Small angle scattering studies of chromatin proteins in the human malarial parasite

Ashley Jordan

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy December 2017

Institute for Science & Technology in Medicine

Keele University

ABSTRACT

This thesis describes the biochemical, biophysical and structural characterisation of two proteins believed to play an important role in active gene silencing mechanisms present in the human malarial parasite, *Plasmodium falciparum*. These investigations were performed using the histone deacetylase enzyme, PfSir2a, and the DNA binding protein, PfAlba3. The initial goal of this PhD project was to obtain structural information on both PfSir2a and PfAlba3, as well as the proposed silencing complex thought to be formed by the two proteins. This information would then aid the development of novel pharmaceuticals with a perspective towards new therapeutics to combat the continued threat of malaria.

Thorough biochemical and biophysical characterisation of both PfSir2a and PfAlba3 was conducted and is described in Chapter 3 of the thesis. These results could not characterise a strong interaction between PfSir2a and PfAlba3 and highlighted several properties exhibited by the proteins that, as a result, proved challenging to characterise by structural methods. Most notably concentration dependent oligomerisation and protein aggregation effects were observed for PfSir2a and PfAlba3 respectively.

Nevertheless, structural studies were performed using SAXS and SANS techniques to investigate the individual proteins to obtain structural information about the solution state of PfSir2a and PfAlba3 and generate *ab initio* models for both proteins and these are presented in Chapter 4. The difficulties presented by the target proteins aided in the development of a new investigative method for bio-SANS experiments, SEC-SANS, and the first example of testing and validation of this technique is presented in Chapter 5. The findings of work conducted for this thesis are summarised in Chapter 6 with an outlook for future work and development.

Table of Contents

Acknowledgements

First and foremost I would like to thank my supervisors, Dr Catherine Merrick, Prof. Trevor Forsyth, Prof. Edward Mitchell and Dr. Michael Haertlein for their continued support, advice, and, what I can only describe as extreme patience! I consider myself very fortunate to have had a supervisory team with such an extensive range of expertise. I am very grateful for the opportunities which have been afforded to me throughout my undertaking of this project. I would particularly like to thank you for all your comments and revisions of this manuscript.

I would also like to extend my thanks to those who have helped me throughout the course of this thesis work. Firstly I would like to thank Dr. Juliette Devos for her help and advice almost every day, as well as her endless enthusiasm which always made it a fun and exciting time spent in the lab! Merci Juliette! Thanks also go to Martine and Valerie, both of you have given me valuable help and advice throughout this project which saved me from tearing my hair out at several stages! A big thank you also goes to the rest of the D-lab members for their help at various stages during this thesis, it has been an absolute pleasure to work with you all and share in adventures outside of the lab as well! I am also grateful to Dr. Anne Martel and Dr. Lionel Porcar, co-responsibles of the ILL beamline D22 for their help in performing my experiments and the data analysis.

Moving onto friends and family, I'd like to say a big thank you to everyone back home in Leicester. Pete, Charlotte, Sam, Kim, Harry and Tom, it's always been good to come home and feel like I haven't been anywhere at all! I'd like to say thanks to Matt as well for providing inspirational music to listen to whilst writing! Special mention goes to Chris for being the best cousin ever and always being able to kick back just like old times! Next up

is a mention to all my fellow Tumbleweeds for providing the best discussions and laughs after work. “Citizens of Dalaran! Raise your eyes to the skies and observe!”

Lastly I’d like to give my most heart warmed thanks to my family for all their love and support. Parents and grandparents alike, you have all helped me over and over again throughout the years. I’d like a huge thank you to be said to Hannah for being the best sister a brother could ever have hoped for!

Mum, Dad, the last few years have perhaps been tougher than most, but your everlasting support has helped in more ways than I can begin to count. I can’t even begin to imagine how I can possibly repay you for all that you have done for me. I can only hope that by achieving this you can be proud of me, as proud as I am, to be your son.

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 4

Chapter 5

Glossary:

A – Amps

A – measured UV absorbance

A_2 - second virial coefficient

A260 – Absorption 260 nm

ACN – acetonitrile

ACS - acetyl-CoA synthase

ACTs – Artemisinin combination therapies

ADP – Adenosine diphosphate

ATS - acidic terminal segment

AUC - Analytical Ultracentrifugation

BL21 (DE3) – *E. coli* bacterial expression strain

β-ME - beta-mercaptoethanol

bp – base pairs

c - protein concentration (g ml^{-1})

°C – Degrees Celsius

cDNA – Copy DNA

CIDRs - Cysteine-rich InterDomain Regions

cm – centimetre

CSA - chondroitin sulphate A

CV – Column volumes

DBL - Duffy Binding Like

ddH_2O – double distilled H_2O

DLS – Dynamic light scattering

D.M.M – deuterated minimal media

DNA - deoxyribonucleic acid

dn/dc – refractive index increment

dNTPs - deoxynucleoside triphosphates

DP – Differential power

dRI - change in refractive index

dsDNA – double stranded DNA

D_T - translational diffusion coefficient

DTT – Dithiothreitol

ε - molar absorption coefficient in L mol^{-1}

EDTA - Ethylenediaminetetraacetic acid

f - frictional coefficient

F_b – Buoyancy force

F_f – Frictional force

f_n – neutron atomic scattering length

f_p – atomic nuclear potential

FPLC - fast protein liquid chromatography

f_s – atomic spin

F_s – Sedimentation force

FT – Flow through

f_x - X-ray atomic scattering length

ΔG – Gibbs free energy

g – relative centrifugal force

GF = Gel filtration

ΔH – Enthalpy

HDACs - histone deacetylases

HF – High fidelity

HPLC – High performance liquid chromatography

ICAM1 - intercellular adhesion molecule 1

IEX – Ion exchange chromatography

IF3 -Bacterial translation initiation factor 3

IHF - Integration Host Factor

IMAC - immobilised metal affinity chromatography

IR – Infrared

IPTG - isopropyl β-D-1-thiogalactopyranoside

ITC - Isotheral Titration Calorimetry

J – Joules

K_A – binding affinity

k_B - Boltzmann's constant

K_B - binding constant

K_D - constant of dissociation

keV – kilo electronvolt

k_i – incident monochromatic wave vector

k_f – scattered wave vector

kV – kilo volt

L – Litre

l – path length

LB – Luria Broth

λ – wavelength

λ_0 – vacuum wavelength of incident light

M – Molar

m – mass of particle

mg – milli gram

Milli-Q H_2O – Milli-Q system filtered H_2O

mM – milli molar

MM – Molar Mass

M.M – minimal media

MRM – most representative model

MST - Micro-scale Thermophoresis

M_w - weight averaged molar mass (g mol^{-1})

MW – Molecular weight (kDa)

η – viscosity

η_0 – solvent refractive index

N_A – Avogadro's number

NAD^+ - nicotinamide adenine dinucleotide

ng – nano grams

Ni-NTA – Nickel-Nitrilotriacetic acid

nm – nano meter

NTS- N-terminal segments

OAADPr - O-acetyl-ADP ribose

$O.D_{600}$ – Optical Density 600 nm

ORF – Open reading frame

ρ – solvent density g ml^{-1}

PBS - phosphate buffered saline solution

PBST - phosphate buffered saline solution with Tween®-20

PCR – Polymerase chain reaction

PCS - photon correlation spectroscopy

PDB – Protein Data bank

PfEMP1 - *Plasmodium falciparum* erythrocyte membrane protein 1

PfSir2a – *Plasmodium falciparum* silencing information regulator protein 2a

PfAlba3 - *Plasmodium falciparum* acetylation lowers binding affinity protein 3

pH – potential of hydrogen

ρ(r) - scatter length density distribution

ρ_s - scatter length density

q - scattering vector

QELS - quasi-elastic light scattering

qPCR – quantitative polymerase chain reaction

r – distance from rotor

R – Gas constant

r_o – Thomson radius for an electron

R_θ - Rayleigh ratio

RBCs - Red blood cells

R_f - relative migration

R_H - hydrodynamic radius

RID - refractive index detector

RNA – Ribonucleic acid

rpm – revolutions per minute

RT – Room temperature (21°C)

s – sedimentation coefficient

ΔS - Entropy

SAS- Small angle scattering

SAXS – Small angle X-ray scattering

SANS – Small angle neutron scattering

SDM - site-directed mutagenesis

SDS-PAGE - Sodium Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis

SEC – Size exclusion chromatography

SEC-MALLS-RI - Size Exclusion Chromatography Coupled with Multi Angle Laser Light Scattering and Refractive Index

SOC - super optimal broth with catabolite repression

ssDNA – single stranded DNA

S_T - Soret coefficient

T – Temperature

TAE - tris/acetate/EDTA

TAREs - telomere associated repetitive elements

TBE - tris/borate/EDTA

TFA - Trifluoroacetic acid

T_m – Melting temperature

TM – Transmembrane

TOF -MS - Time of Flight Mass Spectrometry

TSA - Fluorescence thermal shift assay

TSP – thrombospondin

TT – Tris-tricine

u – velocity

U – Units

µF – micro Farad

µg – micro gram

µl – micro litre

µm – micro meter

UV – Ultra violet

$\bar{v}$ - partial specific volume of particle in $cm^3\ g^{-1}$

V – Volts

ω - rotor speed in rad s^{-1}

XL10 Gold - *E. coli* bacterial transformation strain

Z – atomic number

Chapter 1: Introduction

1.1 An Introduction to Malaria

Human malaria is a disease caused by infection by parasitic protozoan organisms which presents an incredible burden upon global public health. It has been long since known to be a severe disease with accounts of malaria dating back to 2700 B.C (Cox, 2010). It is prevalent in the tropical and sub-tropical regions of the globe with recent figures estimating around 40% of the world's population being currently at risk of contracting malaria (see Figure 1.1). There are an estimated 198 million cases of infection and over 584,000 deaths per year (WHO, 2014). The majority of infections and mortality associated with malaria occurs in the sub-Saharan Africa region with young children and immunocompromised adults being most affected.

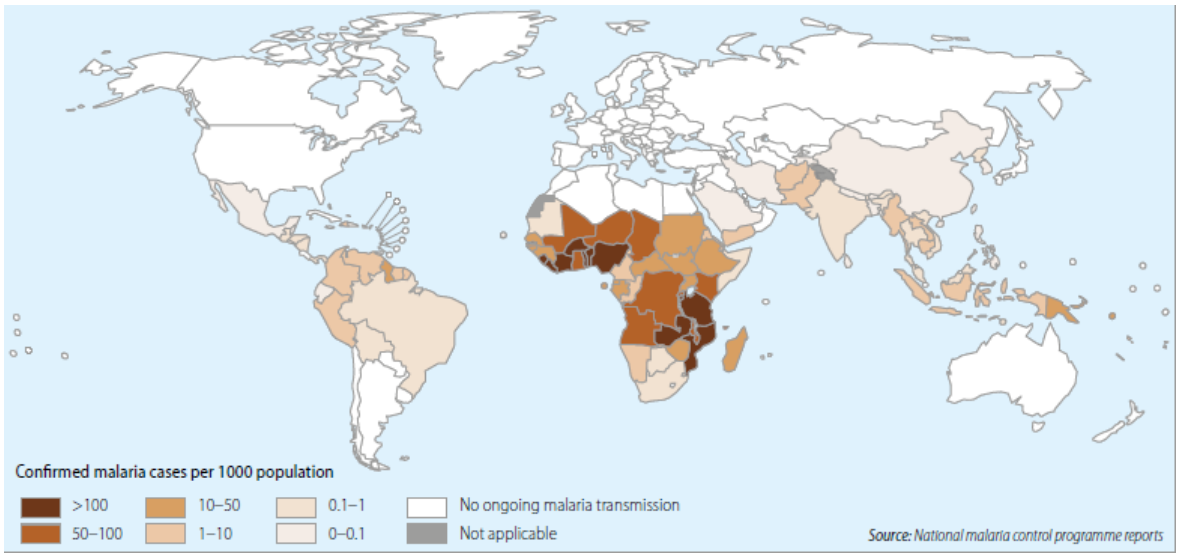


Figure 1.1: Global map highlighting countries with ongoing malaria transmission (WHO, 2014).

The causative organisms are parasitic protozoans of the *Plasmodium* genus belonging to the apicomplexan phylum. There are five main species which are known to infect humans,

P. falciparum, P. malariae, P. ovale, P. vivax and P. knowelsii. *P. falciparum* is the organism which is predominantly located in sub-Saharan Africa and accounts for over 90% of the total deaths associated with malarial infection. This is thought to be due to a combination of the parasites intra-erythrocytic stage during the life cycle and the inability of infected individuals with young, or compromised, immune systems to cope with high parasite burdens (Schumacher and Spinelli, 2012).

Infection with malaria occurs during the blood feeding behaviour by parasite-infected female mosquito hosts from the *Anopheline* genus. Sporozoite stage parasites are transferred from the salivary glands of the mosquito into the human host and rapidly migrate to the liver to invade hepatocytes and undergo asexual reproduction to develop into the merozoite stage (see Figure 1.2).

The merozoites break out of the liver cells and into the blood stream of the host where they attach to and invade circulating red blood cells (RBCs), or erythrocytes. Here they continue the next stage in the parasite life cycle by undergoing asexual reproduction into more merozoites to begin another cycle of re-invasion, or develop into the sexual stage gametocytes which are transferred back to the mosquito hosts during blood feeding to complete the life cycle (Trampuz *et al.*, 2003).

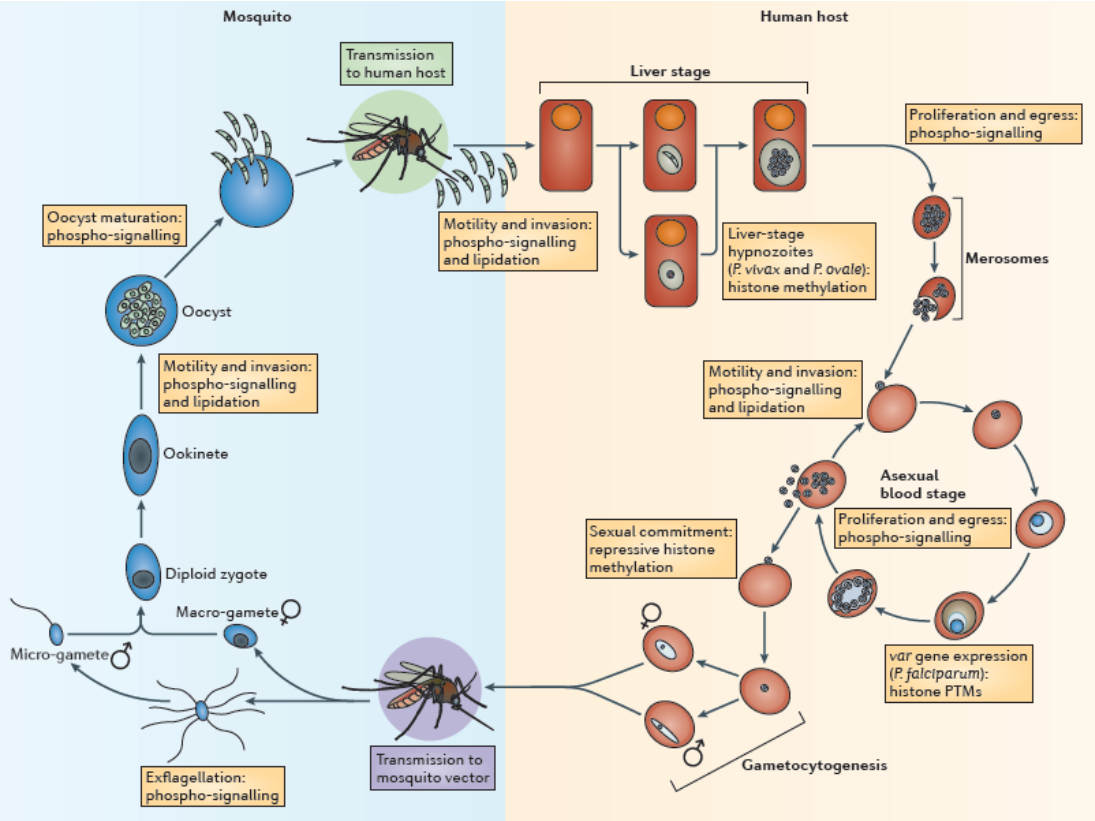


Figure 1.2: Life cycle stages and development of *P. falciparum* parasite in human and mosquito hosts (Doerig *et al.*, 2015).

The duration of this invasion cycle requires only 48 hours to complete and allows for malaria parasites to rapidly increase in number in a very short period of time. It is during this intra-erythrocytic stage that all the symptoms associated with severe malarial infection occur. The high levels of parasitaemia can cause the rapid onset of flu-like symptoms such as fever, nausea and headaches. Individuals who are unable to mount a sufficient immune response cannot control the rapid parasite development and life threatening levels of parasitaemia, between 2 – 5% (equivalent to 100,000 – 250,000 parasites per µl of blood) are reached (WHO, 2010). This results in more serious symptoms such as severe anaemia, clotting, metabolic acidosis and cerebral malaria which even if treated rapidly have a mortality rate of 15-20% (Trampuz *et al.*, 2003; Miller *et al.*, 2013).

Antimalarial drug treatments such as artemisinin combination therapies (ACTs) exist to treat infected individuals and remove parasites from the body. Although these treatments are very effective, *Plasmodium* is well known for rapidly developing resistance to drug treatments used against it. This was especially highlighted when resistance to chloroquine, a long-standing drug treatment for malarial infections, was detected in *P. falciparum* strains in Southeast Asia in the late 1950s. This later was detected in Africa in the 1970s and spread rapidly across the continent within a decade (Wellems and Plowe, 2001). Recent findings have indicated the emergence of resistance to artemisinin and highlight the need for developing additional treatment methods to combat malaria (Ariey *et al.*, 2014).

While drug treatments remain the cornerstone method for targeting malaria parasites in infected individuals, it is important to continue finding suitable candidates to combat the challenge of resistance development. Resistance to antimalarials is attributed to spontaneous and random changes to genes during replication. If a mutation which confers resistance to an antimalarial drug occurs then a selection pressure is created that favours resistance development. Most drugs developed so far against *Plasmodium* target the toxic products of the haemoglobin digestion pathway that is used by parasites to provide amino acids for protein synthesis and parasite development (Bozdech *et al.*, 2003; Miller *et al.*, 2013). Additional avenues of research are focused on developing drugs which are able to interrupt another key area needed for parasite survival: immunoevasion.

1.2 Antigenic Variation: A Complex Problem

By invading host erythrocytes, the parasites are essentially hidden from a large portion of the humoral and physical components of the human defence system which other pathogens would be subjected to, e.g. antibody response. But RBCs are relatively short-lived cells inside the human body, lasting approximately 100-120 days in circulation, and senescent

cells are destroyed by the spleen which metabolizes the contained haemoglobin into its constituent amino acids to be recycled. This process also destroys any malaria parasites that are contained within the erythrocytes because they cannot develop outside RBCs, requiring the metabolism of haemoglobin to survive.

As a result, malaria parasites have developed a unique method of evading host immune responses to avoid destruction and continue the life cycle. To reduce the chance of destruction before malaria parasites can develop inside an infected erythrocyte they produce a polymorphic protein, *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1). This protein is exported to the cell surface membrane of infected RBCs where it serves to bind to host endothelial cell surface receptors, as well as uninfected erythrocytes, to prevent infected RBCs from passing through host capillaries effectively causing clotting. This slows progress towards the spleen and helps to ensure that parasite have sufficient time to go through the intra-erythrocytic development cycle.

PfEMP1 is a relatively large protein, of approximately 200-305kDa in size, and cytoadherence is mediated by a diverse number of Duffy Binding Like (DBL) domains and Cysteine-rich InterDomain Regions (CIDRs) present in the protein sequence. These are able to bind several host receptors, such as CD-36, intercellular adhesion molecule 1 (ICAM1), thrombospondin (TSP) and chondroitin sulphate A (CSA), with varying affinity (K_D values ranging from 12nM to 344μM) (Barnwell *et al.*, 1989; Berendt *et al.*, 1989; Smith, Chitnis and Craig, 1995; Srivastava *et al.*, 2010; Buckee and Recker, 2012; Turner *et al.*, 2013).

The variability in size of PfEMP1and its affinity for host receptors is a result of several different polymorphs of the protein that are encoded by a family of clonally variant *var* genes, with approximately 60 identified variants (see Figure 1.3) (Baruch *et al.*, 1995;

Smith, Chitnis and Craig, 1995; Su *et al.*, 1995; Gardner *et al.*, 2002). PfEMP-1 variants which preferentially bind to ICAM1 and CSA endothelial cell surface receptors are linked more strongly with severe symptoms including cerebral malaria, metabolic acidosis, low birth weight, prematurity and high mortality (Penman and Gupta, 2008).

Figure 1.3: Binding domain and architectures of PfEMP-1 polymorphs. PfEMP-1 proteins have multiple domains; N-terminal segments (NTS), Duffy Binding Like (DBL), Cysteine-rich InterDomain Regions (CIDRs), C2, transmembrane (TM) and acidic terminal segment (ATS) (Kraemer and Smith, 2006).

The expression of *var* genes in *P. falciparum* is tightly controlled, with only one full length transcript being expressed at a time, while the remaining genes are placed in a transcriptionally silent state (Scherf *et al.*, 1998). The majority of *var* genes, approximately 60%, are located in subtelomeric regions of chromosomes that are downstream of six

distinct repetitive deoxyribonucleic acid (DNA) sequences known as telomere associated repetitive elements (TAREs 1-6) (Freitas-Junior *et al.*, 2005). These regions are also populated by other multigene families such as *stevor* and *rifin* which encode other smaller erythrocyte membrane proteins (Scherf, Lopez-Rubio and Riviere, 2008). *Var* genes are divided into three main groups, A, B and C, which are classified by similarities in their chromosomal location and upstream 5' promoter regions, UpsA, UpsB and UpsC (S. A. Kyes, Kraemer and Smith, 2007).

UpsA and UpsB *var* genes are located in the subtelomeric chromosome regions whereas UpsC *var* genes are located towards the centre. UpsA and UpsB *var* genes differ in the way in which they are transcribed with UpsA genes transcribed towards and UpsB transcribed away from the telomere respectively (Kyes et al., 2007). Studies have shown that severe malaria in children is associated with increased transcripts from *var* genes belonging to groups A and B (Jensen *et al.*, 2004; Kyriacou *et al.*, 2006; Rottmann *et al.*, 2006; Merrick *et al.*, 2012). Initially, immune selection pressure and host cell polymorphisms can select for parasites expressing *var* genes that make them better able to survive inside the host bloodstream. But as a host begins to gain naturally acquired immunity, the repertoire of variant surface antigens will become reduced and result in lower parasite survival rates (Nielsen *et al.*, 2002).

Thus the benefit of mutually exclusive *var* gene expression is that it enables parasites to restrict the surface antigens that become exposed to the host immune system, which can develop an antibody response against the parasite-infected erythrocytes. In addition, malaria parasites seem to be able to switch which particular *var* gene is expressed via a process known as 'antigenic variation' and this allows parasites to sustain chronic infection with successive waves of parasitaemia (Scherf, Lopez-Rubio and Riviere, 2008). This

ability to cause recurring infections is indicative of a well-developed and sophisticated immune evasion strategy (Scherf, Lopez-Rubio and Riviere, 2008).

The mechanism that gives rise to this antigenic variation is not yet fully understood, but is thought to arise through changes made to the local nuclear architecture, instead of genetic recombination into a conserved expression site as is seen in another protozoan parasite, *Trypanosoma brucei,* that also utilises antigenic variation (Duraisingh *et al.*, 2005; Freitas-Junior *et al.*, 2005; Lopez-Rubio *et al.*, 2007). Control of transcription of active genes in this manner is believed to be governed by recognition of, and changes to, posttranslational modifications present on parasite chromatin. This is known as 'epigenetic regulation' and has become an active area of research in the last decade to try to better understand the processes involved in this complex regulatory mechanism (Merrick and Duraisingh, 2006; Scherf, Lopez-Rubio and Riviere, 2008; Guizetti and Scherf, 2013).

1.3 Chromatin and Epigenetic Regulation

Organisms across all three kingdoms of life face the complication of containing their genomic DNA within individual cells in order to function. For bacteria and archaea this means accommodating their entire genome, that when laid out as single linear piece of DNA has an approximate end to end length of 1mm, inside a cell only 1μm in diameter. This requires a compaction factor of roughly 10^4 to fit the genome inside the cell alongside the rest of the cellular components (de Vries, 2010). Eukaryotes face a more daunting task of requiring roughly two metres of stretched out DNA to fit inside a nucleus only 10μm in diameter (White and Bell, 2002). Therefore a high degree of nuclear organisation is required to be able to physically fit all the necessary genetic information required for development and continued existence into the cell. This is achieved by utilising a set of

chromosomal proteins, called histones, whose purpose is to package DNA into compact, highly ordered structures.

Euryarchaea and eukaryotes contain proteins which possess a distinctive α-helical fold, known as the histone fold. Approximately 140-200 base pairs (bp) of DNA becomes wrapped around two copies of four histone protein subunits: H2A, H2B, H3 and H4 forming a octamer core nucleosome structure which leads to the compaction of the genomic DNA (Arents and Moudrianakis, 1995; Ramakrishnan, 1997). A fifth linker histone, H1, binds to the nucleosome and forms the first level of organisation, the 10nm fibre 'beads on a string' array (Mariño-Ramírez and Kann, 2005a). This allows for the compaction of meter sized lengths of genomic DNA to become compacted into a region only a few microns across (Ramakrishnan, 1997). Further compaction via octamer-octamer interactions leads to the next level of organisation, the 30nm fibre (Tremethick, 2007a).

Core histones are globular in structure and amongst the most well conserved proteins across eukaryotes due to their essential role in DNA packaging and gene regulation (Mariño-Ramírez and Kann, 2005b). Variations between histone structures exist but all share the common histone fold motif (Arents and Moudrianakis, 1995). This allows for histone proteins to interlock together in a 'handshake' like fashion to form the octamer core of the nucleosome and bind DNA non-specifically (Arents and Moudrianakis, 1995; Ramakrishnan, 1997). In addition to the globular C-terminal regions, histones also possess N-terminal 'tail-like' regions which protrude from the nucleosome and may comprise up to 25-30% of the mass of the histone (see Figure 1.4) (Wolffe and Hayes, 1999; Villar-Garea and Imhof, 2006).

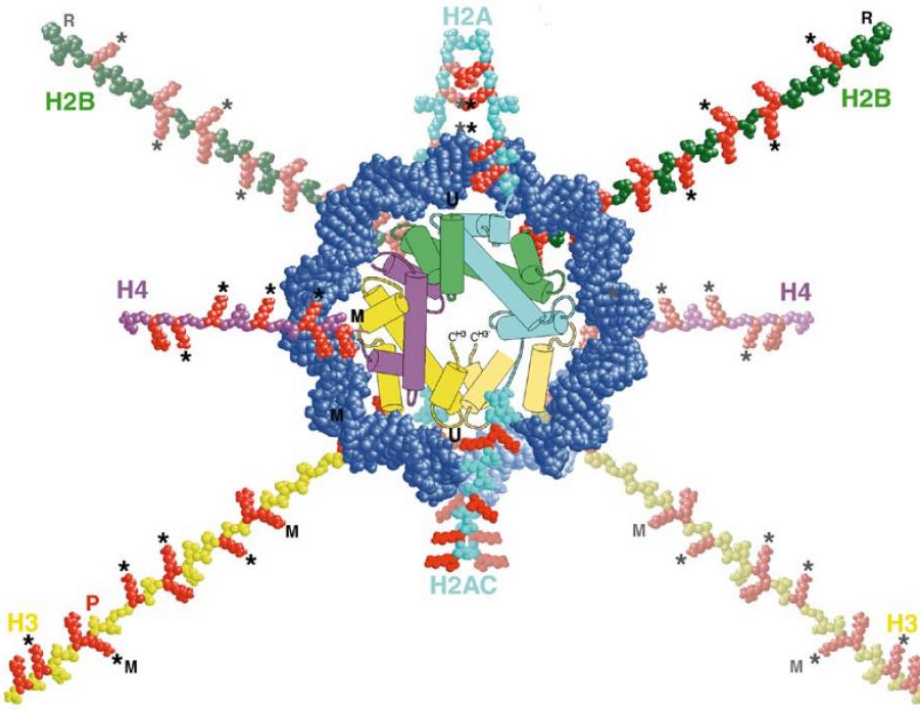


Figure 1.4: Diagram depicting histone packing of chromatin DNA. The histone proteins comprise the central core and have protruding N-terminal tails. The DNA is labelled in blue and wrapped around the octamer core comprised of the four protein subunits (Wolffe and Hayes, 1999).

Posttranslational modifications such as methylation, acetylation, alkylation, glycosylation and oxidation are a major cellular mechanism used in signalling and the control of protein activity within the cell. These occur largely through enzyme-catalysed covalent addition of an electrophilic fragment from a co-substrate to a side chain residue of the protein being modified (Walsh, Garneau-Tsodikova and Gatto, 2005). N-terminal tail regions are no exception to this and have a high capacity for posttranslational modifications. They are highly conserved, much like the globular C-terminal regions, and play an important role in nucleosome structure and function (Wolffe and Hayes, 1999; Tremethick, 2007b). The tail region of subunit H3 contains approximately 38 amino acids, 19 of which can be potentially modified (Villar-Garea and Imhof, 2006). It is also known that histone tails are

important for viability in yeast, as removal of both H3 and H4 subunit tail regions results in a lethal event (Ling *et al.*, 1996).

The ability of histone tails to be modified in such a wide and versatile manner has led to the formation of the idea that there is a histone 'language' which can be read by other proteins or protein complexes (Strahl and Allis, 2000). Commonly referred to as the histone code, this mechanism of signalling and control is thought to play a crucial role in the epigenetic regulation of gene activity. This occurs via changes to the local nuclear architecture instead of the encoding DNA sequence itself, leading to regions of transcriptionally active euchromatin, or 'silenced' heterochromatin (Ramakrishnan, 1997).

One well studied example of posttranslational modification of proteins is the addition of acetyl groups and this has been widely implicated in regulating gene expression (Struhl, 1998). Histones which are acetylated facilitate the transcription of active genes, while under-acetylated histones are correlated with repressed transcription (Braunstein *et al.*, 1993). Acetyl groups are typically found on lysine residues within core histone tail regions. Histone subunits H3 and H4 are particularly enriched in lysine residues and have been shown to be readily interact with and become modified by acetyltransferases and deacetylases (Kornberg and Lorch, 1999).

While initial studies were conducted on yeast models, recent work has shown that parasitic protozoan histones, including *P. falciparum* histones, function in a similar manner (Croken, Nardelli and Kim, 2012). Despite lacking the H1 linker histone, the core histones are present in a homologous form to eukaryotic counterparts and interact to form a nucleosome core which can wrap approximately 150bp of DNA (Duffy *et al.*, 2012). Subunit H3 and H4 N-terminal tail regions are also well conserved and enriched in lysine residues (Miao *et al.*, 2006).

Originally it was shown that only two histone modifications were found to be involved with transcriptionally active genes in *P. falciparum*, methylation of histone 3 lysine 4 (H3K4Me) and acetylation of histone 3 lysine 9 (H3K9Ac) (Salcedo-Amaya *et al.*, 2009). Further studies have since shown that additional residues H4K5Ac, H4K8Ac and H4K4Ac are also involved in gene expression (Luah *et al.*, 2010). Acetylation of these residues leads to transcriptionally active euchromatin structures, and conversely inactive heterochromatin when non-acetylated. Acetyl group removal is carried out by interaction with histone deacetylases (HDACs).

1.4 Role of Histone Deacetylases in Epigenetics

Since the importance of histone modifications was recognised, particularly acetylation, a great deal of research has been invested in understanding the regulatory mechanisms behind these processes. Recent focus has been trained on HDACs to gain insight into their functional role within the epigenetic process. HDACs are grouped into two major families, classical HDACs and the more recently discovered SIR2 family of nicotinamide adenine dinucleotide (NAD^+) dependent HDACs (Ruijter *et al.*, 2003). Previously HDACs had been separated into two groups (class I and class II) based on molecular mass and pairwise sequence similarity (Gregoretti, Lee and Goodson, 2004). NAD^+ dependent HDACs form their own group (class III) and a further group (class IV) was implemented to accommodate the recently discovered human HDAC11 which is distinct from existing class I and class II members (Gregoretti, Lee and Goodson, 2004).

Class III HDACs are more commonly known as ‘sirtuins’ after the discovery of their founding member, ScHst2, in the yeast organism *Saccaromyces cerevisiae*. Here ScHst2 serves to function as a transcriptional repression element at *S. cerevisiae* telomeres, silent mating-type loci and ribosomal DNA loci (Rine and Herskowitz, 1987). Its discovery

highlighted an important role in the epigenetic transcriptional control of the budding yeast genome by acting as a HDAC (Rine and Herskowitz, 1987). This established regions of 'silent' heterochromatin which were repressive for gene transcription at subtelomeres, mating type-loci and ribosomal DNA (Gasser and Cockell, 2001).

Since their discovery, homologues of yeast ScHst2 have been found in a wide range of organisms ranging from archaea to humans, including *P. falciparum* (Frye, 2000). As well as HDAC activity and gene silencing, sirtuins participate in a wide range of cellular processes including cell cycle progression, chromosome segregation, microtubule organisation, protein aggregates transport, genome stability, DNA repair, apoptosis and autophagy (Wenzel, 2006; Rajendran *et al.*, 2011). This has led to sirtuins being the focus of many ageing, life-span and metabolism studies linked to neurodegenerative diseases, cardiovascular disease and cancer (Pruitt *et al.*, 2006; Haigis and Sinclair, 2010).

Initial phylogenetic analysis of sirtuins classified 60 known sirtuins into four main types based upon similar domains to 7 human sirtuins (hSIRT1-7) (Frye, 2000). This sequence similarity analysis initially placed sirtuins into classes I-IV. Each of these classes include at least one of the originally identified 7 human sirtuins. A fifth undifferentiated class, U, was also identified for sirtuin genes with sequence motifs which seemed to be intermediate between classes II and III and the classes I and IV and includes sirtuins found in archaea and bacteria (Greiss and Gartner, 2009).

Certain characteristic amino acid sequences are used to define the classes of sirtuins. In the central core domain of the sirtuins there exists the conserved motifs GAGISXXXGIPXXR, PXXXH, TQNID, HG, two sets of CXXC, FGE, GTS and (I/V)N (Frye, 2000). For class III sirtuins the GAGISXXXGIPXXR motif is typically found as GAGISAESGIPTFR, whereas in class II sirtuins it is found as GAGISTESGIPDYR (Frye, 2000). Sirtuins found

in class U, I and IV also have a GIPD motif within this sequence like those in class II, it is only class II which possesses the GIPT motif. Class IV sequences also have a GVWTL motif present 4 residues C-terminal to the GAGISTXXGIPDFR motif (Frye, 2000).

Since this initial classification a larger number of genome sequences from all kingdoms of life has become available and this has informed more recent reviews (Greiss and Gartner, 2009; Religa and Waters, 2012). These more extensive analyses have further differentiated Class III into further subdivisions (a, b and c) and split the undifferentiated class into four branches (U1-4) as shown in Figure 1.5 (Greiss and Gartner, 2009; Religa and Waters, 2012).

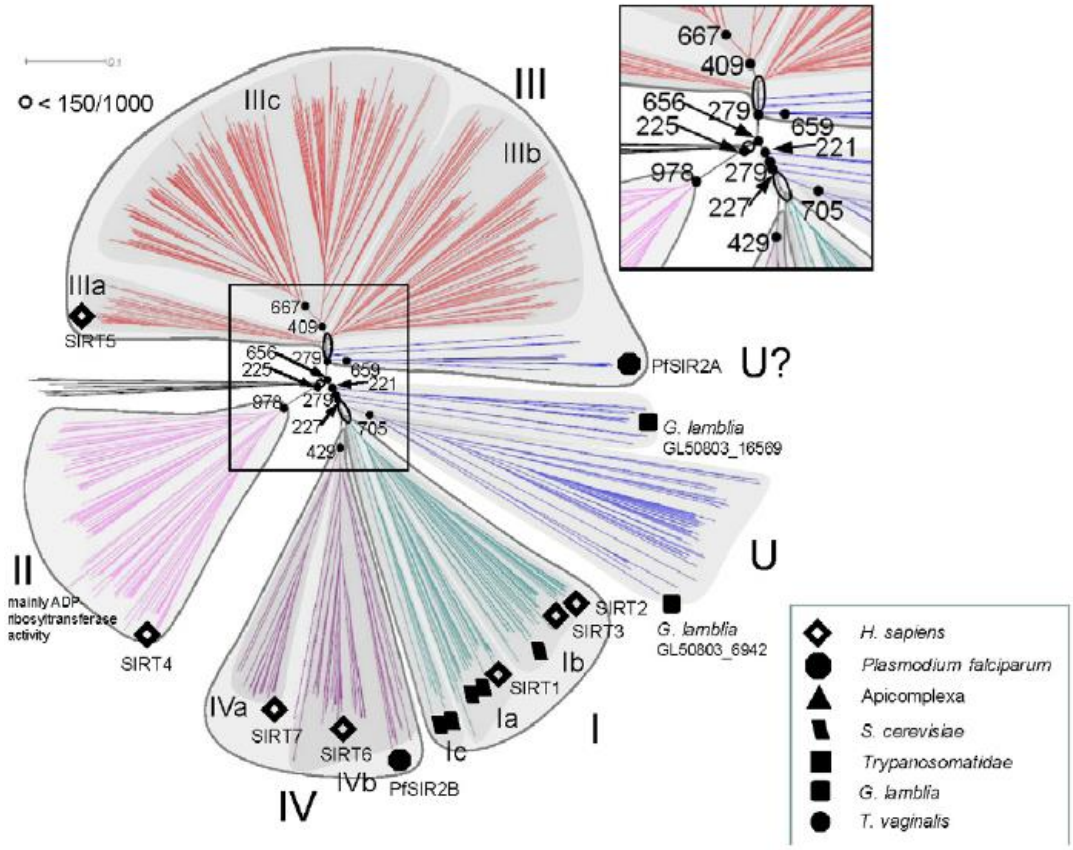


Figure 1.5: Phylogenetic tree of 778 annotated sirtuins from different organisms highlighting the different classifications (Religa and Waters, 2012).

An interesting discovery when conducting these phylogenetic searches was that a number of parasitic organisms were found to contain sirtuin sequences within their genome. Notable examples include *Trypanosoma brucei, Leishmania spp.* and also *P. falciparum*. Investigation of *P. falciparum* sirtuins has revealed their involvement with the epigenetic transcriptional control of a large number of subtelomeric genes which are heavily involved in malaria pathology and parasite survival, including the *var* gene family addressed previously in section 1.2 (Duraisingh *et al.*, 2005; Tonkin *et al.*, 2009). Given sirtuins' involvement in gene silencing as well as the regulation of many cellular functions, their role in pathogenic organisms is intriguing and presents possible treatment opportunities in the form of sirtuin-targeting drugs (Religa and Waters, 2012).

1.5 Structure of Sirtuins

All known sirtuins possess several conserved sequences necessary for structure and function. In eukaryotic organisms these are found within a ~260 amino acid region of homology flanked by N- and C-Terminal extensions (Finnin, Donigian and Pavletich, 2001). Crystallisation of a range of sirtuins has revealed this catalytic core region to contain two domains: a large NAD^+ binding Rossmann fold domain and a small domain comprised of two insertions into the Rossmann fold. This small domain is comprised of a variable $Zinc^{2+}$ binding domain and a helical module (Dutnall and Pillus, 2001; Min *et al.*, 2001; Greiss and Gartner, 2009).

The helical domain connects to the N-terminal of the Rossmann fold via a long loop structure (L-1B) and shows structural flexibility (Min *et al.*, 2001). The helical and Zinc domains become packed together to form a single domain despite arising from different regions of the core domain sequence (Dutnall and Pillus, 2001). The Rossmann fold

domain and the small domain are separated by a cleft in which both NAD^+ and an acetylated peptide substrate bind (see Figure 1.6) (Hoff *et al.*, 2006).

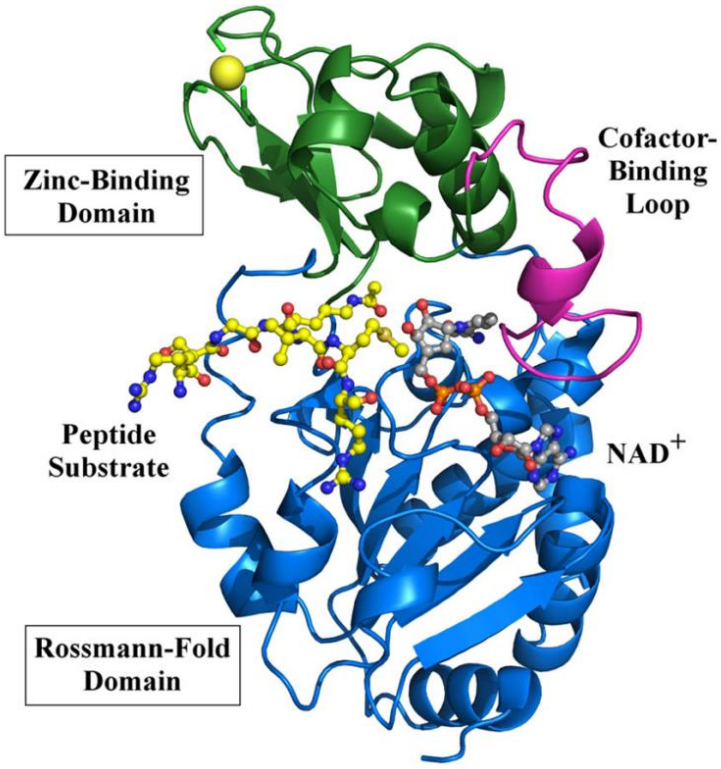


Figure 1.6: Overall structure of human Sirt3 with bound peptide substrate and NAD+ co-factor (Moniot, Weyand and Steegborn, 2012).

NAD^+ binding occurs within the larger domain that is a variant of the Rossmann fold, present in many diverse NAD(H)/NADP(H) binding enzymes (Bellamacina, 1996). This sequence is important for NAD phosphate binding and is typically conserved as a GAGISXXXGIPXXR motif. In class III sirtuins it is usually found as GAGISAESGIPTFR whereas in class II it is GAGISTESGIPDYR (Frye, 2000). These motifs are highly conserved between species, likely because of the structural requirements needed to bind NAD (Dutnall and Pillus, 2001).

The crystal structure of Sir2 from *Archaeoglobus fulgidus* reveals that the adenine base of NAD^+ binds in an inverted orientation compared to other NAD-linked binding enzymes

(Min *et al.*, 2001). The NAD binding pocket can be divided into three regions: site A where the adenine ring is bound; site B where the nicotinamide ribose is bound; and site C which is deep within the pocket (see Figure 1.7) (Min *et al.*, 2001). Site A is mostly hydrophobic and contains the glycine residues which are conserved across Sir2 family members (Min *et al.*, 2001). Sites B and C contain conserved polar residues and have been subject to mutagenesis which have shown reduced enzymatic activity. This indicates that they form the active site of the enzyme (Dutnall and Pillus, 2001).

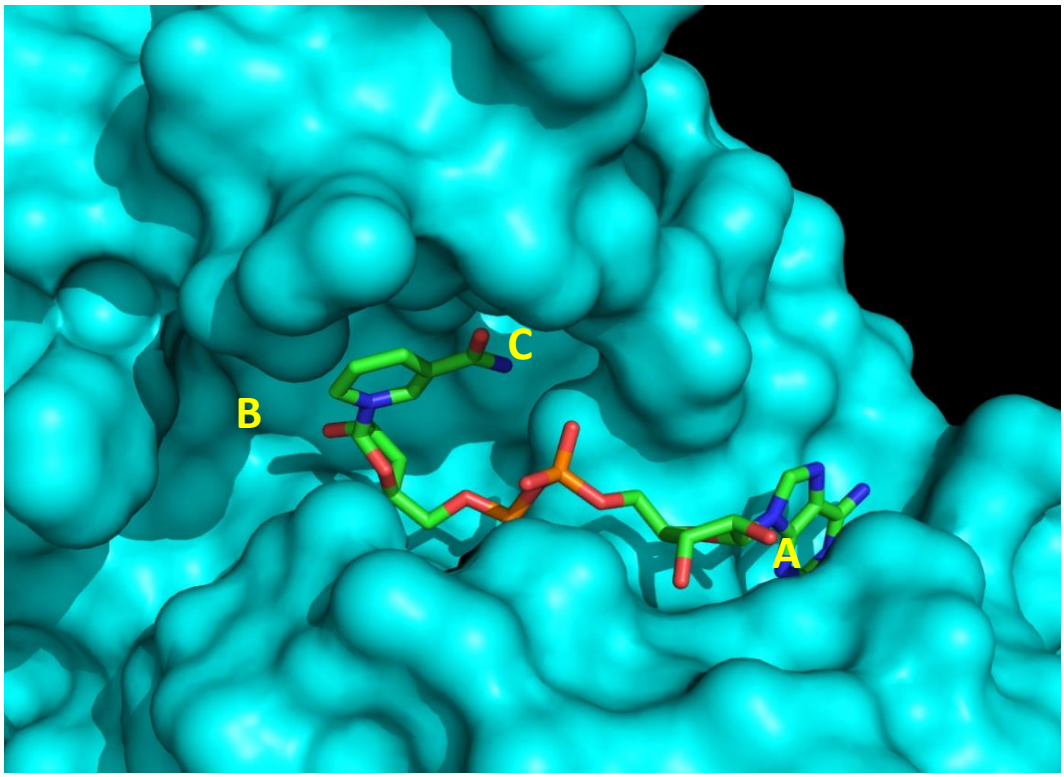


Figure 1.7: NAD binding pocket of Sir2 with bound NAD+ from Archaeoglobus fulgidus by Min et al, 2001 (PDB code 1ICI). (A) The adenine ring binding site, (B) nicotinamide ribose binding site, (C) deep pocket site containing conserved polar residues.

The smaller domain present in sirtuins contains the zinc binding module where a zinc atom is tetrahedrally co-ordinated by four cysteine residues. These are typically conserved in

two sets of CXXC sequences in most sirtuins (Frye 1999). Sir2s with CXXC … CXC, CXXXXC … CXXC and CXXC … CXXXXC are also found in sequence alignments (Chakrabarty and Balaram, 2010). There can be large insertions of amino acids between these two sequences. This suggests that considerable diversity is tolerated within the zinc binding module outside of these residues (Finnin, Donigian and Pavletich, 2001). While the zinc atom is not used in the catalytic activity of sirtuins it does have a fundamental structural role. This has been demonstrated in the case of *P. falciparum* where removal of the metal ion results in loss of enzyme structure and function (Chakrabarty and Balaram, 2010).

1.6 Sirtuin Silencing Mechanism

To inhibit transcription of a gene, a region of 'silenced' heterochromatin must be established. This is accomplished by the deacetylation of N-terminal lysine residues present on histone tails. Class III HDACs, which include sirtuins, differ from other deacetylases in classes I and II, which use a divalent Zn^{2+} ion for catalysis. When present in the active site of the class I and II deacetylases, the zinc ion increases the nucleophilicity of a bound H_2O molecule which hydrolyses the acetyl group present on lysine residues and releases it as a free acetate ion (Finnin, Donigian and Pavletich, 2001). Instead sirtuins use a unique and relatively complex reaction, which requires the presence of the coenzyme NAD^+ and an acetylated substrate (Tanny *et al.*, 1999; Imai *et al.*, 2000).

Via this mechanism ScHst2 cleaves the high energy glycosidic bond that exists between the adenosine diphosphate (ADP)-ribose moiety of NAD to nicotinamide (Landry, Slama and Sternglanz, 2000; Tanner *et al.*, 2000; Tanny and Moazed, 2001). Nicotinamide is then released from NAD^+ and the acetyl group of the substrate is transferred over to the cleaved product of the coenzyme (Tanner *et al.*, 2000; Tanny and Moazed, 2001). This results in

the generation of the metabolite O-acetyl-ADP ribose (OAADPr) (see Figure 1.8. for reaction mechanism) (Tanny and Moazed, 2001).

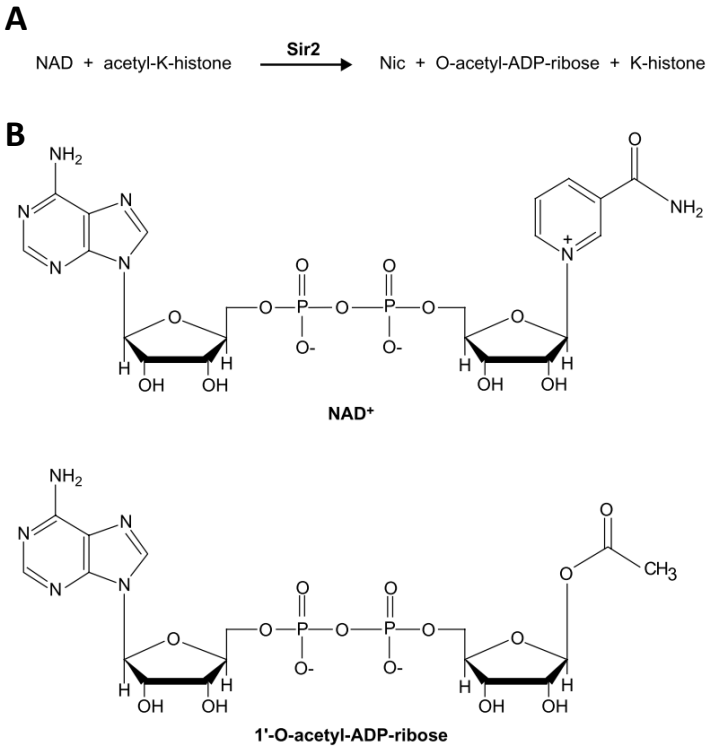


Figure 1.8: (A) Overall reaction scheme for NAD cleavage and acetyl transfer from substrate to ADP-ribose by ScHst2 and ScHst2-like proteins. (B) Structures of the NAD and O-acetyl-ADP ribose compounds (Moazed, 2001).

The removal of the acetyl group from the lysine residues present on histone tails by sirtuins via this mechanism results in an overall positive charge which increases the electrostatic attraction between the histones and DNA backbone of the chromatin structure due to the negative charge of the DNA backbone. This creates a 'silent' chromatin domain which is inaccessible by cellular recombination machinery and enzymes.

The generation of OAADPr allows for sirtuins to also act as ADP-ribosyltransferases (Sauve *et al.*, 2001). In fact, sirtuins were initially reported for their ribosylation activity upon both themselves and their histone substrates (Tanny *et al.*, 1999; Frye, 2000). Both

ScHst2 and HsSirT2 were shown to possess ADP-ribosylation activity rather than deacetylase activity (Tanny *et al.*, 1999). However, the observed effect was very weak and following studies focused on the ability of sirtuins to act as a deacetylases on histones H3 and H4 (Imai *et al.*, 2000). The deacetylase activity was shown to be much more active compared to that of the ribosylation (Imai *et al.*, 2000). This suggested that deacetylase activity is in fact the sirtuins primary activity.

Both ADP-ribosylation and deacetylation by sirtuins are biologically important processes involved in signalling pathways regulating gene expression (Ziegler, 2000; Belenky, Bogan and Brenner, 2007). It has since been determined that some sirtuins, such as ScHst2, are biased towards deacetylation whereas other sirtuins such as HsSIRT6 and HsSIRT4 are biased towards ADP-ribosylation (Liszt *et al.*, 2005; Haigis *et al.*, 2006). The necessity for both reactions to have an acetylated substrate present for activity suggests there is a link between the two activities (Tanny *et al.*, 1999; Imai *et al.*, 2000; Landry, Slama and Sternglanz, 2000; Tanner *et al.*, 2000; Tanny and Moazed, 2001).

Studies in *S. cerevisiae* models have shown that in order to function, different sirtuins interact with partner proteins to bind to the chromatin structure. ScHst2 forms a complex with Sir3p and Sir4p to mediate silencing at mating-type loci and telomeric DNA regions (Moazed *et al.*, 1997). This does not appear to be the case in higher organisms however; human sirtuins are currently not known to form complexes in order to bind to chromatin.

1.7 Sirtuins in *P. falciparum*

Several histone deacetylases and histone acetylases have been found in the *P. falciparum* genome to date suggesting a functional acetylation/deacetylation mechanism as found in other organisms (Religa and Waters, 2012). Two Sir2-like genes have been found based on sirtuin domain homology, *PfSir2a* and *PfSir2b* (Duraisingh *et al.*, 2005; Tonkin *et al.*,

2009). These sirtuins have been characterised and linked with transcriptional silencing of the *var* gene family encoded in subtelomeric regions (Freitas-Junior *et al.*, 2005; Tonkin *et al.*, 2009). Activation of a *var* gene occurs with the establishment of a euchromatin region permissive to transcription.

PfSir2a is perhaps the best characterised of the two *P. falciparum* sirtuins. It is also one of the smallest known eukaryotic sirtuins having only 273 residues and resembles the *A. fulgidus* sirtuin Af1 in size and sequence similarity (30.13%) (Merrick and Duraisingh, 2007). The core domains of PfSir2a are poorly conserved when compared to yeast, human and *Archaeoglobus* sequences (23 -35%) but similarity between PfSir2a and *S. cerevisiae* sirtuin exists in nearly all of the residues considered to be important for catalytic activity in sirtuins (Merrick and Duraisingh, 2007). The four cysteine residues necessary for the zinc binding module are conserved in the CXXC…CXC form as previously mentioned. The NAD^+ binding pocket is broadly conserved with that of the Af2 sirtuin where only three substitutions occur: GAGXS/G<u>S</u>GXS, GIPXFR/<u>N</u>IPXFR and TQNIDL/TQN<u>V</u>DXL (see Figure 1.9) (Merrick and Duraisingh, 2007).

A crystal structure of PfSir2a in complex with histone H3 Lys9 (H3K9) peptide sequence has recently been achieved (Zhu *et al.*, 2012). This work suggested that PfSir2a has weak deacetylase activity compared to other sirtuins and it is more effective at catalysing the hydrolysis of medium and long chain fatty acid groups from lysine residues (Zhu *et al.*, 2012).

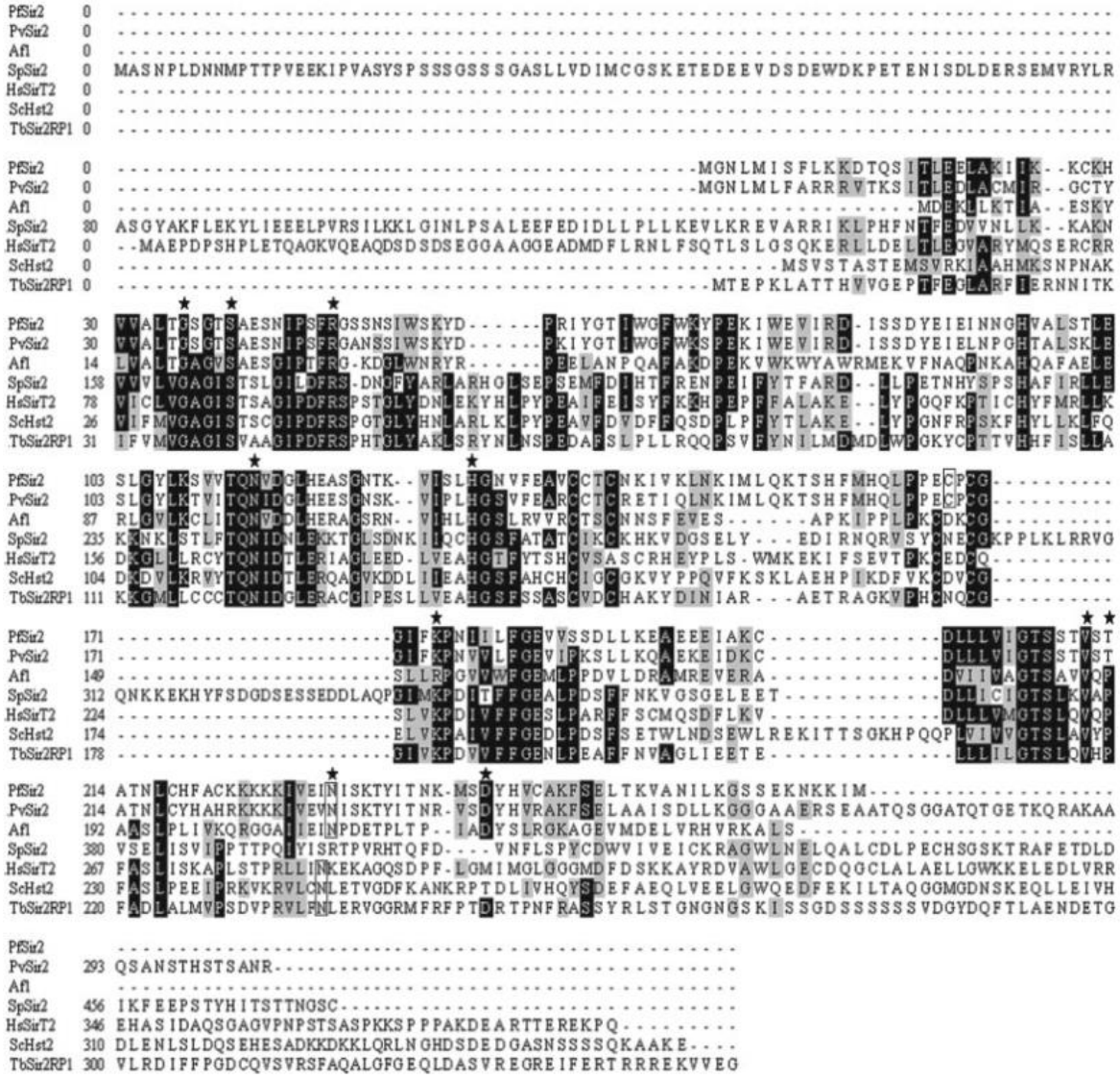


Figure 1.9: Alignment of PfSir2a sequence with that of other known sirtuins (Merrick and Duraisingh, 2007). The predicted sequence of PfSir2a is aligned with that of *A. fulgidus* Afl, *S. cerevisiae* Hst2, *Schizosaccharomyces pombe* Sir2, *Homo sapiens* SirT2 and *Trypansoma brucei* SIR2RP1. Strictly conserved and conserved residues between species are shaded in black and grey respectively. Starred residues are those known to be essential for catalytic activity in ScHst2.

PfSir2a's role in the regulation of *var* genes has been shown via genetic studies in which PfSir2a-lacking mutants showed up-regulation of many *var* gene transcripts simultaneously (Duraisingh *et al.*, 2005). The most predominantly affected *var* genes were

those controlled by UpsA, UpsE and to some extent, UpsC promoters (Duraisingh *et al.*, 2005). A number of neighbouring *rifin* multigene family members were also upregulated in the absence of PfSir2a (Duraisingh *et al.*, 2005). It has previously been mentioned in section 1.2 that severe malaria correlates with the expression of particular *var* genes from *upsA* regions (Jensen *et al.*, 2004; Kyriacou *et al.*, 2006; Rottmann *et al.*, 2006). This highlights the possibility of an important role for PfSir2a in disease pathology.

1.8 Partner Proteins

As mentioned earlier, yeast models have shown that sirtuins can form complexes with partner proteins in order to bind to chromatin regions and perform their function. Whilst homologues to these partner proteins are not seen in higher organisms, several DNA/ribonucleic acid (RNA) binding protein families and transcription factors can be found encoded in the *Plasmodium* genome, which are involved in regulation of gene expression (Cui, Fan and Li, 2002; De Silva *et al.*, 2008; Mancio-Silva *et al.*, 2008). One example which has been recently studied is the archaeal Alba (acetylation lowers binding affinity) DNA-binding protein (Bell *et al.*, 2002).

Alba, formerly known as Sso10b, was initially discovered when several small (~10kDa) DNA binding proteins were isolated from the extreme thermophilic archaea *Sulfolobus solfataricus* (Forterre, Confalonieri and Knapp, 1999). Since its discovery several Alba homologues have been found to be highly conserved in the genomes of both euryarchaea (which encode histone proteins) and crenarchaea (which do not) (Forterre, Confalonieri and Knapp, 1999). The genomes of a range of eukaryotes are found to contain homologues of Alba. These include many higher plants and protists, such as *Leishmania* and Trypanosome species, and *P. falciparum*. Weaker homologues also exist within vertebrates including *Homo sapiens* (Bell *et al.*, 2002).

Several species, including *S. solfataricus*, *S. tokodai, S. acidocaldarius* and *A. fulgidus* have two copies of the Alba gene present within the genome (Wardleworth *et al.*, 2002; Jelinska *et al.*, 2005). The more highly conserved gene represents the sequence for Alba1 which share over 90% homology with one another and ~ 60% for more distant archaeal species (Jelinska *et al.*, 2005). The less conserved gene encodes Alba2 which only shares 30 – 40% homology to Alba1 sequences but are 60 – 70% identical to one another (Jelinska *et al.*, 2005). This divergence suggests that Alba2 may have a different function to that of Alba1 in these organisms. The *P. falciparum* genome also encodes multiple Alba homologues which may have evolved to facilitate different functions (Goyal *et al.*, 2012).

Alba is an abundant protein within archaea; it represents 4% of the total protein in *S. shibatae* (Xue *et al.*, 2000). The presence of Alba of such quantities in thermophilic and hyperthermophilic organisms growing at high temperatures (up to 80°C) suggests a role in DNA protection and maintaining stability, as high temperatures lead to the denaturing of DNA (Jelinska *et al.*, 2005). This is enhanced by the fact that, like eukaryotic histones, Alba binds DNA non-specifically and chromatin immunoprecipitation studies show that it appears to be equally distributed on the chromosome (Xue *et al.*, 2000; Bell *et al.*, 2002).

1.9 Structure of Alba

Structural and *in vitro* studies of Alba have revealed that it exists as a dimeric protein with 10 kDa subunits (see Figure 1.10) both in crystallised structure and solution (Wardleworth *et al.*, 2002; Zhao *et al.*, 2003). Alba possesses a mixed α/β fold structure which is composed of two α-helices and four β-strands. This fold is similar to the N-terminal DNA binding domain of DNase I and the C-terminal domain of the bacterial translation initiation factor 3 (IF3) (Wardleworth *et al.*, 2002; Zhao *et al.*, 2003). This is suggestive that the origin of Alba stems from a group of ancient nucleic acid-binding domains.

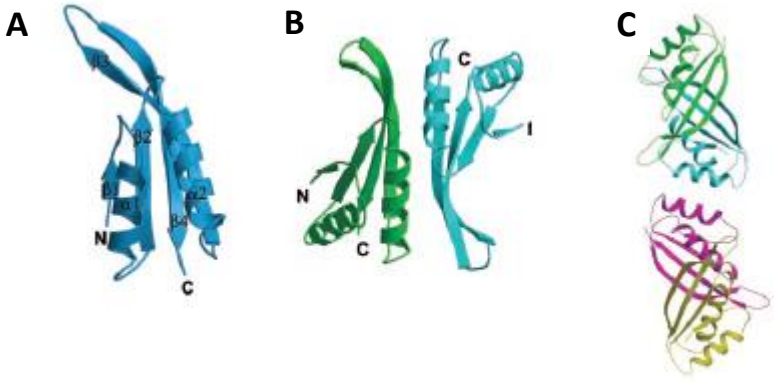


Figure 1.10: Structure of Alba from *A. fulgidus* as (A) monomer (B) dimer (C) in crystal unit cell (Zhao, Chai and Marmorstein, 2003).

Research has also been carried out by analysing known Alba homologues using comparative genomics and sequence profile analysis (Aravind, Iyer and Anantharaman, 2003). Sequences of several RNA binding proteins across archaea and eukaryotes including YhbY, RNase P subunit and IF3 were compared to that of Alba to identify characteristic traits. This investigation revealed that archaeal Alba proteins appear to have originated as RNA-binding proteins which formed various ribonucleoprotein complexes (Aravind, Iyer and Anantharaman, 2003). Ssh10b, an Alba homologue in *S. shibatae*, has been shown to bind RNA and DNA with a similar affinity in vitro and may also have a functional role in chaperoning RNA in vivo (Guo, Xue and Huang, 2003).

In addition to the α/β fold a long β-hairpin comprised of two interacting β-strands extends out from the body of the protein (Wardleworth *et al.*, 2002). Extended β-hairpin arms are a known feature of existing prokaryote DNA binding proteins. Bacterial HU and Integration Host Factor (IHF) specifically bind to DNA and function as architectural factors using two flexible β-hairpin arms which interact with the minor groove of DNA (Rice *et al.*, 1996). In the dimer form of Alba the β-hairpin arms extend out in opposite directions from the

protein and span a distance of ~40 Å where they interact with the minor groove (Wardleworth *et al.*, 2002).

The central body of the Alba dimer consists of two highly conserved loops which are separated by ~20 Å (Wardleworth *et al.*, 2002). These loops contain two consecutive lysine residues (Lys16 and Lys17) which are involved in binding to DNA (Bell *et al.*, 2002). The two β-hairpin arms bring the central body into contact with the major groove of the DNA. Consecutive dimers of Alba can overlap along the DNA duplex with a virtually complete coating of the nucleic acids. This results in a densely coated nucleoprotein filament (Wardleworth *et al.*, 2002).

1.10 Role of Alba in Silencing

As the name suggests, acetylation of Alba results in a lower binding affinity for DNA. This occurs when the lysine residues contained within the loops of the central body of the dimeric protein gain an overall negative charge due to the addition of an acetyl group (Bell *et al.*, 2002; Zhao *et al.*, 2003). Deacetylation releases the acetyl group from the basic lysine residue and results in a polar positive charge which has a higher affinity for DNA. Both acetylated and deacetylated forms of Alba have been shown to exist in *S. solfataricus* (Bell *et al.*, 2002).

It has been shown that Alba can be deacetylated at Lys16 and Lys17 by a Sir2 homologue of *S. solfataricus*, ssSir2, and thereby increase its binding affinity for DNA (Bell *et al.*, 2002). Later work has since shown that a *Sulfolobus* homologue of the bacterial protein acetyltransferase, Pat, from *Salmonella enterica* is able to acetylate Alba on Lys16 (Marsh, Peak-Chew and Bell, 2005). In the bacterial system a sirtuin homologue CobB deacetylates acetyl-CoA synthase (ACS) where Pat acetylates it forming a metabolic regulatory system (Starai *et al.*, 2003). The ability of a Pat homologue to acetylate Alba in *S. solfataricus*

suggests that the metabolic regulatory pathway has been adopted to regulate chromatin structure (Marsh, Peak-Chew and Bell, 2005).

1.11 Alba in *P. falciparum*

As mentioned previously Alba homologues have been found and characterised within the genome of *P. falciparum,* PfAlba1-4 (Chêne *et al.*, 2012; Goyal *et al.*, 2012)). Of the four Alba homologues, PfAlba3 shares a similar number of residues and molecular weight with the archaeal Albas (see Figure 1.11). Several residues are highly conserved across species, indicating a similar function, and molecular modelling from previously defined Alba structures has yielded a similar overall structure with the characteristic β-hairpin loops required to dock with the minor groove of DNA (Goyal *et al.*, 2012). The remaining Albas have significantly larger C-terminal extensions compared to PfAlba3 and contain many RGG box repeats. These are characteristic of many RNA binding proteins and suggestive that they may preferentially bind to RNA (Chêne *et al.*, 2012).

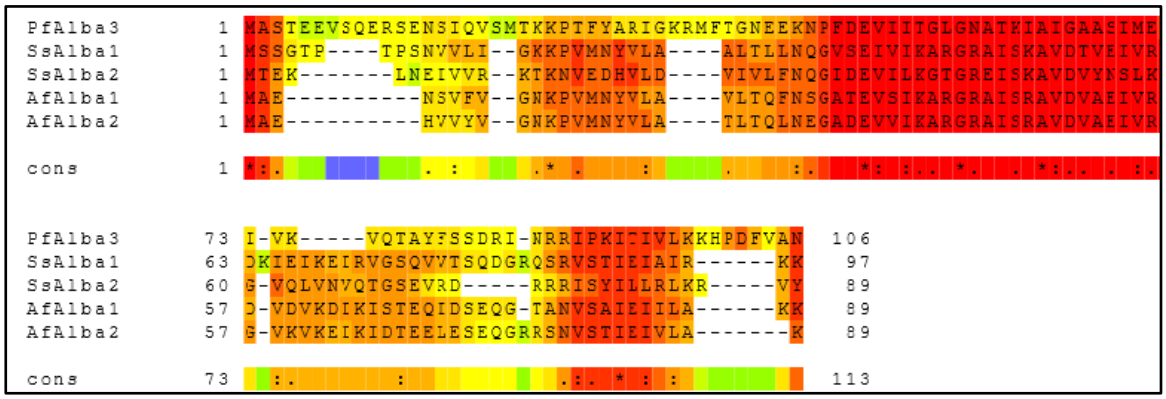


Figure 1.11: T-Coffee amino acid sequence alignments between Alba's from *P. falciparum*, *S. solfataricus* and *A. fulgidus*.

PfAlba3 possesses DNA binding properties similar to that of the archaeal Alba's previously described. It is able to bind to DNA non-specifically and is localised within the nucleus of the parasite where it co-localised with PfSir2a (Goyal *et al.*, 2012). Mutation of the lysine residues in the central domain of PfAlba3 was shown to interfere with DNA

binding properties, but more importantly PfAlba3 was shown to interact with PfSir2a (Goyal *et al.*, 2012). Co-localisation studies suggest that both PfSir2a and PfAlba3 play a functional role in the epigenetic regulation of transcription within *P. falciparum.*

Characterisation of PfAlba1, 2 and 4 show that they do in fact bind RNA and DNA and a cellular redistribution of PfAlbas occurs throughout the development of the parasite (Chêne *et al.*, 2012). PfAlba4 shares some co-localisation with PfSir2a during this development. Given that that PfSir2a is known to interact with PfAlba3, it is possible it may also interact with PfAlba4 to increase its binding affinity. This hints at the possibility of an interesting regulatory mechanism for the transcriptional control of *Plasmodium* virulence genes.

This thesis aims to investigate and characterise the interaction between PfSir2a and PfAlba3 evidenced from the literature. This will be achieved by producing recombinant PfSir2a and PfAlba3 proteins using bacterial expression systems. The recombinant proteins will then be assessed for suitability for structural characterisation using biochemical, biophysical methods and is described in Chapter 3. Chapters 4 and 5 relate to small angle X-ray (SAXS) and neutron (SANS) scattering methods used to determine the solution structures of these proteins and describe their behaviour and interactions.

Chapter 2: Experimental Methods

2.1 Molecular Biology

Several molecular biology techniques were employed for the initial stages of sample preparation and characterisation. This involves the manipulation of DNA and proteins to produce recombinant target proteins for further investigation.

2.1.1 Plasmid DNA Preparation

Plasmids are small, circular, double stranded DNA molecules which can be extracted from prokaryotes. These independently replicating molecules can be used in molecular biology to amplify a gene of interest which can then be translated into a protein or peptide for further study. Specifically designed plasmids (or vectors) also contain antibiotic resistance markers to select for a single population of bacteria in which only cells containing the plasmid will survive when grown in media containing an antibiotic agent (e.g. kanamycin sulphate, ampicillin).

2.1.2 pET28a Plasmid Preparation

Initial cloning work was carried out using the Novagen pET28a (Cat no. 69864-3) kanamycin sulphate resistant expression system. To prepare a maxiprep of this plasmid, 5 µl of pET28a stock plasmids were mixed with 50µl of DH10b electrocompetent cells on ice for 15 minutes. Bacteria were pipetted into 0.1cm spaced Gene Pulser® electroporation cuvettes (Bio-Rad Cat no. 1652089) and electroporated at 1.6kV, 25µF, 200 ohms. Bacteria were recovered in 300 µl super optimal broth with catabolite repression (SOC) media for 45 minutes in a shaking incubator set at 37°C, 200 revolutions per minute (rpm) before plating onto a Luria broth (LB) agar plate containing 35 µg ml^{-1} kanamycin sulphate (Euromedex Cat no. EU0420). The plate was then placed in a 37°C incubator overnight.

A single colony from the overnight incubation was grown up in 5 ml of LB starter culture containing 35 μg ml^{-1} kanamycin sulphate and grown overnight at 37°C, shaking at 200 rpm. Starter cultures were then inoculated into 500ml LB containing 35μg ml^{-1} kanamycin sulphate and placed in a shaking incubator set at 37°C, 200 rpm until a measured optical density 600nm ($O.D_{600}$) of 0.6 (when bacteria is in exponential growth phase) was reached. Bacteria were pelleted by centrifugation for 15 minutes at 5000 rpm, 4˚C. Harvested pellets were prepared by Maxiprep (Qiagen Cat no. 12165) to recover plasmid DNA in 1ml distilled water.

2.1.3 Polymerase Chain Reaction

Target copy DNA (cDNA) sequences were produced using the polymerase chain reaction (PCR) method. Here, the genes of interest are replicated from template DNA using forward and reverse primer sequences specific to the target DNA. The template DNA is placed together in a buffered reaction mixture containing Taq polymerase, deoxynucleoside triphosphates (dNTPs) and two primer sequences which are complementary to the forward and reverse DNA strands. The reaction then undergoes several stages of heating inside a thermal cycler.

An initial denaturation step occurs by heating the DNA to 95°C for 5 minutes to separate the double-stranded DNA (dsDNA) template into single-stranded DNA (ssDNA). An annealing step follows by lowering the temperature of the reaction to between 50°C and 65°C for 20 to 40s to allow for the hybridisation of the primers to the template DNA strands. The temperature must be low enough for the hybridisation process to occur but high enough so that the binding is specific to the target sequences. Adjusting the temperature of the annealing step can be performed to optimise the efficiency of the reaction if required.

After the annealing step an extension/elongation stage is carried out, typically at 72°C which is the optimum temperature for the Taq polymerase enzyme used in the reaction. During this step the polymerase synthesises a new DNA strand by adding dNTPs complementary to the target DNA sequence in the 5' to 3' direction. The extension time for this step depends on the length of the product being amplified and the DNA polymerase being used. Typically the rate of synthesis for a DNA polymerase is 1000bp per minute, and under optimum conditions (no rate limiting step e.g. lack of dNTPs) the amount of product generated is doubled which leads to an exponential increase in amplification of the target DNA.

The cycle of denaturation, annealing and extension is carried out typically between 20-40 times depending on the desired amount of product to be obtained at the end of the reaction. A final elongation step at 70°C - 74°C for 5 minutes is performed to ensure that any remaining single stranded pieces of DNA are fully elongated before the reaction terminates and the product can be recovered.

The PfSir2a sequence, codon optimised for overexpression in *E. coli,* was obtained from Zhu *et al* (Zhu et al., 2012) on filter paper in a pET21a vector. Intact plasmids were unable to be recovered by soaking the filter paper. Instead the PfSir2a sequence was recovered by PCR. GoTaq Green (Promega Cat no. M7121) with T7 forward and reverse primers (MWG) were used in reactions (see Appendix B for primer sequences). The reaction cycle was set up with an initial denaturation step at 95˚C for 5 minutes followed by 20s annealing at 50˚C, 45s extension at 72˚C, 30s at 95˚C for 35 cycles. A final extension step was performed at 72˚C for 5 minutes before placing on hold at 4˚C for product recovery.

2.1.4 Agarose Gel Electrophoresis

Agarose gel electrophoresis is a method employed to separate mixed populations of DNA and also proteins of different sizes in an agarose gel matrix for analysis and/or recovery of samples for further downstream applications. The typical size separation range for DNA molecules in agarose gels is between 50 - 20,000bp; higher percentage gels provide better resolution of smaller DNA fragments and vice versa.

Gels are prepared by adding agarose powder (Sigma Aldrich, Cat no. A9539) to tris/acetate/ethylenediaminetetraacetic acid (EDTA) (TAE) or tris/borate/EDTA (TBE) buffer solutions with typical concentrations between 0.7% - 2% w/v and heating to above 95°C to allow the agarose to dissolve in the buffer and create a uniform conductive gel matrix as it slowly cools. Gels are created in a casting mould to allow for the addition of a gel comb to create the wells in which DNA samples can be placed for separation and later visualisation.

Once set, agarose gels are submerged in a tank apparatus containing the same buffer solution as was used to make the gel. Samples mixed with a dye can be loaded into the pre-formed wells alongside a DNA marker set for estimation of molecular weights. The system is sealed and a voltage is applied across the gel at a rate of between 5-8 V per cm length of the gel that was cast. DNA molecules have a net negative charge and so will migrate from the cathode end of the gel towards the anode.

When the dye front reaches the anode end of the gel, the voltage is turned off and the gel stained with an appropriate agent to visualise the separated DNA. Ethidium bromide or SYBR safe stain (Invitrogen, Cat no. S33102) were used: these intercalate into the major groove of DNA and fluoresce under ultra violet (UV) light. In the case of PCR products, the separated DNA was recovered by extracting the band and the surrounding agarose for

purification via gel extraction kit (Qiagen Cat no. 28704) for use in digestion and ligation reactions.

2.1.5 Plasmid Digestion And Product Ligation

Target cDNA sequences are inserted into expression vectors by first digesting the PCR product and the plasmid intended to be inserted into bacteria for expression. Restriction endonucleases such as NcoI or BamHI are used which cut the target DNA at specific points along the nucleotide sequence known as restriction sites. DNA strands cut by these enzymes leave overhanging nucleotides which can be joined to another DNA strand with the complementary overhang sequence by using DNA ligases. Digestion of the expression vector plasmids with the same restriction enzymes creates the same overhanging sequences in the multiple cloning site ready for ligation.

DNA ligase enzymes join the complementary sequences together by forming phosphodiester bonds between the 3' hydroxyl group and the 5' phosphate group of the adjacent DNA residues. This allows the target gene sequence to be introduced into the expression vector to be made ready for transformation into an expression strain of bacteria.

In this work, the recovered pET28a plasmid concentration was quantified by Nanodrop A260 absorption (where at a concentration of 1 ug/ml and a 1 cm path length, dsDNA has A260 = 50) (Thermo scientific) and digested with restriction enzymes. Plasmids for *Alba3* sequence inserts were digested with NcoI and XhoI restriction enzymes (New England Biosciences Cat no. R1093S, R0146S). Plasmids for *PfSir2a* sequence inserts were digested using XhoI and BamHI (New England Biosciences Cat No. R0136S). Using the restriction sites listed for the respective sequence inserts yields an N-terminal cleavable 6 His-tag on PfSir2a whereas a non-cleavable C-terminal 6 His-tag would be present on PfAlba3. Restriction digests were performed at 37°C for 2 hours. After 1 hour 0.5µl of calf

intestinal alkaline phosphatase (Fisher Cat no. BP3217-1) was added to remove terminal phosphates and thus prevent self-ligation of plasmid vectors. Digested plasmids were run on 0.8% agarose gel at 80V in 1X TAE buffer for 1 hour and products were extracted from the gel via gel extraction kit (Qiagen Cat no. 28704).

Plasmids were digested a second time, as mentioned previously, with the digested product subjected to phenol/chloroform extraction with addition of 50 µl phenol/chloroform and centrifuged at 8000 rpm for 1 minute. The DNA-containing aqueous phase was mixed with 5µl 3M sodium acetate (pH 5.5) and 110 µl 100% ethanol. Precipitations were incubated for 2 hours at -80°C before centrifugation at 14,000rpm, 4°C for 30 minutes. The supernatant was removed and DNA pellets were washed with 100µl 70% ethanol by centrifugation for 5 minutes at 14,000rpm, 4°C. Residual ethanol was removed and tubes air dried for 15 minutes before re-suspending pellets in 10µl double distilled water (ddH_2O).

2.1.6 Site Directed Mutagenesis

In this method a high fidelity (HF) DNA polymerase (*PfuUltra*) is used for a mutagenic primer-directed replication of the both the original plasmid DNA strands. The supercoiled dsDNA vector containing the insert of interest is added to a reaction mix containing two synthetic primers, both containing the desired mutation, and the HF polymerase. The oligonucleotide primers, each complementary to opposite strands of the vector, are extended under thermocycling conditions. The extension of the mutated primers generates a mutated plasmid containing staggered nicks. After the thermocycling, the product is treated with *Dpn* I endonuclease which is specific for methylated and hemimethylated DNA. This digests the parental DNA template and selects for the mutation-containing plasmids as DNA isolated from almost all *E.coli* strains is dam methylated and therefore

susceptible to *Dpn* I digestion. The nicked vector DNA containing the desired mutations can then be transformed into competent cells (see Figure 2.1).

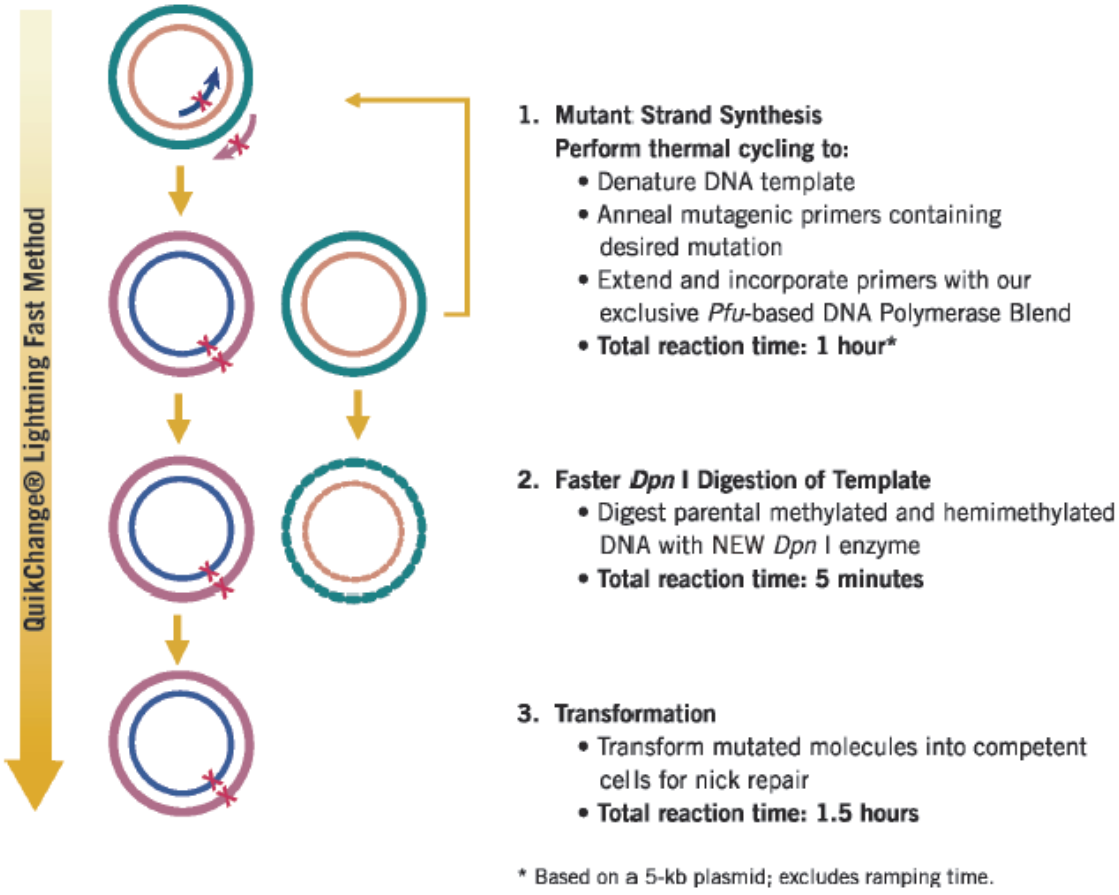


Figure 2.1: Overall scheme of site directed mutagenesis method. DNA is first denatured to allow for mutagenic primers to anneal and begin extension. Parental DNA template is then digested using *Dpn I* enzyme to leave only mutated sequence. Mutated molecules are then transformed into competent cells for nick repair and plasmid recovery. Source: Agilent Technologies QuickChange Lightning Manual (Cat no. 210518).

A 50µl reaction mix was created using 5µl of 10X reaction buffer, 1µl of 100ng μl^{-1} dsDNA (pET28a plasmid with original PfAlba3 insert), 1.25µl of both the forward and reverse primer for the mutagenesis (at 100ng μl^{-1}), 1µl of dNTP mix (as supplied in kit), 1.5µl of Quicksolution reagent, 38µl of autoclaved Milli-Q H_2O and 1µl of QuickChange

Lightning enzyme. The reaction was placed in a thermocycler and PCR was performed using the following protocol. Initial denaturation step at 95˚C for 2 minutes followed by 18 cycles of denaturation at 95˚C for 20s, annealing at 60˚C for 10s and an extension step at 68˚C for 2.5 minutes. A final extension step at 68˚C was performed for 5 minutes before being placed on hold at 4˚C for recovery.

After the PCR cycle was completed, 2µl of *Dpn* I restriction enzyme was added directly to the amplification reaction to digest parental DNA. This was mixed gently by pipetting before being briefly spinning down the contents and incubated at 37˚C for 5 minutes. A 45µl aliquot of XL10-Gold ultracompetent cells was thawed on ice in a 1.5ml Eppendorf tube per transformation reaction. 2µl of beta-mercaptoethanol (β-ME) was added to the cells and incubated on ice for 2 minutes. 2µl of the *Dpn* I treated amplification reaction was added to the XL10-Gold cells and mixed gently before incubating on ice for 30 minutes.

A heat shock transformation step was performed as described in section 2.1.7 below and checked for bacterial growth. Colonies were observed on the LB agar plates after overnight growth and three were chosen for miniprep production to confirm the newly mutated plasmid sequence was correct. Colonies were isolated into separate 10ml LB flask cultures containing 35µg ml^{-1} kanamycin sulphate and left to grow overnight. The bacteria were harvested by centrifugation at 5000rpm for 15 minutes at 4˚C, the excess media was removed and the bacterial pellets were frozen at -20˚C prior to recovery of the plasmids via miniprep. Plasmids were recovered using the Promega Wizard Plus SV miniprep DNA purification system (Cat no. A1460).

2.1.7 Bacterial Transformation

For expression and recovery, the pET28a vector containing the PfSir2a sequence and the mutated pET28a vector containing the PfAlba3 sequence were transformed into an *E. coli* expression strain, BL21 (DE3) (ThermoFisher Scientific, Cat no. C600003). This was achieved by mixing 1µl of 10ng μl^{-1} plasmid stock with 50µl of thawed BL21 (DE3) cells and incubating on ice for 30 minutes. The transformation was initiated by heat shock (immersion in a water bath) at 42˚C for 30 s before being placed back on ice and incubated for a further 2 minutes. After incubation, 0.3 ml of pre-warmed SOC media was added to the transformation reaction and the cells were left to recover at 37˚C, shaking at 200 rpm for 1 hour. 100µl and 200µl of the recovered bacteria was plated (to ensure sufficient colonies without overcrowding) on LB agar plates containing 35µg ml^{-1} kanamycin sulphate. Plates were incubated overnight at 37˚C and checked after 16 hours to see if bacterial colonies have developed which can be used for expression tests.

2.1.8 Sodium Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein expression and purity can be assessed using SDS-PAGE gels which separate proteins by size based on their mobility through a gel matrix. Whole bacterial lysates or purified protein samples are mixed in a sample loading buffer which containing Tris-HCl, glycerol, SDS, bromophenol blue dye and dithiothreitol (DTT). Samples were denatured and linearised by heating to 100°C for 10 minutes before loading onto 12% or 16.5% Tris-Tricine (TT) gels run at a constant voltage of 100V, at a pH value of 8.45. Visualisation of protein migration through the gels was monitored by staining with coomassie blue water based stain protocol. Recipes for gels and sample loading buffer can be found in Appendix A.

2.1.9 Western Blot Protocol

In this protocol samples were run on 16.5% TT gels and run at 100V for 90 minutes before transferring to a nitrocellulose membrane. The transfer was conducted using the semi-dry method with a Trans-Blot® Turbo™ blotting systems (Bio-Rad Cat no. 170-4155) western blot system and its associated Trans-Blot nitrocellulose membrane transfer packs (Bio-Rad Cat no. 170-4158). The SDS-PAGE gel was placed between the nitrocellulose membrane and transfer pads and set to transfer at 1.3A, 25V for 7 minutes.

After the transfer was complete the membrane was removed and placed into a 5% milk powder 1x phosphate buffered saline solution (PBS – Euromedex Cat no. ET330-A) with 0.05% Tween®-20 (Sigma Aldrich Cat no. P2287) (PBST) and incubated at room temperature (RT - 21°C) on a shaking platform set at 150 rpm for 1 hour. After 1 hour the membrane was washed three times by shaking with 1x PBST for 5 minutes each. The membrane was then placed into a 5% milk powder, 1x PBST solution containing a 1:3000 dilution of mouse anti-His primary antibody (GE Healthcare Cat no. 27-4710-01) and left to incubate for 1 hour shaking at RT

After incubation with the primary antibody the membrane was washed again three times with 1x PBST for 5 minutes each. The membrane was then placed in a 1xPBST solution containing a 1:7500 dilution of donkey anti-mouse, alkaline phosphatase conjugated secondary antibody (Promega Cat no. S3721) and incubated for 1 hour shaking at RT. The membrane was removed for a final wash step with 1x PBST, three times for 5 minutes each. The results of the blot were developed by adding 3ml of Western Blue® Stabilized Substrate for Alkaline Phosphatase solution (Promega Cat no. S3841) to the membrane and incubating for 10 minutes at RT. The reaction was inhibited by submersing the membrane in ddH_2O and left to dry before imaging.

2.1.10 Nickel-Nitrilotriacetic acid (Ni-NTA) Affinity Purification

Affinity chromatography is a powerful technique used to separate biomolecules using highly specific covalent interactions between amino acids and metal ions. Here immobilised metal affinity chromatography (IMAC), specifically Ni-NTA affinity chromatography, was used to separate proteins of interest from bacterial lysate after expression. Histidine tags were employed which bind strongly with nickel or cobalt ions covalently attached to agarose beads via a chelating ligand.

The histidine tags present on the target, introduced to the target protein by genetic modification, bind tightly to the resin whilst native host proteins have little or weak binding. The resin is washed with increasing concentrations of imidazole (2-40mM), which serves as a competitor to weakly bound contaminants to improve the purity of the target proteins. After washing the targets are eluted using a high concentration of imidazole (300-500mM) for final purity checks and characterisation.

Purification of PfSir2a and PfAlba3 for biochemical and biophysical characterisation prior to SAXS/SANS experiments was achieved using IMAC with an ÄKTAPrime fast protein liquid chromatography (FPLC) system (GE Healthcare, Cat no. 18-1135-24). A 1ml and 5ml HisTrap™ FF crude column pre-packed with Ni-NTA agarose resin (GE Healthcare, Cat no. 11-0004-58 and 17-5286-01) were used for the purification of the PfSir2a and PfAlba3 respectively. Batches of 6x 1L bacterial cultures were prepared for purification with the aim to provide respectable yields (greater than 1mg of target protein per L of culture used) of high quality protein and bacterial cell pellets were stored frozen at -80˚C prior to purification.

2.1.11 PfSir2a Ni-NTA Purification Protocol

Soluble PfSir2a was obtained by thawing the cell pellets in a lysis buffer containing 50mM sodium phosphate, 500mM sodium chloride, 10mM Imidazole, 5 units (U) ml^{-1} Benzonase nuclease (Novagen® Cat no. 71205-3) 1x cOmplete EDTA-free Protease inhibitor cocktail (Roche Cat no. 11 873 580 11), pH 7.5. Once the cells were thoroughly defrosted and resuspended they were lysed by sonication using a Sonics Vibra cell VC750 sonicator (5 seconds on, 25 seconds off, 20 cycles at 50% amplitude). The cell lysate was centrifuged at 20,000 rpm for 30 min at 4°C and the supernatant was loaded onto a GE HisTrap Ni-NTA column pre-equilibrated with lysis buffer using a 1ml min^{-1} flow rate at 4°C.

Once the cell lysate finished passing through the column the resin was washed with 10 column volumes (CV) of wash buffer containing 50mM sodium phosphate, 500mM sodium chloride, 40mM Imidazole, pH 7.5. A step elution was performed to elute the His-tagged protein in 1ml fractions with an elution buffer containing 50mM sodium phosphate, 500mM sodium chloride, 500mM Imidazole, pH 7.5.

2.1.12 PfAlba3 Ni-NTA Purification Protocol

Soluble PfAlba3 was obtained by thawing the cell pellets in a lysis buffer containing 50mM sodium phosphate, 500mM sodium chloride, 10mM Imidazole, 10U ml^{-1} Benzonase nuclease (Novagen® Cat no. 71205-3) 1x cOmplete EDTA-free Protease inhibitor cocktail (Roche Cat no. 11 873 580 11), pH 7.5. Once the cells were thoroughly defrosted and resuspended they were lysed by sonication using a Sonics Vibra cell VC750 sonicator (10 seconds on, 59 seconds off, 10 cycles at 50% amplitude). The cell lysate was centrifuged at 20,000 rpm for 30 min at 4°C and the supernatant was loaded onto a GE HisTrap Ni-NTA column pre-equilibrated with lysis buffer using a 5ml min^{-1} flow rate at 4°C.

Once the cell lysate finished passing through the column the resin was washed with 100 CV of wash buffer containing 50mM sodium phosphate, 1M sodium chloride, 40mM Imidazole, pH 7.5. A step elution was performed to elute the His-tagged protein in 1ml fractions with an elution buffer containing 50mM sodium phosphate, 500mM sodium chloride, 500mM Imidazole, pH 7.5.

2.1.13 Ion Exchange Chromatography (IEX)

Proteins can also be further purified from residual contaminants by utilising ion exchange chromatography (IEX). Here target proteins were separated by charge differences using a mobile phase and a stationary phase. The stationary phase is a resin which contains a functional group that can be ionised to allow the binding of target molecules via charge interactions. In this work, cation exchange chromatography was used to separate negatively charged targets from residual bacterial contaminants using a GE Mono S 5/50 GL (GE Healthcare) cation exchange column.

This method was used as an additional purification step for PfSir2a. Pooled concentrated elution fractions of PfSir2a were dialysed against IEX buffer containing 50mM sodium phosphate, 50mM sodium chloride, pH 7.5 overnight at 4˚C. A 1ml sample was loaded onto a Mono S 5/50 GL cation exchange column (GE Healthcare) pre-equilibrated in IEX buffer A at a flow rate of 2ml min^{-1}. The ionic strength of the buffer was altered by increasing the ratio of IEX buffer B, containing 50mM sodium phosphate, 2M sodium chloride, pH 7.5, to IEX buffer A. The elution profile of the protein was monitored by UV 280nm and 260nm absorbance and 1ml fractions were collected. The length of elution was 20 CV.

2.1.14 Size Exclusion Chromatography (SEC)

A final purification step undertaken for target proteins was size exclusion chromatography (SEC), also called gel filtration. Here molecules are separated by their size as they migrate through a resin matrix with different pore sizes packed into a column. Large molecules or protein aggregates pass through the pores in the resin more quickly and are eluted first. Smaller molecules spend longer in the column as they pass enter the pores of the resin, eluting at a later time. This allows the removal of protein aggregates and/or the separation of oligomeric states of the target proteins studied. In this thesis GE *Superdex*™ S200 10/300 GL and *Superdex*™ S75 10/300 GL (GE Healthcare) analytical columns were used. The column is connected to a fast protein or high-performance liquid chromatography system (FPLC/HPLC respectively) and pre-equilibrated with a solvent before injecting a sample. It is connected to a fractionation device which separates the recovered sample based on size and can be monitored via UV absorbance (190nm-450nm).

PfSir2a and PfAlba3 were subjected to a further purification step by SEC using a *Superdex*™ S75 10/300 GL analytical column on an ÄKTA purifier HPLC system. A buffer containing 50mM sodium phosphate, 500mM NaCl, pH 7.0 was used for all samples. 250µl samples were centrifuged using a Spin-X centrifugal filter (Corning, Cat no. 8161) at 10,000 rpm for 5 minutes at 4˚C prior to being injected onto the column pre-equilibrated in SEC buffer. A 500µl sample loop was used to load the samples which were injected onto the column at a constant flow rate of 0.5ml min^{-1}. The elution profile of the proteins was monitored by UV 280nm and 260nm absorbance and 1ml fractions were collected. The length of elution was 1.5 CV.

2.2 Biophysical Characterisation

To confirm expression of the correct target proteins, additional methods were used to investigate the physical properties and study their behaviour. These are outlined in brief below.

2.2.1 Time Of Flight Mass Spectrometry (TOF -MS)

TOF-MS was used to calculate the mass of the proteins investigated in the thesis. The molecules present in a sample were ionised before being accelerated by an electric field of known strength before arriving at a detector. Ions with the same charge are accelerated at the same rate and the velocity of the molecule arriving at the detector depends on the mass to charge ratio. By observing the time of flight for particles to arrive at the detector at a known distance, the molecular mass of the protein can be determined. TOF-MS measurements were carried out at the *Institut de Biologie Structurale* (IBS, Grenoble) in collaboration with Luca Signor.

Samples of 30µl at a concentration of 30µM for both PfSir2a and PfAlba3 were prepared in 50mM sodium phosphate, 500mM sodium chloride, pH 7.5 buffer. Samples were then sent to be analysed by the Mass Spectrometry Platform at the Institut de biologie structurale (IBS, Grenoble). A on-line desalting step was performed with a protein trap (Zorbax 300SB-C8, 5um, 5x0.3mm, Agilent Technologies) for 3 minutes at a flow rate of 100 µl min^{-1} with solvent A containing 0.03% Trifluoroacetic acid (TFA), 5% acetonitrile (ACN). This was followed by elution at 50µl min^{-1} with 70% solvent B containing 95% CAN, 5% H_2O and 0.03% TFA. The samples were then analysed on a 6210 LC-TOF mass spectrometer interfaced with LC pump system (Agilent Technologies).

2.2.2 Fluorescence Thermal Shift Assay (TSA)

Fluorescence thermal shift assays (TSA or Thermofluor™) were carried out to determine protein sample stability and optimisation. In this technique, the thermal stability of a protein is assessed by the addition of a fluorescent dye (SYPRO Orange®, Invitrogen). This dye interacts non-specifically with hydrophobic residues of proteins and is strongly quenched by water.

There are two underlying principles behind TSA, the first being that the fluorescent dye has little or no interaction with the protein being studied. The second principle is dependent upon the unfolding of the sample whilst being heated which exposes the hydrophobic core of the protein. This allows for the fluorescent dye to bind to the inner hydrophobic regions and give rise to a fluorescence signal by excluding water. The fluorescence is proportional to the degree of the protein denaturation and optical measurements of protein unfolding as a function of temperature can be made.

This fluorescence can be observed increasing with temperature before reaching a maximum emission and gradually fading back down due to additional quenching of the dye at increased temperatures. This generates a stability curve for a protein and a melting temperature (T_m) can be determined by calculating the midpoint between the temperatures at which the lowest and maximum fluorescence emission occurs.

As the T_m of a protein is closely associated with its stability in solution, many conditions may be screened to find optimised buffers and/or ligands. These can include pH, buffering agent, salt concentration, co-factors and or metal ions. Using a quantitative PCR (qPCR) machine one can efficiently screen many conditions with a small amount of sample (~250µl at 10 - 100mM concentration). The excitation and emission wavelengths for the

SYPRO® dye are 470nm and 570nm respectively for which a MyIQ RT PCR instrument, Bio-Rad was used.

2.2.3 Dynamic Light Scattering (DLS)

DLS, also known as quasi-elastic light scattering (QELS) or photon correlation spectroscopy (PCS) is a physical technique used to estimate the size of particles contained within a solution, typically in the submicron range. In the case of proteins, this provides information about sample homogeneity (e.g. single species or monomer/dimer mixtures) and if aggregates remain within the solution which can hinder further experiments.

DLS uses a coherent monochromatic laser light source to illuminate particles in a sample solution. If the particles are smaller than the wavelength of the incident beam (less than 250nm) the light is scattered in all directions via Rayleigh scattering. As the particles are in a solution they undergo Brownian motion as they are bombarded by solvent molecules and so this causes the distance between scattering particles to be constantly changing with time. The scattered light waves from these particles will then either constructively or destructively interfere with each other, causing a fluctuation in intensity over time which is measured with a detector.

This timescale of these fluctuations provides information about the size of the scattering particles, larger particles will undergo slower Brownian motion than smaller particles which move more rapidly through interactions with the solvent molecules. Intensity patterns produced by a sample over several short time periods are recorded and analysed using an autocorrelation function which compares the intensity at each light spot produced over time. This function allows for the translational diffusion coefficient (D_T) of the particles to be calculated.

Using this coefficient, the size of the particles within the solution may be calculated using the Stokes-Einstein equation:

$$R_H = \frac{k_B \mathrm{T}}{6\pi\eta D_T}$$

Equation 2.1

Where R_H is the hydrodynamic radius of the particle, D_T is the translational diffusion constant, k_B is Boltzmann's constant (1.38 x 10^{-23} J K^{-1}), T is the absolute temperature (K) and η is the viscosity of the solvent (Einstein, 1905).

The size value given by the hydrodynamic radius in Equation 2.1 represents how a particle appears to move through a solution and it is important to accurately know the temperature and viscosity of the solvent to determine accurate values.

In this thesis, DLS experiments were performed on PfSir2a and PfAlba3 after SEC to provide information about the monodispersity and hydrodynamic radius of the proteins in solution in preparation for SAXS/SANS experiments. Typically, if a sample exhibits < 20% polydispersity it can be considered as 'monodisperse' and suitable for continuation into solution scattering investigations. If a sample exhibits > 20% polydispersity it is an indicator that there may be several species present within the sample such as different oligmeric states, contamininats or aggregates and the sample conditions should first be optimised before proceeding to solution scattering experiments. DLS measurements were performed using 60-100µl of protein sample taken immediately after SEC. Samples were filtered using Spin-X centrifugation columns at 10,000rpm for 5 minutes at 4˚C. Samples were then loaded into three-window quartz cuvettes (Hellma, Cat no. ZEN2112) by syringe with a 0.22µM Millex filter (Millipore Cat no. SLGV004SL). Measurements were

performed using a Zetasizer Nano S instrument (Malvern) by a series of 10-15 scans repeated three times for each measurement.

2.2.4 Size Exclusion Chromatography Coupled With Multi Angle Laser Light Scattering And Refractive Index (SEC-MALLS-RI)

In this method, several characteristics of a protein such as its absolute molecular mass, homogeneity and polydispersity can be observed at the same time using a combination of techniques in a multi instrument setup. A typical setup has a size exclusion column connected to a HPLC pump which has been pre-equilibrated in a solvent suitable for investigating the protein/particle of interest. A sample is injected and passes through the size exclusion column and is observed afterwards by a series of detectors.

A UV absorbance detector monitors the elution of macromolecules as they pass through the size exclusion column and generate an elution profile. Next the sample passes through a flow cell surrounded by light scattering detectors which measures the scattered intensity of a laser light source from the particles in the solution across different angles. A refractive index detector (RID) is also used to measure the difference in refractive index of the solvent as the macromolecules pass through the flow cell.

Using the information gathered from both these detectors, the absolute molecular mass of the macromolecule can be derived by using the Zimm equation described below (Zimm, 1945):

$$\frac{K^* c}{R_\theta} = \frac{1}{[M_w P_\theta]} + 2A_2 c$$

Equation 2.2

Where M_w = weight averaged molar mass (g mol^{-1}), R_θ = measured Rayleigh ratio, c = protein concentration (g ml^{-1}), A_2 is the second virial coefficient (a measure of solvent-

solute interaction) and K^* = is an optical parameter for the incident vertically polarised light given by:

$$K^* = \frac{4\pi^2 \eta_0^2 \left(\frac{dn}{dc}\right)^2}{N_A \lambda_0^4}$$

Equation 2.3

Where η_0^2 is the solvent refractive index and (dn/dc) is the refractive index increment, N_A is Avogadro's number (6.022×10^{23}) and λ_0 is the vacuum wavelength of the incident light (Wyatt, 1993). This is calculated across the elution profile and allows the determination of the homogeneity and polydispersity of the sample to be assessed. This can be cross correlated with static DLS measurements for consistency checks and sample stability assessment.

SEC-MALS RI experiments were performed to obtain a estimate for the molecular mass of both PfSir2a and PfAlba3 as they migrated through a SEC column. This technique can also be used to assess the monodispersity of a sample within an elution peak. Initial samples of PfSir2a and PfAlba3 were prepared at a minimum concentration of 2mg ml^{-1} after initial Ni-NTA and SEC purification step. A *Superdex* S200 10/300 GL analytical column was connected to DAWN HELEOS light scattering instrument and an optilab T-rEX refractive index detector (Wyatt Technology Corporation). A buffer solution containing 50mM sodium phosphate, 500mM sodium chloride, pH 7.0 was used to equilibrate the systems for 2 hours prior to sample injection. Sample volumes of 50µl were injected onto the column using a flow rate of 0.5ml min^{-1}.

2.2.5 Analytical Ultracentrifugation (AUC)

Analytical ultracentrifugation is another powerful and versatile method used to characterise the properties and behaviour of macromolecules in solution. AUC analysis is sensitive to the mass and shape of particles and is also non-destructive, enabling recovery of the sample for additional experiments. Two complementary methods of AUC are available, sedimentation velocity and sedimentation equilibrium. Sedimentation velocity experiments provide information about the size and shape of macromolecules based on hydrodynamic principles. Sedimentation equilibrium experiments reveals information on the molar masses of macromolecules in solution using thermodynamic principles. This can be used to calculate stoichiometry and association constants between protein-protein and protein-ligand complexes (Cole et al., 2009).

AUC experiments are performed by loading a sample into centrifuge and subjecting it to high acceleration spinning, typically between 40,000 to 60,000 rpm, which generates a centrifugal field of approximately 250,000*g*. This exerts a high centrifugal force upon the sample and macromolecules suspended in solution which, due to the mass and shape of the particles, will begin to sediment in layers forming boundaries. This causes a balancing of the gravitational potential energy of molecules with their chemical potential energy. By monitoring the rate at which the boundaries of these molecules move when reaching equilibrium, we are conducting sedimentation velocity experiments. If we determine the concentration gradient of macromolecules after the equilibrium point has been reached then we are conducting a sedimentation equilibrium experiment.

For the purposes of this work sedimentation velocity experiments have been carried out. In this case the forces acting upon a particle in solution exposed to a centrifugal field can be considered as shown in Figure 2.2.

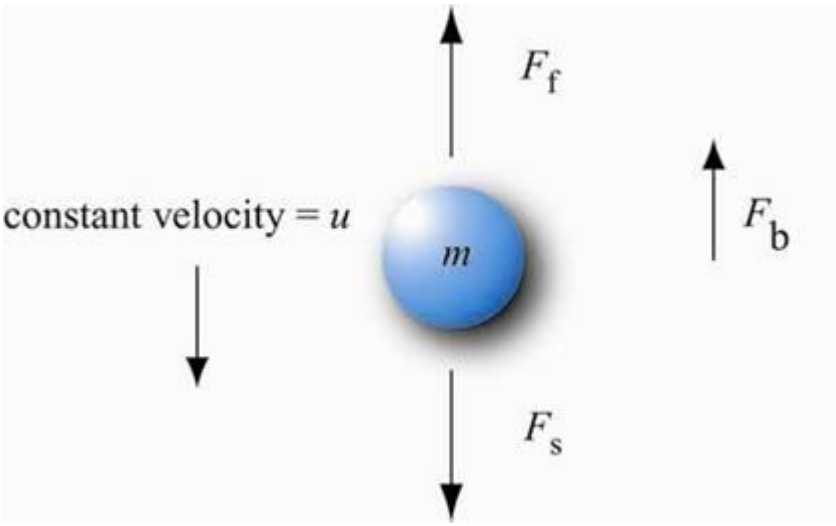


Figure 2.2: Graphical representation of the forces exerted on a particle experiencing high centrifugal force within a solution (Crates, 2003).

The force acting on a particle in a gravitational field is given by:

$$F_s = m\omega^2 r = \frac{M}{N_A}\,\omega^2 r$$

Equation 2.4

Where m is the mass of the particle, ω is the rotor speed in rad s^{-1} ($\omega = 2\pi \cdot \frac{\text{rpm}}{60}$) and r is the distance from the centre of the rotor, M is the molecular weight of the solute in g mol^{-1} and N_A is Avogadro's number. There is also a buoyancy force, F_b, exerted by the solvent on the particle:

$$F_b = -m_0\omega^2 r$$

Equation 2.5

Where m_0 is the mass of the fluid displaced by the particle given by:

$$m_0 = m\bar{v}\,\rho = \frac{M}{N_A}\bar{v}\,\rho$$

Equation 2.6

Here, $\bar{v}$ is the partial specific volume of the particle in $cm^3\ g^{-1}$, this represents the volume in ml that each gram of solute in solution occupies, effectively the inverse density of the particle. For proteins, this value is between 0.70 and 0.75 but for most globular proteins it can be approximated to 0.73. The density of the solvent is ρ in $g\ ml^{-1}$.

If the particle partial specific volume exceeds that of the density of the solvent, the particle will sediment and begin to sediment with a velocity which increases with radial distance travelled. Particles moving through a viscous fluid medium will also experience a frictional drag that is proportional to the velocity. This frictional force, F_f, is given by:

$$F_f = -fu$$

Equation 2.7

Where *f* is the frictional coefficient, which depends on the size and shape of the particle, and *u* is the velocity. Large, bulky or elongated particles will experience more drag force compared to those which are more compact spherical ones.

By collecting the and rearranging terms it can be shown that:

$$\frac{M(1-\bar{v}\,\rho\,)}{N_A f} = \frac{u}{\omega^2 r} \equiv s$$

Equation 2.8

The term ($u/\omega^2 r$) is called the sedimentation coefficient, *s*, and is defined as the ratio of the velocity per particle to the centrifugal field.

As the velocity of individual particles cannot be resolved, this coefficient is calculated by observing the change over time in the position of the boundary region in the sample cell. This is performed with a quartz sample cell and measuring the UV absorbance spectra (see Figure 2.3).

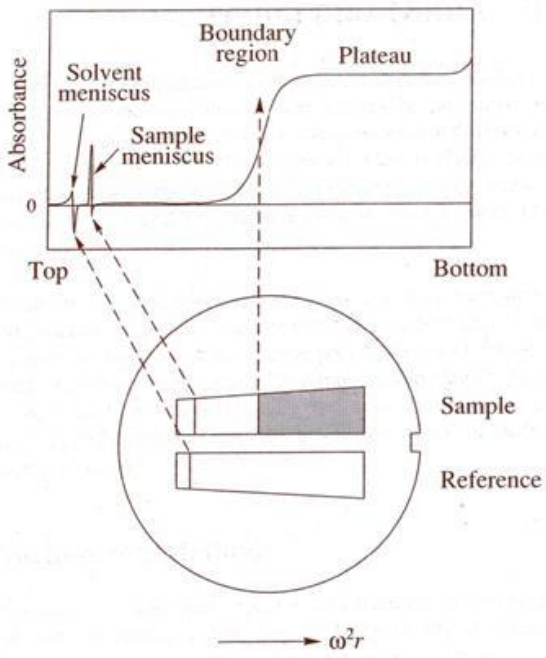


Figure 2.3: AUC sample and reference cell with UV output observing the boundary region between solvent – solute (Crates, 2003).

UV absorbance measurements also allows for the calculation of the concentration distribution of the sample using the Beer-Lambert law:

$$A = \varepsilon c l$$

Equation 2.9

Where A is the measured UV absorbance, ε is the molar absorption coefficient in L mol^{-1} cm^{-1}, c is the concentration in mol L^{-1} and l is the path length in cm (Beer, 1852). This information is particularly needed when conducting sedimentation equilibrium experiments

to find association constants between self-associating systems or macromolecular complexes.

By monitoring the rate of spreading of the boundary it is possible to determine the translational diffusion coefficient, D_T. Again, using the Stokes-Einstein relationship in the form of $D_T = RT/N_A f$ where R is the gas constant (8.314 J mol^{-1} K^{-1}) we can use ratio between s and D_T to find the molecular weight by:

$$M = \frac{s^0 RT}{{D_T}^0 (1 - \bar{v}\rho)}$$

Equation 2.10

The superscript zero for s and D_T indicate they have measured at different concentrations and extrapolated to the zero concentration value which removes the effects between particles and their movements. This enables the use of AUC as another characterisation method for the absolute molecular mass of a macromolecule in solution and determine any other oligomeric states or aggregation behaviour.

AUC experiments conducted for this thesis were performed using an XLI analytical ultracentrifuge, with a rotor Anti-50 (Beckman Coulter, Palo Alto, USA) and double-sector cells of optical path length 12, 3 and 1.5mm equipped with sapphire windows (Nanolytics, Potsdam, DE). PfSir2a and PfAlba3 samples were prepared in a buffer solution containing 50mm sodium phosphate, 500mM sodium chloride, pH 7.0. Acquisitions were made using interference optics at 278, 250 and 230nm UV absorbance at a temperature of 5˚C. The data were processed by Redate software v 0.2.1 and the analysis was done with the SEDFIT software, version 15.01b and Gussi 1.1.0.

This work used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

2.2.6 Isotheral Titration Calorimetry (ITC)

Isothermal titration calorimetry is a biophysical technique aimed at investigating the thermodynamic parameters of interactions between molecules in a solution to determine the constant of dissociation (K_D). Interactions which can studied include protein-ligand, enzyme-inhibitor, antibody-antigen, protein-protein, protein-DNA and protein-metal ions. This method works by detecting the changes in temperature between two identical cells that are made of very efficient thermally conductive, chemically inert materials such as gold or Hastalloy. Thermosensitive circuits detect the differences in temperature, ΔT, between the reference cell containing the buffer or water (ΔT_1), and the sample cell which also contains buffer and the macromolecule of interest (ΔT_2) (see Figure 2.4). Prior to adding any ligands into the sample cell, a small constant power is applied to the reference cell. This differential power (DP) forms part of a feedback circuit which is connected to another heater located on the sample cell used to maintain temperature equilibrium.

Throughout the course of an experiment, ligands are titrated into the sample cell in precisely controlled aliquots. If there is an evolution of heat (exothermic) in the sample cell due to an interaction, then there will be a negative change in the DP provided to the sample cell as it is no longer required to maintain temperature equilibrium. In an endothermic reaction, the opposite is the case where the DP must be increased to maintain constant temperature.

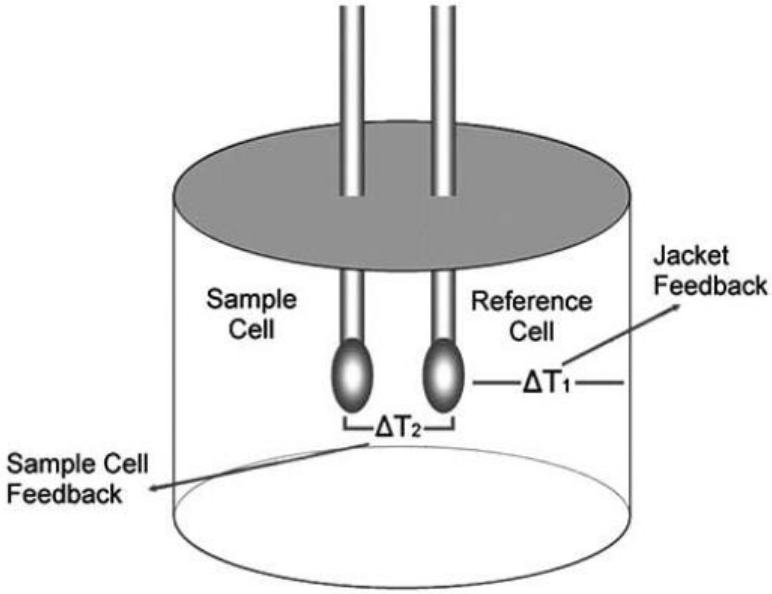


Figure 2.4: Schematic of ITC reference cell and sample cell. The sample cell typically holds a volume of 250µl (Dias, 2009).

Measurements are made of the time-dependant input of power required in order to maintain a temperature equilibrium between the two cells upon ligand addition. This results in a series of peaks and troughs as the system experiences a change in energy and is brought back into equilibrium over time. As the experiment progresses, the macromolecule in the sample cell becomes saturated with the ligand the signal will diminish, as only the heat of the dilution of the system is observed (see Figure 2.5).

The integral of the area under the peaks gives the change in energy ΔH in J, or enthalpy of the system, for each injection. The binding constant, K_B, is derived from the slope of the curve and is equivalent to $1/K_A$, where K_A is the binding affinity for the ligand-macromolecule interaction. This binding affinity is directly related to the total free binding energy available in the system, the Gibbs free energy ΔG in J, as follows:

$$\Delta G = -RT \ln K_A$$

Equation 2.11

Which can also be defined as:

$$\Delta G = \Delta H - T\Delta S$$

Equation 2.12

Where ΔS is the change in entropy of a system in J K^{-1} (Gibbs, 1873). Favoured processes or reactions occur when $\Delta G \leq 0$, with more negative ΔG indicating a higher affinity.

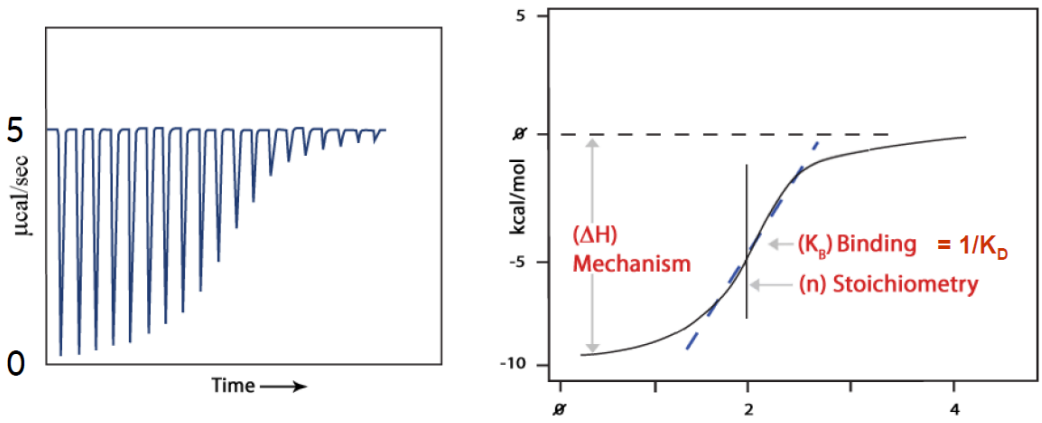


Figure 2.5: Left: graphical illustration of the measure of change in DP over time resulting from ligand addition to macromolecule. Right: Interpretation of data yielding information about the thermodynamics of the system (Milev, 2013).

This method helps to understand the mechanism and to determine the stoichiometry of interactions between macromolecules. In experiments for this thesis, PfSir2a and PfAlba3 were prepared at a concentration of 300µM and 30µM respectively in buffer solution containing 50mM sodium phosphate, 500mM sodium chloride, pH 7.0. A MicroCal iTC200 instrument (Malvern Instruments Ltd) was used to measure the samples. PfAlba3 was placed inside the sample cell and PfSir2a was titrated into the sample cell with 16 injections of 2.5µl volume. A spacing of 120 seconds was used between each injection. A series of experiments were conducted at 4°C and 15°C.

2.2.7 Micro-scale Thermophoresis (MST)

Micro-scale thermophoresis is another method aimed at quantifying biomolecular interactions. It functions in a different manner compared to ITC by measuring the motion of molecules as they move along microscopic thermal gradients induced by an infrared (IR) laser. Changes in the hydration shell, charge and size are observed by using a fluorescence marker covalently attached to a macromolecule of interest.

The MST experimental setup consists of preparing a fluorescently labelled macromolecule of interest at a fixed concentration. A dilution series with the target ligand/macromolecule is then prepared in several glass capillaries. The capillaries are then placed inside a machine in the path of an IR laser. Initially the molecules are evenly distributed in the solution inside of the capillary. Once the IR laser is turned on a local temperature gradient is induced in the capillary that causes the molecules to diffuse out of the heated spot (see Figure 2.6) (Duhr and Braun, 2006).

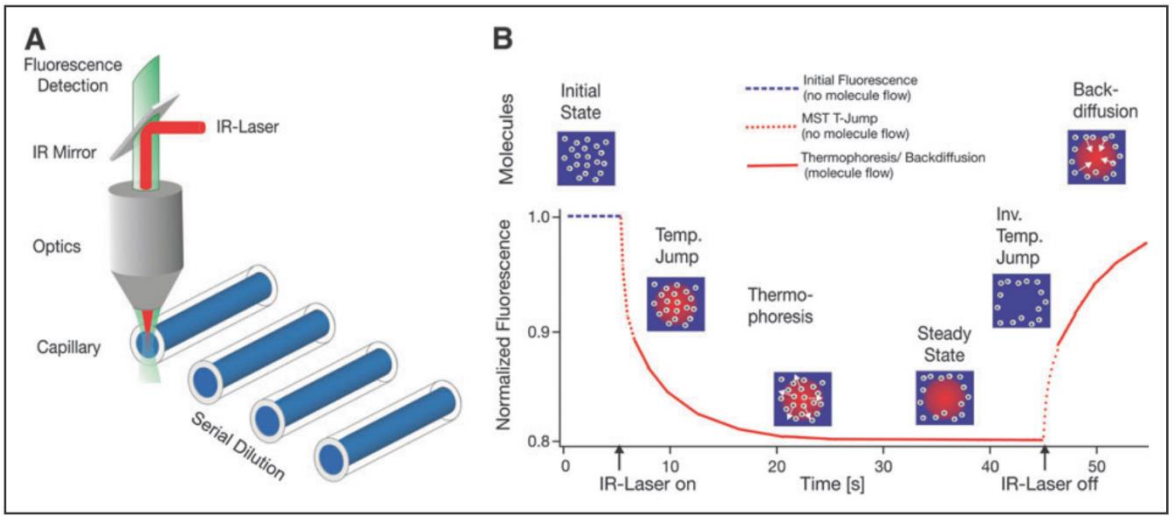


Figure 2.6: Diagram highlighting IR laser heating capillaries containing sample concentration series and components of the recorded signal as a fluorescence time trace (Jerabek-willemsen et al., 2011).

Local concentration changes caused by depletion of the solvated biomolecules in response to this heating is characterised by the following:

$$\frac{c_{hot}}{c_{cold}} = \exp(-S_T \Delta T)$$

Equation 2.13

Where S_T is the Soret coefficient given by:

$$S_T = \frac{D}{D_T}$$

Equation 2.14

In this case D_T is the thermally induced diffusion coefficient. This thermophoretic depletion depends on the interface between the macromolecules and the solvent. Under constant buffer conditions, this probes the size, charge and solvation entropy of the molecules.

Fluorescence in the heated area is measured before, during and after heating through the same objective lens used by the IR laser. Measurements begin at the initial steady state to determine the starting background fluorescence. As the IR laser is turned on, the signal observed decreases as molecules move out of the heated area due to thermal diffusion. After ~30s of heating (inducing a temperature gradient of between 2 - 6°C) the area is depleted of molecules. The signal increases again as the molecules move back into the spot due to normal mass related diffusion.

Analysis of the fluorescent signal produced during these stages across a serial dilution range reveals information about the bound and unbound states of the interacting

macromolecules-ligands. This allows for the characterisation of an equilibrium constant and the quantification of the binding affinity.

Full length recombinant PfSir2a was labelled with an amine reactive dye using the Monolith NT™ Protein labelling Kit Red-NHS (Nanotemper Cat no. L001). A 100µl stock of PfSir2a at a concentration of 20µM was prepared after Ni-NTA and SEC purification in 50mM sodium phosphate, 500mM sodium chloride, pH 7.5 buffer. The dye powder in the kit was suspended to make a 100µl stock volume at a concentration of 75µM in the same buffer as the PfSir2a protein stock. The protein and dye stocks were then mixed and incubated together at a 1:1 ratio for 30 minutes at 21°C, covered in aluminium foil to prevent any light exposure.

The labelled protein was then recovered from the reaction mixture by removing excess unbound dye. This was performed using Column-B supplied in the labelling kit which was washed three times with 3 ml of buffer to equilibrate prior to adding the labelling mixture. The labelled protein and dye mixture was then added to the column and recovered after passing through the column by adding buffer to Column-B and collecting the flow through.

MST experiments were performed with a 16-point serial dilution (1:2) of unlabelled PfAlba3 from 20µM to 0.6nM. This was incubated with a fixed concentration of labelled PfSir2a at 20nM for 30 minutes at 21°C, covered in aluminium foil to prevent light exposure. After incubation, the incubated solution was taken up by capillary action into Nanotemper standard MST capillaries (Nanotemper Cat no. K002) and placed inside a Nanotemper Monolith NT.115 MST device for analysis. A fixed LED power of 50% was used which excites the fluorescent dye, while the infrared laser power which heats the sample was performed at three different settings (MST power).

2.3 Structural Characterisation Methods Using X-ray And Neutron Scattering

To study proteins, DNA or other macromolecules that are of the order of between 1 – 100nm in size, it is important to use methods which are sensitive to these length scales. X-rays and neutron scattering as tools for this purpose has been widely used and characterised previously (Feigin and Svergun, 1987; Jacques and Trewhella, 2010; Svergun and Koch, 2003). They are well suited to this task as their wavelength range can provide high resolution data on atomic distances, in diffraction experiments, as well as lower resolution (in the region of 15–20 Angstroms) structural envelope and size information using small angle scattering experiments. Subsequently the experiments conducted for this thesis have made extensive use of X-ray and neutron scattering techniques and the general principles of the methods and their experimental parameters are outlined below.

2.3.1 Small Angle Scattering (SAS) With X-rays And Neutrons

X-rays and neutrons interact with matter in fundamentally different ways, but their behaviour can be treated in the same mathematical fashion. Both methods have associated advantages and disadvantages as investigative techniques for structural biology. X-ray photons, being electromagnetic radiation, are scattered by a sample via interactions with the electrons of atoms. Typical energies produced in modern synchrotron sources for structural experiments are around 10keV with wavelengths λ between 0.1 - 0.15nm. For neutrons, which are particles that have wave-like properties due to the de Broglie relationship (Broglie, 1924), their interaction occurs with the atomic nuclei of the sample. Thermal neutrons produced from fission or spallation sources for scattering experiments have a typical wavelength range of between 0.2 - 1nm.

The atoms of objects illuminated by an incident monochromatic plane wave with wave vector $k_i = |\boldsymbol{k_i}| = \frac{2\pi}{\lambda}$ become the source of spherical scattered waves. This is a result of the scattering centres appearing as 'point-like' due to the range of the interaction between the incoming X-ray or neutron radiation is much smaller than the wavelength of the incoming radiation. Scattering for X-rays and neutrons can occur by elastic scattering, where the direction of the incoming wave vector is changed but not the magnitude, or inelastic scattering, which causes a change in both the direction and magnitude of the incoming wave vector. Inelastic scattering confers information about the motion and dynamics of molecules but not directly their structure. As such, for the purposes of the experiments conducted in this thesis we will only consider elastic scattering.

In this case the scattered wave k_f is equivalent to the modulus of the incident wave vector $k_f = |\boldsymbol{k_f}| = k_i$. The observed scattering is the amplitude of coherently scattered waves and is given by the scatter length of the atom encountered by the incident radiation. For X-rays the atomic scattering length, f_x, is given by $f_x = Zr_0$ where Z is the atomic number and r_0 is the Thomson radius for an electron (2.82 x 10^{-13} cm). As a result X-ray scattering increases in sensitivity linearly with atomic number, as the number of electrons is proportional/directly related to the atomic number, and is more sensitive to heavy metals than lighter elements such as hydrogen, carbon and nitrogen.

For neutrons the scattering length is given by $f_n = f_p + f_s$. This is due to neutrons interacting with the nuclear potential of an atom, f_p, and the spin of the atom, f_s. Unlike X-rays, f_p does not increase linearly with atomic number does not follow any particular trend with respect to atomic number and lighter elements are as visible to neutrons as heavier elements. The spin term only reveals structural information if the spin of the incoming radiation and the nuclear spins in the sample are aligned, otherwise it only yields a flat

incoherent background. In many cases the incoherent background produced can exceed that of the coherent scattering signal and lead to severe experimental difficulties by degrading the signal to noise ratio.

This has several significant consequences when considering scattering via X-rays or neutrons. Firstly is that neutrons are more sensitive to lighter elements compared to X-rays (see Figure 2.7). Secondly there is a difference between isotopes of the same element from the nuclear potential scattering term. This behaviour also enables neutron scattering to become sensitive to isotope substitution, the most important of which is that of hydrogen and deuterium which have scatter lengths of -0.374 x 10^{-12}cm and 0.667 x 10^{-12}cm respectively.

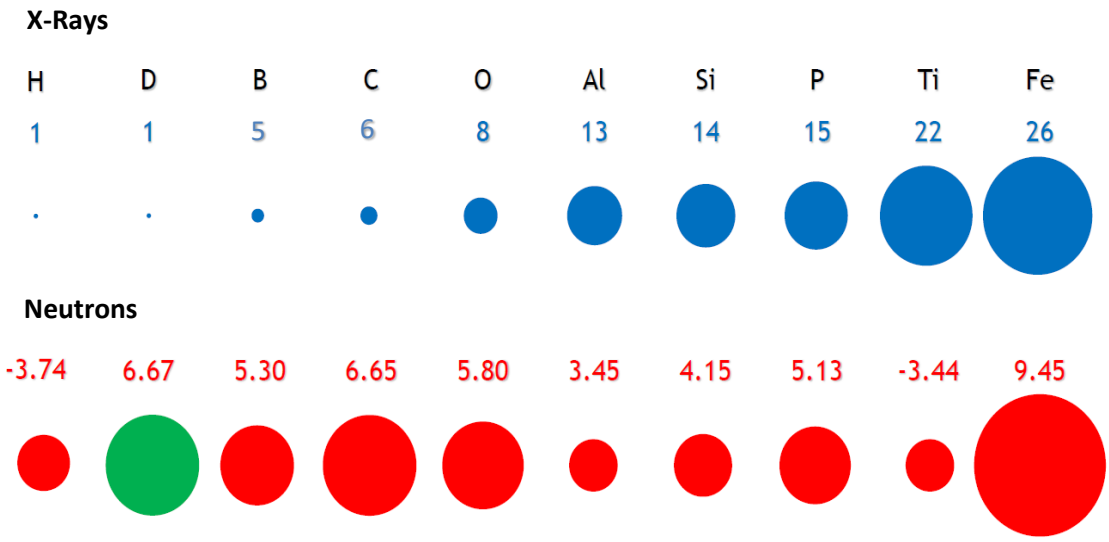


Figure 2.7: Comparison between the scatter lengths of elements by X-rays and neutrons (10^{-12} cm). Source private communication from Professor V.T.Forsyth.

In SAS experiments samples are illuminated by monochromatic X-rays or neutrons produced by sources such as the European Synchrotron Radiation Facility (ESRF, Grenoble) and Insitut Laue-Langevin (ILL, Grenoble). The distribution of scattered

radiation is observed by a detector after the interaction with the sample by the incident X-rays or neutrons. A typical experimental setup can be seen in see Figure 2.8. In this thesis SAXS experiments were performed using the BM-29 beamline (ESRF, Grenoble) with the experimental setup as described in Chapter 4.2. SANS experiments were performed using the D22 beamline (ILL, Grenoble) with the experimental setup as described in Chapter 5.2.

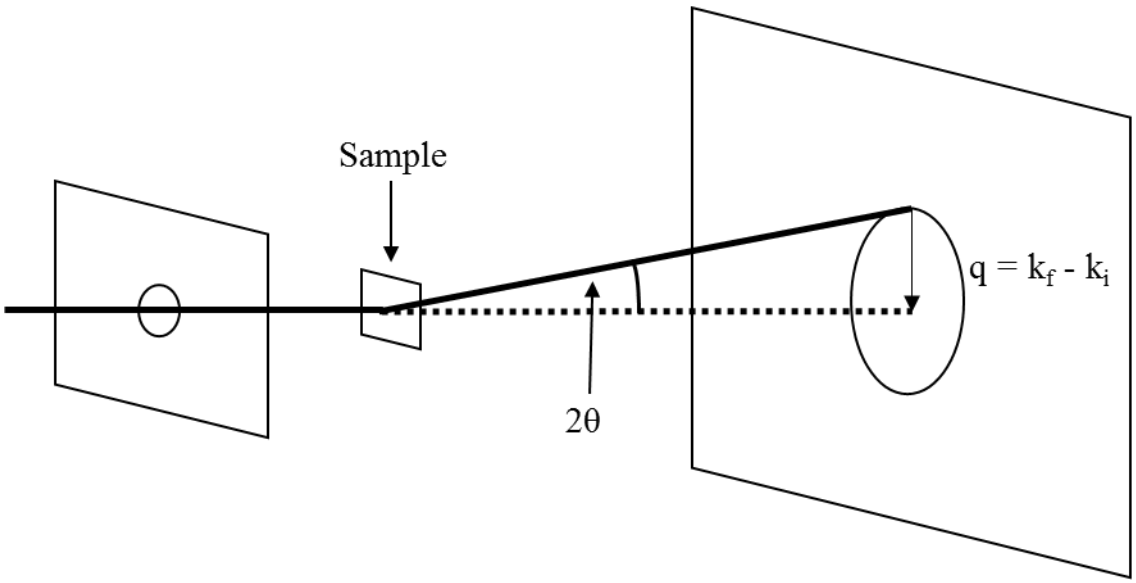


Figure 2.8: Experimental setup for SAS experiment. Incident monochromatic radiation beam incoming from the left interacts with the sample and is scattered in the 2θ direction.

The scattering vector *q* (sometimes denoted as *s* in the case for X-rays) is given by:

$$q = |\boldsymbol{q}| = \left|\boldsymbol{k_f} - \boldsymbol{k_i}\right| = \frac{4\pi \sin\theta}{\lambda}$$

Equation 2.15

Where λ is the wavelength of the incident radiation and the scattering angle is given by 2θ (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013).

When measuring macromolecules, which contain an assembly of many scattering atoms and of different elemental composition, it is more convenient to calculate the average

scatter length density distribution ρ(r). This is equal to the sum of the total scatter lengths of the atoms per unit volume of the macromolecule. If we take a typical sample, i.e. a homogenous protein suspended in a solution, we assume that the solvent will be a featureless matrix of a constant scatter length density, ρ_s.

The difference in scattering amplitudes from a single particle from that of an equivalent solvent volume is given by:

$$\Delta\rho(r) = \rho(r) - \rho_s$$

Equation 2.16

This is related to the scattering amplitudes of the particles in the sample by a Fourier transform:

$$A(q) = \Im[\rho(r)] = \int_V \Delta\rho(r) \exp(i\boldsymbol{qr})\, d\boldsymbol{r}$$

Equation 2.17

The amplitude, *A(q)* cannot be measured directly. However, it is possible to measure the intensity of the scattered waves which is a product of the amplitude and its complex conjugate:

$$I(\boldsymbol{q}) = A(\boldsymbol{q})A^*(\boldsymbol{q})$$

Equation 2.18

If particles in a sample are randomly distributed and their positions and orientations uncorrelated then their scattering intensities are summed. The scattering pattern observed by the detector corresponds to the convolution of the particle density distribution describing the positions and orientations of the particles within the sample. For a homogeneous suspension of biological macromolecules, randomly orientated, this

produces an isotropic scattering pattern from the summed intensities averaged over all orientations of the particle$(q) = \langle I(q) \rangle_{\Omega}$. This represents a particles structure in reciprocal space and by using the Fourier transform we can convert the measured intensities to real space co-ordinates to find its size and low-resolution shape in real space.

The real space structure in biological context refers to the organisation of amino acids into higher order secondary structures (e.g. alpha helices and beta sheets) and/or tertiary and quaternary shapes. This builds the overall topology of biological macromolecules and determines how they may perform functions such as enzymatic reactions or energy transfers which are uniquely interesting. Solution scattering provides the opportunity to observe the overall shape envelope of these molecules and can see changes in this envelope for single macromolecules/complexes as they perform certain functions or interact with binding partners.

2.3.2 Deuteration And Contrast Variation For SANS

As mentioned previously, the key advantage of neutrons over X-rays is that they are scattered more easily by light elements and can distinguish between isotopes. This is exceptionally valuable in the case of hydrogen and its isotope deuterium in the preparation of biological samples. By labelling proteins in a perdeuterated form (100% exchange of hydrogen atoms to deuterium atoms) significant gains can be made in the area of neutron crystallography to isolate and identify important water molecules in protein active sites to understand reaction mechanisms. However, this can be challenging as many proteins can change from the resulting isotope substitution and become unstable, no longer fold/function correctly or no longer crystallise/produce crystals of sufficient volume for use in neutron experiments.

A more useful approach can be used instead by producing partially deuterated proteins (75%) for use in solution contrast match-out experiments. This is particularly useful in the case of protein-protein and protein-DNA complexes. Here we can change the ratio of H_2O:D_2O in the solvent to alter its scatter length density. By doing so, different components of a protein-protein or protein-DNA complex can become 'matched out' as they possess an identical scatter length density to that of the solvent and cannot be distinguished (see Figure 2.9). This allows for individual components of a complex to be resolved to reveal information about the size and shape of the visible partner to identify any structural changes that can occur resulting from any interactions.

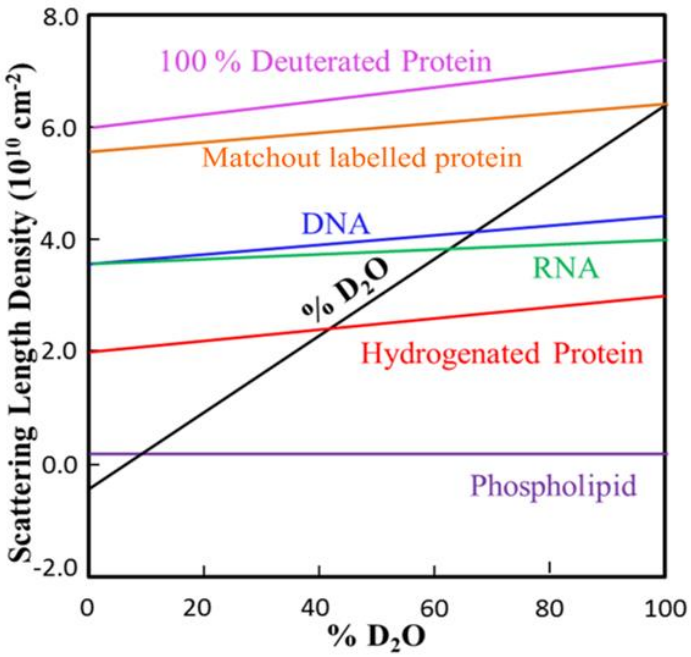


Figure 2.9: Change in scatter length density of H_2O solvent with increasing D_2O ratio and respective match points for protein, DNA/RNA and lipid systems. Components of a system are matched out where they intersect with the increase in %D_2O of the solvent (Haertlein et al., 2016).

For a typical protein-protein complex experiment, one partner will be produced in a 75% deuterated state while the other remains the normal hydrogenated (or protonated) form. In

normal 100% H_2O solvent conditions, both partners of the complex can be seen via neutron scattering providing information about the size and structural envelope of the overall complex. At approximately 40% D_2O content of the solvent, the hydrogenated protein will become 'matched out' leaving only the deuterated partner visible (see Equation 2.16). At 100% D_2O concentration the opposite case is true whereby the deuterated protein will be indistinguishable from the solvent but the hydrogenated protein can be resolved (see Figure 2.10).

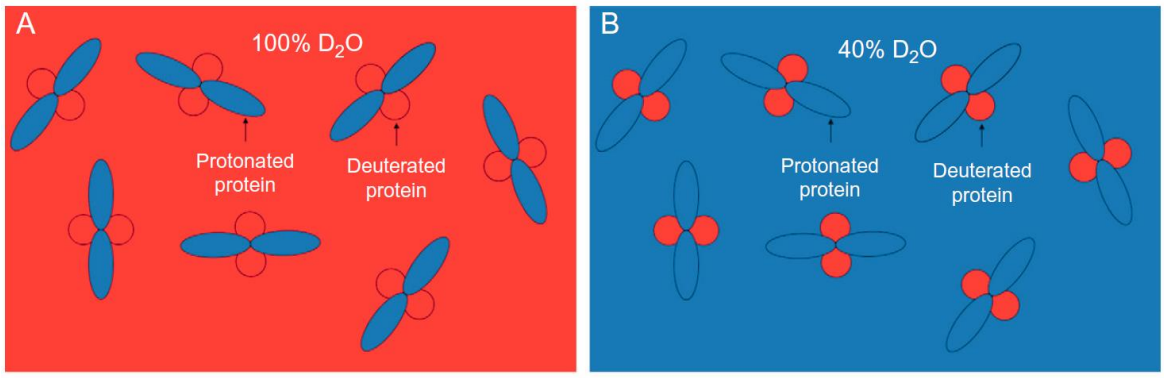


Figure 2.10: Illustration of contrast labelling and 'matched out' proteins (Haertlein et al., 2016). (A) 100% D_2O solvent – deuterated proteins matched out, protonated proteins remain visible. (B) 40% D_2O solvent – protonated proteins matched out, deuterated proteins now become visible.

2.3.3 Deuterated Protein Production

To prepare deuterated proteins in preparation for contrast match out experiments, the bacterial expression system for both PfSir2a and PfAlba3 was adapted to growth in minimal media growth conditions. Firstly, this was accomplished in normal hydrogenated conditions. Here a glycerol stock of the bacterial expression system was inoculated into 10 ml minimal media growth conditions, with 35 mg ml^{-1} kanamycin sulphate, by a series of three culture passages over a 72-hour period (recipe in Appendix A). Using 1 ml of the previous culture, the bacteria were left to grow for 16 hours overnight at 37°C. The growth

of the bacterial cultures was assessed by monitoring the $O.D_{600}$ reading before performing each successive passage.

After 72 hours, expression tests were performed to ensure recombinant protein production was still present. Glycerol stocks using 1 ml of the final culture after 72 hours adaptation to the minimal media growth conditions was prepared for storage at -80°C. This process was repeated for deuterated cultures by using the same minimal media recipe dissolved in an 85% D_2O solution. After the adaptation process, the expression tests were performed again to check for expression of the recombinant proteins and assess solubility levels.

Chapter 3: Biochemical And Biophysical Characterisation

Abstract

This chapter focuses on the biochemical and biophysical investigations performed on PfSir2a and PfAlba3 in preparation for structural characterisation by SAXS/SANS techniques. Expression of recombinant PfSir2a and PfAlba3 was performed using the *E. coli* bacterial expression systems and optimised for the maximum recovery of soluble forms of both proteins. These were subsequently purified using Ni-NTA, IEX and SEC method to achieve high purity samples required for further biochemical, biophysical and structural experiments.

Investigations by DLS, SEC-MALS and TOF-MS indicated both PfSir2a and PfAlba3 samples were monodisperse and of the correct molecular weight expected for the full length recombinant forms of each protein. PfSir2a was found to be generally stable in solution whereas PfAlba3 exhibited a tendency towards rapid aggregation shortly after purification requiring SEC steps immediately before any characterisation methods were attempted. Further analysis by AUC revealed that PfSir2a exhibited a concentration dependent oligomerisation effect whereas PfAlba3 was found to remain as a monomer in solution.

Pulldown experiments using tagged and non-tagged versions of PfAlba3 and PfSir2a respectively indicated an extremely weak interaction between the two proteins. This was investigated using MST and ITC to characterise the binding affinity and stoichiometry of the interaction but could not be successfully performed due to the unstable nature of PfAlba3 and oligomerisation effect exhibited by PfSir2a.

3.1 Introduction

The focus of this chapter is the production, characterisation and optimisation of recombinant target proteins PfSir2a (33.8 kDa) and PfAlba3 (13 kDa). This was essential preparation work for structural experiments carried out using SAXS and SANS. The background to this work and methods has been explained in Chapters 1 and 2 and work was carried out using the facilities at both Keele University (Staffordshire, UK) and the partnership for structural biology (PSB, Grenoble).

The overall strategy for this work was to produce recombinant PfSir2a and PfAlba3 using the pET28a expression system and isolate the individual proteins by means of IMAC and IEX purification. The overview of the plasmid preparation and expression systems used is covered in section 3.2. Further purification was carried out using SEC to prepare samples for further biochemical and biophysical characterisation. This includes TOF-MS, TSA, DLS, SEC-MALS-RI, AUC, ITC and MST and is covered in sections 3.4 onwards.

3.2 Production Of PfSir2a And PfAlba3

This section details the results of the cloning work carried out to prepare PfSir2a and PfAlba3 for expression studies.

3.2.1 PfSir2a Cloning

PfSir2a PCR products were obtained as described in Chapter 2, section 2.1.3 and run on a 0.8% agarose gel at 80V in 1x TAE buffer for 1 hour and recovered by gel extraction kit. Purified products were digested for 2 hours with BamHI/XhoI restriction enzymes, then phenol/chloroform extracted and ethanol precipitated as previously described (Chapter 2, section 2.1.5). Cleaned Sir2a insert was then ligated into double-digested pET28a vector using a 20 µl T4 DNA ligase reaction (ThermoFisher Scientific Cat no. EL0014) at a 1:4 ratio in favour of PfSir2a insert. The ligation reaction was carried out overnight at 16˚C

and ethanol precipitated as previous with the addition of 1µl 10mg ml^{-1} Yeast tRNA (ThermoFisher Cat no. AM7119) to act as a carrier.

The ligated pET28a vector was then transformed into DH10b competent cells and incubated overnight. Single colonies from overnight growth were isolated and subjected to colony PCR to check for the PfSir2a sequence. The colony PCR results of selected PfSir2a colonies can be seen in Figure 3.1. The expected product size for PfSir2a insert plus flanking regions from T7 upstream and downstream primers is 1138bp. Several bands of this approximate size were detected in the gel images in Figure 3.1. However, several reactions show amplification of additional products and are thus not clean PCR reactions. Six selected colony PCR reactions (highlighted in red in Figure 3.1) showed relatively clean products and the respective colonies were chosen for culture and plasmid recovery via miniprep (Qiagen Cat no. 12125).

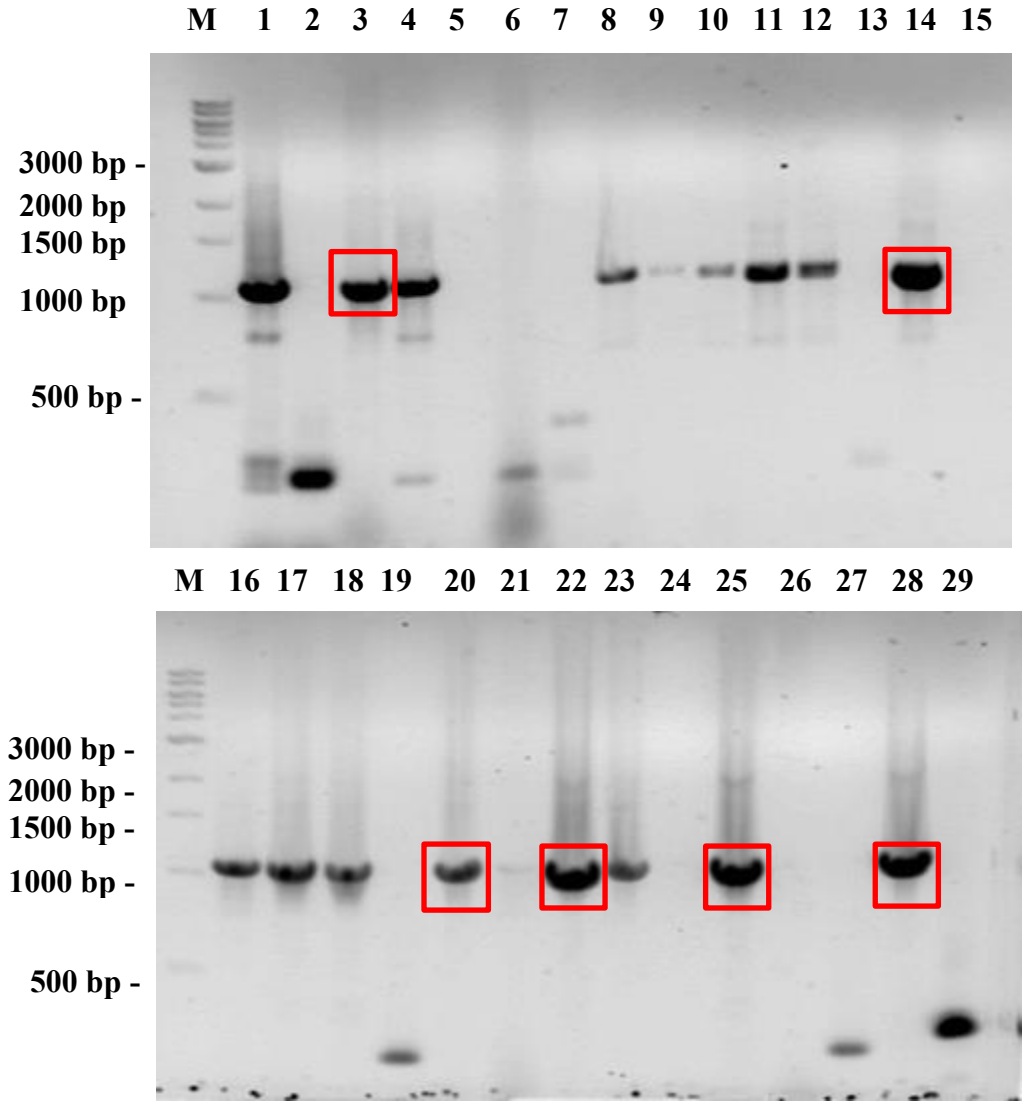


Figure 3.1: Colony PCR products of PfSir2a colonies. Lane 1 positive control (PCR templated on ligated pET28a vector with PfSir2a insert). Lane 2 negative control (unligated pET28a vector). Lanes 3 – 30 individual colony PCR products. Highlighted boxes indicate colonies selected for sequencing.

Plasmids recovered from the overnight cultures were digested with BamHI and XhoI restriction enzymes in 40µl reactions and digested products were run on a 0.8% agarose gel (see Figure 3.2). Digested products of the expected size for the inserted PfSir2a gene, 822bp, were seen in lanes 1, 3, 4 and 5. Respective colonies were then re-grown to harvest more pET28a vector containing PfSir2a insert and were sent off for sequencing.

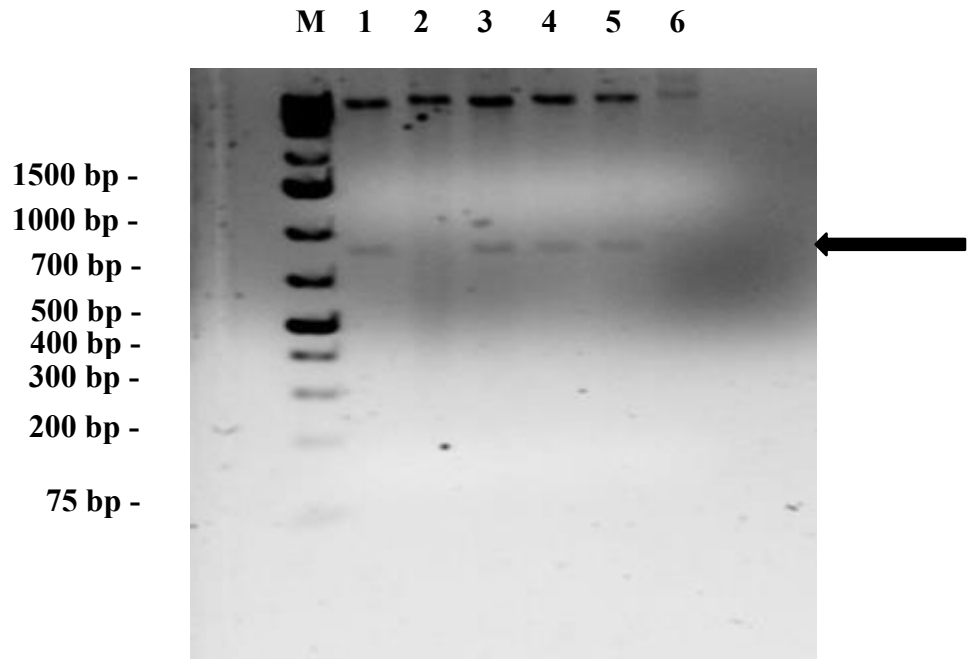


Figure 3.2: PfSir2a DH10b clones, digested with BamHI and XhoI restriction enzymes on 0.8% agarose gel. Digested PfSir2a inserts can be seen in lanes 1, 3, 4 and 5 at just below the 700bp marker as indicated by arrow.

3.2.2 PfAlba3 Cloning

A codon-optimised PfAlba3 gene insert was purchased from Eurofins, MWG Operon. The sequence is available in the Appendix B. This was excised from pEX-A vector (Eurofins, MWG Operon) by restriction digest with NcoI and XhoI enzymes as described in Chapter 2, section 2.1.5. The digested PfAlba3 product was then gel extracted (Figure 3.3 A) and digested a second time before ethanol precipitation and ligation into digested pET28a vector. Figure 3.3 B shows ethanol-precipitated digested pET28a vector in lane 1 and PfAlba3 in lane 2. Approximately 80ng μl^{-1} digested vector and 6.25ng μl^{-1} PfAlba3 insert was recovered after ethanol precipitation. Ligation reaction for pET28a and PfAlba3 was carried out in a 1:5 ratio in favour of PfAlba3 insert (7.8ng pET28a vector, 39ng PfAlba3).

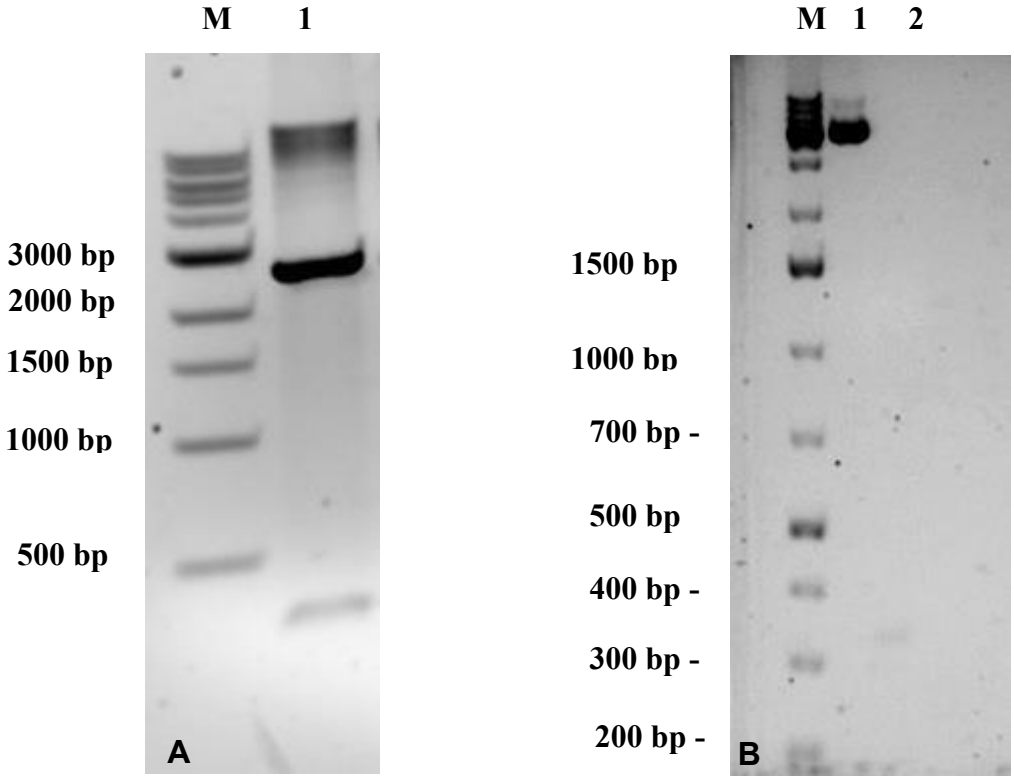


Figure 3.3: (A) pEX-A/Alba3 digestion products on 0.8% agarose gel from 200ng of plasmid DNA. (B) Recovered ethanol-precipitated pET28a and Pfalba3 (lanes 1 and 2 respectively) after digestion and before ligation.

In Figure 3.4, the digested products of DH10b cells transformed with ligated vector containing PfAlba3 insert are shown. Products are seen in each of the 6 recovered plasmid samples digested in the expected size region for PfAlba3, 324bp. Remaining undigested plasmids were sent for sequencing.

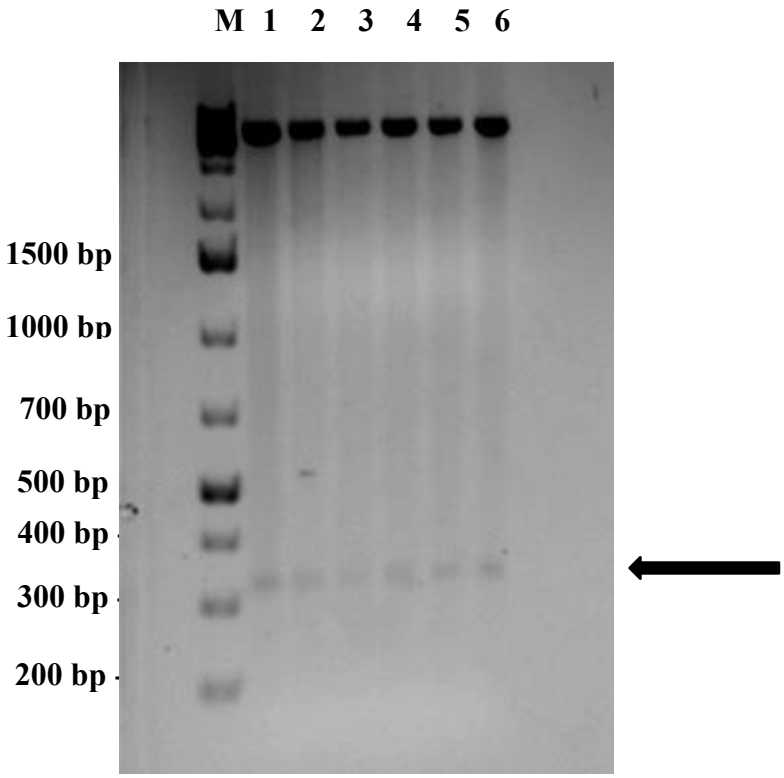


Figure 3.4: Products of digest reaction performed on recovered pET28a plasmids after ligation with PfAlba3 sequence insert and transformation into *E. coli* DH10b cloning strain. Digested PfAlba3 inserts are visible as bands just above the 300 bp marker as indicated by the arrow. Plasmids digests were performed with NcoI and XhoI enzymes for 1 hour at 37°C.

3.2.3 Sequencing Of Expression Vectors

Recovered pET28a plasmids containing sequences for PfSir2a and PfAlba3 were sent for sequencing analysis to confirm insertion of sequences ready to be used for expression tests. The results returned revealed two important features about the sequences which were ligated into the expression vectors. The most significant observation affected the expression vector for PfAlba3. It was found that an error had occurred in the initial sequence and restriction digest design for insertion into the native pET28a vector. The starting Methionine codon in the pET28a vector is included as part of the NcoI cut site

sequence. This was not accounted for in the design of the codon optimised PfAlba3 sequence which, when digested and ligated into the pET28a expression vector, resulted in an offset of the open reading frame (ORF). This meant the PfAlba3 sequence would not be translated correctly (see Figure 3.5).

CCATGGATGGCAAGCACCGAAGAAGTCTCGCAAGAACGCTCAGAAAATAG

NcoI : C'CATGG

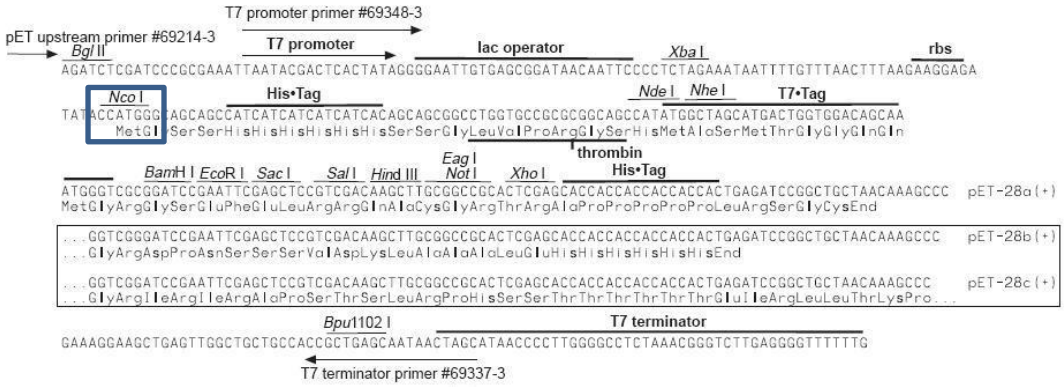


Figure 3.5: N-terminal codon optimised sequence of PfAlba3 with the NcoI cut site highlighted in red and the start ATG codon of the *P. falciparum* sequence underlined in black. When digested and ligated into the pEt28a vector the NcoI site (highlighted in blue) forms the start codon for the sequence to be translated. With the NcoI site added to the beginning of the PfAlba3 sequence this becomes out of frame and cannot be translated correctly.

Additionally, it was noted that the PfSir2a sequence contained a single mutation of an adenine residue to a guanine at nucleotide position 279 in the PfSir2a sequence, position 433 in sequencing data (see Figure 3.6). However, this mutation is in frame with the third codon of the gene sequence. Both the original sequence at this point, GAA, and the mutated sequence, GAG, encode for the same amino acid, Glutamic acid. Further work

was carried out using the mutated sequence as the amino acid composition of the translated sequence would not change.

```
CO PfSir2a    247 ATCCGTGATATTTCTAGTGATTACGAAATCGAAATTAATAACGGCCACGT
                  ||||||||||||||||||||||||||||||.||||||||||||||||||
S27 F         401 ATCCGTGATATTTCTAGTGATTACGAAATCGAGATTAATAACGGCCACGT
```

Figure 3.6: Sequence alignment of codon optimised PfSir2a sequence (CO PfSir2a) vs the returned sequence results for the plasmid recovered from colony 27 (S27F). The adenine in position 279 (highlighted in red) has been mutated to a guanine but is a silent mutation.

However, the frame shift for the PfAlba3 sequence would need to be corrected before expression studies could begin. It became necessary to perform site directed mutagenesis on the sequences to have them in the correct reading frame and allow for the translation of the target protein sequences in the *E. coli* BL21 DE3 expression strain. This was performed using a QuickChange lightning site-directed mutagenesis (SDM) kit from Agilent Technologies (Cat no. 210518) (see Chapter 2, section 2.1.6). It was necessary to design primers to remove the GATG sequence from the pET28a ligated sequence, removing the start ATG codon of the native sequence and the G base from the end of the original NcoI cut site. This would allow for the ATG inside the NcoI cut site to remain as the start codon as intended (see Appendix B for primer sequence).

3.2.4 Transformation And Expression Of PfSir2a And PfAlba3

Transformation of expression plasmids for PfSir2a and PfAlba3were carried out as described in Chapter2, section 2.1.7. Expression tests were conducted by selecting colonies from the LB agar plates and placing them in 10ml flask cultures containing 35µg ml^{-1} kanamycin sulphate. Cultures were grown at 37˚C shaking at 200rpm until an $O.D_{600}$ of 0.6 was reached and a 1ml uninduced sample was taken. The bacteria in the sample were pelleted by centrifugation at 14000rpm for 2 minutes and the excess media removed before

suspension in 50µl of 1x SDS-PAGE sample buffer. Cultures were then induced by the addition of 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and left to grow for 3 hours. After this period, an induced sample was taken and normalised to an OD_{600} of 0.6. The bacterial cultures were harvested by centrifugation at 5000rpm, 4˚C for 15 minutes. The excess LB media was removed and the bacterial pellets were stored at -20˚C for future solubility and extraction studies. Samples were heated to 100˚C for 10 minutes after mixing with SDS-PAGE sample buffer and expression levels of the selected colonies for both PfSir2a and PfAlba3 were analysed via SDS-PAGE gel run at 100V for 90 minutes and visualised by coomassie staining (see Figure 3.7).

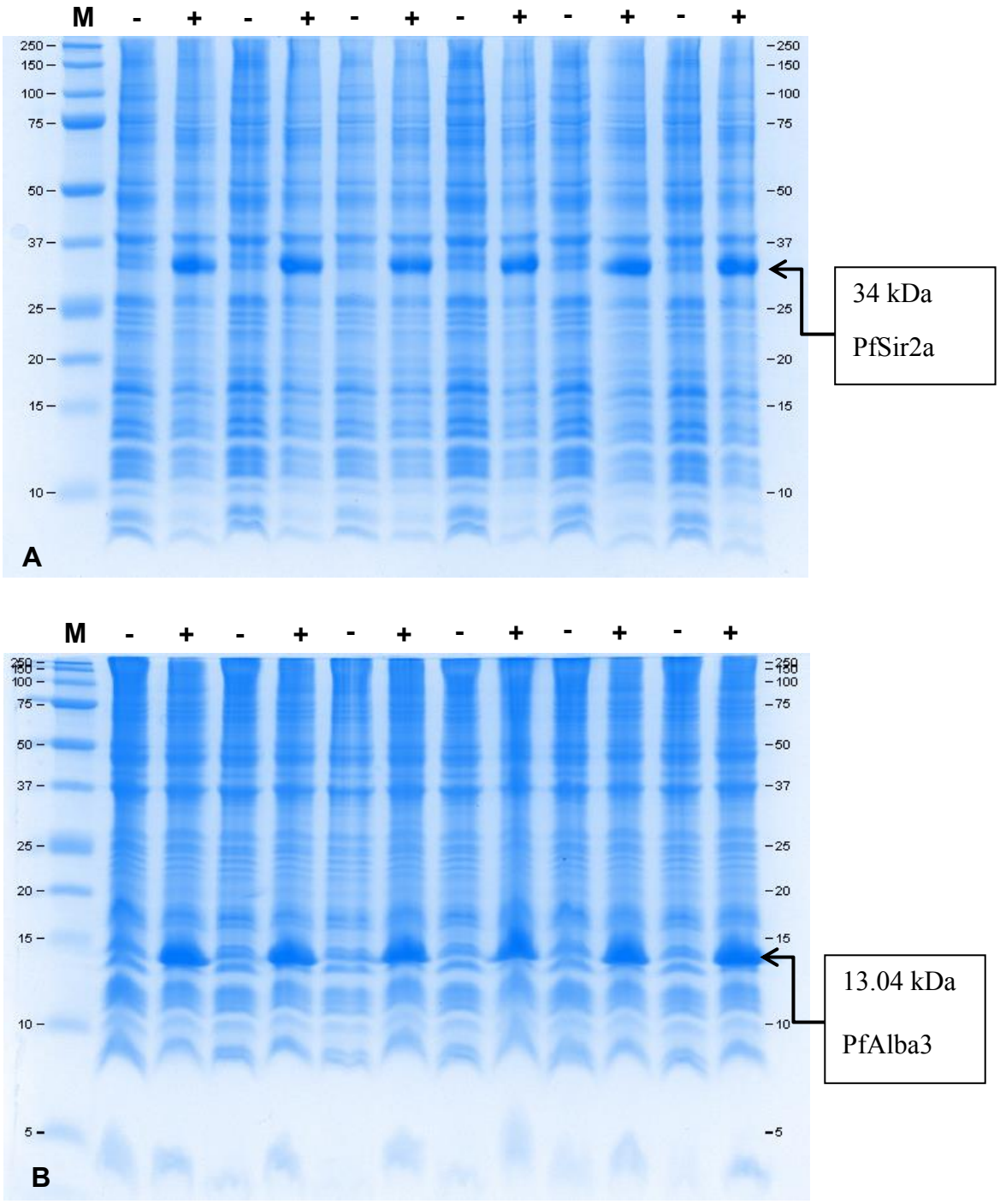


Figure 3.7: PfSir2a (A) and PfAlba3 (B) expression test results from six colonies as analysed by 12% and 16.5% SDS-PAGE respectively (10µl of sample per lane). M indicates molecular weight marker lane. Uninduced and induced lanes marked by – and + respectively.

Expression of the tagged constructs was also confirmed via Western blot analysis as described in Chapter 2, section 2.1.9. Uninduced and induced samples for both PfSir2a and PfAlba3 were analysed and can be seen in Figure 3.8. Here we see a clear induction band for both the recombinant forms of PfSir2a and PfAlba3 in the induced lanes. Additional bands were located at approximately 10 kDa in size in both uninduced and induced lanes which corresponded to a native *E. coli* protein.

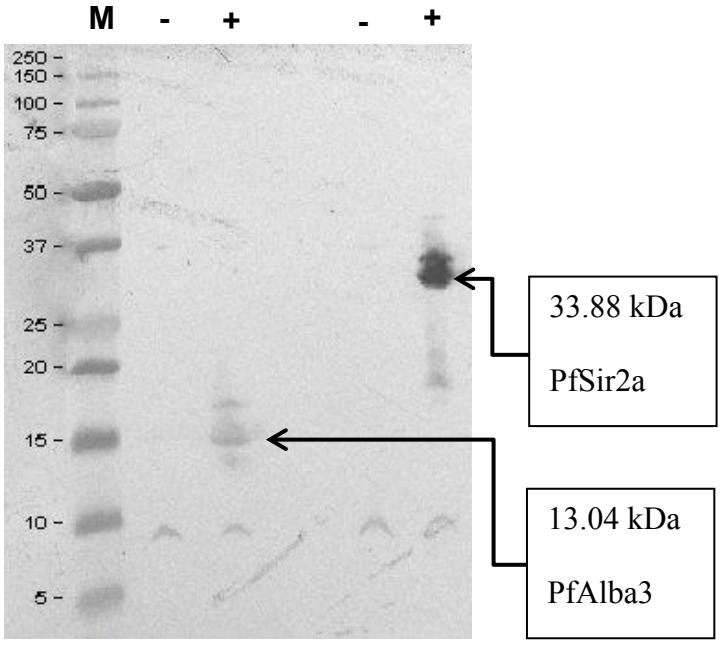


Figure 3.8: Western blot results of PfAlba3 and PfSir2a expression test. M indicates marker lane. Uninduced and induced lanes are marked by (-) and (+) respectively.

3.2.5 Solubility Tests And Optimisation For Large Scale Production

For structural experiments, PfSir2a and PfAlba3 proteins must be recovered in a soluble form and free from contaminants (e.g. host *E. coli* proteins). The solubility of PfSir2a and PfAlba3 was assessed by lysing induced *E. coli* cell cultures and separating the insoluble and soluble fractions. Three different induction temperature conditions were used to assess the impact on solubility, 37˚C, 30˚C and 20˚C with an induction time of 3, 16 and 20 hours

respectively. Uninduced and induced samples were taken and again normalised to an $O.D_{600}$ of 0.6.

For solubility tests, frozen cell pellets of induced cultures were resuspended on ice with 1ml of lysis buffer containing 50mM sodium phosphate, 300mM sodium chloride, 10mM Imidazole, pH 7.5. Once fully resuspended the cells were lysed by the addition of 0.2µm bacterial lysis bio-beads (Scientific Industries, Inc, Cat no. SI-B601) and placed in a cell disruptor genie system (Scientific Industries, Inc, Model No. SI-D256) at 4˚C for 6 minutes. After lysis, 50µl of the recovered suspension was separated into the soluble and insoluble fractions by centrifugation at 14,000rpm for 2 minutes. The soluble fraction containing supernatant was pipetted off and mixed with 3x sample loading buffer in a 3:1 ratio respectively. The insoluble fraction was resuspended in 50µl of 1x sample buffer for direct comparison by SDS-PAGE analysis (see Figure 3.9).

The solubility studies revealed that PfSir2a is generally quite insoluble despite good expression. The optimum condition for expression of soluble protein was determined to be a 3-hour induction at 37˚C. For PfAlba3 an increase in solubility was observed by decreasing the temperature and increasing length of induction to 20˚C and 20 hours respectively. Expression levels remained good and constant across the temperature range tested. As such, large scale production of PfSir2a and PfAlba3 was conducted with these conditions in 1L flask cultures containing 35µg ml^{-1} kanamycin sulphate for further purification.

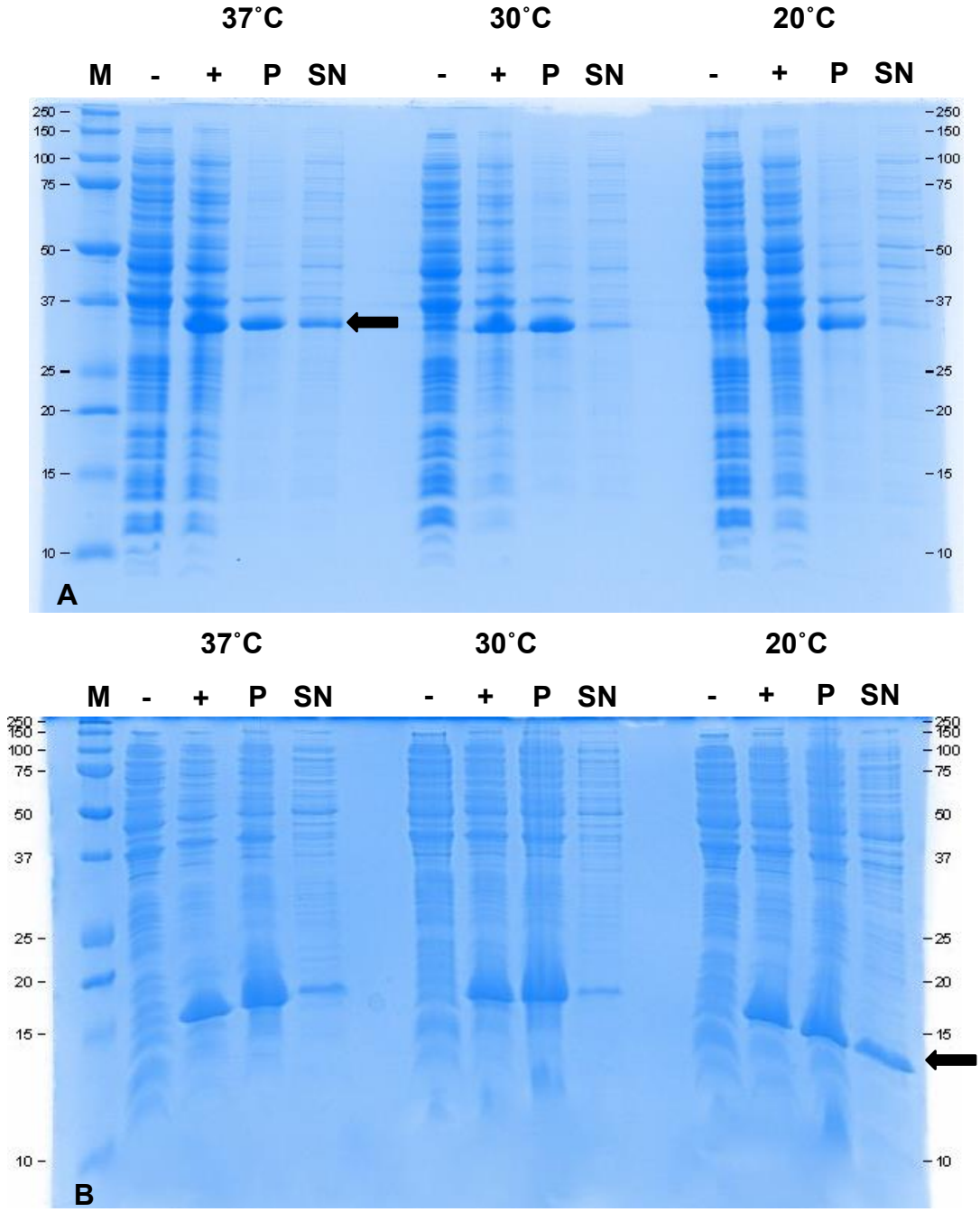


Figure 3.9: Solubility test results for PfSir2a (A) and PfAlba3 (B). (M) Indicates molecular weight marker lane. Induction at each temperature is monitored via uninduced (-) and induced samples (+). Insoluble fraction containing pelleted remains after lysis by bio-bead disruption and the soluble fraction supernatant can be seen in lanes marked P and SN respectively. Conditions with highest levels of recoverable soluble protein for PfSir2a or PfAlba3 indicated by arrow.

3.3 Purification Of PfSir2a And PfAlba3 By Ni-NTA And SEC

This section describes the work carried out to purify the recombinant forms of PfSir2a and PfAlba3 using the expression system a described section 3.2.

3.3.1 Purification Of PfSir2a

The purification of PfSir2a was performed as described in Chapter 2, section 2.1.11 and was monitored by UV_{280nm} absorbance (see Figure 3.10). Throughout each step samples were taken from the supernatant flow through, wash and elution step for purity analysis by SDS-PAGE.

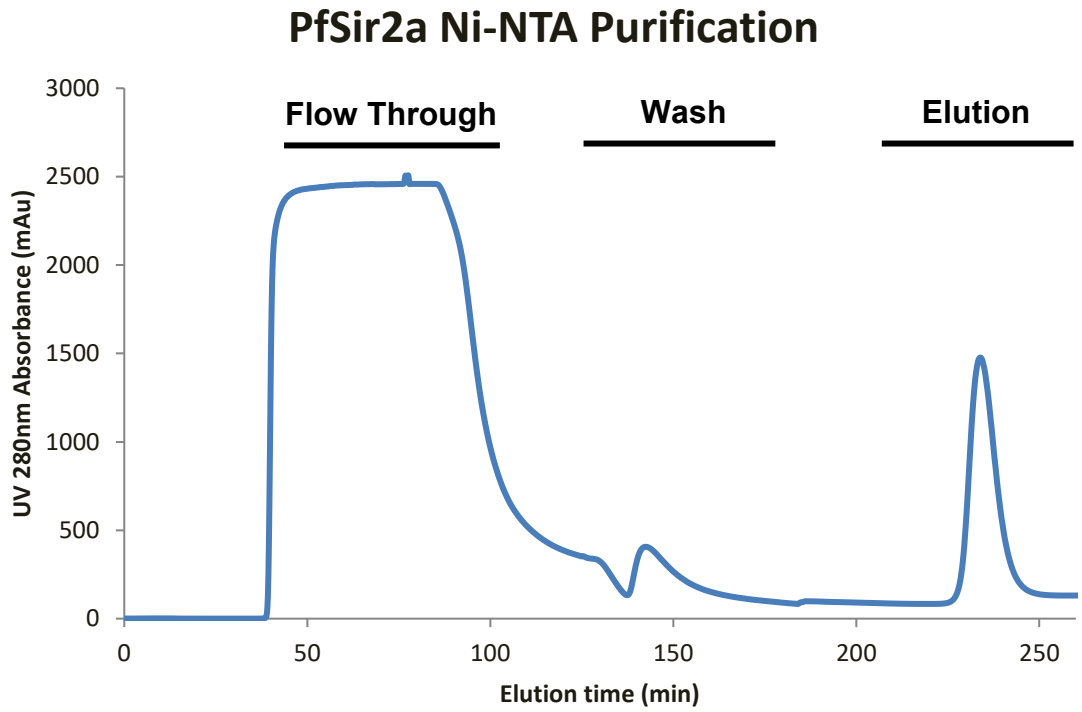


Figure 3.10: UV_{280nm} absorbance monitoring of Ni-NTA purification of PfSir2a using ÄKTAPrime FPLC system.

A flow through sample (FT) was taken to check the protein was efficiently binding to the column and not wasted. Samples from the across the peaks of the wash and elution step (see Figure 3.11) show some PfSir2a being removed with contaminants in the wash step.

The elution samples show a large amount of PfSir2a is retained on the column after the wash until the elution step with some higher molecular weight contaminants.

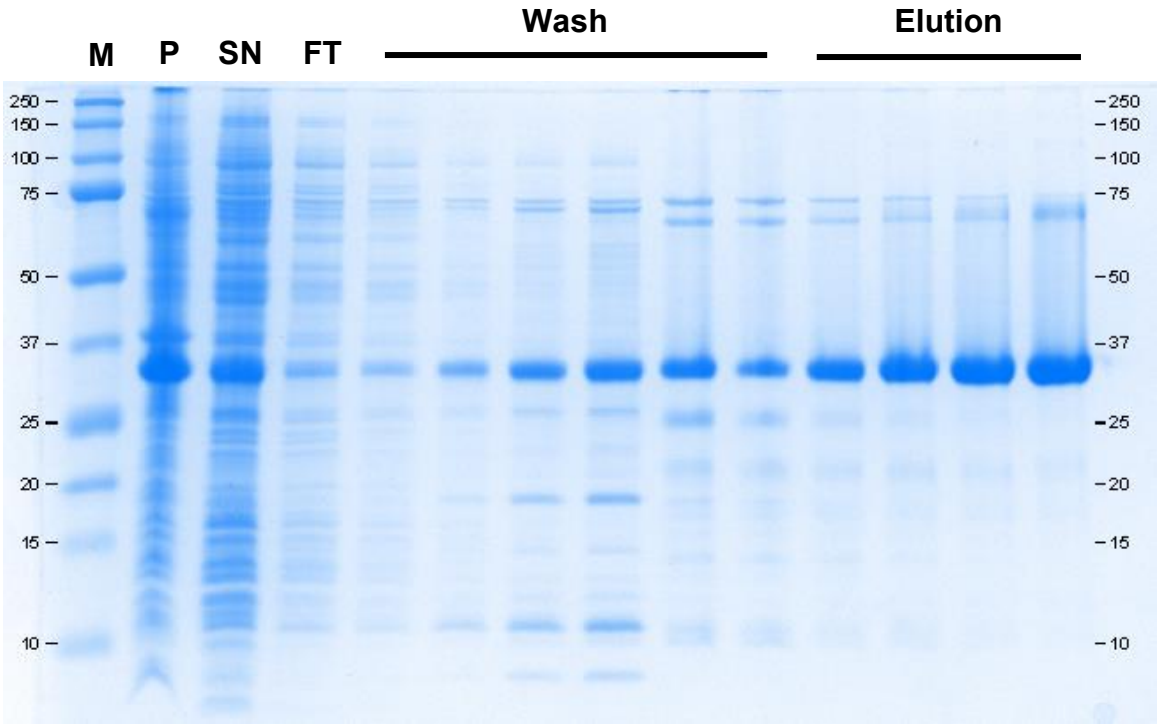


Figure 3.11: 12% SDS-PAGE gel analysis of purification steps. P and SN indicate the insoluble and soluble fractions of the induced protein from the *E.coli* bacterial culture. Flow through (FT), wash and elution step samples are labelled.

The elution fractions were pooled and concentrated using Amicon® Ultra-15 centrifugal filter devices (Merck Millipore (Cat no. UFC901008) before overnight dialysis using 10kDa cut-off dialysis cassettes (Thermo Scientific, Cat no. 66383) into a size exclusion buffer containing 50mM sodium phosphate, 500mM sodium chloride, pH 7.5 at 4˚C. A size exclusion step was performed as described in Chapter 2, section 2.1.14 and samples were taken over the elution volume for purity analysis by SDS-PAGE (see Figure 3.12).

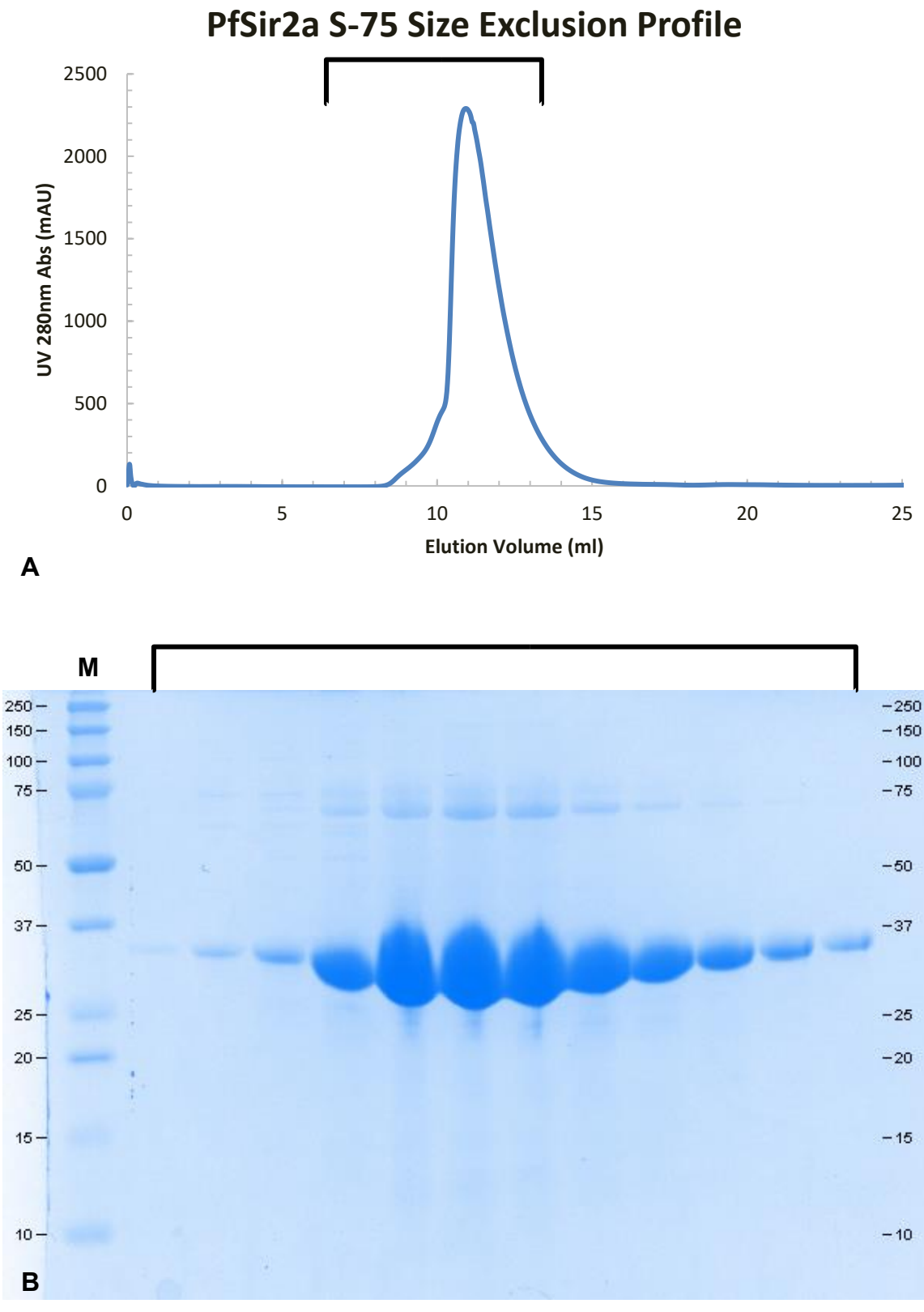


Figure 3.12: (A) SEC migration profile of PfSir2a. (B) Samples collected over elution peak of PfSir2a's migration through the *Superdex*™ S75 10/300 GL analytical column.

Initial analysis reveals PfSir2a as a non-aggregated protein by its SEC elution profile, proteins and aggregates larger than 75 kDa are excluded from the void volume of the *Superdex*™ S75 10/300 GL analytical column (GE Healthcare) at an approximate elution volume of 8.5ml. Analysis using molecular weight gel filtration standards (Bio-Rad Cat no. 151-190) yields an estimated molecular weight of 50-55 kDa which is higher than expected for a monomer of PfSir2a (33.8 kDa) and lower than a dimer (67.6 kDa) (data not shown). This could be suggestive of PfSir2a existing as an elongated monomer or a compact dimer in solution.

Some higher molecular weight contaminants still remained in the fractions and appeared to be migrating with PfSir2a, despite their higher apparent molecular weight as observed by the SDS-PAGE. As such an additional IEX chromatography step was included in the purification process. This step was carried out as described in Chapter 2, section 2.1.13 and the resulting elution profile can be seen in Figure 3.13 A. Samples were taken over the elution volume for purity analysis by SDS-PAGE (see Figure 3.13 B).

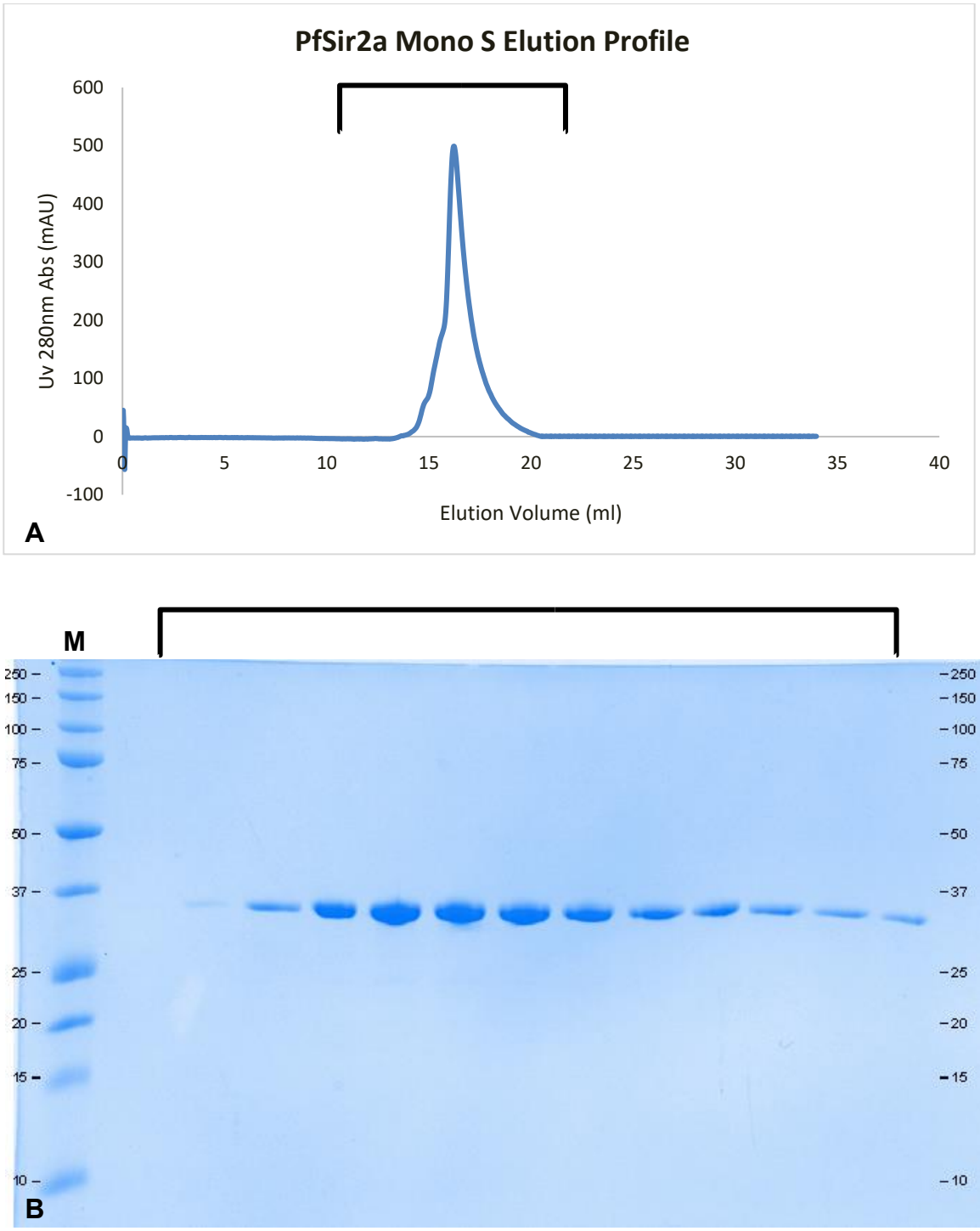


Figure 3.13: (A) Mono S elution profile of PfSir2a. (B) PfSir2a fractions recovered after Mono S IEX purification step.

3.3.2 Purification Of PfAlba3

The purification of PfAlba3 was carried out as described previously in Chapter 2, section 2.1.12. The elution profile was monitored by UV_{280nm} absorbance (see Figure 3.14). Throughout each step samples were taken from the supernatant flow through, wash and elution step for purity analysis by SDS-PAGE.

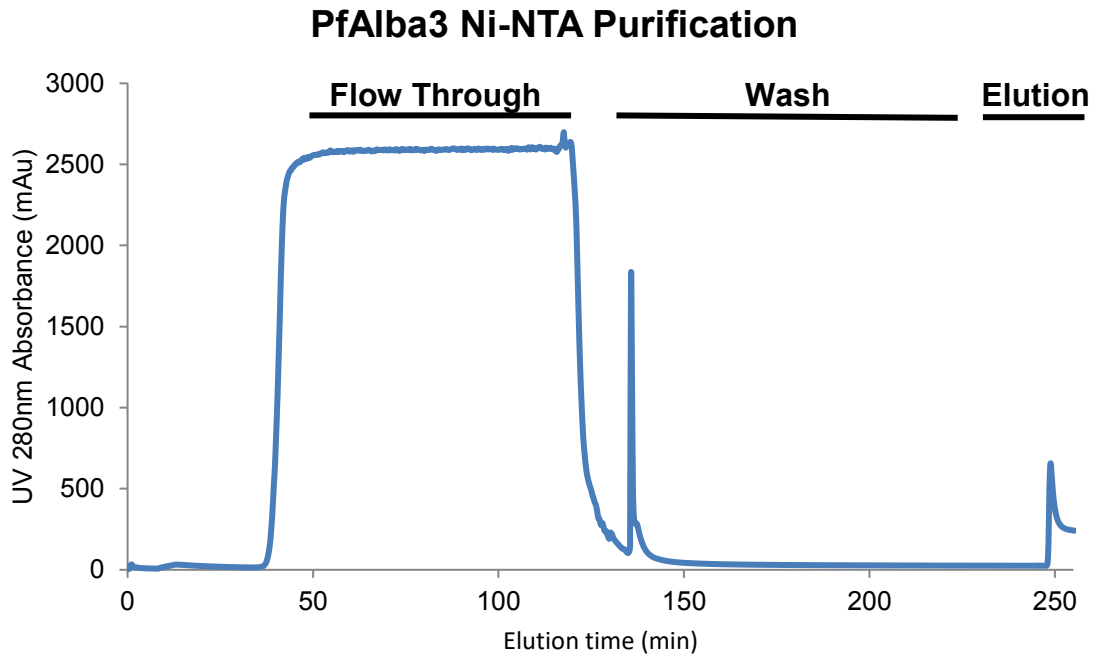


Figure 3.14: UV_{280nm} absorbance monitoring of Ni-NTA purification of PfAlba3 using ÄKTAPrime FPLC system.

In Figure 3.15 we see that there is a signifciant amount of PfAlba3 that is recovered from the lysis protocol and present in the elution fractions.

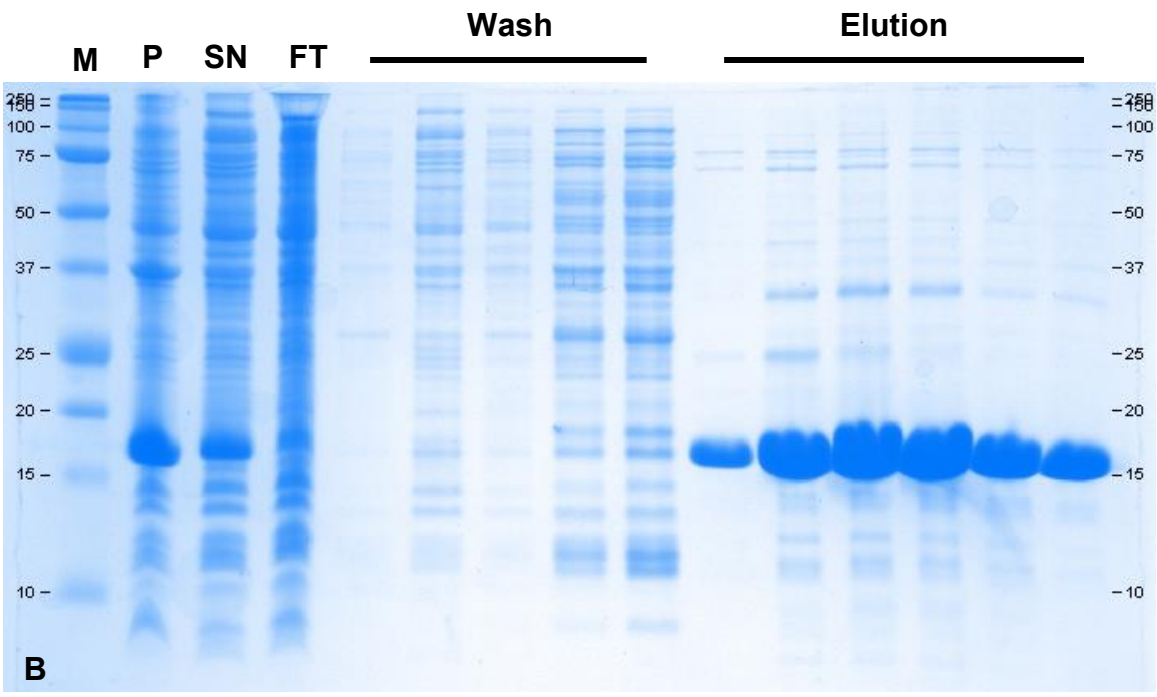
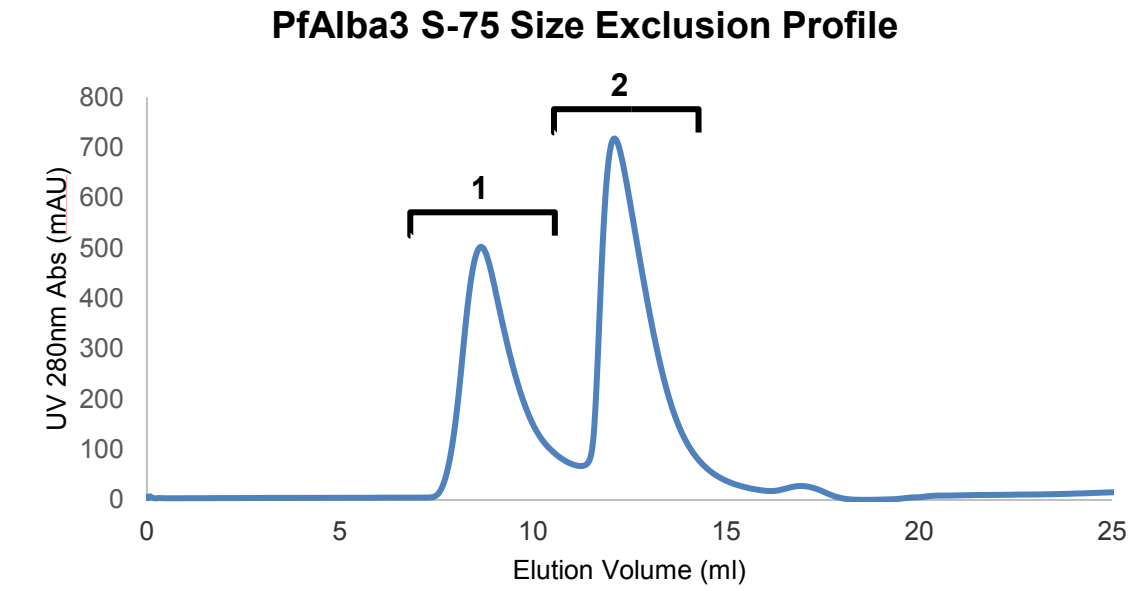
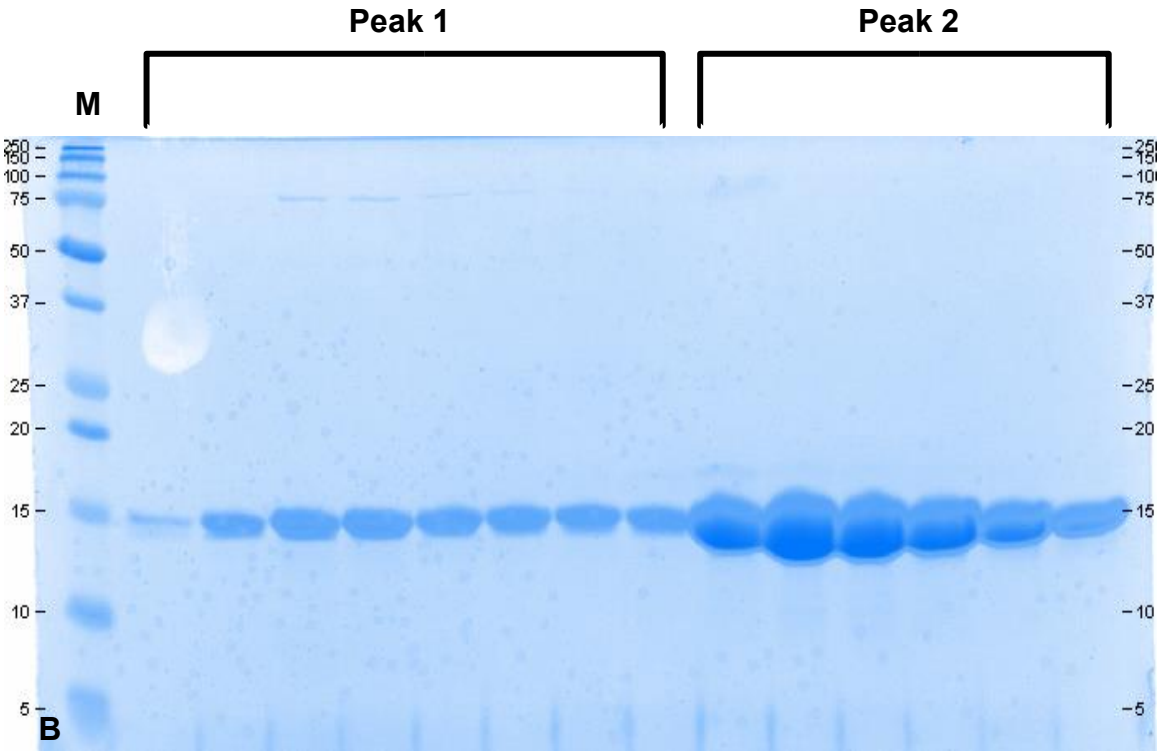


Figure 3.15: 16.5% SDS-PAGE gel analysis of purification steps. P and SN indicate the insoluble and soluble fractions of the induced protein from the *E.coli* bacterial culture. Flow through (FT), Wash and elution step samples are labelled.

Further purification of PfAlba3 was carried out by SEC as described previously in Chapter 2, section 2.1.14. Samples were taken over the elution volume for purity analysis by SDS-PAGE (see Figure 3.16).



A



B

Figure 3.16: (A) S-75 SEC migration profile of PfAlba3. (B) Samples collected over elution peak of PfAlba3's migration through the *Superdex*™ S75 10/300 GL analytical column.

Initial analysis shows that PfAlba3 has oligomerisation/aggregation tendencies as can be seen from the two peaks observed in the SEC elution profile. This is confirmed by SDS-PAGE analysis as PfAlba3 is clearly present in both elution peaks which corresponded to different molecular weights. An estimated molecular weight using gel filtration standards was calculated for PfAlba3 (13 kDa) to be 25kDa. This indicated the protein could be present as a dimer form in solution.

A second SEC experiment was conducted using a *Superdex*™ S200 10/300 GL analytical column that has a void volume which excludes proteins larger than 200kDa. This was to identify if the protein was in an additional oligomeric state or aggregated in the first peak. Sample preparation and injection was carried out as described previously (Chapter 2, section 2.1.14) and the SEC elution profile can be seen in Figure 3.17.

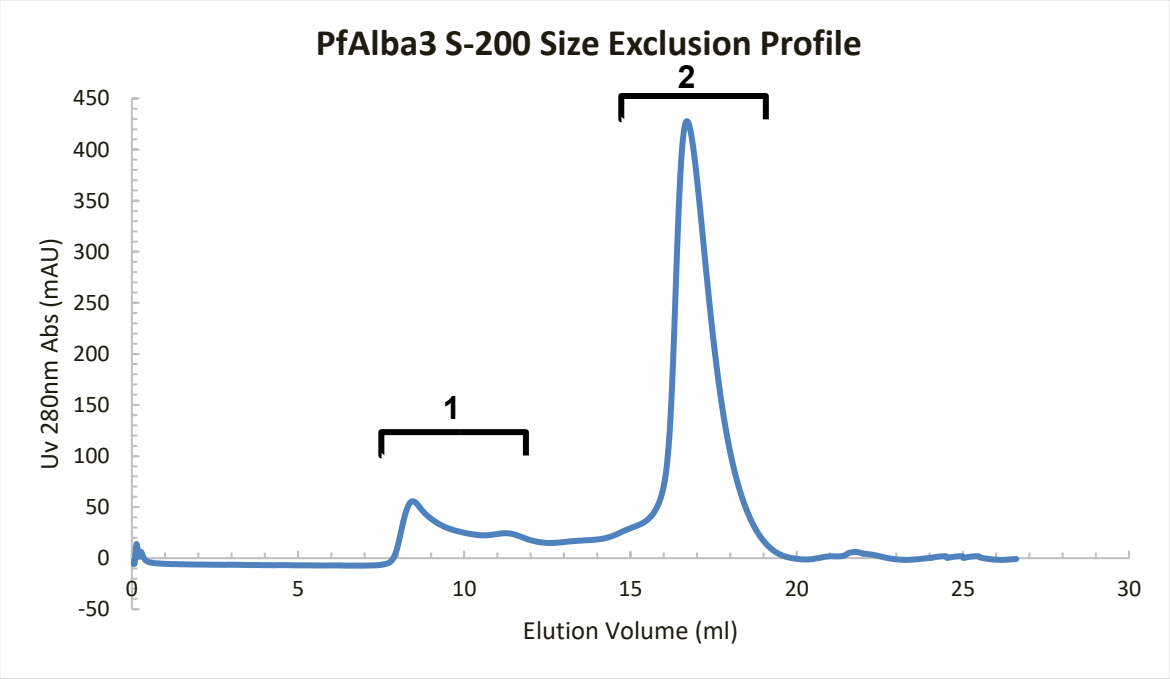


Figure 3.17: S-200 SEC migration profile of PfAlba3. Aggretated PfAlba3 elutes in peak 1 while non aggregated protein elutes in peak 2.

The fractions from both peak 1 and peak 2 from the S-200 SEC profile were analysed by SDS-PAGE to confirm that it is high molecular weight aggregates of PfAlba3 contained within elution peak 1 (see Figure 3.18).

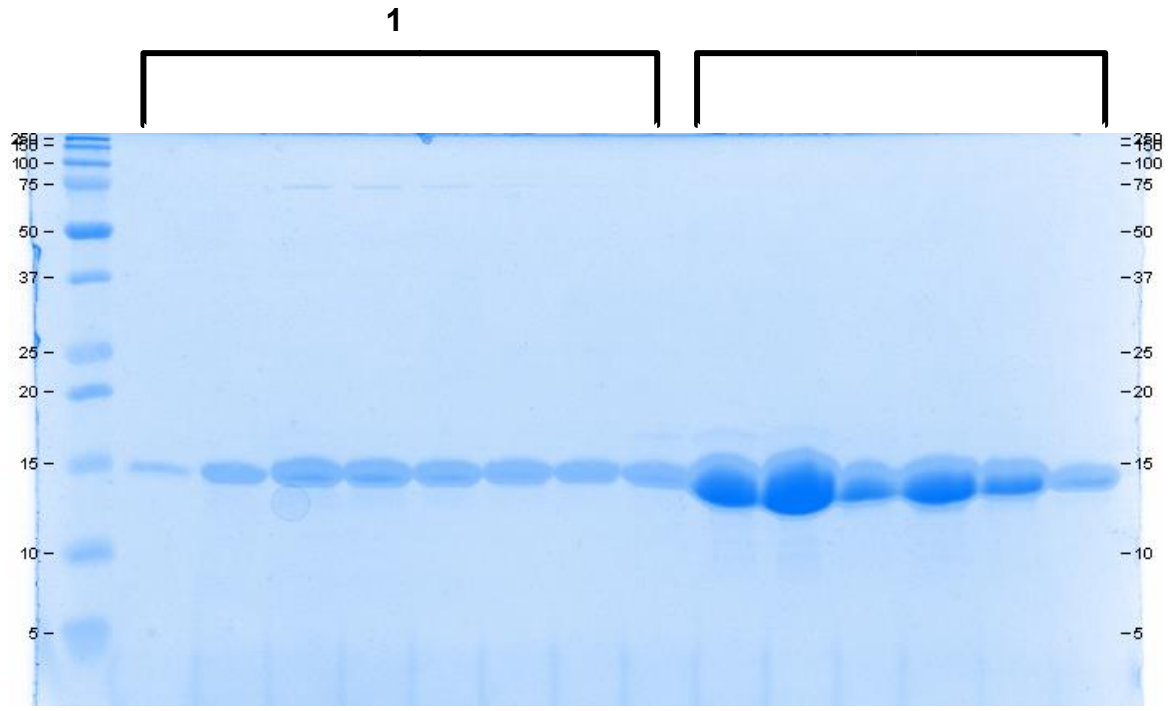
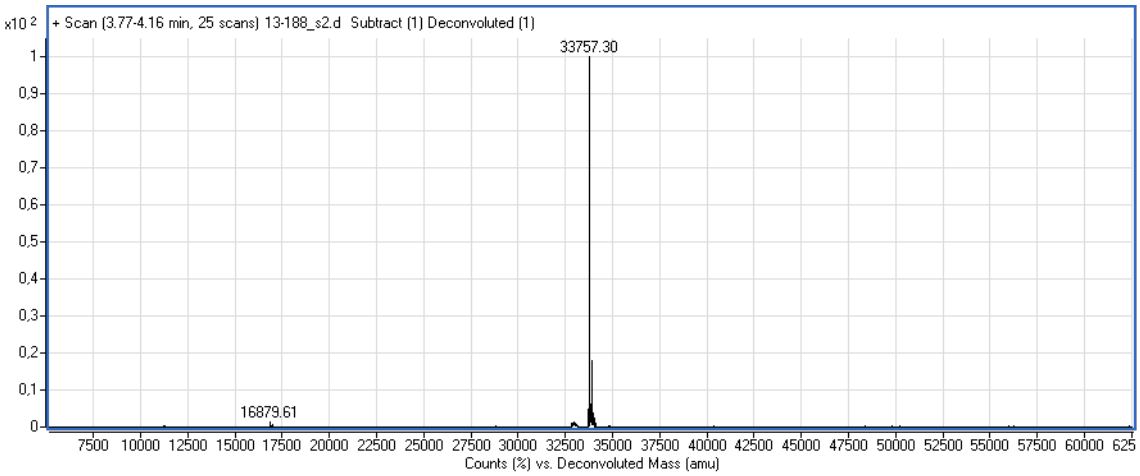


Figure 3.18: Samples collected over elution profile of PfAlba3's migration through the *Superdex*™ S200 10/300 GL analytical column.

3.4 TOF-MS Results

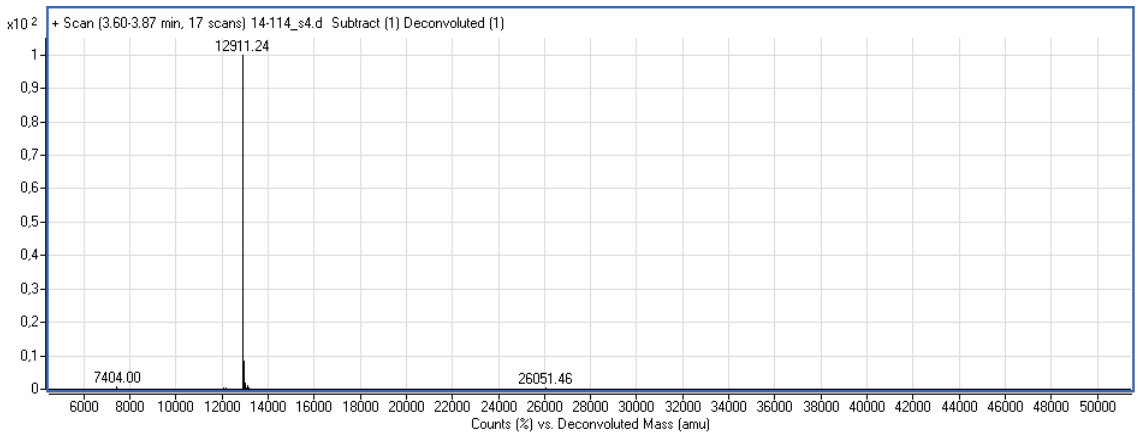
PfSir2a and PfAlba3 were confirmed to have been produced in full length recombinant forms by TOF-MS analysis. The deconvoluted results from the analysis can be seen in Figure 3.19.

A



Sample	Expected Mass (Da)	Observed Mass (Da)	Mass Deviation (Da)	Mass Accuracy (ppm)
PfSir2a	33888.36	33757.30	0.15	4

B



Sample	Expected Mass (Da)	Observed Mass (Da)	Mass Deviation (Da)	Mass Accuracy (ppm)
PfAlba3	13041.91	12911.24	0.53	41

Figure 3.19: Deconvoluted TOF-MS spectrum results for PfSir2a and PfAlba3 in panels A and B respectively.

Results for PfSir2a and PfAlba3 show a lower observed mass than the expected mass as calculated by the amino acid sequence of the recombinant proteins. In both cases this is attributed to the loss of the N-terminal methionine residue (MW = 141.21 Da) from both protein sequences, believed to be caused by the action of methionine aminopeptidase found in *E. coli* (Xiao et al., 2010).

3.5 TSA Buffer Optimisation

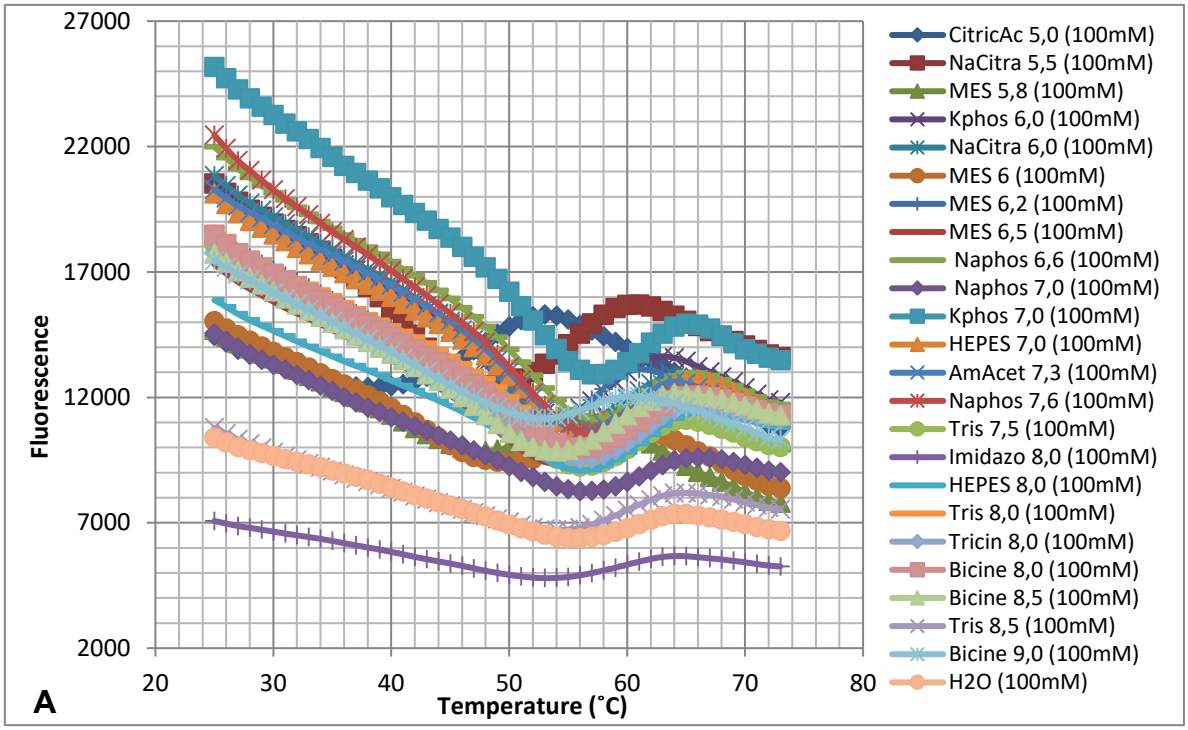
Thermal shift assay experiments were performed to evaluate buffer conditions (buffering agent, pH, salt concentration etc) via a screening method to optimise buffers for protein purification, storage and further experiments. Initial screens were performed in 96-well low-profile, non-skirted, optical reading compatible white plates (Stratagene, Cat no. 410088). Wells were prepared by the additions of 4µl of 5x buffer stocks (final concentration 100mM), 1µl of SYPRO® Orange (Invitrogen Cat no. S6651) 100x stock (final concentration 5x) and 2µl of protein sample stock at 100µM (final concentration 10µM).

Two salt conditions were tried, 100mM and 500mM sodium chloride, by adding 2µl or 5µl of 2M initial stock concentration respectively. The remaining volume for each well was made up to a final volume of 20µl with ddH_2O. Once all wells had been filled with the reaction mix they were sealed with an optical compatible seal (Bio-Rad Cat no. MSB-1001) and placed in a Mx-3005P Q-PCR instrument (Stratagene) and heated from 25˚C to 75˚C at a rate of 1˚C min^{-1}.

3.5.1 Thermofluor Analysis Of PfSir2a

Results for the thermofluor experiments performed on PfSir2a samples can be seen in Figure 3.20. For both salt concentrations tested the analysis is difficult to interpret due to the high initial fluorescence observed at low temperatures. This means that the SYPRO® dye is bound to hydrophobic regions present on the protein surface. Another reason for this could be that the protein is unfolded or aggregated but this is considered unlikely as the previous SEC results indicate the protein is not aggregated, due to being taken immediately after a SEC step was performed. Also, the thermofluor results still show a small but characteristic change in fluorescence due to unfolding with increase in temperature.

PfSir2a appears to be most stable in a buffer pH range of 6.6 – 8.0. The protein was found to be most stable in HEPES buffer, pH 7.0, 100mM sodium chloride with a T_m of 61.5˚C. The protein appears to be generally stable in most buffer conditions tested (e.g sodium phosphate, potassium phosphate, Tris, Bicine, Tricine). There also appeared to be no significant effect of salt concentration on the stability of the protein (significant effect when ΔT_m of greater than 2˚C occurs).



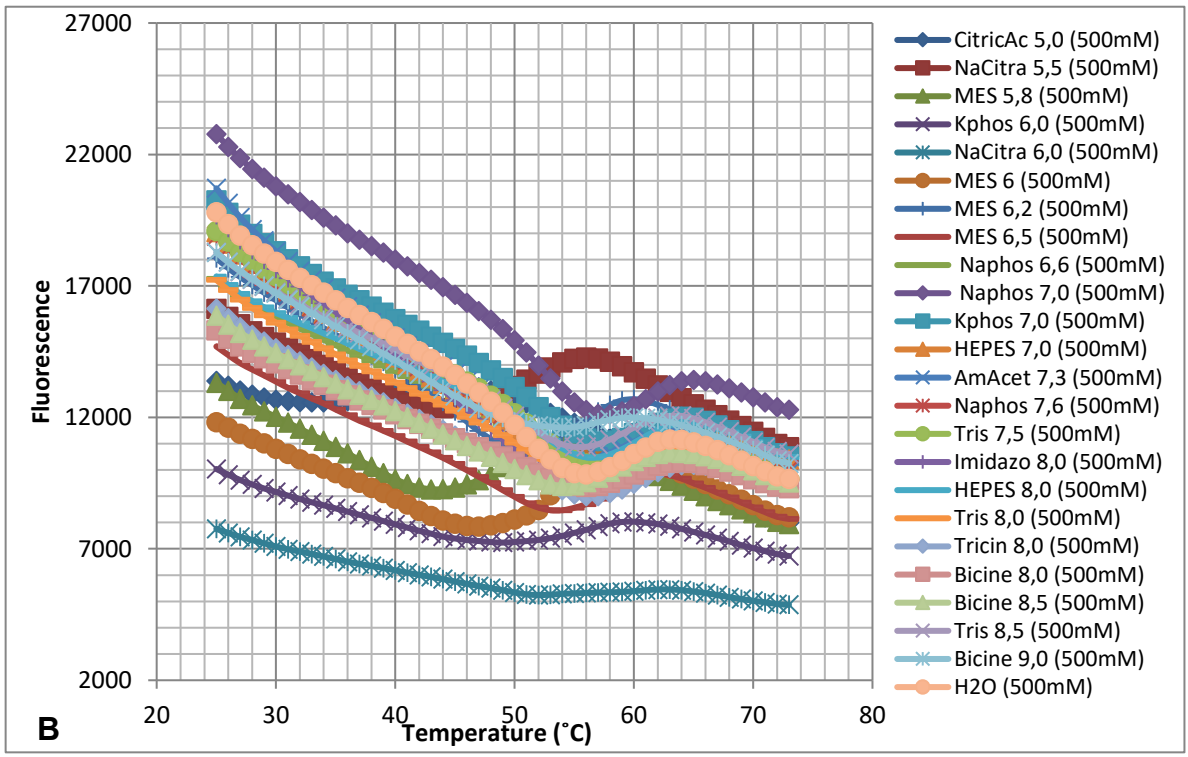
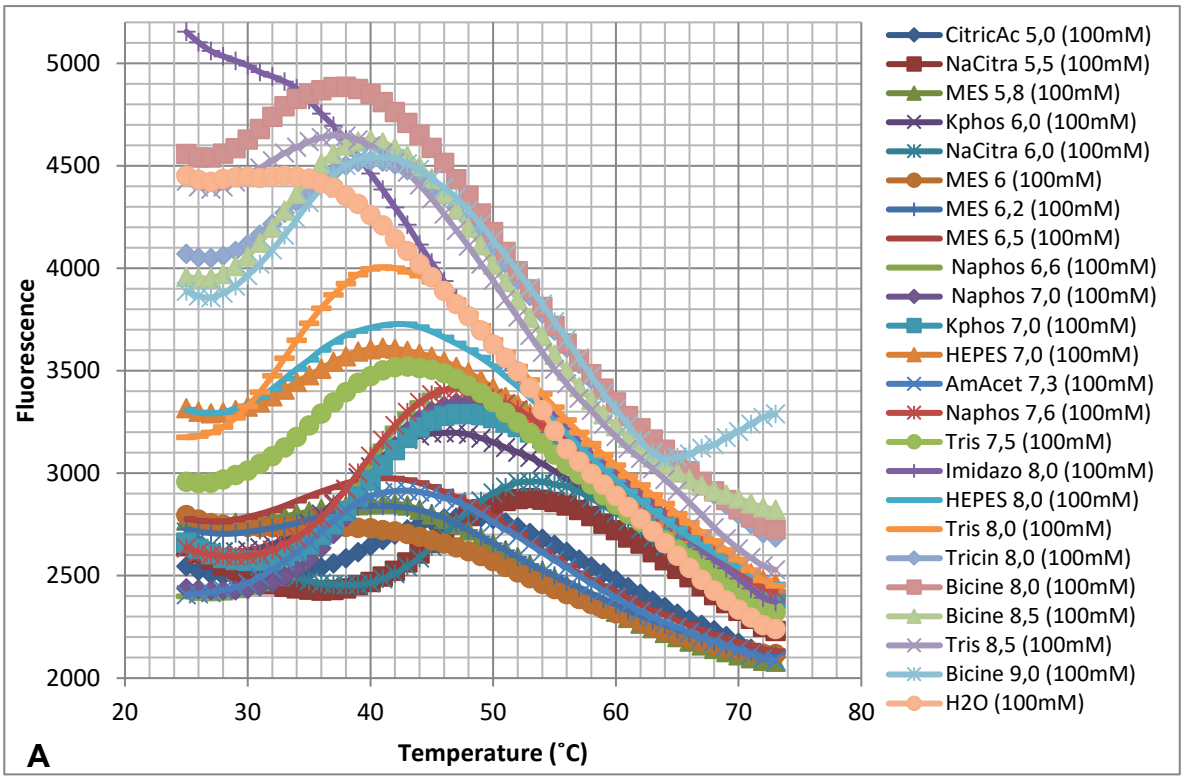


Figure 3.20: Thermofluor results for PfSir2a. Panel A shows the experiments conducted with 100mM sodium chloride concentration and Panel B shows those conducted with 500mM.

3.5.1 Thermofluor Analysis Of PfAlba3

Results for the thermofluor experiments performed on PfAlba3 can be seen in Figure 3.21. A distinct difference is observed between the two salt concentreations tested, those conducted at 100mM sodium chloride concentration again show high initial fluoresence levels indicating that either the SYPRO® dye is binding to several exposed hydrophobic patches, or that the protein is unfolded or aggregated. This is not observed in the 500mM sodium chloride concentreation experiments which show reduced intial fluorescence levels compared to that of the 100mM results. This indicates that there is a salt concentration dependence for PfAlba3 where it is more stable in higher salt concentrations.

PfAlba3 appears to be most stable in a buffer pH range of 5.5-7.0. The optimal condition for the protein was found to be in potassium phosphate, pH 6.0, 500mM sodium chloride with a T_m of 49.5˚C. The protein appeared to be less stable at higher pH conditions as the fluoresence change due to heating produced flat lines with no easily identified thermal shift. Significant changes in the T_m of PfAlba3 were observed between the two salt concentrations in each buffer tested. A dramatic decrease in T_m of 6˚C or more was noticed in almost all conditions tested suggesting that PfAlba3 becomes unstable and begins to aggregated in reduced salt environments.



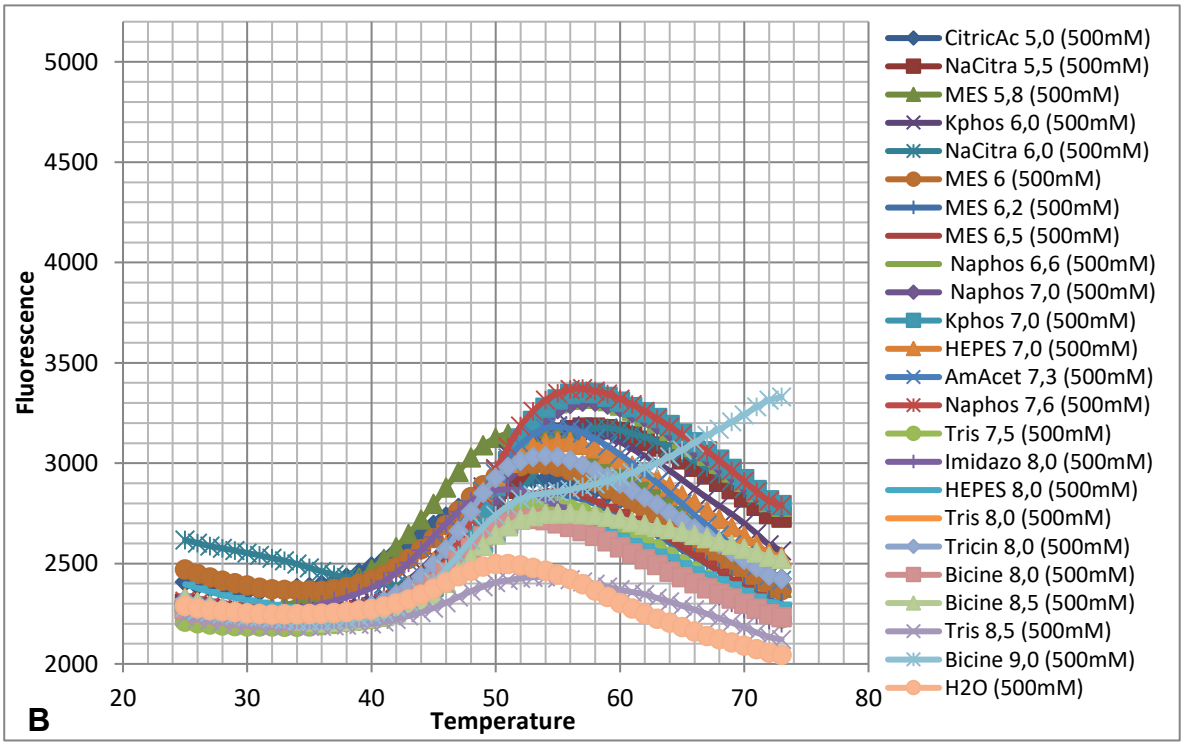


Figure 3.21: Thermofluor results for PfAlba3. Panel A shows the experiments conducted with 100mM sodium chloride concentration and Panel B shows those conducted with 500mM.

3.6 Light scattering results For PfSir2a And PfAlba3

This section details the results obtained for DLS experiments performed using purified recombinant PfSir2a and PfAlba3 as described in Chapter 2, section 2.2.3.

3.6.1 DLS Results Of PfSir2a

Initial experiments to measure the hydrodynamic raidus of PfSir2a were conducted in a buffer containing 50mM HEPES, 100mM sodium chloride, pH 7.5 at a sample concentration of 2mg ml^{-1}. A single broad peak was observed with a calculated R_H of 8.74 ± 5.45nm and a polydispersity of 49.9% (see Figure 3.22). This high polydispersity indicates multiple species are present within the sample and may be a monomer/dimer mixture; other oligomeric states may also exist. This indicates the sample would not be suitable for SAXS/SANS experiments in the current buffer choice.

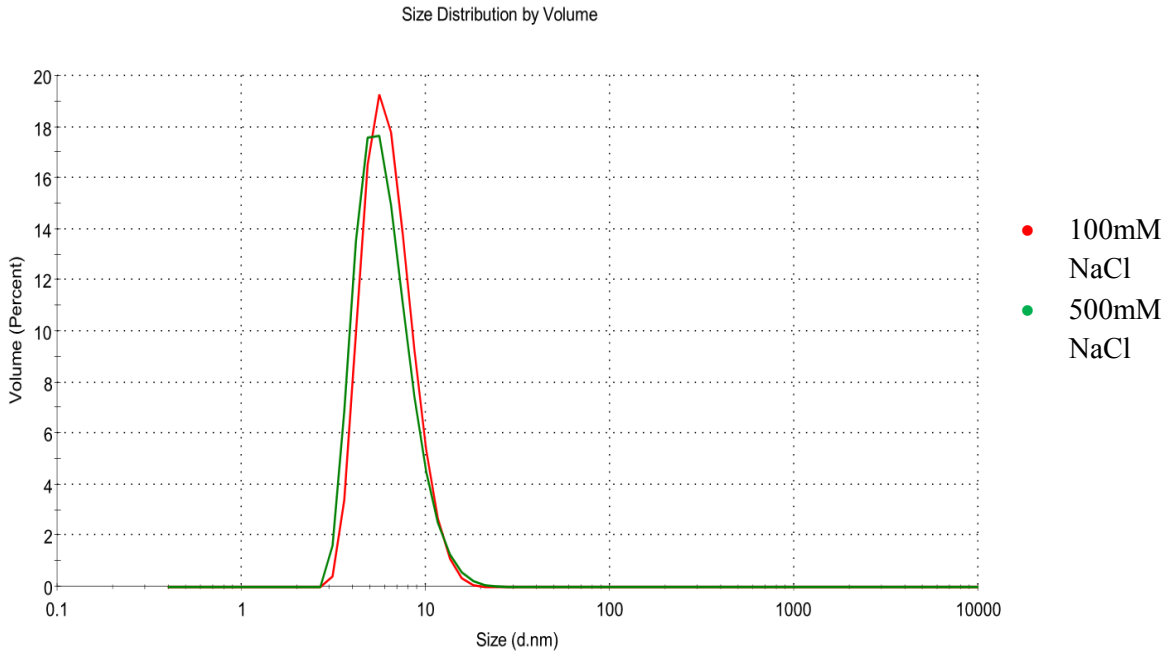


Figure 3.22: DLS results for PfSir2a in 50mM HEPES buffer with 100mM or 500mM NaCl.

The salt concentration was then increased to 500mM sodium chloride with the sample concentration kept at 2 mg ml^{-1} and repeated (Figure 3.22). These results yielded a R_H of 8.72 ± 2.67nm and a polydispersity of 31.1%. This indicates that the higher salt concentration appears to have an effect on the overall polydispersity of the species within the sample but it still remains above the 20% threshold for a monodisperse solution.

Additional buffer conditions were screened to try to improve the overall monodispersity of the sample. Buffer solutions containing 50mM Tris or 50mM sodium phosphate at pH 7.5 and pH 7.0 were tested with 100mM or 500mM sodium chloride concentration (see Figure 3.23). For samples measured in Tris buffer the observed R_H was calculated to be 8.5± 2.72nm and 8.9 ±2.02nm in at a salt concentration of 100mM and 500mM sodium chloride respectively. The polydispersity value for each were determined to be 28.2% for 100mM and 21.2% for 500mM again indicating there is a salt effect causing a reduction in polydispersity.

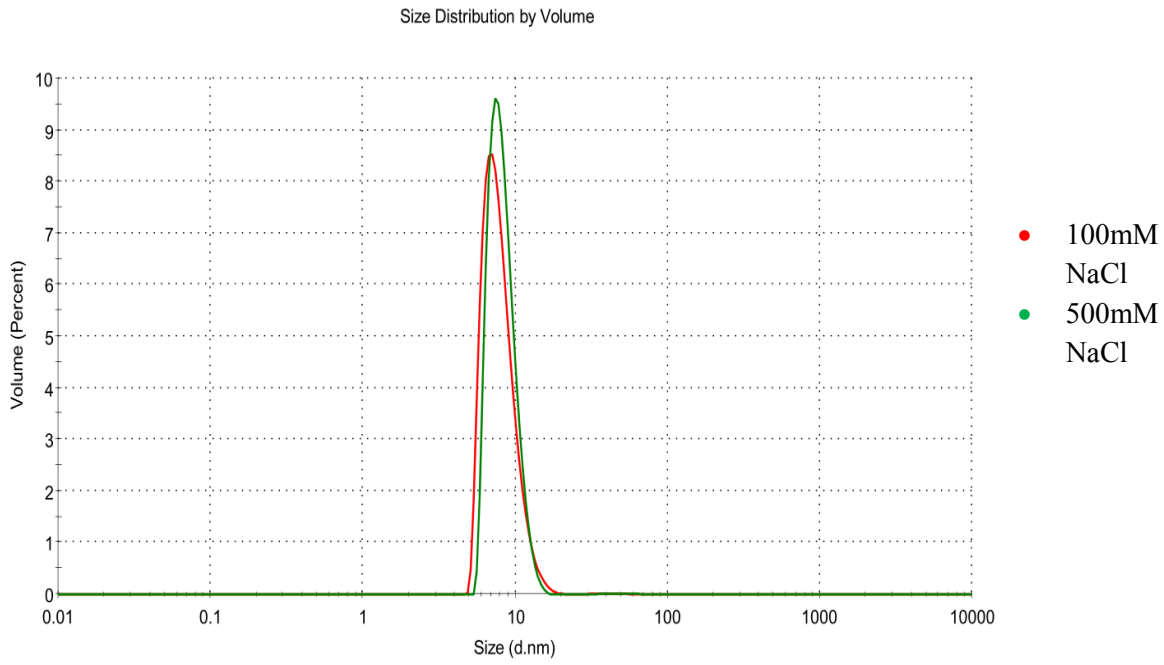


Figure 3.23: DLS results for PfSir2a in 50mM Tris buffer with 100mM or 500mM NaCl.

Samples measured in sodium phosphate buffer (see Figure 3.24) were found to have a R_H of 8.12 ±2.2nm and 9.33 ± 1.36nm in 100mM and 500mM sodium chloride. The polydispersity was found to be 26% for 100mM and 14.5% for 500mM which is acceptable for SAXS/SANS experimentation.

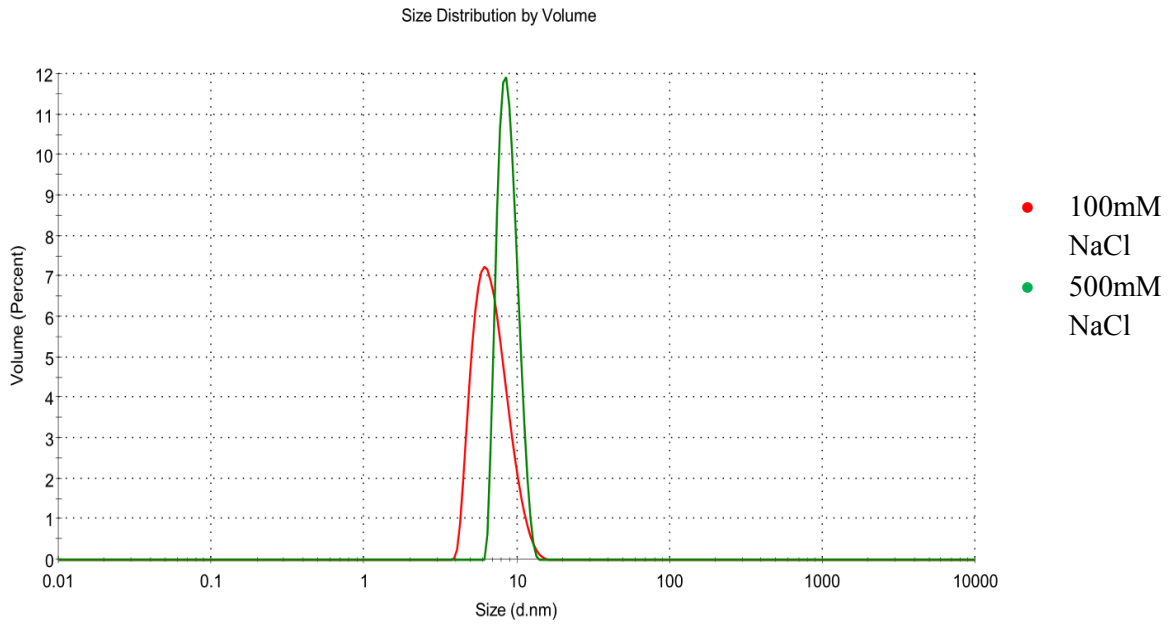


Figure 3.24: DLS results for PfSir2a in 50mM sodium phosphate buffer with 100mM or 500mM NaCl.

3.6.2 DLS Results Of PfAlba3

DLS experiments for PfAlba3 were conducted in buffers containing 500mM sodium chloride as the TSA results indicated it would be prone to aggregation, or become unstable if placed in a low salt buffer environment. Initial measurements were made in buffers containing either 50mM sodium phosphate or 50mM potassium phosphate at pH 7.0 at a protein concentration of 1 mg ml^{-1} (see Figure 3.25). For the sample prepared in potassium phosphate buffer a very large R_H was observed, 58.77 ± 71.22nm with 79.5% polydispersity. This indicated the sample had begun to aggregate after the SEC step (within 1 hour) and visible white precipitate was observed after retrieving the samples confirming

protein aggregation/precipitation. A relatively lower R_H of 7.53 ± 4.31 nm was observed in the sample prepared in sodium phosphate buffer with a polydispersity of 45.4%. This result indicated that PfAlba3 was potentially more stable in sodium phosphate buffer condtions but existed as a mixture of oligomeric species, or was beginning to aggregate at a slower rate compared to that of the sample prepared in phosphate buffer.

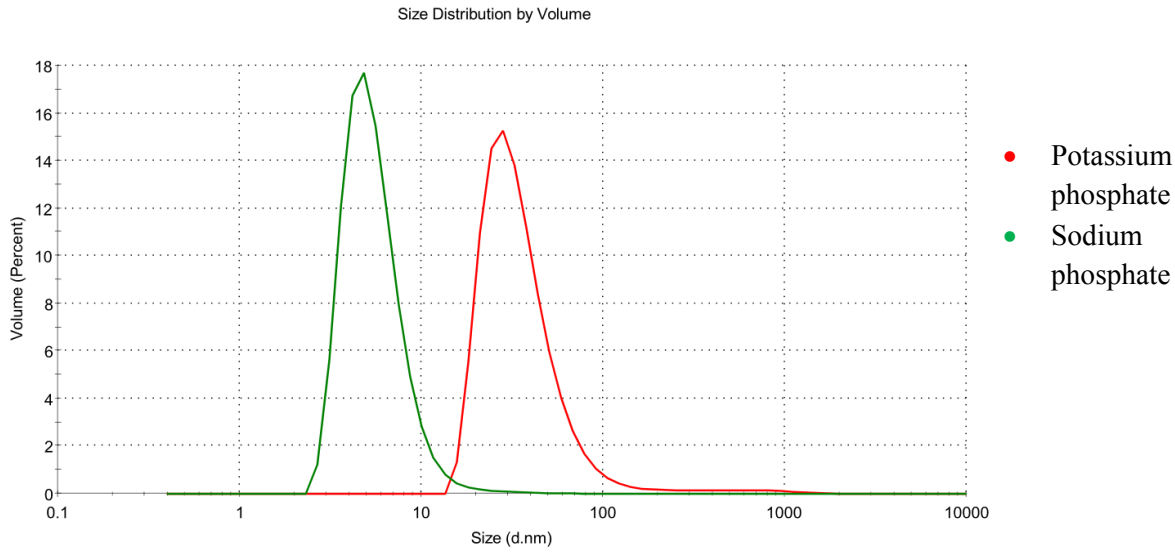


Figure 3.25: DLS results for PfAlba3 in 50mM potassium phosphate or 50mM sodium at 21°C, each containing 500mM NaCl.

Based on the TSA results the samples were remeasured again at 4˚C due to the generally low T_m of PfAlba3. In these experiments a R_H of 6.72 ± 1.87nm was observed for samples in potssium phosphate buffer with a polydispersity of 22.5%. For samples in sodium phosphate buffer a R_H of 7.07 ± 1.15nm with a polydispersity of 15.9% (see Figure 3.26). This represents a marked improvement in the quality of both samples compared to measurements made at 21˚C. The results for the samples in potassium phosphate show a slight tailing of the peak towards particles of larger R_H along with a slightly high polydispersity percentage. This may be still be indicative of aggregates forming or a oligomeric mixture being present.

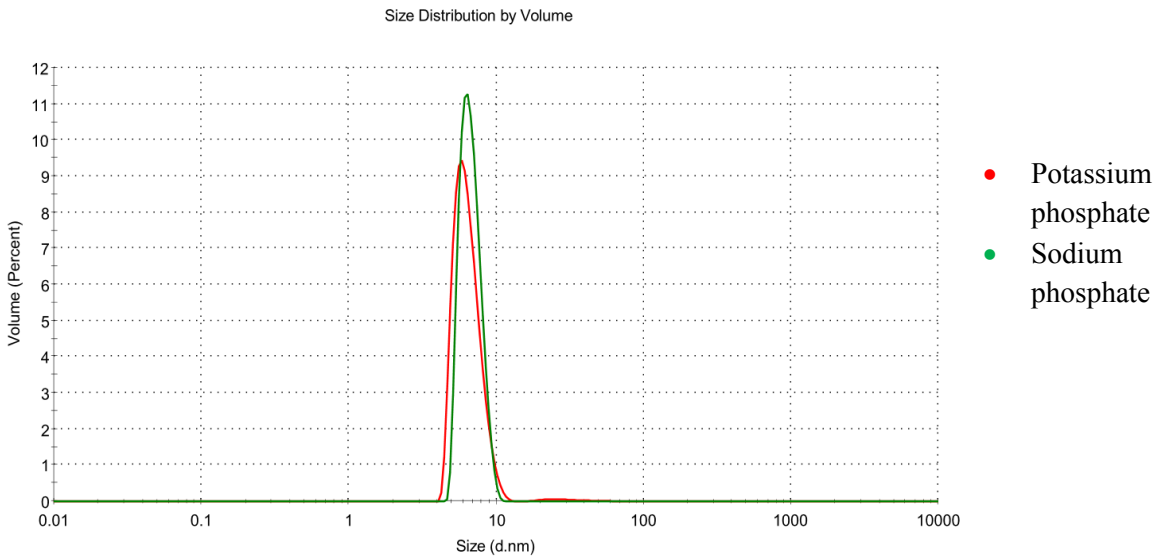


Figure 3.26: DLS results for PfAlba3 in 50mM potassium phosphate or 50mM sodium at 4°C, each containing 500mM NaCl.

Despite this increase in overall sample quality, further measurements revealed that PfAlba3 is still aggregation prone even when purified and stored at 4˚C prior to experiments. This was also observed as a visible precipitate forming in PfAlba3 samples a few hours after purification. The quality of PfAlba3 over time was assessed by making repeated DLS measurements on a sample at set time points after purification, the results of which can be seen in Figure 3.27.

As Figure 3.27 shows, PfAlba3 shifts into a fully aggregated state within the space of a only a few hours. This occurs initially as soluble aggregates of PfAlba3 formed in solution which eventually form into larger aggregates and precipitates. This is observed as a white precipitate within the sample solution. This has several consequences for further biochemical and biophysical characterisation methods which may be affected by the presence of aggregate molecules. It is also a major issue for SAXS/SANS solution experiments which rely on samples being a monodisperse solution of particles, free from aggregates for accurate interpretation and modeling.

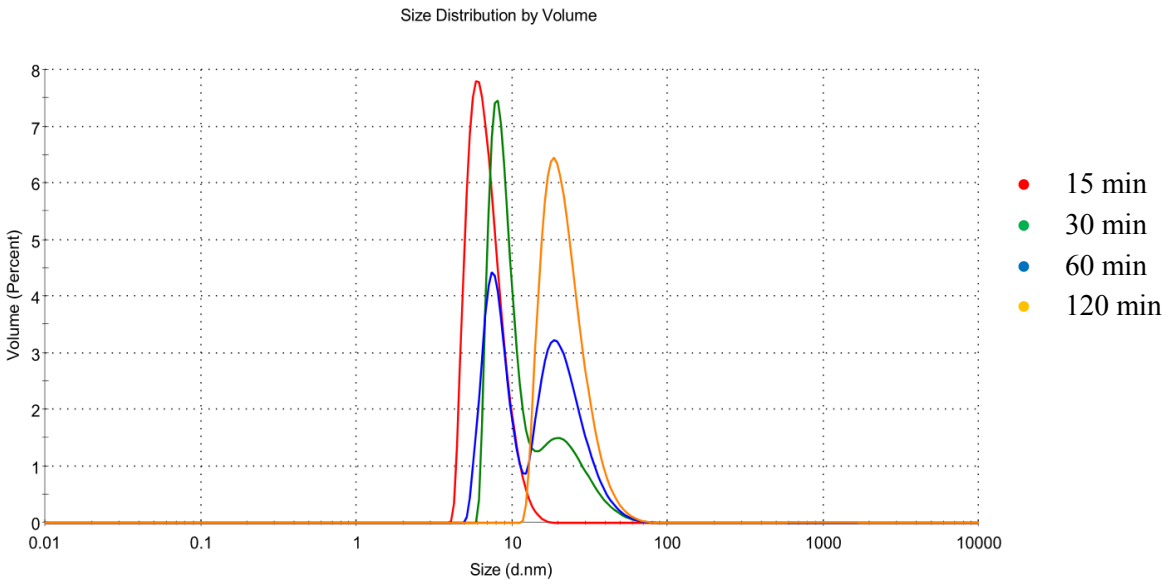


Figure 3.27: DLS monitoring of Alba3 aggregation over 120minute time course. After 15 minutes (red) PfAlba3 single peak becomes more polydisperse. At 30 minutes (green) and 60 minutes (bue) two separate peaks can be observed as the shift towards aggregation occurs. At 120 minutes (yellow) PfAlba3 becomes fully aggregated.

3.7 SEC-MALS RI Analysis Of PfSir2a And PfAlba3

This section details the results obtained during SEC-MALS RI investigations of PfSir2a and PfAlba3 as described in Chapter 2, section 2.2.4.

3.7.1 SEC-MALS RI Results For PfSir2a

Normalised SEC-MALS RI resuslt for PfSir2a can be seen in Figure 3.28. The results show a single large peak in UV 280nm absorbance, Rayleigh ratio and change in the refractive index of the sample (dRI) at an elution volume of 16.7ml. A small shoulder to this can be observed preceding the main peak at an elution volume of 14.6ml. This is more distinguished in a plot containing the UV and Rayleigh ratio curves (see Figure 3.28 B). A large change in the dRI is always observed at the end of the column due to effects of buffer exchange.

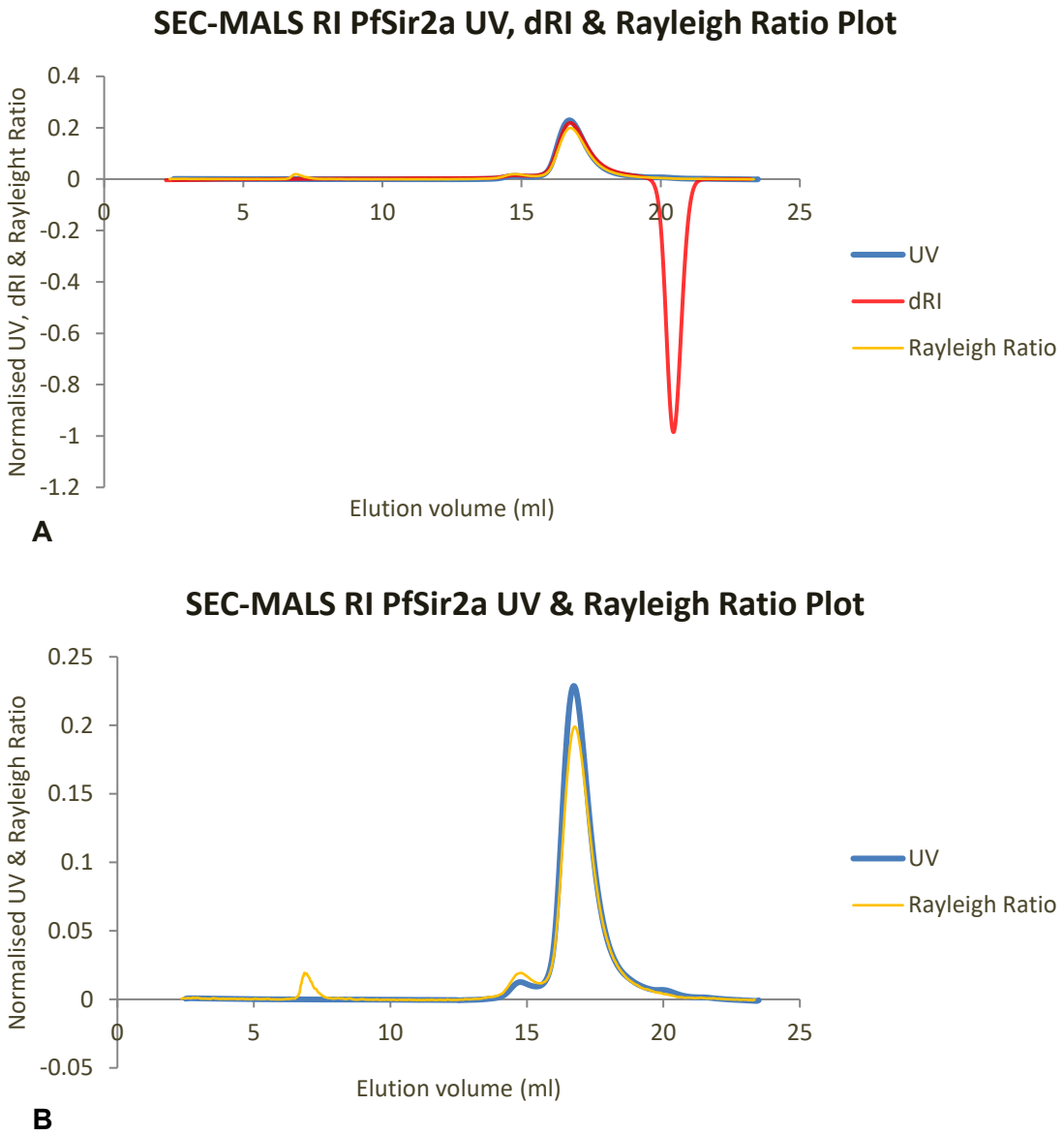


Figure 3.28: (A) Elution profile of PfSir2a SEC-MALS RI experiment showing observed UV_{280nm} absorbance, Rayleigh ratio and dRI. The large change in the dRI at the end of the column is due to buffer exchange at the end of the column. (B) Plot of UV_{280nm} absorbance (for protein elution reference) and Rayleigh ratio highlighting the single major peak at 16.6ml with a small preceding peak at a elution volume of 14.6ml.

The average molar mass of the particles contained within the main peak was found to correspond to a molecular weight (MW) of 34.6 kDa (± 1.937%) with a polydispersity index of 1.000 (± 2.740%) (see Figure 3.29). This indicates a monodisperse sample with a MW close to the expected value of 33.88 kDa for a PfSir2a monomer. The average molar mass of the particles within the shoulder peak was found to correspond to a MW of 49.66 kDa (±2.341%) with a polydispersity index of 1.000 (± 3.539%). This MW does not exactly correspond to a dimer sized particles of PfSir2a which we would expect to be 67.76 kDa and so could be a potential contaminant. The mass fractions for the main and shoulder peak were found to be 97.4% and 2.6% respectively.

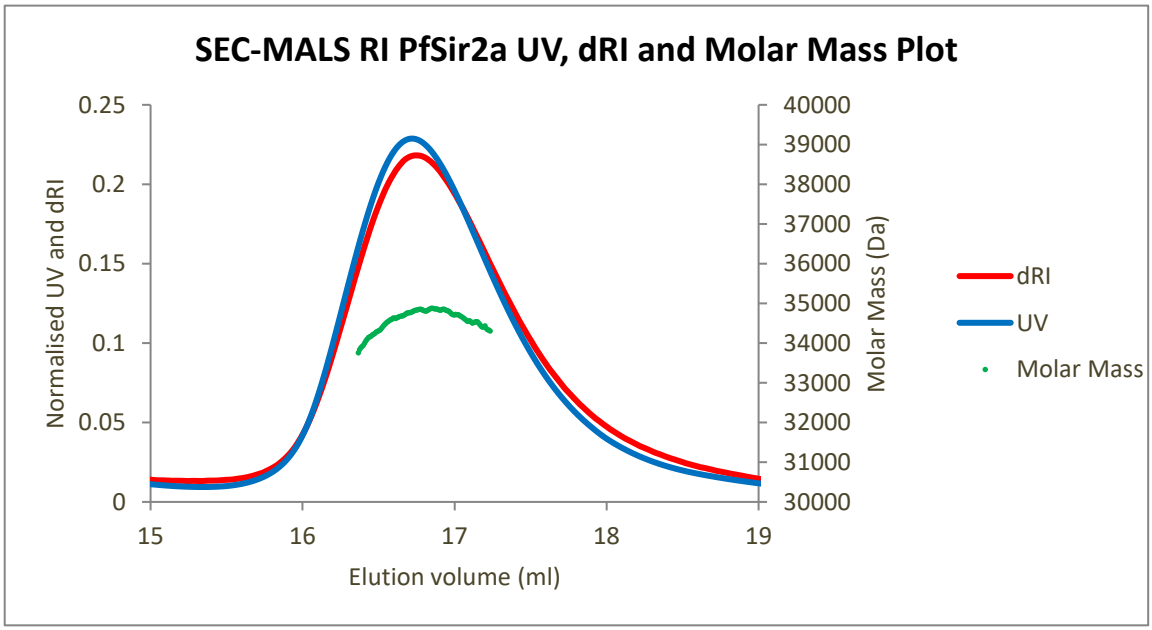


Figure 3.29: Plot of UV_{280nm} absorbance, dRI and molar mass over main peak observed in PfSir2a SEC-MALS RI experiment.

An interesting feature observed is that the plot of molar mass observed over the elution peak shows a curved nature. This results in an increase in observed molar mass as the sample reaches maximum concentration at its peak elution volume, followed by a decrease as the sample passes out of the detector. This plot of molar mass should normally be a straight line with some small variation across the sample elution peak. This observed result indicates that there could be a co-migrating contaminant with a higher molecular weight. Alternatively a concentration depenent dynamic equilibrium could be occuring between aggregated and non aggregated species or between oligomeric states (e.g. monomer – dimer). As previous results have shown PfSir2a to be >95% pure as judged via SDS-PAGE and aggregates have not been observed via SEC or by DLS, this suggests that the more likely explanation is a dynamic equilibrium between oligomeric species.

A repeat experiment was performed using a *Superdex* S7510/300 GL analytical column at a sample concentration of 23.44 mg ml^{-1} which is more representative of sample concentrations required for SAXS/SANS experiments. The SEC-MALS RI results can be seen in Figure 3.30 and the elution profile closely resembles that observed in the previous experiment. The main elution peak (A) is observed at an elution volume of 11.2ml and the shoulder peak (B) at 9.7ml due to the change from the S200 to the S75 column.

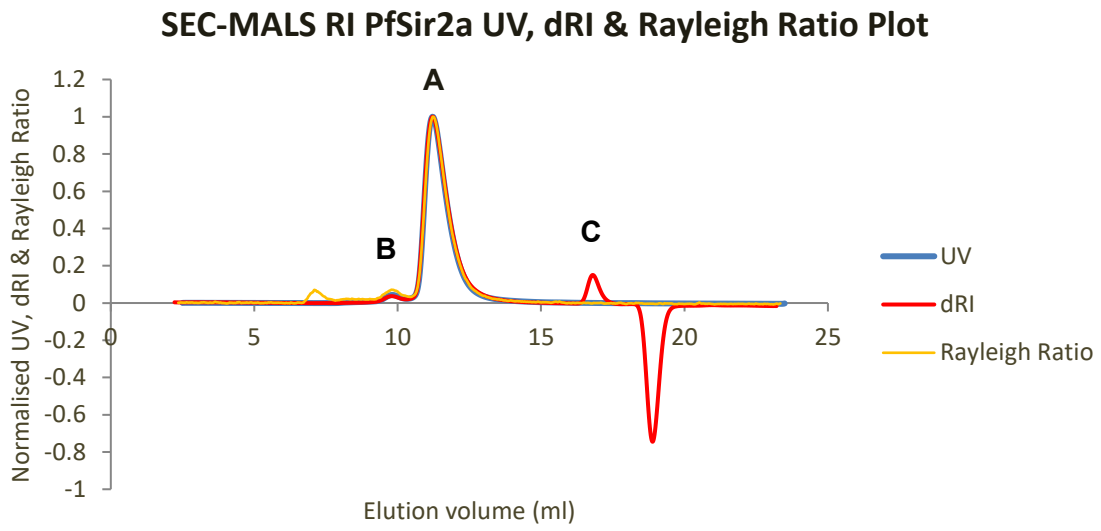


Figure 3.30: Elution profile of PfSir2a SEC-MALS RI experiment showing observed UV_{280nm} absorbance, Rayleigh ratio and dRI. The main elution peak and shoulder peak are indicated (A) and (B) respectively. The large change in the dRI at the end of the column is due to buffer exchange at the end of the column. A third small peak (C) at 17ml is due to small molecular weight contaminant protein.

Here, the average molar mass of the particles contained within the main peak in this experiment was calculated to at 40.29 kDa (± 0.935%) with a polydispersity index of 1.001 (±1.321%) (see Figure 3.31). This is ~6 kDa larger than the expected MW for a PfSir2a monomer and shows a slight increase in polydispersity (a polydispersity index of 1.000 indicates a monodisperse population). The average molar mass of the particles contained in the shoulder peak was found to correspond to a MW of 83.17 kDa (± 4.572%) with a polydispersity index of 1.002 (± 6.491%). The mass fractions for the main and shoulder peak were found to be 98.1% and 1.9% respectively.

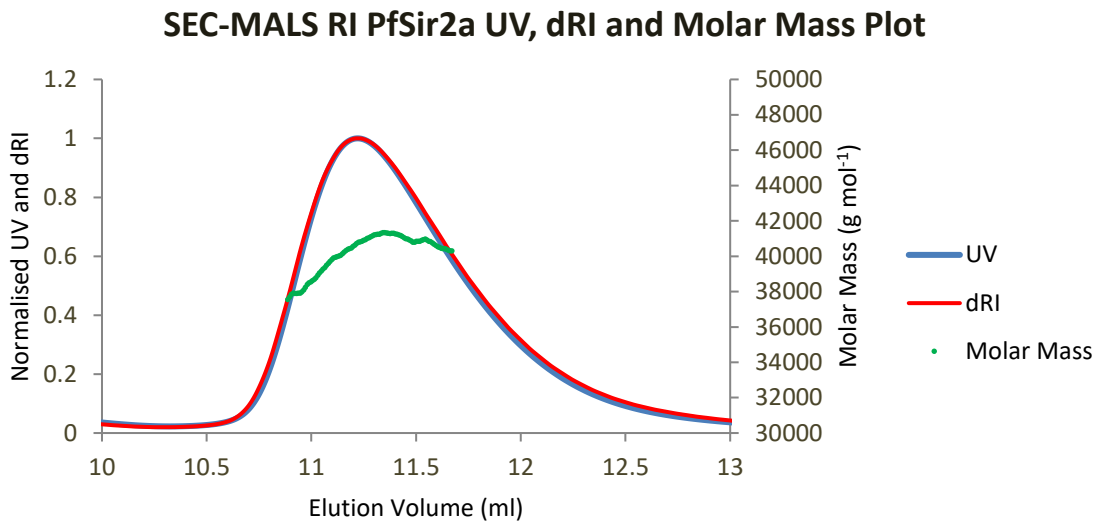


Figure 3.31: Plot of UV_{280nm} absorbance, dRI and molar mass over main peak observed in PfSir2a SEC-MALS RI experiment.

The curved plot of the calculated molar mass of PfSir2a as over the main elution peak is more pronounced with the increase in sample concentration. Initially a a lower concentration the calculated molar mass is found to be 37.56 kDa which corresponds to a similar value for the PfSir2a monomer. The MW increases to a value of 41.34 kDa at the peak elution volume before falling slightly to 40.32 kDa. This is similar to the results obtained for the shoulder peak at 16.6 ml in the initial experiment seen in Figure 3.25. The average MW of 83.17 kDa observed in the shoulder peak in this experiment is larger than an expected MW of 67.76 kDa for a PfSir2a dimer, but smaller than the expected MW of a PfSir2a trimer at 101.64 kDa.

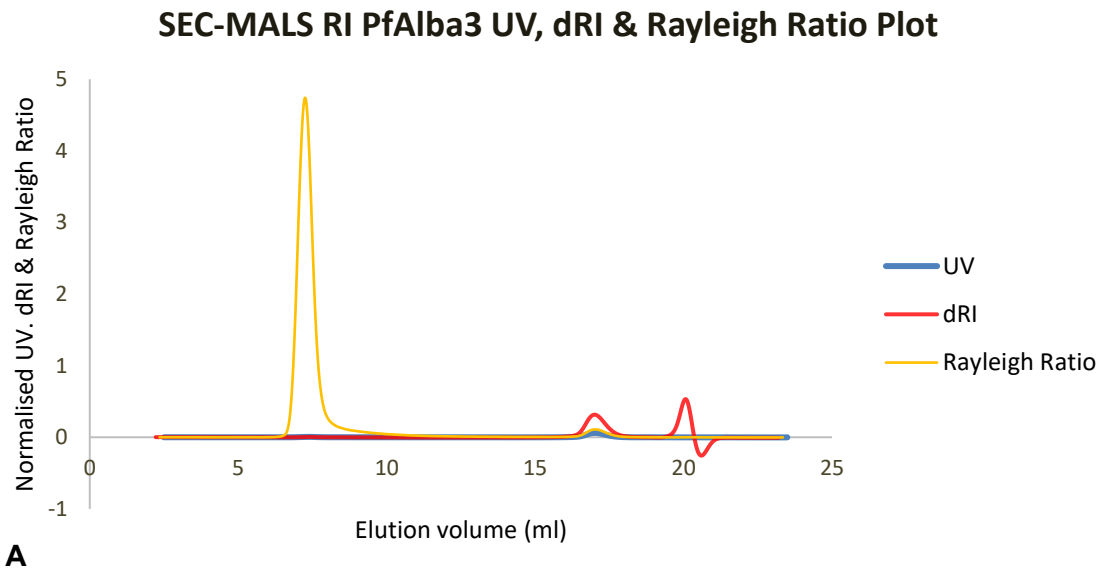
The overall indication is that there is a concentration-dependent dynamic equilibrium between different oligomeric states of PfSir2a. At the low concentrations observed in the intial SEC-MALS RI experiment this appears to be a monomer-dimer equilibrium. For the higher concntration experiment an additional oligomeric state could potentially exist with a

monomer-dimer-trimer equilibrium. Further characterisation is therefore necessary to understand the nature of the species present in PfSir2a samples.

3.7.2 SEC-MALS RI Results For PfAlba3

Normalised SEC-MALS RI results for PfAlba3 can be seen in Figure 3.32. The results show two peaks that are present in the UV 280nm absorbance, Rayleigh ratio and dRI at elution volumes of 7.4ml and 17.02ml. This corresponds to the aggregated and non aggregated PfAlba3 respectively. An important observation is that despite the relatively low UV absorbance observed for the aggregated PfAlba3 peak, compared to the non aggregated peak, the Rayleigh ratio is observed to be far greater (see Figure 3.32 panel B). This indicates that very large aggregates particles are forming in the protein sample as it aggregates over time.

The average molar mass across the main peak was found to correspond to a MW of 13.44 kDa (±1.737%) with a polydispersity index of 1.001 (± 2.463%). This is in agreement with the expected molecular weight of 13.04 kDa for a PfAlba3 monomer (see Figure 3.33). The molar mass calculated for PfAlba3 over the main protein peak also changes across the elution volume where a steady increase is observed. This could be caused by small populations of aggregating proteins as they pass through the column and cause a slight increase in the polydispersity of the sample.



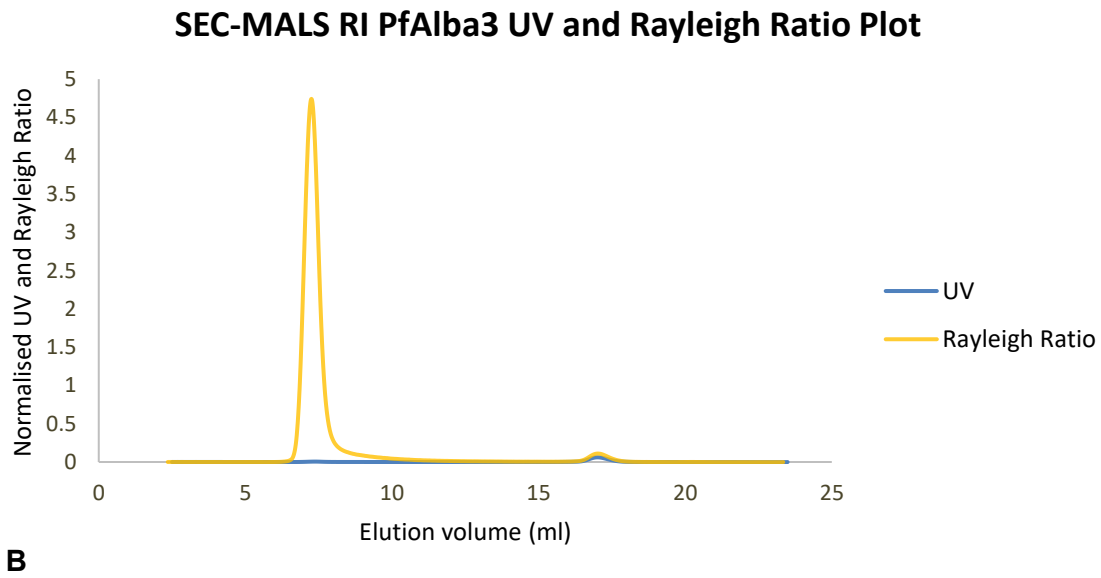


Figure 3.32: (A) Elution profile of PfAlba3 SEC-MALS RI experiment showing observed UV_{280nm} absorbance, Rayleigh ratio and dRI. (B) Plot of UV_{280nm} absorbance and Rayleigh ratio highlihting the aggregate peak at 7.4ml and the main protein peak at 17.02ml. The change in the dRI at the end of the column is due to buffer exchange at the end of the column.

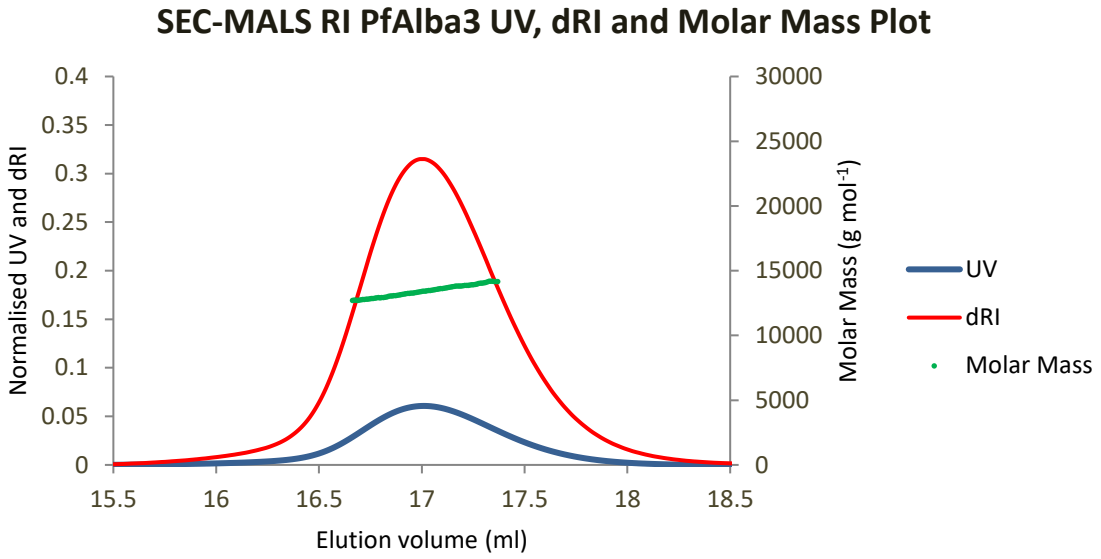


Figure 3.33: Plot of UV_{280nm} absorbance, dRI and molar mass over main peak observed in PfAlba3SEC-MALS RI experiment. An increase in molar mass across the elution peak may indicate the prescence of aggregates or populations of different oligormeric states of PfAlba3.

A repeat experiment was also performed using a *Superdex* S7510/300 GL analytical column at a sample concentration of 9.4 mg ml^{-1}. The SEC-MALS RI results can be seen in Figure 3.34 and the elution profile again closely resembles that observed in the previous experiment. The aggregate peak is observed at 7.8ml and the main protein peak at 12.3ml due to the change from the S200 to the S75 column.

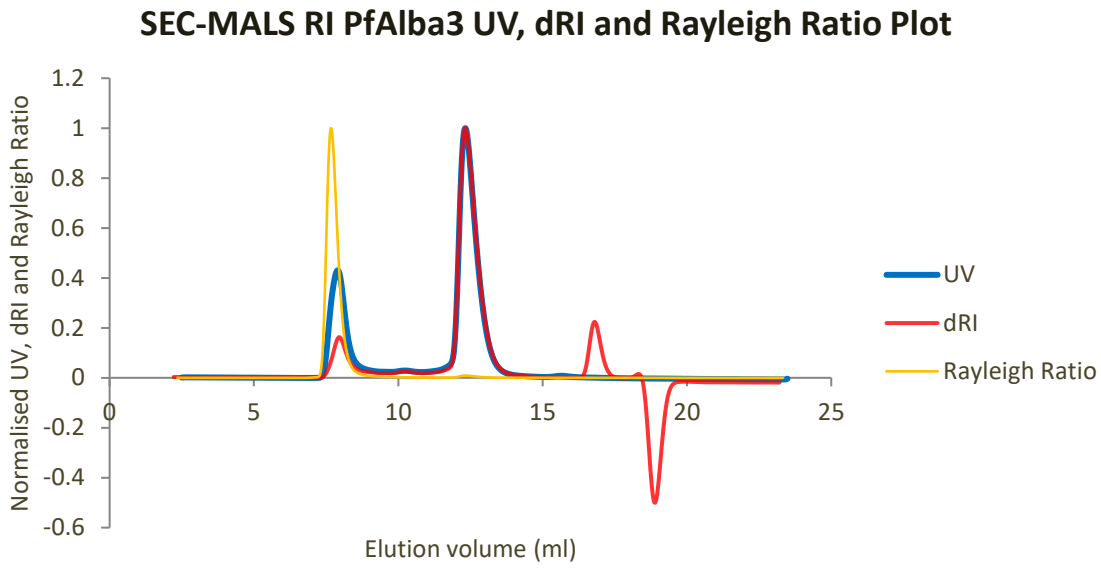


Figure 3.34: Elution profile of PfSir2a SEC-MALS RI experiment showing observed UV_{280nm} absorbance, Rayleigh ratio and dRI.

Here, the average molar mass of the particles contained within the main peak in this experiment was calculated to represent a MW 16.78 kDa (± 1.430%) with a polydispersity indexof 1.000 (±2.023%) (see Figure 3.35). This also shows an upward trend in the observed molar mass upon using a much higher sample concentration. The molar mass continues to change across the peak but begins at the higher value of 17.8 kDa before quickly dropping to 16.5 kDa. A steady increase is then observed again across the elution peak as the sample passes through the column. The mass fractions for the aggregate and protein peak was found to be 16.1% and 83.9% respectively.

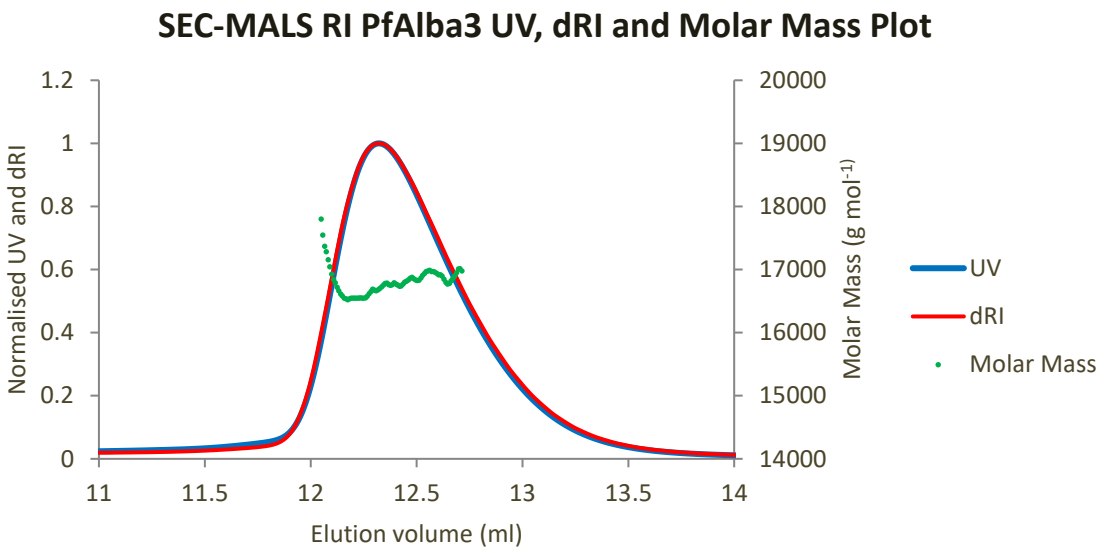


Figure 3.35: Plot of UV_{280nm} absorbance, dRI and molar mass over main peak observed in PfAlba3 SEC-MALS RI experiment.

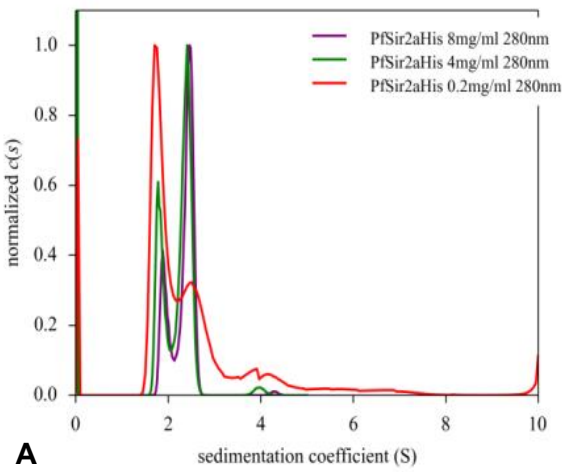
These results also indicate there may be a concentration dependent effect on PfAlba3 causing a shift towards the formation of oligomeric states (e.g. monomer-dimer) or protein aggregates, whilst being few in number, are forming as the sample migrates through the column. Additional characterisation is again needed to understand the composition and behaviour of the sample before SAXS and SANS experimentation.

3.8 AUC Analysis Of PfSir2a And PfAlba3

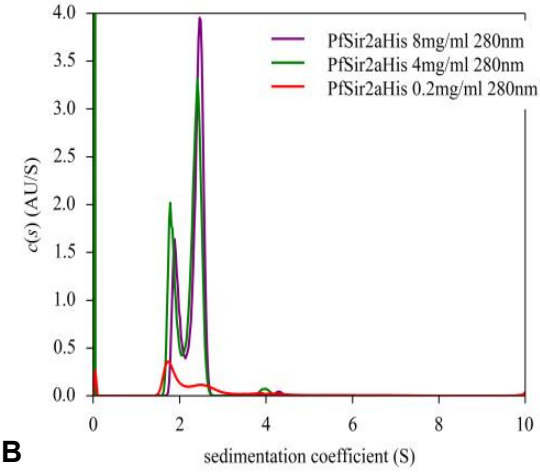
This section details the results observed for PfSir2a and PfAlba3 using AUC as described in Chapter 2, section 2.2.5.

3.8.1 AUC Results For PfSir2a

A concentration series of 0.2, 4 and 8mg ml^{-1} was prepared for PfSir2a as described in Chapter 2. 2.5. Results of the analysis can be seen in Figure 3.36.



A



B

PfSir2aHis	1.87 +/- 0.05 S				2.47 +/- 0.09 S				mean 1.87-2.47S		4.28 +/- 0.21 S				Aggregates
280 nm	s	s20w	signal, 1cm	%	s	s20w	signal, 1cm	%	s	s20w	s	s20w	signal, 1cm	%	%
8 mg/ml	1.93	3.25	2.23	26	2.44	4.12	6.41	74	2.31	3.89	4.29	7.25	0.06	1	0
4 mg/ml	1.84	3.10	1.44	32	2.38	4.02	2.98	67	2.20	3.72	3.97	6.71	0.07	2	0
0.2 mg/ml	1.82	3.07	0.11	52	2.56	4.32	0.06	29	2.09	3.52	4.49	7.57	0.04	18	1

C

Figure 3.36: (A) AUC results of PfSir2a plotted as normalised concentration vs sedimentation coefficient. (B) Total UV_{280nm} absorbance of each sample vs sedimentation coefficient. (C) Table indicating observed sedimentation coefficients and aggregate populations.

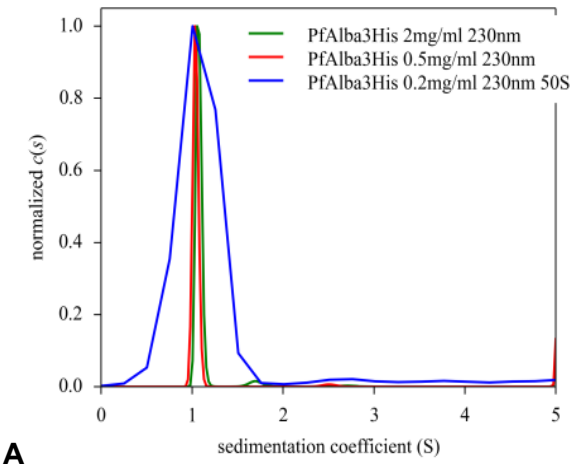
For the AUC results for PfSir2a samples, at the three concentrations 0.2, 4 and 8 mg ml^{-1}, we observe three species at 278nm. Firstly, a contribution at 1.9 +/- 0.1S (s_{20w} = 3.2 +/- 0.1S) is observed which may correspond to a globular compact monomer or an elongated

dimer. The second contribution is observed at 2.5 +/- 0.1S (s_{20w} = 4.2 +/- 0.2S). This s-value may correspond to a globular dimer or an elongated trimer. The third contribution is observed at 4.3 +/- 0.2S (s_{20w} = 7.1 +/- 0.4S) which may correspond to a globular compact tetramer or a more elongated multimer. A clear concentration effect can be observed for the two-main species (1.9S and 2.48S) indicating an equilibrium of association-dissociation. The proportion of the smaller species (1.9S) decreases when the concentration increases (52% at 0.2 mg ml^{-1}, 32% at 4 mg ml^{-1} and 26% at 8mg ml^{-1}) and the proportion of the larger species (2.48S) increases when the concentration increases (30% at 0.2 mg ml^{-1}, 68% at 4 mg ml^{-1} and 74% at 8m ml $^{-1}$).

The calculated MW for the species at 1.9 S is found to be 34.72 kDa (±2.96 kDa) indicating that this is a PfSir2a monomer. The calculated MW for the species at 2.5S is found to be 92.18 kDa (±18.73 kDa). There is a large uncertainty in the MW for this species and, of the expected MW which correspond to either a dimer (67.76 kDa) or a trimer (101.64 kDa) species, the trimer MW falls within the error range. The MW of the species at 4.3S was found to be 174.57 kDa (±129.72 kDa) and is suggested to be the result of reversible aggregation favoured at low concentrations.

3.8.2 AUC Results For PfAlba3

A concentration series of 0.2, 0.5 and 2mg ml^{-1} was prepared for PfAlba3. Results of the analysis can be seen in Figure 3.37.



A

PfAlba3His		1.12 +/- 0.09 S				Aggregates
		s	s20w	signal, 1cm	%	%
2	J	1.10	1.79	4.34	100	0
mg/ml	280nm	1.13	1.85	0.35	90	10
	230nm	1.07	1.75	8.93	95	5
0.5	J	1.03	1.67	0.53	100	0
mg/ml	280nm	1.21	1.97	0.06	81	19
	230nm	1.04	1.69	1.96	97	3
0.2	J	1.15	1.88	0.09	48	52
mg/ml	280nm	1.09	1.78	0.02	85	15
	230nm	1.09	1.78	0.39	90	10

B

Figure 3.37: (A) AUC results of PfAlba3plotted as normalised concentration vs sedimentation coefficient measured using UV_{230nm} absorbance. (B) Table indicating observed sedimentation coefficients and aggregate populations.

For the AUC results for PfAlba3 samples, at the three concentrations 0.2, 0.5 and 2 mg ml^{-1}, we observe a single species at 230nm. This contribution is observed at 1.1 +/- 0.1 S which may correspond to a globular compact monomer or an elongated dimer. Aggregates are also present in different proportions (< 4% above 0.5 mg ml^{-1} and 10% at 0.2 mg ml^{-1}, with aggregate species up to 50S). The fit for the 0.2 mg ml $^{-1}$ data is not very good probably because of the increased ratio between protein-protein aggregates in the sample. The calculated MW for the species observed at 1.1S was found to be 12.7 kDa (± 1.8 kDa). This is consistent with the interpretation of a globular compact monomer of PfAlba3 (MW = 13.04 kDa).

3.9 Interaction Studies Between PfSir2a And PfAlba3

The work presented here was aimed at investigating and characterising the previously described interaction between PfSir2a and PfAlba3 (Goyal et al., 2012). Molecular biology methods such as pull-down experiments were conducted with His-tagged and non-tagged proteins to try to confirm an interaction. This was supported by biophysical investigation methods such as ITC and MST to try to determine the strength and nature of the interaction.

3.9.1 Preparation Of Non-Tagged PfSir2a

The recombinant PfSir2a protein construct expressed by the pET28a vector was designed to have an N-terminal 6xHis tag which can be cleaved by Thrombin protease. This protease has a recognition site of Leu-Val-Pro-Arg-Gly-Ser (LVPRGS) and cleaves the peptide bond between arginine and glycine. The N-terminal sequence for the first 40 residues of the pET28a PfSir2a construct can be seen in Figure 3.38. By this design a 17 AA sequence (MW = 1900 Da) containing the 6 His tag is removed by the action of thrombin protease. This produces a non-tagged version of PfSir2a with a MW of 32.0 kDa from the original 33.88 kDa full length construct.

Figure 3.38: First 50 residues of the N-terminal sequence of recombinant PfSir2a in pET28a construct. The thrombin protease recognition site is highlighted in blue and the cleavage location is marked by the arrow. The starting methionine of the native PfSir2a sequence is highlighted in green.

Initial small-scale thrombin cleavage trials were conducted on PfSir2a using a biotinylated-thrombin cleavage kit (Novagen Cat no 69022-3). Incubation reactions were set up containing 5µl of 10x Thrombin Cleavage/Capture buffer (containing 200mM Tris-HCl, 1.5M sodium chloride, 25mM calcium chloride, pH 8.4), 10µg of the target protein, 1µl of diluted biotinylated thrombin solution (1 U μl^{-1} in 50mM sodium citrate, 200mM sodium chloride, 0.1% PEG-8000, 50% glycerol, pH 6.5) and made to a final volume of 50µl with ddH_2O. The reactions were then left to incubate for 24 hours at 21˚C or 4˚C. At separate time intervals, the biotinylated thrombin was removed by adding 25µl of streptavidin agarose bead 50% slurry. This was incubated for 30 minutes prior to centrifugation at 500*g* for 5 minutes. The supernatant was removed and prepared for analysis by SDS-PAGE (see Figure 3.39).

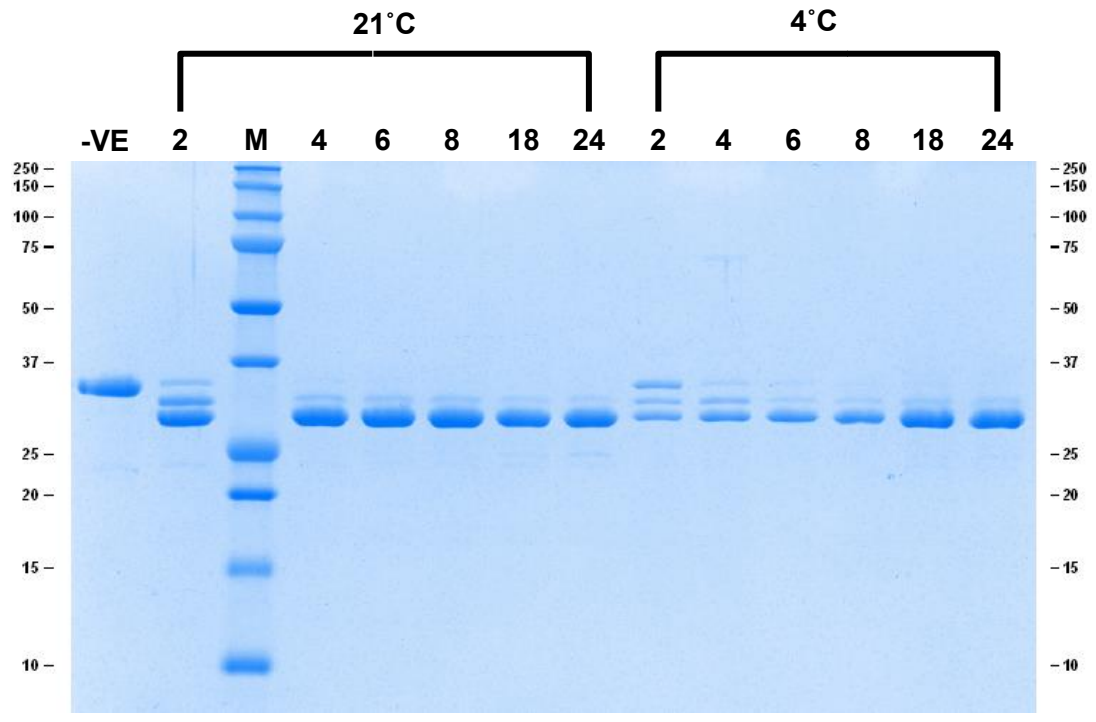
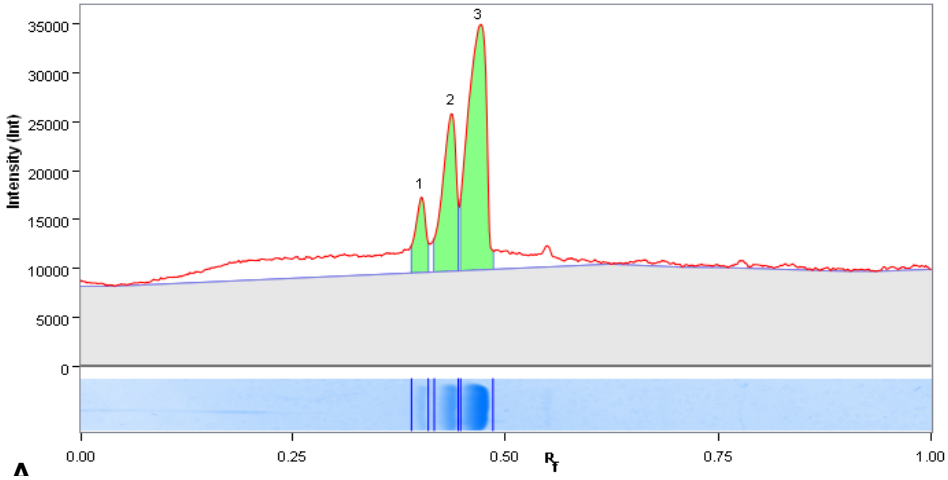


Figure 3.39: Small scale thrombin cleavage of PfSir2a at 21˚C and 4˚C results. Negative control indicated by –ve, MW marker by M and hours of incubation from 2 to 24.

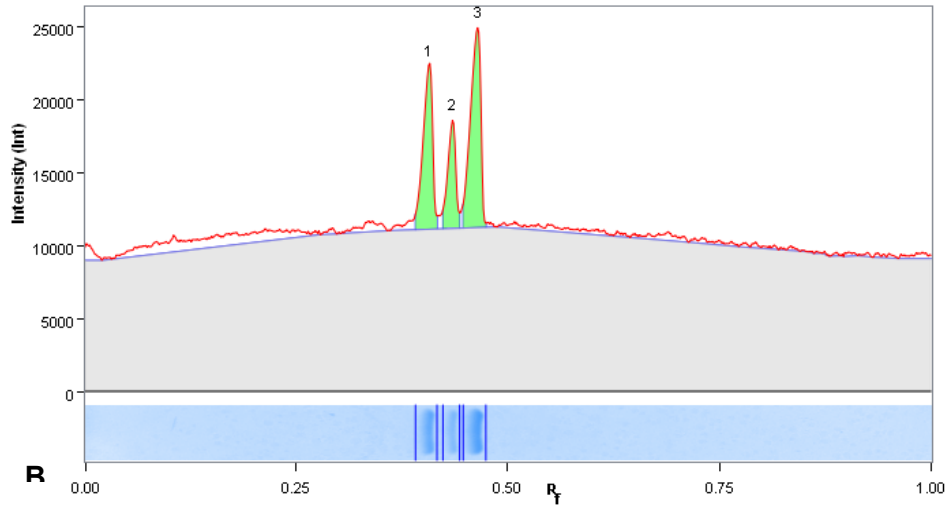
An interesting observation can be made from the SDS-PAGE results seen in Figure 3.36. Three bands are observed in both the 2-hour incubation lanes at 21˚C and 4˚C. This reduces to two bands with increasing incubation time, occurring more rapidly at 21˚C than at 4˚C. An additional faint band can be observed in the 21˚C incubation samples at 18 and 24 hours incubation at an approximate MW of 25 kDa. By using the MW standards as a reference, analysis of the 2-hour incubation samples for both 21˚C and 4˚C reveals the estimated MW of each band present (see Figure 3.40).

The MW for each of the three bands are found to be approximately 34 kDa, 31.3 kDa and 29.1 kDa. The band at 34 kDa represents non-cleaved full length recombinant protein, as compared with the negative control. The expected MW for the thrombin cleaved product is 32 kDa which is most similar to the band observed at 31.1 kDa. To explain the end result species observed at 29.1 kDa after 24 hours of incubation, samples were sent for analysis by TOF-MS. The results returned can be seen in Figure 3.41.



A

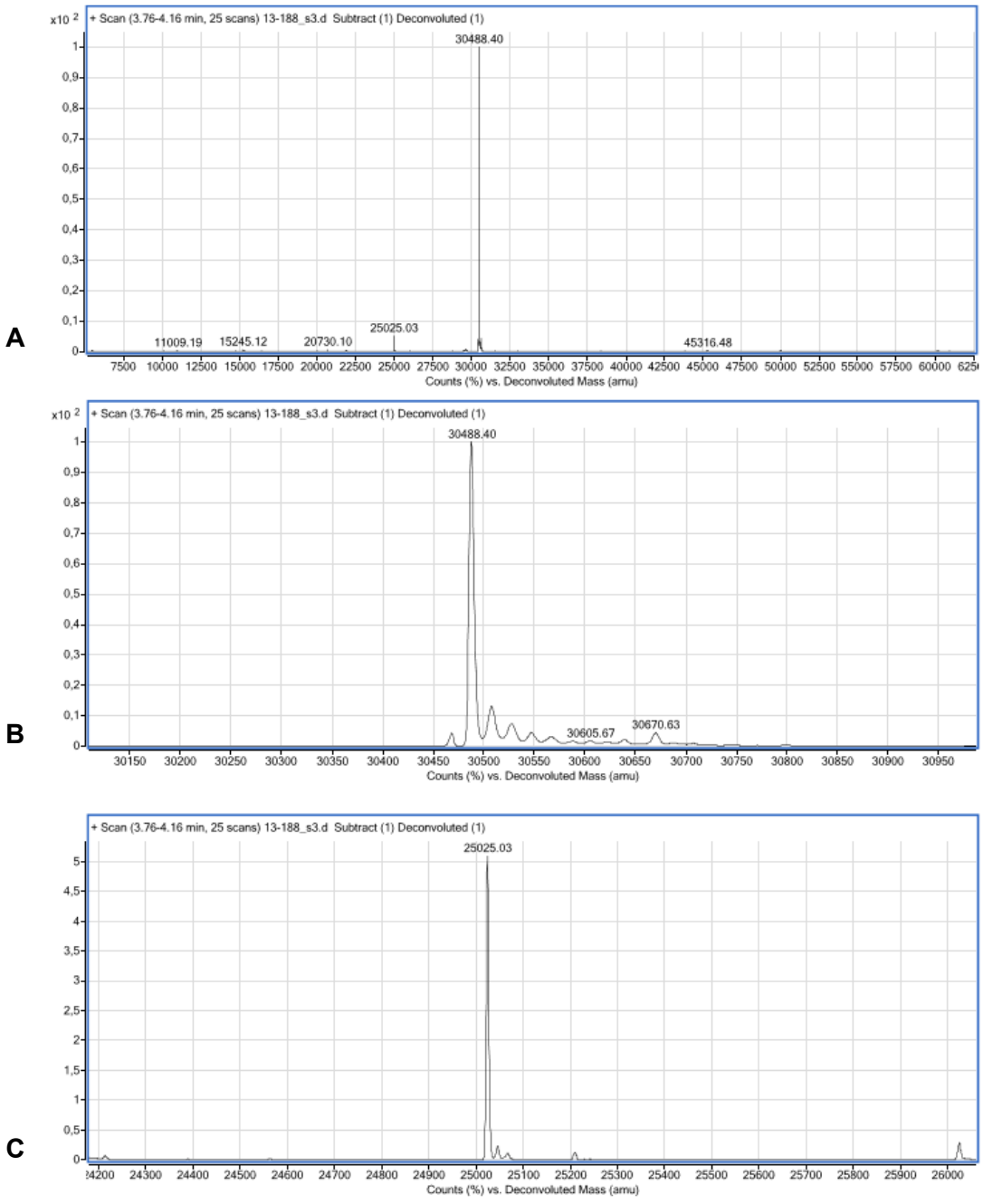
Band No.	Mol. Wt. (KDa)	Relative Front	Volume (Int)	Band %	Lane %
1	34.02	0.401408	6884692	10.56	5.91
2	31.29	0.43662	18742197	28.75	16.09
3	28.97	0.469014	39542344	60.67	33.95



B

Band No.	Mol. Wt. (KDa)	Relative Front	Volume (Int)	Band %	Lane %
1	33.57	0.407042	8797519	34.82	20.05
2	31.39	0.435211	4903846	19.41	11.18
3	29.26	0.464789	11557463	45.75	26.34

Figure 3.40: Molecular weight estimation of multiple species observed in small scale thrombin cleavage tests as a function of intensity vs relative migration (R_f). (A) Analysis for 2-hour incubation lane at 21˚C. (B) Analysis for 2-hour incubation lane at 4˚C.



Sample	Observed mass (Da)	Expected mass (Da)	Mass deviation (Da)	Mass accuracy (ppm)*
Cleaved PfSir2a	**30488.40** 25025.03	**32006.29** **30488.56**	0.16	5

Figure 3.41: TOF-MS analysis of thrombin cleaved PfSir2a sample after 24 hours cleavage reaction. (A) Whole TOF-MS spectra for PfSir2a sample. (B) Zoom in region of predominant peak corresponding to majority of cleaved sample. (C) Zoom in region of secondary peak also present in sample corresponding to product from additional cleavage by Thrombin protease.

The majority of the sample is a species with a MW of 30488.56 Da. This is 1517.54 Da smaller than the expected MW of the thrombin cleaved PfSir2a sequence. It has been observed that whilst Thrombin has a specific recognition site as mentioned previously, additional cleavage may occur at a Gly-Arg-Gly site where the peptide bond is cleaved again after the residue (Gallwitz et al., 2012). A search for this sequence in the full-length recombinant PfSir2a revealed this site to be present (see Figure 3.42).

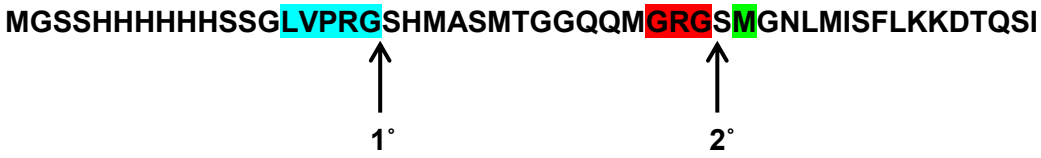


Figure 3.42: Primary and secondary (1˚ and 2˚ respectively) cut sites for Thrombin in PfSir2a AA sequence identified.

Cleavage at this secondary site results in the removal of a 32 amino acid (MW 3417.7 Da) peptide from the N-terminus of the recombinant PfSir2a. This results in an additional 15 amino acid removal from the intended cleaved PfSir2a form and yields a final cleaved product of 275 amino acids (MW = 30488.4 Da). This species matches exactly the MW of the main species observed in the TOF-MS results and differs by only two AA from the native *P. falciparum* sequence (273 amino acids, MW = 30344.4 Da).

Additional cleavage by Thrombin can also sometimes occur at a site with a X-Arg-Gly sequence where X is an aromatic amino acid and cleavage of the peptide bond occurs after the arginine (Gallwitz et al., 2012). A search of the recombinant sequence also reveals a Phe-Arg-Gly site (FRG) starting at position 82. Cleavage at this site would result in a sequence 224 amino acids in length with a calculated MW of 25025.0 Da. This again matches the MW of the species observed in the TOF-MS results, although it is present in a negligible amount.

Therefore, thrombin appears to cleave the recombinant full length PfSir2a in stages, initially removing the first 17 amino acids as intended by the cut-site designed in the pET28a vector, then removing an additional 15 amino acids due to the secondary cleavage site, which produces a smaller product which more closely resembles that of the native *P. falciparum* amino acid sequence. This would be more ideal for interaction studies and crystallisation trials.

Efforts were then made to scale up the thrombin cleavage reaction using several mg of target protein and leaving the cleavage reactions to proceed over 24 hours at 21˚C and 4˚C. The cleavage buffer was kept as used previously in the Novagen Thrombin capture kit and Thrombin from bovine plasma (Sigma-Aldrich Cat no. T4648) was added at a ratio of 1 Unit per 1mg of target protein.

It was discovered when scaling the reactions to more useful quantities of material, a white precipitate was observed forming rapidly (< 1 hour at 21˚C, < 4 hours at 4˚C) in the reactions. This was found to be a precipitated mixture of the two cleaved forms of PfSir2a which could not be recovered. The cleavage buffer was changed to the SEC buffer used for previous experiments (50mM sodium phosphate, 500mM sodium chloride, pH 7.0). Some precipitate was still observed at 21˚C over the course of the reaction incubation but cleaved product could be recovered.

Cleaved PfSir2a was recovered by injecting the reaction onto a 1ml His-Trap™ FF crude column (GE healthcare, Cat no. 11-0004-58) connected in series to a 1ml Benzamidine Hi-Trap™ FF column (GE healthcare, Cat no. 17-5144-01) which was used to remove the Thrombin from the reaction mixture. An ÄKTA Prime FPLC system was used at a flow rate of 0.5 ml $min^{-1}$with a loading and equilibration buffer used for the columns that contained 50mM sodium phosphate, 500mM sodium chloride, 10mM Imidazole, pH 7.0.

Cleaved protein was monitored as it passed through the columns by UV_{280nm} absorbance and recovered in 1ml fractions. Any remaining tagged protein which was not cleaved but attached to the Ni-NTA column was removed using the elution buffer described previously for PfSir2a Ni-NTA purification. The fractions were analysed by SDS-PAGE (see Figure 3.43).

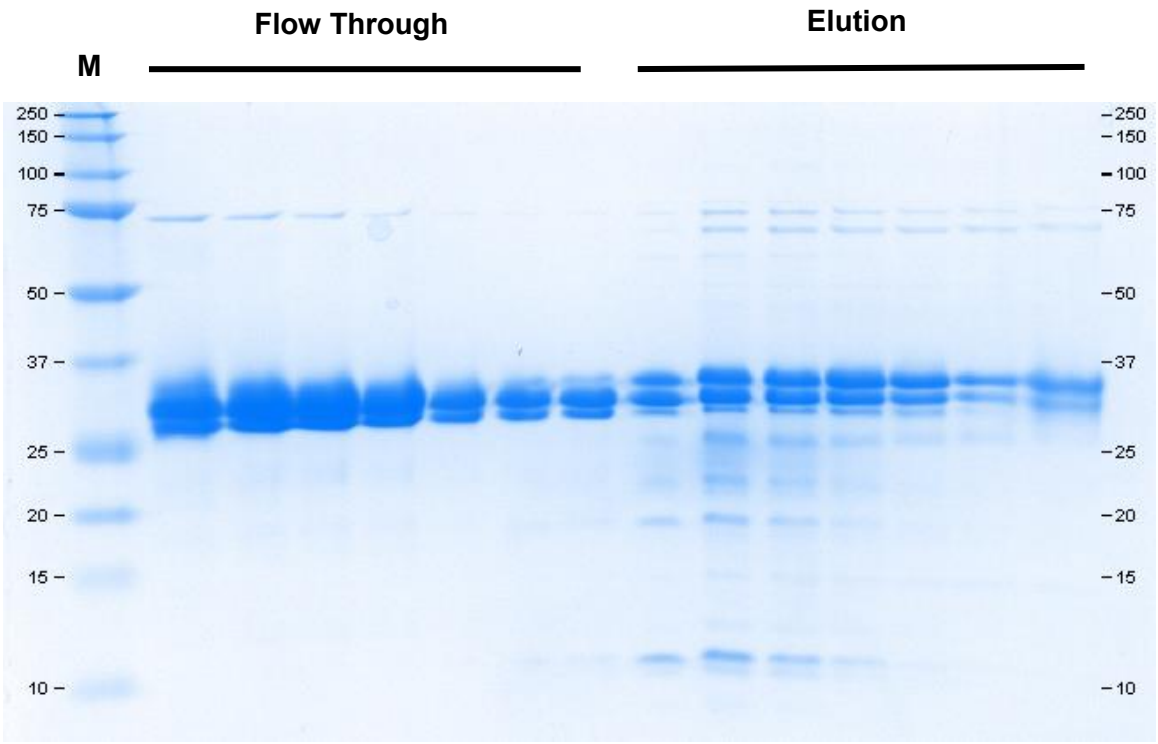


Figure 3.43: SDS-PAGE analysis of PfSir2a cleavage reaction recovery on 12% gel. MW marker is indicated by M, flow through and elution species are labelled respectively.

The SDS-PAGE results indicate that there is still a mixture of species being produced after scaling up the reaction volumes. There also appears to be an interaction effect occurring between the different species: trace amounts of a band corresponding to tagged PfSir2a appear to be washed off the column with cleaved species, as observed in the flow through lanes. Similarly, cleaved PfSir2a can be observed remaining on the column despite no longer possessing its 6 His-tag and this elutes with tagged species. This is probably due to

the interacting nature observed previously in the SEC-MALS RI and AUC results in sections 3.7.1 and 3.8.1 respectively.

The flow through fractions were cleaned up by IEX chromatography as described previously in Chapter 2, section 2.1.13 and an additional SEC step was performed to check for the presence of aggregates (see Figure 3.44). One main peak was present as before and fractions were recovered and again analysed by SDS-PAGE.

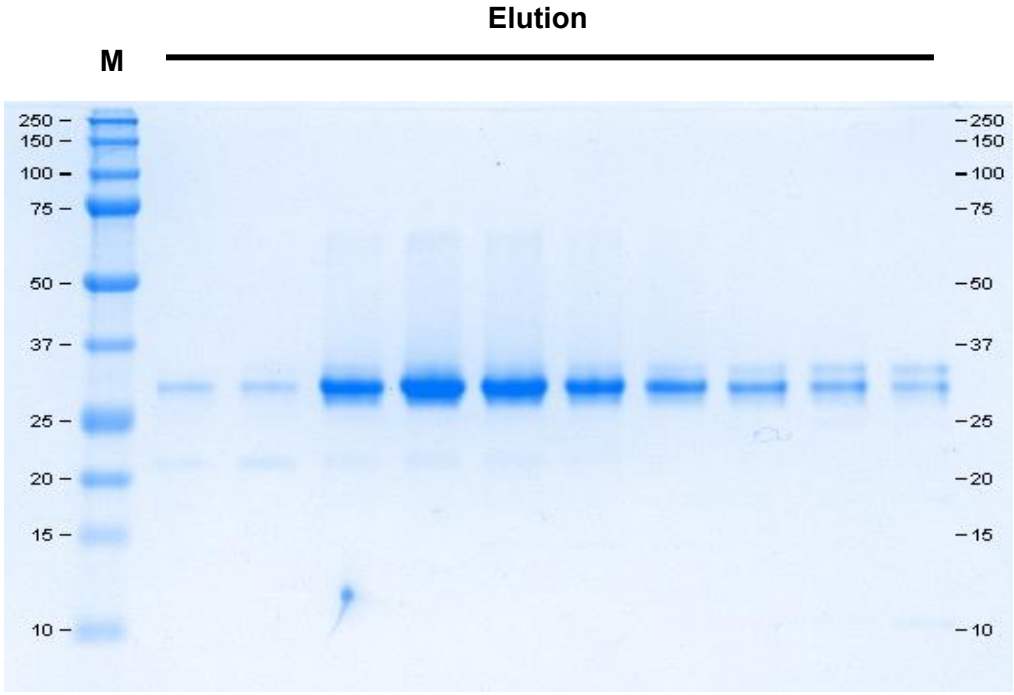


Figure 3.44: Fractions taken across elution peak analysed by 12 % SDS-PAGE.

The results seen in Figure 3.44 indicate that there is still a small presence of the primary cleaved species of PfSir2a which has co-migrated through the additional purification steps but appears to be eluting at a later elution volume than the secondary cleaved species. Fractions appearing to contain only the secondary cleaved PfSir2a were kept for use in pull-down experiments with PfAlba3.

3.9.2 PfSir2a And PfAlba3 Ni-NTA Pulldown Experiments

Pull down experiments to investigate the interaction between PfSir2a and PfAlba3 were performed using 200µl of Protino® Ni-NTA 50% agarose resin slurry (Macherey-Nagel, Cat no. 745400.25). The resin was centrifuged at 700*g* for 2 minutes, leaving a resin volume of approximately 100µl, before washing twice with 1ml of ddH_2O. The resin was centrifuged again at 700*g* for 2 minutes before washing with 1ml of binding buffer containing 50mM sodium phosphate, 500mM sodium chloride, 10mM imidazole, pH 7.5. A 200 µl mixture of cleaved PfSir2a and His-tagged PfAlba3 were incubated together in the binding buffer for 1 hour at room temperature at a concentration of 0.25 mg ml^{-1} each.

The protein mixture was then incubated on the equilibrated resin for 30 minutes shaking at 100 rpm at 21˚C. After this incubation period the supernatant was removed and a sample was taken for SDS-PAGE analysis. The resin was then washed with 200µl of wash buffer containing 50mM sodium phosphate, 500mM sodium chloride, 40mM imidazole, pH 7.5. The resin was centrifuged for 700*g* for 2 minutes and the supernatant removed and a wash sample taken. This wash step was repeated a further two times with an additional wash sample taken from the supernatant of the third wash.

Bound proteins were eluted from the resin with two resin bed volumes of elution buffer containing 50mM sodium phosphate, 500mM sodium chloride, 500mM imidazole, pH 7.5. The resin was centrifuged at 700*g* for 2 minutes and the supernatant removed and a sample taken. Control reactions were set-up using only cleaved PfSir2a or tagged PfAlba3 as negative controls. All samples were loaded on a 16.5% Tris-Tricine gels for SDS-PAGE analysis (see Figure 3.45).

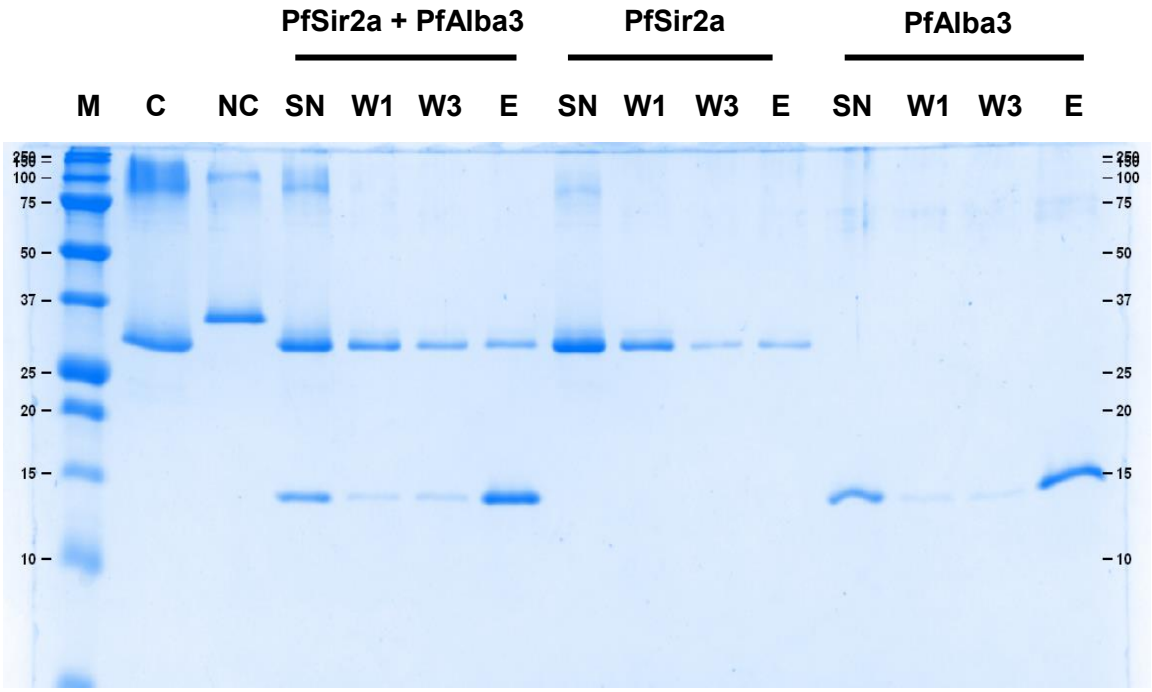


Figure 3.45: 16.5% SDS-PAGE results of Ni-NTA resin pulldown experiment with cleaved PfSir2a and tagged PfAlba3. C and NC indicated cleaved and non-cleaved PfSir2a respectively. Supernatant (SN), 1st and 3rd wash (W1 and W3) and elution (E) fractions indicated for reactions containing cleaved PfSir2a and tagged PfAlba3, cleaved Pfsir2a alone and tagged PfAlba3 alone.

The initial SDS-PAGE results reveal several features about the samples which have been tested via this pull-down technique. Firstly, the cleaved and non-cleaved PfSir2a are clearly distinguished and it is cleaved PfSir2a which has been used for the pull-down with no tagged protein remaining in the samples. Secondly, despite the tag no-longer being present on PfSir2a a significant fraction of the protein remains throughout the wash and elution steps in the cleaved PfSir2a only reactions. A reason for its continued presence could be due to insufficient washing to remove the protein after the initial resin incubation step. Additionally, not all of the PfAlba3 protein has successfully bound to the resin, despite being loaded a concentration well below the total binding capacity.

As a result of the negative controls providing false positives the lanes for the incubation reaction between cleaved PfSir2a and PfAlba3 cannot be interpreted as an indication of a interaction. Improvements to the pull-down protocol were made by increasing the volume of the wash steps to 1ml to fully make sure any free protein was no longer bound or weakly bound to the resin. The results from the modified pull-down protocol can be seen in Figure 3.46.

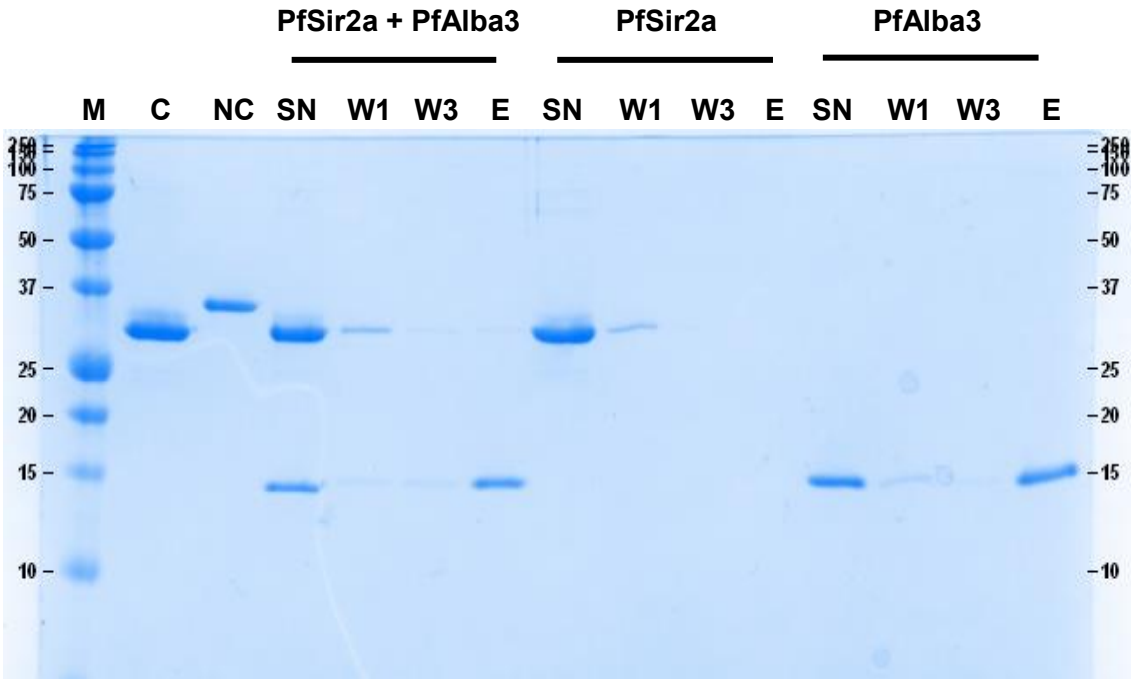
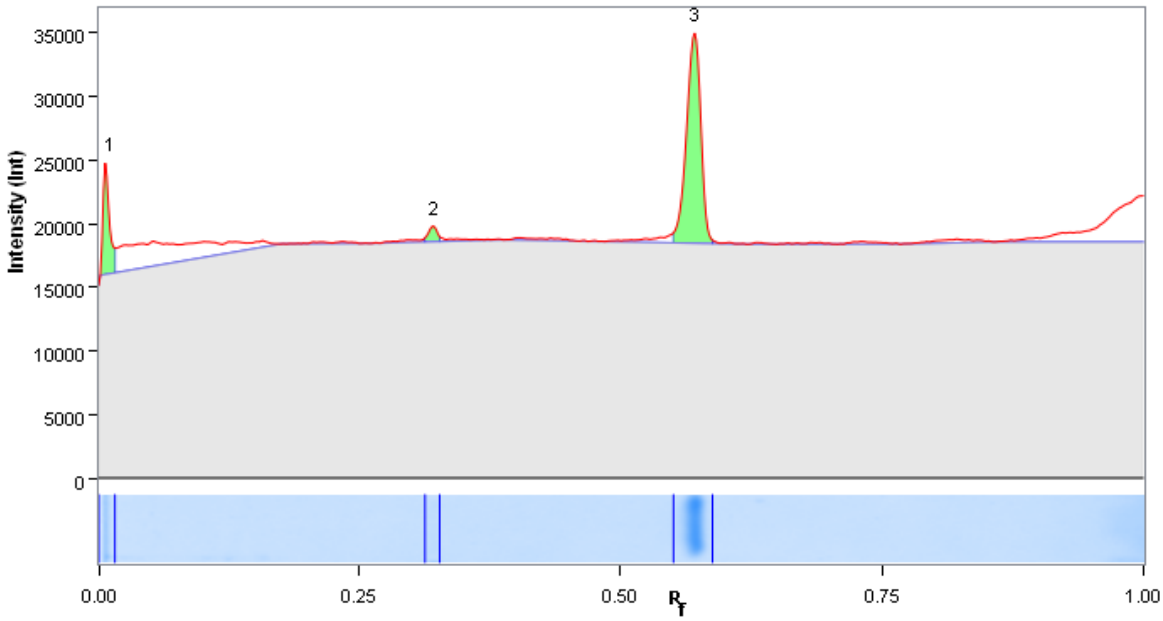


Figure 3.46: 16.5% SDS-PAGE results of repeat Ni-NTA resin pulldown experiment with cleaved PfSir2a and tagged PfAlba3. C and NC indicated cleaved and non-cleaved PfSir2a respectively. Supernatant (SN), 1st and 3rd wash (W1 and W3) and elution (E) fractions are indicated for reactions containing cleaved PfSir2a and tagged PfAlba3, cleaved PfSir2a alone and tagged PfAlba3 alone.

After the increase in washing performed during the protocol there is no longer a band present in the elution lane in the cleaved PfSir2a negative control samples. Tagged PfAlba3 is not completely binding to the resin but is still present in detectable quantity in

the elution lane. For the co-incubated lanes, a large proportion of cleaved PfSir2a is removed in the supernatant, indicating it is not bound to the resin or PfAlba3. A smaller quantity is removed in both wash lanes and is hardly visible in lane W3. In the elution lane we see a similar quantity of PfAlba3 as compared to that eluted in the PfAlba3 only incubation.

An extremely faint band at the MW migration of cleaved PfSir2a is also detected in this elution lane. The intensity for each protein can be observed by lane analysis (see Figure 3.47). This band is not observed in the elution sample lane for the cleaved PfSir2a only. The amount of cleaved PfSir2a is found to be < 2% of the total lane indicating that if there is an interaction between PfSir2a and PfAlba3 it appears to be extremely weak.



Band No.	Mol. Wt. (KDa)	Relative Front	Volume (Int)	Band %	Lane %
1	250	0.007764	3090648	18.73656	8.411959
2	30.27548	0.321429	590364	3.578988	1.606821
3	14.37031	0.571429	12814266	77.68445	34.87718

Figure 3.47: Lane analysis of cleaved PfSir2a and PfAlba3 incubation elution sample with MW of observed species and relative intensities.

3.9.3 Binding Interaction Characterisation By MST

To determine the strength and nature of the interaction between PfSir2a and PfAlba3 samples were prepared for analysis by MST as described previously in Chapter 2, section 2.2.7. The recovered labelled protein was then measured via Nanodrop (see Figure 3.48). The recovered labelled protein concentration was determined to be 0.094mg ml^{-1} or 2.78µM.

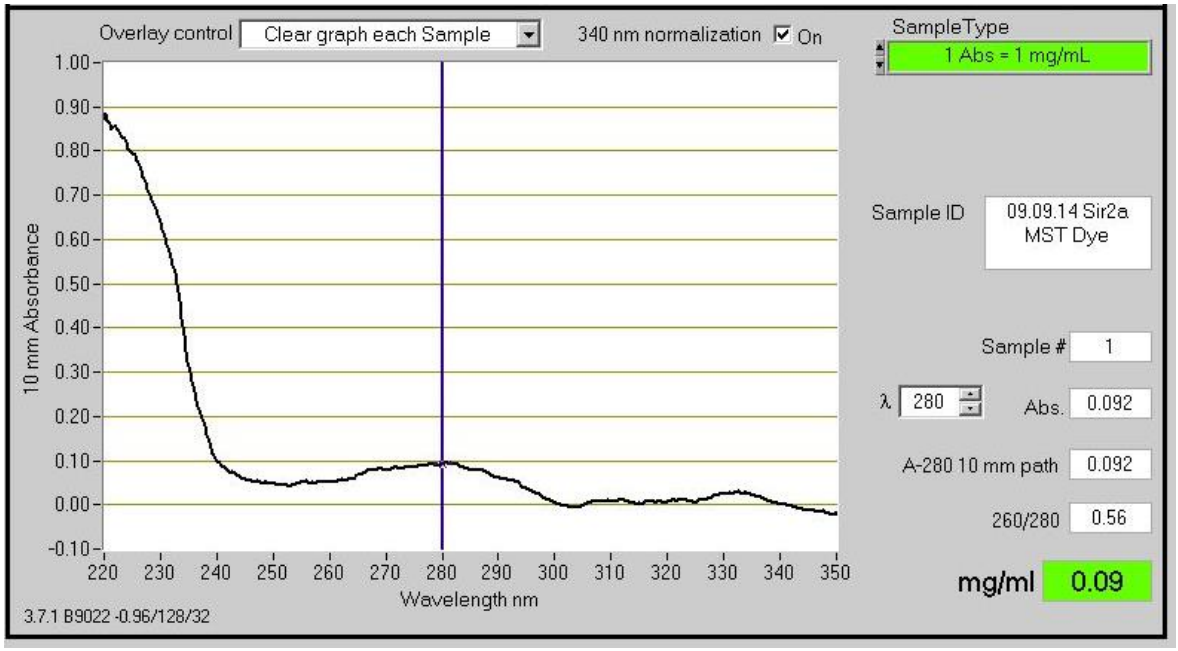


Figure 3.48: Nanodrop spectrum of recovered MST dye-labelled PfSir2a

A 16-point serial dilution (1:2) of unlabelled PfAlba3 from, 20µM to 0.6nM, was prepared with a fixed concentration of labelled PfSir2a at 20µM. The results of the MST experiment can be seen in Figure 3.49.

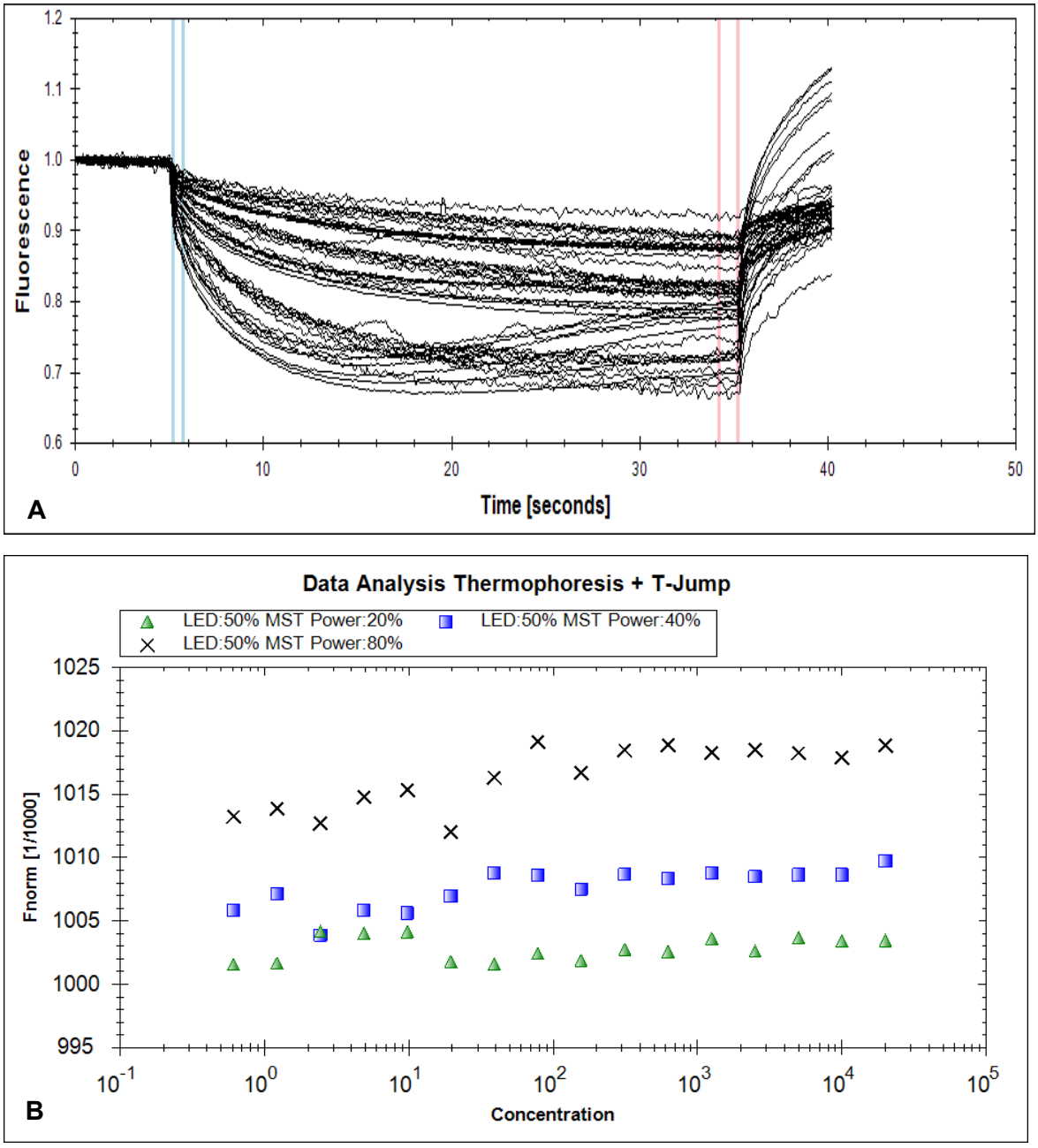
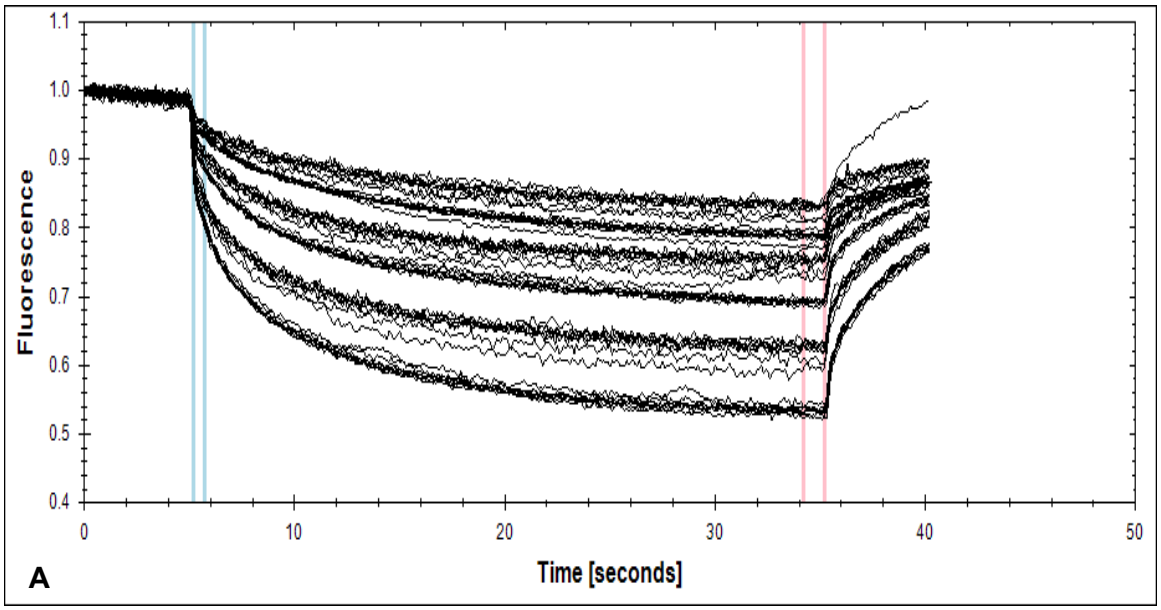


Figure 3.49: MST results for PfSir2a and PfAlba3 interaction studies using standard capillaries at 50% LED power and varied MST power. (A) Raw fluorescence time trace. (B) Thermophoresis and T-Jump results using 20%, 40% and 80% MST power.

The results in Figure 3.49 show a very noisy and bumpy fluorescence time trace throughout the duration of the experiment. This is attributed to the presence of aggregates within the sample capillaries. A fluorescence scan of the capillaries also revealed a 'twin peak' effect from capillary 6 onwards (data not shown). This is indicative of the protein sample sticking to the capillary and is not suitable for MST analysis. As a result the K_D of the interaction could not be calculated. By adjusting the LED power and changing the capillaries to a type with a hydrophobic coating, improvements in the results was observed (see Figure 3.50).

Here we can see an improvement in the fluorescence time trace for the duration of the experiment although it still appears quite bumpy. This indicates that there are still protein aggregates in the sample, likely because of PfAlba3 beginning to aggregate during the experiment based on its known propensity to aggregate rapidly after purification (see section 3.6.2). The capillary scan now reveals there are no longer any 'twin peaks' which occur because of sample sticking (data not shown). The optimal results were found to be in the samples tested at 100% LED power and 20% MST power.



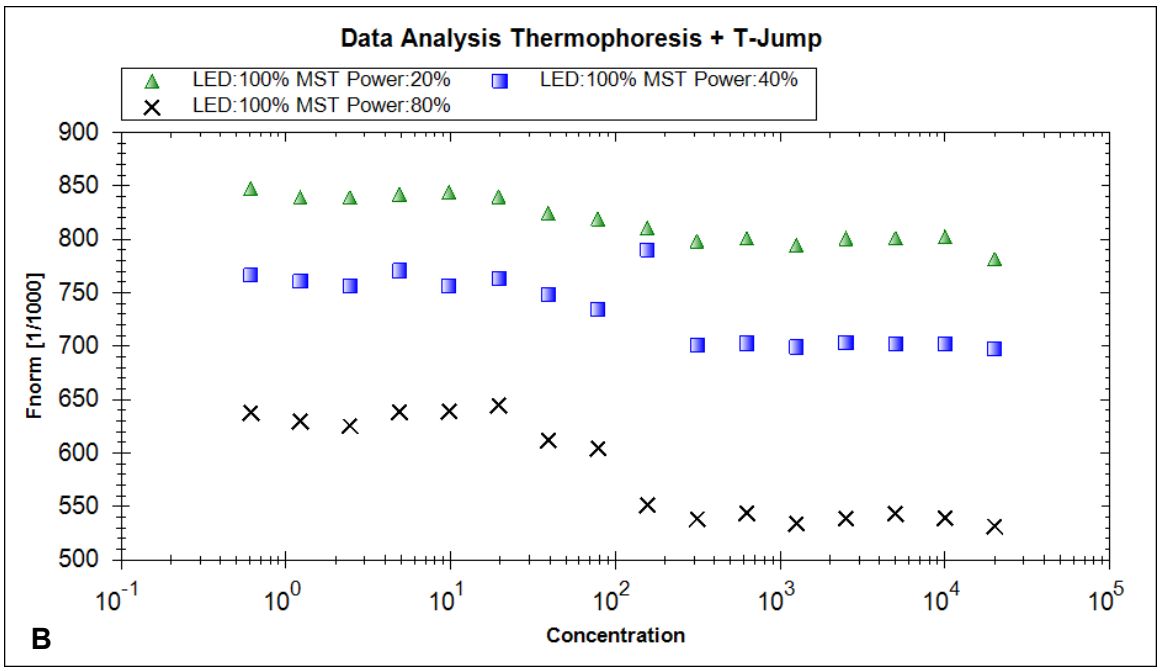


Figure 3.50: MST results for PfSir2a and PfAlba3 interaction studies using standard capillaries at 50% LED power and varied MST power. (A) Raw fluorescence time trace. (B) Thermophoresis and T-Jump results using 20%, 40% and 80% MST power.

A K_D of 29.9 ± 2.44 nM was calculated from these results (see Figure 3.51). This would seem to indicate a tight binding between the two proteins, but the result seems unlikely due to the extremely weak binding observed in the Ni-NTA pulldown experiment. To confirm this result, additional characterisation was needed.

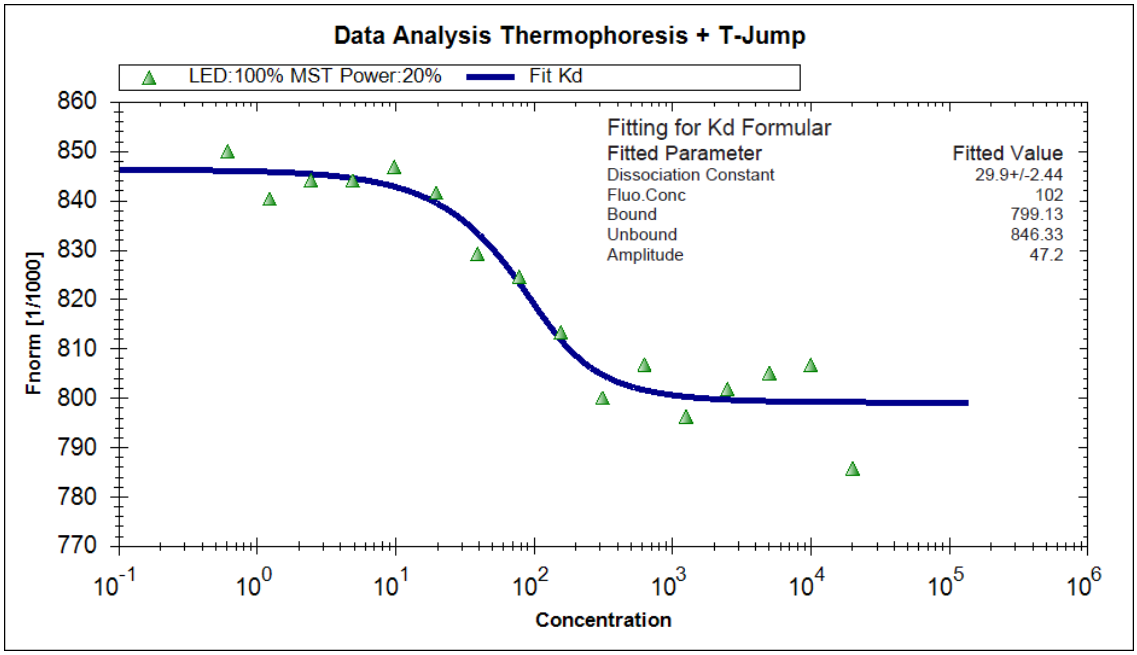


Figure 3.51: K_D fit for PfSir2a and PfAlba3 interaction calculated from MST results.

3.9.4 Binding Interaction Characterisation By ITC

ITC was used as an additional biophysical technique to characterise the interaction between PfSir2a and PfAlba3. Here purified PfSir2a and PfAlba3 samples were prepared as described in Chapter 2, section 2.2.6. The results of the ITC interaction experiments can be seen in Figure 3.52.

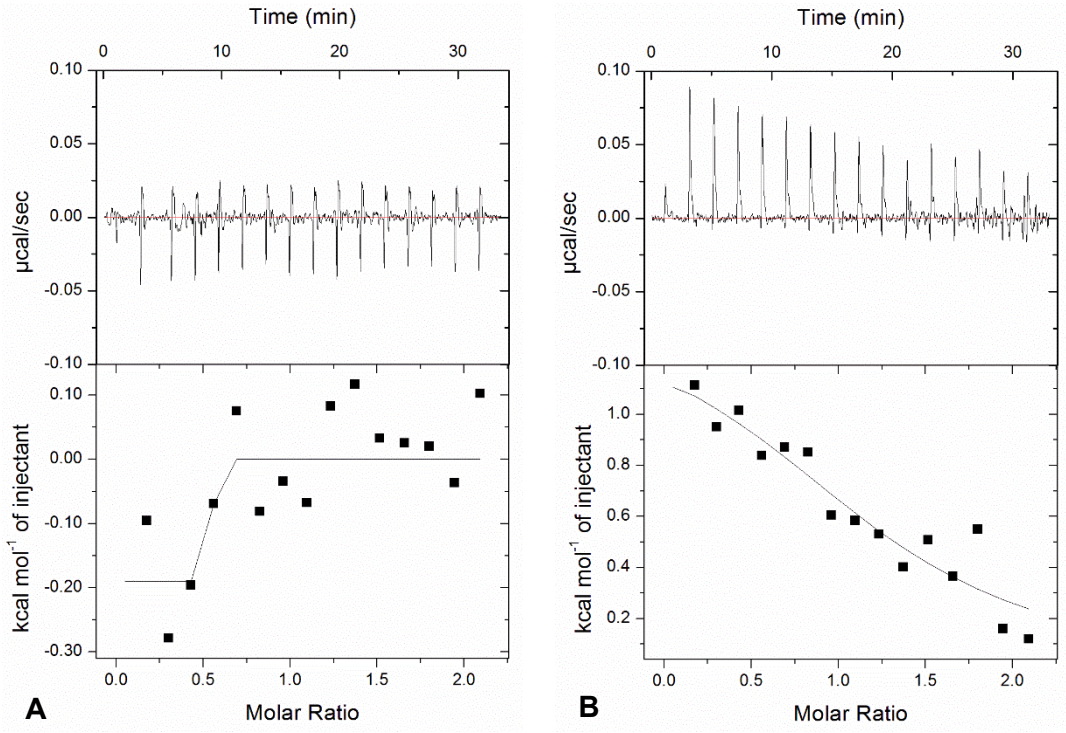


Figure 3.52: ITC results for PfSir2a (300µM) and PfAlba3 (30µM) at 4°C (A) and 15°C (B).

Figure 3.52 shows the results obtained for the ITC experiment performed at 4°C. The change in latent heat of the reaction over time (µcal s^{-1}) is extremely low and does not change with increasing amount of PfSir2a injected into the sample cell. This was repeated with the temperature increased to 15°C (see Figure 3.52 B). A small increase in the latent heat signal was observed at this temperature but a fit of the data could not determine a K_D.

3.10 Summary

In this chapter we have shown the ability to produce recombinant PfSir2a and PfAlba3 using the BL21 DE3 bacterial expression system. These recombinant proteins have been successfully purified by Ni-NTA, IEX and SEC methods. The purified proteins have undergone extensive optimisation and investigation via several biochemical and biophysical techniques in preparation for structural studies.

In these investigations, we find that PfSir2a is generally a stable protein that exhibits some concentration dependent oligomerisation effects. This is evidenced predominantly by the results for the AUC analysis seen in section 3.8.1. For PfAlba3 we find that it is a generally unstable protein with a tendency towards aggregation after purification, as evidenced by the results of the SEC purification and DLS experiments in sections 3.3.2 and 3.6.2 respectively. Despite an interaction between PfSir2a and PfAlba3 reported in the literature (Goyal et al., 2012), attempts to replicate using ITC and MST methods have proved challenging. An interaction was observed using MST that could indicate a tight binding between PfSir2a and PfAlba3 but was not observed using ITC. This might be due to a combination of the nature of the interaction, the limitations on the buffer conditions to keep the proteins stable and the differences between the two methods used.

In MST a local heating effect is created in a very small spot within the sample capillary causing molecules to diffuse out of the heated region before switching off the laser and allowing them to diffuse back. If the interaction between the proteins is ionic in nature, the high salt environment of the buffer used (500mM) will greatly interfere with the proteins ability to bind to one another. During this localised heating, the salt molecules will also diffuse out of the heated spot, creating a salt gradient. In this case, the salt concentration will be much lower than the overall buffer composition and will allow for the proteins to interact with less interference from the high salt content of the buffer. In ITC, the whole solution is heated which would result in the same concentration of proteins and salt within the sample cell, with no local gradients being created.

If an interaction between PfSir2a and PfAlba3 is indeed present, it is not currently possible to investigate by structural methods given the conditions that must be met for keeping the proteins stable and those required for performing solution scattering experiments. Whilst different behaviours are observed for both PfSir2a and PfAlba3 which present challenges

to performing structural investigations, the following chapter demonstrates several methods in which these problems can be tackled to obtain some information about the individual proteins.

Chapter 4: SAXS Characterisation Of PfSir2a And PfAlba3

Abstract

The aim of this chapter was to investigate the structural properties of PfSir2a and PfAlba3 using solution scattering based methods. This was achieved using samples that were produced and characterised as described in Chapter 3. Solution scattering experiments using X-rays were performed using the beamline BM-29 at the ESRF, Grenoble. The end results of the optimisation steps and techniques employed are described here.

From these experiments, we find that PfSir2a and PfAlba3 are challenging targets to characterise by solution based scattering methods. This is due to the different behaviours exhibited by the target proteins such as concentration dependent oligomerisation and rapid tendency towards aggregation for PfSir2a and PfAlba3 respectively. This was further complicated by PfSir2a and PfAlba3 being highly susceptible to radiation damage induced from exposure to X-rays.

Via the use of the on-line SEC system available at BM-29 and modified buffer conditions, we find an R_g of 2.9 ± 0.2 nm and 2.3 ± 0.3 nm for PfSir2a and PfAlba3 respectively. *Ab initio* modelling for PfSir2a and PfAlba3 was carried out using the data collected and indicates both proteins are elongated globular proteins with a D_{max} of 8.1 nm and 7.8 nm for PfSir2a and PfAlba3 respectively.

4.1 Introduction

This chapter focuses on the work and experiments carried out aimed at the characterisation of PfSir2a and PfAlba3 using small angle X-ray scattering. The initial focus was to

investigate the PfSir2a-PfAlba3 complex thought to exist via previous work (Goyal et al., 2012). The results obtained in chapter 3 have outlined the difficulties in the preparation of the samples for structural investigation. PfSir2a was observed to exhibit a concentration dependent oligomeric equilibrium and PfAlba3 and was found to aggregate rapidly following purification. These properties make the formation of a strong, stable complex difficult and affect the feasibility of structural studies on both the individual proteins and the complex. In particular they compromise the possibility of obtaining high resolution data using crystallographic methods.

In parallel to protein crystallisation attempts, which unfortunately did not yield any viable crystals, solution scattering using small-angle X-ray scattering (SAXS) was employed as a strategy for structural work. This approach can provide low-resolution structural information on the shape of the protein in solution. Protein samples are prepared as a monodisperse distribution in a suitable buffer environment for the stability of the protein sample. The sample is then exposed to X-ray radiation to record the total scattering pattern of protein and solvent. Post-data-processing is performed by subtraction of the solvent scattering from that of the sample scattering, yielding the scattering information relating to the protein alone.

4.2 Experimental Setup

Small angle X-ray experiments were conducted at the BioSAXS beamline BM-29 at the ESRF, Grenoble (Pernot et al., 2013). The beamline is situated such that the incoming X-rays from the synchrotron ring are directed to the instrument from a bending magnet (see Figure 4.1). The X-rays are conditioned by a double multilayer monochromator to select desired energy for the incoming beam within the 7-15keV operation range of the beamline. The incoming beam then enters a pair of monochromatic slits and then a 4 mrad toroidal

mirror (1.1m length) for focusing. Exposure to the incoming beam is regulated via control of a fast shutter and passes through an additional set of beam cleaning slits before arriving to the sample. The samples flow through a quartz glass capillary (1.8 mm diameter) to spread the dose of the incoming X-rays and reduce the visible effects of radiation damage.

The incident flux of the X-ray beam at the sample is in the region of 10^{12} photons s^{-1} with a beam size of approximately 700µm x 700µm. The scattering pattern produced by the sample is recorded on a Pilatus 1M detector with a sample to detector distance of 2.867m. The wavelength of the incident X-ray beam was 0.99 Å, this allows for the measurement of the sample over a q-range of 0.025 – 5 nm^{-1}. The total distance from the bending magnet source to the detector is 45m.

This instrument features a robotic sample changer that allows samples to be automatically introduced into the data collection system without any need for direct user intervention in the hutch (Round et al., 2015). For the automated sample changer, samples were prepared by Ni-NTA and SEC purification as mentioned previously (see Chapter 2, sections 2.1.10 – 2.1.14), concentrated with Amicon concentrators and then filtered by centrifugation with Spin-X filters to remove large particles and aggregates. A concentration series of both PfSir2a and PfAlba3 was prepared to rule out or evaluate the effects of concentration dependent inter-particle interaction within the samples which can affect the processed data. Sample volumes of between 20 -100µl were placed in 250µl Eppendorf tubes and maintained at a temperature of 4°C throughout the experiment. A flow rate of between 0.12 and 0.6 ml min^{-1} was used depending on the total volume of the sample being measured. 10 or 20 measurement frames were recorded for each sample with an exposure time of 1s or 0.5s respectively.

The buffer used for initial sample preparation contained 50mM sodium phosphate, 500mM NaCl, pH 7.0, and was filtered using a 0.22µm filter (Corning filter brand) and de-gassed for 30 minutes. Measurements of the buffer alone were collected before and after each sample data collection to verify the cleanliness of the cell and ensure the most reliable measure of the background scattering is measured that must be subtracted from the sample measurements.

For the on-line SEC HPLC system a sample volume of 50µl was prepared for injection onto a pre-equilibrated *Superdex*™ S200 10/300 GL column set at a flow rate of 0.5 ml min^{-1} (Round et al., 2013). The measurements were recorded every 1s over the entire duration of the column elution volume (24ml) - producing ~2400 frames of data.

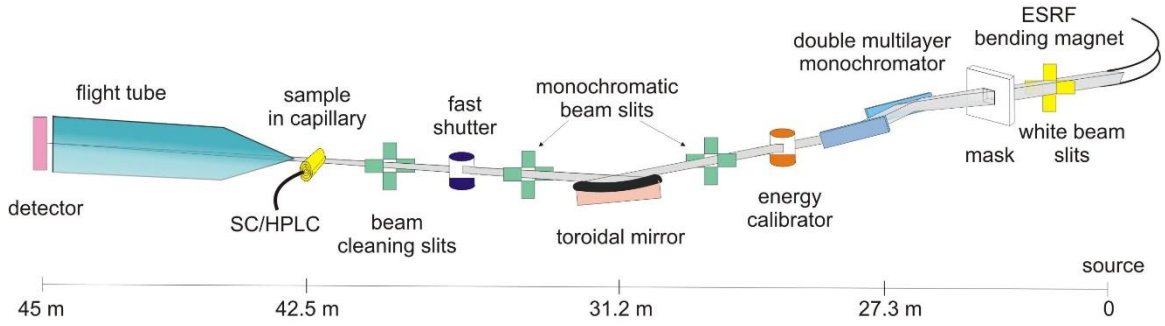


Figure 4.1: Schematic diagram for the layout of BM29 BioSAXS beamline at the ESRF, Grenoble (Pernot et al., 2013).

After the sample has been measured, the raw data is then processed through a data pipeline for reduction. This includes the azimuthal averaging of each individual frame, the creation of a background subtracted scattering curve (*i.e.* subtraction and normalisation against sample concentrations, basic analysis of the scattering curve to provide standard plots and invariants as well as *ab initio* modelling based on basic analysis (Brennich et al., 2016).

4.3 Data Processing

Data processing was carried out using the ATSAS program package for small angle scattering data analysis (Petoukhov et al., 2012). Buffer and sample frames were manually inspected for radiation damage before averaging the data to provide the best signal-to-noise ratio using PRIMUS (Konarev et al., 2003). The buffer scattering frames were then subtracted from the sample frames to yield only the excess scattering relating to the proteins present in the solution. The sample scattering curves were then corrected for concentration and subjected to post-processing analysis. This enables the calculation of some basic structural properties of the particles being studied. This includes the radius of gyration (R_g) refers to distribution of the components of an object around an axis. This is determined by subjecting the data to Guinier analysis at low q (where the Guinier region is $q < 1.3/R_g$). The maximum dimension (D_{max}) and molecular mass (MM) of a particle can also be determined.

The program GNOM (Svergun, 1992) was used to define to D_{max} of the particles from which the P(r) function was calculated. Modelling of treated sample curves was performed using the DAMMIF (Franke and Svergun, 2009) software for ensemble model generation. The various models produced were compared against each other in pairs (for all possible combinations) and the most probable selected using DAMSEL (Volkov and Svergun, 2003). This is calculated by minimising a normalised special discrepancy (NSD) value for

distances between points in each set of pairs and discarding models whose NSD is outside of 2 standard deviations from the mean. Selected models were then aligned and superimposed on one another using DAMSUP before averaging with DAMAVER to calculate a probability map across all the generated models (Volkov and Svergun, 2003). This creates an averaged model from all of the generated structures and a probability map is then used by DAMFILT to create a filtered model with the most likely shape. This is typically the model with the lowest NSD value.

4.3 I-TASSER *In silicio* Modelling And Existing Information

Before proceeding with small angle scattering experiments, some *a priori* information can be deduced about the structural characteristics of PfSir2a and PfAlba3. This is achieved by *in silico* modelling either based on existing crystal structures, or determined by primary sequence information and structural identity against characterised protein structures and folds deposited in the Protein Data Bank (PDB). By using this information, the theoretical small-angle scattering curves for PfSir2a and PfAlba3 can be calculated and compared with real experimental data. The modelling contained within this section was performed using the I-TASSER server (Roy et al., 2010; Yang et al., 2015; Zhang, 2008).

4.3.1 Preliminary Structural Characterisation Of PfSir2a

For a first look at the structural information that can be derived theoretically for PfSir2a, an approximate R_g can be calculated using the following approximation for globular proteins (Hong and Lei, 2007):

$$R_g \text{ (Å)} \sim 6.25 \times M^{1/3}$$

Equation 4.1

where M is the molecular mass in kDa. As R_g is highly dependent on shape, values for an elongated/unfolded proteins or tightly compacted globular protein will have be higher or lower respectively and so this provides an estimation of the expected R_g only. Using the native molecular mass for PfSir2a (30.43 kDa) this gives an approximate R_g of 20 Å (2 nm). We can also use existing structural information for PfSir2a to evaluate an expected R_g value. Two existing crystal structures are already present in the PDB (entries 3U31 and 3JWP). These structures related to a monomeric and trimeric form of PfSir2a as seen in the unit cell of the crystal lattice (see Figure 4.2).

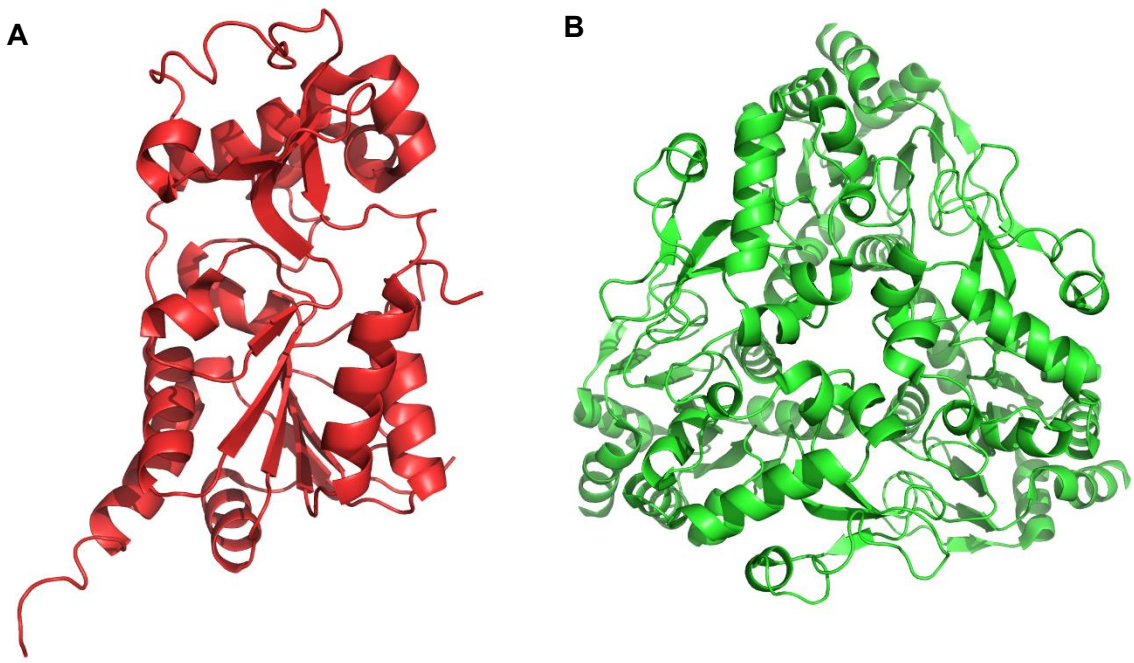


Figure 4.2: Pre-existing crystal structures of PfSir2a (not to scale). (A) Monomeric PfSir2a (PDB ID 3U31). (B) Trimeric PfSir2a (PDB ID 3JWP).

By using the program, CRYSOL we are able to calculate theoretical scattering profiles for both of these entries (Svergun et al., 1995). This program uses the known atomic structure of a protein from crystallography experiments to evaluate its solution scattering curve. This is performed using multipole expansion of the scattering amplitudes to calculate the spherically averaged scattering pattern which also takes into account the hydration shell surrounding the protein in solution (Svergun et al., 1995).

This yields a R_g of 20 Å (2 nm) and 28 Å (2.8 nm) for the monomer and trimer respectively. The estimated R_g of 2nm for the monomeric PfSir2a calculated previously using Equation 4.1 matches that of the theoretical calculation as determined by CRYSOL. By using the same approximation, we can calculate a R_g of 28 Å for the trimeric form of PfSir2a which is also similar to that of the calculated R_g from CRYSOL. These calculations compare to that of the native form of PfSir2a.

For the SAXS experiments carried out in this thesis the full length recombinant form of PfSir2a was used with a mass of 33.88 kDa. This has a higher MM than the native PfSir2a due to the presence of the His-tag and linker amino acids in the construct design. Using this mass value, the estimated R_g for the monomer and trimer form can be calculated at 20 Å and 29 Å respectively. As we have seen from previous chapters, a dimeric state of PfSir2a is believed to exist in a concentration dependent manner in solution. As there is no crystal structure information for a PfSir2a dimer an estimation of the R_g using the mass of the full length recombinant form was calculated as 25 Å (2.5 nm).

Additionally, *in silico* modelling using the I-TASSER server generated several approximations/anticipated structures of the full-length recombinant PfSir2a which can also be used by CRYSOL to calculate a theoretical R_g. This was performed using the PDB result files generated by inputting the amino acid sequence into the I-TASSER server. The

top 5 most probable models are selected based on a scoring system and supplied to CRYSOL. The results returned for the R_g of these models are given in Table 4.1.

PfSir2a Model Or Crystal Structure	CRYSOL R_g (Å)
3U31	20
3JWP	28
Model 1	24
Model 2	24
Model 3	24
Model 4	23
Model 5	22

Table 4.1: Table of theoretically calculated R_g values by CRYSOL for the existing crystal structures of PfSir2a and the I-TASSER generated models.

The results of these theoretically calculated values were used to provide useful starting information relevant to the overall size and oligomeric state of PfSir2a prior to the SAXS solution measurements. From these values, we expect a monomer of the recombinant PfSir2a with to have an R_g in the region of 22 – 24 Å. The dimer is estimated to have an R_g of approximately 25 Å and that for a trimer to be 29 Å.

Figure 4.3 shows an alignment of a single molecule of PfSir2a as seen in the PDB entries 3U3I and 3JWP with I-TASSER generated Model 1 and Model 5. These I-TASSER models represent the overall structure of PfSir2a construct used in this thesis with the largest and smallest calculated R_g using CRYSOL. In general, the structures appear to align to one another with a good agreement. An interesting feature is the positioning of the additional N-terminal sequence of 17 amino acids in the recombinant construct. This position of this with respect to the core PfSir2a features changes dramatically between the I-TASSER generated models and indicates that there may be a high degree of flexibility of this additional linker region which could result in larger R_g values observed in small angle

scattering experiments. This high degree of flexibility may also have interfered with crystallisation attempts by being too flexible and an inability to form crystal contacts.

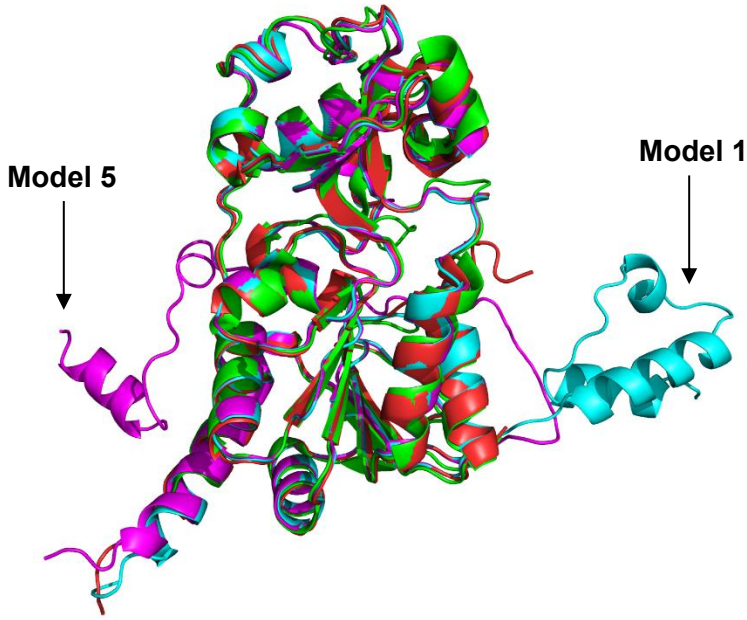


Figure 4.3: Alignment of individual molecules of PfSir2a from existing structures 3U3I (Red), 3JWP (Green), I-TASSER Model 1 (Cyan) and I-TASSER Model 5 (Magenta). The core domain positions appear to be well conserved but the additional N-terminal region used in the PfSir2a recombinant construct appears to be disordered as observed in the positions highlighted for Model 1 and Model 5 respectively.

4.3.2 Preliminary Structural Characterisation Of PfAlba3

As with PfSir2a, we can calculate theoretical structural information for PfAlba3 which can help to interpret the experimental data obtained from solution SAXS. The full length recombinant form of PfAlba3 has a mass of 13.04 kDa which gives an estimated R_g of 15 Å. There is currently no pre-existing crystal structure information for PfAlba3 from which a theoretically calculated R_g can be generated by CRYSOL. As such, only I-TASSER models for the full length recombinant form of PfAlba3 could be used. The results of the CRYSOL runs gave R_g values for these models that are summarised in Table 4.2.

PfAlba3 Model	CRYSOL R_g (Å)
Model 1	16
Model 2	16
Model 3	17
Model 4	16
Model 5	16

Table 4.2: Table of theoretically calculated R_g values by CRYSOL for the I-TASSER generated models of PfAlba3

4.4 PfSir2a SAXS Experimental Results

This section outlines the main results observed from SAXS experiments using purified PfSir2a samples.

4.4.1 Radiation Damage Characterisation

SAXS measurements of PfSir2a were carried out using the automated sample changer system available at BM-29 (see section 4.2). An initial concentration series was carried out over a sample concentration range of 0.5 mg ml^{-1} to 8mg ml^{-1}. Initial analysis of the data indicates that PfSir2a was strongly affected by radiation damage. Figure 4.4 shows the individual frames collected for a PfSir2a sample at a concentration of 8 mg ml^{-1}. At low q ($q < 0.2$ nm^{-1}) we see a shift in the measured intensity with each successive recorded frame, this is a characteristic feature of radiation damage and an ideal sample with no radiation damage would have frames which lie almost directly on top of one another which can be averaged together.

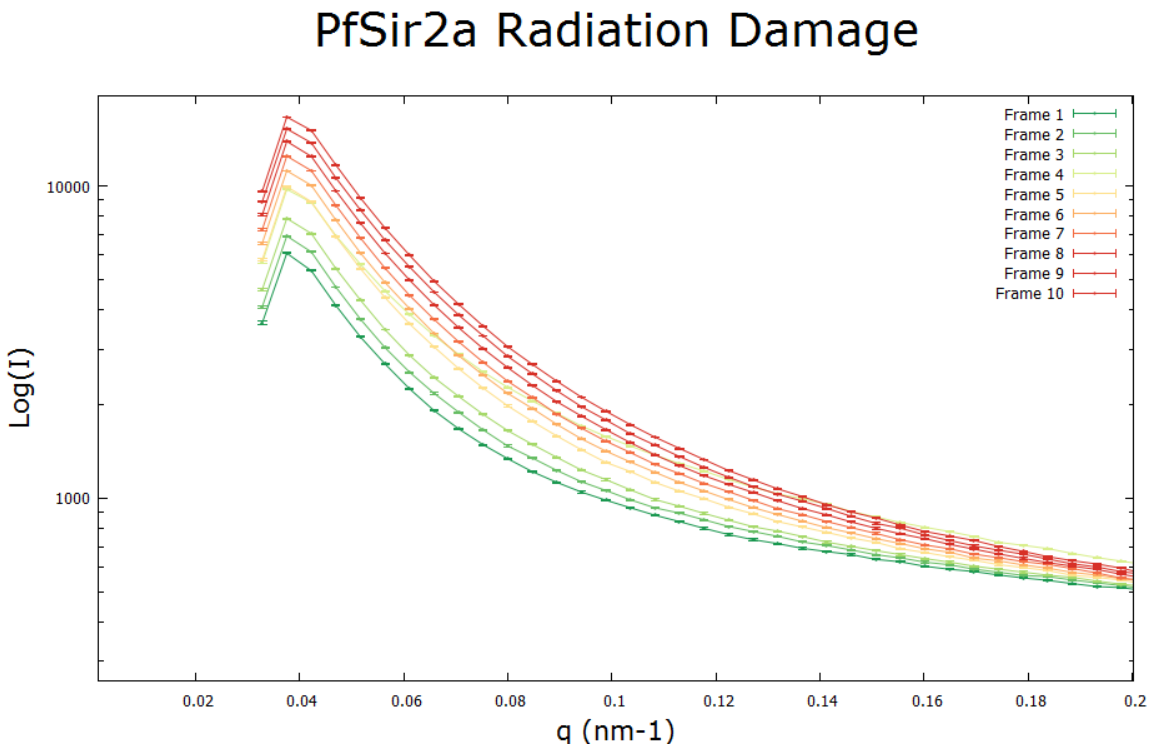


Figure 4.4: Plot of Intensity (I) versus q (nm^{-1}) of ten PfSir2a 8 mg ml^{-1} sample frames with a 1s exposure time on a log scale. A shift in intensity at low q with successive recorded frames is indicative of radiation damage occurring in the sample due to continued exposure to X-rays. Sample frames are coloured from green to red to highlight the progression from non-damaged to radiation damaged respectively.

Radiation damage is a well-known problem encountered during SAXS experiments, particularly when biological samples are studied using high brilliance synchrotron radiation sources such as the ESRF. It is a result of the absorption of high energy X-rays by water molecules present in the sample solution. These water molecules absorb the incoming energy of the X-ray photons and generate hydroxyl or hydroperoxyl radicals that interact with the protein molecules present in the solution to form charge centres which lead to electrostatic attraction and thus aggregation. The aggregated molecules which form

have a significantly increased molecular mass compared to that of the original protein molecules (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013).

The distribution of the aggregated particles also covers a large range and the sample can no longer be considered a homogenous solution. This has a severe effect on further data analysis, causing inaccurate values to be derived for model-independent features of the target protein, such as the average molecular mass and size. This effect typically occurs in the last 2-3 recorded exposures of the sample which can be omitted during the analysis. In the case for PfSir2a it is clear that radiation damage occurs after the first exposure and continues during subsequent sample measurements. The effects of this can be observed in the measurements performed over the concentration series for PfSir2a in Figure 4.5.

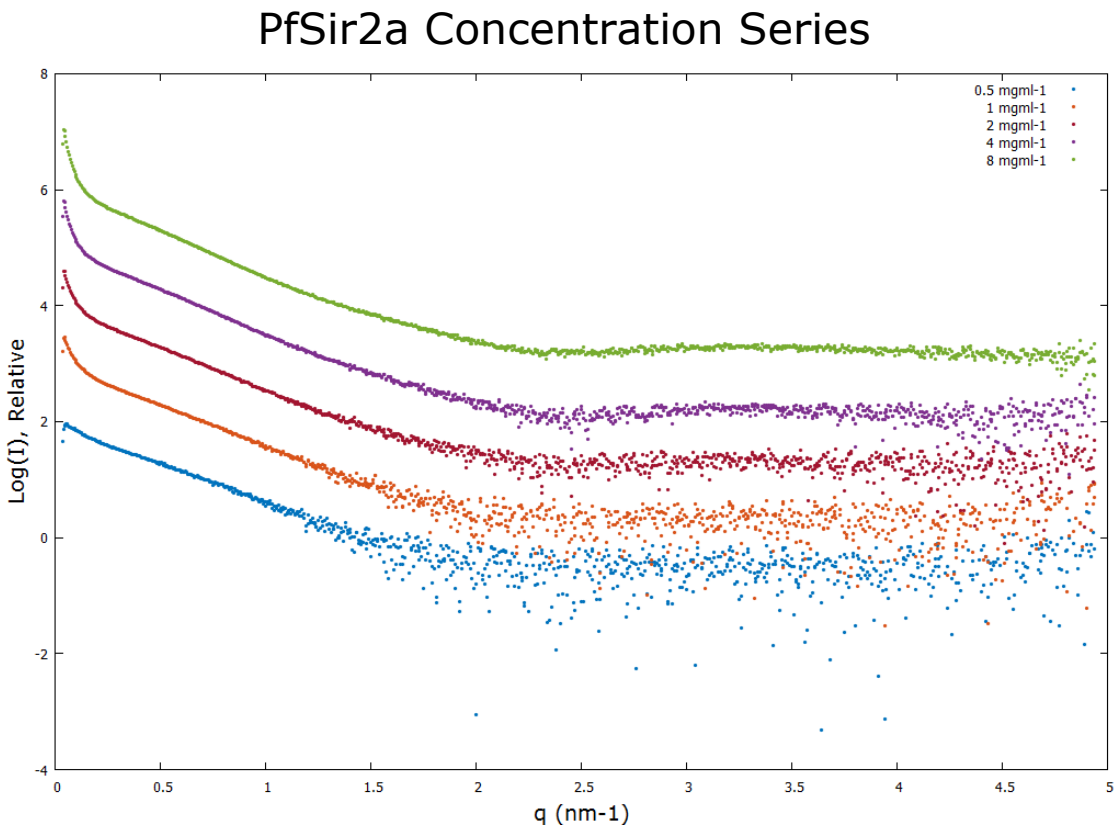


Figure 4.5: Individual processed scattering curves for PfSir2a concentration series measurements of Intensity versus q on a relative log scale (without errors) for better visualisation. A sharp increase in the measured intensity is observed at low q as the concentration of PfSir2a increases.

In this figure, the average scattering curves of PfSir2a over a sample concentration range of 0.5 - 8 mg ml^{-1} are shown. It can be seen in the concentrations of 1 mg ml^{-1} upwards, there is a distinct, sharp increase in the observed intensity at low q ($q < 0.2$ nm^{-1}). This is highly indicative of large protein aggregates being present in the sample. The effect is especially highlighted when normalising for concentration by dividing the buffer-subtracted scattering curves by the sample concentration in mg ml^{-1} (see Figure 4.6).

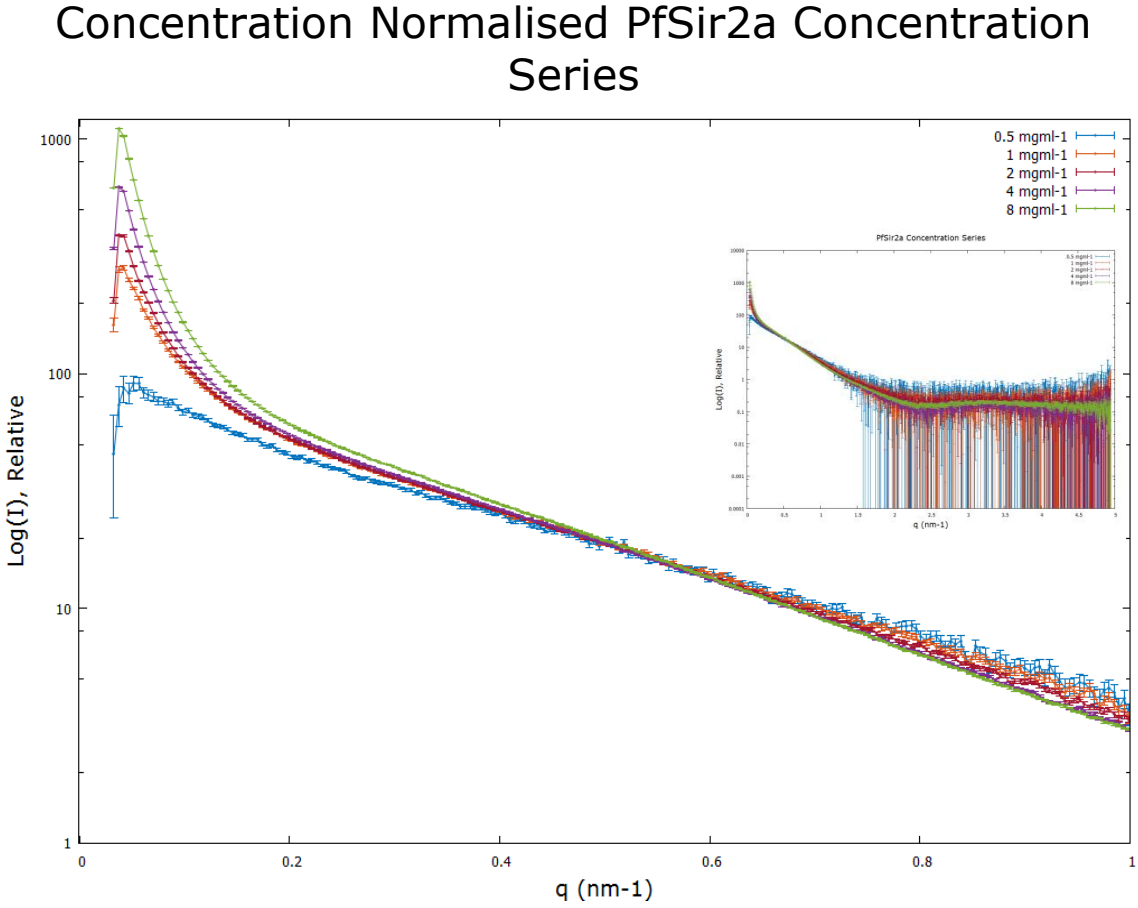


Figure 4.6: Concentration normalised scattering curves for PfSir2a concentration series from 0 to 1 q. A sharp increase in intensity in the low q region is observed across the concentration series. The total scattering curves can be seen inserted on the right for reference.

We can further investigate these results to confirm the effects of radiation damage by using the data to calculate some basic particle properties; the apparent MM and R_g. The initial slope of a scattering curve can be approximated by Guinier's law:

$$I(q) \approx I(0)e^{\frac{-q^2{R_g}^2}{3}}$$

Equation 4.2

where R_g is the radius of gyration and *I(0)* is the zero angle intensity or forward scattering of a particle. These properties can be determined by a plot of ln *I(q)* versus q^2 (Guinier plot) and holds valid for a range of $0 < q < 1/R_g$, known as the Guinier range. For analysis of globular proteins, the limits are usually slightly adjusted to be $q_{min} \leq 0.5/R_g$ and $q_{max} \leq 1.3/R_g$. This plot should be linear at low q and aggregation or intermolecular repulsion are characterised by sharp upswings or downswings in the data respectively (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013).

For MM determination (in kDa), the *I(0)* of each scattering curved is placed on an absolute scale by calibrating to water scattering, measured before each data set collection, using the following equation:

$$MM = \frac{N_A I_0}{{\Delta\rho_M}^2}$$

Equation 4.3

Where, N_A is Avogadro's constant, *I(0)* is the experimentally determined forward scattering intensity (divided by the concentration), $\Delta\rho_M = [\rho_{M,prot} - (\rho_{solv}\ \bar{v})]r_o$ is the scattering contrast per mass, $\rho_{M,prot} = 3.22 \times 10^{23}$ e g^{-1} is the number of electrons per mass of dry protein, $\rho_{solv} = 3.34 \times 10^{23}$ e cm^{-1} is the electron density of the aqueous solvent (water), $\bar{v} = 0.7425$ cm^3 g^{-1} is the partial specific volume of the protein and $r_o = 2.8179 \times 10^{-13}$ cm is the scattering length of an electron. The values determined for the R_g and MM of PfSir2a were determined using a q range of between 0.02 and 0.06 based on the CRYSOL calculated R_g

values and are listed in Table 4.3. As the sample buffer contains a high concentration of NaCl (0.5M) the electron density of the solvent was recalculated to 3.39 x 10^{23} e cm^{-1}.

Sample – Concentration (mg ml^{-1})	R_g (nm)	Error R_g ± (nm)	I(0)	Error I(0) ±	Absolute I(0)	Apparent Molecular Mass (kDa)
Water – N/A	-	-	2.1E+01	4.0E-03	-	-
PfSir2a - 0.5	5.52	0.089	6.67E+01	8.83E-01	5.18E-02	76.51
PfSir2a - 1	7.11	0.076	1.01E+02	1.08E+00	7.84E-02	115.85
PfSir2a - 2	7.39	0.057	1.05E+02	8.18E-01	8.15E-02	120.44
PfSir2a - 4	7.67	0.03	1.15E+02	4.89E-01	8.93E-02	131.91
PfSir2a - 8	8.25	0.016	1.43E+02	3.52E-01	1.11E-01	164.02

Table 4.3: Experimentally determined value for R_g and MM of PfSir2a from processed concentration series data. Values highlighted in red could not be automatically determined using PRIMUS software and were determined manually.

The results in Table 4.3 confirm that the PfSir2a samples experienced heavy radiation damage due to exposure to X-rays. This is seen as an increase in the R_g and molecular mass of the proteins with increasing sample concentration. Higher sample concentrations increase the likelihood of free-radical-induced aggregation and the formation of larger proteins aggregates.

When attempting to fit the data, the only results found to satisfy the Guinier approximation between $q_{min} \leq 0.5/R_g$ and $q_{max} \leq 1.3/R_g$. during analysis was the 0.5 mg ml^{-1} sample. The remaining samples highlighted in red could not be automatically determined using the PRIMUS software and were interpreted manually. The Guinier plot and resulting fit of the 0.5 mg ml^{-1} sample can be seen in Figure 4.7. Despite satisfying these conditions we can see that the Guinier plot is not a linear trend as it tends to low q and we see an upswing in observed intensity. Given the information provided from this analysis, it is not possible to

provide an accurate *ab initio* model for PfSir2a from this data set as the results would be skewed by the effects of radiation-damage-induced protein aggregation.

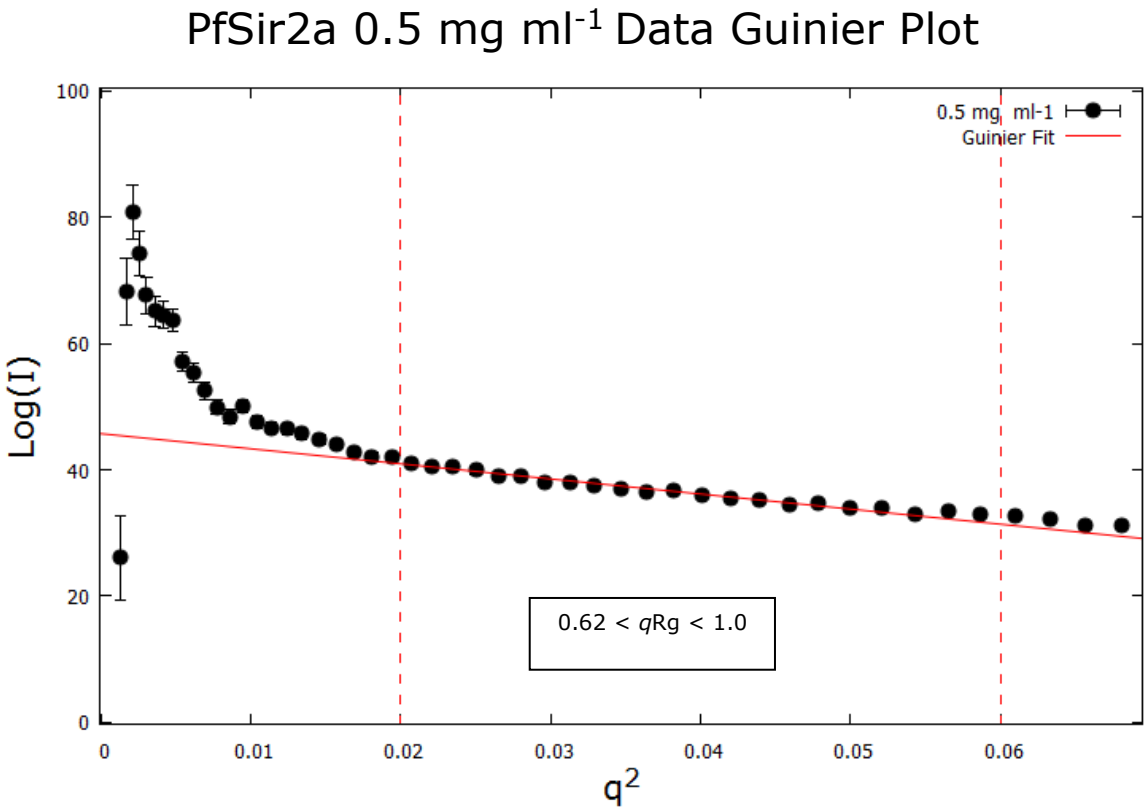


Figure 4.7: Guinier plot of Log(I) versus q^2 for PfSir2a sample at 0.5 mg ml^{-1}. The data points satisfying the Guinier approximation were plotted with a line of best fit. This plot illustrates the non-linearity of the PfSir2a sample in the low q Guinier region and the sharp upwards increase in intensity at low q (q < 0.2) that is indicative of radiation damage or aggregation.

To obtain accurate *ab inito* shape information about PfSir2a in solution, it is necessary to counter the effects of the radiation damage. This may be mitigated in several ways. The first is to reduce the salt content of the buffer in which the sample is measured. High salt concentrations of salt increase the absorbance of the incoming X-ray energy and increase the number of free radicals generated (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013). The addition of free-radical-scavenging reducing agents, such as DTT or

ß-mercaptoethanol to the sample buffer can also be employed and acts as a barrier by soaking up generated free radicals before they can cause damage to the protein molecules.

Initially, the reduction of the salt concentration in the sample buffer in which PfSir2a is prepared was investigated. The NaCl concentration of the buffer was reduced from 0.5M to 0.1M, and the concentration series was prepared again for a repeated SAXS experiment. The results from this experiment can be seen in Figure 4.8 and indicate that there may still be radiation damage occurring from sample exposure to X-rays, signified again by the sharp upturn in I at low q.

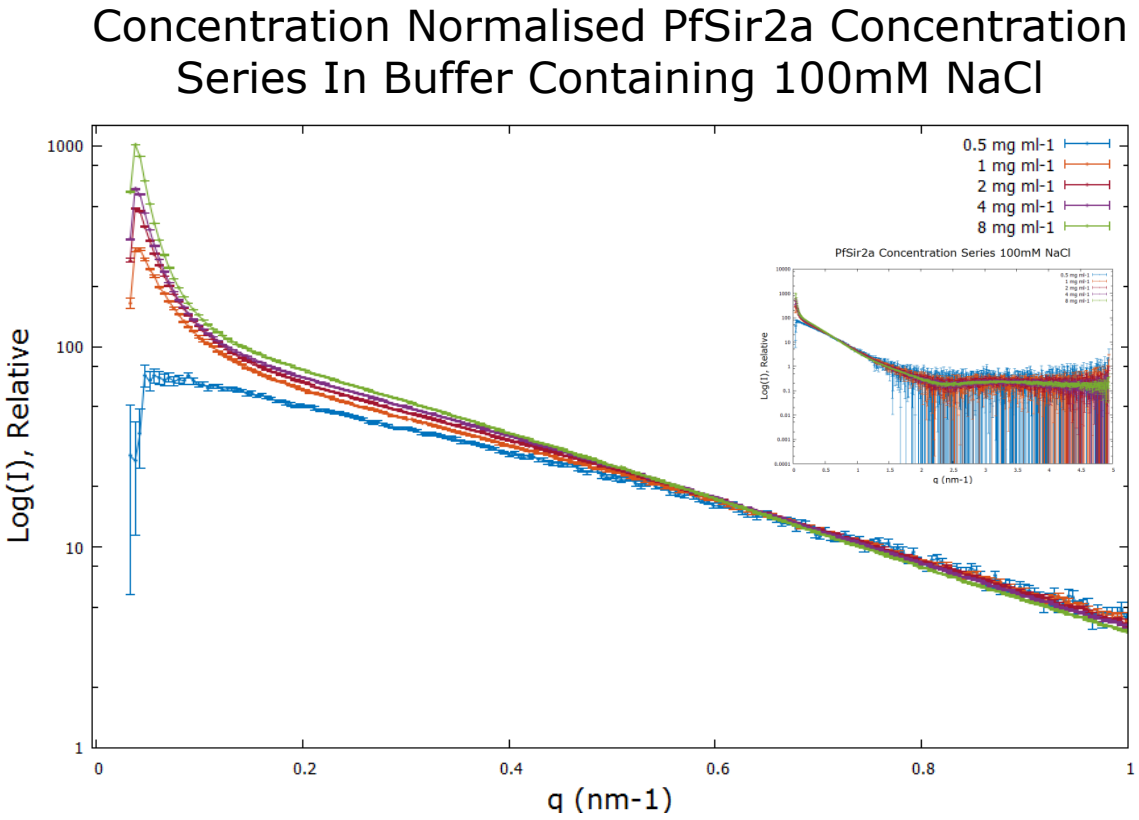


Figure 4.8: Concentration normalised scattering curves for PfSir2a concentration series in buffer containing 100mM NaCl from 0 to 1 q. A sharp increase in intensity in the low q (q < 0.2 nm^{-1}) region is observed across the concentration series. The total scattering curves can be seen inserted on the right for reference.

The shifting of intensity across the individual sample frames was also observed during manual analysis. This is highlighted even in the lowest concentration of PfSir2a measured, 0.5 mg ml^{-1} as can be seen in Figure 4.9.

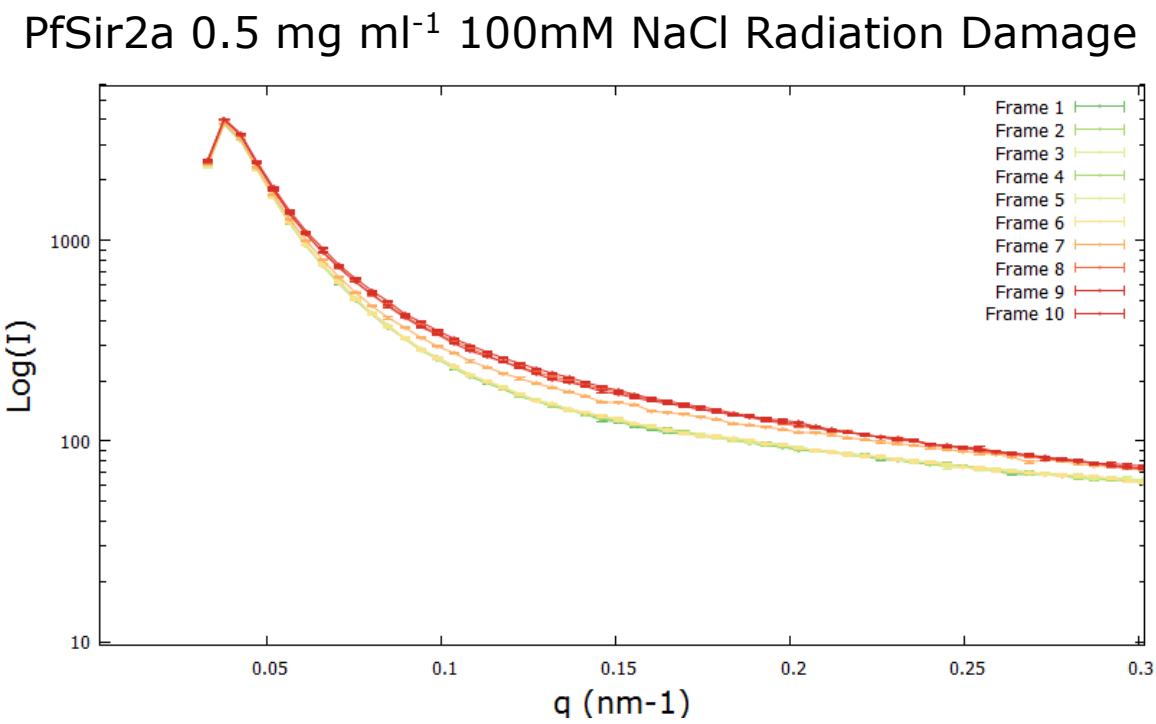


Figure 4.9: Plot of Intensity (I) versus q (nm^{-1}) of ten PfSir2a 0.5 mg ml^{-1} sample frames with a 1s exposure time on a log scale. A shift in intensity at low q (q < 0.3 nm^{-1}) with successive recorded frames is indicative of radiation damage occurring in the sample due to continued exposure to X-rays. Sample frames are coloured from green to red to highlight the progression from non-damaged to radiation damaged respectively.

The experimentally determined values for the R_g and MM of PfSir2a in this lower salt concentration buffer can be seen in Table 4.4. A modified solvent electron density value of 0.335 was used to calculate these results to account for the electron density of water plus that of 100mM NaCl. The flow rate of the sample through the capillary was also increased to 0.6 ml min^{-1}. The results show that there is still evidence of radiation damaged induced

aggregation from the increasing R_g and MM values. Unlike the previous data set, the Guinier approximation held valid for the Guinier range determined by the estimated R_g for each of the scattering curves analysed by the PRIMUS software.

Sample – Concentration (mg ml^{-1})	R_g (nm)	Error R_g ±	I(0)	Error I(0) ±	Absolute I(0)	Molecular Mass (kDa)
Water – N/A	-	-	2.1E+01	4.0E-03	-	-
PfSir2a - 0.5	4.42	1.14E-01	6.61E+01	9.32E-01	5.13E-02	68.40

Table 4.4: Experimentally determined value for R_g and MM of PfSir2a in 100mM NaCl containing buffer at 0.5 mg ml^{-1}.

The value for the R_g and MM is lower than that determined from the previous dataset in Table 4.3. This indicates that although there is some radiation damage still present, it is occurring to a lesser extent than was observed in the initial measurements. This can be seen in the Guinier plot of the 0.5 mg ml^{-1} sample seen in Figure 4.10. Here the Guinier plot is much more linear than that for the 0.5 mg ml^{-1} sample from the previous dataset. However, despite a slight reduction in radiation damage, the Guinier plots for the higher concentration samples were less linear. The previously calculated values for the R_g and MM of PfSir2a also do not correlate with the expected value for a PfSir2a monomer in solution.

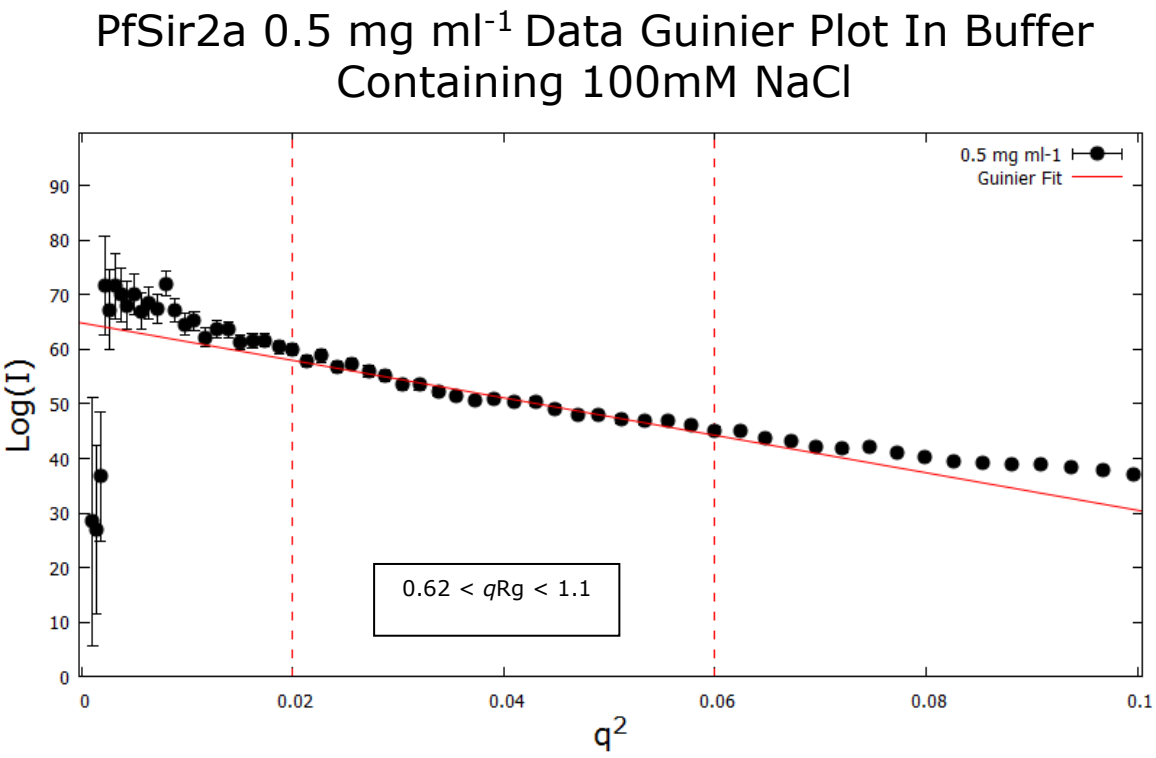


Figure 4.10: Guinier plot of Log(I) versus q^2 for PfSir2a sample at 0.5 mg ml^{-1} in 100mM NaCl buffer. The data points satisfying the Guinier approximation were plotted with a line of best fit. This plot illustrates an improvement to the Guinier plot and data fitting compared to that seen in Figure 4.5 but still shows some non-linearity at low q (q < 0.2 nm^{-1}).

Additional steps to further minimise radiation damage effects were attempted by the addition of 1mM DTT to the sample buffer containing 100mM NaCl to act as a free radical scavenger during the course of the experimental measurements for the PfSir2a concentration series. The results for this data set can be seen in Figure 4.11. A reduction is seen in observed I at low q. Additionally, there is little variation between individual measurements frames compared to that seen previously in Figure 4.9

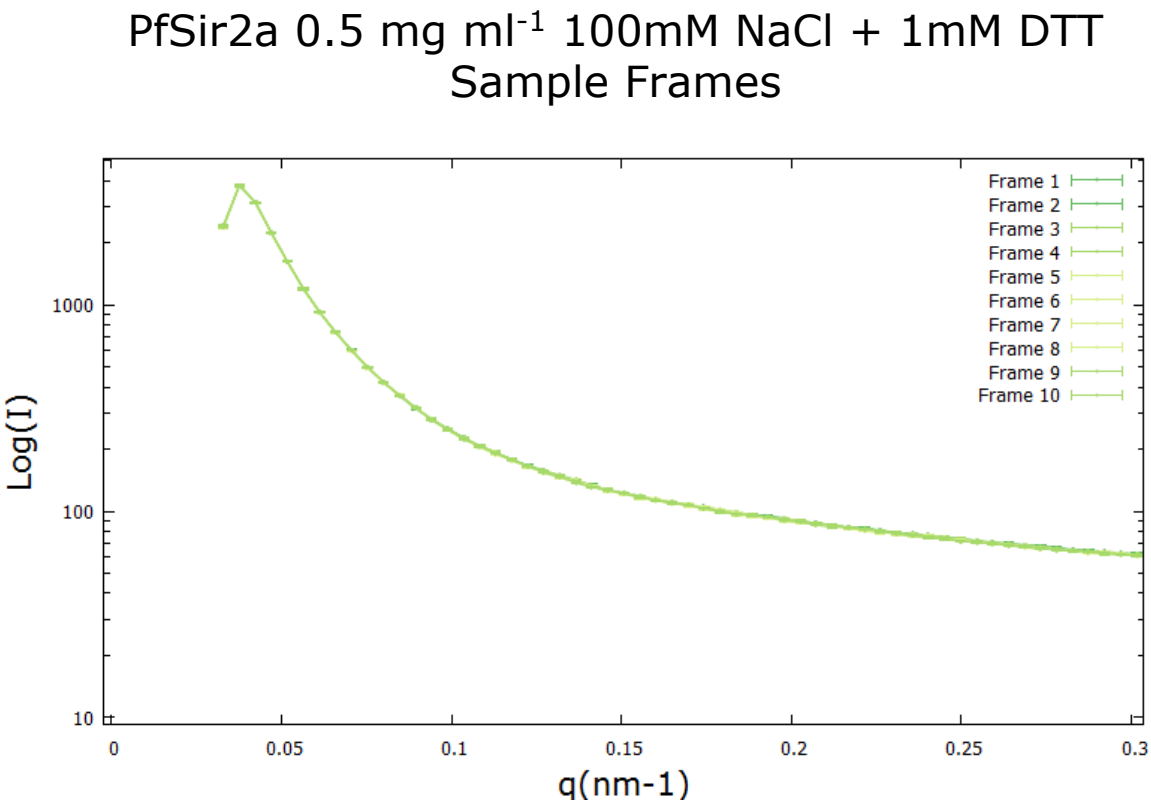


Figure 4.11: Processed scattering curves for PfSir2a concentration series measurements in buffer containing 100mM NaCl and 1mM DTT. Intensity versus *q* is plotted on a relative log scale (without errors) for better visualisation. A reduced upswing in intensity is observed at low *q* due to addition of 1mM DTT to the sample buffer.

The difference in scattering for several concentrations of PfSir2a in buffers containing 100mM NaCl with and without 1mM DTT is especially apparent when comparing the data sets. This is shown in Figure 4.12.

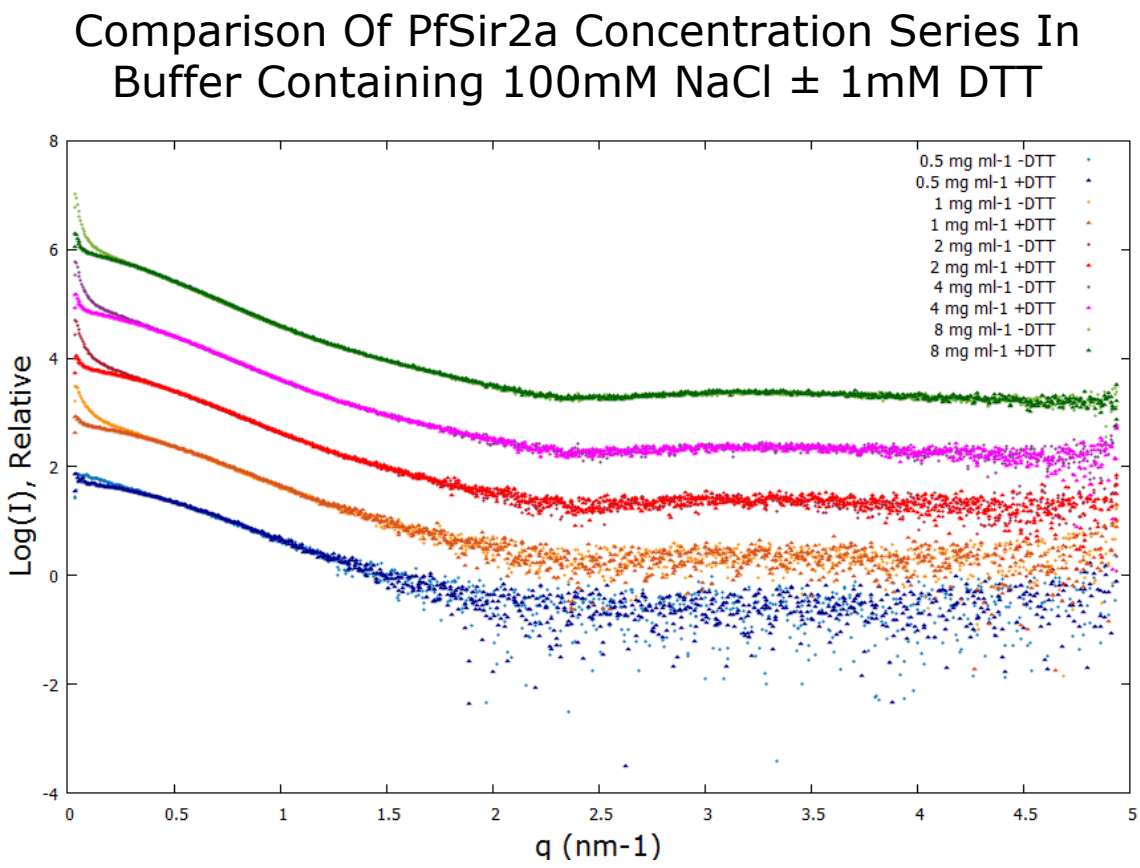


Figure 4.12: Processed scattering curves for PfSir2a concentration series measurements in buffer containing 100mM NaCl ± 1mM DTT. Intensity versus *q* is plotted on a relative log. This plot illustrates the reduction in observed low *q* intensity in the processed sample scattering curves due to addition of 1mM DTT to the sample buffer.

The results shown in Figure 4.12 indicate a significant reduction in radiation damage to the sample concentration range tested from the addition of 1mM DTT to the sample buffer. An improvement is also noticed when normalising each scattering curve against concentration as seen in Figure 4.13. Here there is a much less pronounced increase in intensity in the low *q* region although there is still some discrepancy between the different sample concentration curves.

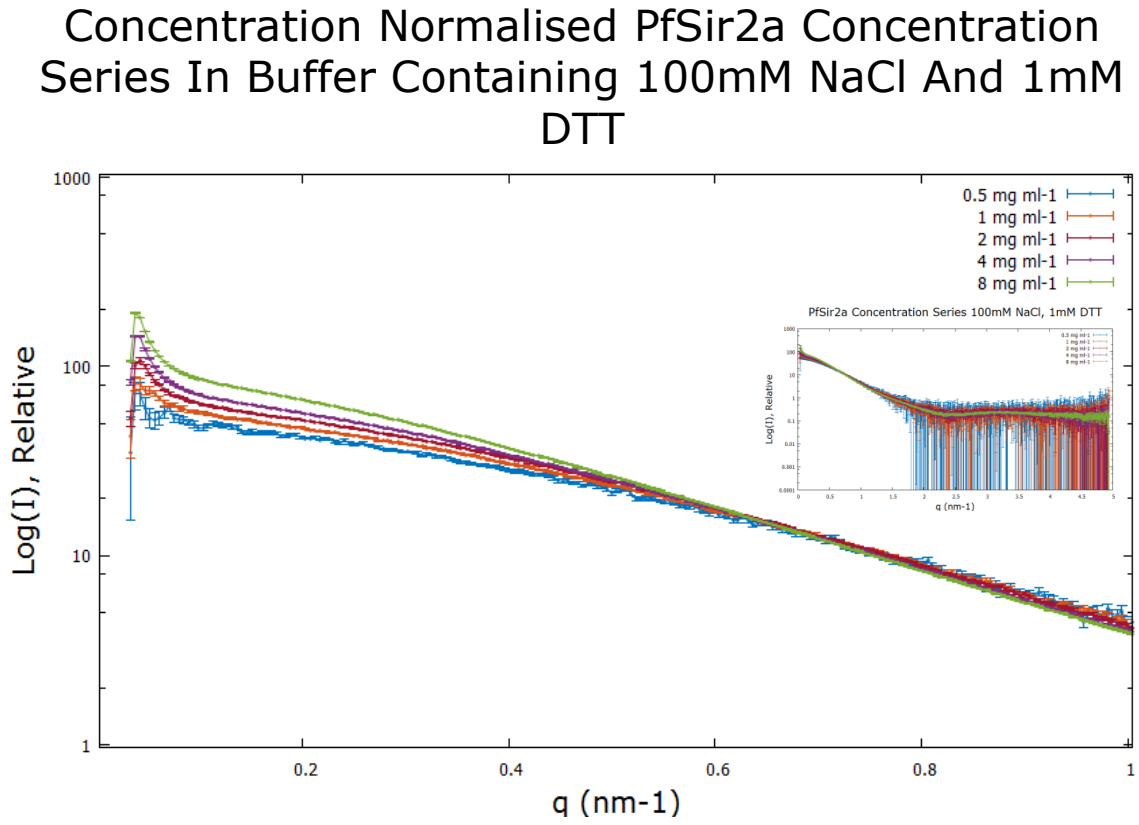


Figure 4.13: Concentration normalised scattering curves for PfSir2a concentration series in buffer containing 100mM NaCl and 1mM DTT from 0 to 1 q. The total scattering curves can be seen inserted on the right for reference.

The experimentally determined values for the R_g and MM of PfSir2a in 100mM NaCl buffer with 1mM DTT were calculated and can be seen in Table 4.5. A significant reduction in the overall R_g calculated for the entire concentration series can be seen. An improvement in the fit of the Guinier plot for the 0.5 mg ml^{-1} sample is also found and is now almost completely linear at low q (see Figure 4.14).

Sample – Concentration (mg ml^{-1})	R_g (nm)	Error R_g ± (nm)	I(0)	Error I(0) ±	Absolute I(0)	Molecular Mass (kDa)
Water – N/A	-	-	2.1E+01	4.0E-03	-	-
PfSir2a - 0.5	3.27	1.30E-01	4.88E+01	5.95E-01	**3.79E-02**	50.50
PfSir2a - 1	3.62	6.50E-02	5.62E+01	8.87E-01	**4.36E-02**	58.16
PfSir2a - 2	3.7	3.30E-02	6.24E+01	9.06E-01	**4.84E-02**	64.57
PfSir2a - 4	3.97	2.00E-02	7.01E+01	9.54E-01	**5.44E-02**	72.54
PfSir2a - 8	4.33	9.00E-03	8.57E+01	9.10E-02	**6.65E-02**	88.68

Table 4.5: Experimentally determined value for R_g and MM from processed concentration series data of PfSir2a in 100mM NaCl and 1mM DTT containing buffer.

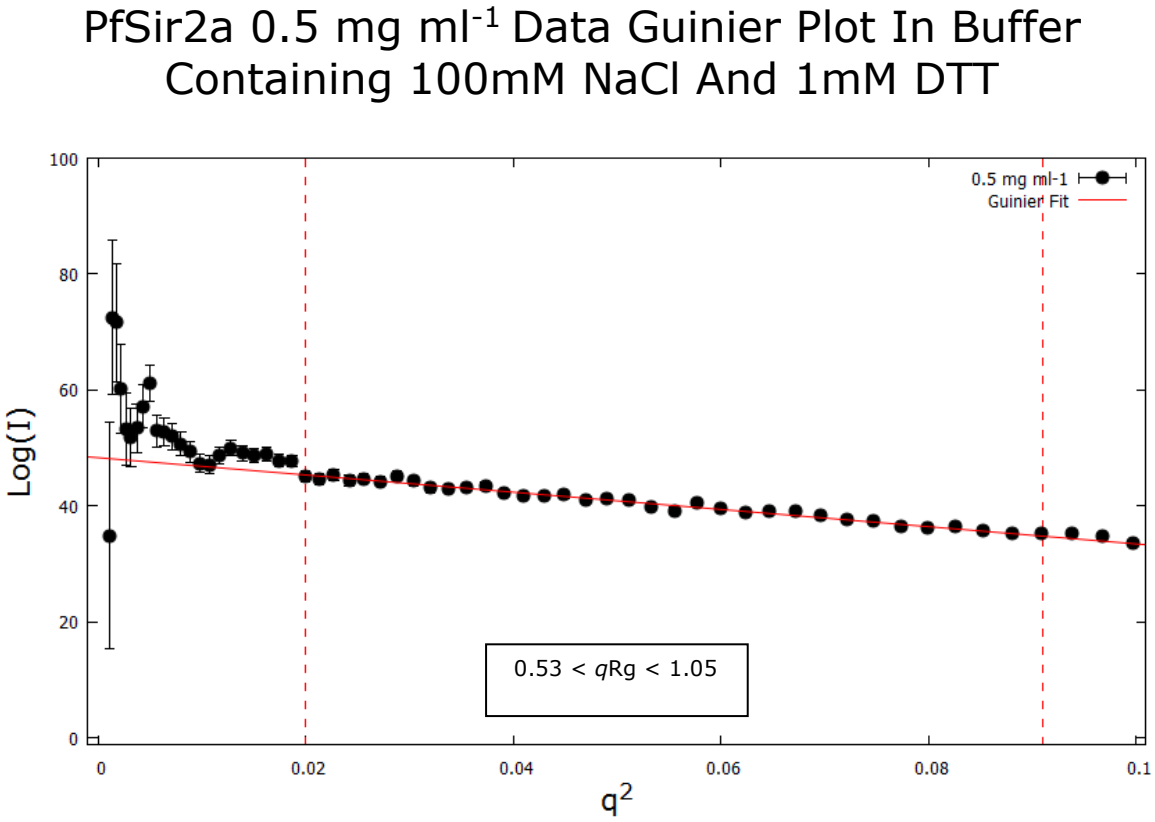


Figure 4.14 Guinier plot of Log(I) versus q^2 for PfSir2a sample at 0.5 mg ml^{-1} in 100mM NaCl, 1mM DTT buffer. The data points satisfying the Guinier approximation were plotted with a line of best fit. A significant improvement in the linearity of the Guinier plot is observed at low q ($q < 0.2$ nm^{-1}).

However, the molecular mass calculation for this concentration yields a value of 50.50 kDa which is still too large for a monomer of PfSir2a and too small for a dimer. A steady increase in R_g and MM is again observed as the concentration of the sample increases. The reason for this continued increase can be attributed to several factors; some small presence of radiation-damage-induced aggregation, sample polydispersity in the adjusted buffer conditions, or a sample concentration effect.

The continued presence of radiation damage was investigated first. As previously shown in Figure 4.4 for the 8 mg ml^{-1} sample from the initial experiment dataset, a plot of the individual sample frames can be seen in Figure 4.15. Here we see there is much less divergence between the individual frames as compared to the previous 8 mg ml^{-1} sample before buffer optimisation. This suggests the effect from radiation damage is minimal and can be mitigated when processing the sample frames. This is reflected in the Guinier plot of the 0.5 mg ml^{-1} concentration sample seen in Figure 4.14.

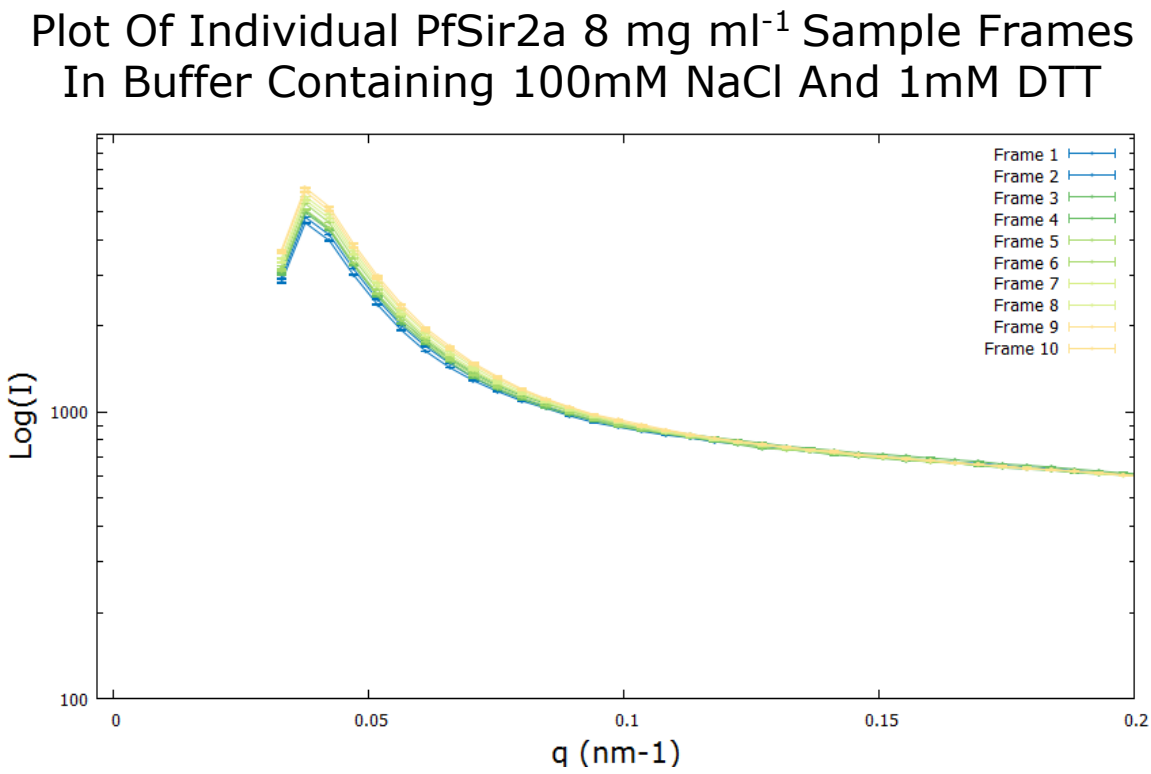


Figure 4.15: Plot of Intensity (*I*) versus q (nm^{-1}) of 10 PfSir2a 8 mg ml^{-1} sample frames with a 1s exposure time on a log scale. A vast improvement is observed in the scattering data with a minimal increase in observed intensity at low *q* observed with successive measurements.

Secondly, the effect of polydispersity was investigated by DLS measurements on samples prepared in the optimised buffer conditions. Previous DLS results seen in Chapter 3, section 3.6.1 indicates that lower salt concentration in the sample buffer increases the polydispersity of PfSir2a in solution. DLS measurements were therefore made with the addition of 1mM DTT to the sample buffer and a PfSir2a sample concentration of 1 mg ml^{-1}; the results of this can be seen in Figure 4.16.

Panel A shows the previous measurements made in sodium phosphate buffers containing 100 or 500mM NaCl without DTT as a comparison with R_H values of 8.1 ±2.2 nm (26%

poldispersity) and 9.3 ± 1.4 nm (15% polydispersity) in 100mM and 500mM NaCl respectively Panel B shows the repeated measurement with the addition of 1 mM DTT to the sample buffer. Values of 8.9 ± 2.5 nm (26% polydispersity) and 10.7 ± 1.6 nm (15% polydispersity) were found for the 100mM and 500mM NaCl-containing buffers respectively. This is in agreement with previous results that reducing the salt content of the buffer increases the overall polydispersity of PfSir2a in solution. The addition of 1mM DTT appears to have had very little effect on the polydispersity of the sample.

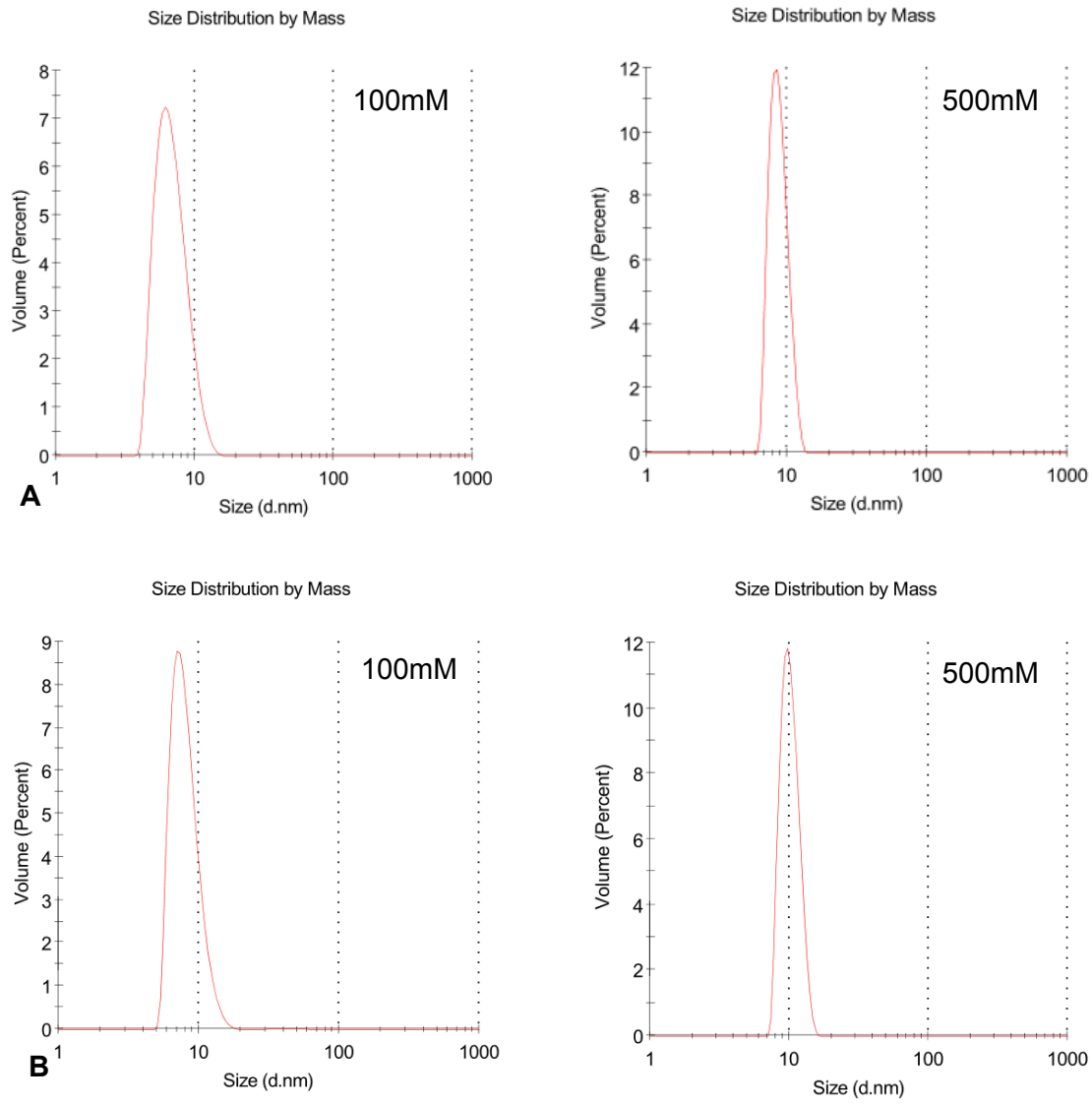


Figure 4.16: DLS results for PfSir2a in sodium phosphate buffer without 1mM DTT (A) and with 1mM DTT (B).with 100mM or 500mM NaCl.

We can see that in this buffer condition, the samples show a polydispersity of greater than 20% and can no longer be considered a monodisperse solution, which is a pre-requisite condition for SAXS experimental analysis.

4.4.2 Concentration Effect Characterisation

With regard to concentration effects in the samples as analysed by SAXS, we know from previous AUC data on PfSir2a that there is a concentration dependent oligomeric state. However, this can only be properly assessed when the effects of radiation damage and polydispersity have been mitigated By modifying the buffer conditions to 50mM sodium phosphate, 500mM NaCl, 1mM DTT a repeat set of measurements was made on the PfSir2a concentration series. The results of this dataset can be seen in Figure 4.17.

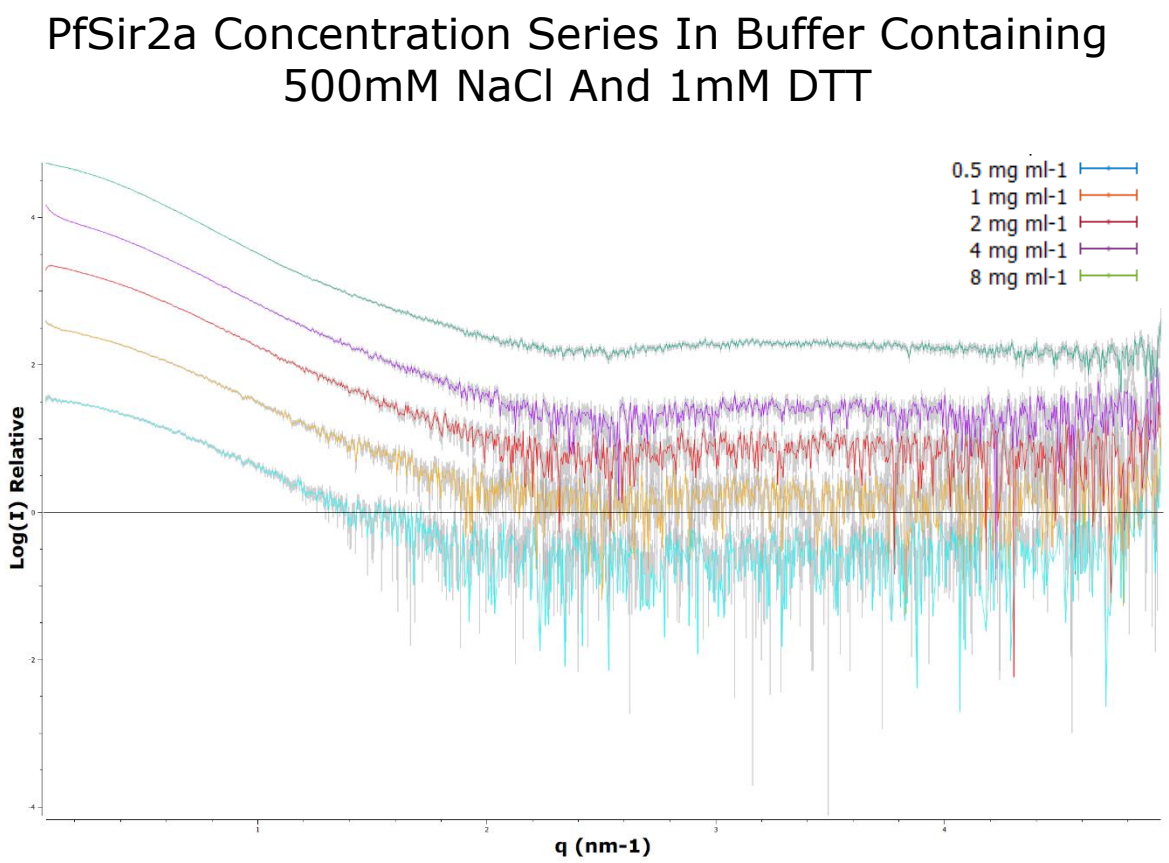


Figure 4.17: Processed scattering curves for PfSir2a concentration series measurements in buffer containing 500mM NaCl and 1mM DTT. Intensity versus q is plotted on a relative log scale. A reduced upswing in intensity is observed at low q due to the addition of 1mM DTT to the sample buffer.

From this repeated data set we can see a marked decrease in intensity in the low *q* region. The addition of 1mM DTT to the sample buffer has significantly reduced the effect of radiation-damage-induced aggregation as also seen in the 100mM NaCl containing buffer. This is again highlighted when comparing the data sets with and without the addition of 1mM DTT to the sample buffer with a salt concentration of 500mM NaCl (see Figure 4.18). An increase in low *q* scattering was still observed for the 4mg ml^{-1} PfSir2a sample.

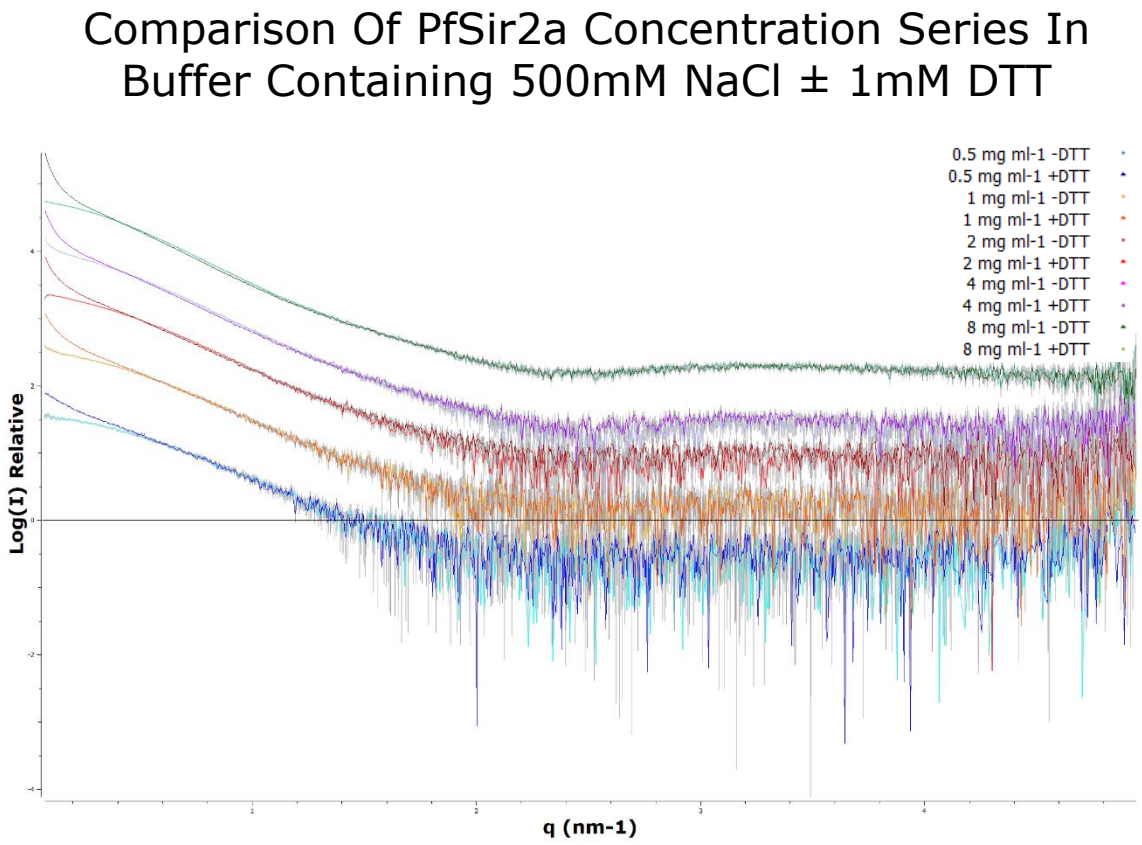


Figure 4.18: Processed scattering curves for PfSir2a concentration series measurements in buffer containing 500mM NaCl ± 1mM DTT. Intensity versus *q* is plotted on a relative log scale. The plot shows a clear improvement in the processed scattering curves for the PfSir2a concentration series from the addition of 1mM DTT to the sample buffer.

By normalising the data against sample concentration, we can see that there is still a divergence at low q between the samples in the concentration range tested (see Figure 4.19). The experimentally determined values for the R_g and MM of PfSir2a in the 500mM NaCl buffer with 1mM DTT were calculated and can be seen in Table 4.6.

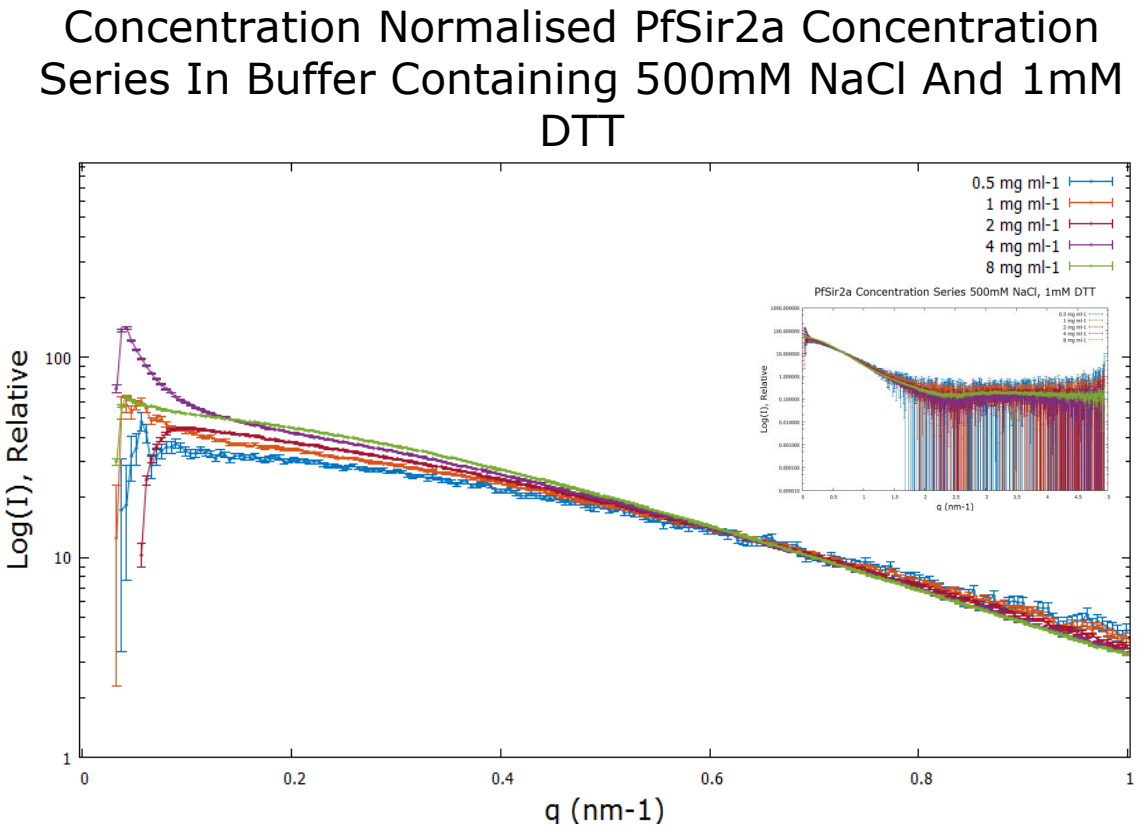


Figure 4.19: Concentration normalised scattering curves for PfSir2a concentration series in buffer containing 500mM NaCl and 1mM DTT from 0 to 1 q. The total scattering curves can be seen inserted on the right for reference. A reduction in low q intensity is observed in the processed concentration series results after the addition of 1mM DTT to the sample buffer. Despite improvements to the sample preparation, a difference is still observed between the scattering curves for the different concentrations of PfSir2a at low q after normalisation.

Sample – Concentration (mg ml^{-1})	R_g (nm)	Error R_g ± (nm)	I(0)	Error I(0) ±	Absolute I(0)	Molecular Mass (kDa)
Water – N/A	-	-	2.1E+01	4.0E-03	-	-
PfSir2a - 0.5	2.94	0.217	3.42E+01	6.28E-01	2.65E-02	39.23
PfSir2a - 1	3.43	0.094	4.03E+01	3.60E-01	3.13E-02	46.22
PfSir2a - 2	3.76	0.04	4.56E+01	1.92E-01	3.54E-02	52.30
PfSir2a - 4	**4.12**	**0.029**	**5.30E+01**	**1.83E-01**	**4.11E-02**	**60.79**
PfSir2a - 8	3.73	0.012	5.36E+01	6.90E-02	4.16E-02	61.48

Table 4.6: Experimentally determined values for R_g and MM of PfSir2a in 500mM NaCl and 1mM DTT containing buffer.

From the results seen in Table 4.6 we can identify several features in the data. Firstly, the calculated values for the R_g and MM of PfSir2a are lower than that which has been determined from the previous datasets. This indicates that the problems of radiation damage and polydispersity from sub-optimal buffer composition have been significantly reduced. However, intrinsic or concentration based issues can still be present. A general increase in the calculated R_g from 2.9 nm to 3.7 nm is found. Additionally, the calculated MM increases from 39.2 kDa to 61.5 kDa.

Further improvements can be seen in the Guinier plot now following a straight line at low q ($q < 0.2$ nm^{-1}) indicating that the Guinier approximation holds valid (see Figure 4.20). However, even at the lowest concentration of recombinant PfSir2a, 0.5 mg ml^{-1}, we see a higher than expected value for both the MM (+5.4 kDa) and the R_g (+0.5 – 0.76 nm).

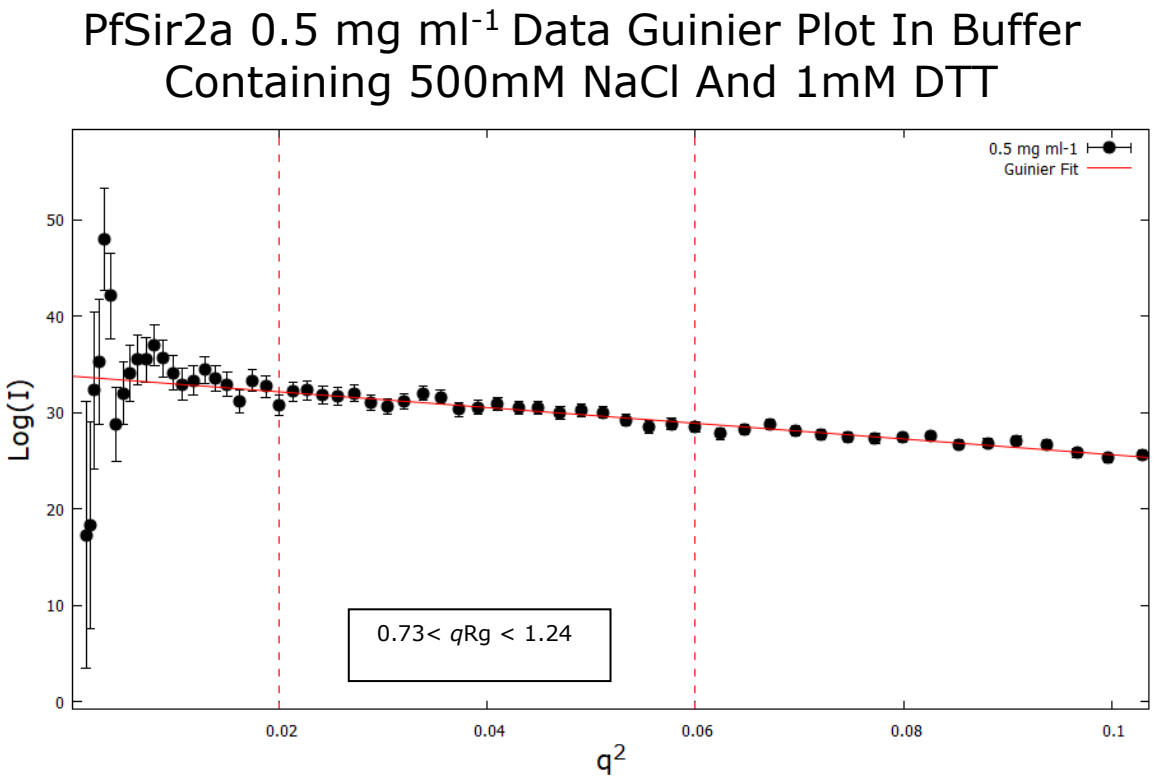


Figure 4.20: Guinier plot of Log(I) versus q^2 for PfSir2a sample at 0.5 mg ml^{-1} in 500mM NaCl, 1mM DTT buffer. The data points satisfying the Guinier approximation were plotted with a line of best fit. A significant improvement in the linearity of the Guinier plot is observed at low q (q < 0.2 nm^{-1}).

Next, the experimental data for the 0.5 mg ml^{-1} sample were fitted against the theoretical scattering curves for the I-TASSER modelled recombinant PfSir2a as well as the two pre-existing crystal structures, 3JWP and 3U31, for data validation and comparison. The data fitting using CRYSOL can be seen in Figure 4.21. The chi-square values determined for all models are artificially low due to the large error range when fitting to the scattering data. However, Of the five I-TASSER models produced, model 2 was found to have the best fit to the experimental scattering data with chi-square fit values of 0.9. The fits for the pre-existing PfSir2a crystal structures for the monomer and trimer form, 3U31 and 3JWP, were found to fit poorly with the experimental scattering data with chi-square fit values of 3.354 and 5.197 respectively.

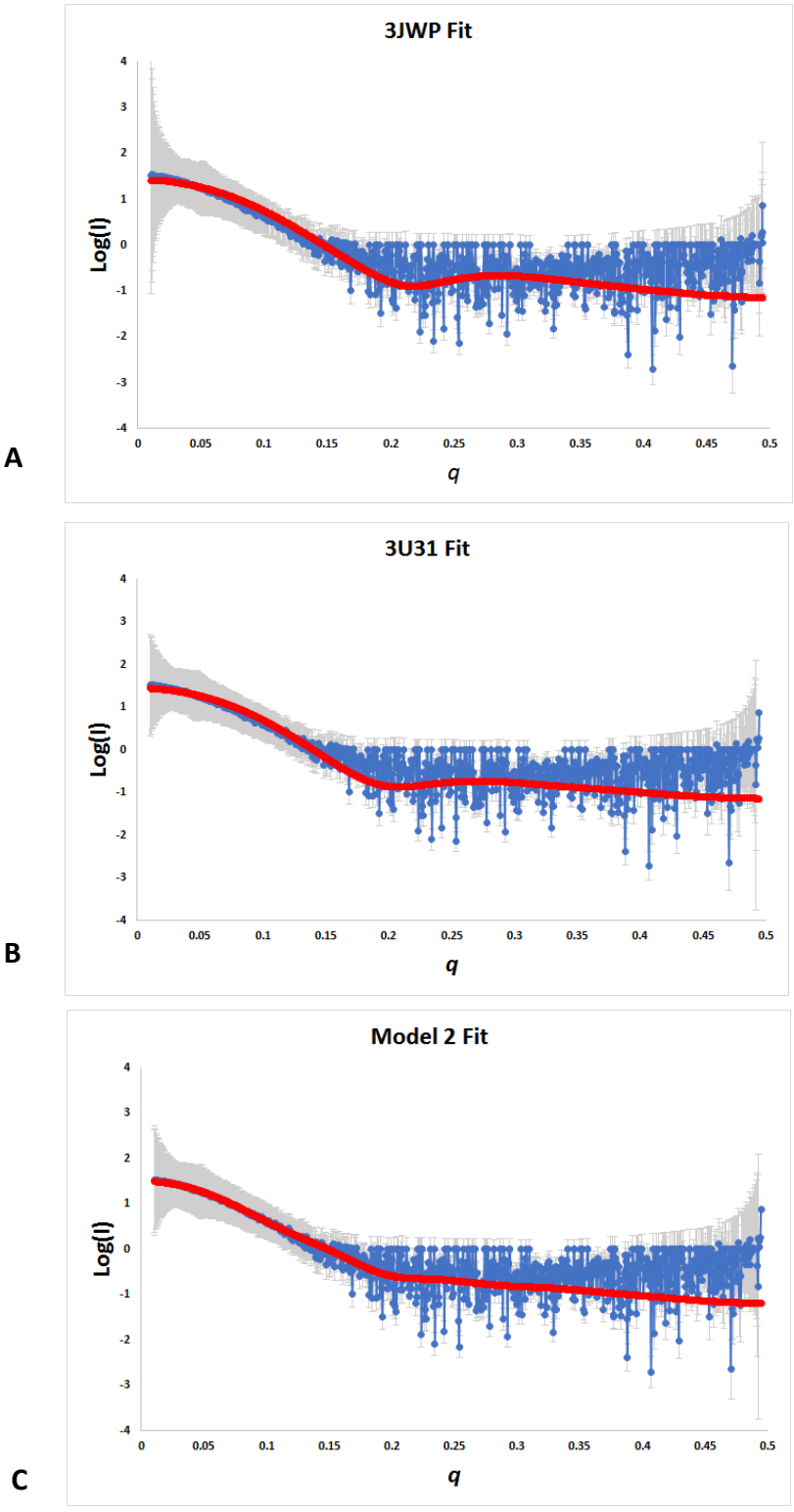


Figure 4.21: CRYSOL output of 0.5 mg ml^{-1} experimental scattering data in 500mM NaCl, 1mM DTT (blue) fit with theoretical scattering curves for I-TASSER model 2, PfSir2a crystal structure 3U31 and 3JWP (red) in panels A, B and C respectively.

Taking the scattering data for the 0.5 mg ml^{-1} sample we can calculate the pair-distance distribution function *p(r)* for the PfSir2a molecule in solution. This is a description of the paired set of all distance between points, *r*, within an object. It is described by the following equation (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013):

$$I(q) = 4\pi \int_0^{D_{max}} p(r) \frac{\sin(qr)}{qr} dr$$

Equation 4.4

This provides information about the overall size of the molecule, D_{max}, and gives an indication of the molecules shape and/or presence of aggregation. The *p(r)* distribution for PfSir2a can be seen in Figure 4.22.

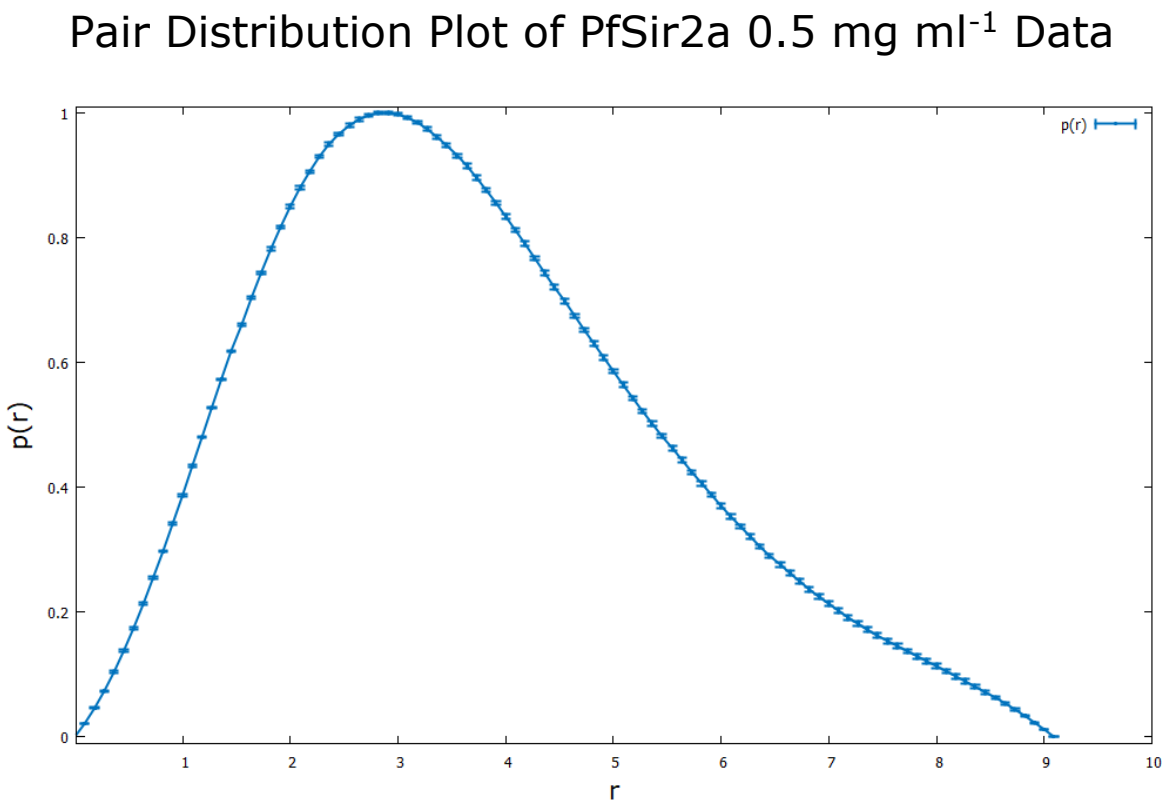


Figure 4.22: Plot of *p(r)* function for PfSir2a 0.5 mg ml^{-1} sample scattering data. The bell-shaped nature of the plot indicates that PfSir2a is a globular protein in solution.

The profile of the *p(r)* plot for PfSir2a shows the expected bell-shape for a globular protein. There is a slight tail to the *p(r)* distribution at larger distances of *r* which can be an indication of aggregation. The D_{max} for PfSir2a at 0.5 mg ml^{-1} was found to be 8.1 nm ± 0.5nm. Further confirmation of whether PfSir2a exists as a globular protein in solution was determined by a Kratky plot of $I(q) \times q^2$ versus q. This can be seen in Figure 4.23. The plot shows a bell-shaped curved with a well-defined maximum which is characteristic for a folded, globular protein. Unfolded and disordered proteins show much slower intensity decay as q increases yielding a plateau instead of a peak.

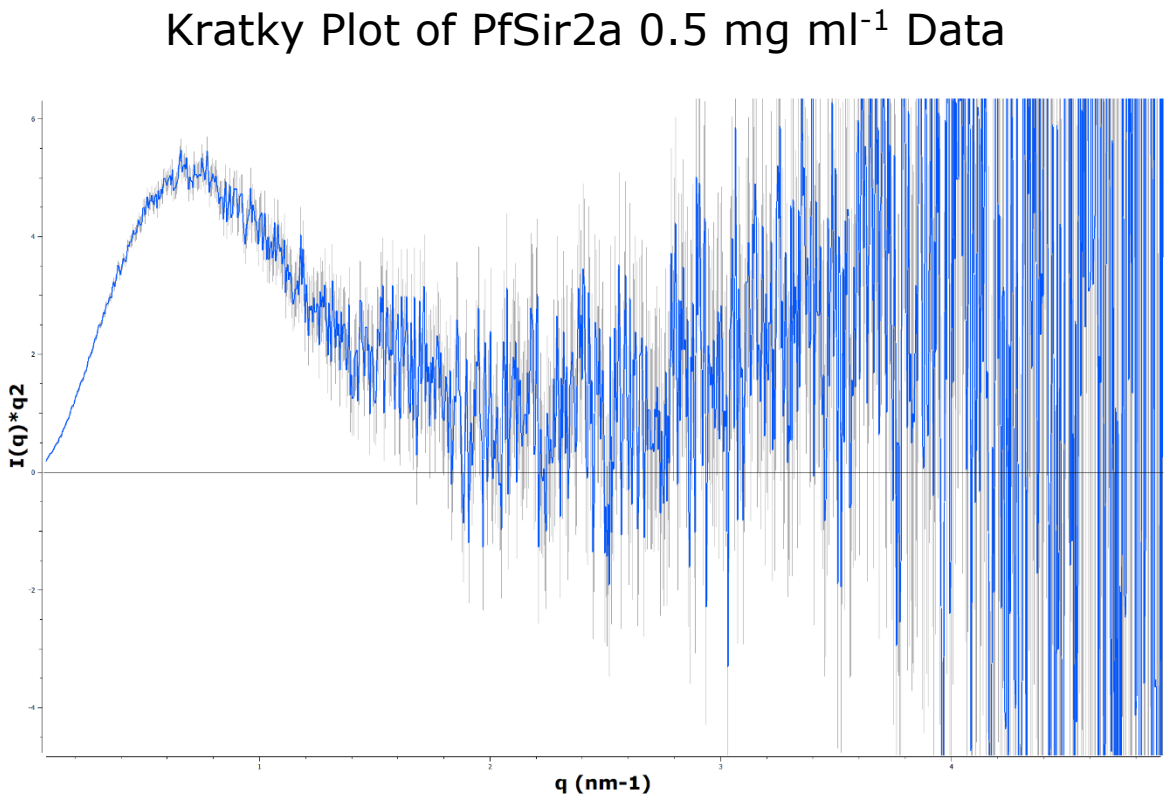


Figure 4.23: Kratky plot of PfSir2a 0.5 mg ml^{-1} sample scattering data. The bell-shaped nature of the plot at low q indicates that PfSir2a is globular in solution. Unfolded proteins typically show a plateau region at higher q.

4.4.3 *Ab initio* Modelling Of SAXS Data For PfSir2a

Ab initio modelling for PfSir2a was performed using the 0.5 mg ml^{-1} scattering data using the DAMMIF program. 40 models were produced from fits to the scattering data which were averaged together after removing least likely models by DAMAVER. The results of the *ab initio* model generation can be seen in Figure 4.24.

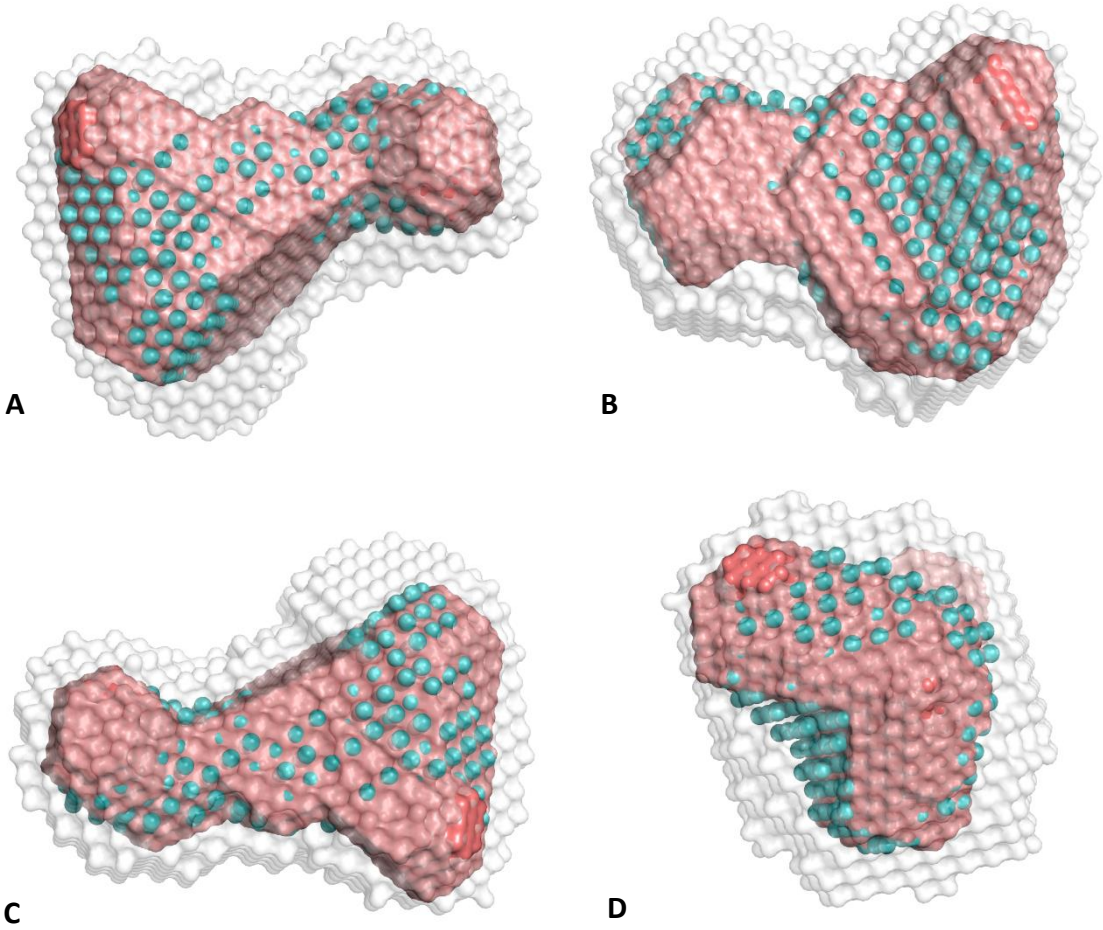


Figure 4.24: Results of *Ab initio* modelling for PfSir2a showing (A) Side view, (B) Alternate side view, (C) Top View and (D) end view. DAMAVER averaged envelope of all models is shown in grey. DAMMIF filtered model is shown in cyan. The most representative model (MRM) is shown in red.

In Figure 4.24 we can see that the SAXS envelope calculated by DAMMIF using the experimental scattering data is significantly larger than the I-TASSER generated model and highlights the variation in the model reconstructions. The filtered model which has low space occupancy areas removed from the averaged model shows the most likely SAXS envelope structure for PfSir2a in solution. However, this does not accurately reflect the experimental scattering data, instead the model with the lowest reported NSD value (i.e the most representative model (MRM) of the protein in solution) is used. This model was compared the I-TASSER model2 in Figure 4.25 which was found to have the best fit to the data in Figure 4.21.

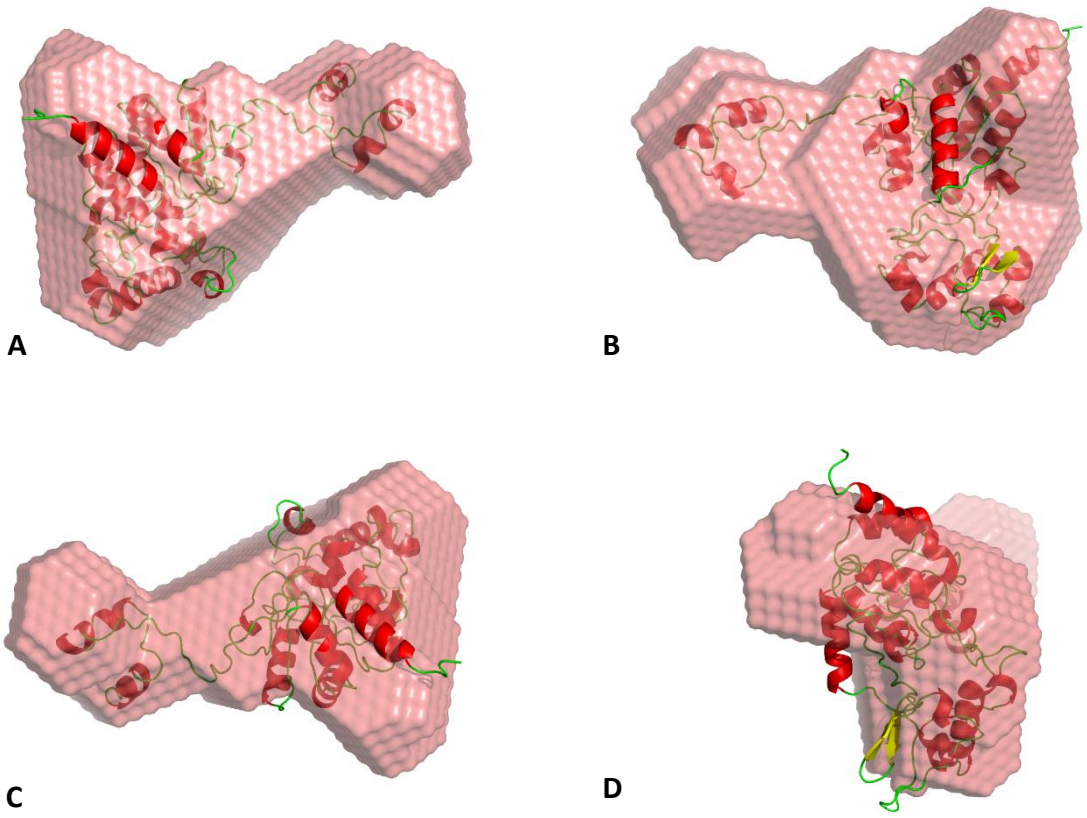


Figure 4.25: Results of *Ab initio* modelling for PfSir2a showing (A) Side view, (B) Alternate side view, (C) Top View and (D) end view. MRM envelope model is shown in red, I-TASSER model2 is shown in cartoon.

The fit of the scattering data for PfSir2a to the MRM as determined by CRYSOL is shown in Figure 4.26.

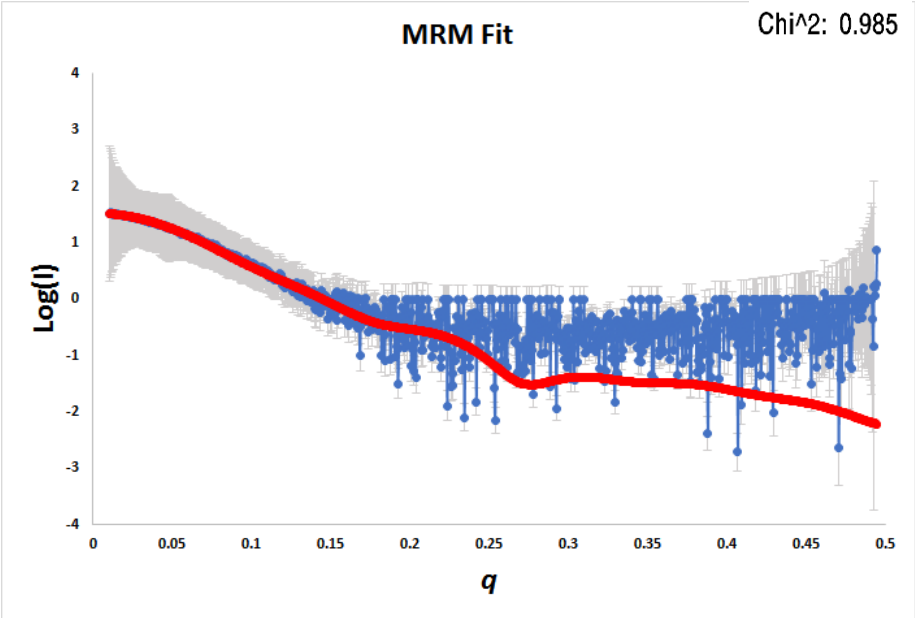


Figure 4.26: CRYSOL output of 0.5 mg ml^{-1} PfSir2a experimental scattering data in 500mM NaCl, 1mM DTT (blue) fit with theoretical scattering curves for MRM (red).

Initially radiation damage had a dramatic impact on the quality of the data obtained from scattering experiments. After reducing the effect of radiation damage, it also became very apparent that concentration effects were also affecting the data for PfSir2a, even at low concentrations (0.5 mg ml^{-1}).

While AUC data supports the idea that there is a monomer-trimer mixture based on the estimated MM (see Chapter 3, section 3.8.1), the R_g of the 0.5 mg ml^{-1} PfSir2a sample exceeds that of the theoretically calculated value for a recombinant PfSir2a trimer, while the molecular mass estimate suggests a predominant monomer population. This R_g further increases in size with the concentration of PfSir2a in solution.

Given the known concentration effects for PfSir2a even at low concentrations, this has presented a significant challenge when attempting to interpret structural information about PfSir2a at higher concentration levels. Without crystal structure information for the recombinant PfSir2a in either monomer or trimer forms it becomes speculation as to the exact nature of the populations and their scattering profiles in solution.

4.5 PfAlba3 SAXS Experimental Results

This section outlines the main results observed from SAXS experiments using purified PfAlba3 samples.

4.5.1 Characterising And Addressing The Rapid Aggregation Of PfAlba3

SAXS measurements of PfAlba3 were also carried out using both the automated sample changer and online SEC system available at BM-29. An initial concentration series was carried out at a sample concentration range of 0.5 mg ml^{-1} to 5 mg ml^{-1}. Initial analysis of the data revealed that PfAlba3 is heavily aggregated by the time the sample could be purified, concentrated and prepared for the measurements using the sample changer. This is again indicated by the sharp rise in intensity at low q (see Figure 4.27).

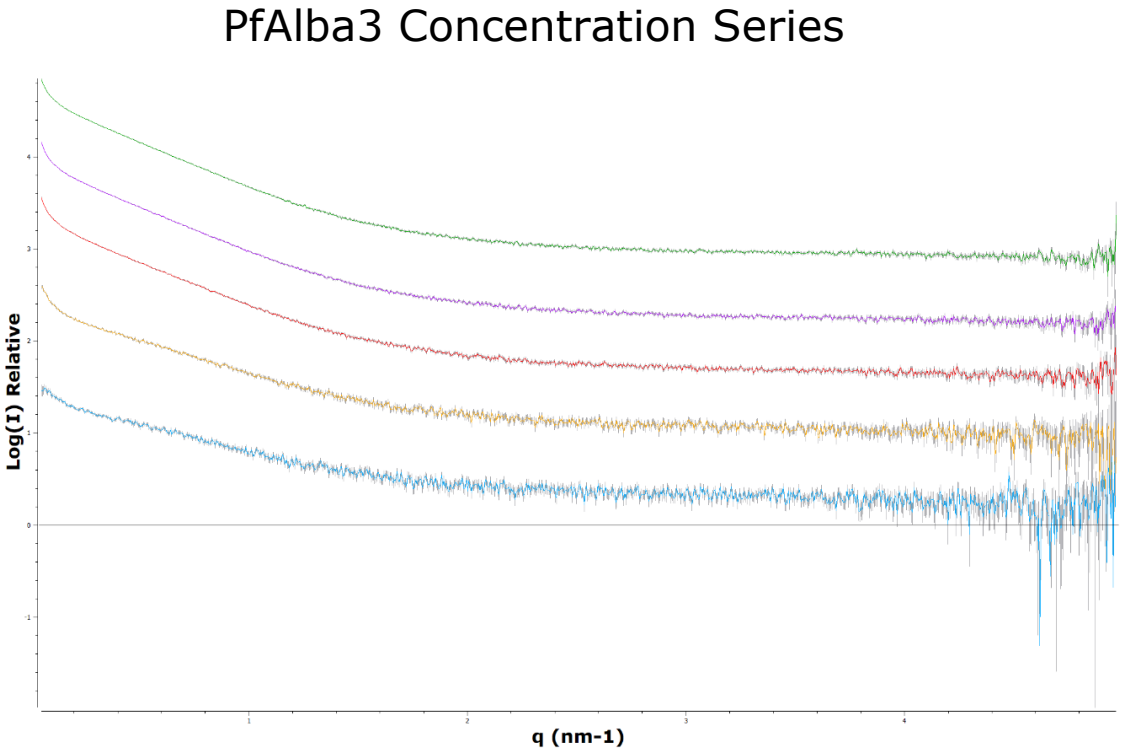


Figure 4.27: Processed scattering curves for PfAlba3 concentration series measurements of Intensity versus *q* on a relative log scale. A sharp upswing in intensity is observed at low *q* that is indicative of sample aggregation.

4.5.2 The Use Of Online SEC-SAXS At BM29 To Address Aggregation

Given the propensity for PfAlba3 to aggregate so rapidly in solution after purification and its low tolerance for concentration before precipitation (< 5 mg ml^{-1}), further sample changer measurements would unlikely give accurate results. To try and overcome this problem of sample stability, the online SEC system present at BM-29 was used (Round et al., 2013). Here a size exclusion column is connected to a HPLC pump system which allows for the injection of a protein sample onto a size exclusion column (varied depending on size of protein). As the protein passes out of the column it passes through the sample capillary where it can be exposed to X-rays for measurements to take place and scattering data be obtained. This offers several advantages such as being able to separate aggregates

from sample proteins, oligomeric species, protein complexes and unbound components as well as UV trace information for accurate protein concentration measurements and peak profile tracking (Round et al., 2013).

PfAlba3 samples at a concentration of 2.5 mg ml^{-1} were injected onto a *Superdex*™ S75 10/300 GL column pre-equilibrated in a size exclusion sample buffer containing 50mM sodium phosphate, 500mM NaCl, 1mM DTT, pH 7.0. A sample volume of 100µl and a flow rate of 0.5 ml min^{-1} were used. During the experiment, the measured *I(0)* of 1 second exposure frames was monitored and plotted against the automatically determined R_g of the protein from the BM-29 data pipeline for 1 column volume. The results can be seen in Figure 4.28.

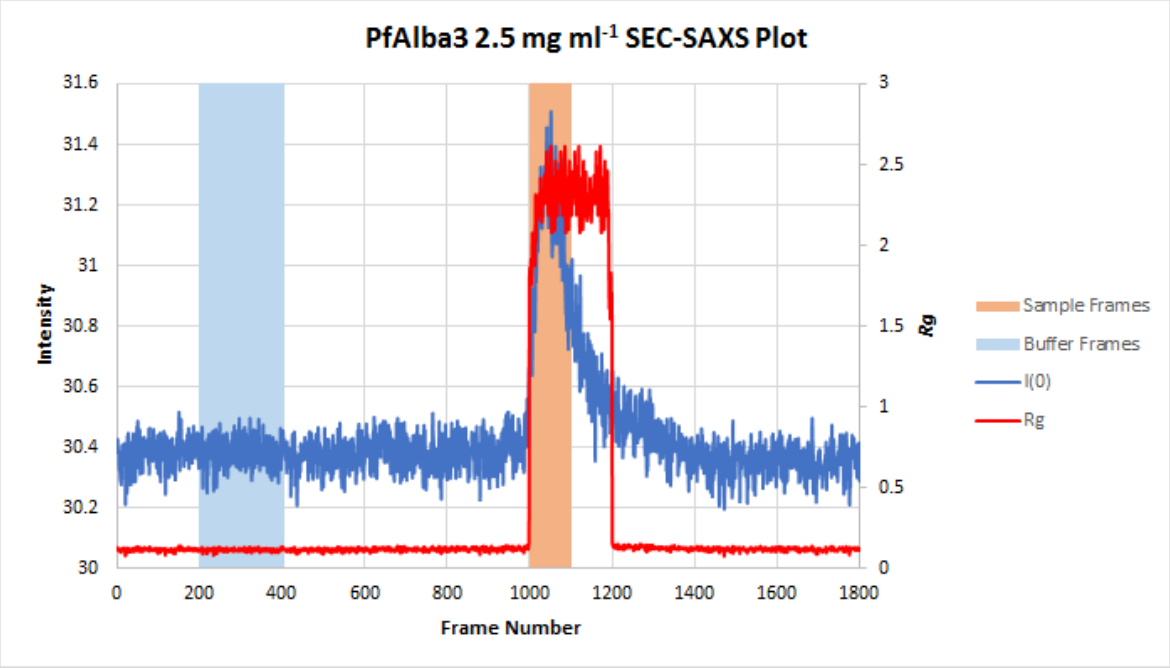


Figure 4.28: Plot of recorded *I(0)*of 1s frames and R_g for every recorded frame during On-line SEC run.

Here we see the expected size exclusion profile for PfAlba3 as it migrates through the column and the observed scattering across the elution volume. A distinct peak can be

observed in the measured *I(0)* of the recorded frames which correlates with the elution profile of PfAlba3. Frames between 1000 and 1100 were taken and processed to provide an average scattering pattern from which an average buffer scattering pattern recorded at the beginning of the experiment (frames 200 – 400) could be subtracted. The processed data for PfAlba3 can be seen in Figure 4.29.

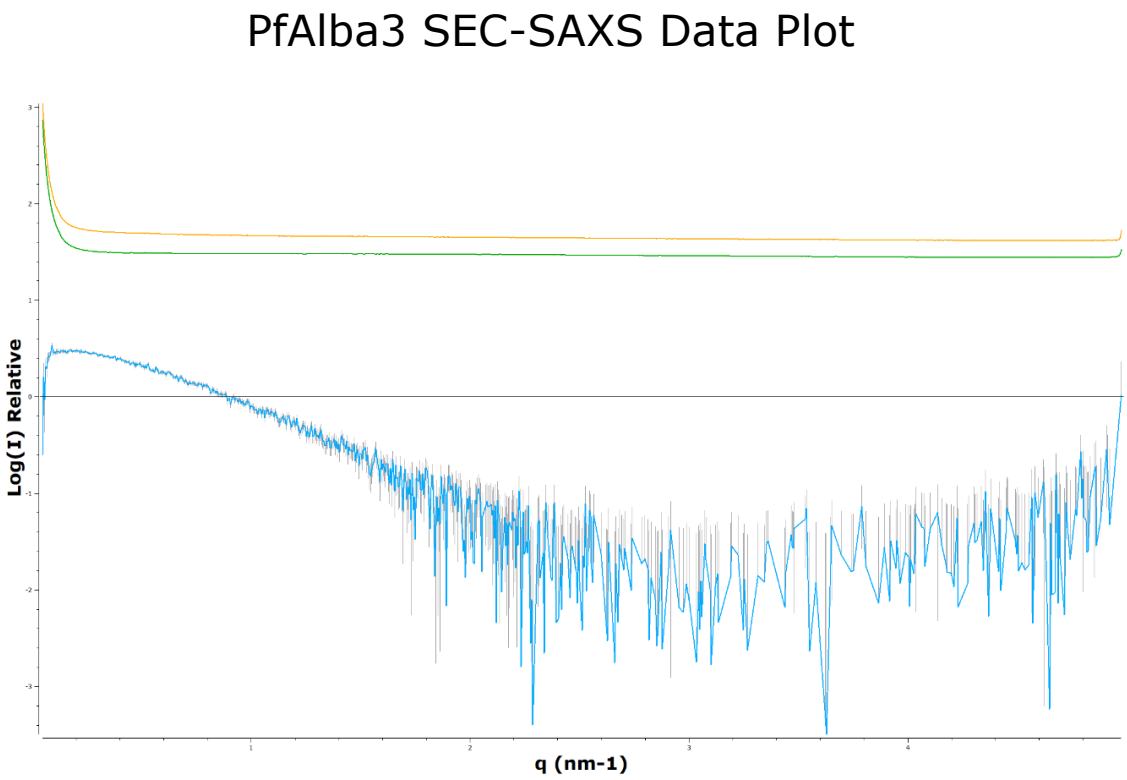


Figure 4.29: Plot of PfAlba3 processed experimental scattering curve. Merged PfAlba3 frames (Orange), merged buffer frames (Green) and subtracted data (Blue).

The results from the processed scattering data seen in Figure 4.29 show several features and improvements. Firstly, there is no sharp increase in intensity at low q ($q < 0.3$ nm^{-1}) as seen previously as a result of large proteins aggregates forming in the sample or as a result of radiation damage from X-ray exposure. The Guinier plot is linear at low q (see Figure 4.30) and values for the R_g and MM were calculated (Table 4.7).

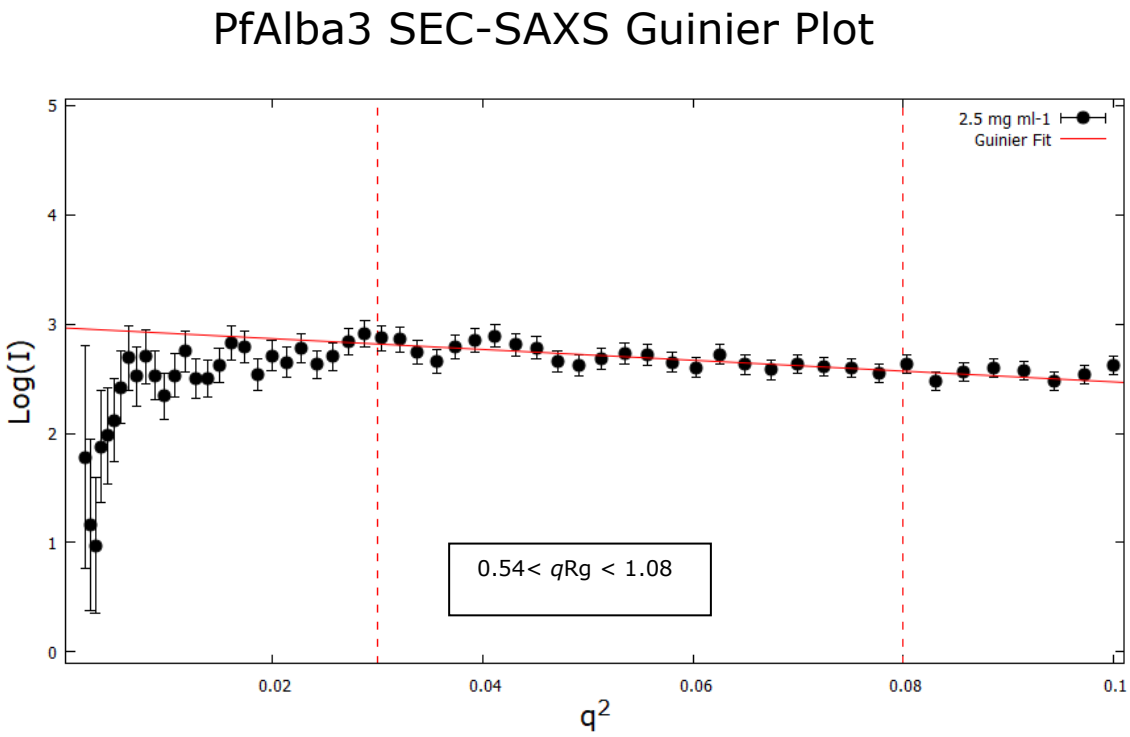


Figure 4.30: Guinier plot of Log(I) versus q^2 for PfAlba3 SEC sample injected at 2.5 mg ml^{-1} in 500mM NaCl, 1mM DTT buffer. The data points satisfying the Guinier approximation were plotted with a line of best fit.

Sample – Concentration (mg ml^{-1})	R_g (nm)	Error R_g ± (nm)	I(0)	Error I(0) ±	Absolute I(0)	Molecular Mass (kDa)
Water – N/A	-	-	2.07E+01	1.20E-02	-	-
PfAlba3 – 2.5	2.35	0.325	2.98E+00	8.80E-02	2.35E-03	3.47

Table 4.7: Experimentally determined value for R_g and MM of PfAlba3 in 500mM NaCl and 1mM DTT containing buffer.

Here we find an experimentally determined R_g for PfAlba3 of 2.35nm, that is larger than the theoretically predicted 1.6 – 1.7 nm determined by CRYSOL. A very low *I(0)* is observed and is reflected in the calculated MM of 3.5 kDa. This could perhaps be attributed in part to some sample loss on the column filter due to large precipitates. Also,

as a low molecular weight protein (13.04 kDa), PfAlba3 has a low overall scattering density which can make it difficult to distinguish above that of the solvent scattering at low protein concentrations. This is more difficult in the case of high salt buffers which have an increased scattering density due to higher electron density compared to that of water or low salt buffers.

4.5.3 *Ab initio* Modelling Of SAXS Data For PfAlba3

Next, the experimental data for the sample were fitted against the theoretical scattering curves for the I-TASSER modelled recombinant PfAlba3. The best fit to the data was found to be model 3 with a chi-square fit value of 1.811 (see Figure 4.31). This appears to fit the experimental scattering curve fairly well.

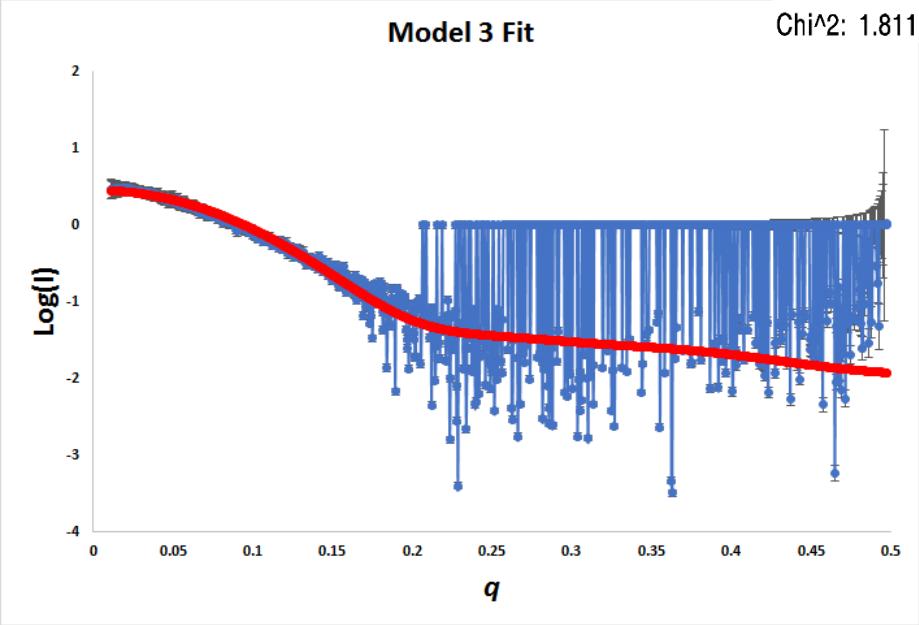


Figure 4.31: CRYSOL output of on-line SEC PfAlba3 sample scattering data in 500mM NaCl, 1mM DTT (blue) fit with theoretical scattering curves for I-TASSER model3 (red).

The *p(r)* function for PfAlba3 was calculated next to obtain the D_{max} of the molecule and can be seen in Figure 4.32. Again, the profile of the *p(r)* plot for PfAlba3 shows the expected bell-shape for a globular protein. The D_{max} for PfAlba3 at 2.5 mg ml^{-1} was found to be 7.8 nm ± 0.2nm.

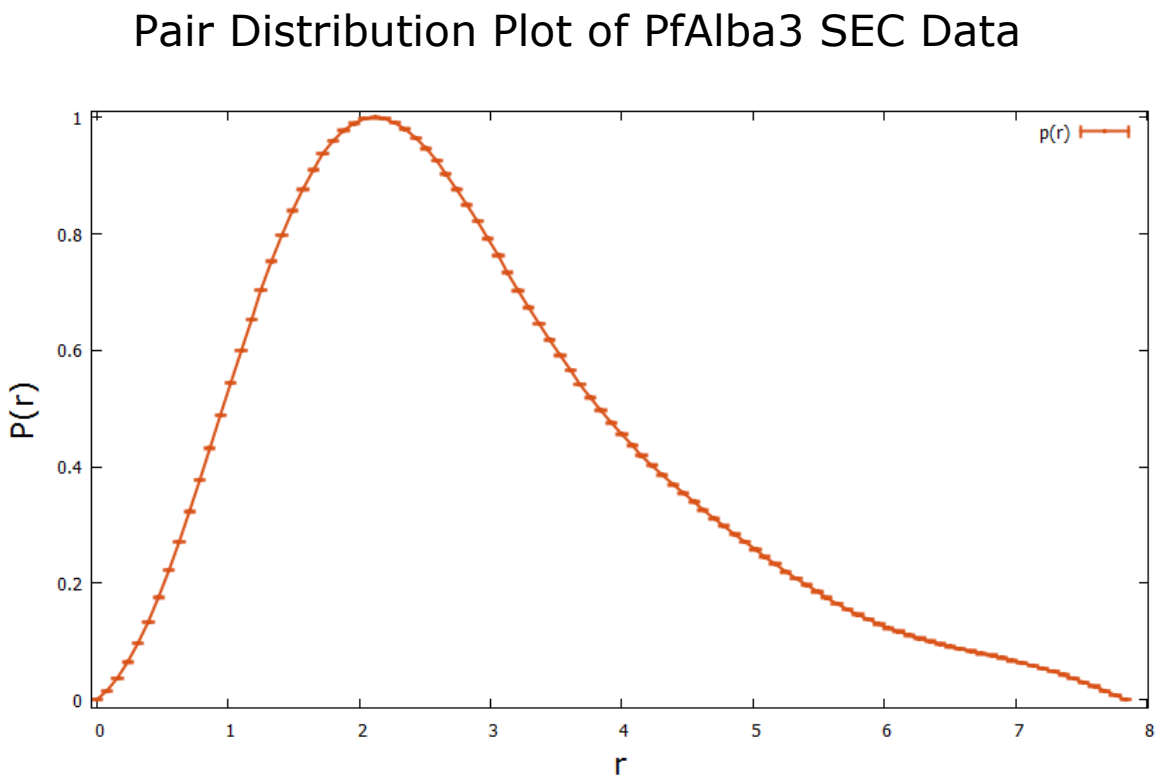


Figure 4.32: Plot of *p(r)* function for PfAlba3 2.5 mg ml^{-1} sample scattering data.

The Kratky plot seen in Figure 4.33 also indicates that the protein is folded and not disordered. *Ab initio* modelling for PfAlba3 was performed using the 2.5 mg ml^{-1} (injected sample concentration) scattering data using the DAMMIF program. 40 models were produced from fits to the scattering data which were averaged together after removing least likely models by DAMAVER.

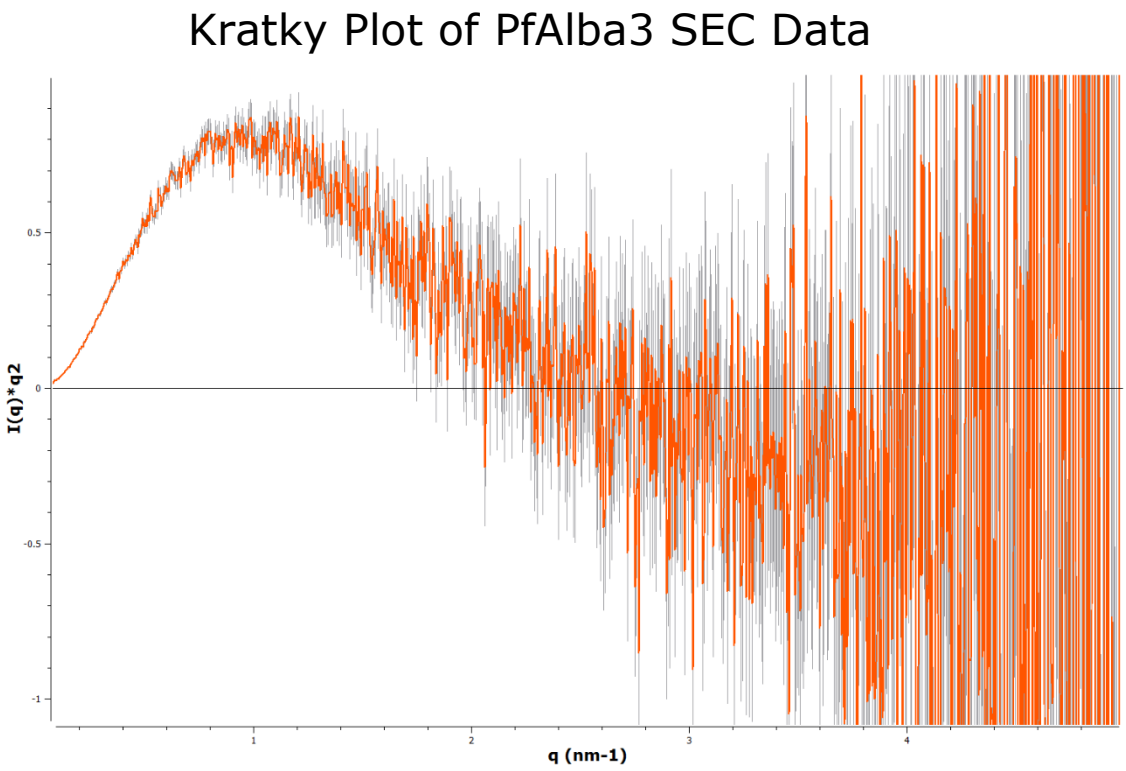


Figure 4.33: Kratky plot of PfAlba3 2.5 mg ml^{-1} scattering data.

The results for the *ab initio* modelling are shown in Figure 4.34 for PfAlba3 and indicate a much more oblate shape compared to that of PfSir2a, despite similar calculated values for the D_{max} of each particle (7.8 and 8.1 nm respectively). The averaged SAXS envelope model seen in Figure 4.34 also indicates a much larger size than the I-TASSER generated model and highlights the variation in the model reconstructions. By removing the areas of low space occupancy we see a likely estimate of the solution based structure of PfAlba3 although, again, it does not fully represent the experimentally obtained scattering data.

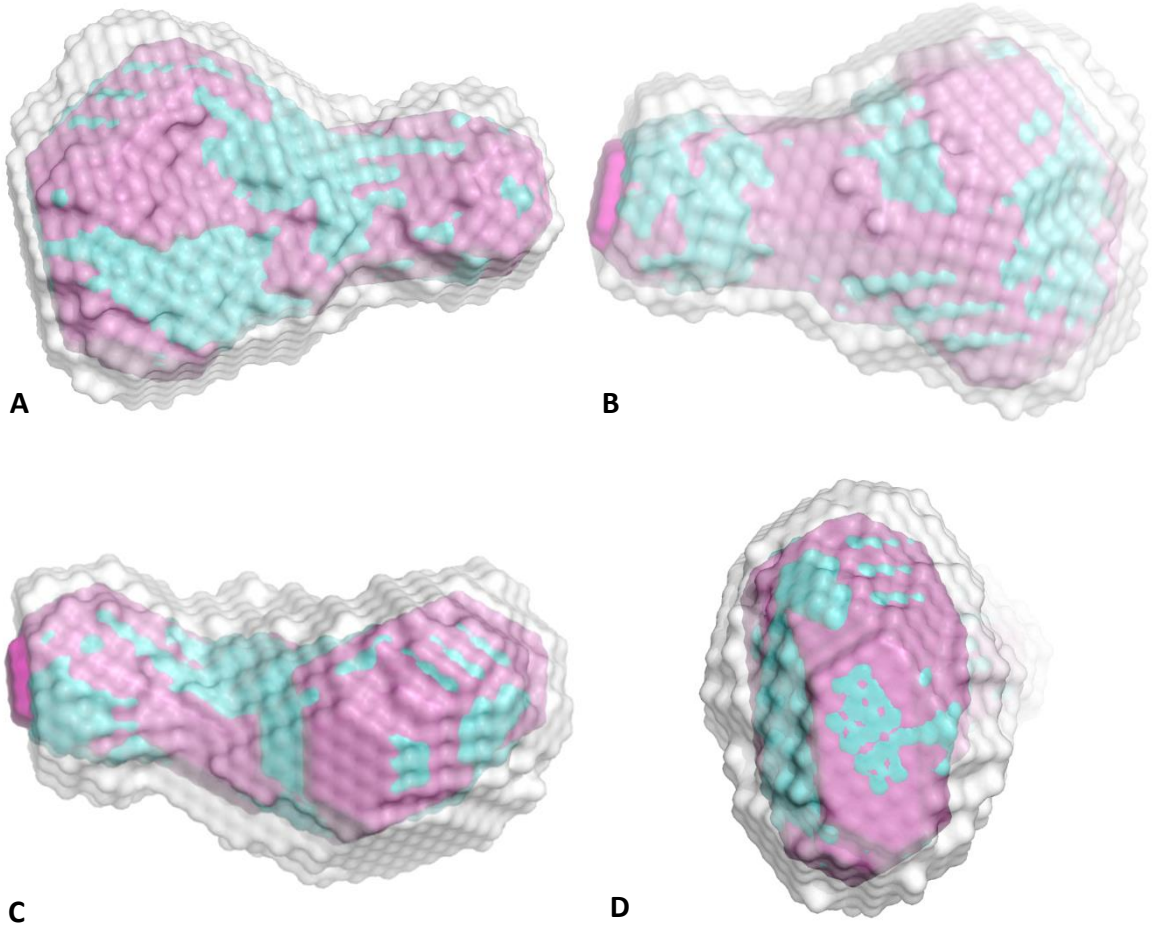


Figure 4.34: Results of *Ab initio* modelling for PfAlba3 showing (A) Side view, (B) Alternate side view, (C) Top View and (D) end view. DAMAVER averaged envelope of all models is shown in grey. DAMMIF filtered model is shown in cyan. The most representative model (MRM) is shown in pink.

As this does not accurately reflect the experimental scattering data, instead the model with the lowest reported NSD value (i.e the most representative model (MRM) of the protein in solution) is used. This model was compared the I-TASSER model3 in Figure 4.35 which was found to have the best fit to the data in Figure 4.31.

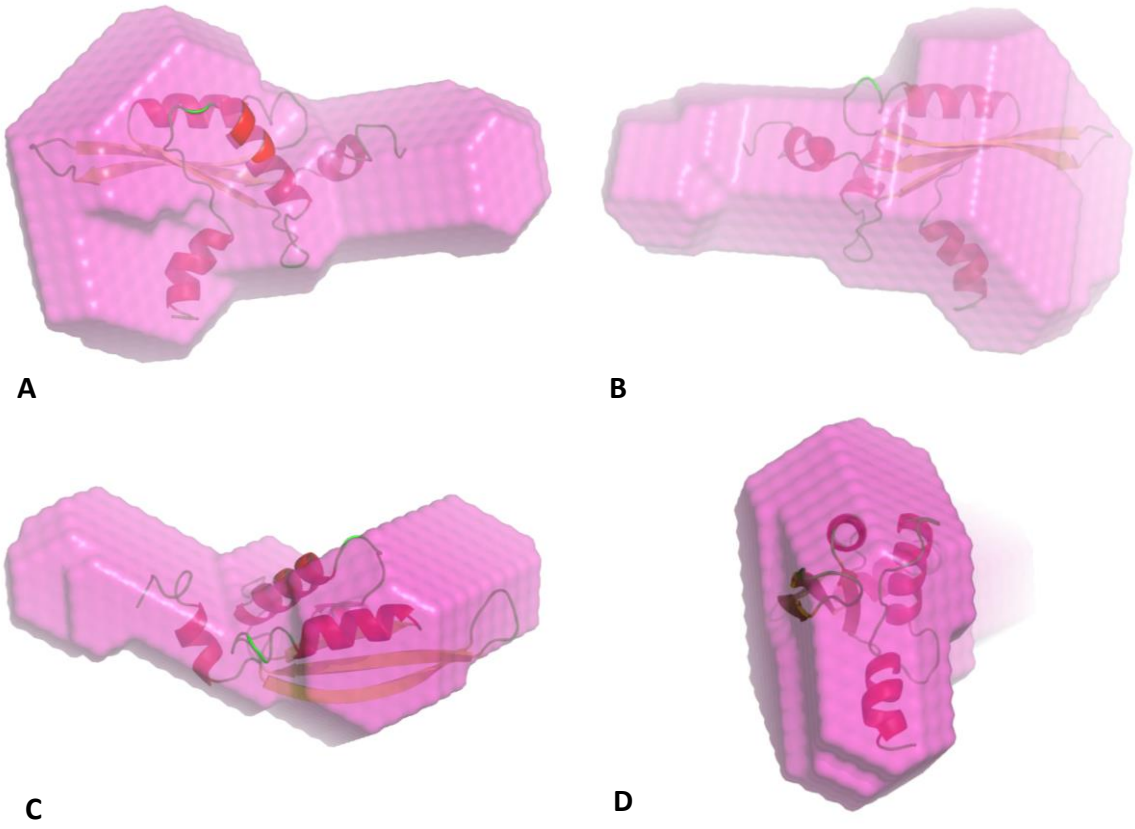


Figure 4.35: Results of *Ab initio* modelling for PfAlba3 showing (A) Side view, (B) Alternate side view, (C) Top View and (D) end view. MRM envelope model is shown in red, I-TASSER model3 is shown in cartoon.

The fit of the scattering data for PfAlba3 to the MRM as determined by CRYSOL is shown in Figure 4.36

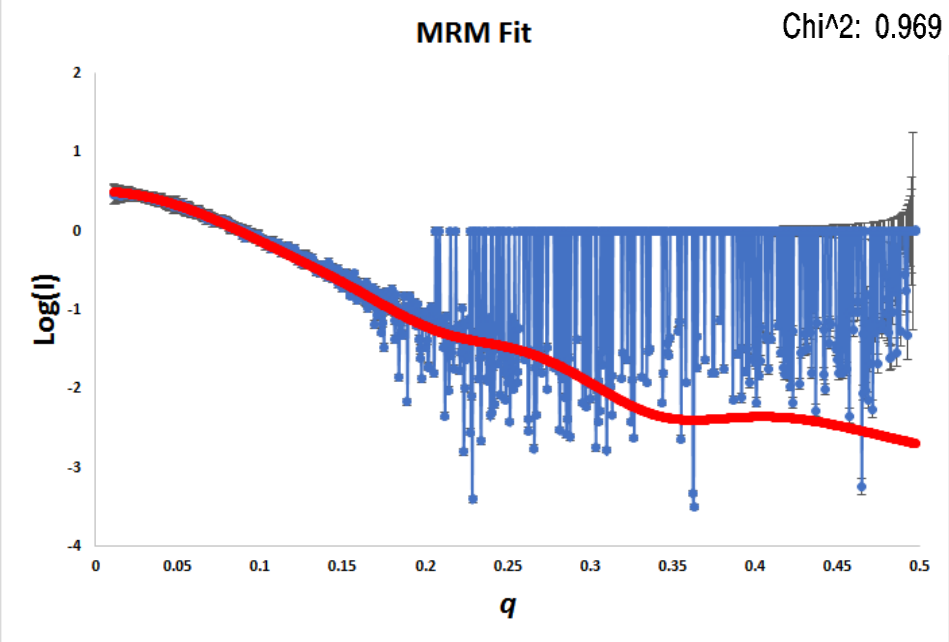


Figure 4.36: CRYSOL output of 2.5 mg ml^{-1} PfAlba3 experimental scattering data in 500mM NaCl, 1mM DTT (blue) fit with theoretical scattering curves for MRM (red).

4.6 Summary

In summary this chapter provides an in-depth analysis of SAXS data gathered on the recombinant PfSir2a and PfAlba3 protein samples that have been produced and characterised extensively by the biochemical and biophysical methods outlined in Chapter 3. We find during the analysis that there are several challenges with each protein which were investigated and optimised to provide the most accurate solution structure analysis for each protein.

For PfSir2a we find that the protein samples do not well-tolerate exposure to X-rays and radiation-damage-induced protein aggregation was contaminating the scattering patterns. Once the problem of radiation damage was mitigated, the analysis of PfSir2a was further

complicated by concentration dependent inter-particle effects between oligomeric species. Given the unknown nature of how strong the interaction between the molecules is at different concentrations and the limited pre-existing crystal structure data available, a low-resolution solution model of PfSir2a was produced by *ab initio* modelling via DAMMIF. Improvements to the quality of the data and analysis could be made by measuring samples at a lower concentration (i.e. < 0.5 mg ml^{-1}) which could help to reduce the aforementioned problems with radiation damage and concentration effects.

Similarly, for PfAlba3 protein aggregation was a major contributing factor to problems with sample preparation and contamination of scattering data. The delicate nature of the sample prompted the use of the on-line SEC experimental set-up available at the ESRF beamline BM-29 for the collection of the sample scattering data. This resulted in successful data collection allowing the use of the *ab initio* modelling software to produce a solution based structure for PfAlba3.

Chapter 5: SEC-SANS Development

Abstract

This chapter focuses on the work carried out for the development of a technique for use in Bio-SANS measurements at the D22 beamline, ILL Grenoble. Here the first use of *in situ* SEC at a SANS beamline is presented for continuous flow measurements. This serves to improve the collection of data on interesting biological samples which often can be unstable or polydisperse in nature over traditional static SANS measurements.

A modified quartz cuvette was placed in a continuous flow cell device directly in the path of a neutron beam. This was connected to an ÄKTAPrime FPLC system for sample injection and flow control. Data was collected on SEC standard samples in both hydrogenated and deuterated buffer conditions for Guinier analysis and *ab initio* modelling. This system was also adapted to explore the solution structure properties of PfSir2a and PfAlba3.

5.1 Introduction

As seen in previous chapters, radiation damage can be a serious issue for biological samples when attempting to analyse SAXS data. This well-known damaging effect can cause significant increases in the R_g of the macromolecules being studied due to radiation damage-induced aggregation. This has severe effects which result in un-interpretable data or a false representation of the sample (Fischetti et al., 2003; Jeffries et al., 2015; Kuwamoto et al., 2004).

SANS can provide information that strongly complements that gained from SAXS, and also offers an alternative approach to analyse samples that can easily suffer from the effects of radiation damage (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013). This is due to the scattering of neutrons by nuclei rather than electrons which is the case for X-rays.

As a result, radiation damage does not occur in the same way for samples analysed by SANS (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013). This offers a unique opportunity to investigate sensitive samples which are of significant biological importance.

Traditionally SANS experiments are carried out with a sample protein solution held in a quartz cuvette having a 1mm path length. For experiments involving deuterium labelled protein-protein, protein-DNA, or protein-RNA samples, the measurements are conducted in a series of different H_2O:D_2O buffer compositions to match out one or several components in the complexes to identify the structural changes that occur upon binding in the unmatched partner. This is compared to the individual proteins to look for changes in the R_g and/or changes in the *ab initio* models which are produced from the processed, buffer subtracted, scattering curves.

It is often the case that many of the samples of interest do not tolerate the high sample concentrations required for SANS experiments (typically 5 mg ml^{-1} or greater is preferred). These high sample concentrations are required to distinguish the excess scattering caused by sample from the high background scattering caused by the incoherent scattering of neutrons by hydrogen atoms contained within normal hydrogenated buffer solutions (see Chapter 2, section 2.3.1). Additionally, even if high concentrations can be reached, samples may not tolerate exposure to D_2O containing buffers which can cause the sample to aggregate or precipitate from solution (Berns et al., 1968; Lee and Berns, 1968). As mentioned previously, the issue of aggregation is a major limitation to the SAS field and methods to improve sample data collection are in demand (Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, 2013; Round et al., 2015).

This chapter describes a novel system that has been devised and built to allow SEC to be combined with SANS measurements. In the SEC-SANS setup presented here, several

developments have been made to facilitate in *situ* SEC on the ILL SANS instrument D22. This approach has been applied previously on SAXS instruments – notably at SWING (SOLEIL, L'Orme des Merisiers Saint-Aubin) and BM-29 (ESRF, Grenoble) and (David and Pérez, 2009; Round et al., 2013). This chapter is focused on the experimental system developed on D22 and on the results obtained from SEC-SANS experiments conducted using gel filtration standards of model proteins for validation of the technique. It was also applied to the PfSir2a and PfAlba3 proteins for comparison to the SAXS data and overall model validation.

5.2 Experimental Set-up

The instrument used for the SANS experiments relevant to this chapter was D22 at ILL, Grenoble. This is a SANS instrument with the highest flux at the sample in a wavelength range of 0.4 to more than 4 nm by using the ILL horizontal cold source. It possesses the largest area detector of all SANS instruments (active area 96 x96 cm) which can be moved laterally by 50 cm at a sample to detector distance of 2.4 – 20 meters. This enables D22 to cover a large q range (4 x 10^{-4} to 0.44 $Å^{-1}$).

Firstly a modified cuvette for measuring the samples in continuous flow cell was designed (see Figure 5.1). Here a Suprasil quartz cell of 1 mm sample thickness was placed in the "stopped-flow head" (Grillo, 2009). This was placed directly after a *Superdex* 75 column 10/300 GL to separate the proteins by molecular weight. The samples were manually injected and then sequentially passed through the size exclusion column using an ÄKTAPrime FPLC system (GE Healthcare, Cat no. 18-1135-24) (see Figure 5.1).

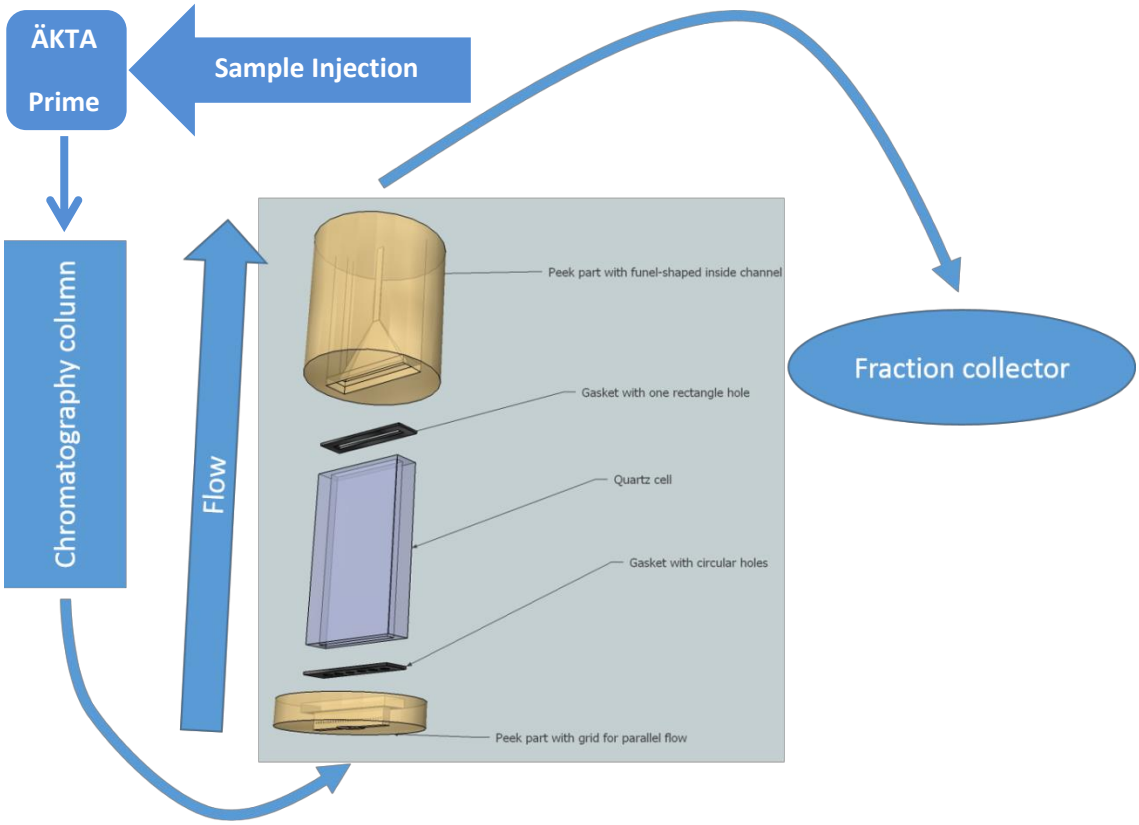


Figure 5.1: Schematic diagram for SEC-SANS experimental setup. Sample injection and constant buffer flow is provided by an ÄKTAPrime FPLC system. This is connected in series to a SEC column which separates proteins contained in the sample based on size. The separated sample enters the quartz flow cell which is placed in the path of the incoming neutron beam to record small angle scattering data on the sample as it passes through the cell. After measurement, the sample is eluted in the fraction collection tray for recovery.

This served to provide a constant flow rate of 0.3 ml min^{-1} throughout each experiment as this was determined to be optimal for appropriate separation of the proteins used and ensured that the samples passed through the flow cell for a sufficient period of time (30 seconds per frame) for each scattering measurement. The UV_{280nm} detector on the ÄKTAPrime system was used to monitor the SEC profile of the proteins after they had passed through the SEC column and the flow cell. The system setup was mounted directly on the experiment platform available at the D22 beamline (see Figure 5.2).

SEC-SANS measurements were carried out using a neutron wavelength of 6 Å ± 10%, and detector distances of 8 m, 4 m or 2 m, a rectangular collimation system of 55 mm x 40 mm having the same length as the sample-detector distance and a rectangular sample aperture of 7 mm x 10 mm. The data recorded were reduced using GRASP software (Dewhurst, 2008) and analysed with Igor software (Kline, 2006). Several corrections were required to correctly process the data including measurements made of the blocked beam, for background subtraction, empty cell scattering, transmission and thickness scaling, absolute intensity calibration using the direct beam intensity and buffer subtraction.

For SEC-SANS the transmission of the buffer, recorded during the elution of the column dead volume, was used for scaling the protein signal. Short 30s exposures were averaged, after normalisation for sample concentration, to increase the signal to noise ratio as necessary. Measurements were performed in SEC buffer described previously (see Chapter 2, section 2.1.14) using100% H_2O or 100% D_2O solvent.

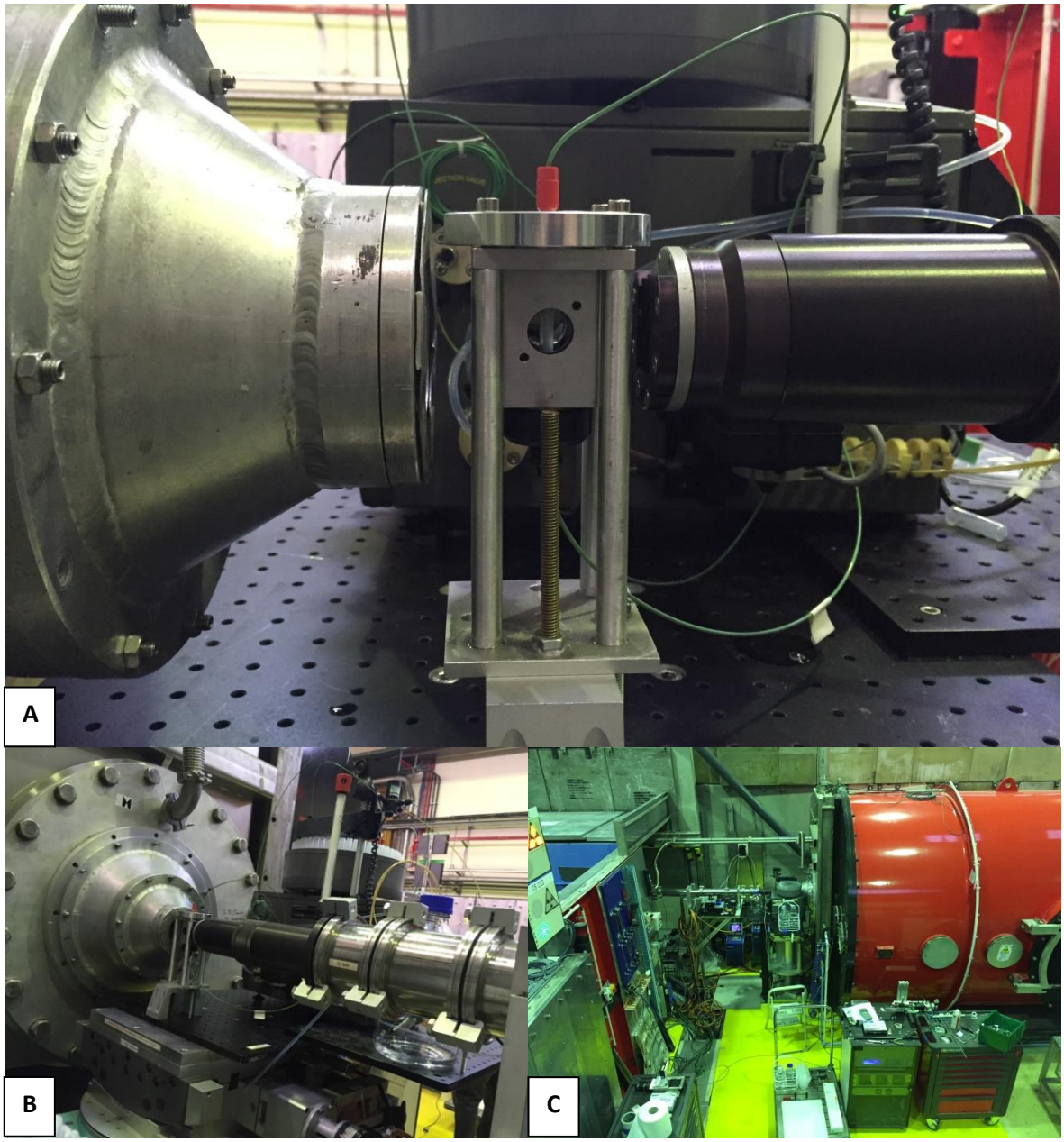


Figure 5.2: (A) Stopped-flow head containing 1mm Suprasil quartz cuvette modified for continuous flow. A window to allow the neutron beam to pass through the quartz flow cell only is present on each side of the flow cell (B) Experimental set-up of ÄKTAPrime and SEC column directly adjacent to the D22 beamline. The neutron beam from the ILL reactor source enters from the right, interacts with the sample contained in the flow cell and enters the detector chamber on the left for recording of the scattering pattern. (C) SEC-SANS system centered in D22 beamline. The detector chamber is located on the right in red.

5.3 Testing And Validation Of The SEC-SANS System Using A Mixture Of Model Proteins

The testing and validation of the SEC-SANS system was carried out using model SEC standard proteins (BioRad, Cat no. 151-190). This sample contains a mixture of bovine Thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken Ovalbumin (44 kDa), horse Myoglobin (17 kDa) and Vitamin B_{12} (1.35 kDa). One 18 mg vial of protein, resuspended in 0.5 ml of buffer, was used per SEC experiment, leading to a total concentration of 36 mg ml^{-1}.

Figure 5.3 shows the initial test results of the SEC-SANS system for SEC standard proteins measured in 100%-H_2O-containing solvent. The average *I(0)* for each frame recorded during the experiment was overlaid against the calculated R_g of the proteins as they passed through the SEC column. It can be seen that there is a clear change in R_g between protein peaks although the data is very noisy and does not appear well separated. Additionally a change in the observed average *I(0)* can be seen as the sample migrates though the column and the flow cell, which is distinguished from the high background scattering of the 100% H_2O containing buffer.

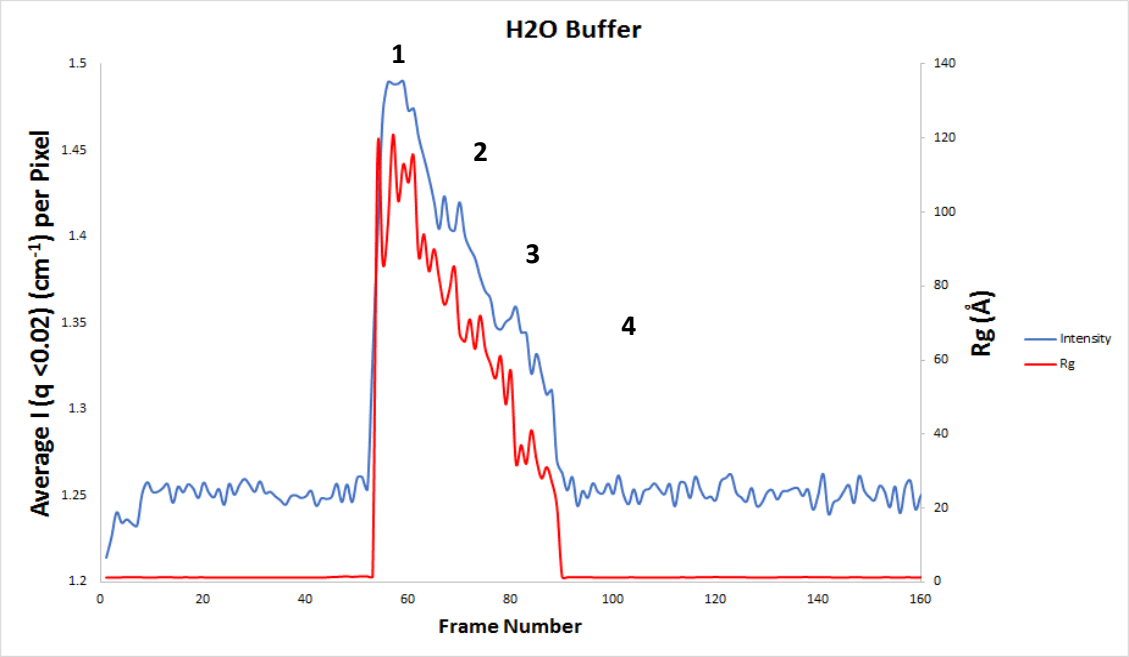


Figure 5.3: Plot of average *I(0)* and determined R_g during SEC-SANS run in 100% H_2O buffer solution (Jordan et al., 2016). Elution Peaks (1) 670 kDa Thyroglobulin and 158 kDa γ-globulin, (2) 44 kDa chicken ovalbumin, (3) 17 kDa horse Myoglobin and (4) 1.35 kDa vitamin B12.

The UV_{280nm} profile for the SEC standards yields four distinct peaks (data not shown). The first peak is the void volume peak containing the 670 kDa Thyroglobulin and 158 kDa γ-globulin proteins. The second peak contains the 44 kDa chicken ovalbumin and the third peak contains the 17 kDa horse Myoglobin. The final fourth peak is that of vitamin B12 1.35 kDa.

While an increase in *I(0)* is observed due to the protein passing through the column and cuvette during the measurements, the peaks are not as well defined as the UV_{280nm} profile. One large peak is observed as the higher MM components of the SEC standard sample elutes, which begins a gradual decline in intensity as the proteins exit the measuring cuvette. We expect this decline in *I(0)* as the scattering power of the smaller protein molecules is not as

strong as that of the larger ones due to the reduction in atomic mass and therefore scattering centres contained in each molecule.

The reason for the poorly defined peaks in the observed scattering from the sample is likely to relate to overlapping contributions from different elution peaks and the limited column resolution. As a result, the scattering contributions of the larger proteins (670 kDa Thyroglobulin and 158 kDa γ-globulin has a significant impact on the observed *I(0)* for the smaller protein molecules due to the slight overlap in the way that the proteins elute through the column. This can lead to increased observed *I(0)* for the proteins, which affects the calculations for molecular mass and R_g.

Additionally, in 100% hydrogenated buffer, the observed increase in scattering is relatively low due to the large incoherent background signal arising from hydrogen atoms in the buffer solution. This results in weak signal visibility and small protein molecules, which do not scatter strongly at relatively high concentrations, can become indistinguishable from the background scattering.

In this case it was not possible to determine accurately the MM or R_g for the samples contained within the first peak of this experiment due to the mixture of the Thyroglobulin and γ-globulin species present in the void volume of the column. The resolution of the column used also meant that the MM and *R*g values for the Ovalbumin protein eluting in peak 2 were found to be larger than expected because of the scattering by larger MM proteins in the tail of elution peak 1.

A repeat experiment was performed using 100%-D_2O-containing buffer as a comparison for the technique. The results of this experiment can be seen in Figure 5.4. The average *I(0)* for each frame recorded during the experiment was overlaid against the calculated R_g for each frame of the experiment. Figure 5.4 shows again that there is a clear separation between the

protein peaks which is expected for SEC standards. Additionally a change in the observed average *I(0)* can be seen as the sample migrates though the column and the flow cell, which is very distinguished from the lower background scattering of the 100% D_2O containing buffer.

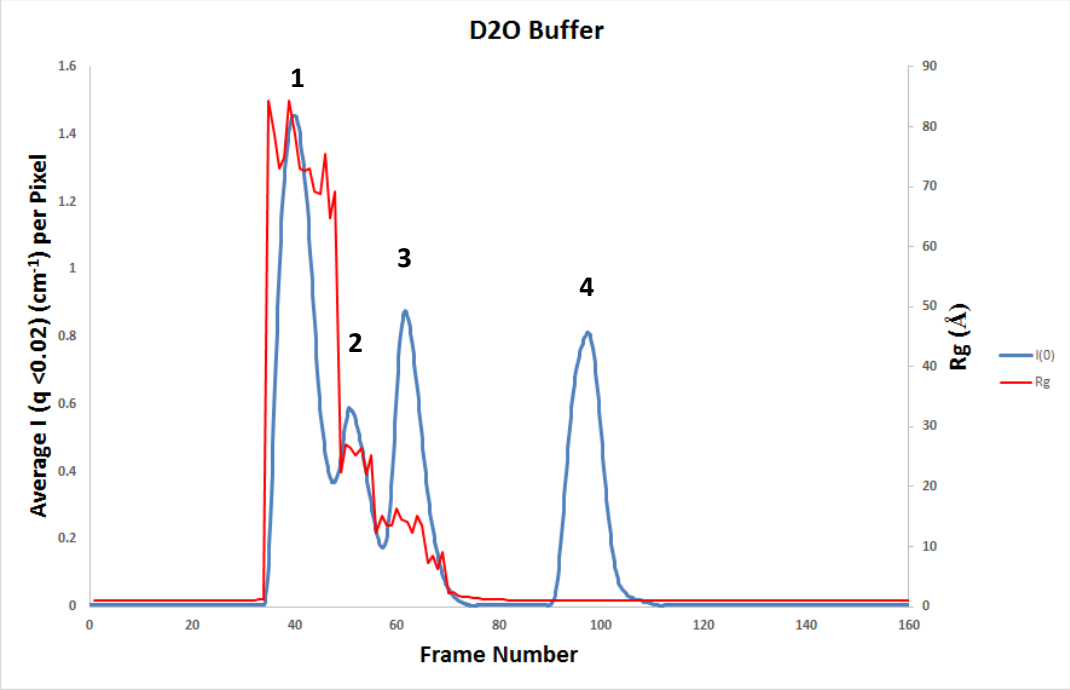


Figure 5.4: Plot of average *I(0)* and determined R_g during SEC-SANS run in 100% D_2O buffer solution (Jordan et al., 2016). Elution Peaks (1) 670 kDa Thyroglobulin and 158 kDa γ-globulin, (2) 44 kDa chicken ovalbumin, (3) 17 kDa horse Myoglobin and (4) 1.35 kDa vitamin B12.

Several improvements are seen as a result of changing the buffer composition from 100% H_2O to 100% D_2O content. Firstly, the overall background scattering level from the buffer is lower due to the lower incoherent scattering of deuterium (essentially negligible by comparison with that of hydrogen). Secondly, the observed peaks of scattering intensity as the proteins pass through the column and into the sample cell are much more clearly defined. This scattering profile more closely resembles the protein elution profile as observed by the

UV 280_{nm} readings. Improvements in the calculated R_g for each protein peak can also be seen due to the improved separation of the intensity peaks for the proteins as they pass through the column.

When analysing the data it was found that the polydispersity problem, associated with the first elution peak, remained due to both the 670 kDa Thyroglobulin and 158 kDa γ-globulin protein molecules both eluting in the void volume of the column. Values for the protein molecules eluting in the second and third peak can be seen in Table 5.1.

	Concentration Determination			Theoretical values (BSLDC)					Experimental values (GRASP)	
Param.	**MW**	**Extinction Coef.**	**Abs 280nm**	**Conc.**	**Spe. Density**	**SLD**	**Contrast**	**$I_{(Q=0)}$ 90% labile H exch.**	$I_{(Q=0)}$	**Rg**
Units	**g/mol**	**$(cm.mg/cm^3)^{-1}$**	**AU**	**g/cm^3**	**g/cm^3**	**cm^{-2}**	**cm^{-2}**	**cm^{-1}**	**cm^{-1}**	**Å**
Ova.	44000	0.7	0.272	0.0019	1.37	2.97E+10	-3.4E+10	0.088	0.1417 +/-0.001	24.0 +/-0.35
Myo.	17000	0.82	0.433	0.0026	1.35	2.95E+10	-3.42E+10	0.048	0.04011 +/- 0.0005	14.0 +/-0.4
Buf.						6.38E+10				

Table 5.1: Theoretical (calculated using BSLDC calculator (Myatt) in green) and experimental (fitted using GRASP (Dewhurst) software, in red) structural parameters derived from the results of in situ SEC-SANS analysis of the BioRAD SEC calibration standards in 100% D2O buffer (Jordan et al., 2016). MW: molecular weight; Extinction coef.: extinction coefficient; Abs: absorbance; Conc: concentration; Spe density: specific density; SLD: scattering length density; I(Q=0): scattering intensity extrapolated at Q = 0; Rg: radius of gyration.

Here five exposures of 30s each, taken symmetrically from the top of the peak, were averaged to build the ovalbumin and Myoglobin SANS curve. The absorbance at UV_{280nm} measured on the same time scale was averaged for the estimation of concentration. The merged data for the buffer and Myoglobin frames, as well as the subtracted data, are shown in Figure 5.5.

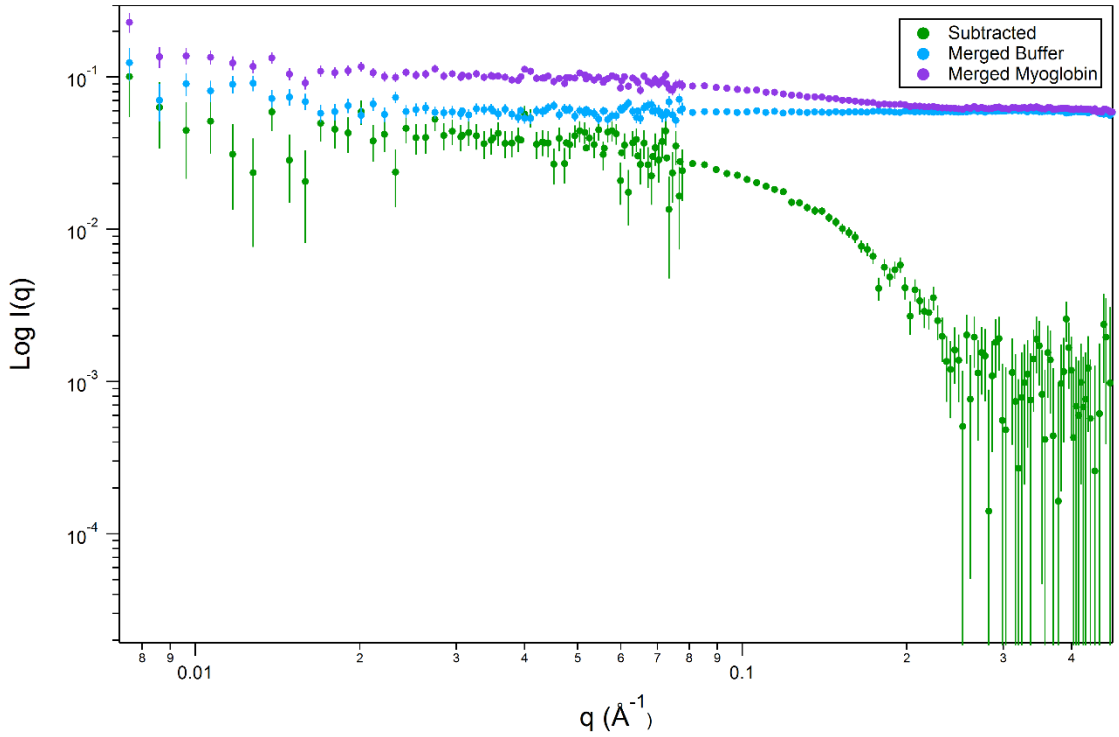


Figure 5.5: Plot of Myoglobin processed experimental scattering curve. Merged Myoglobin frames (Purple), merged buffer frames (Blue) and subtracted data (Green).

These results show that the experimentally determined values of $I_{(Q=0)}$ and *Rg* for Myoglobin are consistent with those in the literature (Goldenberg and Argyle, 2014). However, the values calculated for Ovalbumin are too high. A contribution from another of the proteins is unlikely since UV absorbance profiles show the correct separation. The effect may arise from the presence of dimers or aggregates formed at the exit of the column – such behavior has been noticed previously for this protein (Ianeselli et al., 2010).

By taking the processed data for the Myoglobin from the SEC-SANS curve, basic structural information about the R_g and the *p(r)* can be determined. Using these parameters, an *ab inito* model was generated for the protein in solution. This can be viewed in Figure 5.6. Here we show the SANS envelope determined by DAMMIF for horse Myoglobin.

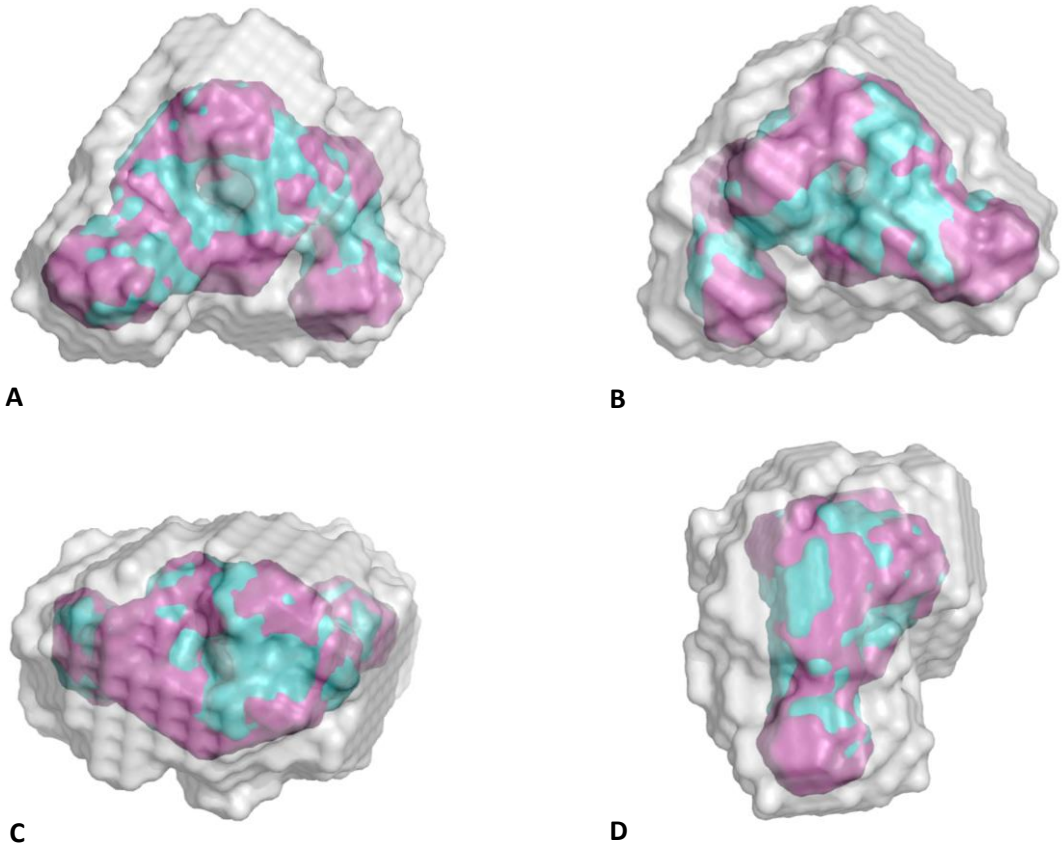


Figure 5.6: Results of *Ab initio* modelling for horse Myoglobin showing (A) Side view, (B) Alternate side view, (C) Top View and (D) end view. DAMAVER averaged envelope of all models is shown in grey. DAMMIF filtered model is shown in cyan. The most representative model (MRM) is shown in pink.

The model with the lowest reported NSD value was then compared to the crystal structure of horse Myoglobin (PDB ID 1WLA) and aligned in Figure 5.7.

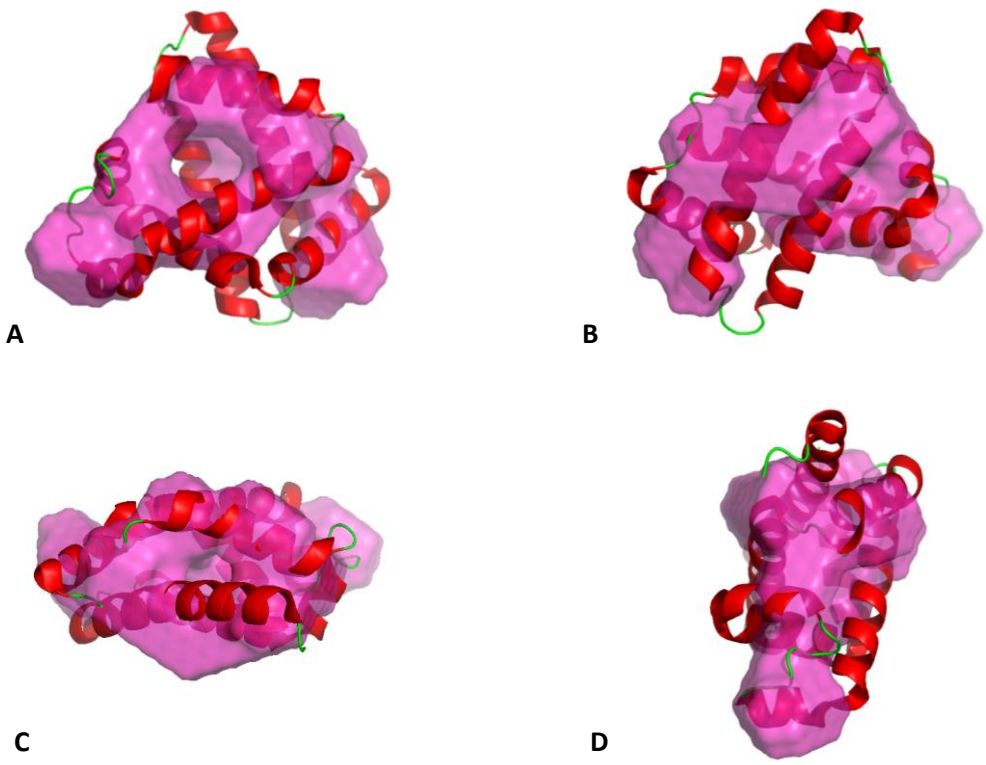


Figure 5.7: Results of *Ab initio* modelling for horse Myoglobin showing (A) Side view, (B) Alternate side view, (C) Top View and (D) end view. MRM envelope model is shown in red, horse Myoglobin crystal structure (PDB 1WLA) is shown in cartoon.

The fit of the scattering data for horse Myoglobin to the MRM as determined by CRYSOL is shown in Figure 5.8. Whilst we have demonstrated that it is possible to determine the R_g of proteins using the SEC-SANS method, the fit of the *ab inito* model to the pre-existing crystal structure data is not good, with a chi-square value of 12.21 reported. In this case, the combination of weak scattering of the small horse Myoglobin molecule above that of the background (see Figure 5.5) in addition to the very short measurements for each separate SEC-SANS frame leaves some uncertainty in the measurements reported. This could be improved with higher sample concentrations and slower flow-rates of the SEC system to improve data quality.

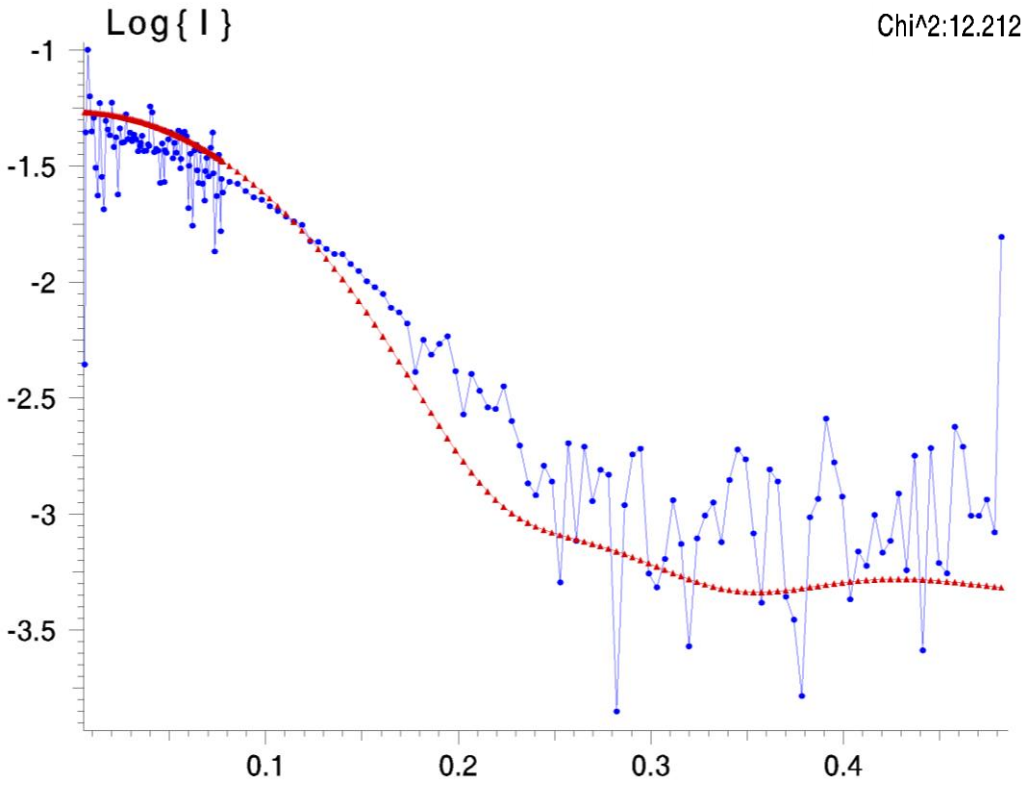


Figure 5.8: CRYSOL output of horse Myoglobin experimental scattering data in 100% D_2O buffer (blue) fit with theoretical scattering curves for MRM (red).

5.4 Deuteration Of PfSir2a And PfAlba3 Systems For SEC-SANS

After demonstrating the feasibility of the SEC-SANS system for determination of solution-based protein structural information, a series of experiments were conducted using recombinant PfSir2a and PfAlba3. This was aimed at investigating whether an interaction could be observed between the recombinant proteins, which were produced in a deuterated environment to enable SANS contrast match-out experiments to take place. This would be used to investigate any changes in structural characteristics of the proteins of interest.

To facilitate these experiments, the production of 75% deuterated PfSir2a and PfAlba3 was performed as described in Chapter 2, section 2.3.3. The results of the adaptation to deuterated minimal media for PfSir2a and PfAlba3 can be seen in Figure 5.9. The optimal growth temperature for expression and solubility of the deuterated forms of PfSir2a and PfAlba3 was found to be 25˚C and 20˚C respectively. Figure 5.9 shows that while expression was achieved for PfSir2a in the minimal media (M.M) and the deuterated minimal media (D M.M) samples, the level of expression was greatly reduced and there is little to no PfSir2a in the soluble fraction. For PfAlba3, we see a similar case where the expression is reduced because of adaptation and growth in both the minimal media environments. However, a relatively strong level of expression is maintained which is also fairly soluble.

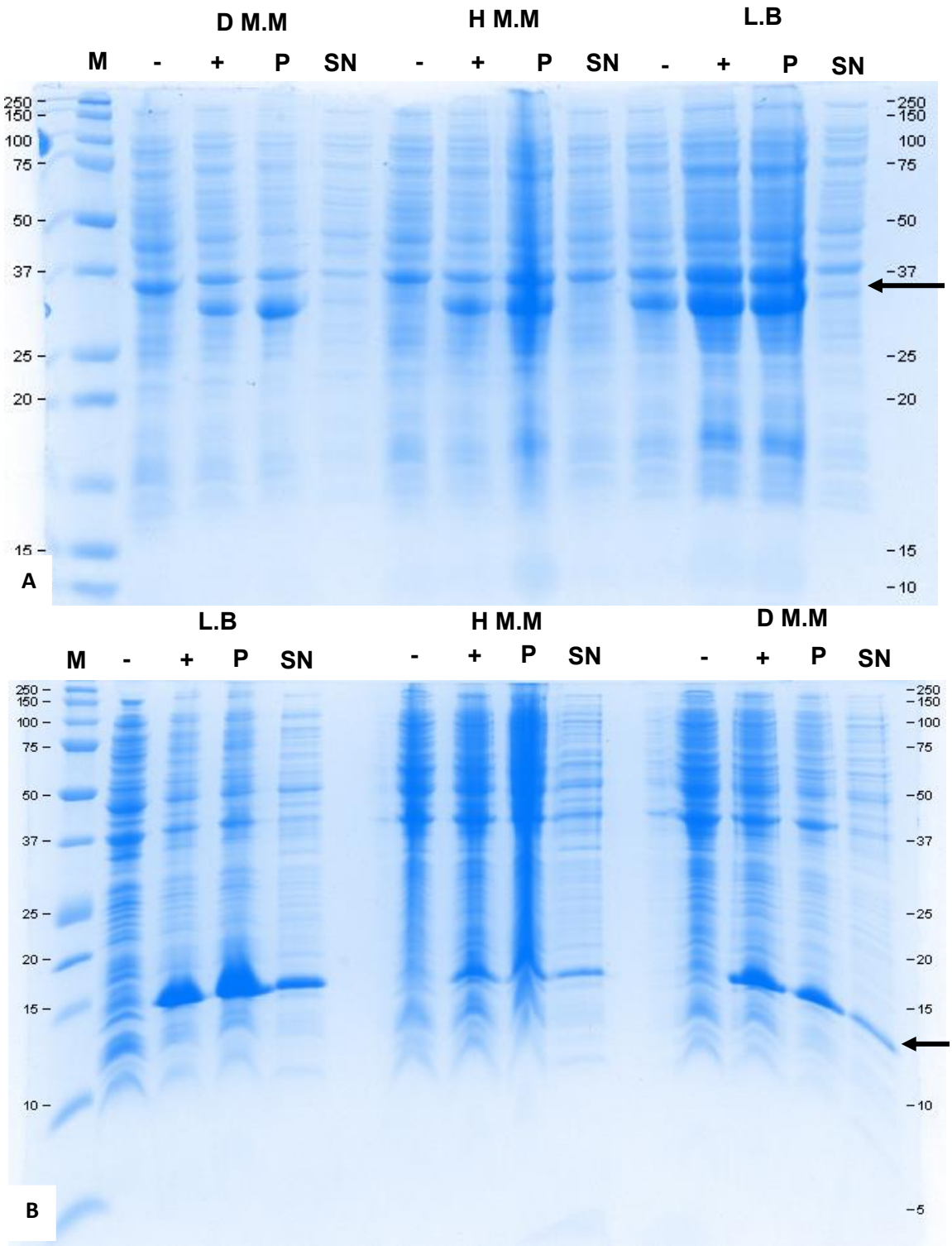


Figure 5.9: (A) PfSir2a expression and solubility tests after adaptation to minimal growth media (M.M) and deuterated minimal growth media (D M.M) at 25˚C with LB expression for comparison. (B) PfAlba3 expression and solubility tests after adaptation to minimal growth media (M.M) and deuterated minimal growth media with L.B expression for comparison. Arrows indicated PfSir2a and PfAlba3 in respective gels.

Large scale production of deuterated biomass for the expression and purification of both PfSir2a and PfAlba3 was performed using the growth conditions determined from the expression and adaptation tests using fermenter cultures as described in Chapter 2, section 2.3.3. The deuterated proteins were purified and recovered by the same method as used for the hydrogenated proteins (see Chapter 2, sections 2.1.11 and 2.1.12). The purification results for deuterated PfSir2a can be seen in Figure 5.10. A decrease in the amount of soluble PfSir2a is observed in the elution fractions and the yield is greatly reduced from that of the hydrogenated recombinant PfSir2a (see Chapter 3, section 3.3.1). While this small amount of PfSir2a could be recovered from the purification process, further purification, IEX and SEC resulted in precipitation of aggregated protein which could not be recovered.

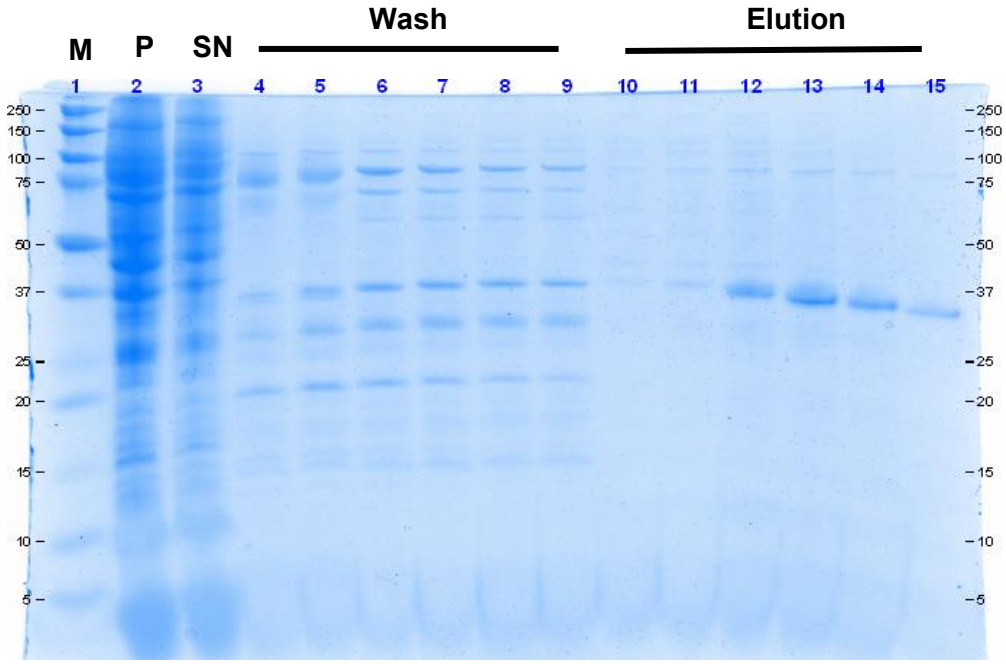


Figure 5.10: 12% SDS-PAGE gel analysis of deuterated PfSir2a Ni-NTA purification steps. M indicates molecular weight marker, P & SN indicate the insoluble and soluble fractions of the induced protein from the *E.coli* bacterial culture. Wash and elution steps labelled respectively.

The purification results for deuterated PfAlba3 can be seen in Figure 5.11. A large amount of soluble deuterated PfAlba3 is recovered from the deuterated fermentation culture cell pellet after lysis. This is comparable to the yield which was observed for the hydrogenated recombinant form seen in Chapter 3, section 3.3.2. It can be seen that the elution fraction bears significantly more contaminants after the purification process, due to increased overall protein synthesis in high cell density culture conditions, which must be removed by further purification.

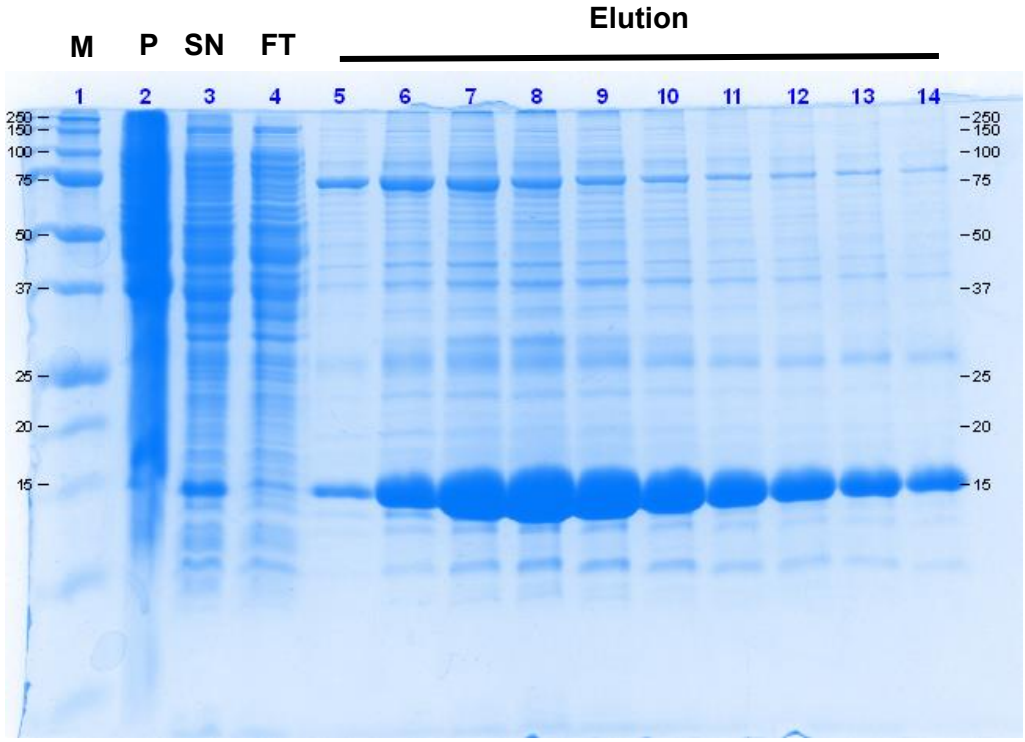


Figure 5.11: 16.5% SDS-PAGE gel analysis of deuterated PfAlba3 Ni-NTA purification steps. M indicates molecular weight marker, P & SN indicate the insoluble and soluble fractions of the induced protein from the *E.coli* bacterial culture. The column flow through (FT) and recovered elution fractions are labelled respectively.

Further purification of deuterated PfAlba3 was performed by SEC using a *Superdex 75* 16/600 PG column using the same buffer conditions as described previously and the results of the SEC profile can be seen in Figure 5.12. Here we see three peaks in the SEC profile instead of the previously observed two. The first peak represents aggregated PfAlba3 and larger MW contaminants left over from the purification process. The second peak is non-aggregated PfAlba3 which possesses a small shoulder at the beginning of the peak which contains PfAlba3 and a second contaminant at approximately 25 kDa. The remainder of the second peak shows PfAlba3 only. The third peak appears to show a very faint band of protein at the same MW as PfAlba3 but eluting much later from the column.

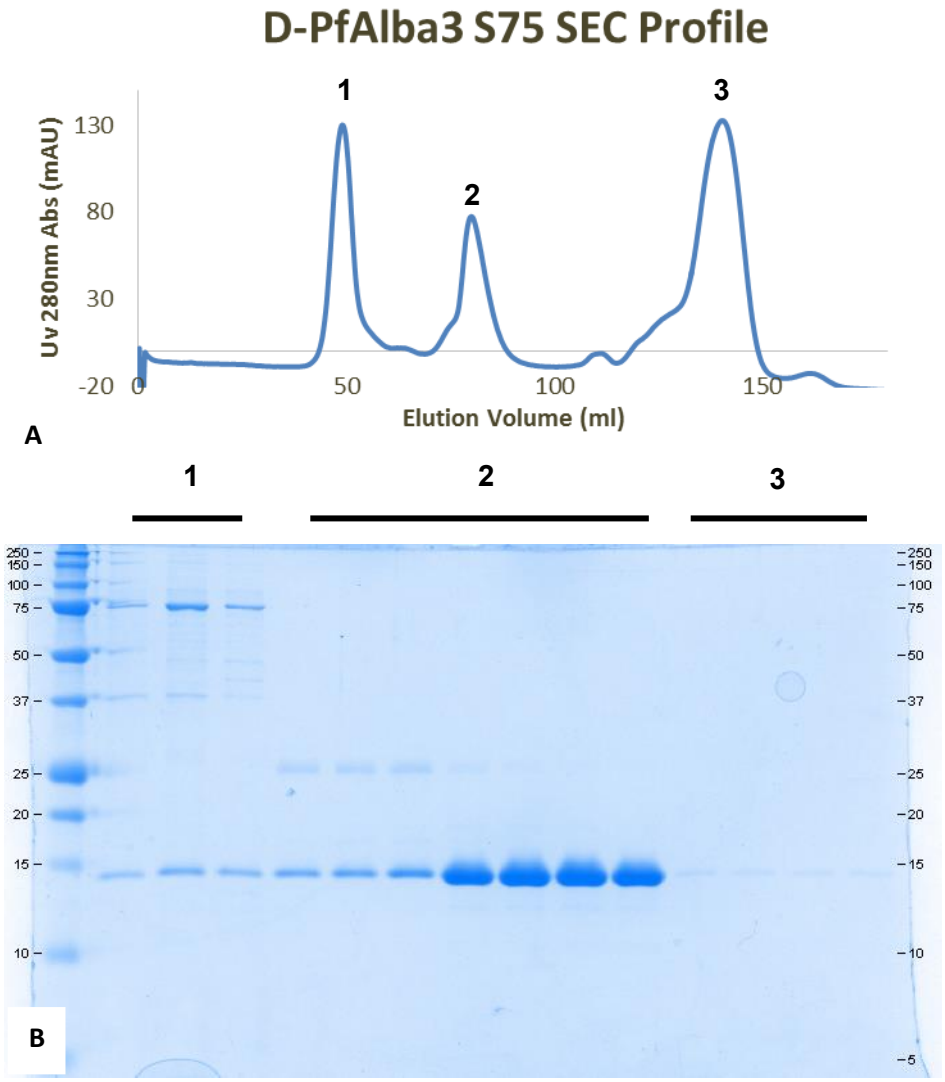


Figure 5.12: (A) SEC profile of deuterated PfAlba3 showing three distinct elution peaks. (B) The first peak (1) represents aggregated protein and high MW contaminants. The second peak (2) contains non-aggregated PfAlba3. The third peak (3) shows a trace amount of PfAlba3 or similarly sized contaminant eluting at a delayed volume – the exact nature of this is not understood.

The purified deuterated PfAlba3 fractions from the second peak of the SEC profile were concentrated to 1 mg ml^{-1} and prepared for analysis by DLS. The results of this can be seen in Figure 5.13. Here a single monodisperse peak for deuterated PfAlba3 with an observed R_H of 7.07 ± 1.33 nm with a polydispersity index of 18.5%. This agrees with the previously observed value of hydrogenated PfAlba3 (see Chapter 3, section 3.6.2). A similar tendency towards aggregation and precipitation of the deuterated PfAlba3 was also observed despite purification and storage of the protein at 4˚C.

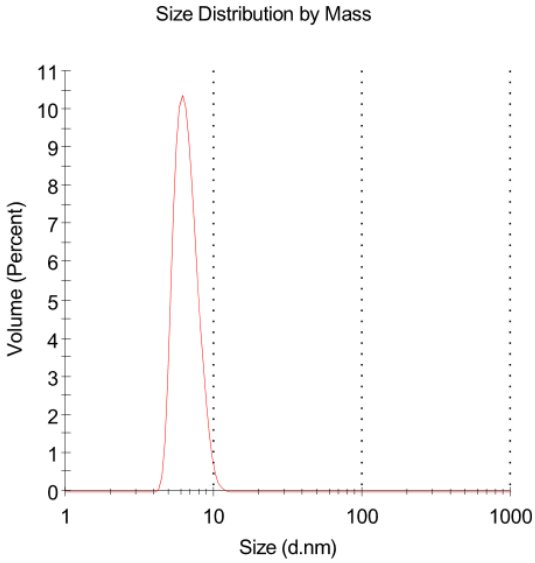
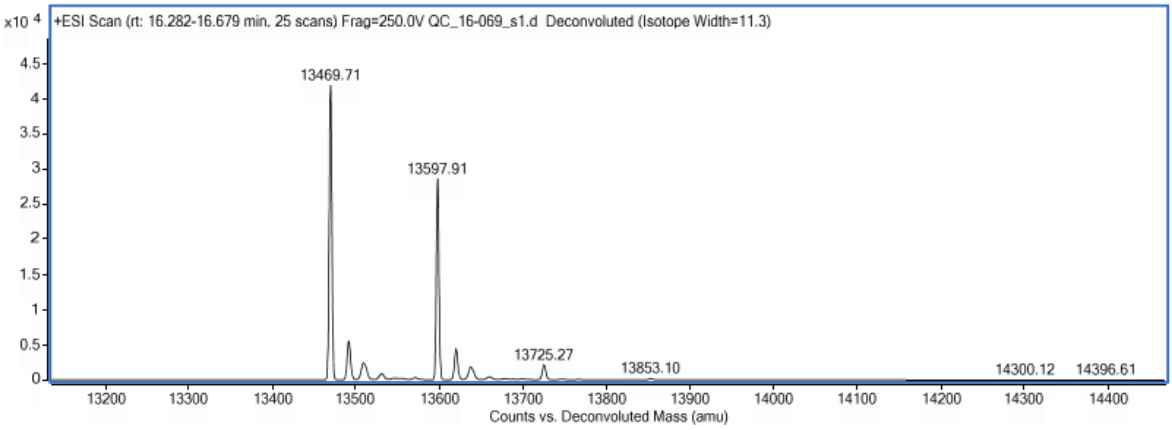


Figure 5.13: DLS results for PfAlba3 in sodium phosphate buffers with 500mM sodium chloride at 4˚C.

Deuterated PfAlba3 was also analysed by TOF-MS for molecular weight and percentage deuteration determination. The results from the TOF-MS experiment can be seen in Figure 5.14. Here we see two observed masses of 13469.71 Da and 13597.91 Da for the deuterated PfAlba3. Using values calculated for the molecular weight of the hydrogenated protein, 13040.97 Da, and the fully deuterated, 13745.97 Da, we find that the final percentage of deuteration is 60.81% and 78.9% respectively. The difference in mass between the two

values, 129.2 Da, suggests that the lower mass of the 13469.71 represents the loss of the N-terminal methionine residue (see Chapter 3, section 3.4).



Sample	Expected Mass (Da)	Observed Mass (Da)
D-PfAlba3	13040.97 13745.97	13469.71 13597.91

Figure 5.14: Deconvoluted TOF-MS spectrum results for deuterated PfAlba3.

A repeat pulldown experiment was performed as described previously (see Chapter 3, section 3.9.2) using hydrogenated PfSir2a after His-tag cleavage and deuterated PfAlba3. The results of this experiment can be seen in Figure 5.15. Here we again see that tagged deuterated PfAlba3 is not completely binding to the resin but is still present in detectable quantity in the elution lane. For the co-incubated lanes a large proportion of cleaved PfSir2a is removed in the supernatant, indicating it is not bound to the resin or PfAlba3. A smaller quantity is removed in both wash lanes and is hardly visible in lane W3. In the elution lane we see a similar quantity of PfAlba3 as compared to that eluted in the PfAlba3-only incubation.

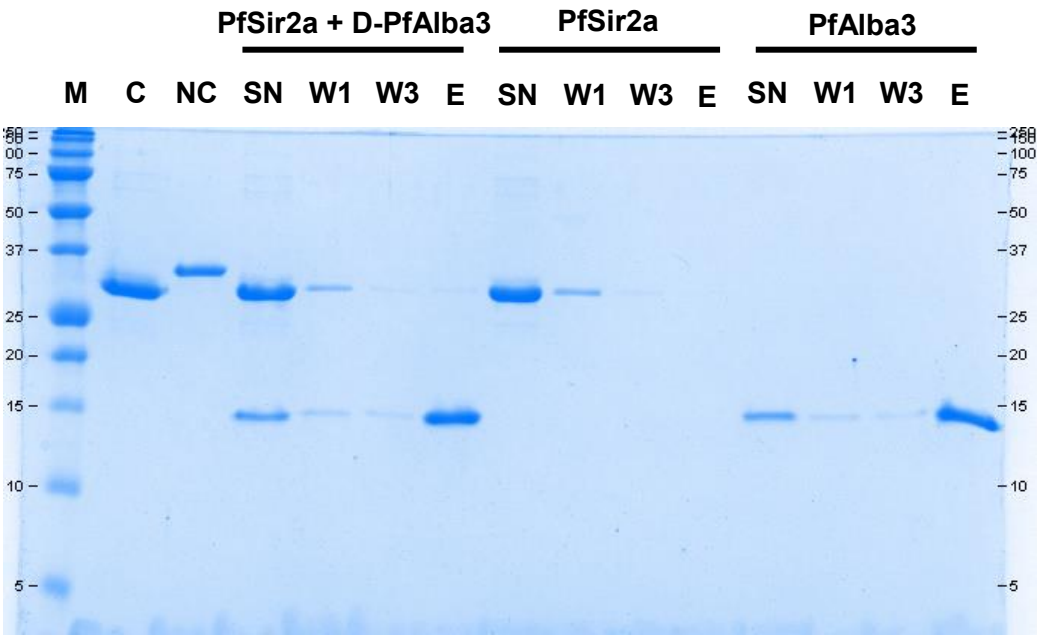
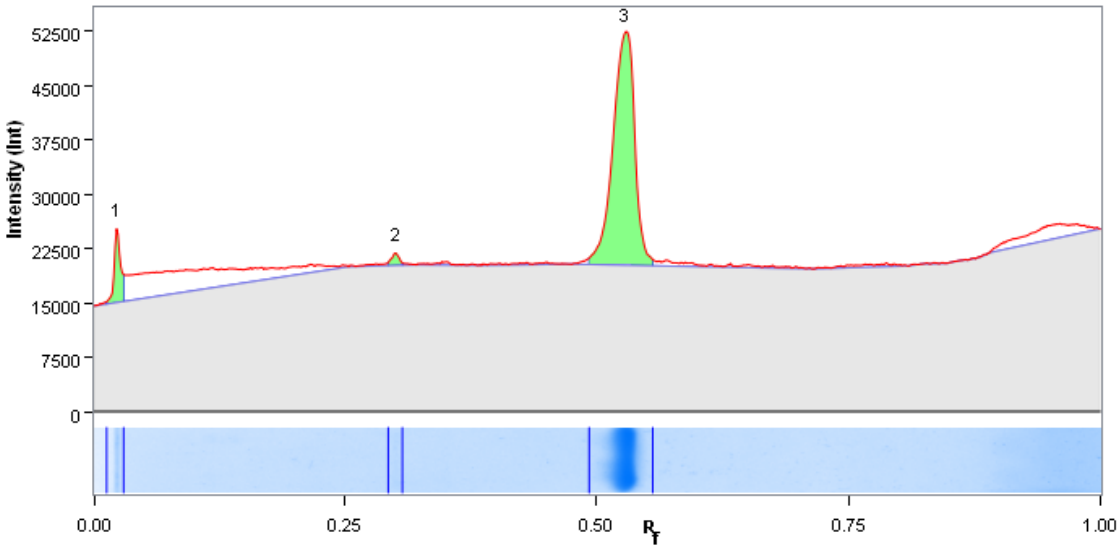


Figure 5.15: 16.5% SDS-PAGE showing the results of a repeat Ni-NTA resin pulldown experiment with cleaved PfSir2a & tagged D-PfAlba3. M indicates molecular weight marker, C & NC indicated cleaved and non-cleaved PfSir2a respectively. Supernatant (SN), 1st & 3rd wash (W1 & W3) & elution (E) fractions are indicated for reactions containing cleaved PfSir2a & tagged D-PfAlba3, cleaved PfSir2a alone & tagged D-PfAlba3 alone.

An extremely faint band at the MW migration of cleaved PfSir2a is also detected in this elution lane. The intensity for each protein can be observed by lane analysis (see Figure 5.16). This band is not observed in the elution sample lane for the cleaved PfSir2a only. The amount of cleaved PfSir2a is found to be < 2% of the total lane indicating that if there is an interaction between PfSir2a & D-PfAlba3 it appears to be extremely weak.



Band No.	Mol. Wt. (KDa)	Relative Front	Volume (Int)	Band %	Lane %
1	206.4176	0.023876	3752920	8.449277	4.742611
2	29.24454	0.300562	766836	1.726445	0.96906
3	14.13174	0.52809	39897300	89.82428	50.41871

Figure 5.16: Lane profile analysis of cleaved PfSir2a & D-PfAlba3 incubation elution sample with MW of observed species and relative intensities.

5.5 Application Of SEC-SANS To PfSir2a And PfAlba3

Due to the weakly interacting nature of PfSir2a and PfAlba3, contrast matchout measurements aimed at investigating the proposed complex as a whole and the two components individually were not viable. Given their sensitivity to radiation damage by X-rays and their aggregation or self-interacting nature, several experiments were conducted to investigate their solution structure properties utilising the SEC-SANS system. For PfSir2a, an investigation into how the R_g of the protein changes as it migrates through the SEC column was performed.

Figure 5.17 shows the results of this experiment using the hydrogenated PfSir2a and a 100% D_2O buffer of the same buffer, salt, and pH conditions as described previously in the related SAXS experiment (see Chapter 4, section 4.4.2). In panel A we see the overlay of the SEC UV_{280nm} trace with that of the azimuthally averaged *I(0)*of the recorded scattering pattern

from the sample as it passes through the SEC column. An interesting feature is a large increase in intensity observed after approximately 70 minutes, due to the buffer exchange of the sample prepared in 100% H_2O with the column buffer containing 100% D_2O.

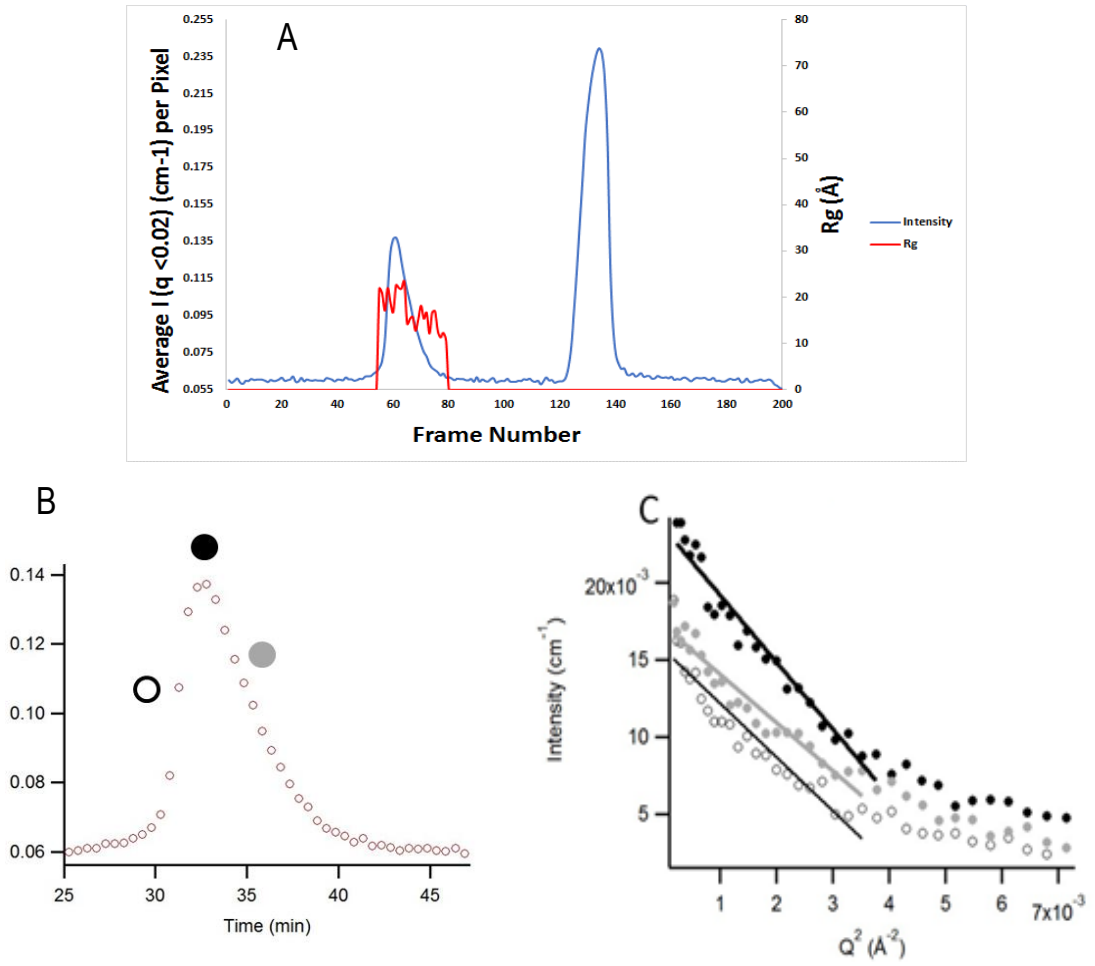


Figure 5.17: A: Results of the SEC-SANS measurement of protonated Sir2a (4 mg ml^{-1}) in 100% D_2O buffer, showing the absorbance of the sample at 280 nm (in blue), and the averaged SANS intensity between Q = 0.012 and 0.041 $Å^{-1}$, (in red). B: magnified plot of SANS intensity of PfSir2A elution peak with symbol for individual curves position. C: Guinier plot of individual SANS curves (30 sec exposure) selected at the beginning (open circles, Rg = 20.2 +/- 0.5 Å), the top (black circles, Rg = 21.3 +/- 0.4 Å) and the end (grey circles, Rg = 18.9 +/- 0.5 Å) of the elution peak (Jordan et al., 2016).

Panel B shows the magnified plot of the SANS intensity curves, of which three were subjected to Guinier analysis which can be seen in panel C. This was performed by subtracting the average buffer scattering taken from the beginning of the SEC run and subtracting it from the peak intensity curves. Here we find that there is a change in the observed R_g for PfSir2a as it migrates through the SEC column starting at R_g = 20.2 +/- 0.5 Å (open circles) before increasing to R_g = 21.3 +/- 0.4 Å (black circles) and later falling to R_g = 18.9 +/- 0.5 Å (grey circles).

These values are lower than those calculated for PfSir2a by SAXS due to the scattering of neutrons by the internal atomic structure rather than the surrounding electron cloud. This marks a significant improvement over the results obtained via SEC-SAXS as there is no radiation damage causing protein aggregates to persist in the sample cell and requiring significant optimization steps and influencing the experimentally derived results (see Chapter 4, section 4.4.1).

SEC-SANS results for PfAlba3 can be seen in Figure 5.18. Here we see the SEC 280_{nm} elution profiles of deuterated PfAlba3 measured in 100% H_2O buffer conditions (panel A) and hydrogenated PfAlba3 measured in 100% D_2O buffer conditions (panel B). The corresponding measured *I(0)* for each 30s time point over the course of the experiment is also plotted for each panel. The subtracted data for the PfAlba3 aggregates (grey open circles) and the non-aggregated PfAlba3 (black circles) can be seen inserted above the UV/Intensity profiles. Here we can clearly see the difference in scattering curves for PfAlba3 by removing the large protein aggregates which show a sharp upswing in intensity at low Q. Using the data from the separated non-aggregated protein peak we find a R_g of 19.06 ± 0.5 Å and 20.03 ± 0.4 Å for the deuterated and hydrogenated PfAlba3 respectively.

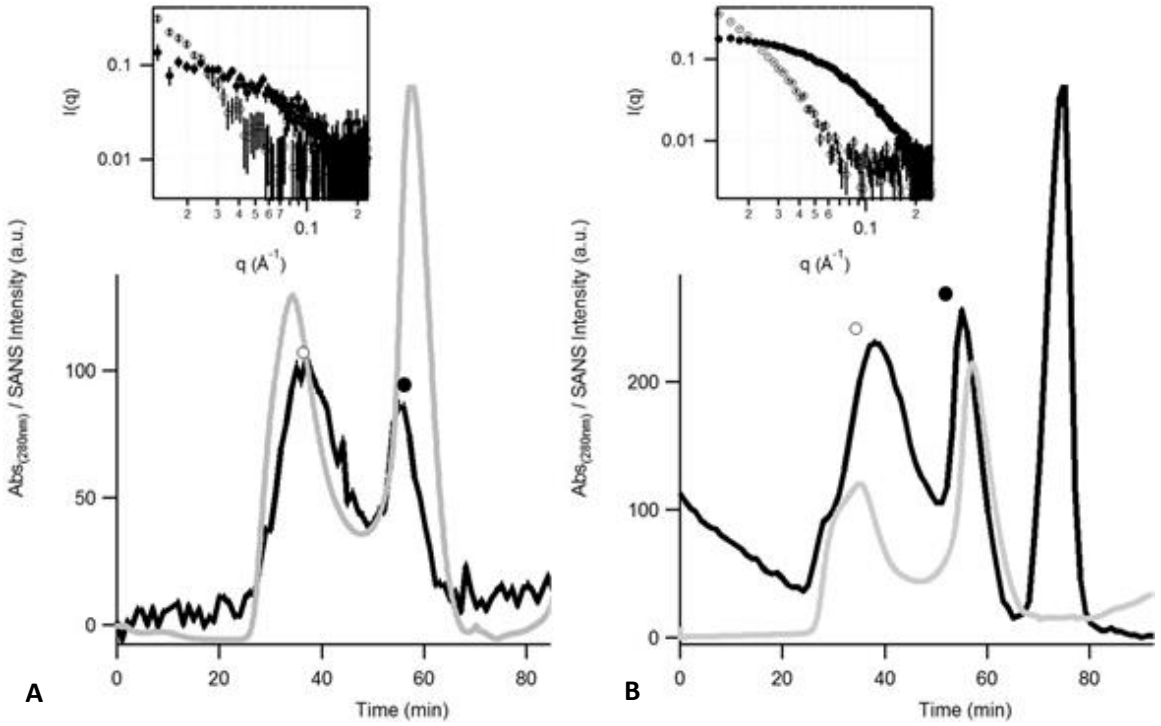


Figure 5.18: (A) Results of the SEC-SANS measurement of deuterated PfAlba3 at 5 mg ml^{-1} in 100% H_2O buffer, showing the absorbance of the sample at 280_{nm} (grey), and the averaged SANS intensity (black). (B) Results of the SEC-SANS measurement of hydrogenated PfAlba3 at 5 mg ml^{-1} in 100% D_2O buffer, showing the absorbance of the sample at 280_{nm} (grey), and the averaged SANS intensity (black) (Jordan et al., 2016).

5.6 Summary

This chapter describes a new method for bio-SANS measurements, which has significant practical application for many areas of biology. This technique enables the separation of samples *in situ* at a SANS beamline, facilitating measurements on mixtures of proteins. The observed scattering data can be correlated with the UV_{280} nm absorption profile to determine individual components and concentration correction in post data processing. Importantly, this methodology also allows the persistent problem of protein aggregation to be tackled which can often occur in fragile systems. By performing SEC directly at the beamline many

systems, which are prone to aggregation, can be separated *in situ* to allow for data collection on the macromolecules of interest as neutron collection rates are much slower compared to that of synchrotron sources. This is especially useful when macromolecules or macromolecular complexes introduced into a non-native D_2O environment that is necessary for contrast matchout experiments for studying protein complexes which can often have a destabilising effect. It also allows for comparative measurements to be made using the same experimental conditions on SEC-SAXS beamlines.

This method also offers an advantage over that used in bio-SAXS studies by avoiding radiation-damage-induced protein aggregation. This problem can have a significant impact on the results obtained for samples and can hamper data processing. There are still some limitations with this system in that relatively high sample concentrations are still required for optimal results. This is due to the high incoherent background scattering of 100% H_2O buffers and the limited window in which samples pass through the SEC column and measuring cell due to the column flow rate.

Chapter 6: Discussion, Conclusions And Future Work

This thesis was intended to investigate the nature of the interaction between the chromatin proteins PfSir2a and PfAlba3 from *P. falciparum* and determine their structural characteristics. A connection established between these proteins indicated possibility of playing a significant role in the antigenic variation immunoevasion strategy employed by malaria parasites to avoid destruction by host immune responses.(Chêne et al., 2012; Goyal et al., 2012; Merrick and Duraisingh, 2007, 2006; Merrick et al., 2012).

This evidence demonstrated that PfSir2a possess several of the important residues necessary for catalytic activity as that of the yeast sirtuin, ScHst2, and was able to deacetylate *Plasmodium* histone substrates in an NAD^+ manner (Merrick and Duraisingh, 2007). Like ScHst2, it was thought that PfSir2a did not directly bind to DNA to perform its silencing role and interacted with other partner proteins to achieve this (Merrick and Duraisingh, 2007). Goyal *et al* demonstrated the interaction of PfSir2a with PfAlba3, which has established DNA and transcriptional repression capabilities (Chêne et al., 2012), using protein pulldown approaches (Goyal et al., 2012). This combined evidence indicated a potential silencing complex between PfSir2a and PfAlba3 is formed to inhibit the transcription of currently active *var* genes, an integral part of the antigenic variation strategy.

While biochemical evidence for the PfSir2a-PfAlba3 interaction was established, structural evidence of the interaction was lacking. This was to be investigated in the project using structural techniques (SAXS/SANS/protein crystallography), supported by data from further biophysical characterisation. It was thought that information obtained from these

techniques would help to understand the nature of the interaction between PfSir2a and PfAlba3, and could yield significant insight into the mechanisms behind the transcriptional silencing activity that takes place in *P. falciparum* to regulate its immune-evasion strategy.

In this thesis, these goals were designed to be achieved in several steps beginning with the production of recombinant forms of both proteins using the pET28a expression system. The proteins were purified and characterised by various methods that are outlined in Chapter 3. The successful production and purification of both PfSir2a and PfAlba3 using the pET28a expression system in *E. coli* BL21 (DE3) strain is demonstrated. Protein expression was optimised using modified bacterial growth conditions to achieve maximum production and recovery of soluble protein via Ni-NTA, IEX and SEC methods. The optimal conditions for solubility and stability of both proteins were found to have a composition of 50mM sodium phosphate, 500mM NaCl, pH 7.0. This relatively high salt concentration was found to reduce the polydispersity of both PfSir2a and PfAlba3 samples as well as increase their stability. Purified PfSir2a and PfAlba3 were determined to be correctly produced in the full length recombinant forms as seen by TOF-MS measurements.

PfSir2a was found to be generally stable in solution but exhibited a tendency towards polydispersity in lower salt conditions as characterised by DLS measurements. An effect on the oligomeric state of PfSir2a was observed by AUC whereby increasing protein concentration causes a shift towards a mixture of oligomeric species. This had a significant effect on downstream structural applications due to sample inhomogeneity, yielding no protein crystals of sufficient quality to be characterised by X-ray crystallography. The polydispersity of PfSir2a also further impeded the results obtained via SAXS experiments as discussed in Chapter 4.

Purified PfAlba3 was shown to be an unstable and highly aggregation-prone protein after purification. Initial attempts led to protein loss due to precipitation and or aggregation immediately after purification. After extensive buffer optimisation screens to find suitable conditions, PfAlba3 could be produced and recovered in a soluble format for a short time before aggregating. This was shown by DLS measurements and proved to be an exceptional challenge for further investigations.

No stability issues appear to be reported for existing archaeal Albas isolated either directly from the native organism or in recombinant forms by bacterial expression(Bell et al., 2002; Marsh et al., 2005; Wardleworth et al., 2001) . However, it is perhaps worth noting that the organisms from which these proteins originate are typically thermophilic in nature which are known to possess generally more stable proteins than their mesophilic counterparts (Razvi and Scholtz, 2006). Contrary to the results obtained for PfAlba3 by AUC investigation by (Chêne et al., 2012), we found that recombinant PfAlba3 exists as a monomer in solution instead of a dimer as previously described by the authors.

As a result of the challenges discovered for both PfSir2a and PfAlba3, significant difficulty was encountered attempting to replicate an interaction between the two proteins as had previously been observed (Goyal et al., 2012). The results obtained for the series of pulldown experiments described in Chapter 3 indicate that under the conditions tested, the interaction between PfSir2a and PfAlba3 suggested that the interaction was extremely weak. Attempts to characterise the strength of the interaction and stoichiometry values by both MST and ITC failed to yield any correlated results between the techniques.

One possible reason for this weak interaction could be the high salt content of the buffers used during the binding, wash and elution steps. The exact conditions of the binding buffer in which the interaction between the GST-PfSir2a and PfAlba3 was reported by Goyal *et al*

using the pull down method does not appear to be listed in the materials or supplementary materials. As with the alternative method that established the PfSir2a-PfAlba3 interaction, using His-tagged PfAlba3 as a bait protein and subsequent incubation with parasite lysate, both techniques shows the interaction of PfSir2a with PfAlba3 via western blot detection. This is suggestive of a weakly bound complex due to the low levels of protein which remains bound to the partner in each method and sensitivity range of the western blot technique. The use of lower salt buffers for these investigations resulted in the near immediate precipitation and loss of PfAlba3.

Additionally, given the nature of PfSir2a as a histone deactylases enzyme, it may display a higher affinity for a construct of PfAlba3 which possesses one or several acetylated lysine residues. Evidence for PfSir2a to interact with short peptide fragments of PfAlba3 with these acetylated residues has been shown previously (Goyal et al., 2012). Development of a PfAlba3 construct with these acetylated residues would be a potential avenue of further research. This could be achieved by chemical acetylation of the recombinant protein post-purification (Fritz, 2013).

Due to the problems posed by concentration-induced oligomerisation causing sample polydispersity for PfSir2a and the rapid self-aggregating nature of PfAlba3, attempts at producing viable protein crystals suitable for X-ray crystallography and structure determination were unsuccessful. The recombinant forms of both proteins used for this thesis may not be favourable to crystallisation: in the case of PfSir2a, multiple cleaved products were produced when attempting to replicate established conditions; while PfAlba3 had a permanent C-terminal 6 His tag on the PfAlba3 construct which may have interfered with the protein stability and crystal formation (Carson et al., 2007; Derewenda, 2004; Waugh, 2005). Future work aimed at the generation of different constructs for both PfSir2a and PfAlba3 is suggested as an avenue for further investigation.

The biochemical and biophysical characterisation undertaken in Chapter 3 highlighted the challenges of producing target protein samples for structural characterisation, and also the importance of having a fully characterised sample before commencing structural studies. The information obtained in Chapter 3 helped to interpret the solution structure data obtained from SAXS experiments which are discussed in Chapter 4.

These solution based structural experiments also encountered several common problems associated with SAXS experiments; radiation-damage-induced protein aggregation was found to be the most challenging of these factors. Both PfSir2a and PfAlba3 were found to be extremely sensitive to radiation damage when exposed to X-rays for structural characterisation. This resulted in many optimisation steps to mitigate this particular challenge and produce results that could be compared to pre-existing or theoretically calculated structural data.

The experimental solution scattering results seen in Chapter 4 yielded an apparent R_g of 2.9 nm and 2.4 nm for PfSir2a and PfAlba3 respectively. In the case of PfSir2a, after mitigating the effects of radiation damage, a major problem was taking into account the effect of concentration-dependent oligomerisation states which were present from sample concentrations as low as 0.5 mg ml^{-1}. Given the lack of pre-existing structural information pertaining to the additional oligomeric states, the final result obtained represents the closest approximation to the structural properties of the monomeric PfSir2a in solution.

For PfAlba3, again after mitigating the effects of radiation damage, the main problem was that of rapidly induced protein aggregation within samples prepared for structural studies. This presented a significant challenge whereby sample preparation must be conducted immediately prior to structural experiments. The use of the BM-29 on-line HPLC system

proved to be the most effective method of gathering data which could be interpreted for this sample.

Unfortunately, given the inherent ambiguity of the PfSir2a results and the evidence for a weak interaction between PfSir2a and PfAlba3, a solution structure for the predicted PfSir2a-PfAlba3 complex could not be established by SAXS under the optimal buffer conditions. Further work on characterising the additional oligomeric forms of PfSir2a and deriving their theoretical structural properties will be required to fully interpret this system. As such, the initial aim of this thesis to characterise the interaction between PfSir2a and PfAlba3, with the intention of investigating the mechanism by which the proposed silencing complex could be disrupted, could not be achieved within the available timescale.

As a result of the biochemical, biophysical and SAXS experiments indicating that PfSir2a and PfAlba3 were extremely challenging samples, the intended SANS contrast match out experiments using deuterated proteins and D_2O also proved a difficult line of investigation. We observed that deuterated PfSir2a becomes extremely insoluble and could not be recovered in the amounts required for SANS experiments. Deuterated PfAlba3 was more easily prepared and extracted to be pure in solution, but suffered from the same problems as the hydrogenated PfAlba3 in that it was very aggregation prone. As such, traditional SANS contrast match out experiments could not be performed using these samples.

The development and investigation of a new method for bio-SANS experiments using an *in situ* SEC system at the neutron beamline D22 was carried out in chapter 5. This was aimed to combat the challenge which is often faced in biological experiments where the most interesting samples are also the most challenging to work with. As mentioned previously, aggregation is a major hindrance to SAS experiments as it affects many of the

experimentally derived values from samples which are used to calculate the solution based structural properties.

This work was found to successfully separate and distinguish between protein species separated via a SEC column immediately before sample measurement using a SANS beamline. It represents the first known example of an experiment to do so. This technique offers the advantage of being able to separate individual proteins and protein complexes in a sample for *ab initio* modelling of the separate components and the overall complex. It also offers the advantage of avoiding radiation damage effects as observed in samples measured using X-rays and hence this technique can be used for sensitive samples. Additionally, samples can be recovered using the fraction collection system and then further analysed by SAXS in joint experiments for comparison and validation of the solution structure of the samples of interest.

This technique is still limited by the need for relatively high sample concentrations, as needed for SANS, and can be challenging when attempting to analyse small proteins in hydrogenated buffer conditions due to the high background scattering from the hydrogen atoms in the buffer and the low scatter length density of the small protein molecules. This results in a low signal to noise ratio which can be improved when using deuterated buffer conditions. The work presented in Chapter 5 represents the proof of principle for this technique which could be developed further using additional instrumentation, such as a DLS device, to provide further complementary information about the samples which are being investigated.

While many challenging factors were encountered throughout the work conducted in this thesis on the investigation of PfSir2a and PfAlba3, the behaviour of these proteins has now been characterised via biophysical methods more extensively than has been reported in the

literature previously. Further investigations to characterise the true nature of the interaction between these proteins could be undertaken by generating several different constructs for each protein, with cleavable and non-cleavable His-tags on both N and C-terminal regions of both proteins for example, to establish the effects of stability and oligomerisation tendencies. A stable, non-aggregating, construct of PfAlba3, would allow for additional investigation and characterisation of its DNA binding ability by other techniques such as nuclear magnetic resonance imaging for example.

APPENDIX A: Media Recipes And Buffers

A1 Bacterial Growth Medium

Luria broth (LB) media was used for both small and large scale cultures of bacterial growth (Bertani, 1951)

For a culture volume of 1 litre, 10 g of bacto-tryptone, 5 g of yeast extract and 10 g of NaCl are mixed together. Flasks were sterilised by autoclaving and stored at room temperature before use.

LB-agar medium was used for Petri dish cultures to plate transformed bacteria. 15 g of agar powder was added to the previous LB medium and sterilised by autoclaving. When cooled to approximately 30˚C, the kanamycin sulphate antibiotic (at 35 μg ml^{-1}) was added. 20 ml of LB agar was poured per petri dish and once set were stored at 4˚C prior to use.

Super optimal broth with catabolite repression (SOC media) (Hanahan, 1983) was used as recovery media for bacterial transformations. The SOC medium contains 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.

Sterilisation of media by autoclaving was carried out at 121˚C for 15 minutes.

A2 Deuterated Media Composition

For 1 litre of Enfors minimal media (using 100% H_2O or 85% D_2O solvent for hydrogenated and deuterated respectively)

6.86 g $(NH_4)_2SO_4$

1.56 g KH_2PO_4

6.48 g $Na_2HPO_4.2H_2O$

0.49 g $(NH_4)_2$-H-citrate

5 g glycerol

Sterilise media by autoclaving and allow to cool down. Once cool, add the following metals salts

0.23 g L^{-1} $MgSO_4$

0.50 g L^{-1} $CaCl_2.2H_2O$

16.7 g L^{-1} $FeCl_3.6H_2O$

0.18 g L^{-1} $ZnSO_4.7H_2O$

0.16 g L^{-1} $CuSO_4.5H_2O$

0.15 g L^{-1} $MnSO_4.4H_2O$

0.18 g L^{-1} $CoCl_2.6H_2O$

20.1 g L^{-1} Na-EDTA

A3 Buffer Composition

PfSir2a Purification buffers

Lysis and binding - 50mM sodium phosphate, 500mM sodium chloride (NaCl), 10mM Imidazole, 5U ml^{-1} Benzonase nuclease (Novagen® Cat no. 71205-3) 1x cOmplete EDTA-free Protease inhibitor cocktail (Roche Cat no. 11 873 580 11), pH 7.5

Wash - 50mM sodium phosphate, 500mM NaCl, 40mM Imidazole, pH 7.5

Elution - 50mM sodium phosphate, 500mM NaCl, 500mM Imidazole, pH 7.5

PfAlba3Purification buffers

Lysis and binding - 50mM sodium phosphate, 500mM NaCl, 10mM Imidazole, 10U ml^{-1} Benzonase nuclease (Novagen® Cat no. 71205-3) 1x cOmplete EDTA-free Protease inhibitor cocktail (Roche Cat no. 11 873 580 11), pH 7.5

Wash - 50mM sodium phosphate, 1M NaCl, 40mM Imidazole, pH 7.5

Elution - 50mM sodium phosphate, 500mM NaCl, 500mM Imidazole, pH 7.5

IEX Buffers

IEX Buffer A - 50mM sodium phosphate, 50mM NaCl, pH 7.5

IEX Buffer B - 50mM sodium phosphate, 2M NaCl, pH 7.5

SEC Buffer

50mM sodium phosphate, 500mM NaCl, pH 7.0

TBE Buffer for Agarose Gel Electrophoresis

For a 1 L 5X stock solution, 54 g of Trise base, 27.5g of boric acid and 20 ml of 0.5M EDTA (pH 8.0) is added. A working solution of 0.5X is used by diluting the 5X stck tenfold.

10X Running Buffers for Tris Tricine SDS-PAGE Gels

10x Cathode Buffer (Load 1X on top into wells, in the middle)

-121.1 g Tris base

-179.2 g Tricine

-10 g SDS

Dilute to 1 litre with ddH_2O

Do not adjust pH

Store at 4˚C

Final concentrations in the 1X buffer are 0.1M Tris, 0.1M Tricine, and 0.1% SDS

10x Anode Buffer (Load 1X at the bottom, on the outside of the gel)

-242.2 g Tris base

Adjust to pH 8.9 with concentrated HCl

Dilute to 1 litre with ddH_2O

Store at 4˚C

Final concentration is 0.2M Tris-Cl, pH8.9

Staining solution (1X)

-2.5 g coomassie Blue G-250

-500 mL of 100 % Ethanol (or **520 mL of 96% ethanol** which is cheaper)

-100 mL of 100% acetic acid

Dilute to 1 litre with ddH_2O

Final concentrations are 10% acetic acid, 50% Ethanol and 0.25% Coomassie

Destain solution (1X)

-300 mL of 100 % Ethanol (or **313 mL of 96% ethanol** which is cheaper)

-200 mL of 100% acetic acid

Dilute to 4 litres with ddH_2O

Final concentrations are 5% acetic acid, 7.5% Ethanol

SDS Tricine Gel Recipes

16.5 % running gel (16.5 % Acry, 1.0 M Tris)

# gels / stock	2 or 3 gels (10 mL)	8 gels (9 gels ?)	12 gels (or 13 gels ?)
Acryl/Bisacryl (30/0.8%) ie 37.5:1 ratio	5.5 mL	16.5 mL	22 mL
Tris-Cl 3M pH 8.45	3.3 mL	10 mL	13.3 mL
10% SDS	100 uL	300 uL	400 uL
Glycerol 100%	1 mL	3 mL	4 mL
ddH2O	100 uL	200 uL	300 uL
10% APS	100 uL	300 uL	400 uL
TEMED	10 uL	30 uL	40 uL

12 % running gel (12 % Acry, 1.0 M Tris)

# gels / stock	2 or 3 gels (10 mL)	8 gels (9 gels ?) (30 mL)	12 gels (or 13 gels ?) (40 mL)
Acryl/Bisacryl (30/0.8%) ie 37.5:1 ratio	4 mL	12 mL	16 mL
Tris-Cl 3M pH 8.45	3.3 mL	10 mL	13.3 mL
10% SDS	100 uL	300 uL	400 uL
Glycerol 100%	1 mL	3 mL	4 mL
ddH2O	1.6 mL	4.8 mL	6.3 mL
10% APS	100 uL	300 uL	400 uL
TEMED	10 uL	30 uL	40 uL

3X Sample Buffer for SDS-PAGE

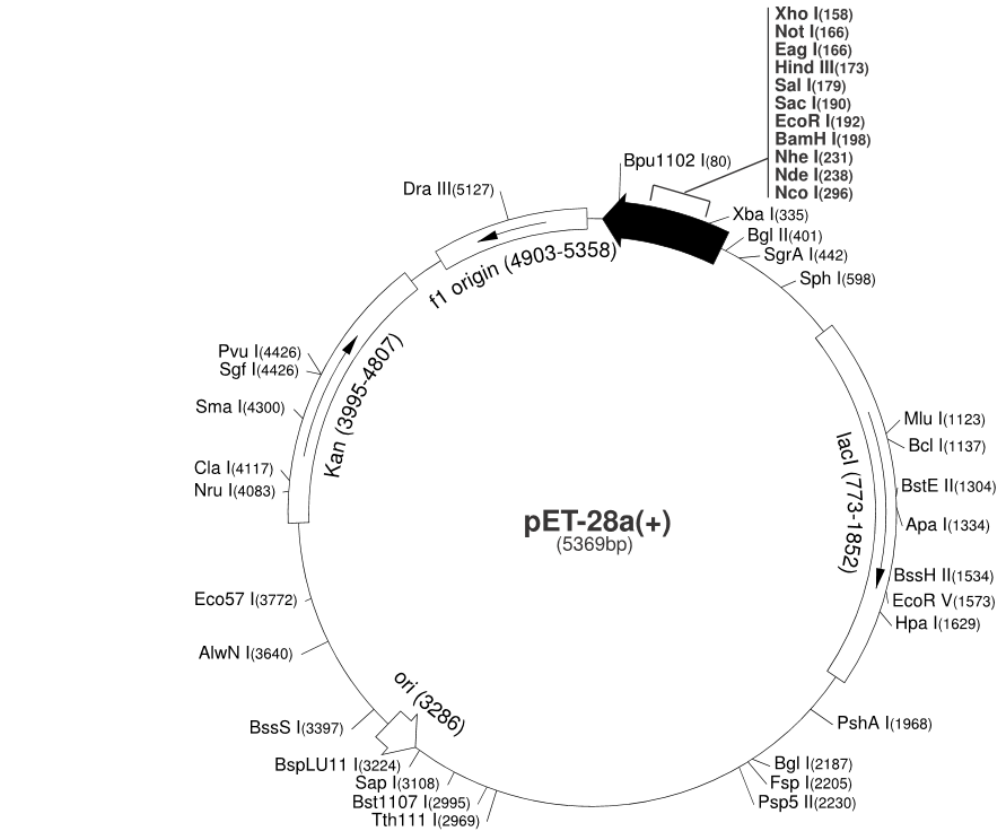
240mM tris HCl, pH 6.8, 6% SDS, 36% Glycerol (W/V), 0.9 mg ml^{-1} Bromophenol blue.

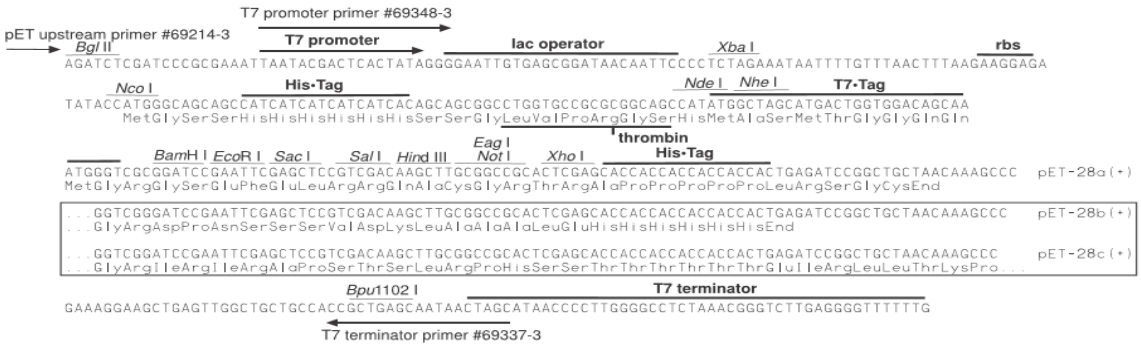
APPENDIX B: Cloning Strategy

T7 primer sequences used to obtain PfSir2a PCR product for cloning

T7 Forward primer: TAATACGACTCACTATAGGG

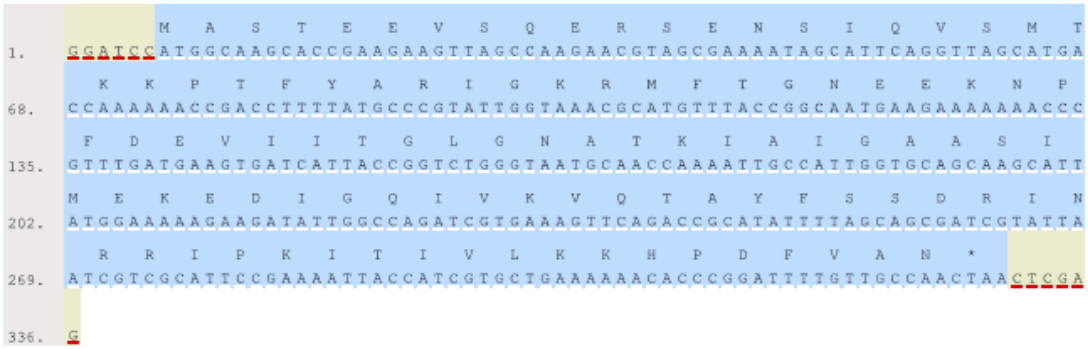
T7 Reverse primer: GCTAGTTATTGCTCAGCGG





pET28a Map And Multiple Cloning Site Schematic

pEX-A PfAlba3 Codon Optimised Sequence



Plasmids for *PfSir2a* sequence inserts were digested using XhoI and BamHI restriction enzymes. This produced a cloning strategy with a N-terminal cleavable His-tag present on PfSir2a.

Plasmids for *Alba3* sequence inserts were digested with NcoI & XhoI restriction enzymes. This produced a cloning strategy with a C-terminal non-cleavable His-tag present on PfAlba3.

PfAlba3 Site Directed Mutagenesis Primers

Forward primer : Alba3_SDM1.for

5'-GAAGGAGATATACCATGGCAAGCACCGAAGAAGTC-3'

Reverse primer : Alba3_SDM1.rev

5'-GACTTCTTCGGTGCTTGCCATGGTATATCTCCTTC-3'

APPENDIX C: Supplemental Information

Recombinant PfSir2a Sequence

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MASMTGGQQM GRGSMGNLMI SFLKKDTQSI TLEELAKIIK

        70         80         90        100        110        120
KCKHVVALTG SGTSAESNIP SFRGSSNSIW SKYDPRIYGT IWGFWKYPEK IWEVIRDISS

       130        140        150        160        170        180
DYEIEINNGH VALSTLESLG YLKSVVTQNV DGLHEASGNT KVISLHGNVF EAVCCTCNKI

       190        200        210        220        230        240
VKLNKIMLQK TSHFMHQLPP ECPCGGIFKP NIILFGEVVS SDLLKEAEEE IAKCDLLLVI

       250        260        270        280        290        300
GTSSTVSTAT NLCHFACKKK KKIVEINISK TYITNKMSDY HVCAKFSELT KVANILKGSS


EKNKKIM
```

Number of amino acids: 307

Molecular weight: 33888.2

Theoretical pI: 8.97

Amino acid composition:

Ala (A)	13	4.2%
Arg (R)	5	1.6%
Asn (N)	16	5.2%
Asp (D)	8	2.6%
Cys (C)	10	3.3%
Gln (Q)	6	2.0%
Glu (E)	19	6.2%
Gly (G)	23	7.5%
His (H)	15	4.9%
Ile (I)	27	8.8%
Leu (L)	24	7.8%
Lys (K)	31	10.1%
Met (M)	10	3.3%
Phe (F)	9	2.9%
Pro (P)	8	2.6%
Ser (S)	35	11.4%
Thr (T)	18	5.9%
Trp (W)	4	1.3%
Tyr (Y)	7	2.3%
Val (V)	19	6.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 27
Total number of positively charged residues (Arg + Lys): 36

Atomic composition:

Carbon	C	1498
Hydrogen	H	2403
Nitrogen	N	409
Oxygen	O	444
Sulfur	S	20

Formula: $C_{1498}H_{2403}N_{409}O_{444}S_{20}$
Total number of atoms: 4774

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 33055
Abs 0.1% (=1 g/l) 0.975, assuming all pairs of Cys residues form cystines

Ext. coefficient 32430
Abs 0.1% (=1 g/l) 0.957, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 38.54
This classifies the protein as stable.

Aliphatic index: 86.97

Grand average of hydropathicity (GRAVY): -0.173

Recombinant PfAlba3 Sequence

```
        10         20         30         40         50         60
MASTEEVSQE RSENSIQVSM TKKPTFYARI GKRMFTGNEE KNPFDEVIIT GLGNATKIAI

        70         80         90        100        110
GAASIMEKED IGQIVKVQTA YFSSDRINRR IPKITIVLKK HPDFVANLEH HHHHH
```

Number of amino acids: 115

Molecular weight: 13041.8

Theoretical pI: 8.96

Amino acid composition:

Ala (A)	8	7.0%
Arg (R)	6	5.2%
Asn (N)	6	5.2%
Asp (D)	4	3.5%
Cys (C)	0	0.0%
Gln (Q)	4	3.5%
Glu (E)	10	8.7%
Gly (G)	6	5.2%
His (H)	7	6.1%
Ile (I)	13	11.3%
Leu (L)	3	2.6%
Lys (K)	10	8.7%
Met (M)	4	3.5%
Phe (F)	5	4.3%
Pro (P)	4	3.5%
Ser (S)	8	7.0%
Thr (T)	8	7.0%
Trp (W)	0	0.0%
Tyr (Y)	2	1.7%
Val (V)	7	6.1%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 14
Total number of positively charged residues (Arg + Lys): 16

Atomic composition:

Carbon	C	574
Hydrogen	H	921
Nitrogen	N	167
Oxygen	O	172
Sulfur	S	4

Formula: $C_{574}H_{921}N_{167}O_{172}S_4$
Total number of atoms: 1838

Extinction coefficients:

This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	2980
Abs 0.1% (=1 g/l)	0.228

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 33.01
This classifies the protein as stable.

Aliphatic index: 78.87

Grand average of hydropathicity (GRAVY): -0.526

References:

Aravind, L., Iyer, L.M., Anantharaman, V., 2003. The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. Genome Biol. 4, R64. doi:10.1186/gb-2003-4-10-r64

Arents, G., Moudrianakis, E.N., 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc. Natl. Acad. Sci. U. S. A. 92, 11170–4.

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C.M., Bout, D.M., Ménard, S., Rogers, W.O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.-C., Fairhurst, R.M., Benoit-Vical, F., Mercereau-Puijalon, O., Ménard, D., 2014. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature 505, 50–5. doi:10.1038/nature12876

Barnwell, J.W., Asch, a S., Nachman, R.L., Yamaya, M., Aikawa, M., 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. J. Clin. Invest. 84, 765–72. doi:10.1172/JCI114234

Baruch, D.I., Pasloske, B.L., Singh, H.B., Bi, X., Ma, X.C., Feldman, M., Taraschi, T.F., Howard, R.J., 1995. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell 82, 77–87. doi:10.1016/0092-8674(95)90054-3

Beer, A., 1852. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten“ (Determination of the absorption of red light in colored liquids).” Ann. der Phys. und Chemie 86, 78–88.

Belenky, P., Bogan, K.L., Brenner, C., 2007. NAD+ metabolism in health and disease. Trends Biochem. Sci. 32, 12–19. doi:10.1016/j.tibs.2006.11.006

Bell, S.D., Botting, C.H., Wardleworth, B.N., Jackson, S.P., White, M.F., 2002a. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. Science 296, 148–151. doi:10.1126/science.1070506

Bell, S.D., Botting, C.H., Wardleworth, B.N., Jackson, S.P., White, M.F., 2002b. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. Science 296, 148–51. doi:10.1126/science.1070506

Bellamacina, C.R., 1996. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. FASEB J. 10, 1257–1269.

Berendt, a R., Simmons, D.L., Tansey, J., Newbold, C.I., Marsh, K., 1989. Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for Plasmodium falciparum. Nature 341, 57–59. doi:10.1038/341057a0

Berns, D.S., Lee, J.J., Sscott, E., 1968. Effect of Deuterium Oxide on Protein Aggregation in Deuterio and Protio Phycocyanin and Other Proteins. pp. 21–30. doi:10.1021/ba-1968-0084.ch003

Bertani, G., 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J. Bacteriol. 62, 293–300. doi:citeulike-article-id:149214

Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J., DeRisi, J.L., 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol. 1, E5. doi:10.1371/journal.pbio.0000005

Braunstein, M., Rose, a B., Holmes, S.G., Allis, C.D., Broach, J.R., 1993. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. Genes Dev. 7, 592–604. doi:10.1101/gad.7.4.592

Brennich, M.E., Kieffer, J., Bonamis, G., De Maria Antolinos, A., Hutin, S., Pernot, P., Round, A., 2016. Online data analysis at the ESRF bioSAXS beamline, BM29. J. Appl. Crystallogr. 49, 203–212. doi:10.1107/S1600576715024462

Broglie, L. De, 1924. Recherches sur la théorie des Quanta Louis De Broglie.

Buckee, C.O., Recker, M., 2012. Evolution of the multi-domain structures of virulence genes in the human malaria parasite, Plasmodium falciparum. PLoS Comput. Biol. 8, e1002451. doi:10.1371/journal.pcbi.1002451

Carson, M., Johnson, D.H., McDonald, H., Brouillette, C., DeLucas, L.J., 2007. His-tag impact on structure. Acta Crystallogr. Sect. D Biol. Crystallogr. 63, 295–301. doi:10.1107/S0907444906052024

Chakrabarty, S.P., Balaram, H., 2010. Reversible binding of zinc in Plasmodium falciparum Sir2: structure and activity of the apoenzyme. Biochim. Biophys. Acta 1804, 1743–50. doi:10.1016/j.bbapap.2010.06.010

Chêne, A., Vembar, S.S., Rivière, L., Lopez-Rubio, J.J., Claes, A., Siegel, T.N., Sakamoto, H., Scheidig-Benatar, C., Hernandez-Rivas, R., Scherf, A., 2012a. PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. Nucleic Acids Res. 40, 3066–77. doi:10.1093/nar/gkr1215

Chêne, A., Vembar, S.S., Rivière, L., Lopez-Rubio, J.J., Claes, A., Siegel, T.N., Sakamoto, H., Scheidig-Benatar, C., Hernandez-Rivas, R., Scherf, A., 2012b. PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. Nucleic Acids Res. 40, 3066–77. doi:10.1093/nar/gkr1215

Cole, J.L., Lary, J.W., Moody, T., Laue, T.M., 2009. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. doi:10.1016/S0091-679X(07)84006-4.Analytical

Cox, F.E., 2010. History of the discovery of the malaria parasites and their vectors. Parasit. Vectors 3, 5. doi:10.1186/1756-3305-3-5

Crates, S., 2003. Analytical Ultracentrifugation [WWW Document].

Croken, M.M., Nardelli, S.C., Kim, K., 2012. Chromatin modifications, epigenetics, and how protozoan parasites regulate their lives. Trends Parasitol. 28, 202–13. doi:10.1016/j.pt.2012.02.009

Cui, L., Fan, Q., Li, J., 2002. The malaria parasite Plasmodium falciparum encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. Nucleic Acids Res. 30, 4607–17.

David, G., Pérez, J., 2009. Combined sampler robot and high-performance liquid chromatography: A fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. J. Appl. Crystallogr. 42, 892–900. doi:10.1107/S0021889809029288

De Silva, E.K., Gehrke, A.R., Olszewski, K., León, I., Chahal, J.S., Bulyk, M.L., Llinás, M., 2008. Specific DNA-binding by apicomplexan AP2 transcription factors. Proc. Natl. Acad. Sci. U. S. A. 105, 8393–8. doi:10.1073/pnas.0801993105

de Vries, R., 2010. DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. Biochimie 92, 1715–1721. doi:10.1016/j.biochi.2010.06.024

Derewenda, Z.S., 2004. The use of recombinant methods and molecular engineering in protein crystallization. Methods 34, 354–363. doi:10.1016/j.ymeth.2004.03.024

Dewhurst, C., 2008. GRASP [WWW Document]. URL https://www.ill.eu/instruments-support/instruments-groups/groups/lss/grasp/home/

Dias, R., 2009. Experimental Approaches to Evaluate the Thermodynamics of Protein- Drug Interactions Experimental Approaches to Evaluate the Thermodynamics of Protein-. doi:10.2174/138945008786949441

Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, R.P.M., 2013. Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules, First edit. ed. Oxford University Press.

Doerig, C., Rayner, J.C., Scherf, A., Tobin, A.B., 2015. Post-translational protein modifications in malaria parasites. Nat. Rev. Microbiol. 13, 160–172. doi:10.1038/nrmicro3402

Duffy, M.F., Selvarajah, S. a, Josling, G. a, Petter, M., 2012. The role of chromatin in Plasmodium gene expression. Cell. Microbiol. 14, 819–28. doi:10.1111/j.1462-5822.2012.01777.x

Duhr, S., Braun, D., 2006. Why molecules move along a temperature gradient 2006.

Duraisingh, M.T., Voss, T.S., Marty, A.J., Duffy, M.F., Good, R.T., Thompson, J.K., Freitas-Junior, L.H., Scherf, A., Crabb, B.S., Cowman, A.F., 2005. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. Cell 121, 13–24. doi:10.1016/j.cell.2005.01.036

Dutnall, R.N., Pillus, L., 2001. Deciphering NAD-dependent deacetylases. Cell 105, 161–4.

Einstein, A., 1905. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Ann. d. Phys. 322, 549–560. doi:10.1002/andp.19053220806

Feigin, L.A., Svergun, D.I., 1987. Structure Analysis by Small Angle X-ray and Neutron Scattering. Plenum Press. doi:10.1002/actp.1989.010400317

Finnin, M.S., Donigian, J.R., Pavletich, N.P., 2001. letters Structure of the histone deacetylase SIRT2 8, 621–625.

Fischetti, R.F., Rodi, D.J., Mirza, A., Irving, T.C., 2003. Research Papers of Protein Solutions : Effect of Beam Dose on Protein Integrity Research Papers. Small 398–404.

Forterre, P., Confalonieri, F., Knapp, S., 1999. Identification of the gene encoding archeal-specific DNA-binding proteins of the Sac10b family. Mol. Microbiol. 32, 669–670. doi:10.1046/j.1365-2958.1999.01366.x

Franke, D., Svergun, D.I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. 42, 342–346. doi:10.1107/S0021889809000338

Freitas-Junior, L.H., Hernandez-Rivas, R., Ralph, S. a, Montiel-Condado, D., Ruvalcaba-Salazar, O.K., Rojas-Meza, A.P., Mâncio-Silva, L., Leal-Silvestre, R.J., Gontijo, A.M., Shorte, S., Scherf, A., 2005. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell 121, 25–36. doi:10.1016/j.cell.2005.01.037

Fritz, K.S., 2013. Chemical acetylation and deacetylation. Methods Mol. Biol. 1077, 191–201. doi:10.1007/978-1-62703-637-5_13

Frye, R. a, 2000. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. Biochem. Biophys. Res. Commun. 273, 793–8. doi:10.1006/bbrc.2000.3000

Gallwitz, M., Enoksson, M., Thorpe, M., Hellman, L., 2012. The extended cleavage specificity of human thrombin. PLoS One 7. doi:10.1371/journal.pone.0031756

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J. a, Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M. a, Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S. a, McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M., Barrell, B., 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419, 498–511. doi:10.1038/nature01097

Gasser, S.M., Cockell, M.M., 2001. The molecular biology of the SIR proteins. Gene 279, 1–16.

Gibbs, J., 1873. A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces. Trans. Connect. Acad. Arts Sci. 382–404.

Goldenberg, D.P., Argyle, B., 2014. Minimal effects of macromolecular crowding on an intrinsically disordered protein: A small-angle neutron scattering study. Biophys. J. 106, 905–914. doi:10.1016/j.bpj.2013.12.003

Goyal, M., Alam, A., Iqbal, M.S., Dey, S., Bindu, S., Pal, C., Banerjee, A., Chakrabarti, S., Bandyopadhyay, U., 2012. Identification and molecular characterization of an Alba-family protein from human malaria parasite Plasmodium falciparum. Nucleic Acids Res. 40, 1174–90. doi:10.1093/nar/gkr821

Gregoretti, I. V, Lee, Y.-M., Goodson, H. V, 2004. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. J. Mol. Biol. 338, 17–31. doi:10.1016/j.jmb.2004.02.006

Greiss, S., Gartner, A., 2009. Sirtuin/Sir2 phylogeny, evolutionary considerations and structural conservation. Mol. Cells 28, 407–15. doi:10.1007/s10059-009-0169-x

Grillo, I., 2009. Applications of stopped-flow in SAXS and SANS. Curr. Opin. Colloid Interface Sci. 14, 402–408. doi:10.1016/j.cocis.2009.04.005

Guizetti, J., Scherf, A., 2013. Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum. Cell. Microbiol. 15, 718–26. doi:10.1111/cmi.12115

Guo, R., Xue, H., Huang, L., 2003. Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. Mol. Microbiol. 50, 1605–1615. doi:10.1046/j.1365-2958.2003.03793.x

Haertlein, F.M., Moulin, M., Devos, J.M., Laux, V., Dunne, O., Trevor, V., Haertlein, M., Moulin, M., Devos, J.M., Laux, V., Dunne, O., Forsyth, V.T., 2016. Biomolecular Deuteration for Neutron Structural Biology and Dynamics.

Haigis, M.C., Mostoslavsky, R., Haigis, K.M., Fahie, K., Christodoulou, D.C., Murphy, A.J., Valenzuela, D.M., Yancopoulos, G.D., Karow, M., Blander, G., Wolberger, C., Prolla, T. a, Weindruch, R., Alt, F.W., Guarente, L., 2006. SIRT4 inhibits glutamate dehydrogenase and opposes the effects of calorie restriction in pancreatic beta cells. Cell 126, 941–54. doi:10.1016/j.cell.2006.06.057

Haigis, M.C., Sinclair, D.A., 2010. Mammalian Sirtuins: Biological Insights and Disease Relevance 253–295. doi:10.1146/annurev.pathol.4.110807.092250.Mammalian

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166, 557–80.

Hoff, K.G., Avalos, J.L., Sens, K., Wolberger, C., 2006. Insights into the sirtuin mechanism from ternary complexes containing NAD+ and acetylated peptide. Structure 14, 1231–40. doi:10.1016/j.str.2006.06.006

Hong, L., Lei, J., 2007. Scaling Law for the Radius of Gyration of Proteins and It's Dependence on Hydrophobicity. J. Polym. Sci. Part B Polym. Phys. 45, 1390–1398. doi:10.1002/polb

Ianeselli, L., Zhang, F., Skoda, M.W.A., Jacobs, R.M.J., Martin, R.A., Callow, S., Prévost, S., Schreiber, F., 2010. Protein−Protein Interactions in Ovalbumin Solutions Studied

by Small-Angle Scattering: Effect of Ionic Strength and the Chemical Nature of Cations. J. Phys. Chem. B 114, 3776–3783. doi:10.1021/jp9112156

Imai, S., Armstrong, C.M., Kaeberlein, M., Guarente, L., 2000. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature 403, 795–800. doi:10.1038/35001622

Jacques, D.A., Trewhella, J., 2010. Small-angle scattering for structural biology — Expanding the frontier while avoiding the pitfalls 642–657. doi:10.1002/pro.351

Jeffries, C.M., Graewert, M.A., Svergun, D.I., Blanchet, C.E., 2015. Limiting radiation damage for high-brilliance biological solution scattering: Practical experience at the EMBL P12 beamline PETRAIII. J. Synchrotron Radiat. 22, 273–279. doi:10.1107/S1600577515000375

Jelinska, C., Conroy, M.J., Craven, C.J., Hounslow, A.M., Bullough, P. a., Waltho, J.P., Taylor, G.L., White, M.F., 2005. Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA Packaging. Structure 13, 963–971. doi:10.1016/j.str.2005.04.016

Jensen, A.T.R., Magistrado, P., Sharp, S., Joergensen, L., Lavstsen, T., Chiucchiuini, A., Salanti, A., Vestergaard, L.S., Lusingu, J.P., Hermsen, R., Sauerwein, R., Christensen, J., Nielsen, M. a, Hviid, L., Sutherland, C., Staalsoe, T., Theander, T.G., 2004. Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. J. Exp. Med. 199, 1179–90. doi:10.1084/jem.20040274

Jerabek-willemsen, M., Wienken, C.J., Braun, D., Baaske, P., Duhr, S., 2011. Molecular Interaction Studies Using Microscale Thermophoresis 342–353. doi:10.1089/adt.2011.0380

Jordan, A., Jacques, M., Merrick, C., Devos, J., Forsyth, V.T., Porcar, L., Martel, A., 2016. SEC-SANS: size exclusion chromatography combined in situ with small-angle neutron scattering 1. J. Appl. Cryst 49, 2015–2020. doi:10.1107/S1600576716016514

Kline, S.R., 2006. Reduction and analysis of SANS and USANS data using IGOR Pro. J. Appl. Crystallogr. 39, 895–900. doi:10.1107/S0021889806035059

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M.H.J., Svergun, D.I., 2003. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 36, 1277–1282. doi:10.1107/S0021889803012779

Kornberg, R.D., Lorch, Y., 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98, 285–94.

Kraemer, S.M., Smith, J.D., 2006. A family affair: var genes, PfEMP1 binding, and malaria disease. Curr. Opin. Microbiol. 9, 374–80. doi:10.1016/j.mib.2006.06.006

Kuwamoto, S., Akiyama, S., Fujisawa, T., 2004. Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: Dose-related considerations and suppression by cryoprotectants. J. Synchrotron Radiat. 11, 462–468. doi:10.1107/S0909049504019272

Kyes, S.A., Kraemer, S.M., Smith, J.D., 2007. Antigenic Variation in Plasmodium falciparum : Gene Organization and Regulation of the var Multigene Family MINIREVIEWS Antigenic Variation in Plasmodium falciparum : Gene Organization and Regulation of the var Multigene Family □ 6. doi:10.1128/EC.00173-07

Kyes, S. a, Kraemer, S.M., Smith, J.D., 2007. Antigenic variation in Plasmodium falciparum: gene organization and regulation of the var multigene family. Eukaryot. Cell 6, 1511–20. doi:10.1128/EC.00173-07

Kyriacou, H.M., Stone, G.N., Challis, R.J., Raza, A., Lyke, K.E., Thera, M. a, Koné, A.K., Doumbo, O.K., Plowe, C. V, Rowe, J.A., 2006. Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. Mol. Biochem. Parasitol. 150, 211–8. doi:10.1016/j.molbiopara.2006.08.005

Landry, J., Slama, J.T., Sternglanz, R., 2000. Role of NAD(+) in the deacetylase activity of the SIR2-like proteins. Biochem. Biophys. Res. Commun. 278, 685–90. doi:10.1006/bbrc.2000.3854

Lee, J.J., Berns, D.S., 1968. Protein aggregation. The effect of deuterium oxide on large protein aggregates of C-phycocyanin. Biochem. J. 110, 465–70.

Ling, X., Harkness, T. a, Schultz, M.C., Fisher-Adams, G., Grunstein, M., 1996. Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation. Genes Dev. 10, 686–699. doi:10.1101/gad.10.6.686

Liszt, G., Ford, E., Kurtev, M., Guarente, L., 2005. Mouse Sir2 homolog SIRT6 is a nuclear ADP-ribosyltransferase. J. Biol. Chem. 280, 21313–20. doi:10.1074/jbc.M413296200

Lopez-Rubio, J.J., Gontijo, A.M., Nunes, M.C., Issar, N., Hernandez Rivas, R., Scherf, A., 2007. 5' Flanking Region of Var Genes Nucleate Histone Modification Patterns Linked To Phenotypic Inheritance of Virulence Traits in Malaria Parasites. Mol. Microbiol. 66, 1296–305. doi:10.1111/j.1365-2958.2007.06009.x

Luah, Y.-H., Chaal, B.K., Ong, E.Z., Bozdech, Z., 2010. A moonlighting function of Plasmodium falciparum histone 3, mono-methylated at lysine 9? PLoS One 5, e10252. doi:10.1371/journal.pone.0010252

Mancio-Silva, L., Rojas-Meza, A.P., Vargas, M., Scherf, A., Hernandez-Rivas, R., 2008. Differential association of Orc1 and Sir2 proteins to telomeric domains in Plasmodium falciparum. J. Cell Sci. 121, 2046–53. doi:10.1242/jcs.026427

Mariño-Ramírez, L., Kann, M., 2005a. Histone structure and nucleosome stability 2, 719–729.

Mariño-Ramírez, L., Kann, M., 2005b. Histone structure and nucleosome stability 2, 719–729.

Marsh, V.L., Peak-Chew, S.Y., Bell, S.D., 2005. Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba. J. Biol. Chem. 280, 21122–21128. doi:10.1074/jbc.M501280200

Merrick, C.J., Duraisingh, M.T., 2007. Plasmodium falciparum Sir2: an unusual sirtuin with dual histone deacetylase and ADP-ribosyltransferase activity. Eukaryot. Cell 6, 2081–91. doi:10.1128/EC.00114-07

Merrick, C.J., Duraisingh, M.T., 2006. Heterochromatin-mediated control of virulence gene expression. Mol. Microbiol. 62, 612–20. doi:10.1111/j.1365-2958.2006.05397.x

Merrick, C.J., Huttenhower, C., Buckee, C., Amambua-Ngwa, A., Gomez-Escobar, N., Walther, M., Conway, D.J., Duraisingh, M.T., 2012. Epigenetic dysregulation of virulence gene expression in severe Plasmodium falciparum malaria. J. Infect. Dis. 205, 1593–600. doi:10.1093/infdis/jis239

Miao, J., Fan, Q., Cui, L., Li, J., Li, J., Cui, L., 2006. The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation. Gene 369, 53–65. doi:10.1016/j.gene.2005.10.022

Milev, S., 2013. Isothermal titration calorimetry: Principles and experimental design [WWW Document].

Miller, L.H., Ackerman, H.C., Su, X., Wellems, T.E., 2013. Malaria biology and disease pathogenesis: insights for new treatments. Nat. Med. 19, 156–67. doi:10.1038/nm.3073

Min, J., Landry, J., Sternglanz, R., Xu, R.M., 2001. Crystal structure of a SIR2 homolog-NAD complex. Cell 105, 269–79.

Moazed, D., 2001. Enzymatic activities of Sir2 and chromatin silencing Danesh Moazed 232–238.

Moazed, D., Kistler, A., Axelrod, A., Rine, J., Johnson, A.D., 1997. Silent information regulator protein complexes in Saccharomyces cerevisiae: a SIR2/SIR4 complex and evidence for a regulatory domain in SIR4 that inhibits its interaction with SIR3. Proc. Natl. Acad. Sci. U. S. A. 94, 2186–91. doi:10.1073/pnas.94.6.2186

Moniot, S., Weyand, M., Steegborn, C., 2012. Structures, substrates, and regulators of mammalian Sirtuins - opportunities and challenges for drug development. Front. Pharmacol. 3 FEB, 1–5. doi:10.3389/fphar.2012.00016

Nielsen, M. a., Staalsoe, T., Kurtzhals, J. a. L., Goka, B.Q., Dodoo, D., Alifrangis, M., Theander, T.G., Akanmori, B.D., Hviid, L., 2002. Plasmodium falciparum Variant Surface Antigen Expression Varies Between Isolates Causing Severe and Nonsevere Malaria and Is Modified by Acquired Immunity. J. Immunol. 168, 3444–3450. doi:10.4049/jimmunol.168.7.3444

Penman, B., Gupta, S., 2008. Evolution of virulence in malaria. J. Biol. 7, 22. doi:10.1186/jbiol83

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L., McSweeney, S., 2013. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J. Synchrotron Radiat. 20, 660–664. doi:10.1107/S0909049513010431

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P. V., Svergun, D.I., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Crystallogr. 45, 342–350. doi:10.1107/S0021889812007662

Pruitt, K., Zinn, R.L., Ohm, J.E., McGarvey, K.M., Kang, S.-H.L., Watkins, D.N., Herman, J.G., Baylin, S.B., 2006. Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation. PLoS Genet. 2, e40. doi:10.1371/journal.pgen.0020040

Rajendran, R., Garva, R., Krstic-Demonacos, M., Demonacos, C., 2011. Sirtuins: molecular traffic lights in the crossroad of oxidative stress, chromatin remodeling, and transcription. J. Biomed. Biotechnol. 2011, 368276. doi:10.1155/2011/368276

Ramakrishnan, V., 1997. Histone structure and the organization of the nucleosome. Annu. Rev. Biophys. Biomol. Struct. 26, 83–112. doi:10.1146/annurev.biophys.26.1.83

Razvi, A., Scholtz, J.M., 2006. Lessons in stability from thermophilic proteins. Protein Sci. 15, 1569–1578. doi:10.1110/ps.062130306

Religa, A. a, Waters, A.P., 2012. Sirtuins of parasitic protozoa: in search of function(s). Mol. Biochem. Parasitol. 185, 71–88. doi:10.1016/j.molbiopara.2012.08.003

Rice, P.A., Yang, S.W., Mizuuchi, K., Nash, H.A., 1996. Crystal structure of an IHF-DNA complex: A protein-induced DNA U-turn. Cell 87, 1295–1306. doi:10.1016/S0092-8674(00)81824-3

Rine, J., Herskowitz, I., 1987. Four Genes Resposnible for a Positions Effect on Expression From HML and HMR in Saccharomyces cerevisiae 22, 9–22.

Rottmann, M., Lavstsen, T., Mugasa, J.P., Kaestli, M., Jensen, A.T.R., M??ller, D., Theander, T., Beck, H.P., 2006. Differential expression of var gene groups is associated with morbidity caused by Plasmodium falciparum infection in Tanzanian children. Infect. Immun. 74, 3904–3911. doi:10.1128/IAI.02073-05

Round, A., Brown, E., Marcellin, R., Kapp, U., Westfall, C.S., Jez, J.M., Zubieta, C., 2013a. Determination of the GH3.12 protein conformation through HPLC-integrated SAXS measurements combined with X-ray crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. 69, 2072–2080. doi:10.1107/S0907444913019276

Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., Blanchet, C.E., Pernot, P., McSweeney, S., Roessle, M., Svergun, D.I., Cipriani, F., 2015. BioSAXS Sample Changer: A robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. Acta Crystallogr. Sect. D Biol. Crystallogr. 71, 67–75. doi:10.1107/S1399004714026959

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. doi:10.1038/nprot.2010.5

Ruijter, A.J.M.D.E., Gennip, A.H.V.A.N., Caron, H.N., Kemp, S., Kuilenburg, A.B.P.V.A.N., 2003. Histone deacetylases (HDACs) : characterization of the classical

HDAC family 749, 737–749.

Salcedo-Amaya, A.M., van Driel, M. a, Alako, B.T., Trelle, M.B., van den Elzen, A.M.G., Cohen, A.M., Janssen-Megens, E.M., van de Vegte-Bolmer, M., Selzer, R.R., Iniguez, a L., Green, R.D., Sauerwein, R.W., Jensen, O.N., Stunnenberg, H.G., 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum. Proc. Natl. Acad. Sci. U. S. A. 106, 9655–60. doi:10.1073/pnas.0902515106

Sauve, A.A., Celic, I., Avalos, J., Deng, H., Boeke, J.D., Schramm, V.L., 2001. Chemistry of Gene Silencing: The Mechanism of NAD $^{+}$ -Dependent Deacetylation Reactions. Biochemistry 40, 15456–15463. doi:10.1021/bi011858j

Scherf, A., Lopez-Rubio, J.J., Riviere, L., 2008. Antigenic variation in Plasmodium falciparum. Annu. Rev. Microbiol. 62, 445–70. doi:10.1146/annurev.micro.61.080706.093134

Scherf, A, Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., Lanzer, M., 1998. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum. EMBO J. 17, 5418–26. doi:10.1093/emboj/17.18.5418

Schumacher, R.-F., Spinelli, E., 2012. Malaria in children. Mediterr. J. Hematol. Infect. Dis. 4, e2012073. doi:10.4084/MJHID.2012.073

Smith, J., Chitnis, C., Craig, A., 1995. Switches in expression of plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 82, 101–110.

Srivastava, A., Gangnard, S., Round, A., Dechavanne, S., Juillerat, A., Raynal, B., 2010. Full-length extracellular region of the var2CSA variant of PfEMP1 is required for specific , high-affinity binding to CSA. doi:10.1073/pnas.1000951107

Starai, V.J., Takahashi, H., Boeke, J.D., Escalante-Semerena, J.C., 2003. Short-chain fatty acid activation by acyl-coenzyme A synthetases requires SIR2 protein function in Salmonella enterica and Saccharomyces cerevisiae. Genetics 163, 545–555.

Strahl, B., Allis, C., 2000. The language of covalent histone modifications. Nature 403, 41–45.

Struhl, K., 1998. Histone acetylation and transcriptional regulatory mechanisms Histone acetylation and transcriptional regulatory mechanisms 599–606.

Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J. a, Peterson, D.S., Ravetch, J. a, 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. doi:10.1016/0092-8674(95)90055-1

Svergun, D., 1992. Determination of the Regnlarization Parameter in Indirect-Transform Methods Using Perceptual Criteria 495–503. doi:10.1107/S0021889892001663

Svergun, D., Barberato, C., Koch, M.H., 1995. CRYSOL - A program to evaluate X-ray

solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. 28, 768–773. doi:10.1107/S0021889895007047

Svergun, D.I., Koch, M.H.J., 2003. Small-angle scattering studies of biological macromolecules in solution 66, 1735–1782.

Tanner, K.G., Landry, J., Sternglanz, R., Denu, J.M., 2000. Silent information regulator 2 family of NAD- dependent histone/protein deacetylases generates a unique product, 1-O-acetyl-ADP-ribose. Proc. Natl. Acad. Sci. U. S. A. 97, 14178–82. doi:10.1073/pnas.250422697

Tanny, J.C., Dowd, G.J., Huang, J., Hilz, H., Moazed, D., 1999. An enzymatic activity in the yeast Sir2 protein that is essential for gene silencing. Cell 99, 735–45.

Tanny, J.C., Moazed, D., 2001. Coupling of histone deacetylation to NAD breakdown by the yeast silencing protein Sir2: Evidence for acetyl transfer from substrate to an NAD breakdown product. Proc. Natl. Acad. Sci. U. S. A. 98, 415–20. doi:10.1073/pnas.031563798

Tonkin, C.J., Carret, C.K., Duraisingh, M.T., Voss, T.S., Ralph, S. a, Hommel, M., Duffy, M.F., Silva, L.M. Da, Scherf, A., Ivens, A., Speed, T.P., Beeson, J.G., Cowman, A.F., 2009. Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum. PLoS Biol. 7, e84. doi:10.1371/journal.pbio.1000084

Trampuz, A., Jereb, M., Muzlovic, I., Prabhu, R.M., 2003. Clinical review: Severe malaria. Crit. Care 7, 315–23. doi:10.1186/cc2183

Tremethick, D.J., 2007a. Higher-order structures of chromatin: the elusive 30 nm fiber. Cell 128, 651–4. doi:10.1016/j.cell.2007.02.008

Tremethick, D.J., 2007b. Higher-order structures of chromatin: the elusive 30 nm fiber. Cell 128, 651–4. doi:10.1016/j.cell.2007.02.008

Turner, L., Lavstsen, T., Berger, S.S., Wang, C.W., Jens, E. V, Avril, M., Brazier, A.J., Freeth, J., Jespersen, J.S., Morten, A., Magistrado, P., Lusingu, J., Smith, J.D., Higgins, M.K., Theander, T.G., 2013. Severe malaria is associated with parasite binding to endothelial protein C receptor 498, 502–505. doi:10.1038/nature12216.Severe

Villar-Garea, A., Imhof, A., 2006. The analysis of histone modifications. Biochim. Biophys. Acta 1764, 1932–9. doi:10.1016/j.bbapap.2006.08.009

Volkov, V. V., Svergun, D.I., 2003. Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Crystallogr. 36, 860–864. doi:10.1107/S0021889803000268

Walsh, C.T., Garneau-Tsodikova, S., Gatto, G.J., 2005. Protein posttranslational modifications: the chemistry of proteome diversifications. Angew. Chem. Int. Ed. Engl. 44, 7342–72. doi:10.1002/anie.200501023

Wardleworth, B.N., Russell, R.J.M., Bell, S.D., Taylor, G.L., White, M.F., 2002. Structure of Alba: An archaeal chromatin protein modulated by acetylation. EMBO J. 21, 4654–4662.

Wardleworth, B.N., Russell, R.J.M., White, M.F., Taylor, G.L., 2001. Preliminary crystallographic studies of the double-stranded DNA-binding protein Sso10b from Sulfolobus solfataricus. Acta Crystallogr. Sect. D Biol. Crystallogr. 57, 1893–1894. doi:10.1107/S0907444901015517

Waugh, D.S., 2005. Making the most of affinity tags. Trends Biotechnol. doi:10.1016/j.tibtech.2005.03.012

Wellems, T.E., Plowe, C. V, 2001. Chloroquine-resistant malaria. J. Infect. Dis. 184, 770–6. doi:10.1086/322858

Wenzel, U., 2006. Nutrition, sirtuins and aging. Genes Nutr. 1, 85–93. doi:10.1007/BF02829950

White, M.F., Bell, S.D., 2002. Holding it together: chromatin in the Archaea. Trends Genet. 18, 621–6.

WHO, 2014. WHO Malaria Report 2014.

WHO, 2010. Guidelines for the treatment of malaria. 2nd Edition. Geneva.

Wolffe, a P., Hayes, J.J., 1999. Chromatin disruption and modification. Nucleic Acids Res. 27, 711–20.

Wyatt, P.J., 1993. Light scattering and the absolute characterization of macromolecules. Anal. Chim. Acta 272, 1–40. doi:10.1016/0003-2670(93)80373-S

Xiao, Q., Zhang, F., Nacev, B.A., Liu, J.O., Pei, D., 2010. Protein N-Terminal Processing : Substrate Specificity of Escherichia coli and 5588–5599. doi:10.1021/bi1005464

Xue, H., Guo, R., Wen, Y., Liu, D., Huang, L., 2000. An abundant DNA binding protein from the hyperthermophilic archaeon Sulfolobus shibatae affects DNA supercoiling in a temperature-dependent fashion. J. Bacteriol. 182, 3929–3933. doi:10.1128/JB.182.14.3929-3933.2000

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat Meth 12, 7–8. doi:10.1038/nmeth.3213\rhttp://www.nature.com/nmeth/journal/v12/n1/abs/nmeth.3213.html#supplementary-information

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 40. doi:10.1186/1471-2105-9-40

Zhao, K., Chai, X., Clements, A., Marmorstein, R., 2003a. Structure and autoregulation of the yeast Hst2 homolog of Sir2. Nat. Struct. Biol. 10, 864–71. doi:10.1038/nsb978

Zhao, K., Chai, X., Marmorstein, R., 2003b. Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding. J. Biol. Chem. 278, 26071–26077. doi:10.1074/jbc.M303666200

Zhu, A.Y., Zhou, Y., Khan, S., Deitsch, K.W., Hao, Q., Lin, H., 2012. Plasmodium falciparum Sir2A preferentially hydrolyzes medium and long chain fatty acyl lysine.

ACS Chem. Biol. 7, 155–9. doi:10.1021/cb200230x

Ziegler, M., 2000. New functions of a long-known molecule: Emerging roles of NAD in cellular signaling. Eur. J. Biochem. 267, 1550–1564. doi:10.1046/j.1432-1327.2000.01187.x

Zimm, B.., 1945. Molecular Theory of the Scattering of Light in Fluids. J. Chem. Phys. 13, 141–145. doi:10.1063/1.1724013