Towards a Structural Model of Adherent Knobs on the Surface of *Plasmodium falciparum* Infected Erythrocytes



Erin Eloise Cutts

Department of Biochemistry

University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford.

Wolfson College

Trinity 2016

I dedicate this thesis to my Mum, Ann Cutts, who taught me that anything is possible if you work hard. I do not think she was imagining a D.Phil from Oxford when she said this to me 20 or so years ago.

Acknowledgements

I would like to extend thanks to my teachers, past and present for helping to train me such that I could produce this thesis. Notably, my supervisors for this thesis, John, Phill and Mark, but also Keith Shearwin and Ian Dodd from Adelaide.

I would also like to thanks my family for making me who I am. My Mum most of all, for being the strongest person I know. But also my sister and brother, who always provided me with challenges growing up.

To everyone who has helped me in the lab, particularly George, and my malarial partners in crime, Jemma and Maddy. To those in and associated with SBCB; Jemma most of all, Phil F. for his coding help, but also Nat, G.B., S.B., Syma, Maria, Anna, Dave, Jo. And to my other Oxford based friends; Chris, Aoife and Robbie

Finally, thanks to the distractions that kept me sane; the films, the food, the drink and occasional excercise.

Publications

Work from this thesis is included in the following publications:

- **Cutts, E. E.**, Laasch, N, Reiter, D. M., Trenker, R., Slater, L. M., Stansfeld, P. J. and Vakonakis, I.,"Interactions of *P. falciparum* KAHRP and *Pf*EMP1 with spectrin suggest a model for adherent knobs on the infected erythrocyte surface", (2016) To be submitted.

- Oberli, A., Zurbrügg, L., Rusch, S., Brand, F., Butler, M. E., Day, J. L., **Cutts, E. E.**, Lavstsen, T., Vakonakis, I., and Beck, H.-P. (2016) "Plasmodium falciparum PHIST Proteins Contribute to Cytoadherence and Anchor PfEMP1 to the Host Cell Cytoskeleton". *Cell. Microbiol.* DOI:10.1111/cmi.12583

- Watermeyer, J. M., Hale, V. L., Hackett, F., Clare, D. K., **Cutts, E. E.**, Vakonakis, I., Fleck, R. A., Blackman, M. J., and Saibil, H. R. (2015) "A spiral scaffold underlies cytoadherent knobs in *Plasmodium falciparum*-infected erythrocytes." *Blood*. DOI:10.1182/blood-2015-10-674002

- Oberli, A., Slater, L. M., **Cutts, E.**, Brand, F., Mundwiler-Pachlatko, E., Rusch, S., Masik, M. F. G., Erat, M. C., Beck, H.-P., and Vakonakis, I. (2014) "A *Plasmodium falciparum* PHIST protein binds the virulence factor *Pf*EMP1 and comigrates to knobs on the host cell surface." *FASEB J.* 28, 4420–4433

The following publications were produced during my D.Phil. but do not include work from this thesis:

- **Cutts, E. E.**, Inglis, A., Stansfeld, P. J., Vakonakis, I., and Hatzopoulos, G. N. (2015) "The centriolar protein CPAP G-box: an amyloid fibril in a single domain." *Biochem. Soc. Trans.* 43, 838–843

- Wojdyla, J. A., **Cutts, E.**, Kaminska, R., Papadakos, G., Hopper, J. T. S., Stansfeld, P. J., Staunton, D., Robinson, C. V., and Kleanthous, C. (2015) "Structure and function of the Escherichia coli Tol-Pal stator protein TolR." *J. Biol. Chem.* 290, 26675–26687

- Hatzopoulos, G. N., Erat, M. C., **Cutts, E.**, Rogala, K. B., Slater, L. M., Stansfeld, P. J., and Vakonakis, I. (2013) "Structural analysis of the G-box domain of the microcephaly protein CPAP suggests a role in centriole architecture." *Structure.* 21, 2069–2077

Abstract

The *Plasmodium falciparum* proteome is enriched in intrinsically disordered proteins, three of which are directly linked to malaria severity; *P. falciparum* Erythrocyte Membrane Protein 1 (*Pf*EMP1), Knob-associated Histidine-rich Protein (KAHRP) and *Plasmodium* Helical Interdispersed Sub-telomeric family member (PHIST) PFE1605w. *Pf*EMP1, KAHRP and PFE1605w all localised to protrusions (knobs) on the *P. falciparum*-infected erythrocyte surface, which enable pathogenic cytoadherence of infected erythrocytes to uninfected erythrocytes and endothelial cells. In this thesis, I have employed a combination of biophysical and computational methods to characterize the interactions of *Pf*EMP1 and KAHRP with host erythrocyte spectrin, and of *Pf*EMP1 with two PHIST family proteins, PFI1780w and PFE1605w. These interactions are likely to assist knob formation and, thus, promote cytoadherence. I derived atomistic structural models of complexes formed between the PHIST domain of PFI1780w and the disordered C-terminal region of the cytosolic domain of *Pf*EMP1, of the folded core of the *Pf*EMP1 cytosolic domain and domain 17 of the host erythrocyte α spectrin, and between the KAHRP 5´ sequence repeat and domains 10-14 of β spectrin. Combining these structural models with previous electron microscopy information allows us to propose a detailed model as to how knobs form and how *Pf*EMP1, KAHRP, PFE1605w, and other possible factors could locate therein.

List of Abbreviations

ATS	Acidic Terminal Sequence (of *Pf*EMP1)
CD	Circular Dichroism
CS-Rosetta	Chemical Shift Rosetta
EM	Electron Microscopy
FP	Fluorescence Polarisation
HSQC	Heteronuclear single quantum coherence
IDP	Intrinsically disordered protein
KAHRP	Knob Associated Histadine Rich Protein
MC	Maurer's Cleft
MD	Molecular Dynamics
NMR	Nuclear Magnetic Reasonance
*Pf*EMP1	*Plamsodium falciparum* Erythrocyte Membrane Protein 1
PHIST	*Plasmodium* Helical Interdispersed Sub-telomeric
PRE	Paramagnetic relaxation enhancement
PV	Parasitic Vacuole
RMSD	Root-mean-square difference
RMSF	Root-mean-square fluctuation
SAXS	Small angle X-ray scattering
SEC-MALS	Size exclusion chromatography multi-angle light scattering
TVN	Tubulovesicular Network

Table of contents

Abstract **v**

Table of contents **vii**

List of figures **xiii**

List of tables **xvi**

1 Introduction **1**

1.1 Malaria 1

1.1.1 Life cycle 1

1.1.2 The Blood Stage 2

1.2 Remodelling 4

1.2.1 Protein Export 4

1.2.2 Adhesion 8

1.2.3 *Pf*EMP1 8

1.2.4 KAHRP 10

1.2.5 PHIST Proteins 11

1.2.6 Adhesion in other *Plasmodium* Species 12

1.3 IDPs in Malaria 13

1.4 Knob Structures 14

1.4.1 Erythrocyte Cytoskeleton and Knob Structures 14
1.4.2 Cytoskeleton Structure 15
1.4.3 Regulation of Cytoskeleton 17
1.4.4 Cytoskeleton Interactions and Modifications in *P. falciparum* 18
1.5 Current View of Knob Structure 20
1.6 Work Presented 21

2 Methods **23**
2.1 Protein Purification 23
2.1.1 *Pf*EMP1 23
2.1.2 PHIST Proteins 23
2.1.3 KAHRP 24
2.1.4 Cytoskeletal Components 25
2.1.5 Cysteine Protein Labelling 26
2.2 NMR 27
2.2.1 HSQC Titrations 27
2.2.2 Paramagnetic Relaxation Enhancement 28
2.2.3 Backbone Assignment 29
2.3 Electron Microscopy 31
2.4 Biophysical Assays and Methods 31
2.4.1 Mass Spectroscopy 31
2.4.2 CD 32
2.4.3 FP Assays 32
2.4.4 SAXS 33
2.5 X-ray Crystallography 33
2.5.1 Structure Determination of α Spectrin Repeats 17 to 19 34
2.5.2 Structure Determination of α Spectrin Repeats 16 and 17 34

2.5.3 Structure Determination of α Spectrin Repeat 17 35
2.6 Computational Methods 35
2.6.1 Homology Modelling 35
2.6.2 CS Rosetta 36
2.6.3 HADDOCK Docking 36
2.6.4 MD Parameter Description 36
2.6.5 MD Run Parameters, Software & Analysis 43

3 *Pf*EMP1-PHIST 45
3.1 Preliminary and Contributions to Data 46
3.2 Localising the ATS-C/PFI1780w Interaction 47
3.3 Modelling and MD of PFI1780w 50
3.4 Alanine Substitution Analysis 54
3.5 Paramagnetic Relaxation Enhancement 57
3.6 Computational Modelling of the Complex 59
3.6.1 Generation of CS-Rosetta Models 59
3.6.2 MD Simulations of CS-Rosetta Models 61
3.6.3 MD Simulations of Model 247 Docked to PFI1780w Surface 65
3.7 General ATS-PHIST Interactions 68
3.8 Interaction Between PFE1605w and Band 3 70
3.9 Discussion 71
3.9.1 MD Simulation 71
3.9.2 A Mutation Resistant Interface 73
3.9.3 PHIST Protein Function 73

4 *Pf*EMP1-Spectrin 75
4.1 Preliminary data 75

4.2 *Pf*EMP1 binds spectrin . . . 76
4.2.1 Localisation of binding . . . 76
4.2.2 N-terminal ATS truncations . . . 80
4.2.3 Localisation of ATS on spectrin with EM . . . 82
4.3 Structural Characterisation . . . 83
4.3.1 CD . . . 83
4.3.2 Crystal Structure of α17-19 . . . 84
4.3.3 Crystal structure of α16-17 . . . 85
4.3.4 Crystal structure of α17 . . . 86
4.3.5 SAXS . . . 89
4.4 Sequence Variants . . . 90
4.5 Interaction Model of *Pf*EMP1 ATS-Core and α17 . . . 91
4.5.1 NMR Shift Perturbation Mapping . . . 91
4.5.2 Complex Docking with HADDOCK . . . 95
4.5.3 MD . . . 97
4.5.4 Model Validation . . . 100
4.6 Discussion . . . 103
4.6.1 Spectrin Structure and Implications . . . 103
4.6.2 Disordered Regions Role in Binding . . . 103
4.6.3 *Pf*EMP1 Localisation . . . 105

5 KAHRP-Spectrin 107

5.1 Contributions to Data . . . 107
5.2 Characterising KAHRP-Cytoskeletal Interaction . . . 108
5.2.1 KAHRP Binds β Spectrin Repeats 10-14 . . . 108
5.2.2 Truncations of KAHRP and β10-14 Reduce Affinity . . . 111
5.2.3 KAHRP Binds Ankyrin Spectrin-Binding Domain via K1D . . . 113

5.3 Complementary Electrostatics Binding Model 115
5.3.1 KAHRP-Spectrin Interaction is Dominated by Electrostatics 115
5.3.2 Analysis of Surface Electrostatics 115
5.3.3 Complementary Path Finding Algorithm 117
5.3.4 Charge Complementation Scoring 119
5.3.5 Analysis of Truncations 121
5.3.6 Charge Distribution is Required for Affinity 125
5.4 Atomistic Modelling of KAHRP K2-4 and β10-14 Spectrin 125
5.4.1 Model Generation 125
5.4.2 Restrained MD Simulations 127
5.4.3 Unrestrained MD Simulations 131
5.5 Discussion 133
5.5.1 The Docking Algorithm 133
5.5.2 MD Simulations 135
5.5.3 Model of KAHRP Binding the Cytoskeleton 137

6 Discussion **139**

6.1 Disordered Proteins 139
6.1.1 Protective to Mutation 139
6.1.2 Energetic Considerations 140
6.1.3 Multivalent Binding 142
6.1.4 Redundancy in Binding 143
6.1.5 Larger Assemblies 143
6.1.6 Additional Roles for Disordered Regions in Binding 144
6.2 The Knob Model 145
6.2.1 Cytoskeletal Interactions 145
6.2.2 Assembly and Timing 147

6.2.3 PHIST Proteins 148
6.2.4 Spiral Scaffold 149
6.2.5 Regulation 150
6.2.6 Blood Disorders 151
6.3 Outlook and Future Work 152
6.3.1 Strain Specific Factors 152
6.3.2 Applications 152
6.3.3 Reconstitution 153

Conclusion **154**

References **156**

Appendix A Constructs **181**

Appendix B HADDOCK **185**

B.1 Active residues used for HADDOCK docking: 185
B.1.1 Chapter 3: PFI1780w/ATS-C 185
B.1.2 Chapter 4: α17/ATS-Core 185

Appendix C Spectrin NMR Overlays **186**

Appendix D Thermostability by CD **199**

Appendix E KAHRP docking Scripts **201**

Appendix F KAHRP/Spectrin Docking **208**

F.1 β10-14 spectrin charge cluster residues 209

List of figures

1.1 The life cycle of *Plasmodium* 3

1.2 Remodelling of infected erythrocyte 5

1.3 Protein export pathways 7

1.4 NMR structure of the ATS of *Pf*EMP1 14

1.5 Simplified view of the eythrocyte cytoskeleton 17

1.6 Current view of knob structure 21

2.1 HSQC pulse sequence 28

2.2 NMR sequential assignment method 30

2.3 Bonded interactions 37

3.1 *Pf*EMP1 ATS binds PFI1780w. 47

3.2 Optimisation of ATS/ PFI1780w FP binding assay. 48

3.3 HSQC spectra of PFI1708w PHIST domain 49

3.4 The binding affinity of ATS-C truncations, determined with NMR ^{1}H-^{15}N HSQC titrations. 50

3.5 PFI1780w is a monomer is solution although crystal structure suggests a dimer. 51

3.6 Molecular Dynamic simulation of the monomeric model of PFI1780w using three different atomic force fields. 53

3.7 Comparison between monomer and dimer of PFI1780w during MD. 54
3.8 Alanine substitution mutants 56
3.9 PRE experiments 58
3.10 CS-Rosetta structure prediction of the ATS-C. 62
3.11 Molecular Dynamics of CS-Rosetta Models. 63
3.12 Molecular Dynamics of CS-Rosetta. 64
3.13 ATS-C docked to PFI1780w PHIST domain. 66
3.14 ATS-C docked to PFI1780w MD analysis. 67
3.15 ATS-C binding to PFE1605w. 69
3.16 ATS-C variant binding to PHIST proteins PFI1780w and PFE1605w. 70
3.17 PFE1605w binds band 3 dimers. 71

4.1 *Pf*EMP1 ATS binds spectrin 77
4.2 *Pf*EMP1 ATS binds spectrin 78
4.3 NMR localisation of binding to ATS-Core and α17 79
4.4 FP binding assay localising binding to ATS-Core and α17 80
4.5 The N-terminal region of the ATS increases binding effects in NMR 81
4.6 EM suggests binding of ATS-N-Core to spectrin 83
4.7 CD of α17-C sub-fragments 84
4.8 Crystal structure of α17-19 85
4.9 Crystal structure of α16-17 87
4.10 Crystal structure of α17 88
4.11 Overlay of crystallographic structures of α17 89
4.12 SAXS data of α16-17 and α17 90
4.13 *Pf*EMP1 sequence variants bind to α17 91
4.14 HSQC shift perturbation of α17 and ATS-Core 93
4.15 Shift perturbations mapping 94

4.16 HADDOCK complex structure prediction 96
4.17 MD on HADDOCK complex 98
4.18 Cluster 1 MD compared to non-docked strutures 99
4.19 MD structural interfaces 100
4.20 Model validation 102
4.21 ATS cytoskeleton interaction model. 105

5.1 KAHRP-spectrin interaction 109
5.2 KAHRP-spectrin NMR 110
5.3 FP assays of KAHRP and spectrin truncations 112
5.4 KAHRP-ankyrin interaction 114
5.5 KAHRP and β10-14 have complementary electrostatics 116
5.6 Path finding 117
5.7 Scoring of paths 120
5.8 Lowest scoring paths over spectrin are similar 122
5.9 Truncation validation 124
5.10 K2-4 sequence scrambles 126
5.11 MD of KAHRP bound to β10-14 spectrin 128
5.12 KAHRP-spectrin structure after MD 129
5.13 KAHRP dynamics in unrestrained versus restrained MD simulation 130
5.14 MD simulations suggest β10-14 spectrin is more dynamic when KAHRP is bound 132
5.15 KAHRP cytoskeleton interaction model 137

6.1 Model for knob structure and assembly 146

D.1 Thermostability of ATS-Core and α17 subsitution mutants 200

List of tables

A.1 Constructs used in experiments 182

A.2 Constructs used in experiments (cont.) 183

A.3 Constructs used in experiments (cont.) 184

F.2 KAHRP K2 Sequences 210

Chapter 1

Introduction

"I see you quiver with antici. . . "

Dr Frank N. Futer,

The Rocky Horror Picture Show

1.1 Malaria

Malaria is a parasitic infection that killed ~1.24 million individuals in 2010, with the majority of deaths being children under the age of 5 [1]. *Plasmodium falciparum* is the cause of the most lethal form of malaria, closely followed by *Plasmodium vivax* [2]. While incidents of death are on the decline due to implementation of insecticides, mosquito nets and anti-malaria agents; drug and chemical resistance is on the rise [3]. Hence, understanding the mechanisms causing severity of *P. falciparum* infection could be key to a new anti-malarial agents and parasite eradication.

1.1.1 Life cycle

The life cycle of *P. falciparum* has 3 distinct stages; the mosquito stage, the liver stage and the blood stage (Fig 1.1A). The mosquito stage begins after a mosquito feeds on a host

infected with *P. falciparum*. If the mosquito ingests the gametocyte forms of the parasite, the gametocytes are able to mate inside the gut of the mosquito and then undergo meiosis. This triggers changes in the parasites enabling them to migrate through the midgut wall of the mosquito to the salivary gland and form oocysts. Motile forms of the parasite, termed sporozoites, develop inside the oocyte, which eventually lyse, and sporozoites are able to be transferred from the mosquito saliva into a human host. The liver stage of development begins after a mosquito bites a human host, injecting sporozoites into the blood stream, where they travel to the liver, and infect hepatocytes. The parasite then replicates and differentiates again to produce merozoites, which are released into the blood stream, to begin the blood stage (Fig 1.1) [4].

1.1.2 The Blood Stage

The blood stage begins when a merozoite binds and invades an erythrocyte. The invasion process itself is complex, comprising of three stages; non-specific binding, re-orientation and entry, with each stage using a set of redundant proteins (for review see [6]). Shortly after entry, the parasite establishes itself in the ring stage, so named as it is observable with microscopy as a ring inside the cell. After 24 hours, this progresses to the trophozoite stage, with the host erythrocyte being remodelled by the parasite to promote growth and to protect the parasite from the host immune system. The schizont stage occurs after 36-48 hours, when the next generation of merozoite have matured and the host cell lyses to release 5-20 parasites [7], enabling another cycle of blood stage replication [5]. A small subset (<10%) of schizonts will commit to sexual development, resulting in the formation of gametocytes on the next round of erythrocyte infection, enabling transmission to mosquitos [8].

As the blood stage results in the destruction of erythrocytes and the presence of immunogens in the blood stream, it causes most of the physiological side effect, such as anaemia, respiratory distress and fever [9]. But in addition to these effects, *P. falciparum* is also

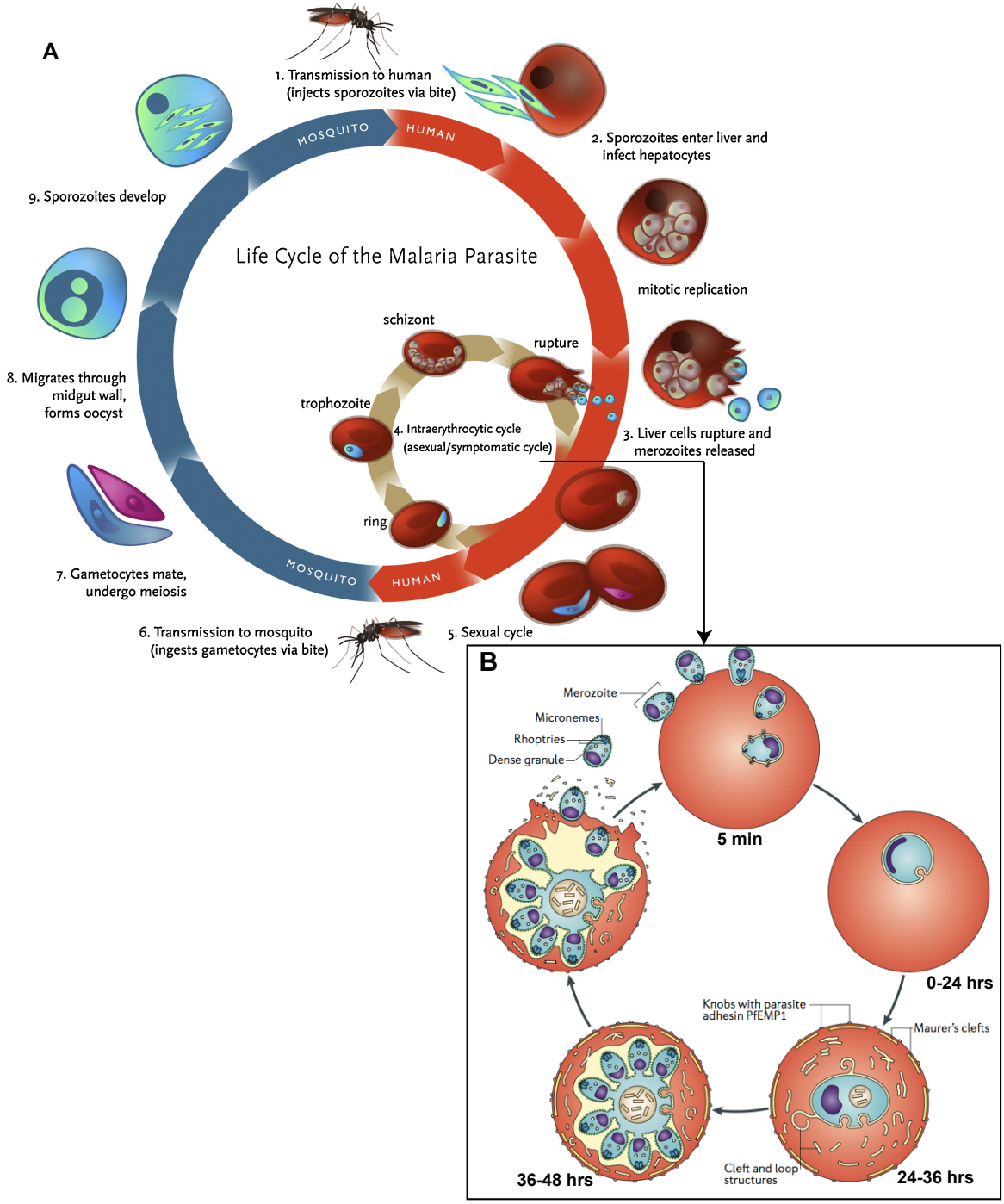


Fig. 1.1: The lifecycle of *Plasmodium*. A: The whole life cycle of *Plasmodium*. Steps 2 and 3 are in the liver, 3-5 in the blood, and 6-1 in the mosquito. Adapted from Klein *et al.* [4]. **B**: The blood stage of *Plasmodium* development, indicating the timing of morphology changes. Adapted from Haldar *et al.* [5].

able to remodel the host erythrocyte such that it adheres to endothelial cells, which line the blood vessel wall, and to other erythrocytes, both infected and uninfected. Adhesion aids parasite growth by preventing the infected erythrocyte from travelling to the spleen, where it would be detected as infected due to changes in deformability properties [10]. Hence, adhesion results in an increase in parasite load on the host. This also causes other physiological symptoms, as accumulation of erythroytes can block the microvasculature. When this occurs in the brain in the case of cerebral malaria, it causes coma and eventually death, making remodelling and adhesion of the host erythrocyte key mechanisms that result in the severity of *P. falciparum* infection.

1.2 Remodelling

P. falciparum extensively modifies the host cell in order to harvest nutrients, proliferate and survive. This results in a change in permeability of host erythrocyte, establishment of new membranous organelles, and the formation of knobs on the surface of the host erythrocyte, all of which help give rise to the symptoms of malarial infection. As this thesis focuses on proteins important for adhesion, the introduction to remodelling will focus on protein export and proteins that play a role in adhesion.

1.2.1 Protein Export

Protein export in blood stage development is still very poorly understood. This is partly due to the many barriers that must be traversed by any protein being exported from the parasite to the host erythrocyte. When a merozoite enters an erythrocyte, not only does it have its own membrane, it also invaginates the host erythrocyte membrane, resulting in a new membrane compartment, referred to as the parasitic vacuole (PV). This membrane forms thin tubular structures in the erythrocyte cytoplasm, referred to as the tubulovesicular network (TVN)

[11] (Fig. 1.2). In order for the parasite to modify the host cell with its own proteins, these proteins must pass from the parasite through the parasite membrane and parasitophorous vacuole membrane (PVM) into the host cytoplasm or to the host membrane.

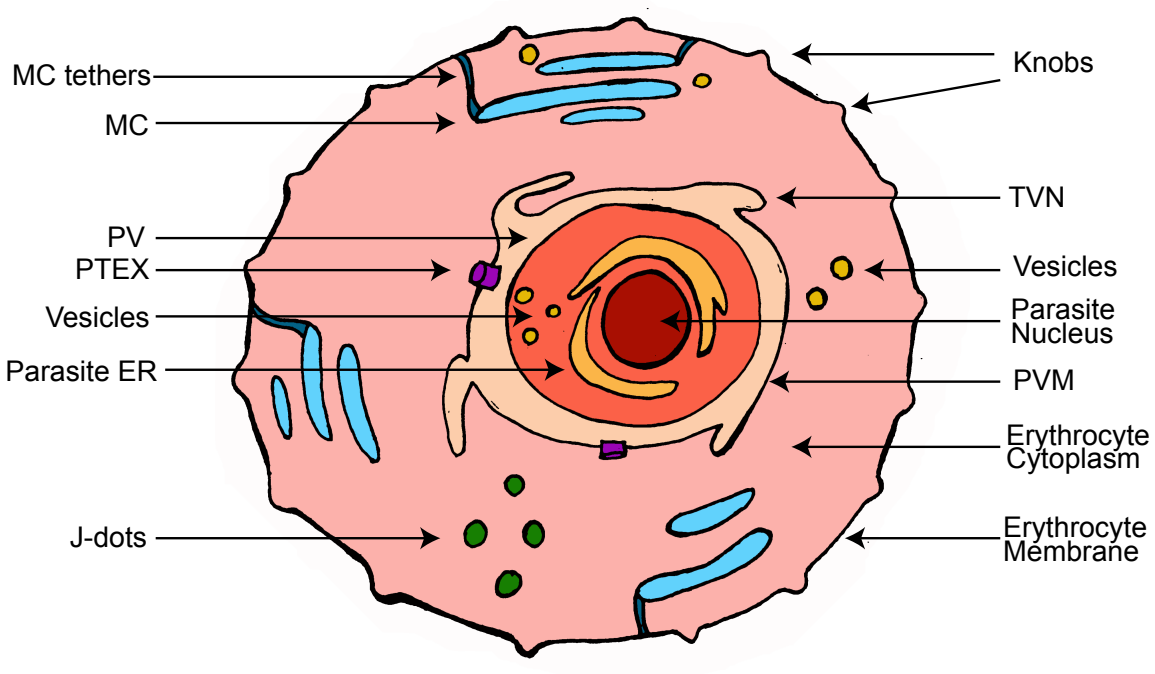


Fig. 1.2: Remodelling of infected erythrocyte. A schematic showing the parasite derived changes to the host erythrocyte at the trophozite stage. The parasite is shown in dark orange. TVN: tubulovesicular network, PV: parasitic vacuole, PVM: parasitophorous vacuole membrane, MC: Maurer's cleft, PTEX: *Plasmodium* translocon of exported proteins, ER: endoplasmic reticulum.

Proteins are produced in the parasite, where they are targeted to the parasite endoplasmic reticulum (ER) via an N-terminal transmembrane hydrophobic sequence. The ER targeting sequence is followed by a pentameric regulatory sequence, RxLxE/Q/D, referred to as *Plasmodium* export element (PEXEL) [12]. This sequence is cleaved by plasmepsin V [13, 14] following the conserved leucine and the new N-terminus acetylated in the ER [15] (Fig. 1.3). From the ER, proteins are transported to the PV, then into the erythrocyte cytoplasm by the *Plasmodium* translocon of exported proteins (PTEX), which spans the PVM [16].

Transport by PTEX is also controlled by the PEXEL motif, although this is poorly understood. Proteins without PEXEL motifs can also be exported, and these PEXEL-negative exported proteins (PNEPs) also rely on PTEX for export out of the PV. PTEX is a large multiple protein assembly, of which 5 subunits have been identified: PTEX150, Heat shock protein 101 (HSP101), Exported protein 2 (EXP2), PTEX88, and thioredoxin 2 (TRX2) [17]. EXP2 forms a pore in the membrane, HSP101 forms a hexameric ring and hydrolyses ATP to unfold protein prior to translocation. Function of other components is not well understood, but both HSP101 and PTEX150 have been shown to be essential for export, while PTEX88 and TRX2 have auxiliary roles, potentially acting as chaperones [18].

In addition to proteins being trafficked from the parasite to the cytoplasm of the erythrocyte, many exported proteins traverse through a parasite derived organelle in the erythrocyte cytoplasm called the Maurer's cleft (MC). This organelle is a dynamic, extended membrane network found beneath the erythrocyte membrane, believed to be formed from the PVM and the TVN budding off [19]. Although descriptions of the MC differ, all include thin tubular structures with surrounding vesicles [20] (Fig. 1.2). These structures do not seem to be linked to each other, as photobleaching experiments resulted in no recovery, suggesting no lateral diffusion with other membrane clefts or the PVM [21]. Linkages between the MC and the host erythrocyte cytoskeleton, and in many cases, knobs, have been observed and are hypothesised to be mediated by actin remodelling [20] or by the parasite, membrane-associated histidine-rich protein-2 (MAHRP2) [22]. Trafficking associated with the MC is sill poorly understood, but MC resident proteins, membrane-associated histidine-rich protein-1 (MAHRP1), Ring exported protein 1 (REX1), skeleton binding protein 1 (SBP1), *Plasmodium*-trafficking protein 1 (PTP1) and *Plasmodium*-trafficking protein 2 (PTP2) have been shown essential for trafficking, with knockdowns resulting in accumulation of protein within the parasite [23–26]. MC also have other regulatory roles, and have been linked to egress pathways [27] and oxidative stress [28].

Exported proteins have also been localised in punctate structures within the host cytosol, referred to as J-dots. J-dots are resistant to staining with BODIPY-ceramide, hence have either a low lipid content, or are composed of a different lipid type than MC. They are highly dynamic, are distinct from the PVM, TVN and the MC and are identified by the presence of Hsp40 chaperone [29]. The role they play in trafficking is still to be determined, but key virulence factors and chaperones have been localised within J-dots [29, 30], suggesting an important role for down stream protein processing.

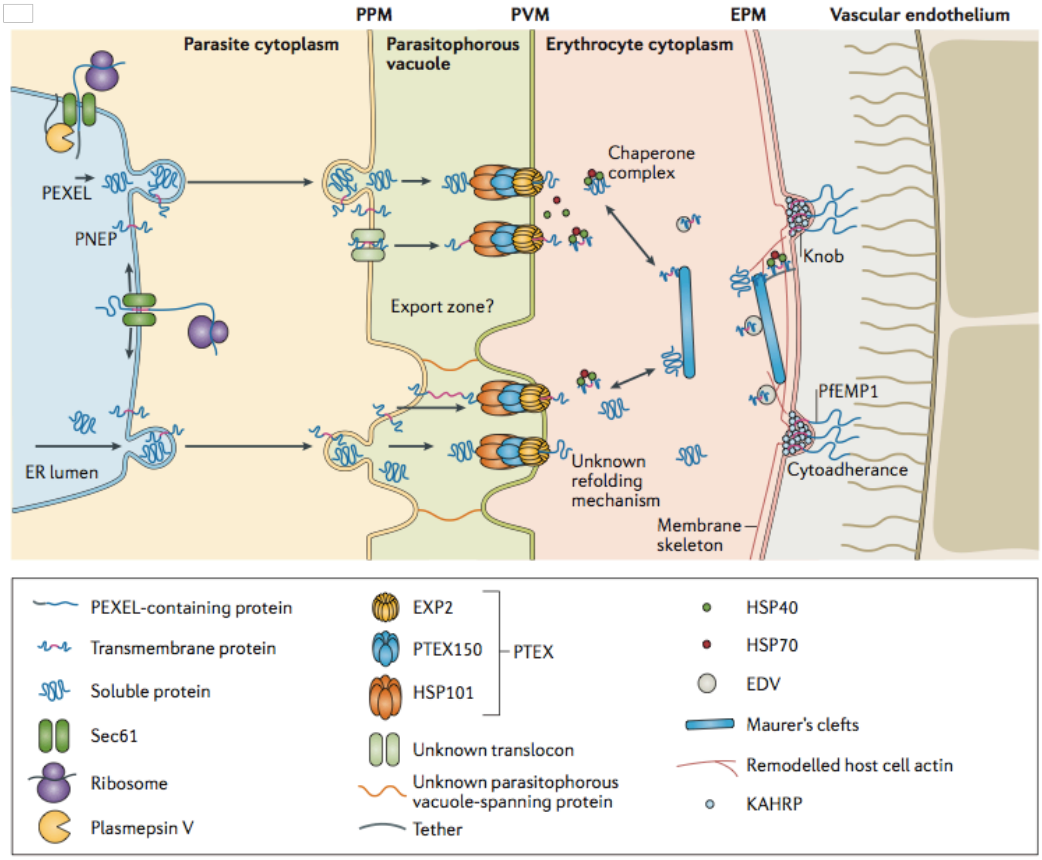


Fig. 1.3: Protein export pathways. Parasite proteins are targeted for export by PEXEL motifs or some other mechanism, exported into the lumen of the parasite, into the PV, then into the cytoplasm of the erythrocyte via PTEX. Protein transport from then is poorly understood, but vesicles and complexes of chaperones, making J-dots, are observed, and some proteins traverse through the MC. Figure adapted from review by de Koning-Ward *et al.* [31].

1.2.2 Adhesion

Proteins exported from the parasite result in the formation of dome-shaped protrusions on the infected erythrocyte surface, known as knobs, which result in adhesion to the host endothelium or other erythrocytes [32, 33]. Adhesion is mediated by proteins from the *Plasmodium falciparum* Erythrocyte Membrane Protein 1 family (*Pf*EMP1) [34]. *Pf*EMP1 is a trans-membrane protein, with a large, highly diverse ectodomain, a transmembrane helix and a conserved disordered intracellular region referred to as the acidic terminal sequence (ATS) [35]. *Pf*EMP1 molecules cluster at the apex of knobs, and while adhesion is still possible without knobs, the cyto-adhesion of knob-less parasitized cells under physiological flow conditions is greatly reduced [36–38]. Hence, understanding how knobs form and function could be key in preventing cytoadhesion. Other proteins found to localise to knobs are knob associated histidine rich protein (KAHRP), which is essential for knob formation [37, 38], and *Plasmodium* helical interdispersed sub-telomeric (PHIST) family member, PFE1605w [39, 40].

1.2.3 *Pf*EMP1

*Pf*EMP1 family members are the product of *var* genes. There are differing numbers of *var* genes identified in different field isolates of *P. falciparum*, with 61 *vars* genes identified in 3D7, but fewer identified in other isolates, such as 53 in HB3, and 48 for IT3 [41]. Each *var* gene is composed of 2 exons, with the ectodomain and the transmembrane helix being the product of exon 1, and intracellular ATS from exon 2. The majority of *var* genes are localised sub-teleomerically, and it is thought that lining up of flanking regions 5' to *var* genes promotes recombination in different groups of *var* genes [42]. Despite recombination events, the ATS remains well conserved [43]. The ectodomain however has very diverse sequence, with variable numbers and isoforms of Duffy binding-like domains (DBL), one or two cysteine-rich, inter-domain regions (CIDR) [34], which the parasite has been able to

adapt to adhere to different receptors, such as Intercellular Adhesion Molecule 1 (ICAM-1) [44], chondroitin sulfate A (CSA) [45] and cluster of differentiation 36 (CD36) [46], amongst other receptors. *Pf*EMP1 expression results in an immune response, and has been the focus of many vaccine based studies [47]. However, *var* gene expression is tightly regulated, such that only one *var* gene variant is expressed at any one time. Together with the large number of variants, and recombination events, the *Pf*EMP1 epitope can easily be changed in order to evade the immune response.

The expression of *var* genes is epigenetically silenced by default; with only one *var* gene being marked for activation on erythrocyte infection. This is achieved through histone modifications, with knockouts of specific histone deacetylases resulting in de-repression of *var* genes [48, 49]. Inactive *var* genes are also co-localised spatially at the edge of the nucleus [50]. Each *var* gene promoter also produces a silencing factor to further ensure monoallelic expression [51]. Hence, *var* gene switching requires removal of sliencing, addition and removal of activation histone markers, and topological changes in DNA location. The mechanism for variant switching is still poorly understood. Switching is observed in cultured parasite, but a higher rate is seen in human infection [52] implying the presence of intrinsic switching, as well as in response to environmental factors.

*Pf*EMP1 lacks a classical PEXEL sequence, but rather has an N-terminal sequence that includes the pentameric sequence, KxLxD, which targets it to the ER, but can not be cleaved by plasmepsin V [53]. *Pf*EMP1 trafficking is dependent on PTEX, with HSP101 knock-downs resulting in reduction of *Pf*EMP1 on the surface and *Pf*EMP1 being trapped inside the parasite [16]. *Pf*EMP1 is trafficked to the MC, where the ectodomain is hidden within the clefts, while the ATS is exposed to the host cytoplasm [54]. It is transported from the MC to the host erythrocyte membrane, most likely via some form of vesicle transport. This final stage is also directed by the N-terminal sequence of *Pf*EMP1, as removal of this region results in cytoplamsic localisation [55]. *Pf*EMP1 has also been localised to J-dots, hence

it is possible that these structures facilitate transport at some stage. Multiple proteins have been shown to be important for correct trafficking of *Pf*EMP1 [56]. These include MC proteins; MAHRP1, REX1, SBP1, PTP1 and PTP2, as well as other trafficking proteins; PTP3, PTP4, PTP5 and PTP6 [31, 56].

1.2.4 KAHRP

KAHRP is exported into the host erythrocyte with the standard PEXEL export signal sequence, where it localises to the periphery of the cell. The sequence of KAHRP contains three notable regions; a histidine rich region (residues 61-116), and two repetitive sequence regions, termed the 5' and 3' repeats (residues 356-456 and 540-600) [57]. Early studies (when KAHRP was still referred to as Histidine-rich protein 1) localised KAHRP to knob structures, attributed the electron density under knob to KAHRP and proposed KAHRP-cytoskeletal interactions [58, 59].

KAHRP has been of interest in cytoadhesion since a knockout line showed KAHRP is essential for knob formation and adhesion under flow [37], but this study also resulted in reduced surface exposure of *Pf*EMP1 and interference of some neighbouring genes [36]. A follow up study by Rug *et al.* [38] was able to produce a single gene knockout of KAHRP, as well as gene truncations. Knockout of full length KAHRP did indeed result in knobs no longer forming on the surface of the erythrocyte and a reduction of adhesion to CD36 to just 5-10% under physiological flow conditions, with *Pf*EMP1 still being present at the surface. Truncation of KAHRP to remove the 3' repeat resulted in ~50% adhesion to CD36, but knob morphology was similar to that of full-length, while removal of the 5' repeat resulted in ~10% adhesion and only slight protrusions instead of knobs. Truncation to a construct containing only the histidine rich region resulted in no knob formation, but localisation appeared the same with immunofluorescence microscopy, suggesting the histidine rich region is sufficient for peripheral localisation, and potentially interacts with the erythrocyte mem-

brane.

The mechanism for how KAHRP aids in knob formation is still unknown. Many studies have suggested interactions with the host erythrocyte cytoskeleton. KAHRP has been found to form protein clusters, which bound to the spectrin-actin-protein 4.1 complex [60, 61]. Other studies suggested an interaction with ankyrin; between the D3 membrane binding domian of ankyrin and a region of KAHRP just N-terminal to the 5' repeat [62]. A spectrin interaction has also been proposed between a region of KAHRP containing the 5' repeat and α spectrin repeat 4 [63]. In addition, interactions between KAHRP and the *Pf*EMP1 ATS have been suggested, involving both the histidine rich region and the 5' repeat of KAHRP [60, 64], although this was not able to be reproduced with solution based binding assays [65].

1.2.5 PHIST Proteins

The PHIST family of proteins was identified in a study of the exported proteins in *Plasmodium*, with all having a PEXEL motif to direct export [12]. Each PHIST family member contains a conserved domain of ~150 residues, consisting of 4 helices, and all genes are localised sub-telomerically, hence the family name; *Plasmodium* Helical Interdispersed Sub-telomeric. There is at least one PHIST family member in each *Plasmodium* species, with *P. falciparum* having the most, with 71 different paralogs. PHIST proteins cluster into three separate groups; PHISTa, PHISTb and PHISTc, based on the position of conserved tryptophan residues within their sequences. In addition, there are architectural differences in these groups, with PHISTa members typically being short, being composed just of the signal sequence for export and the 4 helix region, while both PHISTb and c are more variable in length, some of which have additional domains, including the seven PHISTb members that have DnaJ domains.

PHIST proteins are likely to have diverse functions, but some have been implicated in

trafficking and cytoadhesion as a large scale knockout study resulted in lost of *Pf*EMP1 trafficking, and changes in cell rigidity and adhesion [56]. The presence of DnaJ domains in some PHISTb proteins suggests these may act as co-chaperones, where DnaJ or heat-shock protein 40 (Hsp40) bind to the major chaperone, heat shock protein 70 (Hsp70) to modulate function [66]. As there is an exported parasite Hsp70, these PHIST and DnaJ domain proteins might work with Hsp70 to ease the cellular stress that occurs during *P. falciparum* infection [30]. Many PHIST proteins have been found to localise to the membrane of the host erythrocyte, although their function there is yet to be determined [67].

The PHIST proteins examined in this thesis are PFI1780w and PFE1605w. PFI1780w is in the PHISTc family and interacted with *Pf*EMP1 in a yeast two hybrid screen [68] but attempts to produce a gene knockout have so far been unsuccessful, implying it plays a vital cellular function [65]. Localisation with florescence microscopy, using a GFP fusion, suggested membrane localisation, which is supported by it being found in detergent resistant membrane fractions [69]. PFE1605w is in the PHISTb family and has been shown to localise to the surface of infected erythrocytes, and in some cases to knob structures [39, 40]. Furthermore, deletion of the PFE1605w gene resulted in reduction in adhesion to CD36 of 55%, suggesting a direct role in adhesion [40]. Both PFI1780w and PFE1605w were investigated for potential interactions with *Pf*EMP1.

1.2.6 Adhesion in other *Plasmodium* Species

Avoiding splenic clearance is a challenge all *Plasmodium* parasites must face, whether their host be human, monkey, mouse or other. The formation of cytoadherent knobs however only occurs in *P. falciparum* and its closest genetic relation *P. reichenowi* [70]. However, membrane perturbations similar to knobs are visible in *P. malariae*, *P. brasilianum* [71], *P. fragile* [72], and *P. coatneyi* [73], although cytoadherence properties have not been tested. *P. vivax*, *P.yoelii* [74], *P. berghei* [75], and *P. chabaudi*, have all been shown be be able to

cytoadhere in some cases, and appear to be able to do so without knobs, using receptors for gene families that are distinct from *var* genes [76, 77]. This suggestions that while all parasites evolved adhesion properties, the evolution of knob structures is relatively new and unique to *P. falciparum*, and possibly key to the prevalence and pathogenicity of this parasite species.

1.3 IDPs in Malaria

The four proteins mentioned above; *Pf*EMP1, KAHRP, and PHIST family members PFI1780w and PFI1605w, are all predicted to have extensive disordered regions [78]. This has been confirmed for the ATS of *Pf*EMP1, for which the NMR structure has been determined, suggesting it has a folded core of ~80 residues, with the remaining ~300 residues being disordered [65] (Fig. 1.4).

Intrinsically disordered proteins (IDPs) are acknowledged as key players in essential biological processes, such as signalling, nucleic acid binding and replication, and biomolecular interactions [79]. IDPs contain extensive regions lacking cooperatively-folded structure, and are often characterized by a high proportion of charged and hydrophilic residues and repetitive sequences [80]. Their inherent flexibility renders IDPs and their interactions hard to study by traditional structural methods; thus, resolved protein complexes involving IDPs are scarce [81] and their significance in biology is poorly understood. However, novel methods utilizing Nuclear Magnetic Resonance (NMR), solution X-ray scattering and computational approaches have recently brought a panoply of tools to characterize IDPs [82].

P. falciparum is enriched in IDPs compared to other eukaryotes when considering both the fraction of proteins containing disordered regions and the average length of these regions [78]. IDPs localize in all parasite compartments and host cell organelles, and are present throughout all stages of the parasite lifecycle, with many proteins serving essential functions [83]. A study by Tarr *et al.* on PHIST domains that localise to the erythrocyte

cytoskeleton found that deletion of regions of predicted disorder around the PHIST domain resulted in loss of cytoskeletal localisation, suggesting these disordered regions serve essential functions in binding [67]. The PHIST proteins studied by Tarr *et al.*, were often highly charged, similar to *Pf*EMP1, KAHRP and PHIST proteins that are studied in this thesis, hence this study of *P. falciparum* IDPs may reveal interaction patterns that are applicable to other *Plasmodium* disordered proteins.

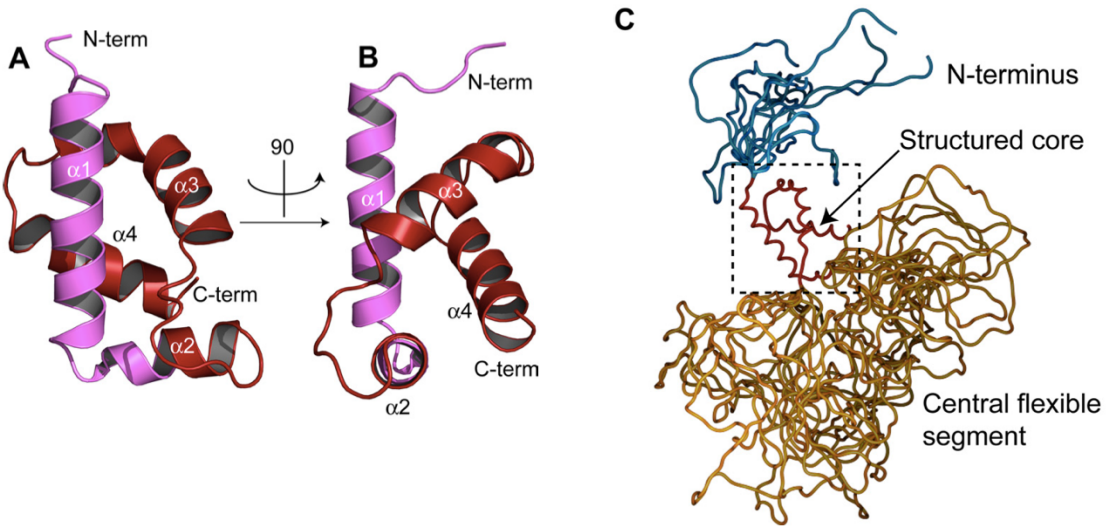


Fig. 1.4: NMR structure of the ATS of *Pf*EMP1, composed of a small structured core, **A, B**: with a disordered N-terminus, middle loop and C-terminus, **C**. Adapted from Mayer *et al.* [65].

1.4 Knob Structures

1.4.1 Erythrocyte Cytoskeleton and Knob Structures

Knob density on the infected erythrocyte surface varies amongst parasite isolates and during intra-erythrocytic parasite development [84]; however, at high density knobs are spaced in regular intervals proportional to the extended length of host spectrin tetramers [38, 84, 85]. This correlation, the observation of changes to host cytoskeleton during parasite growth [86] and the conceptual need to mechanically anchor adhesion molecules to cells in order

to resist blood flow forces suggest the presence of links between knob components and the erythrocyte cytoskeleton. Pull-down assays *in vitro* and from parasitized cells support the presence of interactions between the *Pf*EMP1 ATS and the cytoskeletal junction [60], and between KAHRP, spectrin and ankyrin [62, 63, 87, 88]. Recent nano-scale simulations assaying the contribution of linkages between knobs and the cytoskeleton to cell rigidity showed excellent agreement with experimental data from parasitized cells [89], thereby supporting the notion that knob and cytoskeletal components interact.

1.4.2 Cytoskeleton Structure

In order to understand the structure and significance of *Plasmodium* proteins interacting with the cytoskeleton, first we must understanding the structure of the cytoskeleton in an uninfected erythrocyte. Due to the ease of acquiring human erythrocytes, the basic structure of the erythrocyte cytoskeleton has been long established. The erythrocyte cytoskeleton is uniquely adapted to withstand the shear stresses a circulating blood cell is exposed to [90]. It is composed of long spectrin filaments that are linked to the membrane via interactions with integral membrane proteins. Spectrin has two chains, α and β, which are each composed of structural repeats of three helix bundles, each contain ~100 residues [91]. The α chain has 21 structural repeats, while β chain has 17. Despite these structural repeats having the same fold, the sequence similarity between repeats can be as low as 10% identity, and is not greater than 30% [92]. The α and β chains come together to form an antiparallel dimer of dimers, with one dimerisation interface being composed of one helix from α chain repeat 1 and two helices from β chain repeat 17 [93], and the other dimerisation interface being composed of direct interactions between α repeats 20 and 21 with β repeats 1 and 2. The N-terminus of the β chain has 2 calponin homology (CH) domains, and the C-terminus of α has a calmodulin-like binding domain, both of which mediate attachment to the junctional complex via interactions with actin and band 4.1 [94]. Up to 7 spectrin

tetramers can meet at one junction resulting in an interconnected mesh [95]. Band 4.1 acts as a hub for many other protein interactions, such as band 3, p55, and glycophorin-C (GPC), and hence the exact composition of the junctional complex is not fully understood and is likely to be variable [96]. The minimal junctional complex is composed of a short actin filament, encased by tropomyosin and capped with tropomodulin and adducin [97], linked to the membrane via adapter proteins including band 4.1 and p55 providing connections to integral membrane proteins such as glycophorin C and D (GPC and GPD) and band 3 [98]. Ankyrin also connects spectrin to the membrane. It binds β spectrin repeat 14 and 15 via the ZU5 domain [99], as well as having multiple interaction sites for band 3 in the membrane binding domain (MBD) [100]. Again, the ankyrin complex can also include other factors, such as band 4.2 (thought to bind spectrin, band 3 cytoplasmic domain and ankyrin), CD44 and Rh [96, 101]. In addition, membrane interactions have been proposed for β chain repeat 14, in the absence of ankyrin [102].

Hence, the cytoskeleton is composed of interconnected spectrin filaments, tethered to the membrane. Spectrin acts as springs between these anchor points, potentially being both flexible and being able to unfold and refold in response to shear stress, hence is both strong and flexible [103]. In addition to a structural role, the cytoskeleton can also modulate the diffusion of membrane proteins not directly bound. There is a population of band 3 not associated with the cytoskeleton, but modelling and photo-bleaching demonstrates this population has limited diffusion [104, 105]. Computational models of the cytoskeleton have been able to accurately predict experimental gross-deformation data, using spectrin as elastic springs, and the membrane as resistant to bending [106]. Finer detail has been added to these models to include membrane properties such as membrane protein diffusion, friction and membrane-fluctuations [107]. Coarse grain molecular dynamics has also been used to build a 1 μm^2 patch of erythrocyte membrane and cytoskeleton and has been used to show how defects in cytoskeleton result in changes in cellular physiology [108, 109].

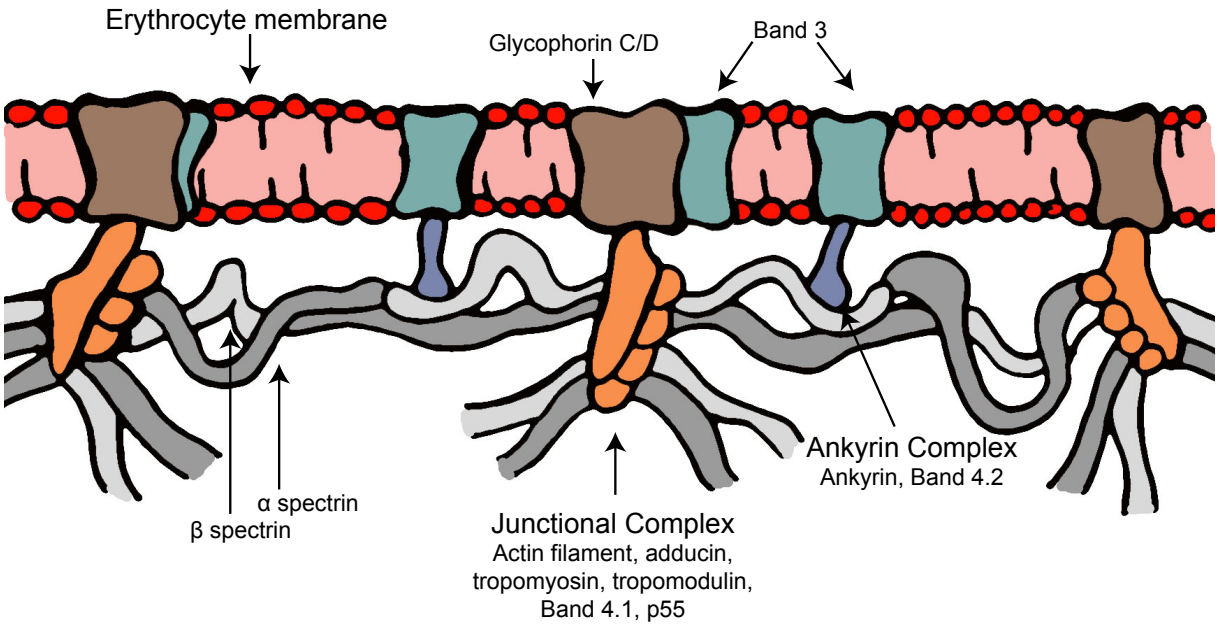


Fig. 1.5: Simplified view of the eythrocyte cytoskeleton. Spectrin tetramers, composed of 2 α and 2 β chains, meet at junctional complexes, which are linked to the membrane by glycophorin and band 3. Additional linkages are made by the ankyrin complex, which binds β spectrin and band 3.

1.4.3 Regulation of Cytoskeleton

Cytoskeletal structure and stability is tightly regulated. This operates by two connected mechanism; Ca^{2+} levels and by phosphorylation. Cytoplasmic concentrations of Ca^{2+} are generally kept at very low levels of just 30-60 nM, despite being at 1.8 mM outside the cell [110]. When Ca^{2+} concentration increases, connections holding together the cytoskeleton start to weaken. In elevated Ca^{2+}concentrations, the FERM domain of band 4.1 can be bound by Ca^{2+}-calmodulin, reducing affinity to band 3 protein, p55, and GPC [111]. Ca^{2+}-Calmodulin also binds to and down-regulates capping activity of adducin thereby regulating actin filament assembly [112]. In addition, Ca^{2+} is required to activate protein kinase C (PKC) which phosphorylates band 4.1, again, resulting in it having reduced affinity for the cytoskeleton [113]. In conditions of stress, tyrosine kinases phosphorylates band 3, greatly

reducing affinity for ankyrin and cell stability [114]. Many other phosphorylation sites exist in cytoskeletal proteins, although the complex nature of regulation by phosphorylation means that the functional interplay is still not well understood [115, 116]. Phosphorylation can also act via phosphatidylinosityl (PI), where phosphatidylinositol 4,5-bisphosphate (PIP_2) is a major PI lipid in the erythrocyte membrane [117]. PIP_2 significantly enhanced the binding of band 4.1 to glycophorin C (GPC) and inhibits the binding of band 4.1 to band 3 *in vitro* [118].

1.4.4 Cytoskeleton Interactions and Modifications in *P. falciparum*

While this thesis focuses on cytoskeleton interactions involved in cytoadhesion, during infection many other parasite protein interactions are occurring with the cytoskeleton. During infection the invading merozoite releases the ring-infected erythrocyte surface antigen (RESA), which has been found to bind the β spectrin repeat 16, stabilising the spectrin tetramer. The increased stability of spectrin tetramers helps to prevent multiple merozoites infecting the same erythrocyte and possibly strengthen the erythrocyte to withstand forces during adhesion [119]. Many MC and trafficking proteins also interact with the cytoskeleton. Skeleton binding protein 1 (SBP1) binds to band 4.1 and spectrin α repeat 4 [120]. *Pf*EMP3 has also been found to bind to the C-terminus of α spectrin and to actin, leading to destabilisation of the cytoskeleton [121, 122]. *P. falciparum* protein 332 (*Pf*332) is the largest exported protein and has been proposed to interact with actin, resulting in increased rigidification of infected erythrocytes [123]. Deletion of SBP1, *Pf*EMP3 or *Pf*332 results in loss of export of *Pf*EMP1 from the MC to the membrane, suggesting protein trafficking from the MC to the membrane and cytoskeleton is a highly regulated and sensitive process. Additionally, many PHIST family members localised to the cytoskeleton or MC, result in impaired trafficking, altered cell deformability or adhesion properties if disrupted [56, 67].

Mature-parasite-infected erythrocyte surface antigen (MESA) has been found to bind

band 4.1 [124–126] and localised to knobs, but disruption did not lead to any change in cytoadhesion, knob morphology or parasite viability [127, 128], hence its role is still unknown. The cytoskeleton interaction motif of MESA is a short sequence present in 13 other proteins, 6 of which were experimentally confirmed to interact with band 4.1 [129], therefore increasing the number of possible cytoskeleton interacting proteins. Some of these proteins may have redundant functions to MESA, which would explain why knocking out MESA has no major effect, despite its specific localisation.

Parasite and host proteases target the cytoskeleton in late stages to help promote egress [130–133]. In addition, merozoite surface protein 1 (MSP1), binding to spectrin results in parasite egress [134], but as egress occurs long after cytoadhesion, is out of the scope of this thesis.

With so many interactions between *P. falciparum* and cytoskeletal proteins it is no surprise that changes to the cytoskeleton are observed during parasite growth. The most extensive study of such changes was presented in Shi *et al.* using atomic force microscopy [86]. This method was able to visualise the cytoskeleton network and showed knobs localised at cytoskeletal junctions with 6-8 spectrin filaments extending out from each knob. With each stage of parasite development the cytoskeleton mesh expanded, with both spectrin filaments appearing longer, and larger holes being present in the mesh. As the volume and surface area of infected erythrocyte does not change significantly during the trophozoite stage [135], this implies loss of cytoskeleton connections in regions without knobs, a hypothesis supported by the decrease in spectrin abundance observed. Examining the distribution of membrane proteins around knobs using freeze fractured samples and electron microscopy show a clearance of membrane proteins around the base of knobs [36, 136]. Proteomic analysis of proteins associated with the erythrocyte membrane suggest loss of cytoskeletal adapter proteins, such as adducin and tropomyosin in late stages of parasite development [137]. Consistent with the loss of actin capping molecules, actin is hypothesised

to be redistributed in infect erythrocytes to provide a linkage between the MC and the host cytoskeleton [20, 138].

1.5 Current View of Knob Structure

The current view of knob structures as reviewed by Maier *et al.* [139], focuses on results suggesting interactions between KAHRP and *Pf*EMP1, and KAHRP and the cytoskeleton (Fig. 1.6A). A major flaw with this model is it suggests an inter-dependent localisation of KAHRP and *Pf*EMP1, but studies have shown that reduction of *Pf*EMP1 at the surface do not affect knob morphology [140] and deletion of KAHRP does not prevent *Pf*EMP1 clustering [36]. This current view also ignores many of the potential interactions suggested above, and furthermore simplifies structures which appear highly regulated. Atomic force microscopy over the surface of infected erythrocytes shows regular height, width and spacing of knob structures [38] (Fig. 1.6B), which is not influenced by multiple parasite infection, suggesting each knob has controlled composition, rather than is depending on the level of gene expression of each component [141]. There is also localised redistribution of intergral membrane proteins around knobs [36, 136].

Furthermore, a recent electron tomography study has shown an organised spiral scaffold underlies the knob (Fig. 1.6C-E). This spiral forms underneath an electron dense layer of material that interacts with the membrane. The electron dense layer is the most recognisable feature of knob structures and has previously been attributed to the presence of KAHRP [59, 87]. Collectively, this data demonstrates the need for a new model for knob structures, that includes a mechanism for the localisation and regulation of key factors in cytoadhesion, such as *Pf*EMP1 and KAHRP.

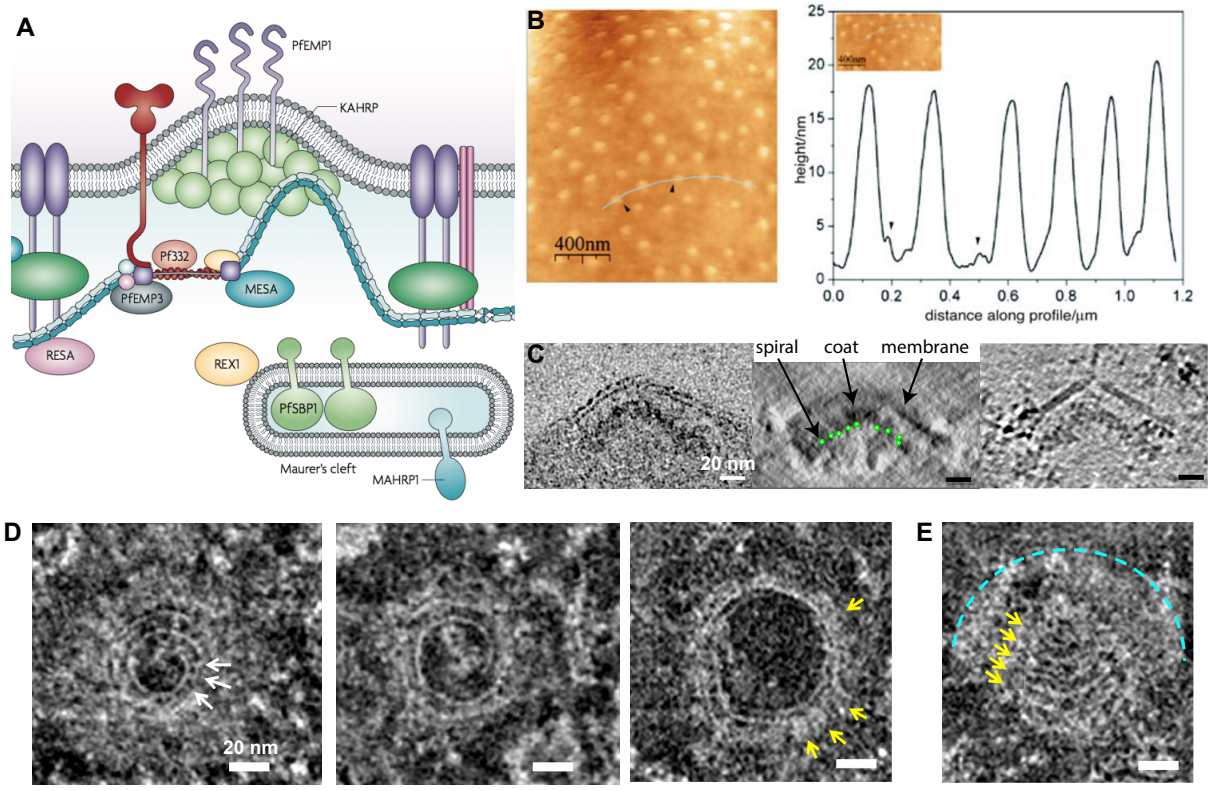


Fig. 1.6: Current view of knob structure. **A**: The most recent model for knob composition and structure. Adapted from Maier *et al.* [139]. **B**: Atomic force microscopy of infected erythrocyte, with the height and spacing of knobs plotted. Adapted from Rug *et al.* [38]. **C**: Internal knob structure in the schizont stage using different methods. Left, prepared from high-pressure-frozen, freeze-substituted infected erythrocytes, middle, negative-stained detergent-insoluble schizont skeleton and right, skeleton in vitreous ice. **D**: Tomogram through knobs, looking down from the apex, taking slices getting closer to the plane of the membrane. The spiral, indicated by arrows, is visible through the different layers of the knob structure. **E**: Slice through the knob structure, from a perpendicular view to **D**, spiral is indicated by yellow arrows. C, D and E adapted from Watermeyer *et al.* [136].

1.6 Work Presented

This thesis characterises the interaction between three classes of *P. falciparum* proteins and host erythrocyte cytoskeleton. Chapter 3 investigates PHIST family member's interaction with *Pf*EMP1 and band 3. Chapter 4 investigates interactions of *Pf*EMP1 with host erythrocyte spectrin. Chapter 5 investigates the interaction between KAHRP and host cytoskeleton.

These results are individually discussed in each chapter, but built together to form a better understanding of how knob structures assembly and how they are possibly regulated, which is discussed in Chapter 6.

Chapter 2

Methods

"Functional perimeters vary from manifestation to manifestation"

'Handbook for the Recently Deceased', Beatlejuice

2.1 Protein Purification

2.1.1 *Pf*EMP1

Recombinant expression of *P. falciparum* Erythrocyte Membrane Protein 1 (*Pf*EMP1) Acidic Terminal Segment (ATS) variants was performed as described in Mayer *et al.* [65] and Oberli *et al.* [39].

2.1.2 PHIST Proteins

Recombinant expression of PFI1780w and PFE1605w was performed as described in Mayer *et al.* [65] and Oberli *et al.* [39].

2.1.3 KAHRP

P. falciparum KAHRP (UniProt accession number Q9TY99) constructs are shown in Appendix A. The scramble sequences were ordered as geneblocks (IDT, California, USA) with sequences shown in Appendix F. KAHRP constructs were cloned in a modified pET16b vector that includes an N-terminal His10-tag and a human rhinovirus (HRV) 3C protease cleavage site, and transformed into *Escherichia coli* strains BL21(DE3) CodonPlus-RP (Agilent Technologies, Stockport UK) or Rosetta2(DE3) (Novagen, Watford UK). Cells were grown at 37 °C in Luria Bertani (LB) media or, for NMR usage, in M9 minimal media supplemented with ^{15}N enriched NH_4Cl and/or ^{13}C enriched D-glucose. The growth temperature was reduced to 18 °C at OD_{600} 0.5, and protein expression was induced at OD_{600} 0.6 with 500 mM final concentration of Isopropyl β-D-1-thiogalactopyranoside (IPTG, Generon, Maidenhead UK) for 16-18 hrs. Cells were harvested by centrifugation and resuspended in 50 mM NaH_2PO_4, 500 mM NaCl, 8 M Urea pH 7.8 buffer. Cells were lysed with sonication and lysates were clarified by centrifugation at 50,000 g prior to loading in Talon metal affinity columns (Clontech, Moutain View CA) equilibrated in lysis buffer. Proteins were eluted by lysis buffer supplemented with 500 mM imidazole, and extensively dialyzed against 500 mM NaCl, 50 mM NaH_2PO_4, 1 mM 1,4-dithiothreitol (DTT), 1 mM ethylenediaminetetraacetic acid (EDTA) pH 6.5 buffer. Cloning tags were removed by cleavage with recombinant HRV 3C protease. Proteins were dialyzed against 150 mM NaCl, 20 mM NaH_2PO_4, 1 mM DTT, 1 mM EDTA pH 6.5 buffer prior to ion exchange chromatography (SP-Sepharose media, GE Healthcare, Little Chalfont UK). Final purification was performed by size exclusion chromatography using Superdex 75 (GE Healthcare) media equilibrated in NMR buffer (20 mM NaH_2PO_4, 50 mM NaCl, 1 mM DTT pH 7) unless otherwise noted. Proteins were concentrated by spin ultrafiltration, and concentrations estimated by UV absorption at 280 nm

2.1.4 Cytoskeletal Components

Human erythrocyte spectrin constructs (UniProt P02549 and P11277, constructs shown in Appendix A) were cloned in modified pET16b (as above) or pET15b (N-terminal His6-tag, thrombin cleavage site) vectors and recombinately expressed in *E. coli* Rosetta2(DE3) in LB media for 4 hrs at 37 °C following induction with 250 mM final concentration of ITPG. Cells were harvested by centrifugation and resuspended in PBS (150 mM NaCl, 20 mM NaH_2PO_4 pH 7.4). Cells were lysed with sonication and lysates clarified by centrifugation and applied to PBS-equilibrated Talon metal affinity columns. Proteins were eluted using PBS supplemented with 500 mM imidazole, dialyzed against 50 mM Tris-Cl, 50 mM NaCl pH 7.5 buffer, and cloning tags were cleaved using HRV 3C or thrombin (Sigma Aldrich, Gillingham UK) proteases. Proteins were further purified by ion exchange chromatography (Q-Sepharose media, GE Healthcare) and size exclusion chromatography (Superdex 75 or 200 media, GE Healthcare) into NMR buffer unless otherwise noted.

Human erythrocyte ankyrin (UniProt P16157) membrane binding domain constructs (D1-D4 and D3-D4 ankyrin repeats, constructs shown in Appendix A) were cloned in a pFloat vector [142] with an N-terminal Glutathione-S-transferase (GST) tag. Protein expression in *E. coli* Rosetta2(DE3) was induced by addition of 0.2 mM final concentration IPTG and left to proceed at 18 °C for 16 hrs. Cells were harvested by centrifugation, resuspended in 50 mM NaH_2PO_4, 300 mM NaCl, 1 mM DTT, 1 mM EDTA pH 7.8 buffer and lysed by sonication. Lysates were clarified by centrifugation and applied to glutathione-sepharose media (GE Healthcare) equilibrated in lysis buffer. Proteins were eluted by lysis buffer supplemented with 15 mM reduced glutathione. The GST-tag was cleaved with HRV 3C protease while dialyzing against lysis buffer, and removed by passage through glutathione-sepharose media. Final protein purification was performed by size exclusion chromatography in Superdex 200 media equilibrated in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 6.8, 1 mM DTT, 120 mM NaCl buffer. Human

erythrocyte ankyrin spectrin binding domain constructs ZU5A and ZU5AB (Appendix A) were produced as described [143]. Construct ZU5B (Appendix A) was cloned in a pFLOAT vector [142] with an N-terminal His6-tag and purified in a manner similar to the ZU5A and ZU5AB constructs.

Band 3 purified as previously described by Zhang *et al.* [144].

Proteins were concentrated by spin ultrafiltration, and concentrations estimated by UV absorption at 280 nm.

2.1.5 Cysteine Protein Labelling

Maleimide is a reactive group that forms an irreversible linkage with reduced cysteines. This chemistry was used for 5-carboxy fluorescein (5-FAM) labelling for fluorescence polarisation assays, attachment of 3-maleimido-PROXYL (2,2,5,5,-tetramethyl-1-1pyrrolidinyloxy) free-radical label for PRE NMR and for gold-nano-particle conjugation for EM. Labelling of proteins was done using previously established protocol [65]. In brief, sample was prepared without reducing agent present, tris(2-carboxyethyl)phosphine (TCEP) was added to a 10-fold or higher molar excess for 1 hour to reduce protein and then label was added in a 10-fold or greater excess and incubated for >3 hours or over night. Unconjugated label was removed with size exclusion chromatography.

Labelling was performed in a site-specific manner using single cysteine residues. In the case of 5-FAM labels, via cysteines introduced at the middle disordered segment of ATS variants [65] or at the protein N-terminus. KAHRP sequence scrambles swapped pre-existing cysteine residues for serine residues and included an N-terminal cysteine for 5-FAM labelling. Incorporation of 5-FAM was confirmed by mass spectroscopy and/or absorption at characteristic wave length, 495 nm. Proteins were concentrated by spin ultrafiltration, and concentrations estimated by UV absorption at 280 nm and 495 nm (label absorption wavelength).

2.2 NMR

All experiments were performed using Bruker Avance II- or Avance III spectrometers with cryogenic TCI probeheads, and used field strengths of 500, 600 or 750MHz. Samples all prepared in NMR buffer supplemented with 5% v/v D2O, 0.02% w/v NaN_3 and 50 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid unless otherwise noted. Date was collected with Topspin software (Bruker), processed with NMRpipe and evaluated with NMRDraw [145] and Sparky [146].

2.2.1 HSQC Titrations

Heteronuclear single quantum coherence (HSQC) spectra are used extensively throughout this thesis to investigate protein-protein interactions, as well as a general indicator of sample quality. HSQC spectra have a peak for each ^{15}N-^{1}H or ^{13}C-^{1}H bond in the sample, therefore on average one peak per amino acid in the backbone, hence is considered to be a protein finger-print. This pulse sequence works by exciting ^{1}H, then transferring of magnetisation via an insensitive nuclei enhanced by polarization transfer (INEPT) to ^{13}C or ^{15}N, a variable evolution time t_1, transfer back on to ^{1}H for detection as a free induction decay, for time, t_2 (Fig. 4.1) [147]. Hence this results in separation of NMR signal into two frequency ranges, from t_1 and t_2. As the resonance frequency of any nuclei depends on its chemical environment, the addition of a binding partner results in a change of resonance frequencies for those regions involved in binding. In practice, this depends on the time scale of the interaction, as fast time scale interactions will result in a peak that has the resonance of the average of bound and unbound states, slow time scale results in separate peaks for bound and unbound states, and intermediate timescales result in peak broadening and loss of signal. The interactions within this thesis are in the intermediate to fast exchange time scale regime, resulting in loss of peak intensity and shift perturbations, respectively. Additional changes to spectral appearance could be caused by the formation of a large complex, as a large

complex tumbles slower with a larger correlation time, τ_C, this results in faster transverse relaxation rates and less signal. Spectral analysis showing loss of peak intensity is shown as 1D slices through HSQC spectra, produced using python package NMRglue [148].

PFI1708w ^{15}N-HSQC spectra were collected with protein concentration of 100 μM collected at 37 °C, with pH varying with experiment. Spectra of the ATS collected as in Mayer *et. al.* [65], essentially at 10 °C for the ATS-N, -M and -C and at 25 °C for the ATS-Core. KAHRP spectra were also collected at 10 °C.

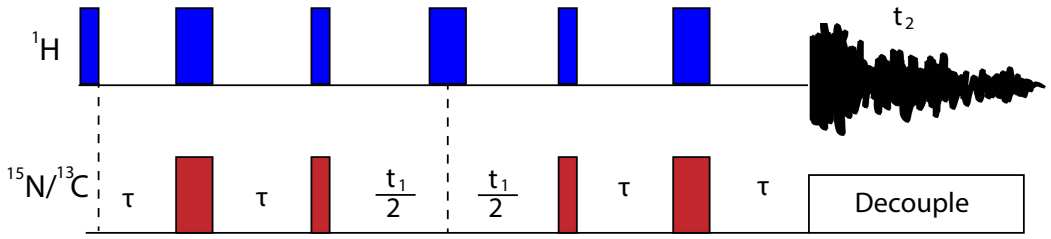


Fig. 2.1: HSQC pulse sequence. The first pulse results in signal on ^{1}H, this is then transferred to the heteronucleus, allowed to evolve for variable time, t_1 and then transferred back to ^{1}H for detection. Figure based on Cavanagh *et al*, 2007 [147]

2.2.2 Paramagnetic Relaxation Enhancement

Paramagnetic relaxation enhancement (PRE) is a method that can can provide information about the orientation of a ligand binding to a NMR labelled protein by incorporating a free-radical label at various positions in the ligand. This free radical reduces NMR signal in a distance dependent manner. The ratio of peak intensities with the label, I, and with the label quenched, I_0, can be expressed as follows [149, 150]:

$$\frac{I}{I_0} = exp\left(\frac{tK}{r^6}\left(4\tau_c + \frac{3\tau_c}{1+\omega_H{}^2\tau_c{}^2}\right)\right)^{-1} \tag{2.1}$$

Where, r is the distance between the radical and the proton, t is the total time the amide proton experiences paramagnetic relaxation, which is 10.6 ms in HSQC experiments, the

parameter K takes the value $1.23*10^{-32}$ cm^6s^{-2}, ω_H is the larmor frequency of the amide proton, 600 MHz (in the experiments in this thesis) and the rotational correlation time, τ_c. Hence, the distance between the label and amide bonds in the structure, r, is proportional to I/I_0. ATS-C had 3-maleimido-PROXYL (2,2,5,5,-tetramethyl-1-1pyrrolidinyloxy, PROXYL) conjugated to engineered cysteines at D328 or E346. ^{15}N-HSQC spectra were collected for 100 μM PFI1780w in NMR buffer at pH 7 in the presence of equimolar ATS-C-PROXYL and then with the PROXYL quenched with 5 mM ascorbic acid. Peak intensity determined with Sparky [146], PRE-effect determined as $1-I/I_0$ and mapped onto the structure of PFI1780w.

2.2.3 Backbone Assignment

Sequence-specific resonance assignments of ATS variant PF08_0141 have been reported previously [65]. Sequence-specific resonance assignments of ATS-Core variant PFF0845c and spectrin α17 were performed using 3D CBCA(CO)NH, CBCANH, HNCO, HN(CA)CO and HBHA(CO)NH pulse sequences [147, 151]. This sequential assignment method works by CBCANH providing cross peaks for Cβ and Cα of residue (*i-1*) and *i* and NH resonance of *i*, which can be lined up to the Cβ and Cα peaks for residue *i* with the NH resonance of (*i+1*) in CBCA(CO)NH, thereby providing an (*i-1*) to *i* and (*i+1*) linkage. This is then repeated throughout the sequence (Fig. 4.2A). The resonance frequency of Cβ and Cα varying with residue type, hence can be matched up to the sequence. Spectral overlap and region of poor signal can result in ambiguous correlation with just two experiments, hence HN(CA)CO and HNCO, which provide similar (*i-1*) and *i* connections, and HBHA(CO)NH, which provides ^{1}Hβ and ^{1}Hα resonances, were collected (Fig. 4.2B,C,D). Spectrin α17 NMR data for assignments were collected at 40 °C in NMR buffer with 300 mM NaCl at pH 6.5. PFF0845c ATS-Core assignment data were collected in the same conditions but at 30 °C. NMR data were processed using NMRpipe [145], analyzed and

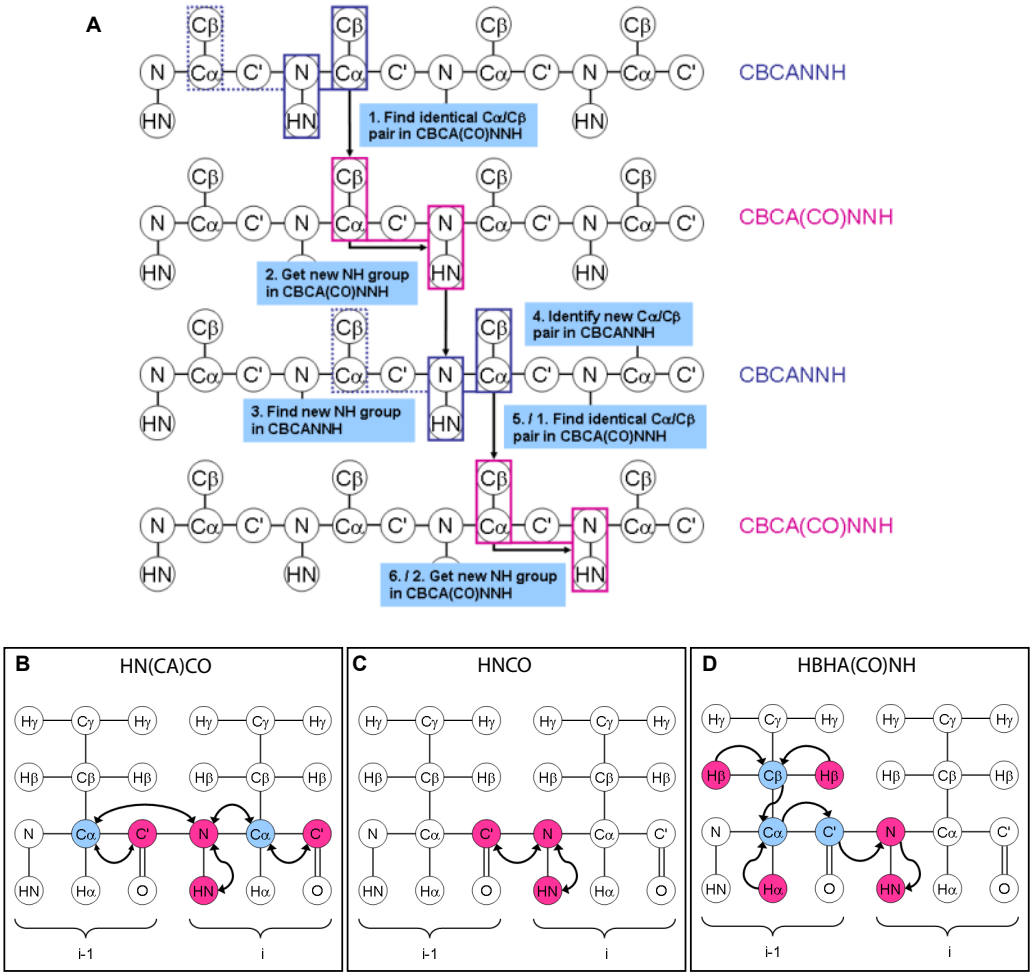


Fig. 2.2: NMR sequential assignment method. **A**: CBCANH provides correlation between (*i-1*) Cβ and Cα to *i* NH, Cβ and Cα. This can then be matched with *i* Cβ and Cα in CBCA(CO)NH, which provides correlation to NH of (*i+1*). **B, C**: HN(CA)CO and HNCO, respectively, also provide (*i-1*) to *i* connections. **D**: HBHA(CO)NH correlate the Hβ and Hα of residue (*i-1*) to the NH of residue *i*. Where the resonance frequency of Hβ and Hα vary depending on the residue, this provides additional information to identify the sequence location. Figure adapted from http://www.protein-nmr.org.uk

assigned using CCPN Analysis [152]. Assignments have been deposited in BioMagResBank under accession numbers 26772 and 26773 for the ATS-Core variant PFF0845c and spectrin α17, respectively. Spectra overlays were prepared with Sparky [146]. Resonance perturbations were mapped using ^{15}N-HSQC and 3D HNCO experiments. Assignment for α17 were transferred from 40 °C to 30 °C via collecting HNCO data at 40, 35 and 30 °C. Perturbations from multiple nuclei types were combined using a sum of absolute differences approach weighted by nuclei-specific factors [153].

2.3 Electron Microscopy

Electron microscopy was performed with a FEI Tecnai 12 Transmission Electron Microscope with CCD (Charge-coupled device) camera. Full length spectrin purified from human erythrocytes was purchased from Sigma. ATS-N-Core was conjugated to 10 nm Maleimide coated gold nano-particles (Sofchip) by an N-terminal cysteine residue. Spectrin was diluted 1/50 in ATS-N-Core-nanoparticle sample straight from size exclusion chromatography, applied to glow discharged 300 mesh formvar-carbon coated cooper grids for 2 minutes, blotted, stained with 2% w/v uranyl acetate for 30 sec, blotted and allowed to dry. Samples of spectrin alone and spectrin with unconjugated nano-particles were prepared as controls.

2.4 Biophysical Assays and Methods

2.4.1 Mass Spectroscopy

Mass of constructs expressed was confirmed where indicated with TOF mass spectrometry using a Micromass LCT.

2.4.2 CD

Circular Dichroism (CD) was used to evaluate the structural state of constructs where indicated. This method is based on the ability of chiral carbons in the protein backbone to absorb circularly polarised light. The wavelength of light absorbed is dependant on the secondary structure of the protein, hence can give an indication of the type, presence, and rough quantity of secondary structure within a protein.

All CD spectra were measured using a Jasco (Eaton, MD) J-815 Spectropolarimeter in a QS glass cuvette with a 1 mm path length (Hellma Analytics) with samples in NMR buffer. Individual spectra were recorded between 260-180 nm at 10 °C with buffer background subtraction. Thermal stability was determined monitoring CD signal at 222 nm, with a spectral range of 250-200 nm and temperatures between 10-80 °C. All data were normalised for sample concentration and number of residues.

2.4.3 FP Assays

Fluorescence polarisation (FP) assays were used to determine binding affinity. These assays are based on the principal that changes in the dynamic behaviour of a protein labelled with a fluorophore, such as when it binds to a large globular protein, result in a change of the fluorescence polarisation.

FP binding assays were performed at room temperature using a CLARIOStar fluorimeter (BMG Labtech, λex= 485 nm, λem = 520 nm). 5-FAM labelled proteins at 0.5 or 1 μM concentration in NMR buffer were titrated with defined concentrations of unlabelled proteins in the same buffer. Changes in fluorescence polarisation were fit using a single binding model in the program Origin (OriginLab, Northampton MA). Unless specified otherwise, error bars show propagated standard error from 4 technical repeats.

2.4.4 SAXS

Small angle X-ray scattering (SAXS) is a solution based method that give low resolution structural information. In this thesis SAXS curves were compared to theoretical scattering curves for crystal structure solved, in order to better understand solution behaviour.

SAXS data were collected at the BioSAXS beamline at ESRF (BM29) at 20 °C and 0.9919 Å wavelength. Samples in NMR buffer with 100 mM NaCl were centrifuged prior to data collection, and supplemented with 5 mM DTT. Sample buffer and protein 2-fold dilution series (starting at 6.3 mg/mL for α17, 10.5 mg/mL for α16-17, 11.7 mg/mL for ATS-N-Core and 10.8 mg/mL for ATS-N(Δ84)-Core) were measured while flowing through a thin capillary. On-site inspection of data showed no indication of radiation damage. A sample of bovine serum albumin was measured as control. Buffer subtraction, intensity normalization, and data merging was performed using PRIMUS [154].

2.5 X-ray Crystallography

In all cases concentration of protein to use in crystallization trial was determined using pre-crystallisation tests (Hampton research). A broad crystallization screening approach was taken using screens JCSG I, II, III, IV (Qiagen) and Morpheous (Molecular dimensions, Suffolk, UK) at room temperature and at 4 °C. A Mosquito robot (TTP LabTech, Melbourn UK) was used to setup 200 nl-size drops with 1:1 volume ratio of protein to mother liquor. All diffracting crystals were obtained using the sitting drop vapor diffusion technique at 4 °C. All structural model quality was accessed with MolProbity [155] and visualized using PyMOL [156].

2.5.1 Structure Determination of α Spectrin Repeats 17 to 19

An initial crystallisation hit of Spectrin α17-19 with ATS-N(Δ84)-Core was observed in Morpheus screen condition C1, which was optimised with pH, precipitant concentration and silver bullet additive screens. Final conditions were 7.5 mg/mL protein concentration, which formed slab cluster crystals when mixed with 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 0.1 M 2-(N-morpholino)ethanesulfonic acid / imidazole pH 6.5 buffer, 23% PEG8K/EDO and Silver Bullet Bio additive screen (Hampton research, California, USA) condition F7 (0.2% w/v of Thymine, Sodium pyrophospahte tetrabasic decahydrate, D-Glyceric acid calcium salt dihydrate, β-cyclodextrin and myo-Inositol, and 0.02 M HEPES sodium pH 6.8). Data were recorded Diamond Light Source (DLS, Harwell UK) beamline I04-1 to 3.7 Å. Space group was determined to be P 21 21 21 with one spectrin α17-19 molecule per asymmetric unit. Data were processed with XIA2 [157], analyzed by CCP4 [158], and the structure was solved by molecular replacement using Bables [159] and Phaser [160]. Model building was performed in Coot [161]. Iterative refinement was performed with Phenix [162]. No density was found for the ATS-N(Δ84)-Core.

2.5.2 Structure Determination of α Spectrin Repeats 16 and 17

Spectrin α16-17 with ATS-N(Δ84)-Core at 4.3 mg/ml concentration formed diffracting crystals when mixed with 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 0.1 M 2-(N-morpholino) ethanesulfonic acid/imidazole pH 6.5 buffer, 20% v/v ethylene glycol and 10% w/v polyethylene glycol 8000. Rod-like crystals developed in 5 days. Crystals were flash cooled in liquid nitrogen and data were collected to 1.54 Å at the Diamond Light Source (DLS, Harwell UK) beamline I04-1. The space group was determined as P 21 21 21 with one spectrin α16-17 molecule per asymmetric unit. Data were processed with XIA2 [157], analyzed by CCP4 [158], and the structure was solved by molecular replacement using Bables [159] and Phaser [160]. Model building was performed in Coot [161]. Iterative refinement was performed

with Phenix [162] and Buster-TNT [163] using automatic TLS restraints. The model and associated crystallographic data have been deposited in the RCSB Protein Data Bank under accession number 5J4O. No density corresponding to the ATS-N(Δ84)-Core was identified in these crystals.

2.5.3 Structure Determination of α Spectrin Repeat 17

Spectrin α17 with ATS-N(Δ84)-Core at 6 mg/mL concentration formed diffracting crystals when mixed with 0.2 M NaCl, 0.1 M HEPES, pH 7.5, 20% PEG3000. Needle cluster crystals developed after 7 days. Crystals were cryo protected with 20% glycerol and flash cooled in liquid nitrogen. Data was collected to a resolution 2.2 Å at ESRF beamline ID30A. The space group was determined as C 2 2 2 with two copies of α17 per unit cell. Data was processed with XDS [164], analyzed by CCP4 [158], and solved with molecular replacement using Phaser [160]. Iterative model building and refinement was done with Coot [161] and Phenix [162] respectively. No density was found for the ATS-N(Δ84)-Core.

2.6 Computational Methods

2.6.1 Homology Modelling

Homology modelling was done with MODELLER [165] using alignments from ClustalΩ [166–168]. PFI1780w monomer was modelled using the crystal structure of the dimer as template (PDB: 4JLE) [39], with loop modelling for residues 190-194. For spectrin, Phyre was used to provide templates for models [169]. The PFF845c ATS-Core was modelled using the NMR structure of ATS-Core of PF08_0141 variant (PDB 2LKL) [65].

2.6.2 CS Rosetta

Standard CS-Rosetta was run on a PC-cluster in parallel [170]. Specific protocol details that differ from standard set-up are provided in text. Processing of decoy structures, rescoring and clustering was done with tools in the Rosetta package.

2.6.3 HADDOCK Docking

HADDOCK [171, 172] was used to model protein complexes based on NMR shift perturbation data. This docking approach uses shift data to define a list ambiguous restraints, which are used to generate possible complexes, which are then scored on how energetically favourable they are likely to be. This was used to predict the complex between ATS-C and PFI1780w, and ATS-Core and α17. In both cases active residues were defined as residues with chemical shift perturbations greater than one standard deviation and also have more than 50 $Å^2$ solvent exposed surface areas as determined by POPS [173]. Active residues used are shown in Appendix B. Docking was done using default parameters with the expert interface of HADDOCK.

2.6.4 MD Parameter Description

2.6.4.1 Method Introduction

Simulating molecular motions using quantum mechanics is currently too computationally expensive to be applied to large protein systems, hence current molecular dynamic approaches involve several simplifications, which are built into a "force field" dictating how atoms interact [174, 175]. In biomolecule simulations this is typically generalised to an energy potential, V(r):

$$V(r) = \sum V_{bonds} + \sum V_{angles} + \sum V_{torsion} + \sum V_{non-bonded} \tag{2.2}$$

2.6.4.2 Bonded Interactions

V_{bonds}, between 2 atoms *i,j* (Fig. 4.3A), and V_{angles}, between three atoms *i,j,k* (Fig. 4.3B), are simplified to elastic springs and described with Hooke's law in which the energy varies with the square of the displacement from the reference bond length *(l_0)* or angle *(θ_0)*. This is multiplied by a force constant, k_B or k_θ, which results in high penalties for unfavourable conformational changes.

$$\sum V_{bonds} = \sum_{bonds} k_B(l_0 - l_{i,j})^2 \tag{2.3}$$

$$\sum V_{angles} = \sum_{angles} k_\theta(\theta_0 - \theta_{i,j,k})^2 \tag{2.4}$$

The result of the high force constants are that most changes during simulation are a result of $V_{torsion}$ and $V_{non\text{-}bonded}$. Torsion angles are defined between sets of 4 atoms, *i,j,k,l* (Fig. 4.3C). These angles have a much lower energetic barrier than bond angles and tend to have periodic nature, hence changes in torsion angle, ϕ_{ijkl}, can be described with cosine, for integer, n, and shift γ. Again, a force constant is applied, K_ϕ.

$$\sum V_{torsion} = \sum_{torsions} K_\phi(1 + cos(n\phi_{ijkl} - \gamma)) \tag{2.5}$$

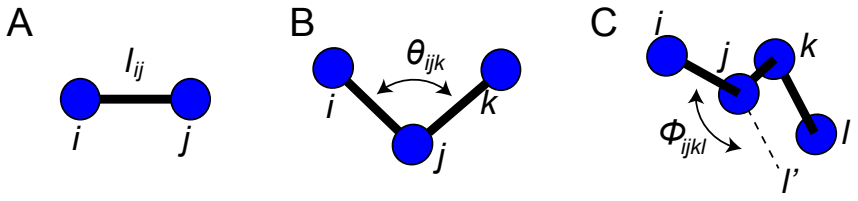


Fig. 2.3: Bonded interactions. **A**: A bond, defined by atoms *i,j* which varies in length, l_{ij}. **B**: Bond angle, defined by atoms *i,j,k* which varies in angle θ_{ijk}. **C**: Torsion angle, defined by atoms *i,j,k,l*, which varies in angle ϕ_{ijkl}.

2.6.4.3 Non-bonded Interaction

Non-bonded terms can be grouped into Van der Waals (VdW) and Coulombic (electrostatic) interactions. Van der Waals interactions are approximated with the Lennard-Jones potential. The Lennard-Jones (LJ) potential is used to approximate the interaction between a pair of neutral atoms, thus the VdW term mentioned here. Included in the potential is the strong repulsive term (r^{12}) (Pauli repulsion), present due to overlapping electron orbitals and long range attractive long term (r^6), originating from short ranged instantaneous dipoles (London dispersion) [175]. The potential has the following form:

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij}\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}}\right]^{6}\right) \tag{2.6}$$

in which σ_{ij} indicates the distance at which the energy between the two atoms in the system is zero, ε_{ij} describing the strength of the interaction. Both parameters are dependent on the species of i & j.

An artefact of the LJ potential is that it has an infinite range, with the force becoming infinitely weak at long distances. Computationally, this is highly demanding as the force must be calculated for every pair of particles within simulation. Thus, a cut-off is introduced to reduce this number of calculations. In this thesis, a cut-off is placed at 12 Å and a smoothing of the potential from 9 Å to reduce complications from a straight cut off.

Charges are treated as point charges and interactions are calculated using Coulombs law:

$$V_C(r_{ij}) = \frac{k_e q_i q_j}{r_{ij}} \tag{2.7}$$

For Coulombs constant, k_e, the particles (i & j), their partial charges (q_i & q_j) and the distance between them (r_{ij}). As V_C decays with increasing r, as opposed to r^6 in the case of the LJ potential, electrostatic effects can be significant over longer distances. This could cause problems in MD, which has a limited dimension. In addition, to apply this rigorously,

it would need to be explicitly defined for every atom pair in the system, resulting in N^2 calculations, for a system of size N. This is dealt with by, first simplifying the system with pair exclusion, which limits non-bonded interaction to only atoms in different molecules, and atoms in the same molecule, but more than 3 covalent bonds apart. Pair-exclusion is also applied to LJ potentials. The long-range effects are dealt with by using the particle-mesh Ewald (PME) method [176–178], which uses the periodicity of simulation to compute the long-range component in Fourier space, while the short-range component is computed in real space. This results in reduction of calculation from N^2 to $N\log(N)$. Although many treatments of electrostatics are available for forcefields, PME is used throughout as it has been shown to out perform cut-off approaches [179].

2.6.4.4 Force Field Type

There are many force fields with different parameterisations of this energy potential, each based on fitting of different experimental data [175]. These can be classified into two types; all-atom force fields, such as AMBER [180] and united-atom force fields, such as GROMOS [181, 182]. United atom force fields combine non-polar hydrogens with carbon atoms, reducing the number of particles in the system. This thesis uses the all-atom force field AMBER99SB-ILDN [183] and united-atom force fields Gromos43a1 [184] and Gromos53a6 [185].

2.6.4.5 Trajectory Calculation

Trajectories are calculated with an iterative methodology.

1. Input initial atomic positions and velocities.
2. Calculate force, F, for each atom from energy potential, V(r).
3. Update configuration using Newton's second law, $F=ma$.

4. Calculate new positions and velocities.

5. Repeat 2-4, until output time step.

6. Repeat 5 for simulation duration.

Force can be determined from the energy potential;

$$F_i = -\nabla_i V(r) \tag{2.8}$$

But also from Newton's second law:

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \tag{2.9}$$

Hence a relationship between force, position and time is derived. In practice, these equations can not be solved analytically due to the complexity of the system. A finite difference approach is taken for time step Δt, using a truncated Taylor series. In this thesis, the leap frog algorithm is used, which is a modified Verlet algorithm based on a forwards and backward Taylor expansions [186].

$$\begin{aligned} r(t+\Delta t) &= r(t) + v(t+\frac{\Delta t}{2})\Delta t \\ v(t+\frac{\Delta t}{2}) &= v(t-\frac{\Delta t}{2}) + a(t)\Delta t \\ v(t) &= \frac{1}{2}\left[v(t+\frac{\Delta t}{2}) + v(t-\frac{\Delta t}{2})\right] \end{aligned} \tag{2.10}$$

All finite differences approaches result in an error, as they truncate what should be an infinite series, and the leap frog algorithm minimises this compared to other approaches [187]. The size of this truncation error is proportional to the time step chosen, such that a large time step could result in rapid increase of energy in the system as high energy atoms overlap. However small time steps make simulations more computationally expensive, so time

step is an important factor to consider in any simulation. To prevent these effects the time step used in these simulations is a 10^{th} of the time of the fastest motion (corresponding to the C-H bond vibration of ~10 fs) [175], hence Δt=2 fs, such that the time-step Δt/2= 1 fs. Fast vibrational bonds are the strongest and result in the least interesting chemistry in a biological system, hence are often restrained. In GROMOS and AMBER this is done with the LINCS method [188]. Further reduction to system complexity is achieved with periodic boundary conditions, which effectively copy the box in all three dimensions, such that as an atom leaves the simulation box in one direction, it will appear in the other, hence produces an approximately infinite solvent system. With the cut-offs mentioned previously, GROMACS uses a boundary condition such that only the particles closest to an interacting box are considered for short-range interactions. Box size should be however ideally be large enough to ensure the soluble protein is not able to interact with itself, which would result in unphysiological strain on the molecule.

2.6.4.6 System Regulation

MD simulates a closed system with a fixed number of particles, N, a fixed volume, V, and a fixed energy, E, due to conservation of energy. As these simulations are used to understand biology system, it is also desirable to fix pressure, P, to atmospheric pressure, and temperature, T, to that relevant for the biological system. As the ideal gas law states:

$$PV = NRT \tag{2.11}$$

for the gas constant, R, this system should be fixed, but in practice, this must be controlled. In this thesis, the Parrinello-Rahman barostat [189] is used to maintain pressure and the V-rescale thermostat [190] is used to maintain temperature. The Parrinello-Rahman Barostat keeps the pressure in a system constant by changing the volume of a system (based on Boyle's law). This is done so by coupling the system to a pressure bath, with the change in

pressure represented as:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p}(P_{bath} - P(t)) \tag{2.12}$$

in which *P(t)* is the pressure of the component at time *t*. P_{bath} is the pressure of the pressure bath & τ_p the coupling constant. For this, the box volume is scaled to allow the constant pressure. Temperature is control by exploiting it relationship with kinetic energy, K.

$$K = \frac{1}{2}\sum m_i v_i^2 = \frac{3}{2}Nk_BT \tag{2.13}$$

where K is the kinetic component, k_B the Boltzmann constant, m is molecular mass, v is the velocity & N the number of molecules. Based on the system, the simplest way to control the temperature is to scale the velocities. In the V-rescale thermostat, the system is coupled to an external heat-bath that is fixed at the chosen temperature T_{bath}. The velocities of each particle are scaled proportionally to the change in temperature of the T_{bath} and the system.

$$\frac{dT(t)}{dt} = \frac{1}{\tau_t}(T_{bath} - T(t)) \tag{2.14}$$

where T is the temperature of the system at time t and τ_t is the coupling parameter of the system to the bath.

2.6.4.7 Water Models

Multiple water models are possible within an atomic force field with a trade off between the models' accuracy and complexity. Within this thesis, the SPC model [191] is used within the GROMOS force field, in which there is no representation of the lone electron pairs on the oxygen, just a negative atom type with the assumption of point charge (this gives an incorrect dipole moment which is altered by changing the bond angle to 109.42°, from

the accurate 104.5°) and TIP3P [192] (which uses three charged points to create 2 dipole moments, but ignores lone pair electrons) within the AMBER simulations.

2.6.4.8 Energy Minimisation

Energy Minimisation (EM) is essential for the start of a MD simulation. This is initially used to find a low-energy structure by reducing steric clashes in the system giving a more stable starting configuration. Two main methods by which this is performed are the steepest decent and conjugate gradient. Both of these methods involve a gradual change in the Cartesian coordinates, reducing the energy of the system to a minimum. As a method, the steepest descent looks for minima using a line-search approach, which is independent of the previous step, thus most effective when the system is far away from the energy minima. This thesis uses both methods of minimisation to ensure energy is minimised as close as possible to 100 $kJmol^{-1}nm^{-1}$ [175].

2.6.5 MD Run Parameters, Software & Analysis

All the simulations performed in this thesis were done so using the GROMACS biomolecular simulation package [193]. Trajectories of *Pf*EMP1 with PFI1780w or spectrin were generated with GROMACS v4.6 and KAHRP and spectrin generated with GROMACS v5.1 gpu [193]. In all cases except β10-14, runs were done in a cubic box, which was 3 nm larger than largest dimension of the molecule. Runs of β10-14 were done in a rectangular box where each dimension was 3 or 4 nm larger than the molecule in each direction. The temperature of simulation in each case was chosen to correspond to experimental NMR or FP binding data, hence run at 37 °C for PFI1780w/ATS-C, and 25 °C for α17/ATS-Core and β10-14/KAHRP. Analysis was conducted either using GROMACS or locally written code. Visualisation and graphics were performed with VMD [194] or PyMOL [156]. The number of salt bridges was determined with VMD based on an oxygen to nitrogen distance

cut-off of 4.5 Å[194, 195]. Convergence of trajectories was determined by monitoring the Cα RMSD from the starting structure. Graphs were plotted using matplotlib [196].

Chapter 3

*Pf*EMP1-PHIST

"Ah, don't pay any attention to them, they're just false alarms. You get a lot of them, especially when you're on the right track "

Hoggle, The Labyrinth

This chapter includes work from the publications;

1. A *Plasmodium falciparum* PHIST protein binds the virulence factor *Pf*EMP1 and comigrates to knobs on the host cell surface. Oberli, A., Slater, L. M., **Cutts, E.**, Brand, F., Mundwiler-Pachlatko, E., Rusch, S., Masik, M. F. G., Erat, M. C., Beck, H.-P., and Vakonakis, I. (2014), *FASEB J.* 28, 4420–4433.

2. *Plasmodium falciparum* PHIST Proteins Contribute to Cytoadherence and Anchor *Pf*EMP1 to the Host Cell Cytoskeleton. AOberli, A., Zurbrügg, L., Rusch, S., Brand, F., Butler, M. E., Day, J. L., **Cutts, E. E.**, Lavstsen, T., Vakonakis, I., and Beck, H.-P. (2016) *Cellular microbiology*.

3.1 Preliminary and Contributions to Data

Initial data showing an interaction between the *Pf*EMP1 ATS and PFI1780w PHIST domain was determined and presented in Mayer *et al.* [65]. The follow up of this work mapped shift perturbations of the full length ATS from *Pf*EMP1 variant PF08_0141 in the presence of the PFI1780w PHIST domain [39], suggesting that the C-terminal disordered region of the ATS preferentially interacts (Fig. 3.1A). Structural work on the PHIST domain of PFI1780w alone was performed by Leanne Slater, in parallel to my work on the ATS interaction, resulting in a crystal structure, SAXS and AUC data, NMR backbone assignments and shift perturbations of the PFI1780w PHIST domain in the presence of the ATS (Fig. 3.1B, C, D). Binding studies of the PFE1605w PHIST domain and the ATS-C were performed by Leanne Slater and Martin Masik, and binding studies of the PFE1605w C-terminal region and Band 3 were performed by Madeleine Butler, under my supervision.

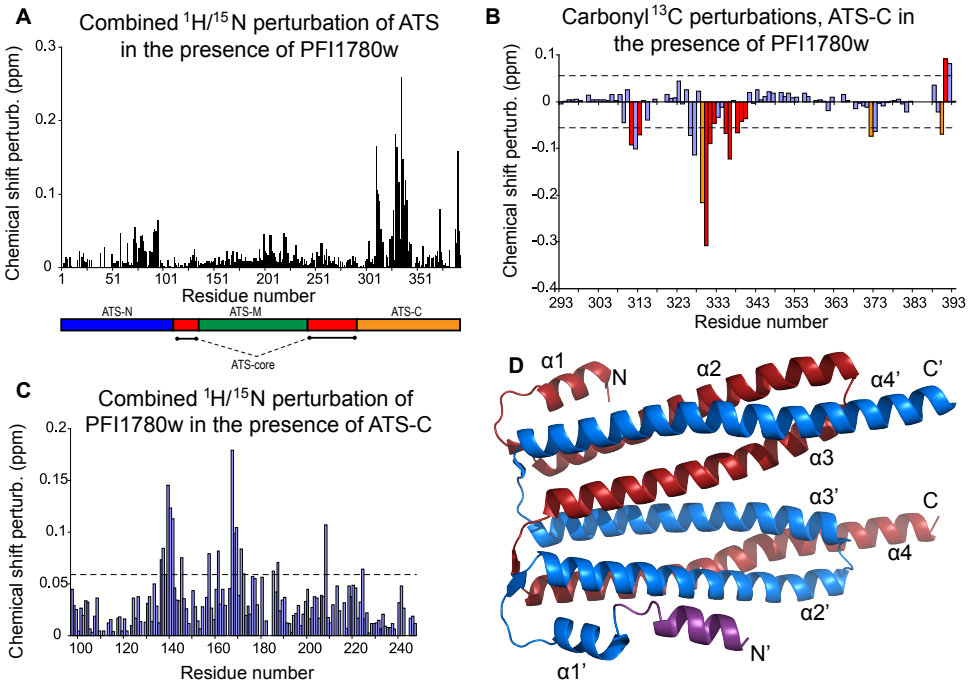


Fig. 3.1: ***Pf*EMP1 ATS binds PFI1780w.** **A**: ^{1}H-^{15}N HSQC of ^{15}N ATS titrated with the PFI1780w PHIST domain, with major perturbations in the ATS-C. **B**: Carbonyl shift perturbations of *Pf*EMP1 ATS-C in the presence of PFI1780w. **C**: ^{15}N PFI1780w PHIST domain HSQC shift perturbation in the presence of ATS-C. **D**: Crystal structure of the PFI1780w PHIST domain (4JLE). Adapted from Oberli *et al* [39]

3.2 Localising the ATS-C/PFI1780w Interaction

The cytoplasmic domain of *Pf*EMP1, known as ATS, is primarily disordered [65]. The ATS comprises a small folded core (ATS-Core) and flexible segments at its N-terminus (ATS-N), middle (ATS-M) and C-terminus (ATS-C) (Fig. 3.2A); this ATS architecture is conserved across *Pf*EMP1 variants [65]. As shift perturbations of ATS-FL were predominantly located at the protein C-terminus (Fig. 3.1A), fluorescently labelled recombinant ATS-FL and ATS-C were incubated with fixed concentrations of recombinant PFI1780w PHIST domain and variations in fluorescence polarization (FP) that result from slower tumbling of the labelled protein upon binding were determined. The binding curve produced was fit, resulting in

K_ds of 22±3 μM and 16±2 μM, for the ATS-FL and ATS-C respectively (Fig. 3.2A). The ATS-C bound with similar affinity as the ATS-FL, suggesting that ATS-C was sufficient for binding. Furthermore, the ATS-C was able to compete with ATS-FL for PFI1780w, resulting in a similar competition K_d of 11±7 μM (Fig. 3.2A).

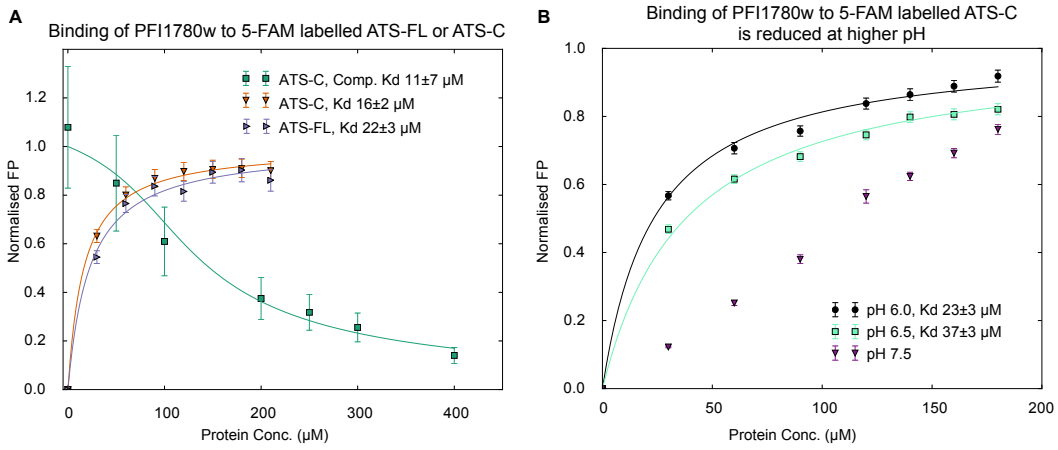


Fig. 3.2: Optimisation of ATS/ PFI1780w FP binding assay. **A**: FP binding assay of PFI1780w with ATS-C (purple) and ATS-FL (orange) and FP competition assay, using 5-FAM labelled ATS-FL, 60 μM PFI1780w with increasing concentrations of ATS-C (green). **B**: The effect of pH on binding of 5-FAM labelled ATS-C to PFI1780w. K_d could not be fit at pH 7.5.

As there are histidine residues near the region of large shift perturbations of PFI1780w (H127 and H167), pH sensitivity was tested by determining the affinity at pH 6, 6.5 and 7.5. The fit K_ds were 22.8±2.9 μM and 37.4±2.7 μM, for pH 6 and 6.5 respectively, while the K_d at pH 7.5 could not be fit (Fig. 3.2B). This suggested that affinity was indeed affected by pH, and that optimal binding was at reduced pH.

To independently verify and further refine this interaction ^{1}H-^{15}N HSQC titrations were performed with ^{15}N labelled PFI1780w PHIST domain and varying truncations of the ATS-C. Addition of the ATS-C residues 309-392 to ^{15}N PFI1780w result in specfic shift perturbations (Fig. 3.3), which can be quantified at different concentrations of ATS-C and fit to determine the K_d (Fig. 3.4A).

Initial experiments were conducted at pH 6.5 with 100 μM of ^{15}N labelled PFI1780w,

resulting in fit K_ds of 340±46 μM and 75±16 μM for ATS-C 309-345 and ATS-C 309-392, respectively (Fig. 3.4A), indicating that ATS regions additional to residues 309-345 are required for full affinity. Further experiments were conducted using 50 μM ^{15}N PFI1780w at pH 6, where the binding affinity is stronger. In these conditions ATS-C 309-392 had a fitted K_d of 23±5 μM (Fig. 3.4B). Removal of regions 345-365 and 345-381 to make ATS-C 309-392Δ(345-365) and 309-392Δ(345-381) resulted in fit K_ds of 20±4 μM and 70±30 μM respectively (Fig. 3.4B).

Together this suggests that the ATS region required for wildtype PFI1780w affinity lie within residues 309-345 and 365-392. Attempts were made to crystallise the complex between PFI1780w and ATS-C 309-392Δ(345-365), but despite rigorous screening and optimising of spherolite forms, no diffracting crystals were achieved.

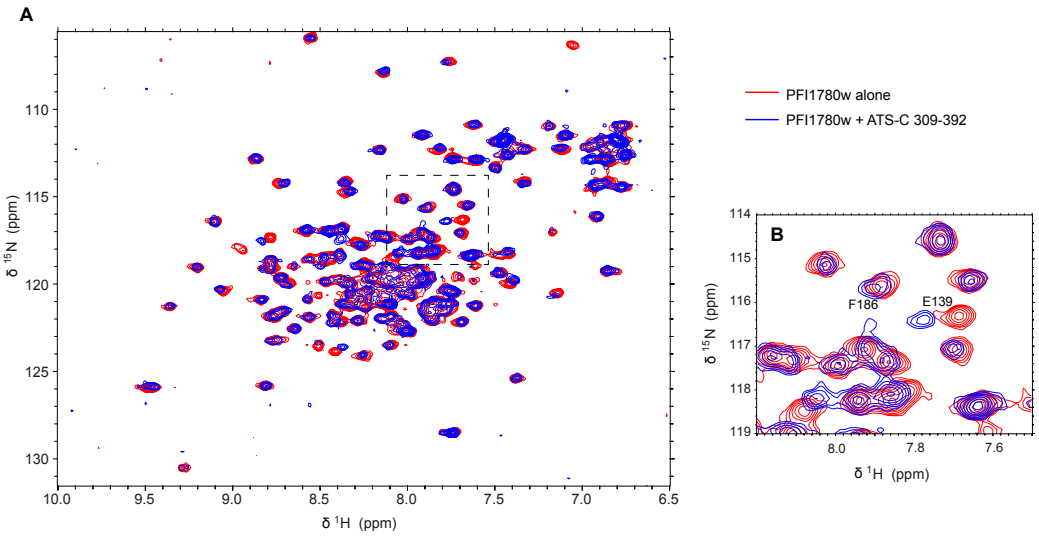


Fig. 3.3: HSQC spectra of PFI1708w PHIST domain. A: ^{1}H-^{15}N HSQC spectra of 100 μM of PFI1780w at pH 6.5, alone (red), and with 200 μM of ATS-C 309-392 (blue). **B**: An enlargement of the area in the dashed box of **A**, with some of the peaks used for K_d fitting labelled.

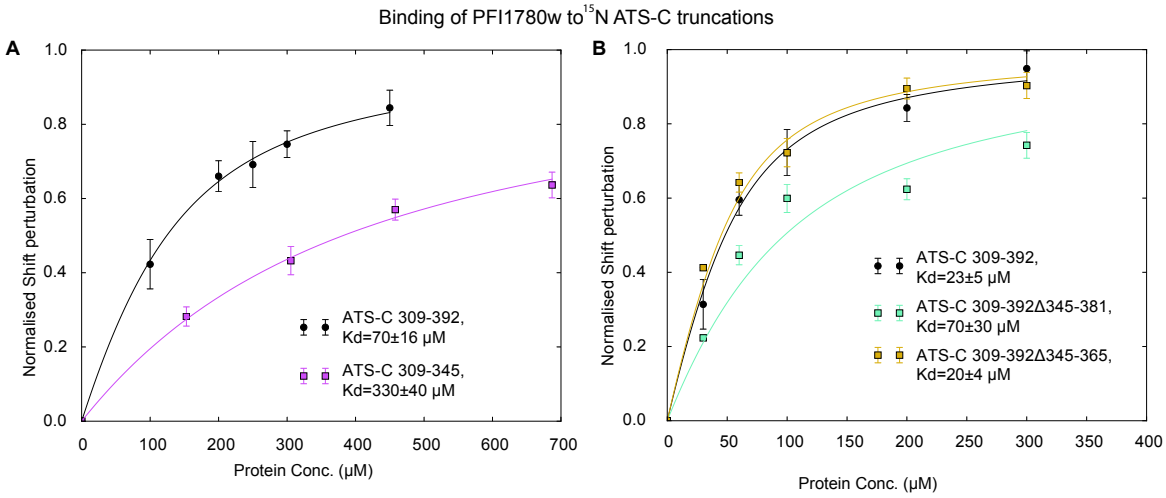


Fig. 3.4: The binding affinity of ATS-C truncations, determined with NMR ^{1}H-^{15}N HSQC titrations. **A**: ATS-C 309-392 (fitting and averaging using peaks E139, H167, V137, R161), ATS-C 309-345 (fitting and averaging using peaks E139, H167, V137, M160, R161). Experiments performed at pH 6.5. **B**: ATS-C 309-392 (fitting and averaging using peaks: E139, E157, N108, R161, H167), ATS-C 309-392Δ(345-381) (fitting and averaging using peaks: E101, F186, N108, not assigned), 309-392Δ(345-365) (fitting and averaged using peaks: N108, F168, E139, E157 and R161). Experiments performed at pH 6.0.

3.3 Modelling and MD of PFI1780w

The PFI1780w crystal structure was resolved as a helix swap dimer at a resolution of 2.3 Å (Fig. 3.1D), with helix 3 and 4 defining the dimerisation interface. However, analytical ultracentrifugation data and small angle X-ray scattering suggested that the size and shape in solution was not consistent with a dimer [39] (Fig. 3.5A, B). Furthermore, shift perturbation mapping indicated that the interaction interface with ATS-C also involved helix 3 and 4 (Fig. 3.1C), suggesting that the dimer is a crystallographic artefact. To investigate how a monomeric structure may behave in solution, monomeric PFI1780w was modelled using the crystal structure of the dimer and the homology modelling software, MODELLER [165]. The helical regions were not altered but the residues 190-194 were modelled with the loop modelling functionality (Fig. 3.5C). This monomer model was used to generate an

ensemble, setting residues 85-97 as flexible and to fit the SAXS data, resulting in a good fit (χ^2 of 1.9) (Fig. 3.5D,E).

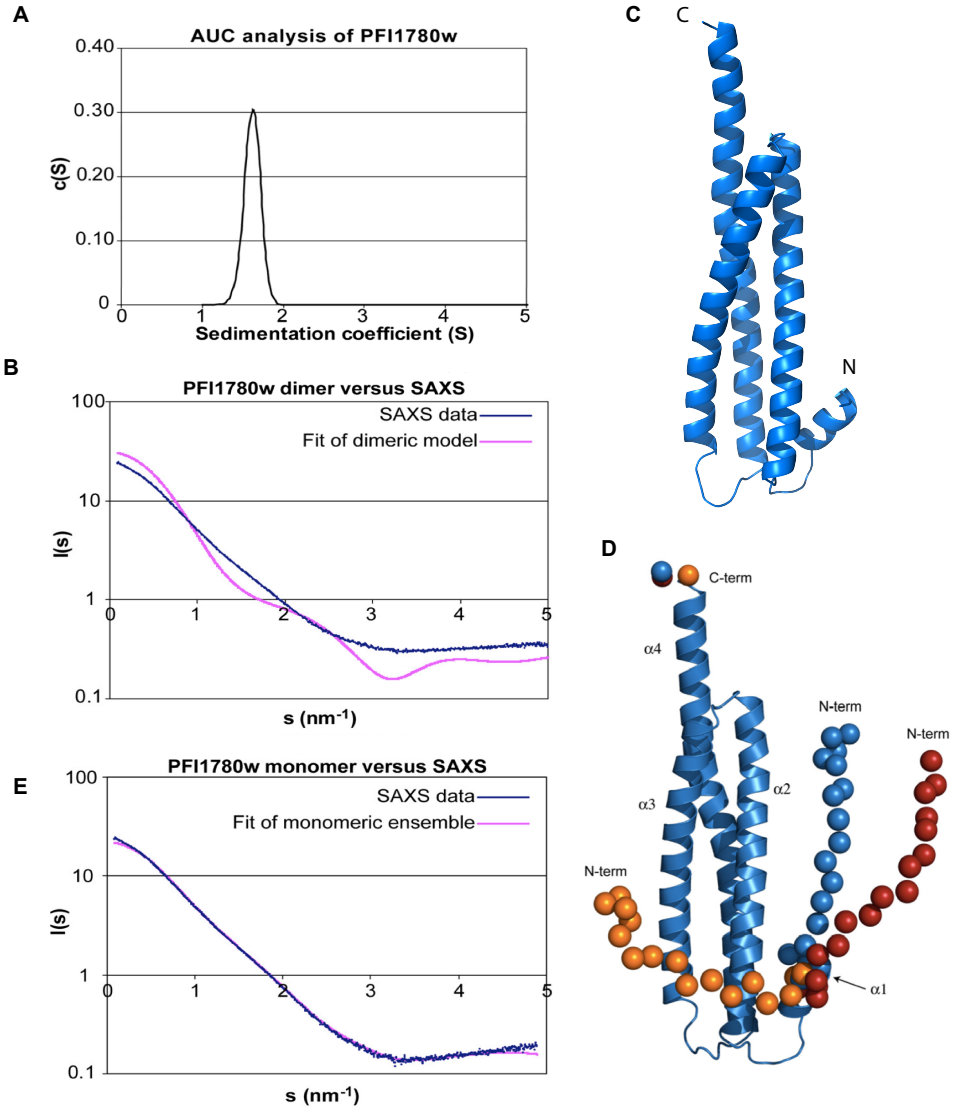


Fig. 3.5: PFI1780w is a monomer is solution although crystal structure suggests a dimer. **A**: AUC velocity experiment of PFI1780w residues 85-247. Analysis of the sedimentation yielded a single, narrow distribution at 1.65 S with a best fit friction ratio of 1.61 with a molecular weight 19,919 Da. The monomeric molecular weight is 20,071 Da. **B**: SAXS data, fit with the dimeric crystal structure, showing a clear lack of fit. **C**: Monomeric model, built from crystal structure using MODELLER. **D**: The monomeric model had residues 85-97 modelled as an ensemble of dynamic spheres. **E**: Ensemble of monomeric models fit SAXS data well. **A**, **B**, **D** and **E** adapted from Oberli *et al* [39].

Molecular dynamics runs were performed on the monomeric model of the PFI1780w PHIST domain with three different force fields; Gromos43a1 [184], Gromos53a6 [185] and AMBER99SB-ILDN [183], running 100 ns simulations in triplicate. To evaluate structural stability and dynamics, the Cα root-mean-square difference (RMSD) from starting confirmation, the number of hydrogen bonds, the Cα root-mean-square fluctuations (RMSF), the secondary structure and accessible hydrophobic surface area were determined (Fig.3.6A-E). In all cases the structure remained reasonably stable, with a consistent number of hydrogen bonds, and a decrease in solvent exposed hydrophobic surface area. Gromos43a1 resulted in the most consistent preservation of secondary structure, with only a slight kink being introduced in helix 4 due to contacts with helix 1, and also buried the hydrophobic surface area more efficiently. For these reasons I chose Gromos43a1 for use in further investigation.

For comparison, a 100 ns simulation was set up for the PFI1780w crystallographic dimer with the Gromos43a1 force field and the equivalent regions analysed (Fig. 3.7A,B,C). While run 1 of the monomer resulted in a larger RMSD and larger RMSF than the dimer, both run 2 and 3 were comparable to the dimer. This suggests that while the dimer may be slightly more stable in simulation, it is only marginally so.

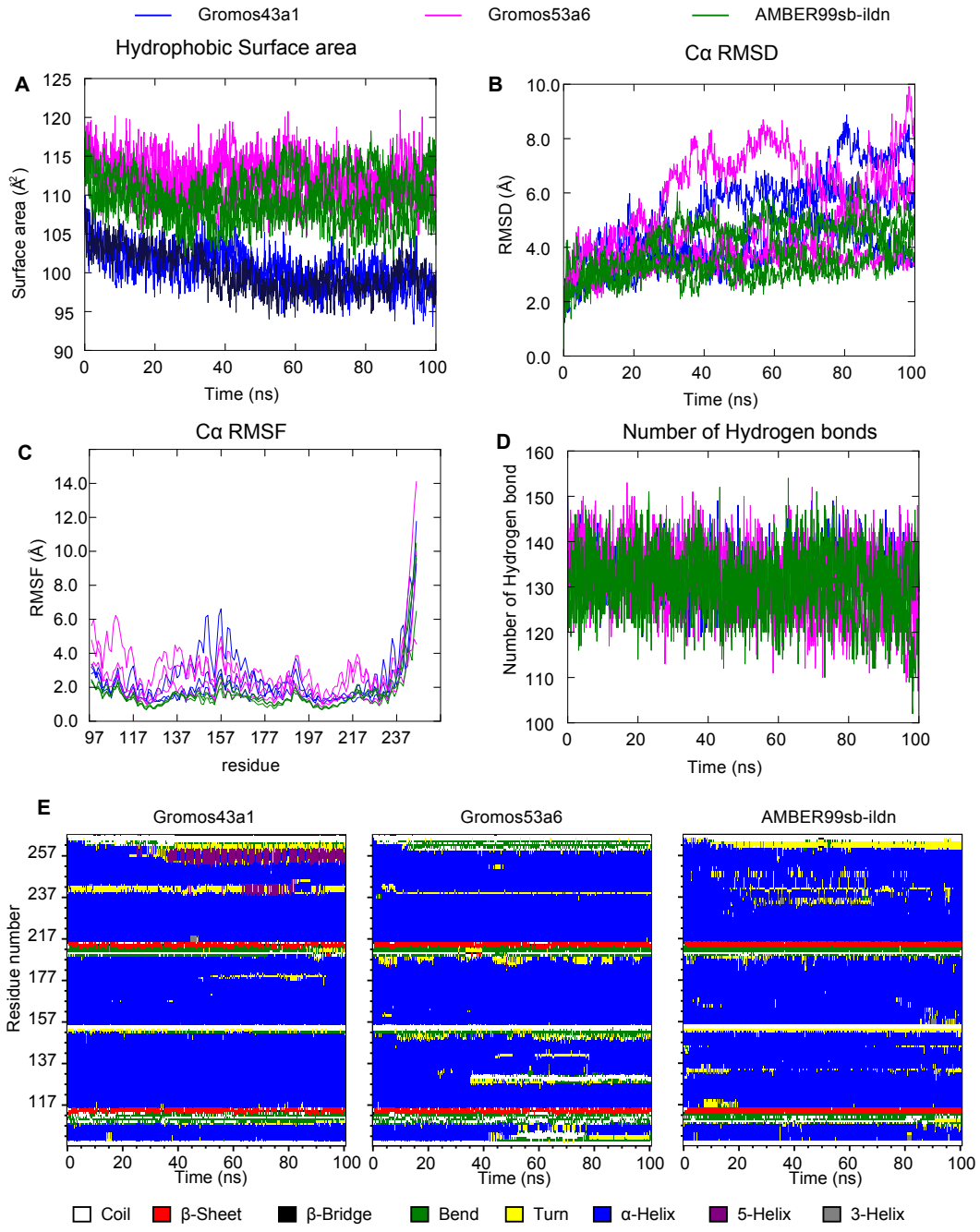


Fig. 3.6: Molecular Dynamic simulation of the monomeric model of PFI1780w using three different atomic force fields. **A**: Exposed hydrophobic surface area as a function of time, with Gromos43a1 having the lowest exposed hydrophobic surface area. **B**: The change in Cα RMSD is not significantly different between force fields. **C**: The change in Cα RMSF is comparable for all force fields, with AMBER99sb-ildn having the least fluctuations. **D**: The total number of hydrogen bonds is comparable between force fields. **E**: Secondary structure analysis shows regions of secondary structure converting from helical to disordered, which seems minimum using the Gromos43a1.

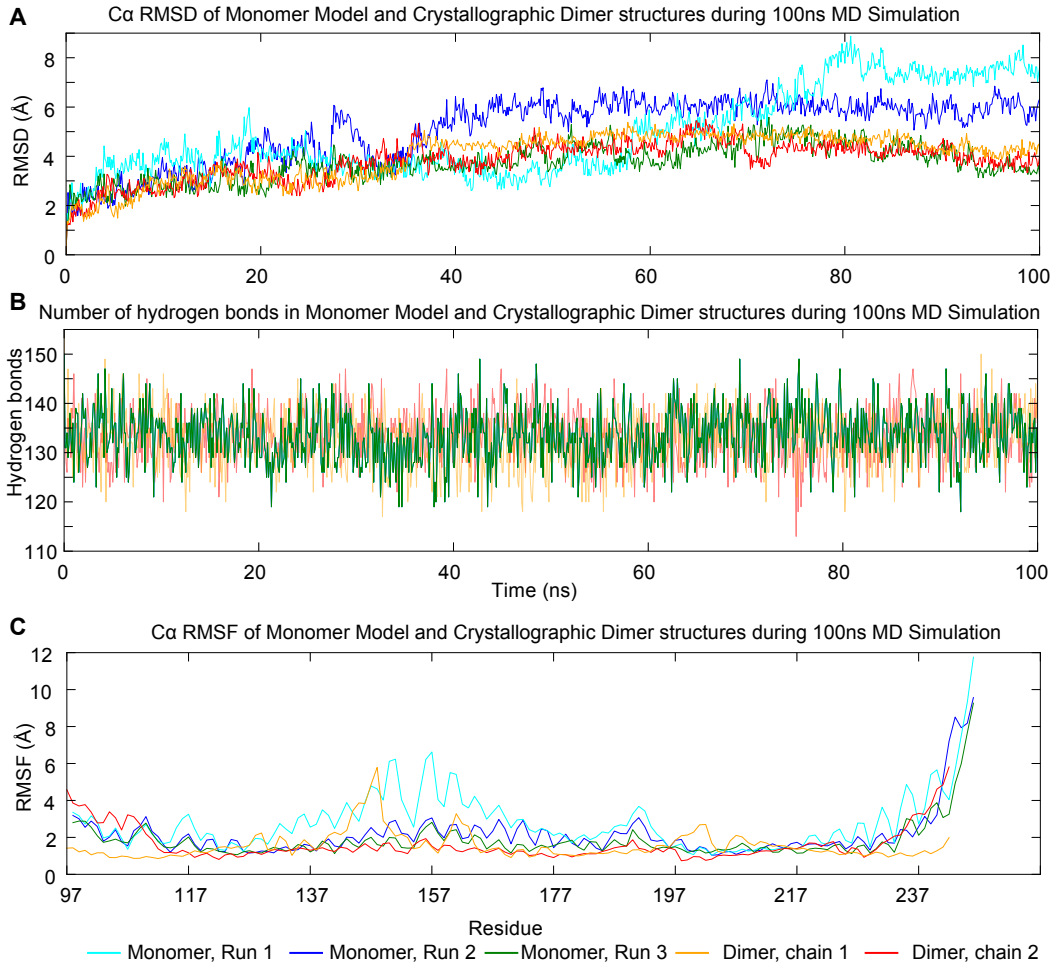


Fig. 3.7: Comparison between monomer and dimer of PFI1780w during MD. A: The change in Cα RMSD is slightly larger during some monomer runs as compared to the dimer. **B**: The total number of hydrogen bonds remains consistent and comparable. **C**: The Cα RMSF generally shows consistent variation, with one loop region being more dynamic in monomer run 1. In both cases, simulations were performed with Gromos43a1 force field.

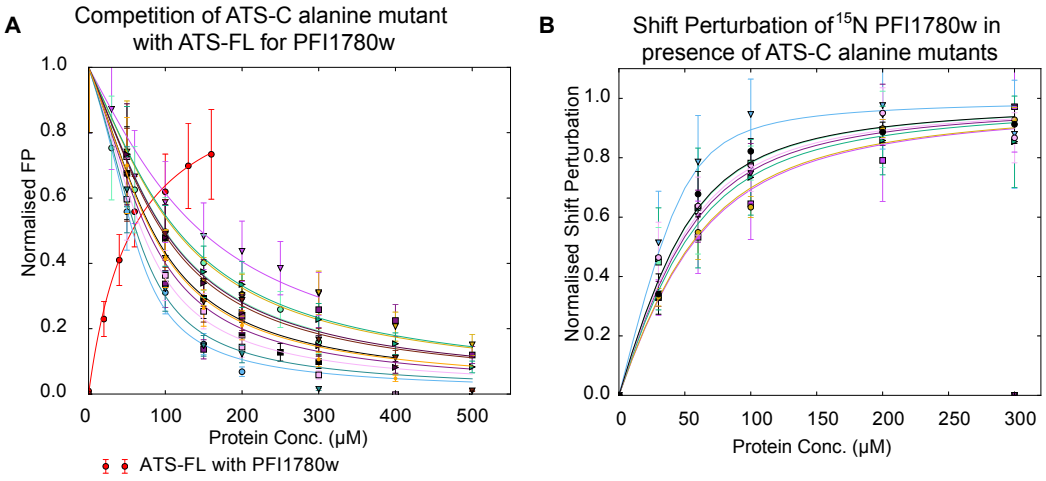
3.4 Alanine Substitution Analysis

Having established an FP assay capable of determining the K_d based on competition with ATS-FL, this was used to assay PFI1780w binding of alanine substitution variants of ATS-C. Residues of ATS-C that had large perturbations in chemical shifts upon PFI1780w binding were chosen for alanine substitution mutagenesis. Eleven of these were successfully cloned

and sequenced, and electro-spray mass spectrometry was performed to verify the correct molecular weight after purification (Appendix A).

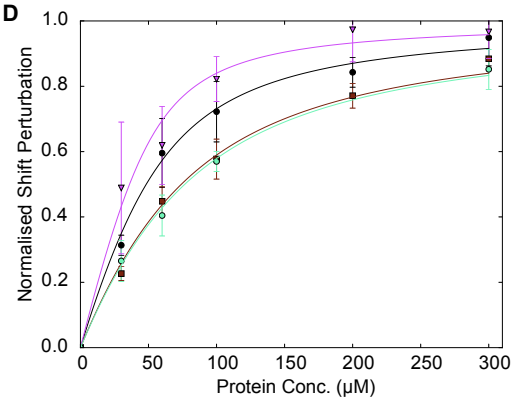
All ATS-C alanine substitution FP binding assays were done at pH 6.5, with 60 μM PFI1780w and 1 μM 5-FAM labelled ATS full length (Fig. 3.8A and C). The largest fold change from ATS-C WT was just 2.4, by W310A, with I330A and L331A both having a roughly 2-fold reduction in affinity. Additionally, ATS-C D332A bound with 2.5 fold higher affinity. All other mutations did not have significantly different K_ds. NMR ^{1}H-^{15}N HSQC titrations were also performed on most substitutions using 50 μM of ^{15}N PFI1780w at pH 6 (Fig. 3.8B) and broadly agreed with the results from FP, showing ATS-C W310A, I330A and D332A all having roughly 2-fold effects with other substitutions having little effect of affinity.

Alanine substitution was also performed on PFI1780w based on residues with large chemical shift perturbation. Again the largest change in affinity was 2 fold, with both PFI1780w F168A and Q171A reducing in affinity by 2-fold, while H167A increased in affinity 2 fold (Fig. 3.8D).



C ATS-C alanine subsitution binding affinity

	Mutant	FP Kd±SD(μM)	NMR Kd±SD (μM)	NMR peaks used
	WT	21±2	17±3	N108, S124, E139,E157, H167, NA
	W310A	51±7	29±10	N108, E139, E157, M160, R161, F163
	I313A	29±5	16±6	N108, E139, E157, NA
	M327A	18±4	20±3	F123, S124, E139, E157, M160, R161
	I330A	35±6	28±5	S124, E139, H167, S216, F229
	L331A	37±6	23±6	N108, F123, S124, V137, E139, E157, R161, S216
	D332A	8±2	7±2	S124, V137, E139, E157, R161, F163, S227
	E335A	27±5		
	D336A	14±3	19±4	S124, V137, E139, M160, R161, H167
	I338A	10±3		
	Y339A	28±5		
	Y340A	20±4		



PFI1780w alanine subsitution binding affinity

	Mutant	NMR Kd±SD (μM)	NMR peaks used
	WT	23±5	N108, E139, E157, R161, H167
	H167A	11±6	E157, L182, S124
	F168A	49±7	E139, V137, L189, D219, F163
	Q171	52±12	S124, E139, NA

Fig. 3.8: Alanine substitution mutants. **A**: FP competition assays using 5-FAM labelled ATS-FL, 60μM of PFI1780w and increasing concentrations of variant ATS-C. **B**: FP results were compared with ^{1}H-^{15}N HSQC titration for some of the mutants. **C**: Fit K_ds from FP and NMR data. **D**: Alanine substitution of PFI1780w tested with ^{1}H-^{15}N HSQC titrations, with fit K_ds.

3.5 Paramagnetic Relaxation Enhancement

PRE experiments are performed in the presence of a paramagnetic label (free radical), which reduces the intensity of NMR resonances in a distance dependent manner (Fig. 3.9A). The paramagnetic label, in this case 3-maleimido-PROXYL (2,2,5,5,-tetramethyl-1-1pyrrolidinyloxy), is linked to cysteine residues introduced into ATS-C constructs via formation of thioether bonds using a maleimide reactive group. ATS-C paramagnetically labelled at D328C or E346C was added to ^{15}N labelled PFI1780w and a ^{1}H-^{15}N HSQC spectrum was recorded. The free radical was then quenched by addition of ascorbic acid, another HSQC spectrum was recorded and the relative intensities of each peak determined.

The PROXYL label at position D328C in ATS-C had the most severe effect, with peaks corresponding to PFI1780w residues F123, S124, K136, E139, M160, F163, H167, G172, I182 and G191 no longer being detectable, (Fig. 3.9A). When the label was at position E346C of the ATS-C it did not attenuate as many peaks on the predicted binding face; however, the intensity of PFI1780w residues F163, H167 and G172 were severely affected (Fig. 3.9B). This suggests that while ATS residue D328 is at or very close to the binding site, residue E346 is further away but is on the same side of helix 3 of PFI1780w. These data supports the interface as suggested by NMR shift perturbations and truncation data, which suggests the region of the ATS required is between 309-345, while 346-365 is less critical, and likely further away from the interface.

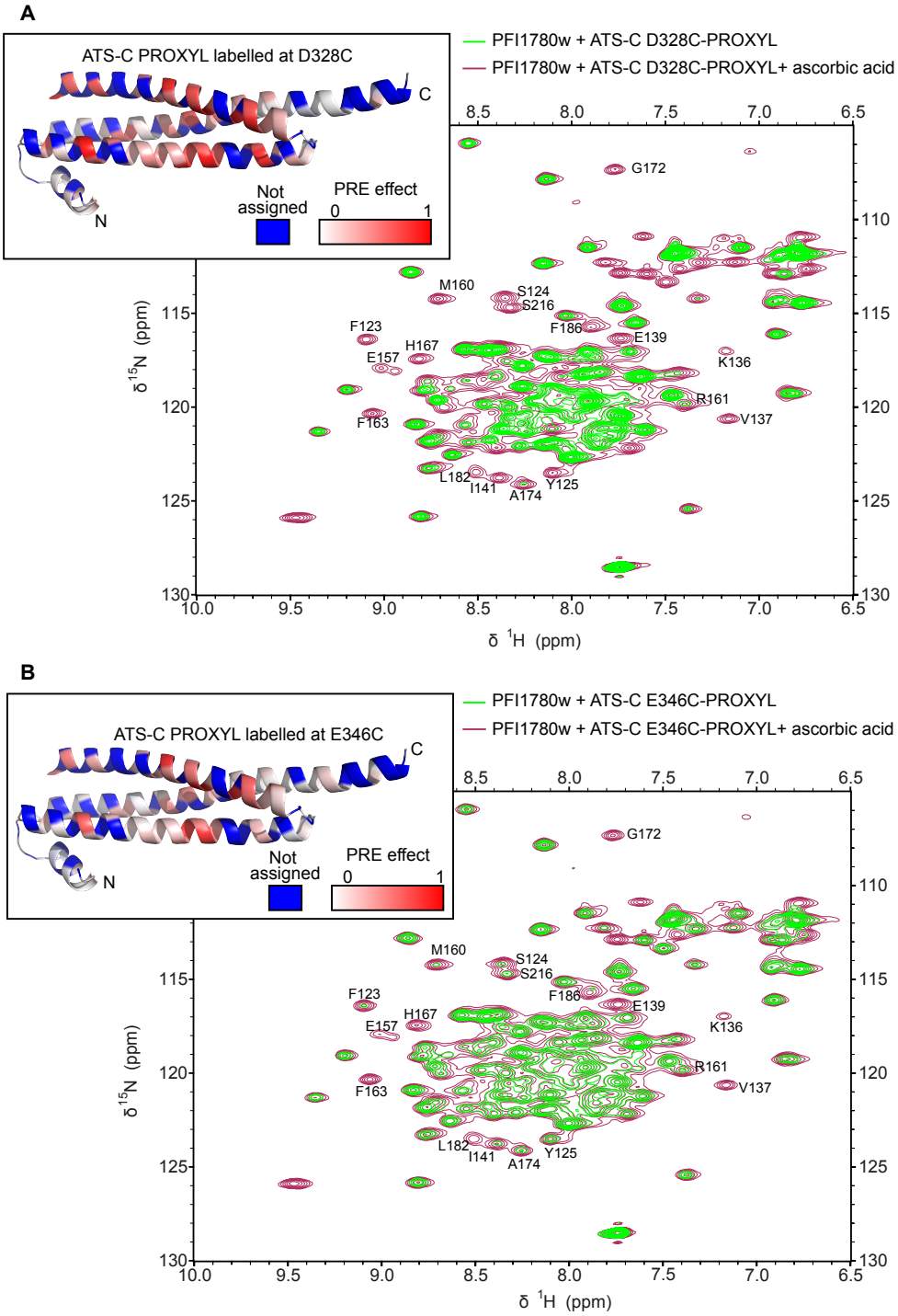


Fig. 3.9: PRE experiments ^{15}N PFI1780w HSQC spectra with ATS-C with an active PROXYL label (green) or PROXYL label quenched with ascorbic acid (magenta). Insert shows these effects mapped onto the monomeric structure. The PRE effect is (1 minus the ratio of non-quenched and quenched label). **A**: Label at position D328. **B**: Label at position E346.

3.6 Computational Modelling of the Complex

3.6.1 Generation of CS-Rosetta Models

Although the ATS-C is predicted to be predominantly disordered, there were negative carbonyl shift perturbations for residues 308-317 and 325-341 in the presence of PFI1780w, (Fig. 3.1B), which suggests the formation of a more extended structure [197]. To explore the possible conformation that the ATS-C could be taking in the bound form, CS-Rosetta [170] was performed on the regions of the ATS-C that were significantly perturbed in NMR.

Chemical shifts for the ATS-C fully bound to the PFI1790w PHIST domain were extrapolated from the shifts of 50 μM ^{15}N labelled ATS-C in the presence of 200 μM PFI1780w. These shifts were analysed with TALOS+ [198] for a prediction of secondary structure and dynamic residues, which suggested the formation of β-strands (Fig. 3.10A). As a large part of the ATS-C was predicted to be dynamic, these regions were removed and structure prediction was performed on ATS-C fragments 304-332 and 306-346. This corresponds to regions that have two and three β-strands, respectively, as predicted by the TALOS+ and as suggested by up-field carbonyl shifts. The sequences for these fragments, along with the corresponding chemical shifts, were then used to select a library of 3 and 9 residue fragments from the CS-Rosetta fragment library, containing proteins from the protein data bank, where the NMR shifts are known. Fragments were assembled into decoy structures using a standard Rosetta *ab inito* relax protocol; generating 8000 decoys for the predicted two stranded ATS-C fragment (residues 304-332) and 12,800 structures for the predicted three stranded fragment (residues 306-346). Should Rosetta prediction be successful, a plot showing the Rosetta scores of all decoys should decrease with the pairwise RMSD of all decoys with the lowest energy one, demonstrating a 'funnel' to the lowest energy structure, which should correspond to the most accurate model (Fig. 3.10B, [170]). Before funnelling can be performed, sets of similar structures must be determined out of the large pool of decoy

structures. This is done by clustering, which determines the pairwise RMSD of structures and groups them such that each cluster contains structures within a given RMSD radius. The lowest energy structure of the largest cluster is then used to determine the pairwise RMSD for all other structures and plot against the score to produce a funnelling plot.

Decoys of the predicted two stranded fragments were clustered at a radius of 3 Å, resulting in two large clusters of 305 and 171 decoys. Examination of the lowest energy decoys in clusters, resulted in three distinct fold types: two or three stranded anti-parallel β-sheets and two strand parallel β-sheets, with the two stranded anti-parallel sheets being the most common. Funnelling was not observed and this was likely due to the unstructured loop regions being divergent in structure; increasing the RMSD and resulting in divergent scores. However, overlays of decoys with two stranded anti-parallel β-sheet, reveals that the stranded regions overlay very well, hence clustering was performed again, at 1 Å, including only the residues from the sheet (residue 309-314 and 321-326). This resulted in a large cluster of 362 members, with lowest energy structure of this cluster, models 3722, being anti-parallel β-sheets. Decoy structures were clipped to remove the unstructured loops, leaving the β-sheet (309-314 and 321-326) and rescored using Rosetta scoring functionality. The pairwise RMSD with all other decoy structures were determined for model 3722 and the score vs. RMSD plot funnelled (Fig. 3.10C, E). Model 1940, which also has an anti-parallel β-sheet, but with a different register, also demonstrated funnelling (Fig. 3.10D, F). Hence, CS-Rosetta was able to predict two potential two strand β-sheet models of the ATS-C.

A similar approach was taken for the predicted three strand fragment. Initial clustering at 3 Å resulted in large clusters, with the largest having 1313 members. The lowest energy models in these clusters ranged from two to four β-stranded sheets, in a mix of anti- parallel and parallel conformations, but anti-parallel sheets were the most common. Re-clustering was performed at 1 Å using residues from the three and four anti-parallel β-strands (where the four strands range from residues 311-314, 319-325, 328-334, 337-344). While four β-

stranded decoys resulted in large clusters, with the largest 719, after clipping, rescoring and plotting the score vs RMSD, no funnelling was observed. Three-strand clipping resulted in much larger clusters, the largest being of 2064 and 1444 members. Plotting score vs RMSD of the lowest energy decoy of the second largest cluster, model 247, after clipping and rescoring, demonstrated funnelling (Fig. 3.10 G, H), giving a very promising prediction of three strands of the ATS-C.

3.6.2 MD Simulations of CS-Rosetta Models

Molecular dynamic simulations were run on CS-Rosetta predicted structures of ATS-C to determine their dynamics in solution. Simulations of the two β-stranded model 3722 and the four-stranded model 247 were run for 100 ns with the Gromos43a1 force field in triplicate. In these simulations the disordered loops and N and C-terminal tail regions were included to determine how they behaved in solution, (Fig 3.11 A, B). Overlays of model 3722 suggest the two strands predominately stay together, but analysis of the secondary structure show the β-sheet confirmation is not preserved in all simulations (Fig. 3.11C, E, G). Overlays of model 247 show that the interface between the second and the third strand came apart, while the first and second, and the third and fourth strands remained together (Fig. 3.11D). Secondary structure analysis suggested that the third and fourth strands remain stable, while the first two strands are less stable, in agreement with the two-stranded model (Fig. 3.11F, H). Cα RMSD, hydrogen bond number and Cα RMSF were also examined as metrics for stability, and both the two stranded model, 3722 and the four stranded model, 247 demonstrated dynamic behaviour with fluctuating RMSD and hydrogen bond number (Fig. 3.12A-D). The RMSF for model 247 was consistently lower than for model 3722 (Fig. 3.12E,F), suggesting the additional strands did improve stability, but any conclusions should be treated carefully as the Cα RMSD suggests that the system has not reached equilibrium (Fig. 3.12A, B).

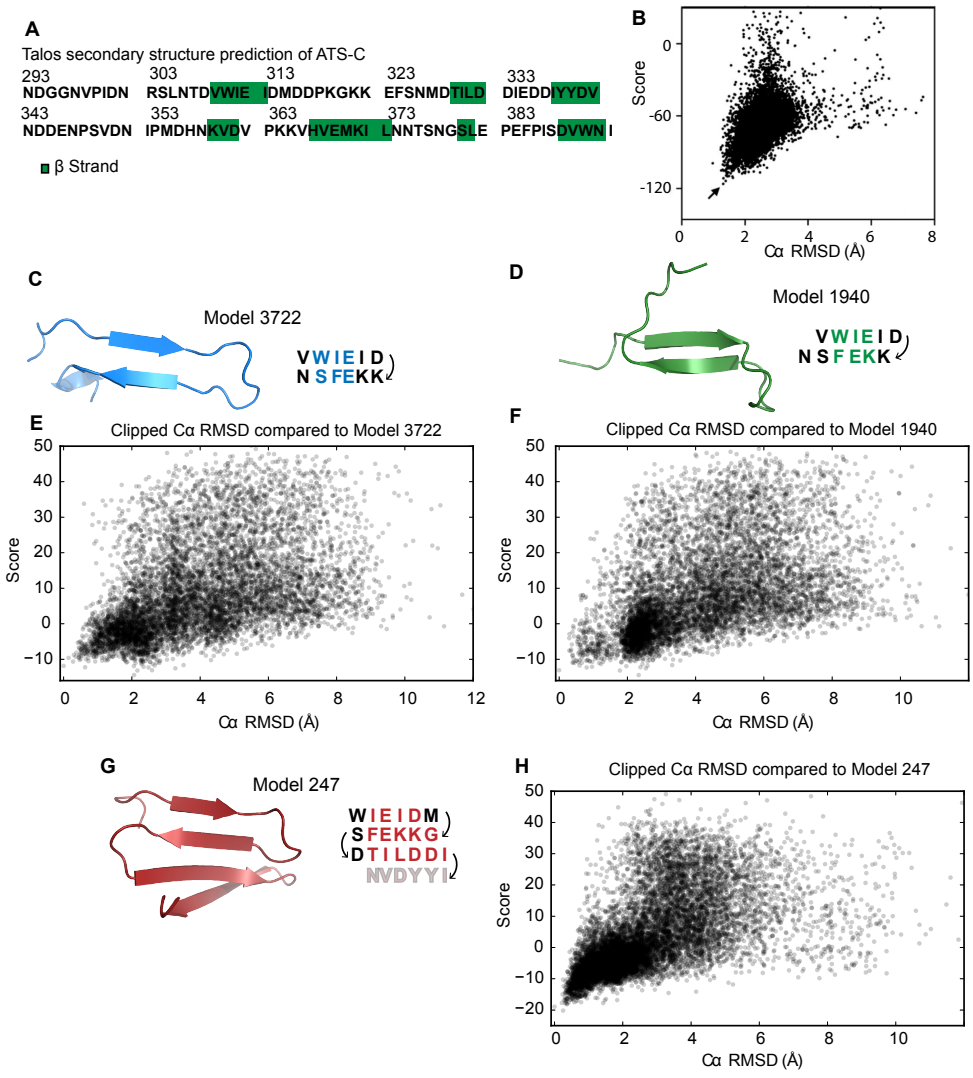


Fig. 3.10: CS-Rosetta structure prediction of the ATS-C. A: TALOS+ structure prediction of β strands based on chemical shifts of ATS-C in the presence of PFI1780w, corrected for saturation. **B**: Example of a funnelling plot, showing lower scores are coupled with lower RMSD, from Shen *et al.* [170]. **C, D**: Two-stranded models predicted by CS-Rosetta, with funnelling plots calculated for clipped stranded region in **E** and **F** respectively. **G, H**: The model generated from the predicted three strand fragment, resulted in structures being predicted to have 4 anti-parallel β strands and its respective funnelling plot calculated for clipped three stranded region.

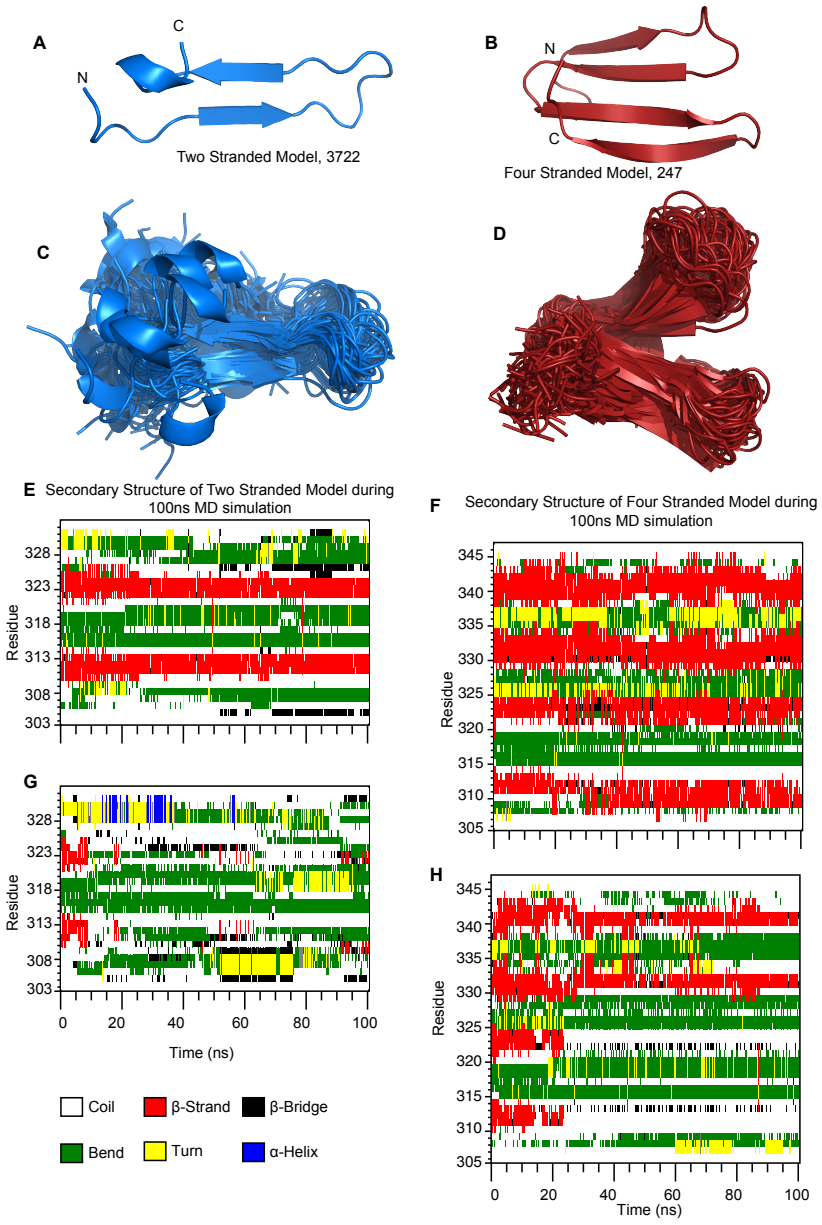


Fig. 3.11: Molecular Dynamics of CS-Rosetta Models. A, B: The initial structures of the two and four stranded model, respectively. **C, D**: Overlay of structures every 1 ns during the 100 ns simulation, with structures aligned as in **A** and **B**. **E-H**: Secondary structure analysis throughout MD simulation. The two-stranded model remained stable in some simulation runs, **E**, but not in others, **G**. This was also true in the four-stranded model, with the structure remaining a β sheet in some runs, **F**, but strand 1 and 2 destabilising in other runs, **H**.

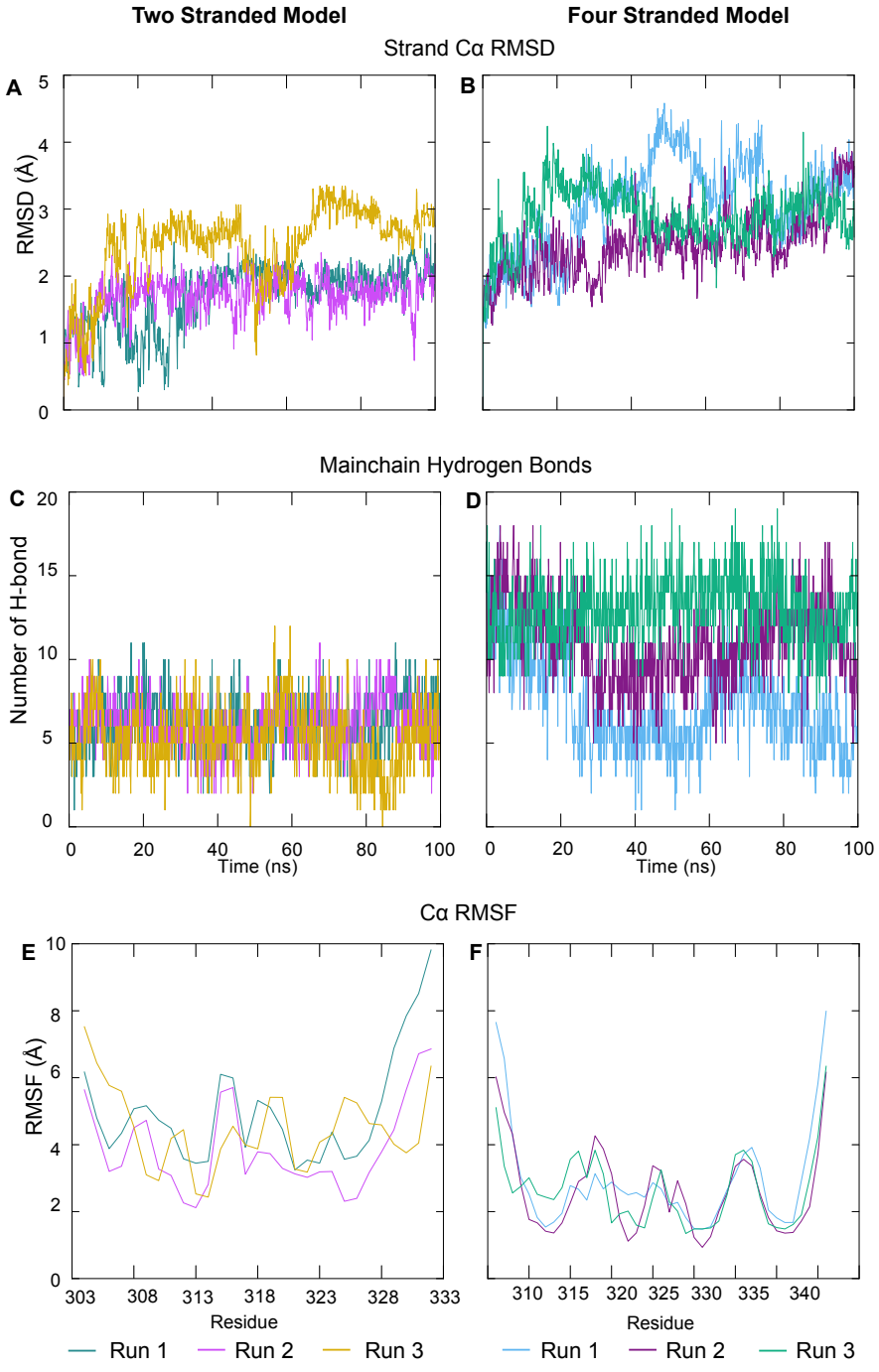


Fig. 3.12: Molecular Dynamics of CS-Rosetta. A,B: Cα RMSD of the stranded region of the two stranded model **A**, is slightly less than the four stranded model. **C, D**: The number of mainchain hydrogen bonds of the four strand model, **D**, is always greater than the two stranded model, **D**. **E, F**: The Cα RMSF for the two stranded model, **E**, is greater than the four stranded models, **F**.

3.6.3 MD Simulations of Model 247 Docked to PFI1780w Surface

While the CS-Rosetta predicted ATS-C structures were not stable in simulations, the extended conformation of the ATS-C only occurs when in complex with the PFI1780w PHIST domain, hence this interaction may provide additional stabilisation. To test this I used the chemical shift perturbations as docking restraints with the HADDOCK server [171, 172] to dock model 247 to the monomeric structure of PFI1780w (docking restraints in Appendix B). This docking resulted in the prediction of just one docked conformation, where the β-sheet lays parallel to helix 2 and 3 (Fig. 3.13A). The interface in the docked structure was extensive, with hydrogen bonds predicted between ATS-C residues K322, D336, D337, D32, D341 and PFI1780w residues R131, K135, E137, E138, K143, K164, E178, K179, as well as the backbone, and a large hydrophobic cluster between ATS-C I311, F324, L331, Y339 and PFI1780w F142, F163, H167 and F168 (Fig. 3.13 B,C).

MD simulations were run on the docked complex, for 100 ns with Gromos43a1 in triplicate. Comparison of the complex structure after 100 ns simulation (Fig. 3.13 D) shows differing orientations and conformations of the ATS on PFI1780w in each run. Despite the dynamic behaviour of the ATS-C on the surface of PFI1780w, the strands were more stable in complex than alone. Secondary structure analysis shows preservation of 4 strands (Fig. 3.13E), hydrogen bond number remains more consistent and RMSD and RMSF are roughly equivalent or lower in the docked structure compared to 247 alone, when the same regions were used for analysis (Fig. 3.14).

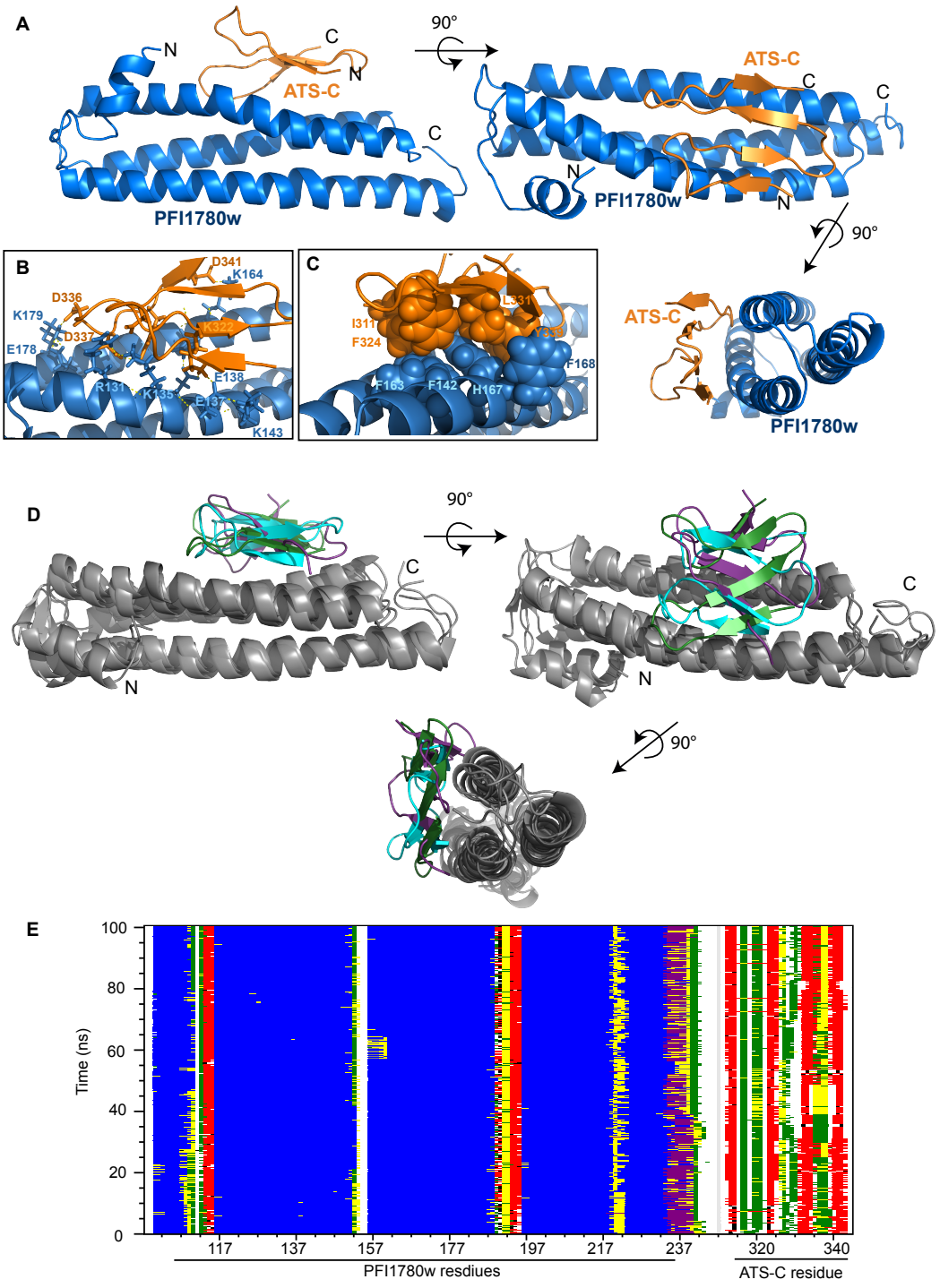


Fig. 3.13: ATS-C docked to PFI1780w PHIST domain. **A**: The model of the ATS-C/PFI1780w complex docked with HADDOCK. **B**: The hydrogen bonds and ionic interactions, and **C**: the hydrophobic interactions made in the docked complex. **D**: An overlay of the three final conformations of the docked structure, showing the ATS-C moves over the surface of PFI1780w. **E**: Despite the ATS-C being dynamic, secondary analysis indicates the structure remains in a β-sheet conformation.

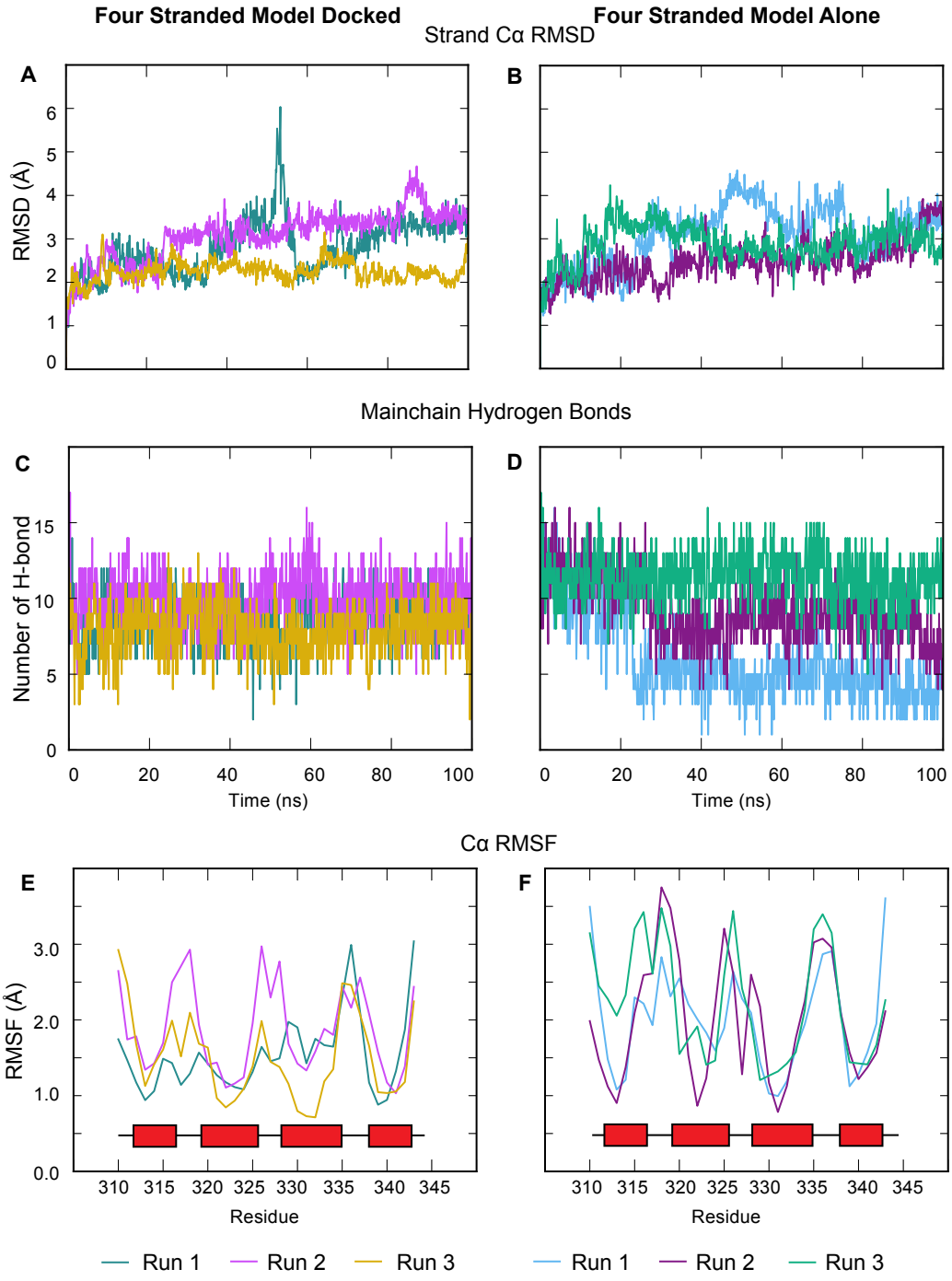


Fig. 3.14: ATS-C docked to PFI1780w MD analysis. **A, B**: The Cα RMSD of the ATS-C when docked to PFI1780w, **A**, is similar to ATS-C alone, **B**. **C, D**: The number of mainchain hydrogen bonds is more consistent in the docked structure, **C**, than the ATS-C alone, **D**. **E, F**: The Cα RMSF of the docked structure, **E**, is smaller than alone, **F**. The red boxes indicate approximate location of β-strands in starting conformation.

3.7 General ATS-PHIST Interactions

Another PHIST protein, PFE1605w, was proposed to interact with the ATS-C. The PHIST domain of PFE1605w was tested for binding to the ATS-C with a FP binding assay, resulting in a fit K_d of 2.1$\pm$0.3 μM, in the same conditions as PFI1780w (Fig. 3.15A). Other PHIST domains were also tested (PFD1170c, MAL8P1.163 and MAL8P1.4) and showed no detectable binding, suggesting the ATS-C does not bind all PHIST domains. NMR ^{1}H-^{15}N HSQC experiments carried out with ^{15}N labelled PFI1780w PHIST domain, suggested that unlabelled PFE1605w PHIST domain was able to compete with PFI1780w for binding to the ATS-C (Fig. 3.15B). Hence, PFE1605w binds the same region of the ATS-C as PFI1780w, possibly via the same mechanism, despite low sequence identity between PFI1780w and PFE1605w of just 18%.

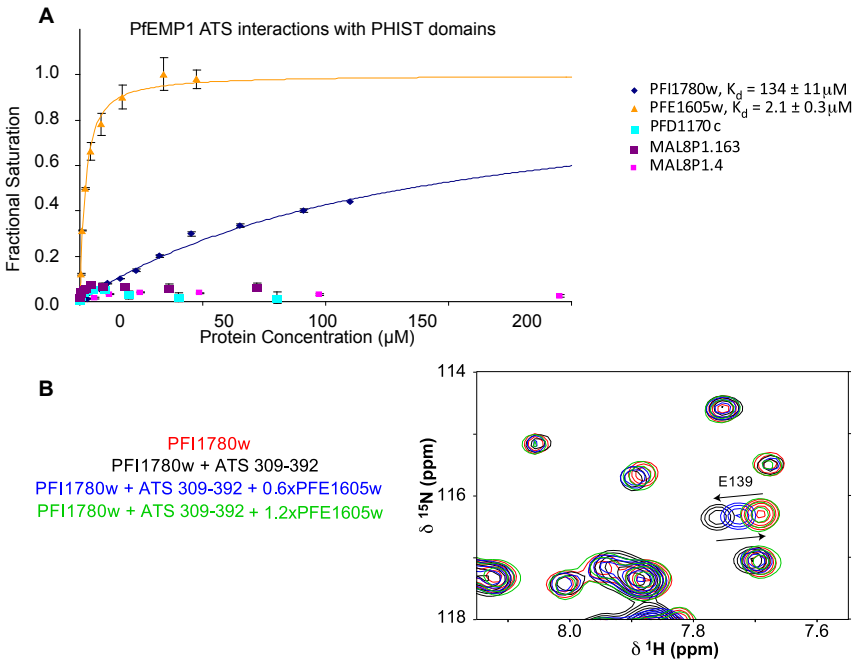


Fig. 3.15: ATS-C binding to PFE1605w. **A**: 5-FAM labelled ATS-FL binds both PFI1780w and PFE1605w, but did not bind other PHIST domains from PFD1170c, MAL8P1.163 and MAL8P.4. **B**: Expansion of overlaid spectra from 50 μM of ^{15}N PFI1780w PHIST domain alone (red), in the presence of 50 μM ATS residues 309–392 (black), after addition of 0.6x stoichiometric ratio of PFE1605w PHIST domain (blue), or after addition of 1.2x stoichiometric ratio of PFE1605w (green). PFE1605w addition reverses PFI1780w perturbations caused by ATS binding, suggesting that the 2 PHIST proteins compete for the same ATS binding site. Figure adapted from Oberli *et al*, 2014 [39].

I attempted to crystallize PFE1605w, both in the presence and absence of ATS-C fragments, but was unable to obtain diffracting crystals.

To further show this interaction was conserved amongst ATS variants and both PFI1780w and PFE1605w, FP binding experiments were performed with five other ATS-C variants, resulting in binding in all cases except to the ATS-C PFF0845c, which is truncated to just 30 residues (Fig. 3.16 A,B). Although binding affinity varied from ~30-70 μM for PFE1605w and ~60-130 μM for PFI1780w.

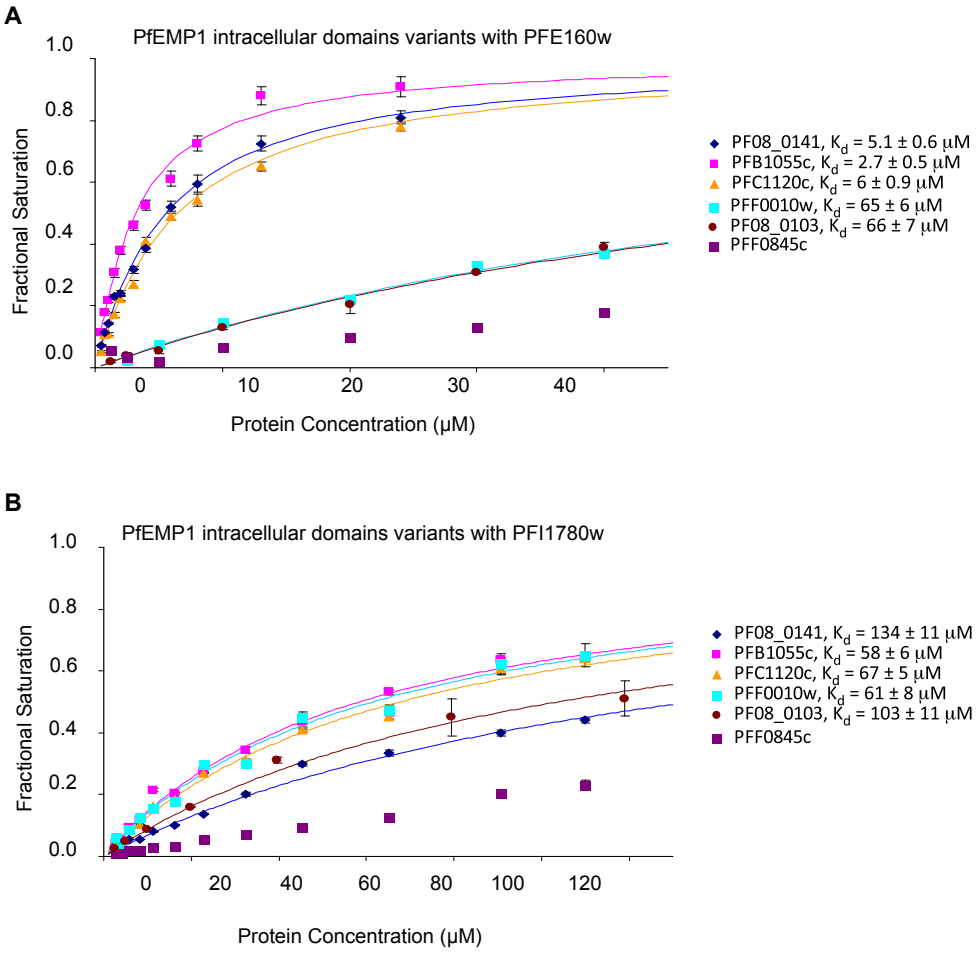


Fig. 3.16: ATS-C variant binding to PHIST proteins PFI1780w and PFE1605w. A, B: 5-FAM labelled ATS variants bind PHIST domains of PFI1780w and PFE1605w, respectively, with different affinities. Figure adapted from Oberli *et al*, 2014 [39].

3.8 Interaction Between PFE1605w and Band 3

An immunoprecipitation assays performed by collaborators on fixed infected erythrocytes followed by mass spectroscopy suggested that PFE1605w could interact with the cytoskeletal proteins band 3, ankyrin, band 4.2 and band 7. FP assay binding assays were performed using the C-terminal, disordered region (residues 300-528) of PFE1605w labelled with 5-FAM. This region of PFE1605w bound to the band 3 cytoplasmic domain with a fitted K_d

of 1.5±2 μM [199] (Fig. 3.17A). The band 3 used for these binding studies was found to be dimeric, with size exclusion chromatography followed by multi angle light scattering (SEC-MALS) predicting a mass of ~80 kDa, where the monomeric mass is 42.6 kDa. Hence, PFE1605w is able to bind band 3 dimers.

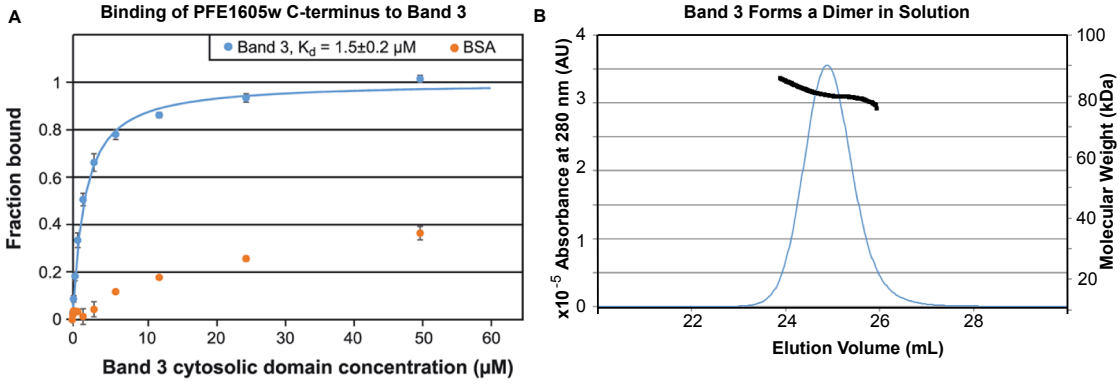


Fig. 3.17: PFE1605w binds band 3 dimers. A: The 5-FAM labelled C-terminal disordered region of PFE1605w binds the cytoplasmic domain of band 3. BSA (bovine serum albumin) was used as a negative control. Adapted from Oberli *et al*, 2016 [199]. **B**: The band 3 cytoplasmic domain is a dimer in solution. SEC-MALS of band 3, suggests a molecular weight of ~80 kDa, where the monomeric mass of band 3 is 42.6 kDa.

3.9 Discussion

3.9.1 MD Simulation

One of the major issues addressed in this work was whether a monomeric model of the PFI1780w PHIST domain based on the crystallographic dimer could be used as a structural model for the monomer observed *in vitro*. This was explored computationally, by modelling the monomer and performing MD simulations using three different force fields. A wide range of force fields are available to describe proteins in MD simulations and, although it is frequently the topic of review and exploration, there still lacks a clear consensus as to which is the best, and in what conditions. For the simulations of the monomer of PFI1780w, I trialled a range of different force fields, namely: Gromos43a1 [184], Gromos53a6 [185]

and AMBER99SB-ILDN [183]. This ensured any inferred stability was independent of force field choice. Previous reports suggest that Gromos force fields perform well, particularly 43a1, and that they do not over sample helical conformations and can fold hairpins [200, 201], but these united atom force fields are often neglected from analysis in favour of all-atom descriptions. Lange *et al* [202] directly compared Gromos and all-atom Amber force fields and found that hydrogen bonds were often underestimated and simulations longer than 100 ns resulted in non-native conformations being sampled in both force fields. Simulations of PFI1780w generally agreed with these conclusions, with Gromos43a1 and AMBER99SB-ILDN reproducing the PFI1780w monomeric stability, with helical regions being maintained through the simulation, while Gromos53a6 resulted in the formation of strand kinks, as hydrophobic residues, previously packed in the dimeric form were repacked into the hydrophobic core at the cost of the rest of the structure (Fig. 3.6).

Simulations involving the ATS-C were performed using Gromos43a1, as Cino *et al.* [201] tested the ability of force fields to fold hairpins, and found Gromos43a1 out-performed AMBER, hence it should offer the best chance for the ATS-C to be stable as a β-sheet. Even using this force field, the ATS-C had limited stability in isolation, with a hairpin of residues 305-331 unfolding. Increasing this to a four stranded ATS-C β sheet, of residues 305-346, seemed to increase the stability slightly, with Cα RMSF being lower. This was further reduced when docked to the PFI1780w PHIST domain, hence PFI1780w could provide a scaffold to enable the ATS-C to form the extended structural conformation inferred from the NMR data. The simulations of the docked structure suggested that the ATS-C remained bound, but the conformation after 100 ns varied greatly, suggesting the ATS-C was dynamic on the surface of PFI1780w, which is consistent with PRE data. Furthermore, the location of residues D328 and E346 in the docked model is consistent with the PRE effects, as D328 is at the centre of the interface, while E346 is at the C-terminal end of the four stranded segment of the ATS-C, pointing away from the main interface.

3.9.2 A Mutation Resistant Interface

Attempts to disrupt this interface with mutagenesis only resulted in 2-fold changes in affinity and, generally, ATS variants were able to bind both PFI1780w and PFE1605w, despite sequence variation. This suggests the interface between the ATS-C and PHIST domains is robust against mutation. It seems to achieve this by having a large and dynamic interface; a common mechanism for interactions of IDPs, which will be discussed in more detail later in this thesis. This mechanism is supported by ATS-C truncations that resulted in reduced affinity and MD simulations showing a large, but dynamic interface. Although this interface seemed to be dynamic, some residues are still important for affinity, for both PFI1780w and PFE1605w. The hydrophobic cluster residues are well conserved, and mutation of some of these residues, such as L331A in the ATS-C and F168A in PFI1780w, each had small effect on binding. The hydrogen bonding or ionic interactions however seem to be far more variable in the different ATS sequences, with PF08_0103 still binding PFI1780w despite having different residues in most of these positions than PF08_0141.

3.9.3 PHIST Protein Function

In this chapter, I demonstrated that the ATS-C interacts with PFI1780w and PFE1605w. There is limited data on how PFI1780w functions, as no attempts to produce a gene knockout have been successful [65]. It has been shown to localise to the erythrocyte periphery and is not co-transported with *Pf*EMP1 [39], hence, it could interact with *Pf*EMP1 at the membrane. More data are available on PFE1605w, which has been shown to localise to the erythrocyte periphery, is co-transported with *Pf*EMP1 and knockouts result in reduced adhesion to CD36 by ~60% and ICAM-1 ~30% [40, 199]. This suggests that PFE1605w plays a significant role in cytoadhesion. FP binding assays suggested the C-terminal region of PFE1605w interacts with the cytoplasmic domain of band 3, hence this interaction would be responsible for peripheral localisation. As PFE1605w is able to bind both band 3 and

the ATS-C, it could anchor and/or localise *Pf*EMP1, enabling *Pf*EMP1 to adhere more efficiently. Band 3 forms oligomers *in vivo* [203], hence clusters of band 3 may be able to recruit many PFE1605w molecules, which in turn helps to cluster and localise *Pf*EMP1. But where the effects of the PFE1605w knock out vary depending on which *Pf*EMP1 variant is expressed, the need for PFE1605w in localising and clustering *Pf*EMP1 is likely to be variable and possibly controlled at a gene level.

Chapter 4

*Pf*EMP1-Spectrin

"Do. Or do not. There is no try"

Yoda, Star Wars

This chapter includes work from the publication;

Interactions of *P. falciparum* KAHRP and *Pf*EMP1 with spectrin suggest a model for adherent knobs on the infected erythrocyte surface, **Cutts, E. E.**, Laasch, N, Reiter, D. M., Trenker, R., Slater, L. M., Stansfeld, P. J. and Vakonakis, I, (2016), to be submitted.

4.1 Preliminary data

Initial work on the project was done by Leanne Slater and Rapheal Trenker. Leanne cloned and expressed the initial sub-fragments of spectrin. Rapheal collected FP titrations and NMR data of spectrin regions that bound the ATS with the strongest affinity. I further localised this interaction, and collected biophysical data to determine a structural model. In all cases I reanalysed the preliminary data, repeated, and added additional experiments in order to attain complete data sets.

4.2 *Pf*EMP1 binds spectrin

4.2.1 Localisation of binding

The ATS, is primarily disordered, comprising of a small folded core (ATS-Core) and flexible segments at its N-terminus (ATS-N), middle (ATS-M) and C-terminus (ATS-C) (Fig. 4.1A) [65]. As the ATS binds components of the erythrocyte spectrin–actin–band 4.1 complex [60], the ability of fluorescently labelled ATS from *Pf*EMP1 variant PF08_0141 to bind spectrin was tested. For this purpose spectrin sub-fragments composed of structural repeats N-terminus-5, 6-11, 12-16 and 17-C-terminus from the α chain (referred to as αN-5, α6-11, α12-16 and α17-C) and N-terminal domain, 1-4, 5-9, 10-14 and 15-C-terminus from the β chain (referred to as βN, β1-4, β5-9, β10-14, and β15-C) were recombinately produced (Fig. 4.1B). The ATS displayed weak affinities for most spectrin sub-fragments; however, it bound more strongly to an α spectrin construct spanning from repeat 17 to the C-terminus (α17-C, K_d = 59$\pm$6 μM, Fig. 4.1C, D).

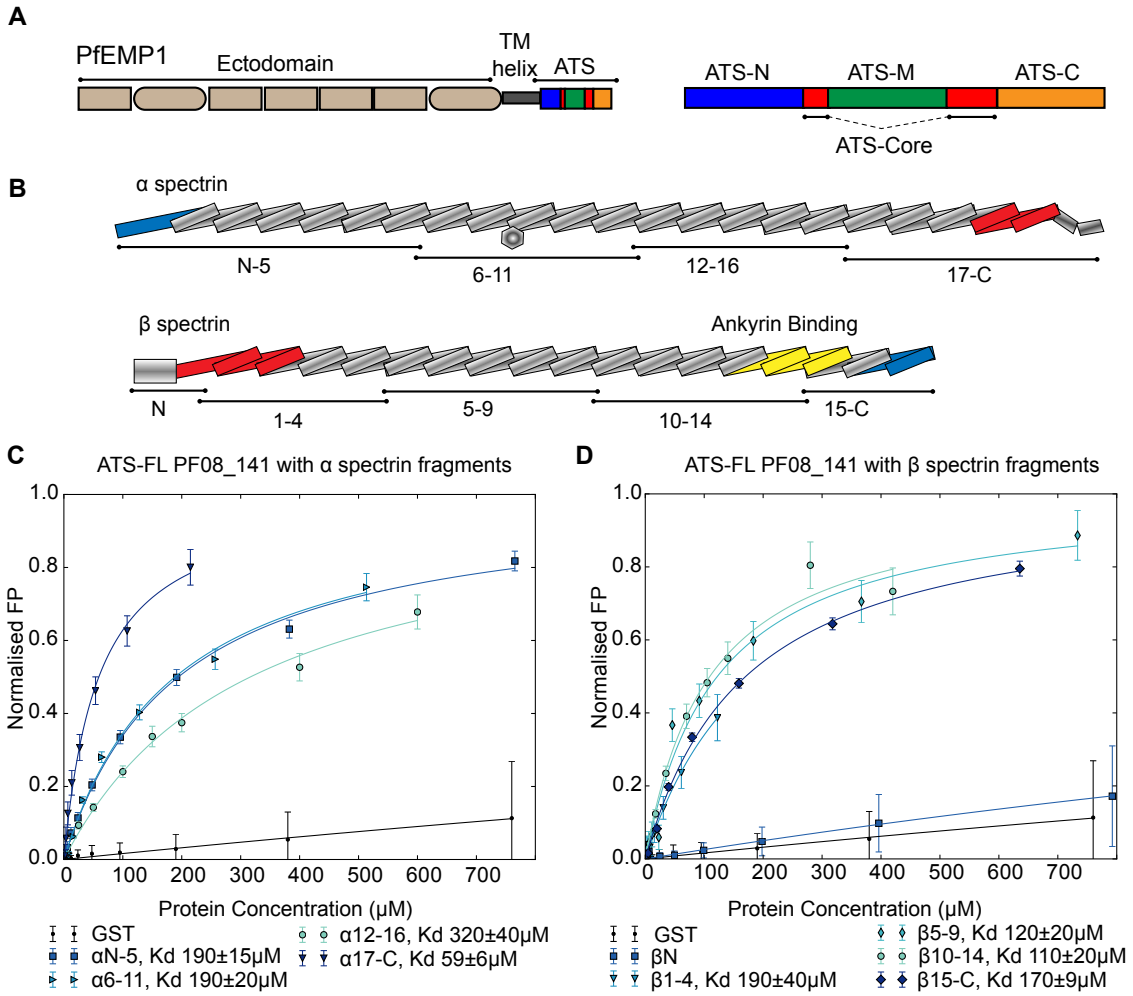


Fig. 4.1: *Pf*EMP1 ATS binds spectrin. **A**: Schematic of *Pf*EMP1 (left), with the ATS-regions (right). **B**: The structural repeat domains of spectrin. The domains in blue are involved in dimerisation, red is involved in tetramerisation and yellow binds ankyrin. **C, D**: FP binding assays for α and β spectrin, respectively, each using 1 μM of 5-FAM labelled ATS.

In order to independently validate the ATS–spectrin association, and to further localise the interaction epitope, NMR ^{1}H-^{15}N HSQC experiments were performed. ^{1}H-^{15}N HSQC spectra were collected of ^{15}N-labelled ATS-N, ATS-Core, ATS-M and ATS-C, titrating in unlabelled spectrin sub-fragments that showed binding in FP titrations (αN-5, α6-11, α17-C, β5-9, β10-14 and β15-C). In these experiments, each peak of the ^{1}H-^{15}N HSQC spectra corresponds to a N-H bond, hence the spectra contains approximately one peak per residue

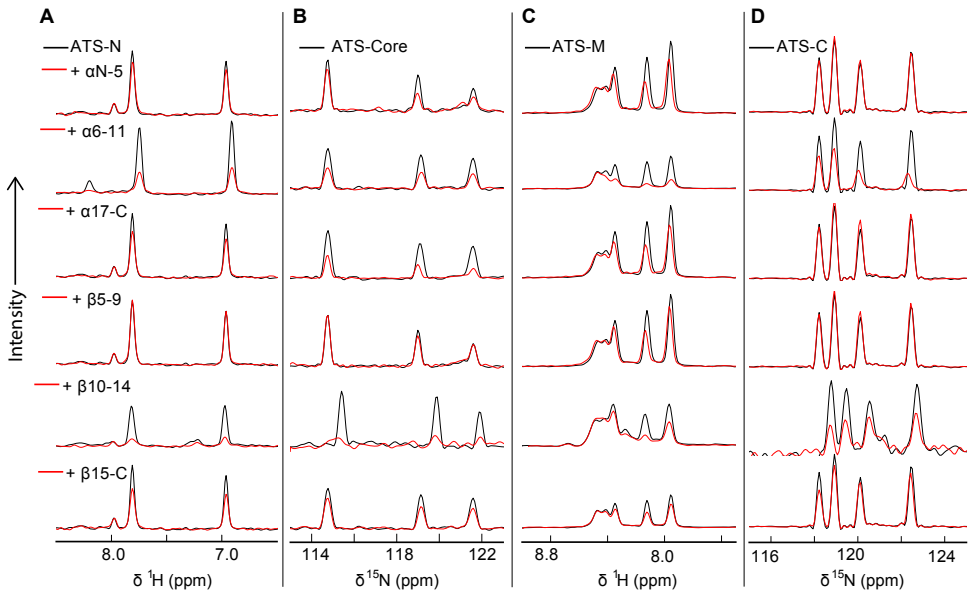


Fig. 4.2: *Pf*EMP1 ATS binds spectrin. 1D slices from ^{15}N-HSQC of ATS-N, -M, -Core and -C in the absence (black) and presence of spectrin sub-fragments (red). **A**: ATS-N, slices at ^{15}N 114.2 ppm. **B**: ATS-Core, slices at ^{1}H 9.2 ppm. **C**: ATS-M, sliced at ^{15}N 119.4 ppm. **D**: ATS-C, sliced at ^{1}H 7.95 ppm. All ATS fragments were used at 100 μM and all spectrin sub-fragments were used at 200 μM, except for α17-C, which was used at 100 μM. Full spectra overlays can be found in Appendix C.

of the ATS regions tested. As these spectrin fragments are large, the complex produced tumbles slower, with a larger correlation time, τ_C than ATS alone, resulting in faster transverse relaxation rates. This results in less NMR signal, and reduced peak intensities in the event of binding. Loss of peak intensity can also be caused by intermediate exchange, where the exchange rate between bound and unbound is similar to the separation frequency between the bound and unbound peak, resulting in peaks broadening. The largest effects seen in the spectra were loss of signal intensity for α6-11, α17-C and β10-14 (Fig. 4.2).

To find whether these binding effects were specific rather than the sum of many weak, transient interactions, α6-11, α17-C and β10-14 were further sub-fragmented and tested for binding with ATS-N, ATS-Core or ATS-N-Core. Sub-fragments of α17-C still resulted in specific reduction in peak intensity, while sub-fragmentation of α6-11 and β10-14 attenuating any binding phenomena (Fig. 4.3). The binding on α17-C was localised on the α17

repeat, with construct α17-19 showing the full binding effect of α17-C, but α18-19 had no binding (Fig. 4.3C). This was confirmed by FP binding assays, titrating similar α17-C sub-fragments into 5-FAM labelled ATS-N-Core (Fig. 4.4). All constructs containing α17 bound the ATS-N-Core with K_d ~230 μM, while constructs lacking α17 did not bind.

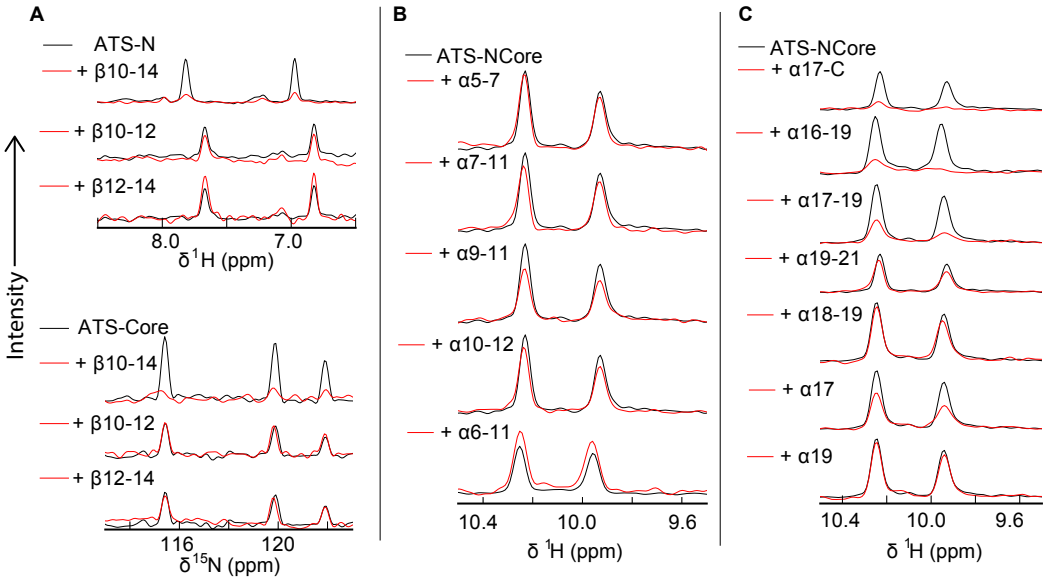


Fig. 4.3: NMR localisation of binding to ATS-Core and α17. All spectrin constructs that showed binding in ^{1}H-^{15}N-HSQC in Fig.4.2 were further sub-fragmented and tested. Spectra with ATS fragment alone (black), and in the presence of spectrin sub-fragments (red). **A**: Sub-fragments of β10-14 tested with ^{15}N ATS-N (top) and ATS-Core (bottom), slices taken at ^{15}N 114.2 ppm and at ^{1}H 9.2 ppm, respectively. **B**: Sub-fragments of α6-11 tested with ^{15}N ATS-N-Core, slices at ^{15}N 129 ppm. **C**: Sub-fragments of α17-C, tested with ^{15}N ATS-N-Core, slices at ^{15}N 129 ppm.

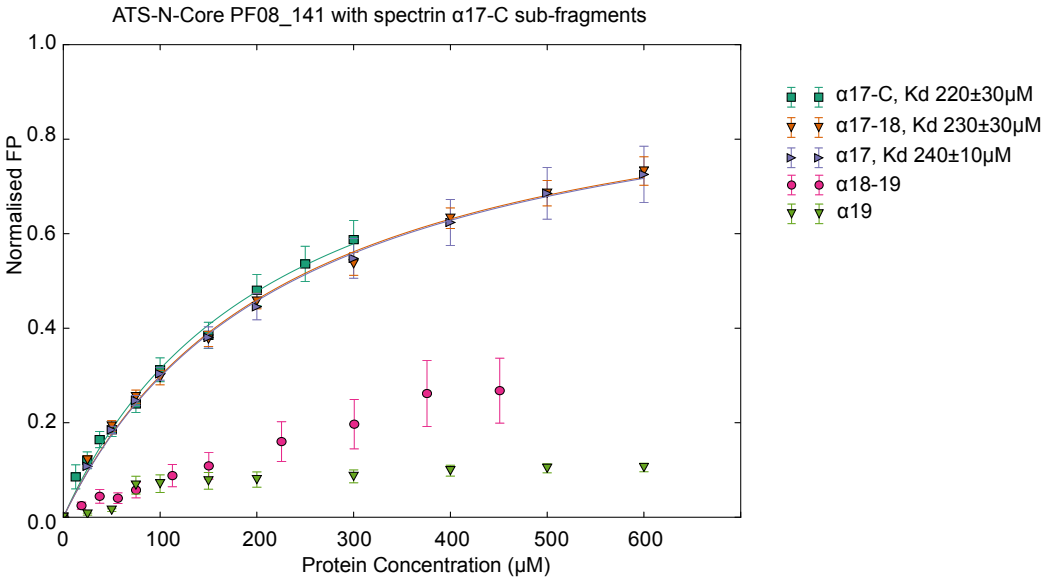


Fig. 4.4: FP binding assay localising binding to ATS-Core and α17. All constructs containing α17 bound with ~230 μM, while those that did not contain α17 did not bind.

4.2.2 N-terminal ATS truncations

As the N-terminal region of the ATS comprises of ~100 disordered residues, N-terminal truncations were made based on ConSurf [204, 205] alignment of 59 ATS sequences from the 3D7 strain of *P. falciparum*. ConSurf estimates which residues are functionally relevant based on conservation and amino acid type. This alignment suggested regions of greater sequence conservation at residues 1-5, 15-50, 60-80 and 85-100 (Fig. 4.5A). N-terminal truncations were made at 15, 35, 59 and 84 residues to remove these conserved patches (resulting in constructs ATS-N(Δ15)-Core, ATS-N(Δ35)-Core, ATS-N(Δ59)-Core and ATS-N(Δ84)-Core, respectively), labelled with ^{15}N and tested for binding with ^{1}H-^{15}N HSQC with α17-19 (Fig. 4.5E-H). While large binding effects are seen in the spectra of the ATS-N-Core (Fig. 4.5 B,C), each truncation slightly reduces these effects suggesting a loss of affinity with each truncation. Although binding is likely to be weaker, the ATS-Core still displays binding effects on the spectra (Fig. 4.5D), similar to that of ATS-N(Δ84)-Core (Fig.

4.5H).

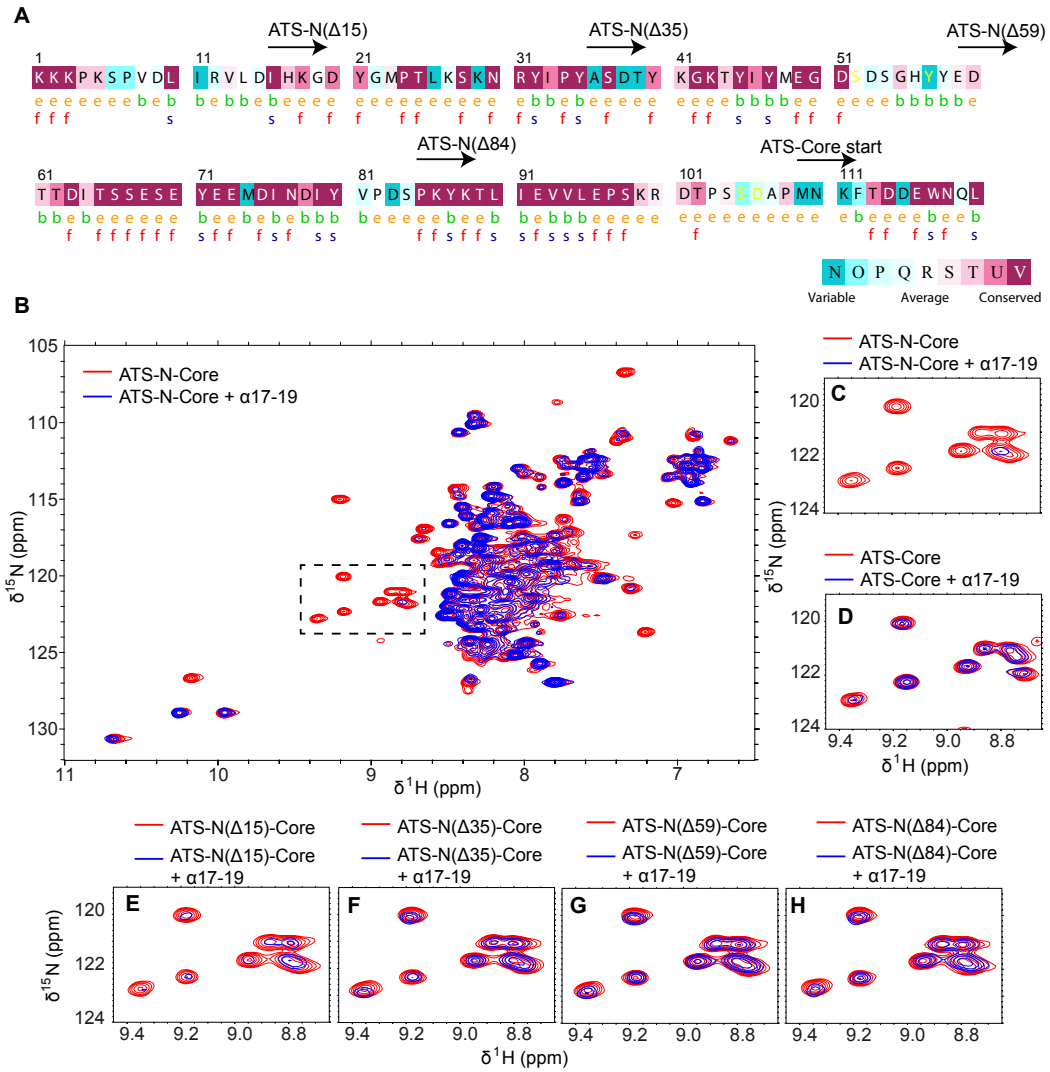


Fig. 4.5: The N-terminal region of the ATS increases binding effects in NMR. A: The N-terminal region of a ConSurf alignment of 59 ATS sequences from parasite strain 3D7 [204, 205]. Residues labelled with (e) are predicted to be exposed, (b) buried, (f) functional and (s) structural. **B**: ^{1}H-^{15}N HSQC of 100 μM ATS-N-Core alone (red) and in the presence of 200 μM α17-19 (blue). **C**: Enlargment of area in dashed line in **B**. **D-H**: The same region of the spectra as the dashed line in **B**, but using ^{15}N ATS-Core, ATS-N(Δ15)-Core, ATS-N(Δ35)-Core, ATS-N(Δ59)-Core and ATS-N(Δ84)-Core, respectively.

4.2.3 Localisation of ATS on spectrin with EM

As all assays so far have used recombinantly produced sub-domains of spectrin, it was desirable to see if binding between ATS-N-Core and spectrin was detectable with full length spectrin. Electron microscopy was performed using ATS-N-Core conjugated to 10 nm gold nanoparticles via a N-terminal cysteine and spectrin purified from human erythrocytes, and stained with 2% uranyl acetate. Spectrin is visualised under negative stain electron microscopy as extended filaments, with a length of ~100 nm for dimers and ~200 nm for tetramers [206]. The length of visible particles was ~100 nm, suggesting dimers are the predominant species. In some cases spectrin filaments were decorated with gold nano-particles (Fig. 4.6 A, B). These predominantly localised at the end of filaments, in agreement with binding data, as α17 would be located at the end of the spectrin tetramer. Two negative controls were also examined; the same concentration of spectrin with nanoparticles without protein conjugated and ATS-N-Core nanoparticles alone (Fig. 4.6 C, D) and neither showed such features. While these results showed promise, the sample preparation was not able to ensure unconjugated particles and protein aggregates were removed. This, together with the short-shelf life of the reactive particles, resulted in insufficient data to perform a rigorous analysis.

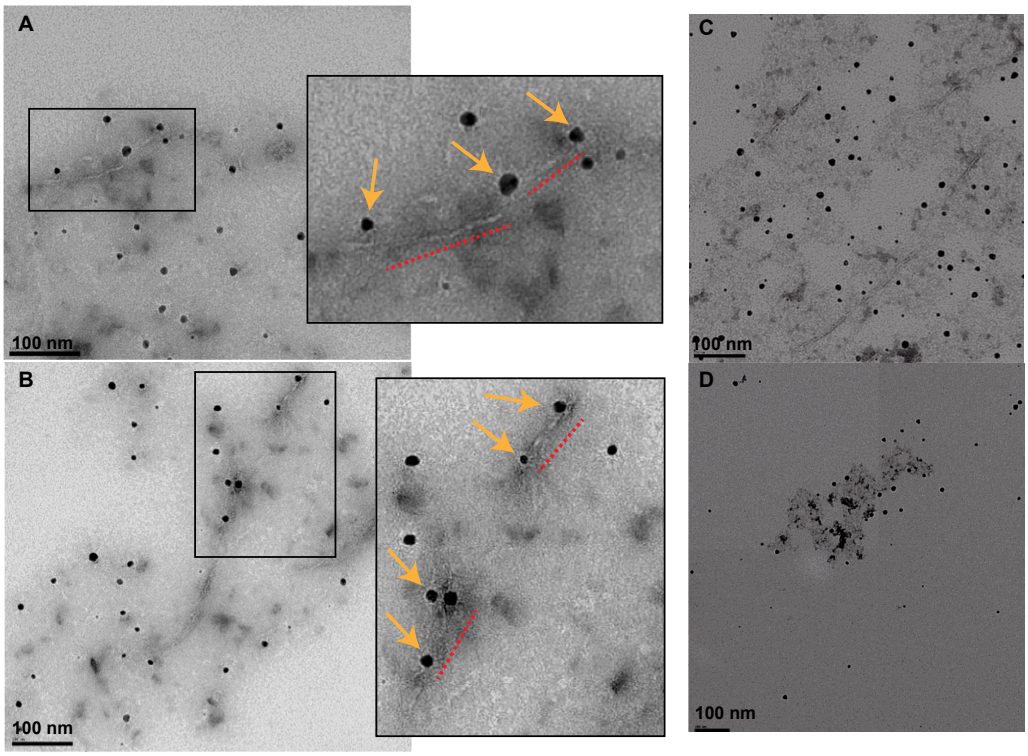


Fig. 4.6: EM suggests binding of ATS-N-Core to spectrin. **A, B**: Spectrin was mixed with ATS-N-Core conjugated to 10 nm gold nano-particles and stained with 2% uranyl acetate. ATS-N-Core nano-particles (orange arrows) can be observed localising to spectrin (dotted red line). **C**: Spectrin mixed with unconjugated gold nano-particles. **D**: ATS-N-Core gold nano-particles alone.

4.3 Structural Characterisation

As binding had been localised between α17 spectrin and ATS constructs containing the core and N-terminal fragments, attempts were made to characterise this complex with X-ray crystallography.

4.3.1 CD

The circular dichroism (CD) spectra were collected for sub-fragments of α17-C (α17-C, α17-19, α19-21, α17 and α16-19) to evaluate the structural state. Characteristic helical

peaks at 208 and 222 nm [207] were observed in all cases (Fig. 4.7A). Deconvulusion of the CD spectra with the Diochroweb server [208] suggest that all subfragments had 86% or greater helicity, consistent with the three-helix repeat of spectrin. CD was also used to assess thermostability of α17 and α17-C. In both cases cooperative unfolding at high temperatures was observed. The longer construct, α17-C, resulted in a melting temperature of ~45 °C, while α17 was more stable at ~60 °C. As constructs were folded, constructs that bound the ATS (α16-17, α17-19 and α17) were carried forward for further structural efforts.

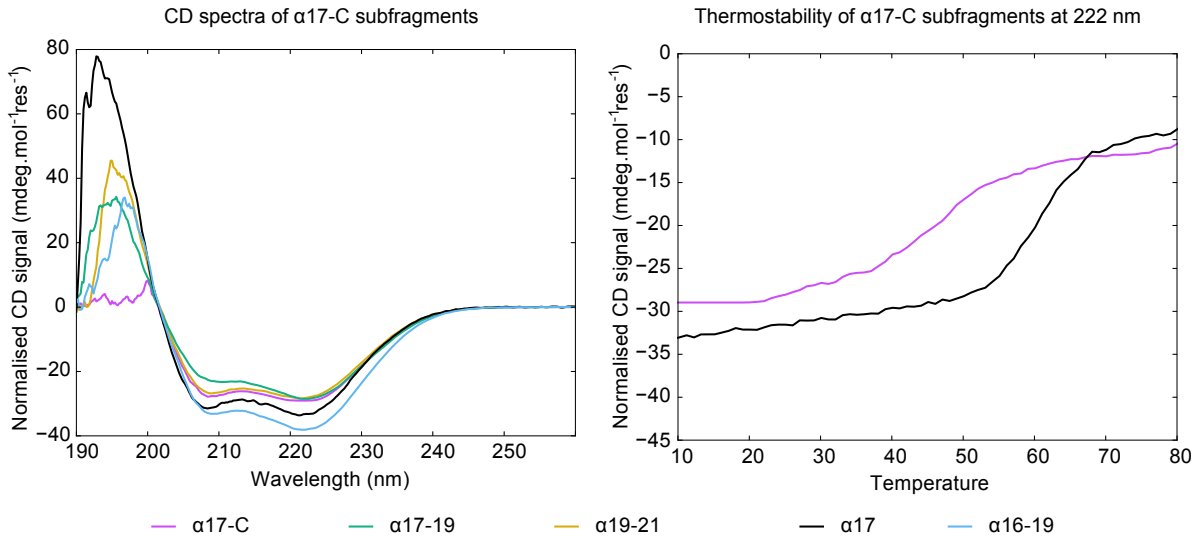


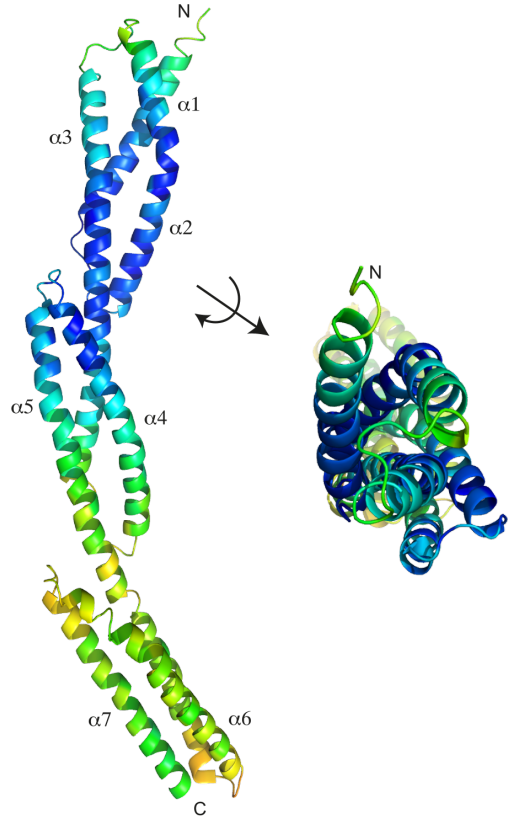
Fig. 4.7: CD of α17-C sub-fragments. A: CD spectra of α17-C, α17-19, α19-21, α17 and α16-19 at 10 μM at 10°C. **B**: Thermo-stability of α17-C sub-fragments, determined at 222 nm.

4.3.2 Crystal Structure of α17-19

Initial crystallisation efforts were focused on α17-19. Screening was set up with molar equivalent of ATS-N(Δ15)-Core or ATS-N(Δ84)-Core to α17-19, with total protein concentration of 6 mg/mL and 7.5 mg/ml respectively. Crystal hits were extensively optimised for α17-19 with ATS-N(Δ84)-Core, resulting in a data set to 3.2 Å resolution. The structure was determined using molecular replacement using a template from the Balbes molecular replacement pipeline, based on PDB 1CUN [159, 209] (Fig. 4.8A). As molecular replace-

ment is challenging at this resolution, the initial solution had less than ideal statistics (Fig. 4.8B). Examination of the B-factors in this structure show that α repeat 19 had poor B-factors, possibly due to poor stability observed for longer constructs in CD or due to protein flexibility. No density was observable for the ATS. Given the poor data quality, further optimisation aimed to increase the stability of the spectrin construct used.

A



B **α17-19 Crystallographic data and refinement statistics**

Data collection statistics	
Space group	P 2_1 2_1 2_1
Unit cell (Å)	a = 39.9, b = 42.2, c = 322.7
Beamline	DLS/I04-1
Wavelength (Å)	0.91741
Resolution range(Å)	53.74-3.21
High resolution shell (Å)	3.33 - 3.21
R_{merge}	0.1752 (0.7575)
Completeness (%)	99.49 (99.79)
Multiplicity	7.0 (7.1)
Mean I/σ(I)	10.13 (2.72)
CC-half	0.997 (0.834)
Refinement statistics	
R_{work} (reflections)	26.8% (9,575)
R_{free} (reflections)	33.9% (445)
Number of atoms	
Protein (excluding H)	2639
Water	10
Ligands	0
Average B factors (Å²)	
Protein (excluding H)	61.5
Water	42.1
Ligands	0
RMSD from ideal values	
Bonds (Å)	0.018
Angles (°)	2.44
Molprobity statistics[a]	
Ramachandran favored (%)	83
Ramachandran disallowed (%)	1.9
Favored rotamers (%)	97.5
Poor rotamers (%)	2.5
Clashscore	62.4
MolProbity score	3.19

Fig. 4.8: Crystal structure of α17-19. **A**: Structure of α17-19 coloured with B-factor heatmap, from 10 (blue) to 122 Å^2 (red). **B**: Crystallographic statistics, values in parentheses correspond to the highest resolution shell, [a] denotes values from MolProbity [155].

4.3.3 Crystal structure of α16-17

Screening of α16-17 was set up with molar equivalent of ATS-N(Δ84)-Core to α16-17, with total protein concentration of 4.3 mg/ml. α16-17 crystallographic data was collected to 1.6 Å resolution and solved with molecular replacement using a template from Balbes

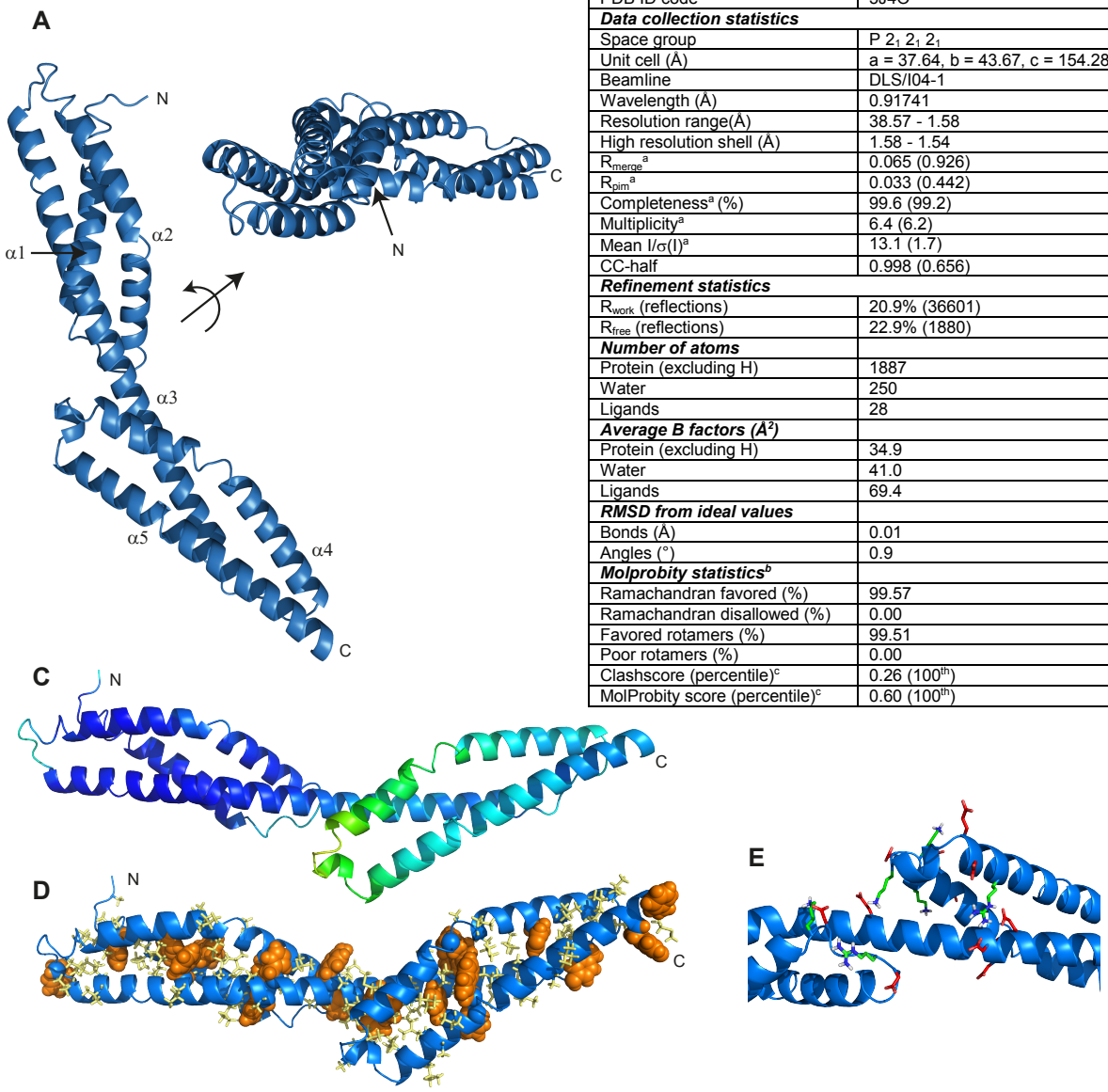
based on PDB 1U5P [159, 210] that contained only helix 1, 2 and 3 (one and one third of a repeat). Helix 5 and 6 were built in Coot (Fig. 4.9A). The structure was refined to an R-free of 22.9%, which is slightly high given the resolution of the data (Fig. 4.9B). Mapping the B-factors onto the structure indicated that repeat 17 has larger B-factors than repeat 16, hence suggesting there could be movement of α17 relative to α16 (Fig. 4.9C).

Structurally, the hydrophobic core of both repeats is stabilised by clusters of aromatic residues, which includes a hydrophobic cluster between the 16th and 17th repeat (Fig. 4.9D). The ends of the repeats had clusters of charged residues, which were engaged in three salt-bridges across helices, although many of these side chains had poorly defined density, thus other contacts could occur in solution (Fig. 4.9E). In both α16 and 17, the second helix is kinked.This is induced by having three hydrophobic residues rather than two on the hydrophobic face of helix two in α16, while in α17, this is induced by a proline residue.

Again no density was observed for the ATS in crystals. This spectrin construct was also set up in crystallisation trials in a large range of conditions and at 4 °C and at room temperature with varying concentration of ATS-N(Δ84)-Core and ATS-N(Δ15)-Core, but no crystals contained the ATS-spectrin complex. The spectrin packing in the resolved crystal structure was too dense to allow space for the ATS.

4.3.4 Crystal structure of α17

α17 readily crystallised with molar equivalent of ATS-N(Δ84)-Core with a total protein concentration of 6 mg/mL and crystallographic data was collected to 2.4 Å resolution. The structure was solved with molecular replacement using repeat 17 from my structure of α16-17 as a template. After phases were attained, the electron density clearly showed the structure was a helix-swap dimer (Fig. 4.10A). This structure is most likely an artefact of the crystallisation condition. Again, no density was observed for the ATS.

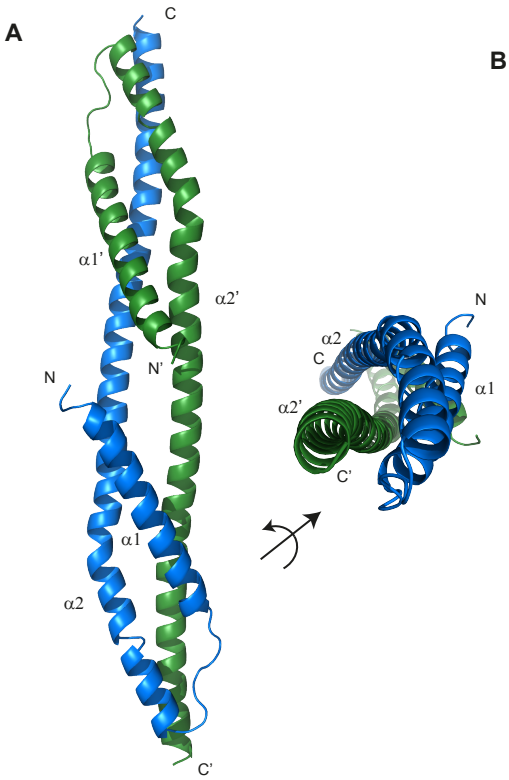


B Crystallographic data and refinement statistics

PDB ID code	5J4O
Data collection statistics	
Space group	P 2_1 2_1 2_1
Unit cell (Å)	a = 37.64, b = 43.67, c = 154.28
Beamline	DLS/I04-1
Wavelength (Å)	0.91741
Resolution range(Å)	38.57 - 1.58
High resolution shell (Å)	1.58 - 1.54
R_{merge}[a]	0.065 (0.926)
R_{pim}[a]	0.033 (0.442)
Completeness[a] (%)	99.6 (99.2)
Multiplicity[a]	6.4 (6.2)
Mean I/σ(I)[a]	13.1 (1.7)
CC-half	0.998 (0.656)
Refinement statistics	
R_{work} (reflections)	20.9% (36601)
R_{free} (reflections)	22.9% (1880)
Number of atoms	
Protein (excluding H)	1887
Water	250
Ligands	28
Average B factors (Å^2)	
Protein (excluding H)	34.9
Water	41.0
Ligands	69.4
RMSD from ideal values	
Bonds (Å)	0.01
Angles (°)	0.9
Molprobity statistics[b]	
Ramachandran favored (%)	99.57
Ramachandran disallowed (%)	0.00
Favored rotamers (%)	99.51
Poor rotamers (%)	0.00
Clashscore (percentile)[c]	0.26 (100th)
MolProbity score (percentile)[c]	0.60 (100th)

Fig. 4.9: Crystal structure of α16-17. **A**: The structure of α16-17. **B**: Crystallographic statistics, [a] denotes values by Aimless [211]; values in parentheses correspond to the highest resolution shell, [b] denotes values from MolProbity [155], [c] percentile values indicate the relative structure quality. 100th percentile suggests the structure is amongst the best of comparable resolution. **C**: B-factors mapped onto structure as a heat map, from 9 (blue) to 118 Å^2 (red). **D**: Hydrophobic core, with aliphatic side chains shown in yellow, and aromatic residues shown in orange. **E**: Charged side chains at interface of repeats, positively charged residues shown in green, negatively charged sidechains shown in red.

Including this structure, three structures were solved of α17. This structure of α17 was overlaid with the other crystallographic structures using helix 1 and the first half of helix 2 of one molecule of the dimer, and the second half of helix 2 from the other molecule of the dimer. Overlaying these structures resulted in excellent agreement for α17, with an RMSD of less than 1 Å (Fig. 4.11). This suggests that swapping of the helices does not result in distortion of the structure.



B **α17 Crystallographic data and refinement statistics**

Data collection statistics	
Space group	C 2 2 2_1
Unit cell (Å)	a = 45.34, b = 72.02, c = 131.74
Beamline	ESRF/ID30
Wavelength (Å)	0.966
Resolution range(Å)	38.37 - 2.2
High resolution shell (Å)	2.26 - 2.2
R_{merge}[a]	0.081 (1.502)
R_{pim}[a]	0.041 (0.944)
Completeness[a] (%)	98.3 (91.7)
Multiplicity[a]	4.6 (3.1)
Mean I/σ(I)[a]	11.7 (0.8)
CC-half	0.998 (0.460)
Refinement statistics	
R_{work} (reflections)	22.5% (11112)
R_{free} (reflections)	25.5% (1054)
Number of atoms	
Protein (excluding H)	1114
Water	108
Ligands	14
Average B factors (Å²)	
Protein (excluding H)	58.2
Water	50.5
Ligands	86.4
RMSD from ideal values	
Bonds (Å)	0.009
Angles (°)	1.16
Molprobity statistics[b]	
Ramachandran favored (%)	98
Ramachandran disallowed (%)	0.00
Favored rotamers (%)	98.1
Poor rotamers (%)	1.9
Clashscore	2.23
MolProbity score	1.20

Fig. 4.10: Crystal structure of α17. **A**: Structure of α17, with monomers coloured separately. **B**: Crystallographic statistics, [a] denotes values by Aimless [211]; values in parentheses correspond to the highest resolution shell and [b] denotes values from MolProbity [155].

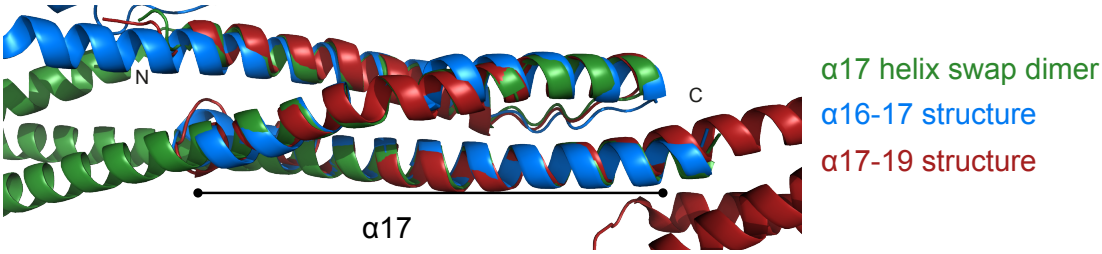


Fig. 4.11: Overlay of crystallographic structures of α17. All crystal structures of α17 agree well with backbone structure.

4.3.5 SAXS

Small angle X-ray scattering (SAXS) is a solution based method that gives low resolution structural information. Experimental data can be compared to theoretical scattering curves from structural models to better understand solution behaviour.

Data was collected for a dilution series of α17, α16-17, ATS-N-Core and ATS-N(Δ84)-Core. In all cases except ATS-N(Δ84)-Core, as the protein concentration was increased the radius of gyration, r_g, also increased, suggesting the formation of higher molecular mass species at high concentrations. SAXS data for α17 and α16-17 at 0.8 and 0.7 mg/mL, respectively, was fitted with theoretical scattering curves from the crystal structures using Crysol [212] (Fig. 4.12A, B). The fit was acceptable for α16-17, with a χ^2 of 0.9. As the crystal structure of α17 was resolved as a helix swap dimer, both the crystallographic dimer and the single α17 repeat from α16-17 were fit, resulting in a χ^2 of 3.0 and 3.7 respectively. Inspection of the data suggests neither model fits well, and there is most likely a mixture of these two, and possibly more species.

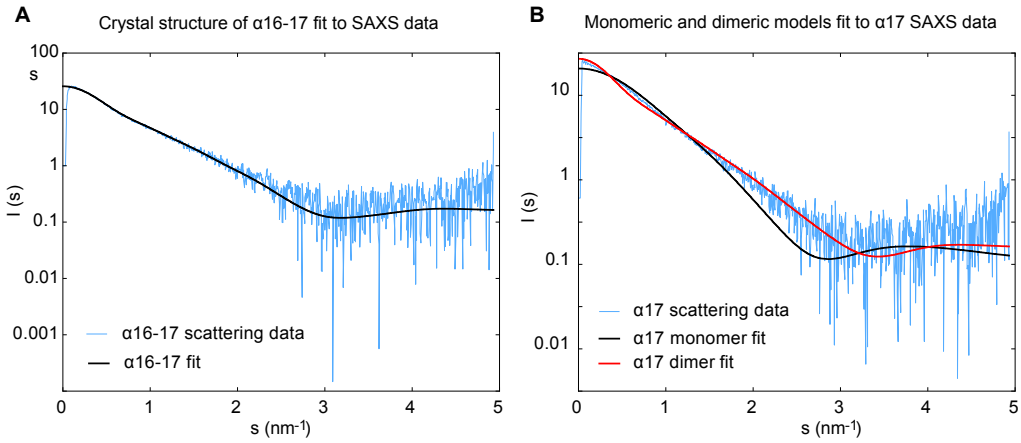


Fig. 4.12: SAXS data of α16-17 and α17. A: SAXS scattering data (blue) of α16-17 at 0.7 mg/mL, with fit curve from crystal structure shown in black, with a χ^2 of 0.9. **B**: SAXS scattering data (blue) of α17 at 0.8 mg/ml, with fit curve from crystal structure of monomer (black) and dimer (red), with χ^2 of 3.7 and 3, respectively.

4.4 Sequence Variants

As crystallography attempts with *Pf*EMP1 variant PF08_0141 were unsuccessful, other sequence variants of the ATS were examined to test whether the ATS-α17 interaction is conserved among *Pf*EMP1 variants. I produced a further five fluorescently labelled full-length ATS domains that show substantial sequence variability [65]. All ATS variants bound α17 with affinities varying between 24 and 200 μM (Fig. 4.13), with PF08_0141 being one of the lowest affinity variants. Further binding investigations were done using the highest affinity variant, PFF0845c.

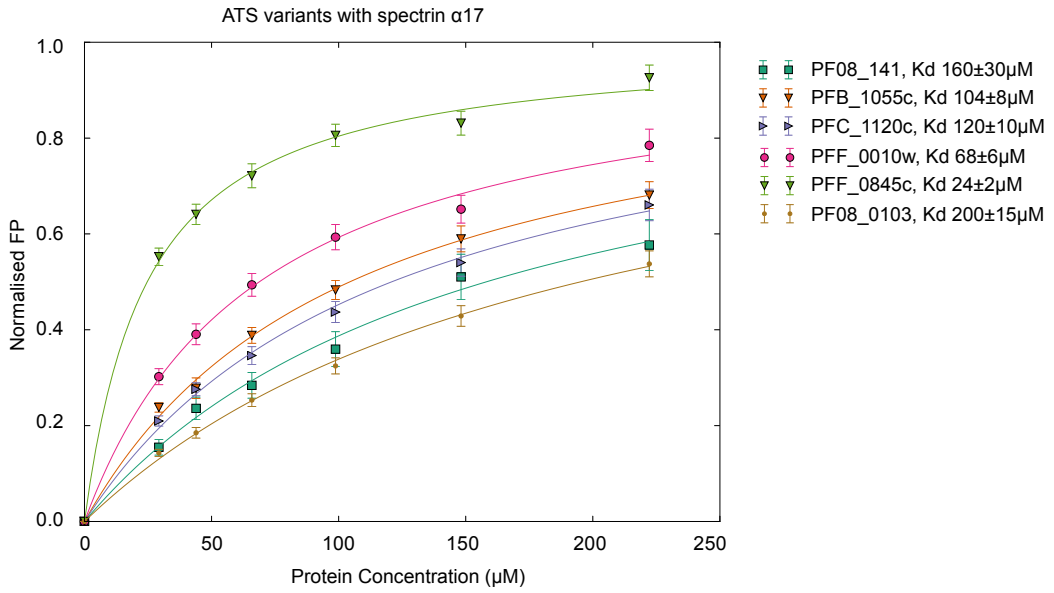


Fig. 4.13: *Pf*EMP1 sequence variants bind to α17. Five additional *Pf*EMP1 ATS variants of diverse sequence were tested for binding with FP binding assays, and all bound with similar or greater affinity than PF08_0141.

4.5 Interaction Model of *Pf*EMP1 ATS-Core and α17

4.5.1 NMR Shift Perturbation Mapping

As ATS variant PFF0845c binds α17 with substantially higher affinity than other members (Fig. 4.13), I optimized the conditions of complex formation to identify specific amino acids at the α17-ATS binding interface. I performed NMR ^{1}H-^{15}N HSQC titrations with ^{15}N ATS-Core PFF0845c, titrating in α17, and ^{15}N α17 titrating in ATS-Core PFF0845c (Fig. 4.14A, B). In both cases significant changes in peak chemical shift are visible, making this interaction amenable to investigation with NMR.

Backbone shifts were assigned using data from 3D experiments CBCA(CO)NH, CBCANH, HNCO, HN(CA)CO and HBHA(CO)NH. ATS-Core PFF0845c data was collected at 30 °C, and HNCO experiments were performed titrating in unlabelled α17, enabling map-

ping of backbone ^{13}C, ^{15}N and ^{1}H shift perturbation (Fig. 4.15A). Experimental conditions had to be optimised for α17, as at high concentrations it tended to form larger complexes, as shown by SAXS, resulting in reduced NMR signal. Increasing the NaCl concentration to 300 mM and the temperature to 40 °C resulted in optimal signal, allowing assignments. Assignments for α17 at 40 °C were transferred to 30 °C and titrations with ATS-Core PFF0845c were performed using the HNCO pulse sequence. This enabled shift perturbation of α17 in the presence of ATS-Core PFF0845c to be determined (Fig. 4.15B). The elevated ionic concentration and temperature resulted in reduced spectral effects compared to the original assay conditions hence to ensure the binding effects were the same at higher temperature, many of the assignments from 30 °C were transferred to a HSQC of ^{15}N α17 collected 25 °C degrees alone and in the presence of ATS-Core PFF0845c. While, in this case, full shift mapping was not possible due to spectral overlap, the perturbations that were mapped corresponded to the same regions of α17 and are generally twice as large (Fig. 4.15B).

As the ATS architecture has been shown to be similar across variants [65] and PFF0845c shares ~70% identity with PF08_0141, the core of PFF0845c was homology modelled from the NMR structure of the core of PF08_0141. Mapping NMR shift perturbations on to this ATS-Core model suggests that the binding interface is located towards the C-terminus on helix 3, helix 4 and the loop between them (Fig. 4.15C). As I solved the crystal structure of several constructs containing the α17 repeat, these shift perturbations were visualised in the context of the crystallographic structure. Doing this analysis using the α17 repeat from the α16-17 crystal structure suggested the binding interface lies predominately on the N-terminus of helix one and helix three (Fig. 4.15D) of α17.

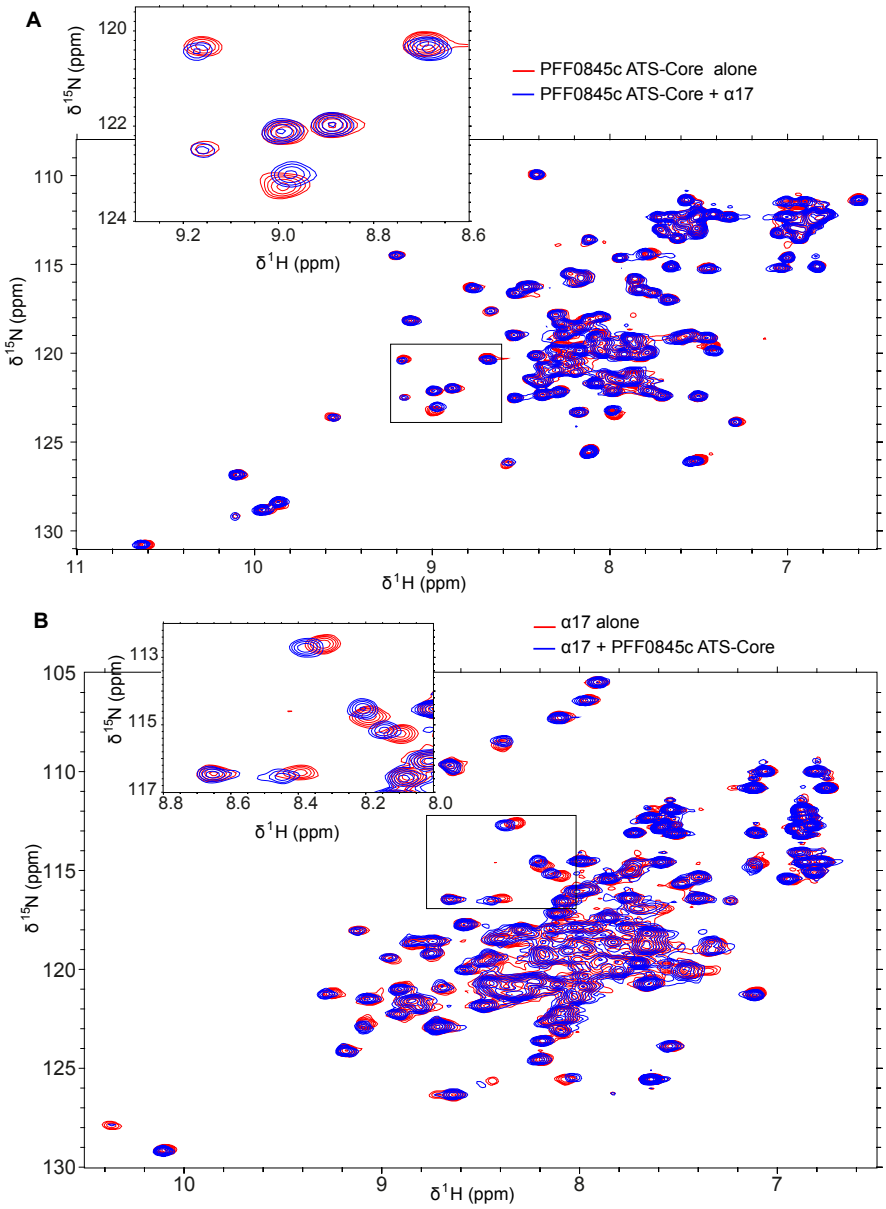


Fig. 4.14: HSQC shift perturbation of α17 and ATS-Core. A: ^{1}H-^{15}N HSQC of 50 μM PFF0845c ATS-Core alone (red) and in the presence of 100 μM α17 (blue). **B**:^{1}H-^{15}N HSQC of 50 μM α17 alone (red) and in the presence of 150 μM PFF0845c ATS-Core (blue). Both **A** and **B** were collected at 25 °C.

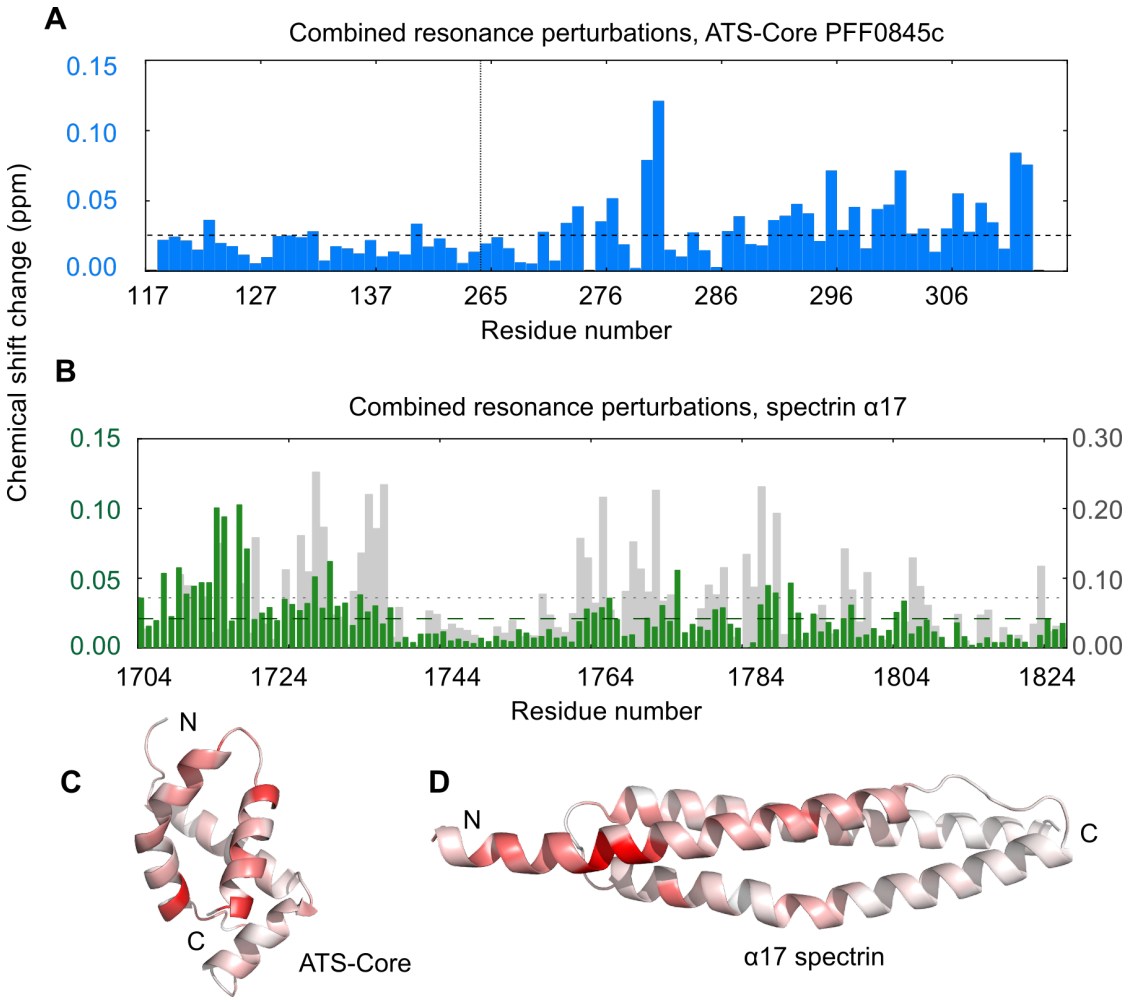


Fig. 4.15: Shift perturbations mapping. A: The combined ^{15}N, ^{13}C and ^{1}H shift perturbations per residue of 50 μM PFF0845c ATS-Core in the presence of 250 μM α17. **B**: The combined ^{15}N, ^{13}C and ^{1}H shift perturbation per residue of 100 μM α17 in the presence of 300 μM PFF0845c ATS-Core at 303 K (green). The combined ^{15}N and ^{1}H shift perturbation per residue of 50 μM α17 in the presence of 150 μM PFF0845c ATS-Core at 298 K (grey). In both **A** and **B** the dashed lines indicate one standard deviation of noise. **C, D**: Shift perturbations of **A** and **B**, mapped onto structural models, coloured from white (0 ppm perturbation) to red (0.1 ppm perturbation).

4.5.2 Complex Docking with HADDOCK

The model of PFF0845c ATS-Core and the crystal structure of α17 were docked together using the NMR chemical shift perturbation as restraints in HADDOCK (Appendix B) [171, 172], resulting in the prediction of two possible complex conformations that are related by approximately 170° rotation of ATS-Core (Fig. 4.16A,D). Cluster 1 had more favourable energy scores for ionic and hydrophobic interactions and a larger buried surface area, while Cluster 2 satisfied more restraints, hence had a more favourable HADDOCK score. Analysis of these interfaces suggest key residues are used in both interfaces; a hydrophobic cluster is formed between W294, M290 and W283 of the ATS-Core and Y1713 and F1716 or L1709 of α17 (Fig. 4.16B,E). There are also extensive hydrogen bonding and ionic interactions, and although these vary between each docked complex, ATS-Core residues R286 and E308 make contacts in both cases. The number of contacts in cluster 1 is greater than cluster 2, so is likely to be more energetically favourable, but as both were deemed possible by HADDOCK, they were both carried forward for further analysis.

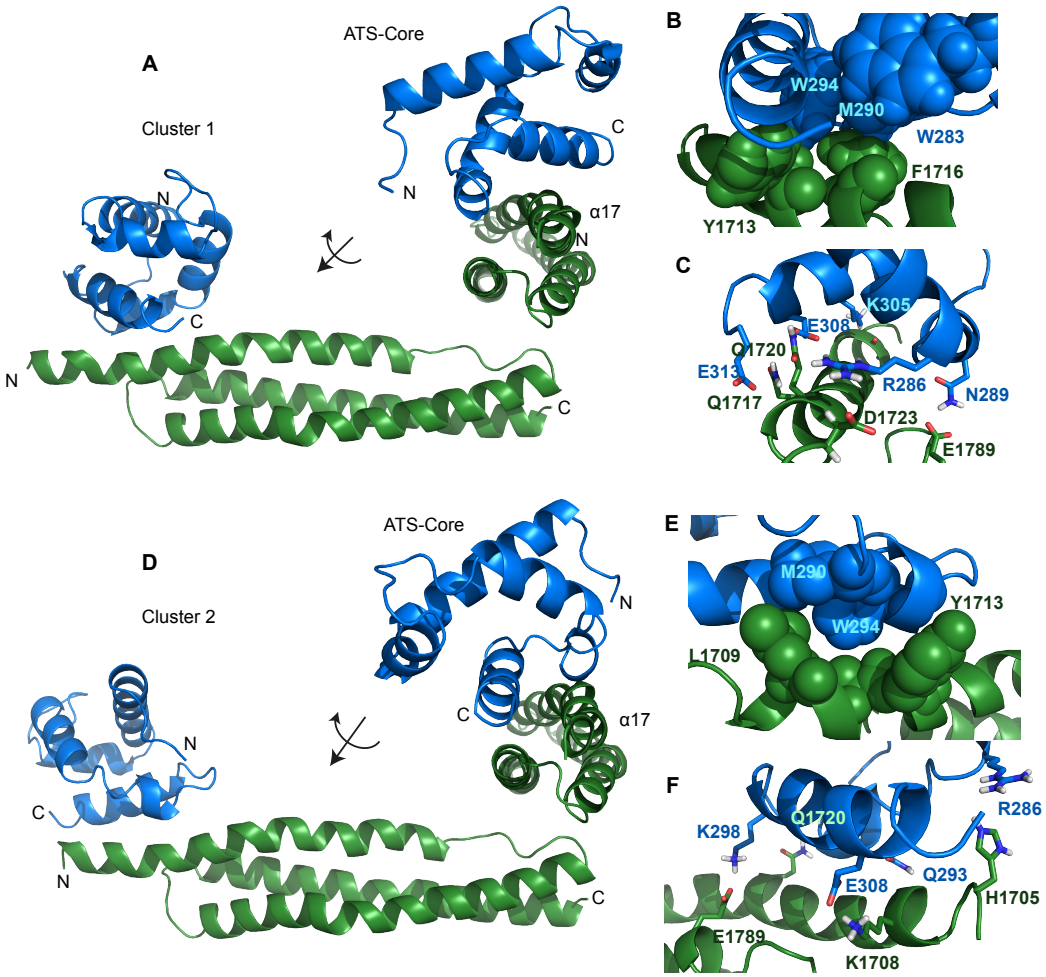


Fig. 4.16: HADDOCK complex structure prediction. **A, D**: The two interfaces predicted with HADDOCK. **B, E**: Hydrophobic interactions. **C, F**: Hydrogen bonding and ionic interactions.

4.5.3 MD

To distinguish between the two possible HADDOCK complexes, I refined both conformations with MD simulations. I observed strengthening of the first complex interface as judged by the formation of additional hydrogen bonds and by an increase in buried surface area during 100 ns of simulation time (Fig. 4.17A, B). In contrast, the second complex interface was destabilized leading to its disruption in run 2. This result suggests that cluster 1 gives a better model of the binding interface and also helps support MD as a method to distinguish between complexes with less than favourable interactions.

Cluster 1 was further analysed to judge the structural stability of the ATS-Core and α17 in the complex. Equivalent MD simulations were set up of α17 and PFF0845c ATS-Core alone. In both cases, the Cα RMSD and the Cα RMSF is comparable for the docked complex and the ATS-Core or α17 alone (Fig. 4.18).

Overlaying cluster 1 structures after 100 ns MD simulation time suggest there is some dynamics in the binding of the ATS-Core to α spectrin (Fig. 4.19A, B), but certain residues of both the ATS-Core and α17 are likely responsible for maintaining the interaction. Analysis of the interfaces after 100 ns found ATS-Core residues H287, M290 and W294 made hydrophobic contacts with α17 residues L1709, F1716, and sometimes with Y1713 or F1719. While ionic and hydrogen bond contacts most often involved ATS residues R286, H287, N289 and K305, and α17 residues Q1717, Q1720, D1722 and D1723, exact interacting pairs altered throughout the simulation. Most of these residues were making contacts in the original HADDOCK structure, hence MD allowed these contacts to relax, while remaining bound.

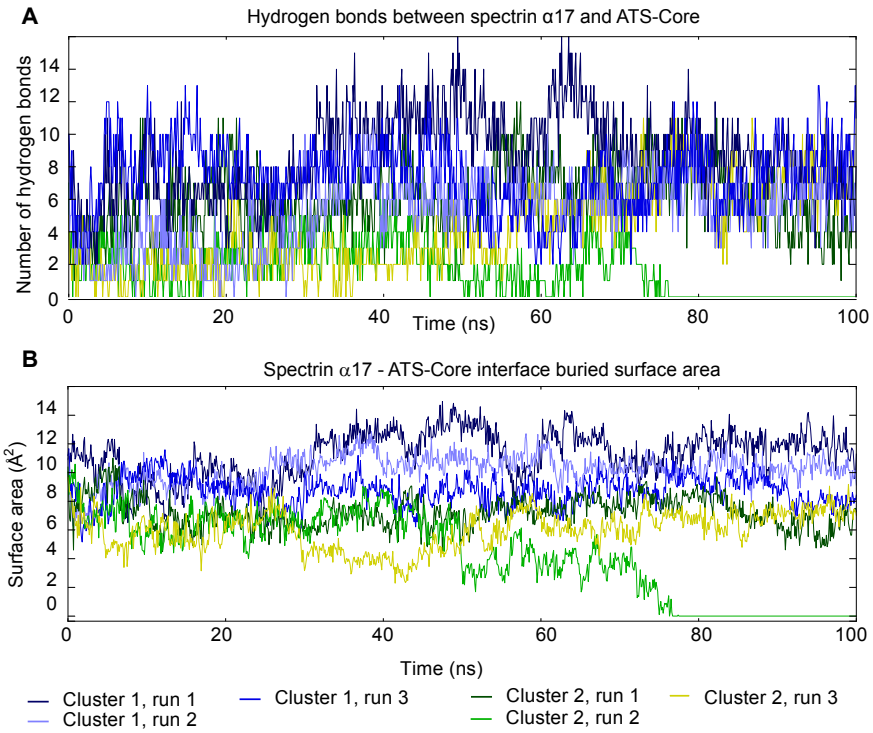


Fig. 4.17: MD on HADDOCK complexes. **A**: The number of hydrogen bonds between α17 and the ATS-Core. **B**: The buried surface area between α17 and the ATS-Core. With both metrics cluster 1 remains fairly consistent during the simulation, but decreases in cluster 2, indicating cluster 2 is less stable. The interaction is completely lost in run 2 of cluster 2, as indicated by loss of hydrogen bonds and buried surface area.

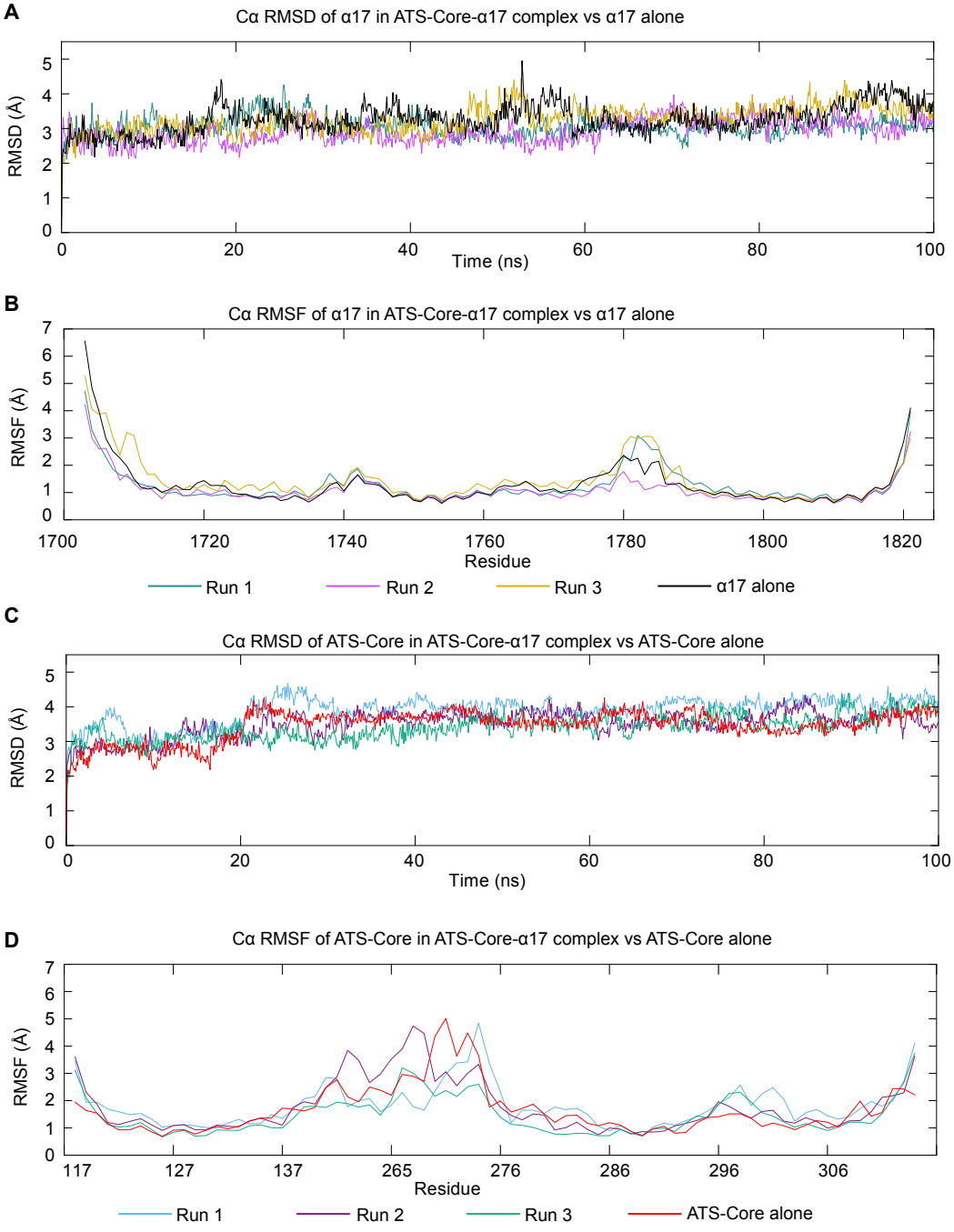


Fig. 4.18: Cluster 1 MD compared to non-docked structures. **A**: Cα RMSD, **B**: Cα RMSF of α17, docked to ATS-Core or alone (black line). **C**: Cα RMSD, **D**: Cα RMSF of ATS-Core, docked to α17, or alone (red line).

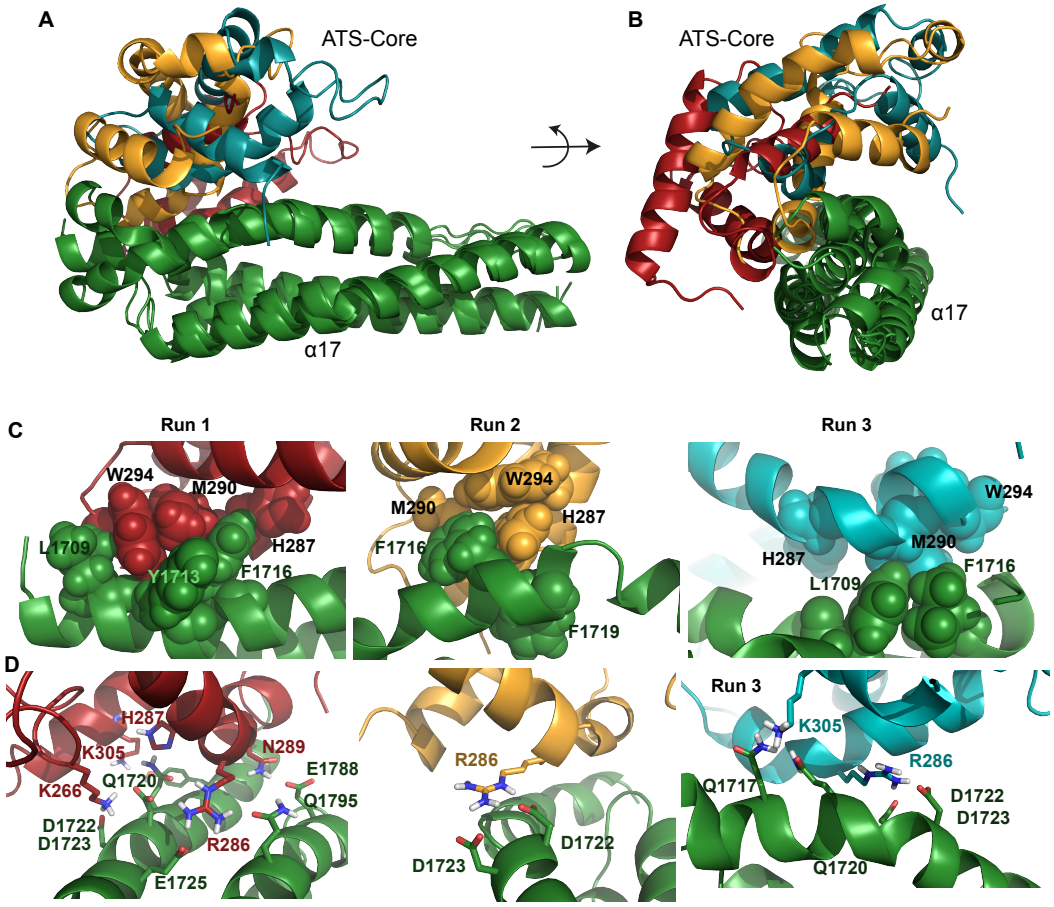


Fig. 4.19: MD structural interfaces. **A, B**: Overlay of ATS-Core-α17 complex after 100 ns MD simulation, showing variable shifting of interface after MD. **C**: Hydrophobic contacts after 100 ns simulation of each run. **D**: Hydrogen bonding and ionic interactions after 100 ns simulation.

4.5.4 Model Validation

Having found specific amino acids that are involved in the binding interface, I set to validate the model of the complex by performing substitution mutagenesis. The observations from docking and MD, as well the sequence alignment of the ATS-Core of the variants was used to select residues for substitution (Fig. 4.20A). To reduce charge interactions and hydrogen bonds that appear in docking and MD, substitutions were produced of α17 residues Q1717A, Q1720A, D1722K and D1723K and of ATS-Core residues K266E,

R286E, N289D, K305A and E308Q, where N289D and E308Q substitute residues from lower affinity variants, charged residues were mutated to the opposing charge and hydrogen bonding side chains to alanine. To disrupt hydrophobic interactions observed in docking and MD, substitutions were produced of α17 residues L1709A, Y1713A, F1716A and F1718A, and ATS-Core residues H287A, M290A and W294A. Substitutions of the ATS-Core were tested using ^{1}H-^{15}N HSQC titration, using ^{15}N α17, while α17 mutants were tested using FP binding assays using 5-FAM PFF0845c ATS-FL.

Single substitutions of specific amino acids of ATS-Core and α17 reduced affinity by up to 10-fold. Notably, residue H287 of ATS-Core, that forms part of a hydrophobic pocket into which F1716 of α17 inserts both result in large affinity effects (Fig. 4.20B,C). ATS-Core residue R286 forms ionic and hydrogen bond interactions with D1722 on α17, and substitution of either residues result in large effects on binding affinity. Mutating N289D also resulted in a significant loss of affinity, hence this residue may be one that results in PFF0845c having stronger affinity than PF08_0141. As expected for a physiologically relevant interaction, the ATS-Core residues important for complex formation, such as R286 and H287, are conserved or conservatively substituted across all *Pf*EMP1 ATS-Core variants. All substitutions that signifcantly effected affinity (ATS-Core R286E, H287A, N289D and K305A and α17 L1709A, Y1713A, F1716A, F1718A and D1722K) were checked for stability using CD thermo-melts. While ATS-Core H287A and K305A, and α17 F1718A had melting temperatures ~10 °C lower than wild-type sequences, these substitutes all had the same amount of helical CD signal at assay temperatures (25 °C), hence all are likely to be correctly folded (Appendix D).

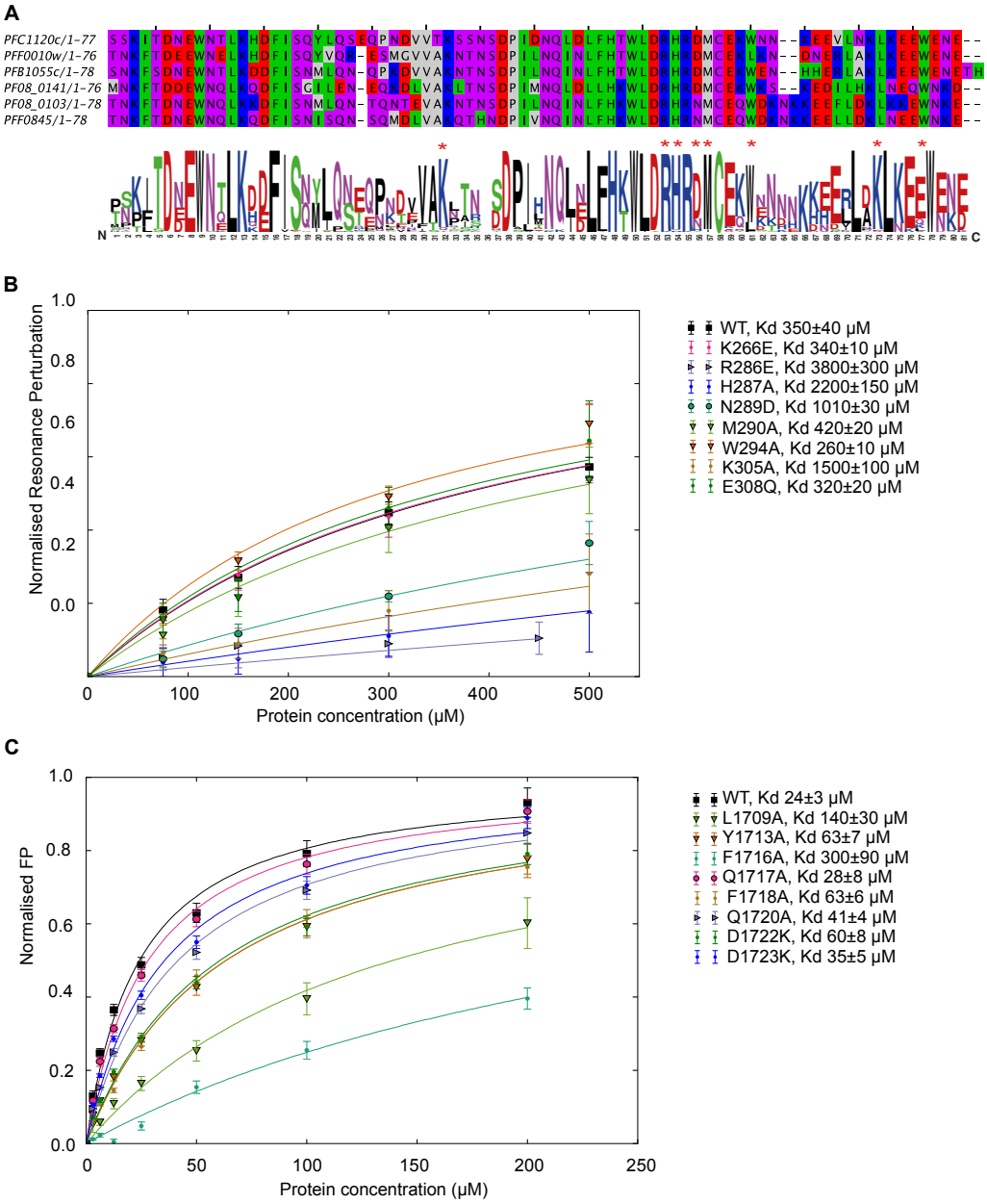


Fig. 4.20: Model validation. **A**: Sequence alignment of ATS-core sequences of variant tested for binding, with consensus of all 59 sequences of the ATS from 3D7 shown underneath, (*), indicate residues mutated. **B**: ATS-Core substitution. Binding affinity was determined by tracking the same shift perturbations of ^{15}N α17 on addition of ATS-Core mutants. **C**: FP binding assay using 5-FAM labelled PFF0845c ATS-FL and α17 mutations. FP data plotted is the mean and standard error derived from four technical repeats and 3 plate repeats.

4.6 Discussion

4.6.1 Spectrin Structure and Implications

The results of this chapter included three structures of spectrin, repeats α17-19, α16-17 and repeat α17. All structures agree well in the overlapping region of α17, despite the α17 structure being resolved as a helix swap dimer. What is clear from the B-factors of α16-17 and α17-19 is that addition of extra repeats results in more structural flexibility. The helix swap dimer form of α17 may have reduced the dynamic movements of the repeat, while allowing it to maximise crystallographic contacts, hence were preferentially crystallised. Examination of this structure via SAXS was not able to determine whether the dimeric or monomeric form is solely present in solution, and could well be a combination of both forms. Higher order molecular weight complexes were observed for both α17 and α16-17 in SAXS, suggesting higher order structures extend beyond just dimer formation for α17. Initial NMR spectra also suggested the formation of higher order structures of α17, as spectral quality was poor at high concentration. Spectral quality was improved by the addition of NaCl and increased temperature, suggesting higher order complexes were ionically driven. It is likely that this effect would have resulted in a decrease in measurable affinity, as higher order structures could hide the binding interface, and also result in reduced NMR signal and produce anomalous FP signal. The FP assay of ATS variants was fit using data less than 200 μM of α spectrin as higher concentrations resulted in a slight upward drift of signal, most likely due to binding to larger α-spectrin complexes. Hence, optimisation and careful analysis ensured that any possible oligomerisation of α17 did not influence conclusions.

4.6.2 Disordered Regions Role in Binding

The affinity of ATS-FL variant PF08_0141 with α17-C has a K_d of ~60 μM, while the ATS-N-Core bound α17-C with K_d of ~200 μM, suggesting the disordered middle loop

and C-terminal region of the ATS had some role in spectrin binding. Furthermore, truncations of the N-terminal disordered region resulted in reduced binding effects. To a certain degree this is counter intuitive. While some flanking regions of these disordered regions could contribute to binding, it is unlikely that all of these disordered regions make specific contacts, and it is surprising disordered regions around the core do not hide the binding interface. Furthermore some regions of the ATS are unlikely to make spectrin contacts *in vivo* as they are tethered at the membrane, such as the first 15 residues of the ATS-N, or because they bind other factors, such as the ATS-C which Chapter 3 shows binds to the PHIST domain of PFE1605w. In addition, all of these disordered regions have net negative charge, like spectrin, so could possibly cause repulsion. One possible explaination is that these regions help orientate the complex, where studies in other systems have shown orientation of binding partners can be pre-determined by electrostatics prior to binding [213]. Such orientation could promote an encounter complex that results in specific residue-residue contacts being made rather than the complex spending time is local lower affinity energy minima. It is also possible these disordered loops help to reduce the movement of the ATS-Core on α17, which was shown to be dynamic in MD runs when no disordered regions were included. Ideally, future experiments should try to account for how these additional disordered regions increase affinity, although this would be challenging. NMR assignments could be extended beyond the ATS-Core to include the disordered regions. This would result in larger complexes and more complex spectra, but using selective labelling and different pulse sequences, such as TROSY, would help resolve these size and spectral overlap issues. MD could also be helpful in examining this issue, but care would have to be taken how the structure of the disordered regions is defined.

4.6.3 *Pf*EMP1 Localisation

Initial FP binding assays (Fig. 4.1C, D) results suggested the ATS bound to many different sub-fragments of spectrin, but seemed to bind with higher affinity to α17. The low affinity ATS has for other regions of spectrin could play a role in *Pf*EMP1 localisation.

As *Pf*EMP1 is in the membrane, it would diffuse in a two dimensional fashion. Attraction to spectrin would change this into linear diffusion along spectrin, potentially enabling it to find its higher affinity site faster.

In context to the whole erythrocyte cytoskeleton, α17 is towards the C-terminal end of the α chain, placing it close to the junctional complex where many spectrin tetramers meet (Fig. **??**). This would result in clustering of *Pf*EMP1 receptors, where clustering is a key step in many adhesion processes. Comparing this to adhesion via integrins, which adhere cells to the extracellular matrix, integrin $\alpha_5\beta_1$ clustering result in it being able to withstand forces six times stronger than alone, while another integrin, $\alpha_v\beta_3$ uses an additional protein, talin, to strengthen interactions with cytoskeletal actin [214]. *Pf*EMP1 may use both these strategies in adhesion. The ATS interaction with α17 spectrin could localise and cluster *Pf*EMP1 and this could be reinforced

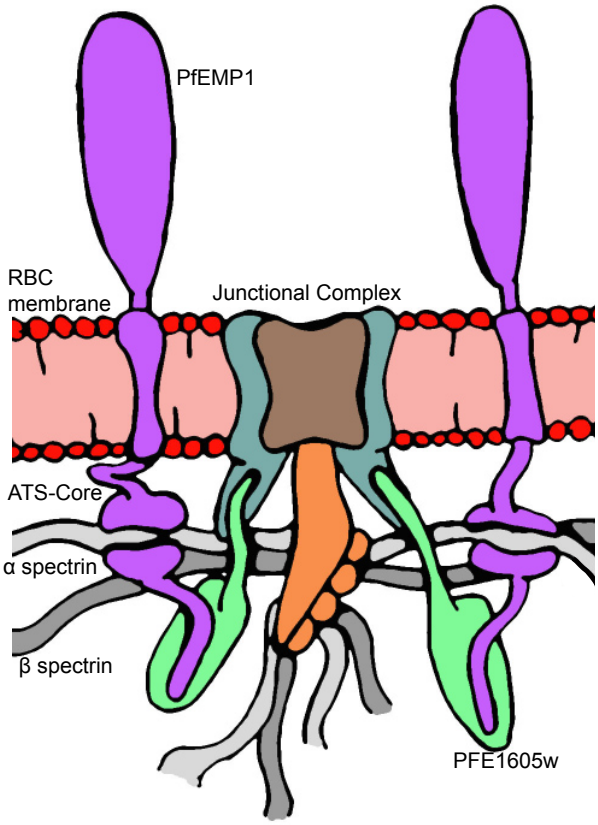


Fig. 4.21: ATS cytoskeleton interaction model. *Pf*EMP1 is localised around the junctional complex via interaction with α 17 spectrin. Additional linkage to cytoskeleton are made via PFE1605w, which binds Band 3 as described in Chapter 3.

with other proteins, such PFE1605w from Chapter 3, which also binds Band 3 and will be discussed further in Chapter 6. Serum proteins such as α2-Macroglobulin and Immunoglobulin have also been found to interact with the ecto-domains of *Pf*EMP1, crosslinking them together [215, 216]. The size and spacing of the spectrin junction, and these serum factors is compatible, hence *Pf*EMP1 could bind both spectrin and serum factors resulting efficient clustering and adhesion.

Chapter 5

KAHRP-Spectrin

"Must go faster"

Jeff Goldblum,

Jurassic Park and Independence Day

This chapter includes work from the publication;

Interactions of *P. falciparum* KAHRP and *Pf*EMP1 with spectrin suggest a model for adherent knobs on the infected erythrocyte surface, **Cutts, E. E.**, Laasch, N, Reiter, D. M., Trenker, R., Slater, L. M., Stansfeld, P. J. and Vakonakis, I, (2016), to be submitted.

5.1 Contributions to Data

Binding assays of wild type KAHRP sequences with spectrin and ankyrin for this work were performed by Niklas Laasch and Dirk Reiter. In all cases, raw data was reanalyzed to ensure consistent processing. XPLOR-NIH structure generation was performed with Ioannis Vakonakis.

5.2 Characterising KAHRP-Cytoskeletal Interaction

5.2.1 KAHRP Binds β Spectrin Repeats 10-14

KAHRP is a ~650 amino acid highly disordered protein (Fig. 5.1A). The N-terminal half, referred to as K1 (Fig. 5.1B), has a histidine-rich region that has been demonstrated to associate with the erythrocyte membrane [38] and a region, referred to as K1D, that binds ankyrin [62, 88]. The C-terminal half of KAHRP, divided into K2 and K3 segments, comprises two sequence repeat elements (5´ and 3´). The KAHRP C-terminal half and fragments therein associate with spectrin [63, 87]. To localize the KAHRP–spectrin interaction within defined sites of the two molecules, fluorescently labelled recombinant KAHRP K2 and K3 were incubated with fixed concentrations of recombinant constructs spanning the spectrin α and β chains (Fig. 5.1C, Appendix A), and variations in fluorescence polarization (FP) that result from slower tumbling of the labelled protein upon binding were determined (Fig. 5.1D). Two spectrin constructs spanning α chain domains 12-16 (henceforth α12-16) and β chain domains 10-14 (β10-14) produced significant variations of polarization in KAHRP K2 indicative of binding. No spectrin association with KAHRP K3 was observed.

Titrations of KAHRP K2 with spectrin α12-16 and β10-14 provided estimates for the dissociation constants, K_d, of 160$\pm$60 μM and 50$\pm$14 μM, respectively (Fig. 5.1E), thereby showing that KAHRP K2 associates with spectrin β10-14 with higher affinity. To independently verify this interaction, ^{1}H-^{13}C-HSQC NMR experiments were performed where ^{13}C-labelled KAHRP K2 was titrated with unlabelled spectrin β10-14. In these assays each resonance peak of the NMR spectrum corresponds to a C-H bond, hence there should be one or more peak for each amino acid of KAHRP K2. When inspecting the Cα-Hα region of the spectra I observed significant variation in the spectral appearance (position and intensity) of specific KAHRP K2 peaks upon addition of spectrin β10-14 (Fig. 5.2A,B) consistent with the interaction of a disordered protein with a folded component.

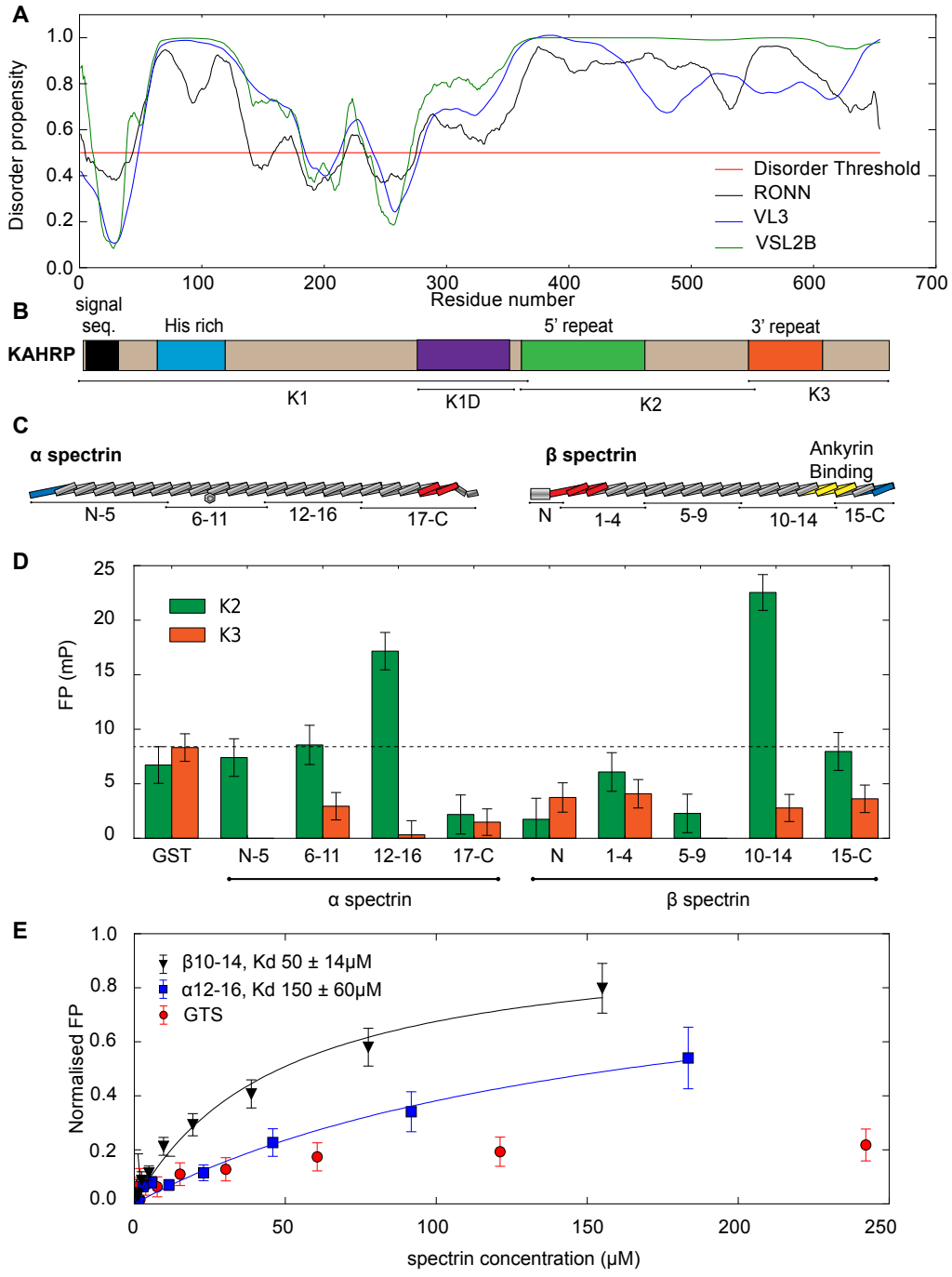


Fig. 5.1: KAHRP-spectrin interaction. **A**: Disorder propensity prediction for KAHRP generated using the RONN [217], VL3 [218] and VSL2B [219, 220] prediction methods, showing that the majority of the sequence is disordered. **B**: Schematic of KAHRP regions. **C**: Schematic of α and β spectrin, showing construct boundaries. **D**: FP assay using 5-FAM labelled KAHRP K2 or K3 fragments with 75 μM of spectrin, indicating only K2 binds α12-16 and β10-14 higher than the negative control, Glutathione S-transferase (GST). **E**: FP titration of 5-FAM labelled K2 with α12-16 and β10-14, with fit K_d, showing β10-14 binds with stronger affinity.

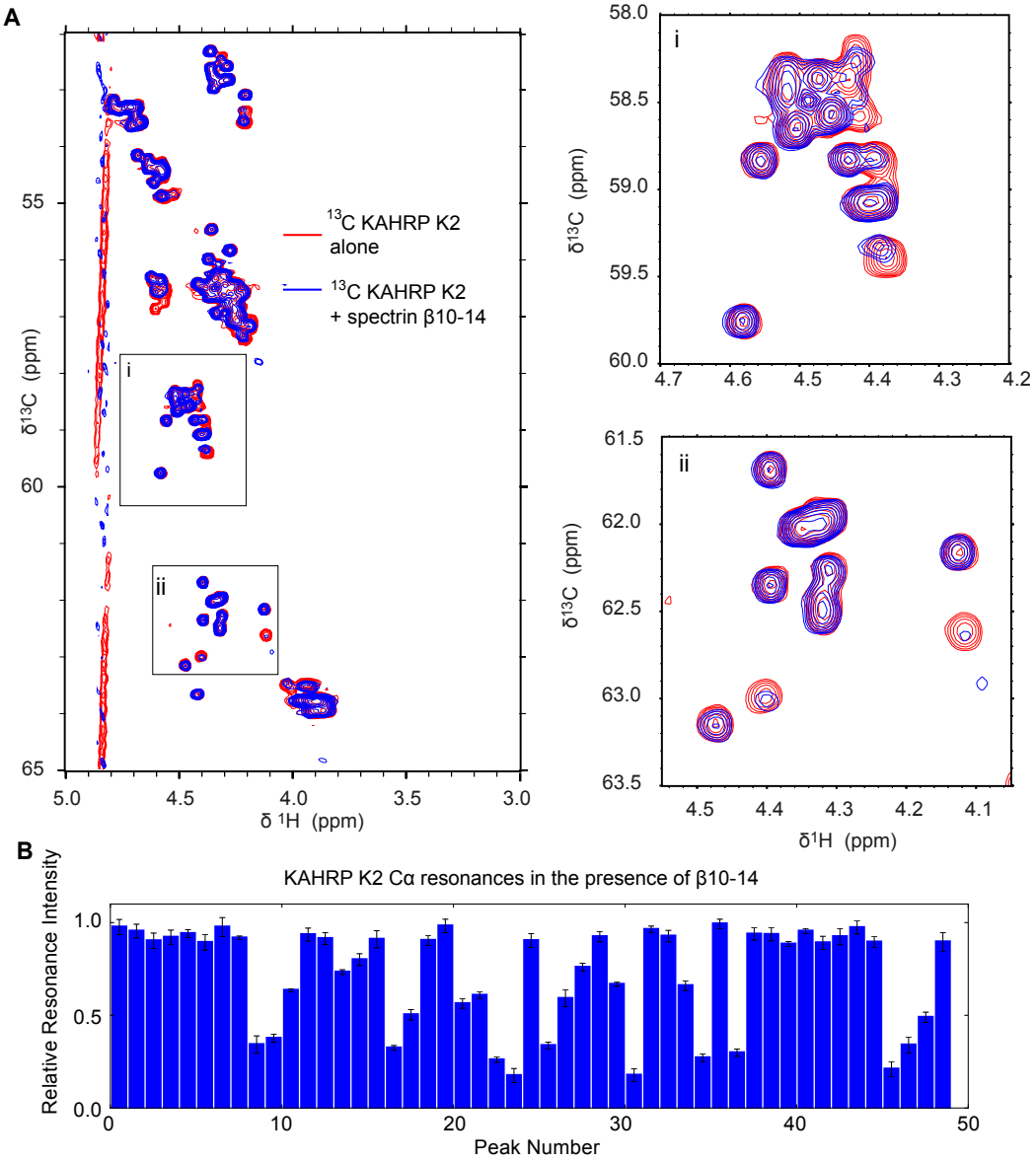


Fig. 5.2: KAHRP-spectrin NMR. A: ^{1}H-^{13}C-HSQC with 50 μM of KAHRP K2 alone (red) or with 3-fold excess of β10-14 spectrin (blue) focusing on the Cα-Hα region of the NMR spectrum. **i** and **ii** show enlarged regions, illustrating reduction in peak intensity. **B**: Quantification of relative peak intensity of isolated peaks in the Cα region of the K2 ^{1}H-^{13}C HSQC spectra, with error bars indicating one standard deviation based on spectra signal to noise.

5.2.2 Truncations of KAHRP and β10-14 Reduce Affinity

To further localize the KAHRP–spectrin interaction, titrations were performed of fluorescently labelled KAHRP K2 sub-fragments with spectrin β10-14, and similar titrations of labelled KAHRP K2 with overlapping spectrin sub-fragments spanning β chain domains 9 to 15 (Fig.5.3A). Truncations of KAHRP K2 to constructs comprising solely the 5´-repeat region did not affect spectrin β10-14 binding (Fig. 5.3B). In contrast, truncations of KAHRP K2 that removed 5´-repeat sequence elements reduced spectrin β10-14 binding in a manner proportional to the number of repeats eliminated. Removal of one or more spectrin repeat domain of β10-14 resulted in a similarly proportioned reduction of KAHRP K2 affinity (Fig. 5.3C). Hence, more sequence repeats of K2 or more structural repeats within β10-14 confer greater affinity. This suggests a binding mechanism where KAHRP K2 repeat region binds along β spectrin repeats 10-14 with an additive binding energy.

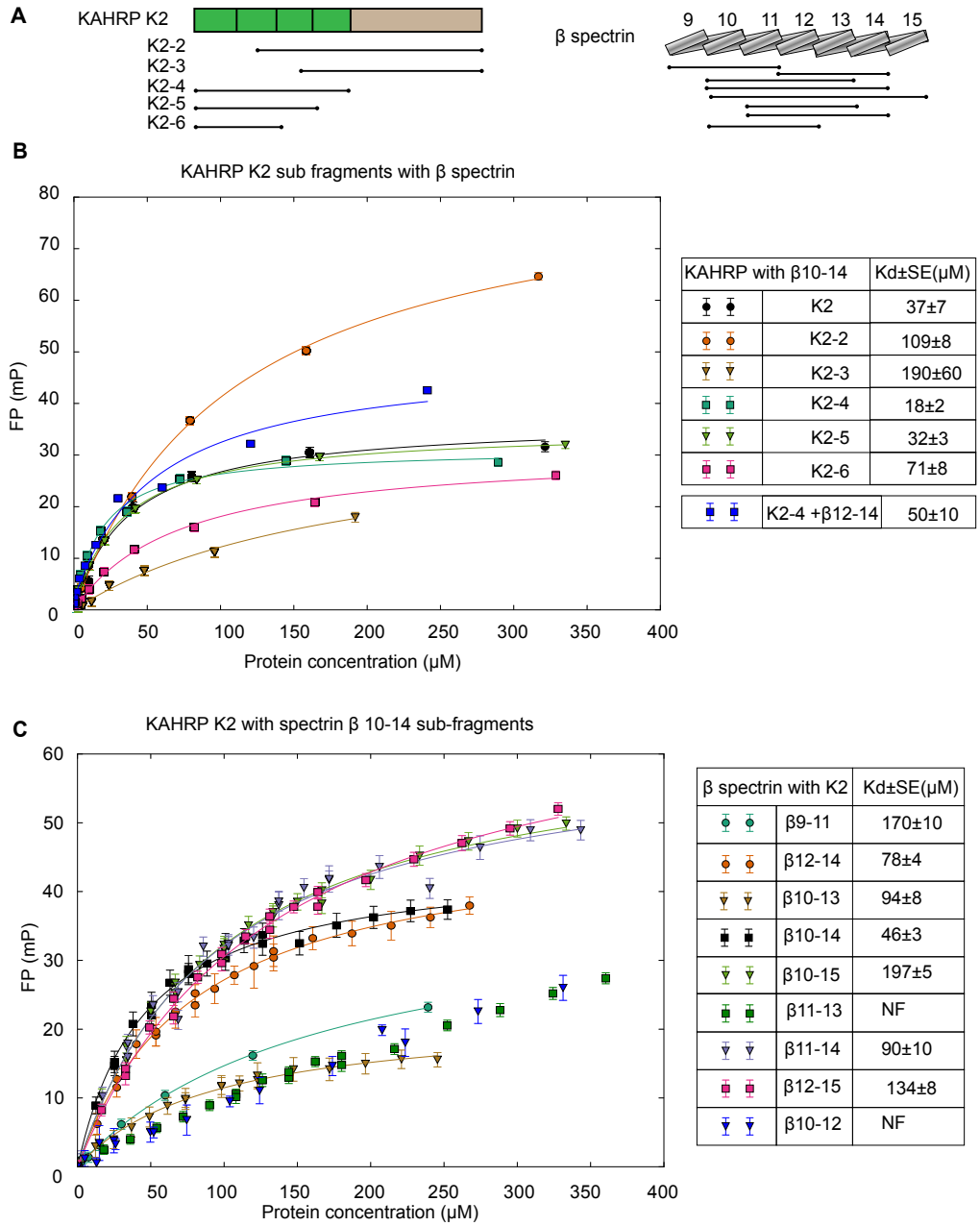


Fig. 5.3: FP assays of KAHRP and spectrin truncations. **A**: Constructs of KAHRP K2 and β9-15 tested for binding. **B**: FP assay of truncations of K2, each performed with 0.5 μM of 5-FAM labelled K2 truncation with increasing β10-14 or 12-14 spectrin. **C**: Truncations of β spectrin, each tested with 0.5 μM of K2 and increasing concentrations of spectrin. NF indicates data that could not be fit. Hence, more sequence repeats of K2 or more structural repeats within β10-14 confer greater affinity.

5.2.3 KAHRP Binds Ankyrin Spectrin-Binding Domain via K1D

Spectrin β10-14 is proximal to the host ankyrin interaction site, β14-15 [99], and previous pull-down studies suggested a KAHRP–ankyrin association via a region just N-terminal to the KAHRP 5´-repeat [99]. Such KAHRP–ankyrin binding would require an anti-parallel orientation of the KAHRP–spectrin β10-14 complex. To test this premise, the ability of a fluorescently labelled KAHRP fragment N-terminal to the 5´-repeat (K1D, Fig. 5.1A) to bind ankyrin domains (Fig. 5.4A) was evaluated. This resulted in specific association of the K1D fragment to the spectrin-binding domain of ankyrin (ZU5A, [99] with K_d of 25$\pm$2 μM, Fig. 5.4B). Thus, the KAHRP 5´-repeat associates in an anti-parallel manner with the region of the spectrin β chain close to the spectrin–ankyrin junction.

As this differs from previous reports suggesting KAHRP K2 bound α spectrin repeat 4 and K1D bound D3 of the membrane binding domain of ankyrin, these two possibilities were investigated with NMR. The ^{1}H-^{15}N HSQC spectra indicated no change in resonance position or peak intensity, suggesting an interaction of the reported affinity (2 μM for K2 with repeat α4 and 5 nM for K1D and D3) is not occurring (Fig. 5.4C,D). It is likely that these reported interactions were artefacts caused from protein immobilisation or aggregation, both of which were avoided by investigating the interaction with FP and NMR. The affinities of KAHRP for spectrin and ankyrin are weak but comparable to the strength of interactions seen between intracellular focal adhesion components in animal cells [221].

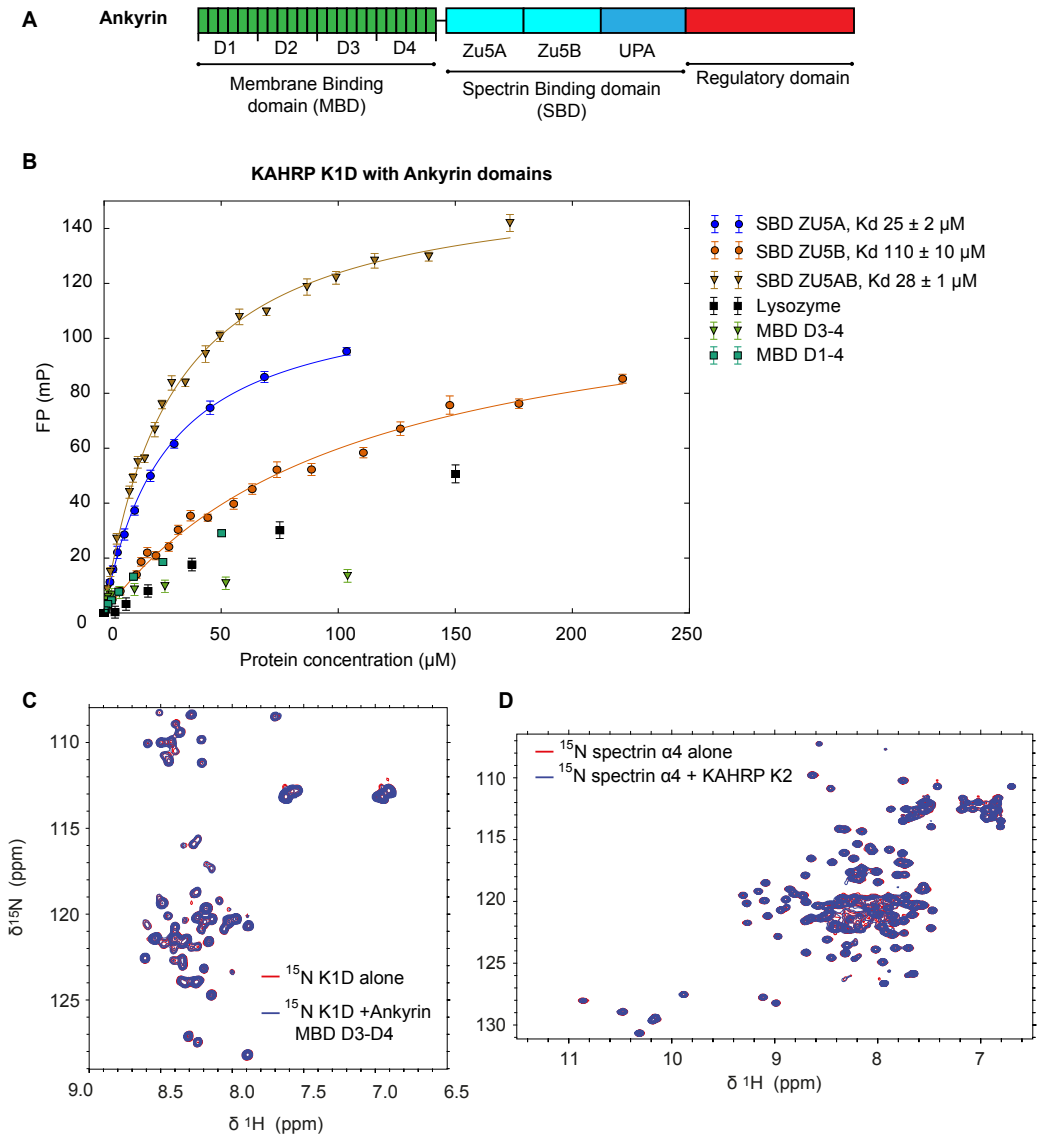


Fig. 5.4: KAHRP-ankyrin interaction. **A**: Schematic of the domains of ankyrin. **B**: FP binding assay using 0.5 μM of 5-FAM labelled K1D and ankyrin regions previously tested for KAHRP interaction, illustrating ankyrin ZU5A provides affinity, but not ankyrin D3 of the MBD. **C**: NMR ^{15}N-HSQC spectra of 50 μM with ^{15}N KAHRP K1D alone (red) and in the presence of 4-fold excess erythrocyte ankyrin MBD repeats 3 and 4 (blue), recorded at 10 °C. **D**: Similar overlay of 100 μM ^{15}N spectrin α spectrin repeat 4 alone (red) or with equimolar of KAHRP K2 (blue). The lack of resonance perturbations in panels **C** and **D** does not support the presence of strong binding for these construct combinations.

5.3 Complementary Electrostatics Binding Model

5.3.1 KAHRP-Spectrin Interaction is Dominated by Electrostatics

The KAHRP 5´-repeat is highly enriched in charged amino acids (Fig. 5.5A and B) suggesting that the KAHRP–spectrin binding is driven by ionic interactions. Indeed, titrations of KAHRP K2 with spectrin β10-14 under increased ionic strength conditions showed decreased affinities (Fig. 5.5C). As spectrin is primarily negatively charged under physiological conditions (calculated pIs of 4.95 and 5.15 for the α and β chains, respectively), while the KAHRP 5´-repeat is positively charged (pI of 9.65), I was intrigued by the mechanism by which the ionically-driven KAHRP–spectrin association gains specificity. That is, how K2 preferentially binds β10-14, despite the all spectrin regions having opposite charge of K2.

5.3.2 Analysis of Surface Electrostatics

To analyse the surface charge distribution for full length β spectrin, I derived a full length model using Phyre2 [169] to predict the structures of regions spanning 3-4 repeats and combining them with MODELLER [165] (Fig. 5.5D). Currently, 7 β erythroid spectrin repeat domain structures have been solved, and many other similar folds determined, making this a high-confidence model. The electrostatic potentials were determined for individual regions with PQR [222, 223] and APBS [224], protein accessible surface meshes were created using Chimera [225] with a default probe radius of 1.4 and a vertex density of 2 per Å, and the electrostatic potential at each mesh grid point was interpolated using the gridData python module from MDAnalysis [226]. Spectrin surface grid points were filtered for potential values above +10 kT/e or below -10 kT/e and clustered using the DBSCAN algorithm [227] with an ε cut-off of 3 Å. The centre of each cluster was determined and clusters were plotted using Matplotlib [196] as a function of distance from the protein N-terminus, with

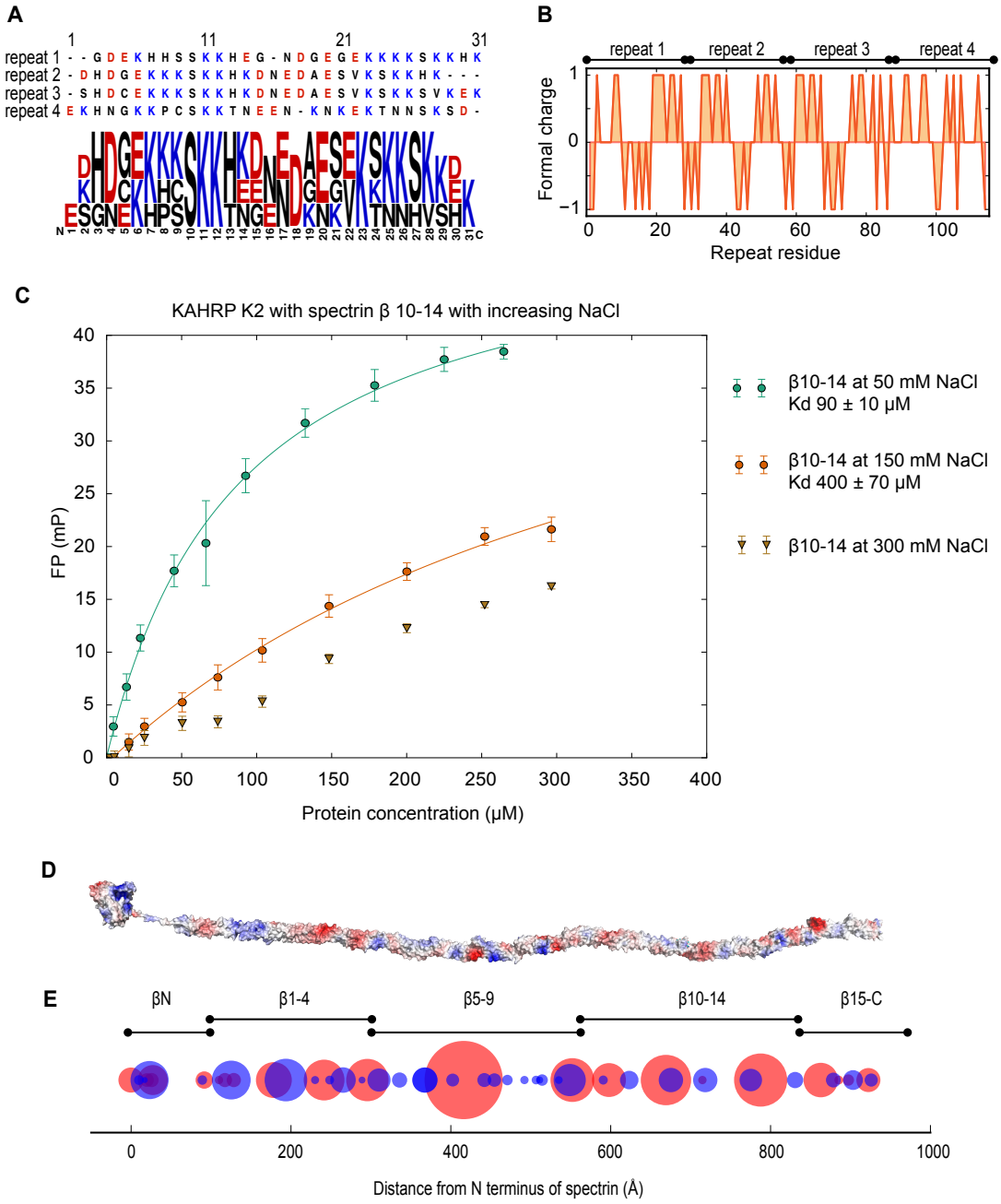


Fig. 5.5: KAHRP and β10-14 have complementary electrostatics. **A**: Alignment of the KAHRP 5´ sequence repeats with sequence consensus (bottom). **B**: Formal charge distribution in the 5´-repeat region of KAHRP derived from the amino acid sequence. **C**: FP titrations of KAHRP K2 with spectrin β10-14 under increasing NaCl concentrations, showing decreased interaction affinity at higher ionic strengths. A binding curve could not be fit at 300 mM of NaCl. **D**: Surface representation of full length β spectrin coloured by electrostatic potential (red, negative potential; blue, positive potential) generated by homology modelling and PyMOL [156]. **E**: Analysis of electrostatic clusters on the β spectrin surface. Each circle represents a location on the β spectrin surface with absolute electrostatic potential greater than 10 kT/e. The circle radius is proportional to the electrostatic cluster size. The boundaries of β spectrin constructs used are indicated.

radius being proportional to how many spectrin surface grid points were in each cluster, hence surface area (Fig. 5.5E). This analysis revealed that repeats β10-14 had both positively and negatively charged regions, which were regularly spaced, hence could provide a complementary charge profile to KAHRP K2.

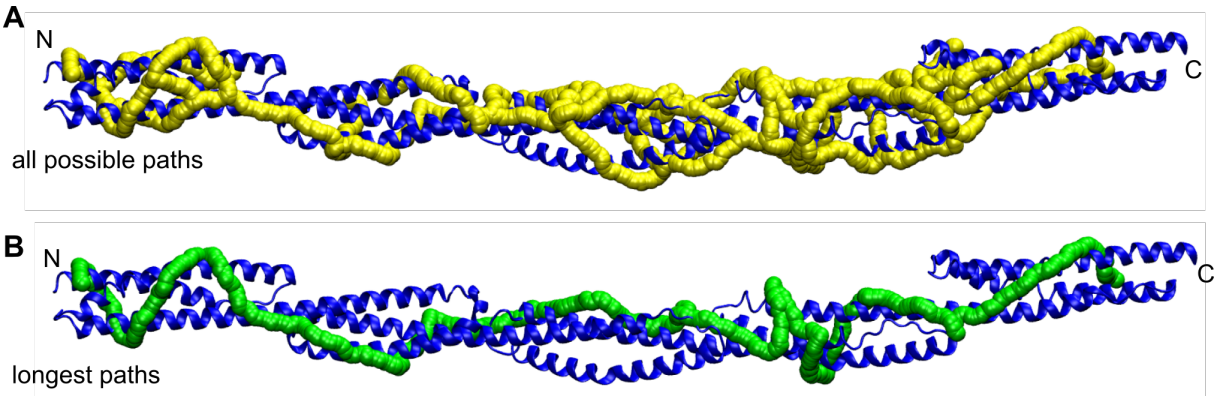


Fig. 5.6: Path finding. **A**: The structure of β10-14 spectrin (blue) with all paths that have complementary charge to KAHRP 5'-repeat shown (yellow). **B**: The longest path (green) predicted with the algorithm.

5.3.3 Complementary Path Finding Algorithm

Working under the hypothesis that KAHRP 5'-repeat region could find a path along spectrin β10-14 with a complementary charge profile, I built an algorithm which would find this path (Appendix E). A simple knowledge based search was undertaken using the information on charge dependence, charge distribution and interface size known for this system. The concept behind this approach was to simplify the system to align the alternating charges of the KAHRP 5'-repeats with the alternating charge patches across the surface of β10-14 spectrin. The following restrictions were applied to path finding by using features of the KAHRP 5'-repeats:

1. KAHRP 5'-repeats have alternating charge, hence the path on the spectrin surface must transverse between positive and negative charge clusters.

2. The periodicity of charge in KAHRP 5'-repeat is ~10 residues, hence the distance between positively and negatively charged regions of KAHRP is up to ~40 Å, assuming an extended protein backbone conformation. Therefore, I selected for possible spectrin surface paths featuring distances greater than 10 Å, but less than 30 Å, between positive and negative patches in Euclidean space.

3. As the length of KAHRP 5'-repeats is similar to that of spectrin β10-14, and removal of any spectrin domains reduces KAHRP affinity, I restricted the possible paths on the spectrin surface to those that do not back-track but instead utilize as many spectrin domains as possible.

4. An anti-parallel binding configuration of KAHRP and spectrin was enforced, derived from the data showing K1D binds ankyrin ZU5A, where ankyrin Zu5 binds spectrin repeats β14 and 15, hence the N-terminus of K2 is located at the C-terminus of spectrin β10-14.

The electrostatic surface mesh was determined for β10-14 spectrin as described above, but in this case, charge patches on spectrin were manually specified as the centre of clusters of up to 4 surface exposed charged residues within a 10 Å radius of each other (Appendix F). The structure of β10-14 was modelled, again using Phyre2 [169] and MODELLER [165], but also using the crystal structure known for β spectrin repeats 13, 14 and 15 [99]. Paths were found using the NetworkX [228] implementation of the Dijkstra's shortest path algorithm [229] between all positive and negative clusters that were greater than 10 Å, but less than 30 Å, in Euclidean space in accordance with criterion 2. These were then added to a new directional network graph, and all the simple paths between nodes 30 Å from the N-terminus and 30 Å from the C-terminus determined were determined, in accordance with criterion 3.

These selection criteria yielded ~2000 paths along the surface of spectrin β10-14. Examination of the nodes that these paths could be composed of showed that some regions of

spectrin are excluded to avoid regions of unfavourable or neutral charge distribution (Fig. 5.6A) but the longest path demonstrates that these paths were able to explore much of the surface of β10-14 (Fig. 5.6B).

5.3.4 Charge Complementation Scoring

Paths were scored based on possible electrostatic interactions of spectrin with the KAHRP 5´-repeat. The score function was based on a simplification of Coulumb's law, which states that any two charged particles experience a force that is proportional to their charge and inversely proportional to the square of their distance, that is:

$$F = k_{\mathrm{e}}\left(\frac{q_1 q_2}{r^2}\right) \tag{5.1}$$

For Coulomb's constant, k_e =8.99x10^9 Nm^2C^{-2}, point charges q_1 and q_2 and distance between the charges, r. This is an attractive force for particles of opposite charge, and a repulsive force for particles of similar charge. While I know the charge on the surface of spectrin; the distance between KAHRP and spectrin is not known, and the conformation of KAHRP is unknown, hence charge distribution of KAHRP cannot be determined. To avoid over biasing the score based on spectrin's charge distribution, a simple two-level score function was used, with a favourable energy for complementary charges and a energy penalty for charge clashes as follows:

- if the charges are opposite, -1 is added to the score.
- if the charges are opposite, and the absolute electrostatic potential of spectrin is greater than 10, a further -1 is added to the score.
- if the charges are like, +1 is added to the score.
- if the charges are like, and the absolute electrostatic potential of spectrin is greater

than 10, a further +1 is added to the score.

This scoring function was applied, enforcing an anti-parallel orientation KAHRP on spectrin from criterion 4, a KAHRP inter-residue spacing of 3.8 Å and a moving window of 15.2 Å to allow for different lengths of paths. Scoring all paths resulted in a roughly Gaussian distribution of scores (Fig. 5.7A), with the lowest score being -324 resulting in a highly complementary path (Fig. 5.7 B, C). Furthermore, the tail of the Gaussian distribution suggests that there are other paths of similar energy, suggesting that KAHRP could adopt multiple, similarly probable configurations whilst bound to spectrin. This could explain why removal of KAHRP sequence repeats, or of spectrin domains, reduces but does not eliminate KAHRP–spectrin binding.

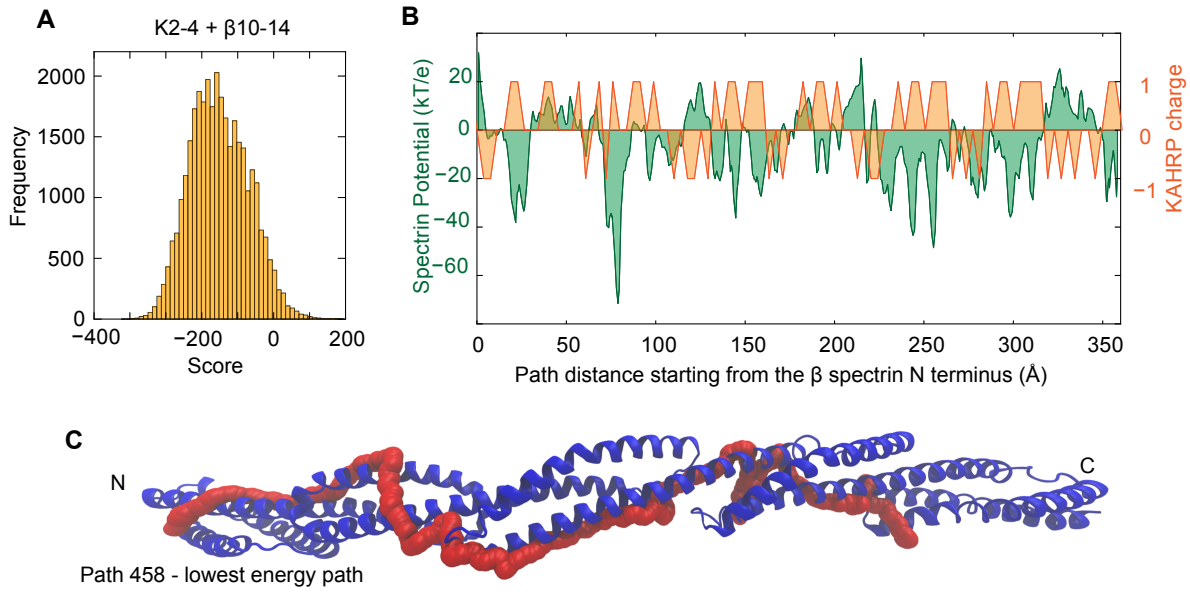


Fig. 5.7: Scoring of paths. **A**: Histogram showing the range of scores determined by the simple score function, with the bulk of them being favourable due to net opposing charge. **B**: The electrostatics of the path with the lowest score (green), with the KAHRP formal charge distribution superimposed (orange), showing charge complementation. **C**: The trajectory of the lowest scoring path wraps around spectrin β10-14.

To examine whether other paths that had favourable scores were unique, and might result in an alternate possible path, all paths were compared to the lowest energy path. A

path similarity score was defined for each path(i) composed of nodes, n_i, with the lowest energy path(0) composed of nodes, n_0, as follows:

$$Path\ similarity\ score\ = \sum_{\forall n_0} (min_{\forall ni}(distance\ between\ n_0\ and\ n_i)) \tag{5.2}$$

Plotting this against the charge complementation scores for each path with the optimal frame shift suggested that as paths diverged in space from the lowest energy path, there was a coupled increase in energy (Fig. 5.8A). There were some cases where a path with a low path similarity score had a high charge complementation score. Visual inspection of these paths found that they typically included additional regions that resulted in charge clashes. The charge complementation scores as a function of path length was plotted alongside, and showed that there was no correlation between path length and energy, and that paths with both low path similarity score and low charge complementation score, being roughly the same size (Fig. 5.8B).

Scoring was initially performed enforcing an anti-parallel orientation, but as the FP binding assays lacked ankyrin and the ankyrin binding domains of both KAHRP and spectrin, the *in vitro* K_d measurement could include conformations binding in a parallel orientation. The parallel orientation of KAHRP 5'-repeat was also scored, resulting in the lowest score being -304, (Fig. 5.9A,B), suggesting the parallel orientation is slightly less favourable. Furthermore, the score distribution of the anti-parallel orientation is skewed towards more favourable scores, while the parallel orientation is not skewed, hence there are more favourable scores is the anti-parallel orientation. This suggests the imperfect charge repeats of K2 KAHRP may have evolved to optimise a directional specific interaction.

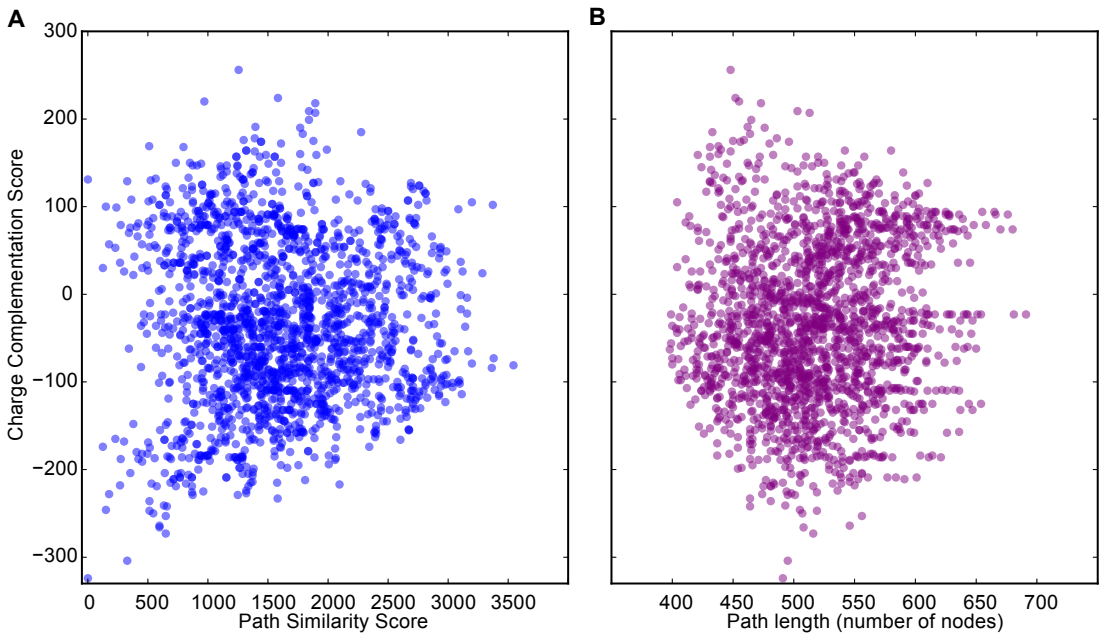


Fig. 5.8: Lowest scoring paths over spectrin are similar. **A**: Plot of the charge complementation score against the path similarity score for the lowest energy path, indicating that a reduction in similarity to the lowest energy path results in a decrease of charge complementation. **B**: Plot of the charge complementation score against path length, which shows no correlation.

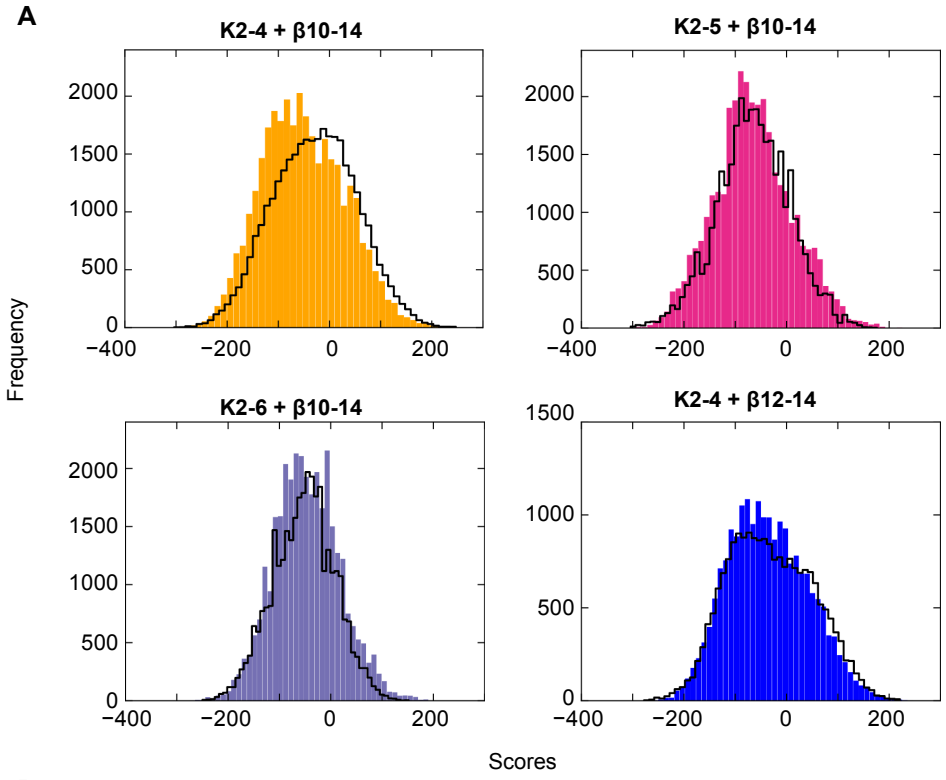
5.3.5 Analysis of Truncations

In order to validate this methodology, I applied the same analysis to KAHRP 5'-repeat and β spectrin truncations, namely K2-5 and K2-6 with β10-14, and K2-4 with β12-14 (Fig. 5.9A). The lowest score for each of these was -293, -275 and -272, which correlates well with the respective K_d of 32$\pm$3, 71$\pm$8 and 54$\pm$13 μM (Fig. 5.9B). Again the parallel orientation was examined, resulting a more favourable score for parallel orientation for the K2-5 with β10-14 and K2-4 with β12-14. Where the score is based on the electrostatic force, which is proportional to electrostatic energy, taking the ratio of the most favourable score attained by each truncation with the wild-type should be roughly equivalent the ratio of the binding energies. Experimentally derived binding energies were determined using Gibbs free energy change:

$$\Delta G = RTln(K_d) \tag{5.3}$$

where R is the gas constant 8.13 $JK^{-1}mol^{-1}$, T is the temperature and K_d is the dissociation constant of KAHRP-spectrin binding. The ΔG ratio for K2-4 binding to β10-14 versus β12-14 was ~1.11, compared to a docking score ratio of ~1.17 (-324 versus -278 for β10-14 and β12-14, respectively, Fig. 5.9B). A similar trend was present for the K2-4 truncation, K2-5 and K2-6, with energy ratios of ~1.06 and ~1.14 and score ratios of ~1.07 and ~1.19. Given the crude nature of the scoring function this correlation is impressive and suggests this simple model can broadly reproduce experimentally-derived data in this system.

A



B

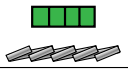
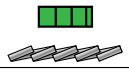
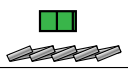
				
	K2-4+ β10-14	K2-5+ β10-14	K2-6+ β10-14	K2-4+ β12-14
Lowest Anti-parallel Score (parallel)	-324 (-304)	-293 (-304)	-275 (-250)	-272 (-278)
Anti-parallel Mean Score	-51	-65	-46	-36
Score ratio	1	1.07	1.19	1.17
Kd ± SE (μM)	18 ± 4	32 ± 3	71 ± 8	54 ± 13
ΔG ($kJmol^{-1}$)	-26.5	-25.1	-23.1	-23.8
ΔG ratio	1	1.06	1.14	1.11

Fig. 5.9: Truncation validation. **A**: Charge complementation score histograms of KAHRP K2-4, K2-5 and K2-6 with β10-14 and K2-4 and β12-14. Filled plot is for anti-parallel orientation, black line is for parallel orientation. **B**: Tabulated comparison of score to binding energy derived from experimental K_d. Score ratio determined by dividing the lowest score for K2-4 with β10-14 by the lowest score in each case, whether anti-parallel or parallel.

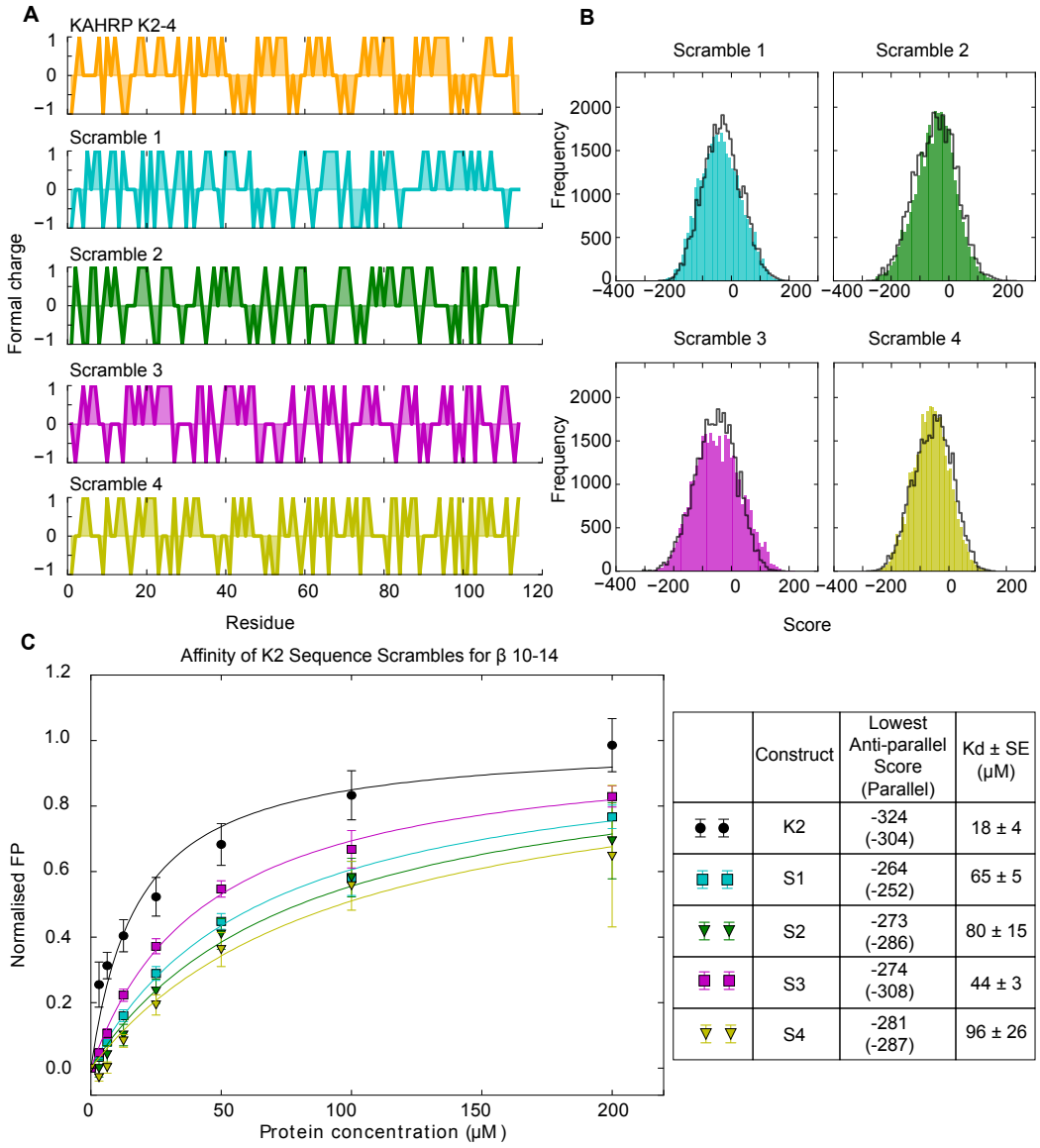
5.3.6 Charge Distribution is Required for Affinity

To further validate this methodology and test the charge complementation hypothesis, the alternating charge distribution of the KAHRP 5'-repeat was removed by randomizing the sequence (sequences are shown Appendix F). This allows for the same net charge on the KAHRP 5'-repeats, hence eliminates the possibility of the interaction being non-specific electrostatic effect. Ten randomised sequences were produced and analysed using the same method above, resulting in a range of scores (Appendix F). The four sequences resulting in the least favourable lowest charge complementation score were recombinately produced, labelled with 5-FAM via an N-terminal cysteine residue and tested for binding to β10-14 with florescence polarization assays. In all cases the sequence scrambles bound with lower affinity (Fig. 5.10C, D), suggesting the alternating charge distribution is involved in binding.

5.4 Atomistic Modelling of KAHRP K2-4 and β10-14 Spectrin

5.4.1 Model Generation

To further visualise the potential complexes being predicted using the path finding algorithm, an atomistic model was generated based on the lowest energy path. This model included the region of KAHRP that was aligned using the charge complementation docking algorithm (residues 362-458, where the 5' repeat spans residues 356-470). An initial molecular ensemble of the KAHRP–spectrin β10-14 complex was derived using XPLOR-NIH [230]. NOE-like distance restraints were applied between KAHRP residues and spectrin β10-14 residues defining surface paths derived by electrostatic-driven docking. Similar restraints were derived from the spectrin β10-14 structure alone and used to limit the conformational space of the spectrin backbone. The complex was further restrained by a confor-



	Construct	Lowest Anti-parallel Score (Parallel)	Kd ± SE (μM)
●●	K2	-324 (-304)	18 ± 4
■■	S1	-264 (-252)	65 ± 5
▼▼	S2	-273 (-286)	80 ± 15
■■	S3	-274 (-308)	44 ± 3
▼▼	S4	-281 (-287)	96 ± 26

Fig. 5.10: K2-4 sequence scrambles. **A**: The formal charge distribution for each of the sequence scrambles compared to the wild-type sequence KAHRP K2-4 sequence. **B**: Charge complementation score histograms for each of the sequence scrambles. Filled plot is for anti-parallel orientation, black line is for parallel orientation. **C**: FP binding assay of K2-4 sequence scrambles, labelled with 5-FAM via N-terminal cysteine, against β10-14, with fit K_d and charge complementation score shown. FP data shown is the mean and standard error of four technical repeats and three plate repeats.

mational database potential [231] and a potential of mean force that conducts a free-search for putative hydrogen bonds during the simulation and optimizes the spatial arrangement of peptidyl backbone units [232]. This resulted in an ensemble of atomistic structures that followed the lowest energy path predicted by the algorithm and had optimised atomic geometry.

5.4.2 Restrained MD Simulations

The lowest energy structure from the ensemble of docked conformations was used to set up 50 ns molecular dynamics (MD) simulations in triplicate using the all atom force field AMBER99SB-ILDN [183] with TIP3P water, 50 mM NaCl and a temperature of 298 K. Positional restraints were placed on the Cα atoms of spectrin and simulations were performed in a box 3 nm bigger than each spectrin dimension.

The Cα RMSD of KAHRP during the MD simulation converged rapidly (~20 ns), Cα RMSF showed limited dynamics for a disordered protein, and the number of ionic interactions was optimised throughout the simulation (Fig. 5.11A, B, C). Inspection of the electrostatics after 50 ns illustrate multiple interactions between charged KAHRP residues and complementary charged clusters of spectrin (Fig. 5.12A). One simulation was extended to 100 ns, with no further changes were observed in the additional time. Overlays from every 1 ns of the simulation show that KAHRP remains bound to the surface, with predominantly local fluctuations (Fig. 5.12B, C). In run 3, the C-terminal region of KAHRP lost ionic interactions with spectrin, but after a brief moment of Brownian motion, found a new electrostatic patch, and stayed bound there for the remainder of simulation (Fig. 5.12C). Hence, although the docked KAHRP remains bound, it may be able to sample other conformations, potentially some of the similar energy paths predicted in docking.

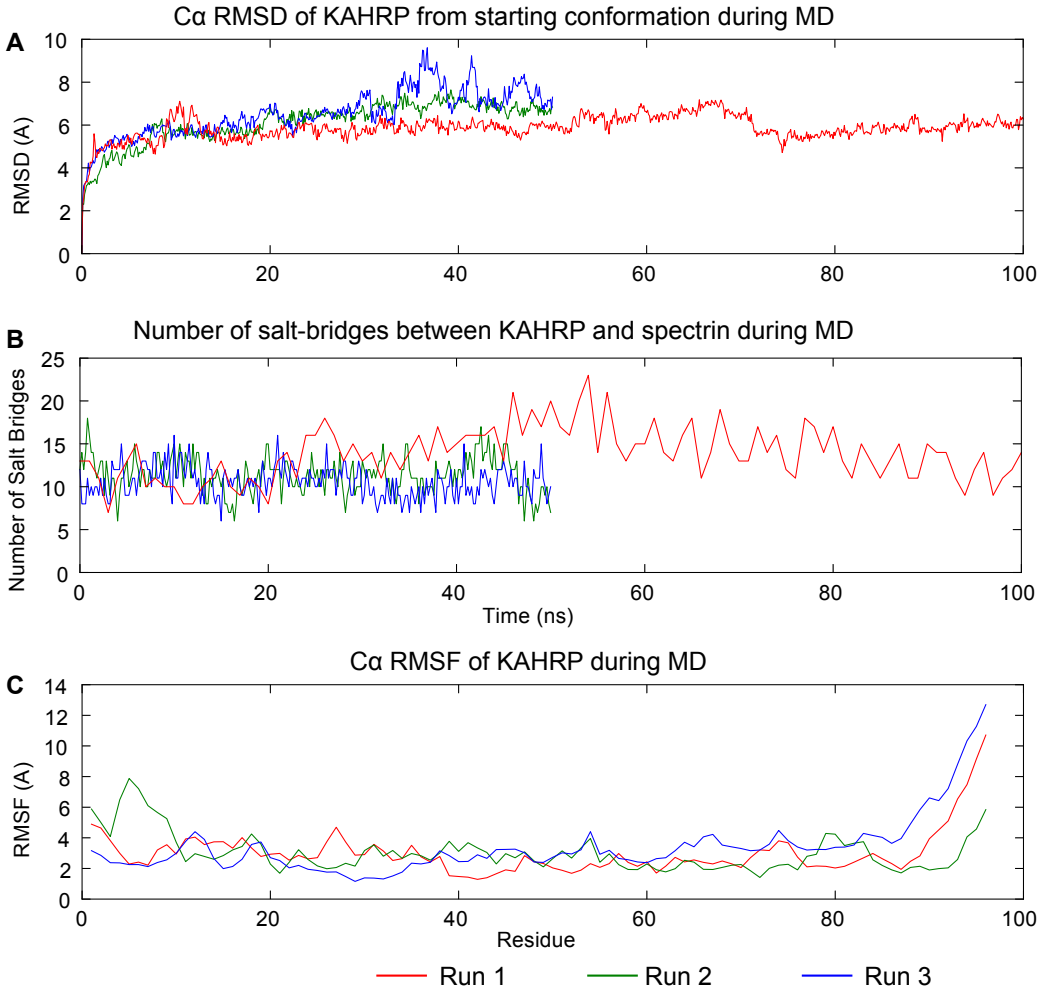


Fig. 5.11: MD of KAHRP bound to β10-14 spectrin. **A**: The Cα RMSD of KAHRP stabilises after 20 ns of MD. **B**: The number of ionic contacts between KAHRP and spectrin during MD simulation. **C**: Cα RMSF per KAHRP residue.

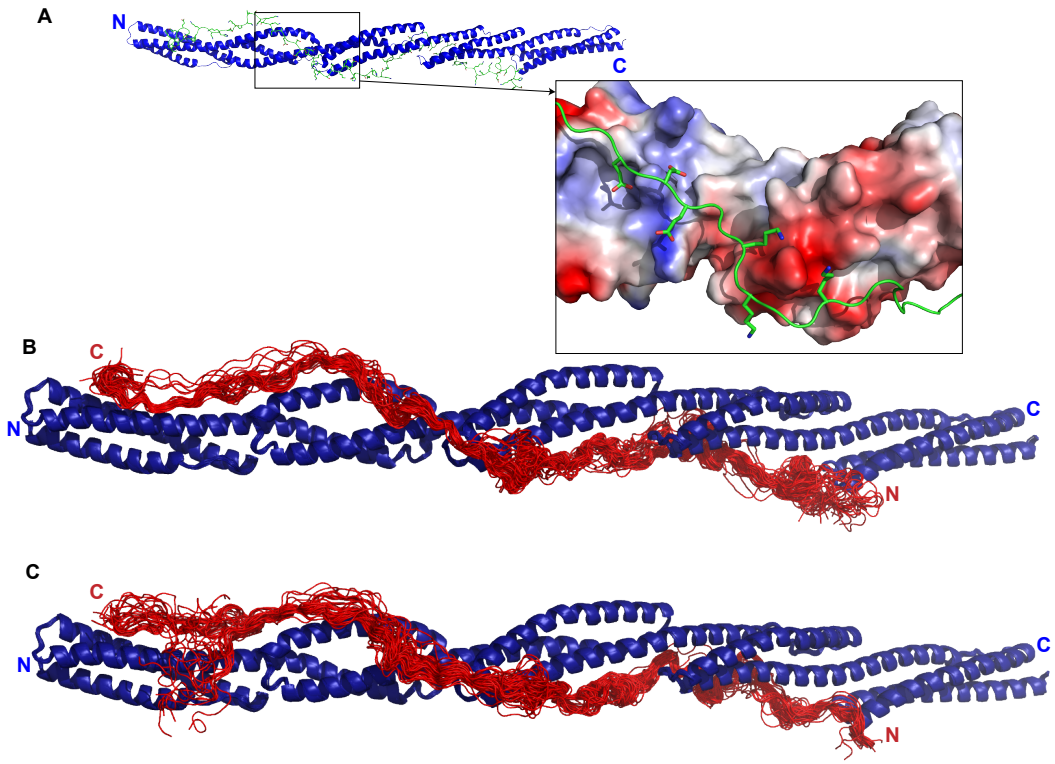


Fig. 5.12: KAHRP-spectrin structure after MD. A: An enlargement showing complementary charge interactions after 50 ns simulation. **B, C**: Overlays from every 1 ns from 50ns simulations for run 1 and run 3 respectively, showing KAHRP (red) remains bound to spectrin (blue). In run 3 (**C**), the C-terminus of KAHRP moves away from original docked position to a new stable position on spectrin.



Fig. 5.13: KAHRP dynamics in unrestrained versus restrained MD simulation. **A**: The Cα RMSD of KAHRP is larger in the unrestrained simulation. **B**: The Cα RMSF of KAHRP is comparable in both restrained and unrestrained simulations for most regions of the KAHRP.

5.4.3 Unrestrained MD Simulations

The positional restraints on β10-14 enforced an extended conformation of spectrin, but spectrin is observed in a condensed form in cryo-electron microscopy of the erythrocyte cytoskeleton [95]. Hence, these positional restraints were removed after 25 ns of restrained simulation and the simulations run for an additional 50 ns. Spectrin β10-14 was run alone as a control. Removal of restraints results in more dynamic behaviour of both KAHRP and spectrin. The Cα RMSD of KAHRP is much greater, but the Cα RMSF is comparable for the majority of the sequence (Fig. 5.13A and B), hence the increase in RMSD likely represent a more dynamic complex as a whole. β10-14 spectrin with KAHRP 5'-repeats bound was also more dynamic, with a larger Cα RMSD and RMSF, and greater changes in the bend angle between the 10th, 12th and 14th repeat (defined by tryptophan residues 1361, 1566, 1705) indicating the structure spends more time in bent conformation (Fig. 5.14A, B and C). The Cα of β10-14 RMSF spectrin alone showed that the loop regions between helices were more dynamic, enabling the bending motions. These loop regions all include charged residues, which make contacts with helices to stabilise the structure. The Cα RMSF of these loops was greater in the presences of KAHRP, as KAHRP competed for these stabilising intra-molecular ionic interactions. Even though there is a net increase in the frequency of bending events, the bend angle when KAHRP is bound is similar to that seen in some of the simulations where spectrin is alone. This suggests KAHRP binding does not lead to gross structural changes in spectrin, although longer simulations should be performed in order to confirm this conclusion.

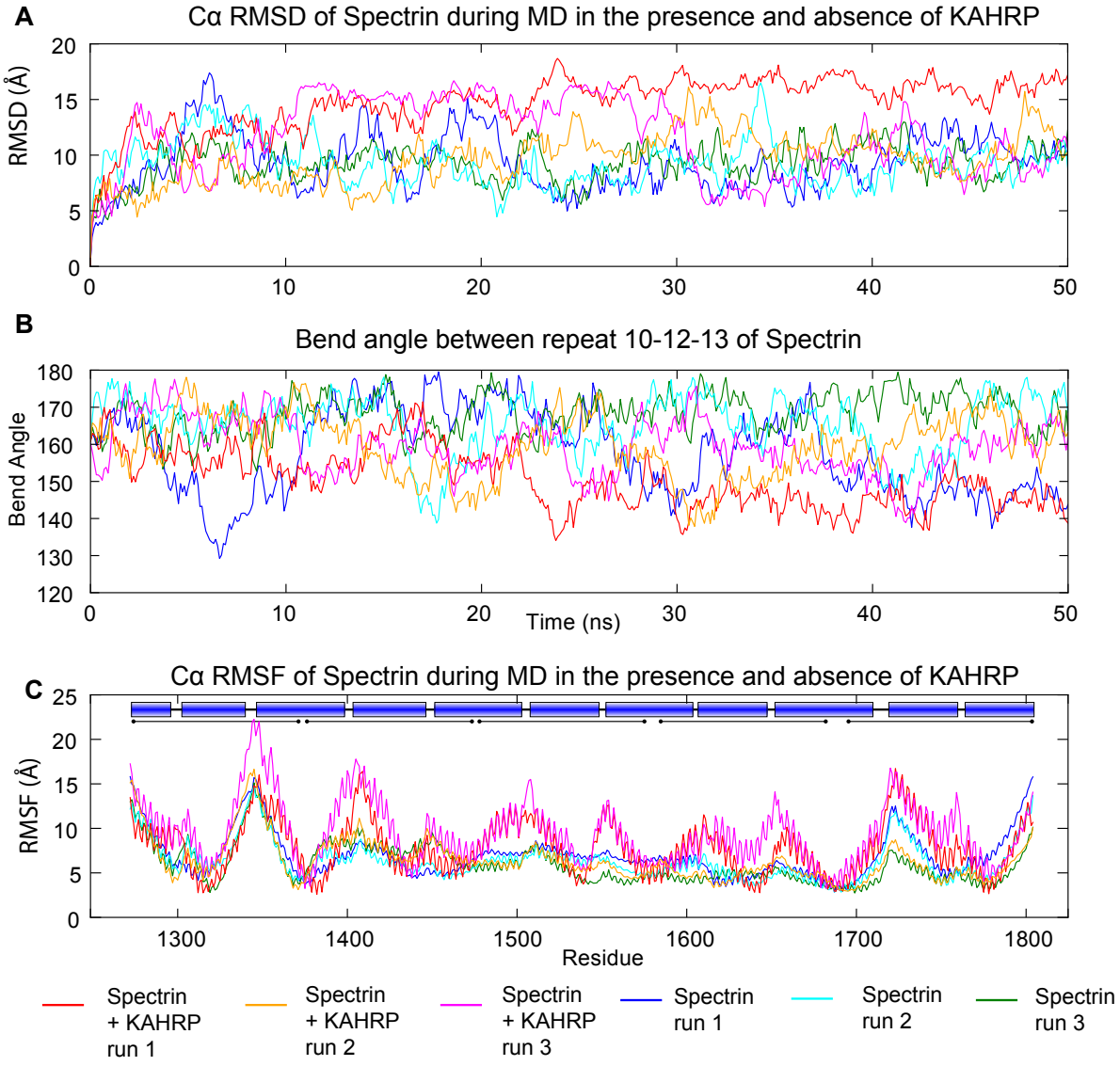


Fig. 5.14: MD simulations suggest β10-14 spectrin is more dynamic when KAHRP is bound. **A**: The Cα RMSD of β10-14 spectrin. **B**: The bend angle between the 10th, 12th and 14th repeat, determined with the Cα residues of tryptophan residues 1361, 1566, 1705. **C**: The Cα RMSF of β10-14 spectrin, blue rectangles indicate the approximate helical regions, with the lines indicating the approximate span of each repeat. The spiky features of the plot are due to the helical structure, with residues forming the hydrophobic core being less dynamic.

5.5 Discussion

5.5.1 The Docking Algorithm

This chapter introduced a novel methodology for docking a large disordered protein to a structured protein, driven by electrostatic interactions. Previous attempts to look at complexes involving disordered proteins have been limited to short peptide, such as FlexPepDock [233] and HADDOCK [171, 172], but this approach docked ~100 residues, the longest disordered protein for which docking has been attempted. This methodology was deliberately kept simple, as the conformational space of a large disordered protein, the possible binding modes and the resultant energy landscape is unfathomably vast and complicated. Energetic contributions from hydrogen bonds and van der Waal interactions were ignored and the electrostatic component was highly simplified. KAHRP was also forced to interact with the surface of spectrin, where in reality it could loop off, hence the inter Cα distance of KAHRP with respect to the path along spectrin is likely fluctuate greatly from the approximation used, 3.8 Å. But even with all the simplifications made to the method, the output compared well to experimental data. Analysis of truncations of spectrin and KAHRP showed a good correlation between score ratios and energy ratios. Although it was expected that truncations would have a lower binding energy, due to smaller interface size, the energy ratio between the 5 repeat spectrin construct, β10-14, and the 3 repeat spectrin construct, β12-14, should be more on the order of 5/3, ~1.67 if construct size was the sole energy determinant, rather than the 1.06 derived from experiments.

The sequence scrambles of the KAHRP 5' repeat with low charge complementation score all bound β10-14 with lower affinity. This demonstrates that the scoring function was also capable of predicting the behaviour of the sequences that sufficiently removed complementary charge distribution in order to reduce binding. In this case there was no correlation between the score and binding affinity, but this is not unexpected. The algorithm

is information driven, using the alternating charge distribution and spacing to find complementary paths. The scrambled sequences should have lost these features, hence there could be other paths that are more complementary to their new charge distribution. Intriguingly, a few randomized sequences were able to attain scores similar to or more favourable than the wildtype KAHRP 5' repeat. This was predominantly due to the path offset allowing charge clashes to be avoided, effectively making the net charge of KAHRP more positive or net charge of spectrin more negative. As an interaction in solution would not be able to avoid charge clashes in this way, it is unlikely the binding affinity of this would be stronger. Although none of the "more favourable" sequences were tested, they do suggest that an interesting future direction of this work would be to use this tool to design an optimized KAHRP sequence that binds even stronger. Given the KAHRP repeats are not perfect repeats, the parasite itself may have optimised the sequence for function over thousands of years of evolution. This idea is supported by the fact the parallel orientation of KAHRP results in less favourable scores for the physiological constructs, K2-4 and β10-14.

There are many improvements that could be made to this approach, such as introducing features like hydrogen bonds. The parameters used, such as spacing of disordered residues relative to the surface, the two-level score function, charge cluster definition, and the cut-off distances used between clusters, could all be further optimised but the data presented here is a solid foundation on a real life example. Ideally, finer development should be done on a more modular system, with perfect sequence repeats, that is amenable to direct investigation by structural biology methods such as crystallography or NMR, such that results are not biased by sequence variation and the predicted solution can be compared to an experimental structure.

5.5.2 MD Simulations

The MD simulations of the docked complex showed how the algorithm was able to locate a binding site on β spectrin where KAHRP could stably bind, but is still able to be dynamic. Simulation were performed in the presence and absence of restraints on spectrin to try to understand the dynamics in solution, but it hard to know which system is most equivalent to how the interaction would occur *in vivo*. In erythrocytes, spectrin in tethered to the membrane in many locations, and carries a net negative charge, hence should repel the similarly charged membrane. This could result in some degree of local restriction, but *in vitro* studies on isolated spectrin suggest it is very dynamic. Structural analysis of spectrin show each repeat is held together by a hydrophobic face on each helix packing together, and ionic interactions between helices and loops at the end of each repeat [91]. Spectrin is able to bend due to ionic bonds shuffling, sliding of hydrophobic faces and conformational changes in the loop regions [209]. As multiple intra-molecular interactions are ionic, the dynamics are also influenced by ionic strength, with the hydrodynamic dimension changing with ionic strength, despite no change in secondary structure [234]. Extraction of spectrin in low ionic strength buffer results in spectrin having an extended conformation [206], but more native studies suggest a condensed structure *in vivo* [95], suggesting low ionic conditions remove shielding effect of ions of the buffer to strengthen intramolecular ionic interactions. Coarse grain simulations also suggests the spectrin structure can be dynamic, but the *in vivo* spectrin spacing corresponds to an over compressed arrangement of spectrin, such that extension results in no energetic penalties [235]. This compaction could also result in local restriction of spectrin dynamics.

Experimental evidence suggest KAHRP binding leads to increased cytoskeletal rigidity [63] while KAHRP truncations that partly remove the protein 5´-repeat reduce cell rigidity and adhesion to the CD36 receptor under shear flow conditions [38]. Further, electron microscopy studies show that the average spectrin tetramer length increases by up to 50%

during parasite growth and that there is a clearance of integral membrane proteins around the edge of knobs, likely as a result of parasite proteins interacting with the cytoskeleton [86, 136].

This chapter tested how KAHRP influences spectrin flexibility in MD simulation. KAHRP binding to spectrin increased the dynamics of spectrin, hence it is unlikely the KAHRP interaction with spectrin results in cytoskeleton rigidity, but rather that stiffening is due to increased connectively of membrane and cytoskeletal as proposed by Zhang *et al.* [89]. The increase in spectrin dynamics on KAHRP binding could explain the extension of the cytoskeleton and the coupled clearance of membrane proteins, as more dynamic spectrin motions could decompress it from its condensed form.

More work needs to be done to further investigate the effects of crowding, of other cytoskeleton proteins and the membrane, and of the concentration of ions. All binding assays and simulations for this chapter were done in relatively low ionic conditions of 50 mM NaCl. Hence to draw significant conclusions, these should be repeated with physiologically relevant salt concentrations. Simulations including the K1D region and ankyrin could also help explore how the larger complex behaves, but the KAHRP-ankyrin interaction would have to be structurally characterised in more detail for such simulations to be meaningful.

The MD simulations also highlight another limitation to the docking algorithm. The algorithm searches for paths between defined charged residues assuming spectrin is completely rigid. The charge cluster spacing and the extent of charge surface exposure is likely to change considerably as spectrin bends, hence binding sites and charge clashes are likely to appear and disappear. Ideally docking could be performed on an ensemble of spectrin structures after MD simulation, with iterative rounds of MD to filter out the most favourable conformations.

5.5.3 Model of KAHRP Binding the Cytoskeleton

Biophysical analysis strongly suggests that KAHRP K1D and 5´-repeat regions associate with host ankyrin and spectrin, respectively, which is in agreement with previous work [62, 63]. Our results, however, differ from these earlier studies in the localization of the KAHRP-binding domains within ankyrin and spectrin. Pei *et al.* suggested that the interaction occurred between the residues 370-441 of the KAHRP K2 region and repeat 4 of α spectrin [63]. This interaction was not observed on the spot FP (Fig. 5.1D) or HSQC experiments (Fig. 5.4D). Both these methods should be able to observe the interaction, given the K_d measured by Pei *et al.* was ~2 μM. Weng *et al.* suggested an interaction between K1D and the D3 ankyrin repeat domain of ankyrin [62]. Again this possibility was tested with both FP and NMR, resulting in no detectable binding despite reported K_d in this case was ~5 nM (Fig. 5.4B, C). The likely explanation for these previous results is the methods used to assess binding were susceptible to aggregation. Both KAHRP and ankyrin constructs are challenging to work with and are prone to aggregation. Our binding studies avoided these problems by using both NMR and FP. The quality of the HSQC spectra of K2 and K1D indicates it is in solution rather than as large aggregates. FP was able to avoid aggregation problems with

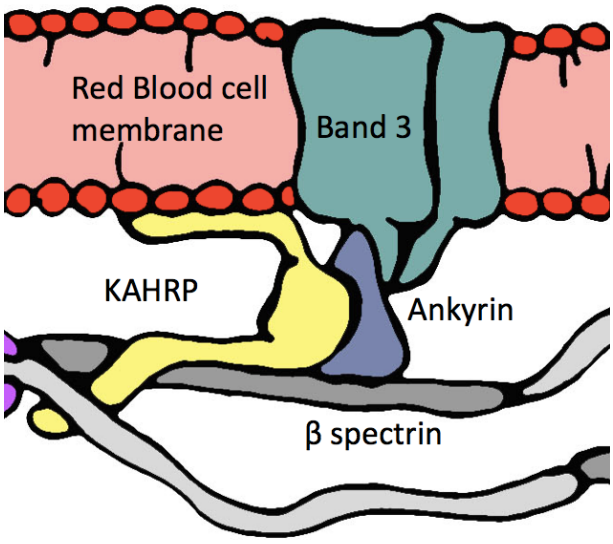


Fig. 5.15: KAHRP cytoskeleton interaction model. Previous studies found the N-terminal histidine rich region was important for membrane interaction. The work of this chapter confirms an interaction between K1D of KAHRP and the spectrin binding domain of ankyrin and K2 binding along β spectrin. The result is a network of interaction, which could afford KAHRP both specificity and affinity in binding.

KAHRP, by ensuring it was kept at low concentration.

I observed KAHRP binding at a cytoskeletal region forming or very near the spectrin–ankyrin complex, that is to β10-14 and ankyrin ZU5A. As ZU5A binds β spectrin at domains 14-15, this suggests that an elegant ternary complex could form between KAHRP, spectrin and ankyrin (Fig. 5.15). Such a ternary complex would strengthen the overall KAHRP association with the cytoskeleton and, by virtue of utilizing a large and promiscuous interaction surface on spectrin β10-14, it would allow the KAHRP–cytoskeletal binding to resist spontaneous mutations of host components. The KAHRP-spectrin docked model suggested in this chapter would not occlude such a complex. Ankyrin binds spectrin via a large negatively charged patch on β repeat 14. This charge patch was included in the path finding algorithm, but was not predicted to be a component of the lowest scoring path. Furthermore, addition of the K1D region should not prevent ankyrin binding spectrin, and this region of KAHRP had roughly neutral charge, hence is unlikely to bind the positively charged face of ankyrin that interacts with spectrin. Ideally, the formation of this tertiary complex should be examined experimentally. As the structure of ankyrin ZU5A with β13-15 has be crystallised, it may be possible to crystallise this in the presence of some region of the KAHRP K1D and 5'-repeat.

Chapter 6

Discussion

"I have to stay functional until my mission is complete"

The Terminator, Judgement Day

6.1 Disordered Proteins

Three interactions were examined in this thesis; *Pf*EMP1 with PHIST domains, *Pf*EMP1 with spectrin, and KAHRP with spectrin. In each case protein disorder was involved. Common themes arise in these interactions, suggesting how protein disorder is used by *Plasmodium* as a tool to modulate interactions.

6.1.1 Protective to Mutation

After characterising the interaction interfaces between *Pf*EMP1 ATS with PHIST domains, *Pf*EMP1 ATS with spectrin, and KAHRP with spectrin, attempts were made to interfere with specific contacts to reduce affinity. This was done with alanine scanning mutagenesis for the ATS/PHIST interactions, with substitute mutagenesis with ATS/spectrin, and sequence scrambling for KAHRP/spectrin. In some cases binding affinity was significantly

affected, but in most cases, binding was still detectable. Other IDPs have been found to be resistant to mutation. One interesting example is Ewings sarcoma fusion proteins (EFP), which is a potent activator of transcription. The Ewings sarcoma activation domain (EAD) is an IDP containing multiple hexapeptide repeats of SYGQQS. Mutational analysis found that while the tyrosines were important, they could be replaced with phenylalanine and that there was little requirement for specific peptide sequences [236]. Another example is yeast regulatory proteins GCN4 and GAL4, which has a disordered region of about 35-40 residues, enriched in acidic residues. Activity seemed to be predominately modulated by the amount of acidic residues, rather than to specific sequence features [237]. Further truncation analysis of GCN4 suggested the acidic region acted in an additive manner, with longer regions providing better activation, even if they have the same or fewer acidic residues than a shorter region [238]. This is analogous of the binding interaction between spectrin and KAHRP, where longer regions of the basic repeat sequence bind more strongly and the charge profile is important in binding.

The major advantage of *P. falciparum* using IDPs is that any mutations, insertions, or deletions that could occur from recombination events (that are common for the subtelomerically located *var* genes and PHIST genes) are less likely to have a detrimental effect. This is supported by the binding studies done in this thesis. Different ATS variants demonstrate how each variant was able to retain affinity to α17 spectrin. To a lesser extent, this was also true with the PHIST interactions with ATS.

6.1.2 Energetic Considerations

The energetics of binding of intrinsically disordered proteins is still under much debate. Under a classical understanding of protein-protein interactions, in order to bind, a disordered, fully flexible peptide must sacrifice some of its conformational flexibility in order to be immobilised in a complex. This model could be energetically favourable, if the bound form

adopts an enthalpically favourable fold, hence the interactions enable the IDP to reach an energy minimum not possible in isolation. This is referred to as "induced fit". Although this in isolation should be entropically unfavourable, many disordered binding reactions are entropically favourable [239]. A possible explanation for how binding overcomes this theoretical entropy cost is that IDPs retain some flexibility on binding, hence form a "fuzzy complex". In this model the entropic cost is reduced by allowing the IDP to retain flexibility [240]. The other major model suggests that IDPs are not fully flexible to begin with, rather they exist in a selection of interconverting, partially folded states, some of which are in a conformation that allows them to bind. This model is termed "conformational selection". In reality binding is probably a combination of all these mechanisms [241]. In cases when entropy is unfavourable the large binding interfaces seen in an IDP could enable the enthalpy to drive binding [242].

A rigorous energetic analysis was not done for the interactions in thesis, but there is evidence to support these binding models. The interaction between PHIST proteins and the ATS-C resulted in shift perturbations suggesting the formation of an extended structure, which was predicted to form a β sheet with CS-Rosetta, hence supports an "induced fit" style model. But PRE experiments resulted in dispersed effects not consistent with one bound conformation and the interface was dynamic in MD simulations, which is more consistent with a "fuzzy complex". This was also the case for the ATS-Core and KAHRP interacting with spectrin, as the interface was dynamic in MD. In addition, the docking approach of KAHRP and spectrin resulted in many alternative paths being predicted, some with similar energy, which could correspond to multiple possible KAHRP poses and form a "fuzzy complex". The KAHRP interaction also highlighted something not often considered in disordered protein interaction; that the dynamics of both binding partners are important. While an IDP is by definition highly dynamic, the dynamics of both proteins must be considered. Spectrin is a highly flexible molecule and simulations suggest that KAHRP fur-

ther modulates its dynamic motions. This increase of dynamics in spectrin could provide a favourable entropy that helps to drive binding, but could also help regulate binding, if some conformations of spectrin have higher affinity for KAHRP than others.

6.1.3 Multivalent Binding

Another common feature in disordered proteins is that they are able to interact promiscuously to modulate function [243]. An interactome study across 4 eukaryotic species suggested that hub proteins, defined as those that interacted with more than 10 different partners, were significantly more disordered [244]. Binding multiple partners helps IDPs to bind both strongly and specifically. A disordered protein can use multiple low-affinity binding sites to result in high-avidity interactions [245]. Although this often occurs with repetitive binding sites, binding many copies of the same protein, it can also occur with different proteins on different regions of the IDP and is termed "allosteric coupling" [246, 247]. This is likely to occur for the interactions investigated in this thesis. KAHRP interacts with two cytoskeletal proteins, spectrin and ankyrin, and is also proposed to interact with the membrane, which ankyrin is linked to via an interaction with band 3. Similarly in the ATS; the ATS-Core binds directly to the cytoskeleton, to spectrin, while the ATS-C is indirectly linked to the cytoskeleton via PFE1605w interacting with band 3. Hence, both the ATS and PFE1605w are localised and linked to the cytoskeleton via two distinct interactions. This possibly enables these proteins to bind more strongly despite the individual interactions being in the μM range. One variant of the ATS, PFF0845c, is not able to interact with PHIST proteins as it lacks the last ~80 residues of the ATS-C. However, this variant bound α 17 with ~10 fold higher affinity than ATS variant PF08_0141 (K_d 24$\pm$2 and 200$\pm$15 μM, respectively). Hence, the PFF0845c variant likely compensates for not being able to bind PHIST proteins by having stronger spectrin affinity.

6.1.4 Redundancy in Binding

An alternative to the model where KAHRP and *Pf*EMP1 binding is due to multivalent interactions is that additional interactions may result in redundancy enabling knob assembly in variable conditions. For example, ankyrin is not present on every β spectrin chain [248], hence KAHRP binding specifically to spectrin may provide an alternative when ankyrin is not present. Later stages of parasite development result in depletion of cytoskeletal components [20, 138], hence the multiple interactions of the ATS and KAHRP, may also enable it to remain linked to the cytoskeleton as it starts to break down. Additionally, *Pf*EMP1 being linked to band 3 via PFE1605w might enable a larger number of *Pf*EMP1 molecules to be localised to knobs, where there is a population of band 3 that is not interacting with the cytoskeleton, but has limited diffusion due to the cytoskeleton meshwork [104, 105].

6.1.5 Larger Assemblies

Sequences of low complexity, that is, repetitive sequences, or those composed of only a few amino acid types, are common in IDPs and are often involved in promiscuous interactions [249]. Furthermore, these tend to result in the formation of aggregate species [250]. Although aggregates are typically thought of as non-functional *in vivo*, condensed aggregate-like protein states can have essential roles. An example of this are RNA granules, which include several proteins with regions of low complexity and are capable of dynamically regulating gene expression [251]. *In vitro* experiments found that proteins that make up RNA granules could form hydrogels that have controlled assembly triggered by phosphorylation [252, 253]. RNA granules can be visualised *in vivo* as organised and other assemblies have been shown to give rise to organised structures *in vitro*, such as amyloids [254]. Phase separation of proteins from the cytosol in this manner is referred to as an "assemblage" and are thought to enable the formation of lipid free organelles, but our understanding is limited by the difficulty of *in vivo* experiments on such assemblages [255].

Although the composition of assemblages is variable, electrostatics have been shown to be able drive the formation of phase separation [256]. Knobs have many features common to assemblages; KAHRP, PFI1780w and PFE1605w have regions of low complexity and are highly charged, and *Pf*EMP1 ATS is also highly charged. KAHRP has previously been shown to form large complexes [60, 61], and the PHIST proteins were both prone to aggregation especially in low salt conditions or in the presence of the ATS. Knobs containing disordered assemblies could explain the electron dense layer observed underneath knobs [59], and this layer might provide a microenvironment allowing the spiral structure to assemble underneath. In this model, the interactions that KAHRP make with spectrin and ankyrin could define the edges, hence the size of the knob structures. Consistent with this idea is data showing that KAHRP is present around the edge of knobs [136] and removal of the spectrin binding region of KAHRP reduces the size and changes the general morphology of knobs [38].

6.1.6 Additional Roles for Disordered Regions in Binding

Unexpected truncation effects were observed for the interactions examined in this thesis. Truncation of the ATS from FL to ATS-N-Core resulted in a reduction in affinity to α 17-C spectrin (K_d 59$\pm$6 and 220$\pm$30 μM, respectively) despite the middle and C terminal regions of the ATS showing little spectrin affinity by NMR. Truncations of the N-terminus of the ATS also resulted in slow loss of binding effects, even though it is unlikely this whole region interacts. Furthermore, the ATS and spectrin both have net negative charge, hence should repel each other via long range electrostatics. An opposing effect was observed in the case of KAHRP and spectrin; the affinity of the full K2 region to β10-14 was less than the 5' repeat region alone (K_d 37$\pm$7 and 18$\pm$2 μM). Hence, it seems binding can be further modulated by neighbouring regions of disordered proteins, despite them not having specific roles in binding. A possible explanation for this are that these additional regions effect how

the protein behaves in solution. The electrostatics of a protein have been shown to have effects of this nature, where regions not directly involved in binding help orientate proteins such that it promotes the formation of a complex [257].

6.2 The Knob Model

6.2.1 Cytoskeletal Interactions

Having mapped the interactions of these three knob proteins, a new model that incorporates all of the current data of knob function can be formulated. Prior to infection the cytoskeleton is a mesh of relaxed spectrin tethered to the membrane at junctional complexes and ankyrin containing complexes (Fig. 6.1A). *Pf*EMP1 binds to α spectrin repeat 17, via the ATS-Core, resulting in ~6 *Pf*EMP1 molecules clustering around the junctional complex as the C-terminal region of spectrin meet at the junctional complex. PFE1605w binds the ATS-C of *Pf*EMP1 and band 3, which is also present at the junctional complex. KAHRP binds to the membrane, ankyrin and spectrin, to be recruited to the ankyrin complex (Fig. 6.1B). KAHRP and possible other proteins can then form the electron dense layer observed in EM, and recruit the spiral forming protein/proteins (Fig. 6.1C).

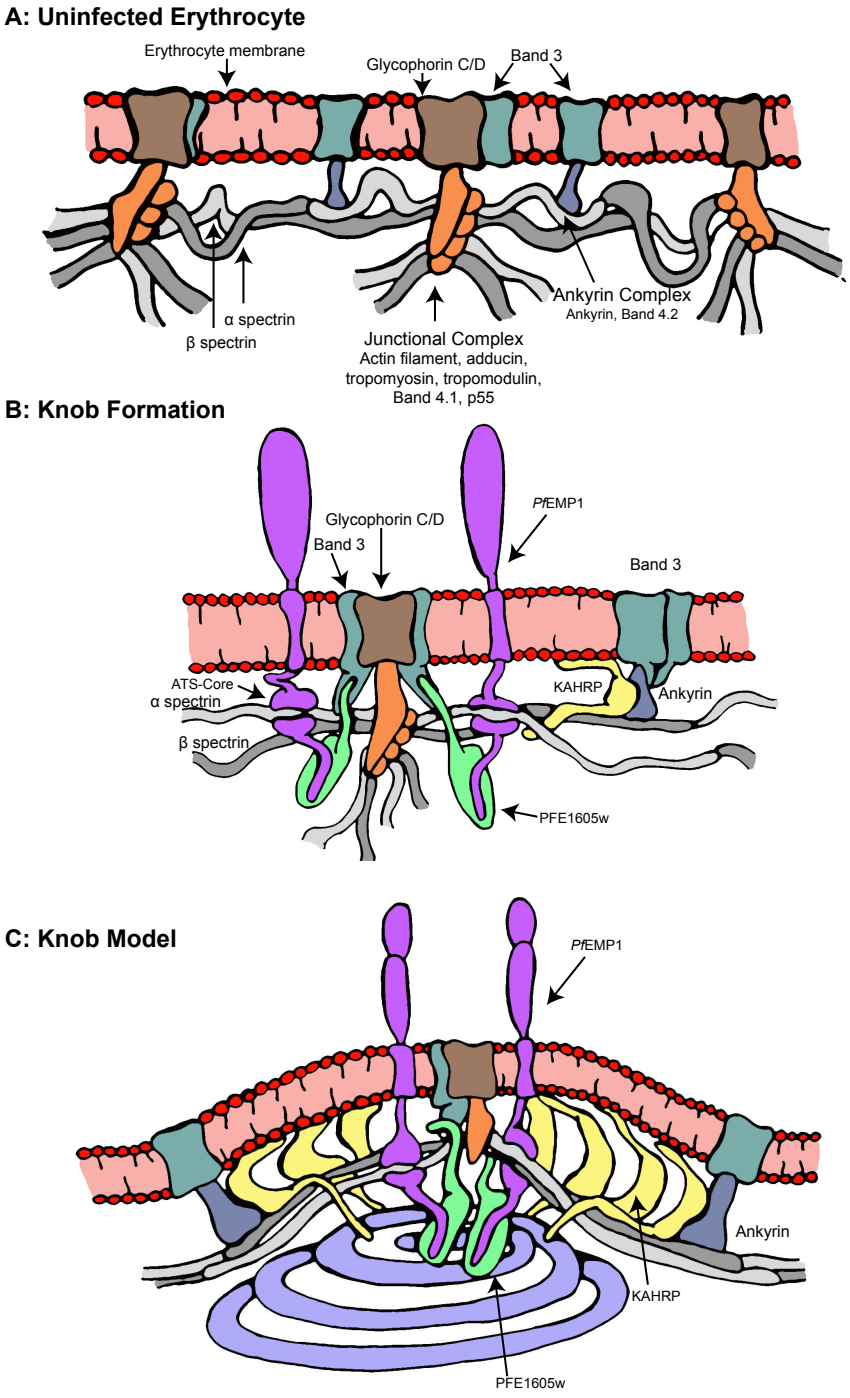


Fig. 6.1: Model for knob structure and assembly. **A**: The uninfected erythrocyte cytoskeleton. **B**: *Pf*EMP1 is able to cluster around junctional complexes due to specific interactions with α17 spectrin and interactions with PFE1605w, which is linked to band 3. KAHRP interacts with the membrane, ankyrin and spectrin. **C**: Accumulation of *Pf*EMP1 and KAHRP could result in assembly of a larger complex of disordered protein, an assemblage, enabling the recruitment of other factors such as the spiral forming protein, resulting in a fully functional knob.

6.2.2 Assembly and Timing

Expression profile data and modelling suggest that *Pf*EMP1 transcription peaks at 10 hours post erythrocyte infection, while KAHRP, PFI1780w and PFE1605w peak around 16 hours post infection [258]. Knob structures form 16-18 hours post infection [141]. Due to its size, *Pf*EMP1 export is delayed, such that it is processed and exported from the Mauer's Cleft with along with PFE1605w and KAHRP [39, 54]. While it is possible that some level of pre-assembly occurs on the Maurer's Cleft, this can not be necessary part of knob formation, as defects that prevent *Pf*EMP1 trafficking do not prevent KAHRP trafficking and knob formation [140], and knockout of KAHRP does not prevent *Pf*EMP1 clustering [36]. The interaction between KAHRP and spectrin/ankyrin in this thesis provides a mechanism of how knobs can be localised in the absence of *Pf*EMP1. And the interaction characterised between the ATS-Core of *Pf*EMP1 and spectrin provide a mechanism of how this clustering could be achieved in the absence of KAHRP. In addition, *Pf*EMP1 interacted with other regions of spectrin, although with lower affinity. These low affinity interactions may be enough for the ATS to associate with the spectrin filaments *in vivo*, and change *Pf*EMP1 diffusion in the membrane from 2 dimensional planar diffusion to linear diffusion, enabling an increase of the local concentration *Pf*EMP1 around the junctional complex.

Previous studies have suggested an interaction between *Pf*EMP1 and KAHRP [60, 64], but this interaction was unable to be reproduced in solution [65]. Considering previous studies determined K_d using methods that immobilised either KAHRP or the ATS, it could be possible that their binding requires clustering of proteins first. Consistent with this idea Waller *et al.* [64] proposed multiple ATS binding sites on KAHRP and KAHRP self associates [60]. Hence, the electron dense layer of KAHRP might form as a result of KAHRP and *Pf*EMP1 accumulation in a small area around the junctional complex.

6.2.3 PHIST Proteins

PFE1605w knockdowns result in reduction of adhesion to CD36 by 64% and to ICAM-1 by 30%, but had no effect on adhesion to CSA [199]. This implies PFE1605w has a role in some, but not all adherent knobs, suggesting it may not interact with all *Pf*EMP1 variants. Oberli *et al.* performed parasite panning to isolate parasites with *Pf*EMP1 variants that only bound to CD36, ICAM-1 or CSA, enabling identification of which *Pf*EMP1 variant is being expressed. Binding assays testing the ATS-C of each *Pf*EMP1 variant suggested that the variant that binds CSA bound the PFE1605w PHIST domain with a lower affinity than the *Pf*EMP1 variants that bound ICAM-1, possibly suggesting that the effects of PFE1605w knockdown are coupled with its binding affinity to *Pf*EMP1. However, the *Pf*EMP1 variant that bound CD36, which was most effected by PFE1605w knockdown, had limited affinity to the PFE1605w PHIST domain. It could be that these proteins interact with some other regions, as only the PHIST domain and the ATS-C was tested, but it could also be an indication that more complicated regulation is occurring or the effects on adhesion are due to some other function of PFE1605w rather than its binding to *Pf*EMP1. With *P. falciparum* having such a repertoire of PHIST proteins exported, it is possible that there is functional redundancy, such that some *Pf*EMP1 variants that bind PFE1605w with lower affinity bind an alternative PHIST protein with higher affinity.

Epigentic regulation of gene families is commonly used in *Plasmodium* [50]. *P. falciparum* uses epigetic control to ensure only one *Pf*EMP1 variant is expressed, and is able to switch variant expression to avoid detection by the immune system. Epigetic control of gene families is more wide spend than just *var* genes, with many gene families being identified as having histone modifications which are changed in an orchestrated manner, including *PHIST* family genes [259]. Hence, it is possible that specific *var*, *PHIST* and other essential knob genes have correlated expression. A study by Claessens *et al.* helps to support this theory, as they analysed the transcriptome of parasites that cytoadhered to brain endothelial cell

lines and found PHISTa family member, PF14_0752, is upregulated [260]. This could explain the different knob morphology that is observed with different *Pf*EMP1 variants [261]. Furthermore, the expression of one PHISTb family member, PFD1170c, has been correlated with adhesion, with deletion resulting in no knobs as a result of incorrect trafficking of KAHRP, and higher expression resulting in higher knob density [56, 262, 263]. No interaction between PFD1170c and the ATS-C has been found [39], hence suggest PHIST domains interact with a wide range of other proteins, and that the four helix bundle structure might act as a scaffold to allow diverse interactions. Other PHIST family members have been implicated in adhesion, such as PHISTb protein PFB0085c [12, 24] and other DnaJ domain containing proteins, such as PF10_0381 [56], suggesting many more PHIST proteins and other factors are involved in knob formation.

6.2.4 Spiral Scaffold

Tomography of knob structures from *P. falciparum* 3D7 ghosts shows a spiral scaffold underlying the knob protrusion with an electron dense layer between the spiral and the membrane [136] (Fig. 1.6C-E). The electron dense layer has long been attributed to KAHRP, as KAHRP knockouts do not have this electron dense layer [36], and tomography using anti-KAHRP anti-body localised KAHRP around the periphery of knobs [136]. While this is consistent with KAHRP making up some component of the electron dense layer, it does not exclude it from containing other factors. This spiral structure was compared to the ESCRT-III (endosomal sorting complex required for transport III) complex that results in membrane budding, but this complex relies on helices embedding into the membrane to bring about its spiral structure and membrane curvature [264, 265]. This is not the case for knob structures, the spiral forming protein must work on a base layer of the electron dense protein. Given that a knockout of KAHRP prevents knob formation, it is likely that KAHRP acts as an adapter protein, either recruiting the spiral protein directly, or translates its effects to the cy-

toskeleton and membrane. This is supported by a study from Subramani *et al.* that panned for specific *Pf*EMP1 using antibodies against specific *Pf*EMP1 ectodomains and analysed knob morphology [261]. One *Pf*EMP1 variant studied, IT4var60 resulted in no knobs, despite the fact KAHRP was still present, while another, IT4VAR32b, resulted in higher than normal knob density. This result suggests that KAHRP alone is not solely responsible for knobs, rather that it is some other protein or proteins that is recruited to knobs, possibly via an interaction with KAHRP, that results in protrusions.

6.2.5 Regulation

An uninfected erythrocyte cytoskeleton can be regulated by phosphorylation and Ca^{2+} concentration [111–113]. More generally phosphorylation can regulate adhesion, as in epithelial cells, tyrosine phosphorylation of cadherins and associated proteins has major effects on the stability of adhesion junctions [266]. It is possible these regulatory mechanisms are high-jacked or perturbed in *Plasmodium* infection. The concentration of Ca^{2+} has been found to be elevated in later stages of infection [267]. Increased Ca^{2+} leads to destabilisation of the cytoskeleton via reduced band 4.1 connectivity [111]. Hence, interactions between KAHRP-spectrin-ankyrin and *Pf*EMP1-PFE1605w-spectrin-band 3 could help stabilise these regions of the cytoskeleton, while other regions of cytoskeleton are stabilised by other factors such as RESA, or possibly allowed to break down to promote egress.

The phosphorylation patterns are also changed in *Plasmodium* infection [116]. Phosphorylation has been reported in MESA and band 4.1, potentially increasing their affinity [268–270], again to potentially stabilise the cytoskeleton. In a study of phosphorylation in *P. falciparum* infection, a large number of cytoskeletal proteins were found to be phosphorylated, including ankyrin, spectrin, band 4.1, band 3, adducin and many more. While many of these proteins are also phosphorylated in uninfected erythrocytes, the phosphorylation pattern was often different, suggesting these are as a response to parasite infection. In

part, this could be due to cellular stress associated with infection, where malarial infection leads to oxidative stress [271]. Many parasite proteins were also found to be phosphorylated in this study, including KAHRP, PFE1605w, other PHIST proteins and some *Pf*EMP1 variants. Phosphorylation of PFE1605w has been found at residue 310, and 317 in three additional studies [272–274]. As residues 310/317 are near the end of the PHIST domain and at the start of the disordered C-terminal region, this phosphorylation could modulate binding to *Pf*EMP1 and band 3. KAHRP phosphopeptides were found extensively through the sequence 5' repeat, hence could potentially be a powerful modulator of binding to spectrin, as this thesis demonstrated that the charge distribution of KAHRP had effects on binding. One study aimed to investigate phosphorylation in knob interactions [275], but their study focused on *in vitro* phosphoylation of *Pf*EMP1. Although *Pf*EMP1 has several predicted phosphorylation sites [35], there is little *in vivo* evidence of these due to the challenges associated with variant switching. But, as treatment with phosphoylation inhibitors reduced adhesion under flow conditions [275], further investigation into the regulation phosphorylation plays in knob structures is warranted.

6.2.6 Blood Disorders

There are many blood disorders that result in a protective effect against malaria [276]. While these are likely to effect all stages of parasite development, many occur in cytoskeletal proteins. In particular, mutations in band 3 result in decrease of cerebral malaria in children [277, 278]. This mutation has actually been shown to increase adhesion in some cases, hence the protective effect is not likely to come as a result of modulation in cytoadhesion [279]. Mutation in hemoglobin (Hb) (β6 Glu3Lys) also protects against severe malarial infection and has been shown to prevent the formation of knobs, potentially reducing cytoadhesion, but this protective effect could also be caused by these erythrocytes having a slower parasite growth rate, reduced remodelling and possibly being resistant to parasite induced lysis [20,

280, 281]. Hence, while it may be tempting to speculate that some protective effects are due to cytoskeleton and knob defects, this is confounded by many other factors.

6.3 Outlook and Future Work

Further investigation into PHIST proteins, gene regulation of knob protein expression, the effects of post translational modification and the role of additional proteins, such as the spiral protein, has already been mentioned. In addition to this, there are other factors and approaches to be considered.

6.3.1 Strain Specific Factors

P. falciparum has evolved many mechanisms to increase its chances of survival. There are different strains, each with different *var* genes, each being able to recombine and shuffle gene expression. This is demonstrated well by the Subramani *et al.* study which demonstrates different *Pf*EMP1 expression yields different knob morphology and receptor binding properties [261]. But this also highlights a limitation in complete understanding of knob structure. Knob studies use different isolates and most likely express a different *var* gene. Due to the organisation and regulation *P. falciparum* gene expression, it is likely that the *PfEMP1* that is expressed and the composition of the knobs is variable. Further studies on knob structure should use the same strain and try prevent swapping of *Pf*EMP1 variants. Although this can be challenging, panning for parasites based on affinity to receptors or with specific anti-bodies has been shown to control this to some extent [199, 261].

6.3.2 Applications

One of the motivations of this work was to provide information that might be able to be used to prevent cytoadhesion. Although interactions involving intrinsically disordered proteins

have long been considered poor drug targets, this view is beginning to be revised [282, 283], hence interference of the interfaces described in this thesis may yet be possible. The interaction between the ATS-core and α spectrin was able to be interfered with by mutation of conserved ATS residues R286 and H287, hence these may define a site that a small molecule could bind to.

6.3.3 Reconstitution

Current EM studies of knob structure have been performed on parasitised erythrocytes [136], but as more components of knobs are determined, reconstitution of knob structures should be possible. Some attempts were made in this thesis towards this, by performing negative stain EM using purified spectrin and ATS-N-Core conjugated to gold nano-particles. This could be extended to full erythrocyte cytoskeleton, where erythrocyte cytoskeletons can be easily purified to produce "ghosts" visualisable with negative stain EM [248]. As membrane interactions are likely to be important, super-resolution fluorescent microscopy approaches like Photo-Activated Localization Microscopy (PALM) and Stochastic Optical Reconstruction Microscopy (STORM) that are able to reach resolutions up to 10 nm, could be used with resealed erythrocytes or inside-out vesicles derived from erythrocytes [284]. Full knob reconstitution would require determination of the spiral protein and all other essential factors, however, this approach would overcome the limitations due to variations in strains and gene expression and be able to confirm the knob model proposed in this thesis.

Conclusion

"...pation!"

Dr Frank N. Futer,

The Rocky Horror Picture Show

This thesis localised and characterised the interactions of three classes of knob proteins from *P. falciparum* with each other, and with host erythrocyte cytoskeletal components:

- *Pf*EMP1 ATS-C with the PHIST domains from PFE1605w and PFI1780w.
- The C-terminal region of PFE1605w with the cytosolic domain of band 3.
- *Pf*EMP1 ATS-Core with α spectrin repeat domain 17.
- The 5' repeat region of KAHRP with β spectrin repeats 10-14.
- The K1D region of KAHRP with the Zu5A domain of ankyrin.

These interactions, along with previously published experimental data suggesting KAHRP self-associates, and a spiral forms underneath the knob, has allowed me to derive a new model of knob architecture (Fig. 7.1). In addition to developing a new model for knob architecture, structural models were derived of three complexes; *Pf*EMP1 ATS-C with the PFI1780w PHIST domain, *Pf*EMP1 ATS-Core with α17 spectrin, and KAHRP 5' repeat region with β10-14 spectrin (Fig. 7.2).

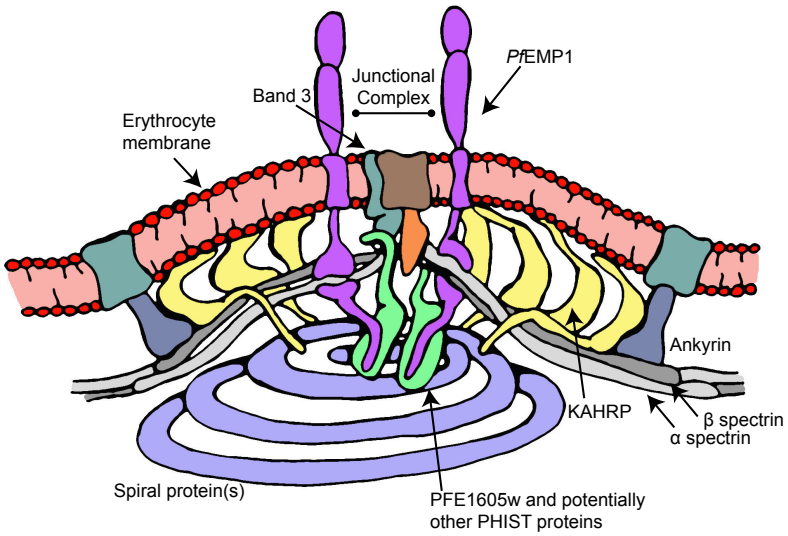


Fig. 7.1: The new knob model. The interactions investigated in this thesis put into context of the knob structure.

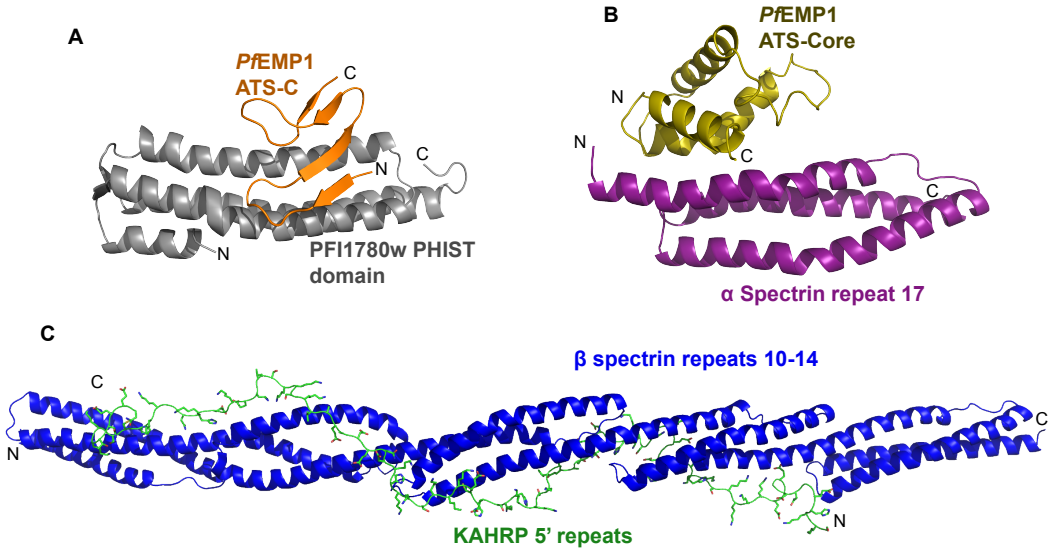


Fig. 7.2: Structural models derived. **A**: Complex between the Rosetta model of PF08_0141 *Pf*EMP1 ATS-C 306-346 and the monomeric model of the PHIST domain of PFI1780w. **B**: Complex between the homology model the PFF0845c ATS-Core and crystal structure of α 17. **C**: The complex between KAHRP 5' repeat residues 362-458 and β10-14 spectrin.

References

[1] C. J. Murray, L. C. Rosenfeld, S. S. Lim, K. G. Andrews, K. J. Foreman, D. Haring, N. Fullman, M. Naghavi, R. Lozano, and A. D. Lopez, "Global malaria mortality between 1980 and 2010: a systematic analysis," *The Lancet*, vol. 379, no. 9814, pp. 413–431, 2012.

[2] L. H. Miller, D. I. Baruch, K. Marsh, and O. K. Doumbo, "The pathogenic basis of malaria," *Nature*, vol. 415, no. 6872, pp. 673–679, 2002.

[3] T. J. C. Anderson and C. Roper, "The origins and spread of antimalarial drug resistance: Lessons for policy makers," *Acta Tropica*, vol. 94, no. 3, pp. 269–280, 2005.

[4] E. Y. Klein, "Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread.," *International journal of antimicrobial agents*, vol. 41, no. 4, pp. 311–317, 2013.

[5] K. Haldar, S. Kamoun, N. L. Hiller, S. Bhattacharje, and C. van Ooij, "Common infection strategies of pathogenic eukaryotes," *Nature Reviews Microbiology*, vol. 4, no. 12, pp. 922–931, 2006.

[6] A. F. Cowman, D. Berry, and J. Baum, "The cellular and molecular basis for malaria parasite invasion of the human red blood cell.," *The Journal of cell biology*, vol. 198, no. 6, pp. 961–971, 2012.

[7] K. Chotivanich, R. Udomsangpetch, J. A. Simpson, P. Newton, S. Pukrittayakamee, S. Looareesuwan, and N. J. White, "Parasite multiplication potential and the severity of Falciparum malaria.," *The Journal of infectious diseases*, vol. 181, no. 3, pp. 1206–1209, 2000.

[8] G. A. Josling and M. Llinás, "Sexual development in Plasmodium parasites: knowing when it's time to commit," *Nature Reviews Microbiology*, vol. 13, no. 9, pp. 573–587, 2015.

[9] K. Marsh, D. Forster, C. Waruiru, I. Mwangi, M. Winstanley, V. Marsh, C. Newton, P. Winstanley, P. Warn, N. Peshu, G. Pasvol, and R. Snow, "Indicators of Life-Threatening Malaria in African Children," *New England Journal of Medicine*, vol. 332, no. 21, pp. 1399–1404, 1995.

[10] A. G. Craig, M. F. M. Khairul, and P. R. Patil, "Cytoadherence and severe malaria.," *The Malaysian journal of medical sciences : MJMS*, vol. 19, no. 2, pp. 5–18, 2012.

[11] B. C. Elford, G. M. Cowan, and D. J. Ferguson, "Parasite-regulated membrane transport processes and metabolic control in malaria-infected erythrocytes.," *Biochemical Journal*, vol. 308 (Pt 2), no. 2, pp. 361–374, 1995.

[12] T. J. Sargeant, M. Marti, E. Caler, J. M. Carlton, K. Simpson, T. P. Speed, and A. F. Cowman, "Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites.," *Genome biology*, vol. 7, no. 2, p. R12, 2006.

[13] J. A. Boddey, A. N. Hodder, S. Günther, P. R. Gilson, H. Patsiouras, E. A. Kapp, J. A. Pearce, T. F. de Koning-Ward, R. J. Simpson, B. S. Crabb, and A. F. Cowman, "An aspartyl protease directs malaria effector proteins to the host cell," *Nature*, vol. 463, no. 7281, pp. 627–631, 2010.

[14] I. Russo, S. Babbitt, V. Muralidharan, T. Butler, A. Oksman, and D. E. Goldberg, "Plasmepsin V licenses Plasmodium proteins for export into the host erythrocyte," *Nature*, vol. 463, no. 7281, pp. 632–636, 2010.

[15] H. H. Chang, A. M. Falick, P. M. Carlton, J. W. Sedat, J. L. DeRisi, and M. A. Marletta, "N-terminal processing of proteins exported by malaria parasites," *Molecular and biochemical parasitology*, vol. 160, no. 2, pp. 107–115, 2008.

[16] B. Elsworth, K. Matthews, C. Q. Nie, M. Kalanon, S. C. Charnaud, P. R. Sanders, S. A. Chisholm, N. A. Counihan, P. J. Shaw, P. Pino, J.-A. Chan, M. F. Azevedo, S. J. Rogerson, J. G. Beeson, B. S. Crabb, P. R. Gilson, and T. F. de Koning-Ward, "PTEX is an essential nexus for protein export in malaria parasites," *Nature*, vol. 511, no. 7511, pp. 587–591, 2014.

[17] T. F. de Koning-Ward, P. R. Gilson, J. A. Boddey, M. Rug, B. J. Smith, A. T. Papenfuss, P. R. Sanders, R. J. Lundie, A. G. Maier, A. F. Cowman, and B. S. Crabb, "A newly discovered protein export machine in malaria parasites," *Nature*, vol. 459, no. 7249, pp. 945–949, 2009.

[18] J. M. Matz, K. Matuschewski, and T. W. A. Kooij, "Two putative protein export regulators promote Plasmodium blood stage development in vivo," *Molecular and biochemical parasitology*, vol. 191, no. 1, pp. 44–52, 2013.

[19] C. Spycher, M. Rug, N. Klonis, D. J. P. Ferguson, A. F. Cowman, H.-P. Beck, and L. Tilley, "Genesis of and trafficking to the Maurer's clefts of Plasmodium falciparum-infected erythrocytes.," *Molecular and Cellular Biology*, vol. 26, no. 11, pp. 4074–4085, 2006.

[20] M. Cyrklaff, C. P. Sanchez, N. Kilian, C. Bisseye, J. Simpore, F. Frischknecht, and M. Lanzer, "Hemoglobins S and C Interfere with Actin Remodeling in Plasmodium falciparum–Infected Erythrocytes," *Science*, vol. 334, no. 6060, pp. 1283–1286, 2011.

[21] E. Hanssen, R. Sougrat, S. Frankland, S. Deed, N. Klonis, J. Lippincott-Schwartz, and L. Tilley, "Electron tomography of the Maurer's cleft organelles of Plasmodium falciparum-infected erythrocytes reveals novel structural features," *Molecular Microbiology*, vol. 67, no. 4, pp. 703–718, 2007.

[22] E. Pachlatko, S. Rusch, A. Müller, A. Hemphill, L. Tilley, E. Hanssen, and H.-P. Beck, "MAHRP2, an exported protein of Plasmodium falciparum, is an essential component of Maurer's cleft tethers," *Molecular Microbiology*, vol. 77, no. 5, pp. 1136–1152, 2010.

[23] C. Spycher, M. Rug, E. Pachlatko, E. Hanssen, D. Ferguson, A. F. Cowman, L. Tilley, and H.-P. Beck, "The Maurer's cleft protein MAHRP1 is essential for trafficking of PfEMP1 to the surface of Plasmodium falciparum-infected erythrocytes," *Molecular Microbiology*, vol. 68, no. 5, pp. 1300–1314, 2008.

[24] M. W. A. Dixon, S. Kenny, P. J. McMillan, E. Hanssen, K. R. Trenholme, D. L. Gardiner, and L. Tilley, "Genetic ablation of a Maurer's cleft protein prevents assembly of the Plasmodium falciparum virulence complex," *Molecular Microbiology*, vol. 81, no. 4, pp. 982–993, 2011.

[25] A. G. Maier, M. Rug, M. T. O'Neill, J. G. Beeson, M. Marti, J. Reeder, and A. F. Cowman, "Skeleton-binding protein 1 functions at the parasitophorous vacuole membrane to traffic PfEMP1 to the Plasmodium falciparum-infected erythrocyte surface.," *Blood*, vol. 109, no. 3, pp. 1289–1297, 2007.

[26] E. Mundwiler-Pachlatko and H.-P. Beck, "Maurer's clefts, the enigma of Plasmodium falciparum.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 50, pp. 19987–19994, 2013.

[27] T. Blisnick, L. Vincensini, J.-C. Barale, A. Namane, and C. Braun Breton, "LANCL1, an erythrocyte protein recruited to the Maurer's clefts during Plasmodium falciparum development," *Molecular and biochemical parasitology*, vol. 141, no. 1, pp. 39–47, 2005.

[28] C. Spycher, N. Klonis, T. Spielmann, E. Kump, S. Steiger, L. Tilley, and H.-P. Beck, "MAHRP-1, a novel Plasmodium falciparum histidine-rich protein, binds ferriprotoporphyrin IX and localizes to the Maurer's clefts.," *The Journal of biological chemistry*, vol. 278, no. 37, pp. 35373–35383, 2003.

[29] S. Külzer, M. Rug, K. Brinkmann, P. Cannon, A. Cowman, K. Lingelbach, G. L. Blatch, A. G. Maier, and J. M. Przyborski, "Parasite-encoded Hsp40 proteins define novel mobile structures in the cytosol of the P. falciparum-infected erythrocyte," *Cellular microbiology*, vol. 12, no. 10, pp. 1398–1420, 2010.

[30] S. Külzer, S. Charnaud, T. Dagan, J. Riedel, P. Mandal, E. R. Pesce, G. L. Blatch, B. S. Crabb, P. R. Gilson, and J. M. Przyborski, "Plasmodium falciparum-encoded exported hsp70/hsp40 chaperone/co-chaperone complexes within the host erythrocyte," *Cellular microbiology*, vol. 14, no. 11, pp. 1784–1795, 2012.

[31] T. F. de Koning-Ward, M. W. A. Dixon, L. Tilley, and P. R. Gilson, "Plasmodium species: master renovators of their host cells," *Nature Reviews Microbiology*, vol. 14, no. 8, pp. 494–507, 2016.

[32] M. E. Wickham, M. Rug, S. A. Ralph, N. Klonis, G. I. McFadden, L. Tilley, and A. F. Cowman, "Trafficking and assembly of the cytoadherence complex in Plasmodium falciparum-infected human erythrocytes.," *The EMBO journal*, vol. 20, no. 20, pp. 5636–5649, 2001.

[33] D. W. Taylor, M. Parra, G. B. Chapman, M. E. Stearns, J. Rener, M. Aikawa, S. Uni, S. B. Aley, L. J. Panton, and R. J. Howard, "Localization of Plasmodium falciparum histidine-rich protein 1 in the erythrocyte skeleton under knobs.," *Molecular and biochemical parasitology*, vol. 25, no. 2, pp. 165–174, 1987.

[34] S. M. Kraemer and J. D. Smith, "A family affair: var genes, PfEMP1 binding, and malaria disease," *Current Opinion in Microbiology*, vol. 9, no. 4, pp. 374–380, 2006.

[35] T. S. Rask, D. A. Hansen, T. G. Theander, A. Gorm Pedersen, and T. Lavstsen, "Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes–divide and conquer.," *PLoS computational biology*, vol. 6, no. 9, p. e1000933, 2010.

[36] P. Horrocks, R. A. Pinches, S. J. Chakravorty, J. Papakrivos, Z. Christodoulou, S. A. Kyes, B. C. Urban, D. J. P. Ferguson, and C. I. Newbold, "PfEMP1 expression is reduced on the surface of knobless Plasmodium falciparum infected erythrocytes.," *Journal of cell science*, vol. 118, no. Pt 11, pp. 2507–2518, 2005.

[37] B. S. Crabb, B. M. Cooke, J. C. Reeder, R. F. Waller, S. R. Caruana, K. M. Davern, M. E. Wickham, G. V. Brown, R. L. Coppel, and A. F. Cowman, "Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress.," *Cell*, vol. 89, no. 2, pp. 287–296, 1997.

[38] M. Rug, S. W. Prescott, K. M. Fernandez, B. M. Cooke, and A. F. Cowman, "The role of KAHRP domains in knob formation and cytoadherence of P falciparum-infected human erythrocytes.," *Blood*, vol. 108, no. 1, pp. 370–378, 2006.

[39] A. Oberli, L. M. Slater, E. Cutts, F. Brand, E. Mundwiler-Pachlatko, S. Rusch, M. F. G. Masik, M. C. Erat, H.-P. Beck, and I. Vakonakis, "A Plasmodium falciparum PHIST protein binds the virulence factor PfEMP1 and comigrates to knobs on the host cell surface.," *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, vol. 28, no. 10, pp. 4420–4433, 2014.

[40] N. I. Proellocks, S. Herrmann, D. W. Buckingham, E. Hanssen, E. K. Hodges, B. Elsworth, B. J. Morahan, R. L. Coppel, and B. M. Cooke, "A lysine-rich membrane-associated PHISTb protein involved in alteration of the cytoadhesive properties of Plasmodium falciparum-infected red blood cells.," *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, vol. 28, no. 7, pp. 3103–3113, 2014.

[41] M. J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M.-S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. A. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell, "Genome sequence of the human malaria parasite Plasmodium falciparum.," *Nature*, vol. 419, no. 6906, pp. 498–511, 2002.

[42] T. Lavstsen, A. Salanti, A. T. R. Jensen, D. E. Arnot, and T. G. Theander, "Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions.," *Malaria journal*, vol. 2, no. 1, p. 27, 2003.

[43] X. Z. Su, V. M. Heatwole, S. P. Wertheimer, F. Guinet, J. A. Herrfeldt, D. S. Peterson, J. A. Ravetch, and T. E. Wellems, "The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes.," *Cell*, vol. 82, no. 1, pp. 89–100, 1995.

[44] A. R. Berendt, D. L. Simmons, J. Tansey, C. I. Newbold, and K. Marsh, "Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for Plasmodium falciparum.," *Nature*, vol. 341, no. 6237, pp. 57–59, 1989.

[45] A. Barrapn, D. Spillmann, J. Carlson, and M. Wahlgren, "Role of glycans in Plasmodium falciparuminfection," *Biochemical Society transactions*, vol. 27, no. 4, pp. 487–493, 1999.

[46] J. W. Barnwell, A. S. Asch, R. L. Nachman, M. Yamaya, M. Aikawa, and P. Ingravallo, "A human 88-kD membrane glycoprotein (CD36) functions in vitro as a receptor for a cytoadherence ligand on Plasmodium falciparum-infected erythrocytes.," *Journal of Clinical Investigation*, vol. 84, no. 3, pp. 765–772, 1989.

[47] J. G. Beeson, J.-A. Chan, and F. J. Fowkes, "PfEMP1 as a target of human immunity and a vaccine candidate against malaria," *Expert Review of Vaccines*, vol. 12, no. 2, pp. 105–108, 2014.

[48] M. T. Duraisingh, T. S. Voss, A. J. Marty, M. F. Duffy, R. T. Good, J. K. Thompson, L. H. Freitas-Junior, A. Scherf, B. S. Crabb, and A. F. Cowman, "Heterochromatin Silencing and Locus Repositioning Linked to Regulation of Virulence Genes in Plasmodium falciparum," *Cell*, vol. 121, no. 1, pp. 13–24, 2005.

[49] C. J. Tonkin, C. K. Carret, M. T. Duraisingh, T. S. Voss, S. A. Ralph, M. Hommel, M. F. Duffy, L. M. d. Silva, A. Scherf, A. Ivens, T. P. Speed, J. G. Beeson, and A. F. Cowman, "Sir2 Paralogues Cooperate to Regulate Virulence Genes and Antigenic Variation in Plasmodium falciparum," *PLoS Biology*, vol. 7, no. 4, p. e1000084, 2009.

[50] J.-J. Lopez-Rubio, L. Mancio-Silva, and A. Scherf, "Genome-wide Analysis of Heterochromatin Associates Clonally Variant Gene Regulation with Perinuclear Repressive Centers in Malaria Parasites," *Cell host & microbe*, vol. 5, no. 2, pp. 179–190, 2009.

[51] R. Dzikowski, M. Frank, and K. Deitsch, "Mutually Exclusive Expression of Virulence Genes by Malaria Parasites Is Regulated Independently of Antigen Production," *PLoS pathogens*, vol. 2, no. 3, p. e22, 2006.

[52] J. Peters, E. Fowler, M. Gatton, N. Chen, A. Saul, and Q. Cheng, "High diversity and rapid changeover of expressed var genes during the acute phase of Plasmodium falciparum infections in human volunteers.," *Proceedings of the National Academy of Sciences*, vol. 99, no. 16, pp. 10689–10694, 2002.

[53] J. A. Boddey, T. G. Carvalho, A. N. Hodder, T. J. Sargeant, B. E. Sleebs, D. Marapana, S. Lopaticki, T. Nebl, and A. F. Cowman, "Role of Plasmepsin V in Export of Diverse Protein Families from the Plasmodium falciparumExportome," *Traffic*, vol. 14, no. 5, pp. 532–550, 2013.

[54] N. Kriek, L. Tilley, P. Horrocks, R. Pinches, B. C. Elford, D. J. P. Ferguson, K. Lingelbach, and C. I. Newbold, "Characterization of the pathway for transport of the cytoadherence-mediating protein, PfEMP1, to the host cell surface in malaria parasite-infected erythrocytes," *Molecular Microbiology*, vol. 50, no. 4, pp. 1215–1227, 2003.

[55] M. Melcher, R. A. Muhle, P. P. Henrich, S. M. Kraemer, M. Avril, I. Vigan-Womas, O. Mercereau-Puijalon, J. D. Smith, and D. A. Fidock, "Identification of a role for the PfEMP1 semi-conserved head structure in protein trafficking to the surface of Plasmodium falciparum infected red blood cells.," *Cellular microbiology*, vol. 12, no. 10, pp. 1446–1462, 2010.

[56] A. G. Maier, M. Rug, M. T. O'Neill, M. Brown, S. Chakravorty, T. Szestak, J. Chesson, Y. Wu, K. Hughes, R. L. Coppel, C. Newbold, J. G. Beeson, A. Craig, B. S. Crabb, and A. F. Cowman, "Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes.," *Cell*, vol. 134, no. 1, pp. 48–61, 2008.

[57] T. Triglia, H. D. Stahl, P. E. Crewther, D. Scanlon, G. V. Brown, R. F. Anders, and D. J. Kemp, "The complete sequence of the gene for the knob-associated histidine-rich protein from Plasmodium falciparum.," *The EMBO journal*, vol. 6, no. 5, pp. 1413–1419, 1987.

[58] D. W. Taylor, M. Parra, G. B. Chapman, M. E. Stearns, J. Rener, M. Aikawa, S. Uni, S. B. Aley, L. J. Panton, and R. J. Howard, "Localization of Plasmodium falciparum histidine-rich protein 1 in the erythrocyte skeleton under knobs," *Molecular and biochemical parasitology*, vol. 25, no. 2, pp. 165–174, 1987.

[59] J. H. Leech, J. W. Barnwell, M. Aikawa, L. H. Miller, and R. J. Howard, "Plasmodium falciparum malaria: association of knobs on the surface of infected erythrocytes with a histidine-rich protein and the erythrocyte skeleton.," *The Journal of cell biology*, vol. 98, no. 4, pp. 1256–1264, 1984.

[60] S. S. Oh, S. Voigt, D. Fisher, S. J. Yi, P. J. LeRoy, L. H. Derick, S. Liu, and A. H. Chishti, "Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton.," *Molecular and biochemical parasitology*, vol. 108, no. 2, pp. 237–247, 2000.

[61] A. Kilejian, M. A. Rashid, M. Aikawa, T. Aji, and Y.-F. Yang, "Selective association of a fragment of the knob protein with spectrin, actin and the red cell membrane," *Molecular and biochemical parasitology*, vol. 44, no. 2, pp. 175–181, 1991.

[62] H. Weng, X. Guo, J. Papoin, J. Wang, R. Coppel, N. Mohandas, and X. An, "Interaction of Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) with

erythrocyte ankyrin R is required for its attachment to the erythrocyte membrane.," *Biochimica et biophysica acta*, vol. 1838, no. 1 Pt B, pp. 185–192, 2014.

[63] X. Pei, X. An, X. Guo, M. Tarnawski, R. Coppel, and N. Mohandas, "Structural and functional studies of interaction between Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) and erythrocyte spectrin.," *The Journal of biological chemistry*, vol. 280, no. 35, pp. 31166–31171, 2005.

[64] K. L. Waller, B. M. Cooke, W. Nunomura, N. Mohandas, and R. L. Coppel, "Mapping the binding domains involved in the interaction between the Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) and the cytoadherence ligand P. falciparum erythrocyte membrane protein 1 (PfEMP1).," *The Journal of biological chemistry*, vol. 274, no. 34, pp. 23808–23813, 1999.

[65] C. Mayer, L. Slater, M. C. Erat, R. Konrat, and I. Vakonakis, "Structural analysis of the Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) intracellular domain reveals a conserved interaction epitope.," *The Journal of biological chemistry*, vol. 287, no. 10, pp. 7182–7189, 2012.

[66] P. Wittung-Stafshede, J. Guidry, B. E. Horne, and S. J. Landry, "The J-Domain of Hsp40 Couples ATP Hydrolysis to Substrate Capture in Hsp70," *Biochemistry*, vol. 42, no. 17, pp. 4937–4944, 2003.

[67] S. J. Tarr, R. W. Moon, I. Hardege, and A. R. Osborne, "A conserved domain targets exported PHISTb family proteins to the periphery of Plasmodium infected erythrocytes.," *Molecular and biochemical parasitology*, vol. 196, no. 1, pp. 29–40, 2014.

[68] D. J. LaCount, M. Vignali, R. Chettier, A. Phansalkar, R. Bell, J. R. Hesselberth, L. W. Schoenfeld, I. Ota, S. Sahasrabudhe, C. Kurschner, S. Fields, and R. E. Hughes, "A protein interaction network of the malaria parasite Plasmodium falciparum," *Nature*, vol. 438, no. 7064, pp. 103–107, 2005.

[69] P. R. Sanders, "Distinct Protein Classes Including Novel Merozoite Surface Antigens in Raft-like Membranes of Plasmodium falciparum," *The Journal of biological chemistry*, vol. 280, no. 48, pp. 40169–40176, 2005.

[70] M. M. Zilversmit, E. K. Chase, D. S. Chen, P. Awadalla, K. P. Day, and G. McVean, "Hypervariable antigen genes in malaria have ancient roots," *BMC Evolutionary Biology*, vol. 13, no. 1, p. 110, 2013.

[71] M. Aikawa, "Morphological changes in erythrocytes induced by malarial parasites.," *Biology of the cell / under the auspices of the European Cell Biology Organization*, vol. 64, no. 2, pp. 173–181, 1988.

[72] H. Fujioka, P. Millet, Y. Maeno, S. Nakazawa, Y. Ito, R. J. Howard, W. E. Collins, and M. Aikawa, "A Nonhuman Primate Model for Human Cerebral Malaria: Rhesus-Monkeys Experimentally Infected with Plasmodium fragile," *Experimental Parasitology*, vol. 78, no. 4, pp. 371–376, 1994.

[73] C. F. Simpson, J. M. Gaskin, and J. W. Harvey, "Ultrastructure of Erythrocytes Parasitized by Haemobartonella felis," *The Journal of Parasitology*, vol. 64, no. 3, p. 504, 1978.

[74] L. Martin-Jaular, M. Ferrer, M. Calvo, A. Rosanas-Urgell, S. Kalko, S. Graewe, G. Soria, N. Cortadellas, J. Ordi, A. Planas, J. Burns, V. Heussler, and H. A. del Portillo, "Strain-specific spleen remodelling in Plasmodium yoelii infections in Balb/c mice facilitates adherence and spleen macrophage-clearance escape," *Cellular microbiology*, vol. 13, no. 1, pp. 109–122, 2010.

[75] F. G. Baptista, A. Pamplona, A. C. Pena, M. M. Mota, S. Pied, and A. M. Vigário, "Accumulation of Plasmodium berghei-infected red blood cells in the brain is crucial for the development of cerebral malaria in mice.," *Infection and Immunity*, vol. 78, no. 9, pp. 4033–4039, 2010.

[76] C. F. Gilks, D. Walliker, and C. I. Newbold, "Relationships between sequestration, antigenic variation and chronic parasitism in Plasmodium chabaudi chabaudi– a rodent malaria model," *Parasite Immunology*, vol. 12, no. 1, pp. 45–64, 1990.

[77] B. O. Carvalho, S. C. P. Lopes, P. A. Nogueira, P. P. Orlandi, D. Y. Bargieri, Y. C. Blanco, R. Mamoni, J. A. Leite, M. M. Rodrigues, I. S. Soares, T. R. Oliveira, G. Wunderlich, M. V. G. Lacerda, H. A. del Portillo, M. O. G. Araújo, B. Russell, R. Suwanarusk, G. Snounou, L. Rénia, and F. T. M. Costa, "On the cytoadhesion of Plasmodium vivax-infected erythrocytes.," *The Journal of infectious diseases*, vol. 202, no. 4, pp. 638–647, 2010.

[78] Z.-P. Feng, X. Zhang, P. Han, N. Arora, R. F. Anders, and R. S. Norton, "Abundance of intrinsically unstructured proteins in P. falciparum and other apicomplexan parasite proteomes.," *Molecular and biochemical parasitology*, vol. 150, no. 2, pp. 256–267, 2006.

[79] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, "Classification of intrinsically disordered regions and proteins.," *Chemical reviews*, vol. 114, no. 13, pp. 6589–6631, 2014.

[80] P. Tompa, "Intrinsically unstructured proteins evolve by repeat expansion," *BioEssays*, vol. 25, no. 9, pp. 847–855, 2003.

[81] T. Le Gall, P. R. Romero, M. S. Cortese, V. N. Uversky, and A. K. Dunker, "Intrinsic disorder in the Protein Data Bank.," *Journal of biomolecular structure & dynamics*, vol. 24, no. 4, pp. 325–342, 2007.

[82] I. C. Felli and R. Pierattelli, eds., *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*, vol. 870 of *Advances in Experimental Medicine and Biology*. Cham: Springer International Publishing, 2015.

[83] A. J. Guy, V. Irani, C. A. MacRaild, R. F. Anders, R. S. Norton, J. G. Beeson, J. S. Richards, and P. A. Ramsland, "Insights into the Immunological Properties of Intrinsically Disordered Malaria Proteins Using Proteome Scale Predictions.," *PloS one*, vol. 10, no. 10, p. e0141729, 2015.

[84] K. A. Quadt, L. Barfod, D. Andersen, J. Bruun, B. Gyan, T. Hassenkam, M. F. Ofori, and L. Hviid, "The density of knobs on Plasmodium falciparum-infected erythrocytes depends on developmental age and varies among isolates.," *PloS one*, vol. 7, no. 9, p. e45658, 2012.

[85] S. C. Liu, L. H. Derick, and J. Palek, "Visualization of the hexagonal lattice in the erythrocyte membrane skeleton.," *The Journal of cell biology*, vol. 104, no. 3, pp. 527–536, 1987.

[86] H. Shi, Z. Liu, A. Li, J. Yin, A. G. L. Chong, K. S. W. Tan, Y. Zhang, and C. T. Lim, "Life cycle-dependent cytoskeletal modifications in Plasmodium falciparum infected erythrocytes.," *PloS one*, vol. 8, no. 4, p. e61170, 2013.

[87] A. Kilejian, M. A. Rashid, M. Aikawa, T. Aji, and Y. F. Yang, "Selective association of a fragment of the knob protein with spectrin, actin and the red cell membrane.," *Molecular and biochemical parasitology*, vol. 44, no. 2, pp. 175–181, 1991.

[88] C. Magowan, W. Nunomura, K. L. Waller, J. Yeung, J. Liang, H. Van Dort, P. S. Low, R. L. Coppel, and N. Mohandas, "Plasmodium falciparum histidine-rich protein 1 associates with the band 3 binding domain of ankyrin in the infected red cell membrane," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1502, no. 3, pp. 461–470, 2000.

[89] Y. Zhang, C. Huang, S. Kim, M. Golkaram, M. W. A. Dixon, L. Tilley, J. Li, S. Zhang, and S. Suresh, "Multiple stiffening effects of nanoscale knobs on human red blood cells infected with Plasmodium falciparum malaria parasite.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 19, pp. 6068–6073, 2015.

[90] W. B. Gratzer, "The red cell membrane and its cytoskeleton," *Biochemical Journal*, vol. 198, no. 1, pp. 1–8, 1981.

[91] H. Kusunoki, R. I. MacDonald, and A. Mondragón, "Structural insights into the stability and flexibility of unusual erythroid spectrin repeats.," *Structure*, vol. 12, no. 4, pp. 645–656, 2004.

[92] J. Leluk, B. Hanus-Lorenz, and A. F. Sikorski, "Application of genetic semihomology algorithm to theoretical studies on various protein families.," *Acta biochimica Polonica*, vol. 48, no. 1, pp. 21–33, 2001.

[93] J. J. Ipsaro, S. L. Harper, T. E. Messick, R. Marmorstein, A. Mondragón, and D. W. Speicher, "Crystal structure and functional interpretation of the erythrocyte spectrin tetramerization domain complex.," *Blood*, vol. 115, no. 23, pp. 4843–4852, 2010.

[94] C. M. Cohen and R. C. Langley, "Functional characterization of human erythrocyte spectrin alpha and beta chains: association with actin and erythrocyte protein 4.1.," *Biochemistry*, vol. 23, no. 19, pp. 4488–4495, 1984.

[95] A. Nans, N. Mohandas, and D. L. Stokes, "Native ultrastructure of the red cell cytoskeleton by cryo-electron tomography.," *Biophysical journal*, vol. 101, no. 10, pp. 2341–2350, 2011.

[96] T. J. Mankelow, T. J. Satchwell, and N. M. Burton, "Refined views of multi-protein complexes in the erythrocyte membrane," *Blood Cells, Molecules, and Diseases*, vol. 49, no. 1, pp. 1–10, 2012.

[97] A. Weber, C. R. Pennise, G. G. Babcock, and V. M. Fowler, "Tropomodulin caps the pointed ends of actin filaments.," *The Journal of cell biology*, vol. 127, no. 6, pp. 1627–1635, 1994.

[98] N. J. Hemming, D. J. Anstee, M. A. Staricoff, M. J. A. Tanner, and N. Mohandas, "Identification of the Membrane Attachment Sites for Protein 4.1 in the Human Erythrocyte," *The Journal of biological chemistry*, vol. 270, no. 10, pp. 5360–5366, 1995.

[99] J. J. Ipsaro and A. Mondragón, "Structural basis for spectrin recognition by ankyrin.," *Blood*, vol. 115, no. 20, pp. 4093–4101, 2010.

[100] V. Bennett and P. J. Stenbuck, "Association between ankyrin and the cytoplasmic domain of band 3 isolated from the human erythrocyte membrane.," *The Journal of biological chemistry*, vol. 255, no. 13, pp. 6424–6432, 1980.

[101] K. Kümpornsin, S. Jiemsup, S. Yongkiettrakul, and T. Chookajorn, "Characterization of band 3-ankyrin-Protein 4.2 complex by biochemical and mass spectrometry approaches.," *Biochemical and biophysical research communications*, vol. 406, no. 3, pp. 332–335, 2011.

[102] M. Wolny, M. Grzybek, E. Bok, A. Chorzalska, M. Lenoir, A. Czogalla, K. Adamczyk, A. Kolondra, W. Diakowski, M. Overduin, and A. F. Sikorski, "Key amino acid residues of ankyrin-sensitive phosphatidylethanolamine/phosphatidylcholine-lipid binding site of βI-spectrin.," *PloS one*, vol. 6, no. 6, p. e21538, 2011.

[103] N. Mohandas and X. An, "New insights into function of red cell membrane proteins and their interaction with spectrin-based membrane skeleton," *Transfusion Clinique et Biologique*, vol. 13, no. 1-2, pp. 29–30, 2006.

[104] A. Tsuji and S. Ohnishi, "Restriction of the lateral motion of band 3 in the erythrocyte membrane by the cytoskeletal network: dependence on spectrin association state," *Biochemistry*, vol. 25, no. 20, pp. 6133–6139, 1986.

[105] H. Li, Y. Zhang, V. Ha, and G. Lykotrafitis, "Modeling of band-3 protein diffusion in the normal and defective red blood cell membrane," *Soft Matter*, vol. 12, no. 15, pp. 3643–3653, 2016.

[106] D. E. Discher, D. H. Boal, and S. K. Boey, "Simulations of the erythrocyte cytoskeleton at large deformation. II. Micropipette aspiration.," *Biophysical journal*, vol. 75, no. 3, pp. 1584–1597, 1998.

[107] Z. Peng, X. Li, I. V. Pivkin, M. Dao, G. E. Karniadakis, and S. Suresh, "Lipid bilayer and cytoskeletal interactions in a red blood cell.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 33, pp. 13356–13361, 2013.

[108] H. Li and G. Lykotrafitis, "Two-component coarse-grained molecular-dynamics model for the human erythrocyte membrane.," *Biophysical journal*, vol. 102, no. 1, pp. 75–84, 2012.

[109] H. Li and G. Lykotrafitis, "Erythrocyte membrane model with explicit description of the lipid bilayer and the spectrin network.," *Biophysical journal*, vol. 107, no. 3, pp. 642–653, 2014.

[110] T. Tiffert, R. M. Bookchin, and V. L. Lew, "Calcium Homeostasis in Normal and Abnormal Human Red Cells," in *Red Cell Membrane Transport in Health and Disease*, pp. 373–405, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[111] C. R. Lombardo and P. S. Low, "Calmodulin modulates protein 4.1 binding to human erythrocyte membranes," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1196, no. 2, pp. 139–144, 1994.

[112] P. A. Kuhlman, C. A. Hughes, V. Bennett, and V. M. Fowler, "A New Function for Adducin Calcium/Calmodulin-Regulated Capping of the Barbed Ends of Actin Filaments," *The Journal of biological chemistry*, vol. 271, no. 14, pp. 7986–7991, 1996.

[113] S. Manno, Y. Takakuwa, and N. Mohandas, "Modulation of erythrocyte membrane mechanical function by protein 4.1 phosphorylation.," *The Journal of biological chemistry*, vol. 280, no. 9, pp. 7581–7587, 2005.

[114] E. Ferru, K. Giger, A. Pantaleo, E. Campanella, J. Grey, K. Ritchie, R. Vono, F. Turrini, and P. S. Low, "Regulation of membrane-cytoskeletal interactions by tyrosine phosphorylation of erythrocyte band 3.," *Blood*, vol. 117, no. 22, pp. 5998–6006, 2011.

[115] E. J. Luna and A. L. Hitt, "Cytoskeleton–plasma membrane interactions," *Science*, vol. 258, no. 5084, pp. 955–964, 1992.

[116] Y. Wu, M. M. Nelson, A. Quaile, D. Xia, J. M. Wastling, and A. Craig, "Identification of phosphorylated proteins in erythrocytes infected by the human malaria parasite Plasmodium falciparum," *Malaria journal*, vol. 8, no. 1, p. 105, 2009.

[117] X. Han and R. W. Gross, "Electrospray ionization mass spectroscopic analysis of human erythrocyte plasma membrane phospholipids.," *Proceedings of the National Academy of Sciences*, vol. 91, no. 22, pp. 10635–10639, 1994.

[118] X. An, X. Zhang, G. Debnath, A. J. Baines, and N. Mohandas, "Phosphatidylinositol-4,5-Biphosphate (PIP 2) Differentially Regulates the Interaction of Human Erythrocyte Protein 4.1 (4.1R) with Membrane Proteins †," *Biochemistry*, vol. 45, no. 18, pp. 5725–5732, 2006.

[119] X. Pei, X. Guo, R. Coppel, S. Bhattacharjee, K. Haldar, W. Gratzer, N. Mohandas, and X. An, "The ring-infected erythrocyte surface antigen (RESA) of Plasmodium falciparum stabilizes spectrin tetramers and suppresses further invasion.," *Blood*, vol. 110, no. 3, pp. 1036–1042, 2007.

[120] L. M. Kats, N. I. Proellocks, D. W. Buckingham, L. Blanc, J. Hale, X. Guo, X. Pei, S. Herrmann, E. G. Hanssen, R. L. Coppel, N. Mohandas, X. An, and B. M. Cooke, "Interactions between Plasmodium falciparum skeleton-binding protein 1 and the membrane skeleton of malaria-infected red blood cells," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1848, no. 7, pp. 1619–1628, 2015.

[121] X. Pei, X. Guo, R. Coppel, N. Mohandas, and X. An, "Plasmodium falciparum erythrocyte membrane protein 3 (PfEMP3) destabilizes erythrocyte membrane skeleton.," *The Journal of biological chemistry*, vol. 282, no. 37, pp. 26754–26758, 2007.

[122] K. L. Waller, L. M. Stubberfield, V. Dubljevic, W. Nunomura, X. An, A. J. Mason, N. Mohandas, B. M. Cooke, and R. L. Coppel, "Interactions of Plasmodium falciparum erythrocyte membrane protein 3 with the red blood cell membrane skeleton," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1768, no. 9, pp. 2145–2156, 2007.

[123] F. K. Glenister, K. M. Fernandez, L. M. Kats, E. Hanssen, N. Mohandas, R. L. Coppel, and B. M. Cooke, "Functional alteration of red blood cells by a megadalton protein of Plasmodium falciparum.," *Blood*, vol. 113, no. 4, pp. 919–928, 2009.

[124] R. L. Coppel, "Repeat structures in a Plasmodium falciparum protein (MESA) that binds human erythrocyte protein 4.1.," *Molecular and biochemical parasitology*, vol. 50, no. 2, pp. 335–347, 1992.

[125] K. L. Waller, W. Nunomura, X. An, B. M. Cooke, N. Mohandas, and R. L. Coppel, "Mature parasite-infected erythrocyte surface antigen (MESA) of Plasmodium falciparum binds to the 30-kDa domain of protein 4.1 in malaria-infected red blood cells.," *Blood*, vol. 102, no. 5, pp. 1911–1914, 2003.

[126] S. Lustigman, R. F. Anders, G. V. Brown, and R. L. Coppel, "The mature-parasite-infected erythrocyte surface antigen (MESA) of Plasmodium falciparum associates with the erythrocyte membrane skeletal protein, band 4.1.," *Molecular and biochemical parasitology*, vol. 38, no. 2, pp. 261–270, 1990.

[127] C. Petersen, R. Nelson, C. Magowan, W. Wollish, J. Jensen, and J. Leech, "The mature erythrocyte surface antigen of Plasmodium falciparum is not required for knobs or cytoadherence," *Molecular and biochemical parasitology*, vol. 36, no. 1, pp. 61–65, 1989.

[128] C. Magowan, R. L. Coppel, A. O. Lau, M. M. Moronne, G. Tchernia, and N. Mohandas, "Role of the Plasmodium falciparum mature-parasite-infected erythrocyte surface antigen (MESA/PfEMP-2) in malarial infection of erythrocytes.," *Blood*, vol. 86, no. 8, pp. 3196–3204, 1995.

[129] G. K. Kilili and D. J. LaCount, "An erythrocyte cytoskeleton-binding motif in exported Plasmodium falciparum proteins.," *Eukaryotic cell*, vol. 10, no. 11, pp. 1439–1447, 2011.

[130] P. Raphael, Y. Takakuwa, S. Manno, S. C. Liu, A. H. Chishti, and M. Hanspal, "A cysteine protease activity from Plasmodium falciparum cleaves human erythrocyte ankyrin.," *Molecular and biochemical parasitology*, vol. 110, no. 2, pp. 259–272, 2000.

[131] M. Hanspal, M. Dua, Y. Takakuwa, A. H. Chishti, and A. Mizuno, "Plasmodium falciparum cysteine protease falcipain-2 cleaves erythrocyte membrane skeletal proteins at late stages of parasite development.," *Blood*, vol. 100, no. 3, pp. 1048–1054, 2002.

[132] S. Dhawan, M. Dua, A. H. Chishti, and M. Hanspal, "Ankyrin peptide blocks falcipain-2-mediated malaria parasite release from red blood cells.," *The Journal of biological chemistry*, vol. 278, no. 32, pp. 30180–30186, 2003.

[133] R. Chandramohanadas, P. H. Davis, D. P. Beiting, M. B. Harbut, C. Darling, G. Velmourougane, M. Y. Lee, P. A. Greer, D. S. Roos, and D. C. Greenbaum, "Apicomplexan parasites co-opt host calpains to facilitate their escape from infected cells.," *Science*, vol. 324, no. 5928, pp. 794–797, 2009.

[134] S. Das, N. Hertrich, A. J. Perrin, C. Withers-Martinez, C. R. Collins, M. L. Jones, J. M. Watermeyer, E. T. Fobes, S. R. Martin, H. R. Saibil, G. J. Wright, M. Treeck, C. Epp, and M. J. Blackman, "Processing of Plasmodium falciparum Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs.," *Cell host & microbe*, vol. 18, no. 4, pp. 433–444, 2015.

[135] A. Esposito, J.-B. Choimet, J. N. Skepper, J. M. A. Mauritz, V. L. Lew, C. F. Kaminski, and T. Tiffert, "Quantitative Imaging of Human Red Blood Cells Infected with Plasmodium falciparum," *Biophysical journal*, vol. 99, no. 3, pp. 953–960, 2010.

[136] J. M. Watermeyer, V. L. Hale, F. Hackett, D. K. Clare, E. E. Cutts, I. Vakonakis, R. A. Fleck, M. J. Blackman, and H. R. Saibil, "A spiral scaffold underlies cytoadherent knobs in Plasmodium falciparum-infected erythrocytes.," *Blood*, pp. blood–2015–10–674002, 2015.

[137] M. G. Millholland, R. Chandramohanadas, A. Pizzarro, A. Wehr, H. Shi, C. Darling, C. T. Lim, and D. C. Greenbaum, "The malaria parasite progressively dismantles the host erythrocyte cytoskeleton for efficient egress.," *Molecular & cellular proteomics : MCP*, vol. 10, no. 12, pp. M111.010678–M111.010678, 2011.

[138] T. F. Taraschi, M. O'Donnell, S. Martinez, T. Schneider, D. Trelka, V. M. Fowler, L. Tilley, and Y. Moriyama, "Generation of an erythrocyte vesicle transport system by Plasmodium falciparum malaria parasites.," *Blood*, vol. 102, no. 9, pp. 3420–3426, 2003.

[139] A. G. Maier, B. M. Cooke, A. F. Cowman, and L. Tilley, "Malaria parasite proteins that remodel the host erythrocyte," *Nature Reviews Microbiology*, vol. 7, no. 5, pp. 341–354, 2009.

[140] B. M. Cooke, D. W. Buckingham, F. K. Glenister, K. M. Fernandez, L. H. Bannister, M. Marti, N. Mohandas, and R. L. Coppel, "A Maurer's cleft-associated protein is essential for expression of the major malaria virulence antigen on the surface of infected red blood cells.," *The Journal of cell biology*, vol. 172, no. 6, pp. 899–908, 2006.

[141] E. Nagao, O. Kaneko, and J. A. Dvorak, "Plasmodium falciparum-infected erythrocytes: qualitative and quantitative analyses of parasite-induced knobs by atomic force microscopy.," *Journal of structural biology*, vol. 130, no. 1, pp. 34–44, 2000.

[142] K. B. Rogala, N. J. Dynes, G. N. Hatzopoulos, J. Yan, S. K. Pong, C. V. Robinson, C. M. Deane, P. Gönczy, and I. Vakonakis, "The Caenorhabditis elegans protein SAS-5 forms large oligomeric assemblies critical for centriole formation.," *eLife*, vol. 4, p. e07410, 2015.

[143] M. Yasunaga, J. J. Ipsaro, and A. Mondragón, "Structurally Similar but Functionally Diverse ZU5 Domains in Human Erythrocyte Ankyrin," *Journal of molecular biology*, vol. 417, no. 4, pp. 336–350, 2012.

[144] D. Zhang, A. Kiyatkin, J. T. Bolin, and P. S. Low, "Crystallographic structure and functional interpretation of the cytoplasmic domain of erythrocyte membrane band 3.," *Blood*, vol. 96, no. 9, pp. 2925–2933, 2000.

[145] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, and A. Bax, "NMRPipe: A multidimensional spectral processing system based on UNIX pipes," *Journal of biomolecular NMR*, vol. 6, no. 3, pp. 277–293, 1995.

[146] T. D. Goddard and D. G. Kneller, *SPARKY 3*. University of California, 2004.

[147] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton, *Protein NMR Spectroscopy*. Principles and Practice, California: Elsevier Academic Press, second edition ed., 2007.

[148] J. J. Helmus and C. P. Jaroniec, "Nmrglue: an open source Python package for the analysis of multidimensional NMR data.," *Journal of biomolecular NMR*, vol. 55, no. 4, pp. 355–367, 2013.

[149] N. U. Jain, A. Venot, K. Umemoto, H. Leffler, and J. H. Prestegard, "Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands," *Protein Science*, vol. 10, no. 11, pp. 2393–2400, 2008.

[150] P. E. Johnson, E. Brun, L. F. MacKenzie, S. G. Withers, and L. P. McIntosh, "The cellulose-binding domains from Cellulomonas fimi β-1,4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations," *Journal of molecular biology*, vol. 287, no. 3, pp. 609–625, 1999.

[151] M. Sattler, "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 34, no. 2, pp. 93–158, 1999.

[152] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue, "The CCPN data model for NMR spectroscopy: Development of a software pipeline," *Proteins: Structure, Function, and Bioinformatics*, vol. 59, no. 4, pp. 687–696, 2005.

[153] F. H. Schumann, H. Riepl, T. Maurer, W. Gronwald, K.-P. Neidig, and H. R. Kalbitzer, "Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions.," *Journal of biomolecular NMR*, vol. 39, no. 4, pp. 275–289, 2007.

[154] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun, "PRIMUS: a Windows PC-based system for small-angle scattering data analysis," *Journal of applied crystallography*, vol. 36, no. 5, pp. 1277–1282, 2003.

[155] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, "MolProbity: all-atom structure validation for macromolecular crystallography.," *Acta crystallographica. Section D, Biological crystallography*, vol. 66, no. Pt 1, pp. 12–21, 2010.

[156] Schrodinger, LLC, "The PyMOL Molecular Graphics System, Version 1.8," 2015.

[157] P. R. Evans, "An introduction to stereochemical restraints," *Acta crystallographica. Section D, Biological crystallography*, vol. 63, no. 1, pp. 58–61, 2006.

[158] Collaborative Computational Project, Number 4, "The CCP4 suite: programs for protein crystallography.," *Acta crystallographica. Section D, Biological crystallography*, vol. 50, no. Pt 5, pp. 760–763, 1994.

[159] F. Long, A. A. Vagin, P. Young, and G. N. Murshudov, "BALBES: a molecular-replacement pipeline.," *Acta crystallographica. Section D, Biological crystallography*, vol. 64, no. Pt 1, pp. 125–132, 2008.

[160] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software.," *Journal of applied crystallography*, vol. 40, no. Pt 4, pp. 658–674, 2007.

[161] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of Coot.," *Acta crystallographica. Section D, Biological crystallography*, vol. 66, no. Pt 4, pp. 486–501, 2010.

[162] P. D. Adams, R. W. Grosse-Kunstleve, L.-W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, and T. C. Terwilliger, "PHENIX: building new software for automated crystallographic structure determination," *Acta crystallographica. Section D, Biological crystallography*, vol. 58, no. 11, pp. 1948–1954, 2002.

[163] G. Bricogne, E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, O. Smart, C. Vonrhein, and T. Womack, "BUSTER-TNT 2.X," 2014.

[164] W. Kabsch, "XDS," *Acta crystallographica. Section D, Biological crystallography*, vol. 66, no. 2, pp. 125–132, 2010.

[165] B. Webb and A. Sali, *Comparative Protein Structure Modeling Using MODELLER*. Comparative Protein Structure Modeling Using MODELLER, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2002.

[166] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.," *Molecular systems biology*, vol. 7, no. 1, pp. 539–539, 2011.

[167] M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, and R. Lopez, "A new bioinformatics analysis tools framework at EMBL-EBI.," *Nucleic Acids Research*, vol. 38, no. Web Server issue, pp. W695–9, 2010.

[168] W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso, and R. Lopez, "The EMBL-EBI bioinformatics web and programmatic tools framework.," *Nucleic Acids Research*, vol. 43, no. W1, pp. W580–4, 2015.

[169] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis.," *Nature protocols*, vol. 10, no. 6, pp. 845–858, 2015.

[170] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, and A. Bax, "Consistent blind protein structure generation from NMR chemical shift data.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 12, pp. 4685–4690, 2008.

[171] S. J. de Vries, M. van Dijk, and A. M. J. J. Bonvin, "The HADDOCK web server for data-driven biomolecular docking.," *Nature protocols*, vol. 5, no. 5, pp. 883–897, 2010.

[172] G. C. P. van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. van Dijk, S. J. de Vries, and A. M. J. J. Bonvin, "The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.," *Journal of molecular biology*, vol. 428, no. 4, pp. 720–725, 2016.

[173] L. Cavallo, J. Kleinjung, and F. Fraternali, "POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level.," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3364–3366, 2003.

[174] J. W. Ponder and D. A. Case, "Force fields for protein simulations.," *Advances in protein chemistry*, vol. 66, pp. 27–85, 2003.

[175] A. R. Leach, *Molecular Modelling: Principles and Applications.* Essex: Pearson Education, second edition ed., 2001.

[176] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems," *The Journal of chemical physics*, vol. 98, no. 12, pp. 10089–10092, 1993.

[177] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *The Journal of chemical physics*, vol. 103, no. 19, pp. 8577–8593, 1995.

[178] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Annalen der Physik*, vol. 369, no. 3, pp. 253–287, 1921.

[179] H. Schreiber and O. Steinhauser, "Molecular dynamics studies of solvated polypeptides: Why the cut-off scheme does not work," *Chemical physics*, vol. 168, no. 1, pp. 75–89, 1992.

[180] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field.," *Journal of computational chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.

[181] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. Van Gunsteren, "The GROMOS Biomolecular Simulation Program Package," *The Journal of Physical Chemistry A*, vol. 103, no. 19, pp. 3596–3607, 1999.

[182] L. D. Schuler, X. Daura, and W. F. Van Gunsteren, "An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase," *Journal of computational chemistry*, vol. 22, no. 11, pp. 1205–1218, 2001.

[183] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 8, pp. NA–NA, 2010.

[184] X. Daura, A. E. Mark, and W. F. Van Gunsteren, "Parametrization of aliphatic CHn united atoms of GROMOS96 force field," *Journal of computational chemistry*, vol. 19, no. 5, pp. 535–547, 1998.

[185] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1656–1676, 2004.

[186] R. W. Hockney, S. P. Goel, and J. W. Eastwood, "Quiet high-resolution computer models of a plasma," *Journal of Computational Physics*, vol. 14, no. 2, pp. 148–158, 1974.

[187] A. K. Mazur, "Common Molecular Dynamics Algorithms Revisited: Accuracy and Optimal Time Steps of Störmer–Leapfrog Integrators," *Journal of Computational Physics*, vol. 136, no. 2, pp. 354–365, 1997.

[188] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *Journal of computational chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[189] M. Parrinello and A. Rahman, "Polymorphic Transitions in Alkali Halides, a molecular Dynamics Study," *Le Journal de Physique Colloques*, vol. 42, no. C6, pp. C6–511–C6–515, 1981.

[190] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling.," *The Journal of chemical physics*, vol. 126, no. 1, p. 014101, 2007.

[191] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," *The Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987.

[192] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.

[193] M. J. Abraham, D. van der Spoel, E. Lindahl, and B. Hess, "GROMACS User Manual version 5.1," 2015.

[194] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[195] D. Frishman and P. Argos, "Knowledge-based protein secondary structure assignment," *Proteins*, vol. 23, no. 4, pp. 566–579, 1995.

[196] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[197] D. Wishart and B. Sykes, "The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data," *Journal of biomolecular NMR*, vol. 4, no. 2, pp. 171–180, 1994.

[198] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax, "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.," *Journal of biomolecular NMR*, vol. 44, no. 4, pp. 213–223, 2009.

[199] A. Oberli, L. Zurbrügg, S. Rusch, F. Brand, M. E. Butler, J. L. Day, E. E. Cutts, T. Lavstsen, I. Vakonakis, and H.-P. Beck, "Plasmodium falciparum PHIST Proteins Contribute to Cytoadherence and Anchor PfEMP1 to the Host Cell Cytoskeleton.," *Cellular microbiology*, pp. n/a–n/a, 2016.

[200] R. B. Best, N.-V. Buchete, and G. Hummer, "Are Current Molecular Dynamics Force Fields too Helical?," *Biophysical journal*, vol. 95, no. 1, pp. L07–L09, 2008.

[201] E. A. Cino, W.-Y. Choy, and M. Karttunen, "Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations," *Journal of chemical theory and computation*, vol. 8, no. 8, pp. 2725–2740, 2012.

[202] O. F. Lange, D. van der Spoel, and B. L. de Groot, "Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data.," *Biophysical journal*, vol. 99, no. 2, pp. 647–655, 2010.

[203] J. R. Casey and R. A. Reithmeier, "Analysis of the oligomeric state of Band 3, the anion transport protein of the human erythrocyte membrane, by size exclusion high performance liquid chromatography. Oligomeric stability and origin of heterogeneity.," *The Journal of biological chemistry*, vol. 266, no. 24, pp. 15726–15737, 1991.

[204] G. Celniker, G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko, and N. Ben-Tal, "ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function," *Israel Journal of Chemistry*, vol. 53, no. 3-4, pp. 199–206, 2013.

[205] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, and N. Ben-Tal, "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.," *Nucleic Acids Research*, vol. 38, no. Web Server issue, pp. W529–33, 2010.

[206] D. M. Shotton, B. E. Burke, and D. Branton, "The molecular structure of human erythrocyte spectrin. Biophysical and electron microscopic studies.," *Journal of molecular biology*, vol. 131, no. 2, pp. 303–329, 1979.

[207] J. T. Yang, C. S. Wu, and H. M. Martinez, "Calculation of protein conformation from circular dichroism.," *Methods in enzymology*, vol. 130, pp. 208–269, 1986.

[208] L. Whitmore and B. A. Wallace, "Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases," *Biopolymers*, vol. 89, no. 5, pp. 392–400, 2008.

[209] V. L. Grum, D. Li, R. I. MacDonald, and A. Mondragón, "Structures of two repeats of spectrin suggest models of flexibility.," *Cell*, vol. 98, no. 4, pp. 523–535, 1999.

[210] H. Kusunoki, G. Minasov, R. I. MacDonald, and A. Mondragón, "Independent movement, dimerization and stability of tandem repeats of chicken brain alpha-spectrin.," *Journal of molecular biology*, vol. 344, no. 2, pp. 495–511, 2004.

[211] P. R. Evans and G. N. Murshudov, "How good are my data and what is the resolution?," *Acta crystallographica. Section D, Biological crystallography*, vol. 69, no. Pt 7, pp. 1204–1214, 2013.

[212] D. Svergun, C. Barberato, and M. H. J. Koch, "CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates," *Journal of applied crystallography*, vol. 28, no. 6, pp. 768–773, 1995.

[213] J. B. Matthew, P. C. Weber, F. R. Salemme, and F. M. Richards, "Electrostatic orientation during electron transfer between flavodoxin and cytochrome c," *Nature*, vol. 301, no. 5896, pp. 169–171, 1983.

[214] P. Roca-Cusachs, N. C. Gauthier, A. Del Rio, and M. P. Sheetz, "Clustering of alpha(5)beta(1) integrins determines adhesion strength whereas alpha(v)beta(3) and talin enable mechanotransduction.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 38, pp. 16245–16250, 2009.

[215] L. Stevenson, P. Huda, A. Jeppesen, E. Laursen, J. A. Rowe, A. Craig, W. Streicher, L. Barfod, and L. Hviid, "Investigating the function of Fc-specific binding of IgM to Plasmodium falciparum erythrocyte membrane protein 1 mediating erythrocyte rosetting.," *Cellular microbiology*, vol. 17, no. 6, pp. 819–831, 2015.

[216] L. Stevenson, E. Laursen, G. J. Cowan, B. Bandoh, L. Barfod, D. R. Cavanagh, G. R. Andersen, and L. Hviid, "α2-Macroglobulin Can Crosslink Multiple Plasmodium falciparum Erythrocyte Membrane Protein 1 (PfEMP1) Molecules and May Facilitate Adhesion of Parasitized Erythrocytes.," *PLoS pathogens*, vol. 11, no. 7, p. e1005022, 2015.

[217] Z. R. Yang, R. Thomson, P. McNeil, and R. M. Esnouf, "RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.," *Bioinformatics (Oxford, England)*, vol. 21, no. 16, pp. 3369–3376, 2005.

[218] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, and A. K. Dunker, "Predicting intrinsic disorder from amino acid sequence.," *Proteins*, vol. 53 Suppl 6, no. S6, pp. 566–572, 2003.

[219] K. Peng, P. Radivojac, S. Vucetic, A. K. Dunker, and Z. Obradovic, "Length-dependent prediction of protein intrinsic disorder.," *BMC bioinformatics*, vol. 7, no. 1, p. 208, 2006.

[220] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, and A. K. Dunker, "Exploiting heterogeneous sequence properties improves prediction of protein disorder.," *Proteins*, vol. 61 Suppl 7, no. S7, pp. 176–182, 2005.

[221] D. A. Calderwood, I. D. Campbell, and D. R. Critchley, "Talins and kindlins: partners in integrin-mediated adhesion," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 8, pp. 503–517, 2013.

[222] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker, "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations.," *Nucleic Acids Research*, vol. 35, no. Web Server issue, pp. W522–5, 2007.

[223] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.," *Nucleic Acids Research*, vol. 32, no. Web Server issue, pp. W665–7, 2004.

[224] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: application to microtubules and the ribosome.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 18, pp. 10037–10041, 2001.

[225] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera–a visualization system for exploratory research and analysis.," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[226] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: a toolkit for the analysis of molecular dynamics simulations.," *Journal of computational chemistry*, vol. 32, no. 10, pp. 2319–2327, 2011.

[227] F. Nelli, "Machine Learning with scikit-learn," in *Python Data Analytics*, pp. 237–264, Berkeley, CA: Apress, 2015.

[228] D. A. Schult and P. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science ...*, 2008.

[229] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[230] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore, "The Xplor-NIH NMR molecular structure determination package.," *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 160, no. 1, pp. 65–73, 2003.

[231] J. Kuszewski, A. M. Gronenborn, and G. M. Clore, "Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases.," *Protein science : a publication of the Protein Society*, vol. 5, no. 6, pp. 1067–1080, 1996.

[232] A. Grishaev and A. Bax, "An Empirical Backbone–Backbone Hydrogen-Bonding Potential in Proteins and Its Applications to NMR Structure Refinement and Validation," *Journal of the American Chemical Society*, vol. 126, no. 23, pp. 7281–7292, 2004.

[233] B. Raveh, N. London, L. Zimmerman, and O. Schueler-Furman, "Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors.," *PloS one*, vol. 6, no. 4, p. e18934, 2011.

[234] C. C. LaBrake, L. Wang, T. A. Keiderling, and L. W. Fung, "Fourier transform infrared spectroscopic studies of the secondary structure of spectrin under different ionic strengths.," *Biochemistry*, vol. 32, no. 39, pp. 10296–10302, 1993.

[235] D. T. Mirijanian and G. A. Voth, "Unique elastic properties of the spectrin tetramer as revealed by multiscale coarse-grained modeling.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 4, pp. 1204–1208, 2008.

[236] K. P. Ng, G. Potikyan, R. O. V. Savene, C. T. Denny, V. N. Uversky, and K. A. W. Lee, "Multiple aromatic side chains within a disordered structure are critical for transcription and transforming activity of EWS family oncoproteins.," *Proceedings of the National Academy of Sciences*, vol. 104, no. 2, pp. 479–484, 2007.

[237] J. Ma and M. Ptashne, "A new class of yeast transcriptional activators.," *Cell*, vol. 51, no. 1, pp. 113–119, 1987.

[238] I. A. Hope, S. Mahadevan, and K. Struhl, "Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein.," *Nature*, vol. 333, no. 6174, pp. 635–640, 1988.

[239] E. B. Gibbs and S. A. Showalter, "Quantitative biophysical characterization of intrinsically disordered proteins.," *Biochemistry*, vol. 54, no. 6, pp. 1314–1326, 2015.

[240] P. Tompa and M. Fuxreiter, "Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions.," *Trends in biochemical sciences*, vol. 33, no. 1, pp. 2–8, 2008.

[241] P. Csermely, R. Palotai, and R. Nussinov, "Induced fit, conformational selection and independent dynamic segments: an extended view of binding events.," *Trends in biochemical sciences*, vol. 35, no. 10, pp. 539–546, 2010.

[242] H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions.," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 3, pp. 197–208, 2005.

[243] A. Cumberworth, G. Lamour, M. M. Babu, and J. Gsponer, "Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes," *Biochemical Journal*, vol. 454, no. 3, pp. 361–369, 2013.

[244] C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, and L. M. Iakoucheva, "Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.," *PLoS computational biology*, vol. 2, no. 8, p. e100, 2006.

[245] K. Van Roey, T. J. Gibson, and N. E. Davey, "Motif switches: decision-making in cell regulation.," *Current opinion in structural biology*, vol. 22, no. 3, pp. 378–385, 2012.

[246] V. J. Hilser and E. B. Thompson, "Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins," *Proceedings of the National Academy of Sciences*, vol. 104, no. 20, pp. 8311–8315, 2007.

[247] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation.," *Nature Reviews Molecular Cell Biology*, vol. 16, no. 1, pp. 18–29, 2015.

[248] S. C. Liu, L. H. Derick, and J. Palek, "Visualization of the hexagonal lattice in the erythrocyte membrane skeleton.," *The Journal of cell biology*, vol. 104, no. 3, pp. 527–536, 1987.

[249] D. B. Lukatsky, A. Afek, and E. I. Shakhnovich, "Sequence correlations shape protein promiscuity.," *The Journal of chemical physics*, vol. 135, no. 6, p. 065104, 2011.

[250] M. A. M. Reijns, R. D. Alexander, M. P. Spiller, and J. D. Beggs, "A role for Q/N-rich aggregation-prone regions in P-body localization.," *Journal of cell science*, vol. 121, no. Pt 15, pp. 2463–2472, 2008.

[251] P. Anderson and N. Kedersha, "RNA granules.," *The Journal of cell biology*, vol. 172, no. 6, pp. 803–808, 2006.

[252] M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko, and S. L. McKnight, "Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels.," *Cell*, vol. 149, no. 4, pp. 753–767, 2012.

[253] T. W. Han, M. Kato, S. Xie, L. C. Wu, H. Mirzaei, J. Pei, M. Chen, Y. Xie, J. Allen, G. Xiao, and S. L. McKnight, "Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies.," *Cell*, vol. 149, no. 4, pp. 768–779, 2012.

[254] A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman, and S. Alberti, "A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.," *Cell*, vol. 162, no. 5, pp. 1066–1077, 2015.

[255] J. A. Toretsky and P. E. Wright, "Assemblages: functional units formed by cellular phase separation.," *The Journal of cell biology*, vol. 206, no. 5, pp. 579–588, 2014.

[256] M. Altmeyer, K. J. Neelsen, F. Teloni, I. Pozdnyakova, S. Pellegrino, M. Grøfte, M.-B. D. Rask, W. Streicher, S. Jungmichel, M. L. Nielsen, and J. Lukas, "Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose).," *Nature communications*, vol. 6, p. 8088, 2015.

[257] L. Le, E. H. Lee, D. J. Hardy, T. N. Truong, and K. Schulten, "Molecular dynamics simulations suggest that electrostatic funnel directs binding of Tamiflu to influenza N1 neuraminidases.," *PLoS computational biology*, vol. 6, no. 9, p. e1000939, 2010.

[258] B. F. C. Kafsack, H. J. Painter, and M. Llinás, "New Agilent platform DNA microarrays for transcriptome analysis of Plasmodium falciparum and Plasmodium berghei for the malaria research community.," *Malaria journal*, vol. 11, no. 1, p. 187, 2012.

[259] N. Rovira-Graells, A. P. Gupta, E. Planet, V. M. Crowley, S. Mok, L. R. de Pouplana, P. R. Preiser, Z. Bozdech, and A. Cortés, "Transcriptional variation in the malaria parasite Plasmodium falciparum," *Genome Research*, vol. 22, no. 5, pp. 925–938, 2012.

[260] A. Claessens, Y. Adams, A. Ghumra, G. Lindergard, C. C. Buchan, C. Andisi, P. C. Bull, S. Mok, A. P. Gupta, C. W. Wang, L. Turner, M. Arman, A. Raza, Z. Bozdech, and J. A. Rowe, "A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 26, pp. E1772–81, 2012.

[261] R. Subramani, K. Quadt, A. E. Jeppesen, C. Hempel, J. E. V. Petersen, T. Hassenkam, L. Hviid, and L. Barfod, "Plasmodium falciparum-infected erythrocyte knob density is linked to the PfEMP1 variant expressed.," *mBio*, vol. 6, no. 5, pp. e01456–15, 2015.

[262] L. Florens, X. Liu, Y. Wang, S. Yang, O. Schwartz, M. Peglar, D. J. Carucci, J. R. Yates III, and Y. Wu, "Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes," *Molecular and biochemical parasitology*, vol. 135, no. 1, pp. 1–11, 2004.

[263] T. Y. Sam-Yellowe, "The role of the Maurer's clefts in protein transport in Plasmodium falciparum," *Trends in Parasitology*, vol. 25, no. 6, pp. 277–284, 2009.

[264] Q.-T. Shen, A. L. Schuh, Y. Zheng, K. Quinney, L. Wang, M. Hanna, J. C. Mitchell, M. S. Otegui, P. Ahlquist, Q. Cui, and A. Audhya, "Structural analysis and modeling reveals new mechanisms governing ESCRT-III spiral filament assembly.," *The Journal of cell biology*, vol. 206, no. 6, pp. 763–777, 2014.

[265] W. M. Henne, H. Stenmark, and S. D. Emr, "Molecular mechanisms of the membrane sculpting ESCRT pathway.," *Cold Spring Harbor perspectives in biology*, vol. 5, no. 9, pp. a016766–a016766, 2013.

[266] J. M. Daniel and A. B. Reynolds, "Tyrosine phosphorylation and cadherin/catenin function," *BioEssays*, vol. 19, no. 10, pp. 883–891, 1997.

[267] J. Adovelande, B. Bastide, J. Deleze, and J. Schrevel, "Cytosolic Free Calcium in Plasmodium falciparum-Infected Erythrocytes and the Effect of Verapamil: A Cytofluorometric Study," *Experimental Parasitology*, vol. 76, no. 3, pp. 247–258, 1993.

[268] G. L. Jones and H. M. Edmundson, "Protein phosphorylation during the asexual life cycle of the human malarial parasite Plasmodium falciparum," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1053, no. 2-3, pp. 118–124, 1990.

[269] M. C. Murray and M. E. Perkins, "Phosphorylation of erythrocyte membrane and cytoskeleton proteins in cells infected with Plasmodium falciparum," *Molecular and biochemical parasitology*, vol. 34, no. 3, pp. 229–236, 1989.

[270] C. Magowan, J. Liang, J. Yeung, Y. Takakuwa, R. L. Coppel, and N. Mohandas, "Plasmodium falciparum:Influence of Malarial and Host Erythrocyte Skeletal Protein Interactions on Phosphorylation Infected Erythrocytes," *Experimental Parasitology*, vol. 89, no. 1, pp. 40–49, 1998.

[271] S. Percário, D. Moreira, B. Gomes, M. Ferreira, A. Gonçalves, P. Laurindo, T. Vilhena, M. Dolabela, and M. Green, "Oxidative Stress in Malaria," *International Journal of Molecular Sciences*, vol. 13, no. 12, pp. 16346–16372, 2012.

[272] B. N. Pease, E. L. Huttlin, M. P. Jedrychowski, E. Talevich, J. Harmon, T. Dillman, N. Kannan, C. Doerig, R. Chakrabarti, S. P. Gygi, and D. Chakrabarti, "Global Analysis of Protein Expression and Phosphorylation of Three Stages of Plasmodium falciparumIntraerythrocytic Development," *Journal of Proteome Research*, vol. 12, no. 9, pp. 4028–4045, 2013.

[273] M. Treeck, J. L. Sanders, J. E. Elias, and J. C. Boothroyd, "The Phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii Reveal Unusual Adaptations Within and Beyond the Parasites' Boundaries," *Cell host & microbe*, vol. 10, no. 4, pp. 410–419, 2011.

[274] E. Lasonder, J. L. Green, G. Camarda, H. Talabani, A. A. Holder, G. Langsley, and P. Alano, "The Plasmodium falciparumSchizont Phosphoproteome Reveals Extensive Phosphatidylinositol and cAMP-Protein Kinase A Signaling," *Journal of Proteome Research*, vol. 11, no. 11, pp. 5323–5337, 2012.

[275] R. Hora, D. J. Bridges, A. Craig, and A. Sharma, "Erythrocytic Casein Kinase II Regulates Cytoadherence of Plasmodium falciparum-infected Red Blood Cells," *The Journal of biological chemistry*, vol. 284, no. 10, pp. 6260–6269, 2009.

[276] P. M. Durand and T. L. Coetzer, "Hereditary red cell disorders and malaria resistance.," *Haematologica*, vol. 93, no. 7, pp. 961–963, 2008.

[277] O'donnell, Allen, Mgone, Martinson, Clegg, and Weatherall, "Red cell morphology and malaria anaemia in children with Southeast-Asian ovalocytosis band 3 in Papua New Guinea," *British Journal of Haematology*, vol. 101, no. 3, pp. 407–412, 1998.

[278] B. Genton, F. AI-Yaman, C. S. Mgone, N. Alexander, M. M. Paniu, M. P. Alpers, and D. Mokela, "Ovalocytosis and cerebral malaria," *Nature*, vol. 378, no. 6557, pp. 564–565, 1995.

[279] A. Cortés, M. Mellombo, C. S. Mgone, H.-P. Beck, J. C. Reeder, and B. M. Cooke, "Adhesion of Plasmodium falciparum-infected red blood cells to CD36 under flow is enhanced by the cerebral malaria-protective trait South–East Asian ovalocytosis," *Molecular and biochemical parasitology*, vol. 142, no. 2, pp. 252–257, 2005.

[280] R. M. Fairhurst, H. Fujioka, K. Hayton, K. F. Collins, and T. E. Wellems, "Aberrant development of Plasmodium falciparum in hemoglobin CC red cells: implications for the malaria protective effect of the homozygous state.," *Blood*, vol. 101, no. 8, pp. 3309–3315, 2003.

[281] J. A. Olson and R. L. Nagel, "Synchronized cultures of P falciparum in abnormal red cells: the mechanism of the inhibition of growth in HbCC cells.," *Blood*, vol. 67, no. 4, pp. 997–1001, 1986.

[282] G. T. Heller, P. Sormanni, and M. Vendruscolo, "Targeting disordered proteins with small molecules using entropy," *Trends in biochemical sciences*, vol. 40, no. 9, pp. 491–496, 2015.

[283] S. Ambadipudi and M. Zweckstetter, "Targeting intrinsically disordered proteins in rational drug discovery," *Expert Opinion on Drug Discovery*, vol. 11, no. 1, pp. 65–77, 2015.

[284] A. M. Sydor, K. J. Czymmek, E. M. Puchner, and V. Mennella, "Super-Resolution Microscopy: From Single Molecules to Supramolecular Assemblies," *Trends in Cell Biology*, vol. 25, no. 12, pp. 730–748, 2015.

Appendix A

Constructs

Table A.1: Constructs used in experiments

Protein name & UniProt number	Construct name	Construct amino acid boundaries
PHIST domains		
PFI1780w	PHIST domain	98-247
PFI1780w	PHIST domain H167A	98-247
PFI1780w	PHIST domain F168A	98-247
PFI1780w	PHIST domain Q171A	98-247
PFE1605w	PHIST domain	122-335
PFE1605w	C-terminal	300-528
Band 3		
Band 3, P02730	Cytoplasmic domain	1-379
Spectrin domains		
α spectrin, P02549	αN-5	1-585
α spectrin, P02549	α6-11	576-1185
α spectrin, P02549	α12-16	1174-1716
α spectrin, P02549	α17-C	1704-2395
α spectrin, P02549	α16-19	1599-2036
α spectrin, P02549	α17-19	1704-2036
α spectrin, P02549	α18-19	1810-2036
α spectrin, P02549	α19-21	1923-2261
α spectrin, P02549	α16-17	1599-1828
α spectrin, P02549	α17	1704-1828
α spectrin, P02549	α19	1923-2036
α spectrin, P02549	α17 L1709A	1704-1828
α spectrin, P02549	α17 Y1713A	1704-1828
α spectrin, P02549	α17 F1716A	1704-1828
α spectrin, P02549	α17 Q1717A	1704-1828
α spectrin, P02549	α17 F1718A	1704-1828
α spectrin, P02549	α17 Q1720A	1704-1828
α spectrin, P02549	α17 D1722K	1704-1828
α spectrin, P02549	α17 D1723K	1704-1828
β spectrin, P11277	βN	1-287
β spectrin, P11277	β1-4	295-747
β spectrin, P11277	β5-9	741-1278
β spectrin, P11277	β10-14	1272-1804
β spectrin, P11277	β15-C	1795-2137
β spectrin, P11277	β10-15	1272-1907
β spectrin, P11277	β10-13	1272-1690
β spectrin, P11277	β11-14	1379-1804
β spectrin, P11277	β12-15	1475-1907
β spectrin, P11277	β9-11	1165-1482
β spectrin, P11277	β10-12	1272-1588
β spectrin, P11277	β11-13	1379-1690
β spectrin, P11277	β12-14	1585-1907

Table A.2: Constructs used in experiments

Protein name & UniProt number	Construct name	Construct amino acid boundaries
Ankyrin domains		
Ankyrin, P16157	D1-4	1-827
Ankyrin, P16157	D3-4	402-811
Ankyrin, P16157	Zu5A	911-1068
Ankyrin, P16157	Zu5B	1071-1233
Ankyrin, P16157	Zu5AB	911-1233
KAHRP fragments		
KAHRP, Q9TY99	K1	38-362
KAHRP, Q9TY99	K2	356-533
KAHRP, Q9TY99	K3	533-654
KAHRP, Q9TY99	K1D	282-346
KAHRP, Q9TY99	K2-2	407-533
KAHRP, Q9TY99	K2-3	424-533
KAHRP, Q9TY99	K2-4	356-470
KAHRP, Q9TY99	K2-5	356-449
KAHRP, Q9TY99	K2-6	356-419
PfEMP1 ATS fragments		
PF08_0141	ATS (for fluorescence)	1-393, Q150C
PF08_0141, Q8IAK2	ATS-N	1-108
PF08_0141, Q8IAK2	ATS-M	137-244
PF08_0141, Q8IAK2	ATS-Core	109-136 and 245-292
PF08_0141, Q8IAK2	ATS-C	293-393
PF08_0141, Q8IAK2	ATS-C 309-392	309-392
PF08_0141, Q8IAK2	ATS-C 309-392 (for PRE)	309-392, D328C
PF08_0141, Q8IAK2	ATS-C 309-392 (for PRE)	309-392, E346C
PF08_0141, Q8IAK2	ATS-C 309-392Δ345-381	309-392Δ345-381
PF08_0141, Q8IAK2	ATS-C 293-393Δ345-365	293-393Δ345-365
PF08_0141, Q8IAK2	ATS-C W310A	293-393
PF08_0141, Q8IAK2	ATS-C I313A	293-393
PF08_0141, Q8IAK2	ATS-C M327A	293-393
PF08_0141, Q8IAK2	ATS-C I330A	293-393
PF08_0141, Q8IAK2	ATS-C L331A	293-393
PF08_0141, Q8IAK2	ATS-C D332A	293-393
PF08_0141, Q8IAK2	ATS-C E335A	293-393
PF08_0141, Q8IAK2	ATS-C D336A	293-393
PF08_0141, Q8IAK2	ATS-C I338A	293-393
PF08_0141, Q8IAK2	ATS-C Y339A	293-393
PF08_0141, Q8IAK2	ATS-C Y340A	293-393
PF08_0141, Q8IAK2	ATS-N-Core	1-136 and 245-292

Table A.3: Constructs used in experiments

Protein name & UniProt number	Construct name	Construct amino acid boundaries
PfEMP1 ATS fragments (continued)		
PFF0845c, C6KT15	ATS (for fluorescence)	1-348, H159C
PFF0845c, C6KT15	ATS-Core PFF0845c	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c K266E	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c R286E	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c H287A	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c N289D	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c M290A	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c W294A	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c K305A	118-145 and 264-313
PFF0845c, C6KT15	ATS-Core PFF0845c E308Q	118-145 and 264-313
PFC1120c, O97312	ATS (for fluorescence)	1-428, Q182C
PFB_1055c, O96296	ATS (for fluorescence)	1-410, Q161C
PFF0010w, C6KSK8	ATS (for fluorescence)	1-402, G156C
PF08_0103, Q8IAS7	ATS (for fluorescence)	1-418, H154C

Appendix B

HADDOCK

B.1 Active residues used for HADDOCK docking:

B.1.1 Chapter 3: PFI1780w/ATS-C

PFI1708w:

133, 136, 138, 139,140,146, 160, 161, 167, 171, 173, 176

PF08_0141, ATS-C:

310, 311, 312, 313, 329, 331, 332, 335, 336, 338, 339, 340

B.1.2 Chapter 4: α17/ATS-Core

PFF0845c, ATS-Core:

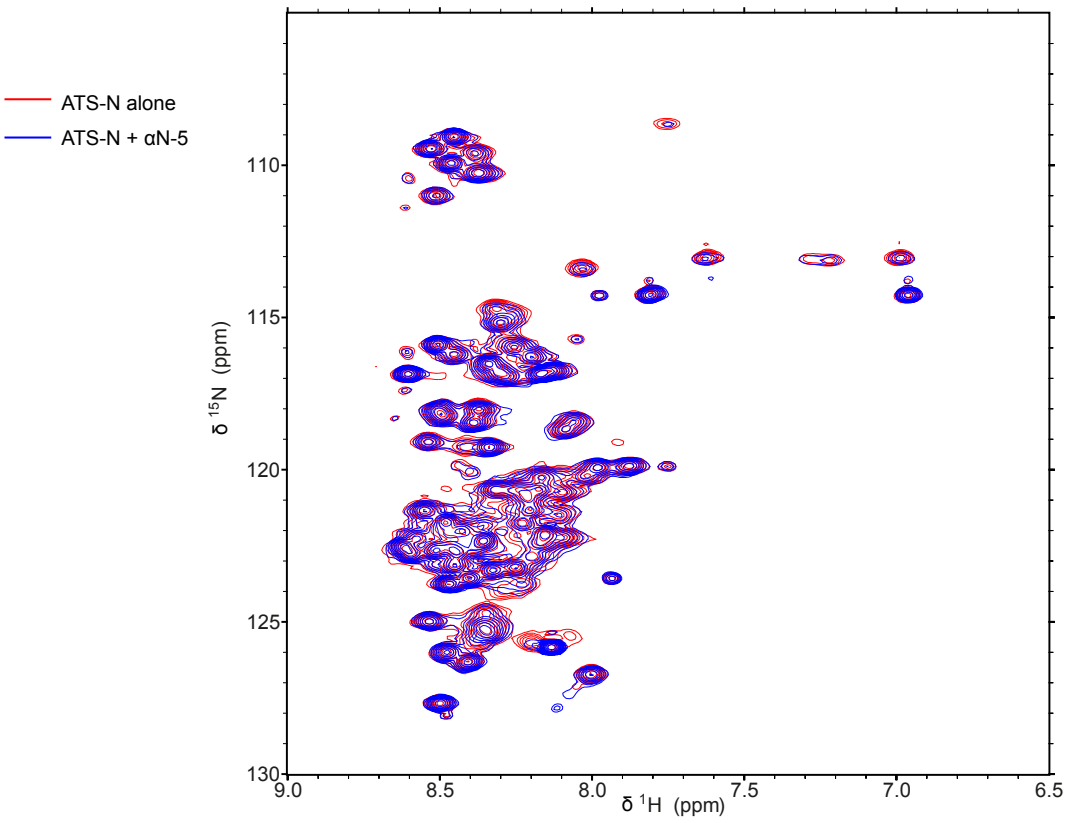
269, 279, 286, 293, 295, 297, 299, 300, 301, 308, 311, 312, 313

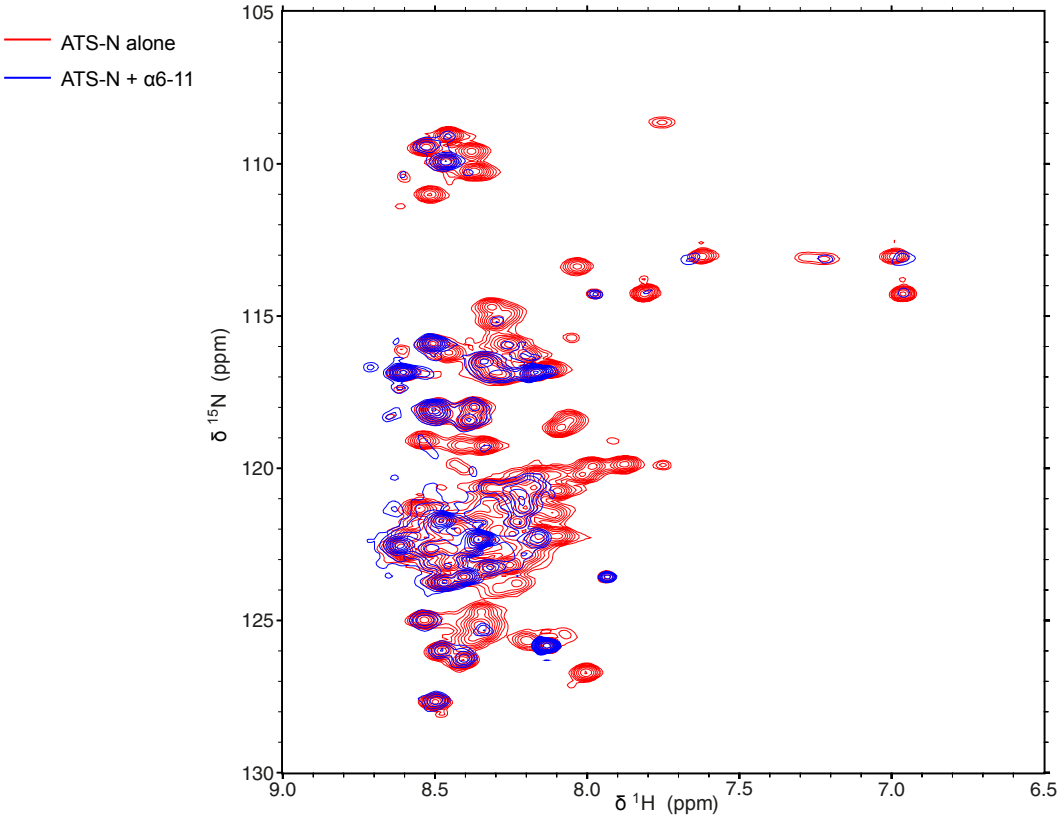
α17 spectrin:

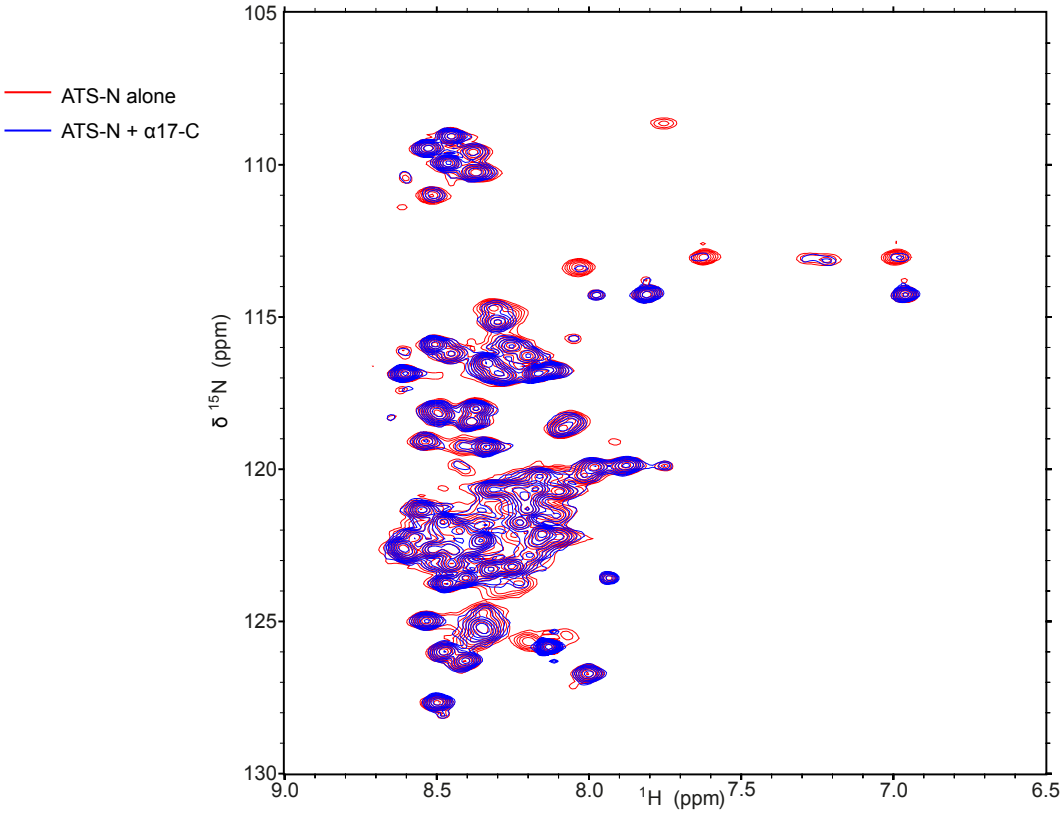
1704, 1705, 1706, 1707, 1708, 1709, 1710, 1712, 1715, 1722, 1723, 1726, 1784, 1787

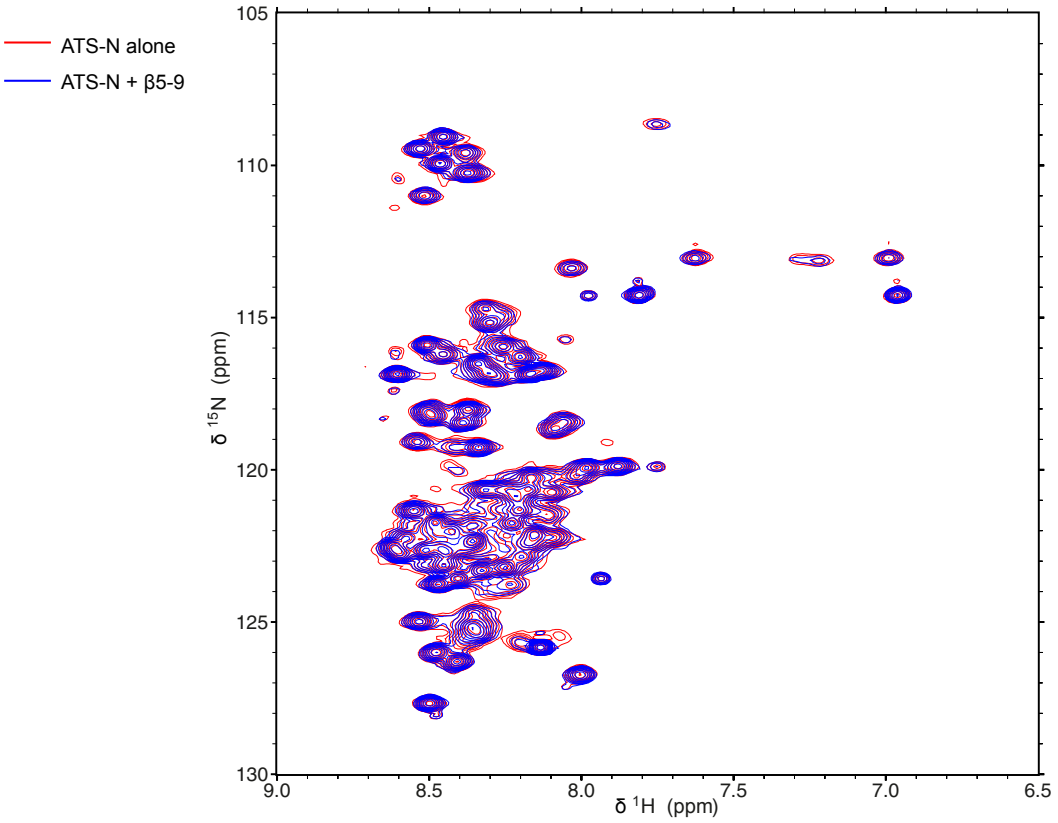
Appendix C

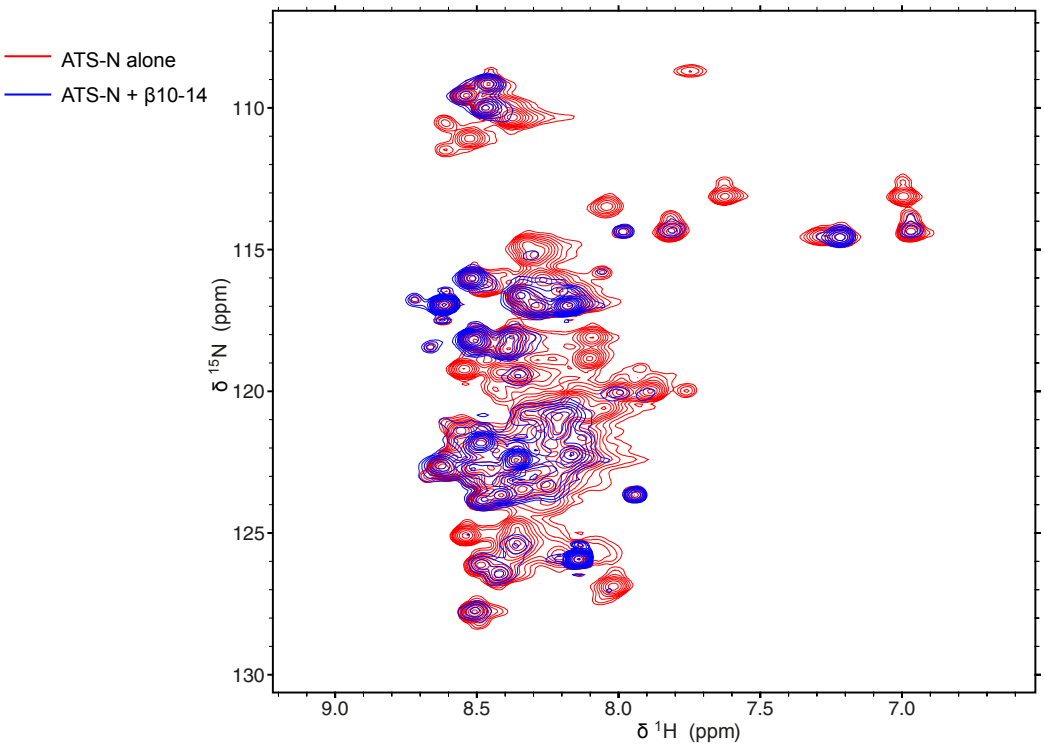
Spectrin NMR Overlays

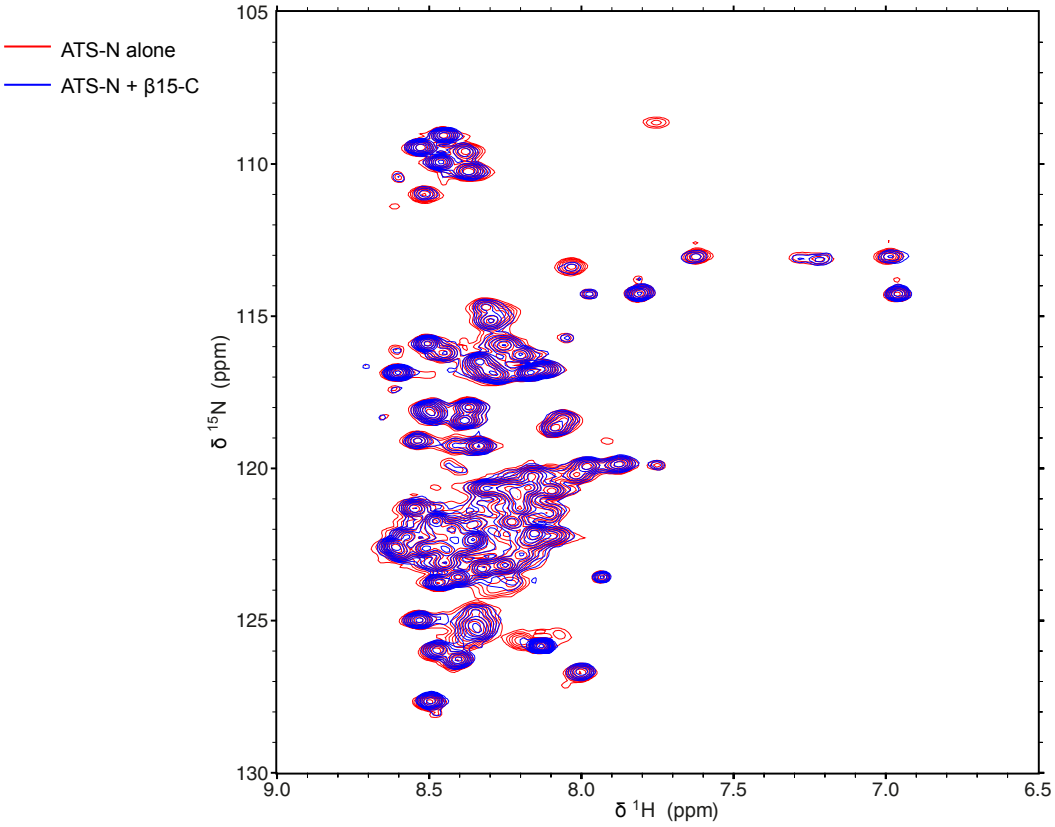

ATS-N alone
ATS-N + αN-5
δ 15N (ppm)
δ 1H (ppm)
110
115
120
125
130
9.0
8.5
8.0
7.5
7.0
6.5

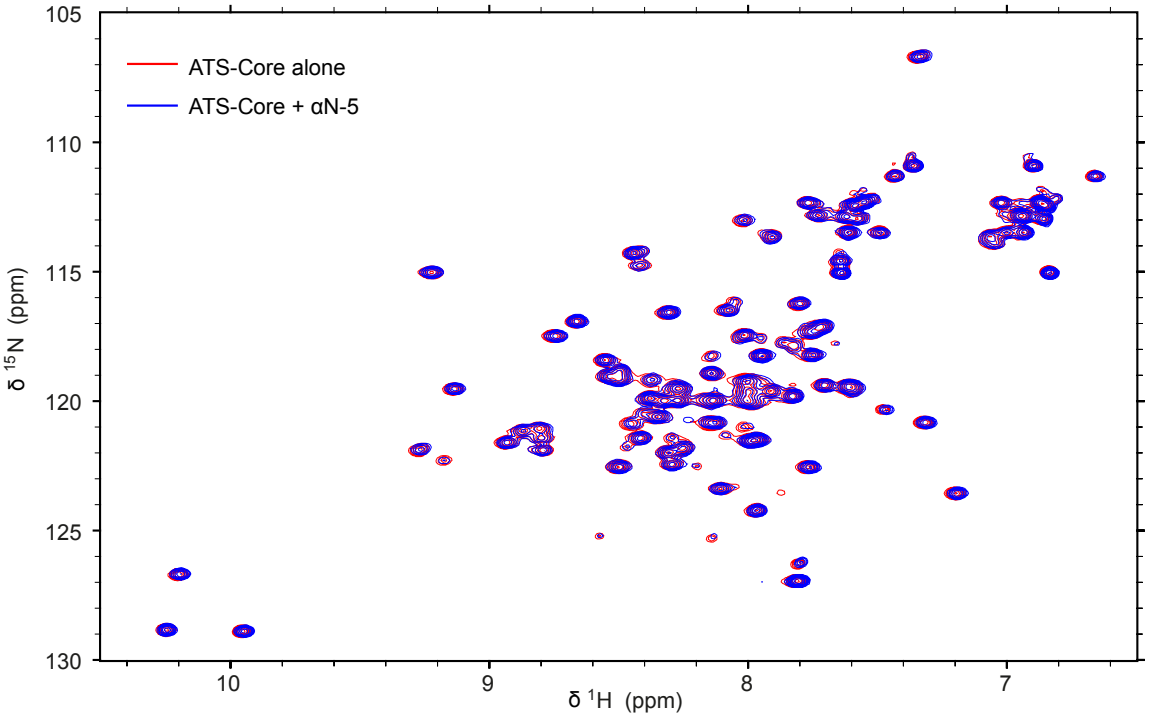

ATS-N alone
ATS-N + α6-11
δ 15N (ppm)
δ 1H (ppm)
105
110
115
120
125
130
9.0
8.5
8.0
7.5
7.0
6.5

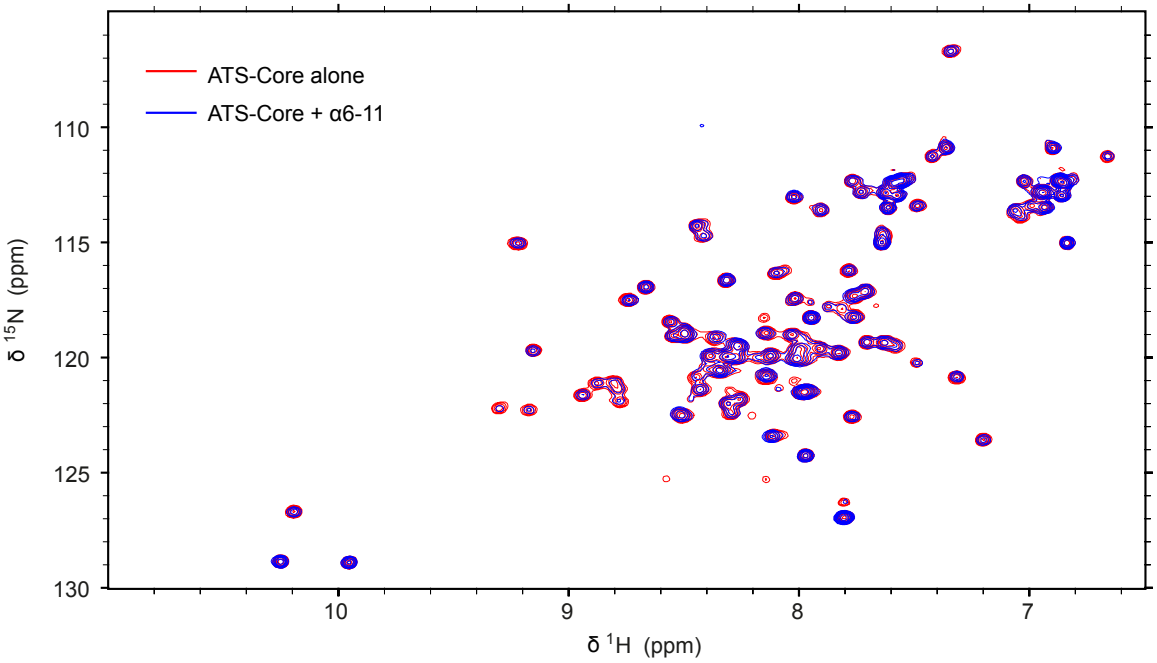

ATS-N alone
ATS-N + α17-C
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
1H (ppm)
7.5
7.0
6.5

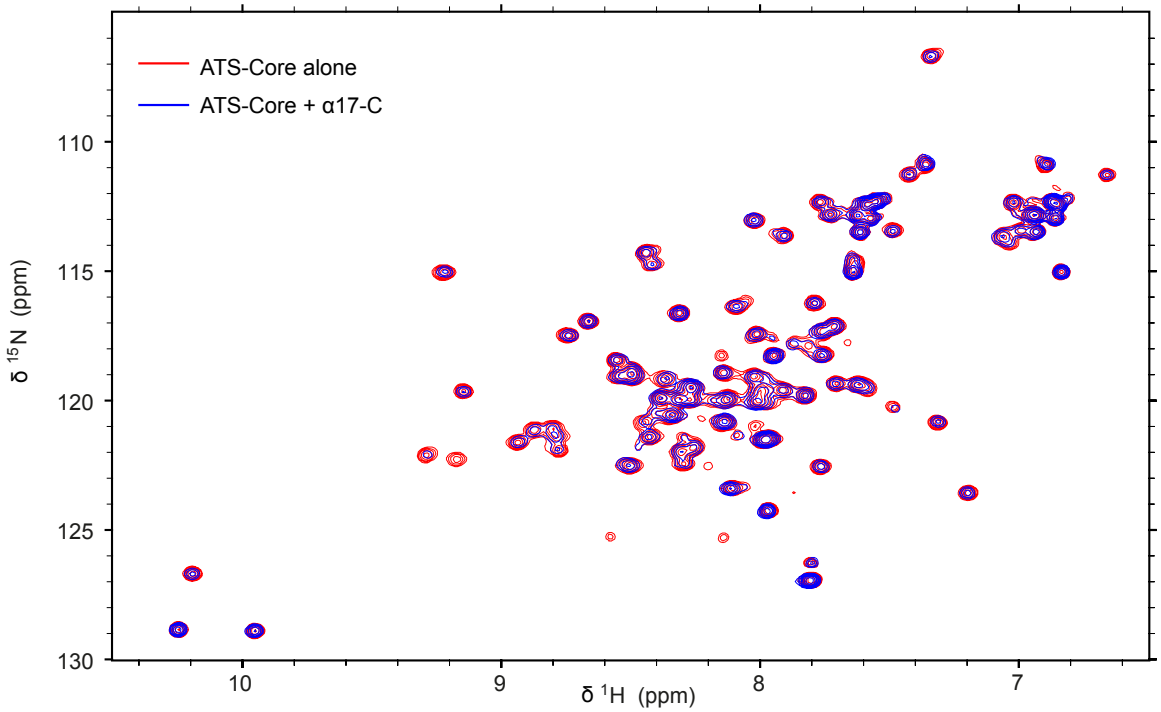

ATS-N alone
ATS-N + β5-9
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
δ 1H (ppm)
7.5
7.0
6.5

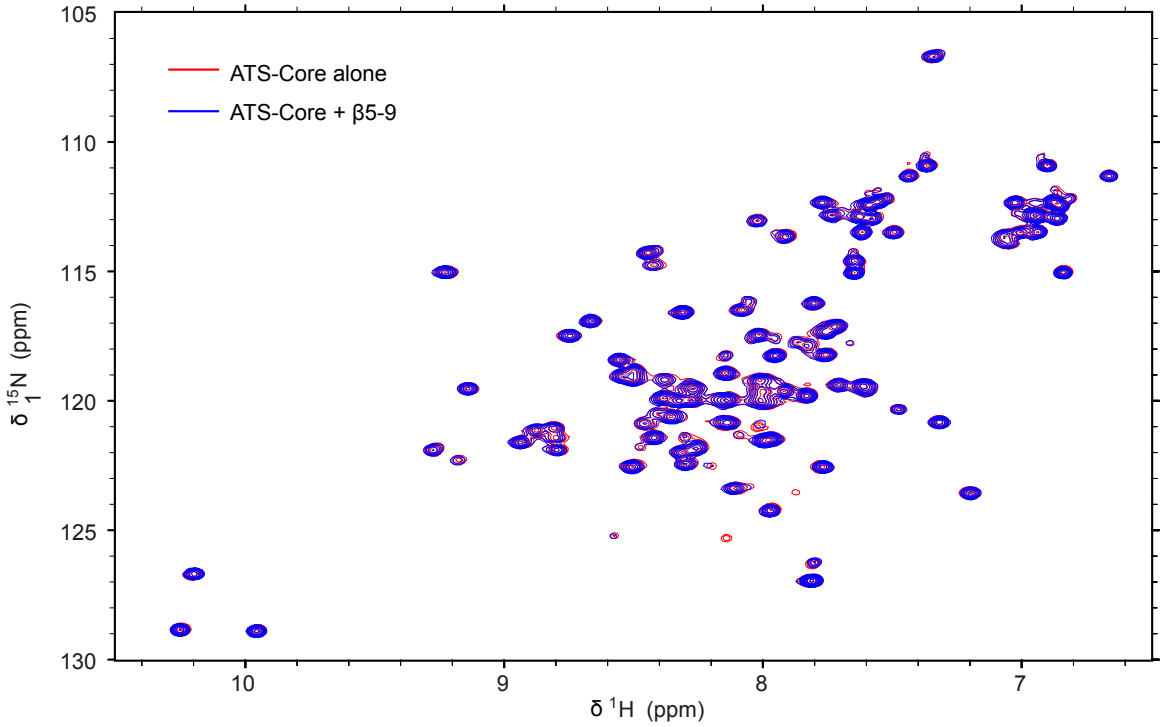

ATS-N alone
ATS-N + β10-14
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
δ 1H (ppm)

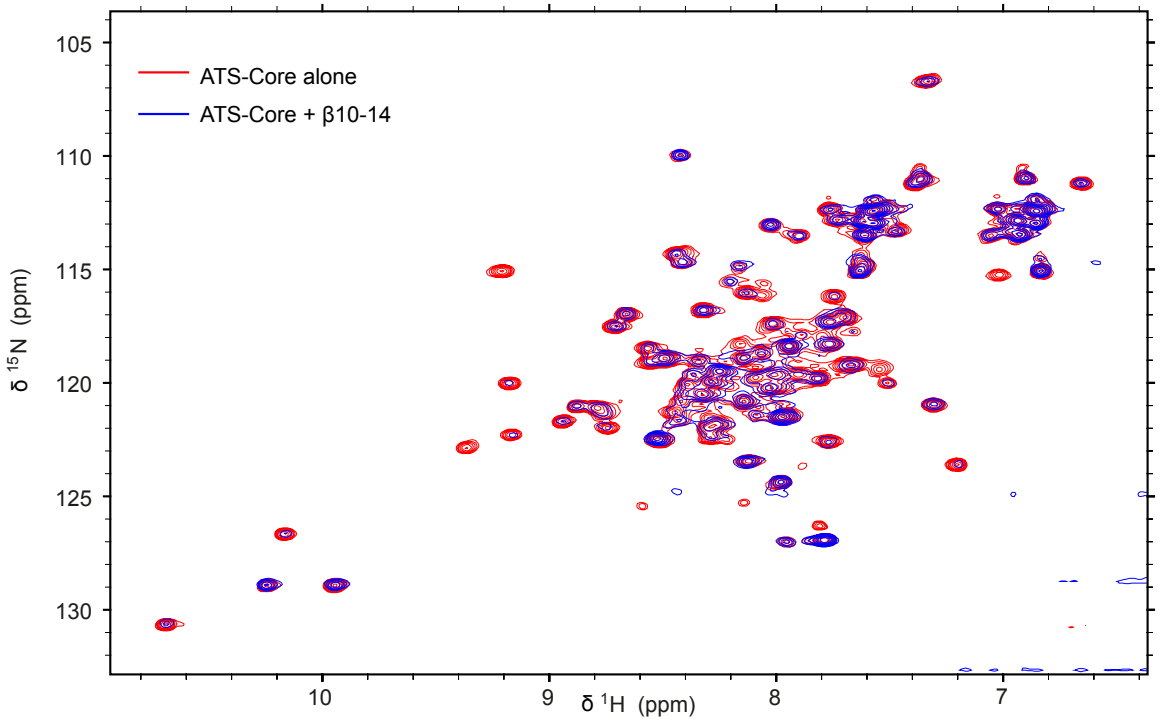

ATS-N alone
ATS-N + β15-C
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
δ 1H (ppm)

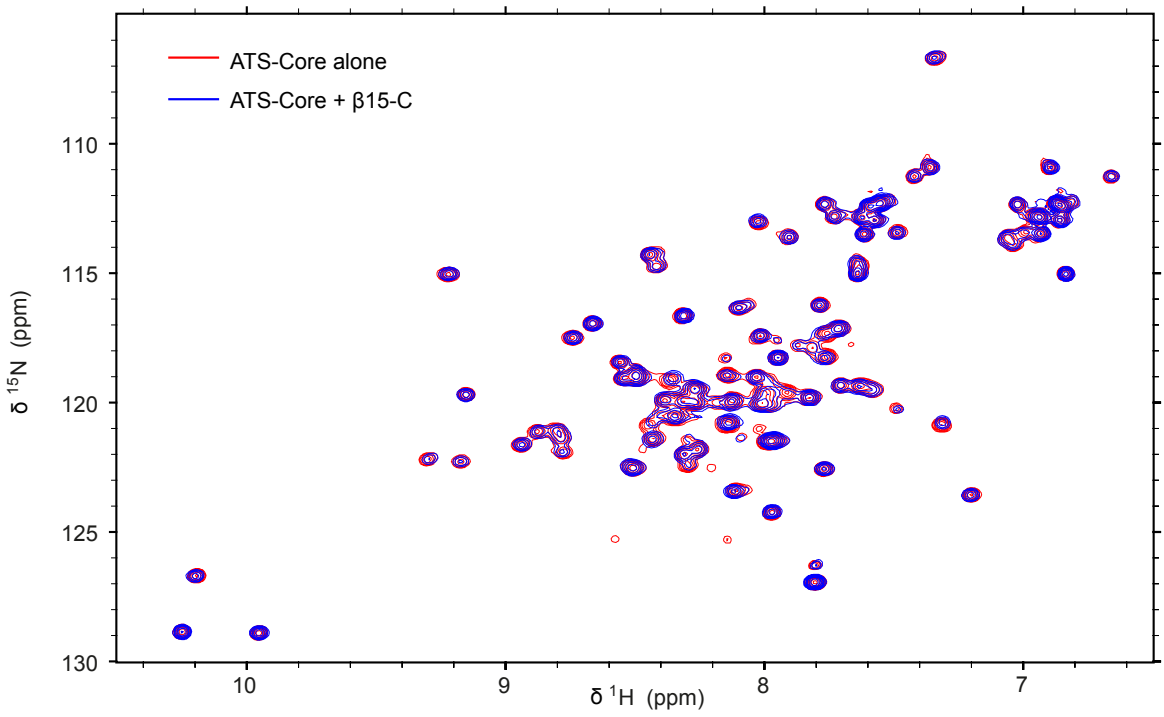

ATS-Core alone
ATS-Core + αN-5
δ 15N (ppm)
δ 1H (ppm)
105
110
115
120
125
130
10
9
8
7

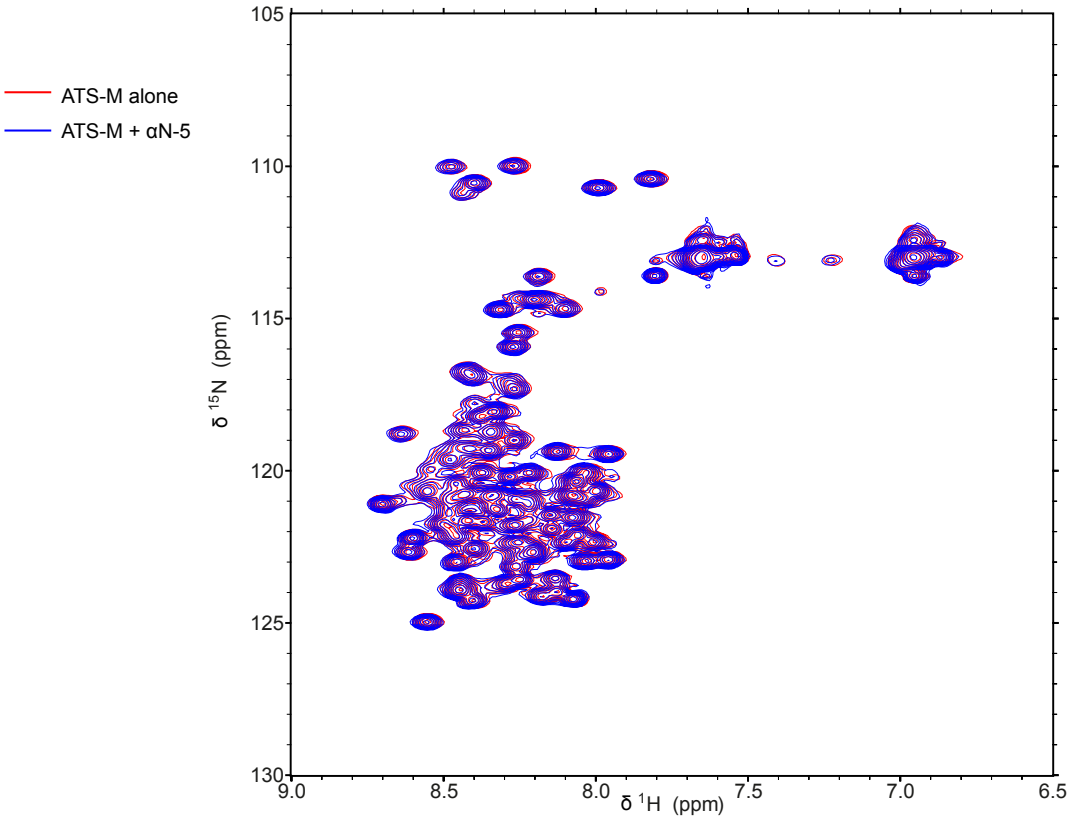

ATS-Core alone
ATS-Core + α6-11
δ 15N (ppm)
δ 1H (ppm)
110
115
120
125
130
10
9
8
7

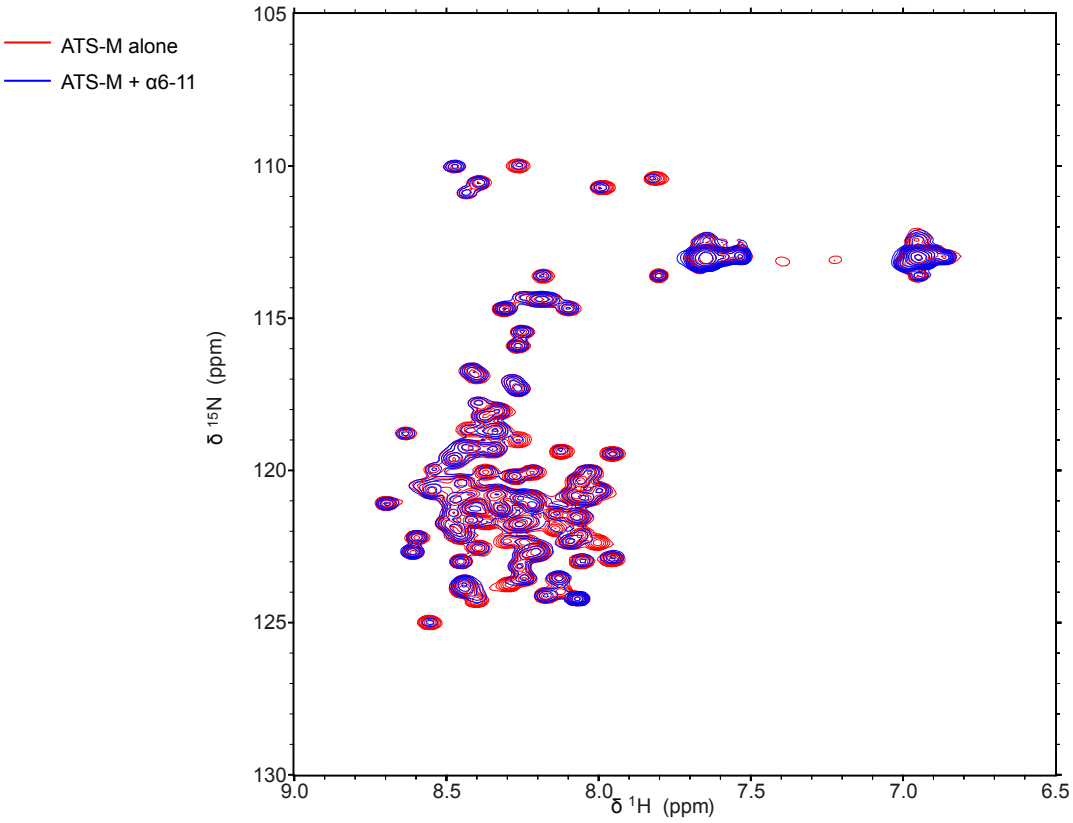

ATS-Core alone
ATS-Core + α17-C
δ 15N (ppm)
110
115
120
125
130
10
9
8
7
δ 1H (ppm)

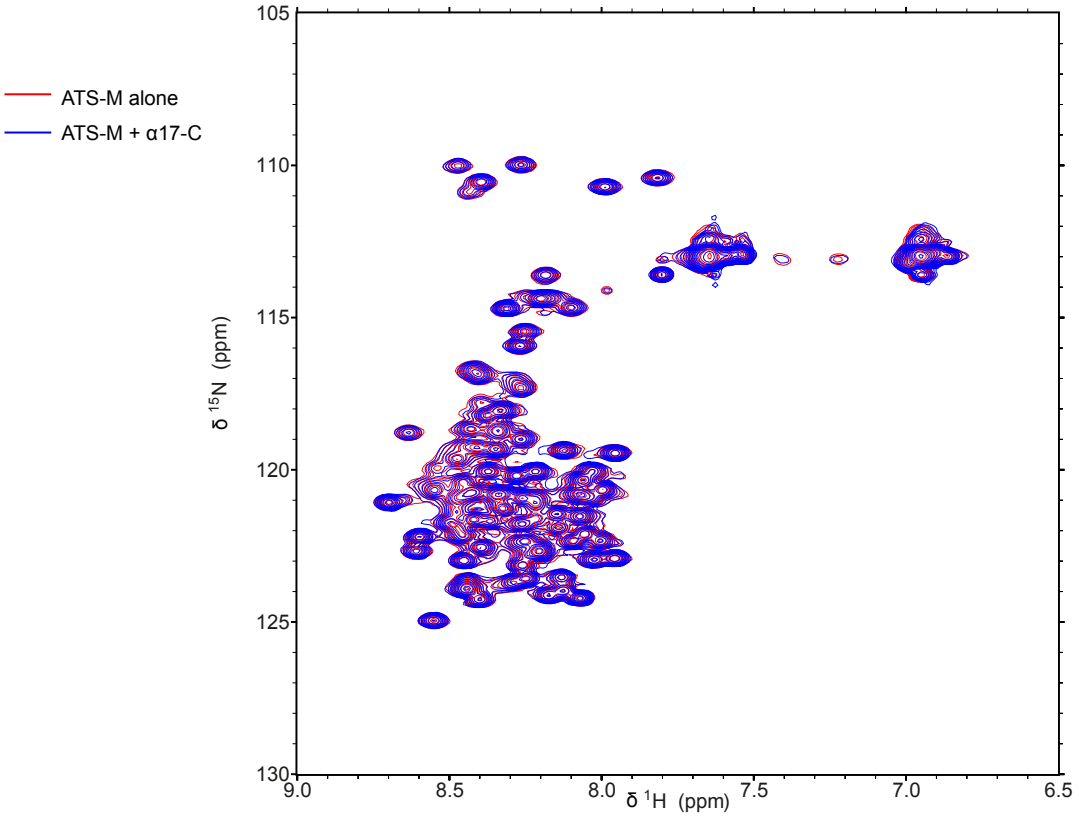

ATS-Core alone
ATS-Core + β5-9
δ 15N (ppm)
105
110
115
120
125
130
10
9
8
7
δ 1H (ppm)

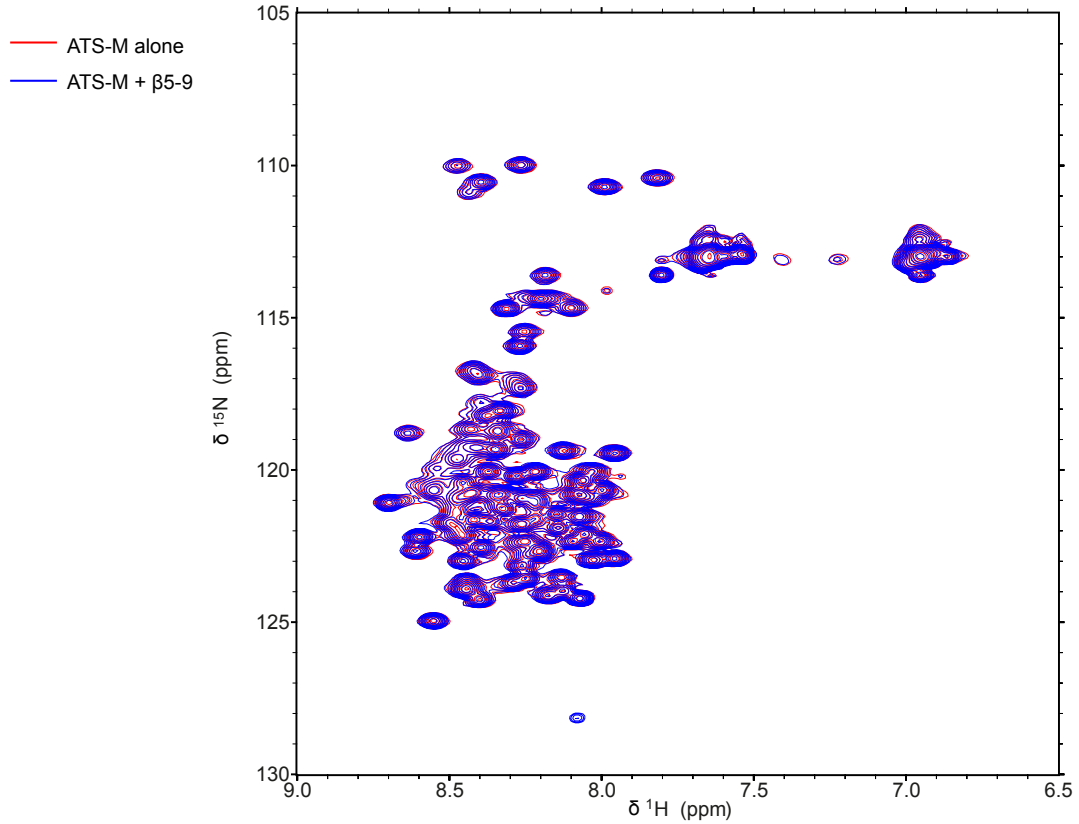

ATS-Core alone
ATS-Core + β10-14
δ 15N (ppm)
δ 1H (ppm)
105
110
115
120
125
130
10
9
8
7

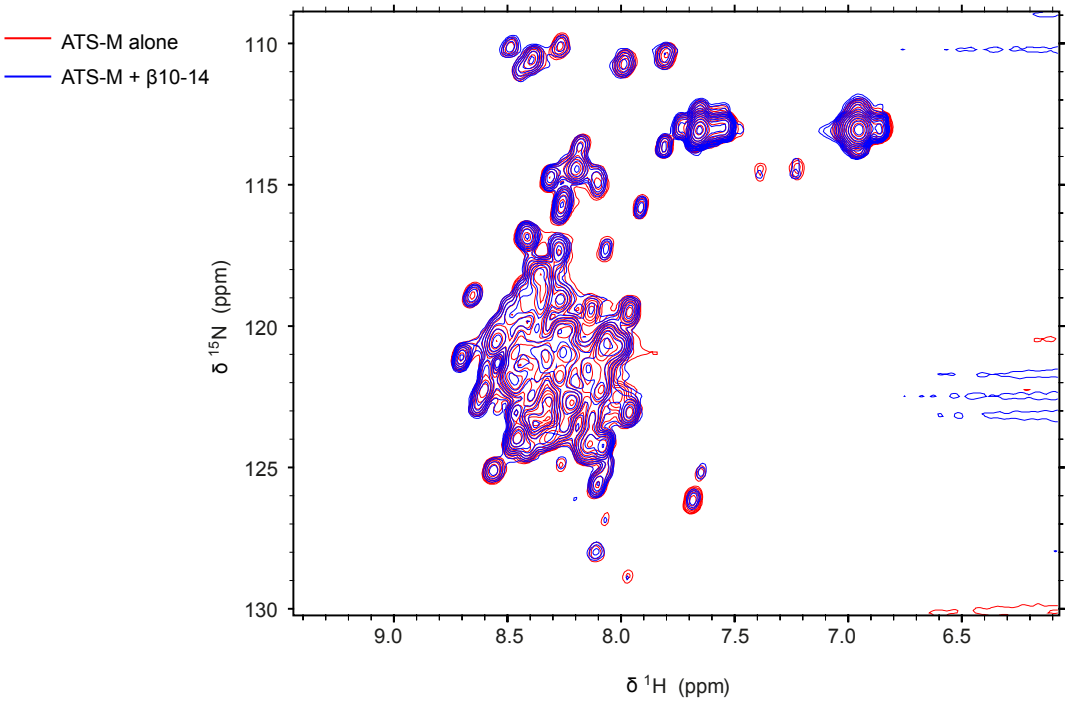

ATS-Core alone
ATS-Core + β15-C
δ 15N (ppm)
δ 1H (ppm)
110
115
120
125
130
10
9
8
7

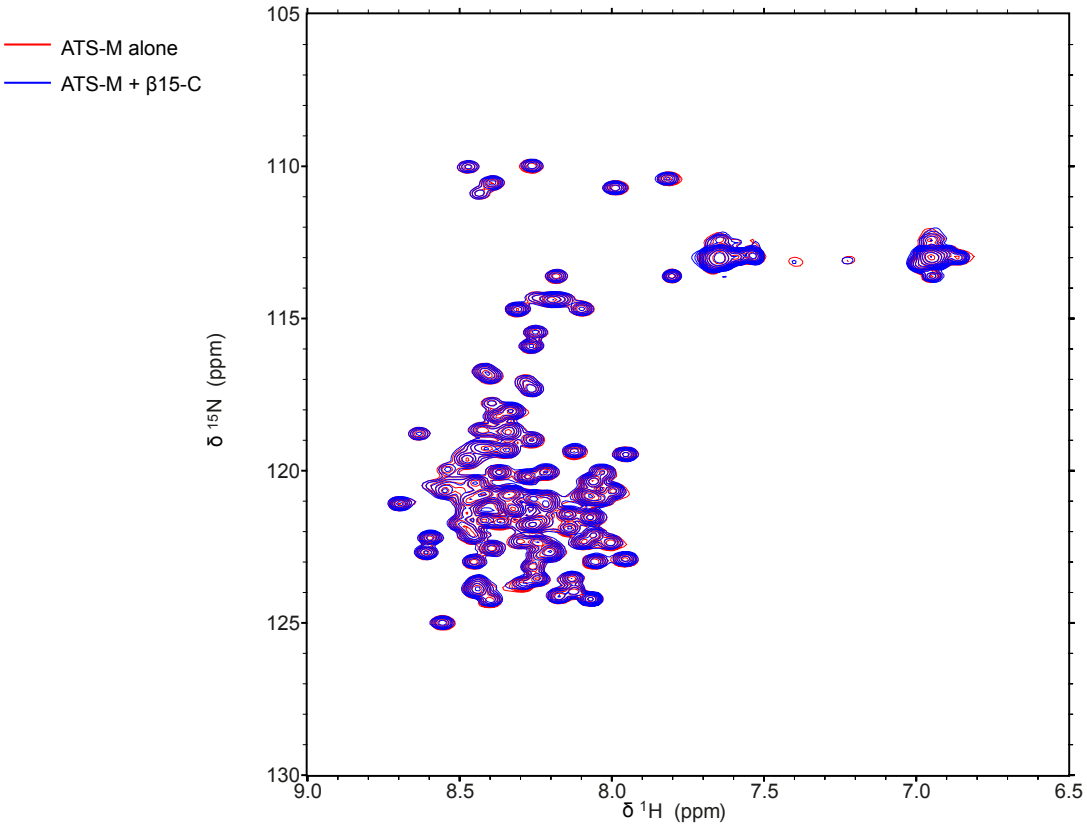

ATS-M alone
ATS-M + αN-5
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
δ 1H (ppm)

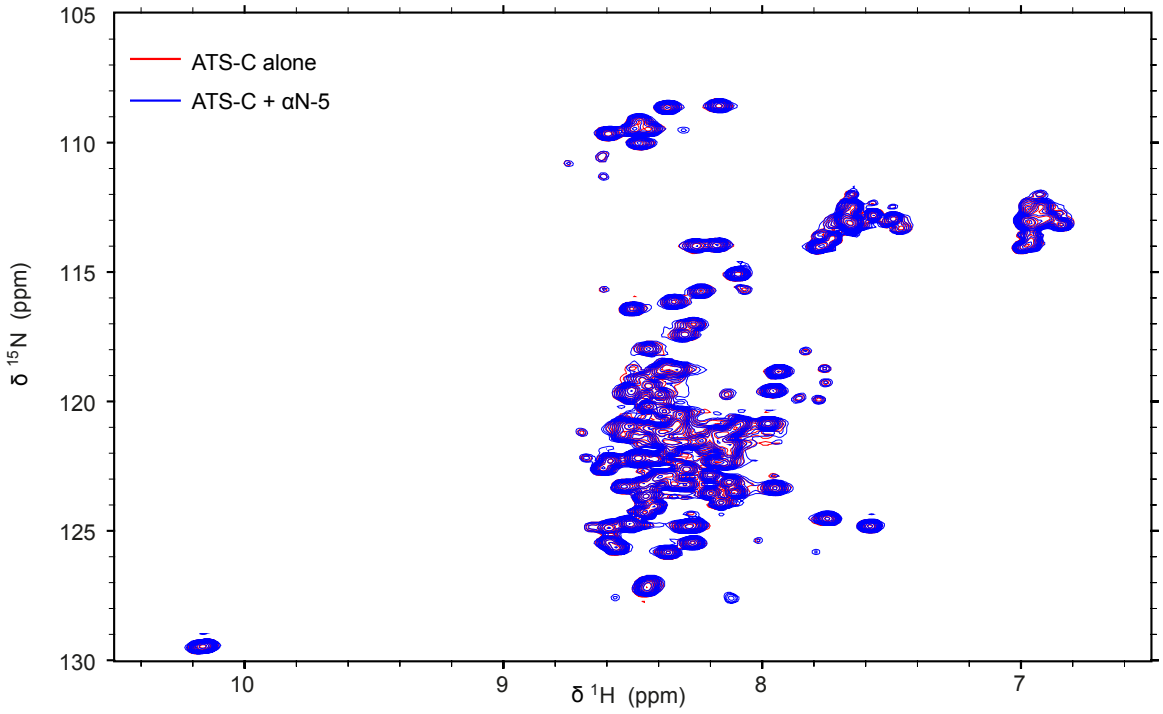

ATS-M alone
ATS-M + α6-11
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
δ 1H (ppm)

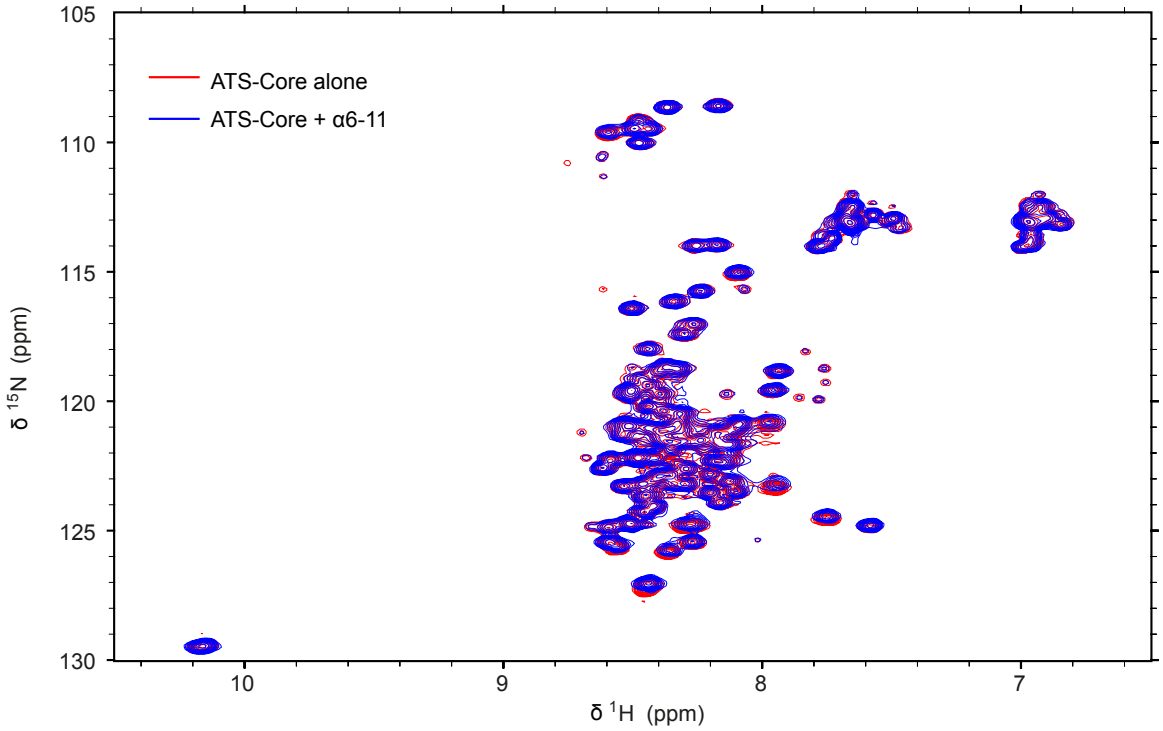

ATS-M alone
ATS-M + α17-C
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
δ 1H (ppm)

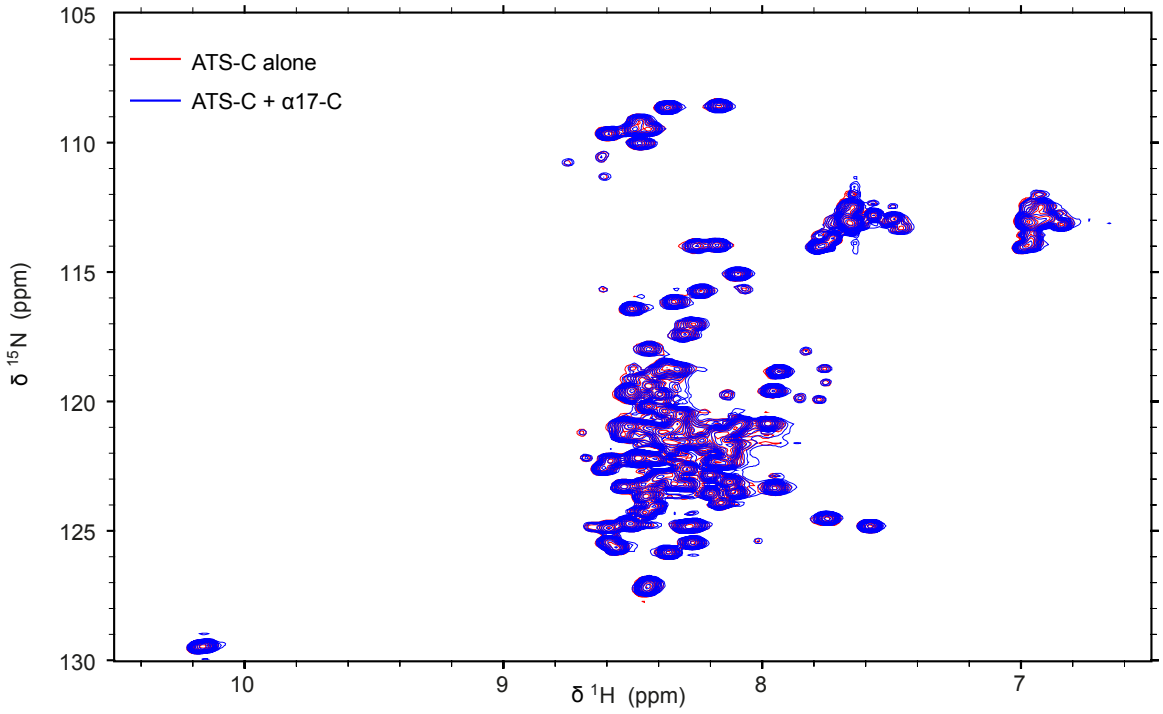

ATS-M alone
ATS-M + β5-9
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
δ 1H (ppm)

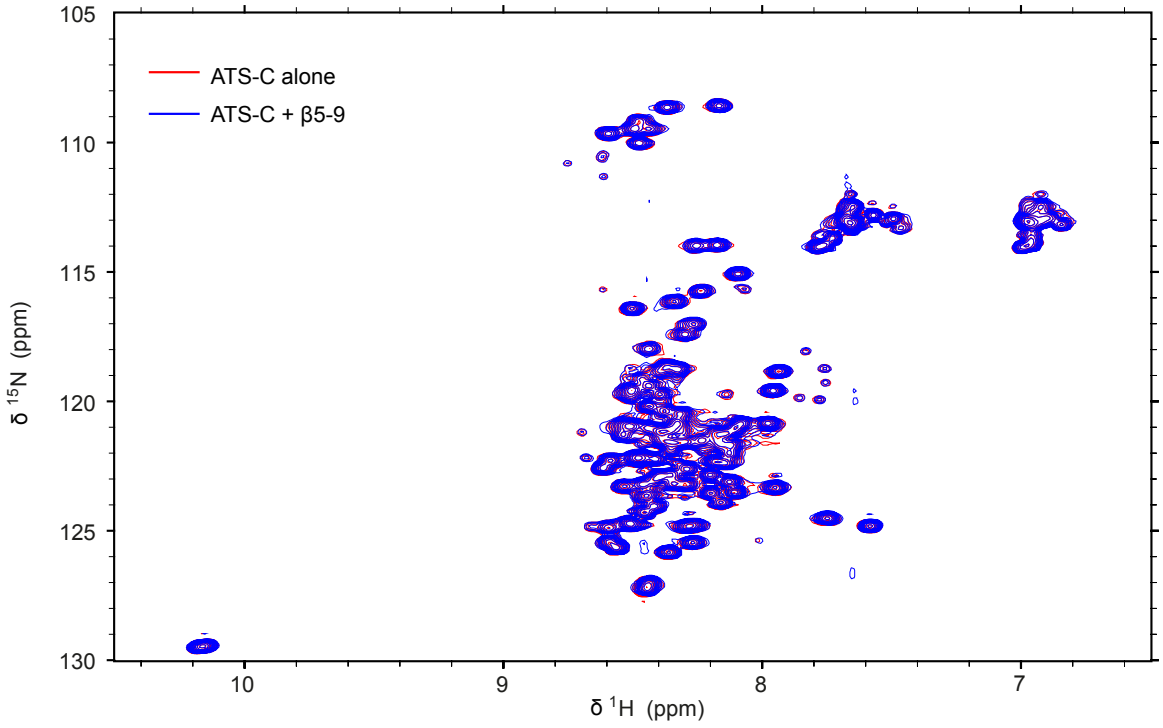

ATS-M alone
ATS-M + β10-14
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
δ 1H (ppm)

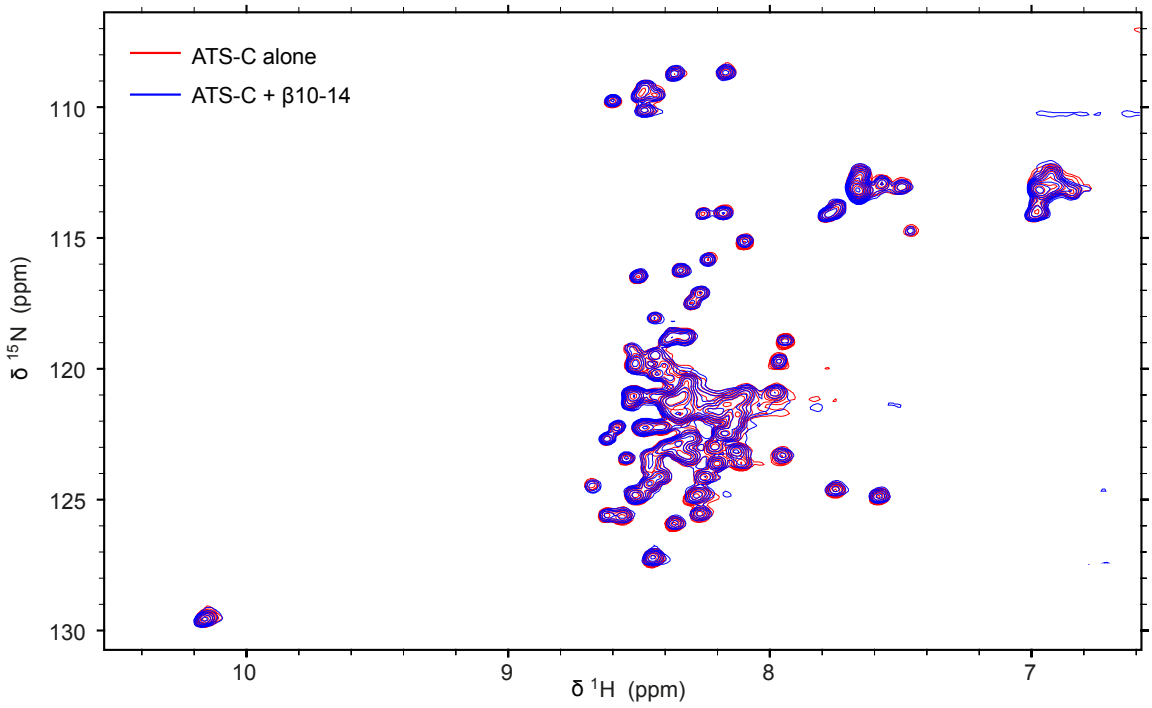

ATS-M alone
ATS-M + β15-C
105
110
115
120
125
130
δ 15N (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
δ 1H (ppm)

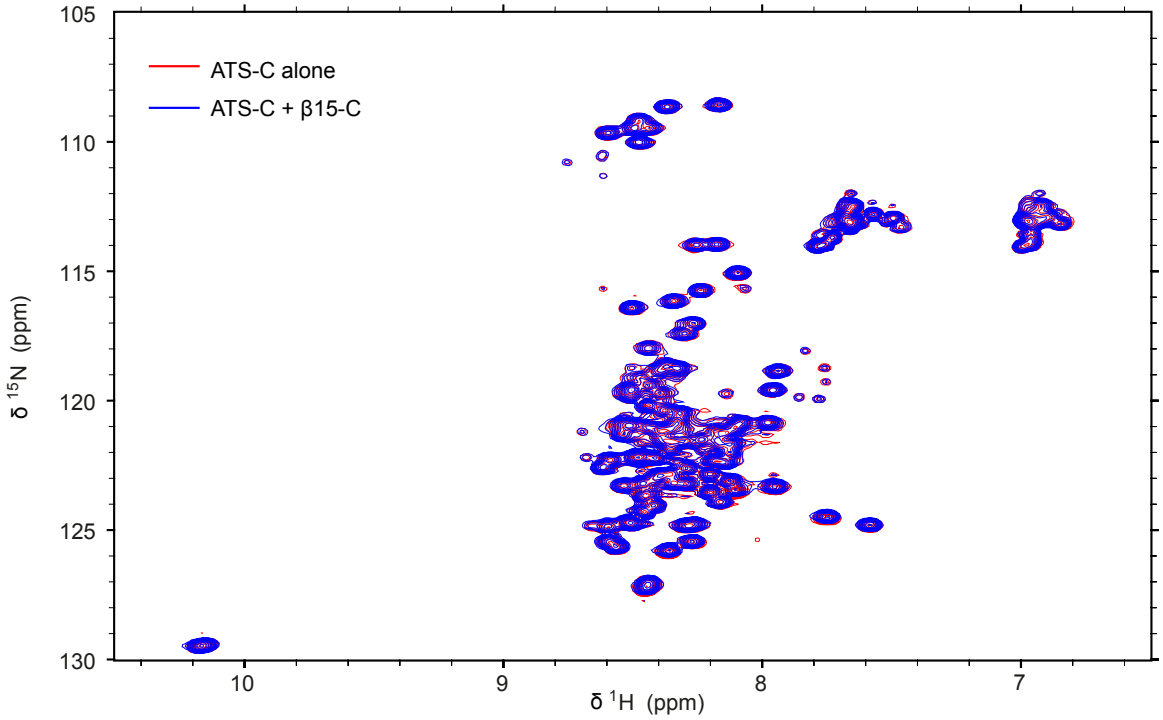

ATS-C alone
ATS-C + αN-5
δ 15N (ppm)
δ 1H (ppm)
105
110
115
120
125
130
10
9
8
7


ATS-Core alone
ATS-Core + α6-11
δ 15N (ppm)
δ 1H (ppm)
105
110
115
120
125
130
10
9
8
7


ATS-C alone
ATS-C + α17-C
δ 15N (ppm)
δ 1H (ppm)
105
110
115
120
125
130
10
9
8
7


ATS-C alone
ATS-C + β5-9
δ 15N (ppm)
δ 1H (ppm)
105
110
115
120
125
130
10
9
8
7


ATS-C alone
ATS-C + β10-14
δ ^{15}N (ppm)
δ ^{1}H (ppm)
110
115
120
125
130
10
9
8
7


ATS-C alone
ATS-C + β15-C
δ ^{15}N (ppm)
δ ^{1}H (ppm)
105
110
115
120
125
130
10
9
8
7

Appendix D

Thermostability by CD

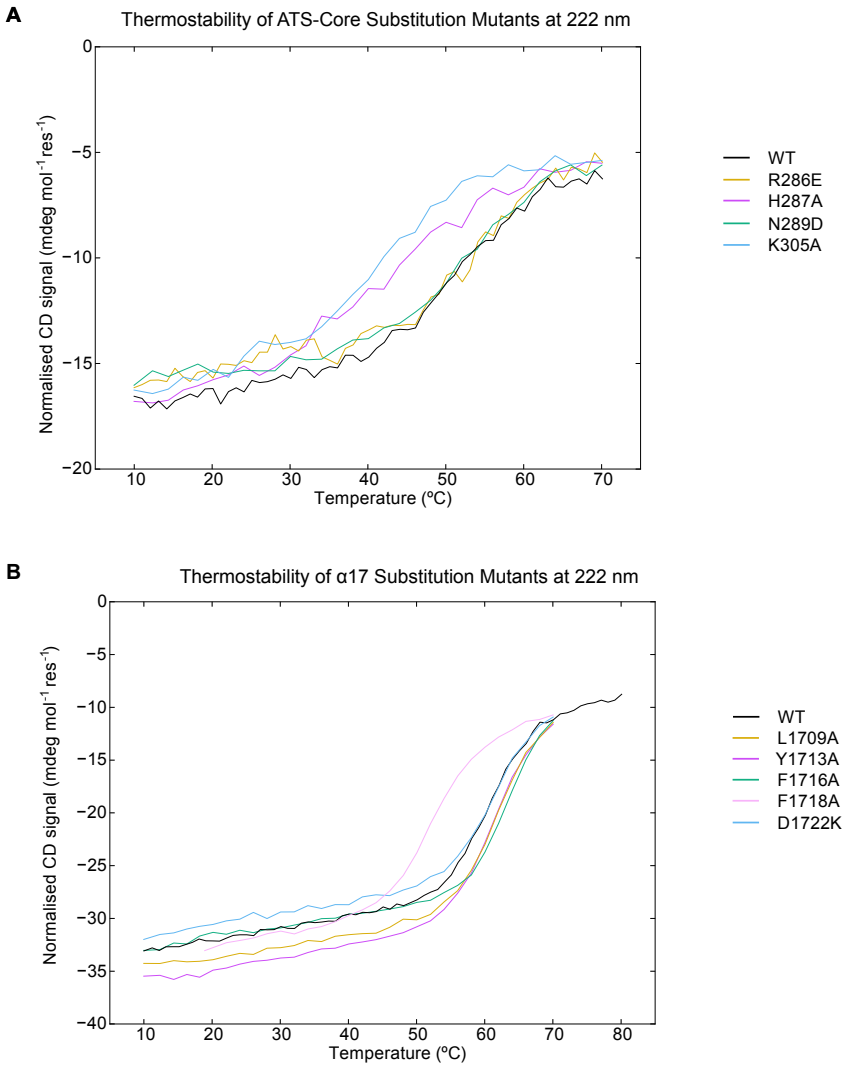


Fig. D.1: Thermostability of ATS-Core and α17 subsitution mutants

Appendix E

KAHRP docking Scripts

The docking script was composed of three parts; path finding, scoring and protein coordinate finding.

Path finding: all_paths_final_b1014.py

Inputs: The surface: out.obj, Positive and negative coordinates: pos_filtered.txt, neg_filtered.txt and the electrostatic potential file: out.dx.

Outputs: Distance along path and potential by dumping text (paths.txt).

Scoring: KAHRP_score_b1014.py

Input: Takes in paths.txt that has been transposed, and zero filled so each column is the same length, KAHRP charge versus distance and shift along KAHRP chain, currently set to every 15.2Å.

Outputs: Best score, which path and shift.

Coordinate finding: coords_search.py

Input: Takes the coordinates of specific path, the beta1014.pdb file and KAHRP charge versus distance.

Output:β10-14 protein atoms corresponding the specfic path.

```
#! /usr/bin/env python
import numpy as np
import networkx
import gridData
from scipy.spatial.distance import euclidean

INPUT1 = open("out.obj",'r') #surface mesh
INPUT2 =open("pos_filtered.txt",'r') #positive clusters
INPUT3 =open("neg_filtered.txt",'r') #negative clusters
OUTPUT1= open("cartesian_points.txt",'w') #output the points of each path

#read surface mesh

# define an empty list of nodes
nodes = []
triangles = []
pos=[]
neg=[]
distance=[]
while True:
    # read the next line
    line = INPUT1.readline()
    if not line:
        break
    # chop the line up based on whitespace
    cols = line.split()
    if cols:
        # look to see if it is vertex
        if cols[0] == "v":
            # append the (x,y,z) to the list
            nodes.append((float(cols[1]),float(cols[2]),float(cols[3])))
        # ..or if it is a triangle line
        elif cols[0] == "f":
            vertex = []
            # assuming the triangles are always index//index i.e. the same
                both sides
            for i in [1,2,3]:
                v = cols[i].split('//')
                vertex.append(int(v[0])-1) # subtract one as looks like
                    Chimera is 1-based, not 0-based!
            triangles.append(vertex)

#Make this a graph
G  = networkx.Graph()
S  = networkx.DiGraph()
S.add_node(0,{"coordinates": (0,0,0)})
S.add_node(10000000,{"coordinates":(0,0,0)})
# loop through all the nodes
for node,coords in enumerate(nodes):
    # add each node to the graph with its coordinates as an attribute
    G.add_node(node,{"coordinates":coords})
# now loop through all the triangles
for sides in triangles:
    # and add all the edges, weighted by the length of each side
    G.add_edge(sides[0],sides[1],{"length": euclidean(nodes[sides[0]],nodes
        [sides[1]])})
    G.add_edge(sides[1],sides[2],{"length": euclidean(nodes[sides[1]],nodes
        [sides[2]])})
    G.add_edge(sides[2],sides[0],{"length": euclidean(nodes[sides[2]],nodes
        [sides[0]])})
```

```
#Find graph node closest to charge cluster point
indpos=[]
while True:
    line = INPUT2.readline()
    if not line:
        break
    col = line.split()
    if col:
        pos.append((float(col[0]),float(col[1]),float(col[2])))
#Determine grid point closest to that charged resiude
for y in xrange(len(pos)):
    for x in xrange(len(nodes)):
        distance.append(euclidean(nodes[x],pos[y]))
    indpos.append(np.argmin(distance))
    distance=[]

indneg=[]
while True:
    line = INPUT3.readline()
    if not line:
        break
    col = line.split()
    if col:
        neg.append((float(col[0]),float(col[1]),float(col[2])))
for y in xrange(len(neg)):
    for x in xrange(len(nodes)):
        distance.append(euclidean(nodes[x],neg[y]))
    indneg.append(np.argmin(distance))
    distance=[]

Nterm=[]
path=[]
subnodes=[]
#Define a N-terminal node of the protein
Nterm.append((333.140991, -414.481995, -49.715000))

#Find all paths from positive to negative nodes that are greater then 10, but
    less than 30 Ang and go from N to C.
#Add these to a new direction graph
for p1 in xrange(len(pos)):
    for n1 in xrange(len(neg)):
        a1=euclidean(Nterm[0],nodes[indpos[p1]])
        b1=euclidean(Nterm[0],nodes[indneg[n1]])
        c1=euclidean(nodes[indpos[p1]],nodes[indneg[n1]])
        if a1 < b1 and c1 > 10 and c1 < 30:
            try:
                path.append(networkx.shortest_path(G,source=indpos[p1],target=
                    indneg[n1], weight="length"))
                path=np.transpose(path)
                for r1 in path:
                    subnodes.append((nodes[r1][0],nodes[r1][1],nodes[r1][2]))
                for node,coords in enumerate(subnodes):
                    S.add_node(int(path[node]),{"coordinates":coords})
                for e1 in xrange(len(path)-1):
                    S.add_edge(int(path[e1]),int(path[e1+1]),{"length":
                        euclidean(subnodes[e1],subnodes[e1+1])})
                path=[]
                subnodes=[]
            except:
```

```
                pass
    #As above, but negative to positive.
    for n2 in xrange(len(neg)):
        for p2 in xrange(len(pos)):
            a2=euclidean(Nterm[0],nodes[indneg[n2]])
            b2=euclidean(Nterm[0],nodes[indpos[p2]])
            c2=euclidean(nodes[indpos[p2]],nodes[indneg[n2]])
            if a2 < b2 and c2 > 10 and c2 < 30:
                try:
                    path.append(networkx.shortest_path(G,source=indneg[n2],target=
                        indpos[p2], weight="length"))
                    path=np.transpose(path)
                    for r2 in path:
                        subnodes.append((nodes[r2][0],nodes[r2][1],nodes[r2][2]))
                    for node,coords in enumerate(subnodes):
                        S.add_node(int(path[node]),{"coordinates":coords})
                    for e2 in xrange(len(path)-1):
                        S.add_edge(int(path[e2]),int(path[e2+1]),{"length":
                            euclidean(subnodes[e2],subnodes[e2+1])})
                    path=[]
                    subnodes=[]
                except:
                    pass

    #Add an N and C terminal node into the directional graph
    for n3 in xrange(len(neg)):
        c3=euclidean(Nterm[0],nodes[indneg[n3]])
        if c3 < 30:
            S.add_edge(0,int(indneg[n3]),{"length": 100000})
        elif c3 > 210:
            S.add_edge(int(indneg[n3]),10000000,{"length": 100000})

    for p3 in xrange(len(pos)):
        c4=euclidean(Nterm[0],nodes[indpos[p3]])
        if c4 < 30:
            S.add_edge(0,int(indpos[p3]),{"length": 100000})
        elif c4 > 210:
            S.add_edge(int(indpos[p3]),10000000,{"length": 100000})
    cartesian=[]
    #load in potential data
    potential = gridData.Grid("out.dx")

    #Find all routes from N to C on the directional graph only containing paths
        that meet the criteria
    for route in networkx.all_simple_paths(S, source=0, target=10000000):
        length=[]
        length.append((0))
        length.append((0))
        route=np.transpose(route)
        pot=[]
        pot.append((0))
        pot.append((potential.interpolated(nodes[route[1]][0],nodes[route[1]][1],
            nodes[route[1]][2])[0]))
        if len(route) > 400:
            for each in xrange(len(route)-3):
                #print each
                length.append((length[each+1]+euclidean(nodes[route[each+1]],nodes
                    [route[each+2]])))
                pot.append((potential.interpolated(nodes[route[each+2]][0],nodes
                    [route[each+2]][1],nodes[route[each+2]][2])[0]))
```

```
                cartesian.append((nodes[route[each+2]][0],nodes[route[each+2]][1],
                    nodes[route[each+2]][2]))
            route=[]
            #output the length, the potential and the coordinates
            print length
            print pot
            print >>OUTPUT1, cartesian
            cartesian=[]
            length=[]
            pot=[]
```

coords_search.py 02/09/2016 10:29

```
import pylab
import matplotlib as plt
import numpy as np
from matplotlib import rc, rcParams
from matplotlib.pyplot import *
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import math
from scipy.spatial.distance import euclidean

INPUT1 = open("458.txt",'r') #the lowest energy path
INPUT2 = open("beta1014.pdb",'r') #the structure provides a list of atomic
    positions
INPUT3 = open("KAHRP4.txt",'r') #provides KAHRP path spacing
point=[]

while True:
    line = INPUT1.readline()
    if not line:
        break
    col = line.split()
    if col:
        point.append((float(col[0]),float(col[1]),float(col[2])))
pdb=[]
while True:
    line = INPUT2.readline()
    if not line:
        break
    col = line.split()
    if col:
        pdb.append((col[3],col[2],int(col[4]),float(col[5]),float(col[6]),
            float(col[7])))
ka=[]
while True:
    line = INPUT3.readline()
    if not line:
        break
    col = line.split()
    if col:
        ka.append((col[0]))

y = 1
score=0
shift=[0, 15.2, 30.4, 45.6, 60.8, 76, 91.2, 106.4, 121.6, 136.8, 152, 167.2,
    182.4, 197.6, 212.8, 228, 243.2]
kx = np.genfromtxt('KAHRP4.txt', usecols=(1))
ky = np.genfromtxt('KAHRP4.txt', usecols=(2))

kcoords=[]
newx0=[]
kpdist=[]
newy0=[]
kdist=[]

i= 9 #shift index goes here
x= 458 #path number goes here
x0 = np.genfromtxt('path.txt', usecols=(2*x))
y0 = np.genfromtxt('path.txt', usecols=(2*x+1))

for k in xrange(len(kx)): #for each point in the kahrp path
```

```
        for each in xrange(len(x0)): #for each point in the spectrin path
            newx0.append((x0[each]+shift[i])) #update the spectrin path to match
                the kahrp path
            if x0[each]==0: #if the spectrin point is zero, ignore it, as the
                paths are zero filled to ensure same length
                pass
            else: #but if the spectrin distance is not zero, find the distance
                between that and the closest Kahrp point
                kdist.append((euclidean(newx0[each],kx[k])))
        m=np.argmin(kdist)
        if m<492: #this is the number of non-zero enteries of the spectrin path
            kcoords.append((point[np.argmin(kdist)])) #print out the coordinates
                along the path of that spectrin distance
            kdist=[]
        else:
            pass
    for c in xrange(len(kcoords)):
        for p in xrange(len(pdb)):
            kpdist.append((euclidean(kcoords[c],(pdb[p][3],pdb[p][4],pdb[p][5]))))
                #Find the closes atom in the pdb to that point on the spectrin
                coordinate
        n=np.argmin(kpdist)
        kpdist=[]
        print ka[c],kx[c],ky[c],kcoords[c],pdb[n][0],pdb[n][1],pdb[n][2],pdb[n][3]
            ,pdb[n][4],pdb[n][5]
```

Appendix F

KAHRP/Spectrin Docking

F.1 β10-14 spectrin charge cluster residues

Positive Residues:	Negative Residues:
1353	1335, 1275, 1342
1311, 1314	1330, 1326
1369, 1379	1292, 1321
1458, 1459	1362, 1363
1472, 1469, 1403	1451, 1447, 1453
1420, 1423, 1422	1437, 1434, 1433
1416, 1478, 1485	1465, 1462, 1399, 1400
1568	1551, 1556, 1557, 1554, 1546
1579, 1576	1488, 1495, 1489, 1496
1519, 1518, 1581	1567, 1571
1663, 1670	1582, 1585
1638, 1641	1594, 1594
1631, 1628, 1625, 1624	1784, 1781, 1710
1683, 1684, 1685	1677, 1680, 1681
1614, 1696, 1697	1604, 1602
1737, 1733, 1729	1617, 1616
	1755, 1759
	1748, 1745, 1698
	1703, 1700, 1770, 1773, 1777

Table F.2: KAHRP K2 Sequences

Name	Score Forward	Score Reverse	Sequence
K2-4	-324	-304	DSKSNNTKEKNKNEENTKKSCPKKGNHKEGKEKVSKKSKVSEADENDKHKKSKKKECDHSKHKKSKVSEADENDKHKKSKKKEGDHDKHKKSKKKKEGEGDNGEHKKSSHHKED
K2-5	-293	-304	DSKSNNTKEKNKNEENTKKSCPKKGNHKEGKEKVSKKSKVSEADENDKHKKSKKKECDHSKHKKSKVSEADENDKHKKSKKKEGDHDKHKKSK
K2-6	-275	-250	DSKSNNTKEKNKNEENTKKSCPKKGNHKEGKEKVSKKSKVSEADENDKHKKSKKKECDHSKHK
Scramble 1	-264	-254	GGSDNSEKNKKEHKKEKSAHEKEKDKKAEHKKDKDGSKKSSEKKKNSTKEHESSEHNHSEHKKSDNHKKKKENKEDDSEKEKNKKSDVVSDNKKTHKKGGKKKSKPKGNKSVHD
Scramble 2	-273	-286	EKSDDKKESKSKHDGSVKKKSEEKKKDNESSTKSDGKNKKNKKHNEHESKHEHKGEGKKKEHGSNEDHKKKNDNEVKGKHKKSPKKKAETKKNSSHDSDKKEKDSSKKVHNEAK
Scramble 3	-274	-308	HDNKGKKSHDSSGDKKHKNKKSKKKKESPTVDKKEKHEVKKKHKNKKDDVHSEEASKEENKKTEKGKNEKESNAKKSKNSHHDSKKGKEHEGSSKNKDHDKKEKNSDESKKGES
Scramble 4	-281	-287	ESNKKNGHEKGGKKSESKHKKKDHAEKEPKNKKGAHHDDSDKSKTKHNSESDEKVHKSESKEKSKDKVGNKGKEKKVSNSEKKDKNKKDHSHSKETKEKDEKDKKHKSSNKENK
Scramble 5	-342		ENPKDKANKNDKKKKNHESEKEENKKESSHHKKGSSKVEHEDKVKKTKEKSESSGEKEKNKKSHHSETSANVKSKKKGGDKSGKEKHKKNHKKHSDKDKNSKKKKSDDHKDDEG
Scramble 6	-294		SSGSHSKHDKKNGTDKKEHKKKDKESKHKKKNEKEKVAKSVNKSKGEDTHGNPKKKKDKEKSNDKKKKSNKDHEKEEEHEKNANVSKKKEKNKGHSSKDEDSHKKSHEDKSSEG
Scramble 7	-350		NKSKEEKKHHNEDDKEGKVEDKKGENKENEHKVEKDKDKHKKNGSPNKESENTHSNSKKGSGDNSVTDDHHKKSKHSKSKKKKSEHKKKKASSKDKDKASKEGKHKSEKKKSKE
Scramble 8	-323		SKSDHKEKAENDEDKSKEKASKKHVKNVEDKHHKSEKNKHEKGKNKKKKHKEGSKKKDDHKKSKPNEEKKTKDEKDHTGGKKKSSHNNKNSNKHSEVSKDKEKESKKSSGSGDE
Scramble 9	-322		KHKDGHSHSDSNNDESSKEVKGKKKHKKNNSGEGAESVDKNSKKSEHKHKEEGKDASHKNKKEKKKTKTKESSKKEEDSKKNKKDPDDKKHSKVKEEEKKHNENKHKDKKGSKS
Scramble 10	-325		DKKKKGKDNESNDEAVDKSKKHSEHKSESPSEKKDKKEEKGHNSDKKSSEEKNKKKKKSEKKKNKKHEDEASHHGKSDNTSKHHKSKGESDVEKKKNKGNDNGKHHEKVTKKS