The roles of the co-chaperone SGTA/Sgt2, the BAG6 complex and E3 Ubiquitin Ligase RNF126 in cytosolic quality control

Krysztofinska, Ewelina Maria

Awarding institution:
King's College London

The roles of the co-chaperone SGTA/Sgt2, the BAG6 complex and E3 Ubiquitin Ligase RNF126 in cytosolic quality control

by

Ewelina Maria Krysztofinska

A thesis submitted in candidature for the degree of Doctor of Philosophy

King's College London

2017

Supervisor: Dr Rivka Isaacson

Abstract

Eukaryotic cells rely on quality control mechanisms to sustain protein homeostasis by regulating protein folding, targeting and degradation. These mechanisms involve the recognition of exposed hydrophobic regions of membrane proteins that have mislocalized to the cytosol to prevent them from misfolding and aggregation. In mammals, the heterotrimeric BAG6 complex, composed of three proteins, BAG6, UBL4A and TRC35, work closely with the co-chaperone SGTA to triage hydrophobic protein clients ensuring their delivery to the ER or to the proteasomal degradation pathway. Recently, RNF126 has been identified as the BAG6-dependent E3 ligase, which ubiquitinates lysine residues on BAG6 associated substrates. However, the decision-making process in classifying hydrophobic protein clients for degradation or rescue are unclear. Therefore, the characterisation of proteins involved in triage system is critical to understanding the mechanisms of protein sorting. The work presented here provides insights into functions of the co-chaperone SGTA and the E3 ubiquitin ligase, RNF126, within the BAG6 quality control module. It also focuses on the additional role of the co-chaperone SGTA and its yeast homologue Sgt2, which is the interaction with molecular chaperones such as Hsp70 (Ssa1 in yeast) and Hsp90 (Hsc82 in yeast) through the central tetratricopeptide repeat (TPR) domain. The solution structures of the N-terminal dimerisation domain of SGTA and N-terminal zinc finger motif of RNF126 E3 ligase are presented in Chapter 3 and 4 respectively. The interactions of both RNF126 and SGTA with the UBL domains of BAG6 and UBL4A are also characterised. This includes the structural models of complexes of RNF126_NZF and SGTA_NT with UBLs using NMR spectroscopy and HADDOCK (Chapters 3 and 4). Chapter 5 contains the x-ray structures of Sgt2_TPR and its complex with the extreme C-terminal of Ssa1. In addition, it shows that Sgt2_TPR interacts with C-terminal fragments of Ssa1, Hsc82 and Ybr137wp (a protein whose function is yet to be elucidated) in the similar binding mode and with comparable binding affinities. The binding studies were performed using biophysical methods, such as isothermal titration calorimetry and microscale thermophoresis. Together, this work aims to extend our understanding of the quality control mechanism in yeast and mammals, by providing molecular details of the components of the SGTA/BAG6 complex quality control module.

Table of Contents

Table of Figures

List of Tables

Acknowledgments

Special thank you to Dr Rivka Isaacson for giving me the chance to work on this project, also for the confidence she has always had in me and great opportunities she has provided me with such as conferences and trainings. I am also grateful for the support throughout my PhD and for proof-reading of my thesis.

Thanks to all members of the Isaacson lab for their practical assistance and suggestions, especially to Dr Santiago Martínez-Lumbreras for help with RNF126 Aria calculations and NOEs assignments and analysis, also to Nicola Evans and Arjun Thapaliya for help with crystallography and Kat Collins for proof reading the chapters of my thesis.

Thanks to Dr Andrew Atkinson and Dr Pete Simpson for their advice and support with NMR experiments and Dr James Murray for advice on crystallography.

Rob, for intense scientific discussions, love and support and cooking during my write-up period.

Finally, thanks to my mum, family and friends, Masooma, Maud, Julia and Cathy for their wonderful friendship and support.

Abbreviations

Abbreviation	Meaning
AIR	Ambiguous Interaction Restraints
ARIA	Ambiguous Restraints Iterative Assignment
ATP	Adenosine triphosphate
BAG6	Bcl2- associated athanogene 6
CAML	Calcium- modulating cyclophilin ligand
CC	Correlation coefficient
COOT	Crystallographic Object-Oriented Toolkit
CSA	Chemical shift anisotropy
CSI	Chemical Shift Index
CSP	Chemical Shift Perturbation
DANO	Anomalous difference
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleid acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
DUB	Deubiquitinase
EDTA	Ethylenediaminetetraacetic acid
EGS	Ethylene glycobis (succinimidylsuccinate)
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation
FID	Free induction decay
FT	Fourier transform
GET	Guided Entry of Tail- anchored proteins
GFP	Green fluorescent protein
HADDOCK	High Ambiguity Driven protein-protein docking
HMQC	Heteronuclear Multi Quantum Coherence
Hsp	Heat shock proteins

HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive nuclei enhancement by polarisation transfer
IPTG	Isopropryl β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
LIC	Ligation independent cloning
MAD	Multi wavelength anomalous diffraction
MLP	Mislocalised proteins
MR	Molecular replacement
MST	Microscale thermophoresis
NCS	Non- crystallographic symmetry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NZF	N-terminal zinc finger
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
ppm	Parts per million
RAMP4	Ribosome associated membrane protein
RCI	Random coil index
RF	Radiofrequency
RING	Really interesting new gene
RMSD	Root mean square deviation
SAD	Single-wavelength anomalous diffraction
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulphate

SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
Sgt2	Small Glutamine - Rich Tetratricopeptide Repeat protein 2
SGTA	Small Glutamine - Rich Tetratricopeptide Repeat protein Alpha
SNARE	Soluble N -ethylmaleimide-sensitive factor activating protein receptor proteins
SRP	Signal recognition particle
TA-proteins	Tail-anchored proteins
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TMD	Transmembrane domain
TPR	Tandem repeat protein
TRC	TMD recognition complex
TROSY	Transverse relaxation optimised spectroscopy
Ub	Ubiquitin
UBD	Ubiquitin-binding domain
UBL	Ubiquitin like domain
UBL4A	Ubiquitin like Protein4A
UPR	Unfolded protein response
WRB	Tryptophan-rich basic protein
ZF	Zinc finger

Chapter 1. Introduction

1.1 Protein quality control in the cytosol and the endoplasmic reticulum

There are several billion protein molecules in human cells which differ in size, assembly, shape, and stability (1, 2). To ensure the correct functioning of these versatile proteins, eukaryotic cells have developed an intricate network of protein quality control pathways. Sustaining protein homeostasis (also known as proteostasis) requires monitoring of protein biosynthesis, by reviewing their folding and localisation in the cell, and degradation of aberrant proteins (3-5). Quality control systems operate in different compartments of the eukaryotic cell, including ribosomes and the endoplasmic reticulum (ER) and at multiple steps of protein biosynthesis. Protein quality checkpoints act during protein translation at the ribosome, and post-translationally, to monitor the nascent polypeptides and avoid the emergence and accumulation of aberrant proteins at this early stage of protein maturation (6-8). Some proteins require a cotranslational folding of their long polypeptides to prevent the formation of intermediate assemblies, which could impede the protein's folding to its native state (9). In eukaryotes, a subset of highly conserved heat shock proteins (Hsp40 and Hsp70 families) can bind to the newly synthesised protein chains and assist their cotranslational folding (10). In human cells, the fully synthesized proteins which do not contain signal sequences for entering the secretory pathway, peroxisomes or mitochondria are released from the ribosomes and their folding can also be supported by the same sets of heat shock proteins in the cytosol (which in yeast are called Ssa1-4) (11, 12). The heat shock chaperones are ubiquitous and can interact with a range of substrate conformers, varying from entirely unfolded to natively folded and even aggregated protein states (13). It has been shown that Hsp70 binding can also induce conformational changes in proteins (14) and support protein disaggregation through one or several ATP-regulated cycles of binding and hydrolysis (15, 16).

Maintaining protein homeostasis is particularly challenging for membrane and secretory proteins because they are imported to the ER for maturation in an unfolded state (17). The long

hydrophobic stretches which characterise integral membrane proteins require protection from the cytosol to avoid the exposure of their hydrophobic residues and subsequent misfolding (6). Furthermore, failure in targeting and mislocalisation of membrane proteins to the cytosol can lead to the formation of aggregates which are toxic to the cell (8, 18, 19). Most secretory membrane proteins are delivered and inserted to the ER membrane cotranslationally and depend on their recognition by the Signal Recognition Particle (SRP) and subsequent targeting to the protein transport Sec61 translocon channel localised at the ER (20-23). The first hydrophobic residues to emerge from a ribosome exit tunnel, either a transmembrane domain (TMD) or a cleavable N-terminal sequence, act as an ER-targeting signal and are recognised and shielded by the 54 kDa subunit of SRP (SRP54) (24-26). This then targets these membrane clients, via the SRP receptor, to the Sec61 translocon for insertion (27-29) The same large multiprotein complex Sec61 is also responsible for the translocation of soluble proteins across the ER membrane (30-32). Nevertheless, inaccessibility or mutations of residues within the central hydrophobic core in the targeting signals severely affect SRP binding (33). Moreover other internal cellular stresses like transcriptional and translational failures, aberrant protein conformational changes, or faulty translocated secretory and mitochondrial precursors (34, 35) can affect the efficiency of SRP-dependent targeting and increase the production of misfolded proteins which can mislocalise to the cytosol. Furthermore, a number of external factors like oxidation, heat, aging, and cancer (36) can result in misfolding of some proteins and increase the danger of protein aggregation (4, 37) which can subsequently lead to development of disease states including neurodegenerative disorders such as Parkinson's disease, Alzheimer's disease and diabetes. Moreover, although the majority of membrane and secretory proteins are targeted and inserted to the ER cotranslationally, a subpopulation inserts posttranslationally via an SRP-independent route and these proteins can also accidently mislocalise to the cytosol (8, 38).

To minimize protein homeostasis perturbations, misfolded polypeptides must be either refolded and re-delivered in the case of mislocalised proteins (MLPs) (39) or, failing correction, they must be targeted for degradation. Central to these quality control mechanisms are networks of

specialised chaperones (40). At the endoplasmic reticulum, terminally misfolded proteins are recognised by ER-resident chaperones and retro-translocated in the ER-associated degradation (ERAD) pathway to the cytosol for degradation. The clearing process of misfolded peptides is constantly performed by a highly conserved set of ERAD components that promote their ubiquitination by specialized E3 ubiquitin ligases, and subsequent digestion by the proteasome (41). The folding status at the ER is also constantly monitored by the Unfolded Protein Response (UPR) to control the accumulation of unfolded and misfolded proteins (42). Degradation of substrates that are misfolded, unfolded, or no longer required by the cell can also be removed by macroautophagy driven lysosomal-mediated hydrolysis (43).

Triage decisions between numerous coexisting quality control pathways, need to be adaptable so they can meet the changing requirements of the cell and allow for correct decisions between chaperone-mediated refolding, stabilisation of misfolded and aggregated proteins, and elimination of damaged polypeptides. The sorting mechanism highly depends on dedicated co-chaperones, E3 ubiquitin ligases and their direct interactions with Hsp70 and Hsp90 chaperones.

1.1.1 TA-proteins

Tail-anchored proteins (TA) belong to the family of membrane proteins whose transmembrane domains (TMD) emerge from the ribosome only after translation terminates and thus their ER-signal sequences are obscured by the ribosome and they cannot be delivered to the ER by the cotranslational mechanism (21, 44-46). Instead, these proteins are inserted into ER membranes posttranslationally in a pathway, called Transmembrane Recognition Complex (TRC) in mammals and Guided Entry of TA proteins (GET) in yeast (46) [described in more detail in section 1.2 and 1.3].

TA proteins constitute approximately 5% of all human membrane proteins, including SNAREs (*N*-ethylmaleimide–sensitive factor attachment protein receptors), ER translocon components, Bcl-2 family members, and signalling proteins involved in fundamental cellular processes, e.g. vesicular traffic, stress-response, electron transfer and apoptosis (47-51). TAs contain their single transmembrane domain within the final ~40 residues of the C terminal, which tethers them to the lipid bilayer such that the majority of the protein is located in the cytoplasm (52). The soluble part of the protein is typically most important for biological function, as in the case of the SNARE proteins, in which the cytoplasmic domain plays a vital role during vesicular transport (53).

In eukaryotes, functionally diverse tail-anchored proteins exist in different organelles including the mitochondria, the peroxisome, and chloroplast (in plants) with the ER containing the largest number of proteins (54). To reach compartments of the secretory pathway, TA proteins are thought to be first delivered to the ER and then forwarded to their final destination by vesicular trafficking (20, 55-57). In contrast, TA proteins destined for the mitochondrial outer membrane and the chloroplast envelope can be targeted directly from the cytosol (46).

Diverse TA proteins are equipped with their own characteristic tail and TMDs also contains the membrane targeting signal (58). Furthermore, the amino acid sequence of the tail can affect both the efficiency with which the protein is inserted into the membrane (47, 59) and its cellular location, e.g. moderate hydrophobicity of these elements shows a predisposition for ER residence (54, 60). Interestingly, in mammals, both TMD hydrophobicity and a tail charge define delivery to individual organelles. For example, positive charge of the tail region seems to be a common property of peroxisomal TA proteins, which differentiates them from mitochondrial and ER tail-anchored proteins. In contrast, the presence of significantly higher TMD hydrophobicity of TAs in endoplasmic reticulum as compared to those destined for mitochondria, indicates that a hydrophobic TMD and a low charge of a tail support ER targeting (61).

The detailed mechanisms of sorting and delivery to peroxisomes and mitochondria are less understood. No components of the mitochondrial TA targeting machinery have yet been identified and it has been proposed that the process can occur via unassisted insertion, with the composition of the mitochondrial outer membrane mediating targeting specificity e.g. for cytochrome b_5 (62-64). However, at the ER, the biogenesis of tail-anchored proteins involves one or more cytosolic factors including HSP70 (65-68).

Despite the importance of TA proteins in numerous cellular processes, their targeting mechanisms and signals are not fully understood. Furthermore it is essential to study TA insertion mechanisms and underlying regulation because mutations in genes encoding tail-anchored proteins have been detected in several inherited disorders (69, 70).

1.2 The BAG6 complex and SGTA in TA targeting

In mammals, TA proteins are targeted and inserted to the ER membrane by a conserved TMD recognition complex (TRC), a pathway in which SGTA (small glutamine-rich tetratricopeptide repeat containing protein alpha) together with the heterotrimeric BAG6 complex, composed of BAG6 (BCL2-associated athanogene 6), TRC35 (transmembrane recognition complex 35) and UBL4A (ubiquitin-like protein 4A) play crucial roles (71-73). TA protein delivery is subject to rigorous quality control to degrade proteins that fail targeting (Figure 1) (74, 75).

It is established that loading of nascent TA proteins onto to the downstream targeting factor TRC40 requires SGTA and the BAG6 complex (72, 76). TRC40 can recognise hydrophobic TA regions (77-79) and promotes their membrane insertion at the ER through components of the transmembrane domain recognition complex, a cognate receptor composed of tryptophan-rich basic protein (WRB) and calcium-modulating cyclophilin ligand (CAML) (80, 81). Absence of any component in the BAG6 complex impairs SGTA-TRC40 bridging and TA protein capture by

TRC40 (73). However, free, newly-synthesised TA proteins are preferentially captured by SGTA which shields their TMD regions from the aqueous cytosol as they are released by the ribosome. SGTA can bind to BAG6 and UBL4A of the Bag6 complex via their UBL (ubiquitin like) domains (71, 82). However, SGTA's interaction with the UBL domain of BAG6 (the N-terminal of BAG6 complex) seem to be dispensable for TRC40 capture because only the C-terminal of BAG6 can form a complex with UBL4A and TRC35 (19, 73). The BAG6 complex acts as a scaffolding binding platform which connects SGTA and TRC40 via the UBL4A and TRC35 subunits (Figure 1) and allows for the transfer of TAs from SGTA to TRC40 (76). The small portion of TA proteins which spontaneously dissociate from SGTA before loading onto TRC40, or those that fail to be delivered to TRC40, can be re-captured by either SGTA or BAG6 and further triaged (73). The conserved ATPase, TRC40 (also called ASNA1) is at the centre of TA targeting in mammals in a process regulated by ATP binding and hydrolysis (66, 79). TRC40 was shown to promote the membrane insertion of several model TA proteins with similarly hydrophobic tail-anchor regions, including RAMP4 and Sec61β (66, 79). Although the mammalian structure of TRC40 has not yet been solved, the structures of yeast (83) and bacterial *Methanothermobacter thermautotrophicus* homologues have been characterised by crystallography (84). Overall the structural and functional analysis of tail-anchored delivery to the ER has been completed in greater detail in yeast as compared to mammals. Although studies of processing the TA proteins by the components of the TRC pathway in mammals have provided considerable mechanistic insights, there are still issues which remain unclear e.g., understanding the function of SGTA in TA protein targeting or details pertaining to degradation. To date there is no structure solved of SGTA with bound TA substrate. Likewise, it is not known if sequences inside the ribosome could promote recruitment of the SGTA and the BAG6 complex.

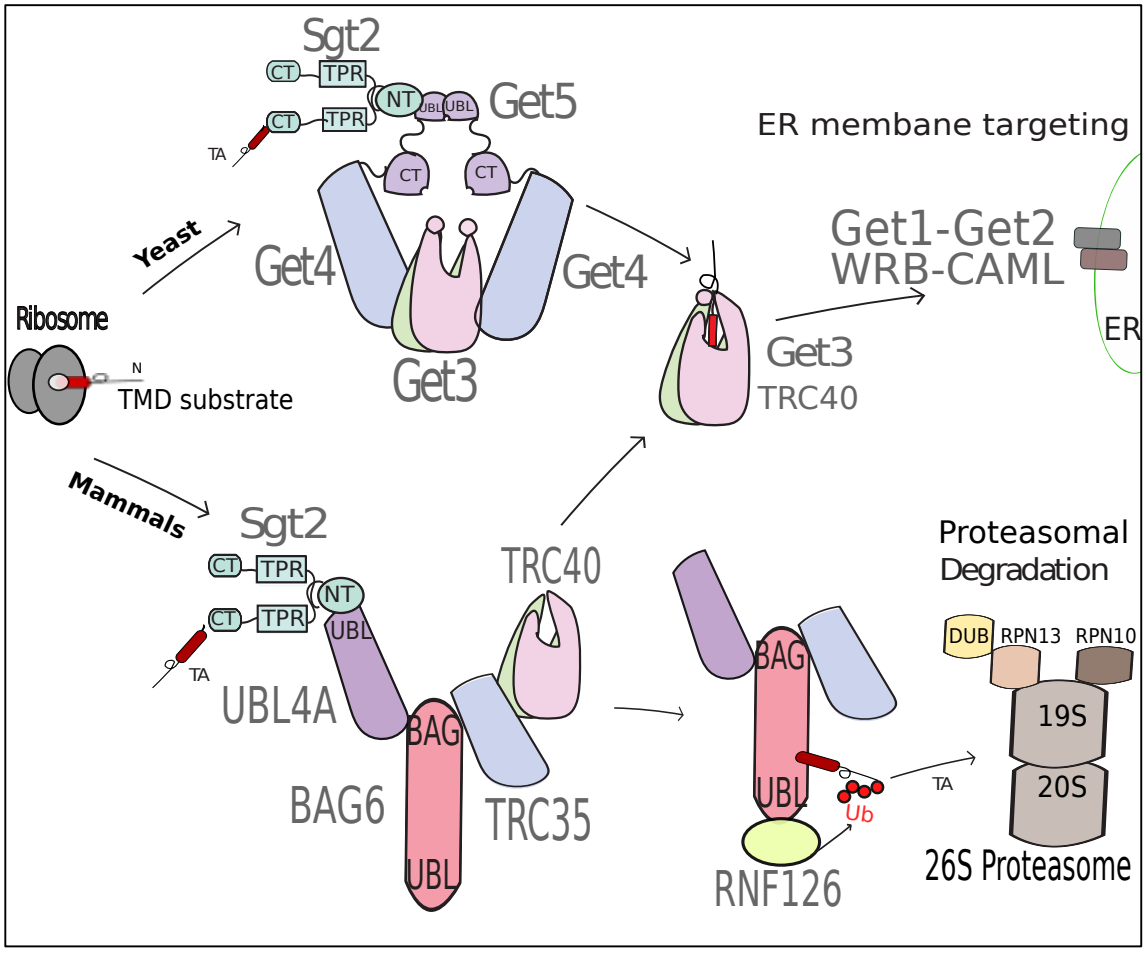

Figure 1: Current models of targeting and quality control of hydrophobic substrates in mammals and yeast. SGTA (Sgt2 in yeast) captures free newly synthesised TA proteins and other hydrophobic substrates (MLPs, ERAD substrates), a process facilitated by its C-terminal domain (58, 85, 86) and hands over the substrates downstream onto components of the TRC pathway (Get pathway in yeast, downstream Get5-Get4-Get3) for delivery to the ER or mammalian quality control pathway for degradation. The BAG6 complex is composed of BAG6, UBL4A and TRC35 proteins, and interacts with SGTA through its UBLs (82). The BAG6 complex connects SGTA and TRC40 via the UBL4A and TRC35 subunits, and allows for a rapid transfer of TA protein from SGTA to TRC40 (73, 79). TA proteins that fail to load onto TRC40, and other hydrophobic substrates which dissociate from SGTA, are re-captured by either SGTA or BAG6 (73). The BAG6 complex associated substrates can be ubiquitinated by the E3 ligase RNF126 and targeted to the proteasome for degradation (75).

1.3 Sgt2 in the GET pathway

TA protein targeting via Guided Entry of Tail- anchored (GET) pathway has been characterised in more detail in *Saccharomyces cerevisiae* than in the mammalian system (**Figure 3**). The structures of many of the components have been solved, and the entire pathway has been

reconstituted from purified recombinant components (58, 78, 87, 88). The tail-anchored targeting mechanism begins with, a TA chaperone, Small, glutamine-rich, tetratricopeptide repeat protein 2 (Sgt2) forming a complex with TA substrates released to the cytosol. BAG6 is found widely across metazoans, but not in fungi and instead the analogous complex contains just two proteins forming a tetramer made of dimers of Get4 and Get5 corresponding to TRC35 and UBL4A, respectively (**Table 1**). Sgt2 escorts TA substrates onto GET components by binding to the UBL domain of Get5 (87, 89, 90). The structure of Sgt2-Get5 complex has been solved by crystallography (87, 89, 90). Get4 and Get5 form a heterodimeric pretargeting complex, with Get4 recruiting the central targeting factor Get3. The role and structures of Get3 at various stages of TA targeting has been well characterised. A 'closed' state formed upon binding of adenosine 5'-triphosphate (ATP) is competent to bind substrates in a process mediated by the Get4-Get5 heterodimeric scaffolding complex (78, 87, 88, 90). Following the release of TA substrates from Sgt2, Get3 binds them within the hydrophobic groove formed by the two helices from the two Get3 subunits in a dimer (78). Several complexes of Get3 with TMDs bound (e.g. Sec22, Pep12) have been crystallised and the solved structures showed that one of the helices can function as a 'dynamic lid' protecting the TMD during transit to ER membrane (78). Conversely, the Get3 'open' state, triggered by the ATP hydrolysis, is followed by TA handover to the heterodimeric Get1-Get2 membrane receptor complex which finally facilitates their insertion into the ER membrane (50, 72, 80, 88). Get3 is recruited to the membrane via its initial interaction with Get2 followed by binding to Get1. This drives TA protein release in an ATPase-dependent reaction. Free Get3 recycles back to the cytosol, prepared for another insertion cycle (91). All Get pathway components are important and there is evidence that deletion of any of the Get genes can result in ER-bound TA proteins misguidedly inserting into mitochondria or aggregating in the cytosol (88).

Thus far, Get3 has been the main focus of high-resolution structural studies, with crystal structures solved in both open and closed nucleotide-bound states (78, 91-94). These structures have provided some molecular insights into the mechanisms that promote Get3 substrate binding and release, and helped to propose models describing how ATP binding and hydrolysis might

affect these steps (83, 95, 96). The 3D structures of Get3 in complex with the cytoplasmic domains of Get1 and Get2 have also been determined and unravelled more details in the next step of TA protein insertion (91). The crystal structures of *S. cerevisiae* Get4 in a complex with an N-terminal fragment of Get5 (87, 89, 97) and a crystal structure of a Get3-Get4-Get5 intermediate complex from S.cerevisiae (98) provided insights into the mechanistic details of pretargeting complexes in the Get pathway. However, the earliest steps involving TA substrate capture by Sgt2 and SGTA are not clear. Sgt2/SGTA facilitates the biogenesis of TA proteins by their high-affinity (with rapid kinetics) interaction with Get5 in the Get4-Get5 complex, ensuring the effective capture of hydrophobic substrates (87, 89, 90). It has also been proposed that SGTA-dependent deubiquitination of TA substrates presents a mechanism for rescue and subsequent handover to downstream targeting factors (86, 99). Although the existence of a pretargeting factor may suggest how substrates are sorted and loaded onto Get3, it is not known how TA proteins are initially captured by Sgt2. The structurally characterized domains of proteins studied in this project and their homologues are shown in **Figure 2**.

Table 1: Analogous protein components of GET pathway in fungi and TRC pathway in mammals.

Fungi	**Mammals**
Sgt2	SGTA
Get5	UBL4A
Get4	TRC35
Get3	TRC40
Get2	CAML receptor (functional homologues)
Get1	WRB receptor
	BAG6

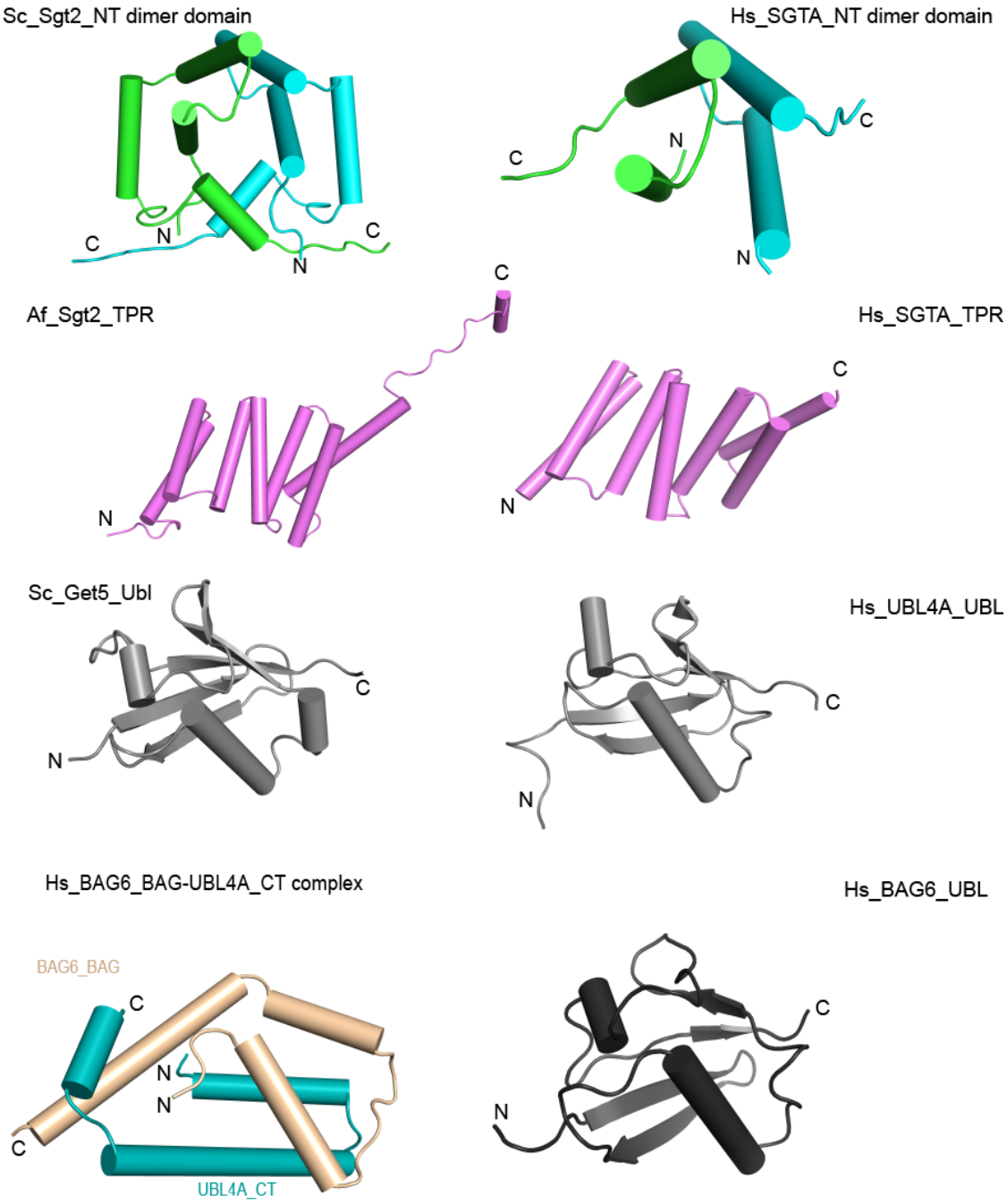

Figure 2: Structures of domains of the proteins studied in this project and their homologues. Sc indicates *Saccharamyces cerevisiae*, Hs indicates *Homo sapiens* and Af indicates *Aspergillus fumigatus*.

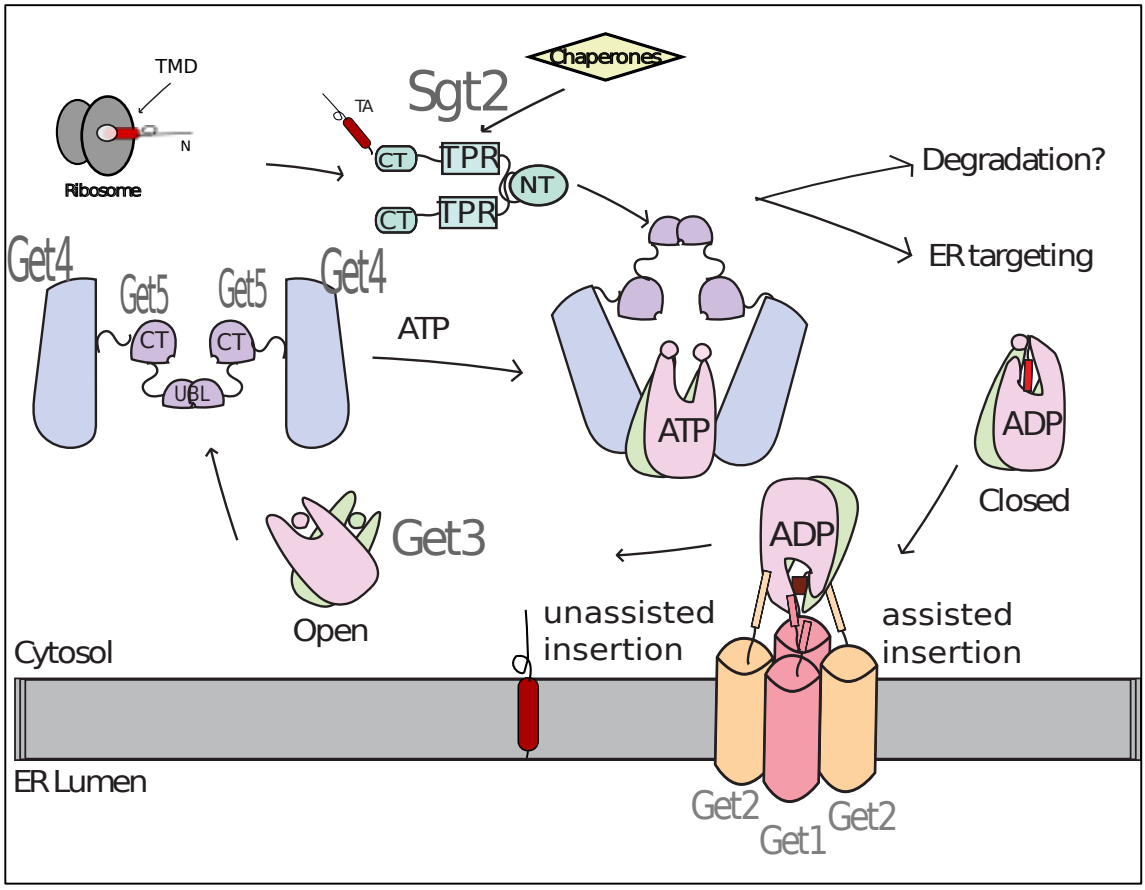

Figure 3: Details of tail-anchored protein targeting to the ER in the GET pathway. In *Saccharomyces cerevisiae,* the chaperone Sgt2, captures and shields ER-destined TA proteins released from the ribosome into the cytosol and hands them over to components of the Get pathway via its interaction with Get5 (88). The pathway is centered around the cytosolic targeting factor Get3, an ATPase that interacts with the heterodimer of Get1-Get2 ER membrane receptor complex to mediate TA protein insertion (88, 91, 93). Loading Get3 with a TA protein requires the heterodimeric Get4-Get5 pretargeting complex (72, 78, 88, 91, 98, 100), which mediates transfer onto dimeric Get3. Get4-Get5 complex recruits Sgt2 in a complex with a TA protein, while Get4 recruits ATP-bound 'open' Get3 (101). Following the ATP hydrolysis, TA substrates are transferred onto 'closed' Get3, which dissociates from Get4 (102-104). No BAG6 homologue exists in the Get pathway and there is no known analogous quality control and degradation pathway in yeast. However, Sgt2's interactions with Hsp70/Hsp90 chaperones (105) via its TPR domain likely provide substrate access to additional quality control networks.

1.4 Degradation of MLPs by the ubiquitin-proteasome system

Failure in targeting of membrane and secretory proteins can result in their mislocalization to the cytosol and accumulation of misfolded proteins that are prone to aggregation poses the danger of developing aggregation related diseases (8, 106). Such aberrant proteins, including MLPs, after recognition by the components of the quality control pathways, must be efficiently eliminated.

The degradation mechanism should be able not only to recognise a broad range of substrates but also to allow for a proficient selection to ensure the correct protein disposal (107, 108). A substantial proportion of intracellular proteins are degraded by the Ubiquitin-Proteasome System (UPS) (109-111). However, extracellular proteins and some cell surface proteins (e.g. cell receptors) are eliminated in a process of endocytosis and degraded within lysosomes (112). In the UPS, damaged or no longer needed, cytosolic proteins are tagged for degradation at the proteasome with a small, 8.5 kDa protein called ubiquitin. The 26S proteasome, a very large multicatalytic protease complex can then recognise ubiquitin, covalently attached to protein clients and degrade these into small peptides. Proteases hydrolyze amino acid chains and break peptide bonds, yielding peptides of about seven to eight amino acids long, which can then be further degraded into shorter amino acid sequences and used in synthesizing new proteins (113). The 26S proteasome in eukaryotes is made of a 20S proteolytic core and a 19S ATP-dependent regulatory cap (114) and is localised in the nucleus and in the cytoplasm (115). The ubiquitin tagging reaction called ubiquitination involves the collaborative action of three proteins: an E1 (ubiquitin-activating enzyme), an E2 (ubiquitin-conjugating enzyme) and an E3 (ubiquitin protein ligase) (116, 117). The client protein is tagged with one or more ubiquitin molecules which can bind to one another, forming polyubiquitin chains. The pattern and distribution of ubiquitination determines the final fate of a protein (118). Notably, ubiquitin molecules can connect mostly via Lys48 or Lys63, with potential to form chains with a variety of linkage types, attached to the protein client which can promote different outcomes for the substrate: Monoubiquitination may facilitate protein recognition, complex formation, or allosteric regulation, whereas polyubiquitination via Lys48 is a signal for proteasomal degradation. In contrast, polyubiquitination of Lys63-linked chains have non degradative roles in cellular signaling, intracellular trafficking and DNA repair (119, 120). Typically, a polyubiquitin chain made of four ubiquitin molecules is sufficient for the complex to be recognized by the 26S proteasome (108, 121, 122).

In the sequential action of the E1/E2/E3 ubiquitination cascade, a free ubiquitin is activated by an E1 ubiquitin-activating enzyme which hydrolyses ATP and adenylylates a ubiquitin molecule. Typically a cysteine at the active site of the E1 enzyme then forms an energy-rich thioester bond with the C-terminal glycine residue of ubiquitin in concert with adenylylation of a second ubiquitin (107, 123). Subsequently the adenylylated ubiquitin is conjugated to the cysteine residue at the active site of an E2 ubiquitin-conjugating enzyme, from where the ubiquitin is transferred onto the lysine side chain of a protein client generating an isopeptide bond, in a process catalysed by E3 ubiquitin ligase, that directly binds to the E2 enzyme (124-127), (**Figure 4**). In multi monoubiquitinated substrates, several Lys residues on the same protein are modified with one ubiquitin and in polyubiquitinated substrates, a lysine residue on a preceding ubiquitin molecule becomes covalently attached to a glycine residue of the next ubiquitin. The formation of polyubiquitin chains, commonly through internal Lys48 residues of ubiquitin, leads to recognition by the proteasome and degradation (108, 121, 122).

Degradation may require additional factors, such as the AAA ATPase p97-Ufd1 (ubiquitin fusion degradation 1)-Npl4 (nuclear protein localization 4) complex, a ubiquitin-selective segregase which functions in the detection of a number of polyubiquitin-tagged proteins and mediates their presentation to the 26S proteasome for degradation (128-131).

Although the whole enzymatic E1/E2/E3 ubiquitination cascade is needed to link ubiquitin molecules to client proteins, the key regulatory enzymes in the process are the E3 ubiquitin–protein ligases, which maintain the specificity of this process by selecting the suitable substrate for ubiquitination via collaboration with several molecular chaperones involved in the recognition and binding of MLPs (4, 132, 133). To determine substrate selectivity, E3 ligases bring together the appropriate E2 with the specific substrate. There are >600 mammalian E3s thought to facilitate recognition of a broad range of substrates, although no biological function, mechanism of action and substrate specificity has been characterised for many of them (127). Nevertheless, E3s are

classified into three main families of E3 ubiquitin ligase classes, each characterised by conserved structural domains and different mechanism by which ubiquitin is transferred from the E2 enzyme to the substrate and the way in which E3 ubiquitin ligases contact the E2 (**Figure 4**). The first type is E3 RING (really interesting new gene) ligases, which typically catalyze direct transfer of ubiquitin from the E2 enzyme to the protein client, simultaneously binding both the E2~Ub cargo and the substrate (134, 135). In some of the E2-E3 pairs, E3 attachment can induce a conformational change in the E2 leading to an increased rate of ubiquitin release (136, 137). The families of HECT (homologous to E6-AP COOH terminus) and the RING-between-RING (RBR) E3 ligases ubiquitinate substrates in a two-step mechanism in which ubiquitin is first transferred from the E2 to a catalytic cysteine of the E3 and then from the E3 to the protein client (138). The HECT domain's active site cysteine can form a thiol-ester bond with ubiquitin (139). There is evidence that some E3 ligases may play a role in regulating chain topology e.g. some of the HECT E3 ligases preferentially form K63-linked chains irrespective of the identity of the E2 (140). Furthermore, a number of E3 ligases have been shown to ubiquitinate misfolded proteins in an Hsp70-dependent manner. Likewise, the mammalian E3 ligase CHIP (carboxy-terminal Hsp70 interacting protein) directly binds to Hsp70 and Hsp90 and can direct some of the Hsp70-bound proteins for degradation (141, 142). However, despite many studies of substrate ubiquitination by the E2-E3 pairing, the catalytic mechanisms by which E3 ligases recognise substrates and regulate ubiquitin transfer remain elusive.

In yeast, some of the ubiquitin ligases can directly interact with misfolded proteins in chaperone-mediated quality control pathways. E3 ligase Ubr1 (ubiquitin protein ligase E3 component N-recognin 1) and Ubr2 can mediate degradation of misfolded proteins in the cytosol while another E3 ligase, San1 (Sir antagonist 1), operates in the nucleus and binds misfolded proteins via its own intrinsically disordered region (143).

1.4.1 Ubiquitin binding domains

The mechanism of proteasomal targeting of polyubiquitinated proteins is not entirely clear. Nevertheless, the ubiquitin modification of target proteins can be specifically recognised by various ubiquitin-binding domains (UBDs) or ubiquitin receptors that contain one or more UBD within their structure (144, 145). UBDs can noncovalently interact with monoubiquitin or polyubiquitin chains and the majority of identified UBDs bind to a hydrophobic area surrounding Ile 44, Leu 8, Ile 44, and Val 70 residues of ubiquitin. The residues flanking the binding surfaces on ubiquitin have also been shown to contribute to the binding specificity of different UBDs (96, 146, 147). Ubiquitin chains attached to proteins destined for proteolysis can be bound directly by a proteasome subunit or by a shuttle factor which links ubiquitinated proteins with the proteasome. The 26S proteasome regulatory subunit Rpn13 and Rpn10 are two intrinsic ubiquitin receptors, present at the 19S regulatory particle of the proteasome that bind mono- and poly-ubiquitin chains as well as UBL domains (148-151). The shuttle factors Rad23 (UV excision repair protein) and Dsk2 (Ubiquitin domain-containing protein), contain both UBDs and Ubiquitin-like (UBL) domains and, through binding both ubiquitin and the proteasome, deliver the polyubiquitinated cargo to the 26S proteasome for degradation (152-155).

The UBL proteins which can bind substrates in a similar manner to ubiquitin are often referred to as type I UBLs (156). Other UBLs, present as functional domains in larger proteins are termed type II UBL domains (157). Many type II UBLs interact with a broad range of UBDs whereas type I UBLs seem to have more specific binding partners. Some of the type II UBLs can interact with ubiquitin binding domains (UBDs) such as UBA (Ubiquitin Associated), UIM (Ubiquitin Interacting Motif), and CUE (Coupling of Ubiquitin to ER degradation) or PAZ (polyubiquitin-associated ZINC FINGER), (144, 158). Ubiquitin, UBLs and UBDs, all have role in the ubiquitin proteasome system and ERAD pathway in fundamental protein-protein interactions (154, 159). The BAG6 complex includes two type II UBLs, one in the UBL4A subunit which does not seem to bind canonical UBDs but does interact with SGTA. The second,

at the N-terminal of BAG6 itself, can bind canonical UBDs but also has the ability to interact with the unusual UBD of SGTA (82, 160).

Work focusing on UBD–ubiquitin interactions, has shown that the majority of them and especially those with monoubiquitin are typically on the low to medium affinity scale. The range of UBD–ubiquitin affinities, varies between K_d values of 10–500 μM (144, 161).

1.4.2 Deubiquitinases

Coupled to the UPS cycle are deubiquitinases (DUBs), proteases, which exist in cells in large numbers and can reverse protein ubiquitination (119). The E3 ligases associate with the client protein for a period of time necessary for the attachment of one or more ubiquitin. Following the ligase dissociation, there is a competition between Ub cleavage by DUBs and ligase reassociation (162). DUBs can hydrolyze the isopeptide or peptide bond between the ubiquitin C terminus and a substrate (e.g., another ubiquitin molecule). The linkage specificity of DUBs and the recognition mechanisms of the highly complex, so-called, ubiquitin codes has been the subject of extensive study (**Figure 4**) (119, 163, 164). The extent of linkage specificity varies between DUB families. Some DUBs can bind directly to a protein client and remove ubiquitin from it while another can recognise the ubiquitin chain itself. The orientation of Ub-binding sites in DUBs determine whether polyubiquitin is removed from the external distal or proximal end, from within a chain or from the branch points in polyubiquitin (164). The role of DUBs in ubiquitin chain cleavage is wide-ranging and involves recycling of ubiquitin from cargo destined for the proteasome or lysosome (165), or freeing up the proteasome from proteins that are difficult to degrade to avoid the accumulation of ubiquitinated proteins (166). Furthermore, a great number of DUBs work as parts of larger macromolecular machinery (e.g., the proteasome), which recruit substrates for deubiquitination. In these cases, DUB–substrate interaction is indirect and may not be governed by the DUB itself (164).

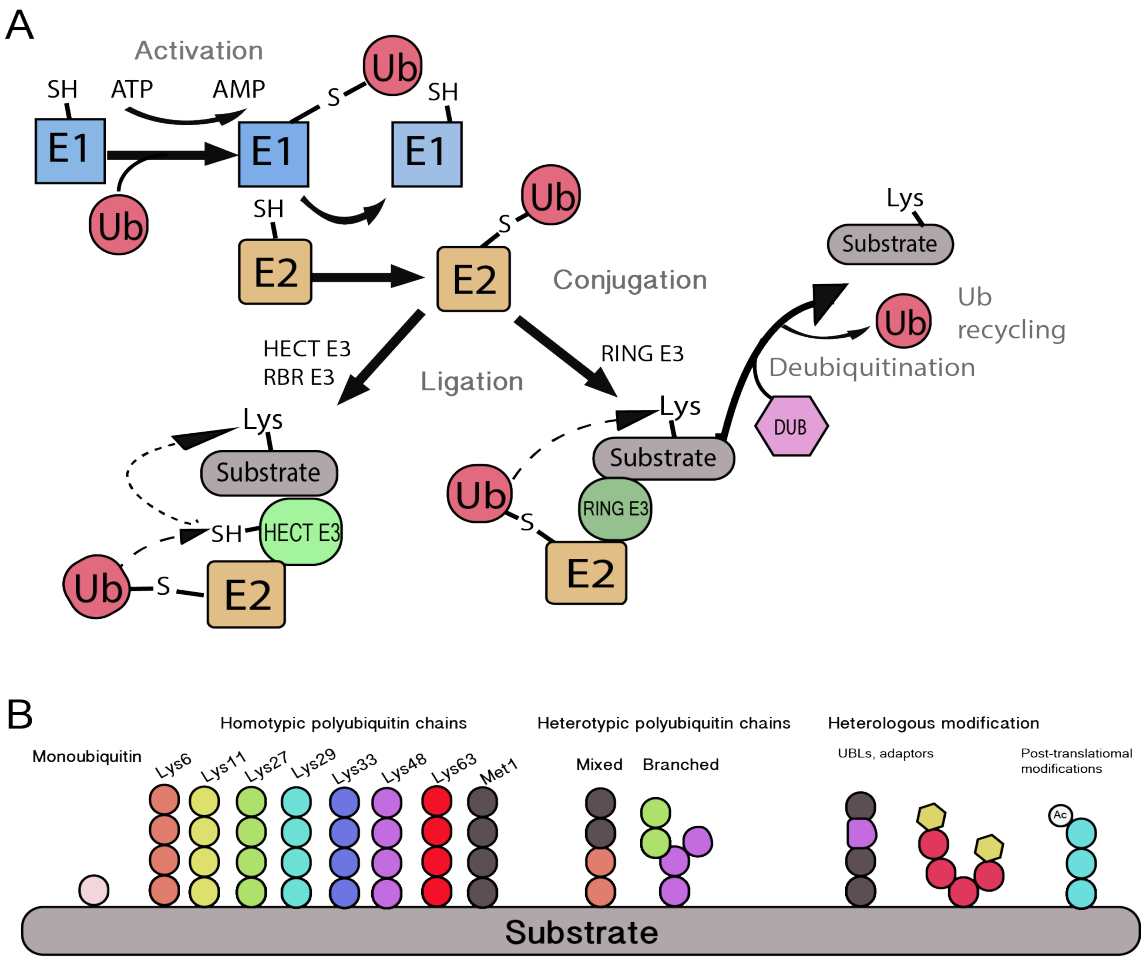

Figure 4: Substrate ubiquitination by the E1/E2/E3 enzymatic cascade. A) The coordination between E1, E2 and E3 enzymes transfers the ubiquitin onto the substrate. Two different modes of action are shown for HECT and RING E3 ligases. Deubiquitinases (DUBs) cleave the ubiquitin from modified proteins. B) The types of ubiquitin linkages. Monoubiquitin, can be extended on one of its eight lysine residues or its N-terminal methionine (Met 1) residue, forming different patterns of homotypic or heterotypic polyubiquitin chains. The heterologous modifications may involve the incorporation of modifiers like ubiquitin-like domains (UBLs) or various adaptors and also post-translational modification like phosphorylation or acetylation (Ac) (adapted from (158) (top) and (164) (bottom).

1.5 SGTA, the BAG6 complex and RNF126 in protein quality control

It has gradually been established that the BAG6 complex (BAG6, UBL4A, TRC35) together with SGTA play a crucial role in the quality control of misfolded and mislocalised proteins (Figure 1). They receive and triage various substrates, each characterised by an exposed hydrophobic region, and maintain them in a soluble state until they can be transferred to the following partner in the relevant pathway (74, 86). The quality control mechanisms, in which SGTA and the BAG6 complex are involved, include transient binding to components of the ribosome (72) and the

proteasome (74, 167, 168). They participate in the DNA damage response when localised to the cell nucleus via the nuclear localization signal on BAG6 (169) and in ERAD, where misfolded proteins are ubiquitinated and retrotranslocated from the ER to the cytoplasm to enable their proteasomal degradation (160, 170, 171). SGTA recognises and binds exposed hydrophobic regions of different targets (TMDs of TA proteins and hydrophobic regions of mislocalised membrane and secretory proteins) and transfers them onto the BAG6 complex to facilitate either the polyubiquitination of hydrophobic substrates and degradation at the proteasome or their appropriate onward delivery to the ER (19, 74, 172). Previous experiments showed that SGTA's depletion from a cell-free translation lysate impeded nascent TA protein capture by TRC40 and BAG6 with subsequent reductions in both ER insertion and ubiquitination (73). Conversely, depletion of the BAG6 complex or TRC40 caused TA proteins to be retained on SGTA (73). This suggests that SGTA facilitates both targeting and degradation, indicating that TA proteins bound to SGTA at this stage are not necessarily committed to either fate (73). A few years ago, RNF126 (ring finger protein 126) was identified as a soluble E3 ligase which is recruited to BAG6 for ubiquitination of BAG6 associated hydrophobic clients that are destined for degradation (75). RNF126 contains two individual domains: an N-terminal zinc-finger domain and a C- terminal RING domain (173) with the former region considered to play an key role in the interaction with the UBL domain of BAG6 (75). Furthermore, SGTA and RNF126, both bind to the same region of the BAG6 complex (174). The precise details of the sorting mechanisms are unclear. Hence it seems that the triage decision between degradation at the proteasome or the dedicated onward delivery to the ER after the recognition of the exposed hydrophobic regions of TAs and MLPs depends on accurate time and kinetics (19, 73). After hydrophobic substrates dissociate from SGTA, they are transferred onto the components of the ER targeting module (C-terminal BAG6/ UBL4A/TRC35/TRC40) and the quality control module (N-terminal BAG6/RNF126). The triage depends on how fast they can bind the hydrophobic substrate and how long they can maintain it bound. The abilities of SGTA, BAG6 and TRC40 to compete for free TA protein are inversely associated to their ability to retain the bound client (73). Hence, TAs transfer from SGTA to TRC40 through the C-terminal of BAG6 is faster than the rate of spontaneous TA protein dissociation from SGTA. Conversely, TAs handover from SGTA to BAG6 is slower and involves

a free TA intermediates (73). SGTA collaborates with the BAG6 complex and links the targeting and ubiquitination pathways. However, the precise function of SGTA in each of these processes is less well established.

The role of SGTA in the specific recognition of hydrophobic substrates has been studied using short hydrophobic fragments and showed that SGTA interacts directly with MLPs, both *in vitro* and *in vi*vo. They highlight the importance of an uninterrupted hydrophobic sequence for this binding to occur (86). It has also been suggested that SGTA can actively promote the deubiquitination of mislocalised proteins by binding their exposed hydrophobic sequences in the cytoplasm and rescuing them from degradation (86, 99). In the case of TAs, it has even been proposed that SGTA may be able to rescue substrates that are prematurely ubiquitinated (19, 74). Furthermore SGTA and the BAG6 complex are also involved in the elimination of aberrantly folded proteins from the ER in the endoplasmic reticulum-associated degradation pathway (ERAD) (160, 171, 175, 176). In this quality control mechanism, SGTA and the BAG6 complex maintain retrotranslocated polypeptides in a soluble state, promoting their turnover (171). Importantly, SGTA deficiency impairs ERAD and induces unfolded protein response, a stress response induced by the accumulation of misfolded proteins at the ER (171).

The importance of the collaborative work of SGTA and the BAG6 complex in the quality control of hydrophobic substrates has been established. However, much structural information on the complete BAG6 complex remains to be obtained. For example, the structure and oligomerization architecture of the 119 KDa BAG6 (also known as Bat3 and Scythe) protein itself has not yet been defined. The non-canonical 'BAG domain' close to the C-terminal part of BAG6 is known to bind directly to a C-terminal region of UBL4A with crystal structures of the minimal complex solved (76, 177). TRC35 also binds to the C-terminal region of BAG6, called the 'nuclear localization sequence' which is close to the 'BAG domain', through its own C-terminal domain. Hence, BAG6 connects TRC35 and UBL4A (76) and allows substrate transfer between SGTA,

BAG6 and TRC40 (73). The BAG domain from BAG6 is called a 'mock BAG' domain because it does not share the function of other BAG domains like the one of BAG1 which can bind Hsp/Hsc70 and inhibit its actions involved in vitro refolding of an unfolded protein substrates (178, 179). Additionally, the BAG domain is structurally different, compared to canonical BAG domains (76, 180). Interestingly, before BAG6 was established as BAG6 family member, the protein was initially associated with apoptosis based on its interaction with an apoptosis-inducing protein, in *Drosophila melanogaster* (called Reaper) (181). The poorly defined middle region of BAG6 is capable of binding hydrophobic substrates (19).

Notably, the activity of the BAG6 complex itself is also subjected to the quality control. The ERAD E3 ligase, GP78, in addition to ubiquitination of ERAD substrates, can also tag UBL4A, a key component of the BAG6 complex. Interestingly, instead of directing UBL4A for proteasomal degradation, polyubiquitination is linked to the irreversible proteolytic processing and inactivation of BAG6. However, ubiquitin molecules can be removed from UBL4A by USP13, a GP78-associated DUB which maintains the functionality of the BAG6 complex (182).

In yeasts, which do not have the BAG6 quality control subunit, cytosolic chaperones like Hsp70 chaperones (Ssa 1-4 in yeast), the Hsp40 chaperone (Ydj1 in yeast) and Hsp104 disagregase, have the ability to extract and refold proteins from aggregates (183-186). The main E3 ligase responsible for ubiquitination of misfolded proteins in the cytosol is the E3 ubiquitin-protein ligase Ubr1 (187), which involves Hsp70 and Hsp110 chaperones in substrate ubiquitination (143, 188). While, at the endoplasmic reticulum, the ER membrane brings E3 ligases together with the components of ERAD such as the Cdc48 (Cell division control protein 48) machinery which typically act in ubiquitination of ERAD substrates but are also capable to target some substrates localised in the cytoplasm for degradation by the proteasome. Other cytoplasmic substrates are transported into the nucleus in a process mediated by chaperones where ubiquitination is facilitated by the E3 ubiquitin-protein ligase San1, which tags misfolded nuclear proteins for

degradation (143, 189). Furthermore, some aberrant proteins at the cytoplasm and nucleus can also be targeted by autophagy and degraded by the vacuole/lysosome in mammalian and yeast cells (4, 190, 191).

1.6 Structural and functional overview of RNF126

RNF126 belongs to the family of RING (really interesting new gene) E3 ligases and comprises two individual zinc finger domains: an N-terminal zinc-finger domain (residues 1-100), and a C-terminal RING domain (residues 229-270) (75, 173). The N-terminal zinc-finger domain is considered to play an important role in the interaction with the UBL domain of BAG6 (75) while the RING domain is required for the E3 ubiquitin ligase activity (173). In addition, RNF126 contains an unusual stretch of ~12 serines at the C-terminal that is evolutionarily conserved between zebrafish and human and is similar to stretches found in transcriptional activators present at RNAPII-dependent promoters (192, 193).

Zinc fingers are small motifs which usually contain multiple finger-like overhangs that interact with their target molecules including RNA, DNA and proteins. They vary in function and topology and require coordination of one or more zinc ions to stabilise their structures. The binding properties of zinc fingers rely on the amino acid sequence of the domain and the linkers between fingers, as well as the three dimensional architecture and the number of fingers (194). They are classified into several different families based on their structure and the residues coordinating the zinc ion, usually a combination of four amino acids, either cysteine or histidine. The most abundant in eukaryotic genomes are proteins containing the Cys2His2 motif many of which function as zinc finger transcription factors that recognise specific DNA sequences (195). Other examples of zinc finger families include Cys4-, Cys6- or Cys3His -types. RNF126 belongs to a family of proteins containing a RING finger domain comprising a Cys3HisCys4 amino acid motif which binds two zinc ions (seven cysteines and one histidine). Most RING domains are small and contain from 40 to 60 residues (196). The solution structure of the mouse RNF126

RING domain has been solved (PDB Accession Number: 2ECT; with no associated publication) and the structure of the N-terminal zinc finger domain is described in chapter 4. However, to date, full-length RNF126 has not yet been structurally characterised. Like RNF126, other proteins containing a RING finger are also involved in the ubiquitination pathway (197). Examples include several functionally diverse RING finger proteins, like BRCA1 (breast cancer type 1 susceptibility protein**)** and TRC8 (Translocation in renal carcinoma on chromosome 8 protein), which can facilitate E2-dependent ubiquitination (198).

Typically, RING domains of E3 ligases transiently interact with the N-terminal helix of E2 conjugating enzymes (199) and may compete with binding of the E1 enzyme to E2 (200). In addition to the E2-binding domain, RING E3 ligases possess domains like regions that can bind substrates destined for ubiquitination. For example, the RING E3 gp78 ligase contains an additional domain, that binds to the E2 β-sheet on the side opposite to the active site (to which free ubiquitin can bind in solution) and can cause allosteric changes in the E2, opening the active site for substrate binding and promoting ubiquitination (136, 201). Although a number of structures of RING domain–E2 complexes have been characterised (199, 202-205), the mechanism by which RING domains facilitate ubiquitin transfer remains elusive.

1.6.1 RNF126 as an E3 ligase

The E3 ligase activity of RNF126 relies on its recruitment to the UBL domain of BAG6 where it can ubiquitinate BAG6-associated hydrophobic clients through its C-terminal RING domain (75). Although the substrates specificity of RNF126 is unclear, it has been reported that it preferentially ubiquitinates lysine residues which are close to the BAG6 binding site, which may suggest that RNF126 recruits the E2-ubiquitin conjugate in close proximity to the BAG6 substrate binding region (75). RNF126 can mediate the formation of K63-linked ubiquitin chains with ubiquitin-conjugating E2 enzymes Ubc13/Uev1a and also K48-linked chains with E2 conjugating enzyme

UbcH5b *in vitro* (206). The N-terminal zinc finger domain is not required for formation of a poly-ubiquitin chains *in vitro* (206).

While BAG6-RNF126 ubiquitination is considered to be a primary substrate degradation pathway, there is also evidence that there must be alternative ubiquitination systems operating in the absence of RNF126 (75). It is unclear whether another E3 ligase, RNF115 (also called Rabring7 or BCA2 - breast cancer-associated gene 2), which is similar to RNF126, can function in a rescue degradation pathway and ubiquitinate the same range of substrates (75). RNF126 shares similar structural features with RNF115, which also contains N-terminal and C-terminal zinc fingers. RNF115 has previously been shown to have RING finger dependent auto-ubiquitination activity (207-209). The malfunction of RNF115 is linked to invasive breast cancer related to the regulation of the estrogen receptor (210, 211). Furthermore, RNF115 promotes the lysosomal degradation of trapped human immunodeficiency virus (HIV-1) virions and also functions as anti-HIV factor by targeting Gag (HIV-1 precursor of the structural proteins) for lysosomal degradation impairing virus assembly and release (212, 213).

1.6.2 The role of RNF126 in cancer

The past few years have seen efforts in identification of E3 ligases as anti-cancer targets (214-216). The activity of RNF126 has been linked to cancers, including breast cancer, prostate cancer and tongue cancer (173, 193). It has been reported that RNF126 can regulate the cell cycle by promoting cancer cell proliferation and targeting the tumor suppressor p21 cyclin-dependent kinase inhibitor for ubiquitin-mediated degradation (173). Furthermore, RNF126 knockdowns inhibit the cell viability of tongue cancer cells and prostate cancer (173, 193). RNF126 also exhibits ubiquitin ligase activity against pyruvate dehydrogenase kinase (PDK) where it is necessary for the tumorigenicity of cancer cells in mice (217) and also targets activation-induced cytidine deaminase (AID), an enzyme that deaminates deoxycytidines in single-stranded DNA (192). Conversely, the function of RNF126 was linked to the regulation of homologous

recombination (HR)-mediated DNA double-strand break repair by promoting the expression of BRCA1 tumor suppressor, in a manner independent of its E3 ligase activity but by direct interaction with the E2F1 transcription factor (218).

1.6.3 RNF126 in the regulation of cell signalling

Both, RNF126 and RNF115 regulate the ubiquitin-dependent endosomal sorting of cell surface receptors, like epidermal growth factor receptor (EGFR) and chemokine receptor type 4 (CXCR4) (206). They associate with EGFR and CXCR4 via their ubiquitin binding N-terminal zinc finger domains and promote ubiquitination leading to lysosomal degradation (206). Furthermore, RNF126 and RNF115 alter the sorting machinery found at the late endosome and inhibit the formation of multivesicular bodies (206). Moreover, RNF126 regulates the retrograde sorting of the cation-independent mannose 6-phosphate receptor (CI-MPR) in the pathway where a protein cargo undergoes retrograde transport from the endosome to the trans-Golgi network within the endocytic system (219).

1.6.4 RNF126 and Frataxin

Friedreich's ataxia is a neurodegenerative disease caused by decreased expression of the mitochondrial protein frataxin of which levels are controlled by the ubiquitin-proteasome system. To date, there is no effective therapy to treat the disease (220-222). Patients live with 5%–30% residual frataxin and the severity of the disease progresses with the reduction of frataxin (223). Frataxin is synthesized in the cytosol in a precursor form and contains an N-terminal mitochondrial localisation signal, which allows the precursor to be targeted to mitochondria (224), where it plays a crucial role in the biosynthesis of iron-sulfur clusters (225, 226) and in iron metabolism (227, 228). RNF126 was shown to preferentially bind to the precursor form of frataxin as compared to the intermediate and mature forms and to promote frataxin ubiquitination and degradation (229). Binding experiments using catalytically inactive RNF126 with mutations of the two critical cysteines in the RING domain show that the mutant maintains the ability to

interact with frataxin, although can no longer ubiquitinate frataxin (229). Further study is necessary to clarify which domain of RNF126 directly interacts with frataxin.

1.7 Structural and functional overview of the co-chaperones SGTA and Sgt2

Small glutamine-rich tetratricopeptide repeat-containing protein α (SGTA) was discovered almost twenty years ago as a novel interaction partner of viral proteins (230, 231). SGTA is conserved amongst species with a common ancestor in animals and fungi (**Figure 5**). The *S. cerevisiae* orthologue to human SGTA (called Sgt2) shares 32% sequence identity (82). Other homologues include *C. elegans, D. rerio*, *M. musculus, X. laevis, P. pastoris* and *A. fumigatus* (**Figure 5**). SGTA is expressed in almost all tissues with the protein levels varying between species (232) and the alternative isoform β-SGT is predominantly brain-specific and shares 60 % amino acid sequence identity with SGTA (232, 233). SGTA has been found to localise primarily to the cytoplasm. However, its presence in the nucleus has also been detected (232).

Since their discovery, the roles of SGTA and Sgt2 have been established in several biological processes. These include the post-translational transport of TA membrane proteins (87, 88, 90, 97, 105), regulation of protein quality control (19, 86, 99, 234), cell cycle and apoptosis (235), intracellular compartmentalization (236, 237), neuronal synaptic transmission (238), viral infection (230, 231, 239-241) and hormone signalling (232, 235, 237, 242-244).

Since SGTA is engaged in such a variety of biological processes, it is not surprising that it has been implicated in many diseases including cancer of the prostate (237, 245-247), ovary (248), liver (249) and oesophagus (250); polycystic ovary syndrome (251); Alzheimer's disease (238) and prion disease (252). Furthermore, a functional *in vitro* study showed that knockdown of SGTA in some cells can reduce cell viability and proliferation (237) and promote cell death (235, 253) while SGTA's ablation in mice models limits the viability and growth of offspring

(254). Given the importance of SGTA in so many diverse cellular processes it is important to understand its mechanism of action in health and disease.

1.7.1 Structural overview of SGTA/Sgt2

SGTA and Sgt2 are both approximately 34kDa dimers and thus far, a number of biophysical techniques have been used to characterise their structure in some detail (**Figure 5**), and to study their interactions with binding partners in diverse biological systems. SGTs consists of three structural domains (**Figure 5**) with individual functions: an N-terminal dimerisation domain, followed by the conserved, central TPR domain and a glutamine rich region towards the C-terminus (82, 87, 90, 101, 234, 255). SGTs of *C. elegans, D. rerio, and A. fumigatus* seem to have longer N-terminal linkers compared to both mammalian and yeast homologues. TPR domains are the most conserved, but the TPR capping helix (helix 7) shows lower conservation in fungi than in mammals (**Figure 5**). The N-terminal domain of SGTA (Sgt2) can directly bind the ubiquitin-like (UBL) domain of BAG6 and UBL4A (Get5 in yeast) and facilitate the handover of TA substrates downstream onto components of the TRC pathway (Get pathway in yeast) for membrane delivery (73, 82, 87, 99). The C-terminal domain includes a glutamine-rich region, which comprises a stretch of 39 amino acids including 12 glutamine residues (231, 234) and is structurally uncharacterised. These domains of both Sgt2 and SGTA bind hydrophobic proteins including the TMDs of TA substrates (171, 234, 240, 256). Tetratricopeptide-repeat (TPR) domains typically consist of three or more tandem repeats of a loosely conserved 34 residue motif (257). Each tandem motif is formed of two anti-parallel α-helices, organised in a helix-turn-helix system, and forming arrays of 3–16 motifs (258). The TPR sequence varies with large and small hydrophobic amino acid residues highly conserved at positions that are important for structure and function. Other positions are less conserved and allow the replacement by amino acids with similar properties (258, 259). TPR domains are well-known for mediating protein-protein interactions (136). The structure of the human SGTA TPR domain was determined previously by X-ray crystallography (240) and has been reported to interact directly with Hsp70/Hsp90

chaperones, the proteasomal subunit Rpn13 and a variety of disease-related proteins (168, 236, 240, 260).

The TPR domain of SGTA exhibits sequence similarity to the TPR domains of Hop (Co-chaperone Hsp70/Hsp90 organizing protein), CHIP (C-terminus of Hsc/p70-Interacting Protein), and TOM70 (Mitochondrial import receptor subunit) (236). These TPR domains are known to interact with Hsp70/Hsp90 proteins. The high-resolution structures of such complexes were reported for HOP TPR1 with an Hsp70–derived peptide, HOP TPR2A with an Hsp90-derived peptide (261), CHIP TPR with an Hsp70 C-terminal (171, 262) and *Af*Sgt2_TPR (from *Aspergillus fumigatus*) modeled with its C-terminal linker (263). The common mode of interaction involves the formation of a carboxylate clamp where both the side-chain and main-chain terminal carboxylate groups of the C-terminal aspartic acids of these peptides form salt bridges with residues within the groove of the co-chaperone's TPR domains (264). To date the structure of the Sgt2 TPR domain from *Saccharomyces cerevisiae* has not been solved.

Notably, the simple architecture of tandem-repeat proteins allows investigation of protein folding kinetics, thermodynamic characteristics (stability, energetics, and equilibrium two/multistate folding) (265-268). Such investigations allow controlled protein design to engineer the folding of repeat proteins, which then can be used in biomaterials (269) antibody substitutes (270) and disease therapeutics (265). Furthermore, specifically designed TPR proteins can serve as scaffolds for protein-ligand interactions and can help in understanding novel binding specificities (265). The structures of full-length SGTs presents a challenge because of their flexible linkers and dynamic C-terminal region. The full-length protein architecture of the dimer has been suggested to be elongated, with the three domains adjacent to each other (82, 234, 271). Evidence for this comes from the reported hydrodynamic radius of Sgt2's *C. elegans* homologue which has been observed to be larger than that of a globular protein (271). This is also supported by small angle X-ray scattering (SAXS) studies of the yeast Sgt2 devoid of its C-terminal domain which also

exhibits an elongated arrangement (263). Overall, the structural information on domain organization in the context of the full-length protein is yet to be discovered. In addition, the lack of conservation of the region between the N-terminal domain and the TPR, together with its length (10-15 amino acids) suggests the presence of a flexible linker between these two regions (260).

1.7.2 The interaction of SGTA and Sgt2 with molecular chaperones

SGTA/Sgt2 are co-chaperones, known to interact with general chaperones. SGTs associate with several heat-shock proteins such as Hsp104, Hsc82 (yeast homologue of Hsp90) and Ssa1/Ssa2 (yeast homologues of Hsp70), which can bind directly to its central TPR domain (105, 253). The Hsp70 and Hsp90 chaperones are important parts of the cellular machinery for protein folding, maturation and structural stability (272, 273). They often associate with co-chaperones containing multiple copies of TPR domains (273, 274) which help them to facilitate correct folding of client proteins (88, 275). The activity of chaperones, Hsp70 and Hsp90 is tightly coupled to ATP binding and hydrolysis, and is subject to complex regulation by a large number of co-chaperones (276, 277). It has been proposed that SGTA can also regulate the ATPase activity and folding rates of Hsp70 (242). General chaperones of the eukaryotic cytosol bind a wide range of substrates, among them proteins involved in signal transduction such as protein kinases and transcription factors. However, the detailed mechanism of the substrate recognition is poorly characterised and may partially reside within specific co-chaperones. While Hsp90 is thought to bind and stabilise partially folded polypeptides of some of its substrates, other proteins may be recognised in elongated conformations (6, 277).

A brain-specific isoform of SGTA, βSGT, shares 60% amino acid sequence identity and can also interact with Hsc70 (233). βSGT has also been discovered to bind the cysteine string protein (CSP), a small protein with string of 14 cysteine residues (244, 272). Modification of the cysteines mediates the attachment of CSP to secretory and synaptic vesicles and binding of the CSP to Hsc70 may recruit the chaperone to these vesicles (278). Trimeric complexes of CSP-Hsc70-SGT were also detected on synaptic vesicles (272). However, the functional implication

of this interaction remains to be elucidated, but the presence of the trimeric complex on synaptic vesicles suggests a specialised role in neuronal function (233).

Yeasts, which do not have the BAG6 quality control subunit, deal with damaged proteins in alternative ways and, for example, employing disaggregases and using general cytosolic chaperones like Hsp70 to refold proteins from aggregates. However it is not yet known whether the interaction between Sgt2 and chaperones is important in this context (183-186).

1.7.3 Role of SGTA in viral cycles

SGTA was first identified in the context of its ability to form a complex with a viral protein, a nonstructural protein of parvovirus H-1 (NS1) which is crucial for viral DNA replication and transcriptional gene expression. Evidence for this derives from data suggesting that SGTA is localised in both the cytoplasm and nucleus of rat fibroblasts, potentially linking SGTA's role to parvoviral replication and/or gene expression (231). Later it was discovered that SGTA has a wider role in viral infections and interacts with other viral proteins too. For example, SGTA also interacts with an accessory severe acute respiratory syndrome coronavirus protein 7a (SARS-CoV 7a), via its TPR domain (241). However, details relevant to mechanisms of this interaction and how SGTA is involved in the life cycle of this virus is poorly characterised.

In contrast, the role of SGTA in human immunodeficiency virus type 1 (HIV-1) particle release has been elucidated in some detail (230, 279, 280). HIV-1 encodes three key structural polyproteins (Gag, Pol and Env), two regulatory proteins (Tat and Rev), and four accessory proteins (Vif, Vpr, Vpu and Nef) (281). The two main roles of Vpu include degradation of CD4, the primary receptor for HIV-1 (282, 283) and enhancement of the viral particles release from the cell surface by inhibiting the activity of the host restriction factor tetherin (213, 280, 284). There is evidence that SGTA is involved in Vpu () mediated enhancement of newly formed viral particle

release. Overexpressed SGTA interacts with HIV-1 Gag (a viral core protein precursor) in a way that can be reversed by competition with Vpu. Interestingly, SGTA can interact with both Vpu and Gag. It has been suggested that in the absence of Vpu, SGTA binds to HIV-1 Gag and forms a complex that inhibits HIV-1 release. However, in the presence of Vpu, the SGTA-Gag interaction is abrogated, possibly caused by SGTA forming a more stable complex with Vpu. An independent study reported that SGTA's overexpression in mammalian cell lines transfected with HIV-1 proviral constructs reduced the efficiency of virus particle release (230). Furthermore, it has been suggested that SGTA may promote shuffling of the viral protein Gag to the plasma membrane where it assemblies into HIV-1 virus capsids (279). However, the interaction between SGTA and Gag is terminated when Vpu is overexpressed in the cell. It has also been shown that Vpu can influence cellular localisation of SGTA and Gag proteins (230, 279). Nonetheless, the TPR domain has been shown to be sufficient for SGTA's binding to Vpu and Gag albeit with lower efficiency than full length SGTA (240). In addition, SGTA can stabilise expression of a non-glycosylated form of the host restriction factor, tetherin, and it is the C-terminal of SGTA which appears to be necessary and sufficient for this action (280). Notably, the non-glycosylated tetherin form appears to partially mislocalise to the cytosol in the presence of Vpu and SGTA, and the relocalisation depends on the presence of the C-terminal domain of SGTA (280).

Interestingly, the role of SGTA in ERAD may be extended into extraction of a hydrophobic protein complex Simian virus 40 (SV40) from the ER into the cytosol during SV40 infection. This way it may be promoting viral infections by assisting in membrane penetration (239, 285). During ER-to-cytosol membrane transport of the nonenveloped polyomavirus SV40, a decisive infection step, a cytosolic complex composed of Hsc70-SGTA-Hsp105 was previously shown to associate with the ER membrane. SGTA-Hsp105 extract SV40 from the ER and transport the virus into the cytosol. However how SGTA assists Hsc70 and Hsp105 in ejecting SV40 into the cytosol, remain to be elucidated (239).

1.7.4 Role of SGTA in the regulation of hormone receptor signalling

Steroid hormone receptors are ligand-dependent transcription factors (typically either cytoplasmic or nuclear) that require the multi-chaperone complexes for their correct folding and transcriptional activity (286-288). There has been growing evidence that the co-chaperone SGTA, through its interactions with chaperones and steroid receptors, may be important in the pathogenesis of androgen-associated disorders like prostate cancer, breast cancer and polycystic ovary syndrome (248-251). Research investigating the role of SGTA has been conducted in several model organisms and cell types, showing that SGTA interacts with several steroid receptor complexes and signaling pathways. SGTA specifically regulates the activity of the androgen receptor (AR), a nuclear transcription factor (289). This occurs through the interaction between the TPR domain of SGTA and the part of the AR localised between the DNA-binding domain and the ligand-binding domain, called a hinge region (236). The role of SGTA in the AR signaling pathway is not clear but it appears to be upregulated in most cancer cell lines (289). SGTA's knockdown/deletion enhances receptor activity, while SGTA overexpression suppresses receptor activity (289). It is also speculated that SGTA dimerization is linked to the suppression of the androgen receptor signalling by negative regulation of AR transport to the nucleus and by inhibiting the connection between the cytoplasmic AR complex and the dynein motor complex (236, 237, 254, 290). The tumor suppressor, REIC/DKK-3 (Reduced Expression in Immortalized Cells/Dickkopf-3) seemingly interferes with SGTA dimerization and rescues AR signaling (247). REIC/DKK-3 is expressed ubiquitously in normal cells, whereas its expression is downregulated significantly in various types of cancer cells, including prostate cancer (291, 292). Furthermore, another protein which is a member of the endogenous retroviruses of the HERV-K (HML-2) family, Rec, has been proposed to regulate AR activity through its interaction with SGTA (290). It has also been shown that SGTA modulates the activity of glucocorticoid and progesterone receptors (289). In addition, SGTA has been linked to mechanisms that involve endocytosis of the growth hormone receptor which is regulated by the ubiquitin-conjugating system. SGTA was shown to interact with growth hormone receptor through its first TPR motif in a study involving using yeast two hybrid and pull-down assays (243). However, in the absence

of molecular details describing the direct interactions of SGTA with hormone receptors, the role of SGTA in pathogenesis of hormone signaling related diseases remain largely elusive.

1.7.5 The interaction of SGTA with the proteasomal ubiquitin receptor Rpn13

Rpn13 and Rpn10 are two intrinsic ubiquitin receptors of the 26S proteasome regulatory subunit 19S that facilitates substrate capture prior to degradation (148, 149). Two years ago it was shown that SGTA associates with the proteasome through a direct interaction between its TPR domain and the C-terminal domain of Rpn13 (167). Furthermore, the BAG6 subunit of the BAG6 complex can bind Rpn10 (293, 294). On that basis, it has been proposed that the BAG6/SGTA quality control cycle may be operating at the 19S regulatory particle of the proteasome, capable of moderating the access of MLPs targeted for degradation to the proteolytic machinery (167). This would allow the substrates several rescue attempts before being terminally committed to degradation. Furthermore Rpn13 also interacts with the deubiquitinating enzyme UCH37 (UCH-L5) in a mechanism that faciliates UCH37 access to its ubiquitinated substrates (295). Hence in the context of the BAG6/SGTA actions, it has been proposed that the binding of SGTA to Rpn13 provides SGTA-associated substrates with access to the UCH37 deubiquitinase and could therefore mediate the rescue of prematurely ubiquitinated proteins (99, 167).

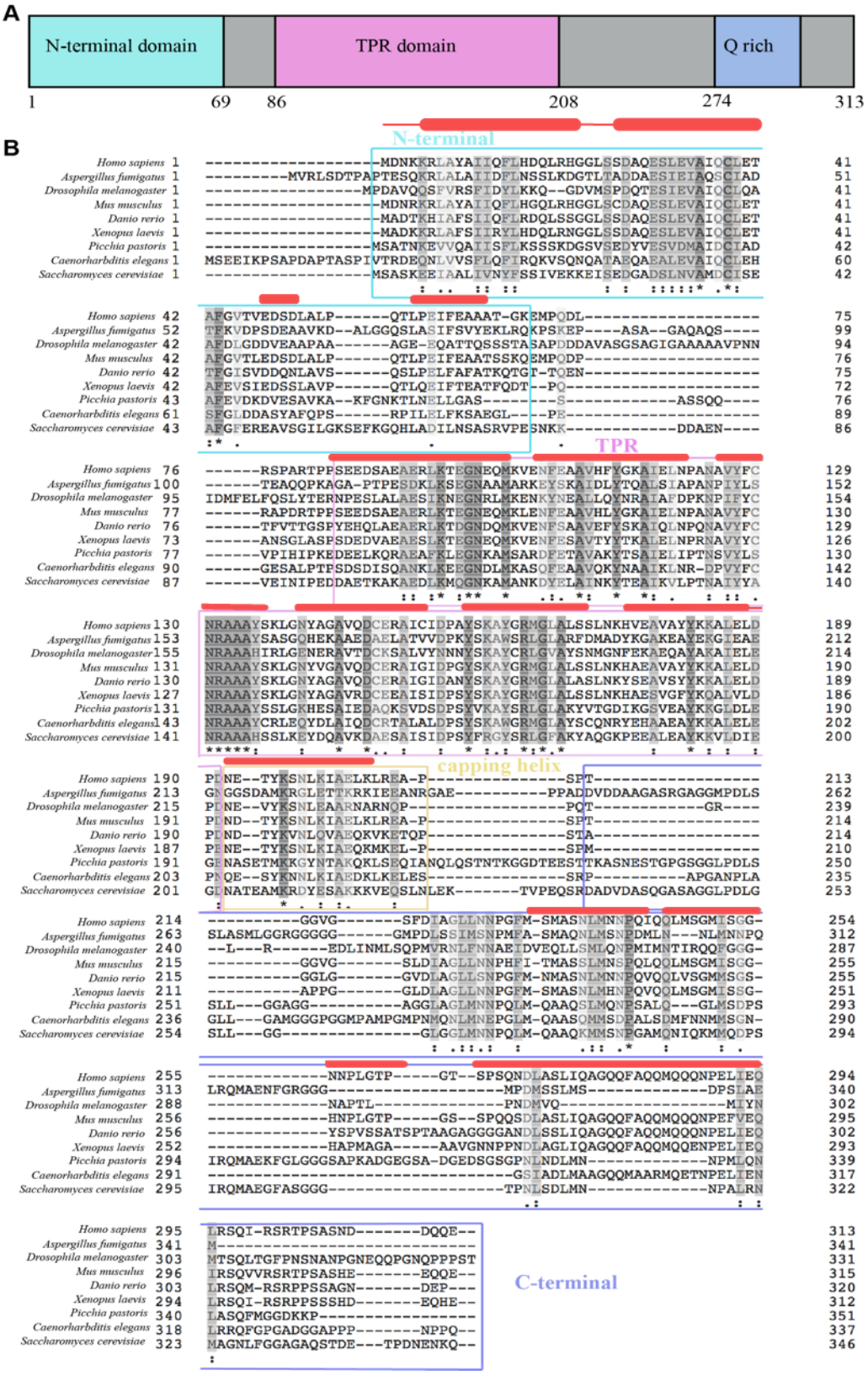

Figure 5: Domain architecture and a sequence alignment of SGTA. A) Full length SGTA contains an N-terminal dimerization domain followed by a 14-residue linker, three TPR repeats with a capping helix and a C-terminal domain comprising a Q-rich region. Numbers indicate the number of amino acids in the proteins. B) Sequences of SGTA and its homologs are aligned, and domains annotated according to the human SGTA sequence. Secondary structure as predicted by Jpred is depicted as red cylinder for helices.

1.8 The SND alternative membrane protein targeting pathway

A new pathway in *Saccharomyces cerevisiae* has recently been discovered which has been proposed to be the back-up mechanism for delivery of substrate proteins to the endoplasmic reticulum and can also operate in the event of GET system failure (296). The three proteins which are part of this pathway, named Snd1, Snd2 and Snd3 (named for SRP-independent targeting), were reported to work together with a possible role in targeting substrates to the translocation machinery Sec61 (296). Snd1 (SRP-independent targeting protein1) is predicted to be soluble, localizes to the cytosol (296, 297) and has been proposed to associate with eukaryotic ribosomal complexes (298). Snd2 (SRP-independent targeting protein2) is predicted to have four transmembrane domains, has been shown to localize to the ER membrane (296) and was previously shown to affect the maturation of the carboxypeptidase Y (299). Snd3 (SRP-independent targeting protein 3), is predicted to have one transmembrane domain (296) and also localizes to the ER (297). All three Snd proteins appear to collaborate and operate in one pathway, in which Snd1 interacts with Snd2 and Snd2 can bind Snd3 (296). Furthermore, Snd2 and Snd3 were found in a complex with the Sec61 translocon and a model substrate (296). Although the relationship between the SND proteins and the SRP and GET pathways has been investigated, the mechanistic details have yet to be elucidated. Nevertheless, discovery of the new pathway and the role of the Snd proteins demonstrates the complexity of the cellular networks involved in efficient targeting.

1.9 Sgt2 and Ybr137wp in the GET pathway

Recently, Sgt2 was reported to interact with Ybr137wp, a protein that is specific to fungi and no homologous sequence has been found in mammals to date. The Sgt2 TPR domain binds to the C-terminal end of Ybr137. Ybr137wp forms a decamer in its crystal form and also in solution as observed by size exclusion chromatography and analytical ultracentrifugation (300). In the crystal structure five Ybr137wp molecules form a homodecamer and each Ybr137wp monomer contains seven α-helices and four β-strands (β1 to β4). The four β-strands form an antiparallel β-sheet

creating the core of Ybr137wp surrounded by α-helices (300). An immunoprecipitation study indicates that Arg171 and Arg175 of the Sgt2 TPR domain interact with Ybr137wp and the two point mutations of these two arginine residues abolish the interaction of Sgt2 with Ybr137wp (88). In addition, isothermal titration calorimetry (ITC) experiments using a Ybr137wp construct lacking the C-terminal SLEEDLNLD motif abolished its binding to the TPR domain of Sgt2, indicating that the C-terminal acidic motif of Ybr137wp is necessary for the interaction (300). The function of Ybr137 has been investigated in the context of the Get pathway and yeast cell viability assays demonstrate that Ybr137wp is not essential for a regular TA protein targeting under normal conditions. However, it is able to rescue the TA protein delivery defect caused by a GET system that is impaired under starvation conditions (300). Furthermore, a deletion of *ybr137w* from the yeast strain promotes the association of heat-shock chaperones Hsp104 or Hsp70 with Sgt2 (88). This suggests that Ybr137wp and the chaperones compete for the same binding site on the Sgt2 TPR domain. Interestingly, Ybr137wp can also form a complex with Kap142p in the cytosol (301). Kap142p is a karyopherin that mediates nuclear export and import (301). However, the role of Ybr137wp in this interaction with a karyopherin is unclear. While it has been shown that Ybr137wp may contribute regulate TA protein targeting under certain condition, its detailed function in this pathway also remains to be uncovered.

1.10 Aims

The overall objective of this PhD project was to contribute to understanding the mechanisms governing the quality control pathways and the decision-making process in classifying hydrophobic protein clients either for degradation or biosynthesis. To achieve this, I aimed to biophysically characterise crucial proteins involved in the triage system, including the co-chaperone SGTA and E3 ligase RNF126, in relation to the BAG6 complex. To study the full length proteins of the SGTA/RNF126/BAG6 complex module in vitro is challenging, because of hydrophobicity/flexibility leading to aggregation or excessive size. Therefore, this study explores the structures and interactions of excised domains. To provide a better understanding of mechanistic details underlying the interactions between these proteins, the structures of the N-

terminal dimerisation domain of SGTA and the zinc finger motif of RNF126 E3 ligase, and their complexes with the UBL domains of BAG6 and UBL4A were determined. Project also focused on an additional role of the co-chaperone SGTA and its yeast homologue Sgt2, which is their interaction with general chaperones Hsp70 and Hsp90. Thus, the structure of the TPR domain of *S.cerevisiae* Sgt2 in complex with the C-terminal peptide of Ybr137wp was determined. In addition, the interactions of the TPR domain of Sgt2 with C-terminal fragments of Hsp70 and Hsp90 were also characterised. The methodology chosen to use in structural experiments included crystallography and nuclear magnetic resonance spectroscopy while the biophysical characterisations mainly involved isothermal titration calorimetry and microscale thermophoresis. Characterisation at the atomic level aimed to expand our understanding of the quality control mechanism in yeast and mammals, by providing molecular details of the components of the SGTA/BAG6 complex quality control module. They also have implications for future research, with the potential for development of therapeutic strategies.

Chapter 2. Methods

The majority of the research in this thesis uses nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, and isothermal calorimetry (ITC). Throughout these studies NMR is used extensively to assess the folding of protein constructs, assign protein resonances to facilitate protein structure solution, map chemically perturbed residues onto structure and determine protein-protein interactions. 1D NMR experiments have also been used to access the stability and folding of mutants (e.g., RNF126 mutants). ITC together with NMR was used to establish the binding affinities and molar ratios of interacting partners. Crystallography was used to solve the structures of a protein in its free state and in complex with a peptide. This chapter gives a brief overview of the theoretical and practical aspects of protein NMR spectroscopy, crystallography and ITC in the context of the experiments I have performed throughout these studies.

2.1 Principles and practice of solution NMR spectroscopy

2.1.1 Overview of basic concepts in NMR spectroscopy

NMR utilizes a physical property of atomic nuclei which means they can absorb and re-emit electromagnetic radiation, when placed in a magnetic field. Magnetic nuclei, composed of neutrons and protons, possess an intrinsic angular momentum known as spin. The spin has a magnitude and a direction and has an associated nuclear spin magnetic moment and a spin quantum number. Elementary quantum mechanics defines the spin quantum number, I, to have a integer or half integer value (e.g., 0, ½, 1,1½, 2, …). Any nucleus containing an odd number of protons or neutrons or both will possess a non-zero spin that can be detected as an electrical signal on NMR spectra. Typical NMR experiments employ nuclei with I = ½ such as ^{1}H, ^{15}N and ^{13}C, ^{19}F, ^{31}P, while ^{14}N and ^{2}H have an I value of 1. A nucleus with spin quantum number I in a magnetic field can occupy $2I$+1 non-degenerate energy levels. Therefore, spins with I of ½ can occupy only two orientations, often labelled α and β. In a static magnetic field B_0, the two allowed orientations of the nuclear magnetic spin moment have slightly different energies; a lower energy orientation aligned parallel to the field B_0 and a higher energy orientation antiparallel to the field.

The energy levels of these orientations are described by the quantum number, *m*, which has values in integer steps from -*I* to *I*. For example, for *I* of ½, m has a value of -½ and ½ and the number of energy levels equals 2. In a static magnetic field, such as is generated by an NMR spectrometer, in a sample at thermal equilibrium, the states are populated according to the Boltzmann distribution. The energy difference between the α and β states is minor and the difference in population is also small. However, there is some excess of spins in the slightly preferred lower energy level, which leads to a bulk magnetization M, aligned parallel to the magnetic field B_0 along the z axis. The direction of the magnetic field B_0 is described by the z axis of an x,y,z Cartesian coordinate axis system, known as the laboratory frame. A pulse is a linearly oscillating magnetic field applied along e.g., the x axis at a frequency (known as the transmitter frequency) typically around several hundred MHz. The pulse leads to the magnetization vector M tilting away from the z axis and a precession of the bulk magnetization in the xy-plane and around the z-axis which can be called the B_0 axis (Figure 6). The frequency of this rotation is known as the Larmor frequency and therefore the NMR frequency of the nucleus in the molecule is determined by its gyromagnetic ratio and the strength of the field. The rate of the Larmor frequency or angular frequency of precession can be defined as $\omega_0 = -\gamma B_0$, where γ is the gyromagnetic ratio of the nuclei under the pulse. The angle of the pulse depends on the angular frequency and the pulse length and can, for example, be equal to π/2 (a 90° pulse) or π (a 180° pulse) as shown in Figure 7. The flipping of the direction of the pulse axis can be achieved by modifying the phase of the radiofrequency field relative to the phase of the reference frequency of the detector. This magnetization produced by the pulse does not precess in the xy plane indefinitely and gradually returns to thermal equilibrium in a process called relaxation.

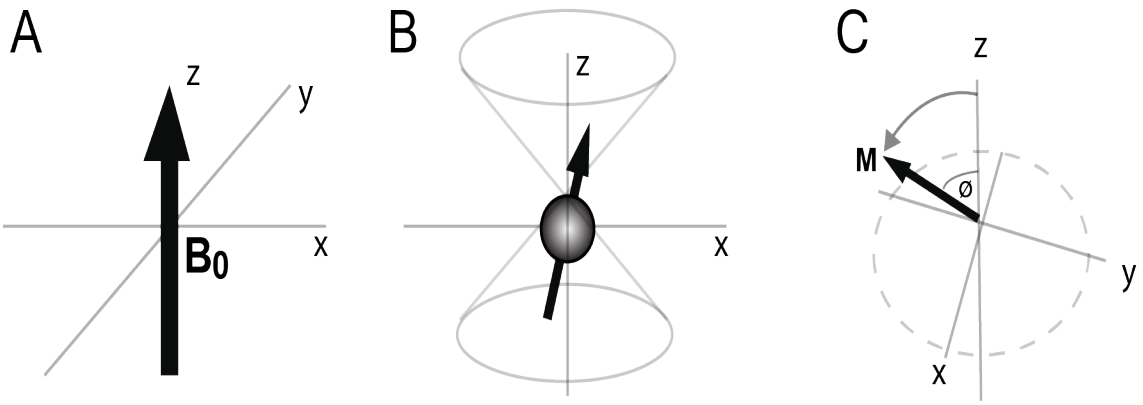

Figure 6: Classical representation of the magnetic moment of a nucleus with a ½ spin. A) Orientation of the B_0 and the bulk magnetisation before the pulse, magnetic moment M is aligned along the z axis of the laboratory frame. B) and C) The effect of the radiofrequency pulse. During the pulse, M precesses around the B_0 field along the x axis.

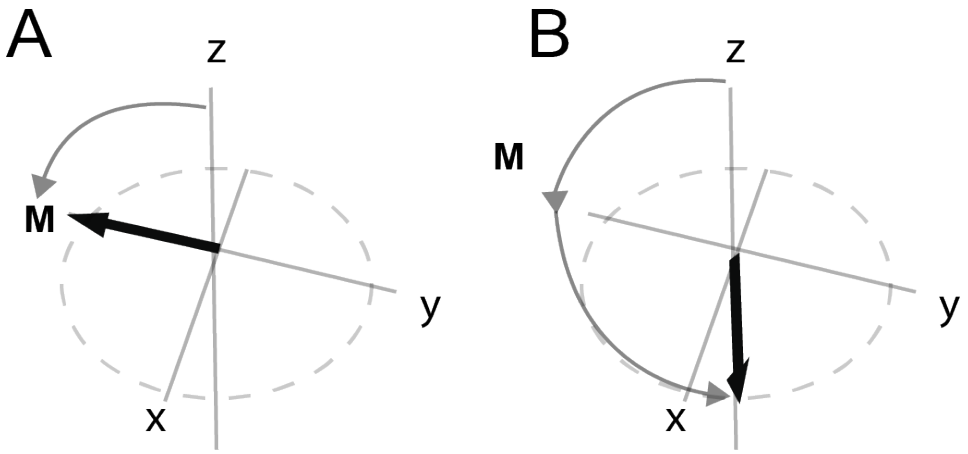

Figure 7: The effect of the radiofrequency pulse on the bulk magnetisation, M. A) after 90° pulse and B) after 180° pulse around the x axis. The flip of the bulk magnetisation is followed by a free precession.

2.1.2 The basic NMR experiment and detection of signals in the spectrometer

A common magnetic field used in NMR experiments is 2.35-23.5 T (tesla), which gives hydrogen nuclei resonance frequency of between 100 – 1000 MHz that is within the radiofrequency region of the electromagnetic spectrum (**Figure 8**). Proteins studied by NMR are commonly produced using isotopically-labeled media to generate proteins that contain ^{13}C and ^{15}N so that the specific quantum mechanical magnetic properties of these atoms can be measured. To date in practical applications, the frequency used has been within a 600–1000 MHz range.

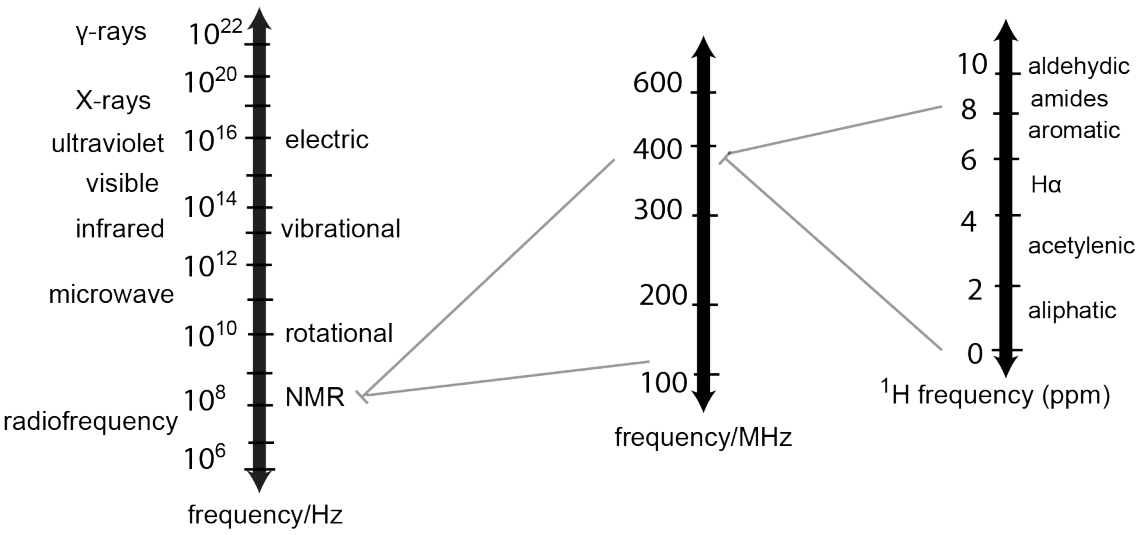

Figure 8: The electromagnetic spectrum and relative frequencies of NMR experiments. The electromagnetic spectrum with NMR operating frequencies compared to other spectroscopic methods (left). In the middle is the enlarged region of NMR frequency close to 400MHz for some nuclei. On the right is an expanded region of 400MHz showing some of the ^{1}H chemical shifts in parts per million, ppm, of a few functional groups of a peptide adapted from (302).

In a typical experiment the sample is placed in a receiver/transmitter coil of wire which is used to produce a short radiofrequency pulse (RF) that excites spins within a molecule and generates a phase coherence of magnetic moments. The coherent precession of spins induces an alternating current in the receiver coil, which is amplified and then detected to give the free induction signal. The signal decays over time which is called a free induction decay (FID) and this is recorded for a time called an acquisition time. Finally, a Fourier transformation of the FID mathematically transforms the NMR signal from its time domain into a frequency domain to obtain the peaks we can observe in the NMR spectrum. The way the NMR spectrum is recorded using a pulsed experiment is shown in **Figure 9**. The initial delays allow the system to return to equilibrium after a period of time typically from microseconds to seconds (called t_r for relaxation time). The NMR signal is usually weak and the spectrum obtained from a single FID is not sufficient for a good quality spectrum and signal-to-noise ratio. To improve the spectral quality, we repeat the experiment several times and add together the resulting FIDs. The FID is the sum of many oscillating waves of different frequencies, amplitudes and phases and, to filter this information,

the spectrometer employs detection channels known as quadrature detection during the Fourier transformation.

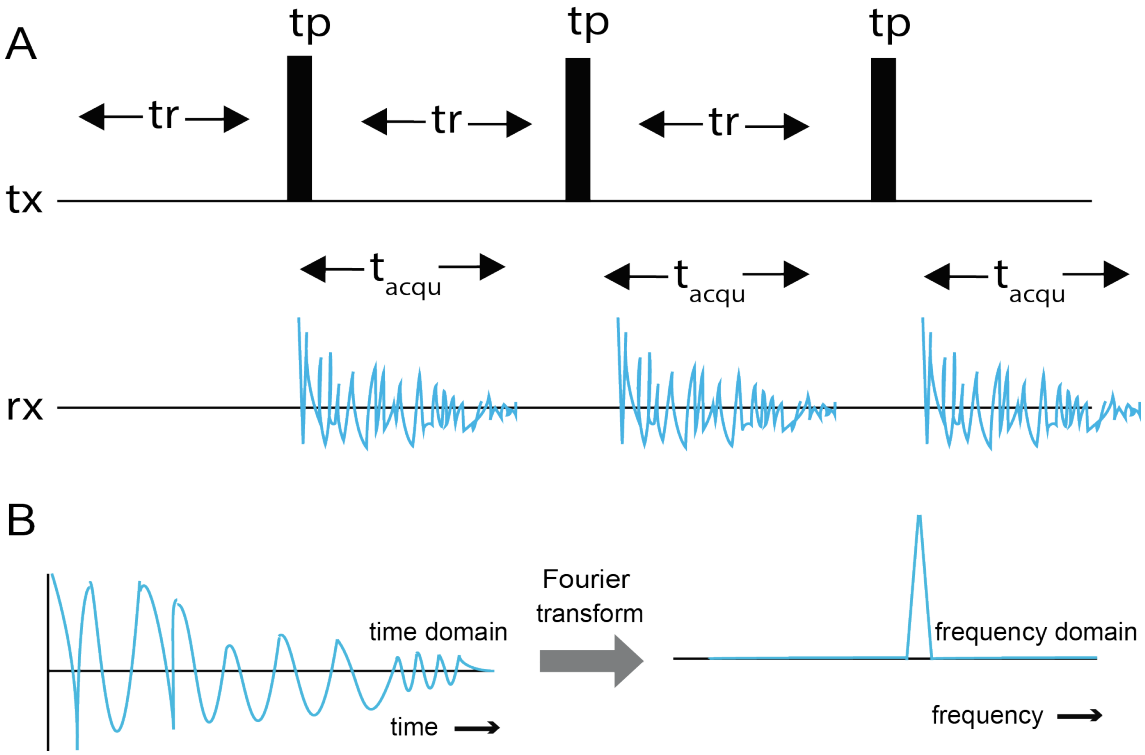

Figure 9: Schematic view of the recording of basic NMR experiments. A) and B) During time tr the spins achieve equilibrium. A short RF pulse is applied for time tp and the subsequent FID is recorded for time t_{acqu}. To collect sufficient information, the processes is repeated several times and FIDs added together. The location of RF pulses is marked as tx (for a RF transmitter) and the NMR signal is detected by a receiver (rx). C) The Fourier transformation is a process which turns a time domain signal, the FID, into a frequency domain, the spectrum (adapted from (303).

There are two distinct types of relaxation, the spin-lattice or longitudinal relaxation (T1) and spin-spin or transverse relaxation (T2). A T1 is a process in which bulk magnetisation returns to alignment along the z axis. It is caused by the local magnetic fields fluctuating at the Larmor frequency of the spins and allow spins to flip between the energy levels until the magnetization reaches equilibrium. A T2 is caused by a loss of phase coherence between spins called transverse magnetization.

Along with the external magnetic field, nuclei also experience different local electromagnetic environments. Hence, nuclei of the same residue but in different chemical environments within the molecule will have different Larmor frequencies. This effect is called the chemical shift. NMR experiments rely on the concept that the spins between different nuclei interact, either through covalent bonds or by existing at close distances to each other. These interactions reveal information that can be used to extrapolate structural features. Hence, the chemical shift of nuclei is influenced by electron shielding, the magnetic anisotropy of neighboring groups and ring current effects. Electron shielding occurs when, in an applied magnetic field (B_0), electrons circulate and produce an electric current which opposes B_0. As a consequence, the effective field at the nucleus is reduced by the shielding effect. Electron density shields nuclei from the external field and the distribution of electrons is influenced by the electronegativities of covalently bound atoms. A nucleus in the vicinity of an electronegative atom experiences reduced electron density and the nucleus is therefore deshielded. Protons attached to more electronegative nuclei which are less shielded have higher Larmor frequencies and will appear at higher ppm values on the chemical shift scale. Furthermore, the distribution of environment around a nucleus is generally anisotropic, i.e. it is not spherical around a nucleus but varies on different sides of the nucleus. This means that the shielding and the chemical shift experienced by a nucleus is also anisotropic and depends on the orientation of a molecule with respect to the magnetic field. This is called chemical shift anisotropy (CSA). Nevertheless, in solution NMR, the anisotropy of the chemical shift is averaged out by rapid tumbling of a molecule through Brownian motion and a single isotropic chemical shift value is observed. In addition to these effects, chemical shift is also influenced by the ring current effects caused by aromatic rings in proteins. Under the influence of external magnetic field B_0, aromatic rings can create their own local magnetic fields which have strong effects on neighboring molecules. A ring current is generated by the delocalised π electrons of the aromatic ring which cause the magnetic field inside the ring to counteract the externally applied field. In contrast, the field generated outside the ring matches the direction of B_0 and reinforces it. As a result, protons within the plane of the aromatic ring and outside the ring resonate at higher frequencies while protons above and below the ring resonate at lower frequencies. This effect is used in molecular structure determination and in assessing protein

folding. The ring protons of aromatic residues in proteins are known to resonate around 7 ppm. In a folded protein aromatic residues are often buried within the hydrophobic core and packed against the aliphatic residues of which the methyl groups which are above or below the aromatic ring will resonate at characteristic lower frequencies (between 0 to -1 ppm).

In systems with two or more nuclear spins, which evidently include proteins, it is necessary to consider the energy state effects that neighboring nuclei have on each other. The α and β energy transition state of one nucleus influences the neighboring spin, causing an effect called a spin coupling, scalar coupling or J-coupling (**Figure 10 A**). The coupling is mediated through a chemical bond and it is useful in determining which nuclei are close to one another. The presence of these couplings generates multiplets in the spectrum. For example, if two ½ spins are coupled, the resonance from each spin splits symmetrically about the chemical shift into two lines called a doublet. Each doublet is split by an equal amount and the distance is called a coupling constant, J. If another coupling forms between one spin couple and the next spin, a multiplet is formed (**Figure 10 B**).

J-coupling and dipolar mechanisms are utilised in correlation spectroscopy to allow generation of the multiple dimensions in NMR spectra. This is beneficial in particular in protein characterisation where it helps to resolve spectral overlap arising from a huge number of spins.

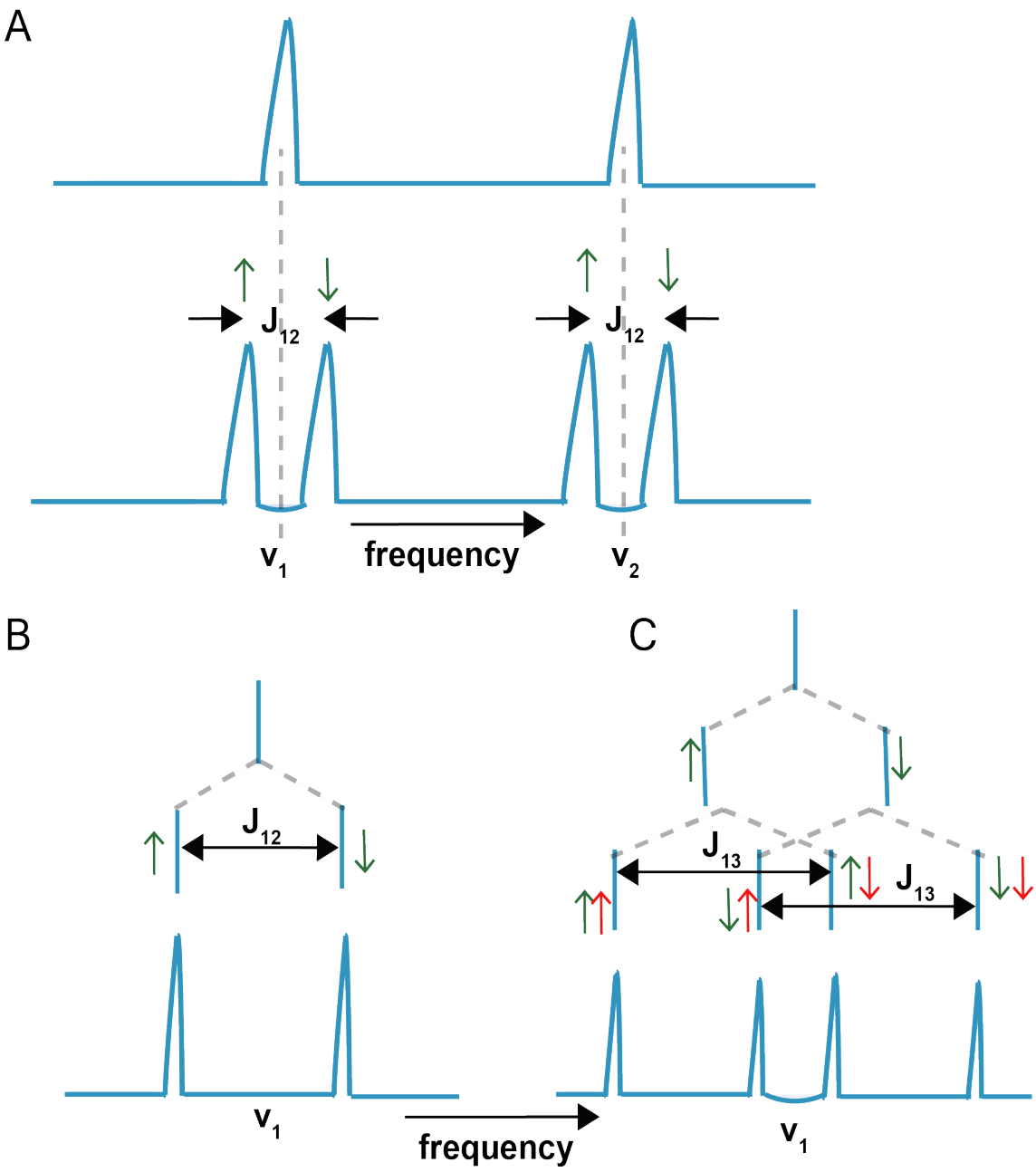

Figure 10: Scalar coupling between spins. A) Singlet peaks, at frequencies ν1 and ν2 from two different spins (top). Two doublets formed by the scalar coupling between peaks. The frequency splits symmetrically into two and gives a coupling constant of J_{12} (bottom). The green arrows indicate α (up) and β (down) energy states of the coupling. B) and C) Schematic representation of multiplet formation using tree diagrams. B) A doublet formed by the coupling of the first spin to a second spin with the coupling constant J_{12}. r coupling between peaks. The frequency splits symmetrically into two and gives a coupling constant of J_{12} (bottom). The green arrows indicate α (up) and β (down) energy states of the coupling. C) The coupling of the first spin to the third spin, with coupling constant of J_{13} causes each line of the doublet to split again. Subsequent formation of the multiplet is called a doublet of doublets. The green arrows indicate α (up) and β (down) energy states of the coupling and the spin states of the second spin are presented as red arrows, α (up) and β (down). The illustration was adapted from (303).

2.1.3 NMR in protein characterisation

NMR is an important tool for the structural biologist. It is a common experimental method used to characterise proteins, nucleic acids, and protein complexes. Since protein structure relates to function, such information can shed light on biological processes in health and disease states. NMR offers a broad range of routinely used 1D, 2D, 3D and even 4D experiments to carry out protein research. These experiments are used for the assignment of atomic chemical shifts, measurements of coupling constants, relaxation experiments, interaction studies and measurements of short proton-proton distances. Rapidly developing NMR research allows for determination of structures, dynamics, and interactions of larger and larger molecules including membrane proteins which are key drug targets.

2.1.3.1 1D NMR

A one dimensional proton NMR experiment (1D) provides a quick method which helps to judge whether a protein is folded and therefore likely to be functional. The basic 1D NMR experiment has been described in the previous section (**Figure 9**). The simplest experiment uses a single 90° pulse (for example applied along the x axis) and it also includes water suppression because the signal of the solvent is significantly stronger than the protein signals (**Figure 11**).

A typical 1D ^{1}H NMR protein spectrum contains many overlapping peaks because each peak corresponds to a particular proton in the protein. This means a spectrum of a protein of 100 residues could feature up to 1000 NMR peaks. The general concept is that if a small protein is properly folded, the peaks in the 1D NMR spectrum are sharp, narrow and well resolved on the chemical shift scale. However, if the protein is partially folded or self-associating, the NMR signals are broader and less dispersed. The main indicator of folded globular proteins is a good dispersion of the NMR signals in the regions of the methyl protons (–0.5 to 1.5 ppm), α-protons (3.5–6 ppm), and amide protons (6–10 ppm).

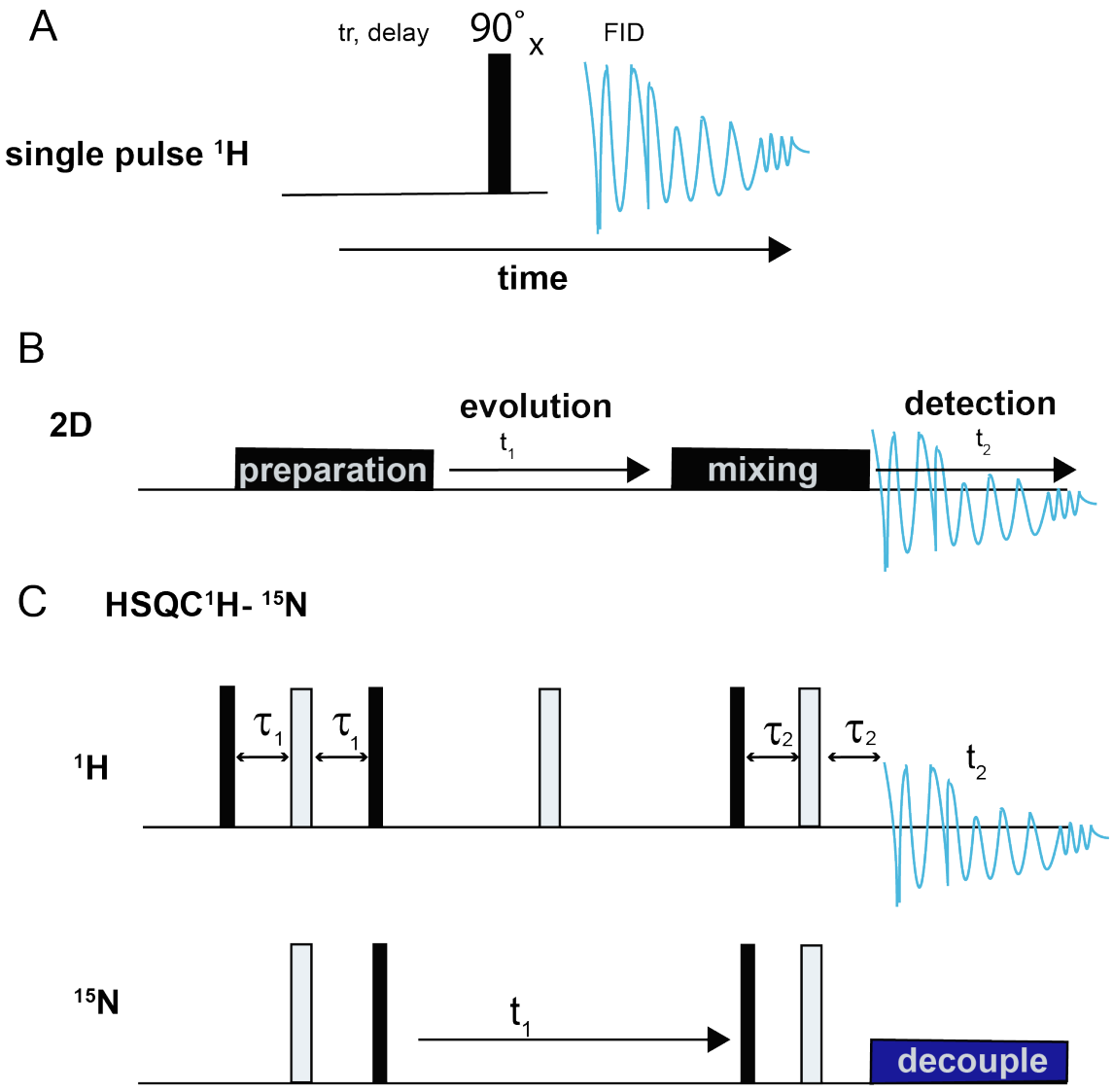

Figure 11: 1D and 2D NMR experiments. A) In a 1D ^{1}H NMR experiment, during the tr time the spins recover to equilibrium. The 90° pulse generates transverse magnetisation and the subsequent FID signal is sampled. B) The general scheme of 2D NMR experiment. The preparation period is followed by the evolution of the signal (t_1), a mixing period and a detection period (t_2). C) Pulse sequences for the HSQC experiment. The frequencies of ^{1}H and ^{15}N are correlated. The sequence starts with the equilibrium magnetization on the ^{1}H spin, which is transferred onto the ^{15}N spin in the Insensitive Nuclei Enhancement by Polarisation Transfer (INEPT) sequence. The ^{15}N magnetization evolves during t_1 with the centrally placed 180° pulse that refocuses chemical shift evolution and the magnetization is transferred back onto spin ^{1}H, where it is observed. Dark blocks correspond to 90° pulses and white blocks to 180° pulses. The illustration was adapted from (303).

2.1.3.2 2D NMR

Modern NMR spectroscopy routinely uses multidimensional experiments which involve generation of ^{1}H, ^{13}C, and ^{15}N nuclei in isotopically labeled proteins. 2D spectroscopy provides better resolution of NMR signals by preferentially selecting them and reducing the signal overlap. The typical 1D NMR experiment produces a spectrum that correlates an intensity of absorption with frequency, while in 2D experiments the intensity is linked to two frequency co-ordinates. It is common to use 2D experiments that correlate the shifts of different types of nuclei such as ^{15}N and proton or ^{13}C and proton. The heteronuclear single quantum coherence (HSQC) experiment is used to correlate ^{1}H nuclei with either ^{15}N nuclei or ^{13}C nuclei, with the proton being the observed nucleus. In this experiment, magnetisation is transferred from the first spin (^{1}H) to the second (^{15}N) through scalar coupling (also called J-coupling) using an INEPT sequence which is the building block of pulse sequences for 2D and 3D experiments. The chemical shift evolves on ^{15}N and then the magnetisation is transferred back onto ^{1}H, where it detected (**Figure 11**)**.** The resulting ^{1}H- ^{15}N-HSQC spectrum has peaks corresponding to all H-N groups, which are mostly the backbone amide groups in the protein and so it is called the protein fingerprint. Hence, this experiment is very useful for protein backbone assignments and to monitor chemical shift changes.

Another type of important experiment in structural studies is the 2D Nuclear Overhauser Enhancement Spectroscopy (NOESY). The Nuclear Overhauser Effect (NOE) used in these experiments relies on the magnetic dipole interaction in which magnetisation transfers in an effect of relaxation. This effect occurs throughout radiofrequency saturation of one spin, which then influences the intensity of other spins. The NOESY spectrum presents cross peaks which are generated by cross-relaxation between nuclear spins instead of coherence through couplings. Hence, NOESY correlates protons which are close in space (up to approximately 5 Å) rather than coupled through-bond. The cross-peak intensities of the correlated signals are equal to a nuclear distance of $1/r^{6}$. This NMR experiment provides useful information about the distances between atoms which can be used as restraints in protein structure calculations.

2.1.4 Chemical shift assignment

The assignments of chemical shifts of atoms within the protein is important in determination of protein structure by NMR or in characterisation of protein interactions. The first step involves a backbone assignment of amino acids within the peptide chain, followed by assignment of the side chains. Furthermore, the measurement of distances between protons, requires identification and assignment of the protons which give rise to the signals in the spectra.

2.1.4.1 Backbone assignment

Assignment of a protein backbone requires knowledge of the sequence of the peptide chain and involves determination of the chemical shifts of NH, Cα, Cβ, and C' atoms. Typical assignment involves an HSQC and triple resonance experiments, that correlate three nuclei (^{1}H, ^{15}N, and ^{13}C) and provide sequential connectivities between amino acids so that the resonances can be linked to particular residues and sequentially assigned. This requires preparation of uniformly isotopically double labeled proteins (^{13}C and ^{15}N) by recombinant expression in minimal media using $^{13}C_6$ glucose as the sole carbon source and $^{15}NH_4Cl$ (or $-SO_4$) as the sole nitrogen source.

The HSQC spectrum shows peaks of the correlated nitrogen and amide protons in the peptide bond of each residue of the protein, except for proline. However, in practice not every peak is always visible due to several factors such as exchange with solvent. Furthermore, sidechains with nitrogen-bound protons can also generate peaks and can be easily recognised in the spectrum due to their characteristic shifts. For example, tryptophan side-chain Nε-Hε groups are typically shifted downfield and can be observed in the bottom left corner of the spectrum. Asparagine and glutamine side-chain Nδ-Hδ2/Nε-Hε2 groups appear as doublets in the top right corner of the spectrum, and an additional smaller and distinctive peak may appear above each peak due to deuterium exchange from the D_2O typically added to NMR samples. The backbone amide peaks of glycine usually appear at the top of the spectrum. In addition, signals corresponding to Arg Nε-Hε can be observed, but as the Nε chemical shift is outside the region typically recorded,

the peaks are ‘folded’ which means that they appear as aliased ppm (often negative peaks). The HSQC spectrum of a folded protein shows well-dispersed peaks where single peaks can usually be distinguished. In contrast, a large area of not resolved and greatly overlapped peaks in the middle of the spectrum, indicates the presence of unstructured regions in the protein. Furthermore, the intensity of peaks varies with their flexibility and so the folded protein regions exhibit similar, fairly strong, peak intensities. In contrast more flexible regions, such as loops tend to exhibit lower or no peak intensities in the spectrum. However, highly flexible regions, such as the C-terminus show very high peak intensities and appear in characteristic, lower positions in the HSQC spectrum. The assignment of the HSQC spectrum, normally requires the CBCANH and CBCA(CO)NH experiments and often, especially in a case of large proteins, HNCO and HN(CA)CO, HNCA, HN(CO)CA spectra. These experiments work in pairs following the concept that e.g., the CBCANH correlates each NH group with the Cα and Cβ chemical shifts of residue ‘i’ (producing a strong peak), and of the preceding residue ‘i-1’ (producing a weak peak). While, CBCA(CO)NH only correlates the NH group to the Cα and Cβ chemical shifts of the preceding residue i-1. This allows the identification of the sequential links of one NH group with the next one within the protein sequence (**Figure 12**). Furthermore, the Cα and Cβ chemical shifts adopt values characteristic for each particular amino acid, which is advantageous for the backbone assignment. For example, residues such as alanine, serine and threonine can be easily distinguished because their Cβ chemical shifts are very distinctive, while Glycine has no Cβ. The Cα chemical shifts of Valine, Isoleucine and Proline exhibit lower chemical shifts compared to those of other amino acids. The pair of HNCO and HN(CA)CO experiments show peaks of correlated NH group of each i residue with the C' of i-1 in the case of (HNCO) or, with C' of both, i and i-1 in the case of HN(CA)CO (**Figure 12**). The HNCO experiment is more sensitive than the CBCANH and is often used to resolve ambiguous assignments from the CBCANH data. Other experiments often used in assignments include HNCA/HN(CO)CA pairs.

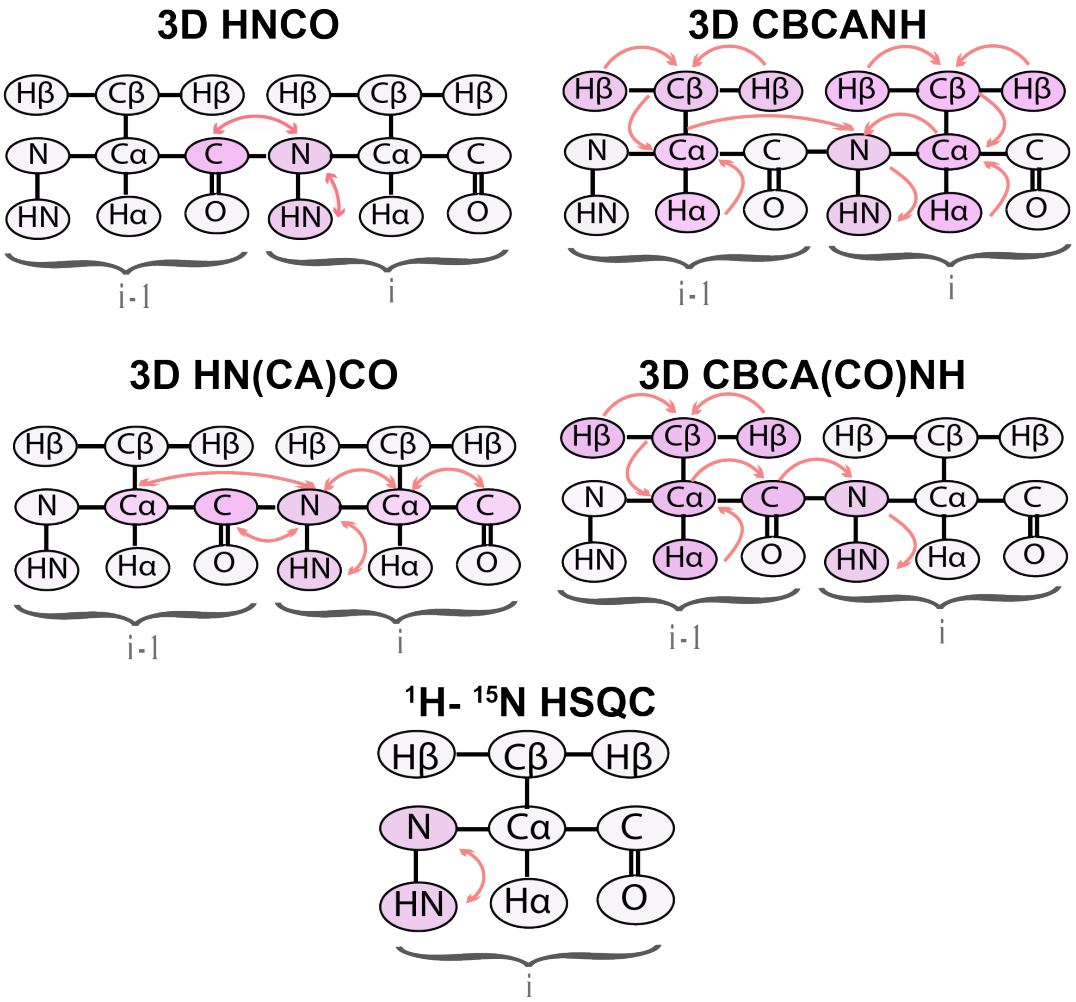

Figure 12: The basic scheme of 3D experiments used in a backbone assignment indicating the direction of magnetization transfer. In the HNCO, HN(CA)CO and HSQC experiments, the magnetisation originates from an amide proton and, via J-coupling, is then transferred to the correlated nucleus for evolution, which may be ^{15}N or ^{13}C, via an INEPT. Next it is transferred back to the proton and the signal is then recorded. In these experiments, the ^{1}H signal is detected in the directly measured dimension, while the chemical shift of ^{15}N or ^{13}C is recorded in the indirect dimension which is produced from the series of experiments. CBCANH and CBCA(CO)NH are recorded differently and the magnetisation originates at $^{1}H\alpha$ and $^{1}H\beta$ and is transferred to $^{13}C\alpha$ and $^{13}C\beta$, respectively, followed by transfer from $^{13}C\beta$ to $^{13}C\alpha$. It is then transferred first to $^{15}N^{H}$ and then to $^{1}H^{N}$ for detection. In CBCANH spectra, for each NH group, there are two Cα and Cβ peaks visible while in CBCA(CO)NH spectra there is only one peak of each.

The typical assignment strategy is shown in **Figure 13 A** and involves using a uniformly $^{13}C^{1}H$-labelled protein and the random coil chemical shifts of Cα, Cβ, Cγ and Cδ as reference shifts (304).

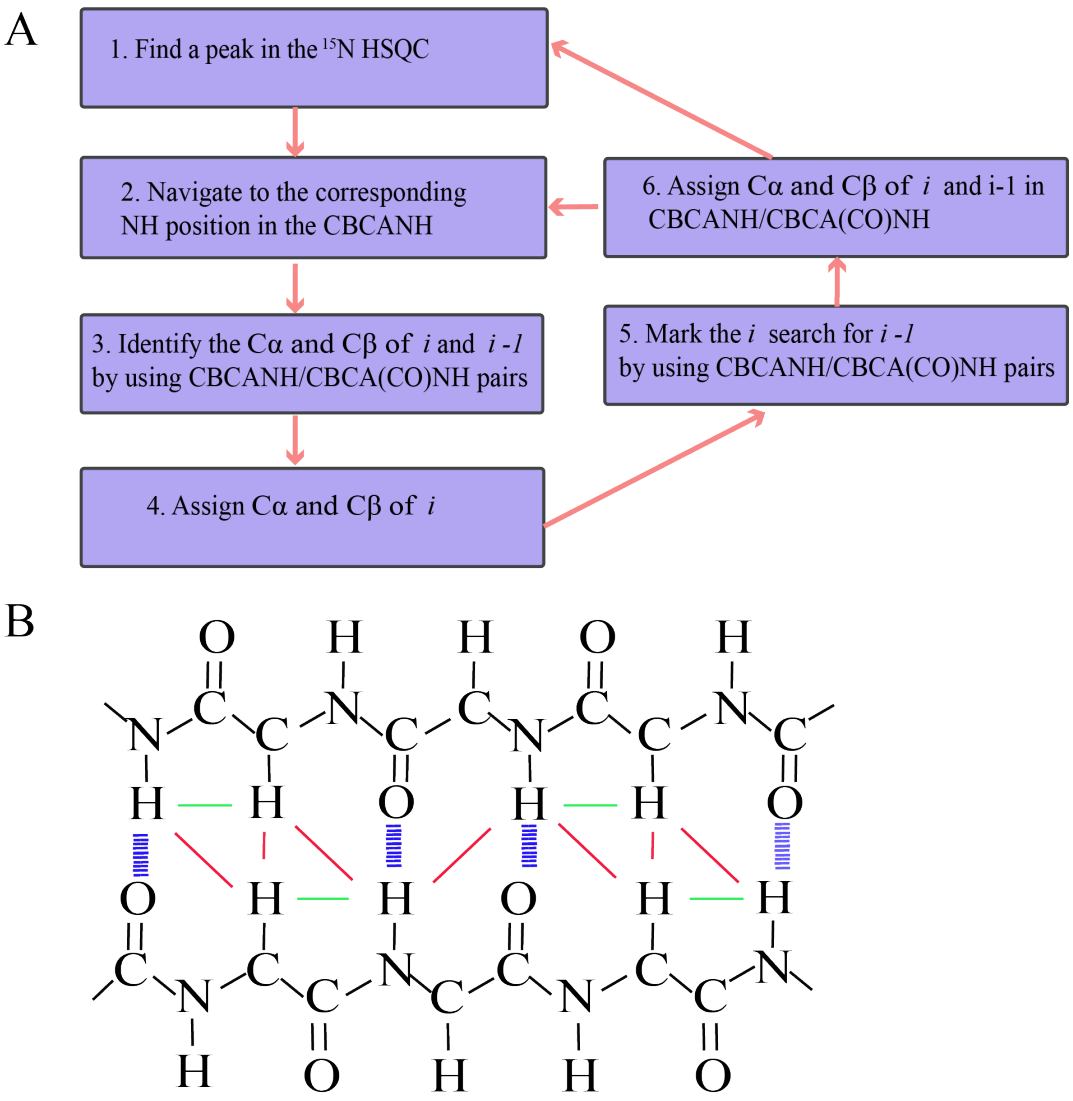

Figure 13: Strategies for sequential assignment. A) Step by step backbone assignment. B) An example of characteristic distances of which NOEs are generated in antiparallel strands of a β-sheet. β-sheet specific NOEs are in red and simple sequential NOEs in green.

Interestingly, it is possible to predict secondary structure of the protein and dihedral φ and ψ angles directly from the chemical shift backbone assignment due to the strong correlation of chemical shifts with local structure. Chemical shift indexing (CSI) identifies the patterns of stretches of α-helix or β-strand by analysis of their chemical shifts. For example, Cα atoms in α-helices, typically have positive secondary chemical shifts and Cα atoms in β-strands are likely to have negative secondary chemical shifts. In contrast, Cβ in β-strands are characterised by positive secondary chemical shifts and negative in α-helices. Furthermore, the chemical shifts of Hα H^N, N^H and C' are also correlated with secondary structure, although to a lesser extent. The

Ramachandran plot of the φ and ψ torsion angles specifies that the secondary structure of a protein depends on the sequential residues comprising similar φ and ψ torsion angles, with a particular area of the plot being characteristic either for α-helices or for β-strands.

2.1.4.2 Sidechain assignment

A common experiment in protein sidechain assignment is a total correlation spectroscopy (TOCSY), which correlates all protons of ^{13}C nuclei. The TOCSY experiment allows for the transfer of magnetization and correlation of ^{13}C nuclei and all protons of protein sidechains. Hence, the spectra may become crowded due to a great amount of information especially in large proteins and may require additional experiments such as NOESYs. The assignments for sidechain resonances in the experimental part of this thesis were achieved via a combination of HBHA(CBCACO)NH [also called HBHA(CO)NH] and 3D HCCH-TOCSY, such as carbon observed (H)CCH-TOCSY and proton observed H(C)CH-TOCSY (**Figure 14).** The HBHA(CBCACO)NH experiment provides chemical shifts of $^{1}H\alpha$ and $^{1}H\beta$ protons and correlates their resonance to the following NH group. This experiment is analogous to the previously described CBCAC(CO)NH in which the magnetisation is transferred from $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$, respectively, and then from $^{13}C\beta$ to $^{13}C\alpha$. Next the magnetisation is transferred to ^{13}CO, and then to $^{15}N^{H}$, followed by the final transfer onto $^{1}H^{N}$ for detection. While, the chemical shift is not evolved at any of the ^{13}C nuclei, it is evolved on the $^{1}H\alpha$ and $^{1}H\beta$, and also the $^{15}N^{H}$ and $^{1}H^{N}$. The subsequent 3D spectrum possesses one nitrogen and two hydrogen dimensions.

Generally, the practice of sidechain assignment involves using the Cα and Cβ chemical shifts assigned during the backbone assignment, and navigating to the corresponding peak in the HCCH-TOCSY. HBHA(CO)NH can be used to confirm the $^{1}H\alpha$ and $^{1}H\beta$. This allows identification of Hα, Hβ, Hγ and Hδ chemical shifts linked to ^{13}C shift which have peaks corresponding to the same H ppm values in this 3D experiment. It is also convenient to use an average shift available for the observable sidechain protons as reference.

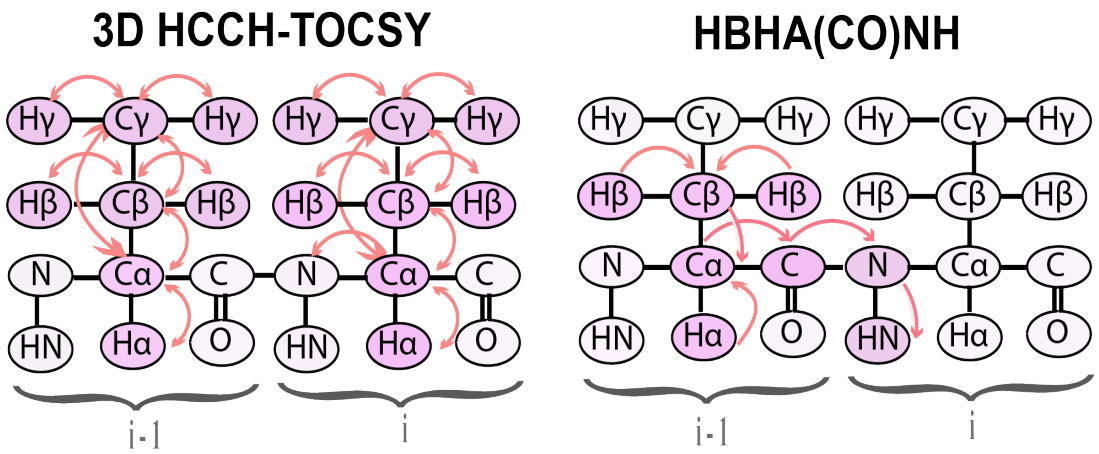

Figure 14: The basic scheme of 3D experiments for sidechain assignment indicating the direction of the magnetisation transfer. In the TOCSY experiment the magnetisation transfers from the hydrogen nuclei of sidechains to their corresponding ^{13}C nuclei (i.e., Cβ, Cγ) . After the subsequent isotropic ^{13}C mixing, the magnetisation is transfered back to the hydrogen atoms of sidechain for detection. In HBHA(CO)NH the magnetisation is transferred from ^{1}Hα and ^{1}Hβ to ^{13}Cα and ^{13}Cβ, respectively, and then from ^{13}Cβ to ^{13}Cα. This is followed by magnetisation transfer to ^{13}CO and then to $^{15}N^{H}$ and $^{1}H^{N}$ for detection. In order for the ^{13}Cα and ^{13}Cβ to appear in the same dimension, the chemical shift is evolved simultaneously on both nuclei. The magnetisation does not evolve on ^{13}CO.

2.1.4.3 NOE assignments

NOESY experiments provide information about distances between protons (NOE distance restraints), which is used in protein structure calculations. A crosspeak appears if protons are close in space and the intensities of the correlated signals are proportional to $1/r^6$, where r is a internuclear distance. NOESY experiments are often used in both backbone and sidechain assignments. NOE distance restraints in the experimental part of this thesis were derived from the most commonly used NOESY experiments such as 2D NOESY and a pair of 3D experiments: ^{15}N-edited NOESY-HSQC and ^{13}C-edited NOESY-HSQC (**Figure 15**).

The 2D NOESY spectrum is typically used in assignment (especially of aromatic residues) and to provide structural restraints. It shows signal for all protons which are in close proximity to one another. While for very large proteins the aliphatic region of the spectrum tends to be very

crowded, the NH and aromatic region provides valuable constraints for methyl and aromatic hydrogens. In addition, the edge of the spectrum often contains some well isolated methyl peaks. The 3D NOESY experiments are also used for both, to obtain restraints for structure calculations and for protein assignment. Each NOE strip of ^{15}N- NOESY-HSQC spectrum contains NOE peaks linking one NH group with all other hydrogen atoms close in space. Valuable hydrogens peaks can be visible in the aliphatic and in the amide and aromatic regions. In addition, each strip of ^{13}C-edited NOESY-HSQC contains NOE signals connecting one CH group to all other neighboring hydrogens. In this spectrum, a double set of peaks usually occurs because a CH_2 group has two hydrogens with different chemical shifts. ^{13}C-edited NOESY-HSQC is usually recorded in 100% D_2O to reduce the signal of noise and artifacts.

In practice the first step of NOEs assignment is to find the sequential neighbours to each residue using the NOEs between their NH groups. The next step is to determine the correct sequential order of the neighbouring residues using NOEs between the NH group of one residue and the sidechain of the other. Once several residues are connected, it is possible to identify where they are within the sequence. Furthermore, α-helices and β-sheets have slightly different patterns of NOEs and once identified, can help with the assignment. For example, characteristic cross peaks in a NOESY spectrum between protons connecting two strands in a β-sheet facilitate the assignment and building of two neighbouring strands in parallel **Figure 13 B**. NH-NH distance is around 2.8 Å in helices and 4.2 Å β-sheets. Additionally, the intensities of peaks are useful in identifying sequential and structural links. The NOEs linking sequential residues are the strongest and cover a short range distance NH (1.8-2.5 Å) while additional NOEs which can be intra- or interresidue are of medium range NH (1.8-3.5 Å) or long-range NOEs NH (1.8-5.0 Å). Especially in α-helical secondary structure it is possible to observe medium and long range NOEs and these are weaker and can be useful in linking several sequential residues. In β-sheet secondary long range NOEs connect to the neighbouring strand which may be of comparable strength to the short range NOEs in a parallel β-sheet or stronger in an antiparallel β-sheet. Generally short range NOEs are important in identifying elements of secondary structure while the long range NOEs

provide crucial information about the tertiary structure of a protein. Furthermore, to decide the direction in the sequential assignment it is useful to know that the Hα of residue i is usually much closer to the NH of the following residue i+1, than the NH of the preceding residue i-1. It is also practical to use a reference of hydrogen chemical shifts to compare the shifts with the average hydrogen chemical shifts. Some amino acids such as alanine, glycine and threonine have unusual patterns of sidechain hydrogen chemical shifts which can be easier to identify. However, not all resonances are always visible in the spectra. Overall, the more unambiguous NOE restraints are assigned, the more precise is the structure. Typically, there are about 15-20 NOE restraints including about 3.5 long range distance restraints per residue in a high-resolution NMR structure. In addition, there are some special NOESY experiments which are often used in structural studies of protein complexes. These include filtered, edited NOESYs which are based on selection of NOEs from two molecules with unique labelling patterns. They require a sample comprising a mixed labelled and unlabeled ligand and allow observation of signals which are only present at the binding interface.

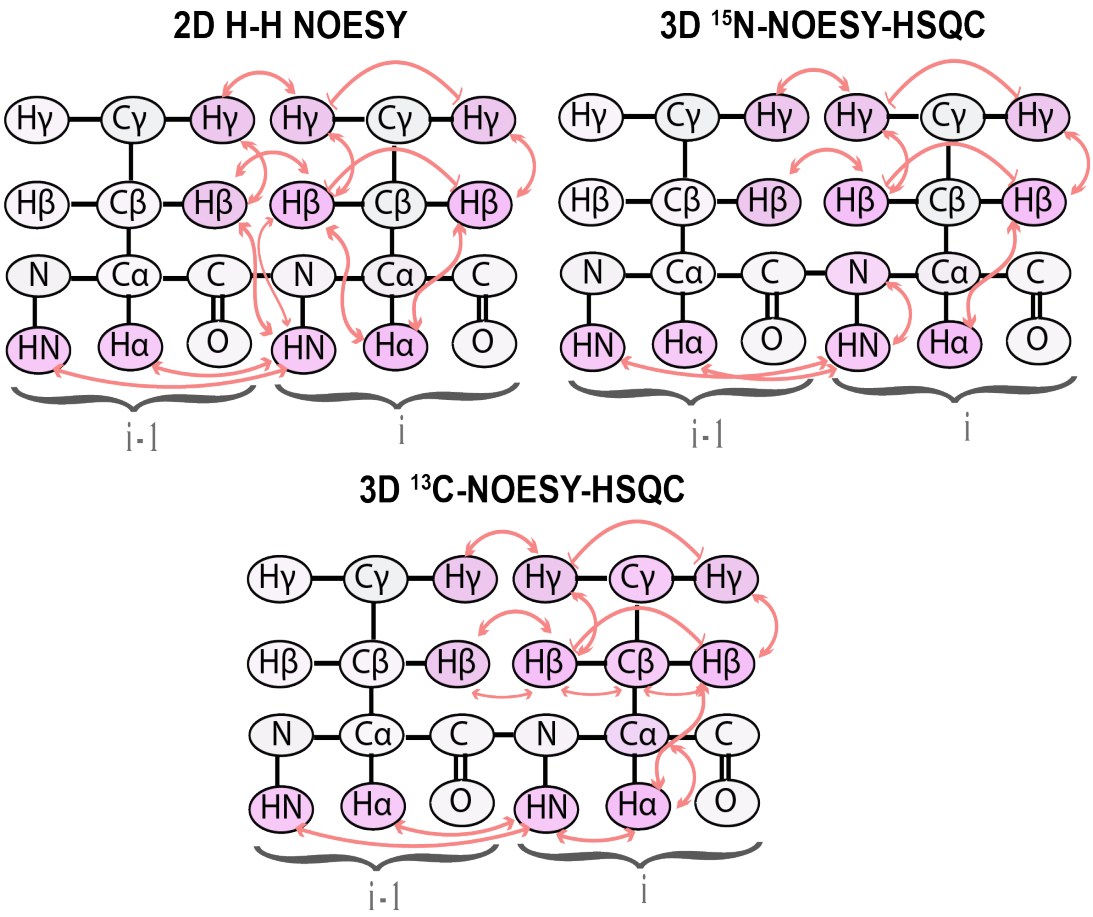

Figure 15: The basic scheme of 3D experiments of NOE assignment indicating the direction of magnetisation transfer. In 2D NOESY experiments, the magnetisation transfer occurs first between all hydrogen nuclei using the NOE. In ^{15}N-edited NOESY-HSQC the magnetisation is subsequently transferred to neighbouring ^{15}N nuclei and back to ^{1}H for detection. In ^{13}C-edited NOESY-HSQC the magnetisation is subsequently transferred to neighbouring ^{13}C nuclei and back to ^{1}H for detection. The direction of transfer depends on the ^{13}C frequency used in the pulse sequence and occurs to/from the aliphatic ^{13}C nuclei or to/from the aromatic ^{13}C nuclei (but not both).

2.1.5 Structure calculations

Following the resonance assignment, the NMR data (through space correlation of NOE intensities and through bond correlation of J-couplings and chemical shifts) need to be converted into distances and torsion angles to generate the structural constraints which then can be used in 3D structure calculations. The torsion and dihedral angles for structural calculations are provided from the assignments and calculated by TALOS+ (305). Initially, only unique and unambiguous NOEs are assigned and proton-proton distance restraints are applied in structural calculations

using a restrained molecular dynamics (MD) or simulated annealing (SA). Hence, calculation of structures involves computational heating/cooling of structures to attempt to find a global energy minimum (in SA) or to satisfy structural restraints and determine the initial structure (in MD). This is followed by many cycles of finding and evaluating the violated NOE restraints in an iterative process. The starting set of structures can be used to filter ambiguous assignment. After completion of the assignment of the remaining NOEs, the water refined ensemble of structures is obtained which typically spans fairly narrow conformational space. Hence, the quality (resolution and precision) of a NMR structure can be judged by the root mean square deviation (RMSD) of each structure from this ensemble to describe how similar these structures are. The RMSD for both backbone and all heavy atoms are usually analysed and the high-resolution structure will have a backbone RMSD ≤ 0.8 Å and heavy atoms RMSD ≤ 1.5 Å. Sometimes only the RMSD of more ordered regions are reported. The NOE assignment is usually accomplished by a combination of automatic and manual methods. The structure calculations are performed using available software such as ARIA (Ambiguous restraints for iterative assignment) (306), CYANA (Combined assignment and dynamics algorithm for NMR applications) (307) or Rosetta (308). Furthermore, the structure is validated by inspecting the stereochemical quality of the protein including backbone phi and psi-angles favoured in a Ramacharndran plot and low deviation from idealised covalent geometry. These statistical parameters are then compared to other deposited structures using various software such as WhatIf or PROCHECK.

2.1.6 Study of protein-protein interactions by NMR

NMR is commonly used in studying protein-protein and protein-peptide interactions. This method is very sensitive and allows easy detection of an apparent change in signal between the free states and the complexes. In NMR, observable parameters such as the chemical shift, relaxation rate and J-coupling may all change upon binding of a ligand (e.g., another protein). Chemical shifts are sensitive to changes in the local magnetic field of a nucleus and can be perturbed by both changes in covalent structure and by non-covalent interactions with binding partners. Therefore, chemical shifts are well suited to identify binding interfaces of interacting protein complexes. Chemical

shift mapping or chemical shift perturbation (CSP) studies usually involve ^{1}H-^{15}N correlational spectroscopy experiments such as HSQC and HMQC. The ^{1}H-^{15}N HSQC is a sensitive experiment that generates well resolved peaks corresponding to each amino acid in the protein. In a good quality spectrum, it is easy to follow any peak changes caused by addition of a ligand. Typically, the perturbations of chemical shifts are monitored in a series of ^{1}H-^{15}N HSQC experiments in which increasing amounts of a ligand (e.g. a peptide) is titrated in to a protein. The resulting CSPs belong to amino acids close to the binding interface. However, if the protein undergoes a conformational change upon complex formation this will reflect in widespread CSP, which can be away from the binding site.

The effect on chemical shifts upon addition of a ligand depends on the type of chemical exchange that affects the nuclei. Overall, there are three exchange regimes: slow, intermediate and fast exchange. Fast exchange occurs in weak interactions (in the mM range) and the chemical shifts change continuously throughout the addition of a binding partner and until the binding site is saturated. These peaks usually shift in a straight line which makes the changes easy to identify. Since this interaction is weak, the rate of exchange between the state of a free and bound nucleus is high. Therefore, the observed single peak corresponds to the weighted average chemical shift in the NMR spectrum for the free protein and the complex. The intermediate regime, is characteristic in stronger interactions (μM range) in which protein resonance peaks may become weaker, then disappear and reappear again when increased amounts of a binding partner is added. Sometimes peaks may completely disappear from the spectrum. This effect is caused by the difference in the populations between the two (free and bound) conformations occurring during the recording of the NMR experiment. This can also result in line broadening and produce peaks so broad, that they are not observed in an NMR spectrum. In a slow exchange regimen, the binding between interaction partners is very strong (nM range) and results in appearance of new resonance peaks. While the intensity of the newly formed peaks increases upon the addition of a binding partner, some original peaks (free protein) gradually decrease in intensity until they disappear. This effect occurs because the binding is very strong and so the exchange rate between the two

conformations is also slow and therefore both, free and bound protein states can be observed simultaneously.

In interaction studies between two proteins, it may be beneficial to isotopically label each of them and perform reciprocal chemical shift perturbation experiments. This allows comparison of the binding effect in both proteins and provides information on possible conformational changes.

2.2 Protein Crystallography

In crystallography the structural information is obtained from electrons. A single protein crystal is placed into a focused X-ray beam causing the electrons to oscillate in the electric field and scatter the X-rays. They produce diffraction images that are recorded on a detector. The electron density corresponding to the atomic structure of protein molecules within the crystal is reconstructed by Fourier transform techniques of the diffraction data and the atomic model is built into a three-dimensional electron density map. Over the past 100 years crystallography has grown significantly and the resolution dramatically improved reaching even 0.48 Å for some structures deposited in the protein data bank. High resolution structures (<1.5 Å) can provide additional information such as the positions of hydrogen atoms, distributions of valence electrons and orientations of bound waters, that is valuable in establishing the dynamic and chemical properties of proteins. In this section I will begin by briefly outlining the fundamental concepts of crystallography and then I will focus on the practical aspects and stages of solving protein structure by molecular replacement.

2.2.1 Fundamental concepts in X-ray crystallography

2.2.1.1 Crystal geometry and symmetry

Protein crystals have the ability to diffract X-rays to produce valuable information about their atomic structures. The typical X-ray wavelength used in laboratories varies between 1.6 Å and

0.5 Å and uses molybdenum or a copper as a beam source. Protein crystals are formed from many copies of the same molecule organised in an ordered fashion. The protein molecules repeat themselves in a self-assembled crystal lattice that can be arranged in many different ways. However, the number of arrangements the molecules can pack into is limited. The collection of all symmetry elements of a repeating pattern in a protein crystal is called a space group. There are 65 possible chiral space groups in which proteins can construct their three-dimensional crystal structure. These different arrangements are possible through combination of symmetry operations such as translation and rotation. The repetitive order of the crystal lattice allows identification of positions that are equal. These positions can be marked with an imaginary point, called a lattice point. In two dimensions, four neighbouring lattice points form a lattice plane. In three dimensions, the neighbouring lattice points can be connected to form a box called a unit cell. In a crystal, unit cells are stacked on a top of each other in every direction. Crystal contacts occur when intermolecular interactions connect the molecules within the crystal. These repeat with the same periodicity as the lattice. The unit cell is a basic building block of a crystal lattice that can be translated up, down or sideways through the lattice without changing the content. Since there is often more than one possible choice of selecting the lattice points and the unit cell, the general practice is to choose the smallest unit that exhibits the maximum symmetry of the structure. Furthermore, within the unit cell even smaller elements can be distinguished, known as the asymmetric unit. It is described as the smallest repeat unit from which the crystal is formed. Each unit cell has its parameters defined including the axes a, b and c and, connecting them, angles α, β and γ which then form the faces A, B and C (**Figure 16 A** and **B**).

Considering the organisation of unit cells within a crystal lattice the crystal geometries are categorized into 7 basic crystal systems and 14 Bravais lattices. The seven crystal systems are cubic, tetragonal, orthorhombic, hexagonal, triclinic, monoclinic and triclinic (**Table 2**). The different unit cell lattices include P, primitive, F-face-centred, C-side-centred, I, body-centred. In the primitive lattice, the lattice points lie only at the corners of the unit cell. In the face-centred lattice, the lattice points lie at the corners of the unit cell and one lattice point lies in the middle

of every face of the unit cell. In the body-centred lattice, the lattice points lie at the corners of the unit cell and one lattice point lies in the middle of the unit cell. In the one side-centred lattice, the lattice points lie at the corners of the unit cell and also one lattice points lies at each of the two relevant faces. The 14 Bravais lattices are defined based on the variations of the parameters of the unit cell and different lattice points (**Table 2**). For example, one of the most common crystal systems, orthorhombic, is characterised by all different lengths of the axes (a, b, c), although all angles are of 90°. The types of Bravais lattices occurring in the orthorhombic system are: the primitive, the body centred and face centred (**Table 2**).

Table 2: The seven crystal systems and fourteen Bravais lattices (309).

Crystal system	Cell parameters	Lattice types
Cubic	$a = b = c$, $\alpha = \beta = \gamma = 90°$	P, I, F
Tetragonal	$a = b \neq c$, $\alpha = \beta = \gamma = 90°$	P, I
Orthorhombic	$a \neq b \neq c$, $\alpha = \beta = \gamma = 90°$	P, I, F, C
Hexagonal	$a = b \neq c$, $\alpha = \beta = 90°$, $\gamma = 120°$	P
Trigonal	$a = b \neq c$, $\alpha = \beta = 90°$, $\gamma = 120°$	P
Monoclinic	$a \neq b \neq c$, $\alpha = \gamma = 90°$, $\beta \neq 90°$	P, C
Triclinic	$a \neq b \neq c$, $\alpha \neq \gamma \neq \beta \neq 90°$	P

Within a unit cell, each asymmetric unit is related by symmetry. The lack of symmetry of the asymmetric unit poses limitations on its position within a unit i.e., it cannot be located at the symmetry elements. Hence in protein crystallography, non-crystallographic symmetry exists when more than one identical molecule is present in the asymmetric unit like in a case of dimers and higher oligomeric structures. There are two types of symmetry element in a crystal lattice,translational and non-translational. Translational symmetry which must be present for a crystal to diffract an X-ray beam and produce reflections is characterised by translation, that is moving an object in a direction (a,b,c). In this symmetry, rotation or reflection combined with

translation gives rise to screw axes (as opposed to simple rotation axes) and glide planes (as opposed to mirror planes), respectively. Screw axes combine a simple rotation with a translation along the direction of the axis. This type of symmetry cause absences in a diffraction data. Systematic absences are the absent reflections in diffraction data that appear when the observed intensity is zero. Since, asymmetric chiral protein motifs limit possible symmetry operations of proteins, mirror planes or glide planes are not allowed. Non-translational symmetry does not cause absences in diffraction data and the symmetry operations include the inversion centre, reflection, rotation and rotation-inversion.

Miller indices are part of the notation system that defines the orientation of a plane or a set of parallel planes of atoms in crystal lattices. These indices describe the position of a face or internal plane of a crystal and are determined on the basis of the reciprocal of the intercept of the face or plane on the crystallographic axes. The planes are determined by three integer values *h*, *k*, and *l*, referred as the Miller indices. The integers are written in brackets (hkl), and define the Cartesian coordinates or the axes of the unit cell, where *h* corresponds to the *x-axis* or *a* direction of the unit cell, *k* corresponds to the *y-axis* or *b* direction and *l* corresponds to the *z-axis* and *c* direction. Furthermore, if the intensities corresponding to a specific Miller index (hkl value) is zero, absences in diffraction data occur.

While there are several different methods of describing crystal geometry, the one that is preferred by crystallographers for practical reasons is the Hermann–Mauguin notation. This system, commonly used to represent the symmetry elements in point groups, plane groups and space groups, includes translational symmetry elements and the directions of the symmetry axes. The first letter of the Hermann–Mauguin notation represents the type of lattice and the following three notations indicate the type of symmetry operation or operations along the corresponding axes *x,y z*. For example, the orthorhombic space group referred as $P2_12_12_1$ indicates that the lattice is of a primitive type with a 2_1 screw axis along each of the *x, y, z* directions. A 2_1 screw axis indicates a

180° rotation followed by a translation upwards of half the unit cell. On the other hand, a space group $P2_1$ belongs to a monoclinic crystal system with a primitive lattice type while it also has a similar screw axis 2_1. For comparison, the P222 space group indicates a primitive lattice with two-fold rotations along each of three axes.

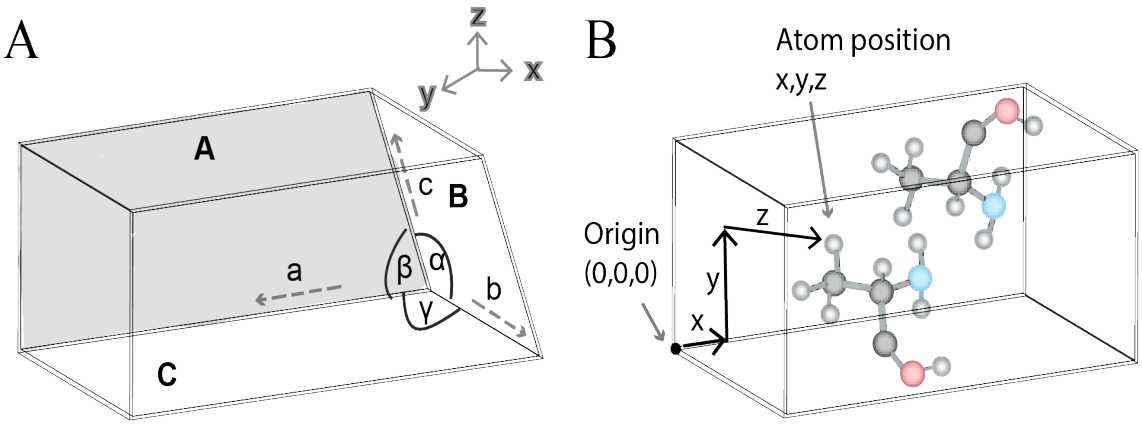

Figure 16: Unit cell. A) the parameters of the unit cell. The axes, angles and faces of the unit cell. A faces are the front and the back of the unit. B faces are the left and the right side of the unit. C faces are the top and the bottom of the unit. The unit cell volume, V = a • (b x c). B) Within one unit cell, the position of an atom can be specified by a set of spatial coordinates x,y,z (309).

2.2.1.2 Theory of X-ray diffraction by a crystal

In crystallographic experiments, copper, molybdenum, silicon and germanium are the most common sources of X-rays. The X-ray generators may include X-ray tubes, rotating anodes and synchrotrons. Synchrotron radiation provides very high energy X-rays by bending the electron paths as they move through a storage ring of bending magnets. Initial X-rays are generated when accelerated electrons move from excited to relaxed states, emitting electromagnetic radiation (**Figure 17 A**). The excited state of electrons can be caused by a collision with a photon. The total amount of energy emitted corresponds to the energy difference between the excited and ground states and can be represented as a function of frequency: $E = hv$, where E is the energy emitted, h is Planck's constant and v is the frequency of the wavelength.

As discovered by Laue and Bragg, X-rays are electromagnetic waves and, as a consequence, get diffracted by the periodically arranged atoms in a crystal. During an experiment, the X-rays enter a single crystal, and interact with oscillating electrons. This cause the beams to be scattered in all directions which forms a diffraction pattern of spots (also called images) on the detector. The intensities and angles of these beams are recorded as the crystal is gradually rotated. Spot are called reflections, because they correspond to the reflection of X-rays from a set of parallel planes passing through lattice planes. To define the plane, three numbers referred to as indices (*hkl)* are used to specify its orientation with respect to the rims of the unit cell.

Bragg's law introduced a simplified interpretation of X-ray diffraction that helps to understand how diffraction patterns and spots are generated in an incident X-ray beam (**Figure 17 B**). The scattering diagram can be interpreted as the reflection or diffraction of X-rays on a set of lattice planes in a crystal. The planes and lines are separated by a distance *d.* When the incident X-rays collide with the crystal lattice at specific points, the X-rays are reflected out from these points generating diffraction spots (**Figure 17 B**). Bragg's equation indicates that diffraction spots appear when $2d\sin\theta = n\lambda$, that is the incident X-ray wavelength is twice the *d* distance (also called spacing or d_{hkl}) multiplied by the sine of the diffraction angle (**Figure 17 B**). Individual diffraction spot is the sum of the diffraction from Miller planes and so can be defined with a *hkl* value that corresponds to Miller indices. Furthermore, each spot can be characterised by an intensity value that depends on the strength of the X-ray diffraction. Some spots are missing if the intensity is zero and these are called absences. Thousands of spots are mathematically reconstructed to produce a diffraction map and subsequently an electron density map. Electrons that scatter the X-rays in an atom are not steady but move around the nucleus in defined orbits. The amplitude of scattered wave depends on how many electrons scatter from a particular position in an atom, which is described by the electron density.

The reconstruction or processing of the sets of diffraction data produces the reciprocal form (**Table 3**). It relies on the concept that for each set of Miller planes (*hkl*) in real space there is a point in reciprocal space and the collection of all of these points can be defined as the reciprocal space (**Table 3**). In X-ray crystallography, the intensities of points are experimentally measured in reciprocal space during data collection. These data are used to generate an electron density map from a crystal in real space which then can be used to solve a protein structure (**Table 3**).

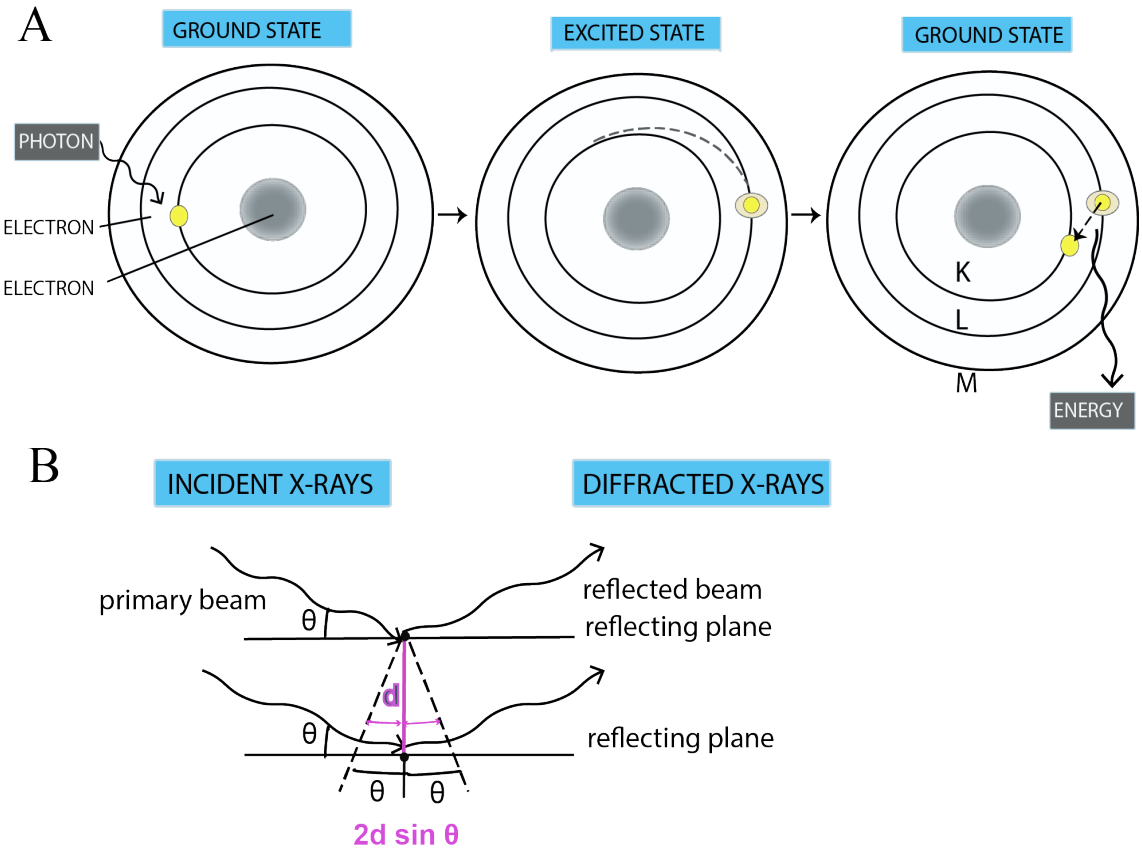

Figure 17: X-ray diffraction. A) Emission of X-rays. When a photon travelling as a wave collides with an electron in an atom, the electron is promoted into an excited state. X-rays are generated when an excited electron relaxes back into its ground state. This causes energy emission that includes X-ray radiation. K, L, M are the atomic shells between which the electron moves and, in an X-ray diffraction experiment, the most valuable energy comes from an electron moving from the L to K shell. This wavelength is called a 'characteristic X-ray'. B) Diffraction according to Bragg's law. The interpretation of scattering as relates the scattering angle θ to the distance d for the set of reflecting planes hkl.

Furthermore, the Ewald sphere introduced by Paul Ewald, (310) demonstrates the relationship between the wave vector of the incident and diffracted x-ray beams, the diffraction angle for a given reflection and the reciprocal lattice of the crystal. It allows us to visualize the properties of Bragg's law and it shows that the form of a diffraction condition: $d^*_{hkl} = 2\sin\theta/\lambda$, where d^* is a distance in a reciprocal space, can be effectively visualised by a geometrical construction. The relationship between the Ewald sphere and Bragg's law (**Figure 18**) shows that diffraction can only occur when a reciprocal lattice point intersects the Ewald sphere (**Figure 18 B, C and D**). Moreover, the detector is flat and therefore the reciprocal lattice points near the rotation axis never pass through the Ewald sphere and hence they contribute to the 'blind region'. In practice, the construction of the Ewald sphere predicts which diffracted waves satisfy Bragg's law, when the information of the beam wavelengths, the angle, the details of the lattice, the distance from detector and the orientation of lattice relative to the detector is provided. Furthermore, the position and intensity of each reflection (spot) in a diffraction pattern, corresponding to the angle at which it reached the crystal, is used to provide three-dimensional stryctural information of proteins, such as atom positions, bond lengths and angles, torsion angles and non-bonded distances . Also, the absences in crystal data relate to lattice types and symmetry elements. They can be used to determine the space group of a crystal lattice as the absences are specific for a particular group (except for a primitive lattice). There are two types of absences in diffraction patterns: general absences that are linked to lattice types and systematic absences, linked to translational symmetry. General absences appear on a diffraction pattern due to destructive interference of X-rays (as opposed to constructive) for non-primitive lattices (I, F, C). The rules characterizing these absences affect all reflections in the diffraction pattern (for all *hkl* values). Systematic absences also appear due to destructive interference of X-rays in relation to specific translational symmetry elements (i.e., screw axes). Therefore, systematic absences are found only in diffraction patterns for translational symmetry. Additionally, intensities in crystal data are associated with the structure factors. X-rays are waves that are scattered (diffracted) by electrons and share the properties of waves such as amplitude, wavelength and the phase. Therefore, a structure factor is applied to describe each reflection including its position, intensity and phase.

Table 3: Correlations between the real space and the reciprocal space (309).

Real space	**Reciprocal space**
Crystal structure	Diffraction pattern
Electron density (atomic parameters)	Amplitudes and phases (X-rays)
Crystal lattice and unit cell	Reciprocal lattice
Coordinates (x, y, z)	Coordinates (*h, k, l*)

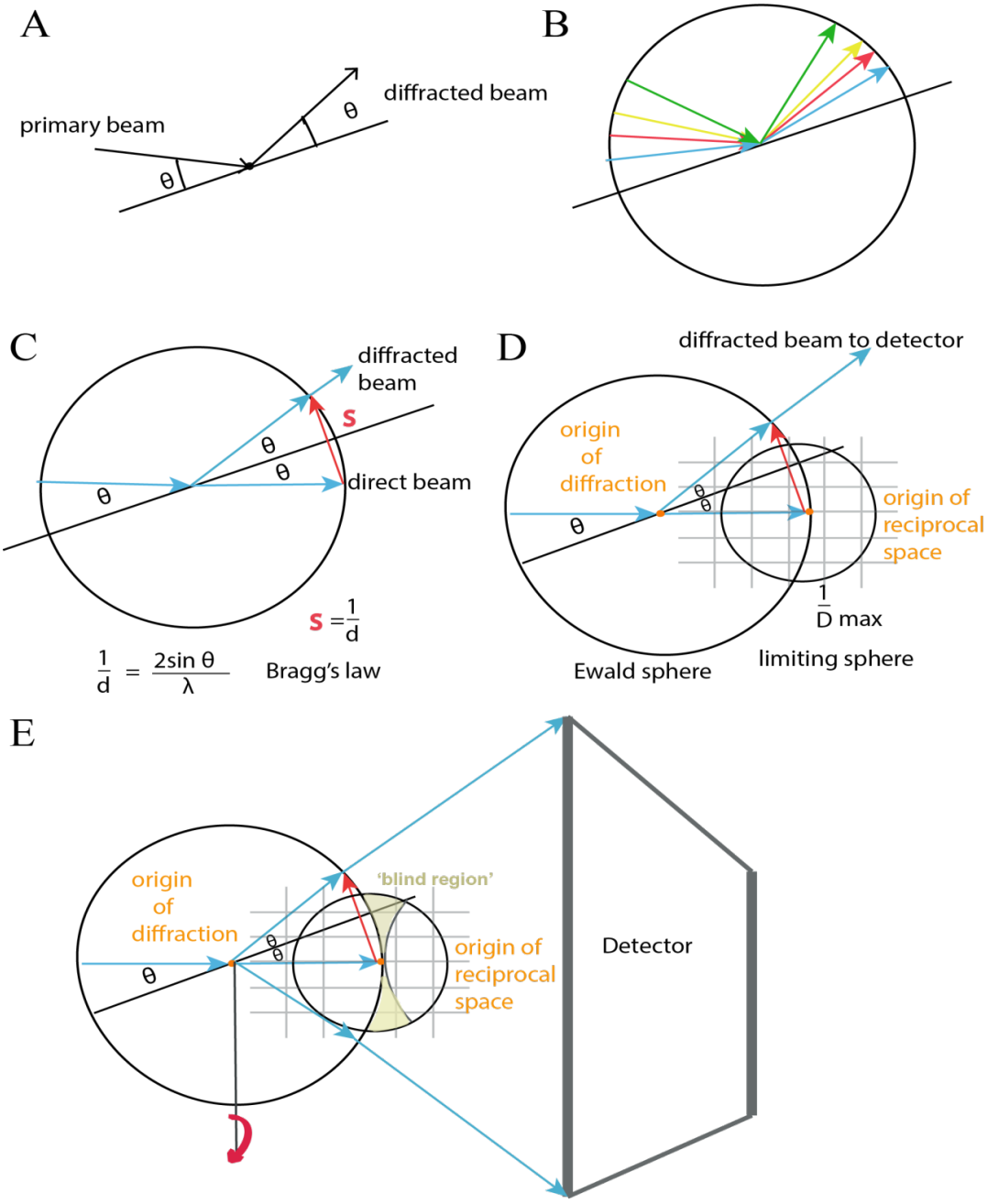

Figure 18: The relationship between the Ewald sphere and Bragg's law. A) Diffraction from a lattice plane. The angles between the primary and diffracted beams and the plane, both equal θ. B) The angles between the lattice plane and the primary and diffracted beams are always θ and diffracted beams are always on a circle. Different colours correspond to different θ angles. C) The construction of the Ewald sphere respects the properties of Bragg's law. The angle between the primary and diffracted beams is always 2θ. A vector connecting the points where the diffracted beam and the primary beam lie on the edges of the circle (S in red). If the radius of the circle equals 1/, λ then the red vector has a length of 2sinθ/λ. The center of the circle is the (real space) origin of diffraction and the point where the horizontal beam exits the circle is the origin of reciprocal space. D) Diffraction occurs when a reciprocal lattice point intersects the Ewald sphere. The corners of each rectangle indicate points of the reciprocal lattice. E) The reflections in a 'blind region' (yellow) do not intersect the reciprocal sphere and are not recorded or detected. Rotating (red arrow) and tilting the crystal helps to increase the number of detected reflections.

2.2.2 X-ray crystallography in practice

2.2.2.1 Crystallisation of proteins

There are several techniques applied in growing protein crystals such as vapour diffusion (hanging drop or sitting drop), evaporation, and seeding. A typical multi-well assay plate contains a reservoir in one well which is the crystallisation solution and a drop of the mixture of protein and reservoir in the second well. Since the plate is sealed, the system equilibrates by slow water vapour diffusion (and other volatile reagents), from a protein solution drop into the reservoir, increasing the concentration of the drop (including the protein). During this process, the drop becomes supersaturated and nucleation can begin which then promotes the crystal growth. Crystallisation solutions include precipitants, additives, detergents and others which aim at reducing protein solubility when it reaches supersaturation. These can be methyl-pentane-diol (MPD), glycerol, phosphate, sulfate and metal cations which often participate in in intermolecular contacts and sometimes are even essential for crystal formation. MPD salts are common precipitants that affect protein solubility in such a way that small amounts often increase it (salting in) and large amounts reduce it. Mild organic precipitants such as polyethylene glycol (PEG) of varying chain lengths (between 200-15 000 Da average molecular weight) are commonly used also as a mild denaturant or even cryprotectant. Crystals come in different forms and morphologies that may or may not reflect in a different space group.

2.2.2.2 Strategy of solving the structure

Prior to the structural data collection, it is beneficial to obtain the unit cell of the selected protein crystal and determine the data collection strategy. This can often be achieved by a combination of experimental and computational methods. The diffraction data are collected by rotating a crystal mounted onto a goniostat and exposed to a focused X-ray beam. The rotation of a crystal is necessary to collect the full set of data. Diffraction images are recorded on a detector during small rotation increments (oscillation) of a crystal and combined into a dataset. Fourier transform is used to mathematically transform the diffraction images from reciprocal diffraction space into direct molecular space. Phases of each diffraction spot are acquired separately and together with

FT allow synthesising or reconstructing the electron density. The position and intensity of each spot is fairly easily calculated, whereas the phases of the waves which formed each spot are more challenging to determine but are necessary to produce an electron density map. In practice solving protein structures by crystallography involves several main steps: indexing, data reduction, molecular replacement and refinement.

Collecting good diffraction data is very important in later stages of structure determination. To allow the data collection the crystal can be rotated to 360° around a single axis in small increments. A diffraction image (also called a frame) is recorded at each stage of rotation. Rotation around a single axis (often perpendicular to the X-ray beam and referred to as φ) is usually sufficient to collect many symmetry related reflections and obtain good multiplicity of a full dataset. For example, it is common to collect 1800 images of 360° crystal rotation with 0.2° increments. However, an initial screening of the data collection parameters can help to optimise the strategy according to the predicted symmetry of a reciprocal space. This is done by a short X-ray exposure time for the crystal and small portions of diffraction data being collected at various angles (i.e., 3 x 45° rotation range for 40msec) followed by data analysis and initial inspection of the diffraction image.

2.2.2.3 Indexing

Based on the Ewald sphere and crystal orientation, the *hkl* reflections are collected on the detector. However, to measure the intensities, the reflections from the reciprocal lattice have to be transformed into real space (**Table 3**). The intensities are determined by the contents of the crystal lattice and unit cell. The intensities (*I*) of all unique reflections (Bragg peaks) are collected in a diffraction experiment and allow the reconstruction of the electron density map of the protein. At this stage it is possible to calculate the dimensions of the unit cell (a,b,c and α, β, γ) and indicate the probable space group. However, the precise space group is only confirmed after the structure of a protein has been solved. Indexing is essentially an assignment of a consistent set of unit cell

vectors matching a diffraction pattern where the diffraction from each Miller plane is integrated across all frames of data and therefore possesses the intensity value (from the multiplicity). A list of intensities can be indexed using several different software such as IMOSFLM, XDS, Xia2. The algorithms applied in these calculations rely on a spot finding routine and a Fourier analysis to find a consistent set of spot positions on different images.

The initially indexed diffraction spots are transformed into a list of appropriately scaled intensities and their estimated standard errors. The integrated data consist of a list of indices, the reflection intensity of each individual reflection, the standard error, and additional batch information for each frame. Following data integration and several iterations of the background, polarization and absorption corrections, the resulting file is referred to as the scaled, unmerged data file or the *hkl* file.

2.2.2.4 Data reduction

Once the complete diffraction dataset is indexed, the data are reduced. The data reduction includes merging and scaling of multiple measurements of equivalent reflections according to the initial indexing. The programs used in this methods, such as AIMLESS, SCALA or TRUNCATE produce a *.mtz* file. The merging statistics allow the analysis of the data quality at a given resolution. These include the R-merge value, signal-to-noise ratio (I/δ) and completeness factor. The data can be reindexed if some of the data have to be removed (cut). To estimate where to 'cut' the data and decide on the appropriate resolution it is common to access the following parameters: I/sigma, to assess background to noise signal, R-merge to assess data consistency from all the reflections and the CC ½ correlation coefficient to judge the effect on refinement (311).

The reduced data are used to determine the space group. The diffraction spots and intensities signify the lattice and the unit cell geometry of the crystal system. To determine the symmetry of the unit cell and the space group of collected data it is necessary to analyse both the position of spots and their absences. Systematic absences are locations of missing *h,k,l* reflections that are absent due to destructive wave interference. The missing reflections are important in assigning the space group because for specific space groups there is a pattern of missing reflections defined. In general, the recommended unit cell is usually the smallest unit that displays the maximum symmetry of the structure. In practice, it is useful to analyse the statistics provided by the software in a table with suggested possible space groups. For example, it is worth analysing how much the RMSD changes between the P1 space group and the preferred, higher symmetry group (this means how much it is necessary to 'bend' the cell unit to fit in the data). If the initially chosen space group was incorrect, the data can be reprocessed with correct symmetry.

When analysing the data quality, we can also assess factors such as the Wilson plot and the presence of twinning or non-crystallographic symmetry (NCS). The Wilson plot is used to determine the absolute scale of the diffracted intensities and to obtain the temperature factor, therefore its value depends on the resolution. Twinning occurs when two separate crystals share some of the same crystal lattice points in a symmetrical manner and two overlapping spots may be detected in a dataset. NCS occurs when there is more than one copy of a molecule in the asymmetric unit. The Matthews coefficient calculator is used to estimate the number of molecules in a volume of the asymmetric unit considering the molecular weight of a molecule.

Data reduction essentially converts intensities (I_{hkl}) into structure factor amplitudes $|F_{hkl}|$. The structure factor is a description of each reflection including its position, intensity and phase. Reconstruction of electron density from these structure factors is crucial to visualize the map in structure determination. To calculate the electron density $\rho\ (xyz)$ and to determine the atomic positions inside the unit cell, we require phases of the various diffracted beams, $\Phi\ (hkl)$.

However, phase information is lost during the physical detection of the diffracted photon experiments (there is currently no experimental technique available to measure the phases) and is called the phase problem. The X-ray phase angle for each reflection (in centrosymmetric structures: 0° or 180°; positive or negative) is therefore supplied from the phasing experiment which involves the calculation of a molecular model that represent the contents of the crystal lattice. The phase angle indicates the angle at which the incident X-ray beam reaches the crystal. The electron density map is created, where the electron density is a Fourier transform of the structure factors, amplitudes and phases at a point (x,y,z) in the crystal lattice.

There are different electron density maps and map coefficients which can be reconstructed from structure factors and amplitudes (F_{calc}), calculated phases (φ_{calc}) and simulated noise intensity data observed from structure factor amplitudes (F_{obs}). These maps employ different combinations of Fourier coefficients and include basic difference density maps (F_{obs} -F_{calc}) • exp ($i\varphi_{calc}$) and combined ($2F_{obs}$-F_{calc}) • exp ($i\varphi_{calc}$), where F_{obs} is the observed structure factor amplitude, and both F_{obs} and φ_{calc} come from the initial phasing model. $2F_{obs}$-F_{calc} map or $2mF_{obs}$–DF_{calc} is usually visualised in blue and F_{obs}-F_{calc} or mF_{obs} - DF_{calc} are usually shown in red and green; m and D are 'de-biasing' coefficients. These maps allow for efficient model building in structural analysis. The maps are usually contoured in units of sigma (RMSD), typically 1.0σ for $2F_{obs}$ -F_{calc} and ±3σ for F_{obs} -F_{calc}. Other modifications involve changing the coefficients and include, for example, omit maps.

2.2.2.5 Molecular replacement

To determine the structure of proteins by crystallography it is essential to solve the phase problem which is referred to as phasing. Techniques for phasing include direct methods (using heavy atoms), experimental phasing, molecular replacement (MR), multi-wavelength anomalous diffraction (MAD) or single-wavelength anomalous diffraction (SAD). Molecular replacement (MR) is beneficial in a situation where there is a previously determined, structurally similar

protein model available, which can be used to calculate initial phases which then can be applied in the reconstruction of the electron density. In MR the pdb file of a similar protein is used a search model, where the sidechains are often removed and a polyalanine backbone chain is created. The starting model, provided using an automated model building software such as an Autobuild, is included in phasing calculations using for example PHASER. MR requires a search for a correct orientation and location of a search molecule with regard to the unknown structure. In the MR calculations, at each grid point of an asymmetric unit of the unknown cell, there is a search model placed in varying orientations and a structure-based correlation score is computed against the observed data. The scored solutions are ranked with the preferred solution having the highest correlation coefficient.

The output files generated by the molecular replacement software are the *.pdb* file and the *.mtz* file. The *.pdb* file contains the atomic coordinates and secondary structure information and the *.mtz* file contains the electron density map into which the protein structural model is built in real space.

2.2.2.6 Refinement

The building of a structural model of a protein can be done using visualisation software such as Coot and involves local real space fitting of residues into an electron density map using global restrained reciprocal space refinement. The cycles of refinement involve 'fitting' the model against the experimental diffraction data (often performed by Refmac5 or PHENIX) using difference maps. The iterations of refinement involve implementing small changes of positional atom coordinates to find the overall energy minima and atomic displacement parameters. Restrained reciprocal space refinement facilitates the correction of the protein's bond lengths and angles, backbone geometry and torsion angles during the model building and the refinement cycles. The model is also improved by building in water molecules and appropriate bound ion metals throughout the refinement.

Furthermore, statistical methods provide measures of how well the model fits the experimental data. One parameter of refinement is the temperature factor (B-factor) that is annotated to each atom in the final model. A high temperature factor can suggest disorder or thermal motion. Disorder indicates that the atom occupied various conformations in different molecules within the crystal, while 'thermal motion' means a vibration of an atom about its static position. Portions of flexible protein chains produce areas of low electron density in the map, making it difficult to assign the positions of corresponding atoms. Therefore, often the tails and loops of a protein chain are sometimes missing from a crystallographic atomic coordinate file. Also, oxygen and nitrogen atoms have similar numbers of electrons and therefore look similar in electron density maps, thus it is sometimes difficult to assign these atoms in sidechains such as asparagines and glutamines. Furthermore, partial occupancy always requires building of a model with care to avoid model bias. For instance, B-values under 10 create a model of the atom that is sharp and values greater than 50 produce a weaker signal or may have missing coordinates. This is common in particular for atoms at the surfaces of proteins, since they can have long sidechains exposed which are free to move in the surrounding water. Furthermore, in some crystals, there are slight differences between each of these molecules. For example, a sidechain on the surface may move between several conformations, or a ligand may bind in two different orientations, or a metal ion may be bound only to a subset of the molecules. The $2F_{obs}$-F_{calc} and F_{obs}-F_{calc} density maps used in model building indicate the missing or additional portions of a model we are building in. They can also show errors in the position of a built atom, errors in occupancy and errors in B-factor.

The are several extra parameters that can be employed in structural refinement. These include non-crystallographic symmetry (NCS) restraints, Translation/Libration/Screw (TLS) refinement and refinement of individual coordinates, B-factors, and occupancies for some atoms. NCS refinement is typically used in non-symmetrical molecules such as dimers, where different copies of a molecule in the asymmetric unit have different overall displacements. TLS is defined as a possible mean square displacements of groups of atoms in the protein model. The refinement of TLS parameters can be used to define a TLS group for rigid side chains. TLS refinement is often

useful if the NCS is present. Additionally, the anomalous difference for anomalous Fourier (DANO) is a difference Patterson electron density map that can be used in determination of the positions of heavy atoms such as metal ions within the structure.

The alternative methods to MR for estimating phases, called isomorphous replacement, involve the addition of a few electron-dense atoms, such as metal ions, to the crystal. The resulting diffraction pattern is compared with that of a similar crystal that does not have the heavy atoms. Comparing the differences, facilitates location of the heavy atoms, and estimation of the phases based on their positions. Other methods include single-wavelength anomalous dispersion (SAD) and multi-wavelength anomalous dispersion (MAD).

2.2.2.7 Structure validation

To assess the quality of a protein model, it is useful to analyse structural features such as the crystal packing (contacts between symmetry molecules), chemical bonds within peptide chains (lengths, angles, chirality, planarity and torsions) and consider possible rotamers, crystallographic parameters (real space fits, B- factor) and statistical parameters (CC ½, R-free).

Cross-validation using the R-free and R-work parameters is a useful method to assess the model. It measures how well the structural model fits the experimental data. The cross-validation relies on the concept of splitting the data into two sets: a first set that is not used in any model optimisation (~5-10% of randomly selected data, R-free), and a second set (the rest of the data, R-work) used for calculations in model building and refinement. It allows two factors, R-free and R-work, to be observed throughout the structure refinement. In a good model ~90% of the work set should fit 10% of the excluded data (random set) well. Since the random set data do not participate in refinement, R-free > R-work. The difference R-free – R-work depends on resolution and ranges from 5-7% (at medium to low resolution) to ~0.15% (at high resolution). Typical values should be approximately 0.20%. However, even if the R-factor (the difference between R-

free and R- work) is appropriate and one would expect a good model it is always crucial to check model-to-map fit, to avoid overfitting.

The CC ½ coefficient is another good parameter that helps to estimate the quality of a model. Thus, CC ½ values greater than ~0.5-0.8 indicate good correlation but may also give high correlation for weak densities. The CC½ is correlated with the B-factor: poorly defined regions typically have low CC½ and high B-factors.

Furthermore, the crucial measure of the accuracy of an atomic model is its resolution, which describes the amount of detail that may be observed in the experimental data. High-resolution structures (resolution of 1 Å or lower) are highly ordered and allow visualisation of each atom in the electron density map. In contrast, lower resolution structures (resolution of 3 Å or higher) display only contours of the protein amino acid chain.

A very useful software to validate crystal structures is Molprobity. It analyses all-atom contacts and geometry and inspects the chemical bonds within peptide chains (lengths, angles, chirality, planarity and torsions). These are compared in a Ramachandran plot, and inspected for outliers. Typically, the presence of some outliers (1-3%) is normal.

The submission of a final model of a crystal structure to the PDB archive usually involves two types of data. The coordinate file (pdb) describes atomic positions of the final model of the structure while the data files (mtz or mmcif) describe the structure factors (the phases and the intensities of the X-ray spots in the diffraction pattern) obtained from the structure solution. As part of the service, the PDB generates validation reports that provide an assessment of structure quality using broadly accepted standards and criteria.

2.3 ITC overview

Isothermal titration calorimetry is a label-free technique used to obtain the thermodynamic parameters of interactions between biomolecules in solution. It is frequently used to study the interaction between proteins, peptides and DNA. ITC directly measures the heat that is either released or absorbed throughout the binding event. It is a quantitative method and measuring the heat transfer during binding allows determination of the binding affinity (K_a), enthalpy changes (ΔH) and binding stoichiometry (n) of the interaction between molecules in solution. From these measurements the Gibbs free energy (ΔG) and entropy changes (ΔS) can be calculated using the formula: $\Delta G = - RTlnK_a = \Delta H - T\Delta S,$ where R is the gas constant and T is the absolute temperature. The dissociation constant (K_d) can be calculated using the relationship, $1/K_a$.

The measurement relies on the principle that a binding event causes heat to either be absorbed or released and this is measured by a sensitive calorimeter during gradual titration of the protein into the sample cell containing the binding partner. The microcalorimeter is composed of two cells, one of which is a reference cell and contains water or buffer, while the second cell contains the sample. The microcalorimeter retains these two cells at the same temperature throughout the time of the titrations. The heat sensing circuits detect the temperature difference between the cells when binding occurs and this triggers the response of the heaters, which return the cells to equal temperature. The instrument calculates how much power is required to maintain identical temperatures between the sample and reference cells. In an exothermic reaction, the temperature in the sample cell rises when the reactant is titrated and this requires lower power resulting in a downward peak in the signal. When the temperature of the two cells returns to equilibrium, the signal returns to the baseline. In contrast, an endothermic reaction, requires more power which is shown as an upward peak in the signal. The resulting experimental (raw) data present a series of injections over a period of time, where each heat signal corresponds to one injection of a reactant. The heat signal reaches baseline as the limiting reactant becomes saturated. The area of each peak

is integrated and plotted versus the molar ratio of binding partner/protein. The resulting isotherm is fitted to an appropriate binding model from which the binding affinity and a reaction stoichiometry can be derived.

2.4 Summary

NMR spectroscopy, X-ray crystallography and cryoelectron microscopy (not discussed here) are the main methods for determining high resolution atomic structures of proteins and their complexes. These methods are used independently or in combination to supplement each other with the complementary information that they provide.

NMR spectroscopy and X-ray crystallography, each have their unique strengths and weaknesses. Protein structures solved by both techniques are often similar, however NMR focuses on structural studies of proteins in solution or in their solid state while crystallography studies proteins in their crystal form. NMR is particularly useful for characterising the structures of well folded proteins with lower molecular weights (<40kDa) as well as proteins with large intrinsically disordered regions. In contrast flexible proteins are far more difficult to study by crystallography because it relies on having many molecules aligned in exactly the same orientation (a repeated pattern in a crystal lattice). It means that the flexible parts of proteins will often be invisible in crystallographic electron density maps. Moreover, some proteins are difficult or impossible to crystallise, because of factors such as intrinsically disordered regions and the heterogeneous conformational properties of samples. Unlike NMR, X-ray crystallography provides limited information about protein dynamics, in the form of the crystallographic B-factor. Furthermore, unlike NMR, X-ray crystal diffraction typically cannot assign the positions of hydrogen atoms or reliably distinguish nitrogen from oxygen and carbon. This means that the chemical identity of the terminal side chain atoms can be unclear, for example for Asp, Gln and Thr. Furthermore, there is a risk that the crystals and protein will be damaged by the X-ray radiation. This is caused

because the electrons from inner atomic shells can be released and the resulting free electrons can directly interact with atoms, break bonds or generate free radicals.

On the other hand, crystallography has the great benefit of being able to study larger proteins and their complexes at high resolution, given that they crystallise. Interestingly, new techniques have developed recently to improve and speed up the screening of protein crystals. This method includes *in situ* data collection. This can be achieved by collecting data directly from crystals growing in thin screen plates without the need for mounting and freezing the crystals. The possibility of using a grid scan and collecting data from a multiple positions of a crystal from different angles allows not only inspection of whether the crystal is formed of protein or salt, but also full data collection. For example, data collected from three different positions on the crystal and from three different angles can be merged into one dataset.

Similarly, NMR has its weak points and one of the main disadvantages of NMR is the limiting size of proteins which nowadays usually cannot be bigger than 40-50kDa. Another drawback is the lack of automated software for the otherwise time consuming methods used in unambiguous assignment of peaks in a spectrum which is required for protein structure solution.

Overall, both crystallography and NMR are effective in studying macromolecular complexes and also the dynamics of binding in the case of NMR. ITC and other biophysical methods are great tools for quantifying interactions and providing thermodynamic and kinetic details by delivering parameters like stoichiometry of binding partners and binding affinity.

X-ray crystallography and NMR spectroscopy have constantly been developing and are now able to provide very high resolution structural data on very complex biological assemblies. Combining them is also highly beneficial in studying structures and functions of proteins and their complexes. Moreover, NMR models can be used as molecular replacement search models or to assess the folding of proteins prior to crystallization. Furthermore, cryo-electron microscopy is gaining

momentum in structure determination, especially of big complexes and the quality is dramatically improving reaching atomic resolution.

The following books have been used as reference to describe the techniques in this chapter. NMR references included: Keeler, J. *Understanding NMR Spectroscopy (Second Edition)* (Wiley, 2010), Rattle, H. *An NMR Primer for Life Scientists* (Partnership Press, 1995), Hore, P.J. *Nuclear Magnetic Resonance* (Oxford University Press, 1995), NMR: The Toolkit by P. J. Hore, J. A. Jones, S. Wimperis Oxford University Press (2000) Rule, G. S., and Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy* (Springer, 2006). J. Cavanagh, W. Fairbrother, A.G. Palmer III and Skleton N.J. *Protein NMR Spectroscopy – Principles and Practice* (Academic Press). Crystallography sections were based on Clegg W. *X-Ray Crystallography* 2/e (Oxford Chemistry Primer, 2015), Li-ling, O. *X-ray Crystallography* (Oxford University Press, 2010), Blow, D. *Outline of Crystallography for Biologists* (Oxford University Press, 2002), Rupp, B. *Biomolecular Crystallography* (Garland Science, 2010).

2.5 Experimental material and methods

2.5.1 Cloning and mutagenesis

Proteins described in Chapter 3 and Chapter 4 were constructed using the following method. Gene fragments encoding human RNF126 (Uniprot code Q9BV68, constructs 1-40, 1-100, 142-218 and 219-289 and RNF full-length) and N-terminal SGTA (Uniprot code O43765, construct 1-86) were PCR amplified from cDNA (Life Technologies), codon optimised for *E.coli,* and cloned into the BamHI/XhoI site of a home-modified pET28 vector which includes a N-terminal thioredoxine A tag followed by a hexahistidine peptide and tobacco etch virus (TEV) protease cleavage site. DNA inserts were PCR-amplified using primers that introduced 5' BamHI and 3' XhoI restriction sites. Both, inserts and pET-28 vector were digested with restriction enzymes to create compatible overlapping ends. Vector and insert were subsequently joined together in a ligation process using T4 DNA ligase. The RNF126 (1-40) construct is later called RNF126_NZF (for N-terminal Zinc finger) in all experiments. GFP versions of RNF126 (1-40) were cloned following a similar

strategy but using a pET28 vector variant containing a N-terminal GPF tag instead of thioredoxine A. The RNF126 mutants (H14A, F36A, E38K and E39K) were obtained by PCR mutagenesis reaction using pET28-GFP-RNF126 (1-40) vector as template and overlapping complementary primers carrying the mutated codons at the centre of the oligonucleotides. The mutagenesis PCR conditions used in this work are described in (**Table 4**) and (**Table 5**). Following PCR amplification, parental, methylated DNA was removed by digestion with 10U of DpnI enzyme at 37^{o}C.

Sequences corresponding to the N-terminal SGTA (1-69) and the UBL domains of BAG6 (Uniprot code P46379, residues 17-101, named BAG6_UBL) and UBL4A (Uniprot code P11441, residues 1-74, named UBL4A_UBL) were cloned using ligation-independent Ek/LIC cloning (Novagen) into pET-46 vector containing a T7lac promoter and an amino-terminal His-Tag coding sequence immediately followed by an Ek/LIC cloning site. DNA inserts were produced in a PCR amplification with appropriately designed primers that include the 5' LIC extensions and generating vector compatible overhangs by treating purified PCR product with T4 DNA Polymerase. Target inserts were then joined with linearised pET46 vector in an annealing process by 5 min of incubation, followed by a transformation. RNF126_FL was cloned via Sequence and Ligation-Independent Cloning (SLIC) (312) because LIC and classic cloning methods have failed. The cloning was performed following the protocol reported here (312). This method involved designing primers with extensions which will produce specific sticky-ends in a PCR followed by the T4 polymerase digestion of 3’ ends (similar to the LIC cloning). However, unlike the LIC cloning it does not involve nucleotide addition and instead it is stopped after a time equal for insert and a vector. The insert with sticky ends self-assembly with a vector that also comprises sticky ends upon mixing. Furthermore, the construction process does not involve a ligase addition and during the transformation the nicks in DNA are repaired by processes of *E.coli*. In the case of yeast constructs of Sgt2 (Uniprot code Q12118, presented in Chapter 5), gene fragments encoding the Sgt2_TPR (residues 93–229 and 96-225 for the shorter construct) from *Saccharomyces cerevisiae* were PCR amplified and cloned into the home-modified pET28 vector

containing thioredoxin and hexahistidine tags (described above). The Sgt2_TPR (93-229) construct was used in structure determination by crystallography and also used in NMR titrations and ITC experiments. The shorter construct (96-225) was used to solve the structure of the Sgt2_TPR/PTVEEVD complex.

Table 4: PCR reaction mixture.

Reagents	Volume in µl
10X QuickChange Lightening Multi reaction buffer	2.5
double-distilled H_2O	18.5
ds- DNA template	1.0
mutagenic primers	1.0
dNTP mix	1.0
QuickChange Lightning Multi enzyme blend	1.0

The reactions were cycled in a thermocycler using the parameters outlined in the table below.

Table 5: PCR reaction conditions.

Segment	Cycles	Temperature	Time
1	1	95°C	2 minutes
2	30	95°C	20 seconds
		55°C	30 seconds
		65°C	30 seconds/kb of plasmid length
3	1	65°C	5 minutes

A full list of primers used in this work can be found in Appendix.

2.5.2 Protein expression

E. coli was used as an expression host to produce all recombinant proteins described in this work. SGTA (1-69) and UBL4A_UBL were transformed into Rosetta cells, while BAG6_UBL, and all plasmids carrying RNF126 and Sgt2_TPR, were transformed into the BL21 (DE3) strain. The transformation protocol involved using a standard method, where 50 µl of E. coli cells were incubated with 0.5-2µl of plasmid and heat shocked at 42 °C for 45 seconds. The mixture was plated onto Lysogeny Broth (LB) agar containing the appropriate antibiotic for plasmid selection (50 µg/ml of carbenicillin for the pET46 vector, 30 µg/ml of kanamycin for the pET28 vector and 34 µg/ml chloramphenicol for Rosetta cells). Transformed cells were used to inoculate 10 or 50 ml of medium (LB or M9 accordingly) containing the necessary antibiotic and the overnight culture, was used to innoculate larger expression cultures. Typically, protein expression was induced by adding 0.3-0.5 mM isopropyl-β- D-thiogalactopyranoside (IPTG) to 1000ml of bacterial cultures at $OD_{600} \approx 0.8$, followed by either 4 hours' incubation at 37 °C (for SGTA, SGT2, UBL4A and BAG6) or overnight incubation at 18 °C (for RNF126). ^{15}N-, $^{15}N/^{13}C$-, and $^{2}H/^{15}N/^{13}C$-labelled protein samples were expressed in M9-based minimal media supplemented with labelled ammonium chloride (>98 % ^{15}N, Sigma-Aldrich) and/or glucose (> 99% U-^{13}C, Sigma-Aldrich). For RNF126 expression, the cultures were supplemented with 10 µM $ZnCl_2$. The cultures were harvested by centrifugation at 5000g and stored at -80 ^{o}C.

2.5.3 Protein purification

Typically, harvested cells were resuspended in lysis buffer (20 mM potassium phosphate, pH 8.0, 300 mM NaCl, 10 mM Imidazole, 250 µM TCEP), supplemented with 1mM PMSF and protease inhibitors (a mix of 0.3 µM Aprotinin, 10 µM Leupeptin and 1 µM Pepstatin A) or Protease Inhibitors EDTA-free tablets (Merck). Cells were lysed by sonication on ice or using a cell disruptor (Constant Systems Ltd). Cell debris and insoluble material were clarified by centrifugation at 17,000 g for 60 mins and overexpressed protein was recovered in a soluble fraction. Recombinant proteins included a his tag and therefore could be purified using nickel

affinity chromatography (HisTrap^{TM} HP 5 ml, GE Healthcare). Typically, proteins were eluted with buffer containing 300 mM imidazole. For proteins containing a TEV cleavage site (all RNF126 constructs, SGT2 constructs and SGTA (1–86)), protein samples were dialyzed against cleavage buffer (20 mM potassium phosphate, pH 8.0 and 300 mM NaCl) and simultaneously digested with homemade TEV protease (≈ 100 μ g/ml) at 4 °C overnight. Following TEV cleavage a second nickel affinity chromatography step was performed to remove fusion protein, his tags, uncleaved protein and TEV protease; the target protein was then recovered in the flow through and loaded onto a gel filtration HiLoad 16/60 Superdex 75 column (GE Healthcare), previously equilibrated in buffer containing 10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 μ M TCEP (or 20mM Tris-HCl pH 7.5 in the case of Sgt2 used in crystallisation experiments and 100 mM pH 6.0 MES buffer, 150 mM KCl in the case of NMR titration and structural experiments of SGTA (1-69) with UBLs). Proteins were concentrated using Vivaspin concentrators (Sartorius Stedin) and sample purity was checked by SDS-PAGE after addition of 5x loading dye and boiling for 10 minutes. The SDS-PAGE was typically run at 160V for 45 minutes using 12-15% gels and stained for 60 min using Quick Coomassie Stain (Generon). Protein concentration was determined from the absorbance at 280 nm (or using a BCA assay) and molecular weights confirmed by mass spectrometry. The lyophilized peptides: PTVEEVD (corresponding to Ssa1 C-terminal; residues 634-640), MEEVD (corresponding to C-terminal of Hsp82; residues 705-709) and SLEEDLNLD (corresponding to C-terminal of Ybr137wp; residues 171-179) were purchased from Alpha BioScience (Birmingham, UK) and resuspended in water or an appropriate buffer before use. All peptides were purified and verified by HPLC (high pressure liquid chromatography) and mass spectrometry.

2.5.4 Estimation of protein concentration by BCA assay

The Thermo Scientific Pierce BCA Protein Assay was used to determine total protein concentration of SGTA (1-69), UBL4A_UBL and BAG6_UBL samples with reference to a series of bovine serum albumin (BSA) protein standards between 25 and 2000 μg/ml. Samples and

protein standards were prepared with BCA reagent according to the manufacturers protocol and their absorbances were measured at 562 nm (A562) after a 15-minute incubation period at 37˚C.

2.5.5 Native PAGE mobility shift assay

The native PAGE mobility shift assay was employed to analyse of protein-protein interactions. The experiment involved incubating 25 µM GFP-tagged RNF126_NZF (wild type and mutant variants) with increasing amounts of BAG6_UBL or UBL4A_UBL (at a concentration range from 0 to 4 molar equivalents) for 10 min at room temperature in a final volume of 20 µl. In the case of competition experiments, GFP-tagged RNF126 was added to the SGTA (1-86) and BAG6_UBL protein mixture at increasing concentrations (between 0 to 4 molar equivalents). In vice-versa titrations, increasing amounts of SGTA (1-86) were added to the RNF126/BAG6_UBL complex. The visualisation of fluorescent bands on the gel was performed using an Amersham Imager 600 (GE Healthcare) with a light source of Epi-RGB green light at 520 nm and Cy3 emission filter. The gels were also stained using Coomassie to detect protein bands. Native gels were made using 10% polyacrylamide in 0.45 M Tris-HCl pH 8.8, 1% ammonium persulphate and 0.02% TEMED and run under native conditions (no SDS), in buffer containing 25 mM Tris, 200mM glycine at 100 V for 180 min

2.5.6 NMR titrations

Protein samples of SGTA (1-69), UBL4A_UBL and BAG6_UBL for NMR titrations were dialysed overnight into 100 mM MES, pH 6.0 with 150 mM KCl. All RNF126 constructs, UBL4A_UBL, BAG6_UBL, SGTA (1-86), Sgt2_TPR (93–229) and peptides were prepared in 10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 µM TCEP). Spectra were recorded in the absence and presence of a binding partner in a suitable range of molar ratios and keeping the concentration of the labelled protein constant. Shift changes were monitored by 1D ^{1}H- and 2D ^{1}H-^{15}N HSQC spectra and experiments were typically recorded for each titration point at 25°C except for SGTA (1-69)/UBL titrations, which were performed at 30 °C. Chemical shift

perturbation (CSP) was calculated for every amide signal using the following formula, where Δδ1H and Δδ15N are the chemical shift differences for the same amide in its free and bound spectra (δfree-δbound) and for proton and nitrogen values respectively:

$$\Delta\delta^{av} = \sqrt{\left((\Delta\delta_{1H})^2 + \left({}^{\Delta\delta_{15N}}/_{5}\right)^2\right) \cdot 0.5}$$

Competition binding experiments of SGTA (1-69), UBL4A_UBL and BAG6_UBL were carried out by titrating ^{15}SGTA (1-69) with BAG6_UBL, followed by the addition of UBL4A_UBL in 1:1 molar ratio. In the reverse titration scenario, ^{15}N SGTA (1-69) was titrated with UBL4A_UBL, followed by the addition of BAG6_UBL. Prior to the NMR binding experiments, the concentration of all the proteins was measured, by comparing the integrated methyl peak of 100 µM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) to that of isolated methyl peak of the protein on 1D spectra.

CSP results were mapped onto the structures using the PyMOL software DeLano Scientific LLC, Palo Alto, CA, USA). For the RNF126 titration with UBL4A_UBL, chemical shift perturbation data were analysed and fitted using the Dynafit software (313).

2.5.7 NMR structural experiments

Samples of SGTA (1-69), UBL4A_UBL and BAG6_UBL used in experiments described in Chapter 3 were buffer-exchanged by dilution/re-concentration into 100 mM pH 6.0 MES,150 mM KCl buffer. NMR experiments were performed on samples at a concentration between 200 and 1000 µM of ^{15}N, ^{13}C-labelled protein in either 5 mm Shigemi (Sigma-Aldrich) or standard 5 mm NMR tubes at 35 or 30°C for SGTA and BAG6/UBL4A samples respectively. SGTA protein backbone assignments were carried out using experiments – HNCO, HN(CA)CO, CBCA(CO)NH, and HNCACB. Assignments for sidechain resonances were executed using a combination of HBHA(CBCACO)NH, (H)CCH-TOCSY, H(C)CH-TOCSY, and amide detected

H(CCCO)NH-TOCSY and (H)C(CCO)NH-TOCSY experiments. NOE distance restraints and assignment of aromatics rings were derived from 3D ^{15}N-NOESY-HSQC and 3D ^{13}C-separated, (^{13}C, ^{15}N-F1)- filtered edited NOESY experiments. Data were mostly collected on 600 MHz Bruker Avance lll and 800 MHZ Bruker Avance II spectrometers at Imperial College London with TCI and TXI cryoprobes respectively, controlled by Topspin3 (Bruker Biospin Ltd). The ^{15}N-filtered NOESY experiment was recorded on a Bruker 700 MHz Avance spectrometer with a cryoprobe at Kings College London. The 3D, ^{15}N-HSQC-NOESY was collected on a homebuilt 950 MHz spectrometer equipped with triple-resonance, triple-axis gradient probehead at the University of Oxford. Data were processed using NMRPipe (314) and analysed in NMRView (315) (One Moon Scientific) and with CcpNMR Analysis (316). The structure calculation was performed using ARIA2 (50) by generating 100 conformers in the final iteration and selecting the 20 best structures with lowest restraint energies for a final step of refinement in water.

NMR experiments described in Chapter 4, including BAG6_UBL and RNF126_NZF proteins, involved using samples prepared at concentrations between 100 and 1000 μM in 10% or 100% D2O (Sigma Aldrich), 10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 μM TCEP buffer (also containing 10 μM DSS for proton chemical shift referencing). NMR experiments were recorded in 5 mm NMR tubes at 25 °C on Bruker Avance spectrometers at 500 MHz and 700 MHz equipped with cryoprobes, controlled by the TopSpin 3.1 software package. Backbone assignments were performed using 3D experiments (HNCO, HN(CA)CO, CBCA(CO)NH, and CBCANH and sidechain resonances were assigned using 3D HCCH-TOCSY experiments for both RNF126_NZF and for the RNF126_NZF/BAG6_UBL complex. All NMR spectra were processed with NMRPipe followed by examination with CcpNMR Analysis. For RNF126_NZF, NOE distance restraints were obtained from the following NOESY experiments: 2D NOESY (in 90% H2O and 100% D2O). The complex study involved using either ^{15}N, ^{13}C labelled BAG6_UBL or RNF126 1–40 and acquiring the two pairs of ^{15}N-edited NOESY-HSQC, ^{13}C-edited NOESY-HSQC and ^{12}C, ^{14}N-filtered-13C-edited NOESY-HSQC experiments. Dihedral constraints (ϕ and Ψ angles) were extracted from the chemical shift values using TALOS+ (305).

The structure calculation was performed using ARIA2 (306) by generating 100 conformers in the final iteration followed by selecting 20 best structures with lowest restraint energies for a final step of refinement in water. The coordination of the zinc atom in RNF126 structure was determined in a first round of calculations using only NOE based distance restraints and once it was confirmed, the zinc atom coordination was added as an additional restraint in calculations that involved using ARIA2 tools. The final ensemble of each structure was deposited with the PDB - accession numbers:2N9O; RNF126_NZF, 2N9P; RNF_NZF/BAG6_UBL.

The samples of Sgt2_TPR used in NMR experiments described in Chapter 5 were prepared at concentrations between 500-3000 μM in 90% H2O/10% D2O (Sigma Aldrich), 10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 μM TCEP buffer. Similarly, to above, NMR experiments were acquired in 5 mm NMR tubes at 25 °C on Bruker Avance spectrometers at 500 MHz and 700 MHz equipped with cryoprobes at Kings College London, operated by the TopSpin 3.1 software package. Backbone assignments were carried out using 3D experiments (HNCO, HNCA, HN(CA)CO, CBCA(CO)NH, and CBCANH) for Sgt2_TPR. All NMR spectra were processed with NMRPipe and analysed with CcpNMR Analysis. Chemical shifts were deposited in the STAR format created in CCPN Analysis at BMRB accession number 27044.

All structures were analysed and represented using PyMOL (317).

2.5.8 Complex docking with HADDOCK

2.5.8.1 SGTA/UBL complex assembly using HADDOCK

The CSP studies provided the information about the interaction surface of SGTA (1_69)/UBL4A_UBL and SGTA (1-69)/BAG6_UBL complexes. These data were implemented in structural calculation of the complex using the HADDOCK approach (318). Assignment of intermolecular NOE provided additional restraints. Calculations of complex structures involved

using the PDB-deposited structures of UBL4A_UBL (2DZI) and BAG6_UBL (4EEW) and our lowest energy NMR structure of SGTA (1-69) dimer. Ambiguous Interaction Restraints (AIRs) were added in HADDOCK according to the standard protocol and included the most perturbed residues in CSP experiment. These were chosen from 20 amino acid residues in SGTA (1-69) (in both titrations) and 26 and 21 in UBL4A_UBL and BAG6_UBL, respectively, which were detected to have chemical shift changes higher than the averages of: 0.055 or 0.083 ppm for SGTA (1-69) and 0.222 and 0.148 ppm for UBL4A_UBL and BAG6_UBL and were used as suitable cut-off values. Following the filtering for a solvent exposed residues greater than 25%, as calculated using the program Naccess, 14 and 8 residues in each SGTA (1-69) monomer, for the UBL4A and BAG6 titrations respectively, were defined as active. These were 34, 38, 40, 41, 43, 44, 47, 48, 51, 54, 55, 60, 62, and 64 for the UBL4A titration and 39, 45, 47, 51, 54, 55, 58, and 60 for the BAG6 titration. Similarly, for UBL4A_UBL 14 residues, namely 15, 16, 17, 18, 19, 21, 23, 49, 54, 55, 56, 73, 75, and 78 were active. Finally, for the BAG6_UBL 16 residues, 22, 23, 24, 25, 26, 58, 62, 63, 65, 67, 76, 83, 85, 86, 87 and 88 were identified as active residues. Filtered NOESY experiments using a 3D ^{13}C F1-filtered, F3-edited NOESY-HSQC pulse sequence were performed on samples comprising unlabeled SGTA (1-69) and 15N, ^{13}C-UBL4A_UBL or ^{15}N, ^{13}C-BAG6_UBL. Obtaining unambigous distance restraints to support docking was not possible due to the ambiguity of the cross peak resonances. This could be due to the chemical shift perturbations upon protein binding. Additional AIRs were used instead as shown in tables in Chapter 3.6. Solvent exposed residues neighbouring the active residues were automatically defined as passive residues by the HADDOCK software. One thousand initial complex structures were produced by rigid body energy minimization, and the top 200 structures defined by total energy were selected for torsion angle dynamics and subsequent Cartesian dynamics in an explicit water solvent. Following the standard benchmarked protocol, cluster analysis of the 200 water-refined structures generated clear ensembles with the lowest HADDOCK score.

2.5.8.2 RNF126_NZF/UBL4A_UBL complex assembly using HADDOCK

To generate a model of RNF126_NZF complex with UBL4A, the same HADDOCK approach was used as for SGTA/UBL complex. For HADDOCK calculations, PDB-deposited structures of UBL4A_UBL (2DZI) and the lowest energy NMR structure from RNF126_NZF structural ensemble were used. AIRs were implemented based on the chemical shift perturbation data and allowed identification of 14 amino acid residues in RNF126_NZF and 21 in UBL4A_UBL with chemical shift changes greater than 0.15 ppm. After sorting for a relative solvent accessibility higher than 45%, 10 residues in RNF126_NZF and 14 in UBL4A respectively, were identified as active. These were RNF126_NZF residues 14, 27, 34, 36, 38, 39 and UBL4A residues 8, 44, 48, 49, 68, 70. Solvent exposed residues near the active residues were automatically named passive residues by the HADDOCK. Following the same protocol as for SGTA/UBL, cluster analysis produced 175 structures in 5 cluster ensembles. The top scoring cluster (lowest energy) was selected as the most reliable result as shown by HADDOCK benchmark testing.

2.5.9 Isothermal Titration Calorimetry

ITC experiments were performed at 30°C for SGTA (1-69) binding assays with UBL4A_UBL and BAG6_UBL and at 25 °C for RNF126_NZF interacting with UBLs and for Sgt2_TPR interacting with peptides. The experiments were carried out using ITC-200 MicroCal microcalorimeter (GE Healthcare) following the standard procedure. The samples of SGTA (1-69), UBL4A_UBL and BAG6_UBL were prepared in 100 mM MES, pH 6.0, 200 mM KCl and the titrations involved 20 injections of 2 µL of SGTA (1-69) (dimer), at a concentration of 500 µM, added to a sample of BAG6_UBL or UBL4A_UBL respectively at 50 µM (monomer).

RNF126_NZF and UBL domains were prepared in 10 mM potassium phosphate pH 6.0, 100 mM NaCl, 250 µM TCEP. In each titration, 20 injections of 2 µL of RNF126_NZF (wild-type or mutant), each at a concentration of 500 µM, were added to a sample of BAG6_UBL or UBL4A_UBL at 50 µM in the reaction cell. In Sgt2_TPR binding experiments with peptides,

proteins were prepared in 10 mM potassium phosphate pH 6.0, 100 mM NaCl, 250 µM TCEP and 20 injections of 2 µL of peptide solution at a concentration of 500 µ M, were added to Sgt2_TPR (residues 93–229) at 50 µM in the reaction cell.

Integrated heat data acquired for the titrations were corrected for heats of dilution and fitted using a nonlinear least-squares minimization algorithm to a theoretical titration curve, using the MicroCal-Origin 7.0 software package. ΔH (reaction enthalpy change in Kcal/mol), Kb (equilibrium binding constant per molar), and n (molar ratio between the proteins in the complex) were the fitting parameters. The reaction entropy, ΔS, was calculated using the $\Delta G = -RT \cdot \ln K_b$ (R = 8.314 J/(mol · K), T 298 K) and $\Delta G = \Delta H - T\Delta S$. Dissociation constants (Kd) are shown for each interaction.

2.5.10 Microscale Thermophoresis

Microscale thermophoresis was used in protein-protein interaction studies and performed on the Monolith NT.115 (Nanotemper Technologies, Munich, Germany) with fluorescently labelled proteins as described (319).

In binding assays of SGTA (1-69) with UBL4A_UBL and BAG6_UBL, purified UBLs were labelled using the Monolith NT protein labeling kit RED-NHS (Amine Reactive) dye (NanoTemper Technologies GmbH). The labelling reaction was completed in 100 mM MES pH 6.0, 150 mM KCl, 0.5 mM TCEP buffer using a concentration of 20 µM protein (molar dye : protein ratio ≈2∶1) at RT for 30 min. Unbound dye was removed with the dye removal columns which were supplied in a reaction kit. The concentration of labelled proteins was adjusted to 0.8 µM using buffer comprising 100 mM MES pH 6.0, 150 mM KCl, 0.5 mM TCEP, 0.0% TWEEN20. Samples of unlabelled SGTA (1-69) were prepared in serial dilutions of equal volumes (1∶1) using the same buffer and yielding SGTA (1-69) concentrations ranging from 800

μM to 48.8 nM. For thermophoresis, each SGTA (1-69) dilution was mixed in 1:1 ratio with one of the labelled UBL proteins, producing a final concentration of 0.4 μM fluorescently labelled protein and final SGTA (1-69) concentrations ranging from 400 μM to 24.4 nM. Following 15 min incubation, approximately 5 μl of each solution was loaded into Monolith NT Standard Capillaries (NanoTemper Technologies GmbH). Thermophoresis rates were measured using a Monolith NT.115 instrument (NanoTemper Technologies GmbH) at room temperature of 25°C with 5 s/30 s/5 s laser off/on/off times, respectively. Instrument parameters used in the experiments were: 10% LED power and 40% IR-laser power. Data from three independently pipetted measurements were analysed (NT Analysis software version 1.2.101, NanoTemper Technologies) using the signal from Thermophoresis + T-Jump.

MST studies of RNF126_NZF with BAG6_UBL and UBL4A_UBL were performed using GFP tagged construct of RNF126_NZF in buffer comprising 10 mM potassium phosphate, pH 6.0, 100 mM NaCl, 250 μM TCEP. Samples for the reaction were prepared by mixing GFP tagged RNF126_NZF with either BAG6_UBL or UBL4A_UBL and maintaining the concentration of RNF126 constant (0.125 or 0.250 μM), while UBL domain concentrations ranged from 25 μM to 0.8 nM (BAG6_UBL), or 50 μM to 1.5 nM (UBL4A_UBL) using 1:1 serial dilutions. Followed 15 minute incubation, ~5 μl of each solution was loaded into Monolith NT Standard Capillaries. Thermophoresis was measured at 25 °C with 5s/30s/5s laser off/on/off times and instrument parameters were adjusted with 10–20% LED power and 20% IR-laser power. Data from three independently pipetted measurements were compared (NT Analysis software version 1.2.101, NanoTemper Technologies) using the signal from Thermophoresis + T-Jump.

2.5.11 Crystallisation

Sgt2_TPR (residues 93–229) was concentrated to 35 mg/ml in 20mM Tris-HCl pH 7.5 buffer and crystals were obtained after four days by the vapor-diffusion method at 293K using MRC plates in 0.1M SPG, pH 6.0, 25% w/v PEG 1500 (ICL7 screen plate from Imperial College London,

based on PACT premier from Molecular Dimensions) at 20°C (drop volume = 400nl). In the case of Sgt2_TPR (96-225)/PTVEEVD complex, protein/peptide mixture was co-eluted from a HiLoad 16/60 Superdex 75 column and concentrated to 20 mg/ml in 20mM Tris-HCl pH 7.5 followed by a further peptide addition (up to protein: peptide =1:3 ratio) preceding crystallization. The lyophilized peptides: PTVEEVD (corresponding to Ssa1 C-terminal; residues 634-640), MEEVD (corresponding to C-terminal of Hsp82; residues 705-709) and SLEEDLNLD (corresponding to C-terminal of Ybr137wp; residues 171-179) were purchased from Alpha BioScience (Birmingham, UK) and resuspended in water or in the same buffer as the protein before use. All peptides were purified and verified by HPLC and mass spectrometry. The complex was crystallised after seven days by the vapor-diffusion method at 293K in 0.2M Zinc Acetate, pH 6.3, 20% w/v PEG 33500 (ICL3 screen plate from Imperial College London, PEG/Ion screen from Hampton Research) producing needles. These were optimised and rod crystals were produced by the vapor-diffusion method in 0.2M Zinc Acetate, pH 7.2, 30% w/v PEG 3350 after 7 days). All crystals were harvested in reservoir solution with 20% glycerol before flash cooling in liquid nitrogen.

2.5.12 Crystallography (Data collection and processing)

A complete dataset was collected from a single crystal on Diamond Beamline I04 for the free Sgt2_TPR dataset and I03 for the Sgt2_TPR/PTVEEVD complex using a Pixel Array Detector (PAD): Pilatus 6M-F detector and a single wavelength 0.920 Å. Data were processed using Xia2 (320) followed by scaling and merging with Aimless and the asymmetric unit contents were estimated using the Matthews coefficient with CCP4 suite (321). In the case of the complex data were reduced with Xia2 and scaled with Aimless. The scaled and unmerged data were selected and one batch of bad images was discarded. Next the data were merged using Aimless providing good statistics.

2.5.13 Crystallography (Structure solution and refinement)

The crystal structure of Sgt2_TPR (93–229) was determined by molecular replacement using Phaser (322) with human SGTA_TPR crystal structure (PDB accession number: 2VYI) used as a search model. The Sgt2_TPR (96-225)/PTVEEVD complex structure was determined by molecular replacement using Phaser and mammalian model (PDB accession: 2VYI) as a search model. The free Sgt2_TPR structure was refined using REFMAC5 (323) and the complex structure, using PHENIX (324) with manual model building using Coot (325). PHENIX was run using a script with automatic NCS and TLS (10 final TLS groups). Before submission, data were extracted into the correct format using PDB_EXTRACT 3.20. A free R-value of 4.9% was used as a cross-validation method for Sgt2_TPR and 4.8% for Sgt2_TPR (96-225)/PTVEEVD. Water molecules, Zn atoms and BO_4 atoms were fitted manually using Coot.

Chapter 3. Structural and binding studies of SGTA, BAG6 and UBL4A

This chapter aimed to contribute to better understanding of the way in which hydrophobic proteins have their fate decided by the collaboration between SGTA and the BAG6 complex. First, a solution structure of the SGTA dimerisation domain was solved. This was followed by using biophysical techniques to investigate the interaction of SGTA's UBD with two different ubiquitin-like domains (UBL) from the BAG6 complex. Moreover, a combination of structural and interaction data was implemented including the HADDOCK protein-protein interaction docking tool to generate models of the SGTA-UBL complexes.

3.1 Protein production

The domains of SGTA (1-69), SGTA (1-86) and UBL4A_UBL constructs were designed based on the previous studies in Isaacson lab. Proteins were successfully overexpressed in *E. coli* Rosetta cells and BAG6_UBL was expressed in BL21 (DE3). The purification steps proved all proteins were well expressed, soluble and found to be stable. The SDS-PAGE profile shows bands corresponding to the correct proteins sizes of approximately 9.3kDa for SGTA (1-69), 9.6kDa for UBL4A_UBL and 10.4kDa for BAG6_UBL, with the BAG6 band consistently running higher than expected on a gel (**Figure 19**). In all cases, overexpressed protein was recovered in the soluble fraction. Enrichment of expressed protein was achieved by affinity purification and size exclusion chromatography (SEC), prepared in a buffer at a pH that was distant from the pI of the individual proteins. Final yields of purified proteins were typically 50-100 mg/L for expression in LB and in M9 minimal media. The high yields of SGTA (1-69) did not bind to the His-Trap column efficiently and protein was typically present in the elution and also wash fraction possibly due to overloading the His-Trap column or non accessible His-tag obscured by the dimer interface of SGTA (**Figure 19**).

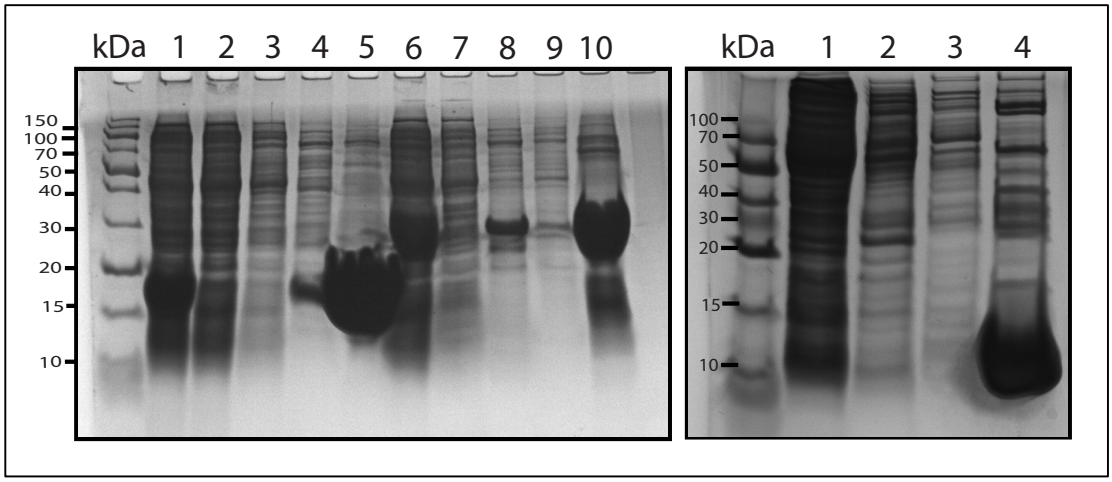

Figure 19: **Purification of SGTA (1-69), BAG6_UBL and UBL4A_UBL.** Left, SDS-PAGE analysis of purification by affinity chromatography of SGTA (lanes 1-5) and BAG6 (lanes 6-10). 1), 6) uninduced cell control 2), 7) flow through after column loaded with the cell lysate, 3), 8) wash with 5 CV of wash buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 μ M TCEP), 4), 9) wash with 2 CV of wash buffer 2 (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 μ M TCEP, 20mM Imidazole) 5), 10) elution with 2 CV of elution buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 μ M TCEP, 300mM Imidazole). Right, SDS-page analysis of UBL4A_UBL purification 1) uninduced cells control 2) flow through after column loaded with the cell lysate, 3) wash with 5 CV of wash buffer, 4) elution with 2 CV of elution buffer.

3.2 SGTA (1-86) ^{1}H, ^{13}C, and ^{15}N Chemical Shift backbone and side chain assignment

The backbone and side-chain assignments of SGTA (1-69) construct were mostly assigned by a Master's student Mr Newran Sriskandarajah employing NMRView modules which provided rapid input for MARS automated assignment (326) and facile handling of sidechain data (327). These assignments for side-chain resonances were performed from a combination of HBHA(CBCACO)NH, (H)CCH-TOCSY, H(C)CH-TOCSY, and amide detected H(CCCO)NH-TOCSY and (H)C(CCO)NH-TOCSY experiments. I have completed the backbone assignment of N-terminal SGTA (residues 1-86) using CCPN analysis, including the assignment of the linker between the N-terminal domain with the TPR domain. The chemical shifts were deposited at BMRB with Accession Number: 19779. The complete set of 3D NMR experiments (HNCA, HN(CO)CA, HNCACB, HN(CO)CACB and also 2D NOESY experiment) were acquired at 700MHz of SGTA (residues 1-86), to facilitate the full backbone assignment. The protein sample

was prepared for NMR spectroscopy in 100 mM MES, pH 6, 150 mM KCL and 10% D_2O buffer. Under these conditions the number and dispersion of the peaks displayed the features of a homogeneous folded protein. ^{15}N HSQC peaks were picked manually and the NH assignment for all missing residues was completed except for Prolines. The full ^{15}N HSQC spectrum with assigned peaks is shown in **Figure 20**. The overlay between SGTA (1-69) and SGTA (1-86) is shown in **Figure 21**. By the analysis of the HSQC spectrum, residues 69-86 were likely to be flexible therefore construct 1-69 was used in structural calculations.

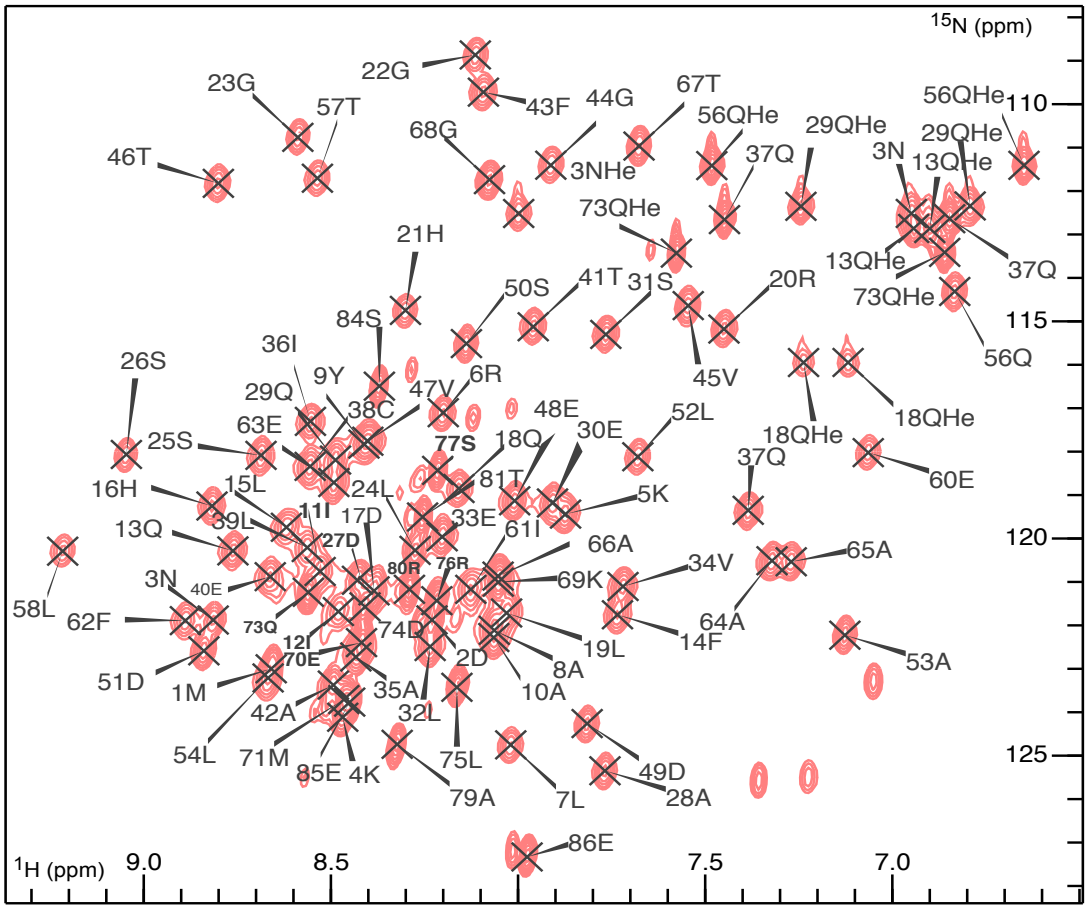

Figure 20: **^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA (1-86).** The assignment of SGTA (1-69) was propagated or reassigned onto the longer construct of SGTA including the missing 17 residues of the linker. 'He' corresponds to side chain protons.

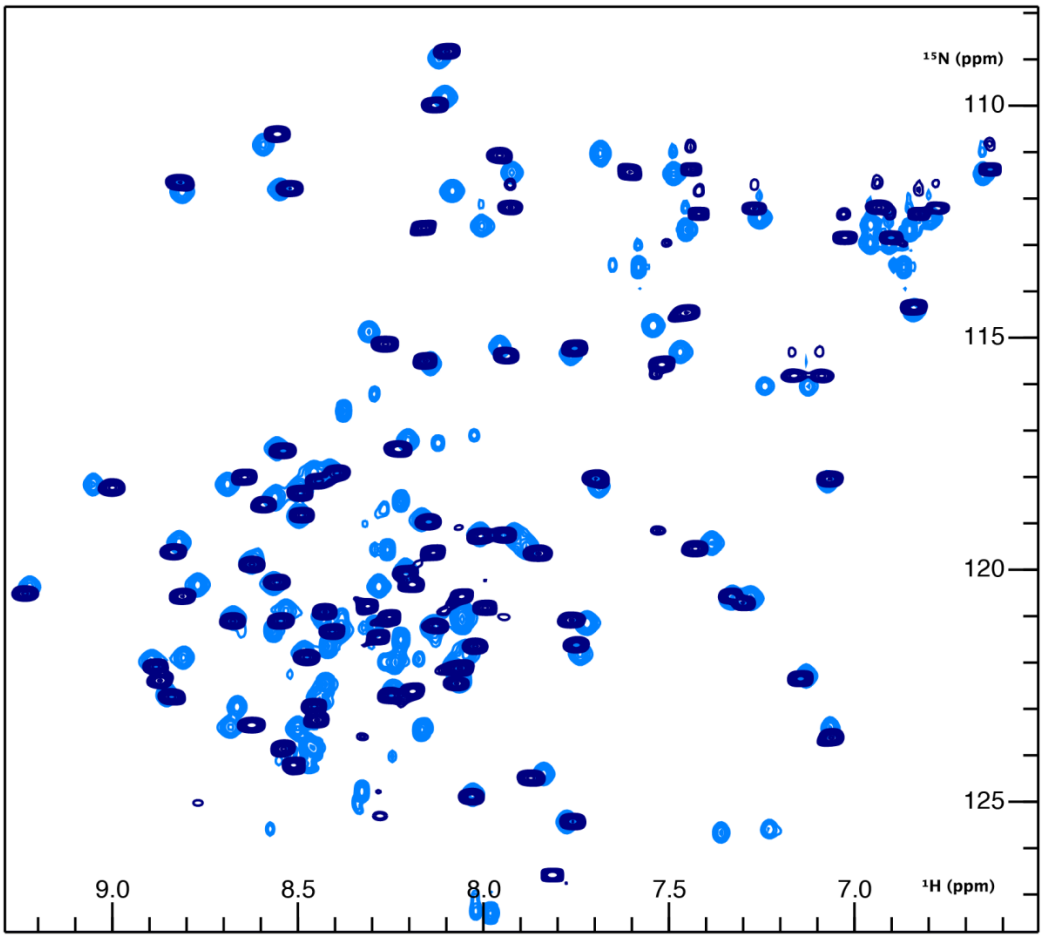

Figure 21: An overlay of ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA (1-69) in dark blue and ^{15}N-labelled SGTA (1-86) in light blue.

3.3 The NMR solution structure of SGTA (1-69)

The SGTA (1-69) construct was used in all structural calculations which have been done in collaboration with Dr John Darby using ARIA2. The backbone and side chain assignment chemical shift lists, dihedral constraints (ϕ and Ψ angles), extracted from the chemical shift values using TALOS+ and manually assigned NOE restraints, were used to calculate the structure of a SGTA (1-69) dimer. NOE distance restraints and assignment of aromatics rings were derived from 3D, ^{13}C-NOESY-HMQC, 3D, ^{15}N-NOESY-HSQC and 3D, ^{13}C-separated, (^{13}C, ^{15}N-*F1*)-reject filtered NOESY and ^{15}N-filtered NOESY experiments. In the initial Aria runs we also allowed for automated nuclear Overhauser effect (NOE)-derived assignment. In each of many cycles of structure calculations with water refinement, we inspected the restraints violations and

manually corrected them. The final Aria run generated 100 conformers in the final iteration and the 20 best structures with lowest restraint energies were selected for a final step of refinement in water. The final ensemble was deposited at the PDB with the Accession Number: 4CPG. The structural statistics including those obtained from PROCHECK are presented in **Table 6**. Like its yeast homologue, SGTA (1-69) forms a tight symmetrical homodimer (**Figure 22**). Each monomer consists of four alpha helices (α1= N3-H21; α2= S26-F43; α3= V47-L52; α4= L58-T67) connected by short loops with helices from each subunit arranged in a head to tail fold form, analogous to that of the Sgt2 N-terminal. The first two helices are of similar length and mediate homodimerization with the third, shorter helix packing against their other side, away from the dimer interface (**Figure 22 B**). The dimer interface is highly hydrophobic resembling the core of a globular protein and spans an area of 1266 $Å^2$ as calculated by PISA (328). The dimer contacts mainly via the α2 helices and includes the interaction of small and conserved residues (Ser31, Ala35, Cys38, and Ala42), resulting in close packing between the main chains of the two subunits. Also, residues with larger hydrophobic side chains (Leu24, Leu39, Thr41, Phe43 and Val34 on α2 and Ile11 and Leu15 on α1) contribute to the formation of the dimer hydrophobic core (**Figure 22 F**). The conserved acidic residues Asp27 and Glu30, Glu33 and Glu40, result in a negatively charged ring surrounding the hydrophobic core (**Figure 22 C**). Our SGTA (1-69) structure aligns with the yeast homologue with an RMSD of 2.4 Å. Minor structural differences are present at helix α3 and the loop between helices α2 and α3 which form part of the UBL interaction site. A structural overlay with the equivalent yeast domain is shown in **Figure 22 D**. The equivalent conserved residues of N-terminal SGTA and Sgt2 of which some are involved in formation of the binding interface with UBL domains are shown in **Figure 22 G** and **H** and the sequence alignment showing conserved, hydrophobic and charged residues is shown in **Figure 22 E**.

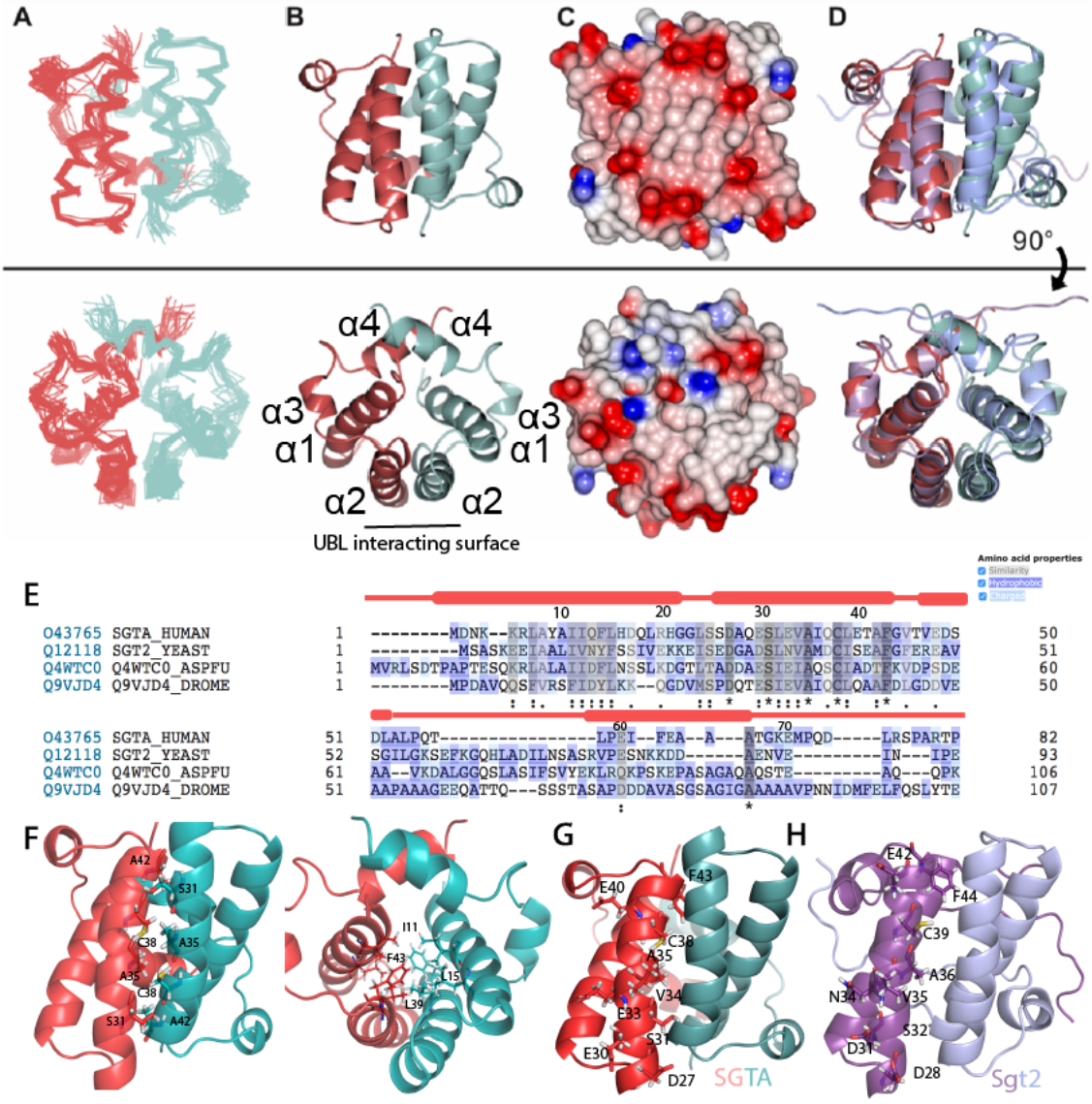

Figure 22: **Orthogonal views of NMR structures of SGTA (1-69) dimer**. A) Ensemble views showing 20 lowest energy ARIA-calculated structures as deposited in the PDB (Accession code: 4CPG); monomers indicated by red and green. B) Cartoon views of a dimer with monomers coloured as in A. C) Electrostatic views ranging from −10 negative charge in red to +10 positive charge in blue modelled using ccp4mg (329) which calculates the charge distribution present on the solvent accessible surface of the protein. D) Structural alignment of SGTA (1-69) (red/ green) with Sgt2 N-terminal (lilac/blue; PDB: 4ASV) superimposed using secondary-structure matching in ccp4mg. The structures align with RMSD of 2.41Å. E) Sequence alignment of SGTA (1-69) homologs. The species are *Homo sapiens*, *S. cerevisiae Aspergillus fumigatus* , *Drosophila melanogaster*. Alignment was made using Clustal and asterisks indicate sequence identity and dots indicate sequence similarity. Residue coloring is based on similarity (grey), hydrophobicity (purple) and charge (blue). The residues are numbered according to the human SGTA sequence. The helices are shown as red cylinders. F) Details at the dimer interface showing small, conserved residues (left) and bulky, hydrophobic residues (right) with side chains highlighted as sticks. G) Conserved residues are shown as sticks on SGTA and equivalent residues on Sgt2 (H) (82).

Table 6: Summary of SGTA (1-69) dimer solution structural statistics of ensemble of 20 structures.

Distance and Dihedral Restraints	
NOEs: Intra-residue	518
Sequential	345
Short range	232
Medium range	106
Long range	183
Inter-monomer	136
Total unambiguous	1520
Ambiguous	694
Total NOE-derived	2214
Hydrogen bond restraints (applied to each monomer)	68 (34 H-bonds)
Dihedral angles (φ + ψ) (applied to each monomer)	126 (63 residues)
Structural Statistics	
Ramachandran Plot (%) (from Procheck-NMR)	
Most favoured	92.2
Additionally allowed	7.2
Generously allowed	0.6
Disallowed	0.0
RMSD from experimental restraints	
Distances (Å)	0.027 ± 0.018
Dihedrals (^{o})	1.24 ± 1.32
RMSD from idealized geometry	
Bonds (Å)	0.0034 ± 0.0007
Angles (^{o})	0.53 ± 0.101
Impropers (^{o})	1.78 ± 0.718
NOE Violations > 0.5 Å	0
> 0.3 Å	0
> 0.1 Å	7.7 ± 3.5
Structural Precision	
RMSD from average structure (Å)	
Backbone (all/2ndary structure)	0.95 ± 0.35 / 0.50 ± 0.16
Heavy atom (all/2ndary structure)	1.35 ± 0.34 / 1.00 ± 0.15

3.4 Probing the binding interface for SGTA (1-69)/UBL4A_UBL by chemical shift perturbation

The NMR structure of the UBL domain from UBL4A which we used in our experiments has been solved by Zhao and colleagues as part of the RIKEN Structural Genomics Initiative and deposited in the Protein Data Bank (Accession Number: 2DZI) with no associated publication to date. Since the NMR chemical shifts were not originally published, we used the backbone assignments of UBL4A_UBL reassigned by Aline Simon in our lab (90) in our interaction experiments. Reciprocal chemical shift perturbation studies (CSP) were carried out by titrating unlabelled UBL4A_UBL into ^{15}N-labelled SGTA (1-69) and vice versa to probe the binding interface between the two domains (**Figure 23** and **Figure 26** peak list can be found in appendix. The interaction was monitored by recording a series of ^{1}H/^{15}N-HSQC experiments until the binding reaction was saturated and no further chemical shift perturbation was observed. The spectra showed binding in a fast and fast-intermediate exchange regimen and CSPs were calculated for each residue in its free and bound position (δfree-δbound) for proton and nitrogen values respectively (**Figure 24 and Figure 29**) and plotted on the structure, with the most perturbed residues shown in dark red (**Figure 27A and C**). The results showed that the interface is analogous to the equivalent complex in yeast Sgt2/Get5 (90), although with some differences. While, only one set of NMR signals from the two domains of the SGTA homodimer was observed in the bound state, it would be expected that a single UBL domain, binding at the homodimer interface should disrupt the chemical equivalence of some NMR signals from the two halves of each dimer, resulting in two sets of signals. This was a similar result to the yeast complex of Sgt2_NT/Get5_UBL. In the Sgt2_NT/Get5_UBL complex, we, and others (263) found that exchange of the UBL between the two equivalent binding sites on the Sgt2 dimer is faster than the NMR chemical shift timescale causing averaging of the frequencies at the elevated temperatures required for analysis (310 K) and it is probable that we observe a similar event in the mammalian complex (82). The 1∶1 (SGTA dimer: UBL monomer) stoichiometry was confirmed by ITC and MST (see later section).

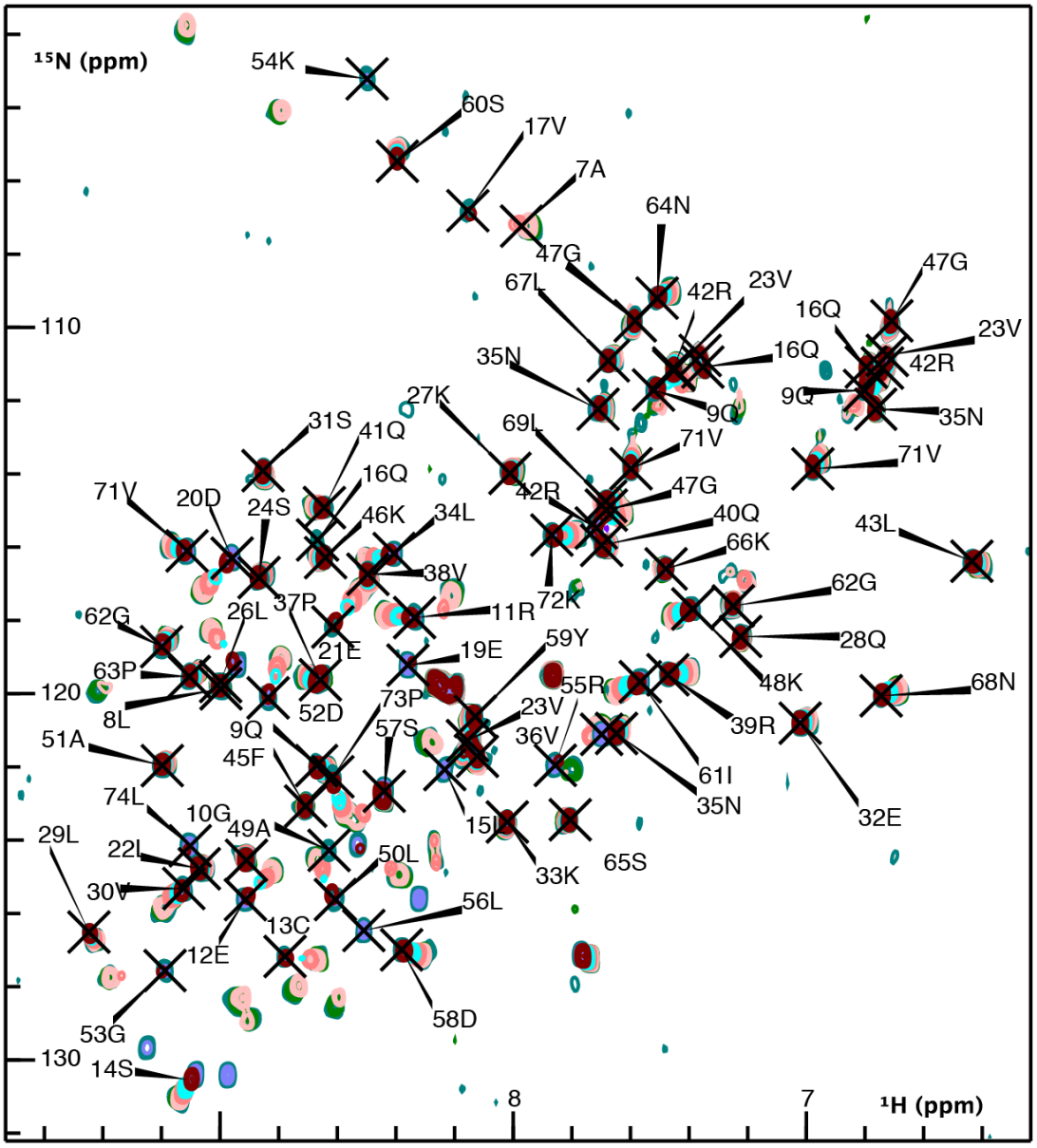

Figure 23: Full ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled UBL4A_UBL titrated with unlabelled SGTA (1-69). The ^{15}N-labelled UBL4A_UBL: SGTA (1-69) protein ratio was used at 1:0 (maroon), 1:0.3 (blue), 1:0.6 (teal), 1:08 (cyan), 1:1 (salmon), 1:1.4 (light pink), 1:1.9 (green).

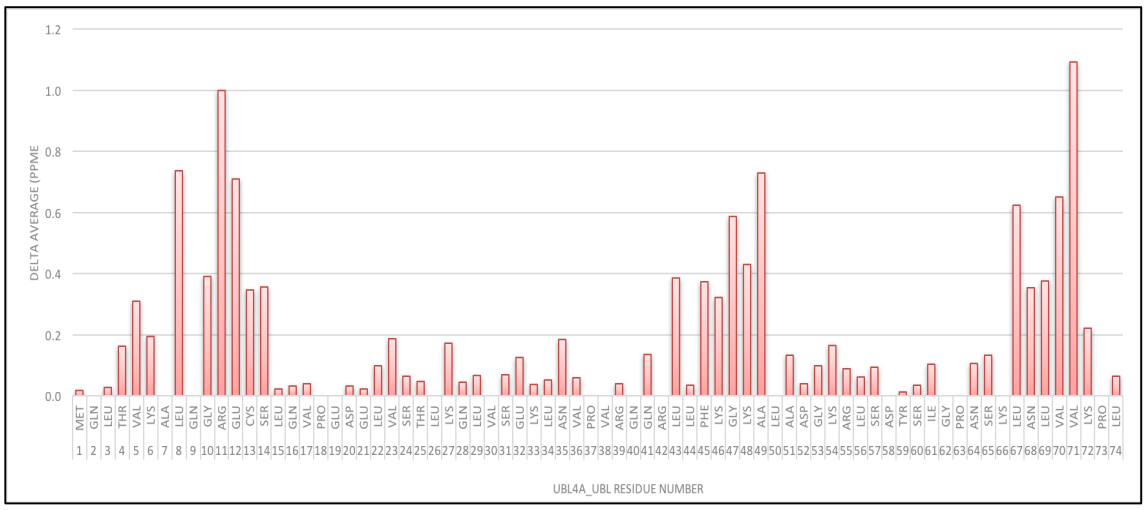

Figure 24: Normalized CSP of ^{15}N-labelled UBL4A_UBL titrated with unlabelled SGTA (1-69). The graph shows the CSP plot for each residue.

3.5 Probing the binding interface for SGTA (1-69)/BAG6_UBL by chemical shift perturbation

We aimed to investigate the interaction of SGTA (1-69) with the second UBL domain from the BAG6 complex, which belongs to the BAG6 protein itself. Since UBL domains of both, BAG6 and UBL4A, are highly conserved and represent ubiquitin-like folds, we wanted to compare the SGTA-UBL interactions. The solution structure of the UBL domain from BAG6 (residues 17-101) was previously solved by Zhao and colleagues and deposited in the Protein Data Bank (Accession Number: 1WX9) and the BMRB (Accession Number: 11263). Furthermore, a crystal structure of this domain was also solved and deposited by Kozlov *et al*. (PDB Accession Number: 4EEW). Neither of these structures is yet associated with a publication. We therefore used the published BMRB assignments in our reciprocal CSP studies. We used the BAG6 STAR file from the BMRB and converted it into an NMRView file (XPK). The numbering we use here corresponds to the BMRB entry. Leu8 is the first relevant amino acid and this corresponds to the first amino acid in the UBL domain or the 17th amino acid in the full length protein. The first 7 amino acids in the BMRB sequence are a Gly/Ser linker from the construct but most of these haven't been assigned. The CSP binding experiments were carried out by titrating ^{15}N-labelled BAG6_UBL into unlabelled SGTA (1-69) and vice versa (**Figure 25** and **Figure 26**; the peak list

can be found in the appendix. Similarly, to the NMR titrations of SGTA and UBL4A, the interaction of SGTA with BAG6_UBL is mostly in a fast exchange rate in the NMR timescale with the same pattern of peak shifts. CSPs were plotted onto the SGTA (1-69) and BAG6_UBL structures (**Figure 28**) and the comparison of shift perturbations highlighted in red in **Figure 27** shows that SGTA (1-69) binds the BAG6 and UBL4A UBLs at the same binding site. However, in contrast to UBL4A_UBL binding, we do observe peak splitting for several residues located in α2 and in the loop connecting α1 and α2 (His21, Gly22, Ala28, Ala35, Gln37 and Ala42), as shown in **Figure 27E** and **Figure 30** and Figure 37**.** The residue A28 is presented here as an example and the remainder of the residues displaying peaks splitting are later displayed on the complex structure. The peak splits may suggest that that exchange of the UBL between the two binding sites on the SGTA (1-69) homodimer is slower than the NMR chemical shift timescale in the case of BAG6_UBL binding. This interesting, although subtle, difference in the dynamic properties of the SGTA/UBL complexes may contribute to further characterization of their binding kinetics.

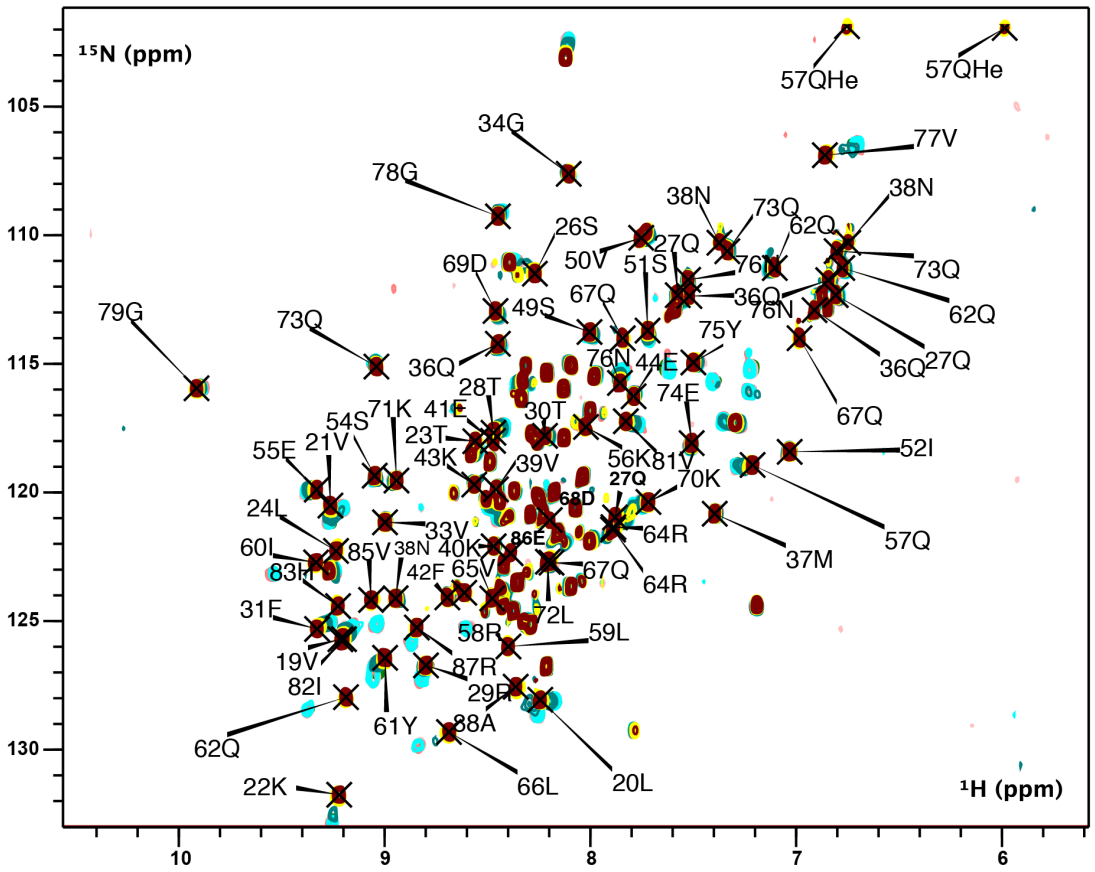

Figure 25: Full ^{1}H-^{15}N HSQC spectra of ^{15}N -labelled BAG6_UBL titrated with unlabelled SGTA (1-69). The ^{15}N-labelled BAG6-UBL: SGTA (1-69) protein ratio was used at 1:0 (maroon), 1:0.1 (yellow), 1:0.3 (blue), 1:0.6 (teal), 1:08 (cyan), 1:1 (salmon), 1:1.4 (light pink), 1:1.9 (green).

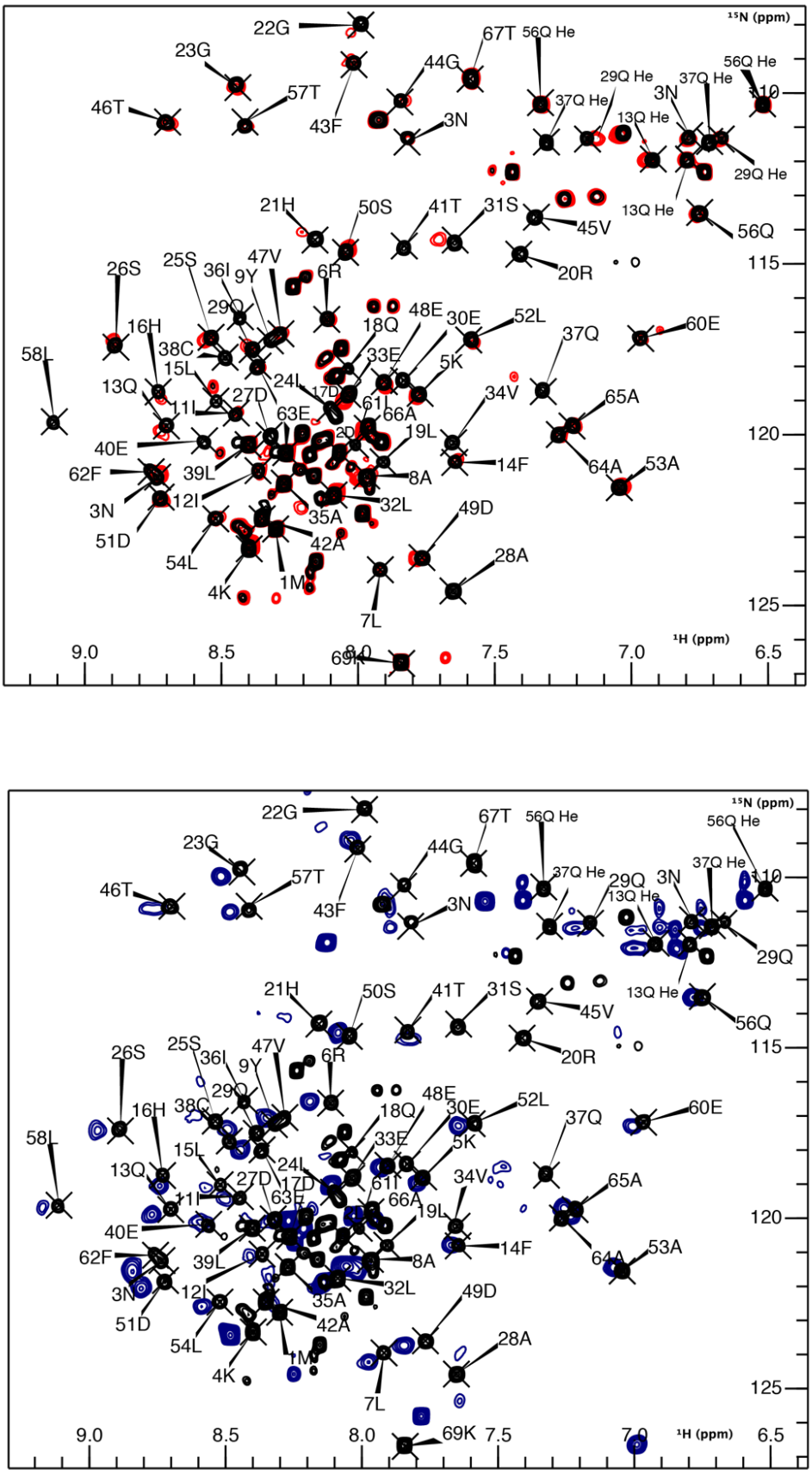

Figure 26: Full ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA (1-69) before (black) and after (red/blue) titration with saturating quantities (1:2 molar ratio) of unlabelled BAG_UBL (bottom) and UBL4A_UBL (top).

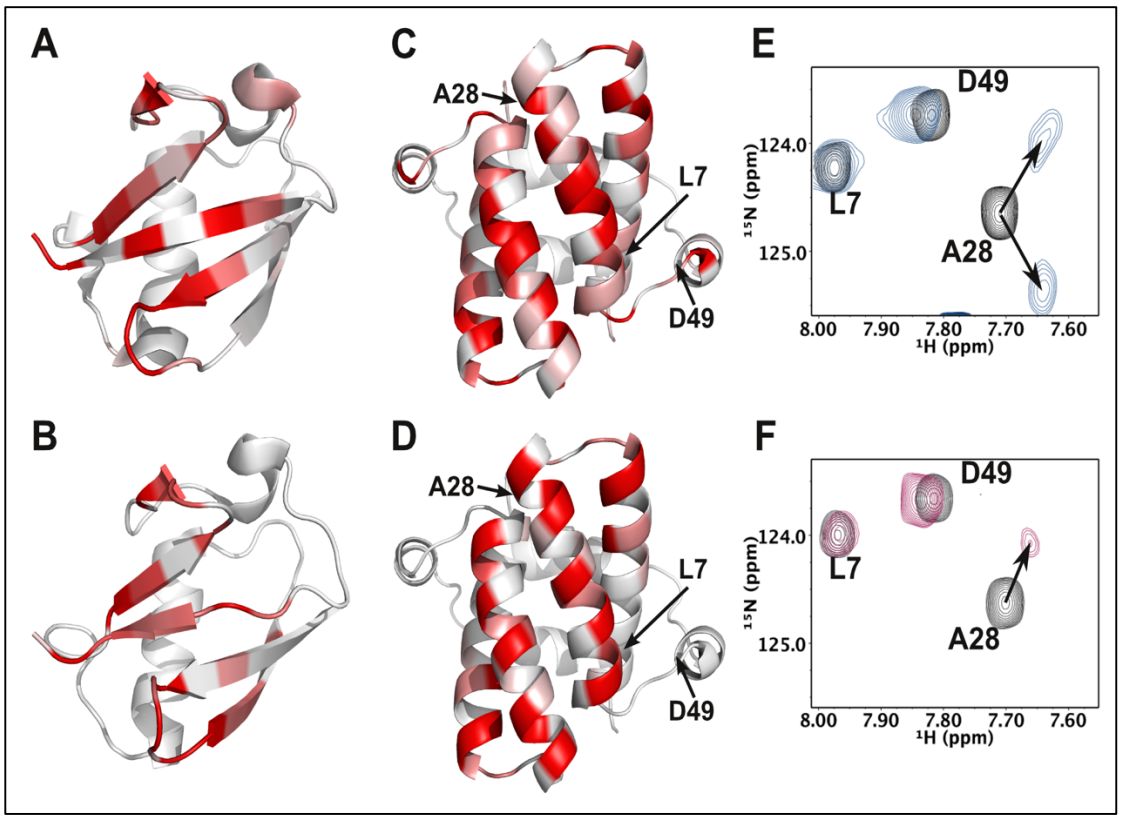

Figure 27: Chemical shift perturbation data for SGTA (1-69)/UBL interactions. A–D: Cartoon views coloured according to reciprocal CSP upon addition of a binding partner. Residues whose shifts are higher than one standard deviation above the mean chemical shift are coloured darkest red. Those below the mean are coloured white and shifts between these points are coloured pink. A) BAG6_UBL B) UBL4A_UBL C) SGTA (1-69) upon binding BAG6_UBL D) SGTA (1-69) upon binding UBL4A_UBL; E–F: Region of ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA (1-69) before (black) and after (blue/maroon) titration with saturating amounts of unlabelled BAG_UBL (E) and UBL4A_UBL (F) Residue Ala28 splits upon binding to BAG6_UBL but not upon binding UBL4A_UBL. Residues Leu7 and Asp49 are displayed on the SGTA structure to show where the neighboring peaks on an HSQC spectrum of the A28 splitting peak are localised on the structure (82).

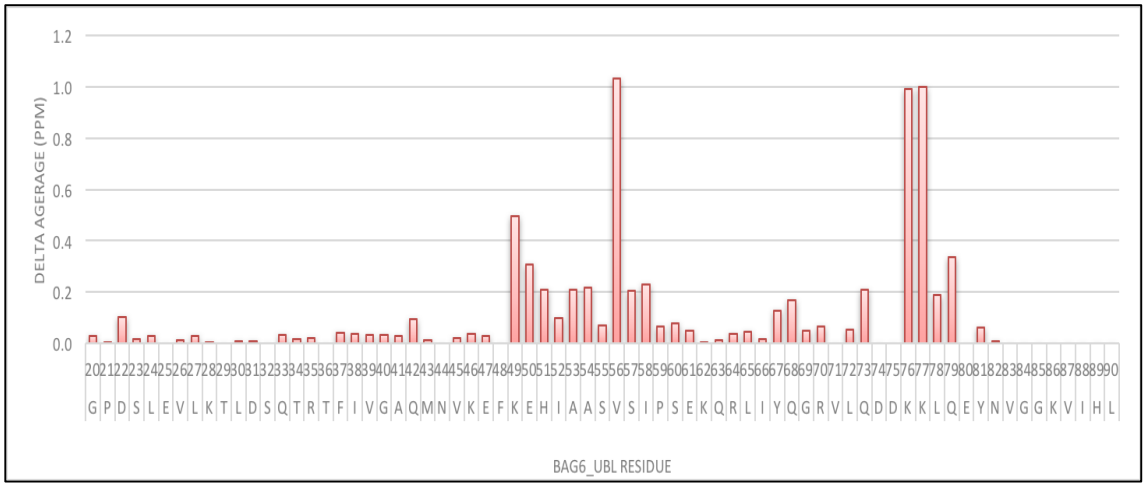

Figure 28: Normalized CSP of ^{15}N-labelled BAG6_UBL titrated with unlabelled SGTA (1-69). The graph shows the CSP plot for each residue.

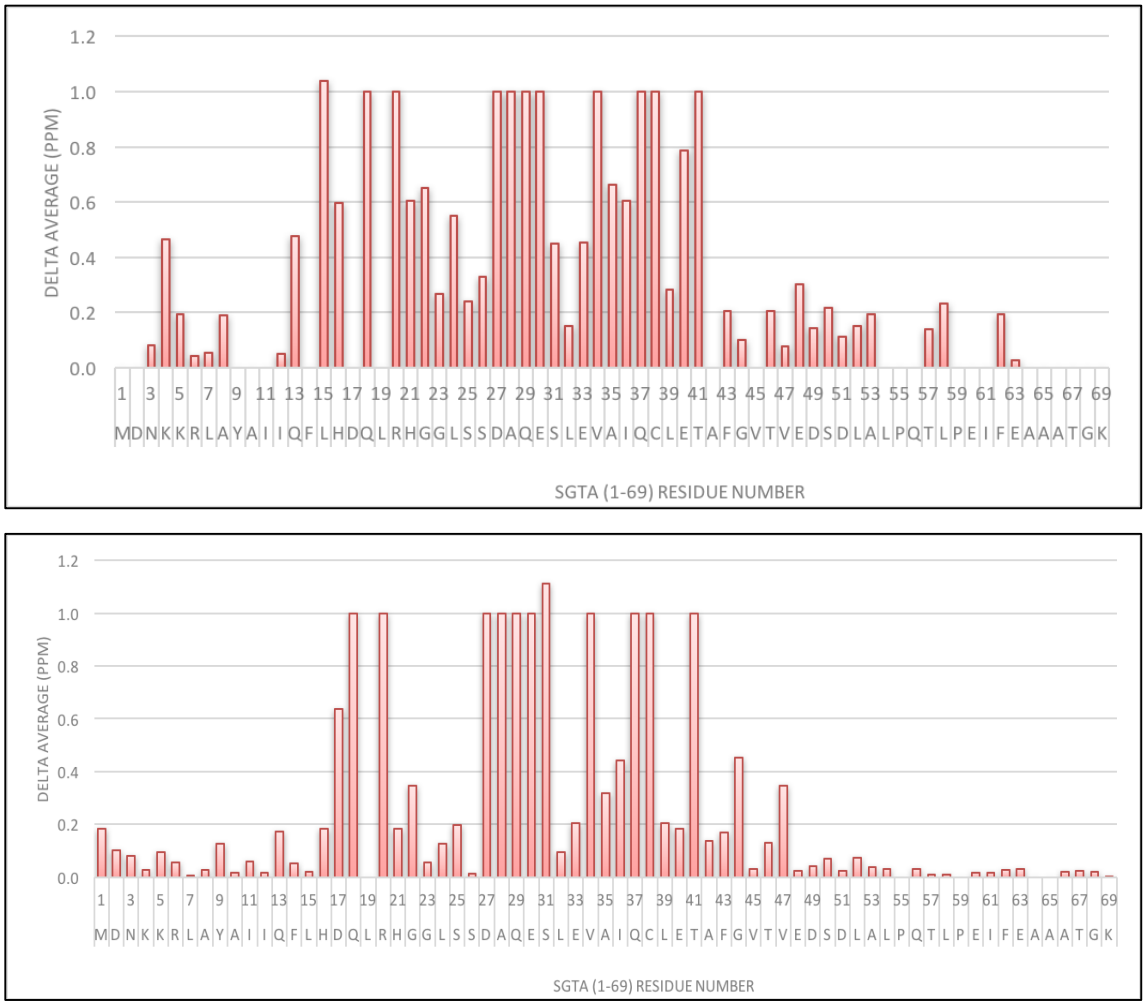

Figure 29: Normalized CSP of ^{15}N-labelled SGTA (1-69) titrated with unlabelled UBLs of UBL4A (top) and BAG6 (bottom). The graph shows the CSP plot for each residue.

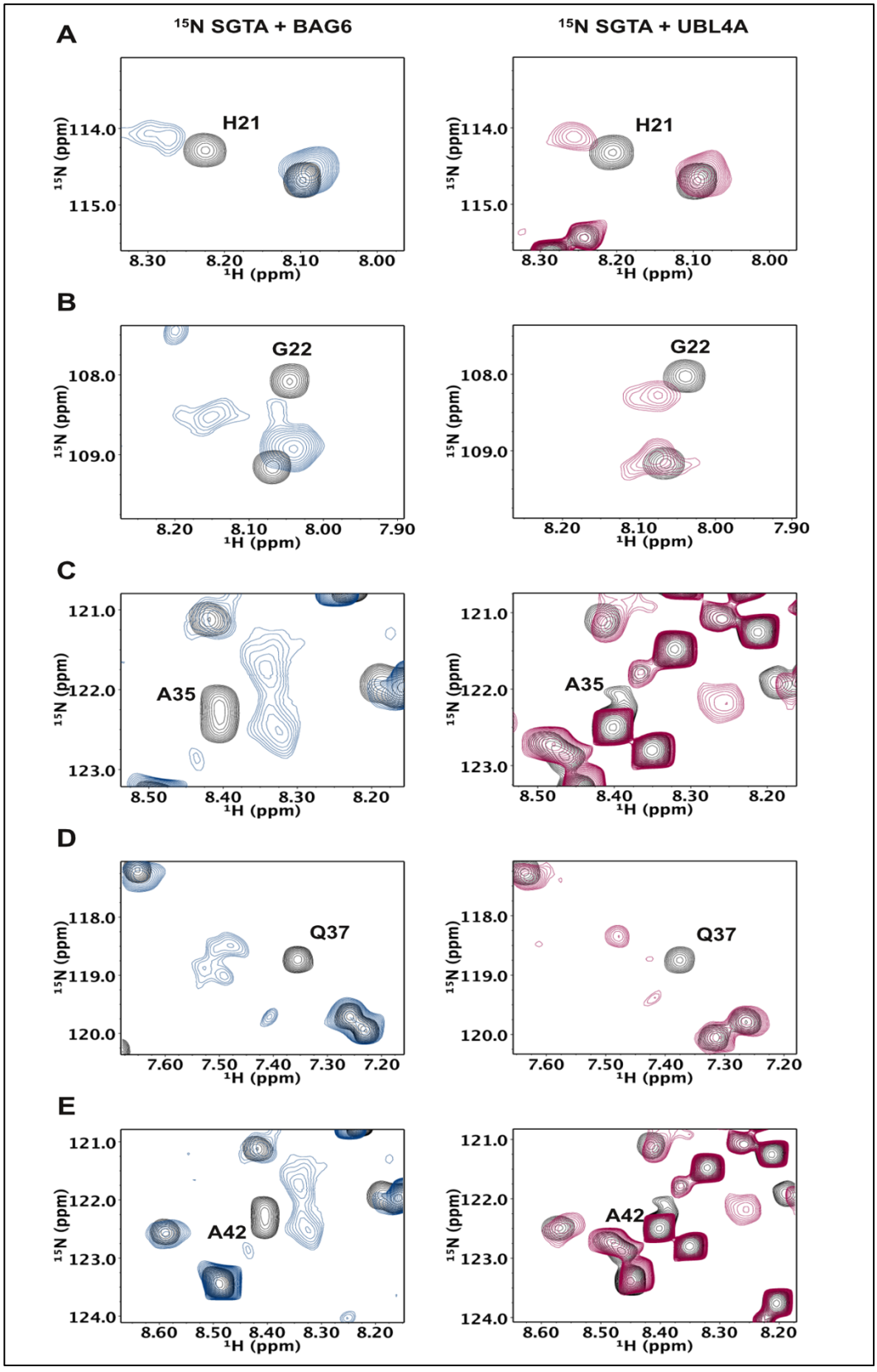

Figure 30: Fragments of ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA (1-69) before (black) and after (blue/red) titration with saturating amounts of unlabelled BAG_UBL (left) and UBL4A_UBL (right). The selected fragments of the spectra indicate amide backbone peaks which show peak splitting, possibly due to the slower exchange of BAG6_UBL between its two orientations on the dimer of SGTA (1-69) (82).

3.6 SGTA (1-69)/UBL4A_UBL and SGTA (1-69)/BAG6_UBL complexes by HADDOCK

The structures of SGTA (1-69) and the UBL domains display high similarity to the yeast equivalent studied previously. However, the CSP experiments indicate some subtle differences in UBL domains which include residues at the start of strands β3 and β5 (residues 39–42 and 66–68 in the sequence alignment shown in **Figure 33 G**). This showed large perturbations in the mammalian but not the yeast system. The binding interface is very similar in both SGTA-UBL complexes although with some differences in the positions of the conserved and charged residues known to be important for the Sgt2/Get5_ubl interaction from previous mutagenesis experiments. Therefore, we aimed to investigate this structurally, by generating models of the SGTA (1-69)/UBL4A_UBL and SGTA (1-69)/BAG6_UBL complexes using HADDOCK-based semi-rigid, data-driven docking. Ambiguous interaction restraints (AIRs) for HADDOCK (330) were defined using the chemical shift perturbation data and residue solvent accessibility analysis. Additionally, we acquired filtered 3D NOESY experiments on complex samples consisting of labelled SGTA (1-69) dimer bound to unlabelled _UBL or unlabelled SGTA (1-69) dimer bound to labelled _UBL to identify distance restraints within the binding surface. These experiments involved using ^{15}N, ^{13}C-UBL4A_UBL or ^{15}N, ^{13}C-BAG6_UBL protein samples. The data generated additional ambiguous restraints which were used in our HADDOCK calculations. We have recorded 3D ^{13}C F1-filtered, F3-edited NOESY-HSQC experiments which allow selection of intermolecular NOEs. However, the signal: noise ratio of these experiments was low even on the most sensitive instrument available to us (a 700 MHz spectrometer equipped with cryoprobe). This was somewhat expected due to the increased T2 relaxation rate of the complex, which causes signal loss during the additional filter elements of the pulse sequence. Of the handful of resolvable peaks observed, we were able to obtain 3 (UBL4A complex) and 4 (BAG6 complex) useful intermolecular NOEs. The NOE restraints are listed in **Table 7**. In HADDOCK runs we used our SGTA (1-69) deposited ensemble, NMR structure of UBL4A_UBL (PDB accession number: 2DZI) and x-ray structure of BAG6_UBL (PDB accession number: 4EEW, A chain). Previously deposited complexes of Sgt2/Get5_ubl (PDB accession number 3ZDM, 4ASW and 2LXC) were

used in comparisons. The input in HADDOCK included CSP residues selected from the ^{15}N BAG6_UBL/SGTA (1-69) titration with peak shifts in HSQC experiments greater than one standard deviation above the mean and solvent accessibility above 25%. The final input for HADDOCK combined the following active residues: 22, 23, 24, 25, 26, 58, 62, 63, 65, 67, 76, 83, 85, 86, 87, 88. In the case of ^{15}N UBL4A _UBL/SGTA (1-69), residues were selected by the same method and final HADDOCK input included: 15, 16, 17, 18, 19, 21, 23, 49, 53, 54, 55, 56, 73, 75, 78, 79. The passive AIRs were assigned automatically.

Table 7: HADDOCK ambiguous interaction restraints (AIRs) obtained from filtered NOE experiments for the SGTA (1-69) and UBL4A_UBL and the SGTA (1-69) and BAG6_UBL complex.

AIR	**UBL4A detected residue**	**SGTA NOE**
1	L15	S26,E30,S31,C38,L39
2	L15	E30,E35,Q37
3	R55	T41
AIR	**BAG6 detected residue**	**SGTA NOE**
1	V64, V80, V84	E30,E35
3	R63,V64	D27,E40
4	R63,V80	V34

We analysed the four lowest energy HADDOCK clusters of SGTA (1-69)/UBL4A_UBL and SGTA (1-69)/BAG6_UBL from our experimentally restrained docking. The top-ranked first two clusters of SGTA (1-69) /UBL4A_UBL and SGTA (1-69) /BAG6_UBL were similar to each other and also the most comparable to the experimentally-derived N-terminal Sgt2/Get5_ubl complex structure. However, cluster 3 and 4 display differences in UBLs orientation on SGTA (1-69)/UBL4A_UBL and SGTA (1-69) /BAG6_UBL models (**Figure 31** and **Figure 32** respectively). HADDOCK statistics are shown in **Table 8** and **Table 9.**

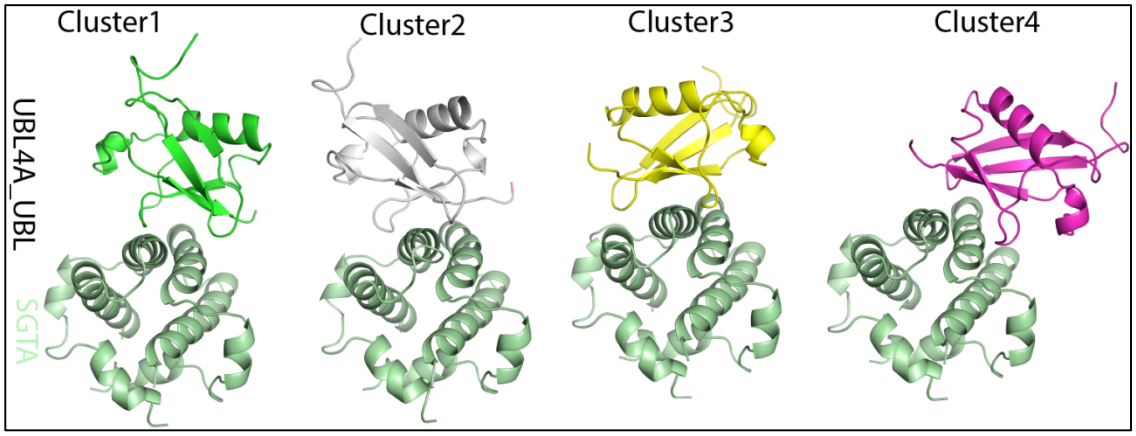

Figure 31: Four top scoring clusters of the SGTA (1-69)/UBL4A_UBL models generated by HADDOCK. The selected cluster 2 displayed the best Z-score and the lowest energy.

Table 8: The statistics of the top 4 clusters SGTA (1-69)/UBL4A_UBL HADDOCK models. The top cluster is the most probably according to HADDOCK. Its Z-score shows how many standard deviations from the average this cluster is located in terms of score (the more negative the better).

Cluster	HADDOCK score	Cluster size	RMSD (Å)	Van der Waals energy	Restraints violation energy	Buried surface área	Electrostatic energy (kcal/mol)	Desolvation energy (kcal/mol)	Z score
Cluster1	-117.1 +/- 8.7	19	10.8 +/- 1.0	-34.8 +/- 10.9	102.4 +/- 5.30	1605.7 +/- 107.8	--671.3 +/- 92.6	41.7 +/- 6.1	-1.5
Cluster2	-121.7 +/- 9.2	25	2.5 +/- 1.6	-37.4 +/- 5.3	115.5 +/- 57.13	1649.3 +/- 72.5	-668.0 +/- 15.9	37.7 +/- 6.4	-1.9
Cluster3	-103.1 +/- 4.2	19	11.1 +/- 0.5	-38.7 +/- 8.2	146.6 +/- 21.44	1510.3 +/- 67.6	-568.4 +/- 50.0	34.6 +/- 2.1	-0.5
Cluster4	-95.8 +/- 13.8	10	8.6 +/- 1.5	-29.1 +/- 4.4	108.0 +/- 34.91	1455.0 +/- 73.0	-579.0 +/- 37.0	38.3 +/- 4.2	-0.1

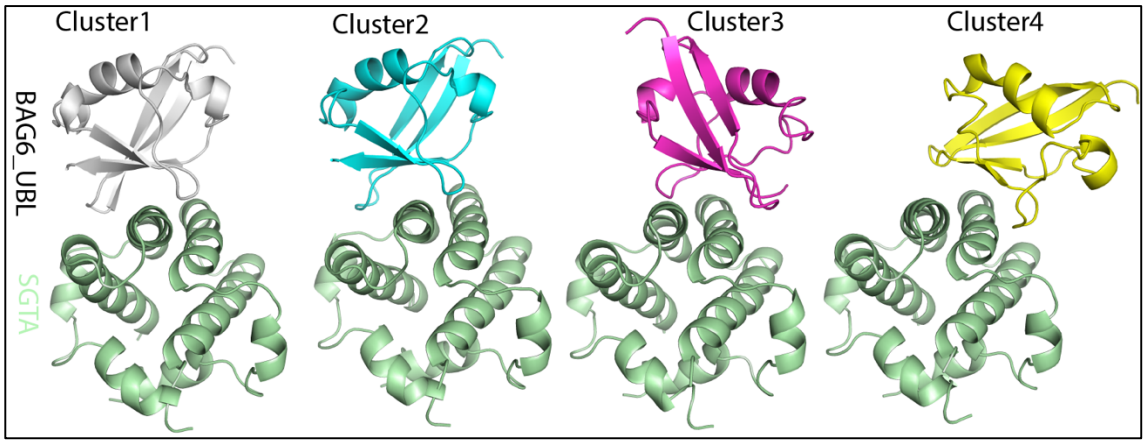

Figure 32: Four top scoring clusters of the SGTA (1-69)/BAG6_UBL models generated by HADDOCK. The selected cluster 1 displayed the best Z-score and the lowest energy.

Table 9: The statistics of the top 4 clusters of SGTA (1-69)/BAG6_UBL HADDOCK models. The top cluster is the most probably according to HADDOCK. Its Z-score shows how many standard deviations from the average this cluster is located in terms of score (the more negative the better).

Cluster	HADDOCK score	Cluster size	RMSD (Å)	Van der Waals energy	Restraints violation energy	Buried surface área	Electrostatic energy (kcal/mol)	Desolvation energy (kcal/mol)	Z score
Cluster1	--107.0 +/- 13.8	35	3.4 +/- 2.0	-34.9 +/- 4.7	86.8 +/- 24.16	1681.7 +/- 73.8	-573.5 +/- 101.9	34.0 +/- 7.8	-1.5
Cluster2	-102.6 +/- 5.7	19	5.0 +/- 0.7	-41.7 +/- 13.9	69.0 +/- 48.35	1694.6 +/- 71.9	-477.4 +/- 110.0	27.7 +/- 5.3	-1.1
Cluster3	-97.5 +/- 18.8	16	6.0 +/- 0.8	-30.3 +/- 2.9	68.0 +/- 19.44	1406.4 +/- 198.2	-557.3 +/- 98.0	37.5 +/- 3.9	-0.7
Cluster4	-95.0 +/- 5.4	57	13.6 +/- 0.2	-52.1 +/- 3.0	99.0 +/- 60.40	1754.6 +/- 176.3	-405.4 +/- 59.6	28.3 +/- 9.0	-0.4

The final HADDOCK models of SGTA (1-69)/UBL4A_UBL, chosen based on the lowest energy scores, and similarities of the orientation of conserved residues at the interface to homologue complexes are shown in **Figure 33 A** and **Figure 35** and SGTA (1-69)/BAG6_UBL complexes are shown in **Figure 33 B** and **Figure 36**. The structural analysis showed that the orientation of the SGTA binding helices relative to the UBL partner in the SGTA mammalian complexes was slightly different and rotated by approximately 45 degrees in comparison to the yeast complex, as highlighted in **Figure 33**. Since the binding was driven by electrostatic interactions, this reflected some differences in the negative charge distribution of the α2 central binding helices of SGTA (1-69) and Sgt2 N-terminal and also in the differences in positive charge distribution on the surface of the UBLs (**Figure 33**). Furthermore, Lys85 in the Get5_UBL corresponds to Glu12 in human UBL4A_UBL and therefore do not form the same salt bridge. Alternatively, UBL4A_UBL Lys66 and Lys46 can tweak the SGTA_NT domain around slightly to satisfy the charge interactions. From the SGT perspective, the four negatively charged amino acids are spread along the helix differently in Sgt2 and SGTA, with Sgt2 displaying two at each end (Asp28/Asp31 and Asp38/Asp42) and SGTA showing three at one end (Asp27/Glu30/Glu33 where the Lys85 interaction was lost) and one at the other (Glu40). The minor differences in charge distribution of important interacting residues can support the slight differences in binding mode that occur in mammalian and yeast systems (82).

The binding interfaces of SGTA (1-69)/UBL4A_UBL and SGTA (1-69)/BAG6_UBL HADDOCK models are similar however, the BAG6_UBL binding site displays a slightly narrower surface spanning 1–2 fewer amino acids at the N-terminal of α2 which is central to the interaction. All of the residues that were identified to be involved experimentally are found at the interface (**Figure 34**). The interface is comprised of the few hydrophobic residues on UBL4A_UBL (Ile61, Leu8) and BAG6_UBL (Leu24 and Val65), that docks against the Val34 and Thr145 from the SGTA (1-69) dimer. The conserved lysine residues on UBL4A_UBL are 6, 46, 42, 48, and 66 and their equivalents on BAG6_UBL 22, 58, 64, 80 make electrostatic contacts with the charged residues on SGTA (1-69), Asp27, Glu30, Glu33, and Glu40. The equivalent of Lys 46 is missing in the BAG6_UBL and is replaced by Gln62. The residues which split in the NMR titration experiments of SGTA (1-69) during the interaction with BAG6 are shown in Figure 37. Some of those residues are present at the binding interface and we suggest that they split because of differences in contact with BAG6_UBL on each SGTA protomer. However, some of them are located away from the binding interface and perhaps split due to differences in their local environment upon binding.

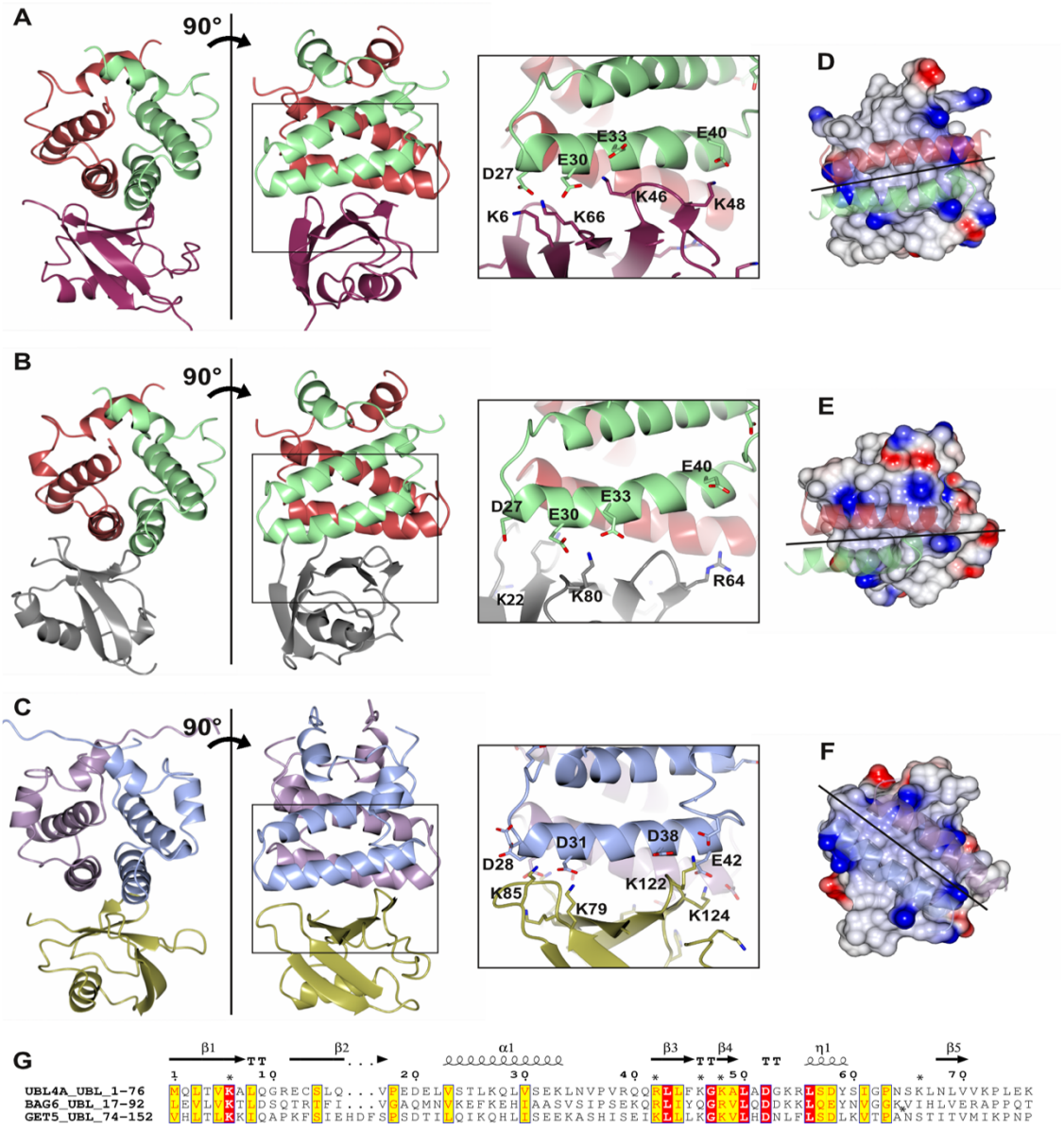

Figure 33: **Cartoon views of the lowest energy complexes of A) SGTA (1-69)/UBL4A_UBL, B) SGTA (1-69)/BAG6_UBL and C) Sgt2_NT/Get5_UBL (90) as calculated by HADDOCK from chemical shift perturbation data and intermolecular NOEs**. SGTA (1-69) is shown in red/ green and SGT2 N-terminal in lilac/blue; the UBLs of UBL4A, BAG6 and GET5 are in purple, grey and yellow respectively. The complexes are aligned according to the SGT domain and detailed views in boxes indicate specific residues involved in each interaction at one side of the SGT dimer. D), E) and F) show the vacuum electrostatic views of the equivalent aligned UBL domains. Surfaces are coloured according to electrostatic charge, exhibiting the positively charged residues (blue) that faciliate the interaction. The binding helices from the relevant SGT proteins are superimposed to show the relative orientations of binding. G) Sequence alignment between the three UBL domains with boxes indicating conserved residues and red highlighting sequence identity. Structural motifs are labelled across the top with 'TT' indicating a β turn. Sequences are numbered according to the UBL4A sequence. Graphic was generated using the ESPript 3.0 server (331). The conserved residues displayed on the interface are indicated with asterisks on the sequence alignment (82).

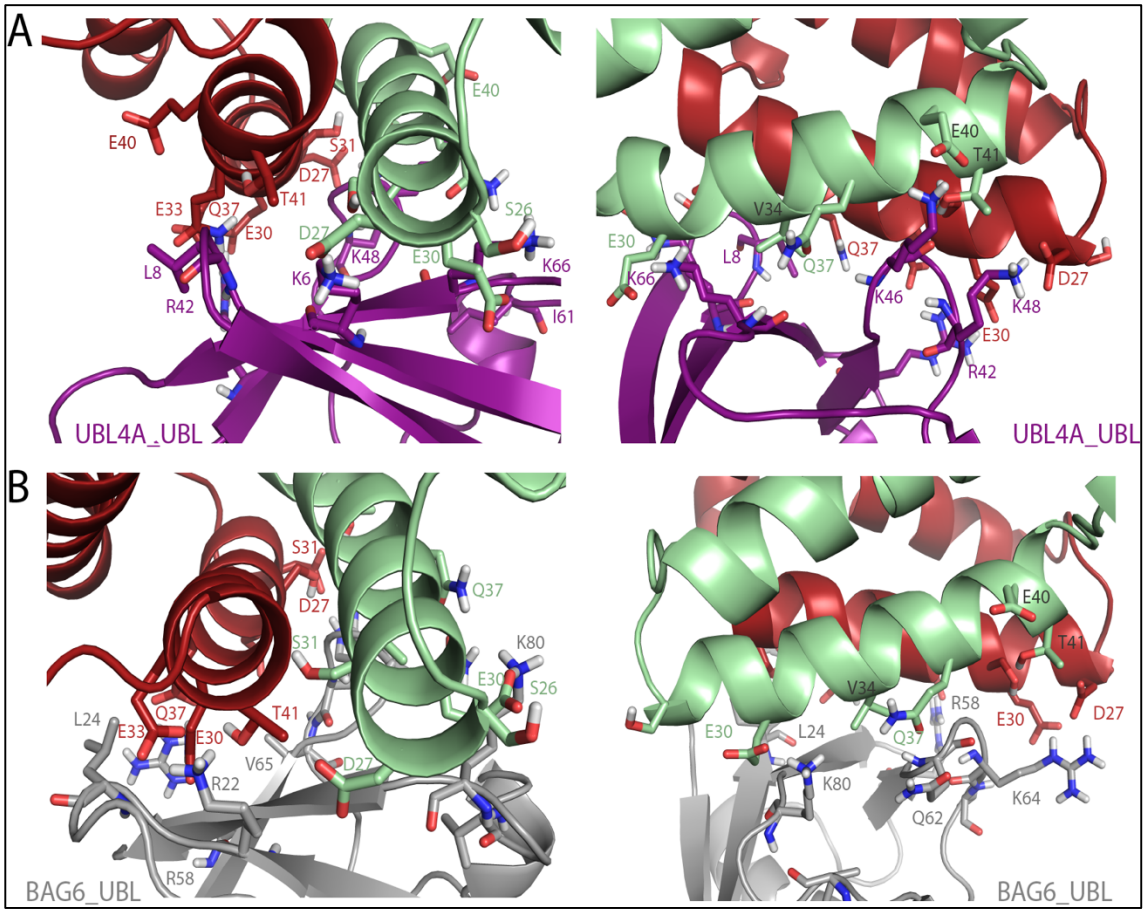

Figure 34: Two detailed views of the binding interface between SGTA (1-69) and (A) UBL4A or (B) BAG6_UBL. All residues at the interface are drawn as sticks.

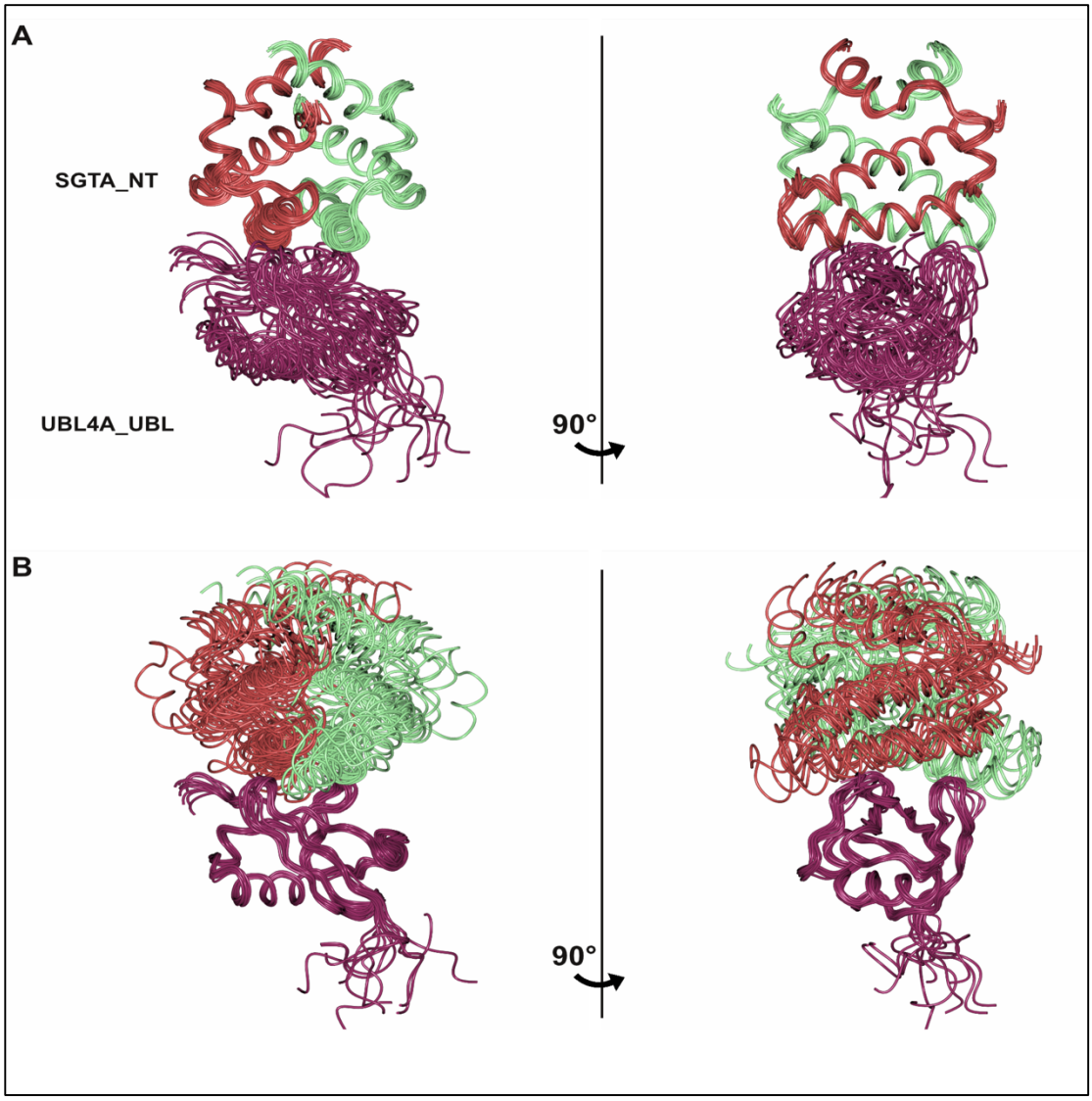

Figure 35: HADDOCK-generated structural ensembles showing the 10 lowest energy structures from the top-scoring cluster for the complex between SGTA (1-69) and UBL4A_UBL aligned according to SGTA (top) and UBL4A (bottom) structures (82).

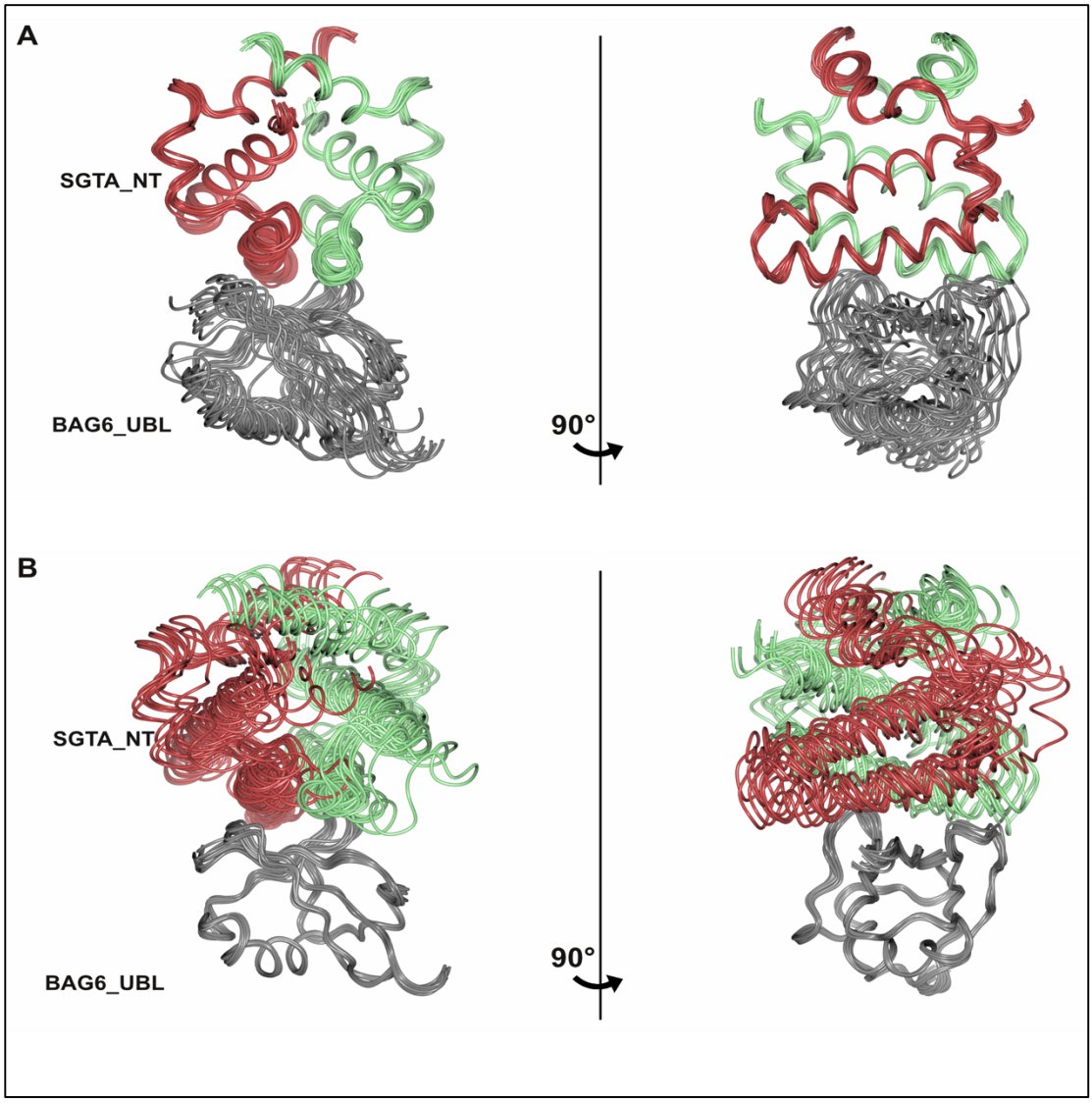

Figure 36: HADDOCK-generated structural ensembles showing the 10 lowest energy structures from the top-scoring cluster for the complex between SGTA (1-69) and BAG6_UBL aligned according to SGTA (top) and BAG6 (bottom) structures (82).

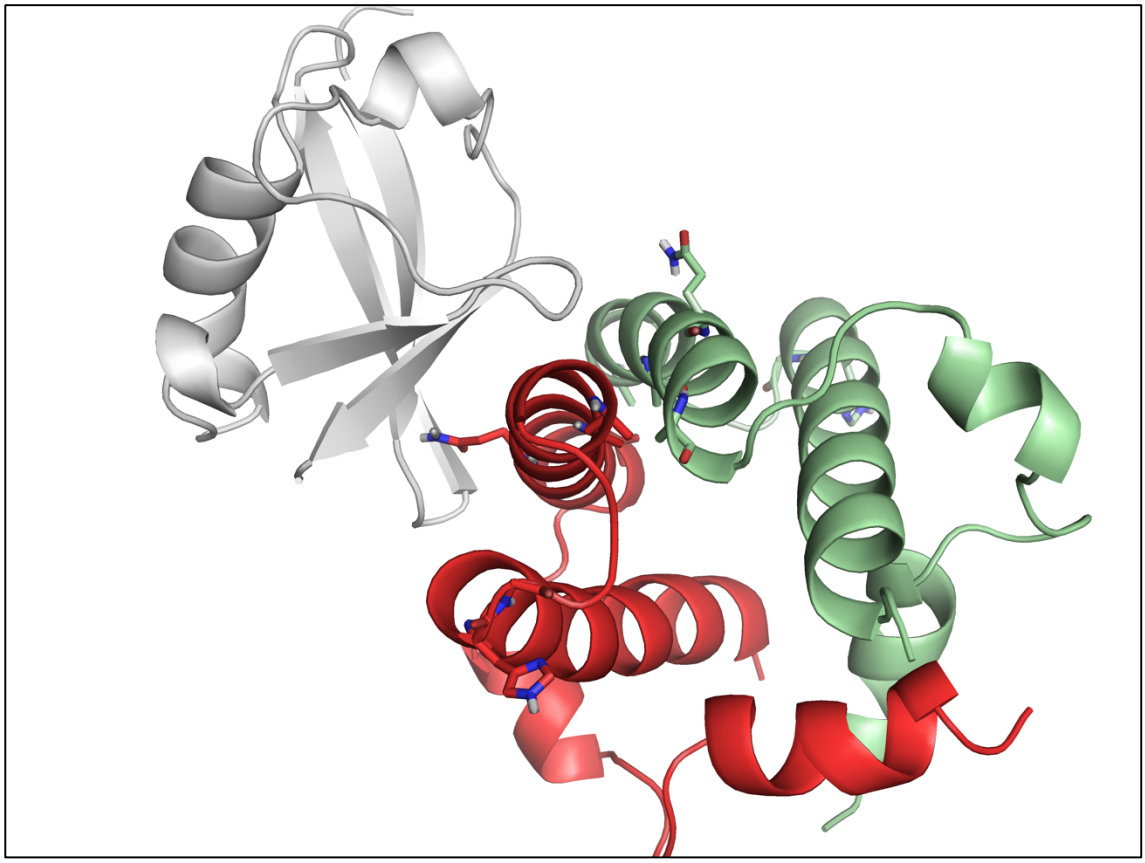

Figure 37: SGTA (1-69)/BAG6_UBL model indicating the residues which split in NMR titration experiments. These residues (His21, Gly22, Ala28, Ala35, Gln37 and Ala42) are shown as sticks.

3.7 BAG6_UBL and UBL4A_UBL competition for the same binding site on SGTA (1-69)

We also wanted to investigate whether the two different UBLs from the BAG6 complex can bind to the SGTA (1-69) dimer simultaneously or if they compete for the same binding site. We recorded an HSQC NMR experiment of ^{15}N-labelled BAG6_UBL titrations with unlabelled SGTA (1-69) up to a 1:2 molar ratio and analysed the characteristic NMR chemical shift perturbations that occurred upon binding (**Figure 38 C**). Next, we added unlabelled UBL4A_UBL to the protein mixture which resulted in the BAG6_UBL peaks shifting back towards their unbound state as the availability of SGTA_NT decreased (**Figure 38 D**). Additionally, we ran the reciprocal experiment by titrating unlabelled SGTA (1-69) into ^{15}N-labelled UBL4A_UBL (**Figure 38 A**) followed by the addition of unlabelled BAG6_UBL (**Figure 37 B**). This indicated that, in excess, the two UBL domains displace each other from

SGTA and compete for the same binding site and do not bind simultaneously. Under the conditions used, the binding affinity of both SGTA/UBL complexes appears similar as observed by analysing the 1D spectra, although it is possible that *in vivo* the relative affinities are influenced by additional cofactors or other domains of BAG6 or UBL4A.

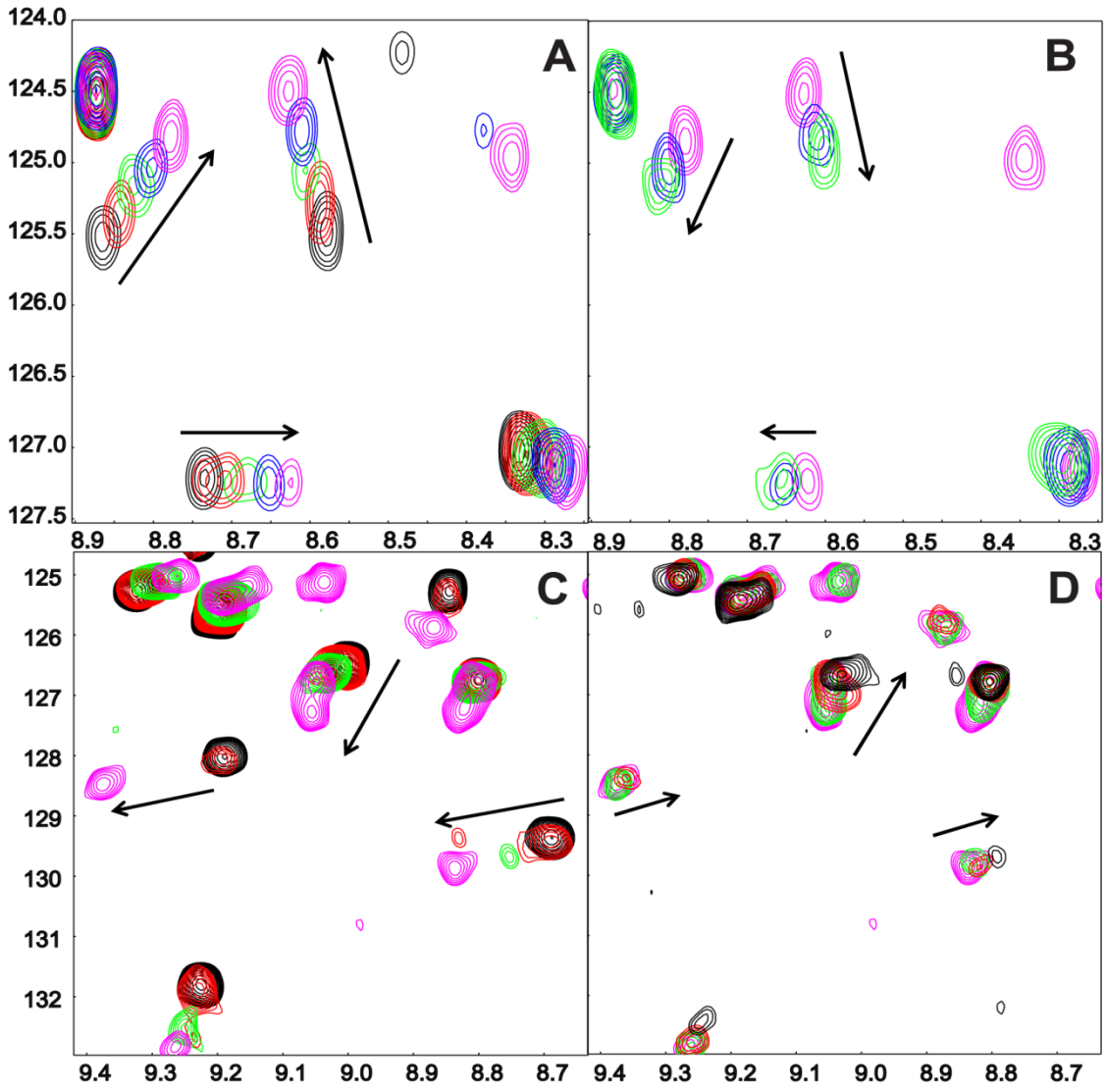

Figure 38: Detailed view of ^{1}H-^{15}N HSQC NMR titrations. A) ^{15}N-labelled UBL4A_UBL free (black) and titrated with increasing concentration of unlabelled SGTA (1-69) in red/green/blue/magenta. B) Saturation point of A with binding competed out through addition of unlabelled BAG6_UBL. C) ^{15}N-labelled BAG6_UBL free (black) and titrated with increasing quantities of unlabelled SGTA (1-69) (red/green/magenta). B) Saturation point of C with binding competed out through addition of unlabelled UBL4A_UBL. In B and D the bound UBL peaks move back towards their free state as their unlabelled competitor binds the SGTA (1-69).

3.8 Interaction study of SGTA (1-69) and UBLs by ITC, MST and SEC

To further characterize the interaction of SGTA (1-69) and UBL domains, we examined the stoichiometry and the binding affinity of the complexes empirically. In the case of UBL4A and SGTA (1-69) binding, we performed Isothermal Titration Calorimetry (ITC) and indicated that, just like in the equivalent yeast system, one UBL4A_UBL monomer interacts with one SGTA (1-69) dimer (N of 0.949 ± 0.013; ΔH of 5.9 ± 110 kcal/mol; **Figure 39**) albeit with a dissociation constant (K_d) of 3.0 μM, an order of magnitude weaker than in the yeast system. In addition, the binding affinity was verified using microscale thermophoresis (MST) which monitors changes in hydration shell, charge and size of molecules across a laser-induced temperature gradient (332). MST produced a comparable K_d value of 3.7 μM for the interaction between SGTA (1-69) and UBL4A_UBL (**Figure 39**).

In the case of BAG6_UBL, ITC experiments confirmed the interaction between SGTA (1-69) and BAG6_UBL and resulted in a sigmoid-shaped binding curve. However, the value of N deviated from 1 and an accurate K_d could not be determined as these data did not fit 1:1 stoichiometry. However, by assuming n of 1, we arrived at a Kd of 6.0 μM, a result fairly similar to the UBL4A/SGTA binding affinity (1-69) (**Figure 40**). Initially, we tried the same parameters as were used for UBL4A_UBL (50mM MES pH 6, 150mM KCl) with 50μM BAG6_UBL in the cell and 500μM SGTA (1-69) in the syringe at 30 ºC. Analysed data indicated an interaction (**Figure 40**), however, fitting the isotherms using either a single-site or two-site model still did not generate a binding curve at n=1 molar ratio. A set of experiments, designed to optimise the method was performed by extensively varying the following parameters alone and in tandem: buffer (20mM Tris pH 7.5, 100mM NaCl, 50mM MES pH 6, 200mM KCl and adding TWEEN at a range of concentrations between 0.04% and 0.1%); temperature (25 ºC and 30 ºC); protein concentration (adding 150-1500 μM of the sample in the syringe to the cell containing protein at concentrations between 15-150 μM and testing each protein in the cell and syringe). Unfortunately, none of these attempts yielded data that could be fitted appropriately. Similarly, MST experiments at different protein concentrations were performed as well as fluorescent labelling either SGTA (1-69) or

BAG6_UBL. A series of dilutions for unlabelled BAG6_UBL was prepared between 80 µM and 2.4 nM (or 20 µM to 0.6 nM respectively) in the presence of labelled SGTA (1-69) at a concentration of 0.4 or 0.2 µM. In addition, we tested the binding by using labelled BAG6_UBL at a concentration of 0.4 µM with SGTA (1-69) serially diluted from 400 µM to 24.4 nM. Unfortunately, no interaction between BAG6_UBL and SGTA (1-69) was observed in any of these experiments. This is despite the NMR chemical shift mapping and filtered NOESY data which indicate a bindingbetween SGTA_NT and BAG6_UBL and the competition titrations which qualitatively indicate a similar binding affinity for the two UBLs. The interaction between SGTA (1-69) and BAG6_UBL can also be seen by size exclusion chromatography as demonstrated in **Figure 41.**

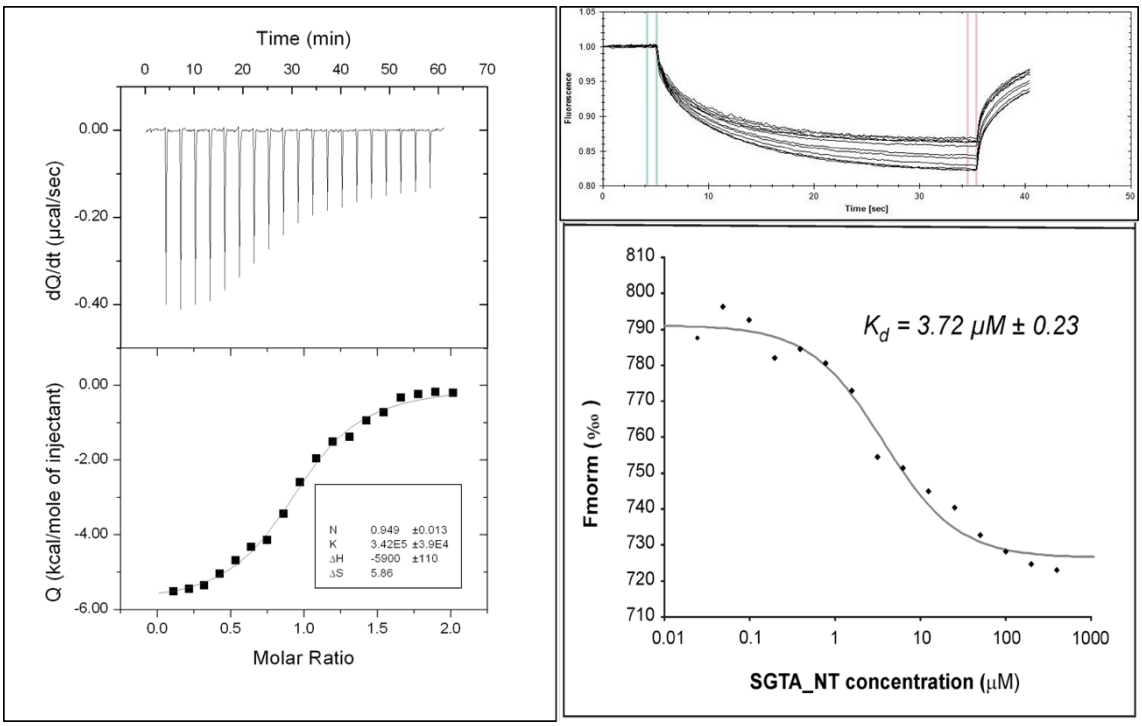

Figure 39: ITC data showing binding of one Ubl4A_UBL domain per dimer of SGTA (left) and MST data showing binding of Ubl4A_UBL domain to SGTA (1-69) (right). In the MST data, the top panel shows raw fluorescence signals during the reaction and the bottom panel shows normalised binding curves plotted against the SGTA concentration. Affinity constants (Kd), determined by ITC and MST respectively 3.0 µM and 3.72 µM.

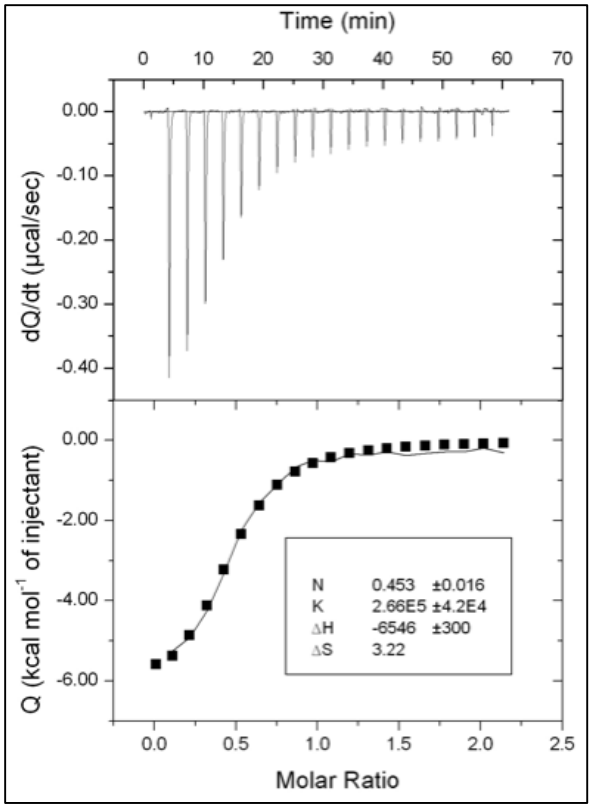

Figure 40: ITC data showing interaction between BAG6_UBL and SGTA (1-69) domains. If n is fixed at 1, Kd = 6.0 µM.

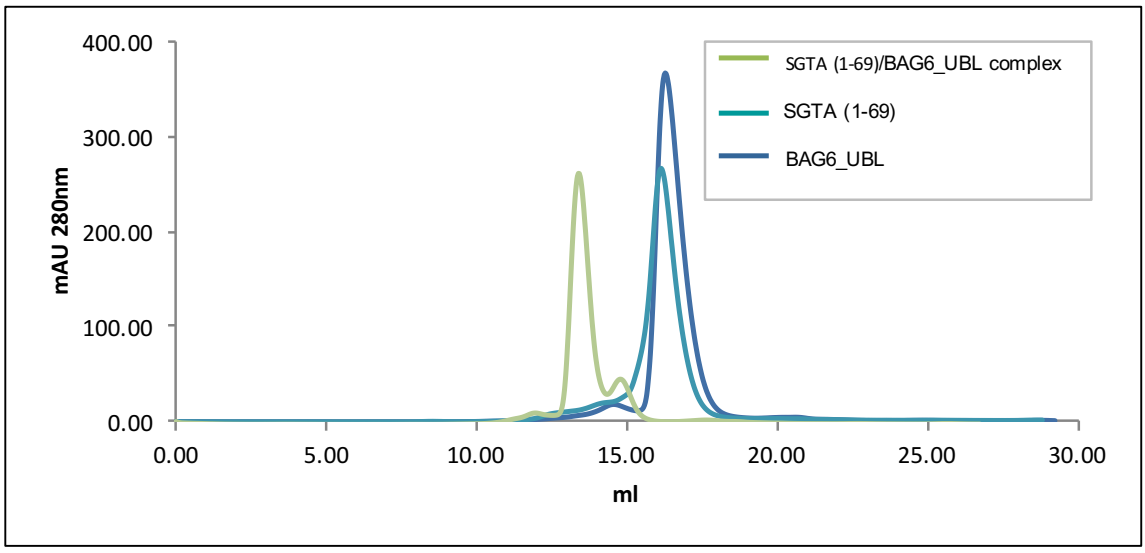

Figure 41: Chromatogram of size exclusion chromatography using Superdex 75 16/60 showing interaction between BAG6_UBL and SGTA (1-69). Elution volume is shown in ml.

3.9 Summary

The backbone and side chain assignment of the N-terminal SGTA dimer, together with NOE assignment were used to determine the structure of SGTA (1-69). Moreover, the backbone of the linker connecting SGTA (1-69) with the TPR domain has also been assigned. The structure of SGTA (1-69) aligned well with the yeast homologue with some minor structural differences, present at helix α3 and the loop between helices α2 and α3 which form part of the interaction site with UBLs.

The analysis of the interaction between SGTA and the UBL domains of BAG6 and UBL4A showed that they do not bind simultaneously but compete for the same binding site. This has been shown in a reciprocal titration experiment, where CSP peaks, where shifting back to the unbound state upon the addition of the competitive protein. Additional biophysical techniques were used to determine binding affinities of these complexes. ITC showed the binding affinity of 3.0 μM for SGTA (1-69)/UBL4A_UBL complex that is an order of magnitude weaker than in the yeast system (Sgt2/Get5_UBL). Furthermore, the binding affinity was confirmed using MST which showed a comparable K_d value of 3.7 μM. In the case of SGTA (1-69)/BAG6_UBL complex, ITC experiments confirmed the interaction but an accurate K_d could not be determined as these data did not fit 1:1 stoichiometry. Therefore, n was fixed to 1, giving a Kd of 6.0 μM that is fairly similar to the UBL4A/SGTA binding affinity (1-69). In addition, HADDOCK, used to generate models of SGTA-UBL complexes showed that the binding interfaces of the two models are similar as the BAG6_UBL binding site exhibited a slightly narrower surface spanning 1–2 fewer amino acids at the N-terminal of α2 which is central to the interaction.

Chapter 4. The structure and interactions of RNF126, E3 ubiquitin ligase with BAG6 and UBL4A

RNF126 E3 ligases comprises two zinc finger domains: an N-terminal zinc-finger domain and a C-terminal RING domain (75, 173). The N-terminal zinc-finger domain has been shown to be involved in the interaction with the UBL domain of BAG6 (75) while the RING domain is required for the E3 ubiquitin ligase activity (173). The work described in this chapter, is a structural and biophysical characterisation of RNF126 and its complex with the UBL domains of BAG6 and UBL4A. Furthermore, to expand our understanding of the ubiquitination and rescue triage mechanism of hydrophobic substrates, we study the competitive interaction between RNF126, SGTA and UBLs of BAG6 and UBl4A using ITC, MST and NMR.

4.1 Construct design of RNF126

The N-terminal region of RNF126 (residues 1–100) was predicted to be a theoretical Zinc finger domain and previously defined as the BAG6-interacting module (75). Therefore, this construct was cloned and used in the preliminary experiments (**Figure 42).** However, the secondary structure prediction using PsiPred showed random coil predisposition for residues 38–100 (**Figure 43**). Thus, a shorter construct was also produced (residues 1–40, hereafter named RNF126_NZF) and later used in structural and interaction studies. To further characterise RNF126 E3 ligase, the following RNF126 constructs were also made: full length RNF126 (named RNF126_FL), a construct containing a predicted hydrophobic region of RNF126 (named RNF126 142-218), and a RING finger domain (construct excluding the Serine rich region, residues 219-289), of which an NMR structure has previously been solved using the mouse sequence (PDB accession 2ECT). The boundaries for the constructs were chosen based on secondary structure prediction (**Figure 43**) and a primary sequence alignment of RNF126 from several species to identify conserved domains (**Figure 44**). The alignment shows that the constructs share two distinct regions of similarity present at their N- and C- termini which are also similar to the equivalent regions in RNF115, a homologue of RNF126 (**Figure 44**). All constructs were cloned

into a pET28 vector which includes a N-terminal thioredoxine A tag and a hexahistidine tag. Furthermore, an N-terminal GFP-fusion of RNF126_NZF was also constructed to be used in MST experiments.

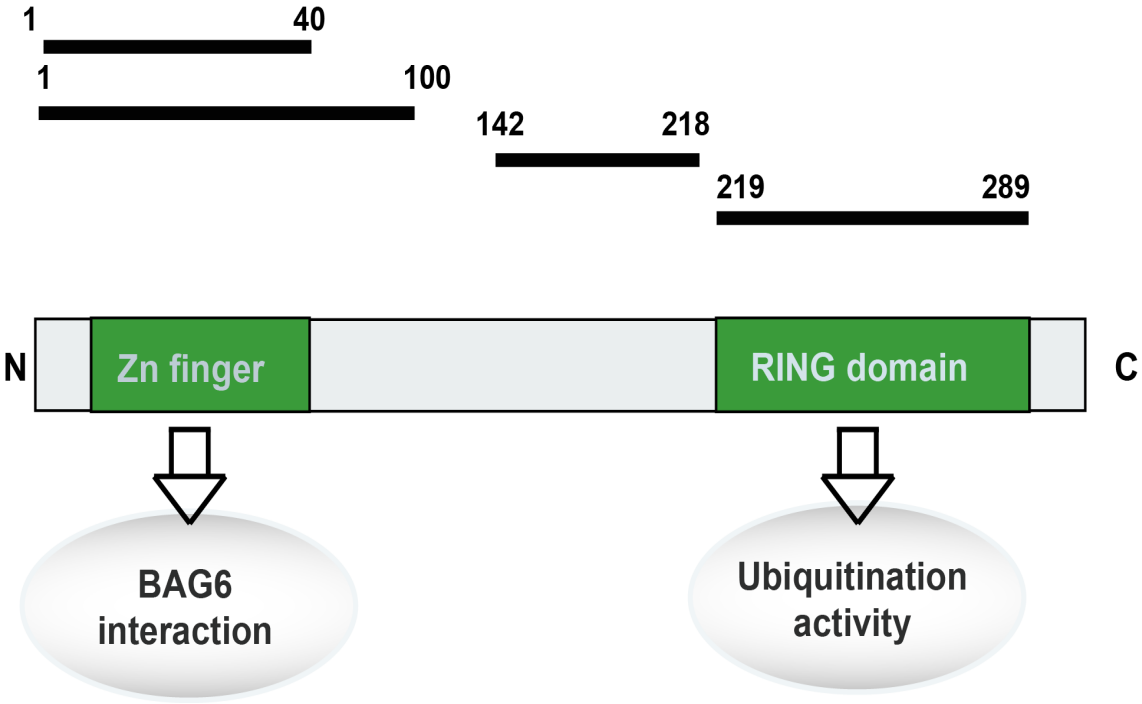

Figure 42: **Construct design of RNF126.** The boundaries of RNF126 constructs used in this project are shown with residue number of N and C termini. The constructs were designed based on previously reported results and new data, including secondary structure prediction and sequence alignment. The two distinct regions of RNF126 include a C-terminal RING domain, responsible of the ubiquitination activity and a N-terminal Zn finger region known as the Bag6_UBL binding platform.

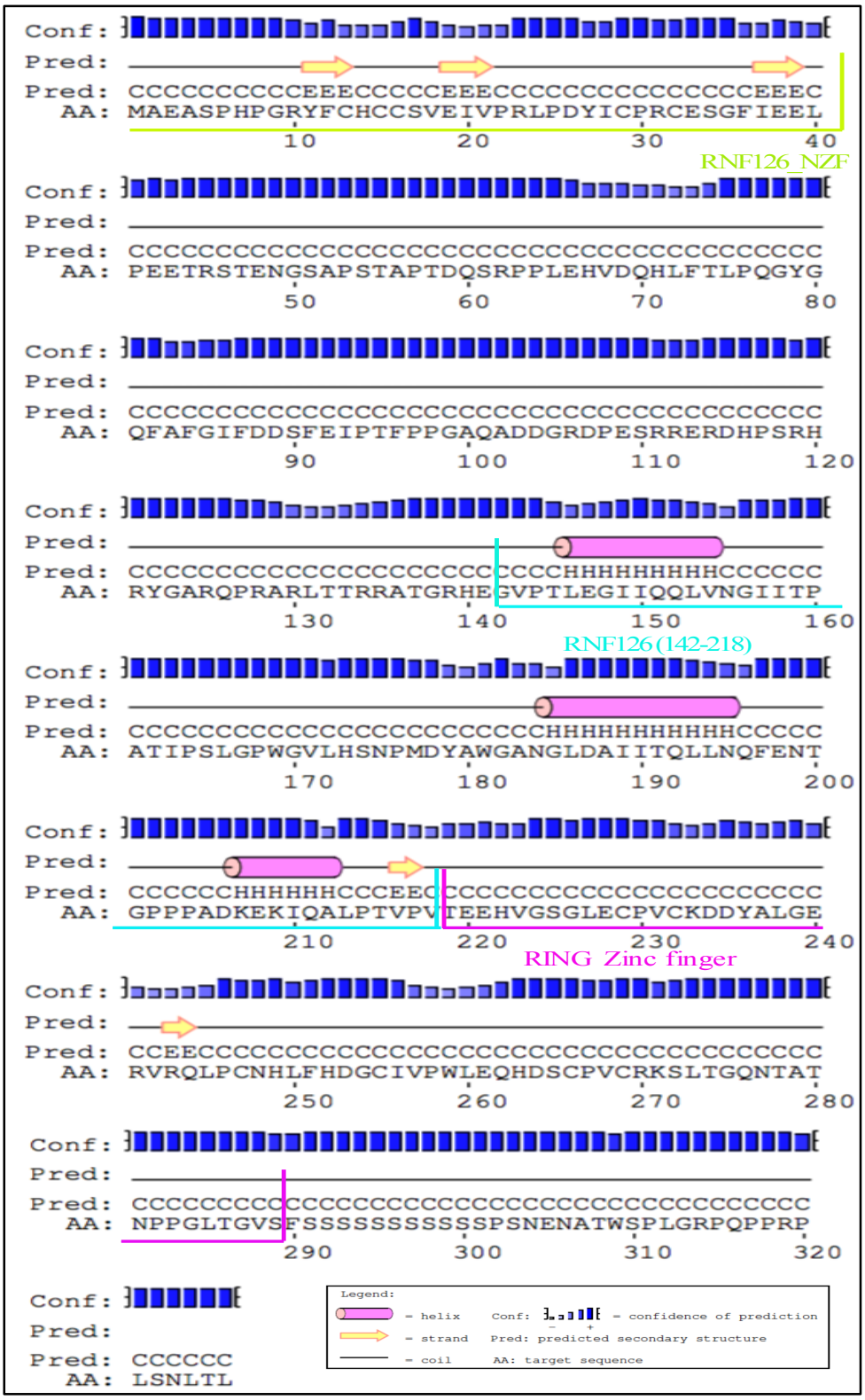

Figure 43: **PSIPRED secondary structure prediction and construct boundaries of the human RNF126.** The software predicts β-sheet features (yellow arrow) in the N-terminal region (residues 1-40) followed by a largely disordered section and hydrophobic region (α-helices shown as pink cylinders) for resides 142-218. The C-terminal region displaying β-sheet predisposition belongs to the RING domain residues (219-289), previously characterised in mouse.

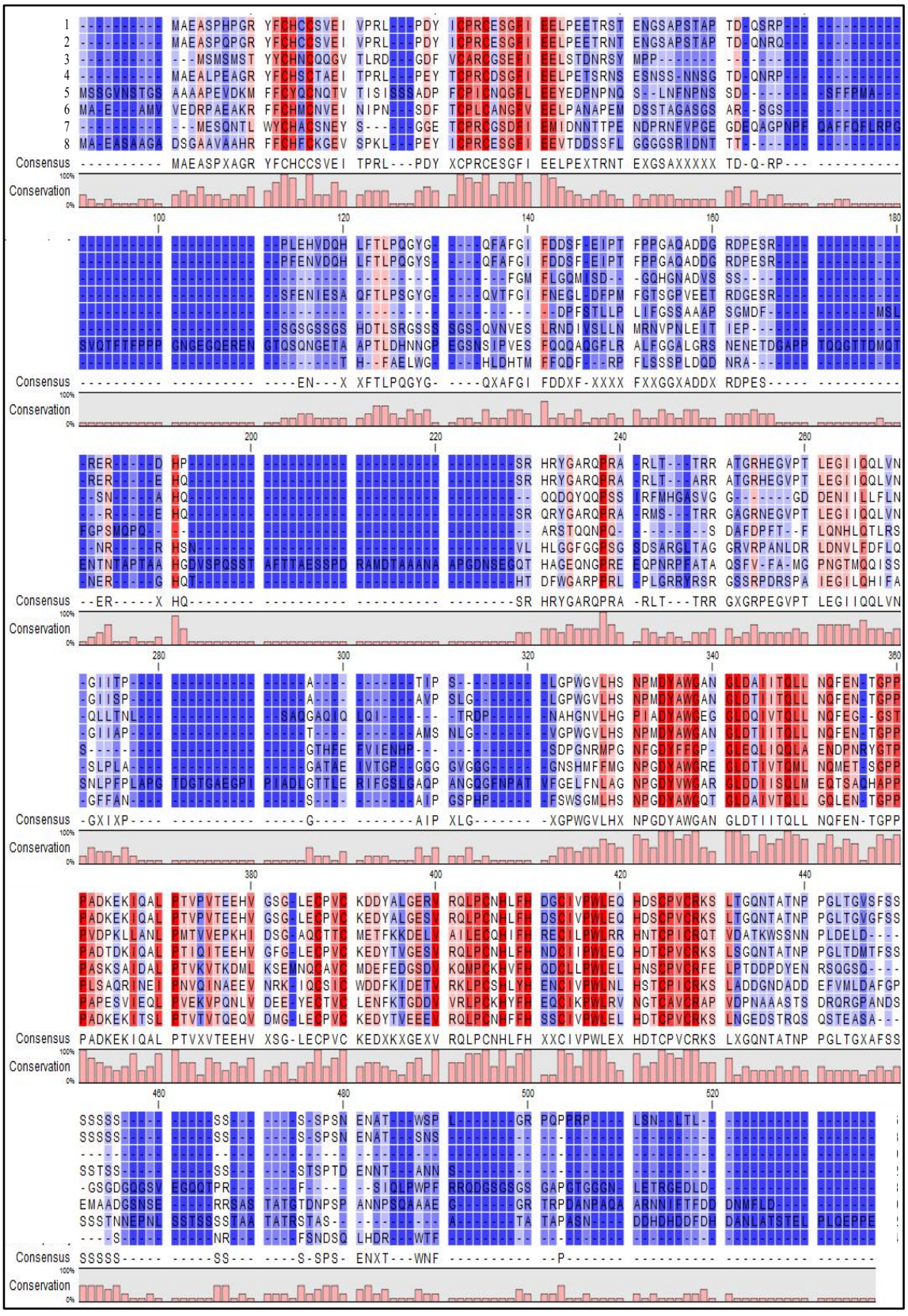

Figure 44: ClustalW multiple sequence alignment of RNF126. 1) Homo sapiens, against 2) Mus musculus, 3) Brugia malayi, 4) Xenopus laevis, 5) Arabidopsis thaliana, 6) Drosophila melanogaster, 7) Schizosaccharomyces japonicas and 8) RNF115 (Homo sapiens). The level of conservation is rated from 0%(blue)-100%(red). Primary sequence alignments were performed using PSI-BLAST and ClustalW and the figure was created using CLC Sequence Viewer Software (CLC Bio).

4.2 Expression and purification of RNF126

During the overexpression of RNF126_NZF, the cultures were supplemented with 10 µM $ZnCl_2$ to support protein folding and solubility. Enrichment of expressed protein was achieved by affinity purification and size exclusion chromatography (**Figure 45**). In the case of RNF126_NZF with a GFP tag, expression and purification including the tag removal step were also monitored by the GFP fluorescence. (**Figure 45 B**). A typical purification profile shows a diffuse RNF126_NZF sample on SDS-PAGE, despite denaturation and the presence of a reducing agent in the loading dye (**Figure 45 A and C**). However, the protein in solution is fairly stable and produces sufficient yields for the required experiments. Moreover, we have noticed that other zinc finger proteins in the lab have a similar SDS-PAGE profile.

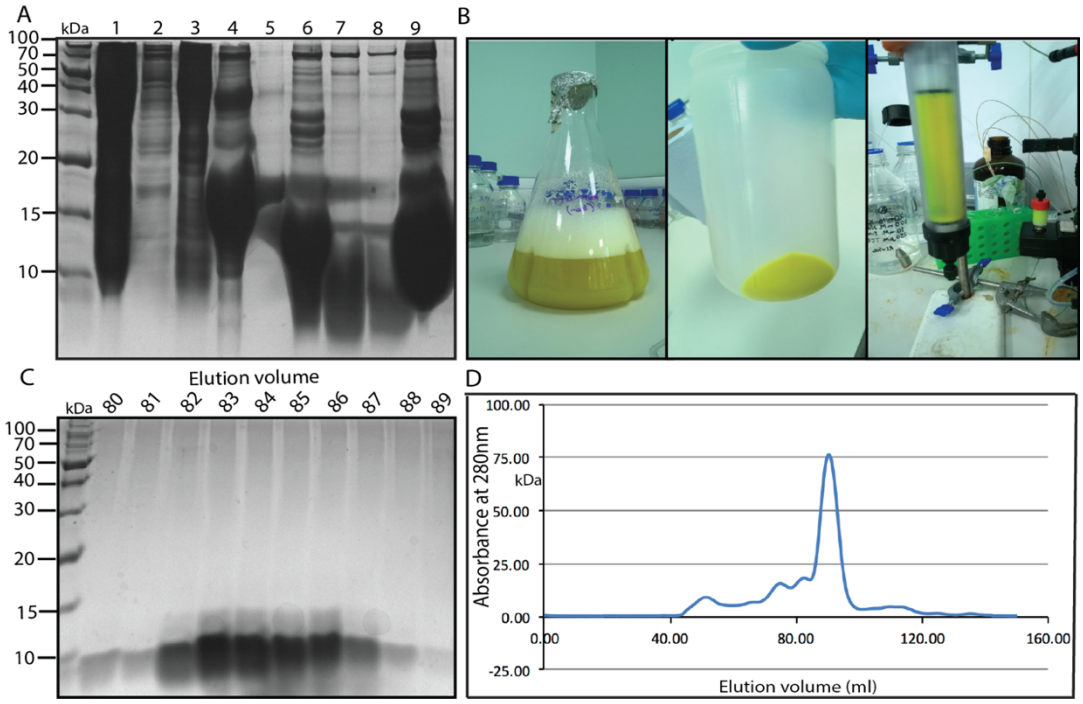

Figure 45: **Purification of RNF126_NZF**. A) SDS-page analysis of purification and the removal of the fusion/histidine tag by affinity chromatography of RNF126_NZF 1) uninduced cells control 2) flow through after column loaded with the cell lysate, 3) wash with 5 CV of wash buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 µ M TCEP), 4) elution with 2 CV of elution buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 µ M TCEP, 300mM Imidazole), 5) additional elution step, 6) total cleavage reaction, 7) flow through after column loaded with total cleavage reaction (cut RNF126_NZF), 8) wash with 2 CV of wash buffer 9) elution with 2 CV of elution buffer (uncut protein and tag). B) Expression and purification of RNF126_NZF_GFP. C) SDS-page analysis and D) Chromatogram of size exclusion chromatography using Superdex 75 16/60. Lanes are identified with elution volume in ml.

4.3 Backbone and sidechain assignment of RNF126_NZF by NMR

4.3.1 Backbone ^{1}H, ^{13}C, and ^{15}N Chemical Shift Assignments of RNF126_NZF

The following triple resonance experiments, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB were acquired at 700 MHz using the shorter construct, RNF126_NZF, to facilitate the full backbone and side-chain assignments. ^{15}N HSQC peaks were picked manually and the NH assignment for all residues except for five prolines was completed using CCPN Analysis. The full ^{15}N HSQC spectrum with assigned peaks is shown in **Figure 46**. The chemical shifts of ^{1}HN, ^{15}N, $^{13}C\alpha$ and $^{13}C\beta$, and $^{13}C'$ were deposited in the BMRB (Accession Number: 25913). Validation Software (AVS) was run automatically after we deposited the assignments and the full report can be seen in Appendix. The $^{1}H^{15}N$ HSQC spectrum of RNF126 (1-100) indicated a partially folded protein and backbone assignment of residues 1–66 showed that residues 1–40 constitute the structured region while the remainder exhibits the chemical shift dispersion typical of disordered proteins. A shorter, RNF126_NZF construct displays comparable spectra to the longer version, showing that the additional truncation has no effect on the folding of the zinc finger region (**Figure 47**). Therefore, this construct was used in full backbone and sidechain assignments, structure calculations and interaction studies.

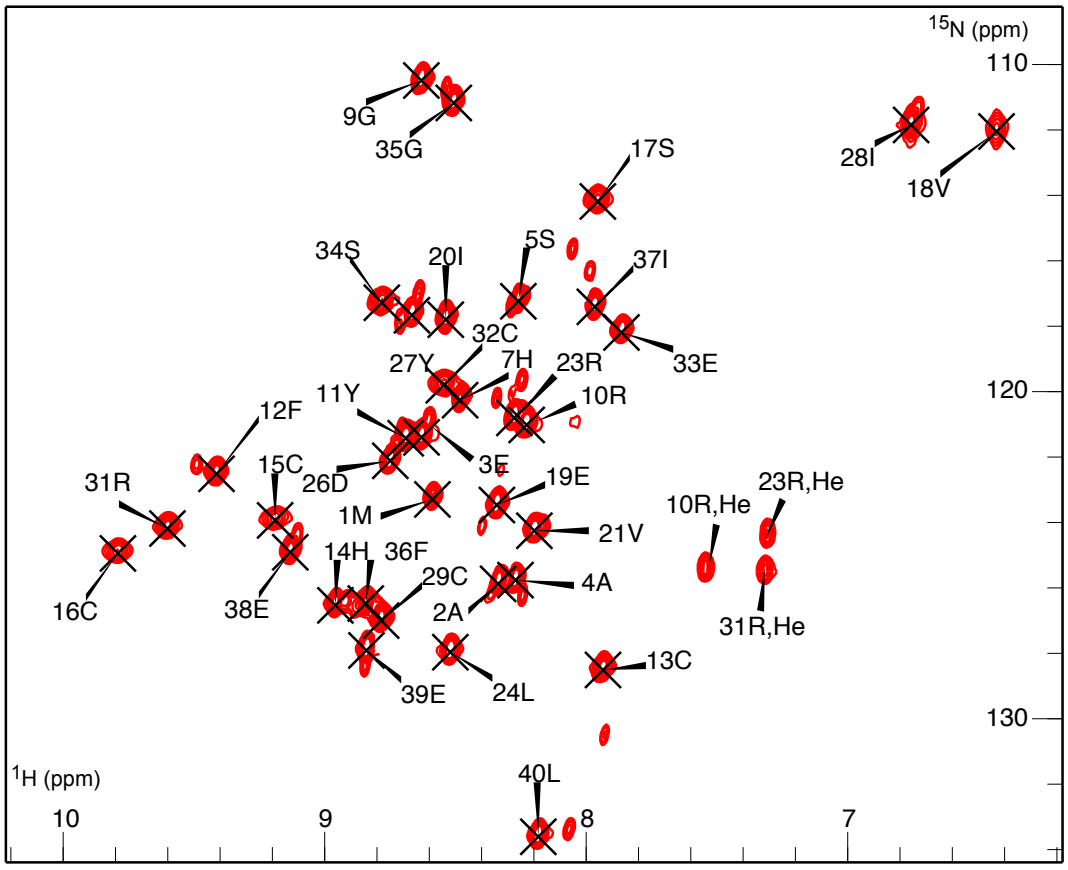

Figure 46: **^{1}H-^{15}N HSQC spectra of ^{15}N-labelled RNF_NZF (1-40).** Assigned peaks from region 1-40 are labelled.

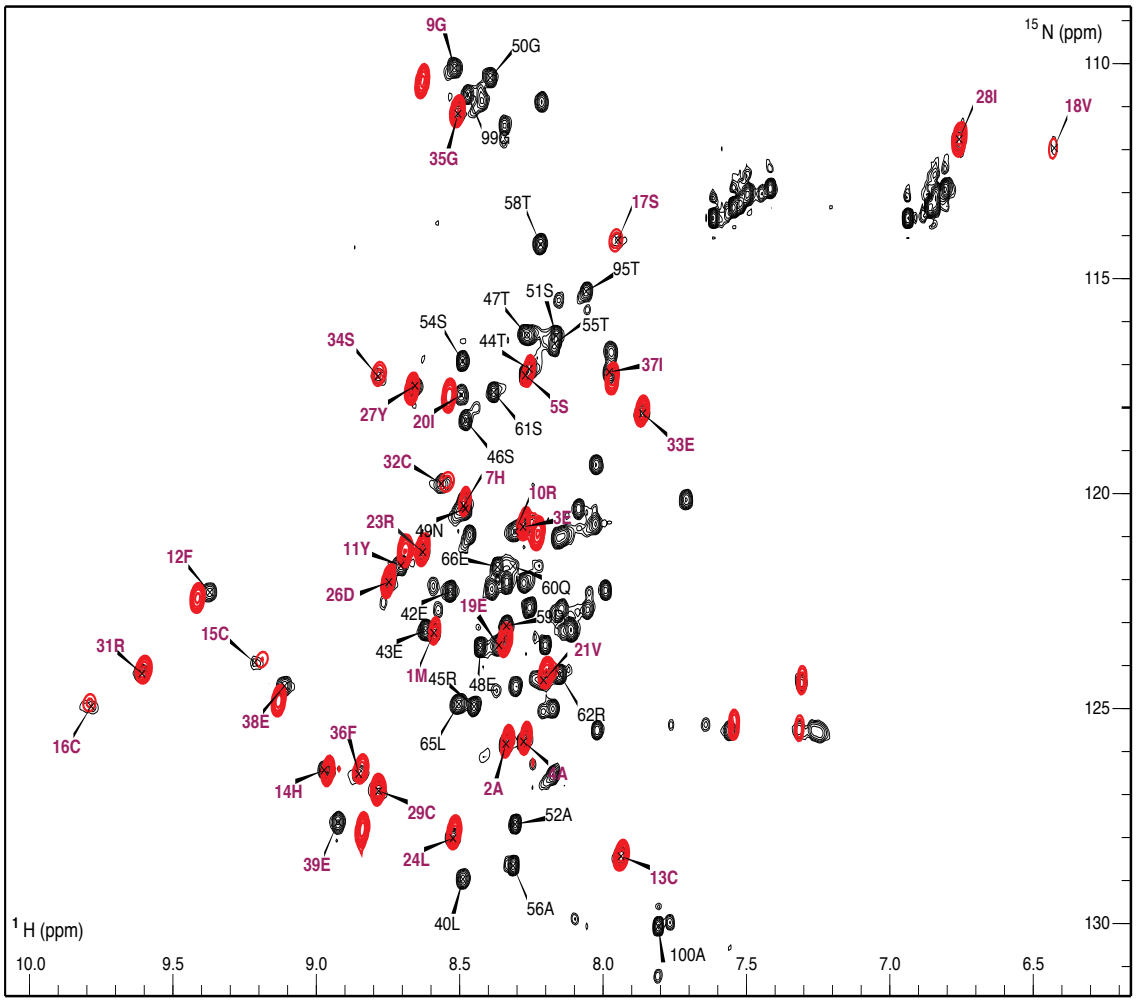

Figure 47: **An overlay of ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled RNF126 (1-100) and (1-40).** ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled RNF126 (1-100) in black and the shorter construct RNF126_NZF (1-40) in red. Assigned peaks from region 1-40 are labelled in red. The comparison of the peaks chemical shifts indicates that the folding of longer and shorter RNF126 constructs is conserved.

4.3.2 Sidechain assignment of RNF126_NZF

Side chain resonances were assigned using 3D HcCH-TOCSY and hCCH-TOCSY together with ^{13}C HSQC and 2D NOESY experiments. The chemical shifts were deposited, together with the backbone assignments (BMRB Accession Number: 25913, and the validation report is shown in appendix). Overall, chemical shifts assignments for all residues were assessed as consistent except for 24 Pro which was marked as suspicious because of the unusual shift of atom HG2 = 0.816, where the expected shift should be 1.93 (Std = 0.31). However, after an additional assessment the resonance was ascertained to be correct. In total there were 183 chemical shifts of ^{13}C, 38 chemical

shifts of ^{15}N and 261 chemical shifts of ^{1}H. Overall, 85% of side chain H were assigned, including 97% of Hα and 90% of Hβ (**Figure 48**). The remaining residues were unassigned due to the poor quality signal to noise ratio in this area of the spectrum and spectral overlap.

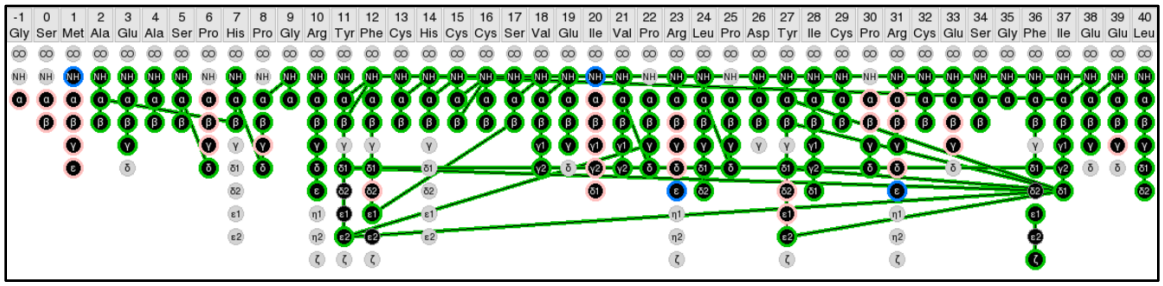

Figure 48: **Chemical shift assignment for heavy atoms of RNF126_NZF.** The completed assignment for the heavy atom is shown in black circles, highlighting with where ^{1}H-^{15}N HSQC (blue), ^{13}C-HSQC (pink) and 2D NOESY were used. Figure was created using CCPN Analysis. A line shows the NOE assigned between atoms.

4.3.3 Chemical Shift Index Analysis

The secondary structure of RNF126_NZF was predicted from the chemical shifts of C′, Cα, Cβ and Hα atoms using the chemical shifts index (CSI) method (304, 333) in CCPN Analysis (**Figure 49**). These chemical shifts are combined to give the CSI values indicating the secondary structure type; -1 indicates alpha-helix and +1 beta-strand. The chart is the graphical representation of the secondary structure prediction for each residue indicating the difference of recorded chemical shift from (sequence adjusted) random coil chemical shift. The structure shows three main β-strands, which is in agreement with the PSIPRED secondary structure prediction, and one additional very short β-strand.

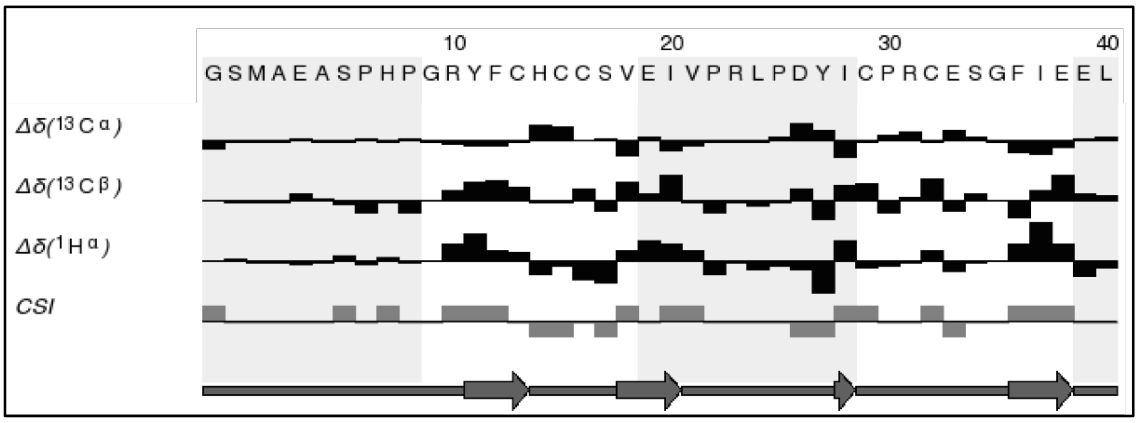

Figure 49: **Secondary structure analysis of RNF126_NZF using chemical shift index of Cα, Cβ, C' and Hα.** CSI values indicate the secondary structure type; -1 indicates α-helix and +1 β-strand. Residues with chemical shifts indicating β-strands are shown as arrows.

4.3.4 TALOS

TALOS+ was used to determine the RNF126 dihedral angle restraints from the chemical shift assignment using standard methods (305). In total there were 34 pairs of dihedral angles (Φ+Ψ) derived by TALOS and used in structure determination (section 4.4).

4.3.5 NOE Distance constraints in intra- and inter-residue assignments

NOE distance restraints were derived from different NOESY experiments: 2D NOESY (in 90% H2O and 10% D2O) for RNF126_NZF. Manual assignments of NOESY cross peaks were performed using CCPN starting with the identification of amino acids, with a characteristic pattern of cross signals, i.e. of glycine, alanine, threonine, valine, leucine and isoleucine, and continuing to the neighbouring residues. The NOEs were assigned for proton cross peaks between β-strands and for cysteines coordinating a zinc atom. All NOE distance restraints were grouped into several classes and a final count includes 719 NOEs including 326 intra-residue signals,119 sequential, 55 medium-range and 219 long-range distances.

4.4 NMR structure of RNF126

The solution structure of RNF126_NZF was solved in collaboration with Dr Santiago Martinez-Lumbreras. The structure calculation was performed using the ARIA2 program (306) providing spectra with chemical shift lists, dihedral constraints (ϕ and Ψ angles) extracted from the chemical shift values using the TALOS+ program and manually assigned NOE restraints. In the initial Aria runs we also allowed for automated NOE assignment. In each of many cycles of structure calculations with water refinement, we inspected the restraints violations and manually corrected them. The final Aria run generated 100 conformers in the final iteration and the 20 best structures with lowest restraint energies were selected for a final step of refinement in water. The coordination sphere of the zinc ion in the structure of RNF126 was obtained in a first round of calculations using only NOE based distance restraints and, once it was confirmed, the zinc coordination was added as an additional restraint in structural calculations by ARIA2. The final structure is in agreement with the structure prediction from PSIPRED and CSI and the structural assembly was deposited in the PDB after validation with PROCHECK-NMR (PDB Accession Number: 2N9O). The statistical parameters are presented in Table 10 and the structure is shown in **Figure 50** A. The folded domain comprises residues Arginine 10 to Leucine 40. This region contains a three-stranded short antiparallel β-sheet (β1 = Y11-C13; β2 = V18-I20; β3 = I37-E39) and a zinc finger comprising four cysteines; one pair from the β1-β2 short loop (C13; C16) and the other pair from the long structured region between strands β2 and β3 (C29; C32). Identification of cysteines that coordinate zinc can be supported by the analysis of chemical shift data of ^{13}Cα and ^{13}Cβ. These can be compared to previously documented characteristic shifts of zinc-ligated cysteines (334). In RNF126_NZF structure the first zinc coordination shell contains four cysteine side-chains in tetrahedral conformation (**Figure 50** B) whose chemical shift values for ^{13}Cα (~59 ppm) and ^{13}Cβ (~31 ppm) compared well to predicted shifts of zinc-binding cysteines (334). The second coordination shell is produced by the formation of three hydrogen bonds between the cysteine sulphur atoms (C13, C29 and C32) and the amide group of the residue at position +2 (C15, R31 and S34 respectively) (Figure 50 B). All parameters of the zinc coordination are presented in **Table 11.** A stable hydrophobic core is formed by the long

structured loop β2-β3 folded against one side of the β-sheet that buries several non-polar residues. The opposite side of the β-sheet is solvent exposed and comprises some glutamic acid side-chains and aromatic rings. This creates a putative platform for protein-protein interaction (Figure 50 C).

Table 10: NMR and refinement statistics for the final 20 structure ensembles of RNF126_NZF. The asterisk means data were obtained from PROCHECK-NMR.

NMR distance and dihedral restraints		
Distance restraints		
Total NOE	719	
Intra-residue	326	
Sequential (|i-j| = 1)	119	
Medium-range (1 < |i-j| < 4)	55	
Long-range (|i-j| > 5)	219	
TALOS derived dihedral restraints		
Total dihedral restraints (Φ+Ψ)	34	
Structure statistics		
Violations (mean and s.d.)		
Number of violated distance restraints per structure (>0.15 Å)	1.2 ± 0.9	
Max. distance restraint violation (Å)	0.19	
Number of violated dihedral angle restraints per structure (>2°)	5.1 ± 0.6	
Max. dihedral angle restraint violation (°)	2.8	
Ramachandran Plot analysis (%)*	*(1–40)*	*(10–40)*
Residues in most favoured regions	77.4	84.8
Residues in additionally allowed regions	19.8	15.2
Residues in generously allowed regions	0.9	0.0
Residues in disallowed regions	1.8	0.0
Derivation from idealized geometry		
Bond length (Å)	0.0041 ± 0.0001	
Bond angles (°)	0.58 ± 0.01	
Averages RMSD to mean structure (range)	*(10–40)*	
Backbone (Å)	0.19 ± 0.05	
Heavy (Å)	0.73 ± 0.11	

Table 11: **Geometry of the RNF126_NZF zinc coordination centre**. Averaged distance (Å) and angular (º) values for first and second coordination shell across the NMR ensemble. Two hydrogen bonds are established in the outer shell of coordination on Cys13 and Cys29 and possibly Cys32.

	Cys 13	**Cys 16**	**Cys 29**	**Cys 32**
First coordination shell				
Sγ-Zn distance	2.29 ± 0.01	2.29 ± 0.01	2.29 ± 0.01	2.29 ± 0.01
Cβ-Sγ-Zn angle	110.0 ± 0.3	109.9 ± 0.3	105.7 ± 0.5	110.1 ± 0.3
Second coordination shell				
Sγ-HN (i+2) distance	2.42 ± 0.08	5.24 ± 0.04	2.49 ± 0.04	3.08 ± 0.07
Sγ-H-N (i+2) angle	130.0 ± 5.3	91.1 ± 0.8	150.1 ± 1.9	127.3 ± 4.1

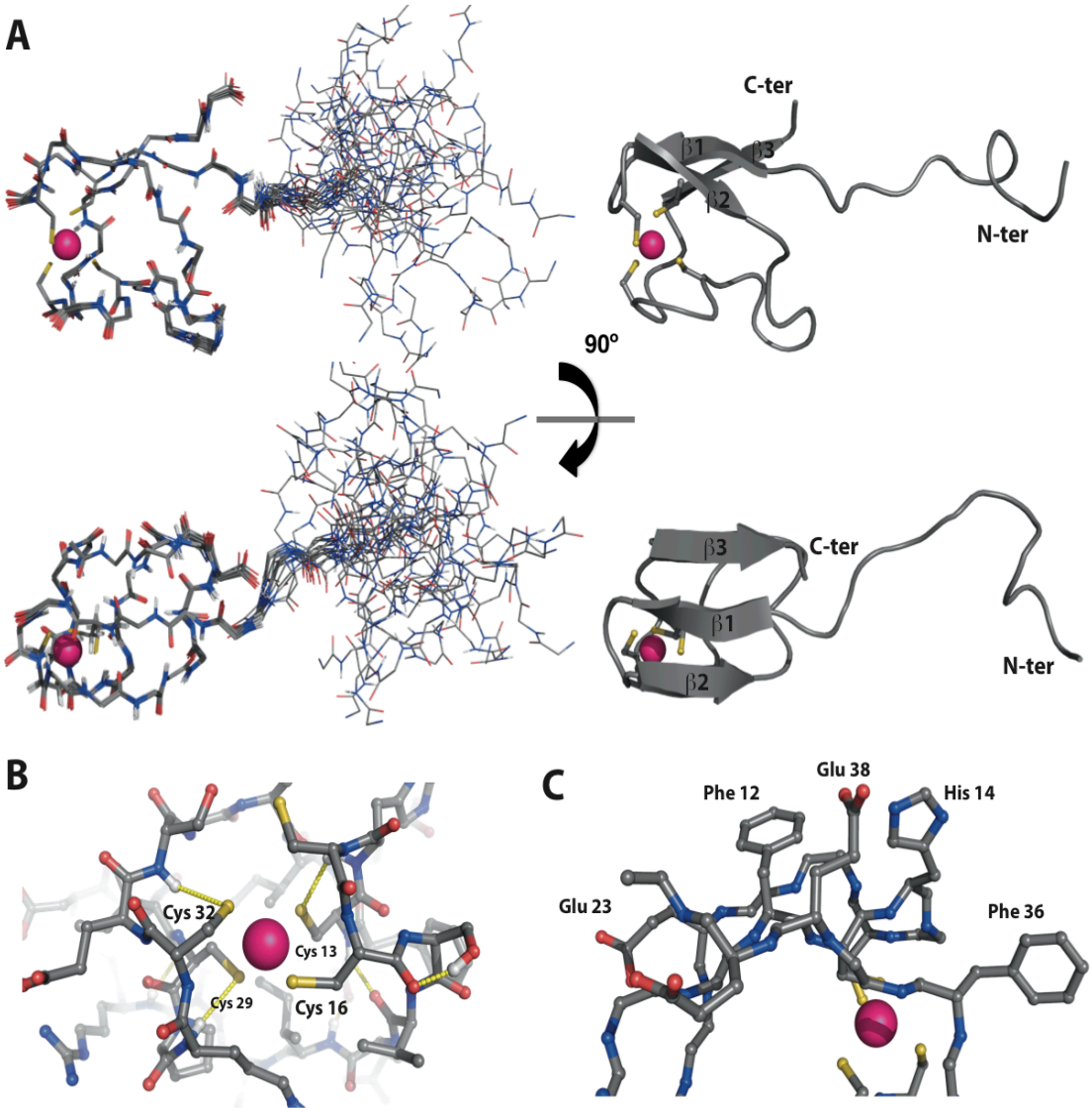

Figure 50: **NMR solution structure of RNF126_NZF.** A) Two views of ensemble backbone (left) and cartoon diagram (right) showing the 20 lowest energy ARIA-calculated structures as deposited in the PDB (Accession number: 2N9O). B) Detailed representation of the zinc finger coordination shell showing the cysteines coordinating the zinc cation; hydrogen bonds in the second coordination shell are shown as yellow dashed lines. C) Detailed view of the solvent-exposed β-sheet interface. Polar and hydrophobic residues present at this interface are illustrated using ball-and-stick. Carbon atoms are coloured in grey, oxygen atoms in red, nitrogen atoms in dark blue and sulphur atoms in yellow; zinc cation in magenta (174).

4.5 Binding study of RNF126_NZF with UBL domains of UBL4A and BAG6

To characterise and compare the interactions between RNF126 and UBL domains, we used a range of biophysical methods including chemical shift perturbation (CSP), ITC, MST and mutagenesis.

4.5.1 Chemical Shift Perturbations of RNF126_NZF interaction with BAG_UBL

CSP studies were carried out to analyse the binding interface between RNF126 and BAG6. The unlabelled N-terminal RNF126 (both constructs: 1–40 and 1–100) was titrated into ^{15}N-labelled BAG6_UBL and vice versa. Both RNF126 constructs have the same effect on the BAG6_UBL spectrum and show similar shifts upon BAG6_UBL binding. This confirms that residues 1–40 are necessary and sufficient for the interaction (**Figure 51**). Hence we used the 1–40 construct in most subsequent experiments. The spectra displayed binding in a slow exchange regime characterised by two independent chemical shifts, one for the residue in the free state and one for the residue in the bound state (for ^{1}H-^{15}N-labelled RNF126_NZF see **Figure 52** and for ^{1}H-^{15}N-labelled labelled BAG6_UBL see **Figure 53)**. This type of CSP suggests a high-affinity binding with a saturation point close to a 1:1 molar ratio. CSPs were calculated for each residue in its free and bound spectra (δfree-δbound) for proton and nitrogen values respectively **(Figure 54)** and plotted onto the structure, with the most perturbed residues shown in dark red (**Figure 55**). We used our backbone assignments described in Section 4.3 together with the assignments for BAG6_UBL deposited by another research group in the BMRB (Accession Number: 11263). The assignment of the bound state spectra (described in Section 4.6) indicated that the β-sheet (predominantly the β3 strand) of RNF126_NZF is influenced upon titration, together with some residues from the β2-β3 loop. The binding region in BAG6_UBL is found along its exposed β-sheet; β3, β4, β5 strands, the loop between β1 and β2 and also includes part of the C-terminal region (**Figure 55**).

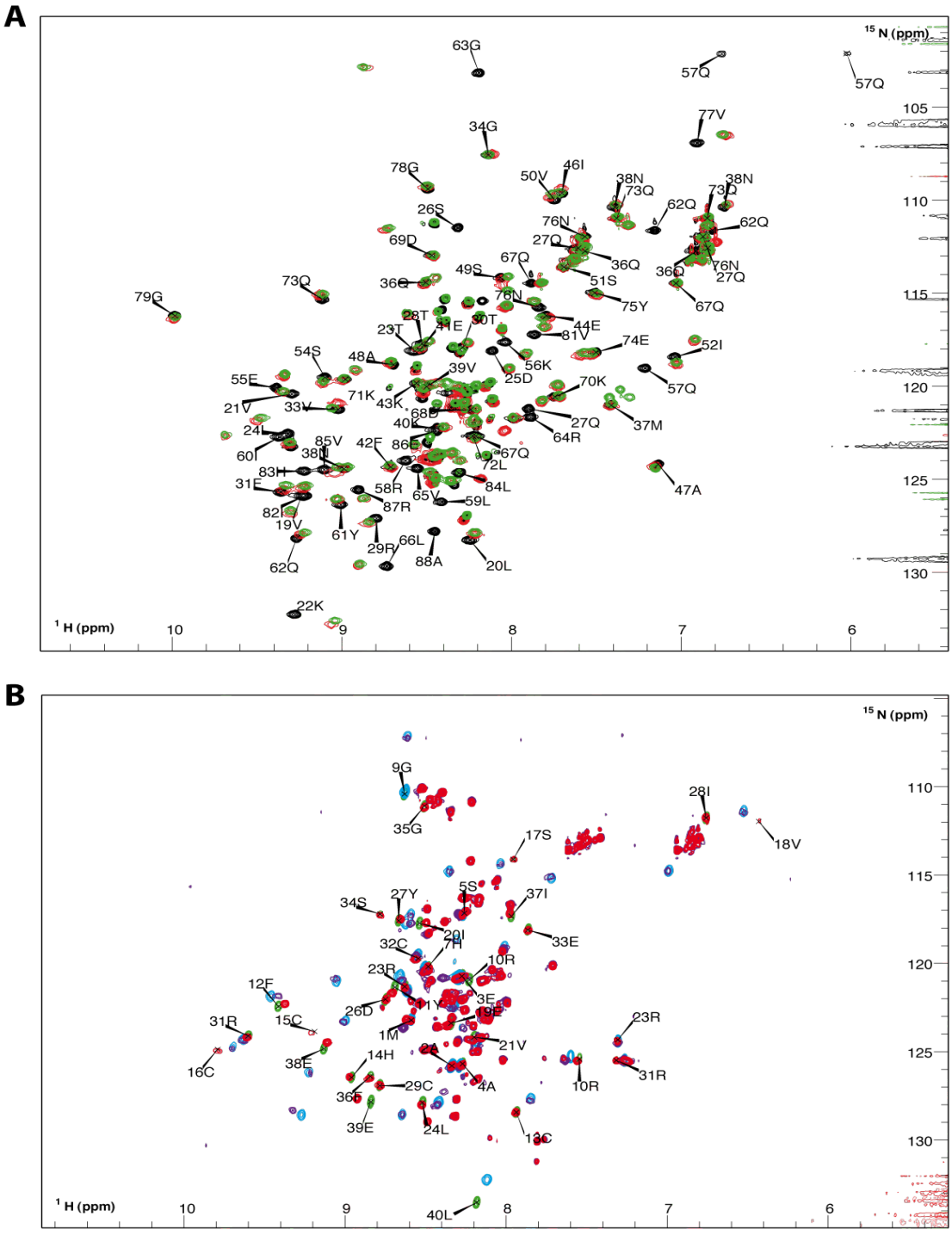

Figure 51: **NMR experiments showing that RNF126_NZF is sufficient for BAG6_UBL interaction**. A) ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled BAG6_UBL in free form (black with assignments), ^{15}N-labelled BAG6_UBL bound to unlabelled RNF126_NZF (1-40) at 1:1.5 molar ratio (green) and ^{15}N-labelled BAG6_UBL interacting with unlabelled RNF126 (1-100) construct at the same molar ratio (red). Bound states of BAG_UBL to both RNF126 constructs are similar. B) ^{1}H-^{15}N HSQC overlapping spectra of ^{1}H-^{15}N-labelled RNF126_NZF (green) and RNF126_1-100 (red) in free form and bound to unlabelled BAG6_UBL at a 1:1.5 molar ratio (NZF construct in sky blue and 1-100 in purple). Both RNF126 constructs interact in the same way with BAG6_UBL as they present the same chemical shift perturbations.

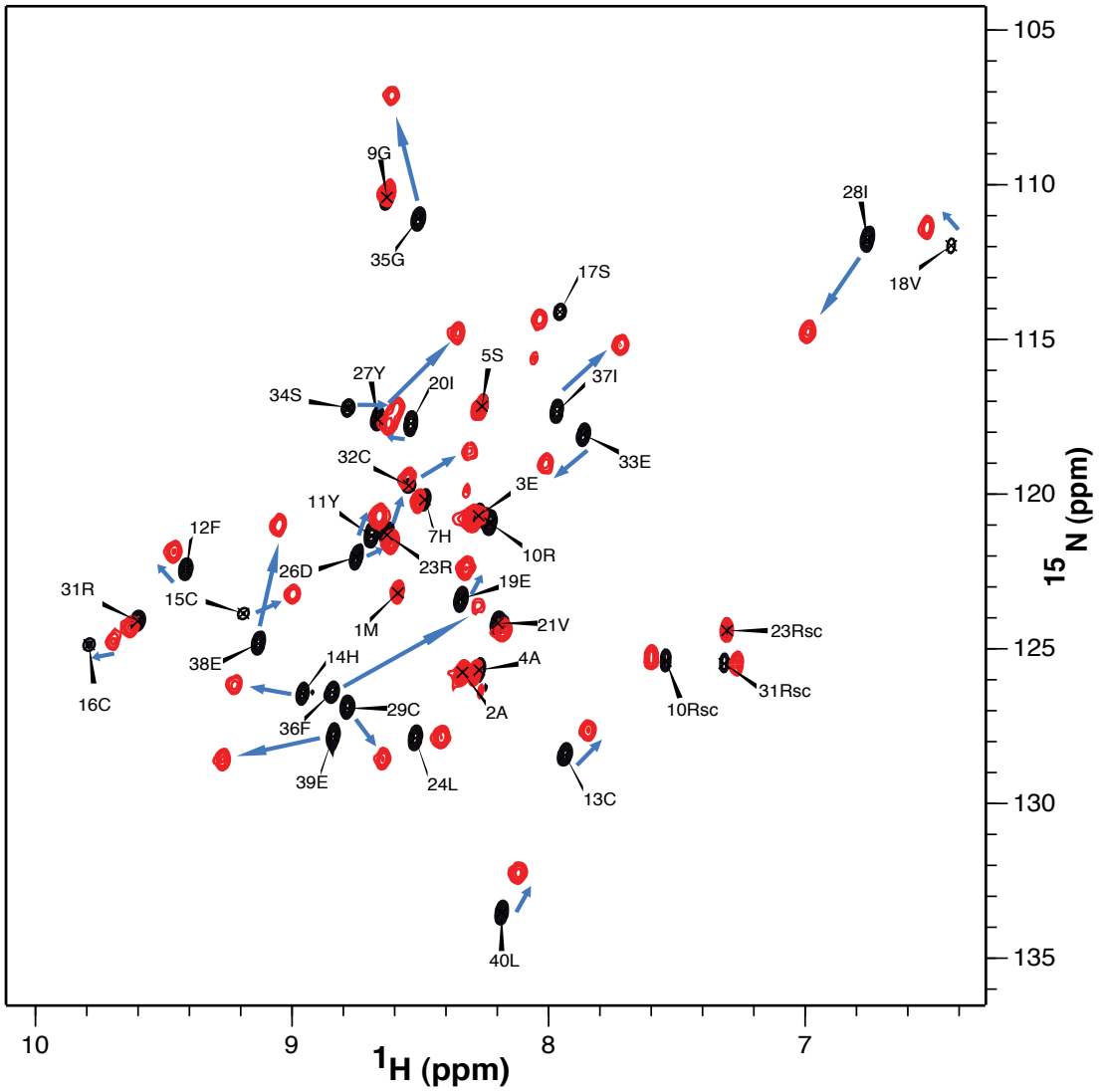

Figure 52: **The ^{15}N-^{1}H HSQC spectrum of ^{15}N- RNF126_NZF titrated with BAG6_UBL**. ^{15}N-^{1}H HSQC spectra of free RNF126_NZF (black, assigned peaks) and RNF126_NZF bound to BAG6_UBL (1:1 molar ratio in red). Blue arrows indicate the CSPs upon titration. Rsc corresponds to the Arginine side chain protons.

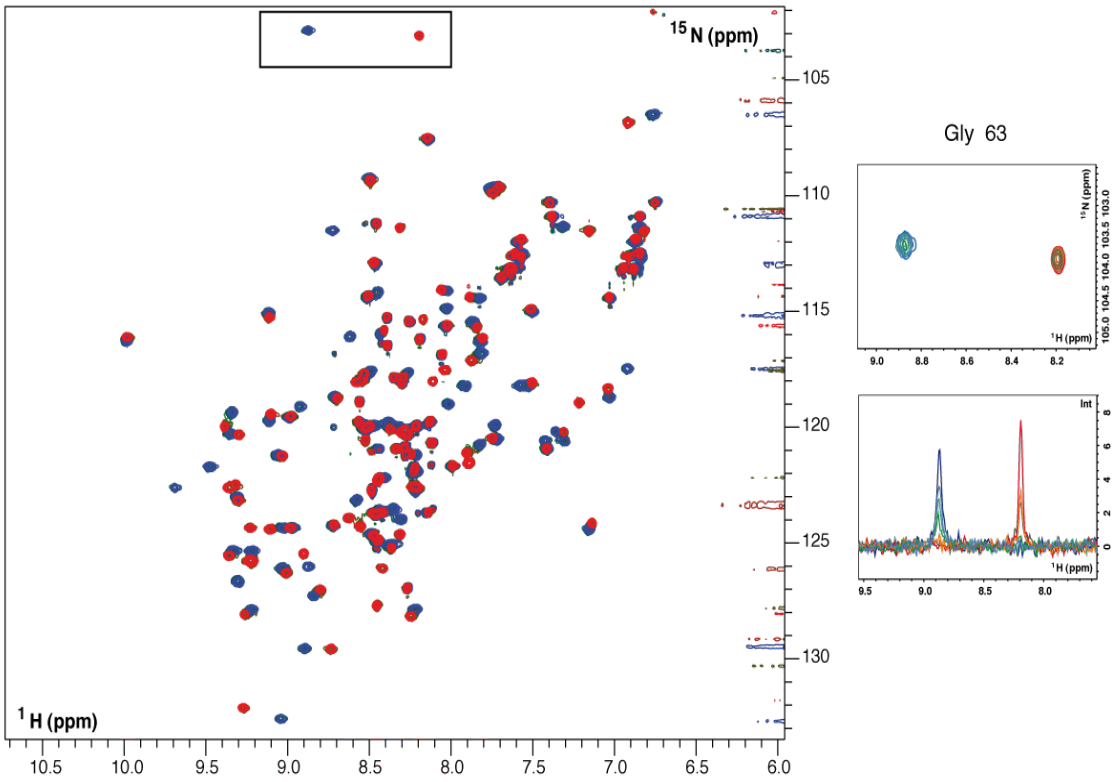

Figure 53: **The ^{15}N-^{1}H HSQC spectrum of ^{15}N- BAG6_UBL titrated with RNF126_NZF**. ^{15}N-labelled BAG6_UBL at different titration points with unlabelled RNF126_NZF (1:0, 1:0.25, 1:0.5, 1:0.75, 1:1, 1:2; in corresponding colours: red, orange, light green, cyan, blue, dark blue). On the right a detail of a spectral region, showing the slow exchange regimen of Gly63 shifting (up) and the peak intensities (down) (174).

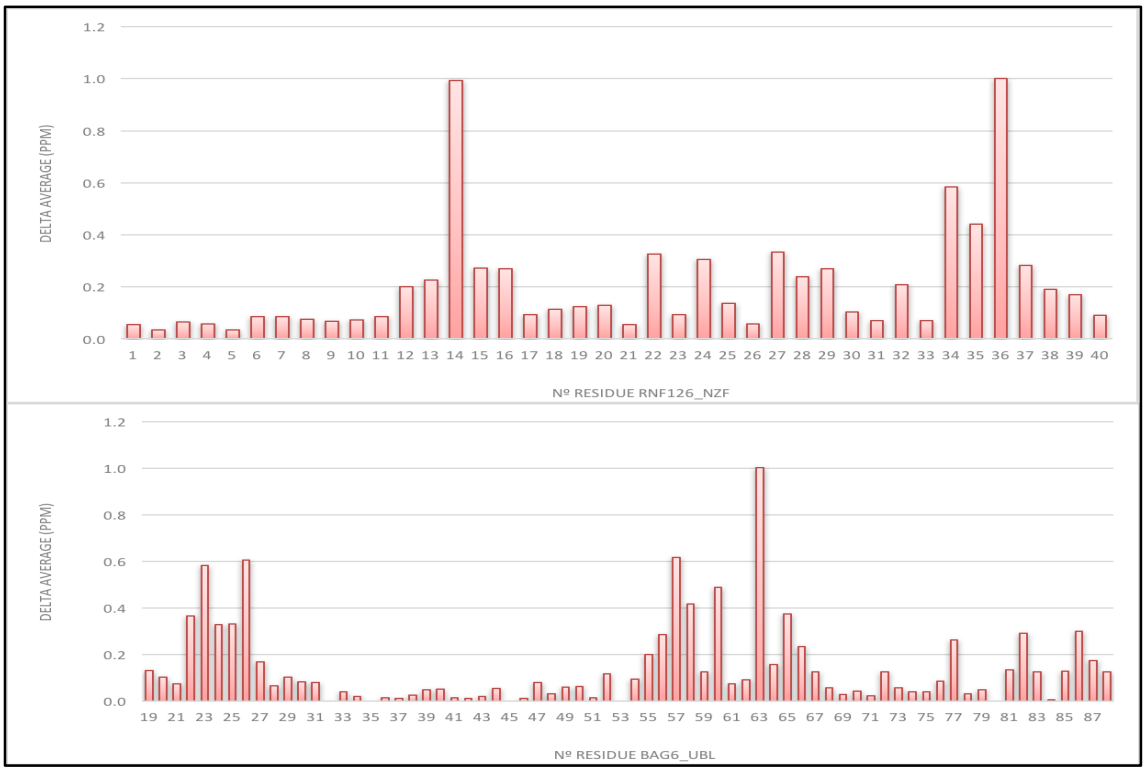

Figure 54: **Normalized CSP of ^{15}N-labbeled RNF126 titrated with unlabelled BAG6_UBL (top) and ^{15}N-labelled BAG6_UBL titrated with unlabelled RNF126_NZF (bottom).** The graph shows the CSP plot for each residue.

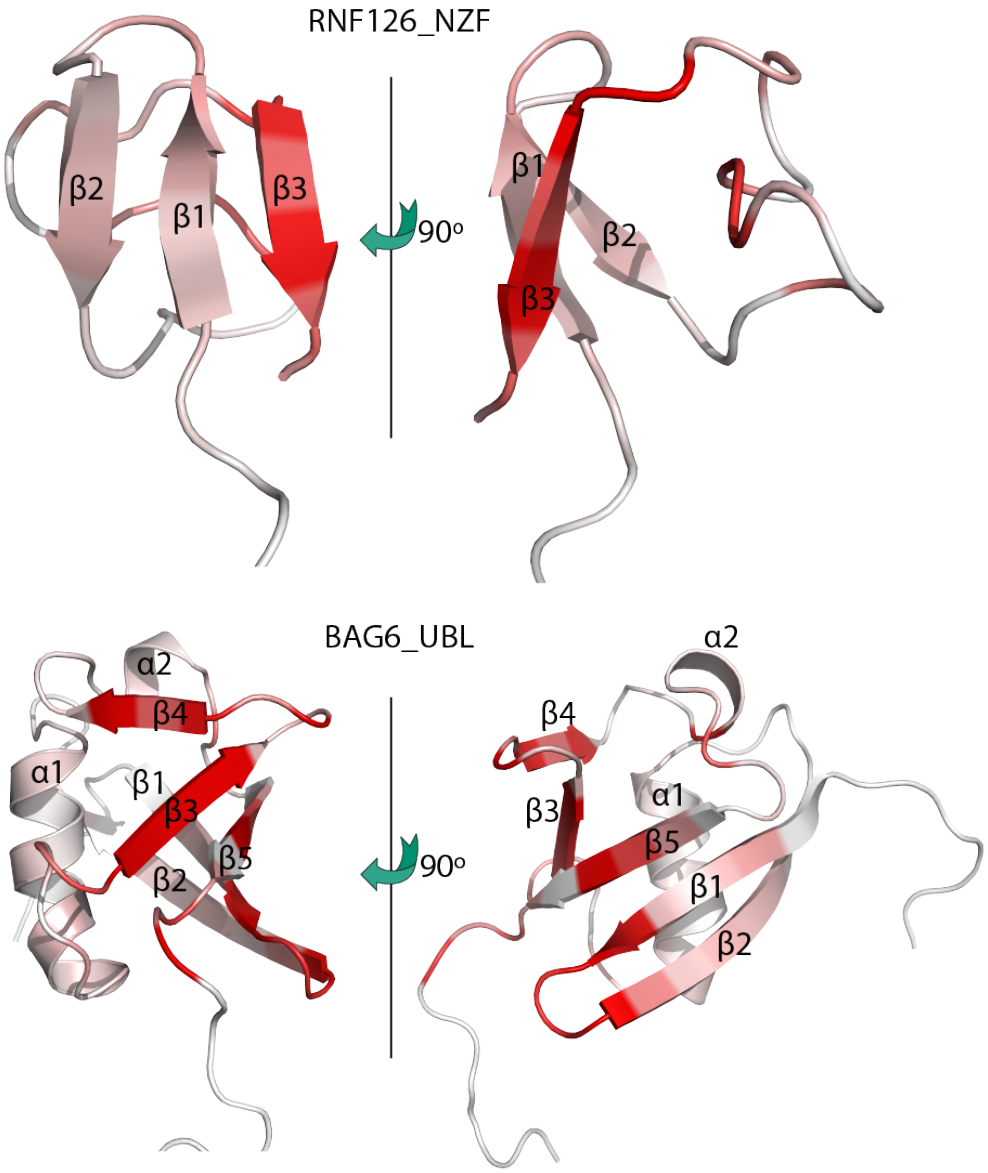

Figure 55: **Orthogonal cartoon views of RNF126_NZF and BAG6_UBL (PDB: 1WX9).** The structures are coloured according to reciprocal CSPs with the most perturbed residues shown in red.

4.5.2 ITC of RNF126_NZF interaction with BAG_UBL

The binding between RNF126_NZF and BAG6_UBL was confirmed and characterised using ITC (**Figure 56**). The results display 1:1 stoichiometry with a dissociation constant of 0.40 ± 0.05 μM in agreement with the slow exchange regime of the NMR titration data. The favourable enthalpy and entropy values obtained from ITC (ΔH = −4.52 ± 0.04 kcal/mol; ΔS = 14.1 ± 0.4 cal/mol·K, -TΔS = -4.21 ± 0.2 kcal/mol, ΔG = −8.73 ± 0.05 kcal/mol) suggest that the complex formation

between RNF126 and BAG6 is driven by the establishment of both hydrogen bonds and hydrophobic interactions.

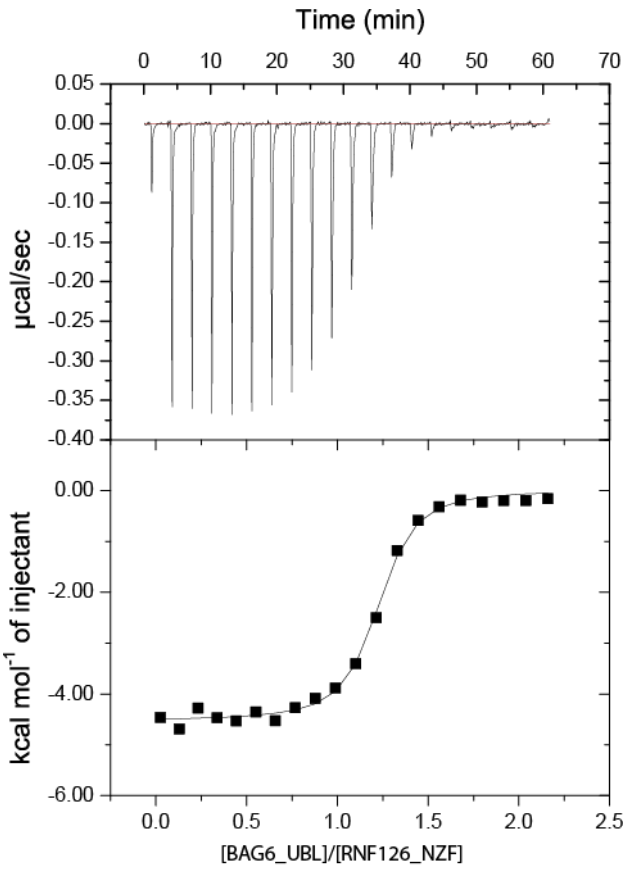

Figure 56: **ITC data showing binding of RNF126_NZF to BAG6_UBL**. The affinity constant, determined by ITC was: Kd = 0.40 ± 0.05 µ. All integrated heat data obtained for the titrations were corrected for heats of dilution (top panel) and fitted using a nonlinear least-squares minimization algorithm to a theoretical titration curve, using the MicroCal-Origin 7.0 software (bottom panel).

4.5.3 Probing the binding interface of RNF126_NZF/BAG_UBL complex by mutagenesis

To further characterize the interface, RNF126_NZF mutants were designed, considering the largest chemical shift perturbations. ITC results showed that the RNF126_NZF binding with UBL domain was driven by hydrophobic and hydrophilic interactions. H14 and F36 residues with exposed side chains at the β-sheet and highly perturbed were selected as good candidates, in addition to E38 and E39 which are involved in electrostatic interactions. The following mutants were produced by mutagenesis: single mutants of H14A and F36A and also a E38R/E39R double

mutant to reverse the charge. The effect of RNF126 mutations on BAG6_UBL interaction were studied using ITC (**Figure 57**), native PAGE mobility shift assay (**Figure 58**) and NMR (**Figure 59**). All mutations result in a drop in binding affinity of at least an order of magnitude, and binding was not detected in the case of a double mutant E38R/E39R, proving that these residues are involved in the interaction with BAG6. The wild-type structure is retained in each mutant, as judged by 1D NMR (**Figure 60**).

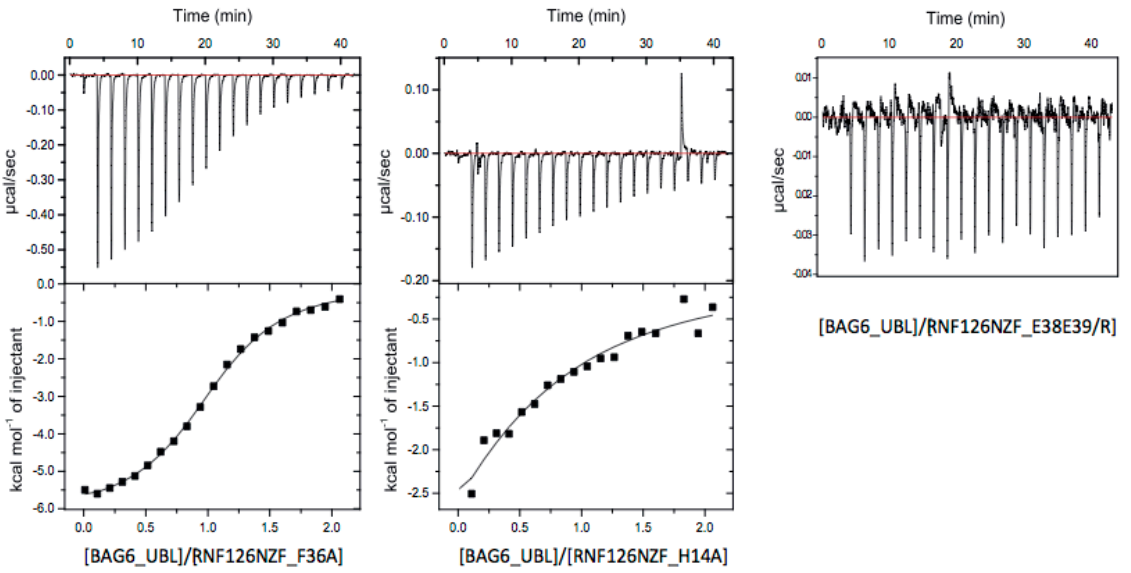

Figure 57: **ITC data indicating binding of RNF126_NZF mutants to BAG6_UBL**. Affinity constants, determined by ITC were: Kd of F36A mutant = 5.0 ± 0.3 µM; Kd of H14A mutant = >50 µM and Kd of E38E39/R mutant could not be calculated. In comparison, Kd of wild type RNF126_NZF = 0.40 ± 0.05 µM.

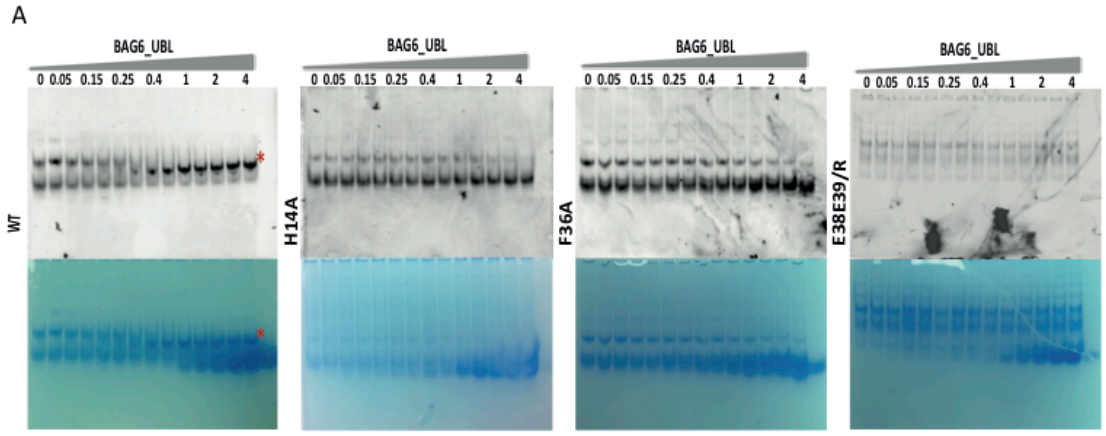

Figure 58: **Native PAGE of GFP-tagged RNF126_NZF**. Native PAGE of GFP-tagged RNF126_NZF_WT and mutants visualised by fluorescence (top) and Coomassie (bottom) titrated with a increasing concentration of BAG6_UBL. The red asterisk indicates the complex formation in the case of

RNF126_NZ WT. However, the RNF126_NZF_H14A and RNF126_NZF_F36A and E38E39/R mutants present lower or no affinity.

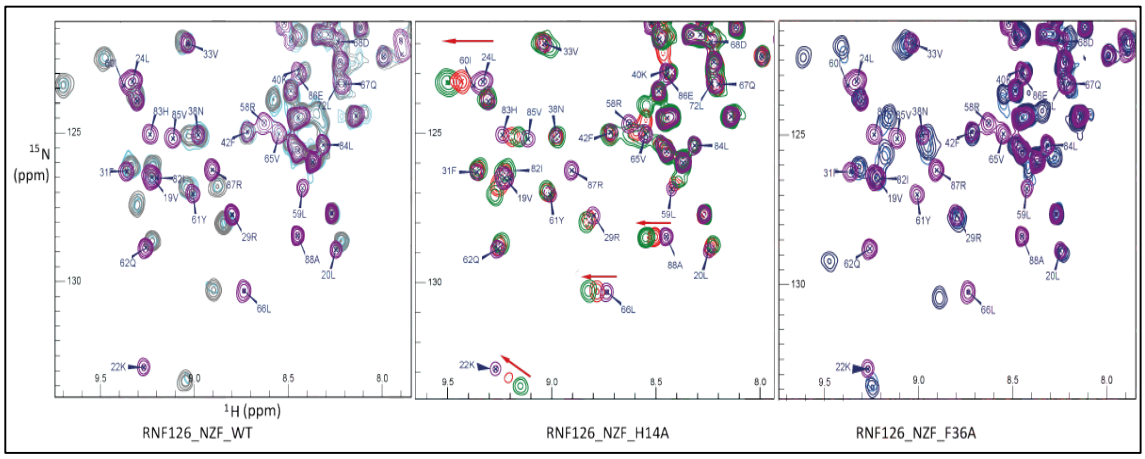

Figure 59: **^{1}H-^{15}N HSQC spectra of labelled BAG6_UBL upon titration with RNF126_NZF_WT, RNF126_NZF_H14A and RNF126_NZF_F36A**. Chemical shift perturbation occupies a slow chemical exchange regime in the case of both wild type and the F36A mutant and a fast chemical exchange regime in the case of mutant H14A indicating a lower affinity (red arrows show the CSPs). The protein concentration of E38E39/R mutant was too low to be used in NMR titrations after it precipitated.

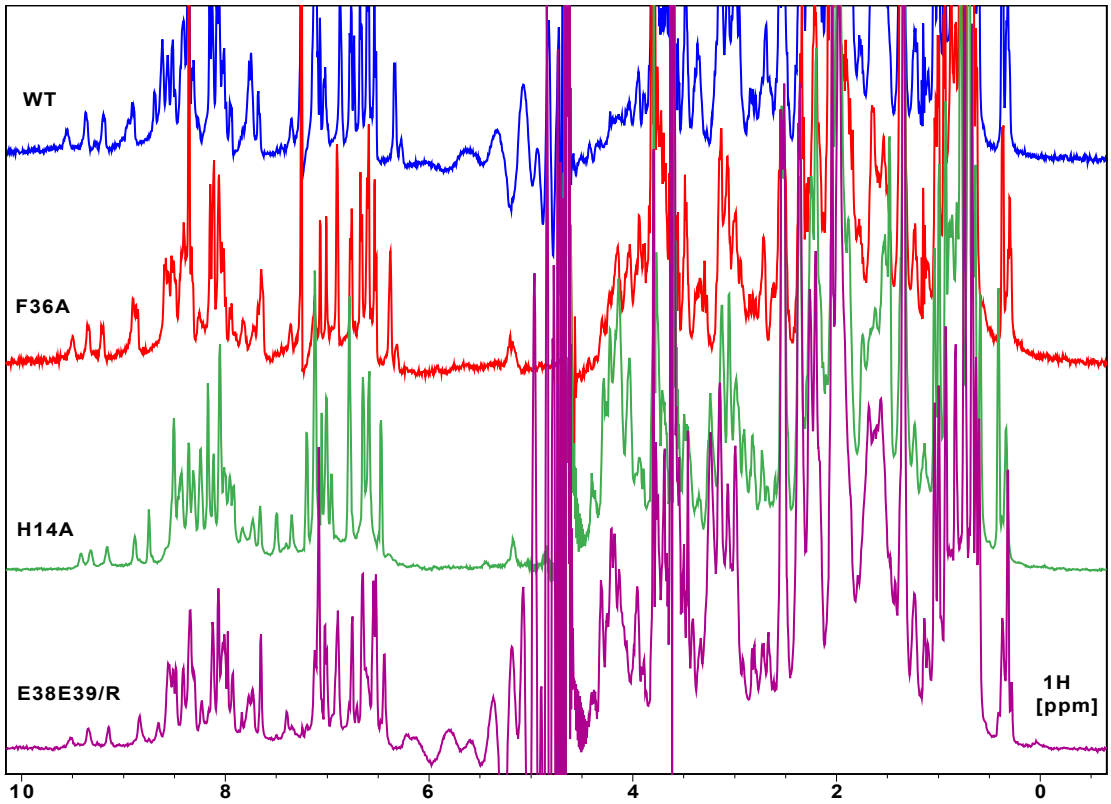

Figure 60: **1D ^{1}H NMR spectra of RNF126_NZF_WT and mutants, showing characteristic shifts of the folded protein.**

4.5.4 Chemical Shift Perturbations of RNF126_NZF interacting with UBL4A_UBL

Since both subunits of the BAG6 complex, BAG6_UBL and UBL4A_UBL, each bind to SGTA (82), we considered that RNF126_NZF and UBL4A_UBL might also interact. First, the interaction was studied using chemical shift perturbations. Upon NMR titration with unlabelled UBL4A_UBL, ^{15}N-labelled RNF_NZF displayed CSPs within the first forty residues that agreed with those shifts that occur upon BAG6_UBL binding (**Figure 61**). The analysis of the reciprocal titration with ^{15}N-labelled UBL4A_UBL was completed using assignments from the BMRB database (Accession Number: 11279) and CSPs were calculated (**Figure 63**) and mapped onto the UBL4A_UBL structure, deposited previously by another group (PDB: 1WX9). The analysis showed that the binding interface on UBL4A is equivalent to the one obtained for BAG6: β3, β4, β5 strands and the loop between β1 and β2 are perturbed upon binding RNF126 (**Figure 62** and **Figure 65**). Both titrations occur in a fast exchange timescale suggesting that this interaction has a lower affinity than that of BAG6_UBL binding. Furthermore, plotting the CSP data against the concentration of RNF126_NZF and fitting the curve to a simple 1:1 equimolar binding model produces a dissociation constant (Kd) in the order of 20 μM, around 50 times higher than that of BAG6 (**Figure 64**).

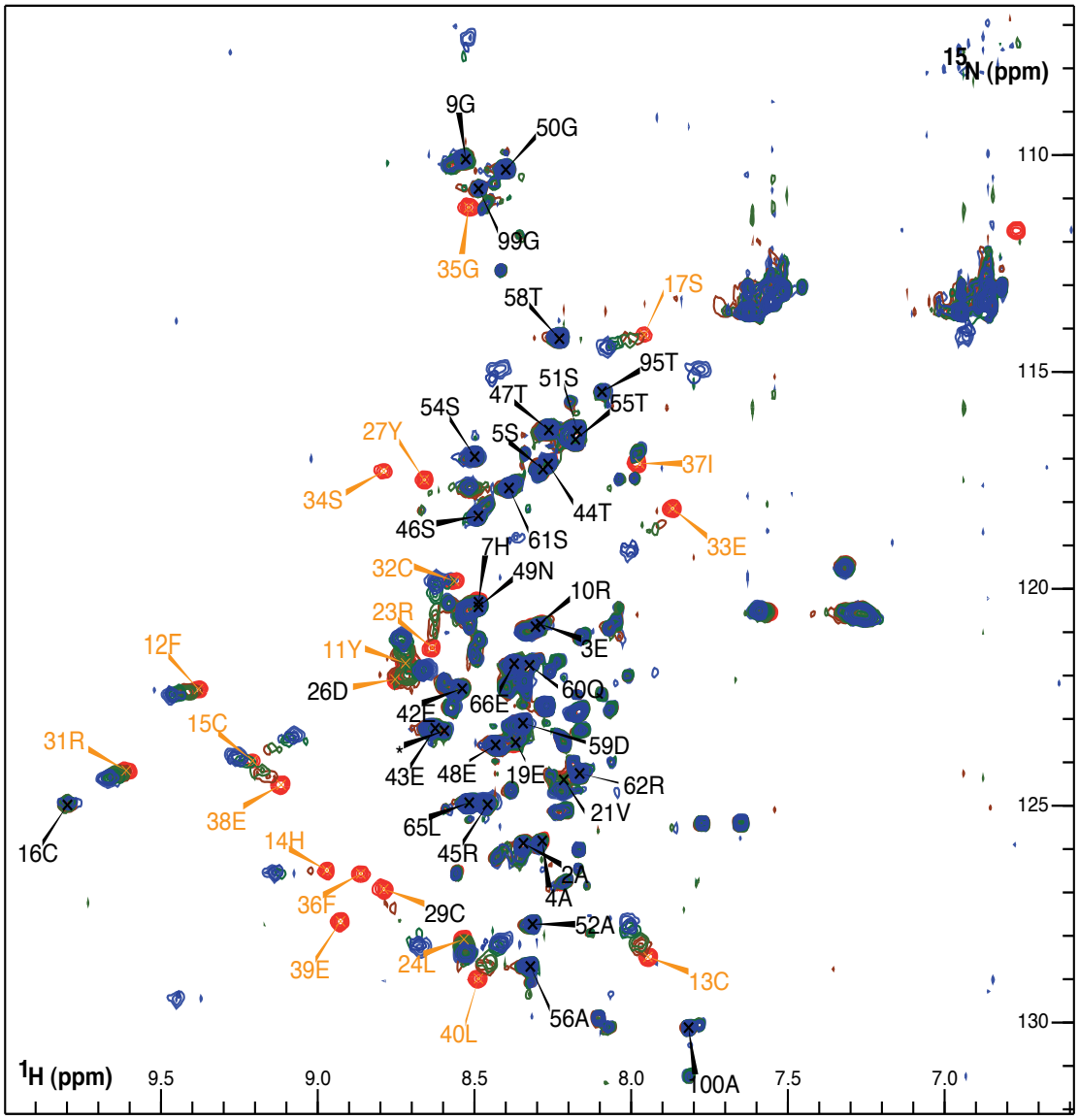

Figure 61: ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled RNF126 1-100 (black) titrated with unlabelled UBL4A_UBL. UBL4A_UBL was added at different molar ratios (RNF126:UBL4A): 1:0 (red), 1:0.125 (brown), 1:0.25 (green), 1:0.75 (light blue) and 1:1 (navy). The residues of RNF126 assigned in orange showed chemical shift perturbations within the first forty residues just like those shifts that occur upon BAG6_UBL binding.

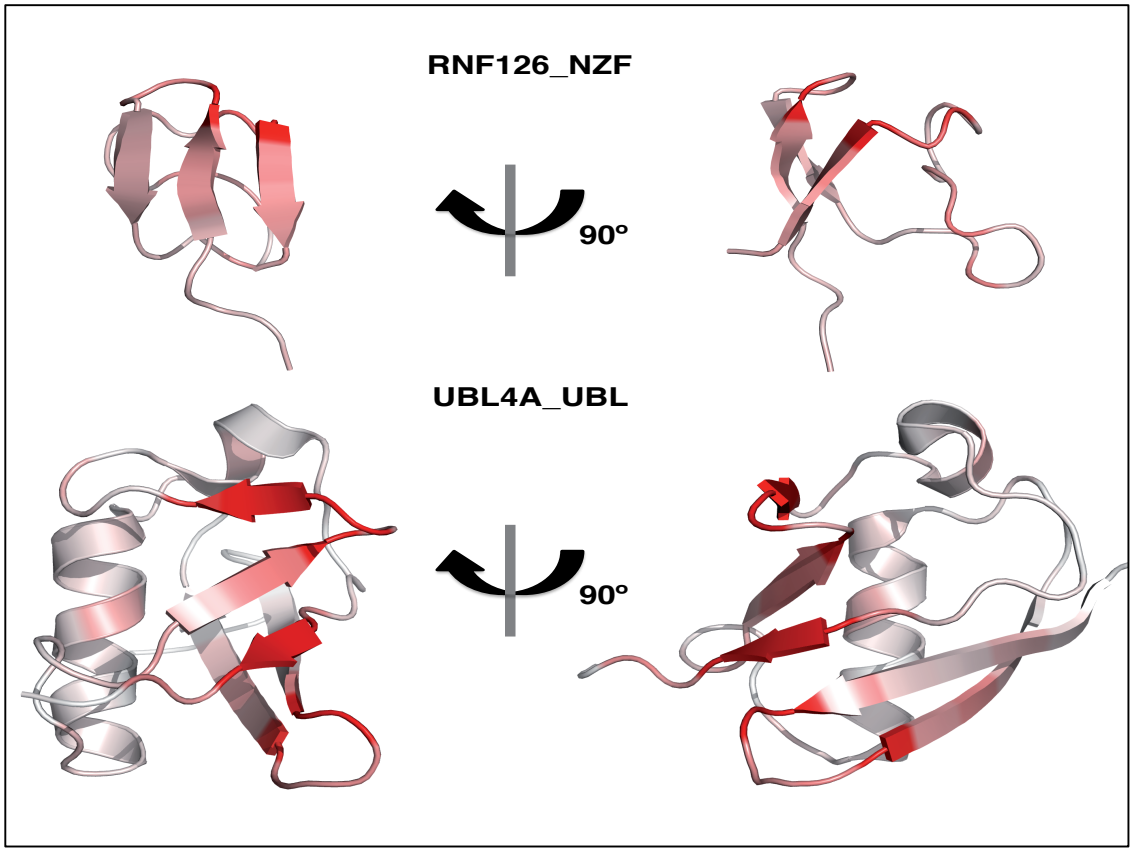

Figure 62: **Cartoon views of RNF126_NZF and UBL4A_UBL coloured according to reciprocal CSP data.** The most perturbed residues are presented in red. UBL4A_UBL PDB accession is 1WX9.

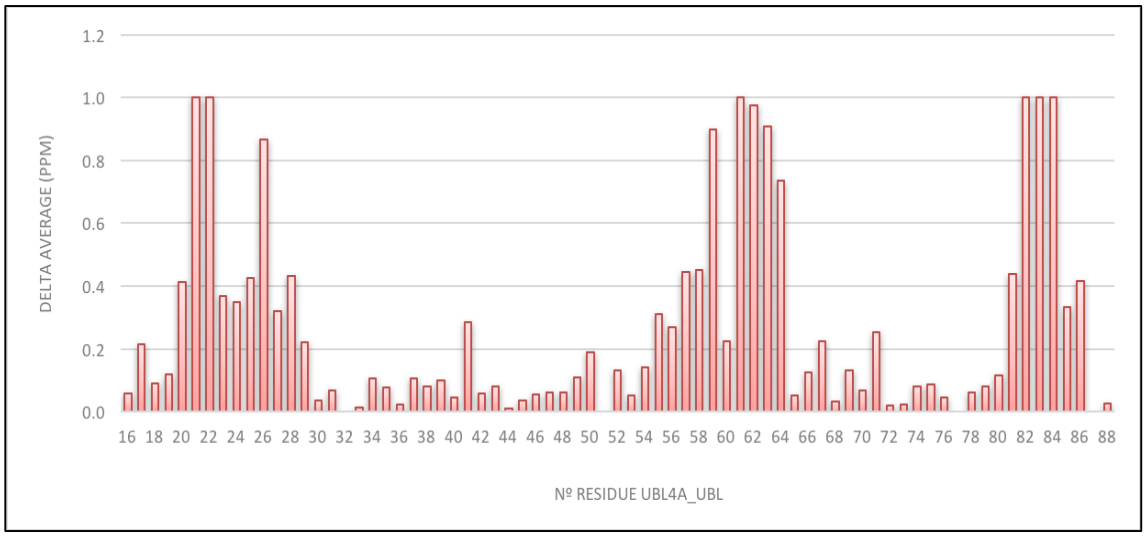

Figure 63: **Normalized CSP of ^{15}N-labelled UBL4A_UBL titrated with unlabelled RNF126_NZF**. The graph shows the CSP plot for each residue.

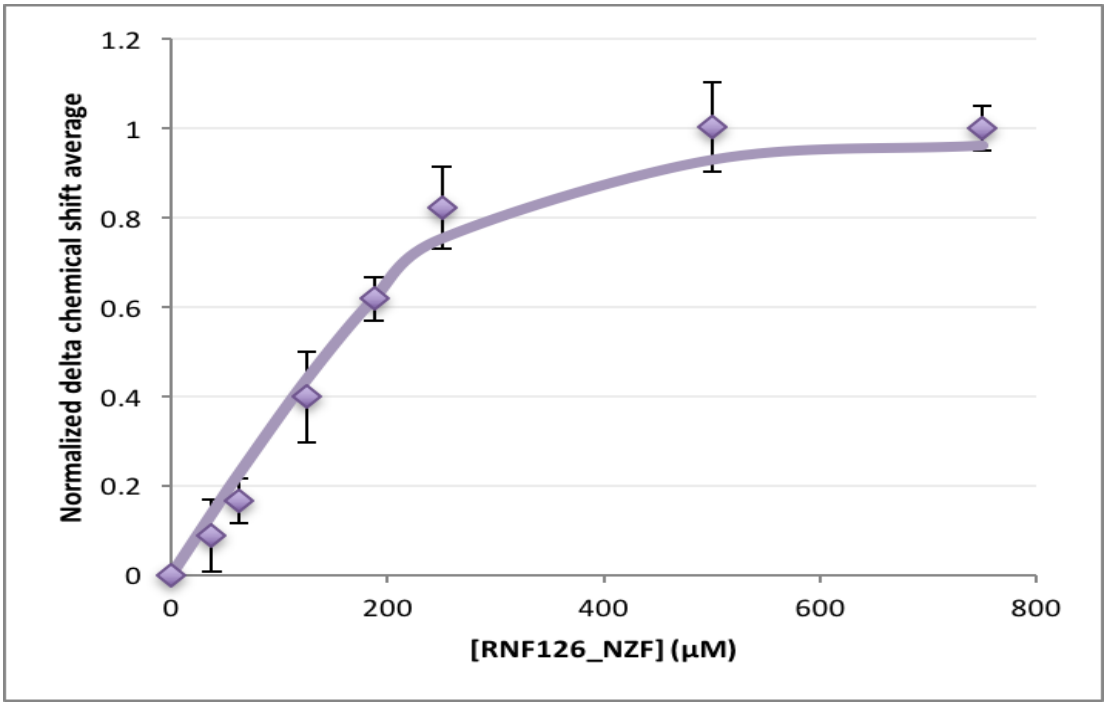

Figure 64: **Plot representing the normalized CSP data of the most perturbed UBL4A_UBL amino acids**. The perturbed UBL4A_UBL residues upon titration with different concentrations of RNF126_NZF were 20, 21, 22, 23, 24, 25, 26, 28, 57, 58, 59, 61, 62, 63, 64, 81, 82, 83, 84, 85, 86. Fitting was performed using the DynaFit program yielding a Kd of 16.6 ± 2.7 µM (174).

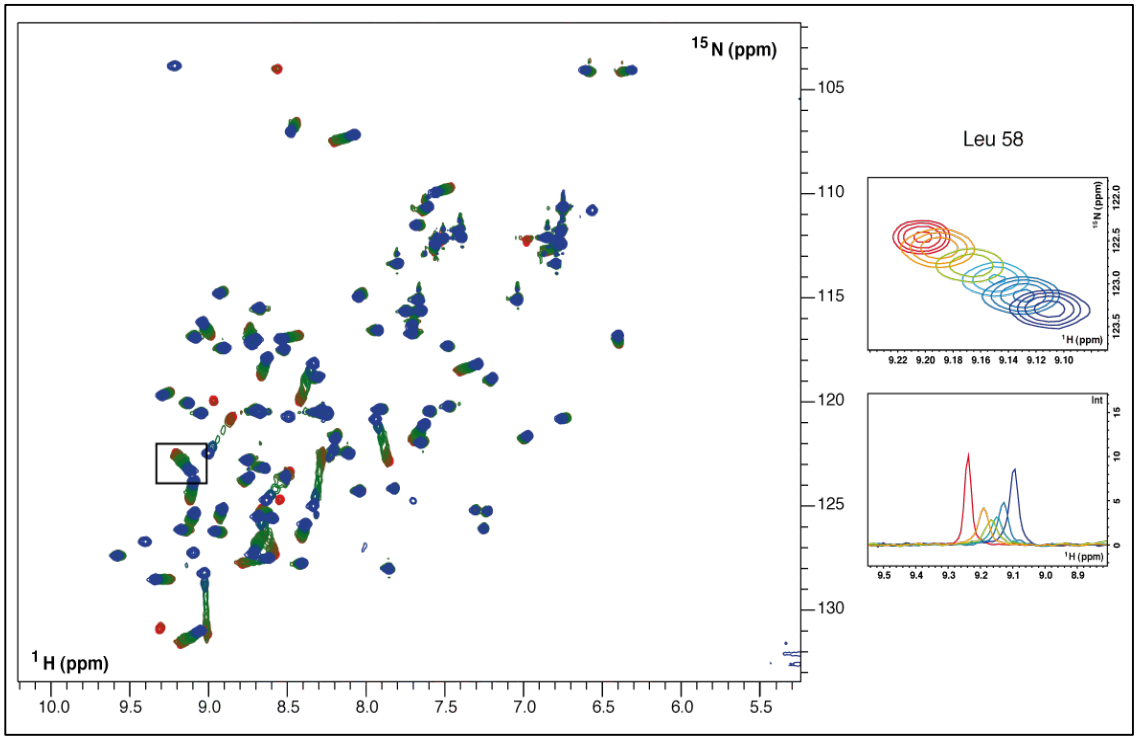

Figure 65: **^{15}N-labelled UBL4A_UBL titrated with RNF126_NZF. RNF126_NZF was used at molar ratio equivalents (1:0, 1:0.25, 1:0.5, 1:0.75, 1:1, 1:2; in corresponding colours: red, orange, light green, cyan, blue, dark blue)**. The detail on the right shows the shifting and peak intensities of the Leu58 signal (174).

4.5.5 Comparison of the RNF126_NZF interaction with UBL domains of UBL4A and BAG6 probed by MST, ITC and native gel electrophoresis

We compared the interaction of RNF126_NZF with UBL domains using ITC, MST and native gels. Microscale thermophoresis (MST) of RNF126_NZF and UBLs binding were performed using GFP tagged version of RNF126_NZF. GFP tagged RNF126_NZF was kept at a constant concentration, while BAG6_UBL concentrations ranged from 25 μM to 0.8 nM and UBL4A at 50 μM to 1.5 nM. The K_d values derived from MST are approximately an order of magnitude lower than the values obtained by ITC indicating higher affinity (**Figure 67**). Experimental constraints necessitated using the labelled protein at a similar concentration to the expected Kd, meaning that small errors in GFP protein concentration have disproportionate impact on the results, potentially leading to overestimation of the affinity. We used 0.25 or 0.125 μM GFP protein while the Kd is 0.4 μM (ITC). We made many attempts to resolve this including using lower GFP concentrations to move away from the Kd value but we were unable to observe sufficient fluorescence signal (even after increasing the LED power). ITC however, proved not sensitive enough to detect the binding of RNF126_NZF with UBL4A_UBL in any of the tested conditions. In the same experimental conditions as were used for BAG6, the observed transition was small and impossible to fit, in keeping with the reduced affinity that we observed (**Figure 66**). The concentration optimization we attempted (500-1000 μM for RNF126_NZF and 25,50 or 100μM in the reaction cell for UBL4A_UBL) failed to improve the results.

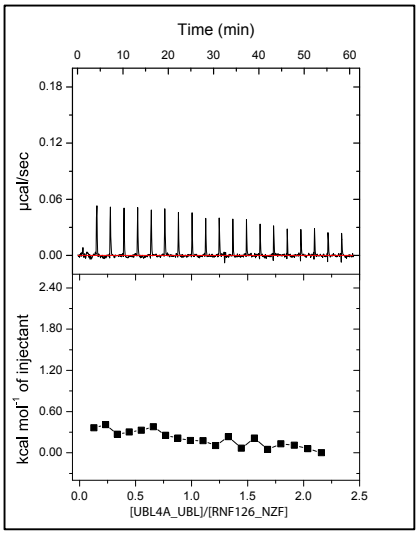

Figure 66: **ITC data showing a titration curve of RNF126_NZF into UBL4A_UBL.**

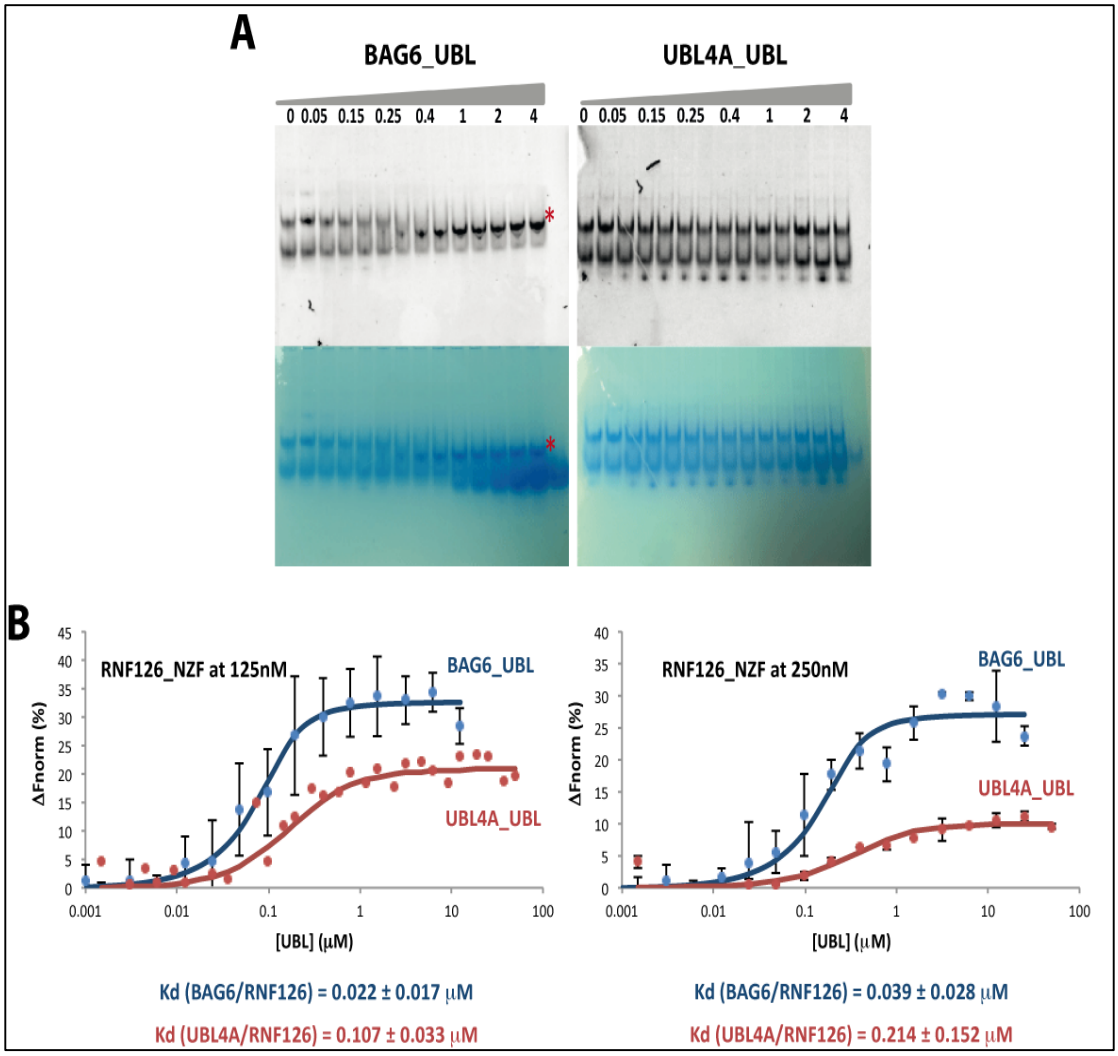

Figure 67: **Binding of RNF126_NZF to UBL domains.** A) Native PAGE of GFP tagged RNF126_NZF binding assay with UBLs (BAG6_UBL on the left, UBL4A_UBL on the right) visualised by fluorescence (top) and Coomassie (bottom). Values on the top of the gel indicate RNF126:UBL molar equivalents. The red asterisk indicates complex formation. B) MST data showing the interaction of GFP tagged RNF126_NZF at two different concentrations (0.125 and 0.250 μM) with BAG6_UBL (blue) and UBL4A_UBL (red). The change in normalized fluorescence is plotted against the concentration of unlabelled ligand and the affinity values obtained from fitting with Dynafit are displayed in the figure.

NMR, MST and native PAGE mobility shift, collectively demonstrate a lower affinity (or no binding detected in the case of ITC and native gels), between UBL4A and RNF126 as compared to the BAG6 binding. However, more sensitive technique of NMR shows that, RNF126 is capable of binding to both of the UBL domain containing subunits of the BAG6 complex, just like SGTA.

4.6 Backbone and sidechain assignment of RNF126_NZF and Bag6_UBL bound states in RNF126_NZF/Bag6_UBL complex

To study the RNF126_NZF and BAG6_UBL interaction in greater detail and to investigate whether there are any structural rearrangements upon complex formation we aimed at solving the structure of the complex by NMR. The stability of the complex and the high quality of the NMR data allowed the experimental structure solution of the complex with no need to resort to in silico methods. The backbone and sidechains were assigned for both free and bound protein. Since the RNF126_NZF and BAG6_UBL interaction displays binding in a slow exchange regime, two independent chemical shifts can be observed (one for the residue in the free state and one for the residue in the bound state). Therefore, peaks corresponding to backbone and sidechain atoms of the bound state of RNF126_NZF and Bag6_UBL in RNF126_NZF/Bag6_UBL complex had to be assigned. We acquired complete triple resonance experiments for backbone and side-chain assignment of the bound state for each protein (by mixing labelled and unlabelled protein) and assigned them in the same way as the free RNF126_NZF. The assignments of ^{1}H, ^{13}C, and ^{15}N chemical shift were deposited at BMRB with the Accession Number: 25914. The Validation Software (AVS) was run automatically after the assignment deposition and the full report can be seen in Appendix.

4.6.1 NOE Distance constraints in intra- and inter- molecular assignments

Filtered NOESY experiments were run on complex samples consisting of $^{15}N^{13}C$-labelled RNF126_NZF bound to unlabelled BAG6_UBL or unlabelled RNF126_NZF bound to $^{15}N^{13}C$–labelled BAG6_UBL to identify distance restraints within the binding interface. NOE restraints were derived from the two pairs of ^{15}N-edited NOESY-HSQC (to see both intra- and intermolecular NOEs), ^{13}C-edited NOESY-HSQC (intra- and intermolecular) and ^{12}C, ^{14}N-filtered-^{13}C-edited NOESY-HSQC (to see only intermolecular cross-peaks) experiments for the complex using either ^{15}N, ^{13}C -labelled BAG6_UBL or RNF126_NZF. Manual assignments of

NOESY cross peaks were performed using CCPNanalysis. In total 48 intermolecular distance constraints were assigned and the numbers of NOE distance restraints are shown in

Table 12.

4.7 NMR structure of the RNF126_NZF/BAG6_UBL complex

The backbone and sidechain assignments of bound states of RNF126_NZF and BAG6_UBL, together with NOE restraint assignments from 3D NOESY experiments, and dihedral constraints (ϕ and Ψ angles), extracted from the chemical shift values using TALOS+, were used in complex structure calculations using the ARIA2. The final Aria run generated 100 structures and the 20 structures with the lowest energy are represented as an ensemble (**Figure 69A**) and were deposited in the PDB with the Accession Number: 2N9P. The statistical parameters are shown in **Table12**. The reciprocal intermolecular NOEs between RNF126 and BAG6 were unambiguously assigned to ensure the accuracy of the structure calculation. The complex structure exhibits overall conservation of the original fold for each protein and a strong match between the binding interface of the obtained structure and the CSP analysis (**Figure 69A** and **Figure 54**). RMSD values between the free form of RNF126_NZF (this work) and the bound state are low (backbone RMSD of 0.62Å) indicating lack of major conformational changes and the zinc coordination parameters are conserved **(Table 13**). BAG6 also presents low values of RMSD comparing free (PDB: 1WX9) and bound states (backbone RMSD of 1.43Å), with a minor alteration of the arrangement of loop β1-β2 (**Figure 68**). Interactions observed at the binding interface of RNF126_NZF/BAG6_ubl complex include hydrogen bond formation and electrostatic and hydrophobic interactions, which agrees with ITC data. Hydrogen bonds are formed between the guanidinium groups of three arginine residues of BAG6 and the carboxyl group of two glutamic acids of RNF126. The side-chain of Arg64 connects with Glu39 via hydrogen bond while Arg58 and Arg87 create a hydrogen bond with Glu38. Furthermore, the Glu38 amide is involved in hydrogen bond formation with the backbone carbonyl group of Gly63 from BAG6. Tyr27 from RNF126 is aligned with Gln62 from BAG6 and stabilises its aromatic ring. Tyr27 also interacts

with Arg64 through its hydroxyl group. Additional interactions are identified for imidazole groups of His14 (RNF126) and His83 (BAG6) (**Figure 69 B**). Hydrophobic contacts are shown for the two components: the aliphatic side-chain of Glu38 from β3 of RNF126 packed against Ile60, Val65 and side-chain of Arg58 from BAG6, and Phe36 in β2 of RNF126 fitted neatly within a hydrophobic pocket formed by Leu24, His83 and Val85 of BAG6 (**Figure 69 C**).

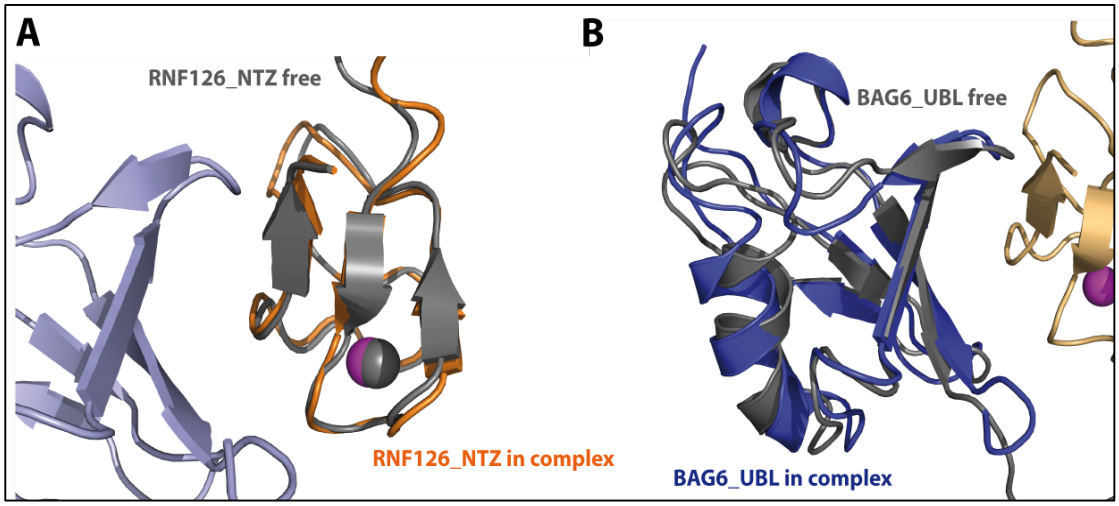

Figure 68: **Structural alignment of RNF126_NZF and BAG6_UBL in free and bound states.** A) free RNF126_NZF (grey) and bound to BAG6_UBL state (orange). B) free BAG6_UBL (grey, PDB: 1WX9) and bound state to RNF126 (blue, this work). Backbone RMSD values are: RNF126_NZF (10-40) RMSD of 0.62Å; BAG6_UBL (17-88) RMSD of 1.43Å (174).

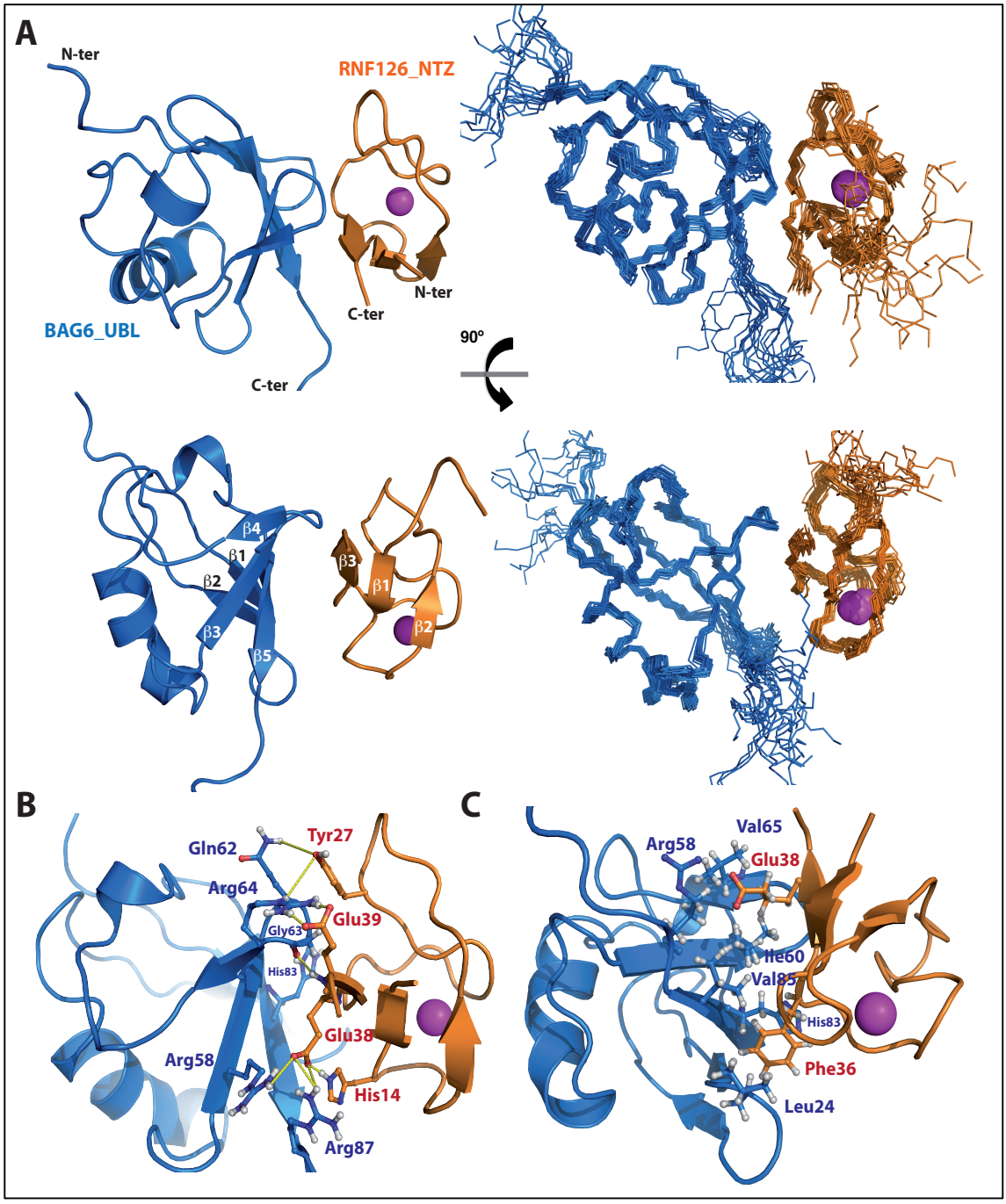

Figure 69: **RNF126_NZF/BAG6_UBL complex structure**. A) Two views of cartoon representations (left) and ensemble backbone (right) showing the 20 lowest energy ARIA-calculated complex structures as deposited in the PDB (Accession number: 2N9P). RNF126_NZF is coloured orange and BAG6_UBL is coloured blue with the coordinated zinc cation shown as magenta sphere. B) Binding interface of the complex displaying interactions. Residues shown in ball and stick representation are involved in the formation of hydrogen bonds or electrostatic interactions (indicated as yellow dashed lines). C) Details of the hydrophobic core at the binding interface. Apolar residues and hydrophobic regions of Glu38 and Arg58 side-chains forming hydrophobic contacts at the complex interface are shown in ball and stick-representation (174).

Table 12: **NMR and refinement statistics for the final 20 structure ensembles of RNF126_NZF/BAG6_UBL complex.**

NMR distance and dihedral constrains		
Intramolecular distance constrains	**RNF126_NZF**	**BAG6_UBL**
Total unambiguous NOEs	461	1283
Intra-residue	252	689
Sequential (|i-j| = 1)	58	266
Medium-range (1 < |i-j| < 4)	23	115
Long-range (|i-j| > 5)	128	213
Ambiguous constrains	2	51
Intermolecular distance constrains	48	
TALOS derived dihedral constrains	**RNF126_NZF**	**BAG6_UBL**
Total dihedral constrains (Φ+Ψ)	33	90
Structure statistics		
Violations (mean and s.d.)		
Number of violated distance restraints per structure (>0.25 Å)	1.2 ± 0.5	
Max. distance constraint violation (Å)	0.28	
Number of violated dihedral angle restraints per structure (>5°)	1.3 ± 0.8	
Max. dihedral angle violation (°)	7.6	
Ramachandran Plot analysis (%)*	**RNF126 (1–40) BAG6 (1–101)**	**RNF126 (10–40) BAG6 (17–88)**
Residues in most favoured regions	65.3	72.3
Residues in additionally allowed regions	28.6	22.5
Residues in generously allowed regions	3.9	3.8
Residues in disallowed regions	2.2	1.3
Derivation from idealized geometry		
Bond length (Å)	0.0086 ± 0.0003	
Bond angles (°)	0.84 ± 0.02	
Averages RMSD to mean structure (range)	**RNF126 (10–40)/BAG6 (17–88)**	
Backbone (Å)	0.97 ± 0.25	
Heavy (Å)	1.64 ± 0.22	

*Obtained from PROCHECK-NMR.

Table 13: Geometry of the zinc coordination centre of RNF126_NZF in the RNF126_NZF/BAG6_UBL complex.

	Cys 13	Cys 16	Cys 29	Cys 32
First coordination shell				
Sγ-Zn distance	2.29 ± 0.01	2.29 ± 0.01	2.30 ± 0.01	2.29 ± 0.01
Cβ-S-Zn angle	108.0 ± 1.5	110.0 ± 1.0	106.4 ± 1.0	108.9 ± 1.0
Second coordination shell				
Sγ-HN (i+2) distance	3.02 ± 0.76	4.74 ± 0.48	2.60 ± 0.58	2.97 ± 0.81
Sγ-H-N (i+2) angle	122.5 ± 7.2	137.3 ± 6.3	129.3 ± 5.0	143.0 ± 9.1

Averaged distance (Å) and angular (º) values for first and second coordination shell across the NMR ensamble. Similar values compared with free form are obtained.

4.8 HADDOCK models of RNF126_NZF/UBL4A_UBL complex

Due to the similarities in interactions between RNF126_NZF and UBL domains of BAG6 and UBL4A in our NMR studies, we aimed at generating a structural model of the RNF126_NZF/UBL4A_UBL complex using HADDOCK-based semi-rigid, data-driven docking (330) rather than experimental structure solution. The CSP data of bound RNF126_NZF and UBL4A_UBL were used as ambiguous restraints and included 10 residues of RNF126_NZF and 14 of UBL4A_UBL with chemical shift changes greater than 0.15 ppm and filtered for a relative solvent accessibility higher than 45%. In calculations we provided PDB-deposited structures of UBL4A_UBL (2DZI) and the lowest energy NMR structure from our family of RNF126_NZF. The three lowest energy HADDOCK clusters of RNF126_NZF/UBL4A_UBL are shown in **Figure 70 A** and parameters are presented in **Table 14**. The top-ranked cluster 1 is highly similar to the experimentally-derived RNF126_NZF/BAG6_UBL structure (**Figure 70 B**) indicating that residues forming contacts with RNF126_NZF are conserved between UBL4A_UBL and BAG6_UBL (**Figure 71 A**, **Figure 70 C** and **D**). The UBL4A_UBL residues shown in the binding interface with RNF126 are Leu8, Arg42, Leu44, Lys48, Ala49, Asn68 and Val70 of UBL4A (**Figure 70 C**), which align with residues Leu24, Arg58, Ile60, Arg64, Val65, His83 and Val85 of BAG6_UBL (**Figure 70 D**). Both of these groups of UBL residues make contact with His14,

Tyr27, Phe36, Glu38 and Glu39 in RNF126_NZF. The positively charged Lys48, Arg42 and Lys72 of UBL4A (equivalent to Arg64, Arg58 and Arg87 in BAG6) form electrostatic interactions with Glu39 and Glu38 in RNF126_NZF (**Figure 71 B**). Phe36 of RNF126_NZF inserts into a pocket formed by Leu8, Asn68 and Val70 of UBL4A forming hydrophobic contacts between aliphatic portions of the side-chains (**Figure 71 C** and **D**).

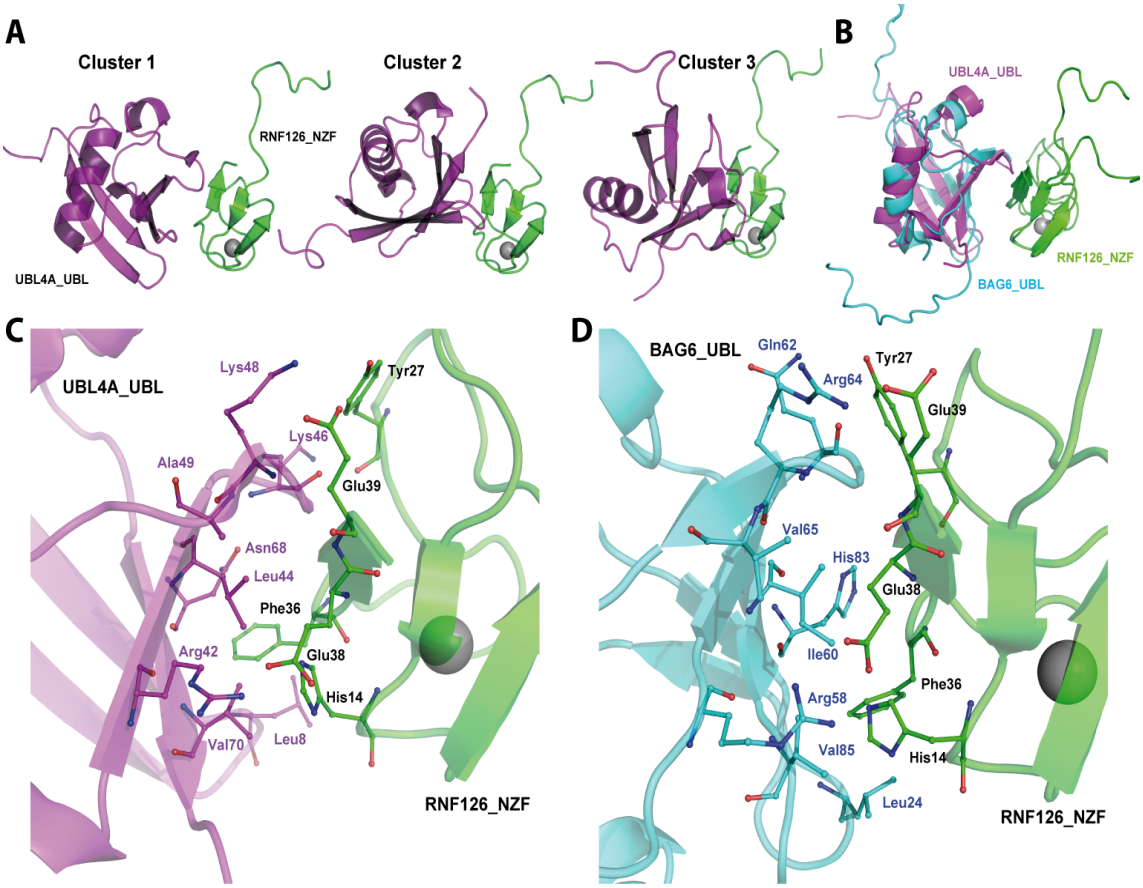

Figure 70: **RNF126_NZF/UBL4A_UBL complex structure modelling results**. A) Three top scoring clusters of the RNF126_NZF/UBL4A_UBL complex generated by HADDOCK. B) Overlay of RNF126_NZF/BAG6_UBL complex structure as calculated using ARIA with the HADDOCK-generated top scoring cluster 1 of RNF126_NZF/UBL4A_UBL complex. C) and D) Details of the binding interface displaying the equivalent related residues as sticks in complexes of RNF126_NZF/UBL4A_UBL (C) and RNF126_NZF/BAG6_UBL (D).

Table 14: Energetic parameters obtained for the three best clusters of RNF126_NZF/UBL4A_UBL complex calculated using HADDOCK.

Cluster	HADDOCK score	Cluster size	RMSD (Å)	Van der Waals energy	Restraints violation energy	Buried surface área	Electrostatic energy (kcal/mol)	Desolvation energy (kcal/mol)	Z score
Cluster1	-91.7 ± 2.7	119	1.5 ± 1.2	-32.9 ± 7.4	5.0 ± 2.0	1017.7 ± 74.4	-297.3 ± 64.4	0.2 ± 4.0	-1.3
Cluster2	-79.1 ± 3.6	23	5.9 ± 0.8	-31.3 ± 5.4	3.3 ± 1.8	1031.1 ± 98.6	-302.9 ± 21.8	12.5 ± 2.8	-0.7
Cluster3	-73.9 ± 6.7	21	9.7 ± 0.8	-25.0 ± 3.0	16.1 ± 17.4	958.2 ± 30.2	-274.5 ± 23.9	4.4 ± 3.0	-0.4

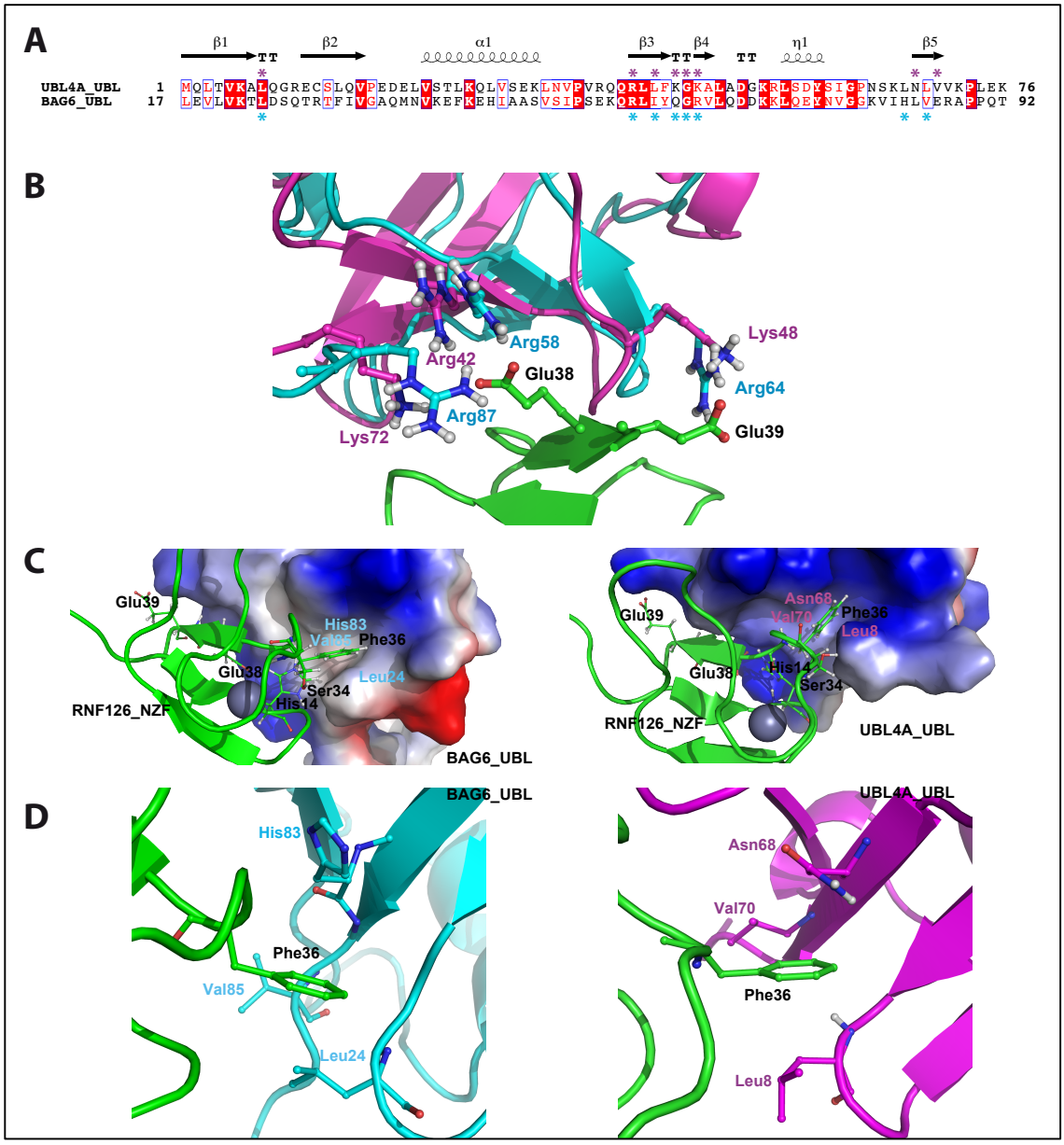

Figure 71: **Structural analysis of RNF126_NZF and UBLs complexes**. A) Structure-based sequence alignment of BAG6_UBL and UBL4A_UBL with the most perturbed residues from the CSP analysis labelled with asterisks; boxes indicate conserved residues while red indicates sequence identity, structural motifs are labelled across the top with 'TT' indicating a β turn. Figure was produced using ESPript 3.0 server. B) Superimposition of the BAG6_UBL complex with RNF126_NZF (experimentally calculated using ARIA) with the HADDOCK generated complex of UBL4A_UBL and RNF126_NZF indicating possible differences in electrostatic interactions at the binding interface. RNF126 is coloured in green, BAG6 in cyan and UBL4A in magenta, with polar residues at the interface shown in ball and stick representation. C) Vacuum electrostatics view of UBL binding pockets (BAG6 on the right and UBL4A on the left) highlighting higher hydrophobicity of BAG6_UBL compared to that of UBL4A_UBL. RNF126 is shown in cartoon representation (green) with the zinc cation shown as a sphere (blue); RNF126 residues involved in interactions are displayed as sticks. D) Detailed cartoon view of RNF126_NZF/UBL hydrophobic interactions showing key differences between residues lining the hydrophobic pocket. The backbone NH group of Leu24 and the aliphatic part of the His83 sidechain that lines the cavity in BAG6_UBL suggests stronger hydrophobic interactions with Phe36 in RNF126_NZF (right panel) compared to those of UBL4A. At the RNF126_NZF/UBL4A_UBL binding interface, the sidechain NH group of Asn68 lines the cavity in UBL4A_UBL and the backbone amino group of Leu8 is solvent exposed (left panel) (174).

4.9 Competition between SGTA (1-86) and RNF126_NZF for the same binding site on BAG6_UBL

Since the RNF126 binding site on BAG6 exhibits some overlap with its SGTA interaction interface (82, 99) we examined whether there is a possibility of simultaneous or competitive binding between these three proteins using NMR and native PAGE mobility shift assay.

4.9.1 CSPs in competitive binding study of SGTA (1-86), RNF126_NZF and BAG6_UBL

First, the direct binding between SGTA (1-86) and RNF126_NZF was tested by reciprocal NMR titrations. It showed unperturbed spectra of each protein upon addition of the other, to a molar ratio of 1:4 (SGTA dimer: RNF) (**Figure 72**), indicating the lack of interaction. Next we analysed the CSP of ^{15}N-labelled SGTA (1-86) upon titration with BAG6_UBL, followed by the addition of RNF126_NZF until saturated. The peaks that are perturbed upon BAG6_UBL binding incrementally return to their free state upon RNF126_NZF addition (detailed panels are shown in **Figure 73 A** and full spectra in **Figure 74**. Similarly, ^{15}N-labelled BAG6_UBL protein was titrated with SGTA and then saturated with RNF126_NZF resulting in the characteristic shifts for SGTA binding pattern changing into the RNF126_NZF binding state upon saturation. Furthermore, the reverse titration was performed by adding RNF126_NZF followed by SGTA, with similar results (**Figure 73 B**) except for the BAG6/RNF126 complex signals not fully disappearing in this case.

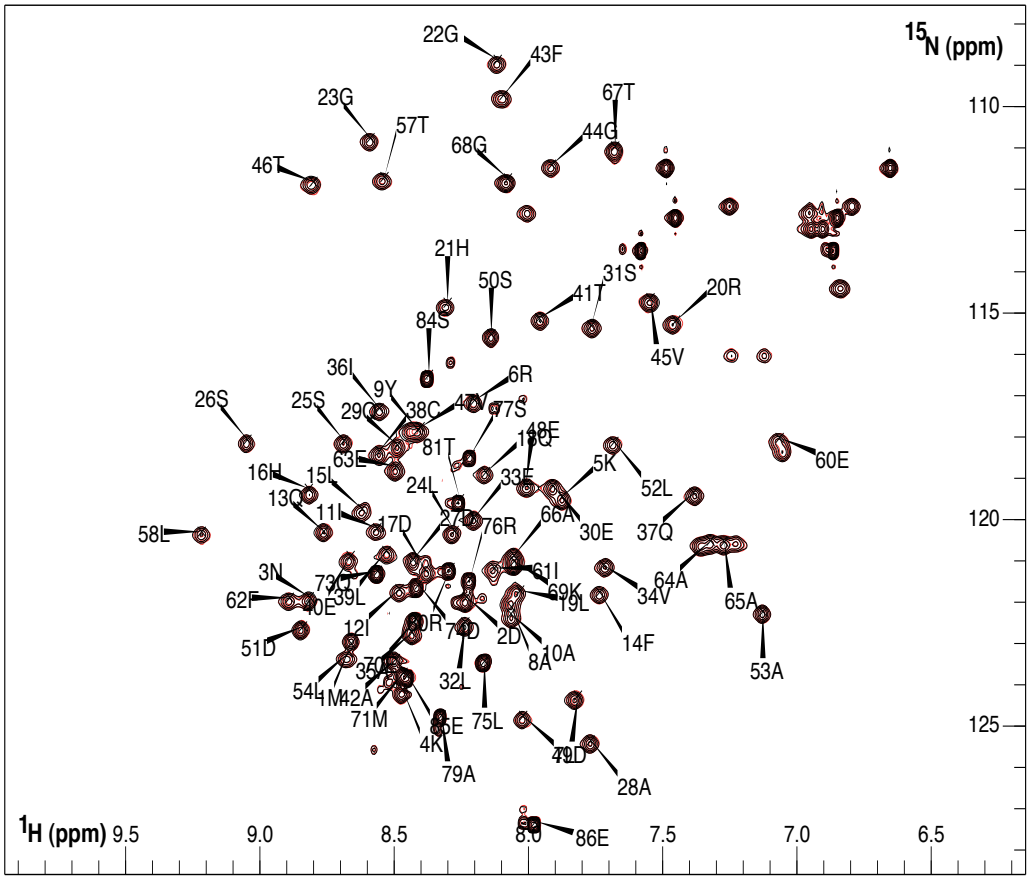

Figure 72: **^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA (1-86) dimer titrated with RNF126_NZF**. ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA_NT dimer (assignment extracted from BMRB accession number 19779) in free form (black), and with a four-fold molar excess of unlabelled RNF126_NZF (red). SGTA (1-86) and RNF126_NZF does not show any evidence of an interaction.

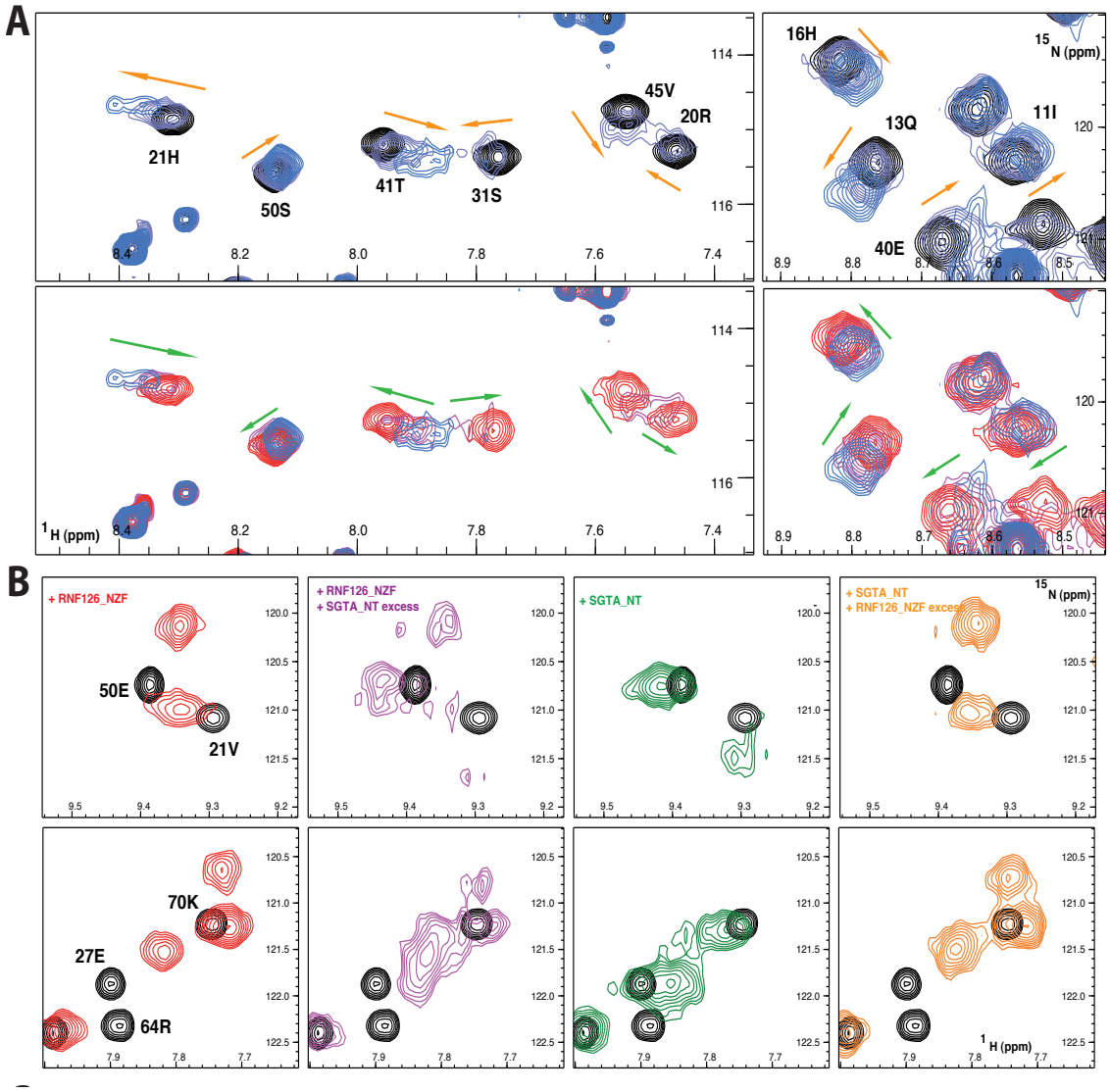

Figure 73: **CSPs in competitive binding study of SGTA, RNF126_NZF and BAG6_UBL.** A) Upper panel - detailed views of overlapping ^{1}H-^{15}N HSQC spectra showing ^{15}N-labelled SGTA (1-86) protein titrated with unlabelled BAG6_UBL at ratios: 1:0 (black), 1:1 (purple) and 1:2 (blue). Lower panel – equivalent views of ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled SGTA in complex with unlabelled BAG6_UBL (1:2 molar ratio) titrated with unlabelled RNF126_NZF at ratios: 1:2:0 (blue), 1:1:1 (magenta) and 1:0.5:1.5 (red). B) Detailed views of ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled BAG6_UBL (black) titrated with RNF126_NZF (red, far left) at a ratio of 1:1 and then saturated with unlabelled SGTA (1-86) (purple) to reach a BAG6:RNF126:SGTA ratio of 1:1:4; or ^{15}N labelled BAG6_UBL titrated first with unlabelled SGTA (1-86) (green) at 1:2 ratio and then saturated with unlabelled RNF126_NZF (orange, far right) at ratios of 1:2:2 (BAG6:SGTA:RNF126). Since SGTA (1-86) dimerises, a two-fold concentration of this protein is necessary to reach equivalent concentrations for BAG6/SGTA complex formation (174).

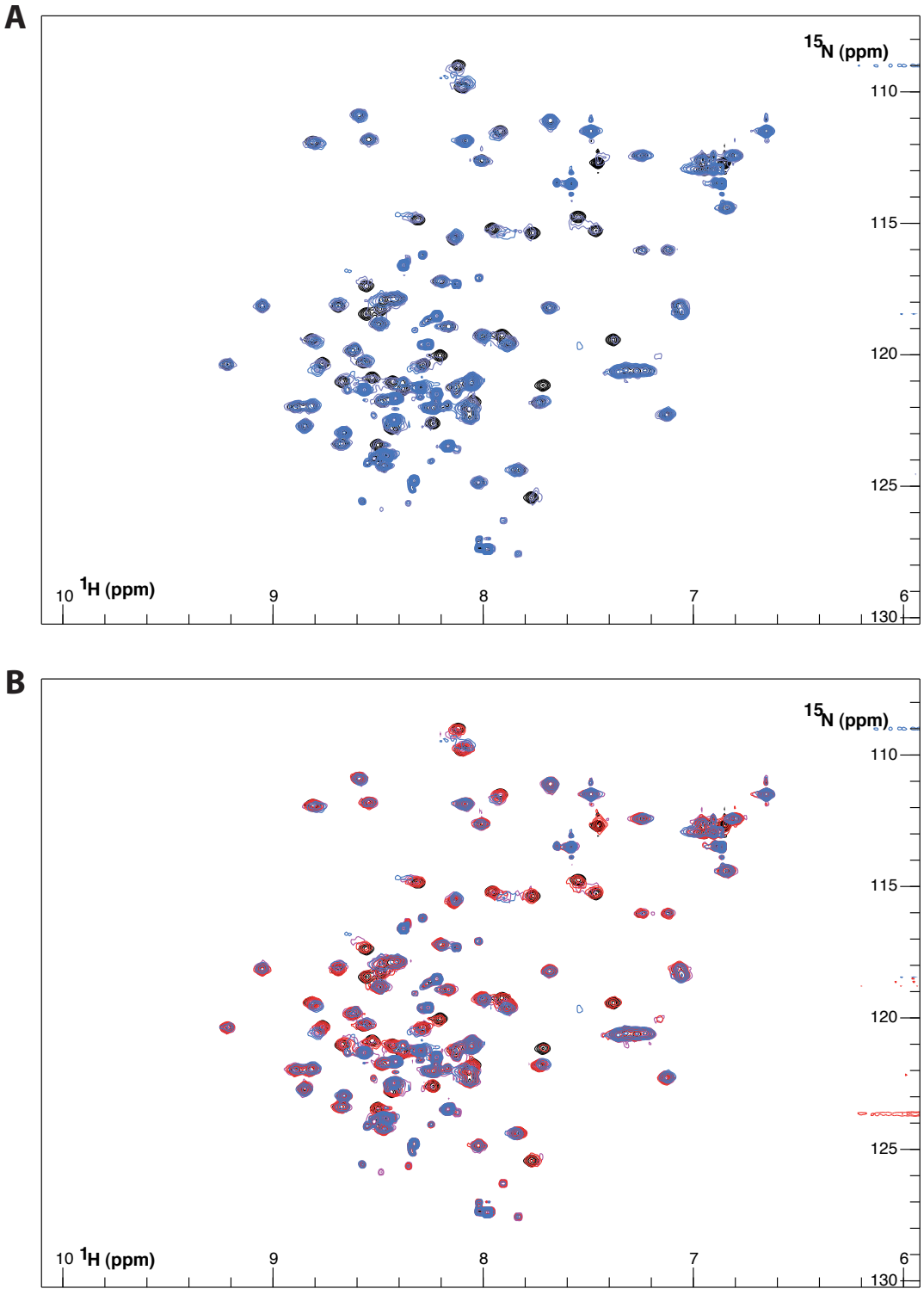

Figure 74: **Expanded ^{1}H-^{15}N HSQC spectra corresponding to the detailed views in Figure 73A**. A) ^{15}N-labelled SGTA (1-86) titration with unlabelled BAG6_UBL. B) ^{15}N-labelled SGTA (1-86) in complex with unlabelled BAG6_UBL titrated with unlabelled RNF126_NZF. Molar ratios and colours as described in the legend for **Figure 73.**

4.9.2 Native PAGE in competitive binding assay of SGTA (1-86), RNF126_NZF and BAG6_UBL

Native PAGE mobility shift assays were employed to further analyse the competition between RNF126_NZF and SGTA (1-86) for BAG6_UBL binding. First, SGTA (1-86) was saturated with BAG6_UBL and a band shift was detected indicating a complex formation. Then, GFP-tagged RNF126 was added into the protein mixture at increasing concentrations which resulted in characteristic band shifts for the free form of SGTA and subsequent complex formation between BAG6_UBL and RNF126. On the contrary, in the reverse titration where BAG6_UBL/RNF126_NZF complex was formed followed by addition of SGTA, the release of RNF126 from the RNF126/BAG6_UBL complex upon competition with SGTA_NT was not observed (**Figure 75**).

Collectively, these experiments are consistent and indicate that RNF126_NZF competes with SGTA (1-86) for the same binding region in BAG6_UBL (and on UBL4A_UBL, **Figure 76**) and exclude the possibility of RNF126 and SGTA binding to BAG6 simultaneously. Furthermore, BAG6_UBL binds more tightly to RNF126_NZF than it does to SGTA (1-86) as the former interaction has Kd ≈ 0.4 μM as calculated from ITC while the ITC data for the latter interaction could not be sufficiently fitted, despite trying a number of experimental conditions. The binding could only be demonstrated by NMR chemical shift perturbation in fast exchange and size exclusion chromatography (82).

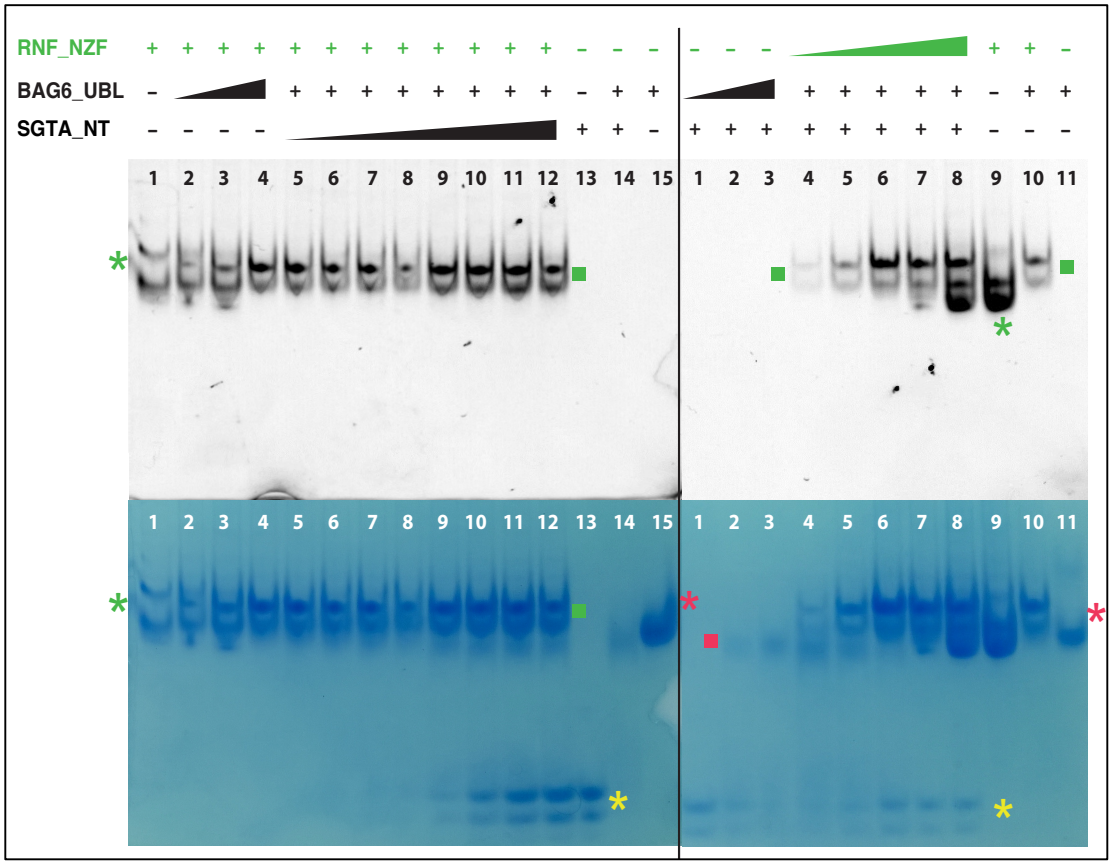

Figure 75: **Native PAGE of GFP-tagged RNF126_NZF competitive binding assay with BAG6_UBL and SGTA_NT visualised by fluorescence (top) and Coomassie (bottom)**. Left panel: lane 1 - free GFP-tagged RNF126_NZF, lanes 2–4 - GFP-tagged RNF126_NZF titrated with BAG6_UBL at 0.2, 0.3 and 1 molar equivalents, lanes 5–12 – GFP-tagged RNF126_NZF in complex with BAG6_UBL (1:1 ratio) titrated with SGTA_NT dimer at 0.1, 0.2, 0.3, 0.5, 1, 2, 3 and 4 molar equivalents, lane 13 – free SGTA_NT, lane 14 – SGTA_NT in complex with BAG6_UBL (1:1) and lane 15 – free BAG6_UBL. Right panel: lane 1 – free SGTA, lanes 2 and 3 – SGTA_NT titrated with BAG6_UBL at 0.3 and 1 molar equivalents, lanes 4–8 - SGTA_NT in complex with BAG6_UBL (1:1) titrated with GFP-tagged RNF126_NZF at 0.2, 0.5, 1, 3 and 4 molar equivalents, lane 9 - free GFP-tagged RNF126_NZF, lane 10 - GFP-tagged RNF126_NZF in complex with BAG6_UBL (1:1), and lane 11 – free BAG6_UBL. Free GFP-tagged RNF126_NZF (green), free BAG6_UBL (red) and free SGTA_NT (yellow) are labelled with asterisks; GFP tagged RNF_126_NZF/BAG6_UBL complex with a green square and SGTA_NT/BAG6_UBL complex with a red square.

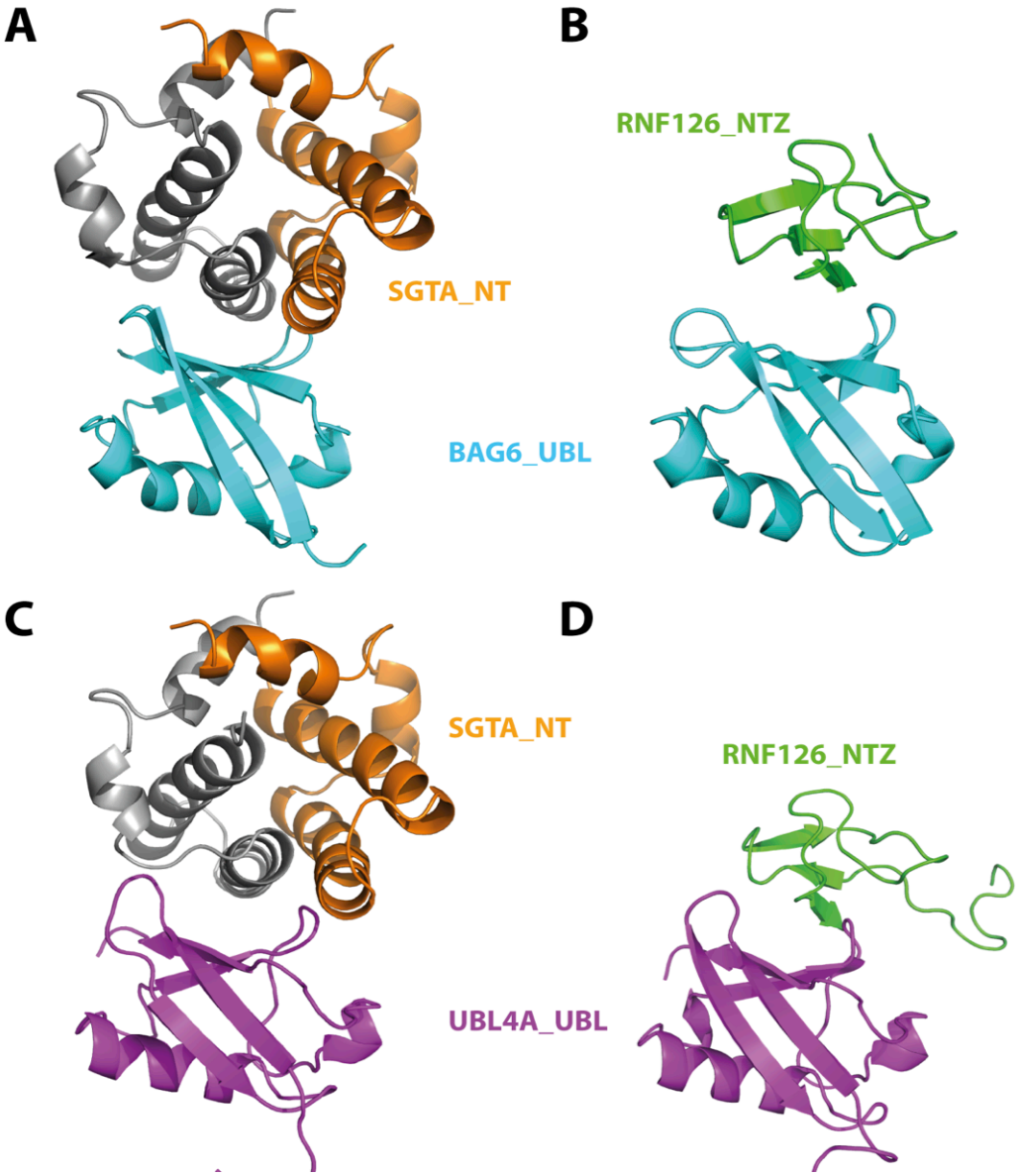

Figure 76: **SGTA_NT and RNF126_NZF interact with the same region of the UBL domains of BAG6 and UBL4A.** A) and B) Cartoon representations of BAG6_UBL in complex with SGTA (A-HADDOCK model (82), Chapter 3) and with RNF126_NZF (B -solution structure). C) and D) UBL4A_UBL in complex with SGTA_NT (C -HADDOCK model (82), Chapter3) and RNF126_NZF (D -HADDOCK model).

4.10 Preliminary expression and purification trials of RNF126_FL and RNF126 (142-218)

The full length RNF126 constructs and hydrophobic region (142-218) constructs were produced to characterise the full length protein and contribute to understanding how RNF126 E3 ligase selects its substrates for ubiquitination. Proteins were expressed in both, Bl-21 and C41 cells to

test expression conditions. Both constructs were not stable after the purification and solubility tag removal and several purification trials yielded none or little protein. The purification and mass spectrometry analysis of both samples are shown on Figure 77**.** Overall the results are not fully conclusive to decide if all the protein sample precipitated or whether the tag removal was not entirely successful. More purification trials are needed in order to find better purification/expression condition.

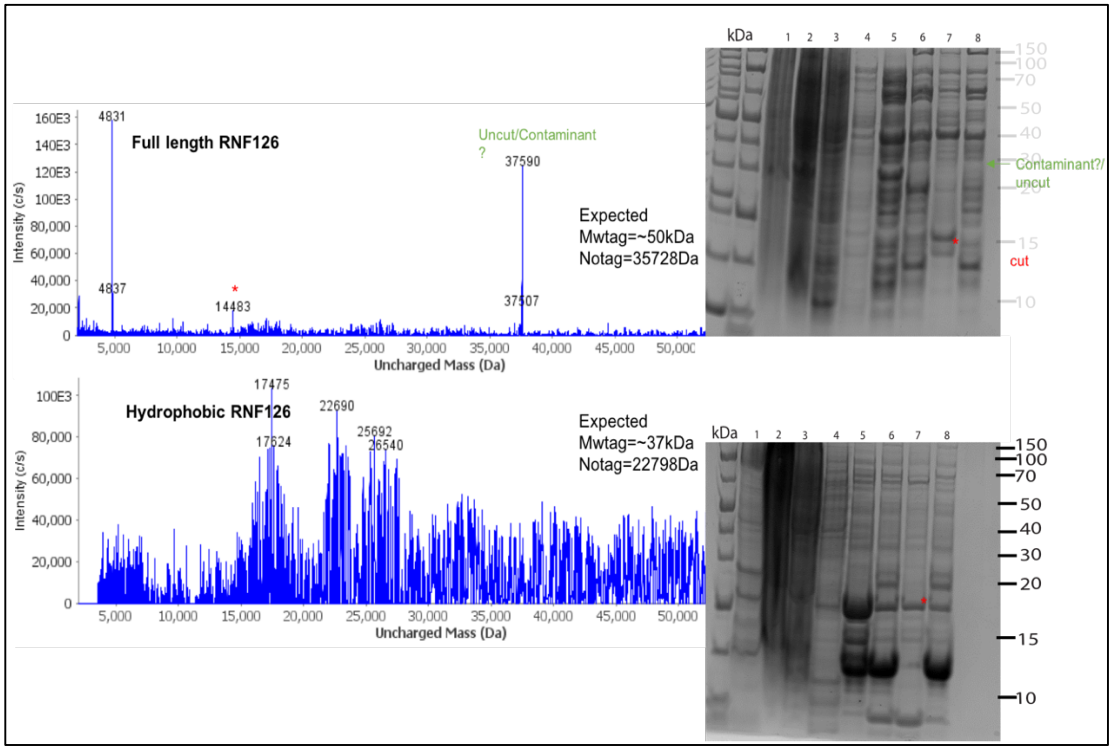

Figure 77: **Mass spectrometry results and SDS-PAGE of RNF126 full length (top) and RNF126 (142-218) purification trials.** Left: annotated mass spectrometry analysis. Right: SDS- PAGE 1) uninduced cells control 2) flow through after column loaded with the cell lysate, 3) wash with 5 CV of wash buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 µ M TCEP), 4) elution with 1 CV of elution buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 µ M TCEP, 300mM Imidazole), 5) additional elution step with 1CV of elution buffer, 6) total cleavage reaction, 7) flow through after column loaded with total cleavage reaction (cut RNF126), 8) elution with 2 CV of elution buffer (uncut protein and tag). Red asterisk indicates cleaved (cut) protein samples, possibly degraded in the case of the RNF126 full length

4.11 Preliminary study of RNF126 interaction with Frataxin

Recently, in vivo studies have shown that the N-terminal portion of Frataxin interacts with full length RNF126 in cell pull-downs (75). However, it is not clear whether it was a full length frataxin precursor or which domain of RNF126 is necessary for the interaction. Since, there were no substrates identified for RNF126 before, we aimed to test whether the two proteins interact directly in solution. Frataxin constructs were kindly provided by our collaborators (Dr Robert Yan from Prof Annalisa Pastore lab, King's College London). The full-length RNF126 could not be produced in sufficient amounts, therefore RNF126_NZF was tested for interaction. The ^{1}H- ^{15}N HSQC spectrum shows some signal broadening upon RNF126_NZF titration into ^{15}N-labelled Frataxin (**Figure 78)** and the reciprocal titration shows similar results (not shown). However, more experiments are needed to verify whether there is an interaction between the two proteins and if it is the N-terminal of RNF126 which interacts.

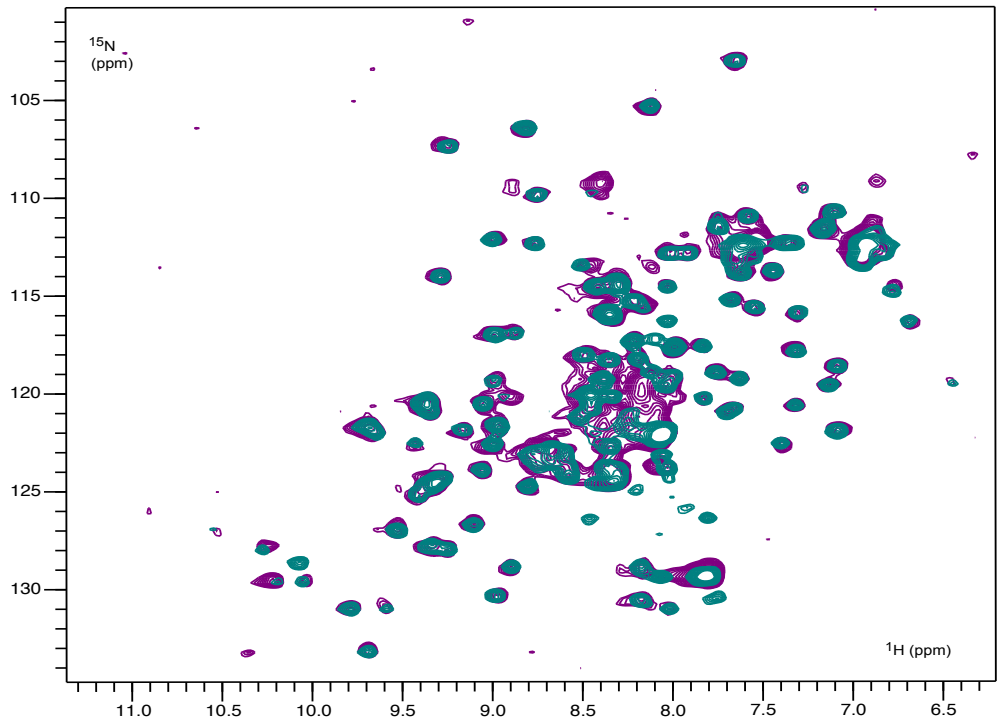

Figure 78: **^{1}H- ^{15}N HSQC of ^{15}N Frataxin (teal) titrated with RNF126_NZF (magenta) in molar ratio 1:2).**

4.12 RNF126 alignment with RNF115

Since, full length RNF126 is not very stable and producing high enough yields for binding studies proved to be challenging, we planned to construct its homologue RNF115 (**Figure 79**). RNF115 is another soluble E3 ligase which shares near identical NZF and RING domains with RNF126 (with 42.3% overall sequence identity and 52.6% similarity). We planned to investigate whether RNF115 can bind to the BAG6_UBL which would support its role in the selective ubiquitination of MLPs.

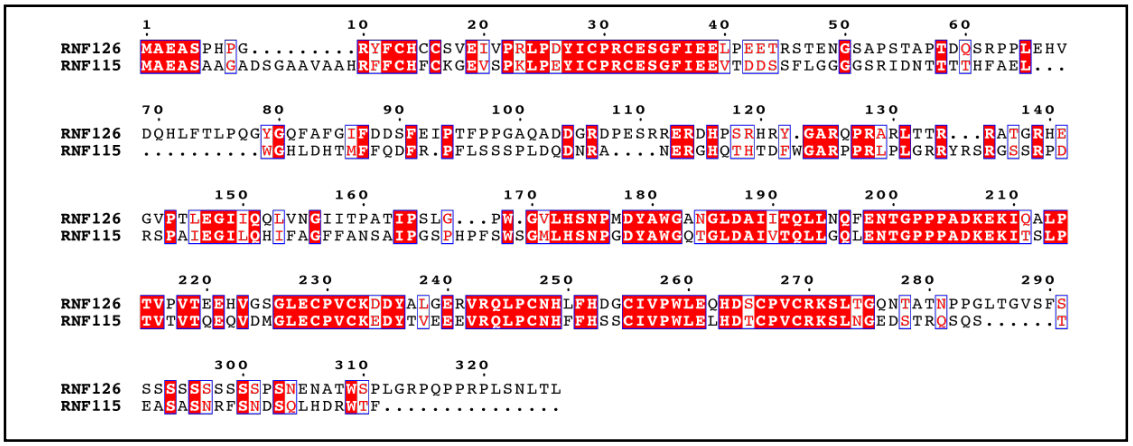

Figure 79: **Sequence alignment of human E3 ligase proteins RNF126 and RNF115.** The proteins present high conservation in the N-terminal zinc finger (10-40 of RNF126) and in a region comprising the RING domain (229-270 of RNF126). Boxes show conserved residues while red highlights sequence identity; figure generated using ESPript 3.0 server.

4.13 Summary

The backbone and side chain assignment of N-terminal RNF126 zinc-finger, together with NOE assignments were used in structural calculations. Moreover, residues 1-40 of RNF126 have been identified to be sufficient to interact with UBL domain of BAG6 and UBL4A. The analysis of the interaction between RNF126_NZF and the UBL domains of BAG6 and UBL4A showed that they do not bind simultaneously but compete for the same binding site, just like SGTA. This has been shown in a reciprocal titration experiment, in which CSP peaks, shifted back to the unbound state upon the addition of the competitive protein.

Binding affinities of these complexes were determined using MST and ITC. MST showed that the binding affinity of RNF126_NZF)/UBL4A_UBL complex was an order of magnitude weaker than the one of RNF126_NZF/BAG6_UBL. However, the binding affinity of RNF126_NZF/BAG6_UBL derived from MST was an order of magnitude stronger compared to the one obtained from ITC (0.4 μM). Furthermore, ITC was not sensitive enough to detect the binding of RNF126_NZF with UBL4A_UBL in any of the tested conditions.

In addition, the structural models generated of complexes of RNF126_NZF with UBLs using NMR (for RNF126_NZF/BAG6_UBL) and HADDOCK (for RNF126_NZF/UBL4A_UBL) showed that the binding interfaces of the two models are essentially similar with some differences discussed in the section 4.8 and in the Discussion (**Chapter 6**).

Chapter 5. Structure and interactions of the TPR domain of Sgt2 with yeast chaperones and Ybr137wp.

The TPR domain of Sgt2 and SGTA is a highly conserved motif known to form transient complexes with molecular chaperones such as Hsp70/Hsp90 (Ssa1 and Hsc82 in yeast) and Ybr137wp (conserved only in yeast). This chapter, describes the first high resolution crystal structure of the free Sgt2_TPR and the same protein in complex with a C-terminal peptide PTVEEVD from Ssa1. Furthermore, it reports the complete assignment of the Sgt2_TPR backbone and, by using NMR and ITC, we demonstrate that Sgt2_TPR interacts with C-terminal fragments of Ssa1, Hsc82 and Ybr137wp with a similar binding mode and comparable binding affinities.

5.1 Protein production

Two Sgt2_TPR constructs (residues 93–229 and 96-225 for the shorter construct) were produced by cloning and overexpression in *E. coli* described in the Methods Section 2.2. Overexpressed protein was recovered in the soluble fraction. Enrichment of expressed protein was achieved by Nickel affinity chromatography, fusion protein and histidine tag removal, and size exclusion chromatography (**Figure 80**). This resulted in pure protein with typical final yields of 1.5g/L for expression in LB and 1g/L for expression in M9 minimal media. Proteins were purified in a buffer at a pH that was kept away from the pI of the individual proteins (for details see Methods).

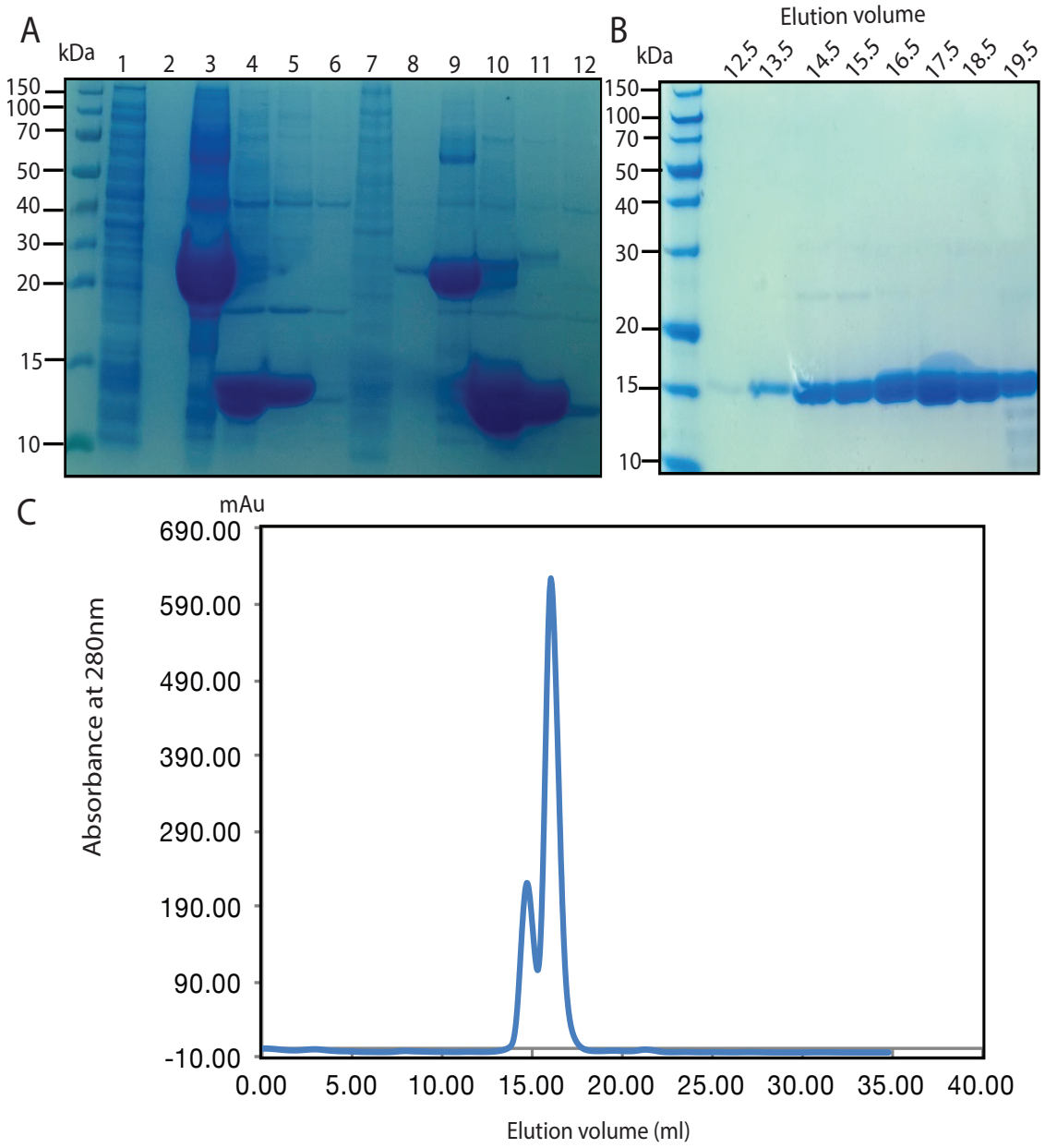

Figure 80: Purification of Sgt2_TPR. A) SDS-page analysis of purification and the removal of the fusion/histidine tag by affinity chromatography of unlabelled Sgt2_TPR (lanes 1-6) and ^{15}N Sgt2_TPR (lanes 7-12). 1) flow through after column loaded with the cell lysate, 2) wash with 5 CV of wash buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 µ M TCEP), 3) elution with 2 CV of elution buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 250 µ M TCEP, 300mM Imidazole), 4) total cleavage reaction, 5) flow through after column loaded with total cleavage reaction, 6) elution with 2 CV of elution buffer. B) SDS-page analysis and C) Chromatogram of size exclusion chromatography using Superdex 75 16/60. Lanes are identified with elution volume in ml.

5.2 Crystallisation of Sgt2_TPR

We aimed to crystallize Sgt2 alone and in complex with C-terminal fragments of Ssa1 (PTVEEVD peptide), Hsc82 (MEEVD peptide) and Ybr137wp (SLEENDNLD) in order to gain insight into how the TPR domain of Sgt2 recognises its targets. The longer construct of Sgt2_TPR (residues 93–229) produced a thick plate crystal (dimensions 200 x 400 μM) in 0.1M SPG, pH 6.0, 25% w/v PEG 1500 within four days (**Figure 81**). It was used to collect a diffraction dataset and solve the X-ray structure of the free Sgt2_TPR. The peptides used for co-crystallisation were prepared as described in Methods (Section 2.1.10). Attempts to crystallise the longer TPR construct in complex with peptides failed, possibly due to the flexible N-terminal and C-terminal ends of the symmetry-related molecules in the crystal occluding the binding interface and preventing complex formation during crystal growth. To demonstrate this, we used an initial model of our x-ray Sgt2_TPR structure and modelled the MEEVD peptide into the binding interface based on the SRMEEVD peptide in TPR_Tah1 complex (PDB accession number: 4CGQ) (**Figure 81**). This supported our theory so we used the model to design a shorter Sgt2_TPR (96-225) construct by removing EDD residues from the N-terminus and EKTV from the C-terminus. This resulted in successful crystallization of the Sgt2_TPR/PTVEEVD complex after seven days. Initially, we observed two different crystal morphologies in the screens including plates and thin needles. The needles crystallized in 0.2M Zinc Acetate, pH 6.3, 20% w/v PEG 3350 and were optimised by varying the PEG concentration from 8-30% and pH from 5.9-7.3. This yielded rod-shaped crystals in 0.2M Zinc Acetate, pH 7.2, 30% w/v PEG 3350 and enabled a dataset collection for the complex (**Figure 81**). We found that zinc was essential for the crystallization of the complex and the crystals produced in conditions without it did not result in the complex formation, suggesting that the metal ion is needed to stabilize complex formation within the crystal. The PTVEEVD peptide was the only one to effectively crystallize in a complex with Sgt2_TPR. All attempts at crystal soaking with increasing concentrations of MEEVD and SLEENDLND peptide solutions resulted in crystal disintegration.

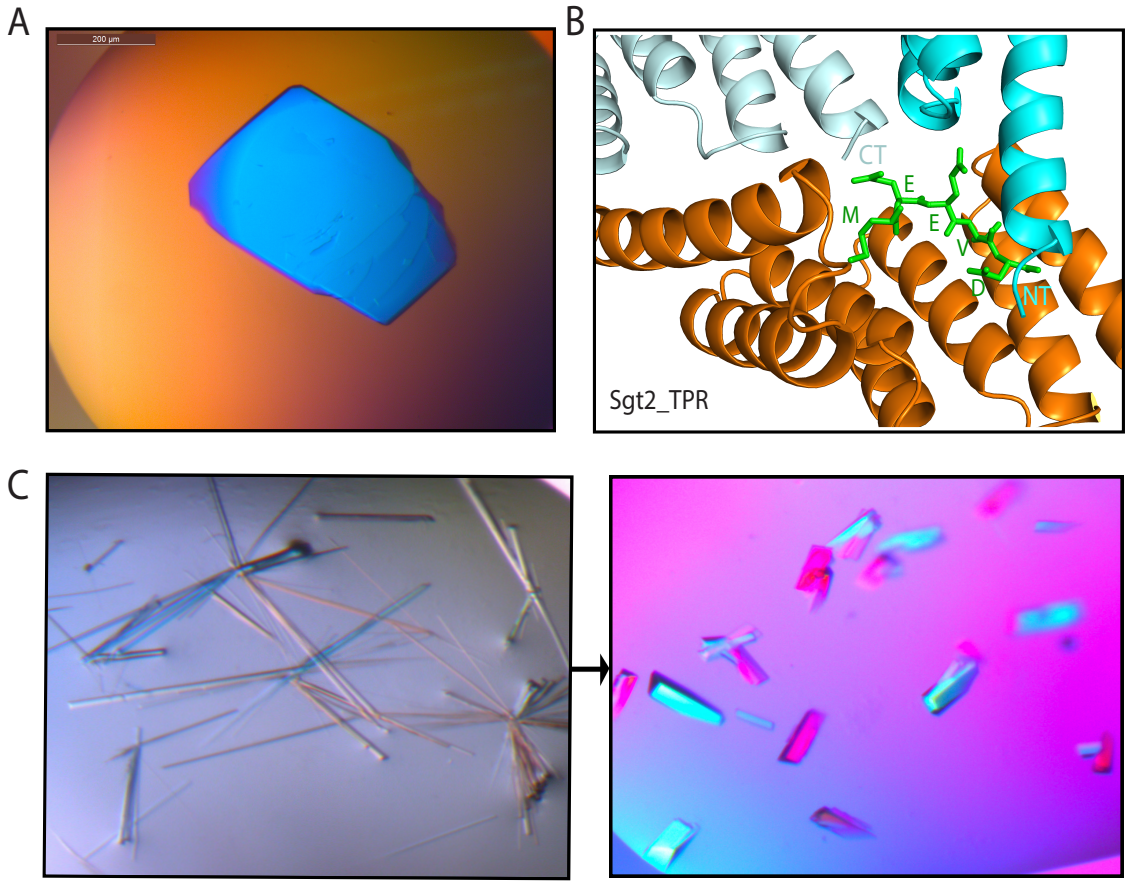

Figure 81: Crystallisation of Sgt2_TPR. A) Crystal morphology of the free Sgt2_TPR, B) A model of Sgt2_TPR with MEEVD peptide modelled in the binding interface based on the SRMEEVD peptide in TPR_Tah1 complex (PDB accession number: 4CGQ) to show the N-terminal (NT) and C-terminal (CT) ends of neighbouring symmetry molecules occluding the binding interface. Model made using Pymol. C) Crystal optimisation from thin needles to rods of the Sgt2_TPR/PTVEEVD complex.

5.3 Overall Sgt2_TPR x-ray structure

The structure of Sgt2_TPR was determined by molecular replacement and refined to 1.55 Å resolution (**Figure 82 A**). The coordinates file and the mtz reflection file were deposited to PDBe after the model was validated by MolProbity (335) (PDB Accession Number: 5LYP; statistical parameters from MolProbity and Aimless in **Table 15**). The coordinates of SGTA_TPR (PDB accession number: 2VYI) were used as a search model due to high structural homology (57% sequence identity) (**Figure 82 C**). The structure was solved with the space group P 21 21 21 and the cell parameters are 36.86Å (a) 50.76Å (b) 67.12Å (c) 90.00° (α) 90.00 ° (β) 90.00 ° (γ). All residues could be built into electron density maps except for the C-terminal Val229 and the solvent-exposed sidechains of Glu93 and Asp94. The final model also contains 90 water

molecules and a single BO_4 ion from the crystallization condition, which were added manually in Coot. The refined structure shows very good stereochemical properties with no Ramachandran outliers, 1% of sidechain outliers and 0.7% of the RSR Z-score (RSRZ) outliers (statistics from the Molprobity report and Aimless shown in **Table 15**). The TPR domain of Sgt2 consists of three TPR repeats, comprising six almost identical α-helices and a C-terminal 'capping' helix (α1=A96-N115; α2= Y118-V131; α3= A136-L149; α4= Y152-I165; α5=F170-183Q; α6=P186-E200; α7=E206-L225) connected by short loops and arranged in an antiparallel fold homologous to that of SGTA_TPR. A structural overlay with the equivalent human domain is shown in **Figure 82 A** (RMSD of 1.13Å over 135Cα).

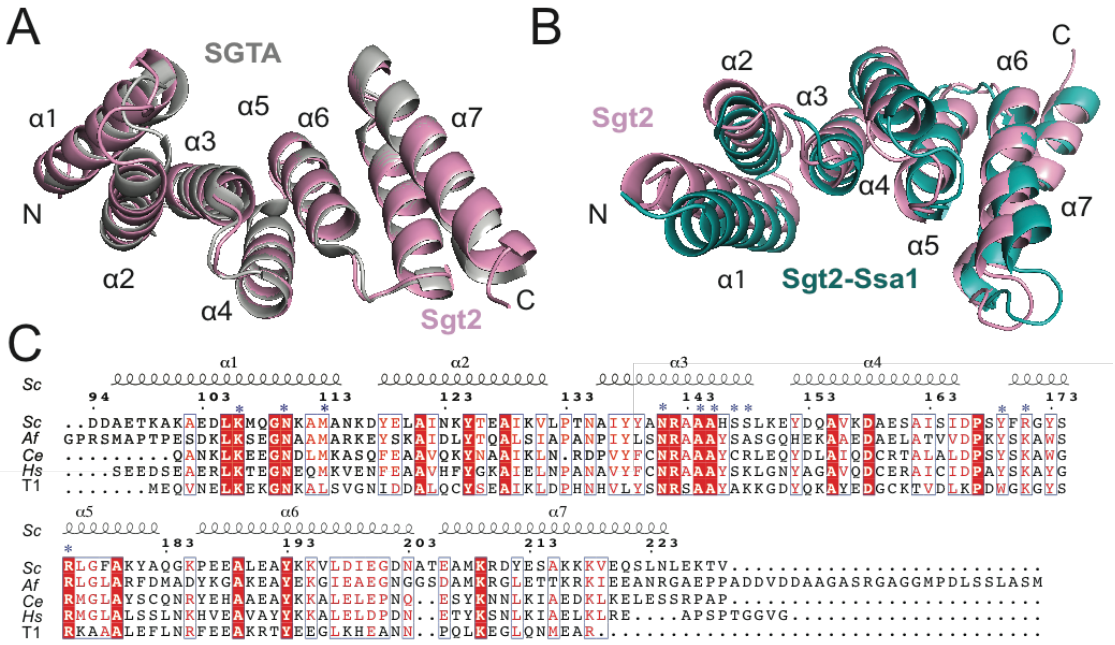

Figure 82: Crystal structure of Sgt2_TPR. A) Superimposition of TPR domains from Sgt2 (magenta, PDB accession code: 5LYN) with human SGTA (gray, PDB accession code: 2VYI) Structures were superposed using secondary-structure matching in ccp4mg (328). The structures align with RMSD 1.13 Å. B). Superimposition of TPR domains from free Sgt2 and Sgt2 bound to the Ssa1 derived peptide PTVEEVD. (teal). The peptides align with RMSD 0.77 Å. C) Structure-based sequence alignment of SGT TPR domains from *S.Cerevisiae* (Sc), *A.fumigatus* (Af), *C.elegans* (Ce), *Homo sapiens* (Hs) and human HOP TPR1, (T1). The residues involved in protein-protein interactions and forming a two-carboxylate clamp are indicated with asterisks; boxes show conserved residues while red highlights sequence identity; structural motifs are labelled on the top. The residue numbering is from *S.Cerevisiae*. Figure generated using ESPript 3.0 server.

Table 15: Data collection and refinement statistics of Sgt2_TPR.

Resolution range	32.31 - 1.55 (1.605 - 1.55)	**CC(work)**	0.939 (0.928)
Space group	P 21 21 21	**CC(free)**	0.921 (0.863)
Unit cell	36.8614 50.7597 67.1162 90 90 90	**Number of non-hydrogen atoms**	1157
Total reflections	232024 (23272)	**macromolecules**	1062
Unique reflections	18908 (1864)	**ligands**	5
Multiplicity	12.3 (12.5)	**Protein residues**	137
Completeness (%)	1.00 (1.00)	**RMS(bonds)**	0.005
Mean I/sigma(I)	12.89 (8.72)	**RMS(angles)**	0.62
Wilson B-factor	8.98	**Ramachandran favored (%)**	1e+02
R-merge	0.2183 (1.061)	**Ramachandran allowed (%)**	0
R-meas	0.2277 (1.104)	**Ramachandran outliers (%)**	0
CC1/2	0.995 (0.975)	**Rotamer outliers (%)**	0.95
CC*	0.999 (0.994)	**Clashscore**	0.95
Reflections used in refinement	18908 (1864)	**Average B-factor**	10.96
Reflections used for R-free	937 (89)	**macromolecules**	10.50
R-work	0.1936 (0.1711)	**ligands**	19.64
R-free	0.2186 (0.2231)	**solvent**	15.95

Statistics for the highest-resolution shell are shown in parentheses.

5.4 X-ray Structure of Sgt2_TPR in complex with Ssa1 C-terminus

The Sgt2_TPR (96-225)/PTVEEVD complex structure was determined by molecular replacement using our recently solved Sgt2_TPR (93–229) structure as a search model (**Figure 83**, PDB

Accession Number: 5LYN, **Table 16** for statistics). The structure was solved with space group P 1 21 1 and cell dimensions, 45.49Å (a) 61.09Å (b) 55.25Å (c) 90.00° (α)108.81° (β) 90.00 ° (γ). Two copies of the TPR domain were present in the asymmetric unit (**Figure 86 A**). All residues of Sgt2_TPR could be built into electron density maps in chain A and chain B. The two peptides (chain C and D) showed partial occupancy at the binding groove and were thoroughly verified after building. They were modelled into a fair electron density (**Figure 83 A**) and then verified by producing a simulated annealing omit map (**Figure 84**). The overall structures of the two chains, A and B, are essentially identical in the case of backbone (RMSD of 0.77Å over 135Cα), with some deviations observed for the R171 sidechains due to their flexibility and significant differences in the modelled C and D peptide (**Figure 86 A** and **B**). The electron density was less unambiguous for peptide C at chain B, especially for Pro1 with some electron density appearing in the Fo-Fc map which could be suggesting the presence of another atom. However, neither the zinc ion or water molecule could be fitted. Nevertheless, all PTVEEVD (1-7) residues were successfully modelled. The occupancy for both peptides (chain C and D) were refined to an Rfactor of 0.158 and an Rfree 0.202, and both converged to occupancy 0.93. The complex was refined with Phenix script (322) using NCS and TLS (321). TLS groups for both protein chains and peptides were automatically identified by Phenix. The NCS option was run using chain Aresidues as a reference and compared with chain B. We choose to automatically check for rotamer consistencies without restraining B_factors. The final model also contains 148 water molecules, nine Zinc ions and a single BO_4 ion added manually in Coot. Zinc ions were added at peaks of the phased anomalous difference map (DANO) (Appendix). The DANO map was analysed in Coot after opening in an expert mode and showed densities for all nine zinc atoms. The refined structure shows good stereochemical properties with no Ramachandran or sidechain outliers and 1.1% of the RSR Z-score (RSRZ) outliers (statistics from the Molprobity report are shown in **Table 16**).

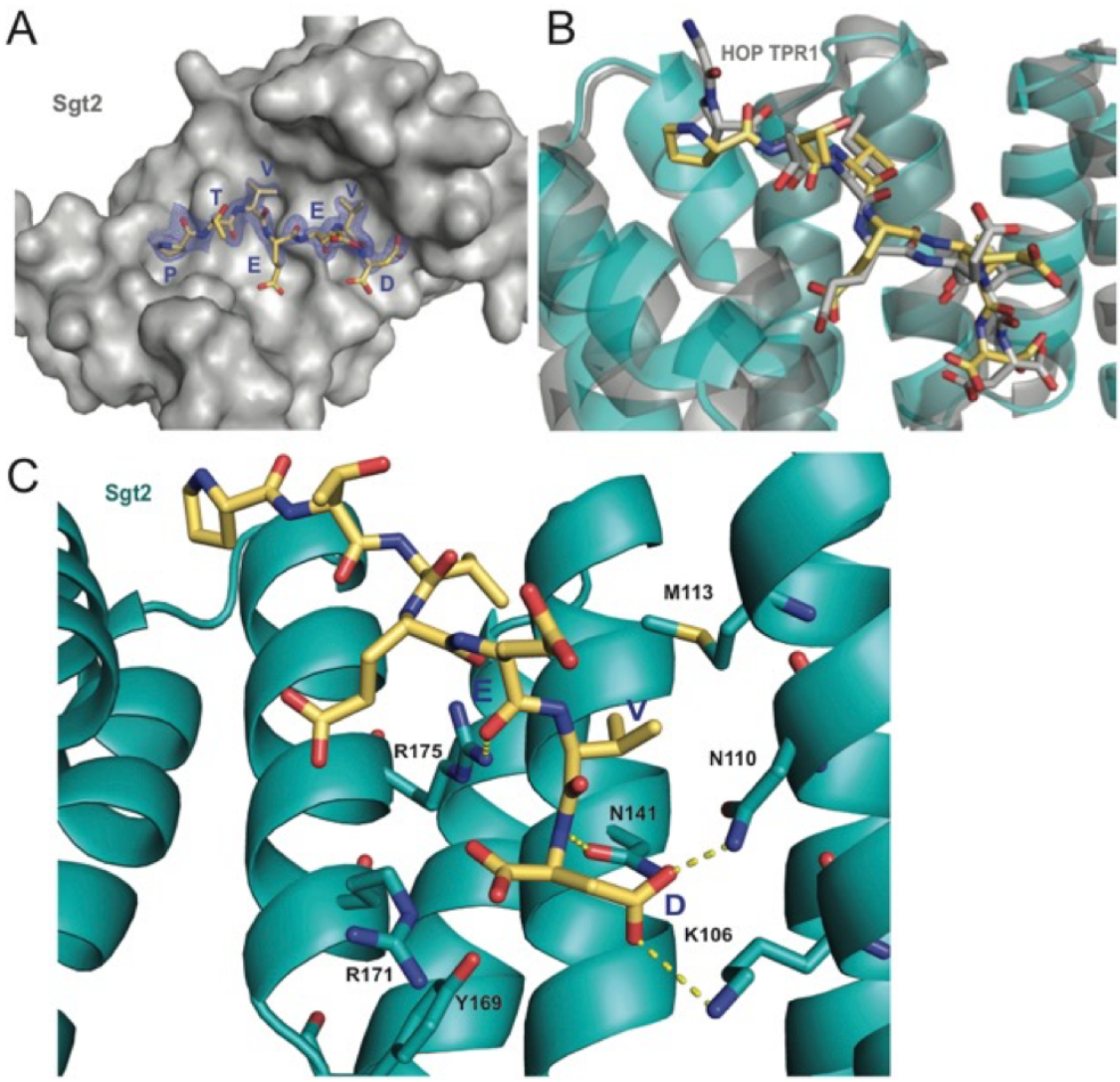

Figure 83: Crystal structure of the Sgt2_TPR/ PTVEEVD complex. A) Surface representation of the Sgt2 hydrophobic groove with bound Ssa1 derived PTVEEVD peptide (PDB accession number 5LYP). The 2Fo-Fc map for the peptide was contoured at 1.0 σ. B) Superimposition of Sgt2_TPR/ PTVEEVD complex (peptide in yellow) onto Hsc70 peptide-bound HOP TPR1A (peptide in grey, PDB accession number 1ELW) highlighting similarities in peptide conformation at the binding interface. The peptides align with RMSD 0.52 Å. C) Network of interactions formed at the complex interface. Residues shown as sticks are involved in the formation of hydrogen bonds or electrostatic interactions (shown as yellow dashed lines). Residues K106, N110, N141, R171, R175 and Y169 are involved in the formation of two-carboxylate clamp and M113 is involved in hydrophobic interactions.

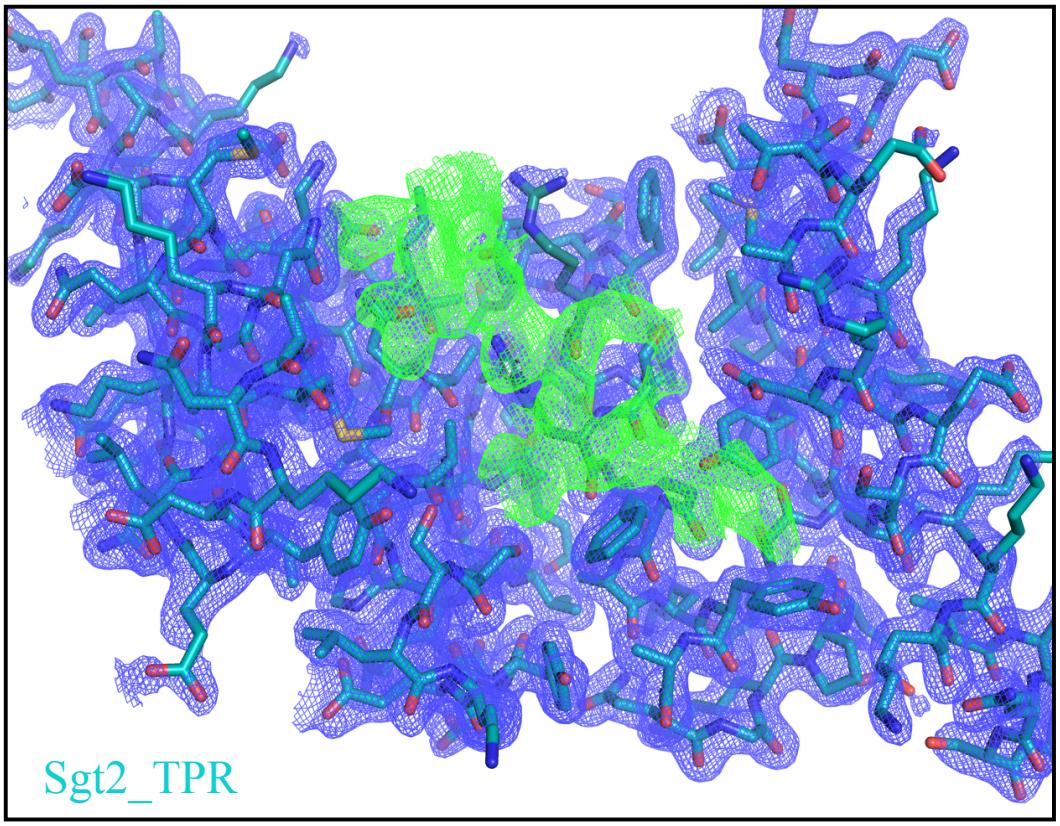

Figure 84: **Sgt2_TPR groove with 1σ 2Fo-Fc map calculated for the peptide.** Fo-Fc map of Sgt2_TPR groove with 1σ 2Fo-Fc map calculated for the peptide region (green) using the final model from which the peptide was omitted following simulated annealing to remove the model bias.

The interaction between the Sgt2_TPR and a PTVEEVD peptide is mostly driven by the formation of a two-carboxylate clamp. Most of the electrostatic interactions between the TPR domain and the peptide occur in the C-terminal EEVD region and anchor the peptide in place. Direct backbone contacts involve hydrogen bond formation between the carboxamide sidechain of Asn141 and Asn110 in Sgt2_TPR, and the sidechain of the terminal Asp7 of the Ssa1 peptide. Also, the sidechain amine of Lys 106 binds to the same carbonyl sidechain of Asp7. The guanidinium group of Arg171 forms a salt bridge with the carbonyl main chain of Asp7 and forms an additional internal contact with Tyr169. The Arg175 sidechain interacts with the carbonyl main chain of Glu4 and Glu5 of the peptide (**Figure 83 C**). Moreover, the N-terminal of the peptide is involved in hydrophobic and van der Waals interactions. Phe178 and Tyr181 contribute to creating hydrophobic pockets and interact with the aliphatic part of Thr2 and Pro1 respectively.

Met113 makes a hydrophobic contact with Val6 of the peptide (**Figure 85**). The 'two-carboxylate clamp' binding mode is characteristic for TPR domains interacting with the conserved C-terminal IEEVD and MEEVD motifs of Hsp70 and Hsp90 chaperones, respectively (261, 336). The comparison of Sgt2_TPR/PTVEEVD binding interface with a previously published complex of HOP TPR1/GPTIEEVD (Hsp70-derivative) (261, 336) shows that the PTVEEVD peptide occupies the same position at the Sgt2 TPR groove as GPTIEEVD. Both peptides essentially overlap except for the difference in the conformation of the main chain and the sidechain of the terminal Asp7 (**Figure 83 B**). This however corresponds to the Asp7 conformation in our chain B peptide bound Sgt2_TPR/PTVEEVD complex structure. Other differences between chain A and chain B are the sidechain conformation of chain B Arg171 is slightly different due to its flexibility and small variations in the orientation of Pro1, Thr2 and Val3 sidechains. The differences between the peptides modelled within the two Sgt2_TPR groove chains cause a change at the interface. In Chain B the sidechains of Asn141 and Asn110 of Sgt2_TPR bind to the main chain of the terminal Asp7 of the Ssa1 (Hsp70) peptide. Also, the hydroxyl group of Tyr169 forms a hydrogen bond with Asp7 peptide sidechain (**Figure 83 C** and **Figure 86 B**).

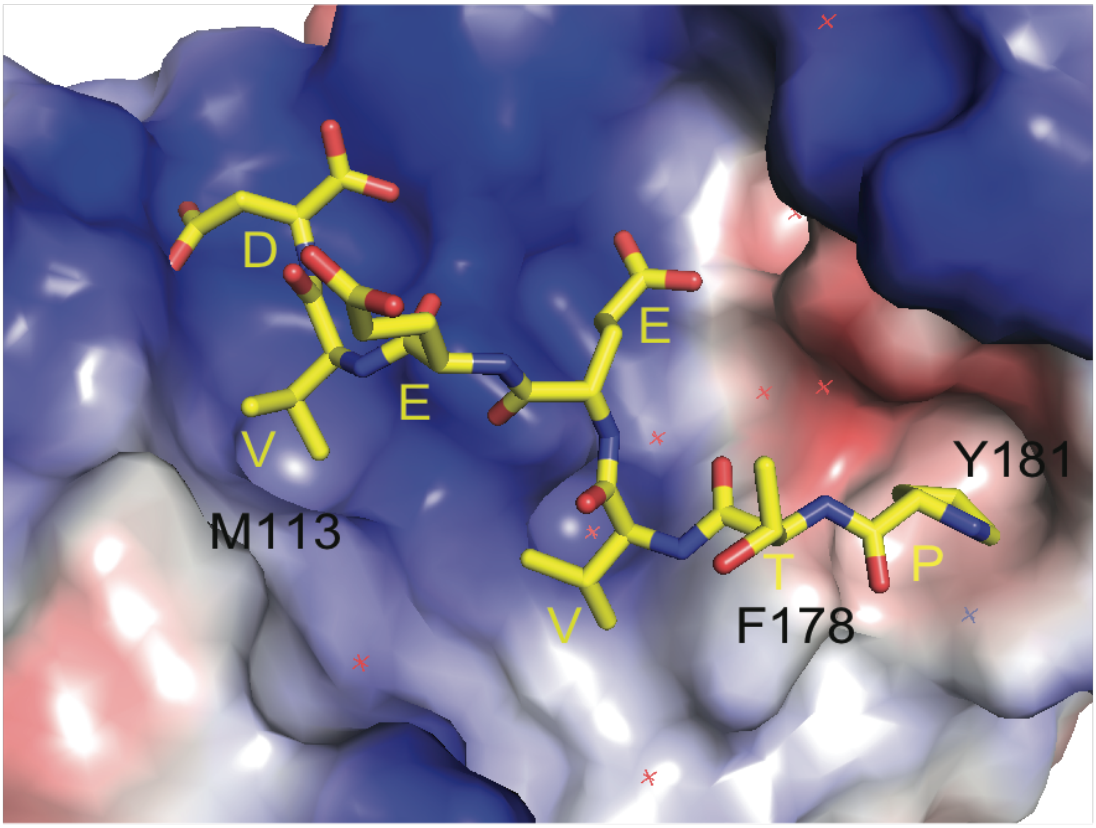

Figure 85: Vacuum electrostatics view of Sgt2_TPR hydrophobic pockets. Vacuum electrostatics view of Sgt2_TPR hydrophobic pockets highlighting hydrophobic interactions between M113, F178 and Y181 and Valine, Threonine and Proline of the PTVEEVD peptide (shown as sticks). The C-terminal end of the peptide is involved in electrostatic interactions. Red corresponds to the negatively charged electrostatic potential on the surface; blue for the positively charged and grey for the neutral. The bound peptides are shown in a stick representation. Red crosses correspond to the water molecules.

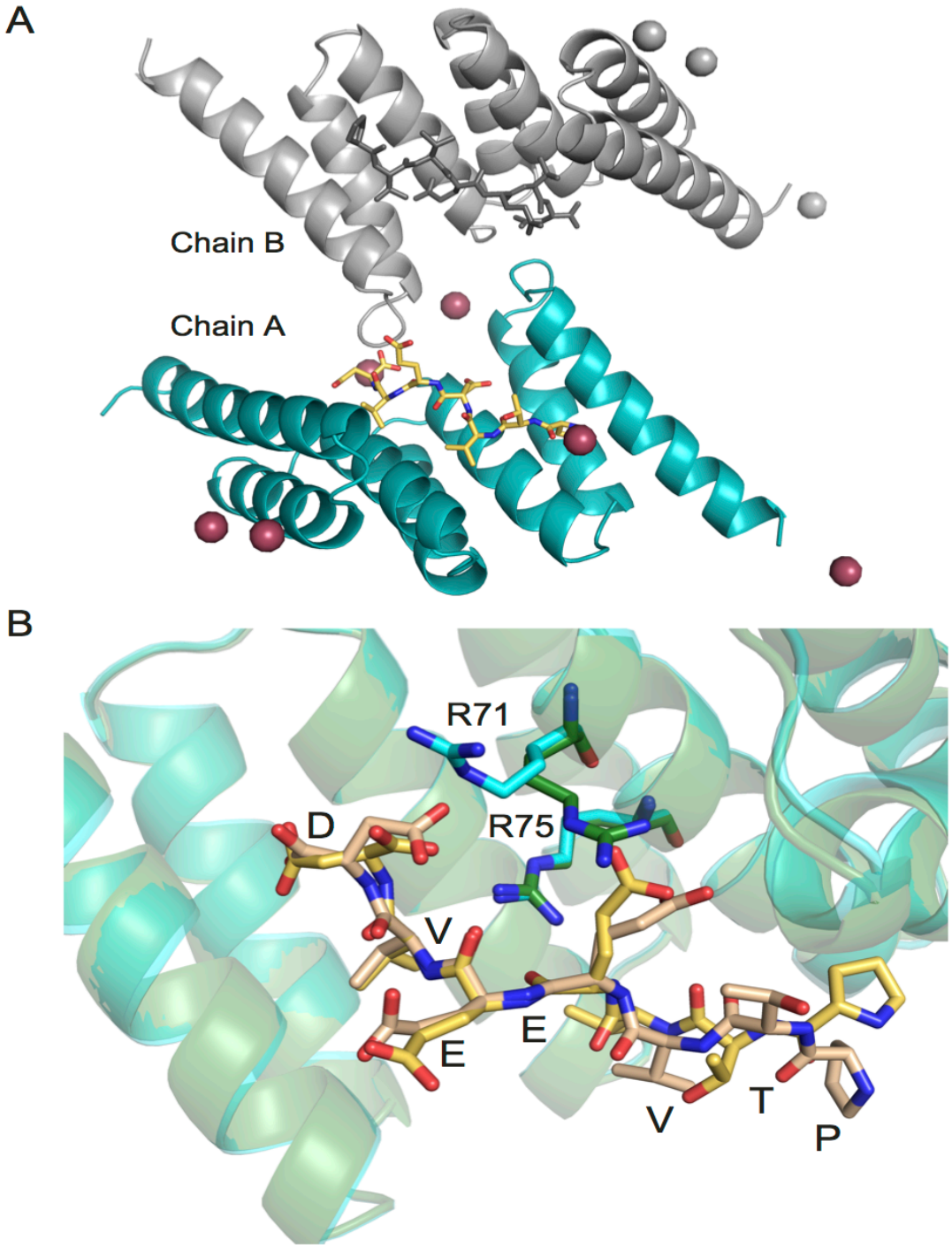

Figure 86: The asymmetric unit of Sgt2_TPR/PTVEEVD complex x-ray structure. A) The asymmetric unit shown as a cartoon representation of Sgt2_TPR chain A (teal) and chain B (grey) and bound peptides (yellow and grey respectively). Zinc ions shown as balls. B) Superimposition of Sgt2_TPR Chain A (teal) and Chain B (green) bound peptides (yellow and wheat respectively) highlighting main differences. Peptides were superposed using secondary-structure CA matching in ccp4mg. The peptides align with RMSD 1.24Å.

Table 16: Statistics for the highest-resolution shell are shown in parentheses.

Resolution range	33.59 - 2.0 (2.071 - 2.0)	**CC(work)**	0.845 (0.345)
Space group	P 1 21 1	**CC(free)**	0.725 (0.476)
Unit cell	45.49 61.09 55.25 90 108.81 90	**Number of non-hydrogen atoms**	2340
Total reflections	117547 (10486)	**macromolecules**	2183
Unique reflections	19189 (1904)	**ligands**	9
Multiplicity	6.1 (5.6)	**Protein residues**	280
Completeness (%)	0.99 (0.97)	**RMS(bonds)**	0.006
Mean I/sigma(I)	6.49 (2.84)	**RMS(angles)**	0.68
Wilson B-factor	25.49	**Ramachandran favored (%)**	1e+02
R-merge	0.354 (0.8515)	**Ramachandran allowed (%)**	0.37
R-meas	0.3851 (0.9392)	**Ramachandran outliers (%)**	0
CC1/2	0.92 (0.531)	**Rotamer outliers (%)**	0
CC*	0.979 (0.833)	**Clashscore**	2.77
Reflections used in refinement	19279 (1903)	**Average B-factor**	30.91
Reflections used for R-free	928 (83)	**macromolecules**	30.60
R-work	0.1576 (0.1903)	**ligands**	50.77
R-free	0.2025 (0.2587)	**solvent**	34.38

Statistics for the highest-resolution shell are shown in parentheses.

5.5 Backbone ^{1}H, ^{13}C, and ^{15}N Chemical Shift Assignments for Sgt2_TPR

The complete battery of NMR triple resonance experiments using the longer construct of Sgt2_TPR (93–229) were performed to facilitate the full backbone assignment. ^{15}N HSQC and the following 3D experiments (HNCO, HNCA, HN(CA)CO, CBCA(CO)NH, and CBCANH) were recorded on a 500 MHz spectrometer. Peaks were picked manually using CCPN Analysis

and the sequential assignments for the shifts of $C\alpha^{i}$, $C\beta^{i}$, $C\alpha^{i-1}$, $C\beta^{i-1}$, C'^{i}, C'^{i-1} were manually performed using the same software. A number of overlapped amide resonances were observed in the middle of the spectrum and were initially difficult to resolve, therefore an additional ^{15}N HSQC, HNCO and CBCANH experiments were recorded at 700 MHz to improve the quality of the data. All the protein residues were assigned except for Prolines and Arg171 for which NH peaks were missing (**Figure 87**).

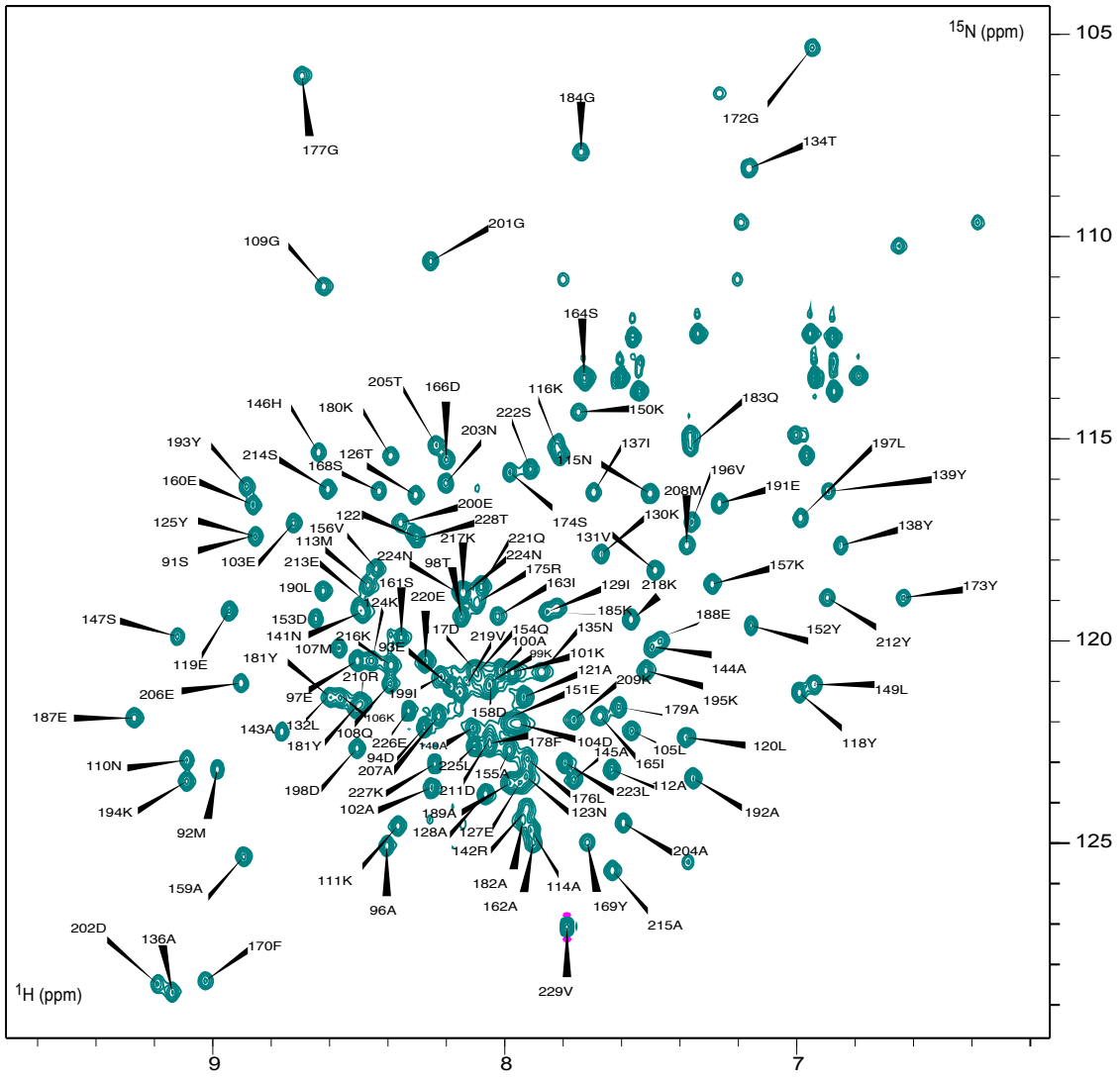

Figure 87: ^{15}N HSQC assigned spectrum of Sgt2_TPR.

Prior to the BMRB submission, the chemical shifts from the following experiments: HNCO, HN(CA)CO, CBCA(CO)NH, and CBCANH were converted into the required STAR format (Appendix). The deposited chemical shifts (accession number BMRB ID: 27044) were checked

for referencing and assignment errors by the software PANAV (run by BMRB team). The validation report indicated no deviant assignments and an unusually high chemical shift for one residue, Glu4 of Cβ = 41.206, where the average chemical shift values for this residue is 30.01. Closer inspection indicated a good connectivity of Cα , Cβ ,and C for the preceding and following residues and therefore the assignment remained unchanged. All chemical shifts were calibrated by the reference offsets: C′: -0.58ppm, Cα: -0.67ppm, Cβ: 0.45ppm, N: -0.75ppm.

5.5.1 TALOS

TALOS+ was run to predict the backbone torsion angles from the chemical shifts assignments and to generate the random coil index (RCI) with a predicted order parameters for each residue and a chemical shift index (CSI) (**Figure 88**). Secondary structure prediction/RCI perfectly agrees with the structure solved by crystallography indicating the presence of seven α helices and the flexible N-terminal and C-terminal ends. .

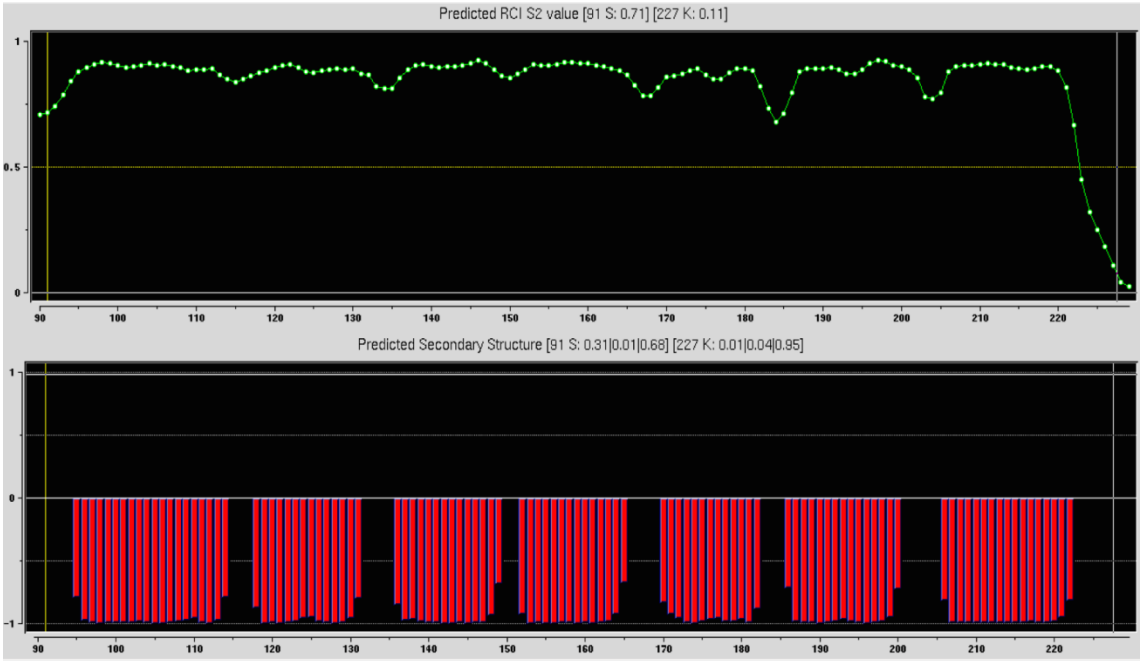

Figure 88: Secondary Structure and Random Coil Index (RCI)-S2 Prediction Window of Sgt2_TPR. The RCI shows the predicted order parameter S2 (upper panel) and CSI-predicted secondary structure (lower panel; red, helix) for all residues. The height of the bars reflects the probability of the neural network secondary structure prediction. The residue number is labelled at the bottom of the panel.

5.6 Interactions of the TPR domain of Sgt2 with yeast chaperones and Ybr137wp

To understand the molecular details of the interactions between Sgt2_TPR with chaperones and Ybr137wp, we focused on probing protein-peptide interactions using the C-terminal fragments of Ssa1 (PTVEEVD), Hsp82 (MEEVD) and Ybr137wp (SLEENDNLD) in binding experiments combining NMR and ITC. NMR allowed us to identify the residues involved in the interaction for all three peptides and ITC delivered binding affinities.

5.6.1 Chemical shift perturbation assays

Chemical shift perturbation (CSP) experiments were carried out by titrating unlabelled peptides PTVEEVD, MEEVD and SLEENDNLD into ^{15}N-labelled Sgt2_TPR (93–229) with up to a 6-fold molar excess of PTVEEVD and MEEVD and a 5-fold excess in the case of SLEENDNLD. The ^{15}N HSQC experiment was used to monitor for chemical shift changes. We could not produce SLEENDNLD at the 6-fold molar excess due to the lower solubility of this peptide in water or buffer. The NMR backbone assignment of Sgt2_TPR (described in the Section 5.5) allowed us to identify the residues involved in the interactions in all three titrations. The full ^{15}N HSQC spectrum and the selected enlarged peaks for the titration experiments are shown in **Figure 89**, **Figure 90** and **Figure 91**. The CSP graphs displaying chemical shifts plotted against the residues of Sgt2_TPR are presented in **Figure 93**. CSP analysis exhibited a similar pattern for all three peptide titrations and showed binding in a fast exchange regimen, with a number of peaks shifting in a non-linear manner suggesting the formation of an intermediate during the titration. We analysed the Sgt2_TPR CSPs by applying a titration cut-off at 3 molar equivalents of the peptide for all the peaks and then dividing them into two groups. The first group consisted of peaks which shift between 0 and 3 molar equivalents and was named the ‘first event’ and the second group, called ‘second event’ comprised the peaks shifting between 3 and 6 (or 5 in the case of SLEELDNLD) molar equivalents (marked as red and black arrows respectively in **Figure 90 B, Figure 91 B Figure 92 B**. The CSP of Sgt2_TPR/PTVEEVD binding for the ‘first event’, plotted

onto our x-ray structure is shown in **Figure 92 A** on the left and the ‘second event’ on the right. The most perturbed resides for the ‘first event’ correspond to the residues at the binding interface in the x-ray complex structure and/or the neighbouring residues. The behaviour of the two subsets of peaks and what it means for the interaction is discussed in the Discussion (Chapter 6).

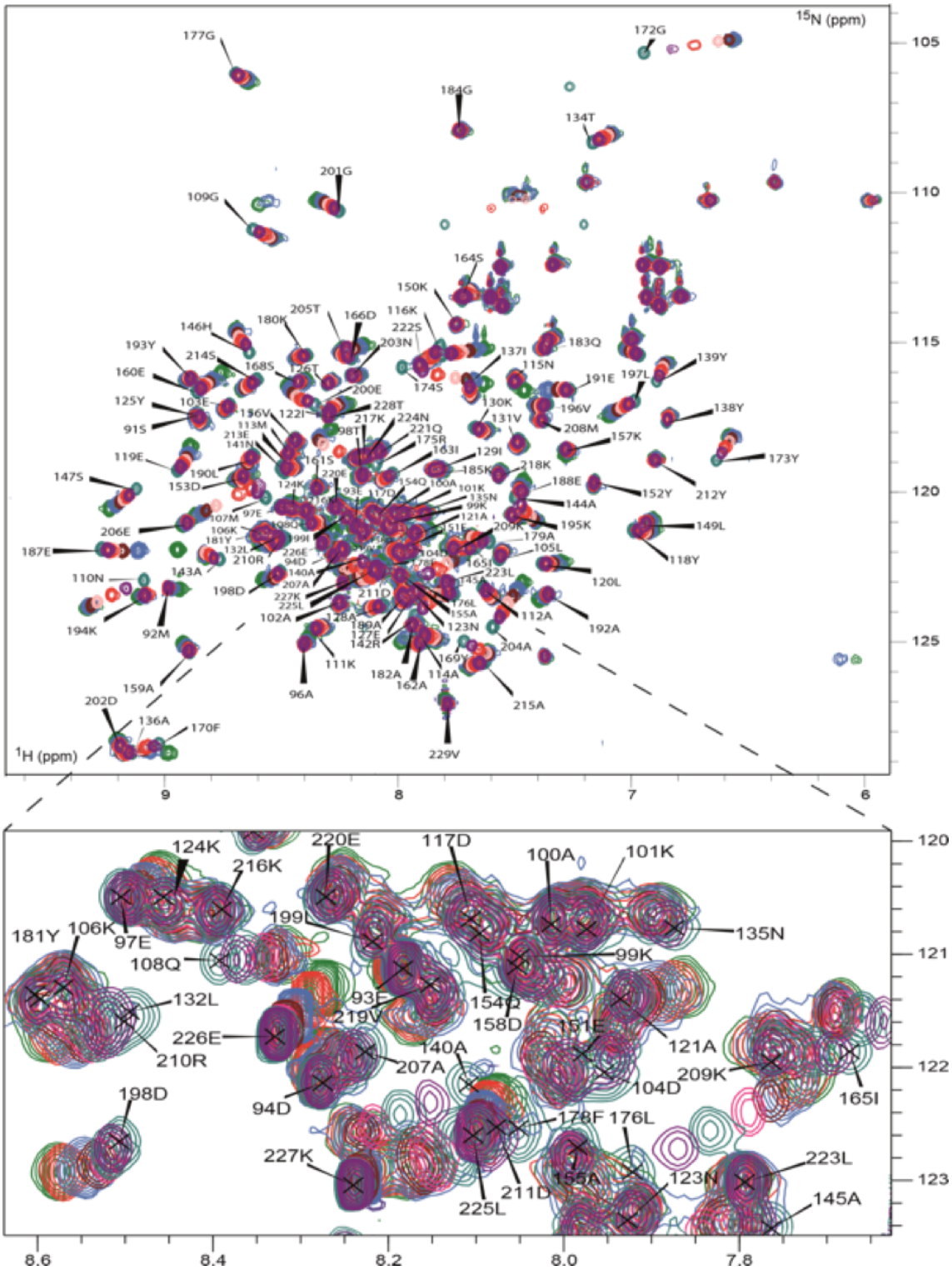

Figure 89: The ^{15}N HSQC spectrum of ^{15}N-labelled Sgt2_TPR titrated with PTVEEVD. The CSP of ^{15}N-labelled Sgt2_TPR peaks upon binding to unlabelled PTVEEVD at ratios: 1:0 (teal, assigned), 1:0.5 (purple), 1:1 (red), 1:2 (light pink), 1:3 (moroon), 1:4 (blue), 1:5 (cyan), and 1:6 (green). Bottom: Enlarged view of the central part of the 15N HSQC spectrum to show detail of the overlapped peaks.

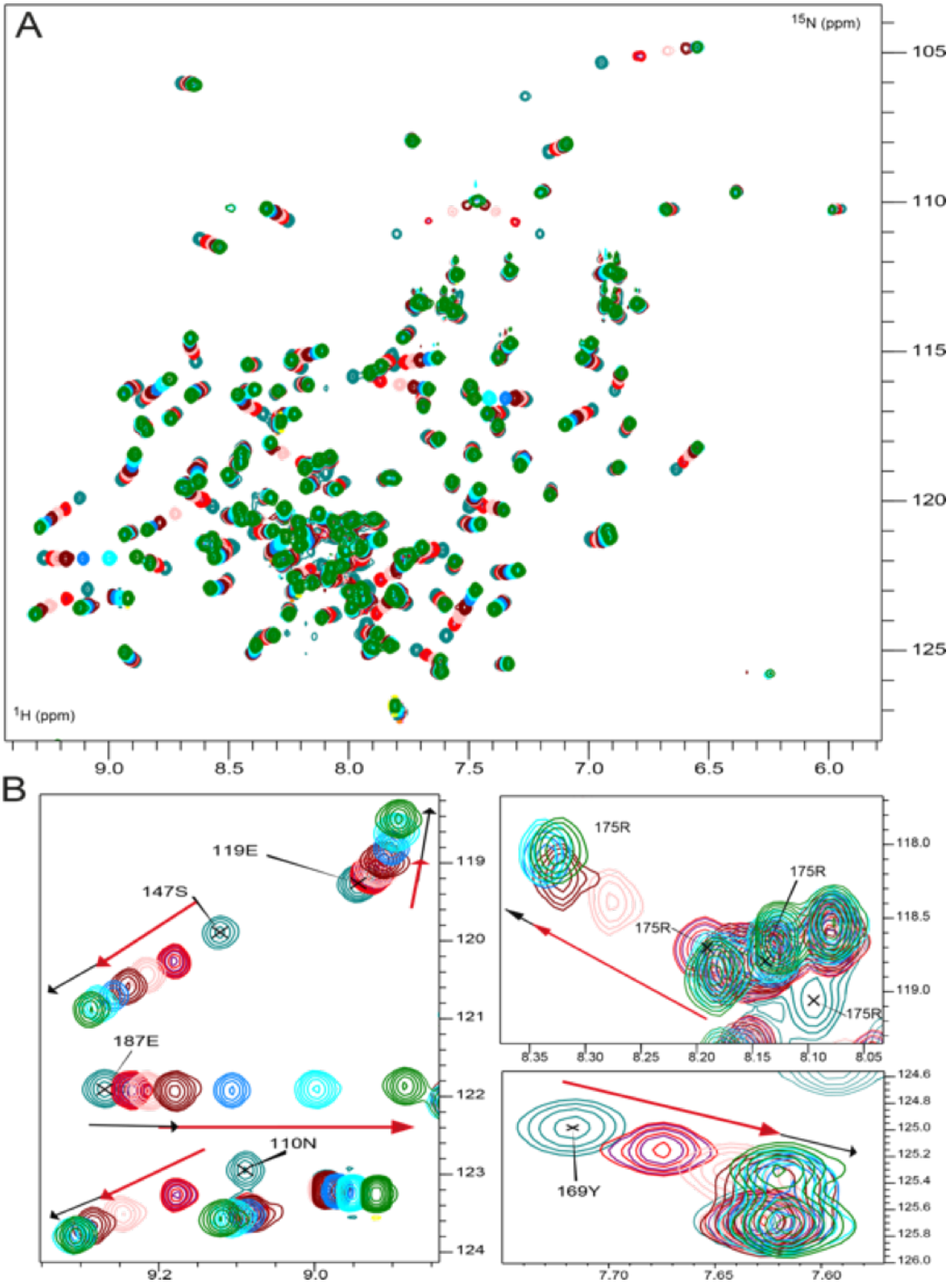

Figure 90: The ^{15}N HSQC spectrum of ^{15}N-labelled Sgt2_TPR titrated with MEEVD. A) The CSP of ^{15}N-labelled Sgt2_TPR peaks upon binding to unlabelled MEEVD at ratios: 1:0 (teal, assigned), 1:0.5 (purple), 1:1 (red), 1:2 (light pink), 1:3 (maroon), 1:4 (blue), 1:5 (cyan), and 1:6 (green). B) Detailed view of ^{15}N HSQC spectrum showing chemical shift perturbations (CSP) for the selected residues. Black and red arrows indicate first and second CSP event upon titration respectively.

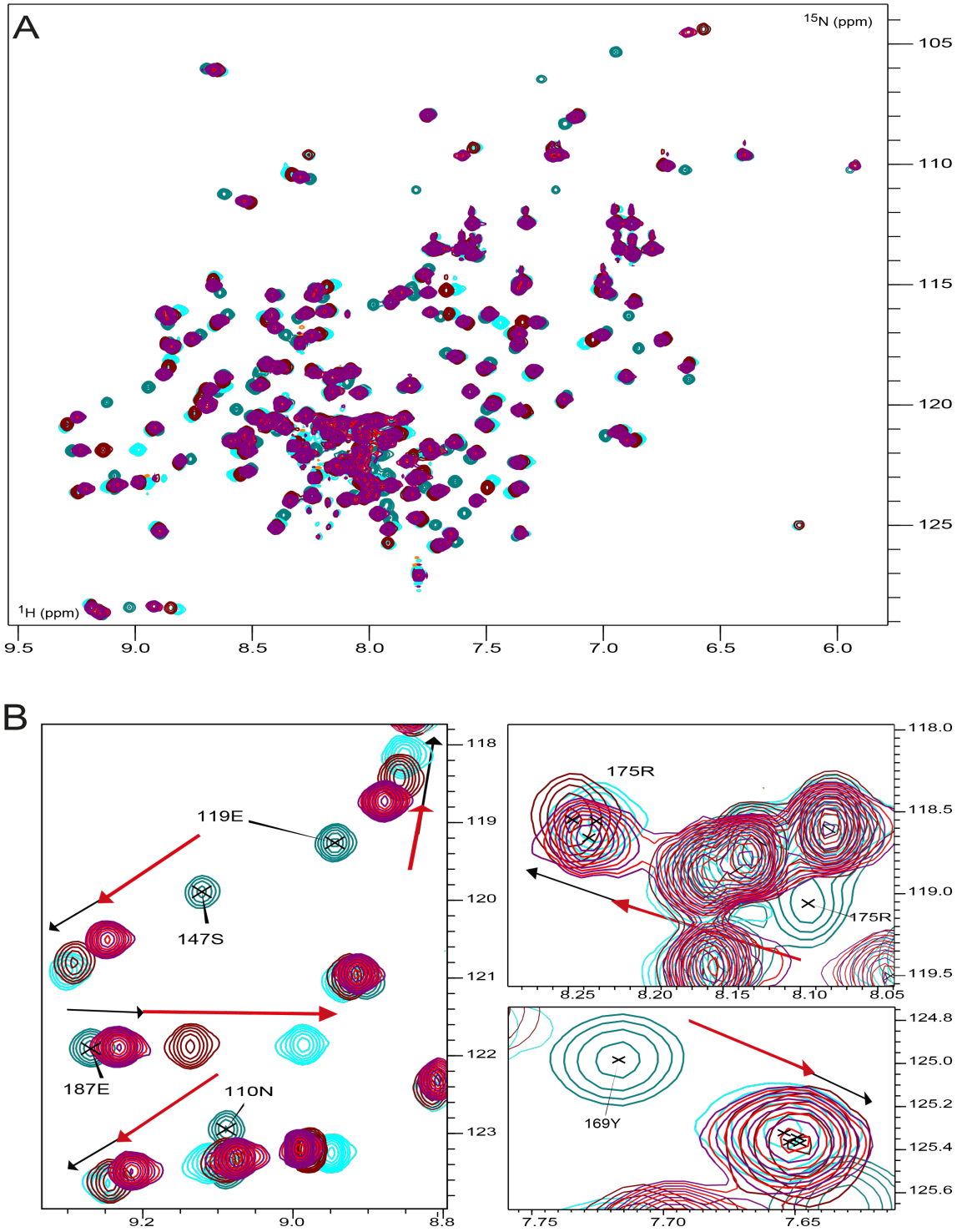

Figure 91: The ^{15}N HSQC spectrum of ^{15}N-labelled Sgt2_TPR titrated with SLEEDNLD. A) The ^{15}N HSQC spectrum of ^{15}N-labelled Sgt2_TPR titrated with unlabelled SLEEDNLD at ratios: 1:0 (teal, assigned), 1:0.5 (purple), 1:1 (red), 1:2 (light pink), 1:3 (maroon), 1:4 (blue), 1:5 (cyan). B) Detailed view of ^{15}N HSQC spectrum the showing chemical shift perturbation (CSP) for the selected residues. Black and red arrows indicate the small and big CSPs upon titration respectively.

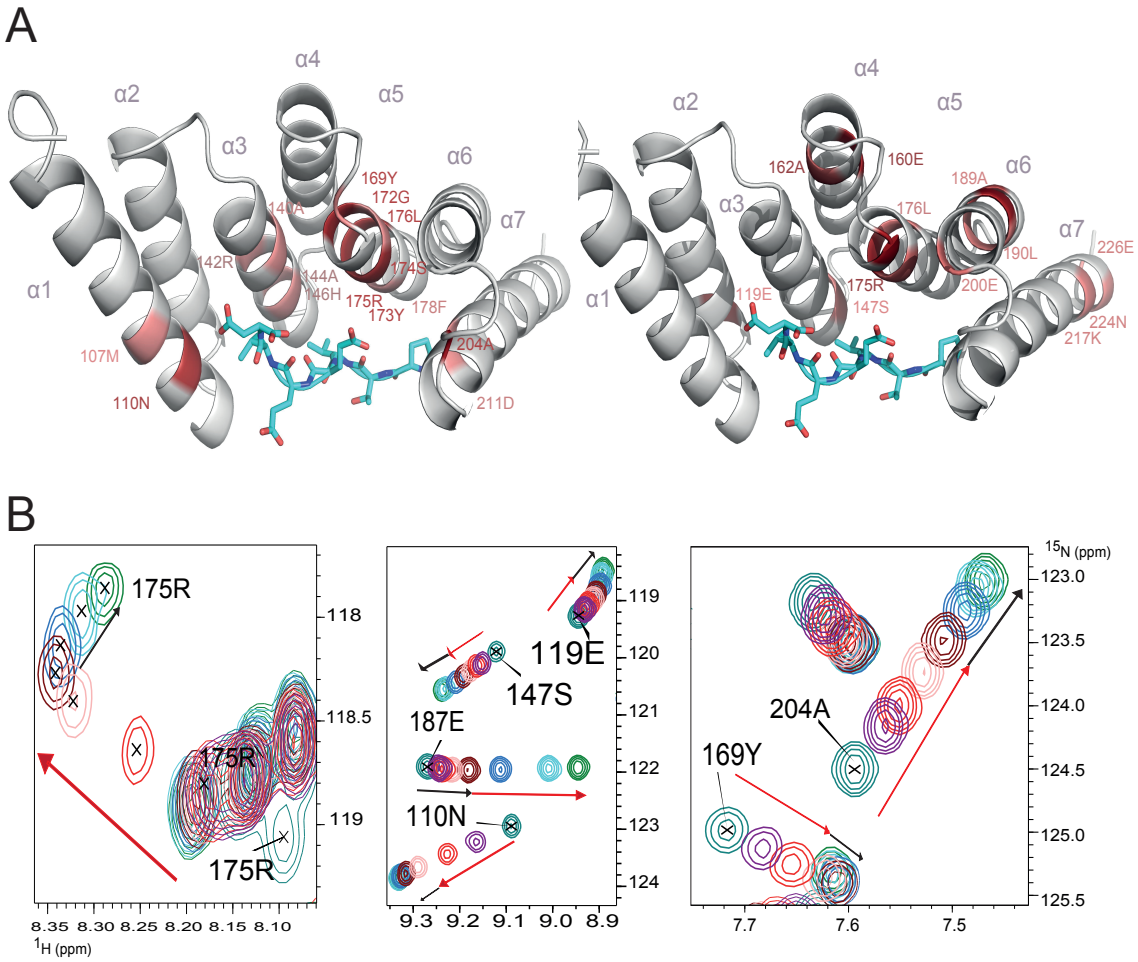

Figure 92: Sgt2_TPR binds yeast chaperones and Ybr137wp. A) Sgt2_TPR domain colored according to chemical shift perturbation (CSP) upon additions of unlabelled PTVEEVD peptide. The model on the left shows the titration points between 0 and 1:3 protein/peptide molar ratio and on the right between 1:3 and 1:6 molar ratio. The most perturbed residues were selected at > 0.4 ppm (red) and >0.3 ppm (light red) for the model on the left and >0.2 (red) and >0.15 (light red) for the model on the right. B) Small regions of the 15N HSQC spectrum of 15N-labelled Sgt2_TPR titrated with unlabelled PTVEEVD at ratios: 1:0 (teal), 1:0.5 (purple), 1:1 (red), 1:2 (light pink), 1:3 (maroon), 1:4 (blue), 1:5 (cyan), and 1:6 (green). Black and red arrows indicate first and second CSP event upon titration respectively.

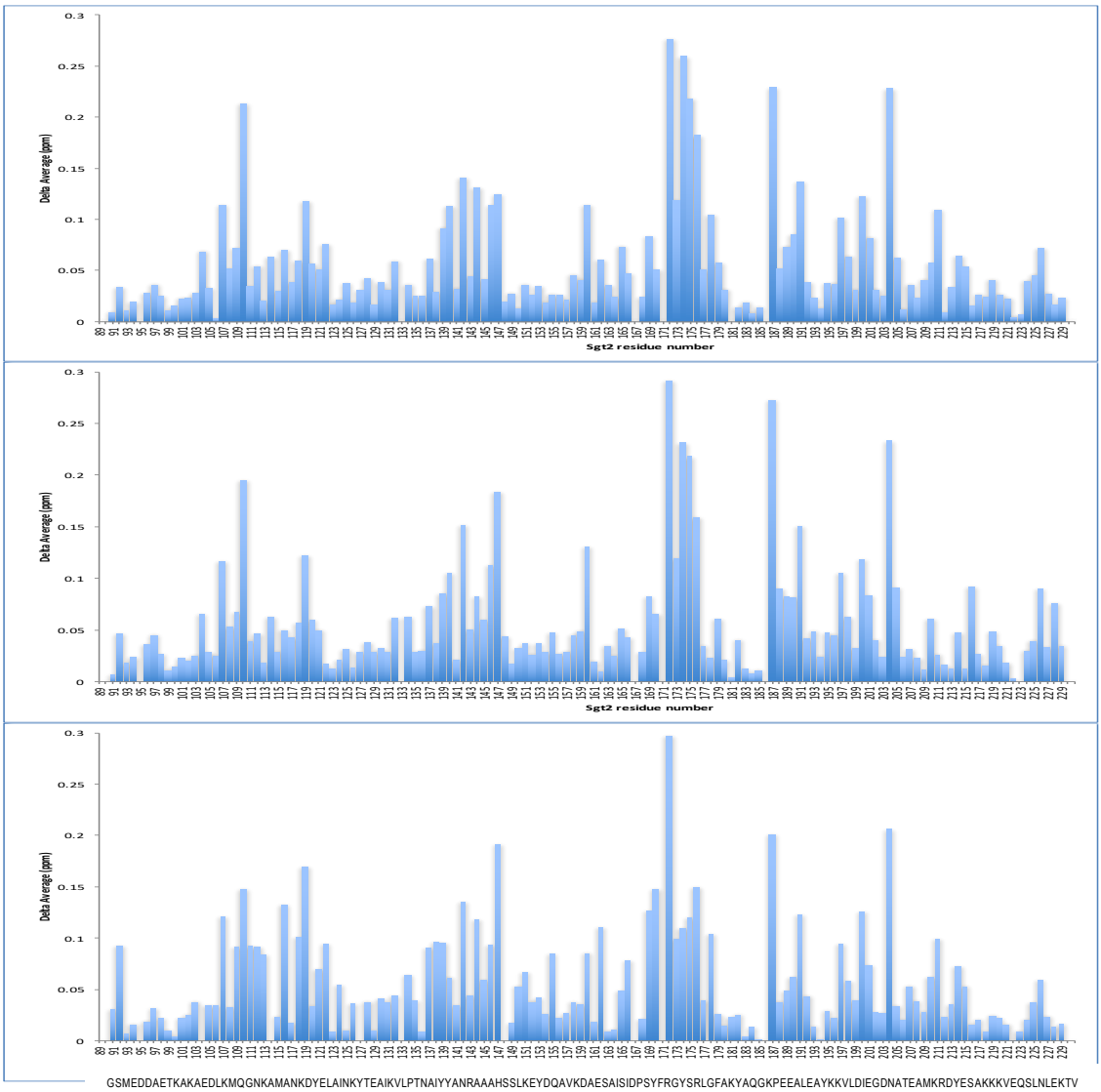

Figure 93: CSP graphs of Sgt2_TPR upon titration with peptides. A) PTVEEVD, B) MEEVD, C) Ybr137wp. The chemical shifts were normalized and calculated by using the equation described in Methods (Section 2) for the last titration point plotted for each residue.

5.6.2 ITC

In addition, we characterised the binding between Sgt2_TPR (93–229) and PTVEEVD, MEEVD and SLEENDNLD peptides by ITC (**Figure 94 C**). The ITC results indicate similar binding affinity for all three complexes with dissociation constants (Kd) of 9.00 ± 0.05 µM for Sgt2_TPR/PTVEEVD, 2.95 ± 0.30 µM for Sgt2_TPR/MEEVD and 1.53 ± 0.05 µM for Sgt2_TPR/SLEEDLNLD. Interestingly the Sgt2_TPR/PTVEEVD complex with the lowest binding affinity was the only one which crystallised. The favourable enthalpy and entropy values

obtained from ITC suggest that all complex formations were driven by the establishment of both hydrogen bonds and hydrophobic interactions in agreement with the x-ray structure of the Sgt2_TPR/PTVEEVD complex (**Figure 94 C**).

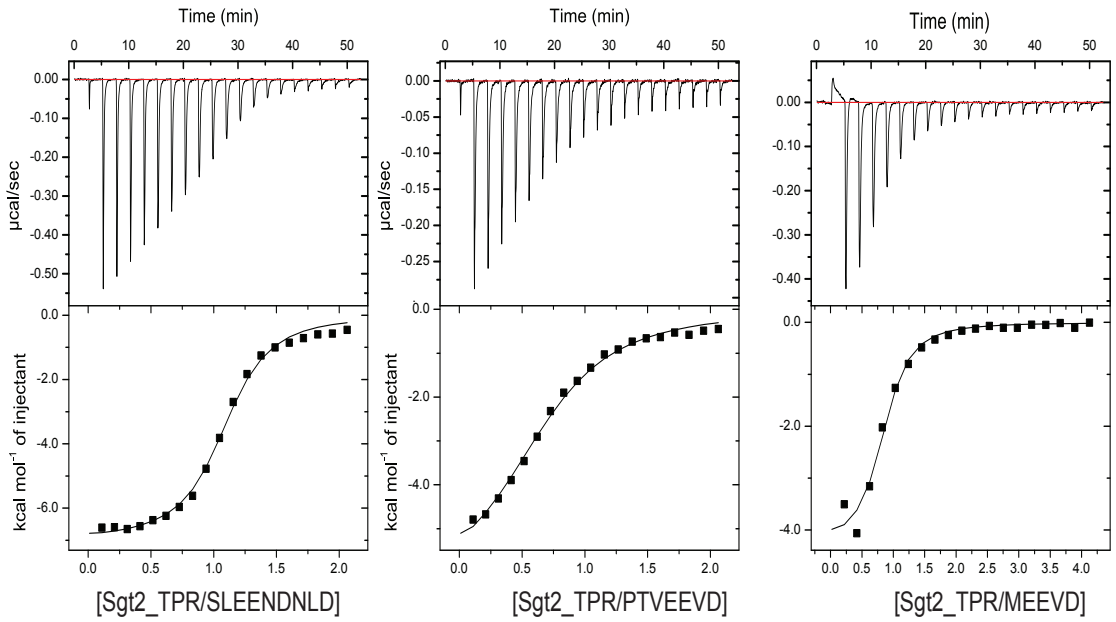

Figure 94: **ITC binding curves for Sgt2_TPR and peptides.** ITC data showing binding of Sgt2_TPR to SLEENDNLD, PTVEEVD and MEEVD. Binding parameters, determined by ITC were Kd= 1.53 ± 0.05, ΔH = − 7.073± 0.99 kcal/mol; ΔS = 3.19cal/mol/deg for SLEENDNLD; Kd= 9.01 ± 0.05 µM, ΔH = − 6.719± 0.07 kcal/mol,ΔS = 140cal/mol/deg for PTVEEVD and Kd= 2.95 ± 0.30 µM, ΔH = − 4.513± 0.03 kcal/mol; ΔS = 10.9 cal/mol/deg, for MEEVD.

5.7 Summary

Chapter 5 reports the structural and biophysical study of Sgt2_TPR and its complexes with C-terminal peptides of Ssa1, Hsc82 and Ybr137wp. The structures of the free Sgt2_TPR and in a complex with PTVEEVD were crystallised and solved. Interestingly, the zinc was essential for the crystallisation of the complex possibly acting as an additive stabilizing the crystal lattice (suggesting that metal ion is needed to stabilize the complex formation within the crystal) and crystals formed in conditions without zinc did not produced the complexes. While the presence of zinc is not biologically relevant, other metal atoms such as nickel are also present in the

homologous complex structures of Hop TPR1/GASSGPTIEEVD, Hop TPR2A/MEEVD and CHIP TPR/DDTSRMEEVD (261, 336). Attempts to co-crystallise MEEVD and SLEELDNLD bound to Sgt2_TPR failed and only the PTVEEVD formed a complex.

Furthermore, the backbone of Sgt2_TPR was assigned using NMR spectroscopy and this allowed us to study the binding with C-terminal fragments of Ssa1 (PTVEEVD), Hsc82 (Hsp90) and Ybr137wp. The structural data confirm that the binding mode of the Ssa1 to the TPR domain of Sgt2 protein is via the carboxylate clamp mechanism and we can speculate that Hsc82 and Ybr137wp interact in a similar mode due to the homology we have found in the binding mechanisms of these proteins by ITC and NMR.

Chapter 6. Discussion

6.1 SGTA, the BAG6 complex and RNF126 in a biological context

Newly synthesized hydrophobic proteins are either targeted to membranes or degraded to maintain cellular homeostasis, but the decision making processes are still not well understood. Over the years, several research studies on the mammalian and the yeast TA targeting pathway(71-73, 76, 88-90, 99, 160, 171, 255, 256) and mammalian ubiquitination (74, 75) have identified core proteins involved in a sorting mechanism for hydrophobic proteins in danger of exposure to the aqueous cytoplasm. They have also provided new mechanistic insights into the co-chaperone SGTA and its relationship with the BAG6 complex (which contains BAG6, UBL4A and TRC35) and the equivalent fungal model (which is Sgt2 and Get4/Get5). SGTA can interact with the UBL domains of both, UBL4A and BAG6, which are displayed at the C-terminal and N-terminal of the BAG6 complex. The ability of SGTA to bind to two UBL domains of the BAG6 complex and the same range of hydrophobic substrates enables the transfer of protein clients onto the BAG6 complex for sorting while determining their fate. Furthermore, a complex between SGTA and the C-terminal domain of BAG6 was shown to recognize and stabilise the transmembrane domains (TMD) of TA proteins for handover to the targeting factor TRC40, which can deliver TA proteins to the ER for insertion. The opposite end of BAG6 on the other hand (the N-terminal domain) can recruit the RNF126 E3 ligase for the ubiquitination of hydrophobic protein clients. Recent depletion-replenishment experiments by Shao *et al.* ((73)), in which endogenous proteins were replaced with recombinant variants, supported the proposed functional roles of the components of the sorting mechanism (**Figure 95**) and showed that SGTA, itself, can facilitate both targeting and ubiquitination, suggesting that TA proteins bound to SGTA are initially uncommitted to either fate. However, it is still unclear how different targets are precisely recognised and triaged between appropriate pathways. In this thesis I consolidate our latest molecular-level contributions to the field by solving the structures of the dimerisation domain of SGTA and the zinc finger motif of RNF126 within the BAG6 quality control module and structural/biophysical characterisation of RNF126 and SGTA complexes with the UBL

domains of BAG6 and UBL4A. I also studied the yeast counterpart of SGTA, Sgt2, and its interactions with molecular chaperones Ssa1, Hsc82 and Ybr137wp through the TPR domain.

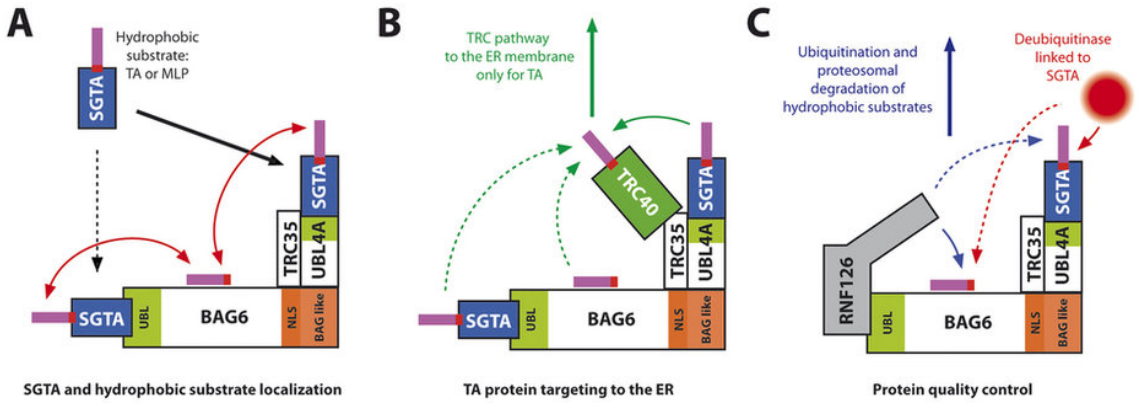

Figure 95: Schematic illustration of the roles of SGTA and the BAG6 complex in quality control pathways for hydrophobic proteins. A) SGTA with bound hydrophobic substrate reaches the BAG6 complex and binds the UBL4A_UBL domain, or the BAG6_UBL. The substrate (TA proteins and/or MLPs) remains bound to SGTA (and handover directly to heat-shock chaperones that bind the TPR) or is passed onto BAG6. B) TA substrates are transferred to TRC40 for ER membrane insertion through SGTA bound to UBL4A which binds TRC35 and recruits TRC40. C) BAG6 also participates in protein quality control of hydrophobic substrates. RNF126_NZF binds BAG6_UBL, likely replacing SGTA and, in collaboration with E2 enzymes could ubiquitinate the hydrophobic proteins from BAG6 or SGTA, targeting them for proteasomal degradation (174)

6.2 SGTA (1-69) dimer in structural context

Mammalian SGTA and its yeast equivalent, Sgt2, share 26.81% sequence identity and a three-domain structure comprised of an N-terminal dimerization domain, a central TPR domain and C-terminal glutamine-rich domain (Figure 96). The tendency of full length SGTA to aggregate in solution and the reported flexibility between the three domains, and at the C-terminal end of SGTA, present significant challenges for structural studies. Thus excised domains of SGTA and Sgt2 have been studied independently by us and others. The TPR domain from SGTA has been solved previously by X-ray crystallography (240). The solution structure of Sgt2 N-terminal dimerisation domain was determined in our laboratory (90). Various other SGTA/Sgt2 structures have also been added to the literature, namely NMR solution (263) and X-ray structures (255) of

partial constructs of the N-terminal domain of Sgt2 in the context of its complex with Get5, and the N-terminal domain of SGTA (89) comprising only the first two or three out of the four helices in each monomer as compared to the full structure presented in this thesis. These structures mostly agree with each other, and confirm the structurally unusual dimer interface, with some differences at the C-terminal end far from the UBL binding interface. Our SGTA (1-69) structure presented in this thesis (82) overlaps well with the yeast homologue with an RMSD of 2.4 Å. There are slight differences between the structures at helix α3 and the loop between helices α2 and α3 **(Figure 21 D)** which are part of the UBL interaction site. These differences may explain the variation in SGTA binding to the UBLs from BAG6 and UBL4A and are discussed later.

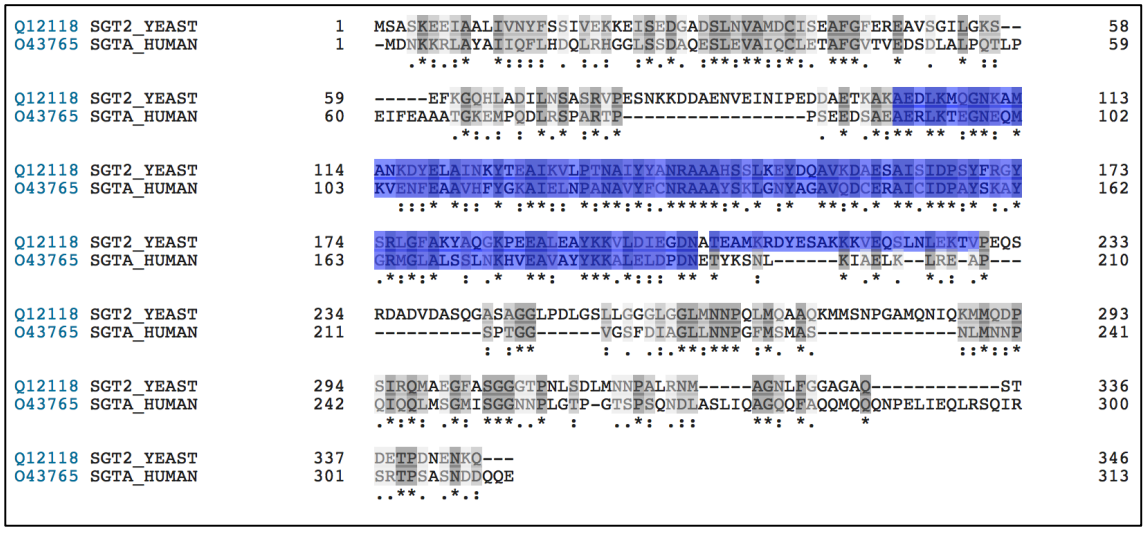

Figure 96: Sequence alignment of SGTA and Sgt2 using Clustal-Omega. Asterisks indicate sequence identity and dots indicate sequence similarity. TPR domains are shown in purple.

6.3 SGTA (1-69) binding to BAG6 and UBL4A

The SGT dimers bind UBLs. In the mammalian system, SGTA binds to the UBLs of BAG6 and UBL4A from the heterotrimeric BAG6 complex with comparable affinity. The yeast system differs in that there is no known homologue for BAG6. Sgt2 binds to the homologue of UBL4A, Get5, which is part of the heterotetrameric Get4-Get5 complex, with considerably higher affinity than the mammalian system. These apparent difference may hint at a more complex role for

SGTA. As noted above, there are small structural differences between SGTA and Sgt2. Could these differences have evolved due to the different demands upon these two proteins in the mammalian and yeast systems. The discussion that follows attempts to evaluate the structures of SGTA, Sgt2 and their complexes with the UBLs of BAG6, UBL4A and Get5 to try to give a structural explanation to the difference in affinities.

The structure of the N-terminal dimerization domain of SGTA contributes structural details into its role in TA membrane protein insertion and its interaction with the BAG6 complex in MLP quality control pathways. The analysis of our mammalian complex structures of SGTA with UBLs indicates that the orientation of the SGTA binding helices relative to the UBL partner in the SGTA complexes is slightly different and rotated by approximately 45 degrees in comparison to the yeast complex, as shown in **Figure 32 D, E, F**. As described in the Results, the binding is driven mostly by electrostatic interactions and therefore these differences in orientation contribute to some differences in the negative charge distribution of the α2 central binding helices of SGTA (1-69) and Sgt2 N-terminal and also to the differences in positive charge distribution on the surface of the UBLs. Namely, the structural equivalent of Lys85 in the Get5_UBL is Glu12 in UBL4A_UBL which cannot form the same salt bridge. Instead UBL4A_UBL Lys66 and Lys46 pull the SGTA_NT domain around slightly to satisfy the charge interactions. On SGT, the four negatively-charged amino acids are located along the helix slightly differect in Sgt2 and SGTA, with Sgt2 displaying two at each end (Asp28/Asp31 and Asp38/Asp42) and SGTA showing three at one end (Asp27/Glu30/Glu33, where the Lys85 interaction was lost, and one at the other end (Glu40). The comparison of binding interfaces between SGTA (1-69)/UBL4A_UBL and SGTA (1-69)/BAG6_UBL HADDOCK models shows similarities however, the BAG6_UBL binding site displays a slightly narrower site. The conserved lysine residues on UBL4A_UBL (6, 46, 42, 48, and 66) and their equivalents on BAG6_UBL (22, 58, 64, 80) make electrostatic contacts with the charged residues on SGTA (1-69), Asp27, Glu30, Glu33, and Glu40. The equivalent of Lys46 is missing in the BAG6_UBL and is replaced by Gln62 which cannot make the same salt bridge.

Analysis of the biophysical properties of the complexes show that one SGTA (1-69) dimer binds to one UBL domain. Our NMR titration experiments indicate that the binding affinities of SGTA

(1-69)/UBL4A_UBL and SGTA (1-69)/BAG6_UBL are comparable however it could not be precisely verified by ITC, where we obtained a molar ratio value,n, of 0.4-0.6 in most titrations for SGTA (1-69)/BAG6_UBL and approaching 1 in the SGTA (1-69)/UBL4A_UBL complex, and the binding affinity could not be derived unambiguously. An additional difference was observed in CSP analysis in which there were several peaks splitting for SGTA (1-69) during the titration with BAG6_UBL but no apparent peak splitting in SGTA (1-69)/UBL4A_UBL titrations. The peaks splitting can be explained by the presence of two chemically different local environments of the same residue on each SGTA monomer in contact with different residues in the BAG6_UBL. For example, residue Q37 on each SGTA protomer interacts with different residues of BAG6 (**Figure 97).** A similar situation occurs at the SGTA (1-69) dimer and UBL4A_UBL binding interface and additionally one Q37 on SGTA contacts the flexible C-terminal of UBL4A_UBL which could potentially explain no peak splitting of that particular residue. The lack of obvious peak splitting in the SGTA (1-69)/UBL4A_UBL interaction might suggest differences in the binding dynamics of the two UBLs where the exchange between the two symmetry-related binding sites on the SGTA (1-69) dimer is significantly slower than the NMR chemical shift timescale in the case of the BAG6_UBL but it is faster for UBL4A. In the case of SGTA (1-69)/UBL4A_UBL we do not observe peaks splits because of the potential averaging of frequencies at the elevated temperatures required for analysis (310 K). If on the other hand we were to expect two peaks at similar chemical shifts to the SGTA(1-69)/BAG6_UBL complex, but only observe one, we could speculate that the absence of the second peak could be attributed to exchange broadening caused by motions on the intermediate timescale of the flexible UBL4A_UBL C-terminus, which is near to Q37 of the second protomer.. We also consider that the binding might be in intermediate exchange for other residues and that signal broadening causes one of the peaks to disappear. Furthermore, HADDOCK models show that first two clusters are similar in both SGTA (1-69)/UBLs and the other two clusters slightly differ in the conformation of UBLs. Since BAG6 could bind in two, considerably different, conformations, of which one might not give as strong a signal in ITC due to fewer electrostatic interactions, decreasing the enthalpy which could explain the impossible molar ratio value of 0.4. In addition, analysis of the binding interface shows that there are fewer Lysine residues on BAG6_UBL in

contact with SGTA (1-69) as compared to UBL4A which may allow for more conformational flexibility in binding. **Figure 98** shows SGTA (1-69)/BAG6_UBL HADDOCK models from cluster 1 and cluster 2 highlighting the positions of splitting residues on SGTA, which might be in different environments in the two different conformations. However, more experiments including mutagenesis would be needed to investigate the differences in SGTA (1-69)/UBL complexes fully.

In the yeast system there is only one type of UBL on the Get4/Get5 complex (located at the centre of each Get5 subunit), and it interacts with Sgt2 with higher affinity. The two UBLs of the larger mammalian BAG6 complex are capable of binding to SGTA, each with comparable affinity that is lower than the K_d reported in yeast. We have also shown that the mammalian UBLs interact with SGTA in a slightly different orientation as compared to fungal equivalent. These differences may reflect the seemingly more complex role of SGTA as compared to Sgt2. Current models shows that both SGTA and Sgt2 capture free TA-proteins before handing them on to BAG6 and heterotetrameric Get4/Get5 complexes (88, 101). In the yeast system, the main function of the Get4/Get5 complex appears to be to deliver TA proteins directly onto Get3 for membrane targeting. The intermediate form of the Get4/Get5 complex was shown to rapidly rearrange to the final Get3·Get4/5 structure for the hand off of TA-proteins to Get3 for ER delivery (78, 98). The mammalian BAG6 complex, however, has the dual role of being able to triage substrates towards distinct targeting or degradation. Hence, we speculated that one of the factors that impacts the fate of SGTA-associated hydrophobic substrates is which one of the two BAG6 complex UBLs first interacts with SGTA prior to substrate release. Thus, one of the UBLs might favor substrate handover on to TRC40 for ER targeting, as seen for TA-proteins (71, 72), whilst the other UBL may promote the ubiquitination of aberrant substrates including MLPs (74) and ERAD substrates in the retrotranslocation process out of the ER membrane (171). A recent study supports our hypothesis. Using depletion-replenishment experiments by replacing endogenous proteins with recombinant variants and photo-cross linking assays, it shows that TA-proteins are captured favorably by SGTA (as compared to direct substrate to Trc40 or BAG6 complex formation). The

TA proteins are then rapidly and preferentially transferred on to TRC40 through the interaction with the C-terminal end of the BAG6 complex (a module comprised of BAG6-C terminal/Ubl4A/Trc35) for ER targeting. The substrate transfer is twice as fast as the rate of spontaneous TA protein dissociation from SGTA. In contrast, client ubiquitination is dictated predominantly by relatively slower substrate dissociation from SGTA followed by a free TA protein intermediate which is captured by the N-terminal of the BAG6 subunit (containing the UBL domain) (73). In cells, the affinity of SGTA towards BAG6 is lower than to UBL4A (73) and the process is likely to be even more complex, since the heterotrimeric BAG6 complex, whose architecture and stoichiometry have been debated, has been shown to form homo-oligomers (176) that would present multiple UBL domains to which SGTA and/or other interacting partners could bind. Meanwhile, the polyubiquitination of UBL4A regulates BAG6 activity via targeted proteolysis (182). It is possible that TRC35 and UBL4A exist as dimers, like their yeast homologues, although UBL4A lacks the N-terminal domain of Get5 which forms the main contact with Get4. Possibly, a part of BAG6 can bring together the UBL4A and TRC35 subunits in the mammalian system. Moreover, it appears that only a small C-terminal fragment of the BAG6 protein that connects the UBL4A and TRC35 is necessary to allow the transfer of TAs from SGTA to the TRC40 delivery complex (73, 76).

Results from our lab and others highlight the structural complexity of the relationship between SGTA and the BAG6 complex and indicate that further structural and functional work is required to completely understand the sorting mechanism. The structure of the full-length heterotrimeric BAG6 complex has yet to be solved and its large size, dynamic behavior and transient interactions are challenging to the structural biologist. Structures of several isolated domains from various species provided some detail on its function at a molecular level. However, the complete structure of the BAG6 protein itself has not yet been solved and there is still much unknown.

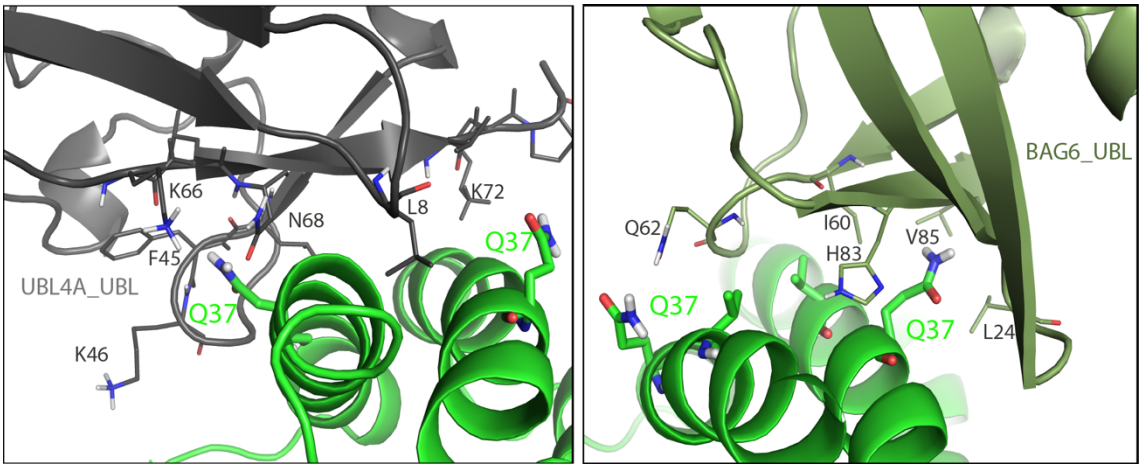

Figure 97: Detail view of the differences in a chemical environment of Q37 residue on SGTA, at the SGTA (1-69)/UBLs binding interfaces, which splits in CSP experiments. Q37 on each SGTA protomer interacts with different residues of UBLs. Additionally, one Q37 on SGTA contacts the flexible C-terminal of UBL4A_UBL which could potentially explain no peak splitting of that particular residue.

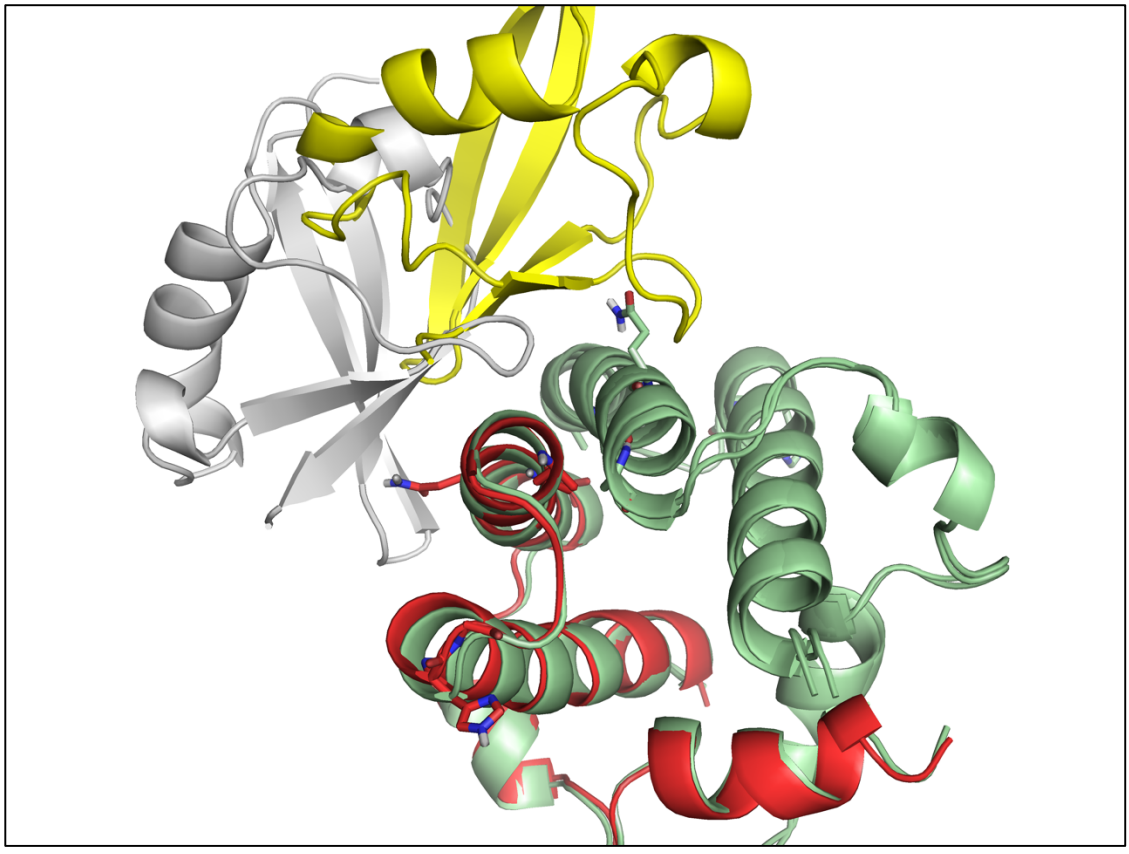

Figure 98: SGTA (1-69)/BAG6_UBL HADDOCK models from cluster 1 and cluster 2 showing splitting residues on SGTA.

6.4 Structure and interactions of RNF126_NZF and the BAG6 complex

The RNF126 E3 ubiquitin ligase belongs to the large RING family of ligases, whose significant feature is a zinc-binding RING finger domain which plays an important role in the ubiquitination pathway by interaction with an E2 (ubiquitin-conjugating) enzyme. There are hundreds of E3 ubiquitin ligases in mammalian cells able to recognize the wide variety of protein degrons (337) but the specific substrate selection mechanisms are unclear. Many of the RING E3 ligases contain an additional zinc-finger motif which can act as a recognition site for ubiquitination clients. RNF126 can bind to the BAG6, and ubiquitinate hydrophobic substrates to mark them for degradation at the proteasome (75). In this work we characterise the interactions between RNF126 and BAG6 at a molecular level and report an additional, lower-affinity RNF126 interaction with UBL4A, another subunit of the BAG6 complex.

We solved the first high-resolution structure of N-terminal RNF126 zinc finger which comprises a three-stranded short antiparallel β-sheet with four cysteine residues coordinating zinc and one solvent exposed side of the β-sheet, which creates a platform for protein-protein interaction. The structure of the RNF126_NZF displays a similar fold to at least two other zinc finger motifs namely, the zinc finger domain of human Pirh2, an E3 ubiquitin ligase (with 27% sequence similarity and RMSD of 2.75 Å) and the Npl4 Zinc-Finger, an Ubiquitin recognition domain (with 26% sequence similarity and RMSD of 2.35 Å). To date, full-length RNF126 has not been structurally characterised, although a solution structure of the mouse RING domain has been solved (PDB Accession Number: 2ECT with no associated publication). The structure solution of the full length RNF126 presents some challenges because of the predicted, long flexible region linking the N-terminal zinc finger with the C-terminal part of the protein and a hydrophobic region towards the C-terminal. Our attempts to express and purify the full length RNF126 and the hydrophobic constructs were not fully successful due to protein instability and more work is necessary to optimise protein production.

In this study we have also identified the minimal domain of N-terminal RNF126 which is sufficient for the interaction with BAG6_UBL. We show that RNF126, although capable of binding to two different UBL domains on the BAG6 complex, has a clear preference i.e. a stronger affinity for the BAG6_UBL. We have compared the structure of RNF126_NZF/BAG6_UBL complex solved by NMR with a HADDOCK model of RNF126_NZF/UBL4A_UBL which we produced to gain some insight into the differences in binding affinities (**Figure 30 C, D**). Although the binding interfaces between RNF126_NZF and the two UBLs is conserved (**Figure 31 A),** the binding affinity between RNF126_NZF and UBL4A_UBL is significantly lower. We observed some differences in the residues on the surface of the UBLs which could account for this difference in affinity **(Figure 31 C, D).** The conformation of the loop between the β1 and β2 strands in UBL4A_UBL is different from that of the BAG6_UBL and the Leu8 sidechain points away from the RNF126_NZF hydrophobic pocket which is formed mainly by Phe36. Furthermore, the backbone NH group of the UBL4A Leu8 is solvent-exposed and lines this cavity. On the other hand, the equivalent backbone NH group of Leu24 in the BAG6_UBL is instead buried within the hydrophobic core. Additionally, the sidechain NH group of Asn68 in the UBL4A_UBL also lines this binding cavity whereas it is the aliphatic part of the His83 sidechain that occupies the binding pocket in the BAG6_UBL (**Figure 31 C, D).** These differences point to a reduced hydrophobicity of the RNF126 Phe36 hydrophobic pocket in the UBL4A_UBL relative to the BAG6_UBL, which likely affects the binding affinity. An additional decrease in affinity may be caused by the lower capacity of Lys48 in UBL4A_UBL to form hydrogen bonds with Glu38 in RNF_NZF as compared to Arg64 in BAG6_UBL. Moreover, the contact between Gln62 in BAG6_UBL and Tyr27 in RNF126_NZF and equivalent Lys53 in UBL4A is reduced. (**Figure 30 C, D**). Taken together, this study illustrates structural differences at the binding interfaces of the two complex structures which may explain a discrepancy in binding affinities, although data are limited to comparison between a higher resolution NMR structure of RNF126_NZF/BAG6_UBL and a HADDOCK model of RNF126_NZF/UBL4A_UBL derived from solved component structures and chemical shift perturbation data.

In chapter 3 we showed that, just like RNF126, cochaperone SGTA can also bind to the UBLs of both BAG6 and UBL4A, in the same binding site but with slightly higher affinity to UBL4A than to BAG6. The SGTA binding interface for UBLs differs from that of RNF126 and exhibits more electrostatic interactions, which are stronger for UBL4A_UBL. We have also shown that RNF126_NZF does not directly bind to SGTA (1-69), which supports the model in which RNF126 binds to BAG6 to get access to its protein clients. Furthermore, we analysed whether SGTA and RNF126 bind to BAG6 _UBL in a competitive manner or if they can bind at the same time. The results show that RNF126_NZF competes with SGTA (1-86) for the same binding site on BAG6_UBL (and on UBL4A_UBL) and cannot bind simultaneously. In addition, the BAG6_UBL binds more tightly to RNF126_NZF than it does to SGTA (1-86). The subtle variations in binding affinities amongst the components of the sorting mechanism might be key in understanding the details pertaining the decision-making processes at the BAG6 complex (**Figure 95**).

Since SGTA can initially capture tail-anchored proteins emerging into the cytosol and it seems to preferentially bind to the UBL4A_UBL subunit of the BAG6 complex (BAG6-UBL4A-TRC35), both in vitro and inside cells (73, 87), the question is how the hydrophobic substrates that are destined for RNF126 mediated ubiquitination are transferred on to the BAG6 subunit? It seems that the fate of hydrophobic clients depends on the relative abilities of SGTA and the BAG6 complex components to capture them and also their ability to retain the bound substrate (73, 86, 99). Given that the transfer of TA proteins from SGTA to TRC40 for ER delivery (via UBL4A) is faster, and TA protein transfer from SGTA to the BAG6 subunit is slower (73), the proteins which fail the initial capture by TRC40 can be then collected by the BAG6 subunit and targeted for quality control (**Figure 95**). However, TA proteins seem to be able to spontaneously associate and dissociate from SGTA more quickly than they are caught by BAG6, allowing for multiple cycles of triage. In contrast, TA proteins once bound to BAG6 subunit, dissociate from BAG6 relatively slowly to allow for quality control (73, 106). The decision making process, and why TRC40 fails to capture hydrophobic substrates independently, can also depend on the type of the

substrate and the biophysical features of the TRC40 binding site, which could have evolved to favor ER-destined TA proteins (78, 256). In the case of TA proteins, their well-defined helical TMDs ensure that they are quickly transferred from SGTA through a UBL4A mediated interaction with the BAG6 complex to TRC40 for ER delivery. In contrast to this, the failure of substrates such as mislocalised proteins with large hydrophobic patches to be efficiently passed from SGTA on to TRC40 would instead target them towards the BAG6 subunit for an alternative quality control pathway. Alternatively, MLPs may not be released to TRC40 if SGTA, with substrate bound, first associates with UBL4A subunit if SGTA can somehow detect unfolded or non-native hydrophobic regions. The BAG6 subunit is capable of distinguishing between different substrates as moderately hydrophobic polypeptides are reported to bind close to its N-terminal UBL domain (133) whereas membrane proteins and heat denatured luciferase bind its central proline rich region (19, 176). As for MLPs, we suggest that competition between SGTA and RNF126 for binding to the BAG6 subunit effects the fate of these substrates since increased residence time of SGTA on BAG6 would reduce the access of RNF126 to its BAG6 associated substrates, thus reducing their ubiquitination and degradation.

Our structural and biophysical study of RNF126 and BAG6 expands our understanding of the way in which hydrophobic proteins are triaged between diverse quality control pathways. Furthermore, there is a growing interests in using E3 ubiquitin ligases as drug targets and prognostic biomarkers since, the client specificity of E3 ligases is proposed to limit the potential side-effects of drugs that target them, thus the new RNF126 structure presented here provides a potential starting point for developing a targeted anti-cancer therapy {Chasapis, 2009 #24553;Zhi, 2013 #20150.

6.5 Structure and interactions of the Sgt2_TPR with yeast chaperones and Ybr137wp

The roles of SGTA and Sgt2 have long been linked to Hsp70 and Hsp90 chaperones {Liou, 2005 #10888}. SGTs can, via their central TPR domain, directly bind to several heat-shock proteins which help them to facilitate correct folding of client proteins (88, 275). It is possible that these chaperones might funnel substrates towards SGTA and enable their access to additional quality control pathways. It has also been proposed that SGTA can regulate the ATPase activity and folding rates of Hsp70 (242). However, the full significance of the relationship between the SGTA/Sgt2 and chaperones, remains to be discovered. Sgt2 was also reported to interact with Ybr137wp, a poorly-characterised protein with its function linked to the Get pathway and TA targeting. However, the exact role of Ybr137wp in the Get pathway is unclear.

In this study we solved the first high-resolution structure of the free Sgt2_TPR domain and the same domain in complex with the last seven amino acids of Ssa1 (Hsp70) using crystallography. We have also assigned the backbone of the Sgt2_TPR using NMR spectroscopy, and characterized the interaction between Sgt2_TPR and the extreme C-terminal fragments of Hsc82 (Hsp90) and Ybr137wp. Analysis of the complex structure shows that Sgt2_TPR binds Ssa1 in well-established carboxylate clamp mode that is common in interactions between heat shock proteins and chaperones (338-340). Therefore, we speculate that Sgt2_TPR may be interacting with Hsc82 and Ybr1372p in a similar binding mechanisms due to their similar behavior as analysed by NMR and ITC.

The structure of Sgt2_TPR complex with Ssa1 indicates at a subtle conformational change spread over the entire structure, which appears to be more compacted upon binding a peptide (**Figure 82 B**). The Sgt2_TPR residues affected by the addition of the MEEVD and SLEENDNLD peptides (**Figure 88 and 90 respectively**) were the same as those perturbed by PTVEEVD (**Figure 88**), with the intriguing double event binding mode in all titrations. Several peaks in the HSQC spectra of the binding experiments shift in a non-linear manner suggesting the presence of an additional

form to completely free and bound states like the ones previously reported in other protein interaction studies (341). This could explain the discrepancies which were observed for the SGTA_TPR binding to the C-terminal peptide from the ubiquitin receptor of the 26S proteasome, Rpn13 (168). The Kd measured by ITC indicated tight binding but the fast exchange regimen in the NMR titrations suggests a medium to low affinity interaction (168). Since the most perturbed residues in the case of the first Sgt2_TPR and peptides binding "event" (using 1:3 protein to peptide molar ratio) correspond to the residues at the binding interface in the x-ray complex structure, and/or the neighboring residues, and we speculate that the residues perturbed the most in the second event may be due to the global TPR conformational change. Similar results were obtained in the interaction of Rpn13 and SGTA_TPR, in which many NMR signals corresponding to amino acid residues widespread along the TPR were affected upon titration and were not localised in a well-defined interface (168). The binding affinity values we obtained from the ITC reactions are comparable to examples in the literature of TPR domains binding HSP-derived peptides such as TPR domains interacting with carboxylate clamps from Hsp70 and Hsp90 chaperones at micromolar scale (261, 271, 342). The orientation of the Hsp peptides at the Sgt2_TPR groove slightly varies between homologues. Although, the conserved carboxylate clamp mode of interaction is consistent, there are some differences observed at the N-terminal end of the peptides, which is unsurprising given that the carboxylate clamp can act as an anchor at the interface. The elongated conformation of the PTVEEVD peptide within the Sgt2_TPR groove in the structure of the complex presented here is similar to that of the HOP TPR1/GPTIEEVD complex (PDB ID 1elw) (261). In contrast, in the two structures of GPTIEEVD/Chip TPR (PDB ID 3q49; (171)and PDB ID 4kbq; (336)) the peptide lines the groove in a curved conformation and also the orientations of the N-terminal proline varies slightly within the binding site **Figure 99**. In addition, in our Sgt2_TPR/PTVEEVD complex structure, we can observe some differences between the Pro1, Thr2 and Val3 sidechains in the two chains (A and B) indicating possible flexibility in the association between TPR and Hsp peptides at the same interface. Sgt2_TPR serves as a binding interface for transient interactions with many proteins including chaperones. Furthermore, the preceding residues to EEVD are also important for the binding affinity and it has been reported that trimming the peptide sequence to EEVD only,

significantly reduced the affinity (261). Since Sgt2 exists as a homodimer it is possible that it can target a broad range of substrates by binding more than one protein simultaneously, either one on each of its two TPR domains or other combinations of domains and bring them into closer proximity promoting interactions.

The role of Ybr137wp in the Get pathway is yet to be established. Our and others' results show that it binds to Sgt2 with the same binding interface as heat shock chaperones (300). Previous ITC experiments showed that one full-length Ybr137w decamer is capable of binding to five Sgt2_TPR dimers with a dissociation constant (Kd) of 1.38 ± 0.09 μM (300). This is almost identical to the ITC results we obtained for the association of Sgt2_TPR with the extreme C-terminal nine- residue Ybr137wp-derived peptide (Kd of 1.53 ± 0.05), suggesting that the SLEEDLNLD fragment is sufficient for the interaction. Moreover, it has been shown that removing ESLEEDLNLD from the C-terminal of Ybr137wp abolished the interaction, confirming that this flexible C-terminal region is also necessary for the interaction (300).

Further work is required to define the exact role of ybr137wp in ER delivery of tail-anchored membrane proteins. To date it is known that the function of Ybr137 is associated with altering the failure of TA protein delivery and cell viability derived by the impaired GET system in cell starvation conditions but not under normal conditions (300). However, it is yet not known whether Ybr137 is linked to any other alternative TA targeting pathways the signal recognition particle (SRP), SRP-independent targeting (SND) proteins or Hsc70 family of chaperones. Thus it has been shown that a deletion of *ybr137w* from the yeast strain promotes the association of heat-shock chaperones Hsp104 or Hsp70 with Sgt2 (88) suggesting that Ybr137wp and the chaperones compete for the same binding site on the Sgt2 TPR domain. Interestingly, Ybr137wp can also form a complex with Kap142p in the cytosol (301). Kap142p is a karyopherin that mediates nuclear export and import (301). However, the role of Ybr137wp in this interaction with a karyopherin, just like in the Get pathway is yet to be discovered.

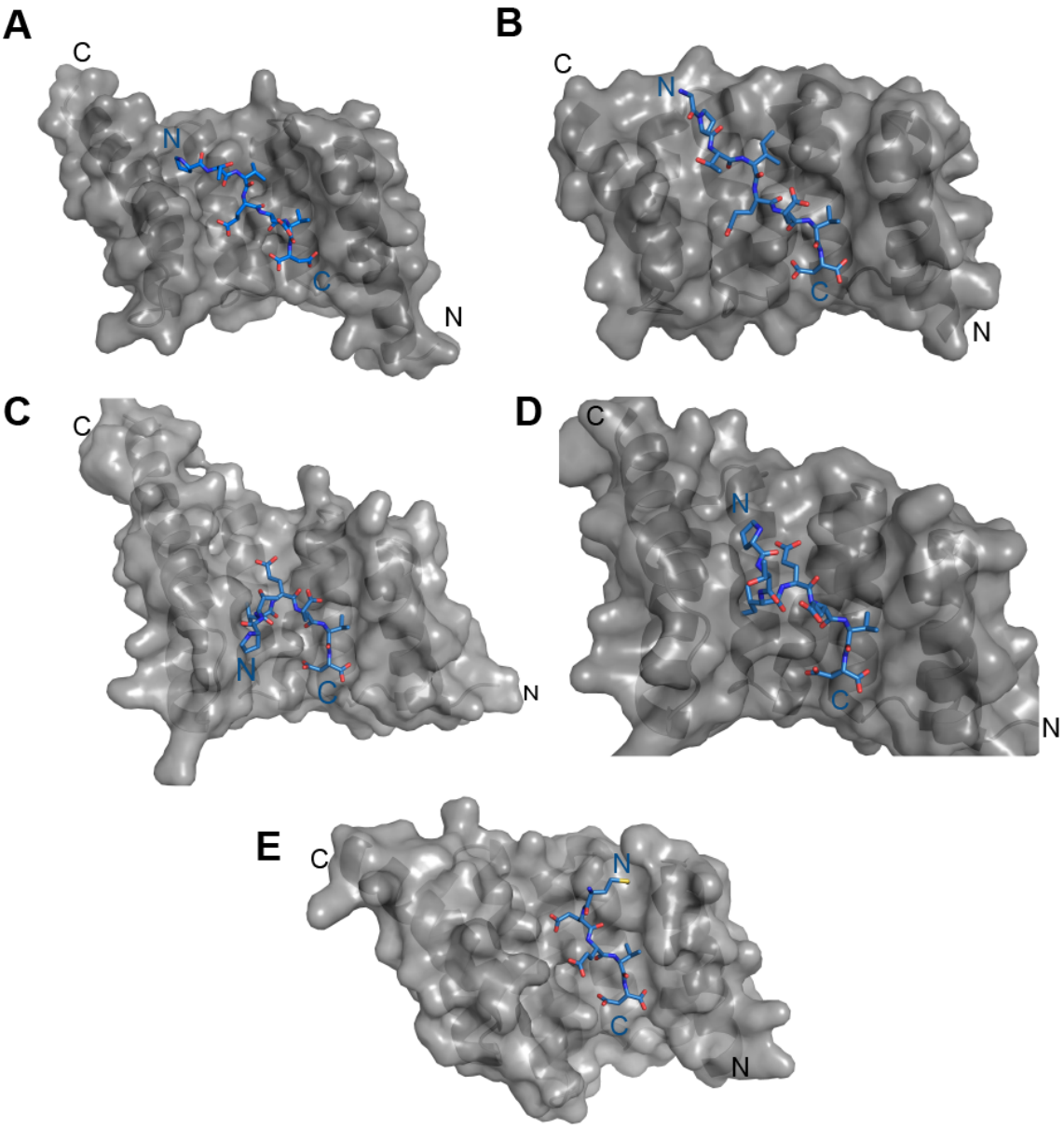

Figure 99: Different conformations of peptide bound to several TPR domains. A) our Sgt2_TPR/PTVEEVD, B) HOP TPR1 /GPTIEEVD (PDB ID: 1elw), C) CHIP TPR/GPTIEEVD (PDB ID: 3q49), D) CHIP TPR/GPTIEEVD (PDB ID: 4kbq) and E) HOP TPR2A/MEEVD (PDB ID: 1elr). Elongated peptide conformations are present in A) and B) and curled conformation are shown in C), D) and E). TPR domain shown as surface representation. The bound peptides are shown in a sticks.

6.6 Overall conclusions and future goals

The key players of protein quality control are chaperone networks and degradation machineries. The failure of sorting mechanisms which maintain homeostasis between protein biosynthesis and disposal of damaged proteins is associated with a broad spectrum of human diseases like cancer or neurodegenerative disorders, hence making them important targets to study. The past years showed that the triage system comprising co-chaperone SGTA, the BAG6 complex and E3 ubiquitin ligase RNF126 is vital in cellular quality control of MLPs, TA protein targeting and also regulation of hormone receptor signaling and viral lifecycles. However, mechanistic details concerning most of these roles are still not well understood and are often limited to characterization of excised subunits from the triage machinery. Moreover, it is still unclear how the particular components of the sorting system recognize distinct types of hydrophobic substrates in order to direct them along appropriate routes like the TRC targeting pathway or proteasomal degradation.

In this work, we combine biophysical and structural approaches to provide insights into the functions of SGTA and RNF126, as part of the BAG6 quality control module. We achieved this by solving the structures of the dimerisation domain of SGTA (Chapter 3) and the zinc finger motif of RNF126 (Chapter 4). We also characterise the interaction of both RNF126 and SGTA with the UBL domains of BAG6 and UBL4A. In addition, we generate structural models of complexes of RNF126 and SGTA with UBLs using NMR spectroscopy and HADDOCK (Chapters 3 and 4). We also focus on an additional role of Sgt2, which is the interaction with molecular chaperones (Hsp70 and Hsp90) and also with Ybr137wp, through its TPR domain. Chapter 5 describes the x-ray structures of Sgt2_TPR and its complex with the extreme C-terminus of Ssa1. In addition, we demonstrate that Sgt2_TPR interacts with C-terminal fragments of Ssa1, Hsc82 and Ybr137wp with a similar binding mode and with comparable binding affinities. Together, this work aims to shed more light on quality control mechanism in yeast and mammals, by describing the relationships between, and providing useful molecular details of, the components of the SGTA/BAG6 complex triage mechanism.

Future work will focus on studying the full-length components of SGTA/BAG6 triage system and even larger complex structures which will add a great value to understanding the sorting mechanism. To date, the full length structures of the BAG6 protein itself and of SGTA have not been solved because of large size and possible oligomerisation in the case of BAG6, and hydrophobicity and lack of structure for the C-terminal domain of SGTA. Nevertheless, it is the structural and functional characterisation of its C-terminus which will increase progress in understanding the function of SGTA, and in particular, of how it binds hydrophobic substrates. This will help to clarify how specific hydrophobic substrates are recognized and triaged by SGTA between the subunits of the sorting system. Furthermore, it will be interesting to determine the structure of full-length SGTA to understand its interactions with all the other binding partners at the macromolecular level especially since SGTA is reported to play an important role in oncogenesis (e.g.of breast cancer, esophageal squamous cell carcinoma and human hepatocellular carcinoma) and is hence suggested to be a potential prognostic marker and therapeutic target (173).

We will also continue with studying full-length RNF126, which so far has been a challenge due to protein instability. The two different constructs we designed and tested in our lab for their expression and purification, showed signs of major degradation, especially after tag removal (as described in Chapter 4). Although, production of a stable construct suitable for further structural studies has not yet been achieved, a number of techniques will be employed with an aim to aid the construction of a more suitable sample for NMR and biophysical studies. Future experiments will involve using RNF126 with Thioredoxin tag to prevent protein degradation. We will also employ Differential Scanning Fluorimetry (DSF) which will give information on the temperature at which a protein unfolds in a particular buffer, salt concentration, detergents and additives. This will enable the screening of many different buffers, which in turn may help to find a better condition that promotes protein stability, and also to overcome the aggregation problem. Other homologous proteins will be also considered. A different E3 ligase RNF115 shares near identical NZF and RING domains with RNF126 (with 42.3% overall sequence identity and 52.6%

similarity) and may also be capable of binding to BAG6 thereby, perhaps, conferring some redundancy on the system. In fact, inhibiting the interaction between RNF126 and BAG6 had only a partial effect on ubiquitination levels in cells (75) which could be a result of RNF115 activity. Therefore, RNF115 will be tested to establish whether this protein is more stable. Moreover, we may consider using eukaryotic organisms as an alternative expression system such as yeast or baculovirus infected insect cells which would allow for post-translational modifications to occur if necessary. It would be also interesting to investigate which (if any) part of RNF126 can bind hydrophobic substrates. Since, there were no direct RNF126 substrates identified before except for the recently reported interaction with Frataxin, we aimed to test whether the two proteins interact directly in solution. However, it is not clear in the literature if it was a full length frataxin precursor and also which domain of RNF126 is necessary for the interaction. The full-length RNF126 could not be produced in sufficient amounts, therefore RNF126_NZF was tested in interaction experiments. ^{1}H- ^{15}N HSQC spectrum displayed some signal broadening upon RNF126_NZF titration into ^{15}N-labelled Frataxin and the vice-versa titration shows similar results. However, more experiments are needed to verify whether there is an interaction between the two proteins and if it is the N-terminal of RNF126 which interacts.

The TPR domains of SGTA and Sgt2 act as a platform allowing for interactions with many different proteins and they physically link the GET pathway to further chaperone pathways. Characterisation of complex structure and interaction of the TPR domain of SGTA with the C-terminus of Ssa1 and comparison to binding with peptides of Hsc82 and Ybr137wp was just a small step towards understanding how these transiently formed interactions maintain the substrate specificity and target recognition. Moreover, further work defining the exact function of Ybr137wp in ER delivery of tail-anchored membrane proteins will clarify whether there is any link between this protein and the recently discovered alternative targeting pathway in yeast.

Overall, it is crucial to further our understanding of the various quality control networks in which the multichaperone triage system composed of the BAG6 complex, SGTA and RNF126 plays a key role, to describe how these pathways are linked in multiple disease states.

References

(1) Finka, A., and Goloubinoff, P. (2013). Proteomic data from human cell cultures refine mechanisms of chaperone-mediated protein homeostasis. *Cell Stress Chaperones* **18**(5), 591-605.

(2) Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A.*, et al.* (2015). Proteomics. Tissue-based map of the human proteome. *Science* **347**(6220), 1260419.

(3) Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.* **70**, 603-47.

(4) Hartl, F.U. (2011). Chaperone-assisted protein folding: the path to discovery from a personal perspective. *Nat. Med.* **17**(10), 1206-10.

(5) Hartl, F.U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**(6), 574-81.

(6) Buchberger, A., Bukau, B., and Sommer, T. (2010). Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell* **40**(2), 238-52.

(7) Fang, N.N., Ng, A.H., Measday, V., and Mayor, T. (2011). Hul5 HECT ubiquitin ligase plays a major role in the ubiquitylation and turnover of cytosolic misfolded proteins. *Nat. Cell Biol.* **13**(11), 1344-52.

(8) Rodrigo-Brenni, M.C., and Hegde, R.S. (2012). Design principles of protein biosynthesis-coupled quality control. *Dev. Cell* **23**(5), 896-907.

(9) Netzer, W.J., and Hartl, F.U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**(6640), 343-9.

(10) Gloge, F., Becker, A.H., Kramer, G., and Bukau, B. (2014). Co-translational mechanisms of protein maturation. *Curr. Opin. Struct. Biol.* **24**, 24-33.

(11) Kampinga, H.H., and Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* **11**(8), 579-92.

(12) Park, S.H., Bolender, N., Eisele, F., Kostova, Z., Takeuchi, J., Coffino, P., and Wolf, D.H. (2007). The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin-proteasome system. *Mol. Biol. Cell* **18**(1), 153-65.

(13) Mayer, M.P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **62**(6), 670-84.

(14) Rodriguez, F., Arsene-Ploetze, F., Rist, W., Rudiger, S., Schneider-Mergener, J., Mayer, M.P., and Bukau, B. (2008). Molecular basis for regulation of the heat shock transcription factor sigma32 by the DnaK and DnaJ chaperones. *Mol. Cell* **32**(3), 347-58.

(15) Diamant, S., Ben-Zvi, A.P., Bukau, B., and Goloubinoff, P. (2000). Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. *J. Biol. Chem.* **275**(28), 21107-13.

(16) Goloubinoff, P.D.L.R.P. (2016). Hsp70 chaperones use ATP to remodel native protein oligomers and stable aggregates by entropic pulling. *Nat. Struct. Mol. Biol.*

(17) Shao, S., and Hegde, R.S. (2011). Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell. Dev. Biol.* **27**, 25-56.

(18) Ast, T., and Schuldiner, M. (2011). Protein degradation: BAGging up the trash. *Curr. Biol.* **21**(18), R692-5.

(19) Leznicki, P., Roebuck, Q.P., Wunderley, L., Clancy, A., Krysztofinska, E.M., Isaacson, R.L., Warwicker, J., Schwappach, B., and High, S. (2013). The association of BAG6 with SGTA and tail-anchored proteins. *PLoS One* **8**(3), e59590.

(20) Kutay, U., Ahnert-Hilger, G., Hartmann, E., Wiedenmann, B., and Rapoport, T.A. (1995). Transport route for synaptobrevin via a novel pathway of insertion into the endoplasmic reticulum membrane. *EMBO J.* **14**(2), 217-23.

(21) Kutay, U., Hartmann, E., and Rapoport, T.A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol.* **3**(3), 72-5.

(22) Wild, K., Halic, M., Sinning, I., and Beckmann, R. (2004). SRP meets the ribosome. *Nat. Struct. Mol. Biol.* **11**(11), 1049-53.

(23) Wild, K., Rosendal, K.R., and Sinning, I. (2004). A structural step into the SRP cycle. *Mol. Microbiol.* **53**(2), 357-63.

(24) Cross, B. (2009). Eeyarestatin I inhibits Sec61-mediated protein translocation at the endoplasmic reticulum. *J. Cell Sci.*

(25) Halic, M., Blau, M., Becker, T., Mielke, T., Pool, M.R., Wild, K., Sinning, I., and Beckmann, R. (2006). Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* **444**(7118), 507-11.
(26) Janda, C.Y., Li, J., Oubridge, C., Hernandez, H., Robinson, C.V., and Nagai, K. (2010). Recognition of a signal peptide by the signal recognition particle. *Nature* **465**(7297), 507-10.
(27) Egea, P.F., Shan, S.O., Napetschnig, J., Savage, D.F., Walter, P., and Stroud, R.M. (2004). Substrate twinning activates the signal recognition particle and its receptor. *Nature* **427**(6971), 215-21.
(28) Focia, P.J., Shepotinovskaya, I.V., Seidler, J.A., and Freymann, D.M. (2004). Heterodimeric GTPase core of the SRP targeting complex. *Science* **303**(5656), 373-7.
(29) Miller, J.D., Wilhelm, H., Gierasch, L., Gilmore, R., and Walter, P. (1993). GTP binding and hydrolysis by the signal recognition particle during initiation of protein translocation. *Nature* **366**(6453), 351-4.
(30) Gorlich, D., Hartmann, E., Prehn, S., and Rapoport, T.A. (1992). A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* **357**(6373), 47-52.
(31) Lingappa, V.R., Katz, F.N., Lodish, H.F., and Blobel, G. (1978). A signal sequence for the insertion of a transmembrane glycoprotein. Similarities to the signals of secretory proteins in primary structure and function. *J. Biol. Chem.* **253**(24), 8667-70.
(32) Wickner, W., and Schekman, R. (2005). Protein translocation across biological membranes. *Science* **310**(5753), 1452-6.
(33) Goder, V., Crottet, P., and Spiess, M. (2000). In vivo kinetics of protein targeting to the endoplasmic reticulum determined by site-specific phosphorylation. *EMBO J.* **19**(24), 6704-12.
(34) Takalo, M., Salminen, A., Soininen, H., Hiltunen, M., and Haapasalo, A. (2013). Protein aggregation and degradation mechanisms in neurodegenerative diseases. *Am J Neurodegener Dis* **2**(1), 1-14.
(35) Voisine, C., Pedersen, J.S., and Morimoto, R.I. (2010). Chaperone networks: tipping the balance in protein folding diseases. *Neurobiol. Dis.* **40**(1), 12-20.
(36) Haigis, M.C., and Yankner, B.A. (2010). The aging stress response. *Mol. Cell* **40**(2), 333-44.
(37) Dobson, C.M. (2003). Protein folding and misfolding. *Nature* **426**(6968), 884-90.
(38) Ast, T., Cohen, G., and Schuldiner, M. (2013). A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. *Cell* **152**(5), 1134-45.
(39) Powers, E.T., and Balch, W.E. (2008). Costly mistakes: translational infidelity and protein homeostasis. *Cell* **134**(2), 204-6.
(40) Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323-55.
(41) Varshavsky, A. (2008). Discovery of cellular regulation by protein degradation. *J. Biol. Chem.* **283**(50), 34469-89.
(42) Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **8**(7), 519-29.
(43) Kroemer, G., Marino, G., and Levine, B. (2010). Autophagy and the integrated stress response. *Mol. Cell* **40**(2), 280-93.
(44) Borgese, N., Brambillasca, S., Soffientini, P., Yabal, M., and Makarow, M. (2003). Biogenesis of tail-anchored proteins. *Biochem. Soc. Trans.* **31**(Pt 6), 1238-42.
(45) Borgese, N., Colombo, S., and Pedrazzini, E. (2003). The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane. *J. Cell Biol.* **161**(6), 1013-9.
(46) Borgese, N., and Fasana, E. (2011). Targeting pathways of C-tail-anchored proteins. *Biochim. Biophys. Acta* **1808**(3), 937-46.
(47) Beilharz, T., Egan, B., Silver, P.A., Hofmann, K., and Lithgow, T. (2003). Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in Saccharomyces cerevisiae. *J. Biol. Chem.* **278**(10), 8219-23.
(48) Fasana, E., Fossati, M., Ruggiano, A., Brambillasca, S., Hoogenraad, C.C., Navone, F., Francolini, M., and Borgese, N. (2010). A VAPB mutant linked to amyotrophic lateral sclerosis generates a novel form of organized smooth endoplasmic reticulum. *FASEB J.* **24**(5), 1419-30.
(49) Navarre WW, S.O. (1994). Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol Microbiol.*
(50) Simpson, P.J., Schwappach, B., Dohlman, H.G., and Isaacson, R.L. (2010). Structures of Get3, Get4, and Get5 provide new models for TA membrane protein targeting. *Structure* **18**(8), 897-902.
(51) Wawrzynow, A., Theibert, J.L., Murphy, C., Jona, I., Martonosi, A., and Collins, J.H. (1992). Sarcolipin, the "proteolipid" of skeletal muscle sarcoplasmic reticulum, is a unique, amphipathic, 31-residue peptide. *Arch. Biochem. Biophys.* **298**(2), 620-3.
(52) Borgese, N., and Righi, M. (2010). Remote origins of tail-anchored proteins. *Traffic* **11**(7), 877-85.

(53) Jahn, R., and Scheller, R.H. (2006). SNAREs--engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* **7**(9), 631-43.
(54) Borgese, N., Brambillasca, S., and Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. *Curr. Opin. Cell Biol.* **19**(4), 368-75.
(55) Jantti, J., Keranen, S., Toikkanen, J., Kuismanen, E., Ehnholm, C., Soderlund, H., and Olkkonen, V.M. (1994). Membrane insertion and intracellular transport of yeast syntaxin Sso2p in mammalian cells. *J. Cell Sci.* **107 (Pt 12)**, 3623-33.
(56) Linstedt, A.D., Foguet, M., Renz, M., Seelig, H.P., Glick, B.S., and Hauri, H.P. (1995). A C-terminally-anchored Golgi protein is inserted into the endoplasmic reticulum and then transported to the Golgi apparatus. *Proc. Natl. Acad. Sci. U. S. A.* **92**(11), 5102-5.
(57) Pedrazzini, E., Villa, A., and Borgese, N. (1996). A mutant cytochrome b5 with a lengthened membrane anchor escapes from the endoplasmic reticulum and reaches the plasma membrane. *Proc. Natl. Acad. Sci. U. S. A.* **93**(9), 4207-12.
(58) Hegde, R.S., and Keenan, R.J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **12**(12), 787-98.
(59) Kim, S.J., Jansson, S., Hoffman, N.E., Robinson, C., and Mant, A. (1999). Distinct "assisted" and "spontaneous" mechanisms for the insertion of polytopic chlorophyll-binding proteins into the thylakoid membrane. *J. Biol. Chem.* **274**(8), 4715-21.
(60) Henderson, M.P., Hwang, Y.T., Dyer, J.M., Mullen, R.T., and Andrews, D.W. (2007). The C-terminus of cytochrome b5 confers endoplasmic reticulum specificity by preventing spontaneous insertion into membranes. *Biochem. J.* **401**(3), 701-9.
(61) Costello, J.L., Castro, I.G., Camoes, F., Schrader, T.A., McNeall, D., Yang, J., Giannopoulou, E.A., Gomes, S., Pogenberg, V., Bonekamp, N.A*., et al.* (2017). Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells. *J. Cell Sci.* **130**(9), 1675-1687.
(62) Kemper, C., Habib, S.J., Engl, G., Heckmeyer, P., Dimmer, K.S., and Rapaport, D. (2008). Integration of tail-anchored proteins into the mitochondrial outer membrane does not require any known import components. *J. Cell Sci.* **121**(Pt 12), 1990-8.
(63) Meineke, B., Engl, G., Kemper, C., Vasiljev-Neumeyer, A., Paulitschke, H., and Rapaport, D. (2008). The outer membrane form of the mitochondrial protein Mcr1 follows a TOM-independent membrane insertion pathway. *FEBS Lett.* **582**(6), 855-60.
(64) Setoguchi, K., Otera, H., and Mihara, K. (2006). Cytosolic factor- and TOM-independent import of C-tail-anchored mitochondrial outer membrane proteins. *EMBO J.* **25**(24), 5635-47.
(65) Colombo, S.F., Longhi, R., and Borgese, N. (2009). The role of cytosolic proteins in the insertion of tail-anchored proteins into phospholipid bilayers. *J. Cell Sci.* **122**(Pt 14), 2383-92.
(66) Favaloro, V., Spasic, M., Schwappach, B., and Dobberstein, B. (2008). Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *J. Cell Sci.* **121**(11), 1832-40.
(67) Rabu, C., Wipf, P., Brodsky, J.L., and High, S. (2008). A precursor-specific role for Hsp40/Hsc70 during tail-anchored protein integration at the endoplasmic reticulum. *J. Biol. Chem.* **283**(41), 27504-13.
(68) Yabal, M., Brambillasca, S., Soffientini, P., Pedrazzini, E., Borgese, N., and Makarow, M. (2003). Translocation of the C terminus of a tail-anchored protein across the endoplasmic reticulum membrane in yeast mutants defective in signal peptide-driven translocation. *J. Biol. Chem.* **278**(5), 3489-96.
(69) Huber, N., Guimaraes, S., Schrader, M., Suter, U., and Niemann, A. (2013). Charcot-Marie-Tooth disease-associated mutants of GDAP1 dissociate its roles in peroxisomal and mitochondrial fission. *EMBO Rep* **14**(6), 545-52.
(70) Shaheen, R., Faqeih, E., Shamseldin, H.E., Noche, R.R., Sunker, A., Alshammari, M.J., Al-Sheddi, T., Adly, N., Al-Dosari, M.S., Megason, S.G*., et al.* (2012). POC1A truncation mutation causes a ciliopathy in humans characterized by primordial dwarfism. *Am. J. Hum. Genet.* **91**(2), 330-6.
(71) Leznicki, P., Clancy, A., Schwappach, B., and High, S. (2010). Bat3 promotes the membrane integration of tail-anchored proteins. *J. Cell Sci.* **123**(Pt 13), 2170-8.
(72) Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**(7310), 1120-4.
(73) Shao, S., Rodrigo-Brenni, M.C., Kivlen, M.H., and Hegde, R.S. (2017). Mechanistic basis for a molecular triage reaction. *Science* **355**(6322), 298-302.
(74) Hessa, T., Sharma, A., Mariappan, M., Eshleman, H.D., Gutierrez, E., and Hegde, R.S. (2011). Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* **475**(7356), 394-7.
(75) Rodrigo-Brenni, M.C., Gutierrez, E., and Hegde, R.S. (2014). Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell* **55**(2), 227-37.

(76) Mock, J.Y., Chartron, J.W., Zaslaver, M., Xu, Y., Ye, Y., and Clemons, W.M., Jr. (2015). Bag6 complex contains a minimal tail-anchor-targeting module and a mock BAG domain. *Proc. Natl. Acad. Sci. U. S. A.* **112**(1), 106-11.
(77) Favaloro, V., Vilardi, F., Schlecht, R., Mayer, M.P., and Dobberstein, B. (2010). Asna1/TRC40-mediated membrane insertion of tail-anchored proteins. *J. Cell Sci.* **123**(Pt 9), 1522-30.
(78) Mateja, A., Paduch, M., Chang, H.Y., Szydlowska, A., Kossiakoff, A.A., Hegde, R.S., and Keenan, R.J. (2015). Protein targeting. Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science* **347**(6226), 1152-5.
(79) Stefanovic, S., and Hegde, R.S. (2007). Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**(6), 1147-59.
(80) Vilardi, F., Stephan, M., Clancy, A., Janshoff, A., and Schwappach, B. (2014). WRB and CAML are necessary and sufficient to mediate tail-anchored protein targeting to the ER membrane. *PLoS One* **9**(1), e85033.
(81) Yamamoto, Y., and Sakisaka, T. (2012). Molecular machinery for insertion of tail-anchored membrane proteins into the endoplasmic reticulum membrane in mammalian cells. *Mol. Cell* **48**(3), 387-97.
(82) Darby, J.F., Krysztofinska, E.M., Simpson, P.J., Simon, A.C., Leznicki, P., Sriskandarajah, N., Bishop, D.S., Hale, L.R., Alfano, C., Conte, M.R.*, et al.* (2014). Solution structure of the SGTA dimerisation domain and investigation of its interactions with the ubiquitin-like domains of BAG6 and UBL4A. *PLoS One* **9**(11), e113281.
(83) Mateja, A., Szlachcic, A., Downing, M.E., Dobosz, M., Mariappan, M., Hegde, R.S., and Keenan, R.J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature* **461**(7262), 361-6.
(84) Sherrill, J., Mariappan, M., Dominik, P., Hegde, R.S., and Keenan, R.J. (2011). A conserved archaeal pathway for tail-anchored membrane protein insertion. *Traffic* **12**(9), 1119-23.
(85) Payapilly, A., and High, S. (2014). BAG6 regulates the quality control of a polytopic ERAD substrate. *J. Cell Sci.* **127**(Pt 13), 2898-909.
(86) Wunderley, L., Leznicki, P., Payapilly, A., and High, S. (2014). SGTA regulates the cytosolic quality control of hydrophobic substrates. *J. Cell Sci.* **127**(Pt 21), 4728-39.
(87) Chartron, J.W., Clemons, W.M., Jr., and Suloway, C.J. (2012). The complex process of GETting tail-anchored membrane proteins to the ER. *Curr. Opin. Struct. Biol.* **22**(2), 217-24.
(88) Wang, F., Brown, E.C., Mak, G., Zhuang, J., and Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* **40**(1), 159-71.
(89) Chartron, J.W., VanderVelde, D.G., Rao, M., and Clemons, W.M., Jr. (2012). Get5 carboxyl-terminal domain is a novel dimerization motif that tethers an extended Get4/Get5 complex. *J. Biol. Chem.* **287**(11), 8310-7.
(90) Simon, A.C., Simpson, P.J., Goldstone, R.M., Krysztofinska, E.M., Murray, J.W., High, S., and Isaacson, R.L. (2013). Structure of the Sgt2/Get5 complex provides insights into GET-mediated targeting of tail-anchored membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**(4), 1327-32.
(91) Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R.S., and Keenan, R.J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **477**(7362), 61-6.
(92) Kubota, K., Yamagata, A., Sato, Y., Goto-Ito, S., and Fukai, S. (2012). Get1 stabilizes an open dimer conformation of get3 ATPase by binding two distinct interfaces. *J. Mol. Biol.* **422**(3), 366-75.
(93) Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y.Y., Schwarz, D., Bomke, J., Hein, C., Lohr, F., Bernhard, F.*, et al.* (2011). Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* **333**(6043), 758-62.
(94) Suloway, C.J., Chartron, J.W., Zaslaver, M., and Clemons, W.M., Jr. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proc. Natl. Acad. Sci. U. S. A.* **106**(35), 14849-54.
(95) Bozkurt, G., Stjepanovic, G., Vilardi, F., Amlacher, S., Wild, K., Bange, G., Favaloro, V., Rippe, K., Hurt, E., Dobberstein, B.*, et al.* (2009). Structural insights into tail-anchored protein binding and membrane insertion by Get3. *Proc. Natl. Acad. Sci. U. S. A.* **106**(50), 21131-6.
(96) Hu, J., Li, J., Qian, X., Jin, Z., Fu, Z., and Sha, B. (2009). Preliminary X-ray crystallographic studies of yeast Get3. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **65**(Pt 5), 489-91.
(97) Chang, Y.W., Chuang, Y.C., Ho, Y.C., Cheng, M.Y., Sun, Y.J., Hsiao, C.D., and Wang, C. (2010). Crystal structure of Get4-Get5 complex and its interactions with Sgt2, Get3, and Ydj1. *J. Biol. Chem.* **285**(13), 9962-70.
(98) Gristick, H.B., Rome, M.E., Chartron, J.W., Rao, M., Hess, S., Shan, S.O., and Clemons, W.M., Jr. (2015). Mechanism of Assembly of a Substrate Transfer Complex during Tail-anchored Protein Targeting. *J. Biol. Chem.* **290**(50), 30006-17.

(99) Leznicki, P., and High, S. (2012). SGTA antagonizes BAG6-mediated protein triage. *Proc. Natl. Acad. Sci. U. S. A.* **109**(47), 19214-9.
(100) Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S., *et al.* (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**(5922), 1693-7.
(101) Chartron, J.W., Suloway, C.J., Zaslaver, M., and Clemons, W.M., Jr. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proc. Natl. Acad. Sci. U. S. A.* **107**(27), 12127-32.
(102) Gristick, H.B., Rao, M., Chartron, J.W., Rome, M.E., Shan, S.O., and Clemons, W.M., Jr. (2014). Crystal structure of ATP-bound Get3-Get4-Get5 complex reveals regulation of Get3 by Get4. *Nat. Struct. Mol. Biol.* **21**(5), 437-42.
(103) Rome, M.E., Chio, U.S., Rao, M., Gristick, H., and Shan, S.O. (2014). Differential gradients of interaction affinities drive efficient targeting and recycling in the GET pathway. *Proc. Natl. Acad. Sci. U. S. A.* **111**(46), E4929-35.
(104) Rome, M.E., Rao, M., Clemons, W.M., and Shan, S.O. (2013). Precise timing of ATPase activation drives targeting of tail-anchored proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**(19), 7666-71.
(105) Liou, S.T., Cheng, M.Y., and Wang, C. (2007). SGT2 and MDY2 interact with molecular chaperone YDJ1 in Saccharomyces cerevisiae. *Cell Stress Chaperones* **12**(1), 59-70.
(106) Shao, S., von der Malsburg, K., and Hegde, R.S. (2013). Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. *Mol. Cell* **50**(5), 637-48.
(107) Ciechanover, A., Heller, H., Katz-Etzion, R., and Hershko, A. (1981). Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system. *Proc. Natl. Acad. Sci. U. S. A.* **78**(2), 761-5.
(108) Ciechanover, A., and Stanhill, A. (2014). The complexity of recognition of ubiquitinated substrates by the 26S proteasome. *Biochim. Biophys. Acta* **1843**(1), 86-96.
(109) Kleiger, G., and Mayor, T. (2014). Perilous journey: a tour of the ubiquitin-proteasome system. *Trends Cell Biol.* **24**(6), 352-9.
(110) Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A.L. (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **78**(5), 761-71.
(111) Voges, D., Zwickl, P., and Baumeister, W. (1999). The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* **68**, 1015-68.
(112) Baehrecke, E.H. (2005). Autophagy: dual roles in life and death? *Nat. Rev. Mol. Cell Biol.* **6**(6), 505-10.
(113) Tanaka, K. (2009). The proteasome: overview of structure and functions. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **85**(1), 12-36.
(114) Pickart, C.M., and Cohen, R.E. (2004). Proteasomes and their kin: proteases in the machine age. *Nat. Rev. Mol. Cell Biol.* **5**(3), 177-87.
(115) Peters, J.M., Franke, W.W., and Kleinschmidt, J.A. (1994). Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. *J. Biol. Chem.* **269**(10), 7709-18.
(116) Hershko, A., and Ciechanover, A. (1992). The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* **61**, 761-807.
(117) Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405-39.
(118) Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**(1), 94-102.
(119) Komander, D., Clague, M.J., and Urbe, S. (2009). Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **10**(8), 550-63.
(120) Komander, D., and Rape, M. (2012). The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203-29.
(121) Finley, D. (2002). Ubiquitin chained and crosslinked. *Nat. Cell Biol.* **4**(5), E121-3.
(122) Finley, D., Ciechanover, A., and Varshavsky, A. (2004). Ubiquitin as a central cellular regulator. *Cell* **116**(2 Suppl), S29-32, 2 p following S32.
(123) Haas, A.L., Warms, J.V., Hershko, A., and Rose, I.A. (1982). Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. *J. Biol. Chem.* **257**(5), 2543-8.
(124) Hamilton, K.S., Ellison, M.J., Barber, K.R., Williams, R.S., Huzil, J.T., McKenna, S., Ptak, C., Glover, M., and Shaw, G.S. (2001). Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. *Structure* **9**(10), 897-904.
(125) Huang, D.T., Hunt, H.W., Zhuang, M., Ohi, M.D., Holton, J.M., and Schulman, B.A. (2007). Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature* **445**(7126), 394-8.
(126) Huang, R.Y., Kowalski, D., Minderman, H., Gandhi, N., and Johnson, E.S. (2007). Small ubiquitin-related modifier pathway is a major determinant of doxorubicin cytotoxicity in Saccharomyces cerevisiae. *Cancer Res.* **67**(2), 765-72.

(127) Lee, I., and Schindelin, H. (2008). Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134**(2), 268-78.
(128) Barthelme, D., and Sauer, R.T. (2012). Identification of the Cdc48*20S proteasome as an ancient AAA+ proteolytic machine. *Science* **337**(6096), 843-6.
(129) Barthelme, D., and Sauer, R.T. (2013). Bipartite determinants mediate an evolutionarily conserved interaction between Cdc48 and the 20S peptidase. *Proc. Natl. Acad. Sci. U. S. A.* **110**(9), 3327-32.
(130) Pye, V.E., Dreveny, I., Briggs, L.C., Sands, C., Beuron, F., Zhang, X., and Freemont, P.S. (2006). Going through the motions: the ATPase cycle of p97. *J. Struct. Biol.* **156**(1), 12-28.
(131) Wilkinson, K.D. (2000). Ubiquitination and deubiquitination: targeting of proteins for degradation by the proteasome. *Semin. Cell Dev. Biol.* **11**(3), 141-8.
(132) Murata, S., Minami, Y., Minami, M., Chiba, T., and Tanaka, K. (2001). CHIP is a chaperone-dependent E3 ligase that ubiquitylates unfolded protein. *EMBO Rep* **2**(12), 1133-8.
(133) Tanaka, H., Takahashi, T., Xie, Y., Minami, R., Yanagi, Y., Hayashishita, M., Suzuki, R., Yokota, N., Shimada, M., Mizushima, T.*, et al.* (2016). A conserved island of BAG6/Scythe is related to ubiquitin domains and participates in short hydrophobicity recognition. *FEBS J.* **283**(4), 662-77.
(134) Budhidarmo, R., Nakatani, Y., and Day, C.L. (2012). RINGs hold the key to ubiquitin transfer. *Trends Biochem. Sci.* **37**(2), 58-65.
(135) Deshaies, R.J., and Joazeiro, C.A. (2009). RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399-434.
(136) Das, R., Mariano, J., Tsai, Y.C., Kalathur, R.C., Kostova, Z., Li, J., Tarasov, S.G., McFeeters, R.L., Altieri, A.S., Ji, X.*, et al.* (2009). Allosteric activation of E2-RING finger-mediated ubiquitylation by a structurally defined specific E2-binding region of gp78. *Mol. Cell* **34**(6), 674-85.
(137) Ozkan, E., Yu, H., and Deisenhofer, J. (2005). Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc. Natl. Acad. Sci. U. S. A.* **102**(52), 18890-5.
(138) Wenzel, D.M., and Klevit, R.E. (2012). Following Ariadne's thread: a new perspective on RBR ubiquitin ligases. *BMC Biol.* **10**, 24.
(139) Scheffner, M., Nuber, U., and Huibregtse, J.M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**(6509), 81-3.
(140) Kim, H.C., and Huibregtse, J.M. (2009). Polyubiquitination by HECT E3s and the determinants of chain type specificity. *Mol. Cell. Biol.* **29**(12), 3307-18.
(141) Arndt, V., Dick, N., Tawo, R., Dreiseidler, M., Wenzel, D., Hesse, M., Furst, D.O., Saftig, P., Saint, R., Fleischmann, B.K.*, et al.* (2010). Chaperone-assisted selective autophagy is essential for muscle maintenance. *Curr. Biol.* **20**(2), 143-8.
(142) Vembar, S.S., and Brodsky, J.L. (2008). One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.* **9**(12), 944-57.
(143) Heck, J.W., Cheung, S.K., and Hampton, R.Y. (2010). Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. *Proc. Natl. Acad. Sci. U. S. A.* **107**(3), 1106-11.
(144) Hicke, L., Schubert, H.L., and Hill, C.P. (2005). Ubiquitin-binding domains. *Nat. Rev. Mol. Cell Biol.* **6**(8), 610-21.
(145) Hurley, J.H., Lee, S., and Prag, G. (2006). Ubiquitin-binding domains. *Biochem. J.* **399**(3), 361-72.
(146) Johnston, S.C., Riddle, S.M., Cohen, R.E., and Hill, C.P. (1999). Structural basis for the specificity of ubiquitin C-terminal hydrolases. *EMBO J.* **18**(14), 3877-87.
(147) Walden, H., Podgorski, M.S., and Schulman, B.A. (2003). Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* **422**(6929), 330-4.
(148) Husnjak, K., Elsasser, S., Zhang, N., Chen, X., Randles, L., Shi, Y., Hofmann, K., Walters, K.J., Finley, D., and Dikic, I. (2008). Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**(7194), 481-8.
(149) van Nocker, S., Sadis, S., Rubin, D.M., Glickman, M., Fu, H., Coux, O., Wefes, I., Finley, D., and Vierstra, R.D. (1996). The multiubiquitin-chain-binding protein Mcb1 is a component of the 26S proteasome in Saccharomyces cerevisiae and plays a nonessential, substrate-specific role in protein turnover. *Mol. Cell. Biol.* **16**(11), 6020-8.
(150) Walters, K.J., Kleijnen, M.F., Goh, A.M., Wagner, G., and Howley, P.M. (2002). Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. *Biochemistry* **41**(6), 1767-77.
(151) Young, P., Deveraux, Q., Beal, R.E., Pickart, C.M., and Rechsteiner, M. (1998). Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. *J. Biol. Chem.* **273**(10), 5461-7.

(152) Elsasser, S., Chandler-Militello, D., Muller, B., Hanna, J., and Finley, D. (2004). Rad23 and Rpn10 serve as alternative ubiquitin receptors for the proteasome. *J. Biol. Chem.* **279**(26), 26817-22.
(153) Elsasser, S., and Finley, D. (2005). Delivery of ubiquitinated substrates to protein-unfolding machines. *Nat. Cell Biol.* **7**(8), 742-9.
(154) Hofmann, K., and Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.* **21**(5), 172-3.
(155) Wilkinson, C.R., Seeger, M., Hartmann-Petersen, R., Stone, M., Wallace, M., Semple, C., and Gordon, C. (2001). Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat. Cell Biol.* **3**(10), 939-43.
(156) Kerscher, O., Felberbaum, R., and Hochstrasser, M. (2006). Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell. Dev. Biol.* **22**, 159-80.
(157) Jentsch, S., and Pyrowolakis, G. (2000). Ubiquitin and its kin: how close are the family ties? *Trends Cell Biol.* **10**(8), 335-42.
(158) Husnjak, K., and Dikic, I. (2012). Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu. Rev. Biochem.* **81**, 291-322.
(159) Harper, J.W., and Schulman, B.A. (2006). Structural complexity in ubiquitin recognition. *Cell* **124**(6), 1133-6.
(160) Xu, Y., Cai, M., Yang, Y., Huang, L., and Ye, Y. (2012). SGTA recognizes a noncanonical ubiquitin-like domain in the Bag6-Ubl4A-Trc35 complex to promote endoplasmic reticulum-associated degradation. *Cell Rep* **2**(6), 1633-44.
(161) Varadan, R., Assfalg, M., Raasi, S., Pickart, C., and Fushman, D. (2005). Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Mol. Cell* **18**(6), 687-98.
(162) Pierce, N.W., Kleiger, G., Shan, S.O., and Deshaies, R.J. (2009). Detection of sequential polyubiquitylation on a millisecond timescale. *Nature* **462**(7273), 615-9.
(163) Komander, D. (2009). The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* **37**(Pt 5), 937-53.
(164) Mevissen, T.E.T., and Komander, D. (2017). Mechanisms of Deubiquitinase Specificity and Regulation. *Annu. Rev. Biochem.* **86**, 159-192.
(165) Koulich, E., Li, X., and DeMartino, G.N. (2008). Relative structural and functional roles of multiple deubiquitylating proteins associated with mammalian 26S proteasome. *Mol. Biol. Cell* **19**(3), 1072-82.
(166) Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477-513.
(167) Leznicki, P., Korac-Prlic, J., Kliza, K., Husnjak, K., Nyathi, Y., Dikic, I., and High, S. (2015). Binding of SGTA to Rpn13 selectively modulates protein quality control. *J. Cell Sci.* **128**(17), 3187-96.
(168) Thapaliya, A., Nyathi, Y., Martinez-Lumbreras, S., Krysztofinska, E.M., Evans, N.J., Terry, I.L., High, S., and Isaacson, R.L. (2016). SGTA interacts with the proteasomal ubiquitin receptor Rpn13 via a carboxylate clamp mechanism. *Sci. Rep.* **6**, 36622.
(169) Krenciute, G., Liu, S., Yucer, N., Shi, Y., Ortiz, P., Liu, Q., Kim, B.J., Odejimi, A.O., Leng, M., Qin, J*., et al.* (2013). Nuclear BAG6-UBL4A-GET4 complex mediates DNA damage signaling and cell death. *J. Biol. Chem.* **288**(28), 20547-57.
(170) Claessen, J.H., and Ploegh, H.L. (2011). BAT3 guides misfolded glycoproteins out of the endoplasmic reticulum. *PLoS One* **6**(12), e28542.
(171) Wang, F., Song, W., Brancati, G., and Segatori, L. (2011). Inhibition of endoplasmic reticulum-associated degradation rescues native folding in loss of function protein misfolding diseases. *J. Biol. Chem.* **286**(50), 43454-64.
(172) Casson, J., McKenna, M., and High, S. (2016). On the road to nowhere: cross-talk between post-translational protein targeting and cytosolic quality control. *Biochem. Soc. Trans.* **44**(3), 796-801.
(173) Zhi, X., Zhao, D., Wang, Z., Zhou, Z., Wang, C., Chen, W., Liu, R., and Chen, C. (2013). E3 ubiquitin ligase RNF126 promotes cancer cell proliferation by targeting the tumor suppressor p21 for ubiquitin-mediated degradation. *Cancer Res.* **73**(1), 385-94.
(174) Krysztofinska, E.M., Martinez-Lumbreras, S., Thapaliya, A., Evans, N.J., High, S., and Isaacson, R.L. (2016). Structural and functional insights into the E3 ligase, RNF126. *Sci. Rep.* **6**, 26433.
(175) Claessen, J.H., Mueller, B., Spooner, E., Pivorunas, V.L., and Ploegh, H.L. (2010). The transmembrane segment of a tail-anchored protein determines its degradative fate through dislocation from the endoplasmic reticulum. *J. Biol. Chem.* **285**(27), 20732-9.
(176) Xu, Y., Liu, Y., Lee, J.G., and Ye, Y. (2013). A ubiquitin-like domain recruits an oligomeric chaperone to a retrotranslocation complex in endoplasmic reticulum-associated degradation. *J. Biol. Chem.* **288**(25), 18068-76.

(177) Kuwabara, N., Minami, R., Yokota, N., Matsumoto, H., Senda, T., Kawahara, H., and Kato, R. (2015). Structure of a BAG6 (Bcl-2-associated athanogene 6)-Ubl4a (ubiquitin-like protein 4a) complex reveals a novel binding interface that functions in tail-anchored protein biogenesis. *J. Biol. Chem.* **290**(15), 9387-98.
(178) Stuart, J.K., Myszka, D.G., Joss, L., Mitchell, R.S., McDonald, S.M., Xie, Z., Takayama, S., Reed, J.C., and Ely, K.R. (1998). Characterization of interactions between the anti-apoptotic protein BAG-1 and Hsc70 molecular chaperones. *J. Biol. Chem.* **273**(35), 22506-14.
(179) Takayama, S., Bimston, D.N., Matsuzawa, S., Freeman, B.C., Aime-Sempe, C., Xie, Z., Morimoto, R.I., and Reed, J.C. (1997). BAG-1 modulates the chaperone activity of Hsp70/Hsc70. *EMBO J.* **16**(16), 4887-96.
(180) Sondermann, H., Scheufler, C., Schneider, C., Hohfeld, J., Hartl, F.U., and Moarefi, I. (2001). Structure of a Bag/Hsc70 complex: convergent functional evolution of Hsp70 nucleotide exchange factors. *Science* **291**(5508), 1553-7.
(181) Thress, K., Henzel, W., Shillinglaw, W., and Kornbluth, S. (1998). Scythe: a novel reaper-binding apoptotic regulator. *EMBO J.* **17**(21), 6135-43.
(182) Liu, Y., Soetandyo, N., Lee, J.G., Liu, L., Xu, Y., Clemons, W.M., Jr., and Ye, Y. (2014). USP13 antagonizes gp78 to maintain functionality of a chaperone in ER-associated degradation. *Elife* **3**, e01369.
(183) Doyle, S.M., and Wickner, S. (2009). Hsp104 and ClpB: protein disaggregating machines. *Trends Biochem. Sci.* **34**(1), 40-8.
(184) Glover, J.R., and Lindquist, S. (1998). Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell* **94**(1), 73-82.
(185) Lum, R., Tkach, J.M., Vierling, E., and Glover, J.R. (2004). Evidence for an unfolding/threading mechanism for protein disaggregation by Saccharomyces cerevisiae Hsp104. *J. Biol. Chem.* **279**(28), 29139-46.
(186) Wendler, P., Shorter, J., Snead, D., Plisson, C., Clare, D.K., Lindquist, S., and Saibil, H.R. (2009). Motor mechanism for protein threading through Hsp104. *Mol. Cell* **34**(1), 81-92.
(187) Khosrow-Khavar, F., Fang, N.N., Ng, A.H., Winget, J.M., Comyn, S.A., and Mayor, T. (2012). The yeast ubr1 ubiquitin ligase participates in a prominent pathway that targets cytosolic thermosensitive mutants for degradation. *G3 (Bethesda)* **2**(5), 619-28.
(188) Nillegoda, N.B., Theodoraki, M.A., Mandal, A.K., Mayo, K.J., Ren, H.Y., Sultana, R., Wu, K., Johnson, J., Cyr, D.M., and Caplan, A.J. (2010). Ubr1 and Ubr2 function in a quality control pathway for degradation of unfolded cytosolic proteins. *Mol. Biol. Cell* **21**(13), 2102-16.
(189) Prasad, R., Kawaguchi, S., and Ng, D.T. (2010). A nucleus-based quality control mechanism for cytosolic proteins. *Mol. Biol. Cell* **21**(13), 2117-27.
(190) He, C., Bartholomew, C.R., Zhou, W., and Klionsky, D.J. (2009). Assaying autophagic activity in transgenic GFP-Lc3 and GFP-Gabarap zebrafish embryos. *Autophagy* **5**(4), 520-6.
(191) Kaganovich, D., Kopito, R., and Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**(7208), 1088-95.
(192) Delker, R.K., Zhou, Y., Strikoudis, A., Stebbins, C.E., and Papavasiliou, F.N. (2013). Solubility-based genetic screen identifies RING finger protein 126 as an E3 ligase for activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci. U. S. A.* **110**(3), 1029-34.
(193) Wang, L., Wang, X., Zhao, Y., Niu, W., Ma, G., Yin, W., and Shi, C. (2016). E3 Ubiquitin ligase RNF126 regulates the progression of tongue cancer. *Cancer Med* **5**(8), 2043-7.
(194) Laity, J.H., Lee, B.M., and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**(1), 39-46.
(195) Pabo, C.O., Peisach, E., and Grant, R.A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* **70**, 313-40.
(196) Klug, A. (1999). Zinc finger peptides for the regulation of gene expression. *J. Mol. Biol.* **293**(2), 215-8.
(197) Joazeiro, C.A., and Weissman, A.M. (2000). RING finger proteins: mediators of ubiquitin ligase activity. *Cell* **102**(5), 549-52.
(198) Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M., Hatakeyama, S., and Weissman, A.M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl. Acad. Sci. U. S. A.* **96**(20), 11364-9.
(199) Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**(4), 533-9.
(200) Eletr, Z.M., Huang, D.T., Duda, D.M., Schulman, B.A., and Kuhlman, B. (2005). E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer. *Nat. Struct. Mol. Biol.* **12**(10), 933-4.
(201) Brzovic, P.S., and Klevit, R.E. (2006). Ubiquitin transfer from the E2 perspective: why is UbcH5 so promiscuous? *Cell Cycle* **5**(24), 2867-73.

(202) Campbell, S.J., Edwards, R.A., Leung, C.C., Neculai, D., Hodge, C.D., Dhe-Paganon, S., and Glover, J.N. (2012). Molecular insights into the function of RING finger (RNF)-containing proteins hRNF8 and hRNF168 in Ubc13/Mms2-dependent ubiquitylation. *J. Biol. Chem.* **287**(28), 23900-10.
(203) Hibbert, R.G., Huang, A., Boelens, R., and Sixma, T.K. (2011). E3 ligase Rad18 promotes monoubiquitination rather than ubiquitin chain formation by E2 enzyme Rad6. *Proc. Natl. Acad. Sci. U. S. A.* **108**(14), 5590-5.
(204) Mace, P.D., Linke, K., Feltham, R., Schumacher, F.R., Smith, C.A., Vaux, D.L., Silke, J., and Day, C.L. (2008). Structures of the cIAP2 RING domain reveal conformational changes associated with ubiquitin-conjugating enzyme (E2) recruitment. *J. Biol. Chem.* **283**(46), 31633-40.
(205) Yin, Q., Lin, S.C., Lamothe, B., Lu, M., Lo, Y.C., Hura, G., Zheng, L., Rich, R.L., Campos, A.D., Myszka, D.G.*, et al.* (2009). E2 interaction and dimerization in the crystal structure of TRAF6. *Nat. Struct. Mol. Biol.* **16**(6), 658-66.
(206) Smith, C.J., Berry, D.M., and McGlade, C.J. (2013). The E3 ubiquitin ligases RNF126 and Rabring7 regulate endosomal sorting of the epidermal growth factor receptor. *J. Cell Sci.* **126**(Pt 6), 1366-80.
(207) Amemiya, Y., Azmi, P., and Seth, A. (2008). Autoubiquitination of BCA2 RING E3 ligase regulates its own stability and affects cell migration. *Mol. Cancer Res.* **6**(9), 1385-96.
(208) Burger, A.M., Gao, Y., Amemiya, Y., Kahn, H.J., Kitching, R., Yang, Y., Sun, P., Narod, S.A., Hanna, W.M., and Seth, A.K. (2005). A novel RING-type ubiquitin ligase breast cancer-associated gene 2 correlates with outcome in invasive breast cancer. *Cancer Res.* **65**(22), 10401-12.
(209) Sakane, A., Hatakeyama, S., and Sasaki, T. (2007). Involvement of Rabring7 in EGF receptor degradation as an E3 ligase. *Biochem. Biophys. Res. Commun.* **357**(4), 1058-64.
(210) Burger, A.M., Kona, F., Amemiya, Y., Gao, Y., Bacopulos, S., and Seth, A.K. (2010). Role of the BCA2 ubiquitin E3 ligase in hormone responsive breast cancer. *Open Cancer J.* **3**(1), 116-123.
(211) Kona, F.R., Stark, K., Bisoski, L., Buac, D., Cui, Q., and Dou, Q.P. (2012). Transcriptional activation of breast cancer-associated gene 2 by estrogen receptor. *Breast Cancer Res. Treat.* **135**(2), 495-503.
(212) Colomer-Lluch, M., and Serra-Moreno, R. (2017). BCA2/Rabring7 Interferes with HIV-1 Proviral Transcription by Enhancing the SUMOylation of IkappaBalpha. *J. Virol.* **91**(8).
(213) Nityanandam, R., and Serra-Moreno, R. (2014). BCA2/Rabring7 targets HIV-1 Gag for lysosomal degradation in a tetherin-independent manner. *PLoS Pathog.* **10**(5), e1004151.
(214) Shangary, S., Qin, D., McEachern, D., Liu, M., Miller, R.S., Qiu, S., Nikolovska-Coleska, Z., Ding, K., Wang, G., Chen, J.*, et al.* (2008). Temporal activation of p53 by a specific MDM2 inhibitor is selectively toxic to tumors and leads to complete tumor growth inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **105**(10), 3933-8.
(215) Soucy, T.A., Smith, P.G., and Rolfe, M. (2009). Targeting NEDD8-activated cullin-RING ligases for the treatment of cancer. *Clin. Cancer. Res.* **15**(12), 3912-6.
(216) Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C.*, et al.* (2004). In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**(5659), 844-8.
(217) Yoshino, S., Hara, T., Nakaoka, H.J., Kanamori, A., Murakami, Y., Seiki, M., and Sakamoto, T. (2016). The ERK signaling target RNF126 regulates anoikis resistance in cancer cells by changing the mitochondrial metabolic flux. *Cell Discov* **2**, 16019.
(218) Wang, Y., Deng, O., Feng, Z., Du, Z., Xiong, X., Lai, J., Yang, X., Xu, M., Wang, H., Taylor, D.*, et al.* (2016). RNF126 promotes homologous recombination via regulation of E2F1-mediated BRCA1 expression. *Oncogene* **35**(11), 1363-72.
(219) Smith, C.J., and McGlade, C.J. (2014). The ubiquitin ligase RNF126 regulates the retrograde sorting of the cation-independent mannose 6-phosphate receptor. *Exp. Cell Res.* **320**(2), 219-32.
(220) Campuzano, V., Montermini, L., Lutz, Y., Cova, L., Hindelang, C., Jiralerspong, S., Trottier, Y., Kish, S.J., Faucheux, B., Trouillas, P.*, et al.* (1997). Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum. Mol. Genet.* **6**(11), 1771-80.
(221) Evans-Galea, M.V., Pebay, A., Dottori, M., Corben, L.A., Ong, S.H., Lockhart, P.J., and Delatycki, M.B. (2014). Cell and gene therapy for Friedreich ataxia: progress to date. *Hum. Gene Ther.* **25**(8), 684-93.
(222) Strawser, C.J., Schadt, K.A., and Lynch, D.R. (2014). Therapeutic approaches for the treatment of Friedreich's ataxia. *Expert Rev. Neurother.* **14**(8), 949-57.
(223) Chutake, Y.K., Lam, C., Costello, W.N., Anderson, M., and Bidichandani, S.I. (2014). Epigenetic promoter silencing in Friedreich ataxia is dependent on repeat length. *Ann. Neurol.* **76**(4), 522-8.
(224) Condo, I., Ventura, N., Malisan, F., Rufini, A., Tomassini, B., and Testi, R. (2007). In vivo maturation of human frataxin. *Hum. Mol. Genet.* **16**(13), 1534-40.

(225) Bulteau, A.L., O'Neill, H.A., Kennedy, M.C., Ikeda-Saito, M., Isaya, G., and Szweda, L.I. (2004). Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity. *Science* **305**(5681), 242-5.
(226) Vaubel, R.A., and Isaya, G. (2013). Iron-sulfur cluster synthesis, iron homeostasis and oxidative stress in Friedreich ataxia. *Mol. Cell. Neurosci.* **55**, 50-61.
(227) Martelli, A., and Puccio, H. (2014). Dysregulation of cellular iron metabolism in Friedreich ataxia: from primary iron-sulfur cluster deficit to mitochondrial iron accumulation. *Front. Pharmacol.* **5**, 130.
(228) Pastore, A., and Puccio, H. (2013). Frataxin: a protein in search for a function. *J. Neurochem.* **126 Suppl 1**, 43-52.
(229) Benini, M., Fortuni, S., Condo, I., Alfedi, G., Malisan, F., Toschi, N., Serio, D., Massaro, D.S., Arcuri, G., Testi, R.*, et al.* (2017). E3 Ligase RNF126 Directly Ubiquitinates Frataxin, Promoting Its Degradation: Identification of a Potential Therapeutic Target for Friedreich Ataxia. *Cell Rep* **18**(8), 2007-2017.
(230) Callahan, M.A., Handley, M.A., Lee, Y.H., Talbot, K.J., Harper, J.W., and Panganiban, A.T. (1998). Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a novel member of the tetratricopeptide repeat protein family. *J. Virol.* **72**(6), 5189-97.
(231) Cziepluch, C., Kordes, E., Poirey, R., Grewenig, A., Rommelaere, J., and Jauniaux, J.C. (1998). Identification of a novel cellular TPR-containing protein, SGT, that interacts with the nonstructural protein NS1 of parvovirus H-1. *J. Virol.* **72**(5), 4149-56.
(232) Philp, L.K., Butler, M.S., Hickey, T.E., Butler, L.M., Tilley, W.D., and Day, T.K. (2013). SGTA: a new player in the molecular co-chaperone game. *Horm. Cancer* **4**(6), 343-57.
(233) Tobaben, S., Varoqueaux, F., Brose, N., Stahl, B., and Meyer, G. (2003). A brain-specific isoform of small glutamine-rich tetratricopeptide repeat-containing protein binds to Hsc70 and the cysteine string protein. *J. Biol. Chem.* **278**(40), 38376-83.
(234) Liou, S.T., and Wang, C. (2005). Small glutamine-rich tetratricopeptide repeat-containing protein is composed of three structural units with distinct functions. *Arch. Biochem. Biophys.* **435**(2), 253-63.
(235) Winnefeld, M., Rommelaere, J., and Cziepluch, C. (2004). The human small glutamine-rich TPR-containing protein is required for progress through cell division. *Exp. Cell Res.* **293**(1), 43-57.
(236) Buchanan, G., Ricciardelli, C., Harris, J.M., Prescott, J., Yu, Z.C., Jia, L., Butler, L.M., Marshall, V.R., Scher, H.I., Gerald, W.L.*, et al.* (2007). Control of androgen receptor signaling in prostate cancer by the cochaperone small glutamine rich tetratricopeptide repeat containing protein alpha. *Cancer Res.* **67**(20), 10087-96.
(237) Trotta, A.P., Need, E.F., Butler, L.M., Selth, L.A., O'Loughlin, M.A., Coetzee, G.A., Tilley, W.D., and Buchanan, G. (2012). Subdomain structure of the co-chaperone SGTA and activity of its androgen receptor client. *J. Mol. Endocrinol.* **49**(2), 57-68.
(238) Fonte, V., Kapulkin, W.J., Taft, A., Fluet, A., Friedman, D., and Link, C.D. (2002). Interaction of intracellular beta amyloid peptide with chaperone proteins. *Proc. Natl. Acad. Sci. U. S. A.* **99**(14), 9439-44.
(239) Dupzyk, A., Williams, J.M., Bagchi, P., Inoue, T., and Tsai, B. (2017). SGTA-Dependent Regulation of Hsc70 Promotes Cytosol Entry of Simian Virus 40 from the Endoplasmic Reticulum. *J. Virol.* **91**(12).
(240) Dutta, S., and Tan, Y.J. (2008). Structural and functional characterization of human SGT and its interaction with Vpu of the human immunodeficiency virus type 1. *Biochemistry* **47**(38), 10123-31.
(241) Fielding, B.C., Gunalan, V., Tan, T.H., Chou, C.F., Shen, S., Khan, S., Lim, S.G., Hong, W., and Tan, Y.J. (2006). Severe acute respiratory syndrome coronavirus protein 7a interacts with hSGT. *Biochem. Biophys. Res. Commun.* **343**(4), 1201-8.
(242) Angeletti, P.C., Walker, D., and Panganiban, A.T. (2002). Small glutamine-rich protein/viral protein U-binding protein is a novel cochaperone that affects heat shock protein 70 activity. *Cell Stress Chaperones* **7**(3), 258-68.
(243) Schantl, J.A., Roza, M., De Jong, A.P., and Strous, G.J. (2003). Small glutamine-rich tetratricopeptide repeat-containing protein (SGT) interacts with the ubiquitin-dependent endocytosis (UbE) motif of the growth hormone receptor. *Biochem. J.* **373**(Pt 3), 855-63.
(244) Tobaben, S., Thakur, P., Fernandez-Chacon, R., Sudhof, T.C., Rettig, J., and Stahl, B. (2001). A trimeric protein complex functions as a synaptic chaperone machine. *Neuron* **31**(6), 987-99.
(245) Azakami, D., Nakahira, R., Kato, Y., Michishita, M., Kobayashi, M., Onozawa, E., Bonkobara, M., Kobayashi, M., Takahashi, K., Watanabe, M.*, et al.* (2017). The canine prostate cancer cell line CHP-1 shows over-expression of the co-chaperone small glutamine-rich tetratricopeptide repeat-containing protein alpha. *Vet. Comp. Oncol.* **15**(2), 557-562.
(246) Kato, Y., Ochiai, K., Kawakami, S., Nakao, N., Azakami, D., Bonkobara, M., Michishita, M., Morimatsu, M., Watanabe, M., and Omi, T. (2017). Canine REIC/Dkk-3 interacts with SGTA and

restores androgen receptor signalling in androgen-independent prostate cancer cell lines. *BMC Vet. Res.* **13**(1), 170.

(247) Ochiai, K., Morimatsu, M., Kato, Y., Ishiguro-Oonuma, T., Udagawa, C., Rungsuriyawiboon, O., Azakami, D., Michishita, M., Ariyoshi, Y., Ueki, H.*, et al.* (2016). Tumor suppressor REIC/DKK-3 and co-chaperone SGTA: Their interaction and roles in the androgen sensitivity. *Oncotarget* **7**(3), 3283-96.

(248) Butler, M.S., Yang, X., Ricciardelli, C., Liang, X., Norman, R.J., Tilley, W.D., and Hickey, T.E. (2013). Small glutamine-rich tetratricopeptide repeat-containing protein alpha is present in human ovaries but may not be differentially expressed in relation to polycystic ovary syndrome. *Fertil. Steril.* **99**(7), 2076-83 e1.

(249) Lu, C., Liu, G., Cui, X., Zhang, J., Wei, L., Wang, Y., Yang, X., Liu, Y., Cong, X., Lv, L.*, et al.* (2014). Expression of SGTA correlates with prognosis and tumor cell proliferation in human hepatocellular carcinoma. *Pathol. Oncol. Res.* **20**(1), 51-60.

(250) Yang, X., Cheng, L., Li, M., Shi, H., Ren, H., Ding, Z., Liu, F., Wang, Y., and Cheng, C. (2014). High expression of SGTA in esophageal squamous cell carcinoma correlates with proliferation and poor prognosis. *J. Cell. Biochem.* **115**(1), 141-50.

(251) Goodarzi, M.O., Xu, N., Cui, J., Guo, X., Chen, Y.I., and Azziz, R. (2008). Small glutamine-rich tetratricopeptide repeat-containing protein alpha (SGTA), a candidate gene for polycystic ovary syndrome. *Hum. Reprod.* **23**(5), 1214-9.

(252) Kiktev, D.A., Patterson, J.C., Muller, S., Bariar, B., Pan, T., and Chernoff, Y.O. (2012). Regulation of chaperone effects on a yeast prion by cochaperone Sgt2. *Mol. Cell. Biol.* **32**(24), 4960-70.

(253) Winnefeld, M., Grewenig, A., Schnolzer, M., Spring, H., Knoch, T.A., Gan, E.C., Rommelaere, J., and Cziepluch, C. (2006). Human SGT interacts with Bag-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. *Exp. Cell Res.* **312**(13), 2500-14.

(254) Philp, L.K., Day, T.K., Butler, M.S., Laven-Law, G., Jindal, S., Hickey, T.E., Scher, H.I., Butler, L.M., and Tilley, W.D. (2016). Small Glutamine-Rich Tetratricopeptide Repeat-Containing Protein Alpha (SGTA) Ablation Limits Offspring Viability and Growth in Mice. *Sci. Rep.* **6**, 28950.

(255) Tung, J.Y., Li, Y.C., Lin, T.W., and Hsiao, C.D. (2013). Structure of the Sgt2 dimerization domain complexed with the Get5 UBL domain involved in the targeting of tail-anchored membrane proteins to the endoplasmic reticulum. *Acta Crystallogr. D Biol. Crystallogr.* **69**(Pt 10), 2081-90.

(256) Leznicki, P., Warwicker, J., and High, S. (2011). A biochemical analysis of the constraints of tail-anchored protein biogenesis. *Biochem. J.* **436**(3), 719-27.

(257) Lamb, J.R., Tugendreich, S., and Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* **20**(7), 257-9.

(258) Blatch, G.L., and Lassle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**(11), 932-9.

(259) Sikorski, R.S., Boguski, M.S., Goebl, M., and Hieter, P. (1990). A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell* **60**(2), 307-17.

(260) Roberts, J.D., Thapaliya, A., Martinez-Lumbreras, S., Krysztofinska, E.M., and Isaacson, R.L. (2015). Structural and Functional Insights into Small, Glutamine-Rich, Tetratricopeptide Repeat Protein Alpha. *Front Mol Biosci* **2**, 71.

(261) Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* **101**(2), 199-210.

(262) Zhang, J., Lu, C., He, S., Wan, C., Zhang, Y., and Cheng, C. (2014). Decreased expression of Small glutamine-rich tetratricopeptide repeat-containing protein (SGT) correlated with prognosis of Hepatocellular carcinoma. *Neoplasma* **61**(1), 83-9.

(263) Chartron, J.W., Gonzalez, G.M., and Clemons, W.M., Jr. (2011). A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex. *J. Biol. Chem.* **286**(39), 34325-34.

(264) Carrigan, P.E., Sikkink, L.A., Smith, D.F., and Ramirez-Alvarado, M. (2006). Domain:domain interactions within Hop, the Hsp70/Hsp90 organizing protein, are required for protein stability and structure. *Protein Sci.* **15**(3), 522-32.

(265) Cortajarena, A.L., Liu, T.Y., Hochstrasser, M., and Regan, L. (2010). Designed proteins to modulate cellular networks. *ACS Chem. Biol.* **5**(6), 545-52.

(266) Javadi, Y., and Main, E.R. (2009). Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. *Proc. Natl. Acad. Sci. U. S. A.* **106**(41), 17383-8.

(267) Lowe, A.R., and Itzhaki, L.S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc. Natl. Acad. Sci. U. S. A.* **104**(8), 2679-84.

(268) Main, E.R., Xiong, Y., Cocco, M.J., D'Andrea, L., and Regan, L. (2003). Design of stable alpha-helical arrays from an idealized TPR motif. *Structure* **11**(5), 497-508.
(269) Galler, K.M., Aulisa, L., Regan, K.R., D'Souza, R.N., and Hartgerink, J.D. (2010). Self-assembling multidomain peptide hydrogels: designed susceptibility to enzymatic cleavage allows enhanced cell migration and spreading. *J. Am. Chem. Soc.* **132**(9), 3217-23.
(270) Zahnd, C., Kawe, M., Stumpp, M.T., de Pasquale, C., Tamaskovic, R., Nagy-Davidescu, G., Dreier, B., Schibli, R., Binz, H.K., Waibel, R.*, et al.* (2010). Efficient tumor targeting with high-affinity designed ankyrin repeat proteins: effects of affinity and molecular size. *Cancer Res.* **70**(4), 1595-605.
(271) Worrall, L.J., Wear, M.A., Page, A.P., and Walkinshaw, M.D. (2008). Cloning, purification and characterization of the Caenorhabditis elegans small glutamine-rich tetratricopeptide repeat-containing protein. *Biochim. Biophys. Acta* **1784**(3), 496-503.
(272) Buchner, E., and Gundersen, C.B. (1997). The DnaJ-like cysteine string protein and exocytotic neurotransmitter release. *Trends Neurosci.* **20**(5), 223-7.
(273) Pratt, W.B., and Toft, D.O. (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp. Biol. Med. (Maywood)* **228**(2), 111-33.
(274) Frydman, J., and Hohfeld, J. (1997). Chaperones get in touch: the Hip-Hop connection. *Trends Biochem. Sci.* **22**(3), 87-92.
(275) Morgan, R.M., Pal, M., Roe, S.M., Pearl, L.H., and Prodromou, C. (2015). Tah1 helix-swap dimerization prevents mixed Hsp90 co-chaperone complexes. *Acta Crystallogr. D Biol. Crystallogr.* **71**(Pt 5), 1197-206.
(276) Hessling, M., Richter, K., and Buchner, J. (2009). Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **16**(3), 287-93.
(277) Taipale, M., Jarosz, D.F., and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* **11**(7), 515-28.
(278) Chamberlain, L.H., and Burgoyne, R.D. (1998). The cysteine-string domain of the secretory vesicle cysteine-string protein is required for membrane targeting. *Biochem. J.* **335 (Pt 2)**, 205-9.
(279) Handley, M.A., Paddock, S., Dall, A., and Panganiban, A.T. (2001). Association of Vpu-binding protein with microtubules and Vpu-dependent redistribution of HIV-1 Gag protein. *Virology* **291**(2), 198-207.
(280) Waheed, A.A., MacDonald, S., Khan, M., Mounts, M., Swiderski, M., Xu, Y., Ye, Y., and Freed, E.O. (2016). The Vpu-interacting Protein SGTA Regulates Expression of a Non-glycosylated Tetherin Species. *Sci. Rep.* **6**, 24934.
(281) Waheed, A.A., Ono, A., and Freed, E.O. (2009). Methods for the study of HIV-1 assembly. *Methods Mol. Biol.* **485**, 163-84.
(282) Chen, M.Y., Maldarelli, F., Karczewski, M.K., Willey, R.L., and Strebel, K. (1993). Human immunodeficiency virus type 1 Vpu protein induces degradation of CD4 in vitro: the cytoplasmic domain of CD4 contributes to Vpu sensitivity. *J. Virol.* **67**(7), 3877-84.
(283) Schubert, U., Anton, L.C., Bacik, I., Cox, J.H., Bour, S., Bennink, J.R., Orlowski, M., Strebel, K., and Yewdell, J.W. (1998). CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin-conjugating pathway. *J. Virol.* **72**(3), 2280-8.
(284) Neil, S.J., Zang, T., and Bieniasz, P.D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**(7177), 425-30.
(285) Walczak, C.P., Ravindran, M.S., Inoue, T., and Tsai, B. (2014). A cytosolic chaperone complexes with dynamic membrane J-proteins and mobilizes a nonenveloped virus out of the endoplasmic reticulum. *PLoS Pathog.* **10**(3), e1004007.
(286) Chen, S., and Smith, D.F. (1998). Hop as an adaptor in the heat shock protein 70 (Hsp70) and hsp90 chaperone machinery. *J. Biol. Chem.* **273**(52), 35194-200.
(287) Echeverria, P.C., and Picard, D. (2010). Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. *Biochim. Biophys. Acta* **1803**(6), 641-9.
(288) Hernandez, M.P., Chadli, A., and Toft, D.O. (2002). HSP40 binding is the first step in the HSP90 chaperoning pathway for the progesterone receptor. *J. Biol. Chem.* **277**(14), 11873-81.
(289) Paul, A., Garcia, Y.A., Zierer, B., Patwardhan, C., Gutierrez, O., Hildenbrand, Z., Harris, D.C., Balsiger, H.A., Sivils, J.C., Johnson, J.L.*, et al.* (2014). The cochaperone SGTA (small glutamine-rich tetratricopeptide repeat-containing protein alpha) demonstrates regulatory specificity for the androgen, glucocorticoid, and progesterone receptors. *J. Biol. Chem.* **289**(22), 15297-308.
(290) Hanke, K., Chudak, C., Kurth, R., and Bannert, N. (2013). The Rec protein of HERV-K(HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT). *Int. J. Cancer* **132**(3), 556-67.
(291) Ochiai, K., Watanabe, M., Azakami, D., Michishita, M., Yoshikawa, Y., Udagawa, C., Metheenukul, P., Chahomchuen, T., Aoki, H., Kumon, H.*, et al.* (2013). Molecular cloning and

tumour suppressor function analysis of canine REIC/Dkk-3 in mammary gland tumours. *Vet. J.* **197**(3), 769-75.
(292) Sakagami, H., Zhou, L., Kawano, M., Thet, M.M., Tanaka, S., Machino, M., Amano, S., Kuroshita, R., Watanabe, S., Chu, Q.*, et al.* (2010). Multiple biological complex of alkaline extract of the leaves of Sasa senanensis Rehder. *In Vivo* **24**(5), 735-43.
(293) Kikukawa, Y., Minami, R., Shimada, M., Kobayashi, M., Tanaka, K., Yokosawa, H., and Kawahara, H. (2005). Unique proteasome subunit Xrpn10c is a specific receptor for the antiapoptotic ubiquitin-like protein Scythe. *FEBS J.* **272**(24), 6373-86.
(294) Minami, R., Hayakawa, A., Kagawa, H., Yanagi, Y., Yokosawa, H., and Kawahara, H. (2010). BAG-6 is essential for selective elimination of defective proteasomal substrates. *J. Cell Biol.* **190**(4), 637-50.
(295) Bhattacharyya, S., Yu, H., Mim, C., and Matouschek, A. (2014). Regulated protein turnover: snapshots of the proteasome in action. *Nat. Rev. Mol. Cell Biol.* **15**(2), 122-33.
(296) Aviram, N., Ast, T., Costa, E.A., Arakel, E.C., Chuartzman, S.G., Jan, C.H., Hassdenteufel, S., Dudek, J., Jung, M., Schorr, S.*, et al.* (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature* **540**(7631), 134-138.
(297) Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. *Nature* **425**(6959), 686-91.
(298) Fleischer, T.C., Weaver, C.M., McAfee, K.J., Jennings, J.L., and Link, A.J. (2006). Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes Dev.* **20**(10), 1294-307.
(299) Ricarte, F., Menjivar, R., Chhun, S., Soreta, T., Oliveira, L., Hsueh, T., Serranilla, M., and Gharakhanian, E. (2011). A genome-wide immunodetection screen in S. cerevisiae uncovers novel genes involved in lysosomal vacuole function and morphology. *PLoS One* **6**(8), e23696.
(300) Yeh, Y.H., Lin, T.W., Li, Y.C., Tung, J.Y., Lin, C.Y., and Hsiao, C.D. (2014). Structural and functional characterization of ybr137wp implicates its involvement in the targeting of tail-anchored proteins to membranes. *Mol. Cell. Biol.* **34**(24), 4500-12.
(301) Yoshida, K., Seo, H.S., Debler, E.W., Blobel, G., and Hoelz, A. (2011). Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **108**(40), 16571-6.
(302) Hore, P.J. (2003). *Nuclear magnetic resonance* (Oxford: Oxford University Press).
(303) Keeler, J. (2013). *Understanding NMR Spectroscopy* (Hoboken: Wiley).
(304) Wishart, D.S., and Sykes, B.D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363-92.
(305) Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**(4), 213-23.
(306) Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**(3), 381-2.
(307) Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**(1), 283-98.
(308) Shen, Y., Vernon, R., Baker, D., and Bax, A. (2009). De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* **43**(2), 63-78.
(309) Ooi, L.-l. (2010). *Principles of x-ray crystallography* (Oxford; New York: Oxford University Press).
(310) Ewald, P. (1969). Introduction to the dynamical theory of X-ray diffraction. *Acta Crystallographica Section A* **25**(1), 103-108.
(311) Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* **336**(6084), 1030-3.
(312) Hill, R.E., and Eaton-Rye, J.J. (2014). Plasmid construction by SLIC or sequence and ligation-independent cloning. *Methods Mol. Biol.* **1116**, 25-36.
(313) Kuzmic, P. (2009). DynaFit--a software package for enzymology. *Methods Enzymol.* **467**, 247-80.
(314) Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**(3), 277-93.
(315) Johnson, B.A., and Blevins, R.A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**(5), 603-14.
(316) Skinner, S.P., Goult, B.T., Fogh, R.H., Boucher, W., Stevens, T.J., Laue, E.D., and Vuister, G.W. (2015). Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallogr. D Biol. Crystallogr.* **71**(Pt 1), 154-61.
(317) Schrodinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

(318) Vangone, A., Rodrigues, J.P., Xue, L.C., van Zundert, G.C., Geng, C., Kurkcuoglu, Z., Nellen, M., Narasimhan, S., Karaca, E., van Dijk, M.*, et al.* (2017). Sense and simplicity in HADDOCK scoring: Lessons from CASP-CAPRI round 1. *Proteins* **85**(3), 417-423.
(319) Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D., and Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun* **1**, 100.
(320) Winter, G., Lobley, C.M., and Prince, S.M. (2013). Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**(Pt 7), 1260-73.
(321) Krissinel, E.B., Winn, M.D., Ballard, C.C., Ashton, A.W., Patel, P., Potterton, E.A., McNicholas, S.J., Cowtan, K.D., and Emsley, P. (2004). The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 12 Pt 1), 2250-5.
(322) McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**(Pt 4), 658-674.
(323) Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**(Pt 4), 355-67.
(324) Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**(Pt 2), 213-21.
(325) Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 12 Pt 1), 2126-32.
(326) Jung, Y.S., and Zweckstetter, M. (2004). Mars -- robust automatic backbone assignment of proteins. *J. Biomol. NMR* **30**(1), 11-23.
(327) Marchant, J., Sawmynaden, K., Saouros, S., Simpson, P., and Matthews, S. (2008). Complete resonance assignment of the first and second apple domains of MIC4 from Toxoplasma gondii, using a new NMRView-based assignment aid. *Biomol NMR Assign* **2**(2), 119-21.
(328) Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**(3), 774-97.
(329) McNicholas, S., Potterton, E., Wilson, K.S., and Noble, M.E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67**(Pt 4), 386-94.
(330) Dominguez, C., Boelens, R., and Bonvin, A.M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**(7), 1731-7.
(331) Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**(Web Server issue), W320-4.
(332) Jerabek-Willemsen, M., Wienken, C.J., Braun, D., Baaske, P., and Duhr, S. (2011). Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**(4), 342-53.
(333) Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123**(13), 2970-8.
(334) Kornhaber, G.J., Snyder, D., Moseley, H.N., and Montelione, G.T. (2006). Identification of zinc-ligated cysteine residues based on 13Calpha and 13Cbeta chemical shift data. *J. Biomol. NMR* **34**(4), 259-69.
(335) Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**(Pt 1), 12-21.
(336) Zhang, M., Windheim, M., Roe, S.M., Peggie, M., Cohen, P., Prodromou, C., and Pearl, L.H. (2005). Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. *Mol. Cell* **20**(4), 525-38.
(337) Ardley, H.C., and Robinson, P.A. (2005). E3 ubiquitin ligases. *Essays Biochem.* **41**, 15-30.
(338) Carrigan, P.E., Nelson, G.M., Roberts, P.J., Stoffer, J., Riggs, D.L., and Smith, D.F. (2004). Multiple domains of the co-chaperone Hop are important for Hsp70 binding. *J. Biol. Chem.* **279**(16), 16185-93.
(339) Panigrahi, R., Whelan, J., and Vrielink, A. (2014). Exploring ligand recognition, selectivity and dynamics of TPR domains of chloroplast Toc64 and mitochondria Om64 from Arabidopsis thaliana. *J. Mol. Recognit.* **27**(6), 402-14.
(340) Prasad, B.D., Goel, S., and Krishna, P. (2010). In silico identification of carboxylate clamp type tetratricopeptide repeat proteins in Arabidopsis and rice as putative co-chaperones of Hsp90/Hsp70. *PLoS One* **5**(9), e12761.
(341) Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1-16.

(342) Brinker, A., Scheufler, C., Von Der Mulbe, F., Fleckenstein, B., Herrmann, C., Jung, G., Moarefi, I., and Hartl, F.U. (2002). Ligand discrimination by TPR domains. Relevance and selectivity of EEVD-recognition in Hsp70 x Hop x Hsp90 complexes. *J. Biol. Chem.* **277**(22), 19265-75.

Appendix A. Primers

Primer Name	Primer Sequence
SGTA (1-69)	5'- GACGACAAGATGGACAACAAGAAGCGCCTGGCC GAGGAGAAGCCCGGTTACTTGCCCGTGGCAGCCGC - 3'
UBL4A_UBL	5'- GACGACGACAAGATGCAGCTGACGGTGAAGGCG GAGGAGAAGCCCGGTTACAGGGGTTTGACCACTAGGTT -3'
BAG6_UBL	5'- GACGACGACAAGATGGAACCGAATGATAGTACC GAGGAGAAGCCCGGTTAACCAGAGCTTGCACCGCT -3'
Sgt2_TPR (96-225)	5'-GCCGCGGATCCATGGCGGAAACTAAAGCCAAAGCTGAGG GGCCGGCTCGAGTTACAGATTCAAAGATTGCTCAACCTTCTT C-3'
RNF126_ H14A	5'-CGGGTCGTTATTTTGTGCTTGTTGTAGCGTTG CAACGCTACAACAAGCACAAAAATAACGACCCG -3'
RNF126_F36A	5'- CGTTGTGAAAGCGGTGCTATTGAAGAACTGTG CACAGTTCTTCAATAGCACCGCTTTCACAACG -3'
E38R/E39R	5'- GTGAAAGCGGTTTTATTCGACGACTGTGAGAAGAAACCCG CGGGTTTCTTCTCACAGTCGTCGAATAAAACCGCTTTCAC-3'
RNF126_1_40 (NZF)	5'- CTTCTCATAATTCTTCAATAAAACCG CGGTTTTATTGAAGAATTATGAGAAG -3'
RNF126_1_100	5'- GCCGCGGATCCATGGCAGAAGCAAGTCCGCATCCGG TCATGCACCAGGCGGAAAGG-3'
RNF126_RING	5'- GCCGCGGATCCATGACCGAAGAACATGTGGGTAGCGGTCTG G GGCCGGCTCGAGttaGCTAACACCGGTCAGACCCGGAGG-3'
RNF126_142_hydro_fr w	5'- GCCGCGGATCCATGGGTGTTCCGACCCTGGAAG GTAT
RNF126_FL	5'- GCCGCGGATCCATGGCAGAAGCAAGTCCGCATCCGG GTCCGCTGAGCAATCTGACCCTGTAACTCGAGCCGGCC -3'
SLIC_RNF126	5'- CATGGATCCCTGGAAATACAGG GATCCGAATTCGAGCTCCGTC -3'

Appendix B. Protein sequences and constructs

SGTA (1-69)

MAHHHHHHVDDDDKMDNKKRLAYAIIQFLHDQLRHGGLSSDAQESLEVAIQCLETAFGVTVEDSDLALPQ
TLPEIFEAAATGK

SGTA_NT (1-86)

MAHHHHHHVDDDDKMDNKKRLAYAIIQFLHDQLRHGGLSSDAQESLEVAIQCLETAFGVTVEDSDLALPQ
TLPEIFEAAATGKEMPQDLRSPARTP

RNF126_NZF (1-40)

MAEASPHPGRYFCHCCSVEIVPRLPDYICPRCESGFIEEL

RNF126_NZF (1-100)

MAEASPHPGRYFCHCCSVEIVPRLPDYICPRCESGFIEELPEETRSTENGSAPSTAPTDQSRPPLEHVDQHLF
TLPQGYGQFAFGIFDDSFEIPTFPPGA

RNF126 (full-length)

MAEASPHPGRYFCHCCSVEIVPRLPDYICPRCESGFIEELPEETRSTENGSAPSTAPTDQSRPPLEHVDQHL
FTLPQGYGQFAFGIFDDSFEIPTFPPGAQADDGRDPESRRERDHPSRHRYGARQPRARLTTRRATGRHEG
VPTLEGIIQQLVNGIITPATIPSLGPWGVLHSNPMDYAWGANGLDAIITQLLNQFENTGPPPADKEKIQAL
PTVPVTEEHVGSGLECPVCKDDYALGERVRQLPCNHLFHDGCIVPWLEQHDSCPVCRKSLTGQNTATN
PPGLTGVSFSSSSSSSSSSSPSNENATWSPLGRPQPPRPLSNLTL

UBL4A_UBL (1-74)

MAHHHHHHVDDDDKMQLTVKALQGRECLQVPEDELVSTLKQLVSEKLNVPVRQQRLLFKGKALADGKRL
SDYSIGPNSKLNLVVK

BAG6_UBL (17-101)

MAHHHHHHVDDDDKMLEVLVKTLDSQTRTFIVGAQMNVKEFKEHIAASVSIPSEKQRLIYQ
GRVLQDDKKLQEYNVGGKVIHLVERAPSGPSSG

Sgt2_TPR (93-229)

EDDAETKAKAEDLKMQGNKAMANKDYELAINKYTEAIKVLPTNAIYYANRAAAHSSLKEYDQAVKDAES
AISIDPSYFRGYSRLGFAKYAQGKPEEALEAYKKVLDIEGDNATEAMKRDYESAKKKVEQSLNLEKTV

Sgt2_TPR (96-225)

AETKAKAEDLKMQGNKAMANKDYELAINKYTEAIKVLPTNAIYYANRAAAHSSLKEYDQAVKDAESAISI
DPSYFRGYSRLGFAKYAQGKPEEALEAYKKVLDIEGDNATEAMKRDYESAKKKVEQSLN

Appendix C. Chemical shifts of RNF126_NZF

	H	N	N_ε	H_α	H_β	H_γ	H_δ	H_ε	H_ζ	C_α	C_β	C_γ	C_δ	C_ε
-1 Gly				3.90						43.50				
0 Ser				4.52	3.87 3.91					58.40	64.08			
1 Met	8.59	123.22		4.49	2.11 2.01	2.56 2.63		2.10		55.77	32.94	32.09		16.97
2 Ala	8.34	125.80		4.30	1.38					52.79	19.28			
3 Glu	8.28	120.73		4.26	2.05 1.95	2.27				56.35	30.51	36.31		
4 Ala	8.27	125.68		4.34	1.39					52.43	19.49			
5 Ser	8.26	117.16		4.74	3.84 3.80					56.34	63.49			
6 Pro				4.41	2.24 1.80	1.97	3.78 3.69			63.28	32.20	27.17	50.73	
7 His	8.49	120.18		4.99	3.22 3.15					53.51	29.18			
8 Pro				4.55	2.34 2.03	2.01	3.60 3.71			63.51	32.39	27.34	50.61	
9 Gly	8.63	110.41		3.95 4.05						45.13				
10 Arg	8.23	120.94	85.43	4.78	1.82 1.64	1.67 1.62	3.23 3.15			55.44	32.99	27.16	43.70	
10 Arg								7.55						
11 Tyr	8.69	121.35		5.33	2.57 2.94		6.95	6.84		56.96	43.10		133.50	118.02
12 Phe	9.41	122.43		4.89	2.77		6.52	7.04		56.98	44.05		131.64	129.74
13 Cys	7.94	128.43		4.65	2.80 2.73					58.36	31.22			
14 His	8.97	126.46		4.32	2.76 3.06					58.71	29.22			
15 Cys	9.20	123.86		4.44	3.28 3.22					61.70	28.37			
16 Cys	9.79	124.87		4.00	2.88 2.63						30.96			
17 Ser	7.95	114.11		3.76	3.89 4.21					58.87	61.70			
18 Val	6.44	111.97		4.46	2.05	0.69 0.79				59.09	36.71	19.39 21.63		
19 Glu	8.35	123.39		5.02	2.05	2.67 2.30				56.70	30.33	36.92		
20 Ile	8.54	117.72		4.74	1.68	1.25 1.07	0.74			58.97	44.64	26.54 21.24	15.33	
21 Val	8.20	124.18		4.59	1.99	1.03 0.97				59.67	32.48	20.94 20.98		
22 Pro				4.06	1.16 1.30	0.82 1.11	3.54 3.46			62.58	32.05	26.63	51.07	
23 Arg								7.31						
23 Arg	8.63	121.32	84.33	4.26	1.51 1.70	1.61	3.14 3.08			55.98	30.90	27.07	43.97	
24 Leu	8.52	127.89		4.17	1.36 1.54	1.41	0.40 0.71			52.96	41.44	27.30	24.75 22.51	
25 Pro				4.34	2.28	1.95 1.71	3.44 3.57			63.67	34.62	24.24	50.29	
26 Asp	8.75	122.04		4.39	2.35 2.60					56.34	41.01			
27 Tyr	8.67	117.59		3.44	3.04		6.78	6.70		60.40	35.03		132.89	118.32
28 Ile	6.76	111.77		4.78	1.41	1.04 0.80 0.63	0.70			57.58	42.06	25.56 17.24	12.58	
29 Cys	8.79	126.92		4.42	3.24 2.81					56.38	32.21			
30 Pro				4.38	2.11	2.38 1.97	4.17 3.87			64.06	32.32	26.81	51.41	
31 Arg								7.32						
31 Arg	9.60	124.11	85.49	4.30	2.12 1.89	1.56	3.06 3.24			58.20	31.65	27.50	43.77	
32 Cys	8.55	119.73		4.84	2.45 3.22					58.56	33.24			
33 Glu	7.87	118.11		4.08	2.30 2.23	2.08 1.98				58.50	26.76	36.85		
34 Ser	8.78	117.20		4.43	4.30 3.86					59.28	65.47			
35 Gly	8.51	111.10		3.77 4.11						45.22				
36 Phe	8.85	126.40		5.15	3.46 3.08		7.28	7.30	7.20	55.24	36.08			
37 Ile	7.97	117.33		5.35	1.68	1.37 1.00 0.82	0.33			58.62	41.13	26.06 18.05	14.29	
38 Glu	9.13	124.83		4.88	2.14	2.31 2.37				54.68	34.56	35.93		
39 Glu	8.84	127.83		3.85	1.91 1.71	2.19 1.91				56.58	30.38	36.22		
40 Leu	8.19	133.53		4.14	1.41	1.40	0.66 0.71			56.33	43.39	27.36	23.01 25.77	

Appendix D. Chemical shifts of Sgt2_TPR (90-229) from the backbone assignment

	H	N	C	C_α	C_β
90 Gly				43.59	
91 Ser	8.85	117.34	175.34	59.51	63.68
92 Met	8.97	123.15	176.83	57.58	29.40
93 Glu	8.18	121.08		54.99	41.21
94 Asp	8.27	122.09		55.53	41.08
95 Asp	8.27	122.09	177.39	55.62	41.07
96 Ala	8.40	125.01	180.69	55.42	18.53
97 Glu	8.49	120.44	179.23	59.46	29.34
98 Thr	8.14	119.39	175.71	66.85	68.72
99 Lys	8.04	121.04	178.04	60.20	32.16
100 Ala	8.01	120.71	180.85	55.24	17.93
101 Lys	7.97	120.73	179.27	59.35	32.49
102 Ala	8.24	123.62	179.24	55.93	19.36
103 Glu	8.72	117.11	179.18	58.66	30.28
104 Asp	7.94	122.02	178.80	57.72	41.59
105 Leu	7.56	122.20	179.00	58.61	41.61
106 Lys	8.56	121.38	178.53	60.17	30.08
107 Met	8.56	120.16	179.70	59.02	31.19
108 Gln	8.38	121.00	180.52	59.65	29.08
109 Gly	8.61	111.24	174.12	47.60	
110 Asn	9.09	122.92	178.85	55.68	36.82
111 Lys	8.36	124.56	178.35	59.58	32.00
112 Ala	7.63	123.14	180.37	55.29	16.86
113 Met	8.46	118.60	180.82	57.99	31.17
114 Ala	7.90	124.67	178.68	54.93	17.72
115 Asn	7.50	116.33	174.00	52.96	39.03
116 Lys	7.81	115.17	173.90	57.46	28.71
117 Asp	8.10	120.68	176.30	51.13	39.26
118 Tyr	6.98	121.19	177.80	60.86	37.63
119 Glu	8.93	119.17	179.03	59.89	29.02
120 Leu	7.36	122.35	178.61	57.49	41.47
121 Ala	7.92	121.32	178.18	56.11	18.19
122 Ile	8.29	117.40	179.58	66.78	37.52
123 Asn	7.92	123.31	177.73	57.17	38.39
124 Lys	8.45	120.50	179.29	57.52	30.36
125 Tyr	8.84	117.29	178.11	59.01	37.46
126 Thr	8.29	116.37	175.68	67.49	68.49
127 Glu	7.94	123.48	178.36	59.61	29.60
128 Ala	8.06	123.77	178.21	56.03	16.46
129 Ile	7.85	119.27	176.50	64.71	39.04
130 Lys	7.66	117.82	179.31	58.97	32.34
131 Val	7.47	118.22	175.78	65.15	32.20
132 Leu	8.49	121.47	179.24	52.35	44.65
133 Pro			178.66	64.99	32.16
134 Thr	7.15	108.28	172.35	60.61	67.95
135 Asn	7.87	120.73	175.44	51.94	39.31
136 Ala	9.13	128.57	178.61	55.71	20.22
137 Ile	7.69	116.32	178.36	64.96	37.43
138 Tyr	6.84	117.61	178.40	57.95	36.89
139 Tyr	6.88	116.22	178.08	62.47	38.53
140 Ala	8.10	122.28	179.41	55.48	17.79
141 Asn	8.48	119.33	177.08	55.20	35.00
142 Arg	7.91	124.05	178.16	61.14	30.26
143 Ala	8.76	122.37	179.57	56.07	18.05
144 Ala	7.49	120.16	180.36	55.01	17.54
145 Ala	7.77	123.20	178.83	56.02	19.34
146 His	8.63	114.98	179.14	60.30	27.78
147 Ser	9.15	120.14		64.12	63.08
148 Ser	7.98	122.04	174.49	62.87	61.49
149 Leu	6.92	121.00	175.56	54.35	42.64
150 Lys	7.75	114.36	175.49	57.43	28.52
151 Glu	7.97	121.85	178.18	53.20	27.03
152 Tyr	7.16	119.72	177.42	61.22	38.01
153 Asp	8.66	119.50	179.48	57.78	39.53
154 Gln	8.08	120.77	177.95	58.67	27.10
155 Ala	8.00	122.83	179.04	56.28	20.57
156 Val	8.44	118.26	177.12	67.62	31.70
157 Lys	7.28	118.54	179.58	59.50	32.05
158 Asp	8.02	121.05	178.36	57.99	40.24
159 Ala	8.89	125.27	179.18	55.74	18.35
160 Glu	8.85	116.55	180.89	59.87	28.75

	H	N	C	C_α	C_β
160 Glu	8.85	116.55	180.89	59.87	28.75
161 Ser	8.34	119.87	175.53	62.69	
162 Ala	7.88	124.93	178.63	55.95	17.79
163 Ile	8.01	119.27	177.05	64.62	39.14
164 Ser	7.72	113.59	176.56	60.87	63.25
165 Ile	7.66	121.83	176.92	64.21	39.03
166 Asp	8.19	115.46		50.92	41.78
167 Pro			178.50	63.72	32.05
168 Ser	8.43	116.26	173.38	58.64	64.32
169 Tyr	7.72	124.92	177.97	55.75	36.60
170 Phe	8.97	128.45	176.56	61.03	38.06
171 Arg			178.26	57.99	29.65
172 Gly	6.93	105.35	174.41	47.90	
173 Tyr	6.62	118.86	178.51	61.42	37.83
174 Ser	7.95	115.81	177.14	63.02	
175 Arg	8.08	118.86	179.37	57.81	29.17
176 Leu	7.91	122.85	177.82	58.24	42.40
177 Gly	8.69	106.04	174.28	47.89	
178 Phe	8.04	122.41	178.13	62.37	39.67
179 Ala	7.59	121.62	178.66	55.26	19.19
180 Lys	8.38	115.42	179.08	57.21	31.76
181 Tyr	8.60	121.40	179.85	62.52	37.82
182 Ala	7.93	124.40	179.25	54.59	16.54
183 Gln	7.35	115.05	176.19	56.10	30.25
184 Gly	7.73	107.95	174.16	45.99	
185 Lys	7.81	119.14	174.62	53.59	31.90
186 Pro			177.41	65.59	31.14
187 Glu	9.22	121.88	179.47	60.57	28.07
188 Glu	7.45	119.91	180.04	59.34	30.01
189 Ala	7.97	123.56	177.52	55.37	18.74
190 Leu	8.62	118.82	177.93	58.96	42.18
191 Glu	7.28	116.59	179.11	59.24	29.46
192 Ala	7.35	123.41	179.44	55.61	17.44
193 Tyr	8.88	116.22	180.15	58.77	37.41
194 Lys	9.10	123.52	177.87	59.25	32.34
195 Lys	7.50	120.73	178.67	58.32	30.77
196 Val	7.37	117.03	177.34	67.96	31.84
197 Leu	7.00	117.08	180.41	57.94	40.15
198 Asp	8.52	122.73	178.78	57.11	40.36
199 Ile	8.22	120.78	178.56	64.93	39.39
200 Glu	8.36	116.82	178.99	57.20	30.55
201 Gly	8.27	110.58	177.13	47.07	
202 Asp	9.18	128.36	176.41	56.19	40.22
203 Asn	8.19	115.92	175.86	53.10	38.93
204 Ala	7.56	124.06	177.42	52.88	19.13
205 Thr	8.21	115.12	175.64	60.17	71.52
206 Glu	8.90	121.08	178.97	59.70	28.92
207 Ala	8.23	121.61	179.32	55.42	18.30
208 Met	7.37	117.60	180.45	58.99	32.08
209 Lys	7.76	121.87	178.36	60.36	31.98
210 Arg	8.51	121.60	179.32	59.35	29.85
211 Asp	8.04	122.39	177.71	57.60	38.49
212 Tyr	6.88	118.87	175.87	61.90	39.19
213 Glu	8.50	119.15	179.47	59.29	29.37
214 Ser	8.61	116.25	177.27	61.52	62.35
215 Ala	7.62	125.64	177.91	55.66	17.68
216 Lys	8.39	120.58	178.48	60.00	32.53
217 Lys	8.14	118.76	178.84	58.88	32.30
218 Lys	7.56	119.42	179.57	58.00	32.18
219 Val	8.15	121.30	178.62	66.64	31.83
220 Glu	8.26	120.43	178.75	59.12	29.47
221 Gln	8.08	118.61	177.64	57.77	28.76
222 Ser	7.90	115.71	175.04	59.90	63.66
223 Leu	7.79	122.94	177.41	55.84	42.36
224 Asn	8.13	118.75	175.31	53.57	38.63
225 Leu	8.09	122.45	177.69	55.57	42.32
226 Glu	8.31	121.62	176.41	56.73	30.01
227 Lys	8.23	122.99	176.64	56.16	33.15
228 Thr	8.29	117.44	173.89	62.10	69.92
229 Val	7.78	127.01	181.00	63.72	33.25

Appendix E. Dano of Sgt2_TPR_PTVEEVD

Fragment of the DANO Anomalous difference for anomalous Fourier (DANO) is a difference Patterson electron density map used to identify the positions of Zinc ions in the structure of Sgt2_TPR/PTVEEVD complex.