS gel of the nickel agarose column purification for Act-EF1234. 1) Molecular weight marker, 4) Elution from first nickel agarose column, 5) TEV cleavage product, 6) Second nickel agarose column elution 7) Third nickel agarose column elution containing pure Act-EF1234. ....	65
<b>Figure 3.11:</b> UV/Visible light absorbance spectra of Act-EF1234 (at 0.51 mM), used to measure protein concentration at 280 nm. ....	65
<b>Figure 4.1:</b> $^{15}\text{N}$ -HSQC spectrum of ZASP-PDZ at pH 6.6 and 27 °C, showing a well-resolved spectrum. The spectrum is labelled with amide backbone resonance assignments, and the solid horizontal lines indicate side-chain amide protons of asparagine and glutamine residues. ....	68
<b>Figure 4.2:</b> Sample of the CBCANH spectra of ZASP-PDZ – red and blue peaks is C $\alpha$ and C $\beta$ resonances respectively, and the residue type and number is indicated at the bottom. The black lines highlight backbone connectivity. ....	69
<b>Figure 4.3:</b> Sample of the $^{15}\text{N}$ -NOESY-HSQC comparison of the unbound and Act-EF1234 bound forms of $^{15}\text{N}$ -ZASP-PDZ - the residue type and number is indicated at the bottom. * Denotes the bound form of ZASP, while red and green peaks denote assigned and unassigned peaks respectively. ....	71

- Figure 4.4:** Schematic of the mechanism of TCEP reduction of disulphide bonds. TCEP reacts with water to reduce the disulphide bonds, resulting in an oxidised phosphorus atom.....72
- Figure 4.5:** Ribbon representation of Act-EF34 (yellow) in complex with titin-ZR7 (green), indicating the position of cysteine 41 (van der Waals representation). Titin-ZR7 may prevent Act-EF34 homodimer formation, as it buries the thiol group between the surface of interaction.....73
- Figure 4.6:** Polyacrylamide SDS gel electrophoresis of A: Sample of freshly prepared Act-EF34 revealing a trace of the dimeric second species (4) using the Broad Molecular Marker (1) for size estimation. B: Non-reducing polyacrylamide SDS gel of an old sample of Act-EF34 (B) in the presence of 3:1 (C) and 10:1 TCEP (D) to protein ratios.....73
- Figure 4.7:**  $^{15}\text{N}$ -HSQC of ‘fresh’ Act-EF34 without TCEP. The protein appears unfolded or aggregated, as the peaks are ‘clumped’ in the centre of the spectrum.....74
- Figure 4.8:**  $^{15}\text{N}$ -HSQC of Act-EF34 with 1:3 ratio of TCEP. The relative intensities of the peaks have increased, and some of the peaks from the second species have disappeared.....74
- Figure 4.9:** Graph showing the change in relative intensity (to the C-terminal residue) on addition of a molar ratio of 3:1 TCEP to Act-EF34. This indicates the formation of the monomeric species from the disulphide-bonded homodimer: line-widths are narrower for smaller molecular species, hence, an increase in intensity indicates a reduction in size.....75
- Figure 4.10:**  $^{15}\text{N}$ -HSQC spectrum of Act-EF34 at pH 6.6 and 27 °C, showing amide backbone resonance assignments. Solid horizontal lines indicate side-chain amide protons of asparagine and glutamine residues, and the arrows indicate assignment of ‘second species’ peaks.....76
- Figure 4.11:**  $^{15}\text{N}$ -HSQC superimposition of ZASP-PDZ bound to  $^{15}\text{N}$ -Act-EF34 (black peaks) and  $^{15}\text{N}$ -Act-EF1234 (red peaks). While some peaks are super-imposable, most are not. See (Joseph et al., 2001) for both Act-EF34 and Act-EF1234 converging upon binding titin-ZR7. ....78
- Figure 5.1:** Secondary structure prediction of ZASP-PDZ domain. H represents  $\alpha$ -helices, and E represents  $\beta$ -strands.....81
- Figure 5.2:** Extract of the  $^{15}\text{N}$ -NOESY HSQC of  $^{15}\text{N}$ -ZASP-PDZ. The  $^{15}\text{N}$ -HSQC (Figure 4.1 –  $^{15}\text{N}$ -HSQC) showed good dispersion of the peaks, therefore the  $^{15}\text{N}$ -



NOESY HSQC was also well resolved. The residue type and number is indicated at the bottom of each strip. ....	82
<b>Figure 5.3:</b> Phylogenetic tree derived from a CLUSTALX alignment of all solved PDZ domains in the PDB, generated using NJplot (Perriere and Gouy, 1996). Although IL16 was the closest relation to ZASP-PDZ, after taking into consideration gaps and insertions, 1BFE/1BE9 was a better template for ZASP-PDZ's homology model.....	84
<b>Figure 5.4:</b> Pair-wise CLUSTALX alignment of ZASP-PDZ and 1BE9/1BFE, showing a 22.5% sequence identity and 47.5% sequence homology. ....	84
<b>Figure 5.5:</b> Homology model of ZASP-PDZ, generated by SWISS-MODEL, based on the alignment with 1BFE – residues coloured in red, orange, and blue represent identical, chemically similar, and no similarity respectively. ....	85
<b>Figure 5.6:</b> Stereogram of the bundle of the best 20 structures from the ZASP-PDZ CANDID automatic structure determination using just NOE distance constraints. The structure is incorrectly folded and parts of the secondary structure have not been formed . ....	87
<b>Figure 5.7:</b> A backbone ensemble stereogram of the structure of ZASP-PDZ (CCANDID). ....	88
<b>Figure 5.8:</b> Ribbon diagram depicting secondary structure elements of ZASP-PDZ (red = $\alpha$ helices, yellow = $\beta$ strands) (CCANDID). ZASP-PDZ has 6 $\beta$ -strands and 2 $\alpha$ -helices forming a $\beta$ -sandwich.....	88
<b>Figure 5.9:</b> Distance constraint plot of the A) manual and the B) CANDID structure calculation. White, light grey, dark grey, and black colours represent intra-residue, short-range, medium-range, and long-range distance connectivities respectively.....	90
<b>Figure 5.10:</b> Structure of ZASP-PDZ. NMR bundles corresponding to the manually (top row) and automatically (bottom row) assigned structure families. From left to right are displayed the bundles without refinement (A and D), with RDC (B and E) and water refinement (C and F) respectively. All structures have no distance violations greater than 0.5 Å and no dihedral angle violations greater than 5 ° from the experimental data. ....	93
<b>Figure 5.11:</b> Comparison of ZASP-PDZ structure calculations (lowest RMSD to the mean) for both manual (top) and CANDID (bottom), after conversion from CYANA to XPLOR (left) and water-refinement (right). The structures of the comparison (labelled) have been superimposed, and colour coded from blue to red to denote the change in RMSD (low to high respectively). ....	95

**Figure 5.12:** Comparison of ZASP-PDZ structure calculations (lowest RMSD to the mean) for both manual (A) and CANDID (B). The result for the CANDID/XPLOR comparison is in black, whilst the result for the XPLOR/water-refinement is in red. .96

**Figure 5.13:** Summary of the average pair-wise RMSD between the structure calculations. The first four values show less difference between the manual and CANDID structures between the conversions, while the last 3 values signify the difference between the manual and CANDID procedures.....97

**Figure 5.14:** A) Comparison of ZASP-PDZ manual and CANDID structure calculations (lowest RMSD to the mean). The result for the CMANUAL/CCANDID comparison is in black, XMANUAL/XCANDID is in red, whilst the result for the RXMANUAL/RXCANDID is in blue. B) The residue-by-residue comparison of the manual and CANDID calculation, in both XPLOR and CYANA and after water refinement. The structures of the comparison (labelled) have been superimposed, and colour coded from blue to red to denote the change in RMSD (low to high respectively). .....98

**Figure 5.15:** Comparison of the backbone and heavy atom RMSD for the best 20 structures from all structure calculations. RMSD calculations include the overall result, and an ordered selection of residues (3-7,9-10,12,15-30,32-37,39-58,60-72, and 76-82). The CANDID group of structures consistently have better precision to their manual calculation counterpart.....99

**Figure 5.16:** Comparison of the WHATIF structure (A) and RMS (B) Z-scores for all structure calculations. The RXCANDID structures give consistently the best structure Z-scores, with the exception of '1<sup>st</sup> generation packing quality', while RXMANUAL and RXCANDID structures give a comparably good 'RMS Z-score' .....102

**Figure 5.17:** Ramachandran plot of RXCANDID generated in PROCHECK-NMR. Squares and triangles represent non-glycine residues and glycine residues respectively.....103

**Figure 5.18:** A) Ramachandran plot comparison of all of the structure calculations, giving the distribution of residues in the most favoured (red), additionally allowed (yellow), generously allowed (light blue), and disallowed regions (white). Graph B has been zoomed to highlight the differences in the additionally, generously, and disallowed regions. RXCANDID has the best overall Ramachandran scores. ....104

**Figure 5.19:** Pair-wise comparison of the manually and automatically assigned bundles, 1kwa and 1qav structures with 1pdr, taken as a reference. The pair-wise Z scores as obtained by WHATIF MOTIF option are plotted versus the sequence of 1pdr. The

secondary structure elements of 1pdr are indicated underneath with yellow ( $\beta$ -strands) and red ( $\alpha$ -helices) boxes.....	107
<b>Figure 6.1:</b> $^{15}\text{N}$ -HSQC titration of $^{15}\text{N}$ -Act-EF1234 with (red) and without (black) ZASP-PDZ (2:1 ZASP-PDZ excess, final titration point). Run at 600 MHz, 300 K, with 0.46 mM (initial) $^{15}\text{N}$ -Act-EF1234.....	112
<b>Figure 6.2:</b> $^{15}\text{N}$ -HSQC of $^{15}\text{N}$ -Act-EF34 with (red) and without (black) ZASP-PDZ (2:1 ZASP-PDZ excess, final titration point). Run at 500 MHz, 300K, 0.60 mM (initial) $^{15}\text{N}$ -Act-EF34. ....	112
<b>Figure 6.3:</b> $\Delta\delta_{\text{weighted}}$ changes in $^{15}\text{N}$ -Act-EF34 upon binding ZASP-PDZ. The changes appear to be localised to Act-EF34's C-terminus. ....	113
<b>Figure 6.4:</b> $^{15}\text{N}$ -HSQC of $^{15}\text{N}$ -ZASP-PDZ with (red) and without (black) Act-EF1234 (2:1 Act-EF1234 excess, final titration point). Run at 600 MHz, 300K, 0.30 mM (initial) $^{15}\text{N}$ -ZASP-PDZ. ....	114
<b>Figure 6.5:</b> $^{15}\text{N}$ -HSQC of $^{15}\text{N}$ -ZASP-PDZ with (red) and without (black) Act-EF34 (2:1 Act-EF34 excess, final titration point). Run at 600 MHz, 300K, 0.37 mM (initial) $^{15}\text{N}$ -ZASP-PDZ. ....	114
<b>Figure 6.6:</b> $\Delta\delta_{\text{weighted}}$ changes of $^{15}\text{N}$ -ZASP-PDZ with Act-EF1234 (white) and Act-EF34 (black).....	115
<b>Figure 6.7:</b> $^{15}\text{N}$ -HSQC of $^{15}\text{N}$ -Act-EF34 with tZR7 (black) and ZR7 (red). Both spectra were run at 600 MHz, 300K, and with an excess of 2:1 ZR7 or tZR7.....	116
<b>Figure 6.8:</b> $^{15}\text{N}$ -HSQC of $^{15}\text{N}$ -Act-EF34/tZR7 with (red – final titration point) and without (black) ZASP-PDZ. Run at 500 MHz, 300K, 0.53 mM (initial) $^{15}\text{N}$ -Act-EF34 with a 2:1 excess of tZR7 and 1.5:1 excess of ZASP-PDZ.....	116
<b>Figure 6.9:</b> $\Delta\delta_{\text{weighted}}$ changes for $^{15}\text{N}$ -Act-EF34/tZR7, after binding ZASP-PDZ. ....	117
<b>Figure 6.10:</b> Comparison of the changes in $\Delta\delta_{\text{weighted}}$ against residue number for the $^{15}\text{N}$ -Act-EF34/titin-ZR7 complex, after binding ZASP-PDZ. ....	117
<b>Figure 6.11:</b> Binding graphs of A) $^{15}\text{N}$ -ZASP-PDZ against Act-EF1234 (from N67), B) $^{15}\text{N}$ -ZASP-PDZ against Act-EF34 (from N67), C) $^{15}\text{N}$ -Act-EF34 against ZASP-PDZ (from S64), and D) $^{15}\text{N}$ -Act-EF34/tZR7 against ZASP-PDZ (from S64). All graphs display stoichiometric binding, leading to an overestimation of the $K_d$ calculated by non-linear regression analysis. ....	119
<b>Figure 6.12:</b> Fluorescence titration of ZASP-PDZ against Act-EF34 (black diamonds), and the Act-EF34/tZR7 complex (white open circles).....	120
<b>Figure 6.13:</b> A) ZASP-PDZ with highlighted $\Delta\delta_{\text{weighted}}$ changes from binding Act-EF34 with residues in intermediate exchange in purple (13, 14, 15), $\Delta\delta_{\text{weighted}}$ changes	

of over 0.3 in red (17, 18), and over 0.2 in yellow (16, 20, 28, 31). W13 is shown in stick representation. B) Act-EF34 (blue) in complex with titin-ZR7 (green) showing changes in its changes in the C-terminus with the same colour scheme as above: in purple - L73; red - G69, S71 and D72, and yellow – L67.....	122
<b>Figure 6.14:</b> Ribbon representation of the Act-EF34/tZR7/ZASP-PDZ trimeric complex, rotated to highlight the ZASP-PDZ/Act-EF34 interaction (A) and the Act-EF34/ZR7 interaction (B).....	123
<b>Figure 6.15:</b> Hydrogen bond network between ZASP-PDZ (yellow ribbons) and Act-EF34 (stick to red ribbon representation) in the complex, forming six possible hydrogen bonds. ....	126
<b>Figure 6.16:</b> Connolly surface representation (solvent accessibility; probe radius of 1.4 Å) of the trimeric complex. This shows electrostatic and surface complementarity between ZASP-PDZ and Act-EF34.....	127
<b>Figure 6.17:</b> Interacting surface between ZASP-PDZ and Act-EF34 coloured by residue property at pH 7: Grey indicates non-polar hydrophobic, yellow - polar, red - negative charge, and blue indicates a positively charged residue. ....	127
<b>Figure 7.1:</b> ClustalX sequence alignment of the results from the BLAST search from ZASP's PDZ domain. The subgroups in the enigma family (enigma, ENH, ZASP, Elfin, RIL and ALP) group together, and are indicated on the alignment. There is almost no variation in sequence in the PDZ domains of each subgroup.....	131
<b>Figure 7.2:</b> Homology models of the PDZ domains of ENH, enigma, elfin, ALP and RIL, generated using SWISS-PDB Viewer and SWISS-MODEL – 'ALL' refers to the sequence conservation across the whole enigma family. Highlighted red and yellow residues represent identical and similar residues respectively. The green arrow on 'ALL' indicates the binding pocket in class I PDZ target recognition.....	132
<b>Figure 8.1:</b> Mapping of point mutations described in (Zhou et al., 2001) onto the structure of ZASP-PDZ. ....	135
<b>Figure 8.2:</b> Ribbon representation of ZASP-PDZ showing A: the conserved residues from the sequence alignment (identical and similar residues in red and orange respectively), and B: the residues most affected upon binding $\alpha$ -actinin-2 (coloured from red, orange and yellow in order of $\Delta\delta_{\text{weighted}}$ change – red being the highest). This shows that the many of the residues directly involved in binding $\alpha$ -actinin-2 are conserved within the enigma family.....	137

---

# List Of Tables

---

**Table 1.1:** SMART domain predictions and information regarding their residue number and molecular weight of the different splice variants of ZASP, cypher (aka oracle). Red hexagons represent PDZ domains, yellow triangles represent LIM domains, and purple and grey sections represent complex and non-domain containing regions respectively. ZASP1, cypher2c and cypher2s do not contain the LIM domains. Note that cypher's isoforms have been characterised in terms of expression in skeletal (s) and cardiac muscle (c), hence the s/c suffix. ....36

**Table 5.1:** Summary of the analysis of the  $^1\text{H}$ - $^1\text{H}$ -NOESY, and  $^{13}\text{C}$  and  $^{15}\text{N}$  NOESY-HSQC spectra: i.e. number of peaks picked, number of assignments made, and percentage assigned. ....81

**Table 5.2:** Summary of the final distance restraints used for the CMANUAL structure calculation. Unique distances are those present after combining the restraints from all of the NOESY spectra, while modified distances are superfluous restraints. The final restraints are the modified restraints after taking into account stereo-specific assignment. ....83

**Table 5.3:** Comparison of the manual and automatic assignment of ZASP-PDZ. \* Only added to the XCANDID and XMANUAL refinements. ....91

**Table 5.4:** Values from the RMSD calculation, WHATIF scores, and Ramachandran plot statistics for each structure calculation. RMSD has been calculated overall and with only the selection of residues 3-7,9-10,12,15-30,32-37,39-58,60-72, and 76-82 (ordered). Backbone atoms are defined by the backbone C $\alpha$ , N, and CO, while C, N, O, and S define the heavy atoms.....100

**Table 5.5:** Results from the DALI search of ZASP-PDZ.....106

**Table 7.1:** Details of the enigma family members including species, tissue type, ability to bind  $\alpha$ -actinin-2, location of  $\alpha$ -actinin-2 binding, and sequence identity and similarity with respect to ZASP-PDZ. \* *H* = *Homo Sapiens*, *M* = *Mus Musculus*, *R* = *Rattus Norvegicus*, *S* = *Salmo Salar*, *G* = *Galus Galus*, \*\* *H* = Heart, *S* = Skeletal, Epithelial. NR = Not Reported .....130

---

# Acknowledgements

---

For me, it is impossible to distinguish between the valuable lessons I have learned here - from the science to the day-to-day 'university of life'. First and foremost, I'd like to thank my supervisor Annalisa Pastore for giving me the opportunity to do this project and encouraging me to work independently throughout my PhD. Her supervision, which has been a great motivating force of the project, has most definitely improved the quality of this work and will, no doubt, enhance the work I will do in the future. Much gratitude to our collaborators Georgine Faulkner and Alberto Pallavicini for providing us with the ZASP constructs, Remo Guerrini for the peptide synthesis, and Steve Martin for the fluorometric titrations; without whom aspects of the work would not have been possible.

*Muchas gracias* to Andres Ramos who has been an excellent office companion; giving me exceptional guidance when required, from thesis-corrections to deciding on peak assignments. *Merci beaucoup* to Andrew Atkinson for being a great co-supervisor, who taught me the NMR ropes at the start of the project. *Danke schöne* to Catherine Joseph for teaching me the 'ins-and-outs' of the wet-lab, and also for her saint-like patience and good humour. *Molte grazie* to Laura Masino for her unswerving scientific advice and naïve *double entendres*. *Dankon* to Jake Grimmer and Jürgen Schmidt and for sorting out the computer network. Спасибо to John McCormick for his general good humour and reliability, for keeping lunchtime regular, and his amazing crosswords skills. *Go raibh maith agat* to the boys at the NMR centre: Geoff Kelly, Fred Muskett, and Tom Frenkiel, for keeping it a smooth operation. Additional kudos to Fred and Geoff for helping me with the finer details of NMR data analysis and teaching me various aspects of structure calculation. 谢谢 to Jim Feeney and Berry Birdsall for their steadfast judgement when called upon. Thanks also go to all of the other members of the lab, past and present, who have helped and supported me throughout my time at the NIMR, and have generally made the place a fun and stimulating place to be.

I'd like to thank my fellow PhD buddies as follows: Carl Shiu for cooking my dinner, teaching me Wing Chun, and getting me out of hairy situations, Eugenio Gutierrez for the tequila and fake Burberry scarves, Steen for humouring my obsession for the Govenators 1980's film quotes and reciprocating with '*The Comic Book Store Guy*', and Julia Pendred and Katie Copeland and for the NIMROD bar reality check. Last but not least, I'd like to say thanks to Mum & Dad, Li-Leng and Lisa for their love and support.

# Abbreviations

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
6HIS	Six histidine (tag)
Act-EF1234	$\alpha$ -actinin-2 EF-hands one to four
Act-EF34	$\alpha$ -actinin-2 EF-hands three to four
ADP	Adenosine Diphosphate
ALP	Actinin-associated LIM protein
ATP	Adenosine Triphosphate
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CANDID	Combined automated NOE assignment and structure determination
CH	Calponin Homology domain
CLP-36	Carboxyl terminal LIM domain protein
COSY	Correlated spectroscopy
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
Da	Dalton
DHR	Disks-large homology region
Dlg	Disks-large protein
DYANA	DYnamics Algorithm for NMR Applications
<i>E. Coli</i>	<i>Escherichia Coli</i>
EF hand	Derived from $\text{Ca}^{2+}$ binding in sheet 5 and 6 in calmodulin
EM	Electron Microscopy
ENH	Enigma Homology Protein
FPLC	Fast Protein Liquid Chromatography
FSSP	Fold classification based on Structure Structure alignment of Proteins
GLGF	Domain with 'Glycine-Leucine-Glycine-Phenylalanine' motif
GLOMSA	Global Method for Stereo-specific Assignments
GST	Glutathione-S-Transferase
<i>h</i> CLIM1	Human carboxyl terminal LIM protein
HMM	Heavy meromyosin
HSQC	Hetronuclear single quantum coherence
IPTG	Isopropyl $\beta$ -D-thiogalactopyranoside
Kd	Dissociation constant
kDa	Kilo-Dalton
LB	Luria Bertani
LIM	Derives from <i>lin-11</i> , <i>isl 1</i> , and <i>mec-3</i> .
LMM	Light meromyosin
LMP	LIM Mineralization protein.
MAGUK	Membrane associated guanylate kinase
MDa	Mega Dalton
$\mu\text{m}$	Micro-metre
$\mu\text{M}$	Micro-molar
mM	Milli-molar
Ni-NTA	Nickel-nitrilotriacetic acid
nM	Nano-molar
NMR	Nuclear Magnetic Resonance

NNOS	Neuronal nitric oxide synthetase
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser Spectroscopy
PDZ	Derives from PSD-95, Dlg, and ZO-1
pH	(The) Power of Hydrogen
Pi	Inorganic Phosphate
PKC	Protein Kinase C
pM	Pico-molar
PSD-95	Post-synaptic protein of density 95 kDa
PTB	Phosphotyrosine-binding domains.
PTP-BL	Protein tyrosine phosphate-bl
REDAC	Redundant Dihedral Angle Restraints
RIL	Reversion induced LIM
RMS	Root Mean Square
RMSD	Root Mean Square Deviation
S1	First subfragment of myosin
S2	Second subfragment of myosin
SH2/3	Src homology 2/3 domain
SMART	Simple modular architecture research tool
TALOS	Torsion Angle Likelihood Obtained from Shift and sequence similarity
TCEP	Tris(carboxyethyl)phosphine
TEV	Tobacco Etch Virus
TEM	Transmission Electron Microscopy
TOCSY	Total Correlation Spectroscopy
tZR7	Minimum binding peptide from the seventh Z-repeat from titin
ZASP	Z-band alternatively spliced PDZ motif containing protein
ZO-1	Zonula occludens-1
ZR7	Seventh Z-repeat from titin



# Chapter 1

---

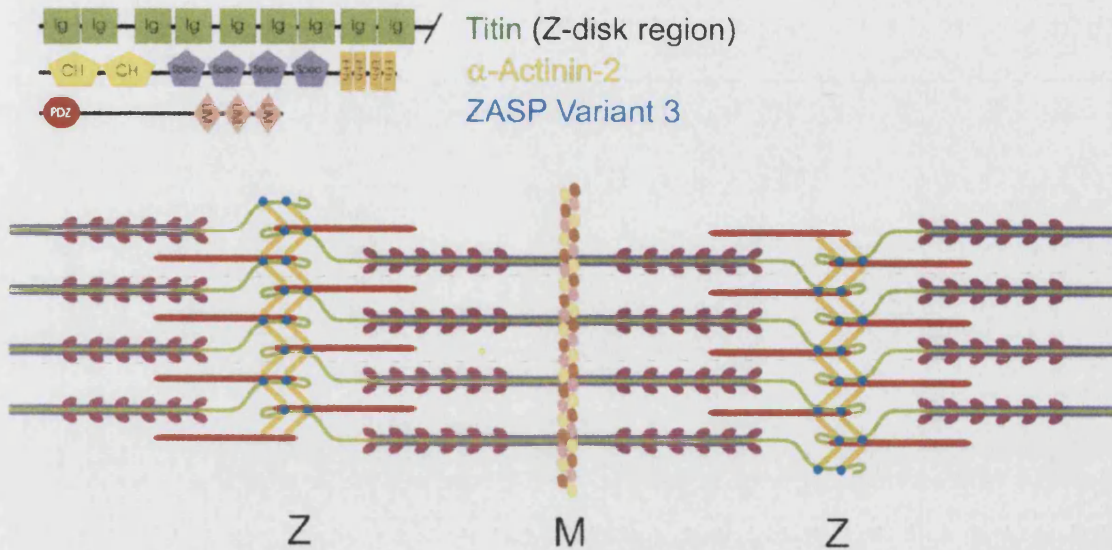
## Introduction

---

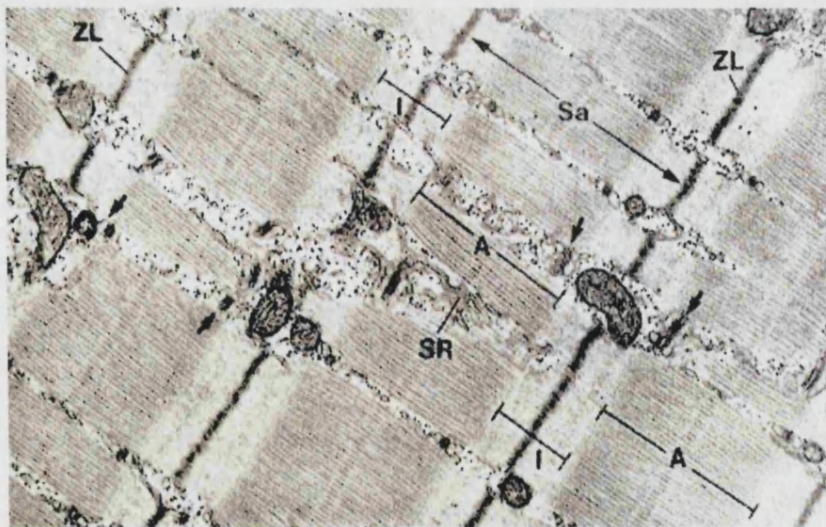
### 1.1 Muscle

Movement plays a vital role for all living systems, whether it is the transport of single molecules, to the entire movement of an organism. In mammalian systems, skeletal muscle contraction is essential for walking, running, swimming, and flying, or even involuntary movements such as the beating heart (cardiac) and peristalsis (smooth muscle). Striated muscle (i.e. cardiac and skeletal muscle) is composed from many muscle fibres, which are huge single cells formed during development by the fusion of many separate cells. These form cylindrical structures known as myofibrils, which are the contractile elements of the muscle cell. Myofibrils in turn are made up of a microscopic contractile element known as the sarcomere, a complex network of proteins and membranes (Figure 1.1). Depending on the organism and tissue type, the sarcomere is approximately 2  $\mu\text{m}$  in length and can shorten to 70% of its original length during contraction.

Transmission Electron Microscopy (TEM) gives an overview of muscle ultrastructure (Figure 1.2), giving rise to the description of the sarcomere in terms of thick and thin filaments, I and A bands, and Z and M lines (Huxley and Niedergerke, 1954; Huxley and Hanson, 1954). The thin filament is primarily composed of actin, tethered to the Z-disk. The thin filaments then interdigitate with the thick filaments - the main constituent being myosin. The thick filament terminates at the M-line. Studies by electron microscopy of the A band cross-section reveal that six thin filaments encircle the thick filament in a hexagonal arrangement. This also shows that cross-bridges, formed by the globular head of myosin, extend to interact with actin.



**Figure 1.1:** a) Schematic representation of the sarcomere. The position of the Z-disk and M-line is indicated. A single molecule titin filament (indicated in green) connects the Z-disk to the M-line. In the Z-disk  $\alpha$ -actinin-2 (in yellow) forms transversal connections between actin filaments (in red) and is bound to ZASP (in blue). Myosin and F-actin are coloured purple and red respectively. b) Domain architecture of the titin region localised in the Z-disk,  $\alpha$ -actinin-2 and ZASP (variant 3).

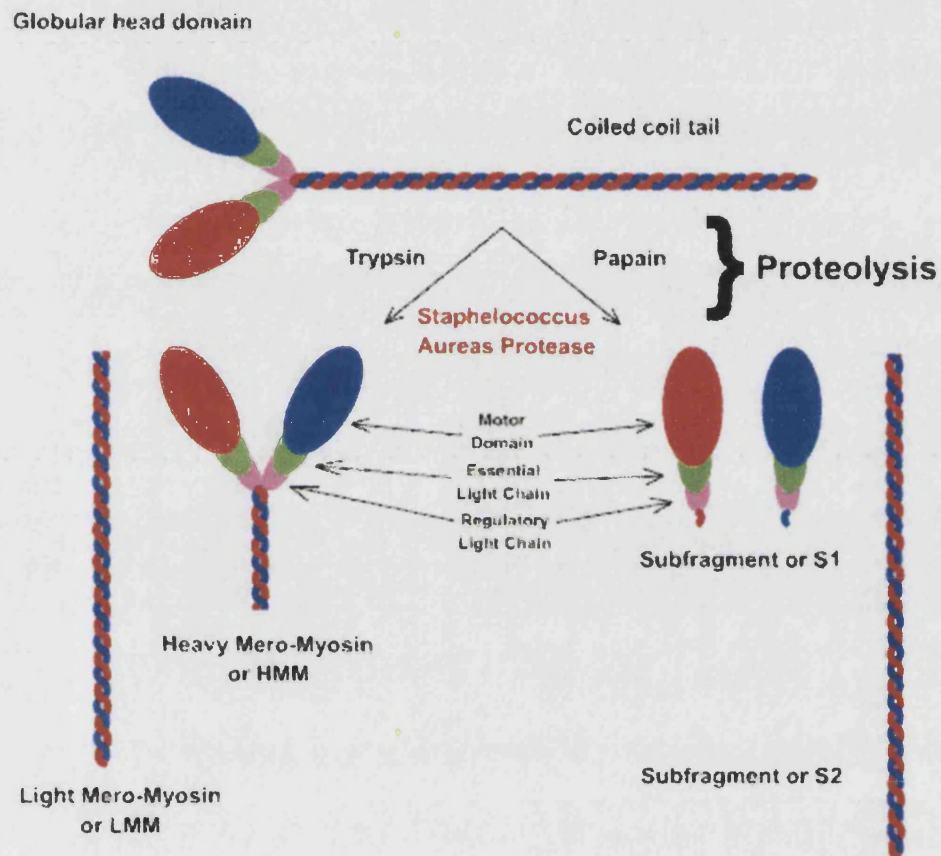


**Figure 1.2:** Transmission Electron Microscope picture of striated mammalian muscle (<http://www.udel.edu/Biology/Wags/>), labelled to indicate the position of the sarcomere, I and A bands, Z and M lines, and the sarcoplasmic reticulum.

Globular actin (G-actin) polymerises to form fibres (F-actin), with a double helical ultrastructure in the thin filament. The helix has thirteen molecules of actin per six turns. F-actin has a structurally polarised orientation, and its monomers are in a constant state of association and dissociation, with the fast growing positive (+ve) 'barbed' end in the Z-disk, and the slow growing negative (-ve) 'pointed' end pointing towards the M-line. The protein CapZ is responsible for capping the barbed end, acting to prevent the addition and loss of actin, whilst tropomodulin is able to carry out the same function at the pointed end. The thin filament also contains two molecules of nebulin, which is thought to act as a 'molecular ruler', regulating the length of the thin-filament. Also associated with the thin filaments are tropomyosin and the troponin complex, which regulates the binding of myosin to actin in a calcium dependent manner.

Myosin forms a bundle of approximately 200-400 molecules per thick filament, the number of which is regulated by C-protein. The length of the thick filament is regulated by titin (see 1.3). There is also a polarity of the thick filaments, with the myosin heads orientating in one direction to face the Z-disk. Each myosin molecule is a complex of six polypeptide chains (Figure 1.3): two 'heavy chains' and two pairs of different light chain – the essential and the regulatory light chains (ELC and RLC). The N-terminal half of the one of the heavy chains form a globular head, to which ELC and RLC associate, while the C-terminal half forms a fibrous alpha-helical tail, which in turn forms a coiled-coil with the second heavy chain (rod-like domain). In 1953, Andrew Szent-Györgi used limited trypsin digestion on myosin to produce 'light' (LMM) and 'heavy' meromyosin (HMM), which contained the rod-like domain and the two globular heads of myosin (Szent-Gyorgyi, 1953) with a section of rod-like domain, respectively. HMM could be further proteolytically cleaved using papain to just the globular head of myosin, termed subfragment 1 (S1), and the remaining section of rod-like domain (S2) (Margossian and Lowey, 1973a; Margossian and Lowey, 1973b).

## The Myosin II molecule

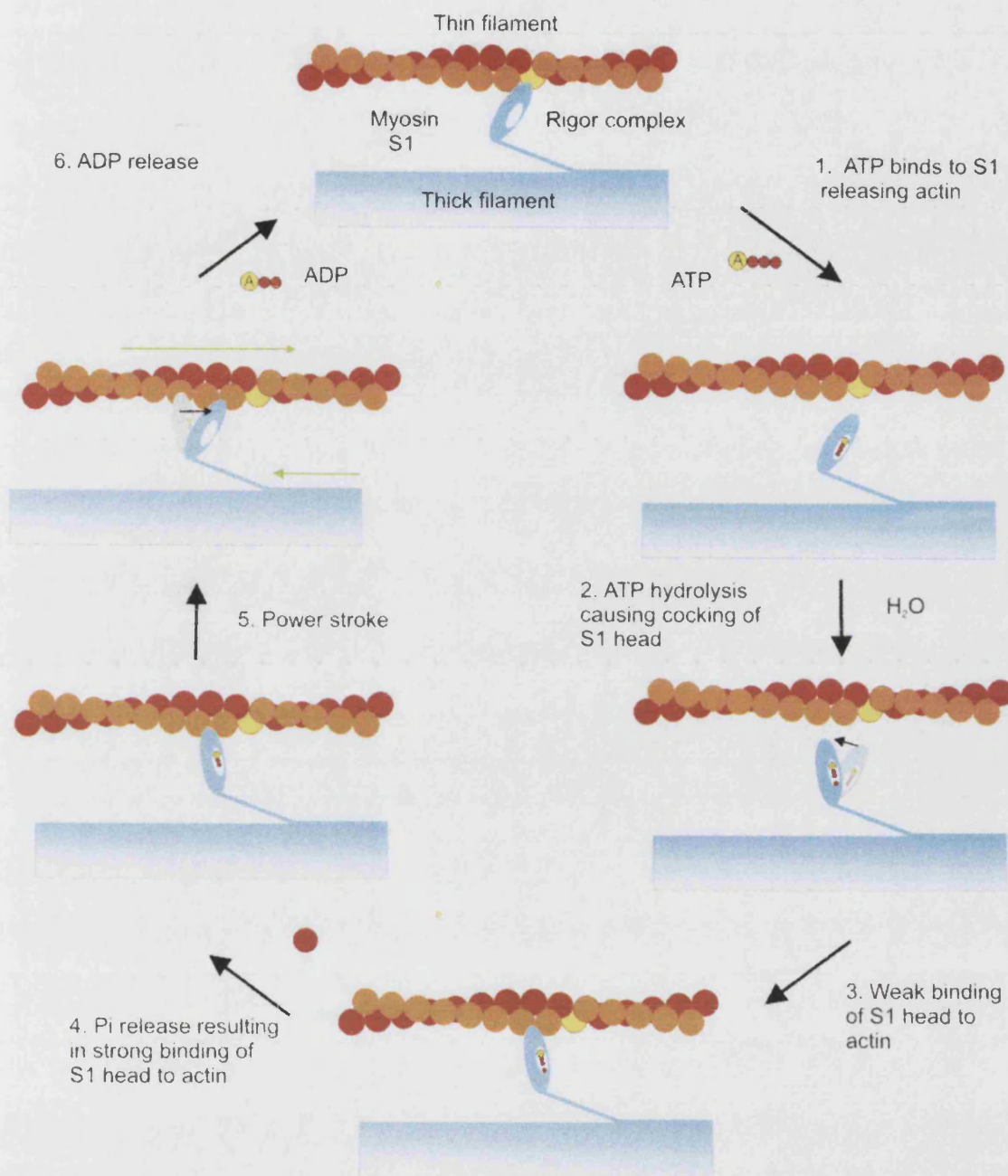


**Figure 1.3:** The myosin II molecule, indicating the production of different subfragments by two types proteolytic cleavage: trypsin cleavage resulting in HMM and LMM, and papain cleavage, producing S1 and S2.

The original sliding filament model (or swinging cross-bridge hypothesis) (Huxley and Hanson, 1954) states that muscle contraction occurs when the myosin filaments slide past the actin filaments, with no shortening of either filament, thus shortening the entire sarcomere. This model for muscle contraction developed to include the role of ATP hydrolysis in 'The Lymn-Taylor cycle' (Lymn and Taylor, 1971). Further refinement of the model was made by the observations in structural and spectroscopic data (Reviewed in (Cooke, 1986)), which defined that the movement between actin and myosin arose from the C-terminal end of S1 acting as a lever ('The Lever-Arm Hypothesis'). (For a review on the mechanism of muscle contraction see (Holmes and Geeves, 2000))

Muscle contraction is a multi-step cycle involving the globular heads of myosin, actin, and adenosine triphosphate (ATP) hydrolysis. The myosin head undergoes a series of ordered conformational changes depending on the presence of ATP, adenosine diphosphate ADP, and inorganic phosphate (Pi) (Figure 1.4). The start of the cycle begins with myosin bound to actin in the rigor conformation. The binding of ATP to myosin reduces the affinity of myosin for actin, and the two proteins separate. Myosin then hydrolyses ATP causing it to adopt a cocked conformational change, priming the 'power stroke'. The weak association of the myosin-ADP-Pi complex with actin results in the concomitant release of Pi, increasing the affinity of myosin for actin. This is immediately followed by a large conformational change in myosin (the power stroke), and ADP is simultaneously released, hence, the initial rigor position is established. This series of reactions results in the force-generation required for filament movement, therefore for the sarcomere to shorten, and ultimately for muscle to contract.





**Figure 1.4:** The lever arm hypothesis of muscle contraction, adapted from (Rayment et al., 1993a). The diagram shows the process between myosin, actin, and ATP. The cycle begins from actin and myosin bound together in the 'rigor complex', after which ATP hydrolysis leads to a series of step-wise conformational changes. The sequence of events ends with the power-stroke, hence, re-starting the cycle.

A major advance in understanding these concerted changes in protein structure came with the structural elucidation of the components involved. The atomic model of the actin filament (Lorenz et al., 1993) was derived from the crystal structure of g-actin solved in complex with DNase I (Kabsch et al., 1990), used in conjunction with x-ray fibre diagrams (Holmes et al., 1990). The structure of rigor state scallop myosin S1 fragment was determined (Rayment et al., 1993b), leading to the atomic model of the rigor conformation of ‘decorated actin’ (i.e. actin fibres with many bound S1) to be established (Rayment et al., 1993a). Different conformational states of myosin were also critical for confirming details in the hypothesis of muscle contraction, such as myosin in the ‘cocked’ position (pre-power stroke – myosin bound to ADP) (Dominguez et al., 1998).

Although interactions between actin and myosin have been well-characterised, structural information of other proteins with a structural and regulatory role within the sarcomere have yet to be determined; this would give us a richer understanding about muscle contraction. An essential component of the sarcomere is a multi-protein complex known as the Z-disk, whose functions are to keep the structure of the sarcomere in register and to transmit tension during muscle contraction. The major component of the Z-disk is the muscle specific  $\alpha$ -actinin-2, a multi-domain protein that tethers major proteins such as actin and titin within the sarcomere (Faulkner et al., 1999; Ohtsuka et al., 1997b). In the next sections I will review what is known about the three structural components of the sarcomere that are the object of this thesis:  $\alpha$ -actinin, titin, and ZASP.

## **1.2 $\alpha$ -Actinin-2**

$\alpha$ -Actinin belongs to the spectrin super-family, defined by the presence of a specific number of spectrin repeats (first identified in spectrin) and their ability to bind actin, of which  $\alpha$ -spectrin,  $\beta$ -spectrin,  $\beta_{\text{heavy}}$  spectrin, and dystrophin are members (Baron et al., 1987; Beggs et al., 1992; Dubreuil, 1991). There are four variants of  $\alpha$ -actinin which are expressed depending on the cell and muscle type:  $\alpha$ -actinin 1 and 4 are non-muscle proteins involved in cytoskeletal interactions, while  $\alpha$ -actinin 2 and 3 are muscle specific.  $\alpha$ -Actinin-2 and 3 are both present in striated muscle, but only  $\alpha$ -actinin-2 is expressed in cardiac muscle.  $\alpha$ -Actinin-3 may be functionally redundant, as its deficiency carries no obvious phenotype;  $\alpha$ -actinin-2 may compensate for its function (Mills et al., 2001). However, the presence of different variants of  $\alpha$ -actinin-3 may be used as an indicator of elite athletic ability (Yang et al., 2003) (Figure 1.5). Chromosomal mapping of  $\alpha$ -actinin-2 shows that the gene's locus is 1q43 (Beggs et al., 1992; Tiso et al., 1999).

$\alpha$ -Actinin-2's exists as a rod shaped anti-parallel homodimer (Blanchard et al., 1989), comprising of two calponin-homology (CH) domains, four spectrin-like repeats and a calmodulin-like domain (Figure 1.6).  $\alpha$ -Actinin-2's main function appears to be to act as a scaffold for the assembly of the Z-disk. It is able to do this by forming crucial interactions between components of the sarcomere at locations along its sequence (Luther, 2000a). The calmodulin-like domain contains four potential EF-hands, which have divergently evolved to not require  $\text{Ca}^{2+}$  to bind its target protein (Nakayama and Kretsinger, 1994).

Currently, the structure of full-length  $\alpha$ -actinin-2 has not been elucidated. However, sections of  $\alpha$ -actinin solved using NMR or x-ray crystallography shows that it is mostly composed of  $\alpha$ -helical domains. These structures include two of the central spectrin-like repeats (R3-R4) (Djinovic-Carugo et al., 1999), subsequently super-ceded by the structure of all four spectrin-like repeats (R1-R2-R3-R4) (Ylanne et al., 2001), and the structure of  $\alpha$ -actinin-2's last two EF-hands in complex with titin's seventh Z-repeat (Atkinson et al., 2001) (Figure 1.7). Spectrin-like motifs are 100-120 residue domains forming an  $\alpha$ -helical triple coiled-coil elongated structure. Each spectrin motif is joined together by a helical segment between the C-terminal helix of the first spectrin domain, and the N-terminal helix of the preceding domain. Both structures of the spectrin domains are anti-parallel in-register homodimers, with a self-association  $K_d$  of 2  $\mu\text{M}$  and 10 pm respectively (Djinovic-Carugo et al., 1999; Flood et al., 1995; Flood et al., 1997).



Both structures show a similar degree of surface burial of 9.2% and 11% of the accessible surface of the monomer, for the two and four spectrin structures respectively. These structures act as models for the dimerisation of  $\alpha$ -actinin-2 *in vivo*; the R1-R4 structure shows a rigid-rod configuration, possibly caused by an end-to-end 90 ° twist.

$\alpha$ -Actinin-2's last two EF-hands (Act-EF34) in complex with titin's seventh Z-repeat (ZR7) was solved by NMR (Atkinson et al., 2001). This complex shows an unconventional binding mechanism, compared to calmodulin, as these EF-hands do not require the presence of  $\text{Ca}^{2+}$  for target binding. Act-EF34 has four  $\alpha$ -helices – two helix-loop-helix motifs joined by a flexible linker. These EF-hands saddle titin-Zr7 in a 'semi-open' conformation, with an EF-hand either side of the target; titin-Zr7 is made to adopt an  $\alpha$ -helical configuration from a previously unstructured conformation (Atkinson et al., 2000a; Joseph et al., 2001). Titin's Z-repeats have been shown to be bound by Act-EF34 at nanomolar affinities with ZR7 being the tightest at 100-250 nm (Joseph et al., 2001).

Although there is no structure of either of  $\alpha$ -actinin's calponin homology domains, the structure of a CH domain derived from  $\beta$ -spectrin (Carugo et al., 1997) has a 57% sequence identity with  $\alpha$ -actinin-2's second CH domain (CH2), so should therefore adopt a similar fold.  $\alpha$ -Actinin-2's first CH (CH1) domain only has a 15% sequence identity with the  $\beta$ -spectrin's CH domain, so may have significant structural differences. The calponin-homology domains are thought to work in tandem to bind actin, as the affinity for both domains with actin is tighter than that of a single CH domain (Way et al., 1992). Actin is thought to bind between the last helix of CH1 and the first helix of CH2. The CH domain is a globular  $\alpha$ -helical protein, which has a unique fold. These structures of  $\alpha$ -actinin-2 mean that only the first CH1 domain and the first two EF-hands of the calmodulin like domain are yet to be structurally described.

At time of writing, there are at least eight proteins found to interact directly with  $\alpha$ -actinin-2 (Figure 1.6), at different regions (Faulkner et al., 2001). Among these, titin and ZASP has been identified to interact with  $\alpha$ -actinin-2 (Faulkner et al., 1999; Ohtsuka et al., 1997a).

# Running in the genes

A SINGLE gene may mark a person out as an elite sprinter or marathon runner, it was claimed yesterday.

The gene comes in two forms that gear an athlete up either for short bursts of speed or long endurance.

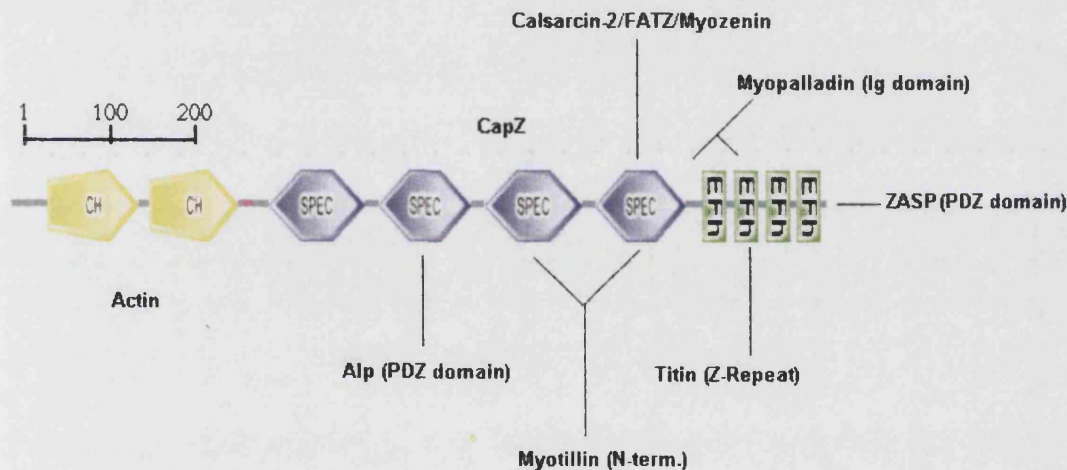
The discovery raises the prospect of identifying potential top athletes at birth, using a DNA test.

Australian researchers, who

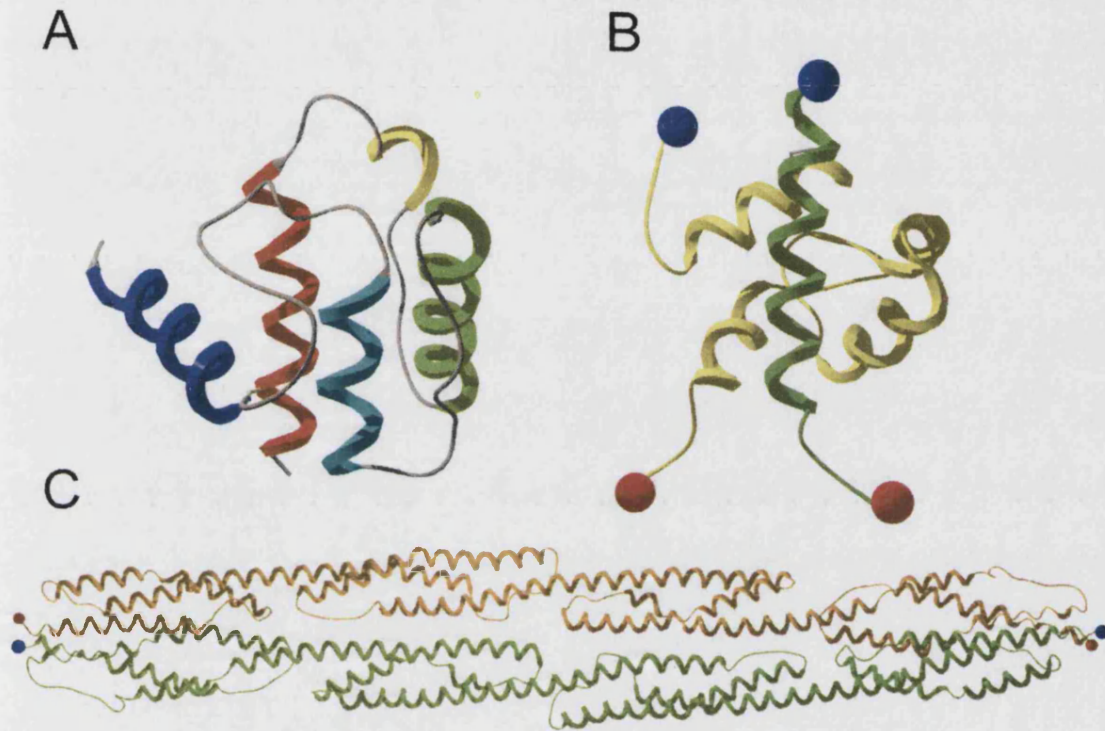
examined the genetic profiles of more than 300 athletes, identified two versions of a gene called alpha-actinin-3, or ACTN3, *New Scientist* magazine reported.

The exact role of actinin-3 is unclear, but researchers believe it may "confer a greater capacity for the absorption or transmission of force during rapid, forceful contraction" of the muscles.

Figure 1.5: The Telegraph 28 August 2003



**Figure 1.6:** Diagram of the domain organisation of  $\alpha$ -actinin-2 and its binding partners. The arrows indicate the region of  $\alpha$ -actinin-2 that the binding partners have been found to interact with - CapZ has only been localised to all of the spectrin-like motifs. The program SMART (Schultz et al., 2000) was used to predict the different domains: calponin-homology domain (CH); spectrin-like domains (SPEC), and EF-hand domains (EFh). Interacting proteins include actin (Panasenkov and Gusev, 2001), titin/connectin (Ohtsuka et al., 1997b) ZASP (Faulkner et al., 1999), ALP (Xia et al., 1997), CapZ (Papa et al., 1999), Myotilin (Salmikangas et al., 1999), Myopalladin (Bang et al., 2001), FATZ/Calsarcin-2/Myozenin (Faulkner et al., 2000; Frey et al., 2000; Takada et al., 2001)



**Figure 1.7:** Solved structures from regions of  $\alpha$ -actinin-2; a) the second calponin homology domain of  $\beta$ -spectrin (1AA2), 57% sequence identity with  $\alpha$ -actinin-2's CH2 domain, coloured by secondary structure succession (from blue to red); b) the last two EF-hands in complex with titin's seventh Z-repeat (1H8B), coloured yellow and green respectively; c) four spectrin-like repeats in an anti-parallel homodimer (1HCI), coloured green and orange to demarcate each monomer. B and C have blue and red spheres to distinguish the N and C termini respectively.

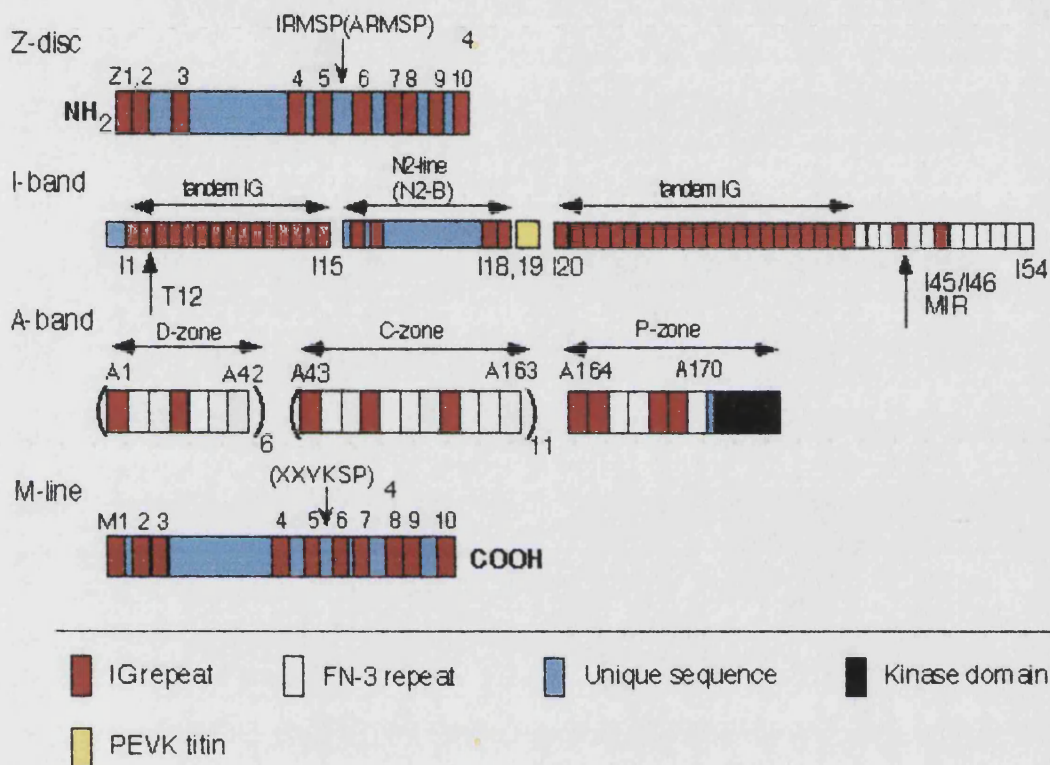
### **1.3 Titin**

Titin (also known as connectin) is a giant elastic filamentous protein (for review, see (Maruyama, 1997) and (Trinick, 1996)). As well as being the largest protein described to date (3 MDa), it is the third most abundant protein in vertebrate striated muscle (Maruyama et al., 1981). Single titin molecules are  $\sim 1 \mu\text{M}$  in length and span half of the sarcomere, with its N-terminus bound in the Z-disk and C-terminus localising in the M-line (Figure 1.1), unambiguously orientating the titin molecule. Its main functions are to help the sarcomere keep alignment during muscle contraction (Maruyama et al., 1981; Trinick et al., 1984), and also to act as a molecular ruler to regulate the assembly and length of the thick filament (Whiting et al., 1989). Therefore, titin is an essential component of the sarcomere's scaffold. Its string-like appearance has been noted by EM, also revealing that it appears to have a globular head consisting of bound M-line proteins.

To achieve its functions, various sections of titin appear to have specific functions: the I band is believed to act as an elastic connection between the thick filaments and the Z-disk; in the A band it is associated with myosin and other constituents of the thick filament; and in the M-line it forms an integral part of an extensive protein meshwork. Sequencing human cardiac titin (Labeit and Kolmerer, 1995) revealed that it is a modular protein, composed of 26,926 residues, of which 90% is made up from  $\sim 100$  (fibronectin type III domain-like (Fn3)) and  $\sim 100$  (immunoglobulin domain-like (Ig)) repeated sequences (Figure 1.8). These repeats make up a complex modular architecture.

Although the structure of full-length titin has not been solved, different sections of titin have been solved independently, including a Fn3-like module by NMR (Goll et al., 1998), several Ig-like repeats by NMR (Improta et al., 1996; Pfuhl et al., 1997; Pfuhl and Pastore, 1995) and x-ray crystallography (Mayans et al., 2001), and the serine kinase domain by x-ray (Mayans et al., 1998). Only  $\alpha$ -actinin-2's last 73 residues are required to bind an N-terminal region of titin referred to as the 'Z-repeats' (Ohtsuka et al., 1997a; Ohtsuka et al., 1997b), between Ig repeats 3 and 4. Z-repeats are 45 residue motifs that vary depending on muscle type, which have been shown to be responsible for determining Z-disk thickness (Gautel et al., 1996). The structure of  $\alpha$ -actinin-2's last two EF-hands (Act-EF34) has been solved by NMR in complex with a region of titin called Z-repeat 7 (ZR7) (Atkinson et al., 2001) – here the EF-hands bind to the target in a 'semi-open' conformation, which may be a general solution for calcium-independent recognition of the target sequence.





**Figure 1.8:** Predicted domain representation of the primary sequence of the cardiac titin filament cDNA (From (Labeit and Kolmerer, 1995)). The immunoglobulin-like domains are represented in red, while the fibronectin type 3-like domains are in white. Irregular areas of sequences are shown in blue, PEVK region in yellow, and titin kinase in black. Titin domains have been labelled according to the main sarcomeric region (Z-line, A-band, I-band, and M-line). Note that the Z-repeats are found between Ig domains 3 and 4 in the Z-disk region of titin.

## **1.4 ZASP**

A protein named ZASP, identified relatively recently, is only found in the Z-disk of skeletal and cardiac muscles (Faulkner et al., 1999). ZASP belongs to the enigma family of cytoskeletal proteins, defined by the presence of a combination of PDZ and LIM domains. The enigma family includes enigma/LMP (Guy et al., 1999), ENH (Nakagawa et al., 2000), ALP (Xia et al., 1997), RIL (Cuppen et al., 1998), and elfin (Kotaka et al., 2001). Most enigma proteins appear to have a relation to cytoskeletal function: ZASP, ALP, and ENH have been found to interact with  $\alpha$ -actinin-2, and whilst enigma has been found to bind  $\beta$ -tropomyosin. Depending on the isoform, ZASP and cypher comprise either one or two known sequence motifs: a N-terminal PDZ alone or in conjunction with three carboxy-terminal LIM domains (Table 1.1).

The interaction between ZASP and  $\alpha$ -actinin-2 has been mapped to ZASP's PDZ domain (ZASP-PDZ) and to the C-terminus of  $\alpha$ -actinin-2 (Faulkner et al., 1999; Zhou et al., 2001). PDZs are protein modules of 80-120 residues involved in targeting and clustering of membrane proteins or directing cellular proteins to multi-protein complexes, named after the first three proteins identified to contain the motif (post-synaptic protein of density 95 kDa, disks-large, and zonula occludens-1) (for a review see (Hung and Sheng, 2002)). Three types of interaction involving PDZ motifs have been described in detail (Hung and Sheng, 2002). The first two types, class 1 and 2, involve binding of carboxy-terminal protein sequences through either an  $-(S/T)-x-(V/I/L)$  or a  $-(F/Y)-x-(F/V/A)$  sequence motif. The third type of recognition involves internal  $-(S/T)-x-V-$  sequences motifs. Over fifty PDZ domains have been solved by either NMR or X-ray crystallography (from the Protein Data Bank – <http://www.rcsb.org>). A structural alignment of these PDZ domains calculated by DALI (Holm and Sander, 1996a) show that approximately twenty of these structures can be represented by three solved structures: DlgA (1pdr), CASK (1kwa), and syntrophin (1qav) (Doyle et al., 1996; Hillier et al., 1999; Morais Cabral et al., 1996). This suggests that most PDZ domains have a fairly consistent architecture (N.B. not all of the PDZ domains have yet been characterised by the DALI server, hence more may be in these three groups). The PDZ domain is a  $\beta$ -sandwich, composed of six strands, flanked at either side by two  $\alpha$ -helices. The second  $\beta$ -strand and the second  $\alpha$ -helix form a groove, which has been observed to be the binding groove for most of the PDZ domains solved in complex with its target protein (PDB Accession Codes: 1B8Q (Tochio et al., 2000), 1BE9 (Doyle et al., 1996), 1MFG (Birrane et al., 2003), 1MFL (Birrane et al., 2003), 1N7T (Skelton et al.,

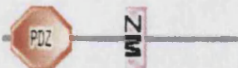


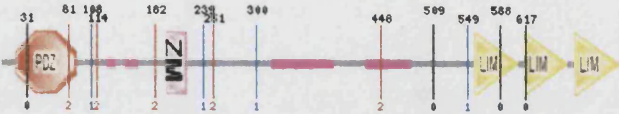



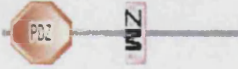


2003), 1OBX (Kang et al., 2003), 1OBY (Kang et al., 2003), 1QAV (Hillier et al., 1999)). These structures of PDZ complexes have all been found solved in the presence of a peptide deriving from their target protein, with the exception of 1QAV; this is a heterodimer between two PDZ domains derived from neuronal nitric oxide synthase (nNOS) and syntrophin (Hillier et al., 1999). In general, PDZ domains show no major change in tertiary structure upon complex formation, for example, as observed between the third PDZ domain from PSD-95 in complex with a C-terminal peptide derived from Cript (Doyle et al., 1996).

The LIM domains are a cysteine rich motif of approximately 50-60 residues, which derived its name from *Lin-11*, *Isl 1*, and *Mec-3* from *Caenorhabditis elegans*, rat, and *C. elegans* respectively (for a review, see (Bach, 2000)). LIM proteins have been found to either be involved in cell line determination or cytoskeletal structure and organogenesis. Cypher's LIM domains have been found to bind various isoforms of Protein Kinase C (PKC), suggesting a putative function of coupling  $\alpha$ -actinin-2 to PKC function. The non-LIM containing isoforms of cypher may function to inhibit LIM containing cypher's activity.

Mutations in ZASP have been linked to dilated cardiomyopathy (DCM), a primary cardiac muscle disease characterised by ventricular dilatation and systolic dysfunction (Dec and Fuster, 1994; Goodwin, 1982; Johnson and Palacios, 1982; Sugrue et al., 1992). DCM is a major cause of morbidity and mortality, affecting 40 people in every 100,000 of the population (Codd et al., 1989). The underlying cause of DCM is heterogeneous, with approximately 30% of genetic origin (Grunig et al., 1998; Keeling et al., 1995; Michels et al., 1992). Autosomal dominant inheritance of DCM is the most common (Towbin, 1999). Multiple genetic loci have been identified for autosomal dominant DCM, while mutations have been identified in actin, desmin, lamin A/C,  $\delta$ -sarcoglycan,  $\beta$ -sarcoglycan,  $\beta$ -myosin heavy chain, cardiac troponin T,  $\alpha$ -tropomyosin, titin,  $\alpha$ -dystrobrevin, and vinculin (Arbustini et al., 2002; Brodsky et al., 2000; Gerull et al., 2002; Ichida et al., 2001; Towbin, 1998; Towbin and Bowles, 2002). These genes contribute to structural support of the myocyte and link the sarcomere to the sarcolemma and extracellular matrix (ECM), as well as to the nuclear lamina (Towbin, 1998). From samples of patients with DCM (Arimura et al., 2003; Vatta et al., 2003), single polymorphisms were identified within in ZASP which were specific to ~1-4% of DCM cases. These mutations include single mutations localised to ZASP4 (T213I, S196L, and I352M), outside the PDZ domain in ZASP1-3 (N115K), and within the 3<sup>rd</sup> LIM domain in ZASP2-4 (D626N). So far the only characterisation of these mutations has been for

D626N found to have a higher affinity for PKC, which may disrupt downstream PKC pathways important for the correct cardiac muscle function (Johnson et al., 1996; Liu et al., 1999). Cypher (the murine ZASP orthologue (aka oracle (Passier et al., 2000)) knockout mice display a severe form of congenital myopathy and die from failure of multiple striated muscles (Zhou et al., 2001), again suggesting that ZASP is an important muscle protein required for maintaining the cytoskeletal ultrastructure during contraction. Structural information about ZASP would be a great aid to understanding its functional relationship with  $\alpha$ -actinin-2, and this may in turn be useful for identifying why mutations in ZASP lead to cases of DCM.



Protein EMBL Acc. (Alias)	SMART prediction	Number of Residues	MW (KDa)
ZASP1 AJ133766		283	30.8
ZASP2 AJ133768		470	50.7
ZASP3 AJ133767		617	66.5
ZASP4 BAA31588 (KIAA0613)		734	77.7
Cypher2c AAO26187		327	35.1
Cypher3c AAO26188 (Oracle2)		661	70.8
Cypher1c AAD42950 (Oracle1)		723	76.4
Cypher2s AAD42951		288	31.4
Cypher3s AAO26190		622	67.1
Cypher1s AAO26189		679	72.3

**Table 1.1:** SMART domain predictions and information regarding their residue number and molecular weight of the different splice variants of ZASP, cypher (aka oracle). Red hexagons represent PDZ domains, yellow triangles represent LIM domains, and purple and grey sections represent complex and non-domain containing regions respectively. ZASP1, cypher2c and cypher2s do not contain the LIM domains. Note that cypher's isoforms have been characterised in terms of expression in skeletal (s) and cardiac muscle (c), hence the s/c suffix.

## **1.5 Nuclear Magnetic Resonance**

There are currently two main methods routinely used for structure determination at atomic resolution: x-ray crystallography and Nuclear Magnetic Resonance (NMR). NMR is an indirect technique, meaning that structures are modelled using data deriving from several variables (such as chemical composition and structure), whereas structures solved by x-ray crystallography originate directly from the electron density of the structure. X-ray crystallography has its limitations such as the ability to crystallise the molecule and artefacts due to crystal packing. However, NMR is also not without its limitations; NMR structure calculation is limited by the size of the macromolecule. Also, the duration of the whole procedure of NMR structure calculation is usually lengthy. Despite these limitations, NMR structures make up a significant percentage of the Protein Data Bank (15% on Jan 2004).

The input data for NMR protein structure calculation derive from geometric measurements such as distances, angles, and relative orientations between the atoms in the molecule. Data from NMR experiments arise from a pattern of peaks related to the atoms chemical environment (i.e. the chemical shift), so before measurement of structural restraints, the peaks need to be assigned to the atoms. This is achieved by acquiring spectra that correlate atoms depending on how they are covalently bonded to each other (through-bond), such as COSY, TOCSY, or 'back-bone' experiments. As mentioned previously, the size of the studied macromolecule limits the NMR structure determination, this is mainly because the number peaks increase with the protein's size leading to peak overlap, which is difficult to deconvolute. Although running two-dimensional NMR experiments improve the overlap problem, a common strategy is to isotopically label the protein with  $^{15}\text{N}$  and  $^{13}\text{C}$ . It is then possible to run spectra to separate the proton signals on whether they are connected to spin active  $^{15}\text{N}$  or  $^{13}\text{C}$ .

After the atoms of the molecule has been assigned to their chemical shift, the through-space Nuclear Overhauser Effect spectroscopy (NOESY) data can be interpreted to measure distances between the atoms; the presence of a NOE signal between atoms can be detected up to a distance of 6 Å. Other experiments can be run to obtain additional distance information, such as hydrogen bonds (Cordier and Grzesiek, 1999). The dihedral angles can be measured using NMR three-bond J-coupling experiments (e.g. the HNHA for  $\phi$  angles), or by using TALOS (Cornilescu et al., 1999), which operates by comparing the chemical shift assignment with a database of known protein assignments with known secondary structure elements. Also, the 'dipolar-coupling' can

be measured, which is a measurement of the angle of a bond vector (e.g. the amide bond) with respect to the external magnetic field; this is a 'long-range' restraint that can orientate structural features of the protein, or align interacting molecules.

All of these restraints are then inputted into a structure calculation program, of which two of the most commonly used are CYANA (Guntert et al., 1997) and XPLOR (Schwieters et al., 2003). Both can calculate structures using torsion angle dynamics with simulated annealing, but XPLOR can also using a simulated annealing protocol in Cartesian coordinate space (for review see (Guntert, 1998)).

Modern computational methods shorten the duration of structure determination by automating the assignment of distance constraints derived from NOESY spectra. Programs such as Ambiguous Restraints for Iterative Assignment (ARIA) (Linge et al., 2003a) and 'Combined Automated NOE assignment and structure Determination' (CANDID) (Herrmann et al., 2002) can substantially accelerate the process and may reduce the human bias during a manual structure calculation. Another advance in NMR structure calculation is the use of 'water-refinement' protocols (Linge et al., 2003b; Spronk et al., 2002), which are designed to tackle problems in NMR structure quality. The problems occur as a result of simplification of the force field for rapid structure calculation. This means electrostatics and van der Waals terms are usually neglected, leading to non-optimal packing, and unsatisfied hydrogen bond acceptors and donors. Refining the NMR calculated structures in a full molecular dynamics force field including electrostatic and Lennard-Jones non-bonded potentials and the interaction with solvent is one way to improve the structures.

## **1.6 Aims**

Structural information helps us to understand how muscle proteins interact to produce the process of muscle contraction – demonstrated by the studies of actin and myosin. However, structural information regarding the scaffolding features of the sarcomere, such as the Z-disk, is still in its infancy. Cryo-EM reconstructions of the Z-line have been studied (Luther, 1991; Luther, 2000b; Luther et al., 2002; Luther et al., 2003; Luther and Squire, 2002). However, finding an atomic resolution model of the sarcomere based on these models has yet to be accomplished, due to the lack of structures defining interactions between components in this region. Hence, the elucidation of high-resolution structures in this region must be established. The interactions between these components should also be studied at sufficient resolution, so the structural details and orientations between these proteins can be clarified.

Despite the crucial structural role that the  $\alpha$ -actinin-2/ZASP interaction may play in the sarcomere, no description of the interaction at atomic resolution is currently available. An important question is whether or not  $\alpha$ -actinin-2/ZASP recognition competes with other  $\alpha$ -actinin-2 interactions, such as those with titin Z-repeats, thus requiring a major rearrangement of the  $\alpha$ -actinin-2 structure to accommodate multiple complexes. It is also interesting to understand if and how other members of the enigma family could compete the  $\alpha$ -actinin-2 interaction. Several point mutations within ZASP have been identified in a form of dilated cardiomyopathy, so structural information regarding its function would further the understanding of its role in this disease.

To address these questions, the structure of ZASP-PDZ was solved by NMR. NMR and fluorescence spectroscopy was applied to study the interaction between  $\alpha$ -actinin-2 and ZASP-PDZ. The possibility of the mutual coexistence of both  $\alpha$ -actinin-2/ZASP and titin/ $\alpha$ -actinin-2 interactions were explored by characterising the binary and ternary complexes. The ZASP-PDZ coordinates were also used as a template to build homology models of other PDZ domains of the enigma family and to map conserved residues that could be functionally important.

## Chapter 2

---

# Materials and Methods

---

### 2.1 Bioinformatics

Protein primary sequences were analysed by different computational services to derive various properties. General protein properties were calculated by ProtParam (<http://us.expasy.org/tools/protparam.html>). Protein sequence similarity searches were performed by the BLASTP program (Altschul et al., 1997) on the non-redundant proteins database nrdb95 running on the EMBL server (<http://www.embl-ebi.ac.uk/~holm/nrdb90/>) (Holm and Sander, 1998); subsequent sequence alignment was done by ClustalX (Thompson et al., 1997). Selected proteins were analysed for domain identification using the Simple Modular Architecture Research Tool (SMART - EMBL Server) (Schultz et al., 2000). Secondary structure prediction was forecast by the PredictProtein PHDsec server (<http://cubic.bioc.columbia.edu/predictprotein/>) (Rost and Sander, 1993). Information regarding tertiary protein properties was also computationally assessed, including the usage of the DALI server to find similar protein structures (Holm and Sander, 1996a). Also, homology modelling was calculated by SWISS-MODEL in combination with Swiss PDB Viewer (Guex and Peitsch, 1997).

### 2.2 Plasmid Production

The constructs were available from previous studies (Faulkner et al., 1999; Joseph et al., 2001). The sequence for human skeletal muscle ZASP-PDZ was derived from GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>) AJ133766 (1-85) with a Serine-2 to Alanine mutation, which was supplied from the group of Georgine Faulkner (ICGEB, Padua, Italy). Human skeletal muscle  $\alpha$ -actinin-2 last four and last two EF-hands, Act-EF1234 and Act-EF34, derived from GenBank M86406 (745-894 and 822-894 respectively). A pET9d plasmid that conferred kanamycin resistance, for transformant selection, was utilised for protein expression. These constructs were designed to be produced as fusion proteins to aid protein purification; this included a N-terminal six-

histidine-tag (6HIS) and *S. Japonicum* glutathion transferase (GST) tag. A sequence encoding a seven-residue Tobacco Etch Virus (TEV) protease recognition site separated the tags from the target protein allowing cleavage between them.

Plasmid purification was achieved using the QIAprep Spin Miniprep Kit (Qiagen) as described in the provided protocol. Single transformed colonies were grown overnight in 5 ml of Luria-Bertani (LB) medium (Maniatis et al., 1989) at 37 °C at 200 rpm. The cells were centrifuged (4000 rpm, 9mins) and re-suspended in re-suspension buffer (P1). The cells were then lysed by addition of alkaline lysis buffer (P2), followed by acid neutralisation by addition of buffer N3. Bacterial debris was pelleted by centrifugation (12000 rpm/10 mins) using a bench micro-centrifuge, and the supernatant was transferred to a QIAprep spin column placed in a 2 ml collection tube. The column was then centrifuged for 13,000 rpm for 1 min. The bound plasmid DNA on the column was then washed with buffer PE, which contains 80 % ethanol, allowing the removal of water insoluble cell debris. Buffer PE was removed by centrifugation and the column was left to rest for 5 mins to allow residual ethanol to evaporate. The plasmid DNA was eluted with 50 µl of nuclease free H<sub>2</sub>O, pre-heated to 70 °C.

## **2.3 Protein Expression**

### **2.3.1 Transformation of Competent Cells**

The constructs were transformed using standard protocols using *E. Coli* BL21(DE3) or BL21(DE3)pLysS competent cells (Novagen). The plasmid DNA (1 µl of between 100-200 ng/µl) was added to 200 µl of competent cells, followed by ice incubation for 30 mins. The cells were heat-shocked for 60 seconds at 42 °C in a water bath followed by incubation on ice for 1-2 mins. 200 µl of SOC medium was added to the sample and left for 1 hour at 37 °C, after which they were spread on LB agar plates containing ampicillin (100 µl/ml) or kanamycin (30 µg/ml), depending on the construct, and left to incubate at 37 °C for 16 hrs.

### **2.3.2 Fermentation**

Unlabelled protein was obtained from cells grown in LB broth (Maniatis et al., 1989) containing kanamycin (30 µg/ml) to select and grow the transformed cells. Typically, a single colony was used to inoculate 25 ml of media per litre of intended

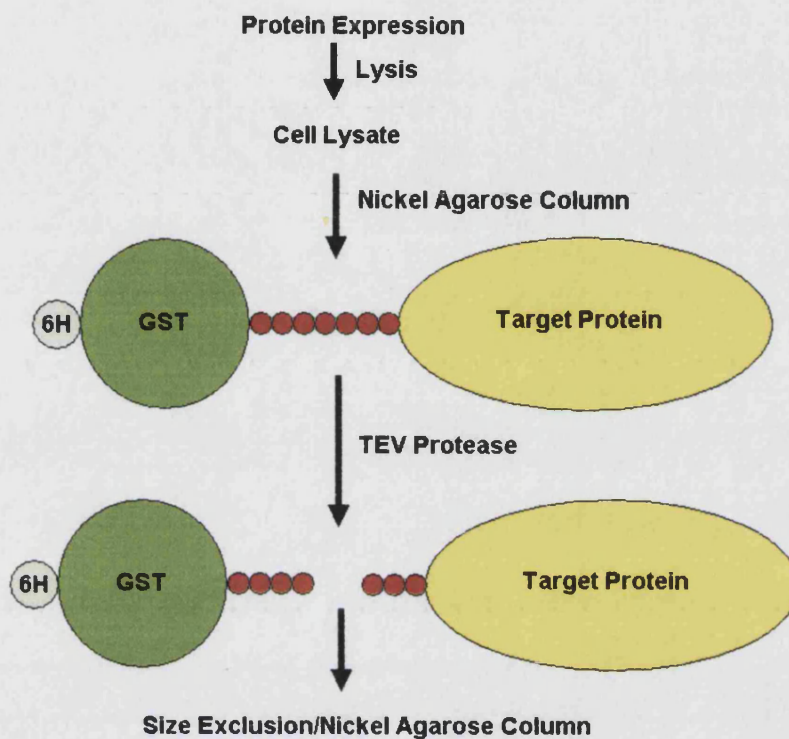
broth fermentation, incubating for approximately 16 hours at 37 °C. The 25 ml growth was then used to inoculate one litre of media and the cells grown at 37 °C. The cells were induced to express the fusion protein by the addition of isopropyl  $\beta$ -D-thiogalactopyranoside IPTG (0.5 mM final concentration) at *E. Coli*'s period of exponential growth ( $\text{O.D}_{595 \text{ nm}} \sim 0.5\text{-}0.6$ ); specifically BL21(DE3) contains a T7 polymerase under a lacZ promoter, which transcribes the target that is under a T7 promoter. The cells were then incubated at 37 °C for an additional three hours to allow for protein expression.

Uniformly labelled  $^{15}\text{N}$  or  $^{13}\text{C}$  labelled protein was produced by using isotopically enriched reagents as a source of the desired isotope, as opposed to using LB broth.  $^{15}\text{N}$ -ammonium sulphate was the only nitrogen source to produce  $^{15}\text{N}$ -labelled protein, while  $^{15}\text{N}/^{13}\text{C}$ -labelled protein was produced by substituting  $^{12}\text{C}$ -glucose with  $^{13}\text{C}_6\text{-D-glucose}$  as the only source of carbon. The protocols for expression and purification were otherwise identical. See Appendix 10.2.1 and 10.2.2 for media compositions.



## 2.4 Protein Purification

Here is a general schematic of the steps for protein purification:



**Figure 2.1:** Schematic of the steps of protein purification for 6HIS (6H), GST, and 'TEV cleavage site' (red circles) tagged proteins. Cells are lysed after protein expression, and the cell lysate is passed down a nickel-agarose column. The fusion protein is recovered and subsequently cleaved using the TEV protease. The target protein is then purified from the 6HIS-GST by size exclusion or another nickel-agarose purification.

### 2.4.1 Cell Lysis

Protein purification was carried out as described in a previous study (Joseph et al., 2001). The fusion protein is expressed in the cytoplasm, so the cells need to be lysed - breaking the bacterial cell wall and plasma membrane. Cells were pelleted after cell growth by centrifugation and re-suspended in lysis buffer (20 mM Tris.HCl pH8, 200 mM NaCl, 5 mM imidazole, 0.2% v/v IPEGAL CA-630), using 25 ml per litre growth of cells. IPEGAL CA-630 prevents hydrophobic membrane proteins aggregating, lysozyme was added as an aid to cell lysis, as it catabolises polysaccharide components of bacterial



cell walls, and DNase I was added to reduce the lysate viscosity. A single tablet of 'Protease Inhibitor Cocktail' (Roche) was also added to prevent protein degradation. The cells were left to incubate for 20 minutes on ice, to allow lysis to occur, followed by homogenisation by sonication (MODEL 250, Branson) using a 50% duty cycle with a 50-watt output for 5 minutes on ice. The supernatant was collected after centrifugation (20,000 rpm for 30 minutes at 4 °C). See Appendix 10.2.3 for solution compositions.

#### **2.4.2 Nickel NTA Agarose Protein Purification From Lysate**

The 6HIS-tagged protein was affinity purified from the cell extract by nickel-nitrilotriacetic acid (Ni-NTA) agarose column (QIAGEN). A column was packed with Ni-NTA (1 ml per litre of growth medium), and equilibrated with five column volumes (CV) of 'Lysis Buffer' (See Appendix 10.2.3 for buffer composition). The cell lysate was passed through the column followed by 5 CV of Wash Buffer, 5 CV of '1M NaCl Wash Buffer', and 5 CV of '30 mM Imidazole Wash Buffer'. These washes are intended to remove any non-specifically bound protein contaminants from the column. The 6HIS-tagged protein was eluted with 5 CV of '300 mM Imidazole Wash Buffer'. All the fractions were checked on 15% acrylamide gel electrophoresis to track the outcome of the purification.

#### **2.4.3 Tobacco-Etch Virus Cleavage**

Cleavage of the 6HIS-GST from protein of interest was achieved by using TEV protease (Gibco BRL®). The TEV protease is specific to a seven-residue tag (GLU-X-X-TYR-X-GLN-SER, cleavage between GLN and SER) between the GST and the desired protein. Approximately 2-3 µM TEV protease was used for the elution from the Ni-NTA column, and this was incubated for approximately 16 hours at room temperature.

#### **2.4.4 Gel Filtration (FPLC) Protein Purification**

The target protein was purified from 6HIS-GST (approx 25 kDa) by gel filtration using a Superdex 75 16/60 column (Pharmacia). The sample was concentrated to approximately 5 ml by using a filter concentrator (Amicon) to improve the separation on the column. The column was typically equilibrated with 20 mM Tris.HCl pH 8, 200 mM

NaCl, 0.2% NaN<sub>3</sub>, which was 0.2 µm filtered and de-gassed – this buffer was also used for the elution. The elution profile was then analysed and the fractions of interest were run 15% acrylamide gel electrophoresis for assessment.

#### **2.4.5 Nickel NTA Protein Purification After TEV Cleavage**

Act-EF1234 was purified using the Ni-NTA column, as there was a problem with aggregation if FPLC was used. Following the TEV cleavage, the product was buffer exchanged to '1M NaCl Wash Buffer' using a PD-10 column (Sephadex G25-M Pharmacia) to remove the imidazole. The sample was then applied to a Ni-NTA column containing the same quantity of Ni-NTA agarose from the purification from cell lysate, this time equilibrated with 5 CV '1M NaCl Wash Buffer'. The flow-through and a further 2 CV of 1M NaCl Wash Buffer was collected. 5 CV of '300 mM Imidazole Wash Buffer' was passed through the column to remove the 6HIS-GST from the column. The sample was then checked by 15% acrylamide gel electrophoresis.

#### **2.4.6 Buffer Exchange Using PD-10 Columns**

Protein samples were then buffer exchanged using a Sephadex G-25M PD-10 column (Pharmacia) using the standard protocol supplied. After equilibration of the PD-10 column with 25 ml of the desired buffer, 2.5 ml of protein solution were passed through. The protein was eluted with 3.5 ml of the desired buffer. All protein samples were buffer exchanged into 20 mM Na<sub>2</sub>PO<sub>4</sub>, 0.02% NaN<sub>3</sub>, and pH 6.6. All NMR samples were prepared with the addition of 10% v/v deuterium oxide (D<sub>2</sub>O) added after buffer exchange.

#### **2.4.7 Concentration Determination Using Absorption Spectroscopy**

Protein concentration was determined using the Beer-Lambert equation (Equation 2.1) using a double beam HELIOS α spectrophotometer (UNICAM). An accurate measurement of the extinction coefficient at 280 nm was determined for each protein by the Edelhoch method (Pace et al., 1995) (Equation 2.2). The absorbance was measured for the protein in the experimental buffer and also in 6M-guanidinium hydrochloride (GuHCl), subtracting the buffer baseline reading. The extinction coefficient is calculated from the amino acid composition using Equation 2.2, and then the extinction coefficient

in the buffer is calculated from Equation 2.3. Typical NMR sample protein concentrations were approximately 0.5 mM, adjusted by utilising a filter concentrator (Amicon).

$$Absorbance = \epsilon \cdot c \cdot l$$

**Equation 2.1:** Beer-Lambert equation. Absorbance is equal to the product of the extinction coefficient ( $\epsilon$ ), concentration ( $c$ ) and path length ( $l$ ).

$$\epsilon_{280, GuHCl} = (\#Trp) \cdot (5685) + (\#Tyr) \cdot (1285) + (\#Cys) \cdot (125)$$

**Equation 2.2:** The Edelhoch Method of extinction coefficient ( $\epsilon$ ) calculation from amino acid composition.

$$\epsilon_{280, buffer} = \frac{(A_{280, buffer}) \cdot (\epsilon_{280, GuHCl})}{A_{280, GuHCl}}$$

**Equation 2.3:** Calculation of extinction coefficient for protein in experimental buffer, from the extinction coefficient calculated by amino acid composition.

## **2.5 NMR**

### **2.5.1 Spectroscopy**

NMR data were recorded on Varian spectrometers: Unity-500, Unity 600, Inova-600, and Inova-800 MHz ( $^1\text{H}$  frequency), all equipped for  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  triple-resonance experiments. The spectrometers were controlled by the VNMR software (<http://www.varian.com/>) installed on SPARC workstations running Sun OS v5.7 (<http://www.sun.com/>).

Details for each spectrum, including the number of points, spectral width, spectrometer  $^1\text{H}$  frequency, and carrier frequency, for each dimension can be found in the Appendix 10.3, while many of the details for the pulse sequences can be found in (Cavanagh et al., 1996). Water suppression was achieved using WATERGATE (Piotto et al., 1992) in all experiments and experiments were run at a temperature of 300 K unless otherwise stated.

Two-dimensional COSY,  $^1\text{H}$ - $^1\text{H}$  NOESY, and  $^1\text{H}$ - $^1\text{H}$  TOCSY spectra were run on unlabelled protein. Different combinations of mixing times (40-150 milliseconds) and temperatures (283-300 K) were tested to determine optimal conditions for TOCSY and NOESY spectra. ZASP-PDZ, Act-EF1234, and Act-EF34 were found to behave well at 300 K, and therefore set for all spectral acquisition.

$^{15}\text{N}$  Heteronuclear Single Quantum Correlation spectra (HSQC) were run on  $^{15}\text{N}$  single or  $^{15}\text{N}/^{13}\text{C}$  doubly isotopically labelled protein. The  $^{15}\text{N}$ -HSQC was used for various purposes, such as sample quality evaluation, titrations, assignment purposes, etc, so acquisition parameters varied between (1024 x 64) to (2048 x 1024) data points.  $^{13}\text{C}$ -HSQC was run on  $^{13}\text{C}$ -labelled protein for assignment purposes.  $^{15}\text{N}$ -TOCSY-HSQC spectra were run on  $^{15}\text{N}$ -labelled ZASP-PDZ and Act-EF34, using mixing times of 100 and 80 ms respectively.  $^{15}\text{N}$ -NOESY-HSQC spectra were run on ZASP-PDZ, using mixing times of 50 and 100 ms, and 80 ms for Act-EF34. HNHA and HNHB experiments were run on both samples of  $^{15}\text{N}$ -ZASP-PDZ and  $^{15}\text{N}$ -Act-EF34. Only  $^{15}\text{N}$ -TOCSY-HSQC and  $^{15}\text{N}$ -NOESY-HSQC spectra were acquired for the assignment of the  $^{15}\text{N}$ -ZASP-PDZ/Act-EF1234 complex, which was present at a 1:2 ratio (see Binding Chapter 6 and Appendix 10.8)

$^{15}\text{N}/^{13}\text{C}$ -labelled protein was used for detecting CBCACONH and CBCANH spectra for both ZASP-PDZ and Act-EF34's backbone atoms assignment, and a HNCOSY spectrum was run for ZASP-PDZ. Side-chain assignments were completed using an

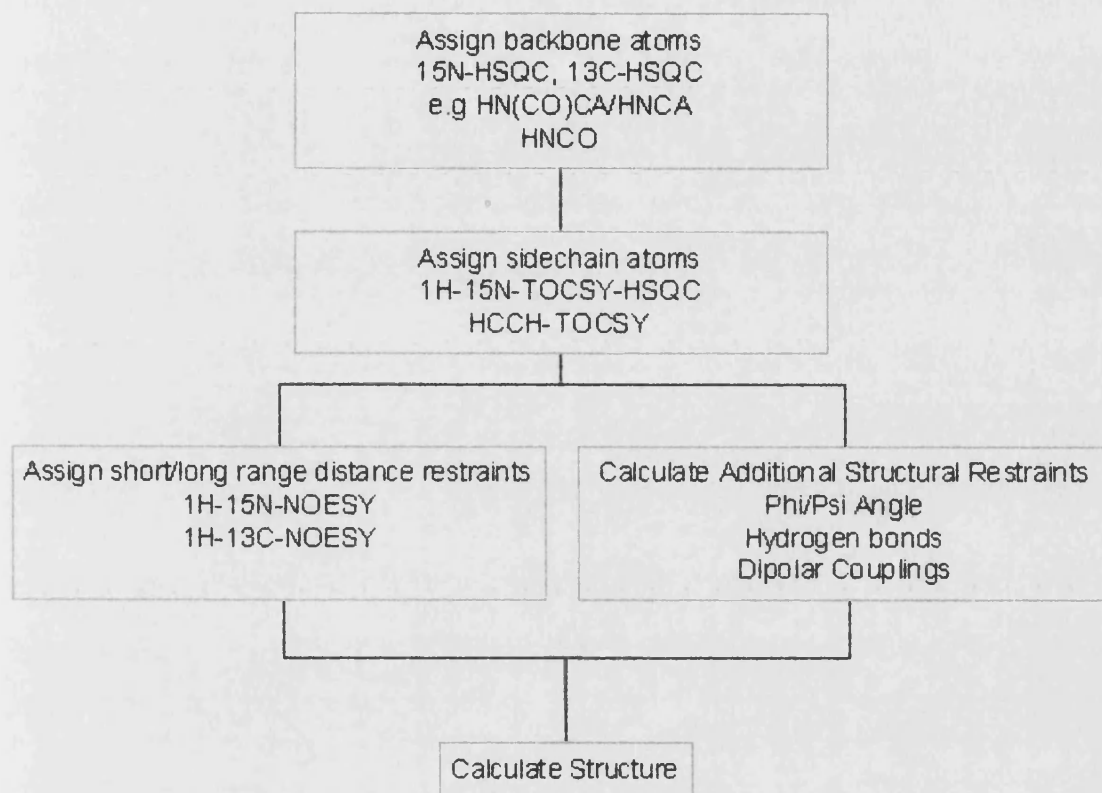
HCCH-TOCSY for ZASP-PDZ. Additional spectra were acquired for the structure of ZASP-PDZ, including a  $^{13}\text{C}$ -NOESY HSQC. T1, T2, and NOE relaxation data spectra were also acquired for ZASP-PDZ and Act-EF34 (see Appendix 10.3.1.1.20 and 10.3.1.2.9 for relaxation delays used).

### **2.5.2 Spectral Processing**

NMR spectra processed using nmrPipe (Delaglio et al., 1995). Typically, the acquisition dimension was multiplied by a Gaussian window function and a 90° shifted sine-bell function in the indirect dimensions to enhance the resolution of the spectra. Polynomial solvent subtraction was used on the acquisition dimension if required, and linear prediction was used on indirect dimension appropriately to improve and increase the resolution of the spectrum respectively. All dimensions were zero-filled to the next appropriate power of two. SPSCAN (<http://www.molebio.uni-jena.de/~rwg/spscan>) was used to convert the nmrPipe data to XEASY (Bartels et al., 1995) for analysis.

## 2.6 NMR Structure Determination

Here is a schematic for NMR structure determination:



**Figure 2.2:** NMR structure determination pipeline. First, the backbone atoms of the protein ‘chemical shift’ are assigned, allowing for sidechain atoms assignment. Structural restraints are then analysed (distance, angular, and orientational), which are then used for structure calculation.

### 2.6.1 NOESY spectra

All NOESY spectra were acquired and processed as outlined in Materials and Methods 2.5. Spectral analysis was performed using XEASY (Bartels et al., 1995). A combination of two and three-dimensional NOESY spectra were analysed for distance constraint assignment. The NOESY spectra included a  $^1\text{H}$ - $^1\text{H}$ -NOESY with a 100 ms mixing time,  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC at 50 ms, and  $^1\text{H}$ - $^{13}\text{C}$ -NOESY-HSQC at 100 ms. The  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC was analysed by using an assigned  $^1\text{H}$ - $^{15}\text{N}$ -HSQC-TOCSY peak list followed by picking all other observable NOESY peaks. The same procedure was implemented for analysing the  $^1\text{H}$ - $^{13}\text{C}$ -NOESY-HSQC, by using HCCH-TOCSY peak list. The  $^1\text{H}$ - $^1\text{H}$ -NOESY was analysed by utilising the assignments made in the  $^{15}\text{N}$  and  $^{13}\text{C}$  NOESY-HSQC and analysing the remaining peaks.

Peak assignment began with assigning short and medium-range NOEs, while initial long-range NOEs assignment was achieved by finding peaks with unique assignment possibility. A homology model of ZASP-PDZ assisted initial long-range NOE assignment, and secondary structure prediction was useful for assigning medium-range NOEs. Peaks with multiple assignments were assigned by observing restraints on preliminary calculated structures. This procedure was then semi-automated by employing MOLMOL (Koradi et al., 1996) to calculate distances less than 7 Å between all atoms in 10 from the best 30 of 100 CYANA generated structures. This distance information was then combined with the possible assignments observed in XEASY, and a Perl script ('dist\_search.pl', Appendix 10.9.5) was used for the quick evaluation of the constraint.

### 2.6.2 Distance Calibration

The distance calibration involves converting peak intensities or volumes into upper-distance limits between atoms (Equation 2.4), both are directly proportional to each other (Equation 2.5), assuming that the line-width for all peaks is identical, so either can be used for peak calibration. Peak integration for the NOESY spectra was calculated using SPSCAN (<http://www.molebio.uni-jena.de/~rwg/spscan>) with five iterations of integration, while peak intensities were directly measured in XEASY. CALIBA (Guntert et al., 1997) was utilised for distance calibration, or alternatively the Perl script 'calibrate.pl' (Appendix 10.9.7) categorised the volumes/intensities into

strong/medium/weak/very weak. Known distances in secondary structure elements (Wuthrich, 1986) verified the calibration.

$$V \propto \left( \frac{A}{r^6} \right)$$

**Equation 2.4:** Peak volume is inversely proportional to distance.

$$V = I \cdot \left( \frac{\pi \cdot r^2}{2} \right)$$

**Equation 2.5:** Relationship between volume and intensity.

Peak overlap was resolved on an individual restraint basis by looking at alternative instances where the overlap was not present; for example, the  $^1\text{H}$ - $^1\text{H}$ -NOESY, in some cases, had less overlap than the three-dimensional spectra due to its higher resolution. If the peak overlap was not resolved by this substitution, the restraint was set to the maximum observable NOE distance of 6 Å. Pseudoatom correction was added to all cases where there was chemical shift redundancy for stereo-pairs that were not stereo-specifically assigned. Stereo-specific assignment was calculated by GLOMSA (Guntert et al., 1991) – this makes stereo-specific assignments of diastereotopic substituents of prochiral centres from calculated structures and distance restraints.

### 2.6.3 Hydrogen Bonds, Dihedral Angles, and Residual Dipolar Couplings

Directly observed hydrogen bonds were used as structural restraints by acquiring and analysing a modified HNCO (Cordier and Grzesiek, 1999). Dihedral angle ( $\phi$  and  $\psi$ ) restraints were calculated using the program TALOS (Cornilescu et al., 1999), which operates by comparing the chemical shift assignment with a database of known protein assignments with known secondary structure elements. Additional and confirmatory  $\phi$  dihedral angles were calculated using HNHA spectra. The ZASP-PDZ NMR sample was partially aligned by using a poly(ethylene glycol) liquid crystalline phase (Ruckert and Otting, 2000): n-dodecyl-penta(ethylene glycol) (C12E5) was added to the NMR sample at a 5% w/v ratio, and n-hexanol was added at a 0.96 molar ratio against the C12E5 to form the lyotropic liquid crystalline phase. Dipolar coupling values for the backbone amide protons were measured by running a J-modulated HSQC experiment (Tjandra et al., 1996).



#### 2.6.4 CYANA

The structure of ZASP-PDZ was calculated with CYANA (Guntert et al., 1997): a program that uses simulated annealing combined with molecular dynamics in torsion angle space. Distance restraints from the  $^1\text{H}$ - $^1\text{H}$ -NOESY,  $^{13}\text{C}$ -NOESY-HSQC, and  $^{15}\text{N}$ -NOESY-HSQC were included for the structure calculation. Restraints from the hydrogen bond experiment, and also the dihedral angles calculated from TALOS and HNHA spectra were inputted only after the NOE data had converged. The simulated annealing protocol involved a high temperature conformational search stage of 2000 steps, followed by slow cooling over 8000 steps with conjugate gradient minimisation at the end. The redundant dihedral angle (REDAC (Guntert and Wuthrich, 1991)) strategy was used to help the convergence of the structures, typically using 5 cycles of 100 structures. Tolerated violations included upper, lower, and van der Waals distance limits of 0.5 Å, and angle violations of 5 degrees.

#### 2.6.5 CANDID

The assignment of ZASP-PDZ and the peak list (both in XEASY format) from the  $^1\text{H}$ - $^1\text{H}$ ,  $^1\text{H}$ - $^{13}\text{C}$ , and  $^1\text{H}$ - $^{15}\text{N}$  NOESY spectra were used for the CANDID module (Herrmann et al., 2002) within CYANA. The chemical shift tolerances used in both dimensions of the  $^1\text{H}$ - $^1\text{H}$ -NOESY were 0.02 ppm, and 0.03 and 0.05 ppm for the proton and  $^{13}\text{C}/^{15}\text{N}$  nuclei in the three-dimensional NOESY spectra respectively. Non-NOE restraints included hydrogen bonds, and the TALOS and HNHA dihedral angles. The default run of seven cycles was used for CANDID structure calculation.

#### 2.6.6 XPLOR

The NIH version of XPLOR (Schwieters et al., 2003) was utilised for the structure calculation of ZASP-PDZ, again using torsion angle dynamics with simulated annealing (Stein et al., 1997). Structural data from CYANA calculations were directly converted for use in the XPLOR calculation: NOE upper-limit boundaries were converted from the CYANA calculation using the Perl script d2x.pl (Appendix 10.9.1). The pseudoatom correction which was automatically added to the distance constraints from the CYANA calculation were subtracted (as XPLOR can add its own pseudoatom correction), and all distances were defined as straight upper-limit boundaries with a 1.8 Å

lower-limit. The stereo-specific assignments were implicit from the converted NOE restraints. Hydrogen bond definitions were inputted with standard XPLOR definitions. Dihedral angles ( $\phi$  and  $\psi$ ) were converted using *aco2tbl.pl* (Appendix 10.9.3), and all angles were inputted with an error of  $\pm 30$  degrees. 78 RDCs were derived from the HSQC-IPAP spectra, but 15 were unused, as they were deemed intrinsically mobile. Mobility was gauged by looking at the relaxation analysis (Appendix 10.3.1.1.20) - if the backbone amide T1/T2 relaxation was outside one standard deviation from the mean, or if the backbone amide had an NOE value of less than 0.7. Tolerated violations included upper, lower, and van der Waals distance limits of 0.5 Å, and angle violations of 5 degrees.

### **2.6.7 CHARM22 Water Refinement**

Twenty of the best structures from both XMANUAL and XCANDID calculations were water-refined using the 'CHARMM22 water-refinement protocol'. The protocol was carried out as detailed in (Spronk et al., 2002), in combination with distance and dihedral angle constraints.

## 2.7 <sup>15</sup>N-HSQC Titration

Unlabelled and <sup>15</sup>N-labelled protein was expressed and purified as detailed in Materials and Methods 2.3 and 2.4. <sup>15</sup>N-HSQC spectra acquisition was carried out as shown in Materials and Methods 2.5, typically acquired using (2048 x 64) points in the direct and indirect dimensions respectively. Titration schemes varied depending on the titration; a typical pair-wise NMR titration involved the incremental addition of unlabelled protein to an amount of <sup>15</sup>N-labelled protein and acquiring <sup>15</sup>N-HSQC spectra for each addition of unlabelled protein. Molar ratios of <sup>15</sup>N-labelled to unlabelled protein ranged between 10:1 to 1:2, or until no further spectral changes were observed (Appendix 10.8).

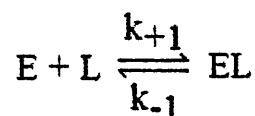
NMR spectra were converted and processed in nmrPipe (Delaglio et al., 1995) as described in Materials and Methods 2.5, and analysis of the changes in <sup>1</sup>H and <sup>15</sup>N chemical shift in the <sup>15</sup>N-HSQC was monitored using XEASY (Bartels et al., 1995). The  $\Delta\delta_{\text{weighted}}$  change was calculated for each peak using Equation 2.6 to simplify the analysis, instead of using <sup>1</sup>H and <sup>15</sup>N chemical shift data. The final concentration point of the titration was used to plot the binding charts ( $\Delta\delta_{\text{weighted}}$  vs. residue number).

$$\Delta\delta_{\text{weighted}} = \sqrt{\Delta\delta_{1H}^2 + \left(\frac{\Delta\delta_{15N}}{10}\right)^2}$$

**Equation 2.6:** Calculation of  $\Delta\delta_{\text{weighted}}$ , which combines <sup>1</sup>H and <sup>15</sup>N changes in chemical shift. The <sup>15</sup>N dimension is divided by ten to account for the difference in gyro-magnetic ratio between <sup>1</sup>H and <sup>15</sup>N.

Dissociation constant (K<sub>d</sub>) calculations were performed to assess the affinity between proteins. Peaks in the spectra were visually inspected to check their suitability for K<sub>d</sub> calculation; to simplify the analysis, only peaks in fast exchange (with respect to the NMR time-scale) were used. For each calculation, a 1:1 binding model was assumed (Equation 2.7), from which the K<sub>d</sub> is determined (Equation 2.8). Equation 2.9 (Roberts, 1993) shows that the K<sub>d</sub> can be solved from the observed chemical shift ( $\delta_{\text{obs}}$ ), the initial chemical shift of the ligand ( $\delta_L$ ), the total amount of enzyme ( $E_T$ ), and the total amount of ligand  $L_T$  – the  $\delta_{EL}$  and K<sub>d</sub> is then solved using non-linear regression analysis (using the computer program Prism (Graphpad).  $\delta_{\text{obs}}$  is a weighted average of the chemical shifts of the protein free in solution ( $\delta_L$ ) and in complex ( $\delta_{EL}$ ) (Equation 2.10); the weighting

being the fractional population of each state ( $\rho_L$  and  $\rho_{EL}$  respectively, defined in Equation 2.11: A and B).



**Equation 2.7:** Simple binding model of an enzyme (E) binding to a ligand (L) forming the complex (EL). The equilibrium constants for association and dissociation dictate where the equilibrium lies.

$$K_d = \frac{k_{-1}}{k_{+1}} = \frac{[E][L]}{[EL]}$$

**Equation 2.8:** The dissociation constant ( $K_d$ ) is calculated from the concentration of the enzyme [E], ligand [L], and the resultant complex [EL].

$$\delta_{obs} - \delta_L = \frac{(\delta_{EL} - \delta_L) \left\{ (E_T + L_T + K_d) - \sqrt{(E_T + L_T + K_d)^2 + 4E_T L_T} \right\}}{2L_T}$$

**Equation 2.9:** This indicates that the  $K_d$  can be determined from  $\delta_{obs}$  (see above for a definition of the parameters). By plotting  $\delta_{obs}$  against  $L_T$ , non-linear regression analysis can be used to determine  $\delta_{EL}$  and  $K_d$ .

$$\delta_{obs} = \delta_L \rho_L + \delta_{EL} \rho_{EL}$$

**Equation 2.10:** Observed chemical shift is the weighted average of the chemical shift values from the bound and free states.

$$\begin{aligned} \text{A) } \rho_L &= \frac{[L]}{[L] + [EL]} \\ \text{B) } \rho_{EL} &= \frac{[EL]}{[L] + [EL]} \end{aligned}$$

**Equation 2.11:** A) The mole fraction of the ligand ( $\rho_L$ ) is calculated from concentration of the free ligand [L], divided by the total amount of ligand – the free ligand and the ligand in complex [EL]. B) The mole fraction of the complex is calculated in an analogous fashion.

## 2.8 Fluorometric Titration

The fluorometric titrations were carried out by Dr. Steve Martin (NIMR, London). Dissociation constants for the interaction of ZASP-PDZ or ZASP-PDZ/tZR7 with Act-EF34 were determined at 20 °C in 20 mM phosphate (pH 6.6) by direct fluorometric titration at 322 nm using a SPEX FluoroMax fluorimeter with  $\lambda_{ex} = 295$  nm. Only one tryptophan in ZASP-PDZ is present, so the titrations of ZASP or ZASP-tZR7 with Act-EF34 (A) were fit to the following equation: Fluorescence =  $F(Z)[Z] + F(AZ)[AZ]$ , where the F values are the molar fluorescence intensities and Z denotes tZR7 and the ZASP-tZR7 complex. A value for the dissociation constant ( $K_d$ ) was obtained from a non-linear least squares fit to this equation (Marquardt method) with concentrations calculated by solving the equation:

$$[AZ]^2 - [AZ](K_d + A_{TOT} + Z_{TOT}) + A_{TOT}Z_{TOT} = 0$$

**Equation 2.8:**  $A_{TOT}$  and  $Z_{TOT}$  are the initial concentrations of unbound Act-EF34 and ZASP-PDZ respectively.  $[AZ]$  is the concentration of the complex.

Derived by the following proof; where Act-EF34 (A) binds to ZASP-PDZ (Z) according to the following scheme:



Total concentrations of A and Z ( $A_{TOT}$  and  $Z_{TOT}$ ) are given by:

$$A_{TOT} = [A] + [AZ]$$

$$Z_{TOT} = [Z] + [AZ]$$

Then:

$$K_d = (A_{TOT} - [AZ])(Z_{TOT} - [AZ]) / [AZ]$$

and:

$$K_d[AZ] = A_{TOT}Z_{TOT} - [AZ](A_{TOT} + Z_{TOT}) + [AZ]^2$$

and:

$$[AZ]^2 - [AZ](K_d + A_{TOT} + Z_{TOT}) + A_{TOT}Z_{TOT} = 0$$

## **2.9 Peptide synthesis and purification of the Minimal Titin ZR7 peptide**

Peptide synthesis was carried out by Dr. Remo Guerrini (University of Ferrara, Italy). The peptide was designed by observing the residues present in the NMR structure of Act-EF34 in complex with ZR7 (Atkinson et al., 2001), plus an addition one further residue of the ZR7 sequence to aid the peptide synthesis. The tZR7 peptide (GKKAEAVATVVAAVDQARVREPR) was synthesized using standard solid-phase synthesis techniques (Atherton and Sheppard, 1989) with a Milligen 9050 synthesizer. Protected amino acids and chemicals were purchased from Bachem, Novabiochem or Fluka (Switzerland). The resin loaded with glutamic acid (Fmoc-Glu(OBut)-PEG-PS) from Millipore (Waltham, MA, U.S.A.) (0.5 g) was treated with piperidine (20%) in DMF. The N<sup>α</sup>-Fmoc amino acid derivatives were sequentially coupled to the growing peptide chain by using a four-fold excess of [O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate] (HATU (Carpino, 1993)) in DMF. Piperidine (20%) in DMF was used to remove the Fmoc group at all steps. The protected peptide-PEG-PS-Resin was cleaved from the resin by treatment with TFA/H<sub>2</sub>O/phenol/ethanedithiol/thioanisole (reagent K) (82.5 : 5 : 5: 2.5: 5; v/v) (King et al., 1990).

Crude peptide was purified by preparative reversed-phase HPLC using a Water Delta Prep 4000 system with a Waters PrepLC 40 mm Assembly column C18 (30 x 4 cm, 300 Å, 15 mm spherical particle size column) using a linear gradient from 10 to 40% of acetonitrile in 0.1% TFA. The molecular weight of the compound was confirmed by a Matrix Assisted Laser Desorption Ionisation Time of Flight (MALDI-TOF) analysis using a Hewlett Packard G2025A LD-TOF system mass spectrometer.

## Chapter 3

---

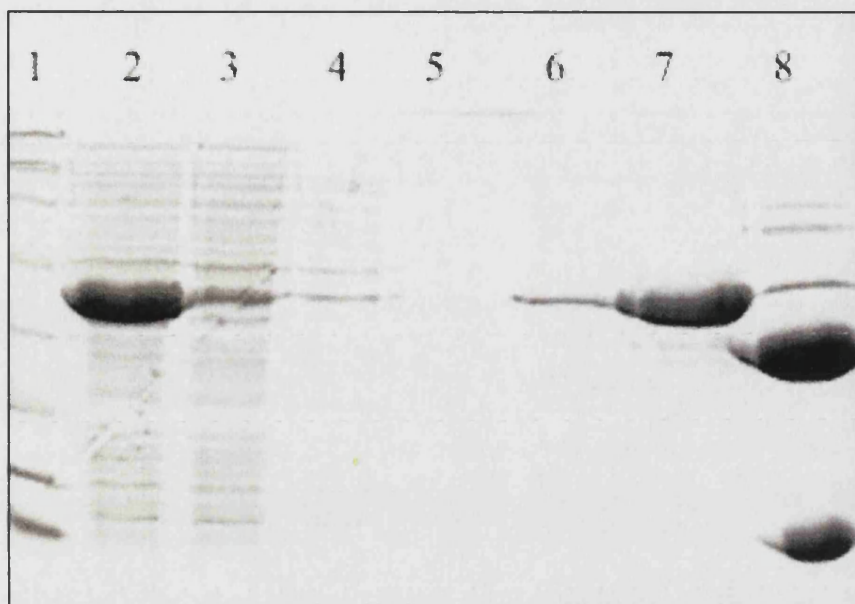
# Protein Expression and Purification

---

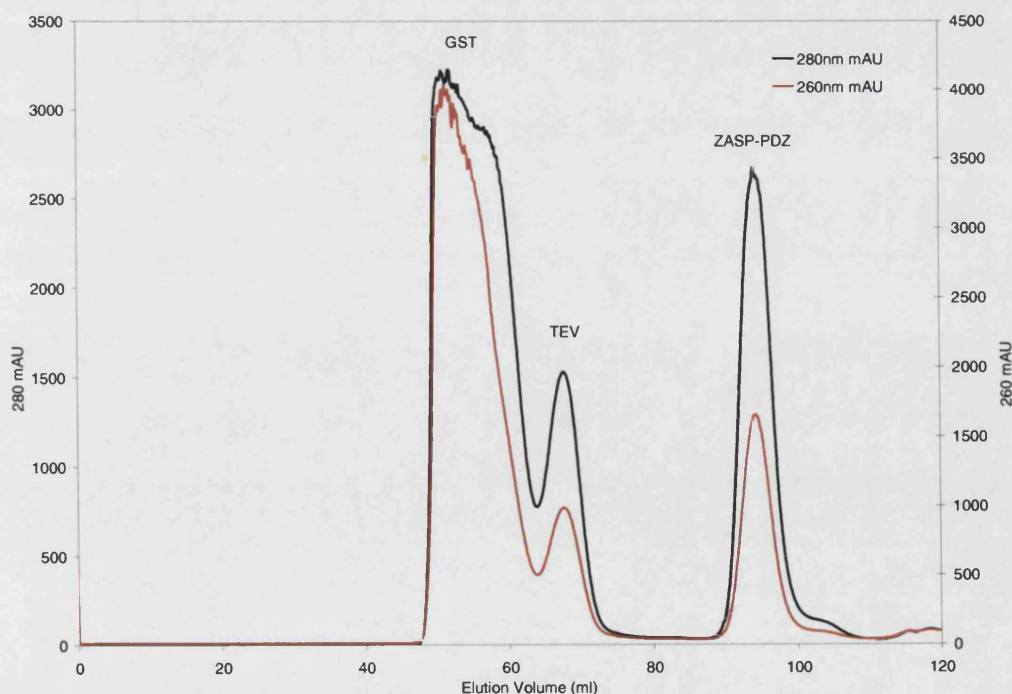
### 3.1 Protein Purification of ZASP-PDZ

All expression and purification was carried out as stated in the Materials and Methods (2.3 and 2.4). Expression of ZASP-PDZ was carried in LB,  $^{15}\text{N}$  minimal media, and  $^{15}\text{N}/^{13}\text{C}$  minimal media. Protein purification after cell lysis was performed by nickel agarose affinity purification to recover the fusion protein: ZASP-PDZ gave a predictable elution profile for each of the purification steps; the gel (Figure 3.1) showed that the first three washes removed the other proteins contained in the cell lysate. The first wash showed an excess of the 6HIS-GST-ZASP-PDZ fusion protein, as some is present in the unbound fraction. The '30 mM Imidazole Wash' showed no detectable impurities, whilst the '300 mM Imidazole Wash' successfully recovered the fusion protein from the column.

The fusion protein was then subsequently cleaved with TEV protease (Figure 3.1), which is apparent from the gel: it appears that >95% of the fusion protein was cleaved, giving the 6HIS-GST, and ZASP-PDZ. ZASP-PDZ was then recovered from TEV protease and 6HIS-GST by size exclusion. The FPLC profile (Figure 3.2) shows well-resolved peaks with GST, TEV, and ZASP-PDZ eluting at 55, 70 and 95 ml respectively. A polyacrylamide SDS gel (Figure 3.3) showed that the ZASP-PDZ contained a single species. The sample was buffer exchanged and concentrated between 0.5 – 1.0 mM, depending on the usage. The UV/VIS spectrum (Figure 3.4) gave a profile from which the concentration of the protein could be determined at 280 nm using the extinction coefficient ( $8250 \text{ M}^{-1} \text{ cm}^{-1}$ ) derived from the Edelhoch method (Materials and Methods 2.4.7 and Appendix 10.1). The yield for the unlabelled ZASP-PDZ protein was approximately 9 mg/L of LB media, whilst either single  $^{15}\text{N}$  or double  $^{15}\text{N}/^{13}\text{C}$  labelled expressions were around 5-6 mg/L of minimal media.

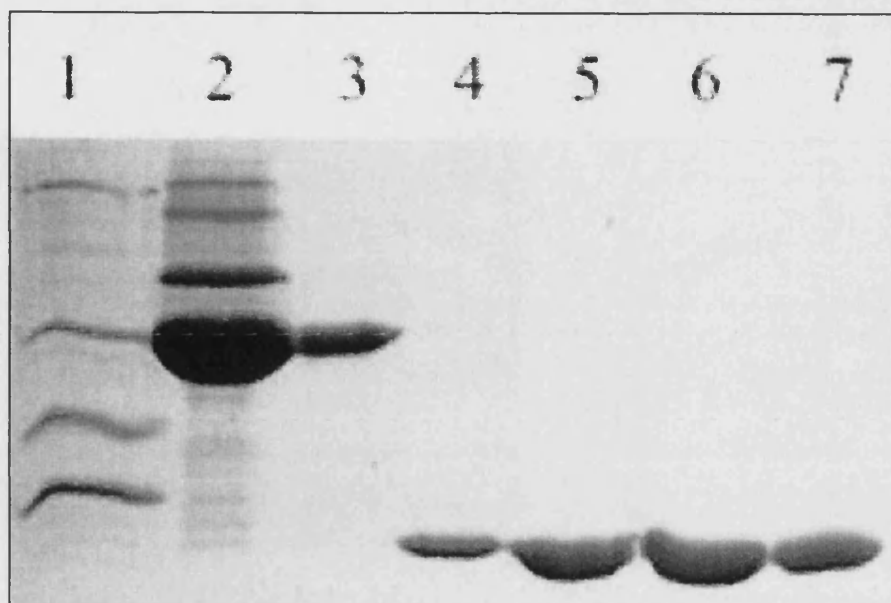


**Figure 3.1:** Polyacrylamide SDS gel of the purification of ZASP-PDZ; from cell lysis up till TEV protease cleavage. 1) Broad molecular weight marker (Appendix 10.2.6), 2) cell lysate supernatant, 3 to 7) nickel agarose purification (in order, unbound fraction, column buffer wash, '1M NaCl' wash, 30 mM imidazole wash, and 300 mM imidazole wash, 8) TEV protease cleavage product containing 6HIS-GST and ZASP-PDZ.

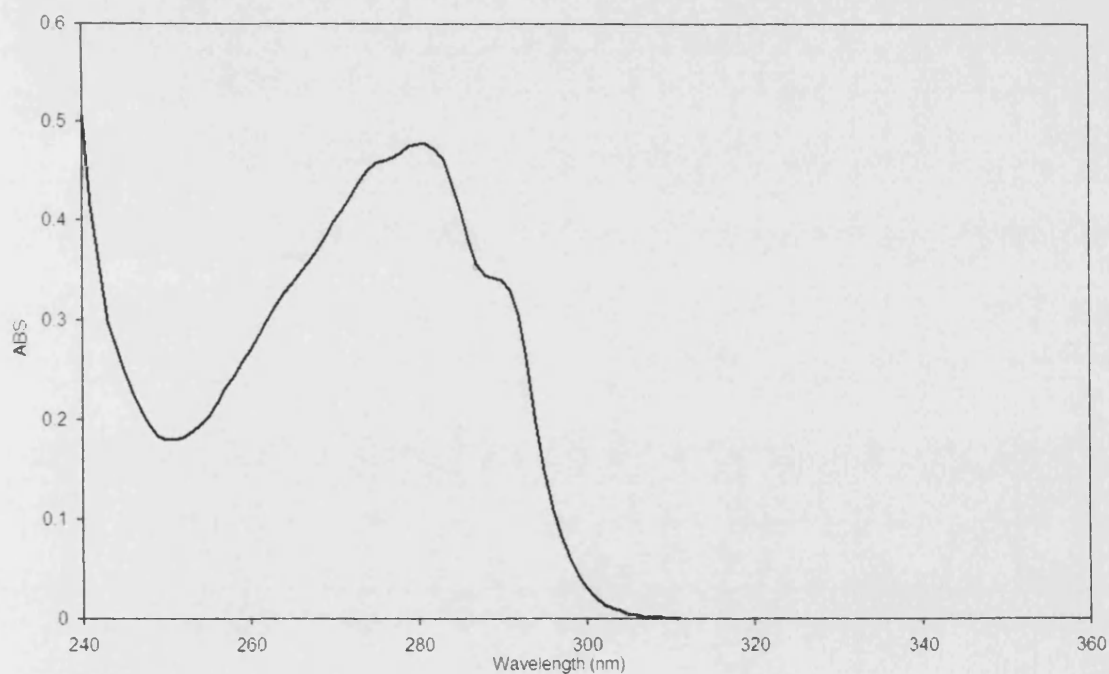


**Figure 3.2:** FPLC elution profile of ZASP-PDZ monitoring fluorescence at 260 and 280 nm.





**Figure 3.3:** Polyacrylamide SDS gel of a ZASP-PDZ FPLC result. 1) Marker, 2) un-cleaved 6HIS/GST/ZASP-PDZ, 3) GST (eluent at 55 ml from Figure 3.2), 4-7) Pure ZASP-PDZ elution (elution volume of 95 ml).

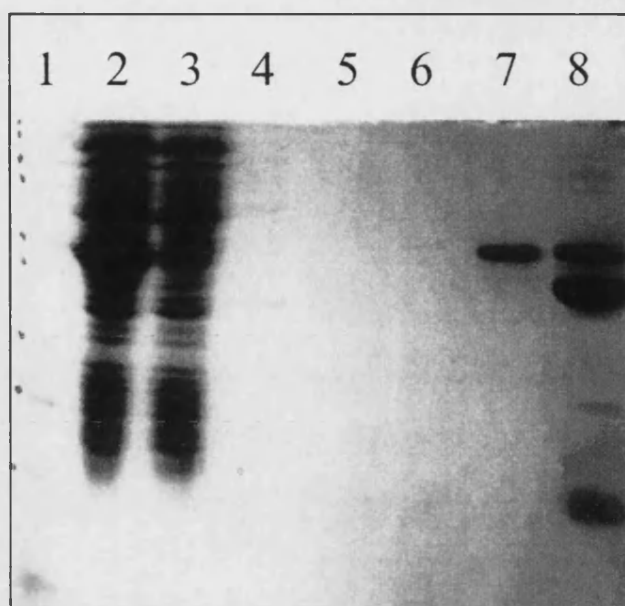


**Figure 3.4:** UV/Visible light absorbance spectra of ZASP-PDZ (at 0.57 mM), used to measure protein concentration at 280 nm.

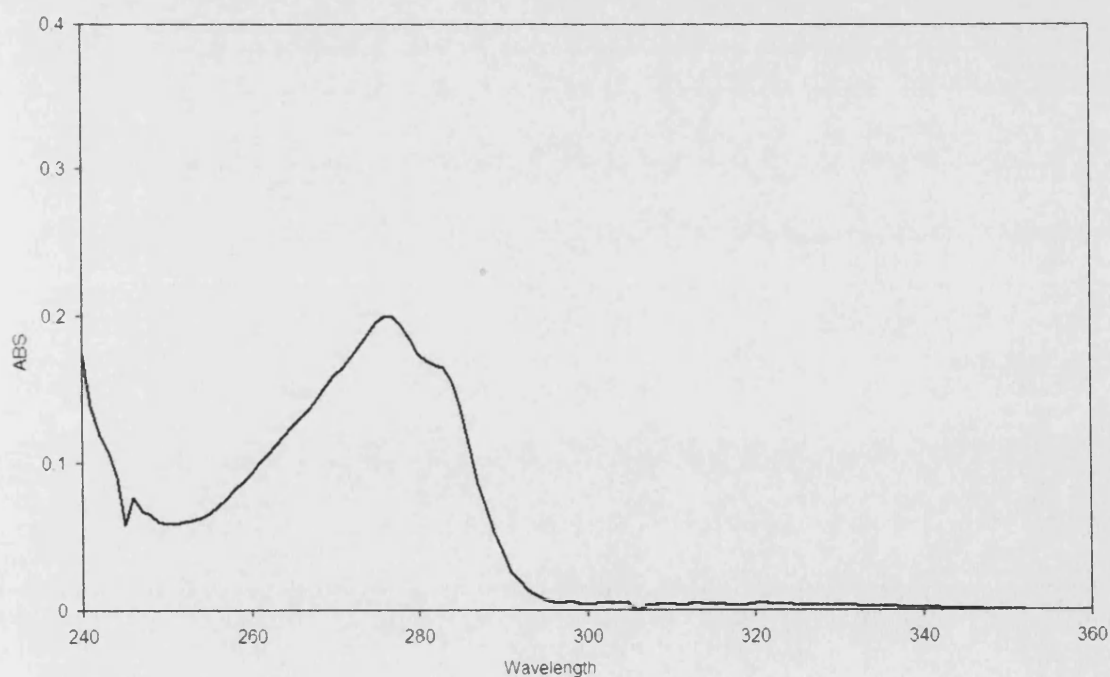
### **3.2 Protein Purification of Act-EF34**

Identical protocols as for ZASP-PDZ were used to express and purify Act-EF34, expressing in LB,  $^{15}\text{N}$ , and  $^{15}\text{N}/^{13}\text{C}$  minimal media. Nickel agarose and TEV protease cleavage was performed to recover the fusion protein, and cleave Act-EF34 from the 6HIS-GST, respectively (Figure 3.5). Initial attempts to express and purify Act-EF34 resulted in co-purification with GST after FPLC purification. However, alterations to the protocol of cell lysis by sonication improved the purification – the heat caused by sonication may have aggregated the protein, so steps were made to limit the sonication time and improve heat dissipation. After these changes were made the construct was well behaved, with the fusion protein being correctly expressed, purified, and cleaved.

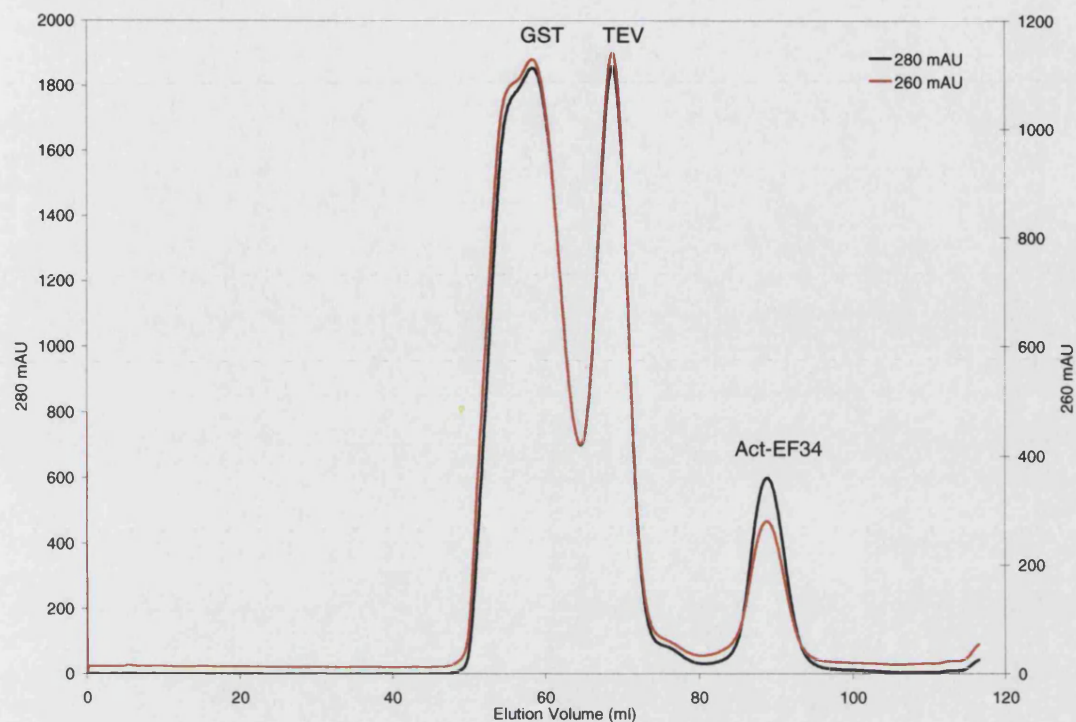
Act-EF34 was purified by size exclusion (Figure 3.7), giving an elution profile showing three distinct peaks for GST, TEV, and Act-EF34, eluting at 55, 70 and 90 ml respectively, which appeared to give a single species (Figure 3.8). The sample was buffer exchanged to NMR buffer and concentrated between 0.5 – 1.0 mM. The concentration of Act-EF34 was deduced from a UV/VIS spectrum at 280 nm (Figure 3.6) using the extinction coefficient (6400  $\text{M}^{-1} \text{cm}^{-1}$ ). The yield of Act-EF34 was approximately 4 mg/L of LB media, while only approximately 2 mg/L of  $^{15}\text{N}$  minimal media.



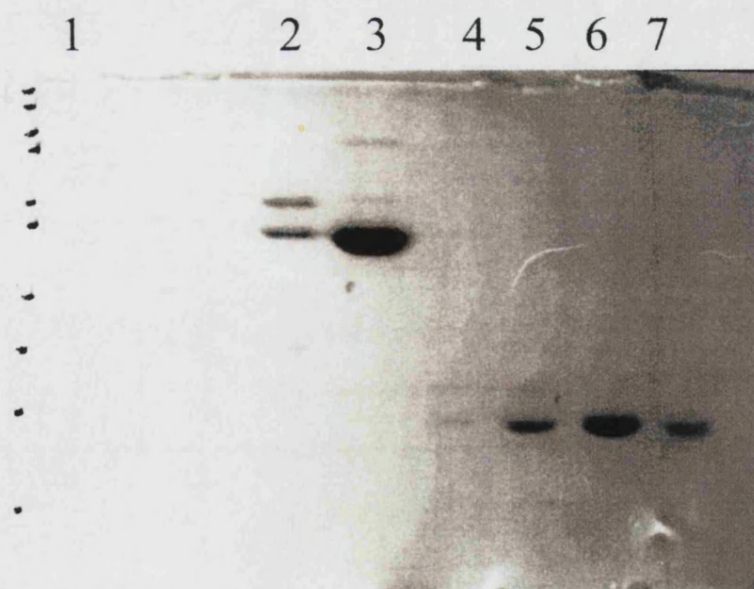
**Figure 3.5:** Polyacrylamide SDS gel of Act-EF34 purification, from cell lysis until TEV protease cleavage. 1) Molecular weight marker, 2) cell lysate, 3 to 7) nickel agarose purification (in order of, unbound fraction, column buffer wash, '1M NaCl' wash, 30 mM imidazole wash, and 300 mM imidazole wash, 8) TEV protease cleavage product.



**Figure 3.6:** UV/Visible light absorbance spectra of Act-EF34 (at 0.27 mM), used to measure protein concentration at 280 nm.



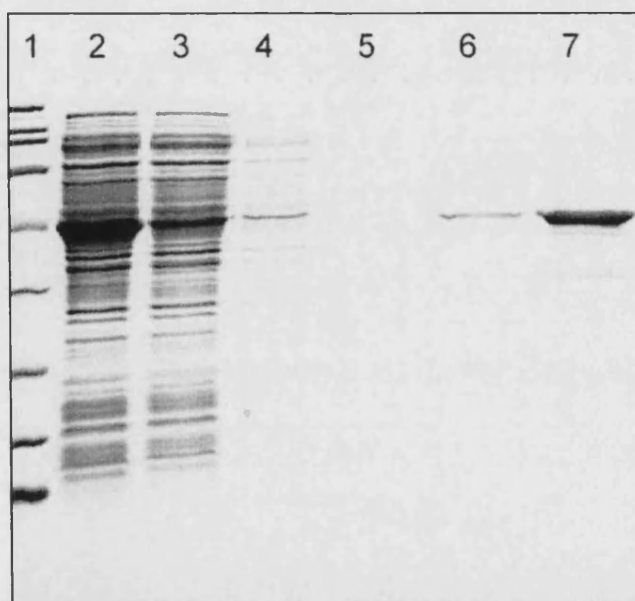
**Figure 3.7:** FPLC elution profile of Act-EF34 monitoring fluorescence at 260 and 280 nm. Act-EF34 is shown here to elute at a volume of 90 ml.



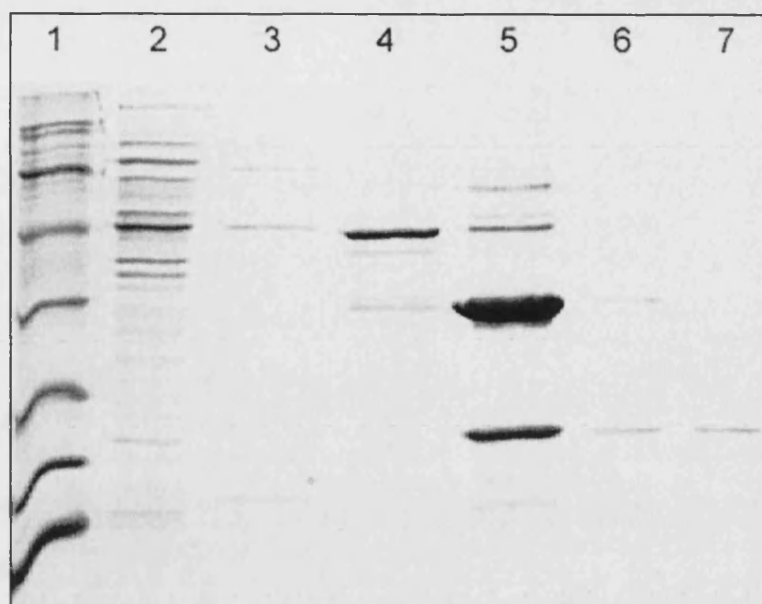
**Figure 3.8:** Polyacrylamide SDS Gel of an Act-EF34 FPLC result. 1) Marker, 2) 6HIS-GST-Act-EF34 fusion protein (elution volume of 55 ml) 3) GST (elution volume of 70 ml), 4-7) Pure Act-EF34 elution (elution volume of 90 ml).

### **3.3 Purification of Act-EF1234**

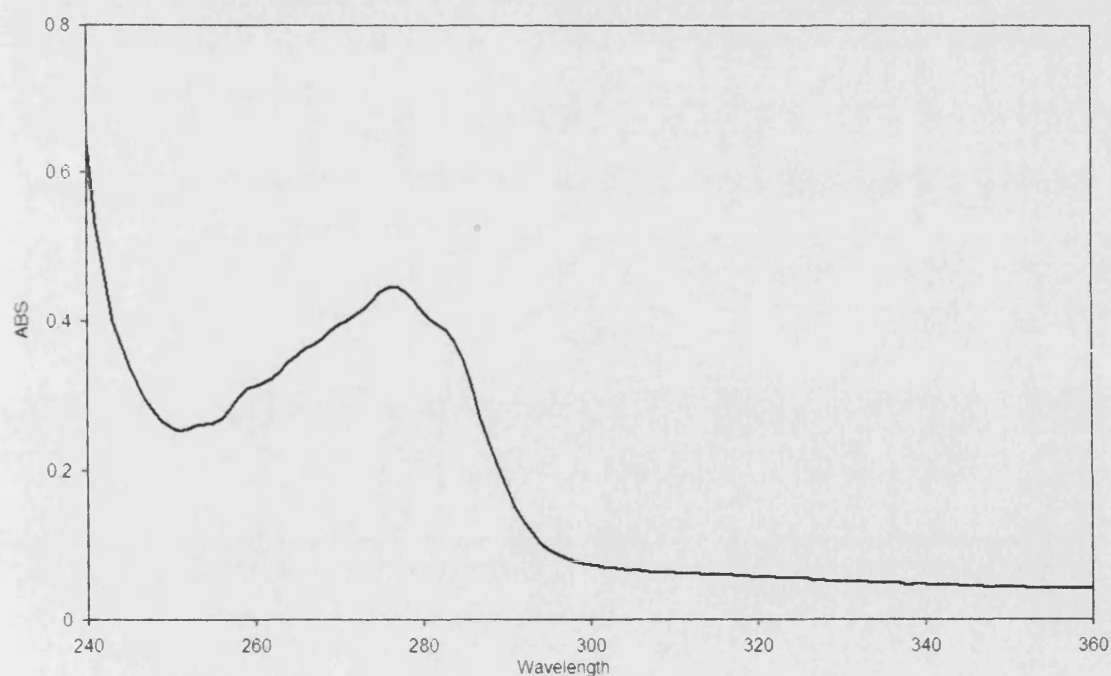
Act-EF1234 was expressed in LB and singularly labelled ( $^{15}\text{N}$ ) minimal media. After cell lysis Act-EF1234 fusion protein was purified by nickel agarose column, followed by TEV protease cleavage, again giving a well-behaved and predictable profile (Figure 3.9). However, Act-EF1234 aggregated on FPLC purification, so imidazole was removed via buffer exchange allowing nickel agarose column purification (Figure 3.10). The sample was buffer exchanged and concentrated, and the concentration was deduced from a UV/VIS spectrum at 280 nm (Figure 3.11), using the extinction coefficient (7800  $\text{M}^{-1} \text{cm}^{-1}$ ). The yield for Act-EF1234 was quite variable, possibly due to the aggregation problems, varying between approximately 2-3 mg/L of LB, and roughly 2 mg/L for  $^{15}\text{N}$  minimal media.



**Figure 3.9:** Polyacrylamide SDS gel of the purification of Act-EF1234 from cell lysis until TEV protease cleavage. 1) Broad molecular weight marker, 2) cell lysate, 3 to 7) nickel agarose purification (in order, unbound fraction, column buffer wash, '1M NaCl' wash, 30 mM imidazole wash, and 300 mM imidazole wash).



**Figure 3.10:** Polyacrylamide SDS gel of the nickel agarose column purification for Act-EF1234. 1) Molecular weight marker, 4) Elution from first nickel agarose column, 5) TEV cleavage product, 6) Second nickel agarose column elution 7) Third nickel agarose column elution containing pure Act-EF1234.



**Figure 3.11:** UV/Visible light absorbance spectra of Act-EF1234 (at 0.51 mM), used to measure protein concentration at 280 nm.

### **3.4 Conclusion**

All three constructs of ZASP-PDZ, Act-EF1234, and Act-EF34 were successfully expressed and purified. A serious complication was that 6HIS-GST-Act-EF1234 TEV cleavage product precipitated upon concentration, thus making purification by size exclusion problematic. Opting for a second nickel-agarose affinity purification to remove the 6HIS-GST, resulting in a pure Act-EF1234 product, circumvented this. After the constructs were purified, buffer exchange into 'NMR buffer' and protein concentration was trivial, allowing for the production of NMR amenable samples. ZASP-PDZ was the best behaved, as protein expression and purification was not problematic, and the yield was the highest – up to 9 mg/L in LB media. ZASP-PDZ was also very stable, as it did not appear to degrade, even left for months at 4 °C. Act-EF34 was the next best behaved, giving fairly predictable expression and purification profiles, and respectable yields. However, there was a problem with the uniformity of the sample, as a second species appeared to form after a few days at 4 °C or room temperature. This problem was later resolved (see Resonance Assignment 4.3). Act-EF1234 was the hardest to handle, as the purification appeared elusive before modification of the purification protocol. Also the yields were not high at 2-3 mg/L - compared to ZASP-PDZ at 9 mg/L from rich media. However, Act-EF1234 was well behaved enough to work with for the given applications.

## Chapter 4

---

# Resonance Assignment: ZASP-PDZ with and without Act-EF1234

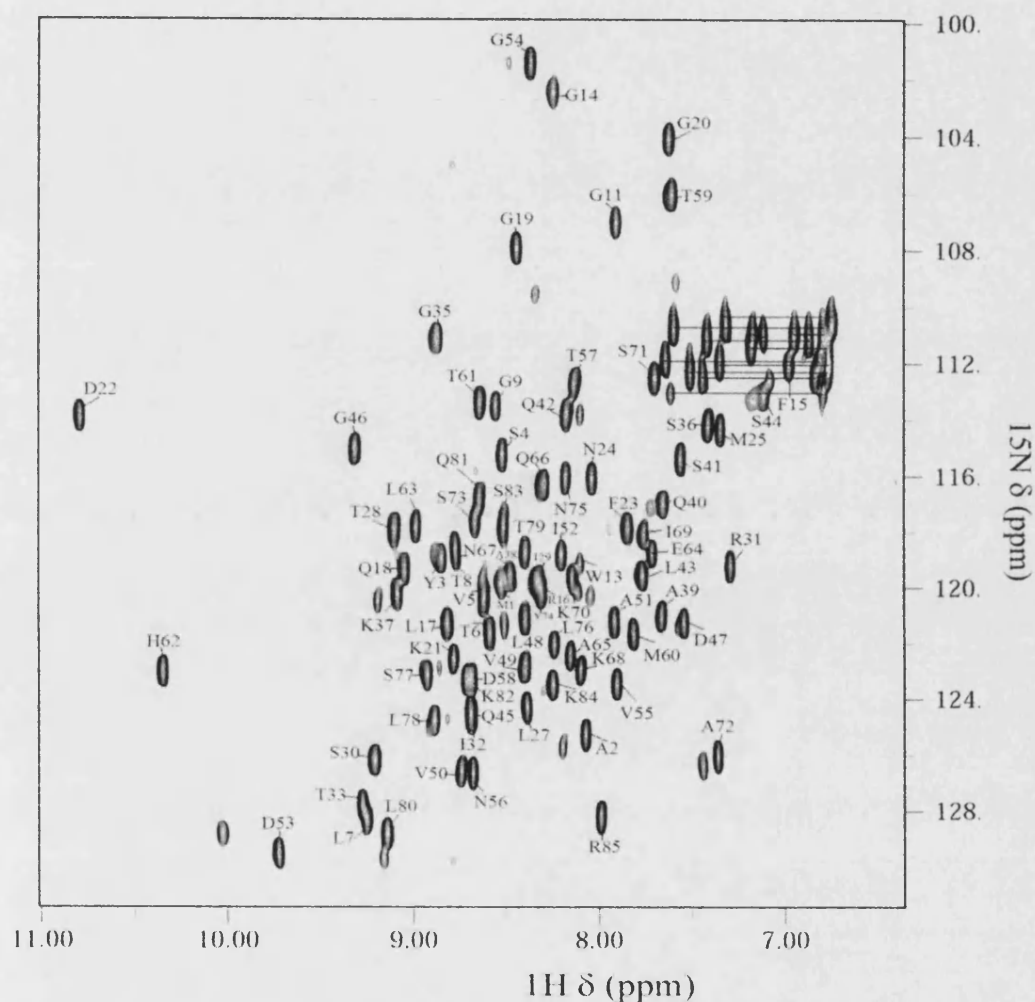
---

Here, the complete  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonance assignments of ZASP-PDZ and the  $^1\text{H}$  and  $^{15}\text{N}$  assignment of ZASP-PDZ domain in complex with Act-EF1234 (forming a 26.3-kDa complex) is reported. Also presented here are the backbone assignments of the 75-residue Act-EF34. The assignment of ZASP-PDZ and Act-EF34 allows the surface of interaction between them to be identified. Also, the ZASP-PDZ assignment is crucial for its structure calculation. Previous assignment of Act-EF34 in complex with titin-ZR7 was available at the beginning of the study (Atkinson et al., 2000b); however, assignment of unbound Act-EF34 was necessary as there is a major conformational rearrangement upon complex formation. By initially assigning the individual components of ZASP-PDZ and Act-EF34, as opposed to the complex, alternative interactions that could arise can be studied in a modular fashion – for example, alternative constructs of  $\alpha$ -actinin-2 (e.g. Act-EF1234) or the binding properties of mutants of either component.

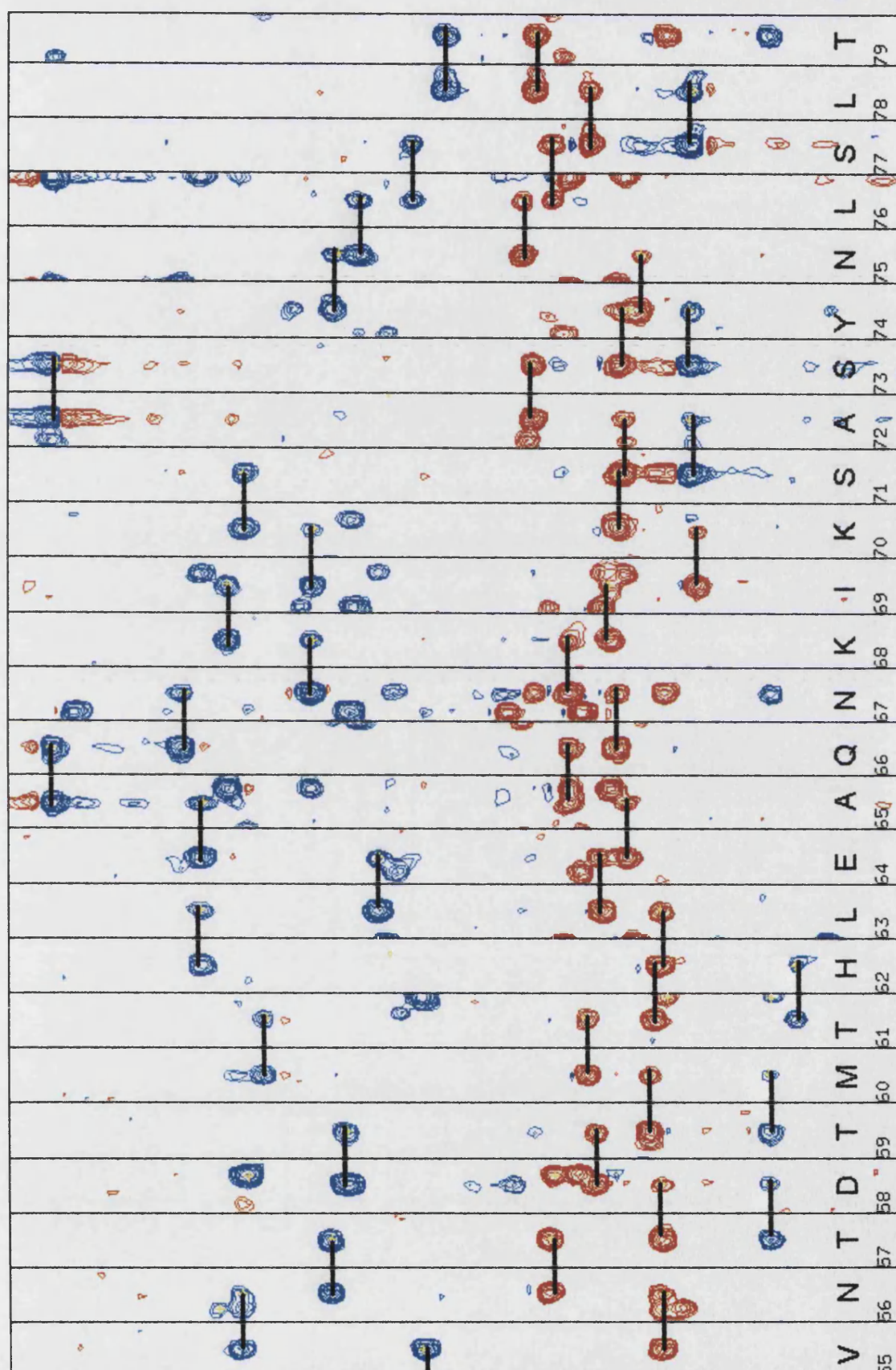
### 4.1 Assignment of ZASP-PDZ

The  $^{15}\text{N}$ -HSQC spectra of the 87-residue ZASP-PDZ displayed 104 peaks, given by the backbone amides from 83 non-proline residues, 10 pairs of glutamine and asparagine side chain amides and one tryptophan indole NH (Figure 4.1). In addition, nine low intensity peaks were observed, derived from a possible minor form of the protein. The  $^{15}\text{N}$ -HSQC spectrum was well dispersed, and the 1D spectrum shows up-field shifted methyl chemical shifts, indicative of a folded protein. Analysis of HNCA, HN(CO)CA, CBCA(CO)NH, and CBCANH (Figure 4.2) spectra allowed the sequential assignment of the ZASP-PDZ, while HNHA and HNCO spectra were used to complete the backbone assignment. A combination of  $^{13}\text{C}$ -HSQC, HCCH-TOCSY, HNHB,  $^{15}\text{N}$ -edited TOCSY, and regular 2D TOCSY spectra were used to assign side-chain  $^1\text{H}$  and  $^{13}\text{C}$  resonances.





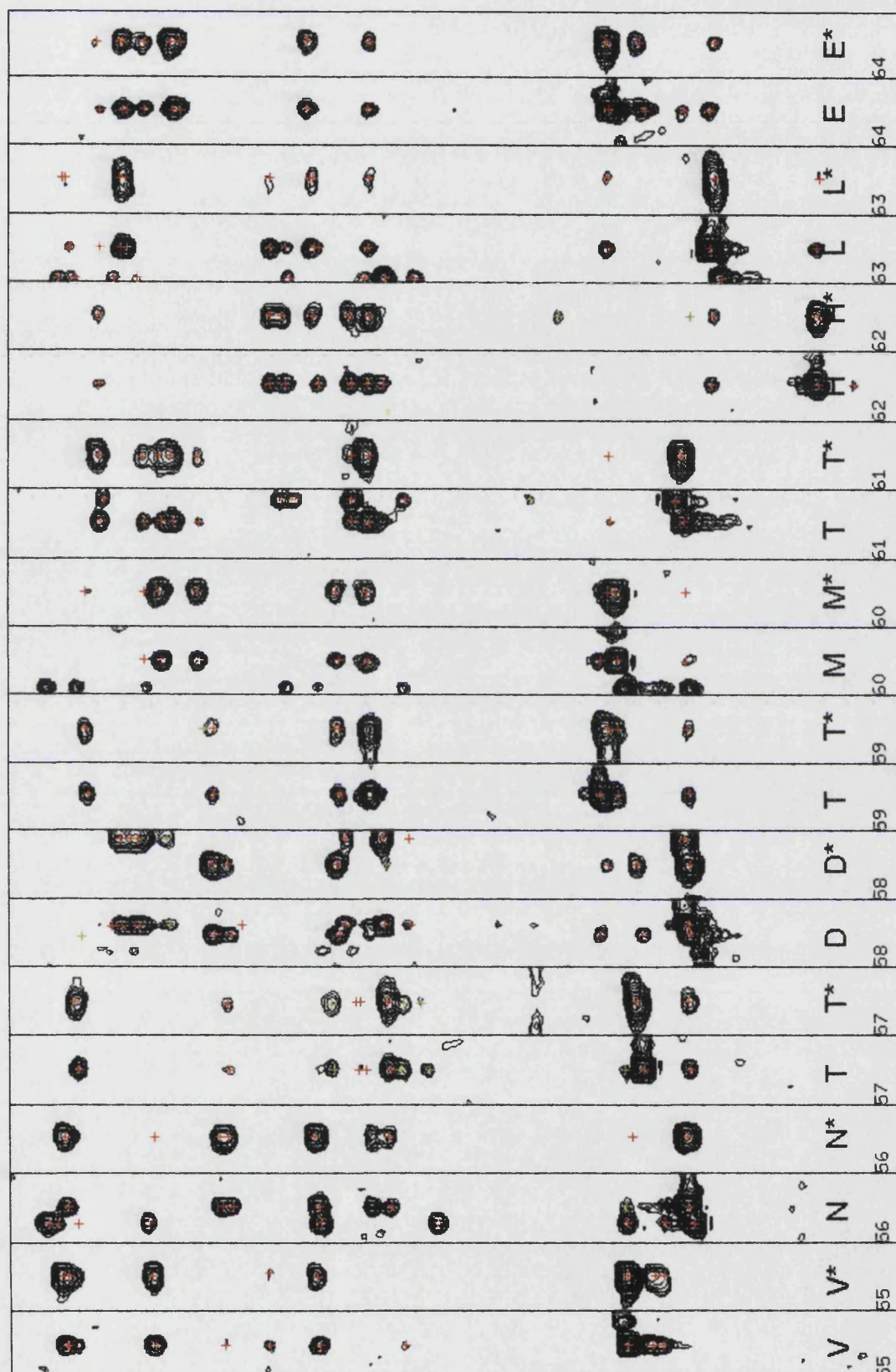
**Figure 4.1:**  $^{15}\text{N}$ -HSQC spectrum of ZASP-PDZ at pH 6.6 and 27 °C, showing a well-resolved spectrum. The spectrum is labelled with amide backbone resonance assignments, and the solid horizontal lines indicate side-chain amide protons of asparagine and glutamine residues.



**Figure 4.2:** Sample of the CBCANH spectra of ZASP-PDZ – red and blue peaks is  $C\alpha$  and  $C\beta$  resonances respectively, and the residue type and number is indicated at the bottom. The black lines highlight backbone connectivity.

## **4.2 Assignment of $^{15}\text{N}$ -ZASP-PDZ bound to Act-EF1234**

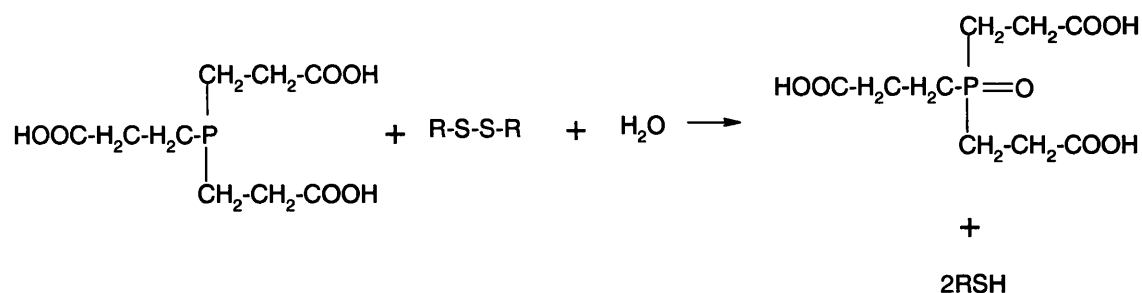
$^{15}\text{N}$ -HSQC spectra were acquired for each point in the titration of  $^{15}\text{N}$ -ZASP-PDZ with unlabelled Act-EF1234 (Section 6.1.2), and  $^{15}\text{N}$ -edited TOCSY and NOESY spectra were subsequently acquired for the complex. The final point of the ZASP-PDZ/Act-EF1234 titration contained 94 peaks, of which 74 were backbone amides resonances and 20 were side-chain peaks. Five of the backbone amides undergo severe exchange broadening during the titration, while others became overlapped (Figure 6.4). Many of the amide  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts could be assigned by following the titration by  $^{15}\text{N}$ -HSQC spectra, and further  $^1\text{H}$  resonances were assigned by comparing  $^{15}\text{N}$ -edited TOCSY and NOESY spectra of the complex with those of unbound ZASP-PDZ (Figure 4.3).



**Figure 4.3:** Sample of the  $^{15}\text{N}$ -NOESY-HSQC comparison of the unbound and Act-EF1234 bound forms of  $^{15}\text{N}$ -ZASP-PDZ - the residue type and number is indicated at the bottom. \* Denotes the bound form of ZASP, while red and green peaks denote assigned and unassigned peaks respectively.

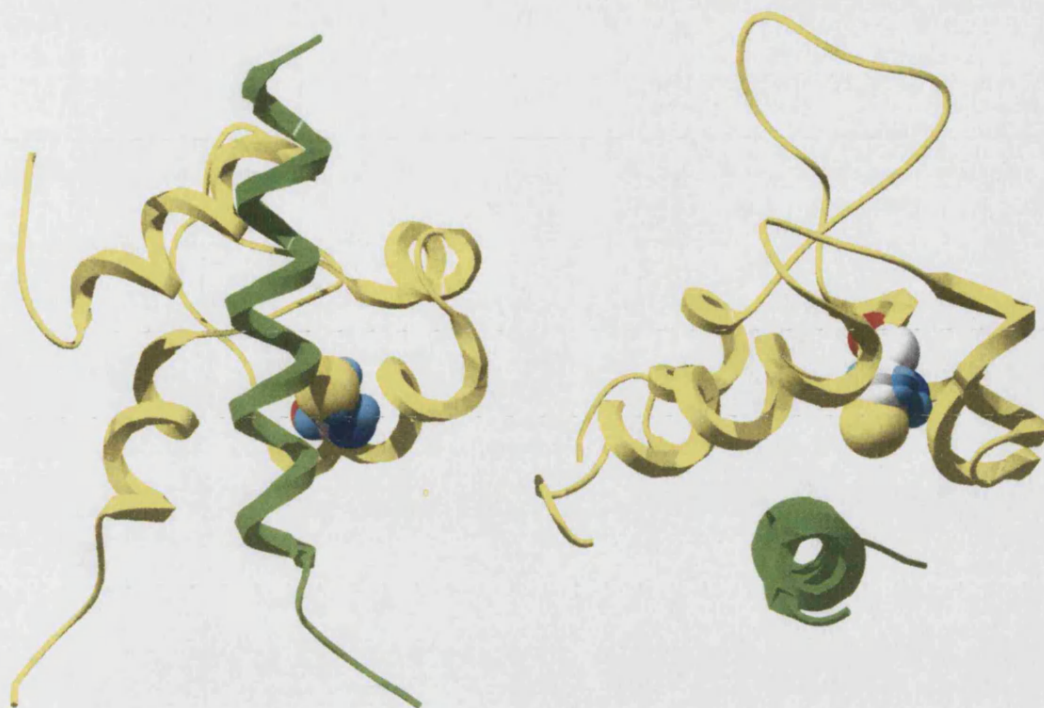
**4.3 Assignment of Act-EF34**

The backbone Act-EF34 atoms were assigned in the same manner as ZASP-PDZ, using a combination of  $^{15}\text{N}$ -HSQC, CBCACONH, CBCANH, HNHA, and HNHB spectra (Figure 4.10). The  $^{15}\text{N}$ -HSQC revealed that a second species appears after a few hours at room temperature, resulting in approximately 32 additional peaks to the 77 expected (from 69 non-proline amides, plus 4 pairs of glutamine side chain amide protons) (Figure 4.7). The secondary species appear to derive from homo-dimer formation from the lone Act-EF34 cysteine residue (C41) forming an intermolecular disulphide bond. Although 2 mM  $\beta$ -mercaptoethanol was present in the buffer to prevent disulphide bond formation, this only appeared to be only partially effective. The CBCACONH/CBCANH spectra revealed 'doubled' backbone amides - these were T4, T6, G56, A66, L67, Y68, and G69. The second species was not observed once Act-EF34 was in complex with ZR7, which could be due to burial of the Act-EF34's cysteine residue (Figure 4.5). A non-reducing gel revealed a peak around 16 kDa, double Act-EF34 alone (Figure 4.6), so triakylphosphine (TCEP, Figure 4.4) added to the sample to reduce the disulphide bond, returning the dimer to a monomeric species. The effectiveness of the treatment is apparent by the improvement of the relative intensities of the monomeric species peaks in the  $^{15}\text{N}$ -HSQC (Figure 4.8 and Figure 4.9).

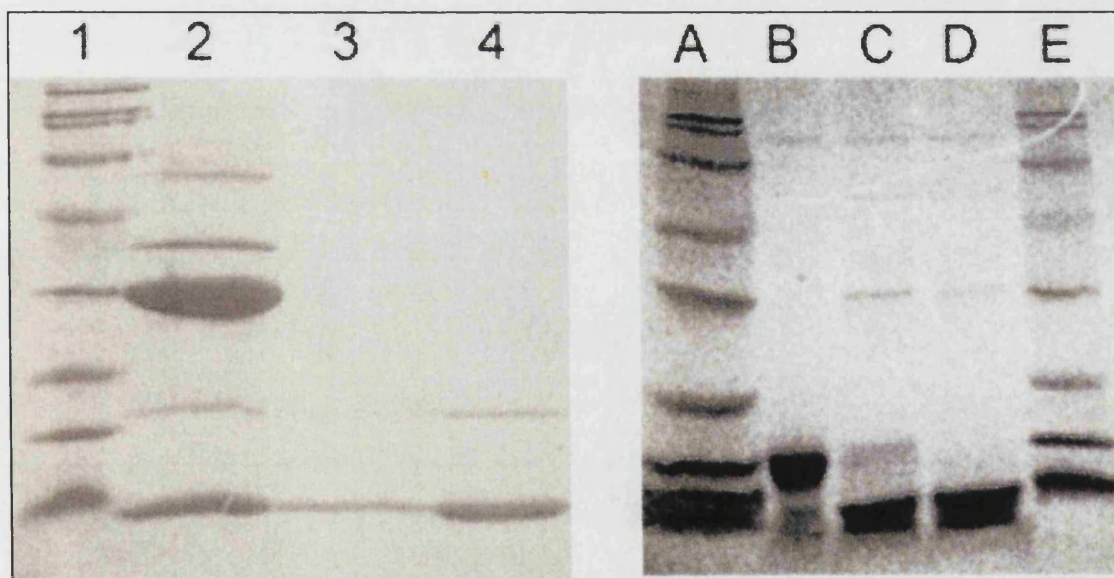


**Figure 4.4:** Schematic of the mechanism of TCEP reduction of disulphide bonds. TCEP reacts with water to reduce the disulphide bonds, resulting in an oxidised phosphorus atom

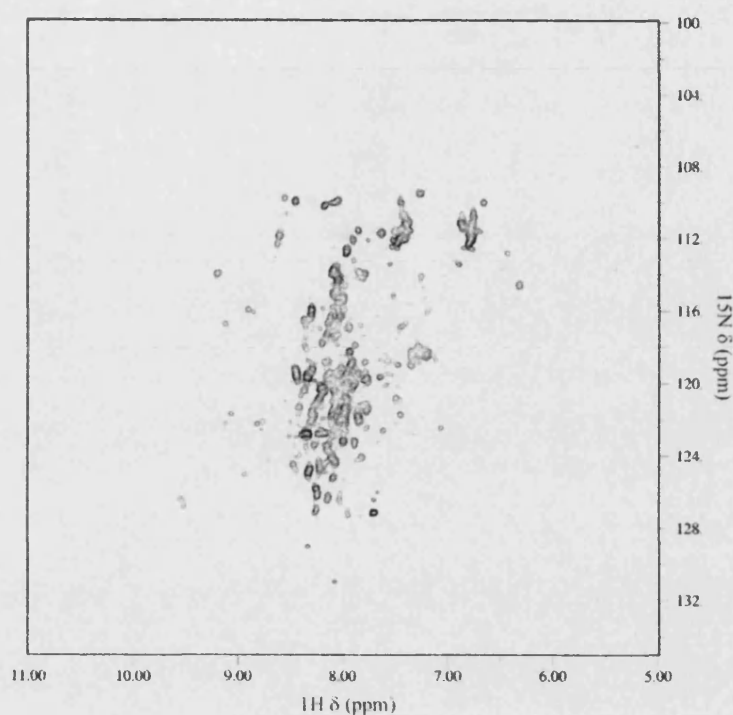




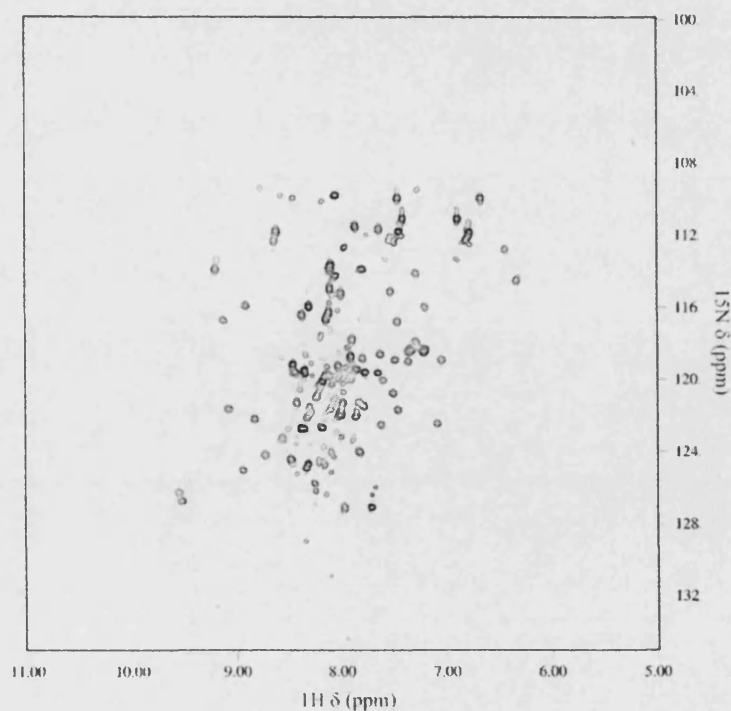
**Figure 4.5:** Ribbon representation of Act-EF34 (yellow) in complex with titin-ZR7 (green), indicating the position of cysteine 41 (van der Waals representation). Titin-ZR7 may prevent Act-EF34 homodimer formation, as it buries the thiol group between the surface of interaction.



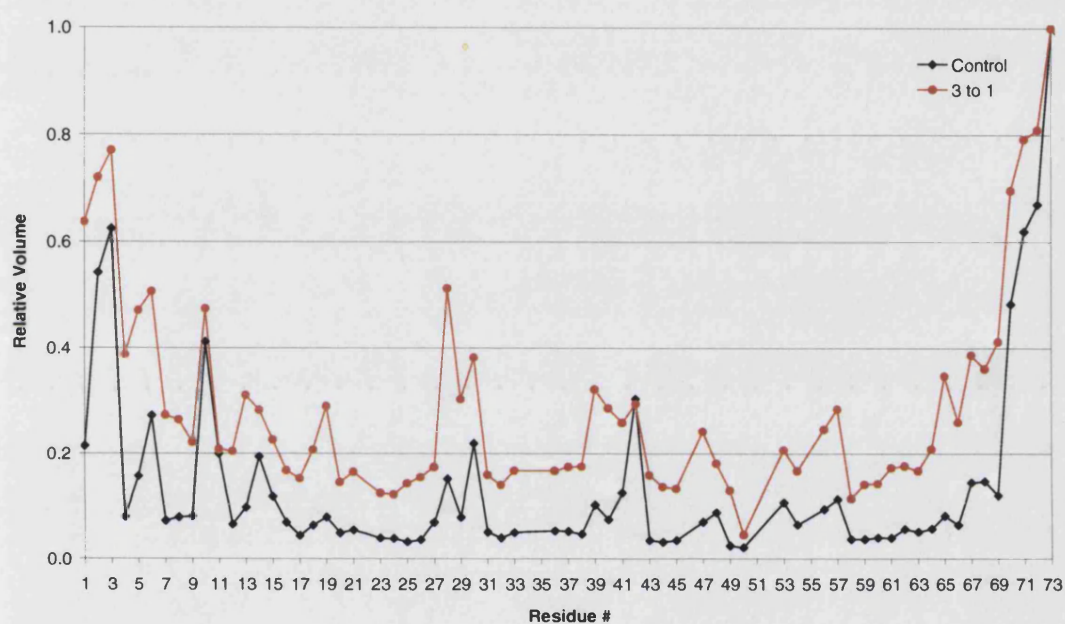
**Figure 4.6:** Polyacrylamide SDS gel electrophoresis of A: Sample of freshly prepared Act-EF34 revealing a trace of the dimeric second species (4) using the Broad Molecular Marker (1) for size estimation. B: Non-reducing polyacrylamide SDS gel of an old sample of Act-EF34 (B) in the presence of 3:1 (C) and 10:1 TCEP (D) to protein ratios.



**Figure 4.7:**  $^{15}\text{N}$ -HSQC of 'fresh' Act-EF34 without TCEP. The protein appears unfolded or aggregated, as the peaks are 'clumped' in the centre of the spectrum.

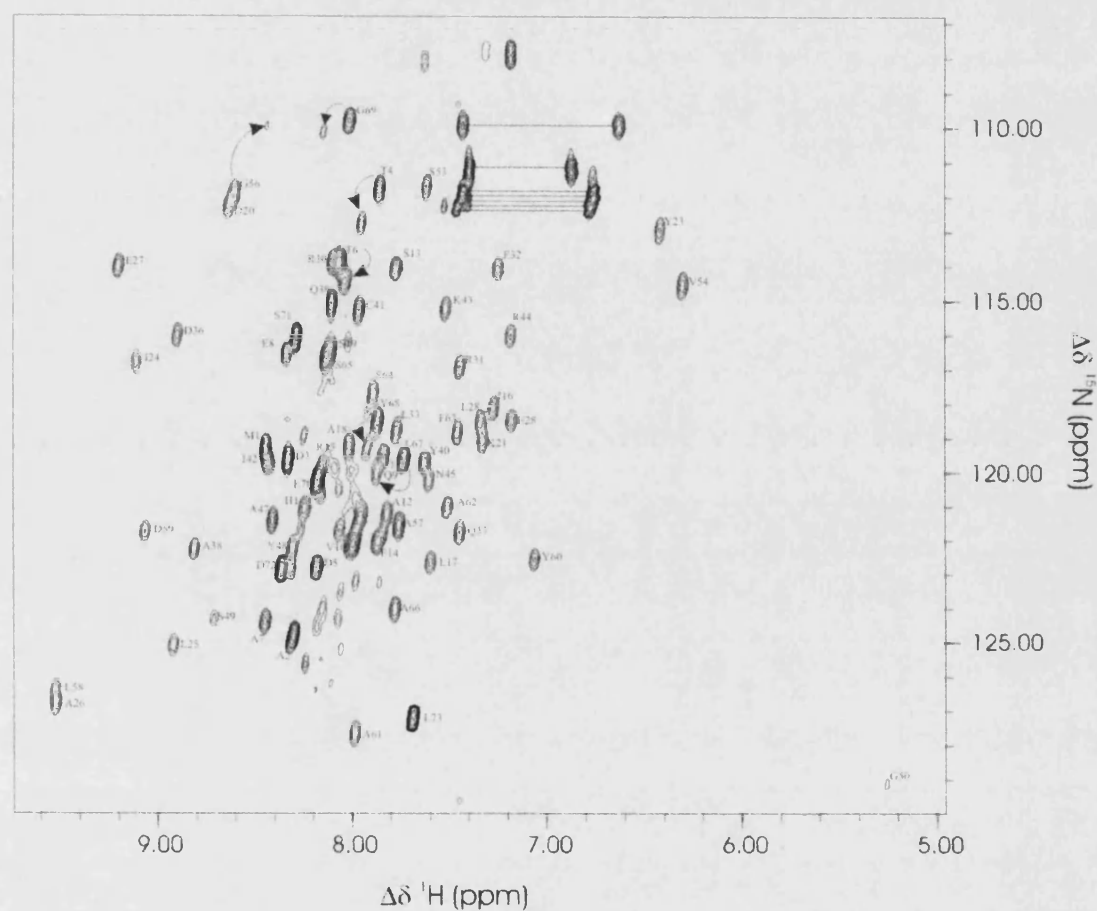


**Figure 4.8:**  $^{15}\text{N}$ -HSQC of Act-EF34 with 1:3 ratio of TCEP. The relative intensities of the peaks have increased, and some of the peaks from the second species have disappeared.



**Figure 4.9:** Graph showing the change in relative intensity (to the C-terminal residue) on addition of a molar ratio of 3:1 TCEP to Act-EF34. This indicates the formation of the monomeric species from the disulphide-bonded homodimer: line-widths are narrower for smaller molecular species, hence, an increase in intensity indicates a reduction in size.

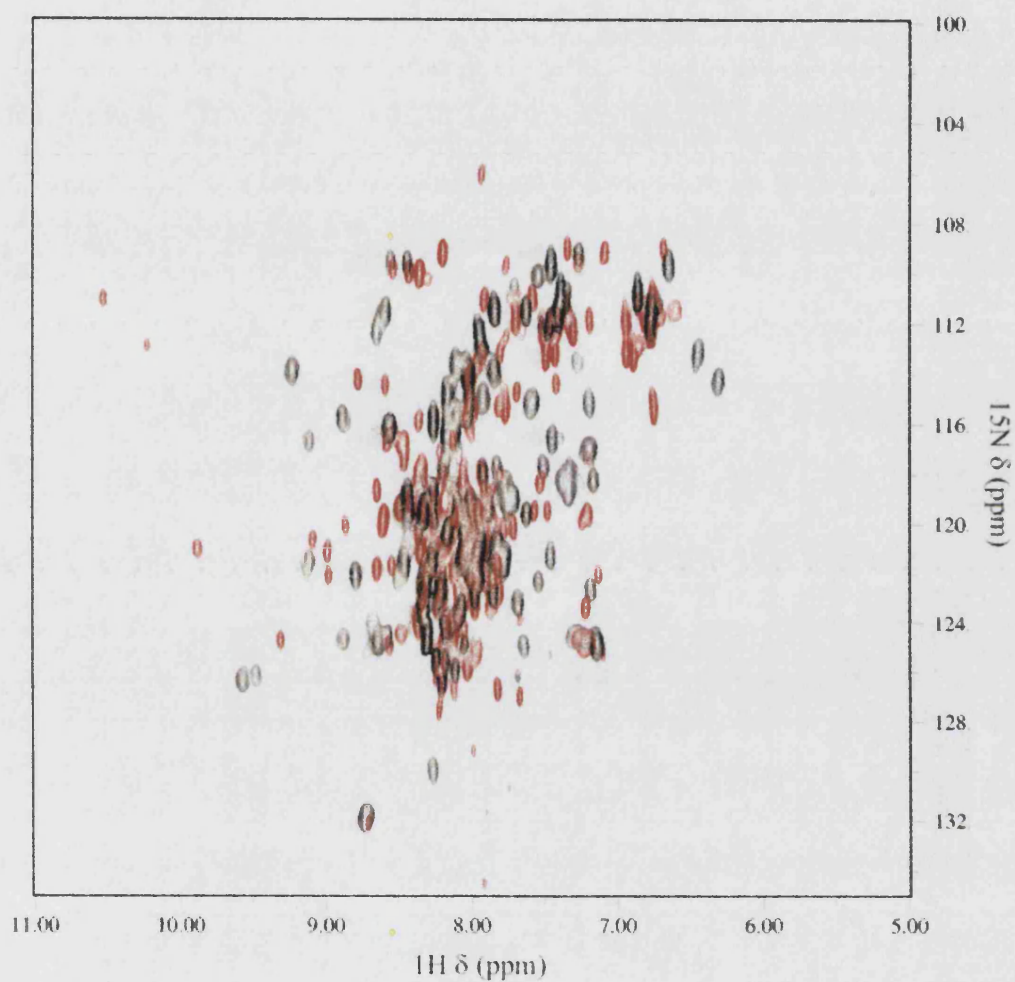




**Figure 4.10:**  $^{15}\text{N}$ -HSQC spectrum of Act-EF34 at pH 6.6 and 27 °C, showing amide backbone resonance assignments. Solid horizontal lines indicate side-chain amide protons of asparagine and glutamine residues, and the arrows indicate assignment of 'second species' peaks.

#### **4.4 Assignment of Act-EF1234**

The assignment of Act-EF1234 alone is unavailable, so an attempt to transfer Act-EF34's assignment by superimposing Act-EF1234 and Act-EF34  $^{15}\text{N}$ -HSQC spectra was made (Figure 4.11). This revealed a conformational difference between them, as many peaks could not be superimposed – one would expect the additional residues between Act-EF1234 to have a simple additive contribution if they were congruous. However, the peaks from Act-EF1234 and Act-EF34 C-terminal residues (specifically G69, S71, D72, and L73 – Act-EF34 nomenclature) were similar enough to be assigned by superimposition, possibly reflecting the lack of structure observed in this region (Atkinson et al., 2001). Although the superimposition revealed this conformational difference, a previous study (Joseph et al., 2001) showed that both constructs converge to the same conformation upon binding to a peptide derived from titin seventh Z-repeat (ZR7), therefore, the two constructs of  $\alpha$ -actinin-2 appear to have the same functional properties.



**Figure 4.11:**  $^{15}\text{N}$ -HSQC superimposition of ZASP-PDZ bound to  $^{15}\text{N}$ -Act-EF34 (black peaks) and  $^{15}\text{N}$ -Act-EF1234 (red peaks). While some peaks are super-imposable, most are not. See (Joseph et al., 2001) for both Act-EF34 and Act-EF1234 converging upon binding titin-ZR7.

## **4.5 Conclusion**

The assignment of the  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances were obtained for the unbound ZASP-PDZ domain. The backbone atoms were fully assigned, with the exception of the backbone carbonyl resonances that are followed by a proline. All side chain resonances were assigned, except for the  $^1\text{H}$  resonances of the N-terminal methionine,  $^{13}\text{C}$  resonances of the aromatic side-chains past their  $\text{C}\beta$   $^{13}\text{C}$  resonances for Leu-7 and His-62 past their  $\text{C}\beta$   $^{13}\text{C}$  and  $^1\text{H}$  resonances for Pro-12, and most exchangeable side-chain protons.

The  $^{15}\text{N}$ -HSQC titration of the ZASP-PDZ/Act-EF1234 complex facilitated the transfer of 94% of the backbone amide assignments. This allowed further assignment of  $^1\text{H}$  resonances by comparing the  $^{15}\text{N}$ -edited TOCSY and NOESY spectra with those of unbound ZASP-PDZ. Residues 13-18 appear to be in chemical exchange upon complex formation; although it was possible to follow their amide resonances in the titration, their side-chain  $^1\text{H}$  resonances could not be observed. Proline residues 10, 12, 26 and 34 and other side-chain assignments could not be completed, leading to the  $^1\text{H}$  and  $^{15}\text{N}$  assignment to be 76% of that of unbound ZASP-PDZ. The assignments have been deposited with the BioMagResBank (<http://www.bmrc.wisc.edu>) with the accession codes BMRB-5696 (unbound ZASP-PDZ) and BMRB-5697 (ZASP-PDZ/Act-EF1234).

The backbone assignment of all of the backbone atoms of Act-EF34 was completed with the exception of the carbonyls, the first two N-terminal residues, and G52. Also, some backbone amide atoms from Act-EF1234 were assigned by comparing the  $^{15}\text{N}$ -HSQC spectra of assigned Act-EF34.

## Chapter 5

---

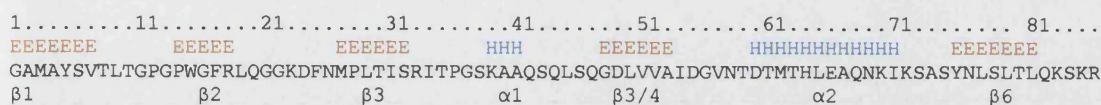
# NMR Structure of ZASP-PDZ

---

Both manual and automatic procedures have been performed for the structure determination of ZASP-PDZ. Manual structure calculation was carried out using NOE constraints derived from  $^1\text{H}$ - $^1\text{H}$ ,  $^1\text{H}$ - $^{13}\text{C}$ , and  $^1\text{H}$ - $^{15}\text{N}$  NOESY spectra, directly observed hydrogen bonds (Cordier and Grzesiek, 1999), TALOS (Cornilescu et al., 1999) and HNHA dihedral angles, and residual dipolar couplings (RDCs). A homology model and secondary structure prediction of ZASP-PDZ were utilised as an initial guide for manual structure determination. CANDID was employed for automated structure calculation, using unassigned peak lists, the chemical shift assignment (Appendix 10.4.1), and dihedral angle and hydrogen bond information defined in the manual calculation. The structure calculation was also undertaken in both CYANA and XPLOR – allowing us to benefit from the speed and accessibility of CYANA, and also the extra functionality of XPLOR (such as the ability to use RDCs). Structural restraints deriving from both manual and automatic structure calculations were exported to XPLOR after convergence within CYANA for re-calculation and subsequent improvement using the ‘CHARMM22 water-refinement’. The quality of the calculated structures was checked using WHATIF (Hooft et al., 1996) and PROCHECK-NMR (Laskowski et al., 1996), while checking them against known PDZ domains assessed the accuracy of the structures. The bundles deriving from different structure calculation protocols were compared and assessed to determine the most effective calculation procedure.

### 5.1 Secondary Structure Prediction of ZASP-PDZ

The secondary structure prediction of ZASP-PDZ reveals six  $\beta$ -strands and two  $\alpha$ -helices, typical of the PDZ domain (Figure 5.1). Although the prediction shows five areas of  $\beta$ -strand, the PDZ domain actually contains six, as the forth predicted  $\beta$ -strand is actually a continuous  $\beta$ -hairpin.



**Figure 5.1:** Secondary structure prediction of ZASP-PDZ domain. H represents  $\alpha$ -helices, and E represents  $\beta$ -strands.

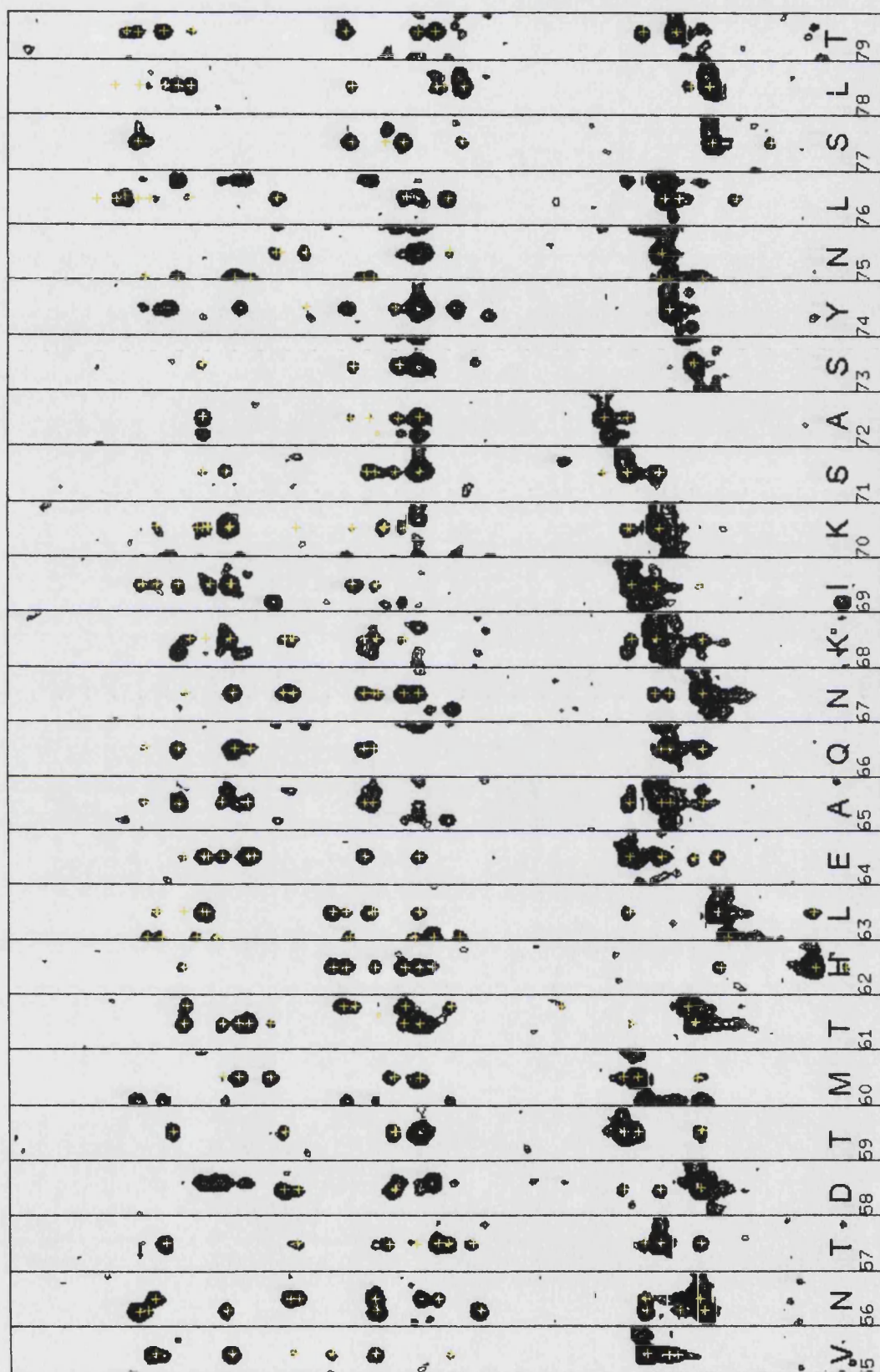
## 5.2 Manual Structure Determination

The manual structure determination of ZASP-PDZ was calculated using CYANA (CMANUAL – Appendix 10.5). The NOESY spectrum had excellent dispersion and resolution (Figure 5.2), allowing for the near to complete assignment (Table 5.1). A total of 1403 unique distances were manually assigned (Table 5.2). These were subsequently modified by CYANA to remove fixed intra-residue distances, and short/medium-range distances above the maximum possible distance. Further alterations of the distance restraints were made by CYANA, by the stereo-specific assignments made by GLOMSA. Additional restraints from the hydrogen bond experiment, and the TALOS/HNHA dihedral angles were also used for the structure calculation. Fifty percent (out of 100 structures) of the CMANUAL structures converged with no upper, lower, van der Waals, or torsion angle violations (0.5 Å, 0.5 Å, 0.5 Å, 5 degrees, respectively). The restraints were subsequently converted to XPLOR format and the structures were recalculated (XMANUAL), after which, twenty of the best of these structures were selected for ‘CHARMM22 water-refinement’ (RXMANUAL).

Spectra	Assigned Peaks	Unassigned Peaks	Total Peaks	% Assigned
$^{15}\text{N}$ -NOESY-HSQC	705	57	762	92.5
$^{13}\text{C}$ -NOESY-HSQC	1402	167	1569	89.4
$^1\text{H}$ - $^1\text{H}$ -NOESY	2180	476	2656	82.1

**Table 5.1:** Summary of the analysis of the  $^1\text{H}$ - $^1\text{H}$ -NOESY, and  $^{13}\text{C}$  and  $^{15}\text{N}$  NOESY-HSQC spectra: i.e. number of peaks picked, number of assignments made, and percentage assigned.





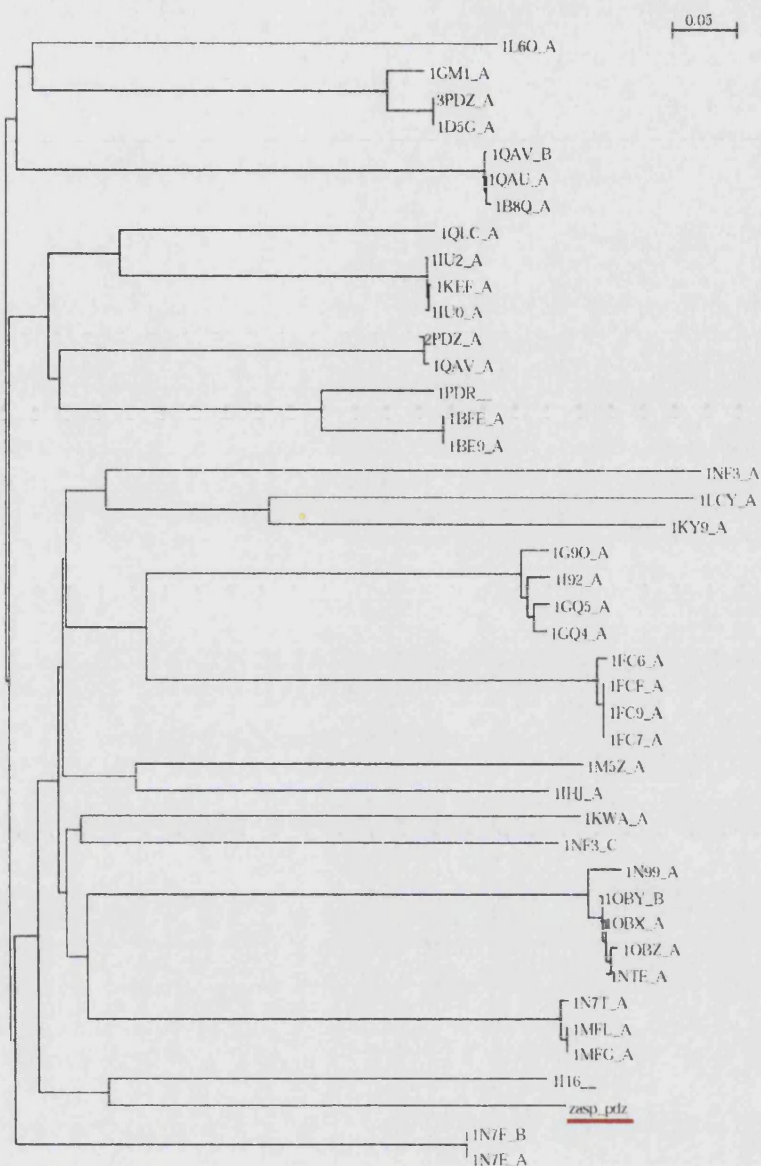
**Figure 5.2:** Extract of the  $^{15}\text{N}$ -NOESY HSQC of  $^{15}\text{N}$ -ZASP-PDZ. The  $^{15}\text{N}$ -HSQC (Figure 4.1 –  $^{15}\text{N}$ -HSQC) showed good dispersion of the peaks, therefore the  $^{15}\text{N}$ -NOESY HSQC was also well resolved. The residue type and number is indicated at the bottom of each strip.

Distance Type	Intra	Short	Med	Long	Total
Unique Distances	722	320	134	227	1403
Modified Distances	216	285	136	246	883
Final Restraints	219	292	143	241	895

**Table 5.2:** Summary of the final distance restraints used for the CMANUAL structure calculation. Unique distances are those present after combining the restraints from all of the NOESY spectra, while modified distances are superfluous restraints. The final restraints are the modified restraints after taking into account stereo-specific assignment.

A homology model of ZASP-PDZ was built as a guide for the NMR structure calculation. The PDB (<http://www.rcsb.org>) was searched for PDZ domains and their protein sequences were subsequently aligned against the ZASP-PDZ domain (Appendix 10.7). A CLUSTALX (Thompson et al., 1997) sequence alignment of the PDZ domain sequences shows that interleukin 16 (Muhlhahn et al., 1998) (PDB accession code IL16) is the closest PDZ domain (Figure 5.3). However, after taking into consideration gaps and insertions, the third PDZ domain from PSD95 (Doyle et al., 1996) (PSD95-PDZ3, PDB accession code 1BFE) was chosen as ZASP-PDZ's template for the homology model. Pair-wise comparisons of the sequences of PSD95-PDZ3 and ZASP-PDZ (Figure 5.4) shows no gaps or insertions after the sixth residue on the ZASP-PDZ construct, and a 22.5% sequence identity and 47.5% sequence homology in the remaining 80 residues of the two proteins. The homology model was generated using SWISS-MODEL (Guex and Peitsch, 1997; Peitsch, 1995; Schwede et al., 2003) (Figure 5.5). PSD95-PDZ3 was also a useful choice as it had been solved in complex with its target peptide derived from Cript (PDB Accession code: 1BE9), which was useful as a guide to study the ZASP/ $\alpha$ -actinin-2 interaction.

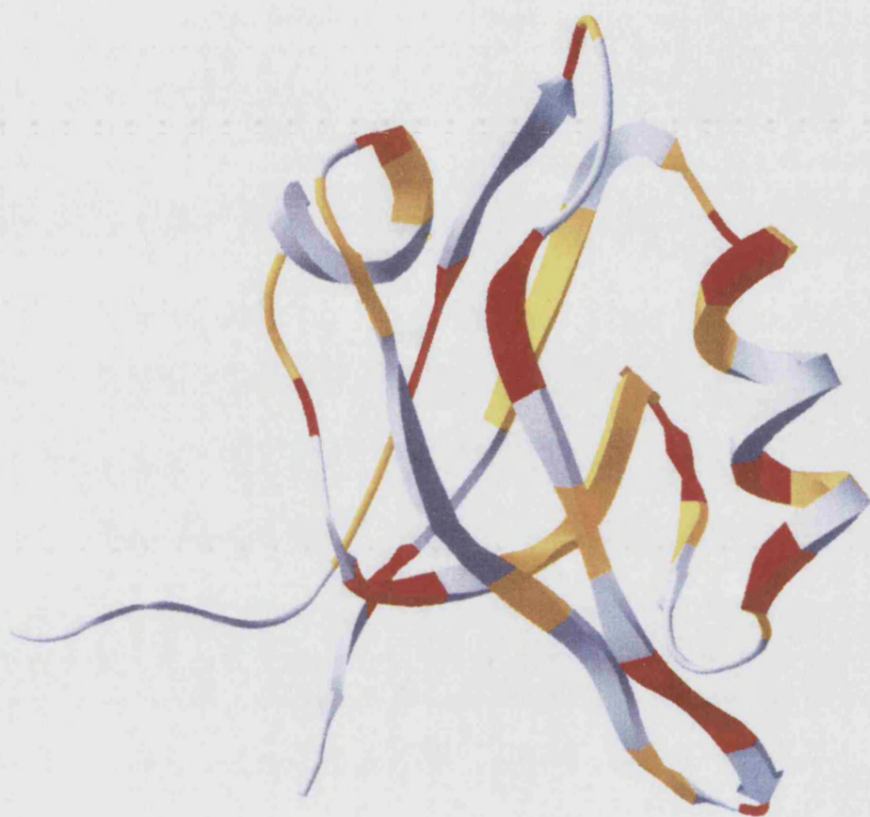




**Figure 5.3:** Phylogenetic tree derived from a CLUSTALX alignment of all solved PDZ domains in the PDB, generated using NJplot (Perriere and Gouy, 1996). Although IL16 was the closest relation to ZASP-PDZ, after taking into consideration gaps and insertions, 1BFE/1BE9 was a better template for ZASP-PDZ's homology model.



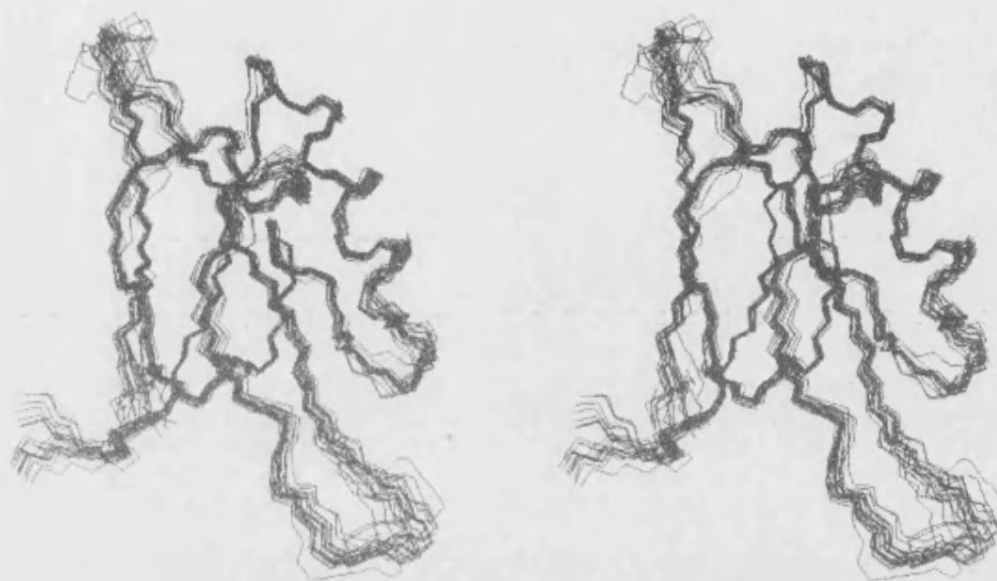
**Figure 5.4:** Pair-wise CLUSTALX alignment of ZASP-PDZ and 1BE9/1BFE, showing a 22.5% sequence identity and 47.5% sequence homology.



**Figure 5.5:** Homology model of ZASP-PDZ, generated by SWISS-MODEL, based on the alignment with 1BFE – residues coloured in red, orange, and blue represent identical, chemically similar, and no similarity respectively.

### **5.3 Automatic Structure Determination**

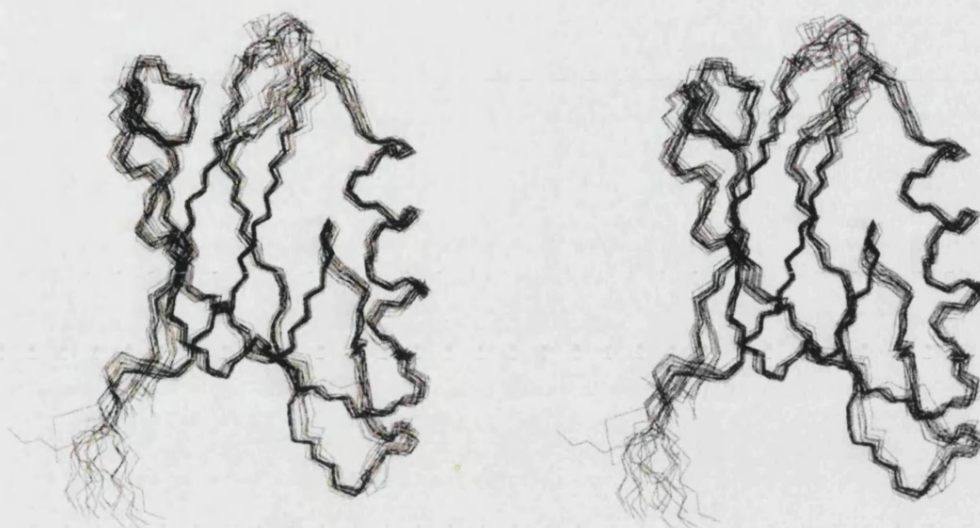
The  $^1\text{H}$ - $^1\text{H}$ -NOESY spectra was utilised for the automatic structure determination using CANDID as it has the most self-consistent peak and chemical shift lists; CANDID makes possible assignments based on chemical shift tolerances, therefore it operates best when there is no deviation between chemical shift and peak lists. The three-dimensional NOESY spectra had lower resolution than the two-dimensional NOESY, leading to less precision in chemical shift: peaks in the manually assigned  $^1\text{H}$ - $^1\text{H}$ -NOESY could be picked within an error of  $\pm 0.01$  ppm, but only  $\pm 0.02$  ppm for the three-dimensional NOESY. This allowed a lower chemical shift tolerance to be used for the CANDID calculation for the homonuclear peak list: tolerances of 0.03, 0.03, 0.04 ppm ( $^1\text{H}$  (direct),  $^1\text{H}$  (indirect),  $^{13}\text{C}/^{15}\text{N}$  (indirect)) was used for the 3D NOESY peak lists, while 0.02 in both direct and indirect dimensions could be used for the 2D NOESY peak list: an additional 0.01 ppm was used to take into account errors in NOESY peak picking. One notable limitation in CANDID was that for correct calculation of ZASP-PDZ, the dihedral angles and hydrogen bonds were needed; to circumvent the problem we added to the hydrogen bonds observed in the HNCO and those inferred from the MANUAL bundle. If these restraints were excluded from the calculation, the algorithm would incorrectly calculate the structure (Figure 5.6) –  $\beta 1$ ,  $\beta 2$ , and  $\beta 3$  have an inconsistent orientation compared with the standard PDZ fold, and the first alpha helix was not correctly formed.



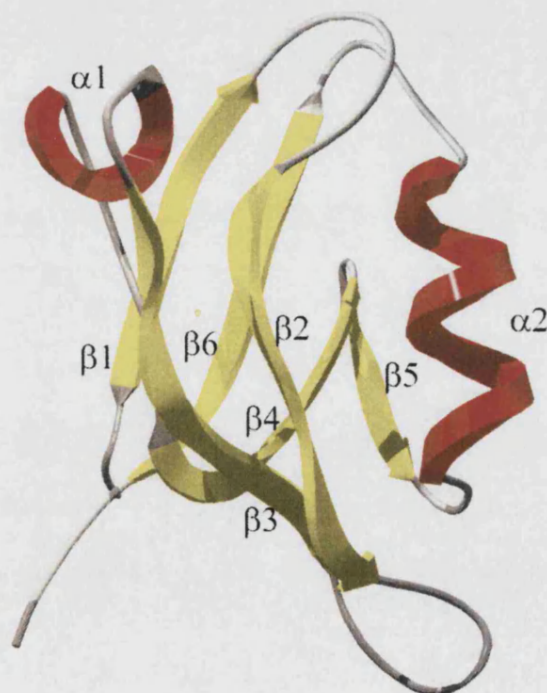
**Figure 5.6:** Stereogram of the bundle of the best 20 structures from the ZASP-PDZ CANDID automatic structure determination using just NOE distance constraints. The structure is incorrectly folded and parts of the secondary structure have not been formed .

#### **5.4 Structure Description**

The structure of ZASP-PDZ adopts a conventional PDZ fold: it is a mixed  $\alpha/\beta$  protein that folds into a six-stranded  $\beta$ -sandwich flanked by and two  $\alpha$ -helices on opposite sides (Figure 5.7). The topology of the anti-parallel  $\beta$  strands is clearly visible (Figure 5.8). The first and last  $\beta$  strands ( $\beta 1$  and  $\beta 6$ ) form an anti-parallel  $\beta$ -sheet, on the opposite side of the  $\beta 2$ - $\beta 3$  anti-parallel  $\beta$ -sheet and  $\alpha 2$ , with  $\beta 4$  going across these structural elements. The groove between  $\beta 2$  and  $\alpha 2$  which starts from the 'GLGF' loop (residues P9, W10, G11 and F12 in ZASP-PDZ) and ends at the C-terminal portion of  $\beta 2$  is the conventional site of interaction for class 1 PDZ binding partners (Figure 7.2). In ZASP-PDZ, this pocket is delimited by the unique tryptophan at the N-terminal end of  $\beta 2$ .



**Figure 5.7:** A backbone ensemble stereogram of the structure of ZASP-PDZ (CCANDID).



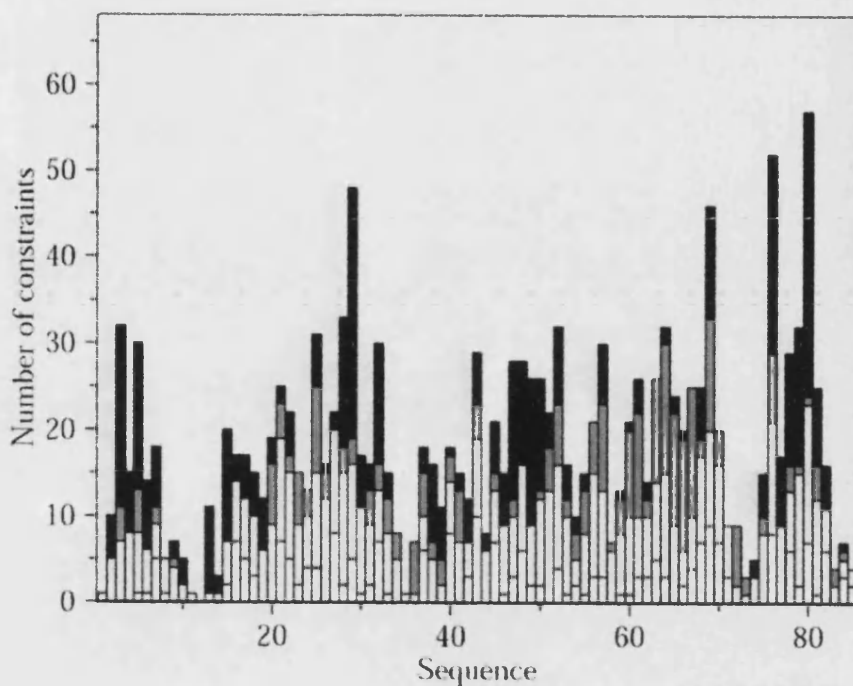
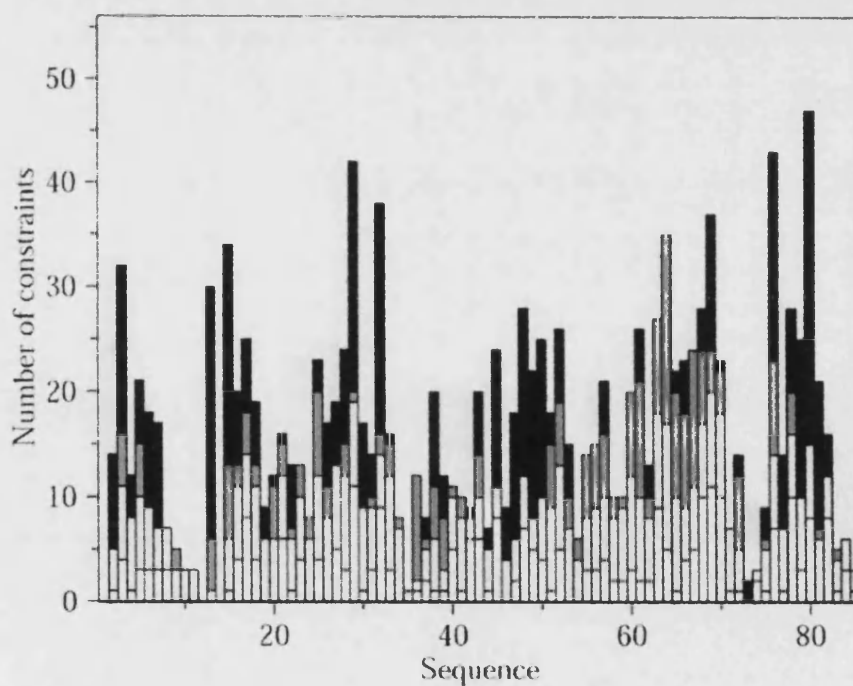
**Figure 5.8:** Ribbon diagram depicting secondary structure elements of ZASP-PDZ (red =  $\alpha$  helices, yellow =  $\beta$  strands) (CCANDID). ZASP-PDZ has 6  $\beta$ -strands and 2  $\alpha$ -helices forming a  $\beta$ -sandwich.



**5.5 Restraint Comparison: Manual vs. Automatic**

A comparison between the NOE assignments used in the structure calculation between the manual and automatic procedure was conducted. Table 5.3 (also Figure 5.9) shows that there were more manual NOE assignments made than the automatic CANDID assignment. The differences in the restraints used in the manual and CANDID calculations were compared on a 'residue-by-residue connectivity' basis: this revealed that CANDID made 52 additional peak assignments – for example, if there were five connectivities between residues 'a' and 'b' made in the automatic assignment, this would be five additional peak assignments more than the manual assignment. Analysis of these peaks in the manually assigned peak list showed that 33 were unassigned, 6 were discarded during the structure calculation but previously assigned the same, and 13 were assigned differently. Of the 13 differently assigned peaks, 8 were short-range, and 5 were long-range: one appeared to be miss-assigned in the manual structure calculation and the four were as good in either structure calculation. Conversely, there were 84 restraints present in the manual restraint list additional to CANDID's, of which 33 were short-range, 22 were additional from  $^{13}\text{C}/^{15}\text{N}$ -NOESY spectra, 18 were unassigned in CANDID, and 11 were assigned differently. Of the 11 differently assigned, all were structurally possible, comprising of short-range and 6 long-range restraints.

There were 290 peaks that CANDID did not assign: 88% were not assigned in the manual assignment and 12% had assignments in the manual assignment. Also, the general distribution of the type of restraint reveals that although CANDID makes more intra-residue assignments (+6.6%) than the manual calculation, but less short-range assignments (-6.0%), there is no significant difference between the distribution of either the medium (-1.2%) or long-range constraints (+0.62). CANDID's method of dealing with stereo-pairs and distance violations appears to result in the average distance being smaller than the manual calculation:  $3.91 \pm 0.57 \text{ \AA}$ , compared with  $5.4 \pm 1.64 \text{ \AA}$ . This will lead to more precise structures, even though CANDID has 5% fewer restraints compared to the manually assigned structure.

**A****B**

**Figure 5.9:** Distance constraint plot of the A) manual and the B) CANDID structure calculation. White, light grey, dark grey, and black colours represent intra-residue, short-range, medium-range, and long-range distance connectivities respectively.

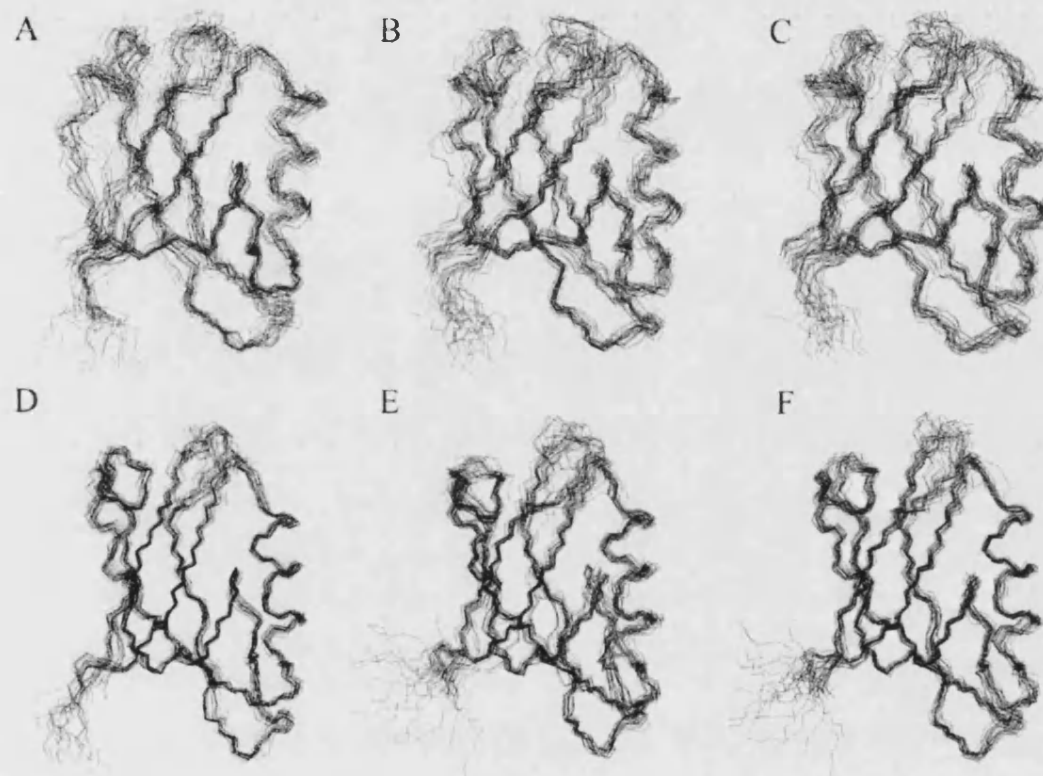
	CANDID	MANUAL
<b>Peak Assignment</b>		
Total Peaks	2655	2655
Without assignment	290	475
With assignment	2365	2180
Restraints not included in opposite restraint file (residue by residue comparison)	52	84
<b>NOESY Distance Restraints</b>		
Intra residue	265 (31.1%)	219 (24.5%)
Short range	227 (26.6%)	292 (32.6%)
Medium range	126 (14.8%)	143 (16.0%)
Long range	235 (27.5%)	241 (26.9%)
<b>Angular restraints</b>		
$\phi$	51	51
$\psi$	46	46
<b>H bonds</b>	30	7
<b>Total No of restraints</b>	980	998
<b>RDCs</b> <sup>*</sup>	63	63
<b>Stereo specific assignments</b>	25	25

**Table 5.3:** Comparison of the manual and automatic assignment of ZASP-PDZ. <sup>\*</sup> Only added to the XCANDID and XMANUAL refinements.



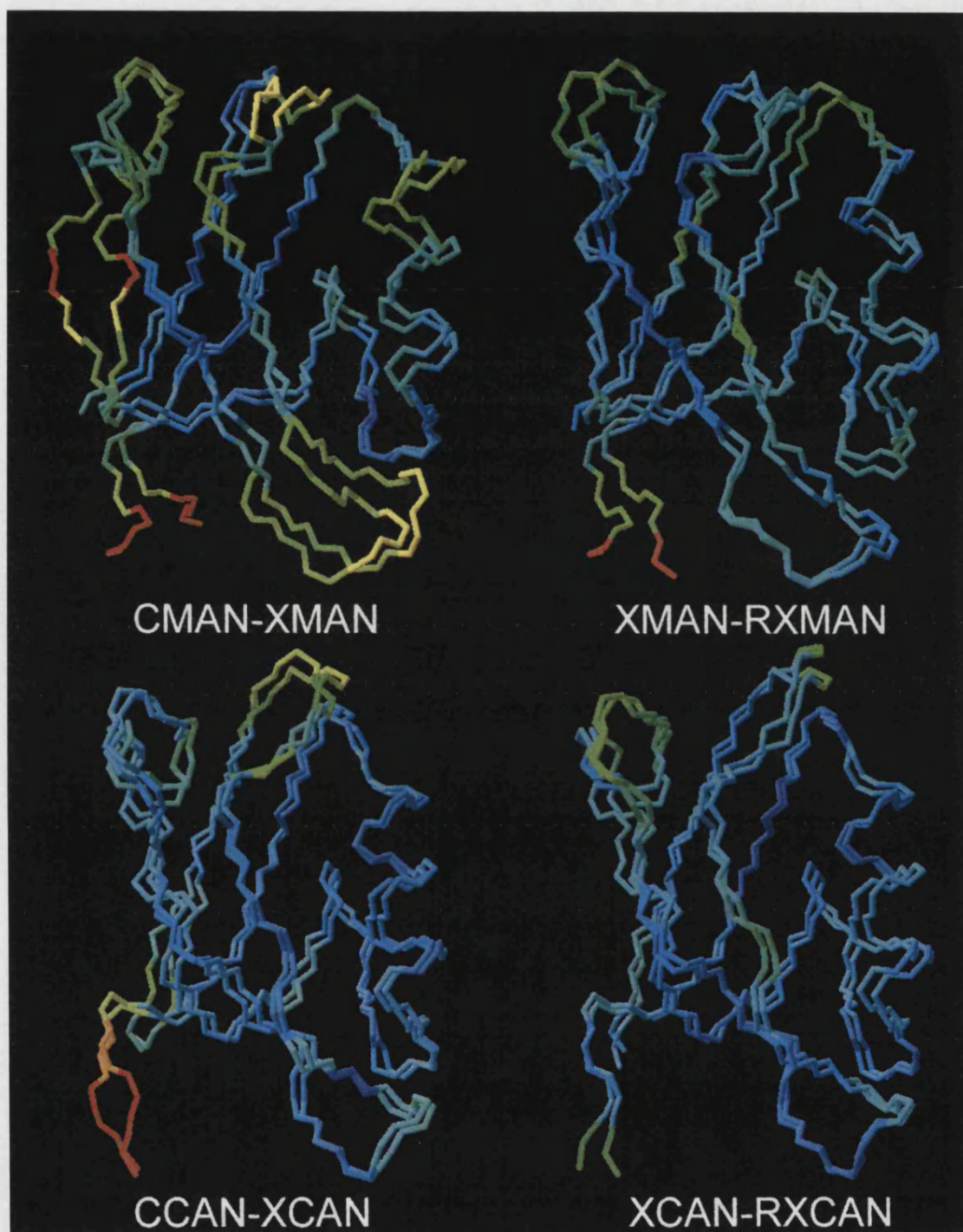
## **5.6 Physical Comparison of ZASP-PDZ structure calculations**

After having setup the conditions with several exploratory runs, we decided to analyse six bundles obtained by different protocols (Figure 5.10): primary manual and automatic structure calculations in CYANA (CMANUAL & CCANDID), which were subsequently recalculated in XPLOR (XMANUAL & XCANDID), and further refined using the 'CHARMM22 water refinement' (RXMANUAL & RXCANDID). Although the fold was the same for all of the calculated structures, the structure details diverged depending on the protocol used. Therefore, a comparison of the structures was undertaken to observe the effect of the CYANA → XPLOR → Water refinement, and also the differences between the manual and automatic calculations. The structure with the lowest RMSD to the mean from the best 20 structures for each bundle was compared for each calculation procedure. RMSD calculations were calculated per residue, using the backbone atoms of residues that did not have a flexible profile from the relaxation analysis (Appendix 10.3.1.1.20): these were residues: 3-7,9-10,12,15-30,32-37,39-58,60-72, and 76-82. Only significant changes in these regions were considered for further investigation, as flexible regions of the protein are likely to have a high RMSD. RMSD changes over 2 Å were considered as a significant change.

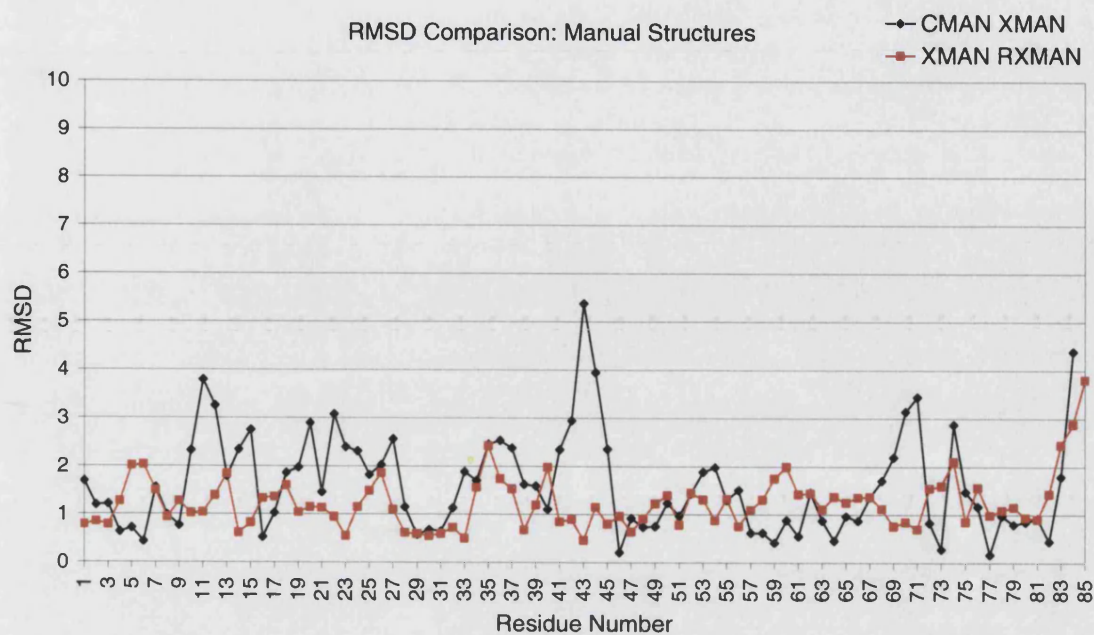
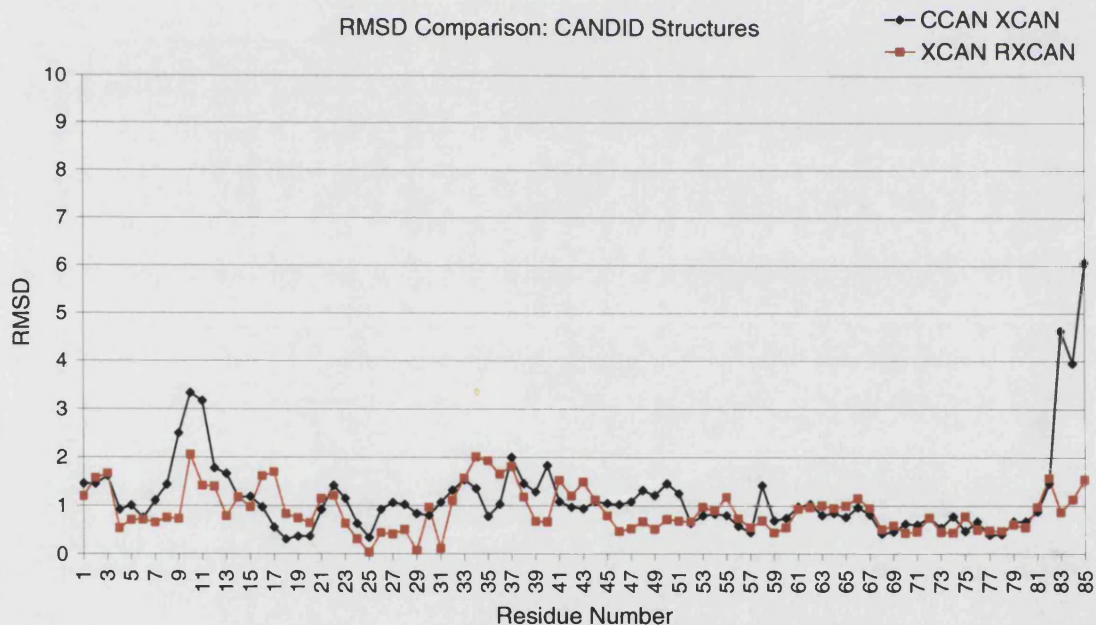


**Figure 5.10:** Structure of ZASP-PDZ. NMR bundles corresponding to the manually (top row) and automatically (bottom row) assigned structure families. From left to right are displayed the bundles without refinement (A and D), with RDC (B and E) and water refinement (C and F) respectively. All structures have no distance violations greater than 0.5 Å and no dihedral angle violations greater than 5 ° from the experimental data.

A series of pair-wise RMSD calculations using WHATIF revealed significant differences in the manual CYANA to XPLOR conversion (Figure 5.11 and Figure 5.12), having significant changes at 10-12, 20, 22-24, 26-27, 35-37, 41-45, 69-71, 73-74, 83-85. Many of these residues are in loops or areas of random coil, and will therefore have few structural restraints. The inclusion of the RDCs appeared to re-orientate the loop after  $\alpha 1$ , bringing it closer to the core of the molecule, and also  $\alpha 2$  appears to shift around  $5^\circ$  from the base of the helix. The proceeding water refinement produced few changes, with only significant variations observed in residues 35, 74, and 83-85. The CCANDID to XCANDID conversion (Figure 5.11 and Figure 5.12) only had significant changes in residues 9-11 (loop) and 82-84 (C-termini), again, both flexible regions. The RDCs appeared to show no notable effects on the structure. The water refinement showed even less of a change than the conversion from CYANA to XPLOR (Figure 5.11 and Figure 5.12), only giving a change in residue 10. Many of the differences between CYANA and XPLOR are probably due to the addition of RDC restraints on conversion to XPLOR and also the difference in the force field used to calculate the structures. Also, since the average distance restraint is looser for the manual calculation (Section 5.5), these differences will have a larger effect on the calculated structures as it is less constrained.



**Figure 5.11:** Comparison of ZASP-PDZ structure calculations (lowest RMSD to the mean) for both manual (top) and CANDID (bottom), after conversion from CYANA to XPLOR (left) and water-refinement (right). The structures of the comparison (labelled) have been superimposed, and colour coded from blue to red to denote the change in RMSD (low to high respectively).

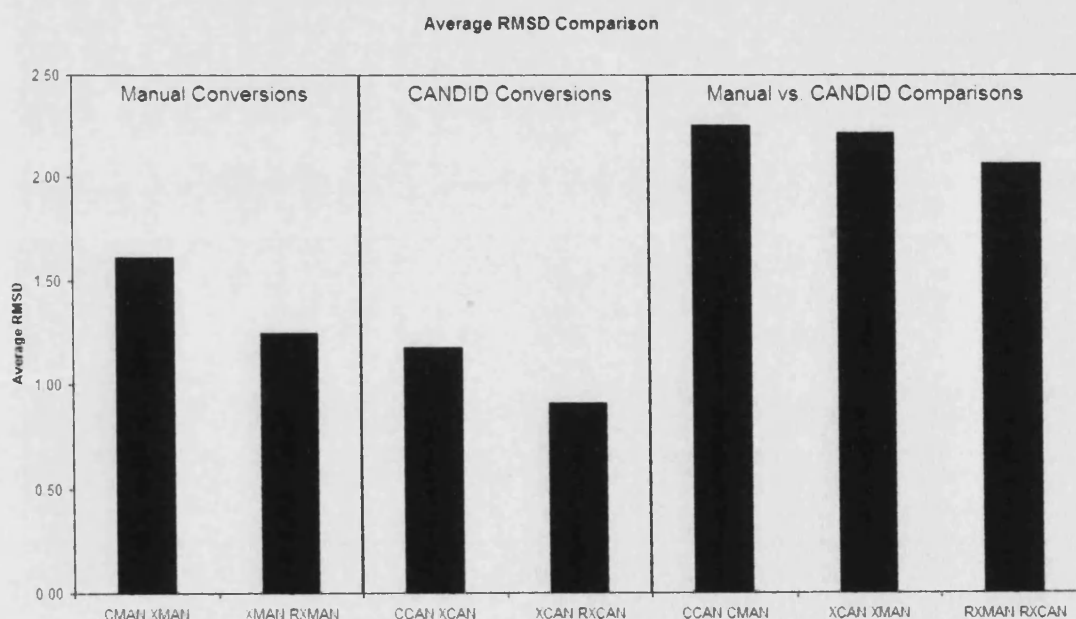
**A****B**

**Figure 5.12:** Comparison of ZASP-PDZ structure calculations (lowest RMSD to the mean) for both manual (A) and CANDID (B). The result for the CANDID/XPLOR comparison is in black, whilst the result for the XPLOR/water-refinement is in red.

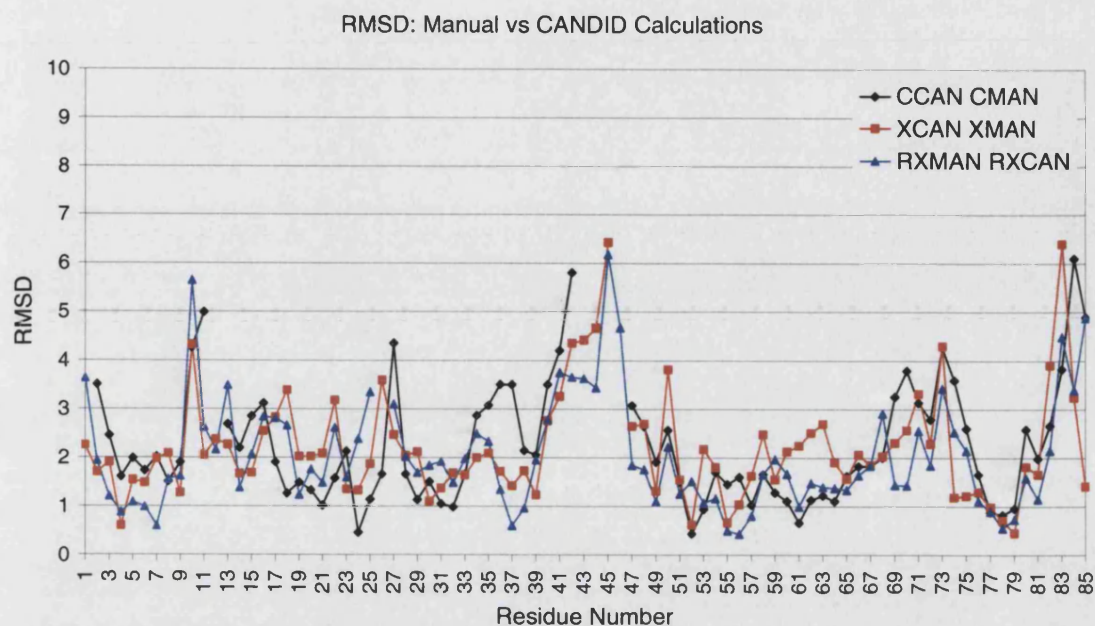
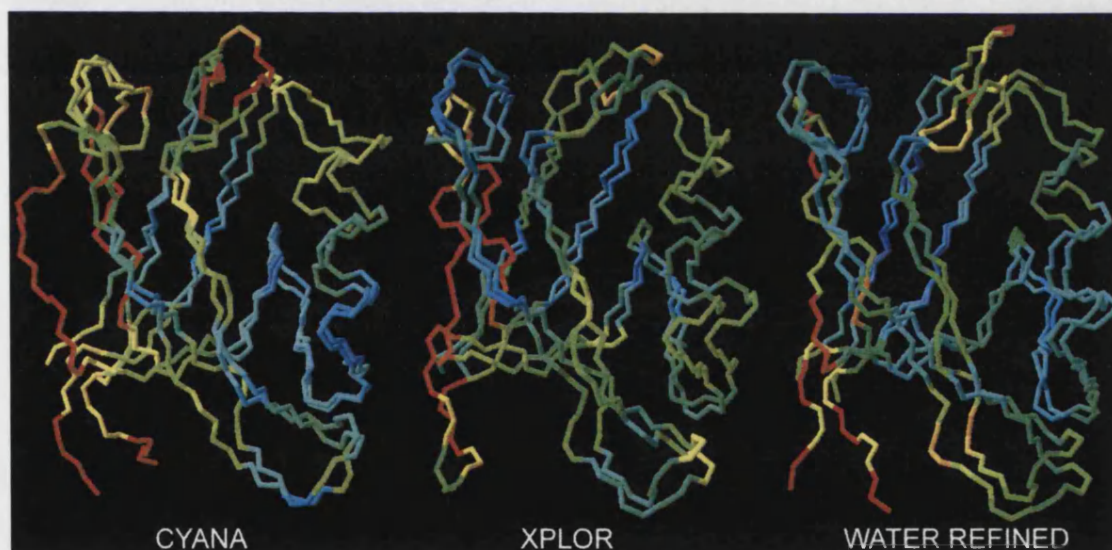


Differences between the manual and the CANDID structure calculation in CYANA (Figure 5.14) show changes in several different places: the general appearance shows that the CANDID result is more compact than the manual. This is mainly due to the orientation of the GLGF loop (13-16), beginning of  $\beta 3$  (27), 34-38, the loop at 40-47, and between  $\alpha 2$  and  $\beta 6$  at 69-77. The differences in the manual and automatic structures after exporting the data into XPLOR appear to show even more variations, the most notable being the start of the helix at 60-63. However, there is improvement of convergence of the loop at 42-45 as it has a lower RMSD value. The water refinement shows even better convergence between 41-46.

Comparing the overall pair-wise RMSD of the structure calculations show that the water refinement improved the convergence between the structures within the manual and automatic calculations (i.e. the difference in the structures between from CYANA to XPLOR is more than the difference between XPLOR and water-refined structures) (Figure 5.13). The water-refinement also appeared to improve convergence between the manual and CANDID structures (i.e. CMANUAL vs. CCANDID, XMANUAL vs. XCANDID, and RXMANUAL vs. RXCANDID).



**Figure 5.13:** Summary of the average pair-wise RMSD between the structure calculations. The first four values show less difference between the manual and CANDID structures between the conversions, while the last 3 values signify the difference between the manual and CANDID procedures.

**A****B**

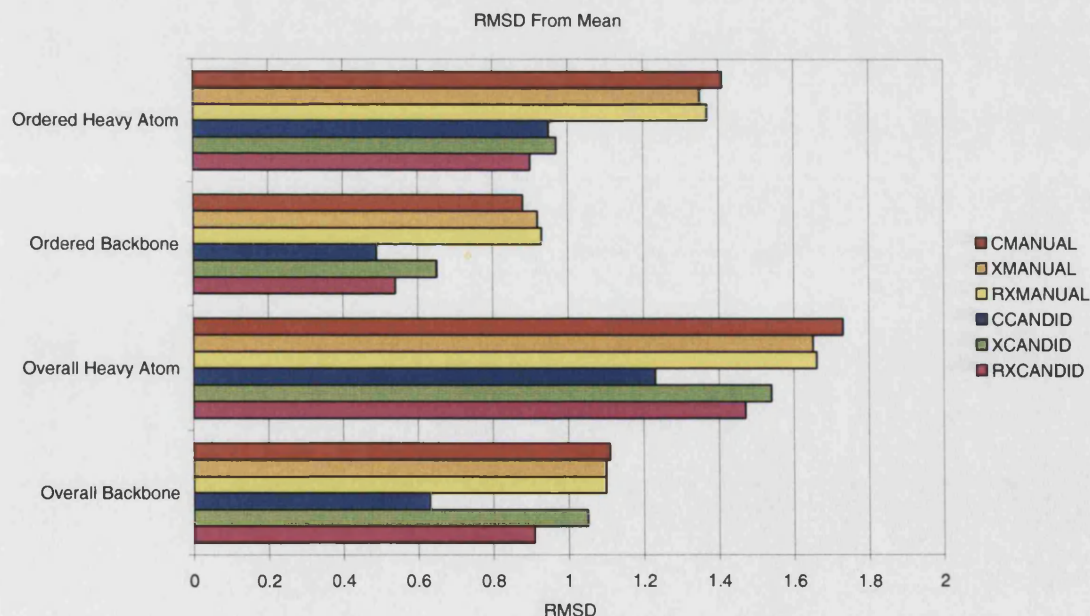
**Figure 5.14:** A) Comparison of ZASP-PDZ manual and CANDID structure calculations (lowest RMSD to the mean). The result for the CMANUAL/CCANDID comparison is in black, XMANUAL/XCANDID is in red, whilst the result for the RXMANUAL/RXCANDID is in blue. B) The residue-by-residue comparison of the manual and CANDID calculation, in both XPLOR and CYANA and after water refinement. The structures of the comparison (labelled) have been superimposed, and colour coded from blue to red to denote the change in RMSD (low to high respectively).

## 5.7 Precision, Quality, and Accuracy Comparison

The calculated structures were compared for precision, quality, and accuracy. Here, an important distinction between precision and accuracy is made. If you have a target, for example – a darts board, then the accuracy is how close your dart is to the bull's-eye, (i.e. the target or 'true' result). The precision is how close your darts are bunched together (i.e. the difference between your results). Therefore, these are two mutually exclusive variables.

### 5.7.1 Precision: RMSD Calculations

An analysis of the RMSD for each structure was performed with and without residues that were identified as flexible from the relaxation analysis (Table 5.4 and Figure 5.15, Appendix 10.3.1.1.20). The CANDID structures had lower RMSD values than the manual structures for each RMSD category. Therefore, the CANDID structures have a higher precision compared to the manual calculation, with the CCANDID structures having the best overall precision (although RXCANDID structures had a slightly better ordered heavy atom RMSD).



**Figure 5.15:** Comparison of the backbone and heavy atom RMSD for the best 20 structures from all structure calculations. RMSD calculations include the overall result, and an ordered selection of residues (3-7,9-10,12,15-30,32-37,39-58,60-72, and 76-82). The CANDID group of structures consistently have better precision to their manual calculation counterpart.

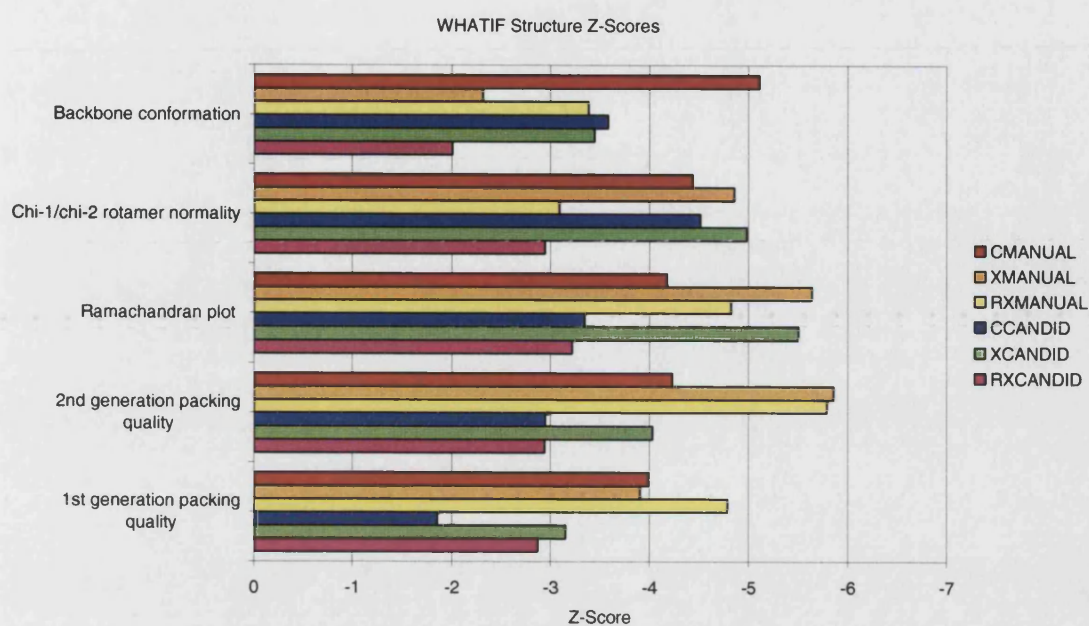
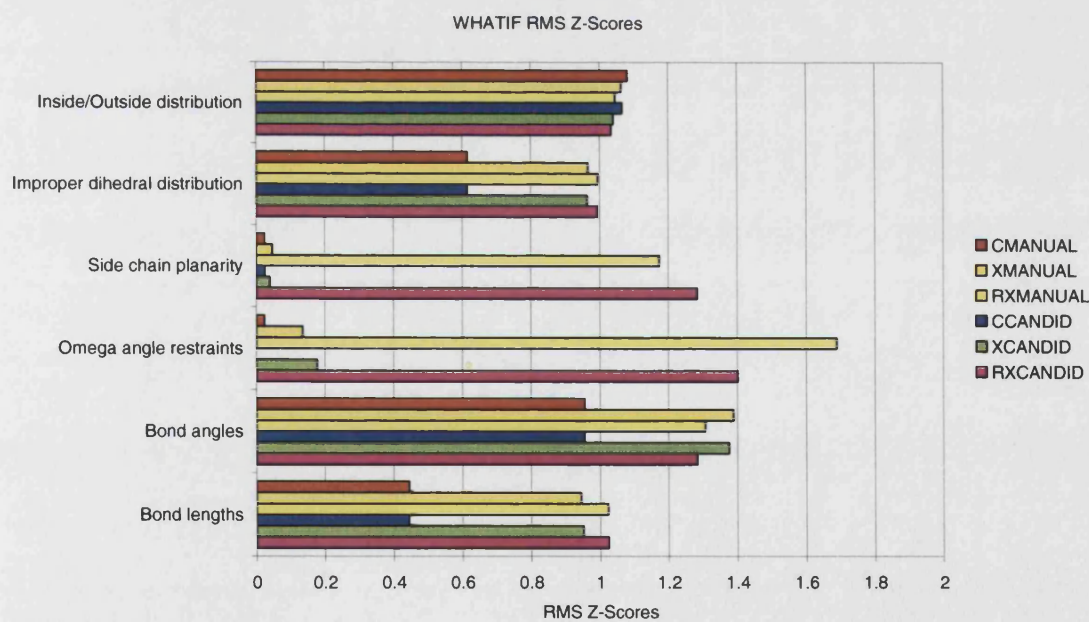


	CMANUAL	XMANUAL	RXMANUAL	CCANDID	XCANDID	RXCANDID
<b>RMSD from mean</b>						
Overall Backbone	1.11	1.10	1.10	0.63	1.05	0.91
Overall Heavy Atom	1.73	1.65	1.66	1.23	1.54	1.47
Ordered Backbone	0.88	0.92	0.93	0.49	0.65	0.54
Ordered Heavy Atom	1.41	1.35	1.37	0.95	0.97	0.90
<b>Structure Z-scores:</b>						
1st generation packing quality	-3.987	-3.903	-4.786	-1.853	-3.151	-2.872
2nd generation packing quality	-4.228	-5.854	-5.787	-2.948	-4.028	-2.941
Ramachandran plot	-4.173	-5.636	-4.82	-3.342	-5.499	-3.217
Chi-1/chi-2 rotamer normality	-4.433	-4.851	-3.09	-4.509	-4.981	-2.943
Backbone conformation	-5.106	-2.316	-3.382	-3.579	-3.446	-2.003
<b>RMS Z-scores:</b>						
Bond lengths	0.444	0.944	1.023	0.443	0.950	1.024
Bond angles	0.955	1.388	1.307	0.954	1.375	1.283
Omega angle restraints	0.022	0.135	1.689	0.000	0.177	1.401
Side chain planarity	0.023	0.046	1.173	0.025	0.039	1.284
Improper dihedral distribution	0.615	0.966	0.994	0.614	0.963	0.993
Inside/Outside distribution	1.082	1.063	1.045	1.066	1.040	1.034
<b>Ramachandran plot (%)</b>						
Most favored regions	70.9	72.5	76.3	73.8	65.1	81.6
Additional allowed regions	25.7	22.9	19.8	24.3	31.6	16.8
Generously allowed regions	2.8	3.7	2.5	1.1	2.5	1.3
Disallowed regions	0.6	0.8	1.3	0.8	0.8	0.3

**Table 5.4:** Values from the RMSD calculation, WHATIF scores, and Ramachandran plot statistics for each structure calculation. RMSD has been calculated overall and with only the selection of residues 3-7,9-10,12,15-30,32-37,39-58,60-72, and 76-82 (ordered). Backbone atoms are defined by the backbone Ca, N, and CO, while C, N, O, and S define the heavy atoms.

### 5.7.2 Quality: WHATIF scores

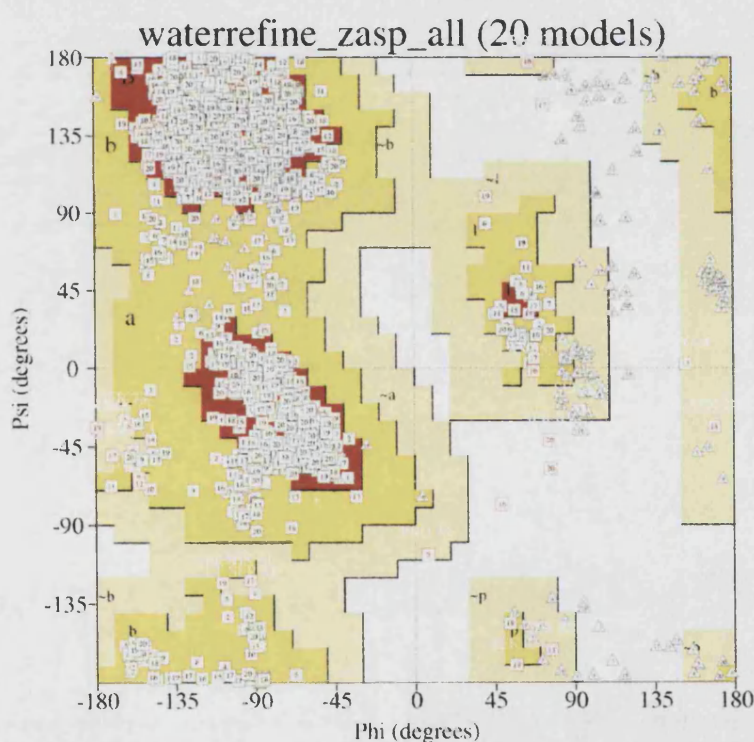
WHATIF (Hoofst et al., 1996) was used to check the quality of various structural parameters such as packing, dihedral angles (i.e. Ramachandran plot ( $\psi$  &  $\phi$ ), chi ( $\chi$ ) angle analysis, and  $\omega$  angle restraints), backbone conformation, and allowed bond lengths, etc (Table 5.4). The graph for the WHATIF 'structure Z scores' (Figure 5.16A) give a result where an average result is 0, a better than average result is positive, and a worse than average result is negative. The WHATIF 'RMS Z-scores' (Figure 5.16B) should be as close to 1 as possible. From the six structure calculations, RXCANDID consistently gave the best 'structure Z-scores'. Both the RXMANUAL and RXCANDID structures give a comparably good 'RMS Z-score', showing definite improvement from their CYANA and XPLOR counterparts. From this analysis, the best quality structure calculation is RXCANDID.

**A****B**

**Figure 5.16:** Comparison of the WHATIF structure (A) and RMS (B) Z-scores for all structure calculations. The RXCANDID structures give consistently the best structure Z-scores, with the exception of '1<sup>st</sup> generation packing quality', while RXMANUAL and RXCANDID structures give a comparably good 'RMS Z-score'

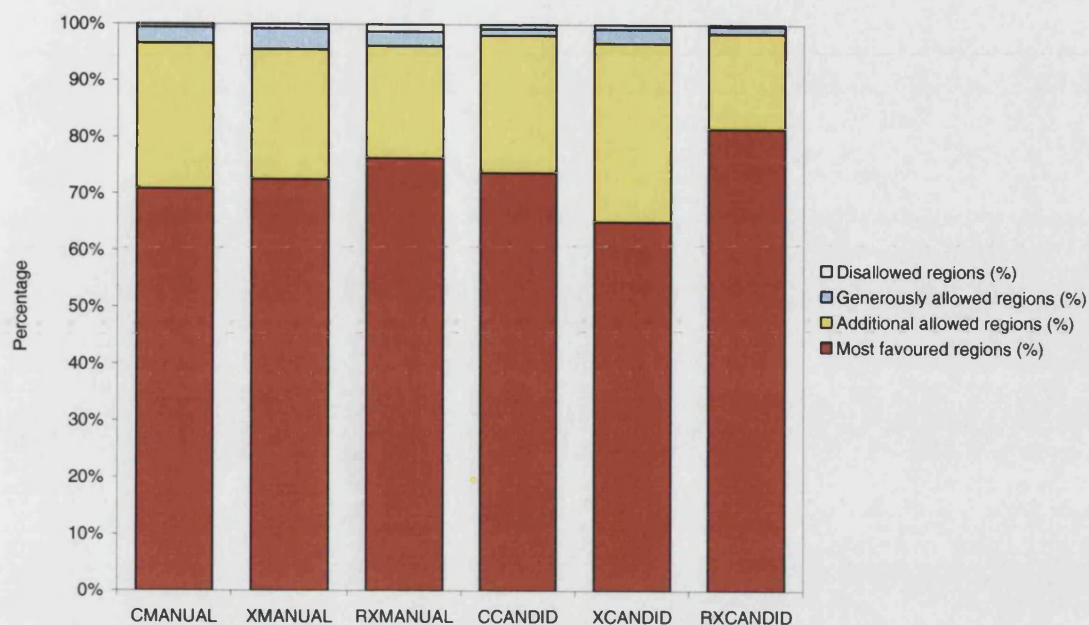
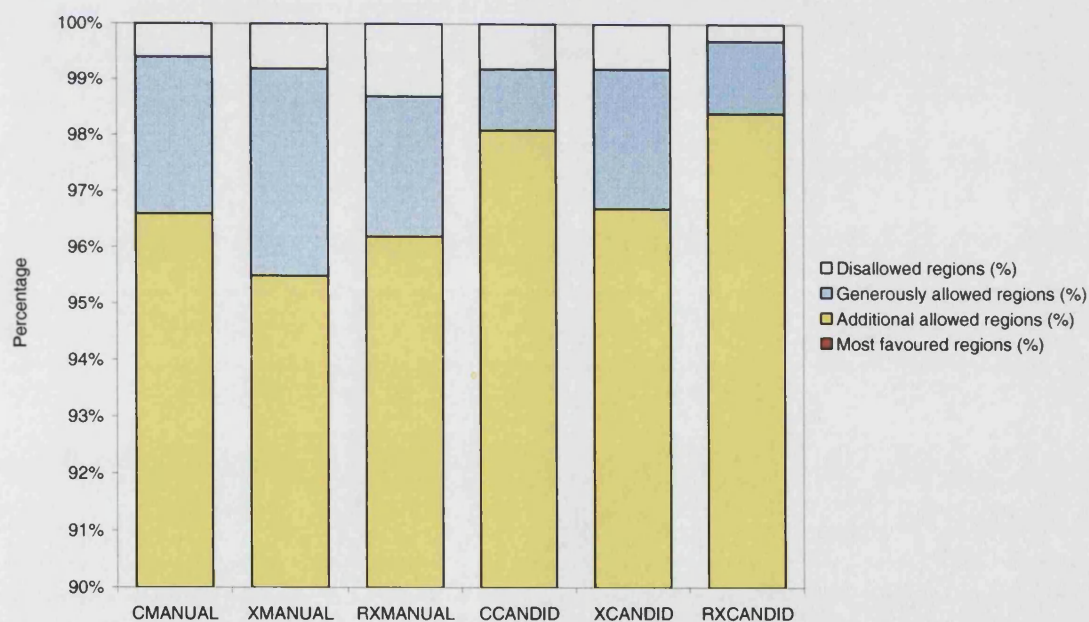
### 5.7.3 Quality: Ramachandran plots

PROCHECK-NMR (Laskowski et al., 1996) was used to conduct the detailed Ramachandran plot analysis of each structure calculation for the best twenty structures, giving an accurate distribution of the residues in the favoured, additional allowed, generously allowed, and disallowed regions (Table 5.4 and Figure 5.18). All structure calculations have over 98% of their residues in the allowed regions and over 95% in the favoured and additionally allowed regions (both excluding glycine and proline residues). Again, the water-refinement appeared to improve the quality of the structures, increasing the percentage distribution of residues in the 'most favoured regions' compared to the CYANA and XPLOR counterparts. However, the RXCANDID again had the best quality, with the highest percentage of residues in the 'most favoured regions' and the least number of residues in the 'disallowed regions' (Figure 5.17).



**Figure 5.17:** Ramachandran plot of RXCANDID generated in PROCHECK-NMR. Squares and triangles represent non-glycine residues and glycine residues respectively.



**A****B**

**Figure 5.18:** A) Ramachandran plot comparison of all of the structure calculations, giving the distribution of residues in the most favoured (red), additionally allowed (yellow), generously allowed (light blue), and disallowed regions (white). Graph B has been zoomed to highlight the differences in the additionally, generously, and disallowed regions. RXCANDID has the best overall Ramachandran scores.

#### 5.7.4 Accuracy: Comparison Of Known PDZ Domains

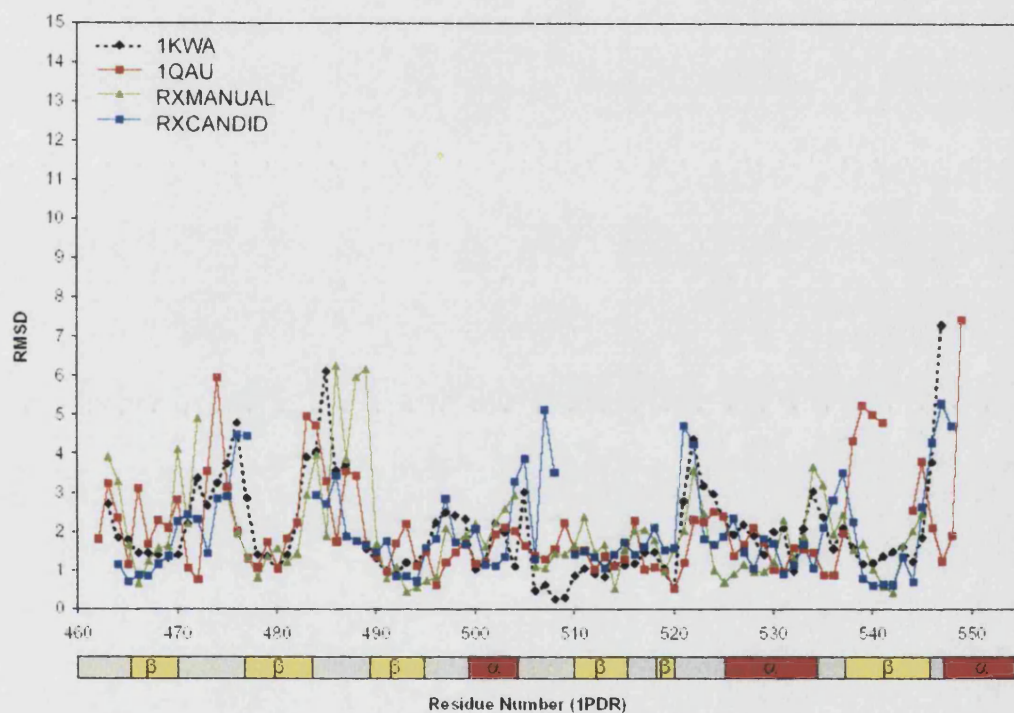
To estimate the structure accuracy of our results, we compared them with the known structures of PDZ domains as suggested in (Walma et al., 2002). Here, they compared their NMR structure of mouse PTP-BL PDZ domain (1gm1) with the NMR structure of human PTP-Bas PDZ2 (3pdz) (Kozlov et al., 2000), to which it shared a 94% sequence identity. The comparison also included the crystal structures DlgA (1pdr - 34% identity), CASK (1kwa - 28% identity), and syntrophin (1qav - 39% identity) (Doyle et al., 1996; Hillier et al., 1999; Morais Cabral et al., 1996). They found that although 3pdz had the highest sequence identity to 1gm1, it ironically had the highest RMSD compared with the crystal structures; hence the 3pdz was not deemed as accurate.

A DALI structural similarity search (Holm and Sander, 1996b) of ZASP-PDZ resulted in six 'hits' with RMSD values between 2.5-3.0 Å and sequence identities ranging between 21-30%, all corresponding to PDZ domains (Table 5.5). The highest similarity is with the crystal structure of the 3rd PDZ domain from human disks large protein (DlgA, entry code 1pdr (Morais Cabral et al., 1996), which superposes ZASP-PDZ with an RMSD of 2.5 Å (Z score of 11.1, 23% sequence identity). Pair-wise comparison of 1pdr with the high resolution crystallographic structures of two representatives of the major interaction classes (1kwa and 1qav (Doyle et al., 1996; Hillier et al., 1999)) shows that the structural variability between the ZASP-PDZ final bundles is in regions with the largest variability for the crystal structures (Figure 5.19). This would suggest that the two bundles have a comparable accuracy.

PDB/Chain	Rep	Z	RMSD	LALI	LSEQ2	%Ide	Protein Name
1PDR	12	8.7	2.5	81	96	27	3rd PDZ from DLG
2PDZ-A	1PDR	8.1	2.7	80	86	30	Syntrophin PDZ domain
1QAV	5	7.8	3.0	83	112	22	Neuronal nitric oxide sythase
1I92	1PDR	7.8	2.7	79	91	28	Na <sup>+</sup> /H <sup>+</sup> exch. regulatory factor
1KWA-A	1KWA	6.9	2.7	78	88	21	Human cask/lin-2 PDZ domain
1KEF-A	1PDR	6.9	3.0	81	93	28	Synapse associated protein-90

Field	Definition
Rep	A number denotes the number of structures in the PDB the fold represents, or a PDB accession code denotes the structure in the PDB that represents it in the FSSP.
Z	Z-score: strength of structural similarity in standard deviations above expected. Only scores >2 are reported.
RMSD	Positional root mean square deviation of superimposed CA atoms in Angstroms.
LALI	Total number of equivalent residues.
LSEQ2	Length of the entire chain of the equivalent structure.
%Ide	Percentage Identity

**Table 5.5:** Results from the DALI search of ZASP-PDZ.



**Figure 5.19:** Pair-wise comparison of the manually and automatically assigned bundles, 1kwa and 1qav structures with 1pdr, taken as a reference. The pair-wise Z scores as obtained by WHATIF MOTIF option are plotted versus the sequence of 1pdr. The secondary structure elements of 1pdr are indicated underneath with yellow ( $\beta$ -strands) and red ( $\alpha$ -helices) boxes.



## **5.8 Conclusion**

The solution structure of ZASP-PDZ was solved using manual and automatic procedures, different structure calculation programs, and different protocols giving a total of six separate results. The structures were compared in terms of precision, accuracy, and quality. The automatic NOESY assignment and structure calculation (CANDID) was faster than the manual procedure, taking hours rather than weeks. RMSD calculations showed that the CANDID structures had higher precision than manually calculated structures. The WHATIF scores also showed that the CANDID structures had higher quality than the manual structures, although the Ramachandran plot showed relatively minor differences between the two methods. Comparing the manual and automatic calculated structures against known PDZ structures, using pair-wise RMSD calculations as a guide, gave an indication of accuracy. Again, the CANDID structures appeared to be more accurate than the manually calculated structures. The RDC constraints had less effect on the CANDID calculation than the manual calculation, meaning they had better agreement with the automatically calculated NOE distance restraints. On these criteria, the CANDID structures are better than those of the manual run. These differences between the automatic and manual calculated structure could be explained by looking at the way the NOESY spectra were analysed. CANDID was able to keep the average distance length lower in comparison to the manual run, which lead to a more compact and therefore more precise, structure. However, there were less NOESY peak assignments made by CANDID in the final structure calculation, which could be down to the elimination of restraint that did not agree with the structure. A limitation with CANDID was that it could only correctly calculate the PDZ fold with restraints additional to the NOE distances – without dihedral angles and hydrogen bonds, the program did not come to the correct solution. Also another limitation of CANDID is the high percentage of proton assignment (typically 90%) needed for it to come to a correct calculation, and within that, a good agreement between the peak list and the assignment list.

The structural restraints were converted from CYANA to XPLOR for structure calculation; hence a comparison was made between the structures calculated by each program. This conversion caused significant differences in the structure, which can be attributed to the difference in force field between CYANA and XPLOR: CYANA uses a ECEPP/2 force field (Momany et al., 1975; Nemethy et al., 1983), whereas XPLOR uses a adapted CHARMM force field (Brooks et al., 1983). The force field defines geometric

parameters, partial atomic charges, non-bonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the amino acids, hence the differences in the results. The comparison of ZASP-PDZ against the known PDZ domains was close between CYANA and XPLOR structures, meaning there was no significant improvement in accuracy on conversion to XPLOR. The 'lowest RMSD to the mean structures' pairwise residue comparison of the CYANA and XPLOR calculations showed no important differences. However, analysis at backbone and heavy atom RMSD calculations over their best twenty structures show that the precision of the CYANA calculated structures are better than XPLOR. This is probably due to the REDAC strategy that reduces the conformations the CYANA calculated structures are allowed to occupy. The REDAC strategy is also a feature in CYANA that improved the convergence of the calculated structures. The WHATIF scores for the XPLOR structures appear to be better than that of CYANA, as it gave better scores in the WHATIF RMS comparison; this could also be attributed to the difference in force field between the two programs.

The CHARMM22 water refinement appears to have improved the precision and quality of the XCANDID structures. All of the RMSD calculations show a small improvement (backbone, and heavy atom, both with and without the selected residues). The PROCHECK-NMR results show a marked improvement after water refinement, as there is an increase in the percentage of residues in the most favoured regions of the Ramachandran plot with the concomitant decrease of residues in the additionally, generously and disallowed regions. WHATIF analysis also shows improvement in every assessment parameter in both the 'Structure Z-score' and 'RMS Z-score' categories. However, the XMANUAL structures did not appear to benefit as much as the XCANDID structures – the precision decreased in categories except for the overall backbone RMSD. The quality of the water refined XMANUAL structures did not improve significantly, as the Ramachandran plot only saw minor variations in the distribution of the residues in the favoured regions. Also, the WHATIF analysis did not show significant improvement in the 'Structure Z-scores', but there were some minor improvements in the 'RMS Z-scores'.

From this analysis, the water refined RXCANDID structures were chosen to represent the solution structure of ZASP-PDZ, as the CANDID structures were better than the manually derived structures in terms of precision, accuracy, and quality, especially after water refinement.

## Chapter 6

---

# Interaction of $\alpha$ -Actinin-2 with ZASP and Titin

---

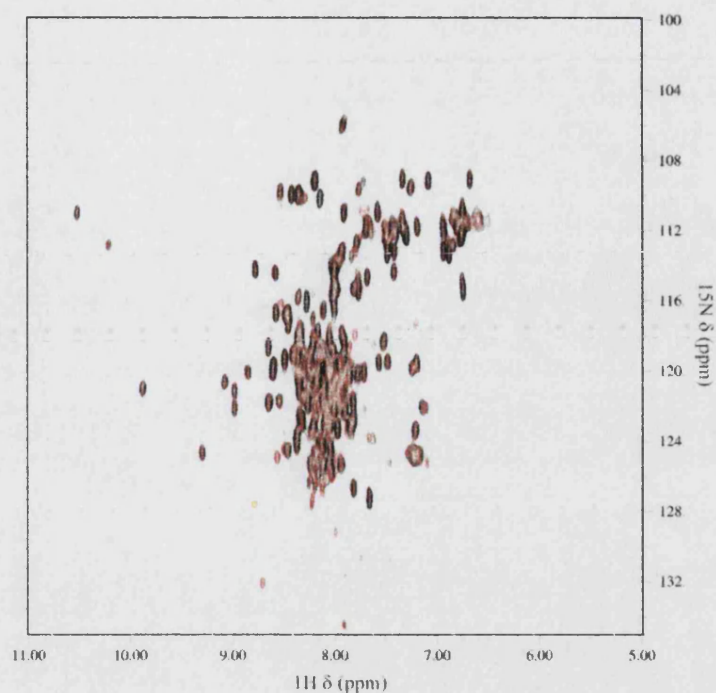
The interaction between  $\alpha$ -actinin-2 with ZASP and titin was characterised by both NMR and fluorometric titration. Two  $\alpha$ -actinin-2 constructs were used: Act-EF1234 and Act-EF34, which contains all four and last two EF-hands of its calmodulin-like domain respectively, as well as the ZASP-PDZ construct. The tertiary complex was formed by first saturating Act-EF34 with tZR7 - a synthetic peptide that spans 24 residues of the seventh Z-repeat of titin: tZR7 was designed to comprise of the region of interaction observed in the structure of Act-EF34/ZR7 complex (See Appendix 10.1.5 and (Atkinson et al., 2001)) – N.B. ZR7 was chosen from seven Z-repeats studied as it had the highest K<sub>d</sub> to Act-EF1234/34 (Joseph et al., 2001). ZASP-PDZ was subsequently titrated against the Act-EF34/tZR7 complex to form the tertiary complex.

The affinity of the interaction between ZASP-PDZ and  $\alpha$ -actinin-2 was measured using fluorescence spectroscopy, by following the signal of the unique tryptophan residue (W13) present in ZASP-PDZ and titrating Act-EF34. The fluorometric titration was then repeated using the Act-EF34/tZR7 complex to check whether the presence of tZR7 would modulate the interaction affinity.

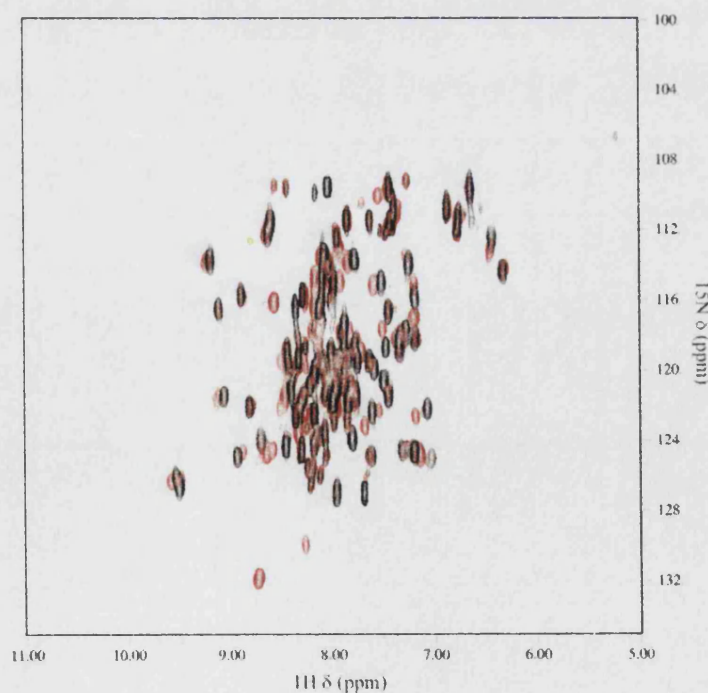
## **6.1 NMR Titrations**

### **6.1.1 $^{15}\text{N}$ -Act-EF1234 and $^{15}\text{N}$ -Act-EF34 with ZASP-PDZ**

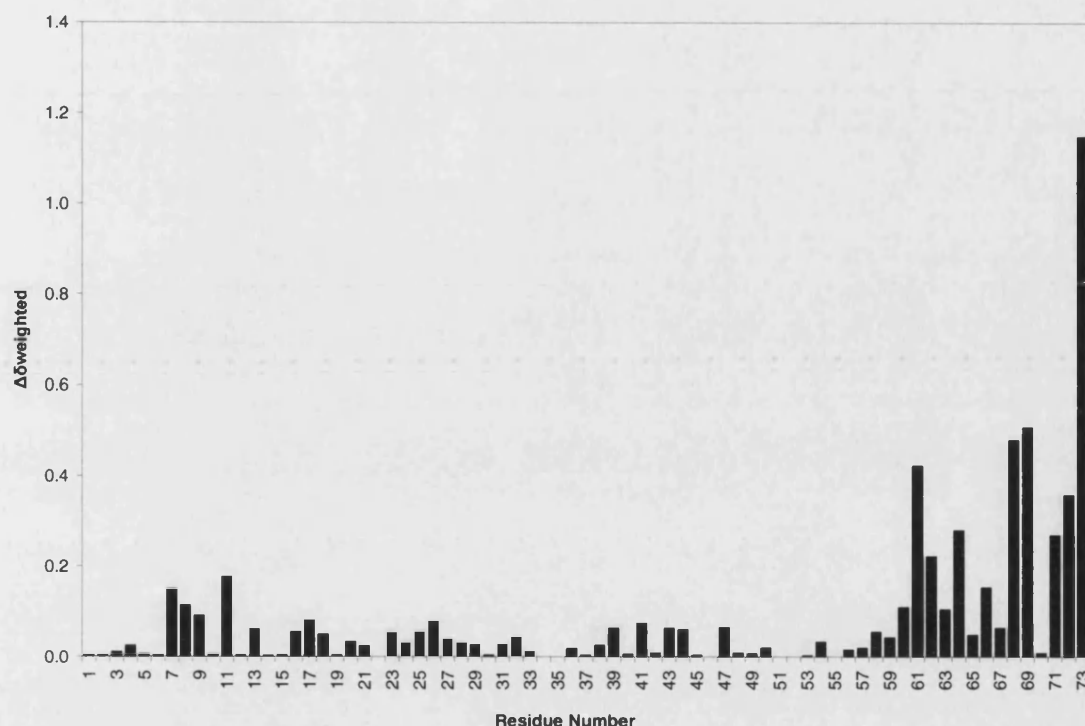
The residues involved in the interaction between Act-EF34 and Act-EF1234 with ZASP were identified by a series of pair-wise titrations. Unlabelled ZASP-PDZ was first titrated against  $^{15}\text{N}$ -Act-EF34 and  $^{15}\text{N}$ -HSQC spectra were recorded for each addition of ZASP-PDZ (Figure 6.2). The observed chemical shift changes confirm that ZASP-PDZ interacts with  $\alpha$ -actinin-2, with the changes mostly being in the intermediate exchange regime. Plotting the  $\Delta\delta_{\text{weighted}}$  against residue number (Figure 6.3) reveal changes localised to Act-EF34's C-terminus section, ranging from the 0 to the -12 position. The same conclusion was made for the titration using the longer  $^{15}\text{N}$ -Act-EF1234 construct (Figure 6.1 and Figure 6.3) and unlabelled ZASP-PDZ. Although the NMR assignment of this section of Act-EF1234 is not available, a partial analysis was conducted by superimposing the assigned  $^{15}\text{N}$ -HSQC spectra of Act-EF34 against Act-EF1234's (Section 4.4 and Figure 4.11). The peaks most affected are those corresponding to Act-EF34's A61, S71, D72, and L73 (Figure 6.13B). The titration shows the direction and  $\Delta\delta_{\text{weighted}}$  changes for these residues are almost identical to the Act-EF34 and ZASP-PDZ titration. No appreciable effect could be observed from the peaks attributed to residues from Act-EF12. Both titrations were taken to an Act/ZASP ratio of 1:2. See Appendix 10.8 for the experimental conditions for the titrations.



**Figure 6.1:** <sup>15</sup>N-HSQC titration of <sup>15</sup>N-Act-EF1234 with (red) and without (black) ZASP-PDZ (2:1 ZASP-PDZ excess, final titration point). Run at 600 MHz, 300 K, with 0.46 mM (initial) <sup>15</sup>N-Act-EF1234.



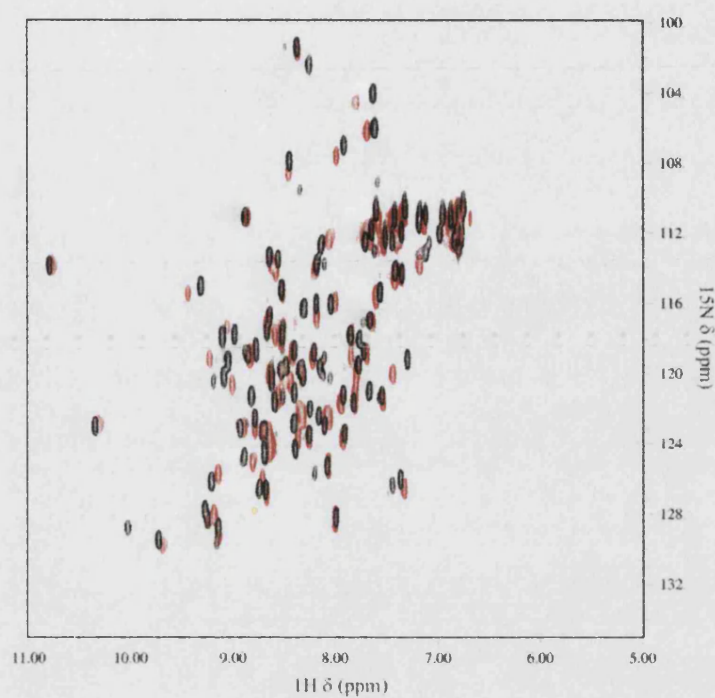
**Figure 6.2:** <sup>15</sup>N-HSQC of <sup>15</sup>N-Act-EF34 with (red) and without (black) ZASP-PDZ (2:1 ZASP-PDZ excess, final titration point). Run at 500 MHz, 300K, 0.60 mM (initial) <sup>15</sup>N-Act-EF34.



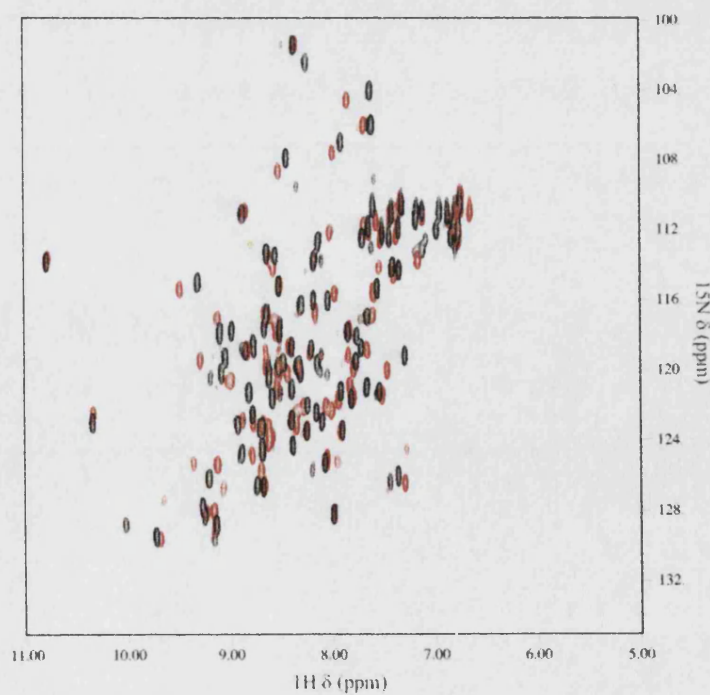
**Figure 6.3:**  $\Delta\delta_{\text{weighted}}$  changes in  $^{15}\text{N}$ -Act-EF34 upon binding ZASP-PDZ. The changes appear to be localised to Act-EF34's C-terminus.

### 6.1.2 $^{15}\text{N}$ -ZASP-PDZ with Act-EF1234 and Act-EF34

The back-titration between unlabelled Act-EF1234 against  $^{15}\text{N}$ -ZASP-PDZ was performed (Figure 6.4) to a ratio of 2:1. Plotting  $\Delta\delta_{\text{weighted}}$  against residue number (Figure 6.6) localises the changes to a defined area of ZASP-PDZ. Residues W13, G14, and F15 from the 'GLGF' motif (PWGF in ZASP) are in intermediate exchange, indicating a significant effect on these residues in the titration. Using a 0.3  $\Delta\delta_{\text{weighted}}$  threshold highlights further changes in the second  $\beta$ -strand: residues L17 and Q18 have large chemical shift changes of  $\sim 6$  ppm in the  $^{15}\text{N}$  dimension for both residues (Figure 6.13A).  $\Delta\delta_{\text{weighted}}$  threshold of 0.2 highlights that residues 16, 20, 28 and 31 are also affected upon complex formation. The titration between  $^{15}\text{N}$ -ZASP-PDZ and Act-EF34 (Figure 6.5) gave an almost identical change in chemical shift (Figure 6.6).

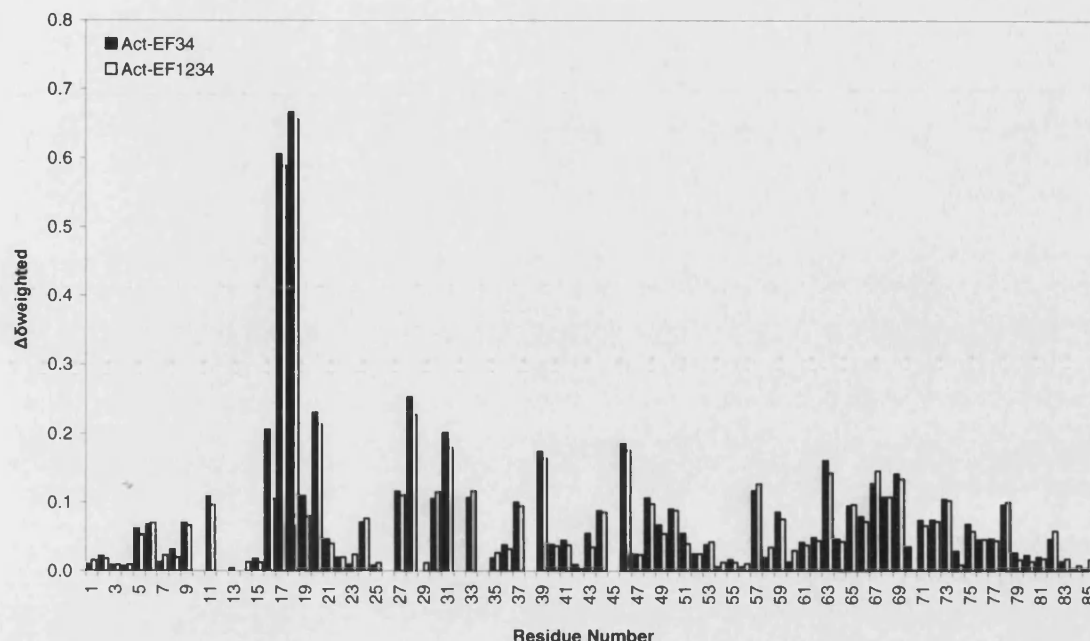


**Figure 6.4:**  $^{15}\text{N}$ -HSQC of  $^{15}\text{N}$ -ZASP-PDZ with (red) and without (black) Act-EF1234 (2:1 Act-EF1234 excess, final titration point). Run at 600 MHz, 300K, 0.30 mM (initial)  $^{15}\text{N}$ -ZASP-PDZ.



**Figure 6.5:**  $^{15}\text{N}$ -HSQC of  $^{15}\text{N}$ -ZASP-PDZ with (red) and without (black) Act-EF34 (2:1 Act-EF34 excess, final titration point). Run at 600 MHz, 300K, 0.37 mM (initial)  $^{15}\text{N}$ -ZASP-PDZ.



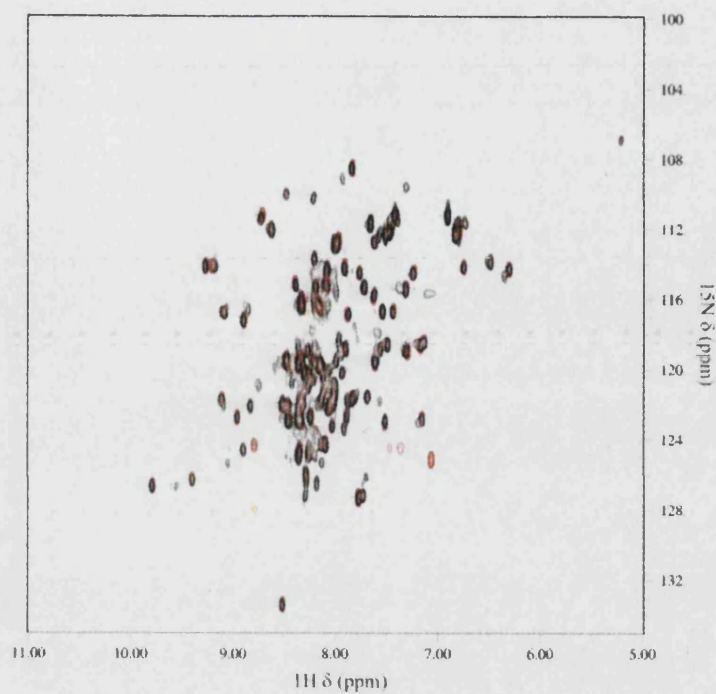


**Figure 6.6:**  $\Delta\delta_{\text{weighted}}$  changes of  $^{15}\text{N}$ -ZASP-PDZ with Act-EF1234 (white) and Act-EF34 (black).

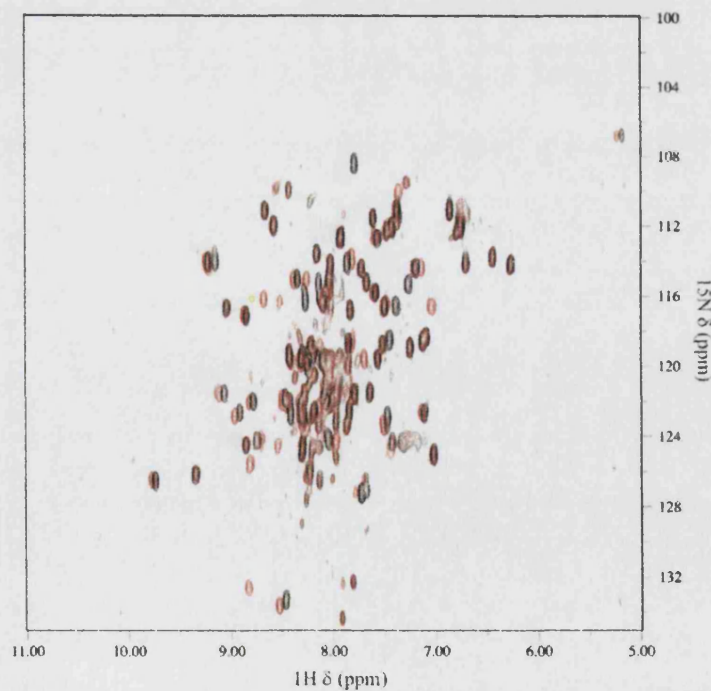
### 6.1.3 $^{15}\text{N}$ -Act-EF34 in complex with Minimal Titin ZR7 and ZASP-PDZ

The titration of the trimeric complex involves components of ZASP,  $\alpha$ -actinin-2, and titin. The  $^{15}\text{N}$ -HSQC from the Act-EF34/ZR7 and Act-EF34/tZR7 final titration points were superimposed (Figure 6.7), revealing no significant difference between the two titin-ZR7 constructs on binding Act-EF34. The trimeric complex titration was carried out between Act-EF34, tZR7, and ZASP-PDZ; this strategy reduced the size of the resulting complex, and the assignment is available for Act-EF34 and ZASP-PDZ - allowing straightforward analysis. Act-EF34 was produced as a  $^{15}\text{N}$ -labelled protein, as it is the common binding partner of ZASP-PDZ and tZR7. A 2:1 ratio of tZR7 was added to  $^{15}\text{N}$ -Act-EF34, followed by titrating ZASP-PDZ to a final Act-EF34/ZASP-PDZ ratio of 2:3 (Figure 6.8). ZASP-PDZ appears to interact with the C-terminal of Act-EF34 between the 0 and -6 residue positions (above 0.2  $\Delta\delta_{\text{weighted}}$ ) with the biggest changes observed occur for residues G69, S71, D72, and L73 ( $\Delta\delta_{\text{weighted}}$  changes above 0.4) (Figure 6.9).

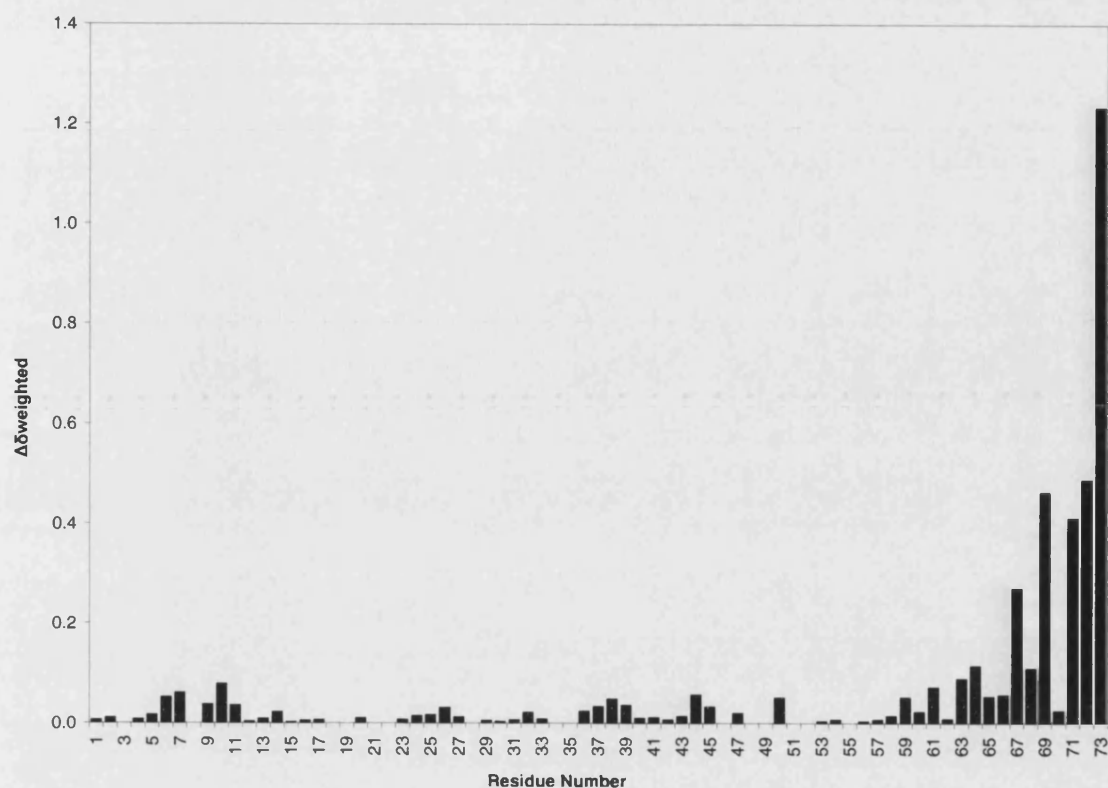




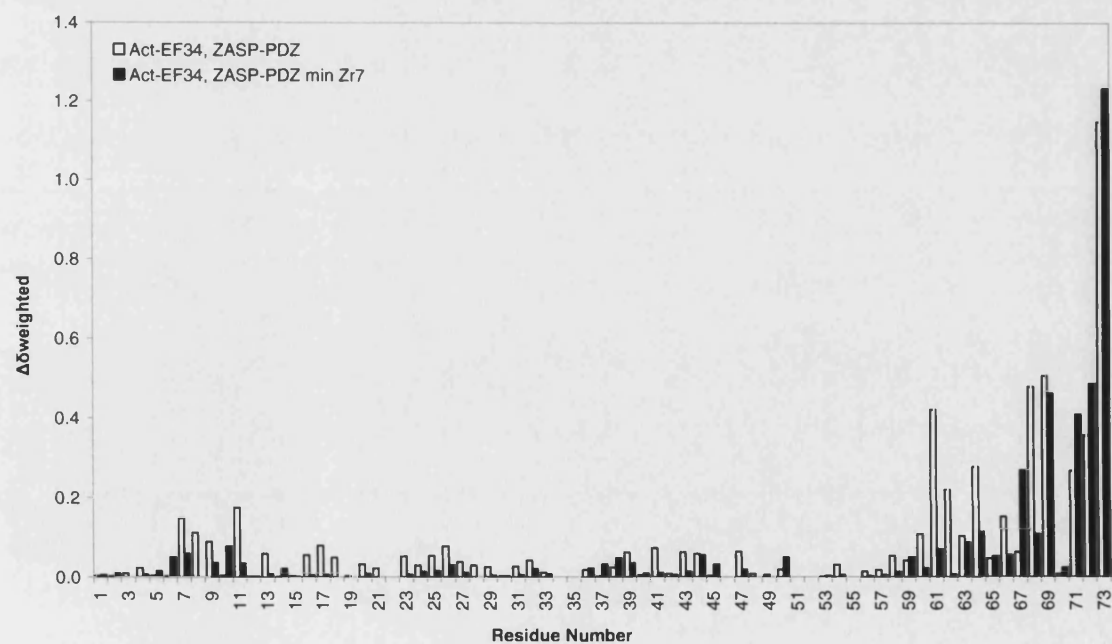
**Figure 6.7:**  $^{15}\text{N}$ -HSQC of  $^{15}\text{N}$ -Act-EF34 with tZR7 (black) and ZR7 (red). Both spectra were run at 600 MHz, 300K, and with an excess of 2:1 ZR7 or tZR7.



**Figure 6.8:**  $^{15}\text{N}$ -HSQC of  $^{15}\text{N}$ -Act-EF34/tZR7 with (red – final titration point) and without (black) ZASP-PDZ. Run at 500 MHz, 300K, 0.53 mM (initial)  $^{15}\text{N}$ -Act-EF34 with a 2:1 excess of tZR7 and 1.5:1 excess of ZASP-PDZ



**Figure 6.9:**  $\Delta\delta_{\text{weighted}}$  changes for  $^{15}\text{N}$ -Act-EF34/tZR7, after binding ZASP-PDZ.

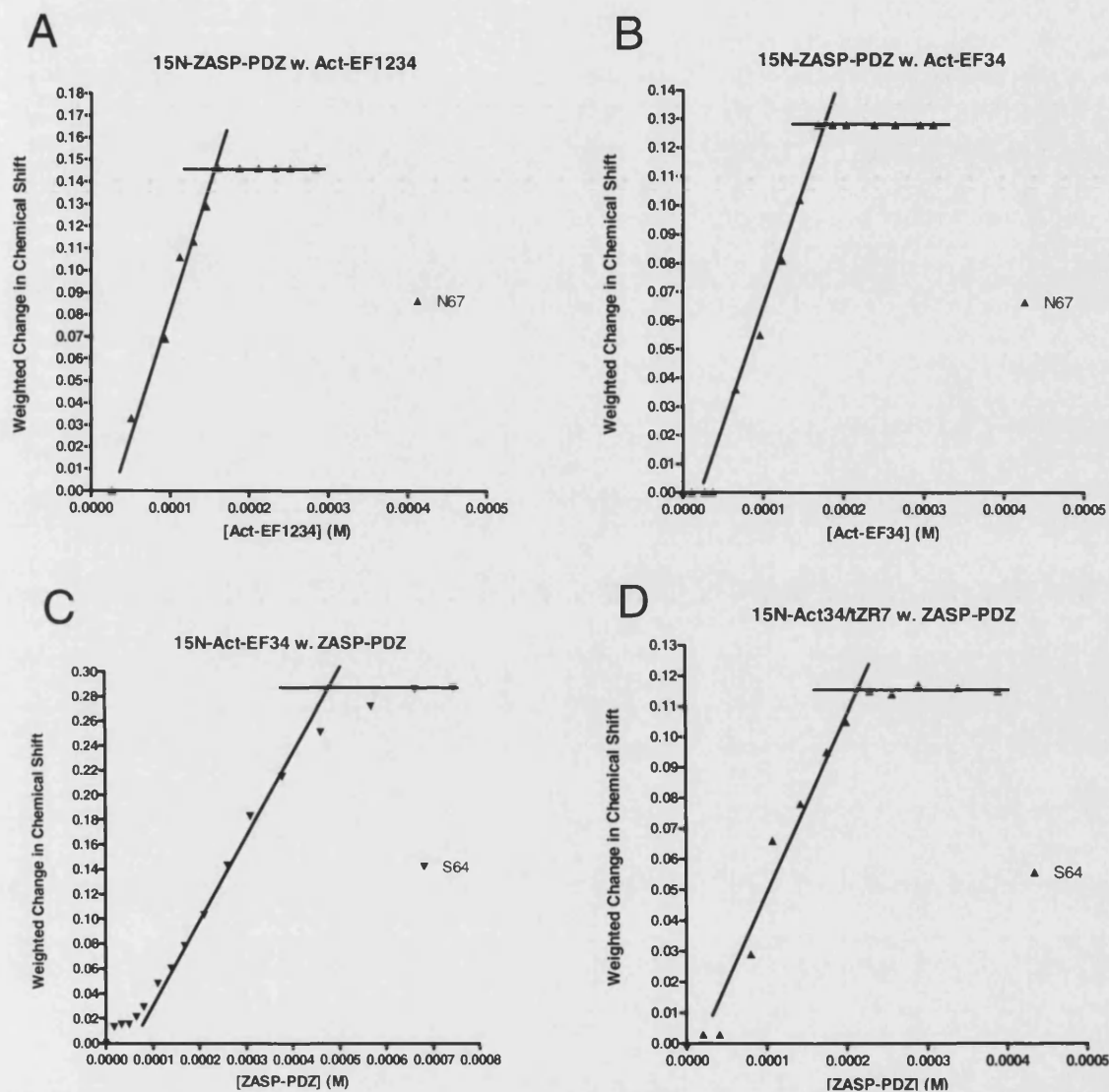


**Figure 6.10:** Comparison of the changes in  $\Delta\delta_{\text{weighted}}$  against residue number for the  $^{15}\text{N}$ -Act-EF34/titin-ZR7 complex, after binding ZASP-PDZ.

#### **6.1.4 NMR Kd calculations between ZASP, $\alpha$ -actinin-2, and Titin**

Titration experiments were carried out on  $^{15}\text{N}$ -ZASP-PDZ's with both unlabelled Act-EF1234 and Act-EF34,  $^{15}\text{N}$ -Act-EF1234 and  $^{15}\text{N}$ -Act-EF34 with unlabelled ZASP-PDZ, and the trimeric complex ( $^{15}\text{N}$ -Act-EF34 saturated with tZR7, titrating in unlabelled ZASP-PDZ). Kd calculations could not be performed between  $^{15}\text{N}$ -Act-EF1234 and unlabelled ZASP-PDZ as all peaks were in intermediate exchange, so did not produce data suitable for the calculation. In all NMR experiments, a stoichiometric binding regime was observed – this is indicated by a linear start of the binding curve, followed by sharp a plateau (Figure 6.11). In stoichiometric binding there is no free ligand, as all added ligand forms a complex with the enzyme. Kd measurements from peaks displaying stoichiometric binding is inappropriate, as this will lead to an overestimation of the Kd by at least a factor of ten (personal communication from Dr. Berry Birdsall): only a limiting value can be determined. Kd calculations for the experiments ranged between  $\sim 200\text{-}500\ \mu\text{M}$ , hence the Kd is expected to be tighter than  $\sim 20\text{-}50\ \mu\text{M}$  as an estimate; any comparison between Kd's from different titrations will be erroneous, as the overestimation is not accurately known. Reducing enzyme concentration (i.e. the labelled component) to the Kd observed in the fluorometric titration ( $\sim 30\ \mu\text{M}$ ) may move the reaction to a non-stoichiometric binding regime, however, this will reduce the signal observed in the NMR experiment.

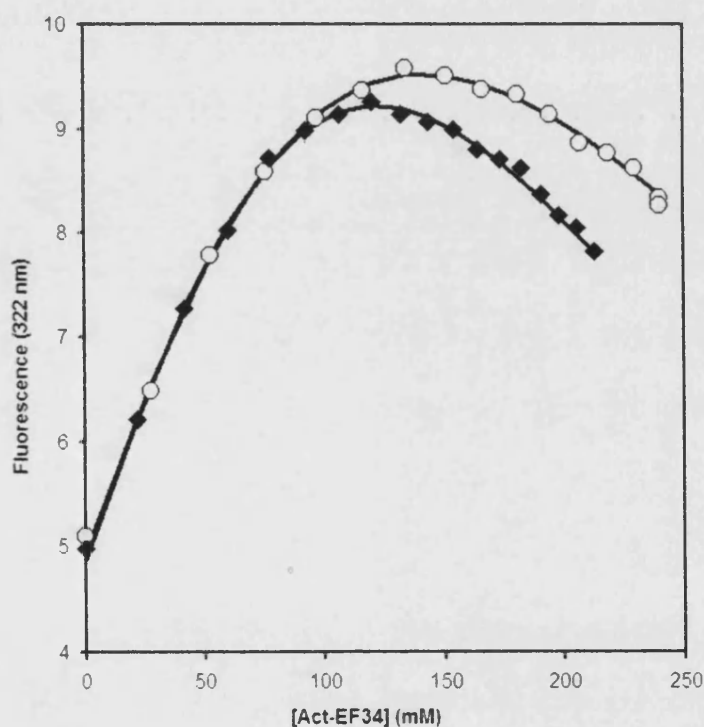
Many limitations in the NMR titrations make it difficult to rely on the accuracy of these Kd calculations. The large amount of protein required for NMR protein/protein titrations mean that ligand concentrations can not be taken to a high excess of the concentration of the fixed protein, required for accurate Kd measurement. Also, the large amount of material required makes it impractical to make repeat titrations. A significant cause of experimental error can derive from the inaccuracy of protein concentration determination and from sample manipulation during the ligand additions. Due to these limitations, the Kd calculations from the fluorometric titration were favoured over the NMR derived values.



**Figure 6.11:** Binding graphs of A)  $^{15}\text{N}$ -ZASP-PDZ against Act-EF1234 (from N67), B)  $^{15}\text{N}$ -ZASP-PDZ against Act-EF34 (from N67), C)  $^{15}\text{N}$ -Act-EF34 against ZASP-PDZ (from S64), and D)  $^{15}\text{N}$ -Act-EF34/tZR7 against ZASP-PDZ (from S64). All graphs display stoichiometric binding, leading to an overestimation of the  $K_d$  calculated by non-linear regression analysis.

**6.2 Fluorometric Titration**

Dr. Steve Martin at the NIMR carried out the fluorometric titration. Both Act-EF34 and the Act-EF34/titin-ZR7 complex were titrated against ZASP-PDZ, producing a titration curve from which the  $K_d$  was calculated (Figure 6.12). The only tryptophan residue from which the change in fluorescence was monitored (W13) is serendipitously located in the binding pocket (Figure 6.13A). The titration between Act-EF34 against ZASP-PDZ yielded a  $K_d$  of  $35.8 \pm 5.3 \mu\text{M}$ , whilst the titration between the Act-EF34/tZR7 complex and ZASP-PDZ gave a  $K_d$  of  $31.6 \pm 3.8 \mu\text{M}$ . From these results, a Student's T test reveals that there is no significant difference between these two results, meaning that the  $K_d$  between ZASP-PDZ and Act-EF34 does not appear to be affected in the presence of tZR7.



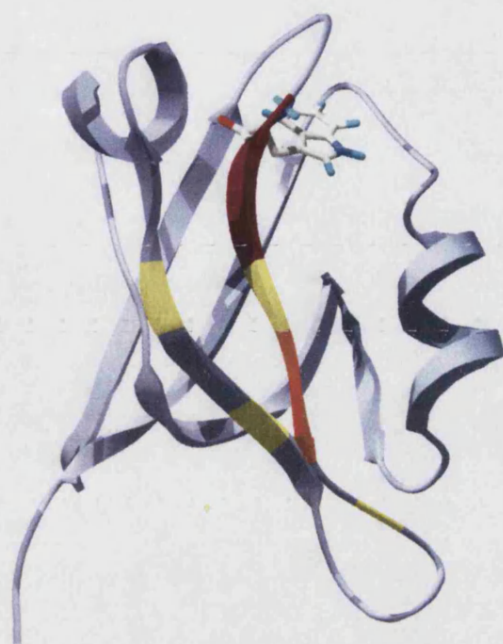
**Figure 6.12:** Fluorescence titration of ZASP-PDZ against Act-EF34 (black diamonds), and the Act-EF34/tZR7 complex (white open circles).

### **6.3 Modelling of $\alpha$ -Actinin, Titin and ZASP Tertiary Complex**

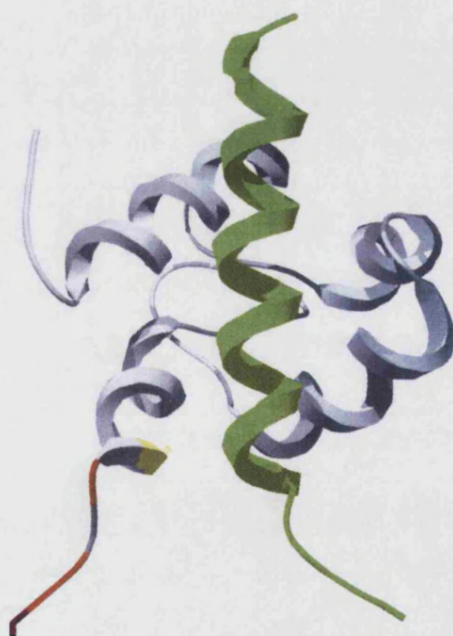
The relatively low affinity between ZASP and  $\alpha$ -actinin-2 prevented the direct elucidation of the complex; the interacting residues were in intermediate exchange, meaning NOE restraints between the complex could not be observed. Therefore we resorted to comparative modelling. From the titration data, the affected peaks were highlighted on the lowest energy structure of ZASP-PDZ from RXCANDID (Section 5.3 and Figure 6.13A) – this was compatible with a class I PDZ to ligand interaction as they include the ‘GLGF’ motif residues within the groove between the  $\beta 2$  and the  $\alpha 2$ . The residues that were affected on Act-EF34 were highlighted on the Act-EF34/ZR7 structure (Figure 6.13B), which was previously solved in our group (Atkinson et al., 2001). This was also consistent with the class I PDZ target proteins, as they were at the C-terminus. The Act-EF34/tZR7/ZASP-PDZ complex was modelled on the third PDZ domain from PSD-95 in complex with a C-terminal peptide derived from Cript (1BE9) (Doyle et al., 1996) – i.e. a class I PDZ domain bound to its target peptide. First, the last four residues on Act-EF34 were rebuilt in an extended conformation compatible with class I PDZ recognition. Secondly, ZASP-PDZ was superimposed onto the structure of 1BE9, using the regions of maximal superposition as indicated by DALI (Holm and Sander, 1996b). Finally, the last four residues of Act-EF34 were aligned with the interacting peptide in 1BE9 - aligning L73 of Act-EF34 to position 0 of the peptide (position 0 is the C-terminal residue, and subsequent residues are -1, -2, etc). The resulting structure appeared to have no apparent clashes, so a GROMOS energy minimization (<http://www.igc.ethz.ch/gromos/>) was applied on the model to correct any distorted geometries arising from the manipulation of the structures (Figure 6.14).



A

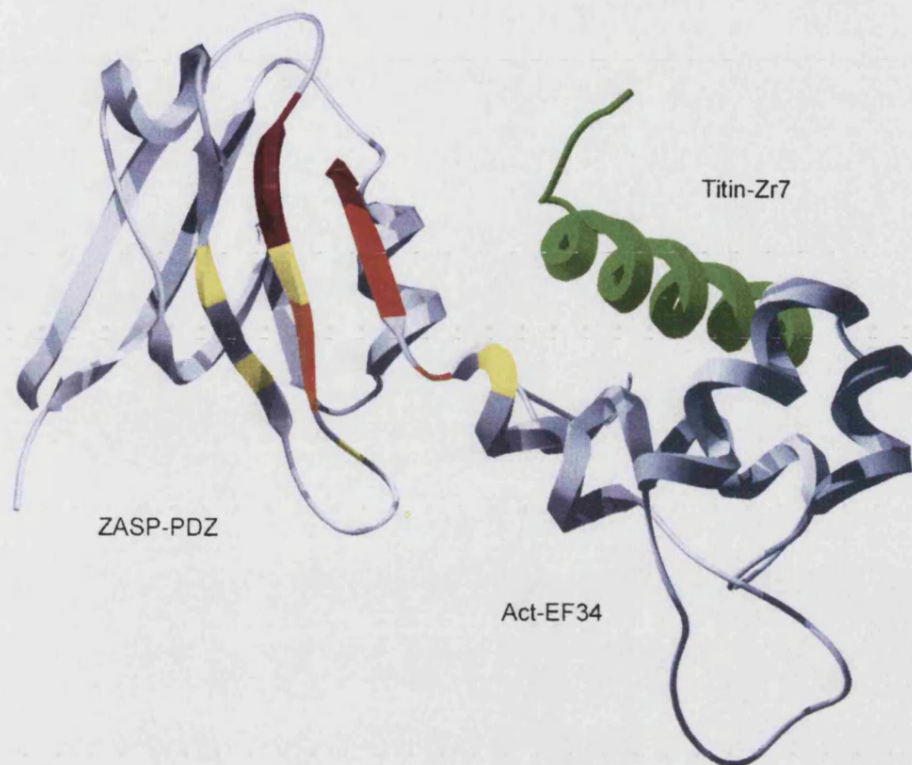


B

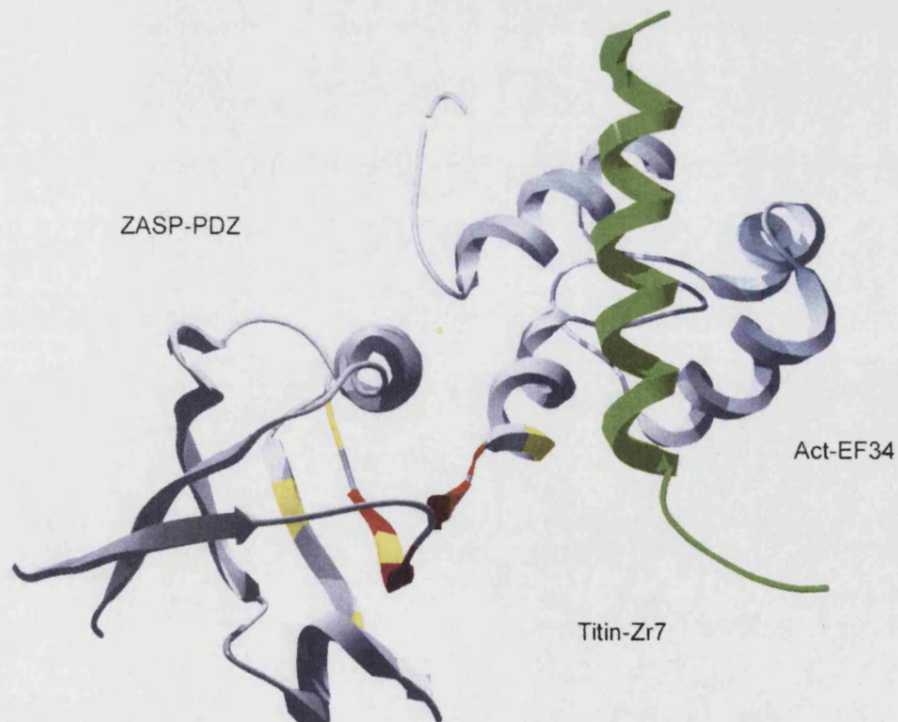


**Figure 6.13:** A) ZASP-PDZ with highlighted  $\Delta\delta_{\text{weighted}}$  changes from binding Act-EF34 with residues in intermediate exchange in purple (13, 14, 15),  $\Delta\delta_{\text{weighted}}$  changes of over 0.3 in red (17, 18), and over 0.2 in yellow (16, 20, 28, 31). W13 is shown in stick representation. B) Act-EF34 (blue) in complex with titin-ZR7 (green) showing changes in its changes in the C-terminus with the same colour scheme as above: in purple - L73; red - G69, S71 and D72, and yellow - L67.

A



B



**Figure 6.14:** Ribbon representation of the Act-EF34/tZR7/ZASP-PDZ trimeric complex, rotated to highlight the ZASP-PDZ/Act-EF34 interaction (A) and the Act-EF34/ZR7 interaction (B).



## **6.4 Conclusion**

We took a modular approach to study the interaction of ZASP and  $\alpha$ -actinin-2 by NMR; the domains of interest were analysed as opposed to the full protein. The limitations become apparent looking at the size of full ZASP and  $\alpha$ -actinin-2 in complex (134 Kda) – a size currently beyond the scope of NMR. In combination with titin for the study of the trimeric complex adds a further 3 MDa.

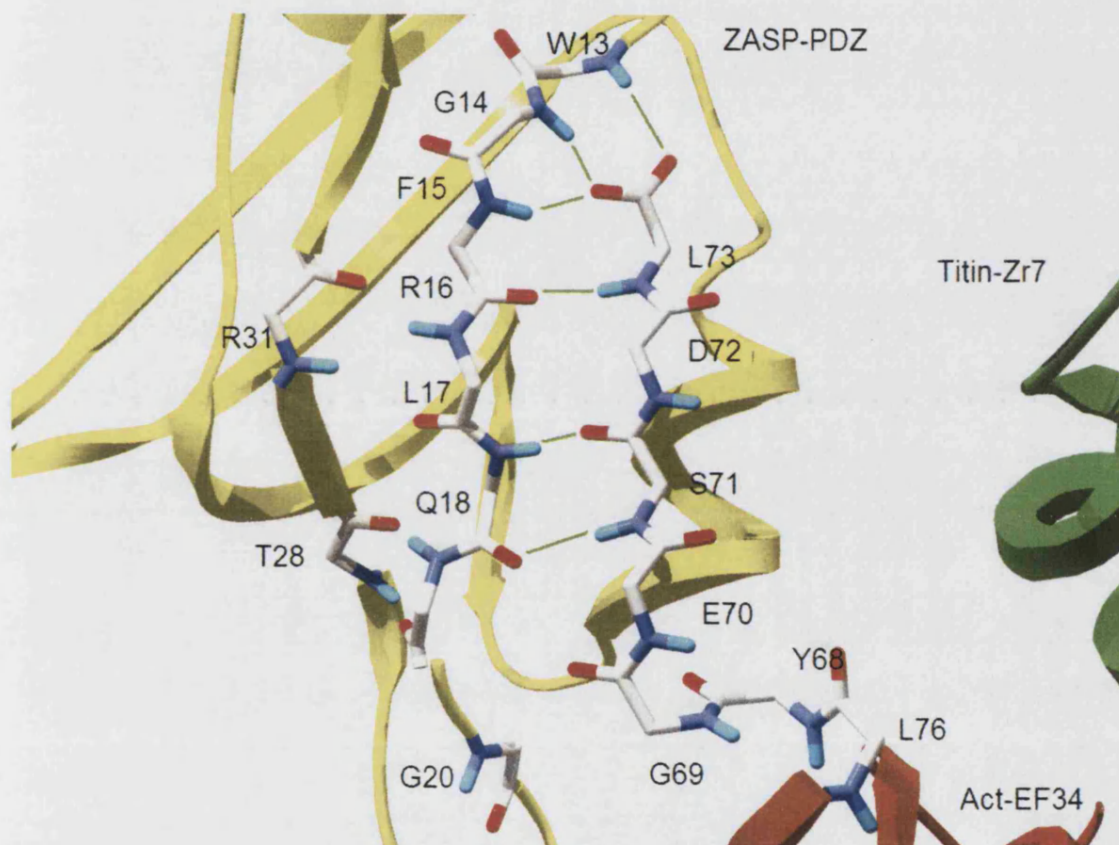
The titration indicates major chemical shift changes of  $\alpha$ -actinin-2's last three C-terminal residues (S71, D72, L73) changing upon binding ZASP-PDZ. These changes are observed for both Act-EF1234 and Act-EF34, signifying that the binding between ZASP-PDZ and Act-EF1234/EF34 is essentially the same. Other significant changes in  $\Delta\delta_{\text{weighted}}$  upon complex formation include A61, F62, S64, Y68, and G69 of Act-EF34: the structure of Act-EF34 in complex with ZR7 (Atkinson et al., 2001) place these residues on one face of the last helix of Act-EF34, suggesting possible secondary interaction with ZASP-PDZ. In total, the interaction with ZASP-PDZ affects the 0 to the -12 residue position of  $\alpha$ -actinin-2 with no affect observed on Act-EF12 (the 0 position is the last C-terminal residue, counting back to negative positions to the N-terminus). The changes observed on the interaction of Act-EF34 and ZASP-PDZ appear to be altered in the presence tZR7: changes above a  $\Delta\delta_{\text{weighted}}$  of 0.2 in the trimeric complex only has significant changes from 0 to the -6 position. This may derive from the  $\alpha$ -helical structural stabilisation caused by the presence of ZR7 observed in a previous study (Atkinson et al., 2000a), however the fluorometric titration result suggests that this is only a minor modification to the interaction between Act-EF34 and ZASP-PDZ as there is no significant difference in affinity between the two components.

The titration between  $^{15}\text{N}$ -ZASP-PDZ and Act-EF1234 showed the biggest changes between residues 13-18, which encompasses the 'GLGF motif' (PWGF in ZASP) that is known to co-ordinate the C-terminal carboxylate group of the target protein in class I PDZ domains (Doyle et al., 1996). Residues 13-18 and 20 correspond to the second  $\beta$ -strand ( $\beta 2$ ), which is modelled to form an intermolecular  $\beta$ -sheet with the target protein. The changes in residues 28 and 31, residues within  $\beta 3$ , were probably caused by the effect of  $\alpha$ -actinin-2 on  $\beta 2$  to which  $\beta 3$  pairs with.

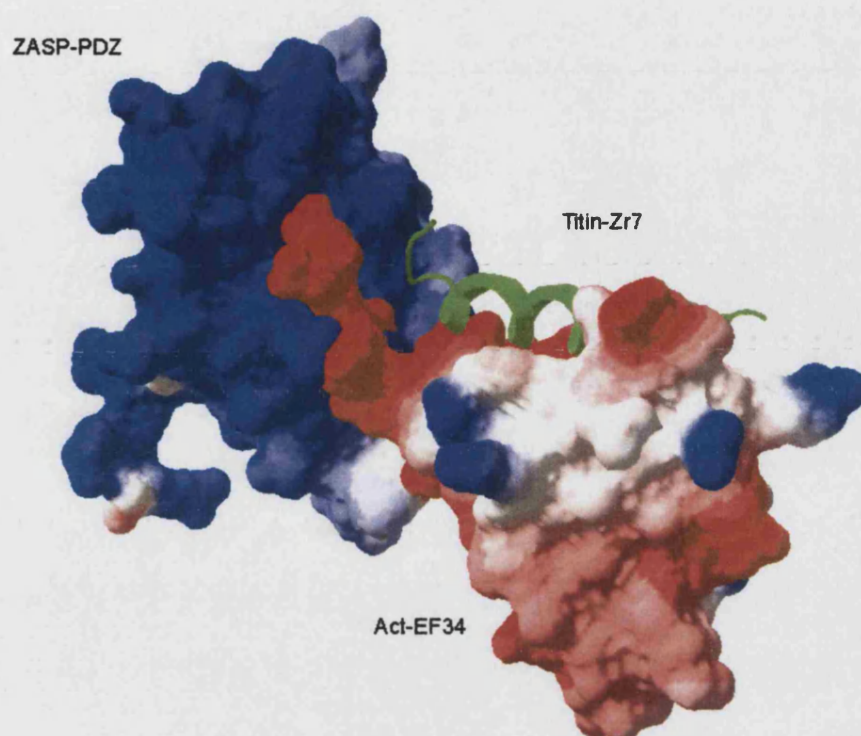
The fluorometric titration gives a  $K_d$  of  $35.8 \pm 5.3 \mu\text{M}$  between Act-EF34 and ZASP-PDZ, and a  $K_d$  of  $31.6 \pm 3.8 \mu\text{M}$  between Act-EF34/tZR7 and ZASP-PDZ. The lack of significance between these  $K_d$  values suggests no effect to the affinity between Act-EF34 and ZASP-PDZ with the presence of tZR7, which indicates that titin and

ZASP bind to  $\alpha$ -actinin-2 in a non-competitive fashion – i.e. they occupy different binding sites.

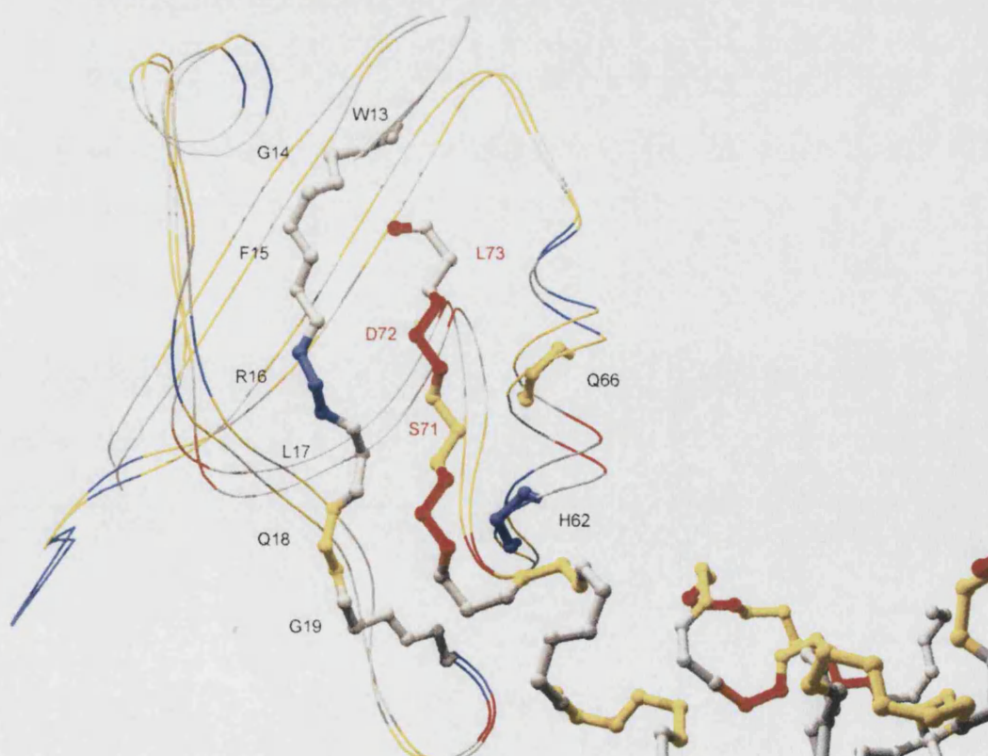
The tertiary complex demonstrates a conventional PDZ to target interaction, with Act-EF34's C-terminus forming an intermolecular  $\beta$ -sheet with ZASP-PDZ's second  $\beta$ -strand and the carboxyl-group of Act-EF34 is co-ordinated within the 'GLGF motif'. An inspection of the model of the tertiary complex reveals six hydrogen bonds forming between Act-EF34 and ZASP-PDZ (Figure 6.15): W13, G14, and F15 backbone amides form a hydrogen bond network with the C-terminal carboxylate group, F15's backbone carboxy group pairs with the 0 position (L73) backbone amide, while L17's carboxy and amide backbone groups pair with -2 (S71) backbone amide and carboxy group respectively. The electrostatic potential at pH 7 of ZASP-PDZ and Act-EF34 was mapped onto the Connolly surface (Figure 6.16). This shows a complementarity between the surface of interaction, with ZASP-PDZ displaying a positive charged surface accommodating the negative charged Act-EF34 C-terminus. An analysis by residue property (Figure 6.17), again highlighting a direct complementarity between the residues at the interacting surface: ZASP W13, G14 and F15 forms a hydrophobic contact with the 0 position (L73) Act-EF34, ZASP R16 forms a electrostatic interaction with the -1 position (D72), Q66 forms a polar interaction with the -2 position (S71), and finally, H62 forms another electrostatic interaction with the -3 position (E70).



**Figure 6.15:** Hydrogen bond network between ZASP-PDZ (yellow ribbons) and Act-EF34 (stick to red ribbon representation) in the complex, forming six possible hydrogen bonds.



**Figure 6.16:** Connolly surface representation (solvent accessibility; probe radius of 1.4 Å) of the trimeric complex. This shows electrostatic and surface complementarity between ZASP-PDZ and Act-EF34.



**Figure 6.17:** Interacting surface between ZASP-PDZ and Act-EF34 coloured by residue property at pH 7: Grey indicates non-polar hydrophobic, yellow - polar, red - negative charge, and blue indicates a positively charged residue.

## Chapter 7

---

# Homology Modelling of the Enigma Family PDZ Domains

---

In general, structure determination is a difficult and laborious process; therefore one of the ways to use available structures to predict information from proteins with unknown structure is 'homology modelling', also known as comparative modelling. Homology modelling is implemented on protein sequences that are homologs of the known protein structure, and one of the main uses of homology modelling is to infer whether related proteins share similar structural features, which may imply comparable functional characteristics. Paradoxically, they can also be used to aid structure determination, as has been utilised in the structure determination of ZASP-PDZ.

Homology modelling of the structure of an undetermined target protein requires a known molecular structure of second protein known as the 'template', with which it shares greater than 30% sequence identity. The procedure involves an alignment of the protein sequences, followed by the direct substitution of the homologous protein onto the template, after which an energy minimisation is performed to correct anomalies in the generated model, such as clashes, etc.

The solution structure of ZASP-PDZ presented in this thesis is the first available structure of an enigma family PDZ domain, so homology models of the other enigma family PDZ domains were produced and used to identify the significance of the conserved residues across the enigma family. This may have implications as to the function of this family of proteins, in particular, their ability to bind  $\alpha$ -actinin-2.

### **7.1 Homology modelling of the enigma family of PDZ**

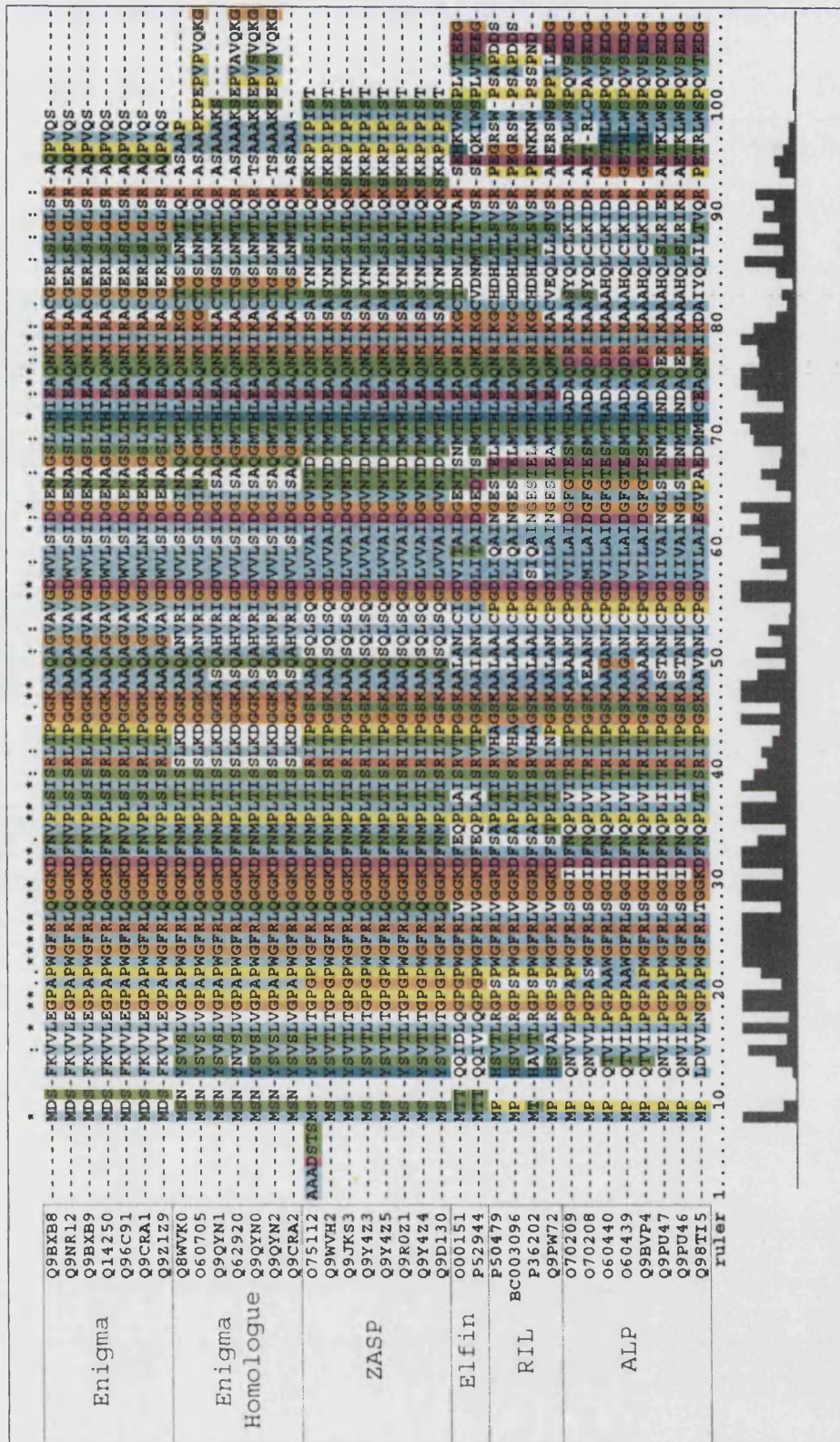
A BLAST search (Altschul et al., 1997) using the ZASP's PDZ sequence recovered fifty sequences. This search revealed that the PDZ domains in all variants of ZASP, cypher, and oracle (see Introduction 1.4) are 100% identical (residues 1-82). Non-mammalian and hypothetical protein sequences were discarded from further analysis to concentrate on the known mammalian proteins. The remaining 36 sequences were then aligned using CLUSTALX (Thompson et al., 1997), which grouped the sequences into their respective sub-groups within the enigma family: enigma, ENH, RIL, ALP, and elfin (Figure 7.1). There was no available structural information regarding any of these sequences at time of writing.

The enigma family PDZ domains have high sequence identity with ZASP-PDZ, with pair-wise sequence identity varying between 54 and 61%, and pair-wise sequence similarity between 71 and 81% (Table 7.1). There was little or no difference in sequence identity within each sub-group, so one representative sequence per group was used for homology modelling. These representative sequences were chosen on criteria such as species and tissue type of origin, and highest sequence identity to ZASP: the PDZ domain sequences in the enigma sub-group family are identical, while ENH1 was chosen as it was the main functioning protein, and each of the human forms of elfin, ALP and RIL were selected. All selected proteins are expressed in mammalian heart and/or striated muscle with the exception of RIL, which has only been reported in epithelial tissue (Cuppen et al., 1998). The SWISS-MODEL server was used to construct the homology models by using the structure of the ZASP-PDZ, accompanied with the alignment of the homologous sequence (Figure 7.2).

Name	Species (# isoforms)*	Tissue **	$\alpha$ -Actinin-2 Binding	Loci on $\alpha$ -actinin	Locus	ID (%)	Sim. (%)
ZASP	<i>H(4), M(4)</i>	H, S	Yes	C-term	Z-disk	100	100
Enigma	<i>H(5), M(2), R(1)</i>	S	NR	NR	Z & I band	56.5	78.8
ENH	<i>H(2), M(3), R(1)</i>	H, S	Yes	NR	Z-disk	60.0	81.2
ALP	<i>H(3), M(1), R(1), S(1), G(2)</i>	H, S	Yes	Spectrin	Z-disk	54.1	76.5
RIL	<i>H(2), R(1), G(1)</i>	E	NR	NR	Epithelial	56.5	71.8
Elfin	<i>H(1), R(1)</i>	H	Yes	C-term	Z-disk	61.2	78.8

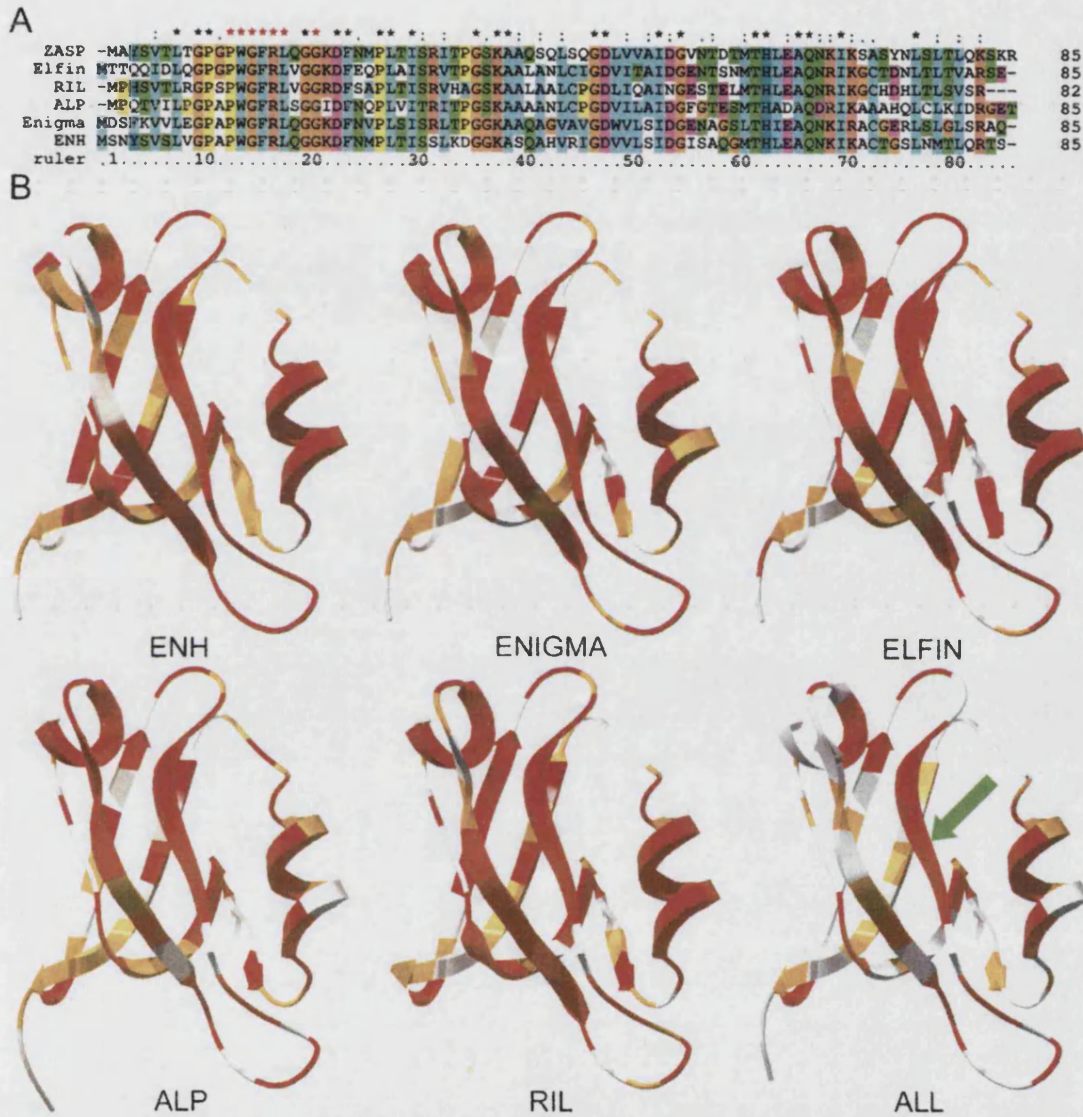
**Table 7.1:** Details of the enigma family members including species, tissue type, ability to bind  $\alpha$ -actinin-2, location of  $\alpha$ -actinin-2 binding, and sequence identity and similarity with respect to ZASP-PDZ. \* *H* = *Homo Sapiens*, *M* = *Mus Musculus*, *R* = *Rattus Norvegicus*, *S* = *Salmo Salar*, *G* = *Galus Galus*, \*\* *H* = Heart, *S* = Skeletal, Epithelial. NR = Not Reported





**Figure 7.1:** ClustalX sequence alignment of the results from the BLAST search from ZASP's PDZ domain. The subgroups in the enigma family (enigma, ENH, ZASP, Elfin, RIL and ALP) group together, and are indicated on the alignment. There is almost no variation in sequence in the PDZ domains of each subgroup.





**Figure 7.2:** Homology models of the PDZ domains of ENH, enigma, elfin, ALP and RIL, generated using SWISS-PDB Viewer and SWISS-MODEL – ‘ALL’ refers to the sequence conservation across the whole enigma family. Highlighted red and yellow residues represent identical and similar residues respectively. The green arrow on ‘ALL’ indicates the binding pocket in class I PDZ target recognition.

## **7.2 Conclusion**

The BLASTP search revealed that the PDZ domains of *enigma*, *ENH*, *ALP*, *RIL*, and *elfin* protein sequences are homologous to that of *ZASP*-PDZ, therefore, homology modelling may help identify structural or functionally relevant residues across the *enigma* family. The sequence alignment of the *enigma* family revealed 27 identical and 44 similar residues (31% and 51% respectively, over the aligned 86 residues), and all generated structures appear to be feasible folds, as there are no large sections of unconserved residues that might disrupt the fold (Figure 7.2). Comparing these homology models with the known structure of the PSD-95/Cript complex (Doyle et al., 1996), show that residues 13-15 of the PDZ form hydrogen bonds with the target's C-terminal carboxylate group, and residues 16 and 19 form additional hydrogen bonds with the target; these residues are well conserved in the *enigma* family. Histidine 62, which is thought to be important for target specificity in PSD-95 PDZ, is also conserved across the *enigma* family. Other conserved residues along the PDZ domain's sequence may serve a more structural purpose. For example, K37 and A38 appear to stabilise the orientation of the 'GLGF' loop. Other similar residues appear to be involved in maintaining the PDZ fold, as they appear to 'pair-up' along stretches of  $\beta$ -strands, and also residues in  $\alpha 2$  have similarity preserved.

## Chapter 8

---

# Discussion

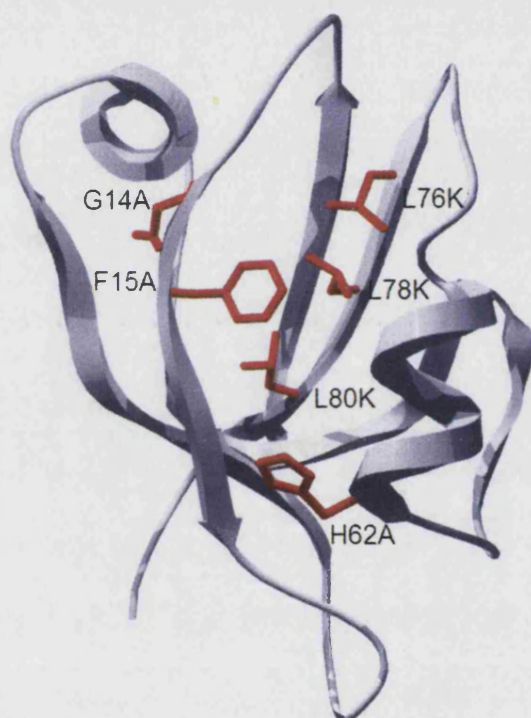
---

This project's main aims were to gain new insights into the muscle ultrastructure by solving the structure of ZASP-PDZ and investigating how it interacts with other proteins that localise in the Z-disk. The last two and four EF-hands of the calmodulin-like domain of  $\alpha$ -actinin-2 and ZASP-PDZ were successfully purified in rich and minimal media – resulting in the near-to-complete NMR assignment of ZASP-PDZ, Act-EF34 backbone assignment, partial assignment of the  $^{15}\text{N}$ -ZASP-PDZ/Act-EF1234 complex, and partial assignment of Act-EF1234. Assignment of ZASP-PDZ allowed its NMR structure calculation by employing a number of protocols, which were subsequently assessed for accuracy, precision, and quality. The high-resolution structure of ZASP-PDZ was then used to model the  $\alpha$ -actinin-2/titin/ZASP-PDZ ternary complex, and also for modelling the PDZ domains in the enigma family of proteins.

The NMR assignments allowed a set of  $^{15}\text{N}$ -HSQC titrations to be completed, characterising the specific interactions between  $\alpha$ -actinin-2, ZASP, and titin. Coupling the NMR and fluorescence titrations with the structures of the components demonstrated that ZASP-PDZ binds to  $\alpha$ -actinin-2 in a class I PDZ interaction, with no significant effect from the presence of tZR7. Subsequent modelling of the  $\alpha$ -actinin-2, titin, and ZASP ternary complex may help us understand how multiple protein interactions contribute towards correct Z-disk function, and it may be useful for further modelling of the Z-disk.

The surface of interaction mapped on ZASP upon complex formation is in excellent agreement with the mutational analysis conducted on cypher, the mouse orthologue of ZASP (Zhou et al., 2001). In this study, residues G14, F15, H62, L76, L78, and L80 of cypher's PDZ were mutated to alanine to test their role in  $\alpha$ -actinin-2 binding (Figure 8.1). The G14A and F15A mutations (within the GLGF motif) abolished binding to  $\alpha$ -actinin-2, consistent with our observation on complex formation. However, the L76K, L78K, and L80K mutants that displayed a reduction in binding affinity to  $\alpha$ -actinin-2 do not appear to be directly involved in the interaction, as they are located on the opposite side of the binding groove. These mutations may only disrupt the PDZ fold, as they are involved in  $\beta$ -sheet formation with  $\beta$ 1. Zhou, et. al. predicted that a H62A

mutation would cause a reduction in affinity for  $\alpha$ -actinin-2 (Zhou et al., 2001) as PSD-95's PDZ domain forms a hydrogen bond; however, the mutation only caused a minor change in affinity to the ligand. It is interesting to note the NMR titration show that L63, directly next to H62, has a large change in chemical shift. The mutational analysis Zhou, et. al. made on  $\alpha$ -actinin-2's C-terminal concurs with the NMR titrations – deletion of these residues lead to abolition of ZASP-PDZ binding. The only human mutation of ZASP linked to DCM (at time of writing) present in the PDZ domain, is the silent mutation V55I – this is a mild mutation as valine and isoleucine are chemically similar.



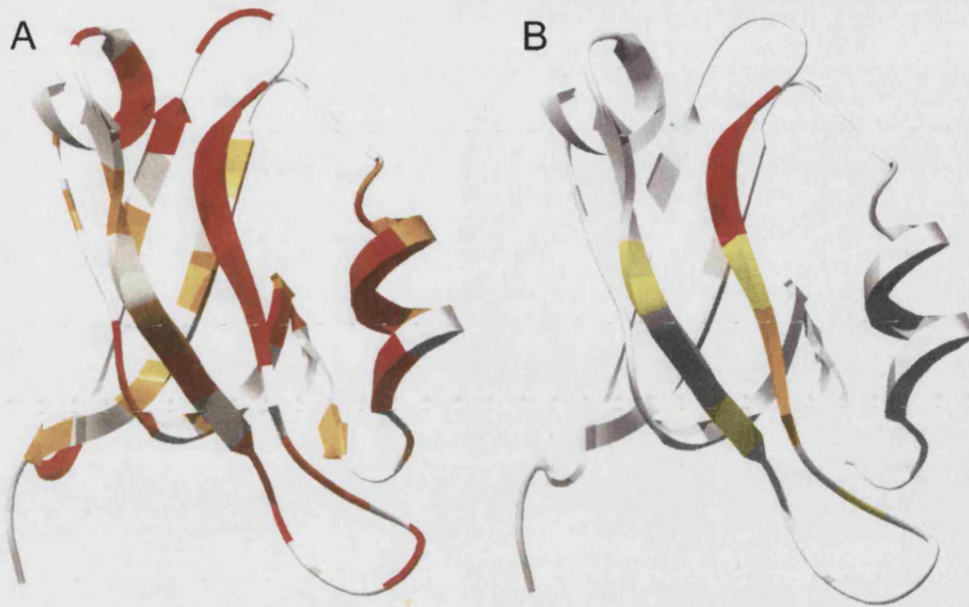
**Figure 8.1:** Mapping of point mutations described in (Zhou et al., 2001) onto the structure of ZASP-PDZ.

As mentioned above, the structure of ZASP-PDZ allowed the homology modelling of the PDZ domains of the other members of the enigma family. As these proteins had high sequence homology and identity with ZASP-PDZ, it is reasonable to suggest functional similarities for these homologs. The large chemical shift variations observed in the  $^{15}\text{N}$ -HSQC of ZASP-PDZ titration with Act-E1234 and Act-EF34 (Section 6.1.2) compare well with the conserved residues in the sequence alignment of the enigma family. The chemical shift changes in 13-17 in the titration parallel the conservation of 13-17. Q18 experiences large  $\Delta\delta_{\text{weighted}}$  upon complex formation, but is

not well conserved. However, the structure of the PSD-95/Cript complex shows only a minor Van der Waals contact between Q18 and the Cript target peptide. Another difference between the titration and the sequence alignment is the absence of large chemical shift change observed in  $\alpha 2$ , with the exception of L63. Again, referring to PSD-95/Cript, the contacts between  $\alpha 2$  and the target are side-chain mediated, therefore reducing the effect on the  $\alpha 2$  amide protons.

Highlighting the conserved residues from the homology models onto the structure of ZASP-PDZ reveal that many of them are in the binding groove (Figure 8.2). These residues, that include W13-L17, G19, G20, D22, F23, H62, A65, Q66, and I69, are also involved in the corresponding residues for PSD-95/Cript recognition (Doyle et al., 1996), suggesting that the PDZ domain of the enigma family may also be able to bind  $\alpha$ -actinin-2's C-terminus. The literature regarding the enigma family proteins contains contradictions regarding their potential binding partners. All of these proteins, with the exception of RIL (epithelial), have been localised to the sarcomeric Z-disk. ZASP, ENH and ALP have been reported to bind  $\alpha$ -actinin-2 via their PDZ domains (Faulkner et al., 1999; Nakagawa et al., 2000; Xia et al., 1997), except for ALP that was reported to bind to  $\alpha$ -actinin-2's spectrin like domains, not the C-terminus. However, the homology model of ALP suggests a PDZ to C-terminal interaction. Enigma has been reported to bind  $\beta$ -tropomyosin, but not  $\alpha$ -actinin-2 (Guy et al., 1999). However, ENH's PDZ domain, sharing a 69% sequence identity to enigma, does not have the ability to bind  $\beta$ -tropomyosin (Nakagawa et al., 2000). The information across elfin/hCLIM1/CLP-36/hCLP-36 is even more confusing. It was reported that hCLIM1 binds to  $\alpha$ -actinin-2 via its LIM domains (Kotaka et al., 2000), hCLP36 binds to  $\alpha$ -actinin-1 via the intervening sequence between the PDZ and LIM domain (Bauer et al., 2000), while CLP-36 was reported to bind to  $\alpha$ -actinin-1 and  $\alpha$ -actinin-4 via its PDZ domain (Vallén et al., 2000). Based on our results and the pattern of sequence conservation, we predict that all currently known members of the family are able to bind  $\alpha$ -actinin-2 *in vitro*. It is possible that *in vivo* other regions of the proteins intervene in a family member specific fashion to modify the signalling properties of the Z-disk through the interaction with other components such as calsarcins and other Z-disk proteins (Frey and Olson, 2002). Further studies will be necessary to test these hypotheses and to assess further the reasons for redundancy of the enigma protein family.





**Figure 8.2:** Ribbon representation of ZASP-PDZ showing A: the conserved residues from the sequence alignment (identical and similar residues in red and orange respectively), and B: the residues most affected upon binding  $\alpha$ -actinin-2 (coloured from red, orange and yellow in order of  $\Delta\delta_{\text{weighted}}$  change – red being the highest). This shows that many of the residues directly involved in binding  $\alpha$ -actinin-2 are conserved within the ZASP family.



## Chapter 9

---

# Bibliography

---

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389-3402.
- Arbustini, E., Pilotto, A., Repetto, A., Grasso, M., Negri, A., Diegoli, M., Campana, C., Scelsi, L., Baldini, E., Gavazzi, A., and Tavazzi, L. (2002). Autosomal dominant dilated cardiomyopathy with atrioventricular block: a lamin A/C defect-related disease. *J Am Coll Cardiol* 39, 981-990.
- Arimura, T., Hayashi, T., Terada, H., Lee, S. Y., Zhou, Q., Takahashi, M., Ueda, K., Nouchi, T., Hohda, S., Shibutani, M., *et al.* (2003). A Cypher/ZASP mutation associated with dilated cardiomyopathy alters the binding affinity to protein kinase C. *J Biol Chem*.
- Atherton, E., and Sheppard, R. C. (1989). *Solid Phase Peptide Synthesis: a Practical Approach (A Practical Approach)*, Oxford University Press).
- Atkinson, R. A., Joseph, C., Dal Piaz, F., Birolo, L., Stier, G., Pucci, P., and Pastore, A. (2000a). Binding of alpha-actinin to titin: implications for Z-disk assembly. *Biochemistry* 39, 5255-5264.
- Atkinson, R. A., Joseph, C., Kelly, G., Muskett, F. W., Frenkiel, T. A., Nietlispach, D., and Pastore, A. (2001). Ca<sup>2+</sup>-independent binding of an EF-hand domain to a novel motif in the alpha-actinin-titin complex. *Nat Struct Biol* 8, 853-857.
- Atkinson, R. A., Joseph, C., Kelly, G., Muskett, F. W., Frenkiel, T. A., and Pastor, A. (2000b). Assignment of the 1H, 13C and 15N resonances of the C-terminal EF-hands of alpha-actinin in a 14 kDa complex with Z-repeat 7 of titin. *J Biomol NMR* 16, 277-278.
- Bach, I. (2000). The LIM domain: regulation by association. *Mech Dev* 91, 5-17.
- Bang, M. L., Mudry, R. E., McElhinny, A. S., Trombitas, K., Geach, A. J., Yamasaki, R., Sorimachi, H., Granzier, H., Gregorio, C. C., and Labeit, S. (2001). Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple roles in Z-disc and I-band protein assemblies. *J Cell Biol* 153, 413-427.

- Baron, M. D., Davison, M. D., Jones, P., and Critchley, D. R. (1987). The structure and function of alpha-actinin. *Biochem Soc Trans* 15, 796-798.
- Bartels, C., Xia, T.-H., Billeter, M., Guntert, P., and Wuthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 5, 1-10.
- Bauer, K., Kratzer, M., Otte, M., de Quintana, K. L., Hagmann, J., Arnold, G. J., Eckerskorn, C., Lottspeich, F., and Siess, W. (2000). Human CLP36, a PDZ-domain and LIM-domain protein, binds to alpha-actinin-1 and associates with actin filaments and stress fibers in activated platelets and endothelial cells. *Blood* 96, 4236-4245.
- Beggs, A. H., Byers, T. J., Knoll, J. H. M., Boyce, F. M., Bruns, G. A. P., and Kunkel, L. M. (1992). Cloning and Characterization of 2 Human Skeletal-Muscle Alpha-Actinin Genes Located on Chromosome-1 and Chromosome-11. *Journal of Biological Chemistry* 267, 9281-9288.
- Birrane, G., Chung, J., and Ladas, J. A. (2003). Novel mode of ligand recognition by the Erbin PDZ domain. *J Biol Chem* 278, 1399-1402.
- Blanchard, A., Ohanian, V., and Critchley, D. (1989). The structure and function of alpha-actinin. *J Muscle Res Cell Motil* 10, 280-289.
- Brodsky, G. L., Muntoni, F., Micić, S., Sinagra, G., Sewry, C., and Mestroni, L. (2000). Lamin A/C gene mutation associated with dilated cardiomyopathy with variable skeletal muscle involvement. *Circulation* 101, 473-476.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* 4, 187-217.
- Carpino, L. A. (1993). 1-Hydroxy-7-Azabenzotriazole - an Efficient Peptide Coupling Additive. *Journal of the American Chemical Society* 115, 4397-4398.
- Carugo, K. D., Banuelos, S., and Saraste, M. (1997). Crystal structure of a calponin homology domain. *Nat Struct Biol* 4, 175-179.
- Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., and Skelton, N. J. (1996). *Protein NMR Spectroscopy: Principles and Practice* (San Diego, CA, Academic Press).
- Codd, M. B., Sugrue, D. D., Gersh, B. J., and Melton, L. J., 3rd (1989). Epidemiology of idiopathic dilated and hypertrophic cardiomyopathy. A population-based study in Olmsted County, Minnesota, 1975-1984. *Circulation* 80, 564-572.

- Cooke, R. (1986). The mechanism of muscle contraction. *CRC Crit Rev Biochem* 21, 53-118.
- Cordier, F., and Grzesiek, S. (1999). Direct observation of hydrogen bonds in proteins by interresidue (3h)J(NC<sup>1</sup>H) scalar couplings. *Journal of the American Chemical Society* 121, 1601-1602.
- Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13, 289-302.
- Cuppen, E., Gerrits, H., Pepers, B., Wieringa, B., and Hendriks, W. (1998). PDZ motifs in PTP-BL and RIL bind to internal protein segments in the LIM domain protein RIL. *Mol Biol Cell* 9, 671-683.
- Dec, G. W., and Fuster, V. (1994). Idiopathic dilated cardiomyopathy. *N Engl J Med* 331, 1564-1575.
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-293.
- Djinovic-Carugo, K., Young, P., Gautel, M., and Saraste, M. (1999). Structure of the alpha-actinin rod: molecular basis for cross-linking of actin filaments. *Cell* 98, 537-546.
- Dominguez, R., Freyzon, Y., Trybus, K. M., and Cohen, C. (1998). Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-power stroke state. *Cell* 94, 559-571.
- Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein- binding domain: molecular basis of peptide recognition by PDZ. *Cell* 85, 1067-1076.
- Dubreuil, R. R. (1991). Structure and evolution of the actin crosslinking proteins. *Bioessays* 13, 219-226.
- Faulkner, G., Lanfranchi, G., and Valle, G. (2001). Telethonin and other new proteins of the Z-disc of skeletal muscle. *IUBMB Life* 51, 275-282.
- Faulkner, G., Pallavicini, A., Comelli, A., Salamon, M., Bortoletto, G., Ievolella, C., Trevisan, S., Kojic, S., Dalla Vecchia, F., Laveder, P., *et al.* (2000). FATZ, a filamin-, actinin-, and telethonin-binding protein of the Z- disc of skeletal muscle. *J Biol Chem* 275, 41234-41242.

- Faulkner, G., Pallavicini, A., Formentin, E., Comelli, A., Ievolella, C., Trevisan, S., Bortoletto, G., Scannapieco, P., Salamon, M., Mouly, V., *et al.* (1999). ZASP: a new Z-band alternatively spliced PDZ-motif protein. *J Cell Biol* 146, 465-475.
- Flood, G., Kahana, E., Gilmore, A. P., Rowe, A. J., Gratzer, W. B., and Critchley, D. R. (1995). Association of structural repeats in the alpha-actinin rod domain. Alignment of inter-subunit interactions. *J Mol Biol* 252, 227-234.
- Flood, G., Rowe, A. J., Critchley, D. R., and Gratzer, W. B. (1997). Further analysis of the role of spectrin repeat motifs in alpha-actinin dimer formation. *Eur Biophys J* 25, 431-435.
- Frey, N., and Olson, E. N. (2002). Calsarcin-3, a Novel Skeletal Muscle-specific Member of the Calsarcin Family, Interacts with Multiple Z-disc Proteins. *J Biol Chem* 277, 13998-14004.
- Frey, N., Richardson, J. A., and Olson, E. N. (2000). Calsarcins, a novel family of sarcomeric calcineurin-binding proteins. *Proc Natl Acad Sci U S A* 97, 14632-14637.
- Gautel, M., Goulding, D., Bullard, B., Weber, K., and Furst, D. O. (1996). The central Z-disk region of titin is assembled from a novel repeat in variable copy numbers. *J Cell Sci* 109 (*Pt 11*), 2747-2754.
- Gerull, B., Gramlich, M., Atherton, J., McNabb, M., Trombitas, K., Sasse-Klaassen, S., Seidman, J. G., Seidman, C., Granzier, H., Labeit, S., *et al.* (2002). Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat Genet* 30, 201-204.
- Goll, C. M., Pastore, A., and Nilges, M. (1998). The three-dimensional structure of a type I module from titin: a prototype of intracellular fibronectin type III domains. *Structure* 6, 1291-1302.
- Goodwin, J. F. (1982). The Frontiers of Cardiomyopathy. *British Heart Journal* 48, 1-18.
- Grunig, E., Tasman, J. A., Kucherer, H., Franz, W., Kubler, W., and Katus, H. A. (1998). Frequency and phenotypes of familial dilated cardiomyopathy. *J Am Coll Cardiol* 31, 186-194.
- Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723.
- Guntert, P. (1998). Structure calculation of biological macromolecules from NMR data. *Quarterly Reviews of Biophysics* 31, 145-237.

- Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273, 283-298.
- Guntert, P., Qian, Y. Q., Otting, G., Muller, M., Gehring, W., and Wuthrich, K. (1991). Structure Determination of the Antp(C39-JS) Homeodomain from Nuclear-Magnetic-Resonance Data in Solution Using a Novel Strategy for the Structure Calculation with the Programs Diana, Caliba, Habas and Glomsa. *Journal of Molecular Biology* 217, 531-540.
- Guntert, P., and Wuthrich, K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J Biomol NMR* 1, 446-456.
- Guy, P. M., Kenny, D. A., and Gill, G. N. (1999). The PDZ domain of the LIM protein enigma binds to beta-tropomyosin. *Mol Biol Cell* 10, 1973-1984.
- Herrmann, T., Guntert, P., and Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology* 319, 209-227.
- Hillier, B. J., Christopherson, K. S., Prehoda, K. E., Brecht, D. S., and Lim, W. A. (1999). Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science* 284, 812-815.
- Holm, L., and Sander, C. (1996a). Alignment of three-dimensional protein structures: Network server for database searching. *Computer Methods for Macromolecular Sequence Analysis* 266, 653-662.
- Holm, L., and Sander, C. (1996b). Mapping the protein universe. *Science* 273, 595-602.
- Holm, L., and Sander, C. (1998). Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics* 14, 423-429.
- Holmes, K. C., and Geeves, M. A. (2000). The structural basis of muscle contraction. *Philos Trans R Soc Lond B Biol Sci* 355, 419-431.
- Holmes, K. C., Popp, D., Gebhard, W., and Kabsch, W. (1990). Atomic model of the actin filament. *Nature* 347, 44-49.
- Hooft, R. W. W., Vriend, G., Sander, C., and Abola, E. E. (1996). Errors in protein structures. *Nature* 381, 272-272.
- Hung, A. Y., and Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. *J Biol Chem* 277, 5699-5702.

- Huxley, A. F., and Niedergerke, R. (1954). Structural changes in muscle during contraction; interference microscopy of living muscle fibres. *Nature* 173, 971-973.
- Huxley, H., and Hanson, J. (1954). Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature* 173, 973-976.
- Ichida, F., Tsubata, S., Bowles, K. R., Haneda, N., Uese, K., Miyawaki, T., Dreyer, W. J., Messina, J., Li, H., Bowles, N. E., and Towbin, J. A. (2001). Novel gene mutations in patients with left ventricular noncompaction or Barth syndrome. *Circulation* 103, 1256-1263.
- Improta, S., Politou, A. S., and Pastore, A. (1996). Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* 4, 323-337.
- Johnson, J. A., Gray, M. O., Chen, C. H., and Mochly-Rosen, D. (1996). A protein kinase C translocation inhibitor as an isozyme-selective antagonist of cardiac function. *J Biol Chem* 271, 24962-24966.
- Johnson, R. A., and Palacios, I. (1982). Dilated Cardiomyopathies of the Adult .1. *New England Journal of Medicine* 307, 1051-1058.
- Joseph, C., Stier, G., O'Brien, R., Politou, A. S., Atkinson, R. A., Bianco, A., Ladbury, J. E., Martin, S. R., and Pastore, A. (2001). A structural characterization of the interactions between titin Z- repeats and the alpha-actinin C-terminal domain. *Biochemistry* 40, 4957-4965.
- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., and Holmes, K. C. (1990). Atomic structure of the actin:DNase I complex. *Nature* 347, 37-44.
- Kang, B. S., Cooper, D. R., Devedjiev, Y., Derewenda, U., and Derewenda, Z. S. (2003). Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. *Structure (Camb)* 11, 845-853.
- Keeling, P. J., Gang, Y., Smith, G., Seo, H., Bent, S. E., Murday, V., Caforio, A. L., and McKenna, W. J. (1995). Familial dilated cardiomyopathy in the United Kingdom. *Br Heart J* 73, 417-421.
- King, D. S., Fields, C. G., and Fields, G. B. (1990). A Cleavage Method Which Minimizes Side Reactions Following Fmoc Solid-Phase Peptide-Synthesis. *International Journal of Peptide and Protein Research* 36, 255-266.
- Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics* 14, 51-&.
- Kotaka, M., Kostin, S., Ngai, S., Chan, K., Lau, Y., Lee, S. M., Li, H., Ng, E. K., Schaper, J., Tsui, S. K., *et al.* (2000). Interaction of hCLIM1, an enigma family protein, with alpha-actinin 2. *J Cell Biochem* 78, 558-565.

- Kotaka, M., Lau, Y., Cheung, K., Lee, S. M., Li, H., Chan, W., Fung, K., Lee, C., Waye, M. M., and Tsui, S. K. (2001). Elfin is expressed during early heart development. *J Cell Biochem* 83, 463-472.
- Kozlov, G., Gehring, K., and Ekiel, I. (2000). Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor. *Biochemistry* 39, 2572-2580.
- Labeit, S., and Kolmerer, B. (1995). Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* 270, 293-296.
- Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular Nmr* 8, 477-486.
- Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003a). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19, 315-316.
- Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M., and Nilges, M. (2003b). Refinement of protein structures in explicit solvent. *Proteins* 50, 496-506.
- Liu, G. S., Cohen, M. V., Mochly-Rosen, D., and Downey, J. M. (1999). Protein kinase C-epsilon is responsible for the protection of preconditioning in rabbit cardiomyocytes. *J Mol Cell Cardiol* 31, 1937-1948.
- Lorenz, M., Popp, D., and Holmes, K. C. (1993). Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. *J Mol Biol* 234, 826-836.
- Luther, P. K. (1991). Three-dimensional reconstruction of a simple Z-band in fish muscle. *J Cell Biol* 113, 1043-1055.
- Luther, P. K. (2000a). Three-dimensional structure of a vertebrate muscle Z-band: implications for titin and alpha-actinin binding. *J Struct Biol* 129, 1-16.
- Luther, P. K. (2000b). Three-dimensional structure of a vertebrate muscle Z-band: implications for titin and alpha-actinin binding. *J Struct Biol* 129, 1-16.
- Luther, P. K., Barry, J. S., and Squire, J. M. (2002). The three-dimensional structure of a vertebrate wide (slow muscle) Z-band: lessons on Z-band assembly. *J Mol Biol* 315, 9-20.
- Luther, P. K., Padron, R., Ritter, S., Craig, R., and Squire, J. M. (2003). Heterogeneity of Z-band structure within a single muscle sarcomere: implications for sarcomere assembly. *J Mol Biol* 332, 161-169.
- Luther, P. K., and Squire, J. M. (2002). Muscle Z-band ultrastructure: titin Z-repeats and Z-band periodicities do not match. *J Mol Biol* 319, 1157-1164.



- Lymn, R. W., and Taylor, E. W. (1971). Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry* 10, 4617-4624.
- Maniatis, I., Fritsch, E. F., and Sambrook, J. (1989). *A Laboratory Manual*, 2nd edition (New York, Cold Spring Harbour Press).
- Margossian, S. S., and Lowey, S. (1973a). Substructure of the myosin molecule. 3. Preparation of single-headed derivatives of myosin. *J Mol Biol* 74, 301-311.
- Margossian, S. S., and Lowey, S. (1973b). Substructure of the myosin molecule. IV. Interactions of myosin and its subfragments with adenosine triphosphate and F-actin. *J Mol Biol* 74, 313-330.
- Maruyama, K. (1997). Connectin/titin, giant elastic protein of muscle. *Faseb J* 11, 341-345.
- Maruyama, K., Kimura, S., Ohashi, K., and Kuwano, Y. (1981). Connectin, an Elastic Protein of Muscle - Identification of Titin with Connectin. *Journal of Biochemistry* 89, 701-709.
- Mayans, O., van der Ven, P. F., Wilm, M., Mues, A., Young, P., Furst, D. O., Wilmanns, M., and Gautel, M. (1998). Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* 395, 863-869.
- Mayans, O., Wuerges, J., Canela, S., Gautel, M., and Wilmanns, M. (2001). Structural evidence for a possible role of reversible disulphide bridge formation in the elasticity of the muscle protein titin. *Structure* 9, 331-340.
- Michels, V. V., Moll, P. P., Miller, F. A., Tajik, A. J., Chu, J. S., Driscoll, D. J., Burnett, J. C., Rodeheffer, R. J., Chesebro, J. H., and Tazelaar, H. D. (1992). The frequency of familial dilated cardiomyopathy in a series of patients with idiopathic dilated cardiomyopathy. *N Engl J Med* 326, 77-82.
- Mills, M., Yang, N., Weinberger, R., Vander Woude, D. L., Beggs, A. H., Easteal, S., and North, K. (2001). Differential expression of the actin-binding proteins, alpha-actinin-2 and -3, in different species: implications for the evolution of functional redundancy. *Hum Mol Genet* 10, 1335-1346.
- Momany, F. A., McGuire, R. F., Burgess, A. W., and Scheraga, H. A. (1975). Energy parameters in polypeptides, VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J Phys Chem* 79, 2361-2381.
- Morais Cabral, J. H., Petosa, C., Sutcliffe, M. J., Raza, S., Byron, O., Poy, F., Marfatia, S. M., Chishti, A. H., and Liddington, R. C. (1996). Crystal structure of a PDZ domain. *Nature* 382, 649-652.

- Muhlhahn, P., Zweckstetter, M., Georgescu, J., Ciosto, C., Renner, C., Lanzendorfer, M., Lang, K., Ambrosius, D., Baier, M., Kurth, R., and Holak, T. A. (1998). Structure of interleukin 16 resembles a PDZ domain with an occluded peptide binding site. *Nature Structural Biology* 5, 682-686.
- Nakagawa, N., Hoshijima, M., Oyasu, M., Saito, N., Tanizawa, K., and Kuroda, S. (2000). ENH, containing PDZ and LIM domains, heart/skeletal muscle-specific protein, associates with cytoskeletal proteins through the PDZ domain. *Biochem Biophys Res Commun* 272, 505-512.
- Nakayama, S., and Kretsinger, R. H. (1994). Evolution of the Ef-Hand Family of Proteins. *Annual Review of Biophysics and Biomolecular Structure* 23, 473-507.
- Nemethy, G., Pottle, M. S., and Scheraga, H. A. (1983). Energy Parameters in Polypeptides .9. Updating of Geometrical Parameters, Nonbonded Interactions, and Hydrogen-Bond Interactions for the Naturally-Occurring Amino-Acids. *Journal of Physical Chemistry* 87, 1883-1887.
- Ohtsuka, H., Yajima, H., Maruyama, K., and Kimura, S. (1997a). Binding of the N-terminal 63 kDa portion of connectin/titin to alpha-actinin as revealed by the yeast two-hybrid system. *FEBS Lett* 401, 65-67.
- Ohtsuka, H., Yajima, H., Maruyama, K., and Kimura, S. (1997b). The N-terminal Z repeat 5 of connectin/titin binds to the C-terminal region of alpha-actinin. *Biochem Biophys Res Commun* 235, 1-3.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4, 2411-2423.
- Panasenko, O. O., and Gusev, N. B. (2001). Mutual effects of alpha-actinin, calponin and filamin on actin binding. *Biochim Biophys Acta* 1544, 393-405.
- Papa, I., Astier, C., Kwiatak, O., Raynaud, F., Bonnal, C., Lebart, M. C., Roustan, C., and Benyamin, Y. (1999). Alpha actinin-CapZ, an anchoring complex for thin filaments in Z-line. *J Muscle Res Cell Motil* 20, 187-197.
- Passier, R., Richardson, J. A., and Olson, E. N. (2000). Oracle, a novel PDZ-LIM domain protein expressed in heart and skeletal muscle. *Mech Dev* 92, 277-284.
- Peitsch, M. C. (1995). Protein Modeling by E-Mail. *Bio-Technology* 13, 658-660.
- Perriere, G., and Gouy, M. (1996). WWW-Query: An on-line retrieval system for biological sequence banks. *Biochimie* 78, 364-369.
- Pfuhl, M., Improt, S., Politou, A. S., and Pastore, A. (1997). When a module is also a domain: The role of the N terminus in the stability and the dynamics of immunoglobulin domains from titin. *Journal of Molecular Biology* 265, 242-256.

- Pfuhl, M., and Pastore, A. (1995). Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the I set. *Structure* 3, 391-401.
- Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-Tailored Excitation for Single-Quantum NMR- Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular Nmr* 2, 661-665.
- Rayment, I., Holden, H. M., Whittaker, M., Yohn, C. B., Lorenz, M., Holmes, K. C., and Milligan, R. A. (1993a). Structure of the actin-myosin complex and its implications for muscle contraction. *Science* 261, 58-65.
- Rayment, I., Rypniewski, W. R., Schmidt-Base, K., Smith, R., Tomchick, D. R., Benning, M. M., Winkelmann, D. A., Wesenberg, G., and Holden, H. M. (1993b). Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* 261, 50-58.
- Roberts, G. C. K. (1993). *NMR of Macromolecules*, Oxford University Press.
- Rost, B., and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* 232, 584-599.
- Ruckert, M., and Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society* 122, 7793-7797.
- Salmikangas, P., Mykkanen, O. M., Gronholm, M., Heiska, L., Kere, J., and Carpen, O. (1999). Myotilin, a novel sarcomeric protein with two Ig-like domains, is encoded by a candidate gene for limb-girdle muscular dystrophy. *Hum Mol Genet* 8, 1329-1336.
- Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000). SMART: A Web-based tool for the study of genetically mobile domains. *Nucleic Acids Research* 28, 231-234.
- Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31, 3381-3385.
- Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* 160, 65-73.
- Skelton, N. J., Koehler, M. F., Zobel, K., Wong, W. L., Yeh, S., Pisabarro, M. T., Yin, J. P., Lasky, L. A., and Sidhu, S. S. (2003). Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain. *J Biol Chem* 278, 7645-7654.

- Spronk, C. A., Linge, J. P., Hilbers, C. W., and Vuister, G. W. (2002). Improving the quality of protein structures derived by NMR spectroscopy. *J Biomol NMR* 22, 281-289.
- Stein, E. G., Rice, L. M., and Brunger, A. T. (1997). Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *Journal of Magnetic Resonance* 124, 154-164.
- Sugrue, D. D., Rodeheffer, R. J., Codd, M. B., Ballard, D. J., Fuster, V., and Gersh, B. J. (1992). The Clinical Course of Idiopathic Dilated Cardiomyopathy - a Population-Based Study. *Annals of Internal Medicine* 117, 117-123.
- Szent-Gyorgyi, A. G. (1953). Meromyosins, the subunits of myosin. *Arch Biochem Biophys* 42, 305-320.
- Takada, F., Vander Woude, D. L., Tong, H. Q., Thompson, T. G., Watkins, S. C., Kunkel, L. M., and Beggs, A. H. (2001). Myozenin: an alpha-actinin- and gamma-filamin-binding protein of skeletal muscle Z lines. *Proc Natl Acad Sci U S A* 98, 1595-1600.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25, 4876-4882.
- Tiso, N., Majetti, M., Stanchi, F., Rampazzo, A., Zimbello, R., Nava, A., and Danieli, G. A. (1999). Fine mapping and genomic structure of ACTN2, the human gene coding for the sarcomeric isoform of alpha-actinin-2, expressed in skeletal and cardiac muscle. *Biochem Biophys Res Commun* 265, 256-259.
- Tjandra, N., Grzesiek, S., and Bax, A. (1996). Magnetic field dependence of nitrogen-proton J splittings in N- 15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *Journal of the American Chemical Society* 118, 6264-6272.
- Tochio, H., Hung, F., Li, M., Bredt, D. S., and Zhang, M. (2000). Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-95. *J Mol Biol* 295, 225-237.
- Towbin, J. A. (1998). The role of cytoskeletal proteins in cardiomyopathies. *Curr Opin Cell Biol* 10, 131-139.
- Towbin, J. A. (1999). Pediatric myocardial disease. *Pediatr Clin North Am* 46, 289-312, ix.
- Towbin, J. A., and Bowles, N. E. (2002). The failing heart. *Nature* 415, 227-233.

- Trinick, J. (1996). Cytoskeleton: Titin as a scaffold and spring. *Current Biology* 6, 258-260.
- Trinick, J., Knight, P., and Whiting, A. (1984). Purification and Properties of Native Titin. *Journal of Molecular Biology* 180, 331-356.
- Vallenius, T., Luukko, K., and Makela, T. P. (2000). CLP-36 PDZ-LIM protein associates with nonmuscle alpha-actinin-1 and alpha-actinin-4. *J Biol Chem* 275, 11100-11105.
- Vatta, M., Mohapatra, B., Jimenez, S., Sanchez, X., Faulkner, G., Perles, Z., Sinagra, G., Lin, J. H., Vu, T. M., Zhou, Q., *et al.* (2003). Mutations in Cypher/ZASP in patients with dilated cardiomyopathy and left ventricular non-compaction. *J Am Coll Cardiol* 42, 2014-2027.
- Walma, T., Spronk, C. A., Tessari, M., Aelen, J., Schepens, J., Hendriks, W., and Vuister, G. W. (2002). Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL. *J Mol Biol* 316, 1101-1110.
- Way, M., Pope, B., and Weeds, A. G. (1992). Evidence for functional homology in the F-actin binding domains of gelsolin and alpha-actinin: implications for the requirements of severing and capping. *J Cell Biol* 119, 835-842.
- Whiting, A., Wardale, J., and Trinick, J. (1989). Does Titin Regulate the Length of Muscle Thick Filaments. *Journal of Molecular Biology* 205, 263-268.
- Wuthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Sons Inc..
- Xia, H., Winokur, S. T., Kuo, W. L., Altherr, M. R., and Bredt, D. S. (1997). Actinin-associated LIM protein: identification of a domain interaction between PDZ and spectrin-like repeat motifs. *J Cell Biol* 139, 507-515.
- Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S., and North, K. (2003). ACTN3 genotype is associated with human elite athletic performance. *Am J Hum Genet* 73, 627-631.
- Ylanne, J., Scheffzek, K., Young, P., and Saraste, M. (2001). Crystal structure of the alpha-actinin rod reveals an extensive torsional twist. *Structure* 9, 597-604.
- Zhou, Q., Chu, P. H., Huang, C., Cheng, C. F., Martone, M. E., Knoll, G., Shelton, G. D., Evans, S., and Chen, J. (2001). Ablation of Cypher, a PDZ-LIM domain Z-line protein, causes a severe form of congenital myopathy. *J Cell Biol* 155, 605-612.

# Chapter 10

## Appendix

### 10.1 Protein Sequences

#### 10.1.1 ZASP [1-85] (ZASP-PDZ)

87 residues, MW = 9267.5 Da, pI = 9.78,  $\epsilon$  = 8250 M<sup>-1</sup> cm<sup>-1</sup>

GA MSYSVTILTGP GPWGFRLQGG KDFNMPLTIS RITPGSKAAQ SQLSQGDLVV AIDGVNTDTM  
THLEAQNKIK SASYNLSLTL QKSKR

#### 10.1.2 $\alpha$ -Actinin-2 [745-894] (Act-EF1234)

154 residues, MW = 17163.2 Da, pI = 4.64,  $\epsilon$  = 7800 M<sup>-1</sup> cm<sup>-1</sup>

G AMGRDAKGIT QEQMNEFRAS FNHFDRRKNG LMDHEDFRAC LISMGYDLGE AEFARIMTLV  
DPNGQGTVTTF QSFIDFMTRE TADTDTAEQV IASFRILASD KPYILAEELR RELPPDQAQY  
CIKRMPAYSG PGSVPGALDY AAFSSALYGE SDL

#### 10.1.3 $\alpha$ -Actinin-2 [823-894] (Act-EF34)

75 residues, MW = 8071.0 Da, pI = 4.34,  $\epsilon$  = 6400 M<sup>-1</sup> cm<sup>-1</sup>

GA MADTDTAEQV IASFRILASD KPYILAEELR RELPPDQAQY CIKRMPAYSG PGSVPGALDY  
AAFSSALYGE SDL

#### 10.1.4 Titin ZR7

53 residues, MW = 5659.3 Da, pI = 5.71,  $\epsilon$  = 3840 M<sup>-1</sup> cm<sup>-1</sup>

GA MGKVGVGKKA EAVATVVAHV DQARVREPRE PGLPEDSYAQ QTTLEYGYKE H

#### 10.1.5 Titin ZR7 minimum sequence

23 residues, MW = 2421.7 Da, pI = 9.98

GKKAEAVATV VAAVDQARVR EPR

**10.2 Media and Buffers****10.2.1 Luria-Bertani Media (LB)**

1 % w/v bacto-tryptone, 0.5 % w/v bacto-yeast extract, 1 % w/v NaCl, 1 mM Tris.HCl

**10.2.2 Preparation of M9 Minimal Media**

Autoclave, and leave to cool before the addition of the following solutions, (\* if appropriate):

Compound	Concentration	Volume (ml)	Final Concentration
MgSO <sub>4</sub>	1 M	1.0	1 mM
CaCl <sub>2</sub>	1 M	0.3	0.3 mM
M9 salts	10 x	100.0	1 x
Trace elements	1000 x	1.0	1 x
Biotin	1 mg/ml	1.0	1 mg /L
Thiamine	1 mg/ml	1.0	1 mg/L
<sup>13</sup> C-glucose*	10.00%	20.0	0.20%
Kanamycin*	30 mg/ml	1 ml	30 mg/L
Ampicillin*	100 mg/ml	1 ml	100 mg/L

**10.2.2.1 10 x M9 Salts with 15N as a Source of Nitrogen**

Compound	Formula Weight (Da)	Final Molarity (M)	Amount (g)/100 ml
Na <sub>2</sub> HPO <sub>4</sub>	141.96	0.423	6.00
KH <sub>2</sub> PO <sub>4</sub>	136.09	0.220	3.00
NaCl	58.44	0.086	0.50
<sup>15</sup> NH <sub>4</sub> Cl	53.49	0.187	1.00

**10.2.2.2 Trace Elements x 1000**

Compound	Formula Weight (Da)	Final Molarity (mM)	Amount (mg)/100 ml
EDTA	397.7	25.00	5ml 0.5M, pH 8 in 80 ml H <sub>2</sub> O
FeCl <sub>3</sub>	162.2	30.83	500
ZnCl <sub>2</sub>	136.3	3.67	50
CuCl <sub>2</sub>	134.5	0.74	10
CoCl <sub>2</sub> ·H <sub>2</sub> O	237.9	0.42	10
H <sub>3</sub> BO <sub>3</sub>	61.83	1.62	10
Na <sub>2</sub> MoO <sub>4</sub> ·2H <sub>2</sub> O	205.9	0.10	2
MnCl <sub>2</sub> ·6H <sub>2</sub> O	125.8	127.19	1600

Adjust to pH 7.5 -8 in 100 ml  
Filter 0.22 µm, Store at 4 °C



**10.2.2.3 Biotin 1 mg/ml**

For 50 ml:

Dissolve 50 mg biotin in 1:1 ethanol/water solution.

Filter 0.22  $\mu$ m, Store at 4 °C

**10.2.2.4 Thiamine 1 mg/ml**

For 50 ml:

Dissolve 50 mg of thiamine into 50 ml water.

Filter 0.22  $\mu$ m, Store at 4 °C

**10.2.2.5 IPTG 0.5 M**

For 50 ml:

Dissolve 5.96g of IPTG into 50 ml water.

Filter 0.22  $\mu$ m, Store at -20 °C in 1 ml aliquots.

**10.2.3 Nickel-NTA Column Buffers**

All buffers are 0.2  $\mu$ m filtered.

**10.2.3.1 Lysis Buffer**

20 mM Tris.HCl pH 8.0, 200 mM NaCl, 5 mM Imidazole, 0.2% v/v IPEGAL CA-630

**10.2.3.2 Wash Buffer**

20 mM Tris.HCl pH 8.0, 200 mM NaCl, 5 mM Imidazole

**10.2.3.3 1M NaCl Wash Buffer**

20 mM Tris.HCl pH 8.0, 1000 mM NaCl, 5 mM Imidazole

**10.2.3.4 30 mM Imidazole Wash Buffer**

20 mM Tris.HCl pH 8.0, 1000 mM NaCl, 30 mM Imidazole

**10.2.3.5 300 mM Imidazole Wash Buffer**

20 mM Tris.HCl pH 8.0, 1000 mM NaCl, 300 mM Imidazole

**10.2.4 FPLC Column Buffer**

20 mM Tris.HCl pH 8.0, 200 mM NaCl, 0.02 mM NaN<sub>3</sub>, 0.2  $\mu$ M filtered and degassed with Nitrogen.

**10.2.5 NMR Sample Buffer**

20 mM Sodium Phosphate Buffer pH 6.6, 0.02 mM NaN<sub>3</sub>, 10% D<sub>2</sub>O.

**10.2.6 Broad Molecular Weight Marker**

Amersham Biosciences: Cat # RPN5800

<b>Protein</b>	<b>Mr</b>	<b>Source</b>
Myosin	205000	Rabbit muscle
$\beta$ -Galactosidase	116000	<i>Escherichia coli</i>
Phosphorylase b	97000	Rabbit muscle
Transferrin	80000	Human serum
Albumin	66000	Bovine serum albumin
Glutamate dehydrogenase	55000	Bovine liver
Ovalbumin	45000	Chicken egg white
Carbonic anhydrase	30000	Bovine erythrocyte
Trypsin inhibitor	21000	Soybean
Lysozyme	14000	Chicken egg white
Aprotinin	6500	Bovine lung

**10.3 NMR Acquisition Parameters****10.3.1 ZASP-PDZ Spectra****10.3.1.1 GM-COSY**

```

-xN      2048      -yN      768      \
-xT      1024      -yT      384      \
-xMODE Complex -yMODE Complex \
-xSW      8000.000 -ySW      8000.000 \
-xOBS      599.924 -yOBS      599.924 \
-xCAR      4.730   -yCAR      4.730   \
-xLAB      H1      -yLAB      H2      \
-ndim      2      -aq2D      States   \

```

**10.3.1.2  $^1\text{H}$ - $^1\text{H}$ -TOCSY**

```

-xN      2048      -yN      768      \
-xT      1024      -yT      384      \
-xMODE Complex -yMODE Complex \
-xSW      6499.838 -ySW      6499.838 \
-xOBS      500.090 -yOBS      500.090 \
-xCAR      4.730   -yCAR      4.730   \
-xLAB      H1      -yLAB      H2      \
-ndim      2      -aq2D      States   \

```

2D-TOCSY spectra were acquired with the above parameters and with these combinations of mixing times and temperatures:

Mixing Time (ms)	Temperature (Celsius)
40	27
50	27
70	10, 27
80	27
100	10, 14, 19, 23, 27

**10.3.1.3  $^1\text{H}$ - $^1\text{H}$ -NOESY**

```

-xN      2048      -yN      768      \
-xT      1024      -yT      384      \
-xMODE Complex -yMODE Complex \
-xSW      6499.838 -ySW      6499.838 \
-xOBS      500.090 -yOBS      500.090 \
-xCAR      4.730   -yCAR      4.730   \
-xLAB      H1      -yLAB      H2      \
-ndim      2      -aq2D      States   \

```

2D-NOESY spectra were acquired with the above parameters and with these combinations of mixing times and temperatures:

Mixing Time (ms)	Temperature (Celsius)
50	27
75	27
100	10, 19, 27
150	10, 27

**10.3.1.4  $^{15}\text{N}$ -HSQC**

-xN	2432	-yN	560	\
-xT	1216	-yT	280	\
-xMODE	Complex	-yMODE	Complex	\
-nSW	8000.000	-ySW	2500.000	\
-xOBS	599.893	-yOBS	60.794	\
-xCAR	4.730	-yCAR	118.088	\
-xLAB	HN	-yLAB	N	\
-ndim	2	-aq2D	States	\

**10.3.1.5  $^{13}\text{C}$ -HSQC**

-xN	2400	-yN	400	\
-xT	1200	-yT	200	\
-xMODE	Complex	-yMODE	Complex	\
-xSW	8000.000	-ySW	10060.36	\
-xOBS	500.087	-yOBS	125.752	\
-xCAR	4.730	-yCAR	40.811	\
-xLAB	HC	-yLAB	C	\
-ndim	2	-aq2D	States	\

**10.3.1.6 HCCH-TOCSY**

-xN	1024	-yN	128	-zN	284	\
-xT	512	-yT	64	-zT	142	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	10559.662	-zSW	8000.000	\
-xOBS	599.891	-yOBS	150.848	-zOBS	599.891	\
-xCAR	4.73	-yCAR	40.931	-zCAR	4.73	\
-xLAB	HC	-yLAB	C	-zLAB	HH	\
-ndim	3	-aq2D	States			\

**10.3.1.7  $^1\text{H}$ - $^{15}\text{N}$ -HSQC-TOCSY: 100 ms Mixing Time**

-xN	1024	-yN	160	-zN	128	\
-xT	512	-yT	80	-zT	64	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	8000.000	-zSW	2027.164	\
-xOBS	500.090	-yOBS	500.090	-zOBS	50.679	\
-xCAR	4.73	-yCAR	4.73	-zCAR	118.089	\
-xLAB	HN	-yLAB	HH	-zLAB	N15	\
-ndim	3	-aq2D	States			\

**10.3.1.8  $^1\text{H}$ - $^{15}\text{N}$ -HSQC-NOESY: 100 ms Mixing Time**

-xN	1024	-yN	168	-zN	128	\
-xT	512	-yT	84	-zT	64	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	8000.000	-zSW	2027.164	\
-xOBS	500.090	-yOBS	500.090	-zOBS	50.679	\
-xCAR	4.73	-yCAR	4.73	-zCAR	118.089	\
-xLAB	HN	-yLAB	HH	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.1.9  $^1\text{H}$ - $^{15}\text{N}$ -HSQC-NOESY: 50 ms Mixing Time**

-xN	1024	-yN	200	-zN	128	\
-xT	512	-yT	100	-zT	64	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	8000.000	-zSW	2431.759	\
-xOBS	599.893	-yOBS	599.893	-zOBS	60.794	\
-xCAR	4.73	-yCAR	4.73	-zCAR	117.941	\
-xLAB	HN	-yLAB	HH	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.1.10  $^{13}\text{C}$ -NOESY**

-xN	1024	-yN	128	-zN	152	\
-xT	512	-yT	64	-zT	76	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	10060.362	-zSW	8000.000	\
-xOBS	500.087	-yOBS	125.752	-zOBS	500.090	\
-xCAR	4.73	-yCAR	40.811	-zCAR	4.73	\
-xLAB	HC	-yLAB	C	-zLAB	HH	\
-ndim	3	-aq2D	States			\

**10.3.1.11 HNCA**

-xN	1024	-yN	164	-zN	128	\
-xT	512	-yT	82	-zT	64	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	4525.400	-zSW	2431.832	\
-xOBS	599.890	-yOBS	150.851	-zOBS	60.794	\
-xCAR	4.730	-yCAR	55.939	-zCAR	117.939	\
-xLAB	H	-yLAB	C	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.1.12 HNCOC**

-xN	1024	-yN	220	-zN	64	\
-xT	512	-yT	110	-zT	32	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	4526.167	-zSW	1945.383	\
-xOBS	599.890	-yOBS	150.851	-zOBS	60.794	\
-xCAR	4.730	-yCAR	55.939	-zCAR	117.939	\
-xLAB	H	-yLAB	C	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.1.13 CBCANH**

-xN	1024	-yN	80	-zN	120	\
-xT	512	-yT	40	-zT	60	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	1945.525	-zSW	9957.680	\
-xOBS	599.924	-yOBS	60.797	-zOBS	150.856	\
-xCAR	4.730	-yCAR	118.067	-zCAR	45.874	\
-xLAB	H	-yLAB	N	-zLAB	C	\
-ndim	3	-aq2D	States			\

**10.3.1.14 CBCACONH**

-xN	1024	-yN	80	-zN	120	\
-xT	512	-yT	40	-zT	60	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	1945.525	-zSW	9957.680	\
-xOBS	599.924	-yOBS	60.797	-zOBS	150.856	\
-xCAR	4.730	-yCAR	118.067	-zCAR	45.874	\
-xLAB	H	-yLAB	N	-zLAB	C	\
-ndim	3	-aq2D	States			\

**10.3.1.15 HNCO**

-xN	1024	-yN	168	-zN	96	\
-xT	512	-yT	84	-zT	48	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	2112.267	-zSW	1945.525	\
-xOBS	599.921	-yOBS	150.850	-zOBS	60.789	\
-xCAR	4.730	-yCAR	176.904	-zCAR	116.073	\
-xLAB	HN	-yLAB	CO	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.1.16 HNHA**

-xN	1024	-yN	244	-zN	84	\
-xT	512	-yT	122	-zT	42	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	8000.000	-ySW	8000.000	-zSW	2431.685	\
-xOBS	599.893	-yOBS	599.893	-zOBS	60.794	\
-xCAR	4.730	-yCAR	4.730	-zCAR	118.089	\
-xLAB	HN	-yLAB	H1	-zLAB	N15	\
-ndim	3	-aq2D	States			\

**10.3.1.17 Hbond (Modified HNCO)**

-xN	2134	-yN	192	\
-xT	1067	-yT	96	\
-xMODE	Complex	-yMODE	Complex	\
-xSW	10666.67	-ySW	2817.001	\
-xOBS	800.321	-yOBS	201.275	\
-xCAR	4.716	-yCAR	176.996	\
-xLAB	H1	-yLAB	C13	\
-ndim	2	-aq2D	States	\

**10.3.1.18 Jmod <sup>1</sup>H-<sup>15</sup>N-HSQC (aligned in PEG)**

-xN	3600	-yN	12	-zN	220	\
-xT	1800	-yT	12	-zT	110	\
-xMODE	Complex	-yMODE	Real	-zMODE	Complex	\
-xSW	11999.400	-ySW	12.000	-zSW	3244.120	\
-xOBS	800.304	-yOBS	1.000	-zOBS	81.103	\
-xCAR	4.754	-yCAR	0.000	-zCAR	118.030	\
-xLAB	HN	-yLAB	t1	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.1.19 Jmod  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (no alignment)**

-xN	3360	-yN	12	-zN	200	\
-xT	1680	-yT	12	-zT	100	\
-xMODE	Complex	-yMODE	Real	-zMODE	Complex	\
-xSW	11999.400	-ySW	12.000	-zSW	3244.120	\
-xOBS	800.304	-yOBS	1.000	-zOBS	81.103	\
-xCAR	4.754	-yCAR	0.000	-zCAR	118.030	\
-xLAB	HN	-yLAB	tl	-zLAB	N	\
-ndim	3	-aq2D	States			\

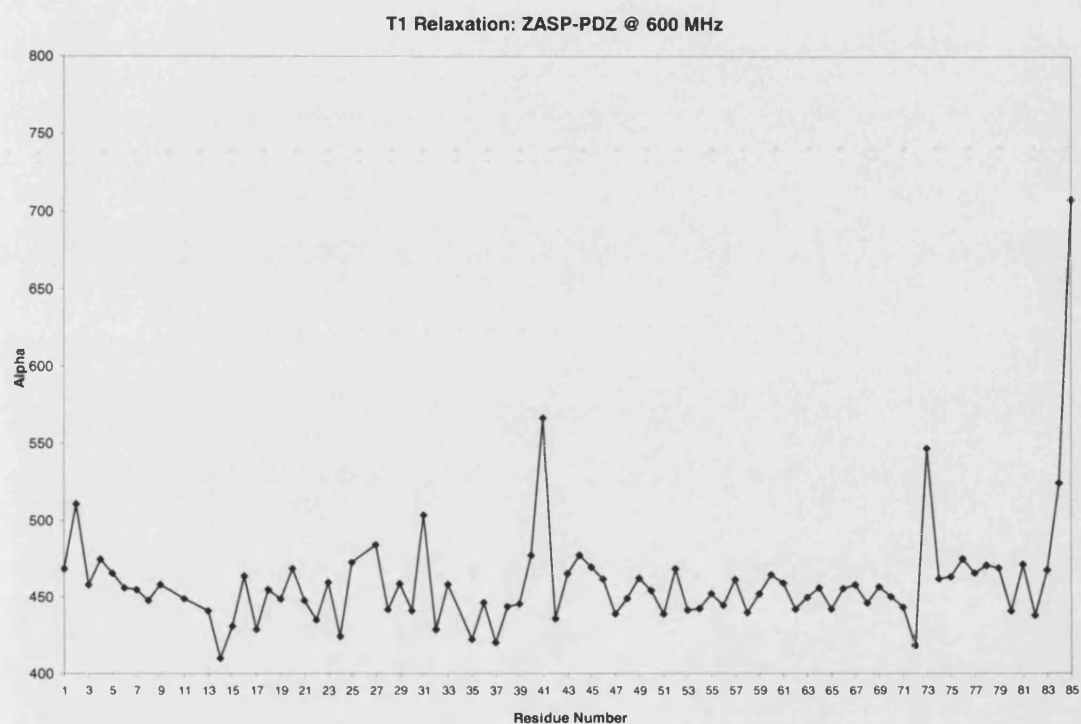
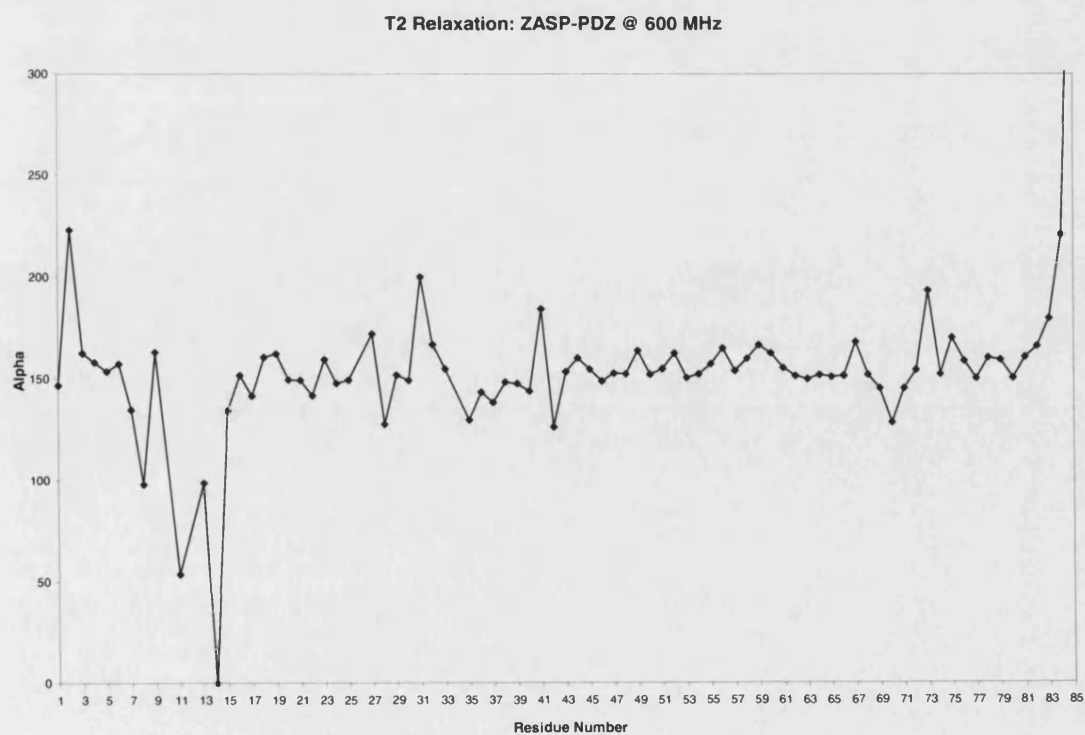
**10.3.1.1.20 T1/T2/NOE Relaxation**

-xN	2048	-yN	14	-zN	256	\
-xT	1024	-yT	14	-zT	128	\
-xMODE	Complex	-yMODE	Real	-zMODE	Complex	\
-xSW	8000.000	-ySW	14.000	-zSW	2000.000	\
-xOBS	599.893	-yOBS	1.000	-zOBS	60.793	\
-xCAR	4.754	-yCAR	0.000	-zCAR	118.282	\
-xLAB	H1	-yLAB	tl	-zLAB	N15	\
-ndim	3	-aq2D	States			\

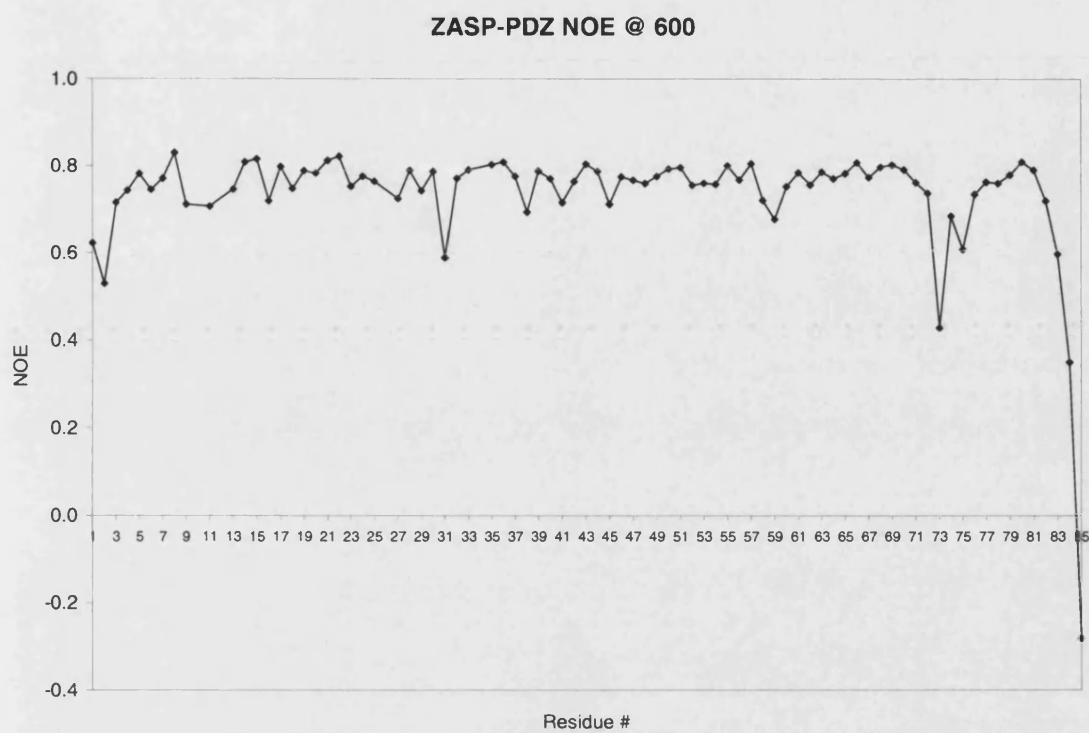
T1 (ms)	T2 (ms)
8.62	8.59
48.74	17.18
96.87	25.78
145.01	34.37
193.15	42.96
249.31	51.55
297.44	60.14
393.72	68.74
498.01	85.92
746.72	103.10
995.43	120.29
1244.13	146.06
1492.84	42.96
249.31	



The following graphs show ZASP-PDZ T1 (A), T2 (B), and NOE (C) relaxation results at 27 °C, 600 MHz. Residues with a backbone amide NOE relaxation of less than 0.7 were excluded from 'ordered' RMSD calculations, as they were deemed too flexible.

**A****B**

C



**10.3.2 Act-EF34 Spectra****10.3.2.1  $^{15}\text{N}$ -HSQC**

-xN	1600	-yN	256	\
-xT	800	-yT	128	\
-xMODE	Complex	-yMODE	Complex	\
-xSW	8000.000	-ySW	2048.026	\
-xOBS	599.924	-yOBS	60.797	\
-xCAR	4.754	-yCAR	118.089	\
-xLAB	H1	-yLAB	N15	\
-ndim	2	-aq2D	States	\

**10.3.2.2  $^{13}\text{C}$ -HSQC**

-xN	2432	-yN	1024	\
-xT	1216	-yT	512	\
-xMODE	Complex	-yMODE	Complex	\
-xSW	8000.000	-ySW	21119.324	\
-xOBS	599.924	-yOBS	150.860	\
-xCAR	4.754	-yCAR	68.818	\
-xLAB	H1	-yLAB	C13	\
-ndim	2	-aq2D	States	\

**10.3.2.3  $^1\text{H}$ - $^{15}\text{N}$ -NOESY-HSQC: 80 ms Mixing Time**

-xN	1920	-yN	256	-zN	48	\
-xT	960	-yT	128	-zT	24	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	6499.838	-ySW	6499.838	-zSW	1250	\
-xOBS	599.924	-yOBS	599.924	-zOBS	60.797	\
-xCAR	4.730	-yCAR	4.730	-zCAR	118.909	\
-xLAB	HN	-yLAB	H1	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.2.4  $^1\text{H}$ - $^{15}\text{N}$ -TOCSY-HSQC: 50 ms Mixing Time**

-xN	1920	-yN	240	-zN	56	\
-xT	960	-yT	120	-zT	28	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	6499.838	-ySW	6499.838	-zSW	1400	\
-xOBS	599.924	-yOBS	599.924	-zOBS	60.797	\
-xCAR	4.730	-yCAR	4.730	-zCAR	118.243	\
-xLAB	HN	-yLAB	HH	-zLAB	N	\
-ndim	3	-aq2D	States			\

**10.3.2.5 CBCANH**

-xN	1024	-yN	48	-zN	120	\
-xT	512	-yT	24	-zT	60	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	6800.408	-ySW	1400.021	-zSW	10000.00	\
-xOBS	599.924	-yOBS	150.856	-zOBS	60.797	\
-xCAR	4.754	-yCAR	41.261	-zCAR	118.439	\
-xLAB	H1	-yLAB	C13	-zLAB	N15	\
-ndim	3	-aq2D	States			\

**10.3.2.6 CBCACONH**

-xN	1024	-yN	50	-zN	120	\
-xT	512	-yT	25	-zT	60	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	6800.408	-ySW	1400.021	-zSW	10000.00	\
-xOBS	599.924	-yOBS	150.856	-zOBS	60.797	\
-xCAR	4.754	-yCAR	41.261	-zCAR	118.439	\
-xLAB	H1	-yLAB	C13	-zLAB	N15	\
-ndim	3	-aq2D	States			\

**10.3.2.7 HNHA**

-xN	1600	-yN	192	-zN	48	\
-xT	8000	-yT	96	-zT	24	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	5499.794	-ySW	2749.897	-zSW	1200.012	\
-xOBS	500.103	-yOBS	500.103	-zOBS	50.681	\
-xCAR	4.754	-yCAR	4.754	-zCAR	118.417	\
-xLAB	HN	-yLAB	H1	-zLAB	N15	\
-ndim	3	-aq2D	States			\

**10.3.2.8 HNHB**

-xN	1600	-yN	192	-zN	48	\
-xT	8000	-yT	96	-zT	24	\
-xMODE	Complex	-yMODE	Complex	-zMODE	Complex	\
-xSW	5499.794	-ySW	2749.897	-zSW	1200.012	\
-xOBS	500.103	-yOBS	500.103	-zOBS	50.681	\
-xCAR	4.754	-yCAR	4.754	-zCAR	118.417	\
-xLAB	HN	-yLAB	H1	-zLAB	N15	\
-ndim	3	-aq2D	States			\

**10.3.2.9 T1/T2/NOE Relaxation**

-xN	1920	-yN	13	-zN	256	\
-xT	960	-yT	13	-zT	128	\
-xMODE	Complex	-yMODE	Real	-zMODE	Complex	\
-xSW	6499.893	-ySW	13.000	-zSW	1400.021	\
-xOBS	599.924	-yOBS	1.000	-zOBS	60.797	\
-xCAR	4.754	-yCAR	0.000	-zCAR	118.953	\
-xLAB	H1	-yLAB	t1	-zLAB	N15	\
-ndim	3	-aq2D	States			\

T1 (ms)	T2 (ms)
1.50	0.00
49.66	8.69
97.82	17.38
145.98	26.06
194.14	34.75
250.33	43.44
298.49	52.13
394.81	60.82
499.16	78.19
747.98	95.57
996.81	121.63
1245.64	147.70
194.14	17.38

**10.4 NMR Spectral Assignment****10.4.1 Assignment of ZASP-PDZ**

Res	HA	HA2	HB	HB2	HB3	HG	HG1	HG2	HG3	HD1	HD2	HD3	HN	N
M1	4.45												8.50	120.10
A2	5.19		1.26										8.06	125.50
Y3	4.86			2.94	2.94					6.74	6.74		8.83	119.20
S4	5.35			3.60	3.60								8.50	115.50
V5	4.45		1.76				0.84	0.83					8.60	120.60
T6	5.04		3.79					0.87					8.57	121.90
L7	4.65			1.43	1.43	1.29				0.80	0.73		9.23	128.50
T8	4.48		4.10					1.34					8.75	119.00
G9	3.82	3.58											8.54	113.80
P10	4.34			2.23	2.12			1.84	1.71		3.47	3.43		
G11	2.32	2.27											7.88	107.30
P12														
W13	4.20			1.97	1.65					7.02			8.09	119.40
G14	3.71	3.71											8.23	102.70
F15	5.71			3.15	2.60					6.82	6.82		6.96	112.30
R16	4.69			1.86	1.60			1.57	1.57		3.15	3.09	8.32	120.00
L17	5.40			1.60	1.32	1.66				0.76	0.76		8.79	121.50
Q18	4.60			1.82	1.82			2.15	2.15				9.02	119.50
G19	4.59	3.31											8.41	108.20
G20	4.87	4.05											7.61	104.40
K21	4.17			1.45	1.38			1.27	1.27		1.73	1.50	8.76	122.80
D22	4.24			2.64	2.43								10.77	114.20
F23	4.50			3.37	2.62					7.09	7.09		7.82	118.10
N24	4.39			3.06	2.71						14.24		8.02	116.30
M25	5.00			2.09	1.75			2.63	2.63				7.33	114.60
P26	4.36			2.23	1.69			1.93	1.88		3.72	3.62		
L27	4.89			1.76	1.56	1.46				0.77	0.76		8.37	124.50
T28	5.26		3.68					0.97					9.08	118.20
I29	4.33		2.12				2.15	0.76		0.24			8.29	119.80
S30	4.50			3.79	3.37								9.19	126.40
R31	4.38			1.70	1.70			1.49	1.49		3.11	3.11	7.27	119.50
I32	4.67		1.65				2.17	0.69		0.58			8.67	124.70
T33	4.40		3.93					1.30					9.26	128.00
P34	4.31			2.31	1.87			2.11	2.00		4.19	3.74		
G35	4.08	3.72											8.85	111.40
S36	4.44			3.78	3.78								7.40	114.40
K37	4.12			2.15	2.07			1.59	1.59		2.00	2.00	9.06	120.50
A38	3.98		1.49										8.45	119.80
A39	4.14		1.59										7.64	121.20
Q40	4.34			2.29	2.05			2.45	2.38				7.64	117.30
S41	4.69			3.94	3.94								7.55	115.60
Q42	4.03			1.96	1.89			2.20	2.15				8.16	114.10
L43	4.28			1.39	1.03	1.68				0.50	0.48		7.74	119.80
S44	4.64			3.58	3.58								7.10	113.40
Q45	3.58			1.89	1.89			2.21	2.15				8.67	124.70
G46	4.38	3.52											9.30	115.30
D47	4.60			2.64	2.45								7.53	121.50
L48	4.56			1.56	1.51	1.55				0.82	0.81		8.38	121.30

Res	HA	HA2	HB	HB2	HB3	HG	HG1	HG2	HG3	HD1	HD2	HD3	HN	N
V49	4.07		1.80				0.60	0.59					8.38	123.10
V50	4.06		1.89				0.75	0.63					8.72	126.80
A51	5.12		1.02										7.90	121.50
I52	4.26		1.23				1.08	0.65		0.65			8.19	119.10
D53	4.21			2.94	2.38								9.71	129.70
G54	3.98	3.41											8.34	101.80
V55	4.03		1.97				0.87	0.85					7.89	123.70
N56	4.94			2.92	2.81						14.41		8.66	126.90
T57	4.22		4.53				5.38	1.03					8.12	113.00
D58	4.31			2.73	2.73								8.70	123.50
T59	4.33		4.30					1.14					7.59	106.40
M60	4.68			2.07	2.07			2.54	2.54				7.79	121.90
T61	4.46		4.67					1.29					8.62	113.60
H62	4.05			3.63	3.44								10.32	123.30
L63	3.99			1.64	1.64	1.58				0.92	0.91		8.97	118.00
E64	3.89			2.22	1.86			2.30	2.30				7.70	119.00
A65	4.03		1.24										8.14	122.70
Q66	3.89			2.03	2.03			2.27	2.27				8.28	116.50
N67	4.48			2.88	2.75						14.32		8.75	118.80
K68	4.10			1.96	1.96			1.39	1.39		1.64	1.64	8.08	123.10
I69	3.75		2.00				2.92	0.92		0.72			7.74	118.30
K70	4.19			1.94	1.94			1.58	1.49		1.67	1.67	8.13	119.90
S71	4.34			4.07	3.94								7.68	112.80
A72	4.42		1.59										7.34	126.20
S73	4.46			3.75	3.75								8.65	117.80
Y74	4.38			3.06	3.06					7.13	7.13		8.29	120.40
N75	5.13			2.67	2.64						14.24		8.16	116.30
L76	4.50			0.48	0.34	0.92				0.65	0.06		8.22	122.30
S77	5.34			3.73	3.73								8.91	123.40
L78	4.96			1.43	1.05	1.23				0.68	0.32		8.87	125.00
T79	4.68		3.64					1.00					8.38	119.00
L80	5.53			1.19	1.19	1.34				0.51	0.50		9.12	129.00
Q81	4.87			1.82	1.82			2.19	2.19				8.63	117.10
K82	4.37			1.78	1.78			1.43	1.43		1.56	1.56	8.66	123.50
S83	4.31			3.82	3.82								8.49	118.10
K84	4.30			1.69	1.61			1.35	1.35		1.78	1.78	8.23	123.70
R85	4.11			1.77	1.63			1.55	1.55		3.14	3.14	7.97	128.40



## Assignment of ZASP-PDZ continued: Additional Sidechain Chemical Shifts

Res	HE	HE1	HE2	HE3	HH2	HZ	HZ2	HZ3	ND2	NE1	NE2
M1											
A2											
Y3		6.29	6.29								
S4											
V5											
T6											
L7											
T8											
G9											
P10											
G11											
P12											
W13		10.01		7.24	6.71		7.25	6.51		129.00	
G14											
F15											
R16	6.80										
L17											
Q18			13.90								111.70
G19											
G20											
K21			2.89	2.89							
D22											
F23		6.78	6.78			6.87					
N24									112.60		
M25											
P26											
L27											
T28											
I29											
S30											
R31											
I32											
T33											
P34											
G35											
S36											
K37			3.07	3.00							
A38											
A39											
Q40			13.88								111.40
S41											
Q42			14.08								112.30
L43											
S44											
Q45			14.26								111.40
G46											
D47											
L48											
V49											
V50											
A51											

Res	HE	HE1	HE2	HE3	HH2	HZ	HZ2	HZ3	ND2	NE1	NE2
I52											
D53											
G54											
V55											
N56									112.20		
T57											
D58											
T59											
M60											
T61											
H62											
L63											
E64											
A65											
Q66			14.08								111.40
N67									111.10		
K68			2.84	2.84							
I69											
K70			2.92	2.92							
S71											
A72											
S73											
Y74		6.79	6.79								
N75									112.70		
L76											
S77											
L78											
T79											
L80											
Q81			14.02								110.90
K82			3.07	3.07							
S83											
K84			2.92	2.92							
R85											

Assignment of ZASP-PDZ Continued:  $^{13}\text{C}$  Chemical Shifts

Res	C	CA	CB	CD	CD1	CD2	CE	CE1	CG	CG1	CG2
M1	174.80	55.20	32.70								
A2	176.90	50.90	20.60								
Y3	172.20	56.20	40.00					38.10			
S4	173.50	56.70	65.50								
V5	174.20	60.10	36.20						21.10	21.00	
T6	173.80	61.10	70.30							21.30	
L7	175.70	52.90	43.10								
T8	175.50	62.00	69.30								21.70
G9		45.10									
P10	176.40	61.90	34.60	49.80					24.30		
G11		42.40									
P12	178.20	64.30	37.30								
W13	176.50	58.40	32.50								
G14	173.20	45.30									
F15	173.10	54.50	42.10								
R16	175.20	53.80	33.80	43.00					27.00		
L17	176.60	53.00	45.50		24.40	25.60			26.40		
Q18	173.90	54.10	32.40						33.10		
G19	172.50	44.50									
G20	176.40	43.20									
K21	178.70	59.70	32.60	29.20			41.90		24.00		
D22	176.10	54.00	36.40								
F23	174.70	57.50	40.40								
N24	173.60	54.00	37.30								
M25		52.90	36.60						31.10		
P26	175.50	62.00	32.10	50.50					26.90		
L27	177.00	54.80	42.50		25.50	25.50			27.40		
T28	174.10	59.40	73.20								21.20
I29	176.50	59.60	36.00		11.30					26.10	17.90
S30	174.60	58.10	64.30								
R31	174.20	55.70	33.50	43.10					26.80		
I32	176.10	59.20	40.00		13.70					28.00	17.00
T33		61.00	69.80								21.00
P34	177.50	63.80	31.60	51.60					27.60		
G35	173.40	45.40									
S36	175.20	57.40	66.80								
K37	180.10	60.00	32.50	29.10			42.40		26.50		
A38	179.30	54.80	19.20								
A39	179.30	54.00	18.10								
Q40	175.90	55.90	28.50						33.80		
S41	172.80	58.10	65.90								
Q42	175.10	56.00	27.90						34.20		
L43	175.90	56.10	43.10		25.00	25.00			27.20		
S44	173.30	56.40	64.90								
Q45	176.30	57.80	27.90						33.80		
G46	173.90	44.30									
D47	174.80	55.40	40.40								
L48	176.60	54.60	42.50		24.90	24.60			28.00		
V49	175.40	61.50	30.60							22.40	20.90
V50	176.50	62.80	32.30							21.10	20.10
A51	175.30	52.20	21.40								

Chapter 10: Appendix

Res	C	CA	CB	CD	CD1	CD2	CE	CE1	CG	CG1	CG2
I52	175.20	60.00	41.20		14.00					26.60	18.00
D53	176.20	55.40	39.70								
G54	173.90	44.80									
V55	175.20	61.60	32.10							21.10	21.10
N56	177.30	53.80	38.50								
T57	176.90	61.50	69.30								21.90
D58	177.10	57.00	39.60								
T59	174.70	60.80	69.00								21.20
M60	177.80	56.30	34.00						33.40		
T61	174.50	61.10	71.00								22.30
H62	177.50	61.50	29.30								
L63	178.90	57.50	41.90		23.70	24.80			26.80		
E64	179.80	59.20	29.50						37.10		
A65	179.20	55.20	18.90								
Q66	179.70	58.50	28.40						33.10		
N67	178.20	55.10	37.20								
K68	179.30	58.10	31.30	28.20			41.90		24.60		
I69	178.20	64.10	37.40		13.60					29.00	17.80
K70	177.70	58.60	32.40	29.30			41.90		25.20		
S71	173.70	59.00	63.70								
A72	177.30	52.70	19.30								
S73	174.50	59.50	63.70								
Y74	175.70	60.30	39.10					38.10			
N75	172.70	52.10	40.60								
L76	174.30	54.00	44.30		23.20	25.90			26.90		
S77	173.30	56.80	63.90								
L78	175.80	52.90	46.60		24.30	25.90			26.60		
T79	173.10	62.00	69.00								21.60
L80	175.50	53.50	45.10		26.60	26.10			27.70		
Q81	175.30	54.10	32.70						37.10		
K82	177.00	55.70	32.70	33.50			43.10		26.80		
S83	174.30	58.20	63.70								
K84	175.30	56.20	32.90	33.00			41.90		24.40		
R85		57.30	31.30	43.30					26.90		

## 10.4.2 Assignment of ZASP-PDZ bound to Act-EF1234

Res	HA	HA2	HB	HB2	HB3	HG	HG1	HG2	HG3	HD1	HD2	HD3	HE1	HN	N
<b>M1</b>														8.53	119.80
A2	5.20		1.24											8.06	125.10
Y3	4.82			2.92	2.92					6.72	6.72			8.84	118.90
S4	5.36			3.59	3.59									8.51	115.20
V5	4.46		1.77				0.85	0.78						8.65	120.00
T6	5.05		3.80					0.85						8.55	121.30
L7	4.65			1.41	1.41	1.27				0.91	0.69			9.24	128.20
T8	4.49							1.33						8.78	118.70
G9	3.78	3.64												8.58	114.20
P10	4.34														
G11	2.20	2.09												7.97	107.50
W13													9.76	8.10	119.30
G14														8.23	102.50
F15														6.87	112.20
R16														8.43	119.90
L17	5.23													8.96	125.00
Q18				1.84	1.84			2.17	2.17					9.26	123.20
G19	4.57	3.40												8.42	108.30
G20	4.85	4.00												7.77	104.50
K21	4.18			1.36	1.34			1.30	1.30		1.69	1.43		8.79	122.90
D22	4.18			2.62	2.41									10.77	113.80
F23	4.43			3.40	2.52									7.84	117.70
N24	4.39			3.04	2.69									7.99	115.90
M25	4.96			2.10	1.79			2.61	2.61					7.35	114.20
P26	4.38			2.22											
L27	4.85			1.73	1.67	1.52				0.72	0.70			8.36	123.50
T28	5.15		3.70					0.89						9.22	118.90
I29								0.72						8.32	119.80
S30	4.54			3.82	3.43									9.15	125.50
R31	4.41			1.71	1.71			1.57	1.57					7.42	119.80
I32	4.65							0.70						8.68	124.60
T33	4.34		3.91					1.26						9.19	127.90
G35	4.04	3.71												8.86	111.00
S36	4.44			3.75	3.75									7.41	114.40
K37	4.10			2.24	2.18			1.62	1.62		2.03	2.03		9.02	120.40
A38														8.48	119.80
A39	4.14		1.59											7.78	121.00
Q40	4.29			2.33	2.02			2.44	2.36					7.63	117.00
S41	4.69			3.91	3.91									7.59	115.40
Q42	3.98			1.96	1.89			2.17	2.13					8.18	113.80
L43	4.23			1.38	1.06	1.64				0.49	0.47			7.79	119.80
S44	4.65			3.58	3.58									7.16	113.60
Q45				1.90	1.90			2.23	2.10					8.69	124.50
G46	4.34	3.53												9.43	115.40
D47	4.60			2.63	2.43									7.53	121.40
L48	4.62			1.58	1.48	1.54				0.80	0.74			8.43	120.50
V49	4.05		1.81				0.55	0.51						8.32	122.30
V50	4.06		1.95				0.65	0.60						8.71	126.00
A51	5.11		1.00											7.93	121.40
I52	4.23		1.20				1.06	0.66		0.69				8.21	118.90
D53	4.21			2.94										9.69	129.50
G54	3.98	3.39												8.35	101.50
V55	4.00		1.96				0.87	0.84						7.91	123.50
N56	4.91			2.91	2.79									8.67	126.60
T57	4.56		4.52					1.00						8.05	112.30
D58	4.28			2.72	2.72									8.69	123.10
T59	4.31		4.26					1.14						7.67	106.10

Res	HA	HA2	HB	HB2	HB3	HG	HG1	HG2	HG3	HD1	HD2	HD3	HE1	HN	N
M60	4.67			2.06	2.06			2.54	2.54					7.80	121.50
T61	4.45		4.69					1.29						8.62	113.50
H62	3.99			3.58	3.44									10.32	122.80
L63	3.95			1.61	1.61	1.56				0.87	0.84			9.03	117.30
E64	3.92			2.20	1.87			2.32	2.32					7.69	118.60
A65	4.05		1.24											8.06	122.30
Q66	3.94			2.00	2.00									8.49	117.00
N67	4.39			2.85	2.71									8.68	117.40
K68	4.00			1.94	1.94			1.26	1.26		1.40	1.40		8.08	122.30
I69	3.68		1.92					1.01		0.78				7.83	118.90
K70				1.88	1.88			1.63						8.14	119.80
S71	4.70			4.03	3.98									7.68	112.00
A72	4.38		1.64											7.32	126.40
S73	4.40			3.71	3.71									8.59	117.70
Y74				3.03	3.03									8.30	120.20
N75	5.14			2.62	2.56									8.17	116.50
L76	4.50			0.55	0.49	0.84				0.65				8.34	121.80
S77	5.35			3.70	3.70									8.89	122.90
L78	4.94			1.43	1.05	1.20				0.80				8.81	124.80
T79	4.69		3.63					1.00						8.41	118.70
L80	5.54			1.17	1.17	1.35				0.52	0.46			9.14	128.80
Q81	4.86			1.82	1.82									8.66	117.00
K82	4.39			1.83	1.83			1.56	1.56		1.69	1.69		8.65	123.30
S83	4.36			3.81	3.81									8.52	117.80
K84	4.28			1.70	1.70									8.24	123.40
R85	4.12			1.65	1.65									7.99	128.20

## 10.4.3 Assignment of Act-EF34

Res	CA	CB	HA	HA2	HB	HB2	HB3	HB#	HG	HG#	HG1#	HG2#	HN	N
G-1	45.00													
A0	52.39	19.18	4.28					1.32					7.97	123.35
M1	55.26	32.82	4.37					1.95		2.45			8.44	119.33
A2	52.16	19.28	4.29					1.33					8.30	124.79
D3	54.61	40.76	4.54			2.68	2.63						8.32	119.61
T4	61.71	69.37	4.26		3.75								7.86	111.78
D5	54.75	41.17	4.63					2.63					8.18	122.79
T6	62.15	69.73	4.24		4.09								8.06	113.91
A7	54.84	17.96	4.00					1.40					8.44	124.37
E8	59.13	29.13	3.99					2.00		2.30			8.33	116.53
Q9	57.97	28.63	4.07					2.08		2.34			7.83	119.49
V10	65.54	30.44	3.57		2.06						0.87	0.81	7.99	122.07
I11	65.32	37.55	3.67		1.89								8.24	121.06
A12	54.49	17.86	4.04					1.40					7.82	121.23
S13	61.00	62.42	4.14			3.81	3.69						7.77	114.09
F14	60.37	38.77	4.04			2.69	2.17						7.87	122.05
R15	59.64	29.76	3.57										8.17	119.99
I16	63.91	37.90	3.79		1.85								7.27	118.08
L17	57.17	41.77	4.00					1.78					7.60	122.64
A18	51.55	18.88	4.22					1.33					8.01	119.22
S19	58.89	61.11	4.06					3.91					8.11	116.41
D20	55.72	40.08	4.07			2.84	2.70						8.63	112.32
K21		31.66	4.38					2.05					7.33	119.05
P22	63.12	31.33												
Y23	54.25	40.68	4.79					2.81					6.42	112.96
I24	58.64	43.34	4.90										9.10	116.78
L25	53.17	43.04	5.04					1.90	1.63				8.92	125.09
A26	56.07	17.45	4.35										9.52	126.89
E27	59.15	29.11											9.20	113.99
E28	58.76	28.99	3.97					1.86					7.17	118.47
L29	58.19	42.30	3.90					2.07					7.34	118.49
R30	58.36	30.89	3.89					1.70					8.10	113.91
R31	58.18	30.84	4.15			1.96	1.75			1.60			7.44	116.92
E32	56.25	32.18	4.46					1.66		2.19			7.24	114.09
L33	52.10	42.71	4.92					1.67	1.33				7.76	118.80
P35	65.98	32.05												
D36	56.84	39.19	4.36			2.69	2.60						8.89	116.00
Q37	58.27	29.31	4.21					2.09					7.44	121.75
A38	56.01	17.56						1.28					8.81	122.24
Q39	58.58	28.39	3.98			2.17	2.08			2.45			8.10	115.09
Y40	61.09	38.59	4.02					3.32					7.63	119.71
C41	63.93	27.62	3.88			3.05	2.45						7.96	115.27
I42	65.42	38.26											8.41	119.76
K43	57.67	32.50	4.04					1.75					7.52	115.22
R44	55.64	31.45	4.25					1.31					7.18	116.04
M45			4.57					1.79		2.53			7.60	120.23
P46	61.14	32.07												
A47	52.48	18.37	4.36					1.32					8.41	121.39
Y48	57.43	38.87	4.59					2.42					8.31	122.19
S49	56.35	63.44											8.71	124.29



Chapter 10: Appendix

Res	CA	CB	HA	HA2	HB	HB2	HB3	HB#	HG	HG#	HG1#	HG2#	HN	N
G50			4.02	3.53									5.19	107.10
G52	45.52													
S53	59.33	62.84	3.70					3.52					7.61	111.71
V54	57.57	33.11	4.67		2.22								6.30	114.56
P55	63.88	31.37												
G56	44.69		4.26	3.67									8.60	111.90
A57	53.04	43.14	4.02					1.06					7.76	121.55
L58	54.17	43.83	4.35					1.45					9.52	126.53
D59	51.88	41.01						3.11					9.06	121.73
Y60	58.41	37.22	4.24			2.77	2.34						7.06	122.55
A61	54.37	17.15	4.20					1.32					7.99	127.64
A62	52.27	18.67	4.30					1.30					7.50	121.03
F63	59.23	40.64	4.46			3.23	3.11						7.46	118.82
S64	59.37	63.64	4.21			3.99	3.81						7.89	117.71
S65	59.86	63.20	4.21					3.90					8.13	116.57
A66	52.77	18.57	4.20					1.28					7.78	124.00
L67	55.59	42.26	4.06					1.21					7.73	119.63
Y68	57.79	38.79	4.49			3.07	2.78						7.86	118.47
G69	45.09		3.95	3.88									8.01	109.82
E70	56.30	30.24	4.28			2.04	1.90			2.23			8.18	120.29
S71	58.10	64.17	4.47					3.81					8.28	116.07
D72	54.14	40.84	4.61			2.69	2.55						8.36	122.83
L73			4.16					1.53					7.69	127.19

**10.5 CYANA Manual Calculation****10.5.1 Init.cya**

This is read when CYANA starts, setting parameters for the structure calculation, such as the name of the protein, violation cut-off limits, restraints, stereo-assignments, and hydrogen bonds, etc.

```
# General Parameters

name:=zasp          # protein name

ang_cut=0.4         # Angle cut-off (Å)
cut_upl=0.5         # Upper-limit cut-off (Å)
cut_lol=0.5         # Lower-limit cut-off (Å)
cut_aco=5.0         # Angle cut-off (degrees)
cut_vdw=0.5         # van der Waals cut-off (Å)

rmsdrange:= 1..85 # Default residue range for RMSD

read lib  file=lib/dyana-1.5/lib/dyana.lib      # read library
read seq  file=RESTRAINTS/$name.seq            # read sequence file
read prot file=RESTRAINTS/$name.prot           # read prot file

# Restraints (NOE, dihedrals)

read upl  file=RESTRAINTS/zasp_15Nnoesy.upl      # read NOE upper limits
read upl  file=RESTRAINTS/zasp_13cnoesy.upl append
read aco  file=RESTRAINTS/hnha.aco              # read HNHA angles
read aco  file=RESTRAINTS/talos.aco             # read TALOS angles

# Stereo assignments (from GLOMSA)

atom swap HA1      46 54
atom swap HB2      10 17 23 30 37 67 76
atom swap HG12     32 69
atom swap HD2      34 10
atom swap HE21     18 45

atom stereo HA1    46 54
atom stereo HB2    10 15 16 17 21 23 30 37 48 56 64 67 76
atom stereo HG2    42 45
atom stereo HG12   32 69
atom stereo QG1    49 50
atom stereo HD2    34 10
atom stereo HE21   18 45

distance modify          # removes irrelevant constraints

# Restraints from H-bond experiment (spec800)

# Confirmed H-bonds

hbond atom1=HN residue1=80 atom2=O residue2=3
hbond atom1=HN residue1=82 atom2=O residue2=1
hbond atom1=HN residue1=3  atom2=O residue2=80
hbond atom1=HN residue1=5  atom2=O residue2=78
hbond atom1=HN residue1=33 atom2=O residue2=14
hbond atom1=HN residue1=18 atom2=O residue2=28
hbond atom1=HN residue1=81 atom2=O residue2=48
```

**10.5.2 Calibration.cya**

Command to calibrate volumes/intensities for distance restraints.

```
read peaks RESTRAINTS/zasp_nhsqc_fifty_volumes.peaks assigned
integrated
caliba dmin=2.2 dmax=6.0 bb=2.27E+07

read peaks RESTRAINTS/zasp_13cnoe_volumes.peaks assigned integrated
caliba dmin=2.2 dmax=6.0 bb=2.65E+07

distance unique
write upl zasp.upl
```

**10.5.3 CYCLE.cya**

Macro that initiates the ANNEAL\_STR.cya REDAC structure calculation. This defines 5 rounds of 100 structure calculations.

```
ANNEAL_STR structures=100 ang=PDB/zasp_001_ ovw=OVW/zasp_001
aco=OVW/zasp_001 start=random steps=10000
read aco OVW/zasp_001.aco

ANNEAL_STR structures=100 ang=PDB/zasp_002_ ovw=OVW/zasp_002
aco=OVW/zasp_002 start=memory steps=10000
read aco OVW/zasp_002.aco

ANNEAL_STR structures=100 ang=PDB/zasp_003_ ovw=OVW/zasp_003
aco=OVW/zasp_003 start=memory steps=10000
read aco OVW/zasp_003.aco

ANNEAL_STR structures=100 ang=PDB/zasp_004_ ovw=OVW/zasp_004
aco=OVW/zasp_004 start=memory steps=10000
read aco OVW/zasp_004.aco

ANNEAL_STR structures=100 ang=PDB/zasp_005_ ovw=OVW/zasp_005
aco=OVW/zasp_005 start=memory steps=10000

structure sort
structure select 1..30
write cor file=zasp_005.cor all append
```

## 10.5.4 REDAC Annealing Protocol (ANNEAL\_STR.cya)

Usage:    anneal\_str    [structures=<n>]    [ang=<file>]    [cor=<file>]  
 [ovw=<file>] [aco=<file>]  
 [start=memory|random|<file>] [steps=<N>] [thigh=<T0>]

Calculate n structures using the simulated annealing protocol. The start structures are read from disk in angle format [file], copied from the structure memory [memory], or randomly generated [random]. The structures are saved to the structure memory and optionally written to angle [ang] or coordinates [cor] files. An overview file is written if [ovw] is specified. An angle constraint file is written if [aco] is specified:

```
var i l2 j iseed nseed info t dt

t = walltime

syntax structures=0<@i ang=* cor=* ovw=* aco=* \
    start*=random steps=@i=10000 thigh=@r=8.0 append

angstat clear

echo := off
info := normal
iseed=seed

do j 1 structures
  if ('$start'.eq.'random') then
    nseed=nint(1000000*rand(iseed,j))
    random nseed
    print "    Random structure $j with seed $nseed created."
  else if ('$start'.eq.'memory') then
    structure copy j 0
  else
    read ang {$start}$j(I3.3).ang
  end if
  anneal steps=steps
  print "    Structure $j minimized, f = {$tf}"
  if ('$ang'.ne.' ') then
    if (append) then
      write ang $ang.ang append
    else
      write ang {$ang}$j(I3.3).ang
    end if
  end if
  if ('$cor'.ne.' ') write cor {$cor}$j(I3.3).cor
  angstat make ang_cut=0.4
  if ('$aco'.ne.' ') write aco $aco.aco redac
  structure copy 0 j name=str$j
end do

overview $ovw.ovw structure=$j range=- hbond vdw full

if (structures.gt.0) then
  t=walltime-t; dt=real(t)/structures
  print "    $structures structures finished in $t s ($dt
s/structure)."
end if
```

**10.6 CYANA Automatic Calculation (CANDID)****10.6.1 Init.cya**

This is read when CYANA starts, setting parameters for the structure calculation, such as the name of the protein, violation cut-off limits, restraints, stereo-assignments, and hydrogen bonds, etc.

```
name:=zasp          # protein name, used for output file names
rmsdrange:=1..85    # residue range for RMSD calculations
nproc:=1            # Number of CPUs

read lib file=RESTRAINTS/cyana.lib # read library
read seq RESTRAINTS/$name.seq      # read protein sequence
```

**10.6.2 CANDID.CYA**

Script to execute the CANDID automatic assignment.

```
peaks      := zasp          # names of peak lists
prot       := zasp          # names of proton lists
tolerance  := 0.02, 0.02, 0.05 # chemical shift tolerances (ppm);
                                     # 1H(a),1H(b),13C/15N(a),13C/15N(b)

subroutine ANNEAL          # structure calculation
  var n
  ./init                  # re-initialize
  read upl cycle$cycle.upl # CANDID NOE Upper Distance Limits
# read upl zasp.upl append # DYANA NOE Upper Distance Limits
  read upl hbonds.upl append # additional upper distance limits
# read lol hbonds.lol append # additional lower distance limits
  read aco talos.aco        # angle constraints
  seed=5671                 # random number generator seed
  n=100                     # number of start conformers
  if (def('nproc')) n=nint(real(n)/nproc)*nproc # number of CPUs
  calc_all structures=n steps=10000 # struct calculation
  overview cycle$cycle structures=20 cor # write ovw & cor
end

candid peaks=$peaks prot=$prot calculation=ANNEAL
```





**10.8 Titration Data**

Tables show the ligand to protein ratios used for the NMR titrations.

**10.8.1 Titration of  $^{15}\text{N}$ -ZASP-PDZ with Unlabelled Act-EF1234**

Data Point	Total Vol. ( $\mu\text{l}$ )	Fin [ZASP] (M)	Fin [Act14] (M)	ZASP/ACT
0	520	2.76E-04	0.00E+00	N/A
1	545	2.64E-04	2.66E-05	9.92
2	570	2.52E-04	5.08E-05	4.96
3	620	2.32E-04	9.34E-05	2.48
4	645	2.23E-04	1.12E-04	1.98
5	670	2.14E-04	1.30E-04	1.65
6	695	2.07E-04	1.46E-04	1.42
7	720	2.00E-04	1.61E-04	1.24
8	770	1.87E-04	1.88E-04	0.99
9	820	1.75E-04	2.12E-04	0.83
10	870	1.65E-04	2.33E-04	0.71
11	920	1.56E-04	2.52E-04	0.62
12	1020	1.41E-04	2.84E-04	0.50

**10.8.2 Titration of  $^{15}\text{N}$ -ZASP-PDZ with Unlabelled Act-EF34**

Data Point	Total Vol. ( $\mu\text{l}$ )	Fin [ZASP] (M)	Fin [Act34] (M)	ZASP/ACT
0	600	3.70E-04	0.00E+00	N/A
1	610	3.64E-04	1.10E-05	33.09
2	625	3.55E-04	2.69E-05	13.24
3	635	3.50E-04	3.70E-05	9.45
4	665	3.34E-04	6.56E-05	5.09
5	700	3.17E-04	9.59E-05	3.31
6	735	3.02E-04	1.23E-04	2.45
7	765	2.90E-04	1.45E-04	2.01
8	800	2.78E-04	1.68E-04	1.65
9	830	2.68E-04	1.86E-04	1.44
10	860	2.58E-04	2.03E-04	1.27
11	930	2.39E-04	2.38E-04	1.00
12	990	2.24E-04	2.64E-04	0.85
13	1070	2.08E-04	2.95E-04	0.70
14	1150	1.93E-04	3.11E-04	0.62
15	1260	1.76E-04	3.40E-04	0.52



**10.8.3 Titration of  $^{15}\text{N}$ -Act-EF1234 with Unlabelled ZASP-PDZ**

Data Point	Total Vol. ( $\mu\text{l}$ )	Fin [ACT] (M)	Fin [ZASP] (M)	Act/ZASP
0	575	4.21E-04	0.00E+00	N/A
1	595	4.07E-04	4.47E-05	9.09
2	605	4.00E-04	6.60E-05	6.06
3	615	3.93E-04	8.65E-05	4.55
4	625	3.87E-04	1.06E-04	3.64
5	635	3.81E-04	1.26E-04	3.03
6	655	3.69E-04	1.62E-04	2.27
7	675	3.58E-04	1.97E-04	1.82
8	695	3.48E-04	2.30E-04	1.52
9	715	3.38E-04	2.60E-04	1.30
10	735	3.29E-04	2.90E-04	1.14
11	755	3.20E-04	3.17E-04	1.01
12	855	2.83E-04	4.36E-04	0.65
13	955	2.53E-04	5.29E-04	0.48

**10.8.4 Titration of  $^{15}\text{N}$ -Act-EF34 with Unlabelled ZASP-PDZ**

Data Point	Total Vol. ( $\mu\text{l}$ )	Fin [ACT] (M)	Fin [ZASP] (M)	Act/ZASP
0	580	6.08E-04	0.00E+00	N/A
1	585	6.03E-04	1.66E-05	36.22
2	590	5.97E-04	3.30E-05	18.11
3	595	5.92E-04	4.91E-05	12.07
4	600	5.88E-04	6.49E-05	9.05
5	605	5.83E-04	8.04E-05	7.24
6	615	5.73E-04	1.11E-04	5.17
7	625	5.64E-04	1.40E-04	4.02
8	635	5.55E-04	1.69E-04	3.29
9	650	5.42E-04	2.10E-04	2.59
10	670	5.26E-04	2.61E-04	2.01
11	690	5.11E-04	3.10E-04	1.65
12	720	4.90E-04	3.79E-04	1.29
13	760	4.64E-04	4.61E-04	1.01
14	820	4.30E-04	5.70E-04	0.75
15	880	4.01E-04	6.64E-04	0.60
16	940	3.75E-04	7.46E-04	0.50

**10.8.5 Titration of  $^{15}\text{N}$ -Act-EF34 with Unlabelled tZR7, followed by ZASP-PDZ**

<b>Data Point</b>	<b>Total Vol (<math>\mu\text{l}</math>)</b>	<b>Fin [Act-EF34] (M)</b>	<b>Fin [Zr7] (M)</b>	<b>Act/Zr7</b>
0	600	5.25E-004	0.00E+000	N/A
1	847	3.72E-004	7.44E-004	0.50

<b>Data Point</b>	<b>Total Vol (<math>\mu\text{l}</math>)</b>	<b>Fin [Act-EF34] (M)</b>	<b>Fin [ZASP] (M)</b>	<b>Act/ZASP</b>
0	600	3.72E-004	0.00E+000	N/A
1	610	3.66E-004	2.11E-005	17.30
2	620	3.60E-004	4.16E-005	8.65
3	640	3.49E-004	8.06E-005	4.33
4	655	3.41E-004	1.08E-004	3.15
5	675	3.31E-004	1.43E-004	2.31
6	695	3.21E-004	1.76E-004	1.82
7	710	3.14E-004	2.00E-004	1.57
8	730	3.06E-004	2.30E-004	1.33
9	750	2.98E-004	2.58E-004	1.15
10	775	2.88E-004	2.91E-004	0.99
11	815	2.74E-004	3.40E-004	0.80
12	860	2.60E-004	3.90E-004	0.67

## 10.9 Perl Scripts

### 10.9.1 D2X.PL: DYANA To XPLOR Distance Restraint Conversion Script

#### 10.9.1.1 Description

Converts DYANA NOE upper-limit constraints (.upl) files into XPLOR format. NOE constraints will be either used as the maximum distance (dmax) or separated into NOE classes. NOE distances above the maximum NOE class will be inputted as a very weak NOE with dmax corresponding to the distance. The conversion has three distinct stages:

- Option to remove of DYANA pseudo-atom correction (generated from 'distance modify): as XPLOR adds its own correction.
- Substitution of DYANA with XPLOR atom nomenclature.
- Conversion of DYANA format into XPLOR format.

#### 10.9.1.2 Usage

The script requires renaming of the input files, and the setting of the options referring to pseudoatom correction, and distance method, as described above.

#### 10.9.1.3 Code

```
#!/usr/bin/perl -w

#Define files:

$name      = 'zasp';           # NAME of the protein
$dyana     = 'zasp.upl';       # DYANA Input
$temp      = 'zasp.psu';       # Pseudoatom correction removal
$temp2     = 'zasp.sub';       # DYANA to XPLOR nomenclature change
$xplor     = 'zasp.tbl';       # XPLOR Output

# Set method, and boundaries (d = distance, $str/med/wea/vwk = bounds):
# 1. Upper boundary alone      (i.e.: 0          0          d      )
# 2. Same as 1, 1.8 lower bound (i.e.: d          (d-1.8)      0      )
# 3. NOE class boundaries      (e.g.: $str1    ($str-1.8) $str2-$str1)

# Set Distance method and pseudoatom correction:

$method    = '2';
$pseudo    = 'y';

# Set NOE class boundaries if method 3 is set:

$str1      = '2.5'; # Strong
$str2      = '2.7';
$med1      = '3.0'; # Medium
$med2      = '3.3';
$wea1      = '4.0'; # Weak
$wea2      = '5.0';
$vwk1      = '4.0'; # Very Weak
$vwk2      = '6.0';
```

```
# A) DYANA Pseudoatom Correction removal

open (IN,          $dyana) or die "cannot open $dyana for read :$!";
open (TEMP, "> $temp") or die "Cannot open $temp for write :$!";

if ($pseudo eq 'y') {

    # Pseudoatom correction executed

    while ($line = <IN>) {
        ($res1,$aa1,$at1,$res2,$aa2,$at2,$dist) = $line=~
        /(\\d+?)\\s+?(\\w+?.)\\s+?(\\w+?.)\\s+?(\\d+?)\\s+?(\\w+?.)\\s+?(\\w+?.)\\s+?(\\d+?)\\s+?(\\d+?)\\s/;

        # First Pseudoatom Correction:

        if ($at1 eq 'QA') { $d1 = $dist - 0.80; }
        elsif ($at1 eq 'QB') {
            if ($aa1 ne 'ALA') { $d1 = $dist - 0.88; }
            else { $d1 = $dist - 1.03; }}
        elsif ($at1 eq 'QG') { $d1 = $dist - 0.88; }
        elsif ($at1 eq 'QG1') { $d1 = $dist - 1.03; }
        elsif ($at1 eq 'QG2') { $d1 = $dist - 1.03; }
        elsif ($at1 eq 'QQG') { $d1 = $dist - 2.31; }
        elsif ($at1 eq 'QD') {
            if ($aa1 ne 'TYR' and $aa1 ne 'PHE') { $d1 = $dist - 0.88; }
            else { $d1 = $dist - 2.13; }}
        elsif ($at1 eq 'QD1') { $d1 = $dist - 1.03; }
        elsif ($at1 eq 'QD2') { $d1 = $dist - 1.03; }
        elsif ($at1 eq 'QQD') { $d1 = $dist - 2.31; }
        elsif ($at1 eq 'QE') {
            if ($aa1 ne 'MET' and $aa1 ne 'TYR' and $aa1 ne 'PHE') { $d1 =
$d1 - 0.88; }
            elsif ($aa1 ne 'TYR' and $aa1 ne 'PHE') { $d1 =
$d1 - 1.03; }
            else { $d1 =
$d1 - 2.13; }}
        elsif ($at1 eq 'QR') { $d1 = $dist - 2.48; }
        else { $d1 = $dist; }

        # Second Pseudoatom Correction:

        if ($at2 eq 'QA') { $d2 = $d1 - 0.80; }
        elsif ($at2 eq 'QB') {
            if ($aa2 ne 'ALA') { $d2 = $d1 - 0.88; }
            else { $d2 = $d1 - 1.03; }}
        elsif ($at2 eq 'QG') { $d2 = $d1 - 0.88; }
        elsif ($at2 eq 'QG1') { $d2 = $d1 - 1.03; }
        elsif ($at2 eq 'QG2') { $d2 = $d1 - 1.03; }
        elsif ($at2 eq 'QQG') { $d2 = $d1 - 2.31; }
        elsif ($at2 eq 'QD') {
            if ($aa2 ne 'TYR' and $aa2 ne 'PHE') { $d2 = $d1 - 0.88; }
            else { $d2 = $d1 - 2.13; }}
        elsif ($at2 eq 'QD1') { $d2 = $d1 - 1.03; }
        elsif ($at2 eq 'QD2') { $d2 = $d1 - 1.03; }
        elsif ($at2 eq 'QQD') { $d2 = $d1 - 2.31; }
        elsif ($at2 eq 'QE') {
            if ($aa2 ne 'MET' and $aa2 ne 'TYR' and $aa2 ne 'PHE') { $d2 =
$d1 - 0.88; }
            elsif ($aa2 ne 'TYR' and $aa2 ne 'PHE') { $d2 =
$d1 - 1.03; }
            else { $d2 =
$d1 - 2.13; }}
    }
}
```

```

elseif ($at2 eq 'QR') { $d2 = $d1 - 2.48; }
else { $d2 = $d1; }

printf TEMP "%3d %-4s %-5s %3d %-4s %-5s
%.2f\n", $res1, $aal, $at1, $res2, $aa2, $at2, $d2;
}
} else {

# No Pseudoatom Correction

while ($line = <IN>) {
    ($res1, $aal, $at1, $res2, $aa2, $at2, $dist) = $line=~
/(\d+?)\s+(\w+?.)\s+(\w+?.)\s+(\d+?)\s+(\w+?.)\s+(\w+?.)\s+(\d+?)\s/;
    printf TEMP "%3d %-4s %-5s %3d %-4s %-5s
%.2f\n", $res1, $aal, $at1, $res2, $aa2, $at2, $dist;
}
}

close IN;
close TEMP;

# B) DYANA to XPLOR atom nomenclature conversion

open (TEMP, $temp) or die "cannot open $temp for read :$!";
open (TEMP2, "> $temp2") or die "Cannot open $temp2 for write :$!";

# Substitution Matrix

my %sub = (

# Substitute

    'HA1' => 'HA2' ,
    'HA2' => 'HA1' ,
    'HB3' => 'HB1' ,
    'HD3' => 'HD1' ,
    'HE3' => 'HE1' ,
    'HG13' => 'HG11' ,
    'HG3' => 'HG1' ,

# No Substitute

    'C' => 'C' ,
    'CA' => 'CA' ,
    'CB' => 'CB' ,
    'CD' => 'CD' ,
    'CD1' => 'CD1' ,
    'CD2' => 'CD2' ,
    'CE' => 'CE' ,
    'CE1' => 'CE1' ,
    'CE2' => 'CE2' ,
    'CE3' => 'CE3' ,
    'CG' => 'CG' ,
    'CG1' => 'CG1' ,
    'CG2' => 'CG2' ,
    'CH2' => 'CH2' ,
    'CZ' => 'CZ' ,
    'CZ2' => 'CZ2' ,
    'CZ3' => 'CZ3' ,
    'HA' => 'HA' ,
    'HB' => 'HB' ,
    'HB1' => 'HB1' ,
    'HB2' => 'HB2' ,

```

```

'HD1'  => 'HD1'  ,
'HD11' => 'HD11' ,
'HD12' => 'HD12' ,
'HD13' => 'HD13' ,
'HD2'  => 'HD2'  ,
'HD21' => 'HD21' ,
'HD22' => 'HD22' ,
'HD23' => 'HD23' ,
'HE'   => 'HE'   ,
'HE1'  => 'HE1'  ,
'HE2'  => 'HE2'  ,
'HE21' => 'HE21' ,
'HE22' => 'HE22' ,
'HG'   => 'HG'   ,
'HG1'  => 'HG1'  ,
'HG11' => 'HG11' ,
'HG12' => 'HG12' ,
'HG2'  => 'HG2'  ,
'HG21' => 'HG21' ,
'HG22' => 'HG22' ,
'HG23' => 'HG23' ,
'HH'   => 'HH'   ,
'HH11' => 'HH11' ,
'HH12' => 'HH12' ,
'HH2'  => 'HH2'  ,
'HH21' => 'HH21' ,
'HH22' => 'HH22' ,
'HN'   => 'HN'   ,
'HZ'   => 'HZ'   ,
'HZ1'  => 'HZ1'  ,
'HZ2'  => 'HZ2'  ,
'HZ3'  => 'HZ3'  ,
'N'    => 'N'    ,
'ND1'  => 'ND1'  ,
'ND2'  => 'ND2'  ,
'NE'   => 'NE'   ,
'NE1'  => 'NE1'  ,
'NE2'  => 'NE2'  ,
'NH1'  => 'NH1'  ,
'NH2'  => 'NH2'  ,
'NZ'   => 'NZ'   ,
'O'    => 'O'    ,
'OD1'  => 'OD1'  ,
'OD2'  => 'OD2'  ,
'OE1'  => 'OE1'  ,
'OE2'  => 'OE2'  ,
'OG'   => 'OG'   ,
'OG1'  => 'OG1'  ,
'OH'   => 'OH'   ,
'SD'   => 'SD'   ,
'SG'   => 'SG'   ,

```

#### # Pseudoatom Substitutions

```

'QA'   => 'HA#'  ,
'QB'   => 'HB#'  ,
'QG'   => 'HG#'  ,
'QG1'  => 'HG1#' ,
'QG2'  => 'HG2#' ,
'QQG'  => 'HG#'  ,
'QD'   => 'HD#'  ,
'QD1'  => 'HD1#' ,
'QD2'  => 'HD2#' ,
'QQD'  => 'HD#'  ,

```

```

        'QE'    => 'HE#' ,
        'QE2'   => 'HE2#' ,
        'QR'    => 'HD#' ,
        'QZ'    => 'HZ#' ,
        'QH1'   => 'HH1#' ,
        'QH2'   => 'HH2#' ,
    );

while ($line = <TEMP>) {
    ($res1,$aa1,$at1,$res2,$aa2,$at2,$dist) = $line=~
/((\d+?)\s+(\w+?)\s+(\w+?)\s+(\d+?)\s+(\w+?)\s+(\w+?)\s+(\d+?)\s+\/;

# Correct Atom 1

if ($at1 eq 'HB3') {
    if ($aa1 ne 'ALA') { $sa1 = $sub{$at1};}
    else { $sa1 = $at1; }
}
elseif ($at1 eq 'HE3') {
    if ($aa1 ne 'TRP' and $aa1 ne 'MET') { $sa1 = $sub{$at1}; }
    else { $sa1 = $at1; }
}
elseif ($at1 eq 'HG13') {
    if ($aa1 ne 'VAL') { $sa1 = $sub{$at1}; }
    else { $sa1 = $at1; }
}
else { $sa1 = $sub{$at1}; }

# Correct Atom 2

if ($at2 eq 'HB3') {
    if ($aa2 ne 'ALA') { $sa2 = $sub{$at2};}
    else { $sa2 = $at2; }
}
elseif ($at2 eq 'HE3') {
    if ($aa2 ne 'TRP' and $aa2 ne 'MET') { $sa2 = $sub{$at2};}
    else { $sa2 = $at2; }
}
elseif ($at2 eq 'HG13') {
    if ($aa2 ne 'VAL') { $sa2 = $sub{$at2};}
    else { $sa2 = $at2; }
}
else { $sa2 = $sub{$at2}; }

# Print Output

printf TEMP2 "%3d %-4s %-5s %3d %-4s %-5s
%.2f\n",$res1,$aa1,$sa1,$res2,$aa2,$sa2,$dist;

}

close TEMP;
close TEMP2;

# C) DYANA to XPLOR Format change

# Method 1: Use upper boundary alone

if ($method == 1 ) {

    open (IN, $temp2) or die "cannot open $dyana for read :$!";
    open (OUT, "> $xplor") or die "Cannot open $xplor for write :$!";

    while ($line = <IN>) {
        ($res1,$aa1,$at1,$res2,$aa2,$at2,$dist) = $line=~
/((\d+?)\s+(\w+?)\s+(\w+?)\s+(\d+?)\s+(\w+?)\s+(\w+?)\s+(\d+?)\s+\/;
    }
}

```



```

printf OUT "assign (segid %s and resid %3d and name %-4s) (segid %s
and resid %3d and name %-4s) 0 0 %.2f\n",
$name,$res1,$at1,$name,$res2,$at2,$dist;

}
close IN;
close OUT;
}

# Method 2: Use upper boundary with 1.8-angstrom lower boundary

if ($method == 2 ) {

    open (IN,      $temp2) or die "cannot open $dyana for read :$!";
    open (OUT, "> $xplor") or die "Cannot open $xplor for write :$!";

    while ($line = <IN>) {
        ($res1,$aa1,$at1,$res2,$aa2,$at2,$dist) = $line=~
/((\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s/;
        $dmin = $dist - 1.8;

        printf OUT "assign (segid %s and resid %3d and name %-4s) (segid %s
and resid %3d and name %-4s) %.2f %.2f 0\n",
        $name,$res1,$at1,$name,$res2,$at2,$dist,$dmin;

    }
    close IN;
    close OUT;
}

# Method 3: Use NOE class boundaries (upper-limits rounded up)

if ($method == 3 ) {

    open (IN,      $temp2) or die "cannot open $dyana for read :$!";
    open (OUT, "> $xplor") or die "Cannot open $xplor for write :$!";

    while ($line = <IN>) {
        ($res1,$aa1,$at1,$res2,$aa2,$at2,$dist) = $line=~
/((\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s/;

        if ($dist <= $str2) {
            $mid = $str1;
            $lower = $str1-1.8;
            $upper = $str2-$str1;
            printf OUT "assign (segid %s and resid %3d and name %-4s) (segid
%s and resid %3d and name %-4s) %.2f %.2f %.2f\n",
            $name,$res1,$at1,$name,$res2,$at2,$mid,$lower,$upper;
        } elseif ($dist > $str2 and $dist <= $med2) {
            $mid = $med1;
            $lower = $med1-1.8;
            $upper = $med2-$med1;
            printf OUT "assign (segid %s and resid %3d and name %-4s) (segid
%s and resid %3d and name %-4s) %.2f %.2f %.2f\n",
            $name,$res1,$at1,$name,$res2,$at2,$mid,$lower,$upper;
        } elseif ($dist > $med2 and $dist <= $wea2) {
            $mid = $wea1;
            $lower = $wea1-1.8;
            $upper = $wea2-$wea1;
            printf OUT "assign (segid %s and resid %3d and name %-4s) (segid
%s and resid %3d and name %-4s) %.2f %.2f %.2f\n",
            $name,$res1,$at1,$name,$res2,$at2,$mid,$lower,$upper;

```

```

    } elsif ($dist > $wea2 and $dist <= $vwk2) {
        $mid    = $vwk1;
        $lower  = $vwk1-1.8;
        $upper  = $vwk2-$vwk1;
        printf OUT "assign (segid %s and resid %3d and name %-4s) (segid
%s and resid %3d and name %-4s) %.2f %.2f %.2f\n",
        $name,$res1,$at1,$name,$res2,$at2,$mid,$lower,$upper;
    } elsif ($dist > $vwk2) {
        $mid    = $vwk1;
        $lower  = $vwk1-1.8;
        $upper  = $dist-$vwk1;
        printf OUT "assign (segid %s and resid %3d and name %-4s) (segid
%s and resid %3d and name %-4s) %.2f %.2f %.2f\n",
        $name,$res1,$at1,$name,$res2,$at2,$mid,$lower,$upper;
    }
}
close IN;
close OUT;
}

```

## 10.9.2 TALOS2DYANA.PL: TALOS To DYANA Dihedral Angle Converter

### 10.9.2.1 Description

This script converts TALOS dihedral angle output to be used in DYANA. All that is required the input and output files to be defined. The script only converts TALOS data that have a 'Good' rating. Be sure to modify the charge states of the residues if you are operating at a non-neutral pH.

### 10.9.2.2 Code

```
#!/usr/bin/perl -w
# Define Input and Output files:
$in = 'pred.tab';
$out = 'talos.aco';

open (IN,$in) or die "Cannot open $in1 for read :$!";
open (OUT, "> $out") or die "Cannot open $out for write :$!";

# Single to Three letter residue conversion:
my %onetothree = (
    'A' => 'ALA',
    'C' => 'CYS',
    'D' => 'ASP-',
    'E' => 'GLU-',
    'F' => 'PHE',
    'G' => 'GLY',
    'H' => 'HIS+',
    'I' => 'ILE',
    'K' => 'LYS+',
    'L' => 'LEU',
    'M' => 'MET',
    'N' => 'ASN',
    'P' => 'PRO',
    'Q' => 'GLN',
    'R' => 'ARG+',
    'S' => 'SER',
    'T' => 'THR',
    'V' => 'VAL',
    'W' => 'TRP',
    'Y' => 'TYR',
);

# Processing of Talos input:
while ($line = <IN>) {
    ($id, $res, $phi, $psi, $dphi, $dpsi, $dist, $no, $rate) = $line =~
/(\d+)\s+(\w+)\s+(\.\d+\.\d+)\s+(\.\d+\.\d+)\s+(\d+\.\d+)\s+?
(\d+\.\d+)\s+(\d+\.\d+)\s+(\d+)\s+(\w+)\s+/;
    if ($rate eq 'Good') {
        $lphi = $phi - $dphi;
        $uphi = $phi + $dphi;
        $lpsi = $psi - $dpsi;
        $upsi = $psi + $dpsi;
        print OUT "$id $onetothree{$res} PHI $lphi $uphi\n";
        print OUT "$id $onetothree{$res} PSI $lpsi $upsi\n";
    }
}
close IN;
close OUT;
```

### 10.9.3 ACO2TBL: DYANA to XPLOR Dihedral Angle Conversion

#### 10.9.3.1 Description

This script converts the dihedral angle constraints from DYANA to be used in XPLOR. All that is required is the input and output files to be defined.

#### 10.9.3.2 Code

```
#!/usr/bin/perl

# Define Input and Output files:

$segid = 'zasp';
$in     = 'hnha.aco';
$out    = 'zasp_hnha.tbl';

# Processing of Talos input:

open (IN,$in)          or die "Cannot open $in1 for read  :$!";
open (OUT, "> $out") or die "Cannot open $out for write :$!";

while ($line = <IN>) {
    ($sid,$res,$type,$amin,$amax) = $line =~
/(\d+)\s+(\w+)\s+(\w+)\s+(\.\d+)\s+(\.\d+)\s+(\.\d+)\s+(\.\d+)\s/;
    $angle = ($amax + $amin)/2;
    $error = ($amax - $amin)/2;

    if ($type eq 'PHI') {
        $sid1 = $sid - 1;
        $at1 = 'C'; $at2 = 'N';
        $at3 = 'CA'; $at4 = 'C';

        printf OUT "assign (segid %4s and resid %4d and name %2s) (segid %4s
and resid %4d and name %2s)
            (segid %4s and resid %4d and name %2s) (segid %4s and resid %4d
and name %2s) 1 %-4.2f %3.2f 2\n\n",
            $segid,$sid,$at1, $segid,$sid,$at2, $segid,$sid,$at3, $segid,$sid,$at4,
            $angle,$error;

        } elsif ($type eq 'PSI') {
            $sid1 = $sid + 1;
            $at1 = 'N'; $at2 = 'CA';
            $at3 = 'C'; $at4 = 'N';

            printf OUT "assign (segid %4s and resid %4d and name %2s) (segid %4s
and resid %4d and name %2s)
            (segid %4s and resid %4d and name %2s) (segid %4s and resid %4d
and name %2s) 1 %-4.2f %3.2f 2\n\n",
            $segid,$sid,$at1, $segid,$sid,$at2, $segid,$sid,$at3, $segid,$sid1,$at4,
            $angle,$error;
        }
    }
close IN;
close OUT;
```

## 10.9.4 COMPARE.PL: Dyana Restraint Comparison Script

### 10.9.4.1 Description

This script compares the DYANA different distance constraint files, outputting the variations in connectivity, ignoring distance length as a difference.

### 10.9.4.2 Code

```
#!/usr/bin/perl -w

$in1 = 'zasp';           # 1st Restraint File
$in2 = 'zaspcomplex';    # 2nd Restraint File
$out  = 'zasp.upl';      # Difference

open (IN1, $in1) or die "Cannot open $in1 for read :$!";
open (IN2, $in2) or die "Cannot open $in2 for read :$!";
open (OUT, "> $out") or die "Cannot open $out for write :$!";

while ($stab1 = <IN1>) {
    ($res1,$aa1,$at1,$res2,$aa2,$at2,$dist1) = $stab1=~
/(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s/;
    while ($stab2 = <IN2>) {
        ($res3,$aa3,$at3,$res4,$aa4,$at4,$dist2) = $stab2=~
/(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s/;
        if ($res1 == $res3 and $at1 == $at3 and $res2 == $res4 and $at2
== $at4) { last }
        elsif ($res1 == $res4 and $at1 == $at4 and $res2 == $res3 and $at2
== $at3) { last }
        else printf OUT "%3d %-4s %-5s %3d %-4s %-5s
%.2f\n", $res1,$aa1,$at1,$res2,$aa2,$at2,$dist;
    }
}

close IN1;
close IN2;
close OUT;
```

## 10.9.5 DIST\_SEARCH.PL: MOLMOL Distance To XEASY Possible Assignment Comparison

### 10.9.5.1 Description

This script searches the MOLMOL distance output deriving from a structure. This was intended to aid the 'possible assignments' command in XEASY for quick evaluation of the presence with connectivity:

- Calculate possible distances from an NMR ensemble (< 7 Angstroms) from all protons in MOLMOL. Write out the output.
- Input the residue-to-residue connectivity as directed by the script. This will search the distance file, and give an output if it is present.

### 10.9.5.2 Code

```
#!/usr/bin/perl

# CalcDist File:
while (1) {

# User Input:
print "Enter residue number: ";
$ref1 = <STDIN>; chop $ref1;
print "Enter atom type: ";
$atom = <STDIN>; chop $atom;

$q = 'n';

while ($q eq 'n') {

$in = 'CalcDist.txt';
open (IN,$in) or die "Cannot open $in for read :$!";
print "Enter residue number: ";
$ref2 = <STDIN>; chop $ref2;
if ($ref1 < $ref2) {
    while ($line = <IN>) {
        ($res1,$aa1,$at1,$res2,$aa2,$at2,$freq) = $line =~
/(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)/
\s/;
        if ($ref1 == $res1 and $atom eq $at1 and $ref2 == $res2) {
            print "$line";
        }
    }
} elsif ($ref1 > $ref2) {
    while ($line = <IN>) {
        ($res1,$aa1,$at1,$res2,$aa2,$at2,$freq) = $line =~
/(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)\s+(\w+?)\s+(\w+?)\s+(\d+)/
\s/;
        if ($ref1 == $res2 and $atom eq $at2 and $ref2 == $res1) {
            print "$line";
        }
    }
}
print "Assignment Found? (y/n): ";
$q = <STDIN>; chop $q;
close IN;
}}
```

### 10.9.6 COMBINE.PL: XEASY Atom List Combiner

#### 10.9.6.1 Description

This script combines atom lists generated from XEASY (prot files). Data is combined when there is an assignment in 'table 2', which is not present in 'table 1'. The output can then be used again with another XEASY atom list. All that is required is that the input and output files are defined. (Suggested order of input: 15N-HSQC, HNHA, CBCACONH, HNCO, HCCH-TOC, 15N-HSQC-NOESY, 13C-HSQC-NOESY).

#### 10.9.6.2 Code

```
#!/usr/bin/perl

$in1 = 'zasp_pdz_hsqc_27.prot';
$in2 = 'zasp_pdz_13cnoesy.prot';
$out = 'junk';

open (IN1,$in1) or die "Cannot open $in1 for read :$!";
open (IN2,$in2) or die "Cannot open $in2 for read :$!";
open (OUT, "> $out") or die "Cannot open $out for write :$!";

while ($tab1 = <IN1> and $tab2 = <IN2>) {
    ($id1, $cs1, $sd1, $aname1, $res1) = $tab1 =~
    /(\d+)\s+(\d+\.\d+)\s+(\d+\.\d+)\s+(\w+)\s+(0|\d+|\-
\d+)\s/;
    ($id2, $cs2, $sd2, $aname2, $res2) = $tab2 =~
    /(\d+)\s+(\d+\.\d+)\s+(\d+\.\d+)\s+(\w+)\s+(0|\d+|\-
\d+)\s/;
    $total ++;
    if ($cs1 == 999 and $cs2 == 999) {
        printf OUT "%4d %7.3f %5.3f %-4s %3d\n", $id1, $cs1, $sd1,
$aname1, $res1;
        $blanks ++;
    } elsif ($cs1 == 999 and $cs2 < 999) {
        printf OUT "%4d %7.3f %5.3f %-4s %3d\n", $id1, $cs2, $sd1,
$aname1, $res1;
        $subs ++;
    } else {
        printf OUT "%4d %7.3f %5.3f %-4s %3d\n", $id1, $cs1, $sd1,
$aname1, $res1;
        $original ++;
    }
}

$ass = ($subs + $original);
$per = 100 * ($ass/$total);

print "$blanks unassigned out of $total\n";
print "$subs substitutions from list 2\n";
print "$original from list 1, $ass ($per%) Atoms Assigned\n";

close IN1;
close IN2;
close OUT;
```



**10.9.7 CALIBRATE.PL: Volume/Intensity To Distance Calibration****10.9.7.1 Description**

This is a simple distance calibration script which takes uses the prot file and peak lists to generate DYANA distance constraint files. The distance calibration is simply estimated using threshold levels of peak intensity/volume to calculate distances. All that is requires are the input and output files are defined, and the intensity/volume threshold is estimated using the lengths of known distances.

**10.9.7.2 Code**

```
#!/usr/bin/perl

$in1  = 'zasp_noec100_27.peaks';
$out1  = 'sort.txt';

$sortit = 'n';
$subit  = 'y';
$seqit  = 'y';
$distcal = 'y';

# Sort of the peak restraints, deletion of unassigned peaks.

if ($sortit eq 'y') {

    open (IN1,$in1) or die "Cannot open $in1 for read :$!";
    open (SORT, "> $out1") or die "Cannot open $out1 for write :$!";

    while ($line = <IN1>) {
        @peaklist = split(" ", $line);
        $peak = $peaklist[0];
        $int  = $peaklist[5];
        $at1  = $peaklist[9];
        $at2  = $peaklist[10];
        if ($at1 > 0 and $at2 > 0 and $at1 != $at2) {
            if ($at2 < $at1) {
                print SORT "$at2 $at1 $int $peak\n";
            } else {
                print SORT "$at1 $at2 $int $peak\n";
            }
        }
    }
    close SORT;
    close IN1;
}

# Substitution of atom numbers from the 'prot file'.

if ($subit eq 'y') {

    $in2  = 'sort.txt';
    $in3  = 'all1.prot';
    $out2 = 'substitution.txt';

    open (IN2,$in2) or die "Cannot open $in2 for read :$!";
    open (SUB, "> $out2") or die "Cannot open $out2 for write :$!";

    while ($peak = <IN2>) {
        @peaklist = split(" ", $peak);
        $atom1  = $peaklist[0];
        $atom2  = $peaklist[1];
    }
}
```

```

$dist = $peaklist[2];
$peakno = $peaklist[3];
open (IN3,$in3) or die "Cannot open $in3 for read :$!";
while ($prot = <IN3>) {
    @protlist = split(" ", $prot);
    $at = $protlist[0];
    $res = $protlist[4];
    $atom = $protlist[3];
    if ($atom1 == $at) {
        $var1 = $res;
        $var2 = $res;
        $var3 = $atom;
    }
    if ($atom2 == $at) {
        $var4 = $res;
        $var5 = $res;
        $var6 = $atom;
    }
    print SUB "$var1 $var2 $var3 $var4 $var5 $var6 $dist $peakno\n";
    close IN3;
}
}
close SUB;
close IN2;
}

# Substitute residue numbers for residue types.

if ($seqit eq 'y') {

    $in5 = 'substitution.txt';
    $in6 = 'zasp.txt';
    $out3 = 'finished.upl';

    open (IN5,$in5) or die "Cannot open $in5 for read :$!";
    open (FIN, "> $out3") or die "Cannot open $out3 for write :$!";

    while ($upl = <IN5>) {
        @upllist = split(" ", $upl);
        $res1 = $upllist[0];
        $type1 = $upllist[1];
        $atom1 = $upllist[2];
        $res2 = $upllist[3];
        $type2 = $upllist[4];
        $atom2 = $upllist[5];
        $dist = $upllist[6];
        $peak = $upllist[7];
        open (IN6,$in6) or die "Cannot open $in6 for read :$!";
        while ($seq = <IN6>) {
            @seqlist = split(" ", $seq);
            $res = $seqlist[0];
            $resno = $seqlist[1];
            if ($type1 == $resno) {
                $aa1 = $res;
            }
            if ($type2 == $resno) {
                $aa2 = $res;
            }
            if ($dist > 0) {
                print FIN "$res1 $aa1 $atom1 $res2 $aa2 $atom2 $dist $peak\n";
            }
        }
        close IN6;
    }
}
}

```

```

close FIN;
close IN5;
}
# Distance Calibration

if ($distcal eq 'y') {

# Filenames
$in7   = 'finished.upl';
$out4  = 'calibrated.upl';

# Volumes
$istr  = '1200'; # Strong
$imed  = '800';  # Medium
$iwea  = '400';  # Weak

# Distances
$str    = '3.0'; # Strong
$med    = '4.0'; # Medium
$wea    = '5.0'; # Weak
$vwk    = '7.0'; # Very Weak

open (IN7,$in7) or die "Cannot open $in7 for read :$!";
open (CAL, "> $out4") or die "Cannot open $out4 for write :$!";

while ($volume = <IN7>) {
    $count ++;
    @vollist = split(" ", $volume);
    $res1 = $vollist[0];
    $type1 = $vollist[1];
    $atom1 = $vollist[2];
    $res2 = $vollist[3];
    $type2 = $vollist[4];
    $atom2 = $vollist[5];
    $vol   = $vollist[6];
    $peak  = $vollist[7];
    if ($vol >= $istr) {
        printf CAL "%3d %-4s %-5s %3d %-4s %-5s    %.2f # peak
%5d\n", $res1, $type1, $atom1, $res2, $type2, $atom2, $str, $peak;
        $strong ++;
    } elsif ($vol < $istr and $vol >= $imed) {
        printf CAL "%3d %-4s %-5s %3d %-4s %-5s    %.2f # peak
%5d\n", $res1, $type1, $atom1, $res2, $type2, $atom2, $med, $peak;
        $medium ++;
    } elsif ($vol < $imed and $vol >= $iwea) {
        printf CAL "%3d %-4s %-5s %3d %-4s %-5s    %.2f # peak
%5d\n", $res1, $type1, $atom1, $res2, $type2, $atom2, $wea, $peak;
        $weak ++;
    } elsif ($vol < $iwea) {
        printf CAL "%3d %-4s %-5s %3d %-4s %-5s    %.2f # peak
%5d\n", $res1, $type1, $atom1, $res2, $type2, $atom2, $vwk, $peak;
        $vweak ++;
    }
}
print "
Strong   = $strong
Medium   = $medium
Weak     = $weak
V. Weak  = $vweak
Total    = $count\n";

close CAL;
close IN7;
}

```