This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

KING'S College LONDON

# Biophysical studies of SGTA, a protein involved in quality control

Thapaliya, Arjun

*Awarding institution:*
King's College London

**The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.**

END USER LICENCE AGREEMENT



**Unless another licence is stated on the immediately following page** this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

**Take down policy**

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 08. Oct. 2023

# Biophysical studies of SGTA, a protein involved in quality control

Arjun Thapaliya




A thesis submitted to King's College London, University of London, in candidature for the degree of Doctor of Philosophy

Department of Chemistry,
King's College London,
Britannia House,
7 Trinity Street,
London SE1 1DB

February 2017

## Abstract

Protein quality control mechanisms are vital in maintaining correct levels of functional proteins in a crowded intracellular milieu. These mechanisms recognise proteins in their non-native state and triage such candidates for either rescue, or channel them towards pathways that lead to their degradation. In the case of secretory and membrane proteins that mislocalize to the cytosol (MLPs), and of newly synthesised tail-anchored (TA) membrane proteins, their exposure of otherwise buried hydrophobic residues necessitates a mechanism of shielding such residues from the aqueous cytoplasm. This task is carried out by the collective actions of SGTA (small, glutamine-rich, tetratricopeptide repeat protein alpha) and the heterotrimeric BAG6 (BCL2-associated athanogene 6) complex, that contribute to a SGTA/BAG6 quality control cycle that can direct hydrophobic substrates towards either ubiquitination and proteasomal degradation or productive membrane insertion.

SGTA is a modular protein consisting of three domains, an N-terminal dimerisation domain, a central TPR domain, and a substrate binding C-terminal domain. The aim of this thesis is to characterise full-length SGTA using a range of biophysical techniques, with an emphasis to understand its C-terminal domain, molecular details of which remain elusive. In addition, this work aims to uncover how SGTA interacts with its hydrophobic substrates, underlying its role in enforcing cytosolic quality control. To this end, a combination of nuclear magnetic resonance (NMR) spectroscopy and native mass spectrometry experiments have identified a constrained conformation of SGTA in solution mediated by C-terminal dimerisation, and circular dichroism (CD) spectroscopy has revealed the presence of alpha helical regions within the C-terminal domain. Furthermore, fluorine-19 NMR has been used to investigate the interaction of TA proteins with the C-terminal domain of SGTA. Finally, results presented herein establish that SGTA interacts with the intrinsic proteasomal ubiquitin receptor Rpn13 via a two-carboxylate clamp mode of molecular recognition, and this interaction has been characterised by solution NMR spectroscopy, size exclusion chromatography, isothermal titration calorimetry (ITC) and mutagenesis experiments, to uncover details of a potential SGTA/BAG6 quality control cycle operating at the 19S regulatory particle of the proteasome.

## Acknowledgements

I would like to thank my supervisors Dr Rivka Isaacson and Professor Ulrike Eggert for giving me the opportunity to work in their labs and for their constant support and guidance throughout my Ph.D. I would also like to extend my gratitude to all members of the Isaacson and Eggert labs, past and present, for their help at different stages of my project. I would like to specially thank Dr Santiago Martinez-Lumbreras for guiding me in the lab, introducing me to protein NMR spectroscopy and for the many useful scientific discussions. I also thank Dr Martinez-Lumbreras for carrying out the NMR backbone assignments presented herein, which have helped this work significantly.

I am indebted to Professor Stephen High, University of Manchester, for stimulating discussions and for his generous provision of materials. I thank Dr. R. Andrew Atkinson at the Centre for Biomolecular Spectroscopy, King's College London, and Dr Geoff Kelly at the MRC Biomedical NMR Centre, The Francis Crick Institute, London for their assistance with NMR experiments. I would also like to thank Dr Dijana Matak-Vinkovic at the Department of Chemistry, University of Cambridge, for carrying out the native mass spectrometry experiments described in this thesis.

I would like to thank Professor Christine Slingsby, Department of Crystallography, Birkbeck College, University of London, for introducing me to the field of protein structure and function and for her guidance and support over the years. I thank my parents Dr Bishnu Thapaliya and Mrs Sunita Khadka Thapaliya for their unwavering encouragement and support. I would like to thank all my friends and colleagues, in particular Mr Om Sapkota for making my time at Britannia House an enjoyable one. Finally, I gratefully acknowledge the Biotechnology and Biological Sciences Research Council (BBSRC) of the United Kingdom for funding my Ph.D. through the London Interdisciplinary Doctoral training (LIDo) consortium.

## Table of Contents


Abstract 2

Acknowledgements 3

List of Figures 8

List of Tables 13

Abbreviations 14

**Chapter 1 Introduction** 17

1.1 Protein homeostasis and macromolecular quality control 18

1.2 The proteostasis network 20

1.3 Small glutamine-rich, tetratricopeptide repeat-containing protein alpha (SGTA) 23

1.4 Structural studies of SGTA 24

*1.4.1 N-terminal domain* 27

*1.4.2 TPR domain* 29

*1.4.3 C-terminal domain* 31

1.5 The heterotrimeric BAG6 complex 32

1.6 The SGTA/BAG6 cycle of MLP quality control 37

1.7 SGTA in Get/TRC pathways of TA membrane protein insertion 40

1.8 SGTA and the ERAD pathway 45

1.9 SGTA mediated quality control at the proteasome 46

*1.9.1 The 26S proteasome* 46

*1.9.2 The intrinsic proteasomal ubiquitin receptor Rpn13* 47

1.10 SGTA in health and disease 50

1.11 SGTA in hormone receptor signaling 51

1.12 The role of SGTA in viral lifecycles 52

1.13 Aims of work described in this thesis 53

**Chapter 2 Materials and methods** 54

2.1 Chemicals 55

2.2 Water 55

2.3 Synthetic genes 55

2.4 Synthetic DNA oligonucleotides 55

2.5 Vectors and plasmids 55

2.6 *E. coli* strains 56
2.7 Bacterial growth media 56
2.8 Synthetic peptides 57
2.9 Chromatography 57
2.10 Agarose gel electrophoresis 57
2.11 SDS-PAGE 58
2.12 Estimation of protein concentration 58
2.13 Restriction endonuclease mediated cloning 59
*2.13.1 Construct design and domain boundaries* 59
*2.13.2 Primer design* 60
*2.13.3 Polymerase chain reaction* 63
*2.13.4 Purification of DNA fragments* 64
*2.13.5 Restriction digestion* 64
*2.13.6 Ligation* 64
*2.13.7 Transformation* 64
*2.13.8 Isolation of plasmid DNA* 65
2.14 Site-directed mutagenesis 65
2.15 Heterologous expression and purification of recombinant proteins 67
*2.15.1 Transformation* 67
*2.15.2 Preparation of overnight pre-cultures* 67
*2.15.3 Expression of recombinant proteins* 68
*2.15.4 Purification of recombinant proteins by nickel affinity chromatography* 68
*2.15.5 Size-exclusion chromatography (SEC)* 69
*2.15.6 Purification of recombinant tail-anchored proteins* 70
2.16 Analytical size-exclusion chromatography 70
2.17 Solution NMR spectroscopy 71
*2.17.1 1D $^{1}$H NMR* 73
*2.17.2 2D $^{1}$H-$^{15}$N HSQC experiments* 73
*2.17.3 NMR sample conditions* 74
*2.17.4 NMR titrations* 74
*2.17.5 Triple resonance experiments for sequential backbone assignment* 75
*2.17.6 $^{15}$N NMR relaxation experiments* 75

*2.17.7* *PRE experiments* 76
*2.17.8* *Fluorine-19 NMR* 78
2.18 Isothermal titration calorimetry (ITC) 79
2.19 Circular dichroism spectroscopy 81
2.20 Native mass spectrometry 82
2.21 High ambiguity driven protein-protein docking 83

**Chapter 3 Biophysical characterisation of SGTA** 85
3.1. Expression and purification of SGTA constructs 86
3.2. Analytical SEC of SGTA constructs 91
3.3. Circular dichroism spectroscopy of SGTA constructs 91
3.4. Characterisation of SGTA constructs by solution NMR spectroscopy 94
3.5. Solution dynamics of SGTA domains 103
3.6 Native mass spectrometry of SGTA constructs 110
3.7. Summary 113

**Chapter 4 Investigation of the interaction between SGTA and hydrophobic substrates** 114
4.1. Expression and purification of SGTA and TA protein constructs for interaction studies 115
4.2. Investigation of the interaction between SGTA and hydrophobic substrates using standard NMR methods 119
4.3. Design and preparation of SGTA C-terminal cysteine mutants for $^{19}$F NMR 126
4.4. $^{19}$F NMR of SGTA C-terminal domain cysteine mutants 128
4.5. $^{19}$F NMR to investigate the interaction between SGTA and Cytb5 133
4.6. Summary 139

**Chapter 5 Biophysical characterisation of the interaction between SGTA and the intrinsic proteasomal ubiquitin receptor Rpn13** 140
5.1 Expression and purification of Rpn13 and SGTA_TPR constructs 141
5.2 SGTA interacts with the C-terminal domain of Rpn13 145
5.3 Characterisation of the SGTA-Rpn13 interaction by solution NMR spectroscopy 147

5.4 The Rpn13 extreme C-terminal DMSLD pentapeptide is necessary and sufficient for its interaction with SGTA_TPR 156

5.5 The SGTA_TPR-Rpn13 interaction occurs via a two-carboxylate clamp mechanism 165

5.6 Intermolecular PRE experiments to delineate the SGTA-Rpn13 binding interface 170

5.7 SGTA_TPR/Rpn13 C-terminal DMSLD pentapeptide complex 174

5.8 Summary 177

**Chapter 6 Discussion** 178

6.1 Structure and dynamics of full-length SGTA *in vitro* 179

6.2 SGTA interacts with hydrophobic substrates via its C-terminal domain 181

6.3 SGTA associates with the intrinsic proteasomal ubiquitin receptor Rpn13 via a carboxylate clamp mechanism 183

6.4 SGTA's role in the quality control of hydrophobic substrates 185

6.5 Concluding remarks 189

**Bibliography** 192

**Appendices** 214

## List of Figures


| | | |
|---|---|---|
| Figure 1.1 | The proteostasis network | 22 |
| Figure 1.2 | Schematic of full-length SGTA | 25 |
| Figure 1.3 | SGTA sequence alignment | 26 |
| Figure 1.4 | SGTA N-terminal domain highlighting its UBL-binding domain | 28 |
| Figure 1.5 | N-terminal domain of SGTA in complex with a cognate UBL | 29 |
| Figure 1.6 | The TPR domain of SGTA | 30 |
| Figure 1.7 | The BAG6 complex | 34 |
| Figure 1.8 | TUGS domain of Ubl4a in complex with BAG-similar domain of BAG6 | 35 |
| Figure 1.9 | The BAG6 UBL domain in complex with the N-terminal zinc finger domain of the E3 ligase RNF126 | 36 |
| Figure 1.10 | The SGTA/BAG6 cycle of quality control in the cytosol. | 39 |
| Figure 1.11 | The TRC pathway of TA protein insertion | 42 |
| Figure 1.12 | Get3 ATPase | 43 |
| Figure 1.13 | SGTA in the ERAD pathway | 45 |
| Figure 1.14 | The 26S proteasome | 47 |
| Figure 1.15 | Structure of full-length human Rpn13 | 48 |
| Figure 1.16 | Rpn13 activation upon binding to the proteasome | 49 |
| Figure 2.1 | Vector map of modified pET28 used for restriction endonuclease mediated cloning | 61 |
| Figure 2.2 | 3-bromo-1,1,1-trifluoroacetone (BTFA) | 79 |
| Figure 3.1 | Schematics of various SGTA constructs. | 87 |
| Figure 3.2 | Recombinant expression of the human SGTA constructs in *E. coli* | 88 |
| Figure 3.3 | Purification of the N-terminal domain of SGTA | 89 |
| Figure 3.4 | Purification of SGTA constructs | 90 |
| Figure 3.5 | Analytical size-exclusion chromatography (SEC) profiles of SGTA constructs | 92 |
| Figure 3.6 | Circular dichroism (CD) spectra of isolated (A) N-terminal, (B) TPR and (C) C-terminal domains of SGTA | 93 |
| Figure 3.7 | $^{1}$H NMR spectra of different SGTA constructs | 96 |
| Figure 3.8 | $^{1}$H-$^{15}$N HSQC spectra of full-length SGTA | 97 |

| | | |
|---|---|---|
| Figure 3.9 | The N-terminal and TPR domains of human SGTA are structurally independent modules | 98 |
| Figure 3.10 | $^{1}H$-$^{15}N$ HSQC spectra of the excised C-terminal domain of SGTA | 99 |
| Figure 3.11 | The C-terminal domain of SGTA is independent of both its N-terminal and TPR domains | 100 |
| Figure 3.12 | Chemical shift perturbation (CSP) analysis of SGTA domains | 101 |
| Figure 3.13 | Chemical shift index (CSI) analysis of full-length SGTA | 102 |
| Figure 3.14 | $^{1}H$-$^{15}N$ heteronuclear NOE enhancement values of various SGTA constructs | 104 |
| Figure 3.15 | $T_1$, $T_2$ and $T_1/T_2$ values of the full-length construct (acquired at 950 MHz) plotted as a function of SGTA residue number. | 105 |
| Figure 3.16 | $T_1$, $T_2$ and $T_1/T_2$ values of the Nter-TPR construct construct (acquired at 700 MHz) plotted as a function of SGTA residue number. | 106 |
| Figure 3.17 | $T_1$, $T_2$ and $T_1/T_2$ values of the isolated Nter domain construct (acquired at 500 MHz) plotted as a function of SGTA residue number. | 107 |
| Figure 3.18 | $T_1$, $T_2$ and $T_1/T_2$ values of the excised TPR domain construct (acquired at 700 MHz) plotted as a function of SGTA residue number. | 108 |
| Figure 3.19 | Dynamic parameters of the TPR domain are highly affected in full-length SGTA. | 109 |
| Figure 3.20 | Non dissociative nano-electrospray ionization mass spectra (Native MS) of SGTA samples | 112 |
| Figure 4.1 | SGTA and TA protein constructs designed to understand the shielding of hydrophobic substrates by SGTA | 116 |
| Figure 4.2 | Purification of the TPR_NNP construct | 117 |
| Figure 4.3 | Purification of tail-anchored (TA) membrane proteins | 118 |
| Figure 4.4 | Identification of backbone amide signals corresponding to C-terminal residues of SGTA from the $^{1}H$-$^{15}N$ HSQC of the TPR_NNP construct | 120 |
| Figure 4.5 | SGTA interacts with cytochrome b5 (Cytb5) | 121 |
| Figure 4.6 | SGTA interacts with synaptobrevin 2 (Syb2) | 122 |
| Figure 4.7 | Hydrophobic substrates interact with residues in the C-terminal domain of SGTA | 123 |

| | | |
|---|---|---|
| Figure 4.8 | Chemical shift perturbation (CSP) delta average values of TPR_NNP amide signals upon interaction with TA protein substrates | 124 |
| Figure 4.9 | Changes in intensity values of TPR_NNP amide signals upon interaction with TA protein substrates | 125 |
| Figure 4.10 | Schematic of SGTA C-terminal domain showing positions of cysteine mutants generated to facilitate the post translational incorporation of fluorine probes (BTFA) | 126 |
| Figure 4.11 | Preparation of SGTA C-terminal domain cysteine mutants | 127 |
| Figure 4.12 | 1D $^{19}F$ fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 218 and 232 flanking the NNP1 motif | 129 |
| Figure 4.13 | 1D $^{19}F$ fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 235 and 248 flanking the NNP2 motif | 130 |
| Figure 4.14 | 1D $^{19}F$ fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 252 and 264 flanking the NNP3 motif | 131 |
| Figure 4.15 | 1D $^{19}F$ fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 272, 281, 297 and 307 | 132 |
| Figure 4.16 | 1D $^{19}F$ fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S218C and S232C, in the presence of Cytb5 | 134 |
| Figure 4.17 | 1D $^{19}F$ fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S235C and S248C, in the presence of Cytb5 | 135 |
| Figure 4.18 | 1D $^{19}F$ fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S252C and S264C, in the presence of Cytb5 | 136 |
| Figure 4.19 | 1D $^{19}F$ fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S272C, A281C, S297C and S307C in the presence of Cytb5 | 138 |
| Figure 5.1 | Expression of the human intrinsic proteasomal receptor Rpn13 | 142 |
| Figure 5.2 | Preparation of recombinant human $Rpn13_{260-407}$ and SGTA_TPR | 143 |
| Figure 5.3 | Size exclusion chromatography of SGTA_TPR and $Rpn13_{260-407}$ | 144 |

| | | |
|---|---|---|
| Figure 5.4 | Complex formation between $Rpn13_{260-407}$ and SGTA_TPR as observed by SEC | 145 |
| Figure 5.5 | Isothermal titration calorimetry (ITC) data showing binding of SGTA_TPR to $Rpn13_{260-407}$ | 146 |
| Figure 5.6 | $^{1}H$-$^{15}N$ HSQC spectra of $^{15}N$-labelled SGTA_TPR overlaid at different titration points with unlabelled $Rpn13_{260-407}$ | 148 |
| Figure 5.7 | Chemical shift perturbation (CSP) delta average values of SGTA_TPR amide signals upon binding to $Rpn13_{260-407}$. | 149 |
| Figure 5.8 | Chemical shift perturbation (CSP) analysis of the SGTA_TPR/ $Rpn13_{260-407}$ interaction | 150 |
| Figure 5.9 | $^{1}H$-$^{15}N$ HSQC spectra of $^{15}N$-labelled $Rpn13_{260-407}$ overlaid at different titration points with unlabelled SGTA_TPR | 152 |
| Figure 5.10- | Chemical shift perturbation (CSP) values of $Rpn13_{260-407}$ amide signals upon binding to SGTA_TPR | 153 |
| Figure 5.11 | Normalised CSP data of the most perturbed $Rpn13_{260-407}$ residues (405, 406, 407) upon titration with different concentrations of SGTA_TPR | 154 |
| Figure 5.12 | $^{1}H$-$^{15}N$ heteronuclear NOE enhancement values of Rpn13 C-terminal domain in its free, and SGTA_TPR bound, states. | 155 |
| Figure 5.13 | Preparation of recombinant human $Rpn13_{260-402}$ | 157 |
| Figure 5.14 | SEC analysis shows abrogation of the SGTA_TPR-Rpn13 interaction in the absence of the Rpn13 extreme C-terminal DMSLD pentapeptide | 158 |
| Figure 5.15 | Rpn13 devoid of its extreme C-terminal residues 403-407 cannot interact with SGTA_TPR | 159 |
| Figure 5.16 | Isothermal titration calorimetry (ITC) binding assay showing that $Rpn13_{260-402}$ does not interact with SGTA_TPR | 160 |
| Figure 5.17 | The Rpn13 extreme C-terminal DMSLD pentapeptide is necessary and sufficient for its interaction with SGTA_TPR | 161 |
| Figure 5.18 | ITC binding assay showing interaction of SGTA_TPR to the Rpn13 derived DMSLD pentapeptide | 162 |
| Figure 5.19 | Comparison of $Rpn13_{260-407}$ vs the DMSLD pentapeptide binding to SGTA_TPR | 164 |

| | | |
|---|---|---|
| Figure 5.20 | The conserved two-carboxlate clamp mode of molecular recognition by TPR domains | 166 |
| Figure 5.21 | Preparation of the K160E/R164E double mutant SGTA_TPR and $Rpn13_{260-407}$ to probe the two-carboxylate clamp mechanism of binding | 167 |
| Figure 5.22 | The K160E/R164E double mutant SGTA_TPR disrupts its interaction with $Rpn13_{260-407}$ | 168 |
| Figure 5.23 | $Rpn13_{260-407}$ and K160E/R164E double mutant SGTA_TPR do not interact | 169 |
| Figure 5.24 | Preparation of an Rpn13 C-terminal cysteine mutant for site-directed spin labelling studies | 171 |
| Figure 5.25 | Intermolecular paramagnetic relaxation enhancement (PRE) experiments | 172 |
| Figure 5.26 | PRE induced line-broadening mapped onto crystal structure of SGTA_TPR | 173 |
| Figure 5.27 | SGTA_TPR/DMSLD peptide complex HADDOCK results | 175 |
| Figure 5.28 | SGTA_TPR/ DMSLD peptide complex generated using the HADDOCK approach | 176 |
| Figure 6.1 | A putative SGTA/BAG6 quality control cycle of MLPs operating at the proteasome | 188 |

## List of Tables


| | | |
|---|---|---|
| Table 1.1 | Summary of high-resolution structures of SGTA/Sgt2 domains | 25 |
| Table 2.1 | Primers designed to clone various constructs into a modified pET28 vector encoding an N-terminal thioredoxin fusion protein followed by a hexa-histidine tag and rTEV cleavage site | 62 |
| Table 2.2 | Combinations of primers used to amplify various cDNA fragments to generate different constructs in the modified pET28 vector. | 63 |
| Table 2.3 | Primers designed to generate various point mutants using quick-change site directed mutagenesis. | 66 |
| Table 2.4 | NMR acquisition parameters | 72 |
| Table 5.1 | Energetic parameters obtained for the three best clusters of SGTA_TPR/DMSLD peptide complex calculated using HADDOCK | 174 |

## Abbreviations

| | |
|---|---|
| BAG6 | BCL2 associated athanogene 6 |
| BCA | bicinchoninic acid |
| BMRB | biological magnetic resonance bank |
| BSA | bovine serum albumin |
| BTFA | 3-bromo-1,1,1-trifluoroacetone |
| CAML | calcium-modulating cyclophilin ligand |
| CcpNMR | Collaborative computational project for NMR |
| CD | circular dichroism |
| CPMG | Carr-Purcell-Meiboom-Gill pulse sequence |
| Cryo-EM | electron cryomicroscopy |
| CSI | Chemical shift index |
| CSP | chemical shift perturbation |
| C-terminus | carboxy-terminus |
| Cytb5 | cytochrome b5 |
| DDM | n-Dodecyl-β-D-Maltoside |
| DEUBAD | deubiquitinase adaptor |
| dNTP | deoxyribonucleotide triphosphate |
| *E. coli* | *Escherichia coli* |
| EDTA | ethylenediaminetetraacetic acid |
| ER | endoplasmic reticulum |
| ERAD | endoplasmic reticulum associated degradation |
| FPLC | fast protein liquid chromatography |
| Get | guided entry of tail anchored membrane proteins |
| HADDOCK | high ambiguity driven biomolecular docking |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| $His_6$ | hexa-histidine tag |
| HIV-1 | human immunodeficiency virus type 1 |
| HOP | Hsp70/Hsp90 organising protein |
| HPLC | high performance liquid chromatography |
| Hsp | heat-shock protein |
| HSQC | heteronuclear single quantum coherence spectroscopy |
| IMAC | immobilized metal ion affinity chromatography |
| INEPT | insensitive nuclei enhanced by polarization transfer |

| | |
|---|---|
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| ITC | isothermal titration calorimetry |
| kDa | kilo Dalton |
| $K_d$ | dissociation constant |
| KPi | potassium phosphate |
| LB | Luria-Bertani broth |
| MCS | multiple cloning site |
| MLP | mislocalised membrane protein |
| MS | mass spectrometry |
| MTSL | 1-oxyl-2,2,5,5-tetramethylpyrroline-3- methyl methanethiosulfonate |
| dMTSL | 1-acetyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methylmethanethiosulfonate |
| NEB | New England Biolabs |
| Ni-NTA | nickel-nitrilotriacetic acid |
| NMR | nuclear magnetic resonance |
| NOE | nuclear Overhauser effect |
| N-terminus | amino-terminus |
| $OD_{600}$ | optical density at 600 nm |
| PCR | polymerase chain reaction |
| PDB | protein data bank |
| PMSF | phenylmethylsulfonyl fluoride |
| PRE | paramagnetic relaxation enhancement |
| Pru | pleckstrin-like receptor for ubiquitin |
| Q-rich | glutamine rich |
| QNP | Quattro-nucleus probe |
| rTEV | recombinant tobacco etch virus protease |
| SARS | severe acute respiratory syndrome coronavirus |
| SAXS | small angle X-ray scattering |
| SDS-PAGE | sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| SEC | size exclusion chromatography |
| SGTA | small glutamine rich tetratricopeptide repeat containing protein alpha |
| SGTA_TPR | TPR domain of SGTA |
| SOC | super optimal broth with catabolite repression |
| SRP | signal recognition particle |
| Syb2 | synaptobrevin-2 |
| TA | tail-anchored membrane protein |

| | |
|---|---|
| TAE | Tris-acetate EDTA |
| TCEP | Tris (2-carboxyethyl) phosphine hydrochloride |
| TEMED | N, N, N, N- tetramethylethylenediamine |
| TMD | transmembrane domain |
| TPR | tetratricopeptide repeat |
| TRC | transmembrane domain recognition complex pathway |
| Tris | tris (hydroxymethyl) aminomethane |
| TxA | thioredoxin A |
| UBL | ubiquitin-like domain |
| UBD | ubiquitin binding domain |
| UPS | ubiquitin-proteasome system |
| WRB | tryptophan-rich basic protein |

# Chapter 1

# Introduction

## 1.1 Protein homeostasis and macromolecular quality control

'Homeostasis' refers to the ability of living organisms to maintain a constant internal environment, a term devised by Walter Cannon to describe the physiological processes responsible for the maintenance of consistent internal chemistry (such as pH, ionic strength, electrolyte composition, temperature, and others) required to sustain life (Cannon, 1939). The concept of homeostasis dates back to the French physiologist Claude Bernard's theory of the '*milieu intérieur*', which states 'The constancy of the internal environment is the condition of a free and independent existence' (Bernard, 1878). This concept has been extended to proteins, the workhorses of a cell. Nascent polypeptides synthesized by the ribosome fold into precise three-dimensional structures to carry out their prescribed functions, and importantly must maintain their structure for the duration of their functional lifetime. However, it must be noted that not all proteins require a precise three-dimensional fold to function, such is the case for intrinsically disordered proteins and molten globules (Dyson and Wright, 2005; Redfield, 1999). Thus cellular 'protein homeostasis' is effected by an integrated network of molecular chaperones and proteases, together with components of the autophagy system, that maintain quality control over protein structure and facilitate the removal of terminally misfolded and aggregated proteins (Hartl *et al.*, 2011; Wickner *et al.*, 1999).

For many small proteins the information contained in their amino-acid sequence is sufficient to attain their correct three-dimensional structure *in vitro* (Anfinsen, 1973; Dobson *et al.*, 1998). However, in a cellular context many proteins require the assistance of molecular chaperones for proper folding to occur on biologically relevant timescales (Hartl, 1996). The term 'molecular chaperone' was coined by Ron Laskey to describe the ability of the nucleoplasmin protein to prevent histone aggregation with DNA during nucleosome assembly (Laskey *et al.*, 1978), and some years later was used

to describe cellular proteins that mediated the folding and assembly of other polypeptide chains (Ellis, 1987). Thus molecular chaperones promote protein folding, and can be defined as 'any protein that interacts with, stabilizes or helps another protein to acquire its functionally active conformation, without being present in its final structure' (Hartl *et al.*, 2011; Hartl, 1996). As protein concentrations in human cells can range from 50 to 300 mg/ml (Wolff *et al.*, 2014), it is unsurprising that a myriad of chaperone networks have evolved to reduce aggregation and maintain proteins in functionally active conformations in such crowded cellular environments.

Molecular chaperones belong to a family of proteins referred to as heat-shock proteins (Hsps). Hsps are present in all domains of life, first appearing in primitive hyperthermophilic archaea, in which their role was to exert cytoprotective effects during conditions of elevated temperatures (therefore called 'heat-shock' proteins). As the family diverged into more developed species, its role has been modified into that of a stress-protein which protects cells from physiological and environmental stress, during which their expression is upregulated (Haslbeck *et al.*, 2005). These chaperones typically recognize surface-exposed hydrophobic regions exposed by non-native proteins, such as folding intermediates and misfolded proteins, frustrating their aggregation and preventing them from degradation. Chaperones such as the Hsp70s, Hsp90s and chaperonins (Hsp60s) promote folding of client proteins via ATP dependent cycles of binding and release (Hartl *et al.*, 2011). On the other hand, ATP independent molecular chaperones such as the small heat shock proteins (sHsps) act as 'holdases' preventing client aggregation before handover to ATP dependent chaperones can take place (Clark *et al.*, 2012).

When molecular chaperones fail to rescue clients by promoting their biologically active conformations, energy-dependent proteases eliminate such irreversibly damaged proteins via the ubiquitin-proteasome system (UPS) (Komander and Rape, 2012). It is important to consider that the UPS cannot limit its remit to the recognition of terminally misfolded proteins, as it also has to eliminate folding intermediates that fail to achieve their native conformations (Rodrigo-Brenni and Hegde, 2012). Consequently, target selection during protein quality control presents a formidable challenge to cellular proteostasis networks, and the kinetics of partitioning between chaperones and the UPS machinery has been suggested to play a role in deciding the fate of client proteins (Shao and Hegde, 2016; Wickner *et al.*, 1999).

When cellular quality control options fail, this subsequently leads to the accumulation of intracellular aggregates, a process that has been associated with many disease phenotypes (Powers *et al.*, 2009). Diseases associated with the failure of cellular protein quality control include neurodegeneration, cardiovascular disease, lung disease, metabolic disorders, and certain cancers (Powers *et al.*, 2009; Chiti and Dobson, 2006). In particular, neurodegenerative phenotypes appear to be closely linked to failures in the cellular quality control machinery, possibly due to the fact that neurons are long-lived cells and are more susceptible to the accumulation of aberrant proteins over time (Shao and Hegde, 2016).

## 1.2 The proteostasis network

The cellular proteostasis network (Fig. 1.1) is an extensive network of quality control factors consisting of around 200 chaperones and co-chaperones, and around 600 UPS and autophagy components. Chaperones exert their effects in maintaining proteostasis by promoting protein folding in collaboration with co-chaperones (such as AHA1 and

SGTA), and other regulatory cofactors such as the BCL2-associated athanogene (BAG) family of proteins (Hartl *et al.*, 2011).

On the other hand, irreversibly misfolded proteins are subjected to proteolytic degradation via the UPS system, which involves the covalent attachment of ubiquitin molecules to lysine residues on substrates via an isopeptide bond, which marks substrates for proteasome degradation (Komander and Rape, 2012). The 76-residue modifier protein ubiquitin is a highly stable protein that presents a compact β-grasp fold (Vijay-Kumar *et al.*, 1987). The attachment of ubiquitin to proteins occurs through a cascade of sequential reactions catalysed by ubiquitin-activating (E1), ubiquitin-conjugating (E2), and ubiquitin ligase (E3) enzymes. This E1/E2/E3 cascade leads to the formation of an isopeptide bond between the C-terminus of ubiquitin and a substrate lysine. Ubiquitin itself contains seven lysine residues which results in the formation of polyubiquitin chains. In particular, K48- and K11-linked polyubiquitin chains mark substrate proteins for proteasomal degradation (Komander and Rape, 2012).

Whereas the degradation of misfolded protein by the UPS requires substrate ubiquitination together with mechanisms to keep clients soluble prior to elimination, it has been observed that terminally aggregated proteins can also be sequestered into perivacuolar inclusions for autophagy (Kaganovich *et al.*, 2008). Autophagy is a major intracellular mechanism that facilitates the degradation of cytosolic components delivered to the lysosome (Mizushima and Komatsu, 2011). These aggregates are sequestered at sites close to the microtubule-organizing centre, and have been referred to as aggresomes, which are suggested to be less toxic to the cell (Hartl *et al.*, 2011; Kopito, 2000). The proteostasis network thus ensures normal cellular physiology through its contributions to the maintenance of a functional proteome.

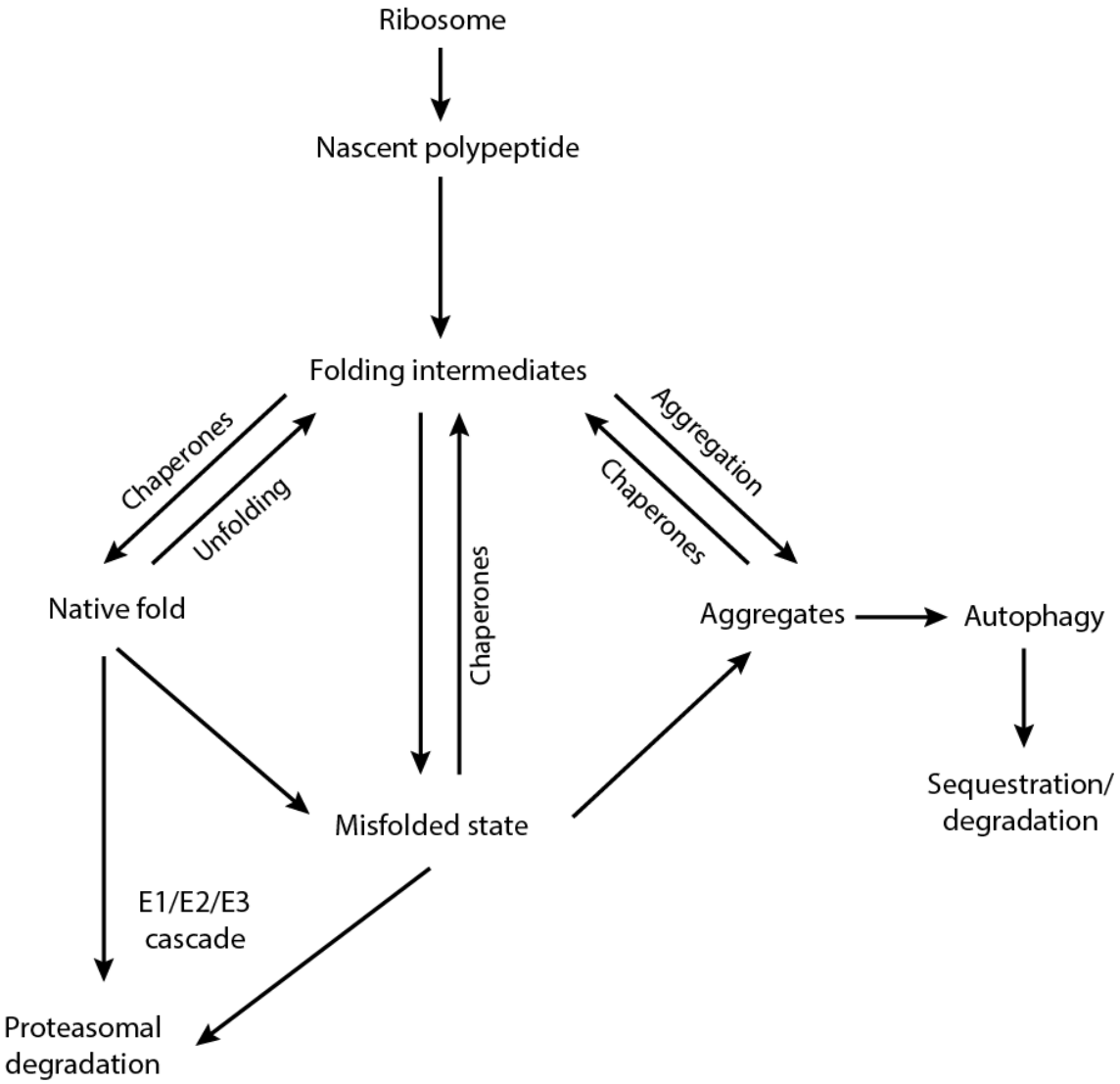



**Figure 1.1- The proteostasis network.** The proteostasis network consists of an extensive network of chaperones, along with the cellular machinery to degrade or sequester terminally misfolded proteins, thereby preventing the accumulation of toxic aggregates in the cytosol. In mammalian cells, there are around 180 different chaperone components, around 600 components of the ubiquitin-proteasome system (UPS) (including the E1/E2/E3 cascade of enzymes) that effect proteasomal degradation, and around 30 components of the autophagy system (Hartl *et al.*, 2011). The co-chaperone SGTA prevents aggregation by maintaining client proteins in a soluble state within the cytosol.

## 1.3 Small glutamine-rich, tetratricopeptide repeat-containing protein alpha (SGTA)

The small glutamine-rich, tetratricopeptide repeat-containing protein alpha (SGTA) is a co-chaperone involved in a specialized branch of the global cellular proteostasis network that decides the fate of secretory and membrane proteins that mislocalise to the cytosol (referred to as MLPs), facilitates the biogenesis of tail-anchored (TA) membrane proteins, and influences ER- associated degradation (ERAD) (Casson *et al.*, 2016; Wunderley *et al.*, 2014; Xu *et al.*, 2012; Hegde and Keenan, 2011). The ability of SGTA to modulate various protein quality control pathways in the cytosol is primarily due to its recognition of surface exposed regions of hydrophobicity (termed 'degrons') on newly synthesized and mislocalised proteins. As a result, SGTA shields these degrons from the surrounding aqueous cytosol, thereby preventing protein misfolding, aggregation, and the formation of aberrant protein-protein interactions. Upon stabilisation, these proteins can have different fates, such as being folded to their native state by molecular chaperones followed by targeting to their final correct destinations, or being marked for degradation by the ubiquitin-proteasome system (UPS). (Casson *et al.*, 2016, Roberts *et al.*, 2015).

SGTA was first identified in complex with viral proteins (Callahan *et al.*, 1998; Cziepluch *et al.*, 1998), and is ubiquitously expressed across tissue types (Philp *et al.*, 2016). In addition to its roles in effecting macromolecular quality control and TA protein biogenesis, SGTA has been implicated in a wide-range of biological processes such as cell-division and apoptosis (Winnefeld *et al.*, 2006), synaptic neurotransmitter release (Tobaben *et al.*, 2001), viral lifecycles (Fielding *et al.*, 2006; Handley *et al.*, 2001; Callahan *et al.*, 1998), and in the regulation of hormone receptor signalling (Philp

*et al.*, 2013; Buchanan *et al.*, 2007). Thus SGTA is emerging as a co-chaperone of importance in diverse pathophysiological contexts underlying health and disease.

## 1.4 Structural studies of SGTA

SGTA is a 34 kDa protein that assembles as a homodimer, with each protomer consisting of three domains: a highly conserved N-terminal dimerization domain, followed a central tetratricopeptide (TPR) domain, and a C-terminal domain which contains three NNP repeat motifs and a glutamine rich region (Figs. 1.2 and 1.3) (Darby *et al.*, 2014; Simon *et al.*, 2013; Dutta and Tan, 2008; Liou and Wang, 2005). High resolution structural studies of SGTA, and its yeast equivalent Sgt2, have thus far focused on its N-terminal and TPR domains (Table 1.1). SGTA's hydrophobic substrate binding capability, pertinent to its function in various quality control related scenarios, is mediated by its C-terminal domain. Hydrophobic substrates specifically recognised by the SGTA C-terminal domain in the cytosol include MLPs, the transmembrane domain (TMD) helices of newly synthesised TA-proteins, and aberrant membrane protein precursors (Casson *et al.*, 2016; Hegde and Keenan, 2011). However, structural details of SGTA's substrate-binding C-terminal domain remain poorly understood. Furthermore, as structural studies carried out to date have focused solely on isolated domains, a detailed picture of SGTA domain organisation and conformations in the context of the full-length protein is yet to be elucidated.

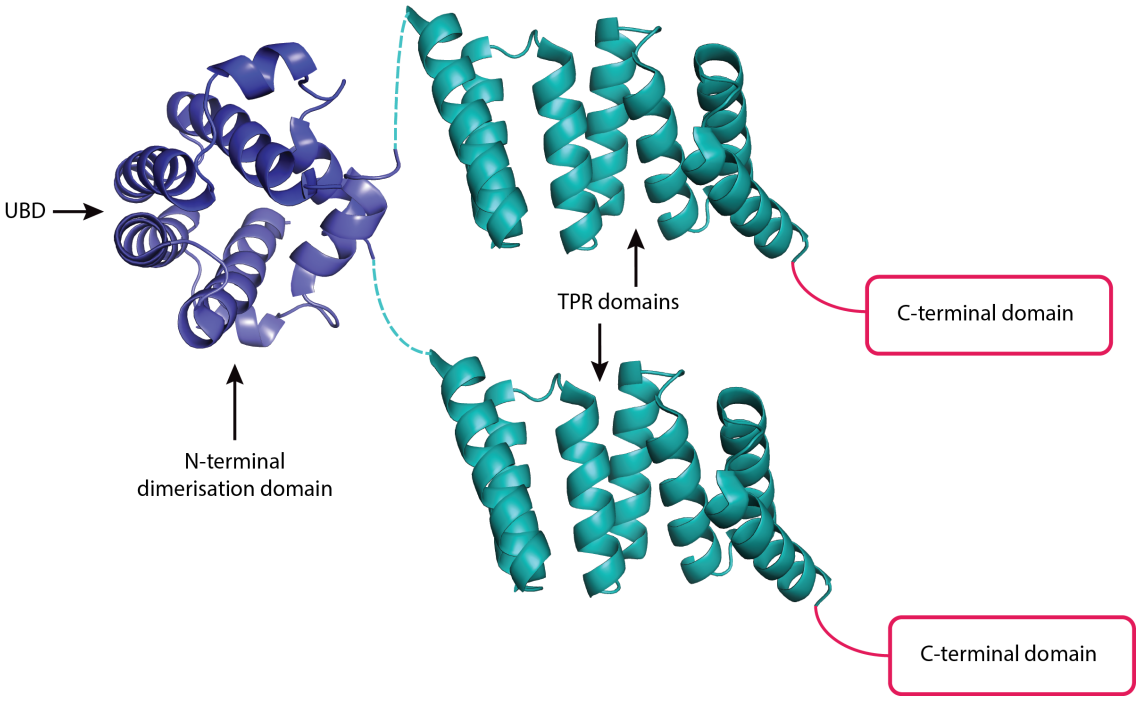



**Figure 1.2 – Schematic of full-length SGTA.** The solution NMR structure of the N-terminal domain of SGTA (PDB accession code: 4CPG) is shown as a dimer with the position of its UBL binding domain (UBD) indicated. The crystal structure of the central TPR domains (PDB accession code: 2VYI) are shown separated with the N-terminal dimer by flexible linkers (dotted lines). Structural details of the C-terminal substrate-binding domain are yet to be determined.

**Table 1.1 - Summary of high-resolution structures of SGTA/Sgt2 domains**

| | Domain | Species | Residues | Technique | PDB |
|---|---|---|---|---|---|
| SGTA | N-terminal (dimer) | *Homo sapiens* | 3-54 | X-ray crystallography | 4GOD |
| | | | 1-69 | Solution NMR spectroscopy | 4CPG |
| | TPR | *Homo sapiens* | 80-210 | X-ray crystallography | 2VYI |
| Sgt2 | N-terminal (dimer) | *Saccharomyces cerevisiae* | 1-78 | Solution NMR spectroscopy | 4ASV |
| | | *Saccharomyces cerevisiae* | 2-72 | Solution NMR spectroscopy | 2LXB |
| | TPR | *Aspergillus fumigatus* | 104-254 | X-ray crystallography | 3SZ7 |

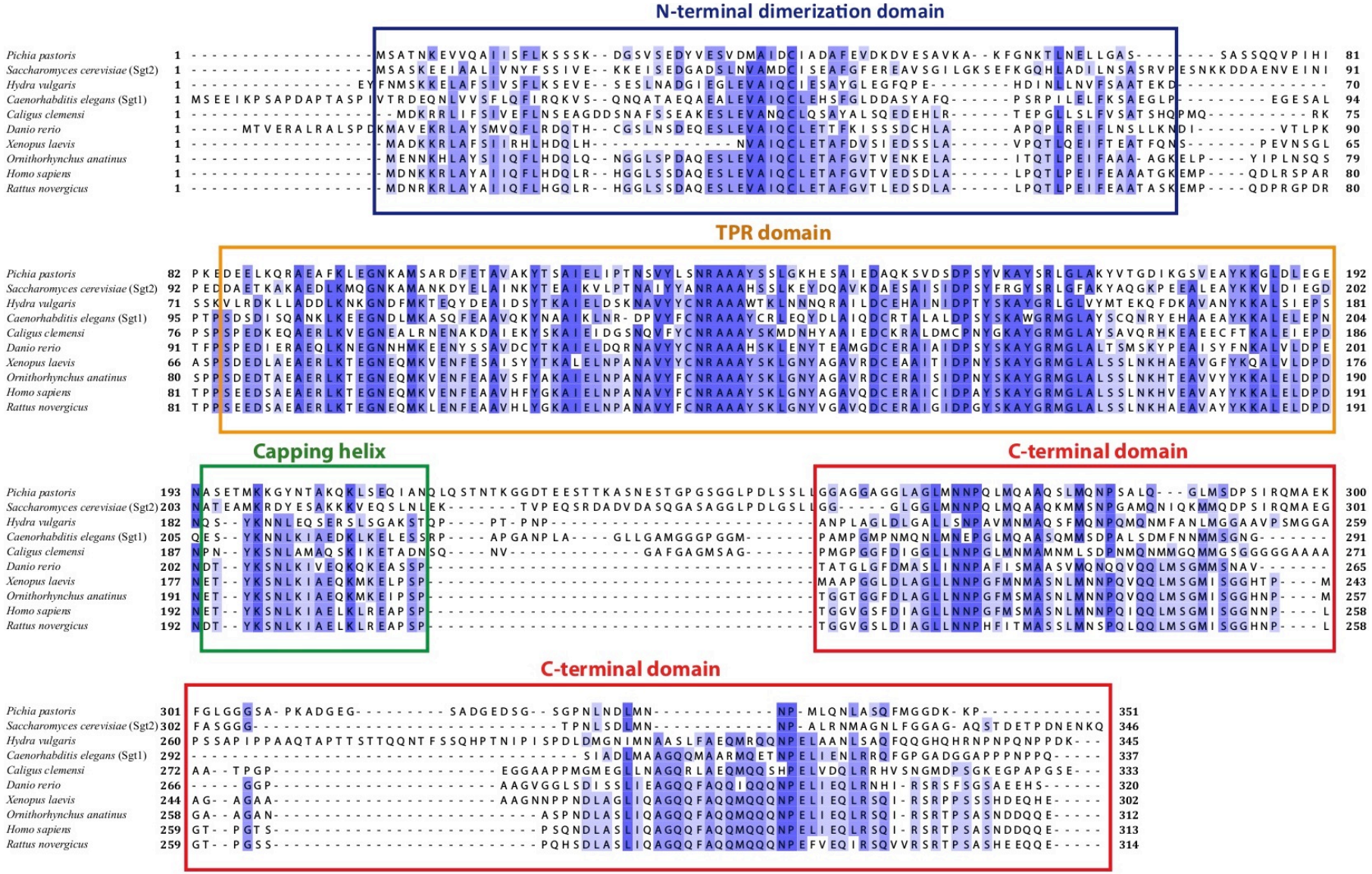



**Figure 1.3– SGTA sequence alignment** (*legend on next page*).

**Figure 1.3– SGTA sequence alignment**. Sequences of SGTA and its homologs across the phyla have been aligned using Clustal Omega multiple sequence alignment. Domains have been annotated with reference to the human SGTA sequence. *C. elegans* and *D. rerio* have longer N-terminal extensions compared to both mammalian and yeast equivalents. The central TPR repeats appear to be the most conserved, however the TPR capping helix (helix 7) is much less conserved in fungal species when compared to metazoans. A glutamine-rich region, present in the C-terminal domain, is more conserved in metazoans. The presence of at least one NNP motif within the C-terminal domain is characteristic across all phyla, with duplication of such motifs occurring in higher metazoans. This sequence alignment was generated using Jalview (Waterhouse *et al.*, 2009).

### *1.4.1 N-terminal domain*

The N-terminal domain of SGTA exists as a homodimer in solution, with each protomer consisting of four α-helices connected by short loops. These protomers arrange into a unique helical fold (Fig. 1.4). The dimeric interface is stabilised by hydrophobic residues that form a tight interaction, reminiscent of the core of a globular protein (Darby *et al.*, 2014). This characteristic N-terminal fold is conserved in the yeast homolog, Sgt2 (Simon *et al.*, 2013). Furthermore, this N-terminal domain provides SGTA with a direct physical link to different subunits of the BAG6 complex, which underlies SGTA's functionality in various cellular quality control pathways. Moreover, this is achieved through the SGTA N-terminal domain's ability to interact with either of the two ubiquitin-like (UBL) domains present in the BAG6 complex (described in Section 1.5) (Darby *et al.*, 2014; Simon *et al.*, 2013).

The ubiquitin-binding domain (UBD) of SGTA is formed by a negatively charged surface on its N-terminal domain dimer. This surface arises as a result of side-chain contributions from charged negatively charged residues that form the second alpha helix ($\alpha_2$) (Asp27, Glu30, Glu33, and Glu40) of each protomer within the dimer. As a result, each N-terminal dimer can recognize a single UBL domain (Figs. 1.4 and 1.5). This is

because UBLs characteristically exhibit a positively charged region (of up to 5 arginine or lysine residues), thus binding is mediated by electrostatic interactions (Darby *et al.*, 2014).

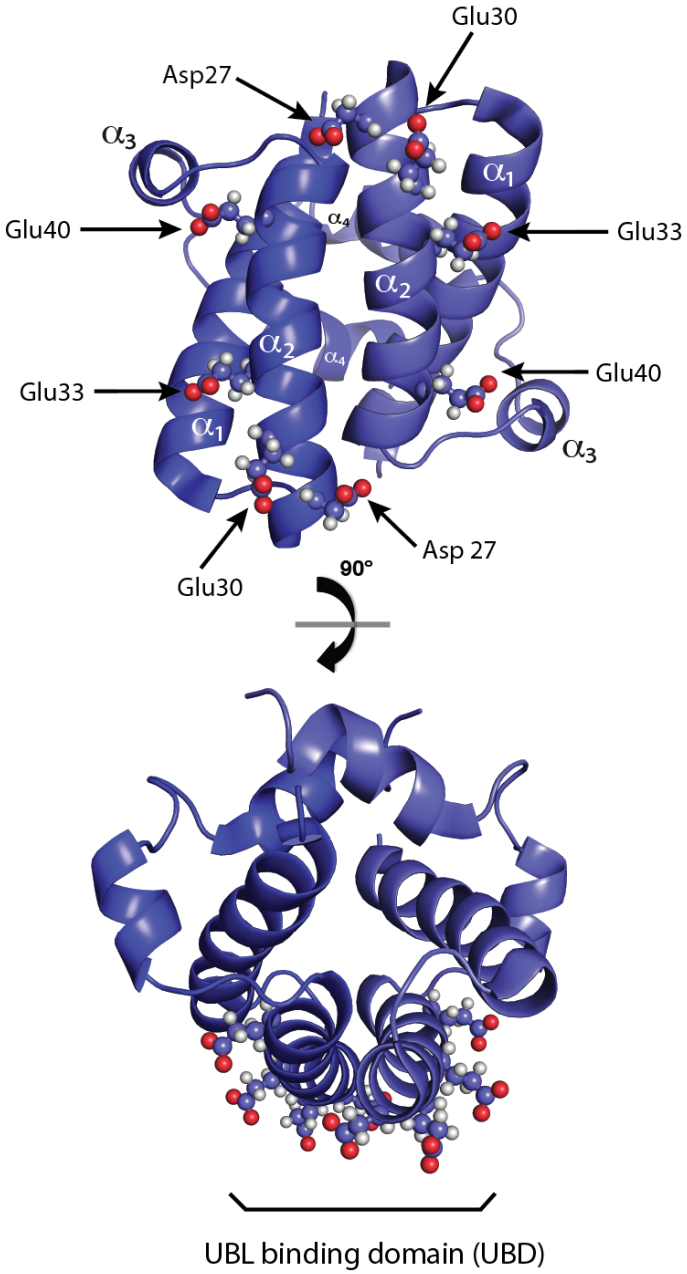



**Figure 1.4– SGTA N-terminal domain highlighting its UBL-binding domain (UBD).** The solution NMR structure of the N-terminal domain dimer of SGTA (PDB accession code: 4CPG) is shown as orthogonal views. The dimer has a negatively charged surface with contributions from side-chains of Asp27, Glu30, Glu33, and Glu40 that form the UBD (shown in ball-and-stick format). This region facilitates the N-terminal domain dimer's interaction with UBLs (from BAG6 and Ubl4a) mediated by electrostatic interactions (Darby *et al.*, 2014).

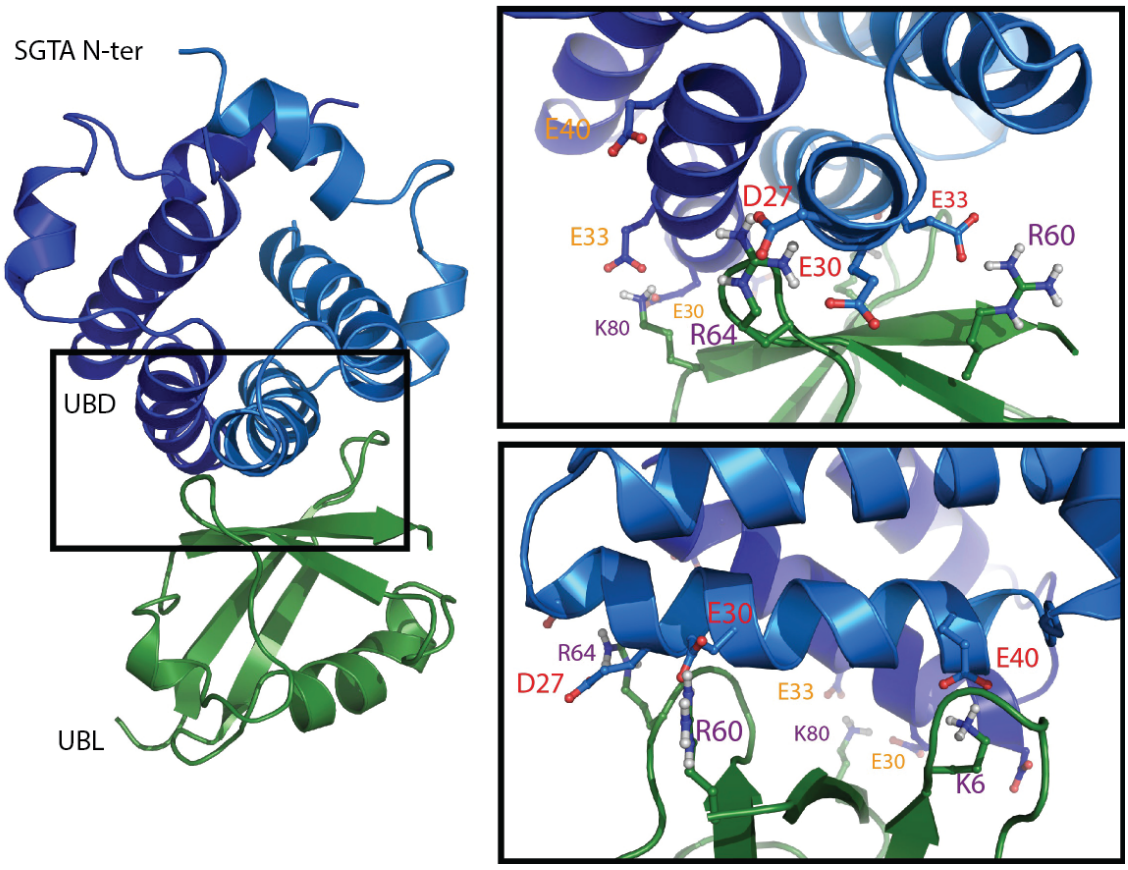



**Figure 1.5- N-terminal domain of SGTA in complex with a cognate UBL.** The UBD of the N-terminal dimer of SGTA (*in blue*) shown in complex with a cognate UBL domain (*in green*). Charged residues facilitating this electrostatic interaction at the binding interface shown as sticks. Inset shows two orthogonal views of the N-terminal dimer/UBL binding interface (PDB accession code: 4CPG) (Darby *et al.*, 2014).

### *1.4.2 TPR domain*

The central TPR domain of SGTA (hereafter referred to SGTA_TPR) consists of three structurally identical helix-turn-helix motifs, in which each pair of α-helices folds in an antiparallel fashion. These TPR motifs are followed by a C-terminal capping helix (α7) that packs against the second helix of the third motif (Fig. 1.6A). These seven helices arrange into a right-handed superhelical structure to form a concave surface lined by helices $\alpha_1$, $\alpha_3$, $\alpha_5$, and $\alpha_7$ (Dutta and Tan, 2008). In

addition, this concave surface forms a groove that consists of conserved basic residues known to facilitate protein-protein interactions (Fig. 1.6B).

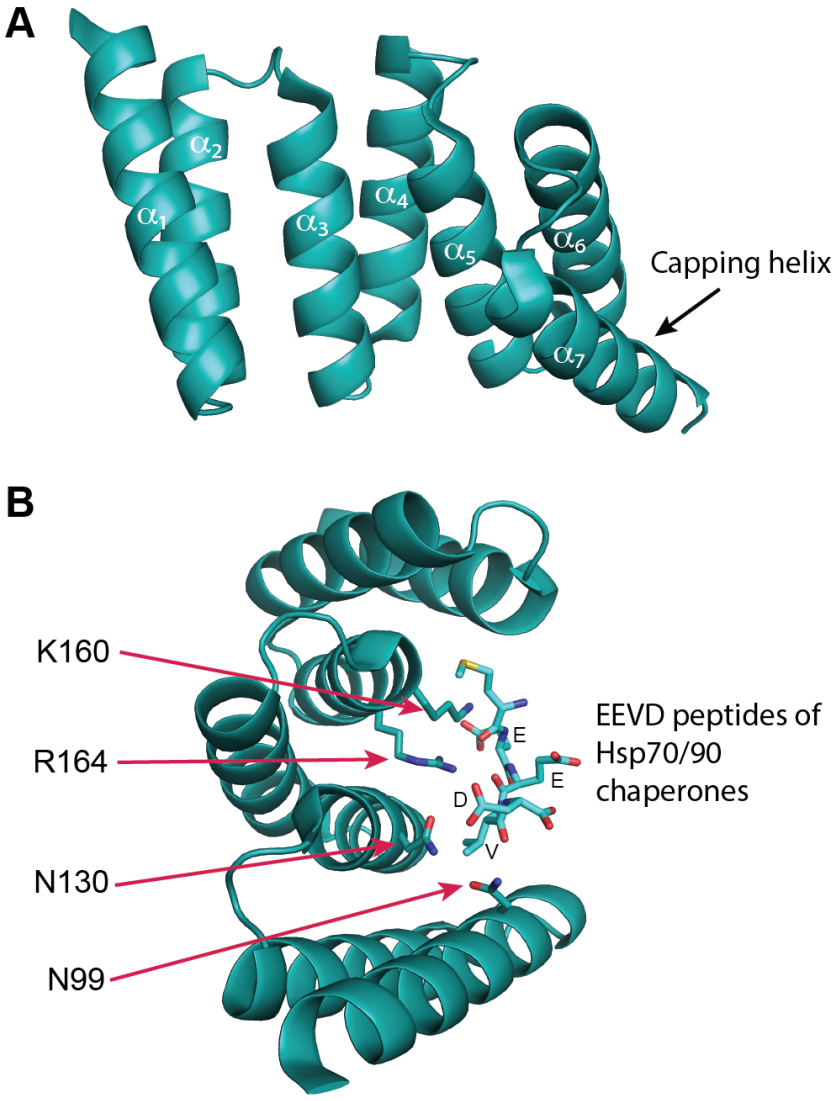



**Figure 1.6- The TPR domain of SGTA.** (A) Crystal structure of the SGTA_TPR domain (PDB accession code: 2VYI), consisting of three helix-turn-helix repeats ($\alpha_1/\alpha_2$, $\alpha_3/\alpha_4$, $\alpha_5/\alpha_6$) followed by a capping helix ($\alpha_7$). (B) Basic residues in the TPR groove present on its concave surface (shown as sticks) provides a platform for protein-protein interactions, such as for the EEVD motifs of the Hsp70/90 molecular chaperones (shown as sticks) (Dutta and Tan, 2008).

The fold of human SGTA_TPR is highly similar to that of TPR domains present in other co-chaperones such as CHIP (the carboxyl-terminus of Hsp70 interacting protein) and HOP (Hsp70/90 organizing protein) (Dutta and Tan, 2008). Both CHIP and HOP TPR domains interact with C-terminal EEVD motifs of Hsp70 and Hsp90 molecular chaperones via a 'two-carboxylate clamp' mechanism of molecular recognition (Scheufler *et al.*, 2000; Zhang *et al.*, 2005). This mechanism is mediated by the conserved basic residues present in the TPR groove, that form a network of interactions to 'clamp' the two free C-terminal carboxylates present on the terminal aspartate of a Hsp70/90 EEVD motif (Scheufler *et al.*, 2000). An overlay of crystal structures of SGTA_TPR with TPR domains of CHIP and HOP bound to EEVD peptides reveal that TPR residues essential to mediate such a 'two-carboxylate clamp' mode of molecular recognition are also present in SGTA_TPR (Dutta and Tan, 2008). This is indicative of human SGTA's ability to associate with Hsp70/90 chaperones in a similar fashion as described for CHIP and HOP TPR domains (Fig 1.6B). Additionally, various independent studies have confirmed that SGTA_TPR interacts with Hsp70 and Hsp90 chaperones, and these studies suggest that binding occurs via the chaperone EEVD motifs (Worrall *et al*., 2008; Liou and Wang, 2005; Liu *et al.*, 1999). Therefore, it appears that SGTA_TPR provides a platform whereby it can interact with ATP-dependent molecular chaperones, and this could provide SGTA bound substrates with access to rescue pathways enabling their acquisition of a native conformation.

### *1.4.3 C-terminal domain*

SGTA's function in quality control relies on its ability to recognise surface exposed regions of hydrophobicity presented by its substrates. This ability of SGTA to recognise hydrophobic degrons has been attributed to its C-terminal domain. The C-terminal domain of human SGTA consists of three NNP motifs, with at least the first NNP motif

conserved in the equivalent region of yeast Sgt2 (Fig. 1.3). NNP motifs have been proposed to support physical interactions within protein complexes (Casey *et al.*, 2011), strongly indicative of a functional role in SGTA. The C-terminal domain of SGTA also consists of a glutamine-rich region, which comprises a stretch of around 39 amino acids interspersed with 12 glutamine residues (Fig. 1.3) (Liou and Wang, 2005).

In agreement with its role in quality control, the SGTA C-terminal domain has been shown to interact with polypeptide fragments containing six or more consecutive hydrophobic residues, and with *in vitro* translated integral membrane protein of the rat type 1 glucose transporter (Liou and Wang, 2005). It is clear that the C-terminal domain is vital to SGTA's function in triaging stray hydrophobic substrates in the cytosol. However, structural details describing this region of SGTA remain elusive. As a result, there are many outstanding questions regarding the nature and extent of this key substrate-binding domain of SGTA. These include mechanisms underlying the binding and release of different classes of mislocalised substrates and aberrant membrane protein precursors, mechanisms of TA-protein recognition and sorting, and details of SGTA's participation in the ERAD pathway.

## 1.5 The heterotrimeric BAG6 complex

SGTA acts in concert with the heterotrimeric BAG6 complex to accomplish its roles in membrane protein quality control and in the biogenesis of tail-anchored proteins (Casson *et al.*, 2016). The heterotrimeric BAG6 complex consists of the BAG6/BAT3/Scythe (BCL2-associated athanogene), TRC35 (transmembrane domain recognition complex 35) and Ubl4a (ubiquitin-like protein 4a) proteins. The BAG6 protein itself consists of an N-terminal UBL domain, followed by a central proline rich region and a C-terminal BAG-similar domain (Fig. 1.7) (Mock *et al.*, 2015; Mariappan *et al.*, 2010). Based on our current understanding of BAG6, it is likely that its central

proline rich region interacts with hydrophobic substrates (Casson *et al.*, 2016). Furthermore, it is known that the N-terminal UBL domain of the BAG6 protein can recruit quality control effectors such as SGTA (Darby *et al.*, 2014) and the E3 ligase RNF126 (Krysztofinska *et al.*, 2016; Rodrigo-Brenni *et al.*, 2014) to act on its substrates. Furthermore, its C-terminal BAG-similar domain has distinct binding sites for the Ubl4a and TRC35 subunits of the heterotrimeric BAG6 complex, and promotes TA protein biogenesis (see Section 1.7) by effecting the transfer of precursor proteins to the ER targeting factor TRC40 (Mariappan *et al.*, 2010).

Structurally, the N-terminal domain of the BAG6 protein is a type II UBL domain. These UBL domains are highly similar to ubiquitin and are present as functional domains within proteins (Darby *et al.*, 2014). The C-terminus of BAG6 is a BAG-similar domain, and interacts with the C-terminal region of Ubl4a, therefore is also referred to as a TUGS (tethering Ubl4a to BAG-similar) domain. The BAG-similar domain of BAG6 consists of three α-helices, with a hydrophobic patch that docks against Ubl4a (Fig. 1.8) (Kuwabara *et al.*, 2015). Unlike its N- and C-terminal domains, structural details with regard to the substrate-binding central proline-rich region of BAG6 remain to be elucidated.

Functionally, the BAG6 complex was identified as an essential component involved in the chaperoning of TA proteins in the cytosol (Mariappan *et al.*, 2010). It was subsequently implicated in the degradation of MLPs, whereby the BAG6 complex recognises regions of exposed hydrophobicity on MLPs and triages such clients down a degradative pathway (Hessa *et al.*, 2011). It is now known that the BAG6 complex carries out ubiquitination of MLPs through the actions of the E3 ligase RNF126, marking such clients for proteasomal degradation (Rodrigo-Brenni *et al.*, 2014).

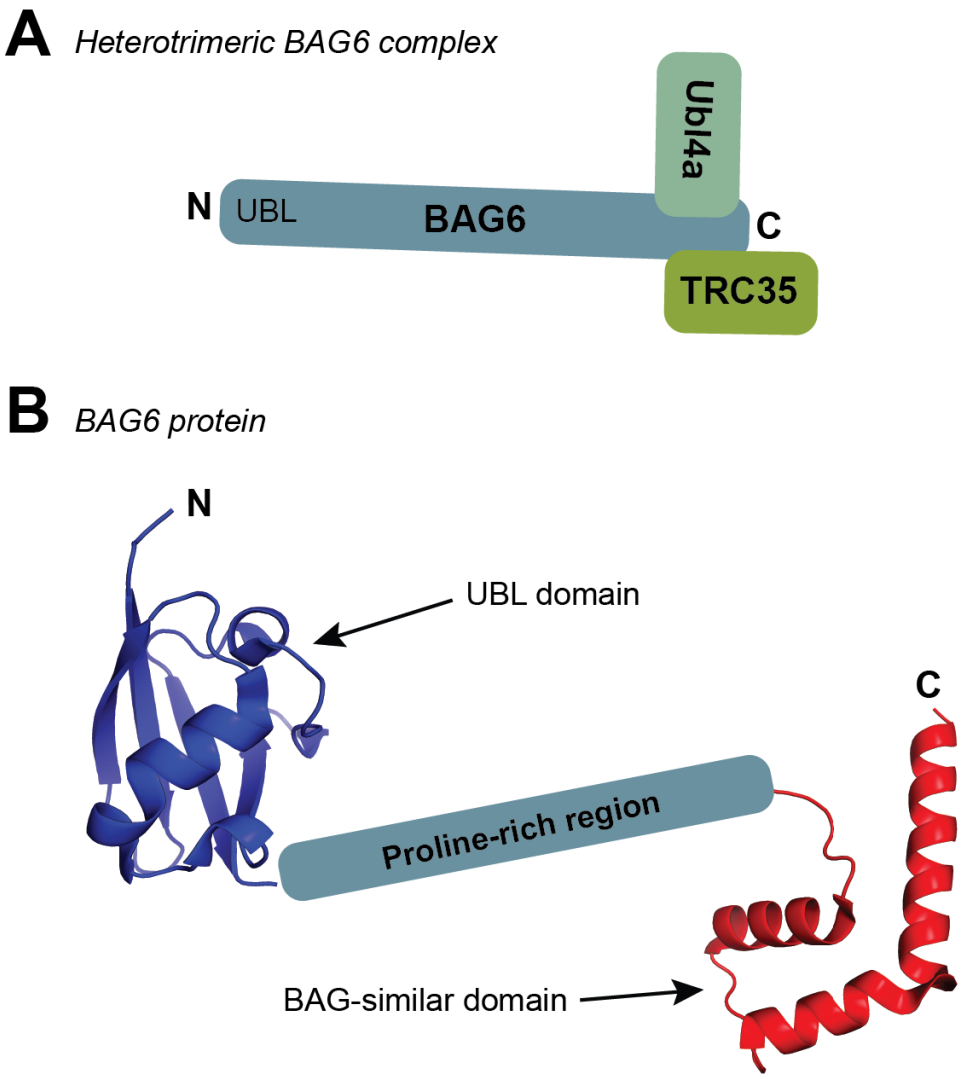



**Figure 1.7- The BAG6 complex.** (A) Schematic of the heterotrimeric BAG6 complex. The TRC35 and Ubl4a proteins interact with the C-terminal domain of the BAG6 protein. (B) BAG6 protein. The crystal structure of the N-terminal UBL domain of BAG6 shown in purple-blue (PDB accession code: 4EEW), this is followed by a central proline-rich region for which high-resolution structural details remain unknown. The C-terminal BAG-similar domain (BAGS) is shown in red (PDB accession code: 4X86) (Kuwabara *et al.*, 2015).

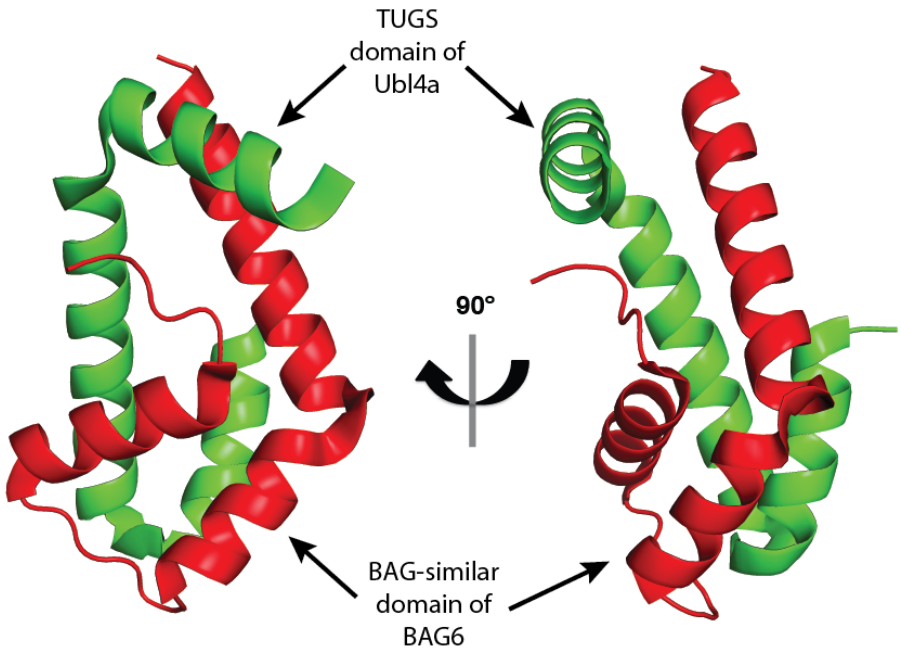



**Figure 1.8- TUGS domain of Ubl4a in complex with the BAG-similar domain of BAG6.** Orthogonal views of the crystal structure of the C-terminal BAG-similar domain of BAG6 (*in red*) in complex with a Ubl4a fragment (*in green*) (PDB accession code: 4X86) (Kuwabara *et al.*, 2015).

The BAG6 complex recruits the E3 ligase RNF126 via its N-terminal UBL domain. The N-terminal region of RNF126 folds into a zinc finger domain in which a zinc cation is coordinated by four cysteine residues. This zinc finger region serves as a platform for its interaction with the UBL domain of the BAG6 protein (Fig. 1.9) (Krysztofinska *et al.*, 2016). As E3 ligases, together with their cognate E2 and E1 enzymes, catalyse the covalent attachment of ubiquitin to their substrates, RNF126 provides a link between the BAG6 complex and the ubiquitin proteasome system (UPS) allowing terminally misfolded substrates bound to BAG6 with access to the proteasome.

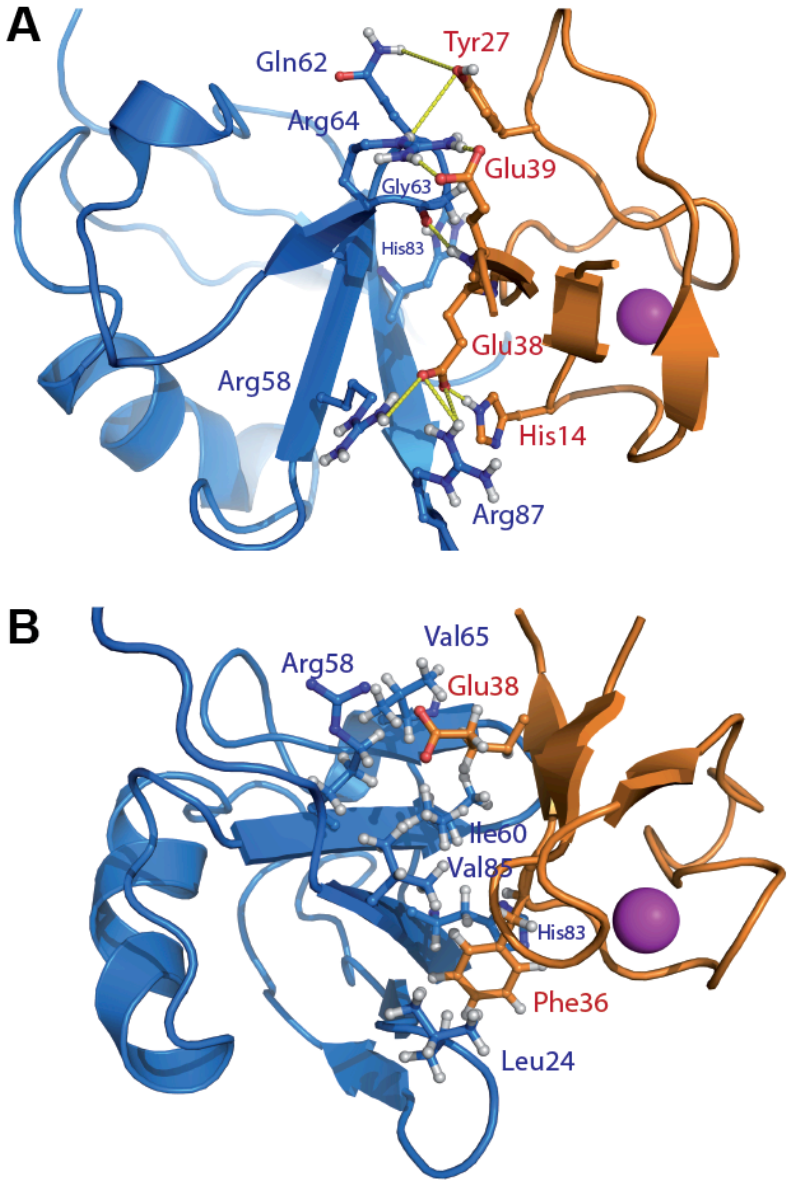



**Figure 1.9- The BAG6 UBL domain in complex with the N-terminal zinc finger domain of the E3 ligase RNF126.** The BAG6 UBL domain is shown in blue, and the N-terminal zinc finger region of RNF126 is shown in orange, both in cartoon representation. (A) Network of hydrogen bonded and electrostatic interactions at the complex interface (B) Hydrophobic interactions driving complex formation (PDB accession code: 2N9P) (Krysztofinska *et al.*, 2016).

## 1.6 The SGTA/BAG6 cycle of MLP quality control

As a result of inefficiencies inherent in the targeting and translocation of newly synthesised proteins to the ER, a certain fraction of these proteins default to the cytosol (Levine *et al.*, 2005). In the cytosol, these mislocalized secretory and membrane proteins (MLPs) expose hydrophobic regions such as signal sequences and transmembrane domains, eventually leading to their aggregation. Furthermore, once these intracellular aggregates form, they can promote the aggregation of other functional proteins in a cell (Chakrabarti and Hegde, 2009). Thus specific quality control pathways have evolved in metazoans that lead to the degradation of MLPs that fail to reach their final cellular destinations, preventing MLP aggregates from accumulating in the cytosol. In particular, SGTA operating in tandem with the BAG6 complex constitutes the SGTA/BAG6 cycle of quality control (Fig 1.10), which determines the fate of stray hydrophobic substrates in the cytosol (Wunderley *et al.*, 2014; Leznicki *et al.*, 2013; Leznicki and High, 2012; Hessa *et al.*, 2011; Mariappan *et al.*, 2010).

Both SGTA and the BAG6 complex interact with a similar set of hydrophobic substrates in the cytosol that include TA proteins and MLPs. Furthermore, in the SGTA/BAG6 cycle, hydrophobic substrates can be transferred between SGTA and BAG6 (Casson *et al.*, 2016). On one hand, SGTA directs bound MLPs towards deubiquitination and hence stabilisation (Wunderley *et al.*, 2014; Leznicki and High, 2012). This would allow viable substrates with a further opportunity at productive folding, and to engage with the corresponding targeting machinery for ER delivery. On the other hand, the BAG6 complex directs MLPs towards polyubiquitination followed by proteasomal degradation (Hessa *et al.*, 2011; Mariappan *et al.*, 2010). Thus for terminally misfolded MLPs and TA proteins, the BAG6 complex triages them onto a

degradative route. Likewise, for prematurely ubiquitinated TA proteins, SGTA dependent deubiquitination could precede transfer to the Ubl4a/TRC35 subunits of the BAG6 complex for TRC40 dependent membrane targeting (see Section 1.7).

Therefore, SGTA is intimately linked with the quality control function of the BAG6 complex in deciding the fate of hydrophobic substrates such as MLPs and TA proteins, by either promoting them towards ER delivery or marking them for proteasomal degradation (Wunderley *et al.*, 2014; Leznicki *et al.*, 2013; Leznicki and High, 2012; Hessa *et al.*, 2011). SGTA facilitates this process by promoting MLPs towards the BAG6 complex, thereby preventing formation of protein aggregates. In the case of terminally misfolded substrates, the BAG6 complex recruits the E3 ubiquitin ligase, RNF126, to mark such MLPs or aberrant precursors for proteasomal degradation (Rodrigo-Brenni *et al.*, 2014). Likewise, SGTA is capable of reversing ubiquitination caused by the BAG6 complex via an as yet unidentified deubiquitinating enzyme thereby rescuing proteins from degradation (Leznicki and High, 2012). Therefore, the SGTA/BAG6 cycle constitutes several successive cycles of substrate ubiquitination and deubiquitination, the outcome of which will eventually decide the fate of MLPs and TA proteins present in the cytosol.

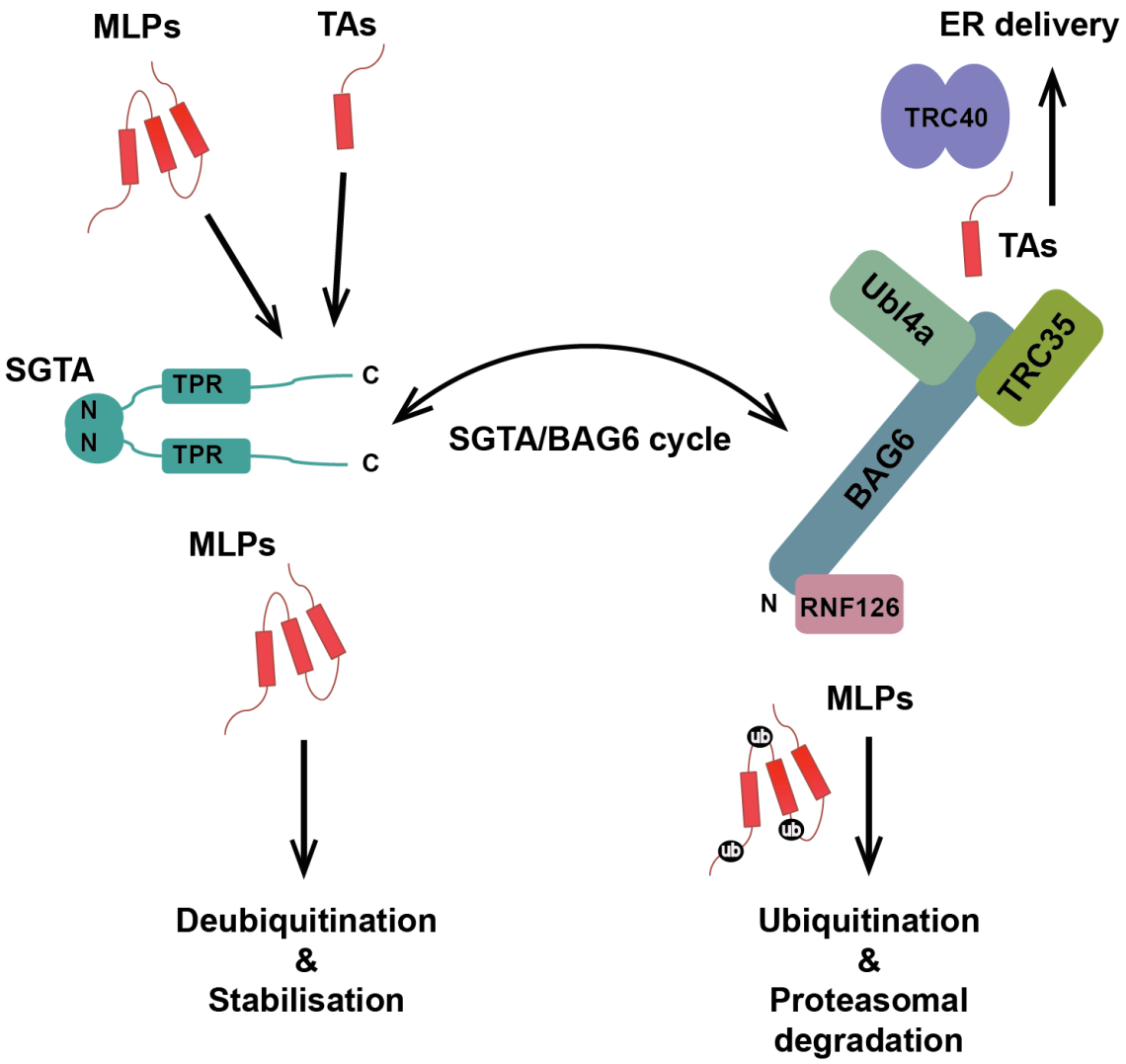



**Figure 1.10- The SGTA/BAG6 cycle of quality control in the cytosol.** SGTA promotes the deubiquitination and stabilisation of MLP substrates, while the BAG6 complex promotes their ubiquitination via the E3 ubiquitin ligase RNF126. BAG6 directs TA protein precursors to the TRC40 targeting factor, which facilitates their ER delivery. (Wunderley *et al.*, 2014; Rodrigo-Brenni *et al.*, 2014; Leznicki *et al.*, 2013; Leznicki and High, 2012; Hessa *et al.*, 2011)

## 1.7 SGTA in Get/TRC pathways of TA membrane protein insertion

Integral membrane proteins participate in a diverse range of cellular activities. They consist of hydrophobic transmembrane regions (TMDs), and are synthesised in the cytoplasm, which necessitates a mechanism whereby they are protected from the aqueous cytosolic environment to ensure targeted delivery and integration into their membrane residences. Most membrane proteins access the signal recognition particle (SRP) dependent co-translational pathway, in which a SRP bound ribosome-nascent chain complex is passed on to the Sec61 translocon, which subsequently facilitates membrane integration (Rapoport, 2007). However, for the tail-anchored (TA) family of membrane proteins, this co-translational pathway turns out to be sterically inaccessible. TA proteins are characterised by the presence of a functional cytoplasmic domain tethered to the lipid bilayer by a single pass TMD helix present at their extreme C-terminus, a region which also includes their membrane targeting signal (Hegde and Keenan, 2011). These TA proteins constitute around five percent of all human membrane proteins, to include SNAREs, ER translocon components and signalling proteins (Simpson *et al.* 2010). TA proteins cannot access the co-translational pathway because their TMD helix will still be present within the ribosomal exit tunnel as translation nears completion. Therefore an alternative, SRP independent, conserved post-translational membrane insertion pathway referred to as the mammalian TRC (transmembrane domain recognition complex) pathway is employed by TA proteins for their correct targeting and integration into the lipid bilayer of the endoplasmic reticulum (ER) (Hegde and Keenan, 2011). The existence of this SRP and translocon independent, post-translational insertion process, was first demonstrated with the SNARE protein synaptobrevin by Kutay *et al.* (1995). However, details pertaining to this post-translational pathway began to emerge with the discovery of the mammalian cytosolic ATPase TRC40 (40 kDa subunit of the transmembrane domain recognition complex),

which was identified to be the central targeting factor of TA proteins to the ER (Stefanovic and Hegde, 2007).

An early step in the mammalian TRC pathway involves the interaction of TMD helices of TA proteins with SGTA, which acts as a pre-targeting factor. SGTA bound TA proteins are then handed over to the TRC40 ATPase as a result of the concerted actions of the SGTA/BAG6 cycle (Mock *et al.*, 2015; Leznicki *et al.*, 2010; Mariappan *et al.*, 2010). The BAG6 protein acts a scaffold for Ubl4a and TRC35 that together ensure effective TA substrate handover from SGTA to TRC40 (Fig. 1.11). The TA-substrate bound TRC40 is then delivered to the ER via its cognate receptor formed by a heterodimeric membrane protein complex of WRB (tryptophan-rich basic protein) and CAML (calcium-modulating cyclophilin ligand) (Fig. 1.11) (Yamamoto and Sakisaka, 2012; Vilardi *et al.*, 2011; Schuldiner *et al.*, 2008).

In yeast, the equivalent TA insertion pathway is referred to as the guided entry of tail-anchored proteins (Get) pathway (Simpson *et al.* 2010). The Get pathway involves handover of TA protein substrates bound to the Sgt2 co-chaperone (the yeast equivalent of SGTA) to the Get3 ATPase (the yeast equivalent of TRC40), a process mediated by the Get4/Get5 complex (homologous to mammalian TRC35 and Ubl4a). Also, there is no yeast equivalent of the BAG6 protein, however the Get4/Get5 complex is sufficient to ensure substrate transfer from Sgt2 to Get3 (Casson *et al.*, 2016). This is followed by subsequent TA-protein release at the ER membrane by the Get1/Get2 receptor complex (the yeast equivalent of mammalian WRB and CAML proteins). Moreover, the Get1/2 receptor complex has been shown to function as a membrane insertase specific for TA proteins (Wang *et al.*, 2014).

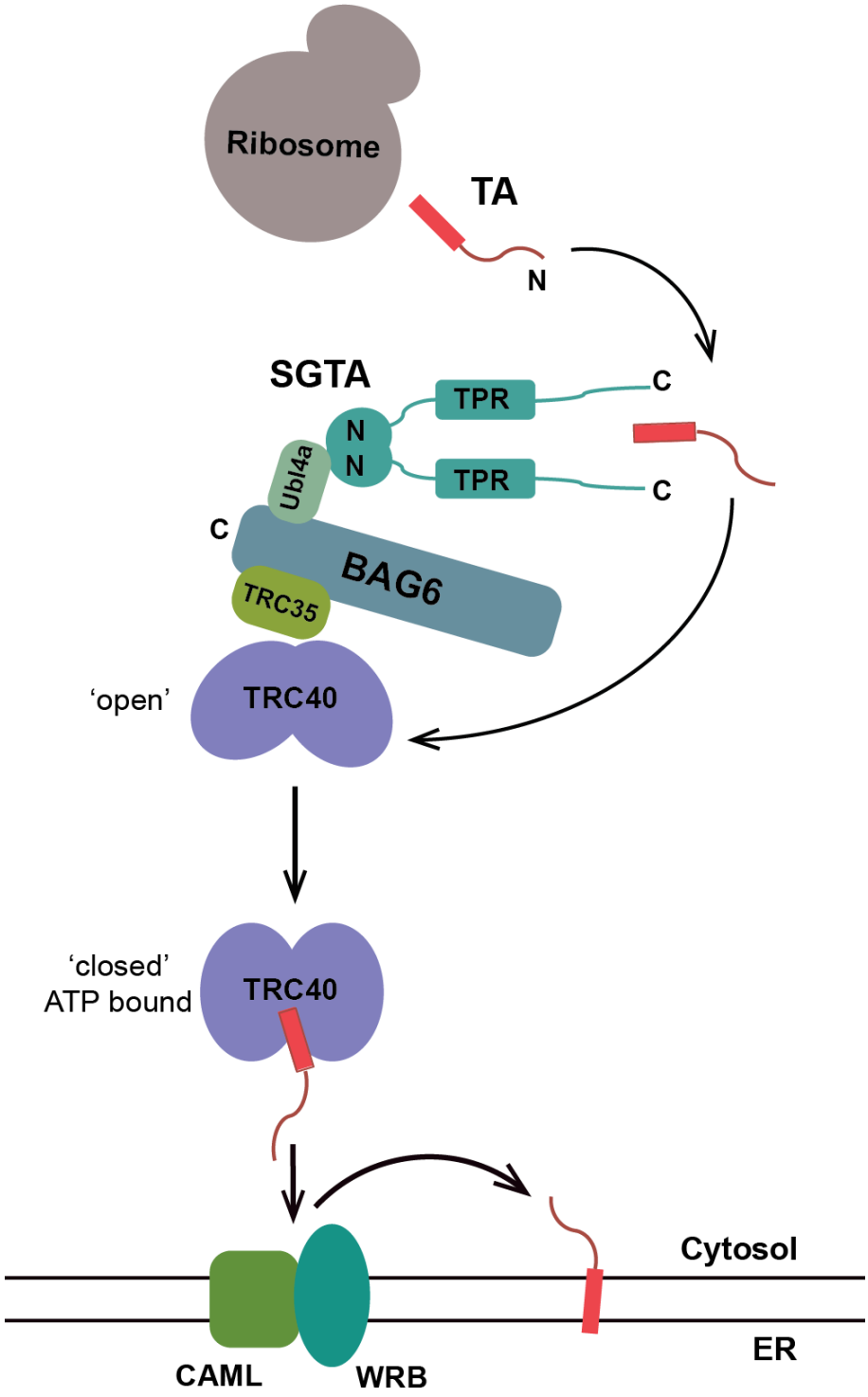



**Figure 1.11– The TRC pathway of TA protein insertion.** The TMD helices of newly synthesised TA membrane proteins are shielded from the aqueous cytosol by SGTA. These proteins are then passed on via the heterotrimeic BAG6 complex to the TRC40 ATPase, which is the central targeting machinery for TA protein insertion. Finally, the TRC40 bound TA protein is handed over to the CAML/WRB receptors for membrane insertion.

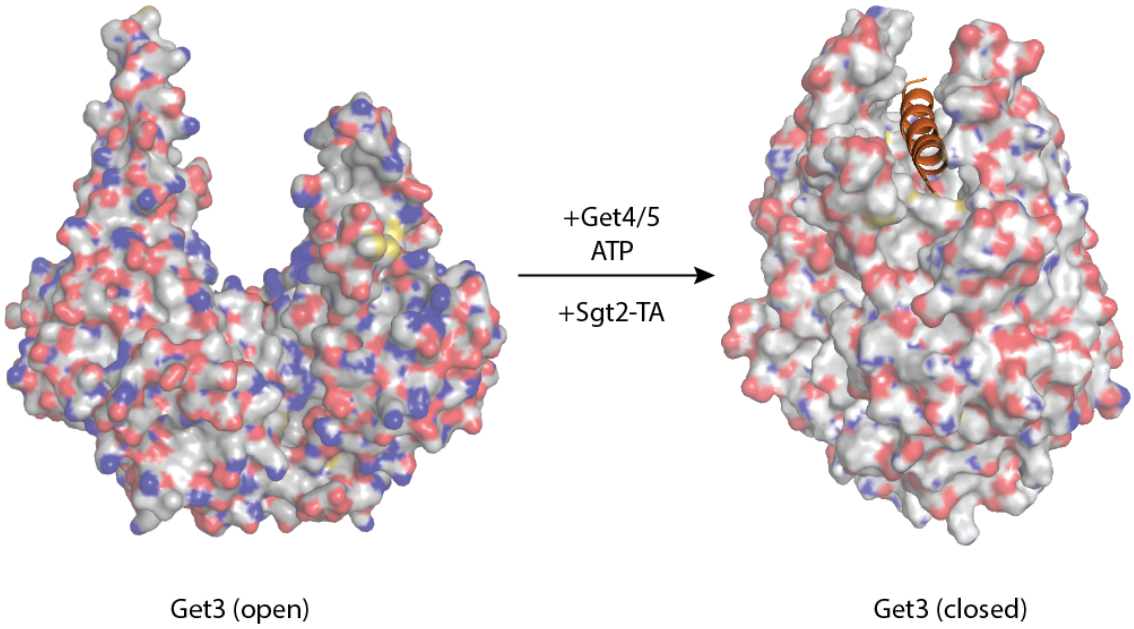



**Figure 1.12- Get3 ATPase.** The Get3 ATPase is the yeast equivalent of mammalian TRC40. The Get3 homodimer transitions from an 'open' to a 'closed' state, driven by ATP binding and contributions from the Get4/5 complex, in order to capture a TMD helix of a TA protein from Sgt2. The groove at the dimer interface effectively buries regions of hydrophobicity presented by the TMD helix. The Get3 homodimer is shown in surface representation, and the TMD helix from the TA protein Pep12 in the 'closed' state is shown in cartoon representation. (PDB accession codes: 5BW8 for 'open' and 4XTR for 'closed' states) (Mateja *et al.*, 2015).

Additionally, structures of the central targeting component of the yeast Get pathway, the Get3 ATPase, have been solved in different nucleotide-bound states and conformations, with the apo 'open' conformation being the ATP-free state. (Gristick *et al.*, 2015; Simpson *et al.*, 2010). More recently, a crystal structure of the Get3 ATPase homodimer in complex with the TMD helix of a TA protein has revealed that ATP binding stabilizes a 'closed' conformation of Get3. In this 'closed' conformation the Get3 homodimer exposes a large hydrophobic groove, capable of interacting with the TMD

helix of the TA protein (Fig. 1.12) (Mateja *et al.*, 2015). This conformation of Get3 protects that hydrophobic TMD helix from the surrounding aqueous cytosol and ensures the delivery of its TA cargo to the ER via the association of substrate-loaded Get3/TRC40 with cytosolic domain of CAML/Get2. Subsequent ATP hydrolysis is required to expose the WRB/Get1-binding site of Get3/TRC40 leading to TA protein release and integration into the ER (Kubota *et al.*, 2012; Mariappan *et al.*, 2011; Stefer *et al.*, 2011).

Thus far, the TRC/Get pathway has been known as the main SRP-independent ER targeting pathway facilitating the post-translational membrane insertion of TA proteins. Recently, a novel SRP-independent membrane targeting pathway has been identified called the SND (for SRP-independent targeting) pathway (Aviram *et al.*, 2016). This newly discovered SND pathway has been shown to be capable of compensating for the loss of SRP-dependent or TRC/Get pathways. Moreover, it has been suggested that the SND pathway functions in parallel to the TRC/Get pathways in targeting proteins with C-terminal hydrophobic regions (such as TA proteins) (Aviram *et al.*, 2016). Additionally, other cytosolic factors such as the ubiquilin family of proteins have also been shown to chaperone TMD helices of both TA and non-TA proteins. (Itakura *et al.*, 2016). In particular, ubiquilins have been proposed to be major TMD binding proteins acting on precursors destined for the mitochondrial membrane (Itakura *et al.*, 2016). It is thus clear that different cytosolic pathways exist to chaperone vulnerable TMD regions on newly synthesised proteins. The possibility of further yet to be identified targeting factors could add to what appears to be a robust network of cytosolic pathways ensuring the delivery of nascent membrane proteins to their correct intracellular destinations.

## 1.8 SGTA and the ERAD pathway

Terminally misfolded proteins in the ER are removed by a process referred to as ER-associated degradation (ERAD). ERAD encompasses protein quality control pathways within the ER that facilitate the removal of misfolded substrates by promoting their retrotranslocation into the cytosol, which is followed by ubiquitin mediated proteasomal degradation. This process is facilitated by the gp78/Hrd1 membrane-bound E3 ubiquitin ligase together with the retrotranslocation driving p97 ATPase (Stein *et al.*, 2014; Xu *et al.*, 2012; Wang *et al.*, 2011). In the cytosol, ERAD substrates are known to interact with BAG6, which maintains them in a soluble state (Payapilly and High, 2014; Wang *et al.*, 2011). This process is facilitated by the recruitment of SGTA to the BAG6 complex via Ubl4a, which has been shown to play a role in maintaining the solubility of retrotranslocated substrates in the cytosol preventing their aggregation, thereby promoting ERAD (Fig. 1.13) (Xu *et al.*, 2012). Hence it is apparent that SGTA, in collaboration with BAG6, can influence the fate of substrates destined for ERAD mediated proteasomal degradation.

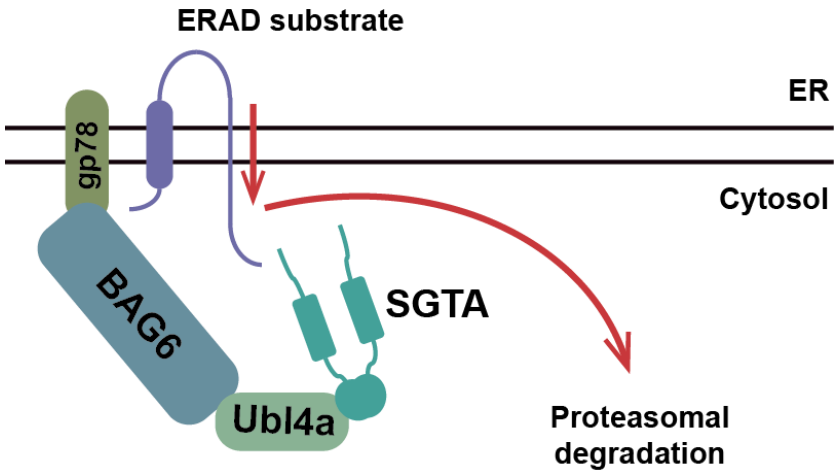



**Figure 1.13- SGTA in the ERAD pathway.** SGTA together with the BAG6 maintains the solubility of ERAD substrates in the cytosol. This is mediated by the Ubl4a dependent recruitment of SGTA to BAG6 (Xu *et al.*, 2012).

## 1.9 SGTA mediated quality control at the proteasome

In a recent study by Leznicki *et al.* (2015), it has been shown that SGTA is recruited to the 26S proteasome via the intrinsic proteasomal receptor Rpn13. This interaction was first identified by a yeast two-hybrid screen, and further pull-down assays showed that the C-terminal domain of Rpn13 was the SGTA interacting region. The same study also demonstrated that the association of exogenous SGTA to Rpn13 could lead to an increase in the steady-state MLP levels. This provides new evidence of SGTA dependent modulation of MLP quality control occurring at the proteasome.

### *1.9.1 The 26S proteasome*

The 26S proteasome is the cellular machinery that degrades proteins conjugated to ubiquitin. It consists of a barrel shaped 20S core where proteolysis occurs, and has its ends capped by 19S regulatory particles that control substrate entry (Fig. 1.14). The 19S regulatory particles contain receptors for ubiquitin that in turn interact with ubiquitinated substrates and shuttle receptors. Substrates bound to the 19S particle are then translocated to the 20S core for degradation (Finley *et al.*, 2016). Substrates delivered to the 19S particle are ultimately recognised by the one of its three intrinsic proteasomal ubiquitin receptors namely Rpn1 (Shi *et al.*, 2016), Rpn10 (Deveraux *et al.*, 1994) and Rpn13 (Husnjak *et al.*, 2008). Also, the 19S particle consists of deubiquitinating enzymes (DUBs) that reverse ubiquitination thus generating free ubiquitin available for conjugation. These DUBs are Ubp6/Usp14, which binds to Rpn1 (Shi *et al.*, 2016) and UCH-L5/Uch37 (Lam *et al.*, 1997) that binds to a deubiquitinase adaptor (DEUBAD) domain of Rpn13 (Sahtoe *et al.*, 2015, VanderLinden *et al.*, 2015). Once ubiquitin molecules are removed from substrates by one of these DUBs, a heterohexameric ring of AAA ATPase proteins present within the 19S particle (Rpt1-

Rpt6) is responsible for the unfolding followed by translocation of substrates into the 20S proteolytic core of the proteasome (Finley *et al.*, 2016).

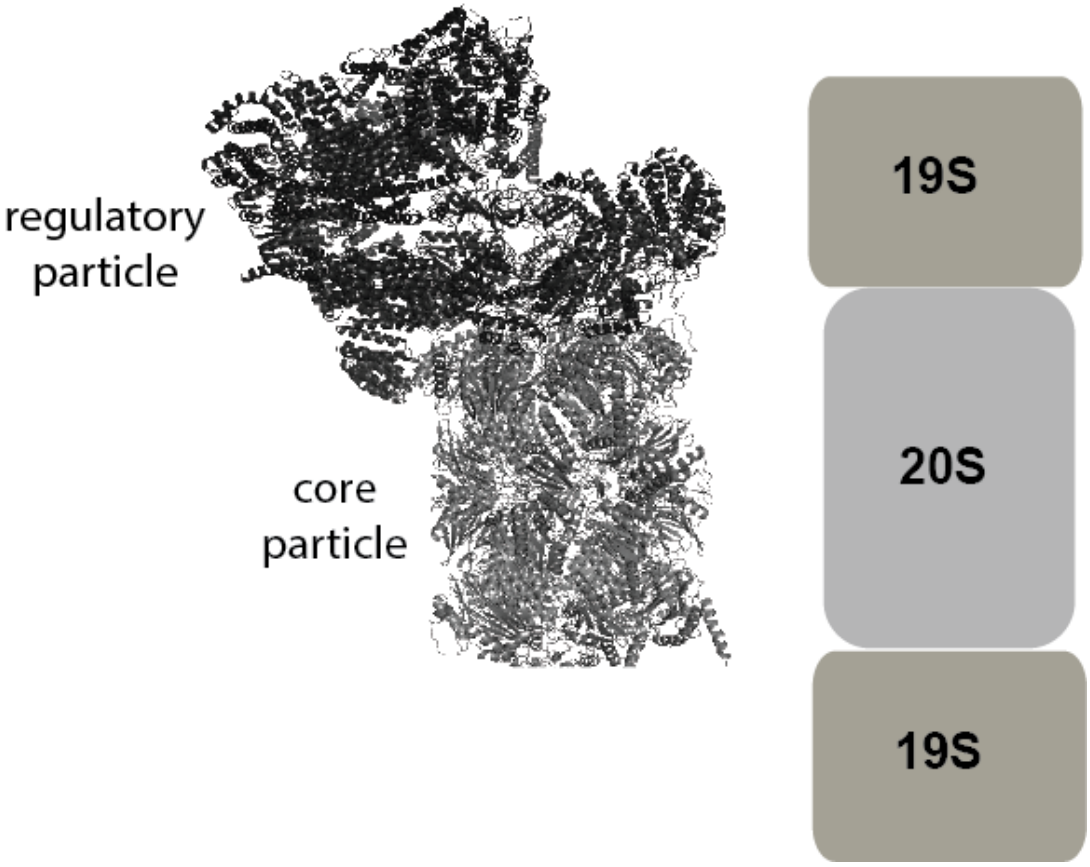



**Figure 1.14- The 26S proteasome.** The 26S proteasome is a multiprotein complex that comprises a barrel shaped 20S core where proteolysis occurs, capped at either end by a 19S regulatory particle that controls substrate entry (Huang *et al.*, 2016; PDB accession code: 5GJR).

### *1.9.2 The intrinsic proteasomal ubiquitin receptor Rpn13*

Rpn13 is a major proteasomal ubiquitin receptor that consists of an N-terminal pleckstrin-like receptor for ubiquitin (Pru) domain that binds to both ubiquitinated substrates and the proteasome, and a conserved C-terminal deubiquitinase adaptor (DEUBAD) domain that binds to the UCH-L5 (UCH37) deubiquitinating enzyme (Chen *et al.*, 2010). Rpn13 is thus capable of facilitating substrate delivery to the

proteasome via its N-terminal Pru domain. This occurs as a result of the Pru domains ability to directly interact with conjugated ubiquitin chains and with shuttle factors. However, when not bound to the proteasome, the Pru and DEUBAD domains associate with each other consequently lowering Rpn13's affinity for ubiquitin (Fig. 1.15) (Chen *et al.*, 2016; Chen *et al.*, 2010).

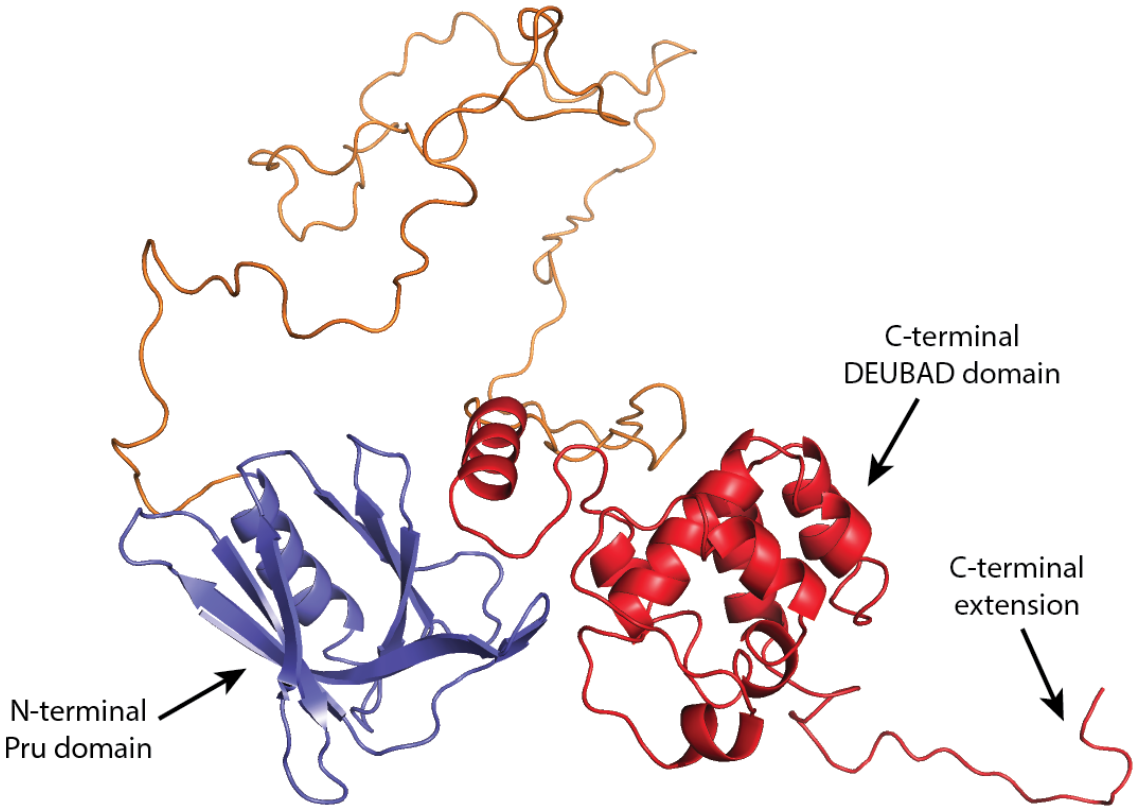



**Figure 1.15- Structure of full-length human Rpn13.** Human Rpn13 consists of an N-terminal pleckstrin-like receptor for ubiquitin (Pru) domain that binds to both ubiquitinated substrates and the proteasome (PDB accession code: 2KR0), a flexible linker, a conserved C-terminal deubiquitinase adaptor (DEUBAD) domain that binds to the UCH-L5 (UCH37) deubiquitinating enzyme which is followed by a flexible C-terminal extension. When not bound to the proteasome, the Pru and DEUBAD domains associate with each other consequently lowering Rpn13's affinity for ubiquitin (PDB accession code: 2KR0) (Chen *et al.*, 2010).

Binding of Rpn13 to the 19S regulatory particle of the proteasome abrogates the interdomain association between its Pru and DEUBAD domains. This results in Rpn13 activation and thereby offers ubiquitinated substrates unimpeded access to its Pru domain (Fig. 1.16) (Chen *et al.*, 2010). The subsequent interaction of ubiquitin with the Pru domain is rather novel, with loop regions involved in the binding as opposed to secondary structural elements (Schreiner *et al.*, 2008).

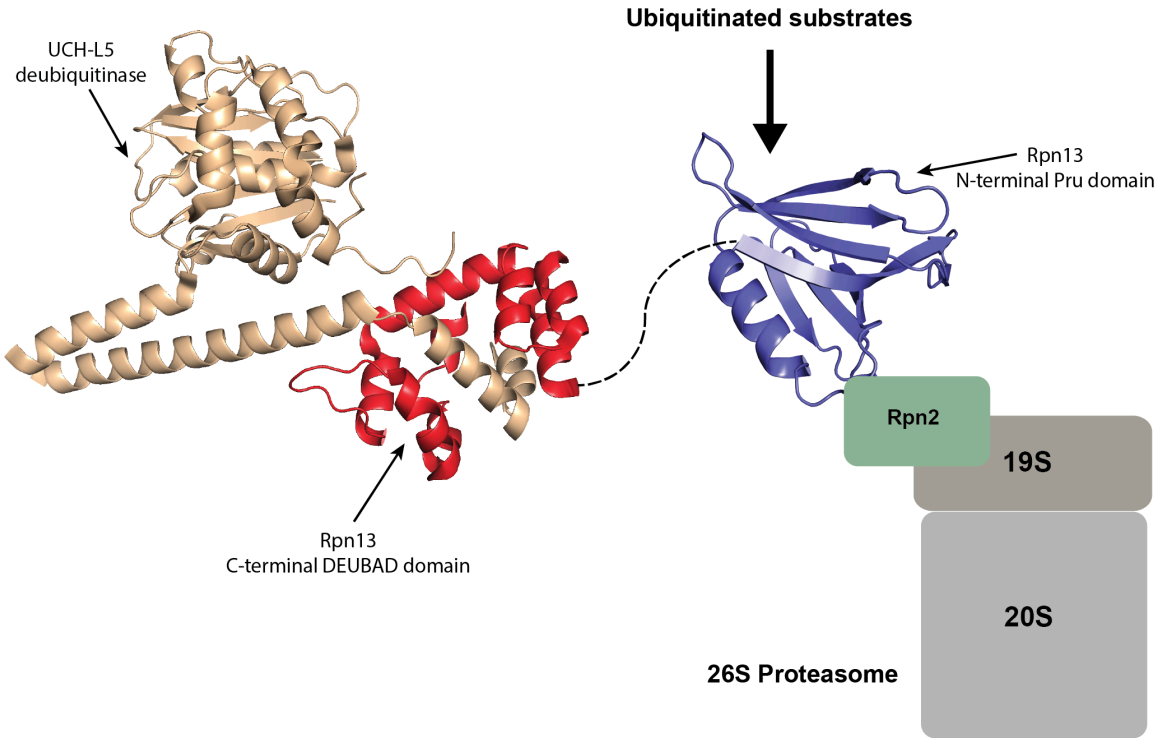



**Figure 1.16- Rpn13 activation upon binding to the proteasome.** Binding of Rpn13 to the Rpn2 subunit at the 19S regulatory particle leads to a dissociation between Rpn13's Pru and DEUBAD domains. This gives ubiquitinated substrates access to Rpn13 via its Pru domain (PDB accession code: 2KR0). Also, upon binding to the proteasome, Rpn13 can interact with the UCH-L5 (UCH37) deubiquitinase via its DEUBAD domain (PDB accession code: 4UEM) (Sahtoe *et al.*, 2015; Chen *et al.*, 2010).

Furthermore, activation of Rpn13 at the proteasome will also enable UCH-L5/UCH37 binding to its C-terminal DEUBAD domain (Fig. 1.16), thus enabling the UCH-L5

deubiquitinase to access potential substrates (Sahtoe *et al.*, 2015, VanderLinden *et al.*, 2015). From a structural point of view, the Rpn13 DEUBAD domain exists as a compact 8-helical bundle in its apo state (Chen *et al.*, 2010) (Fig. 1.15). This helical bundle is fractured to accommodate the C-terminal region of UCH-L5/UCH37 as evident from crystal structures of the UCH-L5/DEUBAD complex (Fig. 1.16) (Sahtoe *et al.*, 2015, VanderLinden *et al.*, 2015). Thus it is apparent that SGTA's interaction with the C-terminal domain of Rpn13 is likely to have implications regarding the ultimate fate of ubiquitinated MLPs that arrive at the proteasome.

In addition to ubiquitin chains and shuttle receptors binding to the Rpn13 Pru domain, it has recently been discovered that the E3 ubiquitin ligase parkin can be recruited to the proteasome via an interaction with the Pru domain. In addition, this association with the proteasome can modulate parkin activity (Aguileta *et al.*, 2015). It is therefore possible for both SGTA and parkin to be present simultaneously at the 19S particle, however the significance of this finding is yet to be understood. As mutations in parkin have been associated with a familial form of Parkinson's disease (Aguileta *et al.*, 2015), exploring possible effects of SGTA on parkin function might be of future interest.

## 1.10 SGTA in health and disease

SGTA has been implicated in the pathogenesis of many human diseases including various cancers, hormonally induced disease states, and neurodegenerative disorders. Thus far, malignancies associated which SGTA include cancers of the prostate (Trotta *et al.*, 2013; Buchanan *et al.*, 2007), ovary (Butler *et al.*, 2013), breast (Zhu *et al.*, 2014), liver (Lu *et al.*, 2014) and oesophagus (Yang *et al.*, 2014). SGTA has also been associated with the pathogenesis of the hormone-related polycystic ovary syndrome (Goodarzi *et al.*, 2008). In addition, SGTA has been linked to suppression of toxicity

associated with human beta amyloid expression (Fonte *et al.*, 2002), and has been implicated in prion diseases (Kiktev *et al.*, 2012). As evidence of SGTA's engagement in various disease processes is rapidly emerging, further investigations will be required to improve our understanding of its involvement in aforementioned pathologies.

## 1.11 SGTA in hormone receptor signaling

SGTA is involved in steroid hormone receptor signalling via its association with the androgen receptor (Paul *et al.*, 2014). The androgen receptor is a nuclear transcription factor and associated signaling pathways are implicated in the pathogenesis of hormonally induced disease states, including prostate and breast cancers. Interestingly, SGTA has been found to be upregulated in such disease states (Goodarzi *et al.*, 2008; Buchanan *et al.*, 2007; Zhu *et al.*, 2014). It has been suggested that SGTA is recruited to the androgen receptor via the SGTA_TPR domain (Buchanan *et al.*, 2007). Additionally, the Rec protein, a member of endogenous retroviruses of the HERV-K (HML-2) family, has been suggested to regulate androgen receptor activity mediated by an interaction with SGTA (Hanke *et al.*, 2013).

Furthermore, SGTA has been implicated in pathways that involve endocytosis of the growth hormone receptor, and in the regulation of glucocorticoid and progesterone receptor activity (Paul *et al.*, 2014; Schantl *et al.*, 2003). With regard to SGTA dependent hormone receptor regulation in general, it has been observed that depletion of SGTA in cells enhances receptor activity, whilst SGTA overexpression leads to suppression of receptor function (Paul *et al.*, 2014). However, in the absence of a detailed characterisation of direct interactions between these components, precise mechanisms as to how SGTA modulates hormone receptor signalling remain unclear.

## 1.12 The role of SGTA in viral lifecycles

SGTA was discovered in complex with a nonstructural protein of parvovirus H-1 (NS1) and was suggested to play a role in viral replication and gene expression (Cziepluch *et al.*, 1998). More recently, SGTA has been identified to associate with the SARS-CoV 7a (severe acute respiratory syndrome coronavirus 7a) protein, and has been suggested to be involved in SARS-CoV virus particle assembly (Fielding *et al.*, 2006). Also, SGTA is involved in promoting viral infections by facilitating ER membrane penetration, as observed for the nonenveloped simian virus 40 (SV40) (Walczak *et al.*, 2014).

In addition, SGTA can influence human immunodeficiency virus type 1 (HIV-1) particle release (Waheed *et al.*, 2016; Handley *et al.*, 2001). HIV-1 is the causative agent of AIDS, and SGTA has been identified as a binding partner of the HIV-1 accessory protein Vpu, and of the HIV-1 structural polyprotein Gag (Dutta and Tan, 2008; Handley *et al.*, 2001; Callahan *et al.*, 1998). SGTA overexpression can inhibit HIV-1 particle release. Moreover, SGTA's interaction with the HIV-1 Gag protein has been suggested to be responsible for this inhibitory effect. However, in the presence of Vpu, SGTA's interaction with Gag is abolished, thus allowing HIV-1 particle release (Dutta and Tan, 2008; Handley *et al.*, 2001; Callahan *et al.*, 1998). Hence it is clear that SGTA plays a role in viral lifecycles, however mechanistic details underlying SGTA's role are yet to be fully understood.

## 1.13 Aims of work described in this thesis

SGTA is emerging as a key effector in a wide-range of biological processes, with fundamental roles in MLP quality control and TA protein biogenesis. However, structural and biophysical studies carried out on SGTA thus far have not been able to address its full-length assembly and conformations. Therefore work described in this thesis aims to characterise full-length SGTA *in vitro*, with a particular focus on its substrate binding C-terminal domain, molecular details of which remain poorly understood (Chapter 3). Also, it is important to study how SGTA binds to its hydrophobic substrates via its C-terminal domain, as this is pivotal to its role in enforcing cytosolic quality control. Thus Chapter 4 investigates the interaction of SGTA with hydrophobic substrates, to understand how SGTA, in particular its C-terminal domain, recognises exposed regions of hydrophobicity on tail-anchored membrane proteins destined for the Get/TRC pathway of membrane insertion. In addition, recent work has suggested that recruitment of SGTA to the 26S proteasome modulates quality control of hydrophobic substrates (Leznicki *et al.*, 2015). Thus biophysical characterisation of the interaction between SGTA and the intrinsic proteasomal receptor Rpn13 is described in Chapter 5 with an aim to further our understanding as to how an SGTA-dependent quality control cycle may operate at the proteasome. Overall, this thesis aims to uncover molecular details underlying SGTA's role in contributing to cellular protein homeostasis.

# Chapter 2

# Materials and methods

## 2.1 Chemicals

All chemicals used were of laboratory grade.

## 2.2 Water

A Milli-Q Academic purification system (Millipore Ltd, Watford, U. K.) was used to purify de-ionised water by reverse osmosis. This was achieved by passing water through a 0.22 μm filter to produce ultra-pure water. Milli-Q purified water was used in the preparation of all aqueous solutions.

## 2.3 Synthetic genes

Synthetic genes were obtained as GeneArt™ Strings DNA Fragments (Thermo Fisher Scientific). These DNA fragments were codon optimised for expression in *E. coli*.

## 2.4 Synthetic DNA oligonucleotides

Short DNA oligonucleotides used as primers for PCR amplification of DNA fragments for plasmid cloning were synthesised by Integrated DNA Technologies (IDT).

## 2.5 Vectors and plasmids

The home-modified pET28 vector derived from the commercially available pET28 vector (Novagen) used for generating various constructs used in this study, together with the plasmid encoding recombinant TEV protease were provided by Dr J.M. Pérez-Cañadillas (Rocasolano Chemical Physical Institute, Madrid, Spain).

## 2.6 *E. coli* strains

Recombinant proteins were expressed in calcium competent *E. coli* BL21 (DE3) Rosetta cells. For routine plasmid cloning, subcloning efficiency NEB 5-alpha calcium competent *E. coli* were used (New England Biolabs). For site-directed mutagenesis, either high efficiency NEB 10-beta calcium competent *E. coli* (New England Biolabs) or XL-1 blue supercompetent *E. coli* (Agilent) were used.

## 2.7 Bacterial growth media

For bacterial growth, agar plates were made with lysogeny broth LB media [1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl, pH 7.4] supplemented with 1.5% (w/v) bacto agar (Miller's LB Agar, Melford) and 50 μg/ml of kanamycin for selection of resistant colonies. The LB media used as rich liquid growth media was the same as described above except bacto agar was not added (LB broth granulated, Melford).

For isotopically ($^{15}N$ or $^{15}N/^{13}C$) labelled proteins, growth was carried out in M9 minimal media. To prepare M9 media, firstly M9 salts [$Na_2HPO_4$ (6 g/L), $KH_2PO_4$ (3 g/L), NaCl (0.5 g/L)] were sterilised by autoclaving. The M9 salt solution was then supplemented with labelled ammonium chloride (>98 % $^{15}N$, Sigma-Aldrich) and/or glucose (>99% U-$^{13}C$, Sigma-Aldrich), 0.1 μM $CaCl_2$, 2 μM $MgSO_4$, BME Vitamins 100x solution (10 ml/L) (Sigma-Aldrich), 10 ml/L trace elements from 10X stock solution (10 mM $FeCl_3 \cdot 6H_2O$, 5 mM $CuCl_2 \cdot 6H_2O$, 10 mM $ZnCl_2 \cdot 4H_2O$, 15 mM $CoCl_2 \cdot 6H_2O$, 2 mM $H_3BO_3$ and 10 mM $Na_2MoO_4 \cdot 2H_2O$), and 50 μg/ml of kanamycin. All M9 supplements were added as filtered sterilised solutions using syringe driven filters with 0.22 μm pores (Millex-GP syringe filter unit, Millipore).

## 2.8 Synthetic peptides

The human Rpn13 derived pentapeptide (Rpn13 residues 403-407; sequence - DMSLD) used in this study was synthesised by Alta BioScience (Birmingham, UK). The purified peptide was verified by HPLC and mass spectrometry.

## 2.9 Chromatography

Chromatography columns were obtained pre-packed. HisTrap columns (5 ml) prepacked with high performance nickel Sepharose were used for purification of histidine-tagged proteins by immobilized metal ion affinity chromatography (IMAC). Size-exclusion chromatography was carried out on either a 120 ml HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare), or a 120 ml HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare). Chromatography was carried out on AKTA prime, AKTA purifier or AKTA pure fast protein liquid chromatography (FPLC) systems controlled by the Unicorn software package (GE Healthcare).

## 2.10 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out on 1% (w/v) agarose gels prepared in 1X TAE (Tris-acetate–EDTA) buffer (40 mM Tris-HCl pH 7.6; 20 mM acetic acid; 1mM EDTA). 6X DNA loading dye (1X TAE, 5% (w/v) glycerol, 0.04% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanol FF) was added to DNA samples prior to loading on agarose gels to track the progress of electrophoresis. Gels were run in 1X TAE at 100 V. DNA was visualized by straining with SYBR safe DNA gel stain (ThermoFisher Scientific), followed by illumination with a safe imager blue-light transilluminator (ThermoFisher Scientific).

## 2.11 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on discontinuous gels (Laemmli, 1970) with a 12% acrylamide resolving gel and a 5% (w/v) stacking gel. These gels were poured in-house, with the resolving gel containing 12% polyacrylamide, 0.45 M Tris-HCl pH 8.8, 1% of SDS, 1% of ammonium persulphate and 0.02% of N,N,N,N tetramethylethylendiamine (TEMED). On the other hand, the stacking gel was made of 5% of polyacrylamide, 0.125 M Tris-HCl pH 6.8, 1% of SDS, 1% of ammonium persulphate and 0.02% of TEMED. Protein samples were denatured by boiling in SDS loading buffer (50 mM Tris-HCl pH 6.8, 1% (w/v) sodium dodecyl sulphate (SDS), 0.01% (w/v) bromophenol blue, 1% β-mercaptoethanol, 10% (v/v) glycerol), then loaded on to the gels that were run in a standard Tris/glycine buffer (Laemmli buffer: 25 mM Tris base, 200 mM glycine and 1% SDS). Electrophoresis was performed in a Mini-PROTEAN tetra vertical electrophoresis cell (Bio-Rad) at a constant current of 42 mA until the dye front had reached the end of the gel. Protein bands were visualised by staining with quick Coomassie stain (Generon) following the manufacturers instructions.

## 2.12 Estimation of protein concentration

Protein concentrations were estimated by measuring sample UV absorbance at a wavelength of 280 nm with a NanoDrop 8000 spectrophotometer (Thermo Scientific). Protein concentrations were calculated using the Beer-Lambert law, given by,

$$A = \varepsilon.c.l$$

where, A is measured absorbance, ε is the molar extinction coefficient, c is sample concentration, and l is the path length through the sample. Theoretical molar extinction coefficients (ε) were obtained from the ExPASy ProtoParm server with amino acid

sequence of the protein of interest used as input (http://web.expasy.org/protparam/). For protein constructs lacking tyrosine or tryptophan residues concentrations were determined using the Pierce BCA assay (Thermo Scientific) calibrated with bovine serum albumin following the manufacturer's protocol.

## 2.13 Restriction endonuclease mediated cloning

Expression plasmids encoding full-length human SGTA, full-length human tail-anchored proteins cytochrome b5 (Cytb5) and synaptobrevin-2 (Syb2), the C-terminal domain of human Rpn13, and variants thereof were generated in a home-modified pET28 vector (kanamycin resistant). In this system, recombinant mRNA transcription is achieved by the IPTG inducible bacteriophage T7 RNA polymerase. All plasmids were constructed to encode fusion proteins consisting of an N-terminal thioredoxin (TxA) protein followed by a hexa-histidine tag ($His_6$) and rTEV protease cleavage site (ENLYFQ ^ GS) followed by the sequence encoding the protein of interest. The N-terminal TxA was engineered to enhance expression levels and improve solubility, the $His_6$ was added to aid affinity purification of the resulting protein, and the rTEV site was inserted to allow for the removal of this N-terminal part of the fusion protein (Fig. 2.1). Cloning was carried out in the steps as described below.

### *2.13.1 Construct design and domain boundaries*

For SGTA constructs, domain boundaries of its N-terminal, TPR and C-terminal domains have been defined as follows. The N-terminal region described in the solution NMR structure of SGTA (PDB accession code: 4CPG; residues 1-69) (Darby *et al.*, 2014) together with the flexible linker extending up to the start of the TPR domain was designated as the N-terminal domain construct (Nter; residues 1-86). Domain boundaries of the central TPR domain were based on the X-ray crystal structure of

human SGTA TPR (PDB accession code: 2VYI; residues 84-211) (Dutta and Tan, 2008). Regions distal to the TPR capping helix were designated as the SGTA C-terminal domain (residues 213-313). Within this C-terminal domain of SGTA, residues 213 to 274 constitute the NNP region (short C-terminal), and residues 275 to 313 form the Q-rich region, with the definition of both regions based on sequence homology.

For tail-anchored membrane proteins, constructs were designed encoding full-length cytochrome b5 (Cytb5; residues 1-134) and full-length synaptobrevin 2 (Syb2; residues 1-116).

For the intrinsic proteasomal ubiquitin receptor Rpn13, a C-terminal domain construct (residues 260-407) encoding its conserved deubiquitinase adaptor (DEUBAD) domain followed by a flexible C-terminal extension was designed, based on domain boundaries as defined in the solution NMR structure of human Rpn13 (PDB accession code: 2KR0) (Chen *et al.*, 2010).

### *2.13.2 Primer design*

Forward and reverse primers were designed (Table 2.1) in such a way that they annealed to the same sequence on opposite strands of the synthetic DNA template (Appendix A). This allowed for the amplification of the DNA sequence of interest using various combinations of primers (Table 2.2). The forward primer had a *Bam*HI restriction site, while the reverse primer contained a *Xho*I restriction site and a stop codon. This was to allow insertion of a restriction-digested fragment into the multiple cloning site (MCS) of the pET28 vector.

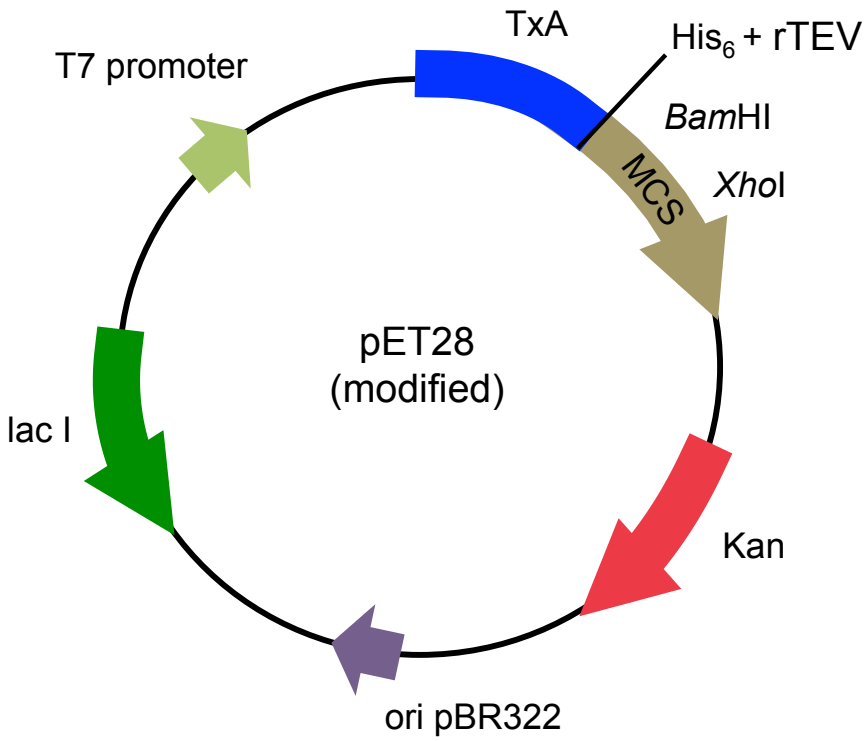



**Figure 2.1– Vector map of modified pET28 used for restriction endonuclease mediated cloning.** A pET28 vector encoding an N-terminally placed thioredoxin A (TxA) fusion protein followed by a hexahistidine tag ($His_6$) and rTEV cleavage site was used for plasmid cloning. The DNA sequence encoding the gene of interest was digested with *Bam*HI and *Xho*I followed by ligation into a linearised version of this vector digested with the same restriction enzymes.

**Table 2.1- Primers designed to clone various constructs into a modified pET28 vector encoding an N-terminal thioredoxin fusion protein followed by a hexa-histidine tag and rTEV cleavage site.**

| | |
|---|---|
| **SGTA constructs** | |
| *Forward primers* | |
| SGTA_1_Fw | 5'- CGC GGA TCC ATG GAC AAC AAG AAG CGC -3' |
| SGTA_84_Fw | 5'- CGC GGA TCC GAG GAG GAC TCA GCA GAG GC -3' |
| SGTA_191_Fw | 5'- CGC GGA TCC GAC AAC GAG ACA TAC AAG TCC -3' |
| SGTA_213_Fw | 5'- CGC GGA TCC ACG GGA GGC GTG GGC -3' |
| *Reverse primers* | |
| SGTA_86_Rv | 5'- CGC TCG AGA TCA CTC CTC GGA AGG CGG GG -3' |
| SGTA_211_Rv | 5'- CGC TCG AGA TCA CGT GGG GCT GGG GGC C -3' |
| SGTA_240_Rv | 5'- CGC TCG AGA TCA ATT GTT CAT TAG GTT CG -3' |
| SGTA_274_Rv | 5'- CGC TCG AGA TCA GAT GAG GCT GGC CAG GTC G-3' |
| SGTA_313_Rv | 5'- CGC TCG AGA TCA CTC CTG CTG GTC GTC G -3' |
| **Tail-anchored proteins Cytochrome b5 (Cytb5) and Synaptobrevin 2 (Syb2)** | |
| *Forward primers* | |
| Cytb5_Fw | 5'-CGC GGA TCC ATG GCA GAA CAG AGT GAT GAA GC-3' |
| Syb2_Fw | 5'-CGC GGA TCC ATG AGC GCA ACC GCA GCC ACC GC-3' |
| *Reverse primers* | |
| Cytb5_Rv | 5'- CGC TCG AGA TTA ATC TTC TGC CAT ATA C -3' |
| Syb2_Rv | 5'- CGC TCG AGA TTA GGT GCT AAA ATA CAC G -3' |
| **Rpn13 C-terminal domain constructs** | |
| *Forward primer* | |
| Rpn13_260_Fw | 5'- GCC GCG GAT CCC AGC CGA TTC AGC TGA GCG ATC TGC -3' |
| *Reverse primers* | |
| Rpn13_407_Rv | 5'- GGC CGG CTC GAG TTA GTC CAG GCT CAT ATC TTC C -3' |
| Rpn13_402_Rv | 5'-GGC CGG CTC GAG TTA TTC CTC TTC GTC TTT TTT ATC TTT GG -3' |

### *2.13.3 Polymerase chain reaction*

DNA fragments used for cloning were amplified by PCR using Q5 high-fidelity DNA polymerase (New England Biolabs). Typically a mix of 50 µl was prepared per reaction containing 0.5 µl polymerase, 10 µl Q5 reaction buffer (5x), 2.5 µl of each primer to generate a specific construct (Table 2.2), 2 µl dNTPs, the corresponding DNA template (Appendix) and 32 µl of nuclease free water. PCR reactions were performed in a DNA Engine Tetrad thermal cycler (Bio Rad) with the following settings: a 30 second activation step at 98°C, followed by 30 cycles of denaturation (10 seconds at 98°C), annealing of primers (30 seconds at 60°C) and elongation (60 seconds per kilobase at 72°C). The reaction was ended with a 2 minute step at 72°C. The presence of successful amplification products was confirmed by agarose gel electrophoresis.

**Table 2.2- Combinations of primers used to amplify various cDNA fragments to generate different constructs in the modified pET28 vector.**

| **SGTA constructs** | **Residues** | **Forward primer** | **Reverse primer** |
|---|---|---|---|
| Full length | 1-313 | SGTA_1_Fw | SGTA_313_Rv |
| N-terminal domain | 1-86 | SGTA_1_Fw | SGTA_86_Rv |
| N-terminal + TPR domains | 1-211 | SGTA_1_Fw | SGTA_211_Rv |
| N-terminal+TPR+short C-terminal | 1-274 | SGTA_1_Fw | SGTA_274_Rv |
| TPR domain | 84-211 | SGTA_84_Fw | SGTA_211_Rv |
| TPR + C-terminal domains | 84-313 | SGTA_84_Fw | SGTA_313_Rv |
| TPR + short C-terminal | 84-274 | SGTA_84_Fw | SGTA_274_Rv |
| TPR capping helix + short C-term | 191-274 | SGTA_191_Fw | SGTA_274_Rv |
| C-terminal domain | 213-313 | SGTA_213_Fw | SGTA_313_Rv |
| **Tail-anchored protein constructs** | | | |
| Cytochrome b5 (Cyt b5) | 1-134 | Cytb5_Fw | Cytb5_Rv |
| Synaptobrevin 2 (Syb2) | 1-116 | Syb2_Fw | Syb2_Rv |
| **Rpn13 constructs** | | | |
| C-terminal domain | 260-407 | Rpn13_260_Fw | Rpn13_407_Rv |
| C-terminal domain (truncated) | 260-402 | Rpn13_260_Fw | Rpn13_402_Rv |

### *2.13.4 Purification of DNA fragments*

Purification of DNA obtained by PCR amplification was carried out using the Wizard DNA clean-up system (Promega) following the manufacturers instructions. This step removed DNA polymerase, primers and dNTPs, with the resulting pure DNA eluted in 50 µl of nuclease free water.

### *2.13.5 Restriction digestion*

The modified pET28 vector was linearised by digestion with *Bam*HI (for 5'-end) and *Xho*I (for 3'-end) restriction enzymes (New England Biolabs). Similarly, DNA fragments amplified by PCR to be used as inserts were digested with the same restriction enzymes. Restriction digestion was carried out at 37°C with 1 µl of each enzyme in compatible buffers followed by purification of resulting DNA as described above. In addition, the linearised vector was dephosphorylated at its 5'-end using shrimp alkaline phosphatase (Roche) for 1 hour at 37°C.

### *2.13.6 Ligation*

Purified DNA inserts were ligated with linerised plasmids with T7 DNA ligase (New England Biolabs). Typically, 0.5 µl of dephosphorylated linearised plasmid (50-100 ng/µl) was mixed with 6.5 µl of a purified DNA insert with sticky ends, 1 µl T7 DNA ligase and 2 µl of T7 ligase buffer. Ligation was carried out overnight at 16°C.

### *2.13.7 Transformation*

The ligation product from cloning was transformed into NEB 5-alpha calcium competent *E. coli* cells (New England Biolabs). Competent cells were thawed on ice to maximise transformation efficiency. 5 µl of the ligation product was added to 50 µl of NEB 5-alpha cells and incubated on ice for 30 minutes. This was followed by a heat-

shock step for 30 seconds at 42°C. Cells were then outgrown at 37°C for 1 hour in 500 µl SOC media on a shaking platform. The cells were then plated on warm LB-agar plates with kanamycin for selection of resistant colonies. These plates were incubated overnight at 37°C.

### *2.13.8 Isolation of plasmid DNA*

Overnight starters cultures were prepared by inoculating LB media (with kanamycin) with a single colony obtained from transformation into NEB-5 alpha cells. These cells were harvested by centrifugation and used for plasmid DNA extraction. This was carried out with the Wizard Plus SV minipreps DNA purification system (Promega) following the manufacturers instructions. Pure plasmid DNA was eluted in 30 µl of nuclease free water and plasmid sequences were validated by Sanger sequencing using the express sequencing service provided by Beckman Coulter Genomics.

## 2.14 Site-directed mutagenesis

Different point mutants were generated by site-directed mutagenesis using the QuikChange approach (Agilent Technologies). PCR mutagenesis reactions were carried out using respective templates and different oligonucleotides carrying the mutated codons as primers to generate the desired mutant (Table 2.3). PCR reactions were carried out with 0.25 µl Q5 high-fidelity DNA polymerase, 5 µl Q5 reaction buffer (5x), 1.25 µl of primer (only one primer used), 0.5 µl template, 1 µl dNTPs and made up to a total reaction volume of 25 µl with nuclease free water. PCR reactions were performed in a DNA Engine Tetrad thermal cycler (Bio Rad) with the following settings: a 30 second activation step at 95°C, followed by 25 cycles of denaturation (30 seconds at 95°C), annealing of primers (1 minutes at 55°C) and elongation (68°C for 6 minutes). The reaction was ended with a 7 minute step at 68°C. The PCR product was then

digested with *Dpn*I for 2 hours at 37°C to remove parental DNA. The *Dpn*I digested product was transformed into either high efficiency NEB 10-beta calcium competent *E. coli* (New England Biolabs) or XL-1 blue supercompetent *E. coli* (Agilent) as described above, followed by isolation of plasmid DNA from positive colonies. All mutant constructs were validated by Sanger sequencing prior to use (Beckman Coulter Genomics).

**Table 2.3- Primers designed to generate various point mutants using quick-change site directed mutagenesis.**

| **SGTA TPR** | Primer for site-directed mutagenesis |
|---|---|
| K160E/R164E | 5'- CCC GGC CTA CAG CGA GGC CTA CGG CGA GAT GGG CCT GGC GCT C -3' |
| **SGTA C-terminal domain** | |
| S218C | 5'- GGA GGC GTG GGC TGC TTC GAC ATC GCC GGC C -3' |
| S232C | 5'- GAA CAA CCC TGG CTT CAT GTG CAT GGC TTC G -3' |
| S235C | 5'- ATG AGC ATG GCT TGC AAC CTA ATG AAC AAT C -3' |
| S248C | 5'- GCA GCT CAT GTG CGG CAT GAT TTC GGG TGG C -3' |
| S252C | 5'- ATG TCC GGC ATG ATT TGC GGT GGC AAC AAC C -3' |
| S264C | 5'- AAC TCC CGG CAC CTG CCC CTC GCA GAA CGA C -3' |
| S272C | 5'- CCC TCG CAG AAC GAC CTG GCC TGC CTC ATC C -3' |
| S297C | 5'- GAT AGA GCA GCT CAG GTG CCA GAT CCG GAG T -3' |
| S307C | 5'- GGA CGC CCA GCG CCT GCA ACG ACG ACC AGC A -3' |
| A281C | 5'- GGG CCA GCA GTT TTG CCA GCA GAT GCA GCA GC -3' |
| **Rpn13 C-terminal domain** | |
| K398C | 5'- CCA AAG ATA AAT GCG ACG AAG AGG AAG ATA TG -3' |

## 2.15 Heterologous expression and purification of recombinant proteins

Recombinant human proteins were expressed in *E. coli* as hexahistidine ($His_6$) tagged thioredoxin A (TxA) fusion constructs that were purified by nickel affinity chromatography followed by tag and TxA removal. In most cases, size-exclusion chromatography (SEC) was used as a polishing step to obtain pure protein. Purification of recombinant protein samples was achieved as described in the following steps.

### *2.15.1 Transformation*

The expression plasmids were transformed into *E. coli* Rosetta (DE3) competent cells (Novagen) by the heat shock method. Calcium competent *E. coli* cells (50 µl, $OD_{600}$ = 0.4) were incubated on ice for 30 min together with an addition of ~100 ng of plasmid DNA. The cells were then heat shocked at 42°C for 30 s. Cells were then recovered in 500 µl SOC outgrown media at 37°C for 1 hour on a shaking platform. The cells were subsequently spread onto LB agar plates and incubated overnight at 37°C with kanamycin (50 µg/ml) used for selection.

### *2.15.2 Preparation of overnight pre-cultures*

For unlabelled protein expression, a few colonies arising from the transformations were picked and used to inoculate 10 mL LB media with antibiotics, which was then incubated at 37°C overnight. This resulting overnight pre-culture was used to inoculate 1L of LB media with antibiotics.

To prepare isotopically labelled samples, colonies from transformations were used to inoculate 50 mL LB media with antibiotics (50 µg/ml), which was then incubated at 37°C overnight. The cells were harvested by centrifugation (4, 000 × g; 15 min; 4°C) then resuspended in 50 mL complete M9 media. The resuspended cells were used to

inoculate 1 L of complete M9 media supplemented with either $^{15}NH_4Cl$, or $^{15}NH_4Cl$ together with $^{13}C_6H_{12}O_6$, to express isotopically ($^{15}N$ or $^{13}C/^{15}N$) labelled recombinant proteins.

To prepare deuterated samples, a few colonies from the transformations were picked and used to inoculate 10 mL LB media with antibiotics, which was then incubated at 37°C overnight. This pre-culture (200 µl) was used to inoculate 50 ml of filter sterilised complete M9 media prepared in 100% $D_2O$, which was then incubated at 37°C overnight. The cells were harvested by centrifugation and resuspended in 50 mL complete M9 media prepared in 100% $D_2O$. These cells were used to inoculate 500 mL of complete M9 media prepared in 100% $D_2O$ supplemented with either $^{15}NH_4Cl$, or $^{15}NH_4Cl$ together with $^{13}C_6H_{12}O_6$ to achieve high levels of deuteration of labelled proteins.

### *2.15.3 Expression of recombinant proteins*

The cells were grown in 1 L cultures at 37°C until their density had reached an absorbance of around 0.4 - 0.6 at 600 nm. Protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.3-0.5 mM. Post induction, recombinant protein was expressed overnight at 18°C, then harvested by centrifugation (5, 000 rpm; 15 min; 4°C) the following day using a JLA-8.1000 rotor (Beckman Coulter).

### *2.15.4 Purification of recombinant proteins by nickel affinity chromatography*

All subsequent steps were performed at 4°C. The cells were re-suspended in 50 ml buffer A [20 mM potassium phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, 250 µM TCEP] with protease inhibitors tablets (complete mini, EDTA-free, Roche) and 1

μg/ml DNase I. All buffers were filtered (pore size 0.2 μm, Sartorius Stedim Biotech) and cooled to 4°C prior to use. The cells were lysed by cell disruption (Constant Systems Ltd.). Cell membranes and insoluble material were removed by centrifugation (17000 rpm; 1 h; 4°C) using a JA25.50 rotor (Beckman Coulter). The resulting supernatant was filtered through a syringe-driven Millipore express membrane (pore size 0.45 μm) and cleared lysate was applied to a 5 ml Hi-Trap nickel affinity column (GE Healthcare) pre-equilibrated in buffer A, at a flow rate of 2 ml/min. Recombinant protein was eluted with a 300 mM imidazole step using 100% buffer B [20 mM potassium phosphate pH 8.0; 300 mM NaCl; 300 mM imidazole; 250 μM TCEP]. The nickel column eluate (100% B) was treated with 100 μM PMSF and 0.5 mM EDTA. This was followed by the addition of rTEV protease to remove the TxA fusion protein and $His_6$ tag. Samples with rTEV were dialysed overnight against 4 L of buffer A using Spectra/Por or Spectra/Por4 dialysis membrane (Spectrum Laboratories, Inc.). The overnight dialysed sample was then loaded on to a 5 ml Hi-Trap nickel affinity column equilibrated in buffer A, and the flow through from this column was collected that contained the desired protein devoid of TxA fusion protein and $His_6$ tag (reverse nickel chromatography step).

### *2.15.5 Size-exclusion chromatography (SEC)*

The purest fractions were identified by SDS-PAGE then subjected to centrifugal concentration (Vivaspin 20, Sartorius Stedim Biotech) prior to size-exclusion chromatography (SEC). SEC was carried out on either a 120 ml HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare), or a 120 ml HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare) pre-equilibrated in gel filtration buffer [10 mM potassium phosphate pH 6.0; 100 mM NaCl; 250 μM TCEP] at a flow rate of 1 ml/min.

### *2.15.6 Purification of recombinant tail-anchored proteins*

To purify full length human tail-anchored proteins cytochrome b5 (Cytb5) and synaptobrevin 2 (Syb2) cell pellets were resuspended in 50 ml lysis buffer [50 mM HEPES pH 7.5; 300 mM NaCl; 10 mM imidazole; 5% glycerol; 1% DDM] with 100 μM PMSF, 1 μg/ml DNase I, and protease inhibitors tablets (complete mini, EDTA-free, Roche). The cells were lysed by incubation with 0.2 mg/ml lysozyme for 1 hr at room temperature on a rotating platform. Lysate was cleared by centrifugation (17000 rpm; 1 h; 4°C) using a JA25.50 rotor (Beckman Coulter). The supernatant was filtered through Millipore express membrane (pore size 0.45 μm) and applied to a 5 ml nickel affinity column equilibrated in TA buffer A [50 mM HEPES pH 7.5; 300 mM NaCl; 10 mM imidazole; 250 μM TCEP; 0.1% DDM]. Elution of $His_6$-tagged recombinant TA proteins with N-terminal TxA fusion protein was carried out by an imidazole step using TA buffer B [50 mM HEPES pH 7.5; 300 mM NaCl; 300 mM imidazole; 250 μM TCEP; 0.1% DDM]. The TxA fusion protein and $His_6$ tag were removed by overnight rTEV digestion. Full-length TA proteins were obtained by reverse nickel chromatography as described above, and sample purity was assessed by SDS-PAGE.

## 2.16 Analytical size-exclusion chromatography

Analytical size-exclusion chromatography (Analytical SEC) can be used to separate proteins based on size. Analytical SEC was performed on either a 25 ml Superdex 200 10/300 GL column (GE Healthcare) or a 120 ml HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with 10mM KPi, 100mM NaCl at pH 6.0. The flow rate was maintained at 1 ml/min. Where necessary molecular mass was estimated based on the migration of protein standards on the SEC column (aprotinin – 6.5 kDa, ribonuclease A – 13.7 kDa, carbonic anhydrase – 29.0 kDa, ovalbumin – 44.0 kDa, conalbumin – 75.0 kDa and aldolase – 158.0 kDa; GE Healthcare).

## 2.17 Solution NMR spectroscopy

Solution NMR spectroscopy is a widely used technique to understand the structure and dynamics of proteins in solution. It is routinely used to characterise protein-protein and protein-ligand interactions, and to investigate biomolecular motions. In addition, NMR spectroscopy can be used to determine atomic resolution structures of small proteins, and to probe structures of larger proteins. Furthermore, NMR can be used to probe protein structure and interactions within living cells. Also, recent years have seen an increase in the use of protein NMR in the discovery of novel therapeutics (Barrett *et al.*, 2013).

Solution NMR experiments were carried out on Bruker Avance spectrometers operating at 400, 500 and 700 MHz at the Centre for Biomolecular Spectroscopy, King's College London. The 500 MHz and 700 MHz spectrometers were equipped with a triple resonance ($^{1}H/^{13}C/^{15}N$) cryoprobe with z-gradients, and a quadruple resonance ($^{1}H/^{13}C/^{15}N/^{31}P$) cryoprobe with z-gradients, respectively. The 400 MHz Bruker Avance spectrometer was equipped with a $^{1}H/\{^{13}C,^{19}F,^{31}P\}$ QNP probe. Further NMR experiments were carried out on the 950 MHz Bruker spectrometer, equipped with a cryoprobe at the MRC Biomedical NMR Centre, The Francis Crick Institute, London. All NMR experiments were recorded at 25°C. The spectrometers were controlled by Topspin version 3.1 (Bruker Biospin Ltd). NMRPipe (Delaglio *et al.*, 1995) and CcpNMR Analysis (Skinner *et al.*, 2015; Vranken *et al.*, 2005) were used for spectral processing and analysis. A summary of acquisition parameters used in key 1D and 2D NMR experiments described in this work is shown in Table 2.4.

**Table 2.4- NMR acquisition parameters**

| Experiment | Temp. | Number of Scans | Spectrometer frequency | Sweep width | | Number of Points | | Dwell time | Recycle delays |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | | |
| 1D $^{1}H$ (zgesgp) | 25°C | 128 | 500 MHz | 16.02 ppm | n/a | 4096 | n/a | 62.4 µs | 1 sec |
| 2D $^{1}H$-$^{15}N$ HSQC (hsqcwg) | 25°C | 4 | 500 MHz | 16.02 ppm | 35 ppm | 2048 | 256 | 62.4 µs | 1 sec |
| | | 4 | 700 MHz | 16.08 ppm | 35 ppm | 2048 | 256 | 44.4 µs | 1 sec |
| | | 4 | 950 MHz | 16.08 ppm | 35 ppm | 2048 | 256 | | |
| 2D $T_1$ (hsqct1etf3gpsi) | 25°C | 8 | 500 MHz | 16.02 ppm | 35 ppm | 2048 | 200 | 62.4 µs | 3 sec |
| | | 16 | 700 MHz | 16.08 ppm | 35 ppm | 2048 | 200 | 51.0 µs | 3 sec |
| | | 16 | 950 MHz | 16.08 ppm | 35 ppm | 2048 | 200 | | |
| For $T_1$ measurements a series of spectra with 30.8, 61.6, 123.2, 246.4, 369.6, 554.4, 739.2, 985.5, 1232, 1386 and 1540 ms delays were recorded. | | | | | | | | | |
| 2D $T_2$ (hsqct2etf3gpsi) | 25°C | 8 | 500 MHz | 16.02 ppm | 35 ppm | 2048 | 200 | 62.4 µs | 3 sec |
| | | 16 | 700 MHz | 16.08 ppm | 35 ppm | 2048 | 200 | 51.0 µs | 3 sec |
| | | 16 | 950 MHz | 16.08 ppm | 35 ppm | 2048 | 200 | | |
| For $T_2$ measurements 16.96, 33.92, 50.88, 67.84, 84.8, 118.72, 152.64, 186.56, 220.48 and 254.4 ms CPMG echo delays were recorded. | | | | | | | | | |
| 2D $^{1}H$-$^{15}N$ Heteronuclear NOE (hsqcnoef3gpsi) | 25°C | 24 | 500 MHz | 16.02 ppm | 35 ppm | 2048 | 256 | 62.4 µs | 4 sec |
| | | 64 | 700 MHz | 16.08 ppm | 35 ppm | 2048 | 256 | 44.4 µs | 4 sec |
| | | 64 | 950 MHz | 16.08 ppm | 35 ppm | 2048 | 256 | | |
| $^{19}F$ 1D (zgflqn) | 25°C | 1024 | 400 MHz | $^{19}F$ 265.63 ppm | | $^{19}F$ 65536 | | 5.0 µs | 1 sec |

### *2.17.1 1D $^{1}H$ NMR*

Proton 1D NMR shows all observable protein signals in the sample. For proteins, these signals are highly overlapped. However, these experiments can still provide information on a protein's degree of structure, or lack thereof. For instance, poor amide dispersion together with an absence of ring current shifted methyl signals is indicative of an unstructured protein. Conversely, good dispersion of amide signals alongside ring current shifted methyl signals can be used as an indication of a structured protein.

### *2.17.2 2D $^{1}H$-$^{15}N$ HSQC experiments*

The $^{1}H$-$^{15}N$ HSQC spectrum of a protein can be described as a 'fingerprint' of its backbone and side chain amide groups. In other words, a cross-peak can be obtained from the amide group of every amino acid except proline. As the HSQC is extremely sensitive to ligand binding, it is widely used to detect and characterise protein-ligand and protein-protein interactions. $^{15}N$ labelled proteins are used routinely in NMR titration experiments to observe changes in the $^{1}H$-$^{15}N$ HSQC spectrum of a protein upon titration with increasing amounts of unlabelled proteins in order to characterise binding, and map resulting binding interfaces. In addition, 3D experiments used for sequential backbone assignment of proteins are built upon a $^{1}H$-$^{15}N$ HSQC with an added dimension corresponding to another nucleus (typically $^{13}C$) (Campbell, 2012).

The HSQC experiment was first described by Bodenhausen and Ruben (1980), and is based on two INEPT (insensitive nuclei enhanced by polarization transfer) type magnetization transfer steps between proton and directly attached heteroatoms (such as $^{15}N$) via a large scalar coupling constant ($J_{H\text{-}N}$ = 90-100 Hz) separated by $t_1$ chemical shift evolution periods. A series of experiments are recorded each with increments in $t_1$, which give rise to the indirect dimension ($^{15}N$ in this case). The $^{1}H$ signal is acquired as

the directly measured dimension of each experiment (Campbell, 2012; Bodenhausen and Ruben, 1980).

### *2.17.3 NMR sample conditions*

Unless stated otherwise, protein samples for solution NMR spectroscopy were dialysed into NMR buffer [10 mM potassium phosphate pH 6.0; 100 mM NaCl; 250 µM TCEP] by overnight dialysis or gel filtration chromatography. Protein samples at concentrations between 200-800 µM were prepared with 10% $D_2O$ (Sigma Aldrich) in 5 mm NMR tubes (VWR International).

### *2.17.4 NMR titrations*

Proteins used for NMR titrations were dialysed overnight against NMR buffer (10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 µM TCEP) and combined in different molar ratios. $^{1}H$-$^{15}N$ HSQC experiments were recorded with a 20 min acquisition for each titration point. Chemical shift perturbation (CSP) values were calculated for each amide signal using the following formula:

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

where $\Delta\delta_{1H}$ and $\Delta\delta_{15N}$ are the chemical shift differences for the same amide in its free and bound state ($\delta_{free}$-$\delta_{bound}$) and for proton and nitrogen chemical shift values respectively. CSP results were mapped on to crystal structures using the PyMOL software (Schrödinger, LLC).

### *2.17.5 Triple resonance experiments for sequential backbone assignment*

Sequential backbone assignment of the SGTA C-terminal domain was carried out using information pertaining to connectivity between amino acids obtained from the following standard triple resonance experiments: HNCO, HN(CA)CO, HN(CO)CA, HNCA, CBCA(CO)NH, and CBCANH (Clubb *et al.*, 1992; Grzesiek and Bax, 1992a; Grzesiek and Bax, 1992b; Grzesiek and Bax, 1992c). These 3D experiments were carried out on $^{13}C/^{15}N$ labeled protein samples. Assignment of backbone amide resonances of the SGTA C-terminal domain was carried out by Dr Santiago Martinez-Lumbreras (KCL).

### *2.17.6 $^{15}N$ NMR relaxation experiments*

$^{15}N$ relaxation experiments are routinely used to characterise molecular motions of proteins in solution. These experiments rely on the generation of a non-equilibrium spin order, and subsequent observation of how this returns to equilibrium. The relaxation parameter $T_1$ describes spins that relax back to equilibrium along the direction of the external magnetic field, and thus referred to as the longitudinal relaxation time. When the magnetization is ordered perpendicular to the external magnetic field, the relaxation time of this spin-order to return back to equilibrium is referred to as the transverse relaxation time $T_2$. $^{15}N$ relaxation experiments can provide information regarding the overall correlation time of a protein (calculated from $T_1$ and $T_2$ values) (Wagner, 1995; Campbell, 2012). Additionally, there is another relaxation parameter referred to as the heteronuclear nuclear Overhauser effect (NOE). This parameter is measured by saturating the proton signal and detecting changes in the $^{15}N$ signal. Heteronuclear NOE values provide valuable information regarding the presence of rapid motion due to local flexibility, as negative values are indicative of high frequency molecular motions (Wagner, 1995).

For $T_1$ measurements a series of spectra with 30.8, 61.6, 123.2, 246.4, 369.6, 554.4, 739.2, 985.5, 1232, 1386 and 1540 ms delays were recorded. On the other hand, spectra series with 16.96, 33.92, 50.88, 67.84, 84.8, 118.72, 152.64, 186.56, 220.48 and 254.4 ms CPMG echo delays were recorded for $T_2$ measurements. $T_1$ and $T_2$ relaxation times were calculated from the single exponential decay fitting of peak intensities of each amide signal. Rotational correlation times ($\tau_c$) of various different constructs have been calculated from the $T_1/T_2$ averaged values using the following equation:

$$\tau_c \approx \frac{1}{4\pi\upsilon_N}\sqrt{6\frac{T_1}{T_2}-7}$$

where $\nu_N$ is the spectrometer frequency (Kay *et al.*, 1989). The rotational correlation time ($\tau_c$) is roughly the time it takes for a molecule to rotate by one radian.

In addition, $\{^1H\}^{15}N$ heteronuclear NOE experiments were performed with a four second saturation transfer or control period. The $\{^1H\}^{15}N$ heteronuclear NOE values are given by:

$$\{^1H\}^{15}N \text{ heteronuclear NOE} = \frac{I_{sat}}{I_{unsat}}$$

where $I_{sat}$ and $I_{unsat}$ are the cross-peak intensities observed during the saturation transfer and control periods, respectively (Kay *et al.*, 1989).

### *2.17.7 PRE experiments*

It is known that the magnetic moment of an unpaired electron is hugely greater (658-fold) than that of a $^1H$ nucleus (Campbell, 2012). As a result, distance-dependent line broadening of NMR signals can be observed in protein spectra in the presence of paramagnetic species (Battiste and Wagner, 2000). These paramagnetic relaxation

enhancement (PRE) effects can provide information about molecular motions and distances between the paramagnetic centre and the observed nucleus. Paramagnetic spin labelling of proteins can be achieved by modifying free cysteines with thiol-reactive nitroxide spin-labeled compounds, and the resulting PRE effect can be detected over distances of around 0 to 25 Å for a proton experiencing relaxation enhancement from a nitroxide spin-label (Barrett *et al.*, 2013).

Intermolecular PRE effects were recorded to understand details pertaining to the binding interface between Rpn13 and SGTA_TPR. For spin labelling, a C-terminal residue of Rpn13 (Lys 398) was mutated to a cysteine, which allowed for the covalent attachment of both paramagnetic and diamagnetic probes. The thiol-specific 1-oxyl-2,2,5,5-tetramethylpyrroline-3- methyl methanethiosulfonate (MTSL) was used as the paramagnetic probe, and the structurally similar 1-acetyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl methanethiosulfonate (dMTSL) was used as a diamagnetic control. Unlabelled mutant $Rpn13_{260-407}$ was treated with excess TCEP for 2 hours to ensure cysteines were reduced, followed by dialysis against 10 mM potassium phosphate pH 6.0, 100 mM NaCl buffer to remove excess TCEP. Mutant samples were then incubated with either diamagnetic (dMTSL) or paramagnetic (MTSL) spin labels (Santa Cruz Biotechnology) overnight at 4°C. Excess dMTSL/MTSL was removed by extensive dialysis against 10 mM potassium phosphate pH 6.0, 100 mM NaCl. For intermolecular PRE experiments, $^{15}$N-labelled SGTA_TPR in 10 mM potassium phosphate pH 6.0, 100 mM NaCl buffer was mixed with each spin labelled Rpn13 sample at a 1:1.2 ratio. A $^{1}$H-$^{15}$N HSQC spectrum was acquired for each 200 μM sample with a 1 hour acquisition.

Proximity dependent line broadening of $^{15}$N-labelled SGTA_TPR backbone amide signals was monitored for intermolecular paramagnetic relaxation enhancement (PRE) effects upon addition of MTSL labelled K398C Rpn13$_{260\text{-}407}$. This would enable mapping of the position of the MTSL spin label within a 25 Å distance of amide signals corresponding to SGTA_TPR residues. PRE intensity ratios were calculated as follows:

$$\text{Intensity ratio} = \frac{I_{para}}{I_{dia}}$$

where $I_{para}$ and $I_{dia}$ are peak intensities measured in the presence of paramagnetic (MTSL) and diamagnetic (dMTSL) probes, respectively.

### *2.17.8 Fluorine-19 NMR*

Fluorine-19 is a spin-½ isotope and has high sensitivity for observation by NMR, corresponding to 83% of $^{1}$H sensitivity (Didenko *et al.*, 2013). Thus 1D fluorine-19 NMR can be used to observe extrinsic $^{19}$F labels that have been incorporated into protein samples. In addition, fluorine-19 chemical shifts are extremely sensitive to the local environment, making $^{19}$F labels on proteins suitable probes to detect protein-protein interactions. Furthermore, in cases where regions on proteins cannot be observed by standard NMR methods, fluorine-19 NMR can provide a way to probe such regions. $^{19}$F probes can be attached to proteins post-translationally, such incorporation can be achieved by chemical conjugation of small molecules containing $^{19}$F with reactive thiol groups on cysteine residues (Horst *et al.*, 2013; Didenko *et al.*, 2013).

Fluorine-19 NMR has been used to investigate the interaction of TA proteins with the C-terminal domain of SGTA. Incorporation of $^{19}$F containing 3-bromo-1,1,1-trifluoroacetone (BTFA) (Fig. 2.2) to cysteine residues positioned at different locations of the SGTA C-terminal domain was carried out in a single-step process resulting in the

formation of a stable thioester bond. All SGTA C-terminal domain cysteine mutants in NMR buffer (10 mM potassium phosphate pH 6.0, 100 mM NaCl and 250 μM TCEP) were treated with 100 μM BTFA at 4°C overnight, then dialysed extensively against NMR buffer to remove traces of free BTFA. For spectra acquisition, 225 μl of the BTFA labeled protein sample was prepared in a 5 mm NMR tube. Free BTFA used for chemical shift referencing was prepared at 100 μM in $D_2O$ in a 3 mm NMR tube, which was placed within the 5 mm tube prior to acquisition. The same 3 mm tube was used as reference for all $^{19}F$ spectra of the different cysteine mutants. 1D $^{19}F$ spectra were recorded at 25°C on a 400 MHz Bruker Avance spectrometer, equipped with a $^{1}H/\{^{13}C,^{19}F,^{31}P\}$ QNP probe, with a 1 hour acquisition for each experiment.

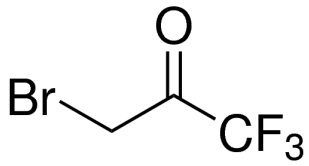



**Figure 2.2- 3-bromo-1,1,1-trifluoroacetone (BTFA).** BTFA was incorporated to cysteine residues positioned at different locations of the SGTA C-terminal domain.

## 2.18 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a technique that can be used to accurately determine binding constants and thermodynamic parameters underlying molecular interactions. ITC provides a direct measurement of heat generated or absorbed as the result of an interaction. This technique is therefore used to obtain the Gibbs free energy (ΔG), which can be subsequently decomposed into enthalpy (ΔH) and entropy (ΔS) thereby providing a complete picture of the thermodynamic profile associated with a binding event (Velazquez-Campoy *et al.*, 2004). Isothermal titration calorimeters typically consist of a reference and a sample cell, both contained within a thermal

insulation jacket. The reference cell contains water or buffer, while the sample cell holds one of the binding partners and a stirring syringe which injects the other binding partner. Heat changes are monitored by measuring the differential power required to maintain a zero temperature difference (i.e. an isothermal condition) between the reference and sample cells, as the binding partners are mixed with each other (Campbell, 2012; Velazquez-Campoy *et al.*, 2004).

ITC experiments were performed using an ITC-200 microcalorimeter from Microcal (GE Healthcare) at 25°C following the manufacturers protocol. Proteins or peptide samples were prepared in 10 mM potassium phosphate pH 6.0, 100 mM NaCl, 250 μM TCEP. In each titration, either 40 injections of 1 μL or 20 injections of 2 μL of SGTA_TPR each, at a concentration of 1 mM, were added to a sample of Rpn13 or the Rpn13 derived DMSLD peptide at a concentration of 50 μM in the sample cell. A nonlinear least-squares minimization algorithm was applied to integrated heat data obtained for the titrations, corrected for heats of dilution, in order to fit the experimentally obtained values to a theoretical titration curve. This was performed using the MicroCal-Origin 7.0 software package. ΔH (reaction enthalpy change in kcal/mol), $K_b$ (equilibrium binding constant in per molar), and n (molar ratio between the proteins in the complex) were used as fitting parameters. The reaction entropy, ΔS, was calculated using the relationships $\Delta G = -RT \ln K_b$ ($R$ = 8.314 J/(mol K), T = 298 K) where $\Delta G = \Delta H - T\Delta S$. Dissociation constants ($K_d$) have been determined for each interaction. Binding was assumed to be at one site (n = 1), to determine the binding affinity ($K_d$) and thermodynamic parameters.

## 2.19 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is based on the differential absorption of protein samples upon interaction with left- and right- circularly polarized light. This can be expressed as

$$\Delta A = A_L - A_R = \Delta \varepsilon c l$$

where $A_L$ and $A_R$ are the absorbances of the left- and right circularly polarized beams, respectively, $\Delta\varepsilon$ is the differential absorption coefficient, c is sample concentration, and l is the path length through the sample. In addition, the differential absorption coefficient is given by $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$ for left- and right circularly polarized beams. Molar ellipticity ($\theta$) is the parameter commonly used to report CD measurements, with $\theta = 3298\Delta\varepsilon$ (degree $cm^2$ $mol^{-1}$). In particular for proteins, mean molar ellipticity per residue is reported, which is given by $\theta$ divided by the number of amino acids in the protein (Campbell, 2012).

For proteins, the spectral region ranging from 170-250 nm is of particular interest as it is dominated by the polypeptide backbone. As the peptide bond is one of the main chromophores of interest in CD spectroscopy, the 170-250 nm region can be used to estimate the secondary structure of proteins. Thus characteristic spectral signatures corresponding to protein secondary structural elements can be obtained by CD spectroscopy (Campbell, 2012; Greenfield, 2007).

CD spectra of various SGTA constructs were acquired using Aviv circular dichroism spectrophotometer, model 410 (Biomedical Inc., Lakewood, NJ, USA). Protein samples were adjusted to 0.5 mg/ml in 10 mM KPi, 100mM NaCl, pH 6.0 buffer and the experiments were recorded using a rectangular demountable Suprasil quartz cell of 0.1

mm pathlength (Hellma Analytics). Each sample was scanned three times from 260 to 190 nm, at 1-nm intervals with an averaging time of 0.5 s. After background subtraction, data was converted to mean molar ellipticity per residue.

### 2.20 Native mass spectrometry

Native mass spectrometry can be used to determine the mass of protein assemblies and their complexes, held together by non-covalent interactions. This approach can therefore be used to detect several different species in a mixture, such as polydisperse samples, and can provide information on the stoichiometry of protein complexes. Furthermore, non-dissociative nano-electrospray ionization mass spectrometry (Native MS) allows for the study of proteins in neutral aqueous buffers in which their structural assembly is preserved (Hilton and Benesch, 2012; van den Heuvel and Heck, 2004).

Native MS of SGTA samples (full-length and C-terminal domain) was carried out by Dr Dijana Matak-Vinkovic at the Department of Chemistry, University of Cambridge. Experiments were carried out on a Synapt HD mass spectrometer (Waters) modified for studying high masses. Protein samples were exchanged into 0.20–0.75 M ammonium acetate (pH 7.0) solution using Micro Bio–Spin 6 chromatography columns (Bio Rad) and diluted to a final concentration of 5–10 μM before analysis. The volume of 2.5 μL of protein solution was electrosprayed from a borosilicate emitter (Thermo Scientific) for sampling. Typical conditions were capillary voltage 1.8–2.5 kV, cone voltage 60–120 V, collision voltage 10−30 V, with backing pressure 3–4 mbar and source temperature of 20 °C. Spectra were calibrated externally using cesium iodide. Data acquisition and processing were performed using MassLynx 4.1. (Waters).

## 2.21 High ambiguity driven protein-protein docking

The high ambiguity driven docking approach (HADDOCK) makes use of protein-protein interaction data obtained from NMR experiments and/or mutagenesis studies, to generate models of protein complexes. In this approach, information on the interacting residues within a complex is introduced as ambiguous interaction restraints (AIRs), which in turn drives the docking. Then the structures are ranked according to their intermolecular energy, which is given by the sum of electrostatic, van der Waals, and AIR energy terms (Dominguez *et al.*, 2003). A HADDOCK approach has been adopted to obtain a protein-peptide complex of the SGTA_TPR domain with the human Rpn13 derived DMSLD peptide based on experimental restraints derived from NMR CSP studies combined with mutagenesis experiments.

The PDB-deposited structure of SGTA_TPR (accession code 2VYI) and the coordinates of the DMSLD pentapeptide from the solution NMR structure of Rpn13 (accession code 2KR0) were used for the protein-peptide complex structure calculation. Ambiguous interaction restraints (AIRs) were implemented using the standard HADDOCK protocol (van Zundert *et al.*, 2016; Dominguez *et al.*, 2003). CSP experiments were used to identify 10 amino acid residues in SGTA_TPR and 3 within the DMSLD peptide with chemical shift changes greater than 0.10 ppm. The Naccess program was used to determine relative solvent accessibility, and residues with higher that 45% values were identified as active (Hubbard and Thornton, 1993). Based on this approach, 10 residues in SGTA_TPR and 3 in the DMSLD peptide, were identified as active. These were SGTA_TPR residues 99, 102, 107, 130, 133, 161, 163, 168, 171, 198 and Rpn13 residues 405, 406 and 407 (for the DMSLD peptide). Solvent exposed residues juxtaposed to active residues were automatically designated as passive residues. Rigid

body energy minimization was used to generate one thousand initial complex structures, from which the best 200 (lowest total energy) were selected for torsion angle and Cartesian dynamics in an explicit water solvent. Default scaling was applied for energy terms. Following the standard protocol, cluster analysis generated 130 structures in 11 cluster ensembles. The top-scoring cluster (lowest energy) was considered as the most reliable result as determined by HADDOCK benchmark testing (van Zundert *et al.*, 2016; Dominguez *et al.*, 2003).

# Chapter 3

# Biophysical characterisation of SGTA

Structural and biophysical studies of SGTA carried out thus far have been limited to its excised N-terminal (Darby *et al.*, 2014) and TPR domains (Dutta and Tan, 2008), with molecular details pertaining to its crucial substrate binding C-terminal domain yet to be elucidated. Furthermore, structural and dynamic properties of SGTA in the context of its full-length assembly are yet to be understood. This chapter presents a detailed biophysical characterisation of full-length SGTA using solution NMR spectroscopy, circular dichroism (CD) spectroscopy and non dissociative nano-electrospray ionization mass spectrometry (Native MS), with an aim to understand SGTA assembly *in vitro*.

## 3.1. Expression and purification of SGTA constructs

To glean insights into the domain organisation and conformations of SGTA *in vitro*, in particular within the context of its full-length assembly, various SGTA constructs have been designed, encoding individual SGTA domains and combinations thereof (Fig. 3.1). Constructs were designed based on sequence homology, and include the extensively characterised N-terminal and TPR domains, followed by the C-terminal domain. The C-terminal domain can be further divided into conserved NNP and Q-rich regions (Fig 3.1).

SGTA constructs were recombinantly expressed in *E. coli* Rosetta cells as fusion proteins with N-terminally placed thioredoxin A (TxA) followed by a hexa-histidine ($His_6$) tag and a rTEV cleavage site. These SGTA fusion proteins were overexpressed at 18°C overnight, and were all recovered in the soluble fraction (Fig. 3.2), allowing for purification by Ni-affinity chromatography (Figs. 3.3 and 3.4).

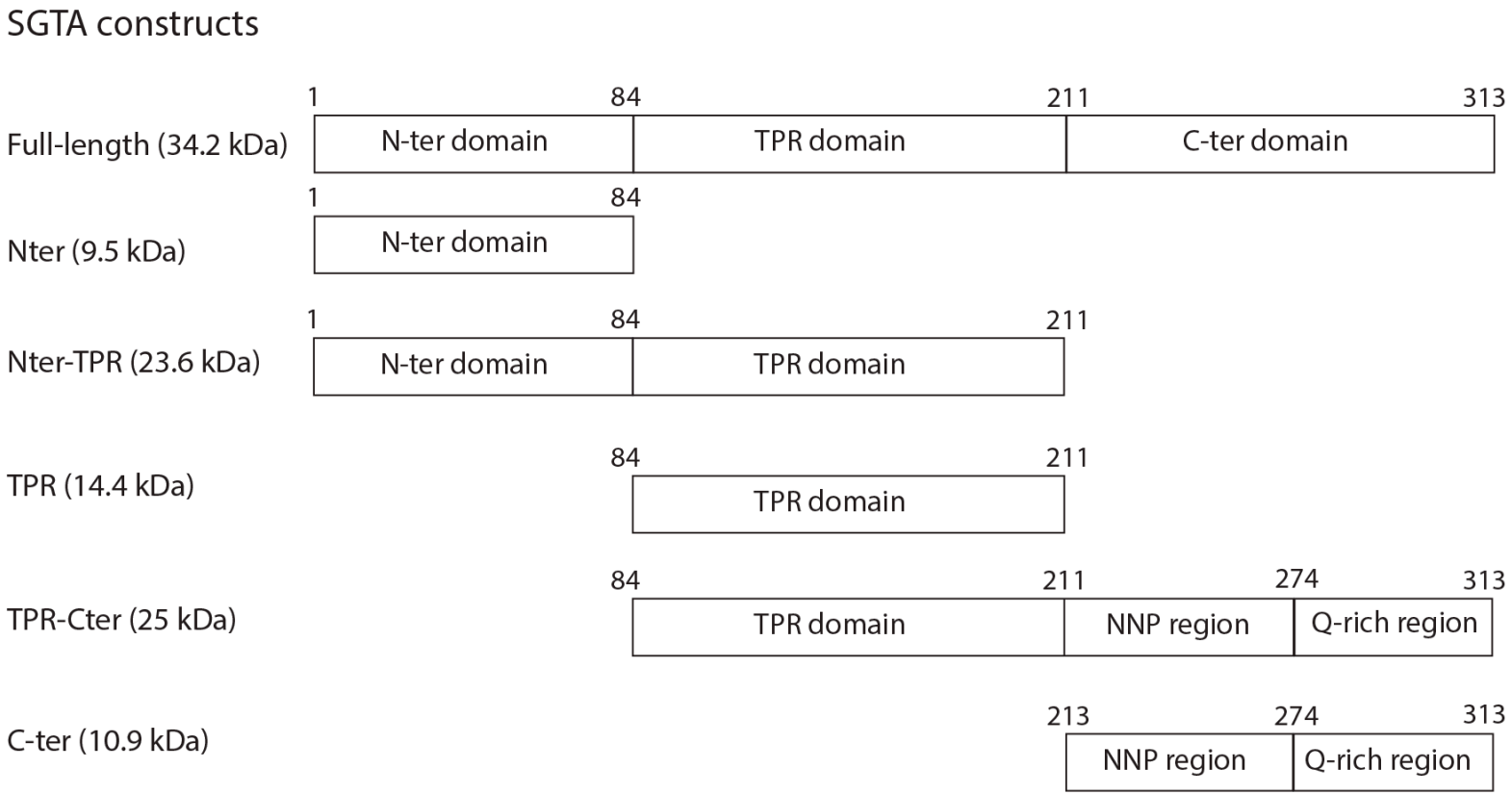



**Figure 3.1- Schematics of various SGTA constructs.** SGTA constructs used in this study are shown with molecular weights of individual protein chains (protomers in the case of dimeric species) together with domain boundaries of each construct. Constructs with the N-terminal domain are known to form dimers.

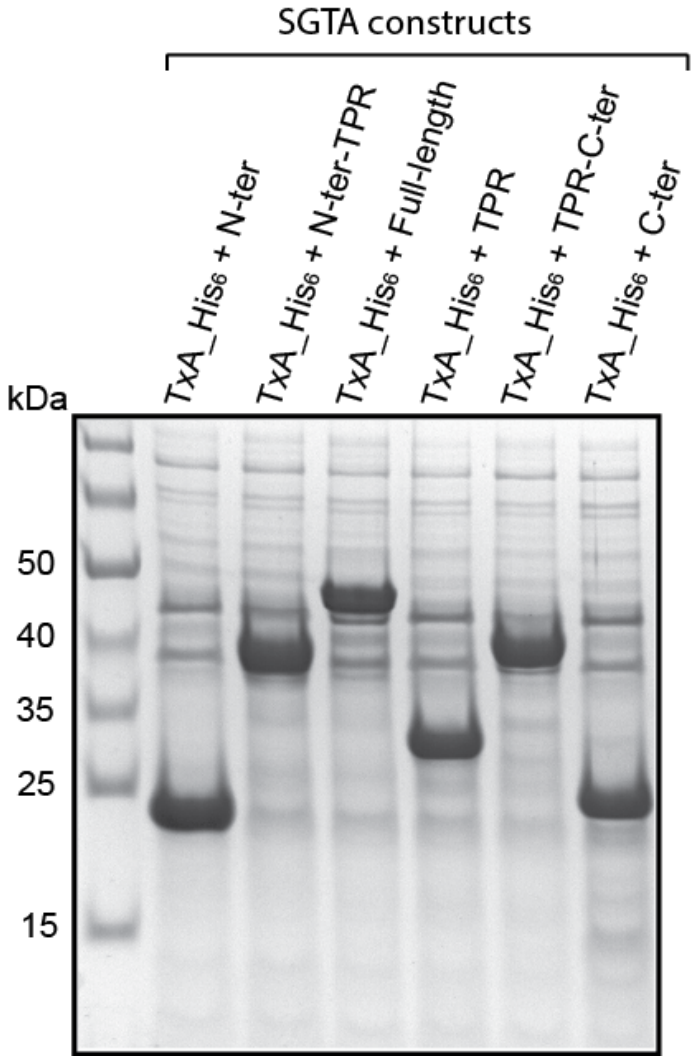



**Figure 3.2- Recombinant expression of the human SGTA constructs in *E. coli.*** SDS-PAGE showing soluble fractions of overexpressed SGTA constructs. Various SGTA constructs including excised individual domains and combinations thereof were expressed in *E. coli* Rosetta cells as a fusion protein with an N-terminally placed thioredoxin A followed by a hexahistidine tag and rTEV cleavage site. All constructs were successfully overexpressed in both rich (LB) and minimal (M9) media at 18°C overnight.

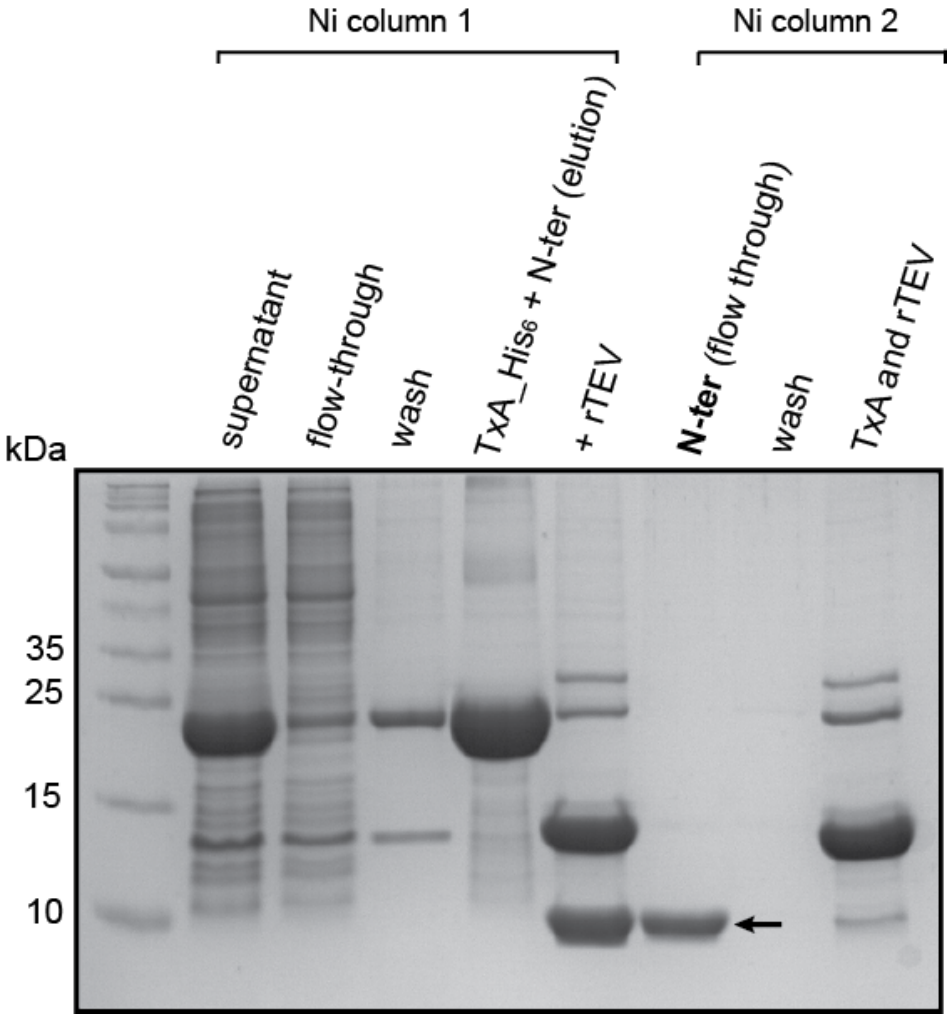



**Figure 3.3- Purification of the N-terminal domain of SGTA.** SDS-PAGE fractions of samples from different stages of Ni-affinity chromatography. Ni- column 1 was used to purify the SGTA construct as a fusion protein with TxA. After subsequent TxA and $His_6$ tag removal using rTEV, the N-terminal domain of SGTA was recovered in the flow-through of Ni- column 2. This purification strategy was used for all other SGTA constructs/variants described herein.

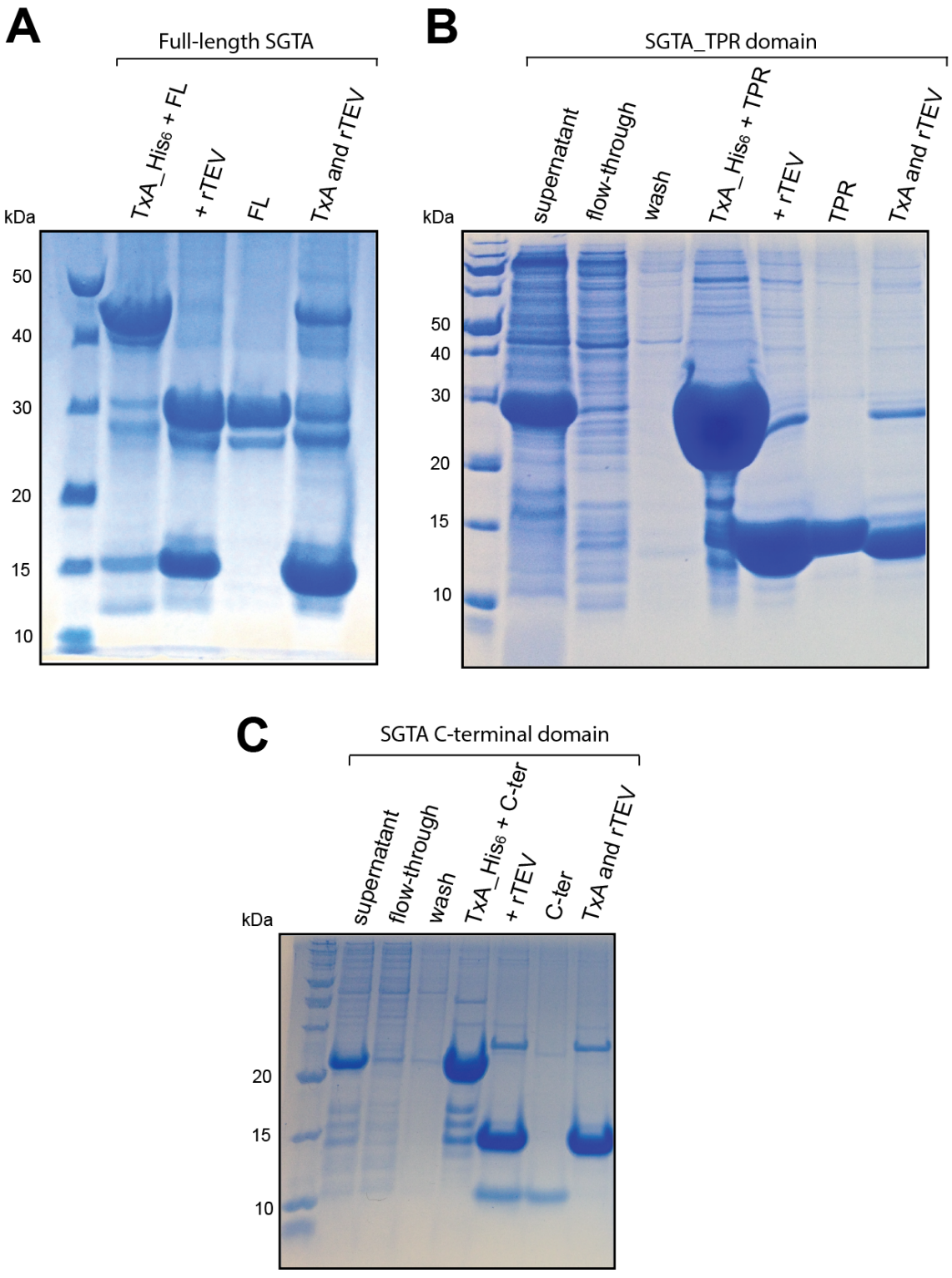



**Figure 3.4- Purification of SGTA constructs.** SDS-PAGE fractions of samples from different stages of Ni-affinity chromatography of (A) full-length (B) the TPR domain, and (C) the C-terminal domain of SGTA. All constructs were purified using a similar strategy as described for the N-terminal domain.

## 3.2. Analytical SEC of SGTA constructs

Different SGTA constructs were subjected to analytical size-exclusion chromatography (SEC) on a 25 ml Superdex 200 10/300 GL column at a flow rate of 1 ml/min. An overlay of the elution profiles of these constructs carried out under identical conditions (Fig. 3.5), reveal that constructs including the C-terminal domain of SGTA (FL and TPR-Cter) appear at lower elution volumes than expected, based on their size and oligomeric state. This is particularly evident with the TPR-Cter construct (residues 84-313; 25 kDa) that lacks the N-terminal dimerisation domain of SGTA, as it appears at an identical elution volume (~14 ml) to the Nter-TPR construct (residues 1-211; 47.2 kDa) that is known to form a dimer via its N-terminal domain (Fig. 3.5). In comparison, both Nter and TPR constructs migrate on the SEC column in a manner indicative of their presence as lower molecular weight species (Fig. 3.5). The isolated C-terminal domain was not suitable for comparison by this method due to a complete lack of aromatic residues. Therefore, based on analytical SEC, the presence of the C-terminal domain in SGTA constructs appears to increase the rate of migration of these constructs on the SEC column, resulting in a shift towards lower peak elution volumes, indicative of the formation of species corresponding to higher molecular weights.

## 3.3. Circular dichroism spectroscopy of SGTA constructs

Circular dichroism (CD) spectroscopy on all three isolated domains of SGTA presented spectra typical of alpha helical proteins with characteristic minimum ellipticity at 208 and 225 nm, and maximum ellipticity at around 190 nm (Fig. 3.6). These findings are in agreement with the solved structures of the N-terminal and TPR domains which are both predominantly alpha helical (Darby *et al.*, 2014; Dutta and Tan, 2008). Interestingly, with regard to the C-terminal domain of SGTA these results suggest, for the first time, the presence of a certain degree of α-helical propensity within this region (Fig. 3.6C).

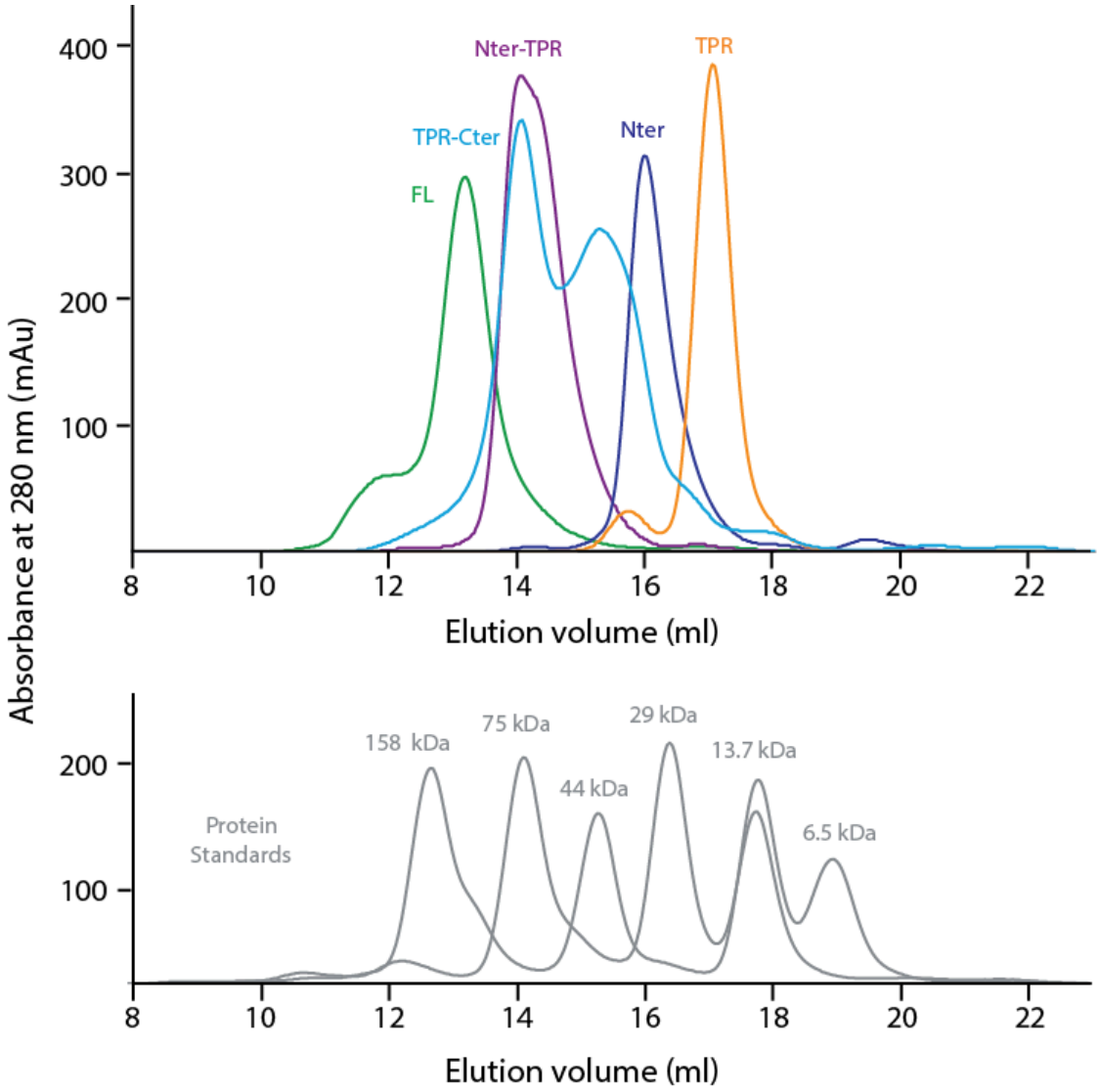



**Figure 3.5- Analytical size-exclusion chromatography (SEC) profiles of SGTA constructs.** Overlay of SEC chromatograms of different SGTA constructs recorded using a 25 ml Superdex 200 10/300 GL column at a flow rate of 1 ml/min. Due to a lack of aromatic residues the C-terminal domain construct was not used for comparison (extinction coefficient: 0). Molecular mass was estimated based on the migration of protein standards on the SEC column (aprotinin – 6.5 kDa, ribonuclease A – 13.7 kDa, carbonic anhydrase – 29.0 kDa, ovalbumin – 44.0 kDa, conalbumin – 75.0 kDa and aldolase – 158.0 kDa).

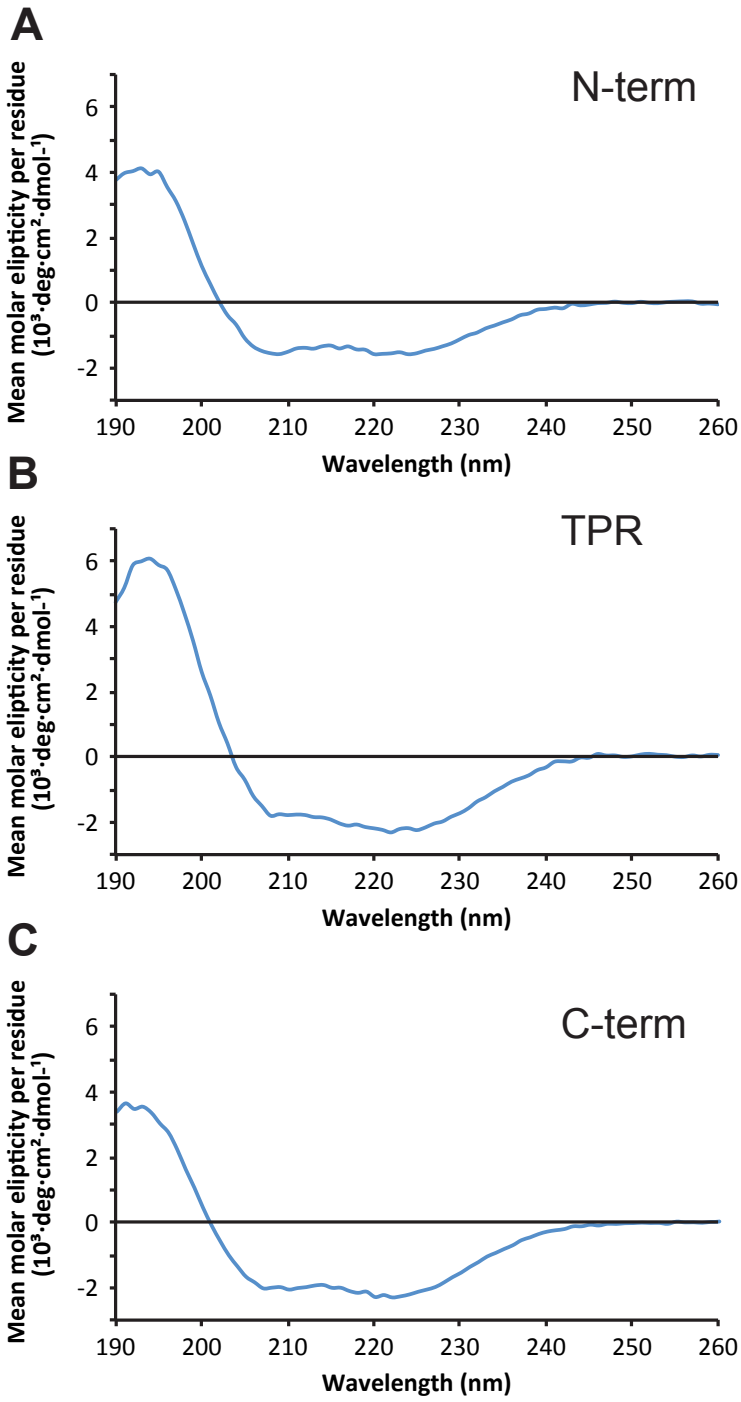



**Figure 3.6- Circular dichroism (CD) spectra of isolated (A) N-terminal, (B) TPR and (C) C-terminal domains of SGTA.** CD spectra show that all constructs present α-helical propensity, including the C-terminal domain.

## 3.4. Characterisation of SGTA constructs by solution NMR spectroscopy

1D $^{1}$H NMR spectra of Nter, TPR, and Nter-TPR constructs (Fig. 3.4) demonstrate spectral features typical for folded proteins, which include ring current shifted methyl signals (typically between 1 and -1 ppm) and dispersion of amide signals (typically higher than 9 ppm). This data is in agreement with with previously determined high resolution structures of these domains (Darby *et al.*, 2014; Dutta and Tan, 2008). On the other hand, the 1D $^{1}$H NMR spectrum of the C-terminal domain construct presents poor dispersion of amide signals, together with an almost complete lack of ring current shifted methyl signals (Fig. 3.7). This finding indicates the possibility that the C-terminal domain of SGTA consists of unstructured regions.

To understand the arrangement of individual domains within the context of full-length SGTA, a series of comparisons were carried out involving $^{1}$H-$^{15}$N HSQC spectra of individual isolated SGTA domains compared to the $^{1}$H-$^{15}$N HSQC of full length SGTA. The $^{1}$H-$^{15}$N HSQC of full-length SGTA displayed considerable peak broadening (Fig. 3.8A), thus sample deuteration was carried out which subsequently led to a marked improvement in spectral quality (Fig. 3.8B). When the $^{1}$H-$^{15}$N HSQC spectra of the isolated N-terminal and TPR domains were compared to that of the full-length protein (Fig. 3.9), it became evident that the fold of each of these isolated domains was conserved in the context of full-length SGTA. Similarly, for the C-terminal domain, sample deuteration was necessary to circumvent the peak broadening effect (Fig. 3.10), possibly arising due to the presence of a Q-rich region, which is likely to have a tendency to aggregate in solution. Furthermore, a comparison of the $^{1}$H-$^{15}$N HSQC of the C-terminal domain to that of full-length SGTA showed that the C-terminal domain is independent of both N-terminal and TPR domains (Fig. 3.11). Also, when the extent of amide signal perturbation of residues from each isolated domain was compared to

equivalent signals present in the $^{1}H$-$^{15}N$ HSQC of full-length SGTA (Fig. 3.12), it was clear that except in boundary regions between domains, the local environment experienced by amides in each of the isolated domains was the same when compared to their presence in full-length SGTA. These finding suggests that the N-terminal, TPR and C-terminal domains of SGTA are structurally independent modules. In addition, this analysis has shown that inter-domain contacts do not exist in full-length SGTA assemblies.

In order to carry out the aforementioned analysis, and subsequent ones described in this chapter, assignments of backbone amide resonances of N-terminal and TPR domains were obtained from the biological magnetic resonance data bank (BMRB) [BMRB accession numbers 19779 (Nter) and 5709 (TPR)]. In addition, standard triple resonance experiments were recorded to obtain the assignments of the SGTA C-terminal domain. Backbone assignments of the missing regions were carried out by Dr Santiago Martinez-Lumbreras (Appendix C). However, due to a complete absence of backbone amide signals 50% of residues within the C-terminal domain could not be assigned. In total, 84% of backbone assignments could be obtained for full-length SGTA.

Using available chemical shift data, the propensity of secondary structure of each domain (CSI) was obtained. As expected, these results show that the N-terminal domain is formed of four α-helices and the TPR domain is formed of seven α-helices, with the two domains separated by an unfolded linker (Fig 3.13). Almost all observable amino acid residues of the C-terminal domain present random coil chemical shift values (Fig. 3.13). This leads to the possibility that regions within the C-terminal domain that lack NMR signals (approximately 50% of C-terminal residues) could have alpha-helical propensity based on data obtained by CD spectroscopy (Fig. 3.6C).

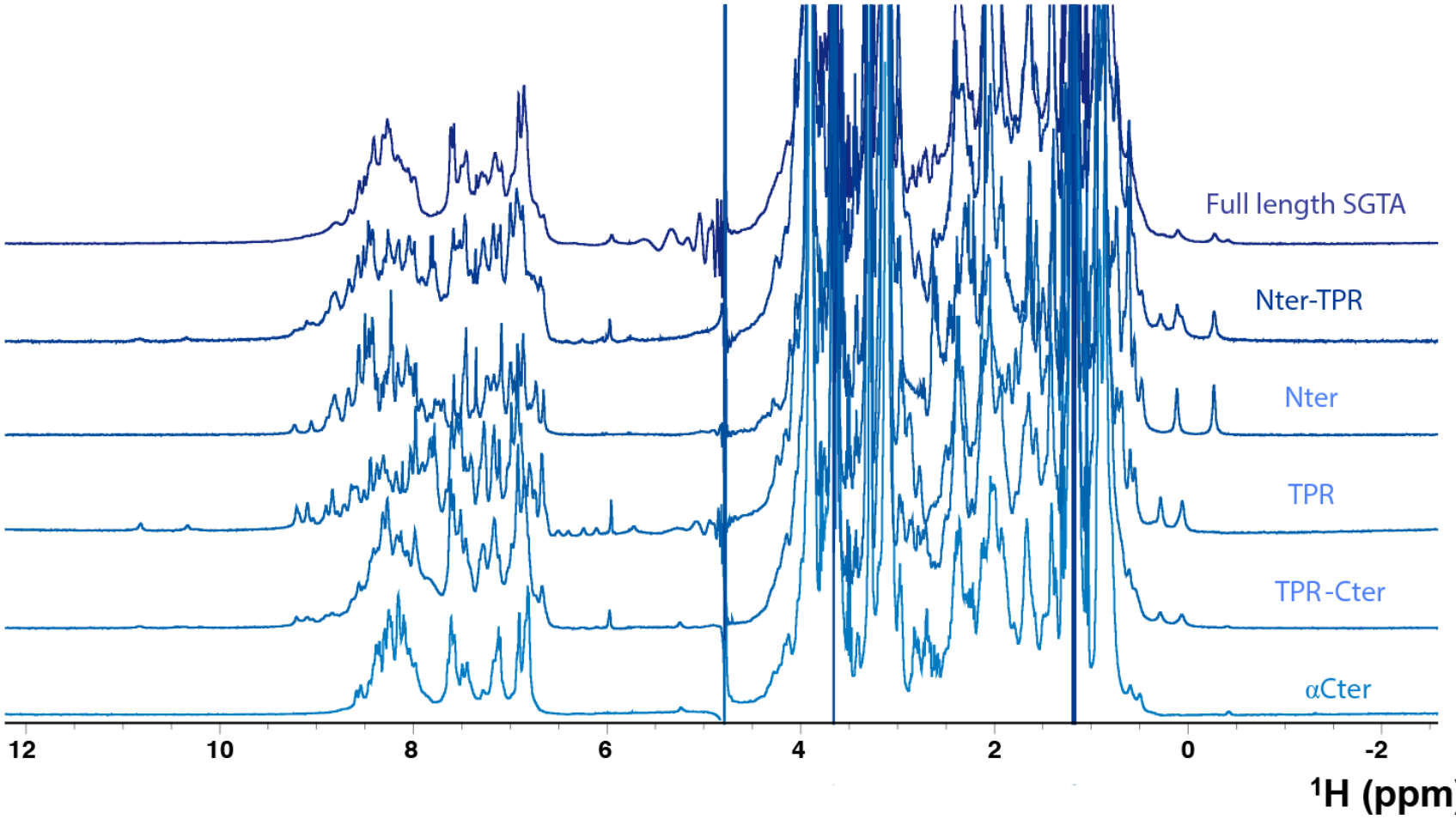



**Figure 3.7- $^{1}$H NMR spectra of different SGTA constructs.** Samples were prepared at concentrations between 500-700 µM in NMR buffer [10 mM potassium phosphate pH 6.0; 100 mM NaCl; 250 µM TCEP]. Spectra was acquired at 25°C on a 500 MHz spectrometer. The αCter construct consists of the TPR capping helix (helix-7) along with the SGTA C-terminal domain.

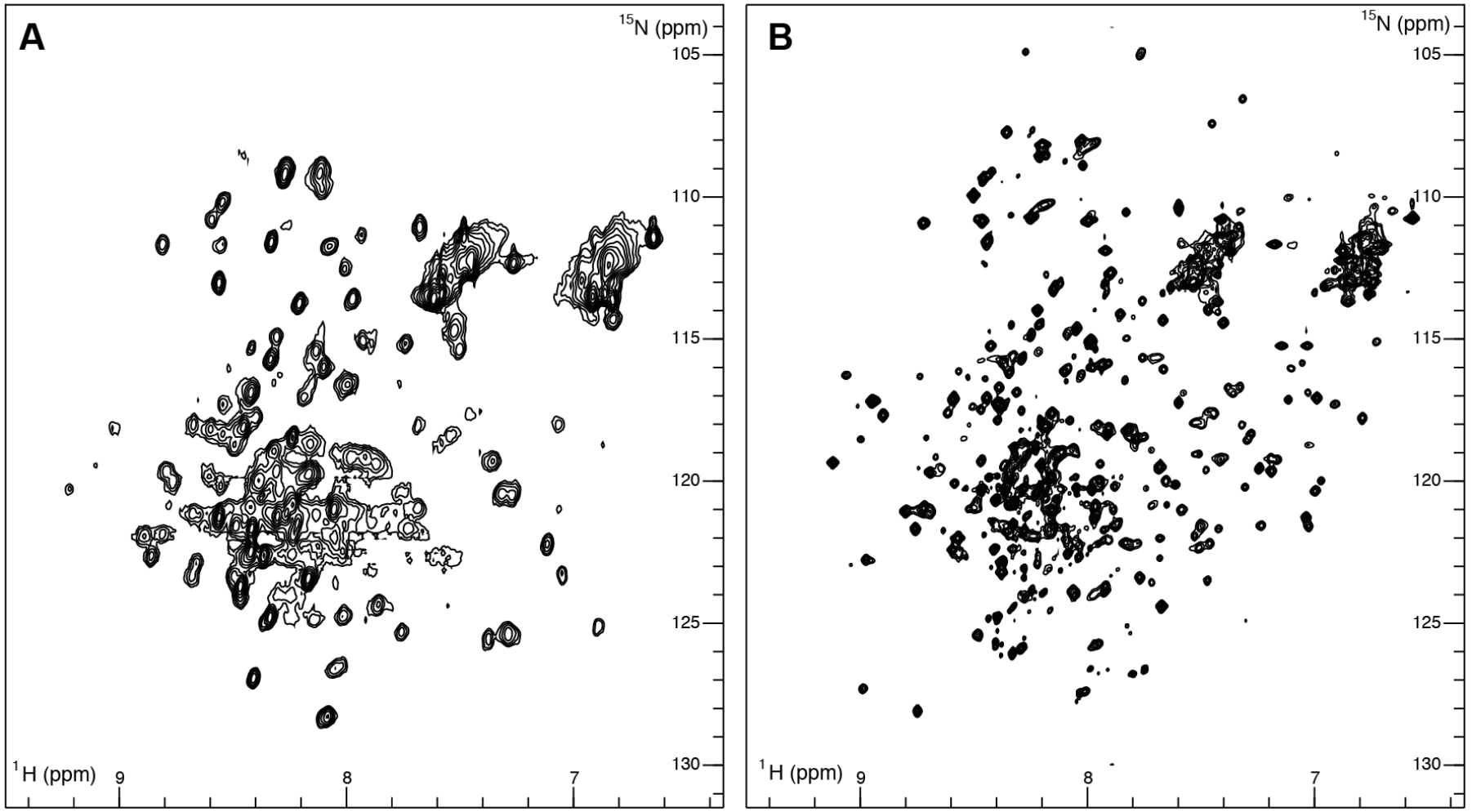



**Figure 3.8- $^{1}$H-$^{15}$N HSQC spectra of full-length SGTA.** (A) $^{1}$H-$^{15}$N HSQC of full-length SGTA recorded on a 700 MHz spectrometer at 25°C. (B) $^{1}$H-$^{15}$N HSQC of deuterated full-length SGTA recorded on a 950 MHz spectrometer at 25°C. Samples were prepared at concentrations between 350-400 μM in NMR buffer.

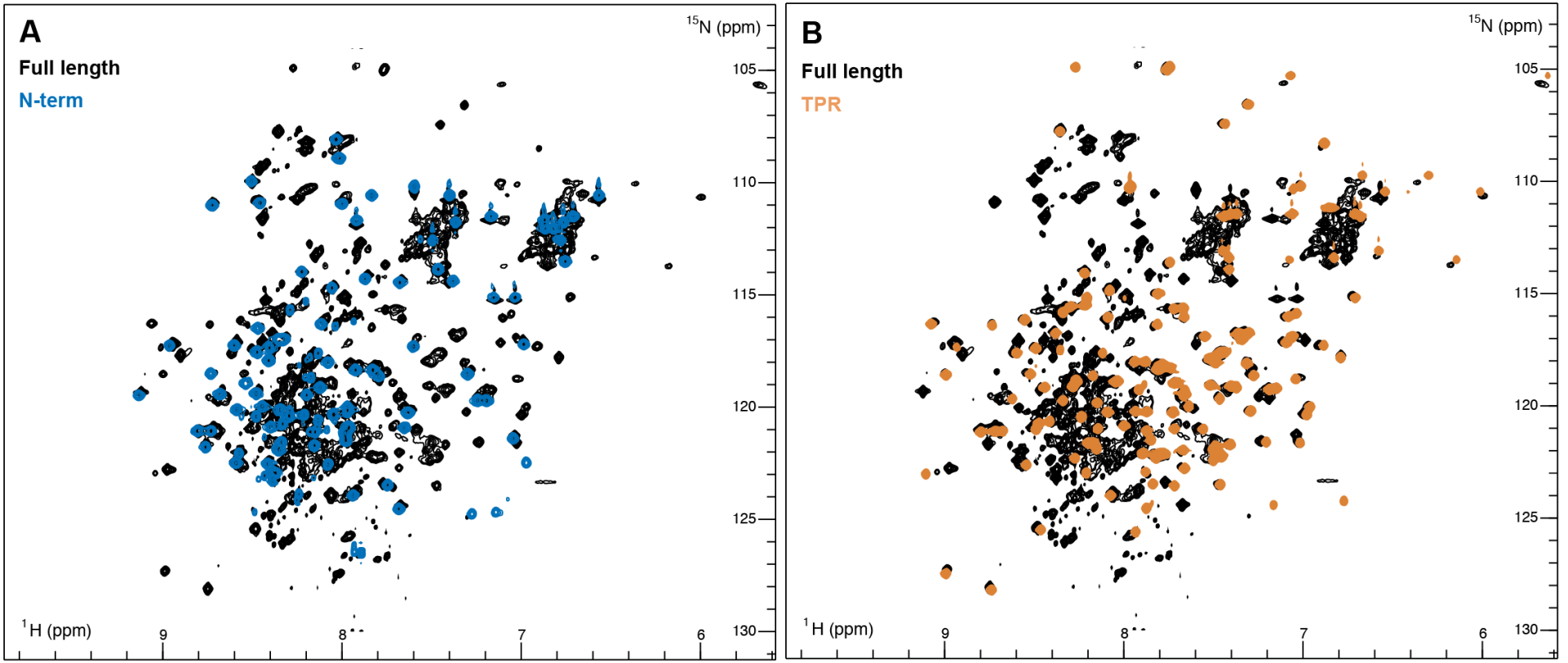



**Figure 3.9- The N-terminal and TPR domains of human SGTA are structurally independent modules.** (A) $^{1}$H-$^{15}$N HSQC spectra of full-length SGTA (acquired at 950 MHz) overlaid with $^{1}$H-$^{15}$N HSQC of the N-terminal domain construct (acquired at 500 MHz). (B) $^{1}$H-$^{15}$N HSQC spectra of full-length SGTA (acquired at 950 MHz) overlaid with $^{1}$H-$^{15}$N HSQC of the TPR domain construct (acquired at 700 MHz). In both cases, backbone amide signals of excised domains almost completely overlap with equivalent signals in the $^{1}$H-$^{15}$N HSQC spectra of full-length SGTA, indicative of their presence as structurally independent modules within the context of full-length SGTA.

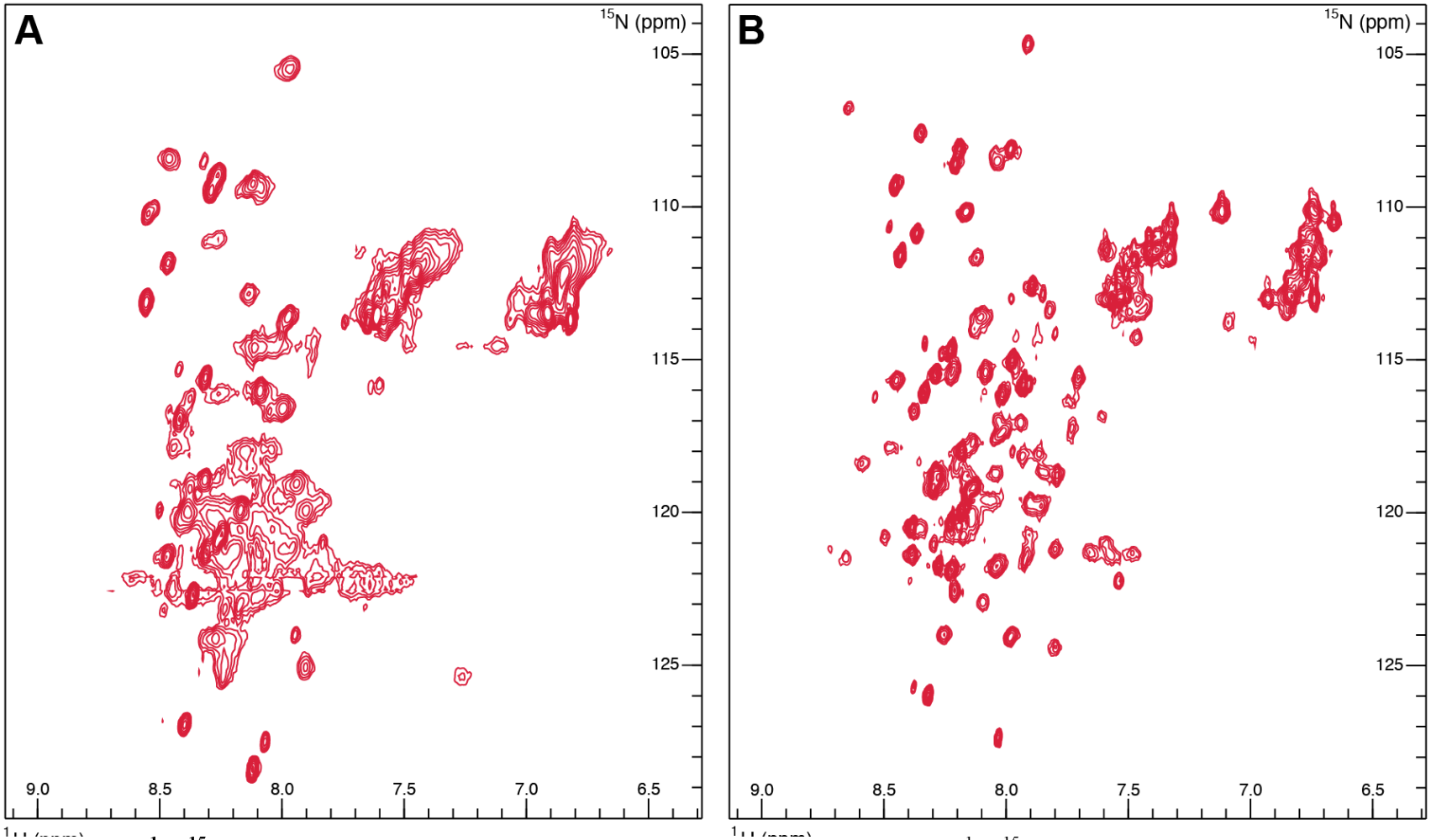



**Figure 3.10- $^{1}$H-$^{15}$N HSQC spectra of the excised C-terminal domain of SGTA.** (A) $^{1}$H-$^{15}$N HSQC of the C-terminal domain of SGTA (B) $^{1}$H-$^{15}$N HSQC of the C-terminal domain of SGTA upon deuteration. In both cases spectra were acquired at 25°C on a 700MHz spectrometer.

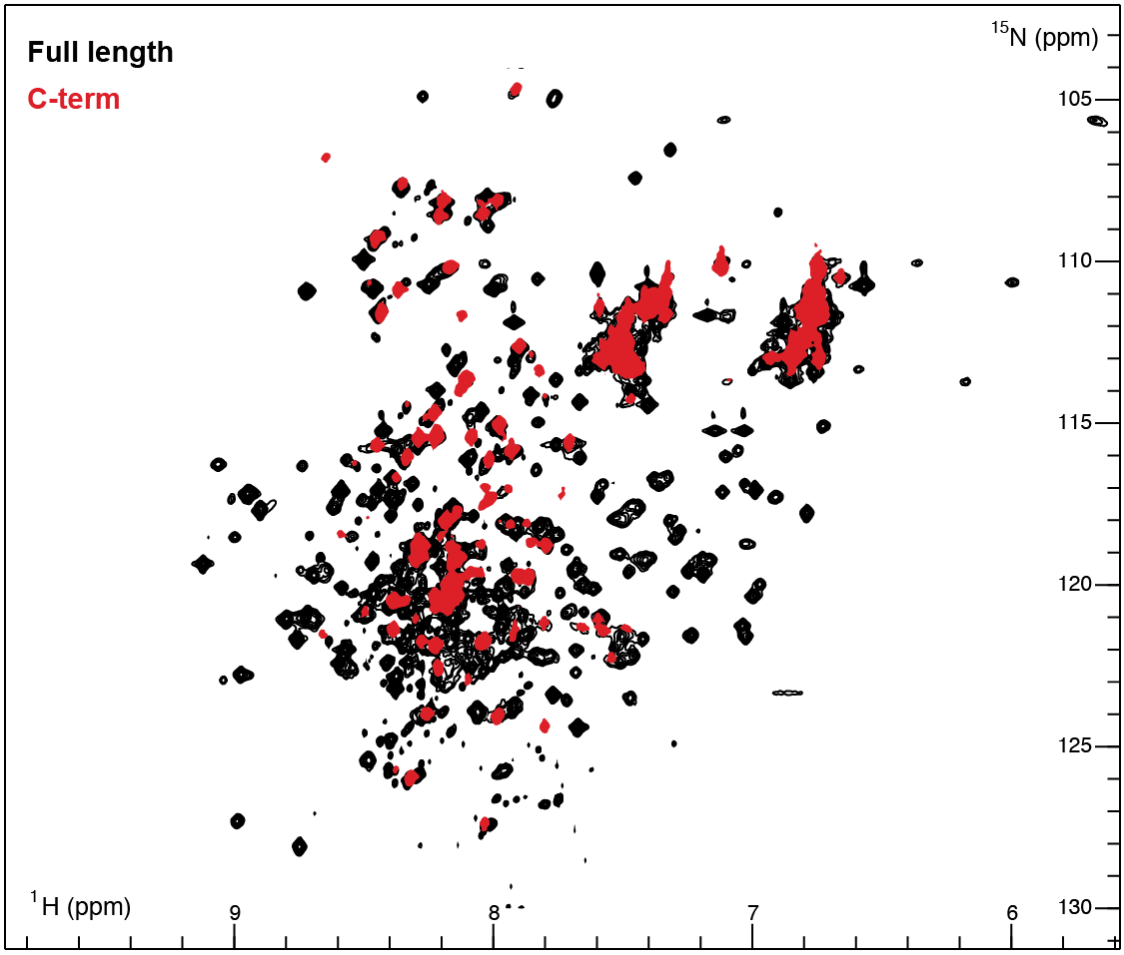



**Figure 3.11- The C-terminal domain of SGTA is independent of both its N-terminal and TPR domains.** $^{1}H$-$^{15}N$ HSQC spectra of deuterated full-length SGTA (acquired at 950 MHz) overlaid with $^{1}H$-$^{15}N$ HSQC of the deuterated C-terminal domain construct (acquired at 700 MHz). Observable amide signals of the excised C-terminal domain almost completely overlap with equivalent signals in the $^{1}H$-$^{15}N$ HSQC spectra of full-length SGTA, indicative of its presence as a structurally independent module within the context of full-length SGTA. In both cases, spectra were recorded at 25°C.

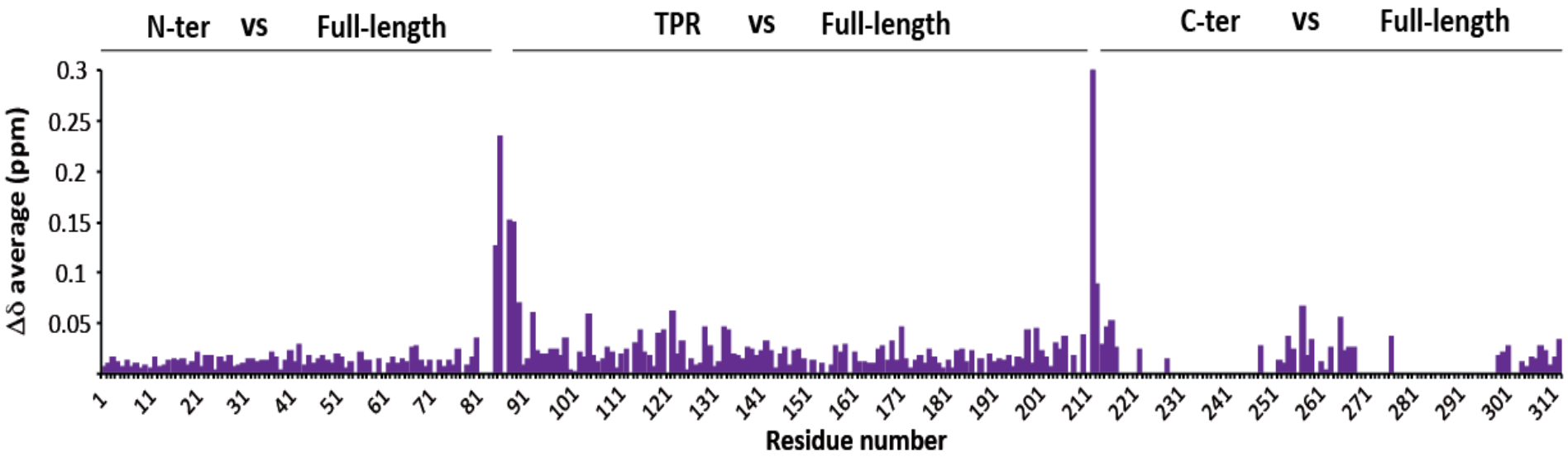



**Figure 3.12- Chemical shift perturbation (CSP) analysis of SGTA domains.** CSP $\Delta\delta^{av}$ values calculated using backbone amide signals belonging to excised N-terminal, TPR and C-terminal domains with respect to corresponding amide signals of the full-length construct.

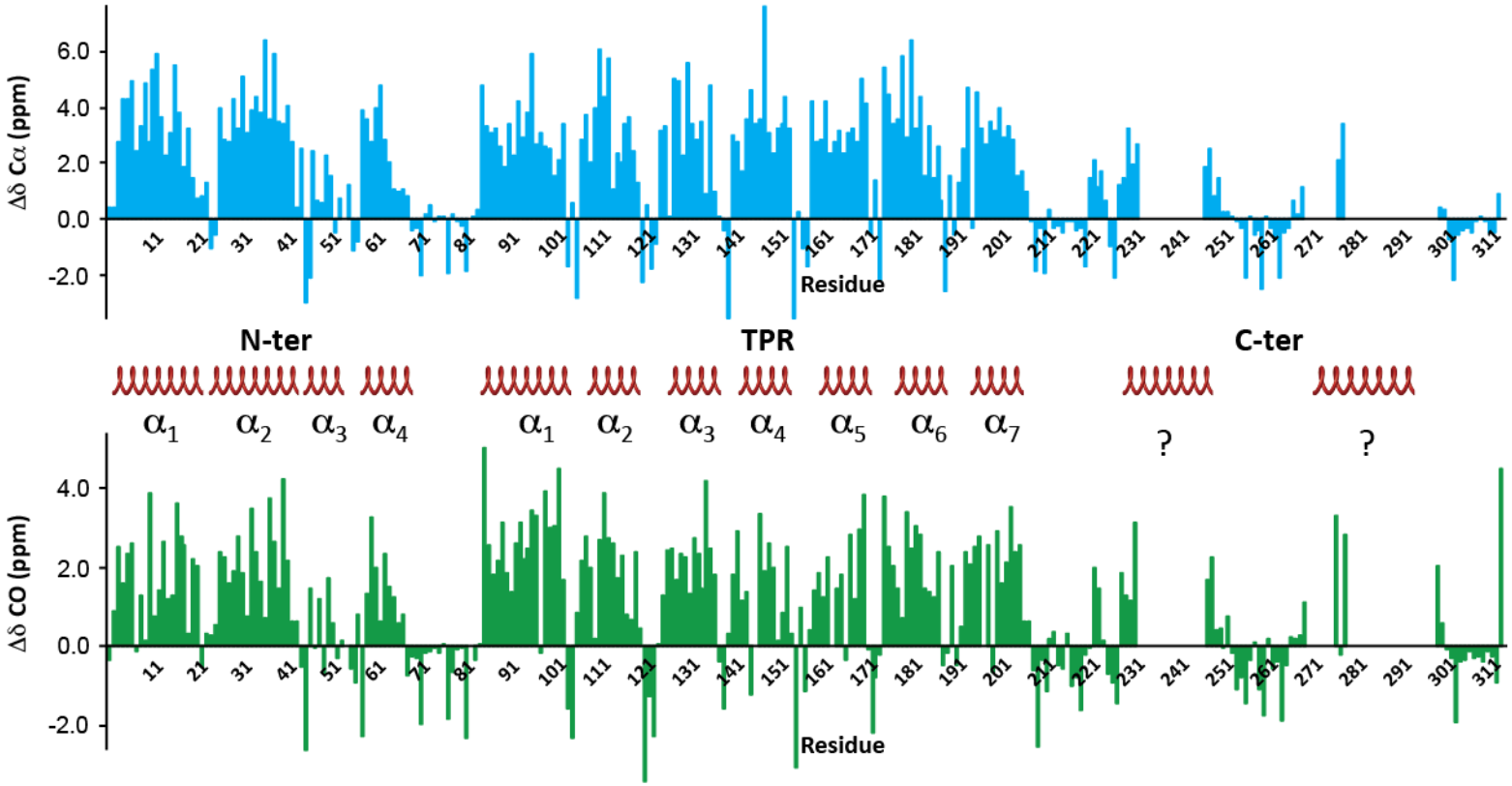



**Figure 3.13- Chemical shift index (CSI) analysis of full-length SGTA**. Alpha carbon and carbonyl chemical shift deviations from a random coil value are shown. The two regions within the C-terminal domain that lack NMR signals, are speculated to be alpha helical based on data obtained by CD spectroscopy.

### 3.5. Solution dynamics of SGTA domains

To investigate the dynamic properties of SGTA in solution, a series of $^{15}N$ NMR relaxation experiments were carried out on full-length SGTA and on isolated domains thereof. Firstly, $^{1}H$-$^{15}N$ heteronuclear NOE experiments were carried out on four constructs (full-length, Nter-TPR, Nter and TPR) to address the motion of NH vectors thereby identifying structured domains and unstructured regions within these constructs. Naturally, the N-terminal and TPR domains appeared as folded domains in all constructs in which they are present, and are separated by flexible linkers in Nter-TPR and full-length constructs (Fig 3.14). The $^{1}H$-$^{15}N$ NOE values corresponding to observable C-terminal residues are typical of flexible proteins, in particular those corresponding to extreme C-terminal residues (Fig. 3.14). In addition, NMR relaxation parameters of N-terminal and TPR domains have been determined in the context of different SGTA constructs (Figs. 3.15 – 3.18). $T_1$ and $T_2$ values and correlation times ($\tau_c$) have been calculated for the N-terminal domain in the context of Nter, Nter-TPR and full-length constructs. Similarly, $T_1$ and $T_2$ values and correlation times ($\tau_c$) were computed for the TPR domain in the context of TPR, Nter-TPR and full-length constructs. These experiments reveal that the N-terminal domain has similar correlation times (~10-11 ns) when present in the context of all three constructs, namely Nter, Nter-TPR and full-length (Fig. 3.19A). On the other hand, TPR domain correlation times appeared to be similar in the context of Nter-TPR and TPR constructs (~9-10 ns). Intriguingly, the correlation time of the TPR domain increases to ~12 ns when it is present in the context of full-length SGTA (Figs. 3.15 and 3.19B). Therefore, it is evident that dynamic parameters of the TPR domain as determined by NMR, are highly affected in full-length SGTA, when compared to the other TPR domain containing constructs (Fig. 3.19). This suggests that within full-length SGTA, the C-terminal domain imposes a viscous drag on the TPR domain, whilst not having an equivalent effect on its N-terminal domain.

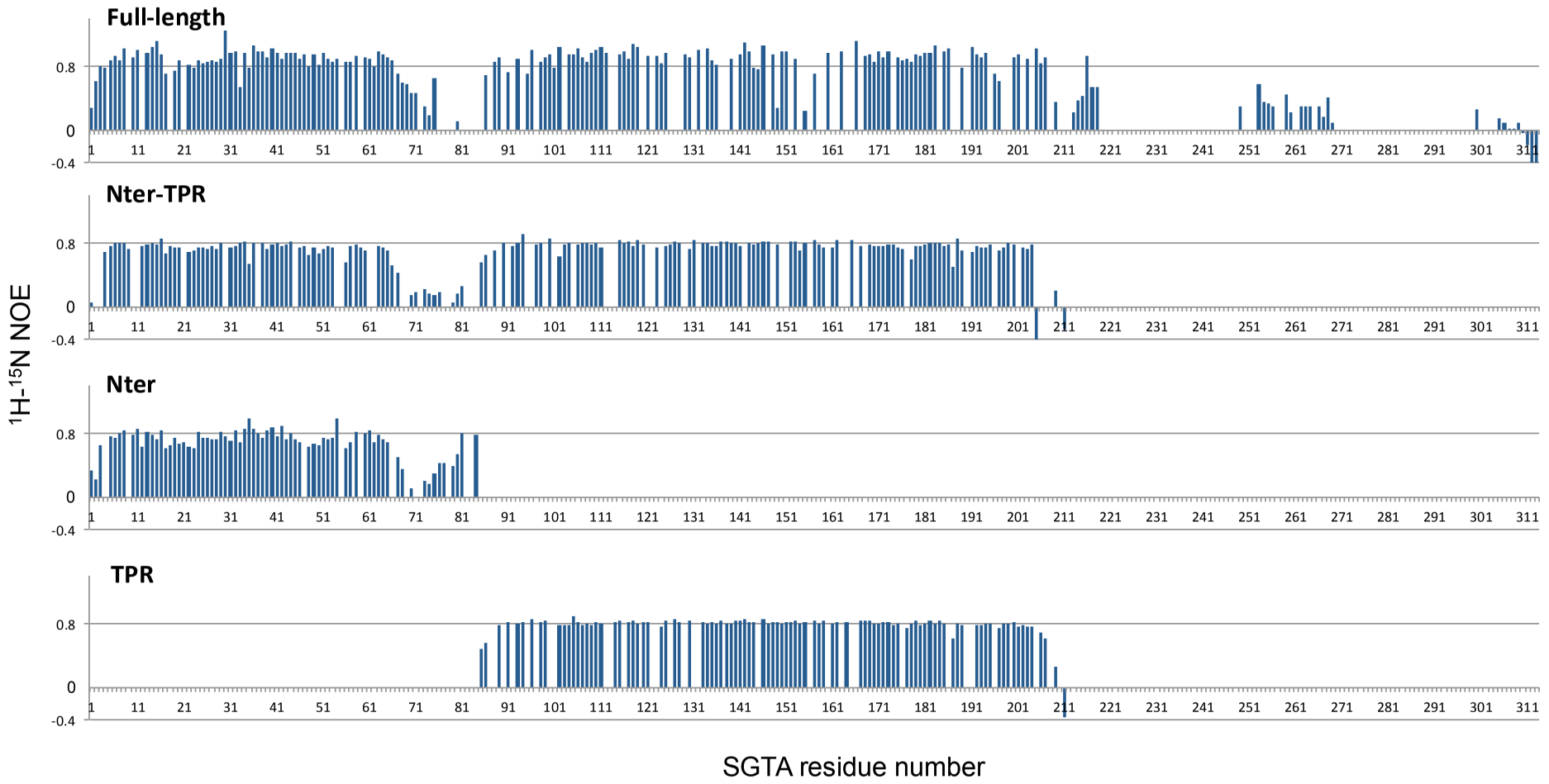



**Figure 3.14- $^{1}$H-$^{15}$N heteronuclear NOE enhancement values of various SGTA constructs.** $^{1}$H-$^{15}$N heteronuclear NOE values of full-length (acquired at 950 MHz), Nter-TPR (acquired at 700 MHz), Nter (acquired at 500 MHz), and TPR (acquired at 700 MHz) constructs are shown, with per residue values obtained from the ratio of cross-peak volumes in two experiments that were carried out with a four second saturation transfer or control period. All experiments were recorded at 25°C.

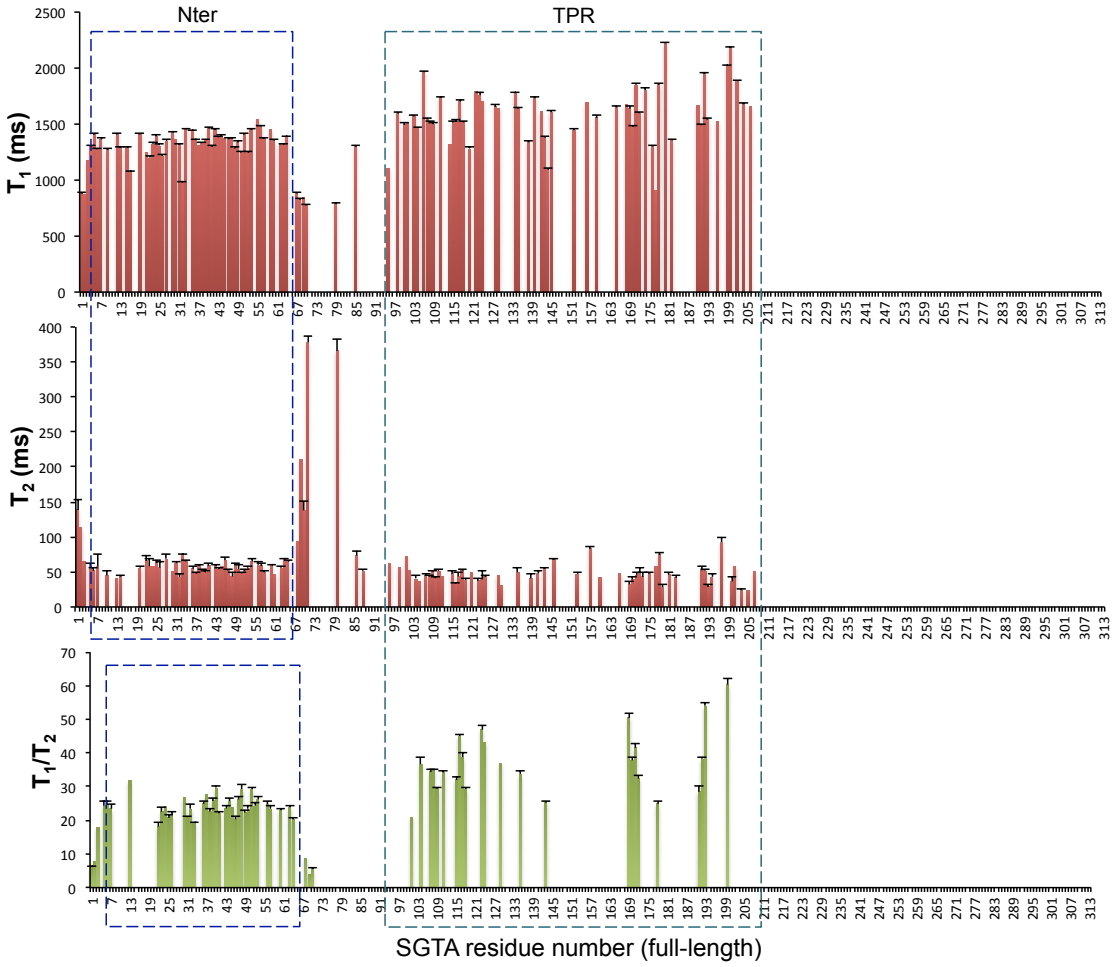



**Figure 3.15- $T_1$, $T_2$ and $T_1/T_2$ values of the full-length construct (acquired at 950 MHz) plotted as a function of SGTA residue number.** $T_1$ and $T_2$ relaxation times were calculated from single exponential decay fitting of peak intensities of each amide signal. Rotational correlation times of the N-terminal domain within the context of full-length SGTA were calculated using $T_1$ and $T_2$ values obtained from N-terminal residues 5 to 65 (*indicated by dotted lines*). Based on an average $T_1$ = 1341.99 ms, and an average $T_2$ = 56.27 ms, an Nter $\tau_c$ = 9.6 ± 2.3 ns was obtained. When based on an average Nter $T_1/T_2$ value of 24.10, an Nter $\tau_c$ = 9.7 ± 1.7 ns was obtained. In addition, rotational correlation times of the TPR domain within the context of full-length SGTA were calculated using TPR domain residues 92 to 205 (*indicated by dotted lines*). Based on an average $T_1$ = 1603.39 ms, and an average $T_2$ = 47.12 ms, a TPR $\tau_c$ = 11.6 ± 4.5 ns was obtained. When based on an average TPR $T_1/T_2$ value of 37.05, TPR $\tau_c$ = 12.1 ± 1.6 ns.

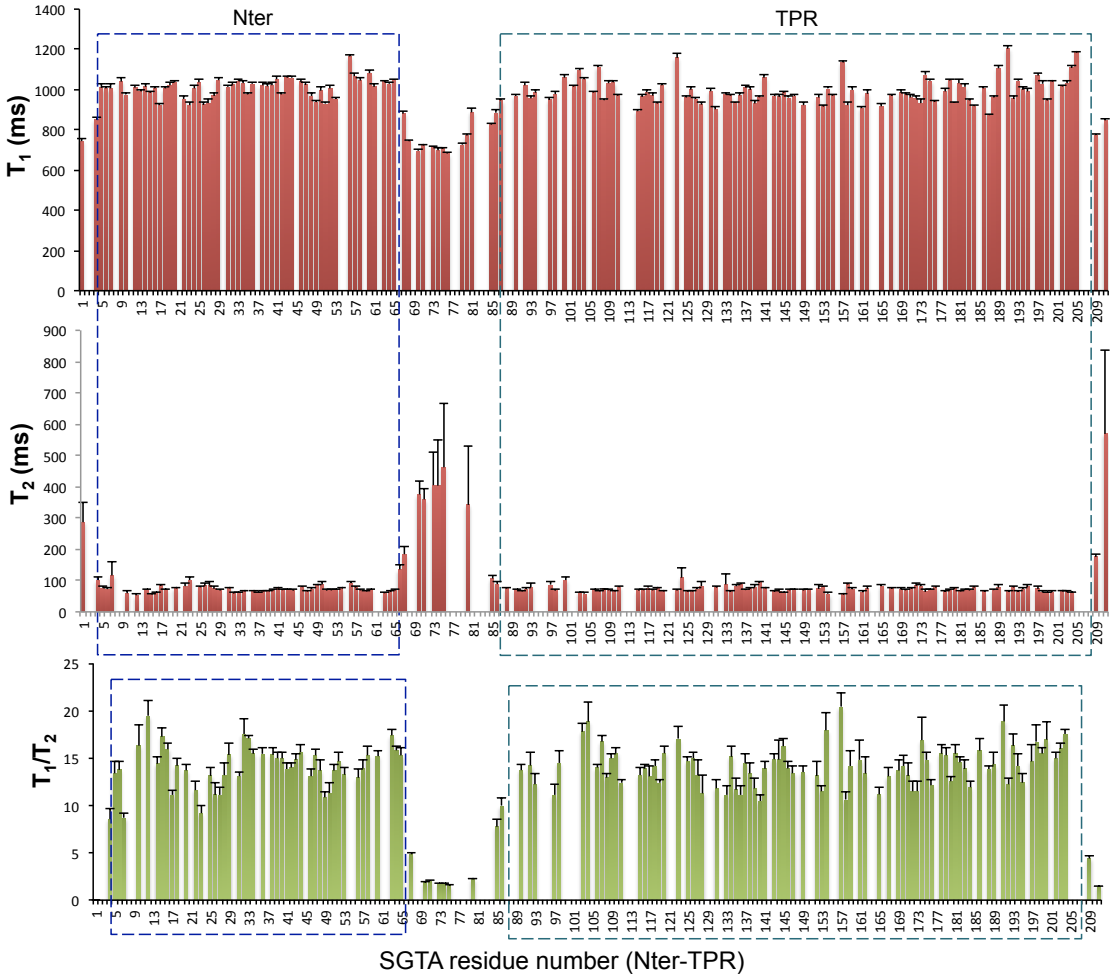



**Figure 3.16- $T_1$, $T_2$ and $T_1/T_2$ values of the Nter-TPR construct construct (acquired at 700 MHz) plotted as a function of SGTA residue number.** $T_1$ and $T_2$ relaxation times were calculated from single exponential decay fitting of peak intensities of each amide signal. Rotational correlation times of the N-terminal domain within the context of the Nter-TPR construct were calculated using $T_1$ and $T_2$ values obtained from N-terminal residues 5 to 65 (*indicated by dotted lines*). Based on an average $T_1$ = 1008.62 ms, and an average $T_2$ = 72.70 ms, an Nter $\tau_c$ = 9.8 ± 1.7 ns was obtained. When based on an average Nter $T_1/T_2$ value of 14.14, an Nter $\tau_c$ = 9.9 ± 1.6 ns was obtained. In addition, rotational correlation times of the TPR domain present within the context of the Nter-TPR construct were calculated using TPR domain residues 87 to 205 (*indicated by dotted lines*). Based on an average $T_1$ = 991.72 ms, and an average $T_2$ = 71.94 ms, a TPR $\tau_c$ = 9.8 ± 1.6 ns was obtained. When based on an average TPR $T_1/T_2$ value of 14.12, TPR $\tau_c$ = 9.9 ± 1.9 ns.

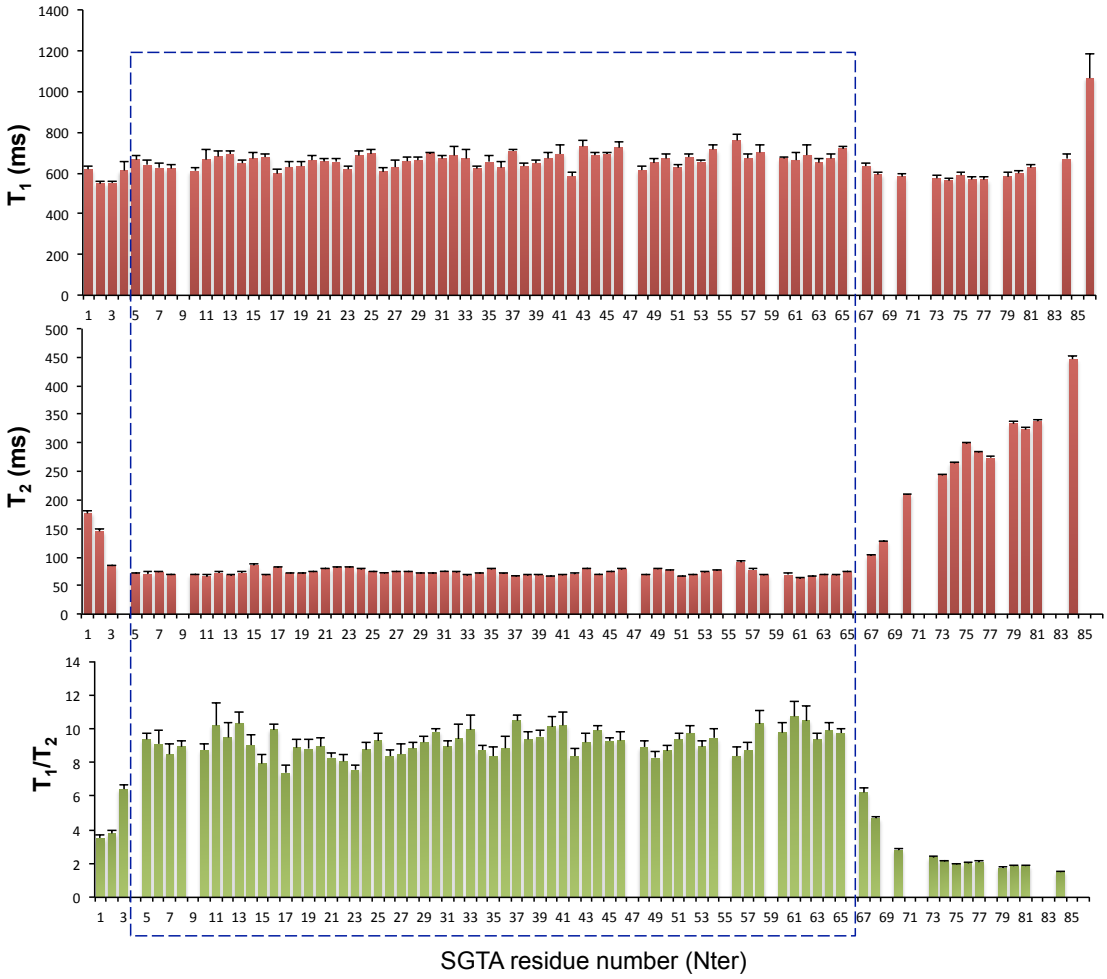



**Figure 3.17- $T_1$, $T_2$ and $T_1/T_2$ values of the isolated Nter domain construct (acquired at 500 MHz) plotted as a function of SGTA residue number.** Rotational correlation times of the N-terminal domain construct were calculated using $T_1$ and $T_2$ values obtained from N-terminal residues 5 to 65 (*indicated by dotted lines*). Based on an average $T_1$ = 662.03 ms, and an average $T_2$ = 72.54 ms, an Nter $\tau_c$ = 10.9 ± 1.1 ns was obtained. When based on an average Nter $T_1/T_2$ value of 9.17, an Nter $\tau_c$ = 10.9 ± 0.9 ns was obtained.

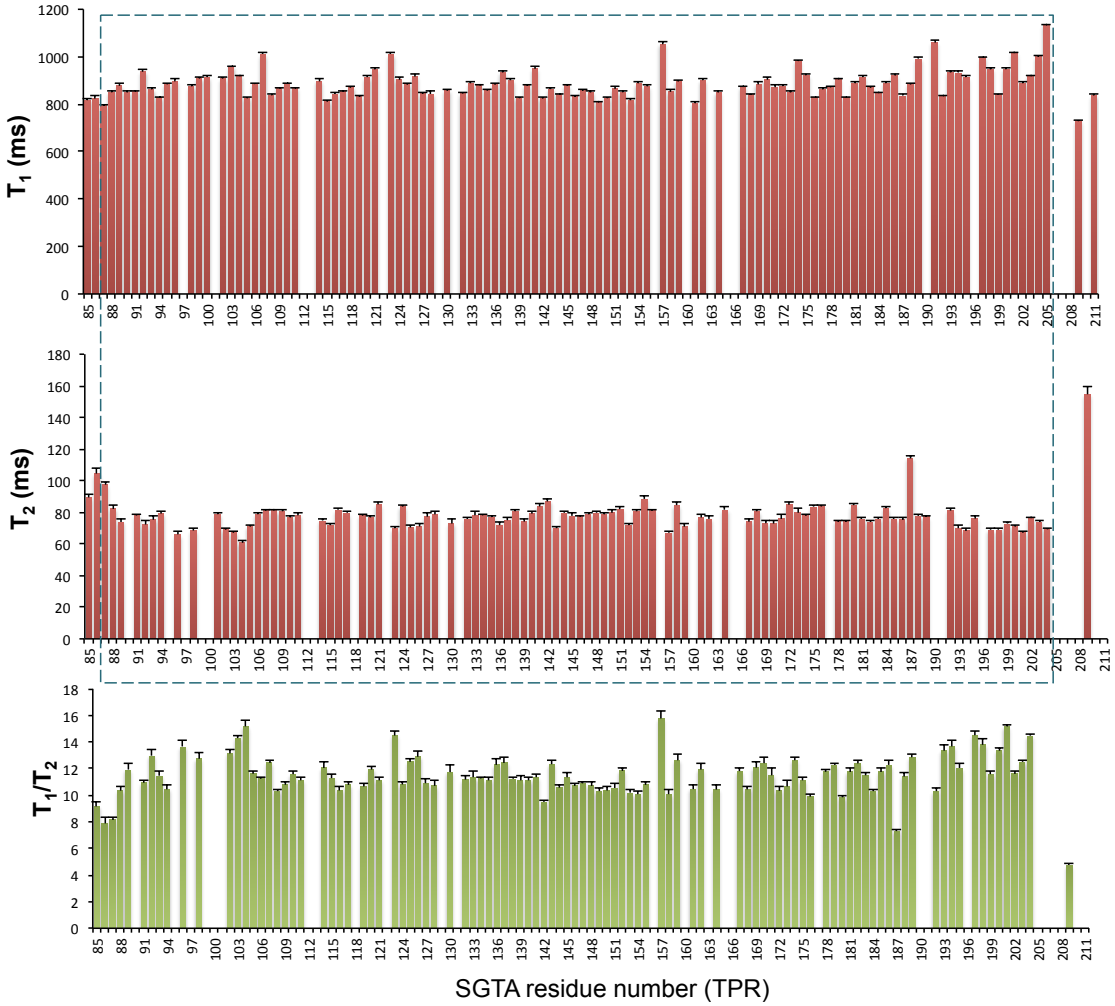



**Figure 3.18- $T_1$, $T_2$ and $T_1/T_2$ values of the excised TPR domain construct (acquired at 700 MHz) plotted as a function of SGTA residue number.** Rotational correlation times of the TPR domain construct were calculated using TPR domain residues 87 to 205 (*indicated by dotted lines*). Based on an average $T_1$ = 888.06 ms, and an average $T_2$ = 76.88 ms, a TPR $\tau_c$ = 8.9 ± 1.4 ns was obtained. When based on an average TPR $T_1/T_2$ value of 11.61, TPR $\tau_c$ = 8.9 ± 0.9 ns.

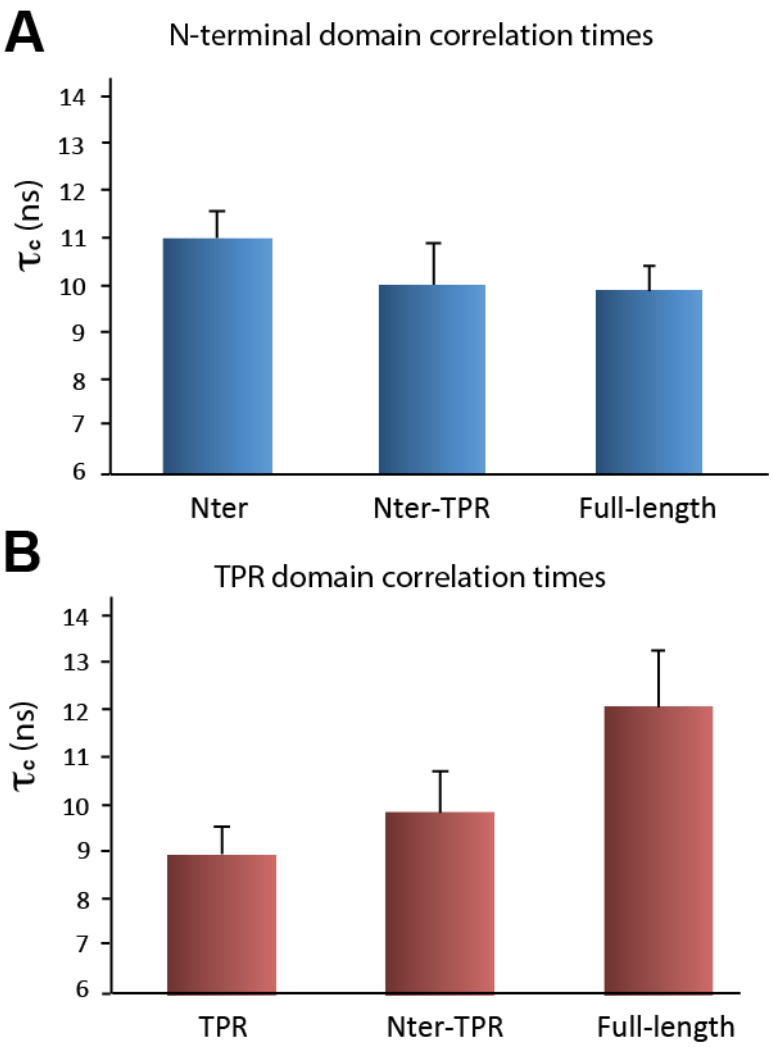



**Figure 3.19- Dynamic parameters of the TPR domain are highly affected in full-length SGTA.** (A) Calculated correlations times ($\tau_c$) for the N-terminal domain in the context of Nter, Nter-TPR and full-length constructs. (B) Calculated correlations times ($\tau_c$) for the TPR domain in the context of TPR, Nter-TPR and full-length constructs.

## 3.6 Native mass spectrometry of SGTA constructs

To investigate possible scenarios that could explain the aforementioned viscous drag experienced by the TPR domain in full-length SGTA, but not in constructs devoid of the C-terminal domain, as observed by NMR relaxation experiments, non dissociative nano-electrospray ionization mass spectrometry (Native MS) has been carried out. As Native MS can be used to monitor intact proteins and provide information about their stoichiometry, full-length and C-terminal domain samples were subjected to this technique.

Native MS experiments were carried out by Dr Dijana Matak-Vinkovic at the Department of Chemistry, University of Cambridge. As expected, full-length SGTA presented dimeric forms as the major species detected (Fig. 3.20A). Interestingly, the isolated C-terminal domain construct also exhibited a dimer as the major species (Fig. 3.20B). This provides a possible explanation for the highly affected dynamic parameters of the TPR domain in full-length SGTA. These results have shown, for the first time, that SGTA is able to dimerise via its C-terminal domain. These findings are also in agreement with analytical SEC experiments (Fig. 3.5) which have suggested that the presence of the C-terminal domain within a construct results in the formation of higher molecular weight species.

Upon further analysis of these datasets, it was apparent that different forms of the C-terminal dimer, varying in molecular weight by multiples of 359 ± 23 Da, were present. Mass spectrometry under denaturing conditions has discounted the possibility of these adducts arising as a result of covalent modifications of the protein. Therefore, these larger species most likely correspond to a non-covalently bound molecule present at a stoichiometry of between one and four times that of the SGTA C-terminal domain

construct. Curiously, these larger species are only observed in the dimeric form of the SGTA C-terminal domain and also appear in the spectrum of full-length SGTA (370 ± 15 Da). None of the buffer components correlate with this mass, hence this compound most likely corresponds to a lipid, or some other hydrophobic molecule present in the *E. coli* expression system.

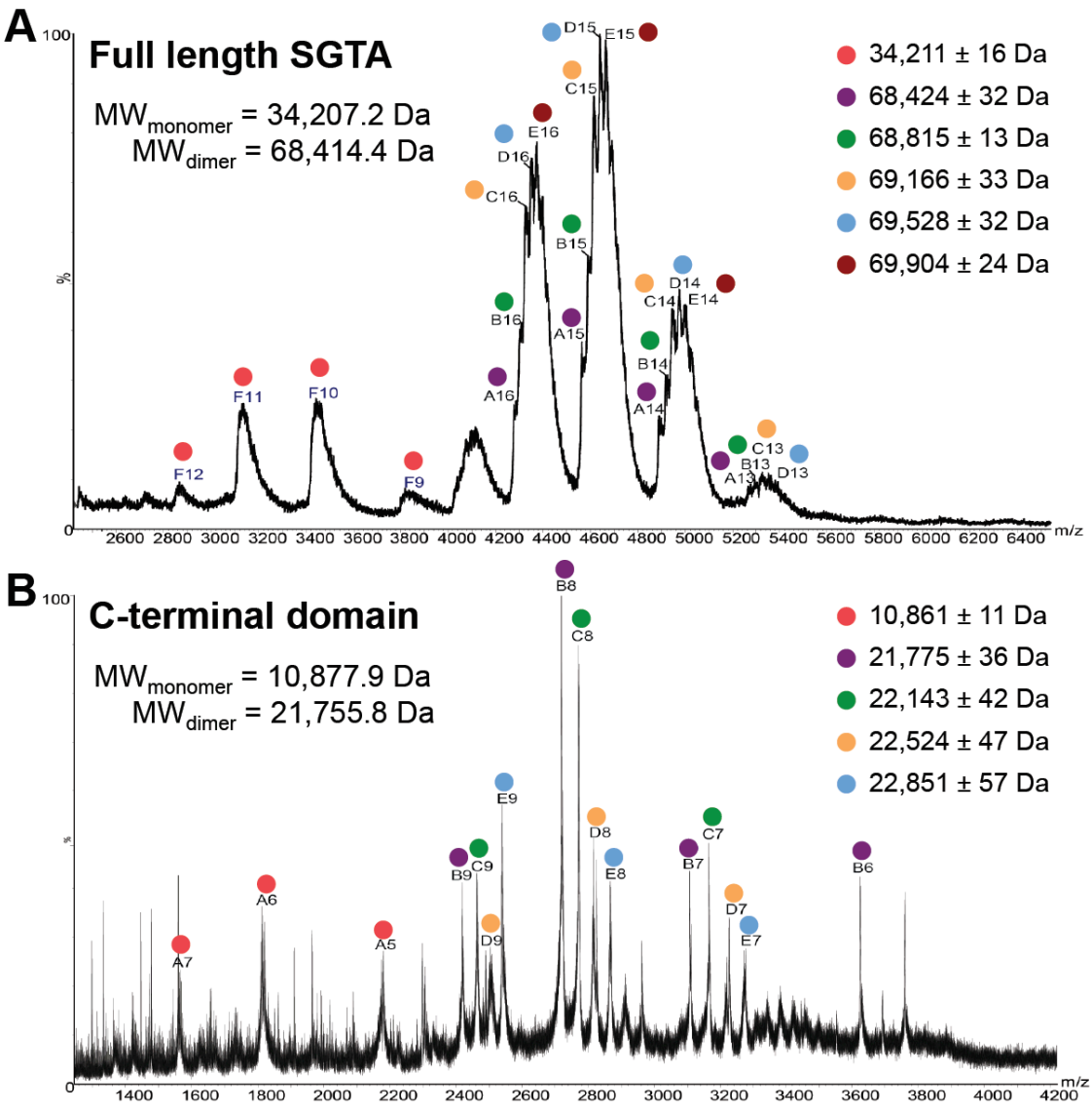



**Figure 3.20- Non dissociative nano-electrospray ionization mass spectra (Native MS) of SGTA samples.** (A) Full-length SGTA (B) C-terminal domain of SGTA. Molecular weights of each species detected in these samples are indicated. Samples were exchanged into 0.20–0.75 M ammonium acetate (pH 7.0) and were present at a final concentration of 5–10 µM prior to analysis. 2.5 µL of protein solution was electrosprayed from a borosilicate emitter for sampling. Native MS experiments were carried out at a capillary voltage of 1.8–2.5 kV, cone voltage of 60– 120 V, collision voltage of 10−30 V, with backing pressure 3–4 mbar and source temperature of 20 °C. Spectra were calibrated externally using cesium iodide. Data acquisition and processing were performed using MassLynx 4.1. (Waters).

## 3.7. Summary

These biophysical experiments provide some novel insights pertaining to the structure and assembly of full-length SGTA *in vitro*. Firstly, CD spectroscopy has indicated the presence of alpha helical secondary structural elements within the C-terminal substrate-binding domain of SGTA. Secondly, characterisation of SGTA constructs by solution NMR spectroscopy has revealed that its N-terminal, TPR and C-terminal domains are structurally independent modules. Also, the same experiments have confirmed the absence of inter-domain contacts in full-length SGTA. In addition, native MS experiments have shown, for the first time, the ability of SGTA to dimerise via its C-terminal region. Finally, a combination of techniques, namely analytical SEC, $^{15}$N NMR relaxation experiments, and native MS, suggest the presence of a constrained conformation of full-length SGTA *in vitro*. In this proposed conformation, C-terminal domains interact with each other closing the dimer at its distal end, thereby constraining the mobility of the central TPR domain. These new findings are likely to have implications with regard to SGTA's role in the recognition of hydrophobic substrates.

# Chapter 4

# Investigation of the interaction between SGTA and hydrophobic substrates

SGTA is known to interact with hydrophobic substrates via its C-terminal domain (Liou and Wang, 2005; Roberts *et al.*, 2015), however molecular level characterisation of this process remains scarce. The C-terminal domain of SGTA can be divided into two regions namely an NNP region (residues 213-274), and an extreme C-terminal Q-rich region (residues 275-313) (Fig. 4.1A). The NNP motifs within the C-terminal region of SGTA are conserved across eukaryotes, with human SGTA containing three such motifs, thus potentially contributing to SGTA's hydrophobic substrate binding functionality. Also, the presence of glutamine residues is also a conserved feature of the C-terminal domain, with possible functional implications. In this chapter I aim to understand the role of the SGTA C-terminal domain with respect to its association with hydrophobic substrates.

## 4.1. Expression and purification of SGTA and TA protein constructs for interaction studies

Hence to investigate the interaction between SGTA and its hydrophobic substrates, firstly a construct encoding the TPR domain of SGTA followed by its C-terminal NNP region (residues 84-274; hereafter referred to as TPR_NNP) was designed to understand the contribution of conserved NNP motifs within the C-terminal region of SGTA (Fig. 4.1A). Furthermore, this TPR_NNP construct would be suitable to understand any contributions of the TPR capping helix (helix-7) arising from its proximity to the C-terminal domain. As tail-anchored membrane proteins are validated substrates of SGTA, the full-length TA proteins cytochrome b5 (Cytb5) and synaptobrevin 2 (Syb2) were used as hydrophobic substrates in this study. Both Cytb5 and Syb2 comprise a C-terminal hydrophobic TMD helix (Fig. 4.1B).

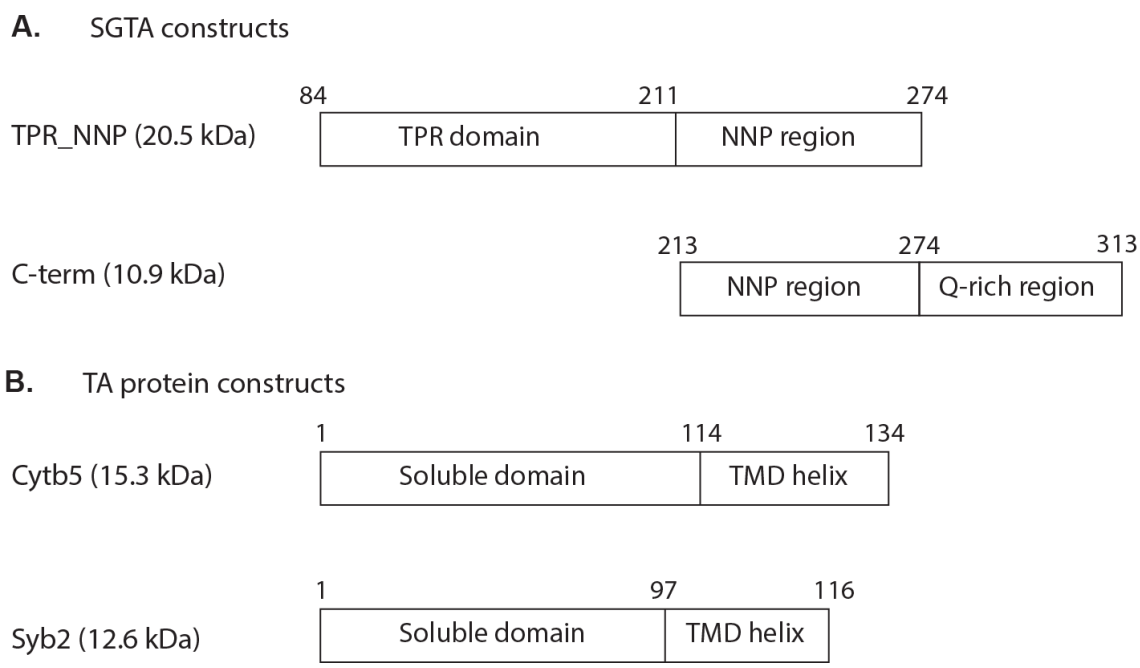



**Figure 4.1- SGTA and TA protein constructs designed to understand the shielding of hydrophobic substrates by SGTA.** (A) Schematics of SGTA constructs incorporating regions of its C-terminal substrate binding domain. The TPR_NNP construct encodes the central TPR domain of SGTA along with the conserved region of the C-terminal domain consisting of three NNP motifs. This construct lacks the C-terminal Q-rich region. The C-term construct encodes the entire C-terminal domain of SGTA that consists of an NNP region and an extreme C-terminal Q-rich region. (B) Schematics of tail-anchored (TA) protein constructs. Both human cytochrome b5 (Cytb5) and synaptobrevin 2 (Syb2) are TA proteins consisting of a functional cytosolic domain (soluble domain) that is tethered to the cell membrane by an extreme C-terminal hydrophobic TMD helix.

All protein constructs were successfully cloned, overexpressed in *E. coli*, and purified, as described in Chapter 2 (*Materials and Methods*) as fusion proteins with N-terminally placed thioredoxin A, which improved recombinant protein expression and solubility, followed by a hexahistidine tag and a rTEV cleavage site. Furthermore, all proteins were recovered in the soluble fraction, and maintained their solubility after removal of thioredoxin A by rTEV digestion (Figs. 4.2 and 4.3).

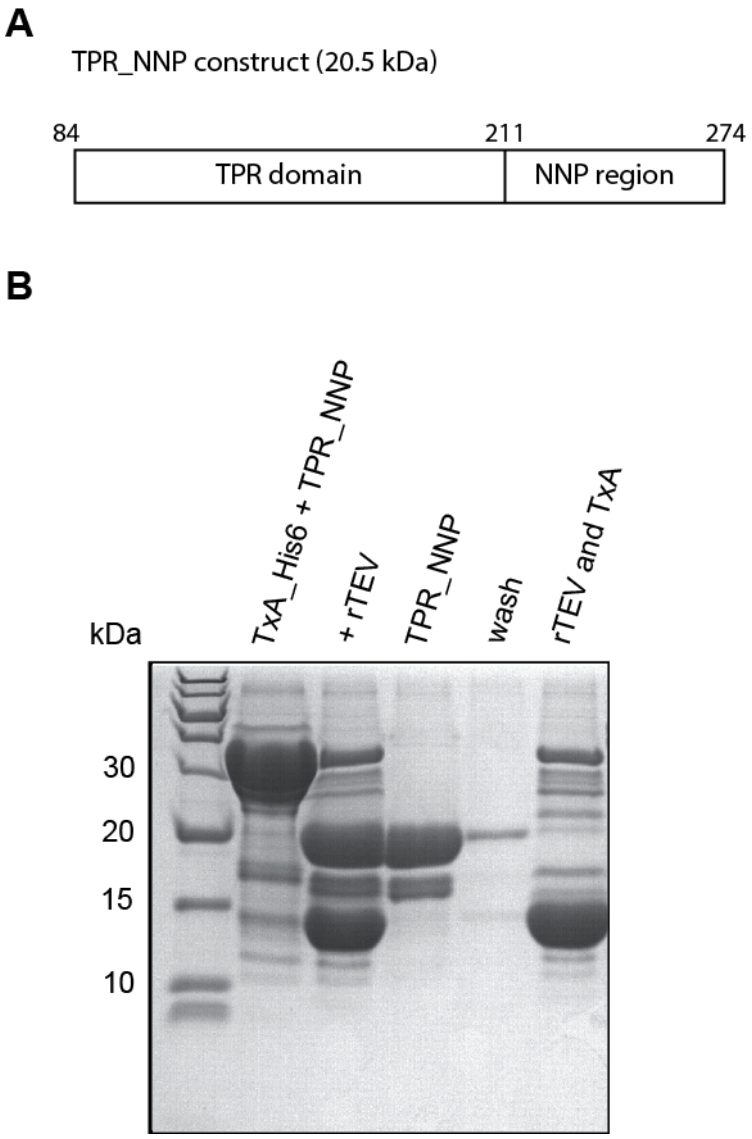



**Figure 4.2- Purification of the TPR_NNP construct.** (A) Schematic of the TPR_NNP construct. (B) SDS-PAGE showing different stages of TPR_NNP purification. TPR_NNP was overexpressed as a fusion protein with thioredoxin A (TxA), purified by nickel affinity chromatography and was recovered in the eluate of the first nickel column (SDS-PAGE: TxA_$His_6$ + TPR_NNP). This was followed by affinity tag and fusion protein removal by treatment with rTEV (SDS-PAGE fraction: +rTEV). Pure protein was obtained by passing the rTEV treated fraction through a second nickel column with pure protein obtained in the flow-through (SDS-PAGE: TPR_NNP). Separated thioredoxin A and rTEV was eluted from the second nickel column (SDS-PAGE fraction: rTEV and TxA).

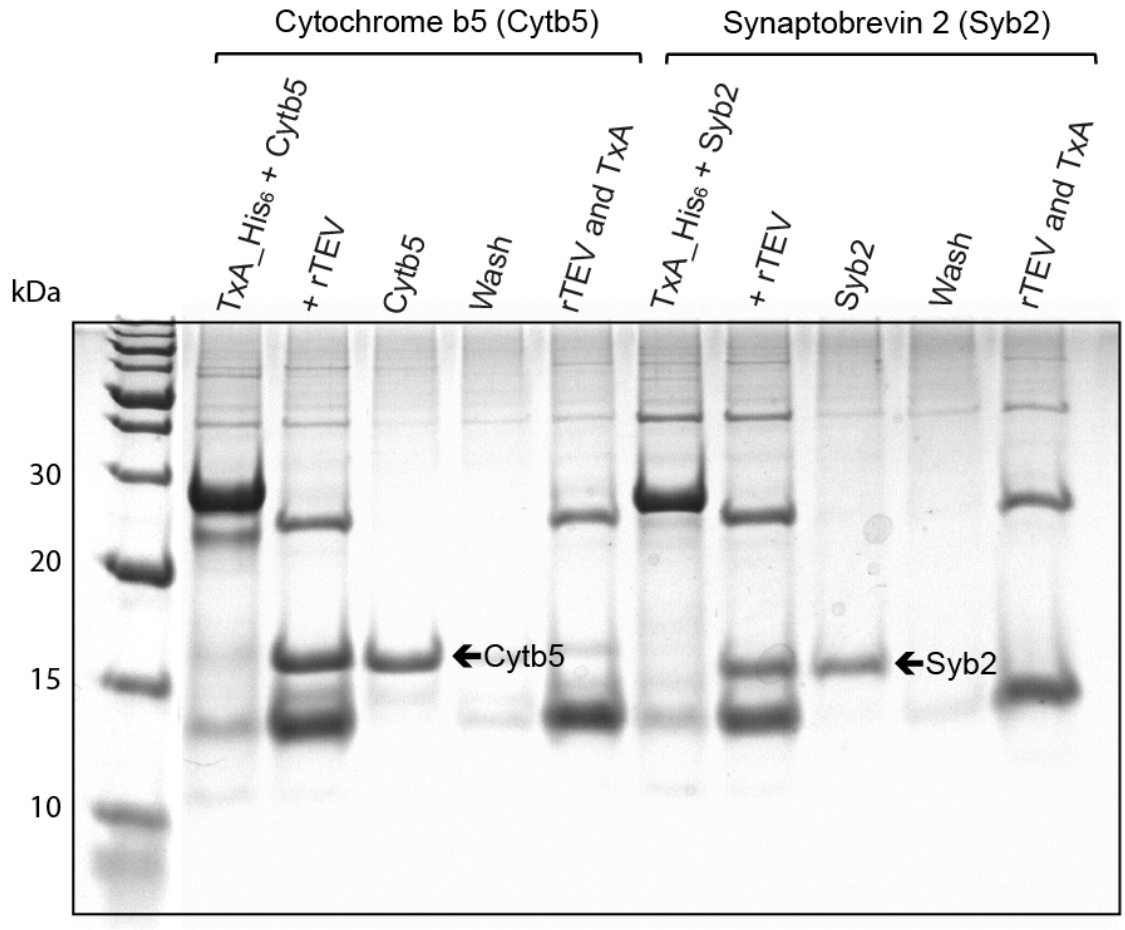



**Figure 4.3– Purification of tail-anchored (TA) membrane proteins.** SDS-PAGE gel showing different stages of TA-protein purification. Both Cytochrome b5 (Cytb5) and Synaptobrevin 2 (Syb2) were expressed as TxA fusion proteins and purified by Ni-affinity chromatography. rTEV protease was used to remove affinity tag and TxA protein yielding pure TA-proteins (indicated by arrows) obtained from the flow through of the second nickel column used in the purification.

## 4.2. Investigation of the interaction between SGTA and hydrophobic substrates using standard NMR methods

Using the abovementioned TPR_NNP construct, $^{1}H$-$^{15}N$ HSQC experiments were carried out, first to identify backbone amide signals that correspond to C-terminal residues of SGTA. An overlay of $^{1}H$-$^{15}N$ HSQC spectra of TPR_NNP with that of the TPR domain of SGTA (SGTA_TPR) is shown in Fig 4.4. This comparison reveals backbone amide signals corresponding to C-terminal residues of SGTA. These backbone amide signals were assigned by Dr Santiago Martinez-Lumbreras using standard triple resonance experiments. Subsequently, using this information, titrations were carried out by adding unlabelled TA proteins to the $^{15}N$ TPR_NNP construct (Figs. 4.5-4.7). Titrations with TA proteins cytochrome b5 (Cytb5) (Fig. 4.5) and synaptobrevin 2 (Syb2) (Fig. 4.6) were both very similar, with chemical shift perturbations observed in backbone amide signals corresponding to some residues present in the TPR capping helix (residues 205-210) along with certain observable C-terminal residues (Figs. 4.7 - 4.9). This revealed some detail with respect to the binding of hydrophobic substrates, describing a direct interaction of such substrates with SGTA regions distal to the TPR capping helix encompassing the NNP region of the C-terminal domain.

However, in $^{1}H$-$^{15}N$ HSQC spectra of both C-terminal domain constructs, namely TPR_NNP and C-term, peak absence has lead to incompleteness in observable amide signals corresponding to regions within the C-terminal domain. In total, 20 residues in the NNP region and 30 residues in the Q-rich region of the C-terminal domain could not be assigned (Chapter 3). Thus, a complete picture of substrate binding could not be gleaned from NMR titrations using $^{1}H$-$^{15}N$ HSQC spectra of SGTA constructs that include, wholly or in part, its C-terminal domain.

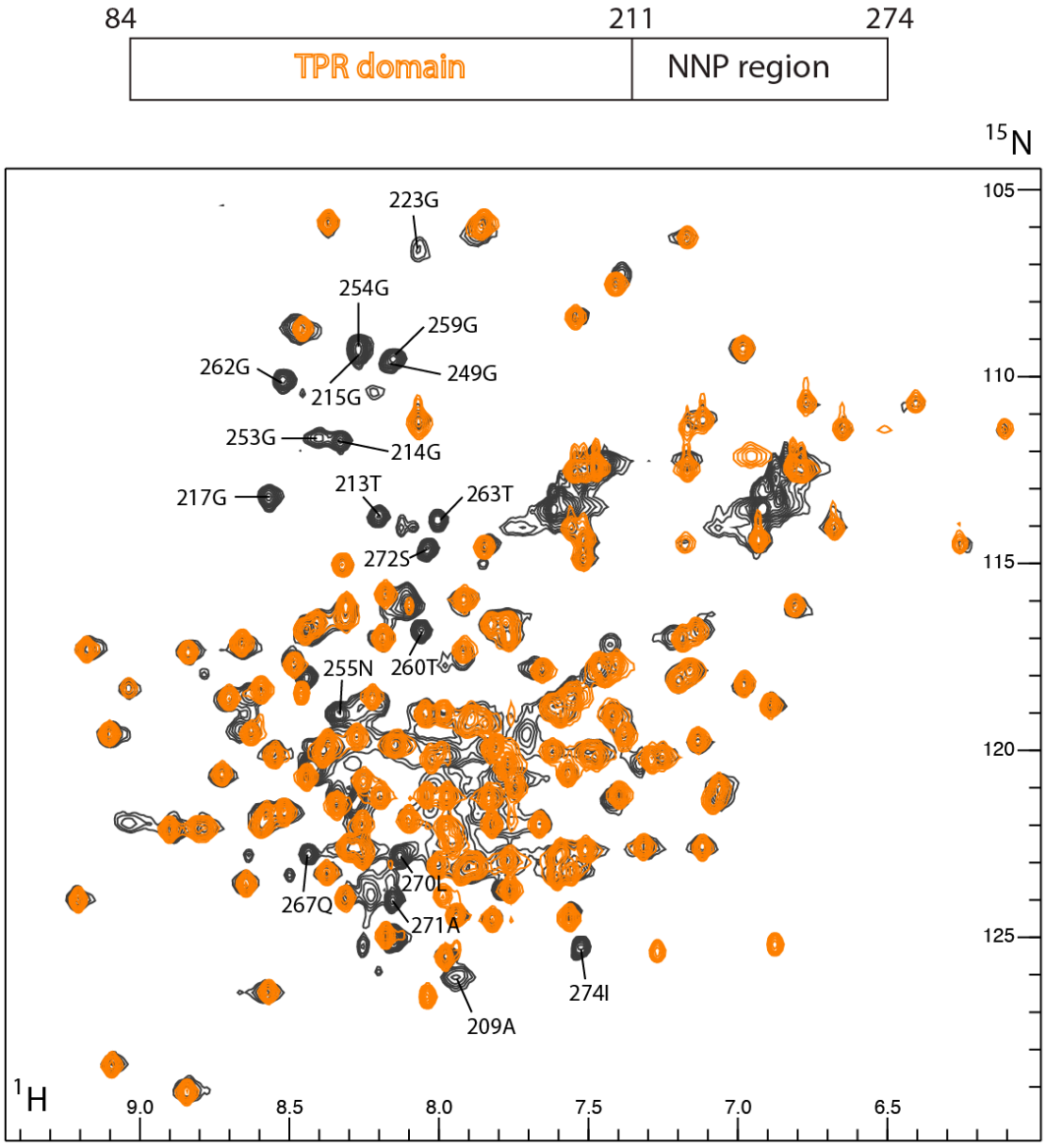



**Figure 4.4- Identification of backbone amide signals corresponding to C-terminal residues of SGTA from the $^{1}$H-$^{15}$N HSQC of the TPR_NNP construct.** Overlay of $^{1}$H-$^{15}$N HSQC of the SGTA_TPR_NNP construct (residues 84-274; *in black*) with the $^{1}$H-$^{15}$N HSQC of SGTA_TPR (residues 84-211; *in orange*), allowing for identification of amide signals that belong to the C-terminal domain (labelled). Spectra were acquired on a 500 MHz spectrometer at 25°C.

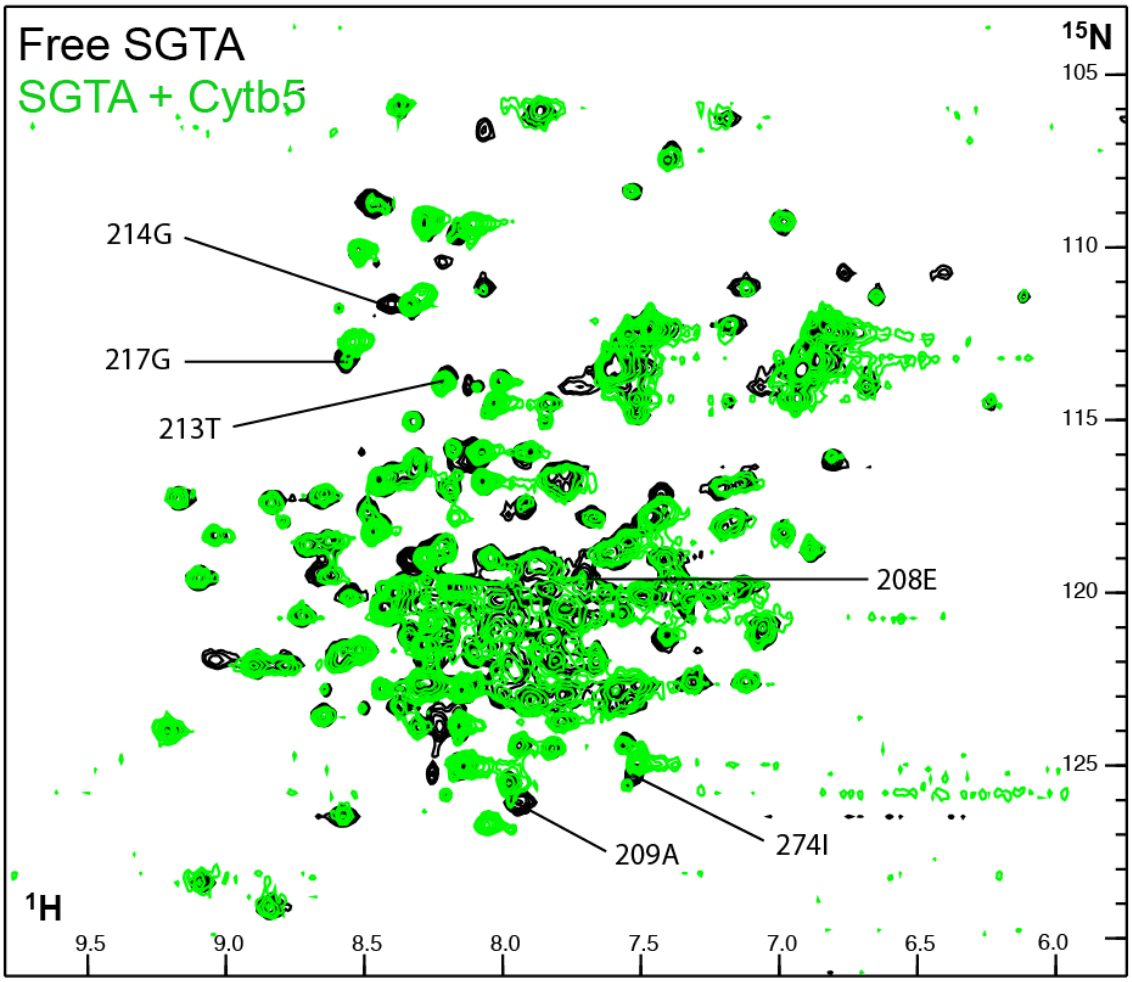



**Figure 4.5– SGTA interacts with cytochrome b5 (Cytb5).** $^{1}H$-$^{15}N$ HSQC of the TPR_NNP construct at 150 µM (*in black*), overlaid with the $^{1}H$-$^{15}N$ HSQC of the TPR_NNP construct in the presence of full-length unlabelled cytochrome b5 (*in green*) at a 1:1.5 molar ratio. Spectra were acquired on a 500 MHz spectrometer at 25°C. Residues for which changes in amide signals are observed upon association with Cytb5 have been labelled.

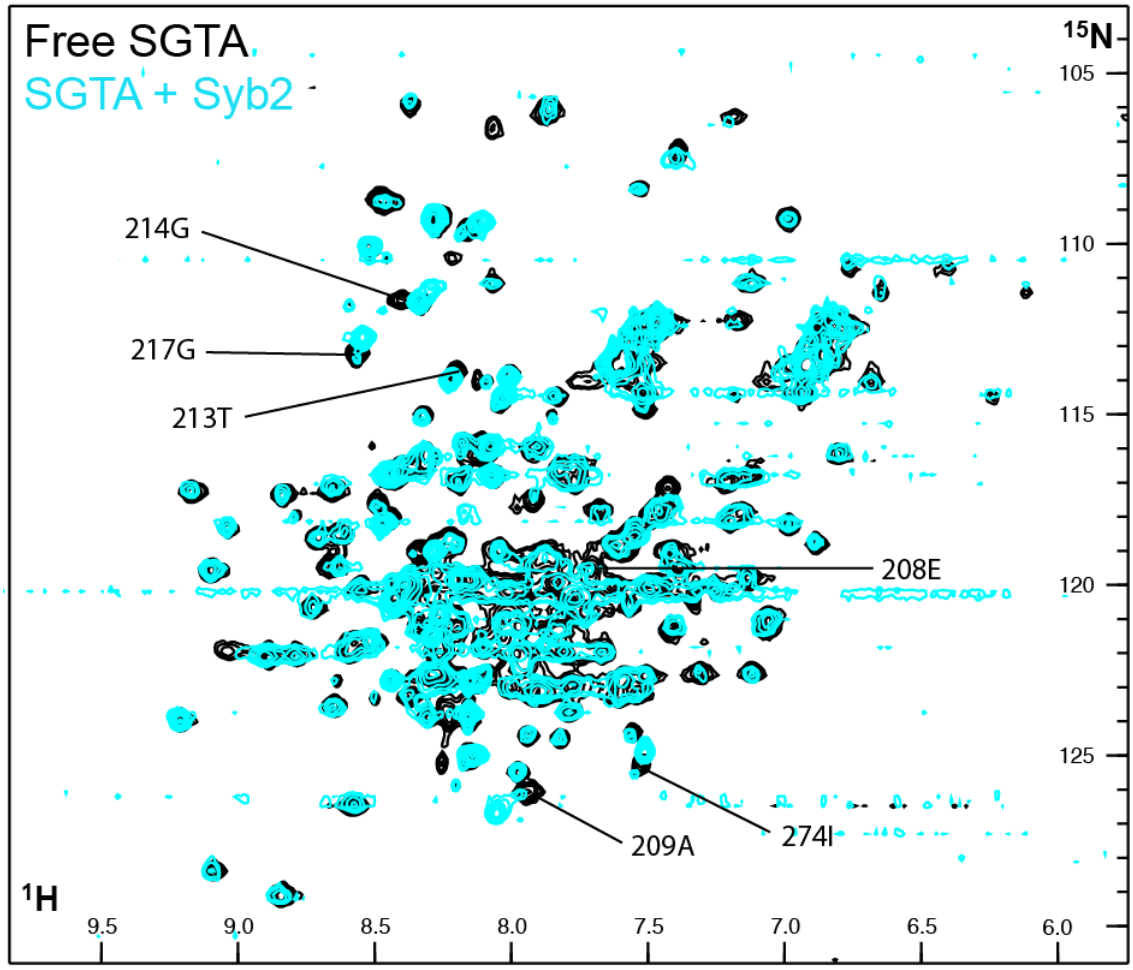



**Figure 4.6– SGTA interacts with synaptobrevin 2 (Syb2).** $^{1}$H-$^{15}$N HSQC of the TPR_NNP construct at 150 µM (*in black*), overlaid with the $^{1}$H-$^{15}$N HSQC of the TPR_NNP construct in the presence of full-length unlabelled Syb2 (*in light green*) at a 1:1.5 molar ratio. Spectra were acquired on a 500 MHz spectrometer at 25°C. Residues for which changes in amide signals are observed upon association with Syb2 have been labelled.

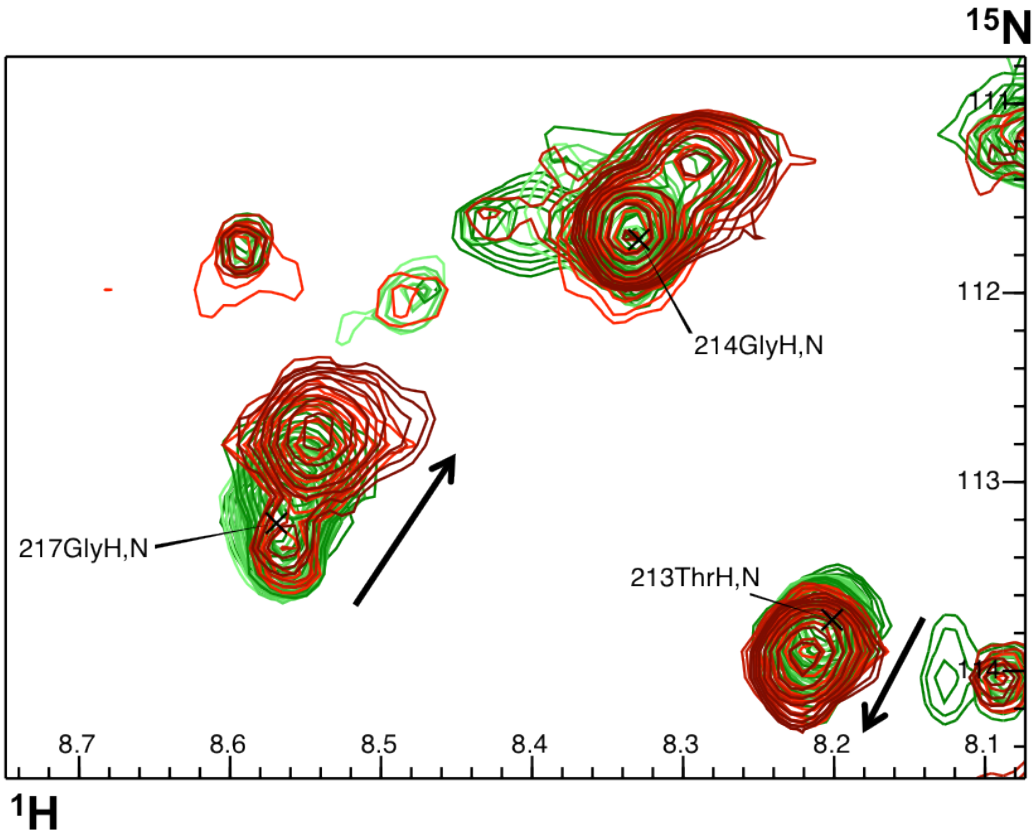



**Figure 4.7- Hydrophobic substrates interact with residues in the C-terminal domain of SGTA.** Region of the $^{1}H$-$^{15}N$ HSQC of the TPR_NNP construct showing perturbation of backbone amide signals corresponding to the C-terminal residues of SGTA upon titration with increasing amounts of Cytb5 with the final titration point approaching a 1:1.5 molar ratio.

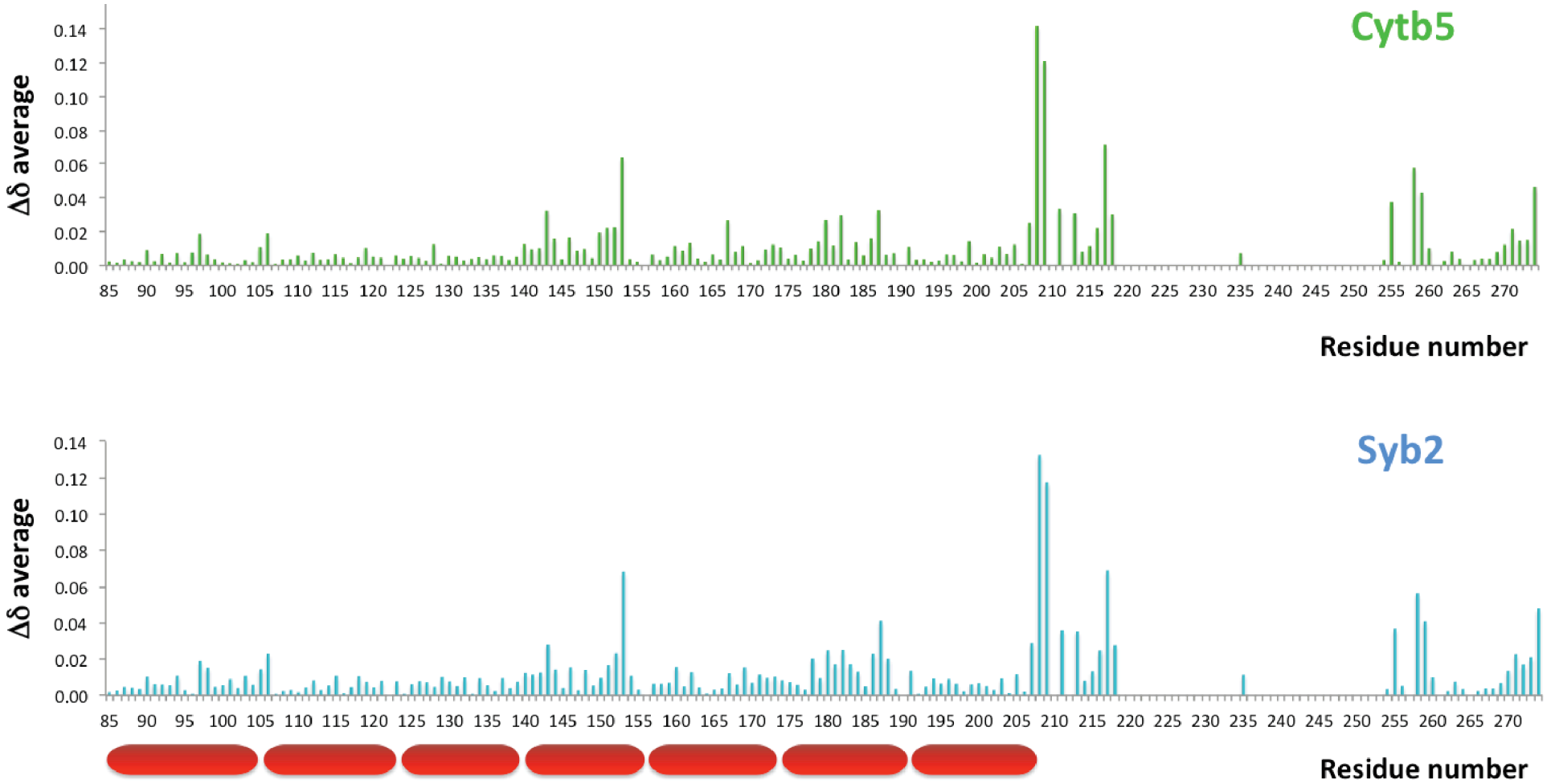



**Figure 4.8- Chemical shift perturbation (CSP) delta average values of TPR_NNP amide signals upon interaction with TA protein substrates.** CSP $\Delta\delta^{av}$ values were calculated for each observable amide signal within the TPR_NNP construct upon interaction with Cytb5 and Syb2. Positions of SGTA_TPR helices denoted in red.

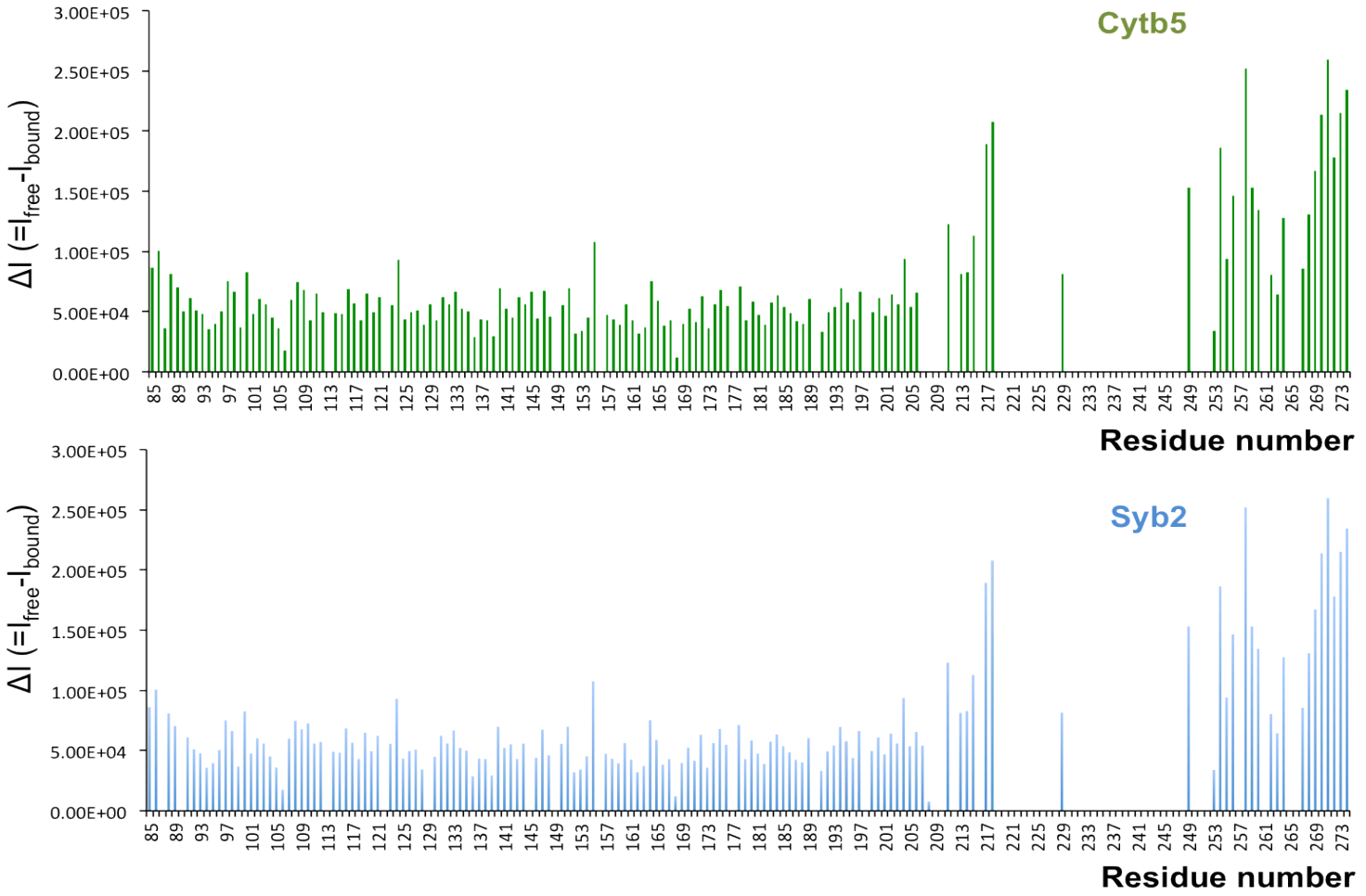



**Figure 4.9- Changes in intensity values of TPR_NNP amide signals upon interaction with TA protein substrates.** ΔI (=$I_{free}$-$I_{bound}$) values were calculated for each observable amide signal within the TPR_NNP construct upon interaction with TA proteins Cytb5 and Syb2.

### 4.3. Design and preparation of SGTA C-terminal cysteine mutants for $^{19}$F NMR

As regions of interest within the C-terminal domain could not be observed by standard NMR methods, a $^{19}$F NMR based approach was adopted to observe extrinsic $^{19}$F labels incorporated into the C-terminal domain via thiol groups on cysteine residues. The C-terminal domain of wild-type SGTA has no cysteine residues. Therefore, SGTA C-terminal domain cysteine mutants were designed, each carrying a single cysteine point mutation, such that the fluorine probes would be present flanking the NNP motifs, and the Q-rich region, in order to understand contributions of these regions to substrate binding. In addition, a cysteine mutant was introduced at the centre of the Q–rich region to further understand its role in shielding hydrophobic substrates. In all but one case, serine residues have been mutated to cysteines to minimise any structural perturbations (Fig 4.10).

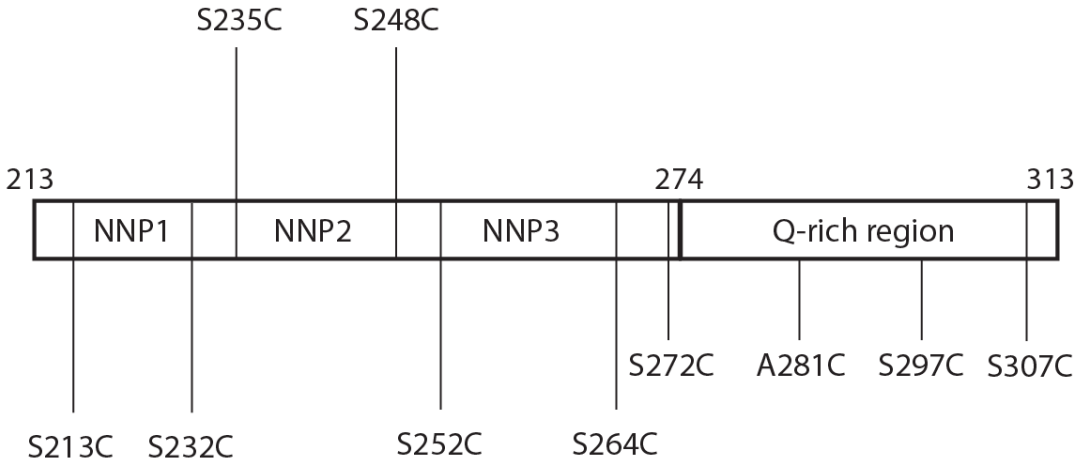



**Figure 4.10- Schematic of SGTA C-terminal domain showing positions of cysteine mutants generated to facilitate the post translational incorporation of fluorine probes (BTFA).** SGTA C-terminal domain cysteine mutants have been designed such that fluorine probes could be incorporated flanking the conserved NNP motifs, and the Q-rich region, in order to understand contributions of these regions to hydrophobic substrate binding.

All SGTA C-terminal domain cysteine mutants were successfully overexpressed and purified in an identical manner as described previously for SGTA constructs (Fig. 4.11). This was followed by the incorporation of 3-bromo-1,1,1-trifluoroacetone (BTFA), used as an extrinsic $^{19}F$ label, which was carried out as described in Chapter 2 (*Materials and Methods*).

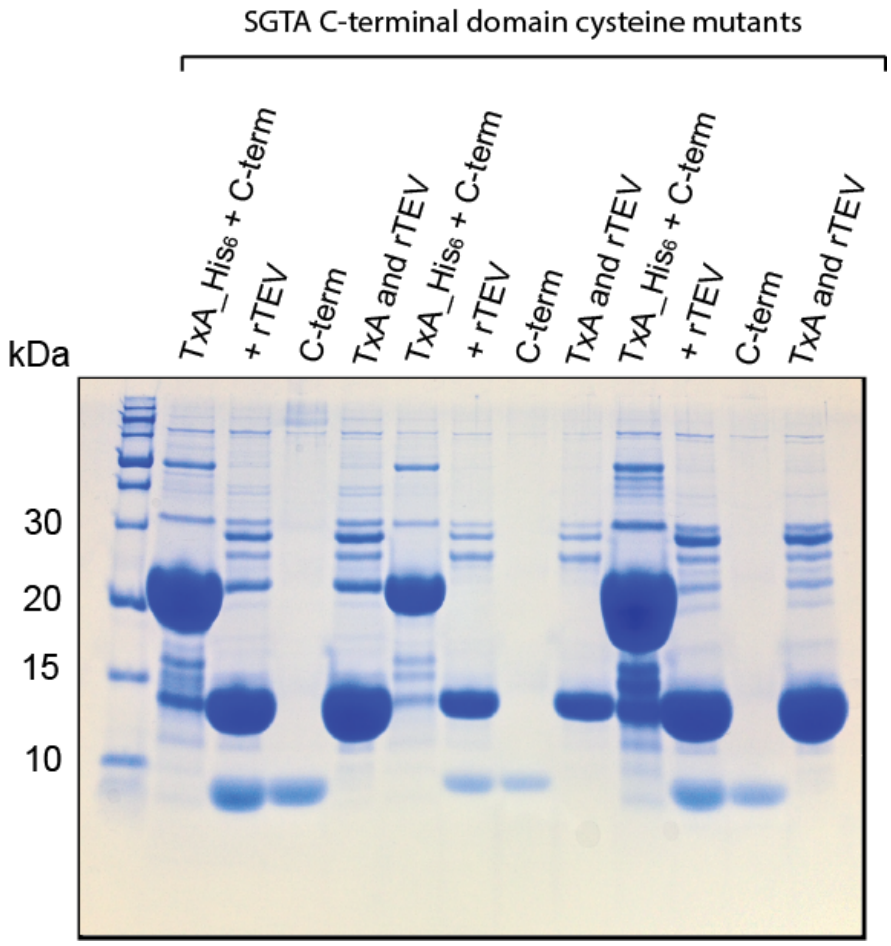



**Figure 4.11- Preparation of SGTA C-terminal domain cysteine mutants.** SDS-PAGE gel showing different stages in the purification of SGTA C-terminal domain cysteine mutants. A total of ten individual point mutants were purified (Fig. 4.9) in a manner identical to the purification of wild-type SGTA. This SDS-PAGE gel shows the preparation of three such cysteine mutants.

## 4.4. $^{19}$F NMR of SGTA C-terminal domain cysteine mutants

Each BTFA incorporated C-terminal cysteine mutant was then subjected to $^{19}$F NMR, with 1D spectra of each BTFA incorporated cysteine mutant acquired at 400 MHz. Free BTFA, used for chemical shift referencing, was prepared at 100 μM in $D_2O$ in a 3 mm NMR tube, which was then placed within the 5 mm tube containing the BTFA incorporated cysteine mutant prior to acquisition. Using this technique, it was possible to obtain 1D $^{19}$F NMR signals that correspond to each BTFA incorporated cysteine mutant. These include signals flanking each of the NNP motifs, namely NNP1 (from S218C and S232C; Fig. 4.12), NNP2 (from S235C and S248C; Fig. 4.13) and NNP3 (from S252C and S264C; Fig. 4.14). Also, with respect to the Q-rich region, all four signals could be observed (S272C, A281C, S297C and S307C; Fig. 4.15). Interestingly, $^{19}$F NMR signals arising from mutants present within close proximity to the C-terminal domain boundaries (S218C and S307C) appear as intense narrow resonances, indicative of an unfolded state. This is in agreement with the random coil-chemical shifts that have been observed for these regions (Fig. 3.13). On the other hand, $^{19}$F NMR signals from S252C and S264C flanking the NNP3 motif, and those from S272C and A281C within the Q-rich region appear to have undergone considerable peak broadening. This effect could possibly be attributed to the fact that these residues, or regions in the vicinity thereof, may be undergoing chemical exchange between different states. Additionally, due to the extremely sensitive nature of $^{19}$F chemical shifts to changes in local environment, the abovementioned $^{19}$F NMR signals provide us with means to understand the somewhat elusive C-terminal domain, regions of which have thus far been intractable to analysis by standard NMR methods. In particular, perturbations of these signals can be monitored upon addition of substrate to gain insight into the nature and extent of substrate binding regions within the SGTA C-terminal domain.

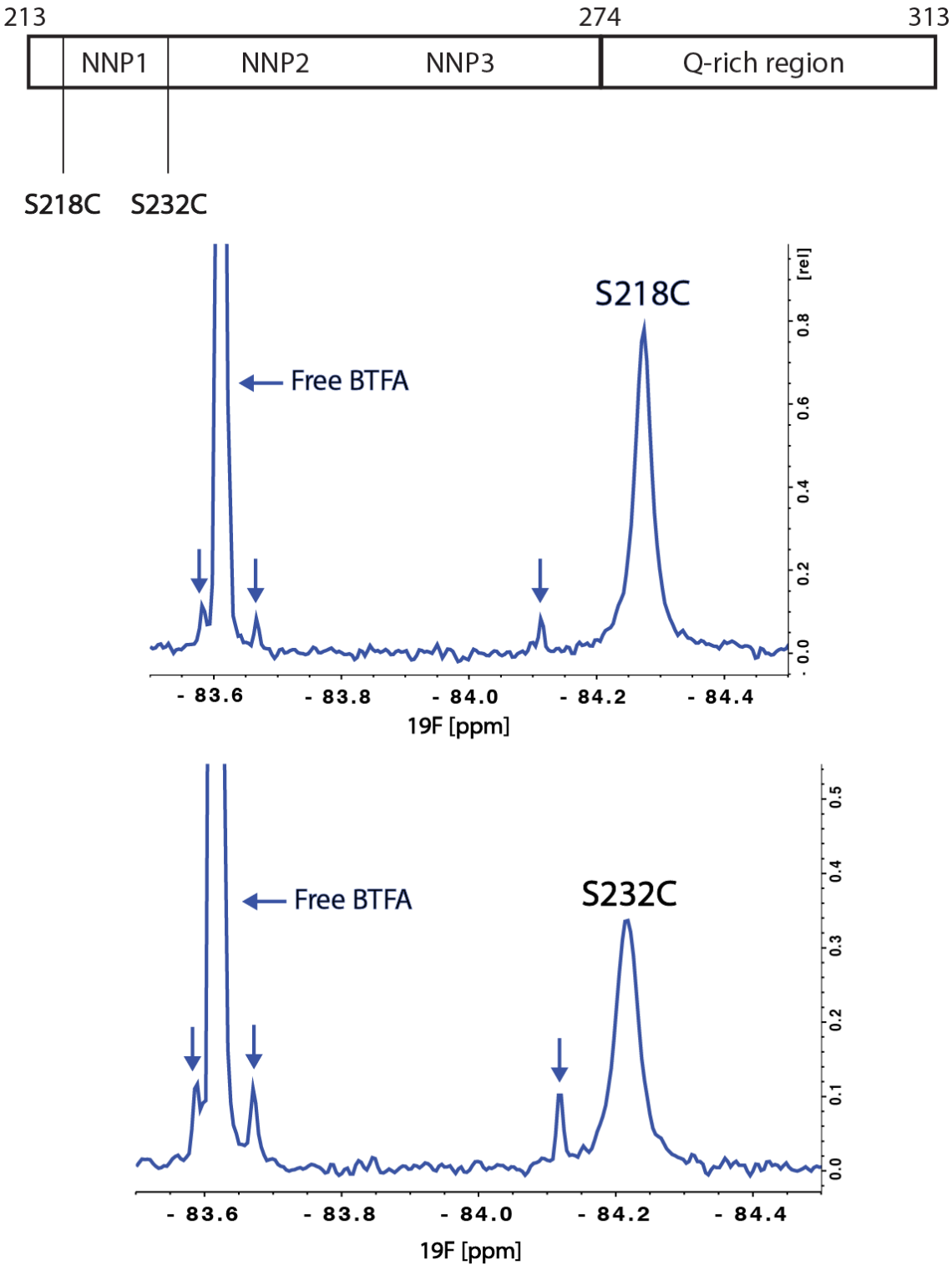



**Figure 4.12- 1D $^{19}$F fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 218 and 232 flanking the NNP1 motif.** Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct was present at a concentration of 250 µM.

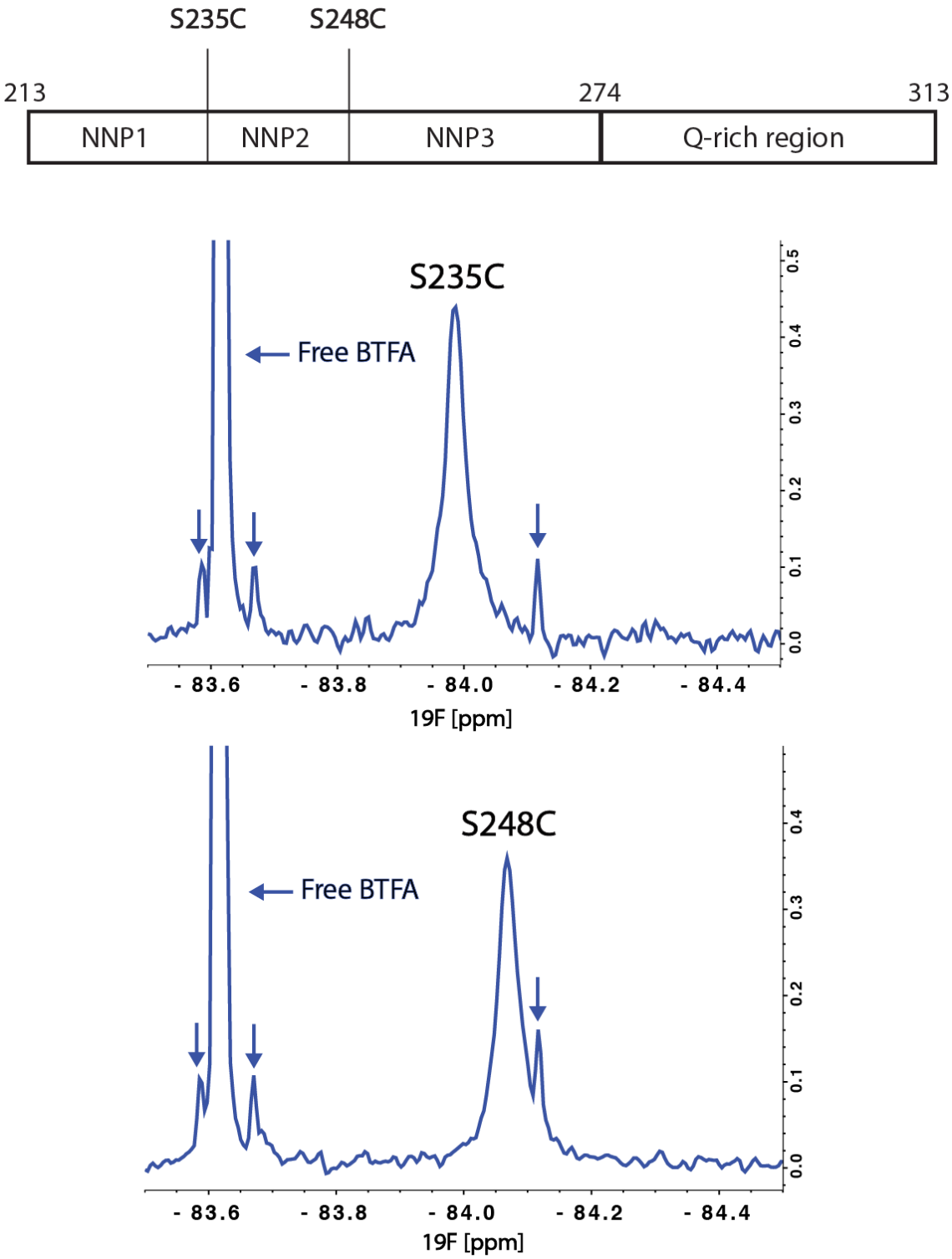



**Figure 4.13- 1D $^{19}$F fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 235 and 248 flanking the NNP2 motif.** Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct was present at a concentration of 250 μM.

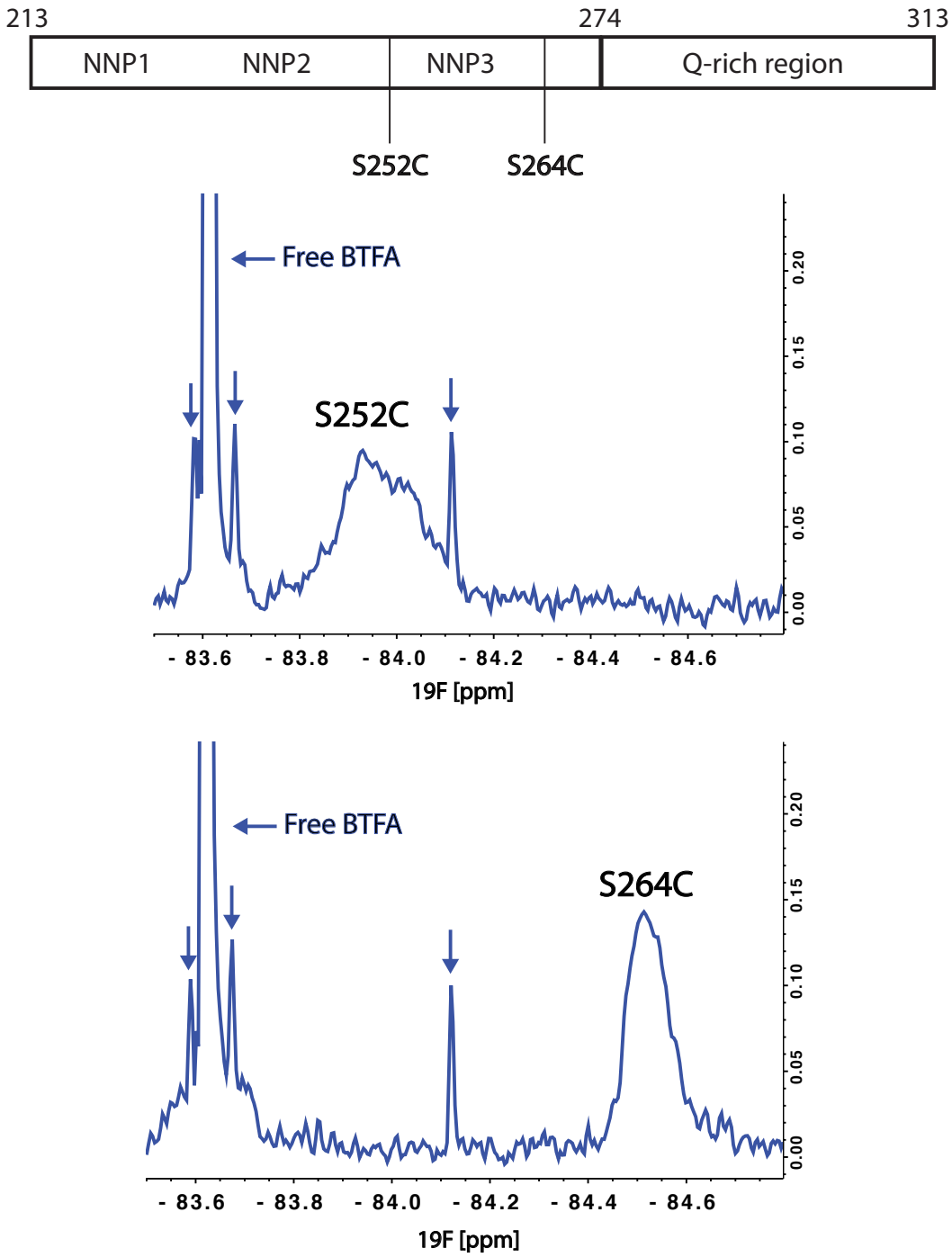



**Figure 4.14- 1D $^{19}$F fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 252 and 264 flanking the NNP3 motif.** Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct was present at a concentration of 250 μM.

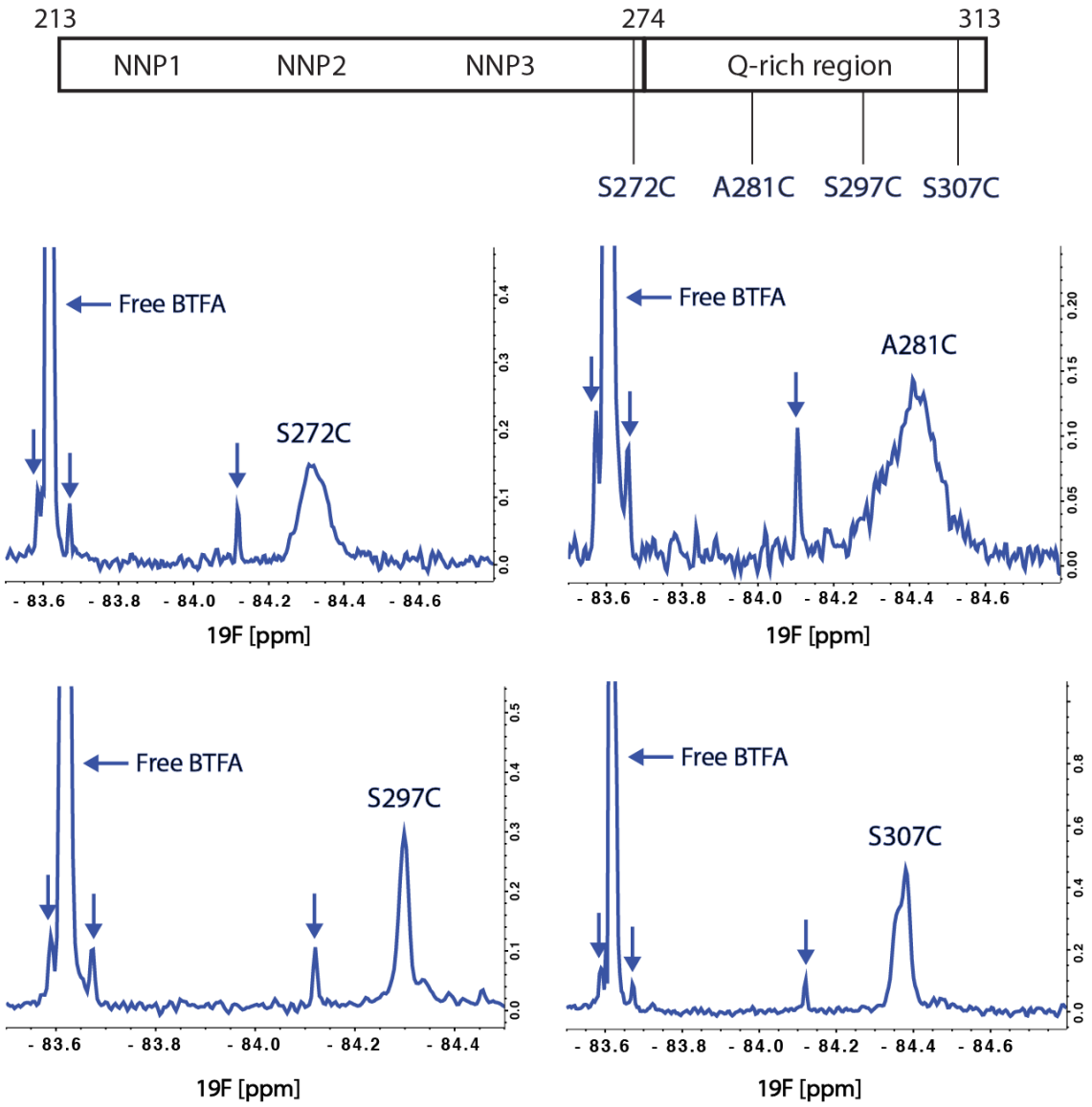



**Figure 4.15- 1D $^{19}$F fluorine spectra of SGTA C-terminal domain cysteine mutants with BTFA conjugated at positions 272, 281, 297 and 307.** Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct was present at a concentration of 250 µM.

## 4.5. $^{19}$F NMR to investigate the interaction between SGTA and Cytb5

Hence to understand the nature and extent of the hydrophobic substrate binding C-terminal domain of SGTA, and contributions of regions therein, aforementioned extrinsic $^{19}$F chemical shifts were monitored for perturbations upon substrate addition. The full-length tail-anchored protein cytochrome b5 (Cytb5) was used as a validated substrate consisting of a hydrophobic TMD helix (Fig. 4.1B). The effect of adding Cytb5 at molar ratios approaching 1:1 with respect to each SGTA C-terminal domain cysteine mutant have been shown in Figs 4.16 - 4.19. With regard to $^{19}$F NMR signals flanking the NNP1 motif, namely ones arising from S218C and S232C, only very subtle perturbations in chemical shifts can be observed, together with a decrease in the intensity of the signal arising from S218C upon addition of substrate (Fig. 4.16). The equivalent effect of substrate addition on $^{19}$F NMR signals from S235C and S248C positioned at either side of NNP2, appears to be different, with chemical shift perturbations of 0.05 ppm observed in both cases. This occurs together with a decrease in the intensity of both signals when Cytb5 is added (Fig. 4.17). Furthermore, the S252C signal flanking the NNP3 motif can be observed transitioning to a narrower lineshape together with a chemical shift change of 0.125 ppm upon association with Cytb5 (Fig. 4.18). As mentioned previously, the relatively broad $^{19}$F NMR signal emanating from S252C could be representative of exchange between different states. Therefore, addition of Cytb5 may have caused this region to select a 'substrate-bound' state, subsequently leading to the appearance of a narrower lineshape. Also, with the other signal flanking NNP3, namely S264C, a 0.1 ppm change in chemical shift can be observed upon Cytb5 addition (Fig. 4.18). Thus it appears that addition of full length Cytb5, incorporating its extreme C-terminal hydrophobic TMD helix, appears to induce changes in the local environment within the NNP region. In particular, the effect of substrate addition appears to be more pronounced for regions in the vicinity of the NNP2 and NNP3 motifs.

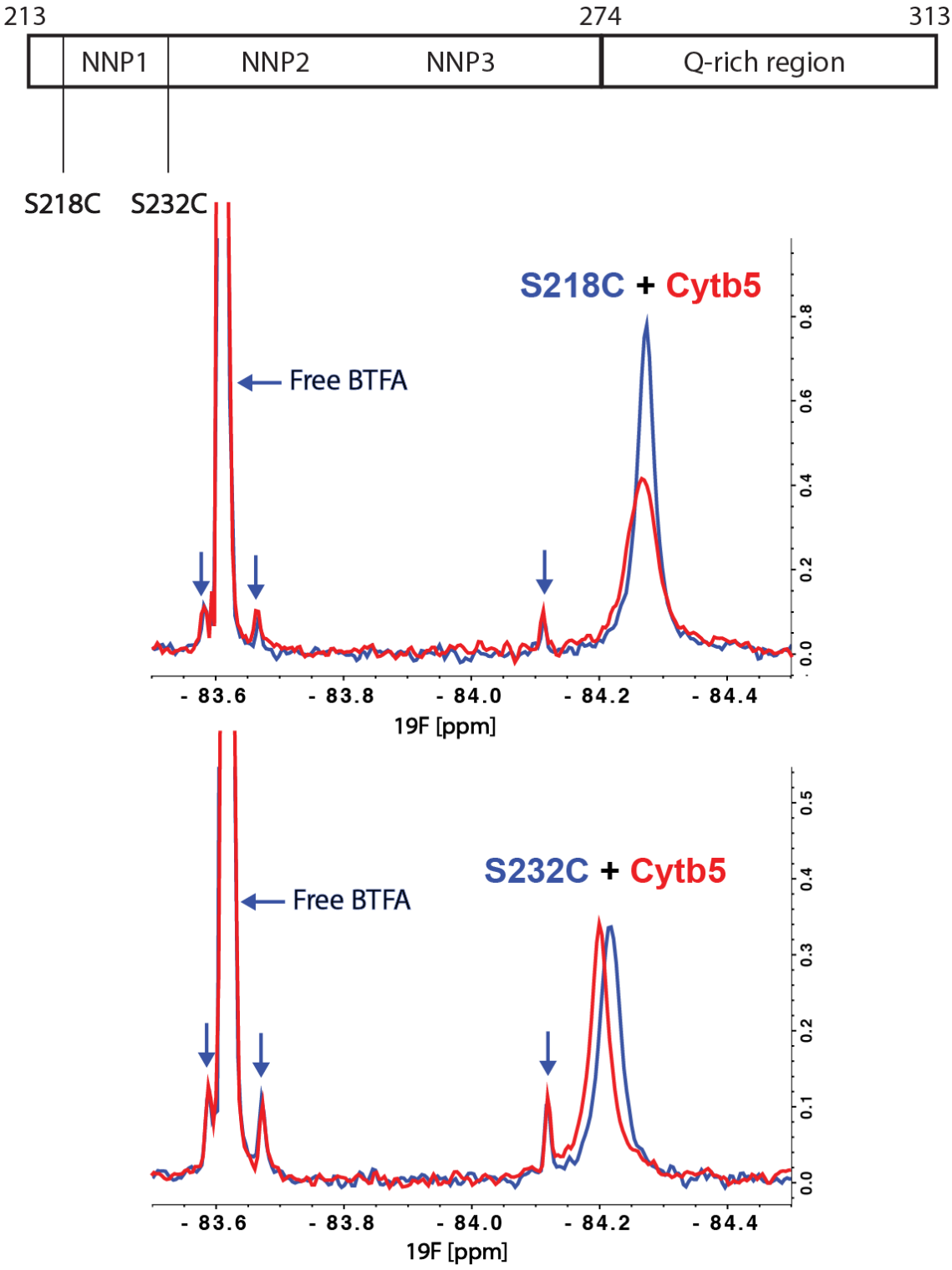



**Figure 4.16- 1D $^{19}$F fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S218C and S232C, in the presence of Cytb5.** Overlay of 1D $^{19}$F fluorine spectra in the absence of Cytb5 (*in blue*) and in the presence of Cytb5 at a 1:1 molar ratio (*in red*). Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct and full-length Cytb5 were each present at a concentration of 250 μM.

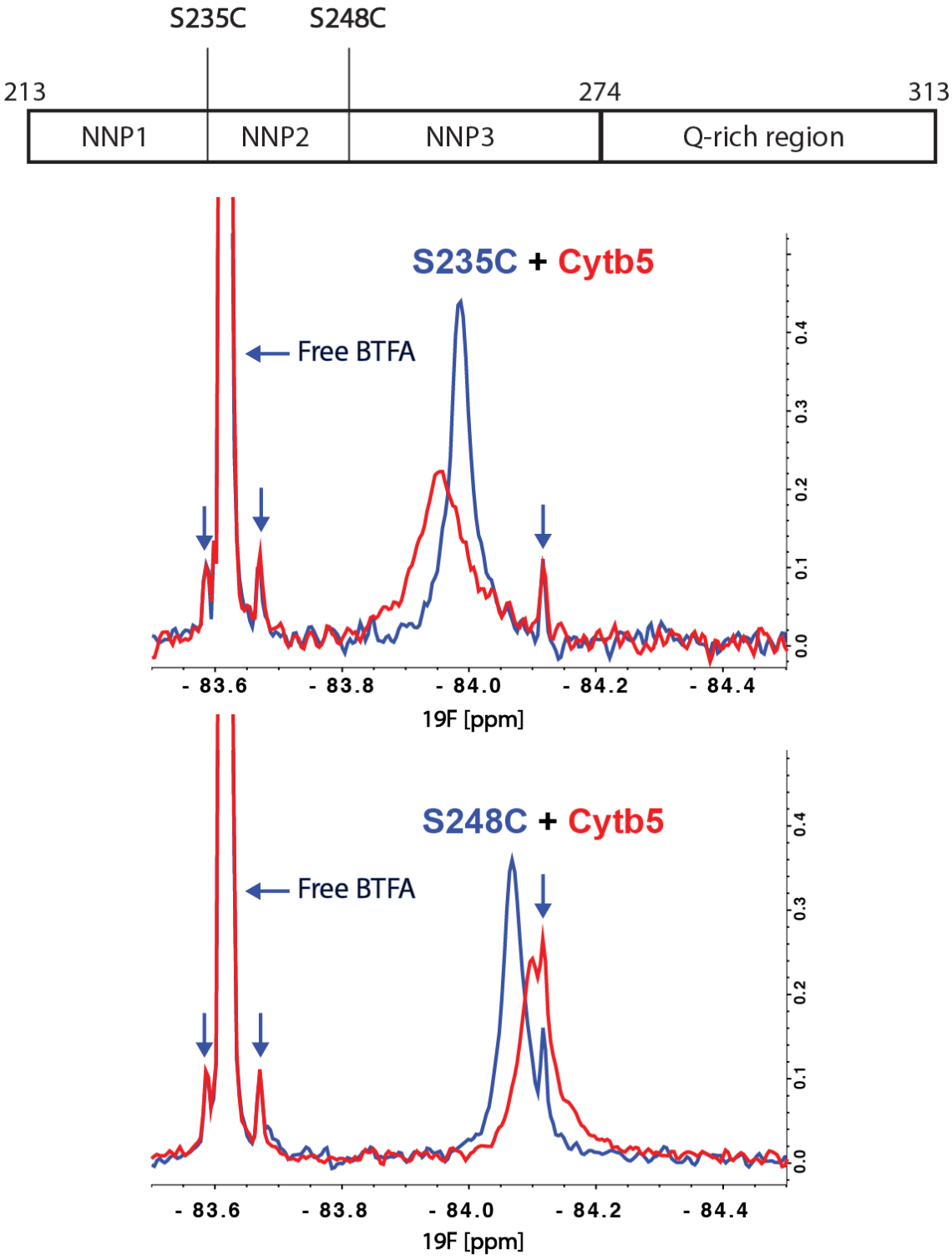



**Figure 4.17- 1D $^{19}$F fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S235C and S248C, in the presence of Cytb5.** Overlay of 1D $^{19}$F fluorine spectra in the absence of Cytb5 (*in blue*) and in the presence of Cytb5 at a 1:1 molar ratio (*in red*). Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct and full-length Cytb5 were each present at a concentration of 250 μM.

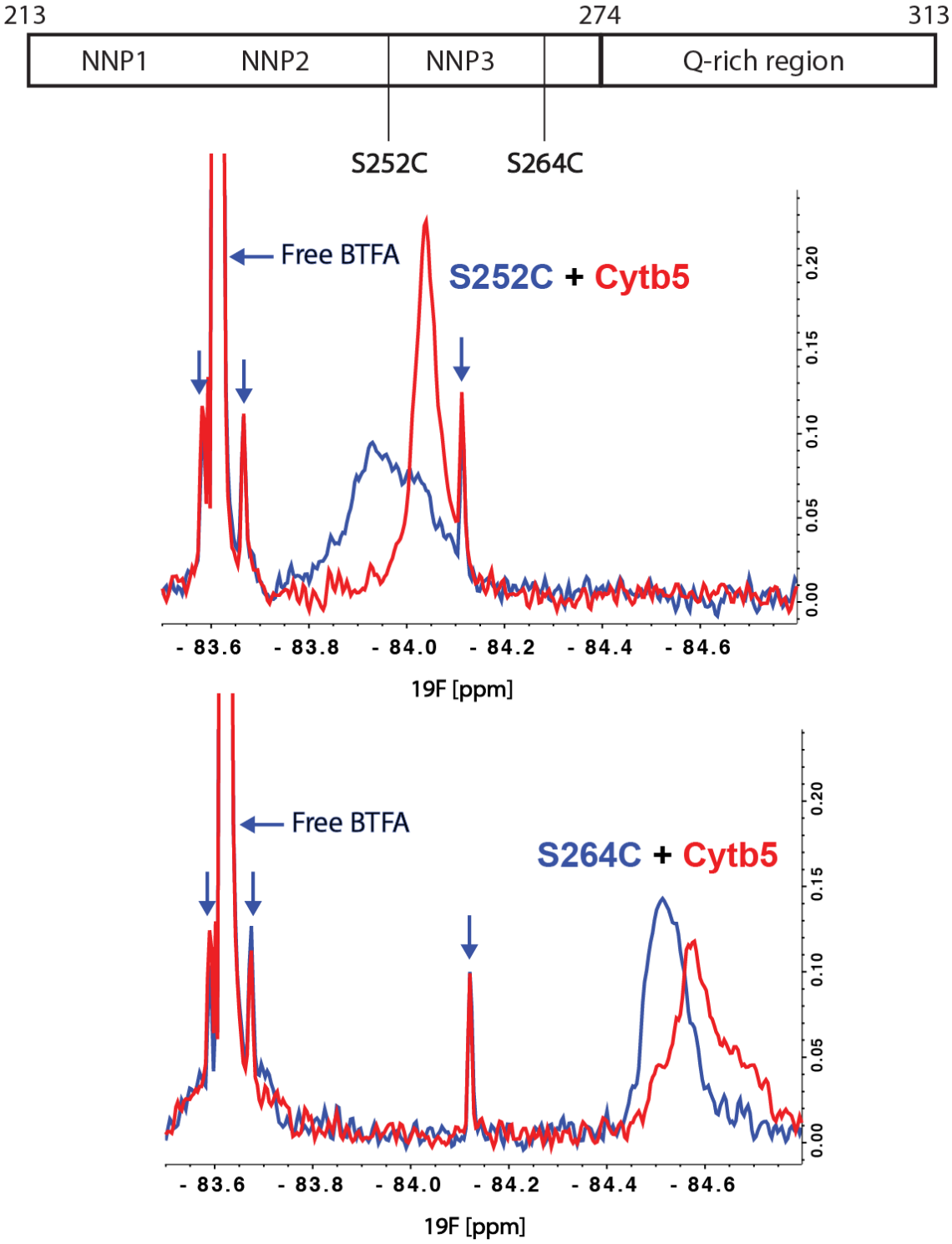



**Figure 4.18- 1D $^{19}$F fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S252C and S264C, in the presence of Cytb5.** Overlay of 1D $^{19}$F fluorine spectra in the absence of Cytb5 (*in blue*) and in the presence of Cytb5 at a 1:1 molar ratio (*in red*). Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct and full-length Cytb5 were each present at a concentration of 250 μM.

A similar set of experiments has been carried out to understand the contribution of the Q-rich region to substrate binding (Fig. 4.19). These experiments show that the $^{19}F$ NMR signal arising from S272C appears narrower upon Cytb5 binding, accompanied by a chemical shift change of around 0.05 ppm. Likewise, a transition to a narrow lineshape can be observed with the signal corresponding to the A281C. As observed in the NNP region, these changes could possibly be indicative of a loss of exchange between different states, occurring within these regions upon substrate binding. Furthermore, perturbations of $^{19}F$ chemical shifts can also be observed with the S297C mutant upon Cytb5 addition, while the chemical shift of the signal arising from S307C mutant does not appear to be perturbed. However, an increase in peak intensity can be observed in the $^{19}F$ NMR signals corresponding to S297C and S307C upon Cytb5 addition.

Therefore, based on the $^{19}F$ NMR observations described above, it appears that the addition of the hydrophobic substrate Cytb5 evokes change in the non-covalent environment of the C-terminal domain of SGTA, to varying extents, throughout both its NNP and Q-rich regions. These findings suggest that the NNP2 and NNP3 motifs are most affected within the NNP region (residues 213-274). In addition, the Q-rich region is also affected by Cytb5 upon interaction, as evident by perturbations of $^{19}F$ NMR signals emanating from this region. Thus it appears that both regions within the C-terminal domain (NNP and Q-rich regions) are necessary to facilitate SGTA's functionality in quality control to triage hydrophobic substrates.

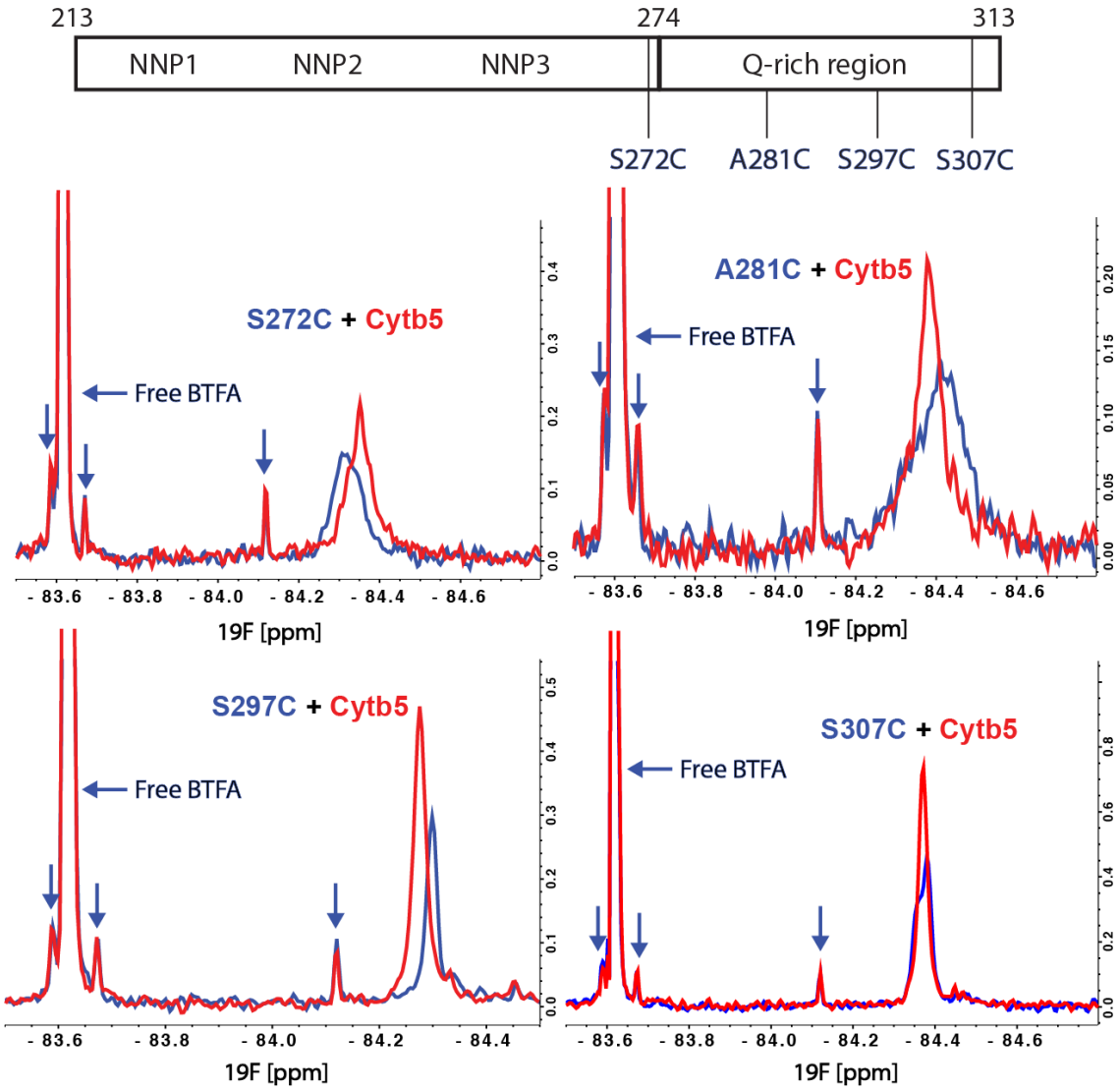



**Figure 4.19- 1D $^{19}$F fluorine spectra of BTFA conjugated C-terminal domain cysteine mutants S272C, A281C, S297C and S307C in the presence of Cytb5.** Overlay of 1D $^{19}$F fluorine spectra in the absence of Cytb5 (*in blue*) and in the presence of Cytb5 at a 1:1 molar ratio (*in red*). Free BTFA was used for chemical shift referencing, and peaks arising from free BTFA are indicated by blue arrows. Spectra were acquired on a 400 MHz spectrometer at 25°C. The SGTA C-terminal domain construct and full-length Cytb5 were each present at a concentration of 250 µM.

## 4.6. Summary

The experiments described in this chapter demonstrate the interaction of hydrophobic TA substrates, namely Cytb5 and Syb2, with SGTA. The use of standard 2D NMR experiments ($^{1}$H-$^{15}$N HSQCs) has provided indications that Cyb5 and Syb2 interact with the distal part of the TPR capping helix and with the NNP region of the SGTA C-terminal domain. Furthermore, the observation of extrinsic $^{19}$F NMR labels introduced through cysteine mutants within the C-terminal domain was, for the first time, able to provide observable NMR signals corresponding to regions within the C-terminal domain that were hitherto inaccessible due to signal absence in $^{1}$H-$^{15}$N HSQC spectra, an absence that persisted even upon deuteration of C-terminal constructs. Subsequently, $^{19}$F NMR findings have provided details regarding the interaction of Cytb5 with different regions within the C-terminal domain, namely its NNP and Q-rich regions. To this end, it appears that both regions within the C-terminal domain contribute to its interaction with Cytb5. These findings thus improve our understanding with regard to this crucial, yet poorly understood, substrate binding C-terminal domain of SGTA, with possible implications pertaining to its biological role in shielding hydrophobic substrates in the cytosol.

# Chapter 5

# Biophysical characterisation of the interaction between SGTA and the intrinsic proteasomal ubiquitin receptor Rpn13

Recent evidence suggests that an SGTA dependent quality control cycle involving hydrophobic substrates could be operating at the 19S subunit of the 26S proteasome (Leznicki *et al.*, 2015). This chapter describes structural and biophysical characterisation of the interaction between SGTA and the proteasome, an interaction mediated via the intrinsic proteasomal ubiquitin receptor Rpn13 at the 19S regulatory subunit. Work described in this chapter has been published in Thapaliya *et al.* (2016).

## 5.1 Expression and purification of Rpn13 and SGTA_TPR constructs

In order to characterise molecular details underlying the interaction between SGTA and the intrinsic proteasomal ubiquitin receptor Rpn13, a C-terminal domain construct of Rpn13 was designed, guided by earlier mapping studies (Leznicki *et al.*, 2015). This construct, encoding C-terminal residues 260-407 of human Rpn13 (hereafter referred to as $Rpn13_{260-407}$) was successfully cloned, overexpressed in *E. coli*, and purified, as described in Chapter 2 (*Materials and Methods*). $Rpn13_{260-407}$ was expressed in *E. coli* Rosetta cells as a fusion protein with an N-terminally placed thioredoxin A followed by a hexahistidine tag and a rTEV cleavage site, to improve recombinant protein expression and solubility. This fusion protein overexpressed in both rich and minimal media, both at 37°C for 4 hrs, and at 18°C overnight (Fig. 5.1). For interaction studies, $Rpn13_{260-407}$ was overexpressed at 18ºC overnight and recovered in the soluble fraction, both in rich and minimal media, allowing for affinity purification, followed by size-exclusion chromatography (SEC) to obtain pure protein samples. Recombinant $Rpn13_{260-407}$ maintained its solubility after removal of thioredoxin A by rTEV digestion and could be obtained in adequate amounts to carry out binding studies using a range of biophysical techniques. SGTA_TPR was purified in a similar manner as described previously. In both cases, sample purity was assessed by SDS-PAGE, with purest samples used for subsequent protein-protein interaction studies (Fig. 5.2). Both

SGTA_TPR (residues 84-211) and $Rpn13_{260-407}$ were successfully separated on a 120 ml superdex 75 column with elution volumes of 77 ml and 69 ml, respectively (Fig. 5.3). The $Rpn13_{260-407}$ has a lower absorbance at 280 nm compared to SGTA_TPR due to a complete lack of tryptophan residues.

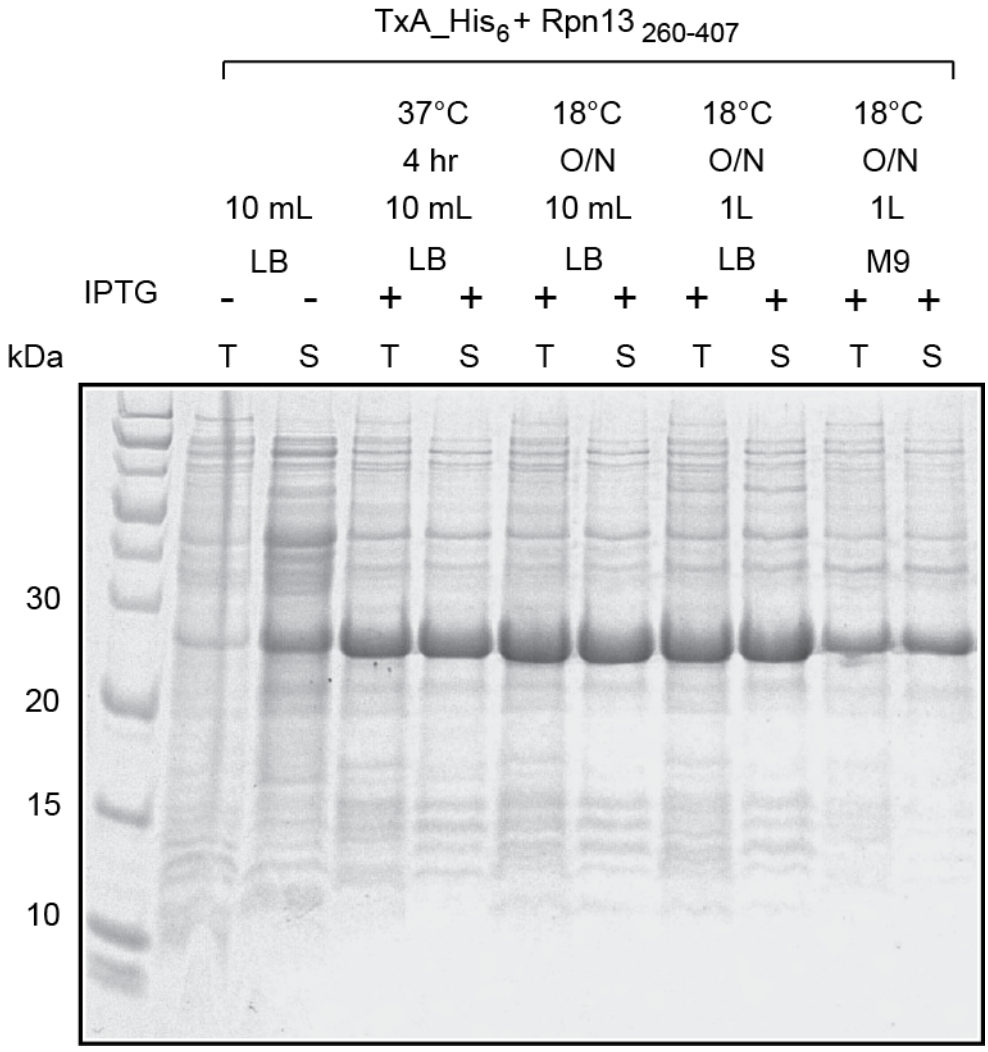



**Figure 5.1- Expression of the human intrinsic proteasomal receptor Rpn13 in *E. coli.*** The C-terminal domain of Rpn13 (residues 260-407) was expressed in *E. coli* Rosetta cells as a fusion protein with an N-terminally placed thioredoxin A followed by a hexahistidine tag ($TxA_His_6$ + $Rpn13_{260-407}$). Tests for the expression of recombinant Rpn13 were carried out at 18°C and 37°C, in both rich (LB) and minimal (M9) media. $TxA_His_6$ + $Rpn13_{260-407}$ could be successfully overexpressed in both rich and minimal media, as evident by the presence of bands on SDS-PAGE corresponding to the size of the fusion protein in both the total (T) and soluble (S) fractions of *E. coli* lysate. Overnight incubation with IPTG at 18°C was used for the subsequent production of Rpn13 in both LB and M9 media.

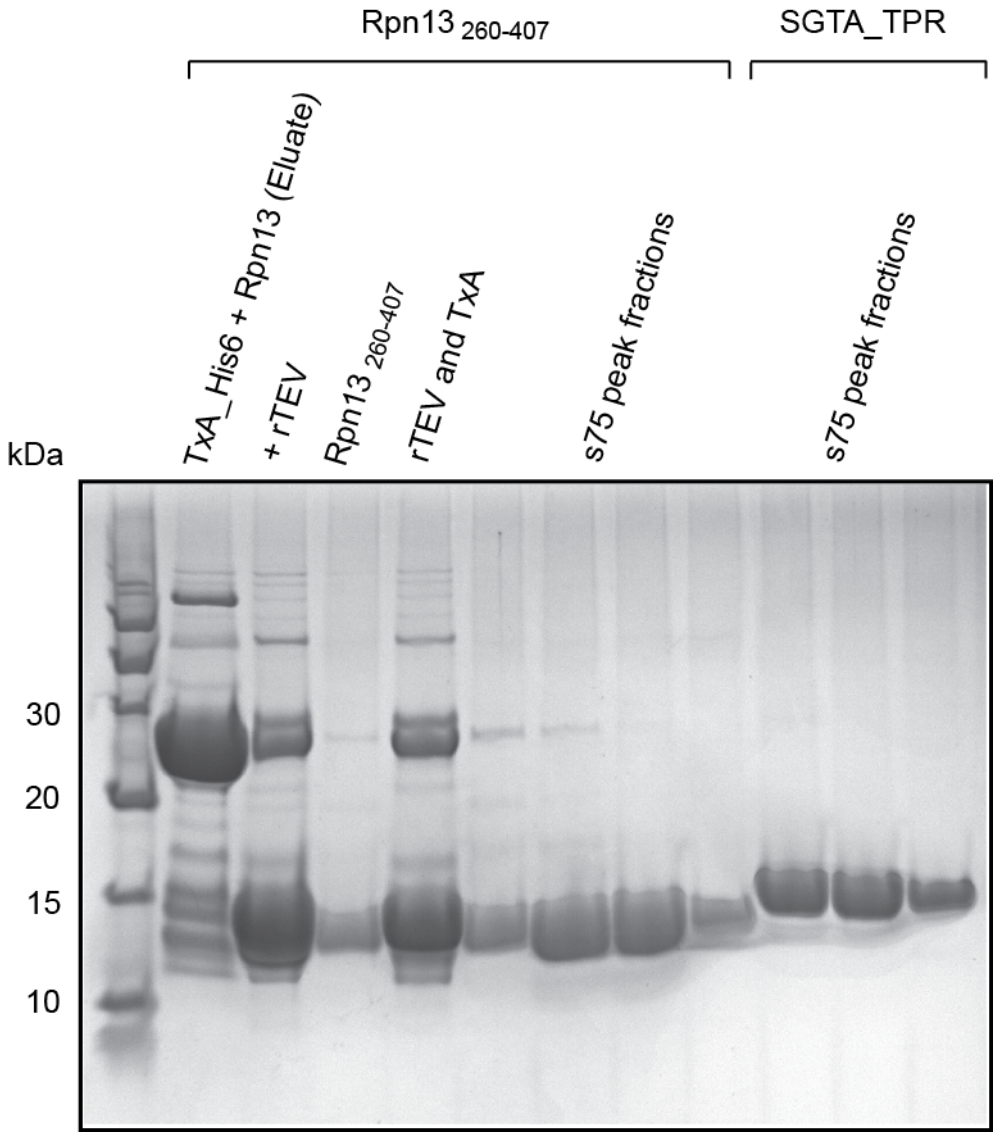



**Figure 5.2- Preparation of recombinant human Rpn13$_{260-407}$ and SGTA_TPR.** Rpn13$_{260-407}$ was overexpressed as a fusion protein, purified by nickel affinity chromatography and was recovered in the eluate of the first nickel column (SDS-PAGE fraction: TxA_His$_6$ + Rpn13). This was followed by affinity tag and fusion protein removal by treatment with rTEV (SDS-PAGE fraction: +rTEV). Pure protein was obtained by passing the rTEV treated fraction through a second nickel column with pure protein obtained in the flow-through (SDS-PAGE fraction: Rpn13$_{260-407}$). Separated thioredoxin A and rTEV was eluted from the second nickel column (SDS-PAGE fraction: rTEV and TxA). SGTA_TPR was purified following a similar protocol as described previously. SEC was used as a final purification step for both proteins (s75 peak fractions).

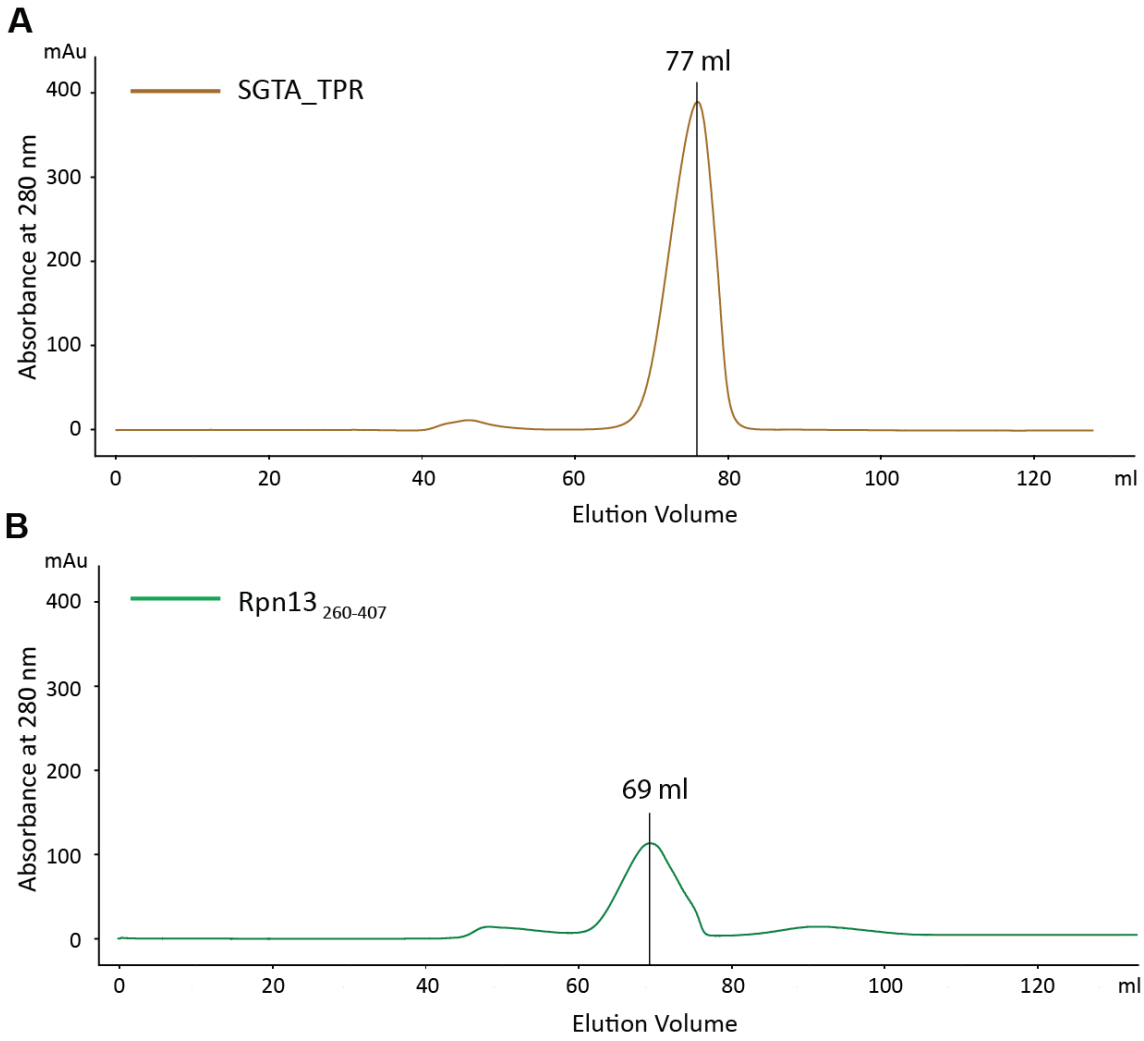



**Figure 5.3- Size exclusion chromatography of SGTA_TPR and Rpn13$_{260-407}$.** SEC was carried out on a 120 ml HiLoad 16/60 superdex 75 prep grade column at a flow rate of 1 ml/min. (A) SGTA_TPR has a peak elution volume of 77 ml (extinction coefficient: 13500 $M^{-1}cm^{-1}$). (B) Rpn13$_{260-407}$ elutes at 69 ml, and has lower absorbance at 280 nm compared to a similar amount of SGTA_TPR due to a complete lack of tryptophan residues (extinction coefficient: 1490 $M^{-1}cm^{-1}$).

## 5.2 SGTA interacts with the C-terminal domain of Rpn13

The interaction between SGTA_TPR and $Rpn13_{260-407}$ was investigated by size-exclusion chromatography (SEC) experiments. SGTA_TPR and $Rpn13_{260-407}$ were combined in a 1:1 molar ratio and analysed by SEC on a 120 ml HiLoad 16/60 superdex 75 column at a flow rate of 1 ml/min. The resulting elution volume (62 ml) was clearly indicative of the formation of a stable complex between SGTA_TPR and $Rpn13_{260-407}$ (Fig. 5.4). This was assessed based on the elution volumes of SGTA_TPR and $Rpn13_{260-407}$ when analysed on the same column under identical conditions.

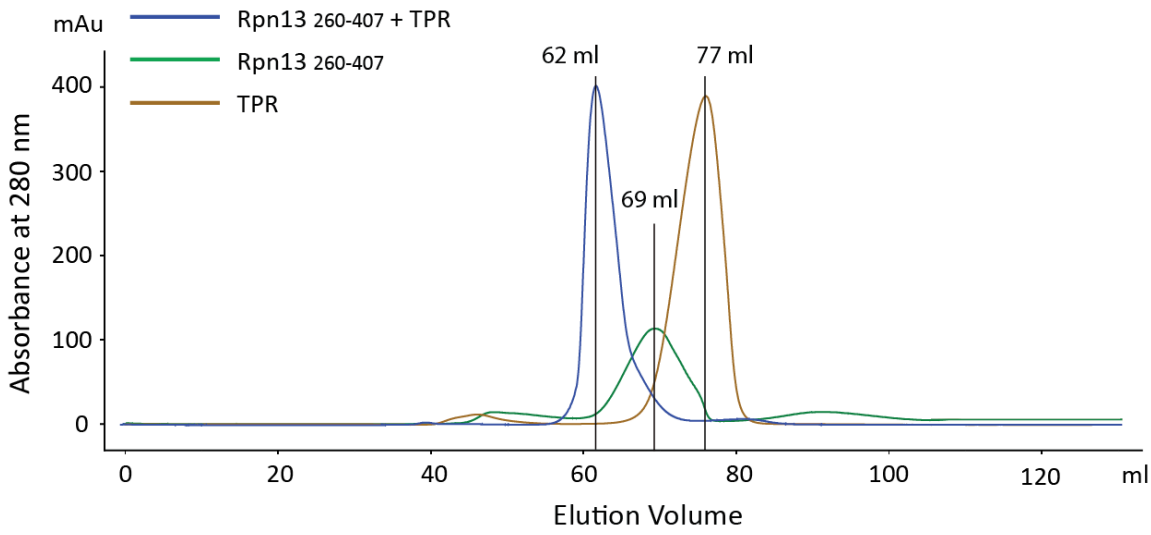



**Figure 5.4- Complex formation between $Rpn13_{260-407}$ and SGTA_TPR as observed by SEC.** SEC experiments carried out on a 120 ml HiLoad 16/60 superdex 75 column at a flow rate of 1 ml/min show the formation of a stable complex between $Rpn13_{260-407}$ and SGTA_TPR as evident by their co-elution at 62 ml when combined at a 1:1 molar ratio. SEC profiles of $Rpn13_{260-407}$ (elution volume: 69 ml) and SGTA_TPR (elution volume: 77 ml) loaded individually under identical conditions have been overlaid for comparison.

Once binding had been established by SEC, isothermal titration calorimetry (ITC) was used to obtain binding constants ($K_d$) and thermodynamic parameters underlying this interaction. ITC experiments yield a $K_d$ of 16.1 ± 1.4 μM with favourable enthalpy and entropy values (ΔH = -3.37 ± 0.08 kcal/mol; ΔS = 10.6 ± 0.8 cal/mol·K) (Fig. 5.5). This is suggestive of an interaction with SGTA_TPR driven by the formation of hydrogen bonds together with hydrophobic interactions.

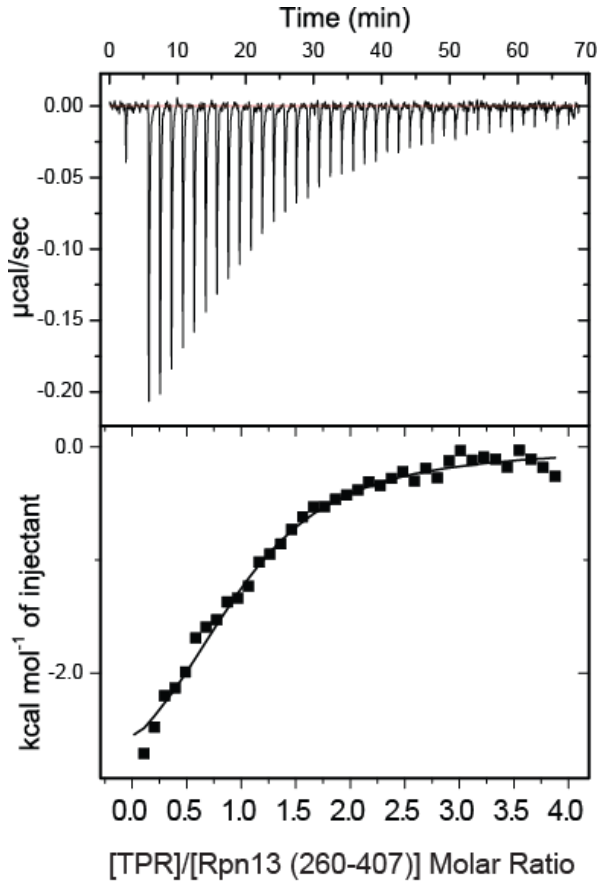



**Figure 5.5- Isothermal titration calorimetry (ITC) data showing binding of SGTA_TPR to Rpn13$_{260-407}$.** ITC experiments were carried out at 25°C, with 40 injections of 1 μL of SGTA_TPR at a concentration of 1 mM added to a sample of Rpn13$_{260-407}$ at a concentration of 50 μM in the reaction cell. The thermodynamic binding constant of this interaction as determined by ITC was $K_d$ = 16.1 ± 1.4 μM, with favourable enthalpy and entropy values (ΔH = -3.37 ± 0.08 kcal/mol; ΔS = 10.6 ± 0.8 cal/mol·K). Binding was assumed to be at one site (n = 1), to determine the binding affinity ($K_d$) and thermodynamic parameters.

## 5.3 Characterisation of the SGTA-Rpn13 interaction by solution NMR spectroscopy

Solution NMR spectroscopy was used to characterise the interaction between SGTA_TPR and $Rpn13_{260-407}$. Reciprocal chemical shift perturbation (CSP) experiments were carried out firstly by titrating unlabelled $Rpn13_{260-407}$ into $^{15}N$-labelled SGTA_TPR, and then by titrating unlabelled SGTA_TPR into $^{15}N$-labelled $Rpn13_{260-407}$. Assignments of backbone amide resonances of SGTA_TPR and $Rpn13_{260-407}$ were obtained from the biological magnetic resonance data bank (BMRB) [BMRB accession numbers 5709 (SGTA_TPR) and 17286 ($Rpn13_{260-407}$)].

The $^{1}H$-$^{15}N$ HSQC spectra of SGTA_TPR displayed binding in a fast exchange regime upon the addition of increasing amounts of unlabelled $Rpn13_{260-407}$ (Fig. 5.6). In this titration, widespread perturbations of backbone amide signals were observed, with certain amide signals not reaching saturation even in the presence of an eight-fold molar excess of unlabelled $Rpn13_{260-407}$. However, the same titration revealed SGTA_TPR backbone amide signals that appear to have reached saturation at lower molar excesses of $Rpn13_{260-407}$. As a result of these differences it was not possible to unambiguously disentangle signals that represented a binding equilibrium from this HSQC spectrum, and consequently this data could not be fit to a meaningful binding isotherm. Furthermore, when CSP analysis of this data (Fig. 5.7) was mapped onto the crystal structure of SGTA_TPR (PDB accession code 2VYI), this mapping could not provide any clear insights regarding the extent of its binding interface with $Rpn13_{260-407}$ (Fig. 5.8).

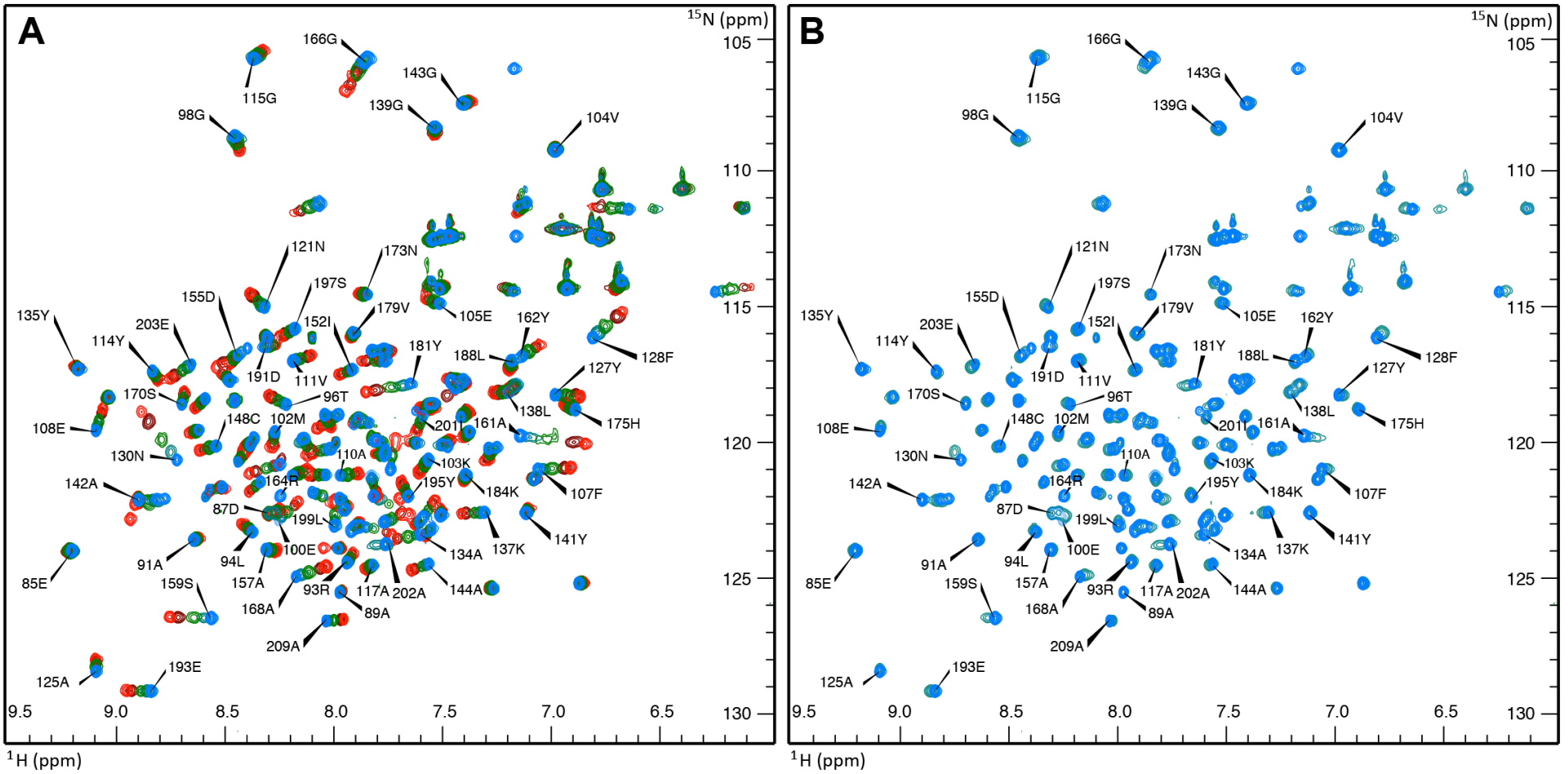



**Figure 5.6- $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled SGTA_TPR overlaid at different titration points with unlabelled Rpn13$_{260-407}$.** (A) For each titration point $^{1}$H-$^{15}$N HSQC experiments were recorded at 25°C on a 700 MHz spectrometer. The titrations were carried out at the following molar ratios of SGTA_TPR to Rpn13$_{260-407}$- 1:0, 1:1, 1:2, 1:3, and 1:8; coloured in blue, teal, green, maroon, and red respectively. $^{15}$N SGTA_TPR was present at 300 µM (for titration points 1:0, 1:1, and 1:2), at 200 µM (for titration point 1:3), and at 100 µM (for titration point 1:8). (B) Titrations points at 1:0 and 1:1 shown for comparison.

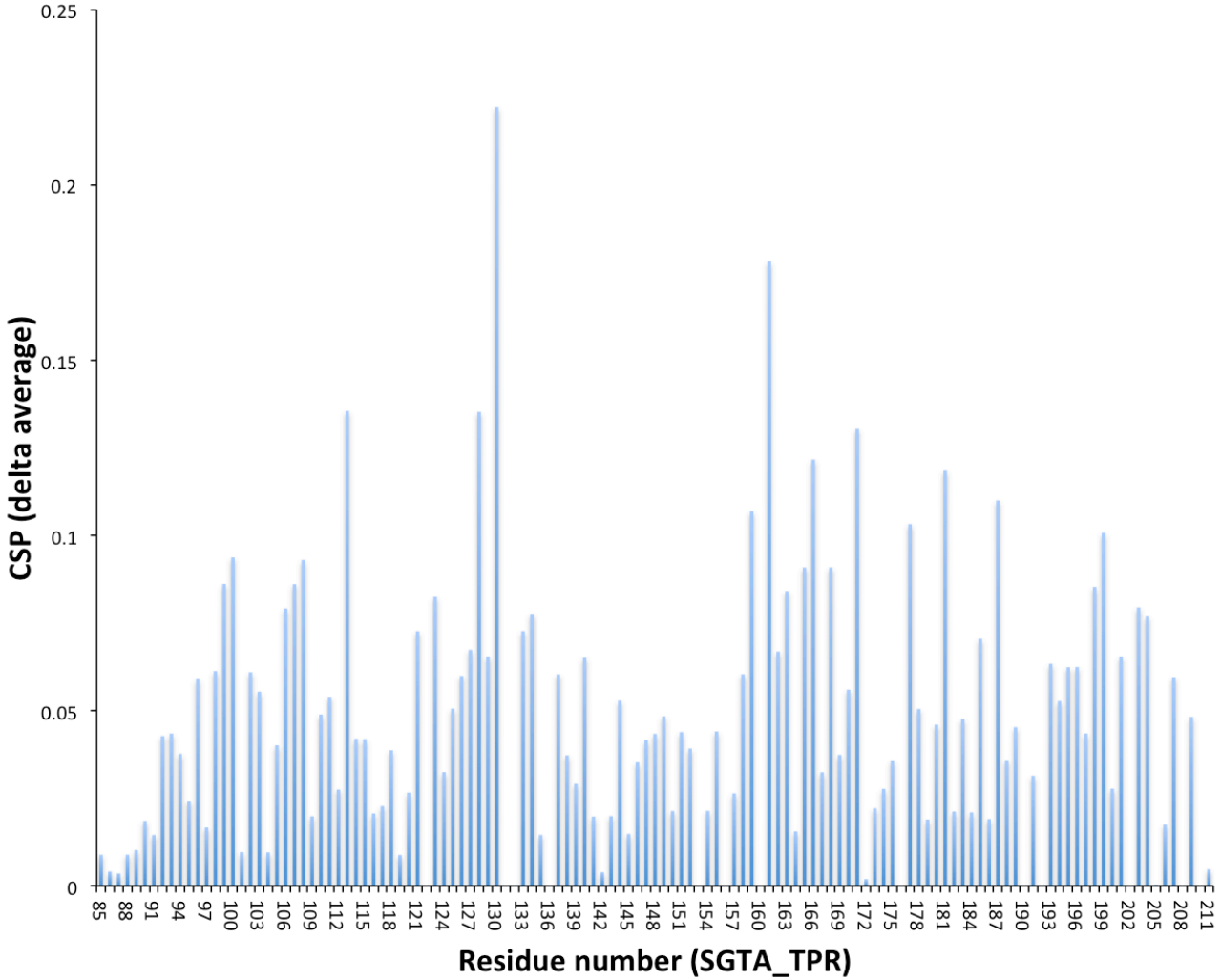



**Figure 5.7- Chemical shift perturbation (CSP) delta average values of SGTA_TPR amide signals upon binding to Rpn13$_{260-407}$.** CSP $\Delta\delta^{av}$ values were calculated for each amide signal.

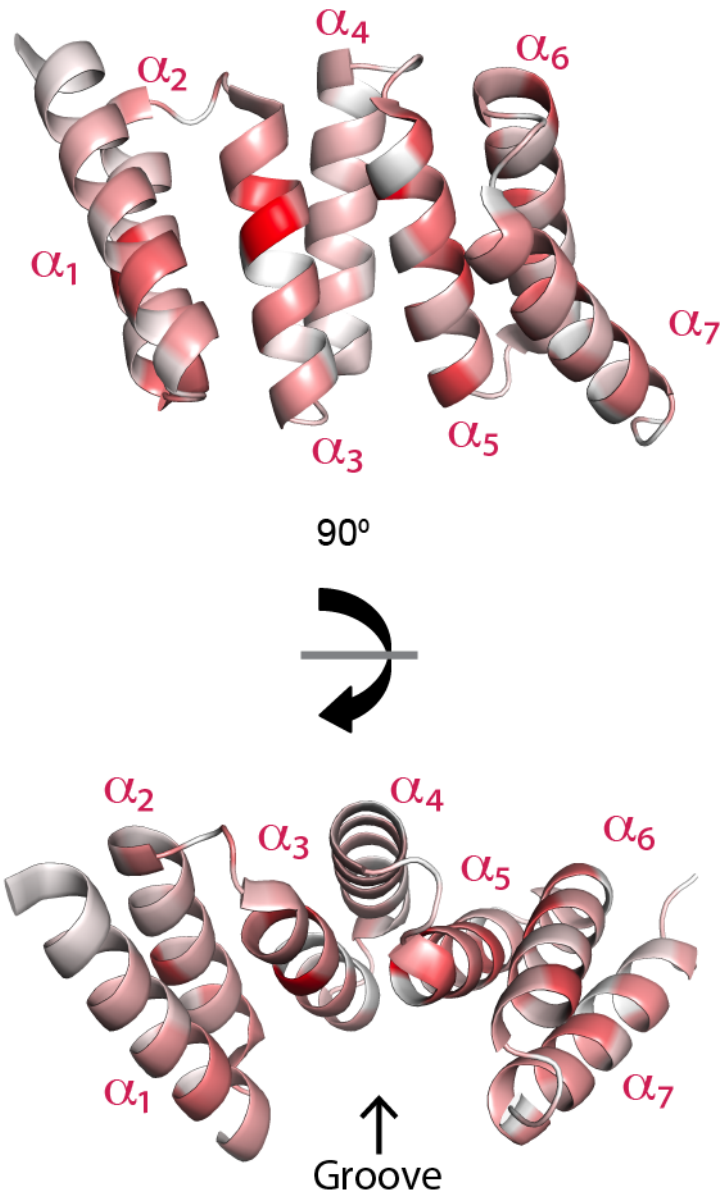



**Figure 5.8- Chemical shift perturbation (CSP) analysis of the SGTA_TPR/ Rpn13$_{260-407}$ interaction.** Normalised CSP $\Delta\delta^{av}$ values of SGTA_TPR backbone amide signals upon binding to Rpn13$_{260-407}$ were mapped on to the crystal structure of SGTA_TPR (PDB accession code 2VYI). Orthogonal cartoon views of SGTA_TPR are coloured according to the normalised CSP $\Delta\delta^{av}$ values with most perturbed residues shown in dark red. SGTA_TPR helices are labelled $\alpha_1$-$\alpha_7$ and the position of the central groove on its concave surface is indicated.

The $^{1}H$-$^{15}N$ HSQC spectra of $^{15}N$-labelled $Rpn13_{260-407}$ also displayed binding in a fast exchange regimen upon addition of unlabelled SGTA_TPR. However, in this titration, backbone amide peaks corresponding to only extreme C-terminal Rpn13 residues, namely Met 404, Ser 405, Leu 406 and Asp 407 were perturbed upon the addition of increasing amounts of SGTA_TPR (Figs. 5.9 and 5.10). This titration approached saturation at a 1:1 molar ratio with SGTA_TPR, with the CSP data (Fig. 5.10) yielding a $K_d$ of 3.27 ± 0.87 μM (Fig. 5.11). This was in agreement with the $K_d$ value obtained by ITC (Fig. 5.5).

In order to further understand the SGTA_TPR binding region of Rpn13, $^{1}H$- $^{15}N$ heteronuclear NOE experiments were carried out with a four second saturation transfer or control period. The extreme C-terminal region of free Rpn13 (residues 403 - 407) was found to be highly dynamic, with negative heteronuclear NOE enhancement values observed in this region of the amino acid sequence (Fig. 5.12; upper panel). Upon addition of SGTA_TPR at a 1:1 molar ratio, these values transitioned to positive ones indicative of the loss of high frequency motions as a result of binding (Fig. 5.12; lower panel). The average heteronuclear NOE enhancement value for Rpn13 residues Asp 403 to Asp 407 in the free and SGTA_TPR-bound state were -0.71 and 0.48, respectively, thus indicating the transition of the Rpn13 extreme C-terminal residues 403 - 407 to a more ordered state upon binding to SGTA_TPR (Fig. 5.12). Taken together with the CSP analysis from the $Rpn13_{260-407}$ perspective, these data clearly implicate five Rpn13 extreme C-terminal residues, namely Asp 403 to Asp 407, as the SGTA_TPR binding region of Rpn13.

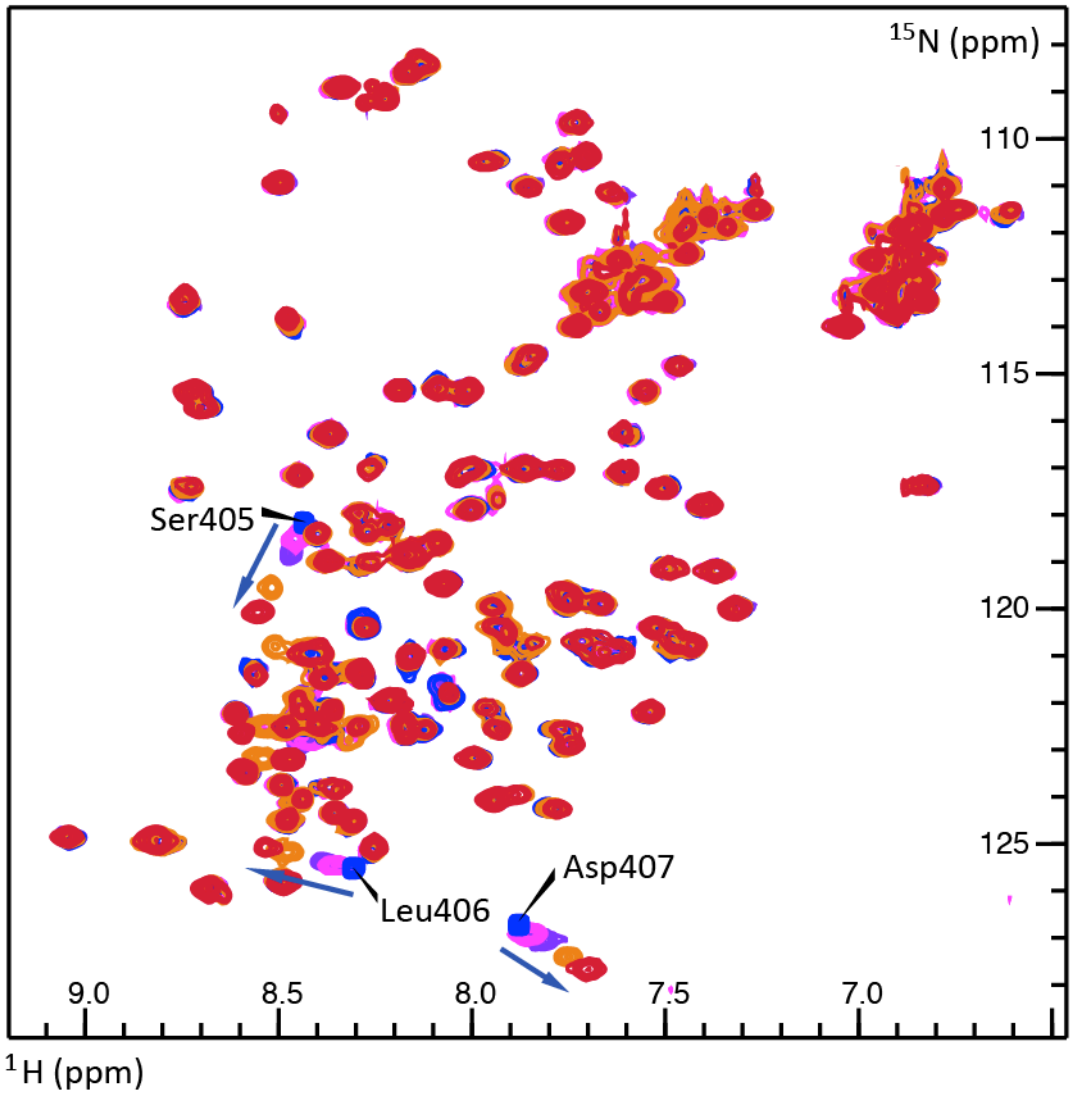



**Figure 5.9- $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled Rpn13$_{260-407}$ overlaid at different titration points with unlabelled SGTA_TPR.** For each titration point (1:0, 1:0.2, 1:0.5, 1:0.7, 1:1; shown in blue, light green, green, maroon, and red respectively) $^{1}$H-$^{15}$N HSQC experiments were recorded at 25°C. $^{15}$N Rpn13$_{260-407}$ was present at a concentration of 370 µM.

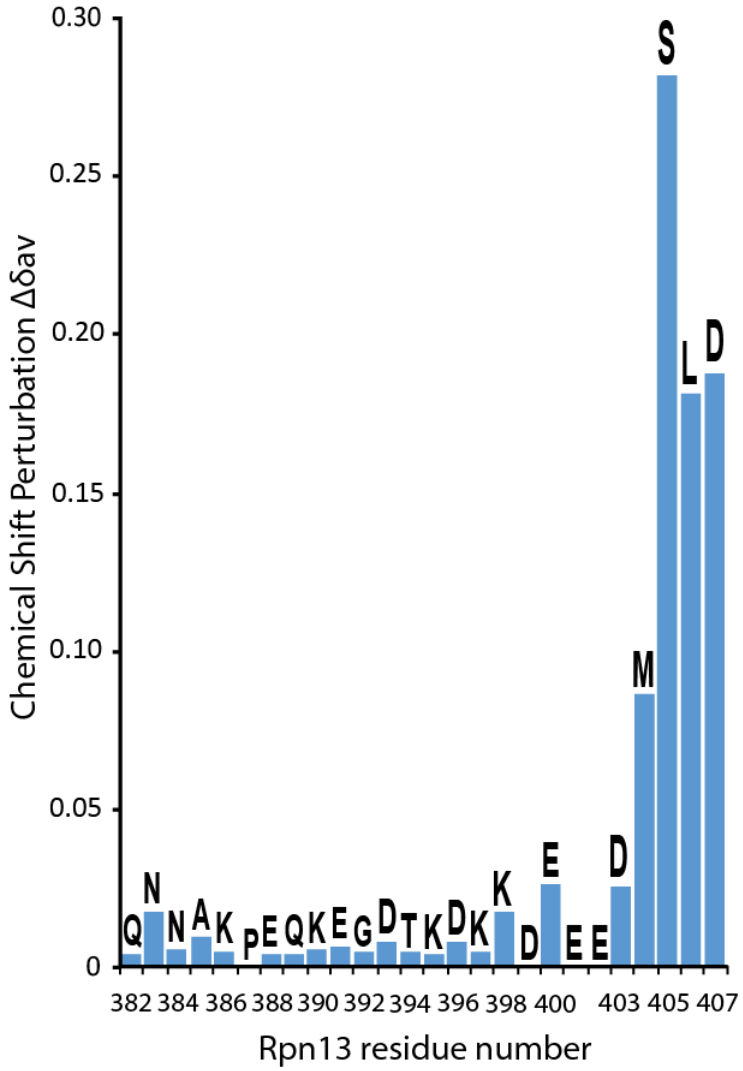



**Figure 5.10- Chemical shift perturbation (CSP) values of $Rpn13_{260-407}$ amide signals upon binding to SGTA_TPR.** CSP $\Delta\delta^{av}$ values were calculated for C-terminal backbone amide signals of $Rpn13_{260-407}$ as described in Chapter 2 (*Materials and Methods*).

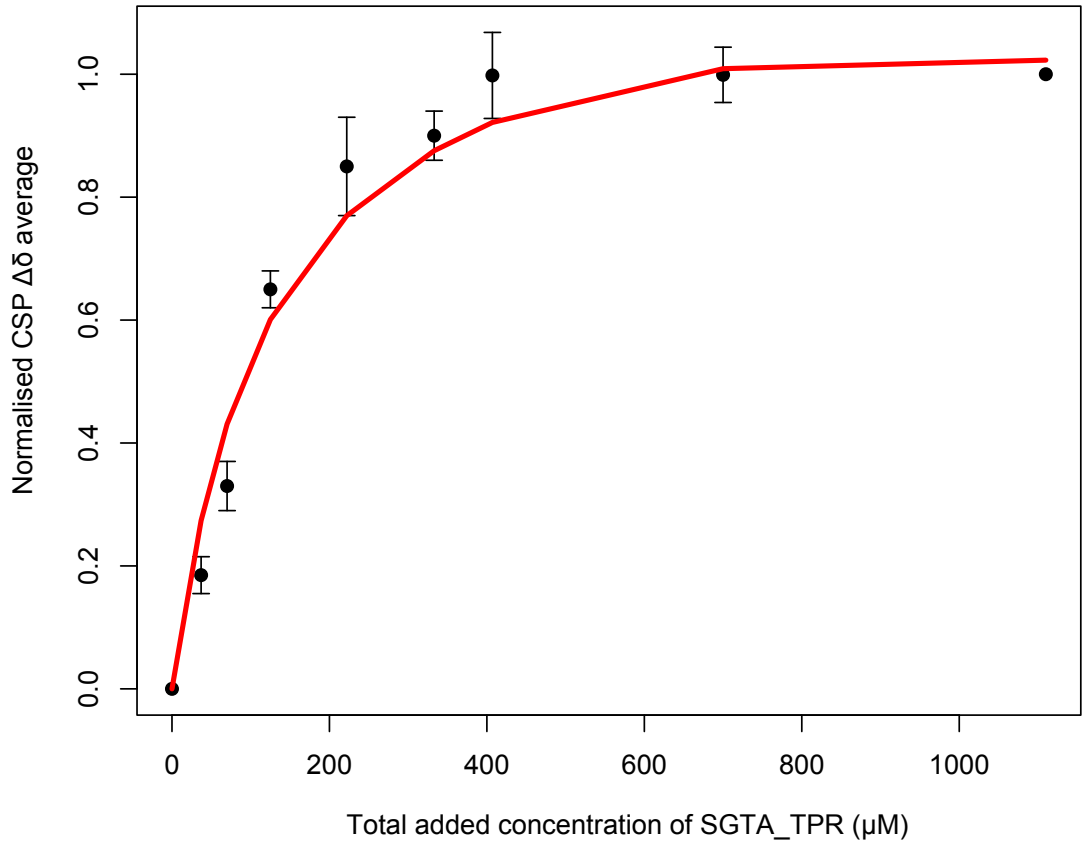



**Figure 5.11- Normalised CSP data of the most perturbed Rpn13$_{260-407}$ residues (405, 406, 407) upon titration with different concentrations of SGTA_TPR**. NMR titrations were carried out at different molar ratios (1:0, 1:0.1, 1:0.2, 1:0.3, 1:0.6, 1:0.9, 1:1.1, 1:1.9 and 1:3). Fitting was carried out using DynaFit (Kuzmic, 1996) to yield a $K_d$ of 3.27 ± 0.87 μM. $^{15}$N Rpn13$_{260-407}$ was present at a concentration of 370 μM.

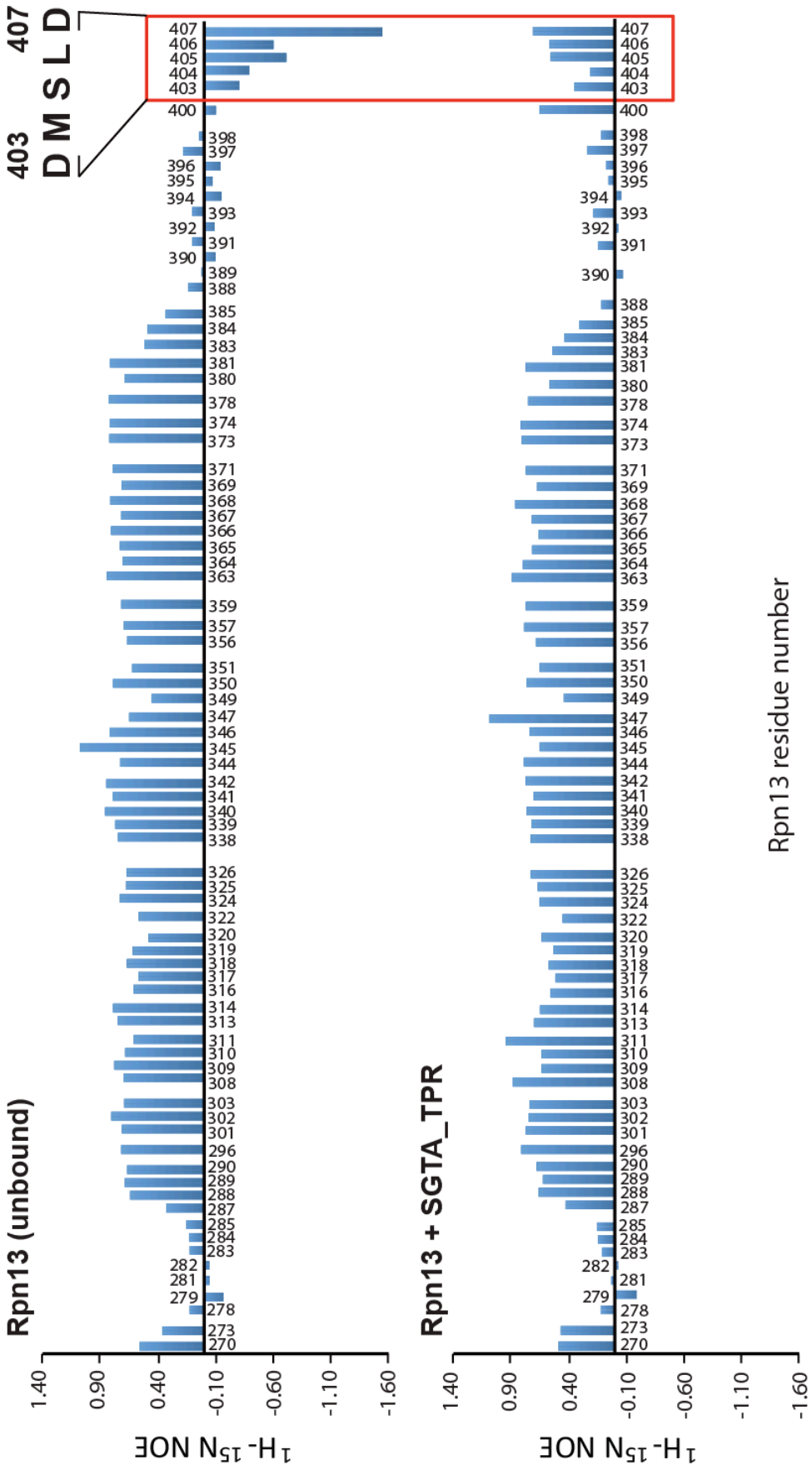



**Figure 5.12- $^{1}$H-$^{15}$N heteronuclear NOE enhancement values of Rpn13 C-terminal domain in its free, and SGTA_TPR bound, states.** The dynamic properties of free and SGTA_TPR-bound Rpn13$_{260\text{-}407}$ were assessed by $^{1}$H-$^{15}$N heteronuclear NOE experiments. The extreme C-terminal residues 403-407 of free Rpn13 were found to be highly dynamic, with negative heteronuclear NOE enhancement values. These values change significantly upon binding to SGTA_TPR.

## 5.4 The Rpn13 extreme C-terminal DMSLD pentapeptide is necessary and sufficient for its interaction with SGTA_TPR

Informed by CSP and $^{1}H$-$^{15}N$ heteronuclear NOE experiments, an Rpn13 C-terminal domain construct devoid of its last five residues was prepared to understand the contribution of this region to its binding with SGTA_TPR. This construct lacking the Rpn13 extreme C-terminal DMSLD pentapeptide, hereafter referred to as $Rpn13_{260-402}$, was expressed and purified (Fig. 5.13) in a manner identical to that described for $Rpn13_{260-407}$.

The $Rpn13_{260-402}$ construct was first analysed by SEC to understand its association with SGTA_TPR. SGTA_TPR and $Rpn13_{260-402}$ were mixed in a 1:1 molar ratio and analysed by SEC on a on a 120 ml HiLoad 16/60 superdex 75 column at a flow rate of 1 ml/min. The resulting elution volumes (69 ml for $Rpn13_{260-407}$; and 77 ml for SGTA_TPR) clearly demonstrated that $Rpn13_{260-402}$ was not capable of interacting with SGTA_TPR, in particular when compared to the elution profile of the SGTA_TPR-$Rpn13_{260-407}$ on the same column carried out under identical conditions (Fig. 5.14).

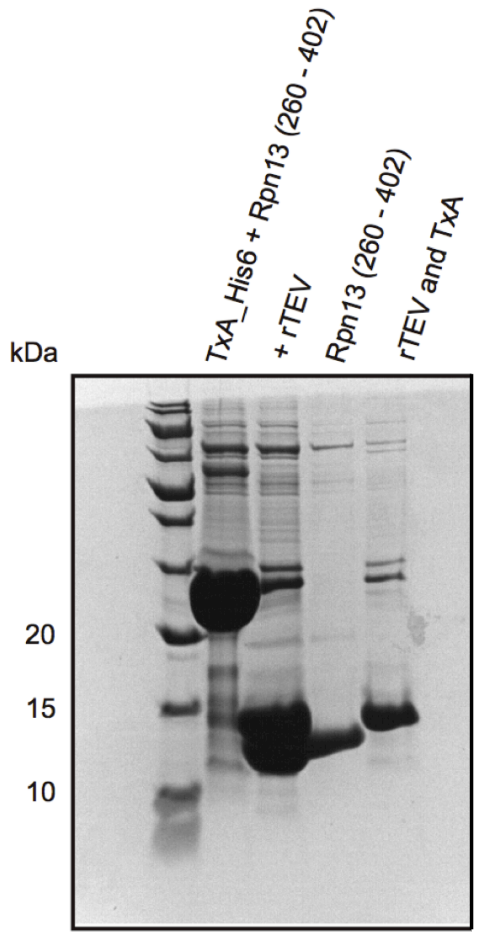



**Figure 5.13- Preparation of recombinant human Rpn13$_{260-402}$.** Rpn13$_{260-402}$ was overexpressed as a fusion protein, purified by nickel affinity chromatography and was recovered in the eluate of the first nickel column [SDS-PAGE fraction: TxA_His$_6$ + Rpn13 (260-402)]. This was followed by affinity tag and fusion protein removal by treatment with rTEV (SDS-PAGE fraction: +rTEV). Pure protein was obtained by passing the rTEV treated fraction through a second nickel column with pure protein obtained in the flow-through [SDS-PAGE fraction: Rpn13 (260-402)]. Separated thioredoxin A and rTEV was eluted from the second nickel column (SDS-PAGE fraction: rTEV and TxA).

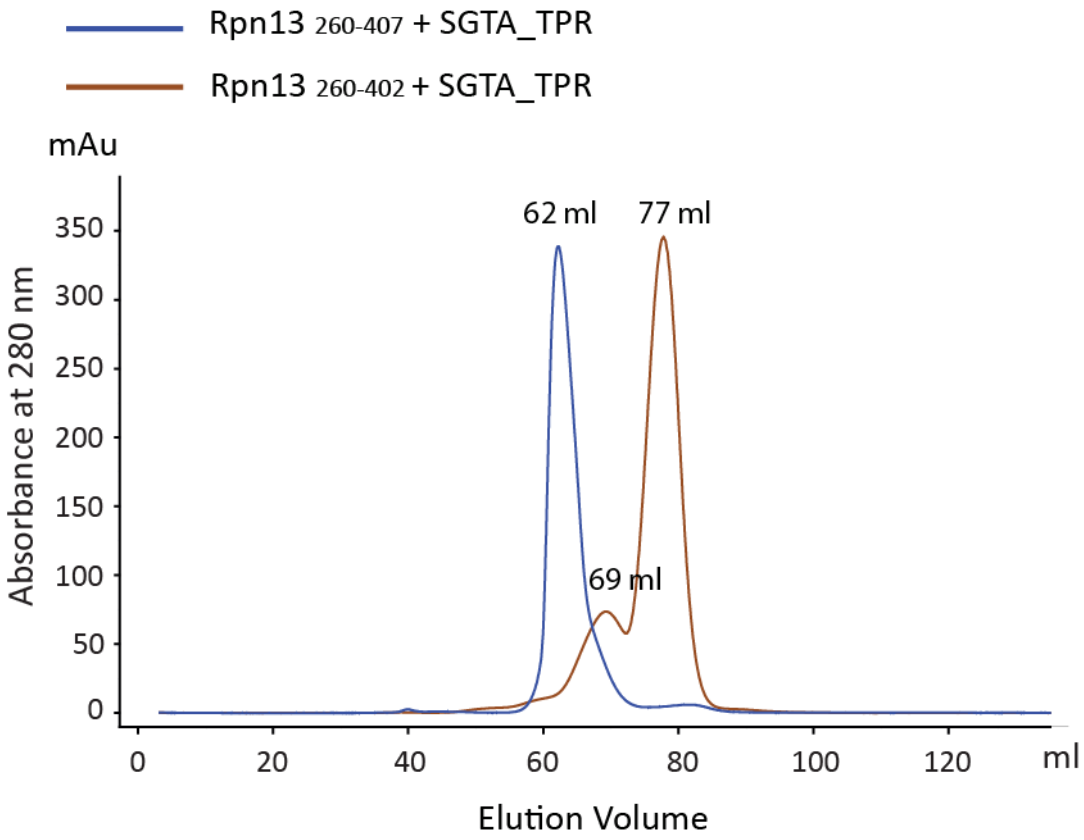



**Figure 5.14- SEC analysis shows abrogation of the SGTA_TPR-Rpn13 interaction in the absence of the Rpn13 extreme C-terminal DMSLD pentapeptide.** SEC experiments carried out on a 120 ml HiLoad 16/60 superdex 75 at a flow rate of 1ml/min show that the disruption of the $Rpn13_{260-407}$-SGTA_TPR complex occurs when the last five residues of Rpn13 (402-407) are absent. $Rpn13_{260-402}$ combined with SGTA_TPR shows no indication of complex formation by SEC. The SEC profile of the $Rpn13_{260-407}$-SGTA_TPR complex is overlaid for comparison.

Futhermore, solution NMR CSP experiments clearly showed no changes in amide chemical shifts in the $^{1}H$-$^{15}N$ HSQC spectra of $^{15}N$-labelled SGTA_TPR even in the presence of a six-fold molar excess of unlabelled $Rpn13_{260-402}$ (Fig. 5.15). In addition, ITC experiments carried out by injecting SGTA_TPR into $Rpn13_{260-402}$ confirmed abrogation of binding with SGTA_TPR (Fig. 5.16).

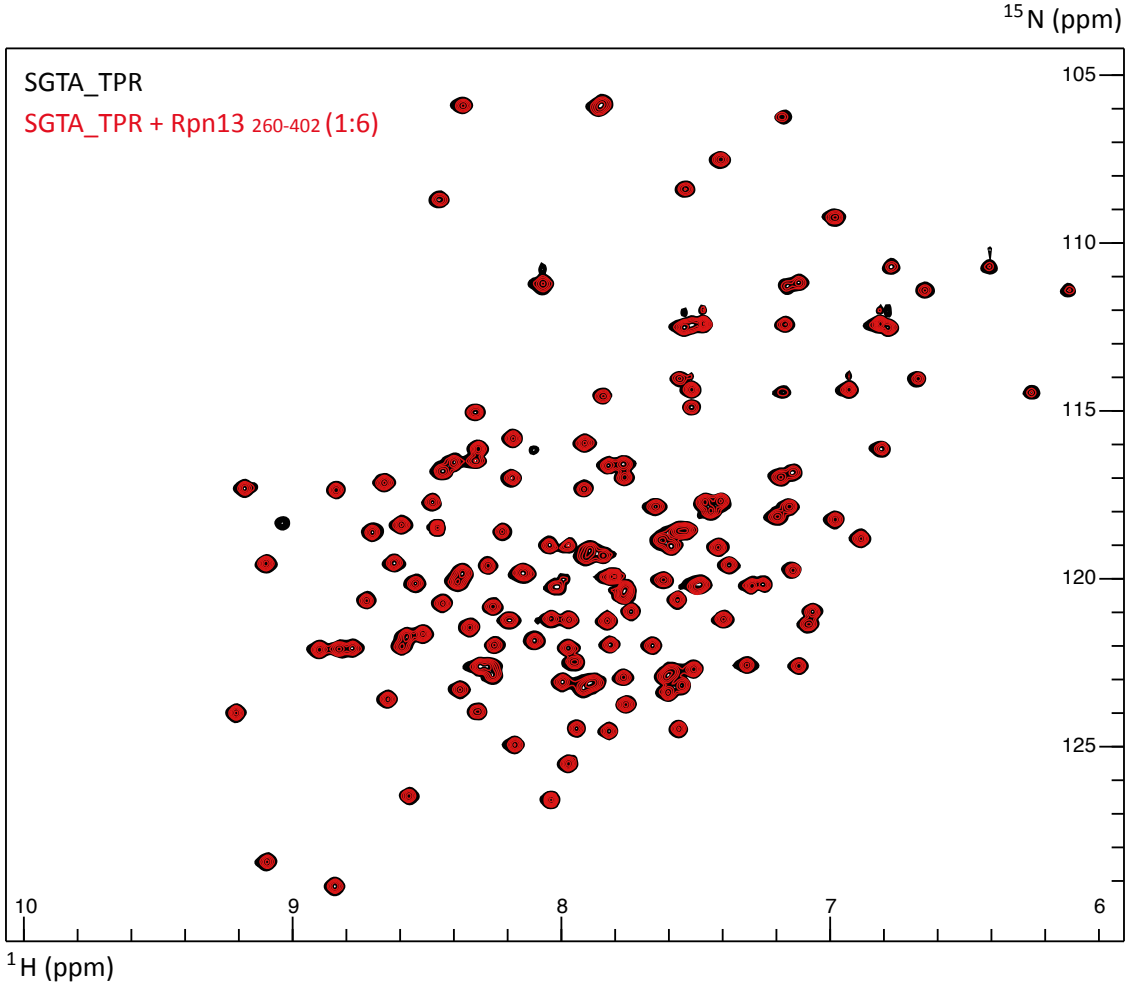



**Figure 5.15- Rpn13 devoid of its extreme C-terminal residues 403-407 cannot interact with SGTA_TPR.** $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled SGTA_TPR in its free form (in black), overlaid with the $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled SGTA_TPR with a six-fold molar excess of unlabelled Rpn13$_{260-402}$ (in red). The absence of the C-terminal pentapeptide of Rpn13 leads to no observed perturbations of SGTA_TPR amide signals. $^{15}$N SGTA_TPR was present at a concentration of 200 µM. Spectra were acquired on a 500 MHz spectrometer at 25°C.

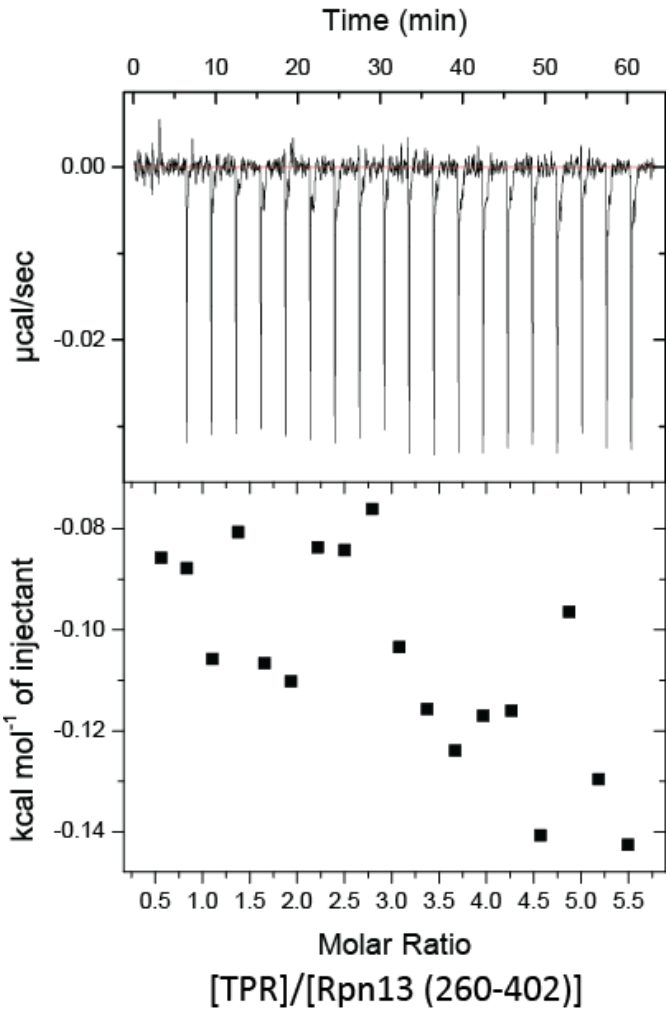



**Figure 5.16- Isothermal titration calorimetry (ITC) binding assay showing that $Rpn13_{260-402}$ does not interact with SGTA_TPR.** ITC experiments were carried out at 25°C, with 20 injections of 2 µL of SGTA_TPR at a concentration of 1 mM added to a sample of $Rpn13_{260-402}$ at a concentration of 50 µM in the calorimeter cell.

Based on insight gleaned from aforementioned experiments, an Rpn13 derived extreme C-terminal synthetic DMSLD pentapeptide was titrated into $^{15}N$-labelled SGTA_TPR. This titration clearly indicated an interaction between the DMSLD pentapeptide and SGTA_TPR with binding in a fast exchange regimen (Fig. 5.17). Additionally, ITC experiments yield a dissociation constant of 71.9 ± 13.3 µM for this protein-peptide interaction (Fig. 5.18), with favourable enthalpy and entropy values ($\Delta H$ = -4.97 ± 0.41 kcal/mol; $\Delta S$ = 2.34 ± 0.04 cal/mol·K).

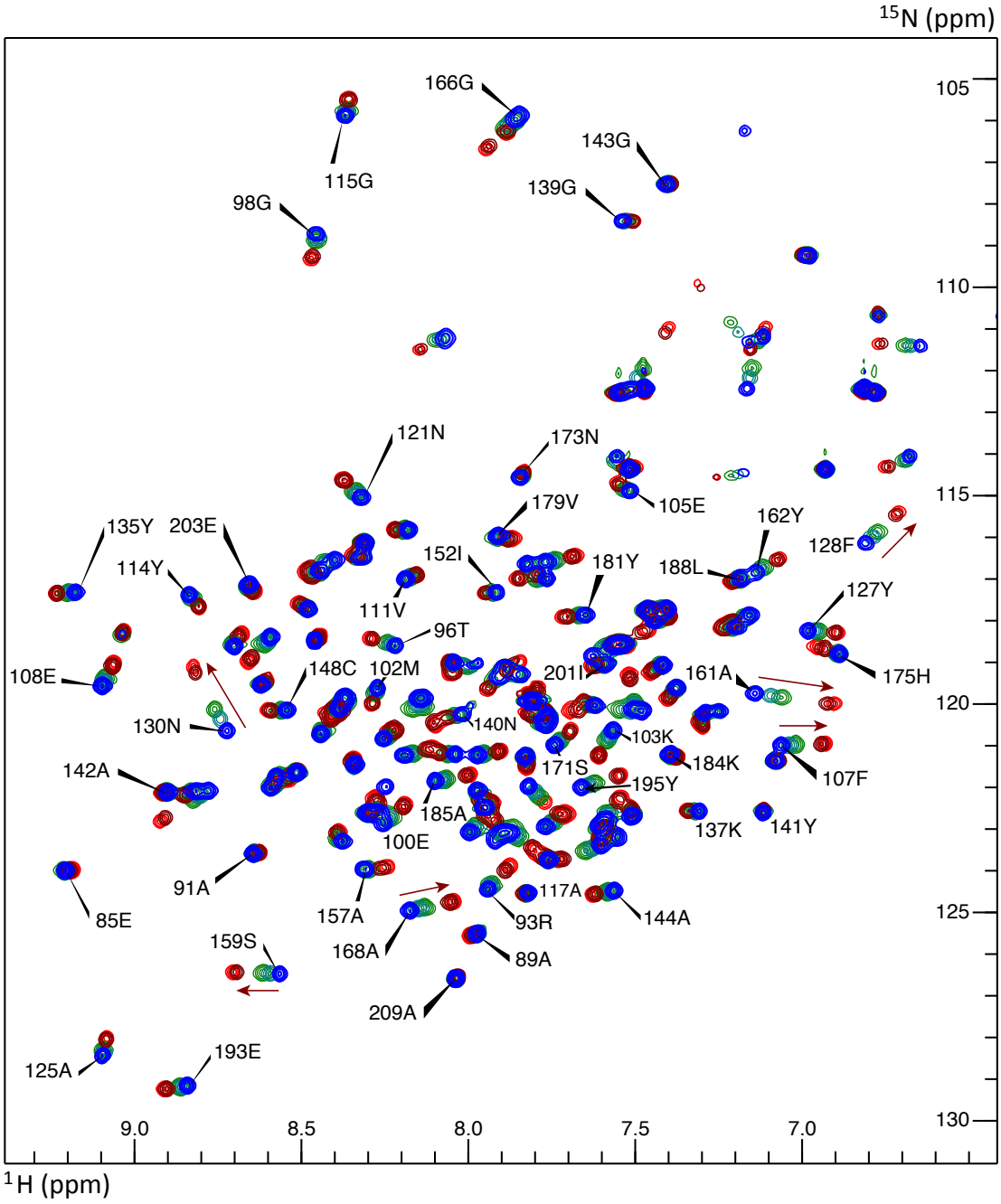



**Figure 5.17- The Rpn13 extreme C-terminal DMSLD pentapeptide is necessary and sufficient for its interaction with SGTA_TPR.** Overlaid $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled SGTA_TPR at different molar ratios with the unlabelled synthetic DMSLD pentapeptide (1:0, 1:0.5, 1:1, 1:6, 1:8; in blue, teal, green, maroon, and red respectively). $^{15}$N SGTA_TPR was present at 300 µM (for titration points 1:0, 1:0.5, and 1:1), and at 100 µM (for titration points 1:6 and 1:8). Spectra were acquired on a 500 MHz spectrometer at 25°C.

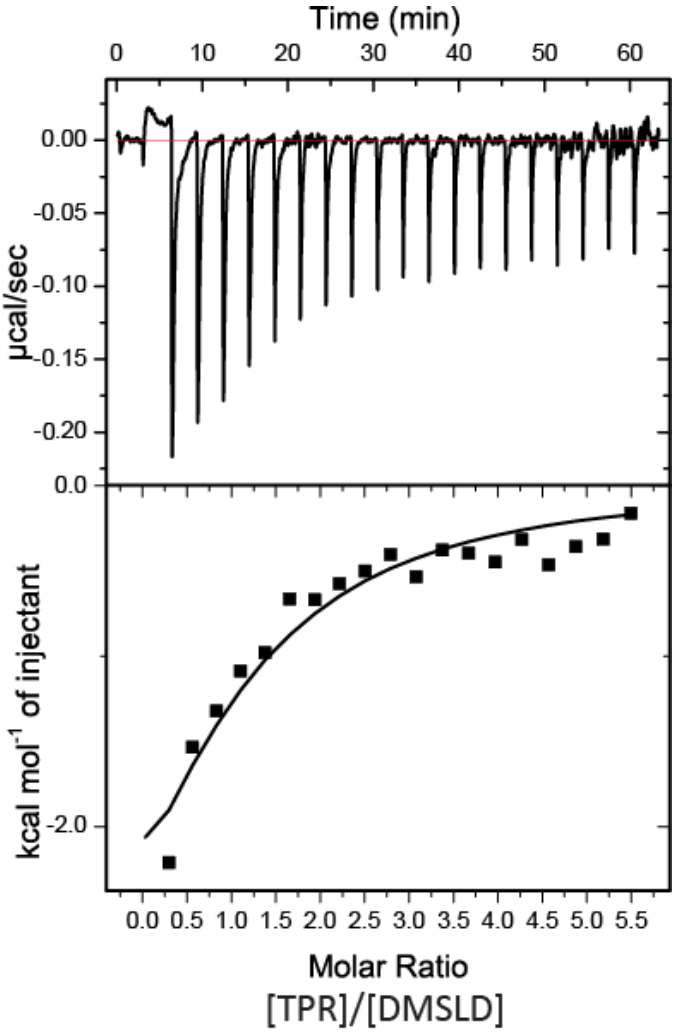



**Figure 5.18- ITC binding assay showing interaction of SGTA_TPR to the Rpn13 derived DMSLD pentapeptide.** The binding constant as determined by ITC for this interaction was, $K_d$ = 71.9 ± 13.3 µM, with favourable enthalpy and entropy values (ΔH = -4.97 ± 0.41 kcal/mol; ΔS = 2.34 ± 0.04 cal/mol·K). Binding was assumed to be at one site (n = 1), to determine the binding affinity ($K_d$) and thermodynamic parameters.

The DMSLD pentapeptide titrated into $^{15}$N-labelled SGTA_TPR resulted in a very similar pattern of CSPs of SGTA_TPR backbone amides amongst the most perturbed signals lining the central TPR groove when compared to $^{1}$H-$^{15}$N HSQC spectra of SGTA_TPR upon binding to $Rpn13_{260-407}$ at equivalent molar ratios (Fig. 5.19). In particular, SGTA_TPR residues Phe 107, Phe 128, Asn 130, Ser 159, Ala 161, and Ala 186 show similar perturbation of amide signals upon binding to both $Rpn13_{260-407}$ and to the DMSLD peptide, suggestive of a common binding modality. Furthermore, it was interesting to observe that amide signals corresponding to residues 203-209 of helix-7 (α7) at the C-terminus of SGTA_TPR were not affected by binding to the DMSLD peptide. This is mainly because these amide signals are perturbed upon binding to $Rpn13_{260-407}$ (Fig. 5.19). Therefore, it appears that C-terminal residues of SGTA_TPR helix-7 (α7) that do not structurally contribute to the central TPR groove, in turn do not form part of the peptide-binding site. This strongly points towards the central SGTA_TPR groove as the Rpn13 binding site, an interaction mediated by the Rpn13 DMSLD pentapeptide. Taken together, these experiments confirm that the extreme C-terminal DMSLD pentapeptide of Rpn13 is necessary and sufficient to facilitate its interaction with SGTA_TPR.

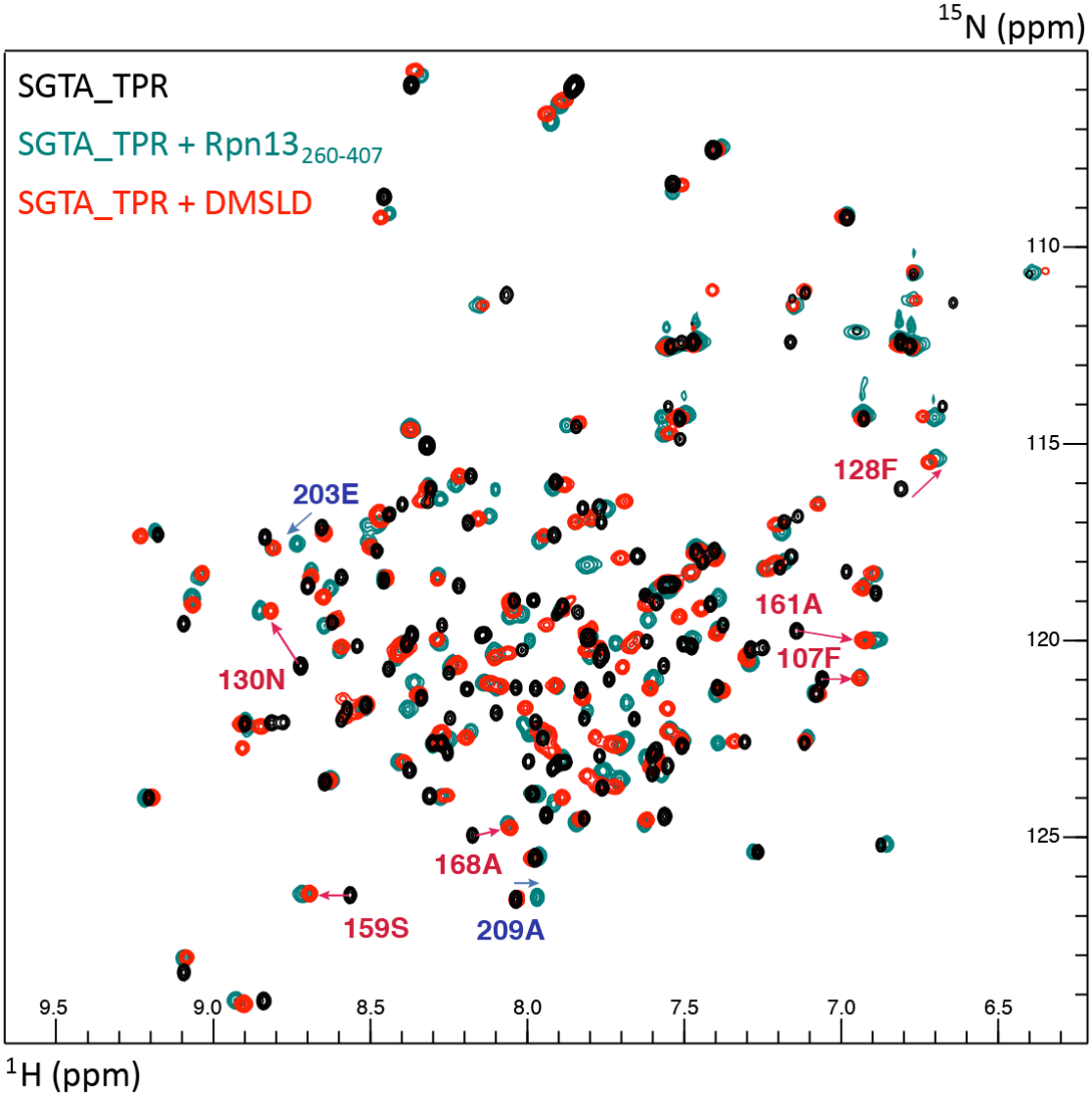



**Figure 5.19- Comparison of Rpn13$_{260-407}$ vs the DMSLD pentapeptide binding to SGTA_TPR.** Overlay of $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled SGTA_TPR in its free state (black), bound to a 3-fold molar excess of unlabelled Rpn13$_{260-407}$ (green), and bound to a 3-fold molar excess of the DMSLD pentapeptide (red). Backbone amide signals corresponding to residues with highest CSP $\Delta\delta^{av}$ values upon binding to either Rpn13$_{260-407}$ or to the DMSLD pentapeptide are indicated with red arrows. Signals corresponding to backbone amides only affected upon binding to Rpn13$_{260-407}$ but not perturbed by the SGTA_TPR/ DMSLD interaction have been indicated with blue arrows.

## 5.5 The SGTA_TPR-Rpn13 interaction occurs via a two-carboxylate clamp mechanism

TPR domains are well-established mediators of protein-protein interactions (Zhang *et al.*, 2005; D'Andrea and Regan, 2003; Scheufler *et al.*, 2000; Das *et al.*, 1998). Their interaction with the C-terminal IEEVD and MEEVD motifs of Hsp70/Hsp90 chaperones is known to occur via a 'two-carboxylate clamp' mode of molecular recognition (Dutta and Tan, 2008; Cliff *et al.*, 2006; Zhang *et al.*, 2005; Scheufler *et al.*, 2000). In this mode of binding, the main chain and side chain carboxylates present on the terminal aspartate of the IEEVD/MEEVD motifs are known to interact with conserved basic residues in the TPR groove. Given the Rpn13 sequence terminates with an aspartate, and the C-terminal DMSLD peptide has been found to be necessary and sufficient for this interaction, the possibility of a two-carboxylate clamp mechanism was explored by site-directed mutagenesis experiments. Structure based alignments of SGTA_TPR with HOP TPR domain was used to identify conserved residues within the TPR groove likely to participate in the formation of a two-carboxylate clamp with interacting peptides. Based on these alignments, residues Lys 160 and Arg 164 of SGTA_TPR were identified as conserved residues within the central TPR groove that could mediate a two-carboxylate clamp mode of binding (Fig. 5.20A). Thus, a K160E/R164E double mutant version of SGTA_TPR (Fig. 5.20B), was used to understand the effect it had on binding to $Rpn13_{260-407}$ by SEC and NMR. The K160E/R164E double mutant SGTA_TPR was successfully expressed and purified as shown in Fig 5.21A. Subsequently, SEC experiments in which the K160E/R164E double mutant SGTA_TPR was mixed with $Rpn13_{260-407}$ showed no evidence of complex formation even when $Rpn13_{260-407}$ was present at a 3-fold molar excess over SGTA_TPR (Fig. 5.21B). Also, reciprocal NMR titrations adding unlabelled $Rpn13_{260-407}$ into $^{15}$N-labelled K160E/R164E double mutant SGTA_TPR with up to a 6-fold

molar excess of $Rpn13_{260-407}$ and vice versa also showed no indication of binding (Figs. 5.22 and 5.23). These experiments therefore confirm that the K160E/R164E double mutant SGTA_TPR is sufficient to disrupt complex formation between SGTA_TPR and $Rpn13_{260-407}$. Consequently, this establishes a two-carboxylate clamp mechanism of binding between the terminal aspartate of Rpn13 and the conserved Lys 160 and Arg 164 residues present in the central grove of SGTA_TPR.

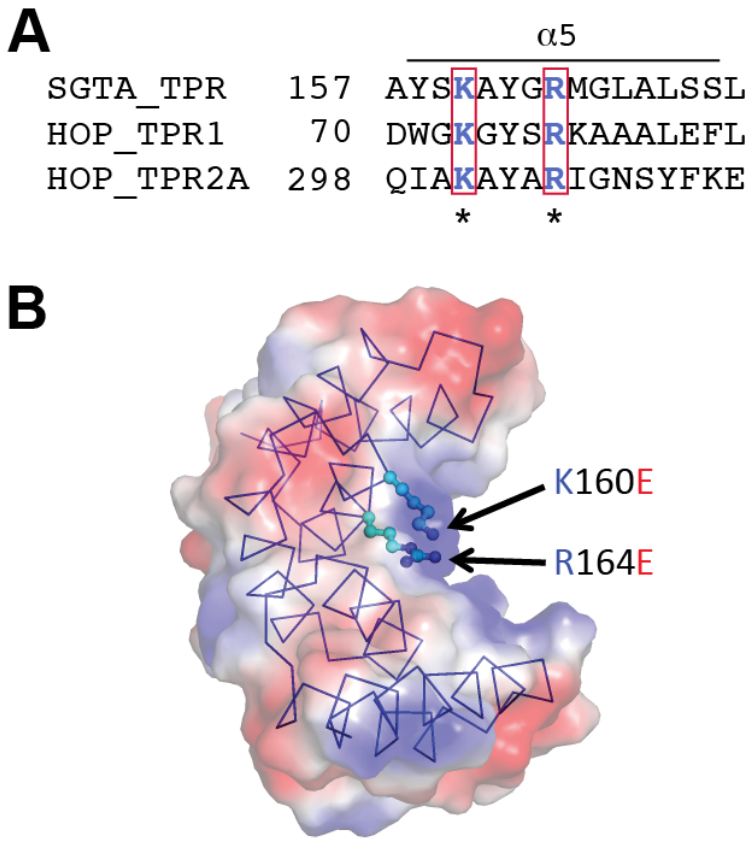



**Figure 5.20- The conserved two-carboxlate clamp mode of molecular recognition by TPR domains.** (A) Sequence alignment comparing helix-5 (α5) of SGTA_TPR to equivalent helices of HOP TPR domains, highlighted with conserved lysine and arginine residues that clamp terminal carboxylates present on binding partners. (B) Conserved residues present in the central groove of SGTA_TPR most likely to facilitate the two-carboxylate clamp mediated recognition of peptides are highlighted in ball and stick representation based on the sequence alignment. Based on crystal structure of SGTA_TPR (PDB accession code: 2VYI).

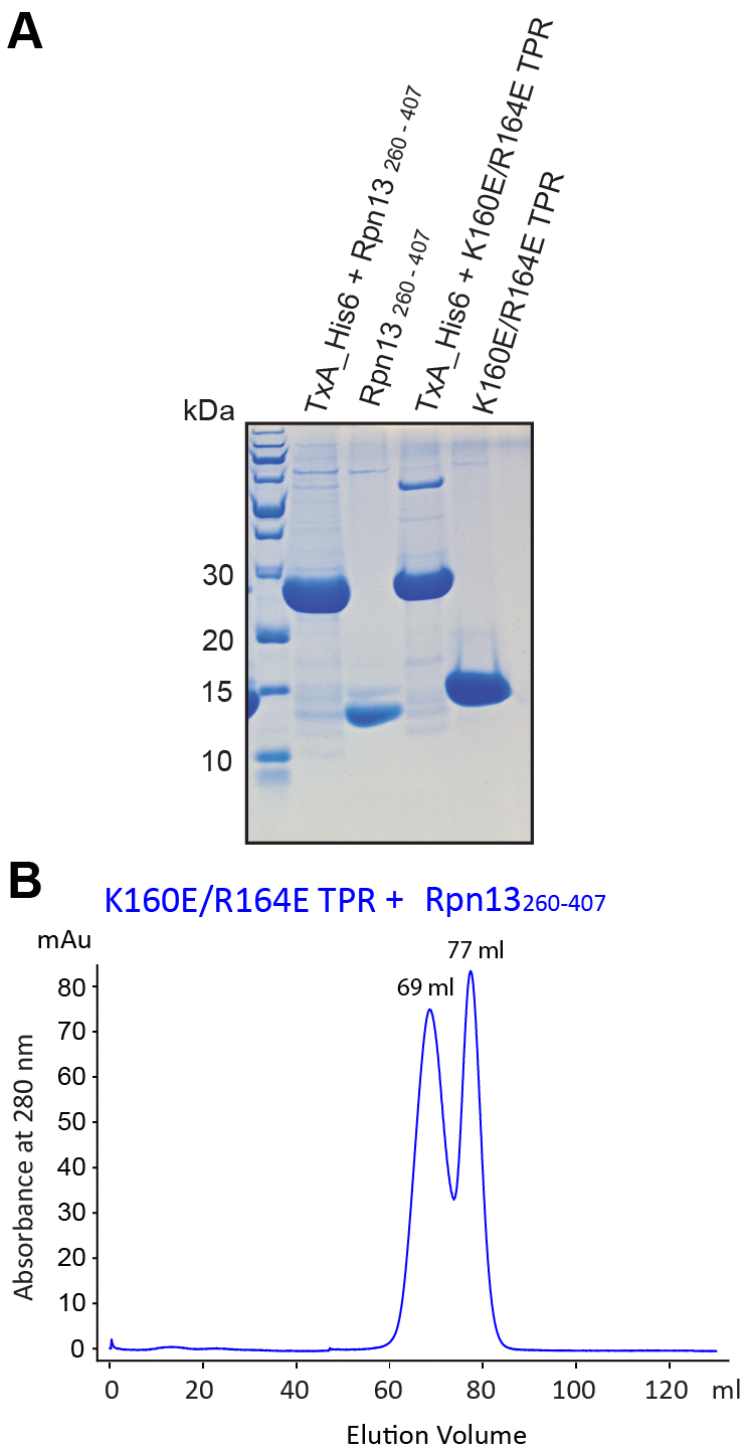



**Figure 5.21- Preparation of the K160E/R164E double mutant SGTA_TPR and $Rpn13_{260-407}$ to probe the two-carboxylate clamp mechanism of binding.** (A) SDS-PAGE showing purification results of $Rpn13_{260-407}$ (lanes: $TxA_His_6$ + $Rpn13_{260-407}$ and $Rpn13_{260-407}$) and of K160E/R164E double mutant SGTA_TPR (lanes: $TxA_His_6$ + K160E/R164E TPR and K160E/R164E TPR). (B) SEC experiments carried out on a 120 ml HiLoad 16/60 superdex 75 at a flow rate of 1ml/min demonstrate that the K160E/R164E double mutant SGTA_TPR does not co-elute with $Rpn13_{260-407}$ when mixed, even with $Rpn13_{260-407}$ present at a 3-fold molar excess over SGTA_TPR.

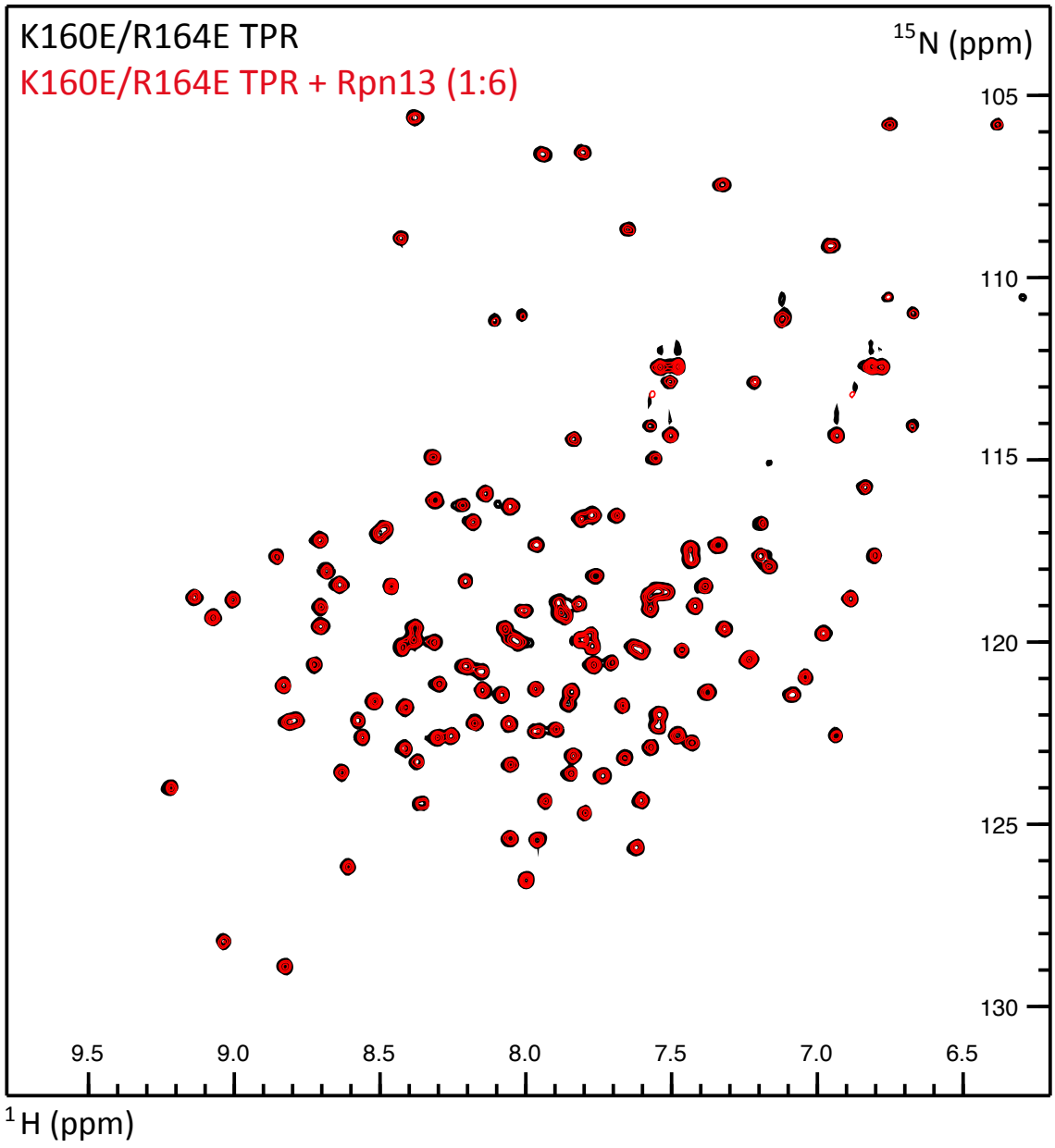



**Figure 5.22- The K160E/R164E double mutant SGTA_TPR disrupts its interaction with Rpn13$_{260-407}$.** Overlay of $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled K160E/R164E SGTA_TPR in its free form (black), and in the presence of a six-fold molar excess of unlabelled Rpn13$_{260-407}$ (red). Spectra were acquired on a 500 MHz spectrometer at 25°C. $^{15}$N K160E/R164E SGTA_TPR was present at a concentration of 200 μM.

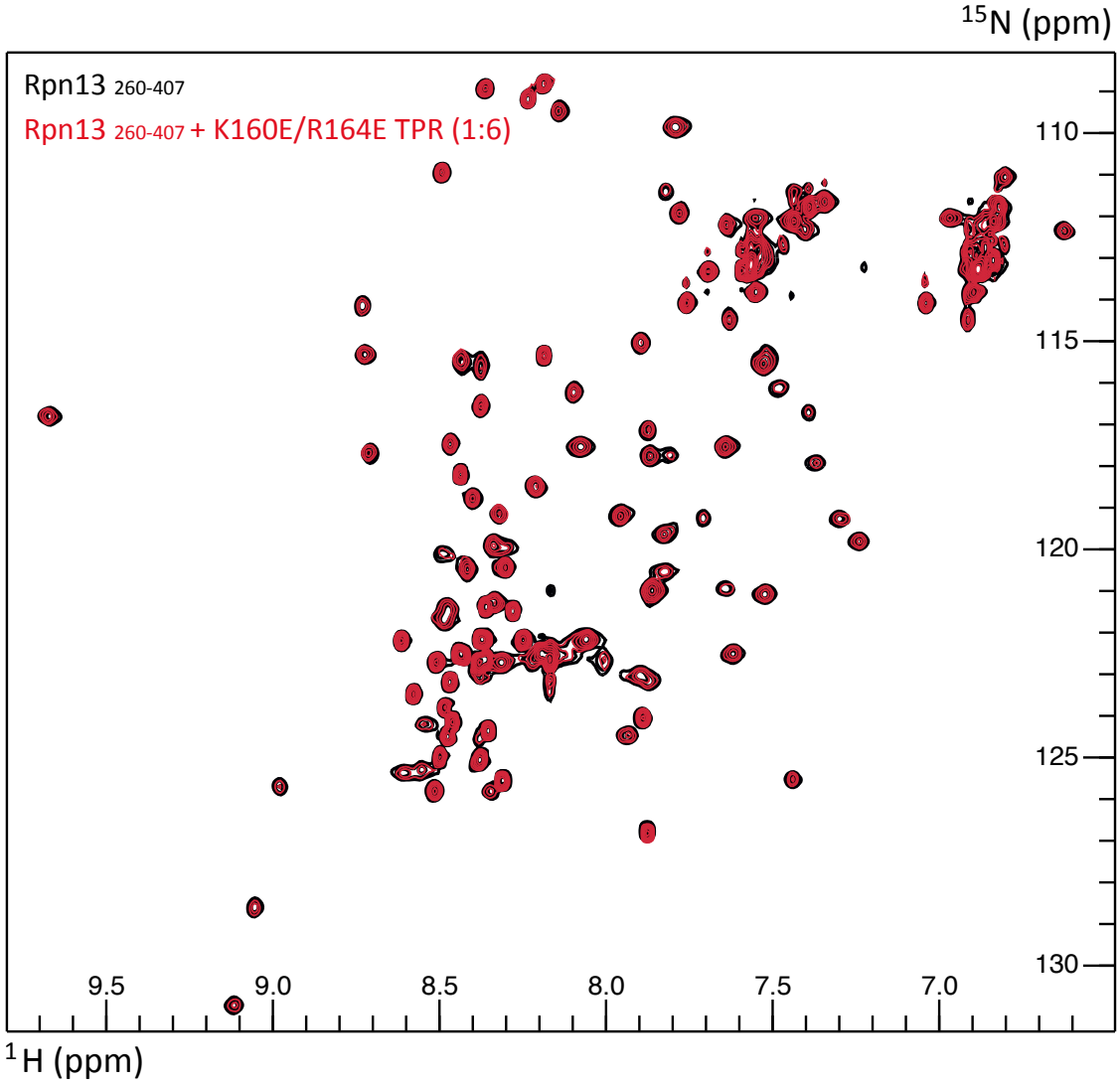



**Figure 5.23- $Rpn13_{260-407}$ and K160E/R164E double mutant SGTA_TPR do not interact.** Overlay of $^{1}H$-$^{15}N$ HSQC spectra of $^{15}N$-labelled $Rpn13_{260-407}$ in its free form (black), and with a six-fold molar excess of unlabelled K160E/R164E SGTA_TPR (red). Spectra were acquired on a 500 MHz spectrometer at 25°C. $^{15}N$ $Rpn13_{260-407}$ was present at a concentration of 200 μM.

## 5.6 Intermolecular PRE experiments to delineate the SGTA-Rpn13 binding interface

Once the two-carboxylate clamp mechanism of binding had been established as described above, intermolecular paramagnetic relaxation enhancement (PRE) experiments were carried out to further understand details pertaining to the SGTA_TPR/Rpn13 binding interface. These were to include details concerning the region of entry of the Rpn13′s C-terminal extension into the central groove of SGTA_TPR, the orientation of the DMSLD pentapeptide within the SGTA_TPR groove, and possible conformations of the pentapeptide within the groove. For these investigations, a site-directed paramagnetic spin-label (MTSL) was introduced via a K398C mutant of $Rpn13_{260-407}$. The position of the K398C mutant of $Rpn13_{260-407}$ is shown in Fig. 5.24A. $^{15}$N-labelled SGTA_TPR and K398C $Rpn13_{260-407}$ were purified in a similar manner as described previously (Fig. 5.24B). Proximity dependent line broadening of $^{15}$N-labelled SGTA_TPR backbone amide signals was monitored for intermolecular PRE effects upon addition of MTSL labelled K398C $Rpn13_{260-407}$ (Fig. 5.25A). Significant line broadening effects on SGTA_TPR backbone amides were observed upon binding to MTSL labelled K398C $Rpn13_{260-407}$ (Fig. 5.25). These effects could be clearly observed in SGTA_TPR backbone amide peaks corresponding to residues Met-102, Lys-103, Glu-105, Lys-137, Ser-170, Asn-173, Ser-197, Lys-200, Ile-201, Glu-203, Ser-211. When PRE intensity ratios (Fig. 5.25B) were mapped onto the crystal structure of SGTA_TPR (PDB code: 2VYI) these residues were all found to be present on the lower edge of its concave surface (Fig. 5.26), lining the entry to the central groove. This finding revealed that the paramagnetic probe on K398C $Rpn13_{260-407}$ was present within a 25 Å distance of these residues. Taken together with CSP data (Fig. 5.19), this confirms that the $Rpn13_{260-407}$ C-terminal extension enters the SGTA_TPR groove via this lower edge of its concave surface. Furthermore, the

position of the paramagnetic spin label confirms the orientation of the DMSLD peptide within the SGTA_TPR groove. In addition, mapping of SGTA_TPR residues affected by intermolecular PRE dependent line broadening onto the crystal structure of SGTA_TPR (Fig. 5.26) reveals that the Rpn13 C-terminal DMSLD pentapeptide could be sampling a lot of conformational space within the SGTA_TPR groove. This has lead to the speculation that the DMSLD peptide may be present in more than one conformation whilst bound to SGTA_TPR.

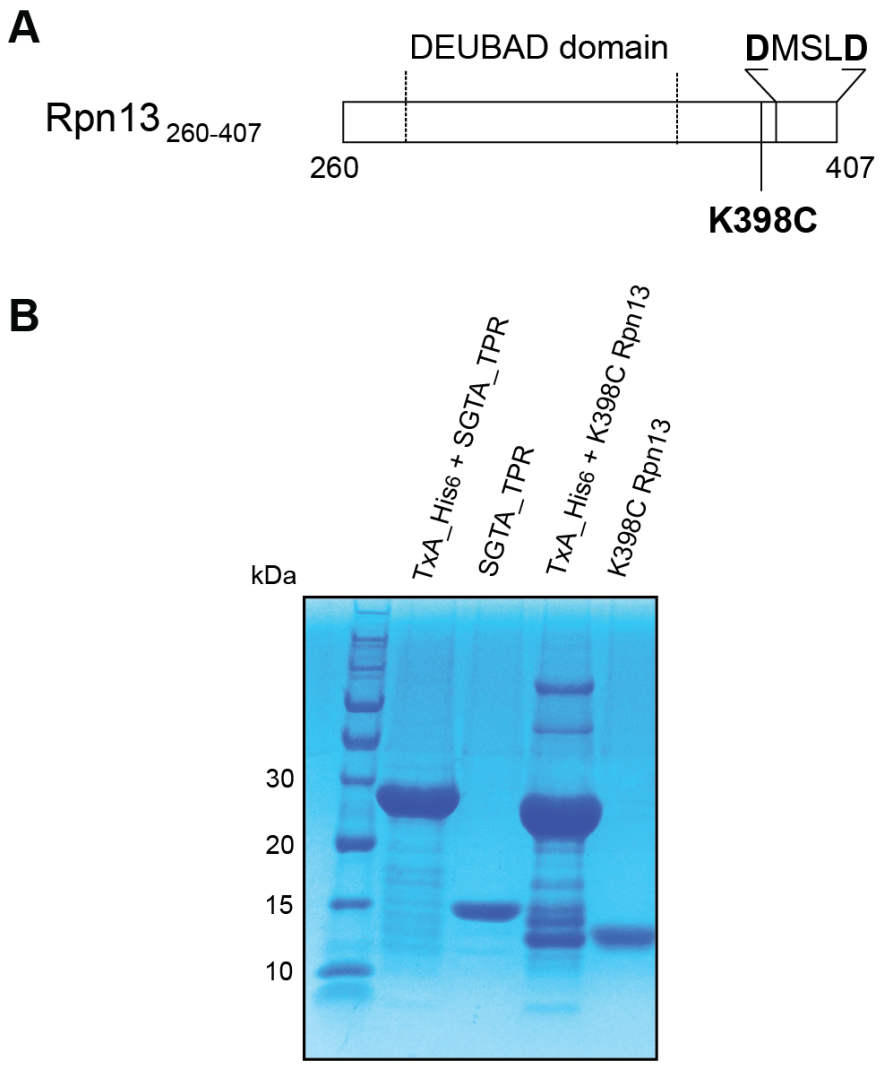



**Figure 5.24- Preparation of an Rpn13 C-terminal cysteine mutant for site-directed spin labelling studies.** (A) Schematic of the Rpn13 C-terminal domain with the position of the K398C used for site-directed paramagnetic spin labelling studies. (B) SDS-PAGE results showing preparation of $^{15}$N SGTA_TPR and unlabelled K398C $Rpn13_{260-407}$ used for intermolecular PRE experiments.

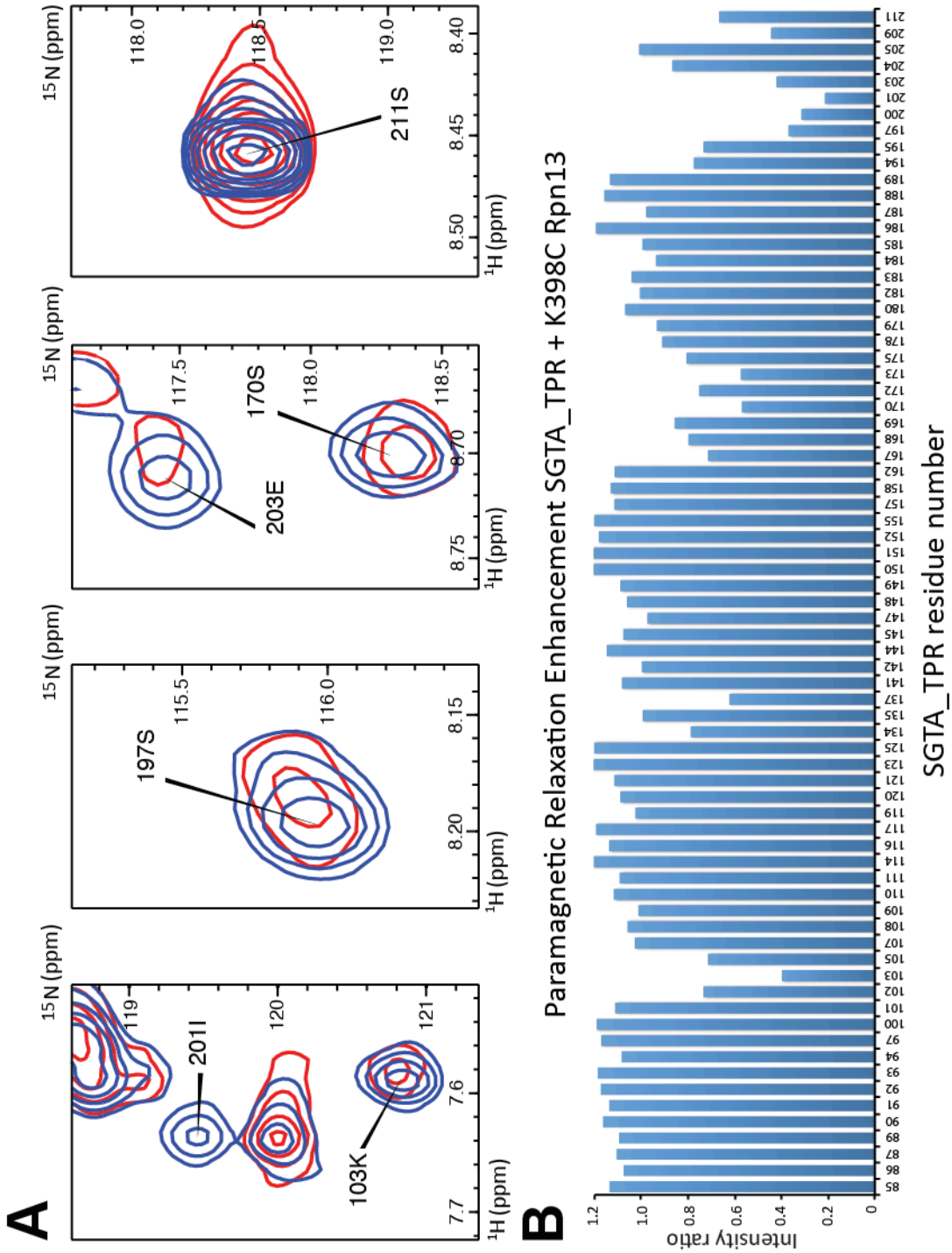



**Figure 5.25- Intermolecular paramagnetic relaxation enhancement (PRE) experiments.** (A) Regions of the $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N-labelled SGTA_TPR upon binding to either the diamagnetic (dMTSL) probe labelled K398C Rpn13$_{260-407}$ (blue), or in the presence of the paramagnetic (MTSL) probe labelled K398C Rpn13$_{260-407}$ (red) showing distance dependent line broadening. (B) Paramagnetic relaxation enhancement (PRE) intensity ratios of SGTA_TPR residues upon binding to the spin labelled K398C Rpn13$_{260-407}$ mutant.

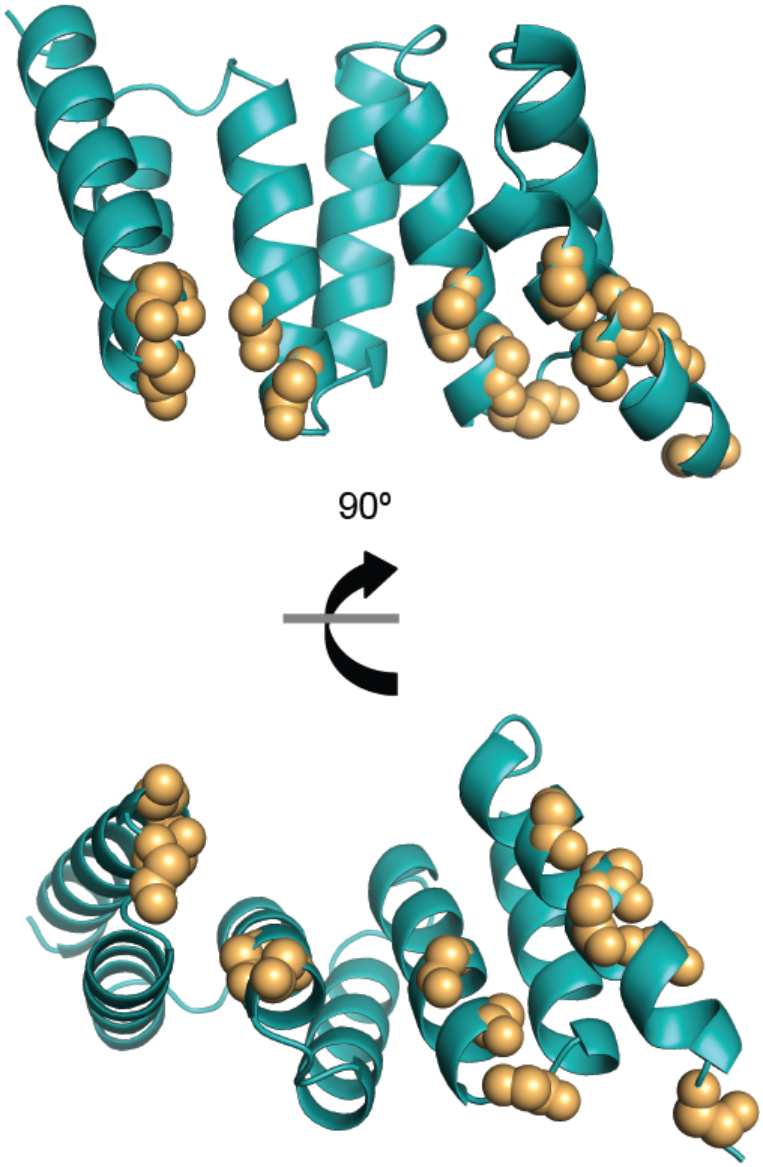



**Figure 5.26- PRE induced line-broadening mapped onto crystal structure of SGTA_TPR.** Orthogonal cartoon views of SGTA_TPR (PDB accession code 2VYI), showing SGTA_TPR residues affected by intermolecular PRE induced line broadening upon binding to MTSL labelled K398C $Rpn13_{260-407}$. Backbone amides within approximately 25 Å of $Rpn13_{260-407}$ residue 398 (spin labelled) are shown as orange spheres.

## 5.7 SGTA_TPR/Rpn13 C-terminal DMSLD pentapeptide complex

CSP experiments, mutagenesis data, together with insights gleaned from paramagnetic spin labelling studies, were used to generate a model of a 1:1 complex using the HADDOCK based semi-rigid, data-driven approach. The coordinates of the DMSLD peptide were obtained from the solution NMR structure of Rpn13 (PDB accession code: 2KR0). On the other hand, the coordinates of SGTA_TPR were obtained from its crystal structure (PDB accession code: 2VYI). The three lowest energy HADDOCK clusters of the SGTA_TPR/DMSLD complex were analysed (Table 5.1; Fig. 5.27).

**Table 5.1-** Energetic parameters obtained for the three best clusters of SGTA_TPR/DMSLD peptide complex calculated using HADDOCK.

| Cluster | HADDOCK score | Cluster size | RMSD (Å) | Van der Waals energy (kcal/mol) | Restraints violation energy (kcal/mol) | Buried surface area (Å²) | Electrostatic energy (kcal/mol) | Desolvation energy (kcal/mol) | Z score |
|---|---|---|---|---|---|---|---|---|---|
| Cluster1 | -107.6 +/-2.1 | 40 | 3.6 +/- 0.1 | -21.1 +/- 4.8 | 33.4 +/- 17.89 | 1119.1 +/- 63.2 | -545.0 +/- 24.0 | 19.2 +/- 4.9 | -1.8 |
| Cluster2 | -98.8 +/- 6.5 | 5 | 1.0 +/- 0.1 | -17.8 +/- 5.8 | 97.8 +/- 19.63 | 1234.2 +/- 12.8 | -533.0 +/- 37.4 | 15.7 +/- 1.5 | -1.4 |
| Cluster3 | -92.2 +/- 18.5 | 4 | 3.9 +/- 0.1 | -22.7 +/- 2.8 | 38.9 +/- 20.28 | 1009.9 +/- 73.8 | -455.2 +/- 93.0 | 17.7 +/- 2.7 | -1.1 |

Based on this model, the SGTA_TPR residues present at the binding interface with the DMSLD peptide are Asn-99, Asn-130, Lys- 160 and Arg-164, with Lys-160 and Arg-164 clamping the two carboxylates present on the terminal aspartate of the DMSLD peptide (Fig. 5.28). Although this structure represents the two-carboxylate clamp mechanism, it must be noted that the peptide is likely to be present in multiple conformations within the SGTA_TPR groove.

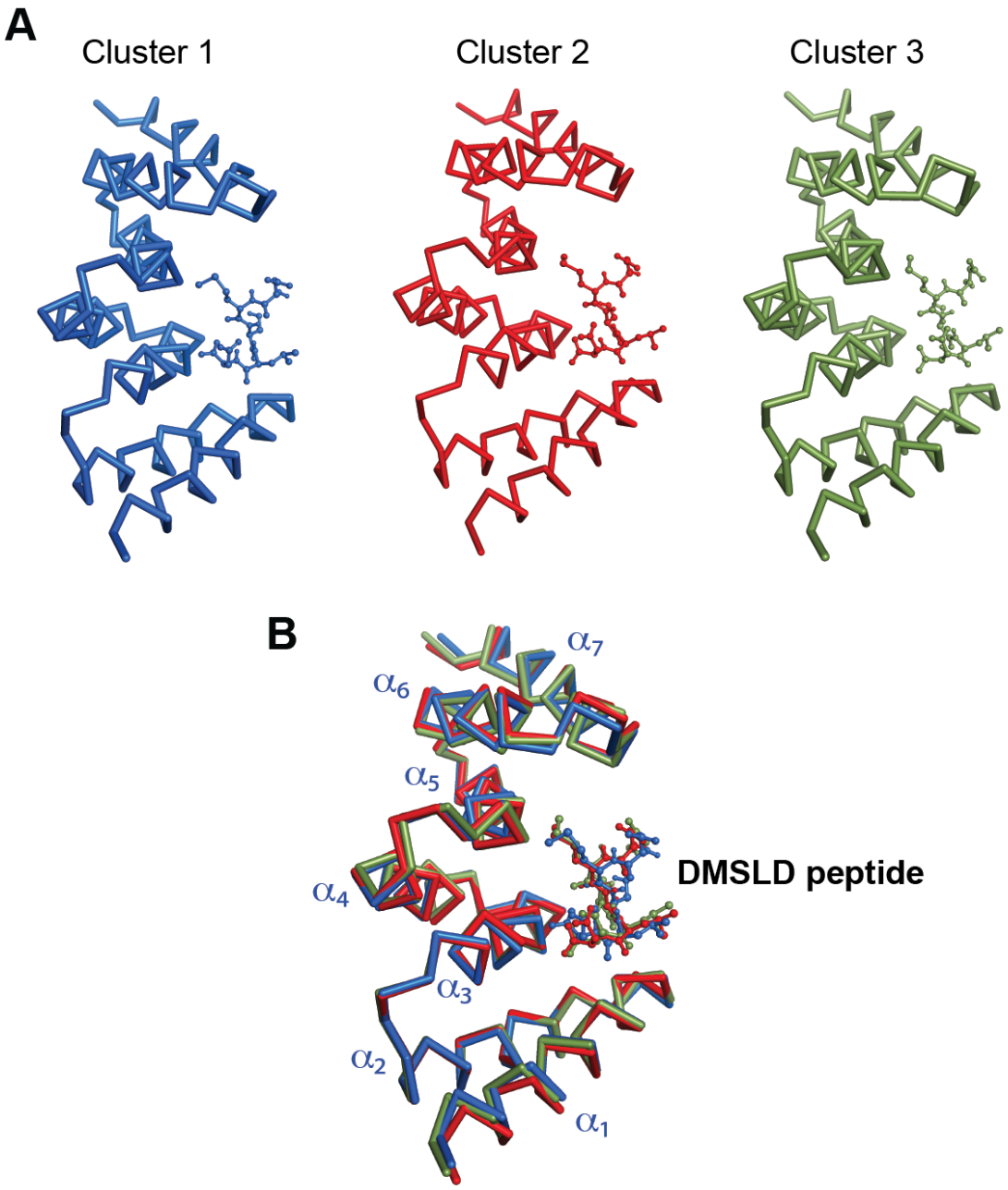



**Figure 5.27- SGTA_TPR/DMSLD peptide complex HADDOCK results.** A) The three top scoring clusters of the SGTA_TPR/DMSLD complex generated by HADDOCK. B) Overlay of HADDOCK-generated top scoring clusters of the SGTA_TPR/DMSLD complex showing the DMSLD peptide bound to the SGTA_TPR groove.

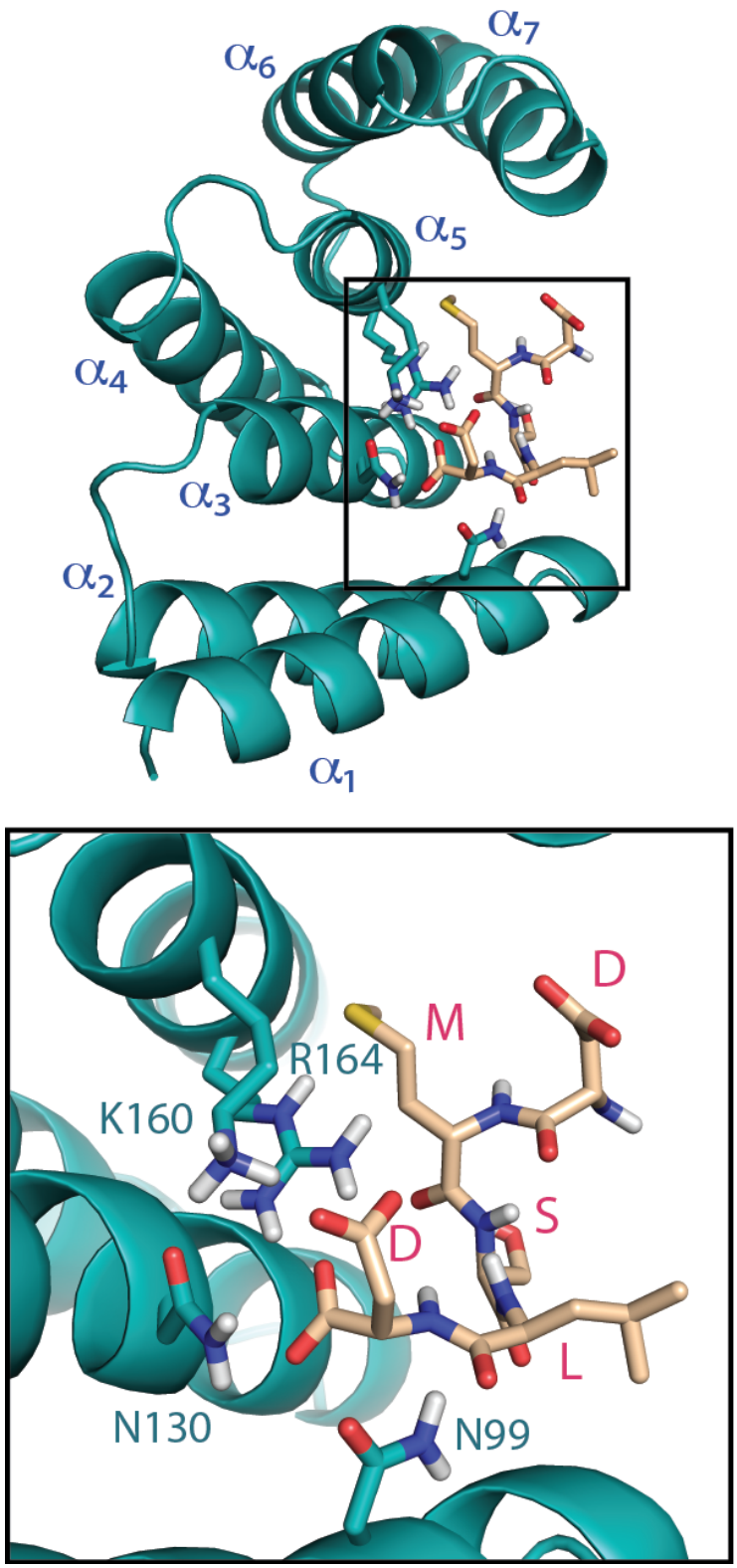



**Figure 5.28- SGTA_TPR/ DMSLD peptide complex generated using the HADDOCK approach.** The SGTA_TPR crystal structure is shown in cartoon representation labelled with TPR helices $\alpha_1$-$\alpha_7$. The Rpn13 derived DMSLD peptide is shown as sticks. The inset highlights SGTA_TPR residues Asn-99, Asn-130, Lys-160 and Arg-164 that form a network of electrostatic interactions with the terminal aspartate on the DMSLD peptide.

## 5.8 Summary

It is clear based on abovementioned experiments that SGTA interacts with the proteasome at its 19S regulatory particle via a two-carboxylate clamp mediated molecular recognition event, involving the extreme C-terminal residues of the intrinsic proteasomal ubiquitin receptor Rpn13 binding to conserved basic residues present in the SGTA_TPR groove. Furthermore, the extreme C-terminal DMSLD pentapeptide of Rpn13 is necessary and sufficient to facilitate this interaction. It also appears that this interaction is dynamic with the DMSLD pentapeptide sampling conformational space within the SGTA_TPR groove. These findings provide molecular details underlying a potential SGTA dependent quality control cycle operating at the 19S regulatory particle of the 26S proteasome, and could have implications in determining the fate of MLPs and TAs before they are committed to the 20S proteolytic core of the proteasome.

# Chapter 6

# Discussion

The aim of this thesis is to extend our knowledge of the SGTA protein's structure and its role in proteostasis, with a key focus on elucidating molecular details underlying its function to triage hydrophobic substrates in the cytosol. The results presented in this work describe the modular assembly of full-length SGTA *in vitro* (Chapter 3), its interaction with hydrophobic substrates (Chapter 4), and details of SGTA's association with the intrinsic proteasomal ubiquitin receptor Rpn13 at the 19S regulatory particle of the 26S proteasome (Chapter 5). In this chapter I critically evaluate aforementioned results and elaborate on their relevance in light of SGTA's biological role in maintaining cytosolic protein homeostasis.

## 6.1. Structure and dynamics of full-length SGTA *in vitro*

Our understanding of the assembly of full-length SGTA *in vitro* has been extended by work presented in Chapter 3. Thus far, biophysical characterization of SGTA was limited to isolated domains, with high-resolution structures of its N-terminal domain determined by solution NMR spectroscopy, and of its central TPR domain by X-ray crystallography (Darby *et al.*, 2014; Dutta and Tan, 2008). The reason as to why characterisation of full-length SGTA has not been possible by conventional structural biology methods alone (such as X-ray crystallography, NMR and cryo-EM) could be due to the intrinsic flexiblilty of linker regions between domains and a particularly challenging C-terminal region that has defied these methods to understand its structure. This study presents a hybrid approach, one that combines a range of biophysical techniques to include solution NMR spectroscopy, CD spectroscopy, and non dissociative nano-electrospray ionization mass spectrometry (Native MS) to reveal significant insights into the domain organization and conformation of SGTA in the context of the full-length protein.

One of the findings of this study is that in full-length SGTA each of its domains are independent of one another. This is likely due to the presence of flexible linkers between folded domains. In addition, CD spectroscopy has identified the presence of alpha-helical regions within the C-terminal domain. Whereas solution NMR studies indicate the presence of unstructured regions within the C-terminal domain, it is likely that these alpha-helical elements lie in regions for which signals have not been visible by standard solution NMR methods (Fig. 3.13). However it has not been possible to pinpoint the exact location of these alpha-helical regions within the C-terminal domain, largely due to the fact that intrinsic aggregation of this domain, together with its tendency to form hydrogels at higher concentrations, has hindered typical high-resolution structural analyses by standard methods.

Also, in this study, native MS experiments have revealed, for the first time, that the excised C-terminal region of SGTA exists as a dimer in solution. Furthermore, these experiments have also confirmed that full-length SGTA exists as a dimer *in vitro*. However, with respect to C-terminal dimerisation, the exact nature of this interaction is yet to be understood. This is mainly due to the lack of NMR signals corresponding to regions within the C-terminal domain, which is suggestive of interaction on an undetectable timescale, and therefore it has not been possible to obtain the precise location of motifs that could be driving this process of dimerisation. Together with information obtained from NMR relaxation experiments, it is evident that full-length SGTA can adopt a 'closed' conformation in solution, driven by C-terminal dimerisation which thereby constrains the mobility of the central TPR domains.

The presence of this closed conformation of full-length SGTA is likely to have implications regarding its recognition of hydrophobic substrates. However it is yet to be

fully understood as to whether this 'closed' state of SGTA is representative of a substrate bound conformation, or if this on the other hand, represents an apo conformation of SGTA. Nonetheless, it is interesting to speculate whether substrate binding and release could result from an interplay between 'closed' and 'open' conformations. One possibility is that the SGTA C-terminal closing action accommodates the substrate, thereby masking it from the surrounding aqueous cytosol, in a manner that has recently been described with regard to the downstream TA targeting factor Get3's ability to bind hydrophobic TMD helices (Mateja *et al*., 2015). In the crystal structure of the Get3-TA complex, a single TMD helix is shielded by a large hydrophobic groove formed at the Get3 dimer interface (Mateja *et al*., 2015). However, in the case of Get3, this substrate bound 'closed' conformation is stabilized by ATP binding, whereas SGTA is devoid of such nucleotide binding or hydrolysis capability. In any case, a tweezer-like opening and closing model of substrate binding remains a possibility and should be pursued further in order to understand the mechanism of substrate binding by SGTA.

## 6.2. SGTA interacts with hydrophobic substrates via its C-terminal domain

A key function of SGTA is to shield hydrophobic substrates from exposure to the aqueous cytosol. The C-terminal domain of SGTA is functionally important for binding MLPs and related hydrophobic substrates, and is responsible for the inhibitory effect of SGTA overexpression on MLP degradation (Leznicki and High, 2012; Liou and Wang, 2005). Work presented in Chapter 4 of this thesis provides further insights with respect to the nature and extent of this substrate-binding region of SGTA.

First of all, 2D NMR titrations of the $^{15}$N labeled TPR_NNP construct with unlabeled TA proteins showed that both substrates associated to a certain extent with the capping

helix of SGTA_TPR. This identifies a potential substrate-binding role of the TPR capping helix that is located adjacent to the C-terminal domain. These experiments have also demonstrated that TA substrates interact with the NNP region of the SGTA C-terminal domain. However, due to a considerable absence of intrinsic NMR signals corresponding to this region, this observation could not be unambiguously verified. Nonetheless, the observation of extrinsic $^{19}F$ NMR labels introduced through cysteine mutants within the C-terminal domain was, for the first time, able to provide observable NMR signals corresponding to both NNP and Q-rich regions within the C-terminal domain. Using Cytb5 as a TA substrate, perturbations of $^{19}F$ NMR signals could be observed within the NNP region, and were particularly evident in signals arising from cysteine mutants flanking the NNP2 and NNP3 motifs within the SGTA C-terminal domain. Similarly, perturbations of $^{19}F$ NMR signals could be observed in the SGTA Q-rich region upon interaction with Cytb5. Thus it appears that in the case of the TA protein Cytb5, consisting of a TMD helix of moderate hydrophobicity, both NNP (motifs 2 and 3) and Q-rich regions are required for its association with SGTA. It could be the case that substrates with differing extents of hydrophobicity require a greater, or lesser, degree of shielding by the C-terminal domain. For instance, MLPs presenting a greater extent of surface exposed hydrophobicity may utilise the entirety of the C-terminal domain to avoid exposure of vulnerable regions. It could also be the case that for such MLPs more than one SGTA dimer may be required to confer sufficient protection.

Previous studies have shown that the C-terminal Q-rich region alone is unable to bind larger hydrophobic substrates such as the GLUT1 membrane transporter, however the same region was shown to be sufficient for associating with smaller hydrophobic peptides (Liou and Wang, 2005). Therefore, it appears that the Q-rich region may act as

an accessory when dealing with larger substrates whereas the NNP region provides the first line of protection for MLP and TA proteins in the cytosol. This could explain the requirement for an expansion of NNP motifs within the C-terminal domain of human SGTA, which may stem from an evolutionary need to triage an increasingly diverse repertoire of hydrophobic substrates, in comparison to substrates presented to its yeast counterpart Sgt2. Unlike the metazoan homologues, Sgt2 only has one NNP motif within its C-terminal domain. Similarly, there appears to have been an evolutionary rearrangement in the Q-rich region of SGTA, as can be gleaned from sequence comparison with Sgt2. Glutamine residues in Sgt2 are more scattered throughout its C-terminal region, rather than being present as stretches consisting of up to three glutamines at a time, as can be observed in SGTA's Q-rich region. This could also increase mammalian SGTA's effectiveness in masking regions of hydrophobicity. Thus, in order to gain a better understanding of SGTA's hydrophobic substrate binding region it would be particularly useful to characterise its binding with different MLP and TA proteins, with varying extents of surface exposed hydrophobicity.

## 6.3 SGTA associates with the intrinsic proteasomal ubiquitin receptor Rpn13 via a carboxylate clamp mechanism

Work presented in Chapter 5 of this thesis was able to conclusively demonstrate that SGTA is recruited to the 19S regulatory particle of the 26S proteasome via the intrinsic proteasomal ubiquitin receptor Rpn13. In particular, this study has determined that the extreme C-terminal pentapeptide of Rpn13 is necessary and sufficient to facilitate this interaction, which occurs through a two-carboxylate clamp mediated molecular recognition event of said pentapeptide by the central TPR domain of SGTA.

The binding affinity between SGTA_TPR and Rpn13, in the context of both the isolated pentapeptide and the $Rpn13_{260-407}$ C-terminal construct, as measured by ITC yields $K_d$ values on the tens of micromolar scale. These $K_d$ values match well with the $K_d$ obtained from our NMR titration of unlabelled SGTA_TPR into $^{15}N$- labelled $Rpn13_{260-407}$ (Fig. 5.9). However, from the perspective of the reciprocal titration (Figs. 5.6 and 5.17), it was not possible to fit the data to a meaningful isotherm, as saturation could not be achieved even after titrating to a 1:8 molar ratio (SGTA_TPR: $Rpn13_{260-407}$ or DMSLD pentapeptide). This suggests weaker binding than that implied by the values obtained using ITC which is intriguing given the results from the reverse experiment. Moreover $K_d$ values obtained by ITC are consistent with reported values of other TPR domains interacting with MEEVD/IEEVD peptides of Hsp70 and Hsp90 molecular chaperones, respectively, via a similar two-carboxylate clamp mechanism (Scheufler *et al.*, 2000). More pertinently, similar $K_d$ values have been reported by ITC with the TPR domain of SGTA's *C. elegans* homolog upon interaction with MEEVD/IEEVD peptides of Hsp70/90 chaperones (Worrall *et al.*, 2008). Therefore, based on the literature around TPR-carboxylate clamps, many low micromolar range $K_d$ values have been reported as measured by ITC (Scheufler *et al.*, 2000; Worrall *et al.*, 2008) and sometimes by CD spectroscopy (Cliff *et al.*, 2005). An explanation for the widespread perturbations of amide signals of $^{15}N$-labelled SGTA_TPR (Figs. 5.6 and 5.17), observed upon binding to unlabeled Rpn13, could be that the titration from the $^{15}N$-labelled SGTA_TPR perspective is not purely representative of a binding equilibrium. There may well be peaks that approach saturation, while ones that defy saturation could represent some other underlying process. However, from the perspective of $^{15}N$-labelled $Rpn13_{260-407}$ (Fig. 5.9), upon titration with unlabelled SGTA_TPR, it is clear that this is representative of a binding equilibrium thus is in agreement with $K_d$ values obtained by ITC.

The significance of the carboxylate clamp dependent binding of SGTA to the C-terminal DEUBAD domain of Rpn13 can be understood from recent work carried out in cultured mammalian cells (Thapaliya *et al.*, 2016; Leznicki *et al.*, 2015). These studies suggest that recruitment of exogenous SGTA to the proteasome can subsequently reduce the degradation of MLPs. It is known that overexpression of exogenous SGTA in mammalian cells results in an increase in the steady state level of model MLPs, an effect that could be attributed to a delay in MLP degradation at the proteasome (Wunderley *et al.*, 2014). Also, it has been shown that the overexpression of an SGTA double mutant with a defective carboxylate clamp does not have the aforementioned stabilization effect on MLPs (Leznicki *et al.*, 2015). In addition, it has been shown that co-expression of SGTA with Rpn13 variants with defective SGTA binding regions leads to a stabilization of MLPs (Thapaliya *et al.*, 2016). Taken together, these findings favour the idea that exogenous SGTA can enhance model MLP levels, an effect that relies on its binding to the extreme C-terminal pentapeptide of Rpn13. As far as endogenous SGTA is concerned, the importance of its interaction with Rpn13 is yet to be fully understood. Nonetheless, is has been suggested that binding of endogenous SGTA to Rpn13 can influence MLP degradation (Leznicki *et al.*, 2015).

## 6.4 SGTA's role in the quality control of hydrophobic substrates

There is a growing body of evidence to illuminate SGTA's role in enforcing quality control of mislocalized hydrophobic substrates, thereby contributing to cytosolic proteostasis. The contribution of SGTA to cytosolic quality control of MLPs was first investigated in a study by Leznicki and High (2012), in which it was shown that SGTA is capable of reversing ubiquitination of MLPs instigated by BAG6. This led to the proposal of a SGTA/BAG6 quality control cycle that decided the fate of MLPs in the

cytosol, leading to either BAG6 dependent ubiquitination and proteasomal degradation, or SGTA dependent deubiquitination. In agreement with this proposed SGTA/BAG6 cycle, the High group have also shown that SGTA overexpression selectively delays the proteasomal degradation of a model mislocalized protein (MLP) but not a related control protein with a disrupted hydrophobic region (Wunderley *et al.*, 2014). In addition, it has also been demonstrated that the overexpression of exogenous SGTA increases steady state MLP levels whilst a knockdown of endogenous SGTA reduces steady state MLP levels (Wunderley *et al.*, 2014). It is also worth noting that a completely independent study by Waheed *et al.* (2016) found that SGTA overexpression in the presence of the HIV Vpu protein leads to the stabilization of a non- glycosylated form of the endogenous type II membrane protein tetherin, which accumulates in the cytosol. For all these reasons, it is becoming increasingly apparent that SGTA is a key effector in deciding the fate of aberrant and mislocalised membrane proteins.

The results presented in this thesis provide insights into the possibility of a SGTA dependent quality control cycle operating at the 19S regulatory particle of the proteasome (Fig. 6.1). As full-length SGTA is present as a homodimer, it is possible for its individual TPR domains to associate with different quality control factors, whilst its C-terminal domain is bound to hydrophobic substrates. It is known that SGTA can form complexes with both Hsp90 and Hsp70 molecular chaperones (Liou and Wang, 2005; Liu *et al.*, 1999), and with the intrinsic proteasomal ubiquitin receptor Rpn13 (Thapaliya *et al.*, 2016; Leznicki *et al.*, 2015) via its central TPR domain. Thus it is plausible to envision a scenario in which SGTA-bound substrates find themselves in close proximity to these ATP-dependent molecular chaperones, whilst at the same time being associated with the 19S regulatory particle of the proteasome. This could provide

substrates that arrive at the proteasome with access to the ATP-dependent (Hsp70/90) chaperone networks.

Furthermore, it is known that BAG6 interacts with the intrinsic proteasomal receptor Rpn10, also present at the 19S regulatory particle, located in close proximity to Rpn13 (Kikukawa *et al.*, 2005; Minami *et al.*, 2010). Thus on one hand it is possible for MLPs at the proteasome to be ubiquitinated by the actions of the E3 ligase RNF126 via BAG6 (Hessa *et al.*, 2011; Rodrigo Brenni *et al.*, 2014). On the other hand, it is also possible for SGTA bound substrates to be subjected to deubiquitination (Leznicki and High, 2012; Wunderley *et al.*, 2014). At the 19S regulatory particle, the mechanism whereby SGTA promotes substrate deubiquitination could be explained by the proximity that SGTA-bound substrates gain to the UCH37 deubiquitinase, which is in turn achieved through SGTA's carboxylate clamp dependent association with Rpn13. As Rpn13 activates UCH37 via an interaction with its C-terminal DEUBAD domain (Sahtoe *et al.*, 2015; VanderLinden *et al.*, 2015), this provides SGTA bound MLP substrates with an opportunity for selective deubiquitination. Taken together, it appears extremely likely that SGTA/BAG6 dependent cycles of MLP substrate ubiquitination and deubiquitination could take place at the 19S regulatory particle of the proteasome, in addition to such cycles occuring in the cytosol (Leznicki and High 2012; Wunderley *et al.*, 2014). The outcome of such a cycle could be either ubiquitination and proteasomal degradation, or deubiquitination and stabilisation (Fig. 6.1).

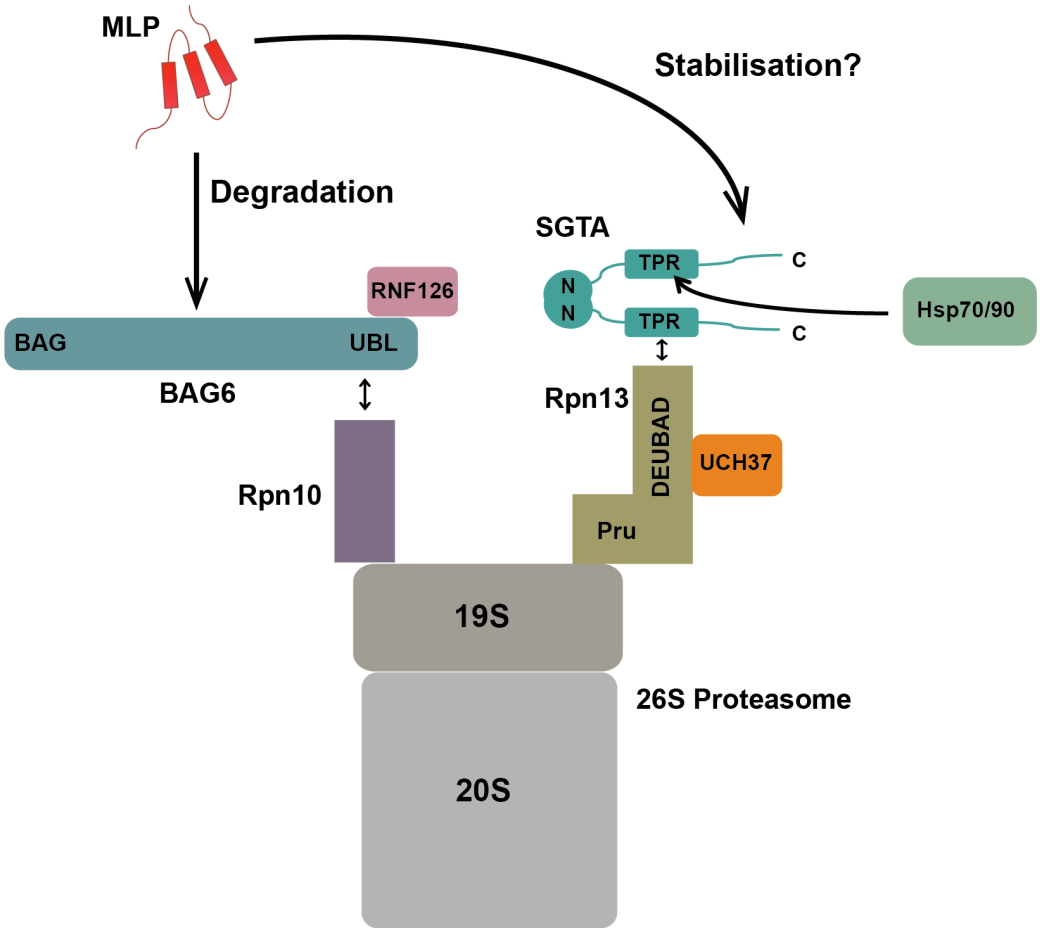



**Figure 6.1.- A putative SGTA/BAG6 quality control cycle of MLPs operating at the proteasome.** MLPs could arrive at the 19S regulatory particle of the proteasome via either BAG6 or SGTA, mediated by intrinsic proteasomal receptors Rpn10 and Rpn13, respectively (Thapaliya *et al.*, 2016; Leznicki *et al.*, 2015; Kikukawa *et al.*, 2005; Minami *et al.*, 2010). MLPs bound to BAG6 are selectively ubiquitinated by the actions of the E3 ligase, RNF126 (Krysztofinska *et al.*, 2016; Rodrigo Brenni *et al.*, 2014). Whereas MLPs bound to SGTA could be subjected to deubiquitination at the proteasome through the actions of UCH37. This could occur due to activation of the UCH37 deubiquitinase upon binding to Rpn13 (Sahtoe *et al.*, 2015; VanderLinden *et al.*, 2015). These concerted actions could lead to cycles of ubiquitination and deubiquitination of MLPs at the proteasome. In addition, the Hsp70/90 molecular chaperones could have a role to play in this process due to their carboxylate clamp dependent binding to the TPR domain of SGTA (Liou and Wang, 2005; Liu *et al.*, 1999).

However, the physiological role of a SGTA/BAG6 cycle operating at the proteasome remains to be understood. It appears unlikely that the physiological role of SGTA's recruitment to the proteasome would be to rescue terminally misfolded MLPs from degradation. This is due to the fact that, undegraded MLPs are known to have an increased propensity to aggregate (Wunderley *et al.*, 2014). For transiently unfolded MLPs and precursors this SGTA/BAG6 cycle at the proteasome could provide a final opportunity at productive folding or membrane targeting before being subjected to proteasomal recycling. Also, it is tempting to speculate that stabilisation of hydrophobic substrates resulting from the SGTA-Rpn13 interaction may constitute a pathway that rescues prematurely ubiquitinated tail-anchored membrane proteins, thereby extending their window of opportunity for productive membrane insertion.

## 6.5 Concluding remarks

Recent years have seen a growing interest to understand molecular mechanisms underlying intracellular protein aggregation, unsurprisingly due to its relevance to the pathogenesis of numerous human diseases. In this regard, SGTA has emerged as a central regulator for the quality control of membrane proteins that mislocalise to the cytosol (Leznicki *et al.*, 2015; Wunderley *et al.*, 2014; Xu *et al.*, 2012; Leznicki and High, 2012). Therefore, this thesis aims to further our knowledge of SGTA by providing a comprehensive biophysical study focusing on quality control pathways within which it operates.

Biophysical characterization of full-length SGTA (Chapter 3) has for the first time provided evidence of a 'closed' conformation of SGTA in solution, mediated by the dimerisation of its C-terminal domain. However, it has not been possible to discern the precise region within the C-terminal domain that facilitates dimerisation. Further work

will be required to identify such regions within the C-terminal domain. One approach would be to generate a number of C-terminal truncated variants of SGTA, followed by native MS analysis to identify the minimal region required for C-terminal dimerisation. Also the existence of 'open' vs 'closed' conformations of full-length SGTA remains a possibility, and should be investigated further in relation to its association with hydrophobic substrates.

Work presented in Chapter 4 investigates the interaction of SGTA with hydrophobic substrates, using tail-anchored membrane proteins as examples. This work has identified regions of the TPR capping helix, together with regions within the C-terminal domain of SGTA, namely the NNP motifs and the Q-rich region, to be necessary for substrate binding. An extension of work presented in this thesis would be to use this information to design mutants targeting regions in the TPR capping helix and in the C-terminal domain of SGTA for overexpression studies in cultured mammalian cells to gain functional insights as to how these regions contribute to the stabilization of model MLPs. Together with further work *in vitro* to understand SGTA's interaction with substrates that present varying degrees of hydrophobicity, functional studies in mammalian cells would be particularly useful in understanding the process of hydrophobic substrate recognition, binding and release by SGTA.

The recruitment of SGTA to the proteasome via a carboxylate clamp dependent binding to the intrinsic proteasomal receptor Rpn13 provides the molecular basis for a putative MLP quality control cycle at the proteasome (Chapter 5). Such a cycle would constitute a process whereby successive cycles of substrate ubiquitination and deubiquitination occurring at the 19S regulatory particle of the proteasome, as a result of the concerted actions of associated BAG6 and SGTA, would eventually determine the fate of MLPs

that arrive at the proteasome. However, further work to reconstitute such a SGTA/BAG6 cycle at the proteasome with purified components, together with assays in cultured mammalian cells, would be necessary to conclusively demonstrate mechanisms of MLP quality control that occur at the proteasome.

# Bibliography

Aguileta M. A., Korac J., Durcan T. M., Trempe J. F., Haber M., Gehring K., Elsasser S., Waidmann O., Fon E. A., Husnjak K. (2015) The E3 ubiquitin ligase parkin is recruited to the 26S proteasome via the proteasomal ubiquitin receptor Rpn13. *J Biol Chem.* **290** (12), 7492-505.

Anfinsen C. B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223-30.

Aviram N., Ast T., Costa E. A., Arakel E. C., Chuartzman S. G., Jan C. H., Haßdenteufel S., Dudek J., Jung M., Schorr S., Zimmermann R., Schwappach B. Weissman J. S., and Schuldiner M. (2016) The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature* **540**, 134-138.

Barrett P. J., Chen J., Cho M. K., Kim J. H., Lu Z., Mathew S., Peng D., Song Y., Van Horn W. D., Zhuang T., Sönnichsen F. D., and Sanders C. R. (2013) The quiet renaissance of protein nuclear magnetic resonance. *Biochemistry* **52** (8), 1303-20.

Battiste J. L., and Wagner G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry.* **39** (18) 5355–5365.

Bernard C. (1878) *Leçons sur les phénomènes de la vie communs aux animaux et aux vegetaux*, Paris, Baillière.

Bodenhausen G. and Ruben D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69** (1) 185–189.

Buchanan G., Ricciardelli C., Harris J. M., Prescott J., Yu Z. C., Jia L., Butler L. M., Marshall V. R., Scher H. I. and Gerald W. L. (2007). Control of androgen receptor signaling in prostate cancer by the cochaperone small glutamine rich tetratricopeptide repeat containing protein alpha. *Cancer Res.* **67**, 10087–10096.

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.

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.

Campbell I. D. (2012) *Biophysical techniques.* Oxford University Press, Oxford. ISBN: 978-0-19-964214-4.

Cannon W.B. (1939) *Wisdom of the Body*, New York, WW Norton.

Casey F., Krogan N., Shields D. C., and Cagney G. (2011) Distinct configurations of protein complexes and biochemical pathways revealed by epistatic interaction network motifs. *BMC Syst Biol.* **5**, 133.

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.

Chakrabarti O. and Hegde R. S. (2009) Functional depletion of mahogunin by cytosolically exposed prion protein contributes to neurodegeneration. *Cell* **137**, 1136–1147.

Chartron J. W., VanderVelde D. G., and Clemons W. M. Jr. (2012) Structures of the Sgt2/SGTA dimerization domain with the Get5/UBL4A UBL domain reveal an interaction that forms a conserved dynamic interface. *Cell Rep.* **2** (6), 1620-32.

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.

Chen X., Randles L., Shi K., Tarasov S. G., Aihara H., Walters K. J. (2016) Structures of Rpn1 T1:Rad23 and hRpn13:hPLIC2 reveal distinct binding mechanisms between substrate receptors and shuttle factors of the proteasome. *Structure* **24** (8) 1257-70.

Chen X., Lee B. H., Finley D., and Walters K. J. (2010) Structure of proteasome ubiquitin receptor hRpn13 and its activation by the scaffolding protein hRpn2. *Mol. Cell.* **38**, 404–415.

Chiti F. and Dobson C.M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.

Clark A. R., Lubsen N. H. and Slingsby C. (2012) sHSP in the eye lens: crystallin mutations, cataract and proteostasis. *Int. J. Biochem. Cell B.* **44** (10) 1687-1697.

Cliff M. J., Harris R., Barford D., Ladbury J. E., and Williams M. A. (2006) Conformational diversity in the TPR domain-mediated interaction of protein phosphatase 5 with Hsp90. *Structure* **14**, 415– 426.

Cliff M. J., Williams M. A., Brooke-Smith J., Barford D., and Ladbury J. E. (2005) Molecular recognition via coupled folding and binding in a TPR domain. *J Mol Biol.* **346**, 717-732.

Clubb R. T., Thanabal V., and Wagner, G. (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C' chemical shifts in 15N-13C-labeled proteins. *J. Magn. Reson.* **97**, 213-217.

Cziepluch C., Kordes E., Poirey R., Grewenig A., Rommelaere J., 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.

D'Andrea L. D., and Regan L. (2003) TPR proteins: the versatile helix. *Trends Biochem Sci.* **28** (12), 655-62.

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., Martínez-Lumbreras S., Thapaliya A., High S., Isaacson R. L. (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.

Das A. K., and Cohen P. W., and Barford D. (1998) The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* **17** (5), 1192-9.

Delaglio F., Grzesiek S., Vuister G. W., Zhu G., Pfeifer J., and Bax A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* **6**, 277-293.

Deveraux Q., Ustrell V., Pickart C., and Rechsteiner M. (1994) A 26S protease subunit that binds ubiquitin conjugates. *J Biol Chem.* **269** (10), 7059-61.

Didenko T., Liu J. J., Horst R., Stevens R. C., and Wüthrich K. (2013). Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Current Opinion in Structural Biology*, **23** (5), 740–747.

Dobson C. M., Sali A., and Karplus M. (1998) Protein folding - a perspective from theory and experiment. *Angew. Chem. Int. Edn Engl.* **37**, 868−893.

Dominguez C., Boelens R. and Bonvin A. M. J. J. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731-1737.

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.

Dyson H. J., and Wright P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197-208.

Ellis J. (1987) Proteins as molecular chaperones. *Nature.* **328**, 378-9.

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.

Finley D., Chen X. and Walters K. J. (2016) Gates, channels, and switches: Elements of the proteasome machine. *Trends Biochem Sci.* **41** (1), 77-93.

Fonte V., Kapulkin W. J., Taft A., Fluet A., Friedman D., Link C. D. (2002) Interaction of intracellular beta amyloid peptide with chaperone proteins. *Proc Natl Acad Sci USA.* **99** (14), 9439-44.

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), 1214-1219.

Greenfield N. J. (2007) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876-2890.

Gristick H. B., Rome M. E., Chartron J. W., Rao M., Hess S., Shan S. O., 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.

Grzesiek S., and Bax A. (1992a) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114,** 6291-6293.

Grzesiek S., and Bax A. (1992b) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201-207.

Grzesiek S., and Bax, A. (1992c) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432-440.

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**, 198–207.

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**, 556–567.

Hartl F. U., Bracher A., and Hayer-Hartl M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324-32.

Hartl F. U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**, 571−580.

Haslbeck M., Franzmann T., Weinfurtner D. and Buchner J. (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Biol.* **12** (10), 842-846.

Hegde R. S. and Keenan R. J. (2011) Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat Rev Mol Cell Biol.* **12**, 787-798.

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**, 394-397.

Hilton G. R., and Benesch J. L. P. (2012) Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J. R. Soc. Interface*, **9** (70), 801–816.

Hore P. J. (1995) *Nuclear Magnetic Resonance*, Oxford University Press, Oxford.

Horst R., Liu J. J., Stevens R. C., Wüthrich K. (2013) β2-adrenergic receptor activation by agonists studied with $^{19}$F-NMR. *Angew Chemie.* **52** (41) 10762-5.

Huang X., Luan B., Wu J., and Shi Y. (2016) An atomic structure of the human 26S proteasome. *Nat Struct Mol Biol.* **23** (9), 778-85.

Hubbard S. J. and Thornton J. M. (1993) 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London.

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**, 481-8.

Itakura E., Zavodszky E., Shao S., Wohlever M. L., Keenan R. J., Hegde R. S. (2016) Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. *Mol Cell.* **63** (1) 21-33.

Kaganovich D., Kopito R., and Frydman J. (2008) Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**, 1088−1095.

Kay L. E., Torchia D. A., Bax A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry.* **28** (23) 8972-9.

Keeler J. (2005) *Understanding NMR Spectroscopy*, Wiley.

Kiktev D. A., Patterson J. C., Müller 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.

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 Journal* **272**, 6373–6386.

Komander D. and Rape M. (2012) The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203-229.

Konarev P. V., Volkov V. V., Sokolova A. V., Koch M. H. J. and Svergun D. I. (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J Appl Cryst.* **36**, 1277-1282.

Kopito R. R. (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10**, 524−530.

Kordes E., Savelyeva L., Schwab M., Rommelaere J., Jauniaux J. and Cziepluch C. (1998) Isolation and characterization of human SGT and identification of homologues in Saccharomyces cerevisiae and Caenorhabditis elegans. *Genomics* **52**, 90-94.

Krysztofinska E. M., Martínez-Lumbreras S., Thapaliya A., Evans N. J., High S., Isaacson R. L. (2016) Structural and functional insights into the E3 ligase, RNF126. *Sci Rep.* **6**, 26433.

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**, 366–375.

Kuwabara N., Minami R., Yokota N., Matsumoto H., Senda T., Kawahara H., 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.

Kuzmic P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237**, 260-273.

Laemmli U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** (5259), 680-685.

Lam Y. A., Xu W., DeMartino G. N., Cohen R. E. (1997) Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. *Nature* **385**, 737-40.

Laskey R. A., Honda B. M., Mills A. D., and Finch J. T. (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature.* **275**, 416-20.

Levine C. G., Mitra D., Sharma A., Smith C. L., and Hegde R. S. (2005) The efficiency of protein compartmentalization into the secretory pathway. *Mol. Biol. Cell.* **16**, 279–291.

Leznicki P., Korac-Prlic J., Kliza K., Husnjak K., Nyathi Y., Dikic I., and High S. (2015) SGTA binding to Rpn13 selectively modulates protein quality control. *J Cell Sci.* **128** (17), 3187-3196.

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

Leznicki P., and High S. (2012) SGTA antagonizes BAG6-mediated protein triage. *Proc Natl Acad Sci. USA* **109**, 19214-19219.

Leznicki P., Clancy A., Schwappach B., and High S. (2010) Bat3 promotes the membrane integration of tail-anchored proteins. *J. Cell. Sci.* **123**, 2170–2178.

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.

Liu F-H., Wu S-J., Hu S-M., Hsiao C-D., and Wang C. (1999) Specific interaction of the 70 kDa heat shock cognate protein with the tetratricopeptide repeats. *J Biol Chem.* **274**, 34425-34432.

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

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**, 61–66.

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.

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.

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**, 637–650.

Mizushima N., and Komatsu M. (2011) Autophagy: renovation of cells and tissues. *Cell* **147** (4), 728–741.

Mock J. Y., Chartron J. W., Zaslaver M., Xu Y., Ye Y., Clemons W. M. (2015) Bag6 complex contains a minimal tail-anchor-targeting module and a mock BAG domain. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 106–111.

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., Buchner J., Chadli A., and Cox M. B. (2014) The cochaperone SGTA (small glutamine-rich tetratricopeptide repeat-containing protein α) demonstrates regulatory specificity for the androgen, glucocorticoid and progesterone receptors. *J Biol. Chem.* **289**, 15297–15308.

Payapilly A. and High S. (2014) BAG6 regulates the quality control of a polytopic ERAD substrate. *J Cell Sci.* **127**, 2898-2909.

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.

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.

Powers E. T., Morimoto R. I., Dillin A., Kelly J. W., Balch W. E. (2009) Biological and chemical approaches to diseases of proteostasis deficiency. *Annu Rev Biochem.* **78**, 959-91.

Rapoport T. A. (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669.

Redfield C. (1999) Molten globules. *Curr Biol.* **9** (9) 313.

Roberts J. D., Thapaliya A., Martínez-Lumbreras S., Krysztofinska E. M., Isaacson R. L. (2015) Structural and functional insights into small, glutamine-rich, tetratricopeptide repeat protein alpha. *Front Mol Biosci.* **2** (71) doi: 10.3389/fmolb.2015.00071.

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.

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

Sahtoe D. D., van Dijk W. J., Oualid F. E., Ekkebus R., Ovaa H., and Sixma T. K. (2015) Mechanism of UCH-L5 activation and inhibition by DEUBAD domains in RPN13 and INO80G. *Mol. Cell.* **57**, 887–900.

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** (3) 855–863.

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**, 199-210.

Schreiner P., Chen X., Husnjak K., Randles L., Zhang N., Elsasser S., Finley D., Dikic I., Walters K. J., Groll M. (2008) Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **453**, 548-52.

Schuldiner M., Metz J., Schmid V., Denic V., Rakwalska M., Schmitt H. D., Schwappach B., Weissman J. S. (2008) The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634–645.

Shao S., and Hegde R. S. (2016) Target selection during protein quality control. *Trends Biochem Sci.* **41** (2), 124-37.

Shi Y., Chen X., Elsasser S., Stocks B. B., Tian G., Lee B. H., Shi Y., Zhang N., de Poot S. A., Tuebing F., Sun S., Vannoy J., Tarasov S. G., Engen J. R., Finley D., Walters K. J. (2016) Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351,** 9421.

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.

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.

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 Cryst* D **71**, 154-161.

Stein A., Ruggiano A., Carvalho P., Rapoport T. A. (2014) Key steps in ERAD of luminal ER proteins reconstituted with purified components. *Cell.* **158** (6), 1375-88.

Stefanovic S., and Hegde R. S. (2007) Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147-1159.

Stefer S., Reitz S., Wang F., Wild K., Pang Y. Y., Schwarz D., Bomke J., Hein C., Löhr F., and Bernhard F. (2011) Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* 333:758–762.

Thapaliya A., Nyathi Y., Martínez-Lumbreras S., Krysztofinska E. M., Evans N. J., Terry I. L., High S., Isaacson R. L. (2016) SGTA interacts with the proteasomal ubiquitin receptor Rpn13 via a carboxylate clamp mechanism. *Sci Rep.* **6**, 36622.

Tobaben S., Thakur P., Fernández-Chacón R., Südhof T. C., Rettig J., and Stahl B. (2001). A trimeric protein complex functions as a synaptic chaperone machine. *Neuron* **31**, 987–999.

Trotta A. P., Need E. F., Selth L. A., Chopra S., Pinnock C. B., Leach D. A., Coetzee G. A., Butler L. M., Tilley W. D., Buchanan G. (2013) Knockdown of the cochaperone SGTA results in the suppression of androgen and PI3K/Akt signaling and inhibition of prostate cancer cell proliferation. *Int J Cancer.* **133** (12), 2812-23.

van den Heuvel R. H. and Heck A. J. R. (2004) Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **8**, 519–526.

van Zundert G. C. P., Rodrigues J. P. G. L. M., Trellet M., Schmitz C., Kastritis P. L., Karaca E., Melquiond A. S. J., van Dijk M., de Vries S. J. and Bonvin A. M. J. J. (2016) The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720-725.

VanderLinden R. T., Hemmis C. W., Schmitt B., Ndoja A., Whitby F. G., Robinson H., Cohen R. E., Yao T., and Hill C. P. (2015) Structural basis for the activation and inhibition of the UCH37 deubiquitylase. *Mol. Cell.* **57**, 901–911.

Velazquez-Campoy A., Leavitt S. A., Freire E. (2004) Characterization of protein-protein interactions by isothermal titration calorimetry. Protein-Protein Interactions: Methods and Applications, *Methods in Molecular Biology*, **261**, 35-54.

Vijay-Kumar S., Bugg C. E., Cook W. J. (1987) Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **194**, 531–44.

Vilardi F., Lorenz H., Dobberstein B. (2011) WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. *J. Cell. Sci.* **124**, 1301–1307.

Vranken W. F., Boucher W., Stevens T. J., Fogh R. H., Pajon A., Llinas M., Ulrich E. L., Markley J. L., Ionides J., and Laue E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* **59** (4), 687-96.

Wagner G. (1995) The importance of being floppy. *Nature Struct Biol.* **2**, 255 – 257.

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.

Walczak C. P., Ravindran M. S., Inoue T., 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**, e1004007.

Wang Q., Liu Y., Soetandyo N., Baek K., Hegde R., and Ye Y. (2011) A ubiquitin ligase-associated chaperone holdase maintains polypeptides in soluble states for proteasome degradation. *Mol Cell.* **42** (6), 758-70.

Wang F., Chan C., Weir N. R., Denic V. (2014) The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature* **512**, 441–444.

Waterhouse A. M., Procter J. B., Martin D. M. A., Clamp M., Barton G. J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.

Wickner S., Maurizi M. R., and Gottesman S. (1999) Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286**, 1888-1893.

Winnefeld M., Grewenig A., Schnölzer 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.

Wolff S., Weissman J. S., and Dillin A. (2014) Differential scales of protein quality control. *Cell.* **157** (1), 52-64.

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**, 496-503.

Wunderley L., Leznicki P., Payapilly A. and High S. (2014) SGTA regulates the cytosolic quality control of hydrophobic substrates. *J Cell Sci.* **127**, 4728-4739.

Xu Y., Cai M., Yang Y., Huang L., 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.

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.

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**, 387–397.

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.

Zhu T., Ji Z., Xu C., Peng Z., Gu L., Zhang R., and Liu Y. (2014) Expression and prognostic role of SGTA in human breast carcinoma correlates with tumor cell proliferation. *J. Mol. Histol.* **45**, 665-677.

# Appendices

# Appendix A

Synthetic gene sequences (with translated sequence) used for plasmid cloning

## A1. Codon optimised sequence encoding full-length human SGTA

```
atggacaacaagaagcgcctggcctacgccatcatccagttcctgcatgaccagctccgg
 M  D  N  K  K  R  L  A  Y  A  I  I  Q  F  L  H  D  Q  L  R
cacggggcctctcgtccgatgctcaggagagcttggaagtcgccatccagtgcctggag
 H  G  G  L  S  S  D  A  Q  E  S  L  E  V  A  I  Q  C  L  E
actgcgtttggggtgacggtagaagacagtgaccttgcgctcccccagactctgccggag
 T  A  F  G  V  T  V  E  D  S  D  L  A  L  P  Q  T  L  P  E
atatttgaagcggctgccacgggcaaggagatgccgcaggacctgaggagccccgcgcga
 I  F  E  A  A  A  T  G  K  E  M  P  Q  D  L  R  S  P  A  R
accccgccttccgaggaggactcagcagaggcagagcgcctcaaaaccgaaggaaacgag
 T  P  P  S  E  E  D  S  A  E  A  E  R  L  K  T  E  G  N  E
cagatgaaagtggaaaactttgaagctgccgtgcatttctacggaaaagccatcgagctc
 Q  M  K  V  E  N  F  E  A  A  V  H  F  Y  G  K  A  I  E  L
aacccagccaacgccgtctatttctgcaacagagccgcagcctacagcaaactcggcaac
 N  P  A  N  A  V  Y  F  C  N  R  A  A  A  Y  S  K  L  G  N
tacgcaggcgcggtgcaggactgtgagcgggccatctgcattgacccggcctacagcaaa
 Y  A  G  A  V  Q  D  C  E  R  A  I  C  I  D  P  A  Y  S  K
gcctacggcaggatgggcctggcgctctccagcctcaacaagcacgtggaggccgtggct
 A  Y  G  R  M  G  L  A  L  S  S  L  N  K  H  V  E  A  V  A
tactacaagaaggctctggagctggaccccgacaacgagacatacaagtccaacctcaag
 Y  Y  K  K  A  L  E  L  D  P  D  N  E  T  Y  K  S  N  L  K
atagcggagctgaagctgcgggaggcccccagccccacgggaggcgtgggcagcttcgac
 I  A  E  L  K  L  R  E  A  P  S  P  T  G  G  V  G  S  F  D
atcgccggcctgctgaacaaccctggcttcatgagcatggcttcgaacctaatgaacaat
 I  A  G  L  L  N  N  P  G  F  M  S  M  A  S  N  L  M  N  N
ccccagattcagcagctcatgtccggcatgatttcgggtggcaacaacccctgggaact
 P  Q  I  Q  Q  L  M  S  G  M  I  S  G  G  N  N  P  L  G  T
cccggcaccagcccctcgcagaacgacctggccagcctcatccaggcgggccagcagttt
 P  G  T  S  P  S  Q  N  D  L  A  S  L  I  Q  A  G  Q  Q  F
gcccagcagatgcagcagcagaacccagagttgatagagcagctcaggagccagatccgg
 A  Q  Q  M  Q  Q  Q  N  P  E  L  I  E  Q  L  R  S  Q  I  R
agtcggacgcccagcgccagcaacgacgaccagcaggagtga
 S  R  T  P  S  A  S  N  D  D  Q  Q  E  -
```

## A2. Codon optimised sequence encoding full-length human cytochrome b5

```
atggcagaacagagtgatgaagccgttaaatactataccctggaagaaatccagaaaacac
 M  A  E  Q  S  D  E  A  V  K  Y  Y  T  L  E  E  I  Q  K  H
aaccatagcaaaagcacctggctgattctgcatcataaagtgtatgatctgaccaaattt
 N  H  S  K  S  T  W  L  I  L  H  H  K  V  Y  D  L  T  K  F
ctggaagaacaccctggtggtgaagaagttctgcgcgaacaggcaggcggtgatgcaacc
 L  E  E  H  P  G  G  E  E  V  L  R  E  Q  A  G  G  D  A  T
gaaaattttgaagatgttggtcatagcaccgatgcacgtgaaatgagcaaaacctttatt
 E  N  F  E  D  V  G  H  S  T  D  A  R  E  M  S  K  T  F  I
atcggtgaactgcatcctgatgatcgtccgaaactgaataaaccgcctgaaaccctgatt
 I  G  E  L  H  P  D  D  R  P  K  L  N  K  P  P  E  T  L  I
accaccattgatagcagcagcagttggtggaccaattgggttattccggcaattagcgca
 T  T  I  D  S  S  S  S  W  W  T  N  W  V  I  P  A  I  S  A
gttgcagttgcactgatgtatcgtctgtatatggcagaagattaa
 V  A  V  A  L  M  Y  R  L  Y  M  A  E  D  -
```

A3. Codon optimised sequence encoding full length human Syb2

```
atgagcgcaaccgcagccaccgcaccgcctgcagcaccagccggtgaaggtggtcctccg
 M  S  A  T  A  A  T  A  P  P  A  A  P  A  G  E  G  G  P  P
gcaccgcctccgaatctgaccagcaatcgtcgtctgcagcagacccaggcacaggttgat
 A  P  P  P  N  L  T  S  N  R  R  L  Q  Q  T  Q  A  Q  V  D
gaagttgttgatattatgcgtgtgaacgtggataaagttctggaacgtgatcagaaactg
 E  V  V  D  I  M  R  V  N  V  D  K  V  L  E  R  D  Q  K  L
agcgaactggatgatcgtgcagatgcactgcaggcaggcgcaagccagtttgaaaccagc
 S  E  L  D  D  R  A  D  A  L  Q  A  G  A  S  Q  F  E  T  S
gcagcaaaactgaaacgtaaatattggtggaaaaacctgaaaatgatgattatcctgggt
 A  A  K  L  K  R  K  Y  W  W  K  N  L  K  M  M  I  I  L  G
gtgatttgcgccattattctgattatcatcatcgtgtattttagcacctaa
 V  I  C  A  I  I  L  I  I  I  I  V  Y  F  S  T  -
```

A4. Codon optimised sequence encoding human Rpn13 C-terminal domain

```
agtccgagtccggcaccgtcaagcggtaatggtgcaagcaccgcagcaagcccgacccag
 S  P  S  P  A  P  S  S  G  N  G  A  S  T  A  A  S  P  T  Q
ccgattcagctgagcgatctgcagagtattctggcaaccatgaatgttccggcaggccct
 P  I  Q  L  S  D  L  Q  S  I  L  A  T  M  N  V  P  A  G  P
gcgggtggtcagcaggttgatctggcaagcgttctgacaccggaaattatggcaccgatt
 A  G  G  Q  Q  V  D  L  A  S  V  L  T  P  E  I  M  A  P  I
ctggccaatgcagatgttcaagaacgtctgctgccgtatctgccgagcggtgaaagcctg
 L  A  N  A  D  V  Q  E  R  L  L  P  Y  L  P  S  G  E  S  L
ccgcagaccgcagatgaaattcagaataccctgaccagtccgcagtttcagcaggcactg
 P  Q  T  A  D  E  I  Q  N  T  L  T  S  P  Q  F  Q  Q  A  L
ggtatgtttagcgcagcactggccagcggtcagctgggtccgctgatgtgtcagtttggt
 G  M  F  S  A  A  L  A  S  G  Q  L  G  P  L  M  C  Q  F  G
ctgcctgcagaagcagttgaagcagcaaataaaggtgatgttgaagcatttgcaaaagcc
 L  P  A  E  A  V  E  A  A  N  K  G  D  V  E  A  F  A  K  A
atgcagaataatgcaaaaccggaacagaaagaaggcgacaccaaagataaaaaagacgaa
 M  Q  N  N  A  K  P  E  Q  K  E  G  D  T  K  D  K  K  D  E
gaggaagatatgagcctggactaa
 E  E  D  M  S  L  D  -
```

# Appendix B

## *B1. NMR theory*

Nuclear magnetic resonance (NMR) is a phenomenon that arises as a result of the reorientation of nuclear spins in an applied magnetic field (Campbell, 2012). Nuclei of certain isotopes have an intrinsic angular momentum referred to as 'spin'. The magnitude of spin angular momentum is given by

$$[I\,(I+1)]^{1/2}\,\hbar$$

and is quantized in units of $\hbar$ (= $h/2\pi$, $h$ is Planck's constant). The spin quantum number $I$ of a nucleus is determined by the number of unpaired protons and neutrons and can take one of the following values

$$I = 0,\ {}^{1}/_{2},\ 1,\ {}^{3}/_{2},\ 2,\ \ldots,$$

A nucleus of spin $I$ can have $2I + 1$ energy levels when placed in a static magnetic field. Many isotopes used in protein NMR spectroscopy have nuclei with $I = {}^{1}/_{2}$ and have only two energy levels. In the absence of a magnetic field all orientations of a spin $I$ nucleus have the same energy. In the presence of a magnetic field ($B_0$), a spin-${}^{1}/_{2}$ nucleus (${}^{1}H$ nucleus for example) can have only two orientations (along or against the direction of the static field) (Fig. B1.1A). These two states are separated by an energy $\Delta E$. The resonance condition occurs when the application of a wave of frequency ν causes nuclei to 'flip' from the lower energy level to the upper one, provided $\Delta E = h\nu$. The large number of spins present in a sample at equilibrium gives rise to bulk magnetization, a vector with magnitude M. The component of $M_z$ along the static field $B_0$ (Fig. B1.1B) is proportional to the population difference between the two energy levels (Hore, 1995; Campbell, 2012).

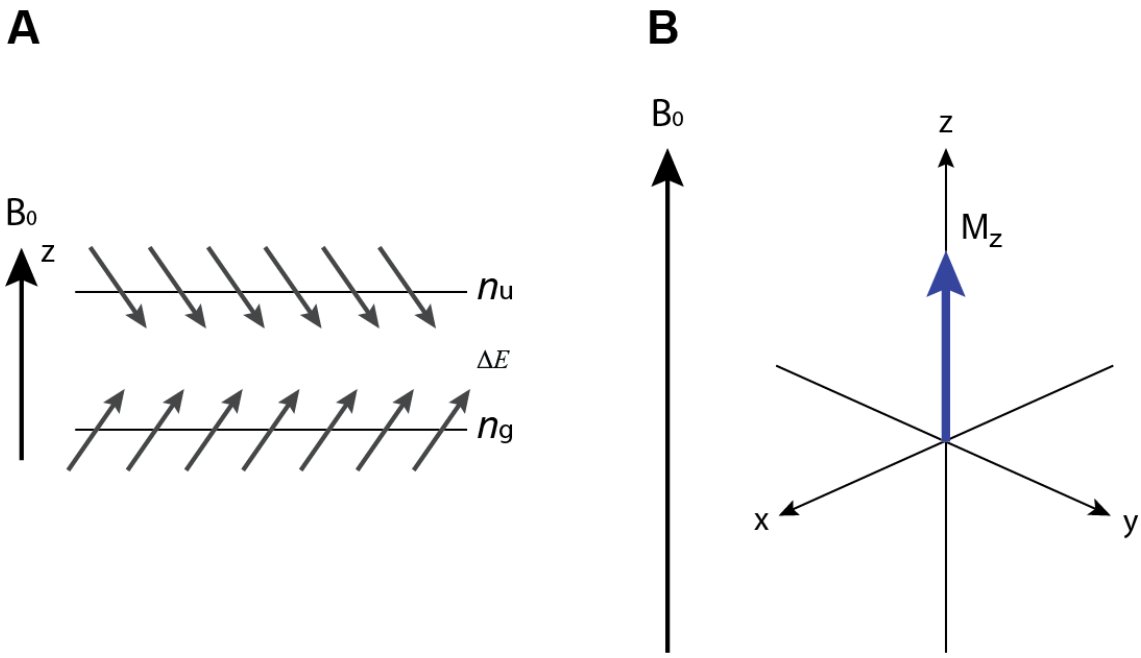



**Figure B1.1- Energy levels and bulk magnetization.** (A) Spin $^1/_2$ nuclei have two possible energy states when interacting with a magnetic field ($B_0$). At equilibrium, magnetic dipoles have a slight preference for the lower energy ground state ($n_g$) over the upper state ($n_u$) due to the Boltzmann distribution. (B) The bulk magnetization vector shown as a blue arrow along the z axis (Campbell, 2012; Keeler, 2005).

A nucleus with finite spin has a magnetic moment (μ) which can interact with an applied static field. μ is directly proportional to $I$, and is given by,

$$\mu = \gamma I$$

where, γ is the gyromagnetic ratio. The 'spin' of μ confers on it the property of angular momentum, which leads to precession (Fig. B1.2). The frequency of precession is given by,

$$\omega = 2\pi\nu = \gamma B_0$$

where ω is the Larmor frequency (angular frequency) and ν is the frequency in cycles per second (Hz). The gyromagnetic ratio varies for different nuclei and gives rise to range of frequencies where NMR signals can be detected (Table B1.1) (Hore, 1995).

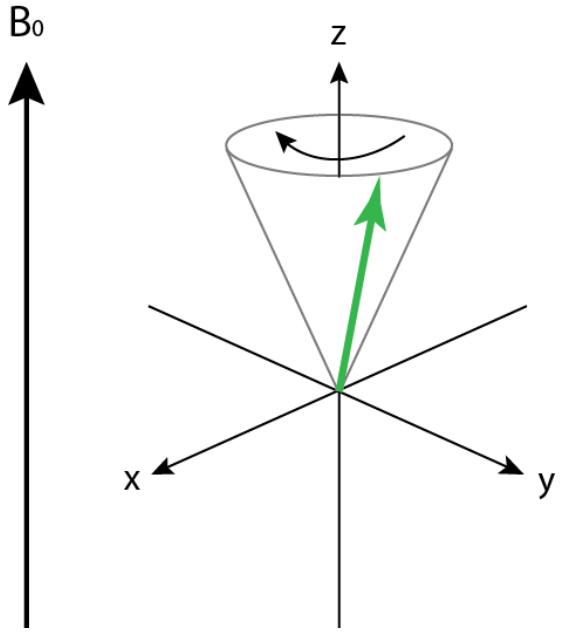



**Figure B1.2- Larmor precession.** The magnetic moment (μ) shown as a green arrow, rotates around the z-axis at a constant angle in the presence of $B_0$ (Keeler, 2005).

**Table B1.1- Spin quantum numbers, gyromagnetic ratios, NMR frequencies (at 9.4 T) and natural abundances of selected nuclei.** Adapted from Hore (1995).

| Nucleus | Spin quantum number (*I*) | Gyromagnetic ratio γ ($10^7 T^{-1}s^{-1}$) | Frequency (MHz) | Natural abundance (%) |
|---|---|---|---|---|
| $^{1}H$ | $^{1}/_{2}$ | 26.75 | 400.0 | 99.985 |
| $^{2}H$ | 1 | 4.11 | 61.4 | 0.015 |
| $^{13}C$ | $^{1}/_{2}$ | 6.73 | 100.6 | 1.108 |
| $^{14}N$ | 1 | 1.93 | 28.9 | 99.63 |
| $^{15}N$ | $-^{1}/_{2}$ | -2.71 | 40.5 | 0.37 |
| $^{19}F$ | $^{1}/_{2}$ | 25.18 | 376.5 | 100.0 |

The application of a radiofrequency pulse at a precise frequency perpendicular to $B_o$ will give rise to a field $B_1$, which in turn induces phase coherence in the *x-y* plane generating transverse magnetization ($M_{xy}$). In a similar manner, the application of different kinds of pulse sequences will effect the bulk equilibrium magnetization ($M_o$)

in ways that can be used to obtain information about the sample under investigation. The $B_1$ field causes $M_o$ to rotate at the Larmor frequency given by,

$$\omega = \gamma B_1$$

This precessing magnetisation gives rise to a transient signal called the free induction decay (FID), and is detected by a coil in the *x*-*y* plane. The FID is a time-domain signal, and can be converted into a spectrum of frequencies by Fourier transformation (Fig. B1.3) (Campbell, 2012; Keeler, 2005).

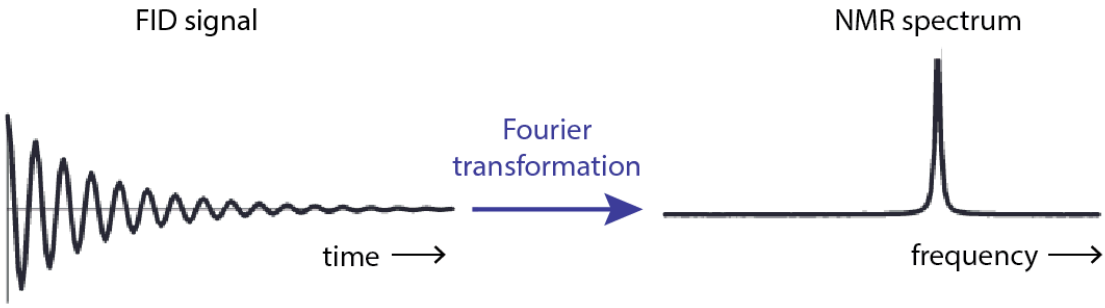



**Figure B1.3- Fourier transformation.** Fourier transformation is a mathematical process by which a time-domain signal such as a free induction decay (FID) can be converted into the frequency-domain, a distribution of frequencies that constitute an NMR spectrum. Figure adapted from Keeler (2005).

The NMR resonance frequency of a nucleus within a molecule is given by,

$$\nu = \frac{\gamma B}{2\pi}$$

This demonstrates that the NMR frequency of a nucleus depends on its gyromagnetic ratio and the field strength it experiences. However, resonances frequencies also depend upon the local electron distribution, which can reduce the effective field experienced by the nucleus. This effect is referred to as the 'chemical shift', and gives rise to separately

detectable NMR signals within a molecule (Hore, 1995; Campbell, 2012). The effective field experienced by a nucleus is therefore given by,

$$B = B_0 (1 - \sigma)$$

where σ is the shielding constant. Thus the actual NMR frequency of a nucleus becomes,

$$\nu = \frac{\gamma B_0 (1 - \sigma)}{2\pi}$$

In practice, however, absolute shifts are not used. Instead chemical shifts are defined in terms of the difference in resonance frequencies between a nucleus of interest and a reference nucleus, and is given by,

$$\delta = 10^6 \frac{(\nu - \nu_{ref})}{\nu_{ref}}$$

and δ is independent of the magnetic field. It is reported in parts per million, or ppm (Hore, 1995). Therefore, resonances can be resolved and assigned based on their chemical shifts, this provides means to understand molecular structure and follow interactions based on perturbations in chemical shifts observed upon binding.

# Appendix C

Assignments of backbone amide resonances corresponding to residues within the SGTA C-terminal domain. These assignments was carried out by Dr Santiago Martinez-Lumbreras at the Department of Chemistry, King's College London.

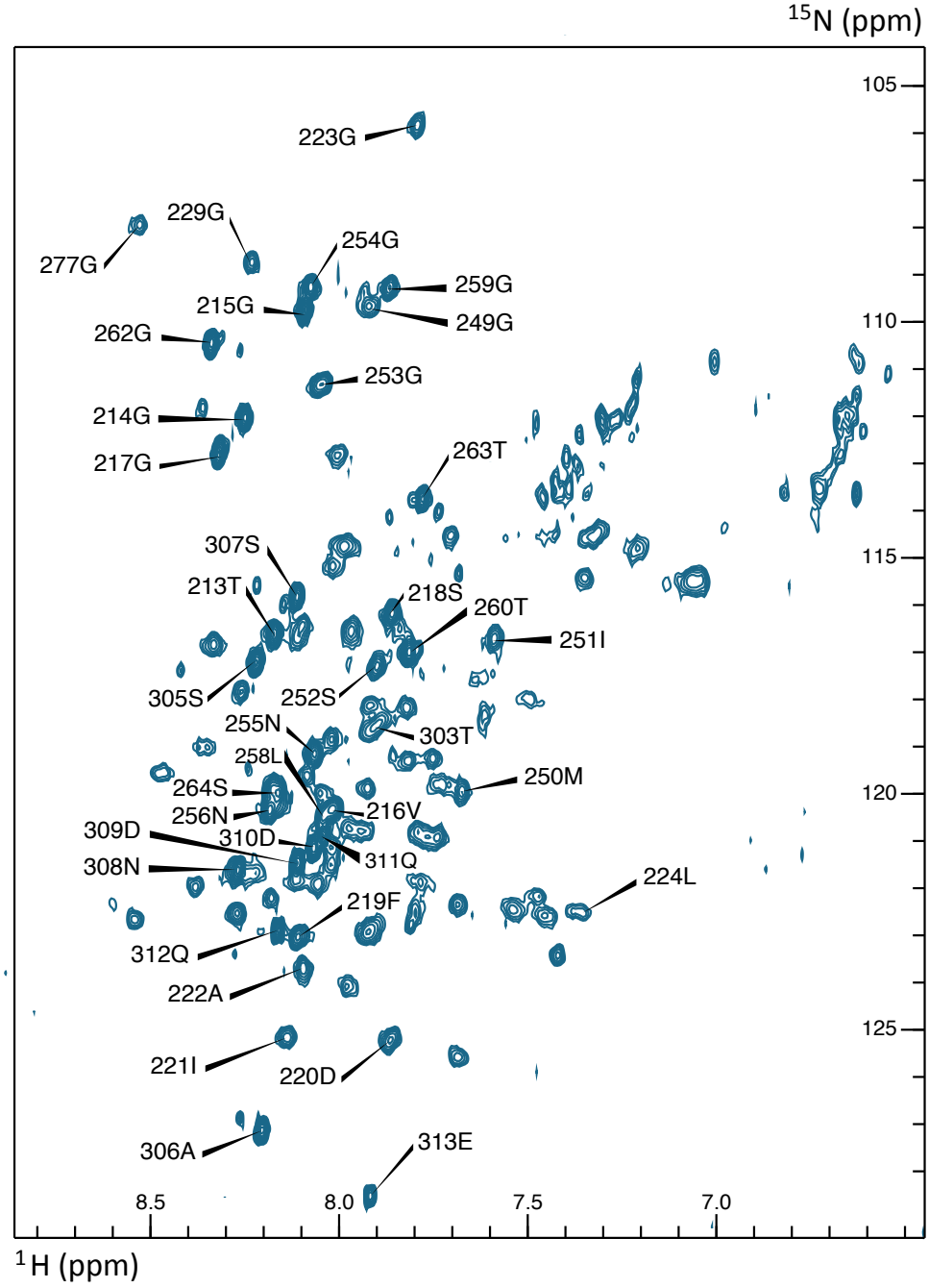



**Figure C1- $^{1}H$-$^{15}N$ HSQC of the C-terminal domain of SGTA (residues 213-313).**

## Appendix D

$^{15}$N relaxation experiments on SGTA constructs

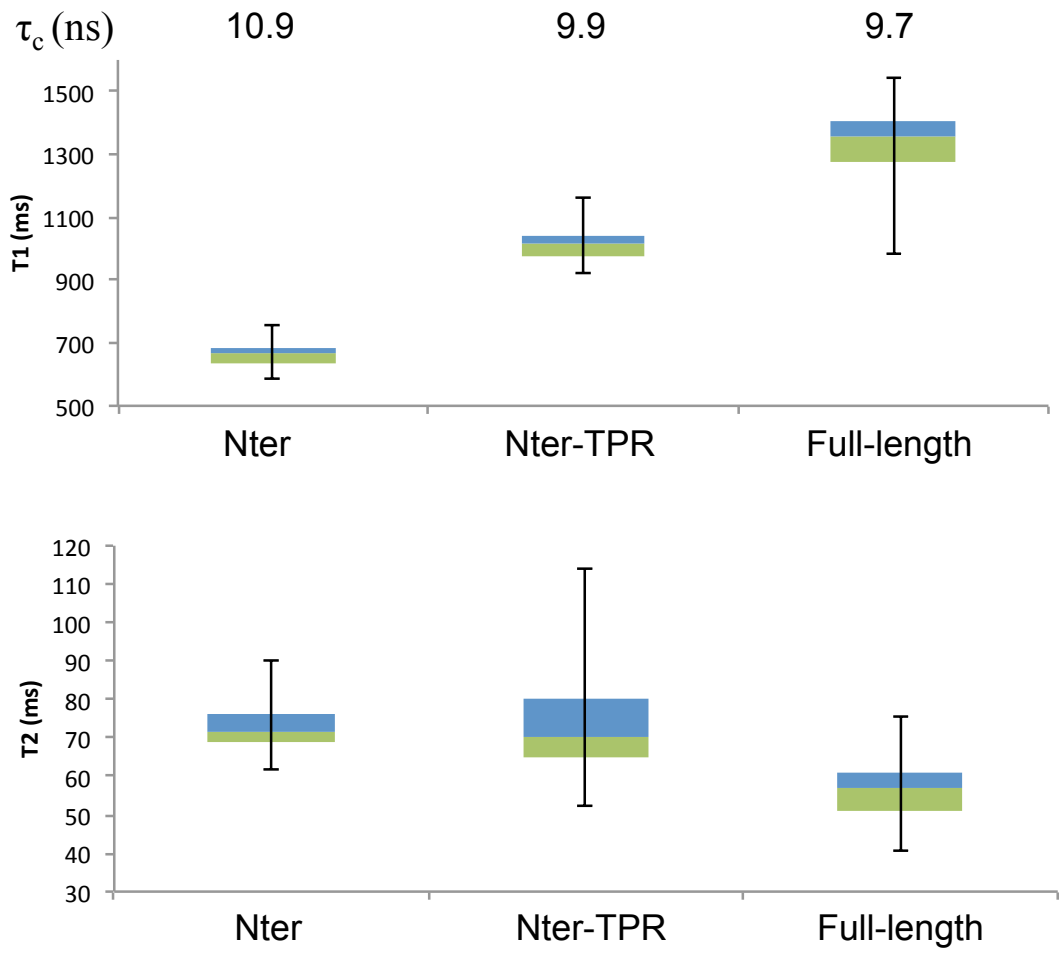



**Figure D1- N-terminal domain correlation times in the context of different SGTA constructs.** Correlation times ($\tau_c$) of the N-terminal domain (shown above) when present in the context of three different SGTA constructs (Nter, Nter-TPR and full-length) were estimated from the $T_1/T_2$ values using the following equation

$$\tau_c \approx \frac{1}{4\pi\upsilon_N}\sqrt{6\frac{T_1}{T_2}-7}$$

$^{15}$N relaxation experiments on the Nter, Nter-TPR and full-length SGTA constructs were acquired at spectrometer frequencies ($\upsilon_N$) of 500 MHz, 700 MHz and 950 MHz, respectively.

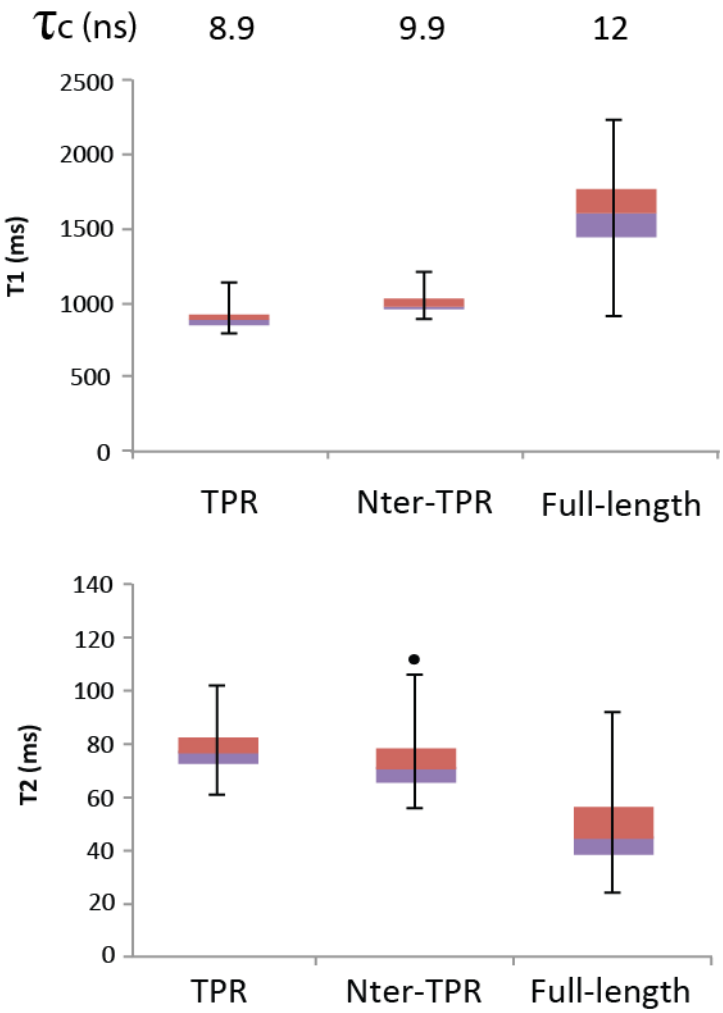



**Figure D2- TPR domain correlation times in the context of different SGTA constructs.** Correlation times ($\tau_c$) of the TPR domain (shown above) when present in the context of three different SGTA constructs (TPR, Nter-TPR and full-length) were estimated from the $T_1/T_2$ values using the following equation

$$\tau_c \approx \frac{1}{4\pi\upsilon_N}\sqrt{6\frac{T_1}{T_2}-7}$$

$^{15}$N relaxation experiments on the TPR, Nter-TPR and full-length SGTA constructs were acquired at spectrometer frequencies ($\upsilon_N$) of 700 MHz, 700 MHz and 950 MHz, respectively.