Structural Characterization of the Protein Tyrosine Phosphatase Shp2 in Solution

by

Piraveen Gopalasingam

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Supervisors:

Professor Michael Overduin and Doctor Mark Jeeves

School of Cancer Sciences

College of Medical and Dental Sciences

University of Birmingham

July 2014

Abstract

Intracellular signalling cascades are mediated by a plethora of receptors, enzymes, adaptors and small molecules. The Protein Tyrosine Phosphatase (PTP) Shp2 is a highly conserved enzyme involved in a myriad of cellular processes including growth, differentiation and apoptosis. Shp2 is multi-domain protein composed of two SH2 domains in tandem, a PTP catalytic domain and a C-terminal tail containing multiple phosphorylation sites and a proline-rich region. The majority of biophysical research has utilised X-ray crystallography to study interactions and effects of mutations at the structural level. To gain a further understanding of Shp2 ligand binding and perturbations caused by disease relevant mutations, a structural investigation was performed with Nuclear Magnetic Resonance (NMR) spectroscopy and Small-angle X-ray Scattering (SAXS) in solution. The NMR signals from the backbone of both SH2 domains were assigned and residue-level interactions and differential SH2 domain specificities with peptides from the novel receptor G6b-B were delineated. In addition, the E76K point mutation that causes Noonan Syndrome and leukaemia was found to have increased conformational dynamics, the first experimental evidence of this phenomenon at the structural level.

Dedicated to Karunanithi Kanagasabai

Acknowledgements

First and foremost I would like to thank Prof. Michael Overduin for his support, guidance and pep talks throughout the project. I would also like to thank Dr. Mark Jeeves for his scientific expertise in all aspects of my project, incredible patience and humour. Because of him I am a better scientist and for that I owe my deepest gratitude. I could not have asked for a better mentor. Thanks also go to Dr. Yotis Senis for all the helpful suggestions and assistance with understanding the messy world of Biology. Thanks also go to the SGC for providing cloning vectors and expertise, Dr. Alistair Barr for providing Shp2 catalytic constructs and Dr. Karen Colwill & the late Prof. Tony Pawson for providing various SH2 domain constructs used throughout this project.

Many thanks also go to current and past members of the Overduin group for their lab assistance, willingness to help, numerous tea breaks, flexibility and for making the daily grind of lab work more than bearable. Thanks to Dr. Timothy “Barthez” Knowles for AUC and SAXS assistance, Marc Lenoir for modelling and docking wizardry. Dr. Claudia Fogl for bringing me back down to Earth and Dr. Fiyaz Mohammed for coaxing me over to the dark side of Crystallography, assistance with understanding crystallographic data and training me for the 2012 Great Birmingham Run. Thanks also go to technicians and co-second mothers Pooja, Sandhya and Jas, HWB-NMR staff and students, in particular Dr. Christian Ludwig and Dr. Sara Whittaker for teaching NMR theory and experimentation, Dr. Rajesh Sundaresan and the biophysics/structural biology community at UoB.

Thanks go next to the Overduin group PhD students: “Catalytic Triads” members Mike ‘Giggsy’ Tong and Riyaz ‘Hocus-Pocus’ Maderbocus for being great friends and labmates. Special thanks go to Caezar ‘Banana-Up’ Al-Jassar for assistance with SAXS analysis. Thanks also go to Vicky and Nazia for helpful discussions around phosphatases and support during write-up. Special thanks also go to my dear friends Penélope “Penny” Rodríguez-Zamora and Daniel Jose Marín-Lámbarri AKA “Dookie”. Honorary mentions also go to Birmingham Cafe Scientifique and Skeptics in the Pub.

During my time here I have been lucky to meet and become friends with many people within the School of Cancer Sciences as well as across the University; the acknowledgements section will not do them justice. Stress relief through nights out in Harborne, Broad Street and outside the West Midlands were made possible by Laura Hindle, Nick Morley, Laura Quinn, Luke Williams, Rachael Cartlidge, Richard Amoroso, Michele Parks, Zineb El-Mihani, Beckie Port and Ditte Hedegaard. Thank you for making life in Birmingham unforgettable, you have all been a genuine joy to share the experience with.

Thanks also go to my friends Kim, Emily and Edd for their camaraderie and support throughout. Special thanks to my friends and housemates Jack and Emma Shepherd, who have put up with my late nights, complaints about noise, general irritability and lack of contribution to house life. I promise to clean the fridge one day!

Thanks to my family and friends in London, my mother and father Jey and Gopal for raising me to value education, my younger brother Chai for his tremendous support and my older brother Adam Deshpande for the many years of friendship and heavy drinking. You all mean the world to me.

Finally, I would like to give my heartfelt thanks to Alexandra Török. I could not have gotten through this period of my life without her love, patience and unwavering friendship.

Contents

List of Figures

List of Tables

Abbreviations

AGE Agarose Gel Electrophoresis

ATP Adenosine Triphosphate

AU Arbitrary Units

AUC Analytical Ultracentrifugation

BEST Band-selective Excitation Short Transient

BSA Bovine Serum Albumin

C6ORF25 Chromosome 6 Open Reading Frame 25

C-terminal Carboxyl-terminal

CTE C-terminal Extension

C2 Calcium (Ca^{2+}) ion binding domain

CD Circular Dichroism

cSH2 C-terminal SH2 domain

CSP Chemical Shift Perturbation

DHAP 2, 6-dihydroxyacetophenone

dITIM double-Immunoreceptor Tyrosine-based Inhibitory Motif

DLS Dynamic Light Scattering

DTT Dithiothreitol

EDTA Ethylenediaminetetraacetic acid

EGFR Epidermal Growth Factor Receptor

EOM Ensemble Optimization Method

ERK Extracellular signal-Related Kinase

FID Free Induction Decay

FL Full-length

FLVR Phe-Leu-Val-Arg

HCl Hydrochloric acid

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HMQC Heteronuclear Multi Quantum Coherence

HSQC Heteronuclear Single Quantum Coherence

IPTG Isopropyl β-D-1-thiogalactopyranoside

INEPT Insensitive Nuclei Enhanced by Polarization Transfer

IRS-1 Insulin Receptor Substrate-1

ITAM Immunoreceptor Tyrosine-based Activation Motif

ITIM Immunoreceptor Tyrosine-based Inhibitory Motif

ITSM Immunoreceptor Tyrosine-based Switch Motif

JAK Janus Kinase

JMML Juvenile MyelomonocyticLeukaemia

LC-MS/MS Liquid Chromatography tandem mass spectrometry

LEOPARD Lentigines, Electrocardiographic conduction abnormalities, Ocular Hypertolerism, Pulmonary stenosis, Abnormal genitalia, Retarded growth and Deafness

LIC Ligation Independent Cloning

MALDI-TOF Matrix Assisted Laser Desorption/Ionisation – Time of Flight

MAPK Mitogen-Activated Protein Kinase

MD Molecular Dynamics

MM Molecular Mass

MS Mass Spectrometry

MUSTER Multi-Sources ThreadER

MWCO Molecular Weight Cut-off

N-terminal Amino-terminal

NMR Nuclear Magnetic Resonance

NRTK Non-Receptor Tyrosine Kinase

NS Noonan Syndrome

NSD Normalized Spatial Discrepancy

nSH2 N-terminal SH2 domain

PCR Polymerase Chain Reaction

PDDF Pair Distance Distribution Function

PDB Protein Data Bank

PDGFR Platelet Derived Growth Factor Receptor

PI-3K Phosphoinositide-3 Kinase

pNPP *p*-nitrophenol phosphate

PPM Parts Per Million

PTK Protein Tyrosine Kinase

PTM Post-Translational Modification

PTB Phosphotyrosine Binding Domain

PTP Protein Tyrosine Phosphatase

pY, pTyr Phosphotyrosine

RMSD Root Mean Squared Deviation

RNA Ribonucleic acid

RPM Revolutions Per Minute

RT Room Temperature

RTK Receptor Tyrosine Kinase

SAXS Small Angle X-ray Scattering

SEC Size Exclusion Chromatography

SH2 Src Homology 2

SH3 Src Homology 3

Shp1 Src Homology 2 containing Phosphatase 1

Shp2 Src Homology 2 containing Phosphatase 2

SPR Surface Plasmon Resonance

SOFAST-HMQC band-Selective Optimized Flip-Angle Short-Transient HMQC

STAT Signal Transducer and Activator of Transcription

Syk Spleen Tyrosine Kinase

TanSH2 Tandem SH2 domain; see also tSH2

TCEP Tris-(2-carboxyethyl) phosphine

TEV Tobacco Etch Virus

TM Transmembrane

TOF Transmitter Offset

tSH2 Tandem SH2 domain; see also TanSH2

WNT int/Wingless

WT Wild-type

XLP X-linked Lymphoproliferative

Zap-70 Zeta-chain Associated Protein kinase 70kDa

Chapter I
Introduction

Phosphorylation

Phosphorylation is the selective and reversible addition of a phosphate moiety onto a target substrate. The substrate may be a small molecule, such as a metabolite or a lipid, or part of a polypeptide. This phenomenon can be observed in all domains of life and as such is a ubiquitous cellular control mechanism (Cohen, 2002).

Serine, threonine and tyrosine residues of proteins are most susceptible to phosphorylation and other post-translational modifications (PTMs), with landmark studies putting the relative Ser/Thr/Tyr phosphorylation levels at 90:10:0.05 (Hunter and Sefton, 1980). These relative levels can increase or decrease depending on signalling cues (Ubersax and Ferrell, 2007). However, histidine, arginine, lysine, cysteine and aspartic acid residues have also been discovered to undergo enzyme-mediated phosphorylation (Matthews, 1995; Lewis *et al*. 1999).

Phosphorylation can modify the function of a protein in a number of ways, by increasing or decreasing its biological activity, initiating or abolishing protein:ligand interactions, aiding or disrupting movement between different cellular compartments and by stabilizing or marking the protein for breakdown. The outlined changes in protein function due to phosphorylation occur as a result of local and/or remote structural changes, causing steric hindrance and charge-mediated attraction and repulsion (Johnson *et al.*, 2001). However, it is possible for phosphorylation to have

no effect on protein structure and function. This is termed ‘silent phosphorylation’ (Huffine, 1996).

It is thought that roughly 30% of the proteins encoded by the human genome are modified by phosphorylation at some point in their cellular existence (Cohen, 2002). Phosphorylation most likely became prevalent as a regulatory mechanism due to the relative abundance of adenosine triphosphate (ATP) as a phosphate donor, along with the simplicity and flexibility of the modification (Pearlman *et al*., 2011).

Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) are a family of enzymes able to phosphorylate the tyrosine residues of specific polypeptide substrates, of which there are receptor bound and cytosolic members. The former have the general domain architecture of an N-terminal extracellular domain, a single transmembrane (TM) domain and a C-terminal cytoplasmic domain. There are 58 receptor PTKs in the human proteome and most of these are activated by a ligand binding to the extracellular domain. These ligands can be growth factors, cytokines, or extracellular domains of other proteins on the same or different cell. Activation causes oligomerization and activation of the cytoplasmic kinase domain by transphosphorylation of the tyrosine residue found in the activation loop (Bossemeyer, 1995). Non-receptor tyrosine kinases (NRTKs) are involved in a myriad of growth factor and signalling cascades. Aside from acting as strict cytoplasmic enzymes, some NRTKs can act as catalytic modulators for membrane-bound receptors that require kinase binding after themselves undergoing phosphorylation (Lemmon & Schlessinger, 2010) . The

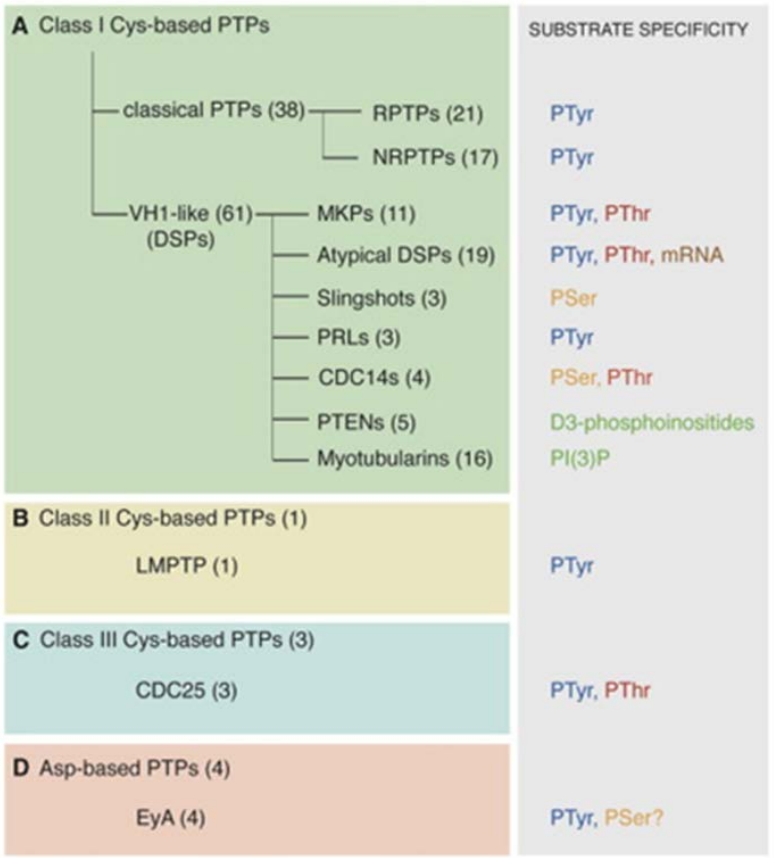
recruitment of NRTKs via PTM adds another layer of regulation to a complex highway of signalling networks (Deribe *et al*. 2010).

Protein Tyrosine Phosphatases

Protein Tyrosine Phosphatases (PTPs) are a highly conserved group of enzymes involved in the removal of a phosphate group from a polypeptide substrate. The various mammalian PTPs are important in the regulation of mammalian cell signalling and due their modular nature many PTPs can exert their functions in specific subcellular compartments at specific times, only being recruited when required and dephosphorylating substrates according to their intrinsic specificity (Zhang *et al*., 1994). As will be explained later, the simple idea that PTKs switch on signalling pathways and PTPs switch them off is now known to be over simplistic and incorrect.

There are 107 PTPs within the human genome, of which 38 make up the 'classical PTPs' which contain catalytic domains of high sequence and structural homology ([Figure 1.1 A]) This is surprising as the number of protein phosphatases is vastly outnumbered by the number of protein kinases, of which there are 518 (Manning, 2002). Clearly there are mechanisms in place to account for this disparity. The classical PTPs can be further separated into two groups, the receptor (rPTPs) and nonreceptor (nrPTPs) phosphatases.

i.



ii.

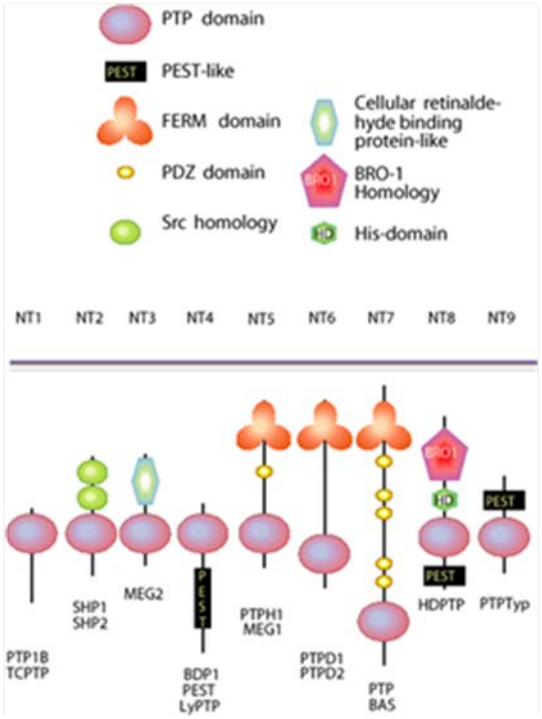


Figure 1.1: The PTPome and Non-receptor PTPs.
(i) The PTPome, classified based on sequence alignments (taken from Alonso *et al.* 2004). **(ii)** Schematic diagram showing the 17 non-receptor PTPs in nine groups. Note that Shp1 and Shp2 are the only members of group NT2. (taken from Andersen *et al.* 2001 and Alonso *et al.* 2004).

PTPs in Signalling

PTPs are involved in most aspects of signalling where phosphorylation is concerned and like PTKs regulate many different aspects of cellular function (Stoker, 2005). PTPs have been found to regulate various intracellular kinase pathways such as mitogen-activated protein kinase (MAPK), extracellular signal-related kinase (ERK), Janus kinase-signal transducer and activator of transcription (JAK-STAT), phosphoinositide 3-kinase PI-3K/Akt and int/Wingless (WNT) signalling, as well as play a particularly crucial role in immune signalling (Pao *et al*. 2007). Their pivotal role in these pathways and the wider function they play in disease contexts is the topic of much research, especially in a therapeutic context (Scott *et al*., 2010; Tonks 2006).

Nonreceptor phosphatases

As seen in Figure 1.1 B, the nrPTP subfamily contains 17 of the classical PTPs. Closer inspection of the 17 nrPTPs shows how diverse they are in modular architecture. It is these modular, independently foldable domains which, through subcellular localisation, increase the specificity of the PTPs they are attached to.

PTPN1, now widely known as PTP1B, was the first tyrosine phosphatase catalytic domain to be purified (Tonks *et al*., 1988). The crystal structure of PTP1B's catalytic domain was solved both free and in complex with a tungstate ion, the latter complex chosen for crystallisation due to sodium tungstate's relatively high IC_{50} (10μM) for PTP1B (Barford *et al*., 1994). The tungstate ion was found within the catalytic cleft interacting with residues important in catalysis, explaining why it had inhibitory effects

on the enzyme. A crystal structure of PTP1B in complex with peptide substrates derived from the Epidermal Growth Factor Receptor (EGFR) intracellular tail was published a year later (Jia *et al*., 1995).

Architecture of the PTP Domain

The catalytic PTP domain is around 280 residues in length and made up of a mixed β-sheet of eight strands, flanked by six α-helices (Figure 1.2 A and B). The active site forms a deep pocket of around 9 Å in depth, located at the base of the catalytic cleft with the conserved "KNRY" motif ensuring selectivity for the side-chain of phosphotyrosine over phosphothreonine and phosphoserine. Selectivity is achieved by the side-chain phenol group of tyrosine within the motif engaging in π-stacking interactions with the aromatic ring of the incoming phosphotyrosine substrate. The loop regions of PTP domains are highly mobile and have been demonstrated to open and close independently of substrate engagement, especially for the WPD loop which has been crystallised as part of PTP domains in open and closed states (Yang *et al*., 2001).

General Mechanism of Dephosphorylation by PTPs

Catalysis is mediated by a number of highly conserved motifs within the phosphatase catalytic domain. These motifs have been identified by sequence alignments, mutagenesis and close inspection of available PTP structures (Alonso *et al*., 2004). The first motif is the catalytic signature motif, which has the consensus sequence CX_5R, where C is the catalytic cysteine and X is any amino acid. In addition a tryptophan-proline-aspartic acid segment (known as the WPD loop) and a glutamine-

containing region dubbed the Q-loop are all positioned near one another (Figure 1.2 C for efficient phosphotyrosine dephosphorylation via the following steps:

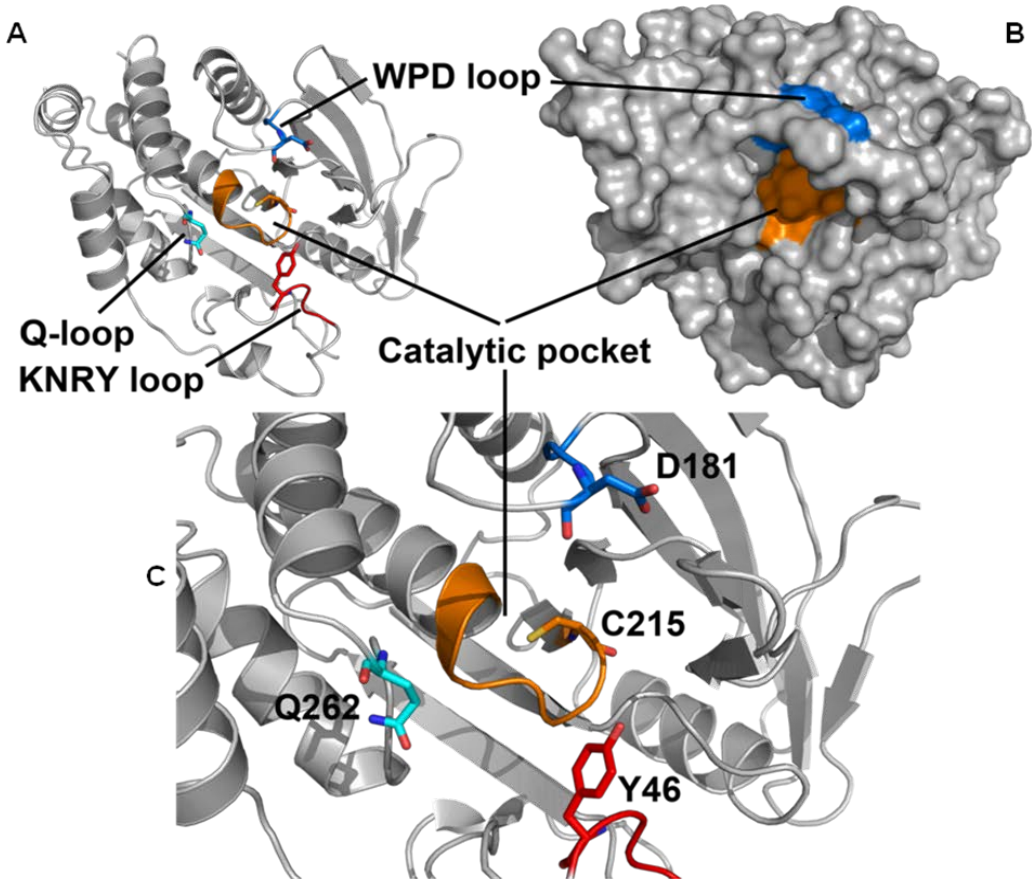


Figure 1.2: Architecture of the PTP Domain.

PTP Domain overview, using archetype PTP1B (PDB ID: 2HNP) **(A)** Cartoon representation (grey) with regions and residues important for catalysis coloured as follows: orange, catalytic signature motif; blue, WPD loop; cyan, Q-loop and red, KNRY loop. **(B)** Surface representation, with catalytic pocket (orange) and WPD loop (blue highlighted for clarity. **(C)** Zoomed region of (A) better highlighting individual residues important for catalysis (labelled in black according to PTP1B numbering).

1) Initial phosphotyrosine binding, aided by specificity pocket adjacent to catalytic pocket.

2) WPD loop transition from the open state to the closed state, bringing the protonated aspartate side-chain into close proximity of phosphate group.

3) Arginine residue within catalytic sequence binds phosphotyrosine

4) Cysteine thiolate ion acts as a nucleophile, attacking phosphate group and forming the phosphocysteine intermediate.

5) Q-loop coordinates a bound water molecule to hydrolyse reaction intermediate

6) Aspartate group of WPD loop acts as a general base to neutralize the leaving tyrosine side-chain.

For more information, see (Barford *et al*., 1998).

Src Homology 2 (SH2) Domains

It is estimated there are over 10,000 sites across the proteome that undergo tyrosine phosphorylation in the eukaryotic cell (Bernard A. Liu *et al*., 2012). Some of these phosphorylated sites serve as docking platforms for transient interactions of varying affinities. The specificity of protein:protein interactions is governed by a number of factors, including spatiotemporal regulation, tissue-specific protein and isoform expression, subcellular localization and degradation rates. However, this is an incomplete view of the complexities behind cell:cell and intracellular communication; the diverse roles observed in multiple signalling pathways are a result of the various

modular domains connected to enzyme effectors, fine-tuning all aspects of signal transduction.

Modular domains have evolved and diverged in order to fulfil a specific function or set of functions in the cell and importantly, within the protein. Some modular domains have enzymatic activity whilst others serve as adaptor proteins capable of regulating cellular processes without explicitly harbouring catalytic activity. These adaptor proteins and domains are non-catalytic but modulate catalysis through their interactions with other domains, proteins and biological ligands as part of the same, or distinct polypeptide chain (Bhattacharyya *et al*., 2006). In the past 30 years there has been an explosion of research into molecular determinants of cell signalling pathways, accelerated further by the use of genome sequencing technologies. Through the boon of knowledge it has provided, as well as the discovery and improvements in sequence alignment algorithms, it has come to light that modular domains are widespread in the mammalian cell. One of these, the SH2 domain is of importance within phosphotyrosine signalling as a crucial regulatory domain involved in bringing numerous proteins together to initiate signalling cascades.

After the initial discovery of the SH2 domain sequence in Fujinami Sarcoma Virus (Sadowski *et al*., 1986), highly conserved SH2 sequences were soon found in other proteins involved in signal transduction. It is now well established that the SH2 domain is a eukaryotic, primarily metazoan protein domain, though some unicellular eukaryotes have been found to contain genes encoding SH2 domains; this being due

to a small number of proteins containing what is known as tandem SH2 domains- two SH2 domains in one polypeptide sequence (Bernard A Liu *et al*., 2012).

SH2 domains are usually found as part of multidomain proteins that have a range of functions, spanning phosphorylation:dephosphorylation to ubiquitination, transcriptional regulation via STAT proteins (Reich & Liu, 2006) and guanine nucleotide exchange. The protein SH2D1A (also known as SAP) only consists of a single SH2 domain and a C-terminal 25 residue unstructured tail, (Cannons *et al*., 2011) highlighting that SH2 domains are not just footnotes in multidomain proteins but can be self-contained functional proteins in their own right.

SH2 domains are folded polypeptides around 100 amino acid residues in length with a highly conserved molecular structure consisting of a central β-sheet formed from multiple antiparallel strands flanked by two α-helices; one on either side of the sheet (Figure 1.3). The helices are named αA and αB while the strands are named βA to βG according to the naming convention by Michael Eck and colleagues (Eck *et al*., 1993). While the SH2 domain secondary structures are largely conserved, a few members of the SH2 family exhibit slight structural differences which relate to their role in signalling pathways (Kaneko *et al*., 2012). A number of conserved sequences within the SH2 domain polypeptide are involved in the recognition and binding of phosphorylated tyrosine motifs and are described below. It is worth mentioning that the SH2 domain is not the only domain capable of binding pY-peptides. Phosphotyrosine binding (PTB) domains (Uhlik *et al*., 2005) have also been shown to bind pY-containing peptides, as has the calcium ion binding (C2) domain of PKCδ (Benes *et al*., 2005) and the HYB domain of E3 ubiquitin ligase Hakai (Mukherjee *et al*., 2012).

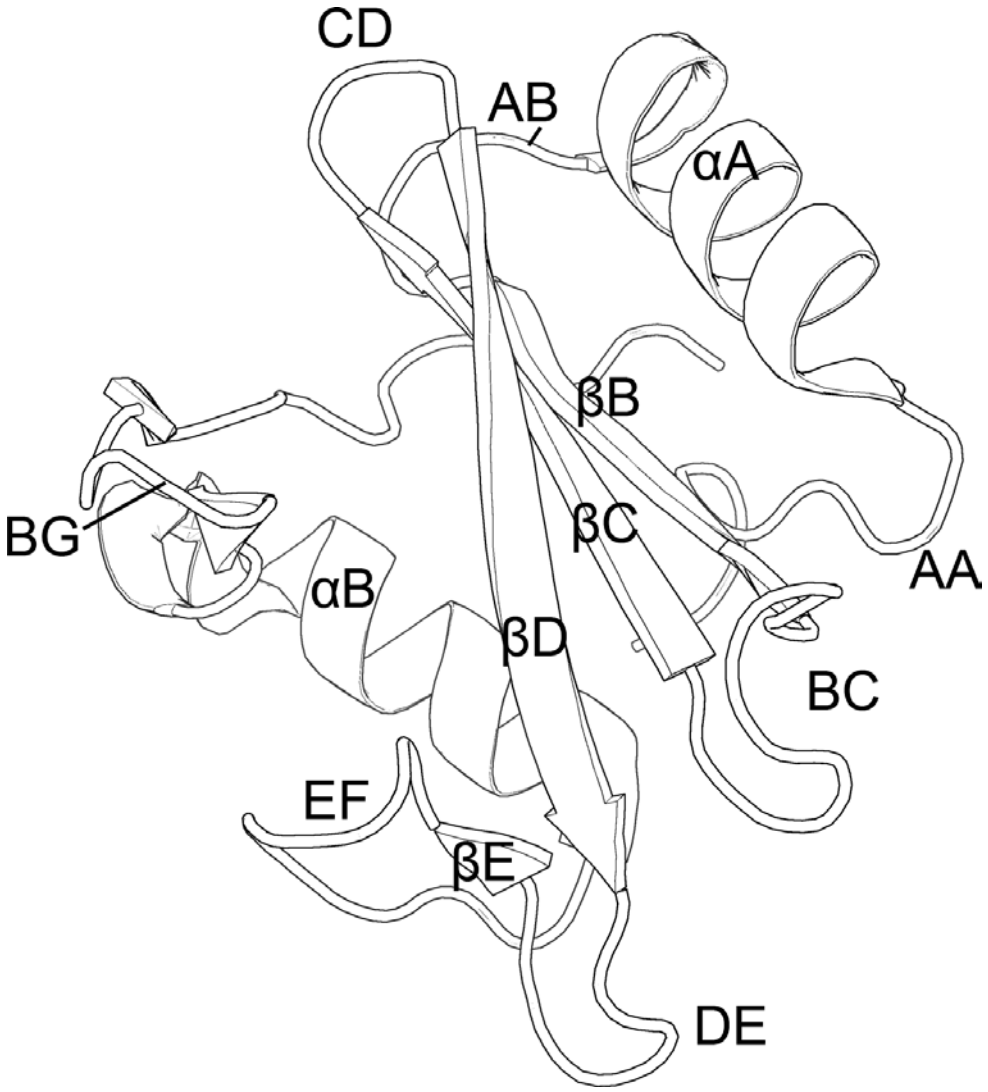


Figure 1.3: Molecular structure of the SH2 Domain.

The SH2 domain (PDB ID: 1AYD, white and black cartoon) is an independently folded modular polypeptide capable of binding pY-containing peptides and is composed of two α-helices flanking a central set of antiparallel β-sheets (βB, βC and βD) with various loop regions playing structural or functional (BC and BG) roles. Secondary structure naming convention as in Eck *et al.* (1993).

The phosphotyrosine binding pocket contains an invariant arginine residue that forms a bi-dentate salt-bridge with the phosphate group of phosphotyrosine in the bound state (Figure 1.4 C). Mutation of this critical residue abrogates phosphopeptide binding (Bibbins *et al*., 1993). This conserved Arg is part of the flavor motif, or FLVR (residues Phe-Leu-Val-Arg, invariant Arg underlined). This sequence makes up most of the βB sheet, with the phosphopeptide binding perpendicular to the entire β-sheet in an extended conformation.

Immediately C-terminal to the flavor motif is the BC loop which acts as a hinge. In the "open" state it is available to bind phosphopeptide, adopting a "closed" state upon interaction with pY phosphate oxygens, its main binding partner in the SH2:phosphopeptide interaction (Figure 1.4 A). An overall positive charge within the pY binding pocket is maintained by various residues. The side-chains of the conserved histidine βD4, lysine βD6 and arginine αA2 also help maintain the charged state for phosphotyrosine docking, though the latter two also interact with the aromatic ring of phosphotyrosine. Thermodynamic studies of SH2:phosphopeptide interactions posit that the phosphotyrosine binding pocket provides around half the binding energy during complex formation (Bradshaw *et al*., 1998), with the remainder driven by burial of hydrophobic phosphopeptide residues into the SH2 domain specificity pocket as well as liberation of bound water from the SH2 pocket into the bulk water.

The EF and BG loops are located in the second half of the SH2 domain sequence and both form portions of the +3 specificity pocket. This pocket is important in defining the specificity for SH2 domains toward certain ligands by the shape and charge of the pocket (Figure 1.4 B and C).

.

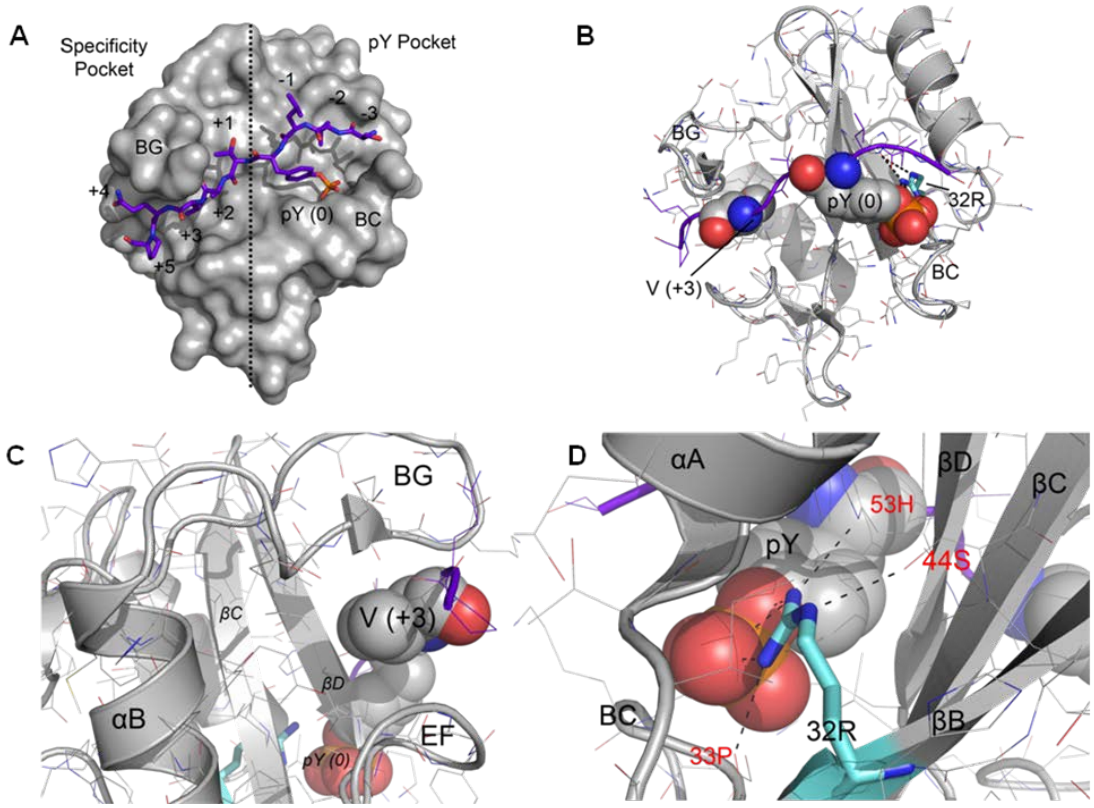


Figure 1.4: A Closer Look at SH2:Phosphopeptide Interactions.

The N-terminal SH2 domain of Shp2 in complex with the 11-mer PDGFR pY-1009 phosphopeptide (PDB ID: 1AYA). **(A)** Surface representation of the SH2 domain (grey), with peptide depicted as purple sticks. In general, the pY pocket (right side of line) binds pY whilst the specificity pocket (left side of line) binds residues C-terminal to pY. Numbers depict peptide residue positions in relation to pY. **(B)** Cartoon and stick form of the SH2 domain (grey) with peptide (purple ribbon) bound. Space-filling spheres depict either pY (0) or V (+3). The invariant Argβ5 (32R) is coloured cyan. Black dashes indicate polar contacts with 32R. **(C)** and **(D)** show zoomed specificity and pY pockets respectively. For (D), red text represents SH2 residues forming salt bridges with 32R.

The residues of the hydrophobic pocket are less conserved than those that make up the positively-charged pocket for phosphotyrosine-docking and have classically been thought to recognise hydrophobic residues at the +3 position, with respect to phosphotyrosine. This has since been studied extensively using combinatorial peptide arrays, which have shown that SH2 domains can sense phosphopeptide residues as far as the +4 and +6 positions (Imhof *et al*., 2006). Mutation of residues within the C-terminal hydrophobic pock*et al*ter peptide binding specificity and affinity (Bradshaw *et al*., 1998), further demonstrating that this pocket plays a critical role in determining the domain's role in signalling pathways. Based on a study by Huang *et al* (Huang *et al*. 2008) using a novel peptide-array approach, the specificity space of the SH2 domain was heavily investigated, leading to new insights regarding the preferences of SH2 domains for amino acids at all positions of a phosphopeptide. The end result of this study was the reordering of the entire SH2 family into new groups based on theses binding preferences. Herein the +3 pocket will be referred to as the "specificity pocket".

The amino acid at position βD5 of the SH2 domain interacts with amino acids at the +1 and +3 position of the phosphopeptide, playing a critical role in determining peptide specificity.

The binding affinity between SH2 domains and their ligands vary considerably; equilibrium dissociation constants between the nanomolar and low micromolar range (Ladbury *et al.*, 1995). This allows for the formation of transient complexes of various residence times, controlling signalling cascades with exquisite temporal precision. However, SH2 domains are capable of binding non-optimal phosphopeptide sequences (Lee *et al*., 1994).

Modifying the on/off rates of SH2:ligand complexes by mutation can lead to changes in normal signalling function (Kaneko *et al*., 2010). In other experiments, it was demonstrated that substituting key residues in an SH2 domain to that of another SH2 domain can alter the phosphopeptide binding preferences of the former to resemble that of the latter (Songyang *et al*., 1995).

Despite the high sequence and structural conservation of SH2 domain family members, it is of note that a few still manage to exert novel functions. The JAK1 SH2 domain exhibits characteristics atypical for SH2 domains. A loss-of-function mutant did not affect subcellular localization, nor did it affect downstream signalling pathways (Rui *et al*., 1997). JAK1 instead plays a structural role with binding partners (Radtke *et al*., 2005). It has also been observed that the SH2-only adaptor protein SAP is capable of binding unphosphorphorylated tyrosine-containing motifs (Poy *et al*., 1999) and the Vav1 SH2 is able to bind two closely spaced pY residues at once; engaging both the pY and specificity pockets (Chen *et al*., 2013). In addition, the c-Abl SH2 domain can bind phosphoinositide 4,5 bisphosphate-containing liposomes (Tokonzaba *et al*., 2006). Clearly, SH2 domain interactions are numerous and versatile in binding partners, deviating from the classical view of the domain in the 1990s.

Mutations in the SH2 domains of certain proteins can lead to certain disease states. For example, deletion or mutation of the SAP SH2 domain causes X-linked lymphoproliferative (XLP) syndrome (Nichols *et al*., 1998) and many mutations within the SH2 domains of Shp2 lead to cancer, with germline mutations causing two related childhood syndromes (Bentires-Alj *et al*., 2004a).

Tandem SH2 domains

A minority of proteins encoded in the human genome contain two SH2 domains: spleen tyrosine kinase (Syk), zeta-chain associated protein kinase 70kDa (Zap-70), SH2-containing Phosphatases 1 and 2 (Shp1 and Shp2, respectively) and the p85α subunit of PI-3K, playing an important role in modulating protein function (Filippakopoulos *et al*., 2009). The yeast genome encodes Spt6, a transcription elongation factor which harbours a tandem SH2 domain containing both a canonical SH2 and noncanonical SH2 domain with novel binding properties (Diebold *et al*. 2010).

For Syk and Zap-70, which make up the Syk family of PTKs, both SH2 domains are separated by a large coiled coil linker region and exhibit conformational flexibility when bound to dually-phosphorylated peptide. Interestingly for Zap-70, the C-terminal SH2 cannot effectively bind phosphopeptide without forming an interface with the N-terminal SH2 (Hatada *et al*., 1995). The same is not true for the Syk tandem SH2 when binding dual phosphopeptides, though there is still an interface between both SH2 domains present (Fütterer *et al*., 1998).

Shp1 and Shp2 tandem SH2 domains lack a long inter-SH2 linker, instead the two SH2 domains are separated by a 6-8 amino acid linker with significant contacts between it and both SH2s (Yang *et al*., 2003a; Hof *et al*., 1998).

The presence of two SH2 domains in proteins is thought to not only increase overall binding affinity between them and cognate ligands in signalling pathways, but contribute additional specificity for ligands whose pY residues are spaced at an optimal distance for binding the SH2 domain-containing proteins of choice. Another

study by Eck *et al*. demonstrated that varying the linker length between two phosphorylated peptides modulates the relative affinity of tandem SH2 domain toward them (Eck *et al*., 1996). The individual SH2 domains within a tandem domain demonstrate different peptide binding specificities, fine-tuning interactions where many potential binding partners are available (Isakov *et al*., 1995; Imhof *et al*., 2006). Phosphopeptide binding orientations of either SH2 domain within a tandem domain can also modulate affinities, only permitting binding where it is spatially favourable (Cowburn, 1996).

Although structures of the Syk and ZAP-70 tandem SH2 domains in complex with dually phosphorylated peptides exist, no such structures have been solved of Shp1 and Shp2 tandem domains.

Shp2

Initial Discovery and Characterisation

The PTP Shp2 was first cloned in and purified the early 1990s (Adachi *et al*., 1992; Freeman *et al*., 1992; Sugimoto *et al*., 1993; Zhao *et al*., 1994) and since then a plethora of studies spanning many fields of biology have been conducted to investigate various aspects of its structure and function (Neel *et al*., 2003; Poole & Jones, 2005). Shp2 is ubiquitously expressed, with high levels observed in the brain, skeletal muscle, heart and testis (Ahmad *et al*., 1993). Homozygous knockout of the *PTPN11*gene leads to embryonic lethality in mice, demonstrating the important role Shp2 plays in development. Knockout mouse studies further established the role of

Shp2 in development, supported by many *in vitro* cell studies of wild-type (WT) and mutant forms of Shp2 (Yang *et al.*, 2006).

Gene and Isoforms

Shp2 is encoded by the *PTPN11* gene, situated on the long arm of chromosome 12 at position 24 (*Ch12q24*) (Isobe *et al*., 1994). There are three known isoforms, of which the most well characterised is the 593 amino acid form. The slightly longer, 597 amino acid form is a result of ribonucleic acid (RNA) splicing and contains a tetrapeptide sequence in the PTP domain capable of reducing catalytic activity (Mei *et al*., 1994), though this study was performed in rats and not repeated in human cell lines.

Similarity to Shp1

Shp1 is another member of the SH2-containing PTP family. The Shp1 and Shp2 protein sequences exhibit 61% overall identity (Figure 1.5) (Stein-Gerlach *et al*., 1998) and the folds of their individual domains are also very similar (Figure 1.6). However, a notable structural deviation resides within the Shp1 C-terminal SH2 (cSH2) domain, which takes an alternative domain orientation to the Shp2 cSH2 domain with respect to the N-terminal SH2 (nSH2) and PTP domains. The overall structures of both cSH2 domains are nearly identical however. Whilst Shp2 is ubiquitously expressed (Adachi et al., 1997), Shp1 expression is confined to certain

blood cells and plays a largely inhibitory role in cell signalling, in contrast to the activatory nature of Shp2 (Haque *et al.*, 1998).

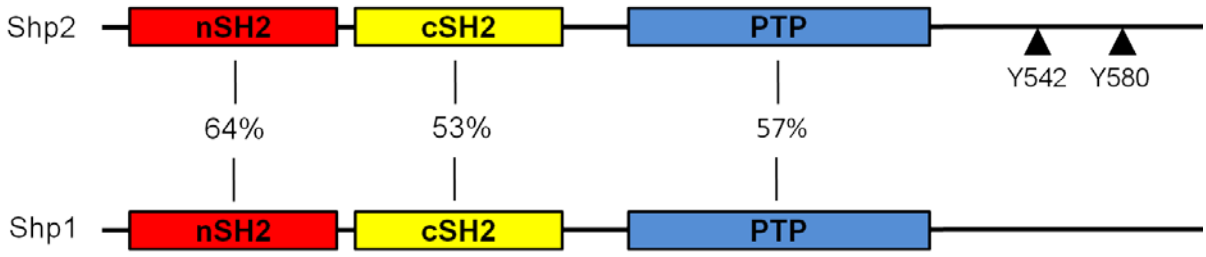


Figure 1.5: Sequence Identity of Shp1 and Shp2 Phosphatases.

Shp1 and Shp2 possess 61% sequence identity overall. Black filled triangles represent important phosphorylation sites in Shp2. Figure adapted from Stein-Gerlach *et al.*, 1998.

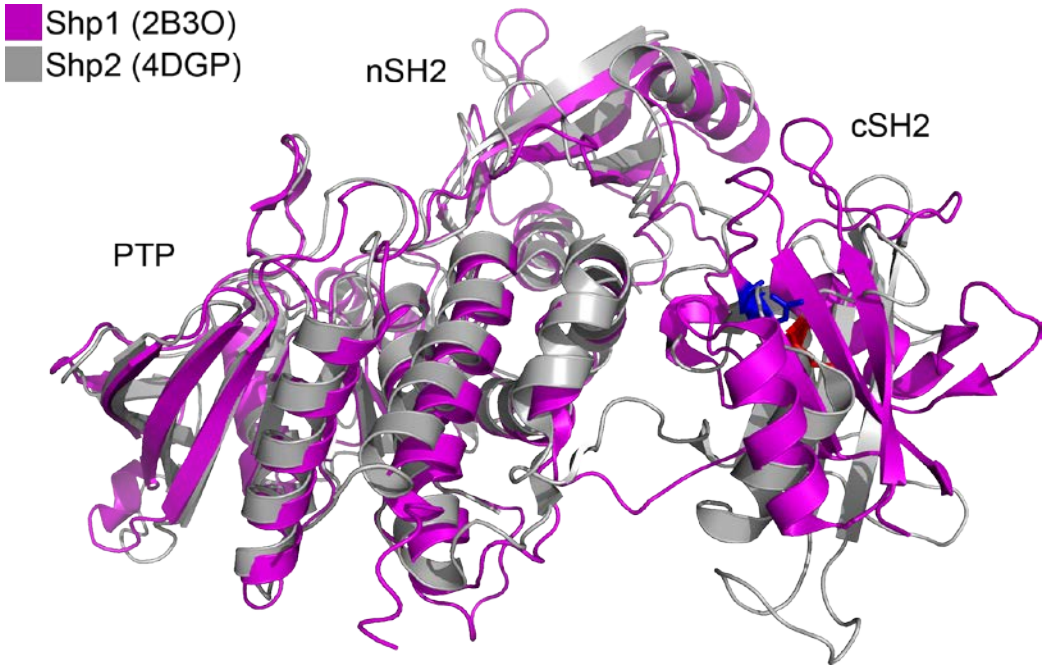


Figure 1.6: Structural Comparison of Current Shp1 and Shp2 Crystal Structures.

Shp1 (PDB ID: 2B3O, purple) and Shp2 (PDB ID: 4DGP, grey) were superimposedusing PyMol v1.6. An RMSD of 1.67 Å was observed, with the nSH2 and PTP domains of both phosphatases aligning well. However, the cSH2 domain orientation is highly divergent, with the Shp1 cSH2 tilting ≈50° toward the nSH2. Shp1 and Shp2 ArgβB5 residues are highlighted in blue and red, respectively.

Shp2 in Cell Signalling

Being a bonafide PTP, Shp2 would be expected to play a generally inhibitory role in cell signalling pathways by remove phosphate groups from ligands that were previously phosphorylated and possible activated by one of many cellular and membrane-bound tyrosine kinases (Lemmon & Schlessinger, 2010). However, many cell-based studies of Shp2 have consistently demonstrated that this is not the case. For example, deletion of Shp2 in mesenchymal cells leads to growth retardation due to downregulated MAPK (Lapinski *et al*., 2013) . Knockout of Shp2 catalytic activity via the well characterised C459S substrate-trapping mutant has also shown largely inhibitory consequences in downstream growth factor pathways (Ren *et al*., 2010). It was later discovered that Shp2 dephosphorylates proteins inhibited by the presence of phosphotyrosine; removal of the phosphate groups serve to activate the inactive substrates and cause the activation and upregulation of downstream effectors (Figure 1.7) (da Silva *et al*., 2013). Shp2 catalytic activity is also not a prerequisite for signalling function, as catalytic-dead Shp2 knockouts still maintain partial function (Yu *et al*. 2003). In addition to MAPK/ERK pathway activation, Shp2 plays a role in regulating JAK-STAT signalling, a pathway important in modulating interferon receptor function (You *et al*., 1999).

Most research has found that Shp2 is localised to the cytosol and inner leaflet of the plasma membrane where cognate ligands and substrates co-localise (Incoronato *et al*., 2004). Shp-2 has also been found to co-localise with lipid rafts (Kim *et al*., 2006). However, roles for Shp2 in the nucleus (Jakob *et al*., 2008) and mitochondria (Büchner *et al*., 2010) have revealed Shp2 function is pivotal to normal cell function.

Not only is signalling activity mediated by the PTP domain of Shp2, but by the C-terminal tail. In particular, two well-studied tyrosines, Y542 and Y580, can be

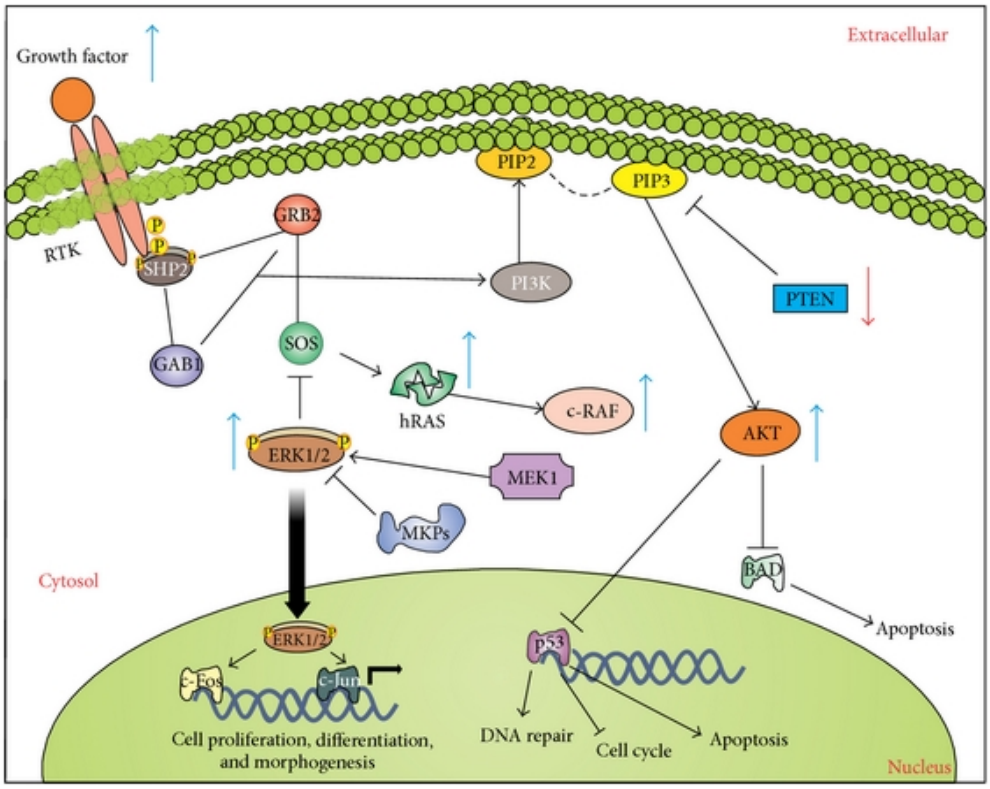


Figure 1.7: Shp2 Signalling.

Phosphorylation of growth factor tyrosine residues (yellow circles, P) causes recruitment of Shp2, in turn recruiting Gab1 and Grb2 adapter proteins which can bind and activate downstream signalling molecules leading to changes in gene expression. Figure taken from de Silva *et al.,* 2013.

reversibly phosphorylated and are even required for normal ERK activation (Araki *et al*., 2003).

In addition to phosphorylation-dependent signalling activity, the Shp2 C-terminal tail contains a proline-rich motif of sequence PLPPCTPTPP which could bind Src Homology 3 (SH3) domains of a variety of signalling proteins such as Src (Walter *et al*., 1999), a central protein in phosphotyrosine signalling. Shp1 however, only contains two consecutive proline residues in its C-terminal tail.

Shp2 can therefore not only act as an enzyme but have functional roles as an adapter protein, since it can bind specific pY-peptides via its two SH2 domains or be bound to by other proteins via their SH2 and SH3 domains. It has also been shown previously that C-terminal residues Y542 and Y580 can be phosphorylated and activate Shp2 catalytic activity (Lu *et al*., 2001; Lu *et al*., 2003), adding another level of complexity, as spatiotemporally separated phosphorylation could fine tune Shp2 function *in vivo*.

Shp2 Structure

At the time of writing there are 22 crystal structures in the protein databank (PDB) of various domains of Shp2 and its homologues, and one additional published crystal structure which has not been submitted to the PDB (Eck *et al*., 1996). These can be summarised in Table 1.

PDB ID	Domain(s)	Ligand	Reference
1AYA	nSH2	PDGFR 1005-1015	Lee, CH.
1AYB	nSH2	IRS-1 887-898 (SPGEpYVNIDFGS)	Lee, CH.
1AYC	nSH2	PDGFR 736-744 (DGGpYMDMSKGS)	Lee, CH.
1AYD	nSH2	N/A (WT Syp)	Lee, CH.
N/A	tSH2	PDGFR 1005-1015	Eck, MJ.
2SHP	tSH2, PTP	Dodecane-Trimethylamine	Hof, P.
3B7O	PTP	Malate ion	Barr, AJ.
3O5X	PTP	3-{1-[3-(biphenyl-4-ylamino)-3-	Zhang, X.
3MOW	PTP	(2Z)-2-[(1R)-3-	Liu, S.
3TKZ	nSH2	Synthetic peptide RVIpYFVPLNR	Zhang, Y.
3TL0	nSH2	Synthetic peptide RLNpYAQLWHR	Zhang, Y.
4DGP	tSH2, PTP	N/A (WT)	Yu, ZH.
4DGX	tSH2, PTP	N/A (Y279C mutant)	Yu, ZH.
4H34	tSH2, PTP	N/A (Q506P mutant)	Qiu, W.
4GWF	tSH2, PTP	N/A (Y279C mutant)	Qiu, W.
4H1O	tSH2, PTP	N/A (D61G mutant)	Qiu, W.
4JE4	nSH2	N/A (monobody)	Sha, F.
4JEG	cSH2	N/A (monobody)	Sha, F.
3ZM0	PTP	N/A (to be published)	Bohm, K.
3ZM1	PTP	N/A (to be published)	Bohm, K.
3ZM2	PTP	N/A (to be published)	Bohm, K.
3ZM3	PTP	N/A (to be published)	Bohm, K.
4JMG	C-terminal	“Clamp Ptpn11_pY580” modified Grb2	Yasui, N.

Table 1: **Details all current crystal structures of various Shp2 domains and their bound ligands, if any, and reference.**

Crystal structure of the Syp-nSH2 domain

The mouse homologue of Shp2, Syp, is 99.5% conserved in primary sequence when compared to the human homologue, amounting to three conservative amino acid substitutions: one of which is close to the PTP signature motif, the other two residing in the C-terminal tail. The SH2 domains of mouse Syp and human Shp2 are 100% identical in sequence, of which the nSH2 domain has been subject to successful crystallisation and structure determination of the unbound, non-specific peptide-bound and two high-affinity peptide bound forms (Lee *et al*., 1994) .The peptides were derived from known tyrosine phosphorylation sites that Syp-SH2 and other SH2 domains bind; these are pY895 of IRS-1 and pY740 and pY1009 of PDGFR.

The crystal structures show a typical SH2 domain fold reminiscent of others previously solved by other groups, even with the low sequence identity, although there are a few differences (Waksman *et al*., 1992; Overduin *et al*., 1992). There is a small insertion within the BG loop that forms an anti-parallel β-strand, which interacts with the phosphopeptide, but this insertion is not only present in the Syp-nSH2 domain. The loops that make up the hydrophobic patch are less conserved in the features they present on the surface, which as discussed earlier plays a critical role in determining peptide specificity.

Another marked difference is the lack of an arginine in the αA2 position; the side-chain of which forms an amino-aromatic interaction with the pY ring as well as other hydrogen bonds with parts of the phosphopeptide. In nSH2-Syp's case the conserved arginine is replaced by a glycine which does not interact with the phosphopeptide and is not compensated by the presence of another basic residue

nearby interacting with the pY. Interestingly, the cSH2 domain of Shp2, as well as both the nSH2 and cSH2 domains of Shp1 also harbour glycine residues at this position instead of Arg at position αA2 (Hof *et al*., 1998; Yang *et al*., 2003b). The presence of a non-basic residue at this point could have an influence on the phosphopeptide binding activity of Shp phosphatases toward their substrate. However, the highly conserved ArgβB5 is present and positioned in a similar fashion to previous pY bound-states solved for SH2 domains, though the absence of ArgαA2 causes a re-orientation of the pY side-chain in order to form the equivalent number of hydrogen bonds to Syp-nSH2 (Machida & Mayer, 2005).

LysβB6 and HisβD4 side-chains wedged with one edge of the pY ring, the histidine being well conserved and critical throughout most SH2 domains, as mutations abolish interactions and normal signalling (Marengere & Pawson, 1992). LysβB6 itself does not interact with phosphopeptide. The LysβB6 conformation is not well conserved, though it is ordered in the structure, hydrogen bonding with two side chains in BC loop, stabilizing loop conformation for phosphate docking. The Syp-nSH2 crystal structures provided more evidence that SH2 domains are well conserved structurally and bind phosphopeptides in a very similar way to each other.

Crystal structure of the Tandem SH2 domain of Shp2

The tandem SH2 domain (TanSH2) of human Shp2 in complex with two mono-phosphopeptides revealed a canonical SH2:ligand interaction between the nSH2 and cSH2 domains connected in tandem by their wild-type linker region. Whilst the TanSH2 of Shp1 also has a short linker between its SH2 domains, it is worth noting

that other TanSH2 domain-containing proteins have much longer linkers, forming stable secondary structure elements (Described in Tandem SH2 Domains section above). In the present TanSH2 structure, the linker is well-ordered and forms part of the interface between the two domains and their pY-binding clefts are far apart, almost perpendicular to one another. Cys104 and Cys174 form a disulphide bridge, which is peculiar since Shp2 is a cytosolic protein and thus expected to be in a reducing environment, making disulphide bridge formation difficult.

Phosphopeptide binding is near enough identical to previously determined SH2:ligand structures; between the two SH2 domains that make up the tandem there is an root-mean-square deviation (RMSD) of 0.56 Å when comparing the 44 Cα that make up the helix A and the β-sheet. The interaction between helix B and the β-sheet is altered due to the presence of an alanine in the βB FLVR sequence, instead of the conserved valine. The βD loops between both SH2 domains also occupy divergent positions.

Interestingly, it was found that changing the length of the linker between two monophosphopeptides connected together altered the rate of phosphatase activity, with four-five amino hexanoic acid molecules being optimal and anything more or less diminishing activation rates. In addition, when testing peptides of known higher affinity toward one SH2 domain over another, single peptides weren't sufficient to fully activate Shp2, nor was a bisphosphorylated peptide where the peptide positions were switched so that the C-terminal peptide would be N-terminal and *vice versa*. Only a bisphosphorylated peptide in the correct orientation with four aminohexanoic acid molecules separating them was sufficient for full activation (Eck *et al.*, 1996).

Crystal structure of Shp2: TanSH2 and PTP Combined

The crystal structure of human Shp2 lacking the last 66 amino acids of the C-terminus showed with no ambiguity how the TanSH2 domains played an important role in regulating phosphatase activity (Hof *et al*., 1998). The nSH2 domain sits on one face of the classical PTP domain (Figure 1.8), forming many intramolecular, polar contacts, amounting to a surface area of 1208 $Å^2$ on each participating subunit. In contrast, the cSH2 domain lies on the opposite side, maintaining a few contacts with the PTP domain. A portion of the nSH2 domain, in particular the DE loop, makes extensive contacts with the catalytic cleft (Figure 1.9 A). The DE loop, made up of residues 58-63, is unique within the SH2 family, only being present in Shp1 and is conserved within Shp-type proteins of other species. This loop, and the β-strands on either side of it, bind deep into the phosphotyrosine binding site of the PTP domain, mimicking substrate interaction and preventing closure of the WPD loop, ensuring an inactive form of the enzyme.

Outside of this region, a number of interactions are formed between the nSH2 and PTP domain, further stabilizing the interdomain closed conformation. (Figure 1.9 B)

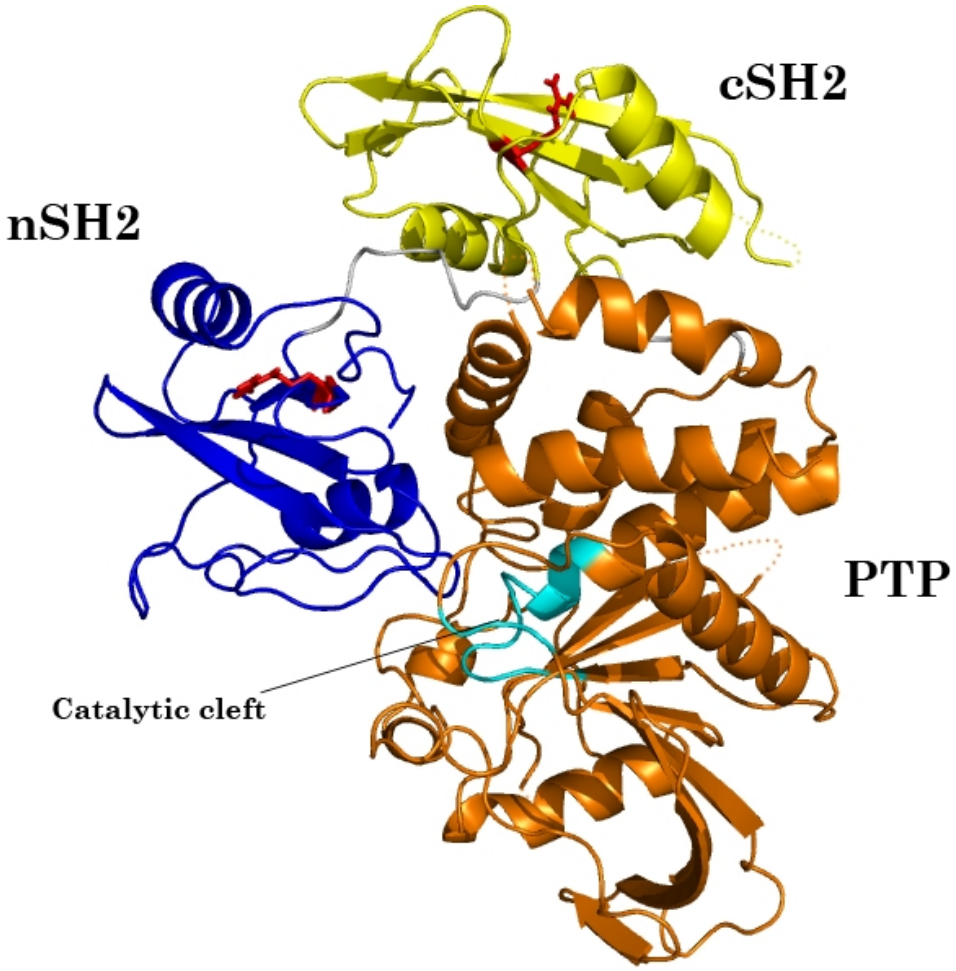


Figure 1.8: Cartoon Representation of Shp2.

(PDB ID:2SHP) Dark blue: nSH2, Yellow: cSH2, Orange: PTP domain, Cyan: catalytic cleft, Red sticks: conserved Arginine βB5 residues, Grey: loops connecting domains, dashed lines show missing loop regions.

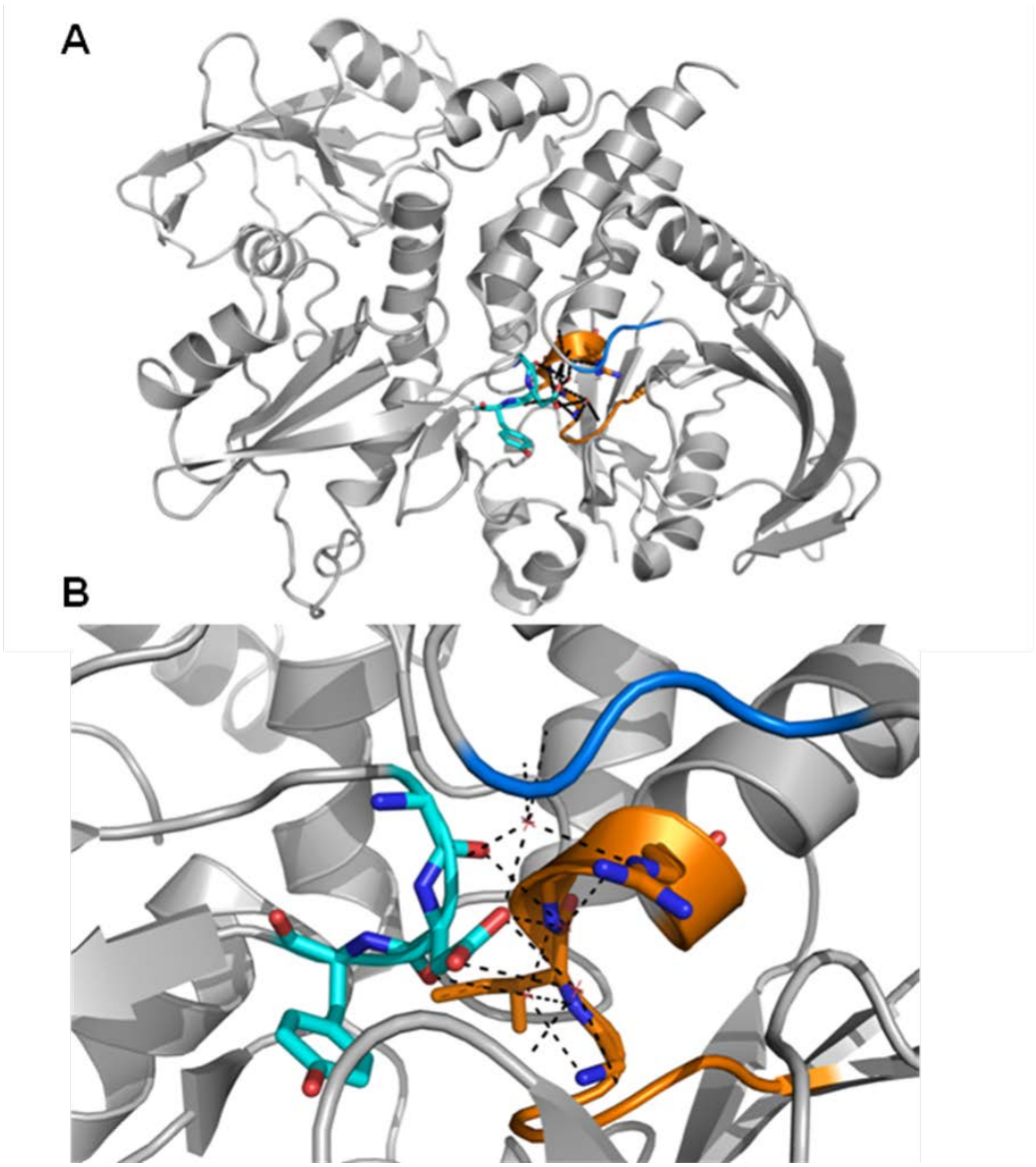


Figure 1.9: Mechanism of Shp2 Inhibition.

(A) Cartoon representation of Shp2 (PDB ID: 4DGP) in the autoinhibited state (cyan sticks: DE loop residues 60G, 61G and 62Y; orange signature catalytic motif 458-466, residues 462-465 shown as sticks; mid-blue: WPD loop; black dashed lines: polar interactions). **(B)** Zoomed region of (A) showing interactions between nSH2-DE loop and PTP-catalytic cleft residues via bound water (red bunched lines).

More recent crystal structures of Shp2, both WT and mutant, have yielded a greater insight into the consequence of disease-relevant mutations (Qiu *et al*., 2014). These are discussed further below.

Shp2 in Disease

Pathogens employ a plethora of strategies to enable successful infection to occur. For many species of bacteria, this involves the secretion of certain virulent effectors that affect host-cell machinery. The gram-negative bacterium *Heliobacter pylori* in particular, using a type IV secretion system, injects many of these effectors into gastric epithelial cells, one of which is called CagA. The strain of *H. pylori* determines the virulence of CagA, Western CagA being far less virulent than Eastern CagA strains. This is due to the presence of certain protein sequences termed EPIYA motifs in Eastern CagA which can become phosphorylated on tyrosine residues by host-cell protein kinases and bind to the cell membrane. Once this has occurred, binding of these motifs by Shp2 occurs, increasing PTP activity and activating RAS/MAPK signalling pathways, leading to growth and proliferation of the host cells. This leads to the initiation of tumours that can lead to gastric cancer (Hatakeyama, 2004; Lochhead & El-Omar, 2007).

There are a large number of germline mutations within the Shp2 gene that have biochemical consequences during development, ultimately leading to a disease phenotype from birth (Huang *et al.*, 2014). Interestingly, two syndromes have been linked to mutations in Shp2, both of which present with similar symptoms but may

have different underlying pathogenesis. These are the Noonan and LEOPARD syndromes. These syndromes fall under the umbrella term ‘RASopathies’, since their pathogenesis can be traced back to mutations in components of the RAS/MAPK signalling cascade. Both syndromes are discussed below, with further analysis of the genetic, biochemical and structural features of mutated Shp2.

Noonan Syndrome

Noonan syndrome is a disorder apparent from birth characterised by congenital heart defects such as pulmonary valve stenosis, short stature, underdeveloped bones and a characteristic facial dysmorphology. Partial deafness and blood disorders are also fairly common (van der Burgt, 2007). It affects between 1:1000 and 1:2500 children born and currently has no cure; some patients go onto develop a rare blood cancer called Juvenile Myelomonocytic Leukaemia (JMML), though this is dependent on the nature of the germline mutation behind Noonan syndrome (Jongmans *et al*., 2011). The molecular basis of Noonan syndrome causation has been elucidated and will be discussed below.

LEOPARD Syndrome

LEOPARD syndrome, as mentioned earlier, has a similar phenotype to Noonan syndrome. LEOPARD is an acronym which stands for Lentigines, Electrocardiographic conduction abnormalities, Ocular Hypertolerism, Pulmonary stenosis, Abnormal genitalia, Retarded growth and Deafness, after some of the

symptoms that patients presented. These are not the only symptoms however, with palpebral ptosis, hypertrophic cardiomyopathy and other heart defects distinct from pulmonary stenosis also concurrent in the syndrome. Phenotype-genotype correlations are difficult for LEOPARD syndrome as there many symptoms and many different amino acid substitutions within the same amino acid position.

Shp2 in Cancer

As well as gastric cancers developing due to Eastern strain *H. pylori* infection, Shp2 has also been implicated in the formation of other cancers. As mentioned previously a subset of JMML patients harbour mutations in their *PTPN11* gene. Sporadic solid tumours of the colon and stomach, as well as the brain have all been found to contain *PTPN11* mutations (Bentires-Alj *et al*., 2004b). Shp2 overexpression has also recently been linked to breast cancer progression (Aceto *et al.,* 2012). As breast cancer is the most common cancer in people, efforts are being made to understand the role of enzymes, particularly PTPs, in breast cancer initiation, progression and metastasis (Aceto *et al.*, 2012).

Functional Consequences of Shp2 Mutation in Disease

Through genetic analysis of Noonan syndrome patients it was discovered that missense mutations were present in the *PTPN11*gene (Tartaglia *et al*., 2003; Gelb and Tartaglia, 2006). Mutations in Shp2 account for approximately 50% of Noonan syndrome cases, with mutations in other RAS pathway members such as the

proteins KRAS, son of sevenless-1 (SOS1) and RAF1 being responsible for other incidences.

Structural mapping of Noonan syndrome mutations onto the Shp2 structure revealed the majority were localised to the nSH2:PTP interface (Tartaglia *et al*., 2002), with one mutation situated in the interSH2-linker region: D106A. Few mutations are found in the cSH2 domain; these localise to a residues within the conserved FLVR motif: E139D and affect pY-peptide binding properties (Keilhack *et al*., 2005). Biochemical analyses of Noonan syndrome Shp2 mutants demonstrate a gain-of-function effect where catalytic activity is markedly increased close to fully-activated levels. Ligand binding is disrupted in some cases (Martinelli *et al*., 2008), causing downstream signalling consequences such as mistargeting of Shp2 to non-canonical binding partners in a mutation-dependent manner (Müller *et al*., 2013). One of the most active mutants observed is E76K, a charge switch which is posited to disrupt nSH2:PTP binding (Darian *et al*., 2011).

The pathogenesis of LEOPARD syndrome with regards to Shp2 mutation is less understood than its Noonan syndrome counterpart. Whilst the latter disorder harbours mutations that increase enzyme activity and alter SH2 domain binding properties, the former seems to manifest from a catalytically inactive phenotype. The majority of LEOPARD syndrome amino acid substitutions in Shp2 are localized to the PTP domain.

There is speculation as to how two syndromes harbouring mutations within the same gene lead to similar phenotypes through different biochemical pathways (Edouard *et al*. 2007). Since Shp2 has both enzymatic and adaptor functions, having a

catalytically deficient PTP domain whilst maintaining phosphopeptide binding capacity via its SH2 domains (and perhaps the PTP active site) could mean LEOPARD syndrome mutants only partially knockout function (Kontaridis *et al*., 2006). The presence of the C-terminal tail also means signalling proteins are still able to phosphorylate residues and bind Shp2 via pSer/pThr/pTyr residues or its proline-rich motif.

Structural Insights into Shp2 Dysregulation

Earlier in 2014, a number of crystal structures of Shp2 were solved characterising Noonan and LEOPARD syndrome mutants at the atomic level (Qiu *et al*., 2014). The authors found that depending on the mutation, different mechanisms leading to the same effect were observed. For D61G, the electrostatic surface of the DE loop where D61 is normally located changes from highly negative to neutral. D61 is said to act as a phosphate mimic and the removal of the negatively charged side chain by genetic mutation ultimately leads to a diminished interaction between the nSH2 domain DE loop and PTP catalytic cleft. For the mutation E139D, no charge removal or reversal is observed, but shortening of the side-chain by one methylene group is enough to cause the BC loop of cSH2-Shp2 to become more ordered than in the WT structure and have higher basal activity. Y279C not only disrupts nSH2:PTP binding but also perturbs the pTyr binding site within the PTP domain, impairing efficient catalysis. Interestingly, although the aforementioned structures were solved and presented, no structural data are available for the E76K mutant which causes JMML and Noonan syndrome.

A molecular dynamics study predicted that E76K mutation would cause the nSH2 domain to release from the PTP domain (Darian *et al.*, 2011). Interestingly the authors found that mutation of D61 may cause reorientation of the connected EF and BG loops into a conformation more amenable to pY-binding or closely resemble the "open" state of the nSH2 domain during uncoupling of the nSH2:PTP domain interaction.

Final remarks

Though much work has characterised Shp2 in various aspects of its biology, including its roles in signalling pathways, biochemical perturbations upon mutation, clinical symptoms in a genetic disorder context, and many crystallographic studies, there has been little biophysical and high-resolution solution structural research dedicated to studying Shp2 and its interactions.

Immunoreceptor Signalling and Protein G6b-B

Membrane bound receptors can receive signals by the interaction of one or more of their extracellular domains with its cognate ligand. The induced conformational change upon ligand binding causes transmission of information through the transmembrane helix resulting in further conformational changes resulting in the exposure of one or more potential binding and/or phosphorylation sites essential for signalling.

A large number of receptors are expressed on the cell surface and the expression of these is often cell-specific. Signalling is invariably a complex process involving hundreds of interactions which together induce a concerted change in gene expression. Delineation of these pathways will allow a wider understanding of signalling in general as well as knowledge of the specific system of interest. This knowledge is essential in understanding those pathways of therapeutic potential. The modulation of immune signalling is a heavily researched field in an industrial context, as many human diseases and ailments have an immune component. For example, immunomodulation during chemotherapy treatment or after organ transplants are important in increasing survival and success rates in these therapies (Kandalaft *et al*., 2010).

Immunoreceptors are a subset of membrane receptors found mainly on the surface of cells such as B-cells, T-cells and platelets, the latter of which has been found to play a role in immune regulation (Smyth *et al*., 2009). As these cells are responsible for processes such as immune memory, antimicrobial defence and blood clotting, careful regulation of the signalling systems is needed to ensure minimal aberrant signalling as they may have serious consequences for health.

Within immunoreceptors there are a number of consensus peptide motifs which when phosphorylated can recruit adaptor proteins and enzymes in a specific manner to initiate a specific signalling cascade. These peptide motifs are termed ITIM (Immunoreceptor Tyrosine-based Inhibitory Motif), ITAM (Immunoreceptor Tyrosine-based Activation Motif) and ITSM (Immunoreceptor Tyrosine-based Switch Motif) and will be discussed further below.

ITIMs

ITIMs are characterised by the consensus sequence: (S/I/V/L)-X-pY-X-X-(I/V/L), containing either a branched chain amino acid or a serine residues at the -2 position, a single phosphotyrosine and a hydrophobic branched chain amino acid at the +3 position. ITIMs inhibit various processes and exert function by binding Shp1, a generally inhibitory phosphatase (Billadeau & Leibson, 2002).

ITAMs

ITAMs are characterised by the consensus sequence: pY-X-X-(L/I)-$X_{6–12}$-pY-X-X-(L/I). The presence of the longer spacer region and dual pY residues could mean that an ITAM could bind two independent proteins each with a single SH2 domain, or binding a single protein with two SH2 domains. ITAMs bind proteins that upregulate signalling cascades leading to activation of downstream cellular processes. ITAMs are thought to bind Syk and/or ZAP-70, tandem SH2 domain containing tyrosine kinases (Barrow & Trowsdale, 2006).

ITSMs

ITSMs are characterised by the consensus sequence T-X-pY-X-X-(V/I) and are able to activate signalling pathways depending on the binding partner they recruit (Chemnitz *et al*., 2004).

Noncanonical ITIMs/ITAMs

Another class of peptide sequences involved in immune regulation, the so-called hemITAMs, are essentially half an ITAM motif i.e they only have one pY and lack the 6-12 amino acid linker followed by another pY that would normally be part of an ITAM sequence. In addition, hemITAMs contain an upstream triacidic amino acid stretch required for signalling (Hughes *et al*., 2013).

There also exist ITAM sequences with ITIM-like function in specific cellular context, these are dubbed inhibitory ITAM or ITAMi sequences (Blank *et al*., 2009).

G6b-B

The G6b protein is encoded by *C6ORF25* (Chromosome 6 Open Reading Frame 25), located in the MHC III cluster region. G6b belongs to the Ig (Immunoglobulin) superfamily and is expressed as seven distinct alternative transcripts – some of which are soluble Research regarding the precise mechanisms of function has focussed on G6b's novel role in blood signalling, specifically on the cell surface of platelets and megakaryocytes (Mori *et al*., 2008). Human G6b-B, a 241 amino acid protein, contains a predicted single extracellular Ig domain followed by a single-pass transmembrane (TM) region and an intracellular tail containing a membrane proximal proline rich region and a number of Tyr, Thr and Ser residues further C-terminal (de Vet *et al*. 2001a). The most well characterised phosphorylation sites at the time of writing are Y211 and Y237. Both are constitutively phosphorylated at these positions (Mazharian *et al*. 2013). Site Y211 is an ITIM whilst Y237 is an ITSM. However, for simplicity they will be referred to as ITIM1 and ITIM2 respectively. The sequence of

ITIM1 is EPSLL(pY)ADLD whilst the sequence of ITIM2 is DASTI(pY)AVVV. Both pY residues are separated by 26 residues and interestingly, there is basic region (RPRR) which could interact with negatively charged species.

G6b-B has been shown by co-immunoprecipitation experiments to bind Shp1 and Shp2 (de Vet *et al*. 2001b). Furthermore, it was recently demonstrated that the ITIMs of G6b-B have varying affinities for Shp1 and Shp2 (Coxon *et al.,* 2012). It was also demonstrated that the SH2 domains of Shp1 and Shp2, apart from nSH2-Shp1, had varying affinities for the various ITIM sequences. More recently, conditional mouse knockout experiments by our collaborators found that mice lacking G6b-B showed bleeding defects as a consequence of diminished platelet function (Mazharian *et al*. 2012). Further experiments demonstrated that Shp1, Shp2 and G6b-B have overlapping and functionally distinct signalling pathways in megakaryocytes and platelets important for function of these cell types (Mazharian *et al*. 2013).

Concluding Remarks

Clearly there is much to be learnt about the interplay of Shp1 and Shp2 in immune signalling pathways. The novel immunoreceptor G6b-B has many unanswered questions associated with it: what are the molecular determinants of Shp1 and Shp2 interaction? Based on the C-terminal tail sequence and structure, can we predict other potential binding partners? On a side note, it has been shown in the prototypical ITIM-containing receptor PECAM-1 is able to bind phospholipid membranes (Paddock *et al*., 2011) and this leads to functional consequences in downstream signalling pathways. As G6b-B contains ITIM sequences and other

similarities with PECAM-1, it would be interesting to see whether phospholipid binding plays a role in G6b-B regulation.

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy, is a powerful technique capable of characterising the structure, dynamics and motions of molecules ranging from small compounds (including water) to megadalton macromolecular complexes (Marion 2013; Yee *et al*. 2006). At the heart of the method is the physical phenomenon known as nuclear magnetic resonance. Exploitation of the so-called NMR has been successful in answering many pertinent biological questions (Grzesiek & Sass, 2009) and has been pivotal in exploring the realms of "Unstructural biology" where proteins with little to no structure can be characterised at the atomic level (Tompa, 2011). In this section the physical basis for the NMR signal is discussed, moving onto chemical shift and chemical exchange phenomena and thereafter an overview of protein NMR techniques used in the present study.

Origin of NMR Signal

For any given atomic nucleus there are a number of intrinsic properties one can use to describe it. One of these properties is the spin quantum number, denoted *I,* and is dependent on the number of unpaired neutrons and protons such that having an odd number of either will usually result in a non-zero non-integer *I* value while having odd numbers of both translate to an integer *I*. Nuclei with an even number of neutrons and protons have an *I* value of 0 (Hore, 1995). Nuclei with *I*=0 are unable to produce an NMR signal as they do not have a magnetic moment; common isotopes with *I*=0

include ^{12}C and ^{14}N. However, ^{1}H, ^{13}C and ^{15}N all have quantum spin numbers of ½ and are NMR-active since the quantum spin number of a nucleus has an associated magnetic moment μ which is proportional to its gyromagnetic ratio γ by the equation:

$$\mu = \gamma I$$

Nuclei are able to populate a number of orientations when placed in an externally applied magnetic field (B_0, the direction of which is the z-axis, by convention) equal to $2I+1$, meaning for I=½ nuclei this is two states, dubbed the α and β-states. This is possible due to interaction of the nuclear magnetic moment with B_0. In the absence of of B_0, a degenerate state of equal energy is populated by the nuclei. In the presence B_0 spin ½ nuclei can orient themselves either parallel (α-state) or antiparallel (β-state) to B_0 and as a consequence they populate one of two discrete energy levels, where α-state is lower and β-state is higher, as it requires more energy to align against B_0. However, nuclei can transition between these two states as they are in thermal equilibrium. States are populated thermally according to the Boltzmann distribution:

$$N_\alpha/N_\beta = \Delta E/2kT$$

Where N denotes the number of nuclei in either α or β-state, k is the Boltzmann Constant, ΔE is energy difference and T is thermodynamic temperature.

The populations of nuclei in α and β-states in B_0 is roughly 50:50, but with slightly more nuclei in the lower (α) state. This energy state anisotropy is miniscule; but it is exactly this that can be exploited in NMR. The presence of a slight net magnetic moment of bulk nuclei in the α-state can be termed a longitudinal magnetic moment. In spectroscopy ΔE is calculated by way of the Planck Relation ($\Delta E = h\nu$, where h is

the Planck constant and *v* is frequency), and ΔE in the context of NMR is the energy difference between the two energy states.

The frequency at which a given nucleus precesses perpendicular to an external magnetic field can be described by the equation:

$v = \gamma B_0$

This is also known as the Larmor frequency and is dependent on the gyromagnetic ratio of the nucleus of interest and the externally applied magnetic field. Due to this relationship, increasing B_0 field strength also increases ΔE between α and β-states (Figure 1.10).

This means the application of energy at a frequency corresponding to ΔE between nuclei in α and β-states causes energy transitions of magnetic nuclei between the two adjacent states. Radiofrequency (RF) pulses are used as they happen to match the frequency at which magnetic nuclei can undergo energy transitions. When applied at 90° to B_0, nuclear magnetisation "flips" from the +z-axis (B_0) to the –y-axis (i.e perpendicular). As the magnetic moment of the nucleus is rotating, a current is induced in a coil surrounding the container in which the nuclei are situated. This current is induced for as long as there is magnetic moment transverse to the z-axis. Transverse relaxation (T_2) occurs, where the nuclear magnetisation gradually returns to the +z-axis; the changes in coil current over time from the end of RF pulsing to the end of transverse relaxation shows a decrease in current induction, this is the NMR signal and is known as Free Induction Decay (FID).

Fourier transformation of the FID results in conversion of the FID from the time-domain to the frequency domain.

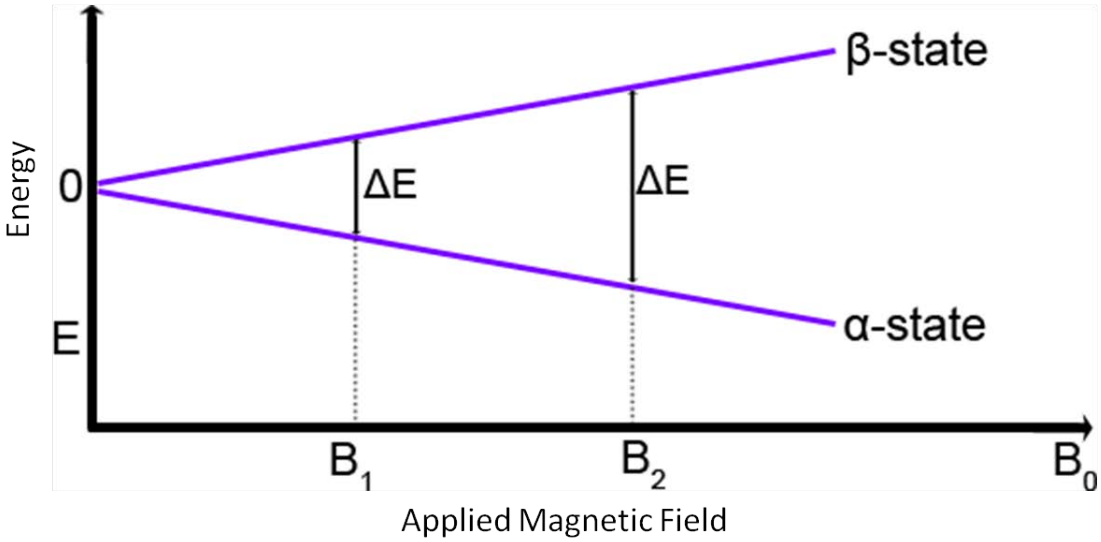


Figure 1.10: Relation Between Field Strength and Nuclear Energy State.

As external magnetic field strength increases (B_0), so does the energy difference (ΔE) between two adjacent energy states of a magnetic nucleus. Adapted from: http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/images/evonb_1.gif.

Chemical Shift

Since the frequency at which an NMR-active nucleus undergoes energy transitions is B_0-dependent, converting the frequencies into dimensionless values independent of B_0 is performed by use of a reference standard such as tetramethylsilane (TMS) or, as is commonly used in protein NMR, the strong water peak. The conversion equation is:

$$\delta = 10^6 [(\nu - \nu_{ref})/\nu_{ref}]$$

Where V is the frequency and V_{ref} is the reference frequency. The unit for δ is parts per million, or ppm and is deemed the chemical shift.

In practice, the chemical shift of a nucleus is affected by the induced magnetic field caused by the electron cloud that surrounds itself as well as electrons from neighbouring nuclei in a chemical. The electron cloud itself is affected by bond geometry, chemical groups nearby, binding partners and transient interactions. This induced magnetic field effectively “shields” the nucleus from the full extent of the external magnetic field such that the local magnetic field experienced is slightly less than the actual B_0 value. It is this subtle effect that allows one to characterise various chemicals by NMR through knowledge of the chemical shift values certain functional groups yield. For example, the region of a ^{1}H NMR spectrum one would normally find amide groups ($-NH_2$) is around 7-10 ppm.

Though it is nuclei-dependent, in general the chemical shift of a nucleus is sensitive to its surrounding, its so-called “chemical environment”. For example, if a nucleus is exposed to the solvent, transient dipoles and interactions may occur, affecting nuclear shielding caused by the changes in electron density distribution. The same

effects caused by the solvent may not apply in a different solvent, where different interactions may occur. Thus the chemical shift of a nucleus is dependent on its surroundings. Motional processes within a molecule may cause its nuclei to experience slightly different chemical environments during its normal internal motions by modulating the extent of shielding processes.

Chemical Exchange Processes

Chemical exchange is the process whereby conformational changes occur around a relatively rigid chemical i.e. a protein backbone, causing fluctuations in local electronic environments.

In the context of proteins, a multitude of exchange processes are known of, ranging from very slow conformational changes on the minutes to seconds timescale, to fast processes such as molecular rotation (ranging milliseconds to nanoseconds), to ultrafast motions such as molecular vibrations on the pico- to femto-second timescale. These processes may have a wider function in protein stability and function and are thus important to understand. NMR is well suited to this challenge as it can be used to observe such processes spanning multiple timescales (Kleckner & Foster, 2011).

Two-site exchange is the process whereby a certain nuclear spin within a molecule experiences two distinct chemical environments, leading to different chemical shift values between the two states. Within this there is symmetrical and asymmetrical two-site exchange. In the former, both states are equally probable whereas in the latter, the probability of one state being populated is higher than the other. With two-

site exchange comes three exchange regimes whereby two-site exchange occurs on different timescales. These are slow, intermediate and fast two-site exchange. In practice the terms do not necessarily mean a transition from one state to another happens slowly, rather that it does not happen often. The actual transition between chemical environments might indeed be fast (µs-ns) but the residence times, i.e the time *spent* in a given chemical environment, might range s-ns.

Crucially, two-site exchange can be observed by NMR, meaning its effects are observable and even quantifiable. The effects of a two-site exchange process on an NMR spectrum are dependent on the chemical shift difference frequency between the two states (A and B) being sampled by the nuclear spin. This is defined as:

$$\Omega_{\Delta} = \Omega^{0}_{A} - \Omega^{0}_{B}$$

where Ω^{0}_{A} and Ω^{0}_{B} are the chemical shift frequencies in Hertz between states A and B. If the exchange rate (*k*) between states A and B ~ Ω_{Δ}, the process is said to be in intermediate exchange. Within this are slow intermediate and fast intermediate exchange where *k* is < or > $|\Omega_{\Delta}/2|$ respectively. k can also equal $|\Omega_{\Delta}/2|$ i.e the "cross-over point" (Levitt, 2007). Collectively, these three exchange phenomena are within the spectral timescale of ms-µs. Figure 1.11 outlines the effect of two-site exchange processes on NMR lineshape.

In the contexts of protein:ligand interactions and secondary conformations, the two-site exchange phenomenon can be observed and can yield information regarding their nature. For example, it is generally accepted that a protein:ligand interaction in slow exchange has an estimated dissociation constant (K_D) of < 3 µM (Williamson, 2013). This makes it fairly straightforward to characterize the binding affinity of a

protein:ligand interaction qualitatively. For a K_D > 3 µM, unbound:bound peaks are likely in fast exchange. In intermediate exchange, the K_D rests somewhere between the aforementioned values; in the high nanomolar to micromolar range.

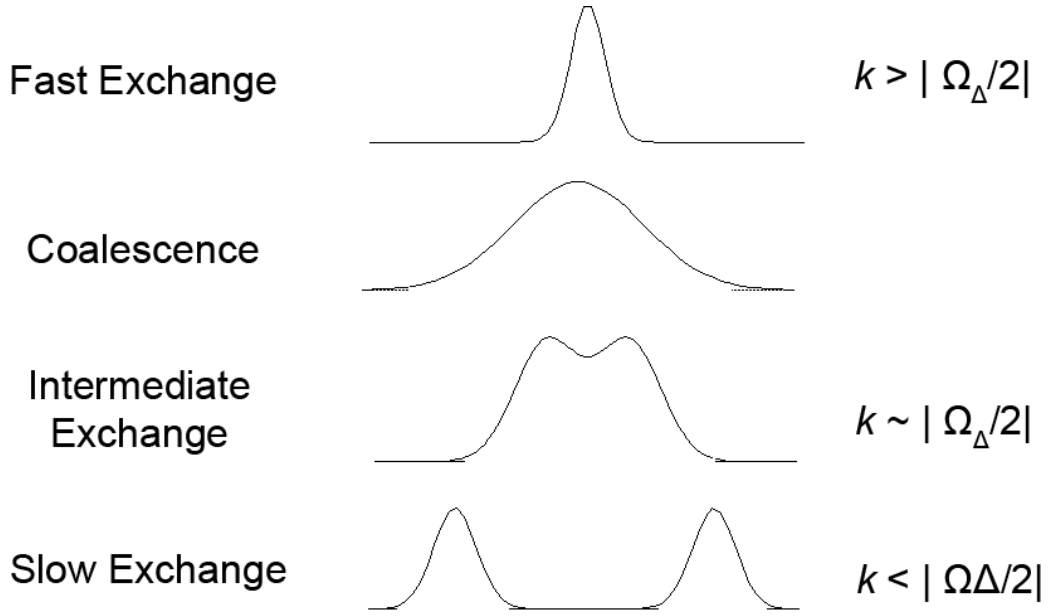


Figure 1.11: Schematic of Two-site Exchange Processes Visualised by NMR.

Range of two-site exchange processes and their effects on an NMR lineshapes. Adapted from http://web.nmsu.edu/~snsm/classes/chem435/Lab8/exchange.gif

Pulse Sequences

NMR experiments are composed of a set of specific pulses for magnetisation transfer and decoupling as well as chemical shift evolution periods, with the acquisition period at the end of an experiment, followed by a recycle delay period (d1, giving the system enough time to reach thermal equilibrium, allowing the restoration of longitudinal magnetisation in time for the beginning of the next set of pulses. Put together, the various components are known as a pulse sequence, of which many exist for breadth of NMR experiments, ranging simple 1D acquisition experiments of ^{1}H NMR data, to highly complex multidimensional experiments for backbone assignment (Reynolds, 2010).

1D-NMR

The first spectrum collected of most protein NMR samples is a 1D ^{1}H spectrum and can quickly tell the user whether the sample is folded or not by checking the dispersion of peaks across the range (1-10 ppm usually) and for a small group of peaks < 0.5 ppm corresponding to high field methyl groups, suggestive of a packed hydrophobic core of a protein. Solvent suppression is performed to reduce the contribution of the water signal to the NMR spectrum, as it dominates the ^{1}H spectrum if left unsuppressed. There are a number of strategies that can be used to achieve good solvent suppression (Price, 1999).

^{15}N,^{1}H Heteronuclear Single Quantum Coherence Experiment (HSQC)

The ^{15}N- HSQC is a mainstay of protein solution studies. Within a matter of hours or less (depending on protein concentration and behaviour) one can record, analyse and draw a number of conclusions from a HSQC.

The HSQC depends on sensitivity enhancement using an Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) pulse sequence which pairs sensitive nuclei (such as protons) with insensitive nuclei such as ^{13}C or ^{15}N. The ratio of gyromagnetic ratios between the sensitive and insensitive nuclei partly determines the signal enhancement factor by transferring magnetisation from the sensitive nucleus to the insensitive one via scalar coupling. The basic architecture of an HSQC pulse sequence is an INEPT pulse followed by chemical shift evolution during t_1 and a reverse-INEPT thereafter, ending with acquisition of the proton FID whilst decoupling of the indirect nuclei (^{15}N) acquisition via a decoupling pulse (Bodenhausen & Ruben, 1980).

^{15}N-HSQC Data: What Can it Tell us?

A ^{15}N-HSQC experiment can provide much information regarding the protein(s) of interest in solution. The basic output of a ^{15}N-HSQC is a spectrum with two dimensions, ^{1}H (direct dimension) and ^{15}N (indirect dimension). Cross-peaks within the spectrum correspond to amino acid main-chain amide groups, one for each amino acid, as well as the side-chain amide groups of Asn, Gln and Trp side chain groups. Proline does not give a cross-peak due to the lack of an amide proton. One can then use this information to assess protein foldedness by inspection of amide

peak dispersion throughout the spectrum. Sharp, poorly dispersed cross-peaks in 8.0-8.5 ppm region of the spectrum likely correspond to amide protons of disordered loops exposed to solvent. Proteins with helical and/or stranded secondary structure character usually show more dispersed cross-peaks. In addition, uniform peak intensities are suggestive of a stably structured protein in long-lived states. Partially unfolded proteins and molten globules give rise to a range of peak intensities across the spectrum. Peak intensities can also hint at dynamic processes, such as conformational changes, monomer:multimer equilibria and aggregative status; comparison of observed and expected peak count (from the equation: No. of amino acids + No. Trp – No. Pro + 2[No. Gln+ No. Asn], where letters represent amino acid three letter codes) can aid such analyses (Kwan *et al*., 2011).

SOFAST-HMQC Experiments for Fast Data Acquisition

In recent years much development has gone into speeding up data acquisition in a protein NMR context (Felli & Brutscher, 2009). The “band-Selective Optimized Flip-Angle Short-Transient”, or SOFAST-HMQC method, was developed to acquire 2D $^{15}N/^{13}C$-^{1}H correlation spectra in shorter lengths of time yet maintain sensitivity and use on standard commercial spectrometers (Schanda & Brutscher, 2005). The SOFAST-HMQC experiment works by using a polychromatic pulse shape (PC9) for band-selective excitation of amide protons, leading to faster ^{1}H spin-lattice relaxation times, allowing for shorter recycle delay periods (Schanda *et al*., 2005).

TROSY-HSQC for Large Proteins

Use of a Transverse Relaxation Optimized Spectroscopy (TROSY) pulse allows the size limit of protein NMR to be extended and is generally recommended for polypeptides larger than 25 kDa. TROSY makes use of the different relaxation mechanisms present in an NMR sample during an experiment, exploiting it to select for sharp signals, raising the signal to noise ratio. (Pervushin *et al*., 1997).

Three Dimensional NMR Experiment for Backbone Assignment

A number of experiments are available to permit unambiguous assignment of backbone amide (HN) peaks of ^{15}N HSQC protein spectra. Other information derived from these experiments include carbon alpha (CA) and beta (CB), as well as carbonyl (CO) chemical shifts. Assignment of these chemical shifts allows for secondary structure prediction, discussed later. These experiments require ^{13}C labelling.

The most widely used are the standard set of backbone assignment spectra, named after the order of nuclei that magnetization is transferred from, starting with the backbone amide proton HN. These are HNCO, HNCACO, HNCA, HNCOCA, HNCACB and HNCOCACB. These will be discussed in more detail below.

HNCO and HNCACO

The HNCO is the most sensitive standard protein backbone experiment and for this reason is usually the first 3D experiment performed when assessing the feasibility of a protein for backbone assignment (Kay *et al*., 2011). The high sensitivity is due to

the large J-coupling constant between CO (i-1) and N (i). The spectrum yields chemical shift data of the preceding (i-1) carbonyl C to the originally magnetized HN (Figure 1.12 A).

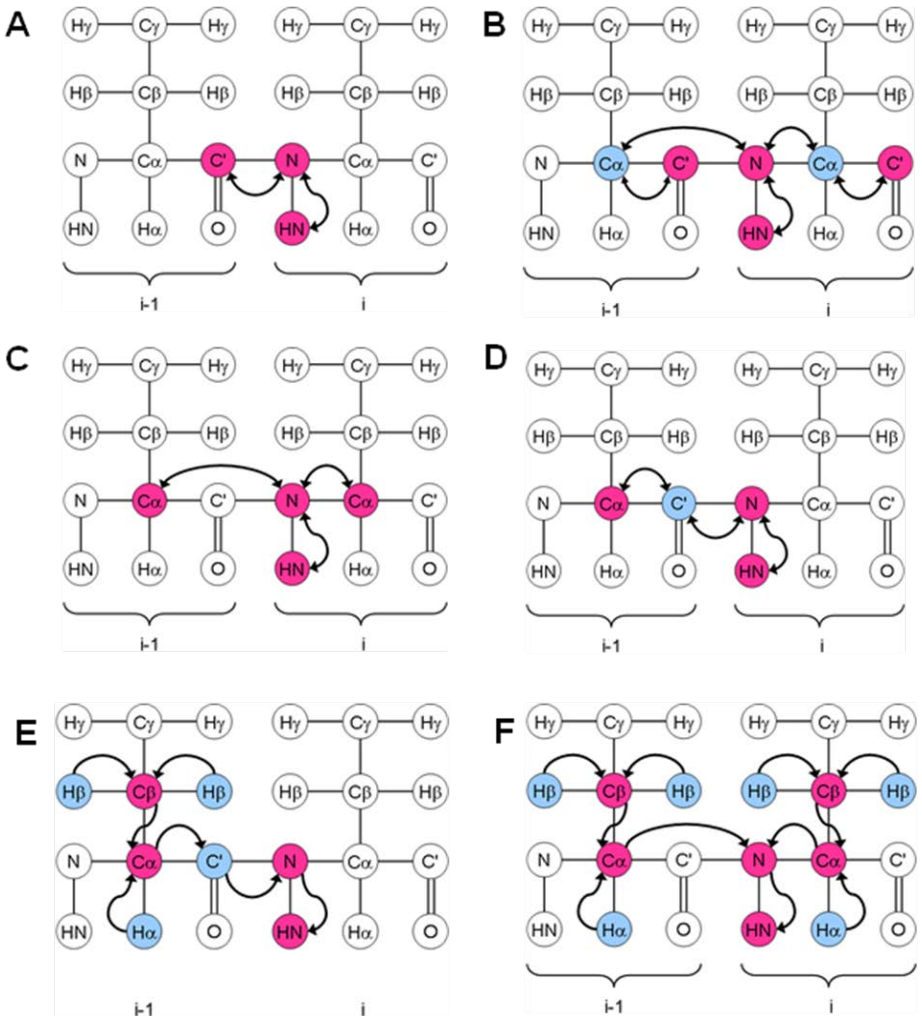


Figure 1.12: Overview of 3D Backbone Assignment Magnetisation Transfer Pathways. Schematic representation of magnetisation transfer pathways of standard 3D protein backbone assignment spectra. Two residues are depicted with atom names in circles. (i) denotes a residue whilst (i-1) denotes the preceding residue to (i). Circles highlighted pink denote atoms where magnetisation is transferred and data indirectly collected, whereas circles highlighted blue denote atoms where magnetisation is transferred but data is not collected. **(A)** HNCO **(B)** HN(CA)CO **(C)** HNCA **(D)** HN(CO)CA **(E)** (HB)CB(HA)CA(CO)NH and **(F)** (HB)CB(HA)CANH. Note that magnetisation begins at atom HN for (A-D) but at the Hα and Hβ for (E-F), depending on whether the experiments are out-and-back or not. Figure taken from protein-nmr.org.uk, Higman V.

The HN(CA)CO, in addition to providing the same information as the HNCO, also provides chemical shift data of the carbonyl C of the originally magnetized HN (Figure 1.12 B). However, due to the much smaller J-coupling between the ^{15}N and $^{13}C\alpha$, sensitivity is markedly reduced.

Taken together, the HNCO and HN(CA)CO can be used to sequentially assign the backbone HN of the protein(s) in question. As Asn and Gln side-chains include a terminal CONH group, chemical shift information of these regions are usually present in HNCO spectra. These correspond to the carbon gamma (Cg) of Asn and carbon delta (Cd) of Gln and can be assigned too, along with their corresponding H-delta/epsilon (HD and HE) chemical shifts. As a ^{15}N HSQC contains a pair of side-chain amide peaks for each Asn/Gln, assignment of both is possible, though this will be ambiguous.

HNCA and HNCOCA

The HNCA experiment provides chemical shift information between HN and the CA of (i) and (i-1) (Bax & Ikura 1991) (Figure 1.12 C).. This is presented as two peaks on a HNCA spectrum, with (i) usually displaying higher intensity than (i-1).

The HN(CO)CA experiment, much like the HNCO, provides chemical shift information between HN and the CA of (i-1) (Figure 1.12 D). When the HN(CO)CA and HNCA spectra are superimposed, unambiguous determination of HNCA (i) and (i-1) peaks are possible. This pair of experiments allows for CA-based backbone assignment.

HNCACB and CBCACONH

The HNCACB experiment provides chemical shift information between HN and the CA and CB of (i) and (i-1). A spectrum will usually contain four peaks: two positive peaks for the CA (i) and (i-1) and two negative peaks for the CB (i) and (i-1) (Figure 1.12 E). The more intense positive and negative peak are likely to be (i), and vice versa for (i-1). The exceptions to this rule are where either (i) or (i-1) are Gly, in which case only 3 three peaks are expected since Gly does not have a CB. In addition, the N-terminal (i) residue will not have an (i-1) peak and thus should only give a maximum of two peaks (one if Gly). However, in most cases the N-terminal residue is unobserved in spectra due to motional dynamics and exchange processes causing broadening of signals.

The HN(CO)CACB experiment provides chemical shift information between HN and the (i-1) CB (Figure 1.12 F) (Grzesiek & Bax, 1992). Depending on how the experiment is set up, Gly residues which do not contain a CB, will be displayed as (often very intense) peaks of opposite phase to other CA peaks, and thus very easy to identify. Superimposition of HN(CO)CACB with HNCACB allows for unambiguous determination of (i-1) CB peaks within the latter spectrum. In addition, both HNCACB and HN(CO)CACB (but particularly the latter) spectra can also provide chemical shift information on side-chain Asn/Gln carbons.

The HNCACB and HN(CO)CACB spectra are the most information rich but may not be enough for unambiguous backbone assignment. For example, a suspected set of weak peaks in the HNCACB can be validated by the HNCA and HNCOCA experiments which have greater sensitivity and better resolution of the Cα. For this

reason, recording and analysing all six backbone experiments mentioned in this section allows for ease of chemical shift assignment by complimenting one another. If most of these spectra, when superimposed, corroborate one another, there is an increased likelihood of present peaks being real.

Whilst the HNCACB and HNCOCACB experiments are “out-and-back” experiments meaning magnetisation begins at the main-chain amide, is transferred all the way to CB and brought all the way back to HN for detection, versions of the experiments exist which begin magnetisation at the HB/HA protons and transfer magnetisation to HN for detection. These are dubbed (HB)CB(HA)CACONH and (HB)CB(HA)CANH and are useful for smaller proteins; larger proteins may be deuterated, preventing magnetisation of HB since they are replaced with non-exchangeable deuterons.

Fast Acquisition of Backbone Assignment Data

Recent advances in multidimensional pulse sequences for protein NMR have permitted collection of entire backbone datasets in > 3 days, instead of 2 weeks when utilising standard pulse sequences. This is achieved by use of a Band-selective Excitation Short Transient pulse, also known as BEST (Lescop *et al*., 2007; Schanda *et al*., 2006) and uses a shaped pulse which reduces the non-specific excitation of water and aliphatic ^{1}H resonances, as well as reduction of signal loss via heterogeneous B_0 and transverse relaxation using broadband ^{1}H inversion pulses. These measures result in dramatic reduction of the recycle delay period of a scan from 1s to ~0.3s. By combining this regime with standard backbone assignment pulse sequences data can be quickly recorded. .

Small-angle X-ray Scattering

SAXS (Small Angle X-ray Scattering) is a structural technique that has undergone significant improvements in the last 20 years. It does not require any prior labelling or tag-inclusion and typically yields data for molecules from the nanometer to sub-micrometer range (Schnablegger & Singh, 2011). Data collection is rapid, ranging from seconds to milliseconds depending on the X-ray source. Whilst laboratory scale SAXS instrumentation exists, higher intensity beams can be taken advantage of at synchrotron sources, providing higher quality data and requiring lower protein concentrations.

Almost all matter large enough will cause X-rays to scatter. There are therefore a number of important steps the SAXS experimenter must take to ensure high quality data which do not contain scattering information from anything but the matter of interest. Specifically, good sample preparation and characterization beforehand will increase the likelihood of quality data. One of the most vital steps is correct solvent matching. In order for a meaningful dataset representing the X-ray scattering of only the particles of interest in the solution, the scattering contribution from the solvent (as well as the sample holder, which may be a tube or cuvette) must be subtracted. This is why solvent blanks are run in between every experimental sample. Samples must be dialysed in the same physical buffer being used for solvent matching, though buffer used in the last purification step, such as size exclusion chromatography (SEC), is generally acceptable.

Samples appropriate for SAXS analysis should be monodisperse, as demonstrated by native polyacrylamide gel electrophoresis (PAGE) and solution-based biophysical

methods such as analytical ultracentrifugation (AUC), dynamic light scattering (DLS), analytical SEC and/or NMR spectroscopy. Samples must also be of high purity, typically above 95% and have no higher molecular weight contaminants which could affect the data. Hence these assays were performed on all the SHP2 constructs studies here.

It is recommended to test samples for aggregation across a range of concentrations.. Soluble aggregates in samples can be minimized by performing SEC purification or high-speed centrifugation immediately prior to SAXS data collection.

SAXS has found use in the physical and material sciences as a technique capable of acquiring size and shape information of matter in almost all forms. Since the 1950s, SAXS has been also been used to probe the structure of biological material at low resolution. Biological SAXS (often abbreviated to BioSAXS) is usually applied to the study of various biomolecules, commonly proteins, DNA/RNA and complexes thereof.

Theory

The spatial resolution of scattered X-rays is dependent on the wavelength of electromagnetic radiation used to study the material of interest. In SAXS, X-rays are utilised; primary (or incident) beam wavelengths between 0.1-0.2 nm are normal and refer to the unscattered, collimated and monochromatic beam needed for SAXS. The dependent variable measured in a SAXS experiment is the intensity (arbitrary units, AU) across the scattering vector q, which is in this study defined as:

$q = (4\pi \sin\theta)/\lambda$

2θ is the scattering angle (<3°) and λ is the wavelength of radiation. q usually ranges from 0.04-6 nm^{-1}, where the lower limit is dependent on the beamstop.

Data collection

The primary beam is passed through a molecule-containing sample causing a fraction of the X-rays to scatter due to interaction with any matter in its path. The scattered X-rays eventually hit a detector, where it is converted into a digital signal and processed further. Any unscattered radiation are absorbed by a beamstop to prevent damage to the detector. This is summarised in Figure 1.13. Since the fraction of scattered X-rays is usually very small in a SAXS experiment, intense X-rays are usually required and provided by synchrotron sources, though in-house sources are available. SAXS experiments are performed in the solution state and data can be acquired in a variety of different buffers, usually containing glycerol and/or a free-radical scavenging reducing agent such as dithiothreitol (DTT) or tris-(2-carboxyethyl) phosphine (TCEP).

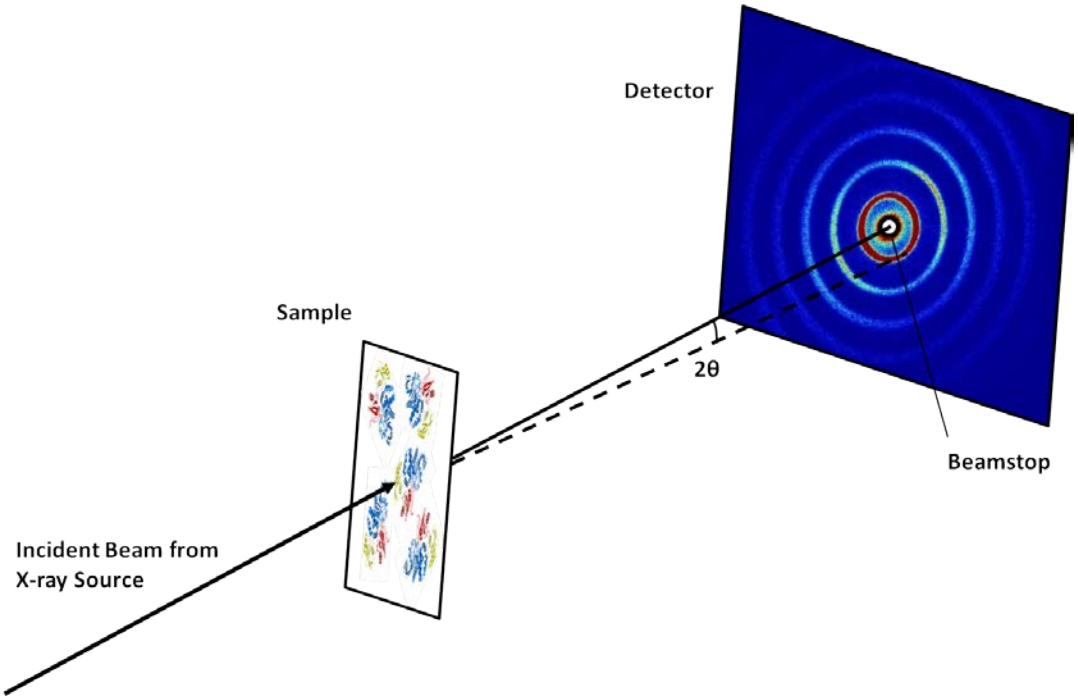


Figure 1.13: Schematic representation of a typical SAXS experiment.

The incident beam (black arrow) passes through the protein-containing sample. A portion of the incident beam scatters (dashed black line) and hits the detector, registering a signal. The remainder of unscattered incident beam hits the beamstop. Detector image adapted from http://www.rigaku.com/en/products/saxs/smax3000/app012 and depicts SAXS pattern of silver behenate.

Post-Data Collection

Radial averaging of the raw data is performed almost immediately after in a step often referred to as data reduction. This increases the signal:noise ratio. Normally 10 frames are collected per sample (for both buffer and macromolecule samples) and inspected automatically by software to assess radiation damage state. After merging of good quality frames, the two buffer-only datasets are then averaged to further increase the signal to noise ratio. After, in a process called buffer subtraction, the scattering data of the buffer is subtracted from the macromolecule-containing sample, yielding a dataset representative of only the macromolecules of interest in solution.

Initial Inspection of SAXS Data

Aggregation can be observed in solvent-subtracted SAXS data as steep slopes in the low-q area of the Guinier plot (explained in below). Using a concentration series one can ameliorate the data by extrapolating it to zero concentration, also known as infinite dilution. Concentration series can also be utilised to rule out interparticle interference of scattering particles, of which there are repulsive and attractive forces, again observable at low-q regions of the Guiner plot. Data indicative of concentration effects can be removed from the low-q region but care must be taken to avoid removing valuable points required for invariant parameter derivation.

Determination of Invariant Parameters

Basic information tractable from SAXS experiments are the radius of gyration (R_g), maximum particle dimension (D_{max}) and intensity at zero angle (I(0)). André Guinier pioneered analysis of SAXS data to determine the R_g to reasonable accuracy by plotting of the data with x-axis as the square of the magnitude of momentum transfer (q^2) and y-axis as the natural logarithm of I(q). This came to be known as the Guinier Law and results in an estimation of the second moment of inertia, or, as put by Konarev P (EMBO SAXS Lecture Series), the "quadratic mean of distances to the centre of mass weighted by the contrast of electron density".

Estimation of the molecular mass (MM) can also be achieved through use of protein standards. SAXS data collection of a protein of known MM such as bovine serum albumin (BSA, ~66 kDa) and subsequent I(0) estimation. Thereafter, MM_{sample} can be determined via the following equation:

$$MM_{sample} = I(0)_{sample} * (MM_{reference}/I(0)_{reference})$$

MM estimation in this manner has an accuracy of around +/- 10% (Mylonas & Svergun, 2007).

Fourier transform of the experimental SAXS profile generates a pair distance distribution function (PDDF). This describes the likely frequency of vector lengths between pairs of atoms within a studied macromolecule and can provide information on shape and excluded volume. For multidomain proteins the PDDF, which is also known as a P(r) function, can be used to investigate subtle conformational changes

upon a change in local environment, such as during a binding event. Changes in domain orientation, and even the spatial distribution of a few amino acid residues, can alter the PDDF and hint toward the nature of structural changes occurring. AUTOGNOM is commonly used to calculate a PDDF for subsequent analysis, including as input into *ab initio* structure determination programs such as DAMMIN/DAMMIF and GASBOR. The online data-driven homology modelling program SAXSTER has an option to submit SAXS data as a PDDF.

Data Analysis

PRIMUS

PRIMUS is an interactive suite of programs and modules for the processing and manipulation of SAXS data (Konarev *et al*., 2003). From here data can be visualised, inspected, subtracted, averaged, merged and extrapolation to zero concentration.

AUTORG and AUTOGNOM

AUTORG and AUTOGNOM are programs developed to approximate invariant parameters such as R_g and D_{max} respectively (Petoukhov *et al*., 2007). Both programs can be accessed via PRIMUS. AUTORG estimates R_g using the Guinier Law, which relates particle size and intensity of scattered radiation (Koch *et al*., 2003). AUTOGNOM is an automated version of the program GNOM which calculates

the PDDF via indirect Fourier transformation of the data. This effectively converts the reciprocal space data into real space. Data processed in this manner can be used as input for *ab initio* shape reconstruction and data-driven modelling.

SUPCOMB

The program SUPCOMB allows one to superimpose (by automatic fitting) and quantify the similarity between atomic and/or bead models (Kozin & Svergun, 2001). SUPCOMB minimizes input structures into an ensemble of points and attempts to find the superimposition that yields the lowest Normalized Spatial Discrepancy (NSD) value. An NSD close to 0 indicates two well superimposed models, above 1 suggests models that consistently differ from one another.

DAMMIF

Ab initio shape reconstruction of SAXS data is achieved using the program DAMMIF and produces a 3D model based on dummy atom beads (Franke & Svergun, 2009b). DAMMIF works by filling a search volume with densely packed space-filling spheres and computing the theoretical scattering amplitude, after which point a simulated annealing algorithm is implemented to generate a model that best fits the data.

Usually 10-20 models are generated with DAMMIF and further manipulated with the DAMAVER suite of programs in the following order: damsel, damsup, damaver and damfilt. Damsel compares all the input models, chooses the most likely one and provides NSD values for all others. Models with high NSD values were designated as

outliers. Damsup then superimposed all models to the reference (i.e. most likely) model and damaver aligns all models. Damfilt then filters the averaged model, removing outlier beads to yield a final envelope.

CRYSOL

Existing atomic models, experimental, comparative or otherwise, can be converted into simulated SAXS data using the program CRYSOL (Svergun *et al*., 1995). CRYSOL works by back-calculating a theoretical SAXS intensity curve and also provides an option to fit an experimental data set against it, providing a means to compare atomic models. χ^2 values are usually given as output, allowing quantification of the goodness of fit between theoretical and experimental datasets.

BUNCH

SAXS data can be used to assist rigid-body modelling of multidomain proteins. BUNCH is a program capable of using multiple experimental and CRYSOL-derived theoretical datasets to model whole or part of a protein utilising a simulated annealing algorithm (Petoukhov & Svergun, 2005). BUNCH is also capable of modelling missing linkers and loop regions of high-resolution structures. Amino acid residues are modelled as dummy residues centered on the approximate Cα position and commonly visualised as spheres.

BUNCH works by global minimization of a target function. Penalties to the score are introduced when steric clashes occur. Additional restraints from high-resolution data also assist in reducing the discrepancy between the model and experimental data.

SAXSTER

SAXS-assisted MUSTER fold-recognition, or SAXSTER, is a web server based program developed for the template-based modelling of proteins using SAXS data as a restraint. Data can be provided in real (as PDDF) or reciprocal space, along with the protein sequence (i.e. query) of the protein construct, the latter passed through MUSTER (MUlti-Sources ThreadER), a protein structure prediction algorithm, to find structures deposited in the PDB with sequence similarity to the query sequence (Wu & Zhang 2008). Highest scoring hits are then used to generate homology models via MODELLER (Eswar *et al*., 2007). Model theoretical SAXS data are computed then compared to experimental SAXS data and ranked.

EOM

The Ensemble Optimization Method (EOM) was developed to better characterize intrinsically disordered proteins (IDPs) and multidomain proteins with flexible linkers in solution by SAXS. It is composed of two programs run consecutively, the first of which is RanCh (Random Chains) which generates a large number of random (usually 10,000) models based on the number of domains, primary protein sequence and regions of flexibility introduced within the protein of interest. These random

models are collectively referred to as the pool. Thereafter a Genetic Algorithm Judging Optimization of Ensembles (GAJOE) is performed, selecting models from the RanCH-generated pool that best fit the experimental SAXS data. The averaged theoretical scattering profiles of the selected ensemble should, in theory, fit the data better than a single profile. R_g and D_{max} values for the pool and ensemble structures are calculated and plotted as frequency diagrams, allowing visual inspection and comparison of the ensemble to the pool. If the frequency plots of both pool and ensemble are similar, this suggests the ensemble samples a broad range of conformations and is either disordered (for IDPs) or very mobile for multidomain proteins. The proteins can also be more compact or extended on average, if the frequency plots for the ensemble lies to the left or right of the pool frequency, respectively (Bernadó *et al*., 2007).

Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC) is a biophysical technique capable of providing information relating to the size distribution and shape of particles in solution. It is performed in an analytical ultracentrifuge fitted with specialised optics adapted to the type of experiment: boundary sedimentation velocity (SV) measures the rate of sedimentation whereas sedimentation equilibrium (SE) measures the concentration distribution achieved when diffusion and sedimentation are balanced. The SV experiment will be discussed further below.

SV experiments are used to measure sedimentation coefficients expressed in Svedbergs (S). The sedimentation of a particle over time whilst under the influence of

a strong centrifugal force is detected by optical techniques: absorbance in the UV-visible domain, fluorescence or interferometry. Many factors affect S, including the shape and mass of the particles in solution. The technique is used in structural biology to assess the formation of complexes, their association constants or the solution state of particles.

Analysis of AUC data is performed using the software package Sedfit (Schuck, 2000). The Lamm equation used to calculate S from the radial concentration distribution. The analysis of AUC data describes the particle and solvent as it is experiencing centrifugal force (Laue & Stafford, 1999).

Circular Dichroism Spectroscopy

The secondary structure content of proteins can be reliably quantified by the collection and analysis of data obtained from circular dichroism spectroscopy, otherwise known as CD. The technique relies on passing circularly polarised light (left and right circularly polarised light are in equal proportions) at wavelengths between 190-260 nm through an aqueous sample containing the molecules of interest. Free passage of left and right forms of light through the sample recombine and result in plane polarised light detected by a photomultiplier. However, the presence of optically active molecules and secondary structural elements leads to the non-equal absorption of left and right circularly polarised light, producing elliptically polarised light. Information regarding the chirality or secondary content of the molecules (Kelly *et al*., 2005) can then be assessed by analysis of the resultant absorption spectrum.

It is normal practice to collect a control sample containing the exact buffer the molecules of interest are dissolved in. For high quality CD data, buffers should contain as few chemical components as possible. Chloride ion (Cl^-) concentration should either be minimal or omitted completely as they strongly absorb at low UV wavelengths (Kelly & Price, 2000). After collecting CD data of the molecules of interest dissolved in buffer, a step known as solvent subtraction is performed where the solvent contribution to the CD spectrum is removed, leaving only information concerning the secondary structure of your molecules of interest. CD spectra are normally presented as molar ellipticity (θ) in millidegrees (mdeg) as a function of wavelength. CD data can be analysed using a variety of available stand-alone software and online web servers; for the present study DICHROWEB (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) was utilised (Whitmore & Wallace, 2004).

Aims of Study

The aims of this study are to overexpress, purify and investigate various Shp2 constructs (i.e. the individual and Tandem SH2 and PTP domains) by solution NMR and SAXS. In addition, ligand interactions of Shp2 with the novel immunoreceptor G6b-B doubly phosphorylated ITIM will be studied using NMR/SAXS and other biophysical methods in solution. The Noonan syndrome/leukaemia mutant (E76K) full length Shp2 as well as the wild-type, as yet uncharacterized structurally, will be studied by SAXS and molecular modelling.

CHAPTER II
MATERIALS AND METHODS

Full-length Shp2

The gene encoding full length Shp2 (*PTPN11*) was codon-optimized for heterologous expression in *E*.coli, synthesised and inserted into the pUC57 vector by *ECORV* blunt ended ligation (Genscript USA Inc., see Appendix 1 for gene sequence).

The *PTPN11* gene was then subcloned by ligation independent cloning (LIC) into a custom-made bacterial expression vector (kindly provided by Dr. Stefan Knapp, Structural Genomics Consortium, Oxford) named pNIC28-BSa4, containing an N-terminal histidine tag sequence and Tobacco Etch Virus (TEV) protease cleavage site (MHHHHHHSSGVDLGTENLYFQS). The pNIC28-Bsa4 vector allows efficient and fast generation of expression constructs. First the vector is linearized and treated with T4 DNA polymerase in the presence of dinucleotide guanine triphosphate (dGTP) to generate 10-15 basepair overhangs. Next specially designed primers with complementary bases to the T4-treated vector are used to amplify your gene of interest. Incubation of both vector and amplified insert allows base-pairing to occur, producing an intact construct ready for bacterial transformation (Savitsky *et al.*, 2010).

An amplification step was carried out using a thermal cycler to produce a DNA fragment with the additional complementary bases required at the 5’ and 3’ ends to mediate ligation.

Forward primer: Full length Shp2:

5’-TACTTCCAATCCATGACCAGCCGTCG-3’

Reverse primer: Full length Shp2:

5’- TATCCACCTTTACTGTTAACGAAAAGATTTCTGC-3’

N.B: The underlined bases indicate where ligation occurs

The KOD polymerase (Merck Chemicals)was used according to standard protocol. Thermal cycler steps were as follows: 1) 94°C for 2 min, 2) 94°C for 15 sec, 3) 44°C for 30 sec, 4) 68°C for 1 min and 5) 68°C for 2 min. Repeat steps 2-4 20 times. Agarose gel electrophoresis (AGE) confirmed presence of a band in correct size range (Between 1500-2000 bp; Full length Shp2 insert ~1800bp). Briefly, samples were run in a 1% agarose gel in Tris-base/acetic acid/EDTA buffer containing 0.5 µg/ml ethidium bromide for 30 min at 100 V, then visualised with an ultraviolet transilluminator. Purification of polymerase chain reaction (PCR) amplification products was carried out using a PCR purification kit (Invitrogen), followed by treatment of 5 µl of the PCR product with T4 polymerase to generate sticky ends. AGEof the full length Shp2 insert and pNIC28-Bsa4 vector was performed to estimate concentration based on a marker (DNA Hyperladder; Bioline). Insert and vector concentration were determined to be 5-6 ng/µl. 1 µl insert was added to 2 µl vector and incubated at room temperature for 10 min before using 2 µl for a bacterial transformation (DH5α, standard protocol) and plated out onto 30 µg/ml kanamycin LB agar plates followed by incubation at 37°C overnight. 7 colonies appeared, of which 2 were used for plasmid amplification/purification (Qiagen Miniprep kit) and sequence confirmation (Functional Genomics, School of Biosciences, University of

Birmingham). (Additional primer required for sequencing: 5'-GATCTGGTGGAACACTATAA-3').

Tandem SH2 domain of Shp2

The LIC subcloning method was also applied to produce a construct containing both the N and C-SH2 domains of Shp2 (residues 4-216 of wild-type) as well as an N-terminal Histidine tag followed by a TEV protease cleavage site.

Primers used to amplify Tandem SH2 of Shp2 DNA:

Tandem SH2 Forward: 5'- TACTTCCAATCCATGCGTCGCTGGTTTCAT-3'

Tandem SH2 Reverse: 5'- TATCCACCTTTACTGTTACAGCGGCTGTTTCA-3'

Tandem SH2 of Shp1 and nSH2 and cSH2s of Shp1 and Shp2

The modified pET28a vectors containing individual nSH2(residues 1-104) and cSH2 (residues 101-222) domains of Shp-1 and Shp2 were a kind gift from Tony Pawson (University of Toronto). The pET28a vector is under T7 promoter control and is thus compatible with lacO-based expression systems. Both constructs feature an N-terminal Histidine tag followed by a TEV protease cleavage site and a C-terminal Leu-Ile-Asn-Glu-Phe (also referred to as LINEF) pentapeptide sequence.

A summary of constructs used in this study can be found in Table 1.2. For a graphical schematic of Table 1.2 see Appendix 4.

Table 1.2. Summary of constructs used.

Construct Name	Vector	Antibiotic Resistance	Affinity Tag	Cleavage Tag
WT FL-Shp2	pNIC28-Bsa4	50 µg/ml Kanamycin	Hexahistidine	TEV Protease
TanSH2-Shp2	pNIC28-Bsa4	50 µg/ml Kanamycin	Hexahistidine	TEV Protease
nSH2-Shp2	pET28a	50 µg/ml Kanamycin	Hexahistidine	TEV Protease
cSH2-Shp2	pET28a	50 µg/ml Kanamycin	Hexahistidine	TEV Protease
nSH2-Shp1	pET28a	50 µg/ml Kanamycin	Hexahistidine	TEV Protease
cSH2-Shp1	pET28a	50 µg/ml Kanamycin	Hexahistidine	TEV Protease
TanSH2-Shp1	pET28a	50 µg/ml Kanamycin	Hexahistidine	TEV Protease

Expression trials

Expression trials were conducted to optimize the yield of protein produced after overexpression in *E. coli*. Three lots of 10-20 ml LB media containing kanamycin at a final concentration of 30 µg/ml in 50 ml falcon tubes were inoculated with single colonies grown on LB-agar-kanamycin plates. Cultures were grown at 37°C and 220 revolutions per minute (rpm) in an Ecotron shaker incubator (Infors-HT). Optical density at 600 nm (OD_{600}) values were measured every hour until values between 0.4-0.6 were reached, at which point the cultures were transferred to incubators set at different temperatures: 18°C, 25°C and 37°C. Cultures were left to acclimatise for 15 min, after which another OD_{600} measurement was recorded. 1 ml of each culture were then removed and transferred to a 1.5 ml microfuge tube. These pre-induction samples were centrifuged at 20000 xg for 1 min, the pellet stored at -20°C and the supernatant discarded. Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM were added to all the main cultures and incubated at their respective temperatures overnight (16 h) whilst shaking at 220 rpm. The next morning, OD_{600} measurements were recorded for all bacterial growths and 1 ml samples were removed and treated in the same way as pre-induction samples. Thereafter, pre-induction pellets were resuspended in 100 µl Laemmli buffer and kept on ice. Post-induction pellets were resuspended in a similar manner except higher volumes of Laemmli buffer were used in resuspension so as to normalise the cell density between pre- and post-induction samples. Volumes for each TanSH2-Shp2 expression trial sample are summarised in Table 1.3 as a typical example.

Table 1.3. **Values measured and calculated for expression trials of TanSH2-Shp2.**

Sample	OD_{600} at pre-induction	OD_{600} 16 h post-induction	Ratio of pre- and post- induction OD_{600} values	Volume of Laemmli buffer added (µl)
18°C 1	0.976	1.751	1.79	179
18°C 2	0.773	1.656	2.14	214
25°C 1	0.621	1.111	1.79	179
25°C 2	0.731	1.217	1.66	166
37°C 1	0.79	1.543	1.95	195
37°C 2	0.705	1.642	2.33	233

All samples were lysed by sonication for 20 sec at 5 microns amplitude in a Soniprep 150 fitted with an exponential probe (MSE). Samples were then loaded onto a pre-cast sodium dodecyl sulphate-PAGE (SDS-PAGE) gel (Criterion Tris HCl 15% resolving/4% stacking gel BioRad) with a molecular weight standard (Precision Plus Protein™ All Blue Standards, Bio-Rad) and run for 40 min at 180 V (~21 V/cm), after which the gel was stained at room temperature and gentle shaking with 30 ml InstantBlue (Expedeon) for 20 min.

Protein expression

A single colony from an agar plate was picked with a disposable inoculating loop and agitated in a 50 ml Falcon tube containing 5-10 ml LB and kanamycin at a final

concentration of 30 μg/ml and left overnight at 37°C in a shaker incubator running at 225 rpm.

The following morning the 5 ml cultures were transferred into a 2 L flask containing 1L of autoclaved LB liquid media at a final kanamycin concentration of 30 μg/ml.

Cultures were grown at 37 °C in either unbaffled flasks at 180 rpm or baffled flasks at 150 rpm until an OD_{600} between 0.4-0.6 was reached, at which point the temperature was dropped to 18°C and cultures induced with the addition of IPTG to a final concentration of 1 mM. Cultures were left overnight for approximately 16 h.

Labelled expression

In order to produce isotopically labelled recombinant proteins for solution NMR studies, use of minimal M9 media with NMR-active isotopes (given as $^{15}NH_4Cl$ for ^{15}N label incorporation, ^{13}C-glucose for ^{13}C incorporation and deuterium oxide, otherwise known as D_2O, for deuteron incorporation). A standard M9 recipe was followed and the same bacterial cell lines and growth protocol were used.

Protein Purification

Buffers used:

- Resuspension buffer: HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, 0.5 mM TCEP and 1 Complete EDTA-free protease inhibitor cocktail (Roche)

- Equilibration buffer: 50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, 0.5 mM TCEP
- Wash buffer: 50 mM HEPES pH 7.5, 500 mM NaCl, 25mM imidazole, 0.5 mM TCEP
- Elution buffer: : 50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole, 0.5 mM TCEP
- Dialysis buffer: 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP
- SEC buffer: 50 mM HEPES pH 7.5, 300 mM NaCl 0.5 mM TCEP
- NMR buffer: 50 mM sodium phosphate pH 7.0, 50 mM NaCl, 0.5 mM TCEP, 0.01% NaN_3.

Cells were transferred to 1 L centrifuge pots and spun at 6000 xg for 20 min at 4 °C. After media were decanted, cell pellets were resuspended in 25 ml/l equilibration buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid [HEPES] pH 7.5, 500 mM NaCl, 5 mM imidazole, 0.5 mM TCEP and 1 Complete ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail (Roche)). Cell suspension was then homogenised and lysed with three passes using Emulsiflex. Lysate was centrifuged immediately afterward at 75000 xg for 45 min at 4°C. Supernatant decanted and passed through 0.45 µm filter (Sartorius) before loading onto a 5 ml Nickel-NTA affinity column (HisTrap, GE Healthcare) pre-equilibrated with five column volumes of equilibration buffer free from protease inhibitors. Supernatant was loaded at 1.5 ml/min and the flow-through collected for SDS-PAGE analysis. 7

column volumes of wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 25 mM imidazole, 0.5 mM TCEP) were then passed through at 2.5 ml/min, with the flow-through collected. Thereafter, elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole, 0.5 mM TCEP) was passed through at 2 ml/min with 1.5 ml fractions collected in microfuge tubes and qualitatively tested for protein concentration using Bradford reagent (Bio-Rad) until no protein eluted. Fractions were prepared for SDS-PAGE analysis by mixing equal parts sample and 2x Laemmli sample buffer prior to being loaded into an 18 or 26 –well SDS-PAGE gel and run and stained in a similar manner as described earlier.

Fractions of the highest purity were then pooled for diafiltration prior to SEC or for tag removal depending on the construct.

His-tag cleavage with TEV protease

500 µl 1.5 mg/ml TEV protease in 50 mM HEPES pH 7.5, 100 mM NaCl and 0.5 mM TCEP in 50% glycerol are added to the pooled fractions, lightly agitated and transferred to Snakeskin dialysis tubing (3 or 10 kDa molecular weight cut-off [MWCO]) and placed in 1 L dialysis buffer overnight at 4°C with mild agitation.

Contents of the dialysis tubing were filtered with a 0.45 µm filter device and syringe, then passed through an equilibrated 5 ml HisTrap Ni-NTA affinity column, with flow-through collected**.** 5 column-volumes of equilibration, wash and elution buffer were sequentially passed through the column, with flow-through collected. SDS-PAGE analysis of flow-through samples immediately followed.

Ion Exchange Chromatography (IEX)

Supernatant was loaded onto a GE 16/60 HiPrep Q XL column pre-equilibrated with two column volumes of 50 mM HEPES pH7.5, 50 mM NaCl, 0.5 mM TCEP and EDTA-free protease inhibitor cocktail (Roche). Flow-through was collected and kept on ice. Two column volumes of elution buffer (50 mM HEPES pH 7.5, 1M NaCl, 0.5 mM TCEP, EDTA-free protease inhibitor cocktail) was applied to the column, flow-through collected and stored on ice.

Size Exclusion Chromatography (SEC)

Protein-containing fractions were concentrated in Amicon concentrators (3 or 10KDa MWCO) to <5 ml and centrifuged at 17,000 g for 10 min. The concentrated sample was then taken up into a Hamilton syringe pre-washed with SEC buffer and injected into the 5 ml sample loop of an AKTA Purifier. A Superdex 75 prep grade 26/60 gel filtration column was equilibrated with 1.2 column volumes of SEC buffer and injected with sample, followed by 1 column volume elution collected in 4 ml fractions. For FL-Shp2 WT and E76K the SEC was performed at 4°C. For TanSH2, nSH2 and cSH2 domains this was performed at room temperature.

Peptides

ITIM1, ITIM1_s, ITIM2 and dITIM peptides were synthesised (Fmoc method) by Genscript (U.S) or Alta Bioscience (Birmingham UK). Properties of the peptides are briefly summarised in Table 1.4 (See Appendix 2 for full G6b-B amino acid sequence).

Peptide Name	Number of residues	Amino acid sequence	Modifications
ITIM1	12	EPSLL**Y**ADLDHL	N-acetylation; C-amidation;
ITIM1_s	10	EPSLL**Y**ADLD	N-acetylation; C-amidation
ITIM2	12	DASTI**Y**AVVV	N-acetylation
dITIM	36	EPSLL**Y**ADLDHLALSRPRRLSTADPADASTI**Y**AVVV	N-acetylation

Table 1.4. **Summary of phosphorylated peptides used. Tyrosine residues with added phosphate groups are highlighted in bold and underlined.**

Mass Spectrometry

Matrix-assisted laser desorption ionisation time of flight (MALDI-TOF) and liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed in collaboration with Dr. Douglas Ward (University of Birmingham, UK) with a ultraflex MALDI-TOF Mass Spectrometer in linear mode (Bruker Daltonics). Samples were

either dialyzed into dilute hydrochloric acid (HCl) overnight prior to MS analysis to remove buffer salts or by use of a zip-tip fitted with a mini desalting medium. Samples were mixed with either sinapinic acid or 2, 6-dihydroxyacetophenone (DHAP) and air-dried on sample plates at room temperature before loading into the mass spectrometer. This was done in collaboration with Dr. D. Ward (School of Cancer Sciences, University of Birmingham)

Phosphatase Activity Assay

The *p*-nitrophenyl phosphate (pNPP) activity assay is a colourimetric technique used to quantify PTP desphosphorylation rates. As pNPP is a non-natural surrogate substrate, not all PTPs show the same magnitude of catalytic activity towards it as they would toward their natural substrates (pY-containing peptides). However, the assay can yield data rapidly and to good quality (Montalibet *et al*. 2005). The pNPP assay is also cheap and straightforward. The assay works by allowing pNPP to be dephosphorylated by the PTP(s) of interest, producing p-nitrophenol, a chromogenic product. Addition of highly alkaline solution (usually between 1-3 M sodium hydroxide) deprotonates the product and in turn produces an intense yellow colour which absorbs optimally at 405 nm. This can be quantified using a spectrophotometer (Lorenz 2011).

Phosphatase activity of Shp2 was determined using previously described protocols (Montalibet *et al*. 2005). Briefly, both PTP-Shp2 and WT and mutant forms of FL-Shp2 were expressed and purified as above and concentrated to act as stock

solutions. Proteins were stored in 25% glycerol at -20 °C when not in use. A 10x assay buffer stock solution was prepared (500 mM HEPES pH 7.5, 5 mM DTT and 20mM EDTA) and filter sterilised. 5 µM assay stock solutions of proteins were diluted to the concentrations required in enzyme dilution buffer (1x assay buffer, 20% glycerol, 0.1% Triton X-100) to make a working buffer of 50 mM HEPES pH 7.5, 2mM EDTA, 0.5 mM DTT, 20% glycerol and 0.1% Triton X-100. *p*-nitrophenol phosphate (PNPP, Sigma Aldrich) was dissolved in enzyme dilution buffer to a stock concentration of 150 mM, filter sterilised (0.45 µm cut-off) and aliquoted into 1 ml fractions in 1.5 ml microfuge tubes and stored at -20°C when not in use. Assays were performed in 96-well flat bottomed polypropylene microplates (Corning) at room temperature. 10 µl of stock PNPP solution was added and quickly pipetted repeatedly into 90 µl of reaction mix to ensure proper mixing, containing no enzyme (negative control) or enzyme of differing concentrations, in the presence or absence of G6B-b peptides mentioned above. The total reaction volume was 100 µl and was left to run for at least 5 min before being quenched with 25 µl 3M NaOH and analysed using a Bio-Rad iMark Microplate Reader at 415 nm single wavelength. 'No enzyme'-containing wells were used as buffer subtraction blanks post data collection.

AUC

Sedimentation velocity experiments were performed in collaboration with Rosemary Parslow (School of Biosciences, University of Birmingham, UK) using a Beckman XL-I Ultracentrifuge. Free and dITIM-bound TanSH2-Shp2 were prepared *via* SEC and dialyzed into 50 mM sodium phosphate buffer, pH 7.0, 50 mM NaCl and 0.5 mM

TCEP. Two concentrations were made of each sample (500 µl total volume) so as to give an absorbance reading (A_{280}) of 0.15 and 0.5 AU.

CD

Anhydrous dITIM peptide was dissolved in 50 mM sodium phosphate (pH 7.4) buffer to final concentrations of 1 and 10 mg/ml and loaded into Starna® 20/C Demountable cells of path lengths 0.1 mm and 0.01 mm respectively. CD data were collected using a Jasco J-810 Spectropolarimeter at 25°C and analysed with DICHROWEB.

NMR

NMR data were collected on various Varian spectrometers at the Henry Wellcome Building for Biomolecular NMR and performed at 298°K unless stated otherwise. VNMRJ was used in all data collection procedures. All data were processed with NMRPipe (Delaglio, Grzesiek *et al*., 1995) and analysed with CCPNMR Analysis versions 1.5-2.2 (Vranken, Boucher *et al*., 2005).

1D ^{1}H-NMR

Acquisition of 1D ^{1}H NMR spectra were performed before higher dimensional experiments to verify the presence of protein in solution. Water suppression was achieved using a presat pulse sequence Optimisation of transmitter offset and 90° pulse calibration was performed. The ^{1}H NMR spectrum was inspected in order to identify possible problems with the experimental set up as well as qualitatively decide whether to carry on with the NMR experiment.

2D ^{15}N-HSQC and SOFAST-HMQC

^{1}H, ^{15}N-HSQC and ^{15}N-SOFAST HMQC experiments (gNfhsqc and sofastNhmqcA in VNMRJ, respectively) from the software Biopack were used. Optimised 90° ^{1}H pulse width (approx 7-11 µs) and transmitter offset (tof) values were copied over from the 1D ^{1}H water experiment. Optimisation of ^{15}N 90° pulse width was performed using the first increment of the 2D experiment and usually fell between 35-40 µs. The 'half dwell time' box was checked and after a final inspection of pulse sequence

parameters, the experiment was initiated. Observation of the probe temperature and program interface for the first minute ensured the experiment was running correctly. HSQCs were performed for all protein constructs apart from E76K FL-Shp2.

Triple resonance assignment data acquisition for nSH2 and cSH2 -Shp2

Acquisition of backbone resonance assignment data for the nSH2 and cSH2 domains of Shp2 were carried out on a Varian DirectDrive 600 MHz spectrometer fitted with z-axis pulse field gradients and a 3-channel cryogenic probe. The samples were prepared as previously stated and dialysed into a buffer containing 50 mM sodium phosphate pH 7.0, 50 mM NaCl, 0.5 mM TCEP and 0.01% NaN_3, made up to a final D_2O concentration of 10% for locking. Final sample concentrations were 1.8 and 2.8 mM for nSH2 and cSH2 Shp2 respectively, at a final volume of 600 µl. Samples were transferred to cleaned NMR tubes via NMR pipettes and lightly centrifuged to migrate all liquid to the bottom of the tube. After an initial $^{1}H,^{15}N$-HSQC to confirm presence of labelled protein, gauge signal intensity and act as a reference spectrum onto which all resonance assignments will be matched to, BEST (Lescop *et al*. 2007)versions of backbone assignment experiments were performed: HNCO, HNCACO, HNCA, HNCOCA, HNCACB and HNCOCACB (Wittekund & Mueller, 1993; Bax & Ikura, 1991); these were obtained from VNMRJ biopack.

Triple resonance assignment data acquisition for TanSH2-Shp2

Acquisition of triple resonance assignment data were performed on a Varian INOVA 800 MHz spectrometer fitted with z-axis pulse field gradients and 4-channel cryoprobe. TROSY versions of HNCO, HNCACO, HNCA, HNCOCA, HNCACB and HNCOCACB experiments were performed after an initial ^{15}N HSQC, using a ~2 mM triple-labelled TanSH2-Shp2 sample with affinity tag left intact at a final volume of 600 µl in 50 mM sodium phosphate buffer pH 7.0, 50 mM NaCl, 0.5 mM TCEP, 0.01% NaN_3 and 10% D_2O. Backbone pulse sequences were obtained from VNMRJ Biopack. Backbone resonance assignments of the individual SH2 domains of Shp2 aided in the assignment of the TanSH2-Shp2 triple resonance data.

NMR Peptide Titrations

NMR peptide titration data were collected on Varian DirectDrive 600 MHz and INOVA 800 MHz spectrometers with ^{15}N-labelled nSH2, cSH2 and TanSH2 domains of Shp2 and cSH2 domains of Shp1 at concentrations ranging from 100-400 µM. All experiments were performed in buffer containing 50 mM sodium phosphate pH 7.0, 50 mM NaCl, 0.5 mM TCEP, 0.01% NaN_3 and 10% D_2O at sample volumes between 550-600 µl in standard NMR tubes. ITIM Peptides were dissolved before each NMR experiment in the same buffer and re-adjusted to pH 7.0. Peptides were added in equal increments up to 2.5-fold excess, ^{15}N-HSQC or SO-FAST HMQC spectra were collected at each ligand concentration. Residue-level interactions can be obtained by calculating chemical shift differences between ^{1}H and ^{15}N using CCPN Analysis software and further statistical analysis in Microsoft Excel 2007.

SAXS

SAXS data were collected on three sites, the X33 Beamline, part of the DORIS III storage ring at DESY (Deutsches Elektronen-Synchrotron), EMBL Hamburg, the P12 beamline on PETRA III, part of the same organisation as the previously mentioned beamline and the BM29 Beamline in ESRF, Grenoble. All measurements were conducted at room temperature using a programmable automated sample changer.

At X33, data were recorded using a single photon counting pixel detector system, 2D Pilatus 1M-W pixel X-ray detector within a momentum transfer range of 0.06 nm^{-1} to 6 nm^{-1} at a distance of 2700 mm and X-ray wavelength of 0.15 nm.

Data collection at P12 utilised a 2D photon counting Pilatus 2M pixel DECTRIS X-ray detector within a momentum transfer range of 0.05^{-1} to 0.35 nm^{-1} at a distance of 3.1 m and X-ray wavelength of 0.124 nm.

At BM29, data were recorded using a single photon counting pixel detector system, PILATUS 1M pixel X-ray detector within a momentum transfer range of 0.01 nm^{-1} to 5 nm^{-1} at a distance of 2841mm and X-ray wavelength of 0.099 nm.

Protein samples were measured between 1 and 10 mg /ml in 50 mM sodium phosphate buffer ranging from pH 7.0-7.4 and 100-200 mM NaCl, with either 1 mM DTT or 0.5 mM TCEP. Automated data processing on-site allowed for fast optimisation of conditions if need be, with further processing and analysis performed using modules from the ATSAS suite of programs (Konarev *et al*., 2006) and Scatter (Classen *et al*., 2013) . R_G and I(0) values for each protein construct or complex were estimated using the AUTORG and AUTOGNOM modules, part of the PRIMUS

program (Konarev *et al*., 2003), with Porod Volume, P(r) distance distribution functions and D_{Max} values estimated using the latter module.

Ab Initio Modelling

DAMMIF and GASBOR were used to produce bead models using only SAXS data. GNOM files generated using AUTOGNOM served as input files for both programs. 10 independent runs of DAMMIF (Franke and Svergun, 2009) in slow mode were performed for each SAXS construct. SUPCOMB (Kozin and Svergun, 2001) was performed on each set of 10 models to align them in 3D space and generate NSD values. DAMSUP, DAMAVER and DAMFILT (Volkov and Svergun, 2003) were then applied sequentially to generate a single, averaged bead model. Models were visualised with PyMol v1.3 (DeLano and Lam, 2005) and Chimera v1.6 (Pettersen *et al*., 2004).

Homology Modelling

Online versions of I-TASSER (Zhang, 2008) and SWISS-MODEL (Arnold *et al*., 2006; Bordoli *et al*., 2009) were utilised to produce homology models of various constructs of Shp1 and Shp2. For both programs, construct protein sequences were uploaded onto the server and a PDB code specified or uploaded as a template for modelling against. In the majority of cases this was carried out in order to model missing atoms and loop regions into already existing crystal structures.

Rigid-body Modelling

SAXS data were used to refine and assist in producing homology models; the programs BUNCH (Petoukhov and Svergun, 2005) and SAXSTER (dos Reis *et al*.,

2011) computed such structures for various constructs used in this study. 10 independent models were generated for each construct and averaged.

Ensemble Optimisation Method

The Ensemble Optimisation Method (EOM) was employed to generate models based on the SAXS data collected in this study and existing atomic coordinates. Unlike rigid-body modelling, EOM allows for flexibility to be accounted, producing models that best represent flexible particles in solution. EOM is split into two distinct parts, RANCH and GAJOE, the former to generate a pool of10,000 random models and their simulated SAXS curves, and the latter to apply a genetic algorithm to select an ensemble of structures from the pool that, when averaged, produce the best-fitting simulated curve when compared to the experimental data (Bernado *et al*., 2007).

CRYSOL

Simulated SAXS curves were generated from calculated high resolution structures, homology models and *ab initio* models were performed using CRYSOL (Svergun *et al*., 1995).

CHAPTER III

CONSTRUCT GENERATION, EXPRESSION AND PURIFICATION

Subcloning

The codon-optimized *PTPN11* gene was synthesised by Genscript and inserted into pUC57. Subcloning TanSH2-Shp2 from the *PTPN11* followed the same protocol as that of subcloning WT FL-Shp2. Sequencing the TanSH2-Shp2 containing plasmid only required two primers (T7 forward and reverse) as the insert was 639 bp in length, permitting full sequence coverage of both 5' and 3' ends of the gene. TanSH2-Shp2 gene insert and vector was digested by the *BamH1* restriction enzyme and analysed by AGE (Figure 3.1), as was FL-PTPN11 (Figure 3.2). FL-PTPN11 required both T7 primers and a specific primer for the middle of the gene.

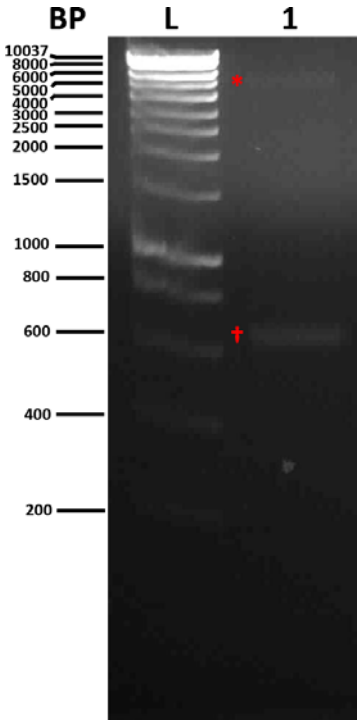


Figure 3.1: DNA AGE Analysis of TanSH2-PTPN11 and Vector.

Analysis of TanSH2-PTPN11 insert and pNIC28-BSa4 vector produced via restriction enzyme digestion by AGE. 1: TanSH2-PTPN11 band is located by †; pNIC28-BSa4 is located by *. Insert is ~660 bp and vector is ~5.2 kbps L: DNA basepair ladder, BP: Basepairs

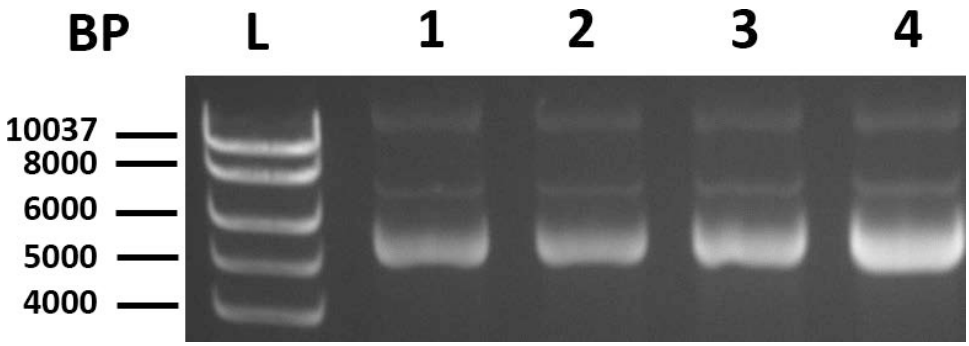


Figure 3.2: DNA AGE Analysis of FL-PTPN11 Plasmid Minipreps.

Analysis of Plasmid Minipreps of four FL-PTPN11 clones by AGE. 1-4: Plasmid clones. Plasmids are estimated to be ~7 kBp. L: DNA basepair ladder, BP: Basepairs. NB: All four plasmids are circular (i.e. not linearized by single digestion).

Expression Trials of WT FL-Shp2 and TanSH2-Shp2

Expression trials were performed for TanSH2-Shp2 in order to maximize the protein yield. These were carried out in BL21 DE3 cells (Agilent). Temperature variation was tested to quickly assess growth conditions for preparation of TanSH2-Shp2.

Expression trials of TanSH2-Shp2 yielded a positive result; the protein was overexpressed in large quantities (Figure 3.3). An overnight incubation temperature of 18°C was selected for TanSH2-Shp2 overexpression.

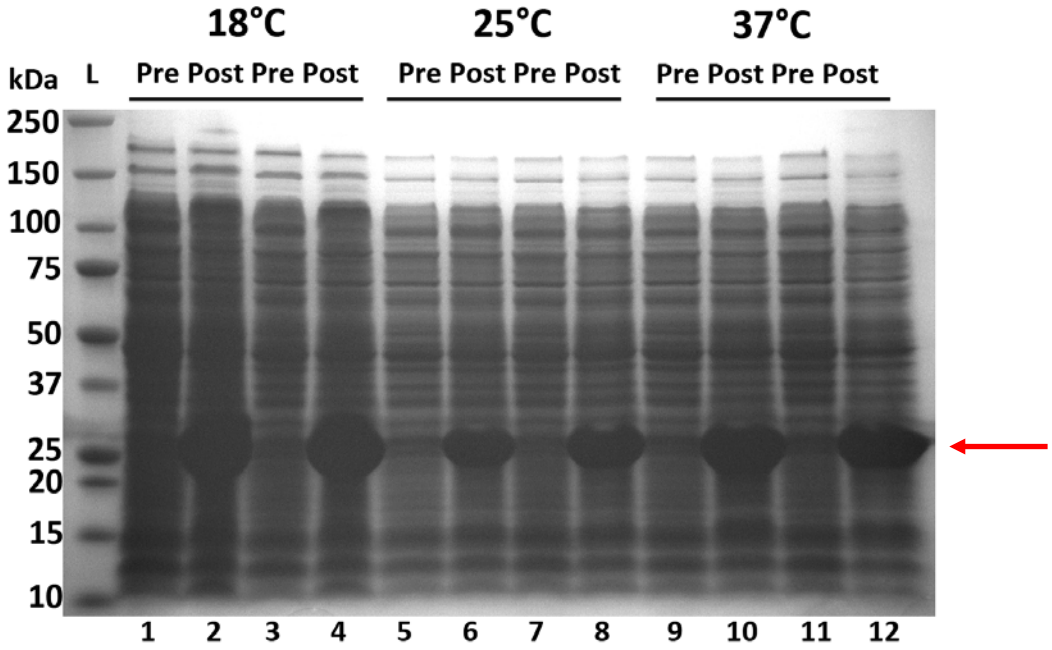


Figure 3.3: Expression trials of TanSH2-Shp2: SDS-PAGE Analysis.

SDS-PAGE analysis of TanSH2-Shp2. The expected mass was estimated by Protparam (Expasy) to be 26.6 kDa. Each induction temperature (18, 25 and 37°C) was tested in duplicate. Odd and even numbers are Pre- and post-induction respectively. kDa: kiloDaltons, L: MW marker, Pre: Pre-induction, Post: Post-induction. Red arrow indicates bands of interest.

Expression trials were also performed for the WT FL-Shp2 construct, indicating that good levels of expression were possible at the three induction temperatures tested (Figure 3.4).

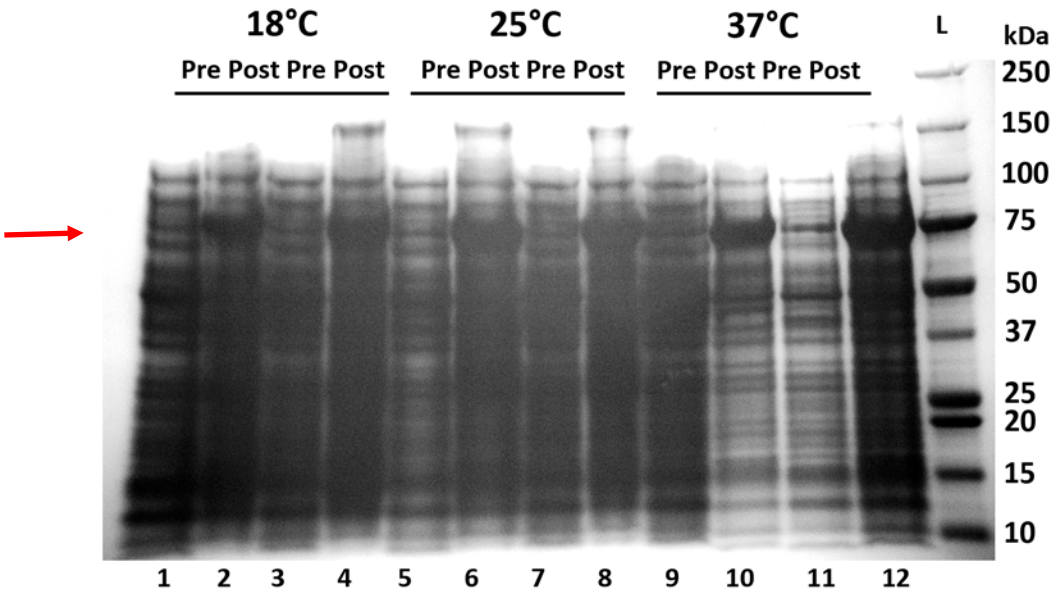


Figure 3.4: Expression trials of WT FL-Shp2: SDS-PAGE Analysis.

SDS-PAGE analysis of WT FL-Shp2. The expected mass was estimated by Protparam (Expasy) to be 70.5 kDa. Each induction temperature (18, 25 and 37°C) was tested in duplicate. Odd and even numbers are Pre- and post-induction respectively. kDa: kiloDaltons, L: MW marker, Pre: Pre-induction, Post: Post-induction.

In this study, no further optimization of expression conditions were performed. Quantities of WT FL-Shp2 were near-identical across all induction temperatures tested; 18°C was chosen as a precaution to decrease the probability of degradation during overexpression. Yields of FL-Shp2 were too low at higher temperatures to warrant induction at 25 or 37 °C for a few hours then harvesting TanSH2-Shp2 concentration seemed highest at 18°C and lowest at 25 °C; this led to an induction temperature of 18°C to be chosen. Protein stability was not assessed post-expression but despite this, at the selected temperatures, protein synthesis in mesophilic bacterial cells would occur at a slower rate than would otherwise result at 25°C or 37°C, decreasing the likelihood of protein misfolding and aggregation. After expression, proteins are at reduced risk of degradation due to bacterial proteases, since these enzymes would be operating at sub-optimal temperatures.

Expression and Purification of FL-Shp2 Constructs

WT FL-Shp2

Scaling up overexpression from 10 ml to litres of growth was not an issue, however, purification of WT FL-Shp2 was problematic. Whilst WT FL-Shp2 was adequately overexpressed, SDS-PAGE analysis of fractions eluted from Ni^{2+}- affinity chromatography (Figure 3.5) indicated a second gel band lower in size to the estimated size of the full length construct (~70.5 kDa). This indicated the presence of either a degradation product or bacterial contaminant. After pooling of fractions for overnight dialysis in the presence of TEV protease and second-pass Ni^{2+} affinity chromatography step, the low-molecular weight band was still present, and in higher

quantities than the previous gel, indicating the band likely corresponds to a degradation product. This band was still present after size-exclusion chromatography, as well as a number of faint bands ranging from 26-50 kDa (Figure 3.6).

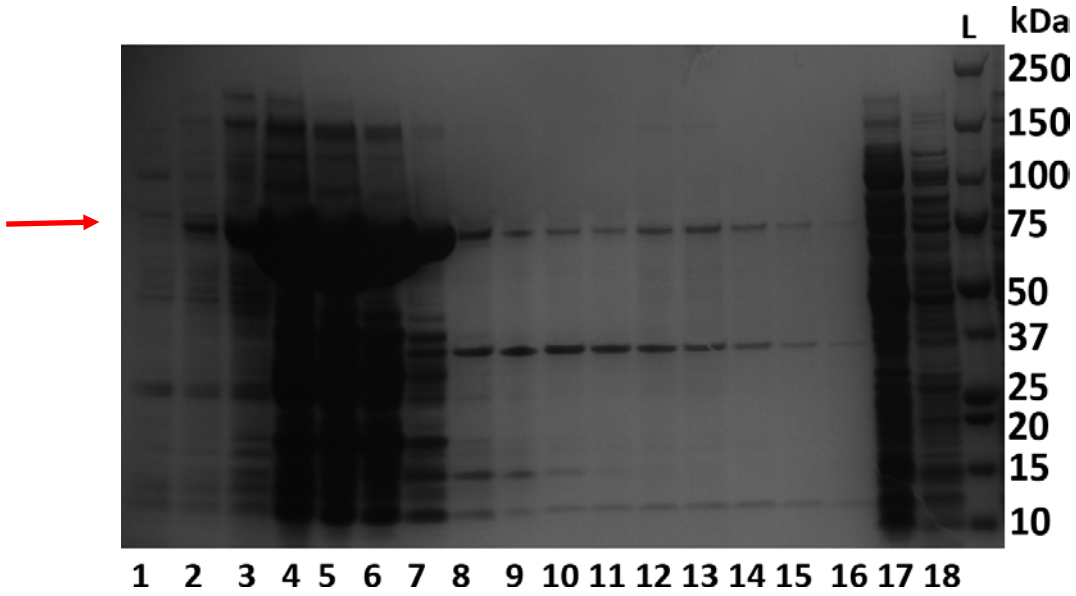


Figure 3.5: SDS-PAGE of WT FL-Shp2 Post-HisTrap Affinity Chromatography.

A 5 ml HisTrap was loaded with the flowthrough of the previous anion exchange chromatography step, containing WT FL-Shp2. Bound protein was then eluted with 250 mM imidazole in 50 mM HEPES pH 7.5, 500 mM NaCl and 0.5 mM TCEP into 1.5 ml fractions (1-16). 17: pre-load sample. 18: Post-load sample. WT FL-Shp2 is observed as a strong band around 75 kDa in size. kDa: kiloDaltons, L: MW marker.

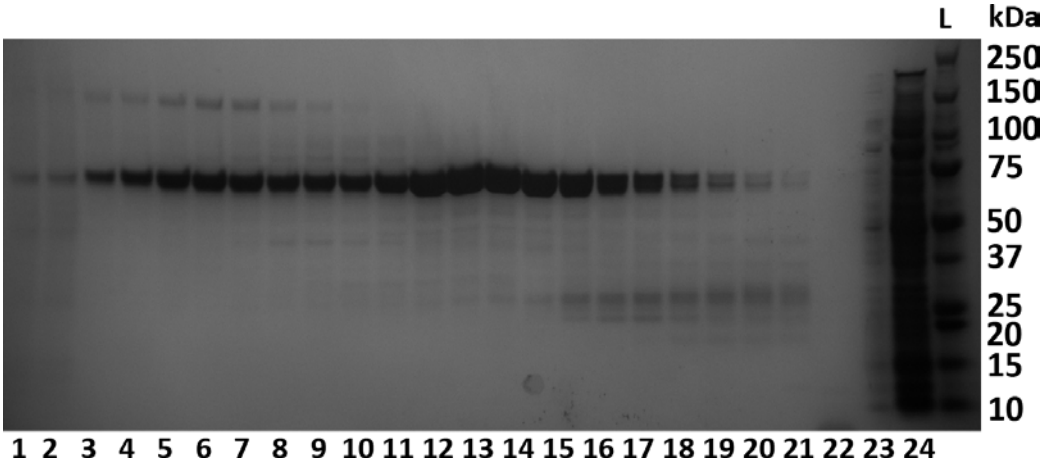


Figure 3.6: Initial SEC Purification of WT FL-Shp2: SDS-PAGE Analysis.

SDS-PAGE analysis of fractions from highest mAU intensity collected post-SEC of WT FL-Shp2. High MW contaminants ~150 kDa are observed in fractions 3-10. Low MW contaminants are observed in nearly all fractions, particularly fractions 15-21. Fractions 23 and 24 contain unpurified supernatant in low and high volumes, respectively. kDa: kiloDaltons, L: MW marker.

Since both bands were present after affinity chromatography, the degradation product could be a C-terminal truncation, as N-terminal degradation would cleave the Hexahistidine tag, decreasing the fusion protein's affinity for the Nickel-NTA resin of a HisTrap.

To decrease the likelihood of degradation, protease inhibitor cocktail tablets (EDTA-free so as not to interfere with Ni^{2+} affinity purification) were added to each buffer. In addition, buffers were kept on ice and measures were taken to minimize exposure of the supernatant and eluted fractions to room temperature. This included performing Ni^{2+} affinity and size exclusion chromatography steps at 4°C. This however, did not suffice to dramatically reduce degradation (results not shown).

An anion exchange chromatography step (a type of ion exchange chromatography) was applied before the first Ni^{2+} affinity chromatography step. This was performed using a HiPrep Q XL 16/10 column. Filtered supernatant was kept on ice throughout and flowthrough was collected during purification (Figure 3.6). After Ni^{2+} affinity purification little to no double bands were observed, indicating that at least some of the proteases were removed. However, subsequent overnight TEV protease cleavage during dialysis increased the proportion of degraded WT FL-Shp2. In addition, incomplete cleavage of the Hexahistidine tag was noted. This led to the removal of the dialysis/tag cleavage step from the purification protocol. Little to no lower molecular weight bands were observed upon SDS-PAGE analysis (Figure 3.7).

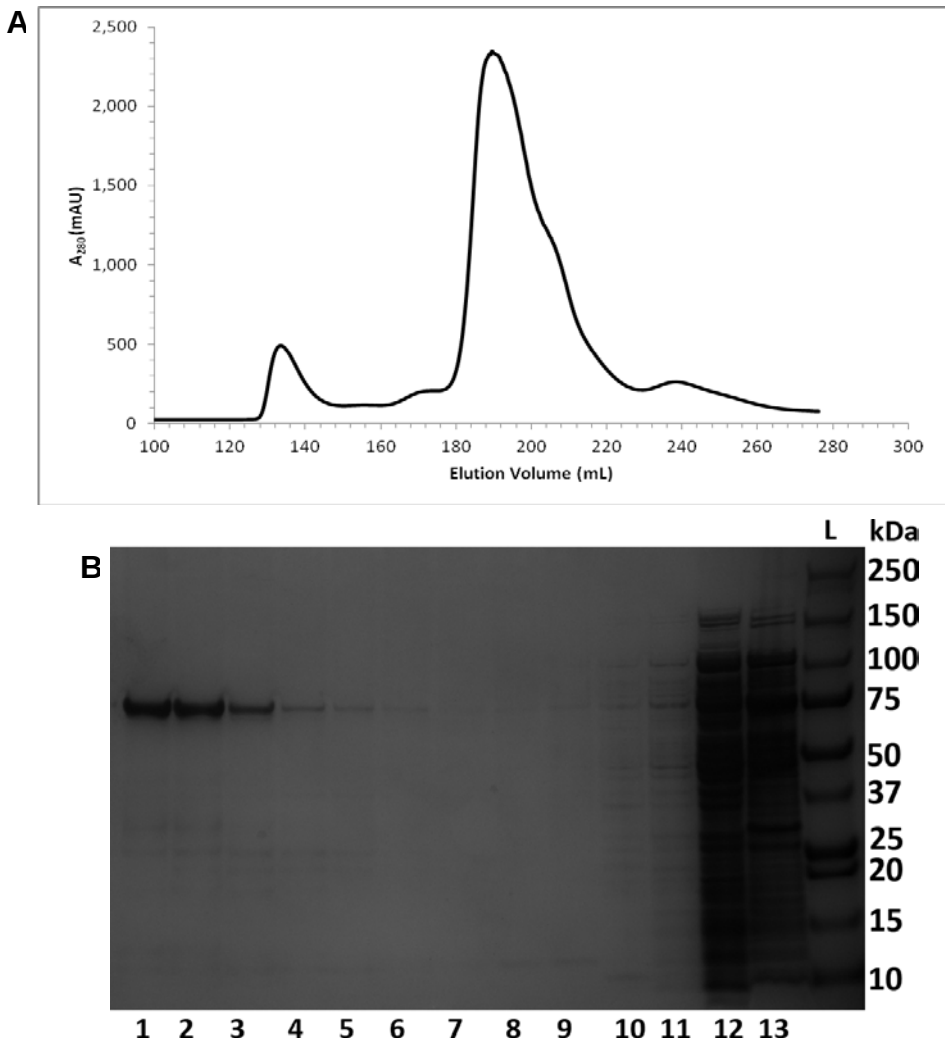


Figure 3.7: Optimized SEC Purification of WT FL-Shp2: SDS-PAGE Analysis.

(A) Elution Profile of WT FL-Shp2 Post-SEC. WT FL-Shp2 was purified at RT on a prep grade S75 column. Fractions of highest UV AU intensities (~180-210 ml) were collected and analysed by SDS-PAGE. **(B)** SDS-PAGE analysis of fractions (1-10) from highest mAU intensity collected post-SEC of WT FL-Shp2. 10-12: Low to high concentrations of pre-purification supernatant. 13: Post-Ni^{2+} affinity chromatography supernatant, with unbound WT FL-Shp2 present Low MW contaminants are observed in nearly all fractions, particularly fractions 1 and 2. Fractions 1-6 were pooled and concentrated for downstream analysis (not shown in gel). kDa: kiloDaltons, L: MW marker.

Expression and Purification of TanSH2-Shp2

Like WT FL-Shp2, scaling up the bacterial overexpression was not an issue, and large quantities of TanSH2-Shp2 were produced. The initial Ni^{2+} affinity chromatography step (Figure 3.8) indicated a high amount of TanSH2-Shp2 had bound to the column, as indicated by the flowthrough post-loading sample. Tag-removal was inefficient; SDS-PAGE analysis of the second Ni^{2+} affinity chromatography step revealed large amounts of protein left after elution with 250 mM imidazole, indicating that the Hexahistidine tag was still present. Optimisation of TEV protease cleavage conditions, such as increasing protease concentration and/or temperature, as well as decreasing NaCl concentration, had little to no effect. To this extent, downstream experiments were conducted with the tag left on, unless specified.

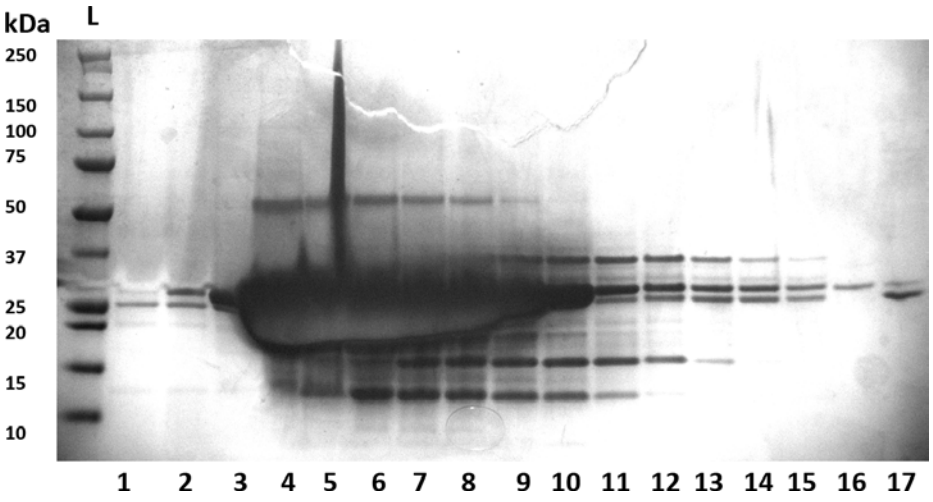


Figure 3.8: SDS-PAGE Analysis of TanSH2-Shp2 Post-Ni^{2+} affinity Chromatography.

A 5 ml HisTrap column was loaded with supernatant containing TanSH2-Shp2 and eluted with 250 mM imidazole in 50 mM HEPES pH 7.5, 500 mM NaCl and 0.5 mM TCEP into 1.5 ml fractions (1-17) Lanes 4-10 contain large quantities of protein at the expected size (~26 kDa), leading to gel overloading. kDa: kiloDaltons, L: MW marker.

TanSH2-Shp2 eluted as a monomer during SEC; this step was performed as a polishing step (Figure 3.9), and early elution fractions (2-7) were pooled and concentrated for use in structural studies. Interestingly, the remaining elution fractions (8-12, Figure 3.9) are contaminated with low molecular weight proteins.

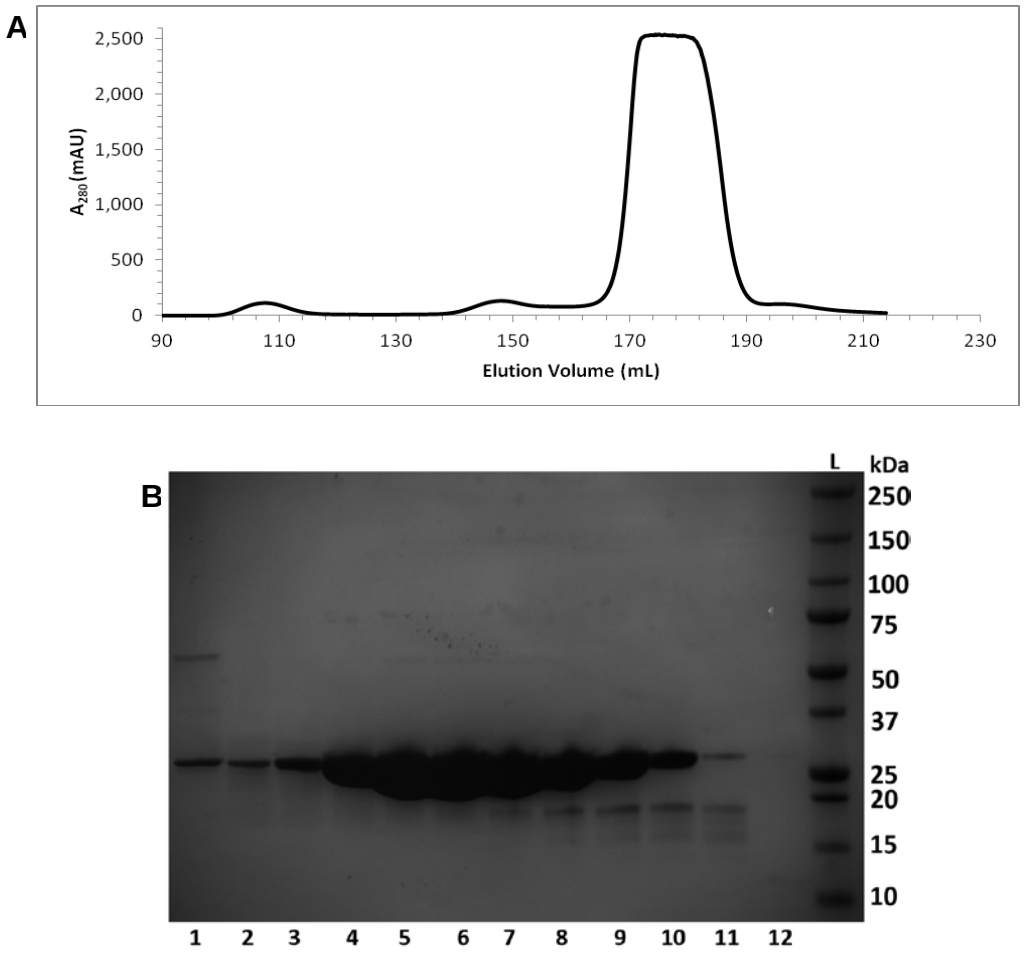


Figure 3.9: SEC Purification of TanSH2-Shp2: SDS-PAGE Analysis.

(A) Elution Profile of TanSH2-Shp2 Post-Size Exclusion Chromatography. TanSH2-Shp2 was purified at RT on a prep grade S75 column. Fractions of highest UV AU intensities (~160-190 ml) were collected and analysed by SDS-PAGE. **(B)** SDS-PAGE analysis of fractions from highest AU intensity collected post-SEC of TanSH2 Shp2. Elution fractions 2-7 were pooled and concentrated for further high-resolution structural studies. Remaining samples were discarded. kDa: kiloDaltons, L: MW marker.

Expression and Purification of Individual SH2 domains of Shp1 and Shp2

Expression trials were performed for these constructs, and expression levels at all temperatures tested were near-identical, leading to the default induction temperature of 18°C to be used for all subsequent growths (Figure 3.10)

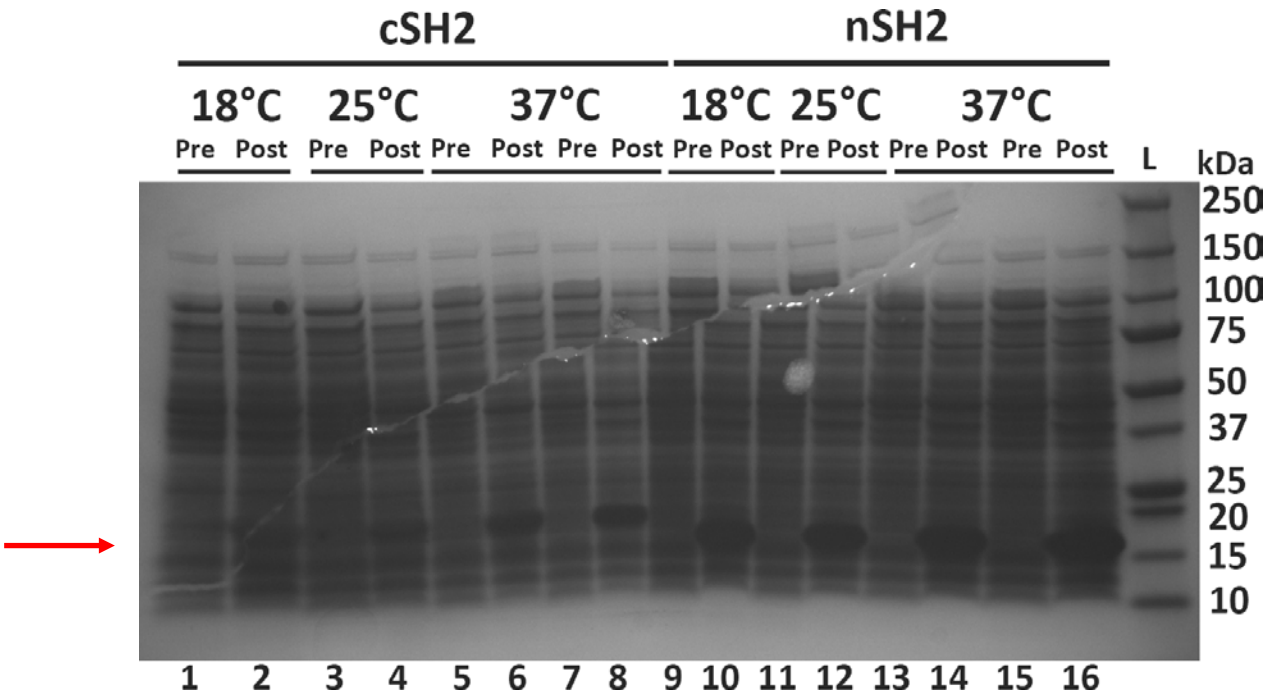


Fig 3.10: Expression trials of n and cSH2-Shp2: SDS-PAGE Analysis.

SDS-PAGE analysis of nSH2 and cSH2 domains of Shp2. The expected mass of the uncleaved proteins was estimated (Protparam, Expasy) to be 15.6 and 17.6 kDa for cSH2 and nSH2 domains respectively. Each induction temperature (18, 25 and 37°C) was tested in duplicate, though only duplicate of 37°C is shown. Odd and even numbers are Pre- and post-induction respectively. kDa: kiloDaltons, L: MW marker, Pre: Pre-induction, Post: Post-induction.

Purification of nSH2 and cSH2-Shp2 by Ni^{2+} affinity chromatography was resulted in all SH2 protein binding the column (Figure 3.11), as was overnight tag-removal by TEV protease and subsequent second step Ni^{2+} affinity chromatography (Figure 3.12), which resulted in a >95% pure sample. However, SEC was performed as a final polishing step so as to remove any higher order soluble aggregates; elution fractions analysed by SDS-PAGE gel. (Figure 3.13 and 3.14). Similar results were obtained for cSH2-Shp1 (Figure 3.15).

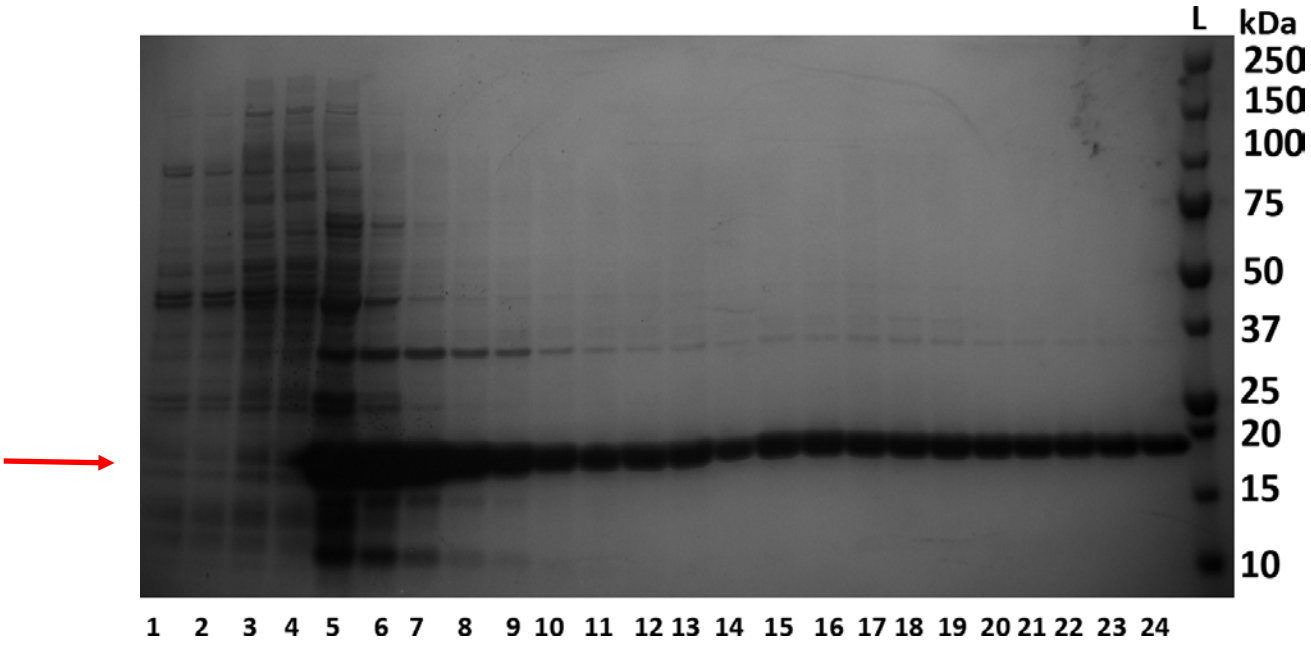


Figure 3.11: SDS-PAGE Analysis of nSH2-Shp2 Post-Ni^{2+} affinity Chromatography.

A 5 ml HisTrap was loaded with supernatant containing nSH2-Shp2 and eluted with 250 mM imidazole in 50 mM HEPES pH 7.5, 500 mM NaCl and 0.5 mM TCEP into 1.5 ml fractions (1-24). nSH2-Shp2 elutes from fractions 5-24, just under the 20 kDa mass marker (computed mass: 15.6 kDa). Elution fractions 1-5 contain no nSH2-Shp2 protein. kDa: kiloDaltons, L: MW marker

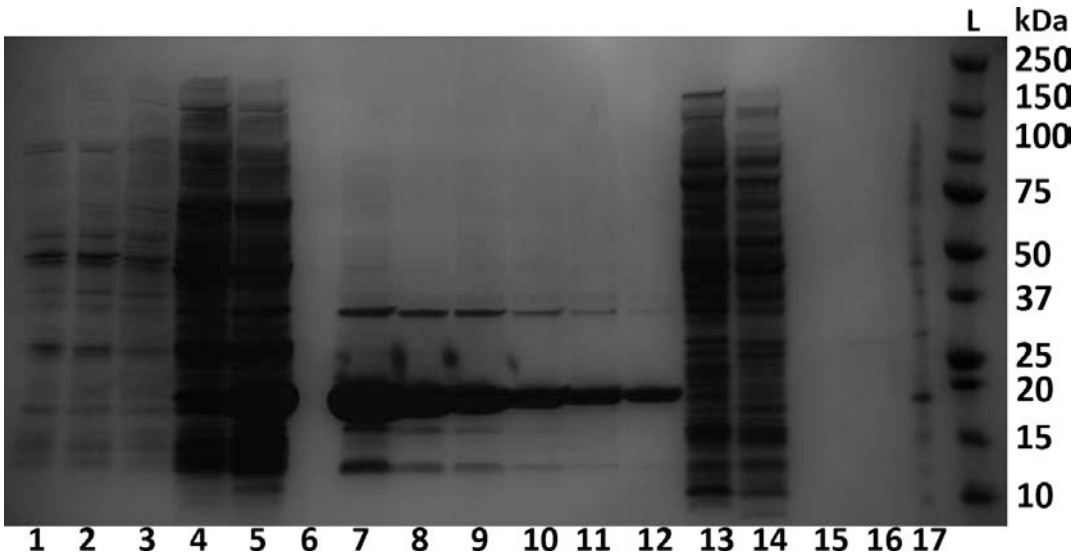


Figure 3.12: SDS-PAGE Analysis of cSH2-Shp2 Post-HisTrap Chromatography.

A 5 ml HisTrap was loaded with supernatant containing cSH2-Shp2 and eluted with 250 mM imidazole in 50 mM HEPES pH 7.5, 500 mM NaCl and 0.5 mM TCEP into 1.5 ml fractions (1-12). cSH2-Shp2 elutes from fractions 5-12, just under the 20 kDa mass marker (computed mass: 17.6 kDa).13: Flowthrough of supernatant post-loading to HisTrap, 14: Flowthrough post-wash step with 25 mM imidazole-containing buffer. kDa: kiloDaltons, L: MW marker

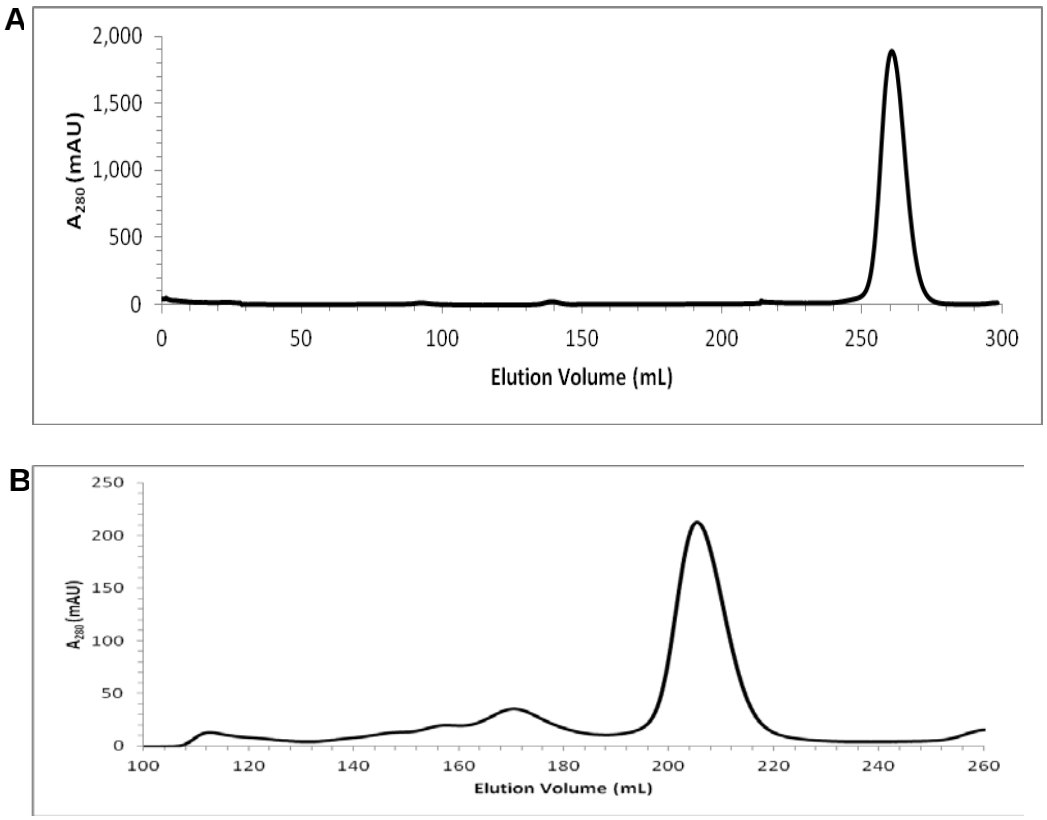


Figure 3.13: Elution Profile of nSH2 and cSH2-Shp2 Post-Size Exclusion Chromatography.

(A) nSH2-Shp2 was purified at RT on a prep grade S75 column. Fractions of highest intensities (~250-275 ml) were collected and analysed by SDS-PAGE. **(B)** cSH2-Shp2 was purified at RT on a prep grade S75 column. Fractions of highest intensities (~195-220 ml) were collected and analysed by SDS-PAGE.

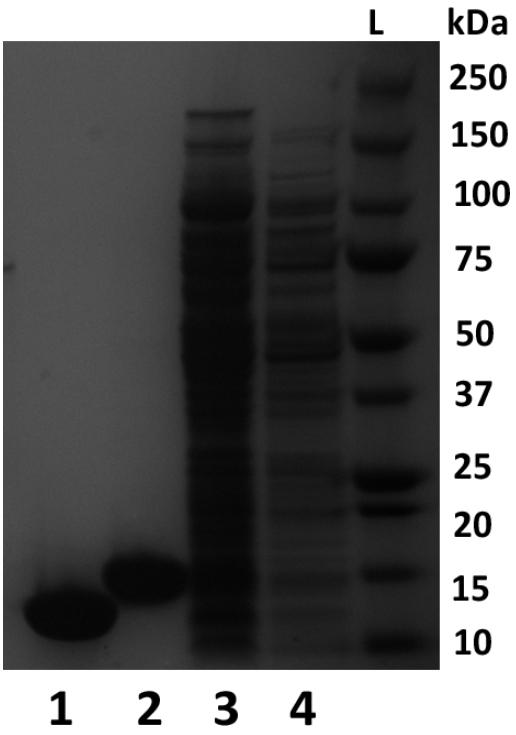


Figure 3.14: Purified nSH2 and cSH2 Shp2: SDS-PAGE Analysis.

SDS-PAGE analysis of nSH2 and cSH2-Shp2 post-purification, prepared for downstream experiments. 1. nSH2-Shp2 , 2. cSH2-Shp2, 3. 250 mM imidazole elution fraction for nSH2-Shp2 HisTrap purification (no nSH2-Shp2 present), 4. 250 mM imidazole elution fraction for nSH2-Shp2 HisTrap purification (no cSH2-Shp2 present) kDa: kiloDaltons, L: MW marker.

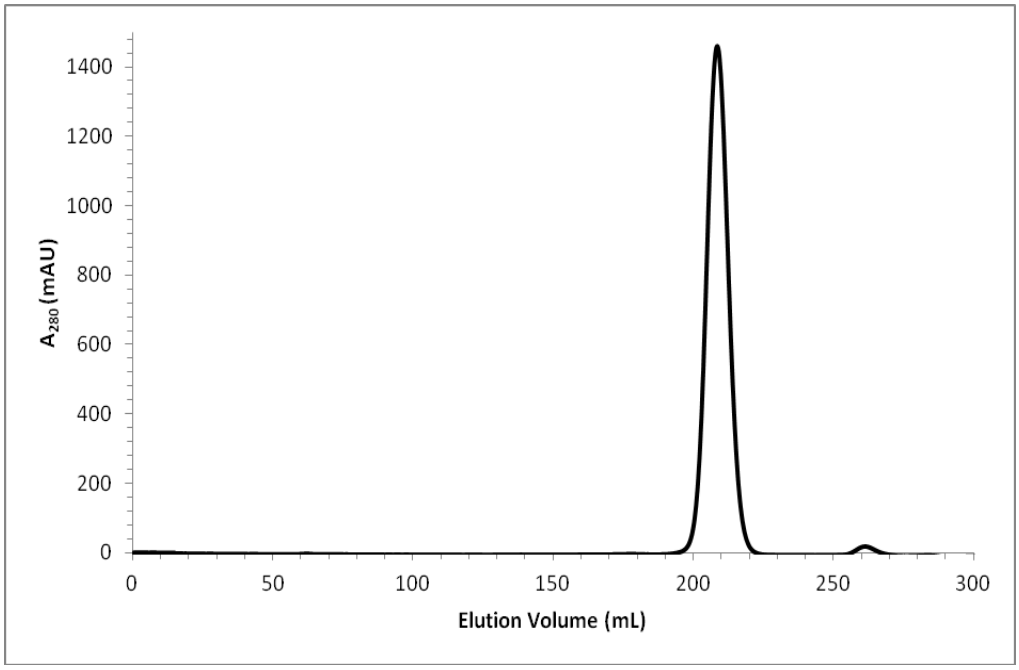


Figure 3.15: Elution Profile of cSH2-Shp1 Post-Size Exclusion Chromatography.

cSH2-Shp1 was purified at RT on a prep grade S75 column. Fractions of highest intensities (~200-220 ml) were collected and analysed by SDS-PAGE.

The nSH2-Shp1 construct, which was expected to behave in a similar fashion to the other three SH2 domains that were purified successfully, yielded an unexpected result. Despite adopting similar expression and purification protocols, little to no protein was identified in eluted fractions post-Ni^{2+} affinity chromatography (results not shown). It was later found to form inclusion bodies and although attempts were made to purify them, the methods employed were not successful This is unusual as the nSH2 domain of Shp2 and cSH2 domains of Shp1 and Shp2 express and purify to adequate levels and purity.

Expression and Purification of PTP-Shp2

The catalytic domain construct of Shp2, PTP-Shp2, was expressed and purified as previously described (Barr *et al*. 2009) excluding the anion exchange step (Figure 3.16). No issues were noted with purification apart from precipitation during diafiltration in preparation for SEC. Any precipitate was removed by centrifugation of the mixture followed by separation of supernatant and pellet by decanting. Highly pure, adequate quantities were produced for NMR and SAXS studies (Figure 3.17).

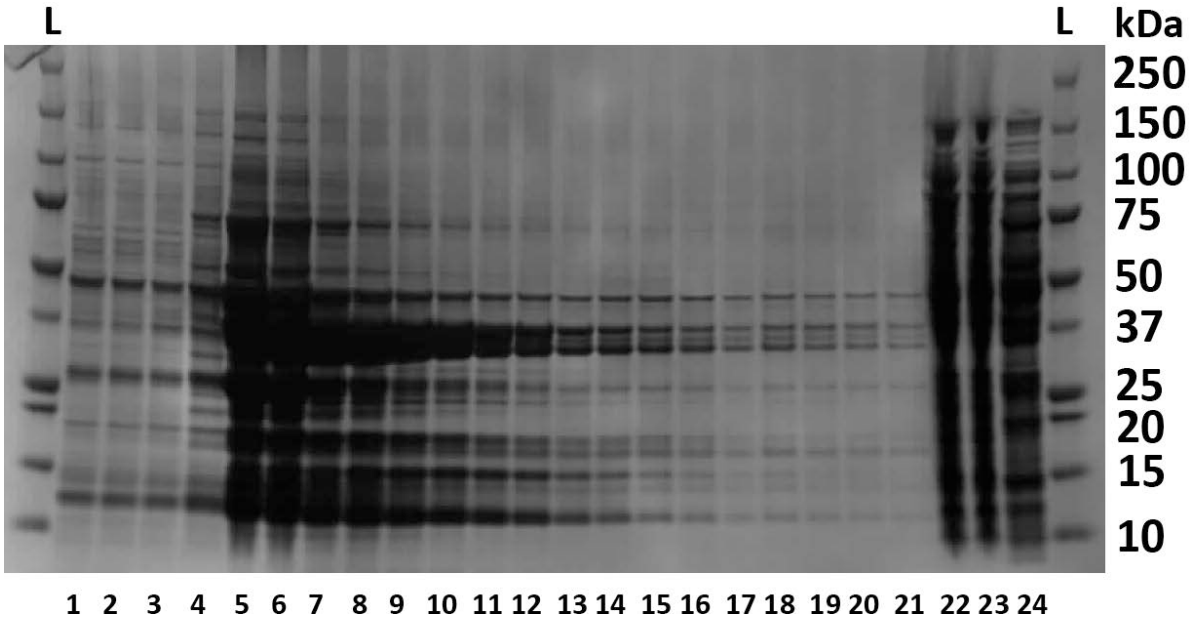


Figure 3.16: SDS-PAGE Analysis of PTP-Shp2 Post-HisTrap Chromatography.

A 5 ml HisTrap was loaded with supernatant containing PTP-Shp2 and eluted with 250 mM imidazole in 50 mM HEPES pH 7.5, 500 mM NaCl and 0.5 mM TCEP into 1.5 ml fractions (1-21). PTP-Shp2 elutes from fractions 5-16, just under the 37 kDa mass marker. 22: supernatant pre-purification, 23: supernatant post-HisTrap purification, 24: flowthrough post-wash with 25 mM-containing buffer as described in Materials and Methods. Fractions 5-16 were pooled for overnight cleavage with TEV protease. (computed mass: 36.8 kDa). kDa: kiloDaltons, L: MW marker.

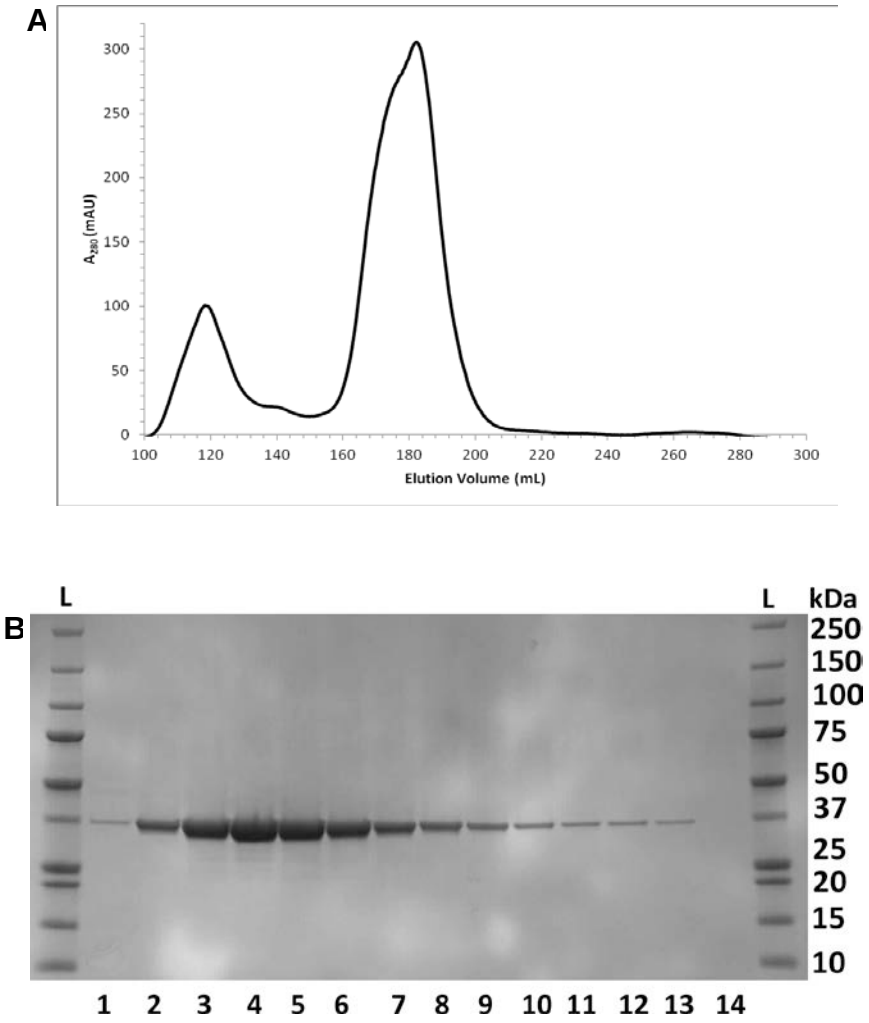


Figure 3.17: SEC Purification of PTP-Shp2: SDS-PAGE Analysis.

(A) Elution Profile of PTP-Shp2 Post-Size Exclusion Chromatography. PTP-Shp2 was purified at 4°C on a prep grade S75 column. Fractions of highest intensities (~160-200 mL) were collected and analysed by SDS-PAGE. **(B)** SDS-PAGE analysis of fractions (1-13) collected post-SEC of PTP-Shp2. All fractions were pooled and concentrated for downstream analysis. kDa: kiloDaltons, L: MW marker.

Mass Spectrometry

Mass spectrometry (MS) was utilised as an accurate polypeptide size determination method for 1) the detection of degradation products for susceptible proteins and 2) to verify the presence of phosphorylation sites.

The mass of WT FL-Shp2 prior to SAXS analysis was assessed by MALDI-TOF MS (Appendix Figure 5) and calculated to be consistent with the mass predicted by sequence analysis to be 70563.5 Da (Artimo *et al*., 2012).

Conclusions

Milligram quantities of pure, stable protein are required for high-resolution structural studies and biochemical analyses. In this chapter, the expression and purification results have been presented. WT FL-Shp2 was subcloned into the pNIC28-Bsa4 vector utilising the LIC protocol provided by the SGC (Gileadi *et al*., 2008) and transformed into high copy number, competent *E. coli* cells. Screening of individual colonies by agarose gel electrophoresis and PCR confirmed the presence of a plasmid around the estimated base-pair size and the correct, in-frame insert sequence. The same was performed with TanSH2-Shp2, with similar results.

As previously mentioned in Chapter II, constructs for nSH2- and cSH2- Shp1 and Shp2, as well as Tan-SH2 Shp1, were kindly donated by Tony Pawson. No further molecular biology steps were required as they were already in-frame within T7-based expression vectors.

CHAPTER IV

NMR ASSIGNMENTS OF THE SHP2 SH2 DOMAINS

The structure of Shp2 in solution has not been investigated extensively. Considering the importance of this PTP in cancer and developmental syndromes, an understanding of Shp2 in solution at the structural level may yield insights into the mechanisms of its biological function and various interactions, as well as providing future drug discovery efforts with data required to design and implement screening programmes effectively.

For the present study various constructs of Shp2 were assessed by solution NMR and potential candidates were taken forward for further investigation. In addition, the cSH2 domain of Shp1 was assessed by solution NMR for comparative purposes.

2D ^{15}N HSQC of WT FL-Shp2

At a mass of 70.5 kDa, FL-Shp2 is too large for conventional NMR experiments. However, as ^{15}N-labelling is relatively inexpensive and straightforward, an attempt was made to collect and analyse a spectrum of this construct. Due to problems with protein solubility, a final FL-Shp2 concentration of 200 μM was used. A ^{15}N TROSY-HSQC spectrum was collected for the WT FL-Shp2 on a 900 MHz Agilent spectrometer; the use of a TROSY pulse sequence as well as a high-field magnet increase the likelihood of acquiring meaningful data with large proteins.

Processing of the data yielded a sparse spectrum (Figure 4.1) containing 47 peaks out of an expected 717. Four peaks lie in the side-chain amide region, with the remainder dispersed along the ^{15}N axis and localised in the random-coil region of the ^{1}H axis, suggesting the peaks represent amino acids within loops segments of FL-Shp2. These peaks could also correspond to amino acids within helices but crucially there is nothing in the NMR spectrum that suggests strands are present. The non-uniform peak intensity throughout the spectrum suggests multiple timescales of motion within FL-Shp2. No further construct or buffer optimization, or additional isotopic labelling (i.e deuteration) was performed for WT FL-Shp2 as the percentage of visible peaks (6.5%) would not be significantly improved to the extent that backbone assignment becomes feasible. At a mass of 70.5 kDa, FL-Shp2 likely tumbles slowly in solution, causing fast T_2 relaxation and thus leading to poor signal. The C-terminal tail may have more independent motion in solution than the other domains of Shp2, which is why sharp peaks in the spectrum are observed. For this reason a “divide and conquer” approach was taken, separating individual domains of WT FL-Shp2 for further study by NMR.

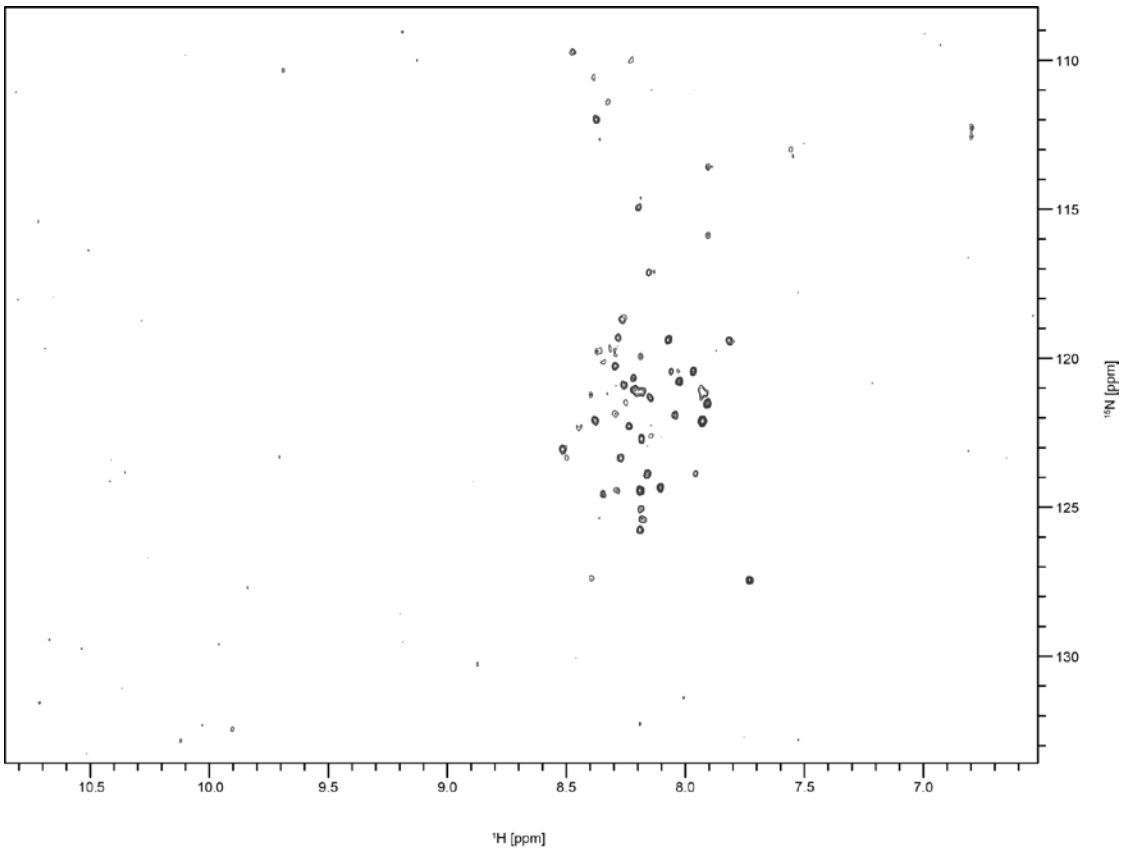


Figure 4.1: ^{15}N-^{1}H TROSY HSQC Spectrum of ^{15}N-Labelled FL-Shp2 WT.

A 2D spectrum was recorded of 200 μM [*U*-^{15}N]-FL-Shp2 WT in NMR buffer at 25°C on a 900 MHz (21.2 T) Varian Spectrometer. Around 47 peaks are observed of an expected 717. Four peaks lie in the side-chain amide region (6.8-7.5 ppm); the remainder in the random coil/α-helical main-chain amide region (7.7-8.5 ppm). Peaks vary in intensity suggesting multiple timescales of motion are present.

2D ^{15}N HSQC of nSH2-Shp2

The ^{15}N HSQC spectrum of nSH2-Shp2 is well dispersed and indicates the presence of folded protein (Figure 4.2). A total of 132 peaks are present out of a possible 133 (>99%), taking into account dual Asparagine/Glutamine side-chain amide and tryptophan side-chain $H^{\varepsilon 1}$ resonances present, as well as the absence of proline resonances. The peaks are near-uniform in intensity, with a few possessing stronger intensities, suggesting variable exchange rates and therefore variable timescale motions in nSH2-Shp2. The two peaks around 10.5ppm in the ^{1}H dimension are in the region Trp $H^{\varepsilon 1}$ resonances are commonly found. Since there is only one tryptophan (6W) in the nSH2-Shp2 sequence, it is reasonable to assume one of the peaks corresponds to 6W-$N^{\varepsilon 1}$. Another scenario is the peak belongs to a main chain amide residue involved in a strongly H-bonded region of β-sheet.

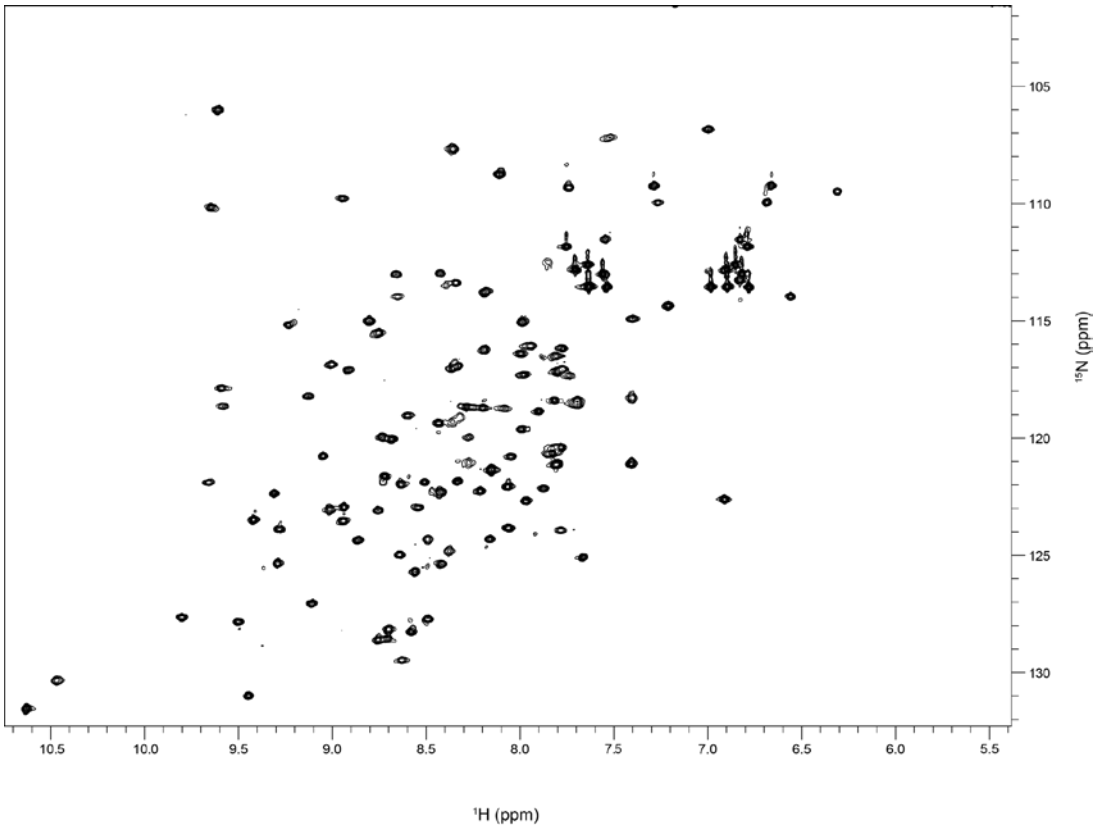


Figure 4.2: ^{15}N-^{1}H HSQC Spectrum of ^{15}N-Labelled nSH2-Shp2.

A 2D spectrum was recorded of 400 µM [*U*-^{15}N]-nSH2-Shp2 in NMR buffer at 25°C on a 600DD MHz (14.1 T) Varian Spectrometer. 132 peaks are observed of an expected 133. Peaks are well dispersed and mostly uniform, indicating a stable globular structure of mixed secondary structure character.

2D ^{15}N HSQC of cSH2-Shp2

Like nSH2-Shp2, the ^{15}N-HSQC of cSH2-Shp2 is also well dispersed, indicating the presence of folded protein (Figure 4.3). 153 peaks out of an expected 149 are present in the spectrum, suggesting cSH2-Shp2 samples multiple conformational states at a slow exchange rate on the NMR timescale. The alpha-helical/unfolded region of the spectrum exhibits more peak overlap then nSH2-Shp2 and this may partly be due to the cSH2 construct being 20 amino acids larger than nSH2-Shp2. Other similarities with the nSH2-Shp2 spectrum include the different peak intensities observed and interestingly, two peaks of which at least one corresponds to 112W $H^{\varepsilon 1}$ resonance.

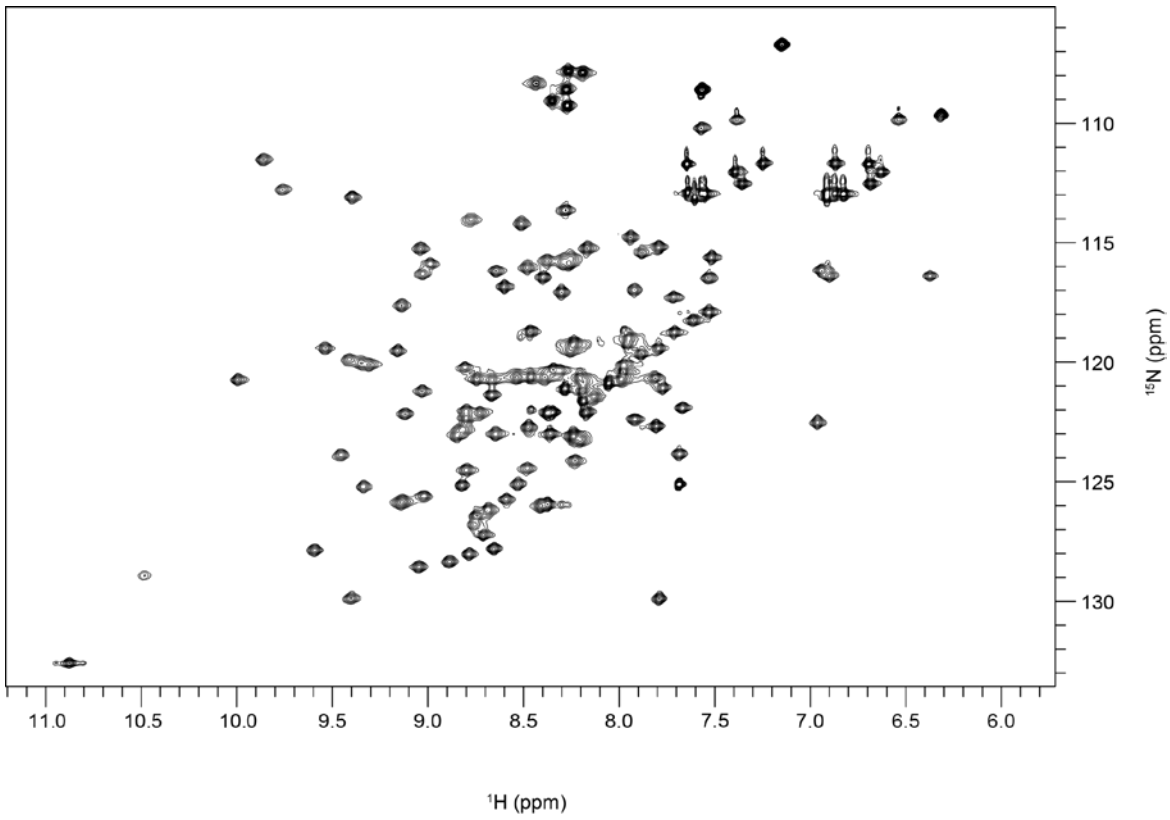


Figure 4.3: ^{15}N-^{1}H HSQC Spectrum of ^{15}N-Labelled cSH2-Shp2.

A 2D spectrum was recorded of 400 µM [*U*-^{15}N]-cSH2-Shp2 in NMR buffer at 25°C on a 600DD MHz (14.1 T) Varian Spectrometer. Peaks are well dispersed and mostly uniform, indicating a stable globular structure of mixed secondary structure character.

2D ^{15}N HSQC of TanSH2-Shp2

His-TanSH2-Shp2

A 400 μM [*U*-^{15}N]-His-TanSH2-Shp2 sample was prepared and a ^{15}N HSQC spectrum acquired thereafter on an 800 MHz Varian spectrometer, Figure 4.4 shows the processed spectrum. 244 out of 266 peaks appear in the spectrum. Peaks are well dispersed, indicative of a folded protein with mixed secondary structure character. Peak intensities vary, suggesting multiple timescales of motion within His-TanSH2-Shp2. With >92% of the expected peak count visible, backbone assignment was deemed to be feasible. However, around 35 peaks were overlapped between 7.5-8.5 ppm in the ^{1}H axis, which complicated the assignment process. Deuteration was used to improve the signal-to-noise ratio and the protein concentration was increased for the collection of assignment spectra. His-TanSH2-Shp2 is 26.2 kDa in size, meaning use of a TROSY-HSQC pulse sequence in the future may enhance peak resolution and signal-to-noise.

The HSQC recorded and analysed in this study resembled that of a previously recorded ^{15}N HSQC (Wu & Guo 2009) though due to differences in spectrometer field strength (600 Vs. 800 MHz) and temperature (17 vs. 25 °C), the HSQC from the current study is higher in resolution and signal-to-noise.

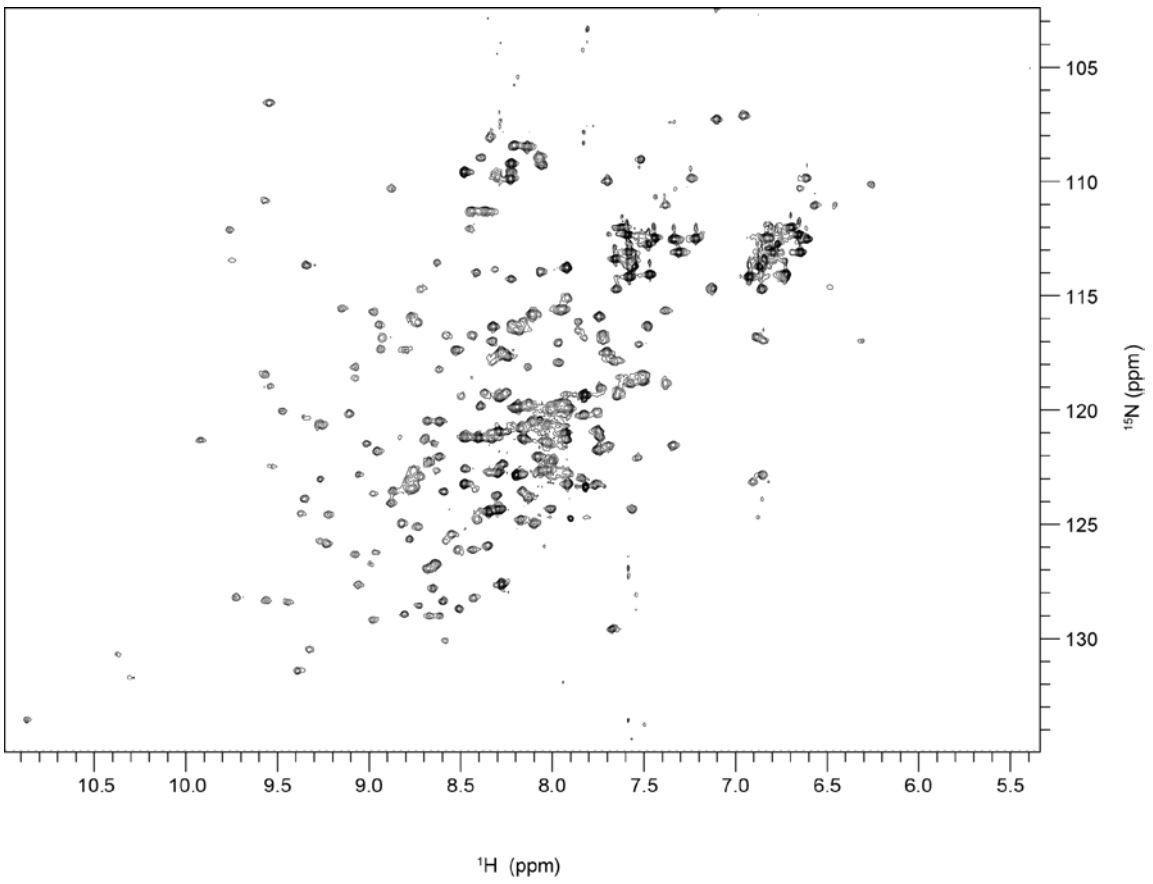


Figure 4.4: ^{15}N-^{1}H HSQC Spectrum of ^{15}N-Labelled His-TanSH2-Shp2.

A 2D spectrum was recorded of 400 µM [*U*-^{15}N]-His-TanSH2-Shp2 in NMR buffer at 25°C on an 800 MHz (18.8 T) Varian Spectrometer. Peaks are well dispersed and mostly uniform, indicating a stable globular structure of mixed secondary structure character.

Differences Exist Between Isolated and Tandem SH2 Domains

Superimposition of HSQC spectra of either nSH2 or cSH2-Shp2 with TanSH2-Shp2 revealed a degree of peak overlap: 53 and 34% of n and cSH2 domain peaks with TanSH2-Shp2, respectively (Figure 4.5).

However, the fact there is limited peak overlap present suggests conformational differences existed between the TanSH2 domain and the isolated SH2 domains, especially the cSH2 domain as the HSQC peaks overlapped less with TanSH2-Shp2 than does the nSH2 domain. Inspection of available crystal structures suggested an interface exists between the SH2 domains when part of the same polypeptide chain.

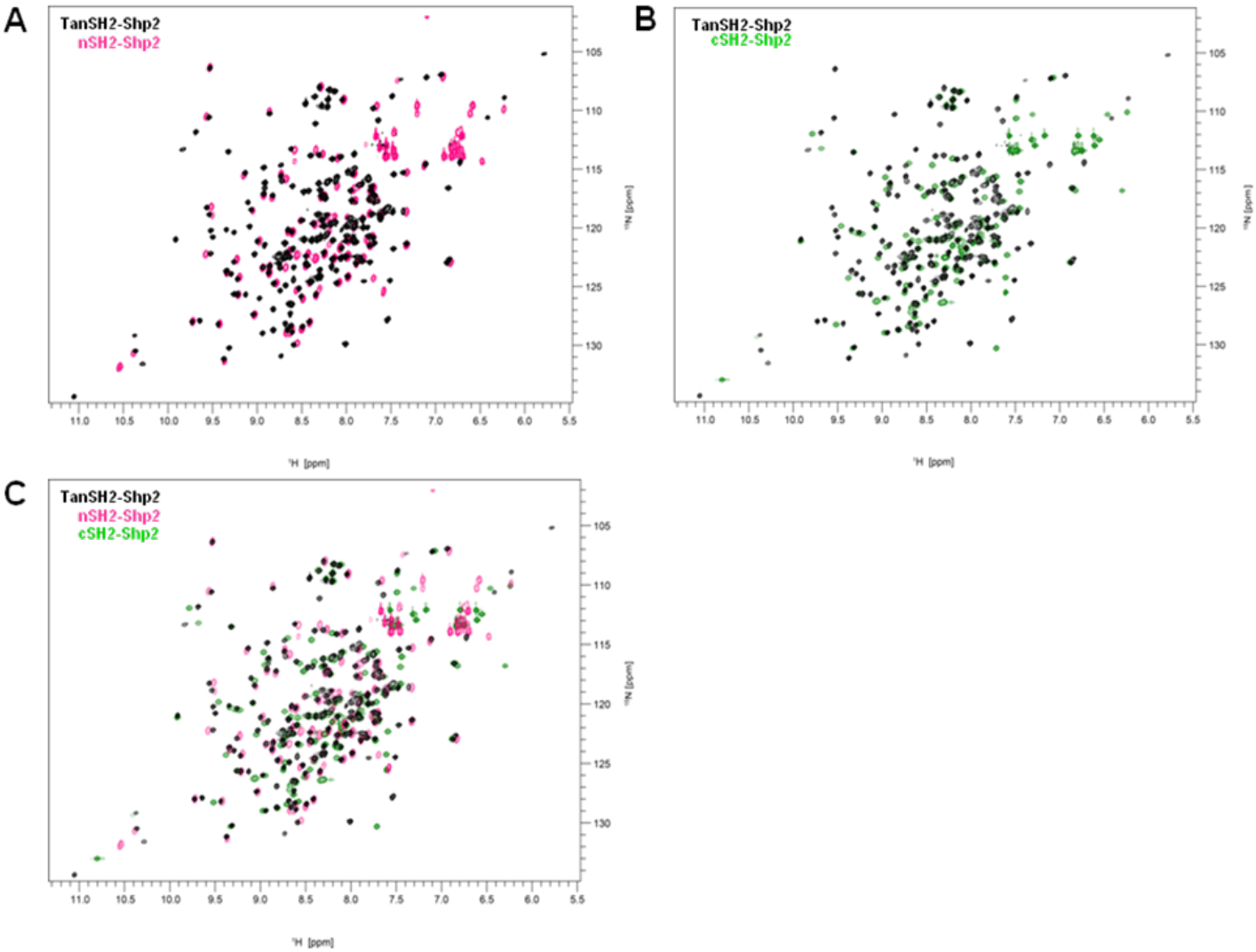


Figure 4.5: Superimposition of nSH2, cSH2 and TanSH2-Shp2 ^{15}N-^{1}H HSQC Spectra.

A 2D HSQC spectrum of His-TanSH2-Shp2 (black) is superimposed with **(A)** nSH2-Shp2 (magenta) **(B)** cSH2-Shp2 (green) and **(C)** both nSH2 and cSH2-Shp2, indicating the chemical shift changes between the individual and continuous SH2 domains, suggesting an intradomain interface may be present.

2D ^{15}N HSQC of cSH2-Shp1

The ^{15}N HSQC of cSH2-Shp1 was well resolved and dispersed, indicative of a folded, mixed secondary structure protein (Figure 4.6). This is consistent with the known structure of SH2 domains. A total of 126 peaks out of 127 are visible (>99%). Around 30 peaks in the spectrum are partially overlapped but easily distinguishable as separate peaks. Interestingly, even though there are two Trp residues within cSH2-Shp1, there seemed to be only one peak in the Trp $N^{\varepsilon 1}$ region, located around 11.2 ppm in the ^{1}H axis, The nSH2 andcSH2 constructs of Shp2 all seem to have two Trp $N^{\varepsilon 1}$ peaks in this region, while TanSH2 demonstrates a possible four Trp side chain peaks. The solution NMR structure of cSH2-Shp1 has been solved both free and in complex with an ITIM peptide (PDB ID: 2YU7; Kasai *et al.,* unpublished data), so backbone assignment is not a priority for this construct, though NMR titrations experiments were performed with ITIM peptides. Surface plasmon resonance (SPR) experiments suggested the SH2 domains of Shp1 and Shp2 were able to bind dITIM-G6b-B. Due to the high sequence identity between the two tandem SH2 domains, one might expect this to be the case (Coxon *et al*., 2012). Observation of chemical shift perturbations (CSP) may yield more information regarding the interactions posited.

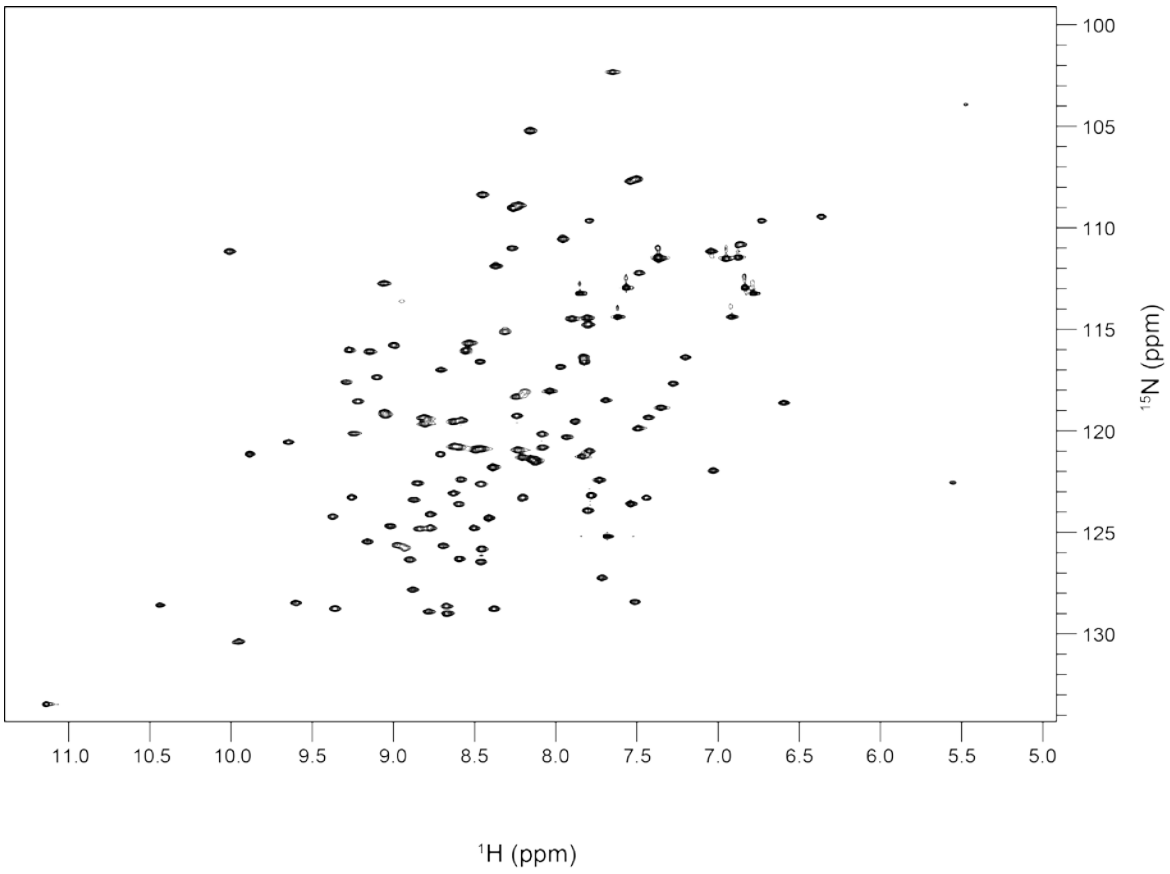


Figure 4.6: ^{15}N-^{1}H HSQC Spectrum of ^{15}N-Labelled cSH2-Shp1.

A 2D spectrum was recorded of 400 µM [*U*-^{15}N]-cSH2-Shp1 in NMR buffer at 25°C on a 600DD MHz (14.1 T) Varian Spectrometer. 126 peaks are observed of an expected 127. Peaks are well dispersed and mostly uniform, indicating a stable globular structure of mixed secondary structure character.

Backbone Assignments of Shp2 SH2 Domains

Isolated SH2 Domains of Shp2

Expression and purification of ^{15}N, ^{13}C-labelled nSH2 and cSH2-Shp2 was straightforward and produced final concentrations of 1.8 and 2.8 mM in 1 ml quantities each, respectively. As the Shp2 SH2 domains are > 15 kDa in mass and stable in solution they did not require deuteration. The small size, high concentration and stability all assist in the fast acquisition of backbone datasets using BEST pulse sequences. For each construct the acquisition of a full set of assignment spectra took 3 days.

92% Sequence-Specific Backbone Assignment of nSH2-Shp2

All six BEST backbone datasets were processed and analysed in order to complete backbone assignment, establishing connectivities between (i) and (i-1) resonances for CA, CB and CO. A C(CO)NH dataset was used to validate backbone assignments by providing additional side-chain carbon chemical shift data for the (i-1) residue (data not shown). Figure 4.7 shows the sequential links between HNCA and HN(CO)CA resonances of residues 46-51 inclusive of nSH2-Shp2. This sequence makes up portions of sheets βC and βD as well as the entire CD loop.

Almost full backbone assignments (H, N, Ca, Cb, C') for 101 out of 110 residues were assigned and are displayed in Figure 4.8. In addition, all 11 Asn/Gln side-chain amides were assigned, one of which (107N) comprises the non-WT C-terminal region, of which full assignments pentapeptide sequence (LINEF) were established.

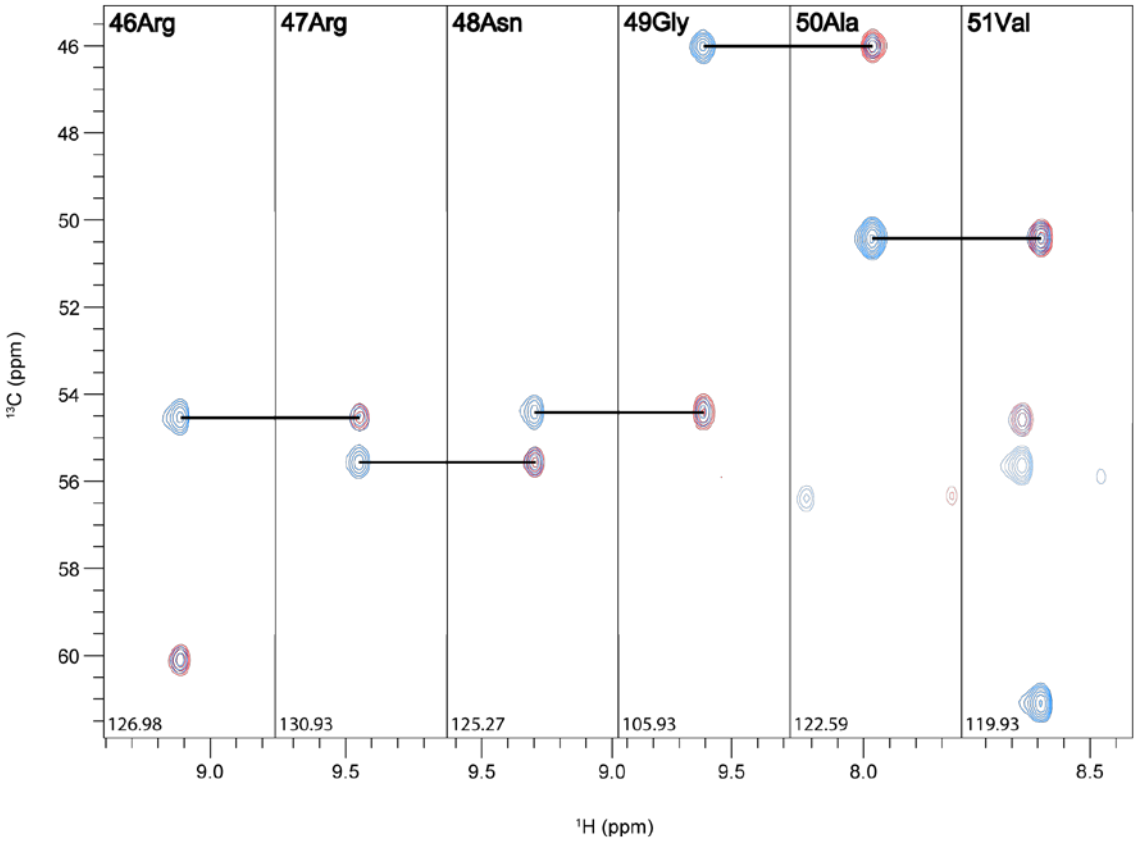


Figure 4.7: Sequential Backbone Assignment of nSH2-Shp2 Using HNCA and HNCOCA Spectra.

Strips of 3D NMR assignment spectra showing connectivities between HNCA (blue peaks) and HNCOCA (red peaks) established using a semi-automated assignment method within CCPN Analysis 2.1.5. Black solid lines show sequential connectivities between HNCA and HNCOCA peaks corresponding to residues at (i) and (i-1) positions for residues 46-51 inclusively for WT nSH2-Shp2, corresponding to the protein sequence RRNGAV. A similar assignment approach was done for HNCO/HNCACO and HNCACB/HNCOCACB spectra in parallel.

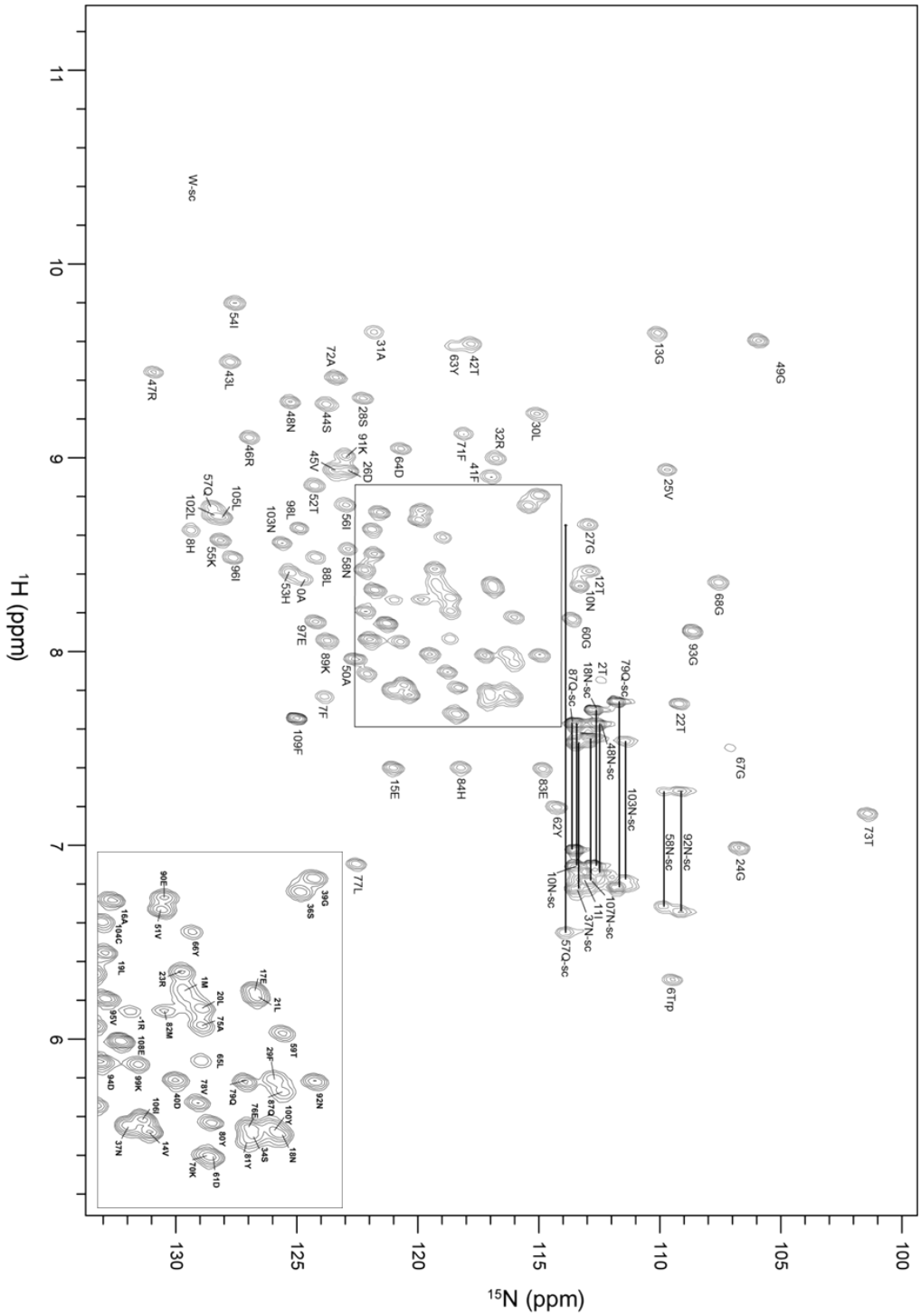


Figure 4.8: Assigned ^{15}N HSQC of ^{15}N, ^{13}C-labelled nSH2-Shp2.

92% nSH2-Shp2 backbone amide signals were assigned using a 1.8 mM sample of uniformly ^{15}N and ^{13}C-labelled protein in NMR buffer at RT using a 600 MHz magnet. The crowded region of the spectrum (boxed, middle) is expanded for clarity (bottom right). “sc” denotes side-chain assignments. Assigned side-chain HN peaks of Asn/Gln are connected via single horizontal black solid lines.

Discounting non-WT residues, a total WT assignment of 94% was reached. Aside from four proline residues, a total of 9 residues remain unassigned, three of which form part of the non-WT N-terminal sequence: -4G, -3S and -2G, 1M marking the beginning of the WT nSH2-Shp2 sequence. The remaining unassigned WT residues are: 3S, 4R, 5R, 35K, 85H and 86G. These residues are highlighted in Figure 4.9 and show the three regions where assignments were unobtainable: 3S-5R are located at the N-terminus, 35K within the BC loop and 85H-86G form part of the loop C-terminal to helix αB. Lack of assignments for peaks in the HSQC were due to the absence of CO, CA or CB (i and i-1) peaks in the 3D data.

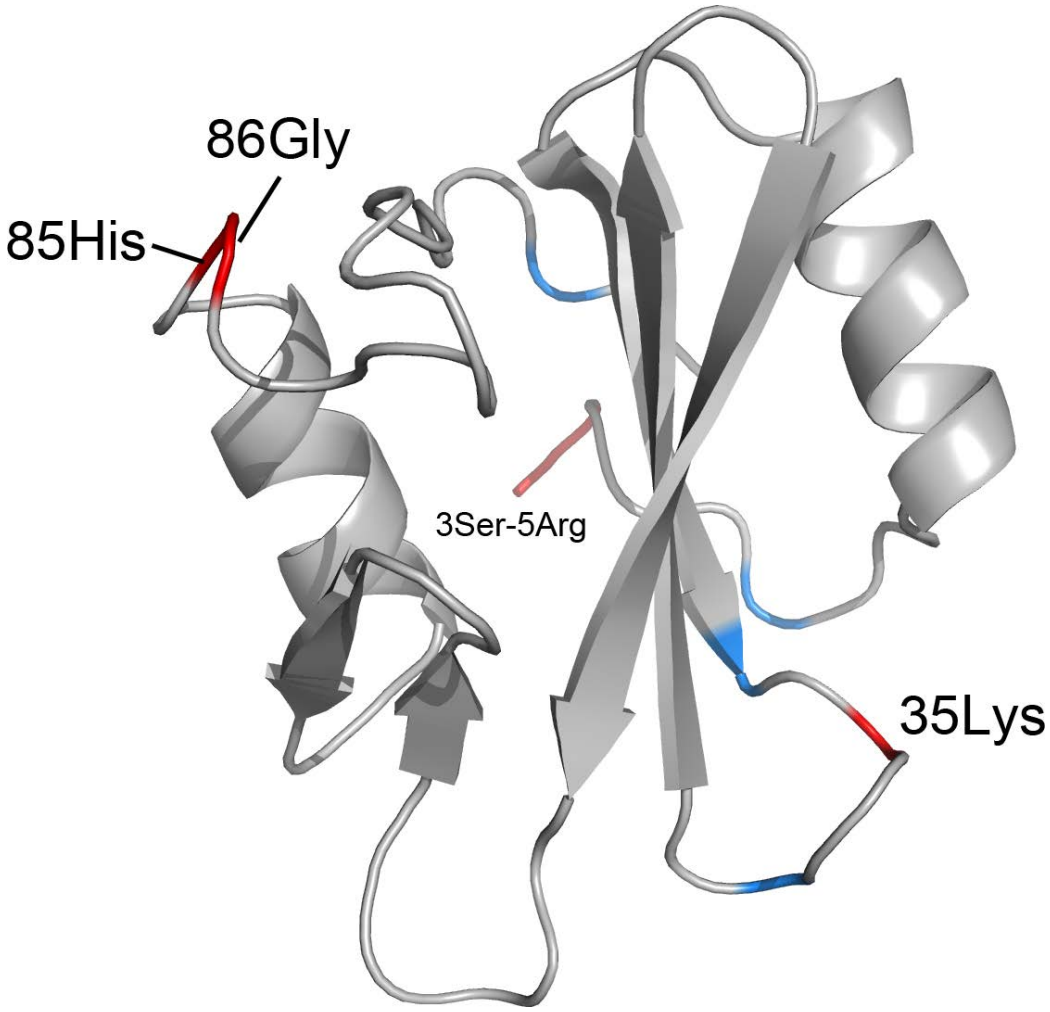


Figure 4.9: Unassigned Regions of nSH2-Shp2.

Six non-Pro residues remain unassigned: 3Ser, 4Arg, 5Arg, 35Lys, 85His and 86Gly (red). All reside within loop regions. Four proline residues (9, 33, 38 and 101Pro) are highlighted in blue. (PDB ID: 4DGP)

95% Sequence-Specific Backbone Assignment of cSH2-Shp2

Assignment of the resonances from cSH2-Shp2 backbone was performed identically to the nSH2-Shp2 assignment. Figure 4.10 shows the connectivities between (i) and (i-1) resonances for the HNCA and HN(CO)CA spanning residues 145-150, which correspond to the sequence GDFVLS. This sequence makes up the portions of the cSH2-Shp2 BC loop and sheet βC.

Backbone resonances for 122 out of 128 WT cSH2-Shp2 residues were assigned (Figure 4.11). Not including the five prolines, five residues remained unassigned. Of these, two are non-WT residues at the N-terminal, the other three being 115G, 141Q and 161N. Discounting non-WT residues, over 97% of the backbone amide signals were assigned unambiguously. These are summarised in Figure 4.12. 115G is located within the N-terminal loop preceding helix αA, 141Gln forms part of the BC loop and 161N is located within the CD loop. These regions are likely disordered and undergoing motions and/or chemical exchange in the intermediate timescale, leading to the lack of assignment. However, side-chain amide assignments for 141Q and 161N were obtained. Of 10 Asn/Gln residues, two lack side-chain amide assignments, these are: 103N and 222N, which are located at the N and C termini of cSH2-Shp2 respectively.

Much as was the case with nSH2-Shp2 a number of peaks within the HSQC were not assigned using 3D data due to the absence of CO, CA or CB (i and i-1) peaks or NH peaks with no apparent connectivities to 3D datasets.

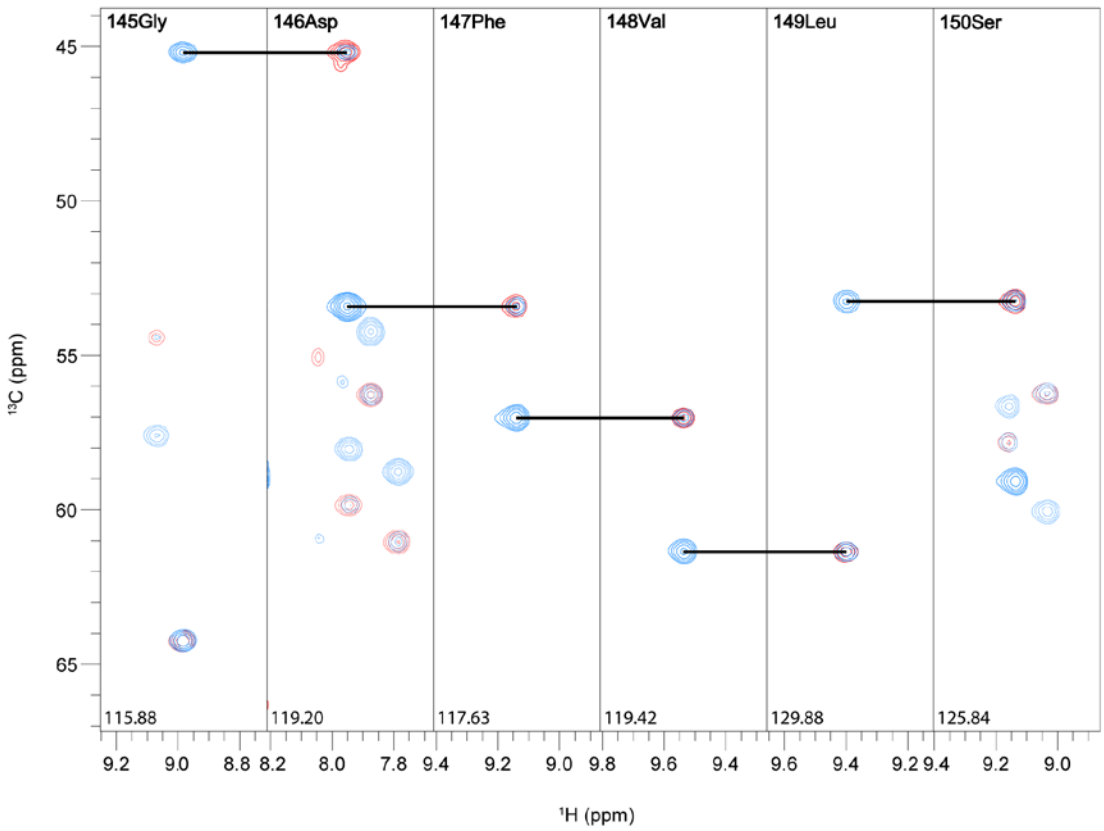


Figure 4.10: Sequential Backbone Assignment of cSH2-Shp2 Using HNCA and HNCOCA Spectra.

Strips of 3D NMR assignment spectra showing connectivities between HNCA (blue peaks) and HNCOCA (red peaks) established using a semi-automated assignment method within CCPN Analysis 2.1.5. Black solid lines show sequential connectivities between HNCA and HNCOCA peaks corresponding to residues at (i) and (i-1) positions for residues 145-150 inclusively for WT cSH2-Shp2, corresponding to the protein sequence GDFVLS. A similar assignment approach was done for HNCO/HNCACO and HNCACB/HNCOCACB spectra in parallel.

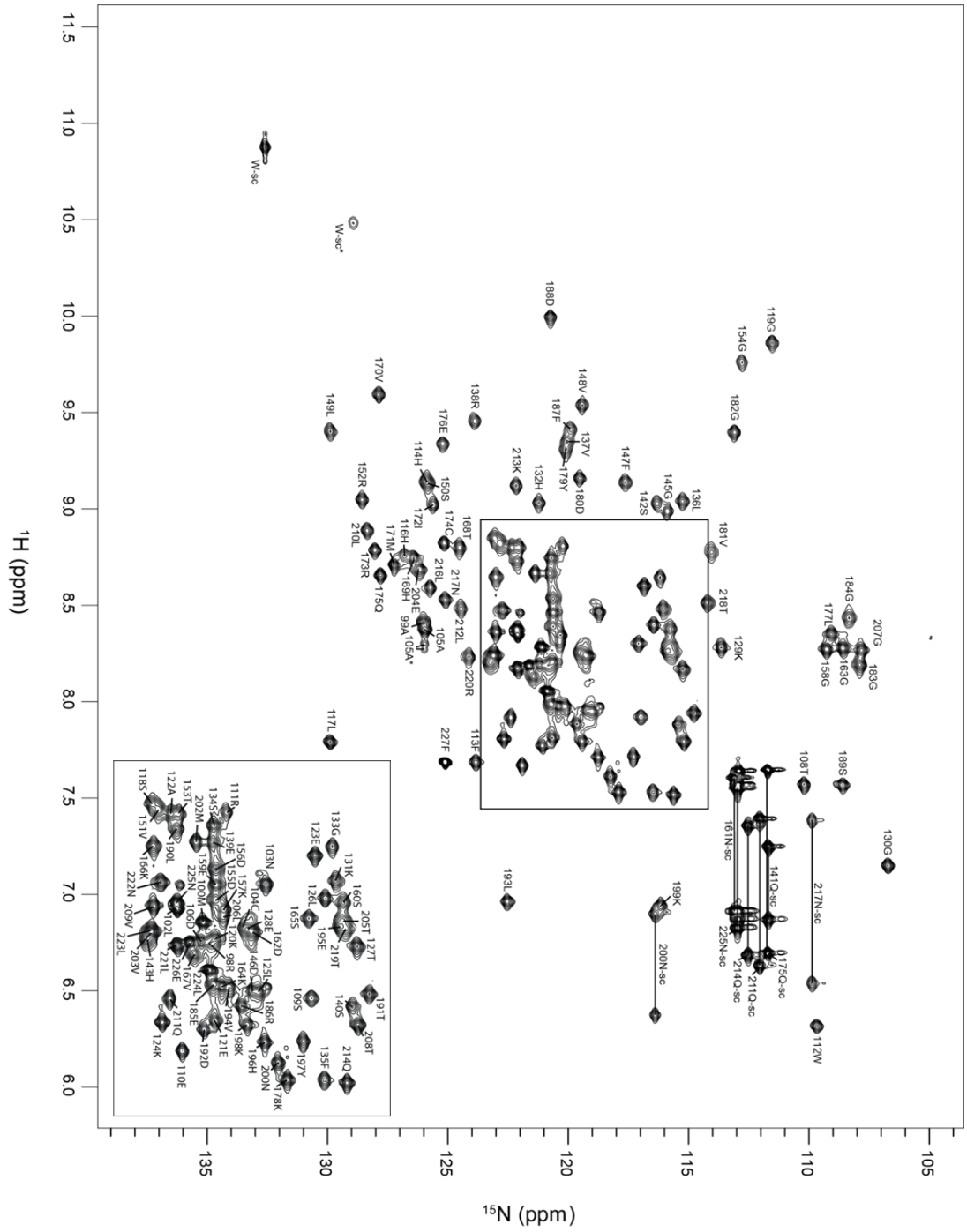


Figure 4.11: Assigned ^{15}N HSQC of ^{15}N, ^{13}C-labelled cSH2-Shp2.

Around 95% cSH2-Shp2 backbone amide signals were assigned using a 2.8 mM sample of uniformly ^{15}N and ^{13}C-labelled protein in NMR buffer at RT using a 600 MHz magnet. The crowded region of the spectrum (boxed, near middle) is expanded for clarity (bottom right). “sc” denotes side-chain assignments. Assigned side-chain HN peaks of Asn/Gln are connected via single horizontal black solid lines. “*” indicates assignment of alternate conformation(s) of residues.

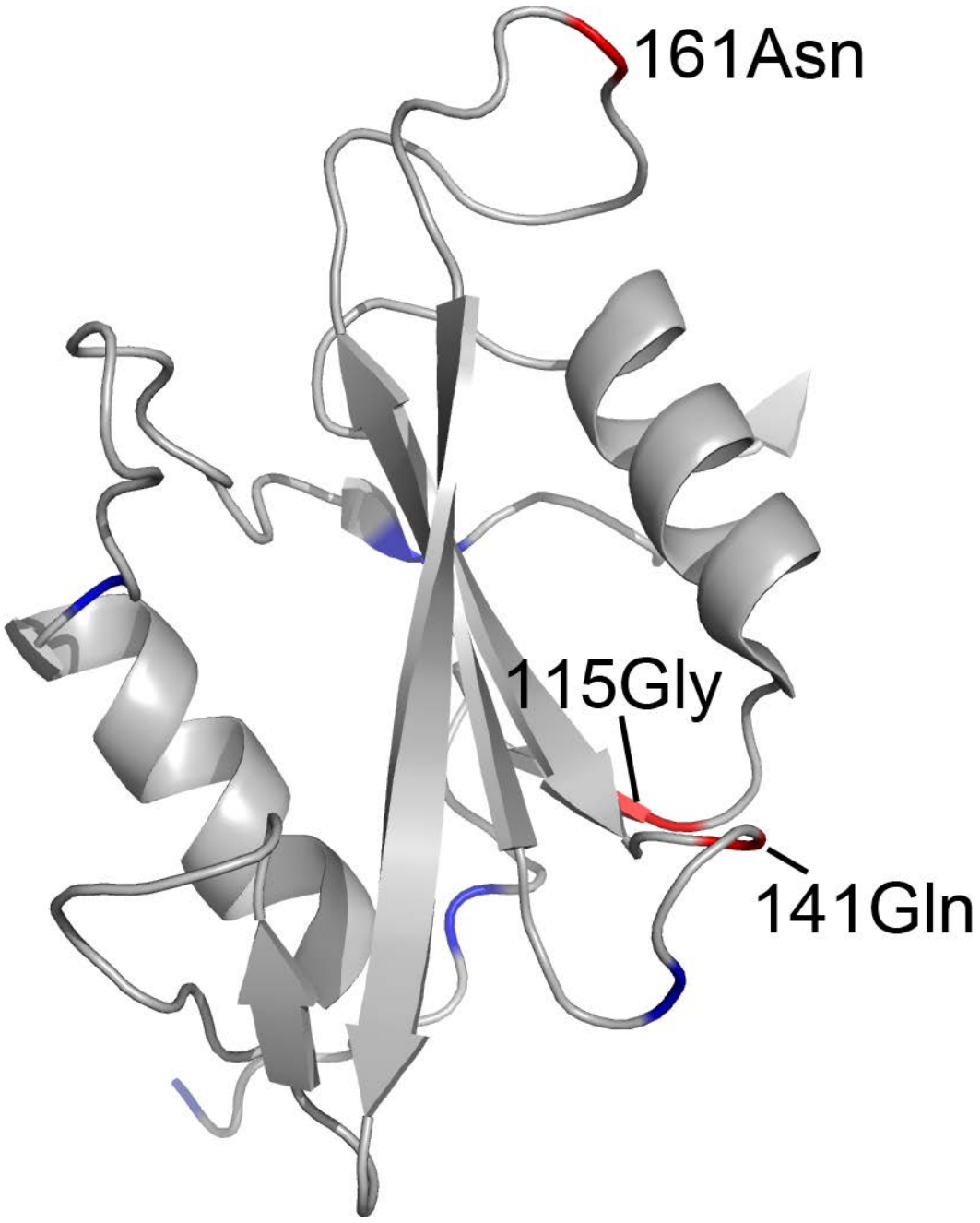


Figure 4.12: Unassigned Regions of cSH2-Shp2.

Three non-Pro residues remain unassigned: 115Gly, 141Gln and 161Asn (red). Both 141Gln and 161Asn reside within loop regions, while 115Gly forms the terminus for sheet βA. Five proline residues (101, 107, 144, 201 and 215Pro) are highlighted in blue. (PDB ID: 4DGP)

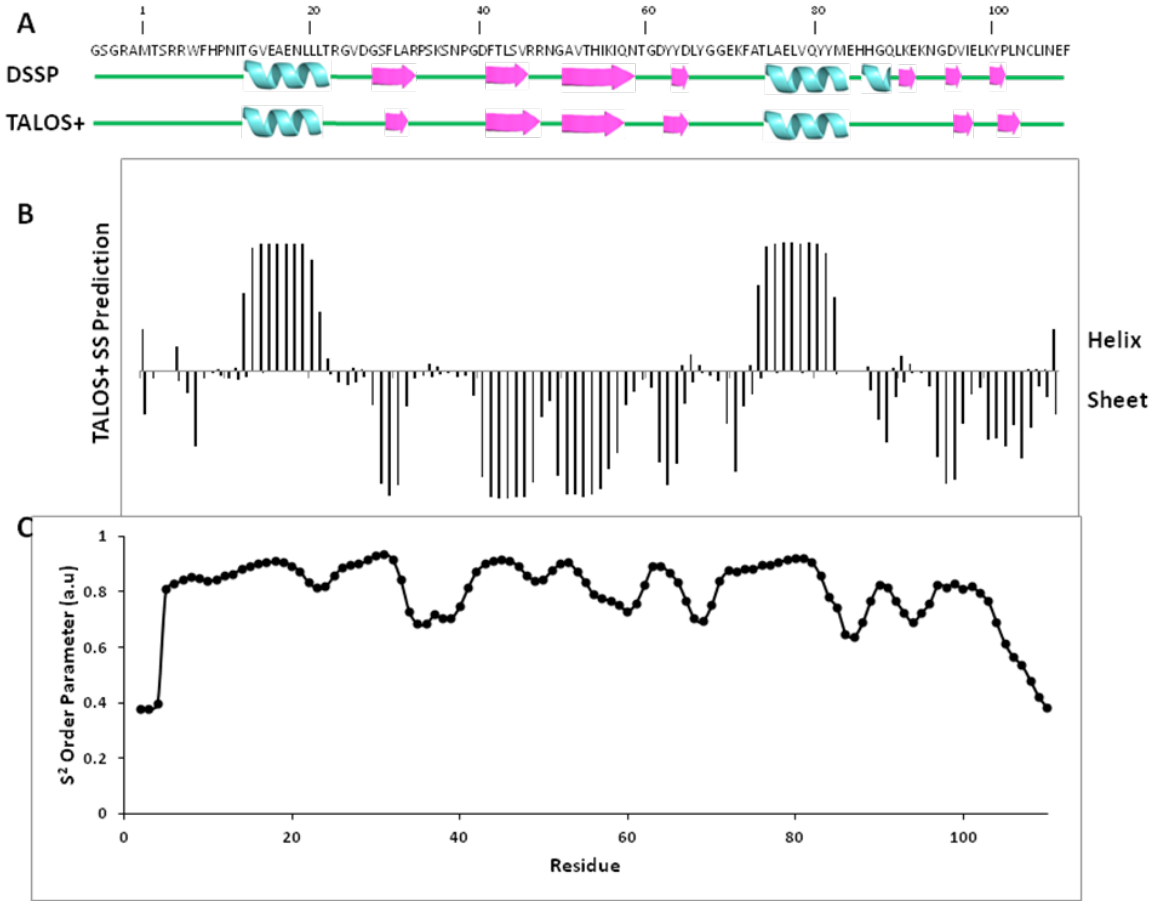
Secondary Structure Prediction of Individual SH2 Constructs

Chemical shift data derived from backbone assignment can be used to predict secondary structure elements and were used in the case of nSH2 and cSH2-Shp2 to further validate assignments. Data were analysed using the TALOS+ system and the results are summarised in Figure 4.13 and Figure 4.14 for nSH2 and cSH2 domains of Shp2, respectively. From this we can conclude that the overall fold of both SH2 domains in solution are consistent with available crystallographic data, with minor variations in helix/strand length.

Alternative Conformational States of nSH2 and cSH2-Shp2 Exist

A number of weak peaks in the nSH2-Shp2 HSQC were identified which had corresponding carbon shifts at positions identical to previously assigned residues. These are for 24Gly and 25Val of the AB loop and 48Asn, 49Gly, 50Ala and 51Val of the CD loop and part of sheet βD. Interestingly, these residues are clustered to one side of nSH2 (Figure 4.15). From this data alone it is not possible to draw conclusions on the precise nature of residue motions or whether the clustering of residues with secondary HSQC peaks is coincidental; NMR-based dynamics experiments can however, shed light on any potential mobility. No significant biological role is apparent for these residues as they are not directly involved in phosphopeptide binding or stabilisation. However, SH2 domains have been demonstrated to have noncanonical roles and functions so this cannot be ruled out entirely (Bernard A. Liu *et al*., 2012).

Alternative conformational states were also observed for residues 102Leu-105Ala, which form the interSH2 linker but are located in the N-terminal of the cSH2-Shp2



construct. This region would presumably display high mobility in the absence of the nSH2 domain, but may become more rigid in the tandem SH2 domain.

Figure 4.13: Secondary Structure Predictions of nSH2-Shp2 using NMR Data.

(A) Secondary structure via DSSP of available crystal structure (1AYD) and prediction using TALOS+. Structural elements are depicted as loops (green solid line), helices (cyan) and sheets (pink). Differences between the crystal and solution structure are apparent, mainly as shortened helices and more β-sheets. **(B)** TALOS+ secondary structure probability histogram, helical propensity is positive, sheet propensity is negative and summarised in (A). **(C)** S^2 order parameter prediction using therandom coil index method. The closer the value toward 0, the more mobile the residue.

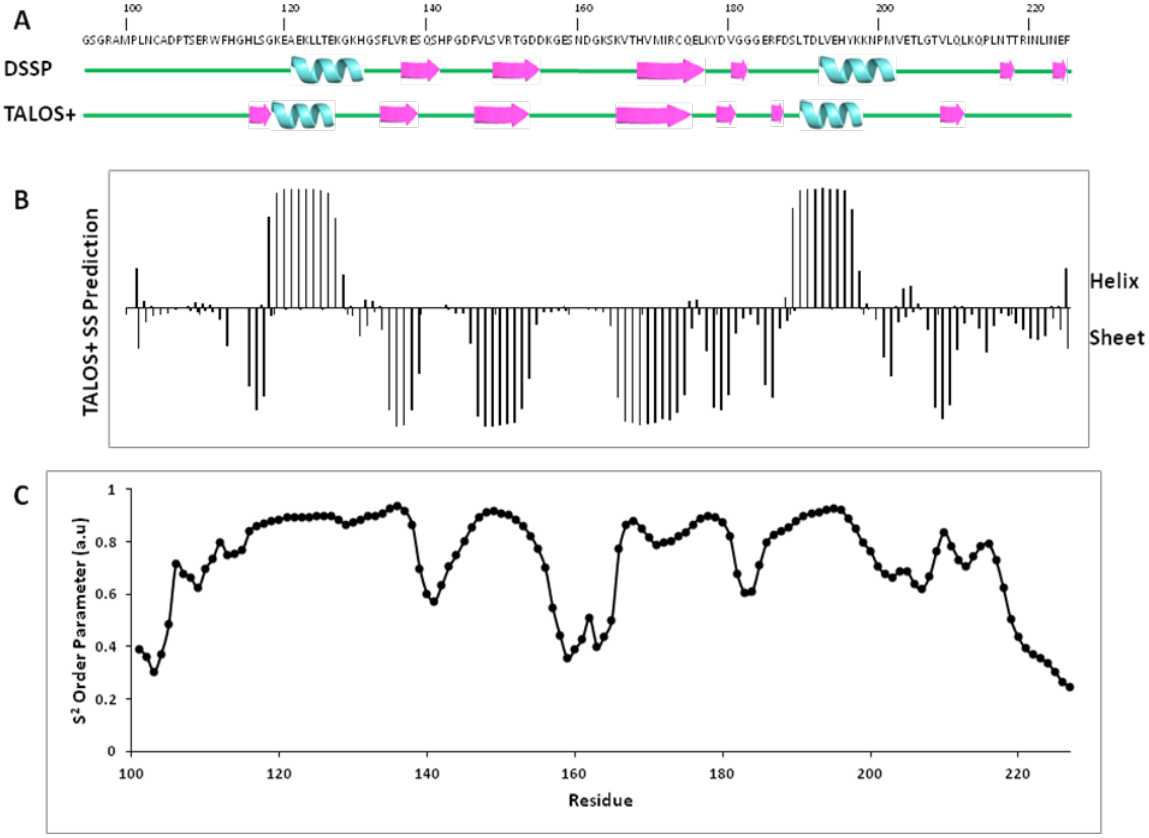


Figure 4.14: Secondary Structure Predictions of cSH2-Shp2 using NMR Data.

(A) Secondary structure via DSSP of available crystal structure (2SHP, chain A) and prediction using TALOS+. Structural elements are depicted as loops (green solid line), helices (cyan) and sheets (pink). Differences between the crystal and solution structure are apparent, mainly as shortened helices and smaller β-sheets. **(B)** TALOS+ secondary structure probability histogram, helical propensity is positive, sheet propensity is negative and summarised in (A). **(C)** S^2 order parameter prediction using the random coil index method. The closer the value toward 0, the more mobile the residue.

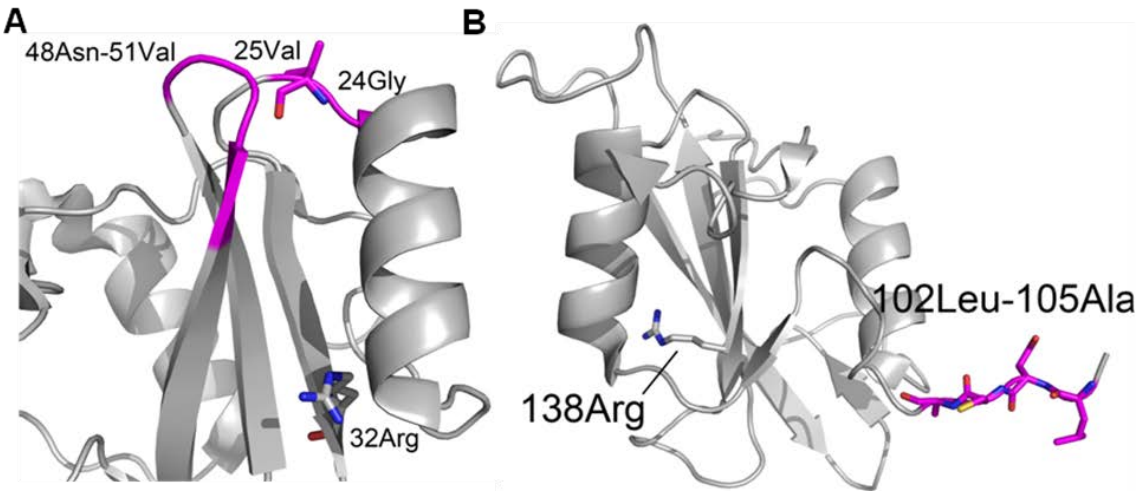


Figure 4.15: Shp2 SH2 Domains Adopt Alternative Conformations in Solution.

Cartoon representation of Shp2 SH2 Domains. **(A)** nSH2-Shp2 (PDB ID: 1AYD, gray) highlighting residues with alternative conformations in solution: 24Gly, 25Val, 48Asn, 49Gly, 50Ala and 51Val (magenta). The invariant arginine is displayed as grey sticks, while 25Val is displayed as magenta sticks. **(B)** cSH2-Shp2 (PDB ID: 4DGP, gray) is coloured similarly, dynamic residues span one region at the N-terminus: 102Leu, 103Asn, 104Cys and 105Ala, which make up part of the interSH2 linker region.

Deuteration of His-TanSH2-Shp2 Improves Spectral Quality

Deuteration and purification of His-TanSH2-Shp2 was straightforward and yielded a pure sample of ^{2}H, ^{15}N and ^{13}C-labelled protein at 2.2 mM concentration in 1 ml. Subsequent ^{15}N-TROSY HSQC collection and analysis reveals a significant increase in resolution due to a decrease in linewidths as a consequence of deuteration and the TROSY pulse sequence; overlapped peaks were more easily distinguishable. 218 peaks are observed (About 82% of the expected peak count if side-chain Asn/Gln peaks are expected, 95% if they are not included). (Figure 4.16).

A standard set of TROSY backbone assignment spectra were collected using this sample.

92% Sequence-Specific Assignment of the Backbone of His-TanSH2-Shp2

For His-TanSH2-Shp2, backbone assignment spectra were processed and analysed to obtain sequence-specific assignments. The assignments from HSQC spectra of the individual Shp2 SH2 domains were utilised to speed up the TanSH2 assignment process and connectivities were made for CA and CB (i) and (i-1) chemical shifts using HNCA-HN(CO)CA and HNCACB and HNCOCACB spectra. Representative HNCA-HN(CO)CA connectivities between (i) and (i-1) resonances for residues 46Arg-51Val (the same as for nSH2-Shp2) can be viewed in Figure 4.17. The HNCO spectrum was of high quality but the opposite was observed for the HN(CA)CO, precluding efficient assignment of CO chemical shifts for the sequential backbone assignment process.

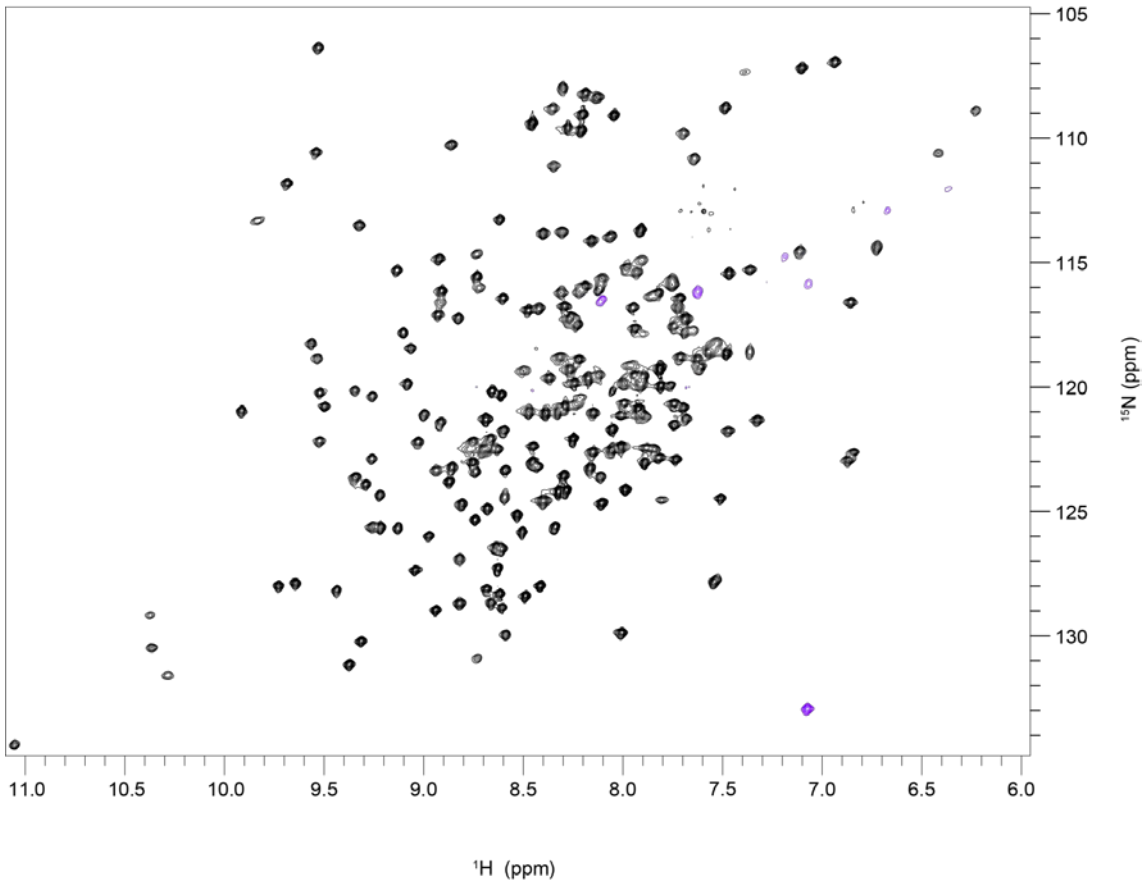


Figure 4.16: ^{15}N-^{1}H TROSY HSQC Spectrum of ^{2}H, ^{13}C and ^{15}N-Labelled His-TanSH2-Shp2.

A 2D spectrum was recorded of 2 mM [*U*-^{2}H,*U*-^{13}C and *U*-^{15}N]-His-TanSH2-Shp2 in NMR buffer at 25°C on an 800 MHz (18.8 T) Varian Spectrometer. Positive peaks appear in black, negative (aliased) peaks appear in purple.

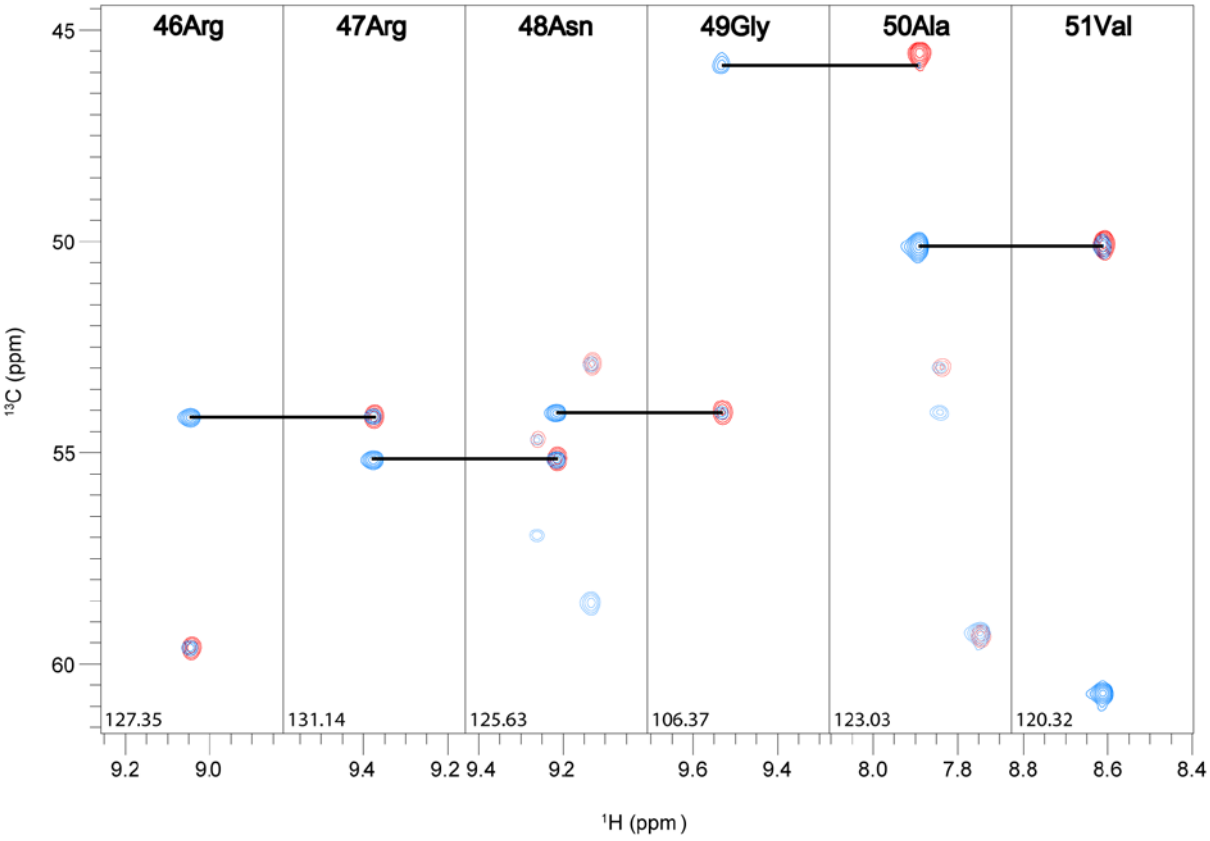


Figure 4.17: Sequential Backbone Assignment of TanSH2-Shp2 Using HNCA and HNCOCA Spectra.

Strips of 3D NMR assignment spectra showing connectivities between HNCA (blue peaks) and HNCOCA (red peaks) established using a semi-automated assignment method within CCPN Analysis 2.1.5. Black solid lines show sequential connectivities between HNCA and HNCOCA peaks corresponding to residues at (i) and (i-1) positions for residues 46-51 inclusively for WT TanSH2-Shp2, corresponding to the protein sequence RRNGAV. A similar assignment approach was done for HNCO/HNCACO and HNCACB/HNCOCACB spectra in parallel.

Of 228 non-prolineresidues within His-TanSH2-Shp2, 209 have been assigned (Figure 4.18), the remaining 19 were unassigned due to poor spectral quality. These numbers reflect the extent of assignment for the entire, His-tag containing construct. When only WT residues are considered, ~95% of these are assigned, as 10/19 unassigned residues are within the His-tag. These 10 residues run consecutively from the N-terminal methionine and incorporate the entire hexahistidine sequence and two linker serine residues, while the TEV cleavage sequence was assigned. His-TanSH2-Shp2 contains eight prolines, all of which were assigned in all backbone atoms apart from main chain imino groups. The remaining unassigned WT residues are coloured on the 4DGP TanSH2 structure in Figure 4.19. These residues are: 4Arg, 5Arg, 35Lys, 85His, 86Gly, 111Arg, 114His, 141Gln and 161Asn. Assignments for most of these residues were not obtained in the individual Shp2 SH2 assignments, the exceptions being 111Arg and 114His, which are both located in the N-terminal loop preceding helix αA. For cSH2-Shp2, both 111Arg and 114His are assigned, but 115Gly was not. A change in the chemical exchange rate and/or dynamics within this loop region may have occurred, causing this variation. No side-chain amide signals were assigned in this procedure due to the nature of the TROSY experiment. Chemical shift tables for all constructs assigned in this study can be found in Appendix 10.

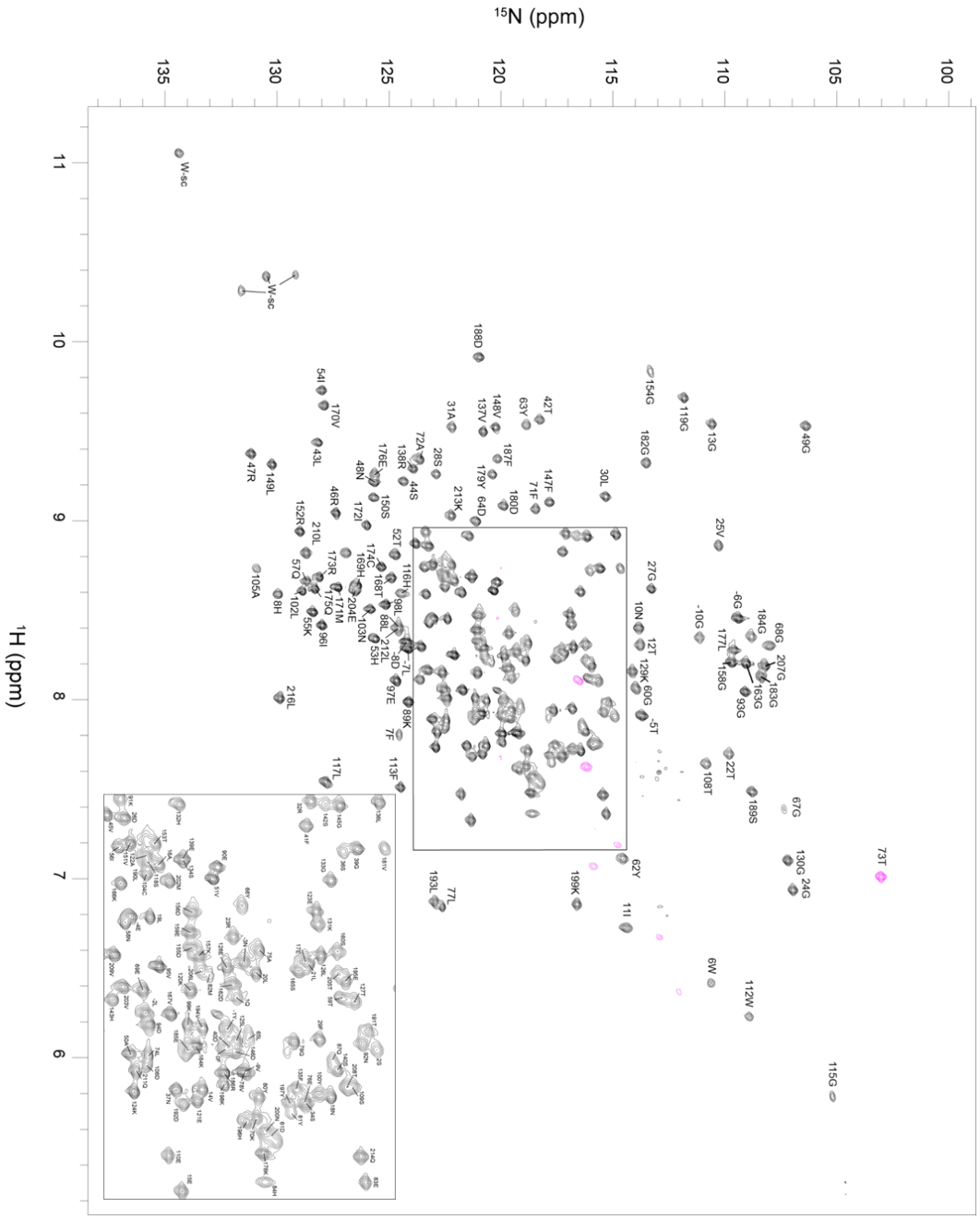


Figure 4.18: Assigned ^{15}N HSQC of ^{2}H, ^{15}N, ^{13}C-labelled TanSH2-Shp2.

92% TanSH2-Shp2 backbone amide signals were assigned using a 2.2 mM sample of uniformly ^{2}H, ^{15}N and ^{13}C-labelled protein in NMR buffer at RT using a 800 MHz magnet.. The crowded region of the spectrum (boxed, near middle) is expanded for clarity (bottom right). “sc” denotes side-chain assignments. 73T is aliased and thus appears as a negative peak (magenta) and has been placed at its original position if unaliased.

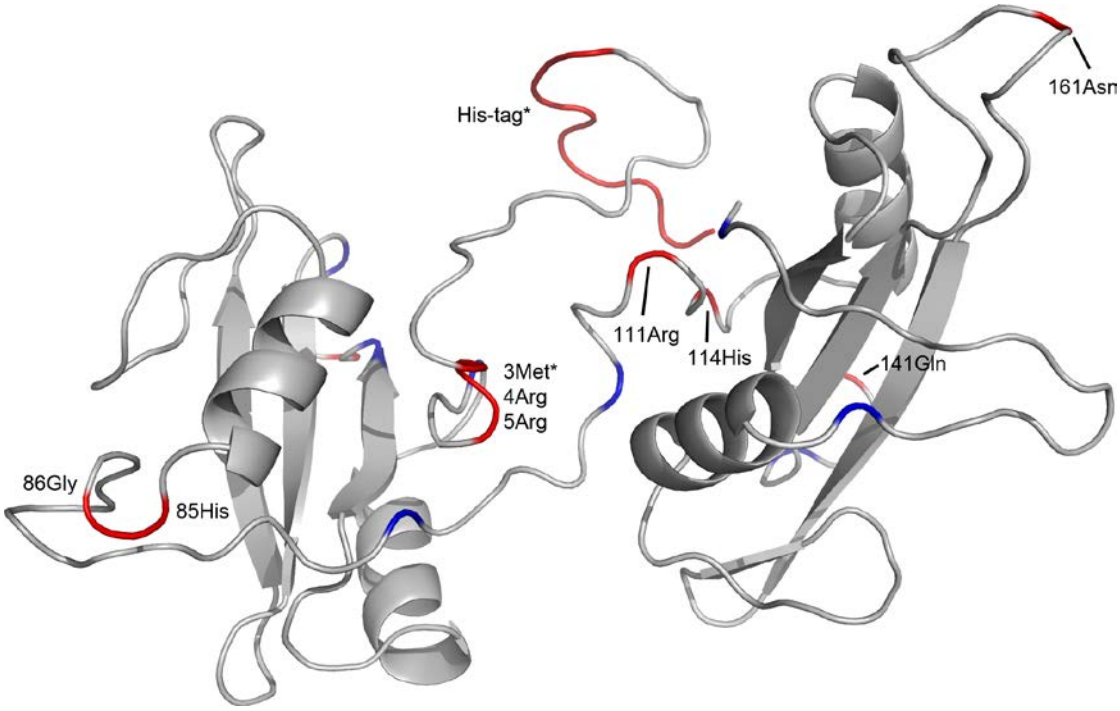


Figure 4.19: Unassigned Regions of TanSH2-Shp2.

Nine non-Pro residues remain unassigned: 4Arg, 5Arg, 35Lys, 85His, 86Gly, 111Arg, 114His, 141Gln and 161Asn (red). Unassigned non-WT residues are indicated by a *. Eight proline residues (9, 33, 38, 101, 107, 144, 201 and 215Pro) are highlighted in blue. (Homology model based on PDB ID: 4DGP)

^{15}N-^{1}H HSQC Spectrum of ^{15}N-Labelled ΔHis-TanSH2-Shp2

Similar conditions were used to acquire NMR data of ΔHis-TanSH2-Shp2 as was previously conducted for His-TanSH2-Shp2. ΔHis-TanSH2-Shp2 is missing the His-tag that TanSH2-Shp2 normally contains. Inspection of the processed ^{15}N-HSQC spectrum yielded similar results to before: 231 peaks are observed of an expected 241 (~96%). Removal of the His-tag does not perturb the NMR spectrum dramatically, however, superimposition of His- and ΔHis-TanSH2-Shp2 spectra (Figure 4.19) reveals around 20 CSPs throughout, varying in shift magnitude with 6W as high as 0.2 ppm in ^{1}H. This indicates that the presence of a His-tag does have a structural effect on the remainder of the TanSH2-Shp2 construct.

The 10 residues with largest CSP are (from largest to smallest): 6W, 7F, 105A, 8H, 104C, 181V, 79Q, 75A, 11I and 32R. Mapping of the perturbed peaks onto the TanSH2 region of 4DGP (Figure 4.20; PTP domain omitted for clarity) localises the affected region to one side of the nSH2 domain, facing and making contacts with the interSH2 linker and cSH2domain. This is expected, as the natural N-terminus of Shp2 is largely buried in between both SH2 domains in the crystal structure of FL-Shp2. 181Val is located in the EF loop of cSH2, far away from the nSH2:cSH2 interface. This residue may form hydrophobic contacts with residues to the C-terminal of the pY in an interacting phosphorylated peptide. Also 32Arg, located in the βB sheet of nSH2, is the highly conserved Arg residue involved in forming a salt-bridge with the pY-phosphate group. Taken together, this suggests that some perturbation of non-interfacial residues is present in His-TanSH2-Shp2. This could be due to direct, non-specific interaction of His-tag residues with both SH2 domains or conformational changes occurring as a result of the His-tag being present.

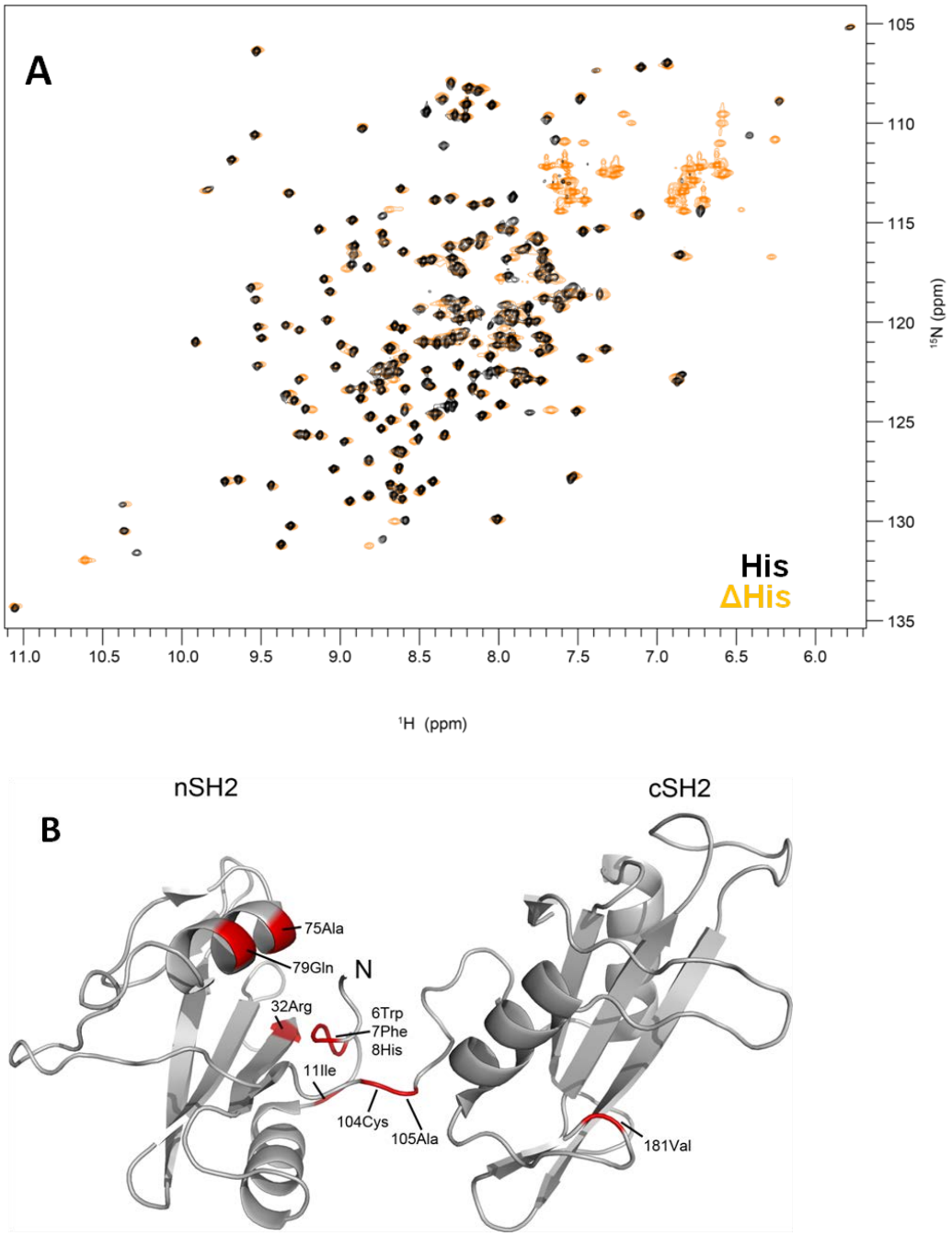


Figure 4.20: The Presence of a His-tag Causes Perturbations to TanSH2-Shp2.

(A) ^{15}N-labelled ΔHis-TanSH2-Shp2 (orange) superimposed with ^{2}H,^{13}C and ^{15}N-labelled His-TanSH2-Shp2 (black). Around 15 cross-peaks have shifted, presumably due to His-tag removal. **(B)** Top 10 residues coloured red on TanSH2 structure (gray, PDB ID: 4DGP). In order of highest to lowest CSP: 6Trp, 7Phe, 105Ala, 8His, 104Cys, 181Val, 79Gln, 75Ala, 11Ile and 32Arg. The majority of CSPs are located on one face of the nSH2 domain or the interSH2 linker region.

Conclusions

cSH2-Shp1 Yields a Well Resolved, Assignable NMR Spectrum

The cSH2-Shp1 construct produces a high quality ^{15}N HSQC. Structures of the former have already been solved by solution NMR in free (PDB ID: 1X6C) and complexed with peptide (PDB ID: 2YU7), though backbone assignment data has not been deposited in the BMRB database.

The cSH2-Shp1, being highly related to the Shp2 SH2 domains, was used for NMR titration experiments. This is further discussed in Chapter 5. However, the nSH2-Shp1 construct did not produce an adequate yield for NMR studies and so cannot be used for comparison in the present study. This is corroborated by another study which also found nSH2-Shp1 insoluble (Coxon *et al*., 2012).

Near-complete Backbone Assignments Were Obtained For All Shp2 SH2 Constructs

94, 97 and 95% of the WT nSH2, cSH2 and TanSH2 domain residues of Shp2 were assigned in the present study. Additional aliphatic carbon assignments were obtained for nSH2 and cSH2-Shp2. Validation of the secondary structure elements of the isolated SH2 domains via TALOS+ and comparison with existing crystallographic data confirms the assignments.

Published backbone assignment data of cSH2-Shp2 (Rubio *et al*., 2013) is consistent with data collected and analysed in the current study, including the non-completeness of sequence-specific assignment. Aside from minor differences in

buffer composition, their study was performed in a similar manner to the present investigation, with additional side-chain proton assignments as well as assignments for a peptide-bound form of cSH2-Shp2. Secondary structure prediction performed by the Rubio *et al*. using the in-program module DANGLE within CCPN Analysis yielded similar results to current crystal data and TALOS+ predictions produced from the current study.

The Individual SH2 Domains of Shp2 Show Conformational Flexibility

A number of residues in both SH2 domains display a second, weaker peak corresponding to an alternate conformational state. 105A is an exception as it displays not only one but two additional peaks in the HSQC and 3D data. 105A is located in the interSH2 linker, but is found in the N-terminal loop for cSH2-Shp2. However, more secondary peaks were observed for nSH2-Shp2 than its C-terminal counterpart. This is strange as characterisation of Shp2 SH2 domains earlier in the study found more peaks than expected in the cSH2 than the nSH2 domain. Whether this is improper peak picking or a phenomenon dependent on other factors not taken into consideration here can be investigated further. The conformational changes observed here could be of biological significance, as it is hypothesised that the nSH2 domain undergoes large scale reorientation to allow PTP activation and as such, conformational flexibility may facilitate this (Keilhack *et al*., 2005).

Interestingly, 35K of the nSH2 domain and 141Q of the cSH2 domain are both located at the same position in the BC loop, 3 residues C-terminal of the invariant arginine residue. In the present study, both 35K and 141Q were unassigned in

individual and tandem SH2 domain 3D spectra. This is likely due to motional dynamics at an intermediate timescale causing line broadening. The BC loop is often highly flexible, as observed in the temperature factors of various SH2 domain crystal structures (Farrow *et al*., 1994), as well as solution NMR structures displaying little NOE information in these regions (Narula *et al*., 1995).

5R and 111R are also in homologous positions for the n and cSH2 domain and form hydrogen bonds with residues adjacent in sequence as well residues within the PTP domain. It is possible that restoration of these hydrogen bonds may alter the exchange dynamics of 5 and 111R. The absence of the Shp2-PTP domain likely perturbs the WT structure and dynamics of these residues.

Comparisons of the 2D ^{15}N-HSQCs of nSH2 and cSH2 domains of Shp2

Two peaks are observed furthest downfield in the in each ^{15}N-HSQC spectrum of the individual SH2 domains of Shp2. As described earlier in this chapter, at least one of these peaks likely belong to the $H^{\varepsilon 1}$ resonances of 6W and 112W (of nSH2 and cSH2 Shp2 respectively) residues. However, it is entirely possible that both resonances in each spectrum correspond to main-chain amide resonances of other amino acids that are uniquely shifted to this region of the spectrum. That said, based on the current data, it is very likely these peaks correspond to Trp side-chains. Since each Shp-2 SH2 spectrum contains double the number of Trp side-chain peaks as expected, unambiguous assignment of these in the TanSH2 construct are impossible based on current data. Differences in the position of these resonances in the single and tandem SH2 also preclude Trp side-chain assignment in the latter.

Collection and analysis of other NMR experiments, including 3D data, or site-directed mutagenesis of the Trp residues and subsequent collection of ^{15}N-HSQCs of said mutants can definitively assign these strange resonances.

The Cysteines of All Shp2 SH2 Constructs are Reduced

TanSH2-Shp2 contains two cysteine residues: 104C in the inter-SH2 linker and 174C in the C-terminus of sheet βD. Both cysteine sulphur atoms are 14.66 Å away from one another in full length crystal structures (PDB IDs: 4DGP, 2SHP) and thus cannot form intradomain disulphide bridges. However, the TanSH2-Shp2 construct, lacking the PTP domain, may display increased domain motions, allowing for positioning of both cysteines in proximity to one another. However, since Shp2 is normally found in the cytosol i.e. a reducing environment, it is unlikely that a 104C-174C disulphide bridge could be formed, especially as both cysteines are at least partially solvent exposed.

An analysis of cysteine atom chemical shift data can give insight into the redox state of the cysteine residue(s) in question (Sharma & Rajarathnam, 2000). More specifically, CB shifts < 32 ppm indicate a reduced cysteine, > 35 ppm for an oxidised cysteine. In all SH2 constructs, CB shifts for 104C and 174C are around 28 ppm. No marked differences are observed between single SH2 construct CA or CB shifts and the TanSH2 construct.

The reduced state of 104 and 174C is in contrast to a paper outlining the crystal structure of TanSH2-Shp2 (Eck *et al*., 1996) which was found to harbour a disulphide bridge between both cysteines under highly reducing (10 mM DTT) conditions,

though the authors suggest this may not be the case *in vivo.* Based on 3D solution NMR data under reducing conditions (0.5 mM TCEP), cysteines 104 and 174 are both reduced and therefore unlikely to form covalent linkages between them. However, the TanSH2-Shp2 structure was solved in the presence of PDGFR-1009 phosphorylated peptide. Complex formation may induce a conformational change bringing both cysteines closer in space for disulphide bridge formation, though a mechanism for oxidation under these conditions remains uninvestigated. PTP-containing pathways can regulated by fluctuating redox potential and though current research implicates the conserved active site cysteine as the commonly modified residue, other solvent exposed cysteines may be redox regulated, perhaps to fine tune spatiotemporal protein-protein interactions (Karisch *et al*., 2011).

The His-tag Perturbs the Chemical Environment of Specific Residues of TanSH2-Shp2

Over 50% of the N-terminal His-tag of His-TanSH2-Shp2 remain unassigned and are located at the extreme N-terminus. Assignments were however, obtained for the TEV cleavage site. Many structural studies of proteins note that residues located at termini are often not observed in electron density for crystal structures and partially assigned in NMR studies due to differences in residue-level dynamics allowing for peaks to be observed in fast chemical exchange. Even still, assignment of extreme terminal residues is uncommon, though the C-terminal residue usually displays a strong peak in HSQC and some 3D spectra.

Removal of the His-tag by repeated TEV cleavage reactions and subsequent NMR ^{15}N HSQC data collection yielded a similar spectrum to His-TanSH2 Shp2, though chemical shift mapping revealed CSPs within the nSH2 domain with preference to the interfacial region between it and the interSH2-linker and cSH2 domain. The partial assignment of the His-tag in His-TanSH2 Shp2 could explain this phenomenon as the tag may have reduced mobility within its constrained environment, leading to fast exchange rates amenable to backbone assignment.

The His-tag is unlikely to interfere with ligand binding as the majority of perturbed residues are located outside of the pTyr and +3 pockets.

The assigned HSQCs of nSH2, cSH2 and TanSH2-Shp2 provide an avenue for exploring various ligand interactions by solution NMR and yielding rich, residue-specific interaction data. In Chapter VI, the work presented here is used to investigate a biological problem.

CHAPTER V

SAXS CHARACTERIZATION OF SHP2 SUB-DOMAINS AND MUTANTS.

To date, almost all high resolution structural information regarding Shp2 has come from X-ray crystallography. This study combines low-resolution solution structural data with high resolution crystallographic data to investigate Shp2 behaviour in the basal and hyperactivated states.

A number of constructs of Shp2 were tested by SAXS at various synchrotron sources in Europe over the course of this project. The following sections concern the in-depth biophysical analysis of the domains of Shp2 as well as full-length constructs. Validation of available crystallographic data with SAXS data in solution will allow the assessment of current structural models.

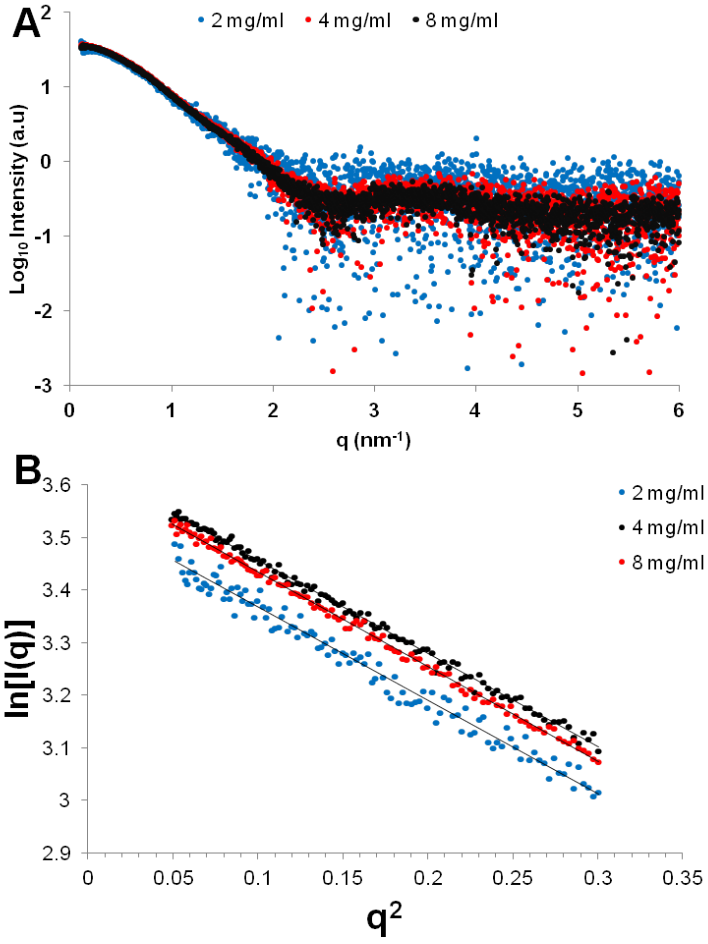
SAXS Analysis of the Tandem SH2 Domain of Shp2

SAXS analyses of the His and ΔHis-TanSH2 Shp2 were performed to complement existing crystallographic data and recently collected NMR data. The sections below reveal insights into these isolated domains in their solution state.

His-TanSH2 Shp2

SAXS data of His-TanSH2 Shp2 were collected at concentrations of 2, 4 and 8 mg/ml. A summary of the data can be found in Figure 5.1 A. Data quality at low-angles were similar for all three concentrations and showed no signs of interparticle effects or aggregation.

Inspection and analysis of the data as a Guinier plot (ln[I(q)] vs. q^2) allows for approximation of I(0) and R_g which is calculated alongside D_{max} and porod volume (V_p) using AUTORG and AUTOGNOM within the ATSAS package (Petoukhov *et al*. 2012) (Figure 5.1 B). Although the Guinier fit at each concentration (Figure 5.1 C) does not suggest any obvious aggregation or interparticle effects, the 2 mg/ml sample shows markedly different D_{max}, I(0) and porod volume (V_p) values in comparison to the 4 and 8 mg/ml datasets. Closer inspection of the low-q range of 2.2 mg/ml revealed large errors in the data and as such the 8 mg/ml data set was used for downstream analysis.



C

Concentration (mg/ml)	R_g (nm)	I(0)	D_{max} (nm)	Porod Volume (nm^3)
2	2.278 ± 0.068	34.363 ± 0.129	7.97	40.11
4	2.313 ± 0.030	37.933 ± 0.057	7.75	45.95
8	2.334 ± 0.014	37.222 ± 0.057	7.65	45.89

Figure 5.1: His-TanSH2 Shp2 SAXS Data.

(A) Buffer Subtracted His-TanSH2 Shp2 SAXS Data. The low-angle data points are similar throughout all concentration tested; the wide-angle data points decrease in quality with increasing concentration. **(B)** Guinier plot and fit of all three concentrations tested. Slopes indicate all samples are free of interparticle interference and aggregation. Data from the 8 mg/ml set was used for downstream analysis. **(C)** Table of calculated parameters from SAXS data. Increases in concentration have an effect on calculated parameters, most markedly the I(0) and Porod volume. Data have not been normalized.

Homology modelling of His-TanSH2-Shp2

Comparison of the experimental SAXS data of His-TanSH2-Shp2 and available crystal structures is hampered by the amino acid sequences of the constructs not matching the crystallographic data, which may affect the results of analyses. His-TanSH2-Shp2 contains an N-terminal His-tag and TEV cleavage site, whereas the crystal structure does not. In addition, WT residues 1-3 inclusive are not part of His-TanSH2-Shp2. For these reasons, a homology modelling approach was used to generate high resolution models using available crystal coordinates of the tandem SH2 domains of Shp2 (PDB IDs: 2SHP and 4DGP) as restraints. I-TASSER (Zhang, 2008; Roy *et al*., 2010) returned five models; the top model having a confidence score of -0.23 where the range is -5 to 2. The higher the confidence score the more reliable the model, and this is also validated by the template modelling or 'TM–score' (Zhang & Skolnick, 2004) which was 0.68 ± 0.12 for the top model. A score of > 0.5 indicates a model of correct topology. In addition, the model had 5107 other similar models within its cluster, meaning these conformations were generated more frequently than others (for comparison, the next best model only had 2480 similar models and a confidence score of -0.95). Visual inspection of the structure generated shows correct topology for both SH2 domains and minimal steric clash (Figure 5.2 C). Based on this, the top scoring I-TASSER model will be used for further analyses.

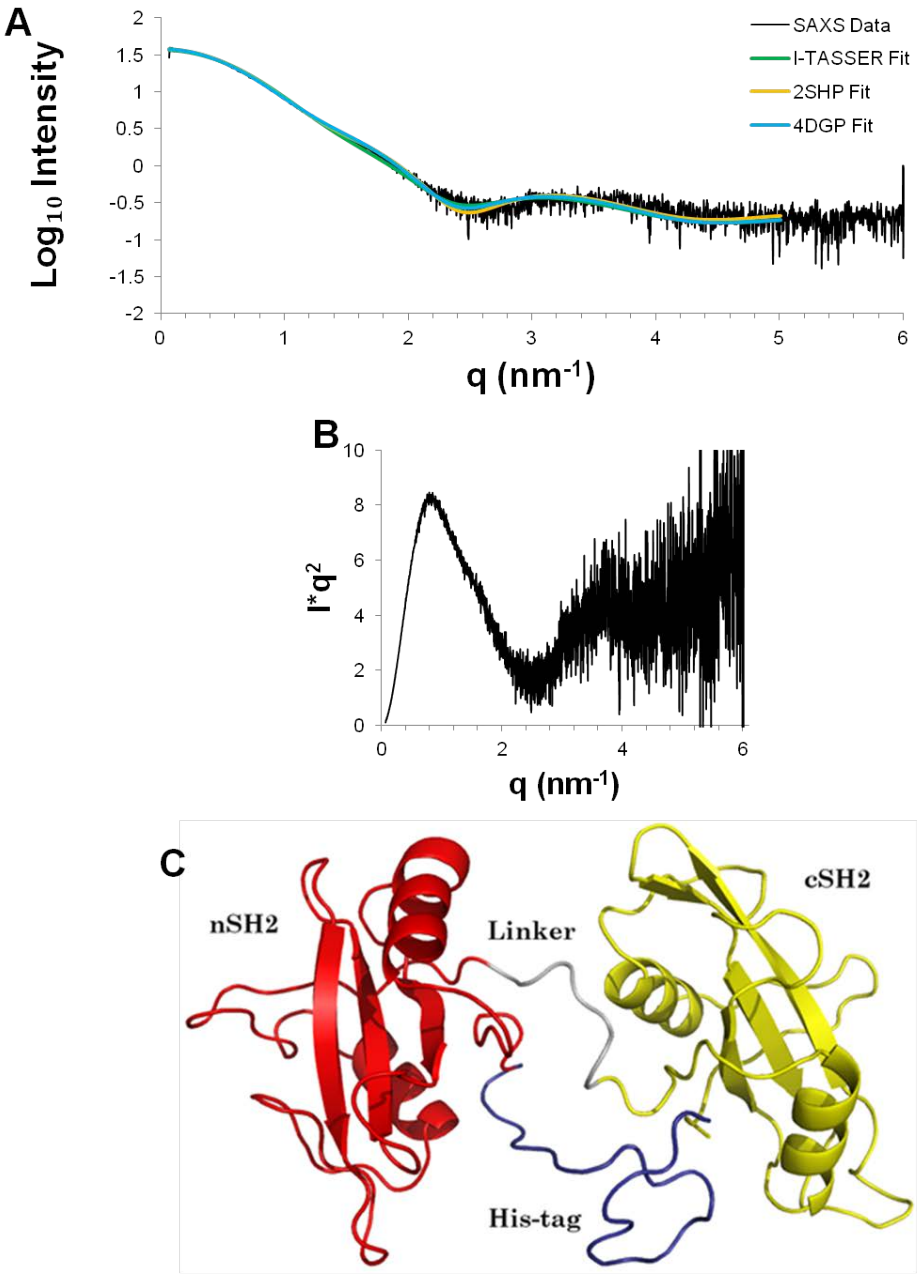


Figure 5.2: Model Fitting of His-TanSH2 Shp2 SAXS Data.

(A) SAXS data of His-TanSH2-Shp2 (black) and CRYSOL simulated curves of: I-TASSER His-TanSH2-Shp2 (χ^2=1.53, green), TanSH2-Shp2 domain of 2SHP (χ^2=1.002, yellow) and 4DGP (χ^2 =0.962, light blue). **(B)** Kratky plot of His-TanSH2-Shp2 SAXS Data, highlighting the presence of a globular domain and flexible region. **(C)** I-TASSER homology model of His-TanSH2-Shp2 (blue: His-tag, red: nSH2, grey: inter-SH2 linker, yellow: cSH2)

CRYSOL Simulation

CRYSOL was used to generate simulated SAXS curves of three sets of coordinates and compare them to the experimental SAXS data acquired of His-TanSH2-Shp2. These are the two crystal structures of tandem SH2 domains (PDB IDs: 2SHP and 4DGP) and the I-TASSER model of His-TanSH2-Shp2. By comparing the overall fit of the simulated SAXS data with the experimental data, as well as inspection of the χ^2 value, one can ascertain the extent of similarity between the two datasets and thus judge similarity between data collected in solution versus crystal conditions.

When compared with the experimental SAXS curve of His-TanSH2-Shp2, a good fit is observed with all three CRYSOL curves (Figure 5.2 A, 2SHP orange, 4DGP light blue and I-TASSER model green), especially at the low-q regions, indicating the overall size and shape match well. The χ^2-values of 2SHP and 4DGP fitted to the experimental data are 1.002 and 0.962, respectively. However, the χ^2-value of the I-TASSER model, which encompasses the entire sequence used in the SAXS study, is 1.53, a worse fit than the two models that do not take the full sequence into account. Visual inspection of the I-TASSER SAXS curve with that of the experimental data shows a slight mis-alignment in the q-range of 1-2 nm^{-1}, and explain the mismatch in χ^2.

CRYSOL calculates curves from static structures and it is possible the lower fit observed for the homology model is due to flexibility in the system not accounted for. The good fit between simulated curves of TanSH2 structures lacking an N-terminal His-tag is surprising, but could be suggestive of His-TanSH2-Shp2 forming structures very similar to those in 2SHP and 4DGP despite the presence of a His-tag. It could

also suggest that the His-tag, being unfolded and only 10% of the entire construct sequence, does not contribute much to X-ray scattering processes overall and instead, the majority of SAXS information derives from the folded domains.

A Kratky plot of the experimental data (Figure 5.2 B) contains a bell shaped curve with a tail region increasing in intensity at higher-q angles, suggesting a mixture of both globular and flexible character of the particle and though qualitative, warrants further investigation.

Ab initio Shape Determination

To investigate the overall shape of His-TanSH2 in solution, *ab initio* shape reconstruction was performed using SAXS data. DAMMIF (Franke & Svergun, 2009a) was used to generate 10 envelopes, all of which had the same overall shape, a rough L-shaped globular particle with a small protrusion at one end (Figure 5.3). Averaging and filtering of these envelopes using DAMAVER and DAMFILT (Volkov & Svergun, 2003) produced a single envelope which was then used in automatic fitting against the I-TASSER model of His-TanSH2 Shp2 via SUPCOMB (Kozin & Svergun 2001), which superimposes any two structures. SUPCOMB gave three optimum orientations, all of which are shown in Figure 5.3, with their corresponding NSD values to the far right. NSD allows one to assess the fit between two 3D models; 0 is a perfect fit and > 1 means the models differ systematically from one another. B(i) is the best fit, with an NSD value of 1.57, though B(ii) and (iii) have comparable NSD values of 1.60 and 1.62. Visual inspection of the superimposed His-TanSH2-Shp2 structure with the DAMMIF-derived envelope demonstrated all three orientations of the tandem SH2 domain fit within the final envelope well, though it is impossible to discern exactly which SH2 domain fits into which spherical portion. Taking B(i) as the reference, B(ii) differs in that the I-TASSER model is rotated about its vertical axis 180°, so that the SH2 domain incorporated into one of two spherical portions is the obverse to the previous fit. B(iii) differs from B(ii) in that there is an 180° rotation about the horizontal axis. Presumably a 4^{th} orientation may involve 180° rotation about the horizontal axis for B(i). Overall, all orientations fit well visually and corroborate the fits of the CRYSOL simulated curves. However, the χ^2 and NSD values between the SAXS data and homology models and crystal structures suggest

that a single rigid model may not be adequate to describe what is happening in solution.

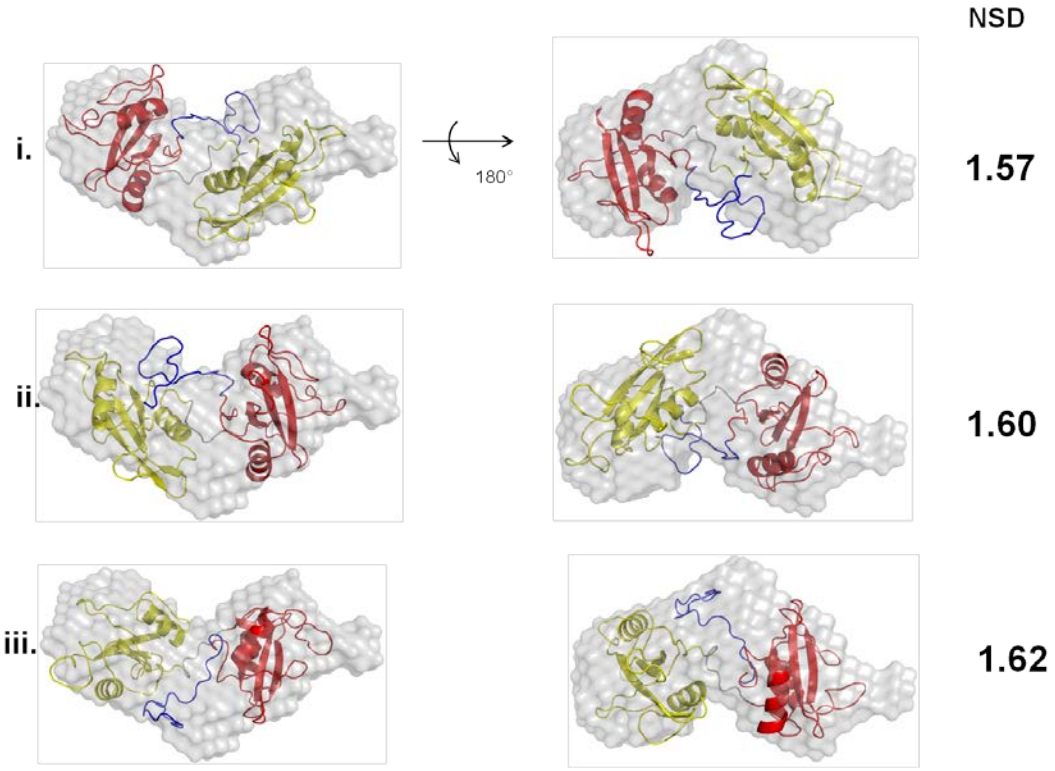


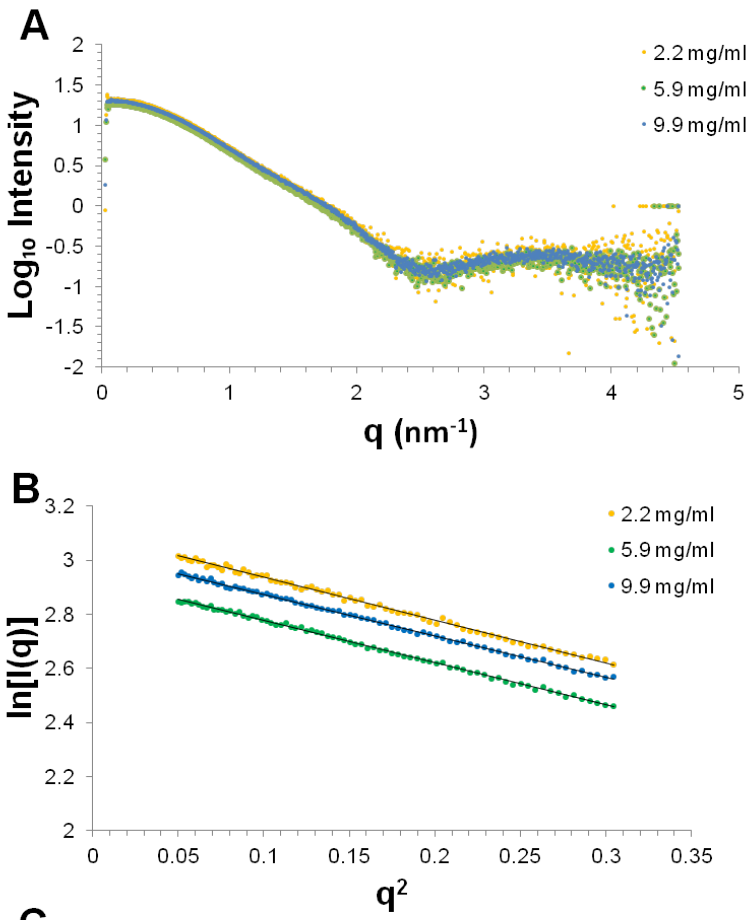
Figure 5.3: *Ab initio* Modelling of His-TanSH2-Shp2.

Averaged and filtered *ab initio* bead model of His-TanSH2-Shp2 with top three (i.-iii.) automated fits of I-TASSER homology model (grey: surface representation of *ab initio* SAXS bead model, blue: His-tag, red: nSH2, grey loop: inter-SH2 linker, yellow: cSH2). Far right: NSD values of automatic fits generated by SUPCOMB, 0 being an ideal fit.

ΔHis-TanSH2 Shp2

The N-terminus of TanSH2-Shp2 is situated in between both SH2 domains. It was reasoned the presence of a His-tag may interfere with the interdomain interface. To address this question, TanSH2-Shp2 was TEV protease treated and SAXS datawere collected at 2.2, 5.9 and 9.9 mg/ml concentrations and summarised in Figure 5.4 A. The difference in data quality is likely not protein specific but due to data collection at two different synchrotron sources.

The Guinier plots of the various concentrations (Figure 5.4 B) in the 0.05-0.3 q^2 region indicate the absence of aggregation and any obvious interparticle interference, though the 9.9 mg/ml dataset slopes less than the other two concentrations tested, suggesting slight interparticle effects with the latter two. Inspection of the parameters calculated from the data using AUTORG and AUTOGNOM shows minor differences in R_g , I(0) and D_{max}. The I(0) is proportional to the molecular weight of the scattering particle but requires calibration against a particle of known molecular weight. Using BSA as a standard the calculated MM is 20.8 kDa; an MM of 24.2 kDa is expected, a 14% error. Owing to high data quality at low-q and the absence of obvious interparticle effects, the 9.9 mg/ml dataset was chosen for further analysis.



C

Concentration (mg/ml)	R_g (nm)	I(0)	D_{max} (nm)	Porod Volume (nm^3)
2.2	2.177 ± 0.006	22.070 ± 0.037	7.62	41.26
5.9	2.165 ± 0.006	18.803 ± 0.006	7.58	36.86
9.9	2.143 ± 0.008	20.637 ± 0.023	7.5	33.67

Figure 5.4: ΔHis-TanSH2 Shp2 SAXS Data.

(A) Buffer Subtracted ΔHis-TanSH2-Shp2 SAXS Data. The low-angle data points are similar throughout all concentration tested, whereas the wide-angle data points increase in quality with concentration. **(B)** Guinier plot and fit of all three concentrations tested. Slopes indicate all three samples are free of interparticle interference and aggregation, and data from 4 and 8 mg/ml are of high quality. These two data sets were merged for downstream analysis. **(C)** Table of calculated parameters. Porod volume and I(0) estimates display concentration dependence, suggestive of interparticle effects.

As stated in the previous section, comparison of the experimental curve with simulated curves from available crystal structures can give insights into the differences in shape and overall fold between the two states. The same crystal structure coordinates for the tandem SH2 domain of Shp2 that were used in the analysis of His-TanSH2-Shp2 SAXS data were used here and were derived from PDB IDs 2SHP and 4DGP. In addition, a homology model best representing the TEV protease cleaved primary sequence of this TanSH2-Shp2 was generated with I-TASSER. The main difference between the homology model and the crystal structure of TanSH2-Shp2 is in the N-terminal region, where the N-terminal 3S (according to WT sequence) in the latter is removed and residues S and M are modelled (coloured blue in Figure 5.5 C). Comparisons of the CRYSOL simulated curves for these structures and the experimental SAXS data of ΔHis-TanSH2-Shp2 are presented in Figure 5.5 A. Visual inspection of the curves show very good fitting in the low-q region with deviations appearing from q=1.5 nm^{-1} onward for all three simulated curves. χ^2 fits suggest 2SHP has the best fit (χ^2=2.071), then 4DGP (χ^2=2.297) and the I-TASSER model (χ^2=2.611). The simulated curve of 4DGP has the worst fit at the q region of 2.5 nm^{-1}, whereas the I-TASSER model has the worst fit at the high-q region (3 nm^{-1}onwards).

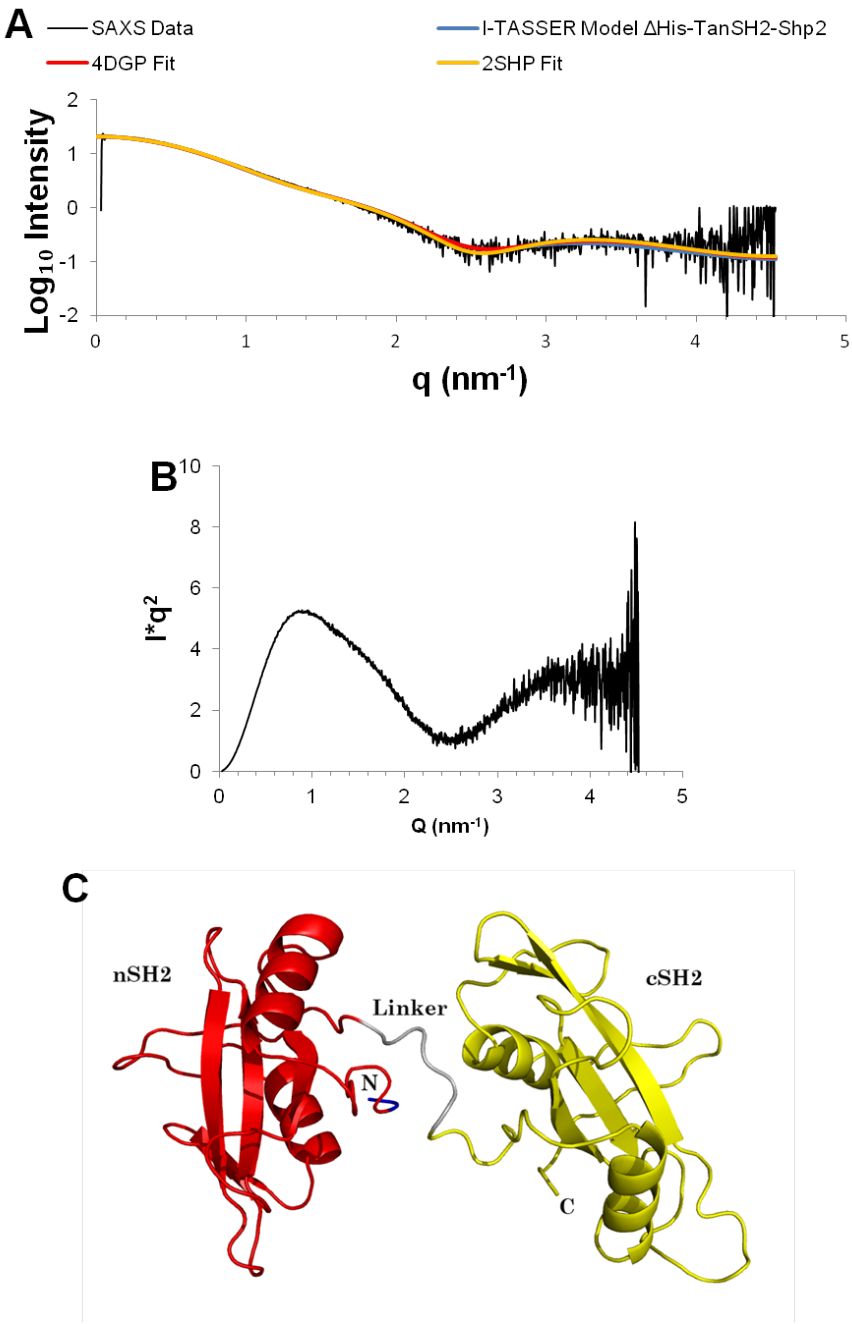


Figure 5.5: Model Fitting of ΔHis-TanSH2 Shp2 SAXS Data.

(A) SAXS data of ΔHis-TanSH2-Shp2 (black) and CRYSOL simulated curves of: I-TASSER His-TanSH2-Shp2 (χ^2=2.969, blue), TanSH2-Shp2 domain of 2SHP (χ^2=2.071, yellow) and 4DGP (χ^2 =2.297, red). **(B)** Kratky plot of ΔHis-TanSH2-Shp2 SAXS Data, highlighting the presence of a globular domain and flexible region. **(C)** I-TASSER model of ΔHis-TanSH2-Shp2. Domains are coloured as follows: Non-wild type residues: blue, nSH2: red, inter-SH2 linker: grey, cSH2: yellow. N and C represent n-terminus and C-terminus.

Ab initio shape determination for the ΔHis-TanSH2 Shp2

In order to determine the *ab initio* shapes of ΔHis-TanSH2-Shp2 based on the SAXS data, DAMMIF was utilised. 10 models were generated, the superimposition of which reveals a high similarity. These envelopes were averaged and filtered as previously described, producing a single dumbbell-shaped envelope. Figure 5.6 shows the final envelope automatically fitted against the ΔHis-TanSH2 Shp2 I-TASSER model, displaying the three orientations with the lowest NSD values. The best fit, with an NSD-value of 1.36, seems to be the optimum superimposition by visual inspection. In comparison, the other two orientations produced NSD-values of 1.46 and 1.60, rotated in the same fashion as the models in Figure 5.6.

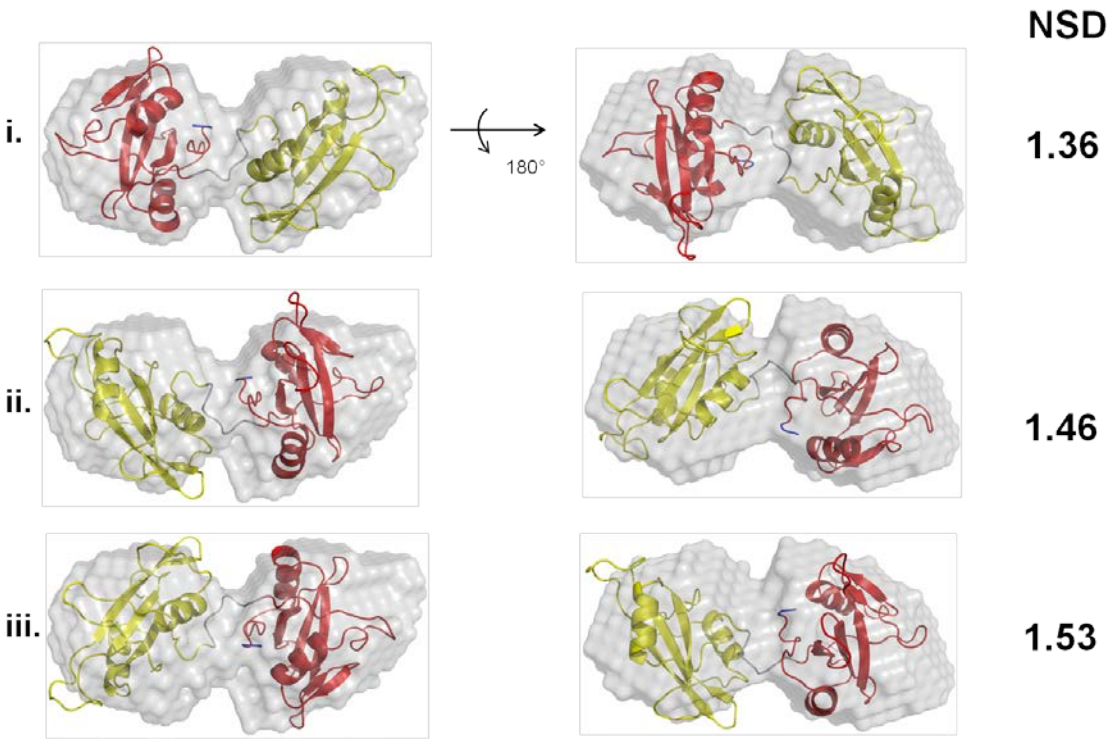


Figure 5.6: *Ab initio* Modelling of ΔHis-TanSH2-Shp2.

Averaged and filtered *ab initio* bead model of His-TanSH2-Shp2 with top three automated fits of I-TASSER homology model (grey: surface representation of *ab initio* SAXS bead model, blue: His-tag, red: nSH2, grey loop: inter-SH2 linker, yellow: cSH2). Far right: NSD values of automatic fits generated by SUPCOMB, 0 being an ideal fit.

Hints at TanSH2 Flexibility in Solution

Qualitative assessment of a Kratky plot of the experimental data (Figure 5.5 B) suggests the presence of both globular and flexible domains as there is a bell shaped curve at the beginning corresponding to the former and a rising tail at higher q that eventually plateaus corresponding to the latter (Doniach, 2001). Given the poor fits between crystal structures, homology models and the SAXS data for ΔHis-TanSH2-Shp2 and the Kratky plot suggest flexible regions, it is possible that a single rigid model cannot fully explain the SAXS data, which gives rise to the poor model fitting observed. In order to full explain the SAXS data for ΔHis-TanSH2-Shp2, flexibility must be accounted for.

EOM Analysis of ΔHis-TanSH2-Shp2

EOM analysis of ΔHis-TanSH2-Shp2 was performed, allowing flexibility in the 8 residue interSH2-linker region, N-terminal S-M residues left over post-TEV protease cleavage and the CD loop of cSH2-Shp2. Results are presented in Figure 5.7. The fit is excellent, with a chi-squared value of 1.102, better than the values obtained for CRYSOL back calculated curves. Inspection of the distributions of R_g and D_{max} reveal that, based on EOM, ΔHis-TanSH2-Shp2 takes on a broad range of states spanning extended and compact. Flexibility of the CD loop and interSH2 linker yielded an excellent fit.

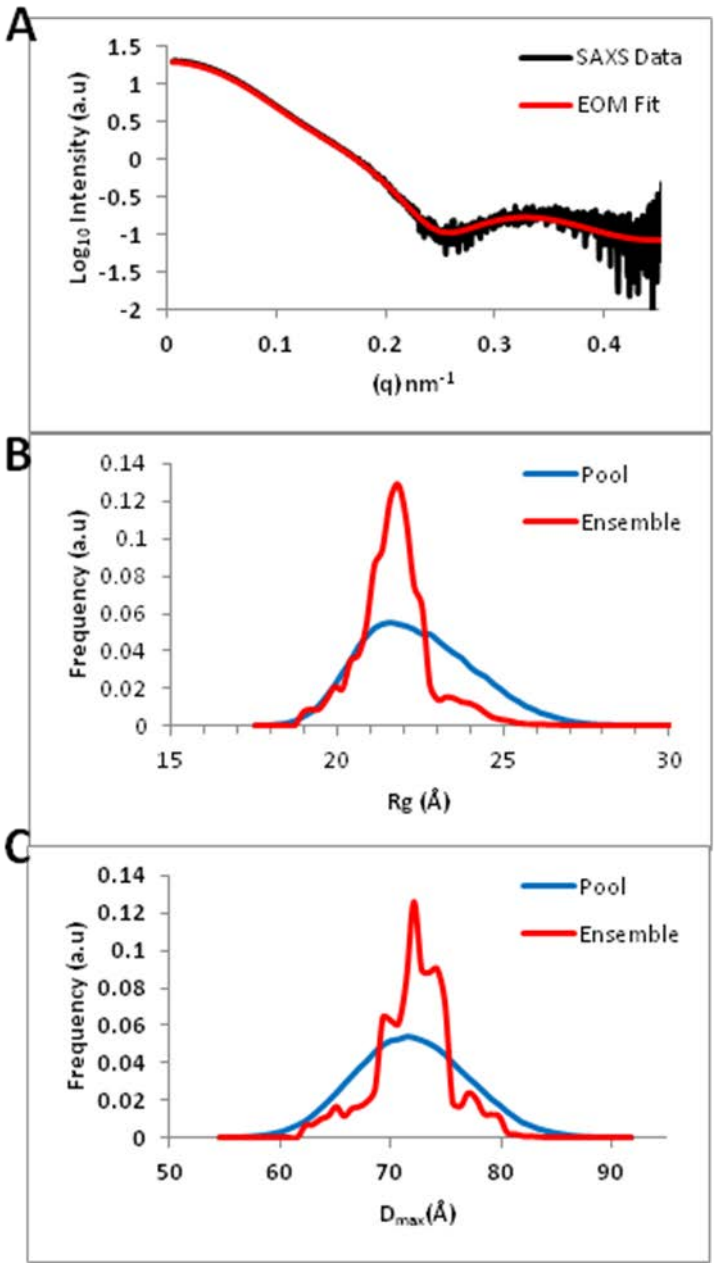
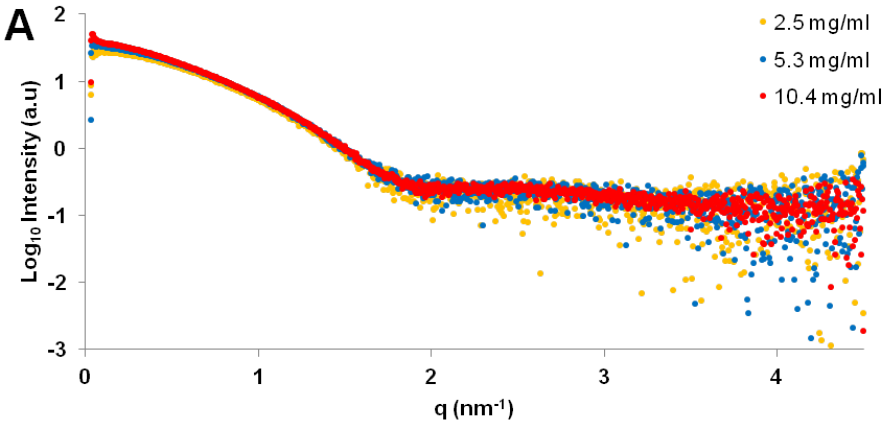


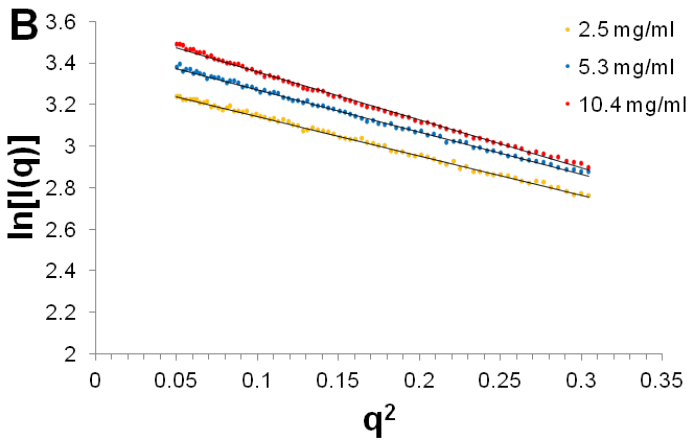
Figure 5.7: EOM Analysis of ΔHis-TanSH2-Shp2.

Fit of ΔHis-TanSH2-Shp2 SAXS data (black) against averaged curve of flexible ensemble selected by EOM (red). X^2=1.102. **(B)** and **(C)** $R_{g\ and}$ D_{max} distributions respectively, of optimized ensemble from EOM. Both indicate ΔHis-TanSH2-Shp2 samples a range of compact and extended states in solution.

SAXS Analysis of the Catalytic domain of Shp2

SAXS of PTP-Shp2 was performed at ESRF at 2.5, 5.3 and 10.4 mg/ml concentrations. (Figure 5.8) The individual buffer subtracted scattering curves are shown in Figure 5.8. Inspection of the Guinier plots show a downward-turn of the linear fit at 10.4 mg/ml, with respect to other concentrations. This indicates aggregation of the sample and should not be used for subsequent analyses, further validated by inspection of the invariant parameters which all increase with concentration.





C

Concentration (mg/ml)	R_g (nm)	I(0)	D_{max} (nm)	Porod Volume (nm^3)
2.5	2.402 ± 0.011	28.109 ± 0.043	8.41	67.11
5.3	2.533 ± 0.032	32.671 ± 0.059	8.87	72.78
10.4	2.614 ± 0.144	35.683 ± 0.103	9.15	75.69
Merged	2.401 ± 0.009	31.592 ± 0.032	8.4	68.58

Figure 5.8: PTP-Shp2 SAXS Data.

(A) Buffer Subtracted PTP-Shp2 SAXS Data. The low-angle data points are similar throughout all concentration tested, whereas the wide-angle data points increase in quality with concentration. **(B)** Guinier plot and fit of all three concentrations tested. Slopes seem to indicate highest concentration is aggregated, therefore data derived from low-q unreliable for further analysis. Data from 5.3 and 2.5 mg/ml are of high quality. These two data sets were merged. **(C)** Table of calculated parameters, in which concentration-dependent increases are observed with increasing concentration.

Merging of the 2.5 and 5.3 mg/ml datasets generated a SAXS curve free from the effects of aggregation (Figure 5.9 A) and similar R_g, D_{max} and Porod volume values to the lowest concentration, assumed to be the least susceptible to aggregation. Interestingly, qualitative inspection of the $q^2 > 3$ region of the Kratky plot of the merged SAXS data (Figure 5.9 B) suggests a degree of flexibility within the PTP-Shp2 particle.

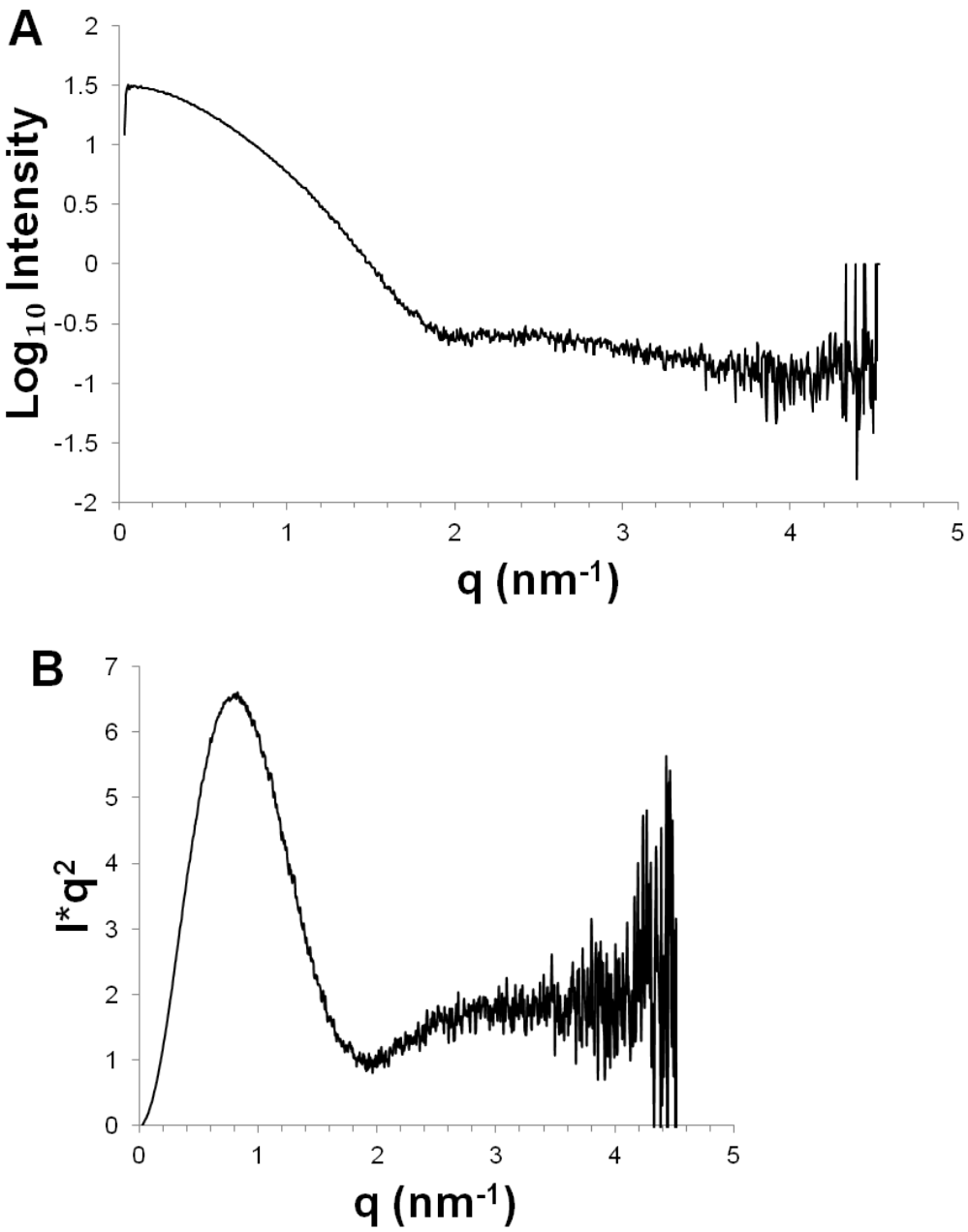


Figure 5.9: Merged SAXS Data of PTP-Shp2.

(A) Merged SAXS data of PTP-Shp2 used for further analyses. **(B)** Kratky plot of merged SAXS data of PTP-Shp2, indicating a globular particle with a degree of flexibility.

Comparison of the merged SAXS data with existing structural models may yield insights into the solution state of the catalytic domain. CRYSOL was used to generate simulated curves of the catalytic domain of Shp2 with an accessible active site (PDB ID: 3B7O), as well as a simulated curve of full length Shp2 (PDB ID: 4DGP) lacking N-terminal residues 3-258 (hereby referred to as 4DGP-PTP) (Figure 5.10). Superimposition and visual inspection of the simulated curves with the merged SAXS data reveals a good fit at the low q region (0.06-1 nm) but gradually worsening from q=1.5 nm onwards. Also, the χ^2 values between 4DGP-PTP or 3B7O and the SAXS data are 3.503 and 3.373 respectively. This suggests that there are differences in the global shape of PTP-Shp2 in the solution state compared to the crystal structures. Indeed, the theoretical R_g values calculated from 3B7O and 4DGP-PTP are 2.11 and 2.12 nm respectively, with the experimental R_g being 2.401 ± 0.009 nm suggesting that in solution PTP-Shp2 is less compact than the available crystal structures tested.

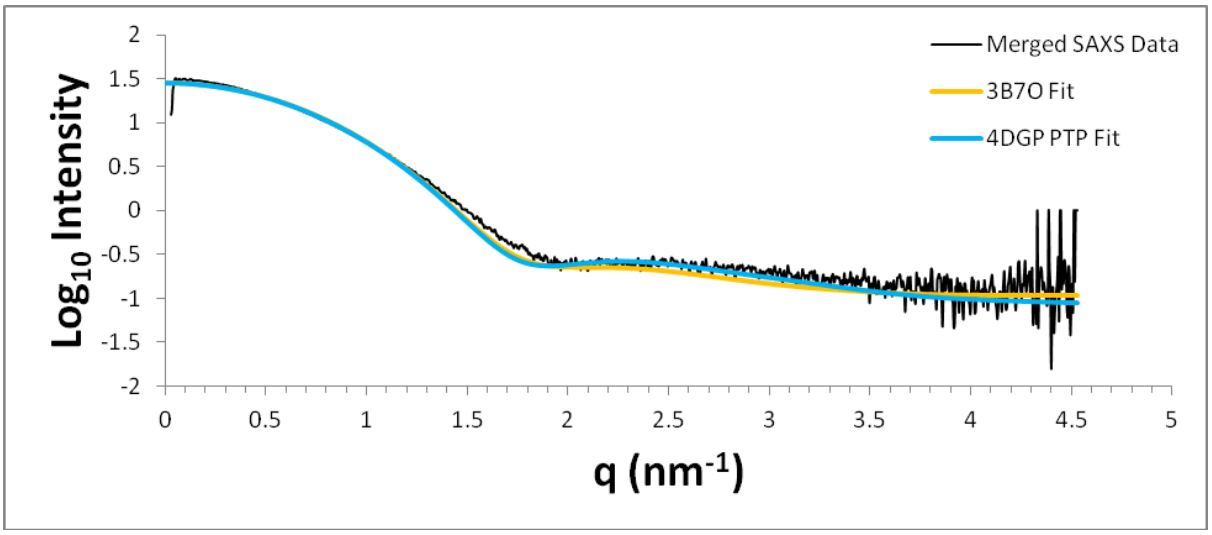


Figure 5.10: Model Fitting PTP-Shp2 SAXS Data.

Merged SAXS curve of PTP-Shp2 (black) was fitted against CRYSOL simulated curves of the catalytic domain of Shp2 with accessible active site (Yellow, χ^2 =3.503, PDB ID: 3B7O) and catalytic domain of Shp2 lacking crystal coordinates between residues 3-258, dubbed 4DGP-PTP in main text (Blue, χ^2 =3.373, PDB ID: 4DGP).

Differences between the simulated intensities and R_g values derived from the experimental and 3B7O simulated curves may be attributed to the complete absence of crystal coordinates for three regions: the 11 N-terminal residues (SMAETTDKVKQ; underlined are non-WT residues), a loop region on the opposite side of the catalytic site separating two beta-sheets (314-324, FETKCNNSKPK) and the extreme C-terminal residue, an isoleucine. Since these regions absent, they are not simulated by CRYSOL and therefore do not contribute to the final simulated curve, a discrepancy should be expected. The 4DGP-PTP crystal structure, though missing the C-terminal isoleucine, encompasses the loop regions from 314-324 as well as the aforementioned N-terminal residues (apart from the non-WT SM residues). One major structural difference between 3B7O and 4DGP-PTP concerns the N-terminal helix of the PTP domain, which is shifted around 50° away (Figure 5.11) from the rest of the domain in the former, but is packed close in the latter, likely due to the presence of the SH2 domains during crystallisation. The resultant simulated curve fits the experimental data in a similar manner to 3B7O, though there is a better fit at the higher q-region (2 nm^{-1} onwards). χ^2-values for both 3B7O and 4DGP-PTP are 3.50 and 3.37, indicating the latter structure fits better. However, a χ^2-value of 3.37 is a relatively poor fit and demonstrates that the back-calculated envelopes produced by CRYSOL do not fit the experimental data and thus current structural models cannot explain the SAXS data well, suggesting that in solution PTP-Shp2 adopts a structure somewhat different to that observed in the crystal structures.

One method that could resolve the missing loop regions of 3B7O is rigid body modelling, where available atomic models and experimental SAXS data are provided to the program BUNCH, part of the ATSAS package (Petoukhov & Svergun, 2005).

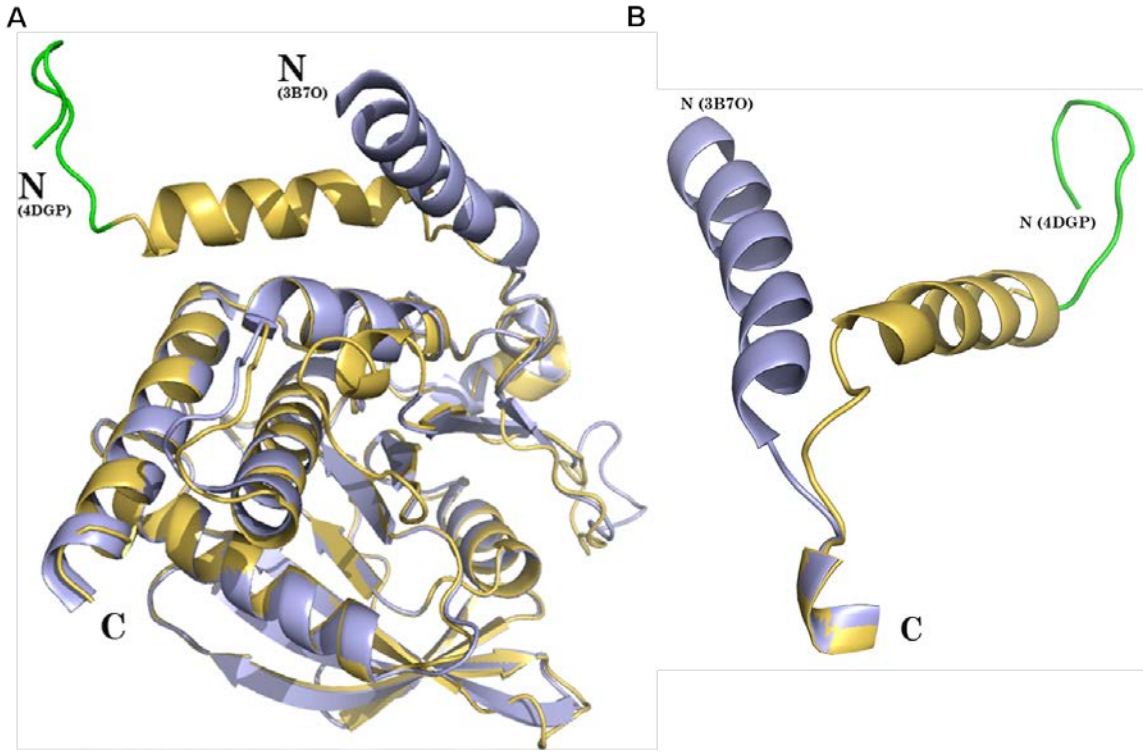


Figure 5.11: Superimposition of Shp2 PTP Domain Crystal Structures.

(A) Superimposition of crystal structures of the PTP domain ONLY (lilac, PDB ID: 3B7O) and PTP domain crystallised as part of a near full-length protein (yellow, PDB ID: 4DGP). The WT N-terminal unstructured loop of 4DGP (green) was included in the 3B7O construct during crystallisation but was not modelled in the final structure of the latter. N: N-terminus, C: C-terminus. **(B)** Zoomed and reoriented region of (A) highlighting the stark differences in N-terminal helix position in both crystal structures.

When provided with these data, BUNCH models missing residues of the model to be fitted in as ‘dummy residues’ which act as a placeholder for the amino acid residue. Steric clashes and improper arrangements are penalized, as would occur in a structure calculation. Placement of the dummy residues and comparison of simulated curves of the resultant model against the experimental data are performed until convergence on a model with the least penalties and lowest χ^2 is created.

BUNCH was run with 3B7O as the starting model, the merged SAXS curve as the experimental data and the missing loop regions defined beforehand. The atomic model was kept rigid throughout each successive BUNCH run, which yielded several models with loops varying in their relative positions to one another but with the rest of the PTP domain holding the same structure. Figure 5.12 A shows a representative BUNCH model with a χ^2-value of 2.470; the corresponding fit is shown in Figure 5.12 B. Seven independent BUNCH models were generated, with an average χ^2-value of 2.433, all of which are superimposed in Figure 5.13, highlighting the relatively high mobility of the loops in question. The mobility of the BUNCH-derived dummy residues is corroborated by the B-factors observed in the crystal structure of the residues adjacent to them, highlighted by the red regions on 3B7O in Figure 5.12 C. Though this is not definitive, this does explain the multiple loop conformations calculated by BUNCH, as well as the absence of definitive density of these regions in the 3B7O crystal structure, which is usually assumed to be due to high degrees of disorder.

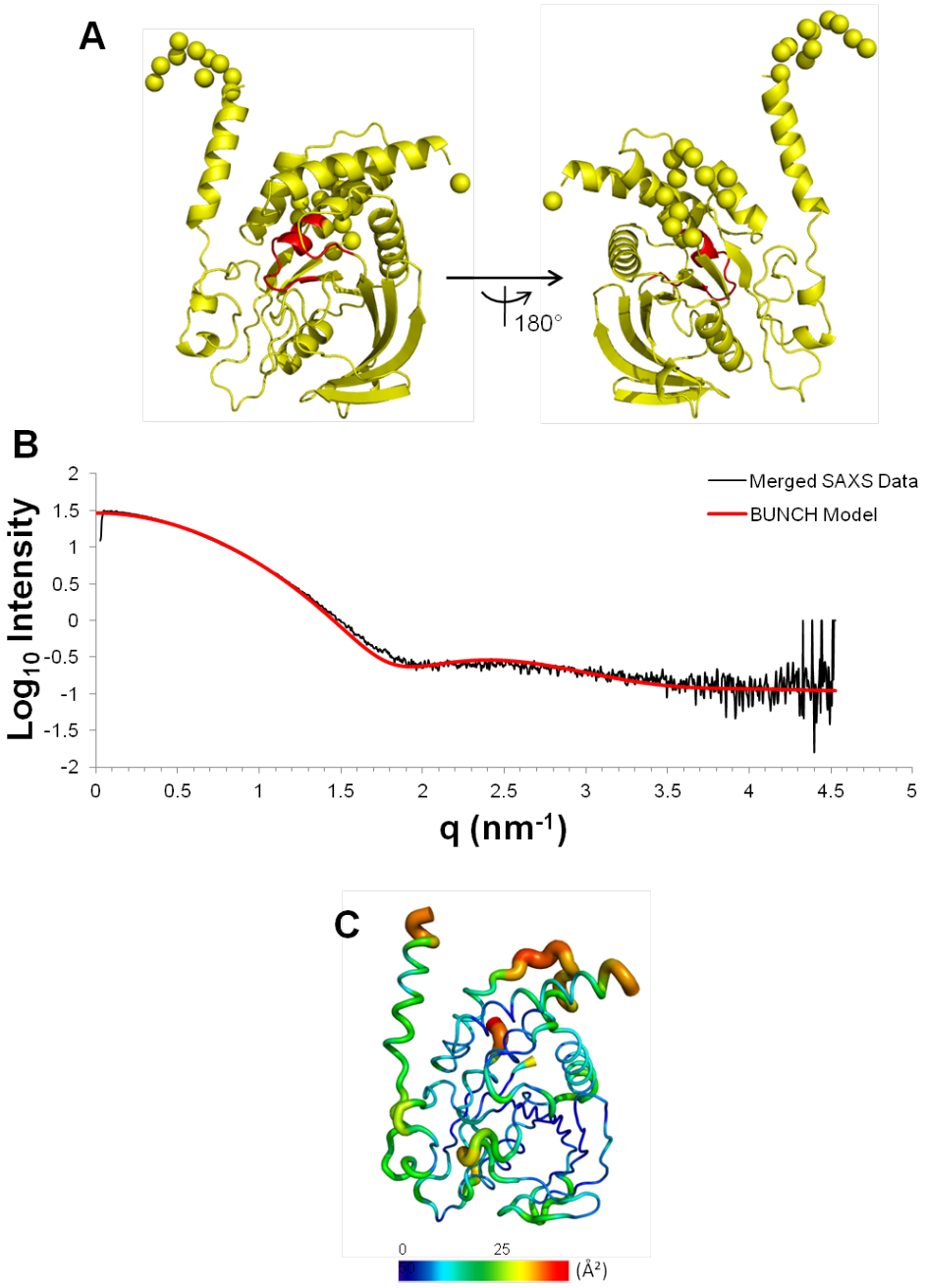


Figure 5.12: Modelling of Missing Loops into PTP-Shp2 structure using BUNCH.

(A) Ribbon diagram of PTP-Shp2 crystal structure (PDB ID: 3B7O). Each sphere represents a missing residue modelled using experimental SAXS data and BUNCH. Red areas indicate residues important for catalysis. **(B)** Merged SAXS data of PTP-Shp2 (black) were fitted against CRYSOL simulated curve of representative BUNCH model (red) in (A) yielding X^2 =2.470. **(C)** Crystallographic temperature factors for 3B7O coloured on cartoon putty diagram of structure for low (dark blue-cyan), medium (green-yellow) and high (orange-red) areas of disorder for PTP-Shp2. Structure oriented as in (A).

Figure 5.13: The Missing Loops of PTP-Shp2 May Be Flexible.

PTP-Shp2 cartoon diagram of multiple aligned BUNCH models. The two modelled loop regions are indicated by asterisks and are depicted in various colours to highlight flexible nature.

Interestingly, two BUNCH simulations where domains either side of the modelled loop were allowed to freely rotate and move with the loop yielded models with χ^2-values of 1.119 and 1.046 (Figure 5.14). Though the structures generated differ substantially from the original crystal structure, an excellent fit between the experimental and simulated data curve (Figure 5.14 A) are observed between q-regions 0.05-1.5 nm, with a reasonable fit thereafter in the high-q regions. The flexible models generated by BUNCH (Figure B and C) show a disrupted active site in that the KNRY-loop containing region is now positioned far from the catalytic site. The flexible models are not accurate representatives of the solution state since activity would be severely impaired, which is not the case when PTP-Shp2 is assayed for enzyme activity. However, the fits between the highly flexible model and the experimental data are much better than CRYSOL fits of the crystallographic data and SAXS data, suggesting that there may be flexibility elsewhere in PTP-Shp2 that has not been accounted for.

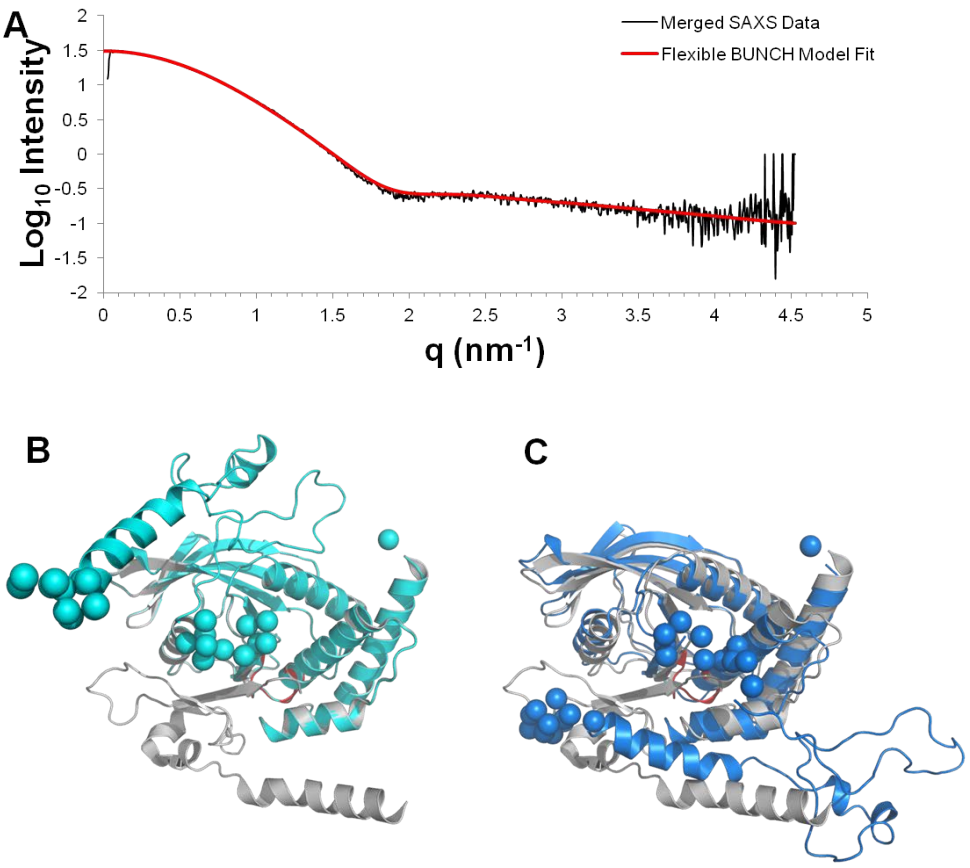


Figure 5.14: Flexible Models of PTP-Shp2 Fit the SAXS Data.

(A) Overlay of merged SAXS data of PTP-Shp2 with CRYSOL simulated curve of best-fitting model generated by BUNCH (red, χ^2= 1.046). **(B)** and **(C)**: Cartoon representations of flexible models generated by BUNCH with original 3B7O PTP-Shp2 structure aligned (grey). Spheres represent modelled residue missing from 3B7O crystal structure. χ^2 values for (B) and (C) are 1.119 and 1.046 respectively. Red region indicates PTP signature motif involved in catalysis.

Ab initio shape determination of PTP-Shp2

DAMMIF was used to generate 10 envelopes (Figure 5.15). Visual inspection of the envelopes indicates two populations. Seven have the general architecture of an oblate ellipsoid with a 42 Å extension protruding almost perpendicular from the surface, hereby referred to as DAM-env-1. The other three resemble an oblate ellipsoid of similar dimensions as the previous seven models but with an extension of 25 Å protruding from one end of the major axis, hereby referred to as DAM-env-2. Interestingly, this extension is globular in nature but with an apparent centre of decreased bead density. Averaging and dummy atom outlier removal of the 10 envelopes using the programs DAMAVER and DAMFILT (Volkov and Svergun, 2003) yields an ellipsoid measuring 70 Å on it's major axis and between 35- 40 Å on its minor axis. Automatic superimposition of the DAMMIF envelope with the original 3B7O crystal structure using SUPCOMB yields an NSD value of 1.718, which suggests a reasonable similarity between the two. However, automatic superimposition of the rigid BUNCH model of PTP-Shp2 with the same DAMMIF envelope leads to a higher NSD value of 2.027, suggesting the addition of the missing loop regions (and flexibility) leads to a worse fit.

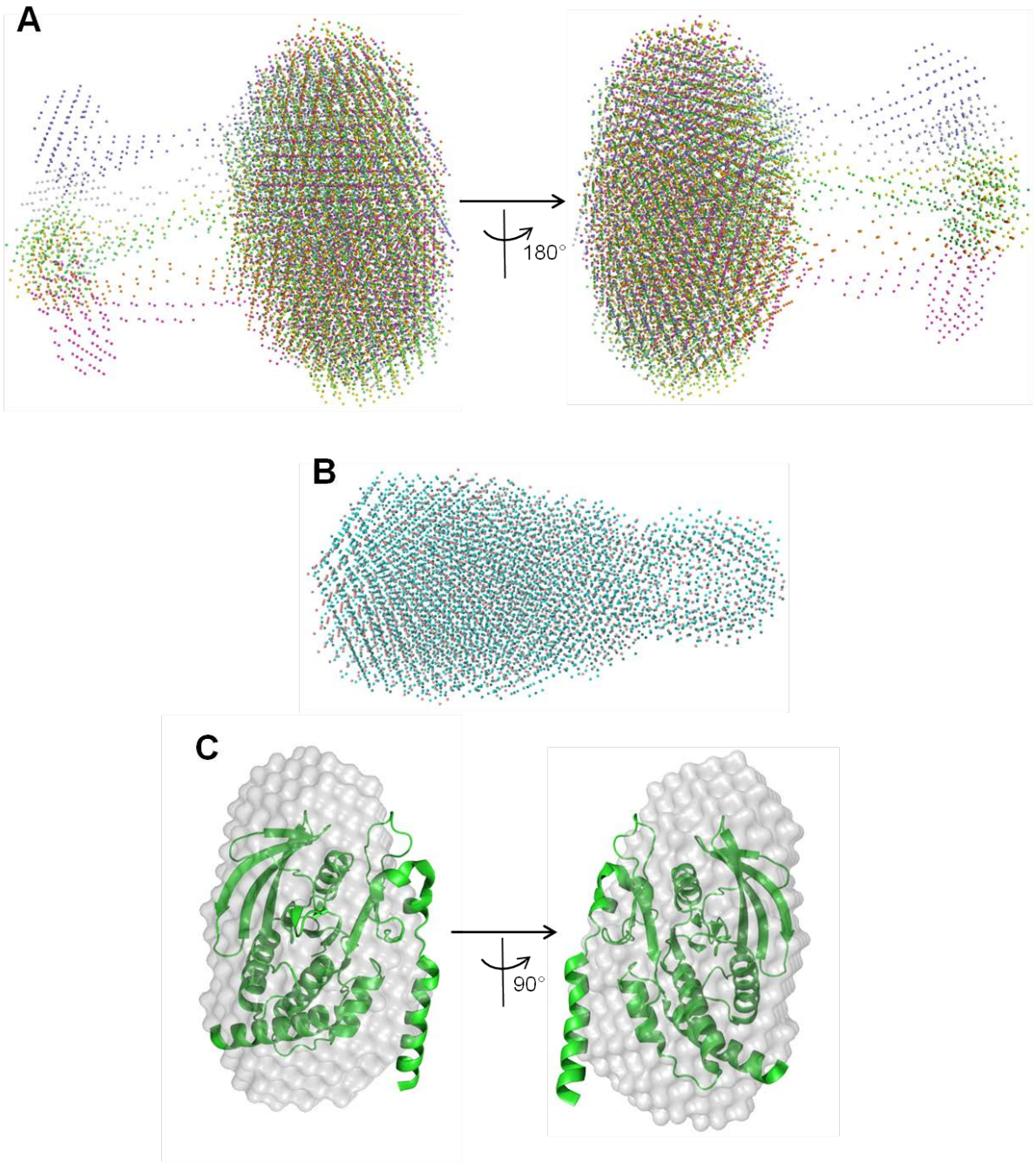


Figure 5.15: *Ab Initio* Models of PTP-Shp2 generated by DAMMIF.

Bead model representation of PTP-Shp2 envelopes generated by DAMMIN. Each colour represents one model. **(A)** contains 1st pool of superimposed models (seven total). **(B)** contains 2nd pool of superimposed models (three total). **(C)** Models averaged and filtered using DAMAVER and DAMFILT respectively to yield final envelope. PTP-Shp2 crystal structure (Green cartoon, PDB ID: 3B7O) automatically fitted with envelope using SUPCOMB; NSD= 1.718.

Fitting of the BUNCH model with DAM-env-1 and 2 leads to NSD values of 2.20 and 2.08 respectively, both of which are reasonable fits, comparable to the averaged DAMMIF envelope fitted against the BUNCH model. Both DAM-env models superimposed with the BUNCH model can be seen in Figure 5.16 and demonstrate that both envelopes can accommodate the BUNCH model well, with no secondary structure elements wildly misplaced outside of the envelope. The distance between the tip of DAM-env-1 and the surface of the ellipsoid is about 44 Å. When this distance is taken from the N-terminal dummy residue in the BUNCH model (Ser), the residues between 42-45 Å away along the polypeptide chain are Tyr263 and Ser264. These two loop regions reside within the loop connecting the N-terminal helix to the next kinked helix; the possibility arises that this region could be responsible for the proposed dynamic nature of the SAXS data. B-factor values for Tyr263 and Ser264 are 23.59 and 22.43 $Å^2$ respectively, which are 17.4 and 11.6% higher than the average across all protein atoms in the structure, 20.09 $Å^2$. It is possible that these residues, or indeed the entire loop, may be flexible in solution in the absence of the tandem SH2 domain and contribute significantly to the flexible nature of PTP-Shp2.

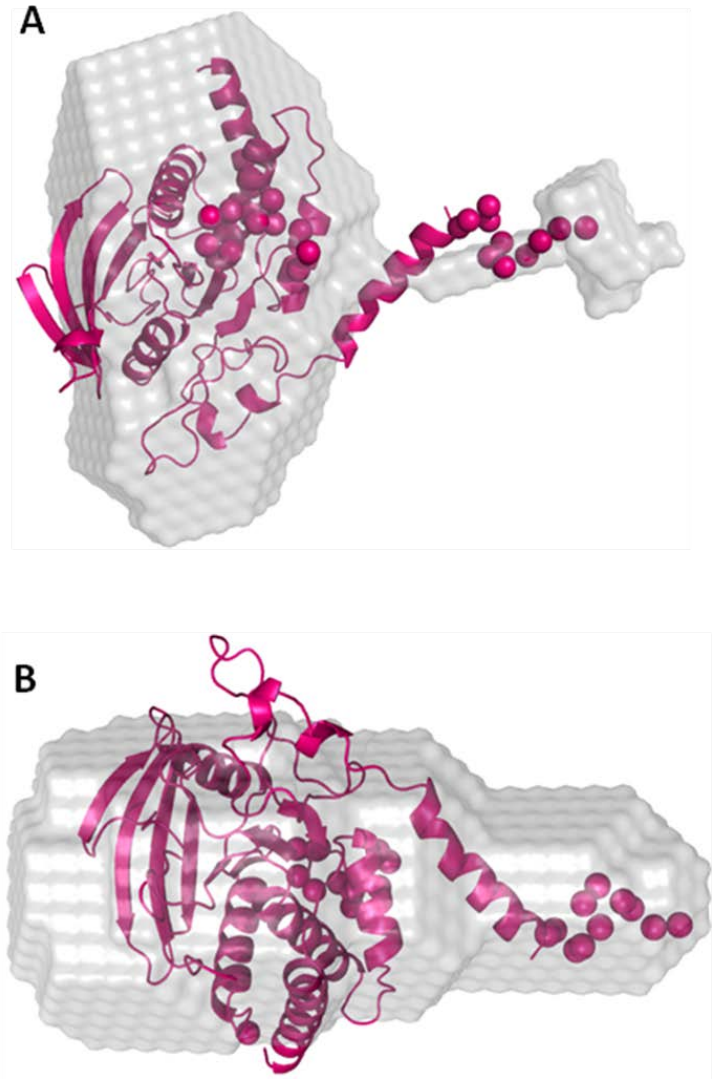


Figure 5.16: BUNCH Model-fitting of PTP-Shp2 to *Ab Initio* Envelopes.

The two distinct species of unaveraged DAMMIN models for PTP-Shp2 were fitted with the previously generated BUNCH rigid body model. **(A)** BUNCH model fit with DAM-env-1 envelope, yielding an NSD of 2.20. **(B)** BUNCH model fit with DAM-env-2 envelope, yielding an NSD of 2.08.

Ensemble Optimisation Method – Analysis of PTP-Shp2

To further assess the flexibility of PTP-Shp2 in solution, the Ensemble Optimization Method (Bernadó *et al*. 2007)was employed via ATSAS online (Petoukhov *et al*. 2007). First, 10,000 rigid structures were generated with flexibility allowed at the N-terminal residues (SMAETTDKVKQ; underlined are non-WT residues) and residues 314-324 using the program RANCH. The merged buffer-subtracted SAXS curve was then used to select a subset of the pool of 10,000 structures that best fit the experimental data, and then calculate R_g and D_{max} distributions for the pool and ensemble. Comparisons of the averaged ensemble fit, R_g and D_{max} distributions may yield more insights into the dynamic behaviour of PTP-Shp2 in solution.

The fit of the averaged curves (Figure 5.17) for the ensemble against the experimental data is poor in the low-q region (<0.6 nm^{-1}), indicating the rigid ensemble yields particles of a smaller R_g than the experimental data suggests. A better fit of the curves is seen thereafter, gradually getting worse in the high-q region. The χ^2 between experimental data and the EOM fit is 2.489, in agreement with the average BUNCH model χ^2 of 2.433. Based on the EOM calculations, a rigid model of PTP-Shp2 in solution does not fit the experimental data, suggesting a more extended, rather than compact set of conformations being taken by the catalytic subunit.

The PTP-Shp2 Domain is Flexible in Solution

A rigid model of PTP-Shp2 did not fit the experimental data in EOM simulations, nor for BUNCH rigid-body modelling. In addition, CRYSOL back-calculated curves of 4DGP-PTP and 3B7O both yielded poor χ^2 values toward the experimental data, and inspection of the theoretical R_g values between the crystallographic data and the experimentally determined R_g found that the latter is larger. Lastly, the Kratky plot of PTP-Shp2 experimental data is suggestive of a partially flexible particle. When taken together, the current evidence point toward PTP-Shp2 being flexible in solution. Flexibility is likely contributed from the N-terminal helix and disordered loop region, as well as highly mobile loop regions nearer the catalytic site of PTP-Shp2. Further optimization of EOM simulations is required for a more accurate understanding of the location(s) of PTP-Shp2 flexibility in solution.

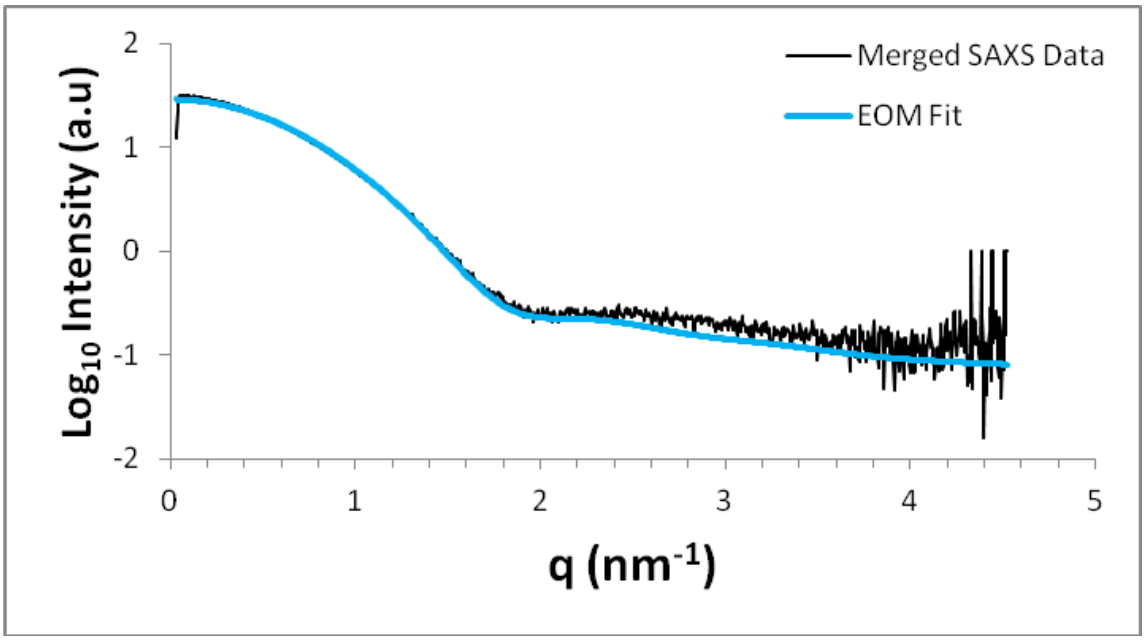


Figure 5.17: Initial EOM Analysis of PTP-Shp2 as a Rigid Model.

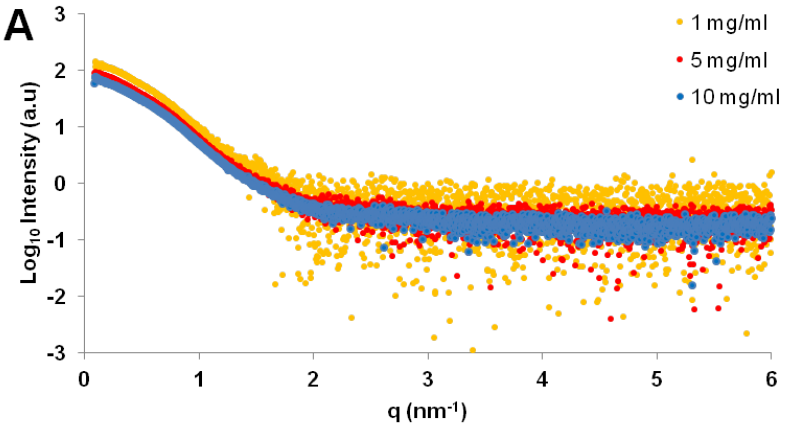
Superimposition of PTP-Shp2 SAXS data (black) with ensemble-averaged simulated curve (blue) generated by EOM assuming a rigid structure, χ^2=2.489, comparable to BUNCH fit (χ^2=2.470). A rigid EOM model does not fit the experimental data well.

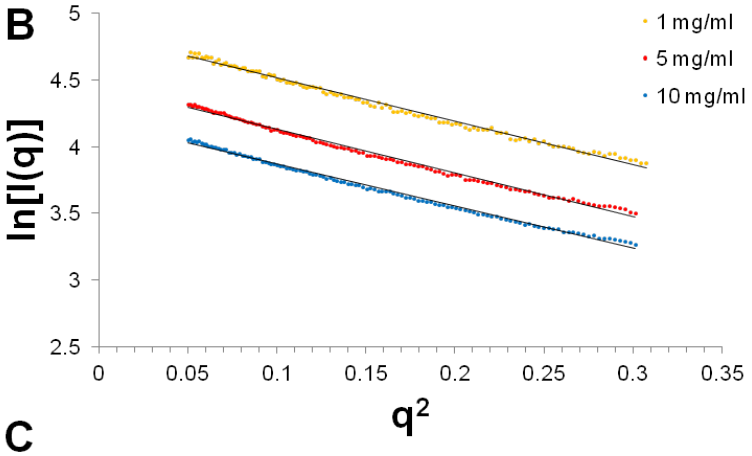
SAXS Analysis of Full length Shp2

Wild-type Shp2

SAXS data of FL-Shp2 WT were collected at 1, 5 and 10 mg/ml and are summarised in Figure 5.18. All three buffer subtracted curves in Figure 5.18 A, when scaled, are near-identical in shape, particularly from the low to mid q-region (0.142-2 nm^{-1}) though slight deviations in intensity are noticed thereafter presumably due to increased signal to noise in the low concentration samples. Calculation and plotting of all three datasets into a Guinier plot (Figure 5.18 B) demonstrate good linearity at q^2 values 0.1-0.25 nm^{-2} but either side of this range the linear fit worsens; this is

dubbed the ‘smiling’ Guinier but is very minute and can suggest aggregation. However, the extremely low q-region in the Log[I(q)] vs. q plot (Figure 5.18 A) (q< 0.14 nm^{-1}) does not reflect this conclusion, no differences were observed between all three concentrations, suggesting no repulsion or aggregation effects were present in the sample. AUTORG and AUTOGNOM did not detect aggregation during parameter calculations. However, some differences in parameters were observed and are summarised in Figure 5.18 C. Extrapolation of the data to infinite dilution generates a curve with comparable R_g and D_{max} values (3.297 ± 0.028 and 10.3 nm respectively) to the 1 mg/ml dataset. The I(0) value is comparable to the 10 mg/ml dataset (71.329 ± 0.265), and the calculated Porod volume is lower than all three datasets, at 105.73 nm^3, approximately 3.2% lower than the 10 mg/ml dataset. This dataset was used for downstream analysis and is summarized in Figure 5.19. For individual SAXS curves of FL-Shp2 WT at all concentrations tested see Appendix figure 6.



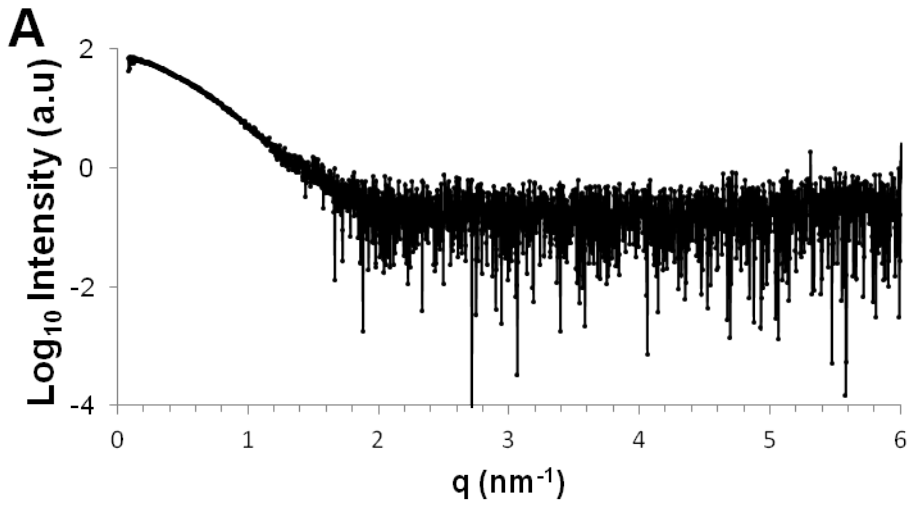


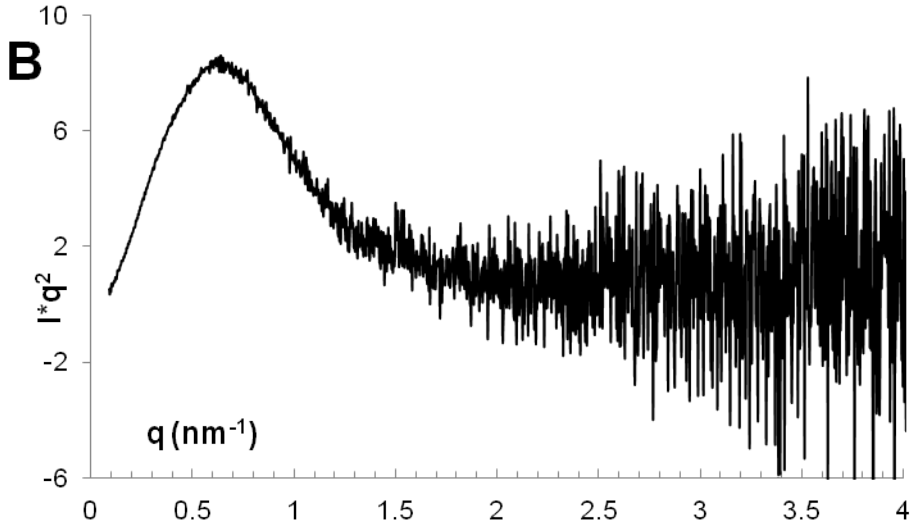
C

Concentration (mg/ml)	R_G (nm)	I(0)	D_{max} (nm)	Porod Volume (nm^3)
1	3.317 ± 0.031	131.994 ± 0.513	11.61	116.91
5	3.408 ± 0.042	91.308 ± 0.719	10.9	112.34
10	3.371 ± 0.014	69.731 ± 0.0173	10.53	109.18

Figure 5.18: SAXS Data of Full-length wild-type Shp2.

(A) Buffer Subtracted Full-length wild-type Shp2 SAXS Data. The low-angle data points are similar throughout all concentration tested, whereas the wide-angle data points decrease in quality with increasing concentration. **(B)** Guinier plot and fit of all three concentrations tested. Slopes seem to indicate all three samples are exhibiting a slight 'smiling Guinier', indicative of aggregation. **(C)** Table of calculated parameters from SAXS data. Note that although the Guinier plot in (B) demonstrates non-linear behaviour indicating aggregation, all calculated parameters decrease as concentration increases, suggesting interparticle interference.





C

Concentration (mg/ml)	R_g (nm)	I(0)	D_{max} (nm)	Porod Volume (nm^3)
Merged Data	3.297 ± 0.028	71.329 ± 0.265	10.3	105.73

Figure 5.19: Merged and Extrapolated SAXS Data of FL-Shp2 WT.

(A) Merged and extrapolated to zero concentration SAXS data of FL-Shp2 WT used for further analyses. **(B)** Kratky plot of SAXS data of FL-Shp2 WT from (A), indicating a globular particle with a degree of flexibility. **(C)** Table of parameters calculated with data from A; also see main text.

Ab initio Modelling of WT Shp2

Information from AUTOGNOM was input into DAMMIF to produce 10 *ab initio* models of FL-Shp2 WT using default parameters. The simulated curves of each model were automatically fit against the experimental curve, allowing for initial visual inspection of the model quality. Fits were deemed satisfactory and inspection of the *ab initio* envelopes reveals a set of related solutions that superimpose well with one another . DAMCLUST, a program that aligns and clusters related shape/atomic models found a mean NSD value of 0.5, indicating a high similarity between all envelopes tested. Averaging and filtering of these models using DAMAVER and DAMFILT led to the envelope used in the next section.

Automatic fitting of this envelope with a Shp-2 crystal structure (PDB ID: 4DGP) encompassing the nSH2, cSH2 and PTP domains using SUPCOMB results in an NSD value of 2.042 for the best solution. Two additional solutions to automatic fitting result in NSD values of 2.095 and 2.142 for solution 2 and 3. All three solutions from SUPCOMB fit the envelope with minimal areas of the Shp-2 structure residing outside this region, as seen in Figure 5.20. As the NSD values are similar for all SUPCOMB solutions and the envelope demonstrates a degree of rotational symmetry, it is impossible to elucidate the orientation of FL-Shp2 subdomains with confidence.

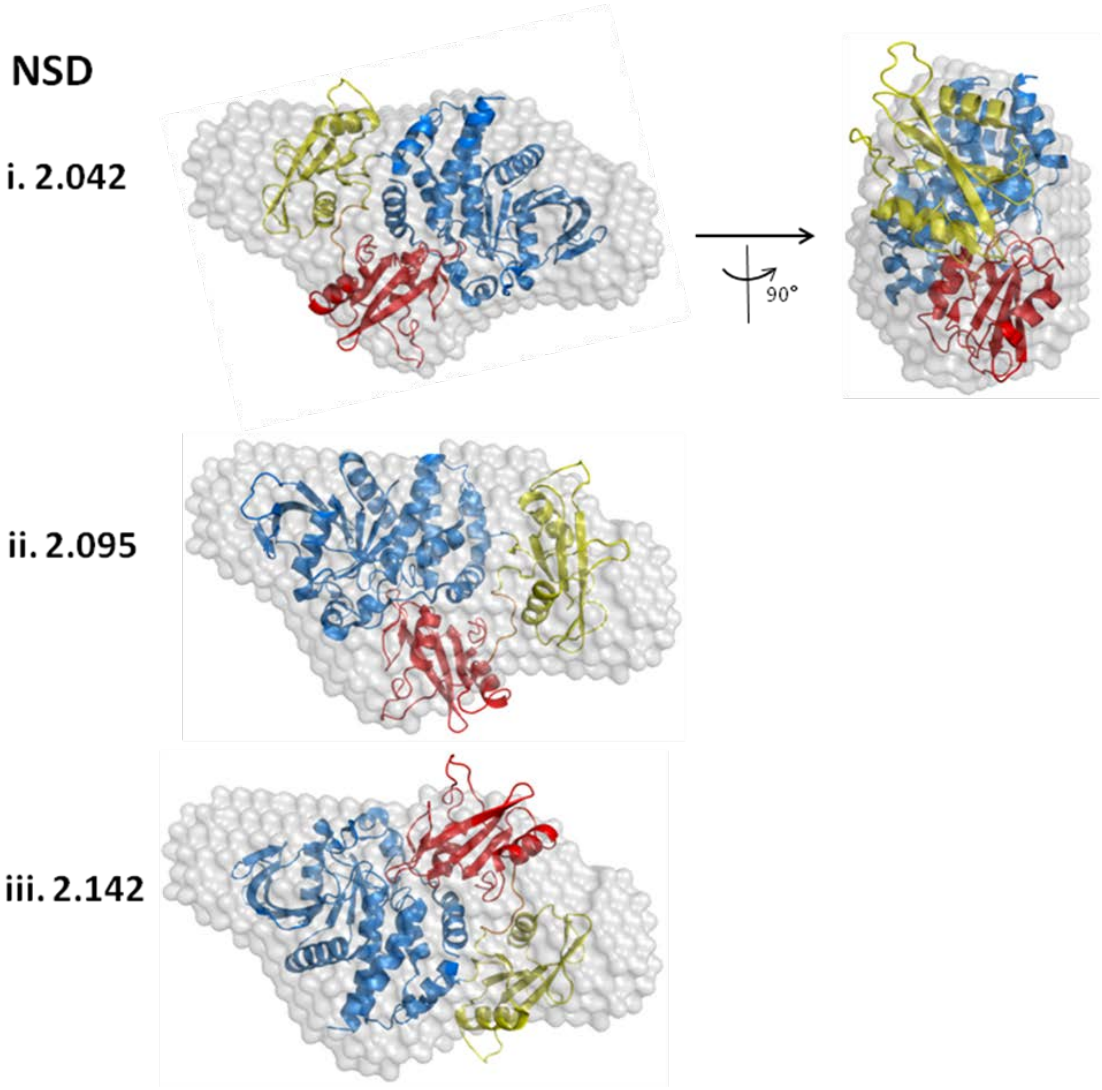


Figure 5.20: Automatic Fitting of FL-Shp2 WT SAXS Envelope with Shp-2 Crystal Structure.

(i-iii) Three automatic fits of Shp-2 crystal structure, in descending order of NSD (PDB ID: 4DGP, grey: surface representation of *ab initio* SAXS bead model, Red: nSH2, Yellow: cSH2, Blue: PTP. Far left: NSD values of fits generated by SUPCOMB, 0 being an ideal fit.

I-TASSER was utilised to generate a homology model of the FL-Shp2 WT construct used in this study and is depicted in Figure 5.21 This was performed to model the missing C-terminal residues absent in the crystal structure. The top model has a C-score of -0.95 and a TM-score of 0.59±0.14, the latter score indicating the correct topology of the model. Structural alignment of the model with crystal coordinates of WT Shp2 residues 1-532 (PDB ID: 4DGP) yields an RMSD of 1.511. The N-terminal hexahistidine tag and TEV protease cleavage site contain a nine residue α-helix, spanning His3-Val11. For the purpose of this study, the model of FL-Shp2 WT generated by I-TASSER is sufficient for further analysis.

Automatic fitting of the I-TASSER FL-Shp2 WT homology model with the DAMMIF-derived SAXS envelope yielded three solutions with NSD values of 1.984, 1.995 and 2.029 for solution 1, 2 and 3 respectively (Figure 5.21). Like the automatic fitting of the Shp-2 crystal structure previously, the solutions differed in their orientation on both axes. The modelled N- and C-terminal residues do not fit well into the envelope in all three solutions generated, though the NSD values suggest a slightly better fit than the experimentally determined Shp-2 structure. In particular, solution 1 for the FL-Shp2 WT homology model is 0.058 lower in NSD than the Shp-2 crystal structure. The improvement is <3% and could be due to more of the DAMMIF envelope being filled by structural elements, regardless of their location within it. As the difference is minor, the orientation and location of the Shp-2 terminal portions require refitting against experimental data for a better fit.

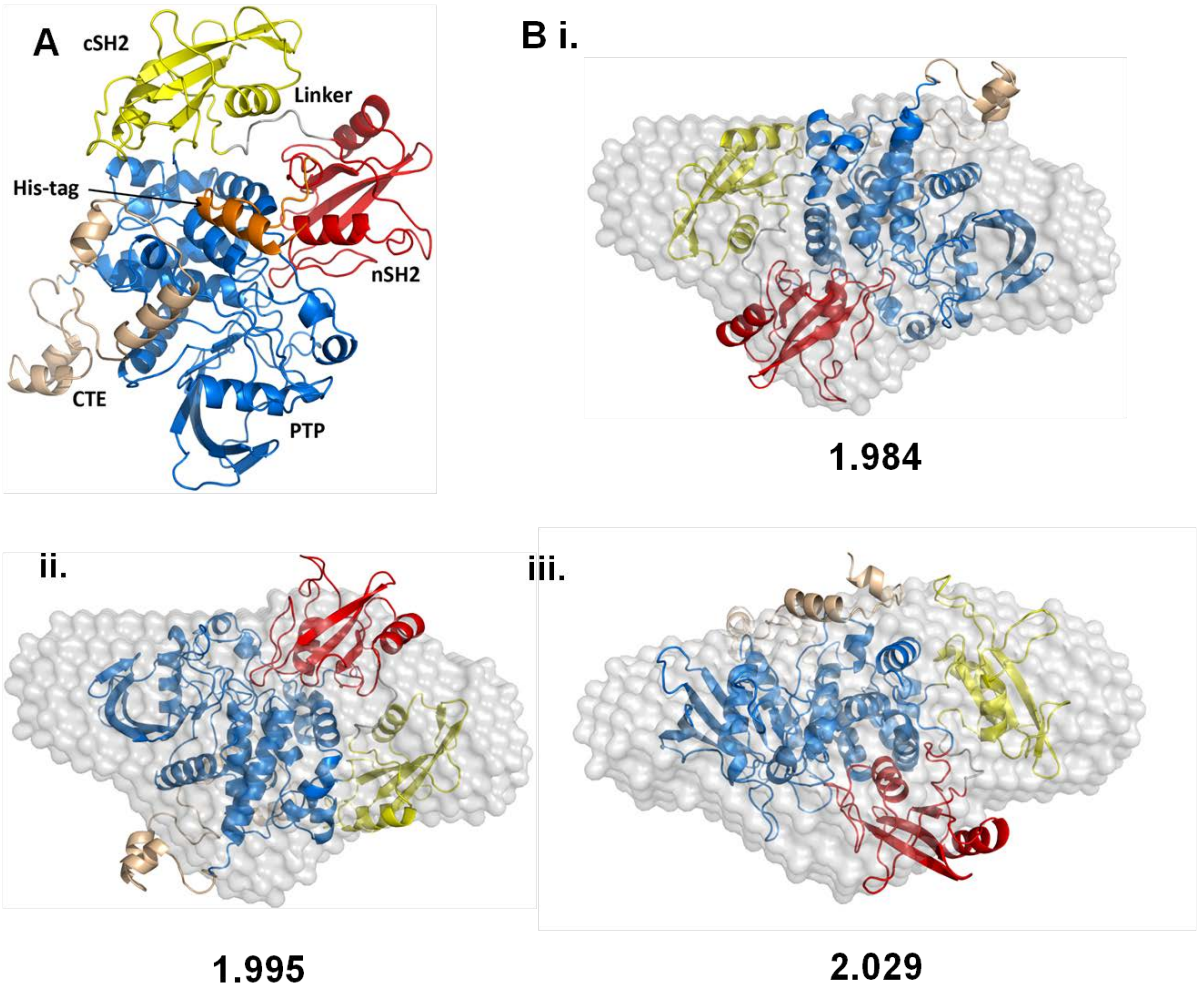


Figure 5.21: Homology Model Fitting Against FL-Shp2 WT SAXS Envelope.

(A) I-TASSER model of FL-Shp2 WT. Domains are coloured as follows: nSH2: red, inter-SH2 linker: grey, cSH2: yellow, PTP: blue, CTE: beige. **(B i-iii)** Automatic fits of the I-TASSER homology model against the DAMMIN-generated SAXS envelope. Values underneath each fit indicate NSD value generated by SUPCOMB, 0 being an ideal fit.

Using a data-driven modelling approach, a more reliable structure may be generated. For this we employed SAXSTER, which utilises P(R) data from SAXS and the protein primary sequence to model the best solution. The top-ranking model from SAXSTER for FL-Shp2 WT aligns with the experimental Shp-2 crystal structure (PDB ID: 4DGP) with an RMSD of 2.317. Comparison of the predicted and experimental P(R) curve shows a reasonable fit of the parabola up to 7 nm, the main deviation being the “tail” region from 8 nm onwards (Figure 5.22). The SAXSTER model overestimates this region, resulting in a D_{max} above 13 nm, contrasting the AUTOGNOM-calculated D_{max} of 10.3 nm.

CRYSOL back-calculation of the SAXSTER model and subsequent comparison with the merged FL-Shp2 WT SAXS data gave rise to an agreeable χ^2 value of 1.073. Visual inspection of the curves shows a misfit at low-q (<0.3 nm^{-1}) but a very good fit thereafter until q>1.5 nm^{-1}, from which point the curves deviate from one another substantially. Taken together the folds of the modelled termini are probably incorrect, whereas the overall shape information of the SAXSTER model is reasonable. This is corroborated by the theoretical and experimental R_g values of 30.99 and 31.89 Å respectively, which differ by <3% to one another. Since the P(r) curves indicate a poor D_{max} prediction and the intensities at low-q do not match well, this is expected.

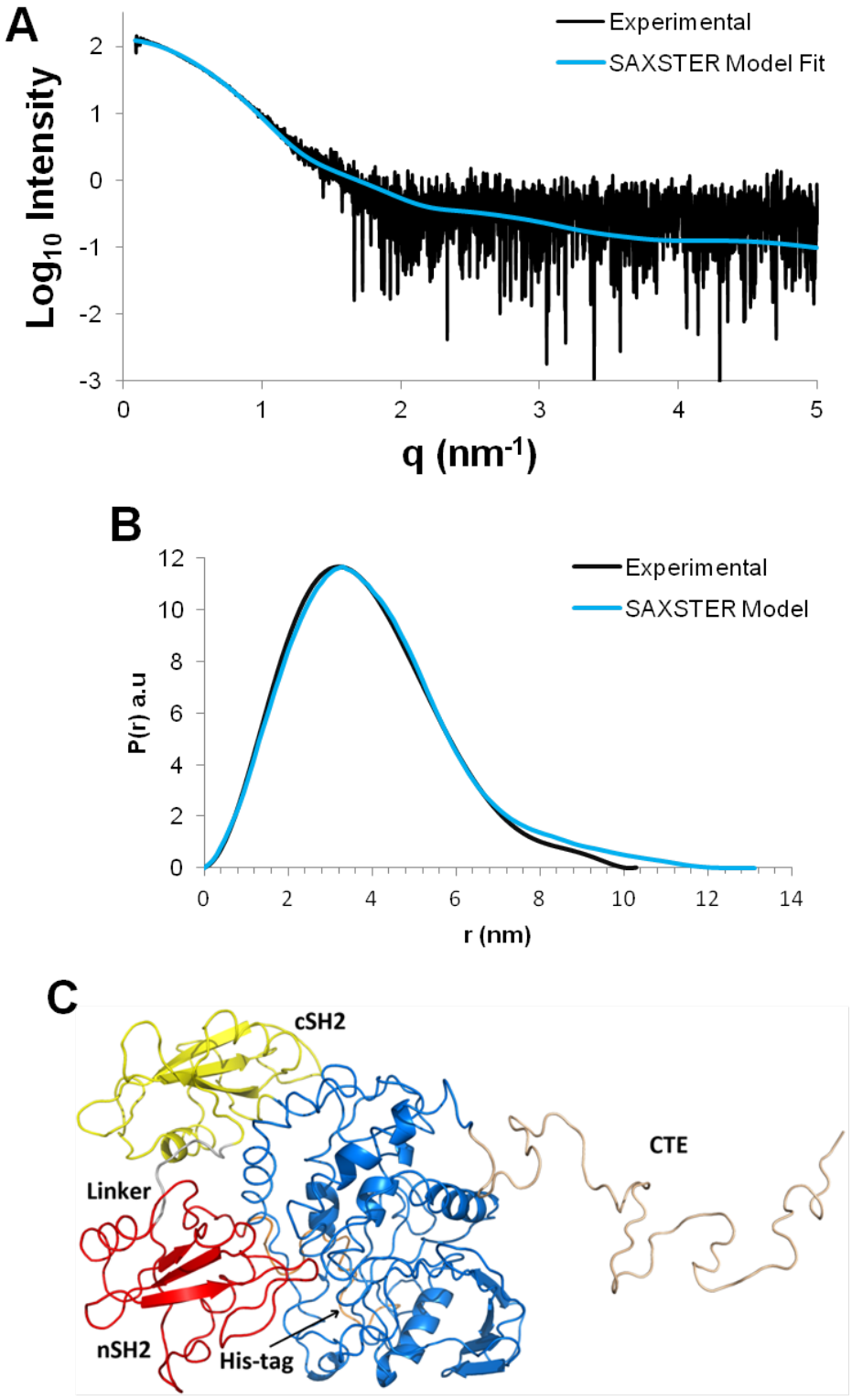


Figure 5.22: Data-driven Modelling of FL-Shp2 WT using SAXSTER.

(A) Overlay of CRYSOL simulated curve of model (blue) against experimental SAXS data (black). **(B)** P(r) curve of experimental and data-driven model. The experimental and model D_{max} values are 10.3 and 13.1 nm respectively. **(C)** Cartoon diagram of SAXSTER-generated model.

SUPCOMB-based superimposition of the SAXSTER model with the DAMMIF-derived envelope yielded three solutions with NSD values 1.876, 1.937 and 2.025 for solutions 1, 2 and 3 respectively (Figure 5.23). These values are slightly better than those of the previous model fits. Interestingly, the C-terminal disordered loop of the SAXSTER model protruding outside the envelope measures around 35 Å in total; when subtracted from the SAXSTER D_{max} prediction of 13.1 nm, a new D_{max} of 9.58 nm is created, which is within ~7% of the AUTOGNOM-calculated D_{max} of 10.3 nm. The data suggests that the C-terminal tail is not as extended as predicted by the SAXSTER model and likely folds back closer to the catalytic subunit.

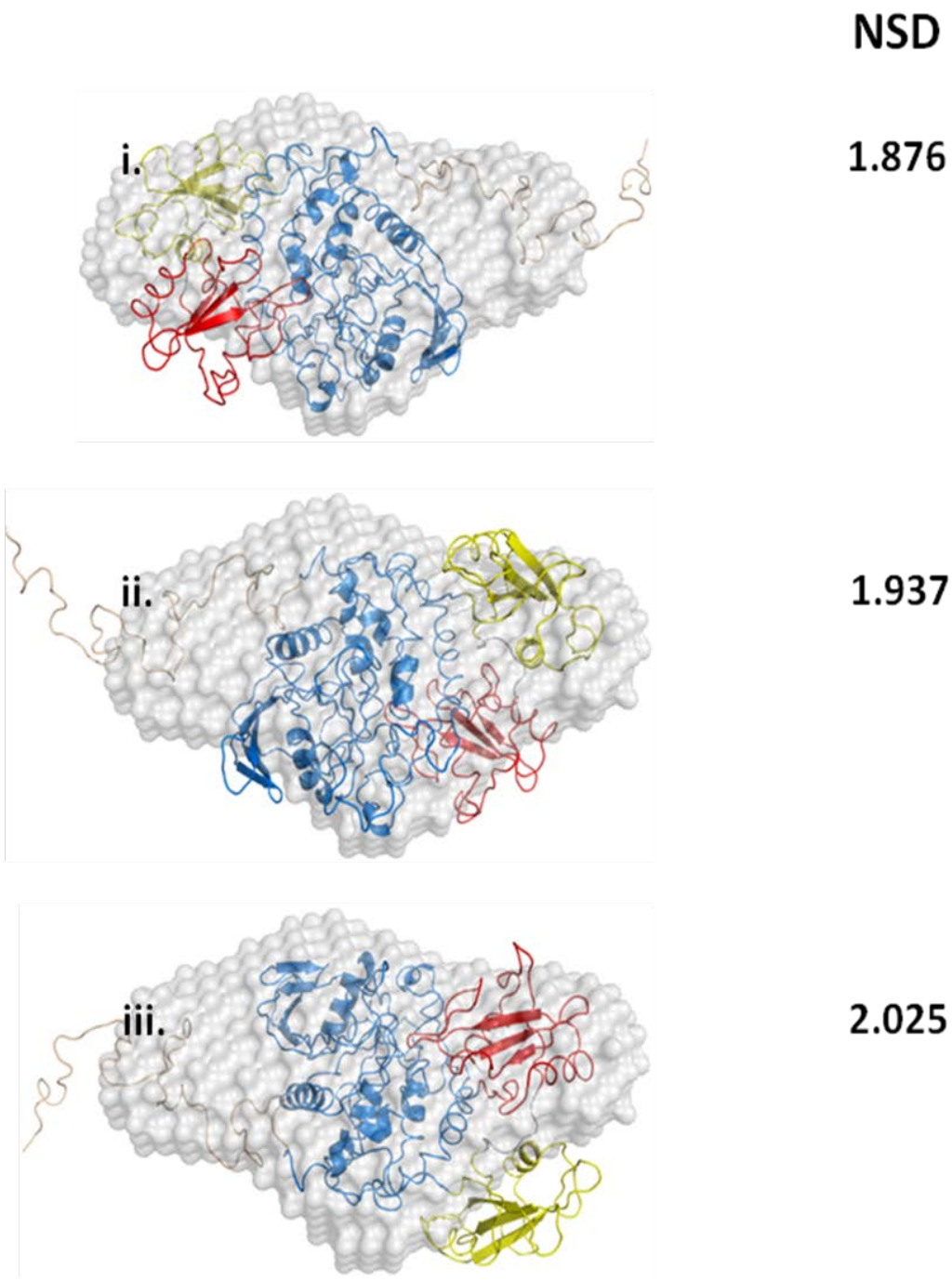
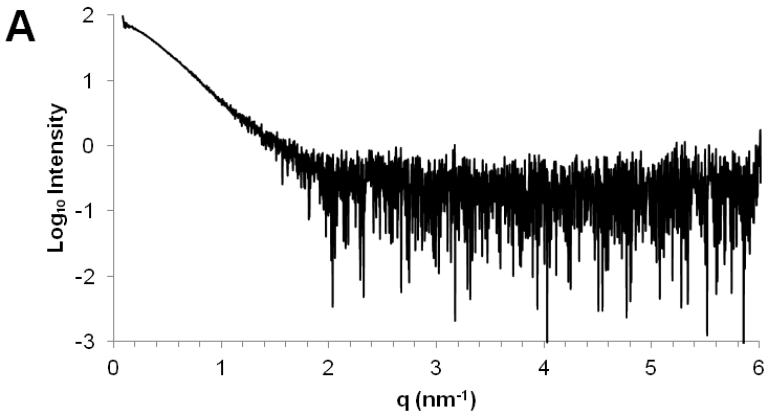


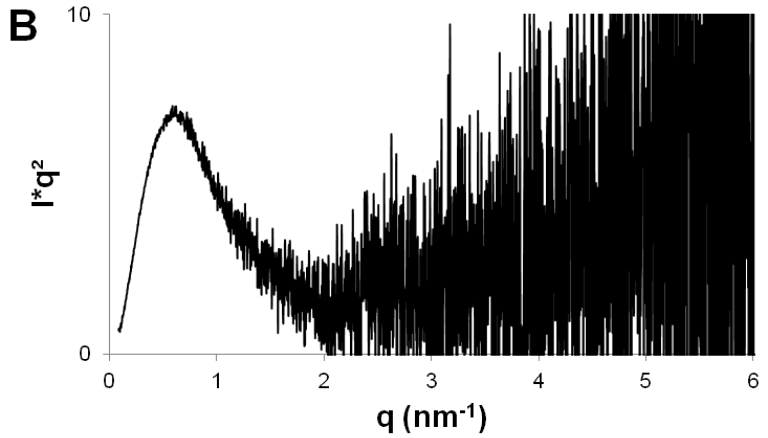
Figure 5.23: Automatic Fitting of FL-Shp2 WT SAXS Envelope with Data-Driven SAXSTER Model.

Three automatic fits (**i-iii**) of SAXSTER-generated model with DAMMIF-generated envelope (grey: surface representation of *ab initio* SAXS bead model, protein domains coloured as in (B). Far right: NSD values of fits generated by SUPCOMB, 0 being an ideal fit.

E76K FL-Shp2

SAXS data of E76K FL-Shp2 were collected at the same X-ray source and concentrations series as WT and displays similar attributes, merged SAXS data are shown in Figure 5.24.





C

Concentration (mg/ml)	R_g (nm)	I(0)
Extrapolated Data	3.588± 0.079	72.125± 0.370

Figure 5.24: Extrapolated SAXS Data of FL-Shp2 E76K.

(A) Merged and extrapolated to zero concentration SAXS data of FL-Shp2 E76K used for further analyses. **(B)** Kratky plot of SAXS data of FL-Shp2 E76K from (A), indicating a globular particle with a degree of flexibility. **(C)** Table of parameters calculated with data from (A).

Extrapolation of the data to zero concentration using PRIMUS results in a very similar curve to the original datasets (Figure 5.24). The R_g value is 3.588 nm, 0.002 nm less than the 1 mg/ml dataset R_g. For individual SAXS curves of FL-Shp2 E76K at all concentrations tested see Appendix figure 7.

Comparisons of FL-Shp2 WT and E76K SAXS Data Reveal Distinct Differences

Superimposition of both WT and E76K scatter profiles is shown in Figure 5.25. Initial fitting of the Shp-2 WT crystal structure (PDB ID: 4DGP) with the extrapolated SAXS datasets for both WT and E76K was performed as a fast quantitative measure of similarity between the two datasets and a reference model. χ^2-values of the WT and E76K SAXS data against 4DGP were 1.388 and 2.410 respectively. A better fit between 4DGP simulated curve and the WT SAXS data is observed than with E76K SAXS data (Figure 5.25. The CRYSOL-derived theoretical R_g for the Shp-2 crystal structure is 27.78 Å; the experimental R_g values for WT and E76K are 30.56 and 32.87 Å respectively. This explains the generally poor fit at low-q for both datasets as size information resides in this region and also explains why E76K has a worse fit than WT. However, it does not explain the poor fit at $q>1$ nm^{-1} for E76K. From this analysis one can conclude that the WT SAXS data fits the most recent Shp-2 crystal structure better than the E76K SAXS data, further suggesting the structure of FL-Shp2 E76K deviates from WT in size, shape and fold. Differences are observed in envelope shape between WT and E76K (automatic fitting between WT and mutant. Solution of E76K envelope does not match well, suggesting a global conformational change between the two constructs in solution.

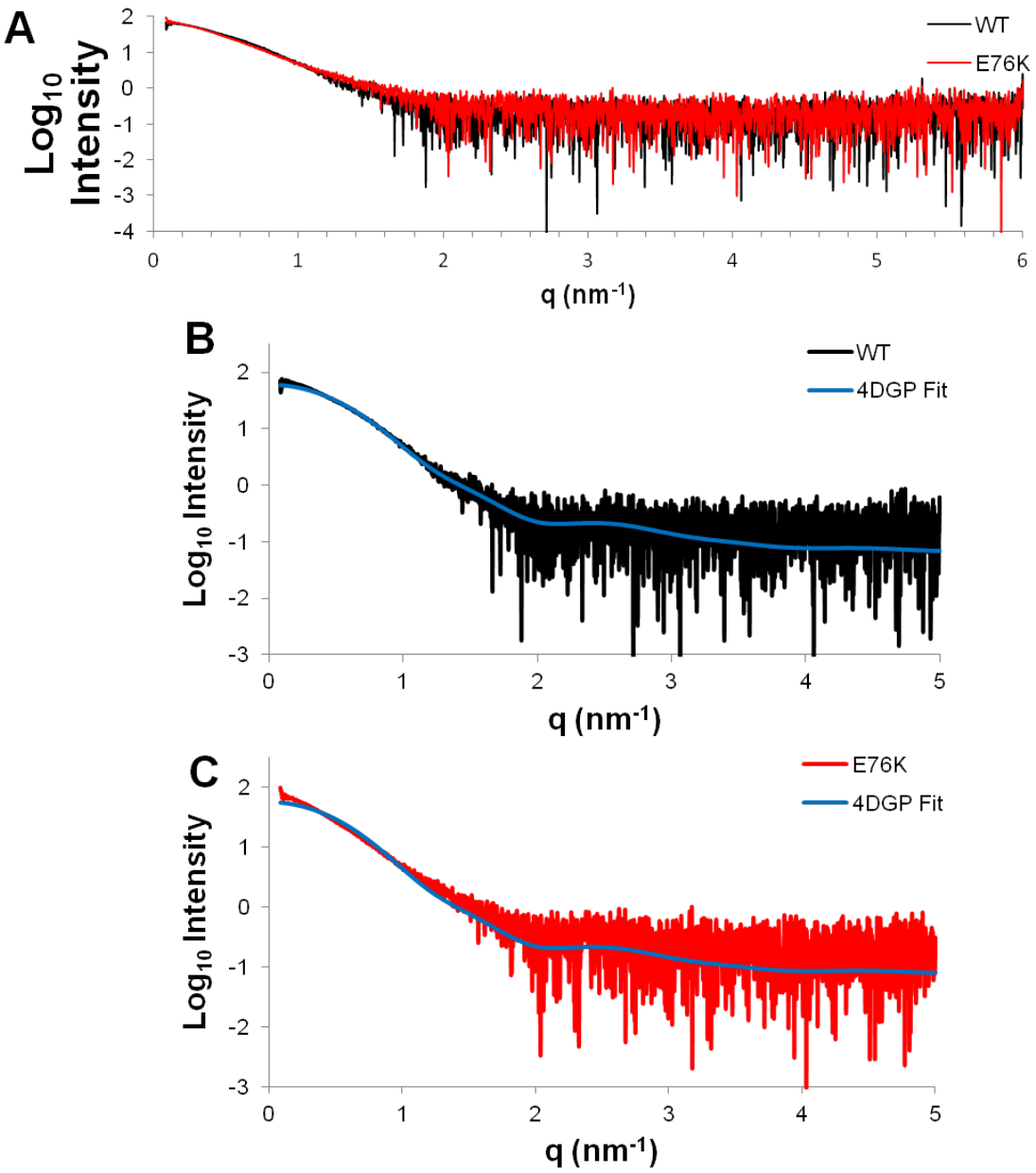


Figure 5.25: Comparison of WT and E76K FL-Shp2 SAXS Data.

(A) Superimposition of final WT (black) and E76K (red) SAXS curves used in all analyses. Differences in intensity and curve shape between both datasets are apparent at $0.5<q<1.6$ nm^{-1}, indicating a change in size and/or shape.**(B)** and **(C)** WT and E76K SAXS profiles fitted to CRYSOL simulated curves of a current Shp-2 crystal structure (PDB ID: 4DGP, blue), yielding χ^2 values of 1.388 and 2.410 respectively.

EOM Analysis of FL-Shp2

As the WT and E76K SAXS data were not yielding good fits against current crystal structures and homology models tested, EOM was employed to investigate the possibility of FL-Shp2 containing regions of flexibility. Since the WT and E76K constructs contain terminal residues that may be unstructured (Poole & Jones 2005), EOM can take this into account to yield an ensemble of structures that best fit the data as opposed to one rigid model.

In the first instance, all structural domains were kept rigid, with residues 1-22 (His-tag) and 552-615 (C-terminal extension, CTE) remaining flexible. This was performed to assess the contribution of the termini to the experimental data and is summarised in Figure 5.26. The fit between the ensemble and experimental curve is shown in Figure 5.26 A and demonstrates a good superimposition, with χ^2=1.146, better than that of 4DGP (χ^2=1.388. Inspection of the models shows a rigid TanSH2+PTP domain with a large flexible CTE. The His-tag region tends to adopt a fully extended conformation localised around the same region of space between the TanSH2 domain on the opposite face to the catalytic site. The R_g and D_{max} distribution plots in Figures 5.26 B and C indicate that in a 3-domain rigid model, FL-Shp2 WT tends to adopt more extended conformations, with average R_g and D_{max} values of 31.6 and 125.4 Å.

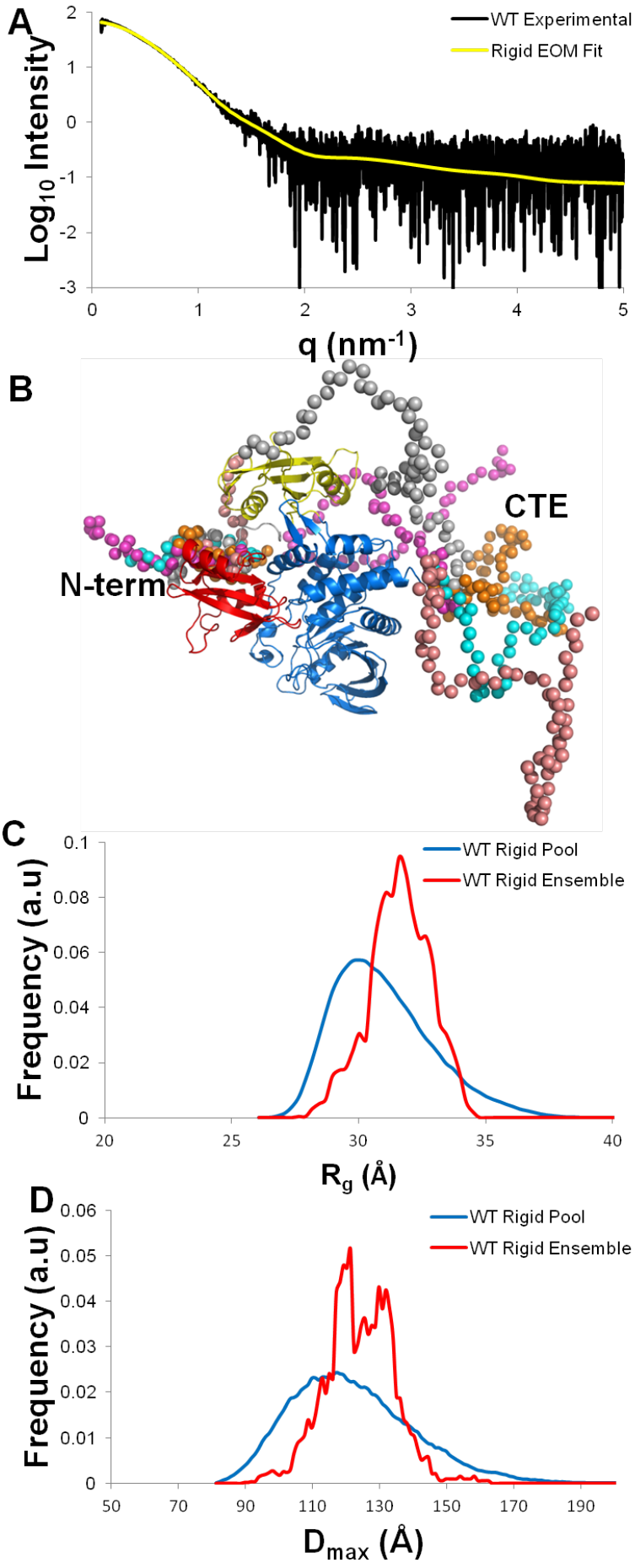


Figure 5.26: EOM Analysis of FL-Shp2 WT Assuming a Rigid Model.

(A) Fit of WT SAXS data (black) against averaged curve of rigid ensemble selected by EOM (yellow). X^2=0.996. **(B)** Superimposition of 5 models from the EOM ensemble. Rigid 3 domains of Shp-2 are coloured as previous; His-tag (denoted as N-term) and CTE are represented as beads, each colour representing a different model. **(C)** and **(D)** R_g and D_{max} distributions respectively. Both indicate the preference for an extended conformation, on average.

Next, EOM was performed with the linker between the two SH2 domains (also called the interSH2-linker here, residues 103-110 of WT sequence inclusively; residues 125-132 of FL construct) kept flexible. This region may behave like a hinge region and thus be involved in the domain reorientation of Shp2 during activation. Results of the analysis are summarised in Figure 5.27 and demonstrate a similar pattern to the rigid model of FL-Shp2. The EOM ensemble fits well to the experimental data (Figure 5.27 A) at all q-values, with a χ^2=0.999, better than the rigid model which gave a value of 1.146. The average R_g and D_{max} values are slightly higher than the rigid model, at 32.9 and 132.7 Å, respectively, and the distribution plots demonstrate that in a flexible interSH2 linker model, the ensemble is more extended on average compared to the pool (Figure 5.26 C and D), consistent with the nSH2 domain being released and allowed to move in space. Models from the ensemble were superimposed and revealed two main populations, four representative models are shown in Figure 5.27 B. In general, the nSH2 domain either remained relatively close to the PTP domain active site or undergoes major domain reorganisation to effectively “peel” away from the PTP domain entirely and reside on the opposite face of the catalytic residues. Of the 15 returned models, 4 were in the latter, peeled configuration, roughly 26% of the ensemble.

Similar results were obtained for the same nSH2-mobile model (where the interSH2 linker was kept flexible) when using E76K SAXS data (Figure 5.27), with a χ^2 between the averaged EOM ensemble and experimental data of 1.106. Comparisons of the R_g and D_{max} distribution plots between WT and E76K (Figure 5.29) reveal that the mutant FL-Shp2 ensemble has higher R_g values than the WT, suggesting that the mutant takes on more extended conformations. However, the D_{max} distribution plot

suggests the opposite, that E76K takes on more compact conformations than the WT. Although these two results seem contradictory, it could point to mutant Shp2 adopting a range of conformations that are lower in maximum particle dimensions to the WT but still undergoing a global reorganisation to yield a higher R_g, though more experiments will be required to substantiate this.

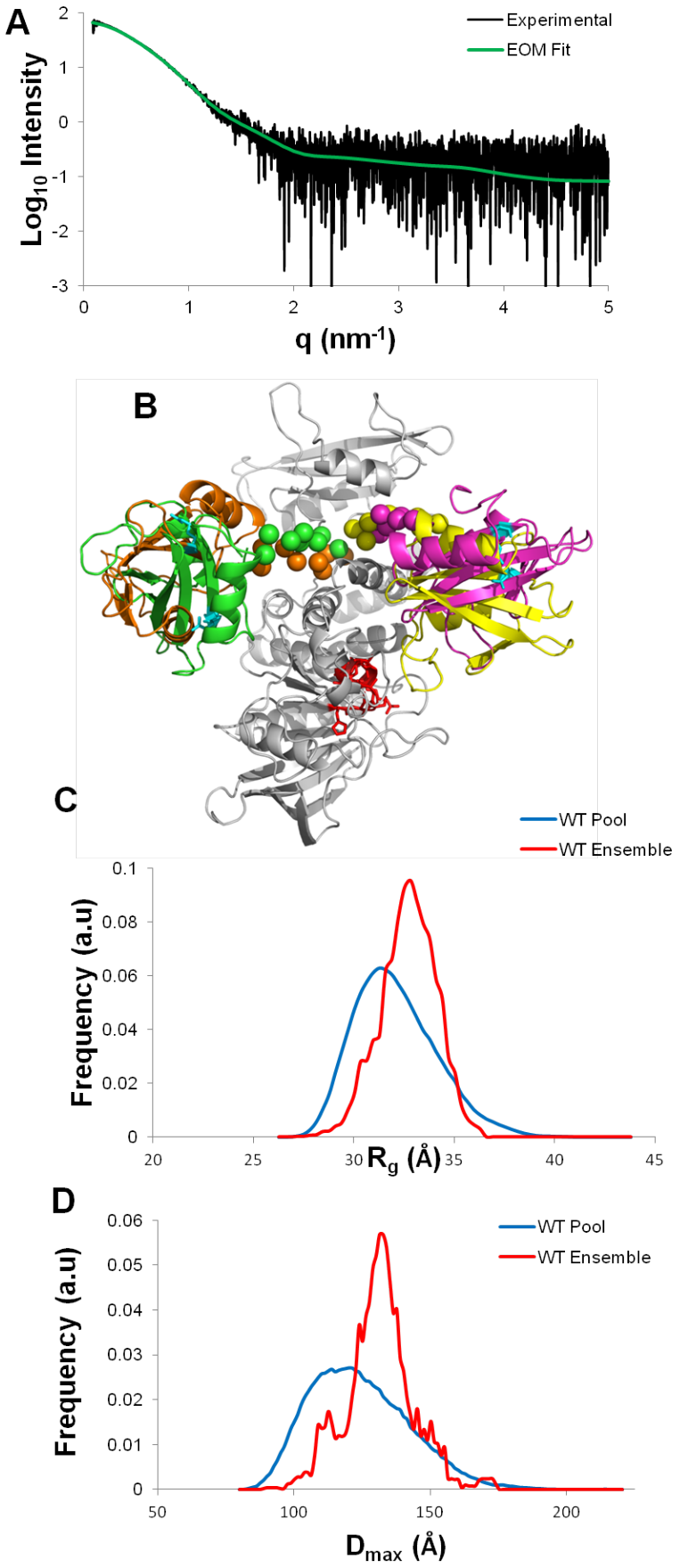


Figure 5.27: EOM Analysis of FL-Shp2 WT Assuming an nSH2-mobile Model.

(A) Fit of WT SAXS data (black) against averaged curve of ensemble selected by EOM (green). X^2=0.999. **(B)** Superimposition of 4 models from the EOM ensemble. His-tag and CTE are omitted for clarity. Grey cartoon: cSH2 and PTP domain, Red sticks: Catalytic residues, Cyan sticks: 32Arg of nSH2 domain, Spheres: interSH2 linker. nSH2 domains and interSH2 spheres are coloured in magenta, yellow, green and orange. **(C)** and **(D)** Distribution plots for R_g and D_{max} of pool (blue) and selected ensemble (red). Both suggest that WT FL-Shp2 adopts, on average, an extended conformation.

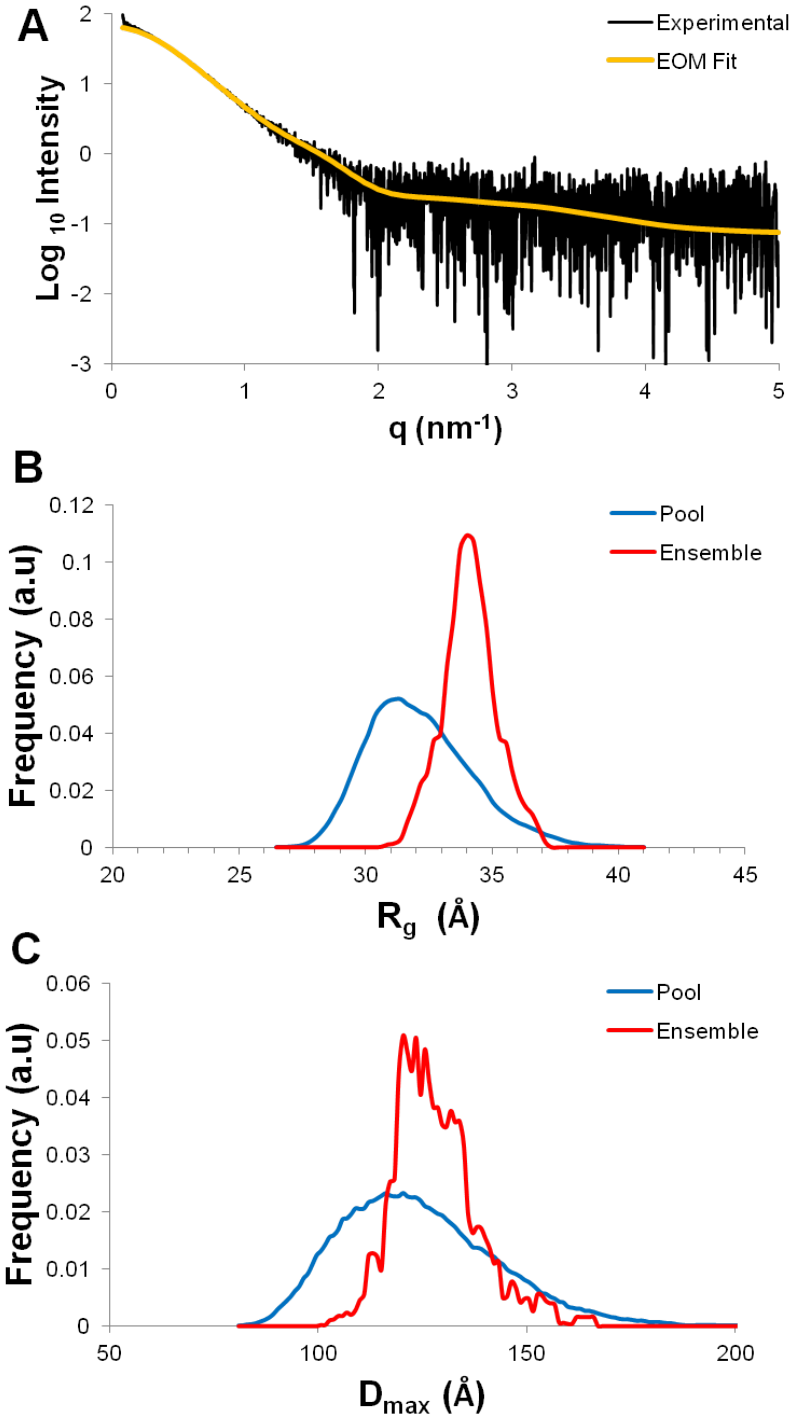


Figure 5.28: EOM Analysis of FL-Shp2 E76K.

(A) Fit of E76K SAXS data (black) against averaged curve of ensemble selected by EOM (yellow). X^2=1.106. **(B)** and **(C)** Distribution plots for R_g and D_{max} of pool (blue) and selected ensemble (red). Both suggest that WT FL-Shp2 adopts, on average, an extended conformation.

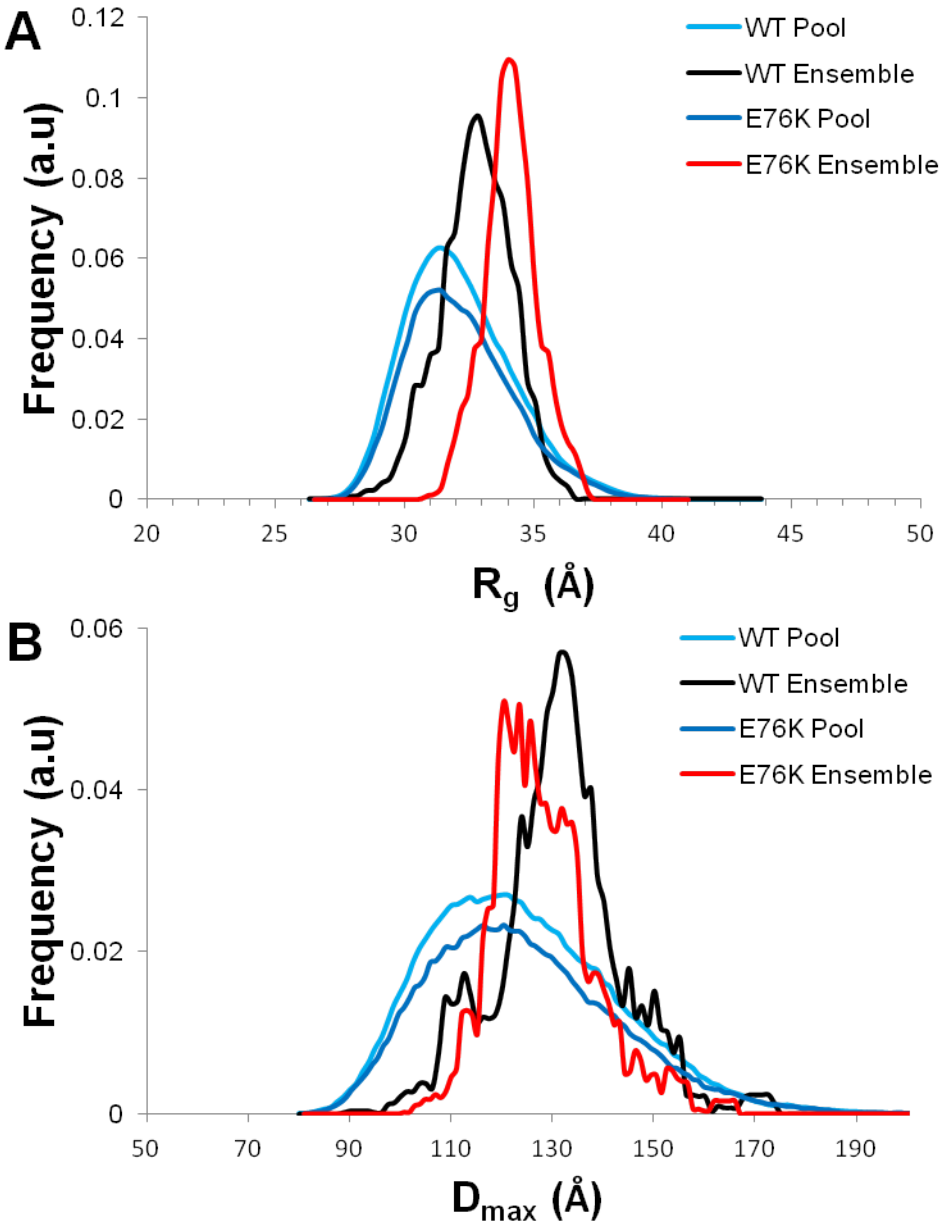


Figure 5.29: Comparison of WT and E76K EOM Fits.

(A) and **(B)** R_g and D_{max} distribution plots for WT and E76K FL-Shp2. The E76K ensemble has a larger mean R_g distribution than the WT, whereas the opposite is true for the D_{max} distribution. (Light blue: WT pool, Black: WT ensemble, Dark blue: E76K pool, Red: E76K ensemble. **(C)** Boxplot Diagram demonstrating differences in EOM Ensemble Cα distances between D61 and C459 for WT and E76K FL-Shp2. The WT takes on a broad distribution of D61-C459 distances, whereas E76K takes on a narrower range of distance distributions that are in general higher than the WT.

It is well known that the active site of Shp2 is occluded by nSH2 domain residues residing in the DE loop (Hof *et al*., 1998). It is thought that the interaction of the DE loop residues with the active site is disrupted upon nSH2 ligand binding, allowing the bound-state nSH2 to reveal the catalytic residues of the PTP domain. Therefore it is reasonable to assume that the distance between the DE loop and catalytic residues could correlate with activation status of the enzyme.

With this in mind, Cα distances were calculated between D61 within the DE loop and the catalytic cysteine C459 for the EOM selected ensembles. This was performed for both WT and E76K, the results of which are summarised in Figure 5.30 . The boxplots clearly reveal a difference between the Cα distances calculated between WT and E76K. The selected ensemble of the mutant contains models with, on average, larger distances than the WT and an overall broader range. The larger distance could be enough to fully activate or prolong FL-Shp2 activation, though more structural studies of the mechanistic consequences of Shp2 mutation are required.

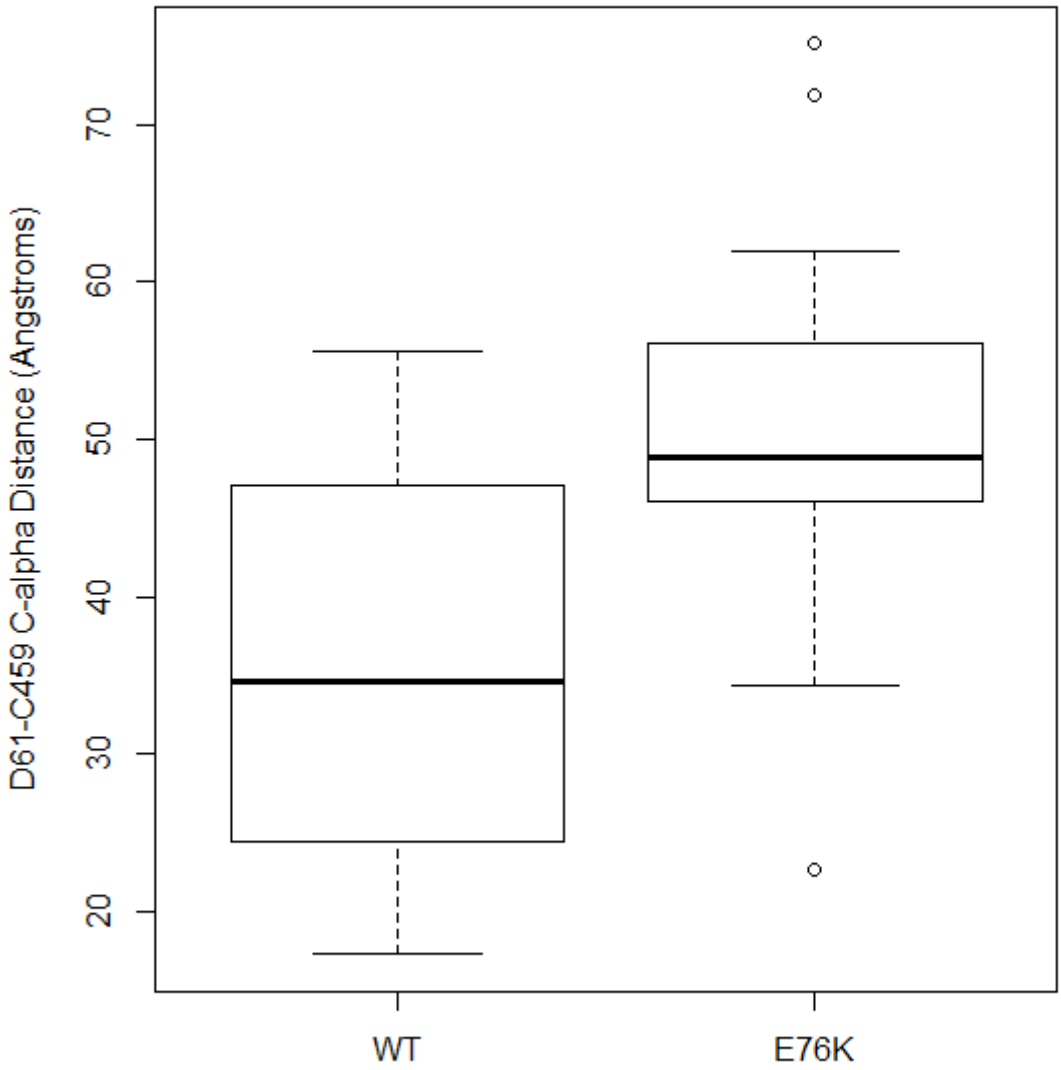


Figure 5.30: Comparison of C-alpha Distances between D61 and C459.

Boxplot showing the distribution of C-alpha distances between C459 and D61 (Catalytic motif and inhibitory loop). The data suggest E76K, on average, takes on a more open form than the WT form. Based on EOM-generated models.

Conclusions

Through SAXS analysis of constructs spanning various Shp2 domains as well as the full-length form itself, a number of insights have been made regarding the structural states of Shp2 in solution. In addition, study of SAXS data of a Noonan Syndrome/leukaemia mutant, E76K FL-Shp2, was performed and also provides insight into Shp2 behaviour in the mutated state. These conclusions are summarised below and discussed further.

Global Envelopes of Shp2 Agree with Published Data

All domains tested yielded high quality models and their shapes largely resemble the crystal state, though the discrepancy in the fits suggests a degree of flexibility in the solution state. In particular, ΔHis-TanSH2 produced an envelope with an asymmetric dumbbell shape, consistent with the fact the cSH2 domain of Shp2 is larger and contains freely mobile loops. While SAXS is unable to definitely resolve which SH2 domain belongs to which spherical portion in the global TanSH2 envelope, it is likely that the cSH2 domain fits the larger sphere. In general the crystallographic and solution data fit well, increasing the confidence of both datasets.

Domain dynamics may play a significant role in Shp1/Shp2 function as no other PTPs have been observed to contain autoinhibitory activity reliant on catalytic cleft occlusion. The dynamic states of Shp2 are expanded on below.

Shp2 and Isolated Domains are Flexible in Solution

The TanSH2 domain was found to be partially flexible in solution. When comparing various flexible models, the best fits were obtained when flexibility in the N-terminus, interSH2 linker and CD loop of cSH2 were introduced. A functional role of the latter is unlikely as the CD loop is non-overlapping to canonical cSH2 pY and +3 pocket residues, however there may be an as of yet unassigned function to the CD loop in Shp2 biology. The relatively long CD loop insertion is atypical of SH2 domains which contain small (3-4 residue) CD loops that play a largely structural role, and thus may confer some function as it is highly conserved throughout vertebrate *PTPN11* sequences.

The PTP domain of Shp2 has been studied structurally both free and in complex with small-molecule inhibitors, though no research has investigated the structure and dynamics of Shp PTP domains in solution as all solved structures were determined by X-ray crystallography. In this present study PTP-Shp2 is observed through *ab initio* modelling to adopt a globular shape that fits reasonably well to the DAMMIF-derived envelope. However, rigid body modelling and existing crystal structures were unable to fit the experimental data, leading to analysis of multiple, flexible ensembles in an attempt to decrease the model:experimental discrepancy.

The results show some preference for flexibility in the structure (in the Flex-N-PTP model), most likely localising to the N-terminal 70 residues. Flexibility can be a means to perform a biological function and is the case for PTP function via switching of the WPD loop between open and closed states (Hu & Stebbins, 2006). With Shp2

in the full-length, PTP domain flexibility could be a means to facilitate nSH2 release from the autoinhibited form into the active state. The flexibility of the PTP domain could also explain why Shp2 in the SH2-ligand free state still displays low basal phosphatase activity; intrinsic motions within the PTP domain may cause spontaneous nSH2 release in the absence of phosphopeptide to bind nSH2. The released nSH2 domain then undergoes conformational changes to optimize phosphopeptide binding and when bound to a ligand, remain in the released state.

Presence of a His-tag alters Global TanSH2 Structure

The DAMMIF envelopes between His and ΔHis TanSH2-Shp2 differ greatly from one another, suggesting the presence of a tag alters the global structure and thus may affect functional aspects of Shp2 function. For ΔHis-TanSH2-Shp2, the crystal coordinates of Shp2's TanSH2 domain fit well within the envelope and yield low χ^2-values, indicating the solution structure is compatible with the solved crystal structure. For His-TanSH2, good fits were observed for atomic models with averaged envelopes (Figure 5.3). The fact a discrepancy between SAXS data and crystallographic data exists suggests that conformational and domain orientation reordering in solution occur.

Wild-type and Mutant Shp2 Differ Structurally in Solution

FL-Shp2 SAXS data revealed some insights into full length structural features, such as the good fit between current crystallographic data (Peter Hof *et al*. 1998) and

DAMMIF envelopes. Flexibility in solution could be due to the His-tag and CTE, which are predicted to be mostly disordered, as well as the CD loop within the cSH2 domain. Additional flexibility may be attributed to the various functional loops within the PTP domain, however this was not tested during modelling.

The C-terminal tail has not been characterized structurally up until now. SAXS has provided a glimpse of this region and it is most likely completely unstructured.

Crucially however, the clear differences in the SAXS data between WT and E76K suggest a definite structural reorganisation, indicated by calculation of invariant parameters. E76K FL-Shp2 may sample more extended conformations compared to the WT and this is consistent with current biochemical evidence (Keilhack *et al*., 2005) and molecular dynamics (MD) studies (Darian *et al*., 2011). This is the first instance of a solution technique probing mutant Shp2 structure, as discussed previously in Chapter I.

Chapter VI

Biophysical Analysis of Shp2 and G6b-B Interactions

G6b-B and Shp2 are known to be binding partners from co-immunoprecipitation (de Vet *et al*. 2001a), SPR and transgenic mouse knockout experiments (Mazharian *et al*. 2012). As a novel membrane receptor with little known of its biological structure and function, the molecular mechanisms behind G6b-B:protein interactions may provide key insights into its role in signalling. With this in mind, experiments were performed in order to shed light on the binding mode of these biomolecules.

Circular Dichroism of the dITIM of G6B-b

Far UV CD was used to characterize the secondary structure of the doubly phosphorylated dITIM of G6B-b prior to high resolution analyses. The dITIM peptide was predicted by the secondary structure prediction program JPred (Cole *et al*., 2008) to be a mixture of unstructured loops and β-strands (See Appendix 5 and figure). Two dITIM concentrations were tested to rule out concentration effects and to select for the concentration that will yield the highest quality data at lower wavelengths. For the present study 1 and 10 mg/ml dITIM was tested by CD. After solvent subtraction and inspection of the high tension values at wavelengths below 200 nm for both concentrations tested, 10 mg/ml was chosen for further analysis using the online CD analysis server DICHROWEB (Whitmore & Wallace, 2004).

Visual inspection of the spectrum revealed that dITIM was unfolded in solution (Figure 6.1, top): the presence of a negative band around 200 nm with no other characteristics suggests little to no secondary structural elements exist. However, estimation of the secondary structure content based on the CD data suggested that whilst dITIM is primarily unfolded, there may be some transient helical and strand content (Figure 6.1, top, inset).

Analytical Ultracentrifugation of the His-TanSH2-Shp2:dITIM Complex

AUC experiments were conducted to investigate His-TanSH2-Shp2 (hereby called His-TanSH2) in the unbound and 1:1 dITIM-bound states. A protein:protein interaction between His-TanSH2 and dITIM should cause changes in the sedimentation profile of the former (Schuck, 2000). Both high (20 µM 0.5 AU A_{280}) and low (6 µM; 0.15 AU A_{280}) concentrations of unbound and dITIM-bound His-TanSH2 were used in sedimentation velocity experiments to rule out concentration effects. The His-TanSH2:dITIM complex was purified by gel filtration (Chapter II and III) to ensure a homogenous solution of d-ITIM bound His-TanSH2 only, excess dITIM peptide eluting last due to its low molecular weight.

Data obtained from sedimentation velocity experiments were processed with Sedfit 12.1b and revealed unbound TanSH2 to have a peak sedimentation coefficient of 2.46 S. The peak sedimentation coefficient for dITIM-bound TanSH2 is 2.77 S (Figure 6.1 B), a clear increase, suggesting complex formation. Interestingly, dITIM-bound TanSH2 displays a much broader sedimentation profile than unbound TanSH2, which may indicate multiple species of varying hydrodynamic properties.

This in turn might be caused by dITIM-bound TanSH2 adopting multiple extended and compact conformation states in solution.

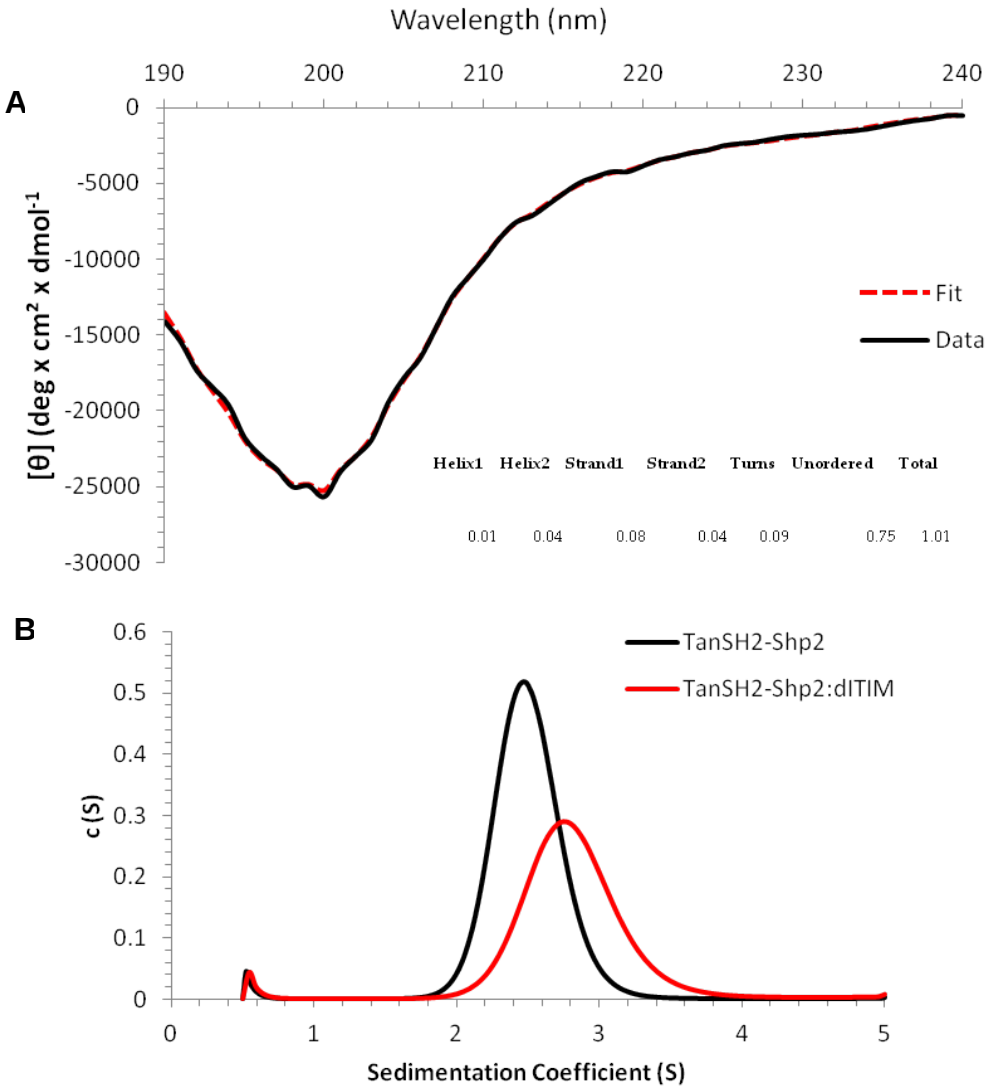


Figure 6.1: CD Data of dITIM of G6B-b and AUC Analysis of Unbound and dITIM-bound Forms of His-TanSH2-Shp2.

(A) CD Spectrum was recorded at RT in 50 mM sodium phosphate buffer. dITIM concentration was 1 mg/ml. Black solid line: experimental data, Red dashed line: Simulated data fit. This CD spectrum is indicative of an unfolded polypeptide. Data were analysed using DICHROWEB. (**B**) Sedimentation analysis (Velocity experiment) of TanSH2-Shp2 both free and in complex with dITIM. The difference in S value is indicative of complex formation as the hydrodynamic radius of TanSH2-Shp2 increases. Black: Unbound TanSH2-Shp2, Red: dITIM-bound TanSH2-Shp2. Data were analysed with SEDFIT 12.1b.

NMR of nSH2:ITIM1s Complex

The ^{15}N HSQC spectrum of nSH2 in the presence of ITIM1s (10 amino acid length, 1:1 ratio) phosphopeptide, like that of the unbound nSH2-Shp2, is well dispersed and indicates the presence of folded protein. A total of 146 peaks are present, out of a possible 133, almost 10% more peaks than would be expected. This could suggest the complex is more flexible, sampling two or more structural states, or, exhibiting slow-intermediate exchange phenomena due to the presence of sub-stoichiometric concentrations of ITIM1s peptide. The peaks are near-uniform in intensity, with a few possessing stronger and weaker intensities, suggesting variable exchange rates and therefore variable timescale motions present in the complex, much like the unbound form which also had peaks of varying intensity. Two peaks between 10.1 and 10.6 ppm in the ^{1}H dimension are present in the bound state, both of which may correspond to the 6W $H^{\varepsilon 1}$ resonance. Interestingly, the unbound nSH2-Shp2 NMR spectrum contains two peaks in this region, both likely to belong to 6W $H^{\varepsilon 1}$ in two conformations as no correlated carbon resonances were observed in 3D assignment spectra in this region. Superimposition of unbound and ITIM1s-bound nSH2-Shp2 HSQCs reveal both 6W $H^{\varepsilon 1}$ peaks have undergone CSPs. The $H^{\varepsilon 1}$ peak at 10.5 ppm has shifted 0.4 ppm upfield, while the $H^{\varepsilon 1}$ peak at 10.6 ppm has shifted 0.01 ppm upfield. The intensities remain relatively constant for the 10.6 ppm peak whereas the 10.5 ppm peak volume has decreased 4-fold. This suggests the conformation(s) corresponding to the latter peak are decreasing in relative populations between the unbound and bound states, with intermediate exchange timescale dynamics being suggested by broadening of resonances.

Based on the results from the nSH2:ITIM1s NMR peptide titration experiments, it was clear that assignment of the bound state would be challenging due to the slow-intermediate exchange regime observed. It was decided to collect a set of BEST backbone assignment spectra to aid the assignment process. Data were obtained on a 900 MHz Varian spectrometer and analysed using a semi-automated assignment method on CCPN Analysis 2.1.5. The results of this assignment are presented below.

Backbone Assignment of nSH2:ITIM1s Complex

Unambiguous assignment of the nSH2:ITIM1s complex backbone was straightforward due to the increased peak resolution observed in the spectra from acquiring data on a high-field magnet. A representative HNCA-HN(CO)CA sequential walk through the sequence is presented in (Figure 6.2 A), showing (i) and (i-1) connectivities for residues 46R-51V, as was the case for nSH2 and TanSH2-Shp2 apo-state assignments. Four prolines (9P in the initial N-terminal loop, 33P at βB6 – immediately C-terminal to the conserved Arg at position βB5, 38P at position BC4 and 101P in the final C-terminal loop) are unassignable using currently collected spectra, leaving 110 assignable residues, of which 100 were assigned in the ^{15}N-HSQC spectrum (Figure 6.2 B). This is an assignment coverage of 90%, unassigned residues are as follows: 3S, 4R, 35K, 54I, 65L, 66Y, 67G, 85H, 86G and 90E. Apart from 54I which resides in sheet βD, all unassigned residues are located in loop regions and is summarised in (Figure 6.3). 65L, 66Y and 67G form part of the EF

loop and were assigned in the unbound state, meaning the SH2:peptide interaction affects the exchange rates of this region such that peaks are no longer observable.

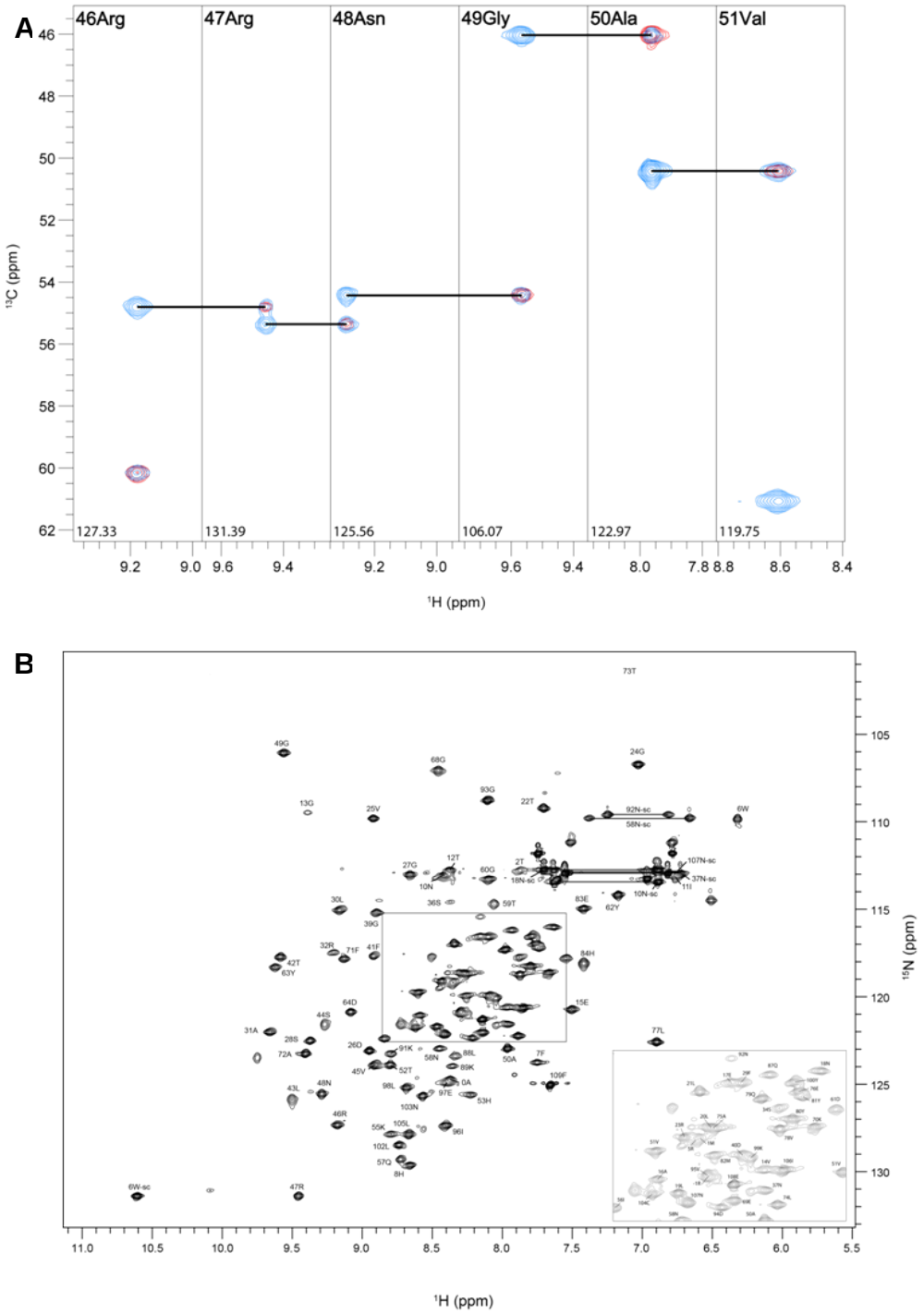


Figure 6.2: Sequential Backbone Assignment of nSH2:ITIM1s

(**A**) **HNCA and HNCOCA Spectra.** Strips of 3D NMR assignment spectra showing connectivities between HNCA (blue peaks) and HNCOCA (red peaks) established using a semi-automated assignment method within CCPN Analysis 2.1.5. Black solid lines show sequential connectivities between HNCA and HNCOCA peaks corresponding to residues at (i) and (i-1) positions for residues 46-51 inclusively for WT nSH2-Shp2, corresponding to the protein sequence RRNGAV. A similar assignment approach was done for HNCO/HNCACO and HNCACB/HNCOCACB spectra in parallel. (**B**): **Assigned ^{15}N HSQC of ^{15}N, ^{13}C-labelled nSH2-Shp2 in Complex with ITIM1s .** nSH2-Shp2 backbone amide signals were assigned using a 2 mM sample of uniformly ^{15}N and ^{13}C-labelled protein. The crowded region of the spectrum (boxed, middle) is expanded for clarity (bottom right). “sc” denotes side-chain assignments. Assigned side-chain HN peaks of Asn/Gln are connected via single horizontal black solid lines. A larger figure can be found in Appendix figure 8.

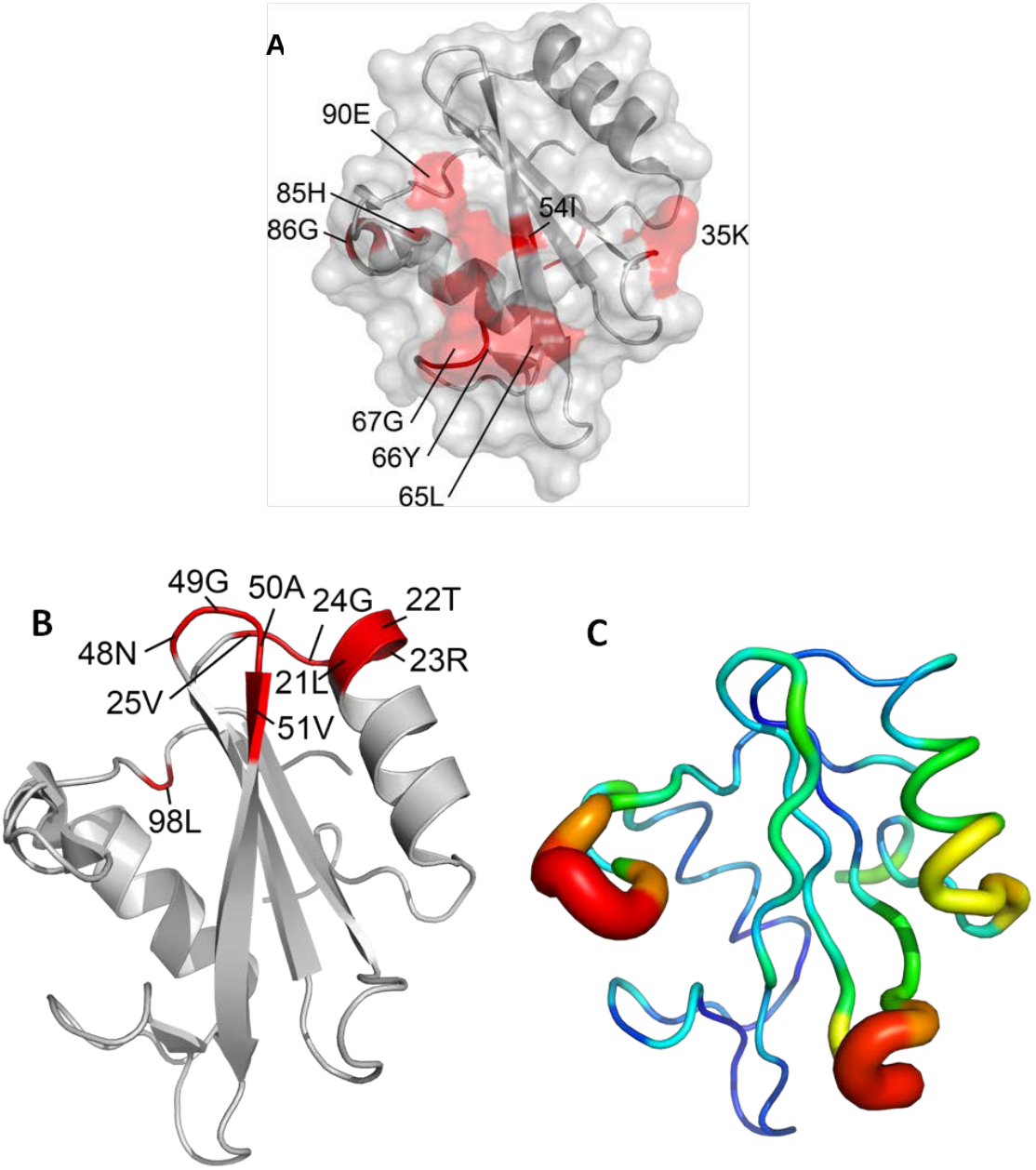


Figure 6.3: Unassigned Residues and Conformational Flexibility in nSH2:ITIM1s

(A) Assignment of eight WT residues was not possible for the nSH2:ITIM1s complex and are highlighted (red) in a cartoon and surface representation (grey, PDB ID 1AYD). 35K forms part of the BC loop, 65G, 66Y and 67G form portions of sheet βE and the EF loop. **(B)** Cartoon representation of nSH2-Shp2 with residues of multiple peaks in ITIM1s-bound HSQC highlighted (red). Apart from 98L, all residues localise to one distinct patch of nSH2 with no overlap to the pTyr and +3 pockets. **(C)** B-factor putty representation of nSH2-Shp2 indicating regions of flexibility. From right to left the red, wide regions correspond to the BC and BG loops respectively (PDB ID: 1AYD)

A number of peaks in the nSH2:ITIM1s complex HSQC were found to adopt alternative conformations. Due to the sub-saturating concentrations of peptide, these alternative conformations may correspond to unbound nSH2 domain. However, comparison of unbound and bound nSH2 HSQC spectra reveals the peaks corresponding to alternative conformations are distinct to both and thus may be a 2nd conformation of unbound and/or bound nSH2. The residues are: 21L, 22T, 23R, 24G, 25V, 48N, 49G, 50A, 51V and 98L. These are summarised in Figure 6.3 B and cluster to one affected region of nSH2-Shp2 which are distinct from pTyr and +3 binding pockets. This area of conformational heterogeneity may represent a normal feature of nSH2-Shp2 regardless of binding status. Chemical shift tables for nSH2:ITIM1s can be found in Appendix 10.

Secondary Structure Prediction of nSH2:ITIM1s

TALOS+ was employed to predict the secondary structure of nSH2-Shp2 in the ITIM1s-bound state. Chemical shift values for CO, CA, CB, N and H derived from backbone assignment analysis of nSH2:ITIM1s were used as input. Figure 6.4 summarises the results. Like the nSH2 and cSH2 domains of Shp2 in the unbound state (Chapter IV), the typical SH2 domain fold is observed: two α-helices flanking three β-sheets, with some additional short β-strands predicted throughout the sequence. Secondary structure elements between the unbound and bound forms of nSH2-Shp2 reveal no significant differences between the two. This suggests the ITIM1s binding does not perturb the secondary structure content of nSH2-Shp2. However, no data are available to elucidate the secondary structure of ITIM1s. Since

CD analysis suggests dITIM is almost completely unfolded, ITIM1s is likely to be unfolded too. While this may change in the bound state, most SH2 domain binding partners adopt an extended conformation (Pawson & Schlessinger 1993).

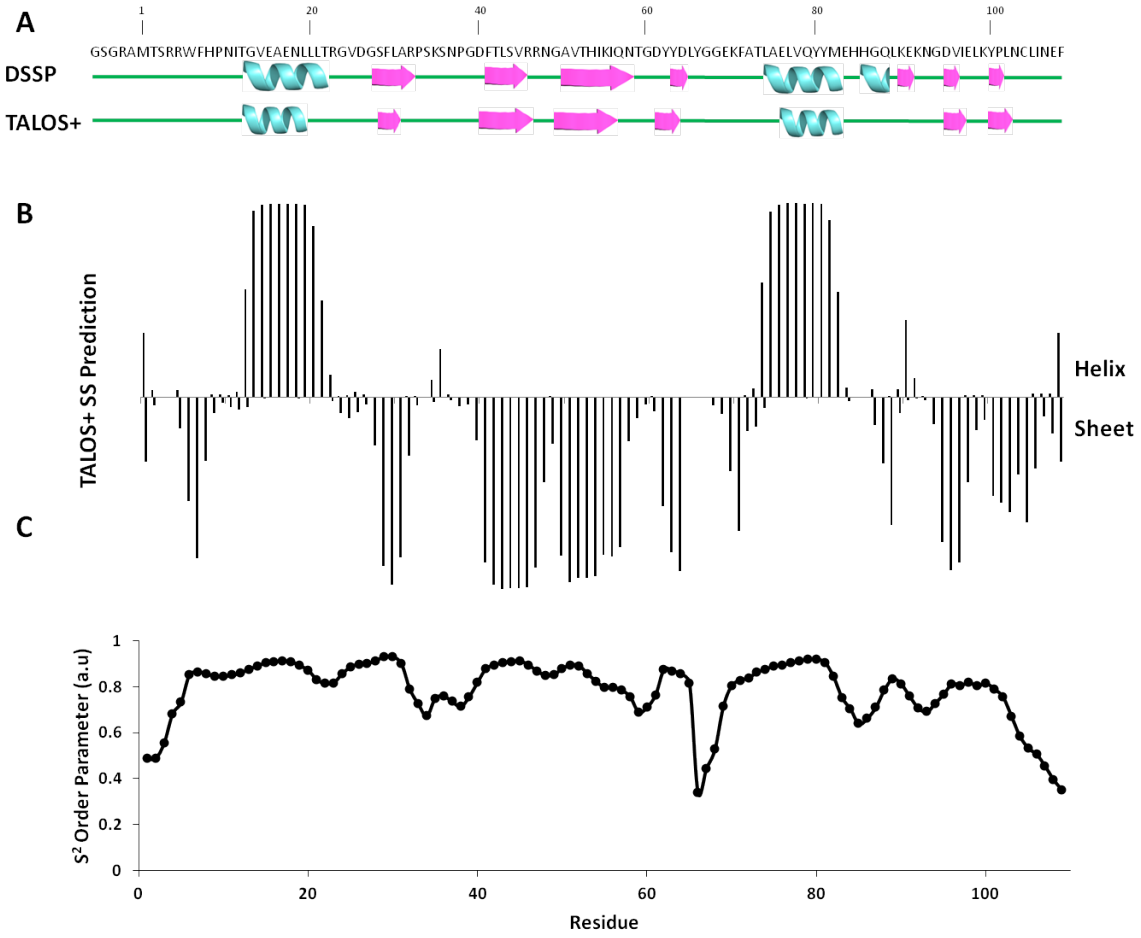


Figure 6.4: nSH2-Shp2 in the ITIM1s Bound State Maintains the Same Secondary Structure to Unbound nSH2-Shp2.

(A) Secondary structure via DSSP of available crystal structure (1AYD) and prediction using TALOS+. Structural elements are depicted as loops (green solid line), helices (cyan) and sheets (pink). Differences between the crystal and solution structure are apparent, mainly as shortened helices and more β-sheets. **(B)** TALOS+ secondary structure probability histogram, helical propensity is positive, sheet propensity is negative and summarised in (A). **(C)** S^2 order parameter prediction using RCI method. The closer the value toward 0, the more mobile the residue.

NMR Titrations of Individual SH2 Domains with ITIM Peptides

Multiple 15N-HSQC titration experiments were recorded on a 600 MHz Varian DD NMR spectrometer in order to identify ligand binding sites. Unlabelled single ITIM peptide was added in 100 µM increments to 2.5-fold excess (unless specified) into an NMR tube containing 400 µM [*U*-^{15}N] cSH2-Shp1 as well as nSH2 and cSH2-Shp2. All experiments were performed in NMR buffer. In-depth analysis of ITIM interactions with various Shp SH2 domains is described further below.

nSH2-Shp2 Binds the ITIM1 in Slow Exchange

More than 30 peaks undergo CSPs upon addition of ITIM1 peptide to 0.25 molar equivalence. All CSPs display slow exchange, indicating a slow on/off rate. By 0.75 molar equivalence most CSPs do not change intensity further (Figure 6.5). CSP analysis reveals that over 40 CSPs above the mean are present, the strongest of which are 54I and 59T. Four peaks corresponding to residues 65L and 66Y (EF loop) as well as 89K and 90E (BG loop), which are known to interact with phosphopeptide residues C-terminal to pY, completely broaden. Figure 6.6 depicts the nSH2-Shp2:ITIM1 interaction and shows many significant perturbations on one face of the SH2 domain, mainly localised to the +3 binding pocket, though CSPs above the mean are noted in C-terminal residues of sheet βB and the BC loop. For the latter, binding information is sparse due to the presence of two proline residues and an unassigned 35K presumably in intermediate exchange in the unbound state. Five new peaks appear throughout the HSQC. Notably, one of these new peaks is located far downfield in the ^{15}N dimension (105.3 ppm) and appears at 0.25 molar

equivalence. Given its location in the HSQC spectrum, this peak could correspond to the previously unassigned 86G residue, though further spectra would need to be recorded to confirm this. ITIM1s and ITIM1 CSPs share similarity, though ITIM1 CSPs undergo slow exchange and stronger shifts.

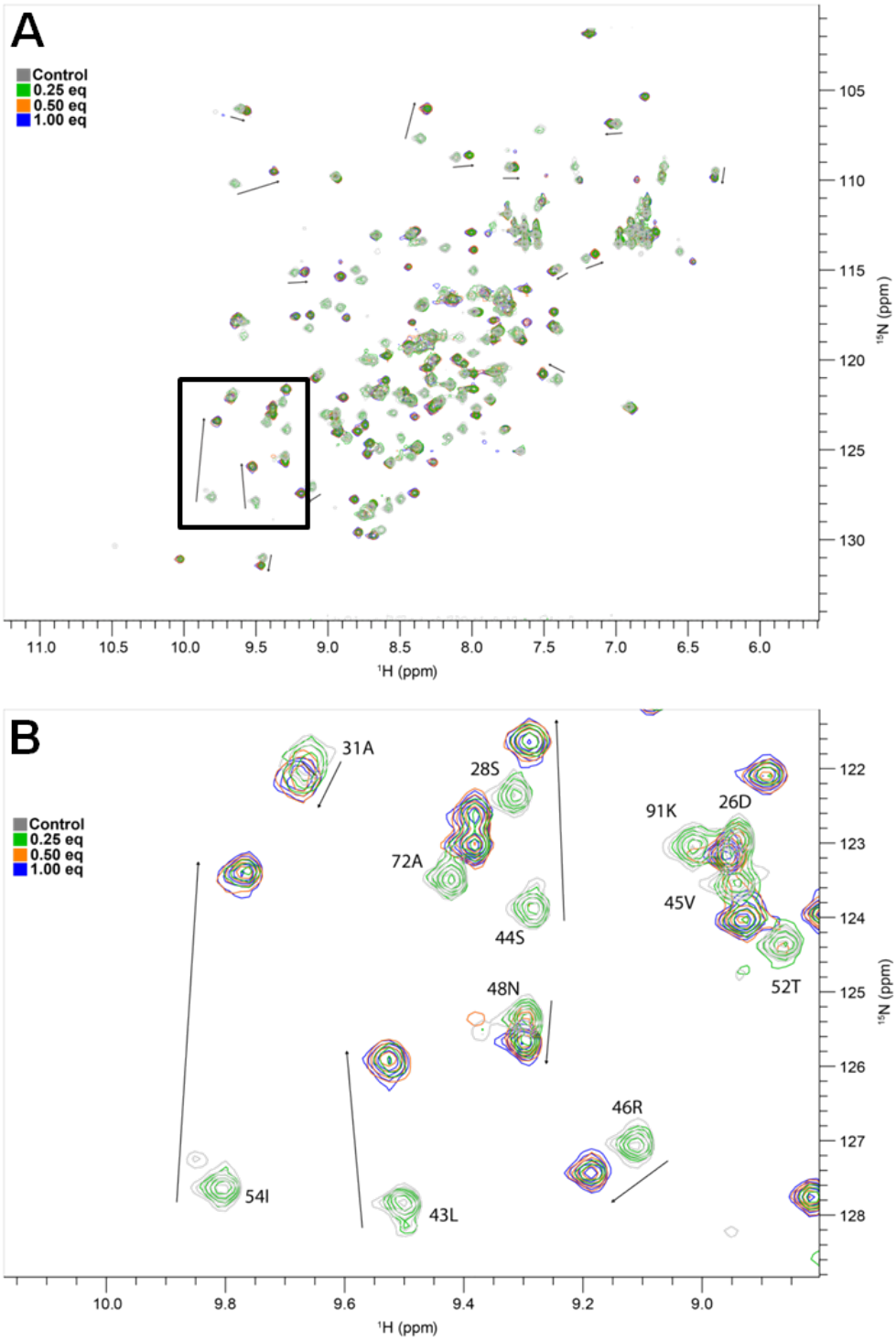


Figure 6.5: NMR Titrations of nSH2-Shp2 with ITIM1 Peptide.

Superimposition of multiple ^{15}N-SOFAST HMQC spectra of 400 µM ^{15}N-labelled nSH2-Shp2 with ITIM1 peptide in 0.25 eq increments. Data were collected on a Varian 600DD NMR Spectrometer at 25 °C in NMR buffer. **(A)** Addition of ITIM1 peptide causes multiple CSPs in nSH2-Shp2. **(B)** Zoomed region of black box in (A) highlighting strongly perturbed residues in slow exchange. Black arrows indicate CSPs as well as relative shift distance.

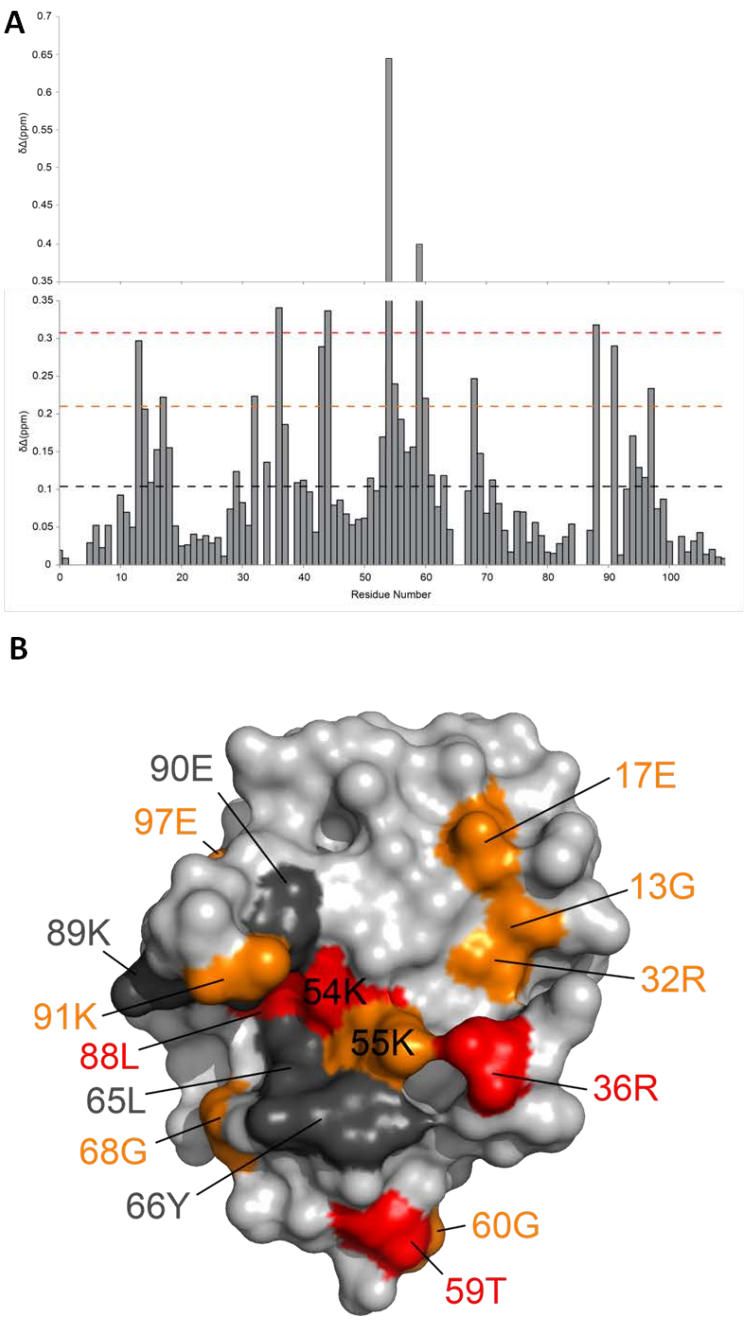


Figure 6.6: nSH2-Shp2 CSPs Induced by ITIM1 Titration.

(A) Histogram (bottom panel) displaying CSPs induced by ITIM1 titration as a function of residue number for 400µM ^{15}N-labelled nSH2-Shp2. Peptide was added to 2.5-fold excess. Dashed lines indicate CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and mean (0.11± 0.10, grey), highlighting strongly perturbed residues. 54I and 59T CSPs are displayed on a separate chart (top panel). **(B)** Surface Representation of nSH2-Shp2 (PDB ID: 1AYD) with ITIM1-induced CSPs mapped on. CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and completely broadened resonances (dark gray) are shown here. Residues 43L and 44S are buried and thus not observed here.

nSH2-Shp2 Binds ITIM2 in Fast Exchange

Upon addition of ITIM2 peptide to 0.25 equivalence to nSH2-Shp2, ≈25 peaks undergo CSPs, with the majority in fast exchange. 43L, 46R, 54I, 68G and 88L shift in intermediate exchange, while 65L and 66Y completely broaden (Figure 6.7). Aside from two new peaks that appear at 2-fold molar excess of ITIM2 peptide, no further CSPs were noted. CSP analysis reveals ITIM2 causes broadening of most assigned signals corresponding to BC loop residues, apart from 34S which undergoes minimal perturbation (Figure 6.8). The largest shifts were noted in sheets βC and βD and the EF loop and show that ITIM2 engages nSH2 in on one face, in the canonical pY- and +3 binding pockets.

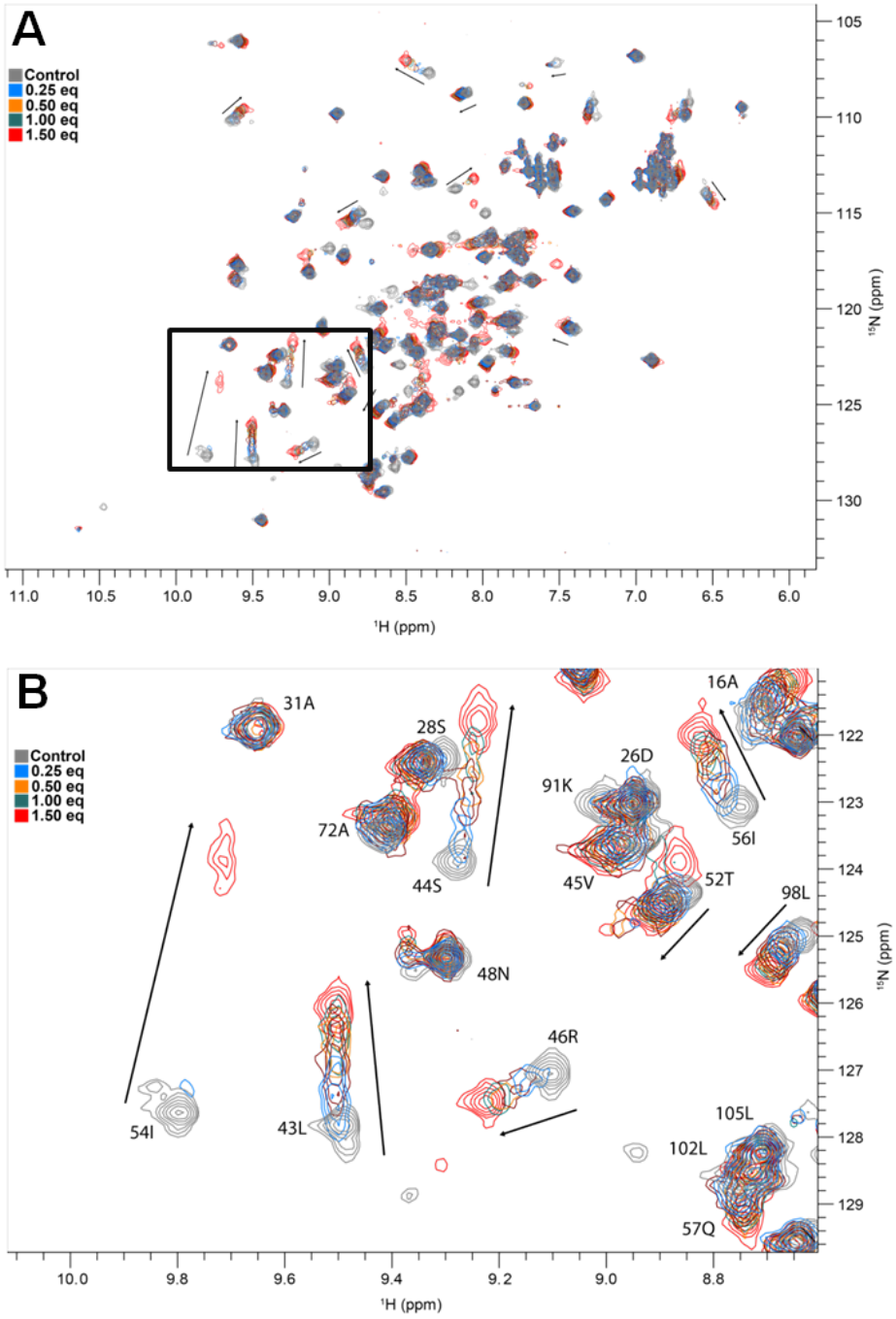


Figure 6.7: NMR Titrations of nSH2-Shp2 with ITIM2 Peptide.

Superimposed ^{15}N-SOFAST HMQC titration series of 400µM ^{15}N-labelled nSH2-Shp2 with unlabelled ITIM2 peptide collected on a Varian 600DD NMR Spectrometer at 25 °C in NMR buffer. **(A)** 0-1.50 eq titration points. Multiple CSPs of different magnitudes and exchange rates are observed, some of which are highlighted with black arrows. CSPs halt after 1.50 eq, Grey and red spectra represent start and end (1.5 eq) points for the titration. **(B)** Zoomed region of black box in (A), highlighting CSPs in intermediate-fast exchange.

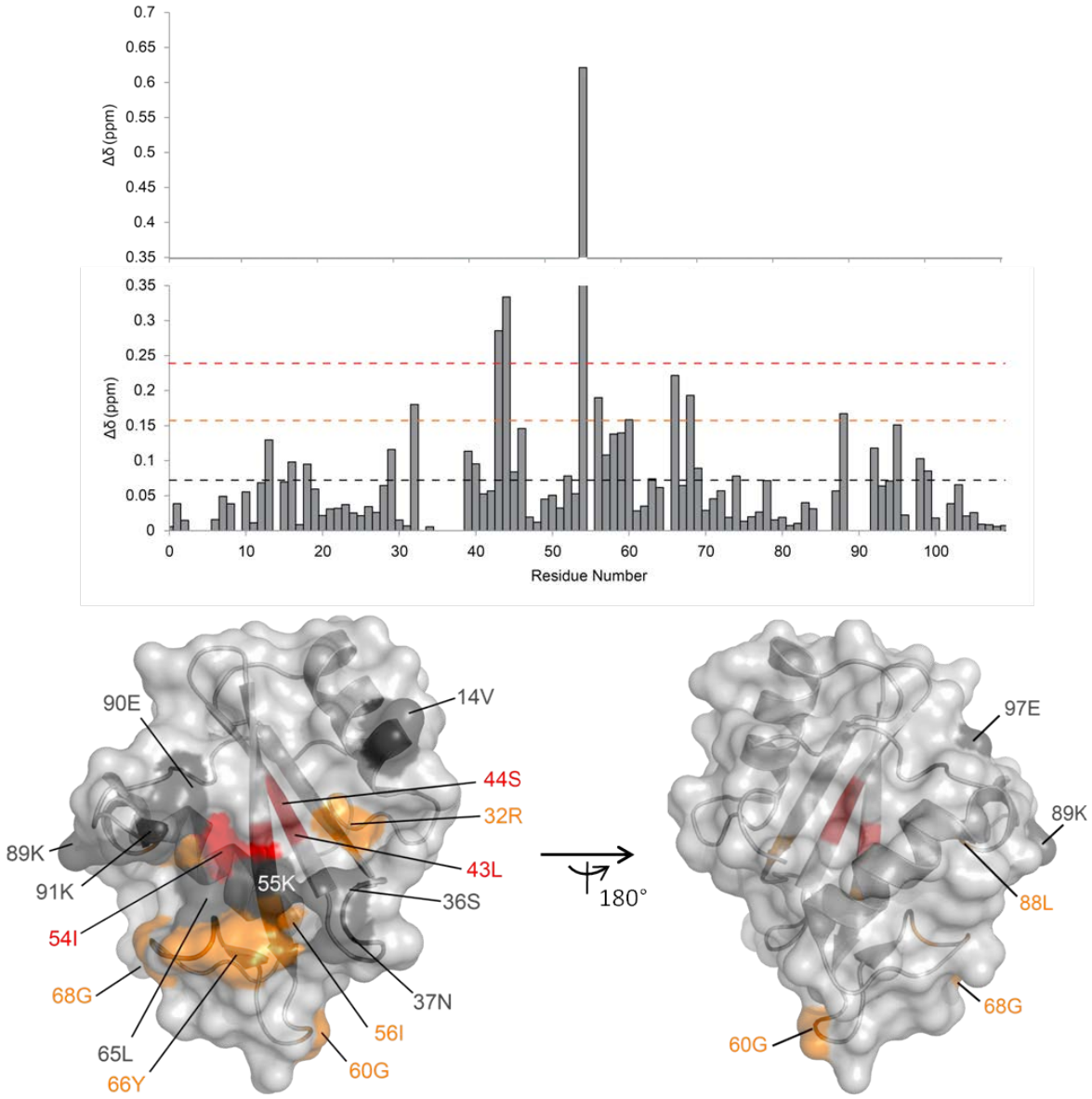


Figure 6.8: Histogram and Binding Surface of nSH2-Shp2 CSPs Induced by ITIM2.

(**A**) The histogram displays CSPs induced by ITIM2 titration as a function of residue number for 400 µM ^{15}N-labelled nSH2-Shp2. Peptide was added to 2.5-fold excess. Dashed lines indicate ≥ mean + 2 S.D (red), ≥ mean+ 1 S.D (orange) and mean (0.071± 0.09, grey) CSPs, highlighting strongly perturbed residues. 54I displayed a CSP of 0.64 ppm and is shown on a separate chart. (**B**) Cartoon and transparent surface representation of nSH2-Shp2 (PDB ID: 1AYD) with ITIM2-induced CSPs mapped on. CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and completely broadened resonances (gray) are shown here.

nSH2-Shp2 Binds ITIM1s in Intermediate-Slow Exchange

The truncated ITIM1 peptide ITIM1s was titrated to 2.5 fold molar excess, although the concentration of [*U*-^{15}N]-nSH2-Shp2 was lower in this experiment (125 μM) than previously. At 0.25 molar equivalence of ITIM1s peptide, ≈20 peaks experience intensity reduction, with minor perturbations in ppm values. 66Y and 67G broaden significantly (Figure 6.9). Further addition of ITIM1s peptide led to CSPs in a mixture of chemical exchange states. Residues 43L and 54I are in slow exchange, whilst 41F is in intermediate exchange, with many other peaks in fast exchange. Negligible CSPs are noted after 1.5-fold molar excess of ITIM2 peptide. CSP analysis reveals that residues within helix αA, sheets βC and βD undergo significant CSPs, with broadening of BC and EF loop signals too (Figure 6.10). Mapping of perturbed residues onto the nSH2 structure (PDB ID: 1AYD) reveals a significant binding patch in residues forming and surrounding the pY-pocket with other perturbations in the BG loop. Interestingly, two residues outside of the aforementioned patches also display binding phenomena. Residue 59T broadens whilst 61D shifts in slow exchange. Both residues are located in the DE loop. Specifically, the main chain amide of D61 shifts upfield and thus becomes shielded in the bound state.

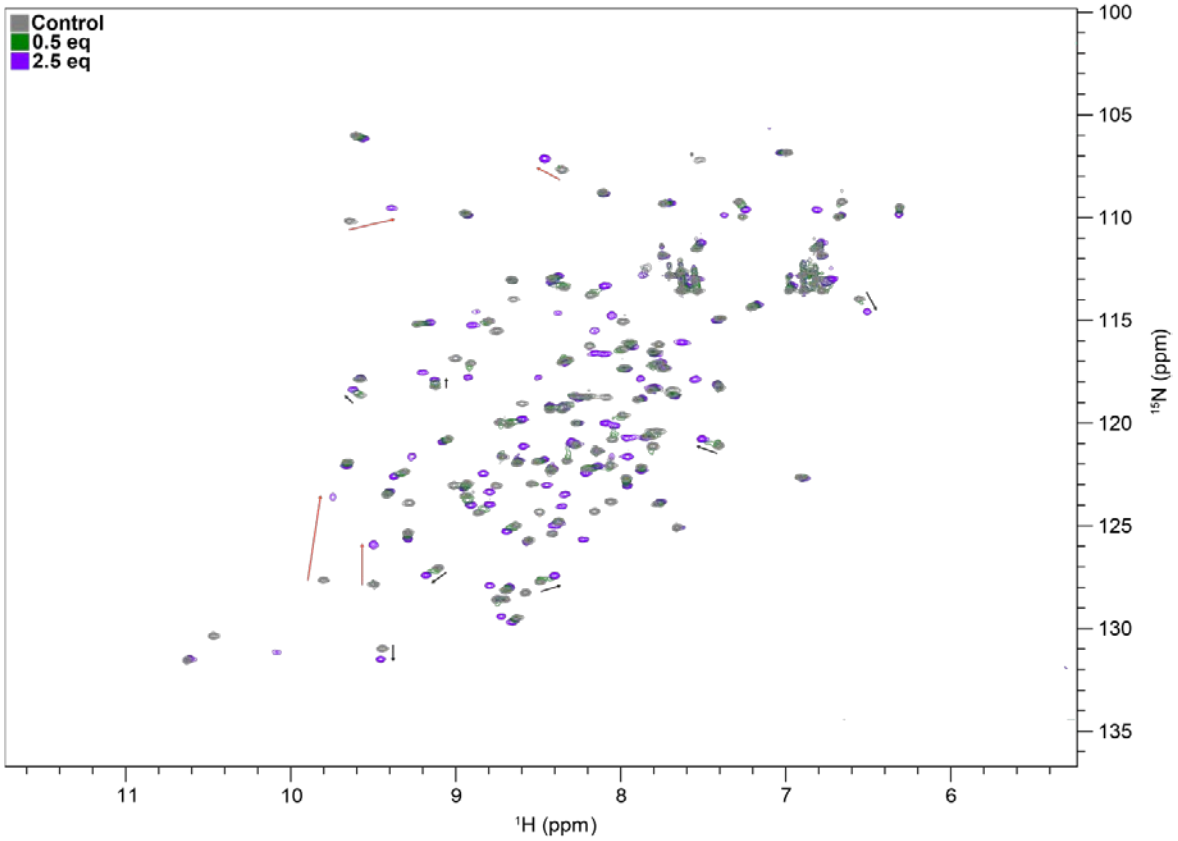


Figure 6.9: nSH2-Shp2 CSPs Induced by ITIM1s Titration.

Superimposed ^{15}N-HSQC spectra of unbound (grey), 0.5 eq (green) and 2.5 eq (purple) 125 µM ^{15}N-nSH2-Shp2 titration experiments with ITIM1s peptide performed at 25 °C on a Varian 600DD NMR spectrometer in NMR buffer. Arrows (red: slow exchange, black: fast exchange) indicate CSPs and direction.

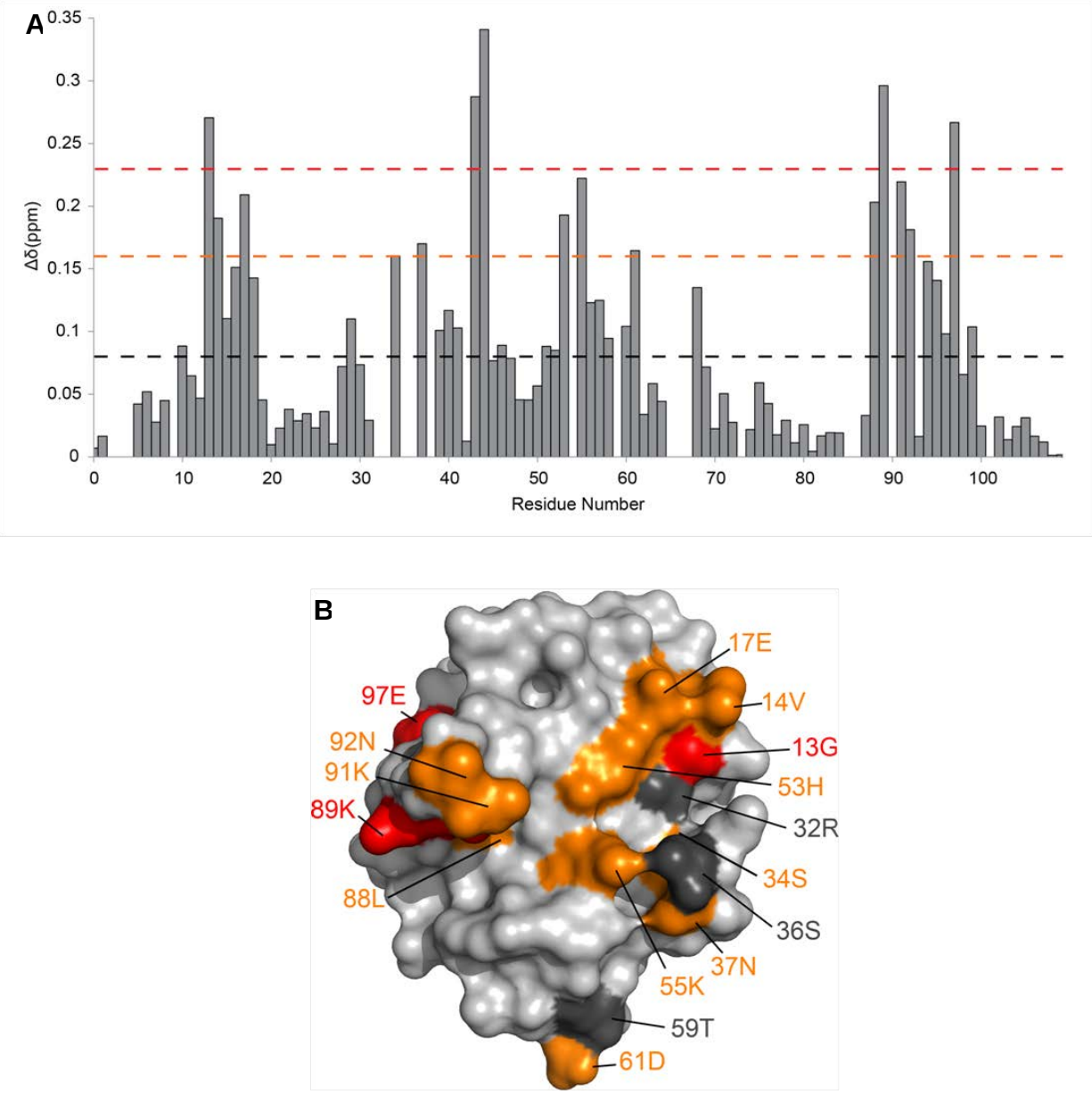


Figure 6.10: Histogram and Binding Surface of nSH2-Shp2 CSPs Induced by ITIM1s.

(A) The histogram displays CSPs induced by ITIM1s titration as a function of residue number for 125 µM ^{15}N-labelled nSH2-Shp2. Peptide was added to 2.5-fold excess. Dashed lines indicate ≥ mean + 2 S.D (red), ≥ mean+ 1 S.D (orange) and mean (0.083± 0.077, grey) CSPs, highlighting strongly perturbed residues. **(B)** Surface Representation of nSH2-Shp2 (PDB ID: 1AYD) with ITIM1s-induced CSPs mapped on. CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and completely broadened resonances (gray) are shown here.

cSH2-Shp2 Binds ITIM1 in Multiple Exchange Timescales

ITIM1 peptide titration causes ≈20 CSPs across all observable NMR timescales, with no further shift changes around 1.75-fold molar excess (Figure 6.11). The majority of CSPs shift in fast exchange; peaks in slow exchange include 119G, 154G, 170V and 183G. These residues are located in helix αA, CD loop, sheet βD and EF loop respectively and apart from 154G, are conserved regions of direct interaction with phosphopeptides. It is therefore likely that conformational changes during ITIM1 binding cause line broadening of 154G. CSP analysis reveals perturbations across most structural elements of cSH2-Shp2, such as helix αA, sheets βC and βD and the BG loop (Figure 6.12). Mapping of significant residues onto the crystal structure (PDB ID: 4DGP) reveals residues perturbed on both faces of the SH2 domain (On the side containing both pY and +3 pockets, perturbed residues populate both regions, though more are observed in the latter). However, perturbations are also observed along BG loop residues which reach behind the SH2 domain and back toward the extreme N-terminal part of the cSH2 sequence.

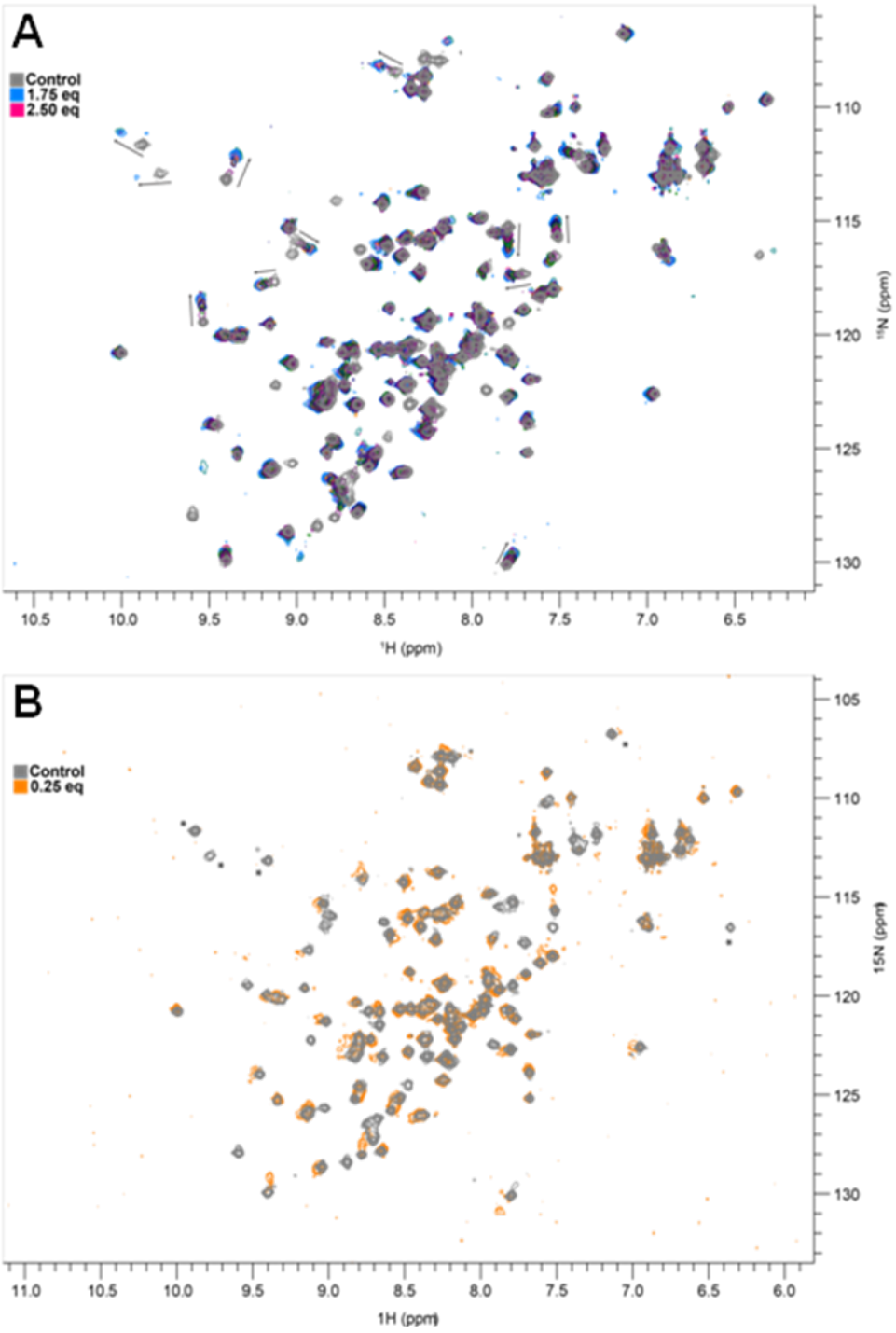


Figure 6.11: NMR Titrations of cSH2-Shp2 with ITIM1 Peptide.

Superimposed ^{15}N-SOFAST HMQC titration series of 400µM ^{15}N-labelled cSH2-Shp2 with unlabelled ITIM1 peptide up to 2.5-fold molar excess. Data were collected on a Varian 600DD NMR Spectrometer at 25 °C in NMR buffer. **(A)** 0-1.75 eq titration points. Multiple CSPs of different magnitudes and exchange rates are observed, some of which are highlighted with black arrows. CSPs halt after 1.75 eq, though 2.50 eq is also shown. Grey and pink spectra represent start and end (2.5 eq) points for the titration. **(B)** Control and 0.25 eq titration points show initial CSPs changes upon ITIM1 binding.

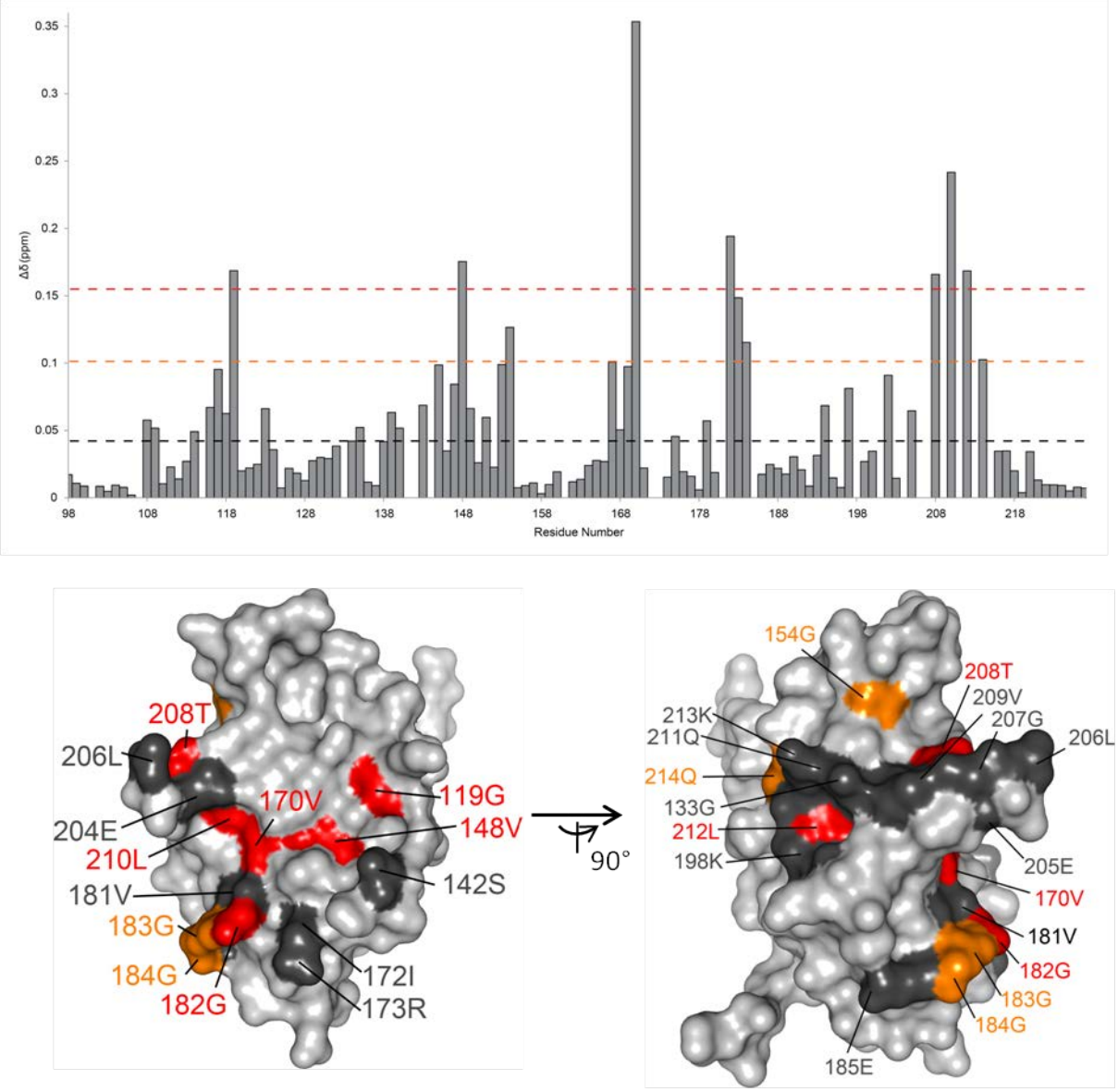


Figure 6.12: Histogram of cSH2-Shp2 CSPs Induced by ITIM1.

(A) The histogram displays CSPs induced by ITIM1 titration as a function of residue number for 400μM ^{15}N-labelled cSH2-Shp2. Peptide was added to 2.5-fold excess. Dashed lines indicate ≥ mean+ 2 S.D (red), ≥ mean + 1 S.D (orange) and mean (0.04± 0.06, grey) CSPs, highlighting strongly perturbed residues. **(B)** Surface Representation of cSH2-Shp2 (PDB ID: 4DGP) with ITIM1-induced CSPs mapped on. CSPs ≥ mean + 2 S.D (red), ≥ mean + 1 S.D (orange) and completely broadened resonances (gray) are shown here.

cSH2-Shp2 Binds ITIM2 in Slow Exchange

Initial addition of ITIM2 peptide to 0.25 molar equivalence causes ≈six intermediate-slow exchange CSPs and complete broadening of signals for residues 119G, 135F, 154G, 170V, 182G, 197Y and 213K (Figure 6.13). Further titration causes multiple peaks to undergo CSPs in slow exchange, with no additional shifts after equimolar concentrations are achieved. Assignment of peaks in slow exchange was performed using data from ITIM1 titration, since it was found that many common residues between both titration series undergo CSPs in the same general direction. CSP analysis thereafter reveals strong perturbations from residues 170V and 210L, the latter of which broadened early in the titration only to reappear in intermediate exchange at 0.75 molar equivalence. ≈25 CSPs are above the mean, with many broadened completely, such as those in BC and BG loops (Figure 6.14). Mapping of significant CSPs onto the cSH2-Shp2 structure reveals a similar pattern as observed for the cSH2-Shp2:ITIM1 interaction. Both pY and +3 pocket residues are perturbed as well as residues on the opposite face, found within the C-terminal of the BG loop.

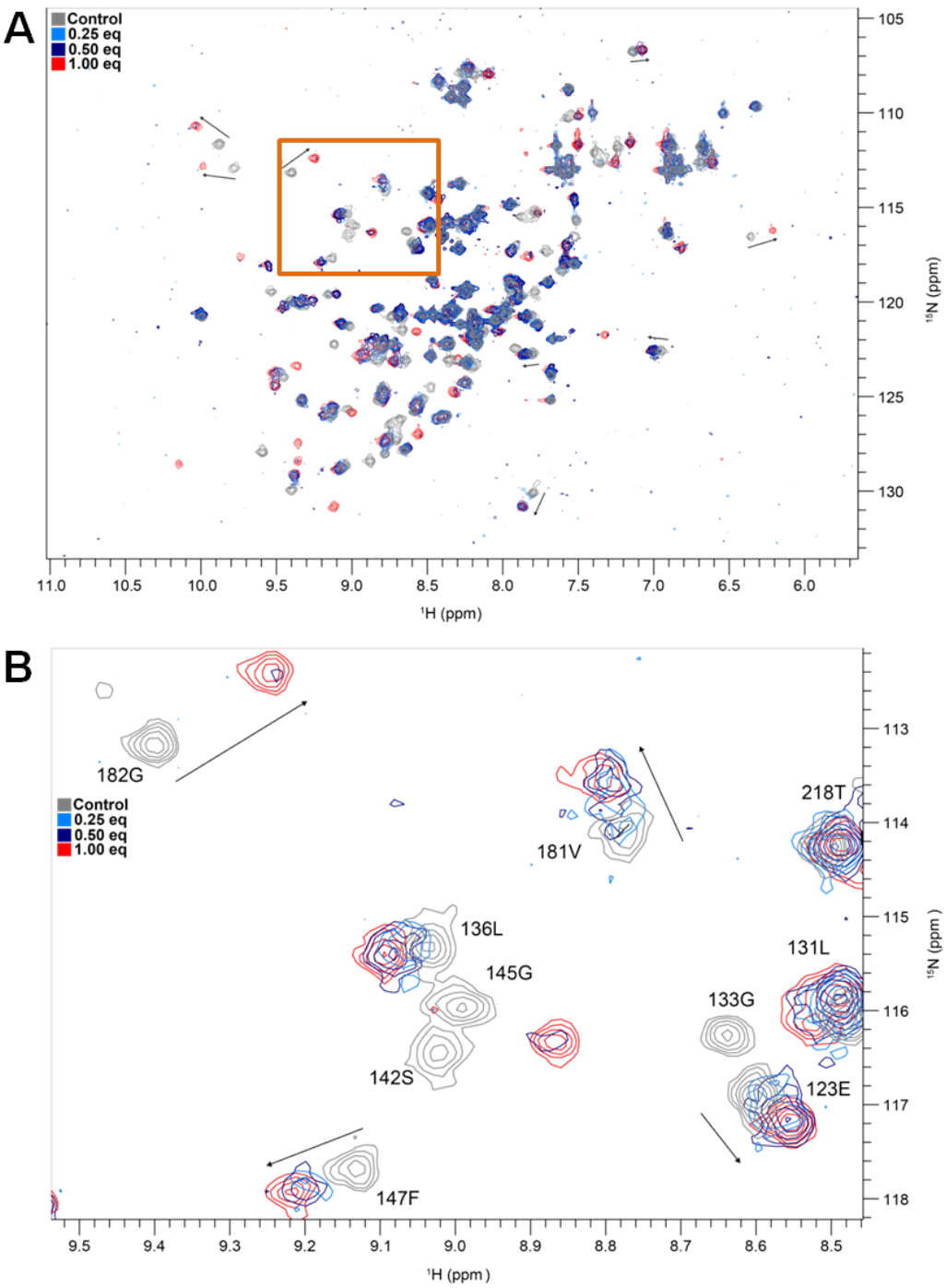


Figure 6.13: NMR Titrations of cSH2-Shp2 with ITIM2 Peptide.

Superimposed ^{15}N-SOFAST HMQC titration series of 400µM ^{15}N-labelled cSH2-Shp2 with unlabelled ITIM2 peptide up to 2.5-fold molar excess. Data were collected on a Varian 600DD NMR Spectrometer at 25 °C in NMR buffer. **(A)** 0-1.00 eq titration points in 0.25 eq increments. Multiple CSPs of different magnitudes and exchange rates are observed, some of which are highlighted with black arrows. CSPs halt by 1 eq. Grey and red spectra represent start and end points. **(B)** Zoomed region from orange box in (A) show CSPs upon ITIM2 binding in more detail.

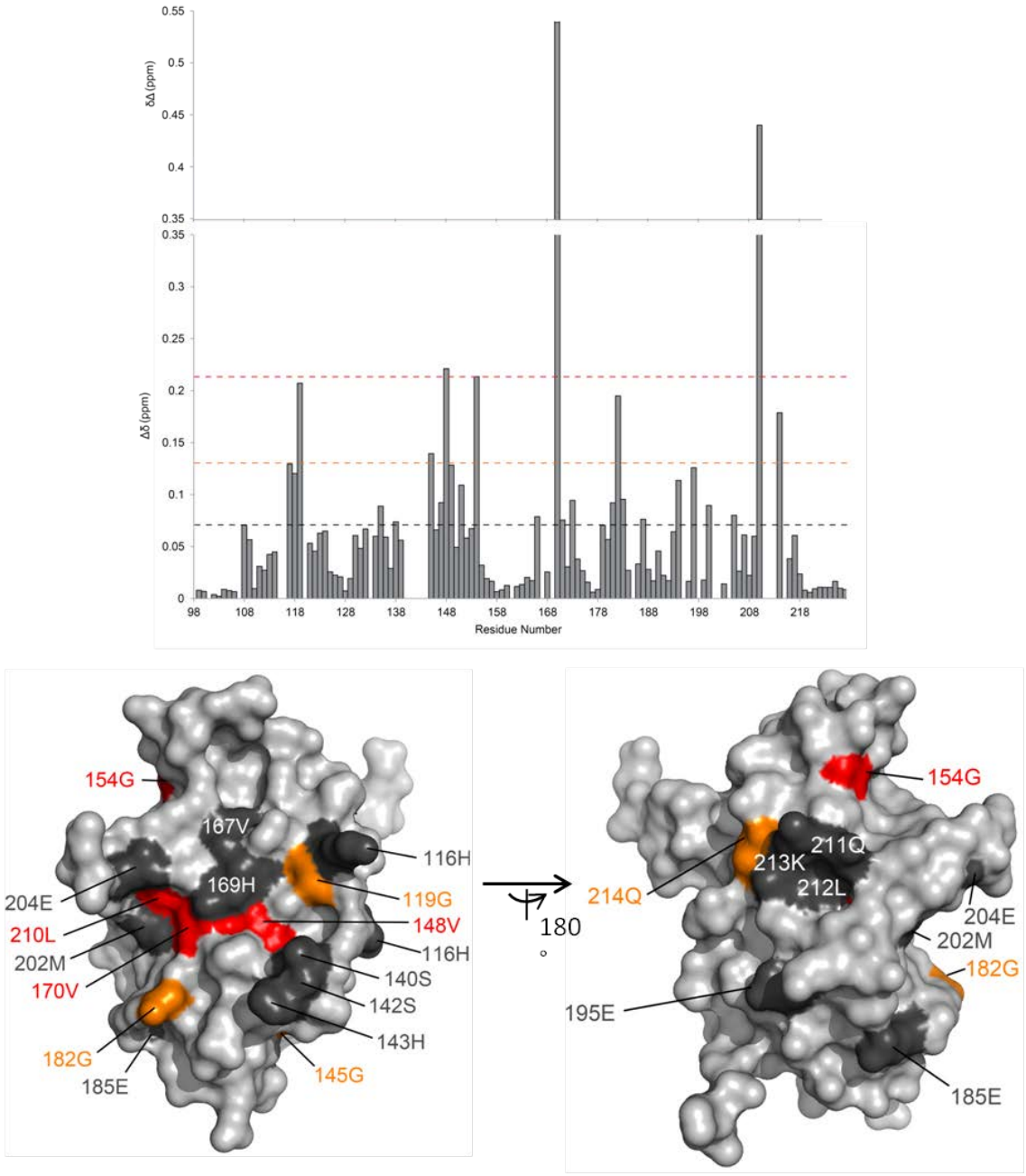


Figure 6.14: Histogram of cSH2-Shp2 CSPs Induced by ITIM2.

(A) The histogram displays CSPs induced by ITIM2 titration as a function of residue number for 400µM ^{15}N-labelled cSH2-Shp2. Peptide was added to 2.5-fold excess. Dashed lines indicate CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and mean (0.06± 0.08, grey), highlighting strongly perturbed residues. 170V and 210L displayed CSPs of 0.54 and 0.44 ppm respectively and are displayed on a separate chart. (**B)** Surface Representation of cSH2-Shp2 (PDB ID: 4DGP) with ITIM2-induced CSPs mapped on. CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and completely broadened resonances (gray) are shown here.

Interestingly, 195E in position αB7 broadens although the side-chain faces away from the SH2 core and is in the middle of a helix. 194V and 197Y are also perturbed above the mean and are also located in helix αB. While the side-chain of 197Y forms part of the +3 pocket, 194V and 195E do not and instead may undergo CSPs due to helix repositioning in the bound state. Since the residues C-terminal to pY in ITIM2 are very hydrophobic (three consecutive valine residues), packing of their side-chains into the +3 pocket may induce such changes.

cSH2-Shp2 Binds ITIM1s in Fast Exchange

cSH2-Shp2 titration with ITIM1s was performed in the same manner as with nSH2-Shp2 insofar as 125 μM [*U*-^{15}N]-protein was titrated step-wise with ITIM1s peptide until no further shifts were observed. ≈20 peaks experienced CSPs on the fast exchange timescale, with 170V the only residue to shift in intermediate-slow exchange (Figure 6.15). No significant shifts were noted up to 3.5-fold molar excess (437.5 μM) of ITIM1s peptide.

CSP analysis reveals 31 CSPs above the mean, mainly localising to sheet βD and the BC and BG loops (Figure 6.16). Interestingly, only two residues from the EF loop region undergo significant CSPs: 181V and 182G. Additionally, 179Y in the preceding short β sheet (βE) is perturbed just above the mean. Mapping of strong perturbations onto the cSH2-Shp2 structure (PDB ID: 4DGP) reveals a binding patch perpendicular to the central three large β-sheets, in both pY and +3 pockets. Additional shifts are displayed on the opposite face to the canonical pockets; these are mainly BG loop residues.

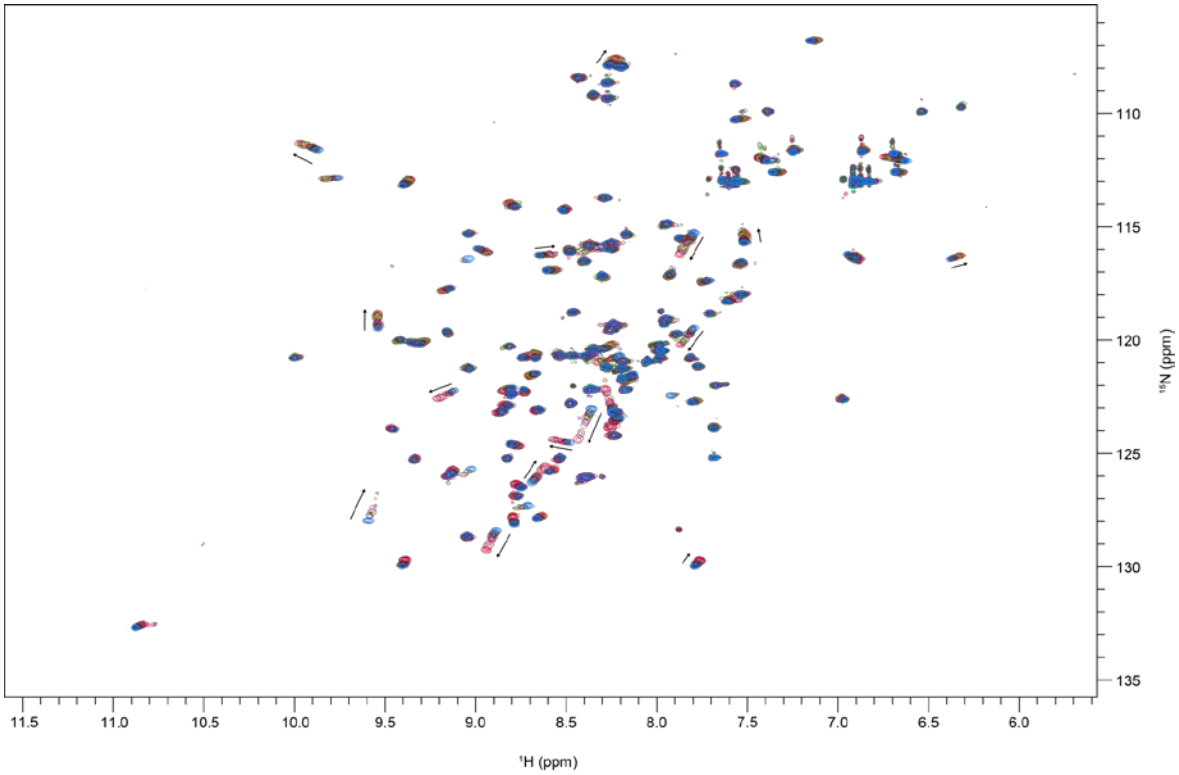


Figure 6.15: cSH2-Shp2 Titration with ITIM1s Peptide by NMR.

Superimposed ^{15}N-SOFAST HMQC titration series of 400µM ^{15}N-labelled cSH2-Shp2 with unlabelled ITIM1s peptide up to 2.5-fold molar excess. Data were collected on a Varian 600DD NMR Spectrometer at 25 °C in NMR buffer. Multiple CSPs of different magnitudes and exchange rates are observed, some of which are highlighted with black arrows. Blue and red spectra represent start and end (2.5 eq) points for the titration.

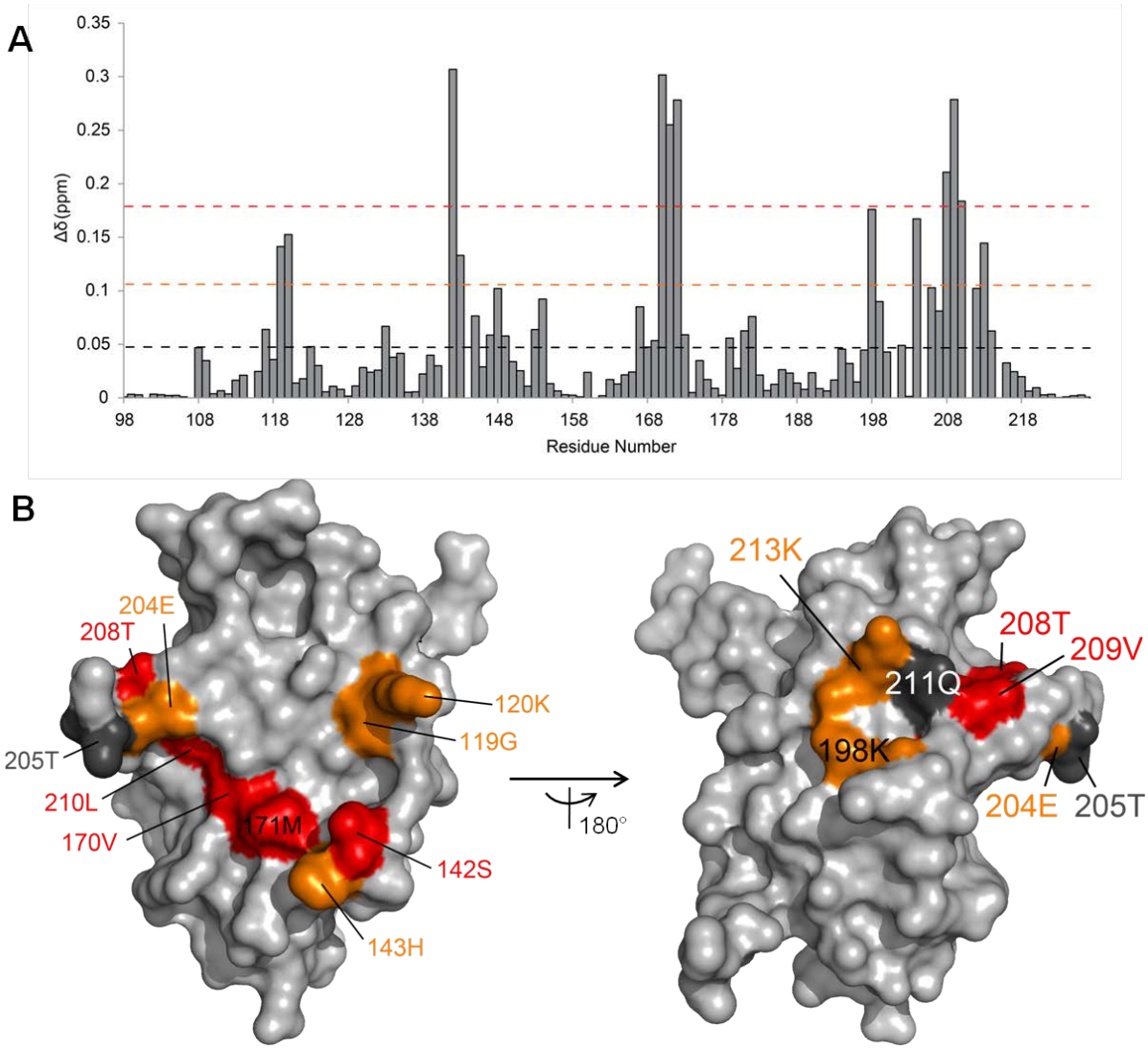


Figure 6.16: cSH2-Shp2 CSPs Induced by ITIM1s Titration.

(A) Histogram displaying CSPs induced by ITIM1s titration as a function of residue number. Dashed lines indicate CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and mean (0.05± 0.07, grey), highlighting strongly perturbed residues. **(B)** Surface representation of cSH2-Shp2 (PDB ID: 4DGP) with ITIM1s-induced CSPs mapped on. CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and completely broadened resonances (gray) are shown here.

cSH2-Shp1 Interactions with ITIM Peptides

For cSH2-Shp1 titrations, 100 µM [*U*-^{15}N]-protein in NMR buffer was titrated with unlabelled ITIM1 or ITIM peptides, as performed in previous experiments. Residue specific assignments were not available for TanSH2-Shp1 due to its poor expression/solubility so only general comments on binding phenomena are described here.

cSH2-Shp1 Binds ITIM1 in Fast Exchange

Addition of ITIM1 peptide to 0.25 molar equivalence results in around 17 weak CSPs in a fast exchange regime (Figure 6.17 A). These peaks continue to be perturbed up until the end of the titration series where 1.5-fold molar excess of ITIM1 peptide had been added. At this point ≈20 peaks displayed significant perturbations. Very weak perturbations were observed for the peak around 11 ppm which likely corresponds to Hε1 of either 110W or 131W. In addition, two side-chain resonances belonging to either Asn or Gln undergo CSPs. It is likely that titration up to saturation was not performed as the experimental end point still showed CSPs; in general titrations are continued until two identical HSQC spectra are observed at the endpoint.

cSH2-Shp1 Binds ITIM2 in Fast-Intermediate Exchange

Initial titration of cSH2-Shp1 with ITIM2 peptide to 0.25 molar equivalence causes ≈35 CSPs, including possible Trp and Asn/Gln side-chain amide resonances (Figure 6.17 B). CSPs continue to shift until equimolar concentrations of ITIM2 are present.

Comparison of ITIM1 and ITIM2 CSPs with cSH2-Shp1 reveal that similar peaks are perturbed by addition of peptide. The majority of peaks move in the same direction regardless of which ITIM peptide is added and in general, ITIM2 causes larger shifts than ITIM1. Although the NMR solution structure of cSH2-Shp1 has been elucidated, assignment data are not yet available on BMRB, PDB or any associated repositories, hindering further analysis.

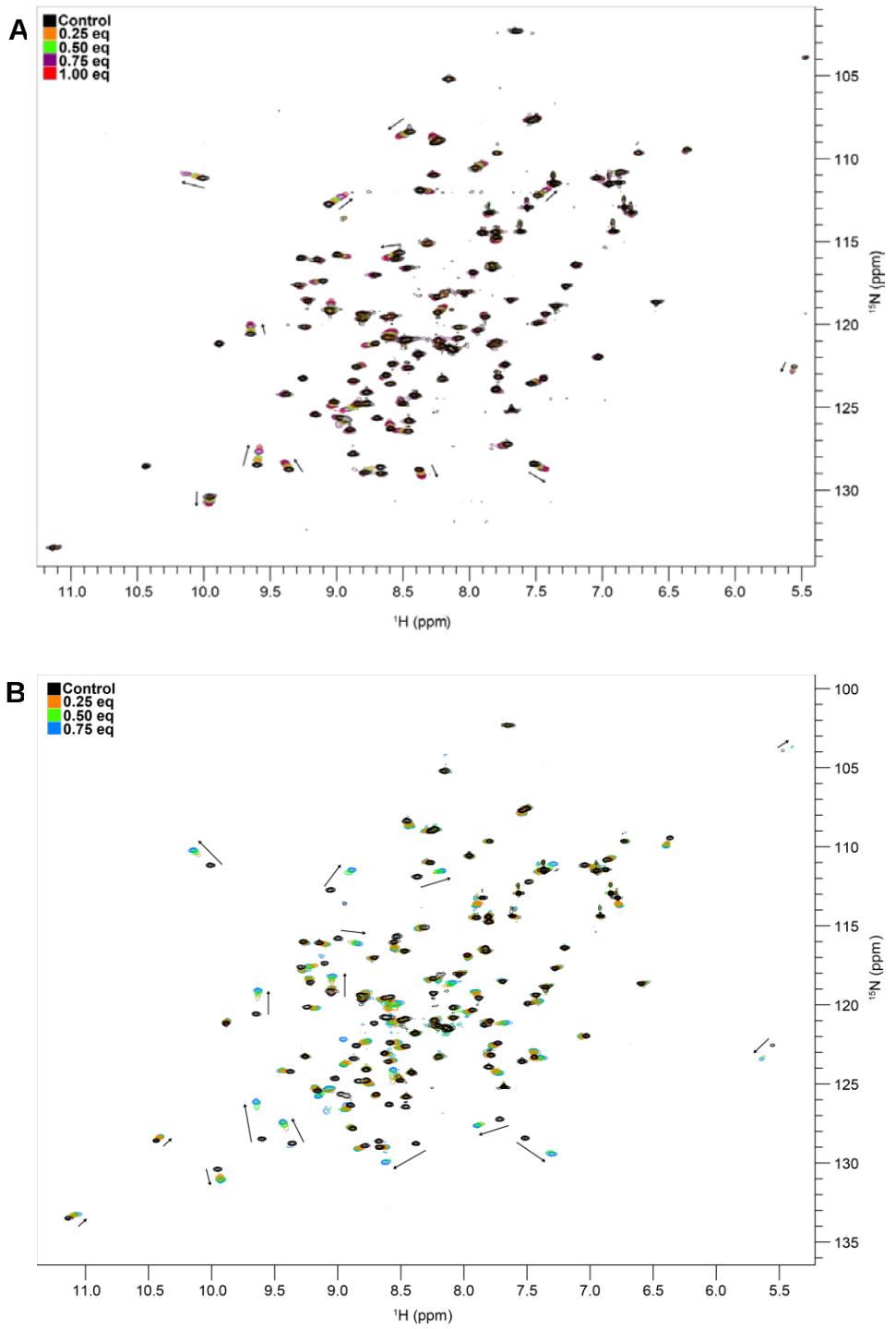


Figure 6.17: cSH2-Shp1 Titrations with ITIM Peptides.

ITIM1 **(A)** and ITIM2 **(B)** peptide were titrated into 100 µM [*U*-^{15}N]-cSH2-Shp1 in 25 µM increments with ^{15}N-HSQC spectra recorded at each point until no further CSPs were observed. Spectra were then superimposed. Black arrows indicate CSPs. Data were collected on a Varian 600DD NMR Spectrometer at 25 °C in NMR buffer.

NMR Titrations of TanSH2 with ITIM Peptides

Multiple ^{15}N-HSQC titration experiments were recorded on an 800 MHz Varian NMR spectrometer. Unlabelled single ITIM or dITIM peptide was added in 100 µM increments to 2.5 and 1.5-fold molar excess respectively into an NMR tube containing 400 µM [*U*-^{15}N] His-TanSH2-Shp2 in NMR buffer. Due to issues with peptide solubility, some precipitation was observed during the titration series, especially with dITIM. In-depth analysis of ITIM interactions with TanSH2:Shp2 is described further below.

TanSH2:ITIM1

Titration of TanSH2-Shp2 with ITIM1 causes around 20 CSPs. Crucially, some of these peaks display CSPs in slow exchange, such as 22T, 48N, 49G and 73T, whilst others show relatively minute CSPs in fast exchange (Figure 6.18), this is particular evident below equimolar concentrations of TanSH2 and ITIM1. A number of peaks also broaden completely such as 7F, 8H, 13G, 31A and 58N. At this point in the titration, CSPs and broadened peaks localise to the nSH2 domain. When molar equivalence is reached, peaks undergoing shifts in slow exchange no longer display intensity changes. Above molar equivalence and up to the end of ITIM1 titration, a number of peaks undergo CSPs in fast exchange (Figure 6.19).

CSP analysis reveals that addition of ITIM1 peptide to equimolar concentration results in perturbations localised to nSH2 domain residues (Figure 6.20). However cSH2 domain CSPs occur above equimolar concentration (1 eq and above) (Figure

6.20). Mapping of residues above mean + 1 and 2 S.Ds reveals around 26 assigned peaks undergo CSPs with Δδ values between the mean and mean + 1 S.D, just over 25% of residues, while 28 peaks completely broadened, including residues between positions 58-61 (DE loop; 59T did not broaden but was unassignable in the bound state), 65-67 (EF loop) and 88-91(BG loop), inclusively. A number of assigned, overlapped peaks could not be assigned in the bound state since there is no way to discriminate which peak shifted, these were: 17E, 21L, 34S, 59T, 74L, 76E, 81Y, 92N, 106D, 109S, 127T, 191T, 197Y and 208T.

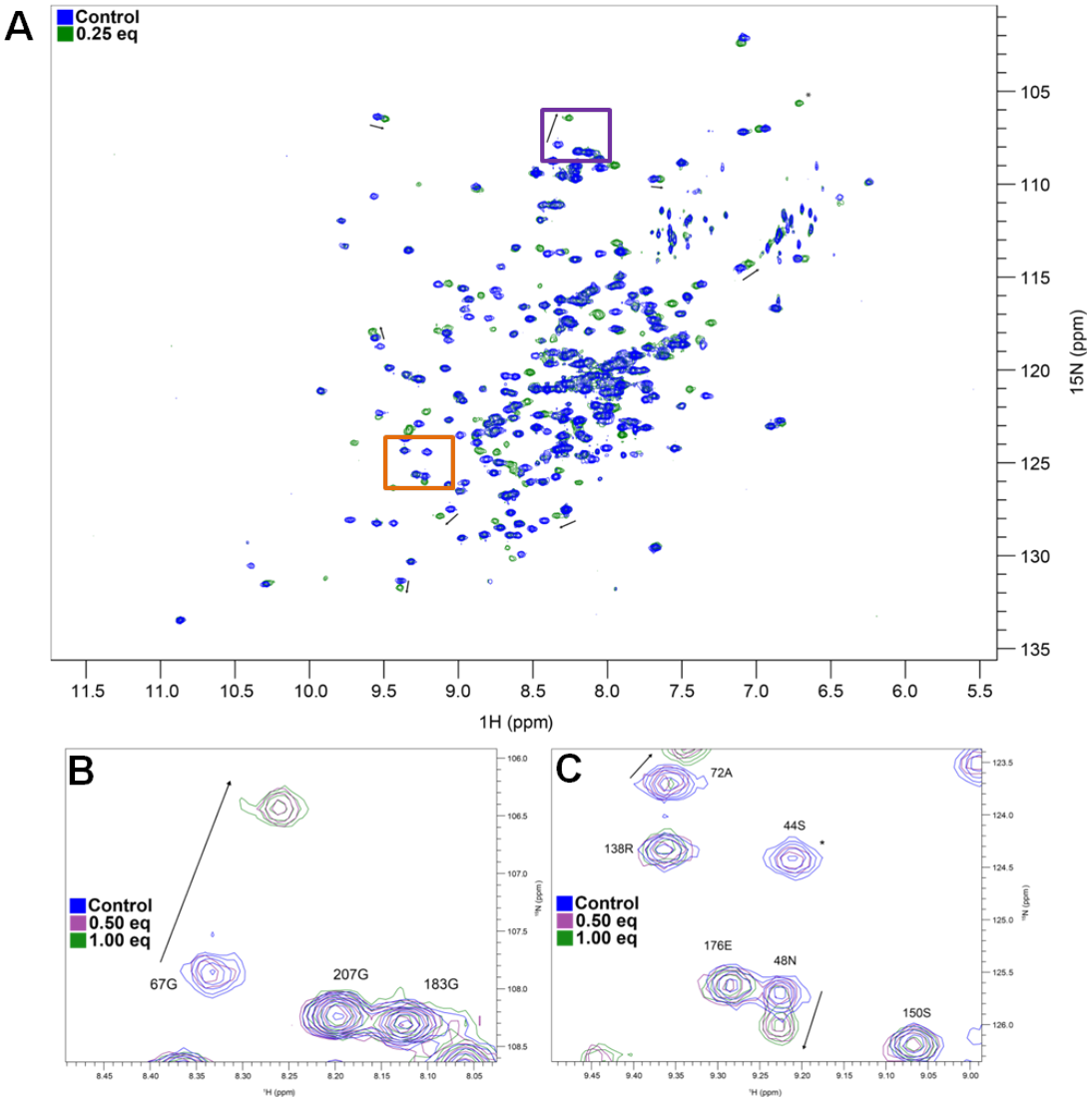


Figure 6.18: TanSH2-Shp2 Interacts with ITIM1 Peptide.

Superimposed ^{15}N-HSQC Spectra of 400 µM ^{15}N-labelled His-TanSH2-Shp2 titrated with varying amounts of ITIM1 peptide. Data were collected on a Varian 800 NMR Spectrometer at 25 °C in NMR buffer. **(A)** Control and ITIM1-bound (1:1 ratio/molar equivalence) His-TanSH2-Shp2. **(B)** Zoomed region of (A), purple box. 67G of the EF loop shifts in slow exchange, whilst cSH2 residues (183G and 207G) do not undergo CSPs. **(C)** Zoomed region of (A), orange box. 48N and 72A of the nSH2 undergo CSPs in slow exchange whilst cSH2 residue peaks such as 138R and 150S do not. Arrows indicate direction of CSPs and asterisks show areas of peak broadening.

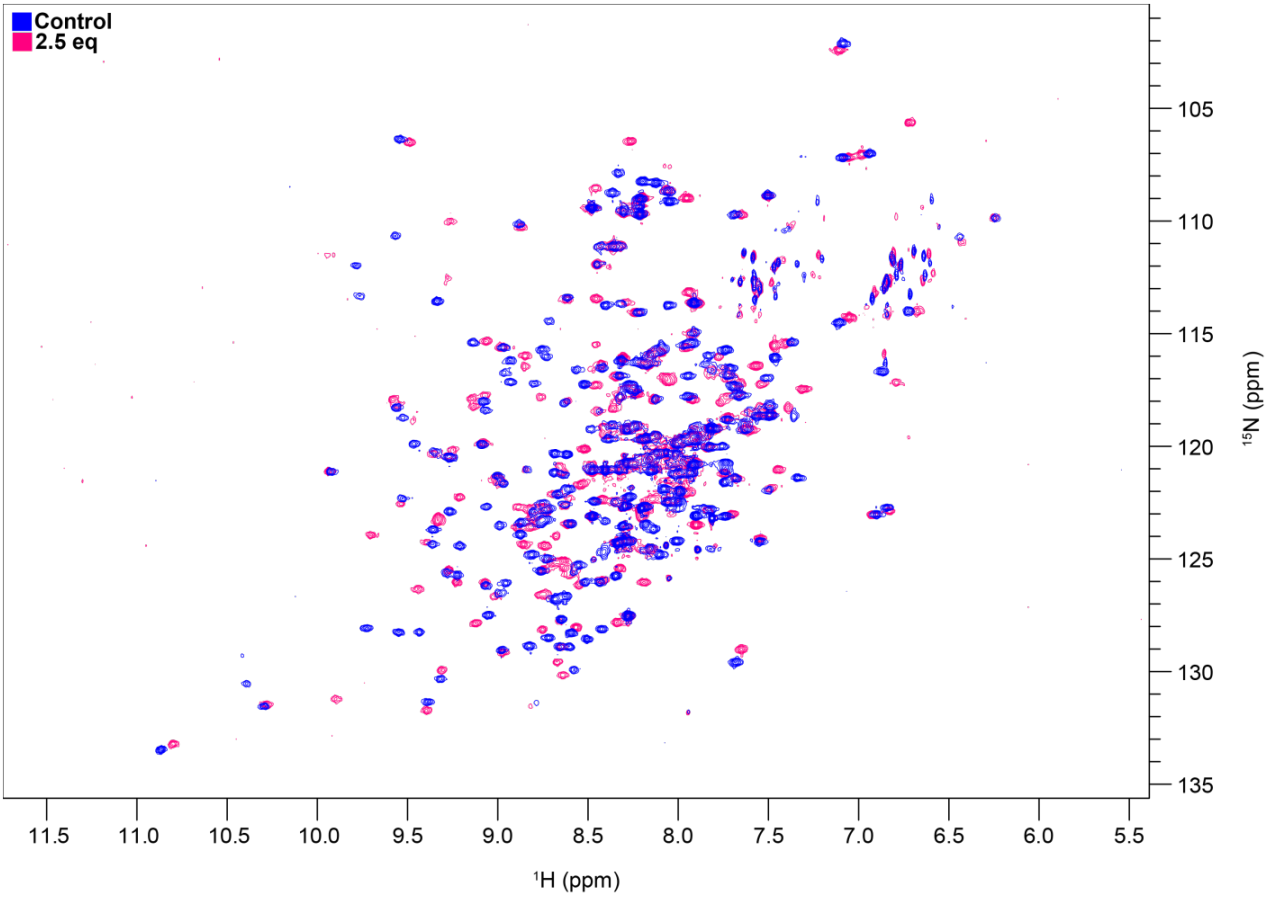


Figure 6.19: Addition of ITIM1 Peptide to Excess Causes Further CSPs in His-TanSH2-Shp2.

400 µM ^{15}N-labelled His-TanSH2-Shp2 (blue) was titrated to 2.5-fold molar excess with ITIM1 peptide (pink), causing many peaks to undergo CSPs in slow and fast exchange regimes. Data were collected on a Varian 800 NMR Spectrometer at 25 °C in NMR buffer.

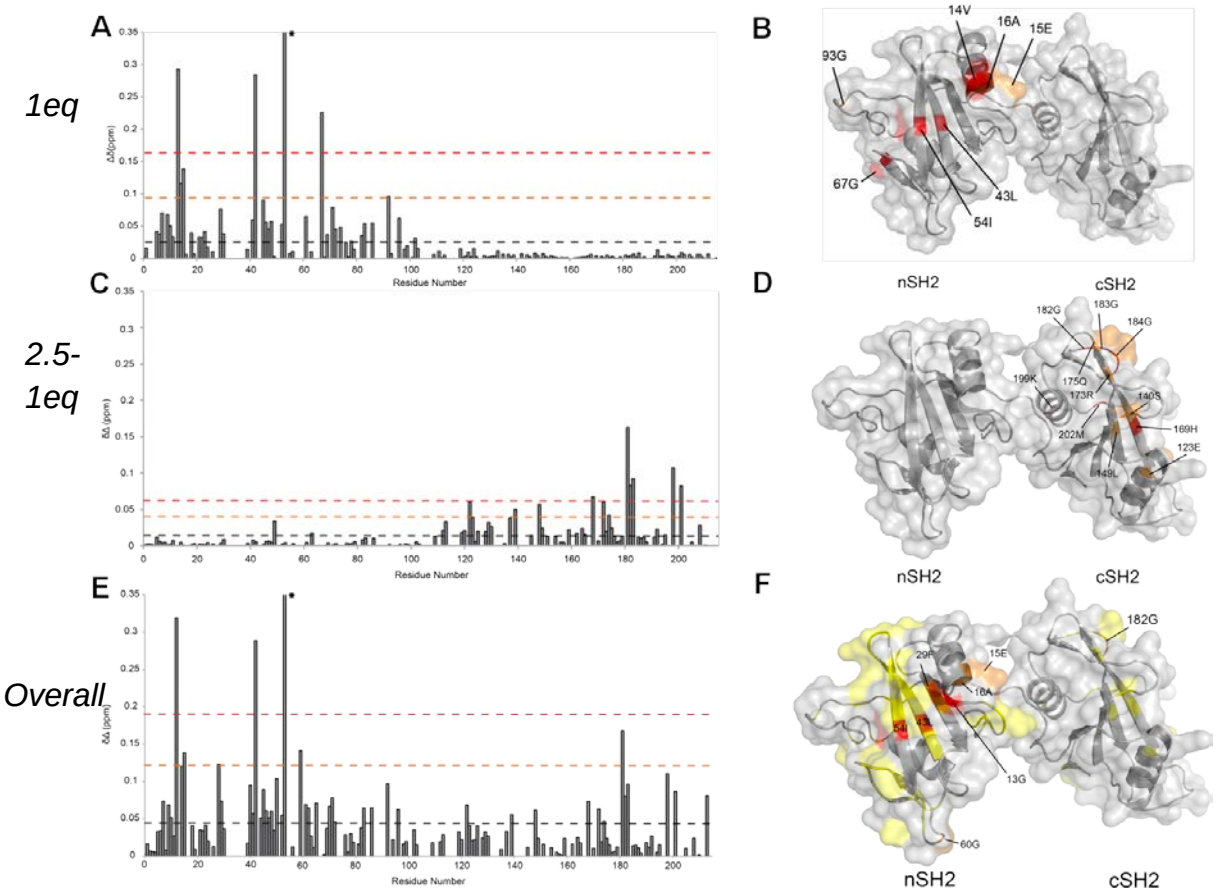


Figure 6.20: TanSH2-Shp2 CSPs Induced by ITIM1 Titration.

(A, C, E) Histograms of CSPs induced by ITIM1 titration as a function of residue number for 400µM ^{15}N-labelled TanSH2-Shp2. Dashed lines indicate CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and mean (grey), highlighting strongly perturbed residues **(B, D, F)** Cartoon and transparent surface representations of TanSH2-Shp2 (PDB ID: 4DGP) with ITIM1-induced CSPs calculated from (A), (C) and (E) mapped on. CSPs ≥ mean + 2 S.D (red) and mean + 1 S.D (orange) are shown, with CSPs ≥ highlighted in yellow for (F). Mean values for (B) (D) and (F) are 0.028± 0.068, 0.017± 0.025 and 0.048± 0.071 respectively.

TanSH2:ITIM2

A similar phenomenon is observed when titrating TanSH2-Shp2 with ITIM2. Around 13 peaks undergo CSPs at 0.25 molar equivalence in a slow exchange regime, including residues 123E, 131K, 152R, 174C and 180D (Figure 6.21 A). A similar number of peaks completely broaden, including 117L, 119G, 143H, 148V, 154G, 171M, 199K, 182G, 203V and 204E (Figure 6.21 B). All the aforementioned residues are located within the cSH2 domain. Slow exchange peaks display no further intensity changes at equimolar concentrations of ITIM2 peptide. At higher concentrations of ITIM2, a number of peaks undergo fast exchange CSPs until the end of the titration course (Figure 6.22). CSP analysis reveals that the cSH2 domain preferentially binds ITIM2 peptide at 1 eq (Figure 6.23 A), though significant CSPs seem to be located throughout the SH2 domain rather than on one face as seen for the ITIM titrations with individual cSH2-Shp2 (Figure 6.14). At 2.5 eq, ITIM2 has caused CSPs in both SH2 domains (Figure 6.23 E and F). Observation of CSPs that occur between 1 and 2.5 eq demonstrate that the nSH2 domain undergoes the most shifts at these concentrations (Figure 6.23 C and D). It is worth noting that CSPs caused by ITIM2 are not as pronounced as those caused by ITIM1 toward TanSH2 (Figure 6.23). However, significant CSPs in the cSH2 domain are located in the EF loop, known to be involved in pY-peptide binding, as well as residues 154G and 199K, located in the CD loop and αB helix. These residues are in regions not normally involved in direct pY-peptide interaction. Binding of ITIM2 likely induces conformational changes throughout cSH2 which could alter relative structural element orientation and dynamics, which may explain the large CSPs observed for these residues.

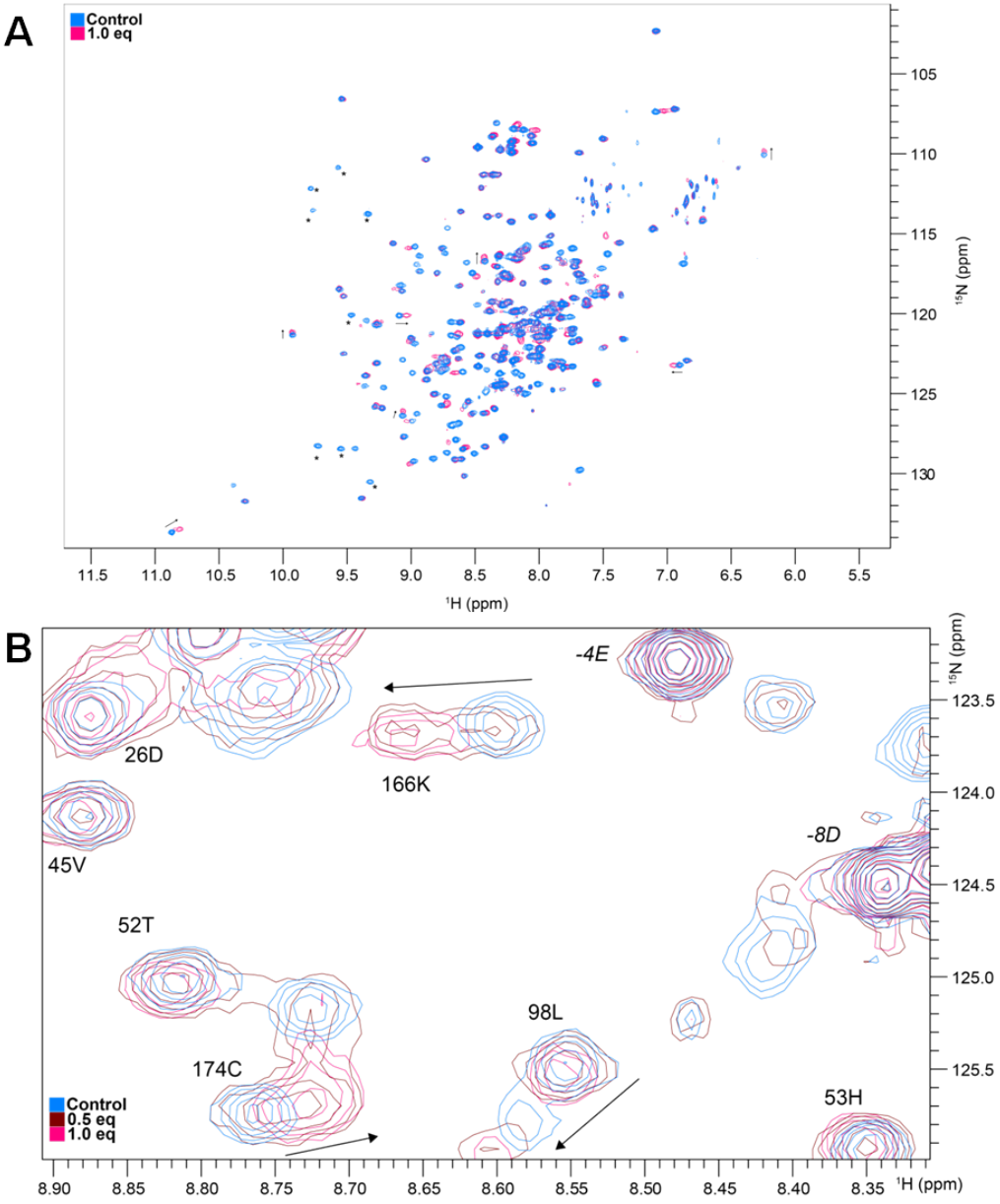


Figure 6.21: TanSH2-Shp2 Titrations with ITIM2 Peptide.

(A) Superimposition of unbound (light blue) and ITIM2-bound (1.0 eq, pink) and **(B)** unbound (light blue), 0.5 eq (brown) and 1.0 eq (pink) TanSH2-Shp2 ^{15}N-HSQC spectra. Arrows highlight CSPs and direction whilst asterisks highlight broadened peaks. For (B), note cSH2 peaks 166K and 174C perturbing in slow exchange, whilst nSH2 residues such as 26D and 45V undergoing no CSPs. 98L is a known nSH2 high affinity site for ITIM peptide interaction and shifts in fast exchange. Residues in italic are part of non-WT His-tag sequence and undergo no CSPs. Data were collected on a Varian 800 NMR Spectrometer at 25 °C in NMR buffer.

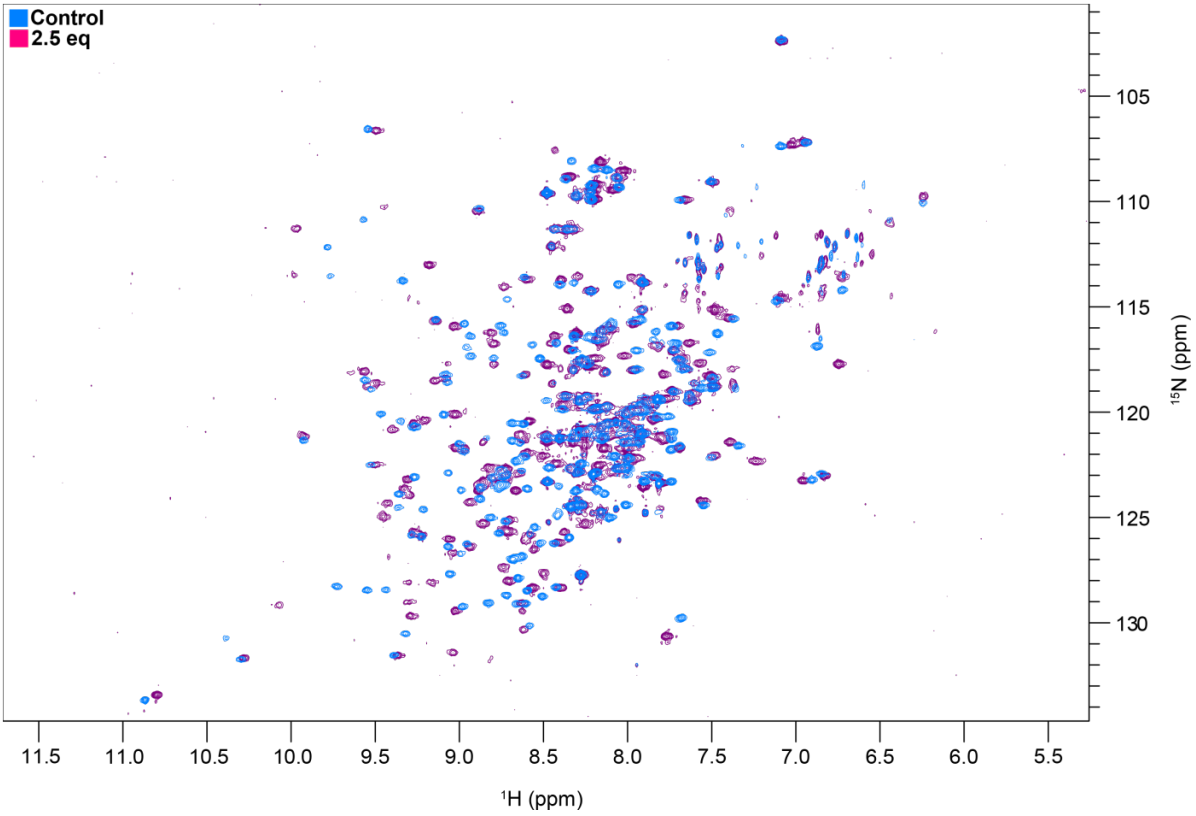


Figure 6.22: TanSH2-Shp2 Interacts with ITIM2 Peptide.

Superimposition of unbound (light blue) and ITIM2-bound (2.5 eq, pink) TanSH2-Shp2 ^{15}N-HSQC spectra, highlighting CSPS throughout the spectrum. Data were collected on a Varian 800 NMR Spectrometer at 25 °C in NMR buffer.

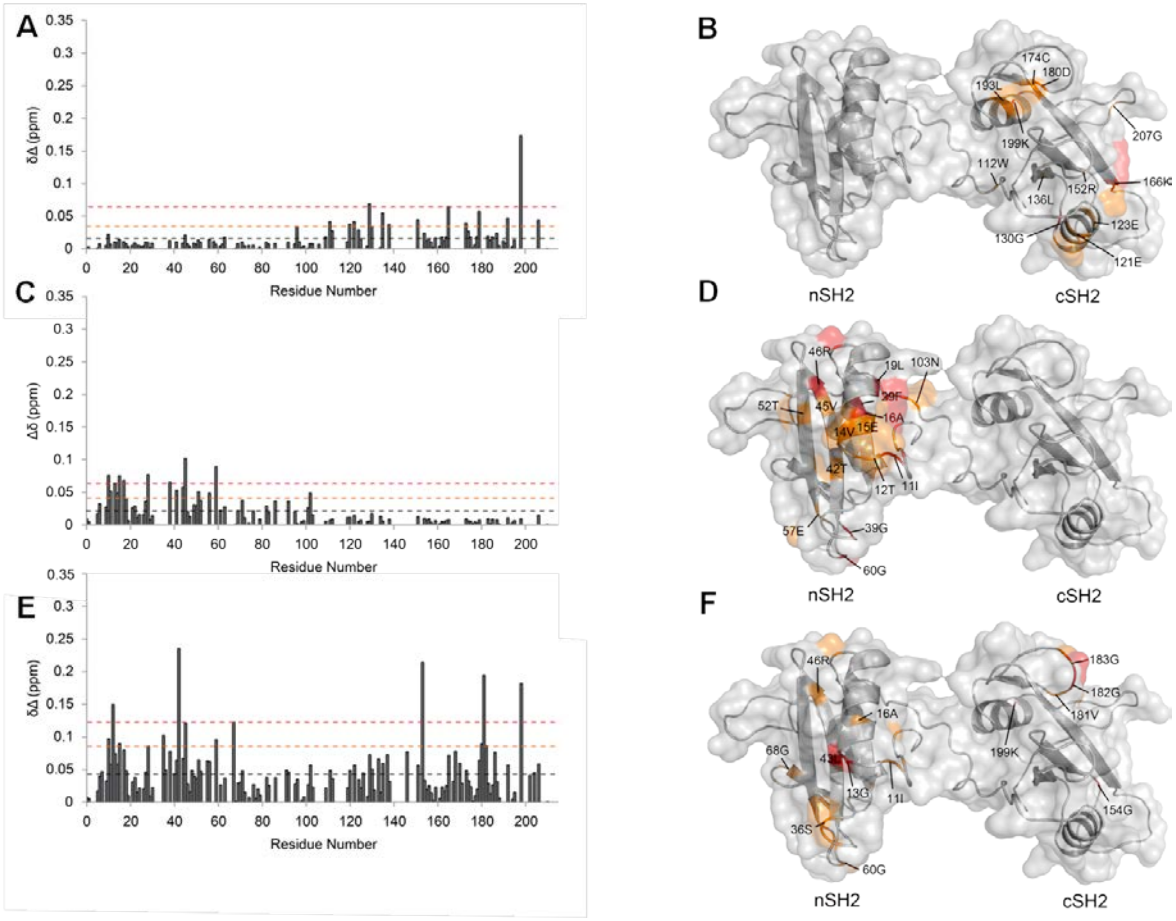


Figure 6.23: TanSH2-Shp2 CSPs Induced by ITIM2 Titration.

(A, C, E) Histograms of CSPs induced by ITIM2 titration as a function of residue number for 400µM ^{15}N-labelled TanSH2-Shp2 at 1 eq, 2.5 eq minus 1 eq and 2.5 eq respectively. Dashed lines indicate CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and mean (grey), highlighting strongly perturbed residues. **(B, D, F)** Cartoon and transparent surface representations of TanSH2-Shp2 (PDB ID: 4DGP) with ITIM2-induced CSPs calculated from (A), (C) and (E) mapped on. CSPs ≥ mean + 2 S.D (red) and mean + 1 S.D (orange) are shown, with CSPs ≥ highlighted in yellow for (F). Mean values for (B) (D) and (F) are 0.016± 0.021, 0.021± 0.022 and 0.046± 0.040 respectively.

nSH2 residues significantly perturbed by saturating concentrations of ITIM2 peptide are located in canonical pY-peptide binding regions: Both BC and EF loop residues and 43L within the βC sheet. However, additional residues outside of these regions, such as 16A and 46R within the αA helix and middle of the CD loop are significantly perturbed too. Minor helical rearrangements in order to accommodate ITIM2 binding could be the cause of this. For 46R, although it is located far from the canonical pY-peptide pockets, the side-chain forms multiple H-bonds with residues nearby: 20L, 21L, 24G, 26D and 28S. It is therefore likely that 46R plays a structure-stabilising role in the nSH2 domain. It was also noted earlier in this chapter that a portion of the aforementioned residues were found to have secondary peaks in both 2D and 3D NMR data, suggesting they were sampling alternative chemical environments. Taken together, it is likely that ITIM2 peptide binding causes reorganization of residues in and around the CD loop, in turn causing 46R to adopt a conformation in a different chemical environment to the unbound state.

TanSH2:dITIM

At 0.25 molar equivalence CSPs are observed in the slow exchange regime (Figure 6.24 B). These correspond to residues 10N, 22T, 49G, 73T, 93G, 152R, 171M, 180D, 181V and 183G. Those CSPs in the nSH2 domain do not localise to a distinctive patch but to loop regions throughout the chain However, early CSPs in the cSH2 domain seem to localise at the +3 pocket.

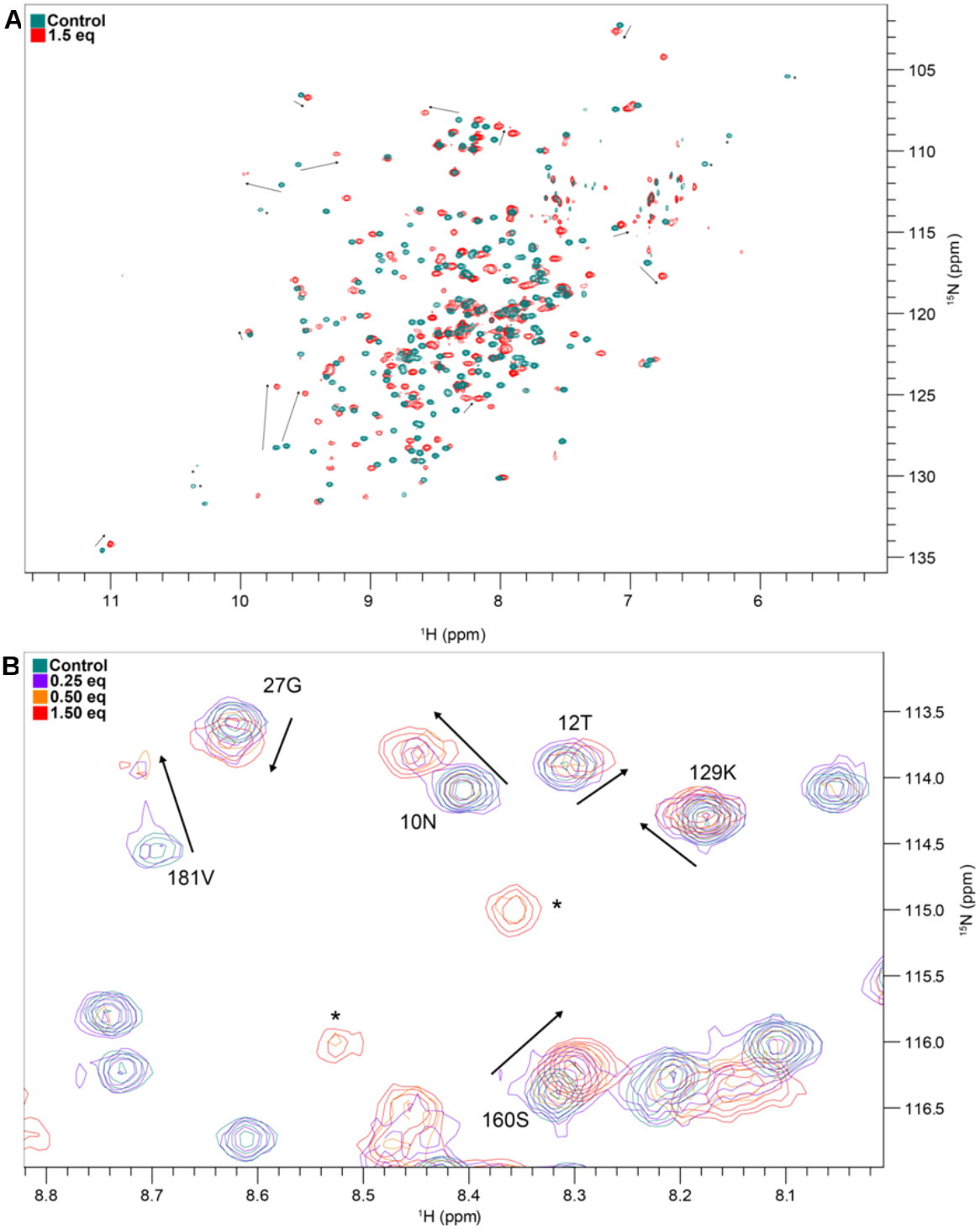


Figure 6.24: TanSH2-Shp2 Titrations with dITIM Peptide.

TanSH2-Shp2 ^{15}N-HSQC spectra of **(A)** Superimposition of unbound (teal) and dITIM-bound (1.5 eq, red) and **(B)** zoomed region of spectrum of unbound (teal), 0.25 eq (purple), 0.5 eq (orange) and 1.5 eq (red). Arrows highlight CSPs and direction whilst asterisks highlight broadened peaks. Data were collected on a Varian 800 NMR Spectrometer at 25 °C in NMR buffer.

At 0.5 molar equivalence peak intensities between the unbound and bound states are roughly equal. Although dITIM peptide was added to 1.5-fold molar excess, no significant shifts were observed after molar equivalence was reached. Multiple peaks underwent CSPs (~60) varying in magnitude. Most CSPs were occurring in slow exchange, making assignment of TanSH2-Shp2 in the bound state difficult. Assignment of the TanSH2:dITIM complex was performed using titration data from previous experiments. A combination of peak overlap and broadening hampered efforts, leading to incomplete assignment of the dITIM-bound state. However, some general features of the interaction can be observed and concluded from the current experiments.

CSPs caused by the dITIM binding TanSH2-Shp2 are larger in comparison to binding single ITIMs (Figure 6.25 A). However, due to dITIM interaction with TanSH2-Shp2 in the slow exchange regime, it was difficult to unambiguously assign peaks in the bound state, even with previously collected NMR titration data. As such, only CSPs for peaks certain to belong to the predicted residue were assigned. In this manner, differentiation between broadened and unassignable peaks was difficult and as a result, peaks in the control spectrum that were no longer present in dITIM-bound spectra were assumed to interact with TanSH2-Shp2. The extent of residue interaction was undeterminable based on current data. The dITIM peptide seems to engage with residues seen in previous titrations (Figure 6.25 B).

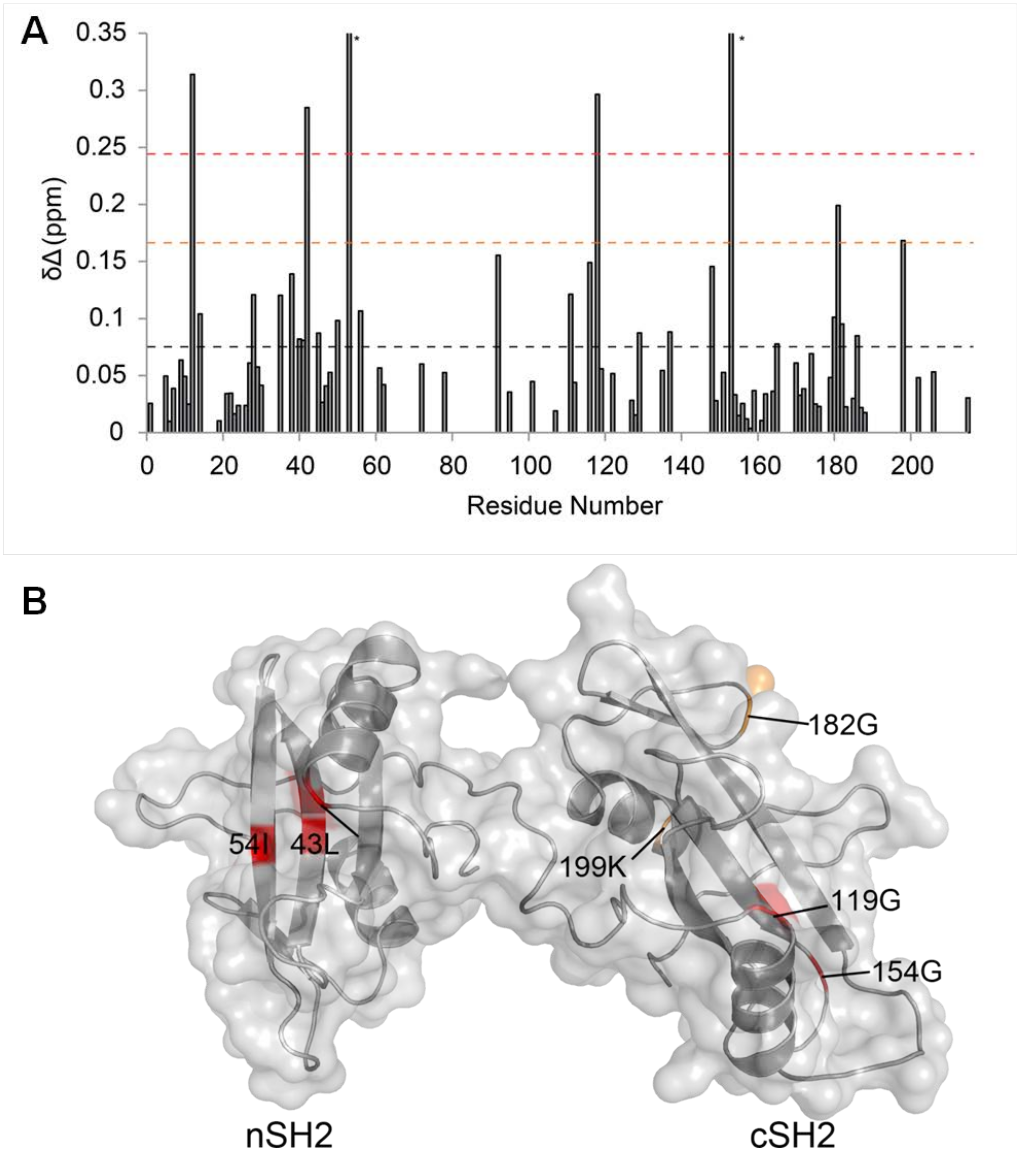


Figure 6.25: TanSH2-Shp2 CSPs Induced by ITIM2 Titration.

(A) Histogram displaying CSPs induced by dITIM titration as a function of residue number for 400µM ^{15}N-labelled TanSH2-Shp2. Peptide was added to 1.5-fold excess. Dashed lines indicate CSPs ≥ mean + 2 S.D (red), mean + 1 S.D (orange) and mean (0.075± 0.087, grey), highlighting strongly perturbed residues. 54I and 154G CSPs (bars next to asterisks) were truncated for clarity, ppm values are 0.56 and 0.36 respectively. **(B)** Surface Representation of TanSH2-Shp2 (PDB ID: 4DGP) with dITIM-induced CSPs mapped on. CSPs ≥ mean + 2 S.D (red) and mean + 1 S.D (orange) are shown here. Mean= 0.08± 0.09.

SAXS Analysis of the Shp2 Tandem SH2:G6b-B dITIM Complex

SAXS data of the tandem SH2 domain of Shp2 (ΔHis-TanSH2 Shp2, hereby referred to as TanSH2) both free and in complex with the dITIM of G6b-B were collected at 2.3, 4.6 and 9.3 mg/ml at the new P12 beamline utilising X-rays produced by the PETRA III storage ring in Hamburg, Germany.

After initial inspection of the data for signs of radiation damage, buffer subtraction and curve plotting as Log[I(q)] vs q (Figure 6.26) were performed. SAXS intensity plots of unbound TanSH2 at all concentrations tested were nearly identical in shape throughout the entire q-range. dITIM-bound TanSH2 SAXS data displayed similar properties. As such the highest concentrations were chosen were further analysis for their high quality at the higher q-angles.

Superimposition of unbound and dITIM-bound SAXS intensity plots reveals differences at the low and high q-range, suggesting a change in particle size upon complex formation. R_g values further indicate this: 2.26 and 2.56 nm for free and dITIM-bound TanSH2, respectively. Invariant parameters are available for bound and unbound TanSH2 (Appendix 9).

Inspection and comparison of the the P(r) curves for unbound and dITIM-bound TanSH2 revealed differences between the two. In particular, TanSH2:dITIM displays a maximum particle dimension of 8.38 nm, 1.15 nm larger than free TanSH2 (7.23 nm). No other significant changes in the P(r) curves are observed. In addition, superimposition of normalised Kratky plots of unbound and dITIM-bound TanSH2 reveal no significant changes in curve shape.

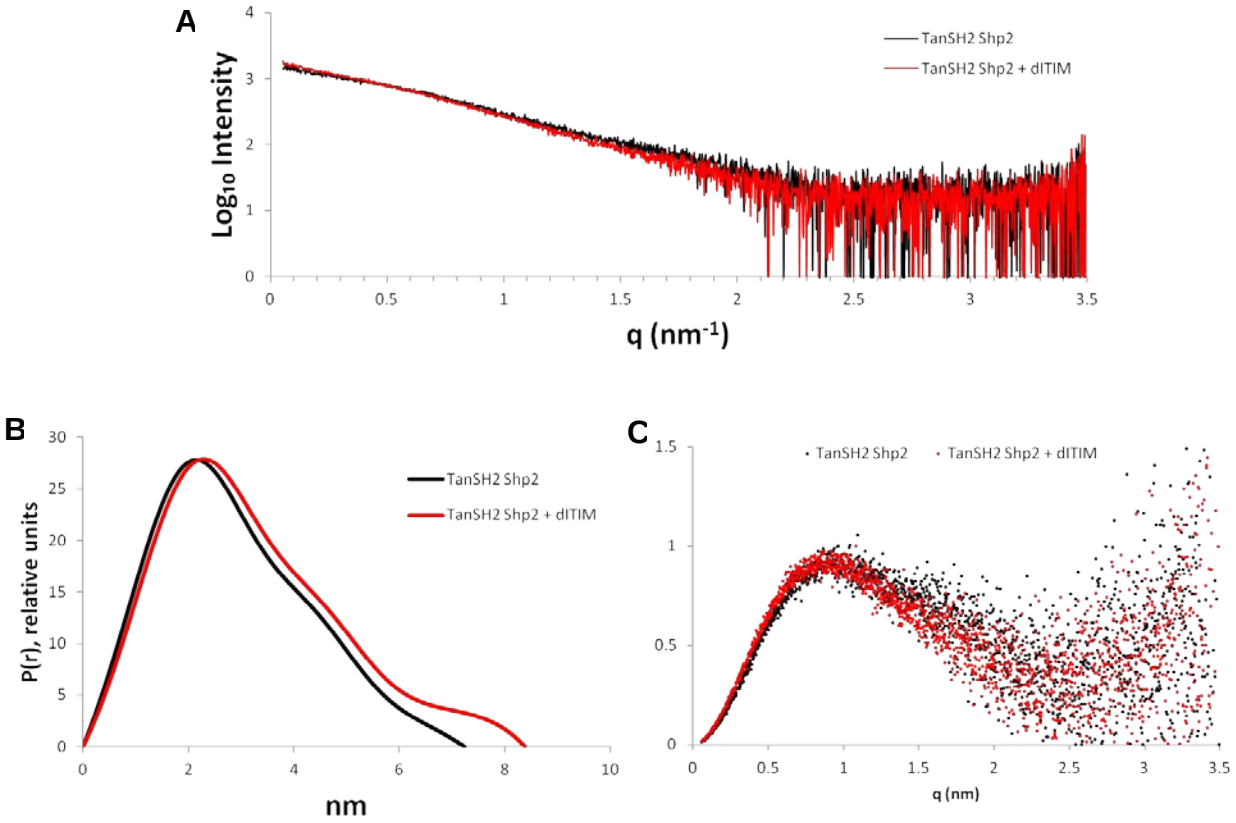


Figure 6.26: SAXS Data of TanSH2-Shp2:dITIM.

Data were collected at beamline P12 within PETRA III storage ring, DESY, Hamburg on a Dectris Pilatus 2M pixel X-ray detector. **(A)** SAXS intensity plot of unbound (black) and dITIM-bound (red) TanSH2-Shp2. Differences between both curves are observed, suggesting a change in global shape and size upon complex formation. **(B)** P(r) curves of unbound and dITIM-bound TanSH2-Shp2, highlighting differences in pair-wise atomic distribution. **(C)** Normalised Kratky plots show no significant changes in unbound and dITIM-bound forms of TanSH2-Shp2.

Taken together, SAXS data of free and dITIM-bound forms of TanSH2 indicate binding of dITIM leads to: 1) An increase in particle size, 2) a change in particle shape, likely due to domain rearrangements and 3) no change in the degree of disorder as qualitatively judged by Kratky plots.

Ab initio Modelling of TanSH2:dITIM Complex

DAMMIF was used to generate 10 *ab initio* models of free and complexed TanSH2 based on SAXS data analysed with GNOM. Averaging and filtering of models with the DAMAVER suite of programs produced rigid-body envelopes for further inspection and analysis.

Unbound TanSH2 adopts a dumbbell-shaped envelope containing two distinct roughly spheroid regions of different dimensions: 34 x 23 x 16 Å and 37 x 33 x 31 Å for the smaller and larger spheroid, respectively (Figure 6.27 A). The larger spheroid has an 18 Å protrusion on the face opposite from its junction with the smaller spheroid. SUPCOMB was utilised to superimpose the envelope with a homology model based on the crystal structure of TanSH2-Shp2 (PDB: 4DGP). The resultant fit has a χ^2 of 1.46, indicating the SAXS-derived envelope and homology model adopt similar global shapes. The larger cSH2 domain fits best in the larger spheroid and *vice versa* for the nSH2 domain. The CD loop within the cSH2 domain fits within the 18 Å protrusion well.

dITIM-bound TanSH2 adopts a similarly shaped dumbbell as in the unbound form, however, both spheroids are of similar dimensions: 33 x 27 x 37 Å and 33 x 26 x 34 Å respectively. In addition, the protrusion has increased in length to 28 Å.

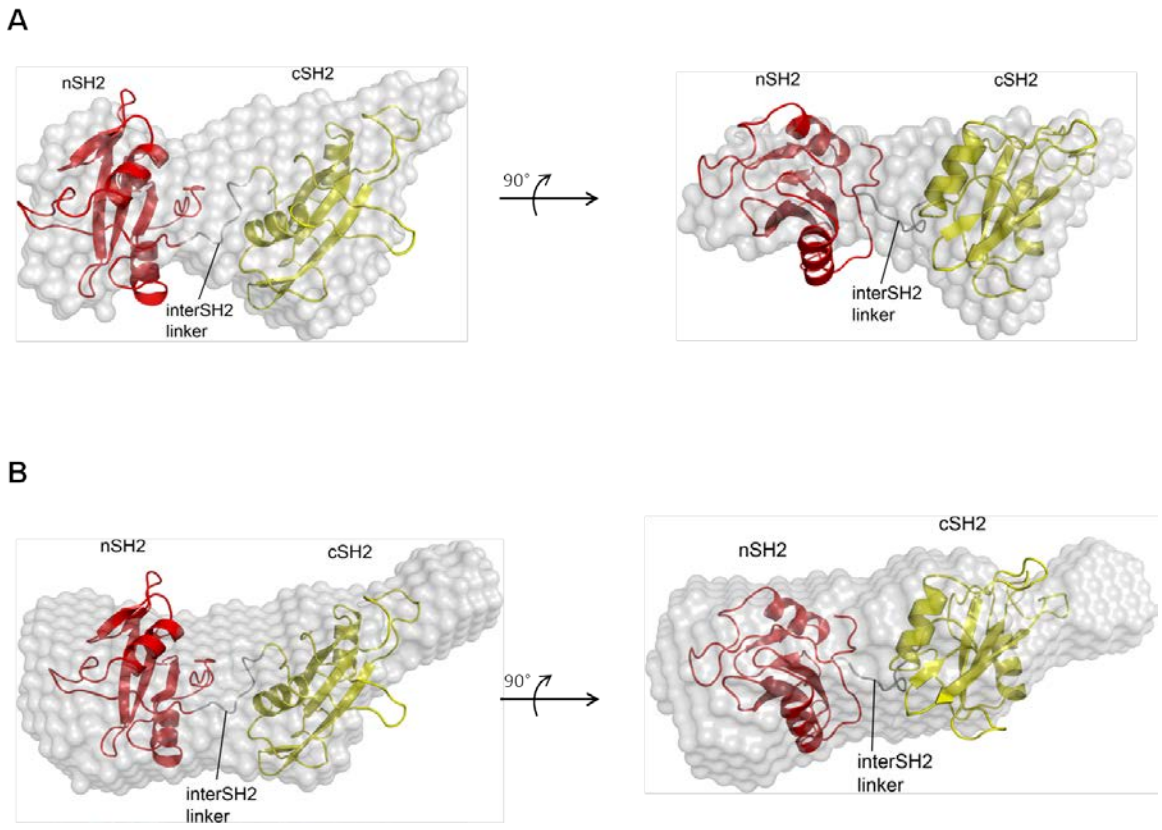


Figure 6.27: *Ab initio* Models of TanSH2-Shp2:dITIM Complex.

DAMMIF-derived envelopes of unbound **(A)** and dITIM-bound **(B)** TanSH2-Shp2 superimposed with homology model of TanSH2 domain (based on PDB ID: 4DGP, cartoon representation) using SUPCOMB (χ^2= 1.46, unbound and 2.14, bound). nSH2, red; interSH2 linker, grey; cSH2, yellow; SAXS envelope, grey transparent surface. Two views are shown.

SUPCOMB was again utilised to fit the TanSH2-Shp2 homology model, yielding a superimposition with χ^2= 2.14 (Figure 6.27 B). This is a worse fit than unbound TanSH2 and indicates that the global shape of the TanSH2:dITIM complex is significantly different from that of the unbound TanSH2 model. The similar dimensions of both spheroids making up the dITIM-bound TanSH2 envelope also make unambiguous assignment of n and cSH2 domains to either one difficult. It is also not known whether the 28 Å protrusion belongs to the CD loop of the cSH2 domain or forms part of the dITIM peptide.

One possible reason for the poor fit between the TanSH2 homology model and the dITIM-bound SAXS envelope is that the former was based on the inhibited state of Shp2 as part of the near-full length protein, which could differ in structure to the ΔPTP domain construct used in this study. To account for this, a second model was generated by reorienting the n and cSH2 domains to resemble the structure published by Eck. *et al*, where the pY binding pockets are on the same face and antiparallel to one another (Eck *et al*., 1996). Automatic fitting yielded a χ^2 of 2.15, slightly worse than the inhibited orientation of TanSH2. Manual fitting of the individual SH2 domains into the dITIM-bound TanSH2 envelope led to a χ^2 of 1.81. While this is better than previous models, the current model does not fully explain the SAXS data. The fact the dITIM peptide, a 36-mer, has not been explicitly accounted for in rigid body modelling could explain the poor fits produced for dITIM-bound TanSH2.

ITIM Peptides Fail to Activate WT FL-Shp2 Catalytic Activity

Previous studies have investigated the effect of both single and dual-pY containing phosphopeptides on the rate of dephosphorylation of various Shp2 constructs. These have shown a high increase in Shp2 catalytic activity with the addition of phosphopeptide (Eck *et al*., 1996; Gavrieli *et al*., 2003). The tight binding of various ITIM peptides to Shp2 SH2 domains observed in previous studies as well as the current investigation suggests G6b-B ITIM peptides should activate Shp2 and thereby increase catalytic activity (Pumphrey *et al*., 1999; Famiglietti *et al*., 1999; Gavrieli *et al*., 2003).

PNPP activity assays were performed to investigate whether the presence of ITIM peptides could increase the catalytic activity of WT FL-Shp2 *in vitro*. Experiments were carried out at RT and as described in Chapter II. Increasing concentrations of dITIM seemed to decrease the extent of PNPP dephosphorylation, contrary to previous research on the effect of pY-peptides on Shp2 activation state (Figure 6.28). However, the high standard deviation for 0.3 and 0.75 µM dITIM concentration obscures effective interpretation of the data. Further activity assays with single ITIM peptides were partially carried out; ITIM1 peptide added to saturating concentrations also failed to increase Shp2 activity as observed by PNPP activity assay (data not shown). The results seem to suggest that dITIM addition leads to minimal increases in activity.

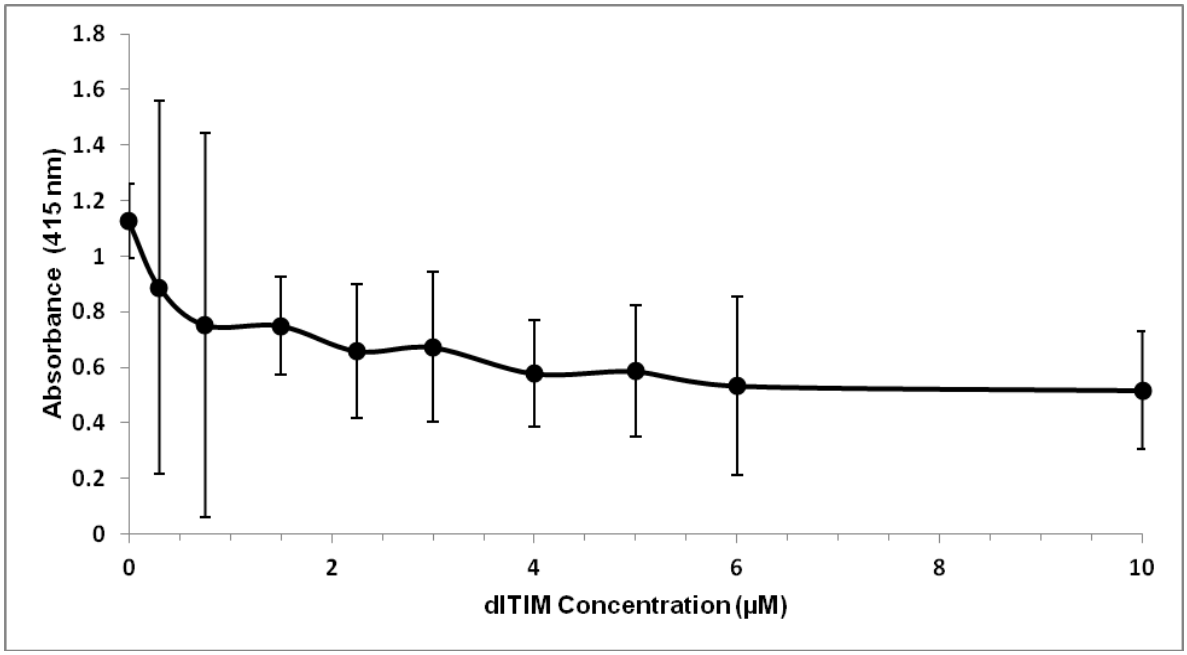


Figure 6.28: PNPP Activity Assays of WT FL-Shp2.

FL-Shp2 (concentration: 3 µM) catalytic activity was assessed with PNPP activity assay. Activity was assessed either in the presence of either increasing amounts of sodium phosphate buffer only or sodium phosphate buffer + dITIM. Each point is an average of 3 repeats within the same experiment, initiated at RT with 10 µl 150 mM PNPP (final concentration of 15 mM), run for 10 minutes and terminated upon addition of 25 µl 5 M NaOH. Data collected using iMark Microplate reader (BioRad) at 415 nm wavelength. Assay buffer: 50 mM HEPES pH 7.5, 2 mM EDTA, 5 mM DTT, 2% glycerol, 0.01% Triton X-100. dITIM stock: 244 µM dITIM in 50mM Na Phosphate pH 7.4.

Chapter Conclusions

Phosphorylated dITIM is Primarily Unfolded

Secondary structure predictions of dITIM suggested a portion of ITIM2 could form a β-strand or extended structure. CD analysis of phosphorylated dITIM indicates a predominantly random coil structure with possible minor secondary structural elements. Since the secondary structure prediction program does not take into account PTMs, prediction of dually-phosphorylated dITIM secondary structure could not be made. This may explain the difference between predicted and experimental data, as phosphorylation has been known to cause changes in secondary structure. However, both ITIMs are constitutively phosphorylated (Mazharian *et al*., 2013) and thus dually-phosphorylated dITIM is more biologically relevant than its unphosphorylated form as the latter is not known to exist at significant levels within cells.

TanSH2-Shp2 Binds dITIM in 1:1 Stoichiometry

As dITIM contains two points of contact for SH2 domain binding, and since TanSH2-Shp2 contains two binding sites for phosphorylated peptides, it was presumed the pair would bind with 1:1 stoichiometry. To validate this hypothesis and to rule out other binding stoichiometries, a solution-based sedimentation technique was employed.

Velocity AUC experiments comparing unbound and dITIM-bound TanSH2-Shp2 reveal an increase in the hydrodynamic radius upon complex formation. The shift in

cS is also consistent with 1:1 binding stoichiometry and such an increase may be indicative of reorientation of SH2 domains to accommodate dITIM.

Shp2 SH2 Domains Display Binding Preferences toward ITIM Peptides

Solution NMR binding experiments showed that the individual SH2 domains of Shp2 are capable of binding ITIM peptides derived from G6b-B. ITIM1 binds the nSH2 domain in slow exchange while binding the cSH2 domain in slow exchange. For ITIM2, the opposite is true: the nSH2 domain binds ITIM2 in fast exchange and binds the cSH2 domain in slow exchange. This indicates that the nSH2 domain has a higher affinity toward ITIM1 than ITIM2 and the cSH2 domain has a higher affinity toward ITIM2 than ITIM1. The ITIM specificity for either SH2 domain thereby suggests that in a cellular environment, there is a preferred domain orientation present in the Shp2:G6b-B interaction. Whether this phenomenon is of functional relevance to Shp2 and indeed the Shp phosphatases in general remains to be elucidated.

ITIM1 Truncation Results in Weaker SH2 Binding

ITIM1s, where ITIM1 was truncated by two C-terminal residues (+5 and +6 positions of phosphopeptide) showed diminished (that is, reduced CSP magnitude) binding to both nSH2 and cSH2 domains, although differential exchange phenomena were still present between the two domains. Peaks of both SH2 domains also underwent smaller CSPs with ITIM1s than with ITIM1. This suggests ITIM1 residues at the +5

and +6 positions i.e. 216H and 217L of WT G6b-B respectively, are important in contributing binding affinity to the Shp2 SH2 domain interaction. A combinatorial screening approach used to study the SH2 domains of Shp1 and Shp2 found they were able to bind phosphopeptide residues beyond the +3 position, up to +6, and each SH2 domain showed preferences for certain amino acids at these positions. For example, nSH2-Shp2 was observed to prefer bulky residues such as F, H and W (in order of highest to lowest preference) at the +5 position (Imhof *et al*., 2006). Interestingly, the cSH2 domains of Shp1 show narrower amino acid binding preferences at the +4, +5 and +6 positions which could confer additional specificity for the tandem SH2 domains when binding peptides.

SH2:ITIM Specificity is Maintained in TanSH2-Shp2

NMR titration experiments of His-TanSH2-Shp2 with single ITIM1 and ITIM2 peptides as well as dITIM demonstrated that TanSH2-Shp2 was capable of binding ligands and was thus functionally active. Closer inspection of CSPs revealed that TanSH2-Shp2 was able to bind ITIM peptides with the same differential exchange rates observed for the individual SH2 domains binding single ITIM peptides. ITIM1 preferentially bound the nSH2 in a slow exchange regime and after 1:1 molar equivalence, presumably when the nSH2 binding site was completely bound, were CSPs observed for the cSH2 domain in a fast exchange regime. In a similar manner, TanSH2-Shp2 titration with ITIM2 causes CSPs in residues within cSH2-Shp2 in a slow exchange regime, albeit these CSPs were markedly reduced compared to ITIM2 binding cSH2-Shp2 alone. In the current experiment, ITIM2-induced CSPs in the

cSH2 domain of TanSH2-Shp2 were weaker in general than for ITIM1 binding the nSH2 domain in TanSH2-Shp2. However, after 1:1 molar equivalence of TanSH2-Shp2 and ITIM2 is reached, fast exchange CSPs are observed for residues within nSH2-Shp2. This clearly demonstrates that peptide specificity is maintained between either SH2 domain within the TanSH2-Shp2 construct.

Differences in SH2:ITIM Peptide Interactions

Whilst both ITIM1 and ITIM2 peptide bind the SH2 domains of Shp2, different subsets of residues underwent CSPs in NMR titration experiments. In particular, the interaction of either ITIM with nSH2-Shp2 caused CSPs in residues within or at the beginning of the DE loop, specifically 59T, 60G and 61D. In WT FL-Shp2, these residues make numerous contacts with the PTP catalytic cleft, particularly 61D which acts as a phosphate group mimic (Hof *et al*., 1998). ITIM1 binding causes significant CSPs or broadening here, although these residues are distinct from the pY and +3 pockets in nSH2-Shp2. It cannot be proven without uncertainty that ITIM1 residues do not make direct contacts with DE loop residues, based on current NMR data. However, taken together with previous phosphopeptide bound nSH2-Shp2 structures, ITIM1-binding likely causes conformational changes in the DE loop. Like ITIM1, ITIM2 binding causes 60G to undergo significant CSPs, though to a lesser extent. Surrounding residues also undergo CSPs above the mean, again to a lesser extent than those seen when ITIM1 peptide had been added. Therefore, ITIM2-induced perturbations is DE loop residue signals could be affinity dependent, only changing conformation once bound to its cognate ligand. Interestingly, cSH2-Shp2

DE loop residue undergo minimal CSPs upon ITIM peptide binding, suggesting DE loop conformational changes are restricted to nSH2-Shp2. nSH2-Shp1 contains a homologous DE loop sequence, suggesting the mechanisms of peptide binding causing DE loop conformational changes is conserved. This has also been proposed in previous studies (Hof *et al*., 1998)

Crystal structures of phosphopeptide-bound nSH2 in isolation (Lee *et al*., 1994) or as part of the tandem SH2 domain (Eck *et al*., 1996) showed differences in DE and EF loop conformation when compared with unbound nSH2 in isolation and as part of WT FL-Shp2. These conformational changes in the bound state would cause steric clash with PTP domain residues and could disrupt the H-bonding network enough to cause nSH2 domain removal from the active site. This mechanism has been suggested and based on crystallographic data and modelling (Hof *et al*., 1998) and is corroborated by the current study of phosphopeptide:Shp2 interactions by solution NMR. Also, molecular modelling suggests Y66 within the EF loop acts as a conformational switch for nSH2-Shp2 (Guvench *et al*., 2007), influencing DE loop residues to undergo local changes and thus could form part of the overall Shp2 activation mechanism. Y66 was consistently found to undergo either significant CSPs or broaden altogether in ITIM titration experiments. While this is not definitive in confirming Y66 acts as a conformational switch in Shp2 activation, these results are promising and warrant further investigation.

dITIM-Binding Causes Domain Rearrangements in TanSH2-Shp2

NMR titration data of the TanSH2:dITIM interaction provided some insights into the binding mode in that the affinity is likely below the micromolar range based on the presence of slow exchange phenomena present during the course of the experiment, implying a slow off rate of dITIM peptide from the TanSH2 domain. Further information obtained from SAXS characterisation of the complex revealed a larger R_g value than for the TanSH2 domain alone, meaning the relative size of the protein has increased and this is likely attributed to complex formation with dITIM. Additionally, the SAXS envelope of dITIM-bound TanSH2 has become more extended, suggesting dITIM terminal residues extending parallel and away from TanSH2 or, more likely, extension of TanSH2 on its major axis caused by large scale domain reorientation of one or both SH2 domains.

The orientation of the Shp2 SH2 domains as part of the full length protein does not appear optimal for binding dually phosphorylated peptide motifs due to both phosphopeptide pockets facing different directions. Initial engagement of one SH2 domain with its cognate ITIM ligand increases the effective local concentration of the second ITIM ligand, increasing the likelihood of an interaction with the remaining unbound SH2 domain. Furthermore, the initial ITIM interaction with its SH2 domain could cause a conformational change in the latter, allowing it to sample more areas of space. An increase in the relative mobility of the interSH2-linker might play a role in facilitating this domain reorientation. An optimal tandem SH2 orientation could form, allowing the remaining SH2 domain to bind the remaining ITIM.

The current understanding of Shp2 structural biology posits that PTP domain-bound nSH2 adopts a distorted conformation unable to effectively bind phosphopeptides. Residues within the DE and EF loops of nSH2 as part of FL-Shp2 show marked differences in Cα positions compared to nSH2 in isolation (Hof *et al,*. 1998). cSH2 however, adopts a conformation more readily able to bind phosphopeptide, and has been hypothesised to be the first point of interaction for peptide ligands (Pao *et al*., 2007). Aside from circumstantial data from crystal coordinates however, no experimental data is available to lend weight to this hypothesis. The AUC and SAXS experiments from the current study demonstrate that TanSH2 in isolation increases in hydrodynamic radius and radius of gyration upon binding dITIM peptide. Qualitative analysis of the intrinsic flexibility of TanSH2:dITIM using Kratky plots reveal that differential mobility may exist between it and unbound TanSH2. It is possible that stable binding of dITIM causes certain mobile elements to rigidify (BC, CD and EF loops of both SH2 domains) whilst causing a large scale domain reorientation to produce an optimal conformation for favourable dITIM binding, the bound state being more extended as a result of interSH2-linker relocation.

nSH2:ITIM1s Complex Formation

The interaction of nSH2-Shp2 with the ITIM1 sequence was judged by SPR (Coxon *et al*., 2012) to be stronger than ITIM2. Slow exchange peaks were difficult to unambiguously assign, leading to use of 3D NMR spectra to assign the backbone to over 90% for nSH2:ITIM1s. Residues involved in ITIM1s high affinity binding were unassignable, such as the EF loop, as well as 35K within the BC loop and 85H and

86G within the BG loop, which were also unassigned in uncomplexed nSH2-Shp2. It is interesting to note that these residues were unassignable in both peptide-bound and unbound forms. For 85H and 86G, neither of which form H-bonds with residues within the nSH2 domain or contacts with phosphopeptides, intermediate exchange processes independent of ligand binding could explain their complete lack of assignments. However, 35K, situated in the highly mobile BC loop, does not appear during ITIM1s interaction. This is consistent with other studies of peptide-bound SH2 domains investigated by NMR which show the BC loop as being highly flexible. While it is thought the BC loop may stabilise in a narrower set of conformations during peptide interaction, the failure to obtain backbone information may suggest no significant change in the dynamic state of 35K upon ligand binding. From NMR titration analysis, both flanking serine residues (34S and 36S) either undergo significant CSPs or broaden completely. This suggests at least part of the BC loop is capable of directly binding and/or changing conformation during phosphopeptide docking, consistent with previous reports of SH2:phosphopeptide interactions (Bradshaw & Waksman, 2002; Frese *et al*., 2006).

Backbone assignment and secondary structure prediction of nSH2:ITIM1s using chemical shift data yielded a typical SH2 domain topology that does not deviate from the unbound nSH2 domain secondary structure predictions made in Chapter IV as well as available crystallographic data (Lee *et al*., 1994).

The EF Loop of nSH2:ITIM1s complex is in intermediate exchange as observed by peak broadening of residues 65L, 66Y and 67G. The EF loop forms portions of the SH2 specificity pocket and is critical in binding hydrophobic residues at the +3 position of phosphopeptides (Songyang *et al*., 1993). ITIM1s residues C-terminal to

pY are Ala-Asp-Leu-Asp. The conserved Leu or other hydrophobic residue of binding phosphopeptides has been shown in previous SH2 domain ligand-bound structures to bind deep into the hydrophobic specificity pocket, contributing entropy-driven binding energy to the interaction (De Fabritiis *et al*., 2008; Ladbury *et al*., 1995). A canonical SH2:peptide interaction is most likely occurring. Interestingly, both ITIM1 and ITIM1s peptide titration series to 2.5-fold molar excess do not cause EF loop residue peaks to return once nSH2-Shp2 is fully saturated. As the ITIM1s peptide has a slightly lower affinity for nSH2-Shp2, a 1:1 ratio of peptide:SH2 may not have been sufficient for saturating nSH2-Shp2 peptide binding sites, which may explain why EF loop residues were in intermediate exchange in nSH2:ITIM1s backbone assignment spectra.

dITIM and Role of Shp2 Interaction

PNPP activity assays were used to test the hypothesis that dITIM peptide interaction with WT FL-Shp2 should increase the enzymatic rate of dephosphorylation caused by release of the autoinhibitory contacts from the nSH2 DE loop with the catalytic cleft. However, little to no activation was noted and instead, a slight inhibitory dITIM concentration-dependent effect was observed. There could be a number of reasons for this result. Firstly, the SH2 domains could be tightly bound to the catalytic domain, exhibiting a very slow off-rate that effectively prevents dITIM binding to the overlapping site. Secondly, the presence of the His-tag could disrupt efficient dITIM binding and PTP activation. Although NMR titration experiments clearly demonstrated His-TanSH2 could bind dITIM, which is corroborated by sedimentation

velocity AUC experiments, in the full-length protein there could be additional points of contact and interaction with the PTP domain and/or C-terminal tail that could prevent domain reorientation for efficient activation.

Thirdly, dITIM could activate WT FL-Shp2 regardless of the presence of an N-terminal His-tag, only to bind and dephosphorylate dITIM peptide not bound to the TanSH2 domain. This is a possibility as saturating concentrations of dITIM peptide were reached after 3 μM, though an inhibitory effect was already observed by this point. Experiments have not been performed to investigate whether the Shp2 PTP domain can indeed bind and dephosphorylate dITIM. Finally, independent repetitions were not performed for this experiment using different batches of WT FL-Shp2, buffers and new stocks of dITIM, limiting the reliability of these initial results. While no positive control was used to assess Shp2 activity during the course of the assay, qualitative PNPP assays (i.e. to observe any colour change after 1h when compared to controls) are performed for each FL-Shp2 batch produced to ensure active protein is present, though a reliable positive control would strengthen the experiment. It is also unclear whether phosphate buffer may inhibit PTP activity and this warrants further investigation.

Additionally, the PNPP assay may not be appropriate to investigate the role of dITIM on FL-Shp2 and instead, the malachite green phosphate detection assay or DiFMUP assay could be utilised (Montalibet *et al*., 2005).

For the reasons outlined above, the activity assay performed is deemed inconclusive and warrants further investigation by way of different assay methods with WT and mutant forms, technical repetitions and His-tag removal.

Chapter VII
Discussion and Conclusions

This project has investigated various questions put forth at the beginning of this work regarding aspects of Shp2 structural biology that were yet to be answered. Through experimentation and modelling, a number of novel discoveries have been made which can now be utilised in building a better understanding of this fascinating molecule in the solution state. In addition, like all research projects, many new questions have arisen and will need to be investigated in the future. The following chapter will attempt to summarise what has been learnt about Shp2 and contemplate the wider implications of the present study.

Shp2 in the Solution State

All solved structures of Shp2 SH2 domains were done so by X-ray crystallography. Such a "rigid snapshot" may not be enough to describe the native state of proteins in solution, especially those with numerous mobile elements, some of which are important for protein function. For SH2 domains in general, mobile loops govern stabilisation and specificity of phosphopeptide interactions. For such a highly conserved protein of potential therapeutic value, increased knowledge of Shp2 in the solution state may be important to consider when designing experiments and inhibitors.

NMR studies of nSH2, cSH2 and tandem SH2 domains of Shp2 reveal the secondary structure of these proteins is consistent with previously published structures of Shp2 as well as other SH2 domain structures solved by both X-ray crystallography and solution NMR. In addition, SAXS analysis of the unbound tandem SH2 domain show that while there is some resemblance to a rigid structure, a flexible model best represents the available data. Dynamic regions of the tandem SH2 domain lie at the termini, interSH2-linker and CD loop of the cSH2 domain. Interestingly, the interSH2-linker has been shown to form numerous contacts between itself and both adjacent SH2 domains and thus was assumed to be fair rigid, also being under 10 residues in length it was thought unlikely to be mobile.

Shp2 exhibits flexibility in solution that would not be detected well by X-ray crystallography. Domains in isolation fit crystallographic data well but with discrepancies that suggest conformational flexibility and loop mobility contribute to the SAXS data.

The C-terminal tail, which has not been structurally characterised up until the present study, is likely to be highly mobile in solution as judged by the SAXS data. Sharp peaks from FL-Shp2 observed in NMR experiments could correspond to the C-terminal tail, though this is somewhat speculative. Given that none of the assigned peaks in either individual SH2 domain and tandem SH2 domain do not overlay with the peaks in the FL-Shp2 spectrum, the peaks in the latter must originate from either the catalytic domain or C-terminal tail, and, more unlikely, the His-tag.

The SH2 Domains of Shp2 exhibit conformational changes in discrete patches, particular for the nSH2 domain where such states were observed outside of the

canonical pockets. This could be due to minor sequence-specific structural rearrangements, explaining why little secondary conformational changes are present in the cSH2 domain.

The interSH2 linker was found to be mobile as part of a TanSH2-only construct. This mobility likely arises due to free movement of the nSH2 in the absence of PTP domain to form an interface with, and thus could correspond to the TanSH2 in the active FL-Shp2 structure.

Conformational flexibility observed in the CD loop by NMR of cSH2 might be of functional relevance, but as of yet no function has been assigned to them. Using the Crk SH2 domain as an example, it was observed that the wild-type structure contains a long polyproline motif–containing insertion within the DE loop that was found to bind the c-Abl SH3 domain (Anafi *et al*., 1996; Donaldson *et al*., 2002). Such a phenomenon in the context of Shp2 is certainly possible due to the solvent accessibility of the CD loop, though it is not rich is proline residues, instead harbouring a number of lysine residues which could form electrostatic interactions with other biological molecules.

Structural Differences between WT and Mutant Shp2 are Present

The first experimental evidence of FL-Shp2 E76K mutation causing structural perturbation has been presented in this study. The likely scenario is the equilibrium between closed and open form could be disrupted allowing FL-Shp2 to take on an extended structure for longer residence times.

A crystal structure does not yet exist for E76K Shp2, though studies have looked at other disease relevant mutants (Yu *et al*., 2013; Qiu *et al*., 2014) but given the importance of understanding highly-activating Shp2 mutants, it was surprising E76K was not structurally characterized and this could be due to the increased mobility caused by E76K mutation.

The Case of the Two-Faced Tryptophans

Current evidence taken from NMR experiments indicates the presence of at least two equally populated conformations of the tryptophan side-chain indole system. It is possible other conformations exist but are not visible at the time-scales mentioned.

Various crystal structures of the SH2 domains of Shp2 show that both 6W and 112W reside in the first N-terminal loop of each individual SH2 domain. Trp residues at this position are highly conserved, suggesting they have a role.

The flanking residues of each Trp are identical, suggesting that this N-terminal portion may be evolutionarily conserved and thus likely serve a function. Since this region is located on the opposite face to the phosphopeptide binding cleft, it is improbable that the role of these Trp residues and indeed the N-terminal loop as a whole are functional in some form. It is therefore likely the loop could play a structural role. The indole ring could engage in multiple hydrophobic interactions and stabilise the SH2 domain core. Both 6W and 112W engage in H-bonding with the main-chain amide groups of 31A and 137V respectively. There could be potential for the indole side chain amide to form H-bonds half of the time in solution. This may be caused by

the rest of the N-terminal loop peeling on and off the back side of the SH2 domains. However, while this could hold true in the individual SH2 domains, the presence of double the expected Trp resonances in the tandem SH2 suggests this phenomenon could still occur as part of a WT protein. The Trp residues within the SH2 domains are both in the same relative position in both SH2 domains (βB4) and are located one residue N-terminal to the highly conserved Arg residues involved in pY engagement. As such, conformational changes in ArgβB5 and any other residue in the conserved FLVR motif could induce CSPs in surrounding residues not directly involved in phosphopeptide binding. This may explain why CSPs are observed in both 6W and 112W main-chain and side-chain amide groups upon ITIM peptide engagement.

Functional Significance of dual ITIM Specificity and Phosphopeptide Engagement

The preferential binding of SH2 domains to specific peptide sequences is well studied and multiple studies have also observed peptide specificity across SH2 domains within the same polypeptide chain (See introduction). The individual Shp2 SH2 domains have been observed to bind ITIM1 and ITIM2 peptides with varying affinities, as well as showing both ITIMs in tandem (dITIM) can bind the tandem SH2 domain with even higher affinity (Coxon *et al*., 2012).

In the present study, the molecular interactions of G6b-B ITIM peptides have been assessed by a variety of biophysical methods. Differential exchange phenomena were observed for both nSH2 and cSH2 domains of Shp2 as well as the cSH2 domain of Shp1 and this was ITIM peptide dependent. This partly corroborates data

from Coxon *et al*. in that differences in relative affinity for n and cSH2-Shp2 existed when studied by SPR and direct-binding assays. However, their study also showed cSH2-Shp2 displayed negligible binding to ITIM1, ITIM2 and dITIM peptides. cSH2-Shp1 was also shown to weakly bind ITIM peptides. By solution NMR experiments, cSH2-Shp2 was found to bind both ITIM1 and ITIM2, the latter with higher affinity. In addition, cSH2 residues from TanSH2-Shp2 bound ITIM2 in slow exchange and ITIM1 in fast exchange, as well as binding dITIM in slow exchange. Taken together, cSH2-Shp2 does in fact bind both ITIM peptides, with a clear preference for ITIM2. cSH2-Shp1 was also found to bind both ITIM1 and ITIM2, also displaying a preference for ITIM2. Coxon *et al*. also observed that nSH2-Shp2 binds dually phosphorylated dITIM with a K_D value of 130 nM. However, TanSH2-Shp2 bound dually-phosphorylated dITIM with a K_D value of 0.48± 0.09 nM i.e almost 271-fold higher in the presence of the cSH2 domain. Given the NMR titration data showing binding and preference for cSH2-Shp2 and ITIM2, the additional binding affinity seen by SPR likely arises from cSH2 binding ITIM2.

The TanSH2 in complex with dually phosphorylated peptide takes on and extended “open” structure and this is the first instance of such an interaction being structurally characterised for Shp family PTPs. Single phosphorylated peptides binding into each SH2 pocket may not induce the same rearrangements caused by dually phosphorylated peptides and thus make the solution characterisation of the TanSH2:dITIM novel. A model for immunoreceptors activation is summarised in Figure 7.0.

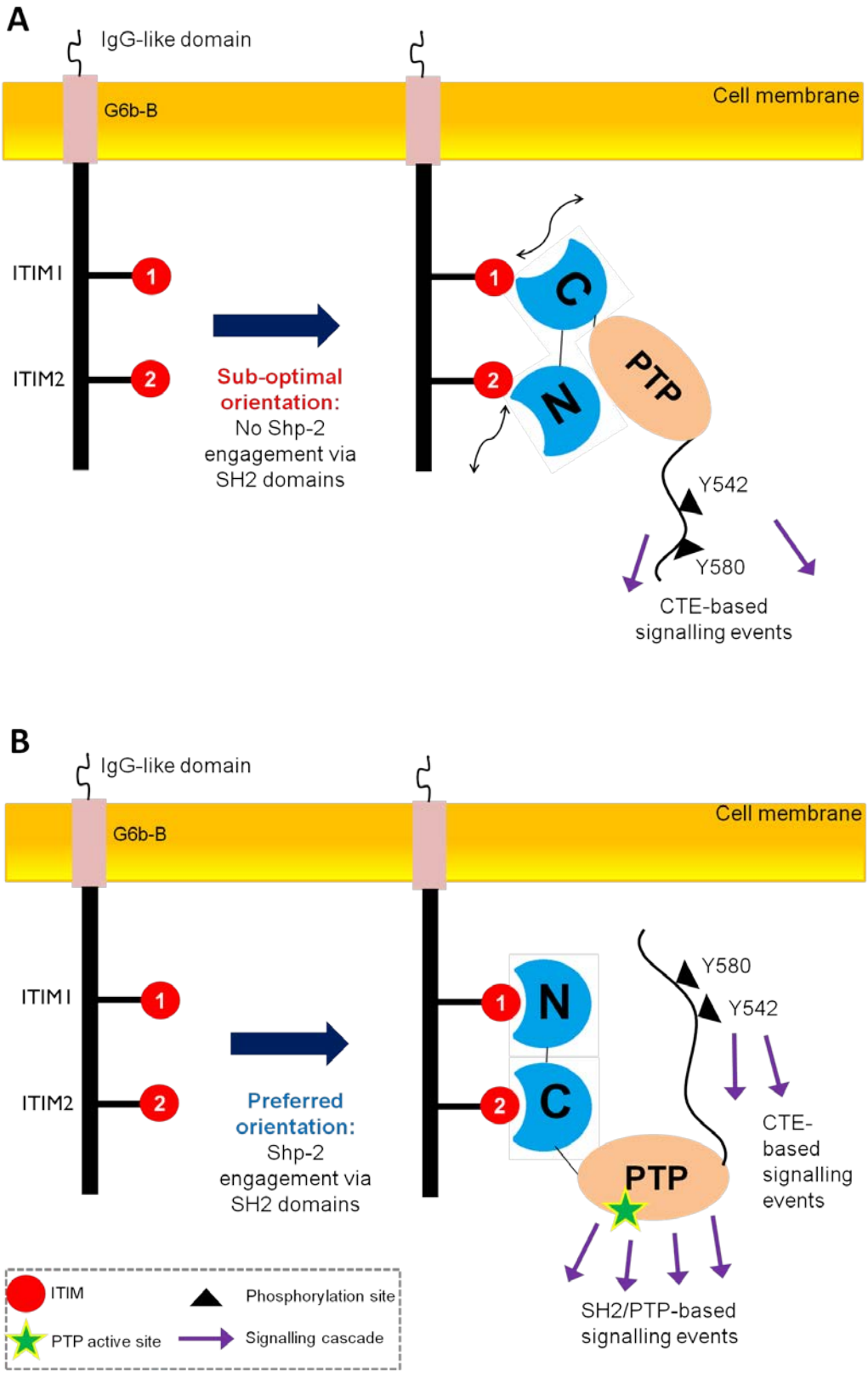

A
IgG-like domain
G6b-B
Cell membrane
ITIM1
ITIM2
1
2
Sub-optimal orientation:
No Shp-2 engagement via SH2 domains
C
N
PTP
Y542
Y580
CTE-based signalling events
B
IgG-like domain
G6b-B
Cell membrane
ITIM1
ITIM2
1
2
Preferred orientation:
Shp-2 engagement via SH2 domains
N
C
PTP
Y580
Y542
CTE-based signalling events
SH2/PTP-based signalling events
ITIM
Phosphorylation site
PTP active site
Signalling cascade

Figure 7.0: Model for Shp2:G6B-b Interaction

(A) Sub-optimal SH2 domain orientation during Shp2:G6B-b docking leads to ineffective signalling. The interaction between the ITIMs and TanSH2-Shp2 is too weak to activate dephosphorylation. The CTE is likely still functional and may become phosphorylated by nearby PTKs, causing Y542 and Y580 to act as binding sites for other proteins.

(B) When the TanSH2 of Shp2 binds G6B-b in its preferred orientation, a stabilized complex can form and lead to a marked increase in PTP activity, affecting downstream signalling cascades.

Limitations

There were a number of limitations of the present studies and these are outlined below.

Although Shp2 was the main focus in this research project, parallel collection and analysis of data from the closely related homologue Shp1 may have proved useful for comparing ITIM peptide binding and mechanisms of activation. More specifically, the nSH2 domain of Shp1 was insoluble post-overexpression in *E. coli*, preventing a direct comparison between it and nSH2-Shp2. Whilst attempts were made to refold nSH2-Shp1 from solubilised inclusion bodies, this was ultimately unsuccessful, though it is well documented for other SH2 domains (Babon *et al*., 2006; Haan *et al*., 1999). These results were corroborated by a recent publication during this study that had similar difficulties with nSH2-Shp1 (Coxon *et al*., 2012). Consequently, low expression yields of TanSH2-Shp1 are likely due to presence of the nSH2-Shp1 and as a consequence this construct was not pursued for further study.

The use of N-terminal His-tagged TanSH2 and various FL-Shp2 constructs resulted in pure and reasonable to high amounts of protein for biophysical studies. However, validation that the His-tag does not interfere with Shp2 ligand binding was not performed. Data suggest that there is a definite structural perturbation caused by the presence of an N-terminal His-tag and this could potentially affect normal ITIM peptide binding. This is an important question that needs to be addressed prior to further studies involving SH2-domain containing phosphatases.

Future Work

Although a myriad of experiments were performed in order to develop an understanding of Shp2 interactions with various biomolecules, many more research questions have been asked and as a result, description of further experimentation is described below to strengthen or otherwise test multiple hypotheses.

Further biophysical characterisation of full length Shp2 WT and mutant via AUC or dynamic light scattering (DLS) will further aid the investigation of structural reorganisation of the sub-domains in disease states. The information derived from these methods, primarily the hydrodynamic radius, may explain available SAXS data as well as provide a basis for further hypothesis testing. In the former scenario, the radius of gyration/hydrodynamic radius (R_g/R_h) ratio can provide information of the flexibility of Shp2 in solution (Bielnicki *et al*., 2011), where ratios close to 1 point toward compact molecules and ratios above 1.5 suggest flexibility (Rubinstein & Colby, 2003). In the latter scenario, hydrodynamic radii of various Shp2 mutants could provide information on the extent of conformational change, measured as a deviation from the WT R_h value.

From this research project, assignment data now exist for the nSH2 and cSH2 domains of Shp2. Introducing disease mutant substitutions into either of these individual domains, as well as the TanSH2 domain, and acquiring ^{15}N-HSQC spectra could yield insights into the extent of conformation change in solution. Deviation from WT can potentially be measured as CSPs depending on how close WT and mutant peaks are. This would yield residue level information concerning the nature of conformational change in various disease mutation contexts. In addition, CD

experiments can yield complementary secondary structure information so as to further investigate structural perturbations.

Following on from the ‘HSQC mutation screen’, one could select promising mutants for further characterisation by way of 3D backbone assignment spectra data collection and analysis. Taken together with side-chain proton assignments and NOE distance restraints a solution structure of a Shp2 SH2 domain can be solved. This is of particular importance as whilst various crystal structures of Shp2 mutants do exist (see Introduction), only two such structures have investigated SH2 domain mutations (Qiu *et al*., 2014)

In addition to NMR studies of unliganded mutant SH2 domains, peptide titration experiments similar to those carried out in this study could be performed for disease-relevant mutants, E76K being a leading choice as it is the most abundant mutation observed in Noonan syndrome and cancer settings (Tartaglia *et al*., 2003).

As described in the previous section, the presence of an N-terminal His-tag may cause aberrant ligand binding. Testing this hypothesis by efficient removal of the His-tag or production of a construct with a C-terminal tag instead would assist validation of efficient construct design.

SH2 domain residues critical in Shp2:G6b-B interactions can be delineated through site-directed mutagenesis of strongly perturbed residues identified by solution NMR binding studies. This may yield further information as to the residue-level determinants of Shp2 specificity for ITIM peptides as well as explain in part the differential affinity of Shp1 and Shp2 toward various biologically relevant peptide motifs. However, with the currently available data from the present and previous

studies, data-driven docking of TanSH2-Shp2 with dITIM may be possible. Using CSP data from solution NMR titration experiments and SAXS data of the TanSH2:dITIM complex, SAXS-assisted docking via HADDOCK could produce a reasonable model (Schmitz *et al*., 2012). However, since no dITIM-residue specific information of the TanSH2:dITIM interaction exists, further experiments such as filtered NOESY experiments may need to be performed in ordered to provide additional NOE restraints. The dITIM peptide is disordered in solution so flexible docking will have to be used to yield a better model (Bonvin, 2006).

Cell-based studies would bolster hypotheses regarding the functional relevance of Shp2 domain orientation within ITIM/ITAM/ITSM interactions. For example, switching of ITIM1 and ITIM2 sequences with one another so that ITIM1 is C-terminal to ITIM2 and observing changes in Shp1/Shp2 binding as well as consequences of downstream signalling pathways could potentially explain the need for peptide motif specificities of SH2 domains within the same polypeptide chain.

SAXS studies of various Shp2 constructs and mutants identified regions of flexibility. However, further validation with complimentary techniques may strengthen these hypotheses. Specifically, the dynamics of TanSH2-Shp2 can be investigated utilising NMR relaxation experiments collecting T_1, T_2 and Heteronuclear NOE data in both unbound and ligand-bound forms (Kay, 2005). A simpler approach employing limited proteolysis and peptide sequencing/mass spectrometry to identify surface exposed regions might also assist in confirming the dynamic nature of Shp2 in solution.

Final Remarks

Whilst X-ray crystallography and MD studies of Shp2 are numerous, little structural research has been performed in the solution state. This project has investigated the behaviour of Shp2 in solution using a variety of biophysical methods. Information has been obtained at the structural level both free and in complex with cognate ligands, further confirming the validity of crystallographic data and discovering the intrinsic flexible nature of Shp2 in wild-type and mutant forms. The knowledge gained from this study will be fed into further biophysical research as well as drug discovery efforts.

References

Aceto, N., and Bentires-Alj, M., 2012. Targeting protein-tyrosine phosphatases in breast cancer. Oncotarget. 2012 May;3(5):514-5.

Aceto, N., Sausgruber, N., Brinkhaus, H., Gaidatzis, D., Martiny-Baron, G., Mazzarol, G., Confalonieri, S., Quarto, M., Hu, G., Balwierz, P.J., Pachkov, M., Elledge, S.J., van Nimwegen, E., Stadler, M.B., Bentires-Alj, M., 2012. Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop. Nature Medicine. 2012 Mar 4;18(4):529-37.

Adachi, M., Sekiya, M., Miyachi, T., Matsuno, K., Hinoda, Y., Imai, K. & Yachi, A., 1992. Molecular cloning of a novel protein-tyrosine phosphatase SH-PTP3 with sequence similarity to the src-homology region 2. FEBS Letters, 314(3), 335–339.

Adachi, M., Iwaki, H., Shindoh, M., Akao, Y., Hachiya, T., Ikeda, M., Hinoda, Y., Imai, K., 1997. Predominant expression of the src homology 2-containing tyrosine phosphatase protein SHP2 in vascular smooth muscle cells. Virchows Archiv, 1997 Apr;430(4):321-5.

Ahmad, S., Banville, D., Zhao, Z., Fischer, E.H. & Shen, S. H., 1993. A widely expressed human protein-tyrosine phosphatase containing src homology 2 domains. Proceedings of the National Academy of Sciences of the United States of America, 90(6), 2197–2201.

Alonso, A., Sasin, J., Bottini, N., Friedberg, I., Osterman, A., Godzik, A., Hunter, T., Dixon, J. & Mustelin, T., 2004. Protein tyrosine phosphatases in the human genome. Cell, 117(6), 699–711.

Anafi, M., Rosen, M.K., Gish, G.D., Kay, L.E. & Pawson, T., 1996. A potential SH3 domain-binding site in the Crk SH2 domain. The Journal of Biological Chemistry, 271(35), 21365–21374.

Araki, T., Nawa, H. & Neel, B.G., 2003. Tyrosyl phosphorylation of Shp2 is required for normal ERK activation in response to some, but not all, growth factors. The Journal of Biological Chemistry, 278(43), 41677–41684.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I. & Stockinger, H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research, 40(W1), W597-W603.

Babon, J.J., McManus, E.J., Yao, S., DeSouza, D.P., Mielke, L.A., Sprigg, N.S., Willson, T.A., Hilton, D.J., Nicola, N.A., Baca, M., Nicholson, S.E. & Norton, R.S., 2006. The structure of SOCS3 reveals the basis of the extended SH2 domain function and identifies an unstructured insertion that regulates stability. Molecular Cell, 22(2), 205–216.

Barford, D., Das, A.K. & Egloff, M.P., 1998. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. Annual Review of Biophysics and Biomolecular Structure, 27, 133–164.

Barford, D., Flint, A.J. & Tonks, N.K., 1994. Crystal structure of human protein tyrosine phosphatase 1B. Science, 263(5152), 1397–1404.

Barr, A.J., Ugochukwu, E., Lee, W.H., King, O.N., Filippakopoulos, P., Alfano, I., Savitsky, P., Burgess-Brown, N.A., Müller, S. & Knapp, S., 2009. Large-scale structural analysis of the classical human protein tyrosine phosphatome. Cell, 136(2), 352–363.

Barrow, A.D. & Trowsdale, J., 2006. You say ITAM and I say ITIM, let's call the whole thing off: The ambiguity of immunoreceptor signalling. European Journal of Immunology, 36(7), 1646–1653.

Bax, A. & Ikura, M., 1991. An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the α-carbon of the preceding residue in uniformly 15N/13C enriched proteins. Journal of Biomolecular NMR, 1, 99–104.

Benes, C.H., Wu, N., Elia, A.E., Dharia, T., Cantley, L.C. & Soltoff, S.P., 2005. The C2 domain of PKCdelta is a phosphotyrosine binding domain. Cell, 121(2), 271–280.

Bentires-Alj, M., Paez, J.G., David, F.S., Keilhack, H., Halmos, B., Naoki, K., Maris, J.M., Richardson, A., Bardelli, A., Sugarbaker, D.J., Richards, W.G., Du, J., Girard, L., Minna, J.D., Loh, M.L., Fisher, D.E., Velculescu, V.E., Vogelstein, B., Meyerson, M., Sellers, W.R. & Neel, B.G., 2004. Activating mutations of the noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia. Cancer Research, 64(24), 8816–8820.

Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. & Svergun, D.I., 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. Journal of the American Chemical Society, 129(17), 5656–5664.

Bhattacharyya, R.P., Reményi, A., Yeh, B.J. & Lim, W.A., 2006. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. Annual Review of Biochemistry, 75, 655–680.

Bibbins, K.B., Boeuf, H. & Varmus, H.E., 1993. Binding of the Src SH2 domain to phosphopeptides is determined by residues in both the SH2 domain and the phosphopeptides. Molecular and Cellular Biology, 13(12), 7278–7287.

Bielnicki, J.A., Shkumatov, A.V., Derewenda, U., Somlyo, A.V., Svergun, D.I. & Derewenda, Z.S., 2011. Insights into the molecular activation mechanism of the RhoA-specific guanine nucleotide exchange factor, PDZRhoGEF. The Journal of Biological Chemistry, 286(40), 35163–75.

Billadeau, D.D. & Leibson, P.J., 2002. ITAMs versus ITIMs: striking a balance during cell regulation. The Journal of Clinical Investigation, 109(2), 161–168.

Blank, U., Launay, P., Benhamou, M. & Monteiro, R.C., 2009. Inhibitory ITAMs as novel regulators of immunity. Immunological Reviews, 232(1), 59–71.

Bodenhausen, G. & Ruben, D.J., 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chemical Physics Letters, 69, 185–189.

Bonvin, A.M., 2006. Flexible protein-protein docking. Current Opinion in Structural Biology, 16(2), 194–200.

Bossemeyer, D., 1995. Protein kinases--structure and function. FEBS letters, 369(1), 57–61.

Bradshaw, J.M., Grucza, R.A., Ladbury, J.E. & Waksman, G., 1998. Probing the “two-pronged plug two-holed socket” model for the mechanism of binding of the Src SH2 domain to phosphotyrosyl peptides: a thermodynamic study. Biochemistry, 37(25), 9083–9090.

Bradshaw, J.M. & Waksman, G., 2002. Molecular recognition by SH2 domains. Advances in Protein Chemistry, 61, 161–210.

Büchner, N., Altschmied, J., Jakob, S., Saretzki, G. & Haendeler, J., 2010. Well-known signaling proteins exert new functions in the nucleus and mitochondria. Antioxidants & Redox Signaling, 13(4), 551–558.

Van der Burgt, I., 2007. Noonan syndrome. Orphanet Journal of Rare Diseases, 2: 4.

Cannons, J.L., Tangye, S.G. & Schwartzberg, P.L., 2011. SLAM family receptors and SAP adaptors in immunity. Annual Review of Immunology, 29, 665–705.

Chemnitz, J.M., Parry, R.V., Nichols, K.E., June, C.H. & Riley, J.L., 2004. SHP-1 and SHP-2 associate with immunoreceptor tyrosine-based switch motif of programmed death 1 upon primary human T cell stimulation, but only receptor ligation prevents T cell activation. Journal of Immunology, 173(2), 945–954.

Chen, C.H., Piraner, D., Gorenstein, N.M., Geahlen, R.L. & Post, C.B., 2013. Differential recognition of syk-binding sites by each of the two phosphotyrosine-binding pockets of the Vav SH2 domain. Biopolymers, 99(11), 897–907.

Cohen, P., 2002. The origins of protein phosphorylation. Nature Cell Biology, 4(5), E127-E130.

Cole, C., Barber, J.D. & Barton, G.J., 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Research, 36, W197-W201.

Cowburn, D., 1996. Much more than the sum of the parts: structures of the dual SH2 domains of ZAP-70 and Syp. Chemistry & Biology, 3(2), 79–82.

Coxon, C.H., Sadler, A.J., Huo, J. & Campbell, R. D., 2012. An Investigation of hierachical protein recruitment to the inhibitory platelet receptor, G6B-b. PLoS ONE, 7(11), e49543.

Darian, E., Guvench, O., Yu, B., Qu, C.K. & MacKerell, A.D.Jr., 2011. Structural mechanism associated with domain opening in gain-of-function mutations in SHP2 phosphatase. Proteins, 79(5), 1573–1588.

Deribe, Y.L., Pawson, T. & Dikic, I., 2010. Post-translational modifications in signal integration. Nature Structural & Molecular Biology, 17(6), 666–672.

Diebold, M.-L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J. & Romier, C., 2010. Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. The Journal of Biological Chemistry, 285(49), 38389–38398.

Donaldson, L.W., Gish, G., Pawson, T., Kay, L.E. & Forma-Kay, J.D., 2002. Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. Proceedings of the National Academy of Sciences of the United States of America, 99(22), 14053–14058.

Doniach, S., 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. Chemical Reviews,101(6), 1763–1778.

Eck, M.J., Pluskey, S., Trüb, T., Harrison, S.C. & Shoelson, S.E., 1996. Spatial constraints on the recognition of phosphoproteins by the tandem SH2 domains of the phosphatase SH-PTP2. Nature, 379(6562), 277–280.

Eck, M.J., Shoelson, S.E. & Harrison, S.C., 1993. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. Nature, 362(6415), 87–91.

Edouard, T., Montagner, A., Dance, M., Conte, F., Yart, A., Parfait, B., Tauber, M., Salles, J.P. & Raynal, P., 2007. How do Shp2 mutations that oppositely influence its biochemical activity result in syndromes with overlapping symptoms? Cellular and Molecular Life Sciences, 64(13), 1585–1590.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper. U. & Sali, A., 2007. Comparative protein structure modeling using MODELLER. Current Protocols in Protein Science 50:2.9:2.9.1–2.9.31.

De Fabritiis, G., Geroult, S., Coveney, P.V. & Waksman, G., 2008. Insights from the energetics of water binding at the domain-ligand interface of the Src SH2 domain. Proteins, 72(4), 1290–1297.

Famiglietti, S.J., Nakamura, K. & Cambier, J.C., 1999. Unique features of SHIP, SHP-1 and SHP-2 binding to FcgammaRIIb revealed by surface plasmon resonance analysis. Immunology Letters, 68(1), 35–40.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J. D. & Kay, L.E., 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry, 33(19), 5984–6003.

Felli, I.C. & Brutscher, B., 2009. Recent advances in solution NMR: fast methods and heteronuclear direct detection. ChemPhysChem, 10(9-10), 1356–1368.

Filippakopoulos, P., Müller, S. & Knapp, S., 2009. SH2 domains: modulators of nonreceptor tyrosine kinase activity. Current Opinion in Structural Biology, 19(6), 643–649.

Franke, D. & Svergun, D.I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography, 42(2), 342–346.

Freeman, R.M., Plutzky, J. & Neel, B.G., 1992. Identification of a human src homology 2-containing protein-tyrosine-phosphatase: a putative homolog of Drosophila corkscrew. Proceedings of the National Academy of Sciences of the United States of America, 89(23), 11239–11243.

Frese, S., Schubert, W.D., Findeis, A.C., Marquardt, T., Roske, Y.S., Stradal, T.E. & Heinz, D.W., 2006. The phosphotyrosine peptide binding specificity of Nck1 and Nck2 Src homology 2 domains. The Journal of Biological Chemistry, 281(26), 18236–18245.

Fütterer, K., Wong, J., Grucza, R.A., Chan, A.C. & Waksman, G., 1998. Structural basis for Syk tyrosine kinase ubiquity in signal transduction pathways revealed by the

crystal structure of its regulatory SH2 domains bound to a dually phosphorylated ITAM peptide. Journal of Molecular Biology, 281(3), 523–537.

Gavrieli, M., Watanabe, N., Loftin, S.K., Murphy, T.L. & Murphy, K.M., 2003. Characterization of phosphotyrosine binding motifs in the cytoplasmic domain of B and T lymphocyte attenuator required for association with protein tyrosine phosphatases SHP-1 and SHP-2. Biochemical and Biophysical Research Communications, 312(4), 1236–1243.

Geahlen, R.L., 2009. Syk and pTyr'd: Signaling through the B cell antigen receptor. Biochimica et Biophysica Acta, 1793(7), 1115–1127.

Gelb, B.D. & Tartaglia, M., 2006. Noonan syndrome and related disorders: dysregulated RAS-mitogen activated protein kinase signal transduction. Human Molecular Genetics, 15(Spec No 2), R220–R226.

Gileadi, O., Burgess-Brown, N.A., Colebrook, S.M., Berridge, G., Savitsky, P., Smee, C.E., Loppnau, P., Johansson, C., Salah, E. & Pantic, N.H., 2008. High throughput production of recombinant human proteins for crystallography. Methods in Molecular Biology, 426, 221–246.

Grzesiek, S. & Bax, A., 1992. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. Journal of the American Chemical Society, 114(16), 6291–6293.

Grzesiek, S. & Sass, H.J., 2009. From biomolecular structure to functional understanding: new NMR developments narrow the gap. Current Opinion in Structural Biology, 19(5), 585–595.

Guvench, O., Qu, C.-K. & MacKerell, A.D., 2007. Tyr66 acts as a conformational switch in the closed-to-open transition of the SHP-2 N-SH2-domain phosphotyrosine-peptide binding cleft. BMC Structural Biology, 7, p.14.

Haan, S., Hemmann, U., Hassiepen, U., Schaper, F., Schneider-Mergener, J., Wollmer, A., Heinrich, P.C. & Grötzinger, J., 1999. Characterization and binding specificity of the monomeric STAT3-SH2 domain. The Journal of Biological Chemistry, 274(3), 1342–1348.

Haque, S.J., Harbor, P., Tabrizi, M., Yi, T., and Williams, B.R. Protein-tyrosine phosphatase Shp-1 is a negative regulator of IL-4- and IL-13-dependent signal transduction. Journal of Biological Chemistry. 1998; 273: 33893–33896

Hatada, M.H., Lu, X., Laird, E.R., Green, J., Morgenstern, J.P., Lou, M., Marr, C.S., Phillips, T.B., Ram, M.K. & Theriault, K., 1995. Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. Nature, 377(6544), 32–38.

Hatakeyama, M., 2004. Oncogenic mechanisms of the Helicobacter pylori CagA protein. Nature reviews. Cancer, 4(9), 688–694.

Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M.J. & Shoalson, S.E., 1998. Crystal structure of the tyrosine phosphatase SHP-2. Cell, 92(4), 441–450.

Hore, P.J., 1995. Nuclear magnetic resonance. Oxford: Oxford University Press.

Hu, X. & Stebbins, C.E., 2006. Dynamics of the WPD loop of the Yersinia protein tyrosine phosphatase. Biophysical Journal, 91(3), 948–956.

Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., Pawson, T., Gao, Y. & Li, S.S., 2008. Defining the specificity space of the human SRC homology 2 domain. Molecular Cellular Proteomics, 7(4), 768–784.

Huffine, M.E. & Scholtz, J.M., 1996. Energetic implications for protein phosphorylation. Conformational stability of HPr variants that mimic phosphorylated forms. The Journal of Biological Chemistry, 271(46), 28898–28902.

Hughes, C.E., Sinha, U., Pandey, A., Eble, J.A., O'Callaghan, C.A. & Watson, S.P., 2013. Critical Role for an acidic amino acid region in platelet signaling by the HemITAM (hemi-immunoreceptor tyrosine-based activation motif) containing receptor CLEC-2 (C-type lectin receptor-2). The Journal of Biological Chemistry, 288(7), 5127–5135.

Hunter, T. And Sefton B.M., 1980. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. Proceedings of the National Academy of Sciences of United States of America. 1980; 77(3): 1311–1315

Imhof, D., Wavreille, A.S., May, A., Zacharias, M., Tridandapani, S. & Pei, D., 2006. Sequence specificity of SHP-1 and SHP-2 Src homology 2 domains. Critical roles of residues beyond the pY+3 position. The Journal of Biological Chemistry, 281(29), 20271–20282.

Incoronato, M., D'Alessio, A., Paladino, S., Zurzolo, C., Carlomagni, M.S., Cerchia, L. & de Franciscis, V., 2004. The Shp-1 and Shp-2, tyrosine phosphatases, are recruited on cell membrane in two distinct molecular complexes including Ret oncogenes. Cellular Signalling, 16(7), 847–856.

Isakov, N., Wange, R.L., Burgess, W.H., Watts, J.D., Aebersold, R. & Samelson, L.E., 1995. ZAP-70 binding specificity to T cell receptor tyrosine-based activation motifs: the tandem SH2 domains of ZAP-70 bind distinct tyrosine-based activation motifs with varying affinity. The Journal of Experimental Medicine, 181(1), 375–380.

Isobe, M., Hinoda, Y., Imai, K. & Adachi, M., 1994. Chromosomal localization of an SH2 containing tyrosine phosphatase (SH-PTP3) gene to chromosome 12q24.1. Oncogene, 9(6), 1751–1753.

Jakob, S., Schroader, P., Lukosz, M., Büchner, N., Spyridopoulos, I., Altschmied, J. & Haendeler, J., 2008. Nuclear protein tyrosine phosphatase Shp-2 is one important negative regulator of nuclear export of telomerase reverse transcriptase. The Journal of Biological Chemistry, 283(48), 33155–33161.

Jia, Z., Barford, D., Flint, A.J. & Tonks, N.K., 1995. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. Science, 268(5218), 1754-1758.

Johnson, L.N. & Lewis, R.J., 2001. Structural basis for control by phosphorylation. Chemical Reviews, 101(8), 2209–2242.

Jongmans, M.C., van der Burgt, I., Hoogerbrugge, P.M., Noordam, K., Yntema, H.G., Nillesen, W.M., Kuiper, R.P., Ligtenberg, M.J., van Kessel, A.G., van Krieken, J.H., Kiemeney, L.A. & Hoogerbrugge, N., 2011. Cancer risk in patients with Noonan syndrome carrying a PTPN11 mutation. European Journal of Human Genetics, 19(8), 870–874.

Kandalaft, L.E., Singh, N., Liao, J.B., Facciabene, A., Berek, J.S., Powell, D.J. Jr., Coukos, G., 2010. The emergence of immunomodulation: combinatorial immunochemotherapy opportunities for the next decade. Gynecologic Oncology. 2010 Feb;116(2):222-33

Kaneko, T., Huang, H., Zhao, B., Li, L., Liu, H., Voss, C.K., Wu, C., Schiller, M.R. & Li, S.S., 2010. Loops govern SH2 domain specificity by controlling access to binding pockets. Science Signaling, 3(120):ra34.

Kaneko, T., Joshi R., Feller, S.M. & Li, S.S.C., 2012. Phosphotyrosine recognition domains: the typical, the atypical and the versatile. Cell Communication and Signaling, 10:32.

Karisch, R., Fernandez, M., Taylor, P., Virtanen, C., St-Germain, J.R., Jin, L.L., Harris, I.S., Mori, J., Mak, T.W., Senis, Y.A., Östman, A., Moran, M.F. & Neel, B.G., 2011. Global proteomic assessment of the classical protein-tyrosine phosphatome and “Redoxome”. Cell, 146(5), 826–840.

Kay, L.E., 2005. NMR studies of protein structure and dynamics. Journal of Magnetic Resonance , 173(2), 193–207.

Kay, L.E., Ikura, M., Tschudin, R. & Bax, A., 2011. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. Journal of Magnetic Resonance, 213(2), 423–441.

Keilhack, H., David, F.S., McGregor, M., Cantley, L.C. & Neel, B.G., 2005. Diverse biochemical properties of Shp2 mutants. Implications for disease phenotypes. The Journal of Biological Chemistry, 280(35), 30984–30993.

Kelly, S.M., Jess, T.J. & Price, N.C., 2005. How to study proteins by circular dichroism. Biochimica et Biophysica Acta - Proteins and Proteomics, 1751(2), 119–139.

Kelly, S.M. & Price, N.C., 2000. The use of circular dichroism in the investigation of protein structure and function. Current Protein & Peptide Science, 1(4), 349–384.

Kim, H.Y., Park, S.J., Joe, E.H. & Jou, I., 2006. Raft-mediated Src homology 2 domain-containing proteintyrosine phosphatase 2 (SHP-2) regulation in microglia. The Journal of Biological Chemistry, 281(17), 11872–11878.

Kleckner, I.R. & Foster, M.P., 2011. An introduction to NMR-based approaches for measuring protein dynamics. Biochimica et Biophysica Acta, 1814(8), 942–968.

Koch, M.H., Vachette, P. & Svergun, D.I., 2003. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Quarterly Reviews of Biophysics, 36(2), 147–227.

Konarev, P. V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I., 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography, 36(5), 1277–1282.

Kontaridis, M.I., Swanson, K.D., David, F.S., Barford, D. & Neel, B.G., 2006. PTPN11 (Shp2) mutations in LEOPARD syndrome have dominant negative, not activating, effects. The Journal of Biological Chemistry, 281(10), 6785–6792.

Kozin, M.B. & Svergun, D.I., 2001. Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography, 34(1), 33–41.

Kwan, A.H., Mobli, M., Gooley, P.R., King, G.F. & Mackay, J.P., 2011. Macromolecular NMR spectroscopy for the non-spectroscopist. The FEBS Journal, 278(5), 687–703.

Ladbury, J.E., Lemmon, M.A., Zhou, M., Green, J., Botfield, M.C. & Schlessinger, J., 1995. Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. Proceedings of the National Academy of Sciences of the United States of America, 92(8), 3199–3203.

Lapinski, P.E., Meyer, M.F., Feng, G.S., Kamiya, N. & King, P.D., 2013. Deletion of SHP-2 in mesenchymal stem cells causes growth retardation, limb and chest deformity, and calvarial defects in mice. Disease Models & Mechanisms, 6(6), 1448–1458.

Laue, T.M. & Stafford, W.F., 1999. Modern applications of analytical ultracentrifugation. Annual Review of Biophysics and Biomolecular Structure, 28, 75–100.

Lee, C.H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S.E. & Kuriyan, J., 1994. Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase. Structure, 2(5), 423–38.

Lemmon, M.A. & Schlessinger, J., 2010. Cell signaling by receptor tyrosine kinases. Cell, 141(7), 1117–1134.

Lescop, E., Schanda, P. & Brutscher, B., 2007. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. Journal of Magnetic Resonance, 187(1), 163–169.

Levitt, M.H., 2008. Spin dynamics: basics of nuclear magnetic resonance, 2nd edition. New Jersey: Wiley-Blackwell.

Lewis, R.J., Brannigan, J.A., Muchová, K., Barák, I. & Wilkinson, A.J., 1999. Phosphorylated aspartate in the structure of a response regulator protein. Journal of Molecular Biology, 294(1), 9–15.

Liu, B.A., Engelmann, B.W., Jablonowski, K., Higginbotham, K. Stergachis, A.B. & Nash, P.D., 2012. SRC Homology 2 Domain Binding Sites in Insulin, IGF-1 and FGF receptor mediated signaling networks reveal an extensive potential interactome. Cell Communication and Signaling, 10(1), 27.

Liu, B.A., Engelmann, B.W. & Nash, P.D., 2012. The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction. FEBS Letters, 586(17), 2597–2605.

Lochhead, P. & El-Omar, E.M., 2007. Helicobacter pylori infection and gastric cancer. Best practice & research. Clinical Gastroenterology, 21(2), 281–297.

Lorenz, U., 2011. Protein tyrosine phosphatase assays. Current Protocols in Immunology, Chapter 11, Unit 11.7.

Lu, W., Gong, D., Bar-Sagi, D & Cole, P.A., 2001. Site-specific incorporation of a phosphotyrosine mimetic reveals a role for tyrosine phosphorylation of SHP-2 in cell signaling. Molecular Cell, 8(4), 759–769.

Lu, W., Shen, K. & Cole, P.A., 2003. Chemical dissection of the effects of tyrosine phosphorylation of SHP-2. Biochemistry, 42(18), 5461–5468.

Machida, K. & Mayer, B.J., 2005. The SH2 domain: Versatile signaling module and pharmaceutical target. Biochimica et Biophysica Acta - Proteins and Proteomics, 1747(1), 1–25.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The protein kinase complement of the human genome. Science. 2002 Dec 6;298(5600):1912-34.

Marengere, L.E. & Pawson, T., 1992. Identification of residues in GTPase-activating protein Src homology 2 domains that control binding to tyrosine phosphorylated growth factor receptors and p62. The Journal of Biological Chemistry, 267(32), 22779–22786.

Marion, D., 2013. An introduction to biological NMR spectroscopy. Molecular & Cellular Proteomics, 12(11), 3006–3025.

Martinelli, S., Torreri, P., Tinti, M., Stella, L., Bocchinfuso, G., Flex, E., Grottesi, A., Ceccarini, M., Palleschi, A., Cesareni, G., Castagnoli, L., Petrucci, T.C.,Gelb, B.D. & Tartaglia, M., 2008. Diverse driving forces underlie the invariant occurrence of the T42A, E139D, I282V and T468M SHP2 amino acid substitutions causing Noonan and LEOPARD syndromes. Human Molecular Genetics, 17(13), 2018–2029.

Matthews, H.R., 1995. Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: a possible regulator of the mitogen-activated protein kinase cascade. Pharmacology and Therapeutics, 67(3), 323–350.

Mazharian, A., Mori, J., Wang, Y.J, Heising, S., Neel, B.G., Watson, S.P. & Senis, Y.A., 2013. Megakaryocyte-specific deletion of the protein-tyrosine phosphatases

Shp1 and Shp2 causes abnormal megakaryocyte development, platelet production, and function. Blood, 121(20), 4205–4220.

Mazharian, A., Wang, Y.J., Mori, J., Bem, D., Finney, B., Heising, S., Gissen, P., White, J.G., Berndt, M.C., Gardiner, E.E., Nieswandt, B., Douglas, M.R., Campbell, R.D., Watson, S.P. & Senis, Y.A., 2012. Mice lacking the ITIM-containing receptor G6b-B exhibit macrothrombocytopenia and aberrant platelet function. Science Signaling, 5(248), ra78.

Mei, L., Doherty, C.A. & Huganir, R.L., 1994. RNA splicing regulates the activity of a SH2 domain-containing protein tyrosine phosphatase. The Journal of Biological Chemistry, 269(16), 12254–12262.

Montalibet, J., Skorey, K.I. & Kennedy, B.P., 2005. Protein tyrosine phosphatase: enzymatic assays. Methods, 35(1), 2–8.

Mori, J., Pearce, A.C., Spalton, J.C., Grygielska, B., Eble, J.A., Tomlinson, M.G., Senis, Y.A. & Watson, S.P., 2008. G6b-B inhibits constitutive and agonist-induced signaling by glycoprotein VI and CLEC-2. The Journal of Biological Chemistry, 283(51), 35419–35427.

Mukherjee, M., Chow, S.Y., Yusoff, P., Seetharaman, J., Ng, C., Sinniah, S., Koh, X.W., Asgar, N.F., Li, D., Yim, D., Jackson, R.A., Yew, J., Qian, J., Iyu, A., Lim, Y.P., Zhou, X., Sze, S.K., Guy, G.R. & Sivaraman, J., 2012. Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin. The EMBO Journal, 31(5), 1308–1319.

Müller, P.J., Rigbolt, K.T., Paterok, D., Piehler, J., Vanselow, J., Lasonder, E., Andersen, J.S., Schaper, F. & Sobota, R.M., 2013. Protein tyrosine phosphatase SHP2/PTPN11 mistargeting as a consequence of SH2-domain point mutations associated with Noonan Syndrome and leukemia. Journal of Proteomics, 84, 132–147.

Mylonas, E. & Svergun, D.I., 2007. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. Journal of Applied Crystallography, 40, s245-s249.

Narula, S.S., Yuan, R.W., Adams, S.E., Green, O.M., Philips, T.B., Zydowsky, L.D., Botfield, M.C., Hatada, M., Laird, E.R., Zoller, M.J., Karas, J.L. & Dalgarno, D.C., 1995. Solution structure of the C-terminal SH2 domain of the human tyrosine kinase Syk complexed with a phosphotyrosine pentapeptide. Structure, 3(10), 1061–1073.

Neel, B.G., Gu, H. & Pao, L., 2003. The “Shp”ing news: SH2 domain-containing tyrosine phosphatases in cell signaling. Trends in Biochemical Sciences, 28(6), 284–293.

Nichols, K.E., Harkin, D.P., Levitz, S., Krainer, M., Kolquist, K.A., Genovese, C., Bernard, A., Ferguson, M., Zuo, L., Snyder, E., Buckler, A.J., Wise, C., Ashley, J., Lovett, M., Valentine, M.B., Look, A.T., Gerald, W., Housman, D.E. & Haber, D.A., 1998. Inactivating mutations in an SH2 domain-encoding gene in X-linked lymphoproliferative syndrome. Proceedings of the National Academy of Sciences of the United States of America, 95(23), 13765–13770.

Overduin, M., Rios, C.B., Mayer, B.J., Baltimore, D. & Cowburn, D., 1992. Three-dimensional solution structure of the src homology 2 domain of c-abl. Cell, 70(4), 697–704.

Paddock, C., Lytle, B.L., Peterson, F.C., Holyst, T., Newman, P.J., Volkman, B.F. & Newman, D.K., 2011. Residues within a lipid-associated segment of the PECAM-1 cytoplasmic domain are susceptible to inducible, sequential phosphorylation. Blood, 117(22), 6012–6023.

Pao, L.I., Badour, K., Siminovitch, K.A. & Neel, B.G., 2007. Nonreceptor protein-tyrosine phosphatases in immune cell signaling. Annual Review of Immunology, 25, 473–523.

Pawson, T. & Schlessinger, J., 1993. SH2 and SH3 domains. Current Biology, 3(7), 434–442.

Pearlman, S.M., Serber, Z. & Ferrell, J.E., 2011. A mechanism for the evolution of phosphorylation sites. Cell, 147(4), 934–946.

Pervushin, K., Riek, R., Wider, G. & Wüthrich, K., 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proceedings of the National Academy of Sciences of the United States of America, 94(23), 12366–12371.

Petoukhov, M. V & Svergun, D.I., 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophysical Journal, 89(2), 1237–1250.

Petoukhov, M. V., Konarev, P.V., Kikhney, A.G. & Svergun, D.I., 2007. ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. Journal of Applied Crystallography, 40, s223-s228.

Petoukhov, M. V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V. & Svergun, D.I., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography, 45(2), 342–350.

Pettersen EF1, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004 Oct;25(13):1605-12.

Poole, A.W. & Jones, M.L., 2005. A SHPing tale: perspectives on the regulation of SHP-1 and SHP-2 tyrosine phosphatases by the C-terminal tail. Cellular Signalling, 17(11), 1323–32.

Poy, F., Yaffe, M.B., Sayos, J., Saxena, K., Morra, M., Sumegi, J., Cantley, L.C., Terhorst, C. & Eck, M.J., 1999. Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. Molecular Cell, 4(4), 555–561.

Price, W.S., 1999. Water signal suppression in NMR spectroscopy. Annual Reports on NMR Spectroscopy, 38, pp.289–354.

Pumphrey, N.J., Taylor, V., Freeman, S., Douglas, M.R., Bradfield, P.F., Young, S.P., Lord, J.M., Wakelam, M.J., Bird, I.N., Salmon, M. & Buckley, C.D., 1999. Differential

association of cytoplasmic signalling molecules SHP-1, SHP-2, SHIP and phospholipase C-gamma1 with PECAM-1/CD31. FEBS Letters, 450(1-2), 77–83.

Qiu, W., Wang, X., Romanov, V., Hutchinson, A., Lin, A., Ruzanov, M., Battaile, K.P., Pai, E.F., Neel, B.G. & Chirgadze, N.Y., 2014. Structural insights into Noonan/ LEOPARD syndrome-related mutants of protein-tyrosine phosphatase SHP2 (PTPN11). BMC Structural Biology, 14(1), 10.

Radtke, S., Haan, S., Jörissen, A., Hermanns, H.M., Diefenbach, S., Smyczek, T., Schmitz-Vandeleur, H., Heinrich, P.C., Behrmann, I. & Haan, C., 2005. The Jak1 SH2 domain does not fulfill a classical SH2 function in Jak/STAT signaling but plays a structural role for receptor interaction and up-regulation of receptor surface expression. The Journal of Biological Chemistry, 280(27), 25760–25768.

Reich, N.C. & Liu, L., 2006. Tracking STAT nuclear traffic. Nature Reviews Immunology, 6(8), 602–612.

Ren, Y., Chen, Z., Chen, L., Fang, B., Win-Piazza, H., Haura, E., Koomen, J.M. & Wu, J., 2010. Critical role of Shp2 in tumor growth Involving regulation of c-Myc. Genes & Cancer, 1(10), 994–1007.

Reynolds, W.F., 2010. NMR pulse sequences. Lindon, J.C., Tranter, G.E. & Koppenaal, D. eds. Encyclopedia of spectroscopy and spectrometry, 2nd edition. London: Academic Press. pp.1554-1567.

Roy, A., Kucukural, A. & Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5(4), 725–738.

Rubinstein, M. & Colby, R., 2003. Polymer physics. Oxford: Oxford University Press.

Rubio, L., Huculeci, R., Buts, L., Vanwetswinkel, S., Lenaerts, T. & van Nuland, N.A.J., 2014. 1H, 13C, and 15N backbone and side-chain chemical shift assignments of the free and bound forms of the human PTPN11 second SH2 domain. Biomolecular NMR Assignments, 8(2), 297-301.

Rui, L., Mathews, L.S., Hotta, K., Gustafson, T.A. & Carter-Su, C., 1997. Identification of SH2-Bbeta as a substrate of the tyrosine kinase JAK2 involved in growth hormone signaling. Molecular and Cellular Biology, 17(11), 6633–6644.

Sadowski, I., Stone, J.C. & Pawson, T., 1986. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. Molecular and Cellular Biology, 6(12), 4396–4408.

Savitsky, P., Bray, J., Cooper, C.D., Marsden, B.D., Mahajan, P., Burgess-Brown, N.A., Gileadi, O., 2010. High-throughput production of human proteins for crystallization: the SGC experience. Journal of Structural Biology. 2010 Oct;172(1):3-13.

Schanda, P. & Brutscher, B., 2005. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. Journal of the American Chemical Society, 127(22), 8014–8015.

Schanda, P., Kupče, E. & Brutscher, B., 2005. SOFAST-HMQC experiments for recording two-dimensional deteronuclear correlation spectra of proteins within a few seconds. Journal of Biomolecular NMR, 33(4), 199–211.

Schanda, P., Van Melckebeke, H. & Brutscher, B., 2006. Speeding up three-dimensional protein NMR experiments to a few minutes. Journal of the American Chemical Society, 128(28), 9042–9043.

Schmitz, C., Melquiond, A.S.J., de Vries, S.J., Karaca, E., van Dijk, M., Kastritis, P.L. & Bonvin, M.J.J., 2012. Protein-protein docking with HADDOCK. In: Bertini, I., McGreevy, K.S. & Parigi, G. eds. NMR of biomolecules: towards mechanistic systems biology. Wenhein: Wiley-VCH. pp. 520–535.

Schnablegger, H. & Singh, Y., 2011. The SAXS guide: getting acquainted with the principles, 2nd edition. Austria: Anton Paar.

Schuck, P., 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophysical Journal, 78(3), 1606–1619.

Scott, L.M., Lawrence, H.R., Sebti, S.M. & Wu, J.,, 2010. Targeting protein tyrosine phosphatases for anticancer drug discovery. Current Pharmaceutical Design, 16(16), 1843–1862.

Sharma, D. & Rajarathnam, K., 2000. 13C NMR chemical shifts can predict disulfide bond formation. Journal of Biomolecular NMR, 18(2), 165–171.

Da Silva, H.B., Amaral, E.P., Nolasco, E.L., de Victo, N.C., Atique, R., Jank, C.C., Anschau, V., Zerbini, L.F. & Correa, R.G., 2013. Dissecting major signaling pathways throughout the development of prostate cancer. Prostate Cancer, 2013, 920612. 23 pages

Smew, P.A., Torok, A.R. & Kaiten, P.G. 2012 SMEW signalling underpins oxytocin release in mating pair interactions. Nature Smews, 1921:3, 87-88.

Smyth, S.S., McEver, R.P., Weyrich, A.S., Morrell, C.N., Hoffman, M.R., Arepally, G.M., French, P.A., Dauerman, H.L. & Becker, R.C. for the 2009 Platelet Colloquium Participants, 2009. Platelet functions beyond hemostasis. Journal of Thrombosis and Haemostasis, 7(11), 1759–1766.

Songyang, Z., Gish, G., Mbamalu, G., Pawson, T. & Cantley, L.C., 1995. A single point mutation switches the specificity of group III Src homology (SH) 2 domains to that of group I SH2 domains. The Journal of Biological Chemistry, 270(44), 26029–26032.

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J., Neel, B.G., Birge, R.B., Fajardo, J.E., Chou, M.M., Hanafusa, H., Schaffhausen, B. & Cantley, L.C., 1993. SH2 domains recognize specific phosphopeptide sequences. Cell, 72(5), 767–778.

Stein-Gerlach, M., Wallasch, C. & Ullrich, A., 1998. SHP-2, SH2-containing protein tyrosine phosphatase-2. International Journal of Biochemistry and Cell Biology, 30(5), 559–566.

Stoker, A.W., 2005. Protein tyrosine phosphatases and signalling. The Journal of Endocrinology, 185(1), 19–33.

Sugimoto, S., Lachleider, R.J., Shoelson, S.E., Neel, B.G. & Walsh, C.T., 1993. Expression, purification , and characterization of SH2-containing protein tyrosine phosphatase, SH-PTP2. The Journal of Biological Chemistry, 268(30), 22771-22776.

Svergun, D., Barberato, C. & Koch, M.H.J., 1995. CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. Journal of Applied Crystallography, 28(6), 768–773.

Tartaglia, M., Kalidas, K., Shaw, A., Song, X., Musat, D.L., van der Burgt, I., Brunner, H.G., Bertola, D.R., Crosby, A., Ion, A., Kucherlapati, R.S., Jeffery, S., Patton, M.A. & Gelb, B.D., 2002. PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-phenotype correlation, and phenotypic heterogeneity. American Journal of Human Genetics, 70(6), 1555–1563.

Tartaglia, M., Niemeyer, C.M., Fragale, A., Song, X., Buechner, J., Jung, A., Hählen, K., Hasle, H., Licht, J.D. & Gelb, B.D., 2003. Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. Nature Genetics, 34(2), 148–151.

Tokonzaba, E., Capelluto, D.G.S., Kutateladze, T.G. & Overduin, M., 2006. Phosphoinositide, phosphopeptide and pyridone interactions of the Abl SH2 domain. Chemical Biology & Drug Design, 67(3), 230–237.

Tompa, P., 2011. Unstructural biology coming of age. Current Opinion in Structural Biology, 21(3), 419–425.

Tonks, N.K., 2006. Protein tyrosine phosphatases: from genes, to function, to disease. Nature reviews. Molecular Cell Biology, 7(11), 833–846.

Tonks, N.K., Diltz, C.D. & Fischer, E.H., 1988. Characterization of the major protein-tyrosine-phosphatases of human placenta. The Journal of Biological Chemistry, 263(14), 6731–6737.

Ubersax, J.A. and Ferrell, J.E. Jr., 2007. Mechanisms of specificity in protein phosphorylation. Nature Reviews Molecular Cell Biology. 2007 Jul;8(7):530-41.

Uhlik, M.T., Temple, B., Bencharit, S., Kimple, A.J., Siderovski, D.P. & Johnson, G.L., 2005. Structural and evolutionary division of phosphotyrosine binding (PTB) domains. Journal of Molecular Biology, 345(1), 1–20.

Vély, F., Nunès, J.A., Malissen, B. & Hedgecock, C.J., 1997. Analysis of immunoreceptor tyrosine-based activation motif (ITAM) binding to ZAP-70 by surface plasmon resonance. European Journal of Immunology, 27(11), 3010–3014.

De Vet, E.C., Aguado, B. & Campbell, R.D., 2001. G6b, a novel immunoglobulin superfamily member encoded in the human major histocompatibility complex, interacts with SHP-1 and SHP-2. The Journal of Biological Chemistry, 276(45), 42070–42076.

Volkov, V. V. & Svergun, D.I., 2003. Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography, 36(3), 860–864.

Waksman, G., Kominos, D., Robertson, S.C., Pant, N., Baltimore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., Resh, M.D., Rios, C.B., Silverman, L., Kuriyan, J., 1992. Crystal-structure of the phosphotyrosine recognition domain (SH2) of the v-src tyrosine kinase complexed with tyrosine phosphorylated peptides. Nature, 358(6388), 646–653.

Walter, A.O., Peng, Z.Y. & Cartwright, C.A., 1999. The Shp-2 tyrosine phosphatase activates the Src tyrosine kinase by a non-enzymatic mechanism. Oncogene, 18(11), 1911–1920.

Whitmore, L. & Wallace, B.A., 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Research, 32, W668-W673.

Williamson, M.P., 2013. Using chemical shift perturbation to characterise ligand binding. Progress in Nuclear Magnetic Resonance Spectroscopy, 73, 1–16.

Wittekund M & Mueller L, 1993. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. Journal of Magnetic Resonance, 101(2), 201–205.

Wu, S. & Zhang, Y., 2008. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins, 72(2), 547–556.

Wu, Y. & Guo, J.F., 2009. Expression and purification of 15N-labeled 2-SH2 protein domain of SHP-2 from Homo sapiens in Escherichia coli for NMR studies and applications. International Journal of Biological Macromolecules, 45(1), 1–7.

Yang, J., Liu, L., He, D., Song, X., Liang, X., Zhao, Z.J. & Zhou, G.W., 2003. Crystal structure of human protein-tyrosine phosphatase SHP-1. The Journal of Biological Chemistry, 278(8), 6516–6520.

Yang, J., Niu, T., Zhang, A., Mishra, A.K., Zhao, Z.J. & Zhou, G.W., 2001. Relation between the flexibility of the WPD loop and the activity of the catalytic domain of protein tyrosine phosphatase SHP-1. Journal of Cellular Biochemistry, 84(1), 47–55.

Yee, A., Gutmanas, A. & Arrowsmith, C.H., 2006. Solution NMR in structural genomics. Current Opinion in Structural Biology, 16(5), 611–617.

You, M., Yu, D.H. & Feng, G.S., 1999. Shp-2 tyrosine phosphatase functions as a negative regulator of the interferon-stimulated Jak/STAT pathway. Molecular and Cellular Biology, 19(3), 2416–2424.

Yu, W.-M., Hawley, T.S., Hawley, R.G. & Qu, C.-K., 2003. Catalytic-dependent and -independent roles of SHP-2 tyrosine phosphatase in interleukin-3 signaling. Oncogene, 22(38), 5995–6004.

Yu, Z., Xu, J., Walls, C.D., Chen, L., Zhang, S., Zhang, R., Wu, L., Wang, L., Liu, S. & Zhang, Z.Y., 2013. Structural and mechanistic insights into LEOPARD syndrome-associated SHP2 mutations. The Journal of Biological Chemistry, 288(15), 10472-10482.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9, p.40.

Zhang, Y. & Skolnick, J., 2004. Scoring function for automated assessment of protein structure template quality. Proteins: Structure, Function and Genetics, 57(4), 702–710.

Zhang, Z.Y., Maclean, D., McNamara, D.J., Sawyer, T.K. & Dixon, J.E., 1994. Protein tyrosine phosphatase substrate specificity: size and phosphotyrosine positioning requirements in peptide substrates. Biochemistry, 33(8), 2285–2290.

Zhao, Z., Larocque, R., Ho, W.T., Fischer, E.H. & Shen, S.H., 1994. Purification and characterization of PTP2C, a widely distributed protein tyrosine phosphatase containing two SH2 domains. The Journal of Biological Chemistry, 269(12), 8780–8785.

Appendix

1. Human PTPN11 cDNA gene sequence (synthesised by Genscript)

ATGACCAGCCGTCGCTGGTTTCATCCGAATATTACGGGTGTTGAAGCGGAAAACCTGCT
GCTGACCCGTGGTGTGGATGGCAGTTTTCTGGCGCGTCCGAGCAAATCTAATCCGGGT
GATTTCACCCTGAGCGTGCGTCGCAATGGCGCAGTTACCCACATTAAAATCCAGAACAC
GGGCGATTATTACGATCTGTACGGCGGTGAAAAATTTGCCACCCTGGCAGAACTGGTTC
AGTATTACATGGAACATCACGGTCAGCTGAAAGAGAAAAACGGCGATGTGATCGAACTG
AAATATCCGCTGAACTGCGCAGATCCGACGAGTGAACGTTGGTTCCATGGTCACCTGAG
CGGCAAAGAAGCGGAAAAACTGCTGACCGAAAAAGGTAAACATGGCTCTTTTCTGGTGC
GTGAAAGTCAGAGCCACCCGGGTGATTTCGTTCTGAGTGTGCGCACGGGTGATGATAA
AGGCGAAAGTAATGATGGCAAAAGCAAAGTTACCCATGTGATGATTCGTTGTCAGGAAC
TGAAATACGATGTTGGCGGTGGCGAACGCTTTGATAGCCTGACCGATCTGGTGGAACA
CTATAAGAAAAACCCGATGGTGGAAACCCTGGGCACGGTTCTGCAGCTGAAACAGCCG
CTGAATACCACGCGCATTAACGCGGCCGAAATCGAAAGCCGTGTTCGCGAACTGTCTAA
ACTGGCGGAAACCACGGATAAAGTGAAACAGGGTTTTTGGGAAGAATTTGAAACCCTGC
AGCAGCAGGAATGCAAACTGCTGTACAGCCGTAAAGAAGGCCAGCGCCAGGAAAACAA
AAACAAAAACCGTTACAAAAACATCCTGCCGTTCGATCATACCCGCGTGGTTCTGCACG
ATGGTGATCCGAACGAACCGGTGTCTGATTACATCAATGCCAACATTATCATGCCGGAA
TTTGAAACCAAATGCAACAACTCTAAACCGAAAAAATCTTACATCGCAACCCAGGGCTGT
CTGCAGAATACGGTGAACGATTTTTGGCGTATGGTTTTCCAGGAAAACTCTCGCGTTATT
GTGATGACCACGAAAGAAGTTGAACGTGGTAAAAGTAAATGTGTGAAATACTGGCCGGA
TGAATACGCCCTGAAAGAATATGGCGTTATGCGTGTGCGCAATGTTAAAGAAAGCGCAG
CGCATGATTACACGCTGCGCGAACTGAAACTGTCTAAAGTTGGTCAGGGCAACACCGAA
CGTACGGTGTGGCAGTATCATTTTCGTACCTGGCCGGATCATGGTGTGCCGAGCGATC
CGGGTGGCGTTCTGGATTTCCTGGAAGAAGTGCATCACAAACAGGAATCTATTATGGAT
GCCGGTCCGGTGGTTGTGCATTGCAGTGCAGGTATCGGCCGTACCGGCACGTTCATCG
TTATCGATATCCTGATCGATATCATCCGCGAAAAGGGTGTGGATTGTGATATTGATGTTC
CGAAAACGATCCAGATGGTTCGTTCTCAGCGCAGTGGCATGGTGCAGACCGAAGCACA
GTATCGTTTCATCTACATGGCGGTGCAGCATTATATTGAAACCCTGCAGCGTCGCATCG
AAGAAGAACAGAAAAGTAAACGCAAAGGTCACGAATACACGAACATCAAATACAGCCTG
GCCGATCAGACCAGCGGCGATCAGTCTCCGCTGCCGCCGTGTACCCCGACGCCGCCG
TGTGCCGAAATGCGTGAAGATAGCGCCCGCGTGTATGAAAACGTTGGCCTGATGCAGC
AGCAGAAATCTTTTCGTTAA

Shp2 protein sequence:

```
>sp|Q06124-2|PTN11_HUMAN Isoform 2 of Tyrosine-protein phosphatase non-
receptor type 11 OS=Homo sapiens GN=PTPN11
MTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSVRRNGAVTHIKIQNTG
DYYDLYGGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPLNCADPTSERWFHGHLSGK
EAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQELKYD
VGGGERFDSLTDLVEHYKKNPMVETLGTVLQLKQPLNTTRINAAEIESRVRELSKLAETT
DKVKQGFWEEFETLQQQECKLLYSRKEGQRQENKNKNRYKNILPFDHTRVVLHDGDPNEP
VSDYINANIIMPEFETKCNNSKPKKSYIATQGCLQNTVNDFWRMVFQENSRVIVMTTKEV
ERGKSKCVKYWPDEYALKEYGVMRVRNVKESAAHDYTLRELKLSKVGQGNTERTVWQYHF
RTWPDHGVPSDPGGVLDFLEEVHHKQESIMDAGPVVVHCSAGIGRTGTFIVIDILIDIIR
```

EKGVDCDIDVPKTIQMVRSQRSGMVQTEAQYRFIYMAVQHYIETLQRRIEEEQKSKRKGH
EYTNIKYSLADQTSGDQSPLPPCTPTPPCAEMREDSARVYENVGLMQQQKSFR

WT FL-Shp2 Sequence:

MHHHHHHSSGVDLGTENLYFQSMTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSVRRNGAV
THIKIQNTG
DYYDLYGGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPLNCADPTSERWFHGHLSGK
EAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQELKYD
VGGGERFDSLTDLVEHYKKNPMVETLGTVLQLKQPLNTTRINAAEIESRVRELSKLAETT
DKVKQGFWEEFETLQQQECKLLYSRKEGQRQENKNKNRYKNILPFDHTRVVLHDGDPNEP
VSDYINANIIMPEFETKCNNSKPKKSYIATQGCLQNTVNDFWRMVFQENSRVIVMTTKEV
ERGKSKCVKYWPDEYALKEYGVMRVRNVKESAAHDYTLRELKLSKVGQGNTERTVWQYHF
RTWPDHGVPSDPGGVLDFLEEVHHKQESIMDAGPVVVHCSAGIGRTGTFIVIDILIDIIR
EKGVDCDIDVPKTIQMVRSQRSGMVQTEAQYRFIYMAVQHYIETLQRRIEEEQKSKRKGH
EYTNIKYSLADQTSGDQSPLPPCTPTPPCAEMREDSARVYENVGLMQQQKSFR

E76K FL-Shp2 Sequence:

MHHHHHHSSGVDLGTENLYFQSMTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSVRRNGAV
THIKIQNTG
DYYDLYGGEKFATLA**K**LVQYYMEHHGQLKEKNGDVIELKYPLNCADPTSERWFHGHLSGK
EAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQELKYD
VGGGERFDSLTDLVEHYKKNPMVETLGTVLQLKQPLNTTRINAAEIESRVRELSKLAETT
DKVKQGFWEEFETLQQQECKLLYSRKEGQRQENKNKNRYKNILPFDHTRVVLHDGDPNEP
VSDYINANIIMPEFETKCNNSKPKKSYIATQGCLQNTVNDFWRMVFQENSRVIVMTTKEV
ERGKSKCVKYWPDEYALKEYGVMRVRNVKESAAHDYTLRELKLSKVGQGNTERTVWQYHF
RTWPDHGVPSDPGGVLDFLEEVHHKQESIMDAGPVVVHCSAGIGRTGTFIVIDILIDIIR
EKGVDCDIDVPKTIQMVRSQRSGMVQTEAQYRFIYMAVQHYIETLQRRIEEEQKSKRKGH
EYTNIKYSLADQTSGDQSPLPPCTPTPPCAEMREDSARVYENVGLMQQQKSFR

His-TanSH2-Shp2

MHHHHHHSSGVDLGTENLYFQSM R R W F H P N I T G V E A E N L L L T R G V
D G S F L A R P S K S N P G D F T L S V R R N G A V T H I K I Q N T
G D Y Y D L Y G G E K F A T L A E L V Q Y Y M E H H G Q L K E K N G
D V I E L K Y P L N C A D P T S E R W F H G H L S G K E A E K L L T
E K G K H G S F L V R E S Q S H P G D F V L S V R T G D D K G E S N
D G K S K V T H V M I R C Q E L K Y D V G G G E R F D S L T D L V E
H Y K K N P M V E T L G T V L Q L K Q P L

ΔHis-TanSH2-Shp2

SM R R W F H P N I T G V E A E N L L L T R G V D G S F L A R P S K S
N P G D F T L S V R R N G A V T H I K I Q N T G D Y Y D L Y G G E K
F A T L A E L V Q Y Y M E H H G Q L K E K N G D V I E L K Y P L N C
A D P T S E R W F H G H L S G K E A E K L L T E K G K H G S F L V R
E S Q S H P G D F V L S V R T G D D K G E S N D G K S K V T H V M I
R C Q E L K Y D V G G G E R F D S L T D L V E H Y K K N P M V E T L
G T V L Q L K Q P L

nSH2-Shp2

GSGRAMTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSVRRNGAVTHIKIQNTG

DYYDLYGGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPLNC**LINEF**

cSH2-Shp2

GSGRAMPLNCADPTSERWFHGHLSGK

EAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKGESNDGKSKVTHVMIRCQELKYD

VGGGERFDSLTDLVEHYKKNPMVETLGTVLQLKQPLNTTRIN**LINEF**

Legend:

Red *single* amino acid: E76K mutation in E76K FL-Shp2

Red in ΔHis and His-TanSH2-Shp2 indicates nSH2 domain

Orange in ΔHis and His-TanSH2-Shp2 indicates cSH2 domain

Purple within nSH2 and cSH2-Shp2 represent terminal non-WT sequences.

Underlined sequences represent N-terminal His-tag and TEV cleavage site (if "SM", remaining TEV site residues post-cleavage.

NB: A similar colour coding system is used in appendix figure 4.

---

2. Full length G6B-b Amino acid Sequence

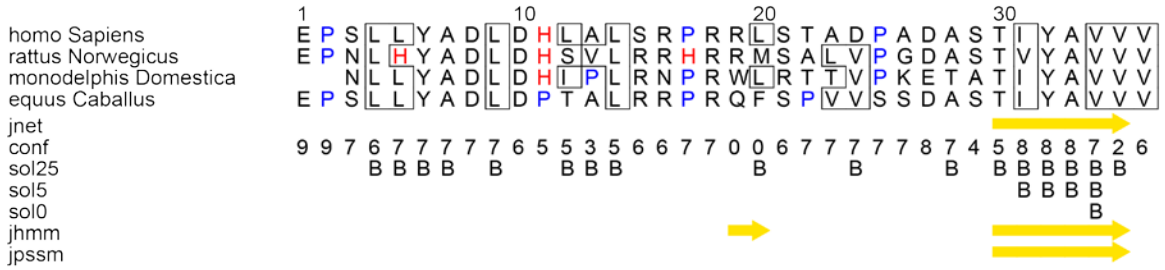
>sp|O95866|G6B_HUMAN Protein G6b OS=Homo sapiens GN=G6B PE=1 SV=1
MAVFLQLLPLLLSRAQGNPGASLDGRPGDRVNLSCGGVSHPIRWVWAPSFPACKGLSKGR
RPILWASSSGTPTVPPLQPFVGRLRSLDSGIRRLELLLSAGDSGTFFCKGRHEDESRTVL
HVLGDRTYCKAPGPTHGSVYPQLLIPLLGAGLVLGLGALGLVWWLHRRLPPQPIRPLPRF
APLVKTEPQRPVKEEEPKIPGDLDQEPSLLYADLDHLALSRPRRLSTADPADASTIYAVV
V

dITIM sequence used in study is indicated in red.

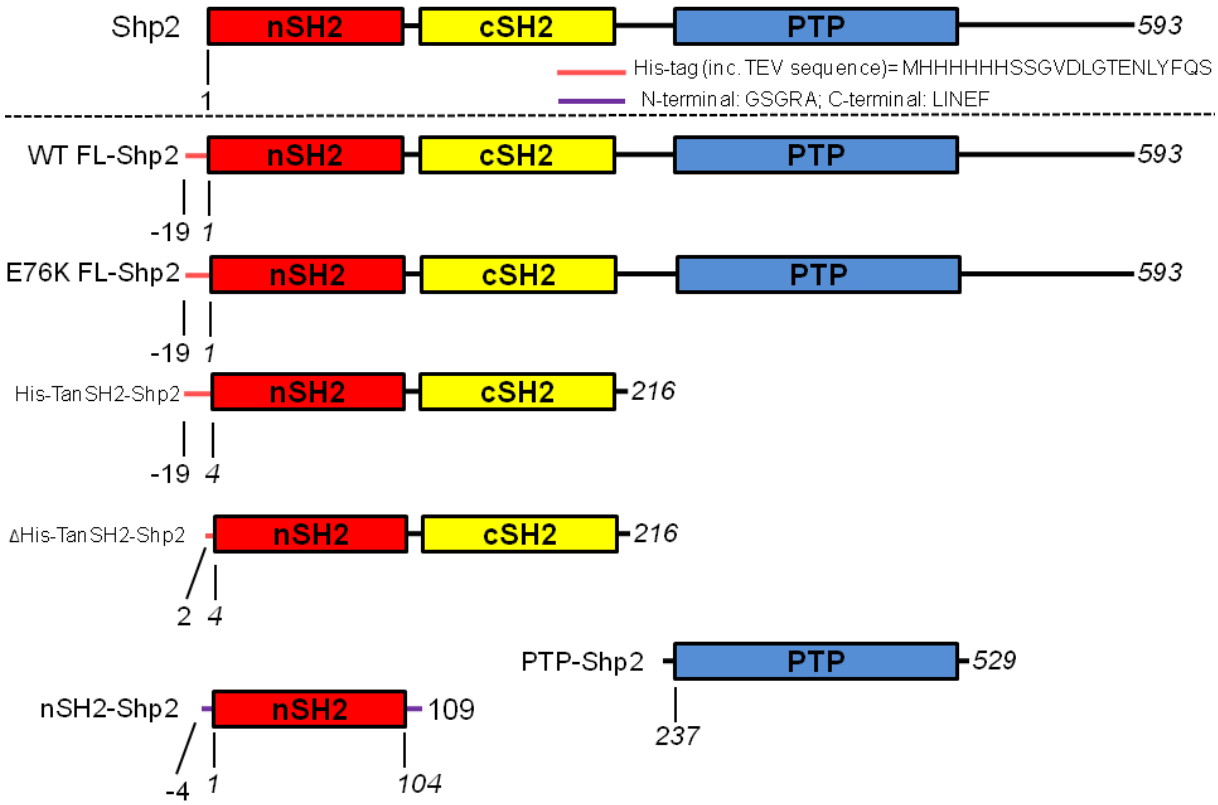
---

3. Chapter VI: Secondary Structure Prediction of G6b-B dITIM.

Jpred was used to predict the presence of helical and sheet character within human G6b-B unphosphorylated dITIM. B= beta sheet, conf= confidence score (0-9). Yellow arrows indicate beta sheet location in relation to sequence.



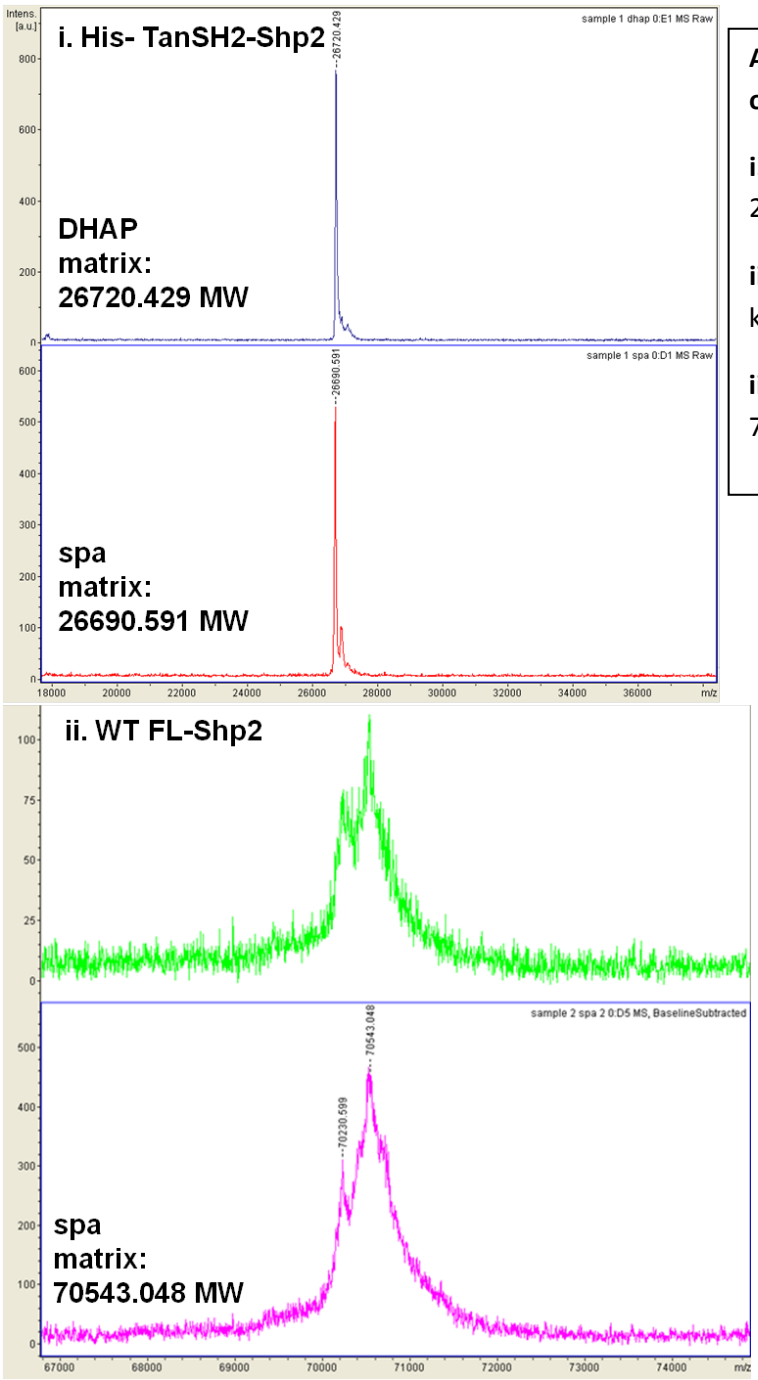
4. Summary of Constructs used in study:



Appendix Figure 4: Summary of Constructs used in study:

Full-length Shp2 is drawn at the top for reference (Shp2). All schematic drawings below dashed line are constructs used in this study, with names to the left of them. Black lines represent wild-type linkers between domains or unstructured elements. Orange lines represent His-tags left on after purification, while purple lines represent non-wild type residues either side of individual SH2 constructs nSH2 and cSH2-Shp2. Red box: nSH2 domain, Yellow box: cSH2 domain, Blue box: PTP domain. Numbers represent amino acid number; those in italics represent wild-type residue numbers, the rest are non-wild-type as part of a stability sequence or hangover post-cleavage.

5. Mass Spectrometry of Selected Constructs:

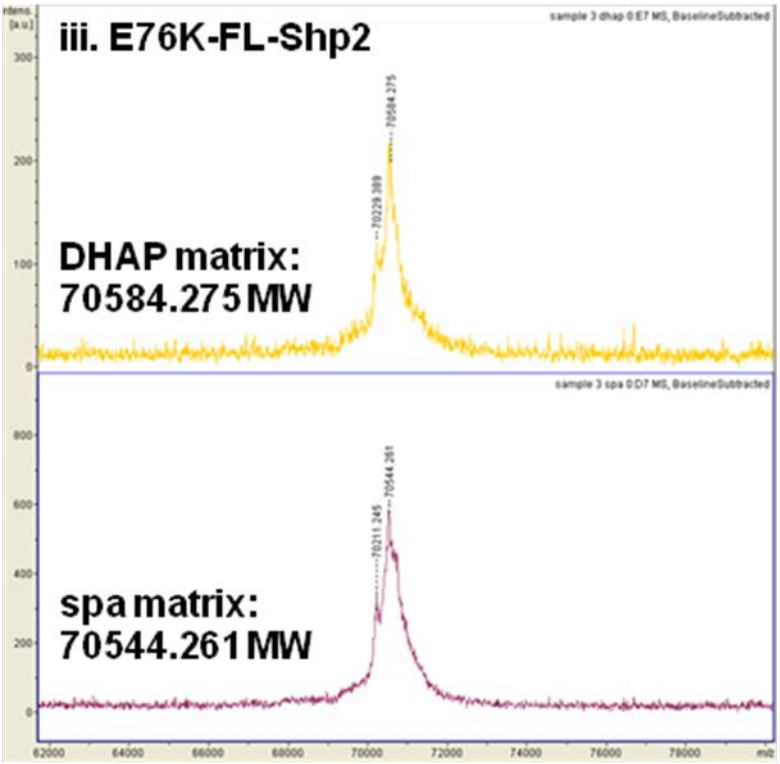


Appendix Figure 5: MALDI-TOF MS of TanSH2, WT FL and E76K FL-Shp2

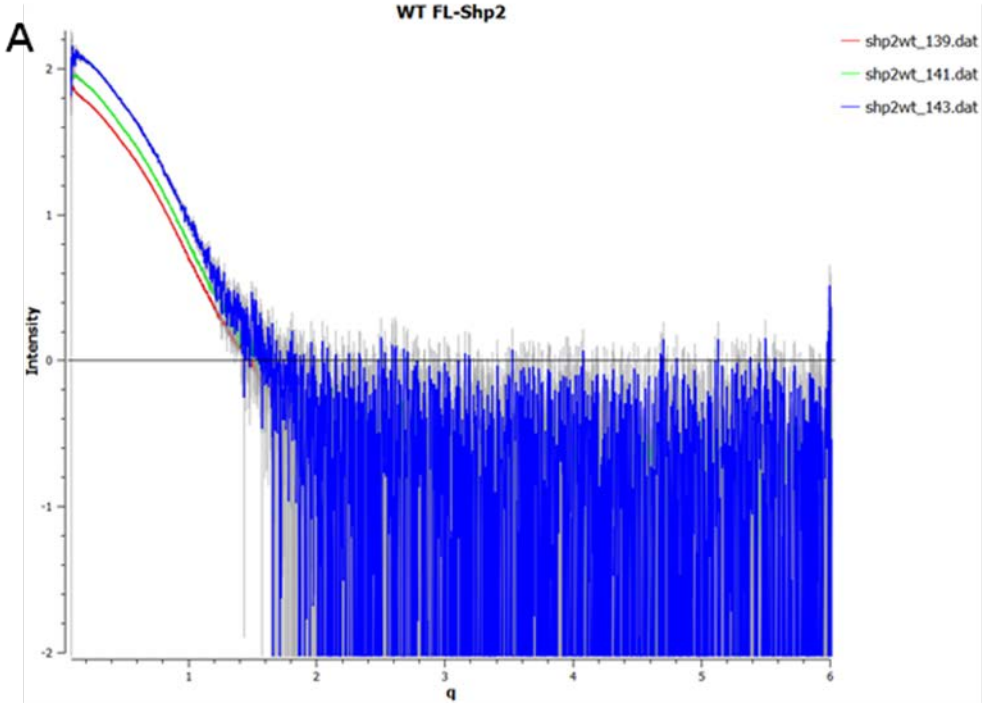
i. His-TanSH2-Shp2: Expected size is 26.2 kDa

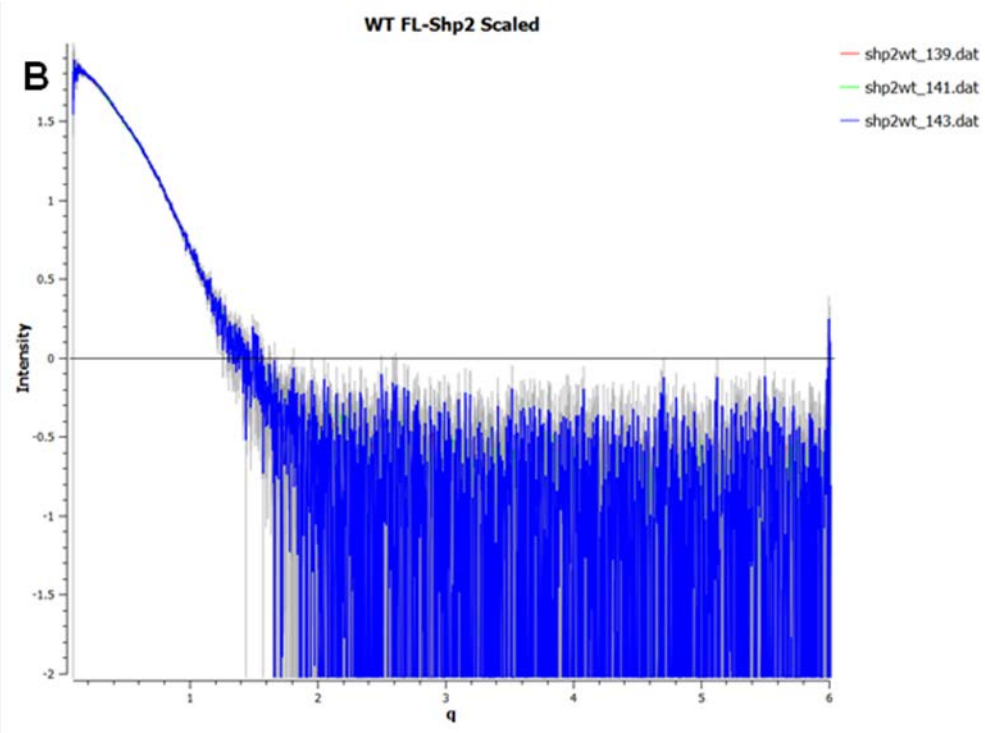
ii. WT FL-Shp2: Expected size is 70.5 kDa.

iii. E76K FL-Shp2: Expected size: 70.5 kDa.



Appendix Figure 6: Chapter V WT FL-Shp2 SAXS Curves





Appendix Figure 6: Chapter V; WT FL-Shp2 SAXS Data

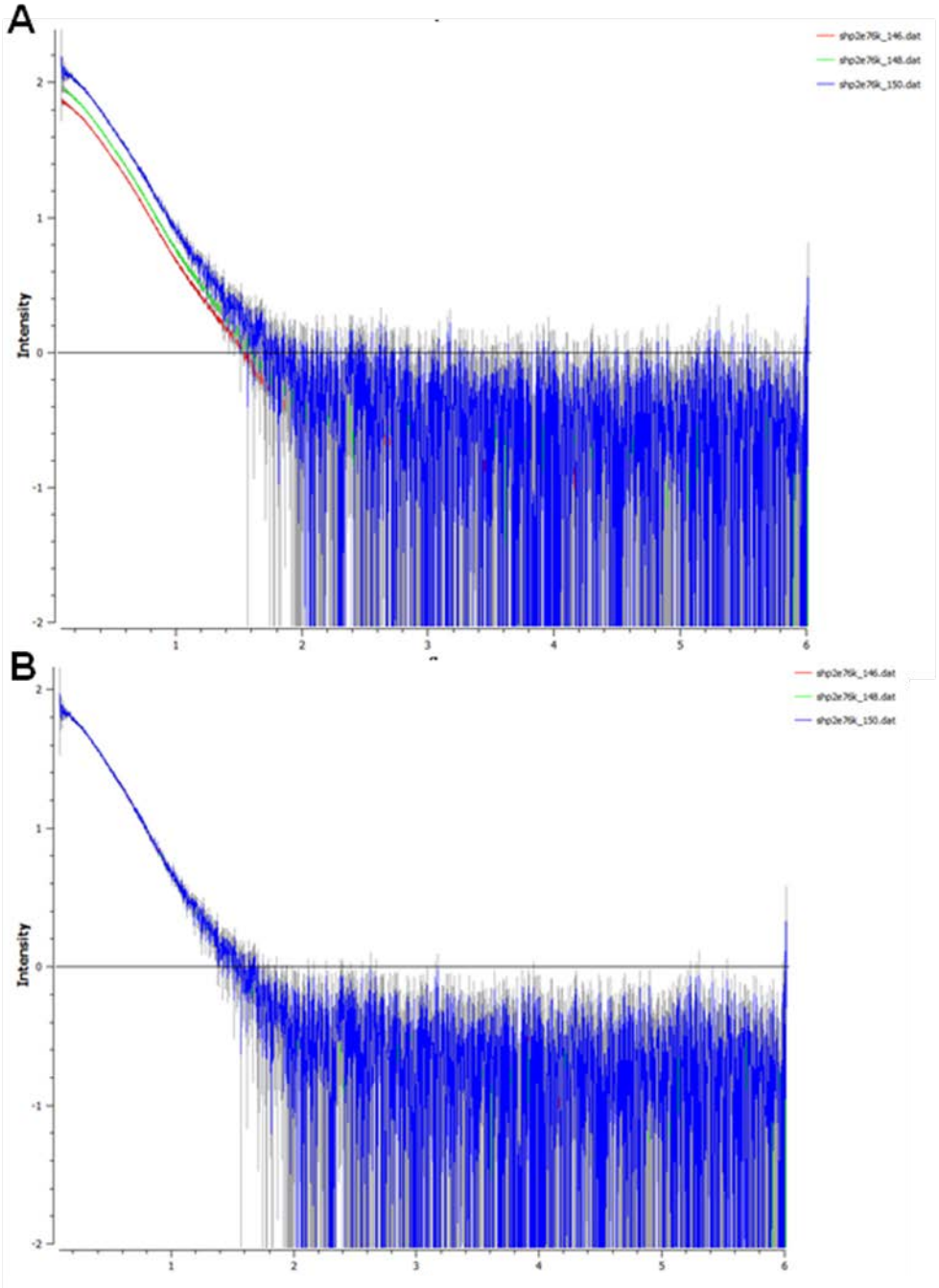
Shp2wt_139, 141 and 143 are 10, 5 and 1 mg/ml respectively.

(A) Unscaled

(B) Scaled

Manipulations were done in Primus Qt.

Appendix Figure 7: Chapter V; E76K FL-Shp2 SAXS Data



Appendix Figure 7: Chapter V; E76K FL-Shp2 SAXS Data

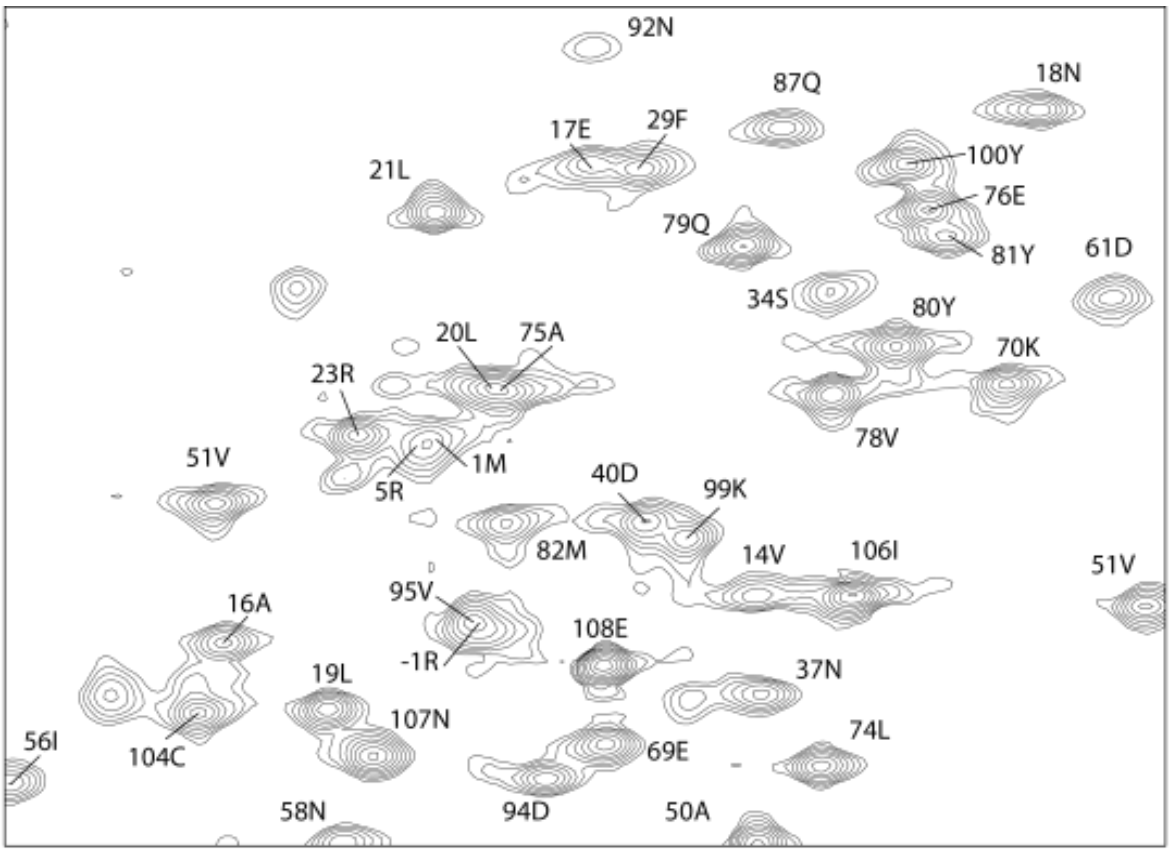
Shp2e76k_146, 148 and 150 are 10, 5 and 1 mg/ml respectively.

(A) Unscaled

(B) Scaled

Manipulations were done in Primus Qt

Appendix figure 8: Chapter VI; Sequential backbone assignment of nSH2:ITIM1s: Inset figure.



9. Chapter VI SAXS invariant parameters for unbound and dITIM-bound TanSH2 Shp2.

State	Conc (mg/ml)	Avg Rg nm)	Avg I(0)	Dmax (nm)
Unbound	9.26	2.25	1213.43	7.23
Unbound	4.63	2.33	1000.16	8.08
Unbound	2.32	2.43	1158.43	7.86
Bound	9.26	2.54	1324.36	8.38
Bound	4.63	2.39	1318.95	7.37
Bound	2.32	2.57	1322.50	7.68

10. Chemical Shift Lists of Assigned Proteins

nSH2-Shp2 Apo

!		H	N	C	CA	CB
-2	Gly	-	-	-	45.46	-
-1	Arg	8.28	120.98	176.19	56.21	30.93
0	Ala	8.38	124.76	177.83	52.74	19.25
1	Met	8.36	119.20	176.33	55.82	32.49
2	Thr	7.85	112.45	173.86	61.67	69.82
5	Arg	-	-	175.10	58.07	30.07
5	Arg	-	-	-	-	-
6	Trp	6.31	109.50	175.00	54.76	30.02
7	Phe	7.77	123.86	173.49	57.30	39.73
8	His	8.63	129.37	173.09	52.35	33.81
9	Pro	-	-	178.09	63.49	32.86
10	Asn	-	-	-	-	-
10	Asn	8.34	113.31	174.77	54.02	39.32
11	Ile	6.83	113.21	175.50	60.55	41.48
12	Thr	8.42	112.90	175.87	60.15	71.99
13	Gly	9.65	110.08	176.01	47.80	-
14	Val	7.78	120.30	178.81	65.74	32.00
15	Glu	7.40	121.02	179.56	59.17	30.01
16	Ala	8.72	121.57	178.83	55.17	18.89
17	Glu	8.35	116.93	177.18	60.20	28.75
18	Asn	7.77	116.09	178.74	56.99	38.72
18	Asn	-	-	-	-	-
19	Leu	8.51	121.78	179.03	58.42	42.67
20	Leu	8.28	118.62	179.37	58.05	42.47
21	Leu	8.32	116.89	177.79	56.76	41.38
22	Thr	7.74	109.22	176.16	63.56	70.59

23	Arg	8.42	119.28	176.57	55.72	31.95
24	Gly	6.99	106.72	170.99	43.87	-
25	Val	8.94	109.67	176.24	59.12	36.12
26	Asp	8.93	122.90	177.00	57.84	38.99
27	Gly	8.66	112.92	175.91	44.84	-
28	Ser	9.31	122.26	172.80	60.62	64.39
29	Phe	8.00	116.31	170.52	55.74	44.37
30	Leu	9.23	115.07	174.49	54.24	45.00
31	Ala	9.65	121.82	174.27	50.06	23.51
32	Arg	9.00	116.80	172.38	53.89	30.19
33	Pro	-	-	176.55	63.11	32.28
34	Ser	7.76	116.97	-	57.19	63.98
35	Lys	-	-	178.36	56.51	32.42
36	Ser	8.75	115.43	175.06	60.61	63.16
37	Asn	-	-	-	-	-
37	Asn	7.81	121.06	170.55	50.83	39.63
38	Pro	-	-	177.42	64.37	31.63
39	Gly	8.81	114.98	173.87	45.39	-
40	Asp	7.99	119.51	176.62	53.63	41.97
41	Phe	8.91	116.99	173.90	57.18	44.37
42	Thr	9.59	117.79	172.92	62.35	72.41
43	Leu	9.50	127.76	175.45	53.39	44.69
44	Ser	9.28	123.78	172.46	58.90	64.75
45	Val	8.94	123.42	173.89	60.12	35.75
46	Arg	9.11	126.97	175.41	54.54	32.11
47	Arg	9.44	130.91	174.83	55.58	32.33
48	Asn	-	-	-	-	-
48	Asn	9.29	125.25	175.43	54.41	37.52

49	Gly	9.61	105.94	172.20	46.03	-
50	Ala	7.96	122.59	175.05	50.45	22.26
51	Val	8.68	119.94	174.76	61.13	34.11
52	Thr	8.86	124.26	171.07	60.08	71.27
53	His	8.41	125.29	175.37	54.77	32.44
54	Ile	9.80	127.56	174.90	60.43	40.72
55	Lys	8.58	128.16	175.39	57.97	33.17
56	Ile	8.76	122.96	175.44	60.22	39.84
57	Gln	8.75	128.49	174.73	55.62	29.02
58	Asn	8.53	122.87	176.04	51.04	40.12
58	Asn	-	-	-	-	-
59	Thr	8.18	116.01	175.08	61.38	69.83
60	Gly	8.17	113.63	173.82	46.00	-
61	Asp	7.68	118.31	174.70	54.69	42.37
62	Tyr	7.20	114.25	172.80	56.70	40.72
63	Tyr	9.58	118.51	175.51	56.42	41.78
64	Asp	9.05	120.69	174.44	53.66	43.65
65	Leu	8.07	118.65	178.26	53.40	41.03
66	Tyr	8.59	118.95	176.00	60.37	35.30
67	Gly	7.51	107.13	174.43	44.64	-
68	Gly	8.36	107.59	174.05	44.84	-
69	Glu	8.22	122.15	174.09	56.73	30.08
70	Lys	7.69	118.52	176.40	54.91	34.89
71	Phe	9.13	118.11	175.95	57.15	44.09
72	Ala	9.41	123.36	177.97	54.46	19.95
73	Thr	7.17	101.40	173.77	58.82	73.71
74	Leu	7.89	122.09	178.10	56.96	41.07
75	Ala	8.21	118.64	179.65	55.49	18.03

76	Glu	7.80	117.05	177.82	59.41	31.00
77	Leu	6.91	122.53	177.26	58.70	41.82
78	Val	7.90	118.76	177.18	67.38	31.66
79	Gln	7.98	117.23	177.72	59.05	28.06
79	Gln	-	-	-	-	-
80	Tyr	7.82	118.31	179.63	62.23	38.91
81	Tyr	7.74	117.20	177.70	62.82	38.06
82	Met	8.27	119.89	177.45	59.35	34.00
83	Glu	7.40	114.85	175.99	56.78	30.57
84	His	7.40	118.20	173.84	54.94	27.00
86	Gly	-	-	174.16	45.32	-
87	Gln	7.94	115.96	175.39	56.24	29.65
87	Gln	-	113.01	-	-	-
88	Leu	8.49	124.23	174.34	54.69	41.59
89	Lys	8.06	123.74	176.91	54.34	36.88
90	Glu	8.73	119.85	178.56	55.45	31.41
91	Lys	9.01	122.96	177.28	59.32	31.99
92	Asn	-	-	-	-	-
92	Asn	7.99	114.95	176.27	52.93	37.58
93	Gly	8.11	108.66	173.91	44.94	-
94	Asp	8.07	121.99	175.74	54.88	40.51
95	Val	8.31	121.78	175.99	63.18	32.90
96	Ile	8.49	127.61	173.98	59.71	39.15
97	Glu	8.16	124.22	175.33	55.10	31.82
98	Leu	8.64	124.88	174.72	53.00	39.85
99	Lys	8.06	120.68	176.37	58.51	34.82
100	Tyr	7.80	116.33	172.37	52.89	40.66
101	Pro	-	-	176.26	62.68	32.14

102	Leu	8.71	128.49	175.57	54.43	41.86
103	Asn	8.56	125.62	175.03	53.51	38.98
103	Asn	-	-	-	-	-
104	Cys	8.63	121.89	174.74	59.18	27.21
105	Leu	8.70	128.05	176.89	55.12	42.01
106	Ile	7.83	120.57	175.47	61.14	38.80
107	Asn	8.42	122.17	174.55	53.21	39.25
107	Asn	-	-	-	-	-
108	Glu	8.15	121.24	174.80	56.67	30.63
109	Phe	7.66	124.98	180.18	58.98	40.54

---

cSH2-Shp2 Apo

!		**H**	**N**	**C**	**CA**	**CB**
97	Gly	-	-	173.83	45.27	-
98	Arg	8.19	120.83	175.84	55.85	31.10
99	Ala	8.41	125.99	177.28	52.37	19.14
100	Met	8.29	121.11	174.21	53.07	32.51
101	Pro	-	-	176.66	63.06	32.06
102	Leu	8.34	122.08	177.21	55.44	42.48
103	Asn	8.46	118.69	174.97	53.24	-
104	Cys	8.26	119.46	173.96	58.64	28.15
105	Ala	8.37	125.93	176.99	52.38	19.24
106	Asp	8.19	121.17	175.39	52.01	41.21
107	Pro	-	-	176.37	63.93	31.49
108	Thr	7.57	110.19	173.96	62.55	68.68
109	Ser	7.92	116.97	174.53	58.47	63.74
110	Glu	7.67	121.91	178.06	55.48	28.08

111	Arg	8.81	120.20	174.94	57.53	29.70
112	Trp	6.31	109.68	176.00	54.34	29.51
113	Phe	7.68	123.83	175.00	57.65	39.23
114	His	9.15	125.85	175.20	56.49	33.76
115	Gly	-	-	173.75	47.64	-
116	His	8.77	126.82	173.01	56.46	28.12
117	Leu	7.78	129.85	174.89	54.68	45.15
118	Ser	8.85	123.05	175.47	57.45	65.63
119	Gly	9.87	111.52	176.46	47.76	-
120	Lys	8.20	120.61	180.06	58.91	32.23
121	Glu	7.81	120.64	179.56	59.09	29.68
122	Ala	8.80	122.32	178.97	55.20	18.39
123	Glu	8.60	116.84	178.81	60.04	28.74
124	Lys	7.80	122.65	178.07	60.03	32.46
125	Leu	7.95	118.98	179.52	58.20	43.02
126	Leu	8.40	116.44	179.61	57.84	42.62
127	Thr	8.17	115.21	175.12	66.32	68.78
128	Glu	8.25	119.31	178.64	58.99	30.58
129	Lys	8.28	113.63	177.17	55.45	34.30
130	Gly	7.15	106.70	170.97	43.92	-
131	Lys	8.48	116.05	175.34	54.01	35.61
132	His	9.03	121.19	177.79	60.55	30.64
133	Gly	8.65	116.16	175.35	45.18	-
134	Ser	8.75	120.67	172.95	60.27	63.28
135	Phe	7.53	116.46	171.08	55.05	44.47
136	Leu	9.04	115.23	174.47	54.11	44.83
137	Val	9.35	120.05	173.00	60.66	34.52
138	Arg	9.45	123.85	174.41	53.06	34.17

139	Glu	8.67	120.73	176.35	56.97	30.70
140	Ser	7.88	115.41	176.51	57.53	63.45
141	Gln	-	-	-	57.24	28.30
141	Gln	-	-	-	-	-
142	Ser	9.05	116.33	175.01	60.40	63.36
143	His	8.21	123.26	170.22	53.16	28.79
144	Pro	-	-	177.91	64.25	31.45
145	Gly	8.98	115.88	173.09	45.22	-
146	Asp	7.96	119.17	176.10	53.44	42.05
147	Phe	9.14	117.62	173.28	57.02	43.61
148	Val	9.54	119.37	174.25	61.37	36.12
149	Leu	9.40	129.85	174.85	53.28	44.32
150	Ser	9.13	125.76	172.94	59.07	64.59
151	Val	8.82	122.82	174.82	59.80	36.01
152	Arg	9.04	128.55	175.07	55.05	31.54
153	Thr	8.80	122.08	174.64	60.96	70.29
154	Gly	9.74	112.72	173.07	45.02	-
155	Asp	8.39	120.65	176.11	54.00	42.26
156	Asp	8.54	120.60	176.46	54.91	41.07
157	Lys	8.34	120.30	177.10	56.30	32.66
158	Gly	8.27	109.26	174.33	45.53	-
159	Glu	8.46	120.57	176.81	56.79	30.21
160	Ser	8.38	115.75	174.68	58.34	63.85
161	Asn	-	-	-	-	-
161	Asn	-	-	-	53.57	38.72
162	Asp	8.24	119.24	176.81	54.42	40.99
163	Gly	8.28	108.59	174.26	45.65	-
164	Lys	7.98	120.39	176.61	56.04	33.20

165	Ser	8.30	117.04	174.16	58.32	64.07
166	Lys	8.65	123.00	174.98	56.08	34.29
167	Val	8.13	121.45	174.33	60.54	34.75
168	Thr	8.80	124.49	171.64	61.54	70.68
169	His	8.74	126.37	174.77	54.79	31.61
170	Val	9.57	127.70	174.88	61.67	33.60
171	Met	8.72	127.22	175.43	56.06	31.54
172	Ile	9.03	125.63	175.52	59.90	39.59
173	Arg	8.79	128.00	175.55	55.48	31.56
174	Cys	8.82	125.13	174.56	57.10	28.15
175	Gln	8.66	127.79	175.05	55.01	31.50
175	Gln	-	-	-	-	-
176	Glu	9.34	125.20	174.93	57.36	27.54
177	Leu	8.35	109.06	175.25	56.60	39.38
178	Lys	7.53	117.91	175.92	53.93	36.16
179	Tyr	9.31	120.09	175.84	56.96	41.88
180	Asp	9.16	119.52	174.02	54.26	45.75
181	Val	8.78	114.09	177.25	60.30	31.81
182	Gly	9.40	113.08	174.90	45.76	-
183	Gly	8.20	107.88	174.39	44.91	-
184	Gly	8.43	108.33	173.69	45.15	-
185	Glu	7.99	120.73	174.28	56.43	30.28
186	Arg	7.88	119.64	176.01	54.37	32.86
187	Phe	9.42	119.88	176.18	57.57	43.69
188	Asp	10.00	120.72	175.29	56.08	41.75
189	Ser	7.57	108.60	173.86	56.42	66.55
190	Leu	8.73	122.12	178.20	57.11	42.15
191	Thr	7.94	114.76	175.29	68.73	67.16

192	Asp	7.77	121.04	177.93	57.28	40.61
193	Leu	6.96	122.53	177.22	58.95	42.14
194	Val	7.98	120.13	177.23	67.20	31.47
195	Glu	8.26	115.82	179.69	58.99	29.05
196	His	7.71	118.68	178.81	60.80	31.11
197	Tyr	7.71	117.28	176.47	61.20	36.90
198	Lys	7.79	119.42	178.23	58.94	33.09
199	Lys	6.94	116.13	175.42	56.68	34.26
200	Asn	-	-	-	-	-
200	Asn	7.61	118.22	170.78	50.57	40.66
201	Pro	-	-	176.62	63.01	32.48
202	Met	8.67	121.36	174.67	55.29	34.91
202	Met	-	-	174.65	-	-
203	Val	8.24	123.08	176.46	61.98	32.61
204	Glu	8.69	126.17	177.38	56.66	30.62
205	Thr	8.29	115.72	175.09	64.48	69.46
206	Leu	8.30	120.32	178.05	55.83	-
207	Gly	8.26	107.83	174.30	45.52	-
208	Thr	7.79	115.19	172.88	62.00	69.86
209	Val	8.36	122.99	175.66	61.76	33.09
210	Leu	8.88	128.34	174.48	53.75	43.39
211	Gln	7.92	122.37	174.26	54.71	30.91
212	Leu	8.48	124.44	176.65	54.58	40.44
213	Lys	9.12	122.16	175.88	57.11	33.21
214	Gln	7.52	115.62	171.84	53.40	29.09
214	Gln	-	-	-	-	-
215	Pro	-	-	176.56	61.62	32.89
216	Leu	8.59	125.72	175.47	53.87	41.68

217	Asn	-	-	-	-	-
217	Asn	8.53	125.09	176.90	52.92	39.52
218	Thr	8.51	114.19	174.35	61.27	69.55
219	Thr	8.25	115.62	174.54	61.74	69.50
220	Arg	8.23	124.11	175.76	56.12	30.88
221	Ile	8.17	122.09	175.54	61.05	38.94
222	Asn	8.47	122.68	174.65	53.11	38.91
223	Leu	-	-	-	-	-
223	Leu	8.24	123.09	176.99	55.21	42.35
224	Ile	8.05	120.85	175.68	61.13	38.69
225	Asn	-	-	-	-	-
225	Asn	8.37	122.10	174.54	53.17	39.20
226	Glu	8.19	121.59	174.91	56.61	30.61
227	Phe	7.68	125.09	180.26	58.93	40.34

nSH2-Shp2:ITIM1s

!		**H**	**N**	**C**	**CA**	**CB**
-2	Gly	-	-	174.30	45.49	-
-1	Arg	8.28	120.98	-	56.23	30.88
-1	Arg	-	-	176.20	-	-
0	Ala	8.38	124.79	177.85	52.77	19.27
1	Met	8.36	119.23	179.40	55.85	32.56
2	Thr	7.87	112.78	175.74	61.79	69.53
4	Arg	-	-	-	55.55	31.36
5	Arg	8.34	119.16	-	58.07	29.97
5	Arg	-	-	175.01	-	-
6	Trp	10.62	-	-	-	-
6	Trp	6.32	109.75	175.20	54.69	29.70

7	Phe	7.76	123.84	173.35	57.14	39.69
8	His	8.66	129.61	173.09	52.32	33.54
9	Pro	-	-	178.00	63.41	32.79
10	Asn	-	-	-	-	-
10	Asn	8.40	113.14	174.66	53.78	39.33
11	Ile	6.78	113.18	175.46	60.37	41.52
12	Thr	8.39	112.81	175.84	60.25	71.51
13	Gly	9.49	109.70	175.60	47.84	-
14	Val	7.93	120.60	178.74	66.18	32.06
15	Glu	7.49	120.82	179.60	59.19	29.98
16	Ala	8.63	121.19	178.82	55.13	19.13
17	Glu	8.20	116.69	177.24	60.93	28.39
18	Asn	7.66	116.07	178.65	56.99	38.82
18	Asn	-	-	-	-	-
19	Leu	8.48	121.75	179.01	58.38	42.68
20	Leu	8.28	118.66	179.45	58.10	42.59
21	Leu	8.34	116.98	177.81	56.81	41.28
22	Thr	7.72	109.24	176.24	63.60	70.45
23	Arg	8.43	119.15	176.57	55.71	32.04
24	Gly	7.02	106.79	171.08	43.89	-
25	Val	8.93	109.81	176.27	59.10	36.05
26	Asp	8.95	123.07	176.99	57.83	38.95
27	Gly	8.66	113.03	175.95	44.84	-
28	Ser	9.36	122.48	172.78	60.70	64.35
29	Phe	8.08	116.55	170.54	55.73	44.42
30	Leu	9.18	115.08	174.56	54.15	45.05
31	Ala	9.65	121.99	174.31	50.07	23.71
32	Arg	9.19	117.41	-	54.11	30.18

33	Pro	-	-	-	63.30	32.66
34	Ser	-	-	-	-	63.97
34	Ser	7.84	117.52	-	57.82	-
36	Ser	8.45	114.77	-	-	-
36	Ser	-	-	175.25	60.67	62.61
37	Asn	-	-	-	-	-
37	Asn	7.94	121.47	170.67	50.72	39.55
38	Pro	-	-	177.54	64.34	31.59
39	Gly	8.87	115.18	174.06	45.40	-
40	Asp	8.07	119.84	176.61	53.56	42.04
41	Phe	8.91	117.50	173.37	57.14	44.25
42	Thr	9.59	117.72	173.92	62.12	73.15
43	Leu	-	-	-	53.25	45.15
43	Leu	9.50	126.47	-	-	-
44	Ser	9.27	122.38	-	58.91	65.02
44	Ser	-	-	172.29	-	-
45	Val	8.93	123.83	173.81	60.19	35.74
46	Arg	9.17	127.29	175.40	54.82	31.81
47	Arg	9.45	131.28	174.86	55.32	32.52
48	Asn	8.38	-	-	-	-
48	Asn	9.30	125.49	175.43	54.41	37.49
49	Gly	9.57	106.06	172.08	46.06	-
50	Ala	7.97	122.92	174.84	50.46	22.25
51	Val	8.62	119.82	174.75	61.08	33.87
52	Thr	8.83	124.07	170.99	60.34	70.75
53	His	8.28	125.53	174.43	54.54	31.86
54	Ile	-	-	175.27	60.28	42.13
54	Ile	9.77	124.84	-	-	-

55	Lys	8.77	127.96	175.67	58.49	33.85
55	Lys	-	-	-	-	-
56	Ile	-	-	-	60.22	-
56	Ile	8.83	122.49	-	-	39.59
57	Gln	8.73	129.17	174.54	55.70	28.46
58	Asn	-	-	176.17	-	-
58	Asn	8.46	122.98	-	50.72	40.19
59	Thr	8.08	114.96	-	61.45	69.74
59	Thr	-	-	175.22	-	-
60	Gly	8.11	113.34	173.80	45.87	-
61	Asp	7.57	117.92	174.86	54.68	42.63
62	Tyr	7.18	114.19	172.53	56.74	40.34
63	Tyr	9.62	118.38	175.70	56.26	41.90
64	Asp	9.08	120.88	174.61	53.70	43.16
67	Gly	-	-	174.19	44.31	-
68	Gly	-	-	174.29	44.89	-
68	Gly	8.43	107.17	-	-	-
69	Glu	8.16	122.08	174.03	56.52	30.15
70	Lys	7.67	118.60	176.36	54.90	35.17
71	Phe	9.13	117.92	175.87	57.13	44.31
72	Ala	9.40	123.29	178.16	54.51	20.02
73	Thr	-	-	173.66	58.84	73.49
73	Thr	7.17	101.60	-	-	-
74	Leu	7.89	122.23	178.18	56.99	41.23
75	Ala	8.24	118.70	-	55.50	18.14
76	Glu	7.77	117.04	177.84	59.41	31.09
77	Leu	6.91	122.60	177.30	58.63	41.80
78	Val	7.88	118.76	177.17	67.36	31.63

79	Gln	7.99	117.32	177.72	59.09	28.16
80	Tyr	7.80	118.31	179.63	62.18	38.99
81	Tyr	7.74	117.28	177.74	62.91	37.99
82	Met	8.26	119.96	177.48	59.20	34.11
83	Glu	7.42	114.93	176.06	56.87	30.62
84	His	7.42	118.12	173.92	54.94	27.03
86	Gly	-	-	174.26	45.50	-
87	Gln	7.94	116.17	175.29	56.32	29.71
88	Leu	-	-	-	-	41.66
88	Leu	8.38	123.58	173.96	54.35	-
89	Lys	8.30	123.94	175.24	54.61	-
89	Lys	-	-	-	-	36.14
90	Glu	-	-	178.51	56.75	-
91	Lys	8.85	123.25	-	-	-
91	Lys	-	-	-	59.73	32.26
92	Asn	-	-	-	-	-
92	Asn	8.09	115.24	176.15	53.20	37.55
93	Gly	8.11	108.76	174.02	44.93	-
94	Asp	8.17	122.26	175.65	54.90	40.83
95	Val	8.31	121.08	176.27	62.11	33.66
96	Ile	8.43	127.48	173.90	60.80	39.03
97	Glu	8.35	124.75	-	55.01	32.00
97	Glu	-	-	175.21	-	-
98	Leu	8.67	125.14	174.67	53.30	40.00
99	Lys	8.05	120.26	176.33	58.47	34.63
100	Tyr	7.79	116.54	172.34	-	-
100	Tyr	-	-	-	52.72	40.66
101	Pro	-	-	176.29	62.64	32.15

102	Leu	8.74	128.54	175.66	54.41	41.97
103	Asn	8.57	125.69	175.20	53.51	38.95
104	Cys	8.63	121.81	174.63	59.15	27.32
105	Leu	8.67	127.92	176.91	55.16	42.02
106	Ile	7.85	120.65	175.49	61.15	38.85
107	Asn	8.42	122.17	174.55	53.24	39.21
107	Asn	-	-	-	-	-
108	Glu	8.15	121.31	174.82	56.67	30.69
109	Phe	7.66	125.02	180.20	58.93	40.41

TanSH2-Shp2

!		**H**	**N**	**C**	**CA**	**CB**
-11	Ser	-	-	-	58.47	63.58
-10	Gly	8.35	111.21	-	45.02	-
-9	Val	7.82	119.25	-	61.59	32.19
-8	Asp	8.33	124.33	-	53.79	40.66
-7	Leu	8.29	124.26	-	55.14	41.25
-6	Gly	8.47	109.49	-	45.35	-
-5	Thr	7.91	113.72	-	62.15	68.96
-4	Glu	8.47	123.13	-	57.23	29.08
-3	Asn	8.28	119.36	-	53.49	38.36
-2	Leu	8.05	122.57	-	55.98	41.08
-1	Tyr	8.00	119.95	-	58.72	37.77
0	Phe	7.91	119.89	-	58.77	38.64
1	Gln	8.13	119.59	-	57.15	28.14
2	Ser	7.91	115.00	-	58.93	63.31
5	Arg	-	-	-	57.43	29.14
6	Trp	6.42	110.81	-	54.82	29.44

7	Phe	7.80	124.57	-	57.11	39.24
8	His	8.60	130.06	-	51.91	33.09
9	Pro	-	-	-	63.09	31.95
10	Asn	8.41	113.79	-	53.24	39.04
11	Ile	6.72	114.22	-	60.18	40.62
12	Thr	8.30	113.76	-	60.00	71.38
13	Gly	9.49	110.52	-	47.33	-
14	Val	7.76	120.89	-	65.30	30.89
15	Glu	7.36	121.34	-	58.67	29.23
16	Ala	8.64	122.08	-	54.73	18.17
17	Glu	8.26	117.32	-	59.78	27.98
18	Asn	7.71	116.54	-	56.51	38.17
19	Leu	8.45	122.44	-	58.00	41.72
20	Leu	8.23	119.09	-	57.59	41.51
21	Leu	8.26	117.27	-	56.45	40.36
22	Thr	7.68	109.81	-	63.10	69.98
23	Arg	8.38	119.67	-	55.31	31.00
24	Gly	6.95	107.08	-	43.33	-
25	Val	8.87	110.31	-	58.74	35.27
26	Asp	8.86	123.35	-	57.35	38.58
27	Gly	8.62	113.44	-	44.42	-
28	Ser	9.27	122.98	-	60.16	64.02
29	Phe	7.97	116.95	-	55.58	43.67
30	Leu	9.12	115.43	-	53.89	44.15
31	Ala	9.53	122.36	-	49.77	23.06
32	Arg	8.93	117.16	-	53.35	29.35
33	Pro	-	-	-	62.71	31.50
34	Ser	7.69	117.34	-	56.93	63.71

35	Lys	-	-	-	55.93	31.18
36	Ser	8.74	116.18	-	60.16	62.92
37	Asn	7.73	121.59	-	50.39	39.16
38	Pro	-	-	-	64.01	30.82
39	Gly	8.75	115.78	-	44.93	-
40	Asp	7.92	119.82	-	53.21	41.42
41	Phe	8.81	117.39	-	56.95	43.75
42	Thr	9.56	118.20	-	61.79	71.61
43	Leu	9.43	127.74	-	53.02	43.63
44	Ser	9.22	124.41	-	58.59	64.35
45	Val	8.88	123.90	-	59.73	34.99
46	Arg	9.07	127.61	-	54.20	31.06
47	Arg	9.38	131.37	-	55.18	31.42
48	Asn	9.22	125.80	-	54.15	37.07
49	Gly	9.52	106.47	-	45.60	-
50	Ala	7.89	123.15	-	50.08	21.65
51	Val	8.60	120.36	-	60.67	33.30
52	Thr	8.82	124.84	-	59.81	70.67
53	His	8.34	125.68	-	54.41	31.82
54	Ile	9.72	127.13	-	60.05	39.71
55	Lys	8.49	128.47	-	57.43	32.35
56	Ile	8.74	123.37	-	59.88	39.15
57	Gln	8.65	128.93	-	55.14	28.23
58	Asn	8.42	123.29	-	50.51	39.58
59	Thr	8.11	116.00	-	61.08	69.31
60	Gly	8.03	113.83	-	45.56	-
61	Asp	7.55	118.45	-	54.49	41.85
62	Tyr	7.10	114.54	-	56.31	40.16

63	Tyr	9.53	118.85	-	55.90	41.21
64	Asp	9.00	121.29	-	53.21	43.14
65	Leu	7.94	119.16	-	52.98	39.98
66	Tyr	8.49	119.34	-	60.10	34.68
67	Gly	7.37	107.33	-	44.18	-
68	Gly	8.32	107.79	-	44.36	-
69	Glu	8.18	122.73	-	56.13	29.24
70	Lys	7.62	119.04	-	54.52	34.09
71	Phe	9.07	118.45	-	56.83	43.56
72	Ala	9.34	123.63	-	54.11	19.39
73	Thr	7.09	105.75	-	-	-
73	Thr	-	-	-	58.51	72.94
74	Leu	7.87	122.54	-	56.55	40.15
75	Ala	8.34	118.99	-	54.89	17.53
76	Glu	7.69	117.32	-	58.79	30.20
77	Leu	6.84	122.77	-	58.29	40.76
78	Val	7.82	119.25	-	67.09	30.78
79	Gln	7.94	117.80	-	58.65	27.26
80	Tyr	7.72	118.84	-	61.66	38.04
81	Tyr	7.65	117.78	-	62.32	37.29
82	Met	8.22	120.52	-	58.80	33.07
83	Glu	7.37	115.40	-	56.36	30.00
84	His	7.36	118.62	-	54.43	26.27
86	Gly	-	-	-	44.97	-
87	Gln	7.85	116.49	-	55.87	28.87
88	Leu	8.40	124.63	-	54.27	40.82
89	Lys	7.99	124.20	-	54.04	36.02
90	Glu	8.66	120.26	-	54.97	30.58

91	Lys	8.96	123.44	-	58.67	31.05
92	Asn	7.93	115.46	-	52.81	37.18
93	Gly	8.03	109.14	-	44.47	-
94	Asp	8.00	122.46	-	54.45	39.96
95	Val	8.26	122.19	-	62.80	32.12
96	Ile	8.42	128.11	-	59.14	38.34
97	Glu	8.12	124.72	-	54.58	30.97
98	Leu	8.55	125.26	-	52.63	39.01
99	Lys	8.00	121.23	-	58.06	34.02
100	Tyr	7.72	116.92	-	52.57	39.98
101	Pro	-	-	-	62.19	31.34
102	Leu	8.61	128.95	-	54.13	41.19
103	Asn	8.52	126.02	-	53.13	38.67
104	Cys	-	-	-	58.77	27.10
104	Cys	8.63	122.64	-	-	-
105	Ala	8.75	131.07	-	51.74	18.74
106	Asp	7.84	122.66	-	51.67	41.05
107	Pro	-	-	-	63.36	30.93
108	Thr	7.64	110.86	-	63.72	68.48
109	Ser	7.75	115.74	-	58.13	63.42
110	Glu	7.48	121.90	-	54.61	27.35
111	Arg	-	-	-	57.53	29.36
112	Trp	6.24	109.34	-	53.30	30.45
113	Phe	7.53	124.39	-	58.72	38.77
114	His	-	-	-	55.13	31.92
114	His	8.82	126.97	-	-	-
115	Gly	5.78	105.20	-	47.41	-
116	His	8.59	124.42	-	55.87	28.05

117	Leu	7.57	128.36	-	54.42	43.96
118	Ser	8.67	122.36	-	57.06	65.40
119	Gly	9.75	111.82	-	47.23	-
120	Lys	8.15	121.13	-	58.66	31.48
121	Glu	7.72	121.02	-	58.59	29.02
122	Ala	8.71	122.60	-	54.65	-
122	Ala	-	-	-	-	17.87
123	Glu	8.49	117.21	-	59.69	28.26
124	Lys	7.73	123.00	-	59.51	31.71
125	Leu	7.96	119.66	-	57.65	41.82
126	Leu	8.31	116.88	-	57.39	41.67
127	Thr	8.10	115.75	-	65.82	68.37
128	Glu	8.24	119.88	-	58.66	29.78
129	Lys	8.19	114.15	-	55.19	33.50
130	Gly	7.08	107.26	-	44.03	-
131	Lys	8.43	116.71	-	53.58	34.54
132	His	8.94	121.56	-	59.99	30.08
133	Gly	8.59	116.54	-	44.67	-
134	Ser	8.67	121.30	-	60.02	62.91
135	Phe	7.72	117.54	-	55.02	43.83
136	Leu	8.95	115.25	-	53.70	44.31
137	Val	9.50	120.82	-	60.27	33.89
138	Arg	9.33	124.10	-	52.34	33.40
139	Glu	8.68	121.30	-	56.51	29.73
140	Ser	7.82	116.20	-	57.23	63.25
141	Gln	-	-	-	56.81	27.47
142	Ser	8.93	116.67	-	60.06	63.24
143	His	8.12	123.65	-	52.96	28.36

144	Pro	-	-	-	63.86	30.67
145	Gly	8.92	116.21	-	44.74	-
146	Asp	7.90	119.58	-	53.14	41.43
147	Phe	9.10	117.96	-	56.90	43.36
148	Val	9.50	120.18	-	61.19	35.01
149	Leu	9.31	130.19	-	53.01	43.39
150	Ser	9.13	125.71	-	58.64	64.36
151	Val	8.76	123.04	-	59.28	35.30
152	Arg	8.96	129.12	-	54.48	30.79
153	Thr	8.75	122.33	-	60.33	70.03
154	Gly	9.87	113.45	-	44.55	-
155	Asp	8.33	121.11	-	53.68	41.84
156	Asp	8.48	121.12	-	54.59	40.61
157	Lys	8.29	120.81	-	55.94	31.89
158	Gly	8.21	109.76	-	45.11	-
159	Glu	8.40	121.12	-	56.31	29.52
160	Ser	8.31	116.23	-	58.07	63.74
161	Asn	-	-	-	53.43	38.35
162	Asp	8.17	119.72	-	54.23	40.51
163	Gly	8.20	109.10	-	45.24	-
164	Lys	7.92	120.92	-	55.81	32.44
165	Ser	8.23	117.61	-	58.02	64.01
166	Lys	8.61	123.46	-	55.78	33.54
167	Val	8.06	121.79	-	60.08	34.06
168	Thr	8.69	124.96	-	61.29	70.05
169	His	8.66	126.63	-	54.35	31.32
170	Val	9.62	128.01	-	61.23	33.10
171	Met	8.65	127.54	-	55.75	30.95

172	Ile	8.96	126.08	-	59.61	38.74
173	Arg	8.70	128.25	-	54.96	30.69
174	Cys	8.74	125.46	-	56.87	28.15
175	Gln	8.61	128.34	-	54.65	30.74
176	Glu	9.27	125.65	-	57.04	26.86
177	Leu	8.29	109.64	-	56.01	38.59
178	Lys	7.49	118.69	-	53.64	35.52
179	Tyr	9.26	120.45	-	56.60	41.38
180	Asp	9.07	119.96	-	53.92	45.30
181	Val	8.72	114.48	-	60.01	31.09
182	Gly	9.30	113.39	-	45.41	-
183	Gly	8.10	108.44	-	44.51	-
184	Gly	8.37	108.80	-	44.65	-
185	Glu	7.92	121.23	-	55.86	29.52
186	Arg	7.82	120.07	-	54.07	31.87
187	Phe	9.35	120.31	-	57.23	43.25
188	Asp	9.92	121.08	-	55.79	41.23
189	Ser	7.49	108.89	-	56.22	66.21
190	Leu	8.69	122.49	-	56.78	41.25
191	Thr	7.98	115.38	-	-	-
191	Thr	-	-	-	66.35	68.14
192	Asp	7.69	121.40	-	57.04	40.15
193	Leu	-	-	-	58.58	41.24
193	Leu	6.89	123.04	-	-	-
194	Val	7.99	120.70	-	66.74	30.72
195	Glu	-	-	-	58.62	28.56
195	Glu	8.18	116.02	-	-	-
196	His	-	-	-	60.29	30.56

196	His	7.62	119.30	-	-	-
197	Tyr	-	-	-	60.81	36.53
197	Tyr	7.69	117.81	-	-	-
198	Lys	7.76	119.99	-	-	32.28
198	Lys	-	-	-	58.60	-
199	Lys	6.84	116.91	-	56.22	33.51
200	Asn	7.57	118.65	-	50.20	40.17
201	Pro	-	-	-	62.75	31.68
202	Met	8.61	121.84	-	54.93	34.22
203	Val	8.18	123.39	-	61.50	31.88
204	Glu	8.61	126.60	-	56.07	29.93
205	Thr	8.21	116.27	-	64.32	68.91
206	Leu	8.24	120.75	-	55.52	40.22
207	Gly	8.19	108.22	-	45.05	-
208	Thr	7.75	115.88	-	61.63	69.33
209	Val	8.30	123.61	-	61.37	32.30
210	Leu	8.82	128.78	-	53.13	42.06
211	Gln	7.82	123.06	-	54.39	30.12
212	Leu	8.40	124.63	-	54.06	39.25
213	Lys	9.04	122.38	-	57.07	32.59
214	Gln	7.47	115.62	-	53.04	28.75
215	Pro	-	-	-	61.21	31.63
216	Leu	8.07	129.34	-	57.13	42.14