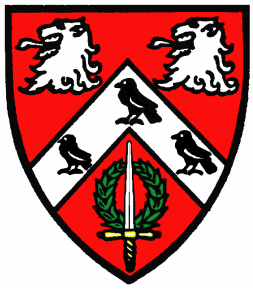
Structural Studies of Malaria Proteins

Christina Mayer



St. Anne's College
University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2012

Wanunda.

A. K.

ABSTRACT

Malaria is a disease of global importance, causing hundreds of thousand of deaths a year. The majority or deaths are caused by *Plasmodium falciparum*, a parasite transmitted by the mosquito *Anopheles*. Its pathogenicity largely results from an ability to transform infected erythrocytes by creating knob-like structures that result in endothelial adhesion. Two major components of these knob structures have been identified as *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1) and knob-associated histidine rich protein (KAHRP). The extracellular fragment of *Pf*EMP1 is responsible for antigenic variability and cytoadherence while its intracellular domain (ATS) connects to the cytoskeleton via interactions with other plasmodium-encoded proteins. In addition, perforin-like proteins (PLPs) with a MACPF domain have been identified in the genome of *Plasmodium*. PLPs are highly conserved and are expressed in various life-cycle stages of the parasite. They are believed to form pores in membranes of the host cell but their structure is yet unknown. The aim of the work in this thesis was to obtain new information about the structure and role of malaria proteins, thus giving a better understanding of the disease and its possible treatment. Studies of numerous designed constructs of the ATS family were carried out using biophysical methods including high resolution NMR, SAXS and CD. These revealed that ATS domains are mainly unstructured with a relatively small folded core, consisting of a bundle of α-helices. Surprisingly, no evidence could be found for ATS binding to KAHRP in solution conditions although previous pull-down data had indicated an interaction. Bioinformatics analysis and yeast-two-hybrid data suggested, however, that there is a conserved protein interaction epitope on the central flexible part of ATS. It was shown, using fluorescence anisotropy measurements that this part of ATS associates with a parasite protein containing a PHIST (Plasmodium helical interspersed sub-telomeric) domain. Expression constructs of the PLP protein family were designed and manufactured, with the aim of enabling structural studies of these putative pore proteins.

ACKNOLEDGEMENTS

Firstly, I would like to thank my supervisors Iain Campbell, Robert Gilbert and Ioannis Vakonakis for their encouragement, patience, understanding and support during my DPhil.

Special thanks go to Ioannis Vakonakis for the supervisory role he played and his constant scientific support, for his practial help with the NMR and in the lab, without him this thesis would not have been possible in this current form. I would like to thank Leanne Slater for providing protein samples and for other help of any kind. I would also like to thank Robert Konrat for this collaboration and for his encouragement and enthusiasm. I greatly appreciate all of these valuable contributions and am very grateful to them for allowing me to reproduce their work here.

Special thanks go to Luke for his help with baculovirus expression and enjoyable coffee breaks. I would like to thank Benjamin and Amber for their help with mammalian protein expression; Weixan for her advice on tissue culture matters. I would also like to thank all my colleagues in the IDC group and in Strubi for making it such an enjoyable place to work, in particular, my friends Ben and Barbara for their support and countless lunch and coffee breaks.

I would also like to take the opportunity here to thank Helen Mott and Matt Higgins for agreeing to examine me on this thesis. I would also like to thank the Wellcome Trust for funding my DPhil and St Anne's College for their support.

I would like to acknowledge Nick Soffe, Jonathan Boyd and Christina Redfield for their assistance with NMR problems of all kinds. I would also like to acknowledge Dave Staunton. Many thanks go to Margaret Jones for her assistance with ordering and keeping Strubi in order and Radu for advice on protein expression.

Finally, I would like to thank my parents, my sister and my brothers, their families, Cory and my friends for their constant support, without them this thesis would not exist. I would like to acknowledge Amreeta, Andreas, Bibi, Christian, Daniel, Kasia, Muj, Rhianedd and Rizvi.

CONTENTS

ABBREVIATIONS

ATS	acidic terminal segment
AUC	analytical ultra centrifugation
BSA	bovine serum albumin
CD	circular dichroism
CDC	cholesterol-dependent cytolysins
C-terminal	carboxy-terminal
DNA	deoxyribonucleic acid
cDNA	complementary deoxyribonucleic acid
DDT	dichlorodiphenyltricholroethane
DSS	2,2-dimethyl-2-silapentane-5-sulphate
DTT	dithiothreitol
DMEM	Dulbecco's modified eagle medium
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetraacetic acid
EOM	ensemble optimisation method
FPLC	fast performance liquid chromatography
GFP	green fluorescent protein
GST	glutathione S-transferase
HEK	human embryonic kidney
HSQC	heteronuclear single- quantum coherence
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IPTG	isopropyl-β-D-1-thiogalactopyranoside
iRBC	infected red blood cells
KAHRP	knob associated histidine-rich protein
LB	Luria Bertani
MAC	membrane attack complex
MRE	mean residue ellipticity
M_w	molecular weight
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
N-terminal	amino-terminal

OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PEXEL	protein export element/host targeting
*Pf*EMP1	*P. falciparum* erythrocyte membrane protein 1
*Pf*EMP2	*P. falciparum* erythrocyte membrane protein 2
*Pf*EMP3	*P. falciparum* erythrocyte membrane protein 3
PHIST	*Plasmodium* helical interspersed sub-telomeric
pI	isoelectric point
PII	protein interaction index
PLP	perforin-like protein
*P*PLP	*Plasmodium* perforin-like protein
ppm	parts per million
PV	parasitophorous vacuole
RESA	ring-infected erythrocyte surface antigen
RMSD	root mean square deviation
SAXS	small angle X- ray scattering
SDS-PAGE	sodium dodecylsulphate polyacrylamide gel electrophoresis
SF-9	*Spodoptera frugiperda*
SMART	simple modular architecture research tool
TRIS	trishydroxymethylaminomethane
UV	ultraviolet
WHO	world health organisation
Y2H	yeast two hybrid
YNB	yeast nitrogen base

CHAPTER ONE

INTRODUCTION

1.1 Malaria

Malaria has been with us for millennia and despite international efforts to control or eliminate the disease we have not yet succeeded. In 1998 the World Health Organisation (WHO) started the latest global initiative to eradicate the deadly disease. The WHO reported 247 million cases of malaria in 2006, the majority emanating from lower-income countries (WHO, 2010). The disease is a major obstacle to the economic improvement of the affected, mostly developing, countries and consigns their inhabitants to a relentless cycle of poverty and disease (Jones and Good, 2006). Malaria takes the lives of more than one million people per year (Murray, 2012), most of whom are children under the age of five in sub-Saharan Africa. Although major efforts are undertaken, there is as yet no vaccine available at present (Hill, 2011). The best available treatment for *P. falciparum* is an artemisinin-based combination therapy (WHO, 2010). However, increasing resistance to different drugs and the business of counterfeit drugs are major obstacles in fighting the deadly disease. Identification of novel therapeutic targets and development of new intervention methods is, therefore, of critical importance.

1.2. *Plasmodium*

The parasite species *Plasmodium* belongs to the phylum *Apicomplexa*; they are transmitted to humans by the female *Anopheles* mosquito. Five species have been reported to cause malaria in humans, namely *P. falciparum, P. vivax, P. ovale, P. malariae and P. knowlesi*. *P. falciparum* and to a lesser extent *P. vivax* are the most common cause of disease and death from malaria (Mendis et al., 2001); whereas, chronic malaria can be attributed to *P. ovale* and *P. vivax*. The later three strains possess the ability to 'hibernate' in hepatocytes, as hypnozoites (Trampuz et al., 2003). Therefore, a relapse of the disease and related symptoms might occur months even years after exposure (Adak et al., 1998).

1.2.1 *Plasmodium falciparum*

The most pathogenic malarial species is *P. falciparum*; it is responsible for 80 % of malaria infections and 90 % of deaths caused by the disease. *P. falciparum* is widespread in Sub-Saharan Africa (Atlas, 2012) (Figure 1.1). If untreated, severe malaria caused by *P. falciparum* can lead to coma and death. Children and non-immune pregnant women are at highest risk of severe malaria (Greenwood et al., 2005).

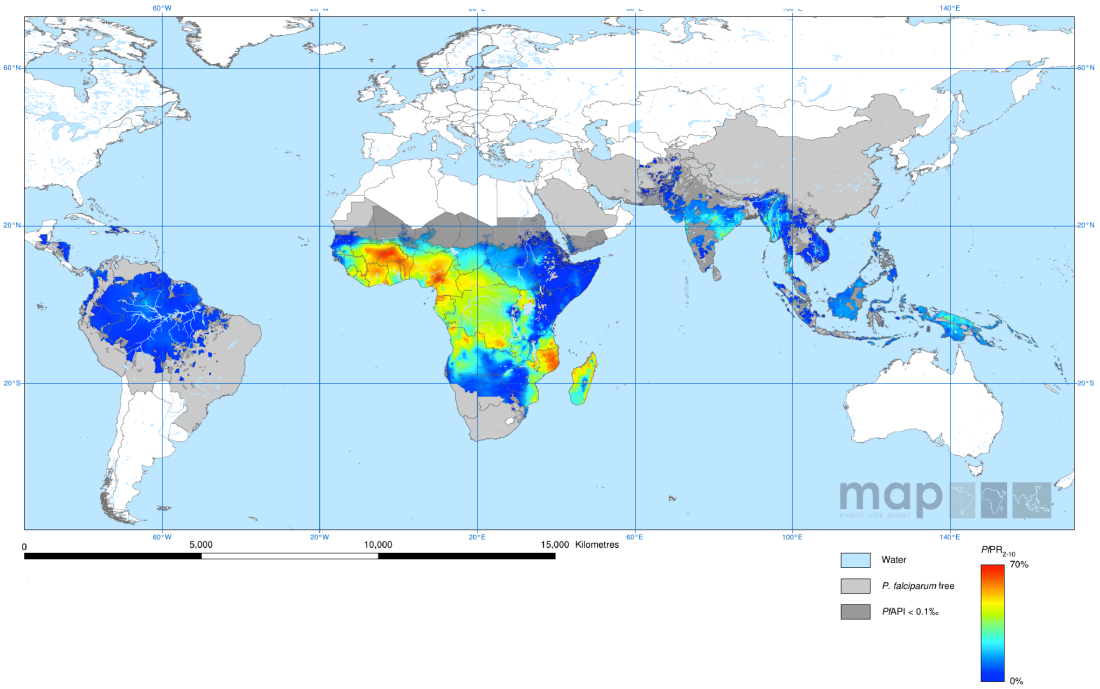


Figure 1.1: The global spatial distribution of endemic *Plasmodium falciparum* malaria in 2010. The map illustrates the estimated proportion of 2-10 year old in the general population that are infected with *P. falciparum* at any one time during 2010. (Graphic adapted from (Malaria (Atlas, 2012) project).

1.2.2. *Plasmodium* – Life cycle

During its life cycle the malaria parasite undergoes different invasive stages (Figure 1.2). The parasite is injected to the host by a bite of the *Anopheles* mosquito through the human dermis in the form of sporozoites. Only a few (10 – 30) sporozoites are transmitted via the saliva of the *Anopheles* (Menard, 2005), which also inhibits blood coagulation. Subsequently, they travel in the lymphatic and vascular system (Amino et al., 2006) and ultimately to the liver, where they infect hepatocytes. Within these cells the first stage of propagation takes place; multiplied *Plasmodia* rupture the infected liver cells and leave the organ in the form of merozoites. This differentiation step is necessary for them to infect their next target – the erythrocytes. Within

minutes of release into the bloodstream merozoite *Plasmodia* enter red blood cells. During the erythrocyte stage, the first overt symptoms of malaria occur. After invasion of the red blood cells, modifications of their membrane can be observed, while at the microscopic level modifications of the red blood cells can be noticed. The infected red blood cells (iRBCs) lose their ability to deform and their shape and architecture alter dramatically. In brief, the host cell is rearranged to provide for the parasite and its needs. More daughter merozoites are produced and about 16 – 32 per erythrocyte are released by bursting of the iRBC (Bannister et al., 2000; Garcia et al., 2008). Each asexual cycle lasts approximately 48 hours and the periodic cell bursts are responsible for the typical malarial fever waves (Miller et al., 2002). In some cases gametocytes develop within the iRBCs and these can be taken up by another mosquito bite. The ingested gametocytes further develop into male and female gametes and fuse into a zygote in the mosquito's gut. The further developed motile ookinete can penetrate the gut's wall and nest there as an oocyst. Within this structure cell division takes place. Once the oocyst ruptures, it releases sporozoites that find their way to the salivary gland of the mosquito and wait for their next human victim. The malarial parasite represents a complex organism that has evolved to overcome several obstacles and face various challenges while migrating from the mosquito vector to the human host organism and back again.

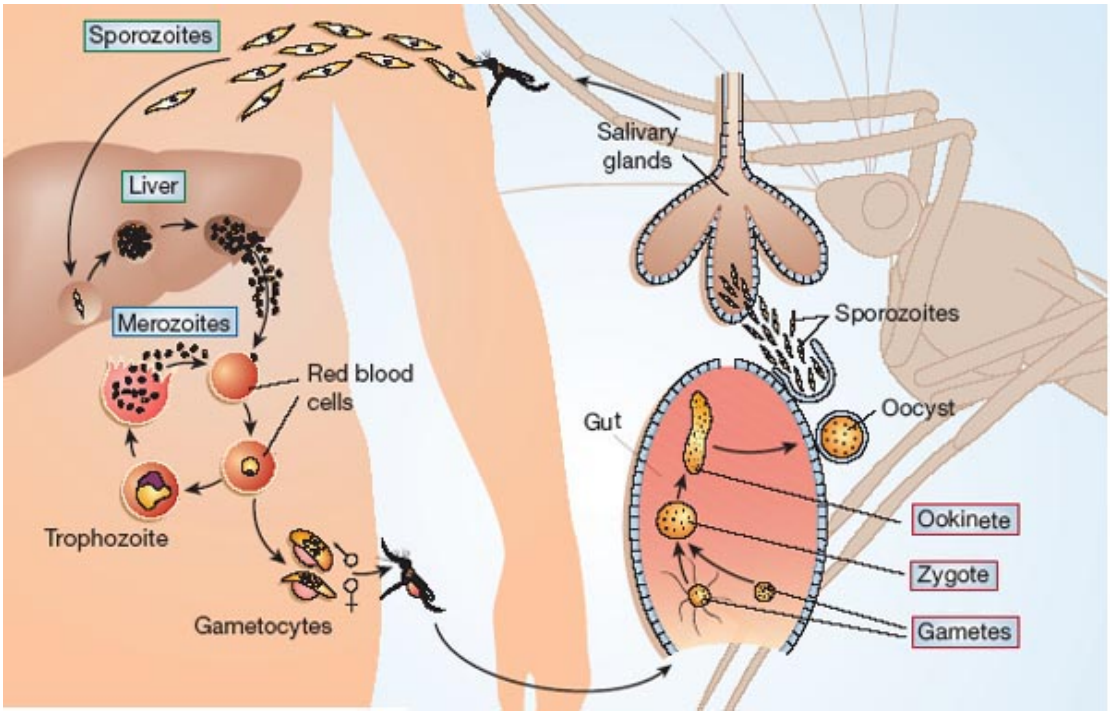


Figure 1.2: The life cycle of *Plasmodium* in humans and in the mosquito vector, highlighting the different asexual stages within the human host and sexual ones in the midgut of the *Anopheles* mosquito. Taken from (Menard, 2005).

1. 3. Cytoadherence

Once the malaria parasite has developed into a merozoite, it can invade erythrocytes and use them for propagation. After infection the apicomplexan parasite resides within a parasitophorous vacuole (PV) from where it takes over the human host cell (Figure 1.4) and morphs through the so-called ring, trophozoite and schizont stages (Bannister et al., 2000; Garcia et al., 2008). Sixteen hours post invasion, *P. falciparum* infected red blood cells show altered shape, decreased flexibility and adherence to host endothelium. The so-called parasite *exportome* is responsible for these morphological changes (Maier et al., 2009); *P. falciparum* encoded proteins are synthesized and subsequently transported to the host cell, including the iRBC membrane, where they form knob-like structures involving cytoskeleton components, including actin and spectrin (Nagao et al., 2000; Trager et al., 1966). These knobs act as a molecular adhesive and mediate binding to the endothelium and other cell types (figure 1.3). The major components of knob-like structures are KAHRP (knob-associated histidine rich protein) and *Pf*EMP1 (*P. falciparum* erythrocyte membrane protein 1). In addition, *Pf*EMP2, *Pf*EMP3, *Pf*332, RESA (ring-infected erythrocyte surface antigen) and other parasite-encoded proteins are associated with these knob structures (Foley et al., 1994; Foley et al., 1991; Ruangjirachuporn et al., 1991). Severe disease forms, such as cerebral the most acute or placental malaria (Scherf et al., 2008), can be attributed to endothelial cytoadhesion of *P. falciparum* iRBC leading to blood vessel occlusion. Cytoadherence is a unique characteristic of *P. falciparum,* which enables the parasite to evade clearance by the spleen and create a microenvironment suitable for merozoite maturation.

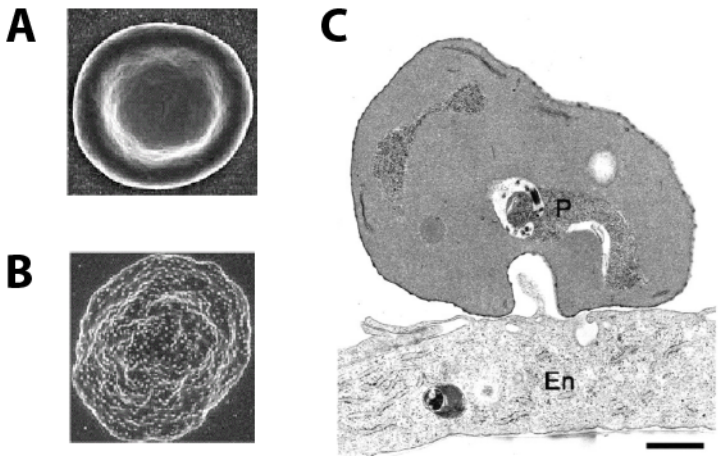


Figure 1.3: *Plasmodium falciparum* infected red blood cells. The transformation of a healthy erythrocyte (A) into a "sticky" *Plasmodium* infected cell (B) Taken form (Rug et al., 2006). (C) Transmission electron micrograph of a parasite (P) infected erythrocyte attaching to an endothelial cell (En).

1.3.1 The *exportome* of *P. falciparum*

During the intra-erythrocyte stage of the *P. falciparum* life cycle a large number of proteins are trafficked into the host cell and its membrane. Up to 8 % of the parasite's gene products have been predicted to be exported to distinct sites in the iRBC (Sargeant et al., 2006). Different sophisticated mechanisms have been reported to facilitate the crossing of the parasite membrane as well as the parasitophorous vacuole (PV) membrane (Cooke et al., 2004; Marti et al., 2005). Extensions of this vacuole membrane - the tubulovesicular network - are supposed to be involved in trafficking lipids and other molecules into the iRBC. Furthermore, Maurer's clefts - disc-shaped membrane structures - appear in the cytosol of the infected erythrocyte, where they manage the transport of virulence proteins to the host membrane (Lanzer et al., 2006; Przyborski et al., 2003). A pentameric amino acid sequence on the *Plasmodium*-encoded proteins acts as a signal sequence for export to the iRBC (Hiller et al., 2004; Marti et al., 2004). The protein export element / host targeting (PEXEL/

HT) motif, R/KXLXE/Q/D can be found on both soluble and membrane proteins of *P. falciparum*. This five amino acid long sequence is recognized and cleaved by a novel endoplasmic reticulum peptidase, then the newly exposed N-terminus is acetylated and supposed to be recognized by a specific transporter in the PV membrane (Boddey et al., 2009; Chang et al., 2008). Interestingly, other proteins lacking this signal sequence (for example *Pf*EMP1) are exported as well (Spycher et al., 2006). Therefore, it is likely that other signal motifs or transport mechanisms still remain to be discovered.

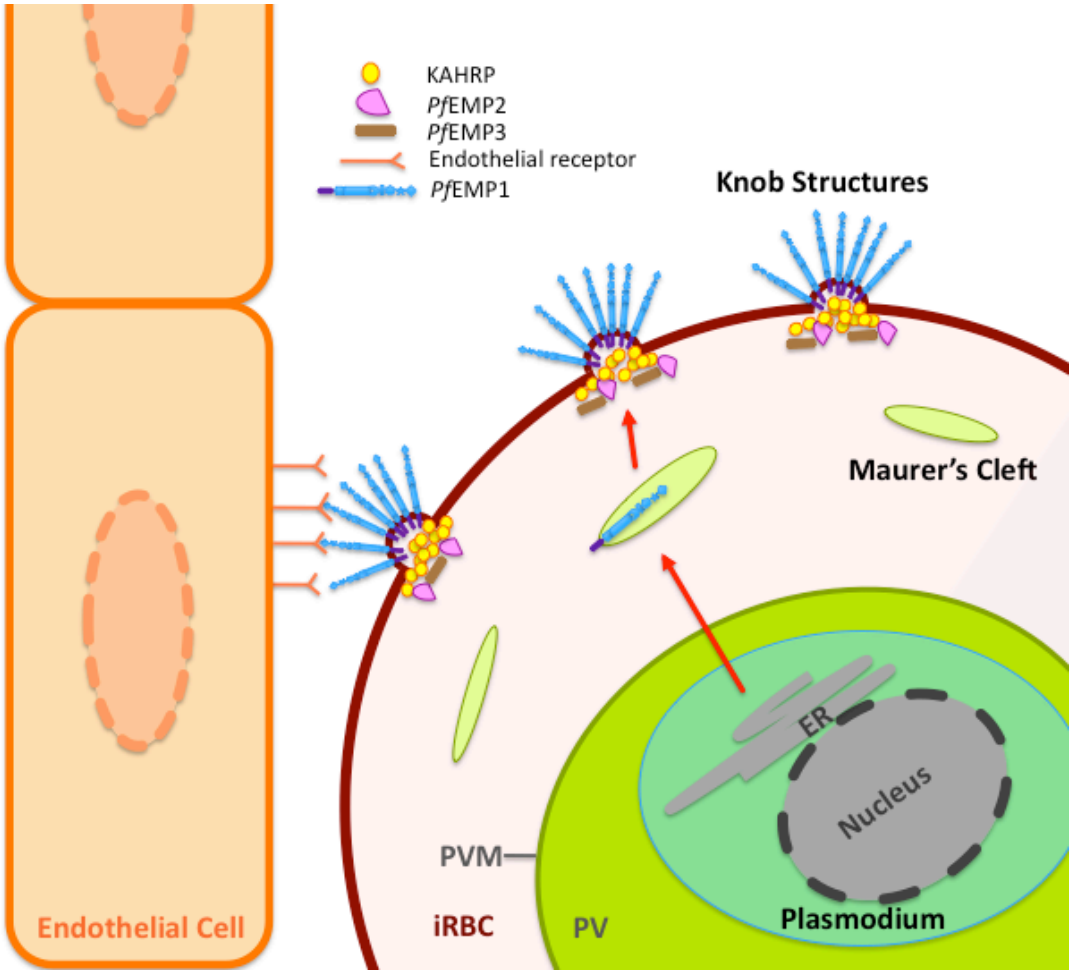


Figure 1.4: Model of the host erythrocyte during *P. falciparum* invasion. During the intraerythrocyte stage the parasite stays within a parasitophorous vacuole (PV) from where it shuttles *Pf*EMP1, *Pf*EMP2, *Pf*EMP3, KAHRP and other proteins (not shown in this picture) to the membrane of the iRBC to build knob structures. *Pf*EMP1 is transported via the Maurer's clefts to its final destination (indicated by the red arrows). Once properly inserted in the membrane, different extracellular domains of *Pf*EMP1 can bind to various endothelial receptors and therefore cause cytoadherence. Nomenclature: ER, endoplasmatic reticulum, iRBC, infected red blood cell; PVM, parasitophorous vacuole membrane.

1.3.2. *Pf*EMP1

A major component of knob structures, apparent in the electron microscope, is the *P. falciparum* erythrocyte protein 1 (*Pf*EMP1) (Miller et al., 2002). This 200 – 350 kDa protein is encoded by the family of *var* genes, of which 59 copies can be found in the 3D7 genome of *P. falciparum* (Bull et al., 2005; Smith et al., 1995; Su et al., 1995). The *var* genes are composed of two exons. The first one encodes the highly polymorphic extracellular domains and a transmembrane segment. The intracellular part termed the acidic terminal segment (ATS) is encoded by the second 1 – 1.5 kb long exon. The variable extracellular domains of *Pf*EMP1 are responsible for adherence and antigenic variation and have been the subject of various studies including high-resolution structural studies (Higgins, 2008; Klein et al., 2008). The different extracellular regions bind to the host via various endothelial molecules, including CD36, ICAM-1 and CSA (Figure 1.5) (Barnwell et al., 1989; Barragan et al., 1999; Berendt et al., 1989), enabling the iRBCs to attach to the blood vessels (Kraemer and Smith, 2006). This promotes the accumulation of iRBC in the microvasculature and can lead to clogging of blood vessels. However, this mechanism is beneficial to the survival of parasite-infected cells as it allows them to be sequestered and hence avoid clearance by the spleen. Evidently, the host immune system is stimulated by the presentation of the foreign proteins on the erythrocytes and responds by antibody release. However, *P. falciparum* can evade the host immune response by antigenic variation – switching to a different variant of *Pf*EMP1 (Kyes et al., 2001).

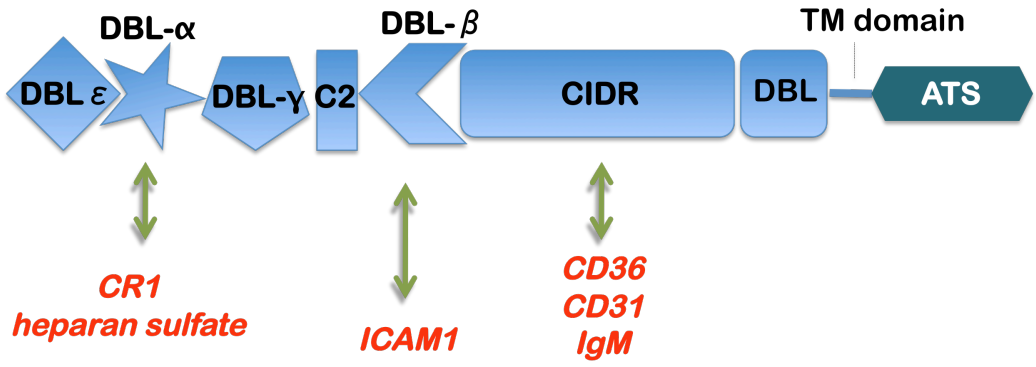


Figure 1.5: Domains of *Pf*EMP1. The multidomain protein is exported to the red blood cell surface. Different parts of *Pf*EMP1 can interact with specific receptors (indicted by arrows) (Maier et al., 2009). Therefore, some variants of *Pf*EMP1 mediate adhesion to the endothelium of distinct organs better than others. Nomenclature: DBL, Duffy-binding-like; CIDR, Cysteine-rich interdomain region; TM, transmembrane; ATS, acidic terminal segment.

The intracellular segment of *Pf*EMP1, ATS, is more conserved than the extracellular domain in the *Pf*EMP1 family of genes (Kraemer and Smith, 2006). The domain is highly acidic and composed of 390 to 500 amino acids (Su et al., 1995). There was until recently little information available on ATS and the intracellular side of the knob-like structures. ATS is presumed to anchor *Pf*EMP1 via other parasite-encoded proteins to the erythrocyte cytoskeleton (Oh et al., 2000; Waller, 2002).

1.3.3. KAHRP

The knob associated histidine rich protein (KAHRP) is the major component of erythrocyte knobs (Culvenor et al., 1987; Pologe et al., 1987). The accumulation of KAHRP at erythrocyte membranes significantly reduces the deformability of the iRBC (Crabb et al., 1997; Glenister et al., 2002). The primary amino acid sequence of KAHRP contains a PEXEL signal (section 1.3.1) and the protein has been shown to be exported across the PV membrane to the erythrocyte membrane (Marti et al., 2004;

Wickham et al., 2001). Previous knockout experiments of KAHRP implied that this protein is essential for functional cytoadherence under flow conditions (Crabb et al., 1997; Rug et al., 2006). It has been described to act as a binding partner for ATS, linking it to the cytoskeleton through interactions with spectrin, actin and ankyrin (Kilejian et al., 1991; Oh et al., 2000; Pei et al., 2005). The histidine-rich N-terminal region of KAHRP has been reported to be responsible for binding to the host cytoskeleton, whereas the C-terminal part is necessary for formation of functional knobs (Rug et al., 2006). Previously published literature (Waller et al., 1999; Waller et al., 2002) defined the binding sites for KAHRP to ATS and *vice versa* under *in vitro* conditions. In short, K1A comprising the histidine-rich domain of KAHRP (residues 38 - 126) interacts with the C-terminus of ATS-FL (residues 215 - 279) and the binding of the KAHRP fragment K2A (residues 356 - 447) has been mapped to an N-terminal part of ATS-FL (residues 118 - 214). The interaction of ATS and KAHRP has been described to an electrostatic driven one (Voigt et al., 2000). Full-length ATS has been reported to bind to KAHRP (comprising the second exon) with an affinity of 10 nM (Oh et al., 2000). In addition, phosphorylation of ATS by the host cell casein kinase II has been proposed to strengthen the binding to KAHRP (Hora et al., 2009).

1.4 Membrane pore formation in *Plasmodium*

1.4.1 The membrane attack complex/ perforin family

Formation of pores in membranes is a phenomenon well studied in the immune response. Components of the membrane-attack complex (MAC) of the complement system and perforin form transmembrane channels. The formation of these pores leads to disruption of the target cell membrane and ultimately cell lysis and death. Perforin is produced by cytotoxic T lymphocytes and natural killer cells and is therefore an important player in adaptive immune response. It is released via intracellular granules across the plasma membrane into an infected cell or pathogen. The activity of perforin is dependent on the presence of calcium: when calcium is present, the protein oligomerizes and this leads to pore formation and subsequent lysis of virus-infected or transformed cells (Bolitho et al., 2007; Liu et al., 1995; Pipkin and Lieberman, 2007; Trapani and Smyth, 2002; Voskoboinik et al., 2006). Perforin and the members of the MAC (C6, C7, C8α, C8β and C9) all encode for a Membrane attack complex/ perforin (MACPF) domain. More than 500 members of the MACPF family have been identified in a wide range of organisms (Altschul et al., 1997; Ponting, 1999; Rosado et al., 2007). These include various pathogens such as *Chlamydia* (Ponting, 1999), *Photorhabdus luminescens* (Rosado et al., 2007), *Toxoplasma (Kafsack et al., 2009)* and *Plasmodium* (Kaiser et al., 2004).

1.4.2 MACPF- domain

The MACPF domain comprises a 40kDa segment, which oligomerises, subsequently inserts in the target membrane and is thus responsible for pore formation (Tschopp et al., 1986). The insertion into the membrane is achieved by a conformational change within the MACPF domain (DiScipio and Berlin, 1999; Pipkin and Lieberman, 2007). The MACPF family shares a sequence identity of 15 - 20% within the domain and a signature motif: Y/W-G-T/S-H-F/Y-X_6-GG (where X is any amino acid) (Ponting, 1999).

1.4.3 MACPF/CDC Superfamily

The X-ray crystal structure of the C8α component of the MAC proved to be structurally homologous to the bacterial, pore-forming, cholesterol-dependent cytolysins (CDCs) (Hadders et al., 2007; Rosado et al., 2007). The family of CDC proteins consists of pore-forming toxins form Gram-positive bacteria, such as *Streptococcus, Listeria, Clostridium and Bacillus*. Pore formation of CDCs is achieved via a dramatic oligmerisation-associated conformational change in the toxin and has been the subject of various studies (Gilbert, 2005; Gilbert et al., 2000; Gilbert et al., 1999; Rossjohn et al., 2007; Shepard et al., 1998; Tilley et al., 2005). Moreover, pore formation is initiated by the oligomerization of CDC molecules to a prepore on the membrane surface (figure 1.6A). The actual pore is created by dramatic rearrangement of the protein domains, most importantly, the transition of helical clusters into membrane spanning, amphipathic β hairpins (figure 1.6B) (Gilbert et al., 1999; Tilley et al., 2005).

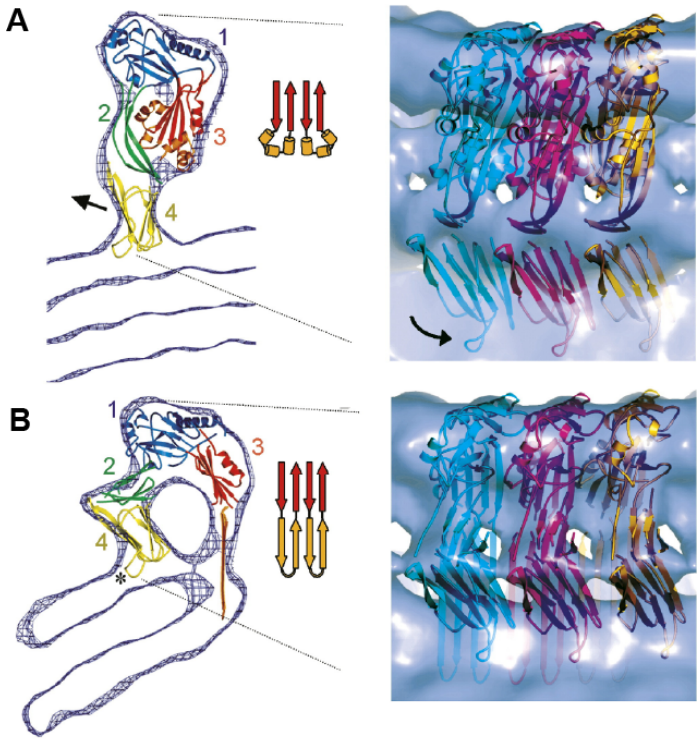


Figure 1.6: Mechanism of pore-formation by pneumolysin. Cross-section trough the **A)** Prepore state **B)** membrane spanning pore state, highlighting the conformational adaptation of domain 3 (shown as cartoon model in red and orange). For both states cryo-electron maps are shown with the fitted domains of one (left) or three subunits (right). Adapted from (Tilley et al., 2005).

Although, members of the MACPF/CDC superfamily share a functional similarity - the formation of pores - and display a highly conserved and complex core fold, only three conserved glycine residues could be identified in both families (figure 1.7) (Rosado et al., 2007). Since MACPF/CDCs have homologous folds and form pores, they should have related mechanisms. However, this assumption still needs to be experimentally confirmed.

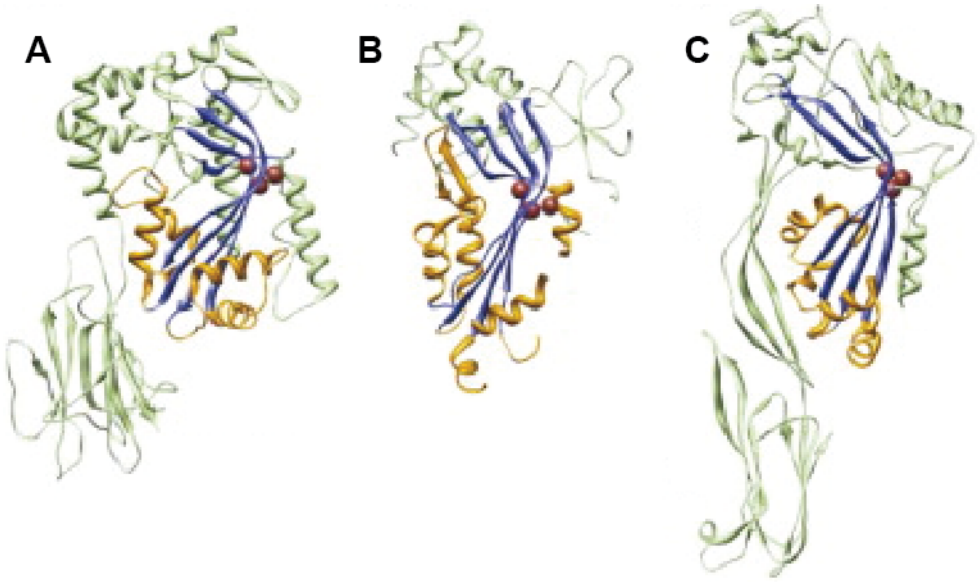


Figure 1.7: Similarities of MACPF domains and CDCs. **A)** *Photorhabdus*-MACPF **B)** human C8α-MACPF and **C)** perfringolysin O. The red spheres indicate the conserved glycine residues. The α-helices which refold during pore formation into the β sheets, are shown in gold. The central kinked β sheet is coloured in blue. Taken from (Lukoyanova and Saibil, 2008).

1.4.4 MACPF- domain in *Apicomplexa*

Interestingly, MACPF- domains could also be located in the genus *Apicomplexa*. Prominent members include the phyla *Toxoplasma* and *Plasmodium*; both are highly pathogenic human parasites. The alignment of apicomplexan Perforin-like Proteins (*Api*PLPs) MACPF- domains revealed a similar but slightly altered MACPF signature motif: W-X_2-F/L-F/Y-X_2-F/Y-G-T-H-X_7-G-G (Kafsack et al., 2009). In addition, studies of the *Toxoplasma* perforin-like protein 1 (*Tg*PLP1) predict a preserved fold for the MACPF-domain and sequence analysis show that formation of a membrane spanning β harpins, necessary for pore-formation, is most likely (Kafsack et al., 2009).

1.4.5 Perforin-like proteins in *Plasmodium*

A family of proteins containing a single MACPF-like domain (membrane attack complex/ perforin pore-forming domain) has been identified in *Plasmodium* species. (Kaiser et al., 2004). Several studies indicate that this family is essential for the invasion and cell traversal of different cell types during the malarial life cycle and hence for malaria infection. *Plasmodium* perforin-like proteins (*P*PLPs) have been identified in the whole *Plasmodium* family and each of the MACPF domains is conserved among *Plasmodium* species, with a sequence conservation of 47 % to 66 % (Kaiser et al., 2004). The *P*PLPs share a common domain architecture, with a variable non-conserved N-terminal segment followed by the MACPF domain and three C - terminal β-pleated sheet domains (figure 1.8A). These C-terminal domains are predicted to be ~55 amino acids long and are unique to apicomplexan PLPs and likely to mediate initial binding of the *P*PLPs to their target membranes (Kafsack and Carruthers, 2010). The existence of an N- terminal signal peptide suggests that all *P*PLPs are secreted (Kaiser et al., 2004).

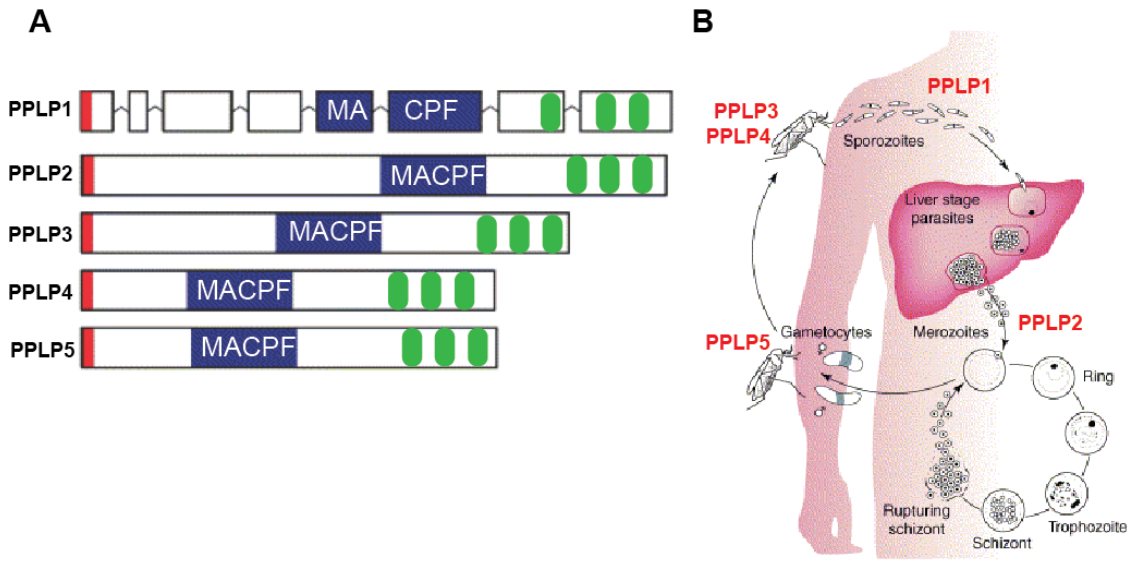


Figure 1. 8: Perforin-like proteins in *Plasmodium*. A) Schematic overview of the identified PLPs in *Plasmodium*, highlighting their MACPF domain (blue). Solely, PPLP1 is encoded by an eight-exon gene, whereas *P*PLP 2-5 are encoded in a single open reading frame (ORF). All genes contain an N-terminal signal peptide, highlighted in red as well as three tandem repeat β-pleated sheet domains (green). Adapted from (Kaiser et al., 2004) **B)** Overview of the life cycle of Plasmodium and the stages of expression of the *P*PLPs are indicated.

Of the five identified *P*PLPs, *P*PLP1 was the first to be described, as SPECT2 (sporozoite protein essential for cell traversal) in *P. berghi*. *P*PLP1 is upregulated in sporozoites (Lasonder et al., 2002) (figure 1.8B) and has been shown to be fundamental for migration of parasites during early stages of mammalian host infection (Sinnis and Coppi, 2007). More precisely, it is necessary for traversal of the sinusoidal cell layer prior to hepatocyte infection (Ishino et al., 2005; Kaiser et al., 2004). Little is known about *P*PLP2; proteomic studies indicate that *P*PLP2 is expressed in the merozoite stage of the malarial life cycle (Florens et al., 2002). However, its function there still remains unkown. Both *P*PLP3 and *P*PLP5 have been reported to be necessary for mosquito midgut invasion (Ecker et al., 2007; Kadota et

al., 2004). Separate knockout experiments of both genes have shown to result in the same phenotype - parasites were not able to pass or invade the midgut epithelia and then develop into infectious sporozoites. Interestingly, both genes are essential for successful parasite development (Ecker et al., 2008; Ecker et al., 2007), which advocates that they may cooperate to succeed in midgut traversal. Furthermore, *P*PLP5 expression has also been shown in gametocytes by proteomic studies (Florens et al., 2002). What kind of function it occupies at this stage still remains to be investigated. Lastly, the expression of *P*PLP4 has been detected in the ookinete form of the parasite as well (Hall and Carlton, 2005; Raibaud et al., 2006). This variant contains a sequence of five positively charged lysine residues near the C-terminus. This region has been speculated to facilitate interaction with the negatively charged groups of phospholipids of cell membranes (Kafsack and Carruthers, 2010). PPLP4 has not been further characterized up to now.

1.5 Objectives

The importance of malaria studies is indisputable for global health. A major factor in the virulence of *P. falciparum* induced malaria can be attributed to cytoadherence. However, many basic molecular aspects of this phenomenon, in particular on the intracellular side, are still poorly understood. When the work for this thesis began, no structural information was available neither on the ATS nor KAHRP or other involved components. Furthermore, no biophysical characterization of ATS and its interactions had yet been performed. This thesis describes an extensive study of the structure of ATS and its interactions with other parasite proteins, especially KAHRP. Furthermore, the structure and function of the all of the *P*PLPs is unknown at the moment. The molecular mechanisms of how *Plasmodium* accomplishes cell traversal and possible pore formation still remain elusive. Understanding the activity of this protein family is very important due to their roles in infection.

1.6 References

Adak, T., Sharma, V.P., and Orlov, V.S. (1998). Studies on the Plasmodium vivax relapse pattern in Delhi, India. The American journal of tropical medicine and hygiene *59*, 175-179.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research *25*, 3389-3402.

Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., and Menard, R. (2006). Quantitative imaging of Plasmodium transmission from mosquito to mammal. Nat Med *12*, 220-224.

Atlas, P.M. (2012). http://www.map.ox.ac.uk.

Bannister, L.H., Hopkins, J.M., Fowler, R.E., Krishna, S., and Mitchell, G.H. (2000). A brief illustrated guide to the ultrastructure of Plasmodium falciparum asexual blood stages. Parasitol Today *16*, 427-433.

Barnwell, J.W., Asch, A.S., Nachman, R.L., Yamaya, M., Aikawa, M., and Ingravallo, P. (1989). A human 88-kD membrane glycoprotein (CD36) functions in vitro as a receptor for a cytoadherence ligand on Plasmodium falciparum-infected erythrocytes. The Journal of clinical investigation *84*, 765-772.

Barragan, A., Spillmann, D., Carlson, J., and Wahlgren, M. (1999). Role of glycans in Plasmodium falciparum infection. Biochemical Society transactions *27*, 487-493.

Berendt, A.R., Simmons, D.L., Tansey, J., Newbold, C.I., and Marsh, K. (1989). Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for Plasmodium falciparum. Nature *341*, 57-59.

Boddey, J.A., Moritz, R.L., Simpson, R.J., and Cowman, A.F. (2009). Role of the Plasmodium export element in trafficking parasite proteins to the infected erythrocyte. Traffic *10*, 285-299.

Bolitho, P., Voskoboinik, I., Trapani, J.A., and Smyth, M.J. (2007). Apoptosis induced by the lymphocyte effector molecule perforin. Current opinion in immunology *19*, 339-347.

Bull, P.C., Berriman, M., Kyes, S., Quail, M.A., Hall, N., Kortok, M.M., Marsh, K., and Newbold, C.I. (2005). Plasmodium falciparum variant surface antigen expression patterns during malaria. PLoS pathogens *1*, e26.

Chang, H.H., Falick, A.M., Carlton, P.M., Sedat, J.W., DeRisi, J.L., and Marletta, M.A. (2008). N-terminal processing of proteins exported by malaria parasites. Mol Biochem Parasitol *160*, 107-115.

Cooke, B.M., Lingelbach, K., Bannister, L.H., and Tilley, L. (2004). Protein trafficking in Plasmodium falciparum-infected red blood cells. Trends in parasitology *20*, 581-589.

Crabb, B.S., Cooke, B.M., Reeder, J.C., Waller, R.F., Caruana, S.R., Davern, K.M., Wickham, M.E., Brown, G.V., Coppel, R.L., and Cowman, A.F. (1997). Targeted gene

disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell *89*, 287-296.

Culvenor, J.G., Langford, C.J., Crewther, P.E., Saint, R.B., Coppel, R.L., Kemp, D.J., Anders, R.F., and Brown, G.V. (1987). Plasmodium falciparum: identification and localization of a knob protein antigen expressed by a cDNA clone. Exp Parasitol *63*, 58-67.

DiScipio, R.G., and Berlin, C. (1999). The architectural transition of human complement component C9 to poly(C9). Molecular immunology *36*, 575-585.

Ecker, A., Bushell, E.S., Tewari, R., and Sinden, R.E. (2008). Reverse genetics screen identifies six proteins important for malaria development in the mosquito. Molecular microbiology *70*, 209-220.

Ecker, A., Pinto, S.B., Baker, K.W., Kafatos, F.C., and Sinden, R.E. (2007). Plasmodium berghei: plasmodium perforin-like protein 5 is required for mosquito midgut invasion in Anopheles stephensi. Exp Parasitol *116*, 504-508.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., *et al*. (2002). A proteomic view of the Plasmodium falciparum life cycle. Nature *419*, 520-526.

Foley, M., Corcoran, L., Tilley, L., and Anders, R. (1994). Plasmodium falciparum: mapping the membrane-binding domain in the ring-infected erythrocyte surface antigen. Exp Parasitol *79*, 340-350.

Foley, M., Tilley, L., Sawyer, W.H., and Anders, R.F. (1991). The ring-infected erythrocyte surface antigen of Plasmodium falciparum associates with spectrin in the erythrocyte membrane. Mol Biochem Parasitol *46*, 137-147.

Garcia, C.R., de Azevedo, M.F., Wunderlich, G., Budu, A., Young, J.A., and Bannister, L. (2008). Plasmodium in the postgenomic era: new insights into the molecular cell biology of malaria parasites. International review of cell and molecular biology *266*, 85-156.

Gilbert, R.J. (2005). Inactivation and activity of cholesterol-dependent cytolysins: what structural studies tell us. Structure *13*, 1097-1106.

Gilbert, R.J., Jimenez, J.L., Chen, S., Andrew, P.W., and Saibil, H.R. (2000). Structural basis of pore formation by cholesterol-binding toxins. Int J Med Microbiol *290*, 389-394.

Gilbert, R.J., Jimenez, J.L., Chen, S., Tickle, I.J., Rossjohn, J., Parker, M., Andrew, P.W., and Saibil, H.R. (1999). Two structural transitions in membrane pore formation by pneumolysin, the pore-forming toxin of Streptococcus pneumoniae. Cell *97*, 647-655.

Glenister, F.K., Coppel, R.L., Cowman, A.F., Mohandas, N., and Cooke, B.M. (2002). Contribution of parasite proteins to altered mechanical properties of malaria-infected red blood cells. Blood *99*, 1060-1063.

Greenwood, B.M., Bojang, K., Whitty, C.J., and Targett, G.A. (2005). Malaria. Lancet *365*, 1487-1498.

Hadders, M.A., Beringer, D.X., and Gros, P. (2007). Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense. Science *317*, 1552-1554.

Hall, N., and Carlton, J. (2005). Comparative genomics of malaria parasites. Current opinion in genetics & development *15*, 609-613.

Higgins, M.K. (2008). The structure of a chondroitin sulfate-binding domain important in placental malaria. The Journal of biological chemistry *283*, 21842-21846.

Hill, A.V. (2011). Vaccines against malaria. Philosophical transactions of the Royal Society of London Series B, Biological sciences *366*, 2806-2814.

Hiller, N.L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science *306*, 1934-1937.

Hora, R., Bridges, D.J., Craig, A., and Sharma, A. (2009). Erythrocytic casein kinase II regulates cytoadherence of Plasmodium falciparum-infected red blood cells. The Journal of biological chemistry *284*, 6260-6269.

Ishino, T., Chinzei, Y., and Yuda, M. (2005). A Plasmodium sporozoite protein with a membrane attack complex domain is required for breaching the liver sinusoidal cell layer prior to hepatocyte infection. Cellular microbiology *7*, 199-208.

Jones, M.K., and Good, M.F. (2006). Malaria parasites up close. Nat Med *12*, 170-171.

Kadota, K., Ishino, T., Matsuyama, T., Chinzei, Y., and Yuda, M. (2004). Essential role of membrane-attack protein in malarial transmission to mosquito host. Proceedings of the National Academy of Sciences of the United States of America *101*, 16310-16315.

Kafsack, B.F., and Carruthers, V.B. (2010). Apicomplexan perforin-like proteins. Communicative & integrative biology *3*, 18-23.

Kafsack, B.F., Pena, J.D., Coppens, I., Ravindran, S., Boothroyd, J.C., and Carruthers, V.B. (2009). Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells. Science *323*, 530-533.

Kaiser, K., Camargo, N., Coppens, I., Morrisey, J.M., Vaidya, A.B., and Kappe, S.H. (2004). A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. Mol Biochem Parasitol *133*, 15-26.

Kilejian, A., Rashid, M.A., Aikawa, M., Aji, T., and Yang, Y.F. (1991). Selective association of a fragment of the knob protein with spectrin, actin and the red cell membrane. Mol Biochem Parasitol *44*, 175-181.

Klein, M.M., Gittis, A.G., Su, H.P., Makobongo, M.O., Moore, J.M., Singh, S., Miller, L.H., and Garboczi, D.N. (2008). The cysteine-rich interdomain region from the highly variable plasmodium falciparum erythrocyte membrane protein-1 exhibits a conserved structure. PLoS pathogens *4*, e1000147.

Kraemer, S.M., and Smith, J.D. (2006). A family affair: var genes, PfEMP1 binding, and malaria disease. Current opinion in microbiology *9*, 374-380.

Kyes, S., Horrocks, P., and Newbold, C. (2001). Antigenic variation at the infected red cell surface in malaria. Annual review of microbiology *55*, 673-707.

Lanzer, M., Wickert, H., Krohne, G., Vincensini, L., and Braun Breton, C. (2006). Maurer's clefts: a novel multi-functional organelle in the cytoplasm of Plasmodium falciparum-infected erythrocytes. International journal for parasitology *36*, 23-36.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., *et al*. (2002). Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature *419*, 537-542.

Liu, C.C., Walsh, C.M., and Young, J.D. (1995). Perforin: structure and function. Immunology today *16*, 194-201.

Lukoyanova, N., and Saibil, H.R. (2008). Friend or foe: the same fold for attack and defense. Trends in immunology *29*, 51-53.

Maier, A.G., Cooke, B.M., Cowman, A.F., and Tilley, L. (2009). Malaria parasite proteins that remodel the host erythrocyte. Nature reviews Microbiology *7*, 341-354.

Marti, M., Baum, J., Rug, M., Tilley, L., and Cowman, A.F. (2005). Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. The Journal of cell biology *171*, 587-592.

Marti, M., Good, R.T., Rug, M., Knuepfer, E., and Cowman, A.F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science *306*, 1930-1933.

Menard, R. (2005). Medicine: knockout malaria vaccine? Nature *433*, 113-114.

Mendis, K., Sina, B.J., Marchesini, P., and Carter, R. (2001). The neglected burden of Plasmodium vivax malaria. The American journal of tropical medicine and hygiene *64*, 97-106.

Miller, L.H., Baruch, D.I., Marsh, K., and Doumbo, O.K. (2002). The pathogenic basis of malaria. Nature *415*, 673-679.

Murray, C.J.a.R., Lisa C and Lim, Stephen S and Andrews, Kathryn G and Foreman, Kyle J and Haring, Diana and Fullman, Nancy and Naghavi, Mohsen and Lozano, Rafael and Lopez, Alan D (2012). Global malaria mortality between 1980 and 2010: a systematic analysis. The Lancet *379*, 413 - 431.

Nagao, E., Kaneko, O., and Dvorak, J.A. (2000). Plasmodium falciparum-infected erythrocytes: qualitative and quantitative analyses of parasite-induced knobs by atomic force microscopy. Journal of structural biology *130*, 34-44.

Oh, S.S., Voigt, S., Fisher, D., Yi, S.J., LeRoy, P.J., Derick, L.H., Liu, S., and Chishti, A.H. (2000). Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. Mol Biochem Parasitol *108*, 237-247.

Pei, X., An, X., Guo, X., Tarnawski, M., Coppel, R., and Mohandas, N. (2005). Structural and functional studies of interaction between Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) and erythrocyte spectrin. The Journal of biological chemistry *280*, 31166-31171.

Pipkin, M.E., and Lieberman, J. (2007). Delivering the kiss of death: progress on understanding how perforin works. Current opinion in immunology *19*, 301-308.

Pologe, L.G., Pavlovec, A., Shio, H., and Ravetch, J.V. (1987). Primary structure and subcellular localization of the knob-associated histidine-rich protein of Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America *84*, 7139-7143.

Ponting, C.P. (1999). Chlamydial homologues of the MACPF (MAC/perforin) domain. Current biology : CB *9*, R911-913.

Przyborski, J.M., Wickert, H., Krohne, G., and Lanzer, M. (2003). Maurer's clefts--a novel secretory organelle? Mol Biochem Parasitol *132*, 17-26.

Raibaud, A., Brahimi, K., Roth, C.W., Brey, P.T., and Faust, D.M. (2006). Differential gene expression in the ookinete stage of the malaria parasite Plasmodium berghei. Mol Biochem Parasitol *150*, 107-113.

Rosado, C.J., Buckle, A.M., Law, R.H., Butcher, R.E., Kan, W.T., Bird, C.H., Ung, K., Browne, K.A., Baran, K., Bashtannyk-Puhalovich, T.A., *et al*. (2007). A common fold mediates vertebrate defense and bacterial attack. Science *317*, 1548-1551.

Rossjohn, J., Polekhina, G., Feil, S.C., Morton, C.J., Tweten, R.K., and Parker, M.W. (2007). Structures of perfringolysin O suggest a pathway for activation of cholesterol-dependent cytolysins. Journal of molecular biology *367*, 1227-1236.

Ruangjirachuporn, W., Udomsangpetch, R., Carlsson, J., Drenckhahn, D., Perlmann, P., and Berzins, K. (1991). Plasmodium falciparum: analysis of the interaction of antigen Pf155/RESA with the erythrocyte membrane. Exp Parasitol *73*, 62-72.

Rug, M., Prescott, S.W., Fernandez, K.M., Cooke, B.M., and Cowman, A.F. (2006). The role of KAHRP domains in knob formation and cytoadherence of P falciparum-infected human erythrocytes. Blood *108*, 370-378.

Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., and Cowman, A.F. (2006). Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome biology *7*, R12.

Scherf, A., Lopez-Rubio, J.J., and Riviere, L. (2008). Antigenic variation in Plasmodium falciparum. Annual review of microbiology *62*, 445-470.

Shepard, L.A., Heuck, A.P., Hamman, B.D., Rossjohn, J., Parker, M.W., Ryan, K.R., Johnson, A.E., and Tweten, R.K. (1998). Identification of a membrane-spanning domain of the thiol-activated pore-forming toxin Clostridium perfringens perfringolysin O: an alpha-helical to beta-sheet transition identified by fluorescence spectroscopy. Biochemistry *37*, 14563-14574.

Sinnis, P., and Coppi, A. (2007). A long and winding road: the Plasmodium sporozoite's journey in the mammalian host. Parasitol Int *56*, 171-178.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I., and Miller, L.H. (1995). Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell *82*, 101-110.

Spycher, C., Rug, M., Klonis, N., Ferguson, D.J., Cowman, A.F., Beck, H.P., and Tilley, L. (2006). Genesis of and trafficking to the Maurer's clefts of Plasmodium falciparum-infected erythrocytes. Molecular and cellular biology *26*, 4074-4085.

Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., Ravetch, J.A., and Wellems, T.E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. Cell *82*, 89-100.

Tilley, S.J., Orlova, E.V., Gilbert, R.J., Andrew, P.W., and Saibil, H.R. (2005). Structural basis of pore formation by the bacterial toxin pneumolysin. Cell *121*, 247-256.

Trager, W., Rudzinska, M.A., and Bradbury, P.C. (1966). The fine structure of Plasmodium falciparum and its host erythrocytes in natural malarial infections in man. Bulletin of the World Health Organization *35*, 883-885.

Trampuz, A., Jereb, M., Muzlovic, I., and Prabhu, R.M. (2003). Clinical review: Severe malaria. Crit Care *7*, 315-323.

Trapani, J.A., and Smyth, M.J. (2002). Functional significance of the perforin/granzyme cell death pathway. Nature reviews Immunology *2*, 735-747.

Tschopp, J., Masson, D., and Stanley, K.K. (1986). Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. Nature *322*, 831-834.

Voigt, S., Hanspal, M., LeRoy, P.J., Zhao, P.S., Oh, S.S., Chishti, A.H., and Liu, S.C. (2000). The cytoadherence ligand Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) binds to the P. falciparum knob-associated histidine-rich protein (KAHRP) by electrostatic interactions. Mol Biochem Parasitol *110*, 423-428.

Voskoboinik, I., Smyth, M.J., and Trapani, J.A. (2006). Perforin-mediated target-cell death and immune homeostasis. Nature reviews Immunology *6*, 940-952.

Waller, K.L., Cooke, B.M., Nunomura, W., Mohandas, N., and Coppel, R.L. (1999). 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 *274*, 23808-23813.

Waller, K.L., Nunomura, W., Cooke, B.M., Mohandas, N., and Coppel, R.L. (2002). Mapping the domains of the cytoadherence ligand Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) that bind to the knob-associated histidine-rich protein (KAHRP). Mol Biochem Parasitol *119*, 125-129.

WHO (2010). World Malaria Report 2010 (WHO).

Wickham, M.E., Rug, M., Ralph, S.A., Klonis, N., McFadden, G.I., Tilley, L., and Cowman, A.F. (2001). Trafficking and assembly of the cytoadherence complex in Plasmodium falciparum-infected human erythrocytes. The EMBO journal *20*, 5636-5649.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Standard molecular cloning techniques

For the various molecular biology procedures, standard protocols were followed according to (Russell, 2000). Polymerase chain reactions (PCR) were catalyzed by VENT polymerase (NEB) unless otherwise stated. Typically, PCR reactions were carried out in a total volume of 50 µl containing: template DNA (~ 2ng/µl), 0.8 pmol/ul of each forward and reverse primer, 1 x VENT polymerase buffer, 0.4 mM dNTPs and 0.5 units of VENT polymerase. All restriction reactions were performed using enzymes and additives purchased from New England Biolabs (NEB). For DNA vector dephosphorylation Antarctic phosphatase (NEB) were used according to the manufacturer's recommendations. DNA ligation reactions were conducted for 10 minutes at room temperature in a total volume of 20 µl with an insert to vector ratio of 3:1. These reactions were catalyzed by the Quick Ligase (NEB). 10 µl of this reaction were then used to transform 50 µl or DH5α competent cells. First, the DNA and the cells were incubated for 45 minutes on ice and then heat shocked for 45 seconds at 42 °C. Then cells were cooled on ice for 2 minutes and 300 µl LB medium was added and incubated at 37 °C for one hour with 200 rpm shaking. Finally, the cells were spread on LB plates containing agar and 50 µg/ml carbenicillin and grown overnight at 37 °C. Furthermore, DNA extraction from agarose, PCR purification and plasmid extractions were performed using nucleotide affinity-based kits

(Qiagen). All manufactured DNA constructs were verified by sequencing (Source BioScience, Oxford) and stored at - 20 °C.

2.1.1 Site-directed mutagenesis

Mutations in the ATS-FL construct for fluorescence labelling were introduced using Phusion high fidelity DNA polymerase and DpnI (both purchased from New England Biolabs) according to the QuikChange method (Stratagene).

2.2 Protein expression

Construct design for the different proteins was done with the aid of different secondary and tertiary structure predictors and disorder prediction tools (section 2.16). Different expression strategies were chosen for the various proteins produced. The expression conditions were optimized individually as described later in this chapter. Primary amino acid sequences of all the proteins studied are provided in Appendix III.

2.2.1 Protein production in *Escherichia coli*

All expression was performed in LB medium (Invitrogen), except for production of $^{15}N/\ ^{13}C$ labelled protein when M9 minimal medium was used (Appendix I). The *E. coli* strains BL21(DE3) and BL21(DE3) Codon+RIPL were used unless otherwise mentioned. Overnight starter cultures, grown at 30 °C and inoculated by a single colony, were prepared prior to the main expression cultures. Protein expression was induced by addition of 0.25 mM isopropyl β-D-1-thiogolactopyranoside (IPTG) at an OD_{600} of ~ 0.6. Cells were grown for a maximal duration of 5 hours at 37 °C and

harvested by centrifugation at 7459 g and 4 °C; pellets were resuspended in their individual lysis buffers and stored at -80 °C as necessary.

2.2.2 Mammalian protein expression

A HEK (human embryonic kidney) 293T mammalian cell line was used for transient protein expression. Cell maintenance and transfection was kindly carried out by Mr Benjamin Bishop (University of Oxford). Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, high glucose, Sigma) supplemented with L-glutamine, non-essential amino acids (Gibco) and 10 % foetal calf serum (FCS, Sigma) at 37 °C, 5 % CO_2. Small-scale expression screening was performed in 6 well plates. Each well contained approximately 3x 10^5 cells incubated for 24 hours in DMEM containing 10 % FCS. Shortly before transfection the media was exchanged for 1.5 ml DMEM 2 % FBS per well. 5 µg of purified DNA were mixed with serum-free DMEM to a final volume of 250 µl; a reaction mix containing 8 µl Lipofectamin 2000 and 242 µl serum-free DMEM was added and incubated for 20 minutes at room temperature. After this incubation step the transfection mix was added to the cells. Cell media and cells were analyzed by Western blotting three days after transfection. The media was centrifuged to remove cell debris and cells were suspended in 1 ml PBS. Then, 10 µl of both media and cells were mixed with SDS loading buffer and prepared for SDS-PAGE and subsequent Western Blot analysis. A more detailed transfection protocol has been well described elsewhere (Aricescu et al., 2006).

2.2.3 Baculovirus expression system

Genes of interest were inserted into the transfer vector pTriEx-2 (Novagen) including a C-terminal His_6-tag and co-transfected with baculoviral DNA, which was provided by Dr Geoff Sutton (University of Oxford). *Spodoptera fugiperda* (*Sf9*) cells were supplied by Dr Weixan Wu (University of Oxford) and maintained in Gibco SF900 II medium (Invitrogen) at 28 °C with shaking at 150 rpm according to standard protocols provided here (http://strubiol.icr.ac.uk/extra/baculovirus/baculo_protocols.html). A volume of 2 ml *Sf*9 cells (approximately 1 x 10^6 cells / ml) was transfected with the vector containing the DNA for the protein of interest. After five days the supernatant was collected and stored at 4 °C for further virus amplification. To verify protein expression cells were harvested by centrifugation and analysed by Western blotting (section 2.4).

2.2.4 Cell–free expression

Small-scale cell-free wheat germ expression tests were performed using the pEU-E01-MCS plasmid (N-terminal His_6-tag, Appendix II) and the WEPRO expression kit (CellFree Sciences) according to the manufacturer's instructions (Endo and Sawasaki, 2005, 2007; Sawasaki et al., 2005). In short, 2 μg of designed DNA plasmids were transcribed *in vitro* using SP6 RNA polymerase. The mRNA obtained was then translated using the 'bilayer reaction method' with the translation mix maintained at 16 °C for 20 hours.

2.2.5 COS-7 cell expression assay and Staining

COS-7 cells were provided by Dr Christian Bell (University of Oxford). For the assay cells were grown on glass coverslips and tranfected with pHLsec constructs containing an N-terminal Flag-tag. After two days, the cells were fixed to the glass surface using 4% paraformaldehyde for 10 minutes at room temperature. Then, the coverslips were rinsed with PBS twice and incubated with pre-block solution (1 % w/v bovine serum albumin (BSA) in PBS buffer) for 15 minutes at room temperature. Afterwards, primary antibody was added (5 μg/ml of monoclonal anti-Flag antibody (Sigma)) and incubated for 1 hour at room temperature. Cells were exposed to the Alexa 488-labelled secondary antibody (5 μg/ml) and an additional stain DAPI (Invitrogen), after three wash steps with washing solution (0.1% w/v BSA in PBS buffer). Following three final washing steps, cells were visualised by fluorescence microscopy.

2.3 SDS – PAGE

The quality of protein expression and purification was tested by SDS-polyacrylamide gel electrophoresis using NuPAGE 10% Bis-Tris gels and NuPAGE MES running buffer. Mark-12 protein molecular weight marker (Invitrogen) served as a molecular weight reference.

2.3.1 Native – PAGE

The NativePAGE Novex Bis-Tris Gel System (3-12% and 4-16%) with the corresponding buffers and NativeMark unstained protein standard were used, according to the manufacturer's instructions (Invitrogen). Gels were run at constant voltage (80V) for 2 hours.

2.4 Western blotting

Protein expression was examined by Western blotting with penta-His (1:1000 dilution, Qiagen) as the primary antibody. Protein bands were transferred from a pre-run SDS-PAGE gel to a PVDF (polyvinylidene difluoride) membrane (Invitrogen) in transfer buffer. After membrane blocking with 5 % milk in PBS and three PBS (0.1 % Tween-20) washing steps, the membrane was incubated with primary antibody for one hour at room temperature or overnight at 4 °C. Three membrane washing steps in PBS (0.1 % Tween-20) followed. As a secondary antibody, goat anti-mouse IgG conjugated with peroxidase (1:2000 dilution, Sigma) was applied for 1 hour at room temperature. Finally, the Amersham ECL detection system and photographic film were used for visualization via chemiluminescence (GE Healthcare).

2.5 Concentration determination

Protein and DNA concentrations were determined according to their absorbance at wavelength λ = 280 nm. Absorbance values were measured by a Nanodrop spectrophotometer ND-1000. The Expasy Protparam tool (www.expasy.org) calculated the extinction coefficients ε based on the amino acid composition of the relevant proteins (Walker, 2005). Fluorescein-5-maleimide (5-FAM) labelled ATS-FL was quantified based on the absorbance of the fluorescein at λ = 495 nm with ε = 79000 M^{-1} cm^{-1}.

2.6 ATS clones

The ATS PF08_0141 variant in a pET28a vector (Novagen) was obtained from Dr Amit Sharma (ICGEB, New Delhi India). Dr Matt Higgins (University of Oxford)

kindly supplied us with the ATS PFF0845c full-length clone in the modified bacterial expression vector pET15 (Novagen). The vector contains a N-terminal His_6-tag and a TEV cleavage site. A *P. falciparum* cDNA library provided by the Malaria Research and Reference Reagent Resource Centre (MR4, ATCC, USA) and *P. falciparum* genomic DNA, provided by Dr Sue Kyes (University of Oxford), were used to amplify the other ATS encoding DNA molecules.

2.7 ATS protein purification under native conditions

ATS PF08_0141 and PFF0845c in the expression vectors pPET28a (C-terminal His_6-tag) and pET15 (N-terminal His_6-tag), respectively, were expressed as described. Cell pellets were resuspended in approximately 30 ml PBS (150 mM NaCl, 20 mM Na_2HPO_4 at pH 7.4) per litre of expression culture supplemented with one complete EDTA free protease inhibitor cocktail tablet (Roche). Cells were lysed by repeated thaw-freeze cycles in liquid nitrogen, DNA was cleaved by addition of 3 units / ml of DNAseI (Sigma), and lysates were centrifuged at 30,000 g at 4 °C for 30 minutes. The supernatant was then used for further purification using a talon metal affinity resin (Clontech) column. Talon beads were equilibrated beforehand with lysis buffer. For elution 500 mM imidazole was added to the lysis buffer. After dialysis into 20 mM Tris-HCl, 1 mM DTT, 5 mM EDTA (0.5 mM EDTA in the case of PFF0845c) for pH 7.8, the His_6-tag was cleaved off by 3C protease or the TEV protease for the PFF0845c variant, overnight at 4 °C. Further purification was achieved by anion exchange chromatography using MonoQ HP columns (GE Healthcare) in the same buffer. ATS was eluted via a NaCl gradient (up to 500 mM) and then further purified by gel filtration chromatography. For this step a HiLoad 26/60 Superdex 75 (GE

Healthcare) was equilibrated in PBS or NMR buffer (50mM NaCl, 20 mM Na_2HPO_4 pH 7.0, 2 mM DTT) on an AKTA FPLC system. Finally, the protein was concentrated with an Amicon ultra centrifugal filter device (Millipore).

2.8 ATS purification from inclusion bodies

ATS PF08_0141, PFB1055c, PFF0010w, PFC1120c, PF08_013 and their shorter fragments were cloned into a pET16b vector (N-terminal His_{10}-tag) that had been previously modified with a 3C protease cleavage site (Anthis, 2009). These proteins were expressed into inclusion bodies and purified under denaturing conditions. The pellets were resuspended in 30 ml (per 1 litre original culture volume) of denaturing lysis buffer (8M Urea, PBS pH 7.4). Cell lysis was performed by repeated freeze-thaw steps in liquid nitrogen and the DNA was sheared by sonication. Centrifugation at 30,000 g at 4 °C for 30 minutes subtracted the cell debris; the supernatant was exposed to a Talon metal affinity resin (Clontech) equilibrated in lysis buffer. The ATS variants were eluted in the same buffer containing 500 mM imidazole (Sigma). Afterwards, proteins were refolded via extensive dialysis against 20 mM Tris-HCl, 1 mM DTT, 5 mM EDTA pH 7.8. Further purification was performed as described above for native conditions (section 2.7).

2.9 Fluorescent labelling of ATS

The Q150C ATS-FL mutant was used for attaching a fluorescent molecule to the protein. Purification was done under denaturing conditions as described above with the difference that gel filtration buffer did not contain DTT. For protection of the reactive cysteine group, tris(2-carboxyethyl)phosphine was added to the protein

solution to maintain its reduced state. Labelling was achieved by addition of a 10-fold excess of fluorescein-5-maleimide (Invitrogen) and incubation at 4 °C overnight. Unbound fluorescent dye was removed by size exclusion chromatography in NMR buffer (50mM NaCl, 20 mM Na_2HPO_4 pH 7.0, 1 mM DTT).

2.10 ATS phosphorylation

Recombinant casein kinase II (CK II) (NEB) and adenosine 5'- triphosphate (ATP) (Sigma) were used for *in vitro* phosphorylation ATS-N according to the manufacturer's recommendations. The proteins were dialysed into the recommended CK II reaction buffer (NEB); the enzymatic reactions took place for 16 hours at 30 °C.

2.10.1 ATS phosphorylation in an NMR tube

100 μM ^{15}N-enriched ATS-N or ATS-FL in 20 mM Na_2HPO_4 buffer pH 7.0 supplemented with 100 μM $MgSO_4$ samples were prepared for *in vitro* phosphorylation monitored by NMR. After initial ^{1}H- ^{15}N- HSQC recordings, 1500 units of CK II and 300 μM ATP were added to start the enzymatic reaction; consecutive ^{1}H- ^{15}N- HSQC spectra were acquired at 20 °C on a Bruker 500 MHz NMR spectrometer.

2. 11 Production and purification of KAHRP fragments

All KAHRP constructs were obtained by PCR amplification using the KAHRP clone (MRA-6) form MR4 (ATCC, USA). Fragment positions, lengths and amino acid compositions are given in Appendix III and have been described previously (Waller

et al., 1999). The fragments of interest were transferred into the pGEX-6P-2 vector (GE LifeSciences) containing an N-terminal GST-tag and transformed into *E. coli* BL21(DE3) cells. Cell pellets were resuspended in PBS at pH 7.4. After initial cell lysis by sonication and centrifugation (as described above), the GST-fusion proteins were separated from other cell products using glutathione sepharose beads (GE Healthcare) equilibrated in PBS buffer. After a washing step with PBS, proteins were eluted in 50 mM Tris-HCl pH 7.7, 12mM reduced Glutathione buffer. The GST fusion tag was removed by overnight cleavage with 3C protease (recombinant, home-made) at 4 °C. For some KAHRP K1A samples the GST-tag was not removed in order to increase the protein solubility. The protein was then further purified and separated from the fusion tag by using a Sephadex G-20 desalting column, a GSTrap column (Amersham Biosciences), and a subsequent gel filtration chromatography step in NMR buffer, using a HiLoad 26/60 Superdex 75 column (GE Healthcare).

2.12 Production and Purification of PFI1780w

Residues 80–280 of PFI1780w were cloned in pGEX-6P-2 using standard techniques and the protein was expressed and purified in a similar way to the KAHRP fragments.

2.13 Bioinformatic tools

Different online tools (Table 2.1) were utilized for amino acid composition, secondary structure and disorder predictions.

Tool	Website	Reference	Application
Globplot	http://globplot.embl.de	(Linding et al., 2003b)	protein disorder predictions
RONN	http://www.strubi.ox.ac.uk/RONN	(Yang et al., 2005)	protein disorder predictions
DISOPRED2	http://bioinf.cs.ucl.ac.uk/dispred	(Ward et al., 2004)	protein disorder predictions
DisEMBL	http://dis.embl.de	(Linding et al., 2003a)	protein disorder prediction
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/	(Jones, 1999)	secondary structure predictions
PROTPARM	http://web.expasy.org/protparam/	(Gasteiger E., 2005)	molecular weight and pI determination
SMART	http://smart.embl-heidelberg.de	(Letunic et al., 2012; Schultz et al., 1998)	analysis of protein domain architecture
Phyre^2	http://www.sbg.bio.ic.ac.uk/phyre2/	(Kelley and Sternberg, 2009)	protein tertiary structure predictions
MULTALIN	http://bioinfo.genotoul.fr/multalin/multalin.html	(Corpet, 1988)	alignment of protein sequences
ESPRIT	http://espript.ibcp.fr/ESPript/ESPript/	(Gouet et al., 1999)	formatting of protein sequence alignments

Table 2.1: Overview of utilised bioinformatic online tools.

2.13.1 Meta-structure analysis

These studies were kindly performed by Prof. Robert Konrat (University of Vienna, Austria). ATS and KAHRP were analysed to explore their secondary structure propensity and identify protein interaction epitopes (Konrat, 2009). He introduced the concept of a 'meta-structure' where the protein structure is viewed as an intricate network of interacting residues (meta- is from the Greek: μετά = "after" or "beyond"). It provides quantitative parameters on compactness, local secondary structure and protein topology. The residue-specific compactness value quantifies the structural complexity of an individual residue in the context of the 3D protein fold, with residues deeply buried in the interior of a structure exhibiting large 'compactness' values. The meta-structure derived secondary structure parameter is defined in analogy to the well-established $^{13}C_\alpha$ NMR secondary chemical shifts, with positive values for a α-helix and negative values indicating the presence of an extended conformation. Details of the calculation procedure have already been described (Konrat, 2009). The availability of quantitative topology information on a per-residue-basis also provides the possibility of performing protein homology searches that go beyond the primary sequence level. This aspect was exploited in a sequence-derived protein-protein interaction site analysis. In brief, sequences of about 1750 protein complexes (3500 individual sequences) were extracted from the PDB database together with the information about residues that form the interaction interface. Residues located in interaction sites were defined based on the criteria that the C_α-C_α distances to the nearest neighbour residues of the binding partner protein are below a cutoff of 8Å. The selected sequences were subjected to a meta-structure calculation. The rationale behind the prediction algorithm is the notion that there is a (limited) set of protein interaction motifs with distinct meta-structural properties.

Despite their different 3D structures proteins exploit similar meta-structural features to accommodate a specific (protein) binding partner. These meta-structure similarities can be quantified and used as the basis for the binding site prediction.

Putative interaction sites are predicted using the following strategy: The sequence of a query protein is aligned with a template sequence from the dataset based on meta-structure similarity. Residues of the query protein aligning with template residues that are part of interaction sites are allocated weights corresponding to the alignment similarity scores. The summation over 3500 individual alignments thus provides a reliable estimation of the protein interaction propensity (protein interaction index - PII). Details of the results together with a statistical analysis will be given elsewhere (*Konrat in preparation*). Briefly, a preliminary statistical survey of the sequence-derived interaction prediction results obtained for the selected sequences (in total about 3500) revealed that about 80% of the interaction site residues have protein interaction scores larger than 100. Protein interaction indexes beyond 1000 are therefore considered to be highly significant.

2. 14 NMR spectroscopy

NMR spectroscopy is based on a quantum mechanical phenomenon of the nucleus called spin (I). The nuclear magnetic moment μ is given by $\mu = \gamma Ih$, where γ stands for the gyromagnetic ratio and h for Planck´s constant. Nuclei possessing a spin of ½ are important in NMR. In biological relevant studies these are mainly ^{1}H, ^{15}N, ^{13}C and ^{31}P. A spin (I) of magnetic quantum number m = ½ can take two different spin states which correspond to two different energy levels: One is termed parallel or α with

m=½; the other is the β antiparallel state with m= - ½. A transition between those states emits or absorbs energy in form of electromagnetic radiation in the radio frequency (*rf*) range. The radio frequency needed for a transition is between 50 and 800 MHz and must match the energy difference. A NMR signal results form the energy difference between the transitions from the lower to the higher state and the one from the higher to the lower. Therefore, only the small difference of populations gives rise to a signal, which makes NMR a very insensitive method. In equilibrium there is small excess population of the lower level, which is expressed by the Boltzmann equation:

$$N^{-}/N^{+} = e^{-E/kT}$$

N^{+} refers to the spins in the lower energy level and N^{-} represents the number of spins in the higher level.

The equilibrium is disturbed by *rf* pulses, which rotate the magnetisation vector by an angle. In NMR experiments one typically uses 90° and 180° pulses, which correspond to defined rotations, achieved by *rf* pulses, whose duration is calibrated. After applying a pulse, the system will return back to the equilibrium state and therefore induces current in the receiver, which is recorded as free induction decay (FID). This time domain function is then transferred to a frequency domain function, the NMR spectrum, by a Fourier transformation. The features of magnetic nuclei - chemical shift, spin coupling and relaxation - reveal essential information gained by NMR.

2.14.1 Chemical Shift

A nucleus is surrounded by electrons, which circulate due to the external applied magnetic field and produce their own small magnetic fields. Thus, each nucleus has a unique chemical environment and effective magnetic field. This is the reason for the distinct absorption frequencies of the different nuclei. If the electron density near the relevant nucleus is high, the attenuation of the magnetic field is stronger (strong shielding). Thus, the resonance of this nucleus is shifted to lower frequencies. The chemical shift indicates the NMR absorption frequency (v) relative to a reference compound (v_{Ref}). The quantity is reported in parts per million (ppm) and given the symbol δ.

$$\delta_{ppm} = 10^6 * (\nu - v_{Ref}) / v_{Ref}$$

As there are many nuclei in a protein, resonances tend to overlap in one – and two - dimensional NMR spectra. This effect can be reduced by adding more dimensions.

2.14.2 Spin – Spin Coupling

It is important to obtain as many of chemical shift assignments as possible in order to calculate an accurate three-dimensional structure. Therefore, ^{13}C and ^{15}N labelled molecules are produced to apply different three-dimensional experiments, which transfer magnetization. Generally, ^{15}N, ^{13}C and ^{1}H are correlated. Triple resonance experiments are essential for backbone assignment of proteins, as they correlate different backbone nuclei. There are two spin coupling mechanisms causing coherence transfer: scalar coupling and dipolar coupling.

2.14.2.1 Scalar Coupling

The interaction of nuclei through-bond is termed scalar coupling, or *J* coupling. The magnetization is transferred via polarization of bonding electrons. *J* couplings over more than three bonds are normally too small to be observed. However, for proteins, which are labelled with ^{15}N and ^{13}C, *J* couplings between ^{1}H, ^{15}N and ^{13}C allow correlations across the peptide bond. Correlation Spectroscopy (COSY) experiments transfer magnetization to different nuclei via J coupling.

2.14.2.2 Dipolar Coupling

Dipolar Coupling stands for the direct interaction of two or more nuclei through space. The correlations between protons in space are crucial for three-dimensional structure determination. They are measured via the nuclear Overhauser effect (NOE). The NOE reflects the magnetization transfer between spins interacting via their associated dipoles. NOEs connect hydrogen atoms separated by less than 5- 6 Å. The intensity of an NOE between two proton spins i and j, separated by the distance r is given by:

$$NOE_{ij} \sim 1/r_{ij}^{6}$$

Accordingly, the intensity of the NOE decreases rapidly with increasing distance. Hence, as the intensity of the correlation is dependent on distance; the NOE gives information about the distance between atoms for subsequent structure calculations. In a folded protein, a proton is surrounded by numerous protons, many of which are not from the neighbouring residues. Therefore, many NOE correlations can be measured and give essential structural information.

For a more detailed introduction to NMR and further understanding, the reader is referred to previously published literature (Cavanagh; Keeler; Pascal, 2008).

All experiments were performed at 25 °C or 10 °C (unless otherwise stated) using NMR spectrometers, with 11.7 T, 14.1 T, 17.6 T or 22.3 T field strengths or a Bruker 11.7 T spectrometer equipped with cryogenic probehead. Unless otherwise indicated, samples were prepared in NMR buffer with 5% v/v D_2O and 50 – 100 μM DSS for ^{1}H spectral referencing. The transmitter ^{1}H frequency was set to the water resonance. ^{15}N and ^{13}C chemical shifts were indirectly referenced to ^{1}H by using the recommended frequency ratios: $^{15}N/\ ^{1}H$ ratio of 0.101329118 and $^{13}C\ /\ ^{1}H$ ratio of 0.251449530 (Markley et al., 1998). For chemical shift assignments and structural restraints $^{15}N/\ ^{13}C$-enriched samples of 0.5 mM to 1 mM concentration were used. Data processing was done using NMRpipe (Delaglio et al., 1995); for further analysis the programs SPARKY (Kneller) and PIPP (Garrett et al., 1991) were used. Subsidiary information from the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu) was incorporated, which offers average amino acid ppm values for resonance assignment. For the identification of unstructured elements, typical random coil chemical shifts were employed for $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$ and ^{13}CO (Appendix V). NMR spectra were recorded by or under the supervision of Dr Ioannis Vakonakis (Univsersity of Oxford).

2.14.3 Sequential backbone assignment

Backbone chemical shift assignments were done using standard triple resonance experiments (Cavanagh 1996). In addition to CBCA(CO)NH (Grzesiek and Bax, 1993) and CBCANH (Grzesiek et al., 1992), HBHA(CO)NH, HBHANH, HNCO, HCACO and HCA(CO)N experiments (Ikura et al., 1990) were acquired due to the lack of chemical shift dispersion of ATS.

2.14.4 Side chain chemical shift assignments and structure calculations

Structure calculations were performed by simulated annealing using the Xplor-NIH software package (Schwieters et al., 2003). ATS-core calculations were based on established simulated annealing protocols. Three-dimensional (3D) (H)C(CCO)NH, 3D H(CCCO)NH, 3D ^{1}H- ^{15}N NOESY-HSQC, 3D ^{15}N HNHA, 3D ^{1}H- ^{13}C NOESY-HSQC (Marion et al., 1989), HCCH-COSY (Clore, 1990) and a 4D ^{13}C- ^{13}C NOESY-HSQC were employed for side chain assignment and collection of NOE-derived distance restraints. Dihedral restraints (ϕ/ψ) were obtained from the TALOS database (Cornilescu et al., 1999). Obtained chemical shift assignments have been deposited in the BioMagResBank (RMRB - http://www.bmrb.wisc.edu) under the accession numbers 16911 and 17999 for the ATS-core and ATS-FL, respectively. In addition, a set of lowest energy structures and structure calculation restraints for the ATS-core can be found at the Protein Databank (PDB – http://www.rcsb.org/pdb) under the accession number 2LKL.

2.14.5 ^{1}H- ^{15}N HSQC titration studies for ATS

^{15}N enriched proteins and unlabelled protein were both prepared in or dialyzed into NMR buffer. In the course of the titration, the desired protein-ligand ratios were calculated and the pH values were checked. ^{1}H- ^{15}N HSQC spectra were recorded for each point in a titration series and analyzed as described before.

2.15 Crystallisation of ATS-core

The favourable protein concentration of 12 mg/ml of ATS-core for protein crystal formation was determined by using Pre-Crystallisation Test (HAMPTON Research) at 10 °C and 20 °C. 200 nl sitting drops containing 100 nl ATS-core (in 50 mM Tris, 50 mM sodium chloride and 1 mM DTT pH 7.3) and 100 nl precipitant were set-up in 96-well Greiner crystallisation plates using a Mosquito (TTP LabTech) robot. Various precipitants were screened in parallel at 4 °C and 21 °C.

2.16 Small angle X-ray scattering

Dr. Ioannis Vakonakis and Ms. Leanne Slater collected small angle X-ray scattering (SAXS) data on the BioSAXS beamline at the ESRF (ID14-3). All ATS variants were measured in PBS buffer at an X-ray wavelength of 0.931 Å and a temperature of 20 °C. An ultracentrifugation step was performed at 189,000 g for 60 minutes beforehand to ensure sample homogeneity, crucial for SAXS measurements. In addition, 5 mM DTT was added shortly before radiation exposure to sequester radicals created by X-rays. Bovine serum albumin (BSA) was measured as a reference for determination of the association states of the proteins. Experiments were carried out at four protein concentrations (2.5 mg/ml, 5 mg/ml, 10 mg/ml, 15 mg/ml).

The scattering intensity was recorded as a function of the momentum transfer, $q = \frac{4\pi \sin\theta}{\lambda}$, where λ is the radiation wavelength, 2θ is the scattering angle.

Subsequent data processing and merging was done using the program PRIMUS (Konarev et al., 2003) from the ATSAS package (Petoukhov et al., 2007) and the scattering profile of the buffer was subtracted. The radius of gyration R_g was deducted using the Guinier approximation (Guinier, 1955). At small angles ($qR_g<1$), the intensity I (q) is

$$I(q) \approx I(0)\exp(-q^2 Rg^2/3)$$

The first parameter characterizes the particle size and the second one its molecular mass. In addition, the R_g was evaluated using the indirect transformation package GNOM (Svergun, 1992), which provides the distance distribution function p(r)

$$p(r) = \frac{r^2}{2\pi^2}\int_0^\infty I(q)\frac{\sin(qr)}{qr}q^2 dq$$

GNOM determines R_g as the second moment of p(r), given by

$$R_g = \sqrt{\frac{\int_0^{D_{\max}} r^2 p(r)dr}{2\int_0^{D_{\max}} p(r)dr}} \quad \text{and} \quad I(0) = 4\pi\int_0^{D_{\max}} p(r)dr$$

The conformation of in solution of different ATS constructs was determined using two *ab initio* approaches: DAMMIN (Svergun, 1999) and GASBOR (Svergun et al., 2001). Both programmes were run using standard parameters. The ensemble optimisation method (EOM) (Bernado et al., 2007) was used for selection of the molecular models that fit the SAXS data (for ATS-25 and ATS-FL). Comprehensive introductions to SAXS have been given in several publications (Bernado and Svergun, 2012; Mertens and Svergun, 2010).

2.17 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) data were collected using a Beckman Optima XL-I analytical ultracentrifuge. Sedimentation equilibrium experiments were performed at 4°C for 48 hours at 15,000 rpm. The protein samples were in PBS buffer at a concentration of 30 μM. The concentration profile was monitored using absorbance at 280 nm; achievement of equilibrium was confirmed by monitoring the profile. Collected data were analyzed and fitted to an ideal monodisperse model with the program Origin (OriginLab).

2.18 Circular dichroism

CD Spectra and Thermal stability measurements were recorded on a Chirascan spectropolarimeter (AppliedPhotophysics, UK) in a 1 mm path length quartz cuvette at concentrations of 10-25 μM protein. Measurements were conducted in PBS with 1mM DTT at pH 7.4. Thermal stability experiments were performed over the temperature range 15 °C to 90 °C with a constant increase of 1 °C / minute. Protein unfolding was observed by CD at 222 nm. Collected data were plotted as a mean residue elipticity (MRE) and wavelength (200 – 260 nm) or temperature [°C], respectively.

2.19 Mass spectrometry

MALDI – TOF Mass Spectrometry experiments were kindly performed by Dr. David Staunton (University of Oxford).

2.20 Fluorescence anisotropy

All measurements were performed on a PHERAstar FS microplate reader (BMG Labtech) in 96-well microplates in NMR buffer and total sample volumes of 140 µl at 25 °C. The instrument was configured for excitation and emission wavelengths of 485 nm and 520 nm. Fluorescein-labelled Q150C ATS-FL was kept at a constant concentration of 0.5 µM; whereas the concentration of different KAHRP fragments (K1, K2, K3, K1A, K2A) and PFI1780w were stepwise increased up to 300 µM. Recombinant expressed GST was used as a negative control. For the deduction of KD-values, binding isotherms from experiments were fitted to the following equation:

$$S = S_0 + (S_{\max} - S_0)\frac{([P_{tot}] + [L_{tot}] + K_D) \pm \sqrt{([P_{tot}] + [L_{tot}] + K_D)^2 - 4[P_{tot}][L_{tot}]}}{2[P_{tot}]}$$

where S (experimental signal) corresponds fluorescence anisotropy. $[P_{tot}]$ and $[L_{tot}]$ denote the total concentration of protein and ligand, respectively.

2.21 Fluorescence microscopy

EGFP-pLEXm constructs and COS-7 cell assays were visualized using a Nikon Eclipse TE20000U fluorescence inverted microscope, equipped with a Hamamatsu C4742-95 Ocra low-noise B&W CCD. Images were taken and refined with IPLab imaging software.

2.22 References

Anthis, N.J. (2009). Structural Studies of Integrin Activation. In Department of Biochemistry (Oxford: University of Oxford).

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta crystallographica Section D, Biological crystallography *62*, 1243-1250.

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc *129*, 5656-5664.

Bernado, P., and Svergun, D.I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Molecular bioSystems *8*, 151-167.

Cavanagh, J., Fairbrother, W., Palmer A. G., Skelton N. J (2007). Protein NMR Spectroscopy - Principles and Practise (Academic Press).

Clore, G.M. (1990). Practical Aspects of Proton-Carbon-Carbon-Proton Three Dimensional Spectroscopy of 13C-Labeled Proteins. Journal of Magnetic Resoncance *87*, 620-627.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of biomolecular NMR *13*, 289-302.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic acids research *16*, 10881-10890.

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

Endo, Y., and Sawasaki, T. (2005). Advances in genome-wide protein expression using the wheat germ cell-free system. Methods Mol Biol *310*, 145-167.

Endo, Y., and Sawasaki, T. (2007). [Wheat germ cell-free protein synthesis]. Seikagaku The Journal of Japanese Biochemical Society *79*, 229-238.

Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, Humana Press (2005) (Humana Press), pp. 571 - 607.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305-308.

Grzesiek, S., and Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. Journal of biomolecular NMR *3*, 185-204.

Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A.M., and Bax, A. (1992). 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. Biochemistry *31*, 8180-8190.

Guinier, A., and G. Fournet (1955). Small Angle Scattering of X-Rays (New York: Wiley).

Ikura, M., Kay, L.E., and Bax, A. (1990). A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry *29*, 4659-4667.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology *292*, 195-202.

Keeler, J. Understanding NMR Spectroscopy (Irvine: University of California).

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols *4*, 363-371.

Kneller, T.D.G.a.D.G. SPARKY 3.

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

Konrat, R. (2009). The protein meta-structure: a novel concept for chemical and molecular biology. Cellular and molecular life sciences : CMLS *66*, 3625-3639.

Letunic, I., Doerks, T., and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. Nucleic acids research *40*, D302-305.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003a). Protein disorder prediction: implications for structural proteomics. Structure *11*, 1453-1459.

Linding, R., Russell, R.B., Neduva, V., and Gibson, T.J. (2003b). GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic acids research *31*, 3701-3708.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., and Clore, G.M. (1989). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry *28*, 6150-6156.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., and Wuthrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. European journal of biochemistry / FEBS *256*, 1-15.

Mertens, H.D., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. Journal of structural biology *172*, 128-141.

Pascal, S. (2008). NMR Primer: An HSQC-based approach (IM Publications).

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., and Svergun, D.I. (2007). ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. J Appl Crystallogr *40*, S223-S228.

Russell, S.a. (2000). Moleular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

Sawasaki, T., Gouda, M.D., Kawasaki, T., Tsuboi, T., Tozawa, Y., Takai, K., and Endo, Y. (2005). The wheat germ cell-free expression system: methods for high-throughput materialization of genetic information. Methods Mol Biol *310*, 131-144.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America *95*, 5857-5864.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003). The Xplor-NIH NMR molecular structure determination package. J Magn Reson *160*, 65-73.

Svergun, D.I. (1992). Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. J Appl Crystallogr *25*, 495-503.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophysical journal *76*, 2879-2886.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophysical journal *80*, 2946-2953.

Walker, J.M. (2005). The Proteomics Protocols Handbook (Humana Press).

Waller, K.L., Cooke, B.M., Nunomura, W., Mohandas, N., and Coppel, R.L. (1999). 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 *274*, 23808-23813.

CHAPTER THREE

ATS STRUCTURAL STUDIES

3.1 Introduction

As mentioned in Chapter one, limited information is available on ATS. Previous published work describes the domain to be largely composed of β-sheets and random coils (Hora et al., 2009). Moreover, ATS-FL (ATS full length) and truncated variants thereof are reported to dimerise in solution (Hora et al., 2009). Previous attempts to obtain a high-resolution structure of ATS by crystallographic studies have been unsuccessful (personal communication, Dr. Amit Sharma, ICGEB). Therefore, no structural information at atomic level is available at present. Our aim was to gain as much insight as possible into the intracellular events occurring in cytoadhering erythrocytes. Therefore, we started to study ATS-FL from a structural perspective.

For our initial structural studies of ATS we used the *Pf*EMP1 variant PF08_0141 (plasmodb.org). Throughout this thesis the residue numbering of this and all other ATS variants starts after the predicted transmembrane segment. In most cases ATS begins with three lysine residues. Therefore, the residues 2466 – 2858 define the ATS-FL of PF08_0141. All relevant protein sequences can be found in Appendix III.

3.2 Bioinformatic analysis of ATS (Sequence-based analysis)

The primary structure of ATS-FL (PF08_0141) is composed of 393 amino acids, of which 77 are negatively charged. The domain is highly acidic. Previous studies (Hora et al., 2009) and disorder predictions (figure 3.1 A and B) suggest that ATS is not a well-folded domain. On the contrary, the prediction tools anticipate flexible stretches on both ends of the domain. Moreover, a meta-structure analysis was conducted, which extracts two parameters for each residue on the basis of the protein sequence; one for compactness and the other for secondary structure (Konrat, 2009). The compactness parameter (C_i) gives a quantitative probability measure for residue burial in the tertiary structure of the protein. High values (> 300) are attributed to residues in the buried core of globular proteins. Low values correspond to surface exposed or flexible residues. The compactness analysis for ATS-FL does not attribute significantly high C_i values for longer sequence stretches (Figure 3.1 C). Much of the ATS sequence falls between the average values calculated for proteins in the PDB and the upper threshold for intrinsically unstructured proteins. The average compactness value (C_{av}) for ATS-FL is ~200, which is comparable to average compactness values (C_{av}) determined for several intrinsically disordered proteins (Bertoncini et al., 2005; Mittag et al., 2010; Mukrasch et al., 2009; Schedlbauer et al., 2008).

Secondary structure prediction results are shown in figure 3. 2. Extended β- sheets are predicted for multiple stretches of ATS-FL, which is in agreement with previous analysis (Hora et al., 2009). In addition, the different prediction algorithms suggested three longer α- helical regions in the middle part of the domain, which have not been reported previously.

The Simple Modular Architecture Research Tool (SMART) does not identify any conserved domains in the primary structure of ATS-FL (Schultz et al., 1998). Phyre^2, a remote homology recognition service able to generate reliable protein models, also did not offer any significant results (Kelley and Sternberg, 2009).

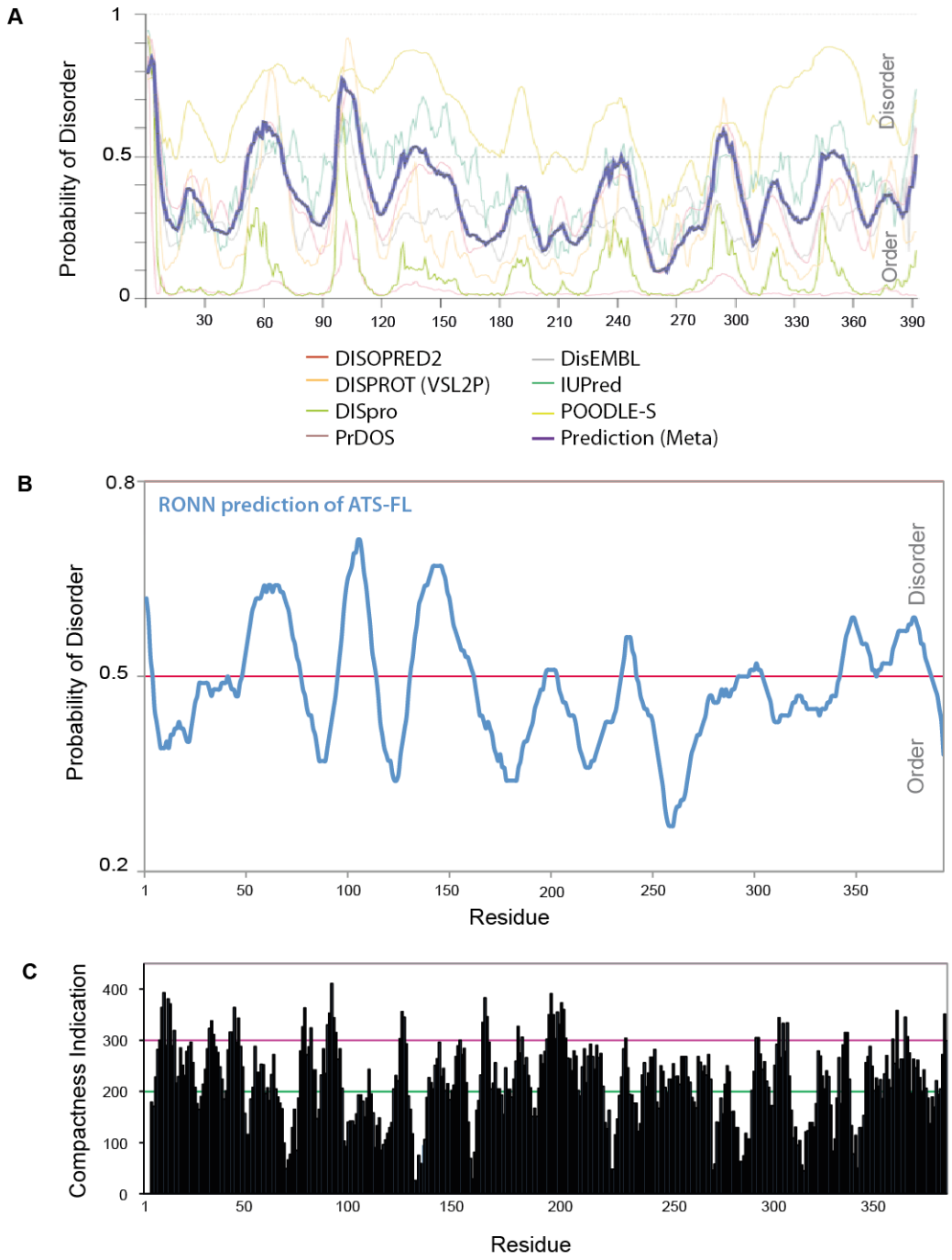


Figure 3.1: ATS-FL based sequence based analysis. **A)** Disorder predictions determined by the indicated algorithms. (Ishida and Kinoshita, 2008) (Cheng et al., 2005; Dosztanyi et al., 2005; Linding et al., 2003; Ward et al., 2004; Xue et al., 2010) **B)** RONN disorder prediction of ATS-FL (Yang et al., 2005). The threshold (0.5) for disorder probability is marked in magenta. Values above have a significant probability of disorder. **C)** Meta structure analysis (Konrat, 2009) – compactness indication per residue. Globular folded proteins have compactness values above 300 (magenta line). Small values predict exposure to solvent, such as residues located on the surface or on flexible loops. The green line represents the upper threshold for disordered proteins.

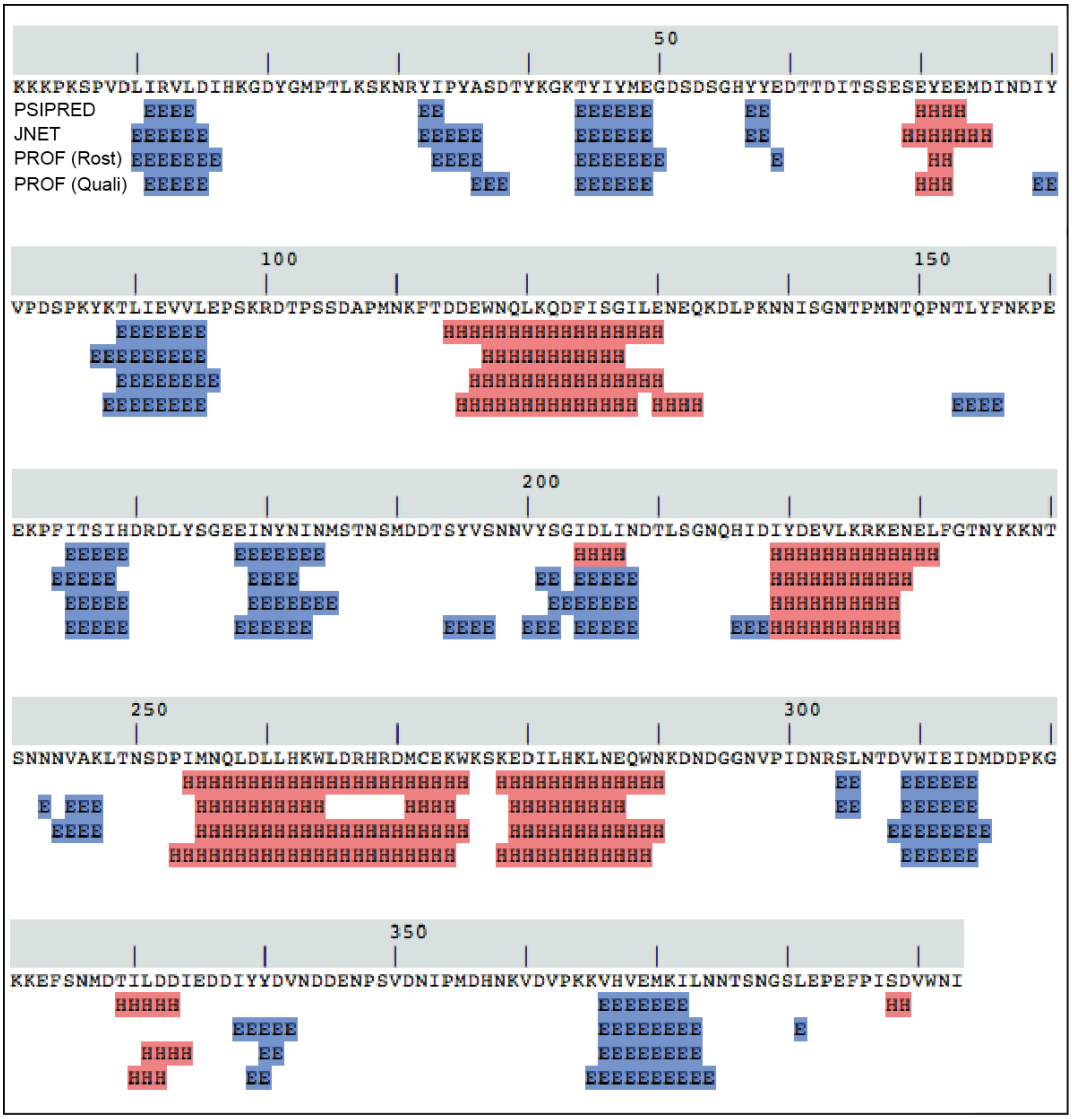


Figure 3.2: Secondary structure prediction for ATS-FL. α-helices predictions are indicated in red and extended (sheet) elements in blue. PSIPRED, JNET, PROF (Rost, Ouali) online prediction tools have been employed (Cuff and Barton, 1999; Jones, 1999; Rost and Liu, 2003); listed from top to bottom, respectively.

3.3 Expression and purification of ATS-FL

3.3.1. Expression in *E. coli*

The pET28a plasmid coding for the ATS-FL (PF08_0141) was obtained from the laboratory of Dr Amit Sharma (New Delhi, India). Initially, the expression of ATS-FL was performed in *E. coli* BL21(DE)RIPL cells (Stratagene). The genome of *P. falciparum* has an AT content of approximately 80 % (Gardner et al., 2002). However, comparing expression yields to standard *E. coli* BL21 (DE) cells did not show any significant difference (data not shown). Therefore, routine expression of ATS-FL was performed in the latter strain.

Initially, ATS-FL purification was performed under native conditions with addition of protease inhibitor cocktail tablets (Roche) to decrease protein degradation. Despite the presence of the inhibitors ATS-FL proteolysis was substantial. Therefore, the DNA segment coding for ATS-FL was cloned in a modified pET16b vector (Novagen) using the restriction enzymes BamHI and XhoI. This plasmid has been optimized to increase protein yields and contains a 3C protease cleavage site to separate the protein product from an N-terminal His_{10}-tag (see Appendix II). For the purification of ATS-FL expressed in pET16b-3C, the first step was performed under denaturing conditions. Cell pellets were resuspended in PBS buffer pH 7.4 containing 8 M urea. After cell lysis and separation from cell debris, metal ion affinity chromatography was conducted still under denaturing conditions. Subsequently, extensive dialysis was used to transfer the fusion protein into native conditions and the His-tag was cleaved off. Protein purity was increased by sequential anion-exchange chromatography and gel filtration (Figure 3.3). Purity and degradation state

of ATS-FL was determined by SDS-PAGE and suitable fractions were pooled and used for further studies.

To produce isotope-enriched $^{15}N/^{13}C$ ATS-FL for NMR studies, over-expression was performed in M9 minimal medium under the same conditions as in LB medium and purification was done as described above.

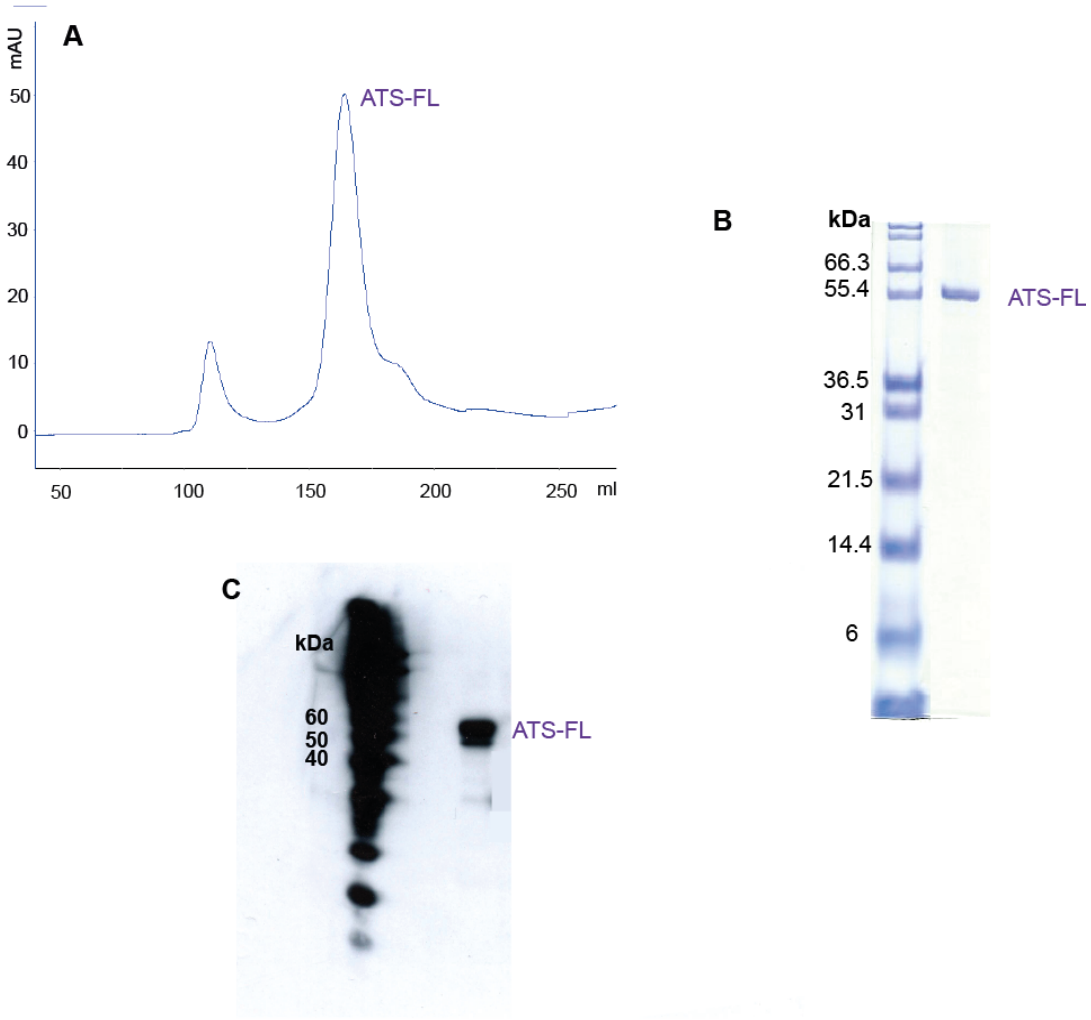


Figure 3.3: PF08_0141 ATS-FL expression and purification. **A)** Gel filtration profile on Sephadex G-75 column of ATS-FL with Mw 45kDa. **B)** Purified ATS-FL loaded on 10 % SDS-PAGE (after final gel filtration purification step). **C)** Western Blot of ATS-FL expression using wheat germ cell free expression system.

3.3.2 *In vitro* protein expression

As an alternative to recombinant protein expression in *E. coli,* we aimed to produce ATS using a eukaryotic expression system. Eukaryotic translation systems have been reported to facilitate co-translational domain folding and peptide growth tends to be five to ten times slower in eukaryotes compared to bacterial translation (Hartl and Hayer-Hartl, 2002; Netzer and Hartl, 1997). We choose the cell free wheat germ coupled transcription/translation system ENDEXT Technology (Endo and Sawasaki, 2007). This system has been used successfully for the production of various *Plasmodium* proteins without the need for codon optimization (Tsuboi et al., 2008). In addition, the *in vitro* system allows isotope labelling for subsequent NMR studies as well as selective amino acid labelling. The ATS-FL was cloned into the pEU vector to obtain the desired protein fused to a N-terminal His_6 tag. We started with a small-scale expression test using a bilayer translation reaction method at 16 °C for 20 hours. Successful initial expression of ATS-FL was visualized by Western blotting (figure 3.3.C). However, protein degradation still occurred and scaling- up attempts to obtain a sufficient ^{15}N labelled ATS-FL NMR sample remained unsuccessful.

To summarize, purification in *E. coli* under denaturing conditions proved to be successful and refolding produced the same protein state as natively purified *E. coli*-made ATS-FL. Figure 3.9 compares ^{1}H-^{15}N HSQC spectra for the domain purified under the two different conditions. Since we could not observe any differences, we concluded that our refolding strategy was effective.

3.4 Experimental characterisation of ATS-FL

Far UV-CD spectroscopy in the range of 190- 240 nm gives valuable information on secondary structure elements in proteins. Figure 3.4 (A) displays the characteristic CD spectra for alpha helix, beta sheet and random coil conformations in proteins. α-helical secondary structure gives rise to two distinctive minima near 208 nm and 222 nm in the CD spectrum. β- sheet and random coil structure minima at 215 nm and near 200 nm can be observed, respectively. An actual protein spectrum can be approximated as a linear combination of the contributions of different structural elements. Previous CD studies of ATS-FL attributed the domain to contain beta sheets and random coils (Hora et al., 2009). However, our recorded CD spectrum of ATS-FL shown in Figure 3.4 (B) suggests otherwise. A characteristic spectrum of an extended coil protein with a minimum at 205 nm was obtained. The absence of a compact molecular architecture helps to explain the considerable proteolytic degradation encountered during protein preparation. Moreover, the CD spectra display a secondary minimum at 222 nm, thereby suggesting an alpha helical propensity for the domain. We also performed thermal stability experiments using CD to study the properties of the structured helical part of ATS-FL and its extended flexible character. The cooperativity of the thermal unfolding reaction can be interpreted qualitatively by the width and shape of the transition. A highly cooperative thermal unfolding indicates that the protein is a folded compact one, while a gradual melting curve suggests that the protein initially existed in an unfolded flexible state. We chose to study the behaviour of ATS-FL at a wavelength $\lambda = 222$ nm during a linear temperature increase to 90 °C to monitor the α-helical feature of the protein. The α-helical signal indicates a decrease upon thermal unfolding, which

again indicates the existence of structured elements in the ATS-FL domain (figure 3.5).

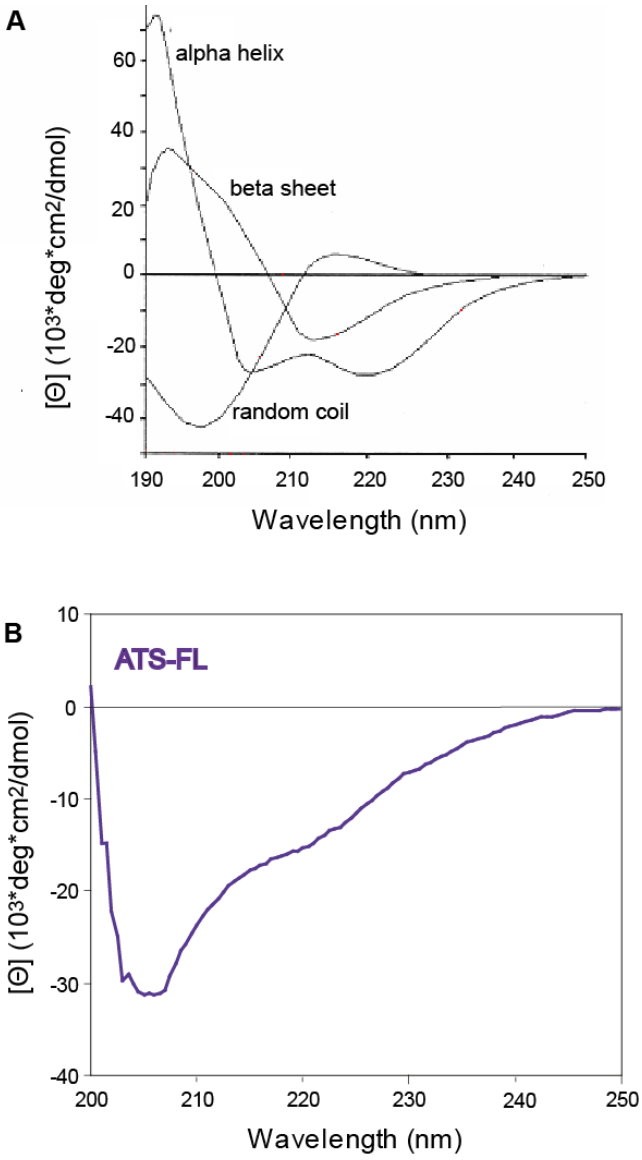


Figure 3.4: CD spectra of ATS-FL. A) Typical CD spectra of poly-L-lysine in α-helix, β-sheet and random coil conformations. **B)** The purple curve represents the CD spectrum of ATS-FL. The spectrum shows evidence for some α-helical content and mainly random coil character with characteristic signals at 222 nm and 205 nm, respectively. Measurements were performed in PBS buffer (containing 1 mM DTT) pH 7.4 at 10°C at a protein concentration of 25 μM.

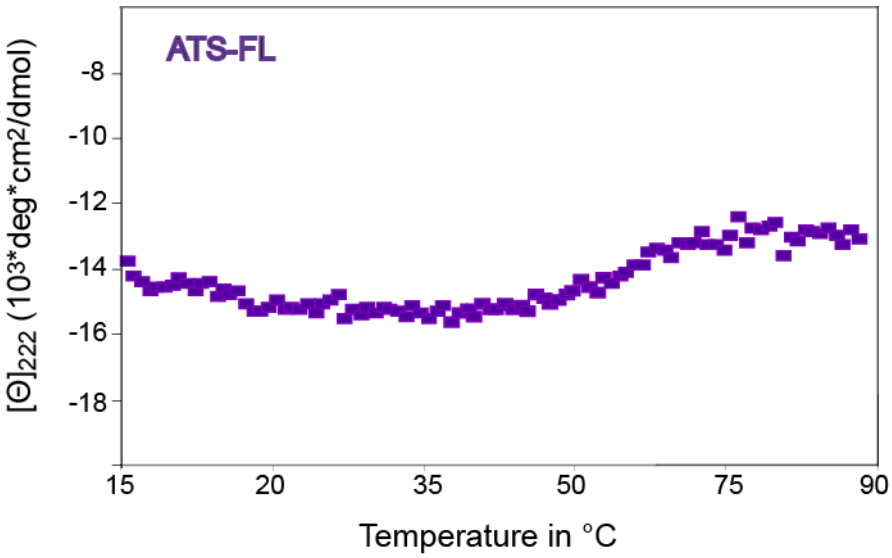


Figure 3.5: Molar ellipticity measured at 222 nm was monitored continuously in the temperature range from 15 to 90 °C. The protein concentration was 25 µM and the pH 7.4 in PBS buffer. Data were recorded at 1 °C intervals.

3.3.1 Oligomeristaion state of ATS-FL

ATS-FL has been described as dimeric under solution conditions. Previously performed gel permeation chromatography experiments suggested that recombinant ATS-FL forms a dimer (Hora et al., 2009). As mentioned before, ATS-FL shows an atypical elution volume (V_E) for a 45kDa protein in size exclusion chromatography. V_E is in a linear relationship with the logarithm of molecular mass (M_w). However, higher values than expected for a protein of a given size can indicate an extended structural state of the protein or the existence of an oligomer (Tompa, 2010). In the case of intrinsically disordered proteins (IDPs) the elution volume can shift by up to six times the expected value (Csizmok et al., 2006). To determine the absolute M_w and oligomerisation state of ATS-FL we performed analytical ultracentrifugation (AUC) sedimentation equilibrium experiments. This type of experiment determines

the M_w of the molecule of interest independent of shape. Interestingly, the results indicate that the recombinant expressed protein remains monomeric under the experimental conditions used (Figure 3.6). We observed minor aggregation for ATS-FL. The proposed flexible nature of ATS-FL is thus most likely to have given rise to the anomalous size exclusion chromatography results and the previous assignment to a dimeric state.

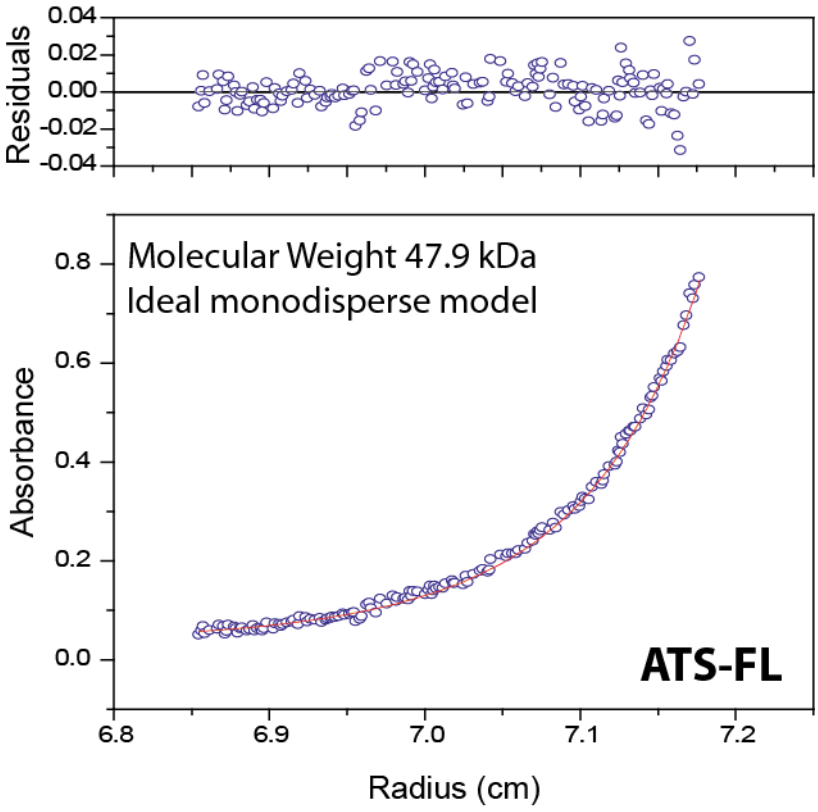


Figure 3.6: Analytical ultracentrifugation equilibrium data of ATS-FL. The UV absorbance at 280 nm of ATS-FL after a 36 hour AUC equilibrium run is plotted in blue circles. The experiment was performed at 4 °C, 15 000 rpm in PBS buffer. The red line shows the fitting of the data to an ideal monodisperse model of 47.9 kDa.

3.3.2 Small Angle X-ray Scattering (SAXS) data on ATS-FL

SAXS data from ATS-FL were recorded in PBS buffer pH 7.4 at various concentrations between 2.5 mg/ml and 15 mg/ml to explore the structural state. The protein behaved well at all concentrations. Therefore, all collected data could be used for further analysis. The Kratky plot displayed in figure 3.7 represents the structural characteristics of ATS-FL scattering data. The Kratky representation gives a first indication for order in proteins. Globular proteins produce a bell-shaped plot with a clear maximum, whereas in the case of IDPs the plot increases monotonically and approaches linearity (Kratky, 1982). A protein with residual structure gives a combination of these two types of curve (von Ossowski et al., 2005). In the case of ATS-FL the Kratky plot exhibits a plateau followed by a monotonic increase, which classifies the protein to be intrinsically disordered. However, the relative maximum on the first half of the curve indicates that ATS-FL is not fully disordered, but contains structured elements.

The scattering data were further employed to estimate the radius of gyration (R_g), using the Guinier approximation (Guinier, 1955). The programme AUTORG (Petoukhov and Svergun, 2007) was employed and R_g for ATS-FL was estimated to be 4.97 nm, a rather high value for a protein of 398 amino acids (393 residues of ATS-FL and a five residue long cloning artefact) (figure 3.7B). For comparison, theoretical values for R_g were determined based on amino acid length. Firstl y, one value for compact protein shape was calculated, approximated using the empirical relationship (Skolnick et al., 1997):

$$R_g = 0.22N^{0.38}$$

In addition, for the calculation of the random coil R_g values, the equation

$$R_g = R_0 N^v$$

was employed, where N is the number of residues, constant $R_0 = 0.1927\ nm$ and $v = 0.588$ (Kohn et al., 2004). In the case of ATS-FL the theoretical estimated R_g values were 2.14 nm and 6.51 nm for a compact and for a chain-like molecule, respectively. The R_g of proteins unfolded by low pH, methanol, urea or guanidinium chloride are typically 1.5 to 2.5 larger than the R_g of the same proteins in their native globular states. Accordingly, the experimentally observed R_g values for ATS-FL suggests that the domain is not a well-folded protein in solution, however not fully extended either.

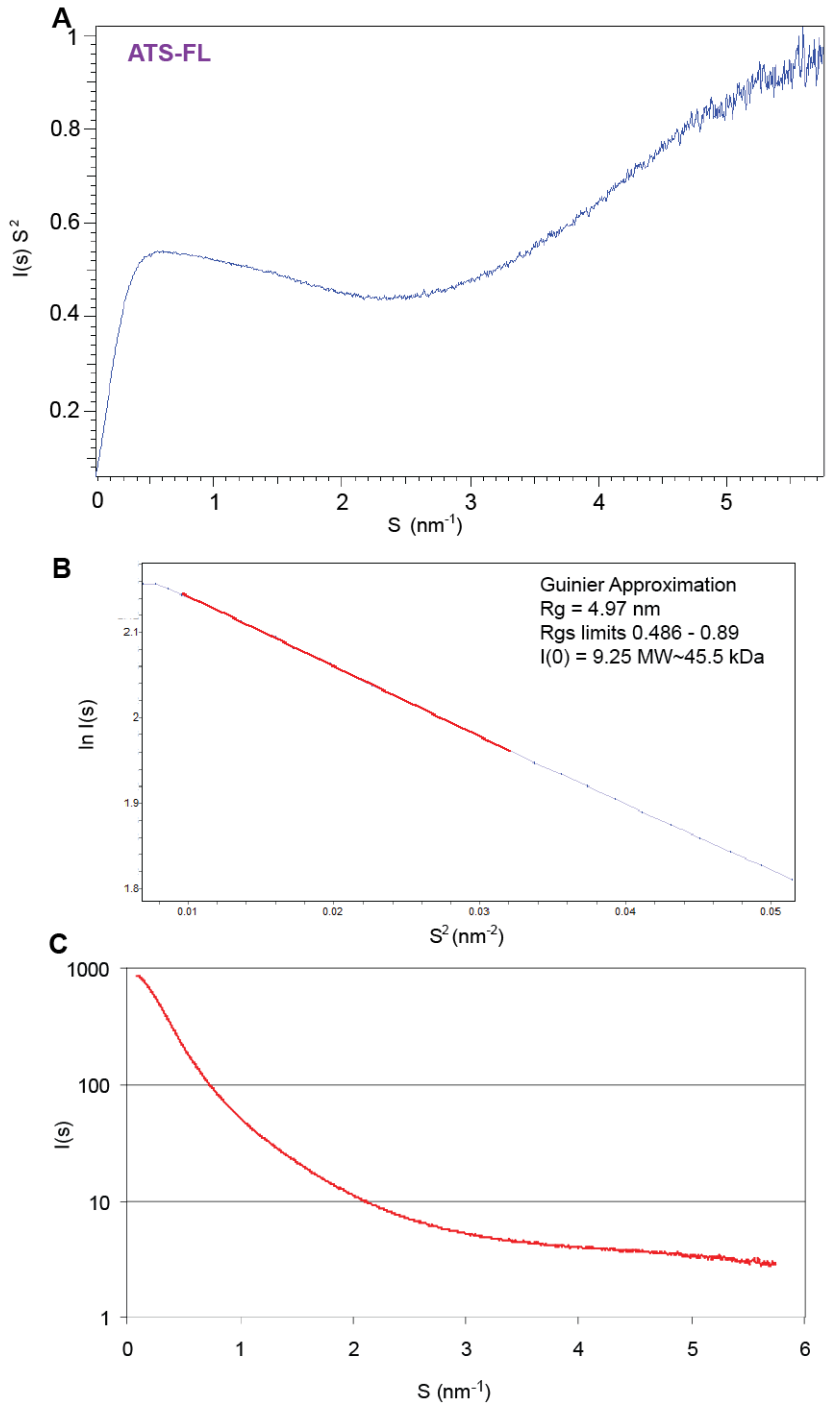


Figure 3.7: Analysis of collected ATS-FL SAXS data. A) Kratky plot of ATS-FL SAXS data collected at 20 °C. **B)** Guinier approximation of the scattering pattern of ATS-FL. The data were fitted to a region suggested by AUTORG (red line). Calculated radii of gyration (R_g) and quality of the fit are also shown. This allowed us to estimate the normalized scattering intensity at zero-angle, I(0), and the particle size using the linear relationship between I(0) and molecular weight. For comparison, I(0) of BSA (66 kDa) was 13.41 in this experiment. The calculated monomeric molecular size is 46.2 kDa for ATS-FL. **C)** X-ray scattering intensity I(s) versus angle in PBS buffer at 20 °C.

Furthermore, the pair distance distribution function P(r) gives information about protein shape and size. P(r) can be deduced by an inverse Fourier transformation of I(q) (section 2.16). In principal the P(r) comprises the same information as the scattering intensity. The representation, however, can give information about the shape of the particle at brief look. Quasi-spherical particles typically have Gaussian P(r) profiles, whereas non-spherical molecules give rise to non- Gaussian functions. The P(r) profile of ATS-FL was calculated using the program GNOM (Svergun, 1991; Svergun, 1992a)(figure 3.8). As anticipated, the curve displays a longer tail, typical for elongated macromolecules.

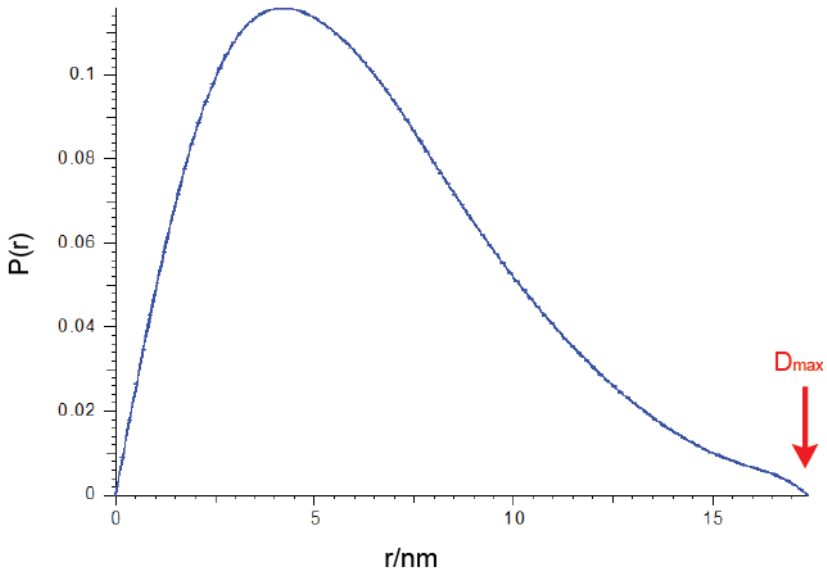


Figure 3.8: Pair distance distribution function for ATS-FL. The graph was generated by GNOM (Svergun, 1991, 1992b). The tail of the functions towards higher r-values is most likely due to the flexible nature of ATS-FL.

3.3.3 Structural analysis of ATS-FL using NMR

The ^{1}H-^{15}N HSQC spectrum of ^{15}N enriched ATS-FL reveals relatively limited chemical shift dispersion. Peaks are concentrated in the 8.0 – 8.5 ppm range in the ^{1}H dimension, which is typical for IDPs. In contrast, amide protons of globular folded proteins give rise to peaks ranging from 6.0 to 10.0 ppm. Interestingly, some of the resonances in spectrum of ATS-FL are more dispersed (Figure 3.9) indicating the existence of structured parts within ATS-FL, as also suggested by the CD experiments. Figure 3.9 compares the ^{1}H-^{15}N HSQCs of ATS-FL expressed as inclusion bodies (A) and produced under native conditions (B). The two spectra are very similar but the quality of the refolded ATS-FL spectrum is considerably better. The natively produced domain produced weaker signals due to lower expression yields and protein degradation. In addition, we carefully inspected the signals from the tryptophan side-chains in both HSQC spectra. In total, six tryptophan residues are present in ATS-FL (W117, W264, W275, W289, W310 and W391). All tryptophan side chain Hε –Nε peaks occupied the same positions in the two spectra, meaning that the chemical environment for these atoms is the same. Therefore, we concluded that the purification under denaturing conditions produced the same protein conformation and that the refolding of ATS-FL was successful.

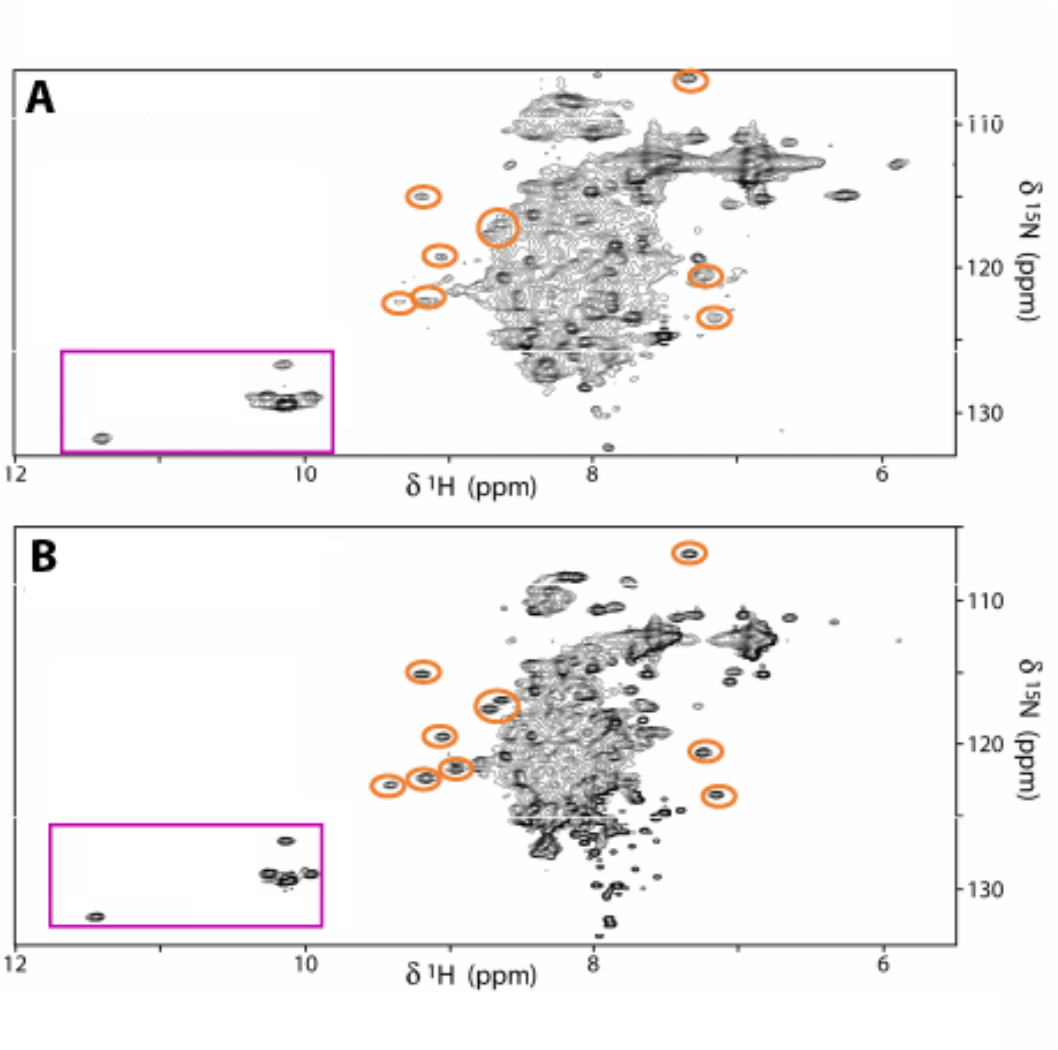


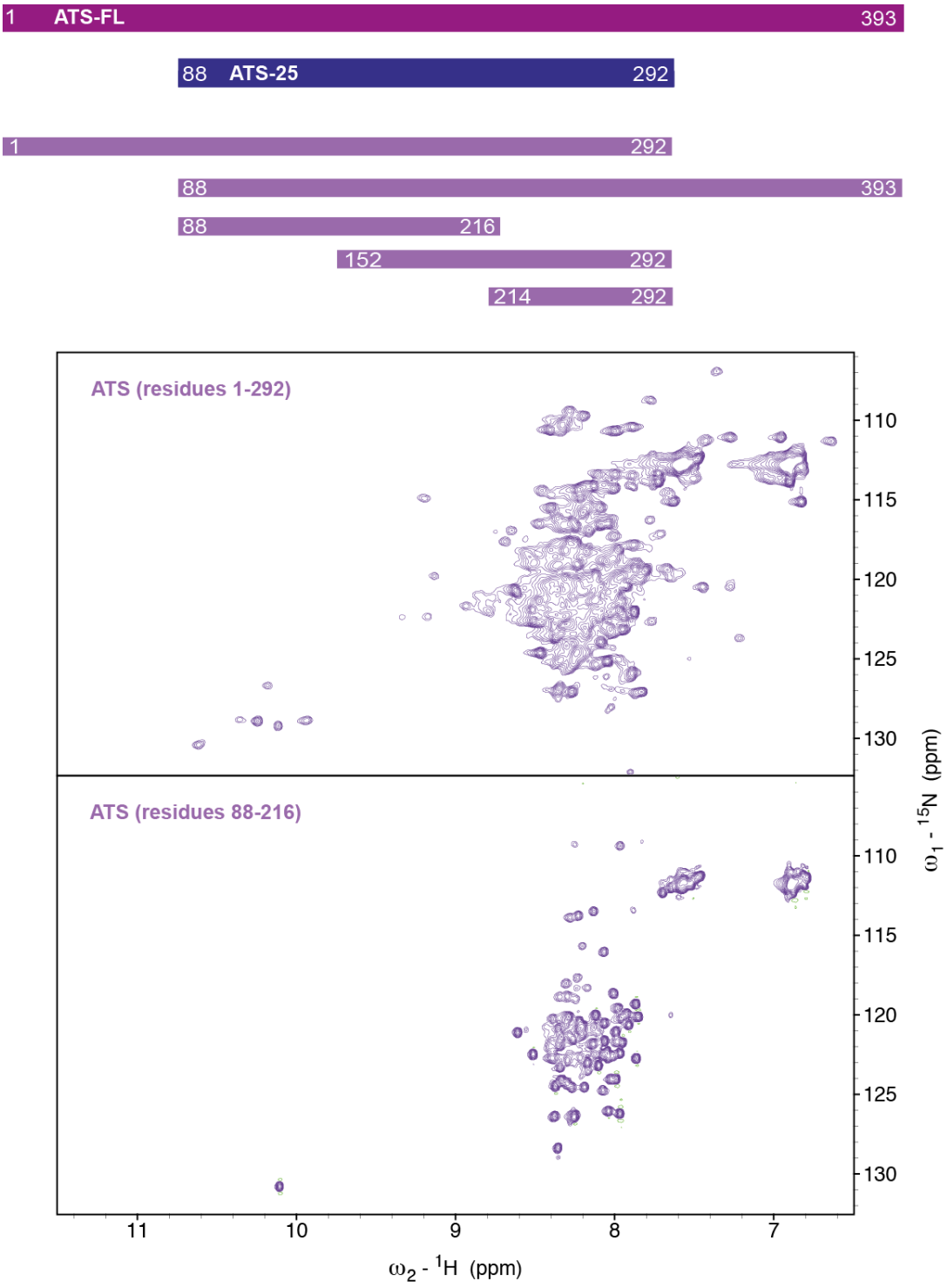
Figure 3.9: The HSQC experiment correlates the backbone amide nitrogen resonances with its attached hydrogen protons. In a ^{1}H-^{15}N HSQC spectrum, the amide bond of each residue gives rise to a peak, except for prolines. Additionally, side chain amide bonds are visible such as tryptophan, asparagine and glutamine. **A)** ^{1}H-^{15}N HSQC spectrum of ATS-FL purified under native conditions. **B)** Spectrum of ATS-FL purified from inclusion bodies under denaturing conditions. Both spectra were recorded at 25 °C in NMR buffer pH 7.0. The peaks originating from Nε-Hε tryptophan side chains are highlighted in magenta boxes. Resonances corresponding to the structured elements in the ATS domain are circled in orange in both spectra.

In summary, disorder predictions, CD, NMR derived data, Kratky plot, pair distribution function and R_g values from SAXS data demonstrate the largely flexible and extended nature of ATS-FL. However, the collected data also supports the existence of structured segments in ATS-FL. The next task was to isolate and characterise these elements.

3.4.4 Determining structured residues of ATS

Using bioinformatic sequence analysis such as secondary structure prediction and disorder prediction tools we designed several truncated variants of ATS-FL to isolate the folded segment of ATS (Figure 3.1 and 3.2). Cloning, protein expression and purification was carried out as for ATS-FL (section 3.3). Constructs containing the N-terminal (residues 1 - 292) or C-terminal part (residues 88 - 393) of ATS-FL did not improve substantially the resonance overlap in the NMR spectra (figure 3.10). A construct of approximately 25 kDa (ATS-25), comprising the middle half of ATS (residues 88 – 292), proved to contain all the structured ATS elements. The ^{1}H-^{15}N HSQC of this construct (Figure 3.11B) shows less crowding in the middle part of the spectrum, indicating reduced flexibility. However, substantial disorder remains in this shortened version of ATS-FL. Therefore, we aimed to identify the structured parts of ATS-25 by further truncating the construct. The construct 152 - 292 was designed according to the RONN disorder prediction (figure 3.1). This part of ATS is supposed to contain the ordered part of the protein according to this algorithm. A further shorting of this construct obtaining a fragment of ATS containing the residues 214 - 292 was performed according to secondary structure predictions (figure 3.2). This region has a high probability to be α-helical. However, these truncations of

ATS-25 produced completely unfolded proteins (Figure 3.10). Therefore, ATS-25 was chosen for further studies.



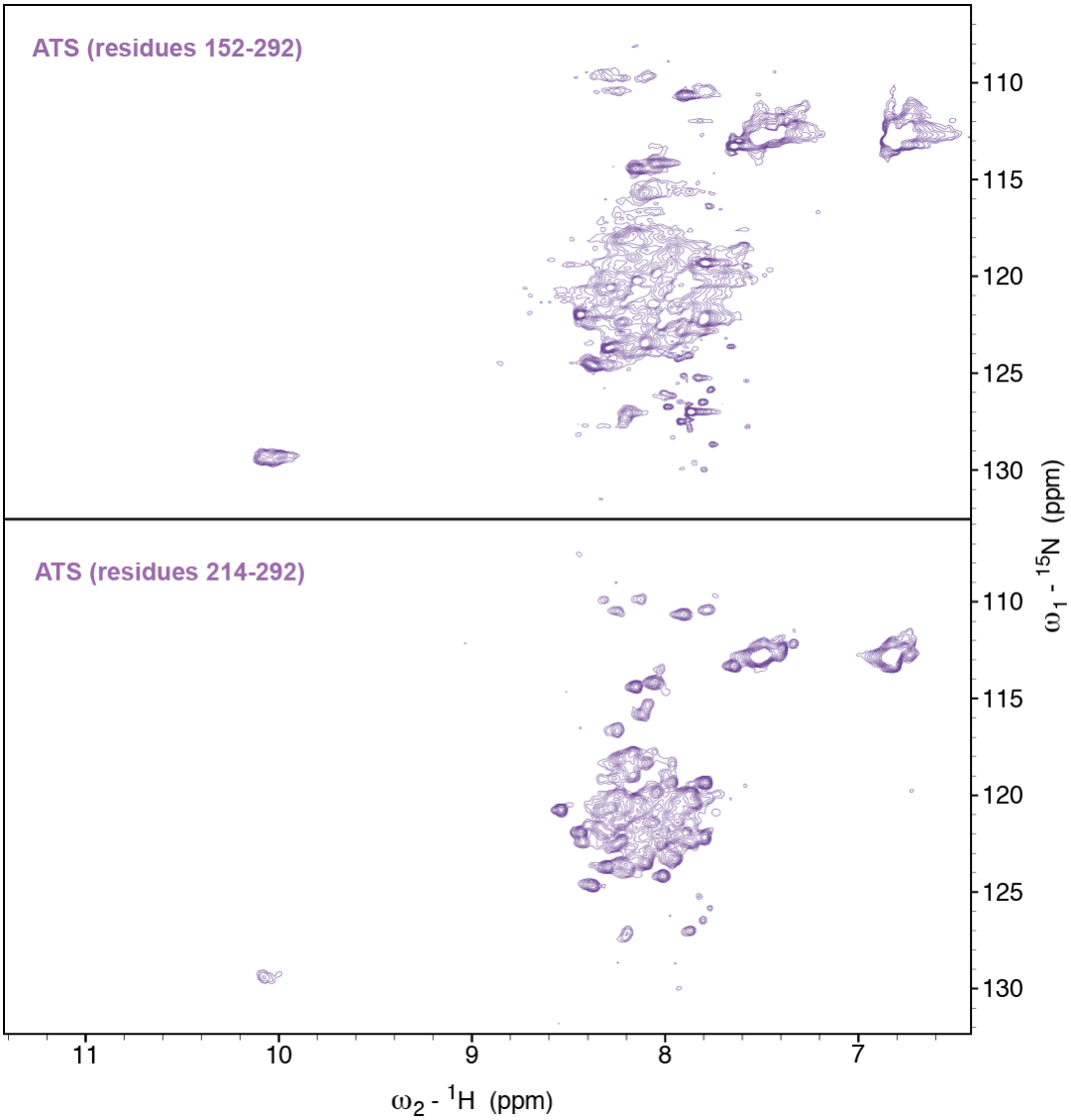


Figure 3.10: ^{1}H- ^{15}N HSQC spectra of various ATS constructs. Schematic representation of the different construct lengths is shown at the top. NMR spectra of 100 µM ^{15}N labelled protein samples at 25 °C recorded in NMR buffer at pH 7.0.

3.4.5 Biophysical characterisation of ATS-25

As mentioned before, ATS-25 contains all the structured parts of ATS-FL as indicated by the HSQC spectra. The highlighted resonance peaks (figure 3.11B) appear in the NMR spectra of both the full-length and ATS-25. Therefore, we concluded that the structured segments must be located in this region of ATS. Analysis of SAXS data of ATS-25 confirmed the existence of the folded residues, as the corresponding Kratky plot displays an increased maximum compared to that of ATS-FL, indicative of a folded conformation. These data suggests that ATS-25 is more compact than ATS-FL. The decrease in resonance overlap in the ATS-25 NMR spectra also suggests decreased flexibility in this construct.

It had been suggested that an ATS dimerisation unit lies within the central part of the protein (Hora et al., 2009); ATS-25 contains this part. However, AUC analysis did not confirm this suggestion (figure 3.11A). ATS-25 remained monomeric under solution conditions (as did ATS-FL, section 3.3.1).

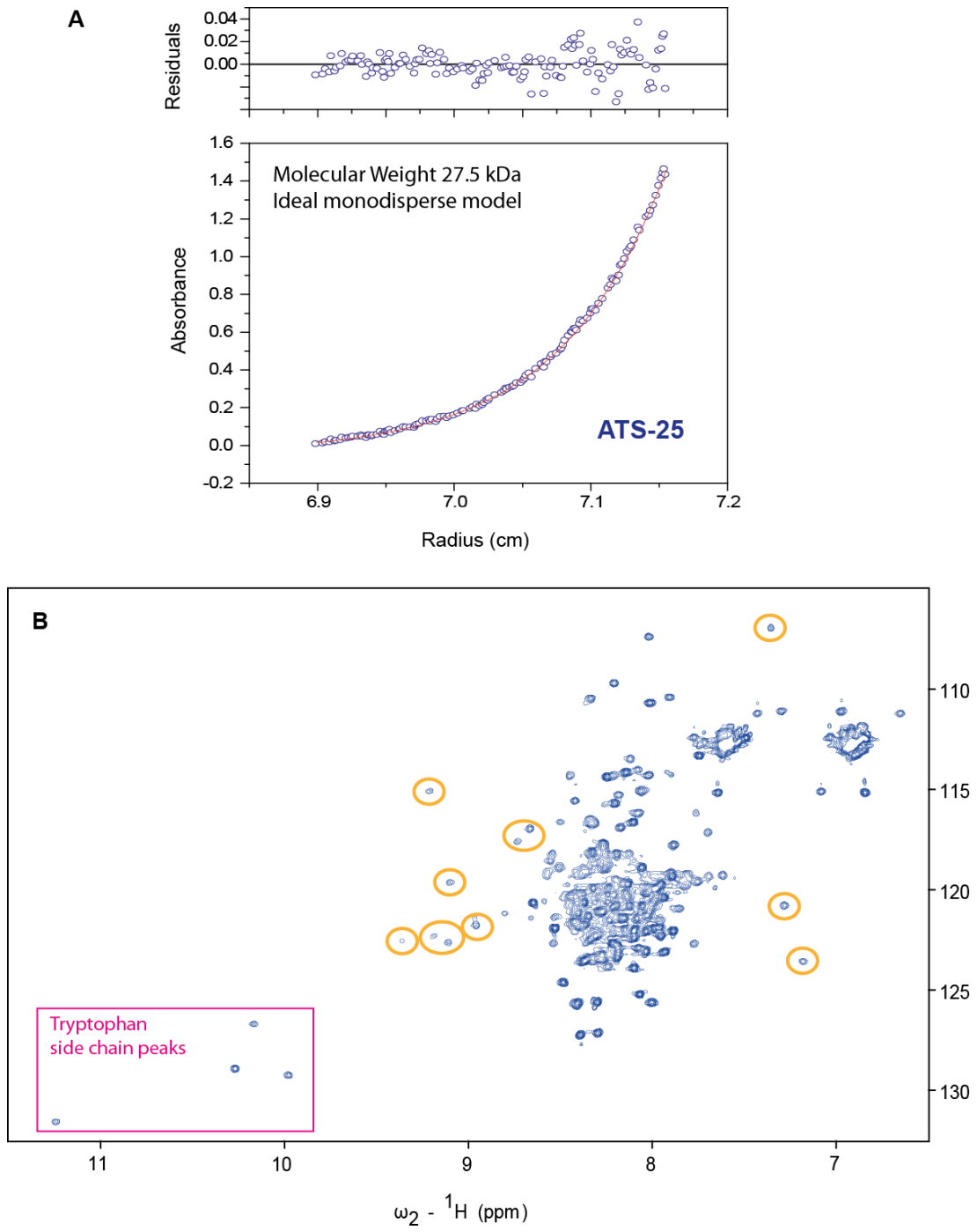


Figure 3.11: Biophysical characterization of ATS-25. A) Analytical ultracentrifugation equilibrium data of ATS-25. The UV absorbance at 280 nm of ATS-FL after a 36 hour AUC equilibrium run is plotted in blue circles. The experiment was performed at 4 °C, 22 500 rpm in PBS buffer. The red line shows the fitting of the data to an ideal monodisperse model of 27.5 kDa. **B)** ^{1}H- ^{15}N HSQC spectrum of ATS-25 recorded at 25 °C pH 7.0. The overlap of peaks in the centre is less compared to ATS-FL and all resonances of structured parts are still present (orange circled peaks).

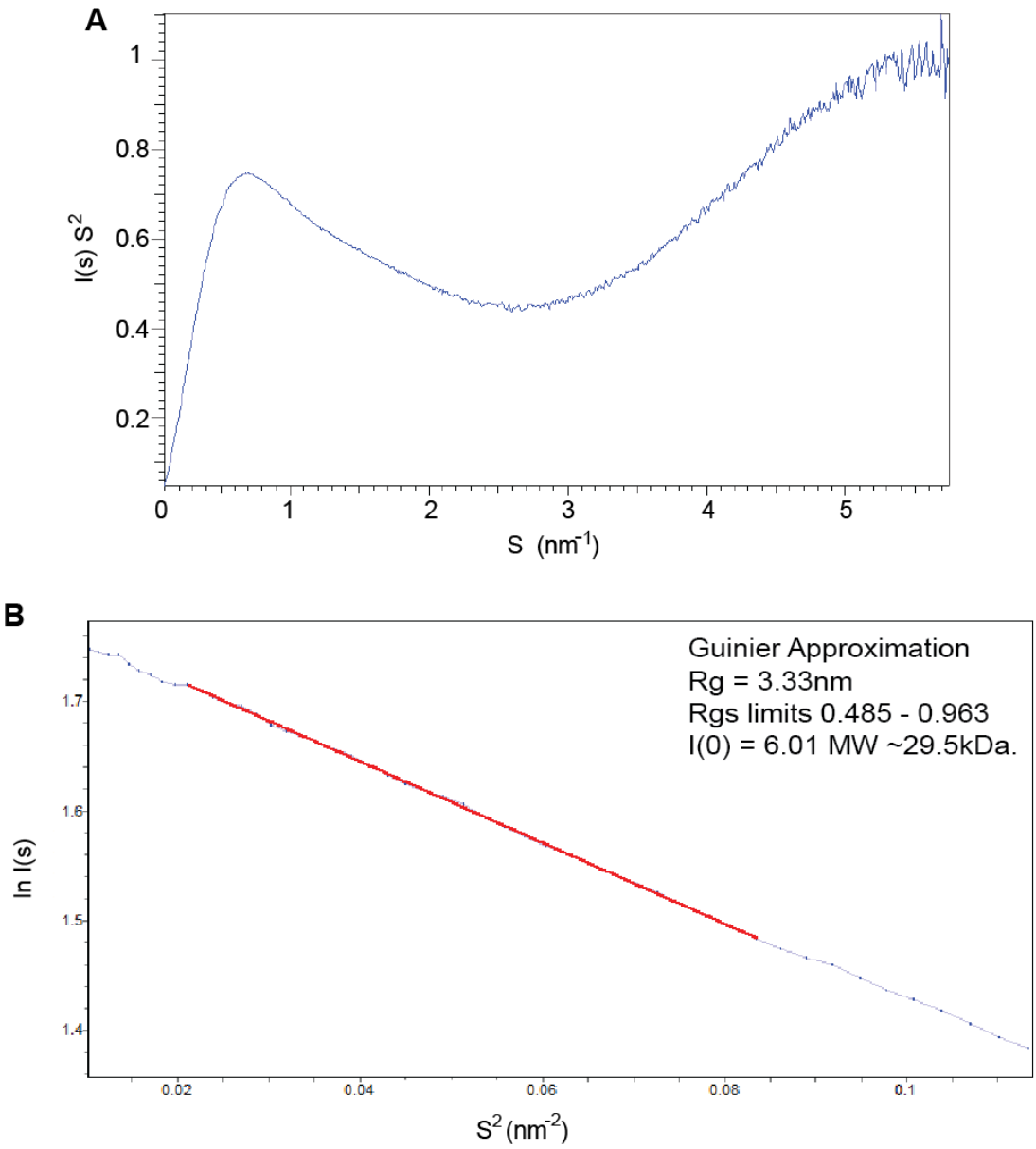


Figure 3.12: SAXS analysis of ATS-25. A) Kratky representation of SAXS data from ATS-25 construct collected at 20 °C. The curve increases monotonically and shows a peak, indicating residual structure. The curvature of the plot is increased compared to that of ATS-FL (figure 3.6A). **B)** Guinier plot of the scattering data from ATS-25. Data fitting was performed with AUTORG.

3.4.6 Triple resonance backbone assignment of ATS

Spectral assignment of NMR resonances can be achieved with a $^{15}N/^{13}C$ labelled sample via a 'sequential walk' through backbone and side-chain resonances using one-bond correlations. In the case of folded proteins successful backbone assignment can be achieved using three-dimensional carbon-resonance experiments that are acquired in pairs: CBCA(CO)NH / CBCANH (Cavanagh, 2007). Carbon frequencies for $^{13}C\alpha$ and $^{13}C\beta$ are unique and dispersed in folded proteins, which allows for fast and accurate assignment. However, the task is more challenging for unfolded proteins as only small differences in carbon chemical shifts between amino acids are observed, which makes the assignment process difficult. For this reason we analysed multiple datasets simultaneously, encompassing both carbon and proton frequencies, to accomplish sequential assignment of ATS-FL (figure 3.15). First, we prepared ^{15}N and ^{13}C double labelled NMR samples of ATS-25, N-terminal ATS (residues 2-112) and C-terminal ATS (293-393) and subsequently recorded a series of three dimensional heteronuclear NMR experiments: CBCANH (Wittekind and Mueller, 1993), CBCA(CO)NH (Grzesiek and Bax, 1992), HBHANH, HBHA(CO)NH (Grzesiek et al., 1993), HCACO, HCA(CO)N and HNCO (Kay et al., 1990). All three constructs were separately analysed and the assignments were later combined. The HSQC spectra for the N-terminal and C-terminal ATS constructs are shown in figure 3.13 and 3.14, respectively. Both spectra show limited chemical shift dispersion in the proton dimension, typical for intrinsically disordered proteins.

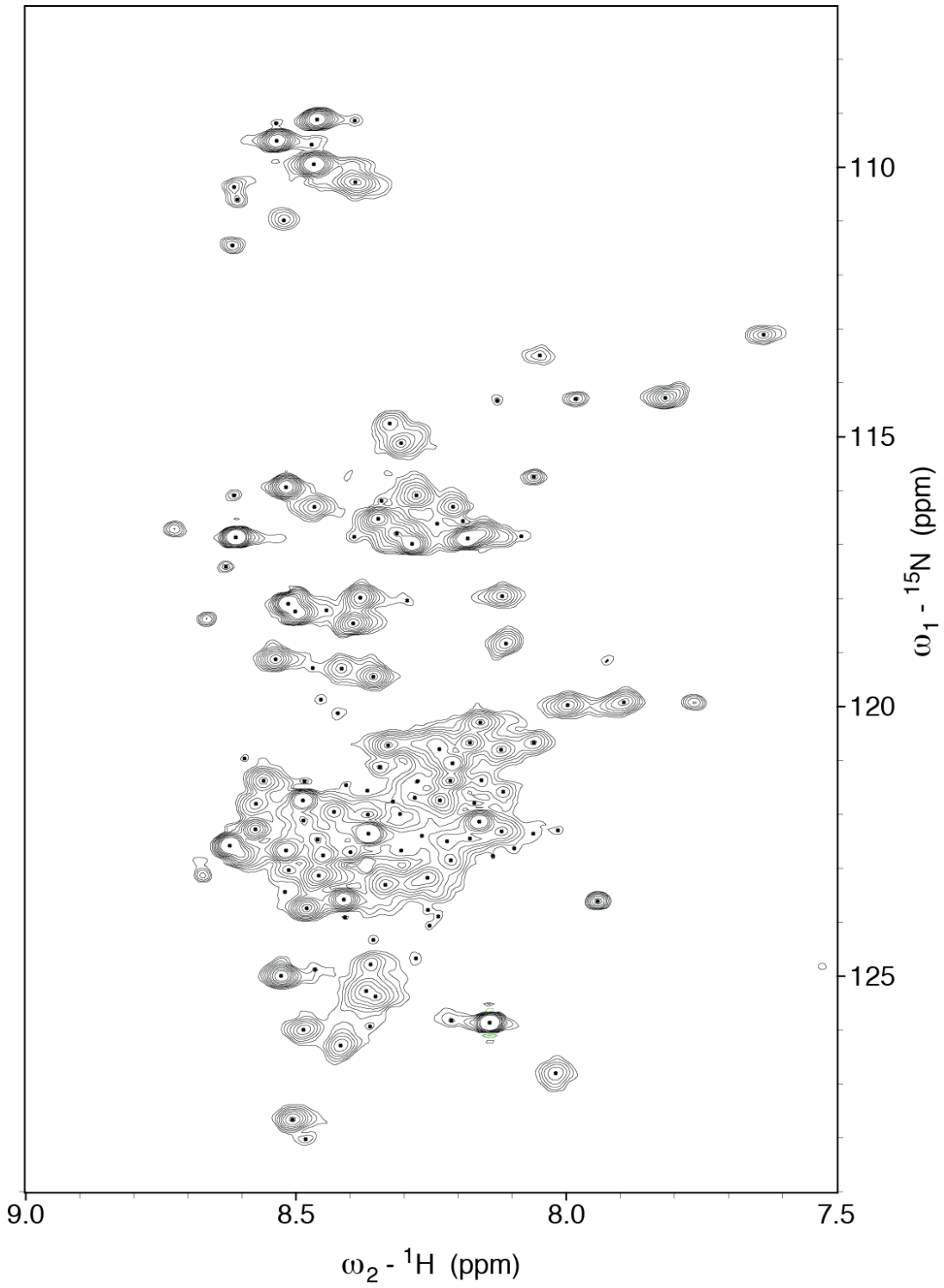


Figure 3.13: ^{1}H- ^{15}N HSQC spectra of N- terminal ATS construct. The construct contains the residues K2 - P108. The experiment was recorded at 10 °C with a 1mM protein sample in NMR buffer.

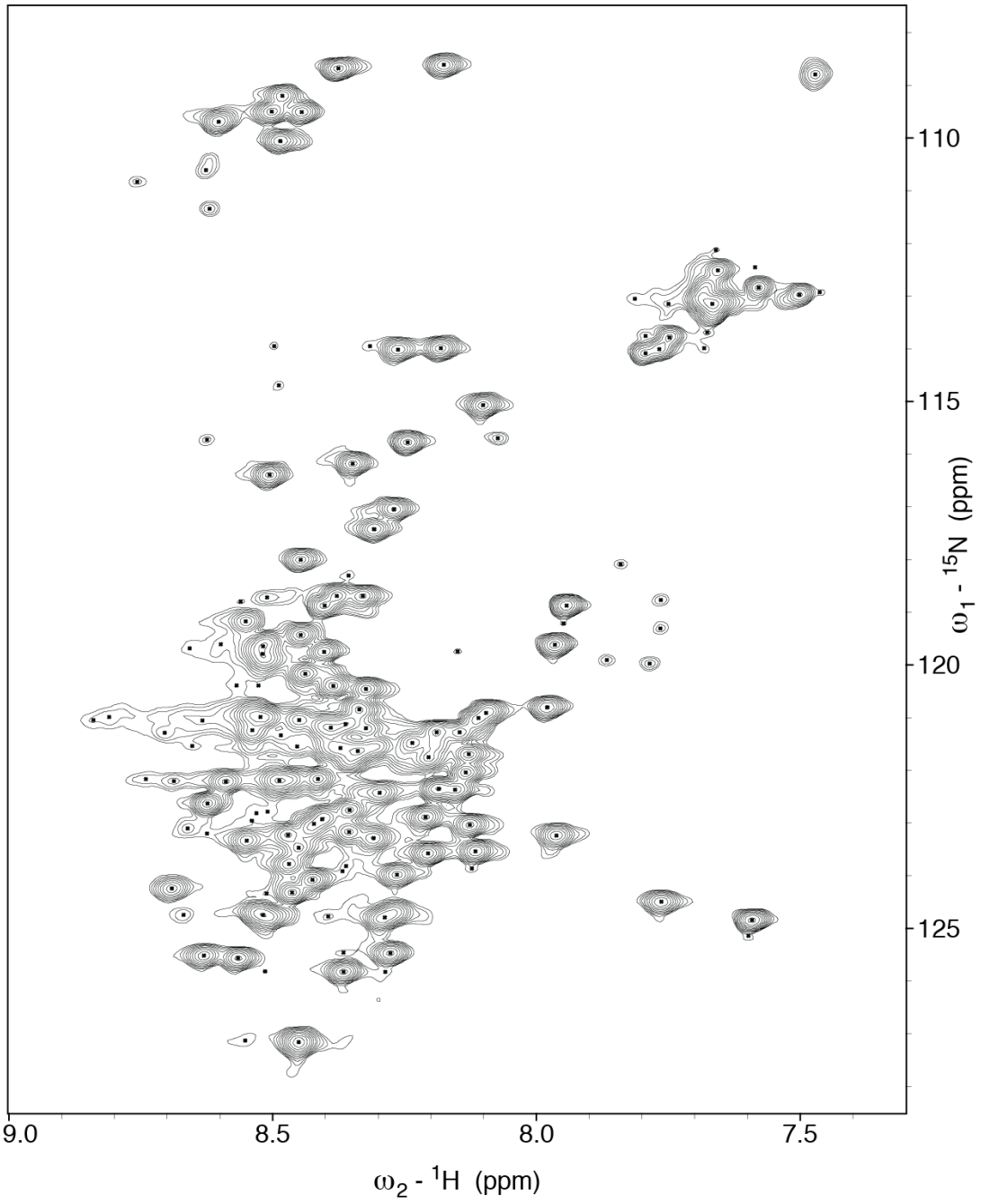
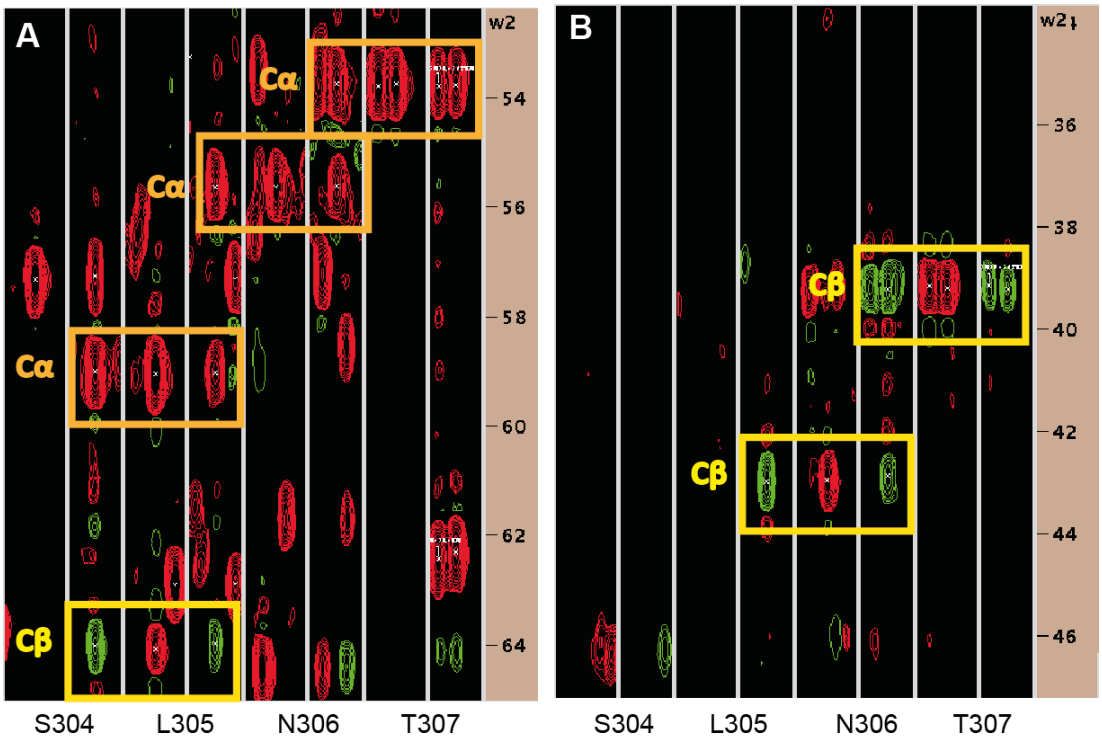
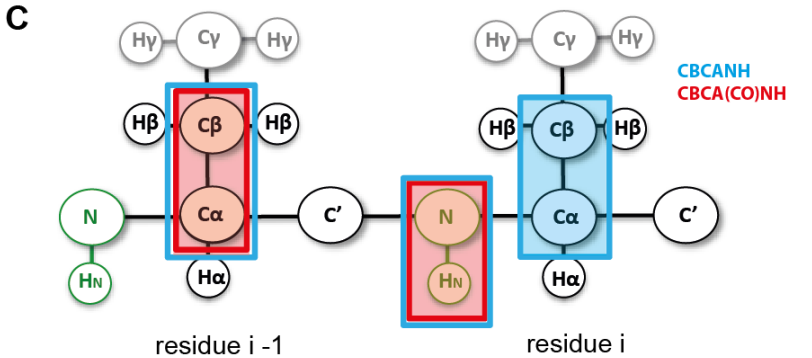
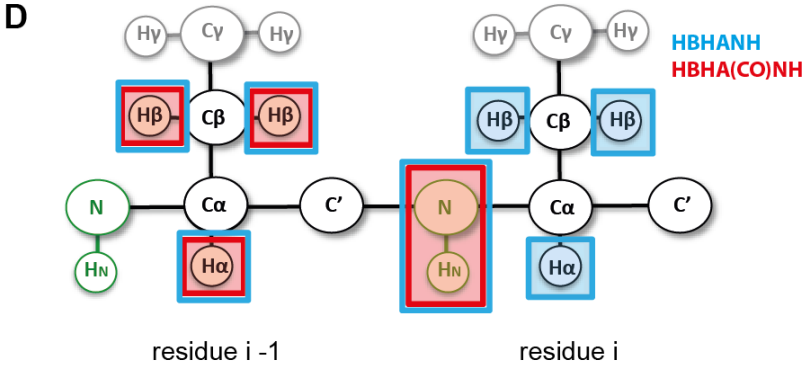


Figure 3.14: ^{1}H- ^{15}N HSQC spectra of C- terminal ATS construct; containing the residues G295 - I393 of ATS-FL. Experimental conditions were the same as for the N- terminal construct.

Assignment stretches were typically started at amino acids with characteristic Cα and Cβ chemical shifts, such as alanine, serine, threonine and glycine. The Cβ chemical shifts of these amino acids stand out and glycine does not have a Cβ resonance. Standard chemical shift ranges for different amino acids were taken from Biological Magnetic Resonance Data Bank as guidance. First, we employed a CBCANH spectrum, which correlates each NH group with the Cα and Cβ resonance of its own residue and the previous residue. CBCA(CO)NH spectra only show the Cα and Cβ chemical shift of the preceding residue. Figure 3.15 (A and B) shows how this connects the backbone of residues S304 to T307. However, in the case of ATS we also needed HBHANH and HBHA(CO)NH spectra, from which we could extract additional Hα and Hβ chemical shift values for each residue. Using the obtained Cα and Hα values we could determine the CO resonance of the residue i and the N resonance of the following residue i+1 using a HCACO and HCA(CO)N spectra, respectively. Subsequently, a HNCO experiment was used to assign the corresponding amide proton frequency of this residue (i+1). An overview of this assignment strategy is displayed in figure 3.15 (C-G). Using this approach we were able to assign over 75 % of the residues of ATS-FL. The determined chemical shifts can be found in Appendix IV and are deposited in the BioMagResBank under accession number 17999.


A
w2
54
56
58
60
62
64
Cα
Cα
Cα
Cβ
S304
L305
N306
T307
B
w2
36
38
40
42
44
46
Cβ
Cβ
S304
L305
N306
T307


C
Hγ
Cγ
Hγ
Hβ
Cβ
Hβ
N
Cα
C'
HN
Hα
CBCANH
CBCA(CO)NH
residue i -1
residue i


D
Hγ
Cγ
Hγ
Hβ
Cβ
Hβ
N
Cα
C'
HN
Hα
HBHANH
HBHA(CO)NH
residue i -1
residue i

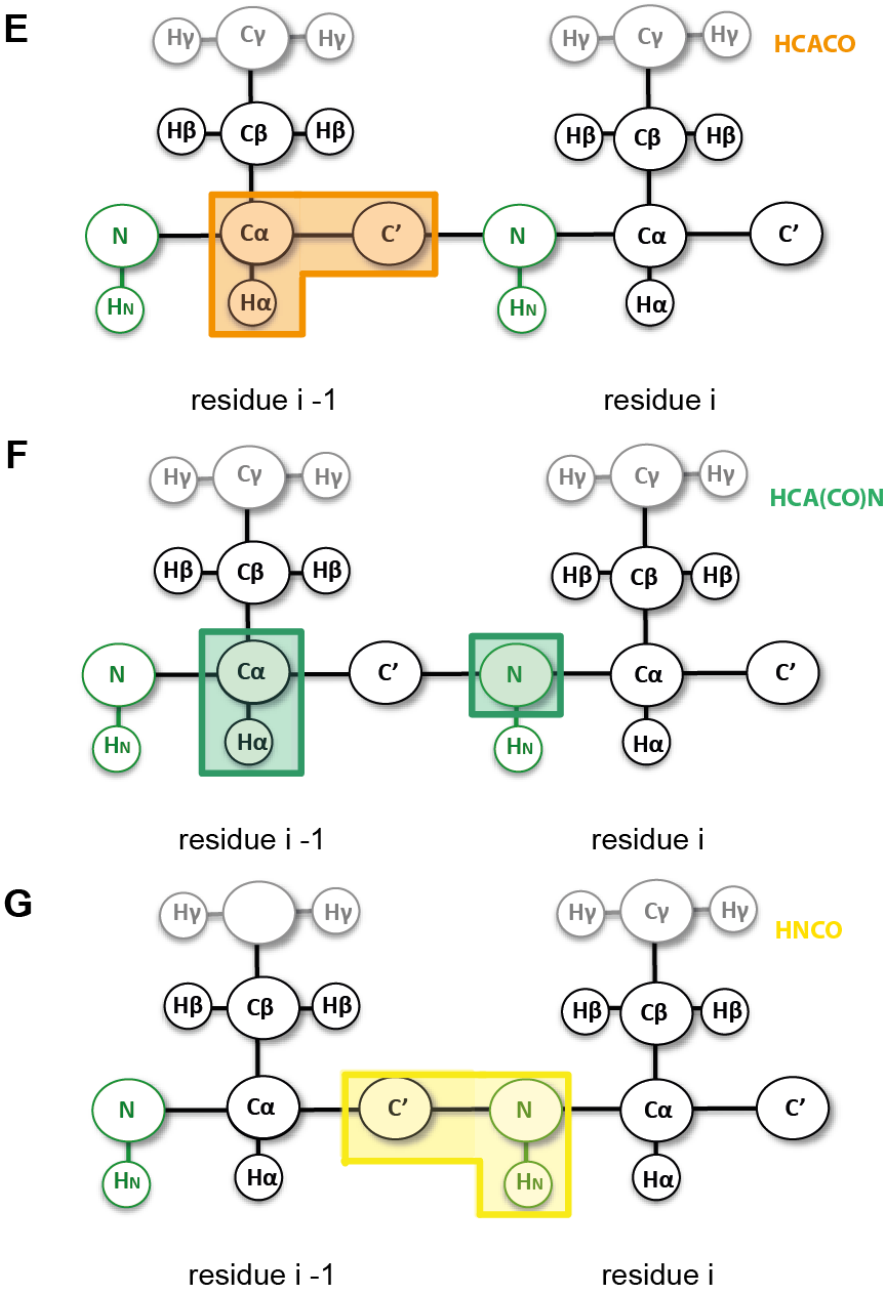


Figure 3.15: NMR backbone assignment strategy for ATS. Due to the lack of chemical shift dispersion in the unstructured segments we used carbon and proton resonances to follow the backbone connectivity. **A** and **B)** show a strip plot (Sparky) for the residues S304 – T307. For each residue a strip of the CBCA(CO)NH and CBCANH spectra are displayed, respectively**. C-G)** Illustration of the stepwise assignment strategy employed, using a combination of the listed triple resonance experiments.

3.4.7 Sequence specific structural information

Assignment of the backbone resonances of ATS enabled us to compare the observed chemical shift values to expected random coil values. The deviation from characteristic random coil values for $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$ and ^{13}CO can be used to reveal local structural propensities. In an α-helix, $^{13}C\alpha$ and ^{13}CO are shifted downfield, whereas $^{1}H\alpha$ and $^{13}C\beta$ are moved upfield (Dyson and Wright, 2002, 2004). The plots of deviations of the $^{13}C\alpha$, ^{13}CO and $^{1}H\alpha$ shifts from random coil values at pH 7.0 for ATS-FL are shown in figure 3.16. Positive deviations for $^{13}C\alpha$ and ^{13}CO relative to random coil values are observed for residues 112-136 and for a second segment stretching from the residues 254 to 292. The values for $^{1}H\alpha$ chemical shift deviations for these two sequences also show significant shifts, however now upfield. Therefore, we concluded that two α-helical structured segments are present in ATS-FL and located them to the residues 112-136 and 254-292. However, the chemical shift data for the rest of the sequence indicated that most of the domain is lacking persistent secondary structure. We can conclude that ATS-FL is composed of three flexible regions: N-terminal (residues 1-111), C-terminal (residues 293-393) and middle part (residues 137-253). These are separated by two distinctly structured elements. These findings were consistent with bioinformatic data provided by Dr. Robert Konrat (University of Vienna, Austria) (Konrat, 2009). His meta-structure secondary structure analysis of ATS-FL showed positive values for the same regions, supporting the α-helical propensity (figure 3.17).

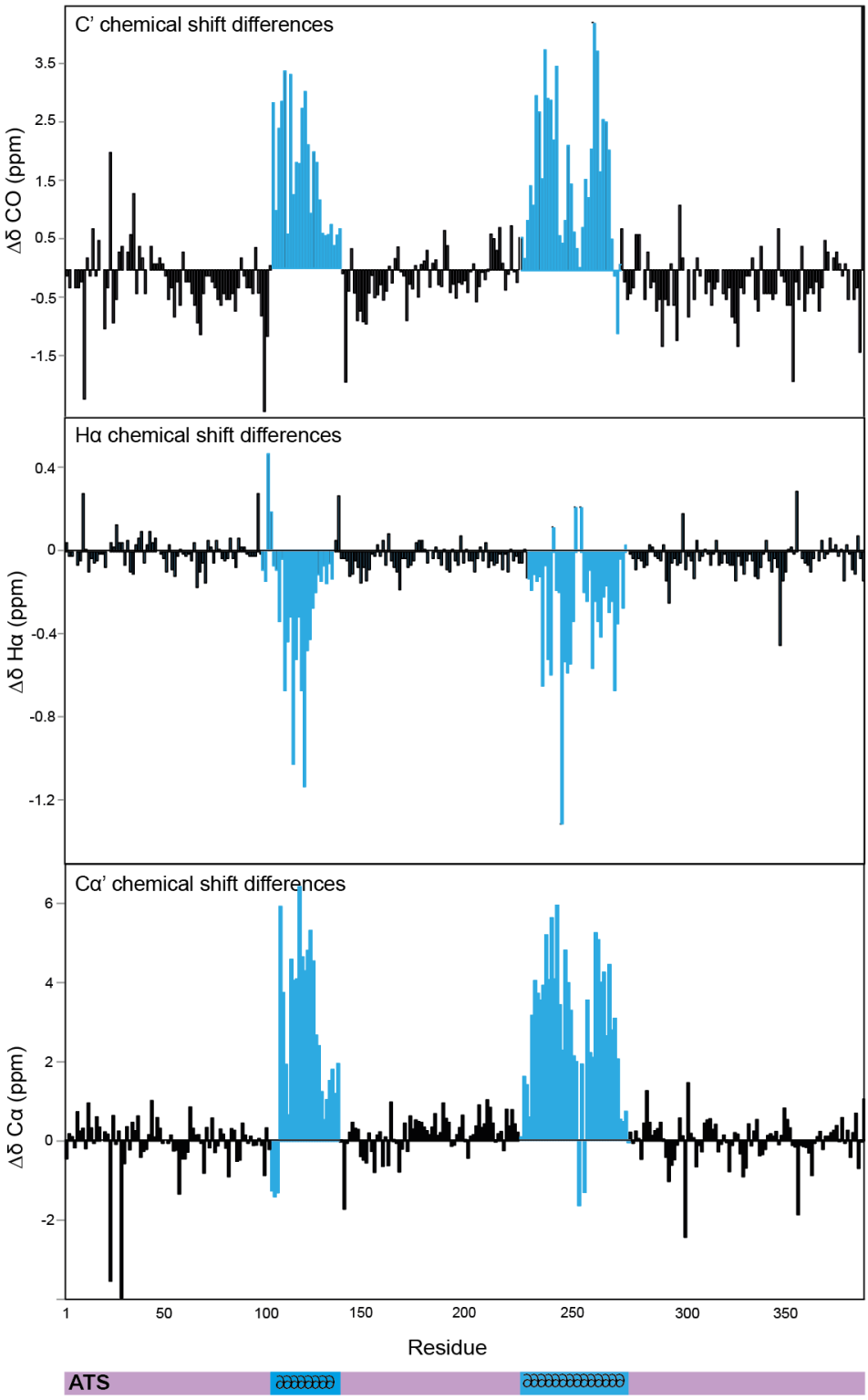


Figure 3.16: Chemical shift deviations (Δδ) of ^{13}CO, $^{13}C\alpha$ and $^{1}H\alpha$ from predicted random coil values for ATS-FL. All shifts were measured from spectra acquired at 25 °C in NMR buffer pH 7.0. Residues in blue highlight the structured elements. Stretches of positive deviations for ^{13}CO and $^{13}C\alpha$ and negative deviations for $^{1}H\alpha$ correspond to α-helical segments.

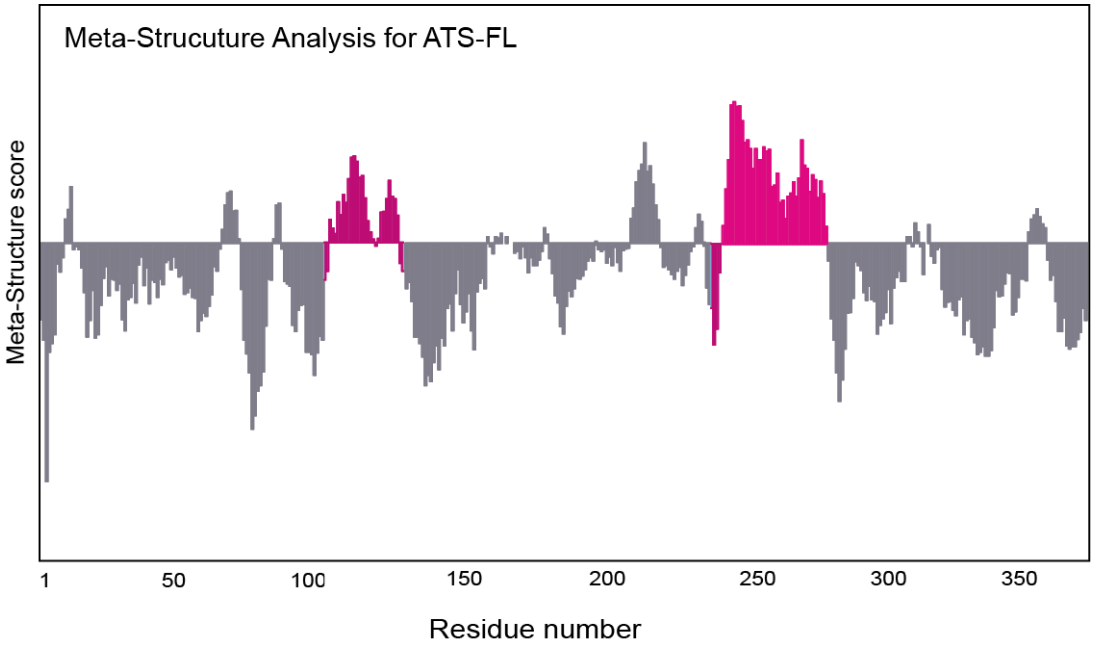


Figure 3.17: Meta-structure analysis for ATS-FL. Positive values correspond to α-helical, negative values to either β-strands or extended elements. The residues determined to have α-helical propensity by NMR chemical shifts are highlighted in magenta (residues 112-136; 254-292).

3.4.8 Characterisation of the structured elements of ATS-FL

NMR backbone assignment and secondary structure determination based on chemical shift analysis revealed the presence and location of α-helical structures within the ATS domain. Using this information we designed four different new ATS constructs based on the PF08_0141 of *P. falciparum* 3D7 (Table 3.1). The corresponding DNA fragments were cloned into the pGEX-6P-2 expression vector. Protein expression was performed in M9 minimal media to obtain ^{15}N-enriched protein samples for NMR studies. All four constructs were expressed and purified in the same manner and ^{1}H-^{15}N HSQC spectra were acquired. The NMR data nicely illustrates the decrease in flexibility for the stepwise shortening of the loop for the respective ATS constructs, as the chemical shift overlapping decreased. Based on these results we decided to concentrate further studies on the shortest construct, as no resonance peak overlapping was observed. The corresponding ^{1}H- ^{15}N HSQC for this construct, which was named ATS-core is shown in figure 3.18.

Residues	Amino acid (aa) length	Predicted mass (kDa)	pI	Length of middle loop (aa)
109 - 292	189	22	4.93	117
109 - 136 217 - 292	109	13	5.35	37
109 - 136 231 - 292	95	11.3	5.88	23
109 - 136 245 - 292	81	9.7	5.24	9

Table 3.1: Overview of designed ATS constructs for characterization of structured parts. Residue numbering corresponds to ATS-FL PF08_0141 (K1 – I393). Molecular weight and theoretical isoelectric point (pI) were calculated using ProtParam (Gasteiger E., 2005).

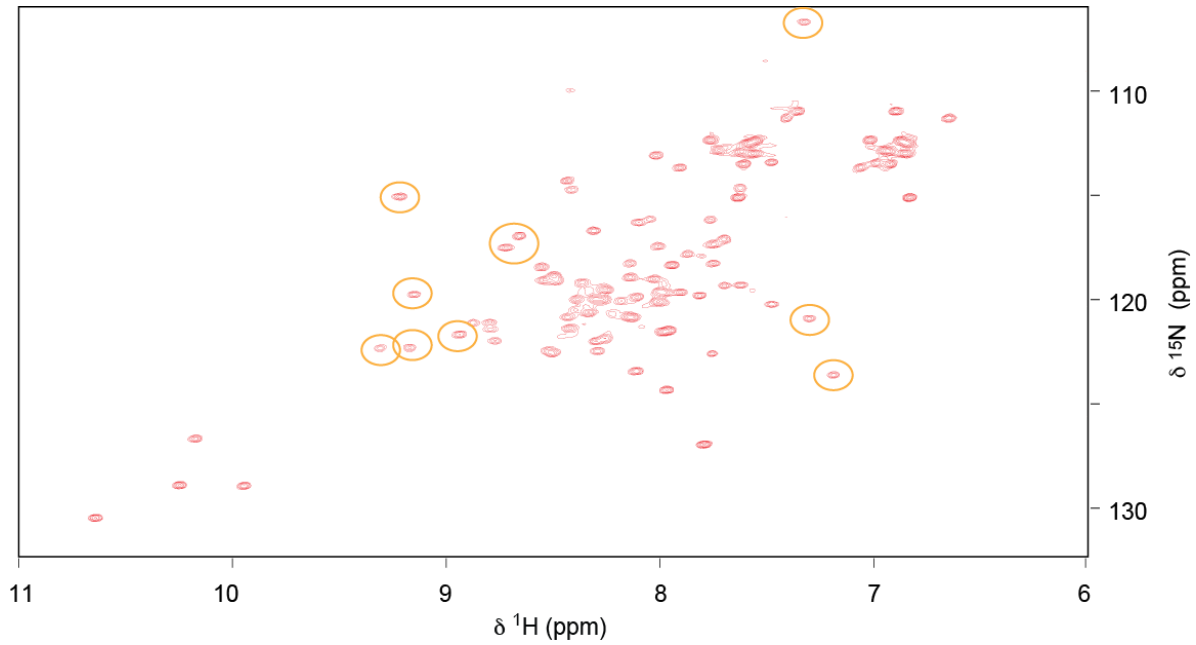


Figure 3.18: NMR ^{1}H- ^{15}N HSQC spectrum of ATS-core at 25 °C. Distinct resonances, which were observed in ATS-FL and ATS-25 are highlighted by orange circles. The chemical shift dispersion pattern indicates that ATS-core is well folded in solution.

3.4.9 Biophysical characterisation of ATS-core

The NMR spectra of ATS-core shows that this construct lacks most flexible elements, as it displays wide chemical shift dispersion (figure 3.18). In addition, we were able to show that the resonances of ATS-core are consistent with those in ATS-FL, which are highlighted in all relevant ^{1}H- ^{15}N HSQC spectra. Therefore, we can conclude that the ATS-core structure is present in the ATS-FL and we are not dealing with an artificially created conformation.

Further studies by SAXS of ATS-core confirmed the fully folded conformation of this construct. The corresponding Kratky plot, shown in figure 3.19A, displays a clear maximum, indicative of a folded protein state. The R_g value of 1.66 nm for ATS-core determined using GNOM, corresponds to the calculated R_g for a globular protein of this size (1.17 nm). As expected, CD measurements illustrate that the ATS-core is

composed of α-helices. The troughs at 220 nm and 208 nm are typical for a highly α-helical structure (figure 3.20A). In addition, thermal unfolding monitored at λ = 222 nm was conducted. The cooperative reaction observed for ATS-core (figure 3.20B) indicated that the protein exists as a compact, well-folded structure.

Findings presented in this chapter propose that ATS-FL exists as a highly flexible, disordered protein in solution. The domain is composed of three large intrinsically unstructured sections at the protein N- terminus, C-terminus and centre and an α-helical core structure.

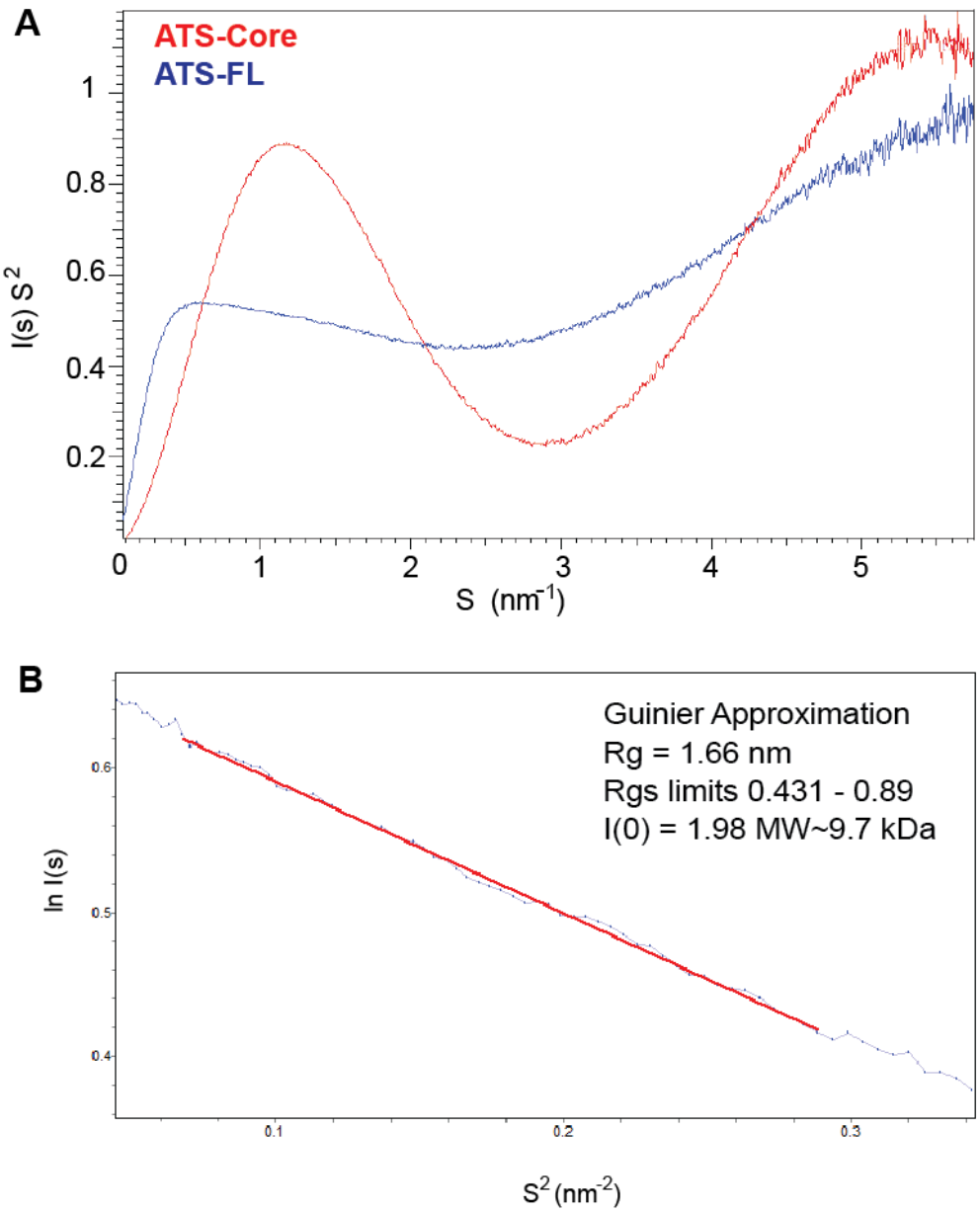


Figure 3.19: SAXS data of ATS-core. A) Kratky plot of ATS-core SAXS data collected at 20 °C. **B)** Guinier approximation of the scattering pattern of ATS-core. The data were fitted to a region suggested by AUTORG (red line). Calculated radius of gyration (R_g) and quality of the fit are shown.

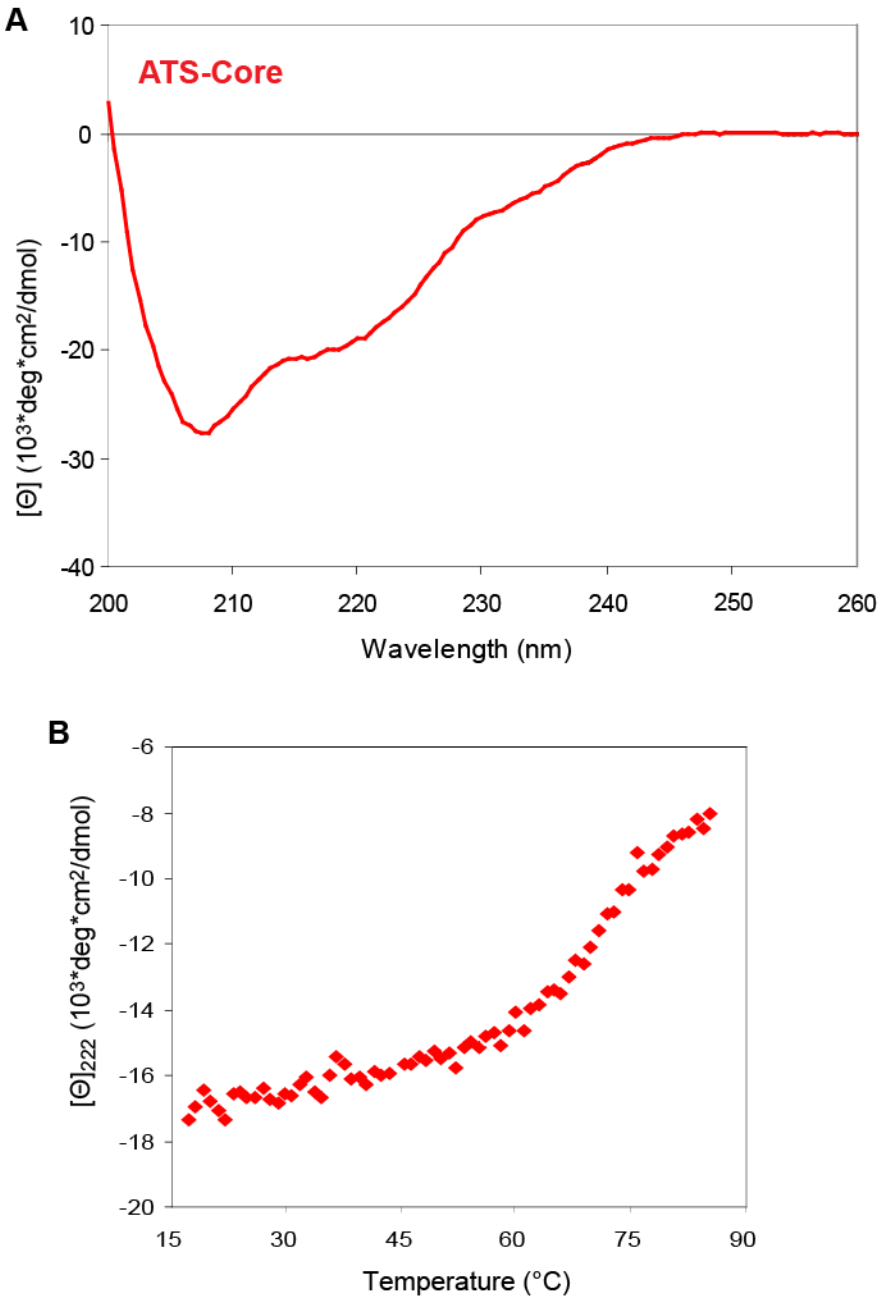


Figure 3.20: CD spectra and thermal unfolding of ATS-core. A) The spectrum was recorded in PBS buffer (containing 1 mM DTT) pH 7.4 at 10°C with a protein concentration of 25 µM. **B)** Molar ellipticity measured at 222 nm is plotted against temperature range from 15 to 90 °C. ATS-core exhibits stability and cooperative transitions during the temperature increase.

3.5 References

Bertoncini, C.W., Jung, Y.S., Fernandez, C.O., Hoyer, W., Griesinger, C., Jovin, T.M., and Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. Proceedings of the National Academy of Sciences of the United States of America *102*, 1430-1435.

Cavanagh, J., Fairbrother, W., Palmer A. G., Skelton N. J (2007). Protein NMR Spectroscopy - Principles and Practise (Academic Press).

Cheng, H., Sen, T.Z., Kloczkowski, A., Margaritis, D., and Jernigan, R.L. (2005). Prediction of protein secondary structure by mining structural fragment database. Polymer *46*, 4314-4321.

Csizmok, V., Szollosi, E., Friedrich, P., and Tompa, P. (2006). A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins. Molecular & cellular proteomics : MCP *5*, 265-273.

Cuff, J.A., and Barton, G.J. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. Proteins *34*, 508-519.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433-3434.

Dyson, H.J., and Wright, P.E. (2002). Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Advances in protein chemistry *62*, 311-340.

Dyson, H.J., and Wright, P.E. (2004). Unfolded proteins and protein folding studied by NMR. Chemical reviews *104*, 3607-3622.

Endo, Y., and Sawasaki, T. (2007). [Wheat germ cell-free protein synthesis]. Seikagaku The Journal of Japanese Biochemical Society *79*, 229-238.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S.*, et al.* (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498-511.

Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, Humana Press (2005) (Humana Press), pp. 571 - 607.

Grzesiek, S., and Bax, A. (1992). Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr. J Am Chem Soc *114*, 6291-6293.

Grzesiek, S., Vuister, G.W., and Bax, A. (1993). A simple and sensitive experiment for measurement of JCC couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. Journal of biomolecular NMR *3*, 487-493.

Guinier, A., and G. Fournet (1955). Small Angle Scattering of X-Rays (New York: Wiley).

Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science *295*, 1852-1858.

Hora, R., Bridges, D.J., Craig, A., and Sharma, A. (2009). Erythrocytic casein kinase II regulates cytoadherence of Plasmodium falciparum-infected red blood cells. The Journal of biological chemistry *284*, 6260-6269.

Ishida, T., and Kinoshita, K. (2008). Prediction of disordered regions in proteins based on the meta approach. Bioinformatics *24*, 1344-1348.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology *292*, 195-202.

Kay, L.E., Ikura, M., and Bax, A. (1990). Proton Proton Correlation Via Carbon Carbon Couplings - a 3-Dimensional Nmr Approach for the Assignment of Aliphatic Resonances in Proteins Labeled with C-13. J Am Chem Soc *112*, 888-889.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols *4*, 363-371.

Kohn, J.E., Millett, I.S., Jacob, J., Zagrovic, B., Dillon, T.M., Cingel, N., Dothager, R.S., Seifert, S., Thiyagarajan, P., Sosnick, T.R., *et al*. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. Proceedings of the National Academy of Sciences of the United States of America *101*, 12491-12496.

Konrat, R. (2009). The protein meta-structure: a novel concept for chemical and molecular biology. Cellular and molecular life sciences : CMLS *66*, 3625-3639.

Kratky, G. (1982). Small Angle X-ray Scattering (Academic Press).

Linding, R., Russell, R.B., Neduva, V., and Gibson, T.J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic acids research *31*, 3701-3708.

Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., Tyers, M., and Forman-Kay, J.D. (2010). Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. Structure *18*, 494-506.

Mukrasch, M.D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2009). Structural polymorphism of 441-residue tau at single residue resolution. PLoS Biol *7*, e34.

Netzer, W.J., and Hartl, F.U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature *388*, 343-349.

Petoukhov, M.V., and Svergun, D.I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. Curr Opin Struct Biol *17*, 562-571.

Rost, B., and Liu, J. (2003). The PredictProtein server. Nucleic acids research *31*, 3300-3304.

Schedlbauer, A., Ozdowy, P., Kontaxis, G., Hartl, M., Bister, K., and Konrat, R. (2008). Backbone assignment of osteopontin, a cytokine and cell attachment protein implicated in tumorigenesis. Biomolecular NMR assignments *2*, 29-31.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America *95*, 5857-5864.

Skolnick, J., Kolinski, A., and Ortiz, A.R. (1997). MONSSTER: a method for folding globular proteins with a small number of distance restraints. Journal of molecular biology *265*, 217-241.

Svergun, D.I. (1991). Mathematical methods in small-angle scattering data analysis. J Appl Crystallogr *24*, 485 - 492.

Svergun, D.I. (1992a). Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. J Appl Crystallogr *25*, 495-503.

Svergun, D.I. (1992b). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr *25*, 495 - 503.

Tompa, P. (2010). Structure and Function of Intrinsically Disordered Proteins (CRC Press).

Tsuboi, T., Takeo, S., Iriko, H., Jin, L., Tsuchimochi, M., Matsuda, S., Han, E.T., Otsuki, H., Kaneko, O., Sattabongkot, J., *et al*. (2008). Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. Infect Immun *76*, 1702-1708.

von Ossowski, I., Eaton, J.T., Czjzek, M., Perkins, S.J., Frandsen, T.P., Schulein, M., Panine, P., Henrissat, B., and Receveur-Brechot, V. (2005). Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulase. Biophysical journal *88*, 2823-2832.

Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., and Jones, D.T. (2004). The DISOPRED server for the prediction of protein disorder. Bioinformatics *20*, 2138-2139.

Wittekind, M., and Mueller, L. (1993). Hncacb, a High-Sensitivity 3d Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. J Magn Reson Ser B *101*, 201-205.

Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., and Uversky, V.N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochimica et biophysica acta *1804*, 996-1010.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics *21*, 3369-3376.

CHAPTER FOUR

ATS FAMILY

4.1 Introduction

*Pf*EMP1 is encoded by the family of *var* genes, of which 59 copies can be found in the genome of the 3D7 *P. falciparum* strain (Bull et al., 2005; Gardner et al., 2002). Other *P. falciparum* genomes are estimated to contain 50 to 60 *var* genes (Mercereau-Puijalon et al., 2002; Voss et al., 2000). The high copy number of *Pf*EMP1 is responsible for antigenetic variation, resulting in strategic evasion from immune system clearance. Although the intracellular part of *Pf*EMP1 is more conserved compared to its extracellular domain, variability exists within the ATS family. Different versions of ATS vary in length and sequence; the domain can be 290 – 500 amino acids long (Su et al., 1995). They share approximately 40 % to 85 % pair wise sequence identity (Lavstsen et al., 2003). Phylogenetic classification, based on the amino acid sequence of all ATS domains, divides the ATS family into four main clusters (Kraemer and Smith, 2003; Lavstsen et al., 2003). So far all our studies have concentrated on one variant of ATS, PF08_0141. Chapter three of this thesis characterises this specific domain as mainly intrinsically disordered but with an α-helical core structure. However, the question arises - is this structural architecture conserved within the ATS family? This chapter presents the results of our structural studies on additional variants of the ATS domain.

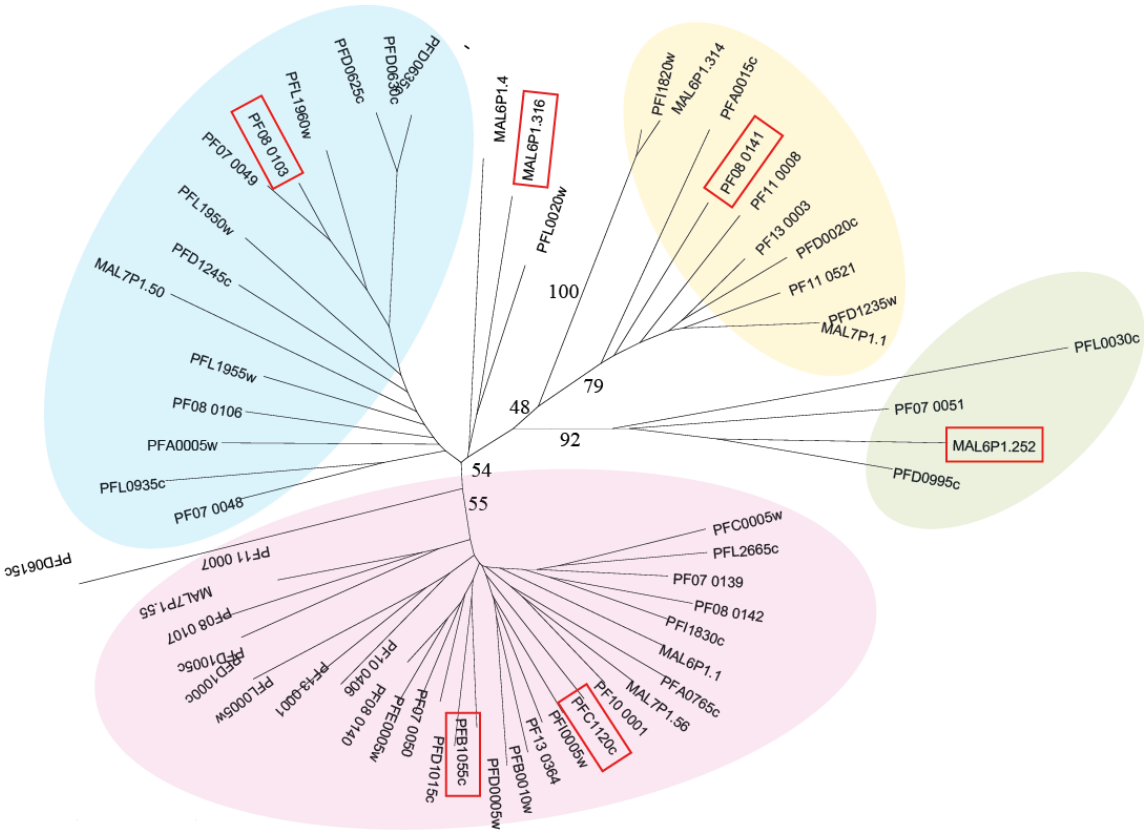


Figure 4.1: Phylogenetic tree of ATS domains of 3D7 *Pf*EMP1 (adapted from (Lavstsen et al., 2003)). The distance tree was generated using the p-distance/NJ method. Plasmodb accession numbers are shown (Bahl et al., 2002). ATS examples chosen to be representative of the different clusters are highlighted in red boxes. MAL6P1.316 and MAL6P1.252 correspond to PFF0010w and PF0845c, respectively.

4.2 Bioinformatic analysis of the ATS-family

Amino acid sequences of ATS domains from 3D7 strain of *P. falciparum* were subjected to disorder prediction by RONN (Yang et al., 2005). The results were consistent for all of the ATS sequences, predicting major parts of the domain to be disordered. For further analysis, we chose representative examples from each branch of the phylogenetic classification for structural characterization. The following five ATS full-length clones were utilized: PF08_0103 (residues 1692 -2109), PFB1055c (residues 1768 - 2177), PFC1120c (residues 1742 - 2169), PFF0010w (residues 2478 - 2879) and PFF0845c (residues 2047 - 2394). Table 4.1 gives an overview of the selected ATS domains and their features.

ATS domain	Amino acid (aa) length	predicted mass (kDa)	pI	Core residues	Core aa length
PF08_0141	393	45.5	4.55	109 -136 245 - 292	76
PFB1055c	410	47.5	4.68	120 - 147 256 - 303	76
PF08_0103	418	48.9	4.78	114 - 141 252 - 301	78
PFC1120c	428	49.1	4.51	140 - 168 277 - 324	77
PFF0010w	402	46.3	4.62	115 - 142 251 - 298	76
PFF0845c	348	40.4	5.18	118 - 145 264 - 313	78

Table 4.1: Overview of the additional six ATS domains of *P. falciparum* 3D7 strain studied. Molecular weight and theoretical isoelectric point (pI) were determined using ProtParam tool (Gasteiger E., 2005).

Multiple alignments of the primary sequence of the selected ATS variants revealed stretches of high conservation in the beginning and middle of the domain (figure 4.2) but the C-terminal region shows high sequence divergence. In total all six domains share pair-wise sequence identity of 65 %. In addition, disorder probabilities shown in figure 4.3 strengthen the assumption that the ATS family architecture is of a flexible nature. To confirm this hypothesis we chose to experimentally analyse the selected ATS domains by CD, NMR and SAXS.

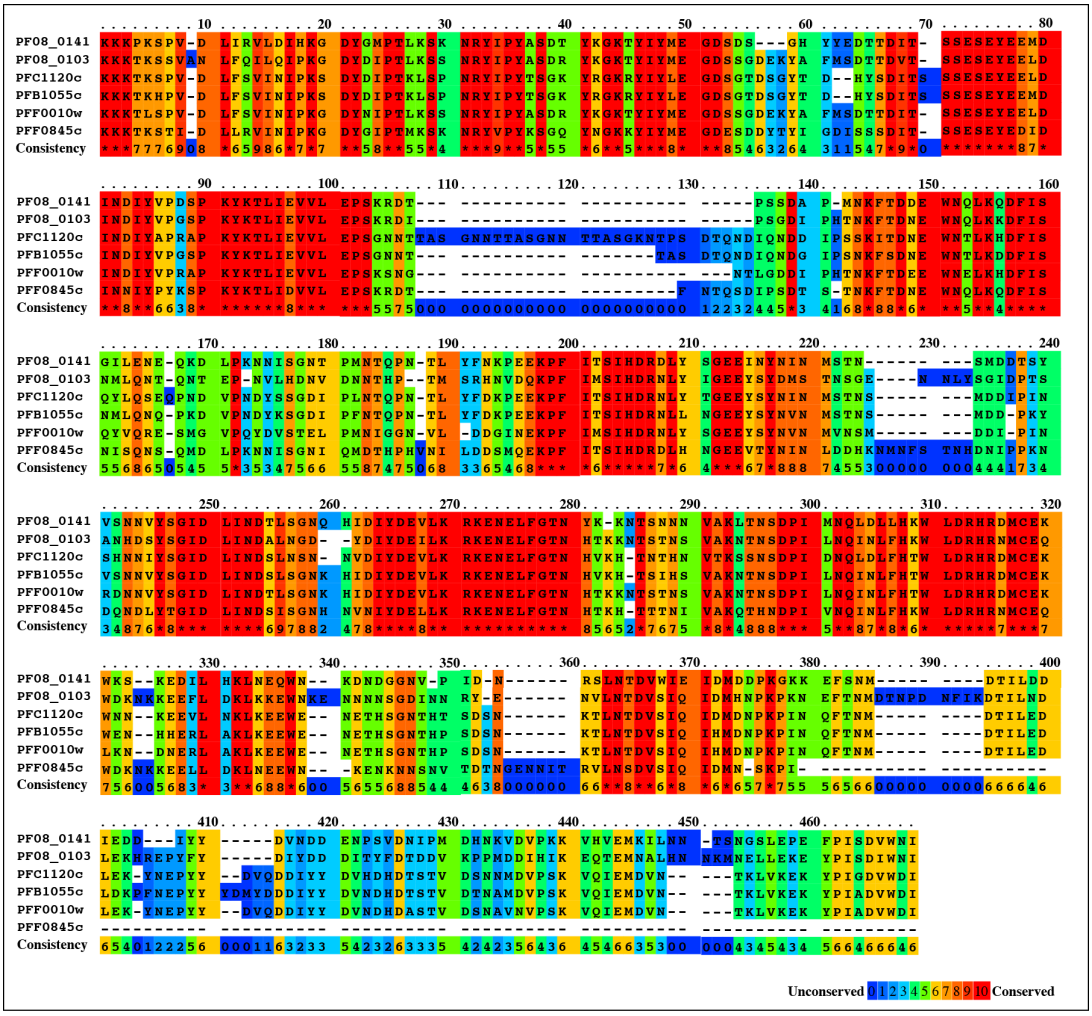


Figure 4.2: Alignment and amino acid conservation of selected ATS domains. The conservation scoring was performed by PRALINE (Simossis and Heringa, 2005). The colour coding of the conservation is shown on the bottom right. The colour red points out the most conserved alignment positions.

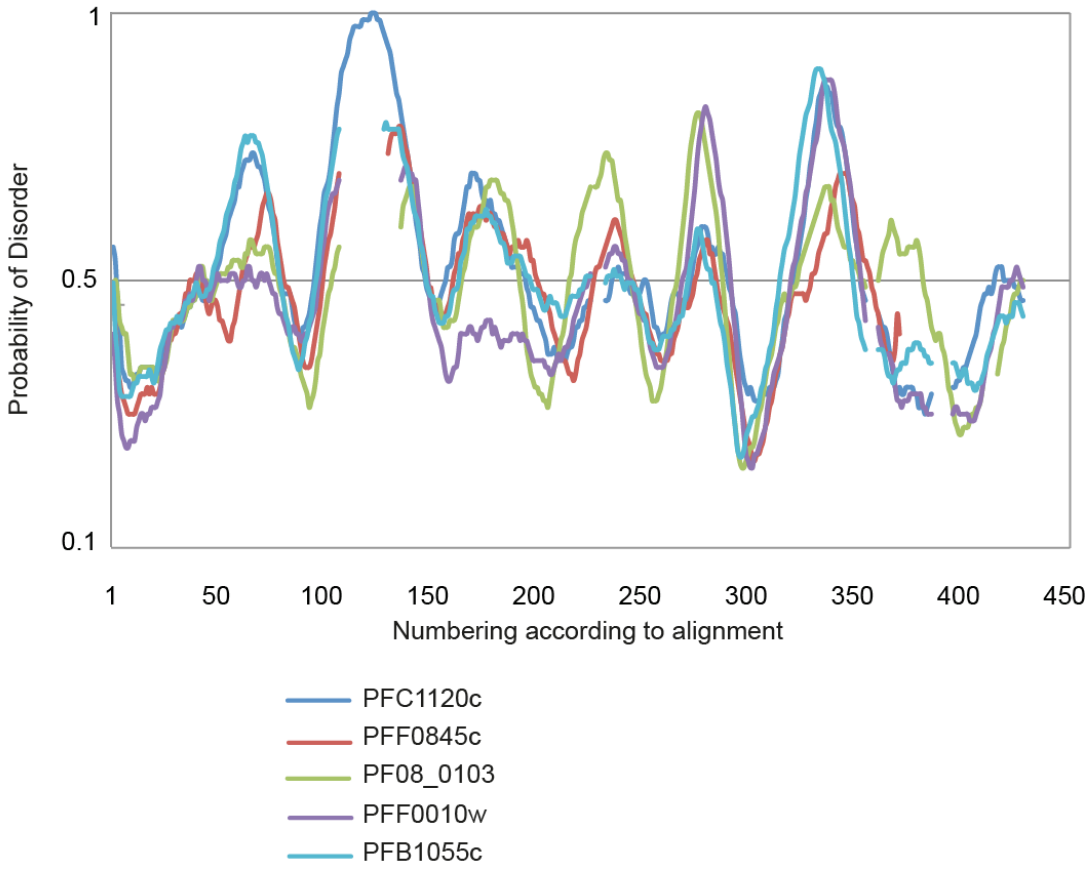


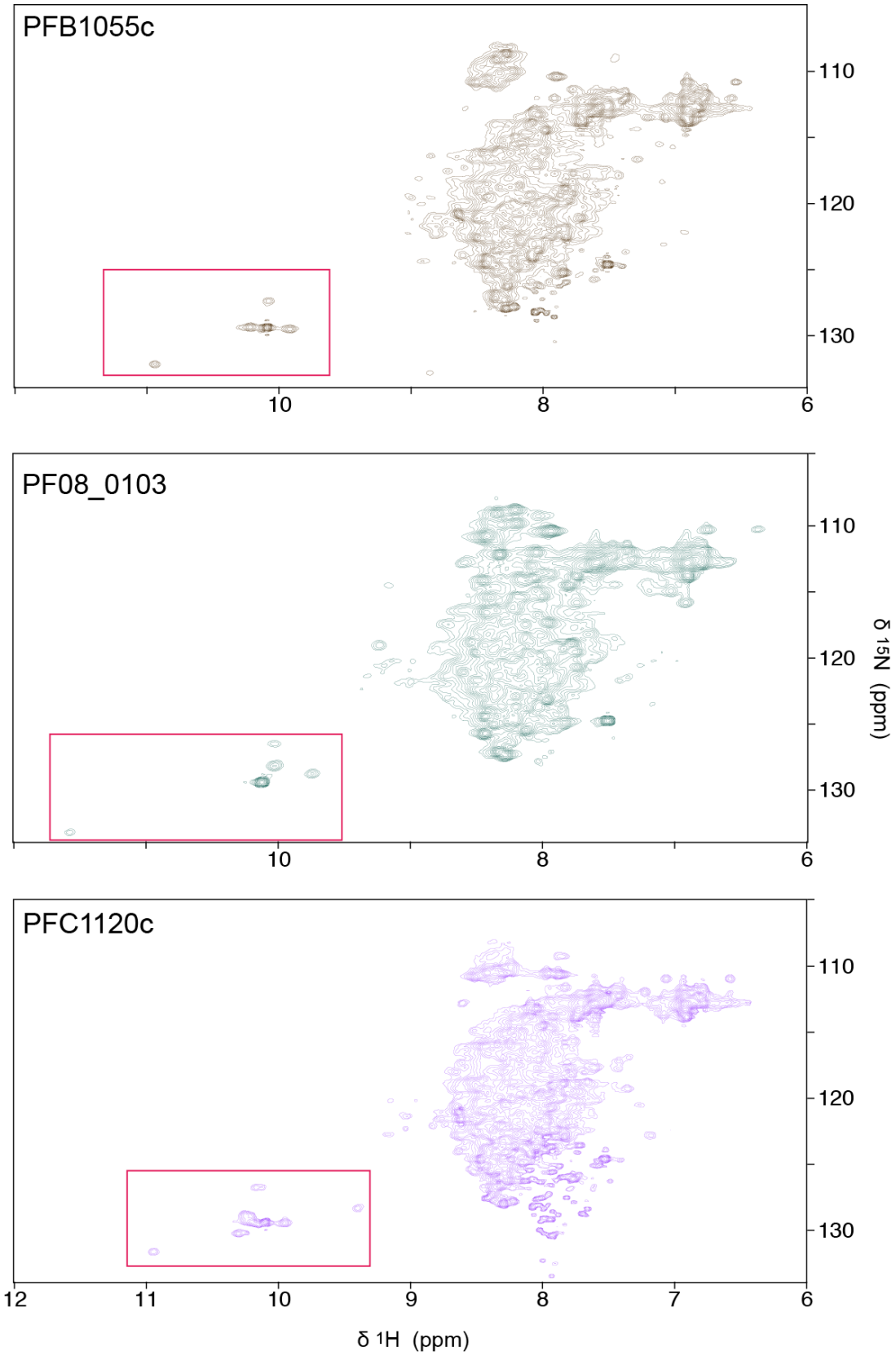
Figure 4.3: Predicted probability of disorder for selected ATS domains. RONN disorder prediction ATS-FL variants (Yang et al., 2005). The threshold (0.5) for disorder probability is marked in red. A score above the threshold 0.5 is considered disordered. Gaps and residue numbering are in accordance with the performed sequence alignment (figure 4.2.).

4.3 Protein production and purification

The relevant DNA sequences for most full-length ATS were amplified from cDNA or genomic DNA and cloned in the pET16b-3C vector. ATS PFF0845c was obtained in a modified pET15 vector from Dr Matt Higgins (University of Oxford). The constructs for the corresponding ATS-core fragments were designed to be homologous to ATS-core PF08_0141 and cloned in pET16b-3C. Table 4.1 lists the residues used for the different constructs. Protein expression was performed in M9 minimal or LB media using the pET16b-3C vector and purification was partly done under denaturing conditions as described before (section 3.3) except for the full-length version of PFF0845c. This protein was expressed under native conditions in the same manner as PF08_0141 (section 3.3).

4.4 NMR data on ATS family

As expected from the disorder predictions, all the ^{1}H- ^{15}N HSQC spectra of ^{15}N enriched ATS domains show limited chemical shift dispersion, particularly in the ^{1}H dimension. Figure 4.4 shows the relevant NMR spectra for the selected full-length ATS domains PFB1055c, PF08_0103, PFC1120c, PFF0010w and PFF0845c. The spectra are highly similar to that of ATS-FL (PF08_0141), showing a high amount of chemical shift overlap indicating a high degree of intrinsic disorder. Moreover, distinct folded resonances are rare in the spectra, demonstrating that only minor parts of the domains are folded, as anticipated. Therefore, we could conclude that the ATS family is dominated by large flexible elements.


PFB1055c
PF08_0103
PFC1120c
110
120
130
10
8
6
12
11
9
7
δ 15N (ppm)
δ 1H (ppm)

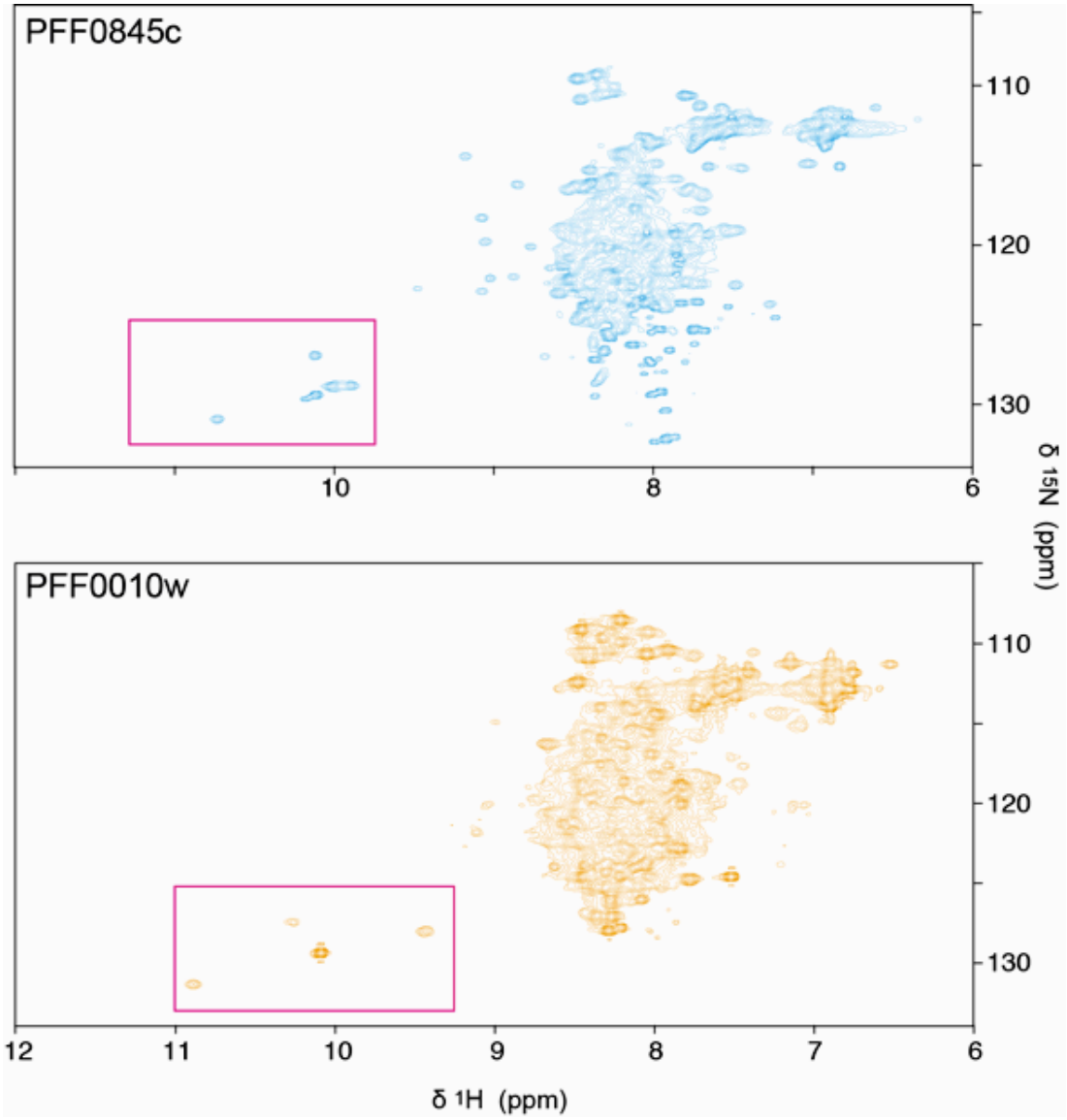


Figure 4.4: ^{1}H- ^{15}N HSQC spectra of full-length ATS domains. The peaks in the proton dimension show lack of chemical shift dispersion. The lack is consistent for all tested ATS variants, suggesting a mainly flexible nature for the ATS family. Only minor parts seem to contain structured elements. The NMR spectra of PFF0845c and PFC1120c show signs of degradation. Tryptophan side-chain peaks are boxed.

4.6 SAXS derived data on PFF0845c

We also collected SAXS data on the PFF0845c full-length domain. The corresponding Kratky plot is shown in figure 4.5C. As expected, the Kratky plot displays a relative minimum and increases monotonically, which typifies a mainly disordered protein. Basically, the plot shows a similar profile to that of PF08_0141 (figure 3.7). Further analysis of the scattering data was performed to determine the approximate radius of gyration (R_g). The R_g for PFF0845c was estimated 4.23 nm using the Guinier approximation. The theoretical value for a compact molecule and a chain-like conformation of this size is 2.05 nm and 6.04 nm, respectively (calculated as described in section 3.3.2). Therefore, analysis of the SAXS data also suggests that the ATS domain PFF0845c is largely extended and disordered. The molecular weight of ATS variant PFF0845c (39.7 kDa) 42.7 kDa was estimated from the experimental I(0) values relative to I(0) of BSA (66 kDa), which was 73.74 in this experiment. This confirms that the protein is monomeric under the given solution conditions. Because ATS PFF0845c was purified under native conditions, these results support our observations that our refolding strategy of other ATS-FL variants was successful and yielded the native state of the protein.

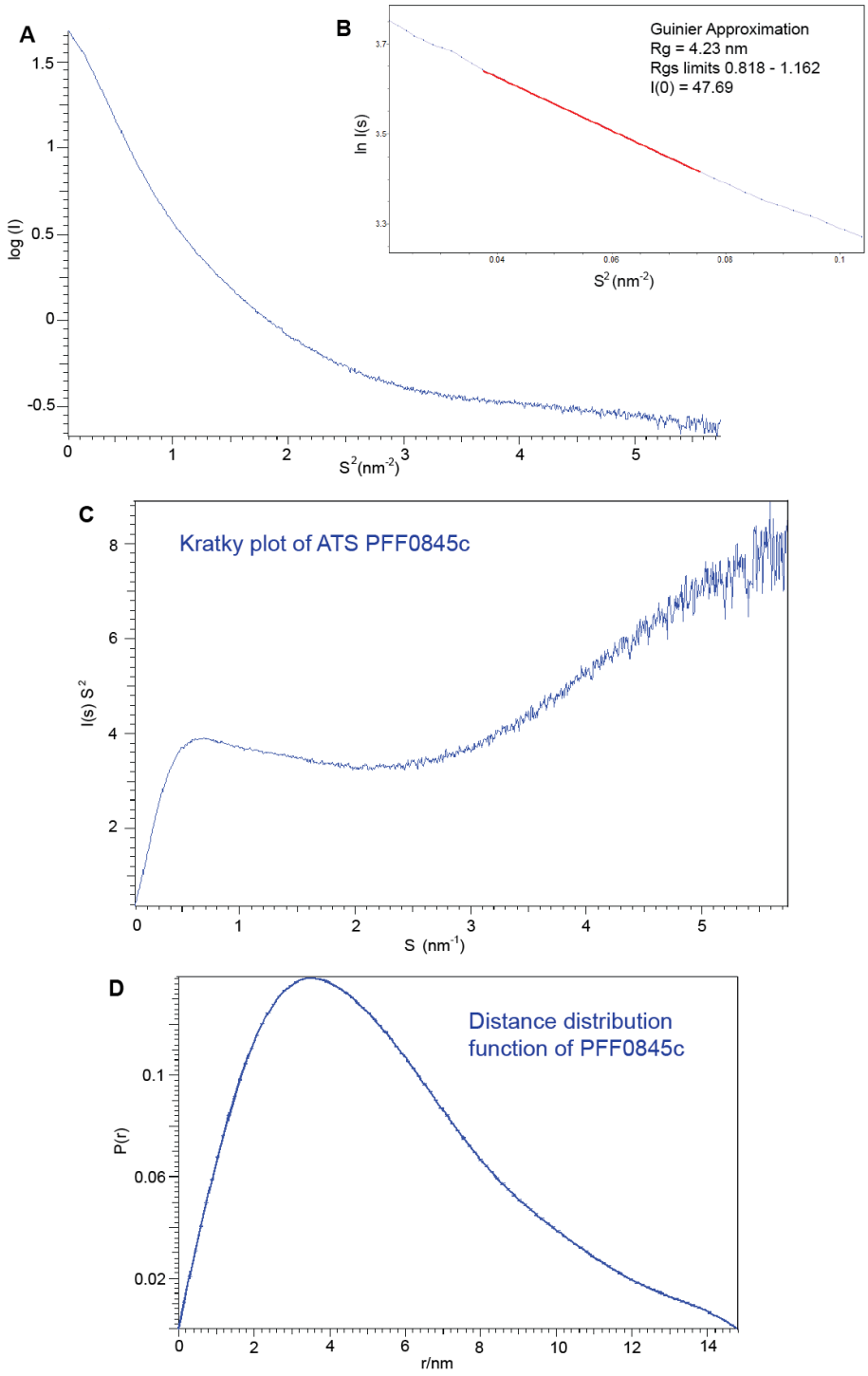


Figure 4.5: SAXS analysis of ATS PFF0845c. **A)** Guinier plot of the SAXS data from the full-length ATS variant. **B)** The section of the the data selected by AUTORG (Petoukhov et al., 2007) as satisfying the Guinier approximation is shown as a red line. **C)** The Kratky representation shows a slight hump and then increases monotonically. **D)** The pair distribution function generated by GNOM (Svergun, 1991, 1992) showing a similar profile to that of ATS-FL PF08_0141.

4.5 Existence of ATS-core structures

We were able to show that the additional ATS domains contain intrinsically disordered regions. Chapter III of this thesis described the analysis of ATS PF08_0141. Our studies revealed a structured core element within the domain of this variant. However, is the α-helical core structure determined for PF08_0141 conserved throughout the family? To answer this question, we designed homologous core constructs of the six additional ATS domains. The far-UV CD spectra of the constructed ATS-cores are displayed in figure 4.6. All variants show a largely α-helical character with typical minima at 222 nm and 208 nm. In addition, CD melting curves (figure 4.6 B), recorded at 222 nm indicated cooperative melting of the α-helical region. Protein unfolding occurs in the temperature range 45 °C to 75 °C. NMR data of the core constructs show typical NMR chemical shift dispersion for folded proteins, with a chemical shift range in the ^{1}H dimension of 6.5 ppm to 10.0 ppm (figure 4.7). In addition, the comparison of the tryptophan side chain residues in the corresponding ATS-FL and ATS-core spectra showed that the resonance positions, and hence the chemical environments for tryptophan residues are similar. This ensures that the ATS-core structured element is present in the full-length constructs. Furthermore, the HSQC spectrum of PF08_0103 core is more crowded in the range of 8.0 ppm to 8.5ppm in the proton dimension. This indicates a more flexible protein state compared to the other ones. Interestingly, this variant was also predicted to be more disordered, according to the RONN algorithm, than the other domains (figure 4.3).

The data presented in this chapter argues for a common architecture of the ATS family. All analysed domains are comprised of large unstructured parts and a folded

helical core. Therefore, we propose that the structural conformation of ATS PF08_141 is conserved throughout the *Pf*EMP1 family.

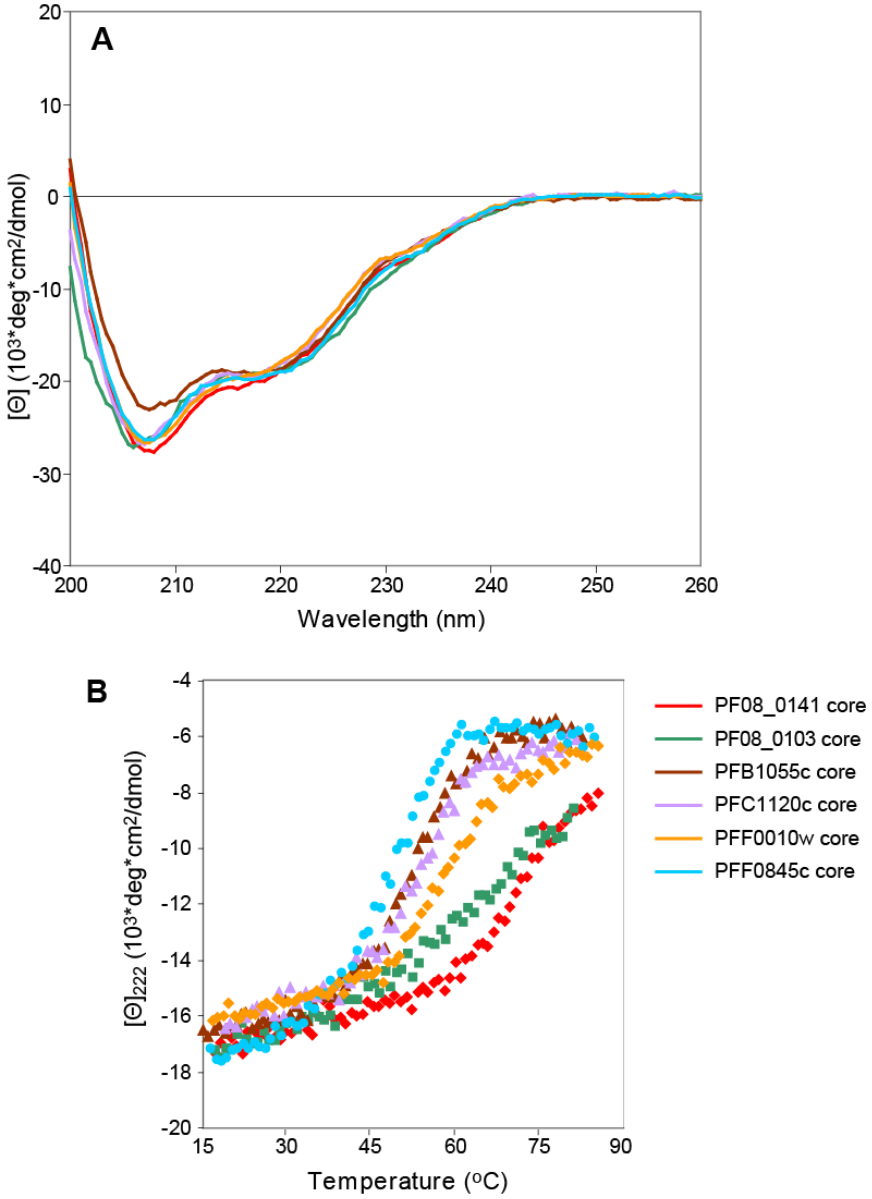
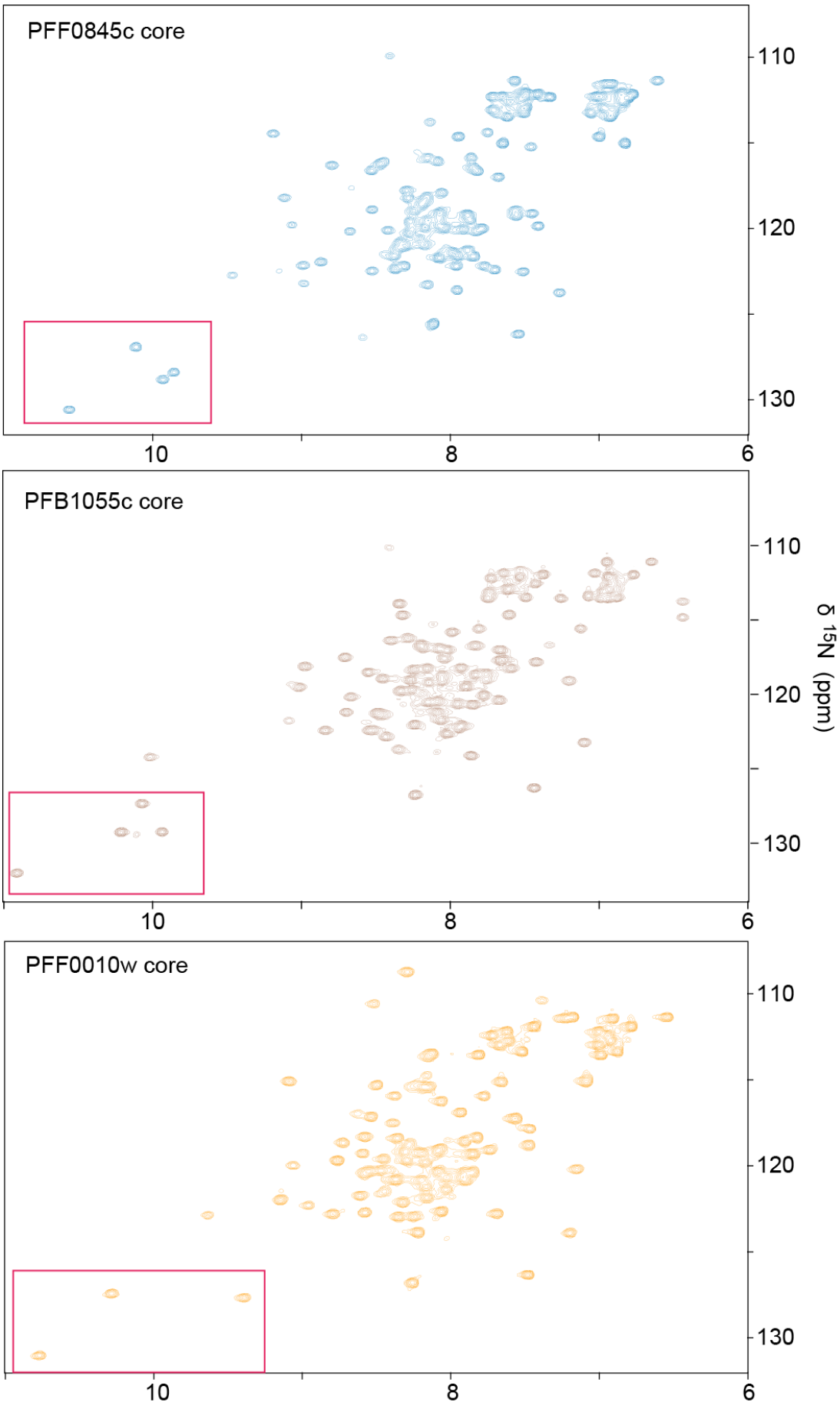


Figure 4.6: CD spectra and thermal stability assays of additional ATS-cores. All experiments were recorded in PBS buffer at 10 °C and at protein concentrations of 25 µM. **A)** The CD data of all ATS-core constructs is plotted as mean residue elipticity (MRE) against wavelength (nm). All spectra show clear evidence of α-helical secondary structure. **B)** Thermal stability data were recorded at 1 °C intervals. Results show that the proteins unfold cooperatively upon heating.


PFF0845c core
PFB1055c core
PFF0010w core
110
120
130
10
8
6
δ 15N (ppm)

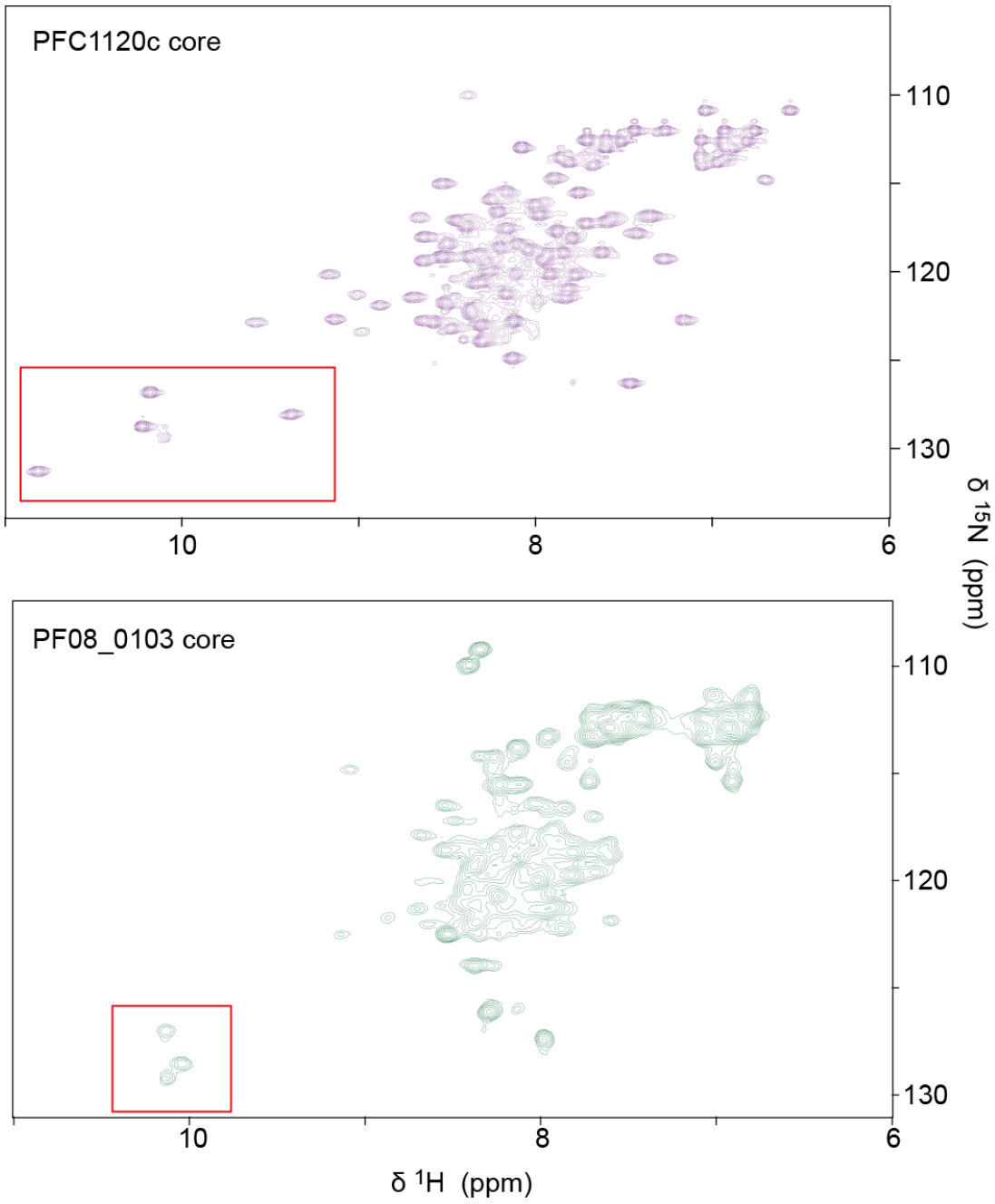


Figure 4.7: ^{1}H- ^{15}N HSQC spectra of ATS-core constructs. The spectra of the uniformly ^{15}N-labelled samples were recorded at 25 °C in NMR buffer pH 7.0. The tryptophan side-chain resonances are boxed.

4.6 References

Bahl, A., Brunk, B., Coppel, R.L., Crabtree, J., Diskin, S.J., Fraunholz, M.J., Grant, G.R., Gupta, D., Huestis, R.L., Kissinger, J.C., *et al*. (2002). PlasmoDB: the Plasmodium genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished). Nucleic acids research *30*, 87-90.

Bull, P.C., Berriman, M., Kyes, S., Quail, M.A., Hall, N., Kortok, M.M., Marsh, K., and Newbold, C.I. (2005). Plasmodium falciparum variant surface antigen expression patterns during malaria. PLoS pathogens *1*, e26.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., *et al*. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498-511.

Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, Humana Press (2005) (Humana Press), pp. 571 - 607.

Kraemer, S.M., and Smith, J.D. (2003). Evidence for the importance of genetic structuring to the structural and functional specialization of the Plasmodium falciparum var gene family. Molecular microbiology *50*, 1527-1538.

Lavstsen, T., Salanti, A., Jensen, A.T., Arnot, D.E., and Theander, T.G. (2003). Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. Malaria journal *2*, 27.

Mercereau-Puijalon, O., Barale, J.C., and Bischoff, E. (2002). Three multigene families in Plasmodium parasites: facts and questions. International journal for parasitology *32*, 1323-1344.

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., and Svergun, D.I. (2007). ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. J Appl Crystallogr *40*, S223-S228.

Simossis, V.A., and Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic acids research *33*, W289-294.

Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., Ravetch, J.A., and Wellems, T.E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. Cell *82*, 89-100.

Svergun, D.I. (1991). Mathematical methods in small-angle scattering data analysis. J Appl Crystallogr *24*, 485 - 492.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr *25*, 495 - 503.

Voss, T.S., Thompson, J.K., Waterkeyn, J., Felger, I., Weiss, N., Cowman, A.F., and Beck, H.P. (2000). Genomic distribution and functional characterisation of two distinct and

conserved Plasmodium falciparum var gene 5' flanking sequences. Mol Biochem Parasitol *107*, 103-115.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics *21*, 3369-3376.

CHAPTER FIVE

ATS SOLUTION STRUCTURE

The ATS-core NMR structure refinement, presented in this chapter, was performed by Dr Michele Erat (University of Oxford) and ATS-25 and ATS-FL NMR structure ensembles were calculated by Dr Ioannis Vakonakis (University of Oxford).

5.1 Crystal trials of ATS-core

X-ray crystallography can be a fast tool to determine protein structures at atomic resolution, and it relies on the formation of crystals. Crystallographic studies were thus attempted to generate a high-resolution structure of ATS-core. The ATS-core construct was expressed and purified as described in section 5.2 and subjected to a pre-crystallization test (Hampton Research) to determine the appropriate sample concentration for subsequent crystallization screening. The protein was concentrated to 12 mg/ml and crystal screens were set up at the Oxford Protein Production Facility (Division of Structural Biology, Oxford). In total 288 conditions were tested at 4 °C and 21 °C. Unfortunately, no crystals were forthcoming in screens.

5.2 Structure determination of ATS-core using NMR

Since we were not fortunate in obtaining protein crystals we chose NMR solution spectroscopy as an alternate route for solving the ATS-core structure. This method does not require crystals and can deal with relatively flexible and small proteins.

5.2.1 Assignment of ATS-core primary sequence

The ATS-core sequence of PF08_0141 corresponds to residues M109 to L136 and V245 to D292, additionally it includes an N- terminal cloning artifact of five residues (GPLGS). Therefore the construct consists of 81 residues with residue M6 corresponding to M109 of ATS-FL. A doubly enriched ^{13}C-^{15}N NMR sample was recombinantly expressed in *E. coli* and standard three-dimensional NMR experiments - CBCA(CO)NH, HNCO and HNCACB - were recorded (Cavanagh, 2007). Usually, the experiments were analysed in pairs with one experiment detecting both intra- and interresidue correlations and the second recording only interresidue correlations. Resonance assignment was accomplished with the help of average amino acid ppm values, available in the Bio Magnetic Resonance Bank (http://www.bmrb.wisc.edu). The backbone ^{13}C, ^{15}N and amide ^{1}H shifts were assigned for residues P2 to D81, as appropriate. The two-dimensional ^{15}N- ^{1}H HSQC spectrum of ATS-core is displayed in figure 5.1.

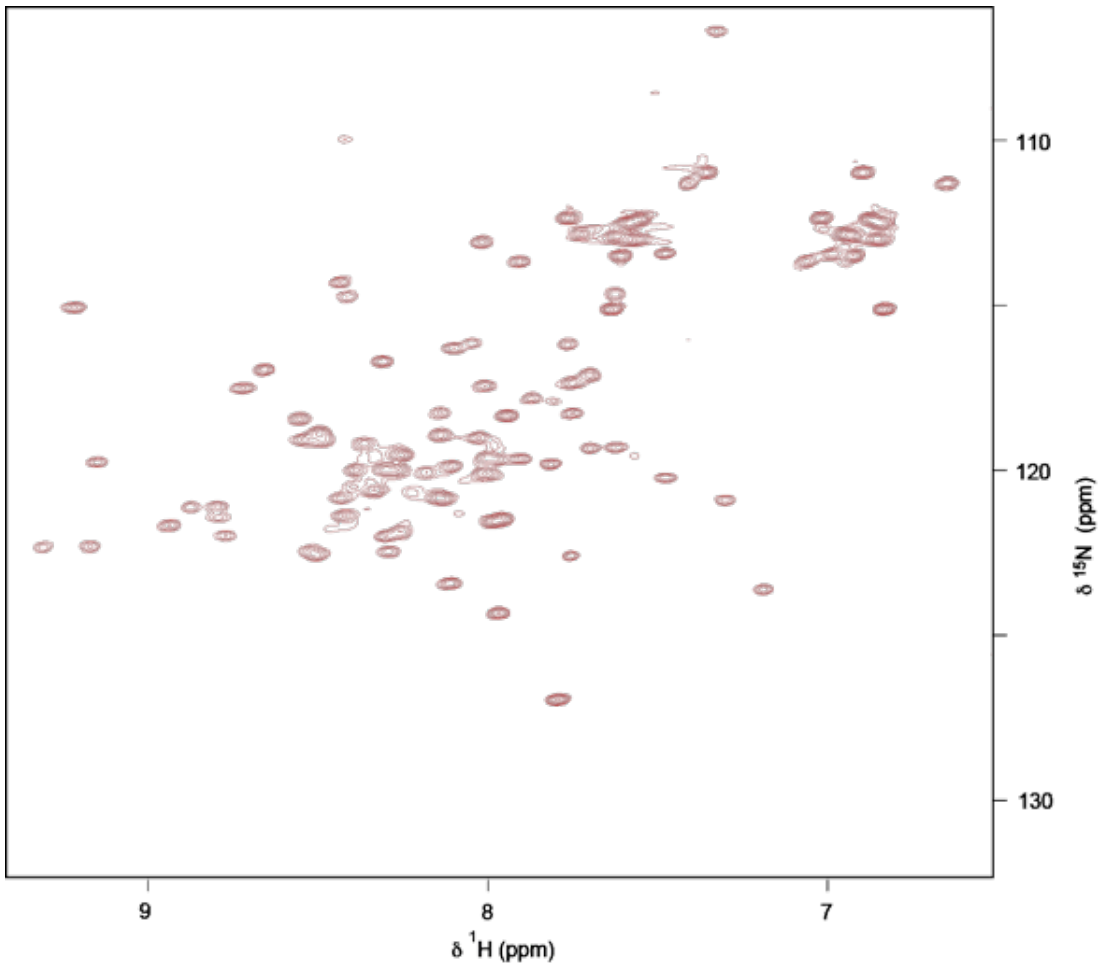


Figure 5.1: The ^{1}H-^{15}N HSQC spectrum of ATS-core recorded at 25°C. Residues M6 - L33 and V35 - D81 of ATS-core construct corresponds to M109 - L136 and V245 - D292 of ATS-FL, respectively.

The TALOS program was used to assess the backbone dihedral angles of ATS-core from the amino acid sequence and chemical shifts (Cornilescu et al., 1999). The TALOS database takes secondary chemical shifts and Hα, Cα, Cβ, CO, N shifts of each residue as well as its preceding and following neighbours, and compares the values to its databank. The ten best matches for each triplet are selected and quantitative predictions for the protein backbone angles Ψ and Φ are made. For ATS-core 106 backbone torsion angle restraints were obtained using TALOS.

5.2.2 Side-chain and structural assignment of ATS-core

The assignment of the side-chains was achieved starting from the sequential assignment of the backbone described above. For this purpose the following NMR experiments were acquired and subsequently analysed: H(C)CH-COSY, ^{1}H-^{13}C NOESY-HSQC, ^{1}H-^{15}N NOESY-HSQC, HBHA(CO)NH, C(CO)NH, H(CCO)NH HNHA, HNHB and a four-dimensional ^{13}C-^{13}C NOESY. Some amino acids were easier to assign from their characteristic frequencies than others. Methyl resonances mainly appear around 0 - 2 ppm in a NOESY spectrum and the Hα resonances can be observed at approximately 4 - 5.5 ppm. In the case of glycine, Hαs usually resonate below 4 ppm and are therefore easier to locate. Additional information was gained by assigning a H(C)CH-COSY and ^{1}H-^{13}C NOESY-HSQC. Complete spin systems were assigned in the H(C)CH-COSY and the relevant residues were also detected in the ^{1}H-^{13}C NOESY-HSQC. The assignment of these spectra was a prerequisite for the collection of further nuclear Overhauser effect (NOE) restraints needed for the later structure calculations. The chemical shift assignments of ATS-core are listed in Appendix IV, and have been deposited in the BMRB under accession number 16911.

5.2.3 Calculation of the structures

Three-dimensional structures were calculated using experimentally observed NOE restraints, backbone torsion angle predictions, experimentally determined sidechain torsion angles, backbone scalar-coupling and residual dipolar coupling restraints. 1145 diverse experimental NOE cross peaks were derived from ^{15}N-edited and ^{13}C-edited NOESY-HSQC spectra. Of those 383 were intraresidue *(|i-j|=0)* and 256 sequential interresidue *(|i-j|=1)*. Furthermore, these NOEs included 310 short range

(|i-j|<5) and 196 long range constraints *(|i-j|≥5)*. As mentioned before (2.14.2) a simulated annealing procedure using the Xplor-NIH software package (Schwieters et al., 2003) was chosen for structure calculations in torsion-angle space, starting from an extended conformation. Further refinement was accomplished by a simulated slow-cooling process in Cartesian space.

Intraresidue	383
Interresidue, sequentiell (|i-j|)	256
Interresidue, medium-range (|i-j|=<5)	310
Interresidue, longrange (|i-j|>=5)	196
Total NOE-distance restraints	**1145**
Backbone Torsions (TALOS)	106
$^{3}J_{HNH\alpha}$ couplings	34
$^{1}D_{HN}$ couplings (RDC)	45
$^{13}C\alpha$, $^{13}C\beta$ shifts	77

Table 5.1: Statistics of NMR derived distance and dihedral angle restraints for ATS-core.

Table 5.1 presents the final NMR restraints used for the calculations. We obtained an ensemble of 39 final structures with a RMSD backbone average of 0.56 ± 0.10 Å and a heavy atoms average of 0.99 ± 0.08 Å (calculated for residues 8 - 80). Figure 5.2 shows one of the calculated ATS-core structures. As the cartoon representations in figure 5.2 shows, the ATS-core is composed of four α-helices (α1 - α4); together they form a compact helix bundle. The N- terminal region (residues M109 - K111) proved to be flexible, as expected from Cα, CO and Hα chemical shift analysis (chapter III). α1 starts with residue F112 as anticipated. A detailed analysis of the NMR structure quality is given in Table 5.2. The calculated structures have been deposited in the RCSB Protein Databank under the accession code 2LKL. Amino acid

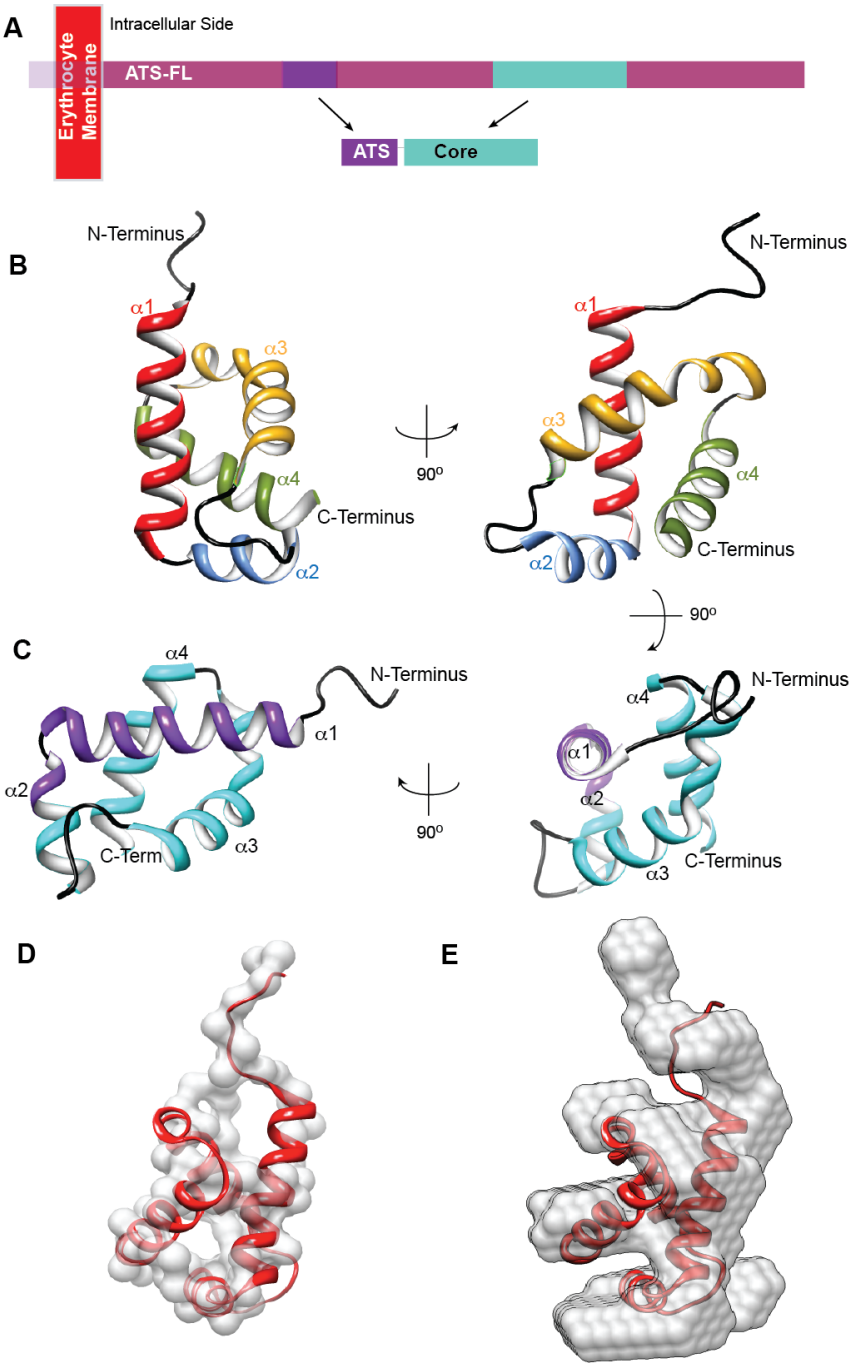
sequence alignments of the ATS region of other variants (figure 4.2) suggest that helices α1, α3 and α4 are conserved. In particular, the residues building the hydrophobic nucleus of the ATS-core are almost entirely conserved or substituted in a conservative way in ATS-domains. Helix α1 and the beginning of α2 derive from residues 112-136, in our construct α2 then continues via residues 245-250. Residues comprising α2 originate from the first and the second structured element determined in ATS-FL, where the flexible middle loop had been excised. However, ATS-FL residues 245-250 do not show substantial helical propensity (figure 3.16), therefore this extension of α2 is probably an artefact of construct engineering. Hence, the finding that helix α2 is not conserved, is not surprising as this secondary structure element was created artificially. A heteronuclear $\{^{1}H\}$-^{15}N-heteronuclear NOE experiment was used to map the internal motion of residues (figure 5.3), showing helices α1, α3 and α4 to be relatively rigid. Lower values indicate regions with fast internal mobility, the majority falling within the loop regions of the protein.

The SAXS data of ATS-core was used to create low-resolution models for the molecule. The programmes DUMMIF (Franke and Svergun, 2009) and GASBOR (Svergun et al., 2001) were utilised for this purpose. DAMMIF performs simulated annealing to construct a compact connected three dimensional model starting from an arbitrary initial bead model, yielding a scattering pattern that fits the experimental data (Franke and Svergun, 2009). GASBOR uses connected dummy residues to find a chain-compatible spatial distribution to fit the SAXS data (Svergun et al., 2001). The obtained low-resolution models were aligned with the high-resolution structure of ATS-core determined by NMR (figure 5.2).

Ramachandran statistics	**(%)**	**Structure precision**	(Å)
most favoured regions	87.2		
allowed regions	9.3	Backbone Atoms	0.56±0.10
generously allowed regions	2.6	Heavy Atoms	0.99±0.08
disallowed regions	0.9		

RMSDs from experimental restraints	
Distance restraints (Å)	0.021±0.002
Distance violations > 0.3 Å	0
Dihedral angles (deg)	0.463±0.074
Dihedral angle violations > 3°	21
RMSDs from idealized geometry	
Bonds (Å)	0.0023±0.0003
Angles (deg)	0.49±0.02
Impropers (deg)	0.39±0.02

Table 5.2: Structure quality statistics for ATS-core. Ramachandran statistics were calculated using PROCHECK (Laskowski et al., 1996). Structure precision was determined for residues 8 - 80 of ATS-core corresponding to ATS -FL F112 - L136 and V245 - K291. Values were calculated for a final ensemble of 39 structures.



ATS-FL and the designed ATS-core construct. The structured segments of ATS-FL are coloured in purple and cyan. **B)** Different orientations of the ATS-core structure. α-helices are highlighted in different colours. **C)** Shown in purple and cyan are the two parts originating from the first and second folded segment of ATS-FL, respectively. N- and C-termini are indicated in all representations. **D)** Atomic model of ATS-core (red) superimposed on to the models of ATS-core obtained by GASBOR and DUMMIF (E). The models were aligned by SUPCOMP (Kozin et al. 2001).

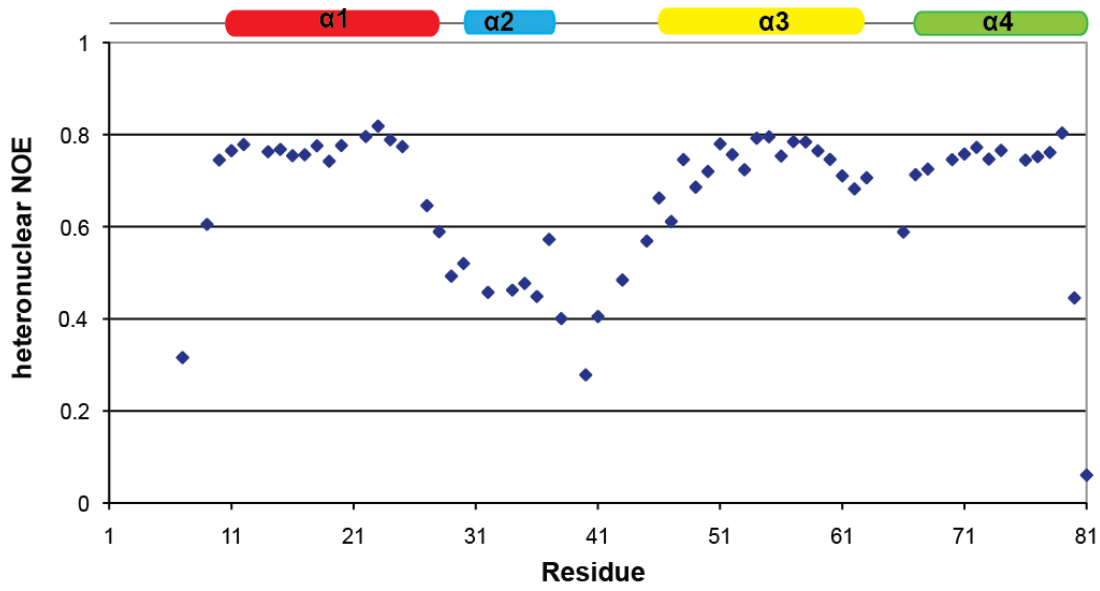


Figure 5.3: {^{1}H}-^{15}N-heteronuclear NOE data for ATS-core. The positions of the α-helices are marked at the top of the figure. Data was recorded at 25 °C in NMR buffer using a Bruker 500 MHz.

5.3 Solution structure of ATS-25

We applied different approaches to obtain a structural model for ATS-25. On one hand we tried to generate an average conformation using the SAXS *ab initio* shape determination programs provided by the ATSAS package (Petoukhov et al., 2007). On the other hand, we calculated a molecular ensemble for ATS-25 based on the structural restraints obtained from ATS-core NMR data and additional chemical shift restraints from the ATS-25 backbone resonances.

Several programs, greatly improved recently, are available for the creation of three-dimensional structure models from one-dimensional SAXS intensities at low resolution. The program DAMMIN creates low-resolution shapes of the protein of interest from SAXS data using simulated annealing to perform minimization (Svergun, 1999). Despite the fact that ATS-25 does not occupy a singular, stable three-dimensional structure, we created average conformations of ATS-25 using the mentioned programs DAMMIN (Svergun, 1999) and GASBOR (Svergun et al., 2001). The programs were run according to standard protocols. The models of ATS-25 obtained displayed overall similar features. Figure 5.4 A and B show representations of the calculated structures; they both appear to be extended with a possible accumulation of density in the central part of the molecule. These results however, should be examined critically. For a monodisperse protein solution the SAXS scattering intensity is proportional to the single molecules scattering over all orientations. However, in the case of IDPs the conformational averaging is not taken into account. Therefore, we used the ensemble optimization method (EOM), which specializes in the analysis of unfolded, flexible proteins (Bernado et al., 2007). The

method allows for the co-existence of multiple conformations in solution (due to protein flexibility). It randomly generates a large pool of models and selects for models covering the protein conformational space. Subsequently, the optimal number of conformers in the ensemble is determined using a genetic algorithm to fit the experimental SAXS data (Bernado et al., 2007). We employed EOM for further analysis of ATS-25 and ATS-FL (section 5.4). The R_g distribution from the selected models was compared to the distribution of the non-optimized initial ensemble (figure 5.5). The R_g distribution offered by EOM shifted to smaller values with an average of R_g ~3 nm. The EOM optimized ensemble fits the experimental data to a resolution of approximately 4 Å.

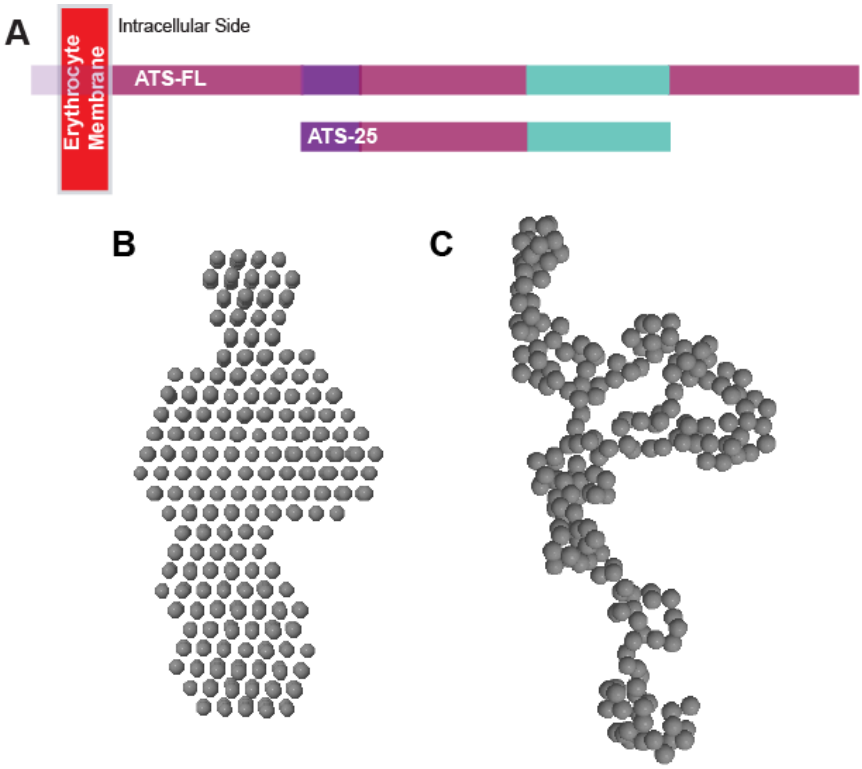


Figure 5.4: ATS-25 solution structure derived from SAXS data. A) Schematic representation of ATS-FL and the ATS-25 construct. The structured segments of ATS-FL are coloured in purple and cyan. **B)** ATS-25 DAMMIN model. **C)** Dummy residue model from solution scattering data obtained by GASBOR.

We also calculated a molecular ensemble for ATS-25 based on the structure of the ATS-core. We used the restraints of ATS-core obtained from NMR and added the missing sequence in a random coil conformation. For the additional flexible parts, we utilized the assigned carbon chemical shifts to derive restraints and, in addition, activated pseudo-potentials for optimizing the Ramachandran and hydrogen bond properties of each residue. Figure 5.5 presents a bundle of structures produced in this way. As expected from previous NMR data analysis of ATS-25 and ATS-FL (section 3.4.7) the middle loop and the N-terminus both lack structural order. The sole

compact unit in ATS-25 is the core structure, which also restrains the flexible middle part to form a closed loop. Computed R_g values for the minimized structures average at 3 nm, which compares well to R_g = 3.33 nm obtained using the Guinier approximation applied to collected SAXS data (section 3.4.5). This agreement between the NMR-based and SAXS-based ensembles suggests that our NMR model of ATS-25 is representative of the protein state in solution.

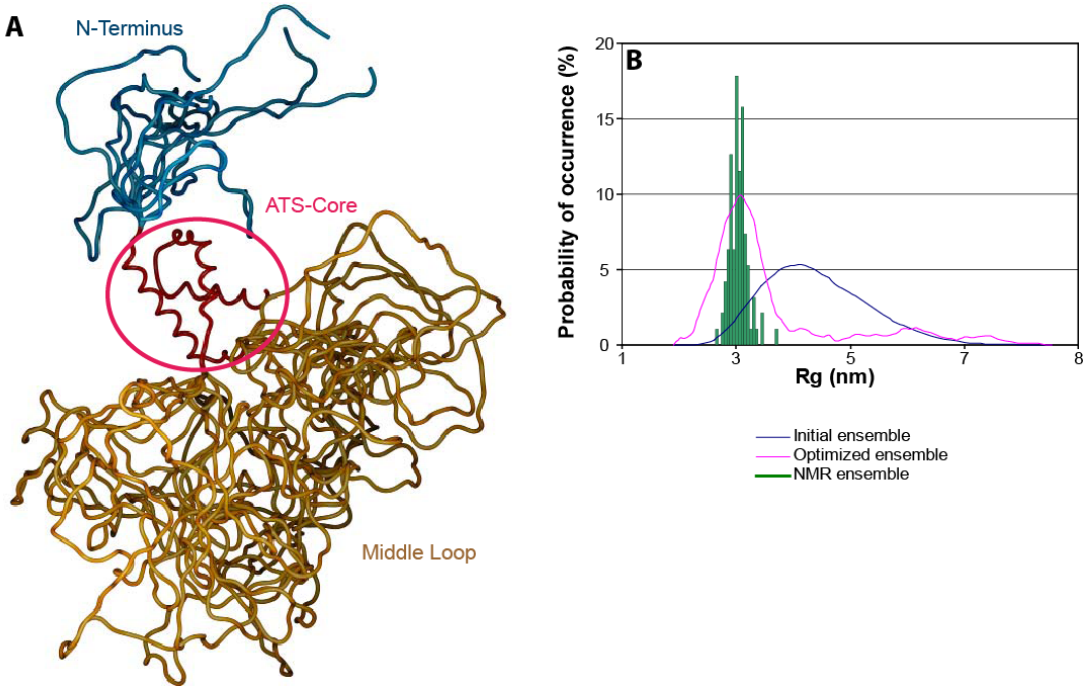


Figure 5.5: ATS-25 structure ensemble based on NMR data. A) The red circle highlights the core structure of ATS. The middle loop is coloured in brown and the N-terminus is shown in blue. The N-terminal part of the protein as well as the middle loop display a high degree of flexibility and therefore occupy different conformations and give rise to the extended shape of the protein. **B)** The R_g probability distribution of the initial pool of ATS-25 ensembles (in blue) and after EOM analysis (magenta). The R_g distribution of the ATS-25 structures calculated from NMR data are shown in green bars.

5.4 Solution structure of ATS-FL

To determine the overall shape of ATS-FL we used a similar method. The SAXS data collected on ATS-FL were also analysed by EOM (Bernado et al., 2007). We were able to obtain good fits to the experimental data with multiple-curve modelling. The R_g distribution of the randomly created pool of conformers is centred around 6 nm, whereas, the distribution of R_g for the selected models (by EOM) shows two populations, one at around 4.6 nm, indicating a more compact state of ATS-FL and a second more extended one at approximately 7.8 nm (figure 5.6C).

In addition, we used simple molecular simulations, incorporating Ramachandran and hydrogen bond pseudo-potentials, and chemical shift restraints starting from the ATS-core structure. Figure 5.6 shows an ensemble of calculated ATS-FL structures. The molecule is extended and N- and C- termini are arranged in opposing directions. An average R_g value of 7.7 nm was determined for this conformation. This ensemble likely corresponds to the larger (7.8 nm) ATS-FL R_g distribution inferred from the SAXS data. According to the determined R_g value distribution of ATS-FL, derived from EOM analysis of the SAXS data, we incorporated an R_g limit of 4.8 nm into our simulations. This R_g limit corresponds well to the R_g of 4.7 nm determined using the Guinier approximation (chapter III). The resulting structure bundle is displayed in figure 5.6A. Evidently, the conformations of the models become more compact and the ends of the protein converge. The ATS-FL central loop is well exposed pointing away from the termini of the protein and, consistent with ATS-25 data, does not show further compaction. The existence and relevance of these two proposed conformation *in vivo*, however is questionable and needs to be considered carefully.

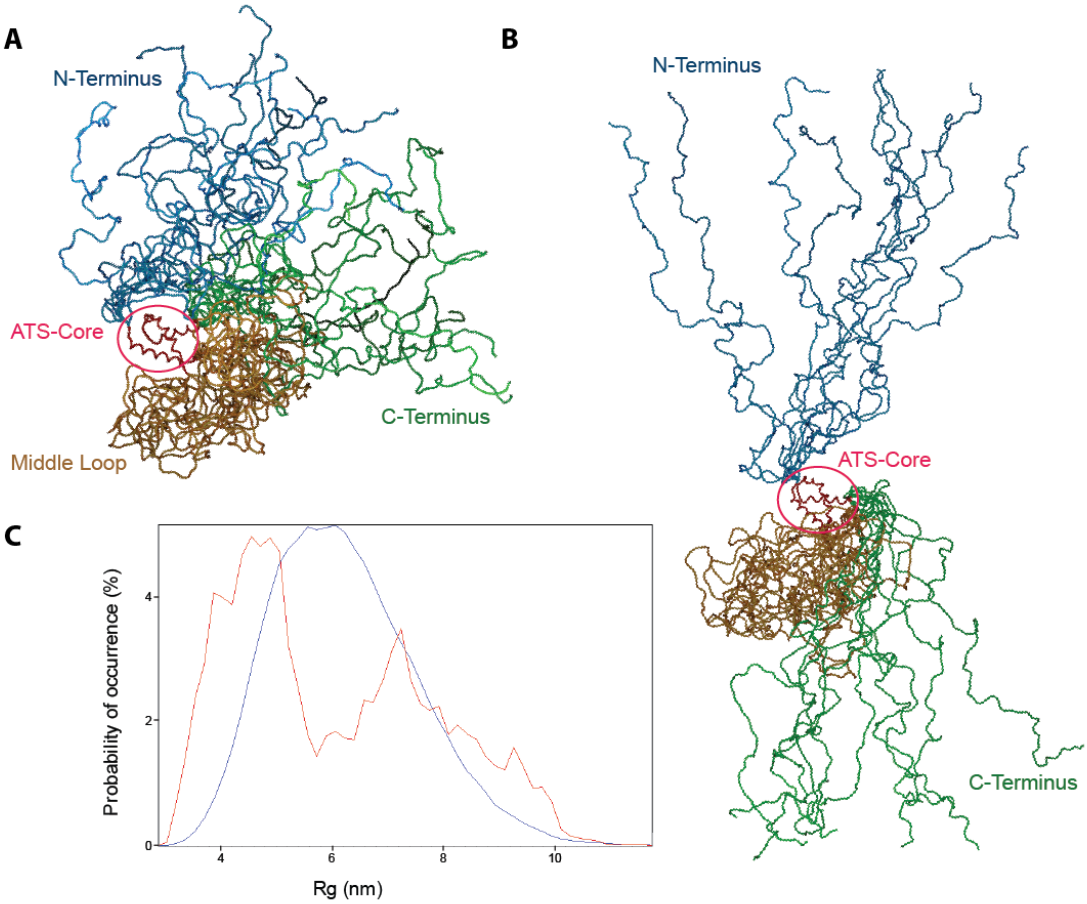


Figure 5.6: Representative members from the ATS-FL structure ensembles calculated from NMR restraints and based on ATS-core. The structured ATS-core is highlighted by a red circle. The central flexible segment is shown in brown and the construct's N- and C-termini are shown in blue and green, respectively. **B)** Molecular ensemble of ATS-FL without additional restraints and **A)** with maximum R_g restraint of 4.8 nm. **C)** The R_g probability distribution of the initial pool of ATS-FL ensembles (in blue) and after EOM analysis (red).

In summary, NMR- based structure determination and SAXS- derived structures for both ATS-core and ATS-25 are in agreement. Therefore, we concluded that the described molecular architecture based on experimental data represents the protein conformation under solution conditions. However, studies on the structure of ATS have proved to be a challenge due to the conformational flexibility of the protein. Since, they cannot be crystallized and instead need to be described in molecular ensemble terms. It is likely that the flexible nature of ATS is largely responsible for the little information available to date for this system.

5.5 References

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc *129*, 5656-5664.

Cavanagh, J., Fairbrother, W., Palmer A. G., Skelton N. J (2007). Protein NMR Spectroscopy - Principles and Practise (Academic Press).

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of biomolecular NMR *13*, 289-302.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Crystallogr *42*, 342-346.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR *8*, 477-486.

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., and Svergun, D.I. (2007). ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. J Appl Crystallogr *40*, S223-S228.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003). The Xplor-NIH NMR molecular structure determination package. J Magn Reson *160*, 65-73.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophysical journal *76*, 2879-2886.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophysical journal *80*, 2946-2953.

CHAPTER SIX

THE ATS - KAHRP INTERACTION

6.1 Introduction

Cytoadherence is mediated by various extracellular domains of *Pf*EMP1, whereas the intracellular domain must be involved in anchoring the protein to the host cell's cytoskeleton. How does ATS anchor the extracellular receptor attachments to the cytoskeleton? A direct mechanical link to the cytoskeleton is crucial for effective adhesion. Several studies suggest that KAHRP is vital for this connection (Oh et al., 2000; Waller et al., 1999; Waller et al., 2002). The major component of knobs is KAHRP, it assembles on the intracellular side of the erythrocyte membrane and is required for the formation of these electron dense structures (Crabb et al., 1997). KAHRP is considered to be an essential scaffolding protein based on knockout (Crabb et al., 1997; Rug et al., 2006) and *in vitro* pull-down experiments (Waller et al., 1999). Previous reports argue that KAHRP interacts with different host cytoskeletal molecules including spectrin, actin, and spectrin-actin band 4.1 complexes (Kilejian et al., 1991; Oh et al., 2000; Pei et al., 2005). Furthermore, two distinct regions of KAHRP have been reported to bind independently to ATS (Waller et al., 1999). The interaction sites on both proteins have been determined and involve a histidine-rich region (named K1A) and lysine-rich region (K2A) on the N- and C-terminus of KAHRP, respectively. The binding properties of ATS to these two KAHRP elements have been analyzed by pull down and surface-attachment experiments, however, *in vivo* or studies in solution have not yet been conducted.

This chapter presents interaction studies of ATS with KAHRP conducted using NMR and fluorescence anisotropy. The binding characteristics of all KAHRP fragments (K1, K2 and K3) as well as two smaller subfragments (K1A and K2A) were tested using these biophysical methods. The fragments K1, K2 and K3 were kindly provided by Ms Leanne Slater.

6.2 KAHRP - primary sequence analysis

KAHRP has two distinct stretches, an N-terminal histidine-rich (residues 60 - 117) and a C-terminal lysine-rich region (351 - 654). The first one is comprised of 53 % histidines, mostly occurring as histidine clusters. 24 % of the longer, lysine-rich regions consist of lysine residues making it positively charged. The SEG algorithm, a program that recognizes regions with low compositional complexity, detected several low complexity regions for these sequences (Wootton and Federhen, 1996). Statistical analysis revealed that low sequence complexity is a common characteristic of intrinsically disordered proteins (IDPs) (Romero et al., 2001). In addition, natively disordered proteins have been shown to contain a higher percentage of 'disorder-promoting' residues (A, R, G, Q, S, P, E, K); and to be significantly depleted in 'order-promoting' amino acids (W, C, F, I, Y, V, L N) (Dunker et al., 2001). As shown in table 6.1, only 21.2 % of the KAHRP sequence is composed of order-promoting residues, compared to an average of 36 % for globular proteins. The fraction of disorder-promoting amino acids is disproportionally high for KAHRP compared to globular proteins (59.9 % versus 47 %). These results suggest that the majority of KAHRP may be disordered.

	order-promoting residues (%)	disorder-promoting residues (%)
KAHRP (FL)	21.2	59.7
KAHRP (38 - 654)	20.3	61.2
KAHRP (38 - 350)	27.2	53.3
KAHRP (351 - 654)	10.8	69.1
Globular proteins	36.2	47.4

Table 6.1: Fractions in percent of order- (W, C, F, I, Y, V, L N) and disorder-promoting (A, R, G, Q, S, P, E, K) residues for KAHRP and subfragments thereof. KAHRP (FL) comprises residues M1 - G654. Reference values for globular proteins are copied from (Tompa, 2010).

Furthermore, disorder prediction algorithms attribute long disordered stretches in KAHRP. The metaPrDOS algorithm, which generates predictions based on a consensus principal between several predictors PrDOS, DISOPRED2, DisEMBL, DISPROT, DISPRO and IUPred, predicts only approximately 150 residues to be ordered and the rest to be disordered (Ishida and Kinoshita, 2008). In particular, the C-terminal half of KAHRP is expected to be completely flexible with predicted disorder probability of around 90% for the last 300 residues (figure 6.1B).

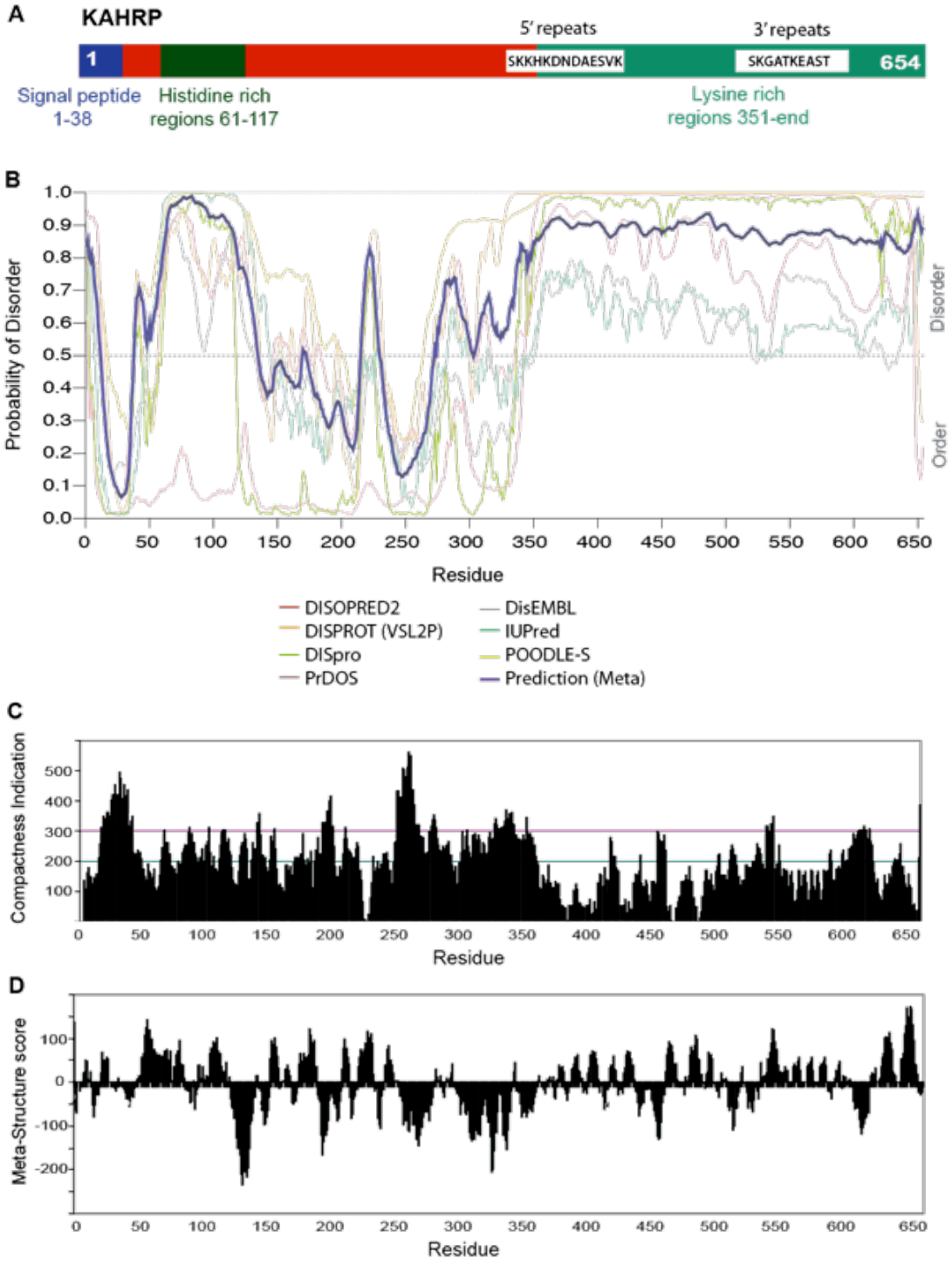


Figure 6.1: Primary structure analysis of KAHRP. A) Schematic representation of full - length KAHRP from *P. falciparum* 3D7 strain. Specific regions of the protein are highlighted in the indicated colours. **B)** Results for KAHRP sequence from different structure prediction algorithms. Values above the 0.5 threshold are predicted to be disordered (Ishida and Kinoshita, 2008). **C)** Meta structure analysis (Konrat, 2009) – compactness indication per residue. Globular folded proteins have compactness values above 300 (magenta line). Small values predict exposure to solvent, such as residues located on the surface or on flexible

loops. The green line represents the upper threshold for disordered proteins. **D)** Meta-structure secondary structure analysis of the KAHRP sequence (Konrat, 2009). Negative values correspond to β-strands or extended coil and positive values predict α-helical structure.

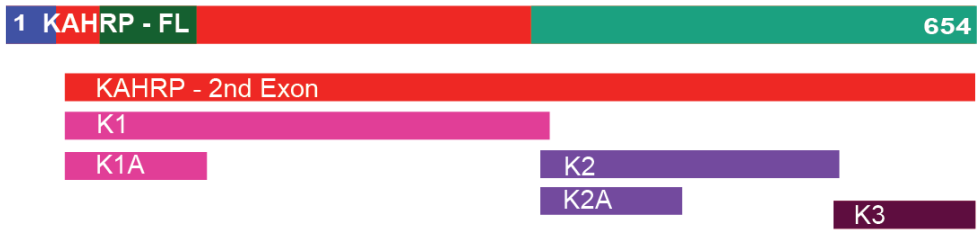
Moreover, a protein meta-structural analysis of KAHRP was performed to study the compactness of the protein (Konrat, 2009). Figure 6.1C shows the average residue compactness for KAHRP; only two short elements have high values (>300), found for residues in the buried core of globular proteins. The N-terminal region (residues 1 - 360) mainly lies between the values calculated for proteins in the RCSB and intrinsically unstructured proteins, whereas for the C-terminal part (residues 360 - end) overall C_i values are lower, indicating that the C-terminus is largely disordered.

In short, the computational sequence analysis of KAHRP suggests that this protein has a high probability to be intrinsically disordered, particularly the C-terminal half of the molecule.

6.3 Protein expression and purification of KAHRP

A clone containing the DNA sequence for full-length KAHRP from *P. falciparum* 3D7 (named MRA-6) supplied by the MR4 (ATCC, USA) was used as a template for PCR to create new expression constructs. Figure 6.2 gives an overview of all the designed and expressed constructs of KAHRP.

A



B

Construct name	Residues	Amino acid length	Predicted mass (kDa)	pI
KAHRP 2E	38 - 654	617	66.9	8.91
K1	38 - 362	325	36.2	6.41
K2	356 - 539	184	20.1	9.59
K3	533 - 654	122	12.1	9.16
K1A	38 - 126	89	10.6	7.29
K2A	356 - 447	92	10.6	9.65

Figure 6.2: Overview of KAHRP fragments used for interaction studies. A) KAHRP fragments are shown with respect to their location in the full-length protein. **B)** Protein characteristics of the relevant constructs. Molecular weight and theoretical pI numbers were calculated by Protparm (Gasteiger E., 2005). Similar construct boundaries have been described previously (Waller et al., 1999).

6.3.1 Expression of KAHRP - full length

As mentioned before, KAHRP is encoded by two exons, of which the first one encodes a hydrophobic core (Wickham et al., 2001) and the PEXEL motif (Marti et al., 2004). Therefore, the first exon is presumed to be cleaved off from the mature protein. For that reason, we aimed to produce the entire second exon of KAHRP (KAHRP 2E), as this part of the protein is most likely to be present at erythrocyte membrane knobs. At first, we started our expression tests using various *E. coli* strains: *E. coli* BL21(DE3), *E. coli* BL21(DE3) Codon+RIPL, *E. coli* CM41(DE3) Rosetta-gami BL21(DE3) cells. However, we were not able to produce this construct in usable amounts using bacterial overexpression. Hence, we chose alternative eukaryotic expression systems. We cloned the relevant KAHRP sequence in the pHLsec expression vector (Aricescu et al., 2006) for small-scale expression tests in mammalian HEK 293T cells. Unfortunately, the attempt using mammalian cells for expression were not successful either.

Finally, we prepared a pEU vector containing the KAHRP E2 sequence for *in vitro* expression using the cell free wheat germ system ENDEXT (Endo and Sawasaki, 2007). As mentioned before (section 3.3.2), this system has shown successful production of multiple *Plasmodium* proteins (Tsuboi et al., 2008) and allows for isotope labelling for NMR experiments. We hoped that an *in vitro* environment would enable us to obtain at least KAHRP E2 protein expression in the nM range. We performed a small-scale bilayer translation test at 16 °C for 20 hours. However, we could not detect any expression of KAHRP on a western blot (data not shown). Given these negative results, we were forced to experimentally analyse KAHRP and the ATS-KAHRP interaction solely using fragments of the former.

6.3.2 Expression of KAHRP fragments in *E. coli*

The different fragments of KAHRP (figure 6.2) were cloned in a pGEX-6G-2 vector, containing an N-terminal GST-tag (GE LifeSciences). All constructs were transformed into *E. coli* strain BL21(DE3) for protein expression in LB medium or minimal media for NMR usage. Cells were harvested by centrifugation and resuspended in PBS buffer pH 7.4 containing protease inhibitor cocktail tablets (Roche). Protein purification was performed using glutathione-sepharose resin. After the GST-tag cleavage by 3C-protesase a desalting and a size exclusion step followed. The GST was separated from the fragments by an additional glutathione-sepharose purification. The quality and purity of the different fragments was assessed by SDS-PAGE. Detailed expression and purification protocols can be found in chapter II.

We started to express and purify K1A and K2A, which have previously been described to bind to ATS (Waller et al., 1999). We were able to produce sufficient amounts of ^{15}N labelled and unlabelled K2A for NMR and other biophysical binding studies. Despite the fact that K1A showed lower expression yields, we were also able to obtain enough material for our experiments. In addition, we prepared the fragments K1, K2 and K3. However, K1 turned out to be very sensitive to degradation and we were not able to prepare solutions with higher concentration than approximately 50 μM for this fragment.

6.4 Interaction studies using NMR

6.4.1 K1A and K2A Interaction with ATS

The interaction of ATS (PF08_0141) with KAHRP has been studied by *in vitro* pull down experiments (Oh et al., 2000; Waller et al., 1999; Waller et al., 2002). The interaction has been mapped to two regions on KAHRP and to the centre of the ATS sequence. The K1A sequence of KAHRP has been reported to bind to residues 215 - 279 of ATS and K2A to residues 118 -214 (Waller et al., 2002). Our construct ATS-25 comprises both these sequences that are supposed to bind to the respective KAHRP fragments. Therefore, we decided to test the binding with ^{15}N-enriched ATS-25 and unlabelled K1A and K2A using NMR. Firstly, ^{1}H- ^{15}N HSQC spectra were recorded of ^{15}N-labelled ATS-25 alone and in the presence of varying amounts of K1A or K2A. Unexpectedly, we could not observe evidence for any binding in solution. Figure 6.4A shows the overlaid NMR spectra of ATS-25 with and without K2A. Similar results were obtained from titration studies involving K1A (figure 6.4B). The spectra do not display any chemical shift changes, neither changes in signal intensity nor line width upon addition of the KAHRP fragments. We repeated the NMR titration experiments using ^{15}N-enriched ATS-core instead of ATS-25. However, we could not observe any interaction of ATS-core with K1A or K2A under the tested conditions (figure 6.5).

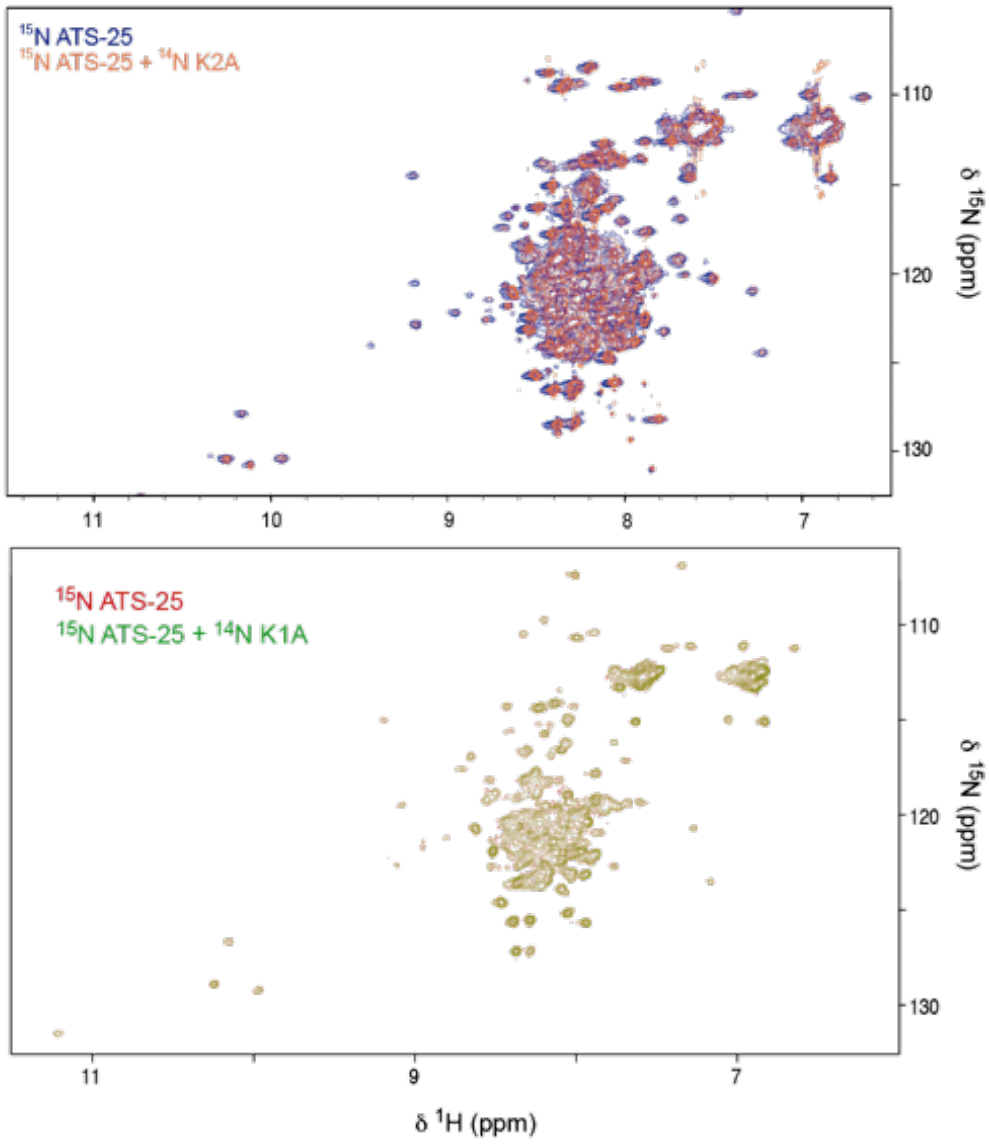


Figure 6.4: NMR titration studies of ATS-25 with K2A and K1A. A) Overlay of ^{1}H- ^{15}N HSQC spectra at of ATS-25 (120 µM) and ATS-25 in the presence of K2A (500 µM) in blue and orange, respectively. **B)** Similar overlay of spectra for 100 µM ATS-25 (red) and ATS-25 in the presence of 100 µM K1A (green). The experiments were recorded in NMR buffer at 25 °C.

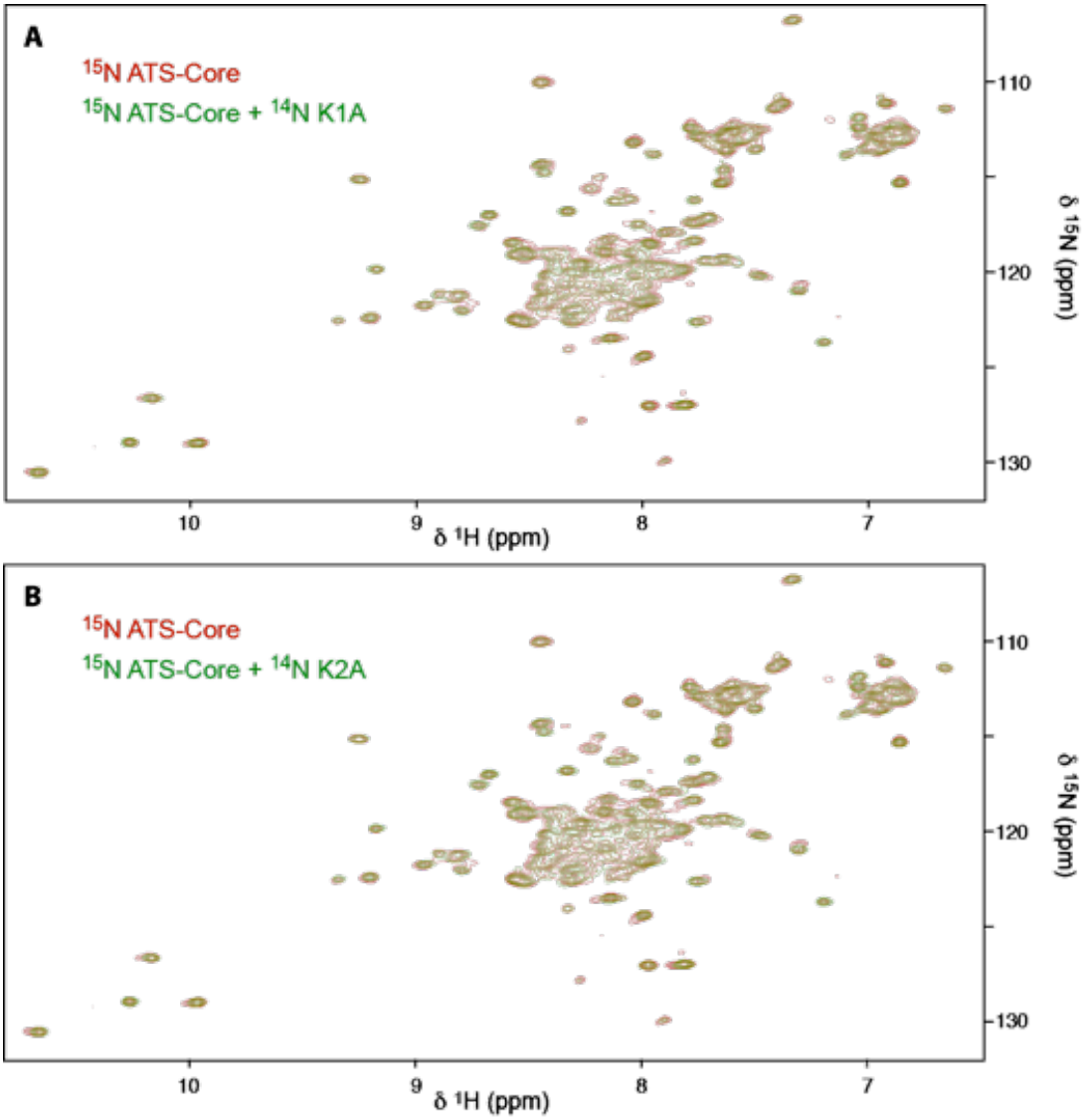


Figure 6.5: NMR titration experiments with ATS-core and K1A and K2A. A) ^{1}H- ^{15}N HSQC spectra of ATS-core (red) and of ATS-core in presence of K1A (green). **B)** Similar superposition of NMR spectra for ATS-core alone (red) or with K2A (green). Neither K1A nor K2A produce significant perturbations in the ATS-core spectra. Both tests were conducted in NMR buffer at 25 °C.

Since the majority of peaks of ATS-25 are overlapping, an exact analysis of resonance peak shifts is unattainable. To address whether spectral changes in ATS-25 are concealed by the lack of chemical shift dispersion of this construct, we isotope labelled the small KAHRP subfragments K1A and K2A. In the NMR spectra of these polypeptides many resonances can be resolved (figure 6.6). However, again, titrations with unlabelled ATS-FL to a concentration of 500 µM did not show significant perturbations.

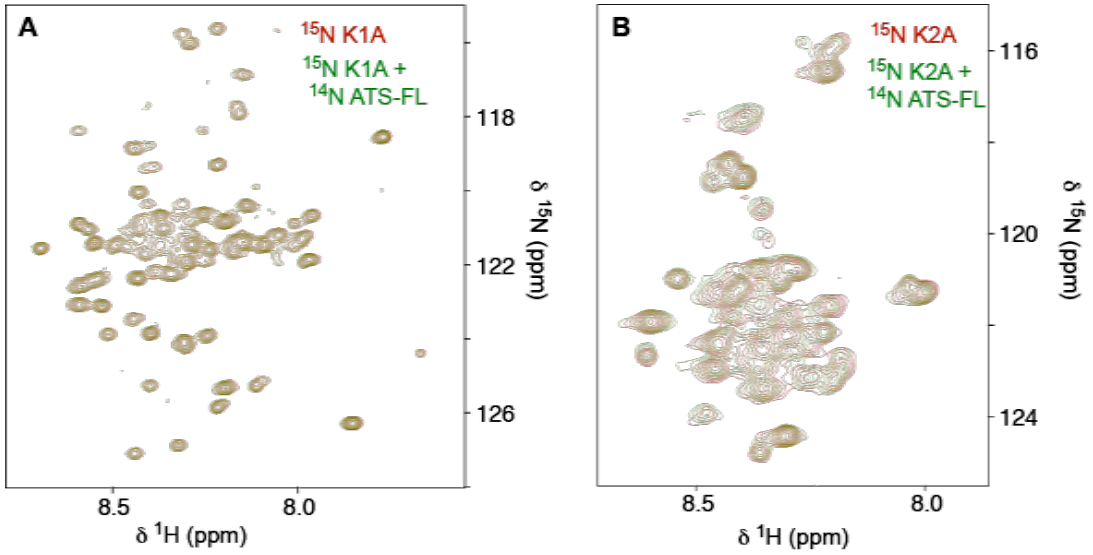


Figure 6.6: Superposition of ^{1}H- ^{15}N HSQC spectra of KAHRP fragments K1A and K2A without and with unlabelled ATS-FL. Experiments were recorded with protein concentrations of 50 µM for K1A/ K2A and 500 µM for ATS-FL at 25 °C. Spectra of the KAHRP fragments alone and in the presence of ATS-FL are coloured in red and green, respectively.

6.4.2 Studies on longer protein constructs of KAHRP

Given the surprising results of the studies using the small fragments K1A and K2A, we decided to take a broader approach to study this reported interaction. We wanted to determine whether segments outside the previously determined interaction domain boundaries (K1A and K2A) affect the interaction (Waller et al., 1999). Therefore, we prepared larger fragments of KAHRP - K1, K2 and K3. Together these three molecules encode for the entire sequence of KAHRP (second exon). We tested them separately for interaction with ^{15}N-labelled ATS-FL, acquiring a series of ^{1}H- ^{15}N HSQC spectra with and without the assumed unlabelled binding partner (K1, K2 and K3). Once more, however, we did not detect any sign for interactions in the NMR spectra under physiological conditions (figure 6.7). Unfortunately, we were not able to conduct studies with higher concentrations of K1, as this constructs shows limited solubility. NMR interaction studies involving full-length constructs for both ATS and KAHRP were not feasible, as the successful expression of KAHRP (second exon) has not yet been possible.

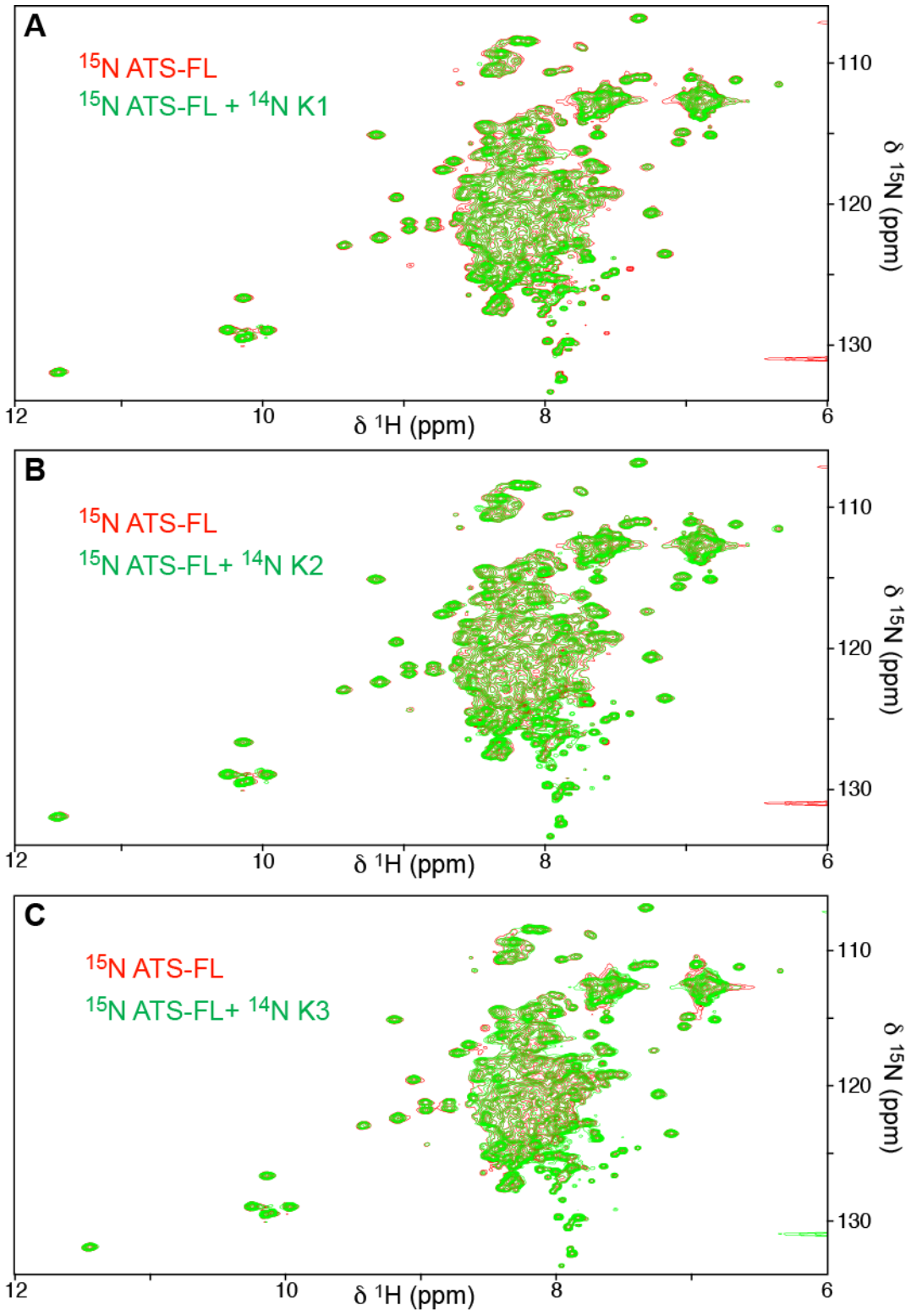


Figure 6.7: NMR binding studies of ATS-FL and longer KAHRP segments. Overlay of ^{1}H- ^{15}N HSQC spectra of ^{15}N enriched ATS-FL alone and in the presence of 50 µM K1 (A), 300 µM K2 (B) and 300 µM K3. All spectra were recorded at 25 °C in NMR buffer at pH 7.0. The limited solubility of K1 restricted us to its low concentration in this study, compared to K2 and K3.

6.5 Interaction studies using fluorescence anisotropy

Since, our NMR interaction studies could not detect any interaction of ATS with the described KAHRP fragments, we decided to examine the binding by an alternative technique - fluorescence anisotropy. This method measures changes of molecular tumbling in solution, caused by the binding of molecules. The fluorophores are selectively excited by polarized light and their emitted light is polarized as well. However, the polarization of emitted light is dependent on the extent of rotational diffusion that occurs between excitation and emission. The binding of a fluorescent labelled protein to another slows down tumbling in solution and, consequently, leads to an increase in fluorescence anisotropy. The method can be used to determine binding constants.

Firstly, we had to introduce a fluorescent label to one of the proteins of interest. ATS-FL only contains one cysteine amino acid C272, yet, this residue is buried in the core structure, therefore, inaccessible. Consequently, we had to mutate one amino acid to cysteine; the substitution Q150C was performed using site-directed mutagenesis and the altered construct was expressed and purified as described before for the wild-type protein (section 2.8), however, without DTT in the final size exclusion chromatography. For the labelling of the ATS-FL mutant we used fluorescein-5-maleimide (Invitrogen), unbound fluorescent label was removed by an additional gel filtration run. The efficiency of the labelling process was verified by absorption measurements at λ = 495 nm, where the fluorescein absorbs. Our labelling strategy was successful and we obtained fluorescent-labelled ATS-FL at a stock concentration of approximately 15 μM. All titration experiments of this ATS-FL variant and

various KAHRP fragments were performed in 96 well plates at 25 °C in NMR buffer and changes of anisotropy were recorded.

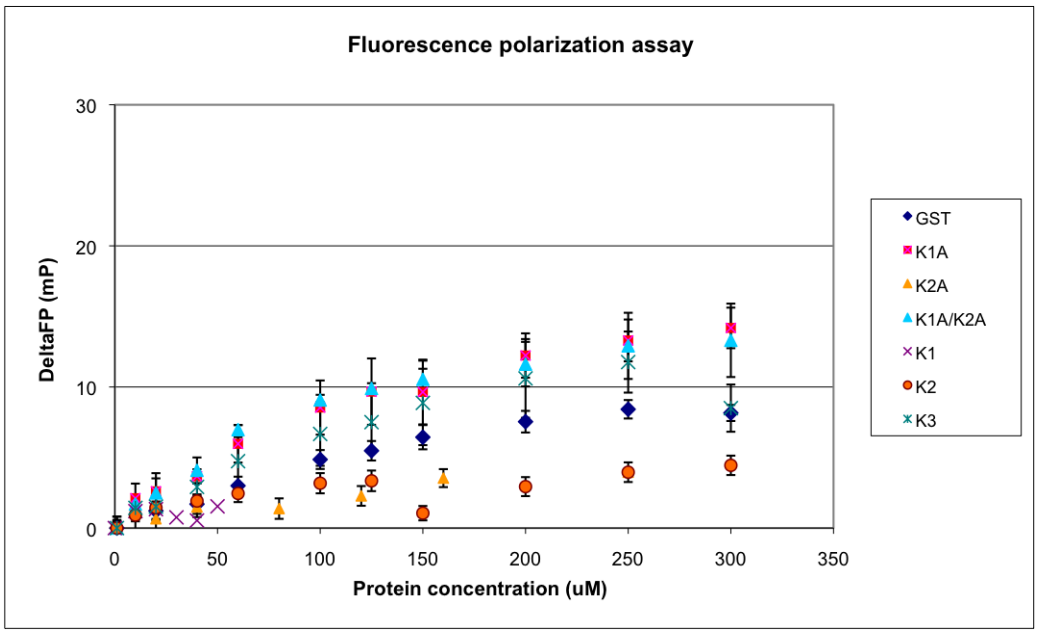


Figure 6. 8: The interactions of labelled ATS-FL with KAHRP fragments or GST. Each data point represents the average of five repetitive measurements, indicated by the error bars. Neither the KAHRP fragments nor GST caused strong anisotropy changes in concentrations up to 300 µM (50 µM for K1 due to limited solubility).

The results of the anisotropy studies confirmed our NMR binding studies: Only small changes of anisotropy were noted when testing all available KAHRP fragments (K1, K2, K3, K1A and K2A) for interaction with ATS-FL. The strongest change of anisotropy of 14 mP was recorded for K1A at a concentration of 300 µM. It is noteworthy that a control experiment using the same molar concentration of GST yielded an anisotropy change of approximately 8 mP (figure 6.8).

6.6 Phoshorylation studies of ATS

Before the current study, ATS phosphorylation by casein kinase II (CK II) has been suggested to strengthen the ATS-KAHRP interaction (Hora et al., 2009). The phosphorylation sites were mainly located within an acidic residue cluster at the N-terminus of ATS - the conserved residues T62, T65, S66, S67, S69 and T329 (Hora et al., 2009). Therefore, we aimed to test the effect of phosphorylation on the ATS N-terminal construct (ATS-N), comprising residues 2-108 of ATS-FL. ATS-N was expressed in *E.coli* and purified as described before (section 2.8). Subsequently, the purified protein was *in vitro* phosphorylated with the commercially purchased CK II (NEB), according section 2.10. Initally, we performed a time course study to confirm and to optimise the conditions of the enzymatic reaction and verified the phosphorylation by SDS- and native PAGE. Figure 6.9 shows successful phosphorylation of ATS-N constructs at 30 °C already after 1 hour. Since, the small-scale tests were successful we prepared two ^{15}N-enriched ATS-N samples, one of them was phosphorylated beforehand and then dialysed into NMR buffer. An overlay of the recorded spectra is shown in figure 6.10. The phosphorylation of ATS-N by casein kinase II caused perturbations in the HSQC spectrum. Due to the lack of chemical shift dispersion and poor NMR spectrum of phosphorylated ATS-N, the interpretation of the chemical shift perturbations was problematic. However, chemical shift changes for the suggested residues T62, T65, S66, S67 and S69 and also for T25, S52 and S54 were observed. The appearing resonance peaks are most likely originating from the modified residues. Protein phosphorylation typically results in large downfield chemical shift perturbations of backbone amide resonances of the affected residues. In addition, we performed NMR experiments to observe the ATS-N phosphorylation in the sample tube. Similar *in situ* observations of protein

phosphorylation have been conducted previously (Selenko et al., 2008). Briefly, the kinase reactions were performed with 100 µM of ^{15}N-enriched ATS-N substrates and 1500 units of CK II in the presence of 300 µM ATP and 100 µM $MgSO_4$ in phosphate buffer (20 mM, pH 7.0). Consecutive ^{1}H- ^{15}N HSQC spectra were acquired while the kinase reaction progressed. Using this method we observed similar chemical shift perturbations. We performed similar experiments for ATS-FL to test the effect of phophorylation on the interaction to KAHRP. The addition of the KAHRP fragment K2 (200 µM) to phosphorylated ATS-FL did not indicate any changes in the spectrum of phosphorylated ATS-FL. These experiments need to be further verified, as the quality of the obtained spectra was poor and substantial protein degradation occurred.

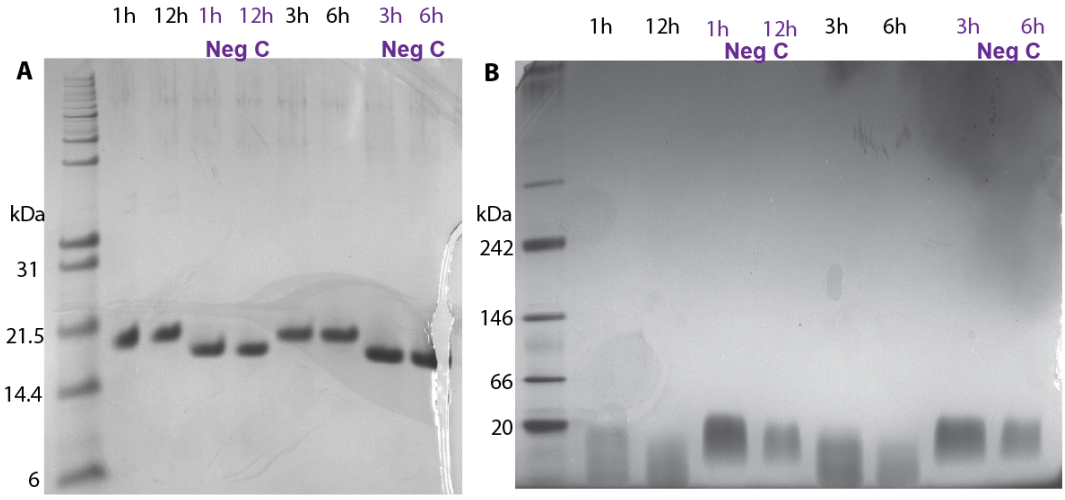


Figure 6.9: *In vitro* phosphorylation of ATS-N observed by SDS- (A) and native PAGE (B). ATS-N was incubated with Casein kinase II (NEB) at 30 °C and samples were taken after 1, 3, 6 and 12 hours and analysed SDS- and native PAGE. Negative controls (neg C) were treated the same way without the addition of enzyme.

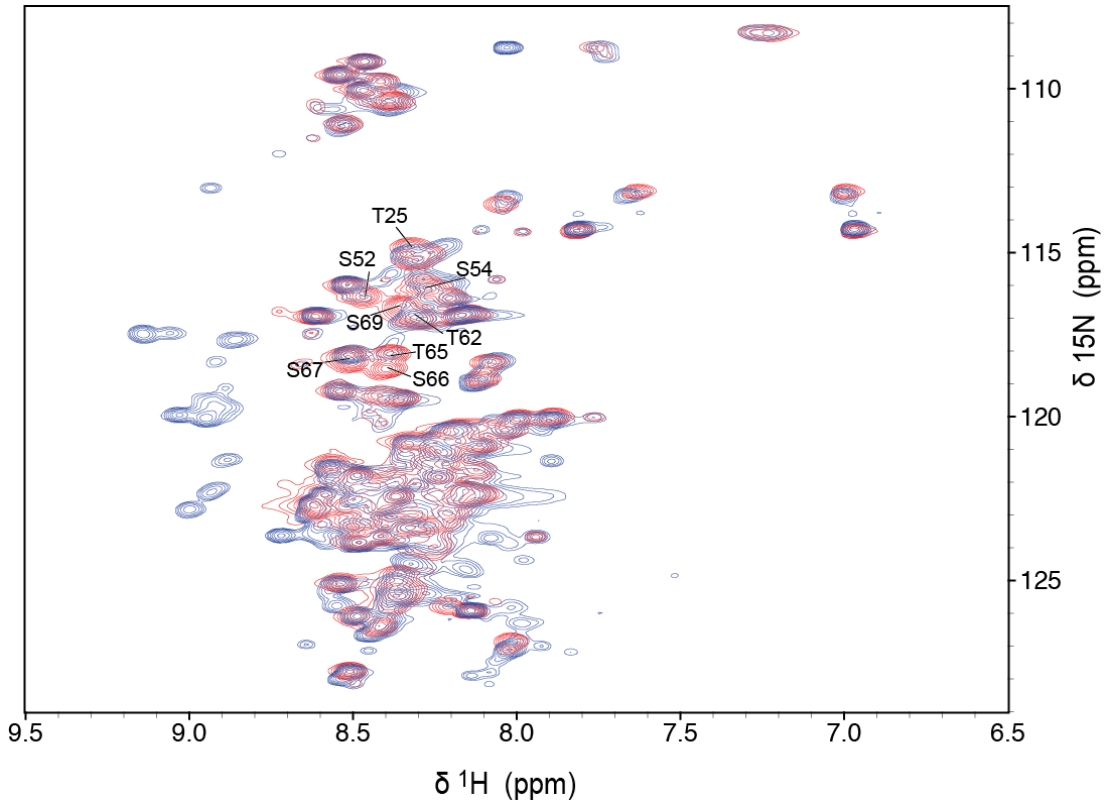


Figure 6.10: Superposition of HSQC spectra of ATS-N and phosphorylated ATS-N. ^{1}H-^{15}N HSQC spectra before (red) and after (blue) the phosphorylation reaction. Both spectra were recorded in NMR buffer (50mM NaCl, 20 mM Na_2HPO_4 pH 7.0, 1 mM DTT) at 20 °C at 500 MHz. Affected serine and tyrosine residues are labelled in the spectrum.

In short, both our NMR investigations and fluorescence anisotropy data, collected so far, do not support the existence of significant ATS-KAHRP interaction in solution. However, we cannot exclude the possibility of a very weak interaction with dissociation constants (K_D) in the mM range, which would be at least 1000 times weaker than previously reported (Oh et al., 2000; Waller et al., 1999; Waller et al., 2002). Moreover, the effects of phosphorylation of ATS and other potential post-translational modifications on ATS/ KAHRP still remain to be investigated further. Another possibility is that the ATS-KAHRP interaction is mediated or strengthened by one or more other proteins in the cell.

6.7 References

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta crystallographica Section D, Biological crystallography *62*, 1243-1250.

Crabb, B.S., Cooke, B.M., Reeder, J.C., Waller, R.F., Caruana, S.R., Davern, K.M., Wickham, M.E., Brown, G.V., Coppel, R.L., and Cowman, A.F. (1997). Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell *89*, 287-296.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., *et al*. (2001). Intrinsically disordered protein. Journal of molecular graphics & modelling *19*, 26-59.

Endo, Y., and Sawasaki, T. (2007). [Wheat germ cell-free protein synthesis]. Seikagaku The Journal of Japanese Biochemical Society *79*, 229-238.

Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, Humana Press (2005) (Humana Press), pp. 571 - 607.

Hora, R., Bridges, D.J., Craig, A., and Sharma, A. (2009). Erythrocytic casein kinase II regulates cytoadherence of Plasmodium falciparum-infected red blood cells. The Journal of biological chemistry *284*, 6260-6269.

Ishida, T., and Kinoshita, K. (2008). Prediction of disordered regions in proteins based on the meta approach. Bioinformatics *24*, 1344-1348.

Kilejian, A., Rashid, M.A., Aikawa, M., Aji, T., and Yang, Y.F. (1991). Selective association of a fragment of the knob protein with spectrin, actin and the red cell membrane. Mol Biochem Parasitol *44*, 175-181.

Konrat, R. (2009). The protein meta-structure: a novel concept for chemical and molecular biology. Cellular and molecular life sciences : CMLS *66*, 3625-3639.

Marti, M., Good, R.T., Rug, M., Knuepfer, E., and Cowman, A.F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science *306*, 1930-1933.

Oh, S.S., Voigt, S., Fisher, D., Yi, S.J., LeRoy, P.J., Derick, L.H., Liu, S., and Chishti, A.H. (2000). Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. Mol Biochem Parasitol *108*, 237-247.

Pei, X., An, X., Guo, X., Tarnawski, M., Coppel, R., and Mohandas, N. (2005). Structural and functional studies of interaction between Plasmodium falciparum knob-associated histidine-rich protein (KAHRP) and erythrocyte spectrin. The Journal of biological chemistry *280*, 31166-31171.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins *42*, 38-48.

Rug, M., Prescott, S.W., Fernandez, K.M., Cooke, B.M., and Cowman, A.F. (2006). The role of KAHRP domains in knob formation and cytoadherence of P falciparum-infected human erythrocytes. Blood *108*, 370-378.

Selenko, P., Frueh, D.P., Elsaesser, S.J., Haas, W., Gygi, S.P., and Wagner, G. (2008). In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. Nature structural & molecular biology *15*, 321-329.

Tompa, P. (2010). Structure and Function of Intrinsically Disordered Proteins (CRC Press).

Tsuboi, T., Takeo, S., Iriko, H., Jin, L., Tsuchimochi, M., Matsuda, S., Han, E.T., Otsuki, H., Kaneko, O., Sattabongkot, J., *et al*. (2008). Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. Infect Immun *76*, 1702-1708.

Waller, K.L., Cooke, B.M., Nunomura, W., Mohandas, N., and Coppel, R.L. (1999). 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 *274*, 23808-23813.

Waller, K.L., Nunomura, W., Cooke, B.M., Mohandas, N., and Coppel, R.L. (2002). Mapping the domains of the cytoadherence ligand Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) that bind to the knob-associated histidine-rich protein (KAHRP). Mol Biochem Parasitol *119*, 125-129.

Wickham, M.E., Rug, M., Ralph, S.A., Klonis, N., McFadden, G.I., Tilley, L., and Cowman, A.F. (2001). Trafficking and assembly of the cytoadherence complex in Plasmodium falciparum-infected human erythrocytes. The EMBO journal *20*, 5636-5649.

Wootton, J.C., and Federhen, S. (1996). Analysis of compositionally biased regions in sequence databases. Methods in enzymology *266*, 554-571.

CHAPTER SEVEN

CONSERVED INTERACTION EPITOPE ON ATS

7.1 Introduction

Since we were not able to confirm the suggested interaction of ATS with KAHRP (Waller et al., 1999; Waller et al., 2002), we attempted to look for alternative binding partners. We assumed that other proteins must be involved, acting as connectors to the host cytoskeleton to ensure functional anchoring of the knob structures. This chapter describes the analysis of the yeast -two- hybrid (Y2H) interactome of *P. falciparum* (LaCount et al., 2005) to find previously unidentified binding partners of ATS. In addition, we employed a novel bioinformatics analysis tool, courtesy of Dr Robert Konrat (University of Oxford), to identify protein interactions based on their primary sequences. Furthermore, we conducted experimental biophysical studies to verify one of the proposed interactions. The production of the PHIST domain of PFI1780w and the NMR experiments presented in this chapter were kindly performed by Ms Leanne Slater and Dr Ioannis Vakonakis (University of Oxford).

7.2 Analysis of Y2H data

We datamined the previously published Y2H interactome of *P. falciparum* (LaCount et al., 2005*),* and identified twelve proteins as interacting with one or more ATS variants of *P. falciparum*. Table 7.1 gives an overview of the proteins, many of which are conserved, but with unknown function. Interestingly, they are all annotated as binding to the central flexible loop of ATS or at least parts of it. Furthermore, bioinformatics analysis, based on the primary sequence, suggests a high interaction probability between the central loop of ATS and other partners. This novel method was developed by Dr Robert Konrat (University of Vienna) to assess the interaction probability of a given sequence. A detailed description of the method, based on a meta structure analysis (Konrat, 2009) was given in section 2.13.1. Briefly, a quantitative prediction of secondary structure for the protein of interest is aligned with similar predictions performed for known protein complexes in the RCSB (Konrat, 2009). The goodness of fit with residues at the complex interaction interface is then expressed as the protein interaction index, PII. PII values greater than 1000 are considered to be highly significant (Konrat, unpublished statistical analysis). Figure 7.2 shows the PII values for ATS (PF08_0141). The graph clearly displays the high PII signal in the central, unfolded segment of the protein, which strongly advocates an interaction site. The anticipated interaction segment comprises residues 161-186 of ATS. Sequence alignment of the ATS variants found in *P. falciparum* 3D7 reveals that this 26 residue long region is highly conserved (figure 7.3). Over half the residues in this part of ATS (14 out of 26) are identical in 95 % of ATS domains, while 25 out of 26 residues are identical or conservatively substituted in all but one of the 59 variants in the *P. falciparum* 3D7 isolate. Therefore, we concluded that the

ATS family has a conserved interaction epitope, located in the middle flexible loop. Interestingly, one of the mapped ATS-KAHRP binding sites proposed by Waller et al. includes this conserved interaction epitope. However, the second interaction site of ATS-KAHRP (residues 215-279 of ATS) does not (Waller et al., 2002).

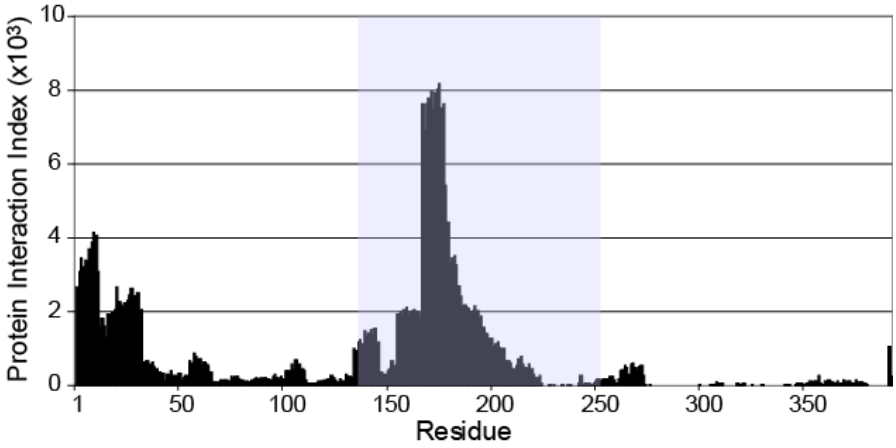


Figure 7.2: Protein interaction index for ATS (PF08_0141). Per residue protein interaction index (PII). The central flexible loop is highlighted in violet.

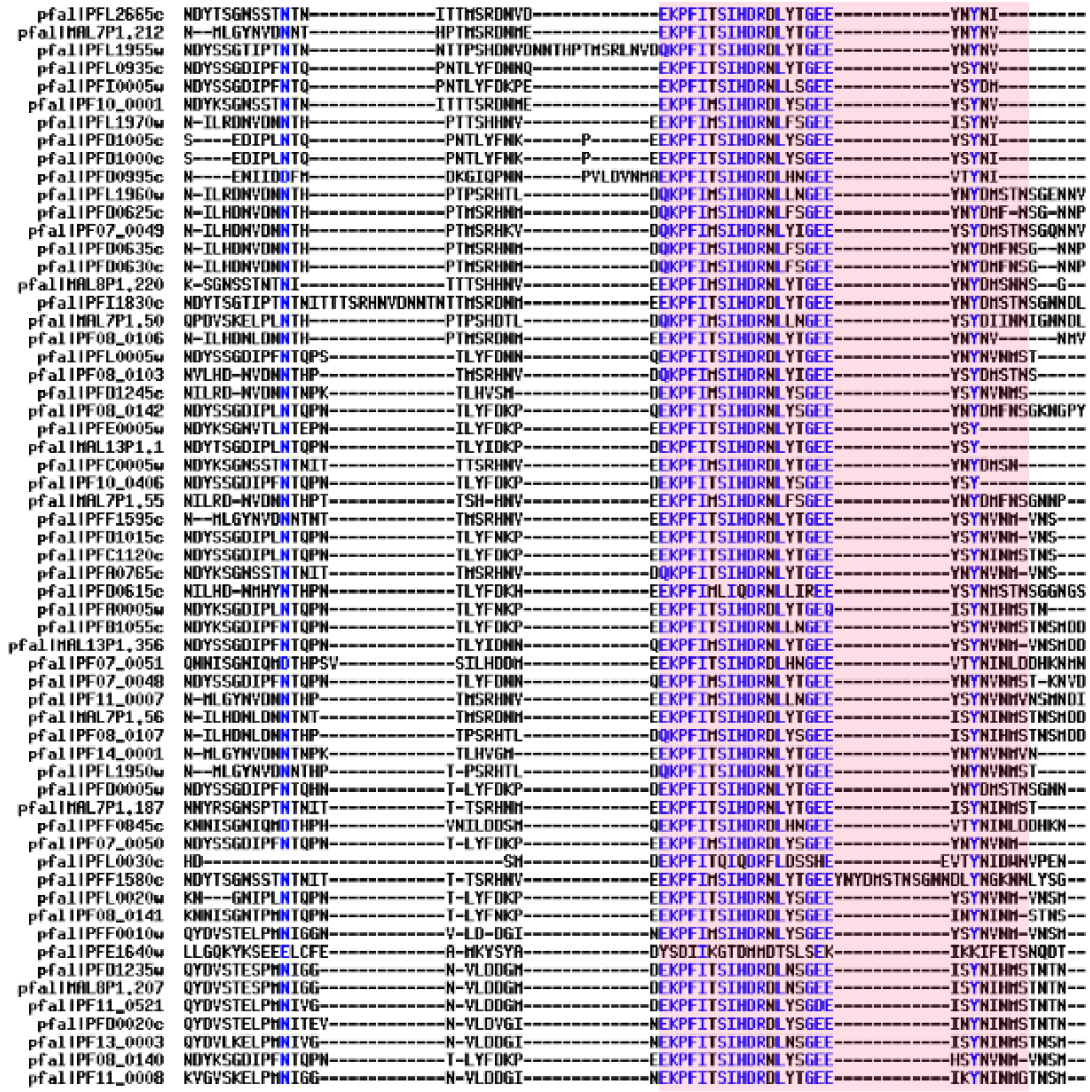


Figure 7.3: Conserved interaction epitope of the ATS family. Multiple sequence alignment (ClustalW) of ATS variants from the 3D7 strain *of P. falciparum*. The accession codes for the different proteins are given, according to plasmodb (Bahl et al., 2002). Only residues above a 50 % conservation are coloured. The conserved ATS region at the protein centre that contains the predicted interaction epitope is highlighted in red.

plasmodb ID	MW (kDa)	remarkable features, other information	putative biological function	localisation	SS prediction of interaction site
PFB0340c	111	serine repeat antigen 5	proteolysis	PV/ iRBC membrane	disordered
PFL2095w	12.9	translation initiation factor	translation initiation	most probably not exported	α-helices, β strands
PFL1485w	71.3	conserved	unknown	unknown	α-helical
PFI0875w	72.4	heat shock protein 70	stress response, ATP binding	unknown	α-helical
PFA_0410w	225	-	unknown	unknown	disordered
PF13_0286	87.5	methyltransferase	RNA methylation	nucleus	α-helices, β strands
PF10_0213	26.7	conserved	unknown	unknown	β strands, extended
PF10_0346	42.2	merozoite surface antigen protein 6	entry into host	unknown	α-helices, extended
PF11_0381	15.5	subtililisin-like protease 2	proteolysis	membrane granule	α-helical
PFL1930w	69.1	conserved	unknown	unknown	α-helical
PFL2120w	15.5	conserved - 2 interaction sites detected	unknown	unknown	α-helical
PFI1780w	45	conserved	unknown	iRBC cytosol/ membrane	α-helical

Table 7.1: Proposed interaction partners of with ATS variants. Overview of *P. falciparum* encoded genes predicted to interact with ATS, based on yeast two hybrid studies (LaCount et al., 2005). Molecular weights were determined using Protparm (Gasteiger E., 2005). Information on the different genes was extracted from plasmodb (Bahl et al., 2002) and ApiLoc (ApiLoc, 2011). Protein secondary structure was predicted by PSIPRED (Jones, 1999). Nomenclature: iRBC, infected red blood cell; PV, parasitophorous vacuole; SS, secondary structure

7.3 ATS interaction with PFI1780w

To verify the interaction epitope of ATS experimentally, we decided to test the predicted interaction partners of ATS by NMR and fluorescence anisotropy. Further analysis of the proposed Y2H binding partners for ATS was performed (as listed in table 7.1). As ATS is active at the host cell membrane, we had to look for an interactor, which is also shuttled there. Hence, we selected proteins, which contain a signal peptide that would allow them to be transported to the same location as ATS. In addition, we classified the proteins according to their likelihood of being structured, since the interaction site on ATS is disordered. Out of the 12 proteins suggested by Y2H screen (LaCount et al., 2005) to interact with ATS, we chose to further investigate PFI1780w. This parasite protein of unknown function is encoded as a two-exon gene and contains a PEXEL (for *Plasmodium* export element) motif (Hiller et al., 2004; Marti et al., 2004). Therefore, we concluded that this protein is very likely to be present in the cytosol of the iRBC. Interestingly, the PFI1780w-GFP-fusion protein was shown to be exported to the host cell cytoplasm (Sargeant et al., 2006). PFI1780w has also been detected in detergent-resistant membrane fractions of infected erythrocytes (Sanders et al., 2005). Moreover, PFI1780w has another interesting feature: it contains a *Plasmodium* helical interspersed subtelomeric (PHIST) domain (Sargeant et al., 2006), a novel, unique domain only found in *Plasmodia*. Previous studies have shown the presence of PHIST domain containing proteins in membrane fractions of parasite-infected erythrocytes (Florens et al., 2004; Florens et al., 2002; Lasonder et al., 2002). The novel PHIST domain family is subclassified into 3 categories PHISTa, PHISTb and PHISTc, the latter includes PFI1780w (Sargeant et al., 2006). A secondary structure prediction for the ~

150 amino acid long PHISTc domain of PFI1780w (residues 89 - 260) suggests that this domain has a mainly α-helical character. The ATS interaction site on PFI1780w was mapped to residues 197 to 276 of the latter. A PFI1780w fragment containing the residues 80-280 was expressed in *E.coli* BL21(DE) cells and purified. Initial CD analysis proved the domain to be helical and cooperatively folded. NMR and fluorescence anisotropy studies testing the binding of the PHIST domain to ATS were performed using the ATS-FL PF08_0141 variant, which was also employed for interaction studies with KAHRP (Chapter IV).

7.3.1 Interaction studies using fluorescence anisotropy

In order to determine the binding of ATS to PFI1780w, we conducted fluorescent anisotropy measurements. The ATS-FL D150C construct was labelled with fluorescein as described earlier (section 2.9). The assay was performed in 96-well plates with a protein concentration of 1 μM for ATS-FL. The change in fluorescence anisotropy caused by the addition of increasing concentrations of PFI1780w is plotted in figure 7.3. Strong changes in anisotropy were observed upon the titration, indicating binding of the two molecules. The data were fitted to a single-site binding model, giving a calculated K_D value of 150 ± 6 μM for the observed binding of ATS-FL with the PHISTc domain of PFI1780w.

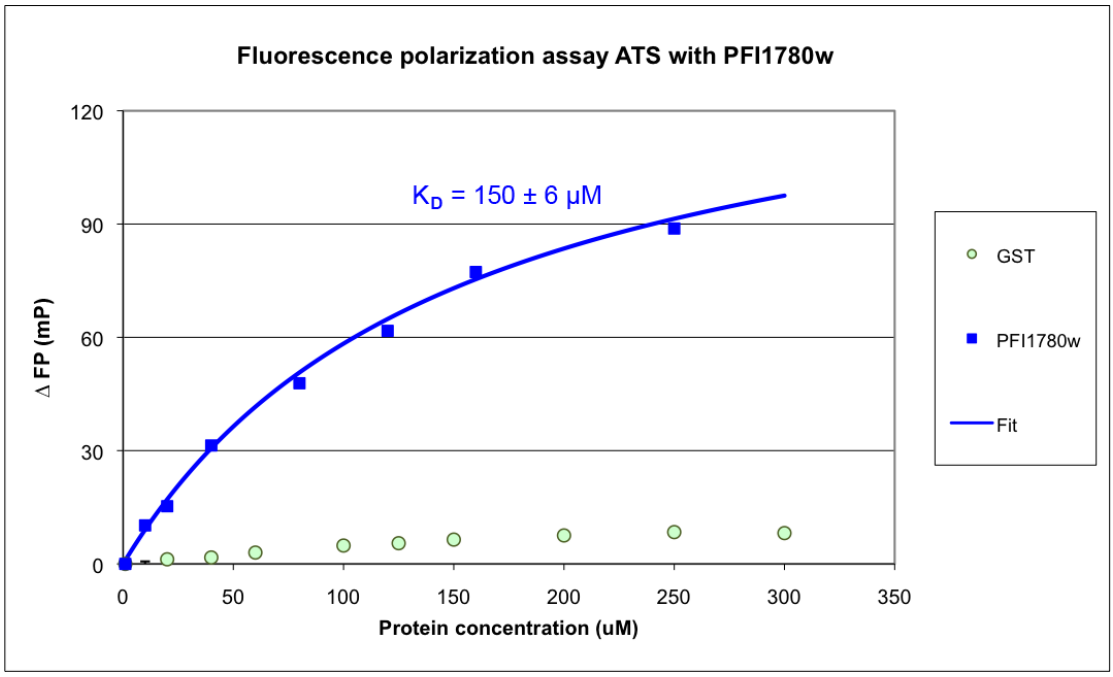


Figure 7.4: Fluorescence polarization assay to measure the interaction of ATS with PFI1870w. PFI1780w (residues 80-280) was added at the given protein concentrations to fluorescently labelled ATS-FL. Purified GST was used as a negative control. The blue line represents the global fit (determined as described in section 2.20). Measurements were conducted at 25 °C in NMR buffer (50mM NaCl, 20 mM Na_2HPO_4 pH 7.0, 1 mM DTT). For each study, five repetitive measurements were taken.

7.3.2 Interaction studies using NMR

^{1}H- ^{15}N HSQC titration studies of ^{15}N-enriched ATS-FL were performed with unlabelled PFI1780w. The experimental conditions were 50 µM ATS-FL in NMR buffer and an excess of 200 µM PFI1780w in NMR buffer (50 mM NaCl, 20 mM Na_2HPO_4 pH 7.0, 2 mM DTT). Interestingly, the addition of PFI1780w to ATS-FL induced substantial changes in the NMR spectra (figure 7.5A). A loss of peak intensity of many ATS-FL peaks was observed. Due to the flexible nature of ATS, producing a mostly overlapping HSQC spectrum, we were not able to pinpoint all

affected residues. However, three glycine residues belonging to the central flexible segment of ATS G203, G213 and G233 disappeared upon addition of PFI1780w, as well as several residues located in the structured core of ATS. This result is consistent with the fluorescence anisotropy interaction studies. Therefore, we were able to confirm the interaction of the two proteins suggested by Y2H. In order to confirm the importance of the flexible middle segment of ATS for this binding, as suggested by bioinformatics, we carried out NMR titration experiments with ATS-25 and ATS-core. ^{1}H- ^{15}N HSQC spectra of ^{15}N-enriched ATS-25 or ATS-core were recorded alone and in the presence of unlabelled PFI1780w. Figure 7.5B shows that the same resonance peaks disappear from the spectra of ATS-25 compared to ATS-FL. However, the addition of PFI1780w to the ATS-core construct did not result in any changes of the NMR spectrum. This result is not surprising, as the ATS-core only contains the structured element of the full-length protein and lacks the central loop. Given these observations, we could confirm the necessity of the flexible middle segment and the conserved interaction epitope of ATS for the interaction with PFI1780w. Detailed analysis of the observed binding is currently being conducted by Ms Leanne Slater.

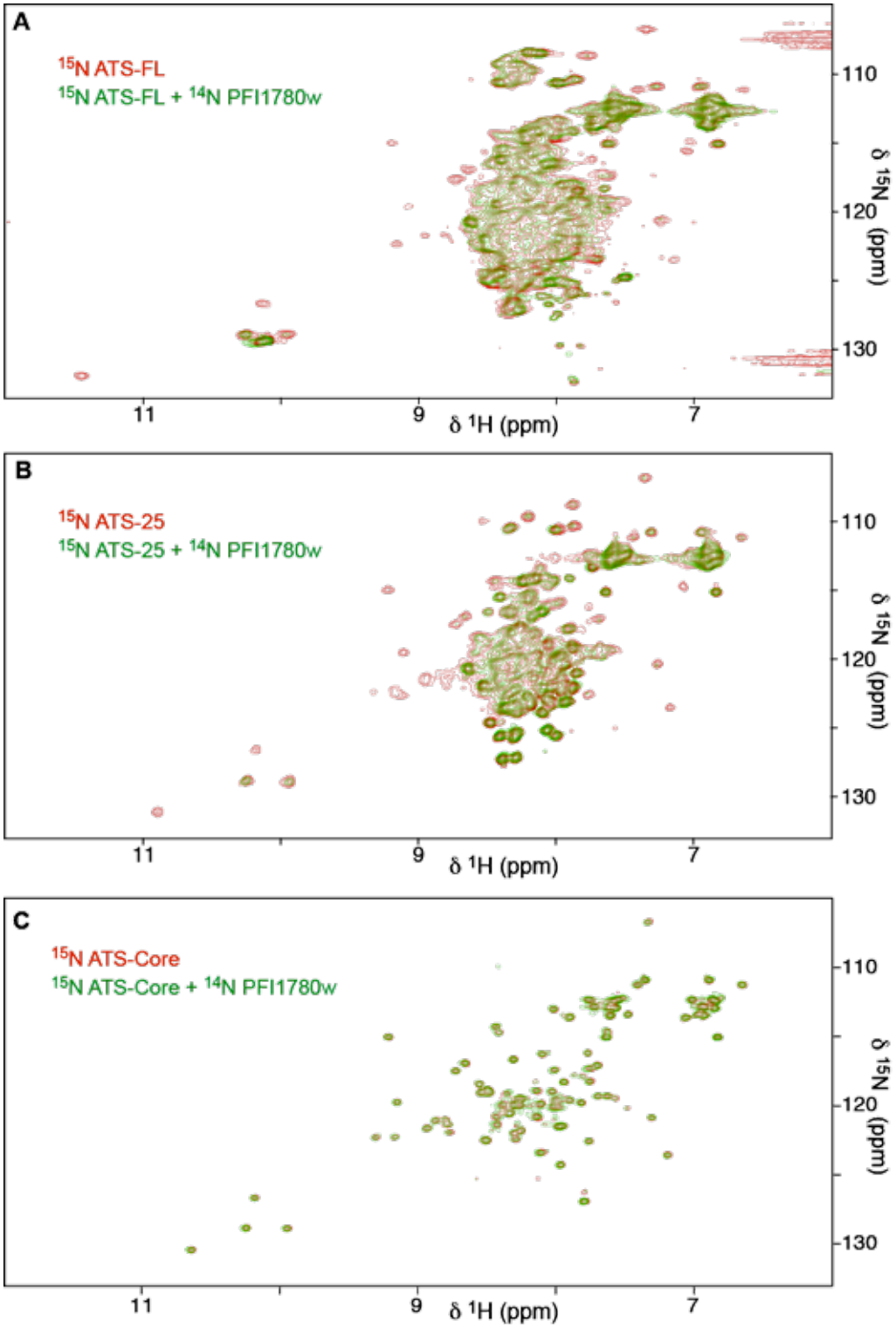


Figure 7.5: NMR binding studies of ATS constructs and PFI1780w segment. A) Overlay of ^{1}H- ^{15}N HSQC spectra of 50 µM ^{15}N enriched ATS-FL (red) alone and in the presence of 200 µM PFI1780w (green). **B)** Similar overlay of ^{15}N enriched ATS-25 alone and with 250 µM of PFI1780w. **C)** Similar overlay of ^{15}N enriched ATS-core alone and in the presence of 250 µM of PFI1780w. All spectra were recorded at 25 °C in NMR buffer at pH 7.0.

7.4 References

ApiLoc (2011). ApiLoc - A database of published protein sub-cellular localisation in Apicomplexa.

Bahl, A., Brunk, B., Coppel, R.L., Crabtree, J., Diskin, S.J., Fraunholz, M.J., Grant, G.R., Gupta, D., Huestis, R.L., Kissinger, J.C., *et al*. (2002). PlasmoDB: the Plasmodium genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished). Nucleic acids research *30*, 87-90.

Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D.J., Yates, J.R., 3rd, and Wub, Y. (2004). Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. Mol Biochem Parasitol *135*, 1-11.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., *et al*. (2002). A proteomic view of the Plasmodium falciparum life cycle. Nature *419*, 520-526.

Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, Humana Press (2005) (Humana Press), pp. 571 - 607.

Hiller, N.L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science *306*, 1934-1937.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology *292*, 195-202.

Konrat, R. (2009). The protein meta-structure: a novel concept for chemical and molecular biology. Cellular and molecular life sciences : CMLS *66*, 3625-3639.

LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., *et al*. (2005). A protein interaction network of the malaria parasite Plasmodium falciparum. Nature *438*, 103-107.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., *et al*. (2002). Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature *419*, 537-542.

Marti, M., Good, R.T., Rug, M., Knuepfer, E., and Cowman, A.F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science *306*, 1930-1933.

Sanders, P.R., Gilson, P.R., Cantin, G.T., Greenbaum, D.C., Nebl, T., Carucci, D.J., McConville, M.J., Schofield, L., Hodder, A.N., Yates, J.R., 3rd, *et al*. (2005). Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of Plasmodium falciparum. The Journal of biological chemistry *280*, 40169-40176.

Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., and Cowman, A.F. (2006). Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome biology *7*, R12.

Waller, K.L., Cooke, B.M., Nunomura, W., Mohandas, N., and Coppel, R.L. (1999). 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 *274*, 23808-23813.

Waller, K.L., Nunomura, W., Cooke, B.M., Mohandas, N., and Coppel, R.L. (2002). Mapping the domains of the cytoadherence ligand Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) that bind to the knob-associated histidine-rich protein (KAHRP). Mol Biochem Parasitol *119*, 125-129.

CHAPTER EIGHT

PERFORIN LIKE PROTEINS IN PLASMODIUM

8.1 Sequence based analysis of *P. falciparum* perforin-like proteins

Perforin-like proteins (PLPs) have been indentified in the genomes of different bacterial and protozoan pathogens (Kaiser et al., 2004; Ponting, 1999; Rosado et al., 2007). MACPF- domains have been described in five highly conserved *Plasmodium* proteins (Kaiser et al., 2004), which are expressed at different stages during the malaria life cycle. Our studies focused on the *P. falciparum* encoded PLPs (*Pf*PLPs). The *Pf*PLPs vary in length from 654 to 1073 amino acids and carry a central MACPF domain (Table 8.1). Protein disorder predictions of all five *Pf*PLPs are shown in figure 8.1. *Pf*PLP 1 - 3 are predicted to have a higher possibility of disorder in their N-terminal segment preceding the MACPF domain, whereas *Pf*PLP4 -5 are predicted to be rather ordered. The N-terminal parts of the *Pf*PLPs display considerable variability in length and sequence. It was suggested that this segment makes a unique contribution to the function of the relevant PLP (Kafsack and Carruthers, 2010).

	Plasmodb accession code	**aa length**	**Molecular Weight**	**pI**	**Chromosome location**	**AT content (%)**
*Pf*PLP1	PfD0430c	842	94.5	4.98	4	70.1
*Pf*PLP2	PfL0805w	1073	124.9	7.85	12	75.1
*Pf*PLP3	PfI1145w	871	93.3	6.61	9	71
*Pf*PLP4	Pf08_0050	654	76.3	8.62	8	74.4
*Pf*PLP5	Pf08_0052	676	79.4	8.98	8	75.6

Table 8.1: Overview *of P. falciparum* perforin-like protein (*Pf*PLP) family
Molecular weights were predicted including signal peptides using Protparm (Gasteiger E., 2005). Sequences were obtained from plasmodb (Bahl et al., 2002). The AT content of the corresponding genes is given in %.

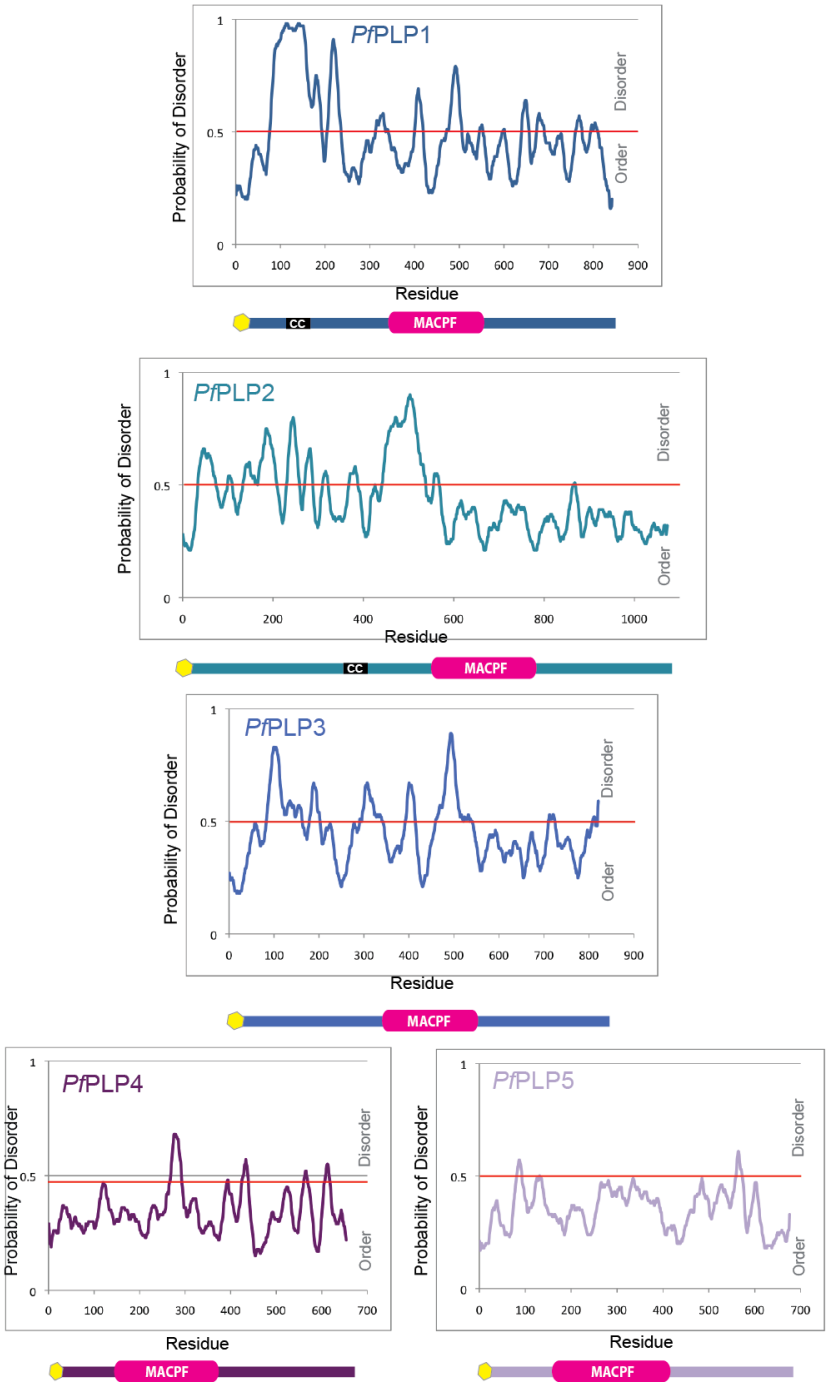


Figure 8.1: Cartoons representing the primary structures of all five *P.falciparum* Perforin-Like Proteins (*Pf*PLPs). All encode for a membrane-attack complex/ perforin-like domain (MACPF) shown in magenta. Domain boundaries were established using SMART (Letunic et al., 2012; Schultz et al., 1998). The yellow boxes indicate the predicted signal peptides/ transmembrane segments and coiled-coil regions are highlighted in black. Plots of predicted disorder for all *Pf*PLPs using the RONN software (Yang et al., 2005) are shown. The threshold for disorder probability is marked in red; scores above the threshold 0.5 are considered disordered.

The central MACF-domain in all variants of the *Pf*PLPs show substantial conservation. A sequence alignment of the *Pf*PLPs with the MACPF- domain of human perforin is given in figure 8.2. The MACPF domains of *P. falciparum* demonstrate a high similarity to MACPF domains found in other organisms (data not shown), including the signature motif W-X_2-[FL]-[FI]-X_2-[FY]-G-T-H-X_7-GG (see Introduction). In addition, homology based tertiary structure predictions (Kelley and Sternberg, 2009) of the *Pf*PLPs, anticipate a conserved MACPF-domain fold for these proteins (shown for *Pf*PLP3 in figure 8.3). The structure prediction model for *Pf*PLP3 displays the structural features characteristic for MACPF-domains - a twisted central β sheet and the two clusters of α helices that are supposed to transform into membrane spanning β barrels.

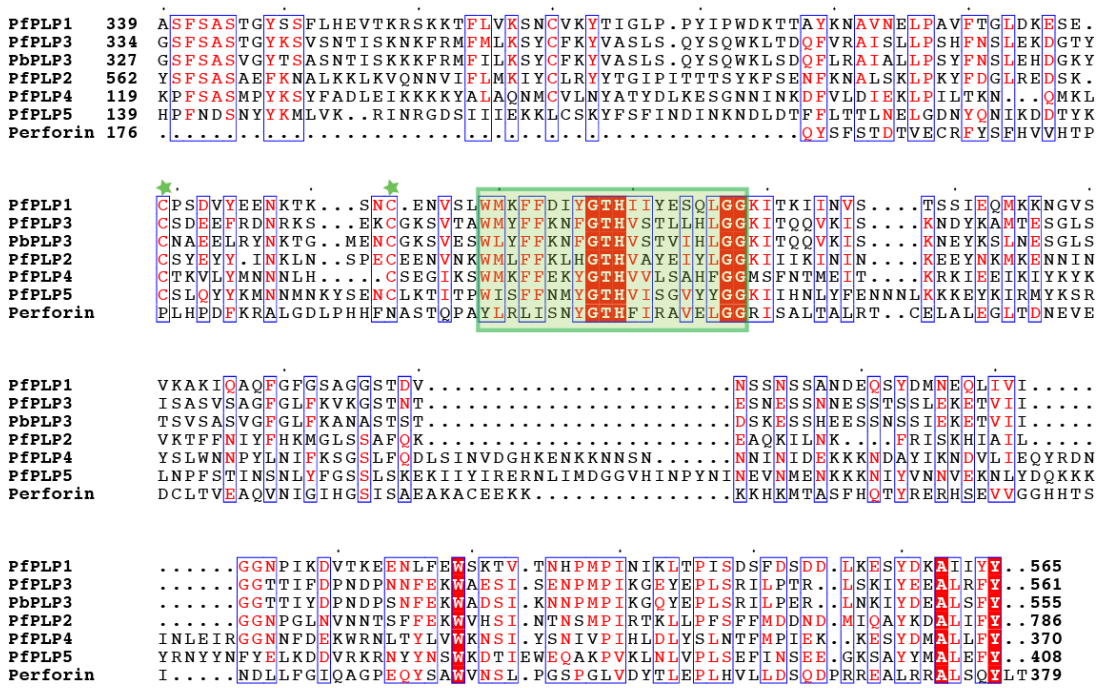


Figure 8.2: Alignment of the predicted MACPF domain of *Pf*PLPs and *Pb*PLP3 with the MACPF domain of human perforin. Identical residues are highlighted in red. The MACPF signature motif is boxed in green. Conserved cysteine residues are indicated by green stars. Residue numbering corresponds to the full-length protein of the *P*PLPs. Sequences were aligned using MULTALIN (Corpet, 1988) and further formatted using ESPRIT (Gouet et al., 1999).

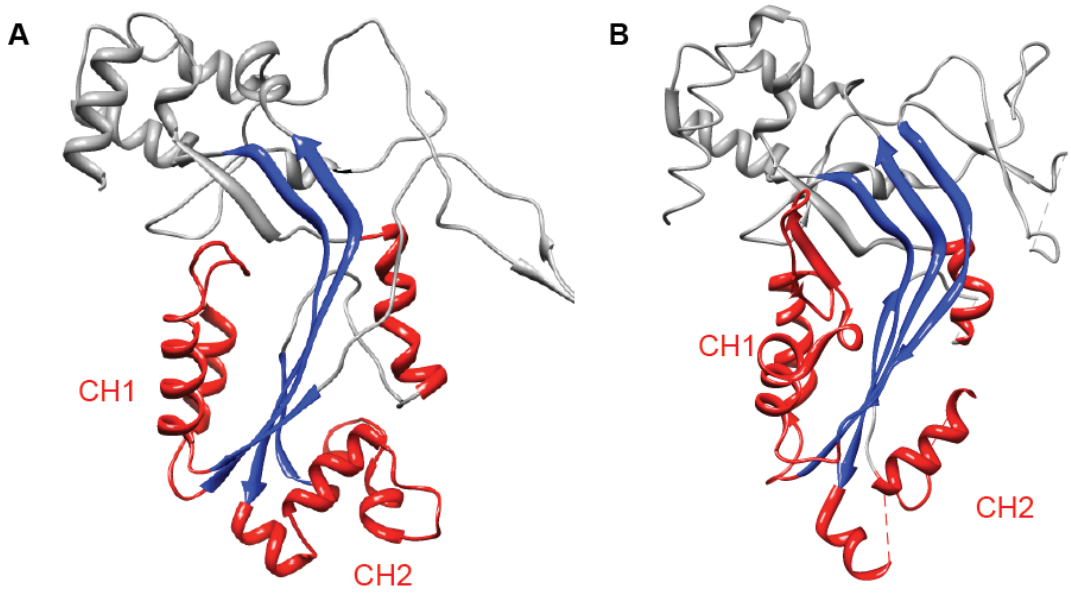


Figure 8.3: Secondary and tertiary structure conservation of the *Pf*PLP3 MACPF domain. A) Structural homology model of the *Pf*PLP3 MACPF domain based on the human C8α-MACPF (PDB ID 2qqh) shown in **(B)**. The central kinked β sheet is coloured in blue and the CH1 and CH2 helical clusters in red. The three-dimensional structure model was calculated by phyre[2] (Kelley and Sternberg, 2009). Nomenclature: CH, Cluster of helices;

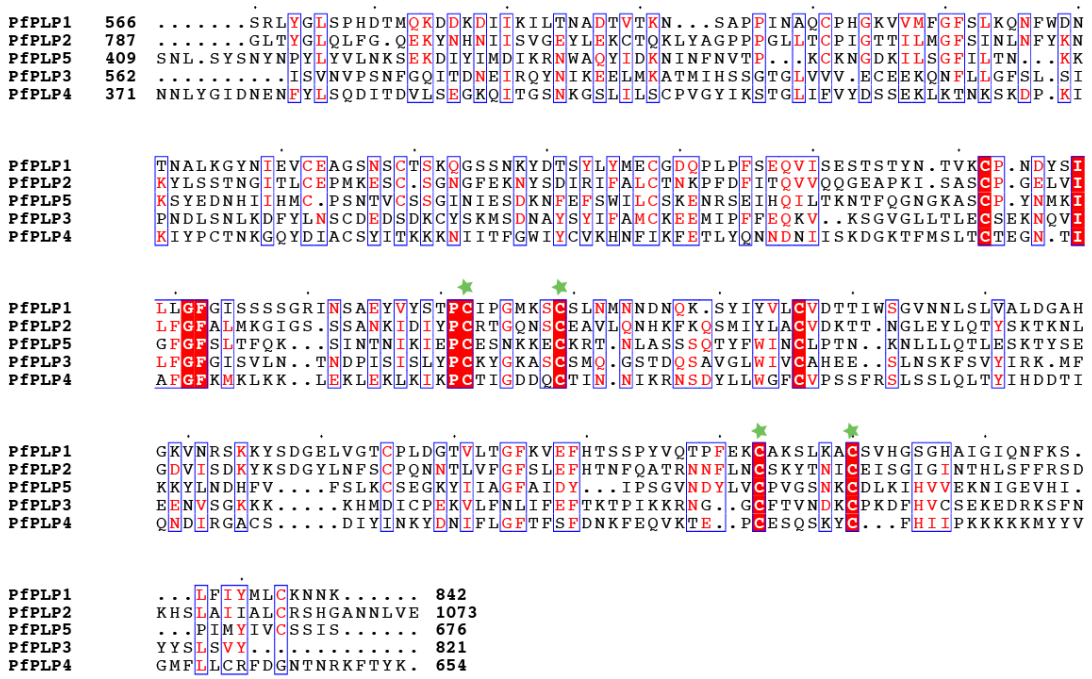


Figure 8.4: Alignment for the C- terminal part following the MACPF domain for all five *Pf*PLPs. The green stars point out the two C-X_6-C motifs, conserved in all *P*PLPs. Sequences were aligned using MULTALIN (Corpet, 1988) and further formatted using ESPRIT (Gouet et al., 1999).

8.2 DNA Cloning of *Pf*PLPs

Since *Pf*PLP2 to *Pf*PLP5 are encoded in a single exon and *Pf*PLP1 has eight exons, we decided to start working with the first mentioned. *Pf*PLP2 - *Pf*PLP5 are encoded in a single open reading frame (ORF) and predicted to be mainly folded, according to the RONN algorithm (Yang et al., 2005) (figure 8.1). The genomic material used as template for this study was provided by MR4 (Malaria Research and Resource Centre) and Dr Sue Kyes (University of Oxford). Initial DNA amplification attempts under standard PCR conditions were unsuccessful, most likely due to the high AT content of the *P. falciparum* genome (Gardner et al., 2002), which is estimated to be above 80%. Some regions of the genome even comprise more than 90 % of AT nucleotides. Furthermore, we employed varying magnesium concentrations and

different DNA polymerases to overcome this obstacle. Besides VENT (NEB, used for standard PCR reactions), we utilized various recommended enzymes, including Taq, Pyrobest (TAKARA), Phusion (Finnzymes) and GoTaq Long PCR Master Mix (Promega). In addition, we also designed longer primers for all genes of interest. Unfortunately, all these attempts failed to produce amplification of the desired DNA molecules. Interestingly, by the suggested lowering of the PCR extension temperature from 72 °C to 65 °C and even 60 °C (Lopez-Barragan et al., 2011; Su et al., 1996) we were able to amplify the entire gene of *Pf*PLP4. The gene encoding for *Pf*PLP4 possesses an AT content of 74.4 % and the reduction of the PCR extension temperature to 60 °C proved to be most successful. However, we were not able to generate any DNA for the other *Pf*PLPs. Therefore, we decided to amplify these *Pf*PLP genes in ~ 500 bp long overlapping DNA fragments and subsequently join them again together by additional PCR steps. This approach proved to be partly successful as we obtained several DNA fragments for *Pf*PLP4 and *Pf*PLP5 (figure 8.2). Furthermore, we were able to amplify the gene for PLP3 of *P. berghi* from genomic DNA provided by Dr Andrew Blaghborough (Imperial College, London). Regrettably, attempts to obtain PLP genes by PCR amplification from another rodent-infecting parasite, *P. yoelii*, were unsuccessful. As a consequence of these ineffective attempts, the DNA amplification process of the *Pf*PLP genes and others turned out to be quite time consuming and expensive and not as successful as anticipated. Therefore, synthetically created DNA constructs were purchased, containing the sequence for *Pf*PLP1 and *Pf*PLP3. Both genes were synthesized by GeneArt (Life Technologies); the DNA sequence encoding for *Pf*PLP1 was additionally optimized for mammalian expression using the GeneOptimizer (Life Technologies) (Fath et al., 2011).

8.3 Expression of *P*PLPs

8.3.1 Expression in bacterial systems

Various DNA segments of PLP genes were cloned in the expression vectors pET15b (Novagen) and pTriEx-2 (Novagen), and subsequently transformed into *E. coli Rosetta-gami* B (DE3)pLysS cells and *Origami B* cells, respectively. Both cell strains allow for disulfide bond formation and encode for codons rarely used in *E. coli*. Such details are likely to be favourable for expression trails of the kinds being undertaken here. Moreover, the pTriEx-2 vector allows for expression in different host systems - *E. coli*, insect and vertebrate cells. Unfortunately, test expressions in *E.coli* cells at varying temperatures and IPTG concentrations remained unsuccessful. Therefore, alternative expression strategies were employed.

8.3.2 Expression in HEK 293T cells

For small-scale expression tests in mammalian HEK 293T cells, we transferred the relevant DNA segments into the expression vector pHLsec (Aricescu et al., 2006). We were able to successfully express full-length *Pf*PLP3, a C-terminal fragment containing the whole MACPF domain and a more C-terminal segment in HEK 293T cells (figure 8.3). The full-length *Pf*PLP3 (90 kDa) is possibly post-translationally processed in the cells, as it appeared in a shortened version on a Western blot (figure 8.4). Interestingly, the expected secretion of the expressed proteins did not occur, neither for constructs carrying their native signal peptide nor for proteins fused to a mammalian secretory signal (encoded by pHLsec).

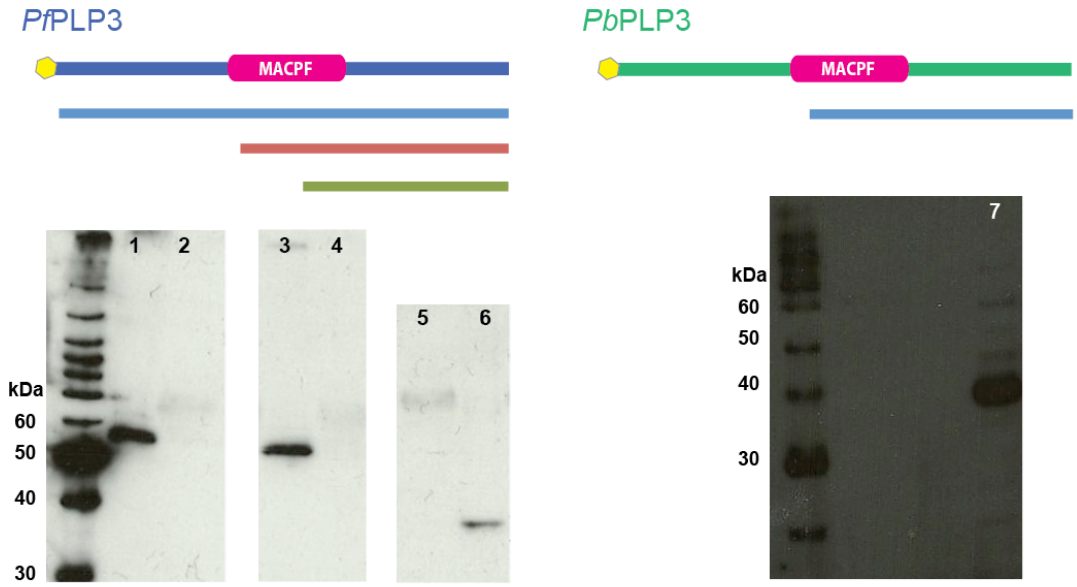


Figure 8.4: Expression of *Pf*PLP3 and *Pb*PLP3 in HEK 293T cells and *Sf*9 cells, respectively. Construct boundaries and lengths are shown with respect to the MACPF domain for both proteins and the signal peptide (yellow). Protein products were visualized using anti-His_6 antibody by western blotting. Lane: (1) *Pf*PLP3 (23-821) cells; (2) *Pf*PLP3 (23-821) media; (3) *Pf*PLP3 (334-821) cells; (4) *Pf*PLP3 (33-821) media; (5) *Pf*PLP3 (538-821) cells; (6) *Pf*PLP3 (538-821) media; (7) *Pb*PLP3 (346-815) expression in Sf9 cells.

8.3. 3 Insect cell based baculovirus expression

At the same time, constructs of the *Pb*PLP3 and *Pf*PLP3 were cloned into the pTriEx-2 vector (Novagen) including a C-terminal His_6-tag. For baculovirus driven expression a volume of 2ml *Sf*9 cells (10^6 cells/ml) were transfected with the sequence verified vectors. After 4 days the supernatant was collected and used for further virus amplification. The remaining cell pellet was analyzed for protein expression via Western blotting. Up to now, a construct comprising *Pb*PLP3 (residues 346-815) could be produced in *Sf*9 cells (figure 8.4).

8.3.4 Cell-free protein expression

In vitro transcription-translation systems have been shown to be successful for different membrane and cytotoxic proteins. In addition, the optimized wheatgerm expression system has proved to be successful for the expression of several *P. falciparum* proteins, with codon optimisation of the genes not making a difference in expression yields (Mudeppa et al., 2007; Tsuboi et al., 2008). For these reasons, we prepared constructs for full-length *Pf*PLP3 and *Pf*PLP4 in the pEU vector for *in vitro* expression using the cell-free wheatgerm system ENDEXT (Endo and Sawasaki, 2007). A small-scale bilayer translation test at 16 °C for 20 hours was performed. However, protein expression could not be verified by Western blotting.

8.4 Localisation studies of *Pf*PLPs

We could not detect any secreted protein from mammalian expression, even though the pHLsec vector encodes for a signal peptide. Therefore, *Pf*PLP3 constructs in the EGFP-pLEXm expression vector were designed. The protein of interested is fused to a C-terminal green fluorescent protein (GFP), which can be used to detect the fusion-protein in HEK 293T cells. Figure 8.5 displays the intracellular location of different *Pf*PLP3-GFP constructs. This result confirms that the protein is not secreted under experimental conditions and suggests that *Pf*PLP3 may be membrane-associated. Therefore, we wanted to test if the *P*PLPs, and *Pf*PLP3 at least, may be membrane-associated. To test this theory, we performed expression tests in COS-7 cells. Briefly, COS-7 cells were grown on glass coverslips and transfected with *Pf*PLP3 (residues 334-561) carrying an N-terminal Flag-tag. Two days post transfection cells were fixed and stained with anti-Flag primary antibody (Sigma) and Alexa 488-labelled secondary antibody (Invitrogen). In addition, the nuclei of the COS-7 cells were counterstained with DAPI (Invitrogen). The cells were visualized by fluorescence microscopy. The anti-Flag primary antibody binds to the N-terminal Flag-tag of the protein of interest, but only if it is exposed on the cell surface. A preliminary result is shown in figure 8.6, which does not argue for the possibility of the exposure of the N-terminal Flag-tag on the extracellular side of the cell membrane. At the time of writing, we are working on refinement of expression in both baculovirus-driven insect and mammalian cell lines. The preliminary aim is to obtain purified protein, which can be used in cell binding and lysis assays. The PLPs are so poorly characterized to date, and their effects in malaria so marked, that any structural or functional information we are able to obtain will be a major advance in the field.

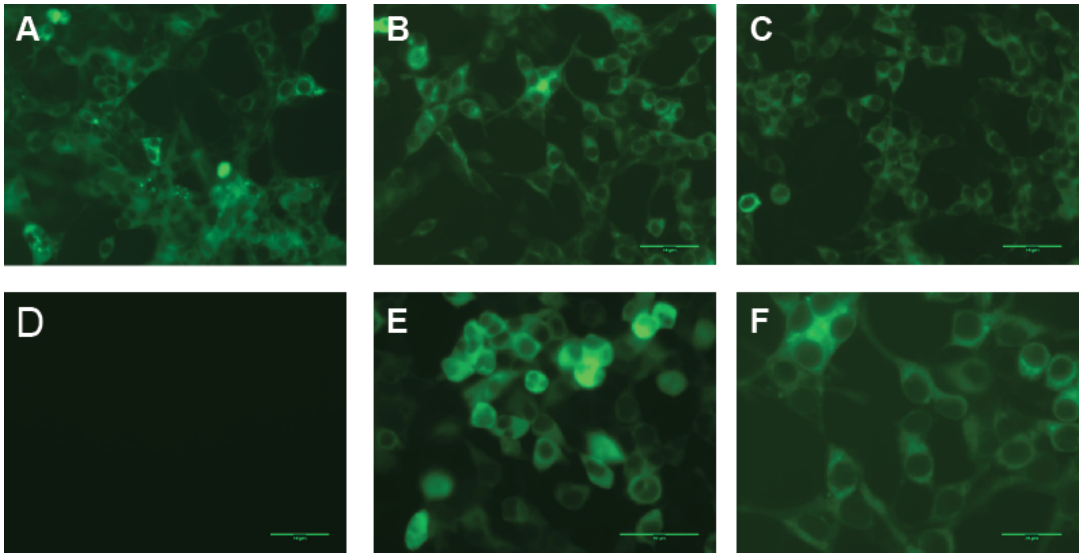


Figure 8.5: Expression of *Pf*PLP3-GFP fusion proteins in HEK 293T cells. A) *Pf*PLP3-GFP (residues 23-821) **B)** and **E)** *Pf*PLP3-GFP (residues 334-821) **C)** and **F)** *Pf*PLP3-GFP (538 -821) **D)** negative control (non-infected HEK 293T cells). Cells shown in **(E)** and **(F)** have been magnified 40x. Scale bar, 40 μm.

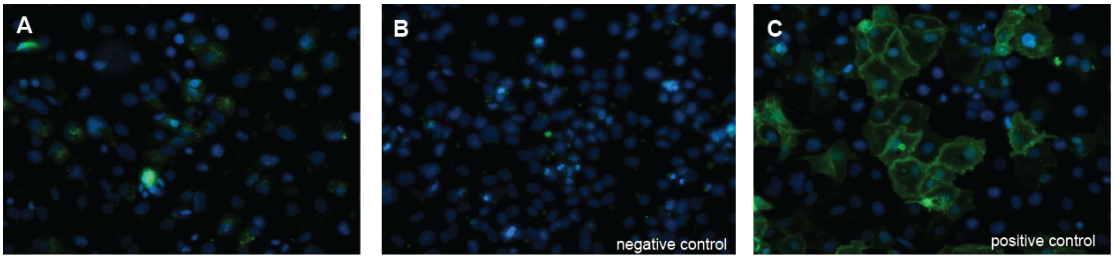


Figure 8.6: Expression of *Pf*PLP3 (MACPF) in COS-7 cells, stained with anti-FLAG antibody. Cell nuclei were stained blue with DAPI. **A)** *Pf*PLP3 MACPF domain (residues 334 -561) **B)** Negative control **C)** Positive control (provided by Dr Christian Bell, University of Oxford).

8.6 References

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta crystallographica Section D, Biological crystallography *62*, 1243-1250.

Bahl, A., Brunk, B., Coppel, R.L., Crabtree, J., Diskin, S.J., Fraunholz, M.J., Grant, G.R., Gupta, D., Huestis, R.L., Kissinger, J.C., *et al*. (2002). PlasmoDB: the Plasmodium genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished). Nucleic acids research *30*, 87-90.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic acids research *16*, 10881-10890.

Endo, Y., and Sawasaki, T. (2007). [Wheat germ cell-free protein synthesis]. Seikagaku The Journal of Japanese Biochemical Society *79*, 229-238.

Fath, S., Bauer, A.P., Liss, M., Spriestersbach, A., Maertens, B., Hahn, P., Ludwig, C., Schafer, F., Graf, M., and Wagner, R. (2011). Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. PloS one *6*, e17596.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., *et al*. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498-511.

Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, Humana Press (2005) (Humana Press), pp. 571 - 607.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305-308.

Kafsack, B.F., and Carruthers, V.B. (2010). Apicomplexan perforin-like proteins. Communicative & integrative biology *3*, 18-23.

Kaiser, K., Camargo, N., Coppens, I., Morrisey, J.M., Vaidya, A.B., and Kappe, S.H. (2004). A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. Mol Biochem Parasitol *133*, 15-26.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols *4*, 363-371.

Letunic, I., Doerks, T., and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. Nucleic acids research *40*, D302-305.

Lopez-Barragan, M.J., Quinones, M., Cui, K., Lemieux, J., Zhao, K., and Su, X.Z. (2011). Effect of PCR extension temperature on high-throughput sequencing. Mol Biochem Parasitol *176*, 64-67.

Mudeppa, D.G., Pang, C.K., Tsuboi, T., Endo, Y., Buckner, F.S., Varani, G., and Rathod, P.K. (2007). Cell-free production of functional Plasmodium falciparum dihydrofolate reductase-thymidylate synthase. Mol Biochem Parasitol *151*, 216-219.

Ponting, C.P. (1999). Chlamydial homologues of the MACPF (MAC/perforin) domain. Current biology : CB *9*, R911-913.

Rosado, C.J., Buckle, A.M., Law, R.H., Butcher, R.E., Kan, W.T., Bird, C.H., Ung, K., Browne, K.A., Baran, K., Bashtannyk-Puhalovich, T.A., *et al*. (2007). A common fold mediates vertebrate defense and bacterial attack. Science *317*, 1548-1551.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America *95*, 5857-5864.

Su, X.Z., Wu, Y., Sifri, C.D., and Wellems, T.E. (1996). Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. Nucleic acids research *24*, 1574-1575.

Tsuboi, T., Takeo, S., Iriko, H., Jin, L., Tsuchimochi, M., Matsuda, S., Han, E.T., Otsuki, H., Kaneko, O., Sattabongkot, J., *et al*. (2008). Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. Infect Immun *76*, 1702-1708.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics *21*, 3369-3376.

CHAPTER NINE

DISCUSSION

9.1 The flexible nature of ATS and KAHRP

The ATS domains are the intracellular part of the *Pf*EMP1 family of proteins, which are unique to *P. falciparum*. This protein family contributes to the virulence of this *Plasmodium* strain, which is responsible for hundreds of thousands of deaths per year. 59 variants of this protein are found in the genome of *P. falciparum*. We have performed extensive structural studies on ATS (PF08_0141). Recombinantly expressed full length and shortened versions of ATS remain monomeric in solution, as determined by AUC. Previous gel permeation chromatography results (Hora et al., 2009) had been interpreted as showing that ATS forms a dimer. This discrepancy can be explained by the largely unstructured nature of ATS. Experimental data and bioinformatics analysis show that the domain is highly flexible, with a small central folded core. Because ATS is mainly unfolded it has a large hydrodynamic radius, which probably explains its atypical elution volume. We extended our investigations to look at more variants of the ATS-family to see if the disordered nature is a common theme in this family. NMR, CD and SAXS analysis of five additional representative ATS domains confirmed the mainly disordered state of the domain. Therefore, we concluded that the molecular architecture determined for ATS-FL of PF08_0141 is conserved throughout the *Pf*EMP1 family.

Bioinformatics analysis of KAHRP also characterized this protein as intrinsically disordered, in particular the C-terminal region (residues 351-654). Intrinsically disordered proteins are defined as lacking a folded structure and they have extended conformations with high intramolecular flexibility and little secondary structure (Tompa, 2002) under physiological conditions. Disorder prediction of entire genomes suggests that disorder is a common feature in most species analyzed so far. Notably, eukaryotic genomes show a higher frequency of disordered proteins compared to bacterial ones. Intrinsically disordered regions (IDR) of more than 30 consecutive residues have been predicted to exist in over 30 % of eukaryotic proteins (Ward et al., 2004). Using a different prediction method IDRs were estimated to be as high as 63 % in *Drosophila melanogaster* proteins (Dunker et al., 2000).

Interestingly, analysis of pathogenic organisms revealed that IDPs are also prevalent in this class. The phylum *Apicomplexa* contains the highest percentage of structural disorder within this group of pathogens (figure 9.1) (Feng et al., 2006). Members of this phylum include highly pathogenic organisms such as *Toxoplasma, Trypanosoma* and *Plasmodium*. In addition, the number of IDPs is predicted to be higher in *Plasmodium* species infecting mammals than those infecting rodents (Feng et al., 2006). Disorder predictions for the proteome of *P. falicparum* estimate 34 % of encoded proteins to be natively disordered (Mohan et al., 2008). In particular, over 90 % of all proteins encoded on chromosome 2 and 3 in *P. falciparum* are predicted to contain low complexity regions (Pizzi and Frontali, 2001). Furthermore, the proteome of four different life stages of *P. falciparum* has been analyzed. This study revealed that proteins expressed during the sporozoite phase are the most enriched in IDPs (Feng et al., 2006).

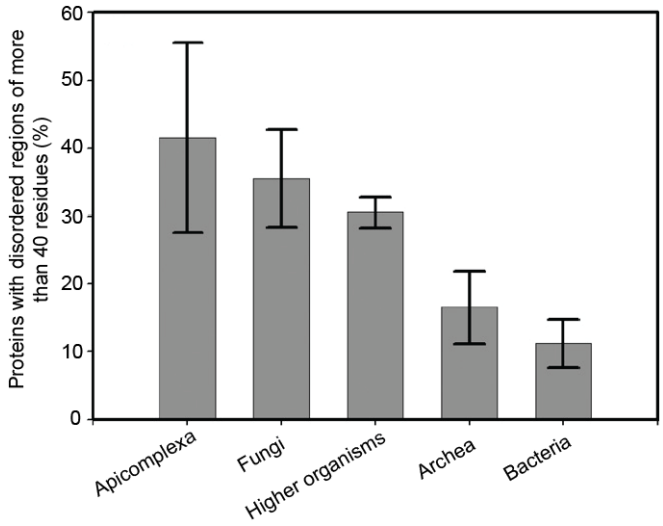


Figure 9.1: Comparison of different groups of organisms according to the average percentage of proteins that contain long disordered regions. The apicomplexan species include: *P. falciparum, P. yoelii, T. gondii, C. parvum, C. hominis, T. parvum*. The higher organisms include: *A. thaliana, B. rerio, R. norvegicus, C. elegans, D. melanogaster, M. musculus, H. sapiens*. Adapted from (Feng et al., 2006).

Given the high content of IDPs in the proteome of *P. falciparum*, it is not surprising that two of the proteins studied in this thesis, ATS and KAHRP, are mainly disordered. We were only able to experimentally confirm structured elements for 19 % of ATS. Full length KAHRP was resistant to heterologous expression and, therefore, could not be analyzed experimentally. However, prediction tools suggest that approximately 60 % of KAHRP is natively unfolded in physiological conditions. Interestingly, other *Plasmodium* encoded proteins involved in cytoadherence such as *Pf*EMP2 and *Pf*EMP3 also display an extremely high probability of being intrinsically disordered (figure 9.2). What are the benefits of structural disorder for these proteins? It has been suggested that IDRs may be advantageous for protein-protein interactions. A correlation study of the interactome of *P. falciparum* proposes

a weak association between intrinsic disorder and protein-protein binding (Mohan et al., 2008). The question arises, how does intrinsic disorder facilitate protein-protein interactions? On one hand, IDRs can experience binding-induced order promoting transitions, which are accompanied by a large decrease in conformational entropy that makes specific interactions reversible (Kriwacki et al., 1996). In addition, IDRs provide larger surfaces for intermolecular interactions due to their extended conformation compared to globular proteins composed of the same number of residues (Gunasekaran et al., 2003). Larger surfaces provide the possibility for simultaneous binding to several binding partners, which would facilitate the formation of multi-partner complexes (Meszaros et al., 2007). The benefits of an extended binding surface are obvious for proteins acting as assemblers - bringing different proteins together. Functional cytoadherence requires the ATS domain to properly anchor the extracellular domains to the host skeleton. Clearly, several interactions of ATS and / or the other involved proteins (KAHRP, *Pf*EMP2, *Pf*EMP3 etc.) would therefore be beneficial for creating a functional and strong 'knob-structure network'. Interestingly, it was suggested that larger complexes are likely to be built from proteins that are prone to be more disordered (Hegyi et al., 2007). Disordered proteins, compared to stably folded proteins, are assumed have the same interaction potential but with smaller protein size. This allows for the same functional capacity, however, less genomic space is necessary (Gunasekaran et al., 2003). In the case of *Plasmodium,* which has to adapt to different hosts and to go through several different life stages, saving genomic space would certainly be advantageous.

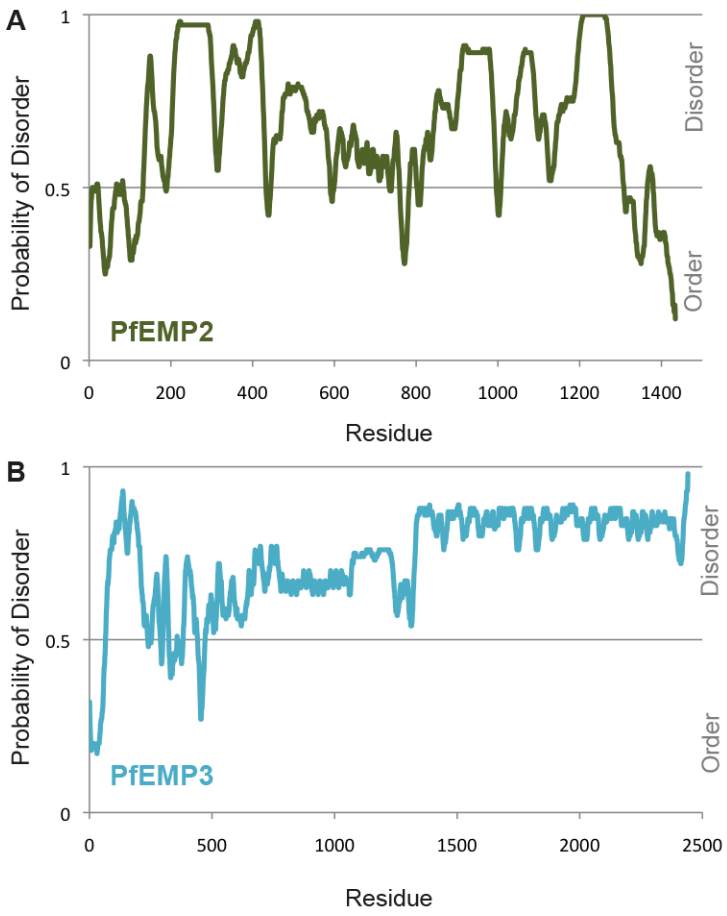


Figure 9.2: RONN disorder prediction for *Pf*EMP2 (A) and *Pf*EMP3 (B) (Yang et al., 2005). The threshold for disorder probability is 0.5, values above that have a significant probability of disorder.

An extended nature not only allows for promiscuity, but also gives a greater 'capture radius' (Gsponer and Babu, 2009). This so-called 'fly-casting' mechanism, assumes that IDPs can search a larger volume of space to look for binding partners compared to compact structured proteins of the same size and thus increase the speed of interaction (Levy et al., 2004; Pontius, 1993; Shoemaker et al., 2000). Moreover, the 'fly-casting' concept implies that the initial recognition of the flexible segment is a weak non-specific interaction, such as electrostatics, followed by folding and specific binding (Sugase et al., 2007). Our studies showed that the middle segment of ATS is

disordered and that this disorder is conserved. In addition, this IDR contains an interaction epitope that binds to the novel protein domain PHIST of PFI1780w. Therefore, we can conclude that the IDR of ATS is of advantage for its interaction properties. However, the structured core of ATS is necessary for the arrangement and therefore exposure of the flexible middle loop. In this case not solely the IDR, but also the orientation and presentation of the flexible sequence by the structured part of the ATS domain seem favourable for the binding. Furthermore, extensive IDRs could enable proteins to wrap around their binding partners, creating a high-affinity binding, potentially in the nM range (Dames et al., 2002). Both ends of the ATS domain as well as the C-terminal part of KAHRP display extensive flexibility that would allow for such an interaction mechanism. However, this is just one hypothetical possibility of how components of the knob structures, such as KAHRP, *Pf*EMP2 and *Pf*EMP3 could assemble.

In summary, protein flexibility comes with functional advantages for protein-protein interactions (Tompa, 2003). However, the study of IDPs is challenging for structural biology. Firstly, protein expression is problematic due to the fact that IDPs are more prone to proteolytic degradation, which results in low expression yields. Secondly, this class of proteins is resistant to protein crystallization and therefore structure determination by X-ray crystallography is not feasible. For these reasons, structural information on parts of ATS, KAHRP, *Pf*EMP1 and *Pf*EMP2 and other *Plasmodium* proteins has been limited. However, in recent years, developments in NMR and SAXS have opened up new possibilities for structural studies of IDPs (Meier et al., 2008; Mittag and Forman-Kay, 2007; Mukrasch et al., 2007; Theillet et al., 2011; Wu et al., 2008). These advances will hopefully lead to a better understanding of the

intracellular arrangements and mechanisms of knob-like structures and provide information on potential drug targets.

9.2 Does ATS interact with KAHRP?

Cytoadhesion is mediated by the various extracellular domains of *PfEMP1*, whereas the intracellular ATS domain must be involved in anchoring the protein to the iRBC's cytoskeleton. Several studies suggest that KAHRP is an important player in this crucial connection (Oh et al., 2000; Voigt et al., 2000; Waller et al., 1999). KAHRP is essential for knob formation and knockout experiments have shown that it is crucial for functional cytoadherence of iRBCs (Crabb et al., 1997). For these reasons KAHRP is believed to bind directly to ATS (Maier et al., 2009). Published pull-down experiments involving full-length ATS and fragments of ATS and KAHRP determined two interaction sites on both proteins. Under the reported experimental conditions, K_D values of 0.2 μM - 1.1 μM were deduced (Waller et al., 1999). A different study described the ATS-KAHRP interaction as even stronger, with an estimated $K_D \sim 10$ nM. In the malaria field, the lack of an interaction between ATS and KAHRP seems rather inconceivable. However, despite the potentially vital nature of this association no supporting *in vivo* data have been presented to date. We have extensively investigated the proposed interaction by NMR titration experiments as well as fluorescence anisotropy measurements using various ATS and KAHRP constructs. Our studies involved protein concentrations more than two orders of magnitude larger than the reported affinity constants. Therefore, we conclude that an interaction is very unlikely, or too weak to be detected under solution conditions. This means that the dissociation constant K_D would be in the order of several mM, which might be physiologically irrelevant.

The ATS domain has been previously reported to be phosphorylated by casein kinase II (CKII) (Hora et al., 2009), which also phosphorylates a wide range of cytoskeletal components, including spectrin, ankyrin and adducsin (Fukata et al., 1999; Ghosh et al., 2002; Patel and Fairbanks, 1981). Moreover, phosphorylated ATS has been described as being composed of β sheets and random coils with altered binding affinities of KAHRP fragments K1A and K2A (Hora et al., 2009). Our CD measurements of ATS did not indicate the existence of β sheets, but revealed extensive disorder and minimal α-helical elements. These observations could be confirmed by the determination of the ATS-core structure, which comprises four α-helices. Protein phosphorylation is known to play an important role in almost all cellular processes (Cohen, 1982; Pawson and Scott, 1997, 2005). IDPs are predicted to be enriched with phosphorylation sites (Dunker et al., 2008; Iakoucheva et al., 2004) and post-translational modifications have been shown to modify structural and folding properties (Wright and Dyson, 2009) and are key to their function and regulation (Bode and Dong, 2004; Brooks and Gu, 2003; Kouzarides, 2007). Our results confirm the phosphorylation of ATS by CKII under *in vitro* conditions. In addition, we were able to observe the *in situ* phosphorylation reaction by solution NMR spectroscopy. Preliminary titration studies of ^{15}N-enriched and phosphorylated ATS with K2, could not, however, detect an interaction between ATS and K2. However, further investigations will be necessary to clarify the role and effect of ATS phosphorylation on the binding properties.

To sum up, we could not confirm an interaction between ATS and KAHRP under our experimental conditions and significant binding seems rather unlikely. However, we were not able to investigate the full-length version of KAHRP and its potential

binding characteristics to ATS, due to the failure of successful protein expression. It is possible that the entire protein displays other essential features, for example a folded segment, which would promote the interaction to ATS, or induce some local folding/ ordering upon molecular recognition. However, this assumption is rather unlikely and we believe that, if such an interaction exists, it could be mediated or strengthened by one or more additional, unknown proteins. For this reason we searched for alternative binding partners of ATS.

9.3 The conserved interaction epitope in the ATS domain

A large-scale yeast-two-hybrid interactome of *P. falciparum* suggested thirteen alternative ATS binding targets (LaCount et al., 2005). Remarkably, all of these interactions involved at least part of the flexible ATS internal loop, which is substantially conserved despite its lack of structure. Interestingly, bioinformatics analysis of the ATS sequence, also predicted an interaction epitope located in the internal loop. While yeast-two-hybrid and bioinformatics each offer circumstantial evidence their independent convergence is highly suggestive of the importance of this loop in protein interactions. The conservation of IDRs has been described previously and most conserved IDRs exist in proteins associated with DNA/RNA binding, ribosomal function and signalling functions (Chen et al., 2006a, b). We hypothesize that the only structured element of ATS - the α-helical core structure - forces the flexible middle segment to form a loop structure that mediates the presentation of the conserved interaction epitope in an optimal way. The restraint provided by the ATS-core could potentially lower the entropic penalty for protein binding, thus enhancing the binding affinity.

In contrast to previous reports of an ATS-KAHRP interaction, we argue that the loop structure of ATS mediates protein contacts with different putative targets. For most of the proteins suggested to bind to the ATS domain by the Y2H study (LaCount et al., 2005), little or no information is available. While their function and location are unknown, they are predicted to be conserved according to plasmodb (www.plasmodb.org) (Bahl et al., 2002). Our studies show that at least one of the twelve proteins interacts with ATS (PF08_0141) in a manner requiring the central

flexible segment. NMR and fluorescence anisotropy experiments revealed the binding of PFI1780w to ATS. The function of this protein is not yet known, although studies located it to the erythrocyte cytosol and it was detected in membrane fractions (Sanders et al., 2005). The existence of a PEXEL motif strongly suggests that PFI1780w is exported to the iRBC. Interestingly, PFI1780w comprises a PHIST-type domain, a recently identified novel gene family (Sargeant et al., 2006). Three subgroups make up the PHIST family and PFI1780w contains a PHIST-c type domain. Moreover, the domain is widespread within *P. falciparum* and *P. vivax* proteins (Sargeant et al., 2006), yet no function has been attributed to a domain of this type. Therefore, PFI1780w is an exciting novel player in cytoadherence of *P. falciparum*. The importance of PHIST domains for knob formation has been shown recently - a gene knockout of PFD1170c, which contains a PHISTb domain, resulted in a 'knobless' pheonotype of *P. falciparum* infected erythrocytes (Maier et al., 2008). In addition, a PHIST-c type encoding protein has been reported to be crucial for *Pf*EMP1 transport to the host cell membrane (Maier et al., 2008). We hypothesize that in PFI1780w, the PHIST domains potentially serve as connector linking parts of the parasite intracellular knob structure network. Future studies will be necessary to determine interaction properties of other proteins involved in cytoadherence, for example KAHRP and host skeletal components. The interaction of ATS with the PHIST domain of PFI1780w is only one piece of the jigsaw puzzle - intensive studies of the molecular interactions are necessary to fully understand the rearrangement of the red blood cell by *P. falciparum*.

9.4 Expression of pore-forming proteins in *P. falciparum*

Pore-forming protein families have been discovered in different pathogenic organisms (Almeida-Campos et al., 2002), including *P. falciparum* (Kaiser et al., 2004). However, the mechanism of pore-formation involving these proteins is still not well understood and no structural information is available at present. Our aim was to express these proteins and conduct structural and functional studies. However, the amplification of relevant DNA molecules from *P. falciparum* proved difficult, most possibly due to the AT-rich sequence of *P. falciparum* (Su et al., 1996) and full-length genes could not be amplified from genomic material except for *Pf*PLP4. In addition, attempts to express the proteins *Pf*PLPs (and KAHRP) in various *E.coli* cell strains were unsuccessful. Interestingly, *P. falciparum* proteins have been described to be resistant to heterologous expression in *E.coli* (Mehlin et al., 2006). Higher molecular weights, greater protein disorder, a more basic isoelectric point and a lack of homology to *E. coli* proteins were all highly and independently correlated with difficulties in expression. Surprisingly, codon usage and the percentage of adenosines and thymidines did not appear to play a significant role (Mehlin et al., 2006). Small-scale expression tests in mammalian cells turned out to be successful for constructs of *Pf*PLP3 (not codon-optimized). Surprisingly, codon usage and the high AT percentage did not hinder the expression of *Pf*PLP3 constructs, which is consistent with previous studies (Mehlin et al., 2006). The expression of full-length *Pf*PLP3 yielded a fragment of lower molecular weight than expected. One possibility is that post-translational processing in the cell resulted in a shortened protein. Interestingly, the *Toxoplasma* perforin-like protein 1, which has been shown to be responsible for egress of *Toxoplasma tachyzoites* from the host cell (Kafsack et al., 2009) is proteolytically processed in the N-terminal region during trafficking (Kafsack and

Carruthers, 2010). Initial studies of the *Pf*PLPs also suggest that they might be membrane-associated rather than being a classical pore-forming protein. Other pore-forming proteins have been localized to the plasma membrane, including perforin-2 from macrophages (Podack et al., 2007). The *P*PLPs might operate as fusion proteins to enable the seamless traversal of the parasite in and out of cells. The work presented in this thesis has just laid the groundwork for the study of PLPs encoded in *Plasmodium*. Future work will include the optimization of the expression of different *Pf*PLP3 and *Pb*PLP3 constructs in mammalian and Sf9 cells. We have already cloned *Pf*PLP1 (codon-optimized) in fragments of 12 different lengths, which we are expressing using baculovirus in *Sf*9 cells and HEK293T cells by transient transfection. Successful recombinant expression of one of the *P*PLPs will enable us to characterize the structures formed by these proteins on target membranes and so determine their mechanism of action. Understanding the activity of these proteins is of great importance, as they are essential for invasion of and replication within human and mosquito hosts and therefore they have great potential for use in vaccines and in molecular targeting (Grode et al., 2005; Tchilian et al., 2009).

9.5 References

Almeida-Campos, F.R., Noronha, F.S., and Horta, M.F. (2002). The multitalented pore-forming proteins of intracellular pathogens. Microbes and infection / Institut Pasteur *4*, 741-750.

Bahl, A., Brunk, B., Coppel, R.L., Crabtree, J., Diskin, S.J., Fraunholz, M.J., Grant, G.R., Gupta, D., Huestis, R.L., Kissinger, J.C., *et al*. (2002). PlasmoDB: the Plasmodium genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished). Nucleic acids research *30*, 87-90.

Bode, A.M., and Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. Nature reviews Cancer *4*, 793-805.

Brooks, C.L., and Gu, W. (2003). Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Current opinion in cell biology *15*, 164-171.

Chen, J.W., Romero, P., Uversky, V.N., and Dunker, A.K. (2006a). Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. Journal of proteome research *5*, 879-887.

Chen, J.W., Romero, P., Uversky, V.N., and Dunker, A.K. (2006b). Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. Journal of proteome research *5*, 888-898.

Cohen, P. (1982). The role of protein phosphorylation in neural and hormonal control of cellular activity. Nature *296*, 613-620.

Crabb, B.S., Cooke, B.M., Reeder, J.C., Waller, R.F., Caruana, S.R., Davern, K.M., Wickham, M.E., Brown, G.V., Coppel, R.L., and Cowman, A.F. (1997). Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell *89*, 287-296.

Dames, S.A., Martinez-Yamout, M., De Guzman, R.N., Dyson, H.J., and Wright, P.E. (2002). Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. Proceedings of the National Academy of Sciences of the United States of America *99*, 5271-5276.

Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., and Brown, C.J. (2000). Intrinsic protein disorder in complete genomes. Genome informatics Workshop on Genome Informatics *11*, 161-171.

Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., and Uversky, V.N. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. BMC genomics *9 Suppl 2*, S1.

Feng, Z.P., Zhang, X., Han, P., Arora, N., Anders, R.F., and Norton, R.S. (2006). Abundance of intrinsically unstructured proteins in P. falciparum and other apicomplexan parasite proteomes. Mol Biochem Parasitol *150*, 256-267.

Fukata, Y., Oshiro, N., Kinoshita, N., Kawano, Y., Matsuoka, Y., Bennett, V., Matsuura, Y., and Kaibuchi, K. (1999). Phosphorylation of adducin by Rho-kinase plays a crucial role in cell motility. The Journal of cell biology *145*, 347-361.

Ghosh, S., Dorsey, F.C., and Cox, J.V. (2002). CK2 constitutively associates with and phosphorylates chicken erythroid ankyrin and regulates its ability to bind to spectrin. Journal of cell science *115*, 4107-4115.

Grode, L., Seiler, P., Baumann, S., Hess, J., Brinkmann, V., Nasser Eddine, A., Mann, P., Goosmann, C., Bandermann, S., Smith, D., *et al*. (2005). Increased vaccine efficacy against tuberculosis of recombinant Mycobacterium bovis bacille Calmette-Guerin mutants that secrete listeriolysin. The Journal of clinical investigation *115*, 2472-2479.

Gsponer, J., and Babu, M.M. (2009). The rules of disorder or why disorder rules. Progress in biophysics and molecular biology *99*, 94-103.

Gunasekaran, K., Tsai, C.J., Kumar, S., Zanuy, D., and Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. Trends in biochemical sciences *28*, 81-85.

Hegyi, H., Schad, E., and Tompa, P. (2007). Structural disorder promotes assembly of protein complexes. BMC structural biology *7*, 65.

Hora, R., Bridges, D.J., Craig, A., and Sharma, A. (2009). Erythrocytic casein kinase II regulates cytoadherence of Plasmodium falciparum-infected red blood cells. The Journal of biological chemistry *284*, 6260-6269.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z., and Dunker, A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic acids research *32*, 1037-1049.

Kafsack, B.F., and Carruthers, V.B. (2010). Apicomplexan perforin-like proteins. Communicative & integrative biology *3*, 18-23.

Kafsack, B.F., Pena, J.D., Coppens, I., Ravindran, S., Boothroyd, J.C., and Carruthers, V.B. (2009). Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells. Science *323*, 530-533.

Kaiser, K., Camargo, N., Coppens, I., Morrisey, J.M., Vaidya, A.B., and Kappe, S.H. (2004). A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. Mol Biochem Parasitol *133*, 15-26.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kriwacki, R.W., Hengst, L., Tennant, L., Reed, S.I., and Wright, P.E. (1996). Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proceedings of the National Academy of Sciences of the United States of America *93*, 11504-11509.

LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., *et al*. (2005). A protein interaction network of the malaria parasite Plasmodium falciparum. Nature *438*, 103-107.

Levy, Y., Wolynes, P.G., and Onuchic, J.N. (2004). Protein topology determines binding mechanism. Proceedings of the National Academy of Sciences of the United States of America *101*, 511-516.

Maier, A.G., Cooke, B.M., Cowman, A.F., and Tilley, L. (2009). Malaria parasite proteins that remodel the host erythrocyte. Nature reviews Microbiology *7*, 341-354.

Maier, A.G., Rug, M., O'Neill, M.T., Brown, M., Chakravorty, S., Szestak, T., Chesson, J., Wu, Y., Hughes, K., Coppel, R.L., *et al*. (2008). Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes. Cell *134*, 48-61.

Mehlin, C., Boni, E., Buckner, F.S., Engel, L., Feist, T., Gelb, M.H., Haji, L., Kim, D., Liu, C., Mueller, N., *et al*. (2006). Heterologous expression of proteins from Plasmodium falciparum: results from 1000 genes. Mol Biochem Parasitol *148*, 144-160.

Meier, S., Blackledge, M., and Grzesiek, S. (2008). Conformational distributions of unfolded polypeptides from novel NMR techniques. The Journal of chemical physics *128*, 052204.

Meszaros, B., Tompa, P., Simon, I., and Dosztanyi, Z. (2007). Molecular principles of the interactions of disordered proteins. Journal of molecular biology *372*, 549-561.

Mittag, T., and Forman-Kay, J.D. (2007). Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol *17*, 3-14.

Mohan, A., Sullivan, W.J., Jr., Radivojac, P., Dunker, A.K., and Uversky, V.N. (2008). Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. Molecular bioSystems *4*, 328-340.

Mukrasch, M.D., Markwick, P., Biernat, J., Bergen, M., Bernado, P., Griesinger, C., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2007). Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. J Am Chem Soc *129*, 5235-5243.

Oh, S.S., Voigt, S., Fisher, D., Yi, S.J., LeRoy, P.J., Derick, L.H., Liu, S., and Chishti, A.H. (2000). Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. Mol Biochem Parasitol *108*, 237-247.

Patel, V.P., and Fairbanks, G. (1981). Spectrin phosphorylation and shape change of human erythrocyte ghosts. The Journal of cell biology *88*, 430-440.

Pawson, T., and Scott, J.D. (1997). Signaling through scaffold, anchoring, and adaptor proteins. Science *278*, 2075-2080.

Pawson, T., and Scott, J.D. (2005). Protein phosphorylation in signaling--50 years and counting. Trends in biochemical sciences *30*, 286-290.

Pizzi, E., and Frontali, C. (2001). Low-complexity regions in Plasmodium falciparum proteins. Genome research *11*, 218-229.

Podack, E.R., Deyev, V., and Shiratsuchi, M. (2007). Pore formers of the immune system. Advances in experimental medicine and biology *598*, 325-341.

Pontius, B.W. (1993). Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. Trends in biochemical sciences *18*, 181-186.

Sanders, P.R., Gilson, P.R., Cantin, G.T., Greenbaum, D.C., Nebl, T., Carucci, D.J., McConville, M.J., Schofield, L., Hodder, A.N., Yates, J.R., 3rd, *et al*. (2005). Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of Plasmodium falciparum. The Journal of biological chemistry *280*, 40169-40176.

Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., and Cowman, A.F. (2006). Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome biology *7*, R12.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proceedings of the National Academy of Sciences of the United States of America *97*, 8868-8873.

Su, X.Z., Wu, Y., Sifri, C.D., and Wellems, T.E. (1996). Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. Nucleic acids research *24*, 1574-1575.

Sugase, K., Dyson, H.J., and Wright, P.E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature *447*, 1021-1025.

Tchilian, E.Z., Desel, C., Forbes, E.K., Bandermann, S., Sander, C.R., Hill, A.V., McShane, H., and Kaufmann, S.H. (2009). Immunogenicity and protective efficacy of prime-boost regimens with recombinant (delta)ureC hly+ Mycobacterium bovis BCG and modified vaccinia virus ankara expressing M. tuberculosis antigen 85A against murine tuberculosis. Infect Immun *77*, 622-631.

Theillet, F.X., Binolfi, A., Liokatis, S., Verzini, S., and Selenko, P. (2011). Paramagnetic relaxation enhancement to improve sensitivity of fast NMR methods: application to intrinsically disordered proteins. Journal of biomolecular NMR *51*, 487-495.

Tompa, P. (2002). Intrinsically unstructured proteins. Trends in biochemical sciences *27*, 527-533.

Tompa, P. (2003). The functional benefits of protein disorder. J Mol Struc-Theochem *666*, 361-371.

Voigt, S., Hanspal, M., LeRoy, P.J., Zhao, P.S., Oh, S.S., Chishti, A.H., and Liu, S.C. (2000). The cytoadherence ligand Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) binds to the P. falciparum knob-associated histidine-rich protein (KAHRP) by electrostatic interactions. Mol Biochem Parasitol *110*, 423-428.

Waller, K.L., Cooke, B.M., Nunomura, W., Mohandas, N., and Coppel, R.L. (1999). 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 *274*, 23808-23813.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of molecular biology *337*, 635-645.

Wright, P.E., and Dyson, H.J. (2009). Linking folding and binding. Curr Opin Struct Biol *19*, 31-38.

Wu, K.P., Kim, S., Fela, D.A., and Baum, J. (2008). Characterization of conformational and dynamic properties of natively unfolded human and mouse alpha-synuclein ensembles by NMR: implication for aggregation. Journal of molecular biology *378*, 1104-1115.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics *21*, 3369-3376.

Appendix I - Media

LB Medium

LB (Granules Fischer)
for LB agar plates 7.5 g BactoAgar (Benton Dickenson) were added per litre medium

M9 Minimal Medium

47.9 mM Na_2HPO_4 x $2H_2O$
22 mM KH_2PO_4
8.6 mM NaCl
after autoclaving supplemented with:
2 mM $MgSO_4$
0.4 % w/v glucose / $^{13}C_6$ D- glucose
0.1 mM $CaCl_2$
0.17 % w/v Yeast Nitrogenous base (YNB)
0.15 % w/v $^{15}NH_4Cl$

Antibiotic Concentrations

Carbenicillin 100 mg/ l
Chloramphenicol 25 mg/ l in 100 % ethanol
Kanamycin 50 mg/ l

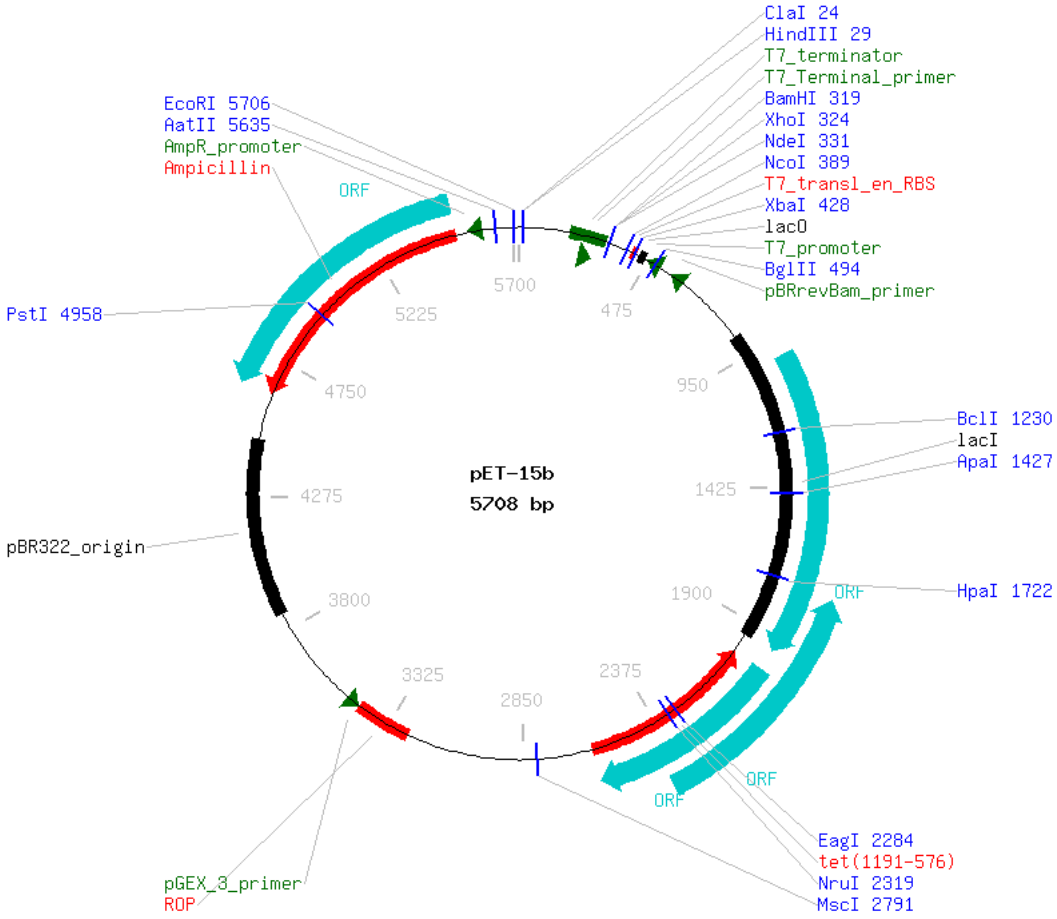
Appendix II - Expression vectors

pET-16b-3C

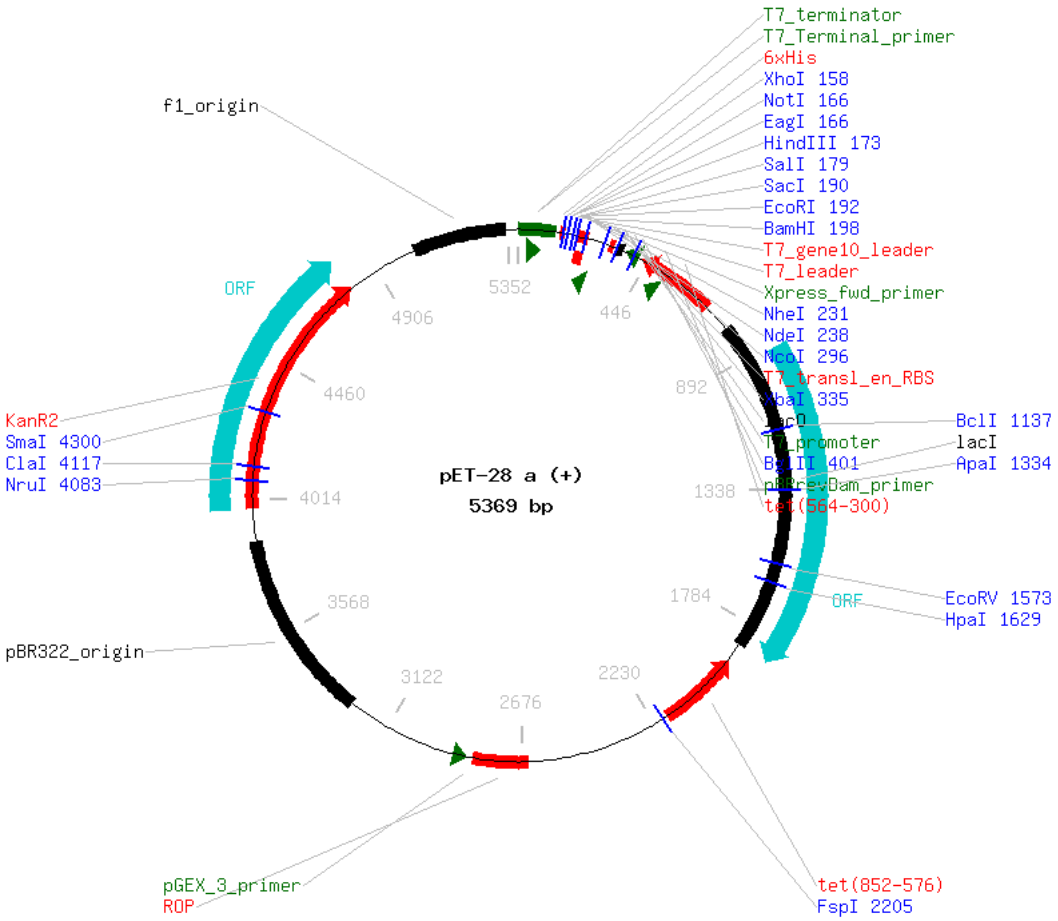
Modified from pET16b (Novagen) by insertion of pGEX-6P-2 3C cleavage site and cloning region. Only unique restriction sites are marked.

```
             ....|....| ....|....| ....|....| ....|....| ....|....|
                      10         20         30         40         50
pET-16b-3C   atctcgatcc cgcgaaatta atacgactca ctatagggga attgtgagcg

             ....|....| ....|....| ....|....| ....|....| ....|....|
                      60         70         80         90        100
pET-16b-3C   gataacaatt cccctctaga aataattttg tttaacttta agaaggagat

             ....|....| ....|....| ....|....| ....|....| ....|....|
                     110        120        130        140        150
pET-16b-3C   ataccatggg ccatcatcat catcatcatc atcatcatca cagcagcggc
                [NcoI]
                  MetGl yHisHisHis HisHisHisH isHisHisHi sSerSerGly

             ....|....| ....|....| ....|....| ....|....| ....|....|
                     160        170        180        190        200
pET-16b-3C   catatcgaag gtcgtcatat gctggaagtt ctgttccagg ggcccctggg
                               [NdeI ]                              [BamHI
             HisIleGluG lyArgHisMe tLeuGluVal LeuPheGlnG lyProLeuGl
                [ Factor Xa ]       [      3C Protease       ]

             ....|....| ....|....| ....|....| ....|....| ....|....|
                     210        220        230        240        250
pET-16b-3C   atccccagga attcccgggt cgactcgagc ggctgctaac aaagcccgaa
          BamHI]               [SalI ]
                           [SmaI]     [XhoI]
             ySerProGly IleProGlyS erThrArgAl aAlaAlaAsn LysAlaArgL

             ....|....| ....|....| ....|....| ....|....| ....|....|
                     260        270        280        290        300
pET-16b-3C   aggaagctga gttggctgct gccaccgctg agcaataact agcataaccc
             ysGluAlaGl uLeuAlaAla AlaThrAlaG luGlnEnd

             ....|....| ....|....| ....|....| ....|.
                     310        320        330
pET-16b-3C   cttggggcct ctaaacgggt cttgaggggt tttttg
```

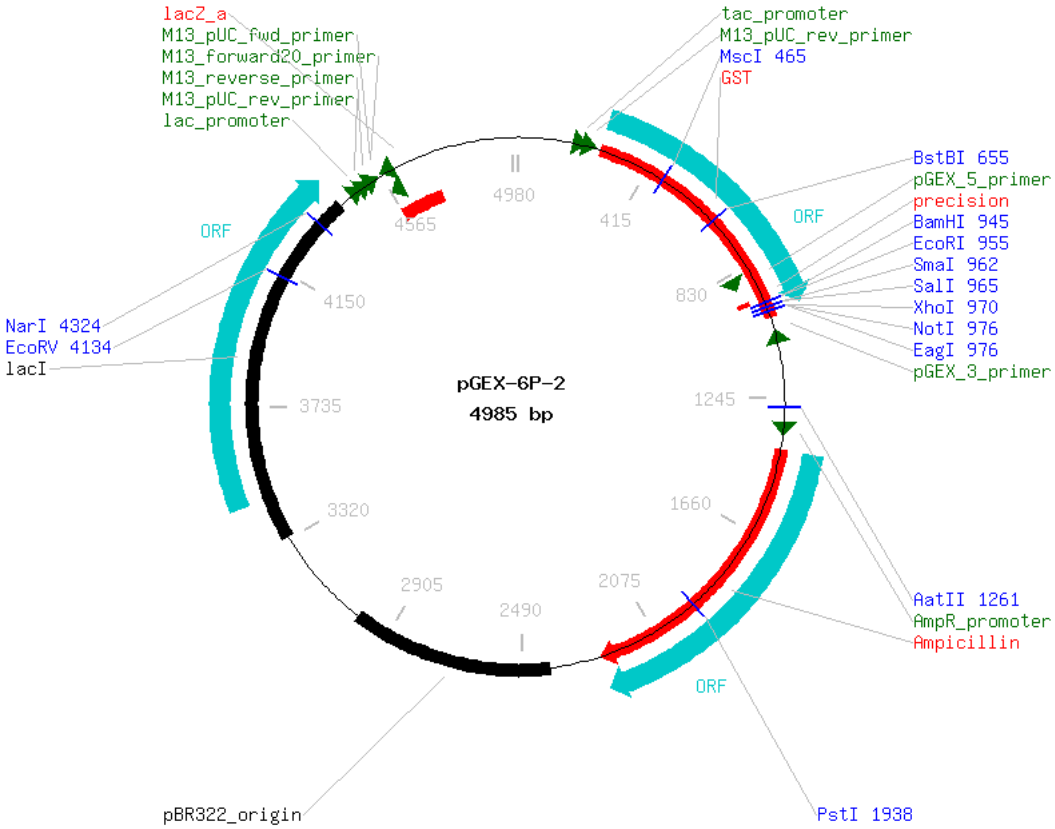
pET15b
(Novagen)



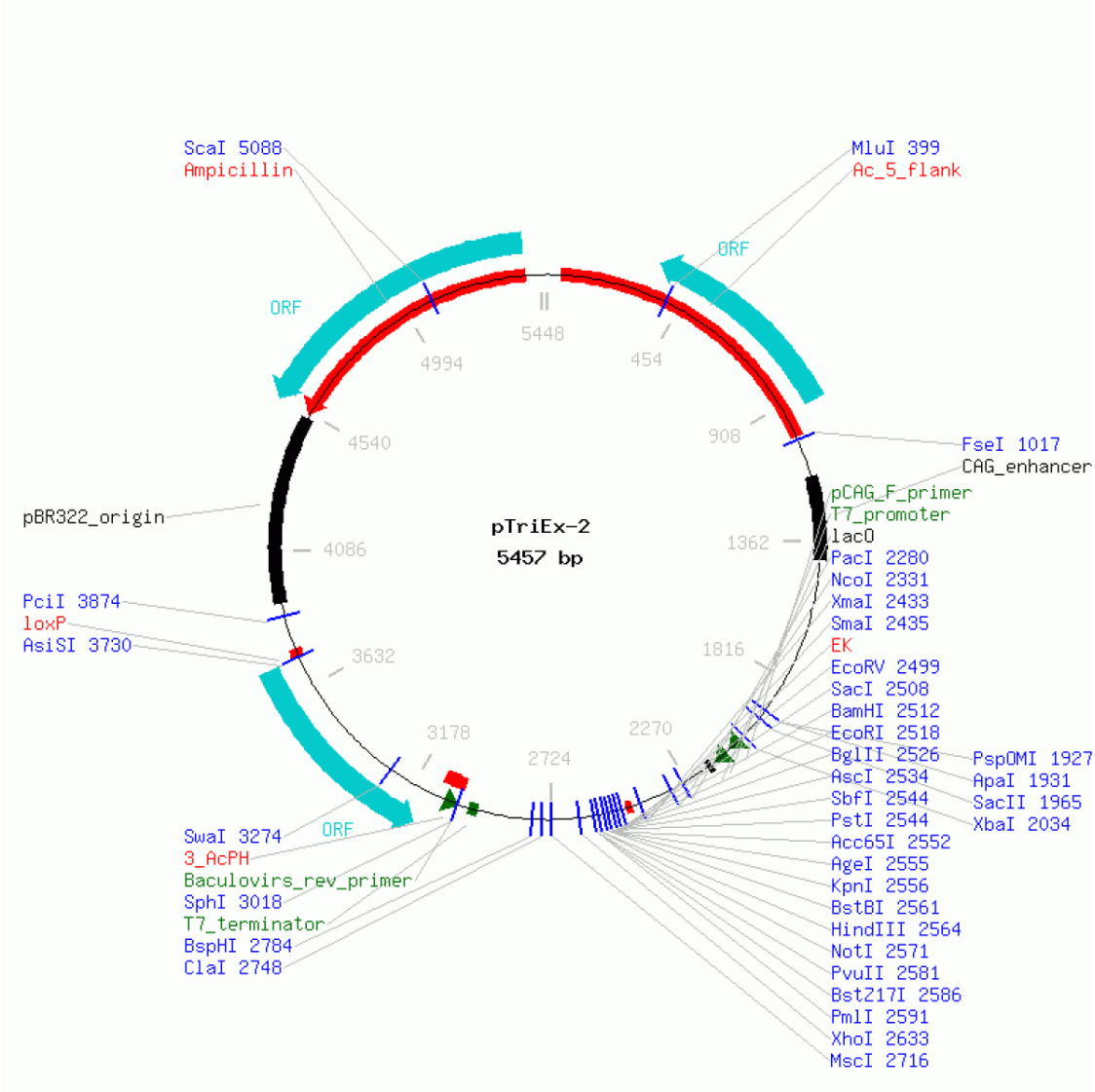
pET28a
(Novagen)



pGEX-6P-2
(Amersham Biosciences)



pTriEx-2
(Novagen)

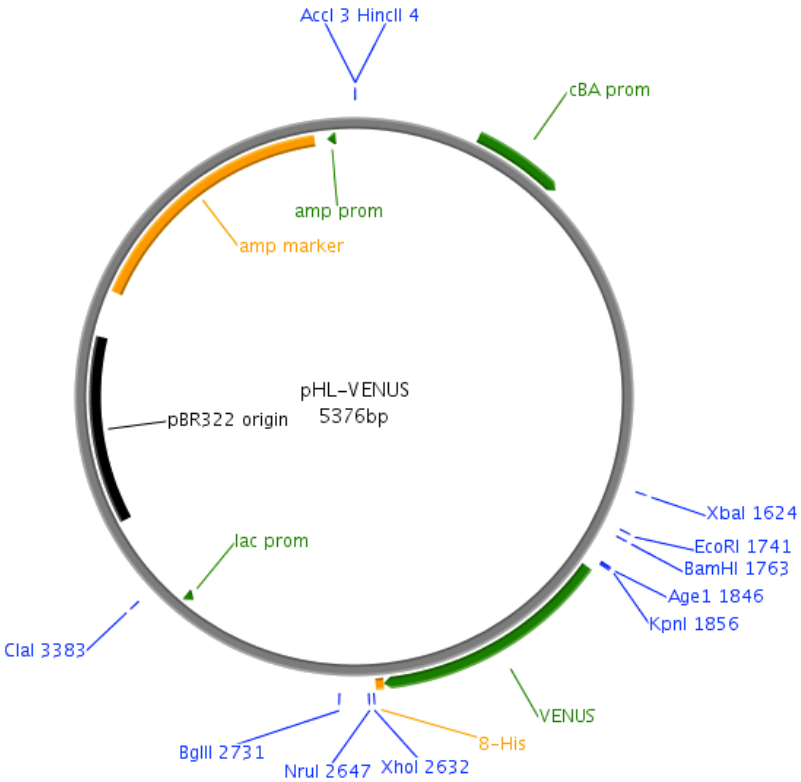


pHLsec

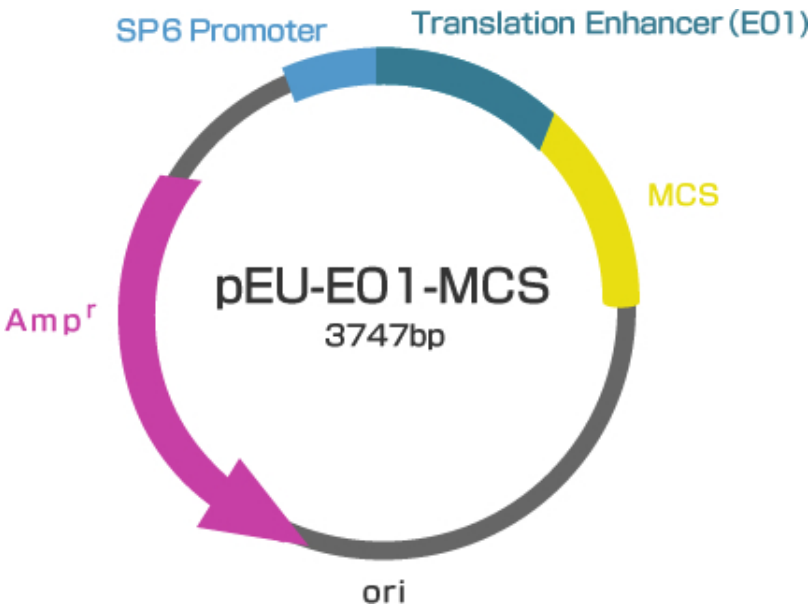
This is a modified version of pLEXm and kindly provided by Dr Aricescu (University of Oxford).

pHL-VENUS

This vector was provided by Dr Aricescu (University of Oxford).



pEU-E01-MCS
(CellFree Sciences)



Appendix III - Protein sequences

ATS-FL (PF08_0141)

KKKPKSPVDLIRVLDIHKGDYGMPTLKSKNRYIPYASDTYKGKTYIYMEGDSDSGHYYEDTTDITSSES
EYEEMDINDIYVPDSPKYKTLIEVVLEPSKRDTPSSDAPMNKFTDDEWNQLKQDFISGILENEQKDLPK
NNISGNTPMNTQPNTLYFNKPEEKPFITSIHDRDLYSGEEINYNINMSTNSMDDTSYVSNNVYSGIDLI
NDTLSGNQHIDIYDEVLKRKENELFGTNYKKNTSNNNVAKLTNSDPIMNQLDLLHKWLDRHRDMCEKWK
SKEDILHKLNEQWNKDNDGGNVPIDNRSLNTDVWIEIDMDDPKGKKEFSNMDTILDDIEDDIYYDVNDD
ENPSVDNIPMDHNKVDVPKKVHVEMKILNNTSNGSLEPEFPISDVWNI

ATS-FL (PFC1120c)

KKKTKSPVDLFSVINIPKSDYDIPTKLSPNRYIPYTSGKYRGKRYIYLEGDSGTDSGYTDHYSDITSSS
ESEYEELDINDIYAPRAPKYKTLIEVVLEPSGNNTTASGNNTTASGNNTTASGKNTPSDTQNDIQNDDI
PSSKITDNEWNTLKHDFISQYLQSEQPNDVPNDYSSGDIPLNTQPNTLYFDKPEEKPFITSIHDRNLYT
GEEYSYNINMSTNSMDDIPINSHNNIYSGIDLINDSLNSNNVDIYDEVLKRKENELFGTNHVKHTNTHN
VTKSSNSDPIDNQLDLFHTWLDRHRDMCEKWNNKEEVLNKLKEEWENETHSGNTHTSDSNKTLNTDVSI
QIDMDNPKPINQFTNMDTILEDLEKYNEPYYDVQDDIYYDVHDHDTSTVDSNNMDVPSKVQIEMDVNTK
LVKEKYPIGDVWDI

ATS-FL (PFF0845c)

KKKTKSTIDLLRVINIPKGDYGIPTMKSKNRYVPYKSGQYNGKKYIYMEGDESDDYTYIGDISSSDITS
SESEYEDIDINNIYPYKSPKYKTLIDVVLEPSKRDTFNTQSDIPSDTSTNKFTDNEWNQLKQDFISNIS
QNSQMDLPKNNISGNIQMDTHPHVNILDDSMQEKPFITSIHDRDLHNGEEVTYNINLDDHKNMNFSTNH
DNIPPKNDQNDLYTGIDLINDSISGNHNVNIYDELLKRKENELFGTNHTKHTTTNIVAKQTHNDPIVNQ
INLFHKWLDRHRNMCEQWDKNKKEELLDKLNEEWNKENKNNSNVTDTNGENNITRVLNSDVSIQIDMNS
KPI

ATS-FL (PF08_0103)

KKKTKSSVANLFQILQIPKGDYDIPTLKSSNRYIPYASDRYKGKTYIYMEGDSSGDEKYAFMSDTTDVT
SSESEYEELDINDIYVPGSPKYKTLIEVVLEPSKRDIPSGDIPHTNKFTDNEWNQLKKDFISNMLQNTQ
NTEPNVLHDNVDNNTHPTMSRHNVDQKPFIMSIHDRNLYIGEEYSYDMSTNSGENNLYSGIDPTSANHD
SYSGIDLINDALNGDYDIYDEILKRKENELFGTNHTKKNTSTNSVAKNTNSDPILNQINLFHKWLDRHR
NMCEQWDKNKKEEFLDKLKKEWNKENNNNSGDINNRYENVLNTDVSIQIDMHNPKPKNEFTNMDTNPDN
FIKDTILNDLEKHREPYFYDIYDDDITYFDTDDVKPPMDDIHIKEQTEMNALHNNKMNELLEKEYPISD
IWNI

ATS-FL (PFF0010w)

KKKTLSPVDLFSVINIPKGDYNIPTLKSSNRYIPYASDRYKGKTYIYMEGDSSGDEKYAFMSDTTDITS
SESEYEELDINDIYVPRAPKYKTLIEVVLEPSKSNGNTLGDDIPHTNKFTDEEWNELKHDFISQYVQRE
SMGVPQYDVSTELPMNIGGNVLDDGINEKPFIMSIHDRNLYSGEEYSYNVNMVNSMDDIPINRDNNVYS
GIDLINDTLSGNKHIDIYDEVLKRKENELFGTNHTKKNTSTNSVAKNTNSDPILNQINLFHTWLDRHRD
MCEKLKNDNERLAKLKEEWENETHSGNTHPSDSNKTLNTDVSIQIHMDNPKPINQFTNMDTILEDLEKY
NEPYYDVQDDIYYDVNDHDASTVDSNAVNVPSKVQIEMDVNTKLVKEKYPIADVWDI

ATS-FL (PF1055c)

KKKTKHPVDLFSVINIPKSDYDIPTKLSPNRYIPYTSGKYRGKRYIYLEGDSGTDSGYTDHYSDITSSS
ESEYEEMDINDIYVPGSPKYKTLIEVVLEPSGNNTTASDTQNDIQNDGIPSNKFSDNEWNTLKDDFISN
MLQNQPKDVPNDYKSGDIPFNTQPNTLYFDKPEEKPFITSIHDRNLLNGEEYSYNVNMSTNSMDDPKYV
SNNVYSGIDLINDSLSGNKHIDIYDEVLKRKENELFGTNHVKHTSIHSVAKNTNSDPILNQINLFHTWL
DRHRDMCEKWENHHERLAKLKEEWENETHSGNTHPSDSNKTLNTDVSIQIHMDNPKPINQFTNMDTILE
DLDKPFNEPYYYDMYDDDIYYDVNDHDTSTVDTNAMDVPSKVQIEMDVNTKLVKEKYPIADVWDI

KAHRP (PFB0100c)

MKSFKNKNTLRRKKAFPVFTKILLVSFLVWVLKCSNNCNNGNGSGDSFDFRNKRTLAQKQHEHHHHHHH
QHQHQHQAPHQAHHHHHHGEVNHQAPQVHQQVHGQDQAHHHHHHHHHQLQPQQPQGTVANPPSNEPVVK
TQVFREARPGGGFKAYEEKYESKHYKLKENVVDGKKDCDEKYEAANYAFSEECPYTVNDYSQENGPNIF
ALRKRFPLGMNDEDEEGKEALAIKDKLPGGLDEYQNQLYGICNETCTTCGPAAIDYVPADAPNGYAYGG
SAHDGSHGNLRGHDNKGSEGYGYEAPYNPGFNGAPGSNGMQNYVPPHGAGYSAPYGVPHGAAHGSRYSS
FSSVNKYGKHGDEKHHSSKKHEGNDGEGEKKKSKKHKDHDGEKKKSKKHKDNEDAESVKSKKHKSHDC
EKKKSKKHKDNEDAESVKSKKSVKEKGEKHNGKKPCSKKTNEENKNKEKTNNSKSDGSKAHEKKENETK
NTAGENKKVDSTSADNKSTNAATPGAKDKTQGGKTDKTGASTNAATNKGQCAAEGATKGATKEASTSKE
ATKEASTSKEATKEASTSKEATKEASTSKGATKEASTTEGATKGASTTAGSTTGATTGANAVQSKDETA
DKNAANNGEQVMSRGQAQLQEAGKKKKKRGCCG

PFI1780w

MAVSTYNNTRRNGLRYVLKRRTILSVFAVICMLSLNLSIFENNNNNYGFHCNKRHFKSLAEASPEEHNN
LRSHSTSDPKKNEEKSLSDEINKCDMKKYTAEEINEMINSSNEFINRNDMNIIFSYVHESEREKFKKVE
ENIFKFIQSIVETYKIPDEYKMRKFKFAHFEMQGYALKQEKFLLEYAFLSLNGKLCERKKFKEVLEYVK
REWIEFRKSMFDVWKEKLASEFREHGEMLNQKRKLKQHELDRRAQREKMLEEHSRGIFAKGYLGEVESE
TIKKKTEHHENVNEDNVEKPKLQQHKVQPPKVQQQKVQPPKSQQQKVQPPKSQQQKVQPPKVQQQKVQP
PKVQKPKLQNQKGQKQVSPKAKGNNQAKPTKGNKLKKN

*Pf*PLP1 (PFD0430c)

MKLRILKKHYYVVFILLYLYDISCFKCIRLNNRSIYKNKYKNNVHIGTNENIRSIEKYSNVLCNSILCK
NDKISSFINQRKNVDDDDESENDDMYESTTAGSSSETDNESDEEENDSSDNNNSDEEQIENSNNNNSDE
EQNDSSSNDNNDEENEEQDDVMDNDQNDKKIKHSFNLANESKHTKEERVKEEKKLKIYDFINDKEKRLN
FNGDQKDEDNEENDDKDENTLENRNIISKHTSVFPGLYFIGIGYNLLFGNPLGEADSLIDPGYRAQIYL
MEWALSKEGIANDLSTLQPVNGWIRKENACSRVESITECSSISDYTKSLSAEAKVSGSYWGIASFSAST
GYSSFLHEVTKRSKKTFLVKSNCVKYTIGLPPYIPWDKTTAYKNAVNELPAVFTGLDKESECPSDVYEE
NKTKSNCENVSLWMKFFDIYGTHIIYESQLGGKITKIINVSTSSIEQMKKNGVSVKAKIQAQFGFGSAG
GSTDVNSSNSSANDEQSYDMNEQLIVIGGNPIKDVTKEENLFEWSKTVTNHPMPININKLTPISDSFDSD
DLKESYDKAIIYYSRLYGLSPHDTMQKDDKDIIKILTNADTVTKNSAPPINAQCPHGKVVMFGFSLKQN
FWDNTNALKGYNIEVCEAGSNSCTSKQGSSNKYDTSYLYMECGDQPLPFSEQVISESTSTYNTVKCPND
YSILLGFGISSSSGRINSAEYVYSTPCIPGMKSCSLNMNNDNQKSYIYVLCVDTTIWSGVNNLSLVALD
GAHGKVNRSKKYSDGELVGTCPLDGTVLTGFKVEFHTSSPYVQTPFEKCAKSLKACSVHGSGHAIGIQN
FKSLFIYMLCKNNK

*Pf*PLP2 (PFL0805c)

MNILLKYHLSYIFFFVHLIVINIKCKEINNNLVNDEKDSINNLEEPSPFHILKKKNSNNIENHSNKINN
IFKNNYSYSYSYSLLDNSLYDNKYESDFFPQKKNNKINEEKYKGTHGTNFAIINEHMNEIHNDMLGQNE
ILSNELDTTKDIIKEEHVKNKKIKNHSFYMNKENEKKKFLDNTFPENKENNKNKRNSKFIEENDLNSSG
ENYFENDKFIEIEYPIKNKDIDEYEEGINRNHKEEDNKNRTYHSNGEDSIKSLNFIKYMNMDKEKKGKK
KSYNRKKKENYNNDNYNINDSDLINVLKQGDYSPDCDTCNKNINGHMNKGRGIRRRMYDRNYYKDFNLR
DALPINKTHYSMLLKSDDDTIIKEKNENDVDDIITKEKRKENLKEVYNDKKNNYLSLKYLGLGYDIIMG
NPEGDPTLNVDPGFRGPVLQINLKEISLNSYDNHKNDDNHMNDDIHKNDDNHMNSDNHKNDEYHKNDEY
HKNDEDHNDDNYDNHADGDKRSKERKSKIQTTNESMKPWVIPEHSCSQSKNVEEIRNLEQYKLELLSDV
KVSTPSSFPYSFSASAEFKNALKKLKVQNNVIFLMKIYCLRYYTGIPITTTSYKFSENFKNALSKLPKY
FDGLREDSKCSYEYYINKLNSPECEENVNKWMLFFKLHGTHVAYEIYLGGKIIIKININKEEYNKMKEN
NINVKTFFNIYFHKMGLSSAFQKEAQKILNKFRISKHIAILGGNPGLNVNNTSFFEKWVHSINTNSMPI
RTKLLPFSFFMDDNDMIQAYKDALIFYGLTYGLQLFGQEKYNHNIISVGEYLEKCTQKLYAGPPPGLLT
CPIGTTILMGFSINLNFYKNKYLSSTNGITLCEPMKESCSGNGFEKNYSDIRIFALCTNKPFDFITQVV
QQGEAPKISASCPGELVILFGFALMKGIGSSSANKIDIYPCRTGQNSCEAVLQNHKFKQSMIYLACVDK
TTNGLEYLQTYSKTKNLGDVISDKYKSDGYLNFSCPQNNTLVFGFSLEFHTNFQATRNNFLNCSKYTNI
CEISGIGINTHLSFFRSDKHSLAIIALCRSHGANNLVE

*Pf*PLP3 (PFI1145w)

MTFTRGSFIFLYILIIILNTIYQNIFDNIDNVSFISKIFSSKIFVSTKGNDYIENTVENIPQIENGKIL
NILQHDQIINKVNHNTIHTGALYTQHNKKSKNRKTIKSSQRKSHPNLVNVKNNINFNEIKNGHETNTND
SLFKTEEIKKIEATEKKECKGSAKNCIRGNVKSLINVMKGMEINKNEKDNSTLDREENIYDDEHHTVIV
FDGIDENKDLYNKYKNIEKNKNGMDSIVIQGTEYLGVGYDFIFGNPIGDPFLKVDPGYRDSIIKLTYPK
SDEDYPDNYMNINPNGSFVRNEISCNRSEKESEISTMSEYTKELSVDASIGASYGLFGSFSASTGYKSV
SNTISKNKFRMFMLKSYCFKYVASLSQYSQWKLTDQFVRAISLLPSHFNSLEKDGTYCSDEEFRDNRKS

EKCGKSVTAWMYFFKNFGTHVSTLLHLGGKITQQVKISKNDYKAMTESGLSISASVSAGFGLFKVKGST
NTESNESSNNESSTSSLEKETVIIGGTTIFDPNDPNNFEKWAESISENPMPIKGEYEPLSRILPTRLSK
IYEEALRFYISVNVPSNFGQITDNEIRQYNIKEELMKATMIHSSGTGLVVVECEEKQNFLLGFSLSIPN
DLSNLKDFYLNSCDEDSDKCYSKMSDNAYSYIFAMCKEEMIPFFEQKVKSGVGLLTLECSEKNQVILFG
FGISVLNTNDPISISLYPCKYGKASCSMQGSTDQSAVGLWIVCAHEESLNSKFSVYIRKMFEENVSGKK
KKHMDICPEKVLFNLIFEFTKTPIKKRNGGCFTVNDKCPKDFHVCSEKEDRKSFNYYSLSVY

*Pf*PLP4 (PF08_0050)

MMIFSNYLFNILIVCYLFIYIGSFSICHDIKTDEKTNEIFSKYLGKGYDILFGYPLPNNELIDDPGFKE
VIIDTQLSIDNISNYICKKEEYVDVIEDINDIGYLGMQKINIDDLDNRIKPFSASMPYKSYFADLEIKK
KKYALAQNMCVLNYATYDLKESGNNINKDFVLDIEKLPILTKNQMKLCTKVLYMNNNLHCSEGIKSWMK
FFEKYGTHVVLSAHFGGMSFNTMEITKRKIEEIKIYKYKYSLWNNPYLNIFKSGSLFQDLSINVDGHKE
NKKNNSNNNININIDEKKKNDAYIKNDVLIEQYRDNINLEIRGGNNFDEKWRNLTYLVWKNSIYSNIVPIH
LDLYSLNTFMPIEKKESYDMALLFYNNLYGIDNENFYLSQDITDVLSEGKQITGSNKGSLILSCPVGYI
KSTGLIFVYDSSEKLKTNKSKDPKIKIYPCTNKGQYDIACSYITKKKNIITFGWIYCVKHNFIKFETLY
QNNDNIISKDGKTFMSLTCTEGNTIAFGFKMKLKKLEKLEKLKIKPCTIGDDQCTINNIKRNSDYLLWG
FCVPSSFRSLSSLQLTYIHDDTIQNDIRGACSDIYINKYDNIFLGFTFSFDNKFEQVKTEPCESQSKYC
FHIIPKKKKKMYYVGMFLLCRFDGNTNRKFTYK

*Pf*PLP5 (PF08_0052)

MKIKKSCIFFGYAFALFCTVCCFSKKDIFNPKLYTSDVEYKEREYFLKYLGMSYDIIKGNPWGDPIYVI
DLGYRRNVLKKRKLNSDNNIKDDIVKFKVGEASKIKCIDTIKENVIDNLCDINKEYERSYSVSSINDDI
HPFNDSNYYKMLVKRINRGDSIIIEKKLCSKYFSFINDINKNDLDTFFLTTLNELGDNYQNIKDDTYKC
SLQYYKMNNMNKYSENCLKTITPWISFFNMYGTHVISGVYYGGKIIHNLYFENNNLKKKEYKIRMYKSR
LNPFSTINSNLYFGSSLSKEKIIYIRERNLIMDGGVHINPYNINEVNMENKKKNIYVNNVEKNLYDQKK
KYRNYYNFYELKDDVRKRNYYNSWKDTIEWEQAKPVKLNLVPLSEFINSEEGKSAYYMALEFYSNLSYS
NYNPYLYVLNKSEKDIYIMDIKRNWAQYIDKNINFNVTPKCKNGDKILSGFILTNKKKSYEDNHIIHMC
PSNTVCSSGINIESDKNFEFSWILCSKENRSEIHQILTKNTFQGNGKASCPYNMKIGFGFSLTFQKSIN
TNIKIEPCESNKKECKRTNLASSSQTYFWINCLPTNKNLLLQTLESKTYSEKKYLNDHFVFSLKCSEGK
YIIAGFAIDYIPSGVNDYLVCPVGSNKCDLKIHVVEKNIGEVHIPIMYIVCSSIS

*Pb*PLP3 (PBANKA_082420)

MMFKRKYLFFPCFFLIIVNILHKLDNLGADESLLYFHIRILSMNKNDTQELYNNKVKIFNNEILSGLQH
EHTIEDERIRKFVSNDADSKHGRRIKKVTLKLPKNRVDRMQNIRKINFNEMRQNDDANPEINKPNITHE
TYPALEPSKHNKNKQMSNVKNKVNNFVKLMKDINDNMENEKEKERNNPNTEDNIYNDEHHTVVIYDGLD
ENKDLYNKYKIETENNMDTKIIQGTEYLGVGYDFLFGNPIGDPFLKVDPGYRDSVIKLTYPKSDVDYPD
NYININPNGSYVRNEISCNRSENETEISTMSEYTKELSVDASLGGSYGLFGSFSASVGYTSASNTISKK
KFRMFILKSYCFKYVASLSQYSQWKLSDQFLRAIALLPSYFNSLEHDGKYCNAEELRYNKTGMENCGKS
VESWLYFFKNFGTHVSTVIHLGGKITQQVKISKNEYKSLNESGLSTSVSASVGFGLFKANASTSTDSKE
SSHEESSNSSIEKETVIIGGTTIYDPNDPSNFEKWADSIKNNPMPIKGQYEPLSRILPERLNKIYDEAL
SFYVSLNVPINIGSKTDGEIKNYNIKDQLMKSKMIHGSGSGLVVLECEDKQNFILGFSLSIPNDLSNLK
DFYMNTCDEESDKCYSKMSDNAYSYIFAMCNDDSIPYLEQKAKSGTGLLTLECSEKNQIILFGFGISVL
NSNDPISIAIYPCKNGKTSCSMQGSTDQSAVGLWIVCGHEESLNSNFSVYVRKLEKDNVSGKKKKKIDI
CPTEVLFNLIFEFTKTPINKRNGKCFTVNDICPKDFHVCSDNKGIKSFNYYSVSVY

Appendix IV - Chemical Shifts

Chemical shift assignments have been deposited in the BMRB under accession numbers 1611 (ATS-core) and 17999 (ATS-FL).

ATS-core

2	PRO	HA	4.47	6	MET	CA	55.6	10	THR	H	9.21
2	PRO	HB2	1.94	6	MET	CB	32.77	10	THR	HA	4.52
2	PRO	HB3	2.31	6	MET	CE	17.21	10	THR	HB	4.81
2	PRO	HD2	3.55	6	MET	CG	32.07	10	THR	HG2	1.42
2	PRO	HD3	3.55	6	MET	N	121.6	10	THR	C	175.68
2	PRO	HG2	2	7	ASN	H	8.36	10	THR	CA	61.36
2	PRO	HG3	2	7	ASN	HA	4.65	10	THR	CB	70.66
2	PRO	C	177.09	7	ASN	HB2	2.64	10	THR	CG2	22.02
2	PRO	CA	63.16	7	ASN	HB3	2.79	10	THR	N	114.98
2	PRO	CB	32.33	7	ASN	HD21	7.56	11	ASP	H	9.16
2	PRO	CD	49.65	7	ASN	HD22	6.85	11	ASP	HA	4.57
2	PRO	CG	27.07	7	ASN	C	174.07	11	ASP	HB2	2.76
3	LEU	H	8.51	7	ASN	CA	53.27	11	ASP	HB3	2.86
3	LEU	HA	4.36	7	ASN	CB	39.28	11	ASP	C	178.68
3	LEU	HB2	1.61	7	ASN	N	119.57	11	ASP	CA	58.43
3	LEU	HB3	1.67	7	ASN	ND2	112.96	11	ASP	CB	39.41
3	LEU	HD1	0.93	8	LYS	H	7.97	11	ASP	N	122.26
3	LEU	HD2	0.89	8	LYS	HA	4.18	12	ASP	H	8.49
3	LEU	HG	1.66	8	LYS	HB2	1.58	12	ASP	HA	4.55
3	LEU	C	177.97	8	LYS	HB3	1.71	12	ASP	HB2	2.63
3	LEU	CA	55.45	8	LYS	HD2	1.59	12	ASP	HB3	2.67
3	LEU	CB	42.28	8	LYS	HD3	1.59	12	ASP	C	179.14
3	LEU	CD1	24.91	8	LYS	HE2	2.93	12	ASP	CA	57.56
3	LEU	CD2	23.55	8	LYS	HE3	2.93	12	ASP	CB	40.71
3	LEU	CG	27.06	8	LYS	HG2	1.26	12	ASP	N	118.86
3	LEU	N	122.31	8	LYS	HG3	1.3	13	GLU	H	7.96
4	GLY	H	8.41	8	LYS	C	176.67	13	GLU	HA	4.03
4	GLY	HA2	3.98	8	LYS	CA	55.66	13	GLU	HB2	2.52
4	GLY	HA3	3.98	8	LYS	CB	33.81	13	GLU	HB3	2.05
4	GLY	CA	45.34	8	LYS	CD	28.77	13	GLU	HG2	2.32
4	GLY	N	109.87	8	LYS	CE	42.02	13	GLU	HG3	2.45
5	SER	H	8.2	8	LYS	CG	24.4	13	GLU	C	179.96
5	SER	HA	4.43	8	LYS	N	119.64	13	GLU	CA	59.16
5	SER	HB2	3.86	9	PHE	H	8.78	13	GLU	CB	30.53
5	SER	HB3	3.86	9	PHE	HA	5.06	13	GLU	CG	37.16
5	SER	CA	58.38	9	PHE	HB2	2.95	13	GLU	N	121.47
5	SER	CB	63.92	9	PHE	HB3	3.54	14	TRP	H	8.93
5	SER	N	115.59	9	PHE	HD1	7.36	14	TRP	HA	4.61
6	MET	H	8.41	9	PHE	HD2	7.36	14	TRP	HB2	3.27
6	MET	HA	4.47	9	PHE	C	178.62	14	TRP	HB3	3.41
6	MET	HB2	1.99	9	PHE	CA	57.01	14	TRP	HD1	7.34
6	MET	HB3	1.99	9	PHE	CB	40.31	14	TRP	HE1	10.18
6	MET	HE	2.08	9	PHE	CD1	131.66	14	TRP	HE3	7.07
6	MET	HG2	2.5	9	PHE	CD2	131.66	14	TRP	HH2	6.94
6	MET	HG3	2.58	9	PHE	N	121.35	14	TRP	HZ2	7.39

14	TRP	HZ3	6.88	18	LYS	HB3	0.98	21	PHE	N	122.51
14	TRP	C	176.68	18	LYS	HD2	1.93	22	ILE	H	8.79
14	TRP	CA	58.75	18	LYS	HD3	1.93	22	ILE	HA	3.11
14	TRP	CB	30.2	18	LYS	HE2	2.5	22	ILE	HB	1.74
14	TRP	CD1	126.85	18	LYS	HE3	2.98	22	ILE	HD1	0.31
14	TRP	CE3	121.62	18	LYS	HG2	0.26	22	ILE	HG12	0.69
14	TRP	CH2	123.61	18	LYS	HG3	0.26	22	ILE	HG13	1.82
14	TRP	CZ2	114.43	18	LYS	C	178.38	22	ILE	HG2	0.76
14	TRP	CZ3	121.7	18	LYS	CA	61.63	22	ILE	C	177.33
14	TRP	N	121.6	18	LYS	CB	32.11	22	ILE	CA	65.91
14	TRP	NE1	126.62	18	LYS	CD	32.05	22	ILE	CB	38.3
15	ASN	H	8.65	18	LYS	CE	41.4	22	ILE	CD1	13
15	ASN	HA	4.11	18	LYS	CG	24.57	22	ILE	CG1	30.82
15	ASN	HB2	2.99	18	LYS	N	116.12	22	ILE	CG2	17.11
15	ASN	HB3	2.82	19	GLN	H	7.7	22	ILE	N	121.04
15	ASN	HD21	7.76	19	GLN	HA	3.85	23	SER	H	8.43
15	ASN	HD22	7.01	19	GLN	HB2	2.16	23	SER	HA	4.02
15	ASN	C	178.5	19	GLN	HB3	2.16	23	SER	HB2	3.9
15	ASN	CA	56.55	19	GLN	HE21	7.55	23	SER	HB3	3.9
15	ASN	CB	37.77	19	GLN	HE22	6.86	23	SER	C	176.58
15	ASN	N	116.89	19	GLN	HG2	2.46	23	SER	CA	61.87
15	ASN	ND2	112.33	19	GLN	HG3	2.46	23	SER	CB	62.83
16	GLN	H	7.81	19	GLN	C	178.72	23	SER	N	114.24
16	GLN	HA	3.93	19	GLN	CA	58.8	24	GLY	H	7.33
16	GLN	HB2	2.15	19	GLN	CB	27.85	24	GLY	HA2	3.67
16	GLN	HB3	2.15	19	GLN	CG	33.69	24	GLY	HA3	3.83
16	GLN	HE21	7.63	19	GLN	N	117.08	24	GLY	C	175.59
16	GLN	HE22	6.83	19	GLN	NE2	112.32	24	GLY	CA	46.55
16	GLN	HG2	2.4	20	ASP	H	8.3	24	GLY	N	106.64
16	GLN	HG3	2.4	20	ASP	HA	4.34	25	ILE	H	7.3
16	GLN	C	177.25	20	ASP	HB2	2.95	25	ILE	HA	3.77
16	GLN	CA	58.71	20	ASP	HB3	2.57	25	ILE	HB	1.48
16	GLN	CB	28.36	20	ASP	C	179.31	25	ILE	HD1	-0.17
16	GLN	CG	33.9	20	ASP	CA	57.59	25	ILE	HG2	0.46
16	GLN	N	119.76	20	ASP	CB	40.2	25	ILE	HG12	0.46
16	GLN	NE2	115.05	20	ASP	N	121.97	25	ILE	HG13	0.46
17	LEU	H	7.76	21	PHE	H	8.5	25	ILE	C	178.2
17	LEU	HA	4.04	21	PHE	HA	3.99	25	ILE	CA	61.79
17	LEU	HB2	1.36	21	PHE	HB2	3.08	25	ILE	CB	36.82
17	LEU	HB3	1.71	21	PHE	HB3	2.79	25	ILE	CD1	10.59
17	LEU	HD1	0.82	21	PHE	HD1	6.88	25	ILE	CG1	27.23
17	LEU	HD2	0.78	21	PHE	HD2	6.88	25	ILE	CG2	17.87
17	LEU	HG	1.6	21	PHE	HE1	6.6	25	ILE	N	120.79
17	LEU	C	179.4	21	PHE	HE2	6.6	26	LEU	H	8
17	LEU	CA	58.38	21	PHE	HZ	6.41	26	LEU	HA	4.08
17	LEU	CB	41.59	21	PHE	C	177.9	26	LEU	HB2	1.47
17	LEU	CD1	24.44	21	PHE	CA	61.6	26	LEU	HB3	1.7
17	LEU	CD2	25.75	21	PHE	CB	38.69	26	LEU	HD1	0.72
17	LEU	CG	27.79	21	PHE	CD1	131.72	26	LEU	HD2	0.76
17	LEU	N	122.53	21	PHE	CD2	131.72	26	LEU	HG	1.67
18	LYS	H	8.04	21	PHE	CE1	131.26	26	LEU	C	178.76
18	LYS	HA	3.36	21	PHE	CE2	131.26	26	LEU	CA	56.55
18	LYS	HB2	1.94	21	PHE	CZ	130.53	26	LEU	CB	41.45

26	LEU	CD1	22.54	31	LYS	HB2	1.83	35	ALA	CB	18.73
26	LEU	CD2	25.61	31	LYS	HB3	1.83	35	ALA	N	124.25
26	LEU	CG	26.68	31	LYS	HD2	1.65	36	LYS	H	7.86
26	LEU	N	119.6	31	LYS	HD3	1.65	36	LYS	HA	4.2
27	GLU	H	7.75	31	LYS	HE2	2.96	36	LYS	HB2	1.75
27	GLU	HA	4.16	31	LYS	HE3	2.96	36	LYS	HB3	1.75
27	GLU	HB2	2.03	31	LYS	HG2	1.43	36	LYS	HD2	1.4
27	GLU	HB3	2.03	31	LYS	HG3	1.43	36	LYS	HD3	1.4
27	GLU	HG2	2.24	31	LYS	C	176.98	36	LYS	HE2	2.69
27	GLU	HG3	2.31	31	LYS	CA	57.63	36	LYS	HE3	2.69
27	GLU	C	177.19	31	LYS	CB	32.73	36	LYS	HG2	1.28
27	GLU	CA	57.7	31	LYS	CD	29.1	36	LYS	HG3	1.36
27	GLU	CB	29.76	31	LYS	CE	42.2	36	LYS	C	177.33
27	GLU	CG	36.23	31	LYS	CG	24.73	36	LYS	CA	57.23
27	GLU	N	118.24	31	LYS	N	120.76	36	LYS	CB	32.89
28	ASN	H	7.76	32	ASP	H	8.25	36	LYS	CD	28.97
28	ASN	HA	4.63	32	ASP	HA	4.58	36	LYS	CE	42.01
28	ASN	HB2	2.8	32	ASP	HB2	2.65	36	LYS	CG	24.82
28	ASN	HB3	2.8	32	ASP	HB3	2.65	36	LYS	N	117.75
28	ASN	HD21	7.6	32	ASP	C	176.86	37	LEU	H	8
28	ASN	HD22	6.92	32	ASP	CA	55.06	37	LEU	HA	4.41
28	ASN	C	175.75	32	ASP	CB	40.76	37	LEU	HB2	1.61
28	ASN	CA	54.07	32	ASP	N	119.47	37	LEU	HB3	1.78
28	ASN	CB	39.34	33	LEU	H	7.97	37	LEU	HD1	0.87
28	ASN	N	117.29	33	LEU	HA	4.21	37	LEU	HD2	0.92
28	ASN	ND2	113.46	33	LEU	HB2	1.72	37	LEU	HG	1.72
29	GLU	H	8.33	33	LEU	HB3	1.59	37	LEU	C	177.68
29	GLU	HA	4.21	33	LEU	HD1	0.9	37	LEU	CA	55.45
29	GLU	HB2	2.02	33	LEU	HD2	0.83	37	LEU	CB	42.35
29	GLU	HB3	2.13	33	LEU	HG	1.64	37	LEU	CD1	23.23
29	GLU	HG2	2.25	33	LEU	C	178.27	37	LEU	CD2	25.46
29	GLU	HG3	2.32	33	LEU	CA	56.82	37	LEU	CG	26.93
29	GLU	C	177.17	33	LEU	CB	42.17	37	LEU	N	120.05
29	GLU	CA	57.68	33	LEU	CD1	25.08	38	THR	H	8.01
29	GLU	CB	29.94	33	LEU	CD2	23.9	38	THR	HA	4.32
29	GLU	CG	36.52	33	LEU	CG	27.02	38	THR	HB	4.29
29	GLU	N	120.51	33	LEU	N	121.47	38	THR	HG2	1.22
30	GLN	H	8.3	34	VAL	H	7.94	38	THR	C	174.51
30	GLN	HA	4.27	34	VAL	HA	3.88	38	THR	CA	62.32
30	GLN	HB2	2.07	34	VAL	HB	2.14	38	THR	CB	69.87
30	GLN	HB3	2.18	34	VAL	HG1	1	38	THR	CG2	21.56
30	GLN	HE21	7.56	34	VAL	HG2	0.94	38	THR	N	113
30	GLN	HE22	6.84	34	VAL	C	177.13	39	ASN	H	8.34
30	GLN	HG2	2.41	34	VAL	CA	64.23	39	ASN	HA	4.72
30	GLN	HG3	2.41	34	VAL	CB	32.02	39	ASN	HB2	2.86
30	GLN	C	176.74	34	VAL	CG1	21.75	39	ASN	HB3	2.86
30	GLN	CA	56.61	34	VAL	CG2	21.32	39	ASN	HD21	7.61
30	GLN	CB	28.79	34	VAL	N	118.26	39	ASN	HD22	6.93
30	GLN	CG	34.03	35	ALA	H	7.96	39	ASN	CA	53.75
30	GLN	N	119.92	35	ALA	HA	4.21	39	ASN	CB	38.96
30	GLN	NE2	112.44	35	ALA	HB	1.42	39	ASN	N	120.23
31	LYS	H	8.13	35	ALA	C	178.53	39	ASN	ND2	112.84
31	LYS	HA	4.17	35	ALA	CA	53.66	40	SER	H	8.16

40	SER	HA	4.43	45	ASN	H	8.35	49	LEU	HG	1.89
40	SER	HB2	3.88	45	ASN	HA	4.75	49	LEU	C	180.3
40	SER	HB3	3.88	45	ASN	HB2	2.95	49	LEU	CA	58.2
40	SER	C	173.94	45	ASN	HB3	2.95	49	LEU	CB	41.97
40	SER	CA	58.6	45	ASN	HD21	7.18	49	LEU	CD1	25.77
40	SER	CB	63.9	45	ASN	HD22	6.94	49	LEU	CD2	23.68
40	SER	N	114.84	45	ASN	C	176.05	49	LEU	CG	27.13
41	ASP	H	8.11	45	ASN	CA	53.74	49	LEU	N	119.57
41	ASP	HA	4.89	45	ASN	CB	38.96	50	LEU	H	8.28
41	ASP	HB2	2.6	45	ASN	N	119.07	50	LEU	HA	4.23
41	ASP	HB3	2.78	45	ASN	ND2	112.79	50	LEU	HB2	1.4
41	ASP	CA	52.94	46	GLN	H	8.42	50	LEU	HB3	2
41	ASP	CB	41.66	46	GLN	HA	4.22	50	LEU	HD1	0.75
41	ASP	N	123.39	46	GLN	HB2	2.14	50	LEU	HD2	0.6
42	PRO	HA	4.47	46	GLN	HB3	2.14	50	LEU	HG	1.71
42	PRO	HB2	1.98	46	GLN	HE21	7.36	50	LEU	C	179.16
42	PRO	HB3	2.35	46	GLN	HE22	6.89	50	LEU	CA	58.71
42	PRO	HD2	3.81	46	GLN	HG2	2.39	50	LEU	CB	42.11
42	PRO	HD3	3.88	46	GLN	HG3	2.39	50	LEU	CD1	25.42
42	PRO	HG2	2.04	46	GLN	C	177.44	50	LEU	CD2	25.26
42	PRO	HG3	2.04	46	GLN	CA	58.5	50	LEU	CG	27.71
42	PRO	C	177.75	46	GLN	CB	29.05	50	LEU	N	122.35
42	PRO	CA	63.99	46	GLN	CG	34	51	HIS	H	8.55
42	PRO	CB	32.21	46	GLN	N	120.74	51	HIS	HA	4.13
42	PRO	CD	50.83	46	GLN	NE2	110.91	51	HIS	HB2	3.13
42	PRO	CG	27.41	47	LEU	H	8.26	51	HIS	HB3	3.28
43	ILE	H	8.38	47	LEU	HA	4.17	51	HIS	HD2	7.1
43	ILE	HA	4.12	47	LEU	HB2	1.89	51	HIS	HE1	7.67
43	ILE	HB	1.97	47	LEU	HB3	1.73	51	HIS	C	177.86
43	ILE	HD1	0.88	47	LEU	HD1	0.94	51	HIS	CA	62.68
43	ILE	HG12	1.28	47	LEU	HD2	1.03	51	HIS	CB	31.5
43	ILE	HG13	1.48	47	LEU	HG	1.85	51	HIS	CD2	117.56
43	ILE	HG2	0.9	47	LEU	C	178.71	51	HIS	CE1	138.13
43	ILE	C	176.96	47	LEU	CA	58.17	51	HIS	N	118.41
43	ILE	CA	62.04	47	LEU	CB	41.45	52	LYS	H	8.14
43	ILE	CB	38.14	47	LEU	CD1	24.08	52	LYS	HA	4.26
43	ILE	CD1	12.98	47	LEU	CD2	24.86	52	LYS	HB2	2.08
43	ILE	CG1	28.13	47	LEU	CG	27.3	52	LYS	HB3	2.08
43	ILE	CG2	17.86	47	LEU	N	119.96	52	LYS	HD2	1.82
43	ILE	N	119.91	48	ASP	H	8.14	52	LYS	HD3	1.82
44	MET	H	8.14	48	ASP	HA	4.54	52	LYS	HE2	3.06
44	MET	HA	4.43	48	ASP	HB2	2.76	52	LYS	HE3	3.06
44	MET	HB2	2.08	48	ASP	HB3	2.76	52	LYS	HG2	1.59
44	MET	HB3	2.08	48	ASP	C	179.28	52	LYS	HG3	1.75
44	MET	HE	2.1	48	ASP	CA	57.66	52	LYS	C	179.53
44	MET	HG2	2.58	48	ASP	CB	40.86	52	LYS	CA	59.88
44	MET	HG3	2.67	48	ASP	N	118.92	52	LYS	CB	32.73
44	MET	C	176.5	49	LEU	H	7.9	52	LYS	CD	29.39
44	MET	CA	56.65	49	LEU	HA	4.21	52	LYS	CE	42.26
44	MET	CB	32.76	49	LEU	HB2	2.05	52	LYS	CG	25.82
44	MET	CE	17.19	49	LEU	HB3	1.75	52	LYS	N	118.23
44	MET	CG	32.41	49	LEU	HD1	1.01	53	TRP	H	8.42
44	MET	N	120.75	49	LEU	HD2	1.02	53	TRP	HA	4.18

53	TRP	HB2	3.6	56	ARG	N	119.3	61	CYS	CA	62.63
53	TRP	HB3	3.78	57	HIS	H	8.09	61	CYS	CB	27.66
53	TRP	HD1	7.53	57	HIS	HA	4.55	61	CYS	N	116.17
53	TRP	HE1	10.62	57	HIS	HB2	2.59	62	GLU	H	7.81
53	TRP	HE3	7.31	57	HIS	HB3	3.35	62	GLU	HA	3.85
53	TRP	HH2	7.21	57	HIS	HD2	6.65	62	GLU	HB2	1.88
53	TRP	HZ2	7.08	57	HIS	C	174.56	62	GLU	HB3	1.88
53	TRP	HZ3	7.17	57	HIS	CA	57.49	62	GLU	HG2	1.88
53	TRP	C	179	57	HIS	CB	29.23	62	GLU	HG3	2.28
53	TRP	CA	62.96	57	HIS	CD2	122.61	62	GLU	C	177.25
53	TRP	CB	29.45	57	HIS	N	116.22	62	GLU	CA	58.14
53	TRP	CD1	127.5	58	ARG	H	7.19	62	GLU	CB	29.32
53	TRP	CE3	120.31	58	ARG	HA	3.11	62	GLU	CG	37.42
53	TRP	CH2	123.92	58	ARG	HB2	1.44	62	GLU	N	117.83
53	TRP	CZ2	113.94	58	ARG	HB3	0.95	63	LYS	H	7.25
53	TRP	CZ3	124.37	58	ARG	HD2	1.48	63	LYS	HA	4.01
53	TRP	N	121.33	58	ARG	HD3	2.39	63	LYS	HB2	1.4
53	TRP	NE1	130.37	58	ARG	HE	7.03	63	LYS	HB3	1.47
54	LEU	H	9.29	58	ARG	HG2	-0.17	63	LYS	HD2	1.46
54	LEU	HA	3.79	58	ARG	HG3	0.93	63	LYS	HD3	1.46
54	LEU	HB2	1.28	58	ARG	C	177.15	63	LYS	HE2	2.83
54	LEU	HB3	2.17	58	ARG	CA	60.53	63	LYS	HE3	2.83
54	LEU	HD1	0.37	58	ARG	CB	30.64	63	LYS	HG2	1.22
54	LEU	HD2	0.63	58	ARG	CD	43.59	63	LYS	HG3	1.14
54	LEU	HG	1.91	58	ARG	CG	26.14	63	LYS	C	176.97
54	LEU	C	179.82	58	ARG	N	123.54	63	LYS	CA	57.35
54	LEU	CA	59.1	58	ARG	NE	85.25	63	LYS	CB	32.27
54	LEU	CB	42.33	59	ASP	H	8.3	63	LYS	CD	28.98
54	LEU	CD1	23.78	59	ASP	HA	4.15	63	LYS	CE	42.13
54	LEU	CD2	25.2	59	ASP	HB2	2.52	63	LYS	CG	24.95
54	LEU	CG	27.86	59	ASP	HB3	2.52	63	LYS	N	117.23
54	LEU	N	122.23	59	ASP	C	178.43	64	TRP	H	7.57
55	ASP	H	9.15	59	ASP	CA	56.95	64	TRP	HA	4.86
55	ASP	HA	4.75	59	ASP	CB	39.88	64	TRP	HB2	3.34
55	ASP	HB2	2.96	59	ASP	N	116.63	64	TRP	HB3	3.34
55	ASP	HB3	2.78	60	MET	H	7.69	64	TRP	HD1	7.33
55	ASP	C	179.78	60	MET	HA	3.94	64	TRP	HE1	10.24
55	ASP	CA	58.45	60	MET	HB2	1.98	64	TRP	HE3	7.74
55	ASP	CB	39.42	60	MET	HB3	2.1	64	TRP	HH2	7.16
55	ASP	N	119.72	60	MET	HE	2.06	64	TRP	HZ2	7.52
56	ARG	H	7.62	60	MET	HG2	2.43	64	TRP	HZ3	7.13
56	ARG	HA	4.17	60	MET	HG3	2.58	64	TRP	CA	56.82
56	ARG	HB2	1.63	60	MET	C	177.77	64	TRP	CB	30.57
56	ARG	HB3	1.76	60	MET	CA	58.22	64	TRP	CD1	125.87
56	ARG	HD2	3.11	60	MET	CB	32.45	64	TRP	CE3	121.45
56	ARG	HD3	3.11	60	MET	CE	17.02	64	TRP	CH2	124.34
56	ARG	HG2	1.59	60	MET	CG	31.96	64	TRP	CZ2	114.54
56	ARG	HG3	1.59	60	MET	N	119.28	64	TRP	CZ3	122.21
56	ARG	C	176.89	61	CYS	H	7.77	64	TRP	N	119.42
56	ARG	CA	58.21	61	CYS	HA	4.24	64	TRP	NE1	128.83
56	ARG	CB	29.7	61	CYS	HB2	2.9	65	LYS	H	7.19
56	ARG	CD	43.8	61	CYS	HB3	3.34	65	LYS	HA	4.32
56	ARG	CG	27.51	61	CYS	C	175.79	65	LYS	HB2	1.89

65	LYS	HB3	1.96	69	ASP	CA	56.27	73	LYS	CB	30.85
65	LYS	HD2	1.57	69	ASP	CB	41.09	73	LYS	CD	27.54
65	LYS	HD3	1.57	69	ASP	N	121.79	73	LYS	CE	41.74
65	LYS	HE2	2.93	70	ILE	H	8.25	73	LYS	CG	24.23
65	LYS	HE3	2.93	70	ILE	HA	3.65	73	LYS	N	119.99
65	LYS	HG2	1.39	70	ILE	HB	2.07	74	LEU	H	8.87
65	LYS	HG3	1.39	70	ILE	HD1	1.1	74	LEU	HA	4.14
65	LYS	C	176.65	70	ILE	HG12	1.47	74	LEU	HB2	2.04
65	LYS	CA	58.5	70	ILE	HG13	2.01	74	LEU	HB3	1.24
65	LYS	CB	33.49	70	ILE	HG2	1.2	74	LEU	HD1	0.64
65	LYS	CD	29.09	70	ILE	C	177.64	74	LEU	HD2	0.66
65	LYS	CE	41.92	70	ILE	CA	65.68	74	LEU	HG	1.93
65	LYS	CG	24.94	70	ILE	CB	38.15	74	LEU	C	179.28
65	LYS	N	116.09	70	ILE	CD1	14.73	74	LEU	CA	58.24
66	SER	H	8.4	70	ILE	CG1	30.29	74	LEU	CB	42.91
66	SER	HA	4.67	70	ILE	CG2	18.95	74	LEU	CD1	22.7
66	SER	HB2	4.21	70	ILE	N	121.81	74	LEU	CD2	25.51
66	SER	HB3	4.06	71	LEU	H	8.53	74	LEU	CG	26.65
66	SER	CA	57.79	71	LEU	HA	4.12	74	LEU	N	121.04
66	SER	CB	64.61	71	LEU	HB2	1.89	75	ASN	H	8.48
66	SER	N	114.66	71	LEU	HB3	1.61	75	ASN	HA	4.59
67	LYS	H	8.94	71	LEU	HD1	0.96	75	ASN	HB2	3.01
67	LYS	HA	4.14	71	LEU	HD2	0.98	75	ASN	HB3	2.76
67	LYS	HB2	1.99	71	LEU	HG	1.82	75	ASN	HD21	7.41
67	LYS	HB3	1.93	71	LEU	C	179.67	75	ASN	HD22	6.65
67	LYS	HD2	1.72	71	LEU	CA	58.67	75	ASN	C	177.77
67	LYS	HD3	1.72	71	LEU	CB	40.92	75	ASN	CA	57.43
67	LYS	HE2	2.96	71	LEU	CD1	25.36	75	ASN	CB	38.8
67	LYS	HE3	2.96	71	LEU	CD2	23.6	75	ASN	N	119.04
67	LYS	HG2	1.53	71	LEU	CG	27.56	75	ASN	ND2	111.26
67	LYS	HG3	1.43	71	LEU	N	119.02	76	GLU	H	8.12
67	LYS	C	177.33	72	HIS	H	8.02	76	GLU	HA	4.09
67	LYS	CA	59.51	72	HIS	HA	4.42	76	GLU	HB2	2.15
67	LYS	CB	32.88	72	HIS	HB2	3.28	76	GLU	HB3	2.15
67	LYS	CD	29.29	72	HIS	HB3	3.28	76	GLU	HG2	2.1
67	LYS	CE	41.87	72	HIS	HD2	7	76	GLU	HG3	2.32
67	LYS	CG	25.22	72	HIS	C	178.32	76	GLU	C	179.13
67	LYS	N	124.09	72	HIS	CA	59.8	76	GLU	CA	59.45
68	GLU	H	8.72	72	HIS	CB	29.82	76	GLU	CB	29.98
68	GLU	HA	4.13	72	HIS	CD2	119.23	76	GLU	CG	36.61
68	GLU	HB2	1.96	72	HIS	N	118.97	76	GLU	N	119.9
68	GLU	HB3	1.96	73	LYS	H	8.17	77	GLN	H	8.01
68	GLU	HG2	2.24	73	LYS	HA	3.94	77	GLN	HA	4.12
68	GLU	HG3	2.37	73	LYS	HB2	1.53	77	GLN	HB2	2.4
68	GLU	CA	58.77	73	LYS	HB3	1.26	77	GLN	HB3	2.48
68	GLU	CB	29.87	73	LYS	HD2	1.16	77	GLN	HE21	7.9
68	GLU	CG	36.58	73	LYS	HD3	1.16	77	GLN	HE22	7.05
68	GLU	N	117.47	73	LYS	HE2	2.42	77	GLN	HG2	2.55
69	ASP	H	7.08	73	LYS	HE3	2.42	77	GLN	HG3	2.55
69	ASP	HA	4.56	73	LYS	HG2	0.62	77	GLN	C	178.05
69	ASP	HB2	2.66	73	LYS	HG3	1.16	77	GLN	CA	57.86
69	ASP	HB3	2.84	73	LYS	C	180.34	77	GLN	CB	28.49
69	ASP	C	177.85	73	LYS	CA	58.18	77	GLN	CG	33.85

77	GLN	N	117.39	78	TRP	CZ3	120.32	80	LYS	HD3	1.73
77	GLN	NE2	113.57	78	TRP	N	121.88	80	LYS	HE2	3.04
78	TRP	H	8.77	78	TRP	NE1	128.82	80	LYS	HE3	3.04
78	TRP	HA	4.03	79	ASN	H	7.63	80	LYS	HG2	1.55
78	TRP	HB2	3.22	79	ASN	HA	4.42	80	LYS	HG3	1.55
78	TRP	HB3	3.35	79	ASN	HB2	2.66	80	LYS	C	175.53
78	TRP	HD1	6.92	79	ASN	HB3	2.86	80	LYS	CA	56.92
78	TRP	HE1	9.94	79	ASN	HD21	7.48	80	LYS	CB	33.36
78	TRP	HE3	6.34	79	ASN	HD22	6.98	80	LYS	CD	29.38
78	TRP	HH2	6.75	79	ASN	C	175.11	80	LYS	CE	42.25
78	TRP	HZ2	7.14	79	ASN	CA	54.08	80	LYS	CG	24.55
78	TRP	HZ3	7.3	79	ASN	CB	39.36	80	LYS	N	120.2
78	TRP	C	176.63	79	ASN	N	114.58	81	ASP	H	7.79
78	TRP	CA	59.7	79	ASN	ND2	113.44	81	ASP	HA	4.39
78	TRP	CB	29.76	80	LYS	H	7.47	81	ASP	HB2	2.58
78	TRP	CD1	127.01	80	LYS	HA	4.29	81	ASP	HB3	2.66
78	TRP	CE3	122.18	80	LYS	HB2	1.92	81	ASP	CA	56.15
78	TRP	CH2	123.71	80	LYS	HB3	1.92	81	ASP	CB	42.33
78	TRP	CZ2	114.03	80	LYS	HD2	1.73	81	ASP	N	126.91

ATS-FL

2	LYS	H	8.211	15	ASP	HA	4.646	25	THR	C	174.7
2	LYS	HA	4.363	15	ASP	C	176.5	25	THR	CA	61.88
2	LYS	C	176.5	15	ASP	CB	41.15	25	THR	CB	69.88
2	LYS	CA	55.95	15	ASP	CA	54.06	25	THR	N	114.8
2	LYS	CB	33.3	16	ILE	H	8.06	26	LEU	H	8.364
2	LYS	N	121.1	16	ILE	HA	4.085	26	LEU	CB	42.4
3	LYS	H	8.523	16	ILE	C	176.3	26	LEU	CA	55.33
3	LYS	HA	4.591	16	ILE	CA	61.75	26	LEU	N	124.7
3	LYS	C	174.5	16	ILE	CB	38.49	27	LYS	HA	4.248
3	LYS	CA	54.3	16	ILE	N	120.7	27	LYS	C	175.6
3	LYS	CB	32.52	17	HIS	HA	4.703	27	LYS	CB	33.19
3	LYS	N	124.9	17	HIS	C	174.8	27	LYS	CA	56.46
4	PRO	CA	62.9	17	HIS	CB	29.27	28	SER	H	8.116
4	PRO	CB	32.27	17	HIS	CA	55.6	28	SER	CB	64.52
4	PRO	HA	4.453	18	LYS	H	8.221	28	SER	CA	55.48
4	PRO	C	176.8	18	LYS	HA	4.293	28	SER	N	116.9
4	PRO	N	116.8	18	LYS	CA	56.9	29	LYS	HA	4.318
5	LYS	H	8.576	18	LYS	C	177	29	LYS	C	176.3
5	LYS	HA	4.302	18	LYS	N	122.5	29	LYS	CB	32.9
5	LYS	C	176.6	19	GLY	HA2	3.925	29	LYS	CA	56.64
5	LYS	CA	56.4	19	GLY	HA3	3.925	30	ASN	H	8.423
5	LYS	CB	33.19	19	GLY	C	173.7	30	ASN	HA	4.683
5	LYS	N	122.3	19	GLY	CA	45.34	30	ASN	CA	53.3
6	SER	H	8.536	20	ASP	H	8.159	30	ASN	C	174.9
6	SER	HA	4.782	20	ASP	HA	4.59	30	ASN	N	119.3
6	SER	C	172.8	20	ASP	C	176.2	37	SER	HA	4.512
6	SER	CA	56.39	20	ASP	CA	54.17	37	SER	C	176.6
6	SER	CB	63.46	20	ASP	CB	41.13	37	SER	CB	63.95
6	SER	N	119.1	20	ASP	N	120.3	37	SER	CA	58.37
7	PRO	CA	63.65	21	TYR	H	8.257	38	ASP	H	8.4
7	PRO	CB	32.19	21	TYR	HA	4.515	38	ASP	HA	4.658
7	PRO	HA	4.461	21	TYR	C	176.4	38	ASP	C	175.4
7	PRO	C	177.2	21	TYR	CA	58.26	38	ASP	CA	54.6
7	PRO	N	117.2	21	TYR	CB	38.55	38	ASP	CB	41.25
8	VAL	H	8.111	21	TYR	N	120.6	38	ASP	N	122.7
8	VAL	HA	4.062	22	GLY	H	8.391	39	THR	H	8.049
8	VAL	C	176	22	GLY	HA2	3.87	39	THR	HA	4.474
8	VAL	CA	62.59	22	GLY	HA3	3.87	39	THR	C	174.2
8	VAL	CB	32.55	22	GLY	C	173.8	39	THR	CA	58.83
8	VAL	N	118.8	22	GLY	CA	45.36	39	THR	CB	70.79
9	ASP	H	8.286	22	GLY	N	110.3	39	THR	N	113.5
9	ASP	HA	4.599	23	MET	H	8.12	42	GLY	HA2	3.942
9	ASP	C	176.1	23	MET	HA	4.732	42	GLY	HA3	3.942
9	ASP	CA	54.52	23	MET	CA	53.52	42	GLY	C	173.7
9	ASP	CB	41.17	23	MET	CB	32.27	42	GLY	CA	45.26
9	ASP	N	123.4	23	MET	N	120.8	43	LYS	H	8.345
10	LEU	H	8.13	24	PRO	HA	4.502	43	LYS	HA	4.364
10	LEU	HA	4.604	24	PRO	C	177	43	LYS	CA	55.95
10	LEU	C	175.4	24	PRO	CB	32.08	43	LYS	CB	33.42
10	LEU	CA	55.39	24	PRO	CA	63.34	43	LYS	N	121.1
10	LEU	CB	42.37	25	THR	H	8.327	49	GLU	CB	30.04
10	LEU	N	122.3	25	THR	HA	4.338	49	GLU	C	176.9

49	GLU	CA	57.1	61	THR	CA	61.73	70	GLU	C	176.1
50	GLY	HA2	3.958	61	THR	CB	69.57	70	GLU	CA	56.72
50	GLY	HA3	3.958	61	THR	N	115.1	70	GLU	CB	30.27
50	GLY	CA	45.24	62	THR	H	8.32	70	GLU	N	122.8
51	ASP	HA	4.684	62	THR	HA	4.297	71	TYR	H	8.178
51	ASP	C	176.7	62	THR	C	174.3	71	TYR	HA	4.582
51	ASP	CB	41.45	62	THR	CA	62.74	71	TYR	C	175.6
51	ASP	CA	54.33	62	THR	CB	69.72	71	TYR	CA	57.91
52	SER	H	8.466	62	THR	N	116.8	71	TYR	CB	39.04
52	SER	HA	4.414	63	ASP	H	8.469	71	TYR	N	120.7
52	SER	C	174.6	63	ASP	HA	4.672	72	GLU	H	8.255
52	SER	CA	58.82	63	ASP	C	176.3	72	GLU	HA	4.274
52	SER	CB	63.85	63	ASP	CA	54.29	72	GLU	C	175.8
52	SER	N	116.3	63	ASP	CB	41.09	72	GLU	CA	56.15
53	ASP	H	8.494	63	ASP	N	123.2	72	GLU	CB	30.79
53	ASP	HA	4.69	64	ILE	H	8.214	72	GLU	N	123.2
53	ASP	C	176.6	64	ILE	HA	4.263	73	GLU	H	8.459
53	ASP	CA	54.59	64	ILE	C	176.8	73	GLU	HA	4.243
53	ASP	CB	41.23	64	ILE	CA	61.47	73	GLU	C	176.4
53	ASP	N	122.2	64	ILE	CB	38.58	73	GLU	CA	56.51
54	SER	H	8.277	64	ILE	N	121.4	73	GLU	CB	30.26
54	SER	HA	4.383	65	THR	H	8.38	73	GLU	N	122.5
54	SER	C	175.2	65	THR	HA	4.388	74	MET	H	8.429
54	SER	CA	58.94	65	THR	C	174.8	74	MET	HA	4.457
54	SER	CB	63.81	65	THR	CA	62.22	74	MET	C	175.7
54	SER	N	116.1	65	THR	CB	69.9	74	MET	CA	55.51
55	GLY	H	8.402	65	THR	N	118	74	MET	CB	33.46
55	GLY	HA2	3.877	66	SER	H	8.394	74	MET	N	121.9
55	GLY	HA3	3.877	66	SER	HA	4.527	75	ASP	H	8.604
55	GLY	C	173.7	66	SER	C	174.7	75	ASP	HA	4.651
55	GLY	CA	45.36	66	SER	CA	58.38	75	ASP	C	176.6
55	GLY	N	110.2	66	SER	CB	63.92	75	ASP	CA	54.15
56	HIS	H	8.118	66	SER	N	118.5	75	ASP	CB	41.33
56	HIS	HA	4.631	67	SER	H	8.502	75	ASP	N	122.7
56	HIS	C	175.4	67	SER	HA	4.47	76	ILE	H	8.232
56	HIS	CA	55.11	67	SER	C	174.8	76	ILE	HA	4.155
56	HIS	CB	29.52	67	SER	CA	58.4	76	ILE	C	176.2
56	HIS	N	118	67	SER	CB	63.9	76	ILE	CA	61.77
58	TYR	HA	4.575	67	SER	N	118.3	76	ILE	CB	38.61
58	TYR	C	175.5	68	GLU	H	8.529	76	ILE	N	121.7
58	TYR	CB	39.04	68	GLU	HA	4.338	77	ASN	H	8.558
58	TYR	CA	57.88	68	GLU	C	176.7	77	ASN	HA	4.72
59	GLU	H	8.242	68	GLU	CA	56.85	77	ASN	C	175
59	GLU	HA	4.263	68	GLU	CB	30.17	77	ASN	CA	53.7
59	GLU	CA	56.07	68	GLU	N	122.7	77	ASN	CB	39.18
59	GLU	N	123.4	69	SER	H	8.348	77	ASN	N	121.4
60	ASP	HA	4.695	69	SER	HA	4.44	78	ASP	H	8.237
60	ASP	C	176.7	69	SER	C	174.5	78	ASP	HA	4.635
60	ASP	CB	41.32	69	SER	CA	58.6	78	ASP	C	175.9
60	ASP	CA	54.21	69	SER	CB	63.95	78	ASP	CA	54.43
61	THR	H	8.306	69	SER	N	116.5	78	ASP	CB	41.18
61	THR	HA	4.441	70	GLU	H	8.45	78	ASP	N	120.9
61	THR	C	174.9	70	GLU	HA	4.263	79	ILE	H	7.893

79	ILE	HA	4.131	88	LYS	C	176.3	97	PRO	CB	32.19
79	ILE	C	175.8	88	LYS	CA	56.3	97	PRO	HA	4.405
79	ILE	CA	61.29	88	LYS	CB	33.32	97	PRO	C	177.2
79	ILE	CB	38.83	88	LYS	N	122.8	97	PRO	N	116.8
79	ILE	N	119.9	89	THR	H	8.209	98	SER	H	8.519
80	TYR	H	8.37	89	THR	HA	4.299	98	SER	HA	4.402
80	TYR	HA	4.59	89	THR	C	174.2	98	SER	C	174.8
80	TYR	C	175	89	THR	CA	62.16	98	SER	CA	58.69
80	TYR	CA	57.83	89	THR	CB	69.86	98	SER	CB	63.74
80	TYR	CB	38.76	89	THR	N	116.3	98	SER	N	115.9
80	TYR	N	125.2	90	LEU	H	8.353	99	LYS	H	8.412
81	VAL	H	8.019	90	LEU	HA	4.394	99	LYS	HA	4.376
81	VAL	HA	4.284	90	LEU	C	177	99	LYS	C	176.5
81	VAL	C	173.8	90	LEU	CA	55.23	99	LYS	CA	56.29
81	VAL	CA	59.4	90	LEU	CB	42.36	99	LYS	CB	33.04
81	VAL	CB	33.01	90	LEU	N	125.4	99	LYS	N	123.6
81	VAL	N	126.8	91	ILE	H	8.172	100	ARG	H	8.364
82	PRO	CA	63.2	91	ILE	HA	4.178	100	ARG	HA	4.357
82	PRO	CB	32.13	91	ILE	C	176	100	ARG	C	176
82	PRO	HA	4.161	91	ILE	CA	60.81	100	ARG	CA	56.12
82	PRO	C	176.6	91	ILE	CB	38.69	100	ARG	CB	31.06
82	PRO	N	117.5	91	ILE	N	122.4	100	ARG	N	122.4
83	ASP	H	8.356	92	GLU	H	8.486	101	ASP	H	8.487
83	ASP	HA	4.554	92	GLU	HA	4.315	101	ASP	HA	4.655
83	ASP	C	175.9	92	GLU	C	176.1	101	ASP	C	176
83	ASP	CA	54.32	92	GLU	CA	56.38	101	ASP	CA	54.38
83	ASP	CB	40.88	92	GLU	CB	30.58	101	ASP	CB	41.16
83	ASP	N	119.4	92	GLU	N	126	101	ASP	N	121.7
84	SER	H	8.155	93	VAL	H	8.334	102	THR	H	8.183
84	SER	HA	4.732	93	VAL	HA	4.081	102	THR	HA	4.604
84	SER	C	173	93	VAL	C	175.8	102	THR	C	172.8
84	SER	CA	56.42	93	VAL	CA	62.41	102	THR	CA	59.81
84	SER	CB	63.49	93	VAL	CB	32.8	102	THR	CB	69.92
84	SER	N	116.8	93	VAL	N	123.3	102	THR	N	116.9
85	PRO	CA	63.6	94	VAL	H	8.417	103	PRO	CA	63.28
85	PRO	CB	32.1	94	VAL	HA	4.093	103	PRO	CB	32.27
85	PRO	HA	4.396	94	VAL	C	175.9	103	PRO	HA	4.484
85	PRO	C	177	94	VAL	CA	62.29	103	PRO	C	177.1
85	PRO	N	117.5	94	VAL	CB	32.72	103	PRO	N	118.6
86	LYS	H	8.329	94	VAL	N	126.3	104	SER	H	8.617
86	LYS	HA	4.184	95	LEU	H	8.507	104	SER	HA	4.454
86	LYS	C	176.5	95	LEU	HA	4.401	104	SER	C	174.98
86	LYS	CA	56.57	95	LEU	C	176.9	104	SER	CA	58.35
86	LYS	CB	32.89	95	LEU	CA	54.71	104	SER	CB	63.95
86	LYS	N	120.7	95	LEU	CB	42.52	104	SER	N	116.89
87	TYR	H	8	95	LEU	N	127.7	105	SER	H	8.515
87	TYR	HA	4.596	96	GLU	H	8.478	105	SER	HA	4.453
87	TYR	C	175.7	96	GLU	HA	4.598	105	SER	C	174.21
87	TYR	CA	57.71	96	GLU	C	174.6	105	SER	CA	58.43
87	TYR	CB	38.9	96	GLU	CA	54.34	105	SER	CB	63.87
87	TYR	N	120	96	GLU	CB	29.82	105	SER	N	118.11
88	LYS	H	8.214	96	GLU	N	123.7	106	ASP	H	8.363
88	LYS	HA	4.344	97	PRO	CA	63.53	106	ASP	HA	4.62

106	ASP	C	175.52	116	GLU	HA	4.034	125	ILE	H	8.79
106	ASP	CA	54.07	116	GLU	C	179.96	125	ILE	HA	3.106
106	ASP	CB	41.08	116	GLU	CA	59.16	125	ILE	C	177.33
106	ASP	N	122.39	116	GLU	CB	30.53	125	ILE	CA	65.91
107	ALA	H	8.155	116	GLU	N	121.47	125	ILE	CB	38.3
107	ALA	HA	4.576	117	TRP	H	8.928	125	ILE	N	121.04
107	ALA	C	175.39	117	TRP	HA	4.615	126	SER	H	8.432
107	ALA	CA	50.65	117	TRP	C	176.68	126	SER	HA	4.024
107	ALA	CB	18.12	117	TRP	CA	58.75	126	SER	C	176.58
107	ALA	N	125.39	117	TRP	CB	30.2	126	SER	CA	61.87
108	PRO	N	120	117	TRP	N	121.6	126	SER	CB	62.83
109	MET	H	8.408	118	ASN	H	8.652	126	SER	N	114.24
109	MET	HA	4.469	118	ASN	HA	4.113	127	GLY	H	7.33
109	MET	CA	55.6	118	ASN	C	178.5	127	GLY	HA2	3.83
109	MET	CB	32.77	118	ASN	CA	56.55	127	GLY	HA3	3.83
109	MET	N	121.6	118	ASN	CB	37.77	127	GLY	C	175.59
110	ASN	H	8.357	118	ASN	N	116.89	127	GLY	CA	46.55
110	ASN	HA	4.652	119	GLN	H	7.814	127	GLY	N	106.64
110	ASN	C	174.07	119	GLN	HA	3.927	128	ILE	H	7.301
110	ASN	CA	53.27	119	GLN	C	177.25	128	ILE	HA	3.77
110	ASN	CB	39.28	119	GLN	CA	58.71	128	ILE	C	178.2
110	ASN	N	119.57	119	GLN	CB	28.36	128	ILE	CA	61.79
111	LYS	H	7.972	119	GLN	N	119.76	128	ILE	CB	36.82
111	LYS	HA	4.176	120	LEU	H	7.756	128	ILE	N	120.79
111	LYS	C	176.67	120	LEU	HA	4.037	129	LEU	H	8.004
111	LYS	CA	55.66	120	LEU	C	179.4	129	LEU	HA	4.085
111	LYS	CB	33.81	120	LEU	CA	58.38	129	LEU	C	178.76
111	LYS	N	119.64	120	LEU	CB	41.59	129	LEU	CA	56.55
112	PHE	H	8.785	120	LEU	N	122.53	129	LEU	CB	41.45
112	PHE	HA	5.055	121	LYS	H	8.042	129	LEU	N	119.6
112	PHE	C	178.62	121	LYS	HA	3.356	130	GLU	H	7.752
112	PHE	CA	57.01	121	LYS	C	178.38	130	GLU	HA	4.159
112	PHE	CB	40.31	121	LYS	CA	61.63	130	GLU	C	177.19
112	PHE	N	121.35	121	LYS	CB	32.11	130	GLU	CA	57.7
113	THR	H	9.205	121	LYS	N	116.12	130	GLU	CB	29.76
113	THR	HA	4.519	122	GLN	H	7.699	130	GLU	N	118.24
113	THR	C	175.68	122	GLN	HA	3.852	131	ASN	H	7.755
113	THR	CA	61.36	122	GLN	C	178.72	131	ASN	HA	4.633
113	THR	CB	70.66	122	GLN	CA	58.8	131	ASN	C	175.75
113	THR	N	114.98	122	GLN	CB	27.85	131	ASN	CA	54.07
114	ASP	H	9.163	122	GLN	N	117.08	131	ASN	CB	39.34
114	ASP	HA	4.572	123	ASP	H	8.299	131	ASN	N	117.29
114	ASP	C	178.68	123	ASP	HA	4.338	132	GLU	H	8.332
114	ASP	CA	58.43	123	ASP	C	179.31	132	GLU	HA	4.211
114	ASP	CB	39.41	123	ASP	CA	57.59	132	GLU	C	177.17
114	ASP	N	122.26	123	ASP	CB	40.2	132	GLU	CA	57.68
115	ASP	H	8.49	123	ASP	N	121.97	132	GLU	CB	29.94
115	ASP	HA	4.546	124	PHE	H	8.501	132	GLU	N	120.51
115	ASP	C	179.14	124	PHE	HA	3.993	133	GLN	H	8.296
115	ASP	CA	57.56	124	PHE	C	177.9	133	GLN	HA	4.272
115	ASP	CB	40.71	124	PHE	CA	61.6	133	GLN	C	176.74
115	ASP	N	118.86	124	PHE	CB	38.69	133	GLN	CA	56.61
116	GLU	H	7.964	124	PHE	N	122.51	133	GLN	CB	28.79

133	GLN	N	119.92	152	ASN	C	175.56	161	GLU	C	176.13
134	LYS	H	8.134	152	ASN	CA	53.33	161	GLU	CA	56.34
134	LYS	HA	4.17	152	ASN	CB	38.66	161	GLU	CB	30.43
134	LYS	C	176.98	152	ASN	N	118.81	161	GLU	N	122.95
134	LYS	CA	57.63	153	THR	H	8.113	162	LYS	H	8.45
134	LYS	CB	32.73	153	THR	HA	4.244	162	LYS	HA	4.596
134	LYS	N	120.76	153	THR	C	174.31	162	LYS	C	174.38
135	ASP	H	8.251	153	THR	CA	62.07	162	LYS	CA	54.25
135	ASP	HA	4.579	153	THR	CB	69.75	162	LYS	CB	32.5
135	ASP	C	176.86	153	THR	N	114.66	162	LYS	N	123.84
135	ASP	CA	55.06	154	LEU	H	8.193	163	PRO	CA	63.11
135	ASP	CB	40.76	154	LEU	HA	4.243	163	PRO	CB	32.06
135	ASP	N	119.47	154	LEU	C	176.74	163	PRO	HA	4.382
136	LEU	H	7.973	154	LEU	CA	55.32	163	PRO	C	176.42
136	LEU	HA	4.205	154	LEU	CB	42.33	163	PRO	N	113.38
136	LEU	C	178.27	154	LEU	N	124.1	164	PHE	H	8.355
136	LEU	CA	56.82	155	TYR	H	8.047	164	PHE	HA	4.598
136	LEU	CB	42.17	155	TYR	HA	4.511	164	PHE	C	175.55
136	LEU	N	121.47	155	TYR	C	175.2	164	PHE	CA	57.72
143	GLY	H	8.44	155	TYR	CA	57.58	164	PHE	CB	39.39
143	GLY	HA2	3.985	155	TYR	CB	38.87	164	PHE	N	120.69
143	GLY	HA3	3.985	155	TYR	N	120.01	165	ILE	H	8.136
143	GLY	CA	45.41	156	PHE	H	8.086	165	ILE	HA	4.202
143	GLY	N	110.88	156	PHE	HA	4.546	165	ILE	C	175.89
144	ASN	H	8.363	156	PHE	C	174.92	165	ILE	CA	60.65
144	ASN	HA	4.79	156	PHE	CA	57.42	165	ILE	CB	38.99
144	ASN	C	175.14	156	PHE	CB	39.8	165	ILE	N	123.78
144	ASN	CA	53.15	156	PHE	N	121.35	166	THR	H	8.28
144	ASN	CB	38.92	157	ASN	H	8.364	166	THR	HA	4.331
144	ASN	N	118.68	157	ASN	HA	4.597	166	THR	C	174.34
145	THR	H	8.271	157	ASN	C	174.28	166	THR	CA	61.9
145	THR	HA	4.599	157	ASN	CA	53.04	166	THR	CB	69.8
145	THR	C	172.79	157	ASN	CB	39.39	166	THR	N	118.9
145	THR	CA	60.02	157	ASN	N	120.66	167	SER	H	8.434
145	THR	CB	69.74	158	LYS	H	8.281	167	SER	HA	4.517
145	THR	N	117.26	158	LYS	HA	4.538	167	SER	C	174.66
149	THR	H	8.166	158	LYS	C	174.41	167	SER	CA	57.95
149	THR	HA	4.324	158	LYS	CA	54.39	167	SER	CB	64.07
149	THR	C	174.34	158	LYS	CB	32.59	167	SER	N	118.62
149	THR	CA	61.75	158	LYS	N	123.15	168	ILE	H	8.259
149	THR	CB	69.78	159	PRO	CA	63.37	168	ILE	HA	4.106
149	THR	N	114.3	159	PRO	CB	32	168	ILE	C	176.18
150	GLN	H	8.456	159	PRO	HA	4.385	168	ILE	CA	61.7
150	GLN	HA	4.616	159	PRO	C	176.98	168	ILE	CB	38.42
150	GLN	CA	53.99	159	PRO	N	114.58	168	ILE	N	122.42
150	GLN	CB	28.87	160	GLU	H	8.722	169	HIS	H	8.362
150	GLN	N	123.64	160	GLU	HA	4.225	169	HIS	HA	4.673
151	PRO	HA	4.395	160	GLU	C	176.51	169	HIS	CB	30.23
151	PRO	C	176.67	160	GLU	CA	56.6	169	HIS	N	122.19
151	PRO	CB	32.18	160	GLU	CB	30.16	174	TYR	H	8.235
151	PRO	CA	63.33	160	GLU	N	121.04	174	TYR	HA	4.63
152	ASN	H	8.653	161	GLU	H	8.46	174	TYR	C	176.09
152	ASN	HA	4.703	161	GLU	HA	4.27	174	TYR	CA	57.86

174	TYR	CB	38.72	183	ILE	CB	38.8	192	ASP	CA	54.45
174	TYR	N	119.97	183	ILE	N	120.67	192	ASP	CB	41.14
175	SER	H	8.36	184	ASN	H	8.518	192	ASP	N	121.25
175	SER	HA	4.431	184	ASN	HA	4.698	193	THR	H	8.322
175	SER	C	174.99	184	ASN	C	175.29	193	THR	HA	4.316
175	SER	CA	58.43	184	ASN	CA	53.45	193	THR	C	175.11
175	SER	CB	63.98	184	ASN	CB	38.76	193	THR	CA	62.48
175	SER	N	117.95	184	ASN	N	121.77	193	THR	CB	69.52
176	GLY	H	8.024	185	MET	H	8.351	193	THR	N	114.69
176	GLY	HA2	3.938	185	MET	HA	4.523	194	SER	H	8.391
176	GLY	HA3	3.938	185	MET	C	176.45	194	SER	HA	4.411
176	GLY	CA	45.38	185	MET	CA	55.58	194	SER	C	174.22
176	GLY	N	110.75	185	MET	CB	32.61	194	SER	CA	58.88
177	GLU	H	8.292	185	MET	N	121.14	194	SER	CB	63.81
177	GLU	HA	4.284	186	SER	H	8.412	194	SER	N	118.24
177	GLU	C	176.59	186	SER	HA	4.522	195	TYR	H	8.109
177	GLU	CA	56.43	186	SER	C	174.93	195	TYR	HA	4.556
177	GLU	CB	30.51	186	SER	CA	58.64	195	TYR	C	175.66
177	GLU	N	120.24	186	SER	CB	63.79	195	TYR	CA	58.26
178	GLU	H	8.615	186	SER	N	116.89	195	TYR	CB	38.69
178	GLU	HA	4.256	187	THR	H	8.27	195	TYR	N	122.34
178	GLU	C	176.5	187	THR	HA	4.396	196	VAL	H	7.973
178	GLU	CA	56.7	187	THR	C	174.41	196	VAL	HA	4.042
178	GLU	CB	30.2	187	THR	CA	61.91	196	VAL	C	175.82
178	GLU	N	122.21	187	THR	CB	69.65	196	VAL	CA	62.21
179	ILE	H	8.216	187	THR	N	115.48	196	VAL	CB	33.04
179	ILE	HA	4.005	188	ASN	H	8.486	196	VAL	N	122.78
179	ILE	C	175.54	188	ASN	HA	4.75	197	SER	H	8.36
179	ILE	CA	61.39	188	ASN	C	175.31	197	SER	HA	4.382
179	ILE	CB	38.64	188	ASN	CA	53.32	197	SER	C	174.39
179	ILE	N	122.54	188	ASN	CB	38.96	197	SER	CA	58.26
180	ASN	H	8.401	188	ASN	N	121	197	SER	CB	63.86
180	ASN	HA	4.705	189	SER	H	8.441	197	SER	N	119.29
180	ASN	C	174.96	189	SER	HA	4.425	198	ASN	H	8.489
180	ASN	CA	52.77	189	SER	C	174.77	198	ASN	HA	4.75
180	ASN	CB	38.83	189	SER	CA	58.64	198	ASN	C	174.96
180	ASN	N	122.53	189	SER	CB	63.79	198	ASN	CA	53.32
181	TYR	H	8.27	189	SER	N	116.72	198	ASN	CB	38.96
181	TYR	HA	4.516	190	MET	H	8.499	198	ASN	N	121.02
181	TYR	C	175.58	190	MET	HA	4.478	199	ASN	H	8.421
181	TYR	CA	58.48	190	MET	C	176.05	199	ASN	HA	4.696
181	TYR	CB	38.75	190	MET	CA	55.73	199	ASN	C	175.02
181	TYR	N	121.67	190	MET	CB	32.54	199	ASN	CA	53.35
182	ASN	H	8.386	190	MET	N	121.77	199	ASN	CB	38.81
182	ASN	HA	4.67	191	ASP	H	8.271	199	ASN	N	119.21
182	ASN	C	175.27	191	ASP	HA	4.614	200	VAL	H	8.067
182	ASN	CA	53.33	191	ASP	C	176.02	200	VAL	HA	4.059
182	ASN	CB	38.83	191	ASP	CA	54.43	200	VAL	C	175.92
182	ASN	N	120.04	191	ASP	CB	41.19	200	VAL	CA	62.55
183	ILE	H	8.019	191	ASP	N	120.85	200	VAL	CB	32.6
183	ILE	HA	4.107	192	ASP	H	8.34	200	VAL	N	120.29
183	ILE	C	175.94	192	ASP	HA	4.657	201	TYR	H	8.402
183	ILE	CA	61.63	192	ASP	C	176.97	201	TYR	HA	4.62

201	TYR	C	175.89	210	THR	C	175.23	232	PHE	CB	39.46
201	TYR	CA	57.9	210	THR	CA	62.74	232	PHE	N	119.05
201	TYR	CB	38.76	210	THR	CB	69.7	233	GLY	H	8.263
201	TYR	N	124.06	210	THR	N	114.38	233	GLY	HA2	3.988
202	SER	H	8.355	211	LEU	H	8.26	233	GLY	HA3	3.988
202	SER	HA	4.42	211	LEU	HA	4.379	233	GLY	CA	45.46
202	SER	C	174.7	211	LEU	C	177.93	233	GLY	N	109.85
202	SER	CA	58.12	211	LEU	CA	55.58	234	THR	H	8.198
202	SER	CB	64.05	211	LEU	CB	42.03	234	THR	HA	4.262
202	SER	N	118.68	211	LEU	N	123.45	234	THR	C	174.68
203	GLY	H	7.859	212	SER	H	8.274	234	THR	CA	62.49
203	GLY	HA2	3.919	212	SER	HA	4.396	234	THR	CB	69.7
203	GLY	HA3	3.919	212	SER	C	175.32	234	THR	N	113.6
203	GLY	CA	45.5	212	SER	CA	58.77	235	ASN	H	8.56
203	GLY	N	110.42	212	SER	CB	63.82	235	ASN	HA	4.751
204	ILE	H	7.969	212	SER	N	115.8	235	ASN	CA	53.31
204	ILE	HA	4.179	213	GLY	H	8.401	235	ASN	CB	38.68
204	ILE	C	175.86	213	GLY	HA2	3.983	235	ASN	N	120.42
204	ILE	CA	61.31	213	GLY	HA3	3.983	246	ALA	H	8.4
204	ILE	CB	38.89	213	GLY	CA	45.52	246	ALA	HA	4.278
204	ILE	N	119.49	213	GLY	N	110.55	246	ALA	C	177.59
205	ASP	H	8.461	214	ASN	H	8.363	246	ALA	CA	52.72
205	ASP	HA	4.6	214	ASN	HA	4.696	246	ALA	CB	19.1
205	ASP	C	176.02	214	ASN	C	175.31	246	ALA	N	127.58
205	ASP	CA	54.31	214	ASN	CA	53.15	247	LYS	H	8.284
205	ASP	CB	41.05	214	ASN	CB	38.92	247	LYS	HA	4.299
205	ASP	N	123.87	214	ASN	N	118.68	247	LYS	CA	56.23
206	LEU	H	8.204	215	GLN	H	8.455	247	LYS	CB	33.06
206	LEU	HA	4.301	215	GLN	HA	4.277	247	LYS	N	120.64
206	LEU	C	177.56	215	GLN	C	175.66	251	SER	H	7.983
206	LEU	CA	55.37	215	GLN	CA	55.96	251	SER	HA	4.455
206	LEU	CB	42.29	215	GLN	CB	29.48	251	SER	CA	58.88
206	LEU	N	122.6	215	GLN	N	120.37	251	SER	N	114.33
207	ILE	H	8.15	216	HIS	H	8.533	252	ASP	H	8.63
207	ILE	HA	4.081	216	HIS	HA	4.682	252	ASP	HA	4.491
207	ILE	C	176.24	216	HIS	CA	55.88	252	ASP	CA	55.47
207	ILE	CA	61.54	216	HIS	CB	29.64	252	ASP	N	121.02
207	ILE	CB	38.33	216	HIS	N	120.46	253	PRO	HA	4.468
207	ILE	N	121.31	218	ASP	H	8.587	253	PRO	C	177.75
208	ASN	H	8.468	218	ASP	HA	4.63	253	PRO	CB	32.21
208	ASN	HA	4.705	218	ASP	C	176.23	253	PRO	CA	63.99
208	ASN	C	175.21	218	ASP	CA	54.24	254	ILE	H	8.38
208	ASN	CA	53.44	218	ASP	CB	41.38	254	ILE	HA	4.12
208	ASN	CB	38.9	218	ASP	N	124.99	254	ILE	C	176.96
208	ASN	N	121.99	219	ILE	H	8.095	254	ILE	CA	62.04
209	ASP	H	8.432	219	ILE	HA	4.09	254	ILE	CB	38.14
209	ASP	HA	4.664	219	ILE	CA	61.74	254	ILE	N	119.91
209	ASP	C	176.91	219	ILE	CB	38.64	255	MET	H	8.138
209	ASP	CA	54.64	219	ILE	N	120.76	255	MET	HA	4.427
209	ASP	CB	41.11	232	PHE	H	8.136	255	MET	C	176.5
209	ASP	N	121.6	232	PHE	HA	4.614	255	MET	CA	56.65
210	THR	H	8.212	232	PHE	C	176.55	255	MET	CB	32.76
210	THR	HA	4.311	232	PHE	CA	57.94	255	MET	N	120.75

256	ASN	H	8.349	264	TRP	N	121.33	273	GLU	CB	29.32
256	ASN	HA	4.746	265	LEU	H	9.294	273	GLU	N	117.83
256	ASN	C	176.05	265	LEU	HA	3.787	274	LYS	H	7.254
256	ASN	CA	53.74	265	LEU	C	179.82	274	LYS	HA	4.01
256	ASN	CB	38.96	265	LEU	CA	59.1	274	LYS	C	176.97
256	ASN	N	119.07	265	LEU	CB	42.33	274	LYS	CA	57.35
257	GLN	H	8.423	265	LEU	N	122.23	274	LYS	CB	32.27
257	GLN	HA	4.219	266	ASP	H	9.147	274	LYS	N	117.23
257	GLN	C	177.44	266	ASP	HA	4.75	275	TRP	H	7.573
257	GLN	CA	58.5	266	ASP	C	179.78	275	TRP	HA	4.86
257	GLN	CB	29.05	266	ASP	CA	58.45	275	TRP	CA	56.82
257	GLN	N	120.74	266	ASP	CB	39.42	275	TRP	CB	30.57
258	LEU	H	8.261	266	ASP	N	119.72	275	TRP	N	119.42
258	LEU	HA	4.173	267	ARG	H	7.622	276	LYS	H	7.193
258	LEU	C	178.71	267	ARG	HA	4.173	276	LYS	HA	4.317
258	LEU	CA	58.17	267	ARG	C	176.89	276	LYS	C	176.65
258	LEU	CB	41.45	267	ARG	CA	58.21	276	LYS	CA	58.5
258	LEU	N	119.96	267	ARG	CB	29.7	276	LYS	CB	33.49
259	ASP	H	8.142	267	ARG	N	119.3	276	LYS	N	116.09
259	ASP	HA	4.54	268	HIS	H	8.091	277	SER	H	8.403
259	ASP	C	179.28	268	HIS	HA	4.552	277	SER	HA	4.669
259	ASP	CA	57.66	268	HIS	C	174.56	277	SER	CA	57.79
259	ASP	CB	40.86	268	HIS	CA	57.49	277	SER	CB	64.61
259	ASP	N	118.92	268	HIS	CB	29.23	277	SER	N	114.66
260	LEU	H	7.901	268	HIS	N	116.22	278	LYS	H	8.936
260	LEU	HA	4.21	269	ARG	H	7.189	278	LYS	HA	4.137
260	LEU	C	180.3	269	ARG	HA	3.115	278	LYS	C	177.33
260	LEU	CA	58.2	269	ARG	C	177.15	278	LYS	CA	59.51
260	LEU	CB	41.97	269	ARG	CA	60.53	278	LYS	CB	32.88
260	LEU	N	119.57	269	ARG	CB	30.64	278	LYS	N	124.09
261	LEU	H	8.279	269	ARG	N	123.54	279	GLU	H	8.719
261	LEU	HA	4.234	270	ASP	H	8.302	279	GLU	HA	4.135
261	LEU	C	179.16	270	ASP	HA	4.151	279	GLU	CA	58.77
261	LEU	CA	58.71	270	ASP	C	178.43	279	GLU	CB	29.87
261	LEU	CB	42.11	270	ASP	CA	56.95	279	GLU	N	117.47
261	LEU	N	122.35	270	ASP	CB	39.88	280	ASP	H	7.085
262	HIS	H	8.546	270	ASP	N	116.63	280	ASP	HA	4.559
262	HIS	HA	4.133	271	MET	H	7.694	280	ASP	C	177.85
262	HIS	C	177.86	271	MET	HA	3.942	280	ASP	CA	56.27
262	HIS	CA	62.68	271	MET	C	177.77	280	ASP	CB	41.09
262	HIS	CB	31.5	271	MET	CA	58.22	280	ASP	N	121.79
262	HIS	N	118.41	271	MET	CB	32.45	281	ILE	H	8.248
263	LYS	H	8.136	271	MET	N	119.28	281	ILE	HA	3.655
263	LYS	HA	4.257	272	CYS	H	7.767	281	ILE	C	177.64
263	LYS	C	179.53	272	CYS	HA	4.235	281	ILE	CA	65.68
263	LYS	CA	59.88	272	CYS	C	175.79	281	ILE	CB	38.15
263	LYS	CB	32.73	272	CYS	CA	62.63	281	ILE	N	121.81
263	LYS	N	118.23	272	CYS	CB	27.66	282	LEU	H	8.533
264	TRP	H	8.417	272	CYS	N	116.17	282	LEU	HA	4.116
264	TRP	HA	4.176	273	GLU	H	7.806	282	LEU	C	179.67
264	TRP	C	179	273	GLU	HA	3.846	282	LEU	CA	58.67
264	TRP	CA	62.96	273	GLU	C	177.25	282	LEU	CB	40.92
264	TRP	CB	29.45	273	GLU	CA	58.14	282	LEU	N	119.02

283	HIS	H	8.023	291	LYS	N	120.2	301	ASP	H	8.463
283	HIS	HA	4.423	292	ASP	H	7.794	301	ASP	HA	4.611
283	HIS	C	178.32	292	ASP	HA	4.392	301	ASP	C	176
283	HIS	CA	59.8	292	ASP	CA	56.15	301	ASP	CA	53.82
283	HIS	CB	29.82	292	ASP	CB	42.33	301	ASP	CB	41.18
283	HIS	N	118.97	292	ASP	N	126.91	301	ASP	N	124.3
284	LYS	H	8.173	293	ASN	H	8.4	302	ASN	H	8.521
284	LYS	HA	3.941	293	ASN	HA	4.773	302	ASN	HA	4.698
284	LYS	C	180.34	293	ASN	C	175.3	302	ASN	C	175.8
284	LYS	CA	58.18	293	ASN	CA	53.2	302	ASN	CA	53.42
284	LYS	CB	30.85	293	ASN	CB	39.05	302	ASN	CB	38.77
284	LYS	N	119.99	293	ASN	N	118.9	302	ASN	N	121
285	LEU	H	8.865	294	ASP	H	8.514	303	ARG	H	8.357
285	LEU	HA	4.142	294	ASP	HA	4.64	303	ARG	HA	4.27
285	LEU	C	179.28	294	ASP	C	177	303	ARG	C	176.9
285	LEU	CA	58.24	294	ASP	CA	54.43	303	ARG	CA	56.83
285	LEU	CB	42.91	294	ASP	CB	41.11	303	ARG	CB	30.46
285	LEU	N	121.04	294	ASP	N	120.96	303	ARG	N	121.1
286	ASN	H	8.484	295	GLY	H	8.502	304	SER	H	8.348
286	ASN	HA	4.588	295	GLY	HA2	3.965	304	SER	HA	4.412
286	ASN	C	177.77	295	GLY	HA3	3.965	304	SER	C	174.6
286	ASN	CA	57.43	295	GLY	C	175.1	304	SER	CA	58.57
286	ASN	CB	38.8	295	GLY	CA	45.66	304	SER	CB	63.6
286	ASN	N	119.04	295	GLY	N	109.5	304	SER	N	116.2
287	GLU	H	8.116	296	GLY	H	8.375	305	LEU	H	8.209
287	GLU	HA	4.085	296	GLY	HA2	3.928	305	LEU	HA	4.365
287	GLU	C	179.13	296	GLY	HA3	3.928	305	LEU	C	177.1
287	GLU	CA	59.45	296	GLY	C	174	305	LEU	CA	55.21
287	GLU	CB	29.98	296	GLY	CA	45.08	305	LEU	CB	42.37
287	GLU	N	119.9	296	GLY	N	108.7	305	LEU	N	123.6
288	GLN	H	8.011	297	ASN	H	8.379	306	ASN	H	8.45
288	GLN	HA	4.125	297	ASN	HA	4.733	306	ASN	HA	4.762
288	GLN	C	178.05	297	ASN	C	175	306	ASN	C	175.5
288	GLN	CA	57.86	297	ASN	CA	53.2	306	ASN	CA	53.32
288	GLN	CB	28.49	297	ASN	CB	39.05	306	ASN	CB	38.86
288	GLN	N	117.39	297	ASN	N	118.7	306	ASN	N	119.4
289	TRP	H	8.774	298	VAL	H	8.186	307	THR	H	8.181
289	TRP	HA	4.035	298	VAL	HA	4.409	307	THR	HA	4.328
289	TRP	C	176.63	298	VAL	C	174.4	307	THR	C	174.4
289	TRP	CA	59.7	298	VAL	CA	60.03	307	THR	CA	61.88
289	TRP	CB	29.76	298	VAL	CB	32.58	307	THR	CB	69.57
289	TRP	N	121.88	298	VAL	N	122.3	307	THR	N	114
290	ASN	H	7.629	299	PRO	CA	63.05	308	ASP	H	8.407
290	ASN	HA	4.423	299	PRO	CB	32.08	308	ASP	HA	4.608
290	ASN	C	175.11	299	PRO	HA	4.438	308	ASP	C	176.1
290	ASN	CA	54.08	299	PRO	C	176.9	308	ASP	CA	54.6
290	ASN	CB	39.36	299	PRO	N	119.8	308	ASP	CB	41.03
290	ASN	N	114.58	300	ILE	H	8.375	308	ASP	N	122.9
291	LYS	H	7.471	300	ILE	HA	4.09	309	VAL	H	7.967
291	LYS	HA	4.292	300	ILE	C	176	309	VAL	HA	4.083
291	LYS	C	175.53	300	ILE	CA	61.28	309	VAL	C	175.6
291	LYS	CA	56.92	300	ILE	CB	38.92	309	VAL	CA	62.21
291	LYS	CB	33.36	300	ILE	N	121.55	309	VAL	CB	32.86

309	VAL	N	119.6	320	GLY	HA3	3.941	329	THR	HA	4.289
310	TRP	H	8.366	320	GLY	C	174.2	329	THR	C	174.3
310	TRP	HA	4.691	320	GLY	CA	45.46	329	THR	CA	62.19
310	TRP	C	175.6	320	GLY	N	108.6	329	THR	CB	69.9
310	TRP	CA	57.19	321	LYS	H	8.096	329	THR	N	115.1
310	TRP	CB	29.7	321	LYS	HA	4.298	330	ILE	H	8.294
310	TRP	N	125.8	321	LYS	C	176.8	330	ILE	HA	4.162
311	ILE	H	7.769	321	LYS	CA	56.3	330	ILE	C	176.2
311	ILE	HA	4.042	321	LYS	CB	33.01	330	ILE	CA	61.12
311	ILE	C	175.1	321	LYS	N	120.9	330	ILE	CB	38.4
311	ILE	CA	60.44	322	LYS	H	8.472	330	ILE	N	124.8
311	ILE	CB	39.16	322	LYS	HA	4.283	331	LEU	H	8.451
311	ILE	N	124.5	322	LYS	C	176.6	331	LEU	HA	4.39
312	GLU	H	8.282	322	LYS	CA	56.27	331	LEU	C	177
312	GLU	HA	4.124	322	LYS	CB	33.09	331	LEU	CA	54.86
312	GLU	C	176.1	322	LYS	N	123.2	331	LEU	CB	42.43
312	GLU	CA	56.32	323	GLU	H	8.468	331	LEU	N	127.2
312	GLU	CB	30.23	323	GLU	HA	4.23	332	ASP	H	8.303
312	GLU	N	125.4	323	GLU	C	175.8	332	ASP	HA	4.576
313	ILE	H	8.211	323	GLU	CA	56.38	332	ASP	C	175.98
313	ILE	HA	4.12	323	GLU	CB	30.33	332	ASP	CA	54.45
313	ILE	C	175.8	323	GLU	N	122.16	332	ASP	CB	41.43
313	ILE	CA	60.98	324	PHE	H	8.361	332	ASP	N	121.5
313	ILE	CB	39.15	324	PHE	HA	4.666	333	ASP	H	8.397
313	ILE	N	122.8	324	PHE	C	175.8	333	ASP	HA	4.588
314	ASP	H	8.527	324	PHE	CA	57.7	333	ASP	C	176.1
314	ASP	HA	4.613	324	PHE	CB	39.71	333	ASP	CA	54.32
314	ASP	C	176.4	324	PHE	N	121.1	333	ASP	CB	41.07
314	ASP	CA	54.19	325	SER	H	8.308	333	ASP	N	121.1
314	ASP	CB	41.19	325	SER	HA	4.432	334	ILE	H	8.146
314	ASP	N	124.7	325	SER	C	174.1	334	ILE	HA	4.164
315	MET	H	8.496	325	SER	CA	58.3	334	ILE	C	176.4
315	MET	HA	4.425	325	SER	CB	64.07	334	ILE	CA	61.1
315	MET	CA	55.8	325	SER	N	117.4	334	ILE	CB	38.78
315	MET	CB	32.72	326	ASN	H	8.543	334	ILE	N	121.3
315	MET	N	122.2	326	ASN	HA	4.725	335	GLU	H	8.567
316	ASP	HA	4.592	326	ASN	C	175.4	335	GLU	HA	4.305
316	ASP	C	175.7	326	ASN	CA	53.41	335	GLU	C	176.2
316	ASP	CB	41.17	326	ASN	CB	38.76	335	GLU	CA	56.32
316	ASP	CA	54.4	326	ASN	N	120.9	335	GLU	CB	30.39
317	ASP	H	8.129	327	MET	H	8.385	335	GLU	N	125.6
317	ASP	HA	4.81	327	MET	HA	4.492	336	ASP	H	8.414
317	ASP	C	175.1	327	MET	C	176.2	336	ASP	HA	4.58
317	ASP	CA	52.15	327	MET	CA	55.6	336	ASP	C	175.8
317	ASP	CB	41.48	327	MET	CB	32.57	336	ASP	CA	54.38
317	ASP	N	121.7	327	MET	N	120.4	336	ASP	CB	41.39
318	PRO	N	118.7	328	ASP	H	8.452	336	ASP	N	122.1
319	LYS	HA	4.244	328	ASP	HA	4.631	337	ASP	H	8.336
319	LYS	C	177.7	328	ASP	C	176.3	337	ASP	HA	4.58
319	LYS	CB	32.24	328	ASP	CA	54.73	337	ASP	C	176
319	LYS	CA	56.81	328	ASP	CB	41.07	337	ASP	CA	54.43
320	GLY	H	8.177	328	ASP	N	121	337	ASP	CB	41.12
320	GLY	HA2	3.941	329	THR	H	8.101	337	ASP	N	120.8

338	ILE	H	7.982	346	GLU	N	120.4	356	ASP	H	8.391
338	ILE	HA	4.044	347	ASN	H	8.438	356	ASP	HA	4.225
338	ILE	C	175.6	347	ASN	HA	4.98	356	ASP	CA	54.32
338	ILE	CA	61.18	347	ASN	C	172.9	356	ASP	CB	41.08
338	ILE	CB	38.71	347	ASN	CA	51.32	356	ASP	N	121.3
338	ILE	N	120.8	347	ASN	CB	39.06	357	HIS	HA	4.6
339	TYR	H	8.27	347	ASN	N	120.2	357	HIS	C	174.8
339	TYR	HA	4.532	348	PRO	CA	63.33	357	HIS	CB	29.3
339	TYR	C	175	348	PRO	CB	32.23	357	HIS	CA	56.13
339	TYR	CA	57.68	348	PRO	HA	4.483	358	ASN	H	8.551
339	TYR	CB	38.9	348	PRO	C	177.1	358	ASN	HA	4.653
339	TYR	N	124.6	348	PRO	N	116.9	358	ASN	C	175.1
340	TYR	H	7.962	349	SER	H	8.507	358	ASN	CA	53.55
340	TYR	HA	4.49	349	SER	HA	4.458	358	ASN	CB	38.81
340	TYR	C	174.6	349	SER	C	175	358	ASN	N	119.2
340	TYR	CA	57.5	349	SER	CA	58.23	359	LYS	H	8.204
340	TYR	CB	39.3	349	SER	CB	63.94	359	LYS	HA	4.327
340	TYR	N	123.2	349	SER	N	116.4	359	LYS	C	176.4
341	ASP	H	8.355	350	VAL	H	8.235	359	LYS	CA	56.4
341	ASP	HA	4.54	350	VAL	HA	4.174	359	LYS	CB	32.97
341	ASP	C	176	350	VAL	C	175.9	359	LYS	N	121.8
341	ASP	CA	53.86	350	VAL	CA	62.21	360	VAL	H	8.199
341	ASP	CB	41.44	350	VAL	CB	32.77	360	VAL	HA	4.137
341	ASP	N	123.2	350	VAL	N	121.5	360	VAL	C	175.7
342	VAL	H	8.117	351	ASP	H	8.363	360	VAL	CA	62.13
342	VAL	HA	4.076	351	ASP	HA	4.598	360	VAL	CB	32.94
342	VAL	C	175.9	351	ASP	C	175.9	360	VAL	N	121.4
342	VAL	CA	62.38	351	ASP	CA	54.43	361	ASP	H	8.506
342	VAL	CB	32.65	351	ASP	CB	41.15	361	ASP	HA	4.628
342	VAL	N	121	351	ASP	N	122.8	361	ASP	C	175.7
343	ASN	H	8.589	352	ASN	H	8.329	361	ASP	CA	54.24
343	ASN	HA	4.731	352	ASN	HA	4.651	361	ASP	CB	41.29
343	ASN	C	175.1	352	ASN	C	174.8	361	ASP	N	124.7
343	ASN	CA	53.34	352	ASN	CA	53.22	362	VAL	H	8.154
343	ASN	CB	39.26	352	ASN	CB	38.9	362	VAL	HA	4.391
343	ASN	N	122.2	352	ASN	N	118.7	362	VAL	C	174.4
344	ASP	H	8.359	353	ILE	H	8.126	362	VAL	CA	59.99
344	ASP	HA	4.601	353	ILE	HA	4.419	362	VAL	CB	32.55
344	ASP	C	176.2	353	ILE	C	174.5	362	VAL	N	122.4
344	ASP	CA	54.5	353	ILE	CA	58.84	363	PRO	N	119.7
344	ASP	CB	41.15	353	ILE	CB	38.6	364	LYS	C	176.7
344	ASP	N	121.8	353	ILE	N	123	365	LYS	H	8.451
345	ASP	H	8.338	354	PRO	CA	63.2	365	LYS	HA	4.32
345	ASP	HA	4.531	354	PRO	CB	32.11	365	LYS	C	176.4
345	ASP	C	176.5	354	PRO	HA	4.397	365	LYS	CA	56.13
345	ASP	CA	54.71	354	PRO	C	177	365	LYS	CB	33.14
345	ASP	CB	41.07	354	PRO	N	120.2	365	LYS	N	123.4
345	ASP	N	120.9	355	MET	H	8.538	366	VAL	H	8.295
346	GLU	H	8.325	355	MET	HA	4.409	366	VAL	HA	4.087
346	GLU	HA	4.229	355	MET	C	175.9	366	VAL	C	175.8
346	GLU	C	176.1	355	MET	CA	55.55	366	VAL	CA	62.13
346	GLU	CA	56.6	355	MET	CB	33.14	366	VAL	CB	32.97
346	GLU	CB	30.27	355	MET	N	121.2	366	VAL	N	122.4

367	HIS	H	8.691	374	ASN	CB	38.87	382	GLU	C	176.5
367	HIS	HA	4.683	374	ASN	N	119.9	382	GLU	CA	56.77
367	HIS	C	174.5	375	ASN	H	8.528	382	GLU	CB	30.11
367	HIS	CA	55.77	375	ASN	HA	4.732	382	GLU	N	121.1
367	HIS	CB	29.93	375	ASN	C	175.7	383	PRO	N	121.2
367	HIS	N	124.2	375	ASN	CA	53.2	387	ILE	HA	4.199
368	VAL	H	8.265	375	ASN	CB	38.88	387	ILE	C	176.5
368	VAL	HA	4.057	375	ASN	N	119.6	387	ILE	CB	38.77
368	VAL	C	175.6	376	THR	H	8.264	387	ILE	CA	61.34
368	VAL	CA	62.04	376	THR	HA	4.405	388	SER	H	8.403
368	VAL	CB	33.04	376	THR	C	175	388	SER	HA	4.464
368	VAL	N	124	376	THR	CA	62.01	388	SER	C	174.1
369	GLU	H	8.631	376	THR	CB	69.8	388	SER	CA	58.08
369	GLU	HA	4.261	376	THR	N	114	388	SER	CB	63.99
369	GLU	C	176.2	377	SER	H	8.447	388	SER	N	119.8
369	GLU	CA	56.19	377	SER	HA	4.483	389	ASP	H	8.425
369	GLU	CB	30.36	377	SER	C	174.6	389	ASP	HA	4.563
369	GLU	N	125.5	377	SER	CA	58.53	389	ASP	C	176.3
370	MET	H	8.55	377	SER	CB	63.76	389	ASP	CA	54.36
370	MET	HA	4.454	377	SER	N	118	389	ASP	CB	41.05
370	MET	C	176	378	ASN	H	8.54	389	ASP	N	123.02
370	MET	CA	55.34	378	ASN	HA	4.769	390	VAL	H	7.943
370	MET	CB	32.89	378	ASN	C	175.4	390	VAL	HA	4.023
370	MET	N	123.3	378	ASN	CA	53.38	390	VAL	C	175.8
371	LYS	H	8.469	378	ASN	CB	38.79	390	VAL	CA	62.5
371	LYS	HA	4.321	378	ASN	N	120.9	390	VAL	CB	32.5
371	LYS	C	176.2	379	GLY	H	8.448	390	VAL	N	118.8
371	LYS	CA	56.24	379	GLY	HA2	3.978	391	TRP	H	8.115
371	LYS	CB	33.03	379	GLY	HA3	3.978	391	TRP	HA	4.727
371	LYS	N	123.8	379	GLY	C	174.4	391	TRP	C	175.8
372	ILE	H	8.308	379	GLY	CA	45.53	391	TRP	CA	56.89
372	ILE	HA	4.127	379	GLY	N	109.5	391	TRP	CB	29.68
372	ILE	C	176.2	380	SER	H	8.244	391	TRP	N	123.5
372	ILE	CA	60.98	380	SER	HA	4.452	392	ASN	H	8.189
372	ILE	CB	38.44	380	SER	C	174.9	392	ASN	HA	4.71
372	ILE	N	123.2	380	SER	CA	58.42	392	ASN	C	173.8
373	LEU	H	8.432	380	SER	CB	63.88	392	ASN	CA	53.13
373	LEU	HA	4.388	380	SER	N	115.8	392	ASN	CB	38.9
373	LEU	C	176.9	381	LEU	H	8.427	392	ASN	N	121.3
373	LEU	CA	54.89	381	LEU	HA	4.353	393	ILE	H	7.596
373	LEU	CB	42.4	381	LEU	C	177.7	393	ILE	HA	4.037
373	LEU	N	127.04	381	LEU	CA	55.51	393	ILE	C	181.1
374	ASN	H	8.513	381	LEU	CB	42.22	393	ILE	CA	62.94
374	ASN	HA	4.705	381	LEU	N	124.1	393	ILE	CB	39.58
374	ASN	C	174.9	382	GLU	H	8.413	393	ILE	N	124.8
374	ASN	CA	53.19	382	GLU	HA	4.221				

Appendix V

RANDOM COIL CHEMICAL SHIFTS

Wishart, Bigam, Holm, Hodges, Sykes (1995) J. Bio. NMR., 5, 67-81.

		Hn	N	Ca	Cb	Co	Ha	Hb		
ALA	A	8.24	123.8	52.5	19.1	177.8	4.32	1.39	0.00	
CYS	c	8.32	118.8	58.2	28.0	174.6	4.55	2.93	2.93	red
CYS	C	8.43	118.6	55.4	41.1	174.6	4.71	3.25	2.99	oxi
ASP	D	8.34	120.4	54.2	41.1	176.3	4.64	2.72	2.65	
GLU	E	8.42	120.2	56.6	29.9	176.6	4.35	2.06	1.96	
PHE	F	8.30	120.3	57.7	39.6	175.8	4.62	3.14	3.04	
GLY	G	8.33	108.8	45.1	-	174.9	3.96	-	-	
HIS	H	8.42	118.2	55.0	29.0	174.1	4.73	3.29	3.16	
ILE	I	8.00	119.9	61.1	38.8	176.4	4.17	1.87	0.00	
LYS	K	8.29	120.4	56.2	33.1	176.6	4.32	1.84	1.75	
LEU	L	8.16	121.8	55.1	42.4	177.6	4.34	1.62	1.62	
MET	M	8.28	119.6	55.4	32.9	176.3	4.48	2.11	2.01	
ASN	N	8.40	118.7	53.1	38.9	175.2	4.74	2.83	2.75	
PRO	P	0.00	0.0	63.3	32.1	177.3	4.42	2.29	1.94	
GLN	Q	8.32	119.8	55.7	29.4	176.0	4.34	2.12	1.99	
ARG	R	8.23	120.5	56.0	30.9	176.3	4.34	1.86	1.76	
SER	S	8.31	115.7	58.3	63.8	174.6	4.47	3.89	3.87	
THR	T	8.15	113.6	61.8	69.8	174.7	4.35	4.24	0.00	
VAL	V	8.03	119.2	62.2	32.9	176.3	4.12	2.08	0.00	
TRP	W	8.25	121.3	57.5	29.6	176.1	4.66	3.29	3.27	
TYR	Y	8.12	120.3	57.9	38.8	175.9	4.55	3.03	2.98	