Functional interactions of the C-terminus of the complement regulator Factor H with its ligands

Thesis Presented for the Degree of Doctor of Philosophy

By
Orla Dunne

Department of Structural and Molecular Biology
University College London

October 2015

I, Orla Dunne, confirm that work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis

Abstract

The complement system plays a crucial role in immunity by eliminating disease-causing pathogens. Its major regulator, complement Factor H (CFH) with 20 short complement regulator (SCR) domains, prevents the destruction of host cells through excessive complement activation. CFH achieves this in the alternative pathway of complement activation through its discrimination between activator (pathogen) and non-activator (hosts) cells. Mutations and polymorphisms within CFH can lead to severe disease. Three interactions that enables CFH to protect host cells were investigated in this thesis.

CFH contains a weak dimerisation site that potentially enables high concentrations of CFH to accumulate at host cell surfaces. The C-terminal SCR-16/20 domains contains an unknown dimerization site. Using seven C-terminal CFH fragments, a combination of size exclusion chromatography (SEC), analytical ultracentrifugation (AUC) and small angle X-ray scattering (SAXS) showed that this dimer site is located in SCR-17/18. Modelling suggested that the dimer was formed by an anti-parallel length-wise interaction. Because disease-causing mutations in SCR-17/18 occurred at the predicted dimer interface, this explains how these mutations may lead to disease.

CFH interacts with pentameric C-reactive protein (CRP). The interaction between the seven CFH C-terminal fragments and CRP was examined to identify the CRP binding site on CFH. A combinantion of SEC, AUC, fluorescent AUC, surface plasmon resonance, microscale thermophoresis and electron microscopy revealed a weak binding site in SCR-19/20 for pentameric CRP. This interaction would allow CFH to bind to CRP at high concentrations on host cell surfaces during an acute phase response with elevated CRP plasma levels, thus protecting CRP-decorated host cells from excessive immune activation.

The CFH SCR-19/20 complex with the C3d domain of C3b has been studied by crystallography to reveal either a 1:1 or 1:2 binding stoichiometry. To resolve this ambiguity, deuterium-labelled SCR-19/20 was produced to carry out contrast variation experiments by small-angle neutron scattering and SAXS. The results showed that one SCR-19/20 interacts with two C3d molecules in solution providing an extra mechanism for CFH to protect host cells during excessive inflammation.

Acknowledgements

Firstly, I would like to thank both of my PhD supervisors, Prof. Stephen J. Perkins and Prof. Trevor Forsyth for all their help, advice and support throughout my PhD.

I would like to thank our collaborators in Australia, Dr. Penny Adamson and Prof. David Gordon, who kindly provided us with protein constructs. I would also like to thank Dr. Wai Li Ling at the Institut de Biologie Structurale in Grenoble for all her time and effort with the electron microscopy experiments. I am grateful to my thesis chair Dr. Andrew Martin and my graduate tutor Prof. Snezana Djordjevic for all their time, help and advice throughout my time at UCL. I am also grateful to Dr. Anne Martel and Dr. Adam Round for their advice and help during beamtime and beyond.

All the members of the Life Science's group at the ILL helped and supported me immensely throughout my PhD which I really appreciate, in particular Dr. Michael Haertlein, Martine Moulin, Valerie Laux, Dr. Estelle Mossou and Dr. Samuel Lenton. I am particularly grateful to Trevor, Michael and Martine who dedicated a lot of time to help me through my scientific struggles. I am also grateful to the members of the SJP lab at UCL both past and present, in particular Jayesh Gor for all his advice and work with the AUC, Dr. Ruodan Nan, Dr. David Wright, Dr. Lindsay McDermot, Gary Fung, and Nilufar Kadkhodayi-Kholghi, for both their scientific and personal support.

Lastly I would like to thank my Family and Granny (for the endless candles) who have supported me throughout all of my studies and have pushed me to achieve my goals. For that I am eternally grateful.

Table of Contents

List of Figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Chapter Seven

List of Tables

Abbreviations

Å	Angstrom
aHUS	Atypical haemolytical uraemic syndrome
AMD	Age-related Macular Degeneration
AP	Alternative Pathway
AUC	Analytical Ultracentrifugation
CRP	C-reactive Protein
CR1	Complement Receptor type 1
CR2	Complement Receptor type 2
CR3	Complement Receptor type 3
CR4	Complement Receptor type 4
CUB	Complement C1r/C1s, UEGF, BMP1 domain
C1INH	C1 Inhibitor
C4bp	C4b-Binding Protein
Da	Dalton
DAF	Decay-Accelerating Factor
DNA	Deoxyribonucleic Acid
E.coli	Escherichia coli
EM	Electron Microscopy
EDTA	Ethylenediaminetetraacetic Acid
ESRF	European Synchrotron Radiation Facility
f	Frictional coefficient
f_0	Frictional coefficient for a compact, spherical particle
f/f_0	Frictional ratio
FHL-1	Factor H Like Protein 1
FHR	FH-Related Protein

GAGs	Glycosaminoglycans
GeV	Giga electron volts
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazin-ethanesulphonic acid
HSA	Human Serum Albumin
ILL	Institute Laue Langevin
Ig	Immunoglobulin
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgM	Immunoglobulin M
k_i	Wave vector for incident radiation beam
k_s	Wave vector for scattered radiation beam
K_D	Dissociation constant
kDa	Kilo Dalton
λ	Wavelength (nm)
L	Length
MAC	Membrane attack complex
MASP	MBL-Associated Serine Protease
MBL	Mannose-Binding Lectin
MCP	Membrane Cofactor Protein
MST	Microscale Thermophoresis
M	Molar
mM	Milli Molar
μM	Micro Molar

M_r	Molecular Mass
MS	Mass Spectrometry
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
mRNA	Messenger Ribonucleic Acid
nm	Nanometre
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffered Saline
PDB	Protein Data Bank
PEG	Polyethylene Glycol
P(r)	Paired-distance distribution function
Q	Scattering vector
RCA	Regulators of Complement Activation
R_G	Radius of Gyration
RNA	Ribonucleic Acid
RPE	Retinal Pigment Epithelium
rmsd	Root mean square deviation
RU	Resonance Units
R_{XS}	Cross-sectional radius of gyration
S	Svedburg unit (10^{-13} seconds)
$s_{20,w}$	Sedimentation coefficient corrected for water at 20 °C
SAS	Small-Angle Scattering
SANS	Small-Angle Neutron Scattering
SAXS	Small-Angle X-ray Scattering
SCR	Short Complement Regulator

SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SLD	Scattering Length Density
SPR	Surface Plasmon Resonance
SV	Sedimentation velocity
TED	Thioester containing domain
UV	Ultraviolet
2D	Two Dimensional

Units

Å	Angstrom
M	Molar
mM	Millimolar
μM	Micromolar
M_r	Molecular mass
nm	Nanometre

Amino acid and monosaccharide abbreviations

Amino acid	Three Letter code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Monosaccharide	Three Letter code
Mannose	Man
Galactose	Gal
N-Acetylglucosamine	Nag
Sialic acid	Neu

Chapter One

The Complement System

1.1 The Immune System

The immune system is the body's defence against disease. It recognises pathogens within the body and removes them before they can cause infection. The immune system consists of the adaptive immune system and the innate immune system. Adaptive immunity is the portion of the immune system which is highly specific to the inducing pathogen and which confers long term protection through immune memory. For example, if a host has been infected with a particular virus once, then often upon future invasion this virus will be rapidly recognised and removed before re-infection can occur. The adaptive response is highly specific to the pathogen which triggered it. The adaptive immune response is carried out by two types of white blood cells or lymphocytes; T-cells and B-cells. B-cells produce antibodies called immunoglobulins which are specific to the pathogens that induced their production. The secreted antibodies circulate in the blood stream and bind to their specific antigens. Antigens are specific components or fragments of the pathogen which has been presented to the T-cells which leads to the secretion of various cytokines which then activate B-cells to produce specific antibodies. Antibody binding to antigens on the pathogen prevents it from interacting with host cells and ultimately targets them for destruction by cells of the innate immune system. T-cell lymphocytes act directly on antigens which are presented on the surface of host cells, in the cell-mediated adaptive response. These cells have T-cell receptors on their surfaces which recognise specific antigens. There are many different types of T-cells which function in different ways to remove the pathogen and to prevent further infection. For example, cytotoxic or killer T-cells will destroy host cells that have been infected with a virus through recognition of viral antigens on its surface.

Innate immunity is the non-specific component of the immune system which provides an immediate response and ultimately functions to remove pathogens before they cause harm. It is the first line of defence in protecting the host against disease. Physical barriers formed by epithelial cells and molecules such as the antimicrobial defensins, are the first components of the innate immune system. If pathogens pass this barrier into the blood stream, then the innate immune system can produce an

inflammatory response to localise infection and can trigger phagocytosis to remove the invading pathogens. Unlike the adaptive immune system, innate components recognise molecules which are common to many pathogens but not to host cells. Cells of innate immunity include natural killer cells, basophils, macrophages and neutrophils. Cells of the innate immune system have pattern recognition receptors which recognise pathogen-associated molecular patterns. One example is the toll-like receptor which is expressed on the surface of many immune cells such as on macrophages. Toll-like receptors recognise many pathogenic features including bacterial cell wall lipopolysaccharides. A major function of the innate immune system is to activate the adaptive response through antigen presentation. Phagocytic cells such as macrophages will present an antigen from a digested pathogen to the T-cells of the adaptive immune system, triggering the adaptive response. A crucial component of the innate immune system which functions to remove pathogens from the blood stream and to target them for immune destruction is the complement system (Meyers, 2007).

1.2 The Complement System

In 1901, Bordet first recognised the complement system as a group of proteins rather than antibodies which carried out an immune function (Bordet, 1901). With the development of protein science and the work of Manfred Mayer it was recognised that the complement system was an enzymatic cascade (Mayer, 1958). Many studies were carried out throughout the remainder of the century which aided in the understanding of the complement system, its components and its role in the immune system. The complement system is part of the innate immune system and consists of a group of proteins which circulate in the blood in an inactive form. It is a multi-step enzymatic cascade which involves the complement proteins C1, C2, C3, C4, C5, C6, C7, C8 and C9. The primary function of the complement system is to remove pathogens or cellular debris before they can cause disease. It does this by a number of mechanisms including eliciting the inflammatory response, opsonisation, and triggering chemotaxis. The complement system is activated through three different pathways, the classical pathway, the alternative pathway and the lectin binding pathway. Despite different mechanisms of activation, the pathways converge to form the C3 convertase enzyme which ultimately

leads to membrane attack complex (MAC) formation. The MAC complex then performs its function of cell lysis and death (Figure 1.1) (Rother, 1998).

1.2.1 The Classical Pathway

The classical pathway of complement activation begins with activation of the C1 protein. C1 is a multi-subunit protein consisting of C1q, C1r and C1s. C1q is a large protein (459 kDa) with a 'tulip-like structure'. It has six globular head domains which are connected via collagen-like domains and is made up of A, B and C chains (Knobel, *et al.,* 1975). C1r and C1s are serine proteases which circulate as pro-enzymes whose activation requires the cleavage of a peptide bond to produce a two chained enzyme linked via a disulphide bridge (Arlaud *et al.,* 1989). C1r and C1s both contain a large globular head with an elongated region connected to a smaller globular domain. The large globular domains contain the active sites (Reid, 1977). Two C1r and two C1s molecules form a tetrameric complex in the presence of calcium which along with one C1q molecule forms C1. C1q interacts with the two C1r and C1s molecules through its collagen like domains (Siegel & Schumaker 1983). Activation of the classical pathway occurs when C1q binds to an activator molecule such as bacterial lipopolysaccharides, cardiolipin, or muccopolysaccharides. Binding occurs through its collagen-like domain. C1q can also bind C-reactive protein through this region and a region close to the globular domain (Loos, 1982; Sim & Reid, 1991; Jiang *et al.*, 1992; Gewurz *et al*., 1993). During an immune response, C1q's globular head domains can bind the Fc region of antigen bound immunoglobulin M (IgM) or IgG (Painter, 1993). Upon ligand binding C1q undergoes a conformational change, which causes the auto activation of C1r, exposing its active site. C1r then cleaves the second C1r molecule resulting in activated C1r. Activated C1r cleaves C1s which is in turn activated. Activation of C1s results in active C1. Activated C1, through its C1s serine esterase activity, cleaves the complement C4 protein into C4a and C4b and complement C2 into C2a and C2b. C4b will bind to the activator surface through its exposed thioester, acting as an opsonin (Law, 1983). C4b is structurally similar to C3b (Mortensen *et al*., 2015). C2a will associate with surface bound C4b and the activator surface. This association forms the classical pathway C3 convertase (C4bC2a). The enzymatic activity of the C3 convertase

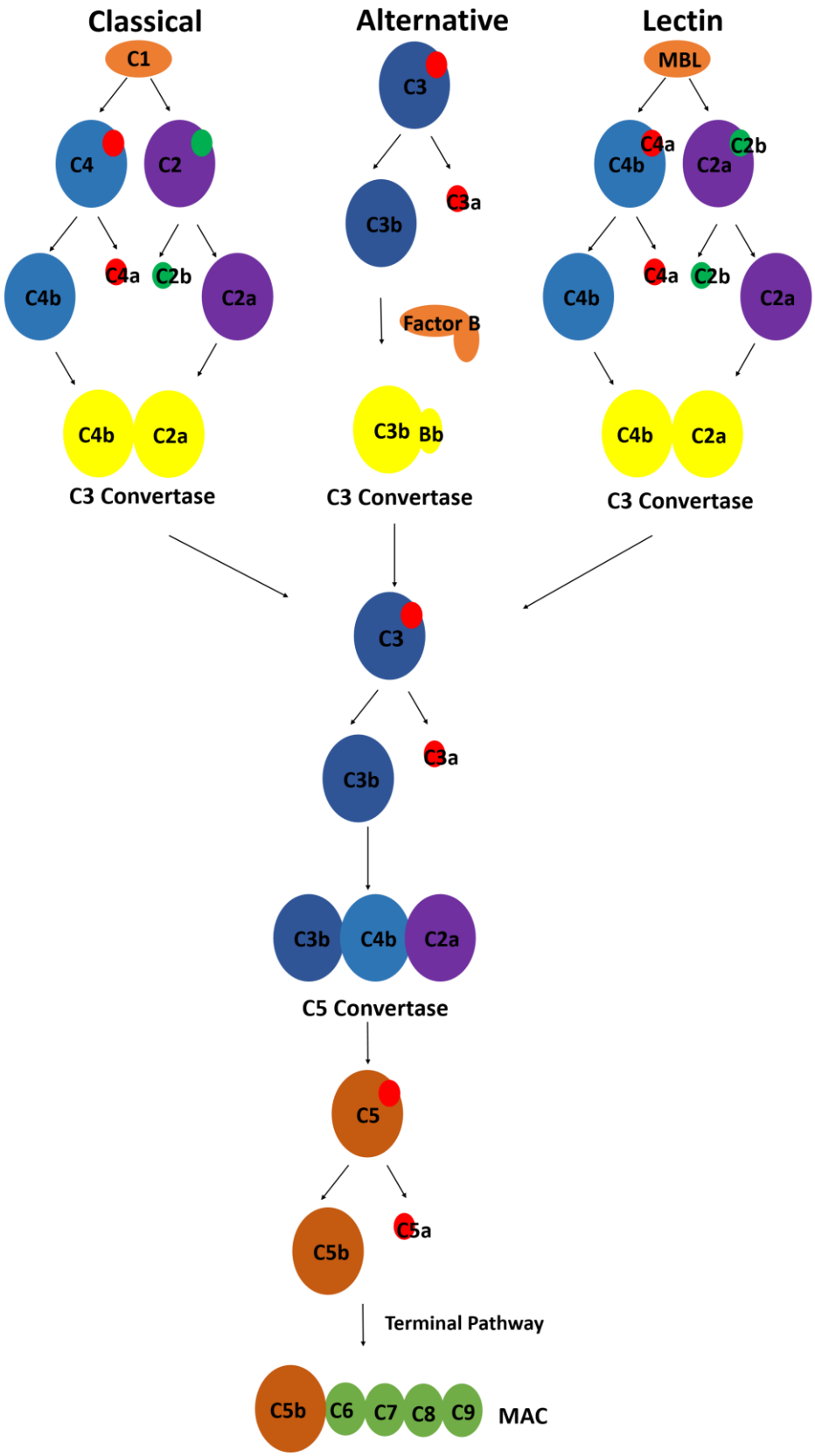

Classical
Alternative
Lectin
C1
C4
C2
C4b
C4a
C2b
C2a
C4b C2a
C3 Convertase
C3
C3a
C3b
Factor B
C3b Bb
C3 Convertase
MBL
C4b
C4a
C2b
C2a
C4b C2a
C3 Convertase
C3
C3a
C3b
C3b C4b C2a
C5 Convertase
C5
C5a
C5b
Terminal Pathway
C5b C6 C7 C8 C9
MAC

Figure 1.1. A schematic representation of the complement system showing the three pathways of activation and their convergence at the formation of the C3 convertase. The classical and lectin pathways have different activator molecules whose binding induces conformational changes in their first component proteins. The alternative pathway is spontaneously activated. All three pathways converge at C3 convertase formation.

will decay rapidly through its release of C2a (Kerr, 1980). The C3 convertase acts on C3 to produce C3a, an anaphylatoxin and C3b. The release of C3a causes C3b to expose its thioester bond allowing it to react with amine and hydroxyl groups of molecules on exposed cell surfaces (Figure 1.2) (Law & Levine, 1977). The C5 convertase (C4bC2aC3b) is formed by C3b binding to C4b of the C3 convertase through a covalent bond between its reactive acyl group and a serine on C4b.

1.2.2 The Lectin Pathway

The lectin pathway of complement is activated by the binding of the mannose-binding lectin protein (MBL) to mannose and other polysaccharides on pathogenic surfaces. MBL exists as a trimer with identical chains of 32 kDa each. It has a collagen-like region, a neck region, and a lectin region responsible for carbohydrate binding (Lu *et al*., 1990; Matsushita *et al.*, 2000). MBL associates through its collagen-like regions to two serine proteases; MBL-associated serine proteases, MASP-1 and MASP-2 which function similarly to C1s and C1r. Carbohydrate binding results in a conformational change in MBL which brings about the activation of MASP-1 through proteolysis to a two chain active enzyme. MASP-1 in turn activates MASP-2 which functions similarly to C1s in its cleavage of C4 to C4a and C4b and C2 to C2a and C2b (Matsushita & Fujita 1992; Thiel *et al*., 1997; Matsushita *et al.*, 1998). The pathway then proceeds similarly to the classical pathway with the formation of the C3 convertase followed by C5 convertase formation (Figure 1.1).

1.2.3 The Alternative Pathway

The alternative pathway (AP) of complement activation provides a more general defence against invading pathogens in that it does not require specific targets such as antibodies in the classical pathway or mannose in the lectin pathway, for its activation. AP activation is spontaneous, consisting of an amplification loop of C3b formation by C3 hydrolysis. Therefore regulation of the AP is therefore crucial in preventing host cell destruction. The spontaneous hydrolysis of C3 into C3a and C3b is the first step of AP activation. C3 is an acute phase protein whose serum concentrations will be elevated in an inflammatory response and is the most abundant complement protein in serum. It is a

185 kDa glycoprotein with two chains α and β linked via a disulphide bond. It contains 13 domains with each chain composed of six domains. The 13th domain is formed by residues of both the α and β chains (Janssen *et al.,* 2005). Spontaneous hydrolysis of the C3 thioester (C3(H_2O)) causes a conformational change allowing Factor B to bind. Factor B is a glycoprotein of 93 kDa whose binding then signals Factor D cleavage of Factor B to Ba and Bb. Bb contains the active protease. Bb remains associated with C3(H_2O) resulting in the formation of the fluid phase C3 convertase (C3(H_2O)Bb) (Figure 1.2). This enzyme can then cleave C3 into the anaphylatoxin C3a (9kDa) and C3b (175 kDa). C3b contains 12 domains which upon activation undergo structural rearrangement to expose various binding sites. Similar to C3 but with the loss of the C3a domain, C3b contains eight macroglobulin domains, a thioester domain (TED), a C345C domain, a complement CUB domain and a linker domain (Janssen *et al*., 2006). The TED is exposed 85 Å away from its position in inactive C3. The cleavage of C3 to C3b and subsequent TED exposure allows C3b to reactively bind to cell surfaces through the formation of an ester or amine bond with a variety of biological molecules, both on pathogenic and host cells (Figure 1.2). Non-bound C3b will be rapidly inactivated by hydrolysis. Surface bound C3b binds Factor B which is then cleaved to Ba and Bb by Factor D. In the presence of the protein properdin on the surface, C3b will associate with Bb (C3bBb) with properdin recruiting fluid phase C3b to the cell surface, promoting the association of Factor B and C3b to form the AP C3 convertase *(*Smith *et al.,* 1984; Hourcade, 2006). This then leads to further C3 hydrolysis thereby amplifying C3 convertase production. In addition to the formation of the C3 convertase causing further C3b production and binding, bound C3b also acts as an opsonin targeting the surface for removal. The C5 convertase of the AP is formed by C3bBb associating with a second C3b ($C3b_2Bb$).

1.2.4 The membrane attack complex

From the formation of the C3 convertase, each of the pathways proceed in a similar manner. The C3 convertase of the classical and lectin pathways has a similar activity to that of the AP C3 convertase which is also the case for the C5 convertase. The C5 convertase binds C5 through a C3b binding site allowing it to be cleaved by C2b

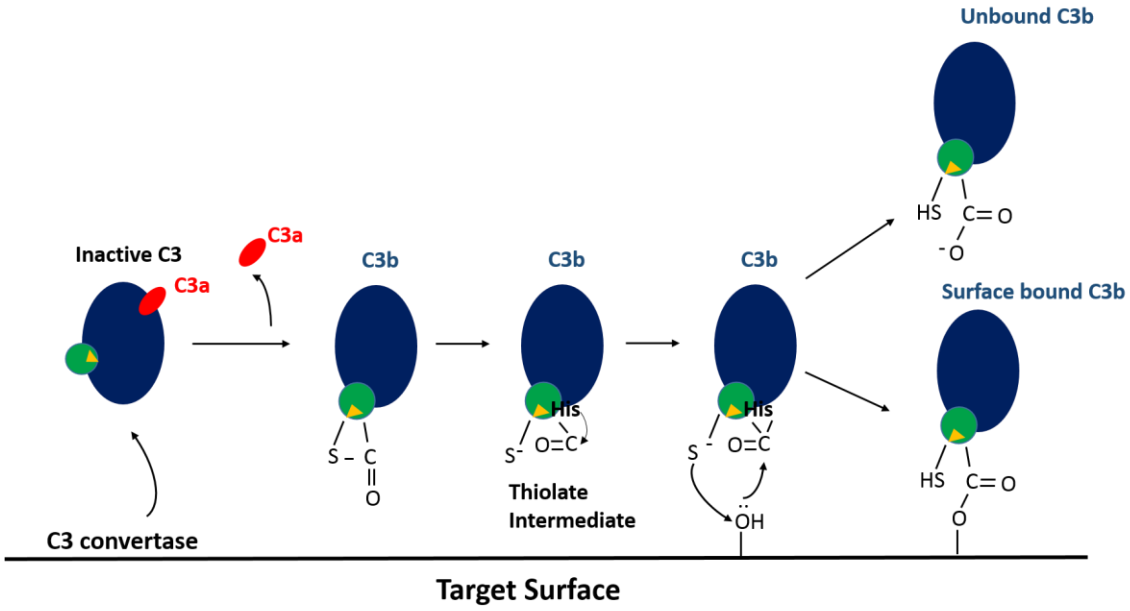


Figure 1.2. C3 activation in the alternative pathway of the complement system. Activation by the C3 convertase results in the release of C3a, inducing a conformational change resulting in the exposure of the thioester domain, ultimately allowing C3b to bind to target surfaces. Unbound C3b will be rapidly inactivated by hydrolysis. C3 is composed of 13 domains, however for simplicity only C3a and the TED are labelled here.

or Bb into C5a and C5b. Cleavage only occurs if this additional C3b is present (Kinoshita *et al*., 1988). C5 is homologous to C3 and C4, and has a similar domain arrangement to C3 but lacks the TED (Discipio *et al.,* 1981; Fredslund *et al.,* 2008). The generation of C5b initiates the formation of the membrane attack complex (MAC) through binding to complement C6. C6C5b then binds C7 which through a conformational change exposes a hydrophobic site. Complement C8 is composed of two proteins, C8α-γ and C8β. C8β binds to C5b in the C5bC6C7 complex. This binding allows the C8α-γ to insert into the lipid bilayer of the targeted cell through its hydrophobic domain. This insertion recruits multiple molecules of C9 which results in the formation of the pore in the membrane. The channel formed is hydrophilic on the inside. The formation of the MAC disrupts the homeostatsis of the cell and allows degradation promoting enzymes to pass into the cell, ultimately leading to cell destruction.

1.3 Regulation of the complement system

Tightly controlled complement regulation is required to ensure that only non-host cells are targeted and removed. Impaired complement regulation is implicated in many immune diseases. A primary method of regulation is the ability of the components of the complement system to distinguish between host and non-host cells. An example of this is seen in the lectin pathway whereby MBL can only interact with mannose exposed on non-host cells as host cells shield their mannose components by other surface carbohydrates. The binding of the complement components often requires stabilising factors on cell surfaces which when not presented results in their decay. For example, the C3 convertase requires properdin to be present on its target surfaces to prevent its quick decay. Some host cell surfaces contain complement receptors which ensure that the complement system will be quickly down-regulated upon binding to these cells. There is also a number of complement factors involved in the regulation of the complement system. Factor I and Factor H both work to regulate the complement system through cleavage and binding of complement components. The complement system is also regulated at the level of the MAC. Complement regulation can occur

either in the solution phase or on the surfaces on which the complement cascade acts and can directly regulate the complement components or their effector molecules.

One of the primary regulators of the classical pathway is the soluble C1 inhibitor (C1INH), which functions to inhibit the C1r and C1s serine proteases. The cleavage of C1INH results in the formation of a stable complex between its C-terminal and C1r and C1s. C1INH will bind to activated C1, bringing about the dissociation of the C1-antibody-antigen complex (Ziccardi & Cooper, 1979). In a similar manner, C1INH acts on MASP-1 and MASP-2 in activated MBL (Kerr *et al.,* 2008). Dissociation of these activator complexes prevents over activation of C4 and C2. The C4b binding protein (C4bp) regulates C4b of the classical and lectin pathways by binding C4b and inhibiting C3 convertase formation (Gigli *et al.*, 1979; Suankratay *et al.*, 1999). C4bp also acts as a cofactor for Factor I which promotes the decay of the C3 convertase through the cleavage of C4b in the complement and lectin pathways and C3b in the AP (Scharfstein *et al.*, 1978; Blom *et al.,* 2003). Complement Factor I facilitates the cleavage of C4b and C3b from the C3 convertase of all three pathways. Factor I has a number of cofactors which it requires for cleavage including C4bp, Factor H, membrane co-factor protein and surface or membrane bound complement receptor 1 (CR1) to carry out its function. Factor I is a serine protease which cleaves the α chain of C3b and C4b preventing the formation of C3 convertase (Davis & Harrison 1982). There are also regulators which act on the later components and on the terminal complex of complement. Complement Factor H related protein 1 (CFHR1) inhibits both C5 convertase formation and MAC formation. It binds C5 and C5b6 preventing C5 dissociation into C5a and C5b thereby preventing further C5 convertase formation. CFHR1 also inhibits the formation of the terminal MAC (Heinen *et al.,* 2009).

In addition to these soluble regulators there are also surface bound receptor molecules which are crucial to complement regulation. There are four complement receptors (CR1-4) which are located on the cell membranes of host cells. CR1 (also known as CD35) is a receptor for C3b, C4b, iC3b (inactive C3b) and C1q. It is present on numerous host immune cells including erythrocytes, neutrophils, macrophages and B lymphocytes (Law & Reid, 1995). Its biological role depends on the cells it is expressed

on. Erythrocytic CR1 will bind C3b and C4b which are covalently attached to surfaces facilitating their removal from circulation. It can also facilitate the proteolysis of surface bound C3b and C4b as a cofactor for Factor I, leading to the deposition of C3d and C4d on the surfaces which are ligands for the other complement receptors (Krych-Goldberg & Atkinson 2001). CR2 (also known as CD21) is a receptor for C3dg, C3d and iC3b. CR2 is present on B cells and T cells. CR2 plays an important role in linking the complement system with adaptive immunity. CR2 bound to C3d will form a complex with CD19 and CD81 with membrane IgM involvement causing an enhancement in immune activation and IgM antigen binding (Carter *et al.,* 1991). CR3 and CR4 are both receptors for iC3b promoting phagocytosis of the opsonised cells (Vik & Fearon 1987). CD55 or the decay accelerating factor (DAF) is a cell surface receptor for the C3 convertases. It is present on a number of surfaces including endothelial and epithelial cells. CD55 promotes the decay of the C3 convertase by binding to C3b and C4b preventing the cleavage of C2 to C2b and Factor B to Bb preventing further formation of C3 convertase (Lublin & Atkinson 1989). CD59 (protectin) is present on the surfaces of erythrocytes and nucleated cells. It prevents the formation of the MAC by binding C9, preventing its association with C8 (Miwa & Song 2001). Figure 1.3 summarises the regulators involved in complement control and the complement components upon which they act.

1.4 The Complement System in Disease

Deficiencies of C1, C2 and C4 of the classical pathway are associated with systemic lupus erythematosus (SLE), commonly known as lupus. It is an autoimmune disease which can affect many of the internal organs and is associated with improper clearance of apoptotic cells. The main symptoms of SLE are joint pain, fatigue and skin rash with more severe cases involving the internal organs. Approximately 15,000 people in the UK have SLE with the majority of cases occurring in women. The protein deficiencies are mostly hereditary and autosomal recessive. A deficiency of C1q is most commonly linked to SLE although reduced levels of C1r and C1s are also observed. A reduction of C4 as a result of the deletion of the C4A gene is the most common reason for a reduction in C4 levels. An individual with reduced C4 levels will be predisposed to

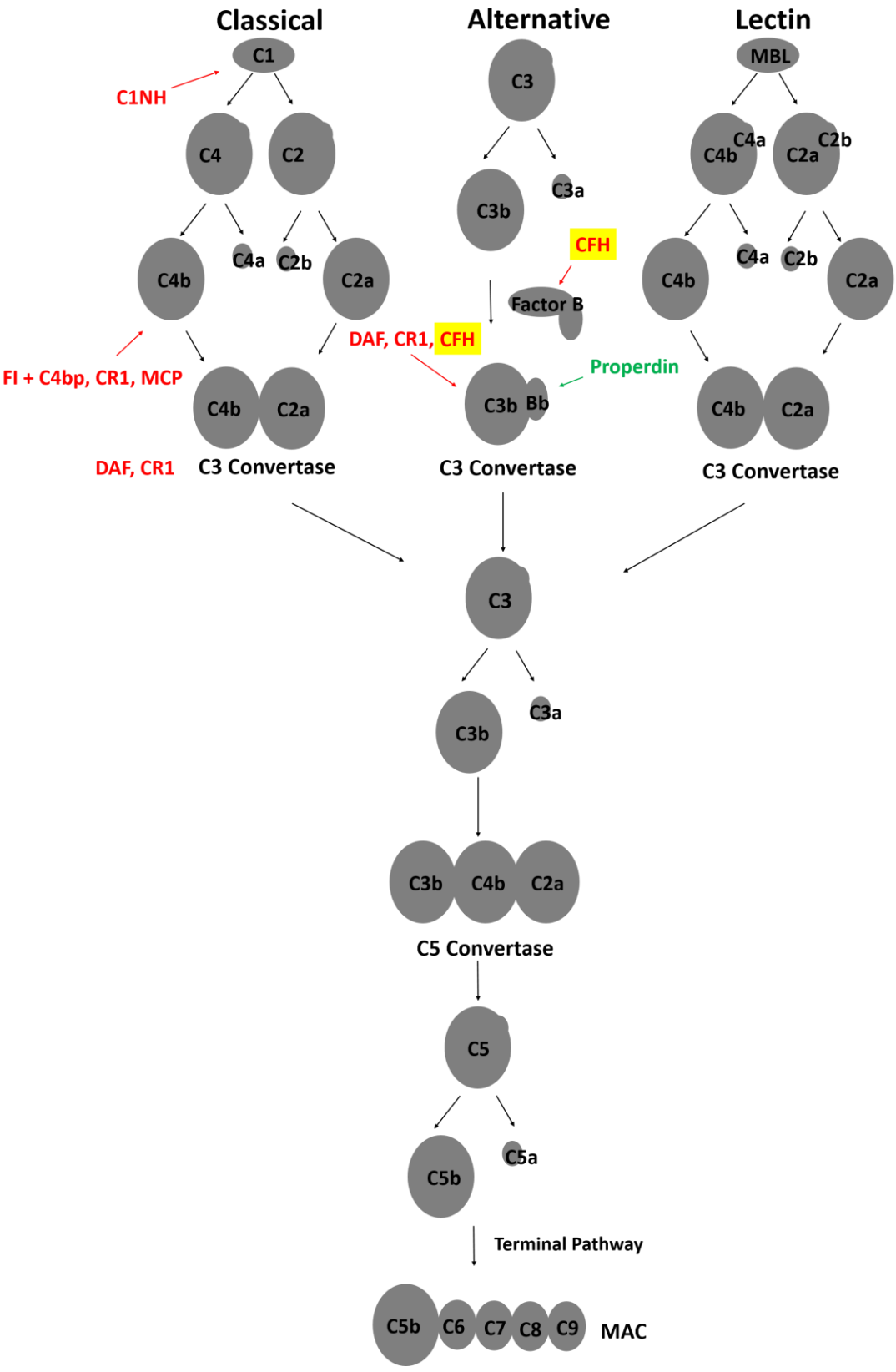

Classical
Alternative
Lectin
C1
C1NH
C4
C2
C4b
C4a
C2b
C2a
FI + C4bp, CR1, MCP
C4b
C2a
DAF, CR1
C3 Convertase
C3
C3b
C3a
CFH
Factor B
DAF, CR1, CFH
Properdin
C3b Bb
C3 Convertase
MBL
C4b
C4a
C2b
C2a
C4b
C2a
C3 Convertase
C3
C3b
C3a
C3b
C4b
C2a
C5 Convertase
C5
C5b
C5a
Terminal Pathway
C5b
C6
C7
C8
C9
MAC

Figure 1.3. A schematic representation of complement regulation. Complement regulation is tightly controlled to prevent over activation and destruction of host cells. The positive regulators are shown in green i.e. those which promote activation, while the negative are shown in red.

Table 1.1 Complement Associated Diseases

Disease	Associated complement molecule	Primary Cause	Main symptoms	Prevalence*
Systemic Lupus Erythematosus	C1,C2,C4	Molecule Deficiency Gene deletion	Inflammation throughout the body	~15,000
Paroxysmal Nocturnal Haemoglobinuria	CD55, CD59	Molecule Deficiency	Anemia Thrombosis	<1,000
Hereditary Angioedema	C1 inhibitor	Molecule Deficiency	Swelling of skin, respiratory and gastrointenstinal tracts	>10,000
Membraneproliferative Glomerulonephritis	Complement Factor H	Molecule Deficiency	Water retention, Blood in urine followed by kidney failure	~ 6,000
atypical Haemolytic Uraemic Syndrome	Complement Factor H	> 100 mutations	acute kidney failure	170
Age-related Macular Degeneration	Complement Factor H	Y402H Polymorphism	Drusen deposits in the retina of the eye Vision loss	>500,000

* Prevalence in the UK, taken from the national health service

developing SLE. It is also associated with other diseases such as scleroderma and IgA nephropathy (Law & Reid, 1995). A deficiency of C3 is linked to SLE-like illness and can also lead to an increased susceptibility to pyogenic infections such as pneumonia and meningitis (Botto *et al*., 1992). Deficiencies in Factor H, properdin and Factor I are also associated with increased susceptibility to pyogenic infections (Lachmann, 2002; Józsi & Zipfel 2008). Deficiencies of the components of the MAC are associated with recurring *Neisseria meningitidis* and *Streptococcus pneumoniae*. Deficiencies or mutations in the regulatory proteins of complement will lead to excessive activation of complement and destruction of host cells. Deficiencies of CD55 and CD59 are associated with paroxysmal nocturnal haemoglobinuria (PNH) where the red blood cells are destroyed due to a lack of CD55 and CD59, allowing C3 convertase formation and therefore destruction of the cells. PNH is a rare disease which affects less than 1.5% per million people however it is a severe disease with patients susceptible to blood clots which can be life threatening. Deficiencies of C1INH transmitted as autosomal dominant lead to hereditary angioedema which can be severe depending on the location of the infection. Hereditary angioedema is extremely rare with 1 in 50,000 people affected worldwide. The symptoms of the disease vary with swelling commonly observed in the sub-dermal tissues. Risk of death arises when angioedema progresses rapidly with a high risk of blocking airwaves. Low levels of Factor H are associated with the development of membraneproliferative glomerulonephritis (MPGN) or dense deposit disease which is due to uncontrolled regulation of C3 activation and amplification of the alternative pathway. MPGN affects approximately 1 in 10,000 people with children and young adults primarily affected. The majority of cases are mild however in some cases MPGN can lead to kidney failure. Mutations and polyporphisms in Factor H are associated with atypical haemolytic uraemic syndrome (aHUS) and age-related macular degeneration (AMD). aHUS is a rare disease with 170 cases currently reported in the UK. AMD is one of the leading causes of blindness in individuals above the age of fifty with over 500,000 cases reported in the UK. Both AMD and aHUS are discussed in more detail in Chapter Two (Klein *et al.,* 2005; Haines *et al.,* 2005; Edwards *et al.*, 2005; & Hageman *et al.,* 2005; Zipfel *et al*., 2006).

Chapter Two

Complement Factor H

2.1 Complement Factor H biosynthesis and Expression

Complement Factor H (CFH) is a 154 kDa glycoprotein found in human plasma at concentrations of 2-8 mg/L (Saunders *et al*., 2006; Fenaille *et al*., 2007). It was first identified in 1965 by Nilsson and Muller-Eberhard as β1H globulin (Nilsson & Muller-Eberhard, 1965). CFH belongs to a gene family which includes the Factor H-like proteins and the complement Factor H–related (CFHR) proteins. This group of proteins have conserved structural elements with high sequence identity between some but not all domains (Józsi & Zipfel 2008). The CFH gene is located in the regulator of complement activation (RCA) gene cluster on chromosome 1q32 and is encoded by a single gene called *HF1* (Rodríguez De Córdoba *et al.,* 1999). *HF1* has 23 exons over a region of 94 kb of genomic DNA (Male *et al.,* 2000). Each of the short complement regulator (SCR) domains of CFH are encoded by a single exon with SCR-2 encoded by exons 3 and 4 (Rodríguez De Córdoba, Esparza-Gordillo, Goicoechea De Jorge, Lopez-Trascasa, & Sánchez-Corral, 2004). CFH expression is regulated and induced by the proinflammatory cytokine IFNγ and the T-cell growth factor (Brooimans *et al.,* 1989). Expression occurs primarily in the liver but has also been detected in other cells including myoblast, neurons, glial cells and in the retinal pigment epithelium and choroid (Schwaeble *et al.*, 1987; Friese *et al.,* 1999; Wistow *et al.*, 2002).

2.2 Complement Factor H structure

2.2.1 Factor H SCR domains

CFH is comprised of 20 short complement regulator (SCR) domains which are also found in other proteins of the complement system (Figure 2.2, 2.3). SCR domains are connected by inter-SCR linkers which vary in their flexibility depending on the length and amino acid content (Perkins *et al.,* 2002). Each SCR domain contains approximately 60 amino acids with the inter-SCR linkers ranging from three to eight amino acids. Each SCR domain contains four conserved cysteine residues that form two disulphide bridges through a 1-3 and 2-4 arrangement. Each domain also contains a conserved tryptophan residue and a number of conserved hydrophobic residues (Figure 2.1). Each SCR domain contains between six to eight β strands with two anti-parallel strands which run along its axis (Figure 2.3) (Rodriguez *et al.,* 2014). The β strands are linked by loop regions with β2 and β3 linked by a hypervariable loop with non-conserved residues. CFH contains nine glycoyslation sites, eight of which have

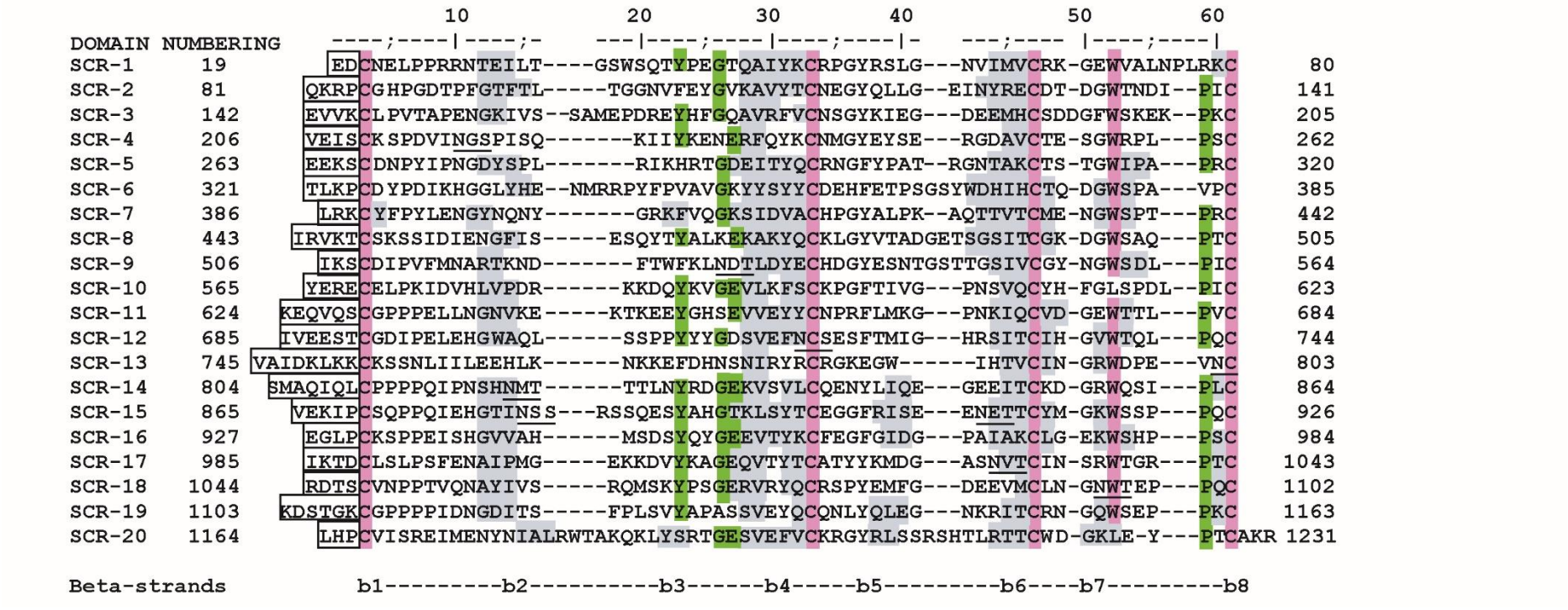


Figure 2.1 Sequence alignment of the 20 SCR domains of Factor H. The conserved cysteine residues and tryptophan residues are shown in pink boxes. Additional conserved residues are shown in green. The grey boxes indicate the assigned β-strand secondary structures and are labelled below the alignment.

been shown to be occupied with the SCR-4 site unoccupied (Figure 2.2) (Kristensen & Tack, 1986; Aslam & Perkins 2001; Fenaille *et al*., 2007).

The 20 SCR domains of Factor H are arranged in a bead like structure with the inter-SCR linkers providing inter-domain flexibility (Figure 2.3). Due to the size, glycosylation and flexibility of CFH, attempts to solve the full length structure have proved difficult. Structural analysis has been carried out on several of the SCR domains using a range of techniques which show that it is difficult to predict the orientation between two SCR domains (Perkins *et al*., 2002). Nuclear magnetic resonance (NMR) and X-ray crystallography have been used to determine high resolution structures for the SCR domains but are generally limited to one to two domains. Electron microscopy (EM) has been used to examine the full length structure but is limited to a static image of CFH which is often obtained under harsh conditions. Multi-disciplinary approaches have been taken which are most effective for the elucidation of structural information about full length CFH. Solution studies using small angle scattering, analytical ultracentrifugation (AUC) and EM have shown that CFH adopts a partially folded back structure which is not fully extended in solution (Perkins *et al.*, 1991; DiScipio 1992). Studies carried out in solution on full length CFH reveal that weak electrostatic interactions between the SCR domains are responsible for this folded back structure (Okemefuna, Nan, Gor, & Perkins, 2009). High resolution information is available for some of the SCR domains including NMR structures for SCR-1/3 (Hocking *et al*., 2008), SCR-5 (Barlow *et al.*, 1992), SCR-10/11 and SCR-11/12 (Makou *et al.*, 2012) and SCR-15/16 (Barlow *et al.*, 1993). X-ray crystallography has also provided high resolution information for SCR-6/8 (Prosser *et al.*, 2007; Herbert *et al.*, 2007) SCR-18/20 (Morgan et al., 2012), and SCR-19/20 (Jokiranta *et al.* 2006). Low resolution solution structures are available for SCR-1/5 and SCR-16/20 (Okemefuna et al., 2008). SCR-1/5 adopts a bent back structure in solution as does SCR-16/20 where the SCR domains are not fully extended. Table 2.1 summarises the current structures available for CFH that have been deposited in the protein data bank (PDB).

2.2.2 Factor H C-terminal structure

Structural information available for the CFH C-terminal region reveals that it has a bent back structure which most likely originates from the flexible linker between

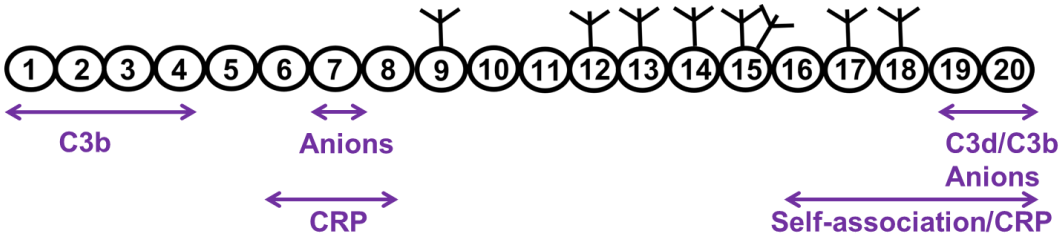


Figure 2.2 Schematic illustration of the 20 Complement Factor H short complement regulator (SCR) domains and their binding properties. Glycosylation sites are shown (**Y**) along with the binding partners associated with the SCR domains. The self-association site refers to the Complement Factor H dimerisation site.

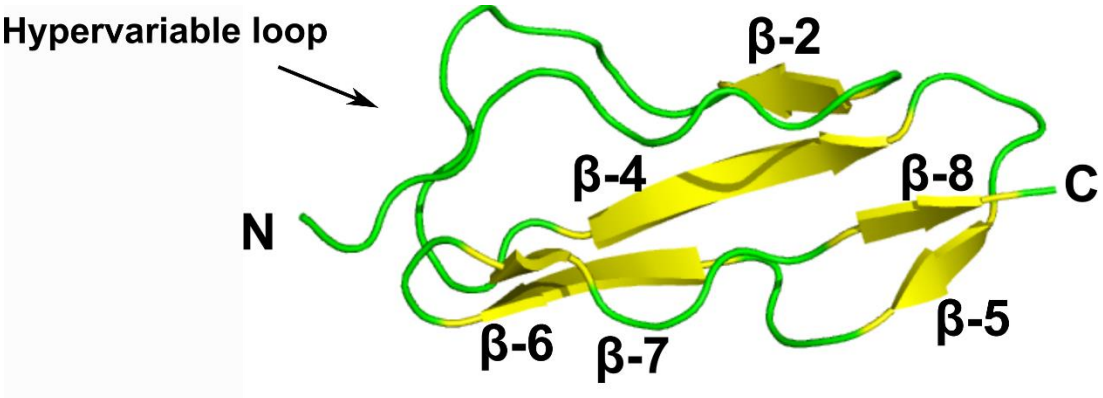


Figure 2.3. A typical short complement regulator (SCR) domain. SCR domains commonly occur in proteins of the complement system. The domains all contain similar structural elements. Each domain contains between six to eight β-strands and a hypervariable loop which is observed here between β2 and β4. Each SCR domains has two disulphide bridges. CFH is composed of 20 SCR domains.

Table 2.1. The structures available for CFH which have been deposited in the protein data bank (August 2015).

CFH domains	PDB code	Method of determination[a]	Reference
SCR-1/2	2RLP	NMR	Hocking *et al*, 2008
SCR-2/3	2RLQ	NMR	Hocking *et al*., 2008
SCR-1/5	2QFG	SAXS	Okemefuna *et al*., 2008
SCR-6/8	2IC4	SAXS	Fernando *et al.*, 2007
SCR-6/8	2UWN	Crystallography	Prosser *et al*., 2007
SCR-7	2JGW	NMR	Herbert *et al*., 2007
SCR-15	1HFI	NMR	Barlow *et al*., 1993
SCR-15/16	1HFH	NMR	Barlow *et al*., 1993
SCR-16	1HCC	NMR	Barlow *et al.*, 1993
SCR-16/20	2QHF	SAXS	Okemefuna *et al*., 2008
SCR-18/20	3SW0	Crystallography	Morgan *et al*., 2012
SCR-19/20	2BZM	NMR	Herbert *et al.*, 2006
SCR-19/20	2G7I	Crystallography	Jokiranta *et al*., 2006
SCR-1/20	1HAQ	SAXS	Aslam and Perkins, 2001
SCR-1/20	3GAV	SAXS	Okemefuna *et al.*, 2009

[a] Abbreviations: NMR, nuclear magentic resonance, SAXS, small angle X-ray scattering

SCR-18 and SCR-19 (Okemefuna *et al*., 2008; Morgan *et al*., 2012). The crystal structure of SCR-18/20 shows a bend of 122º between SCR-18 and SCR-19. The structures show that this bend is exaggerated in the crystal structure with the three domains adopting multiple conformations in solution. SCR-16, SCR-18 and SCR-19 contain the typical β-sheet structure of an SCR domain while SCR-20 contains a short α-helix (Barlow *et al*., 1993; Jokiranta *et al*., 2006; Morgan *et al*., 2012). No high resolution information is available for SCR-17, SCR-16, SCR-17, SCR-18 and SCR-19 have charges at physiological pH of -1, -1, -3 and 0, respectively. SCR-20 contains a patch of basic residues which at physiological pH has a charge of +8 which has the highest charge of any of the CFH SCR domains (Jokiranta *et al*., 2006; Okemefuna *et al*., 2009).

2.2.3 Factor H self-association

The first evidence of CFH self-association came from small angle scattering experiments which showed that at concentrations of 13 µM CFH existed as monomer and dimer (Perkins *et al*., 1991). The result of this was disputed as the concentrations used were significantly higher than physiologically relevant (DiScipio, 1992). However with developments in AUC data analysis, it was confirmed that CFH exists in a monomer-dimer equilibrium with a K_D of 28 µM, implying that at plasma concentrations of between 4-15 % it will exist as a dimer (Nan *et al*., 2008). CFH contains two self-association sites at SCR-6/8 and SCR-16/20 (Fernando *et al*., 2007; Okemefuna *et al*., 2008). Both fragments exist in a weak monomer-dimer equilibrium with the SCR-16/20 self-association site stronger with a dissociation constant (K_D) of 16 µM. SCR-6/8 dimer formation has an estimated K_D of 40 µM.

In this thesis, the C-terminal SCR-16/20 dimer site was further investigated. It was initially proposed that self-association was mediated through an interaction between the SCR-20 domains of CFH (Okemefuna *et al*., 2008). Using size exclusion chromatography, analytical ultracentrifugation, and small angle X-ray scattering we have shown that the SCR-17/18 domains contain the C-terminal CFH dimer site which may explain the presence of disease-associated mutations within this region.

2.3 Factor H regulation in complement

2.3.1 Factor H regulates C3b

CFH down-regulates the alternative pathway of complement activation through regulation of the C3 convertase (C3bBb) and C3b in plasma and on host surfaces. CFH binds to C3b, competing with Factor B and acting as a cofactor for Factor I mediated proteolysis of C3b to iC3b (Whaley & Ruddy, 1976; Pangburn *et al*., 1977). Through its decay accelerating activity, CFH brings about the dissociation of Bb from the C3 convertase (Weiler *et al.,* 1976). CFH readily binds fluid phase C3b but will only interact with surface bound C3b when specific markers are present on the surface which distinguishes it from non-host. CFH distinguishes between host and non-host surfaces by interacting with host surface sialic acids, glycosaminoglycans, and sulphated polysaccharides such as heparin sulphate. The presence of these molecules on surfaces increases CFH affinity for C3b (Fearon, 1978). CFH binds to anionic surfaces through its SCR-7 and SCR-20 domains (Blackmore et al. 1996, 1998; Oppermann *et al.,* 2006; Blaum *et al.,* 2015) (Figure 2.3). SCR-1/4 contains a C3b binding site and the decay accelerating activity of CFH (Gordon *et al.,* 1995; Kühn & Zipfel, 1996). All four of the SCR-1/4 domains make contact with regions of C3b which have become exposed upon C3a removal. The binding site on C3b for SCR-1/4 of CFH overlaps with the C3b binding site for Factor B explaining the competition for C3b binding between Factor H and Factor B (Forneris *et al.,* 2010; Wu *et al.,* 2009). SCR-19/20 binds the C3d (TED) region of C3b and SCR-6/10 binds the C3c region of C3b (Lambris *et al.,* 1988; Sharma & Pangburn, 1996; Jokiranta *et al.,* 2000, 2001). In order to understand how CFH regulates C3b it is necessary to understand how CFH interacts with C3b through each of these binding sites.

2.3.2 Factor H interacts with C3d

C3d remains bound to the surface after the sequential degradation of C3b to iC3b, C3c, C3dg and finally C3d by Factor I. C3d functions in bridging the innate and adaptive immune systems through its interaction with complement receptor 2 (CR2) which is expressed on B-cell surfaces (Weis *et al.,* 1984). Surface bound C3d binds CR2, increasing B-cell sensitivity thereby initiating the adaptive immune response. The structure of C3d has been solved by X-ray crystallography revealing that C3d has an α-α barrel structure which contains five short segments of 3_{10} helices (Figure 2.4) (Nagar *et al.,* 1998). One end of the barrel is acidic while the other end contains the TED and the residues involved in covalent attachment to antigens including Cys17, Gln20 and His133. Cleavage of C3 results in the Gln1013 (Gln20 of cleaved C3d) and

Cys1010 (Cys17 in cleaved C3d) side chains being brought into close proximity allowing them to form a reactive thiolate. His1126 (His133 in cleaved C3d) undergoes nucleophilic attack from the formed reactive thiolate resulting in the formation of an acyl-imidazole intermediate which results in the release of the Cys1010 thiol. This thiolate anion then catalyses the attack of the cell surface nucelophiles. The CR2 binding site is located on the acidic end of the barrel which is accessible when C3d is surface bound (van den Elsen & Isenman, 2011).

SCR-19/20 of CFH was recognized as a C3d binding site by SPR (Jokiranta *et al*., 2000). Initial structural studies on the CFH and C3d interaction were carried out using the SCR-16/20 fragment of CFH. Solution studies of the SCR-16/20 and C3d interaction reveal a multimeric interaction. AUC, SPR, solution scattering and molecular modelling showed that dimeric SCR-16/20 bound C3d with a stoichiometry that could not be described as a 1:1 binding. This was hypothesised to result from the binding of one C3d molecule to each of the CFH molecules in the dimer (Okemefuna *et al.*, 2009). Subsequent to this solution study, two crystal structures were published which showed different binding stoichiometries between C3d and SCR-19/20 (Kajander *et al,*. 2011; Morgan *et al.*, 2011). The Kajander crystal structure revealed two C3d molecules bound to one SCR-19/20, with a binding site in SCR-19, and an additional binding site in SCR-20. SCR-19 binds the C3d component of a C3b molecule with SCR-20 capable of binding an additional C3d molecule. This group suggested that 1:2 binding could play a role in target discrimination in the complement system. Surface bound C3d binding to SCR-20 would replace SCR-20 binding of anionic surfaces, protecting the cell from immune destruction. In contrast, Morgan *et al*. showed that only one C3d molecule interacted with SCR-19/20 and that this binding site was located in the SCR-19 domain. It was claimed that the second C3d binding site in SCR-20 was an artefact of crystallisation. A third crystal structure of the C3d and SCR-19/20 interaction with a sialylated trisaccharide was published in 2015 (Blaum *et al*., 2015). This crystal structure showed that in the presence of sialic acid molecules, SCR-19 binds C3d while SCR-20 interacts with the sialic acid through both hydrogen bond and salt bridge formation. This study confirmed that in the presence of host surface recognition markers the SCR-20 domain of CFH binds these while SCR-19 binds either C3d which has been cleaved from C3b, or C3d which still remains part of C3b. Binding studies in solution suggested that SCR-19/20 can bind

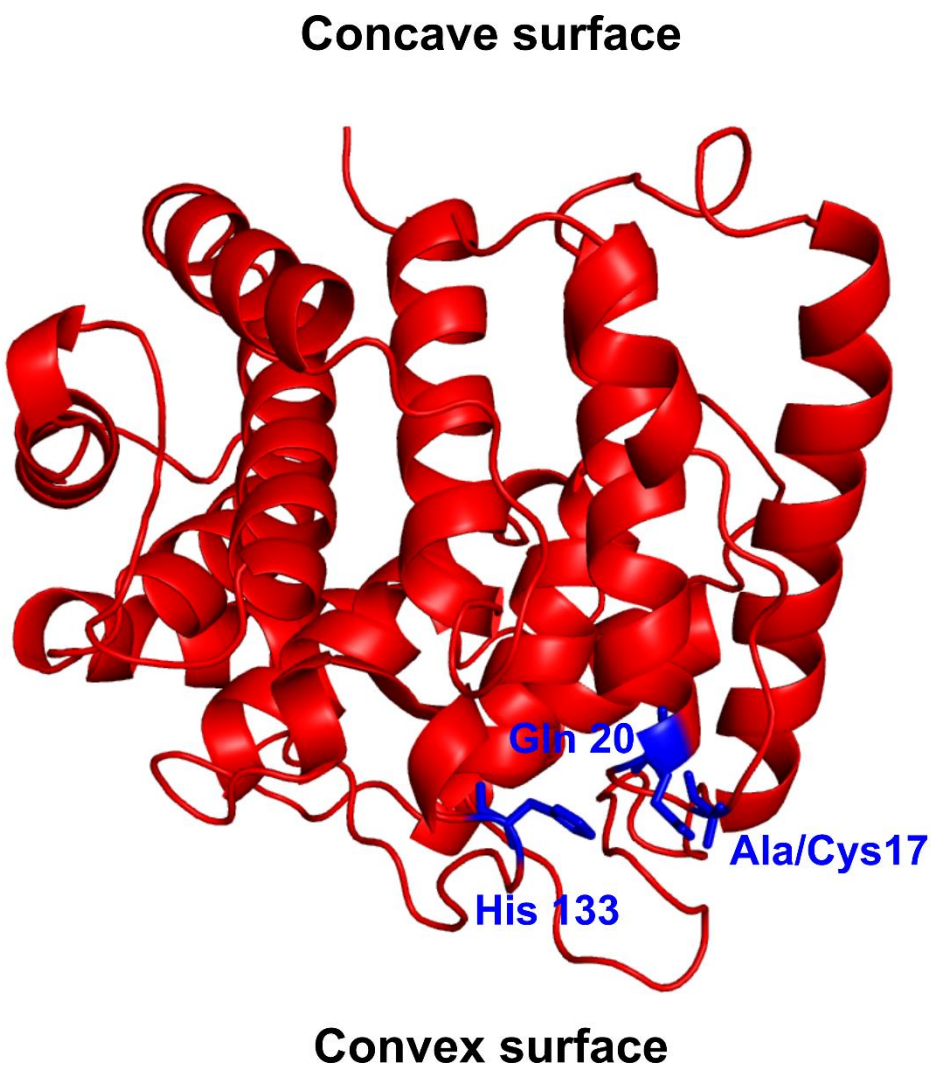


Figure 2.4 Ribbon representation of the crystal structure of C3d showings its α-α barrel structure. The thioester residues responsible for surface binding are shown in blue at the convex surface, these are Ala17 (Cys17 in the wild-type), Gln 20 and His 133. The concave surface contains the CR2 binding site along with the SCR-19/20 binding site (PDB Code: 1C3D).

C3d in both a 1:1 and a 1:2 stoichiometry with dissociation constants (K_D) of 8 μM and the other 1 μM. This study also proposed that the SCR-19 and C3d interaction is driven by hydrogen bond formation and hydrophobic interactions while the SCR-20 and C3d interaction is driven by electrostatic interactions (Rodriguez, unpublished 2014).

In this thesis we confirm that SCR-19/20 and C3d interact with both 1:1 and 1:2 binding stoichiometries in solution in near physiological conditions (in the absence of sialic acids). We used small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) to show that in the absence of other factors, a 1:2 binding stoichiometry is prevalent. Our results confirm that SCR-20 binding to C3d is driven by electrostatic interactions which are weakened with increasing NaCl concentration.

2.4 Factor H and C-reactive protein

2.4.1 C-reactive protein

C-reactive protein (CRP) is a 115 kDa Ca^{2+} dependent acute phase protein involved in the activation of the complement system via the classical pathway (Volkanis, 1978, 1982). Regular serum levels of CRP are between 0.8 mg/L and 10 mg/L (Pepys & Hirschfield, 2003). During an acute inflammatory response - such as from an infection - the CRP concentration can increase up to 10,000 times. The CRP gene is located on chromosome 1 between 1q21 and 1q23 and it has one intron with 1131 bp (Woo *et al.*, 1985). CRP is primarily produced in the liver with transcription controlled by interleukin-6 (IL-6) (Toniatti *et al.*, 1990). Transcription of CRP is also enhanced by IL-1, glucocorticoids and complement activation products through their interaction with IL-6 (Ganapathi *et al.*, 1991; Szalai *et al.*, 2000). CRP is a member of the pentraxin family of proteins characterised by its cyclic pentameric structure consisting of five identical protomer subunits. The crystal structure of CRP has been solved to 3 Å (Figure 2.5). Each protomer subunit has 206 amino acids with a molecular weight of 23 kDa. The subunits are arranged in a cyclic pentamer with five-fold symmetry. Each protomer contains two anti-parallel β-sheets with a single helix. The calcium binding site is located on the opposite side to the helix, at the calcium binding loop. Each protomer coordinates two Ca^{2+} ions. The phosphocholine binding site is located away from the calcium-binding site in a deep cleft on the pentameric face, suggesting that calcium is not directly involved in binding but may act as an allosteric effector (Shrive *et al.*, 1996). CRP exists in a pentamer-decamer equilibrium

in physiological conditions with a K_D value of approximately 23 μM. The decamer forms only in the presence of Ca^{2+} and is formed by an interaction of the two faces of the pentamer (Okemefuna *et al*., 2010).

2.4.2 CRP in complement

CRP functions as an activator of the classical pathway of complement. CRP recognises pathogens and damaged host cells primarily through its interaction with exposed phosphocholine. However, it also interacts with chromatin, histones, fibronectin, microbial surface proteins and polycations (Volanakis & Kaplan, 1971, Black *et al.,* 2004). In its surface bound state, CRP interacts with the C1q component of C1 initiating the activation of the classical pathway (Siegel *et al.*, 1975). Each CRP protomer contains a C1q binding site on the opposite side of the protomer to the phosphocholine binding site (Agrawal *et al.,* 2001). CRP can also interact directly with phagocytic cells through the Fc receptors for IgG (Crowell *et al.,* 1991; Marnell *et al.*, 1995; Gershov *et al.,* 2000). CRP interactions with these cells lead to phagocytosis of the targeted cell. CRP inhibits complement activation through the alternative pathway (Mold *et al.*, 1999). The mechanism for this remains largely unclear although it has been suggested that it is mediated through CRP induced up-regulation of the decay accelerating factor (DAF), membrane co-factor protein (MCP) and CD59 (Li *et al.*, 2004). It has however also been suggested that its interaction with CFH enables its inhibitory role in complement (Mold *et al.,* 1999; Geshov *et al*., 2000).

2.4.3 Factor H interaction with CRP

The CRP-CFH interaction was first identified in 1999, using enzyme-linked immunosorbant assays (ELISA) in which CRP was immobilised to the ELISA wall (Mold *et al*., 1999). However, later studies revealed that CRP had become denatured upon immobilisation and so the existence of the interaction came into question. Using a combination of biophysical methods in solution and on surfaces it was shown in 2009 that this interaction did exist in a Ca^{2+} dependent manner (Okemefuna *et al*., 2009). This study determined the stability of CRP in various buffer conditions revealing that CRP requires both physiological salt (140 mM) and 2 mM Ca^{2+} to remain in its native pentameric structure. Under these conditions, it was confirmed that CFH binds CRP. CFH binds CRP with a binding affinity of 3.4 μM. CFH contains two distinct binding sites for CRP, one in the SCR-6/8 region and one in the C-terminal SCR-16/20 region. The functionality of the CFH-CRP interaction remains relatively

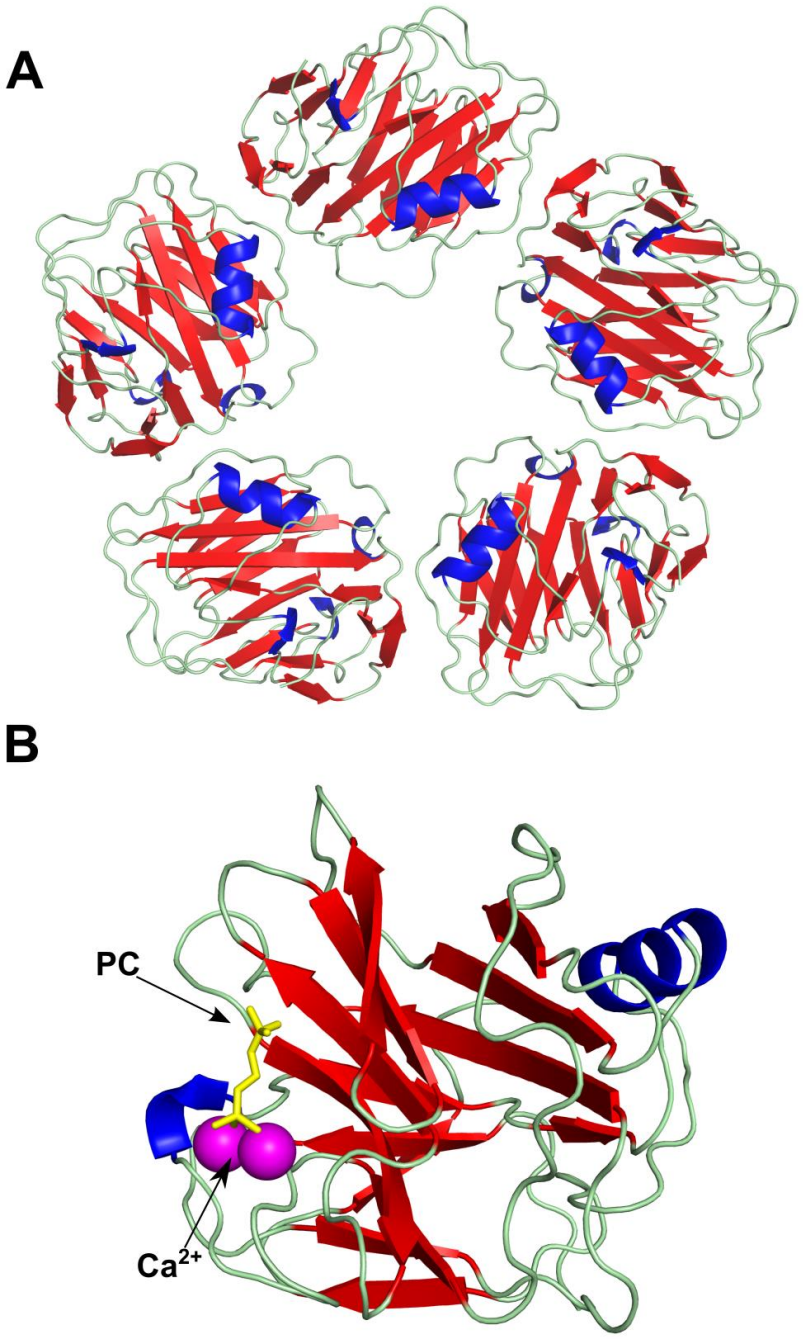


Figure 2.5. The crystal structure of the pentameric C-reactive protein. *A*, The five identical protomer subunits of CRP. The β-sheets are shown in red, α-helix in blue and the loop structures in green. *B*, A CRP protomer with bound Ca^{2+} (pink spheres) and phosphocholine (PC; molecule in yellow) with the binding sites indicated (PDB code: 1GNH).

unknown. It is thought that through its interaction with CFH, CRP inhibits complement activation via the alternative pathway (Mold *et al*., 1999; Geshov *et al*., 2000). The CRP-CFH interaction has been linked with age-related macular degeneration (AMD) with binding studies showing that the AMD-associated H402 and the wild-type Y402 variants bind CRP with different affinities (Laine et al. 2007; Sjöberg *et al.*, 2007; Skerka *et al.*, 2007; Yu *et al.*, 2007). The H402 variant has a weaker affinity for CRP than the Y402 variant suggesting that the CRP and CFH interaction is functionally important (Okemefuna *et al*., 2009). The weaker binding affinity of the disease-associated variant may cause insufficient CRP and CFH binding and therefore complement regulation. AMD patients who have the H402 variant have elevated levels of CRP in the choroid of the eye (Johnson et al., 2006).

While the SCR-6/8 CFH binding site for CRP has been studied, little information is available on the SCR-16/20 binding site. SCR-16/20 contains ayptical haemolytic uraemic syndrome (aHUS) associated mutations Hence a study of its interaction with CRP may shed light on the disease mechanism of aHUS (Saunders *et al.*, 2006; Rodriguez *et al.*, 2014). In this thesis the interaction between SCR-16/20 and CRP was studied. The results reveal that the SCR-19/20 domains of CFH contain a weak CRP binding site. This weak binding suggests that SCR-6/8 contains the primary binding site for CRP. The initial binding of SCR-6/8 may bring the C-terminal domains close to the target surface allowing SCR-19/20 to bind surface bound CRP. aHUS mutations in this region may impair SCR-19/20 binding to CRP thereby impairing complement control on host surfaces (Chapter Five).

2.5 Complement Factor H-associated diseases

Complement Factor H plays an essential role in regulating the complement system both in the fluid phase and on host surfaces. Deficiencies of CFH and mutations and polymorphisms within the protein sequence can cause severe inflammatory and autoimmune diseases. Mutations within CFH are classified as type I or type II mutations. Type I mutations are described as those that cause a reduction in the protein plasma concentrations such as mutations which bring about defects in secretion. Type II are those which show a reduction in CFH function (Perkins & Goodship, 2002). Over 130 mutations and polymorphisms within CFH have been linked to severe disease (Saunders *et al*., 2006; Rodriguez *et al*., 2014). Type I mutations in CFH are associated with membrane proliferative glomerulonephritis (MPGN) and type II are

Figure 2.6. Schematic diagram of Complement Factor H showing the aHUS disease associated mutations.

associated with aHUS and AMD (Klein *et al.*, 2005; Haines *et al*., 2005; Edwards *et al.*, 2005, Hageman *et al.*, 2005; Dragon-Durey *et al.*, 2005). These mutations are located throughout CFH, although many of them are clustered in the C-terminal SCR-16/20 domains, a region whose binding properties are crucial for proper complement regulation (Figure 2.6). AMD is one of the leading causes of blindness in individuals aged over 50. It is characterised by the formation of drusen deposits between the retinal pigment epithelium and choroidal vasculture of the eye (Bird, 1992; Bird *et al*., 1995). Drusen deposits are composed of aggregated proteins and lipids and are similar to amyloid deposits (Anderson *et al*., 2001). The Y402H polymorphism of SCR-7 is strongly correlated to AMD. This is in close proximity to the heparin binding site. Mutations and polymorphisms within the C-terminal SCR-16/20 domain are commonly linked to aHUS, with CFH mutations identified in approximately 40 % of patients with familial aHUS (Noris *et al.*, 2010; Maga *et al*., 2010; Fremaux-Bacchi *et al.*, 2013). aHUS is characterised by autoimmune damage of the epithelial lining of the kidney wall due to improper complement regulation. This can result in kidney failure and is often fatal (Kavanagh *et al.*, 2006). aHUS primarily affects young children. aHUS-associated mutations are mainly located in the SCR-19/20 domains which in addition to C3d and surface binding localise CFH to kidney cells (Clarke *et al.*, 2013). Mutations are also observed in SCR-16/18 which may impair CFH ability to self-associate (Chapter Four).

Chapter Three

Methods for the study of protein structure and protein interactions

3.1 Introduction

Protein structure and function are closely related. Hence in order to understand how a protein functions it is highly advantageous to know as much as possible about its structure. A wide range of techniques are available to study protein structure - each having advantages and limitations. Many techniques are complementary to each other and often a multidisciplinary approach is highly desirable. X-ray crystallography is a common method of structure determination with approximately 93,000 structures solved by X-ray crystallography currently deposited in the protein data bank (PDB). An X-ray beam is diffracted by a protein crystal producing a diffraction pattern consisting of reflections having certain intensities and positions, but where the phase information for each reflection is not determined experimentally. Various methods exist for phase determination including heavy atom phasing, anomalous scattering approaches, and molecular replacement. Once the phases are determined, an electron density map (essentially the Fourier transform of the diffraction amplitude/phase data) can be calculated and interpreted. X-ray crystallography provides high resolution information from which it is usually possible (with the aim of reasonable sequence information and stereochemical restraints) to determine a reliable molecular structure, sometimes at atomic resolution. X-ray crystallography is limited by its need for protein crystals. Large proteins or proteins with regions of high flexibility or inhomogeneity such as glycosylation or large protein complexes often do not form crystals. Crystallisation is often achieved using harsh buffer conditions which can alter the proteins native state such that the obtained protein crystals may not be a true representation of the protein in its native physiological state (Rhodes, 2006). The use of solution techniques in parallel with crystallography can overcome this issue and it is becoming increasingly popular to use multi-disciplinary approaches. The emergence of centres that harness the range of techniques necessary (eg the Partnership for Structural Biology (PSB) in Grenoble (www.psb-grenoble.eu), the Research Complex at Harwell (RCaH) at RAL (www.rc-harwell.ac.uk), The Francis Crick Institute, UK (www.crick.ac.uk)) reflects this change of emphasis in recent years.

Nuclear magnetic resonance (NMR) is another technique which can yield atomic resolution information. It is based on the phenomena of magnetic resonance of the atomic nuclei with an externally applied magnetic field. Not all nuclei display magnetic resonance with ^{1}H, ^{13}C and ^{15}N being the most biologically relevant isotopes

for NMR. Protein NMR is carried out in solution bypassing the need for crystals and allowing the protein to adopt its preferred conformation. However it is limited to smaller proteins as the complexity of the NMR spectra increases with the number of nuclei. Molecular weights above 50 kDa become difficult to analyse. Isotopic labelling is often required with high concentrations of protein needed (Lian & Roberts, 2011).

Electron microscopy (EM) can be used to produce low resolution structures with the use of an electron beam to probe the structure. Proteins studied by negative stain EM generally need to be larger than 100 kDa, and fixation and dehydration of the sample using heavy metal salts are required to produce sufficient contrast for the interacting electron beam. This results in the molecule losing its hydration shell and being exposed to harsh chemicals which can often perturb the native structure. Cryo-EM can be used to obtain medium resolution structures where the protein is studied at cryogenic temperatures therefore bypassing the need for stains and fixatives (Serdyuk *et al.,* 2007). Instrument developments have led to an increase in resolution with recently published structures being solved to 2.8 Å resolution (Campbell *et al.*, 2015). Cryo-EM is limited to large proteins and complexes which are in general above 300 kDa.

Another technique that can be used for structure determination at low resolution is small angle scattering (SAS) which can be used to obtain structural information for large proteins and complexes, flexible proteins and glycosylated proteins in the solution state in near native conditions. SAS techniques using X-rays and neutrons were utilised in this thesis to study the structure of protein complexes and so this technique is discussed in greater detail in this chapter.

Protein-protein interactions are involved in many biological processes. The interaction of proteins can result in changes of protein function for example, through conformational changes or the assembly of proteins into an active complex. In the complement system, the interactions that complement Factor H undergoes with proteins on host cell surfaces, provides protection to these surfaces. Therefore understanding how proteins interact to form complexes is crucial for an understanding of functional relevance. Protein interactions can be stable or transient depending on their biological context. Protein-protein interactions are in general non-covalent and can be driven by hydrogen bonding, hydrophobic forces, electrostatic interactions or Van der Waals forces (Fersht, 1999). The strength of the interaction is commonly

reported as the equilibrium dissociation constant (K_D) which is reported in molar units. The higher the K_D *i.e.* in the μM range, the weaker the interaction.

There are many methods available to study protein interactions and their binding properties. Methods to identify binding partners include yeast two hybrid screens, pull down assays and co-immunoprecipitation. Methods to characterise the kinetics, affinities and binding stoichiometry's of the identified interactions include surface based methods such as surface plasmon resonance (SPR) and dual polarisation interferometry (DPI) where one of the binding partners is immobilised to a sensor surface. Surface free methods are also available including analytical ultracentrifugation (AUC), isothermal titration calorimetry (ITC) and microscale thermophoresis (MST). In this thesis AUC, SPR and MST were employed to analyse protein-protein interactions.

3.2 Analytical ultracentrifugation

AUC is a technique which separates particles in solution under high centrifugal forces. The rate of sedimentation of a particle in suspension is dependent on its size and shape and is monitored using an optical system attached to the ultracentrifuge. A sedimentation coefficient and a diffusion coefficient are obtained which yield size and shape information about the system being studied. AUC is therefore useful for studying the sizes and shapes of biomolecules in solution and for examining biomolecular interactions, sample purity, apparent molecular weights, and the oligomerization states of molecules. Since AUC is carried out in solution, near physiological conditions can be used, and the effects of changing these conditions can be determined with respect to the shape of the biomolecule. AUC was largely developed by Svedberg in the early 20th century with his work on the characterization of colloids under centrifugal forces. In 1926 he won the Nobel prize in chemistry for his work. The centrifuge was further developed by Beams and Pickles in the 1930's who introduced a vacuum which allowed the centrifuge to reach speeds of 40,000 rpm. Modern day ultracentrifuges are developed and produced by Beckman Coulter with speeds of up to 150,000 rpm achievable. They have also developed various optical systems to monitor the sedimentation of biomolecules. Developments in software such as SEDFIT (Schuck, 2000) which allow the simple and rapid analysis of AUC data have also made the technique more accessible to the field of biophysics.

3.2.1 Theory of Sedimentation

Sedimentation is a process whereby a particle in suspension settles to the bottom of its container, over time, due to the gravitational force. This particle will be acted upon by three forces – sedimentation, buoyancy, and friction (Figure 3.1).

The sedimentation force, F_s, which is proportional to the mass of the particle and to the acceleration it is undergoing. F_s is;

$$F_s = m\omega^2 r = \frac{M}{N}\omega^2 r \qquad \text{(Eq.3.1)}$$

where m is the mass of a single particle, r is the radius, ω is the angular velocity, M is the molar weight of the particle, and N is Avogadro's number.

The buoyancy force, F_b, which is derived from Archimedes principle acts in the opposite direction to the sedimentation force and is given by;

$$F_b = -m_0\omega^2 r \qquad \text{(Eq. 3.2)}$$

where m_0 is the mass of the fluid displaced by the particle. m_0 is given by;

$$m_0 = mv\rho = \frac{M}{N}\upsilon\rho \qquad \text{(Eq. 3.3)}$$

where v is the partial specific volume and ρ the density of the solvent. The partial specific volume is the volume that the particle takes up in the solution. The density and viscosity of the solution used can be determined experimentally, while v for a protein is calculated from its sequence.

The final force which acts upon the sedimenting particle is the frictional force, F_f. When the particle begins to sediment, its velocity will increase as it moves away from the meniscus, due to the increasing acceleration. As it moves, the particle will undergo a frictional force which acts in the opposite direction to sedimentation.

F_f is given by;

$$F_f = -fu \qquad \text{(Eq.3.4)}$$

where f is the frictional coefficient. f is dependent on the size and shape of the molecule with smaller more compact or spherical particles experiencing less of a frictional drag than an elongated or bulky molecule. u is the constant velocity. These three forces will arrive at equilibrium described by the following equation;

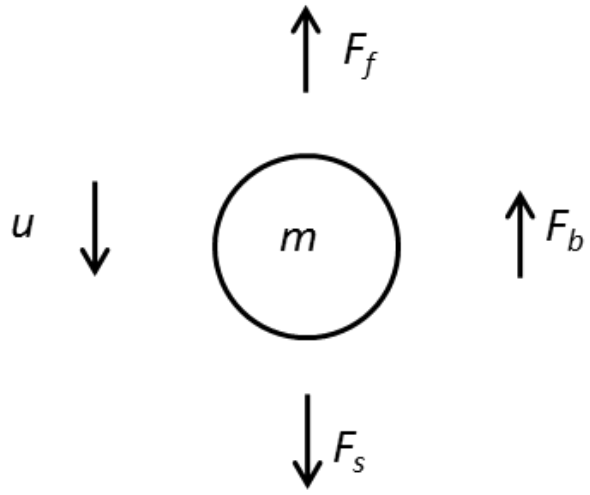


Figure 3.1 The forces acting on a sedimenting particle of mass, *m*. $F_{f,}$ is the frictional force and F_b is the buoyancy force both of which act against sedimentation. F_s is the sedimentation force with *u* the constant velocity. *Adapted from Ralston, 1993*.

$$\frac{M\,(1-\upsilon\rho)}{Nf} = \frac{u}{\omega^2 r} \equiv S \qquad \text{(Eq. 3.5)}$$

s is the sedimentation coefficient and is dependent on the properties of the particle including molecular weight, shape of the particle (frictional coefficient) and its partial specific volume. Particles with different molecular weights or different sizes and shapes will have different sedimentation coefficients. The units for the sedimentation coefficient are in seconds with the Svedberg unit defined as 10^{-13} seconds. The Svedberg unit is conventionally used to express the sedimentation coefficient and so if a coefficient of 3.0 S is derived this is 3 x 10^{-13} seconds. The Svedberg equation is derived from the rearrangement of Eq. 3.5;

$$\frac{M(1-\upsilon r)}{RT} = \frac{S}{D} \qquad \text{(Eq. 3.6)}$$

where R is the gas constant and T is the temperature in Kelvin and D is the diffusion coefficient. The sedimentation coefficient is generally expressed as the $s_{20,w}$ which is the S value converted to standard conditions in water at 20 ºC (Balbo & Schuck, 2005).

3.2.2. Instrumentation

The instrument used for AUC is similar to a preparative centrifuge with the addition of an optical system for detecting the sedimentation of the particles (Cole & Hansen, 1999). Rotors are required which are capable of withstanding high speeds, and special cells are used which allow the optical systems to monitor the particles sedimentation. The rotors are within an evacuated chamber to reduce the friction and the turbulence. Sector shaped cells are used which contain two sectors one for buffer and one for sample. Each sector is held between two thick quartz windows with a volume of 400 µl (Figure 3.2 B). The sectors allow sedimentation velocity experiments to be carried out as the sample will sediment along a radial line. Figure 3.2 shows the sedimentation profile obtained from a sedimentation velocity experiment. Both the sample and solvent or buffer meniscus will be visible in the profile corresponding to the meniscus at the top of the cell. The reaction boundary is indicated as the position at which the sample has reached equilibrium between sedimentation and diffusion.

Three types of optical systems are available for AUC experiments which are chosen depending on the experimental conditions and requirements. Absorbance optics monitor the absorption of the sample over a range of wavelengths between 190

and 800 nm. When the cell passes beneath it, a high intensity xenon lamp will flash; this lamp will scan along the sector of the cell, detecting the different positions of the sample (Figure 3.3). It is therefore possible to carry out AUC experiments over a range of wavelengths or to pick one wavelength and scan the cell. This is useful for studying biomolecules that have strong absorption properties, for example proteins can be monitored at 230 nm or 280 nm and DNA at 260 nm. The disadvantage of using absorption optics is that at high concentrations the signal can become saturated and other optical systems have to be used to obtain accurate data.

Interference optics are often used in AUC and can be collected simultaneously with absorbance data. In the Beckman Coulter range of AUC instruments, a Rayleigh interference system is used. Rayleigh interference is based on the changes of the refractive index within the system and is recorded using a Rayleigh laser interferometer. Two waves of monochromatic light pass through each of the sectors of the cell *i.e.* one through the solvent alone and one through the solvent and the solute. When the two waves emerge from each of the sectors, they undergo interference, with the wave from the sample shifted relative to the solvent wave. This change represents a change in the refractive index which is due to the difference in solute concentration between the sample and the solute and results in an interference pattern as the solute concentration changes over time *i.e.* as it sediments. Interference can be used for samples which do not have chromophores or for samples which have high concentrations which saturate the absorbance detection system.

A third optical system which can be used is the fluorescence detection system (FDS). This requires additional steps in sample preparation such as the addition of fluorescent labels on proteins or biomolecules to allow for their specific detection. This method is particularly useful when studying complex systems where it is difficult to identify the different peaks present in the size distribution *c(s)*. For example when studying protein-protein interactions it may be such that one of the interacting partners oligomerises with a similar sedimentation coefficient and size as the expected complex. By labelling the protein which does not oligomerise it will be possible to detect its presence in the third peak in a highly sensitive manner. The FDS was designed by the company Aviv Biomedical and is compatible with the existing Beckman AUC models. A laser light source is used which provides excitation at 488 nm. The emitted light will then pass back through a dichroic filter capturing the

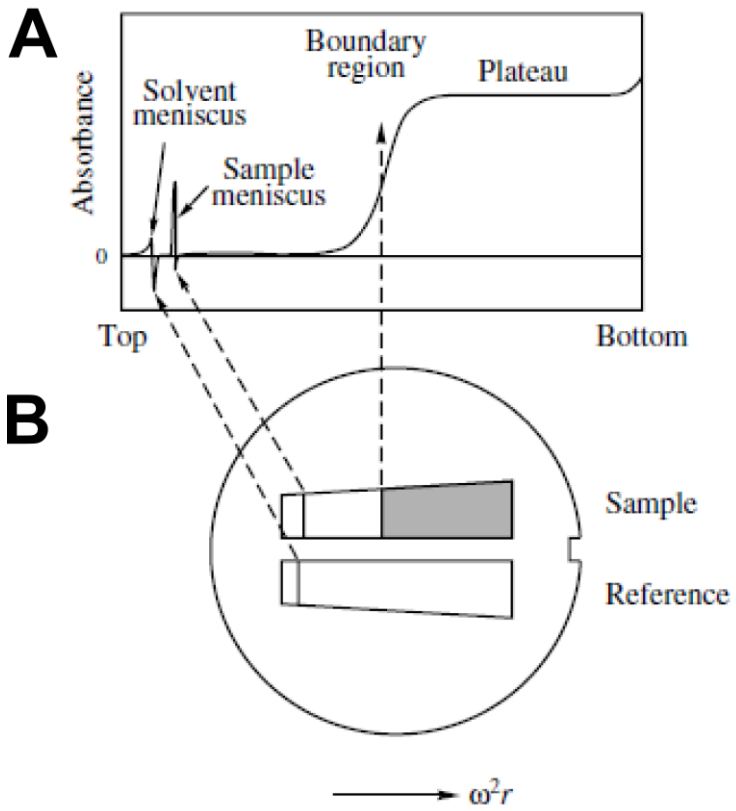


Figure 3.2 A sedimentation velocity experiment. *A*, The sedimentation profile obtained from a sedimentation velocity experiment with the sample and solvent meniscus present and the boundary region of the sample. *B*, The two sector AUC cell is schematically represented with one sector containing sample and the second containing the reference or the buffer. The sample will sediment towards the right of the cell (as shown). (Image adapted from Ralston, 1993).

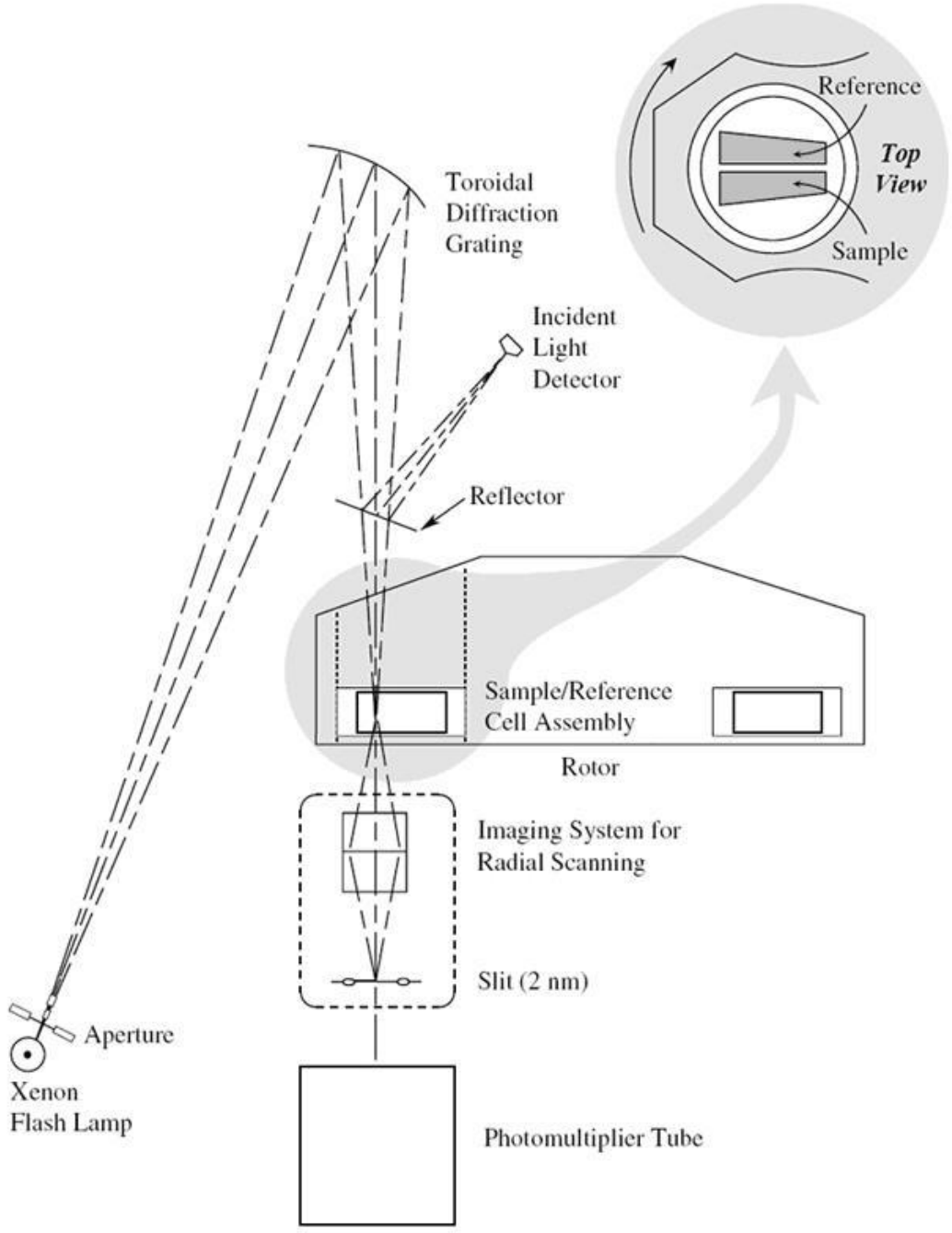


Figure 3.3 Schematic representation of the AUC optical system of the Beckman Optima™ XL-A centrifuge (Adapted from Ralston, 1993).

emitted light. Photo bleaching and quenching may occur when using fluorescence optics both of which will result in a decrease of detection.

3.2.3. Types of AUC experiments

Two types of AUC experiments can be carried out; sedimentation velocity and sedimentation equilibrium. Sedimentation equilibrium (SE) is a thermodynamic method that is carried out at low speeds and will provide information on the equilibrium formed when the centrifugal and diffusional forces are balanced. This method provides information on the molecular mass of the species present and can also provide information on the association of interacting systems. This method was not used in this thesis and is not discussed in detail.

Sedimentation velocity (SV) experiments are carried out at high speeds of up to 60,000 rpm depending on the system being studied. SV experiments provide hydrodynamic information about the size and shape of the molecules present in the sample. At the beginning of the experiment the particle will be suspended at a uniform concentration throughout the solvent within the cell. By spinning the sample at high centrifugal forces, the sedimenting particle will move from the meniscus toward the bottom of the cell. The concentration therefore decreases at the meniscus and increases at the bottom of the cell (Figure 3.2 A). Over time, a boundary will form between the depleted region and the concentrated region. This boundary will move as the concentration changes due to diffusion across the cell and radial dilution. By monitoring the rate and shape change of the boundary over time it is then possible to obtain the sedimentation coefficient and the diffusion coefficient. Different species of different size or shape will have different reaction boundaries in an SV experiment (Lebowitz *et al.,* 2002; Cole *et al.,* 2008).

3.2.4. Analytical ultracentrifuge data analysis

In general, SV experiments using interference, absorbance or fluorescence are analysed in a similar way. There are a number of approaches and software available to analyse AUC data with DCDT+ and SEDFIT (refs) being two of the most popular. With DCDT+, the dc/dt approach is used (Stafford, 2000; Philo, 2006;). Analysis is carried out by subtracting pairs of SV scans yielding the various rates at which the different species present are sedimenting. This information is then transformed into apparent sedimentation coefficient distribution *g(s*).* Individual peaks in the

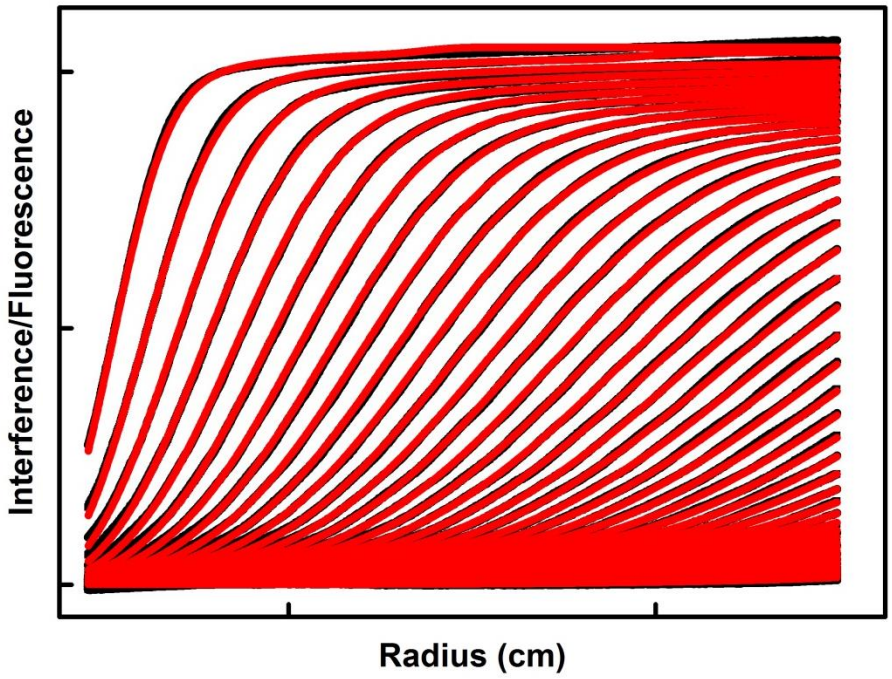


Figure 3.4 The movement of the boundary down the cell in a sedimentation velocity experiment. Black lines represent the experimental data while the red represent the goodness of fit.

distribution correspond to individual species and from this the sedimentation coefficient and the diffusion coefficient are obtained.

The program used in this thesis was SEDFIT, which in comparison to DCDT+, allows the sedimentation coefficient to be calculated directly from the data. This is described as the size distribution *c(s)* analyses and is based on the Lamm equation (Schuck, 2000; Dam & Schuck, 2005). The sedimentation boundaries of species which have different sedimentation coefficients are overlapped (Figure 3.4). Each of sedimentation boundaries are fit to the Lamm equation. The Lamm equation relates the law of diffusion with the centrifugal migration (Lamm, 1929). The SEDFIT *c(s)* method takes the diffusion into account by assuming that the frictional ratio f/f_0 is the same for each species present and that the shape is constant (Schuck, 2000). f/f_0 is a measure of the elongation of the molecule with f as the observed frictional coefficient and f_0 the frictional coefficient of a sphere that has the same volume of the macromolecule in the hydrated state. In SEDFIT the Lamm equation is fit to the experimental data. The *c(s)* distribution has peaks which represent the individual sedimenting species in the sample. The *c(s)* data can be converted to the *c(M)* to calculate the molecular mass of the sedimenting particle. AUC data analysis becomes more sensitive when looking at interacting species in particular rapidly interacting species where reaction boundaries will be present (Dam & Schuck, 2005; Brown & Schuck, 2006). A reaction boundary is a broad peak which corresponds to the free and complexed species. They will occur at well-defined sedimentation coefficients of the free and complexed species.

3.3. Surface Plasmon Resonance

SPR is a biophysical technique used to characterise biomolecular interactions. SPR measures the changes in the refractive index on the surface of a thin conducting film. One molecule (ligand) is immobilised on the film sensor surface while the other is in the solution phase (analyte). Changes in the refractive index provide information about the interactions of molecules on the film surface in real time. SPR can be used to test the binding between two molecules, to determine the kinetics and affinity of an interaction which can ultimately lead to the calculation of the K_D and also to calculate the concentration of the molecules present in the sample. Using SPR as a tool to study the interaction of biomolecules was largely developed in the early 1990's by Pharmacia Biosensor AB who produced the first commercial SPR machines (Jonsson

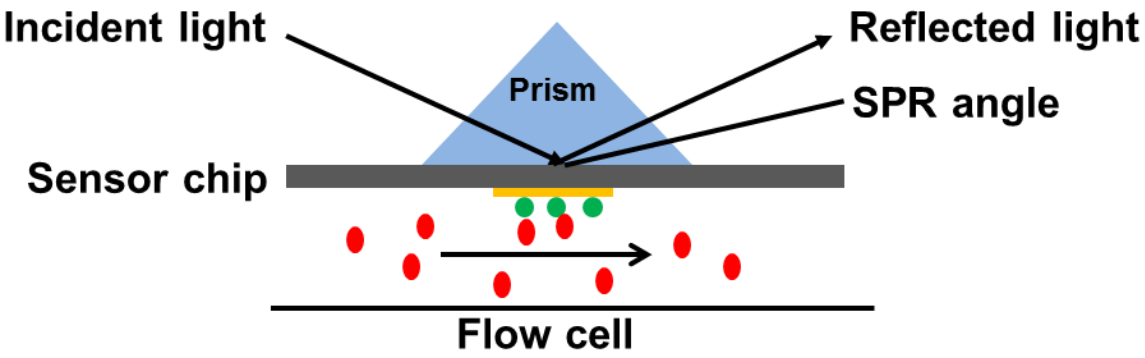


Figure 3.5 A schematic representation of the setup of an SPR experiment. The sensor chip with the gold metal layer (yellow) with immobilised molecules (green). Binding of molecules from the flow sample (green) result in changes of the refractive index causing the incident light reflectivity to decrease with a defined angle called the SPR angle.

et al., 1993). There are now many SPR machines on the market with the most common being the BIAcore® range produced by GE healthcare who also produce numerous sensor chips and kits required to carry out SPR experiments.

3.3.1 Theory of surface plasmon resonance

When polarized light passes through a medium with a high refractive index to a media with a lower refractive index, the light will be partly reflected and partly refracted. If the incident light is at a certain defined angle, light will not be refracted and only internal reflections are observed. When total internal reflection occurs an evanescent wave field (plasmon) is generated across the interface. The plasmon wave resonation results in the absorption of light causing the intensity of the reflected light to decrease at a defined angle called the SPR angle. The SPR angle is highly sensitive to the refractive index and so changes in refractive index at the sensor surface will lead to a shift in the SPR angle. The SPR angle is dependent on the wavelength and the incident angle of the light and on the refractive index at the surface (Figure 3.5). The wavelength and the incident light are fixed and so changes in SPR angle will result only from changes in the refractive index.

During an SPR experiment, the instrument constantly measures the SPR angle of the monochromatic incident light. The interface is formed between a gold coated immobilising sensor chip and the sample/buffer solution (Figure 3.5). Changes of the refractive index upon analyte binding will result in a change in the SPR angle. The change in SPR angle is given as response units (RU) with 1 RU corresponding to a shift of 0.0001º. The change in SPR angle is given by the equation;

$$\Delta\theta\,(\lambda) = C_1\,\Delta n + C_2\Delta d, \qquad \text{(Eq. 3.7)}$$

where Δn is the change in the refractive index and Δd is the change in thickness. When Δd changes due to binding at the surface a refractive index is predicted from the Lorentz-Lorenz equation;

$$\Delta n = -\frac{1}{6n}(n^2+2)^2\left(\frac{n^{2}-1}{n^{2}+1} - \frac{n_{w-1}^2}{n_{w-2}^2}\,\frac{V_p}{V}\right)\frac{\Delta d}{d}\ , \qquad \text{(Eq. 3.8)}$$

where n is the refractive index of the molecule, n_w is the refractive index of water, V_p is the volume of the molecule and V is the volume of the layer (Boussaad *et al.*, 2000).

3.3.2 Sensor surfaces and immobilisation

During an SPR experiment one molecule (ligand) is immobilised on the chip surface and the second molecule (analyte) is injected into the flow cell. The sensor chip is held in the flow cell of the instrument with buffer constantly flowing through. If they analyte interacts with the ligand this will result in an increase in the refractive index. The changes in refractive index upon association and dissociation of the analyte are measured in real time and are shown as a sensorgram which is the RU against time (s) (Figure 3.6). The sensor chip consists of a thin layer of gold (~ 50 nm) coated onto a glass slide. In general a carboxymethylated dextran matrix is covalently attached to the gold layer. The ligand is then immobilised to this either by covalent attachment or through the use of a capturing molecule. The immobilisation must be sufficiently strong so that the bound ligand is not removed under the experimental conditions.

One of the most commonly used sensor chips for covalent attachment is the CM series of chips. The length of the dextran polymer varies depending on which CM chip is used with the length required being dependent on the system being studied. CM chips are popular as the dextran provides a hydrophilic environment in which many biomolecules are stable. Immobilisation using the CM chip series is through attachment of the molecule to the dextran for example through its amine or thiol groups. Amine immobilisation is generally carried out at low pH and salt conditions which may not be suitable for some molecules. Covalent attachment through amine or thiol groups will result in random orientations of the immobilised ligand potentially shielding or exposing binding sites additionally the amine or thiol groups which are bound may be part of the binding site thereby changing the nature of binding. Many of these issues can be overcome using affinity capture methods in which the specific binding nature of the ligand is exploited.

A common method is to capture histidine tagged recombinant proteins using a sensor chip which has nitrilotriacetic acid (NTA) bound to its dextran matrix. For recombinant proteins with other tags e.g. glutathione-s-transferase (GST) or c-myc tags, specific antibodies to these tags can be immobilised onto to the dextran matrix and used to capture the tagged molecule.

Another commonly used affinity capture sensor chip is the streptavidin (SA) chip which has streptavidin covalently bound to the dextran matrix (Figure 3.7). Streptavidin binds biotinylated molecules with a high affinity. An advantage to using SA chips is that immobilisation can be carried out in the buffer of choice such that

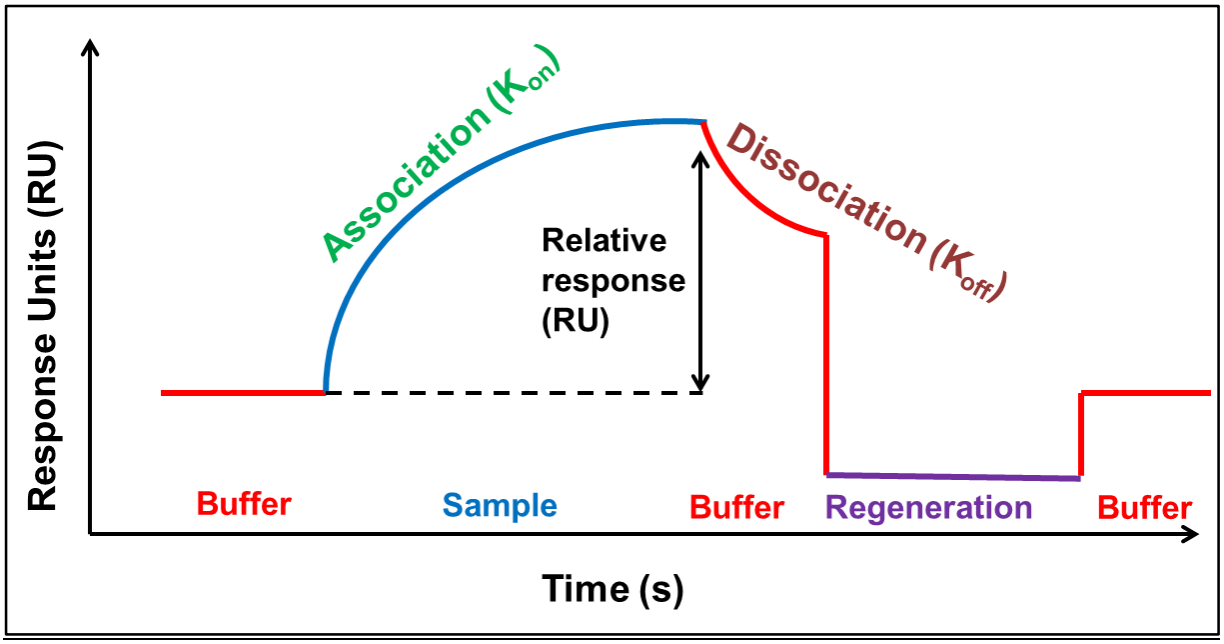


Figure 3.6 The SPR sensorgram of response units versus time. The solution passing over the sensor surface is indicated at each point. The association (K_{on}) and dissociation (K_{off}) are indicated and can be used to calculate the equilibrium dissociation constant (K_D) of the interaction.

optimal conditions for the ligand stability can be retained. However it does require the ligand to be labelled with biotin prior to immobilisation. Biotin labelling is carried out in the buffer of choice and reacts with the primary amine groups of lysine residues therefore possessing the same disadvantages to CM sensor chip immobilisation. To minimise these effects labelling is optimised to produce one molecule of ligand with one molecule of biotin bound which is sufficient to immobilise the ligand to the surface.

The density of ligand to be bound to the sensor chip is dictated by the type of experiment to be carried out. For affinity and kinetic measurements low densities are required in order to avoid complications from mass transfer or steric hindrance. Low densities will also allow the ligand to become saturated with analyte on a realistic time frame. In general a maximum response of 200 response units (RU) is desired for affinity or kinetic measurements. This is calculated by the following equation;

$$R_{ligand} = \frac{Rmax.Mrligand}{Mranalyte.Valency\ ligand} \qquad \text{(Eq. 3.9)}$$

where R is the number of response units and Mr is the molecular mass. If unknown the valency is assumed to be 1. Due to the high affinity of the immobilisation strategies, low starting protein concentrations (< 50 µg/ml) are injected onto the chip to be immobilised. To achieve the calculated maximum response (R_{max}) the contact time of the ligand with the surface can be increased along with the injection volume and if necessary the ligand concentration can also be increased. When the R_{max} has been achieved an SPR binding experiment can be carried out.

The sensor surface is firstly equilibrated with the experimental flow buffer. The analyte can then be injected to test for binding or a series of different concentrations can be injected for affinity or kinetic measurements. After each binding event the surface of the chip needs to be regenerated *i.e.* the analyte needs to be removed while the ligand stays immobilised to the surface in its active state. Regeneration largely depends on the type of interaction between the ligand and analyte. For example if the interaction is electrostatic, a buffer containing a high concentration of NaCl will be sufficient to remove the bound analyte but not the immobilised ligand (BIAcore ® Sensor Surface Handbook).

3.3.3 Surface plasmon resonance data analyses

Three types of experimental analyses can be carried out. Binding analyses, concentration analyses and equilibrium analyses. Binding analyses are used to determine the binding specificity between two partners. The analysis is based on report point values rather than on the association or dissociation of the binding. It is often used as a first step in identifying binding partners and is employed in drug screening. Concentration analyses is whereby the concentration of the analyte can be calculated from a calibration curve of known standards. This method can be used to calculate the specific protein concentration within a sample which is crucial in drug development where this knowledge is needed to optimise drug design.

Equilibrium analyses are employed in this thesis and so are discussed in detail. Successive injections of the analyte at varying concentrations is required for equilibrium analyses. Regeneration is carried out after each injection. Information on both the kinetics and the affinity of the interaction can be obtained from equilibrium analyses. The K_D can be calculated either from the kinetic rate constants or from the steady state affinity. Kinetic measurements involve measuring the kinetic rates of association ($M^{-1}S^{-1}$) (K_a) and dissociation (S^{-1}) (K_d) with the $K_D = K_d/K_a$. The sensorgram can be fit to a number of pre-defined mathematical models to yield these values. The simplest model is that for simple 1:1 binding;

$$A+B \leftrightarrow AB, \qquad \text{(Eq.3.10)}$$

A will reversibly bind to B with a 1:1 ratio with the K_D defined as;

$$\frac{[A]x\ [B]}{[AB]} = K_D \qquad \text{(Eq. 3.11)}$$

Where [A] [B] and [AB] are the concentrations of each species.

$$\frac{d[AB]}{dt} = K_a.\ [A]\ .\ [B], \qquad \text{(Eq. 3.12)}$$

describes the association of the ligand and analyste while,

$$\frac{-d[AB]}{dt} = K_d.\ [AB] \qquad \text{(Eq. 3.13)}$$

describes the dissociation of the complex [AB]. Combined, these two equations describe the formation of the complex over time. From the sensorgram, the values of

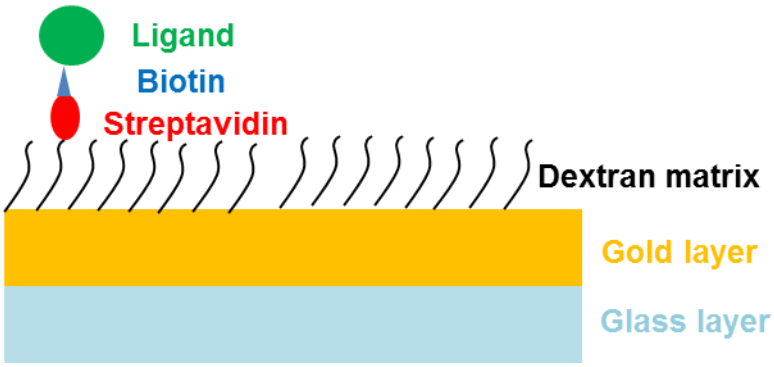


Figure 3.7. A cartoon representation of immobilisation to a streptavidin sensor chip. Capture method of immobilisation of a ligand labelled with biotin to a streptavidin coated sensor surface.

the K_a and K_d are calculated making it possible to calculate the K_D (Figure3.6). If the binding has a fast on and off rate *i.e*. little curvature within the sensorgram then it will not be possible to carry out kinetic analyses and so steady state affinity analyse will be required in order to calculate the K_D.

Affinity measurements involve measuring the affinity constants either from the steady-state binding which assumes a 1:1 binding model or from the calculated kinetic constants. The steady state is defined as the response at which the net association and dissociation rates are zero. By plotting the steady state response against the analyte concentration the equilibrium K_D is calculated. The fitting of the data is assessed by the goodness of fit to the model by χ^2. K_D calculations from affinity measurements require that the highest concentrations measured be above the K_D. This method also requires that the steady state is reached at each concentration. It is usually not possible to carry out both kinetic and affinity measurements on the same experiment as the contact times for kinetics experiments generally do not allow the binding to reach the steady state (BIAcore® Assay Handbook).

3.4. Microscale Thermophoresis

3.4.1 Theory of microscale thermophoresis

MST is a surface free technique which is used to measure the affinity of biomolecular interactions in solution. It is based on the phenomena of thermophoresis. Thermophoresis was first described by Carl Ludwig in 1856 and is known as the Ludwig-Soret effect. Thermophoresis is the movement of molecules with respect to a temperature gradient. The theoretical explanation for thermophoresis has been the subject of some debate and is still not fully understood. It is generally accepted that the thermophoretic effect observed is related to the solvation entropy and the hydration shell of the molecules (Duhr & Braun, 2006). The velocity (v) of the molecule is linearly dependent on the temperature gradient (gradT). This is proportional to the thermal diffusion coefficient, $D_{Ti,}$ described by;

$$v = - D_{Ti} gradT \qquad \text{(Eq. 3.14)}$$

A temperature difference within a region of molecules will result in the movement of those molecules away from the region of elevated temperature. This movement is quantified by the Soret coefficient S_T;

$$S_T = D_{Ti} / D_i, \qquad \text{(Eq. 3.15)}$$

where D_i is the diffusion coefficient. The S_T is dependent on the size and charge of the molecule and also on the solvent entropy. The phenomena of thermophoresis is sensitive to the changes that occur when two biomolecules interact, thus enabling thermophoresis to be employed as a tool for probing biomolecular interactions (Jerabek-Willemsen *et al.,* 2011; 2014).

MST is a relatively new technique which has been developed and patented by the company NanoTemper Technologies in Munich, Germany. The first instruments became commercially available in 2010. Since then MST has grown in popularity among the biophysical community with numerous publications in the past few years; (Alexander *et al.,* 2013; Ascher *et al.,* 2014; Korkhov *et al.,* 2014).

3.4.2 A microscale thermophoresis experiment

During an MST experiment the fluorescence of the molecules is measured in order to detect the molecules' thermophoresis. There are two methods of detection; a label-free method which relies on the intrinsic florescence of the protein and a method which involves labelling the protein with a fluorescent molecule. The label-free method requires the protein to contain aromatic residues while the labelling method requires residues available for covalent attachment of the fluorophore. In this thesis work, the labelling method was used. This involved the covalent attachment of a fluorescent molecule to the lysine residues of one of the binding partners. Labelling kits are available from NanoTemper but commonly used fluorescent dyes can also be used, provided the MST machine has the correct wavelength for excitation. The labelled protein is kept at a constant concentration while the unlabelled binding partner is titrated with the highest concentration 10-fold higher than the expected K_D. MST experiments are carried out in glass capillaries with a volume of approximately 10 µl. Standard glass capillaries, hydrophobic and hydrophilic coated capillaries are available. It is essential to determine which capillary is best suited to the system being studied as it is possible for some proteins to 'stick' or absorb to the capillary wall which will affect the MST signal.

The MST instrument has a capacity for 16 capillaries which are laid out vertically. It contains a fluorescence system (LED system) which is focused onto the capillary using a dichroic mirror (Figure 3.8 A). The fluorescence of the labelled protein alone is measured to ensure that the fluorescence count is sufficient for the experiment and to ensure that the correct capillaries have been used. The fluorescence

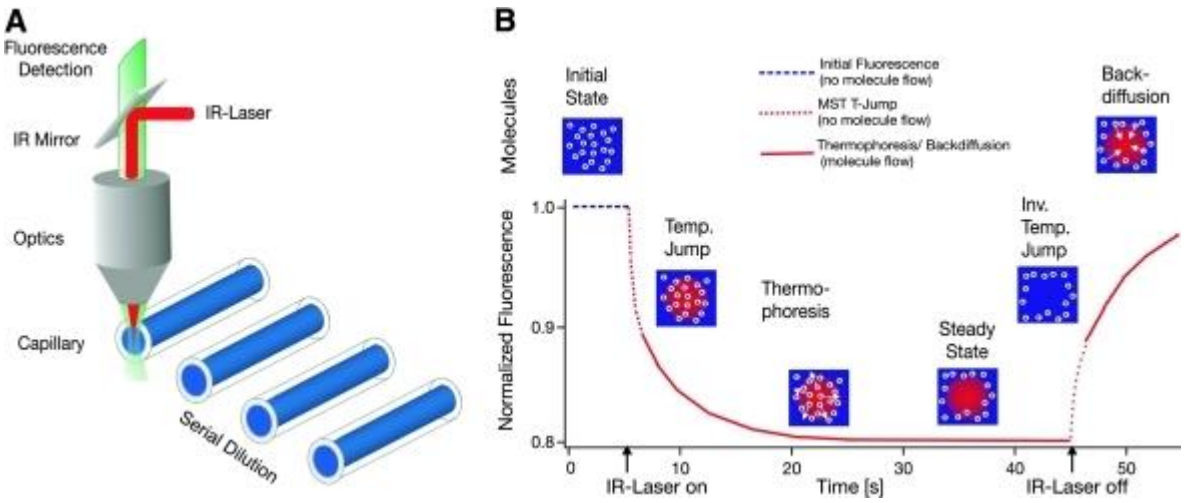


Figure 3.8 A microscale thermophoresis experiment. *A*, The general experimental set up of an MST machine highlighting the focusing of the fluorescence detection system and the Infrared (IR) laser onto the glass capillaries using a dichroic mirror. *B*, A typical MST signal showing the different stages of an MST experiment as observed from the fluorescence time trace (Image adapted from Monolith NT.115 user manual, 2011).

count should be between 80 and 1500 counts with labelled protein concentration chosen depending on this. The LED power can also be optimised to ensure the fluorescence is at the correct intensity. The fluorescent excitation wavelength must be selected according to the label used. The LED power must also be selected, the default value of 50% is generally sufficient but depending on the fluorescence level of the labelled molecule this can be varied to achieve either a lower fluorescence count (below the required 1500) or a higher one (above 80). The microscale temperature gradient is produced by an infrared (IR) laser which is focused onto the capillary via the same dichroic mirror as the fluorescence system. The laser power *i.e.* the heating power can be varied with 20%, 40% and 80% as the default. Measurements can be carried out at all three or the optimum can be selected. The default start temperature is 24 ºC and can be changed depending on the requirements of the system being studied. Initial fluorescence is measured (F_{cold}) before the temperature gradient is applied. Switching on the IR laser results in the temperature change which is referred to as the T-jump and is defined as the change in fluorescence upon temperature *i.e.* before thermophoresis occurs. Following this the fluorescence change due to thermophoresis, which will reach a steady state, F_{hot} is measured. The IR laser is then switched off which results in an inverse T-jump and back diffusion of the molecules (Figure 3.8 B). The thermophoresis of the fluorescent molecule will change upon binding. The ratio of F_{hot}/F_{cold} is described as the normalised fluorescence (F_{norm}). This ratio changes if binding occurs and so this is used to determine the extent of binding (Jerabek-Willemsen *et al*., 2011; Monolith NT.115 user manual, 2011; Jerabek-Willemsen *et al.*, 2014) .

MST data is analysed using the NTAnalysis software from NanoTemper. F_{norm} is plotted against concentration which, if binding has occurred will result in a binding curve. The K_D of the interaction is then calculated from this curve. Similar to SPR, saturation is required to obtain an accurate K_D and so it is often necessary to carry out test experiments to get an estimate of the K_D in order to optimise sample concentration.

3.5 Solution Scattering

3.5.1 Introduction to solution scattering

Solution scattering techniques provide low resolution structural information about biomolecules in solution. Light scattering and small angle X-ray and neutron scattering are the main types of solution scattering techniques used for protein

structure determination. Light scattering experiments give information about the size of molecules suspended in solution. Static or dynamic light scattering experiments can be carried out with different size information obtained depending on the method used. Small angle scattering (SAS) with X-rays and neutrons is a more sensitive technique that allows information to be obtained about the overall shape and size of molecules in solution (Glatter & Kratky, 1982: Perkins, 1988). Using the information obtained from SAS it is possible to generate low resolution (2-4 nm) models of the overall protein structure.

3.5.2 Small angle scattering

SAS experiments can be carried out using X-rays (SAXS) or neutrons (SANS). Similar size and shape information is obtained from X-rays and neutrons which allows low resolution models or structures to be computed. One of the main advantages to using these techniques is that they are carried out in solution. Experiments in solution do not suffer from the packing effects that occur in crystals, and allow flexibility and variability within the structure. Additionally, most buffer conditions are suitable for solution scattering techniques. Buffer can be used which mimics physiological conditions in terms of pH and ionic strength, and the effect of changing these conditions on the protein structure can therefore also be examined. SAS can provide structural information where high resolution techniques are not possible. The major limitation of SAS is that it is only possible to obtain low resolution models that lack atomic detail. High resolution techniques such as crystallography and NMR are complementary to SAS and where high resolution structures are available (either complete or partial) these can be used to refine the scattering models (Perkins *et al.,* 2008; Blanchet & Svergun, 2013).

3.5.2.1 Theory of small angle scattering

While X-rays and neutrons have different properties and require different sources and instrumentation, the theory behind a small angle scattering experiment remains similar regardless of the radiation used. An incoming monochromatic beam will interact with and be scattered by the molecules suspended in solution. A detector is placed at a certain distance from the sample to record the scattering intensities at small angles (Figure 3.9 A). The scattered intensity $I(Q)$ is measured as a function of the scattering angle Q. This creates a scattering vector;

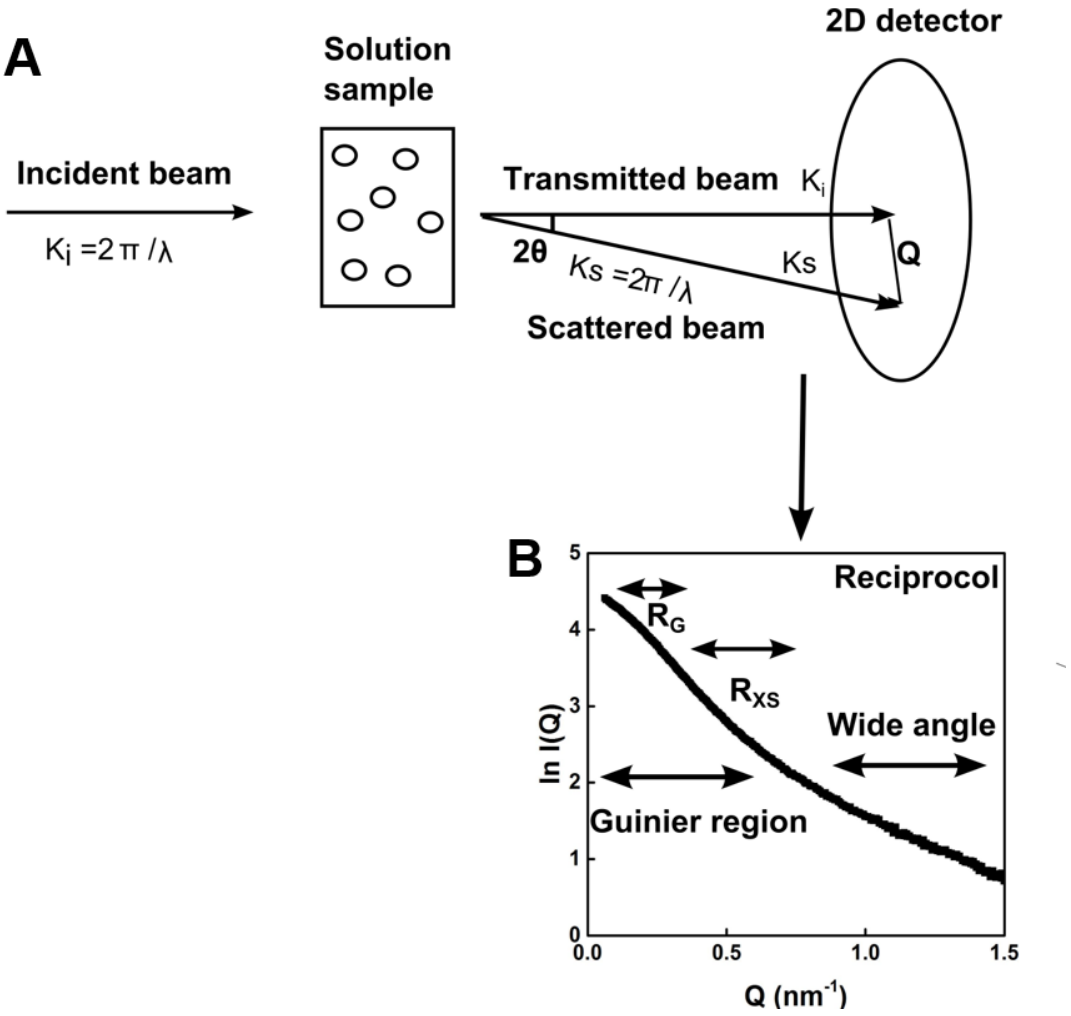


Figure 3.9. A schematic representation of a small angle scattering experiment. *A*, An incoming incident X-ray or neutron beam will be scattered by the particles which are suspended in solution in random orientations. The 2D detector at a defined distance will record the angle at which the beam is scattered (*Q*) and the intensity (*I*). *B*, The resultant scattering curve *I(Q)* which yields information about the size and shape of the particle. In the low *Q* region the Guinier analysis is carried out yielding the radius of gyration (R_G) and the radius of the cross section (R_{XS}).

$$Q = \frac{4\pi \sin\theta}{\lambda} \quad \text{(Eq. 3.17)}$$

where Q is the scattering angle and λ is the wavelength. θ represents the angle between the incident beam and the scattered beam (Figure 3.9 A). This equation is derived from Bragg's Law which describes diffraction as;

$$\lambda = 2d\ sin\theta \quad \text{(Eq.3.18)}$$

where 2θ is the scattering angle and d is the diffraction spacing. In crystallography the diffraction spacing represents the interplanar distances in crystal lattices, while for solution scattering it represents the larger dimensions of 1-100 nm. The Debye equation takes into account the differential orientations of the particles within the sample through rotational averaging in space:

$$I(Q) = \sum_{p}\sum_{q} f_p f_q \frac{\sin(rQ)}{rQ} \quad \text{(Eq. 3.19)}$$

where f_p and f_q are the scattering lengths of the electrons (or b for the case of the nuclei in a neutron experiment where b is the neutron scattering length) at points p and q within the biomolecule. p and q are separated by a distance, *r*. A 2-dimensional scattering curve *I(Q)* is obtained which contains information about the average size and shape of the particles in solution (Perkins *et al.,* 2008).

3.5.2.2 Small angle X-ray scattering

X-rays are electromagnetic waves which are located after ultraviolet radiation in the electromagnetic spectrum with only gamma rays having higher frequency /shorter wavelength. Table 3.1 highlights some of the main properties of X-rays. X-rays have a wavelength range of 0.1 nm to 10 nm, a frequency range from 3 x 10^{16} Hz to 3 x 10^{19} Hz and an energy range from 120 eV to 120 KeV. 'Hard X-rays' are used for structural biology with wavelengths of 0.01 nm to 0.1 nm with energies in the range 10 to 120 KeV. These high energies allow the X-ray's to penetrate deep into the sample with the wavelength comparable to the sizes and interatomic distances observed in biomolecules. Due to the high energy and ionising properties of X-rays, radiation damage often occurs which limits the exposure times and sometimes the samples being studied. X-rays are scattered by the electrons of the sample. The

scattering length will increase with atomic number as the number of electrons increase (Table 3.1 B). The scattering length, *f*, is defined as the amplitude of the wave observed per unit incident flux. The scattering cross section, σ, describes the region about the electron or nucleus i.e. the scattering point, that will be scattered by the incoming X-ray beam. For X-rays another factor must be taken into account, the form factor. X-rays will be scattered with the same amplitude by each of the electrons. However, due to the comparable size of the electron cloud the X-ray wavelength, the scattering amplitude of a given atom will decrease with scattering angle. The dependence of the amplitude on the scattering angle is the form factor of the atom.

3.5.2.3 X-ray instrumentation for SAXS

X-rays can be generated using a bench top rotating anode instrument whereby electrons are accelerated into a metal target anode creating X-rays. However, most commonly used for SAXS are X-rays which are produced at synchrotron sources due to the higher flux of X-rays obtained (~ 10^{14} photons/sec). The flux is defined as the number of X-rays (or neutrons) to pass a defined point. A synchrotron is a large circular particle accelerator where electrons are accelerated to high energies to produce X-rays. Figure 7.10 A, shows a schematic representation of the synchrotron at the European Synchrotron Radiation Facility (ESRF) in France and its associated beam lines. Electrons are firstly injected into a storage ring from a smaller booster synchrotron where energies of up to 6 GeV are obtained. Electrons are injected in bunch modes which are then accelerated around the storage ring using bending magnets to direct the beam and undulators (magnetic device) to refocus it. The storage ring then feeds each of the beamlines with X-rays (Figure 7.10A). There are numerous synchrotron facilities around the world including Diamond in the United Kingdom, Swiss Light Source in Switzerland, Spring 8 in Japan, and SLS in the USA. Each of these facilities have a variety of beamlines each of which are dedicated to a particular type of experiment (Perkins *et al*., 2008).

At the ESRF, BM29 is the dedicated BioSAXS instrument. BM29 has been optimized for biological macromolecules in solution (Pernot *et al.,* 2013). Figure 3.10 B shows a schematic representation of BM29. The X-ray beam is delivered from the storage ring via a bending magnet, the white or polychromatic X-ray beam is guided to the first monochromator for wavelength selection, which then passes through an energy calibrator. A second monochromatic slit allows the beam to pass through where it is

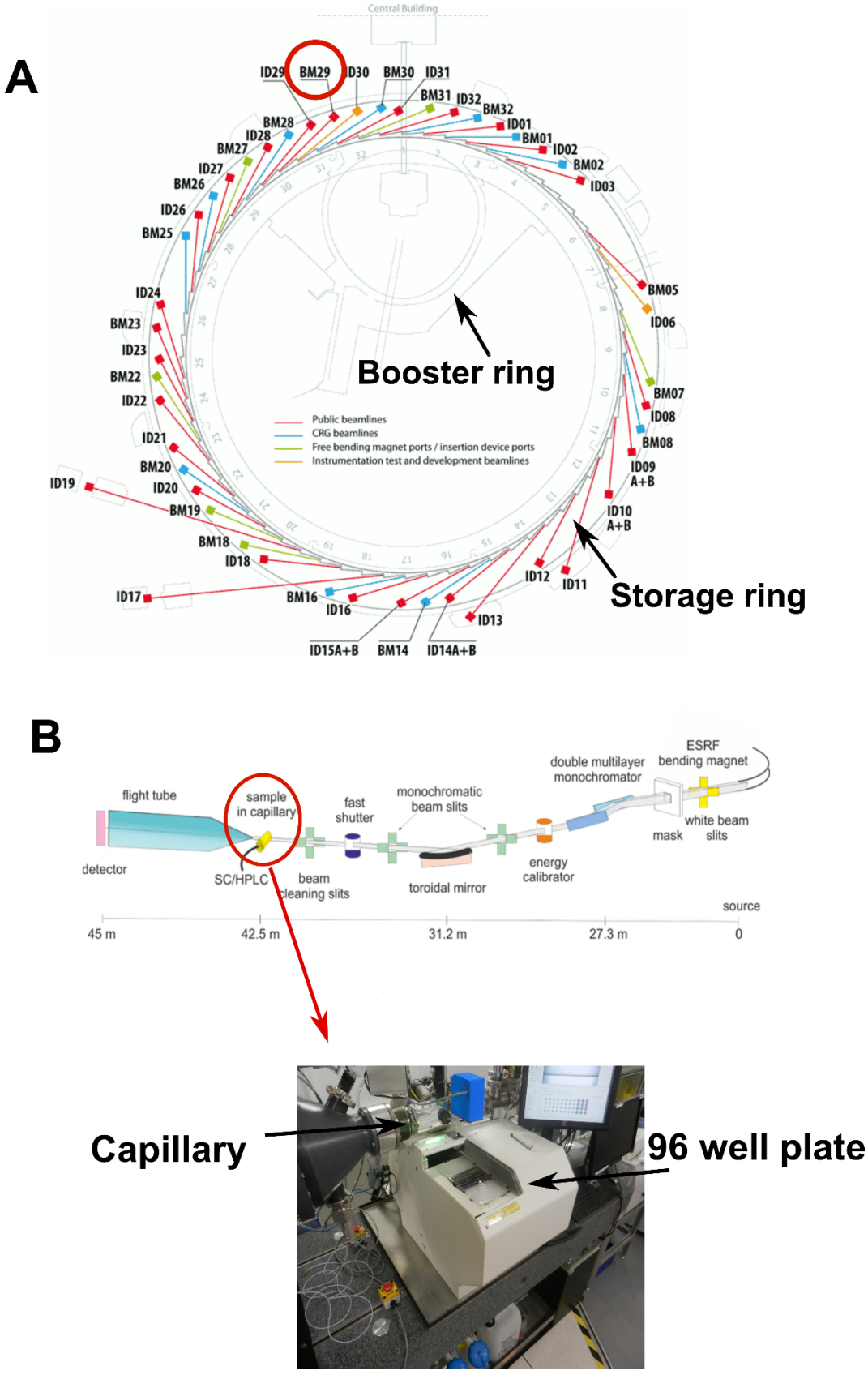


Figure 3.10 legend overleaf

Figure 3.10 Schematic representation of the ESRF and BM29. *A,* Set up of the synchrotron at the ESRF showing the numerous beam lines. The booster ring and storage ring are indicated. BM29 is circled in red (Image from www.esrf.eu). *B,* An illustration of the layout of BM29 from the incoming polychromatic beam (right) to the detection (left). An image of the automated sample changer on BM29 which loads the samples directly into the capillary (Pernot *et al*., 2013) (Image adapted from www.esrf.eu).

focused onto the plane of the detector by a toroidal mirror. It then passes through a final monochromatic slit resulting in a monochromatic and focused beam reaching the sample. A fast shutter is present which when open allows the beam to pass through the sample. The direct and scattered beam is then detected by a Pilatus detector in a vacuum tube. BM29 sample-to-detector distance is set at 2.86 M. A beam stop is required which prevents damage to the detector from the non-scattered direct beam. The Pilatus detector is a semiconductor hybrid pixel detector which uses a sensor and readout chips to detect the incoming X-rays and to convert them into charge (Pernot *et al*., 2013). On BM29 there are two methods of sample loading. A high-throughput automated capillary setup where samples are loaded into tubes or 96-well plates, with the sample automatically injected into the capillary with buffer collection before and after each sample or as required (Figure 3.10 B). The sample flows through the capillary with 10 frames collected on different sections of the sample minimizing sample exposure time and therefore radiation damage. BM29 also contains an online high performance liquid chromatography system which allows the elution volume from a column to pass directly through the beam on a short timescale (minutes). This system can serve as an online purification system (remove aggregation) or for studies with unstable biomolecular interactions where it may be possible to isolate a protein complex from a mixture. X-ray data is then collected continuously on the elute from the column while the UV trace is simultaneously measured. For data collected via the standard capillary method it is necessary to check for radiation damage. Each of the ten frames one-dimensional scattering curves *I(Q)* are compared to ensure no radiation damage has occurred, damaged frames are removed and the remainder are averaged. The corresponding buffer frames are averaged and the buffer is subtracted to yield the scattering curve.

3.5.2.4 Small angle neutron scattering

Neutrons are subatomic particles which have no net charge and which display wave-particle duality. Neutrons have a wavelength range of 0.04 nm to 3 nm, with corresponding energies ranging from 0.1 to 500 meV. Neutrons are classified as cold, thermal or hot according to their energies which are dependent on temperature. For solution scattering experiments, thermal neutrons are used which have wavelengths of between 0.1 nm to 0.3 nm and energies of 5 to 100 meV at temperatures of 60–1000 K. The wavelengths are similar to those used for X-ray experiments but neutrons are

(A)

X-rays	**Neutrons**
Scattered by electrons	Scattered by nuclei
high energies (120 eV -120 KeV)	low energies (5 -100 meV)
Data aquisition time ~10 seconds	Data aquisition time dependent on the neutron flux can be up to one hour at low flux source
Radiation damage occurs, samples not normally reusable	No radiation damage, samples can be reused
H_2O buffers used	D_2O buffers used which can cause protein aggregation
Contrast variation is generally not carried out	Contrast variation to match out individual components is possible

(B)

Element	**Isotope**	**Atomic number**	**X-ray scattering length (*f*) (fm)**	**Neutron scattering length (*b*) (fm)**
Hydrogen	^{1}H	1	2.81	-3.742
	^{2}H	1	2.81	6.671
Carbon	^{12}C	6	16.9	6.651
Nitrogen	^{14}N	7	19.7	9.4
Oxygen	^{16}O	8	22.5	5.8
Phosphorus	^{31}P	15	42.3	5.1

Table 3.1 Different properties of X-rays and neutrons. (*A*) A comparison of X-ray and neutron scattering techniques. (*B*) The X-ray and neutron scattering lengths for the most biologically relevant atoms.

non-ionising and are typically obtained at much lower fluxes, with the result that radiation damage to the sample is generally avoided when using neutrons. This provides an advantage over X-rays when radiation sensitive samples are to be studied. An incoming neutron beam will be scattered by the atomic nuclei of the sample. Unlike X-rays, the scattering lengths (*b*) do not increase with atomic number but rather have specific scattering amplitudes depending on the nuclei and the variation of the neutron scattering length over the periodic table is apparently haphazard (Table 3.1 B). The atoms found in biomolecules all have comparable scattering lengths with the exception of hydrogen. Hydrogen has a negative coherent scattering length (-0.37×10^{-2} cm) and it also contributes strongly to incoherent scattering. Incoherence is used to describe the case when the resulting scattered wave's intensity is the result of the sum of the scattering intensities of individual atoms rather than the coherent case where each of the scattered atoms intensities have been 'added' together to give a single wave in a given direction. In SANS experiments, incoherent scattering from hydrogen will largely contribute to the background noise. H_2O has a scattering length density (SLD) *i.e.* the sum of all the scattering lengths within the molecule divided by its volume, of close to zero (Table 3.1 B) (Perkins, 1988). Deuterium (^2H or D), hydrogen's heavy isotope has a positive scattering length (0.667×10^{-12} cm) and will therefore contribute largely to the coherent scattering. Consequently D_2O has a much higher SLD than H_2O. For this reason, SANS experiments are mainly carried out in buffers containing D_2O to minimize the incoherent background.

3.5.2.5 Neutron instrumentation for SANS

Neutrons can be produced either via a spallation source or by nuclear reaction. At a spallation source neutrons are produced by accelerating protons to high energies and focusing them on a heavy metal target, such as tungsten. This causes an excitation of the nuclei which ultimately results in the production of neutrons. Spallation sources such as ISIS at Didcot in the United Kingdom produce neutrons in bunch modes with a relatively low flux. For high flux production of non-pulsed neutrons, nuclear reactors using fission reactions are required such as the Institut Laue-Langevin (ILL) in Grenoble, France. The fission reaction of Uranium^{235} produces neutrons in the reactor core. A nuclear fission reaction occurs when the uranium nucleus is split into two fission products producing heat and neutrons. This reaction results in the further splitting of uranium nuclei making the reaction continuous. The reactor core is cooled

by the surrounding D_2O pool with magnetic cadmium safety rods present which if required can break the fission reaction. Produced neutrons are then cooled to the required wavelengths and energies and are then directed into the corresponding guide halls to be fed into the various beamlines.

There are three small angle scattering instruments at the ILL; D11, D22 and D33. D22 is dedicated to biomolecules but all three instruments can be used for bioSANS measurements. A schematic of D22 is shown in Figure 3.11, which shows the general layout of a small angle neutron scattering instrument. Neutrons from the reactor which have been moderated to wavelengths of 0.45 – 4 nm are guided down the beamline where the wavelength is selected using a velocity selector as the speed at which the neutrons travel is dependent on their wavelength. The selected wavelength depends on the velocity selector's rotational speed and on how much it is tilted with respect to its axis of rotation. The wavelength will vary +/- 10 % which must be corrected for in the data reduction. Collimation is required to align the beam in the correct plane *i.e.* to focus it onto the sample, with removable collimation sections depending on the focusing required. The maximum flux on D22 is $1.2x10^8$ neutrons/cm^2/s. The beam arrives at the sample which is contained in a precision quartz cuvette (1.00 or 2.00 mm) (Figure 3.11). Each of the ILL SANS beamlines have a sample holder which can hold up to 22 samples each. A ^{3}He multidetector is located inside the vacuum tube and can be arranged at sample to detector distances from 1.1 m to 17.6 m depending on the required *Q* range. ^{3}He converts the neutrons into charged particles which are detected using a Geiger-Muller counter. Similar to the X-ray beamline a beam stop is required to prevent damage to the detector from the directly transmitted (non-scattered) beam. To measure the transmission, attenuators are inserted into the vacuum tube to prevent damage to the detector from the direct beam. D11 and D33 have a similar layout to D22 with the main differences being the flux at each instrument and the *Q* range available due to differences in sample-to-detector distances (Perkins et al., 2008). On D11 the flux is $1x10^8$ n cm^{-2} s^{-1} and available sample-to-detector distances of 1.2 to 39 M. D33 has a flux of $4.1x10^7$ n cm^{-2} s^{-1} anda sample-to-detector distance of 1.2 to 12.8 M. Additionally D33 has two detectors which extends the *Q* range (Dewhurst, 2008).

Sans2d at ISIS is a time of flight SANS instrument with a wavelength range from 0.2 nm – 1 nm. Sans2d contains a spinning disk chopper which will remove every

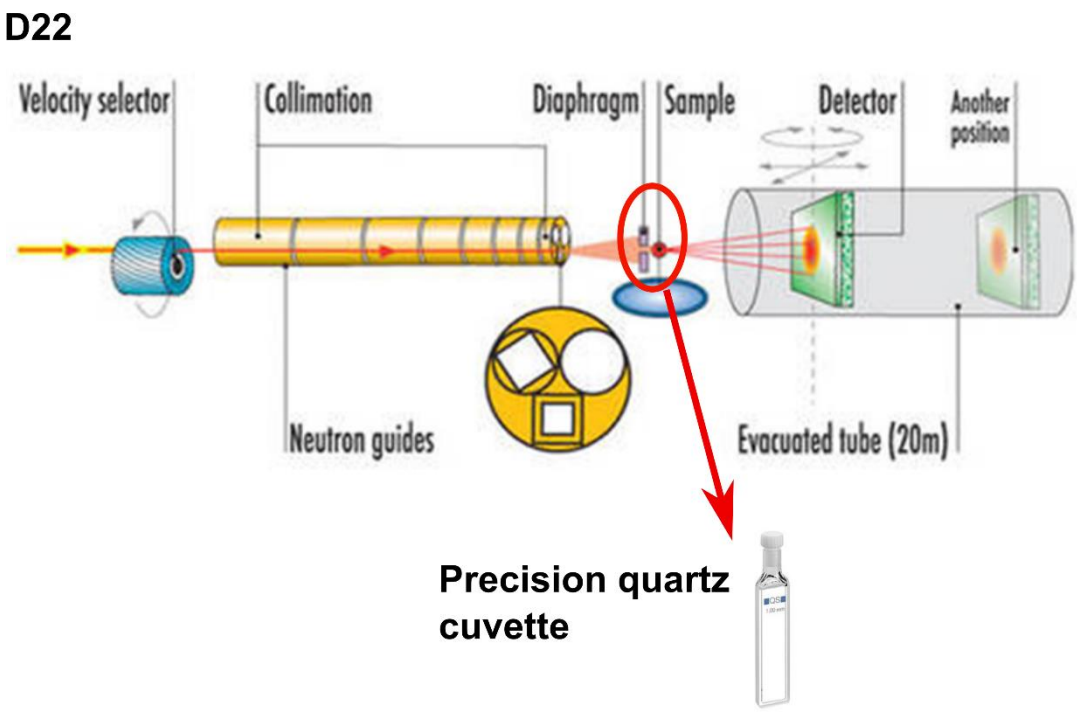


Figure 3.11. A schematic layout of the D22 SANS instrument at the ILL. The neutron beam is arriving from the left. The velocity selector selects the wavelength of the beam which is then focused onto the sample by the collimators. The sample is held in a precision quartz cuvette. The detector in the vacuum tube records the scattering at a certain distance from the sample. (Image adapted from www.ill.fr)

second pulse in order to prevent overlapping between frames, apertures to moderate the size of the beam and five 2 m collimation sections to focus the beam. Sans2d has an automated sample changer which can hold up to 20 samples. It contains two moveable $1m^2$ detectors in a 13 M vacuum giving Sans2d a large accessible *Q* range. The advantage to using a time of flight instrument is that the entire scattering curve within the achievable *Q* range is measured.

During a SANS experiment cadmium, teflon, an empty cuvette and empty sample holder are measured for the data normalisation process. Cadmium measurements are used to monitor the electronic background noise of the detector, teflon is used to align the beam and to mask the beamstop, the empty cell is measured for background subtraction and the empty cell for the measurement of any reflections or straying of the beam. Scattering and transmission are measured for both the sample and the buffer. Transmission measurements are required to calculate the match points and also to confirm the success of dialysis into D_2O buffers as D_2O will have a higher transmission than H_2O. The raw data will be corrected for empty cell and background scattering (cadmium). Buffer subtraction results in the 1D scattering curve *I(Q)* (Perkins *et al*., 2006; Ghosh *et al*., 2006). There are many programs available for SANS data reduction with GRASP and LAMP widely used at the ILL and MANTID at ISIS (Richard *et al.*, 1996).

3.5.3 Contrast Variation with SANS

The scattering of a particle in solution arises due to the difference in SLD between the particle and the buffer *i.e.* the contrast between the two. Contrast is valid at low resolutions where a group of atoms in a volume are considered together with a continuous distribution. The contrast is described as being the difference in the scattering density of the solvent and the scattering particle;

$$\rho = (b_k/v_k) - \rho^0 \qquad \text{(Eq. 3.19)}$$

where ρ is the contrast and b_k is the scattering length, v_k is the volume and ρ^0 is the buffer SLD. As described above (section 3.5.2.4) hydrogen and deuterium have different scattering lengths with hydrogen having a negative scattering length while deuterium has a larger positive scattering length (Table 3.1 B). Consequently H_2O and D_2O have vastly different SLD's. The differential ways in which neutrons interact with hydrogen and deuterium provide SANS with its main advantage over SAXS through

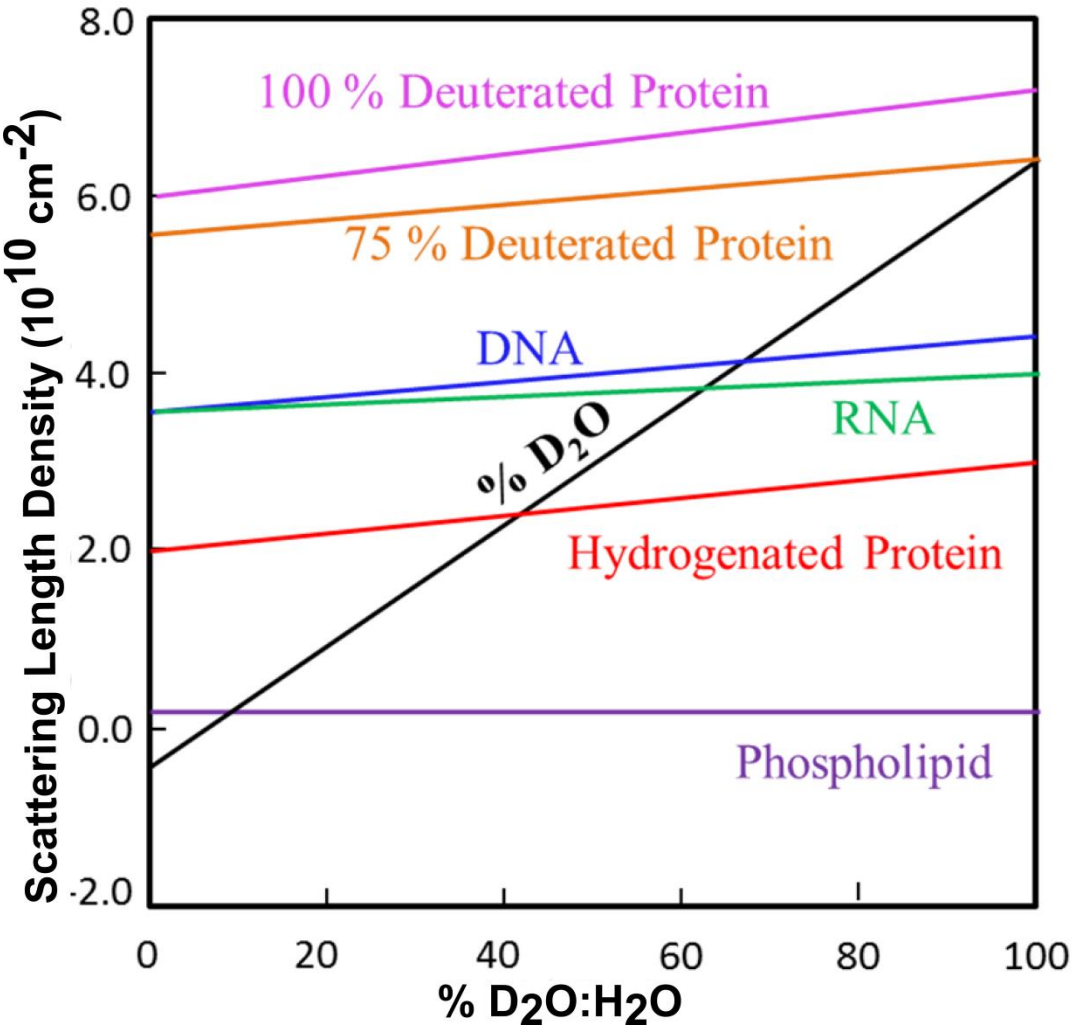


Figure 3.12 An illustration of the scattering length densities of different classes of biomolecules as a function of the percentage of D_2O in H_2O/D_2O solvent. This assumes that 90% of exchangeable hydrogens are exchanged. The black line shows the liner relationship of D_2O percentage with its scattering length density. The match point of the biomolecule is the intersect of this line i.e. when the scattering length density of the molecule is the same as the solvent (Adapted from Jacrot, 1976).

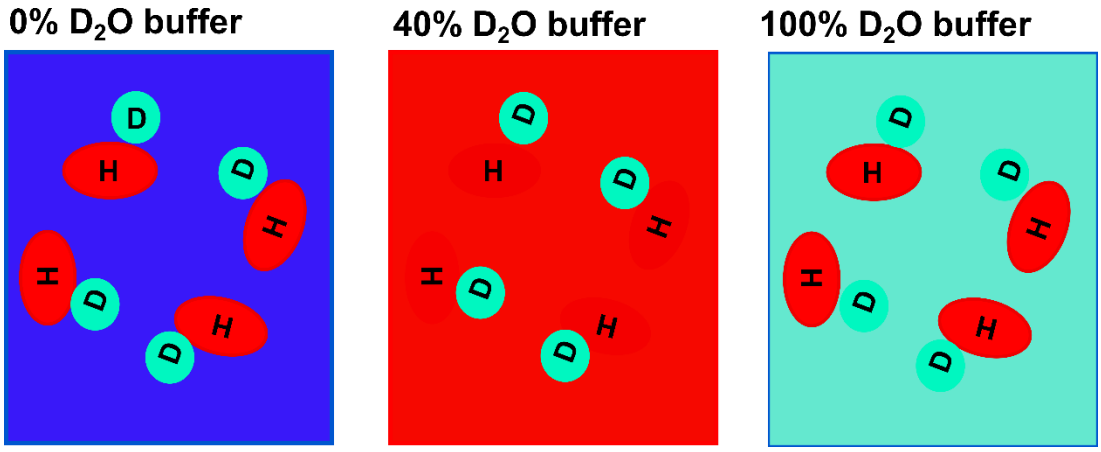


Figure 3.13 A schematic representation of a SANS contrast variation experiment. 'H' represents a hydrogenated protein while 'D' represents a 75% isotopically deuterated protein. In buffer containing 0% D_2O both proteins will be visible to the neutron beam, in 40% D_2O buffer only the 75% deuterated protein will be visible and in 100% D_2O buffer only the hydrogenated protein will be visible.

contrast variation experiments. Naturally occurring differences in neutron SLD exist within the major classes of biomolecules with proteins, lipids, and nucleic acids all having different SLD's (Figure 3.12). Altering the ratio between H_2O and D_2O in the buffer changes the SLD of the buffer subsequently changing the contrast. The SLD of the buffer can be altered to match that of the biomolecule meaning that there is no contrast between the two and therefore no scattering for that biomolecule is observed. The percentage of D_2O (v/v) with respect to H_2O in the buffer, required for this to occur is described as the molecules 'match point'. Contrast variation is a powerful method when examining the complexes formed between biomolecules allowing for individual classes to be selectively observed. For complexes consisting of the same class of biomolecule, for example a complex formed between two proteins, it is necessary to change the SLD of one of the proteins in the complex by isotope labelling with deuterium. Replacing 75% of a proteins non-exchangeable hydrogens with deuterium significantly increases its SLD and therefore match point with respect to a hydrogenated protein (Figure 3.12). This means that it is possible to carry out contrast variation experiments on a protein complex. Figure 3.13 shows an illustration of a contrast variation experiment of a protein complex consisting of one hydrogenated and one 75% isotopically deuterated protein. In a buffer containing 0% D_2O both proteins will be visible to the incoming neutron beam and the resultant scattering curve will contain information about the overall shape of the complex. In 40% D_2O only the 75% deuterated protein will be visible with the resultant scattering curve representing only the deuterated protein. Likewise in 100% D_2O buffer, only the hydrogenated protein will be visible to the beam. Protein deuteration and contrast variation were used in this thesis to examine the interaction between C3d and SCR-19/20 of complement factor H. The method of deuteration and contrast variation are discussed in detail in Chapter Six.

3.5.4 SAS data analysis

Analysis of the 1D scattering curve (*I(Q)*) is similar for both SAXS and SANS. At low *Q* values the Debye equation is a Guinier approximation. The Guinier approximation gives the radius of gyration (R_G) and the *I(0)* or the extrapolated scattering intensity at zero scattering angle which is described as;

$$\ln I(Q) = \ln I(0) - (1/3)R^2_G Q^2 \qquad \text{(Eq. 3.20)}$$

This equation is valid provided $Q.R_G$ is less than 1.3. A guinier plot is obtained by plotting ln *I(Q)* against Q^2. From this;

$$I(0) = \exp(\text{intercept}) \quad \text{(Eq. 3.21)}$$

and;

$$R^2_G = 3 \times \text{slope} \quad \text{(Eq.3.21)}$$

The R_G describes the degree of elongation of the molecule being studied while it is possible to calculate the relative molecular weight by normalizing the *I(0)* to the concentration (mg/ml). Concentration effects should be analysed by comparing the R_G of a concentration series. Interparticle effects will result in changes in the Guinier region (low *Q*) and subsequently the R_G will change. The Guinier region will be affected by sample aggregation with exaggerated R_G values. For macromolecules which have an elongated shape it is possible to calculate the mean cross-sectional radius of gyration (R_{XS}) from a fit in a larger *Q* range than the R_G (Figure 3.9 B) (Perkins *et al*., 2008) according to;

$$\ln [I(Q)Q] = \ln [I(Q)(Q)]_{Q\to 0} - \frac{R^2_{XS}\,Q^2}{2} \quad \text{(Eq. 3.22)}$$

The R_G/R_O anisotropy ratio, where R_O corresponds to the theoretical R_G of a sphere of the same hydrated or unhydrated mass, indicates the level of macromolecular elongation. A globular protein will have a R_G/R_O ratio of 1.28 (Perkins, 1988).

The 1D scattering curve *I(Q)* obtained is in reciprocal space, subsequently Guinier analysis at low *Q* is also carried out in reciprocal space. Information about the molecules in the solution in real space can also be obtained by the distance distribution function *P(r)*. The *P(r)* is carried out by an indirect Fourier transform on the scattering curve *I(Q)* which transforms the scattering curve from reciprocal space (nm^{-1}) into real space (nm). This transformation is carried out using the program GNOM which takes into account the fact that the low *Q* data points close to the beam stop will be missing and that high *Q* is limited to the detector area (a Fourier transform requires that $0 \leq Q \geq \infty$) (D. I. Svergun, 1992). The R_G is calculated from the *P(r)* and should be in agreement with reciprocal space R_G. The D_{max} is also obtained from the *P(r)* which is the maximum dimension of the molecule observed in real space. The D_{max} is the point on the x-axis (r (nm)) at which the *P(r)* curve reaches zero on the y-coordinate.

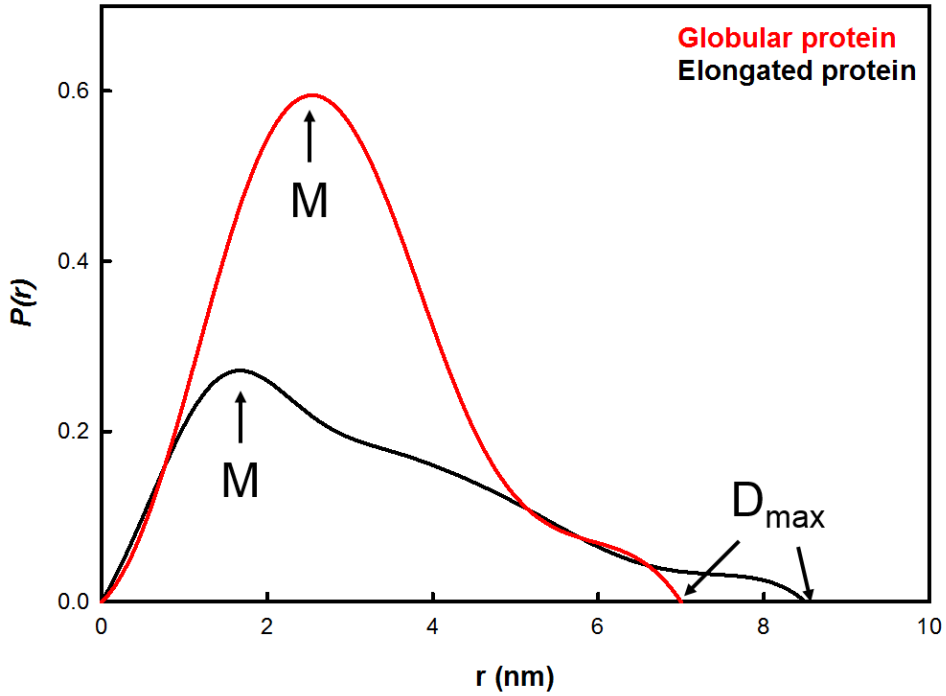


Figure 3.14 Distance distribution function *P(r)* curve. The *P(r)* curve typical of an elongated protein (black) and a globular protein (red). The D_{max} (maximum dimension) is calculated as the intercept of the x-axis at y=0, while M is the most frequently observed intermolecular distance and is calculated as the peak maxima.

Another parameter is obtained from the *P(r)* which is M or the interatomic distance most frequently observed in the molecule. This is read from the x-axis value (nm) of the maximum of the observed peak. The shape of the *P(r)* curve can also indicate the shape of the molecule being studied and changes in this can indicate changes in the molecules shape. Figure 3.14 shows the characteristic curves for an elongated protein (black) and for a globular protein (red). Globular proteins have a symmetric shaped bell curve which tails off slightly while elongated proteins have a more asymmetrical shape.

3.6 Biomolecular modelling of SAS data

Low resolution three dimensional models can be obtained from SAS data provided it is of sufficiently high quality. The constructed model will represent the average of the scattering molecules within the sample which means that aggregation or sample polydispersity will limit the construction of a reliable model. Various modelling software programs are available to model good quality SAS data and are chosen depending on the system being studied and on what information has previously been determined. The ATSAS package developed at the European Molecular Biology Laboratory (EMBL) in Heidelberg provides programs which carry out *ab initio* modelling i.e. no prior information is known, and programs using rigid-body modelling where complete or partial (subunits or domains) high resolution models can be used to produce SAS models. *Ab initio* modelling programs such as DAMMIN and DAMMIF can be used to model SAS data of a single-component system. A sphere with a defined radius and filled with so called dummy atoms or spheres also with a defined radius are used as the starting search volume. This shape is then reconstructed to fit the experimental SAS data using various penalties to do so. For example one such penalty is the looseness penalty which prevents the production of a non-compact or non-interconnected model. Simulated annealing is then employed to perform global minimization of the defined space (Svergun, 1999). Other programs are available such as GASBOR which uses a similar method to DAMMIN but uses chain ensembles of dummy residues with each dummy residue approximating the position of the carbon backbone chain of the protein (Svergun *et al.*, 2001). Rigid body modelling programs such as SASREF and MASHA allow modelling of multicomponent systems where high resolution information is available for one or more of the subunits. Refinement using the high resolution structures against the scattering data can be carried out. These

programs also allow for the validation of high resolution structures. CRYSOL and CRYSON are two programs available which calculate the theoretical X-ray or neutron scattering curves from high resolution PDB models.

Atomistic modelling approaches are also available such as those carried out by the program CCP-SASSIE. A high resolution structure or homology model is used as a start point for atomistic modelling. A monte carlo simulation is used to produce a library of models based on defined input parameters for example regions of known flexibility are specified and will be varied when the library is constructed while the domains are considered static. The theoretical scattering curves of the produced models are then calculated using the SCT program. These are then fit to the experimental data and the best fit models are chosen (Curtis *et al.*, 2012; Wright & Perkins, 2015).

3.7 Thesis Aims

The overall aim of this thesis is to examine and characterise the interactions of the C-terminal of CFH in order to understand how it regulates the complement system. The C-terminal SCR-16/20 domains of CFH contain a weak self-association. The aim of Chapter Four is to further locate this site within the C-terminal and to examine its structure using size exclusion chromatography, AUC and SAXS. There is also evidence to suggest that the C-terminal SCR-16/20 domains contain a binding site for CRP. Previous work on the CFH-CRP interaction has been controversial and so the aim of Chapter Five is to determine which of the C-terminal domains interacts with pentameric CRP and to determine the strength of the interaction under native conditions. Size exclusion chromatography, AUC, FDS-AUC, MST, SPR and negative stain EM are employed to characterise this interaction. The interaction between the C-terminal SCR-19/20 domains and C3d are better characterised, however crystal structures have provided ambiguous results with respect to binding stoichiometry's. The aims of Chapters Six and Seven were to firstly express deuterated protein with could be used for SANS experiments and then to use this protein to characterise the interaction between SCR-19/20 and C3d in solution using contrast variation SANS and SAXS.

Chapter Four

A novel dimerisation site at SCR-17/18 in Factor H may explain its disease-causing mutations and a new mechanism for regulatory control

4.1 Introduction

The complement system is an enzymatic cascade in the innate immune system which functions to remove invading pathogens or damaged cells before they can cause infection. In the alternative pathway of complement activation, complement C3 is spontaneously hydrolysed to active C3b which binds exposed surfaces targeting them for immune destruction. Generation of C3b brings about the formation of the C3 convertase which further hydrolyses C3, amplifying the complement response.

Complement Factor H (CFH) prevents complement-mediated host cell destruction through the interaction of its C-terminal with surface bound C3b on non-activator surfaces (Rother *et al*., 1998). CFH acts as a cofactor for Factor I which cleaves C3b to inactive iC3b (Pangburn *et al*., 1977). CFH recognises host cells through interactions with their anionic surfaces. It is a 154 kDa glycoprotein with nine N-linked glycosylation sites (Fenaille *et al*., 2007). It is composed of 20 short complement regulator (SCR) domains which in turn are composed of approximately 60 amino acids and are linked by three to eight amino acids, having varying degrees of flexibility (Soares & Barlow, 2005). It is difficult to obtain structural information for full length CFH due to its size, glycosylation and flexibility. High resolution structural information is available for some of the domains with SCR-6/8, SCR-18/20, and SCR-19/20 solved by X-ray crystallography (Jokiranta *et al*., 2006; Prosser *et al*., 2007; Morgan *et al*., 2012) and SCR-1/3, SCR-5 and SCR-15/16 solved by NMR (Barlow, 1992; Barlow, 1993; Hocking *et al*., 2008). Electron microscopy and small angle scattering indicates that full length CFH exists with a folded back structure at either its N- or C- terminals (Okemefuna *et al*., 2009b; DiScipio, 1992). The CFH C-terminal SCR-19 and SCR-20 domains interact with C3b and its thioester domain C3d (Sharma & Pangburn, 1996; Jokiranta *et al*., 2000). SCR-20 interacts with the cell surface glycosaminoglycans (Blackmore *et al*., 1998; Blaum *et al*., 2015).

Atypical haemolytic uraemic syndrome (aHUS) is a rare disease which is characterised by damage to the endothelial cells of the kidney through impaired complement regulation. It can lead to renal failure and because of this is often fatal (Kavanagh *et al*., 2006). aHUS is linked to mutations in CFH (Saunders *et al*., 2006; Rodriguez *et al*., 2014). Up to 113 aHUS associated mutations have been described for CFH. Fifty of these are located in the five C-terminal SCR domains with 22 in SCR-20, 9 in SCR-19 and 19 in SCR-16/18 (Figure 4.1 C) (Rodriguez *et al*., 2014).

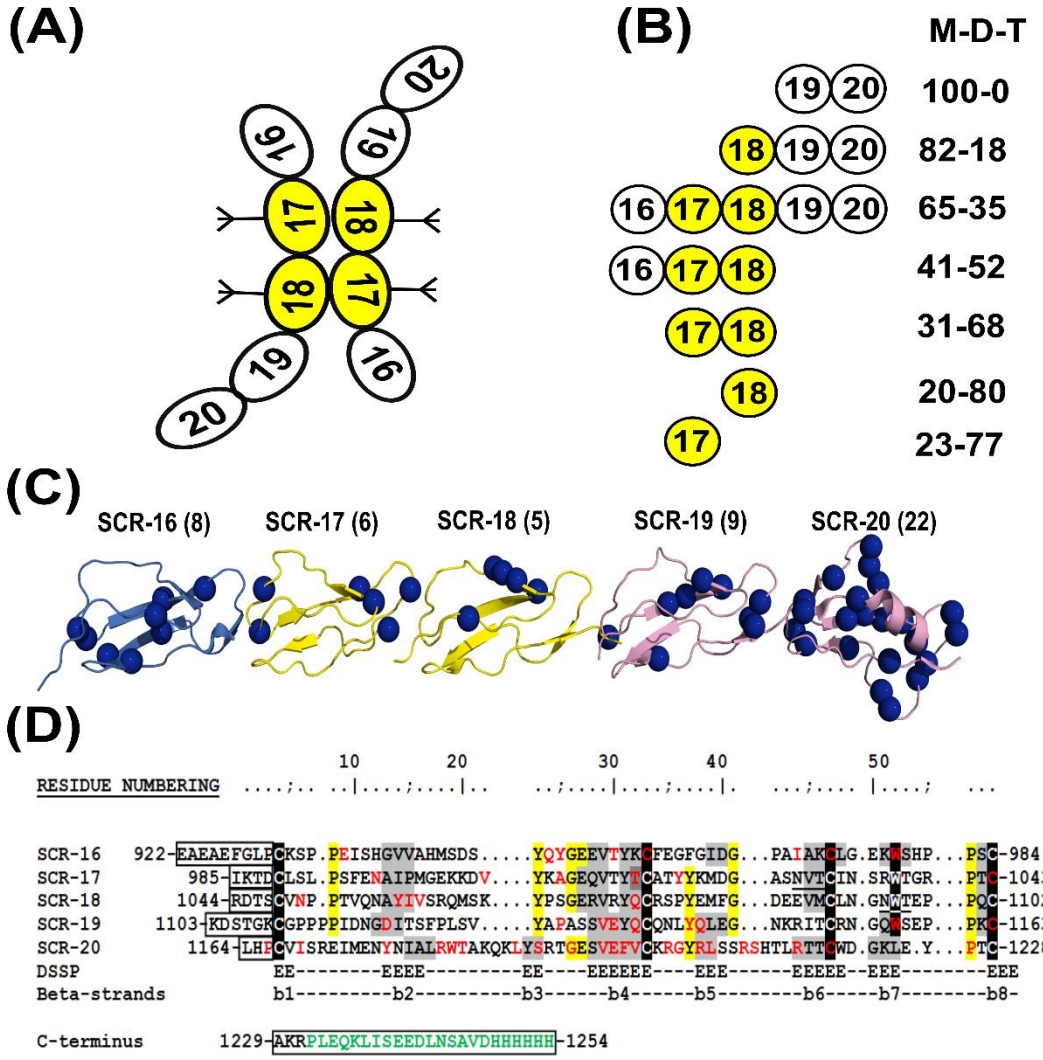


Figure 4.1 The CFH C-terminal region. *A*, A schematic diagram of the SCR-16/20 dimer formed by the side-by-side association of SCR-17 and SCR-18. The two SCR-17 and SCR-18 glycosylation sites are indicated by Y like symbol. *B*, Schematic diagram of the seven SCR fragments expressed for this study, with SCR-17 and SCR-18 highlighted as the dimer site. The right hand side shows the relative percentage of monomer (M), and dimer (D) from the analytical ultracentrifugation analyses. *C*, The structural arrangement of each of the domains with the disease-associated mutations shown as blue spheres. The number of mutations in each domain is shown in brackets beside the domain label. The domains are arranged individually in a line for visualisation purposes and do not represent the actual structure of SCR-16/20. SCR-16 is from its NMR structure (PDB code: 1HCC), SCR-17 is a homology model, SCR-18 and SCR-19/20 are from the crystal structures (PDB Codes: 3SW0 and 2G7I,

Figure 4.1 continued. respectively). *D,* The sequences of each of the SCR16/20 domains. Highlighted in black are the five conserved Trp and Cys residues, yellow shows other conserved residues, the glycosylation sites are underlined, and the linker regions are boxed. Residues highlighted in grey have been assigned β-strand secondary structure using the Define Secondary Structure Prediction tool (Okemefuna *et al*., 2008) with the eight assigned β-strands shown below the sequence. The disease-associated mutations are coloured in red. The C-terminal peptide is indicated below the sequence. For SCR domains expressed with a hexa-histidine tag, the sequence shown in green will be present at the C-terminus. The EAEAF sequence at the N-terminus is part of the α-factor secretion signal sequence and the sequence of the EcoRI restriction enzyme cleavage site.

The majority of these mutations are in SCR-19/20, domains which possess the binding sites for C3b, C3d and surface glycosaminoglycans. Hence mutations here will interrupt these interactions, impairing CFH ability to recognise and protect host cells. SCR-16/18 is a region which contains no reported binding sites; however the presence of mutations in this region indicate that it may be of functional importance.

CFH possesses the ability to self-associate in a weak reversible manner (Perkins *et al*., 1991). CFH forms dimers although higher oligomers have also been observed (Nan *et al.,* 2008). It was shown that the SCR-6/8 and SCR-16/20 domains contained two CFH self-association sites. SCR-16/20 was shown to exist in a weak monomer-dimer equilibrium using analytical ultracentrifugation (AUC) and small angle X-ray scattering (SAXS) (Fernando *et al.,* 2007; Okemefuna *et al.,* 2008). From these results it was proposed that the dimer was formed by an end-to-end association of two SCR-20 domains. Preliminary characterisation of the C-terminal SCR domains four SCR fragments indicated that the dimer site may be within the SCR-16/18 domains and not within SCR-19/20 (Orla Dunne, *MRes dissertation*).

To identify which of the SCR domains within the CFH C-terminal SCR-16/20 contained the dimer site we expressed seven recombinant fragments of the C-terminus (Figure 4.1 B), and carried out size exclusion chromatography (SEC), analytical ultracentrifugation (AUC), and small angle X-ray scattering (SAXS) experiments. We show that the CFH self-association site is within the SCR-17/18 domains. From the AUC and SAXS results in both 137 mM NaCl and 50 mM NaCl buffers we hypothesise that it is formed by a side-by-side weakly hydrophobic anti-parallel association of the SCR-17/18 domains (Figure 4.1 A). The results may explain why aHUS mutations are seen in the SCR-17/18 region of CFH, and suggest that self-association is important for CFH function.

4.2 Materials and Methods

4.2.1. Expression of SCR Fragments

SCR-19/20, SCR-16/20, SCR-18/20, SCR-16/18H, SCR-17/18H, SCR-17H and SCR-18H were all expressed and purified for this study (Cheng *et al*., 2005). Figure 4.1 (B) shows an illustration of the fragments expressed for this study. The ‘H’ indicates the presence of a histidine tag (His-tag). Cloning and transformations were carried out by Dr. P. Adamson and Prof. D.L. Gordon at the Flinders Medical Centre

and Flinders University, Australia. The SCR fragments were cloned into the *Pichia pastoris* expression plasmid pPICZαA which was then transformed into wild-type X33 cells. Expression was carried out according to Invitrogen guidelines (EasySelect™ *Pichia* expression kit manual). Cells which had the plasmid incorporated into their genomes were selected on agar plates containing zeocin antibiotic as pPICZαA contains a zeocin resistant gene. Cell growth was maintained in media containing 2 % glycerol for 4 four days. Cells were then induced using 0.5 % methanol (v/v). Expression was maintained by addition of 0.5% methanol every 24 hours for 96 hours. Cells were removed by centrifugation and the supernatant containing the secreted SCR domains was concentrated using a 5 kDa MWCO membrane in preparation for purification.

4.2.2. Purification and identification of SCR fragments

The four fragments which were expressed with C-terminal His-tags were purified using a 5 ml HiTrap Nickel column (GE Healthcare). The supernatant was dialysed against the wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0.) and loaded onto the column using an AKTA purifier system (GE Healthcare, Buckinghamshire, UK) which had been equilibrated with the same buffer. Five column volumes of nickel wash buffer removed any non-specifically bound protein. The His-tagged SCR fragments were eluted using 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0. For the three non-tagged SCR fragments ion exchange chromatography was used. As SCR-19/20, SCR-16/20 and SCR-18/20 have theoretical isoelectric points of 9.05, 8.04 and 7.69 respectively (calculated from their primary sequences using ExPASy), cation exchange chromatography was used. The supernatant was dialysed against 50 mM Tris-Cl, 25 mM NaCl, 1 mM EDTA, pH 7.4 and loaded onto a SP FF column (GE Healthcare) which had been equilibrated with the same buffer. After loading, the column was washed with five column volumes of the initial buffer. Elution was achieved by applying a salt gradient from 25 mM NaCl to 1 M NaCl (final concentration). For all seven purifications, protein elution was monitored by measuring the absorbance at 280 nm. SDS-PAGE was used to confirm the presence of the SCR fragments with fractions pooled and concentrated using Amicon ultra centrifugal filters with a molecular weight cut-off of 10 kDa or 3 kDa depending on the SCR fragment size. SEC was carried out as a final purification step to remove any remaining impurities and aggregation. Protein samples were loaded

onto a Superdex 75 (GE Healthcare) column which had been equilibrated with 10 mM Hepes, 137 mM NaCl, pH 7.4. SDS-PAGE was used to analyse sample purity. Western blot analysis was carried out using anti-Factor H polyclonal goat antibodies (CALBIOCHEM, Hertfordshire, UK) to confirm the identity of the CFH SCR fragments. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy was carried out on SCR-18H to confirm its mass at 11,180 Da as this was not visible on the western blot. Mass spectroscopy was performed through the Mass Spectroscopy platform of the Partnership for Structural Biology (PSB), Grenoble, France operated by Dr. Luca Signor.

The sequence alignment for the five C-terminal SCR-16/20 domains was carried out on the sequence taken from Uniprot KB. The N- and C- terminal sequences of the expressed fragments depend on the fragment. For all seven SCR fragments the N-termini contain the sequence –EAEAF corresponding to the α-factor secretion signal and the EcoRI restriction enzyme site. The SCR-18H, SCR-17H, SCR-17/18H and SCR-16/18H C-termini contain the first four amino acids of the next linker region followed by the myc tag and His tag sequences – ALEQKLISEEDLNSAVDHHHHHH. The SCR-18H, SCR-17H, SCR-18/20 and SCR-17/18H N-termini contain the last residue of the previous linker. SCR-16/18H and SCR-16/20 contain the last three residues of the previous linker while SCR-19/20 contains the last four residues of the linker at its N-terminus.

4.2.3. Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) data were collected for each of the expressed SCR fragments in 10 mM Hepes, 137 mM NaCl, pH 7.4 and in 10 mM Hepes, 50 mM NaCl, pH 7.4. Data were collected in a concentration series ranging from 0.2 – 3 mg/ml for each of the SCR fragments. Sedimentation velocity experiments were carried out at 20 ºC using an AnTi50 rotor at 50,000 rpm. Interference and absorbance optics at 280 nm were used for detection depending on the sample concentrations (absorbance optics were saturated at an absorbance > 1.5). Experiments were carried out in two sector cells (buffer and sample) with column heights of 12 mm. Size distribution *c(s)* analysis of the obtained data was carried out using the program SEDFIT (version 14.6) (Schuck, 1998, 2000). SEDFIT calculates the *c(s)* by fitting the experimental data to the Lamm equation (Lamm, 1929). 80-100 scans were used for *c(s)* analysis with the frictional ratio (f/f_0), meniscus and baseline

all floated for final analysis. *c(s)* plots were converted to molar mass mass distribution *c(M)* which gave an estimate of the molecular mass of each sedimenting species. The partial specific volume $\bar{v}$ was calculated for each SCR fragment using the program SLUV (Perkins, 1988). The calculated $\bar{v}$ were 0.7277 ml/g for SCR-19/20, 0.7151 ml/g for SCR-16/20, 0.7164 ml/g for SCR-18/20, 0.7071 ml/g for SCR-16/18H, 0.7021 ml/g for SCR17/18H, 0.6954 ml/g for SCR-18H and 0.6993 mg/g for SCR-17H. The buffer densities (ρ) were determined using a densitometer (Qualtech Products Industry, UK) and were calculated at 1.00487 g/ml for 137 mM NaCl and 1.00196 for 50 mM NaCl. The buffer viscosity (η) was 1.002 cp (Hardy & Cottington, 1949).

4.2.4. Small angle X-ray scattering

SAXS experiments were performed on each of the SCR fragments in both 10 mM Hepes, 137 mM NaCl, pH 7.4 and 10 mM Hepes, 50 mM NaCl, pH 7.4 over a concentration series of 0.5 – 2 mg/ml with concentrations varying depending on the SCR fragment. Data were collected on the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF) France. The X-ray wavelength was 0.9919 Å, with energy of 12.5 KeV (Pernot *et al.*, 2013). All experiments were carried out at 20 ºC. The automated capillary flow sample changer on BM29 was used in which the buffer was measured before and after the sample. 50 µl sample volumes were used. 10 frames with one second per frame were collected for each sample. Frames which showed no radiation damage or aggregation were averaged. Averaged buffer frames were then subtracted from averaged sample scattering curves. Guinier analyses at low *Q* were then performed according to the Guinier equation;

$$\ln I(Q) = \ln I(0) - R_G^2\, Q^2/3 \qquad \text{(Eq. 4.1)}$$

where *Q* is the scattering vector ($4\pi \sin\theta/\lambda$), *I* is the intensity as a function of *Q* and R_G is the radius of gyration. The R_G was calculated, which is a measure of the protein's overall elongation. This was calculated from the linear portion of the Guinier plot (ln *I(Q)* v Q^2) within the $Q.R_G$ limit of 1.5. The forward scattering intensity at zero angle *I(0)* was also calculated. For elongated molecules, the radius of the cross-section (R_{XS}) can be calculated at larger *Q* values according to the equation;

$$\ln [I(Q).Q] = \ln [I(Q).Q]_{q\to 0} - R_{XS}^2\, Q^2/2 \qquad \text{(Eq. 4.2)}$$

Initial data subtraction and Guinier R_G analyses was carried out using the software Primus (Konarev *et al*., 2003).

The program GNOM was used to transform the reciprocal scattering data (*I(Q)*) into real space via an indirect Fourier transform to give the distance distribution *P(r)* function (Svergun, 1992);

$$P(r) = \frac{1}{2\,\pi^2} \int_0^\infty I(Q) r \sin(Qr) dQ \qquad \text{(Eq. 4.3)}$$

The *P(r)* yields the real space R_G, the maximum observed length of the molecule (L) and the interatomic distance most frequently observed in the molecule (M).

4.2.5 Modelling of the C-terminal dimer

Sequence alignment of the SCR-1/2 domains of complement Factor H related-1 (CFHR1) protein and SCR-17/18 was carried out for the sequences obtained from Uniprot KB using the EMBOSS water sequence pairwise alignment tool from EMBL (Longden & Bleasby, 2000). PyMOL was used to structurally align the SCR-17/18 domains with the CFHR1 SCR-1/2 domains. The crystal structure of the CFHR1 SCR-1/2 dimer was used (PDB code: 3ZD2). SCR-17 was a homology model from the solution structure of SCR-16/20 while SCR-18 was taken from the crystal structure of SCR-18/20 (Okemefuna *et al.*, 2008; Morgan *et al.*, 2012). Alignment was carried out by aligning the residues of the β-4 strand of each of the domains. SCR-17 was aligned with SCR-1 and SCR-18 was aligned with SCR-2.

4.3 Results

4.3.1. Purification of the seven SCR Fragments

The non-tagged SCR-19/20, SCR-18/20 and SCR-16/20 fragments were successfully purified from the *Pichia pastoris* growth media supernatant by cation exchange chromatography. SCR-16/18H, SCR-17/18H, SCR-17H and SCR-18H were successfully purified from the yeast growth media supernatant by nickel affinity chromatography. SEC separates proteins according to their size and shape. Figure 4.2 (A) shows the elution profiles of each of the SCR fragments with elution volume against the absorbance at 280 nm. Molecular weight standards were used to calibrate the Superdex 75 size exclusion column. The standards comprised of globular proteins

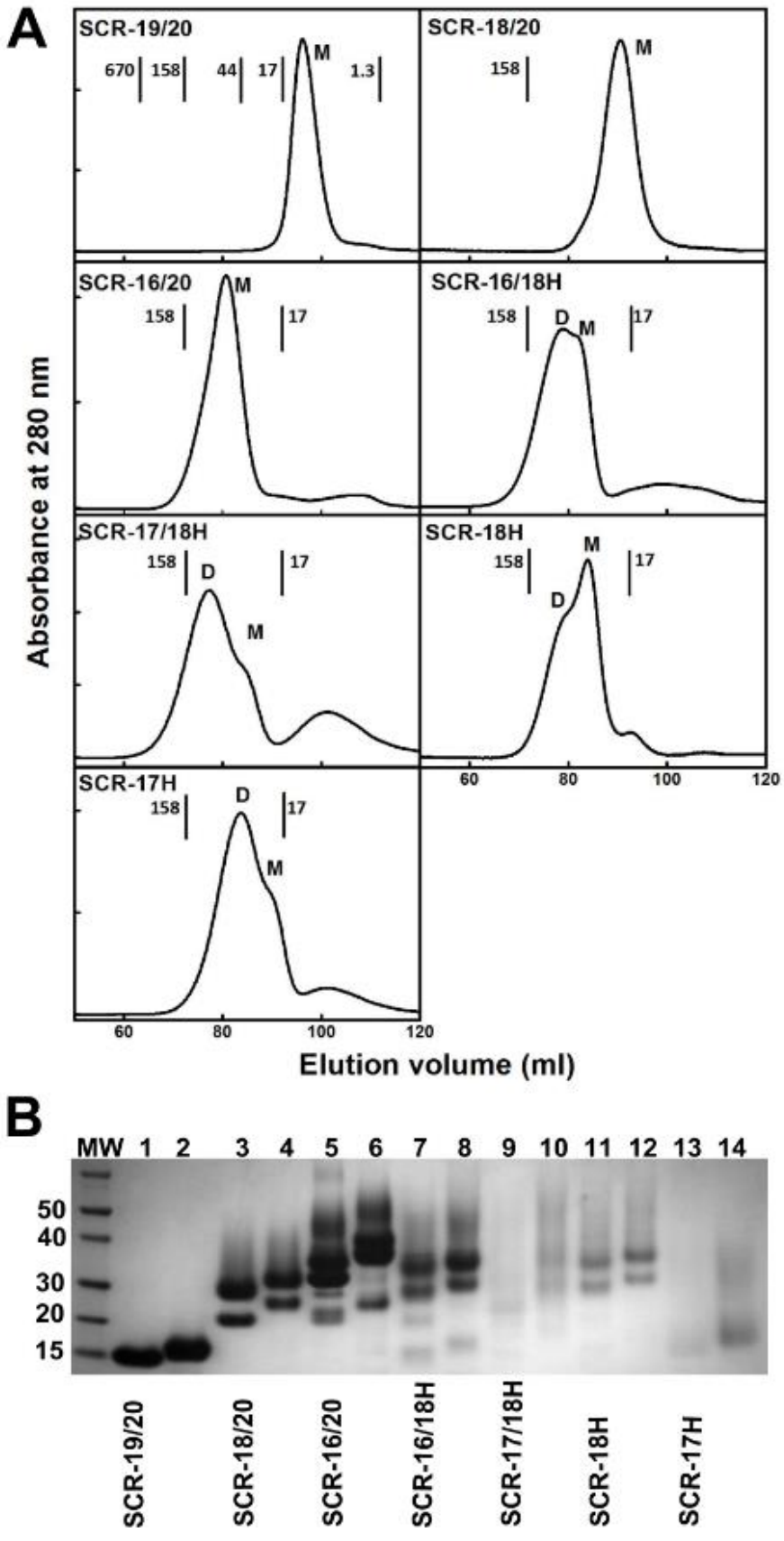


Figure 4.2 legend overleaf

Figure 4.2. SCR fragment purification. *A,* Size exclusion chromatography elution profiles for each of the complement Factor H SCR fragments studied. The molecular mass of the five calibration standards are shown for SCR-19/20. SCR-19/20 elutes as a single monomeric peak, SCR-18/20 and SCR-16/20 elute with broad peaks, while SCR-16/18H, SCR-17/18H, SCR-18H and SCR-17H all elute with peaks partially resolved into two components designated as monomer (M) and dimer (D).

B, SDS-PAGE analysis of each of the SCR fragments studied. Lane 1, SCR-19/20 non-reduced; lane 2, SCR-19/20 reduced; lane 3, SCR-18/20 non-reduced; lane 4, SCR-18/20 reduced; lane 5, SCR-16/20 non-reduced; lane 6, SCR-16/20 reduced; lane 7, SCR-16/18H non-reduced; lane 8, SCR-16/18H reduced; lane 9, SCR-17/18H non-reduced; lane 10, SCR-17/18H reduced; lane 11, SCR-18H non-reduced; lane 12, SCR-18H reduced; lane 13, SCR-17H non-reduced; lane 14, SCR-17H reduced; Molecular weight standards (MW) shown in kDa. Streaking and multiple bands are evident due to varying glycosylation patterns on the SCR fragments.

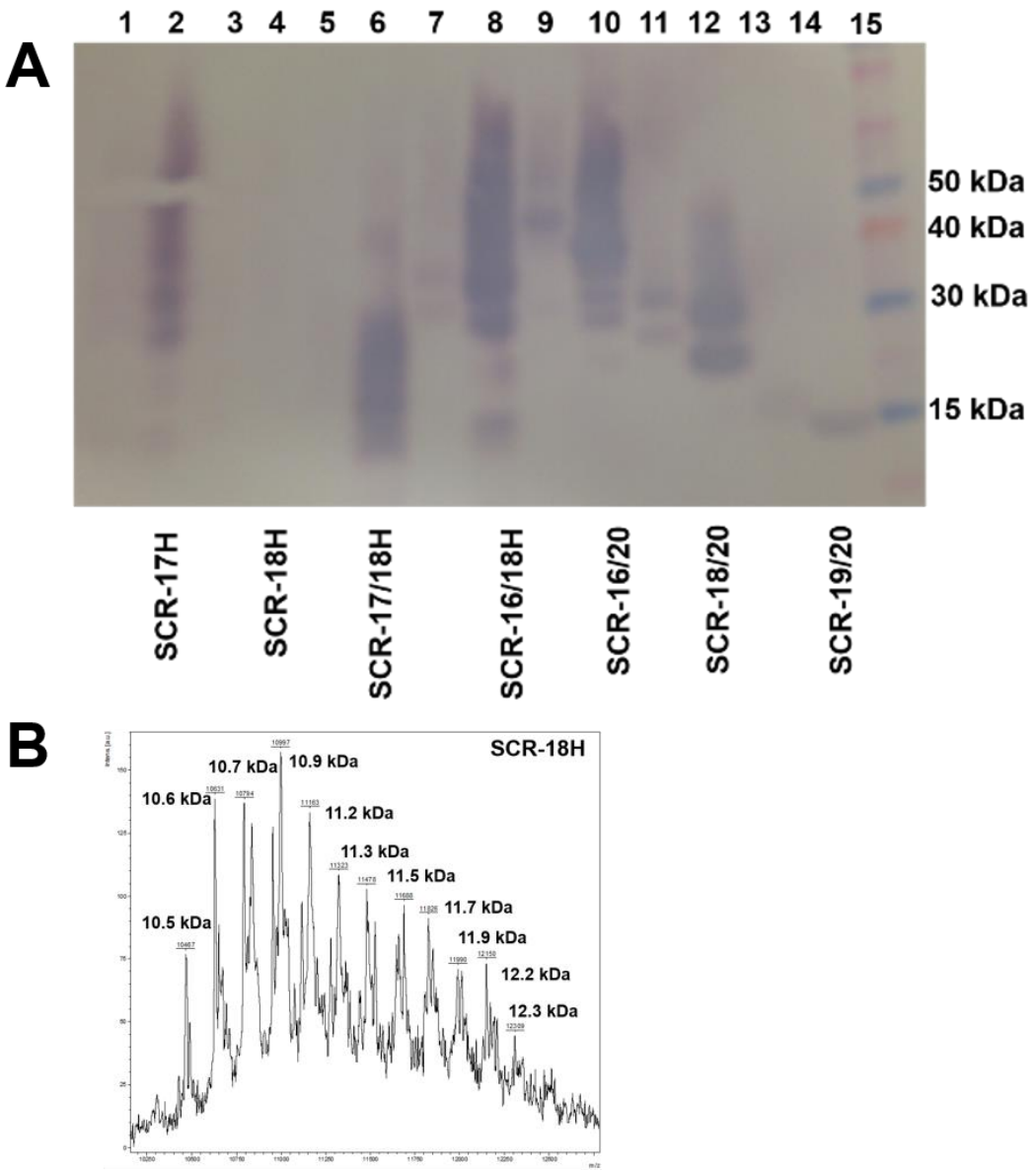


Figure 4.3. Confirmation of the seven SCR fragments. *A*, Western blot analysis. A anti-Factor H goat antibody was used. Lane 1, SCR-17H reduced; lane 2, SCR-17H non-reduced; lane 3, SCR-18H reduced; lane 4, SCR-18H non-reduced; lane 5, SCR-17/18H reduced; lane 6, SCR-17/18H non-reduced; lane 7, SCR-16/18H reduced; lane 8, SCR-16/18H non-reduced; lane 9, SCR-16/20 reduced; lane 10, SCR-16/20 non-reduced; lane 11, SCR-18/20 reduced; lane 12, SCR-18/20 non-reduced; lane 13, SCR-19/20 reduced; lane 14, SCR-19/20 non-reduced; lane 15, pre-stained molecular weight standards (MW). *B*, MALDI-TOF mass spectroscopy spectrum for SCR-18H which shows a variation in mass due to variations in the glycosylation pattern.

of known molecular mass (BioRad Gel Filtration Standard, Hertfordshire, UK) and allowed the apparent molecular mass of each SCR fragment to be estimated. SCR-19/20 elutes as a single symmetrical peak with an apparent molecular mass of 10 kDa, which is comparable to the expected 14.7 kDa. SCR-18/20 elutes with a single broad peak with a small shoulder peak on the left as does SCR-16/20. SCR-18/20 has an apparent molecular mass of 77 kDa and SCR-16/20 of 71 kDa, which are much larger than the expected mass of 24 kDa and 38 kDa respectively. SCR-16/18H, SCR-17/18H, SCR-17H and SCR-18H all consistently eluted with two partially merged peaks corresponding to dimer (D) and monomer (M). SCR-16/18H, SCR-17/18H and SCR-17H all eluted with a larger dimer peak with respect to monomer. For SCR-16/18H, the dimer peak had an apparent molecular mass of 141 kDa, while the monomer was 120 kDa. SCR-16/18H has a calculated molecular mass of 29 kDa. SCR-17/18H dimer peak had an apparent molecular mass of 102 kDa and a monomer peak at 84 kDa. SCR-17/18H has a calculated molecular mass of 22 kDa. SCR-17H showed apparent molecular masses of 48 kDa and 23 kDa for the dimer and monomer, respectively. It has a calculated molecular weight of 10 kDa. SCR-18H showed a larger peak for monomer than dimer with apparent molecular masses of 79 kDa for the dimer and 48 kDa for the monomer. SCR-18H has a calculated molecular weight of 11 kDa. The molecular weight standard proteins used to calibrate the size exclusion column were a set of globular proteins while the SCR fragments are elongated accounting for the discrepancies between observed and calculated molecular masses.

Protein purity was analysed by SDS-PAGE (Figure 4.2 B). The SCR-17 and SCR-18 domains contain N-linked glycosylation sites and hence six of the SCR fragments (with the exception of SCR-19/20) run with streaking on the SDS-PAGE gel which is characteristic of glycosylation. Multiple bands are also evident for the glycosylated fragments which are due to variations in the glycosylation pattern which is often observed for glycoproteins expressed in *Pichia pastoris.* Western blot analysis using an anti-CFH goat antibody confirmed the identity of the SCR fragments (Figure 4.3 A). The signals for the reduced samples were significantly weaker than those for the non-reduced. Each SCR domain contains two disulphide bridges, therefore the non-reduced form in which these bonds are intact may be important for antibody recognition. SCR-18H was not visible in the western blot, but it was confirmed by mass spectroscopy at the correct molecular mass of 11,180 Da with mass variations

observed (figure 4.3 B). This variation was attributed to differences in the glycosylation pattern arising from *Pichia pastoris* by comparison with that attributable to higher eukaryotic systems (Bretthauer & Castellin, 1999).

4.3.2. Analytical ultracentrifugation data in 137 mM NaCl

AUC separates molecules according to their size and shape. The presence of different species within a sample is detected from the peaks in the size distribution *c(s)* analysis. The sedimentation coefficient at 20 ºC and corrected for the density of water ($s_{20,w}$) is calculated and from this the experimental fractional ratio (f/f_o) where *f* is the calculated frictional coefficient of the protein and f_0 is the frictional coefficient for a spherical protein with the same mass. The f/f_o ratio is a measure of the elongation of the protein with a globular protein having a f/f_o ratio of 1.2. The f/f_o ratio will indicate the change of shape or elongation upon dimerisation of the protein. The relative percentage of the main species present in the sedimenting sample are calculated from SEDFIT as the area under the *c(s)* peak curve. Conversion of the *c(s)* distribution to the molar mass *c(M)* distribution gives an approximation of the molecular mass of each species present. From the peak integrations, the dissociation constant (K_D) for dimer formation can be estimated.

Absorbance and interference optics were used with interference only for the non-His tagged fragments SCR-19/20, SCR-18/20 and SCR-16/20.

(i) SCR-19/20 showed only one peak visible in the *c(s)* plot (Figure 4.4) corresponding to 100 % monomer with an $s_{20,w}$ value of 1.6 S (Figure 4.4, Figure 4.7 A). The *c(M)* gave a calculated molecular mass ranging between 15 kDa and 18 kDa, which was in good agreement with the expected mass of 15 kDa. The f/f_o for SCR-19/20 was 1.12 showing that it had a relatively compact shape (Table 4.1).

(ii) SCR-18/20 showed one peak with an $s_{20,w}$ value of 2.4 S corresponding to the monomeric protein. At higher concentrations (above 1.5 mg/ml) a shoulder peak was evident with an $s_{20,w}$ of 3.1 S corresponding to dimer (Figure 4.4, Figure 4.7 A). This shoulder peak accounted for 10 % and 18 % of the sample for 2 mg/ml and 3 mg/ml, respectively (Figure 4.7 B). The calculated molecular mass for the monomer ranged from 26 to 34 kDa (expected 24 kDa) and 58 kDa for the dimer peak. SCR-18/20 had a f/f_0 of 1.13 for monomer and 1.56 for the dimer peak, showing that the dimer was more elongated than the monomer which had a relatively compact shape

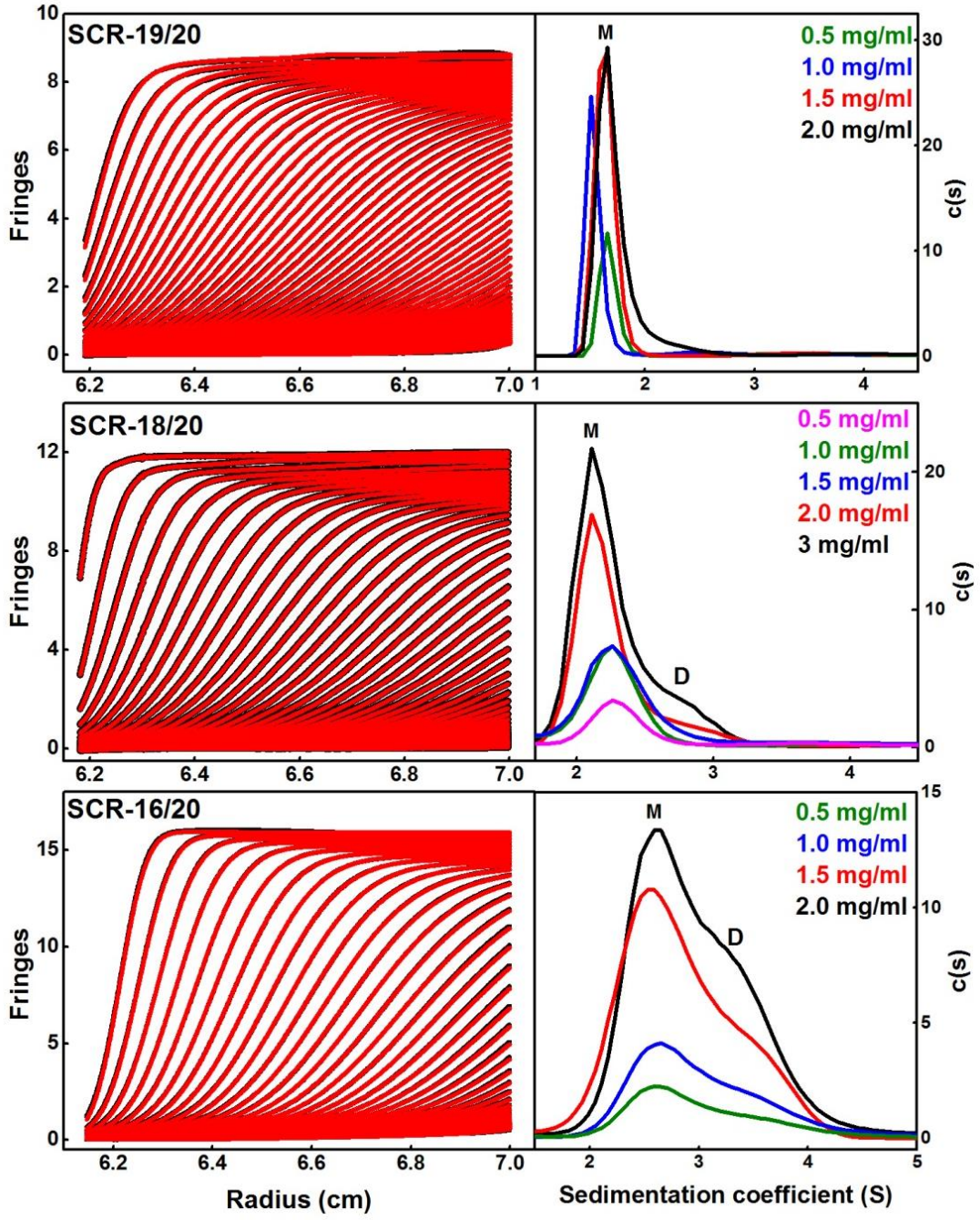


Figure 4.4. Size distribution *c(s)* analyses of SCR-19/20, SCR-18/20 and SCR-16/20 in 137 mM NaCl. Shown in the left panel are the scans with the boundary fits using interference optics shown in red. In the right panel the *c(s)* distribution is shown over the concentrations series used. Monomer and dimer are denoted by M and D respectively. The concentrations are shown in each panel, highlighted according to their colours in the *c(s)* distributions.

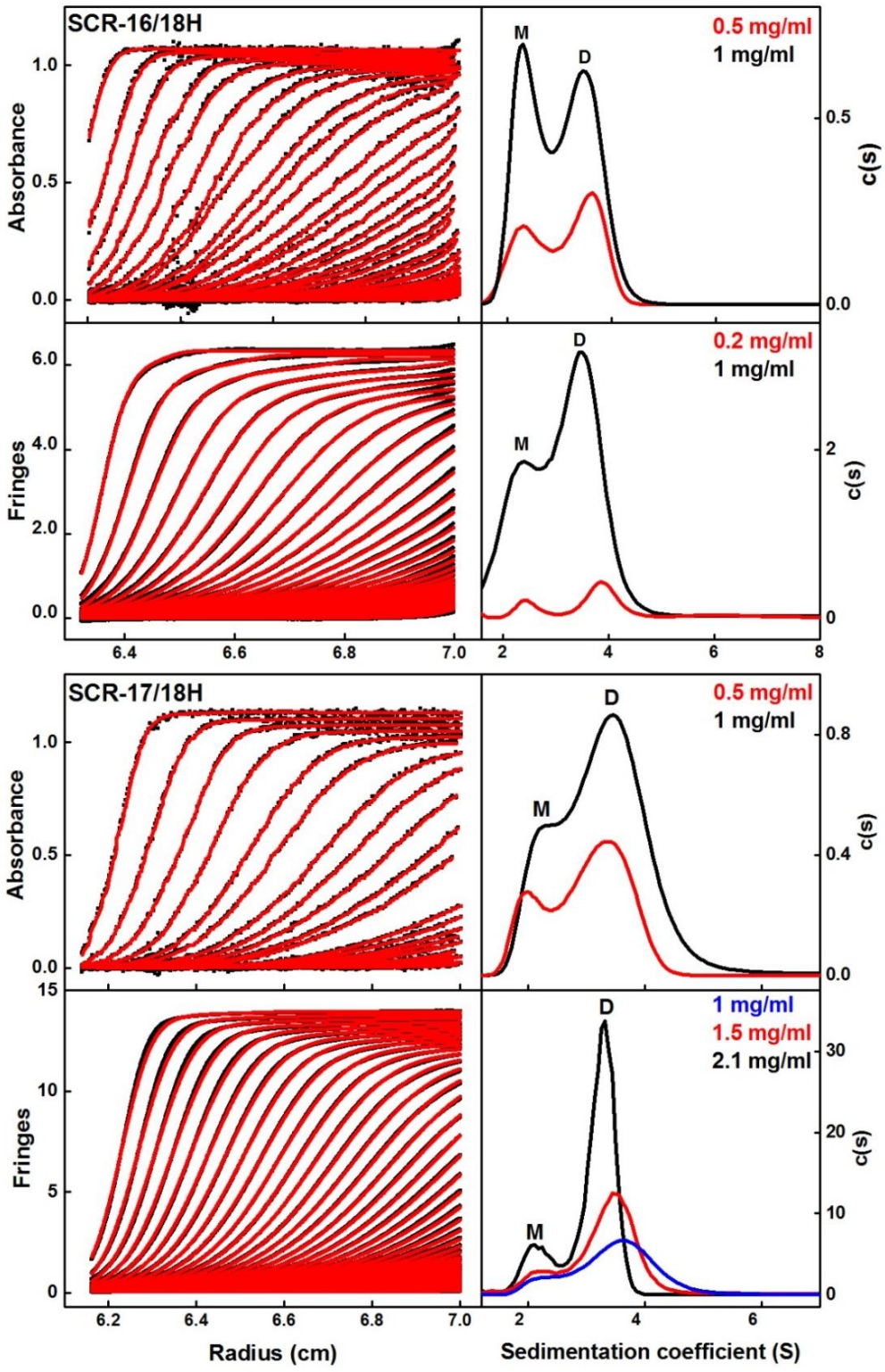


Figure 4.5. Size distribution *c(s)* analyses of SCR-16/18H, and SCR-17/18H in 137 mM NaCl. Shown in the left panel are the scans with the boundary fits shown in red (top absorbance, bottom interference). In the right panel the *c(s)* distribution is shown over the concentrations series used. Monomer and dimer are denoted by M and D respectively. The concentrations are shown in each panel, highlighted according to their colours in the *c(s)* distributions.

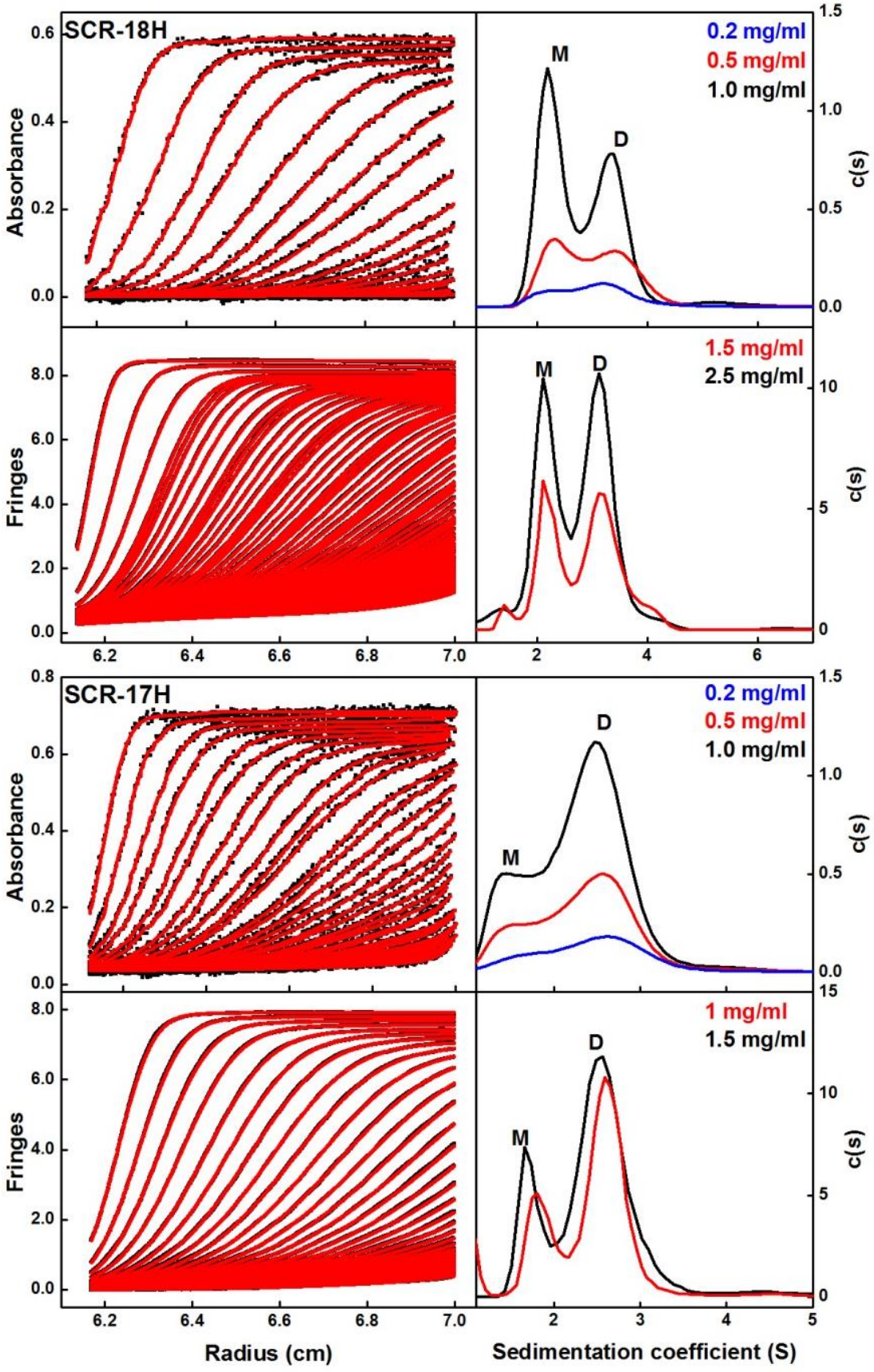


Figure 4.6 legend overleaf

Figure 4.6 Size distribution *c(s)* analyses of SCR-18H and SCR-17H in 137 mM NaCl. Shown in the left panel are the scans with the boundary fits shown in red (Absorbance, top panel; interference, bottom panel). In the right panel the *c(s)* distribution is shown over the concentrations series used. Monomer and dimer are denoted by M and D respectively. The concentrations are shown in each panel, highlighted according to their colours in the *c(s)* distributions.

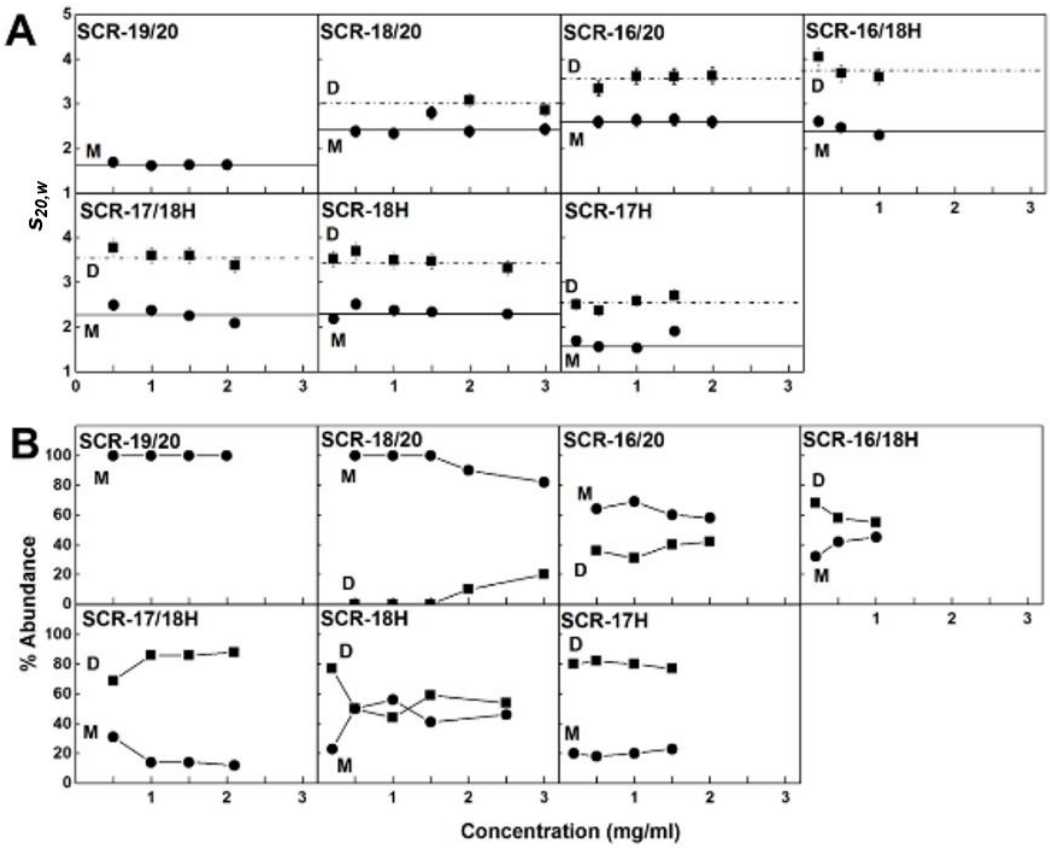


Figure 4.7. Concentration-dependence of the sedimentation data for each of the SCR fragments. *A*, The sedimentation coefficients $s_{20,w}$ for the monomer is denoted by filled circles, and for the dimer with filled squares. Statistical error bars are shown where visible.

B, The relative percentage of monomer and dimer present as a function of concentration in the *c(s)* analyses.

value (Table 4.1). The K_D for the SCR-18/20 dimer was estimated to be 585 ± 150 µM.

(iii) SCR-16/20 existed as a monomer and dimer with two partially merged peaks corresponding to the weak monomer-dimer equilibrium with the monomer at 2.6 S and dimer at 3.6 S (Figure 4.4, Figure 4.7 A) (Okemefuna *et al*., 2008). The relative percentage of monomer to dimer did not significantly change with concentration with approximately 60 % monomer to 40 % dimer (Figure 4.7 B). The apparent molecular masses ranged from 41 to 48 kDa for the monomer and 65 to 66 kDa for dimeric SCR-16/20 in agreement with the expected mass of 38 kDa for monomeric protein. The f/f_0 for SCR-16/20 monomer was 1.5 and for the dimer was 1.7, indicating that the dimer is more elongated than the monomeric protein (Table 4.1). For the SCR-16/20 monomer-dimer equilibrium, the K_D was estimated to be 31 ± 14 µM.

(iv) SCR-16/18H existed as a monomer and dimer with two distinct peaks apparent, even at low concentrations of 0.2 mg/ml (Figure 4.5 top panel). Using absorbance optics, 1 mg/ml and 0.5 mg/ml yielded $s_{20,w}$ values of 2.5 S and 3.6 S for the monomer and dimer, respectively. Interference gave $s_{20,w}$ values of 2.6 S and 2.3 S for the 0.5 and 1 mg/ml monomer and for the dimer, 4.0, and 3.6 S (Figure 4.6, 4.7A). The calculated apparent molecular mass from the *c(M)* analyses for the monomer gave a mass range of 21 to 32 kDa and for the dimer 40 to 55 kDa in good agreement with the expected mass of 29 kDa for monomer. SCR-16/18H existed as approximately 60 % dimer and 40 % monomer (Figure 4.7B). The f/f_o value was calculated to be 1.3 for the monomer and 1.4 for the dimer, indicating that the protein became slightly more elongated upon dimer formation. The K_D value for the SCR-16/18H dimer formation was estimated at 6 ± 5µM.

(v) For SCR-17/18H, absorbance and interference showed that it existed as monomer and dimer (Figure 4.5). For the monomer peak the $s_{20,w}$ value ranged from 2.1 S to 2.5 S and 3.4 S to 3.6 S for the dimer peak (Figure 4.7 A). Calculated apparent molecular masses were 21 to 24 kDa for the monomer and 44 to 49 kDa for the dimer agreeing with an expected mass of 22 kDa for the monomer. SCR-17/18H existed as 80 % dimer (Figure 4.7 B). The f/f_0 was calculated to be 1.3 for both monomer and

dimer indicating no change in elongation upon dimer formation. The K_D for SCR-17/18H dimer formation was estimated at 3 ± 1 µM.

(vi) SCR-18H also existed as two species with two peaks evident in the *c(s)* distribution (Figure 4.6). The first peak showed $s_{20,w}$ values which ranged from 2.2 S-2.5 S and the second peak had values of 3.5 S-4.0 S depending on concentration and optics used. The two peaks varied with concentration but were approximately 50 % in each (Figure 4.7 A,B). *c(M)* analyses gave molecular masses of 20 to 31 kDa for the first observed peak and 37 to 57 kDa for the second peak. Both of these values were larger than the expected size of 11 kDa for monomeric SCR-18H. The maximum sedimentation coefficient possible for a compact sphere (s_{sphere}) of mass 11 kDa was calculated from the Svedberg equation (Chapter Three, Equation 3.6) (Svedberg & Pederson, 1940). The theoretical sedimentation coefficient of SCR-18H was found to be 2.03 S. Therefore the experimental sedimentation coefficients calculated were too large to correspond to monomeric SCR-18H. Likewise for a mass of 22 kDa (dimer) the s_{sphere} was 3.2 S, with the experimental sedimentation coefficient of the second peak being too large to correspond to dimer. The variation in glycosylation of SCR-18H may have affected its sedimentation properties causing the larger than expected values or it may exist as a dimer and tetramer rather than monomer and dimer. From the AUC results the oligomeric state of SCR-18H was not clear; however, it does exist in two oligomeric states. The f/f_0 value was calculated to be 1.2 for the first peak and 1.3 for the second assuming that SCR-18H existed as monomer and dimer. The K_D value for SCR-18H oligomer formation was estimated to be 37 ± 27 µM.

(iv) Two peaks were also evident in the *c(s)* distribution for SCR-17H (Figure 4.6). The first peak had $s_{20,w}$ values of 1.5 S - 1.9 S with the second peak at 2.4 S - 2.7 S depending on concentration and the optics system used. SCR-17H exists with 80% dimer (Figure 4.7 A, B). Calculated apparent molecular masses ranged from 17 kDa to 23 kDa for monomeric SCR-17H and 30 kDa to 41 kDa for the dimeric form. The *c(M)* molecular weights were significantly larger than expected for the 10 kDa monomeric and 20 kDa dimeric SCR-17H. However, SCR-17H is also glycosylated which may give rise to the larger than expected mass calculations. The $s_{20,w}$ values were in the expected range for monomeric and dimeric SCR-17H which strongly suggested that it exists as a monomer and dimer. The f/f_0 value was calculated for monomeric and dimeric SCR-17H at 1.14 for the monomeric peak and 1.6 for the

dimeric peak indicating elongation upon dimer formation. The K_D value for the SCR-17H dimer formation was estimated at 5 ± 4 µM.

With the exception of SCR-19/20, all the SCR fragments *c(s)* analyses showed two distinct $s_{20,w}$ values corresponding to monomeric and dimeric protein. The single domains SCR-17H and SCR-18H and the SCR-17/18H fragment both oligomerize with SCR-17/18H showing the lowest K_D value. It was concluded that these two domains form the C-terminal CFH dimer site.

4.3.3 Analytical ultracentrifugation in 50 mM NaCl

A similar analysis was carried out on each of the SCR fragments in buffers containing 50 mM NaCl in order to act as a control for the 137 mM NaCl data and to investigate possible ionic strength effects.

For all of the SCR fragments with the exception of SCR-16/20, interference and absorbance optics were used, with interference only for SCR-16/20.

(i) The *c(s)* distribution for SCR-19/20 showed only one peak corresponding to 100 % monomer with an $s_{20,w}$ of 1.8 S and 1.7 S independent of concentration when measured by interference and absorbance optics, respectively (Figure 4.8, Figure 4.12 A, B). The apparent molecular mass ranged from 15 kDa to 18 kDa. The f/f_0 value was 1.04 showing that SCR-19/20 had a compact shape (Table 4.1).

(ii) SCR-18/20 showed one peak in the *c(s)* distribution with an $s_{20,w}$ value of 2.3 S by interference optics. A shoulder peak at concentrations above 1 mg/ml was observed by corresponding to the SCR-18/20 dimer with a $s_{20,w}$ value of 2.9 S. Absorbance optics showed a single broad peak which had an $s_{20,w}$ value of 2.2 S (Figure 4.8, 4.12 A). The monomer accounted for approximately 80 % of total protein (Figure 4.12 B). The monomer peaks had calculated molecular weights ranging from 27 to 30 kDa with the dimer peak at 34 kDa. The f/f_o ratio was 1.13 for monomeric protein and 1.6 for the dimeric protein which showed that the dimer is elongated with respect to the monomer (Table 4.1). The K_D value for SCR-18/20 dimer formation was estimated at 184 ± 125 µM.

(iii) SCR-16/20 showed two partially resolved peaks corresponding to monomer and dimer. The monomeric protein had an $s_{20,w}$ value of 3.1 S and 2.8 S for of 0.5 mg/ml and 2 mg/ml respectively while the dimeric protein had an $s_{20,w}$ of 4.2 S (Figure 4.9, Figure 4.12 A). The percentage of dimer increased with concentration

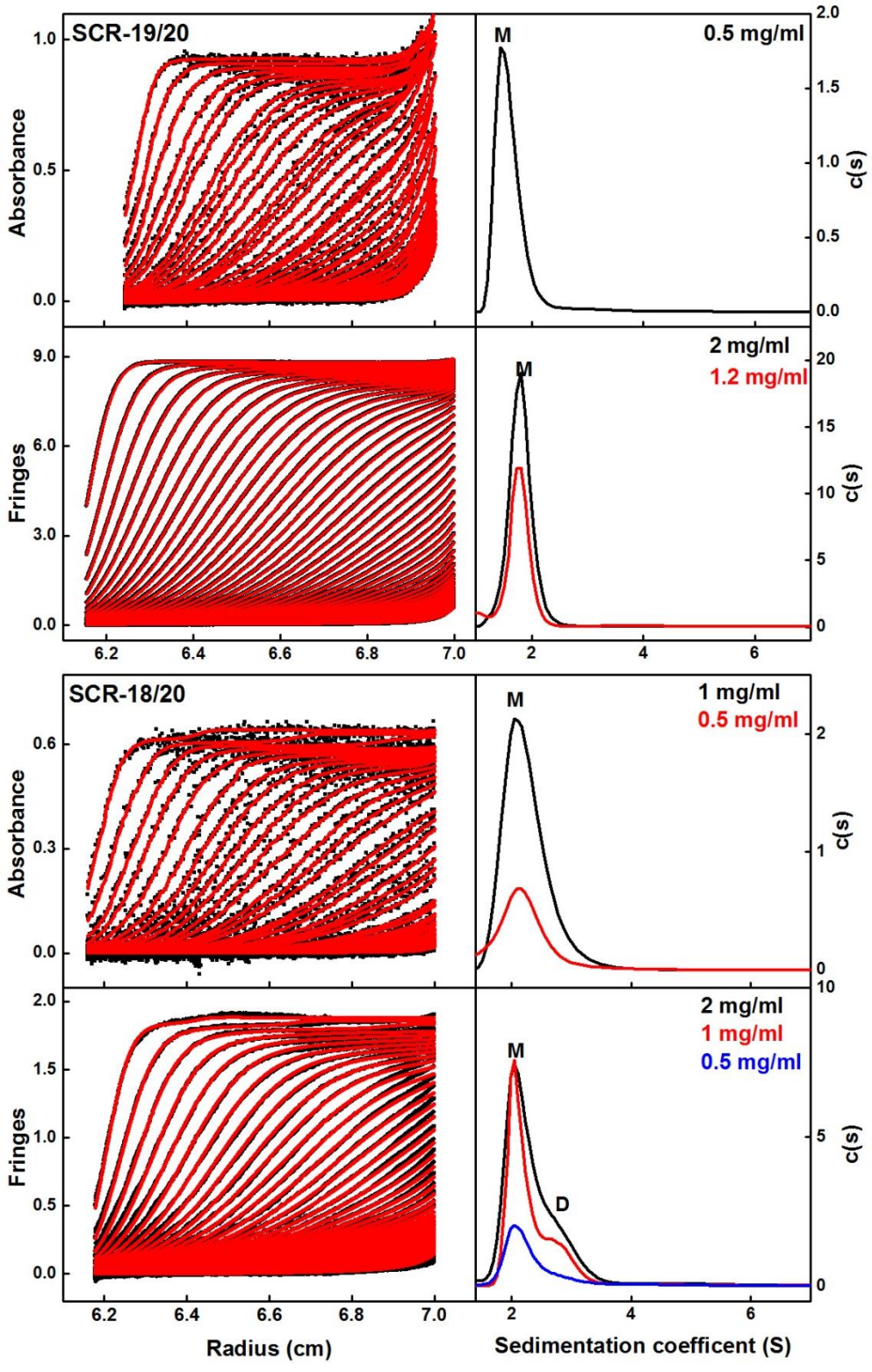


Figure 4.8 *c(s)* distribution of SCR-19/20 and SCR-18/20 in 50 mM NaCl. Shown in the left panel are the scans with the boundary fits are shown in red (absorbance, top; interference, bottom). In the right panel the *c(s)* distribution is shown over the concentrations series used. Monomer and dimer are denoted by M and D. The concentrations are shown in each panel, highlighted according to their colours.

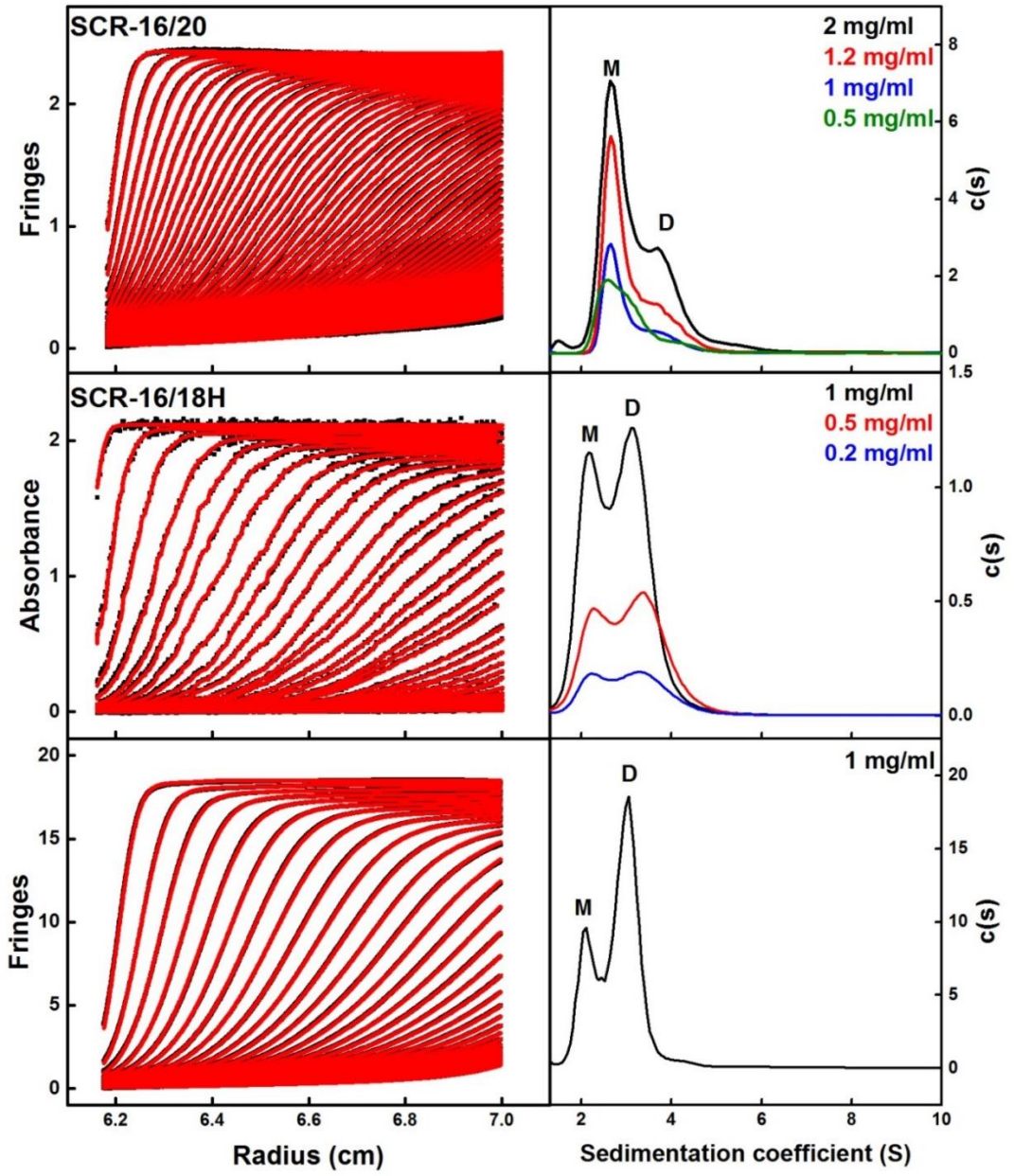


Figure 4.9. *c(s)* distribution of SCR-16/20 and SCR-16/18H in 50 mM NaCl. Shown in the left panel are the scans with the boundary fits are shown in red (Interference only for SCR-16/20; absorbance and interference for SCR-16/18H, top and bottom respectively). In the right panel the *c(s)* distribution is shown over the concentrations series used. Monomer and dimer are denoted by M and D respectively. The concentrations are shown in each panel, highlighted according to their colours in the *c(s)* distributions.

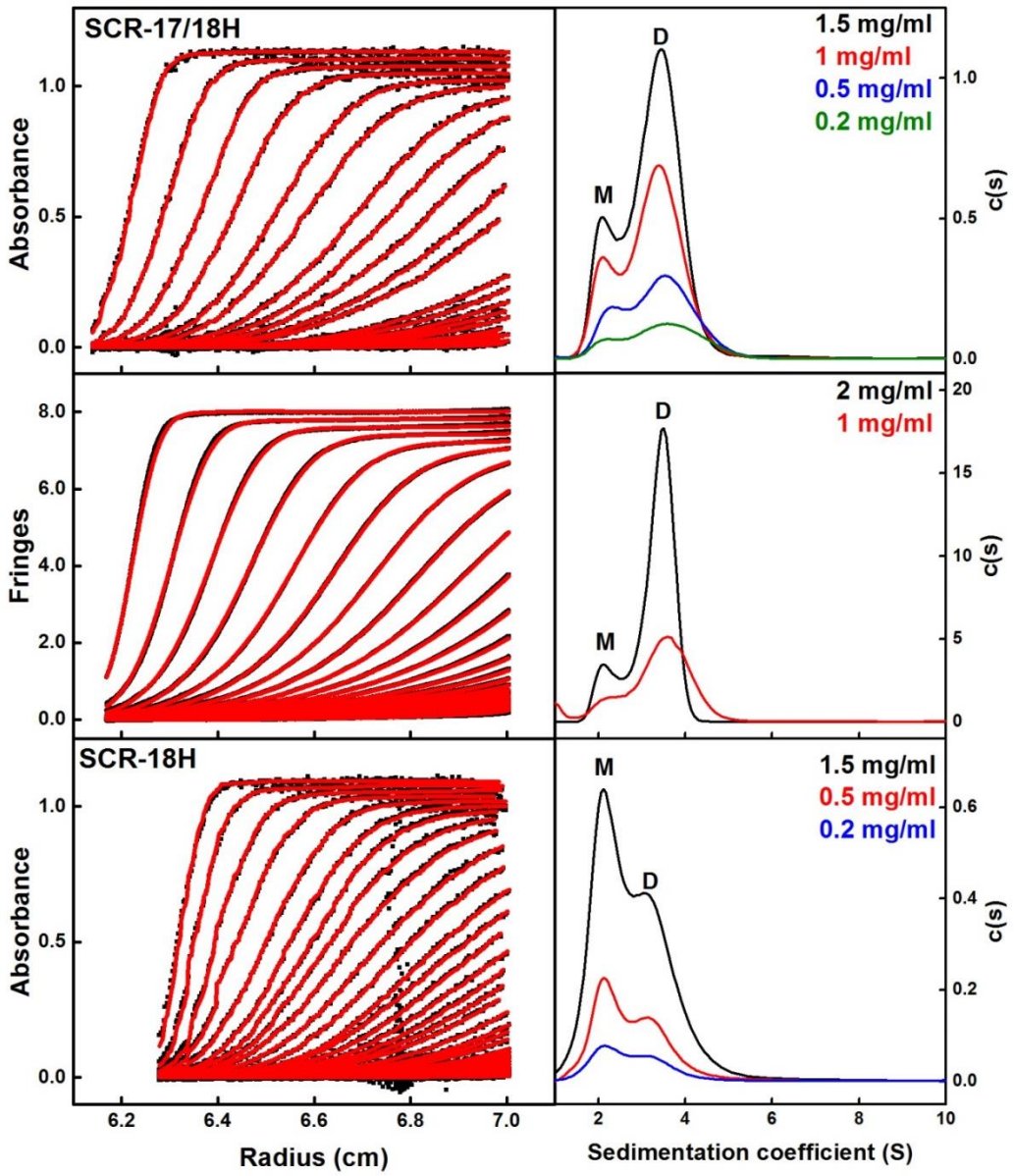


Figure 4.10 *c(s)* distribution of SCR-17/18H and SCR-18H in 50 mM NaCl. Shown in the left panel are the scans with the boundary fits are shown in red (absorbance, top; interference, bottom for SCR-17/18H and absorbance only for SCR-18H). In the right panel the *c(s)* distribution is shown over the concentrations series used. Monomer and dimer are denoted by M and D respectively. The concentrations are shown in each panel, highlighted according to their colours in the *c(s)* distributions.

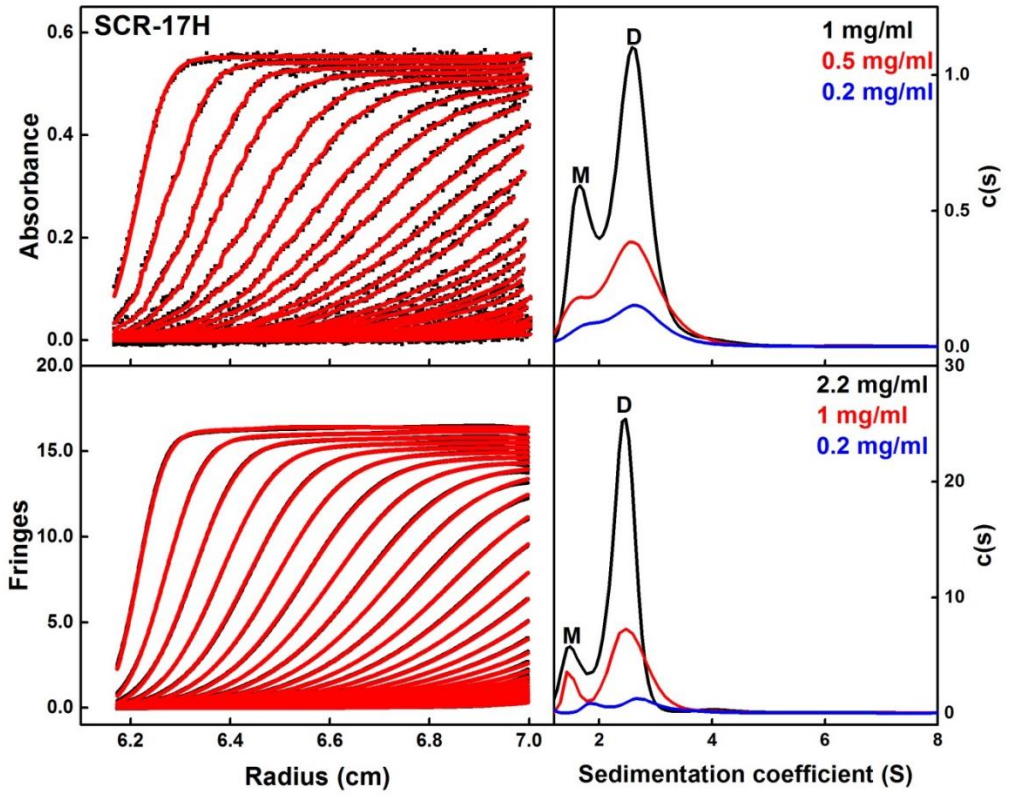


Figure 4.11. *c(s)* distribution of SCR-17H in 50 mM NaCl. Shown in the left panel are the scans with the boundary fits are shown in red (absorbance, top; interference, bottom). In the right panel the *c(s)* distribution is shown over the concentrations series used. Monomer and dimer are denoted by M and D respectively. The concentrations are shown in each panel, highlighted according to their colours in the *c(s)* distributions.

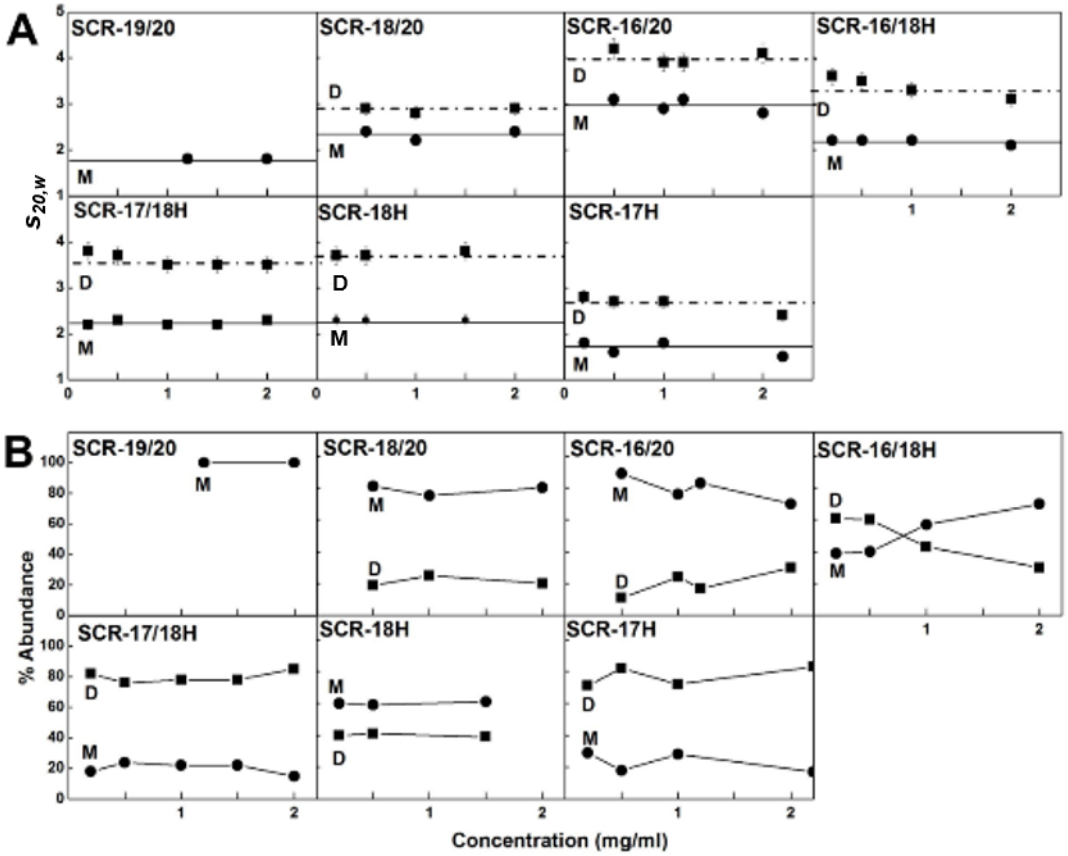


Figure 4.12 Concentration-dependence of the sedimentation data for each of the SCR fragments in 50 mM NaCl. *A*, The sedimentation coefficients $s_{20,w}$ for the monomer is denoted by filled circles, and for the dimer with filled squares. *B*, The relative percentage of monomer and dimer present in the *c(s)* analyses as a function of protein concentration. Statistical error bars are shown where visible.

from 11 % to 30 % (Figure 4.12 B). The calculated molecular weights ranged from 41 to 48 kDa for the monomer and 66 kDa for the dimer. The f/f_0 value was 1.3 for the SCR-16/20 monomer and 1.5 for the SCR-16/20 dimer showing an increase in elongation upon dimer formation. The K_D value for SCR-16/20 dimer formation was estimated to be 86 ± 20 µM.

(iv) SCR-16/18H had two peaks corresponding to monomer and dimer with the monomeric protein having an $s_{20,w}$ value of 2.2 S, and the dimeric protein at 3.5 S and 3.3 S for 0.2 mg/ml and 1 mg/ml (Figure 4.9, Figure 4.12 A). The percentage of dimer decreased at higher concentrations from 60 % to 30 % which was not expected and did not agree with the analysis in 137 mM NaCl (Figure 4.12 B). The calculated molecular weights ranged from 35 to 37 kDa for monomeric SCR-16/18H and 60 to 73 kDa for the dimeric protein. SCR-16/18H had a f/f_0 of 1.5 for monomeric protein and 1.5 for the dimeric protein which indicated no significant change in elongation upon dimer formation with respect to the monomer in 50 mM NaCl. The K_D of SCR-16/18H dimer formation was estimated at 10 ± 11 µM.

(v) SCR-17/18H showed two peaks in the *c(s)* plot corresponding to monomer and dimer. Monomeric protein had a $s_{20,w}$ of 2.2 S while dimeric had an $s_{20,w}$ value of 3.8 S and 3.5 S for 0.2 mg/ml and 2 mg/ml, respectively (Figure 4.10, Figure 4.12 A). The percentage of dimer was approximately 80 % and was unaffected by concentration (Figure 4.12 B). The molecular weights for monomeric SCR-17/18H ranged from 25 to 29 kDa and for dimeric from 46 to 60 kDa. The f/f_0 was 1.3 for both the monomeric and dimeric forms of SCR-17/18H which showed no change in elongation upon dimer formation. The K_D value for SCR-17/18H dimer formation was estimated as 2 ± 1 µM.

(vi) SCR-18H exhibited two peaks in the *c(s)* analyses. The first peak had an $s_{20,w}$ value of 2.3 S and the second peak had a $s_{20,w}$ of 3.7 S (Figure 4.10). The relative percentage remained unchanged with concentration (Figure 4.12 B). The calculated molecular weight for the first peak was 34 kDa and the second was 70 kDa. Similar to the 137 mM NaCl *c(s)* analysis the $s_{20,w}$ and the molecular weights calculated were larger than expected for a protein of 10 kDa indicating that SCR-18H may exist as dimer and tetramer or that the glycosylation may affect its sedimentation properties. The f/f_0 was calculated for both peaks as 1.3 and were calculated as dimer and tetramer (Table 4.1). The K_D value for SCR-18H dimer formation was estimated to be 59 ± 54 µM.

(vii) SCR-17H had two peaks in the *c(s)* distribution. The first peak had an $s_{20,w}$ value of 1.8 S and 1.5 S for 0.2 mg/ml and 2.2 mg/ml, respectively. The second peak SCR-17H had an $s_{20,w}$ value of 2.9 S and 2.5 S for 0.2 mg/ml and 2.2 mg/ml, respectively (Figure 4.11, 4.12 A). The relative percentages did not change significantly with concentration with SCR-17H existing as 70 % dimer. The calculated molecular weight for the first peak ranged from 17 to 23 kDa and from 34 to 47 kDa for peak two. Similar to SCR-18H, these values are larger than expected but this may be attributed to the additional bulk of the carbohydrates on SCR-17H. The f/f_0 were calculated for monomeric and dimeric SCR-17H as 1.16 and 1.02, respectively. The K_D value for SCR-17H dimer formation was calculated to be 5 ± 3 µM.

The results obtained in 50 mM NaCl buffer were in agreement with the results of the 137 mM NaCl buffer study. SCR-17 and SCR-18 were observed to contain the dimer site with the individual domains and the two domain SCR-17/18H fragment forming dimers. Neither the relative percentages of monomer and dimer nor the K_D value estimations changed significantly in 50 mM NaCl with respect to 137 mM NaCl. This suggests that electrostatic interactions are not responsible for dimer formation as at lower ionic strength it would be expected that the amount of dimer would increase.

4.3.4. Small angle X-ray scattering in 137 mM NaCl

SAXS is a low resolution technique which yields size and shape information about macromolecules in solution (Perkins *et al.*, 2011). SAXS data with good signal to noise was obtained for the seven SCR fragments in 137 mM NaCl buffer. Guinier analysis were carried out on the subtracted curves to calculate the R_G which is a measure of the overall elongation of the molecule (Figure 4.13) and the radius of the cross section (R_{XS}) (Figure 4.14). Successful Guinier analyses for each of the SCR fragments were carried out within the fit limits of the $Q.R_G$ and $Q.R_{XS}$ values. The R_G/R_O ratio is a measurement of the elongation of the protein with respect to a sphere. R_O is the R_G of a perfect sphere which has a volume equal to the volume of the hydrated protein (R_O). A typical globular protein will have a R_G/R_O ratio of 1.28 (Perkins *et al*., 1986). The information obtained from a solution scattering experiment represents an average of the sizes and shapes present in the sample and it is therefore not possible to distinguish between monomer and dimer. However with the prior information obtained from AUC about the percentage of dimer present at each concentration it is

possible to extract information on the change in size and shape associated with dimer formation.

4.3.4.1 Guinier analyses in 137 mM NaCl

Guinier analysis was carried out for each of the seven SCR fragments.

(i) For SCR-19/20, the R_G value was 2.4 ± 0.1 nm and the R_G/R_O ratio was 1.7 indicating that it was elongated with respect to a globular protein of the same size. SCR-19/20 had an R_{XS} value of 0.86 ± 0.04 nm. Neither the R_G nor the R_{XS} changed significantly with respect to concentration as expected for the monomeric SCR-19/20 fragment (Figure 4.15 A, B).

(ii) For SCR-18/20 the R_G value increased with concentration from 3.2 ± 0.1 nm to 3.6 ± 0.03 nm (Figure 4.15 A). The R_G/R_O increased from 2.0 to 2.2 which showed that SCR-18/20 had an elongated shape which is further elongated upon dimer SCR-18/20 showed an increase in R_{XS} from 0.9 ± 0.04 to 1.2 ± 0.06 nm (Figure 4.19 B)

(iii) The R_G value for SCR-16/20 showed an increase in the R_G with concentration from 4.7 ± 0.06 nm to 6.0 ± 0.1 nm (Figure 4.15 A). The R_G/R_O increased from 2.4 to 3.1, with SCR-16/20 becoming more elongated upon dimer formation at formation at higher concentrations. The R_{XS} value increased with concentration from 0.87 ± 0.04 nm to 1.2 ± 0.06 nm (Figure 4.15 B). higher concentrations. The R_{XS} value increased with concentration from 1.6 ± 0.08 nm to 2.04 ± 0.1 nm (Figure 4.15 B). For SCR-16/18H, SCR-17/18H, SCR-18H and SCR-17H the R_G did not change significantly with concentration (Figure 4.15 A, B).

(iv) SCR-16/18H had an R_G value of 4.7 ± 0.07 nm and an R_G/R_O of 2.7. SCR-16/20 and SCR-16/18H had similar results indicating they have a similar shape. The R_{XS} value for SCR-16/18H was 1.6 ± 0.08 nm which did not change significantly.

(v) SCR-17/18H had an R_G value of 2.8 ± 0.1 nm and a R_G/R_O of 1.7. SCR-17/18H had a R_{XS} value which increased at higher concentrations from 0.51 ± 0.03 nm to 0.73 ± 0.04 nm (Figure 4.15B). The R_G value was larger than that of the two domain SCR-19/20 due to its larger size and glycosylation, however it had a similar degree of elongation (Figure 4.15 A, B).

(vi) SCR-18H had an R_G value of 2.3 ± 0.03 nm and a R_G/R_O of 1.4. It had a R_{XS} value of 2.3 ± 0.1 nm. The R_G value was similar to that of SCR-19/20 with two

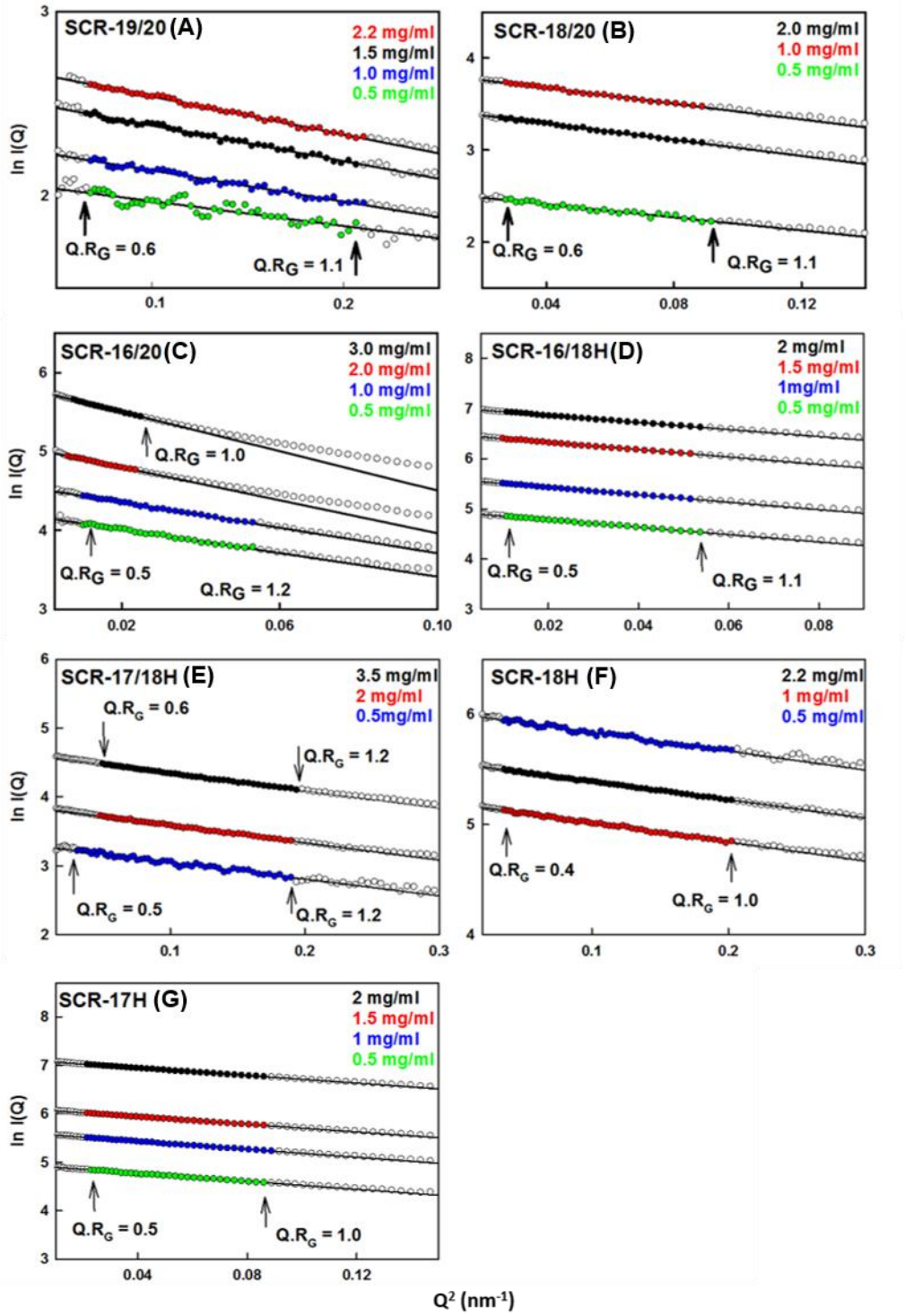


Figure 4.13. Guinier R_G analyses for each of the SCR fragments in 137 mM NaCl. Coloured circles represent the *I(Q)* values used to determine the R_G with the line representing the linear fit to this region. The $Q.R_G$ limits of the linear fit are shown. The *Q* ranges used for the R_G fits were 0.24-0.44 nm^{-1} for SCR-19/20 (A), 0.14-0.3 nm^{-1} for SCR-18/20 (B), 0.10-0.22 nm^{-1} SCR-16/20 (C) monomer and SCR-16/18H (D), 0.10-0.14 nm^{-1} SCR-16/20 dimer, 0.20-0.44 nm^{-1} for SCR-17/18H (E), 0.17-0.45 nm^{-1} for SCR-18H (F) and 0.14-0.28 nm^{-1} for SCR-17H (G).

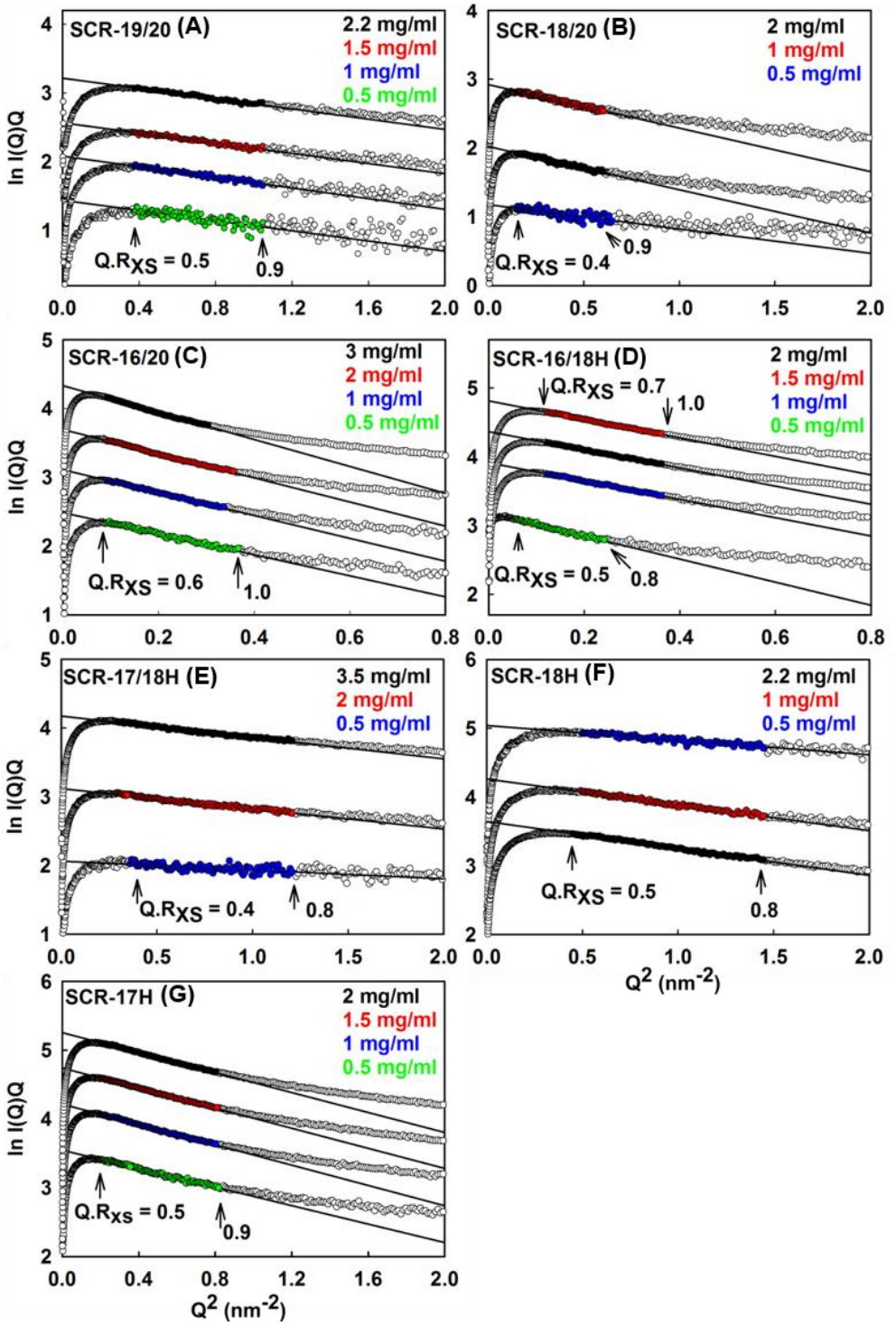


Figure 4.14. Guinier R_{XS} analyses for each of the SCR fragments in 137 mM NaCl. Coloured circles represent the *I(Q)* values used to determine the R_{XS} with the line representing the linear fit to this region. The $Q.R_{XS}$ limits of the linear fit are shown. the ranges for the R_{XS} fits were 0.55-1.02 nm^{-1} for SCR-19/20 (A), 0.39-0.79 nm^{-1} for SCR-18/20 (B), 0.28-0.6 nm^{-1} SCR-16/20 (C) monomer,0.28-0.50 nm^{-1} for SCR-16/20 dimer, 0.22-0.49 nm^{-1} for SCR-16/18H (D) monomer, 0.35-0.60 nm^{-1} for SCR-16/18H dimer, 0.53-1.09 nm^{-1} for SCR-17/18H (E), 0.45-0.88 nm^{-1} for SCR-18H (F) and 0.7-1.19 nm^{-1} for SCR-17H.

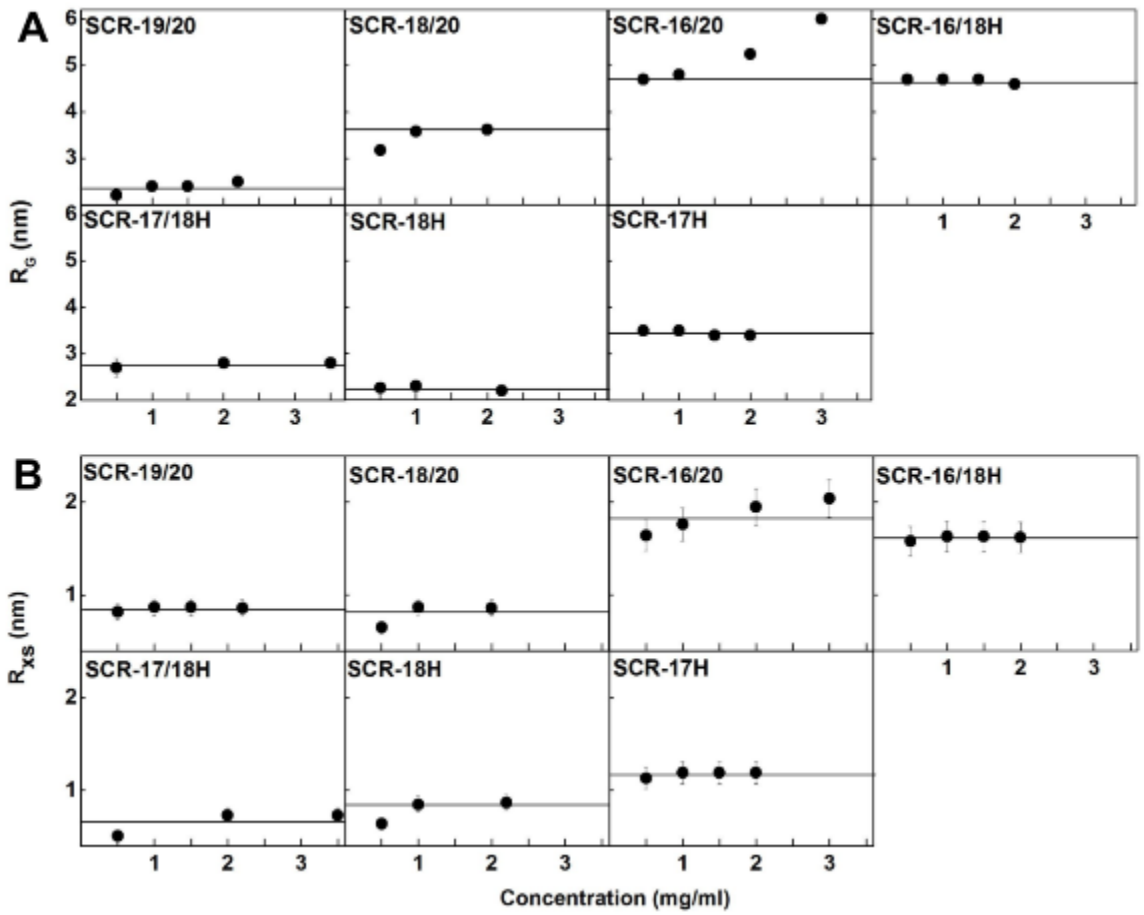


Figure 4.15 Concentration dependence of the R_G and R_{XS} value of the SCR fragments in 137 mM NaCl. *A*, The X-ray R_G values are shown as a function of concentration. The line denotes the mean value. The error shown is the standard deviation of the mean (where visible).

B, The X-ray R_{XS} values are shown plotted as a function of concentration. The line denotes the mean value with the error shown as the standard deviation of the mean.

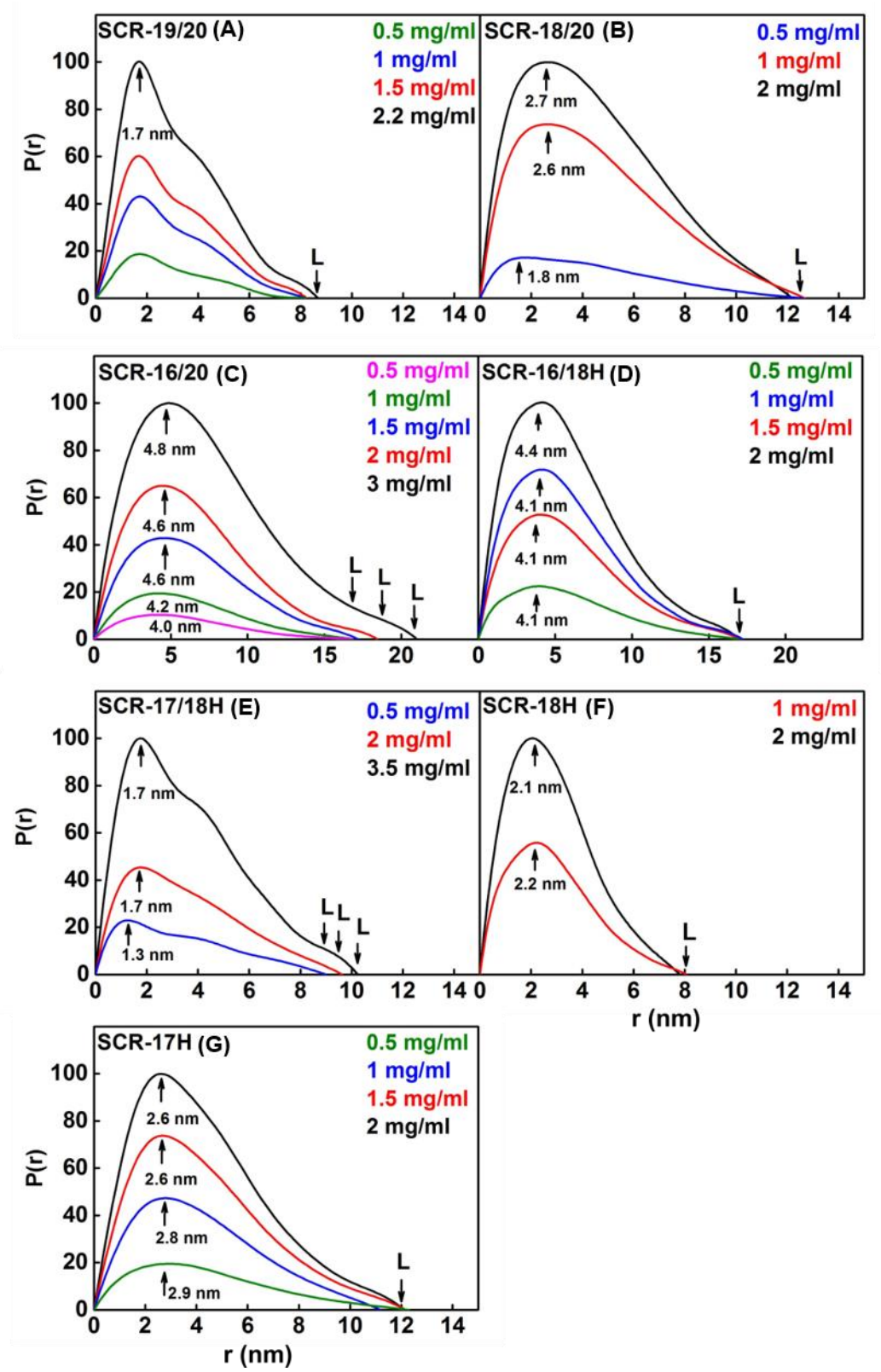


Figure 4.16. Distance distribution function *P(r)* analyses for each of the seven SCR fragments. The arrow under each peak represents M, the most frequent distance within the protein, and L represents the maximum observed dimension of the SCR fragment.

domains which indicated that it may exist as monomer and dimer (rather than dimer and tetramer) (Figure 4.15). *c(s)* analysis shows that SCR-18H exists in two oligomeric states with approximately 50 % each. It is therefore possible that SCR-18H also formed a dimer through an end-to-end interaction and the effect on the R_G is less dramatic than with SCR-17H due to the lower amount of dimer present in SCR-18H. The R_G/R_O was not supportive of this theory as it suggests that the protein had a more compact shape which would be better explained by the formation of the dimer through a side-by-side interaction. However, glycosylation may have also caused discrepancies within the SAXS analysis and so the oligomerisation state of SCR-18H was deemed ambiguous.

(vii) SCR-17H had an R_G value of 3.5 ± 0.02 nm and an R_G/R_O ratio of 2.2. The R_{XS} value was 3.5 ± 0.2 nm. The R_G value was significantly larger than that of SCR-19/20 and SCR-17/18H both of which have an additional domain (Figure 14.5 A). SCR-17H existed primarily as a dimer explaining the larger than expected R_G value for a single domain protein. This large R_G value and the R_G/R_O also indicated that the dimer of SCR-17H formed by an end-to-end interaction where it would appear much larger. The level of glycosylation of SCR-17H may also contribute to the larger than expected R_G value.

4.3.4.2 Distance distribution *P(r)* in 137 mM NaCl

The distance distribution curve *P(r)* was calculated by an indirect Fourier transform using GNOM (Figure 4.16) (Svergun, 1992). The real space R_G values were in good agreement with those calculated by Guinier analysis. The *P(r)* provides a measure of the maximum dimension or length of the macromolecule observed (L) and information of the interatomic distance most frequently observed within the molecule (M).

(i) SCR-19/20 had a maximum length of 8.2 ± 0.4 nm and an M value of 1.7 ± 0.09 nm, neither changed significantly with concentration as expected for the monomeric protein.

(ii) SCR-18/20 had maximum length of 12.2 ± 0.6 nm with an increase in the observed M value from 1.8 ± 0.1 nm to 2.7 ± 0.1 nm with increasing concentration. This indicated that there was no increase in the maximum length of the protein with dimer formation. However the error was high at higher concentrations and they

percentage of dimer formed as determined by AUC is low, so a dramatic increase was not expected.

(iii) The maximum length observed for SCR-16/20 increased with concentration from 16.8 ± 0.8 nm to 21 ± 1 nm. The M values also increased from 4.0 ± 0.2 nm to 4.8 ± 0.2 nm. This indicated that the SCR-16/20 dimer had a more elongated structure than the monomer as the percentage of dimer increases with concentration so does the maximum length.

(iv) SCR-16/18H had maximum length of 17 ± 0.8 nm, with an M value of 4.1 ± 0.2 nm with no observed concentration dependent changes.

(v) SCR-17/18H increased from 9 ± 0.5 nm to 10.2 ± 0.5 nm with M values of 1.7 ± 0.09 nm for each. This slight increase may be as a result of small changes upon dimer formation, however the R_G value does not change and the error is large and so it was concluded that no major changes are observed upon dimer formation of SCR-17/18H.

(vi) SCR-18H had a maximum length of 8 ± 0.4 nm with an M value of 2.1 ± 0.1 nm which did not change with concentration. This value was similar to that of SCR-19/20 and indicated that the SCR-18H dimer was formed by an end-to-end interaction.

(vii) SCR-17H had an L value of 12 ± 0.6 nm with an M values of 2.6 ± 0.1 nm. This maximum length was similar to that of SCR-18/20 and indicated the SCR-17H dimer was formed by an end-to-end interaction.

The SAXS results confirmed that SCR-19/20 is monomeric. Both SCR-16/20 and SCR-18/20 showed an increase in size with increasing dimer formation. SCR- 16/18H and SCR-17/18H showed no significant increases, which is expected as their dimer forming properties were not concentration dependent at the concentrations used here (as shown by AUC). Similarly SCR-18H and SCR-17H did not show significant changes. However their larger sizes indicated that both of the fragments dimerise through end to end interactions rather than side-by-side. The SAXS results also suggest that both SCR-17H and SCR-18H exist as monomer and dimer as much large size parameters would be expected if dimer and tetramer were present.

4.3.5. Small angle X-ray scattering in 50 mM NaCl

SAXS experiments on each of the SCR domains were also carried out in 50 mM NaCl to analyse the effects of a lower salt concentration. A similar analysis to the samples in physiological NaCl was carried out. There were no significant changes as a function of NaCl concentration.

4.3.5.1 Guinier analyses in 50 mM NaCl

The Guinier R_G analysis fits are shown in Figure 4.17 while the R_{XS} fits are shown in Figure 4.18.

(i) SCR-19/20 had an R_G value of 2.2 ± 0.1 nm and an R_G/R_O ratio of 1.6 indicating that it is relatively elongated. The R_{XS} value was 0.8 ± 0.04 nm. No changes were observed with concentration (Figure 4.19).

(ii) SCR-18/20 showed an increase in the R_G with concentration from 3.4 ± 0.1 to 3.6 ± 0.03 nm. The R_G/R_O increased from 2.1 to 2.2 indicating that similar to 137 mM NaCl, SCR-18/20 became more elongated upon dimer formation. SCR-18/20 showed an increase in R_{XS} value from 0.9 ± 0.04 nm to 1.2 ± 0.06 nm (Figure 4.19).

(iii) SCR-16/20 showed an increase in the R_G value with increasing concentration from 4.8 ± 0.1, to 6.1 ± 0.1 nm, with the R_G/R_O increasing from 2.5 to 3.2. This indicated that similar to the 137 mM NaCl study, the SCR-16/20 dimer was more elongated than the monomeric fragment. The R_{XS} value increased from 1.7 ± 0.08 to 2 ± 0.1 nm (Figure 4.19).

(iv) SCR-16/18H had an R_G value of 4.5 ± 0.1 nm and a R_G/R_O of 2.5. SCR-16/18H had an R_{XS} value of 1.6 ± 0.1 nm. No changes with concentration were observed (Figure 4.19).

(v) SCR-17/18H had an R_G value of 2.8 ± 0.1 nm and an R_G/R_O ratio of 1.4. SCR-17/18H had an R_{XS} value of 0.7 ± 0.04 nm with no concentration effects observed (Figure 4.19).

(vi) SCR-18H had an R_G value of 2.2 ± 0.02 nm and an $R_G/R_{O\,\text{ratio}}$ of 1.5. SCR-18H had an R_{XS} value of 0.9 ± 0.05 nm. No concentration effects were observed (Figure 4.19).

(vii) SCR-17H had an R_G value 3.2 ± 0.01 nm, and an R_G/R_O of 2.0. SCR-17H had an R_{XS} of 1.2 ± 0.06 nm. No changes with concentration were observed for SCR-17H (Figure 4.19).

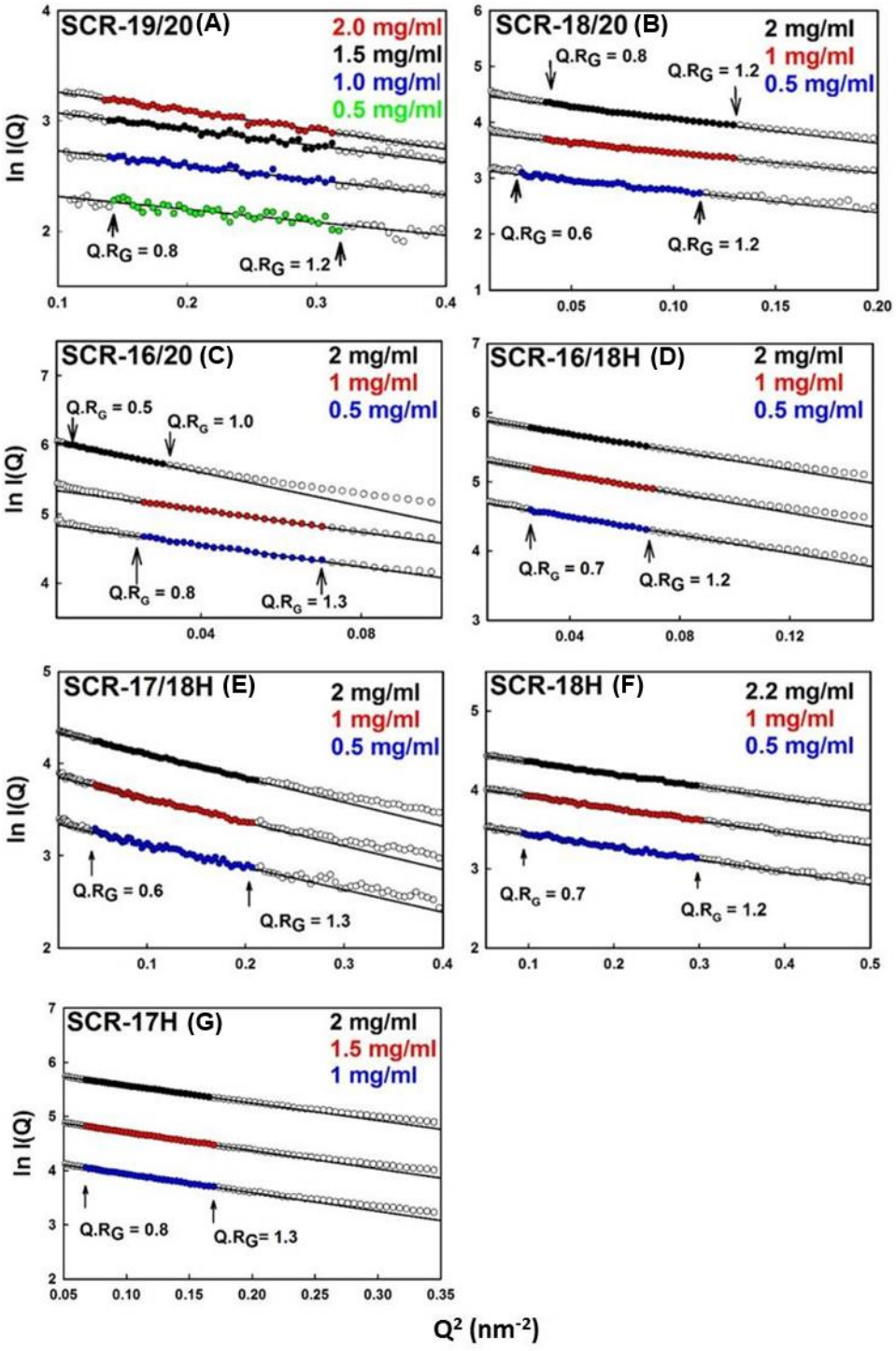


Figure 4.17. Guinier R_G analyses for each of the SCR fragments in 50 mM NaCl. Coloured circles represent the *I(Q)* values used to determine the R_G with the line representing the linear fit to this region. The $Q.R_G$ limits of the linear fit are shown. The *Q* ranges used for the R_G fits were 0.24-0.44 nm^{-1} for SCR-19/20 (A), 0.14-0.3 nm^{-1} for SCR-18/20 (B), 0.10-0.22 nm^{-1} SCR-16/20 (C) monomer and SCR-16/18H (D), 0.10-0.14 nm^{-1} SCR-16/20 dimer, 0.20-0.44 nm^{-1} for SCR-17/18H (E), 0.17-0.45 nm^{-1} for SCR-18H (F) and 0.14-0.28 nm^{-1} for SCR-17H (G).

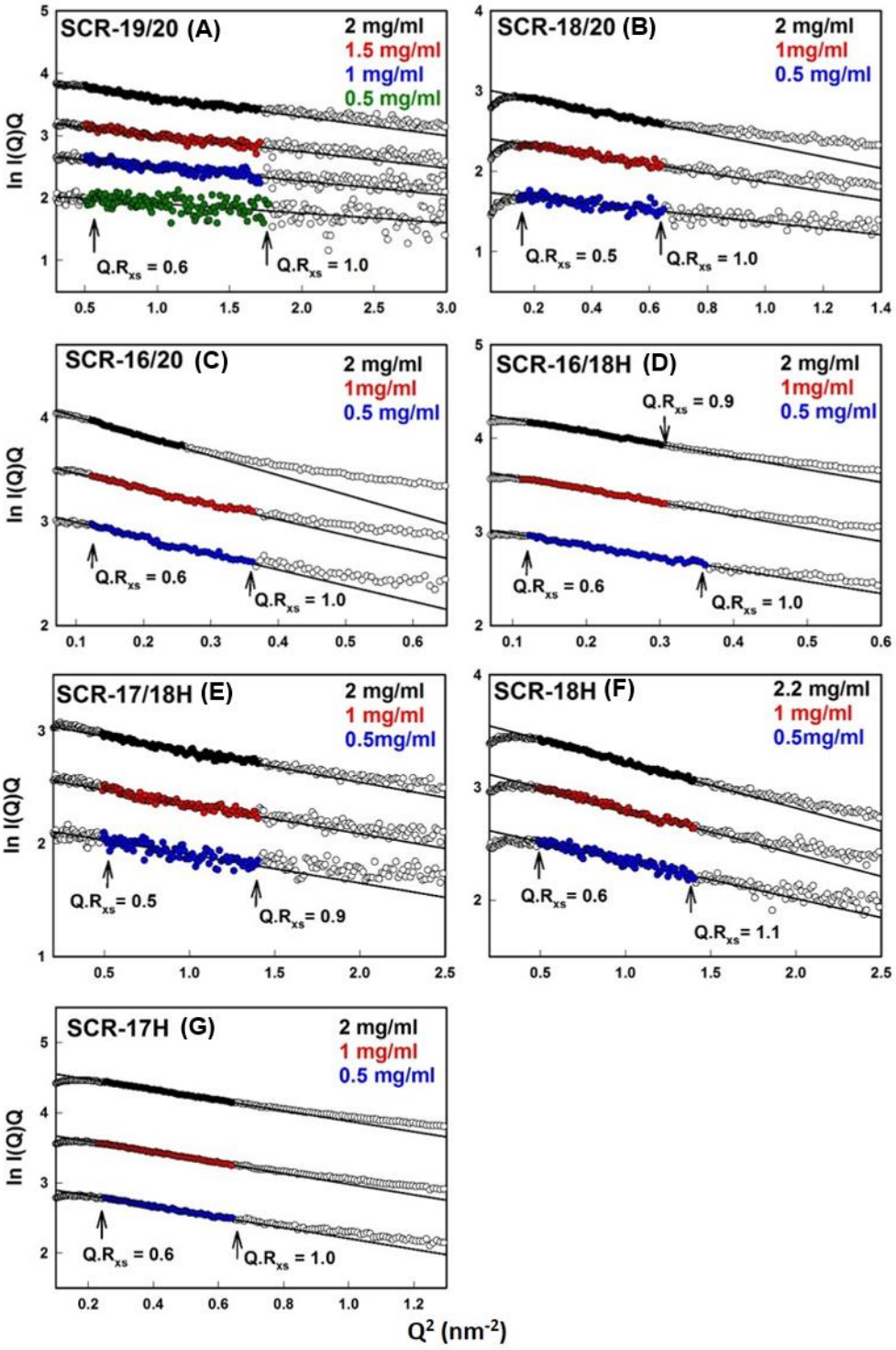


Figure 4.18. Guinier R_{XS} analyses for each of the SCR fragments in 50 mM NaCl. Coloured circles represent the *I(Q)* values used to determine the R_{XS} value with the line representing the linear fit to this region. The $Q.R_{XS}$ limits of the linear fit are shown. The ranges for the R_{XS} fits were 0.55-1.02 nm^{-1} for SCR-19/20, 0.39-0.79 nm^{-1} for SCR-18/20, 0.28-0.6 nm^{-1} SCR-16/20 monomer,0.28-0.50 nm^{-1} for SCR-16/20 dimer, 0.22-0.49 nm^{-1} for SCR-16/18H monomer, 0.35-0.60 nm^{-1} for SCR-16/18H dimer, 0.53-1.09 nm^{-1} for SCR-17/18H, 0.45-0.88 nm^{-1} for SCR-18H and 0.7-1.19 nm^{-1} for SCR-17H.

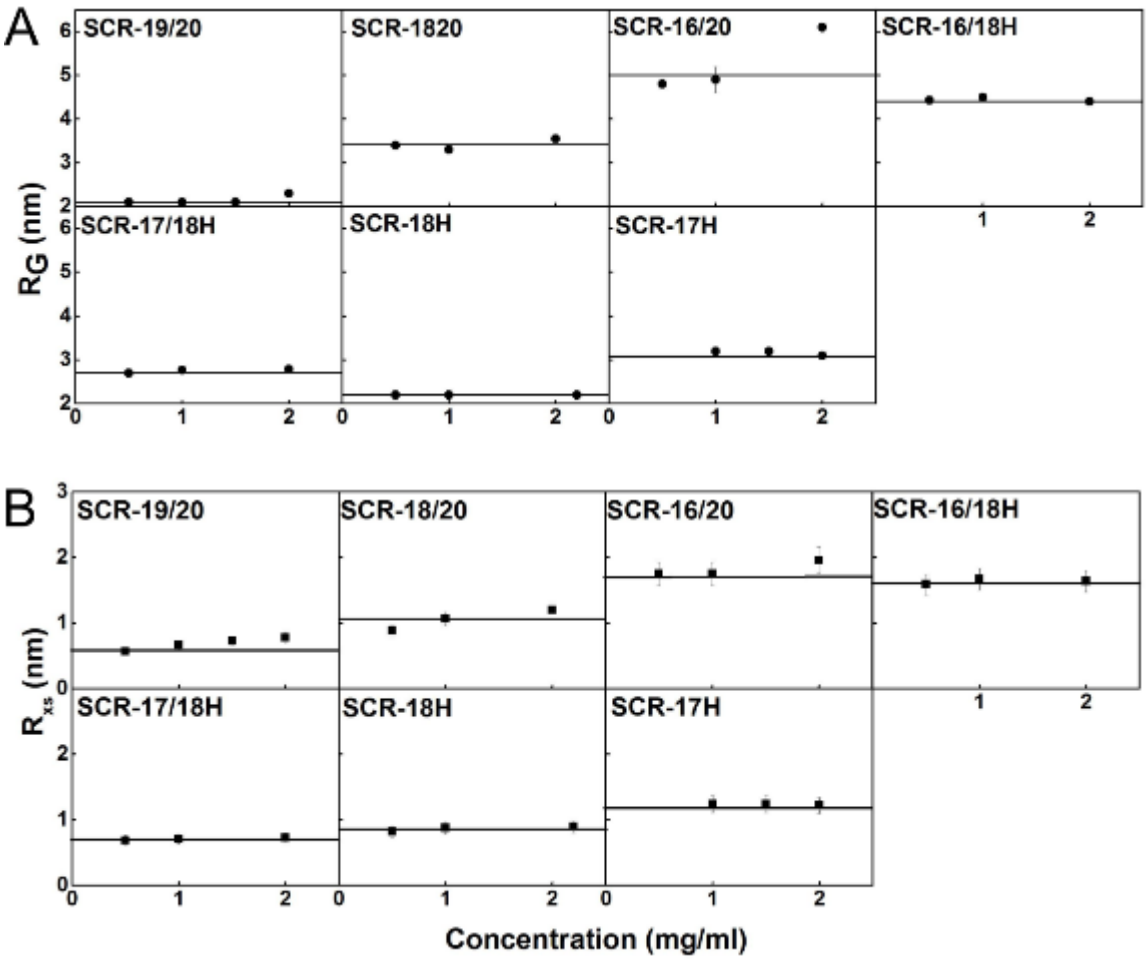


Figure 4.19. Concentration-dependence of the R_G and R_{XS} value of the SCR fragments in 50 mM NaCl. *A*, The X-ray R_G values are shown as a function of concentration. The line denotes the mean value with standard deviation shown as the error. *B*, The X-ray R_{XS} values are shown plotted against the concentration. The line denotes the mean value with the standard deviation as error.

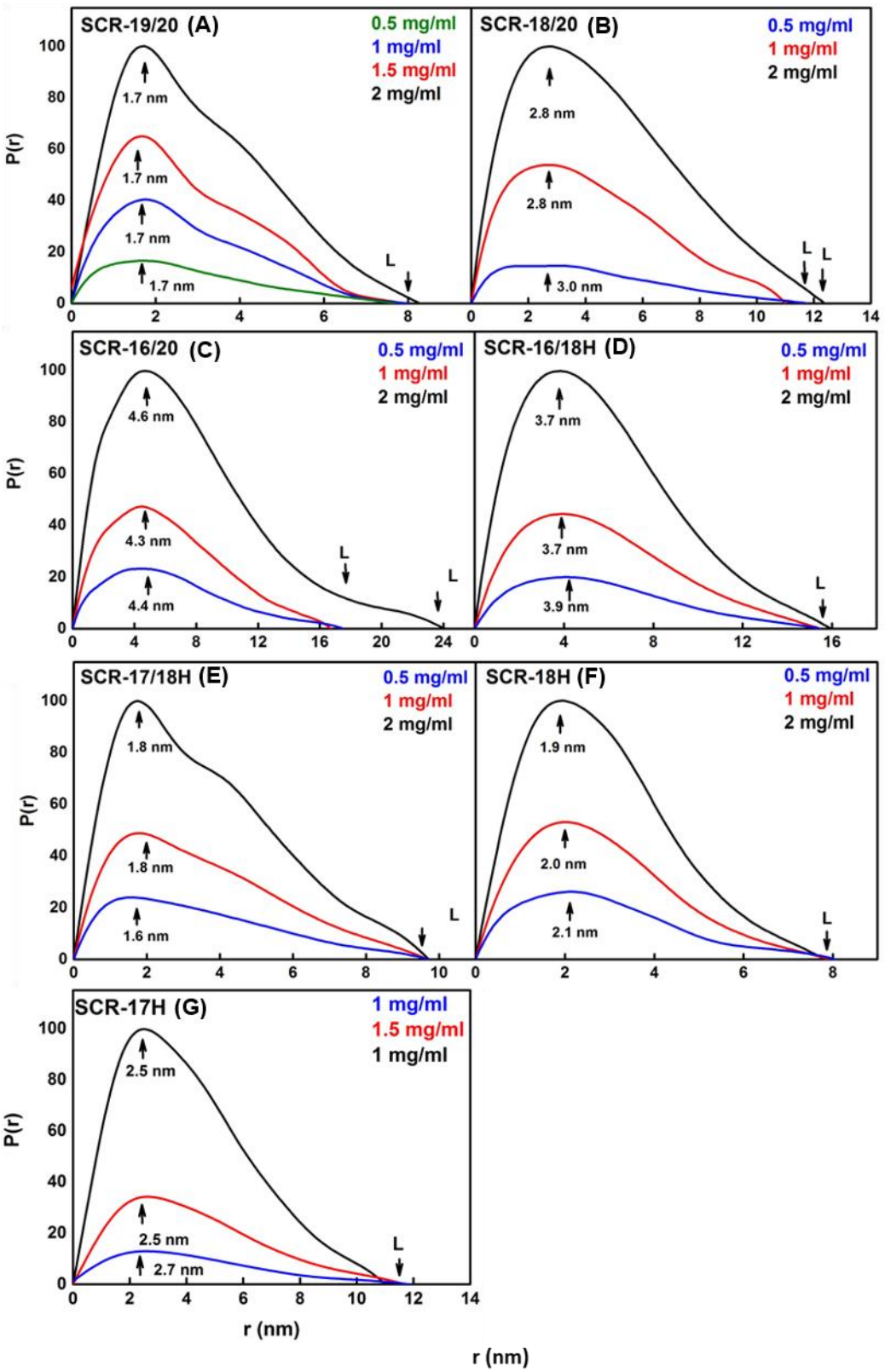


Figure 4.20 Distance distribution function *P(r)* analyses for each of the seven SCR fragments in 50 mM NaCl. The arrow under each peak represents M, the most frequent distance within the protein, and L represents the maximum observed dimension of the SCR fragment.

4.3.5.2 Distance distribution *P(r)* in 50 mM NaCl

The distance distribution *P(r)* curves for each of the SCR domains were calculated (Figure 4.20). Reciprocal and real space R_G values were in good agreement.

(i) The maximum length of SCR-19/20 increased with increasing concentration, from 7.7 ± 0.4 nm to 8.2 ± 0.4 nm which may have been due to small amounts of aggregates present in the sample. The M value did not change at 1.7 ± 0.1 nm.

(ii) SCR-18/20 had a maximum length, which increased with concentration, from 11.7 ± 0.6 nm to 12.3 ± 0.6 nm. The observed M value also increased from 1.8 ± 0.1 nm to 2.7 ± 0.1 nm with increasing concentration.

(iii) SCR-16/20 showed a significant increase in maximum dimension with respect to an increase in concentration. It increased from 17.5 ± 0.8 nm, to 24 ± 1.2 nm. The M also increased from 4.4 ± 0.2 nm, to 4.6 ± 0.2 nm.

(iv) SCR-16/18H had a maximum length of 15.9 ± 0.8 nm which is lower than that for 137 mM NaCl however the error was high and so statistically the reduction in size was insignificant. The M value was 3.8 ± 0.2 nm. No changes with concentration were observed.

(v) SCR-17/18H had a maximum length of 9.7 ± 0.5 nm similar to that observed for SCR-17/18H in 137 mM NaCl buffer. The M value was 1.7 ± 0.1 nm. No concentration affects were observed for SCR-17/18H in 50 mM NaCl buffer.

(vi) SCR-18H had a maximum length of 7.9 ± 0.4 nm, similar to that in 137 mM NaCl and an M value of 2 ± 0.1 nm neither of which were affected by concentration.

(vii) SCR-17H showed an increase in its maximum length with a decrease in concentration from 11.9 ± 0.6 nm to 11 ± 0.6 nm with a high error on both values. The M value did not significantly change with concentration and was calculated to be 2.6 ± 0.1 nm.

No significant size differences were observed for the seven SCR fragments in 50 mM NaCl buffers with respect to 137 mM NaCl buffers. This is in agreement with the AUC results which showed that lowering the ionic strength of the buffer did not significantly alter the fragments dimer forming capabilities.

Table 4.1 Comparison of analytical ultracentrifugation parameters of SCR domains in 137 mM and 50 mM NaCl

Domains*	$s^0_{20,w}$ (S) 137 mM	$s^0_{20,w}$ (S) 50 mM	f/f_0 137 mM	f/f_0 50 mM	R_G (nm) 137 mM	R_G/R_O 137 mM	R_G (nm) 50 mM	R_G/R_O 50 mM	L (nm) 137 mM	L (nm) 50 mM	K_D (μM) 137 mM
SCR-19/20 (monomer)	1.64	1.8	1.1	1.04	2.4 ± 0.09	1.7	2.2 ± 0.1	1.6	8.2 ± 0.4	8.2 ±0.4	na
SCR-18/20 monomer	2.4	2.4	1.1	1.1	3.2 ± 0.06	2	3.4 ± 0.1	2	12.2 ± 0.6	11.7 ±0.6	na
SCR-18/20 dimer	2.9	2.9	1.6	1.6	3.62 ± 0.03	2.2	3.6 0.03	2.2	12.7 ±0.3	12.3 ±0.6	585 ± 150
SCR-16/20 monomer	2.59	3.1	1.5	1.3	4.7 ± 0.06	2.4	4.8 ± 0.1	2.5	16.8 ±0.8	17.5 ±0.8	na
SCR-16/20 dimer	3.63	4.2	1.7	1.5	6 ± 0.1	3.1	6.1 ± 0.1	3.2	21 ± 1	24 ±0.2	31 ± 14
SCR-16/18H monomer	2.47	2.2	1.3	1.5	4.7 ± 0.04	2.7	4.4 ± 0.1	2.5	17 ± 0.8	15.5 ± 0.8	na
SCR-16/18H dimer	3.68	3.5	1.4	1.5	4.7 ± 0.06	2.7	4.4 ± 0.01	2.5	17 ± 0.5	15.9 ± 0.8	6 ± 5
SCR-17/18H monomer	2.49	2.2	1.3	1.3	2.7 ± 0.2	1.7	2.7 ± 0.1	1.4	9 ± 0.5	9.7 ± 0.5	na
SCR-17/18H dimer	3.77	3.8	1.3	1.3	2.8 ± 0.01	1.7	2.8 ± 0.01	1.4	10.2 ± 0.5	9.7 ± 0.2	3 ± 1
SCR-18H monomer	2.51	2.32	1.2	1.3	2.2 ± 0.03	1.4	2.2 ± 0.03	1.5	8 ± 0.4	7.9 ± 0.4	na
SCR-18H dimer	3.49	3.69	1.3	1.3	2.2 ± 0.01	1.4	2.2 ± 0.02	1.5	8 ± 0.1	7.9 ± 0.3	37 ± 27
SCR-17H monomer	1.56	1.82	1.2	1.7	3.5 ± 0.03	2.2	3.2 ± 0.01	2	12 ± 0.6	11.9 ± 0.6	na
SCR-17H dimer	2.37	2.86	1.6	1.6	3.4 ± 0.01	2.2	3.2 ± 0.1	2	12 ± 0.1	11 ± 0.6	5 ± 4

* For SAXS results monomer is taken as the lowest concentration (i.e. lowest dimer percentage) and dimer is taken as the highest concentration

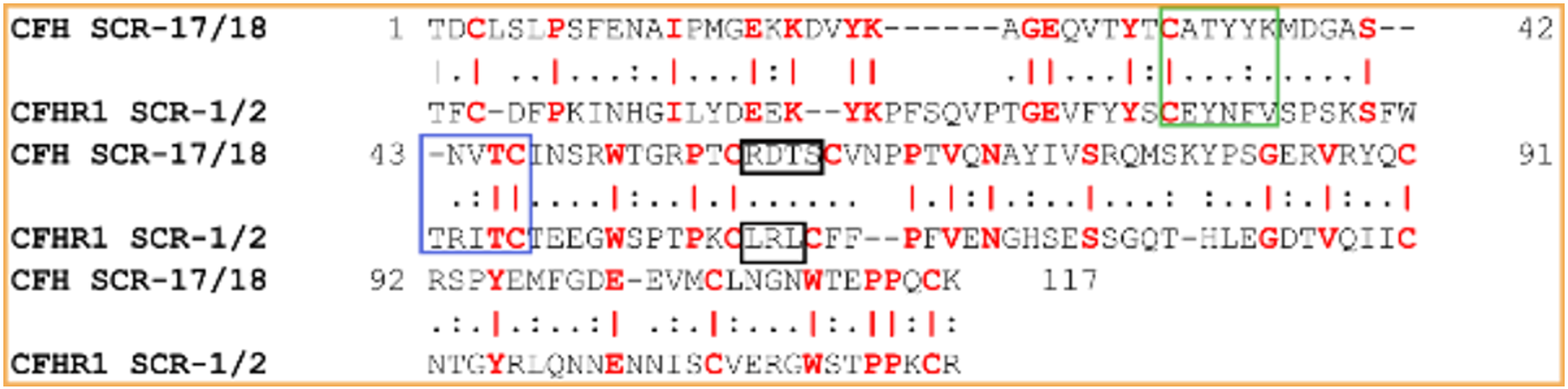


Figure 4.21 The sequence alignment between CFHR1 SCR-1/2 domains and SCR-17/18H from CFH. The green box highlights the loop region with the blue box highlighting the β-4 strand, which were used to align the structure of SCR-1/2 with SCR-17/18. The linker regions of SCR-1/2 and SCR-17/18 are highlighted in the black box. Conserved residues are shown in red

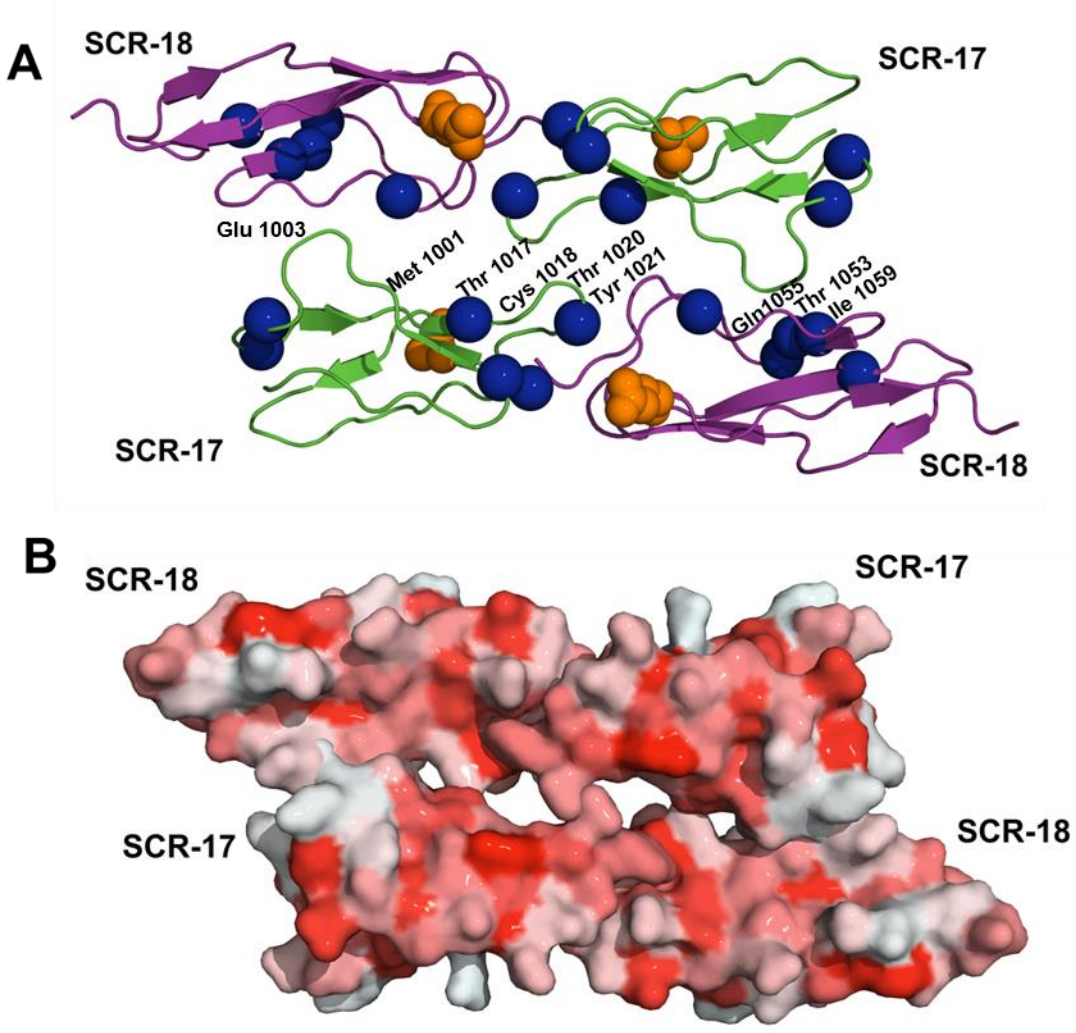


Figure 4.22 A model of the SCR-17/18 dimer interface. A model constructed of SCR-17/18 dimer of complement factor H using the CFHR1 SCR-1/2 crystal structure as a template. *A*, The ribbon structure showing the antiparallel arrangement of the domains. Blue spheres represent the disease-associated mutations of SCR-17/18. The glycosylation sites are highlighted in orange. *B,* A hydrophobicity map of the SCR-17/18 structure illustrating a weak hydrophobic interface with white non-hydrophobic and red hydrophobic.

4.3.6 Modelling of the SCR-17/18 dimer

In order to understand how SCR-17/18 interacts to form a dimer, a model was constructed using existing information from other SCR domain containing complement proteins. The SCR-17/18 domain structure was composed of the SCR-18 crystal structure and a homology model of SCR-17. Using this, a model of the SCR17/18H dimer interface was constructed using the crystal structure of CFHR1 as a template. The SCR-1/2 of CFHR1 forms a dimer by an anti-parallel side-by-side hydrophobic interaction with SCR-1 interacting with the loop region of the second SCR-1 molecule. A multiple sequence alignment was carried out between CFHR1 SCR1/2 and SCR-17/18 of CFH. Using the EMBOS water local sequence alignment tool a sequence similarity of 41 % was obtained (Figure 4.21) indicating a relatively low sequence homology between the two fragments however the four cysteine residues are conserved in each domain with CFHR1 SCR-1/2 linked by three amino acids while SCR-17/18 has a linker region of four residues. Despite the low sequence homology, the SCR domains have a high degree of structural similarity with each SCR domain generally containing six to eight β strands with a hypervariable loop between β2 and β3 (Saunders *et al*., 2007; Rodriguez *et al.*, 2014). SCR-17 of CFH was aligned with SCR-1 through the central β strand (β-4) (Figure 4.21) and through the loop region which contains the Tyr 39 residue essential for CFHR1 dimerisation (Figure 4.22). This model revealed a weak hydrophobic interface with disease-associated mutations clustered at the interface. The two glycosylation sites on SCR-17 and SCR-18 are distinct from the interface region meaning that they will not interfere with dimerisation (shown in Figure 4.22 A). The residues identified directly within the interface region were Met1001, Glu1003, Lys1004, Lys1005, Val1007, Thr1015, Thr1017, Cys1018, Ala 1019, Thr1020 and Tyr1021 from SCR-17 and Val1049, Asn1050, Pro1051, Pro1052, Thr1053, Gln1055 and Ile1059 from SCR-18. The hydrophobicity of the interface was analysed which showed that a hydrophobic interaction existed between the loop region of SCR-17 and SCR-17 of the adjacent anti-parallel molecule along with weak hydrophobic interactions throughout the interface (Figure 4.22 B). This model gives us an indication as to the interface and structure of the SCR-17/18 dimer. The AUC and SAXS results were in support of this model.

4.4 Discussion

In both physiological and low NaCl buffers our results suggest that SCR-17/18 of CFH contains its C-terminal dimerisation site. This provides us with a possible explanation for the aHUS-associated mutations located in SCR-17 and SCR-18 and also suggests a novel mechanism by which CFH could protect host cells during an inflammatory event. SEC and AUC show that the SCR fragments that contain domains 17 and 18 dimerise, and that only SCR-19/20 exists solely as a monomer. SCR-16/18H exists with less dimer than SCR-17/18H, suggesting that SCR-16 is not involved in dimer formation. SCR-17/18H and SCR-17H both exist with a higher percentage of dimer than SCR-18H or SCR-18/20, indicating that SCR-17 contains the main self-association site. However, the estimated K_D values show that SCR-17/18H forms the strongest dimer indicating that both SCR-17 and SCR-18 are involved. No significant differences were observed between the 137 mM and 50 mM NaCl concentrations indicating that electrostatic interactions are not responsible for SCR-16/20 dimer formation.

Figure 4.23 shows a cartoon illustration of each of the seven SCR fragments in their dimeric forms (apart from SCR-19/20). This cartoon hypothesises as to how each of the fragments dimerise based on the size and shape parameters obtained from the AUC and SAXS experiments. From our AUC experiments, SCR-16/20, SCR-16/18H and SCR-17/18H all have similar shapes with respect to elongation as observed from the f/f_0 values. This suggests that SCR-16/20 may have a bent back structure in which its five SCR domains are not fully elongated (Okemefuna *et al*., 2008; Okemefuna *et al*., 2009b). The similarity of the monomeric three domain SCR-16/18H to the monomeric two domain SCR-17/18H indicates that SCR-16/18H is not fully extended and may also possess a bent back structure. The crystal structure of SCR-18/20 shows a bent-back arrangement at its SCR-18/19 linker region, suggesting that the bent-back conformation of the C-terminal SCR-16/20 arises through the SCR-18/19 linker region (Morgan *et al.*, 2012). Our results suggest that although this bent back arrangement of SCR-18/20 may indeed exist, the solution state of SCR-16/18 also exists in a similar bent configuration. The elongated nature of the SCR-17/18H fragment indicates that it is in an extended conformation and hence the bend in SCR-16/18H is likely to arise from the four amino acid linkers of SCR-16 and SCR-17. SCR-18H has a similar $s_{20,w}$ to SCR-17/18H, which indicates that it may exist as a dimer and tetramer rather than

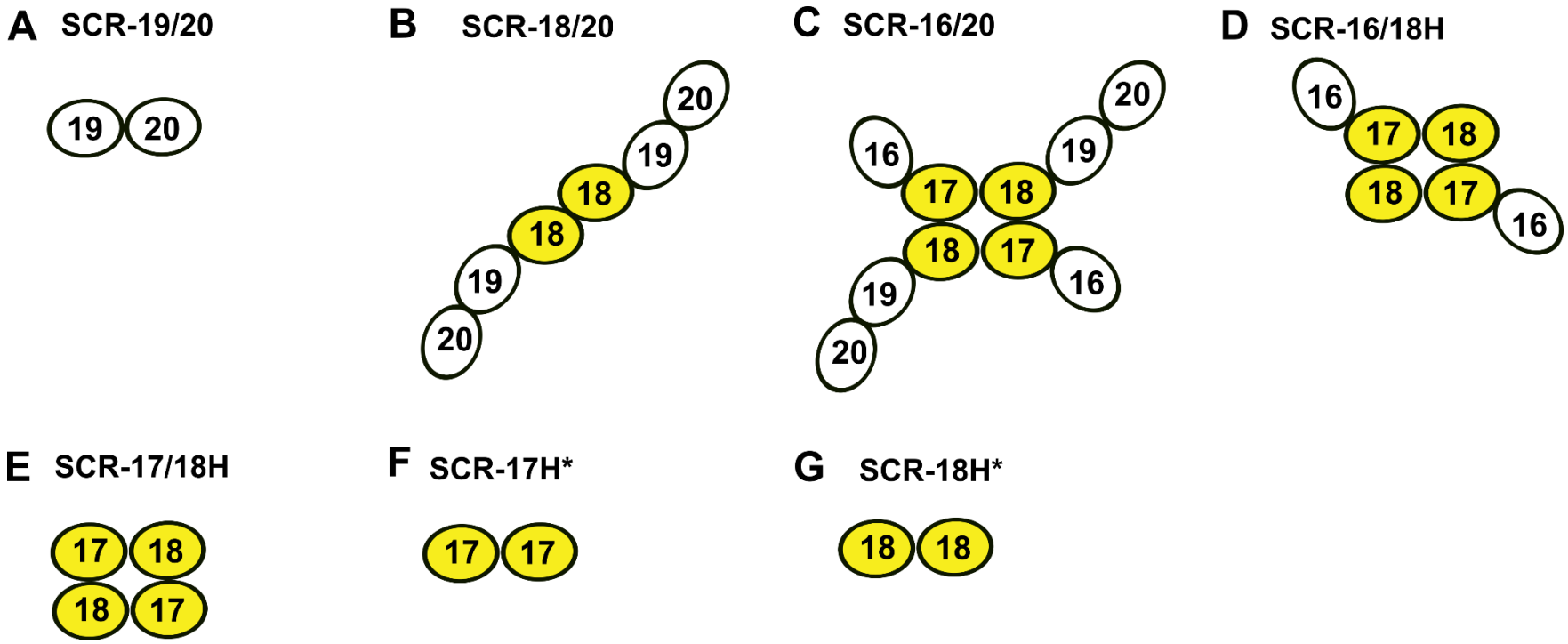


Figure 4.23 A schematic cartoon summary showing the shape of the SCR fragments as suggested from the obtained AUC and SAXS parameters. *A*, Monomeric SCR-19/20 in an elongated arrangement. *B*, SCR-18/20 dimer which is elongated with respect to its monomeric form which has a bent back arrangement. *C*, The possible shape of the SCR-16/20 dimer which is more elongated than its monomeric form. Monoemric SCR-16/20 has a bent back arrangement. *D*, Dimeric SCR-16/18H which is slightly elongated in its dimeric form in comparison to its monomer which is slightly bent between SCR-16/20. *E*, The SCR-17/18H dimer formed by an anti-parallel side-by-side arrangement where each of the monomers are relatively elongated. *F*, The possible dimer formed by two SCR-17H molecules. *G*, The possible dimer formed by two SCR-18H molecules. *The results obtained for both SCR-17H and SCR-18H by AUC and SAXS were difficult to analyse and so it is therefore difficult to hypothesise as to how their respective dimers are formed

monomer and dimer; however its glycosylation may also account for its larger than expected size and so it is difficult to confirm its oligomeric state. SCR-17/18H and SCR-19/20 also show different sedimentation properties, with SCR-17/18H larger and more elongated than SCR-19/20 despite the fact that both contain two SCR domains. This is most likely to be due to the two glycosylation sites on SCR-17/18H, which add a significant amount of bulk to the molecule, and the presence of the His-tag on SCR-17/18H. Similarly, SCR-17H has similar sedimentation properties to SCR-19/20, which again may be accounted for by glycosylation effects. The f/f_0 values indicate that the C-terminal dimer is formed by a side-by-side interaction of SCR-17/18H. The f/f_0 values for SCR-16/20, SCR-18/20, SCR-16/18H, SCR-17H and SCR-18H shows that they are more elongated in their dimer states. However, SCR-17/18H remains unchanged, strongly suggesting that it is formed by a side-by-side interaction. This also suggests that the dimer may form by an anti-parallel interaction which would account for the increase in elongation for all dimer forming fragments apart from SCR-17/18H in which both domains are involved in forming the dimer interface. Our work used a non-His tagged form of SCR-16/20 while in the Okemefuna *et al*., 2008 study a His-tagged variant was used. Both studies show similar results indicating that the His-tag has no effect on dimerisation.

From our AUC experiments it was possible to estimate the K_D value for dimer formation for each of the SCR fragments, the error on the K_D is high, such that the values obtained can only be used as an indication of the strength of the interaction. Little significant changes in the K_D are observed in 50 mM NaCl buffer with respect to 137 mM NaCl buffer confirming that while the method has large errors it is reproducible. This also confirmed that NaCl concentration had no effect on dimer formation, again suggesting that electrostatic interactions are not involved. The K_D values suggested that SCR-17 is the main dimer site. SCR-17/18H forms the strongest dimer (3μM), with SCR-17H only slightly weaker (5 μM). SCR-18H alone is significantly weaker (37 μM). This result suggests that SCR-17H is the main dimer site in the C-terminal of CFH but that SCR-18H also plays a role in strengthening the interaction between the two molecules.

Our SAXS data are consistent with the AUC results, showing that SCR-19/20 is monomeric, and that SCR-16/20 and SCR-18/20 became more elongated with dimer formation. For the remaining fragments, the percentage of monomer and dimer did not

significantly change with concentration as calculated by AUC, and hence information about shape or size changes associated with dimer formation is limited. However by comparing the size and shape of each SCR fragment with monomeric SCR-19/20 and monomeric SCR-18/20 (at low concentrations), and factoring in the amount of dimer present in the sample, it is possible to elucidate some information about the size and shape of the fragments. The size parameters obtained by SAXS for the SCR-18H monomer-dimer mixture has a similar size to SCR-19/20. SCR-18H exists with approximately 50% dimer and hence the similarity to the SCR-19/20 monomer suggests that it forms dimers through an end-to-end interaction. This size similarity also suggests that SCR-18H does exist as monomer and dimer rather than dimer and tetramer. SCR-17/18H is also similar to SCR-19/20, with larger size parameters (due to the carbohydrates) but with a similar degree of elongation that is consistent with side-by-side dimerisation. SCR-17H has similar size parameters to the SCR-18/20 monomer, which indicates that its dimer formation is not a simple side-by-side interaction. SCR-16/18H is much larger than the SCR-18/20 monomer indicating the dimeric form of SCR-16/18H is more elongated than a monomeric three domain protein. This is in agreement with our AUC results. The lowest concentration of SCR-16/20 (*i.e.* the highest percentage of monomer) has similar size parameters to SCR-16/18H indicating that the SCR-16/20 monomer is not fully elongated. The increase in size with concentration (and therefore the amount of dimer present) indicates that it increases in length upon dimer formation.

From SAXS it is often possible to obtain low to medium resolution models of biomolecules in solution. The resultant *I(Q)* scattering curve obtained from a SAXS experiment represents an average of the scattering molecules. It is therefore difficult to obtain models for a mixture, unless structural information is known about each of the individual components. With a lack of structural information about the SCR-16/20 dimer, it was therefore not possible to produce models from our SAXS data using traditional methods. In order to investigate how the CFH C-terminal dimer is formed we used the CFHR1 SCR-1/2 crystal structure as a template to construct a model of the SCR-17/18 dimer. Sequence homology between the SCR-1/2 of CFHR1 and SCR-17/18 of CFH is relatively low with the SCR-1/2 linker containing three amino acid residues while SCR-17/18 linker contains four. Additionally, SCR-1 of CFHR1 contains an alpha helix which is not observed in the SCR-17 homology model.

Nonetheless, the β strand arrangement is similar with the hypervariable loop in a similar orientation. The three essential residues (Tyr 34, Ser 36 and Tyr 39) for CFHR1 dimer formation were not conserved in SCR-17/18 which contains threonine residues at these positions. An anti-parallel arrangement agreed with our AUC and SAXS data which showed that, with the exception of SCR-17/18H, each of the dimer forming SCR fragments increase in elongation upon dimer formation. SCR-16/20 and SCR-16/18H becoming more elongated upon dimer formation while SCR-17/18H remains the same. The SCR-16/18H increase is small; however the AUC and SAXS data indicates that it forms a folded back structure which would account for the absence of large size changes. Both SCR-17H, SCR-18H and SCR-18/20H form dimers showing that SCR-17H and SCR-18H alone can dimerise. The dimerisation of SCR-17 alone agrees with the anti-parallel arrangement whereby the loop regions are in contact (Figure 4.22 A). The AUC and SAXS data indicate that it forms a dimer through an end to end interaction. The oligomeric state of SCR-18 is ambiguous as is its conformation and hence a definitive conclusions is difficult. It is possible however that the SCR-18 oligomerisation is non-specific, and results from the removal of SCR-17 exposing its hypervariable loop which would allow it to form a similar dimer to SCR-17. Additionally, SCR-17 and SCR-17/18 exist with a higher percentage of dimer suggesting SCR-17 is the main site. The model is validated by the distinct positions of the carbohydrates away from the interface region and by the presence of disease-associated mutations at or near the interface. The interface observed on our constructed model is weakly hydrophobic which is also validated by our results and shows that the interaction is not affected by NaCl and unlikely to be an electrostatic interaction (Figure 4.22 B). Val1007, Asn1050 and Ile 1059 are all directly located within the interface region. Mutations of these residues are associated with aHUS (Rodriguez *et al.,* 2014). SCR-17 disease-associated mutations are Asn997Thr, Val1007Cys, Ala1010Thr, Thr1017Ile, Tyr1020Phe and Cys1043Arg and Asn1050Tyr, Tyr1058His, Ile1059Thr, Val1060Ala/Leu and Gln1076Glu in SCR-18. Some of these mutations are of hydrophobic residues which may impact on the formation of the SCR-17/18 dimer.

Full length CFH forms weak dimers with an estimated 15 % dimer at CFH serum concentrations with self-association sites at SCR-6/8 and SCR-17/18 (Fernando *et al*., 2007; Okemefuna *et al.,* 2008, Nan *et al*., 2008; Chapter Four). Dimerisation is

observed in other members of the CFH gene family. The CFHR proteins belong to the same gene family as CFH (Diaz-Gullen *et al.*, 1999). CFHR1, CFHR2 and CFHR5 all exist in a dimeric form mediated by their SCR-1/2 N-terminal domains. Dimerisation is thought to strengthen their interactions on cell surfaces when competing with CFH for binding (de Jorge *et al*., 2013). CFH functions to protect host cells from destruction through its C-terminal interactions with C3b and host cell surfaces mediated by the SCR-19 and SCR-20 domains. The location of the dimer site in SCR-17/18 may provide CFH with a mechanism through which it can concentrate on host surfaces during an inflammatory response. The dimer site is independent to the C3b and surface binding sites in the C-terminal, suggesting that the binding site will still be available when CFH is bound to a surface allowing additional CFH molecules to be recruited eliciting further protection of the surface. SCR-19 and SCR-20 will remain free to bind the host surface and C3d/C3b which has been deposited. This would elicit further protection to the host cell from immune attack. The importance of C-terminal dimerisation is demonstrated by the fact that mutations are located in this region (Chapter Two; Figure 2.6). From our model these mutations are located at or close to the dimer interface and will most likely effect dimer formation. We present here a novel mechanism by which CFH protects host cell surfaces. Further studies are needed to experimentally investigate the effect of the aHUS mutations in SCR-17/18 on dimer formation which will further shed light on the importance of CFH dimerisation.

Chapter Five

SCR-19/20 contains the C-terminal complement Factor H binding site for pentameric C-reactive protein

5.1 Introduction

The alternative pathway of complement is activated by the spontaneous hydrolysis of C3 which ultimately results in the formation of the fluid phase C3 convertase. This C3 convertase cleaves C3 to C3a (an anaphylatoxin) and C3b. C3b binds to exposed cell surfaces through its thioester domain (TED) targeting them for immune destruction. Complement Factor H (CFH) regulates the alternative pathway of complement activation. CFH binds to C3b which is bound to a non-activator surface. It recognises host cells (*i.e*. non-activator cells) through interactions with glycosaminoglycan on their surfaces. CFH acts as a cofactor for Factor I, promoting the conversion of C3b to inactive iC3b, and preventing further C3 convertase formation and therefore complement activation.

CFH is a 154 kDa glycoprotein composed of 20 short complement regulator (SCR) domains. Each domain contains approximately 60 amino acids with the domains linked by regions containing three to eight residues which vary in flexibility and orientation (Perkins *et al.,* 2002; Soares & Barlow, 2005). The CFH C-terminal SCR-19/20 domains bind C3d and cell surfaces (Blackmore *et al*., 1998; Jokiranta *et al*., 2000, 2001; Blaum *et al*., 2015). CFH exists primarily as monomer with approximately 15 % existing as dimer with SCR-6/8 and SCR-17/18 providing the self-association sites (Fernando *et al*., 2007; Okemefuna *et al*., 2008; this thesis, chapter four). Due to its size, flexibility and glycosylation, attempts to solve the structure of full length CFH have proved unsuccessful. X-ray crystallography and NMR have provided high resolution information on fragments of the CFH SCR domains including SCR-1/3, SCR-5, SCR-6/8, SCR-15/16, SCR-18/20, and SCR-19/20 (Barlow, 1992; Barlow, 1993; Jokiranta *et al*., 2006, Prosser *et al*., 2007; Hocking *et al*., 2008). Solution scattering models are also available for full length CFH and SCR-16/20, showing that CFH has a bent back structure at its C-terminus (Figure 5.1 B) (Okemefuna *et al*. 2008, 2009b). CFH mutations and polymorphisms have been strongly correlated to age-related macular degeneration (AMD) and atypical haemolytic uremic syndrome (aHUS). AMD is characterised by the formation of drusen deposits in the retina and is one of the leading causes of blindness to individuals above 50 years of age. aHUS is mostly observed in children and can lead to renal failure following damage of the endothelial cells of the kidney. The Y402H polymorphism in SCR-7 has been strongly correlated to AMD, while mutations within

the SCR-16/20 domains at the C-terminal are correlated to aHUS (Klein *et al.*, 2005; Saunders *et al*., 2007; Rodriguez *et al*., 2014).

C-reactive protein (CRP) is a 115 kDa Ca^{2+}-dependent acute phase protein. It has a serum concentration of 0.8 mg/L which can increase 1000-fold during an inflammatory response. CRP is a member of the pentraxin family of proteins (Pepys and Hirschfield, 2003). It exists as a pentamer composed of five identical non-covalently bound monomers (Figure 5.1 A). CRP exists in a rapid pentamer-decamer equilibrium under physiological conditions (Okemefuna *et al*., 2009). It recognises and interacts with exposed molecules on damaged or apoptotic cell surfaces and on pathogens. It has a number of ligands including phosphocholine, phosphoethanolamine, microbial surface proteins, chromatin, histones, fibronectin, small nuclear ribonucleotides, lamin and polycations (Black *et al*., 2004). Ligand bound CRP interacts with the classical complement protein C1q, inducing conformational changes which ultimately activate the classical complement pathway. CRP has been implicated in AMD with elevated levels seen in patients with AMD; however its role in its pathogenesis is largely unknown (Seddon *et al*., 2004).

The existence of an interaction between CFH and CRP has been debated. The first evidence of a CFH-CRP interaction came from an enzyme-linked immunosorbant assay (ELISA) whereby CRP was immobilised on the plastic well of the ELISA plate (Mold *et al*., 1999). Studies using surface plasmon resonance (SPR) with CRP immobilised to the SPR chip followed, also providing evidence for a CRP-CFH interaction. These experiments suggested that CFH SCR-7 and SCR-8/11 contained the CFH binding sites for CRP (Jarva *et al*., 1999). The weaker binding of the CFH H402 variant to CRP was suggested to be a mechanism of drusen deposit formation in AMD, whereby CRP binding caused an increase in complement activation and inflammation. However, it later emerged that upon immobilisation to the ELISA well and immobilisation to the SPR chip, CRP had dissociated into its monomers and that monomeric CRP bound to CFH; hence the existence of a physiologically relevant interaction came into question (Biro *et al*., 2007; Hakobyan *et al*., 2008). A study was carried out which revealed that CRP was stable in its pentameric state in 140 mM NaCl, and 2 mM Ca^{2+} (Okemefuna *et al*., 2010a; Okemefuna *et al.*, 2010b). Under these conditions, binding studies in solution and on surfaces using analytical ultracentrifugation (AUC), SPR and small angle X-ray scattering (SAXS) confirmed

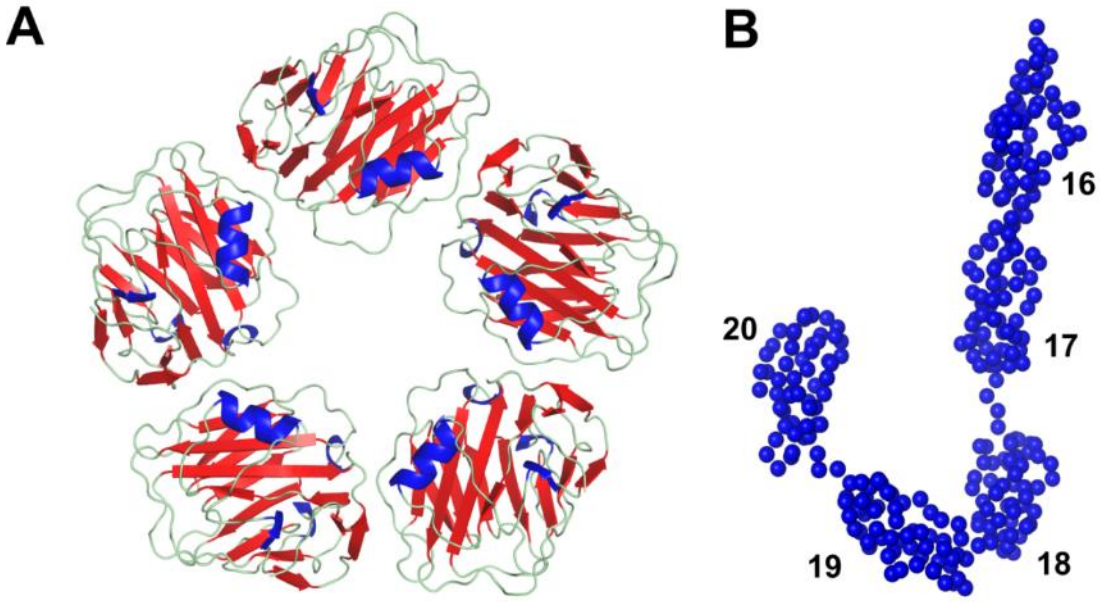


Figure 5.1 CRP and the CFH C-terminal SCR-16/20 domains. *A*, The structure of pentameric CRP from its crystal structure (PDB code: 1GNH). The pentamer is composed of five identical monomers each containing a Ca^{2+} binding site and a distinct ligand binding site. *B*, The low resolution solution structure of SCR-16/20 showing its bent back structure.

that CFH interacts with pentameric CRP under physiological conditions. AUC and SPR revealed that full length CFH binds to CRP. The SPR data fit a 1:1 kinetic model with dissociation constant (K_D) of 4.2 μM. SAXS experiments indicated that CFH-CRP complex formation resulted in a more compact structure than those of the unbound proteins. Additionally, SPR confirmed that wild type SCR-6/8 and the Y402H polymorphism bound to immobilised native CRP. The wild-type protein bound with a K_D of 3.9 μM while the Y402H variant bound with a K_D of 11.9 μM, suggesting that the CRP and CFH interaction may play a role in the disease mechanism of AMD. Interestingly, it has been show that individuals containing the Y402H variant have elevated levels of CRP in the eye (Johnson *et al*., 2006). The study by Okemefuna *et al*.(2010a) also revealed a novel binding site in the SCR-16/20 CFH C-terminus which bound to immobilised CRP with a K_D of 15.3 μM.

The aim of this part of the thesis was to determine which of the five C-terminal SCR-16/20 domains contain the pentameric CRP binding site. SCR-16/20, SCR-19/20, SCR-16/18H and SCR-17/18H were expressed and tested for binding to CRP using size exclusion chromatography (SEC), AUC, SPR, and microscale thermophoresis (MST). Negative stain electron microscopy (EM) images were also collected for a mixture of CRP and SCR-16/20. No complex formation was observed using SEC while AUC experiments using a fluorescence detection system (FDS-AUC) showed weak complex formation between SCR-16/20 and CRP but no interaction for the other SCR fragments. SPR revealed that SCR-16/20 and SCR-19/20 weakly bound to CRP suggesting that the binding site was within the SCR-19/20 binding domains. MST experiments also showed no SCR-16/18H and SCR17/18H binding with weak SCR-19/20 and SCR-16/20 binding observed further suggesting that SCR-19/20 contains the C-terminal CFH binding site. Negative stain electron microscopy (EM) revealed images of the CRP pentamer and decamer. Comparison of images of CRP with and without CFH suggests that two SCR-16/20 molecules may bind across two CRP pentamers which have formed an open decamer in which two monomeric subunits from each pentamer are in contact.

5.2 Materials and methods

5.2.1 Expression and purification of SCR fragments

SCR-19/20, SCR-16/20, SCR-16/18H and SCR-17/18H fragments of the CFH C-terminus were expressed and purified from the yeast *Pichia pastoris* using methods

described in Chapter Four (sections 4.2.1 and 4.2.2). SEC, SDS-PAGE analyses and protein concentration calculations were also carried out as described in Chapter Four (section 4.2.2).

5.2.2 C-reactive protein purification

Human CRP that had been isolated and purified from human pleural fluid was obtained from Lee Biosolutions (USA). SEC was carried out using a Superdex 200 16/600 column (GE Healthcare) in 10 mM Hepes, 137 mM NaCl, 2 mM $CaCl_2$ pH 7.4. SDS-PAGE analysis was used to assess sample purity using 12% Tris-Tricine gels. CRP concentrations were calculated using an absorption coefficient of 17.5 (1%, 280 nm, 1 cm path length). CRP has a calculated molecular mass of 115.1 kDa in its pentameric form.

5.2.3 Size exclusion chromatography

Samples were prepared with equimolar concentrations. 50 µl 1:1 molar mixtures (22 µM: 22 µM) of CRP with either SCR-19/20, SCR-16/20 or SCR-16/18H were loaded onto a superdex 200 increase 3.2/300 column with a bed volume of 2.4 ml. The column was calibrated using a set of molecular weight standards of globular proteins (BIO-RAD). A High Performance Liquid Chromatography (HPLC) system was used to minimise the dilution of the sample (Shimadzu Nexera Quaternary System).

5.2.4 AUC data collection and analyses

Sedimentation velocity experiments were carried out at 50,000 rpm as per the method described in Chapter Four (section 4.2.3). Data was collected on unbound SCR-19/20, SCR-16/20, SCR-16/18H, SCR-17/18H and CRP in 10 mM Hepes, 137 mM NaCl and 2 mM $CaCl_2$, pH 7.4 using interference optics. 1:1 molar mixtures of CRP (15 µM) and each of the SCR fragments (15 µM) were analysed by AUC using interference optics. SEDFIT (v14.6) was employed to calculate the size distribution *c(s)* analyses. (Shuck 1998, 2000). 60-90 scans were loaded, with the frictional ratio (f/f_0), meniscus and baseline floated for data fitting. The partial specific volumes ($\bar{v}$) calculated in Chapter Four for the SCR fragments were used here (Section 4.2.3). They were 0.7277 ml/g for SCR-19/20, 0.7151 ml/g for SCR-16/20, 0.7071 ml/g for SCR-16/18H and 0.7021 ml/g for SCR17/18H. The $\bar{v}$ for CRP was 0.7408 ml/g which was calculated from its primary sequence (obtained from UniProt KB) using SLUV

(Perkins, 1988). The buffer density was 1.00492 g/ml with a buffer viscosity of 1.002 cp (Hardy & Cottington, 1949).

HYDROPRO was used to predict the sedimentation coefficients ($s_{20,w}$) for the possible complexes formed between CRP and each of the SCR fragments.. No crystal structures were available for a CRP-SCR fragment complex. However individual crystal structures were available for CRP (PDB code: 1GNH), SCR-19/20 (PDB code: 2G7I), and a solution structure for SCR-16/20 (PDB code: 2QFH). Using these existing structures, models of the possible complexes formed between CRP and each of the SCR fragments were constructed using PyMOL. SCR-16/18H was constructed by removal of SCR-19/20 from SCR-16/20. Two possible models for each SCR fragment were constructed - one in which the SCR fragment lies across the CRP pentamer face (*i.e.* the most compact), and one in which the SCR fragment interacts with one monomer of the CRP pentamer and protrudes from the side of CRP (*i.e.* the most elongated). The theoretical $s_{20,w}$ was calculated for each.

5.2.5 FDS-AUC data collection and analyses

SCR-19/20, SCR-16/20, SCR-16/18H, and SCR-17/18H were fluorescently labelled with fluorescein isothiocyanate (FITC) according to the manufacturers guidelines (Sigma-Aldrich). FITC has excitation and emission wavelengths of 492 nm and 518 nm respectively. In brief, the SCR fragments were dialysed against 0.1 M sodium carbonate-bicarbonate buffer, pH 9.0. 5 mg of each SCR fragment was incubated with 1 mg of FITC in a 1 ml reaction volume at room temperature with gentle stirring for two hours in the dark. Labelled protein was separated from free FITC using a sephadex G-25M column (Sigma-Aldrich). The molar fluorescence/protein ratio was calculated from UV absorbance measurements at 280 nm and 495 nm. Labelled protein was stored at 4 °C in the dark.

FDS-AUC velocity experiments were carried out at 50,000 rpm using an Optima XL-I analytical ultracentrifuge (Beckman Coulter) equipped with a 10 mW laser with an emission wavelength of 488 nm. FITC-labelled SCR fragments were initially centrifuged at 3,000 rpm to optimise the FDS-AUC settings for the fluorescence intensity of each sample depending on concentration. FDS-AUC sedimentation velocity data were collected for each of the FITC-labelled SCR fragments alone at 15 μM (0.23 mg/ml for SCR-19/20, 0.57 mg/ml for SCR-16/20, 0.4 mg/ml for SCR-16/18H and 0.35 mg/ml for SCR-17/18H) as a control. FDS-AUC

data was collected on molar ratio mixtures of CRP (15 μM) and the FITC-labelled SCR fragments with ratios of 1:1 and 1:2 with the SCR fragment concentration varied. All experiments were carried out in 10 mM Hepes, 137 mM NaCl, 2 mM $CaCl_2$ pH 7.4. A described above, and in Chapter Four, *c(s)* distribution analyses were obtained using SEDFIT (v14.6).

5.2.6 SPR data collection and analyses

A Biacore X100 SPR instrument with evaluation software, version 1.1 (GE Healthcare) was used. CRP was labelled with EZ-Link Sulfo-NHS-LC-Biotin according to the manufacturers guidelines (Thermo Scientific). A 10 mM Hepes, 137 mM NaCl, 2 mM $CaCl_2$ pH 7.4 buffer was used throughout the labelling procedure. A 1:1 molar mixture of CRP and biotin was prepared and incubated on ice for 2 hours. To remove free biotin, the mixture was extensively dialysed against 10 mM Hepes, 137 mM NaCl, 2 mM $CaCl_2$, pH 7.4. The biotin to protein ratio was determined using a Pierce Biotin Quantitation Kit (Thermo Scientific). Biotinylated CRP was coupled to the flow cell of a strepavadin (SA) research grade sensor chip (GE Healthcare) with each conjugated biotin binding one avidin molecule. A 3 μg/ml sample of biotinylated CRP in 10 mM Hepes, 137 mM NaCl, 2 mM $CaCl_2$, pH 7.4. was injected over flow cell two for 30 seconds until the appropriate level of response units (RU) was obtained. In between experiments the SA chip with immobilised CRP was stored in the same buffer. A control surface was prepared on flow cell one without CRP immobilisation. To alleviate non-specific binding to the reference cell (flow cell one), human serum albumin (HSA) was bound to flow cell one using the same method as used for CRP immobilisation in flow cell two. Each of the SCR fragments were flowed over the chip at rates of 10-30 μl/min at 25 ºC. Regeneration was achieved after each run using 10mM Hepes, 2 M NaCl, 2 mM CaCl2, pH 7.4. Affinity fits were performed using the Biacore evaluation software, BIaEvaluation 3.0 (GE Healthcare).

5.2.7 MST data collection and analyses

MST data were collected using a Monolith NT.115 instrument (NanoTemper Technologies, Germany). 19 μM CRP was fluorescently labelled using the Monolith Protein Labelling Kit (NanoTemper Technologies, Germany). The fluorescent label was coupled to the primary amines of the protein through its reactive NHS-ester group with unbound label removed by a size exclusion column. The label was resuspended

in 10 mM Hepes, 137 mM NaCl, 2 mM $CaCl_2$, pH 7.4, and mixed in a 1:3 molar ratio with CRP, with the label in excess. The mixture was incubated for 30 minutes at room temperature in the dark. The ratio of protein to label was calculated by measuring the absorbance at 280 nm and 495 nm. The integrity of labelled CRP was confirmed using size exclusion chromatography. The dilution of labelled CRP was tested using Monolith standard capillaries to ensure that the fluorescence count was sufficient. The concentration of labelled CRP was kept constant at 18 nM with the concentration of the SCR fragments varied. A 1:2 serial dilution series of the SCR fragments was prepared in 10 mM Hepes, 137 mM NaCl, 2 mM $CaCl_2$, pH 7.4, with concentrations in the micromolar range. 16 dilutions were performed with MST measurements recorded in triplicate. LED power was at 50% and MST power was at 40%. Monolith software was used to analyse the MST data and to obtain the K_D of the interaction (NT Analysis 1.5.41).

5.2.8. Negative stain electron microscopy

Negative stain EM was carried out in collaboration with Dr. Wai Li Ling in the Electron Microscopy and Methods group at the Institute de Biologie Structurale (IBS) in Grenoble. The IBS, along with the ILL, ESRF, and EMBL, is a partner in the Partnership for Structural Biology (PSB), based on the European Photon and Neutron (EPN) campus in Grenoble.

To prepare the sample for EM measurements, ~4 µl of the protein sample was applied to a mica sheet covered with a film of evaporated carbon. The carbon film was then floated off the mica in ~100 µl of 2% sodium silicotungstate (SST) and retrieved on a 400-mesh copper EM grid and let dry. Imaging was performed on an FEI F20 transmission electron microscope operating at 200 kV. Images were recorded on a 4k x 4k Eagle digital camera (CCD) with a pixel size of 2.17 angstroms. CRP alone was at 8 µM (0.9 mg/ml) in 10 mM Hepes, 50mM NaCl, 2 mM $CaCl_2$, pH 7.4. Mixtures of CRP with SCR-16/20 were prepared at 1:1 ratios in the same buffer at a total protein concentration of 5 µM. For the CRP sample alone, 8812 particles were selected semi-automatically with EMAN (Tang *et al.*, 2007) from 24 CCD frames. For the CRP sample with SCR-16/20, 5601 particles were selected from 50 CCD frames. The sub-frames were corrected for the contrast transfer function with CTFFIND3 (Mindell & Grigorieff, 2003) and sorted with reference-free 2D class averaging in RELION (Scheres, 2012).

5.3 Results

5.3.1 Protein expression and purification

Each of the SCR fragments were successfully expressed and purified as described in Chapter Four (section 4.3.1). SEC elution profiles for each of the SCR fragments is shown in Figure 5.2 A, alongside the CRP elution profile (left in Figure 5.2 A). The SEC results for CRP resulted in a single peak in the elution profile corresponding to a protein of approximately 115 kDa (Figure 5.2 A). CRP appeared as a clean band at 25 kDa on the SDS-PAGE gel corresponding to monomeric CRP (Figure 5.2 B).

5.3.2 SEC on CRP and SCR fragment mixtures

SEC was carried out on mixtures of CRP with SCR-16/20, SCR-19/20 and SCR-16/18H to investigate complex formation. In theory if a complex formed, a third peak would be seen in the elution profile. Using a regular superdex 200 10/300 (~24 ml) column, no third peak was osbserved (data not shown). As the CRP and SCR interaction was assumed to be weak, (K_D ~15 μM, Okemefuna *et al.*, 2010a) an analytical Superdex-200 Increase column was employed. The design of the superdex-200 Increase (which has a bed volume of 2.4 ml) reduces the amount of sample dilution while retaining the resolution such that high concentrations can be loaded with minimal on-column dilution (GE Healthcare, Superdex 200 increase manual). This enhances the possibility of observing a peak associated with a weak complex. CRP alone was injected onto the column which, in a similar manner to results from the preparative Superdex 200, showed a single peak with slight broadening due to the formation of the CRP decamer (Figure 5.3). Mixtures with total protein concentration at 44 μM (total protein; CRP at 2.4 mg/ml, SCR-19/20 at 0.35 mg/ml, SCR-16/20 at 0.8 mg/ml and SCR-16/18H at 0.55 mg/ml) were injected onto the column. For CRP and SCR-16/18H and CRP and SCR-16/20 mixtures, one asymmetric peak was observed due to the overlap of the two individual protein peaks (Figure 5.3). SCR-16/18H and SCR-16/20 elute earlier than expected due to their elongated shapes and dimerisation (Chapter Four). No third peak was observed. For the CRP and SCR-19/20 mixture, two peaks were observed, one corresponding to CRP, and one to SCR-19/20, with no third peak observed. SEC indicated either that complex formation did not take place or that it was too weak to

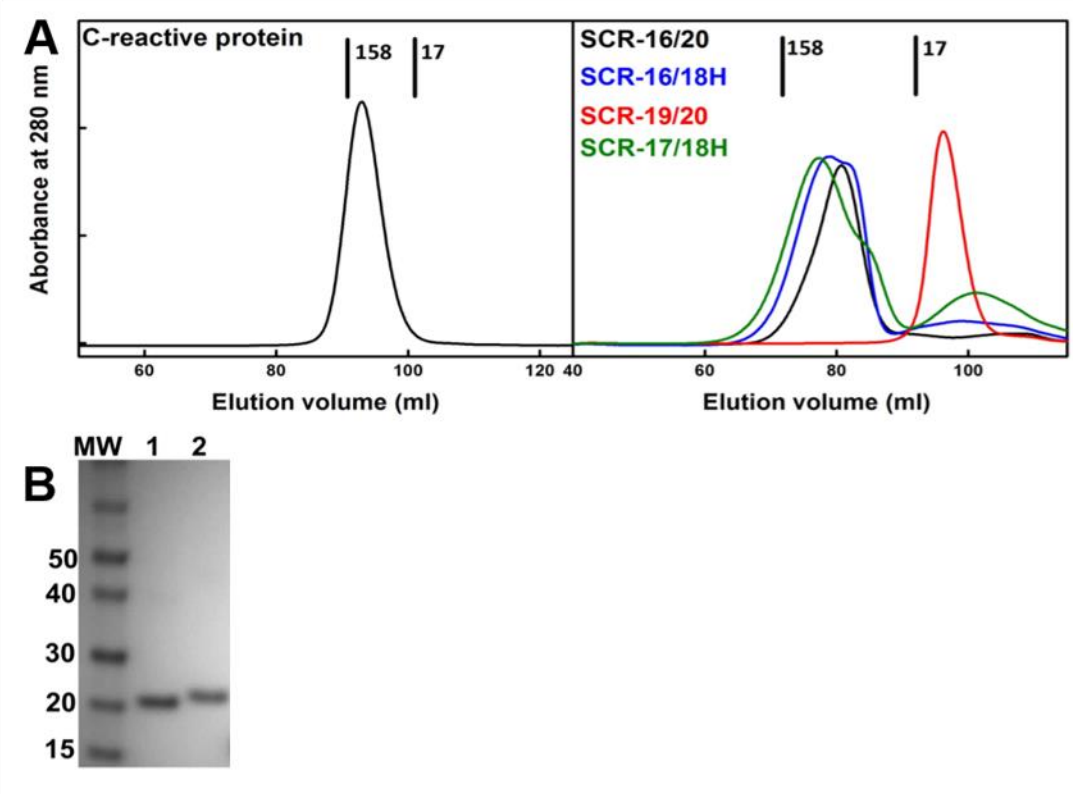


Figure 5.2 Purification of CRP and the SCR fragments. *A*, Size exclusion chromatography elution profile of CRP (left panel) using a Superdex-200 column. The right panel shows the elution profiles of each of the SCR fragments from a superdex-75 column. Molecular weight standards are indicated in both profiles in kDa. *B*, SDS-PAGE of CRP after size exclusion chromatography. Lane 1 and 2, CRP non-reduced and reduced, respectively.

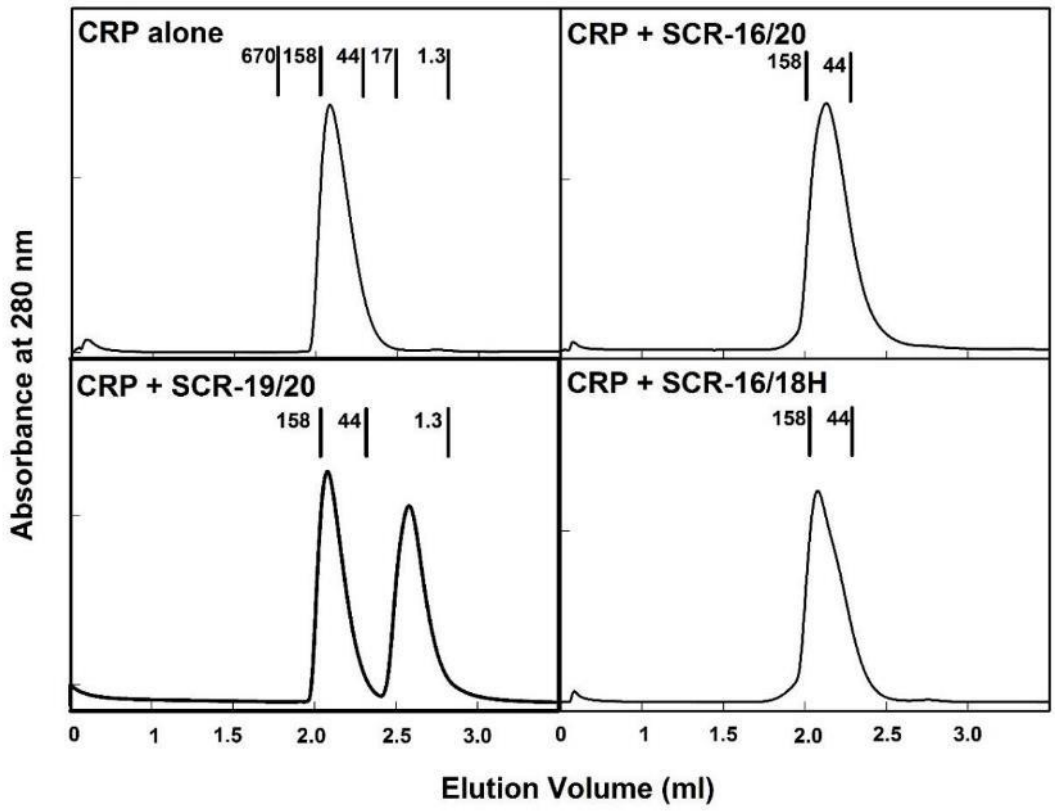


Figure 5.3 Size exclusion chromatography of CRP and SCR fragment mixtures. A Superdex-200 increase column (2.4 ml) was employed to investigate complex formation. The elution profiles monitored by 280 nm absorbance are shown. The molecular weight standards for a set of globular proteins are indicated in each profile in kDa.

be observed by this method. It is also possible that the column resolution was not sufficient to separate complexes of 153 kDa (CRP-SCR-16/20 complex), 143 kDa (CRP-SCR-16/18H complex), and 130 kDa (CRP-SCR-19/20) or that the shape did not change significantly so that the peak associated with the complex occurred within the CRP and SCR mixture peak.

5.3.3 AUC size distribution *c(s)* analyses

Molecules of different sizes and shapes will be separated by AUC. Size distribution *c(s)* analyses reveal the different species present in the sedimenting sample by the number of peaks in the distribution. From the *c(s)* analysis the sedimentation coefficient ($s_{20,w}$) and the f/f_0 are calculated, both provide information about the size and shape of the molecule. Conversion of the *c(s)* analysis to the *c(M)* analysis yields information about the molecular weight of the sedimenting species.

5.3.3.1 Theoretical sedimentation coefficient calculation using HYDROPRO

Theoretical $s_{20,w}$ values for the possible complexes formed between CRP and each of the SCR fragments were firstly calculated. HYDROPRO calculates the hydrodynamic properties of proteins from the atomic coordinates of a crystal structure (Garcia de la Torre *et al*., 2000). Theoretical models of the potential complexes were constructed using the CRP crystal structure and crystal and solution structures available for the CFH C-terminus. These were constructed using PyMOL and were used to give an indication of the expected $s_{20,w}$.

(i) The theoretical sedimentation coefficient for a CRP-SCR-19/20 complex where SCR-19/20 bound across the face of CRP (*i.e.* making a compact shape) was calculated to be 6.8 S. If SCR-19/20 bound to a CRP monomer and protruded at the side of the protein (*i.e.* forming an elongated shape) the theoretical S value was estimated to be 6.2 S.

(ii) For a SCR-16/20-CRP complex formed with a compact shape, the theoretical $s_{20,w}$ value was calculated to be 5.6 S while for an elongated complex in which SCR-16/20 protrudes from CRP, the theoretical value was estimated to be 5.8 S.

(iii) For an SCR-16/18H and CRP complex, two similar structures were constructed using SCR-16/18H constructed from the solution structure of SCR-16/20. This estimated a theoretical $s_{20,w}$ value of 6.8 S for a compact complex and a theoretical

$s_{20,w}$ value of 5.7 S for an elongated complex.

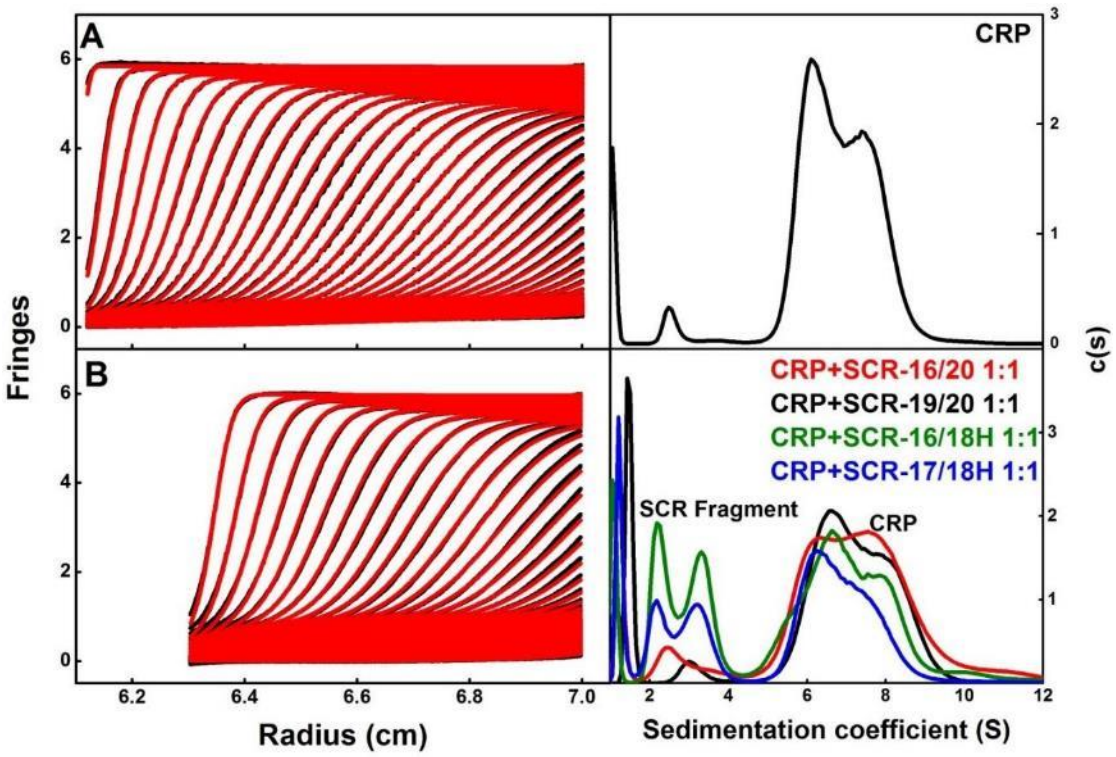


Figure 5.4 Sedimentation velocity size distribution *c(s)* analyses of CRP and SCR fragment mixtures using interference optics. *A*, CRP alone. The interference optics boundary fit (left panel) yields the *c(s)* analyses on the right showing two partially resolved peaks. *B*, For CRP and each of the SCR fragments in 1:1 molar mixtures, the *c(s)* analyses from the interference optics boundary fits showed peaks which correspond to both the SCR fragments and to CRP.

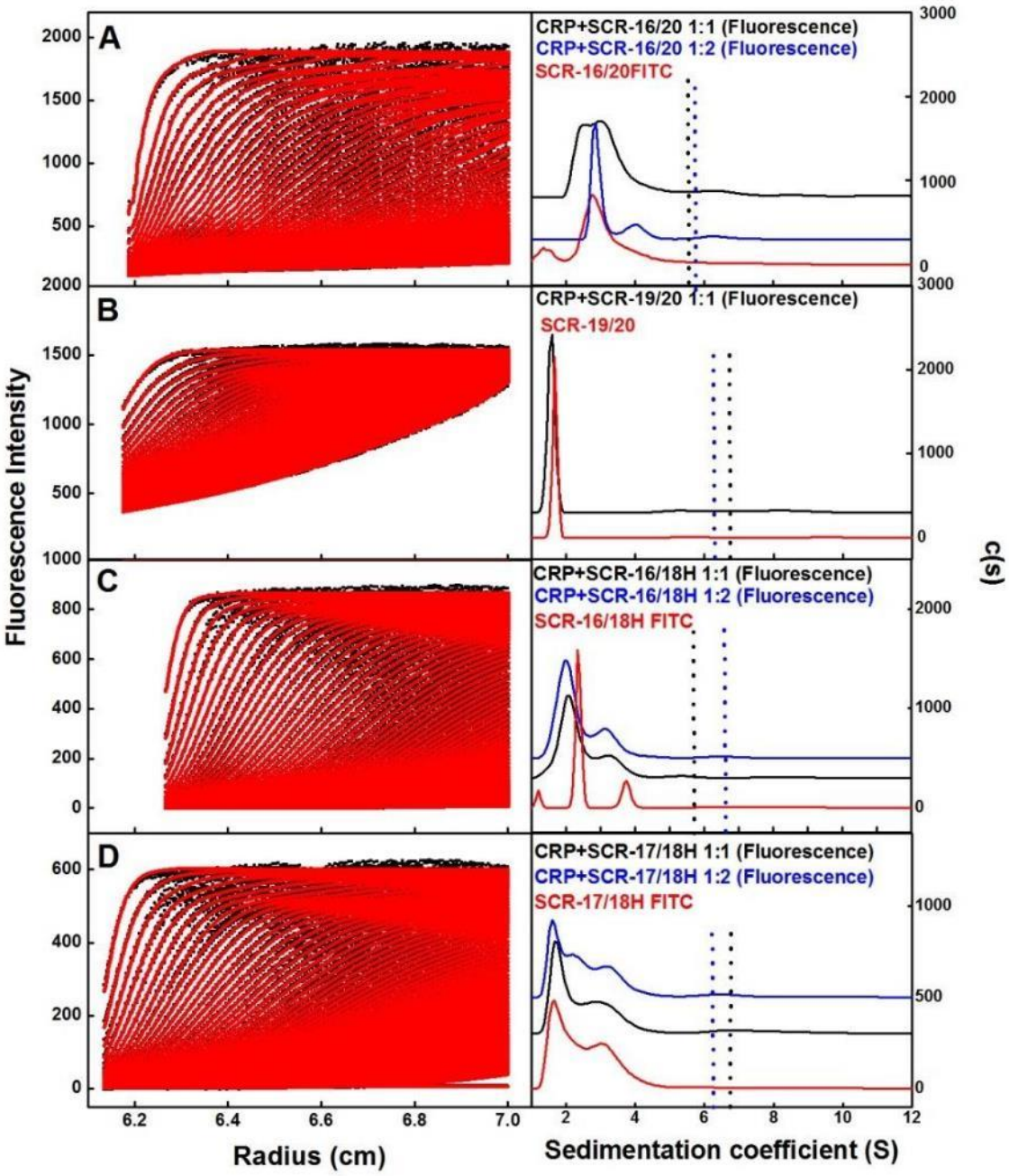


Figure 5.5 Sedimentation velocity size distribution *c(s)* analyses of CRP and SCR fragment using FDS-AUC. The SCR fragments are fluorescently labelled and consequently only these are visible in the *c(s)* plot. The black dotted lines represent the theoretical sedimentation coefficient for a complex formed with a compact shape. The blue dotted line represents the theoretical sedimentation coefficient for a complex with an elongated shape. *A*, CRP and SCR-16/20. Labelled SCR-16/20 alone gave one broad peak corresponding to SCR-16/20 monomer and dimer (red). CRP and SCR-16/20 mixed in a molar ratio of 1:1 (black) showed a double peak corresponding to

(Figure 5.5 continued) monomeric and dimeric SCR-16/20. A third peak was also visible corresponding to the CRP and SCR-16/20 complex. 1:2 ratio (blue) showed two peaks corresponding to SCR-16/20 monomer and dimer. A third small peak was visible corresponding to the CRP and SCR-16/20 complex.

B, Labelled SCR-19/20 (red) alone gave a single monomer peak. CRP and SCR-19/20 mixed in a 1:1 ratio (black) also showed a single monomeric peak.

C, Labelled SCR-16/18H alone (red) showed two peaks corresponding to monomer and dimer SCR-16/18H. CRP and SCR-16/18H in a 1:1 ratio (black) showed two peaks corresponding to monomer and dimer SCR-16/18H. 1:2 ratio (blue) also showed two peaks corresponding to monomer and dimer SCR-16/18H.

D, Labelled SCR-17/18H alone (red) showed two peaks corresponding to monomer and dimer SCR-17/18H. 1:1 molar mixture of CRP and SCR-17/18H (black) showed two peaks corresponding to SCR-17/18H monomer and dimer. 1:2 ratio (blue) showed two peaks corresponding to monomer and dimer SCR-17/18H.

(iv) Both SCR-19/20 and SCR-17/18H contain two SCR domains, and so the same theoretical values of 6.8 S and 6.2 S were also used as an indication of the theoretical $s_{20,w}$ value of a possible CRP- SCR-17/18H complex.

5.3.3.2 AUC data analyses using interference optics

Each of the calculated theoretical values from above were within the same range as those observed for the CRP pentamer and decamer (6.4-8.0 S) (Figure 5.4 B) which due to its weak nature exists with two partially resolved broad peaks (Figure 5.4 A) (Okemfuna *et al*., 2010b). This meant that it would be difficult to distinguish between peaks corresponding to the CRP pentamer and decamer and a CRP and SCR fragment complex. Nevertheless size distribution *c(s)* analyses were carried out on 1:1 molar mixtures of CRP and each of the SCR fragments using interference optics.

(i) The *c(s)* analysis for CRP alone showed two partially resolved peaks (Figure 5.4 A) with $s_{20,w}$ values of 6.4 S and 7.1 S corresponding to pentameric and decameric CRP, respectively, in good agreement with previously published work (Okemefuna *et al.*, 2010).

(ii) For a 1:1 molar mixture of SCR-19/20 and CRP (15 µM:15 µM), three peaks were observed with $s_{20,w}$ values of 1.5 S, 6.6 S and 7.9 S corresponding to monomeric SCR-19/20 and the CRP pentamer-decamer equilibrium (Figure 5.4 B).

(iii) For 1:1 molar ratios of CRP and SCR-16/20, three peaks were observed with $s_{20,w}$ values of 2.5 S, 6.8 S and 7.9 S. The first peak is broad and corresponded to the weak monomer-dimer equilibrium of SCR-16/20 while the second two peaks were assigned to the CRP pentamer and decamer (Figure 5.4 B).

(iv) The mixture of CRP and SCR-16/18H showed four peaks corresponding to monomeric and dimeric SCR-16/18H, with $s_{20,w}$ values of 2.18 S and 3.3 S respectively. The second two peaks corresponded to the pentamer-decamer of CRP with $s_{20,w}$ values of 6.7 S and 7.9 S respectively (Figure 5.4 B).

(v) SCR-17/18H and CRP showed four peaks in the *c(s)* analysis. Peaks one and two corresponded to monomeric and dimeric SCR-17/18H with $s_{20,w}$ of 2.2 S and 3.2 S, while the last two peaks corresponded to the CRP pentamer and decamer with $s_{20,w}$ of 6.3 S and 7.5 S (Figure 5.4 B).

From this analysis it was impossible to determine which of the SCR fragments formed a complex with CRP due to the potential overlap of $s_{20,w}$ values for the CRP

pentamer/decamer and for a CRP-SCR fragment complex. It was evident that both proteins retain the ability to oligomerise in the presence of each other.

5.3.4. FDS-AUC size distribution *c(s)* analyses

FDS-AUC was employed to overcome the difficulties discussed above for AUC using interference optics. Each of the SCR fragments were fluorescently labelled with only the sedimentation of the labelled molecule detected by the FDS. This meant that the CRP peaks would not be seen allowing the observation of a potential CRP-SCR fragment peak. *c(s)* analyses were carried out on each of the labelled SCR fragments alone to confirm that addition of FITC did not alter their sedimentation properties (Figure 5.5). Fluorescently labelled SCR-19/20 was monomeric, with an $s_{20,w}$ value of 1.78 S. The $s_{20,w}$ values for fluorescently labelled SCR-16/20 was 2.85 S and 3.6 S corresponding to monomeric and dimeric SCR-16/20. Fluorescently labelled SCR-16/18H had $s_{20,w}$ values of 2.59 S and 3.69 S, corresponding to its monomeric and dimeric forms. SCR-17/18H which was fluorescently labelled had $s_{20,w}$ values of 2.01 S and 3.51 S for monomer and dimer respectively. These values were in good agreement with the non-labelled SCR fragments (Chapter Four), confirming that labelling did not affect the SCR fragments.

The labelled fragments were mixed in molar ratios with CRP in 1:1 and 1:2 molar ratios, with the exception of CRP and labelled SCR-19/20, which was only carried out in 1:1 molar ratio due to labelled SCR-19/20 aggregation at higher concentrations.

(i) SCR-16/20 and CRP were mixed in a molar ratio of 1:1 (15 μM:15 μM). Two major peaks were visible using the FDS-AUC (Figure 5.5 A). These were partially resolved at 2.5 S and 3.48 S, corresponding to the SCR-16/20 monomer and dimer. A smaller third peak was visible at 6.52 S. Conversion to the *c(M)* yielded an apparent molecular mass of 155 kDa which corresponded to a CRP and SCR-16/20 complex. This peak corresponded to 9 % of the total sample, indicating a weak interaction. A 1:2 molar ratio of CRP to SCR-16/20 (15μM:30μM) showed a similar result with two major peaks, and a small third peak. The two major peaks corresponding to SCR-16/20 monomer and dimer were shifted with respect to the 1:1 complex and had larger $s_{20,w}$ values than expected (2.8 S and 3.0 S respectively). The third peak had an $s_{20,w}$ value of 6.4 S which when converted to *c(M)* gave an apparent molecular weight of 147 kDa, similar to the 1:1 molar mixture.

(ii) To determine if SCR-19/20 was involved in CRP binding, a 1:1 molar ratio mixture was analysed (15 μM: 15 μM). This showed only one peak by FDS optics, with an $s_{20,w}$ of 1.64 S corresponding to monomeric SCR-19/20 (Figure 5.4, B).

(iii) CRP and SCR-16/18H 1:1 and 1:2 molar mixtures with SCR-16/18H in excess (15 μM: 15 μM or 30 μM) resulted in two peaks corresponding to monomeric and dimeric SCR-16/18H with $s_{20,w}$ values of 2.1 S and 3.2 S. No additional peaks were observed.

(iv) Similarly for CRP and SCR-17/18H (15 μM: 15 μM or 30 μM), two peaks were observed corresponding to the SCR-17/18H monomer and dimer with $s_{20,w}$ values of 1.6 S and 2.9 S for 1:1 and 1.6 S and 3.2 S for 1:2. No additional peaks were observed.

From the FDS-AUC results it was not clear which of the SCR fragments of the CFH C-terminal interacts with CRP. SCR-16/20 and CRP shows a small peak in the region of the theoretical $s_{20,w}$ value, suggesting weak binding between CRP and SCR-16/20. It is expected that either SCR-19/20 or SCR-16/18H would bind depending on which domains contain the binding site. However no binding is observed for these fragments by AUC. It is possible that all five domains of SCR-16/20 are required for binding in a synergistic manner explaining the lack of binding seen for the shorter SCR fragments.

5.3.5 SPR analyses of CRP and SCR fragment binding

SPR measures the interactions between biomolecules where one partner is bound to the surface of a sensor chip (ligand) and its partner (analyte) is injected across the surface in solution. Binding events at the chip surface result in changes in the refractive index which is proportional to the number of molecules bound. Rates of association and dissociation are measured and information about the steady state of binding are used to calculate the K_D value of the interaction. SPR measurements were carried out to test the binding of the SCR fragments to CRP immobilised on a streptavidin (SA) chip. To ensure CRP maintained its native pentameric structure, all stages of immobilisation and data collection were carried out in buffers containing 137 mM NaCl and 2 mM $CaCl_2$ at pH 7.4. SCR-16/20, SCR-19/20, SCR-16/18H and SCR-17/18H were initially tested for binding at concentrations of 20 μM and 40 μM. SCR-16/18H and SCR-17/18H showed no binding (Figure 5.6 A) while both SCR-16/20 and SCR-19/20 bound with a low response of ~ 10 RU (initial binding test not shown).

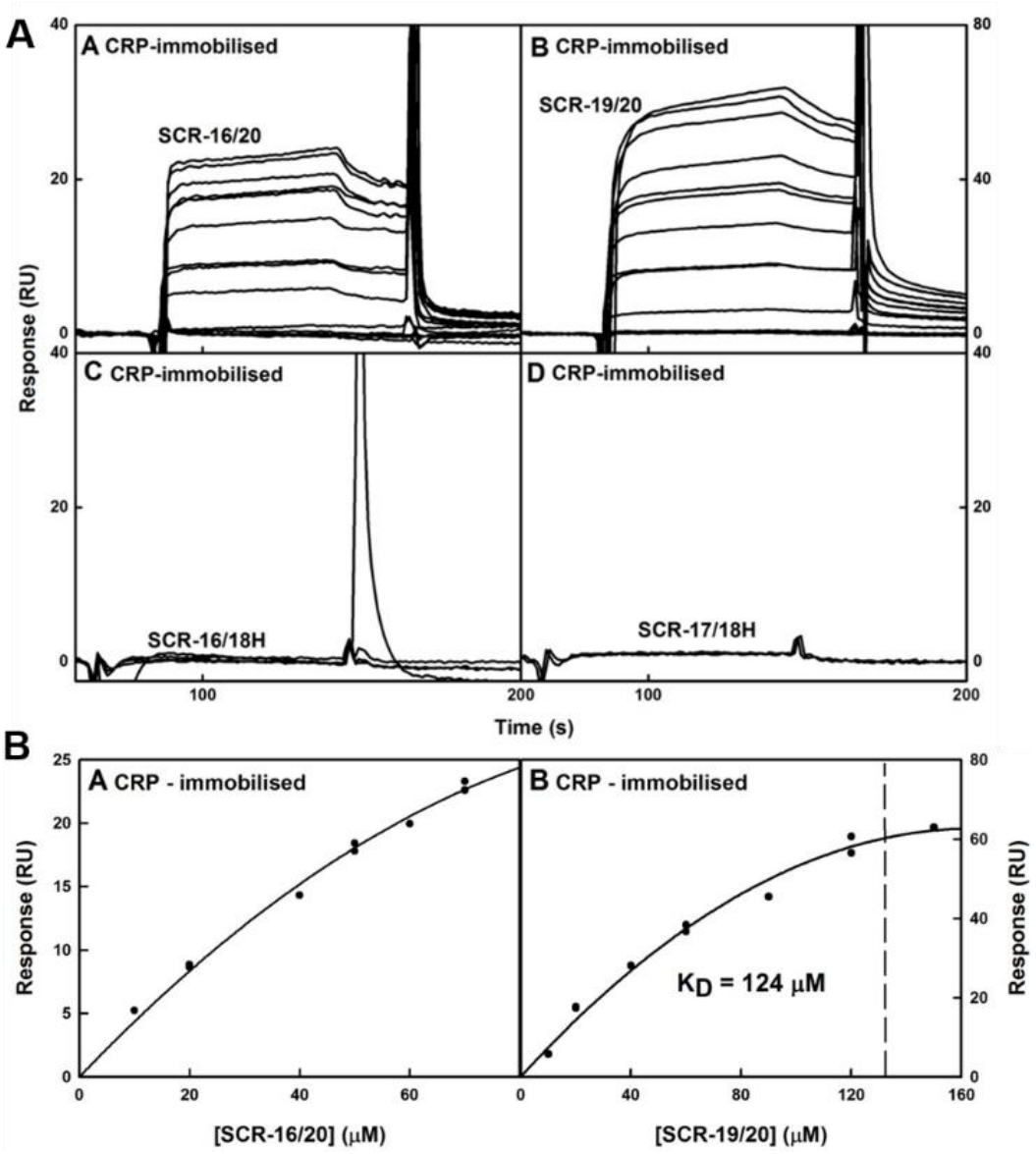


Figure 5.6 Surface Plasmon Resonance with CRP immobilised to a streptavidin chip. The SCR fragments were flown across the chip surface over a concentration series. A, The SPR binding response to each of the protein fragments. *A*, CRP immobilised with SCR-16/20 over a concentration range of 0, 10, 20, 40, 50, 60 and 70 µM. *B*, CRP immobilised with SCR-19/20 over a concentration range of 0, 10, 20, 40, 60, 90, 120 and 150 µM. *C*, CRP immobilised with SCR-16/18H at 40 µM. *D*, CRP immobilised with SCR-17/18H at 40 µM. B. *A*, The binding affinity curve for CRP and SCR-16/20. *B*, The binding affinity curve for CRP and SCR-19/20. The dashed line represents the K_D.

These results suggested that SCR-19/20 contains the CRP binding site in the CFH C-terminal. To determine the K_D value of the binding for SCR-16/20 and SCR-19/20 to CRP, a concentration series for each fragment was flown over immobilised CRP (Figure. 5.6 A). The sensorgrams showed a fast on and fast off rate for both interactions and it was therefore not possible to fit either of the sensorgrams to a kinetic model. Steady state affinity analyses were instead performed to calculate the K_D value.

(i) Concentrations of up to 70 μM of SCR-16/20 did not saturate the binding response (Figure 5.6 B) therefore a K_D value could not be calculated however the curve shows that the K_D value will be above 70 μM. Self-association of SCR-16/20 may explain why it does not reach saturation with CRP binding occurring followed by SCR-16/20 associating with a bound SCR-16/20 molecule.

(ii) SCR-19/20 binding was approaching saturation at 150 μM and the K_D value was estimated to be 124 μM (Figure 5.6 B). SCR-19/20 does not self-associate (chapter four) and similar to SCR-16/20 did not reach saturation at 70 μM indicating that the non-saturation of SCR-16/20 may be a result of weak binding rather than self-association effects.

The weak binding observed by SPR is similar to the AUC results which show that the interaction between SCR-16/20 and CRP is weak. In addition, binding is observed between SCR-19/20 and CRP strongly suggesting that the binding site is located in this region. No binding was observed for SCR-17/18H or SCR-16/18H. SPR will be affected by self-association, however weak binding was observed by AUC for SCR-16/20 and CRP which is not affected by SCR-16/20 propensity to dimerise.

5.3.6 MST analyses of CRP and SCR fragment binding

MST is a technique carried out in solution that measures the affinity of biomolecular interactions. MST was used to investigate the binding of CRP to the SCR fragments and to obtain the K_D value for the SCR fragments which bound. CRP was fluorescently labelled. Labelling and measurements were carried out in buffer containing 137 mM NaCl and 2 mM $CaCl_2$ at pH7.4 to ensure CRP remained in its native pentameric state. CRP concentration was kept constant at 18 nM while the SCR fragment concentration varied. Initial MST measurements were carried out with a start concentration of 100 μM for the four SCR fragments with a dilution series of 1:2 over 16 samples with the lowest concentration at 10 pM. Little or no binding was observed for these experiments (not shown). These experiments were repeated using a higher

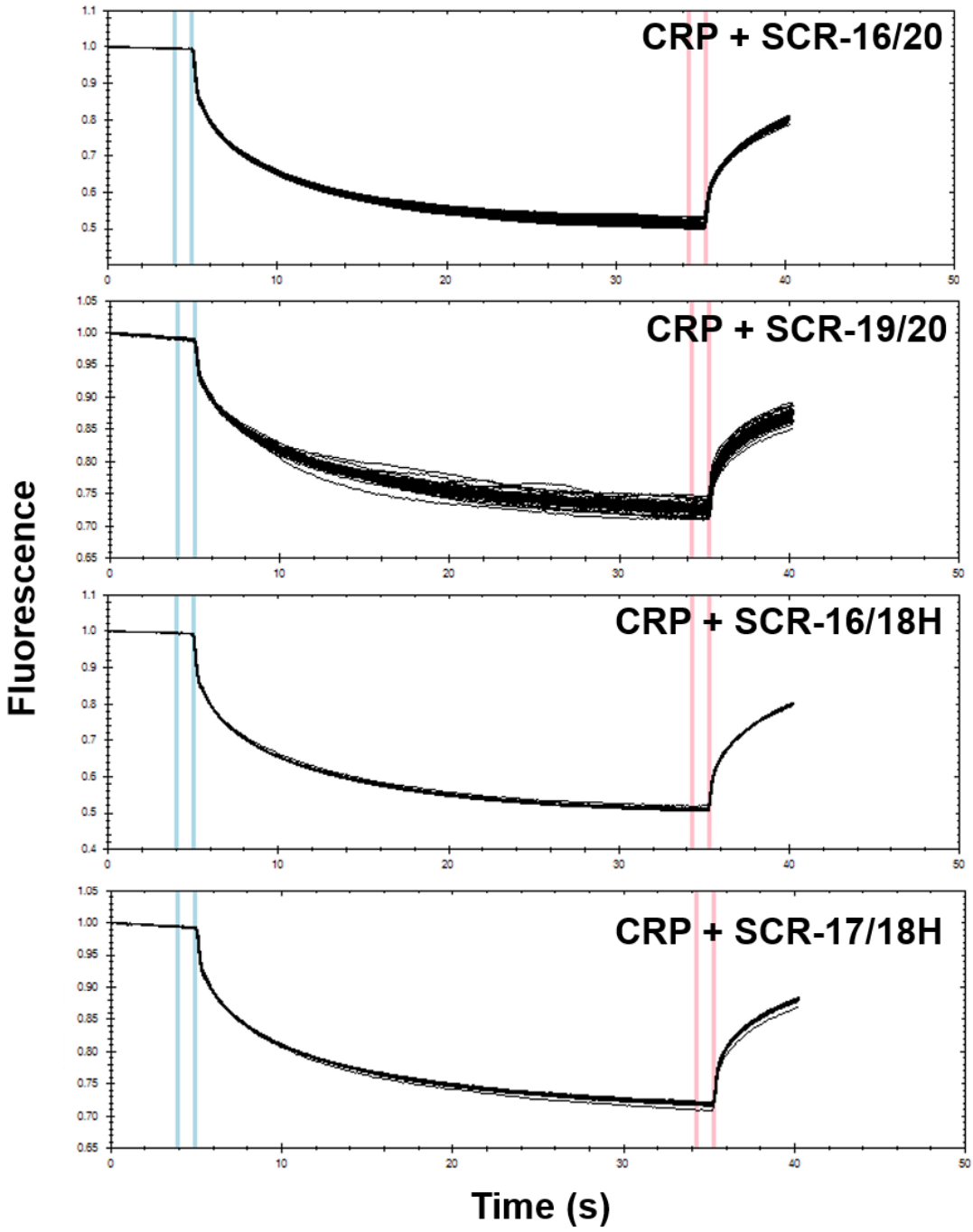


Figure 5.7 Fluorescent time traces of the MST experiment. The blue lines indicates the region where F_{cold} is measured while the red indicates the region where F_{hot} is measured. The drop in fluorescence is observed when the temperature gradient is applied. The steady state is reached immediately prior to measuring F_{hot}.

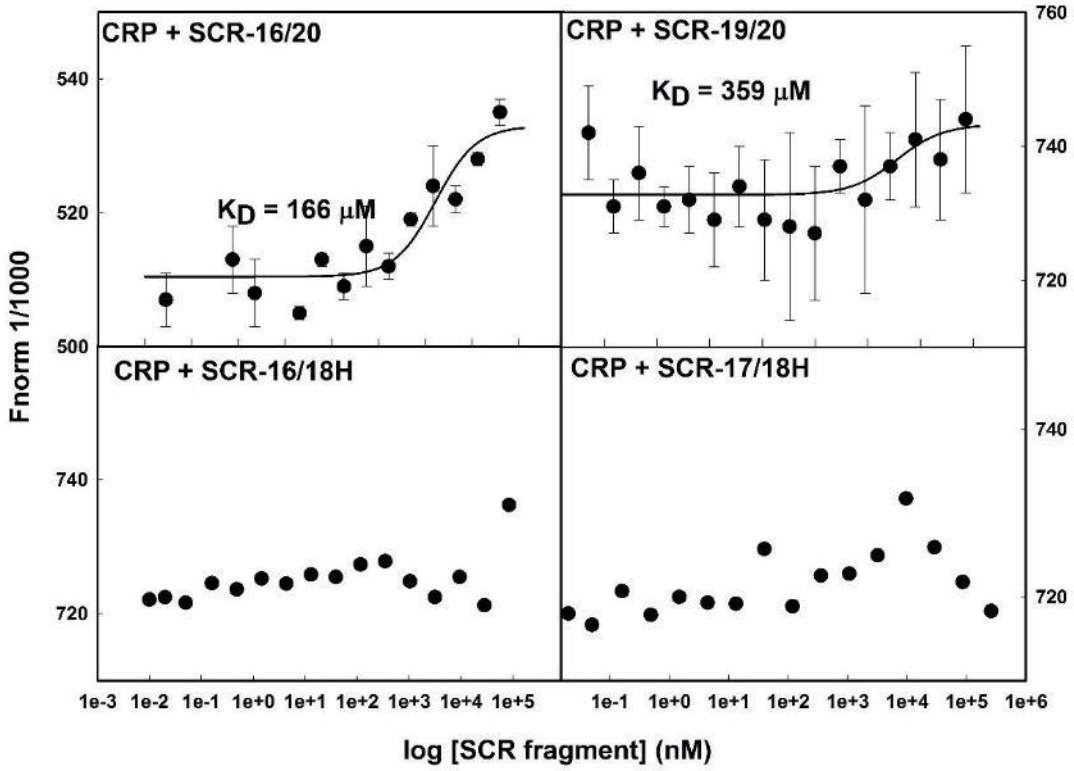


Figure 5.8 MST binding curves obtained for each of the SCR fragments and CRP. Fits were possible for both SCR-16/20 and SCR-19/20 with error calculated as the standard error of the mean of three measurements. Fits were not possible for SCR-16/18H and SCR-17/18H due to little or no change in thermophoresis indicating no binding took place.

start concentration of SCR fragment.

(i) SCR-19/20 had a start concentration of 704 μM with a 1:2 serial dilution carried out over 16 samples with the lowest concentration at 50 pM. Aggregation was evident from the fluorescent time trace, which was not smooth (Figure 5.7). A K_D value of 359 μM was calculated; however aggregation will affect the result and hence this cannot be taken as accurate. However it does indicate a K_D value in the high micromolar region (Figure 5.8).

(ii) SCR-16/20 had a start concentration of 400 μM with a 1:2 serial dilution carried out over 16 samples with the lowest concentration at 30 pM. The K_D was calculated to be 120 μM. This value was in agreement with the SPR results that show the K_D is in the high micromolar range. This K_D value could only be taken as an estimate as at 400 μM of SCR-16/20 the reaction had not reached saturation. This may be as a result of a weak interaction or may be an artefact of SCR-16/20 or CRP self-association, which prevent saturation of the binding curve.

(iii) SCR-16/18H had a start concentration of 82 μM, with a 1:2 serial dilution carried out. The lowest concentration was 10 pM. No change in thermophoresis and therefore no binding was observed.

(iv) SCR-17/18H had a starting concentration of 240 μM with a 1:2 serial dilution carried out over 16 samples. The lowest concentration of SCR-17/18H was 20 pM. No binding was observed between SCR-17/18H and CRP.

MST is affected by protein self-association and aggregation and so the K_D values obtained cannot be taken as accurate but they do indicate weak binding of CRP to SCR-16/20 and SCR-19/20 similar to the results seen by SPR. MST also confirms that neither SCR-16/18H nor SCR-17/18H binds to CRP suggesting that SCR-19/20 contains the CRP binding site.

5.3.7 Negative stain electron microscopy

Negative stain EM was used to obtain images of the CRP and SCR-16/20 complex. Images were acquired for CRP alone and for mixtures of CRP and SCR-16/20 of CFH. Samples were prepared in buffer containing 50 mM NaCl as the image quality was much better with this NaCl concentration than with 137 mM. CRP which has a molecular mass of 115 kDa was clearly visible in the negative stain EM images however SCR-16/20 which is 38 kDa was more difficult to identify. Figure 5.9 (A)

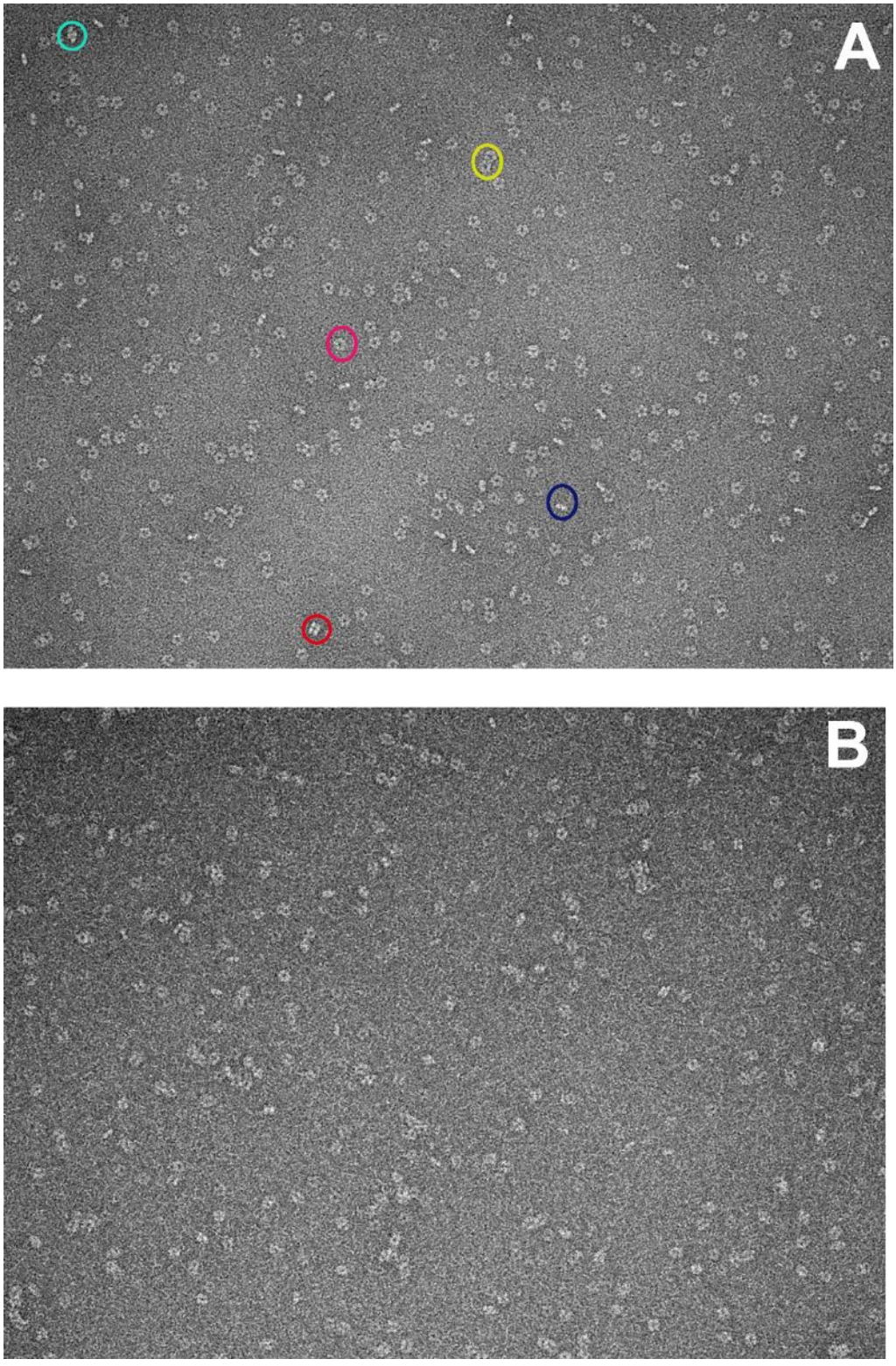


Figure 5.9 legend overleaf

Figure 5.9 Negative stain electron microscopy images of CRP and SCR-16/20 of CFH stained with 2% SST. *A,* A typical image of the sample containing CRP only in 50 mM NaCl. The top view of the CRP pentamer is circled in pink, the side view is circled in blue, the CRP decamer formed by two pentamers side by side is circled in yellow, a side view of the decamer formed by the stacking of two pentamers is circled in red, the side view of the formed decamer which is tilted is circled in turquoise. *B,* A typical image of the CRP and SCR-16/20 mixture with less defined particles evident.

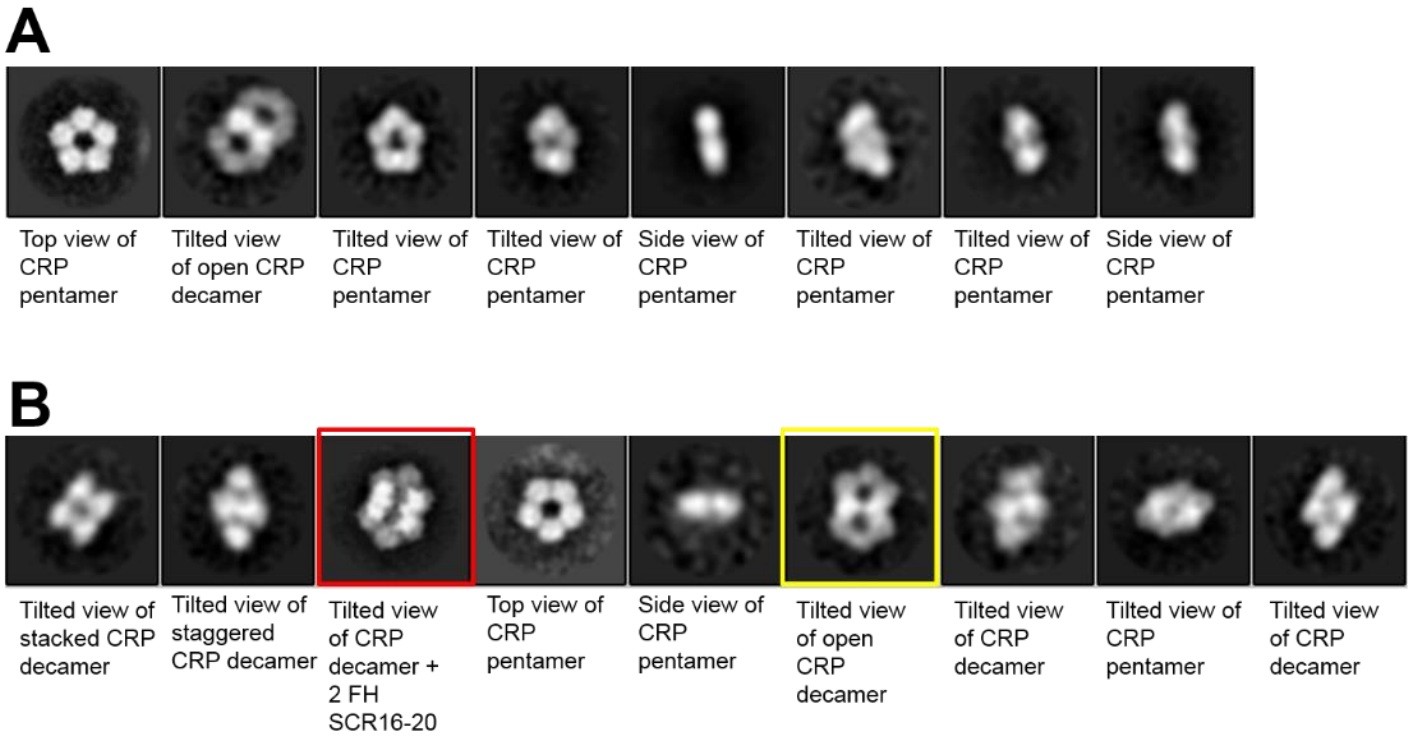


Figure 5.10 legend overleaf.

Figure 5.10 Class averages from the negative stain electron microscopy of CRP and SCR-16/20. *A,* The class averages from the images of CRP alone. Each of the particles were classified and particles belonging to the same class were averaged and are shown here. Highlighted in yellow is the open CRP decamer *B,* The class averages from the images of CRP and SCR-16/20. Highlighted in yellow is the open CRP decamer. Highlighted in red is an open CRP decamer which appears to have two molecules of SCR-16/20 bound.

shows a negative stain EM image of CRP alone. Multiple sub-frames of individual particles were extracted from the images and underwent reference-free 2D classification. Images from the same class were then averaged (Figure 5.10 A). A top view of the CRP pentamer is most prominent in (Figure 5.9 A, pink circle; Figure 5.10 A), however side views of the pentamers were also observed (Figure 5.9 A, blue circle). Because of their abundance, structural detail can be discerned in the top view class average in which the five subunits are clearly visible. Interestingly, three types of decamer were present as minority species. The stacked decamer was observed where two pentamers are stacked directly on top of each other (Figure 5.9 A, red circle) and a staggered decamer (Figure 5.9 A, turquoise circle; Figure 5.10) where the two pentamers were not fully aligned. Additionally, a third type of decamer was observed which we termed an 'open decamer' where two adjacent CRP monomers of two pentamers interact (Figure 5.10 A). Features were much less well defined for this class average with respect to the class average of the pentamer top view as the number of particles in this class are much smaller, *i.e*. the images showed a lot more pentamer top view than open decamer.

For samples containing mixtures of SCR-16/20 and CRP, the particles appeared less well defined (Figure 5.9 B) with the proportion of CRP pentamers diminished with respect to the sample of CRP alone. The open decamer, on the other hand, became more predominant suggesting that the presence of SCR-16/20 stabilises the CRP decamer (Figure 5.10). Similar to the images for CRP alone the sub-frames of individual particles were extracted and classified in an attempt to locate SCR-16/20 in a complex with CRP. The class average of the pentamer top view showed less structural detail than in the CRP alone sample as it was no longer a prominent species (Figure 5.10 B). Besides the classes of the pentamer top, side, tiled views and the side views of the three types of decamer observed in the sample of CRP alone (Figure 5.10 A), a new class emerged (Figure 5.10 B, boxed in red). This class was not observed in the CRP alone sample. This class showed an open decamer of CRP which appears to have two additional particles bound on top of the decamer. We suggest that this may be two SCR-16/20 molecules bound to the open CRP decamer. The poor resolution means it is difficult to confirm if this is the case however it seems likely given that our AUC, SPR and MST results all show SCR-16/20 binding. Similarly, due to the poor

resolution it is difficult to obtain information about the orientation of SCR-16/20 within the complex.

5.4 Discussion

The aim of this study was to investigate the CFH C-terminal binding site for pentameric CRP. SCR-16/20 was identified as a CRP binding site by SPR with a K_D value of 15 μM (Okemefuna *et al.,* 2010). In this study we expressed four SCR fragments, SCR-16/20, SCR-19/20, SCR-16/18H and SCR-17/18H and tested their ability to interact with CRP. Our results revealed that both SCR-16/20 and SCR-19/20 bind weakly to CRP confirming a novel pentameric CRP binding site in SCR-19/20 of CFH. Neither SCR-16/18H nor SCR-17/18H bind to CRP. We attempted to use negative stain EM to capture images of CRP with bound SCR-16/20. From this we observe images of the CRP decamer which may have two SCR-16/20 molecules bound.

From the initial SEC experiments no complex formation was observed between CRP and the SCR fragments. However, on-column dilution and limited resolution may be the reason for this. The CRP and SCR fragment interaction was initially thought to be weak and therefore on-column dilution may well affect complex formation. Secondly the difference in size between CRP and a CRP and SCR fragment complex may not be large enough to be resolved on the SEC column. Therefore more sensitive techniques were employed to identify the binding site. Binding experiments and affinity measurements are complicated by the dimer-forming properties of the SCR fragments (with the exception of monomeric SCR-19/20), and the decamer forming capabilities of CRP (Okemefuna *et al*., 2008; Okemefuna *et al*., 2010b, this thesis, Chapter Four). AUC is not affected by this as the oligomers will be separated under the high centrifugal force, allowing the observation of complex formation. For this reason AUC was used as a starting point to determine which of the C-terminal domains contain the CRP binding site. Conventional AUC using interference optics was not sufficient to observe complex formation between CRP and the SCR fragments since the predicted similarities in sedimentation coefficient between the possible complexes formed and the CRP pentamer-decamer. Therefore FDS-AUC was employed where only the sedimentation of the fluorescently labelled SCR fragments was monitored. AUC was carried out on mixtures with molar ratios at concentrations corresponding to and above the reported K_D value. However no binding was observed for SCR-19/20,

SCR-16/18H or SCR-17/18H. A small peak was visible corresponding to a SCR-16/20 and CRP 1:1 complex in 137 mM NaCl buffers. FITC labelling did not significantly change the sedimentation properties of the SCR fragments although they were found to be prone to aggregation which prevented us from carrying out FDS-AUC experiments at higher concentrations. The FDS-AUC results were inconclusive as to which domains contain the C-terminal binding site as it would be expected that either SCR-16/18H or SCR-19/20 would bind CRP, or that they both would bind depending on which of the domains were required.

SPR experiments were carried out with CRP immobilised on the chip surface. Initial SPR experiments on CRP and CFH used CRP which had dissociated into its monomers on the sensor chip surface, yielding false positive results (Jarva *et al*., 1999). In this study, buffers containing 137 mM NaCl and 2 mM $CaCl_2$ at pH 7.4 were used to ensure CRP remained stable in its pentameric form during labelling and immobilisation. Binding tests revealed that SCR-17/18H and SCR-16/18H do not bind CRP while SCR-16/20 and SCR-19/20 bind weakly. The binding of SCR-16/20 to immobilised CRP was weaker than originally reported (Okemefuna *et al.,* 2010a). Saturation was not reached at concentrations of 70 μM indicating that the K_D is much weaker than 15 μM. Binding experiments can be affected by protein self-association and it is therefore possible that SCR-16/20 dimerisation prevented binding from reaching saturation. The corresponding experiment with SCR-19/20 which does not self-associate showed a similarly weak binding event with an estimated K_D of 124 μM, indicating that the SCR-16/20 and CRP interaction may be similarly weak and that it is not a consequence of self-association.

MST was employed as a surface-free technique to test the binding of the SCR fragments to CRP and to determine the K_D. Similar to SPR it revealed no interaction between CRP and SCR-17/18H or SCR-16/18H. A K_D of 166 μM was estimated for the interaction between SCR-16/20 and CRP which is consistent with the weak binding observed in the SPR experiments. Consistent with the SPR results, SCR-16/20 binding did not reach full saturation even at concentrations of 400 μM. MST measurements are also affected by self-association and it may be that SCR-16/20 dimer formation prevented saturation of the binding curve and result in an inaccurate K_D. The interaction between CRP and SCR-19/20 was difficult to analyse due to aggregation. However it did confirm that the K_D was also in the high micromolar

range, similar to that seen with SPR. The reason for aggregation is relatively unknown. However aggregation was not observed for the SCR-16/20 sample, suggesting that all five domains are required for stabilising the interaction between CRP and SCR-16/20. SCR-19/20 alone is capable of binding to CRP while SCR-16/18H and SCR-17/18H are not - which suggests that the primary binding site of the C-terminus is within SCR-19/20. The hypothesis that the five domains are required for stable binding is corroborated by the fact that full length CFH binds CRP with a K_D of approximately 4 µM in physiological buffer conditions (Okemefuna *et al.*, 2010). The reported binding is much stronger than the binding reported here indicating that physiological binding requires the entire protein. SCR-6/8 also binds CRP with an affinity of 3.9 µM.

Negative stain EM images were taken for CRP alone and for CRP in a mixture with SCR-16/20. The size limit for negative stain EM is approximately 100 kDa so it was therefore not possible to identify SCR-16/20 alone. Images were taken with CRP in 50 mM NaCl, 2 mM $CaCl_2$ buffers as the images obtained under these conditions were of much better quality than those in 137 mM NaCl, 2 mM $CaCl_2$. Images for CRP alone revealed the pentamer as the predominant species with a minority of decamer observed. Besides the observed stacked decamer and the staggered decamer that have been reported in the literature (Ramadan, *et al.*, 2002; Gillon *et al.*, 2015), we also observed an open clam-like shaped decamer where two adjacent CRP monomeric units of two pentamers interact (Figure 5.10 A). The significance of the different conformations of the CRP decamer is not known. However the ability of CRP to form decamers may allow it to further decorate damaged host or pathogenic surfaces, amplifying the immune response to that surface. By comparing images with and without SCR-16/20 we were able to identify images of CRP which appear to have SCR-16/20 bound. This seems to show two SCR-16/20 molecules bound to the open CRP decamer. From our AUC, SPR and MST which confirms that SCR-16/20 binds CRP it seems likely that the images we have observed by EM are of the formed complex. However with the low resolution, and technical limits associated with negative stain EM, this cannot be confirmed. Our SPR and MST results indicate that the binding site is within the SCR-19/20 domains but because of the low resolution of the EM images it is impossible to tell which of the CFH domains are involved in binding or what the binding orientation of CFH is. Interestingly no class was observed

in which one SCR-16/20 molecule was bound to one molecule of CRP. If SCR-16/20 binding only occurs with decameric CRP this could explain the weak K_D calculations from SPR. CRP was immobilized to the SPR chip in very low concentrations in which little decamer would exist on the surface. This would not be an issue in the MST or AUC experiments where CRP is free to form its decamer allowing SCR-16/20 to bind. This again suggests that the binding observed is weak and is not due to protein self-association impairing the binding experiments. Our negative stain EM results provide an interesting insight into how CRP and the CFH SCR-16/20 may interact. Due to the size limits associated with negative stain EM, from our experiments it is difficult to confirm the SCR-16/20 interaction. However an interesting future experiment which would help clarify the CRP and CFH interaction would be to use full length CFH which has been shown previously to be visible by negative stain EM (DiScipio, 1992).

A study in 2009 was published which examined the binding of CFH C-terminal SCR fragments to both pentameric CRP and monomeric CRP (individual 25 kDa subunits) (Mihlan *et al.*, 2009). Like the results described here, this study revealed weak binding of a SCR-15/20 fragment to pentameric CRP using ELISAs with no binding observed for SCR-15/18. This is in agreement with our results where no binding is observed for SCR-17/18H or SCR-16/18H with weak binding of SCR-19/20 observed. Interestingly, this study showed that monomeric CRP bound to SCR-15/20, SCR-15/19 and SCR-19/20 stronger than pentameric CRP. It also revealed that the introduction of aHUS mutations into SCR-20 significantly reduced its binding to monomeric CRP suggesting that the interaction plays a role in disease pathogenesis. Monomeric CRP has been detected on the surfaces of damaged cells and so its ability to bind CFH may play an important role in modulating inflammatory activity (Mihlan *et al.*, 2011).

From our results, which suggest that SCR-19/20 contains the C-terminal pentameric CRP binding site in CFH, we can hypothesize as to how CFH and CRP may interact on cell surfaces. With such weak binding affinities, it seems unlikely that SCR-19/20 will initiate CRP binding. However the much stronger binding of SCR-6/8 may bring the C-terminal into closer proximity with CRP allowing them to interact. Figure 5.11 shows the possible ways in which full length CFH may interact with pentameric CRP during an inflammatory response. It is possible that binding is 1:1 with the stronger binding site at SCR-6/8 initiating binding followed by binding of

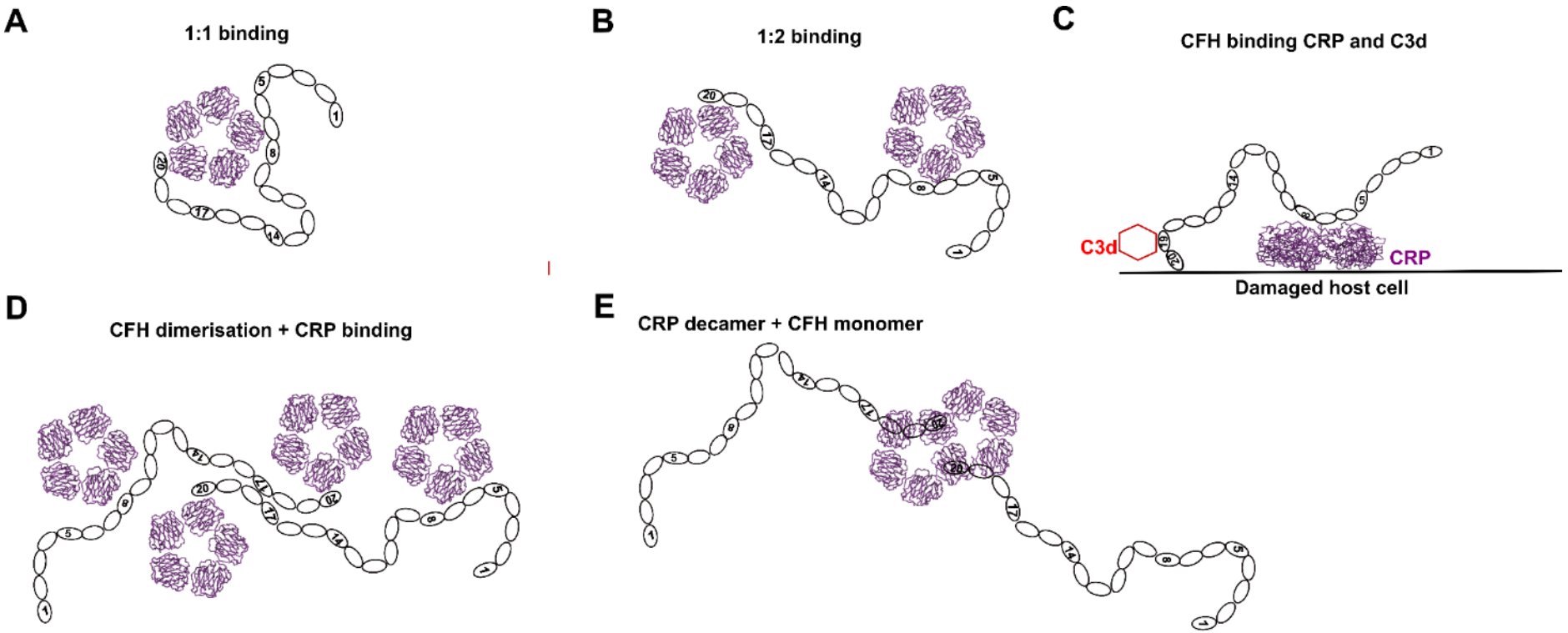


Figure 5.11 legend on the next page

Figure 5.11 A schematic cartoon representation of the possible interactions between CRP and CFH. *A,* An aerial view of a simple 1:1 interaction between CFH and CRP on host surfaces. CFH binds one CRP through its two binding sites, SCR-6/8 and SCR-19/20. *B,* An aerial view of a 1:2 interaction where CFH binds two CRP molecules through each of its binding sites. *C,* A side view of the interaction between CFH and CRP. The weak binding of SCR-19/20 most likely means that when C3d and glycosaminoglycans such as heparin sulphate are present SCR-19/20 will interact with these protecting the surface from complement activation through the alternative pathway. SCR-6/8 is free to bind CRP. *D,* An aerial view of the CFH and CRP interaction with CFH dimerising at SCR-17/18 allowing it to bind four CRP molecules protecting the surface from complement activation. *E,* An aerial view of the possible interaction between CFH and CRP in its open decamer conformation with two CFH molecule bound as observed by negative stain electron microscopy.

SCR-19/20 to the same CRP molecule on the surface (Figure 5.11 A). It is also possible that binding occurs in a 1:2 stoichiometry, and that SCR-6/8 binding site recruits CFH to CRP bound surface, allowing the SCR-19/20 domains to bind to an additional CRP molecule, further protecting the surface from damage (Figure 5.11 B). The weak affinity of SCR-19/20 for CRP means that in the presence of C3d on the damaged host surfaces, SCR-19/20 may interact with C3d preventing excessive complement activation on the surface while SCR-6/8 interacts with CRP (Figure 5.11 C). The CFH C-terminal site is distinct from the CFH C-terminal dimerisation site at SCR-17/18, allowing CFH to concentrate on the surfaces while retaining the capability to bind CRP and protect the surface from immune destruction (Figure 5.11 D) (Okemefuna *et al*., 2008; this thesis, Chapter Four). From our negative stain EM images we suggest a fifth possibility for CFH binding to CRP which has formed its open decamer arrangement on a damaged surface (Figure 5.11 E). In this representation we show two full length CFH binding to decameric CRP which would further protect that surface from excessive immune activation. It is important to note that in Figure 5.11 we use a schematic representation in which CFH is almost fully elongated.

CRP binds damaged host surfaces and pathogenic surfaces triggering the activation of complement through the classical pathway while CFH binds to C3b which has been deposited on host surfaces thereby inhibiting the activation of complement through the alternative pathway. The functionality of the interaction between CFH and CRP is relatively unknown; however there is increasing evidence that impairment of this interaction is involved in AMD disease progression, indicating that it plays an important regulatory role in preventing excessive inflammation and immune destruction (Johnson *et al*., 2006). The AMD associated polymorphism Y402H causes a weakening in the binding of the SCR-6/8 region of CFH to CRP, indicating that this region plays an important role in CRP binding (Okemefuna *et al*., 2010). The reduction of binding in the Y402H SCR-6/8 fragment highlights the potential importance of the CRP and CFH interaction in protecting against immune destruction. Elevated CRP levels have been observed in the choroid of patients with the H402 variant of CFH, suggesting that this reduction in binding may play a role in AMD disease progression with CFH unable to properly regulate CRP at cell surfaces (Johnson *et al*., 2006). A reduction in binding may reduce the amounts of CFH

recruited to damaged host surfaces, leaving them vulnerable to immune attack.

The identification of a binding site in SCR-19/20 also gives an important insight into the possible function of the interaction between CRP and CFH and the role it may play in aHUS. SCR-19/20 is responsible for host cell recognition and C3d binding (Sharma and Pangburn, 1996; Blackmore *et al*., 1998; Jokiranta *et al*., 2000; Blaum *et al*., 2015). It potentially has 31 aHUS-associated mutations (Rodriguez *et al*., 2014). These mutations may interrupt CRP binding during an inflammatory response, leaving the host epithelial cells susceptible to inflammatory attack; they have been shown to interrupt SCR-19/20 binding to monomeric CRP (Mihlan *et al.*, 2009). Concentrations of CRP can rise dramatically during an inflammatory response (up to 500 mg/L (4 μM)) while CFH can have plasma concentrations up to 600 mg/L (3 μM). This means that at the observed K_D for the interaction between full length CFH and pentameric CRP, 24-40% complex formation could occur, thereby preventing excessive immune attack on surfaces in the region of the acute inflammatory response (Okemefuna *et al*., 2010a).

Chapter Six

Deuterium labelling of complement C3d and complement factor H SCR-19/20 for small angle neutron scattering

6.1 Introduction

Small angle neutron scattering (SANS) provides low resolution structural information on biomolecules in solution. Similar to small angle X-ray scattering (SAXS), SANS provides size and shape information about macromolecules that is used to obtain low resolution models (Glatter & Kratky, 1982; Perkins, 2008). SANS is particularly powerful when studying biomolecular interactions, through the use of contrast variation. Using contrast variation methods it is possible to selectively obtain structural information about individual components within a macromolecular complex. The neutron scattering length density (SLD) is described as the sum of all the atomic scattering lengths within the molecule divided by its volume (Perkins, 1988). In a SANS experiment, the resultant scattering is observed due to the difference in SLD between the solvent (buffer) and the macromolecule. If the solvent and macromolecule have the same SLD then no coherent scattering is observed when the buffer is subtracted. For biological systems contrast variation arises from the differential neutron scattering properties of hydrogen (^{1}H) and its heavy isotope deuterium (^{2}H or D). Hydrogen has a negative neutron scattering length and H_2O has a SLD close to zero while deuterium has a positive scattering length with D_2O having a much higher SLD. The SLD of a solvent with an increasing percentage of D_2O:H_2O (v/v) is linear (Figure 6.1). The match point of a molecule is described as the percentage D_2O which has the same SLD as the molecule which means that in a buffer containing that percentage of D_2O no coherent scattering will be observed *i.e.* the molecule will be 'invisible' to the neutron beam. Different classes of biomolecules have different SLD's. During a contrast variation experiment, the macromolecular complex is measured in buffers containing different percentages of D_2O, with measurements taken at the match point of each of the individual components as well as at other contrasts. Babinet's principle validates the use of contrast variation by stating that for a two phase system (buffer and macromolecule), reversing the SLD between the two components will result in the same scattering *i.e.* scattering will be the same if the SLD is changed to be above the match point or below the match point of the molecule (Schoenborn, 1984).

In this thesis we applied this method of contrast variation to study the interaction between complement C3d and the short complement regulator (SCR)

domains 19/20 of complement factor H (SCR-19/20) (Chapter Seven). Two crystal structures of their formed complex have provided ambiguous results, with one showing one C3d molecule bound to one SCR-19/20 (1:1 stoichiometry) and the other showed two C3d molecules bound to one SCR-19/20 (1:2 stoichiometry). It was claimed that the second C3d binding site resulted from crystal packing effects and may not be biologically relevant (Kajander *et al*., 2011; Morgan *et al.*, 2011). Understanding this interaction is important in the context of complement factor H regulation of the complement system. Studies in solution are necessary to clarify the binding stoichiometry and structure of the complex. To carry out contrast variation experiments on a protein-protein complex it is necessary to isotopically label one of the interacting partners with deuterium so that its SLD is changed relative to that of the hydrogenated partner protein. Figure 6.1 shows the calculated SLD for both C3d and SCR-19/20. Both proteins in their hydrogenated states (denoted H-C3d, H-19/20) have a similar SLD, and therefore similar match points at approximately 40% D_2O. By replacing 75% of C3d and SCR-19/20 non-exchangeable hydrogens with deuterium (denoted D-C3d and D-19/20), the SLD's are significantly increased with both proteins having a match point close to 100% D_2O. 100% deuterated C3d and SCR-19/20 (in which 100% of the non-exchangeable hydrogens are replaced by deuterium) are also shown in Figure 6.1 both of which will have a SLD higher than 100% D_2O. By expressing 75% deuterium isotope labelled C3d and SCR-19/20 it will then be possible to carry out a full contrast variation series on complexes of hydrogenated and deuterated SCR-19/20 and C3d. This study will allow us to selectively examine the structure of C3d and SCR-19/20 within the complex in solution. Methods to recombinantly express both proteins in their hydrogenated forms are well established with C3d expressed in *E. coli* and SCR-19/20 expressed in the yeast *Pichia pastoris* (Guthridge *et al*., 2001; Cheng *et al.*, 2005; Herbert *et al*., 2006; Li *et al*., 2008). However no methods have been described to express 75% deuterated C3d or 75% deuterated SCR-19/20, which was the aim of the work presented here.

The gram-negative bacteria *E. coli* is a well-established system for recombinant protein expression. Numerous host strains and expression plasmids are available to produce high yields of recombinant protein (Olins & Lee, 1993). *E. coli* provides a quick and cheap method to produce protein with expression achieved in a matter of hours. The dependence of D_2O content in *E. coli* growth medium on the non-

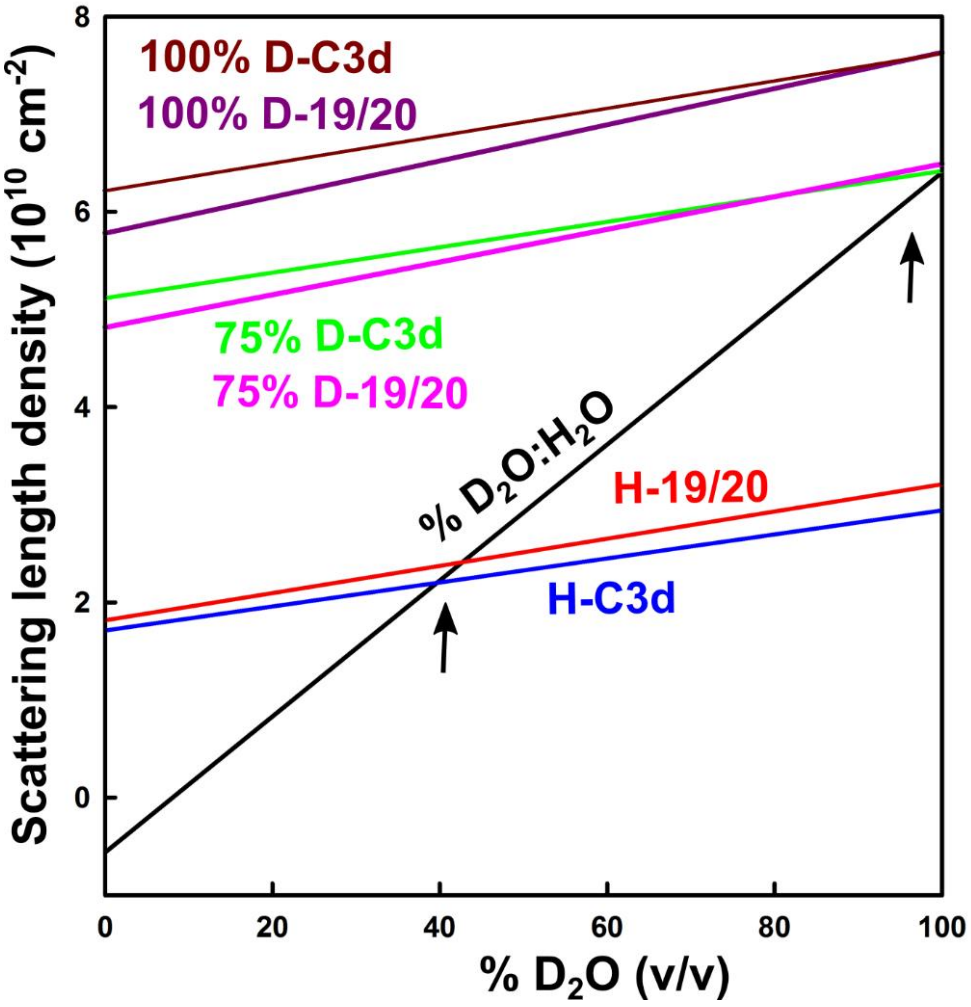


Figure 6.1. An illustration of protein scattering length densities. Hydrogenated C3d (H-C3d), hydrogenated SCR-19/20 (H-19/20), 75% deuterated C3d (75% D-C3d), 75% deuterated SCR-19/20 (75% D-19/20), 100% deuterated C3d (100% D-C3d) and 100% deuterated SCR-19/20 (100% D-19/20) are shown. Scattering length densities are plotted as a function of the volume percentage of D_2O in H_2O/D_2O solvent, assuming 90% of exchangeable hydrogens are exchanged (Perkins, 1986). The black line shows the linear relationship of D_2O percentage with its scattering length density. The match point of each protein is the intersection of this line (arrowed) with that plotted for the protein itself *i.e.* when the scattering length density of the molecule is the same as that of the solvent.

exchangeable hydrogens of proteins has been studied in detail by Lederer *et al*. (1986) as well as by Leiting *et al*. (1998). Media containing the minimal amount of nutrients are required to ensure deuterium incorporation into the proteins amino acids or alternatively deuterated rich media such as Silantes media which contain amino acids in their deuterated forms. To produce a 100% deuterated protein, deuterated carbon sources are required while to produce a protein containing less than 85% deuterium, minimal media in D_2O using hydrogenated carbon sources are sufficient (Leiting *et al*., 1998; Artero *et al.*, 2005). Here we adapt an existing method to express complement C3d in which 75% of its non-exchangeable hydrogens have been replaced by deuterium, denoted D-C3d.

While the use of *E. coli* is the simplest and quickest way to express recombinant protein, issues with insoluble expression due to improper folding and a lack of post-translational modifications means that it can be necessary to use eukaryotic expression systems. The eukaryotic yeast *Pichia pastoris* expression system is a well characterised system with numerous host strains and expression plasmids available for protein expression (Baneyx, 1999; Macauley-Patrick *et al*., 2005). *Pichia pastoris* allows secretion of the expressed protein avoiding issues associated with incorrect protein folding. Secretion facilitates the formation of disulphide bonds which are often required for correct folding. Additionally, since it is eukaryotic it has the ability to produce proteins with post-translational modifications such as glycosylation which is often necessary for structure-function investigations. Studies have not been carried out on the dependence of D_2O content in *Pichia pastoris* growth media on expressed proteins non-exchangeable hydrogens. However its ability to grow in fully deuterated media has been confirmed (Haon *et al*., 1993). Here we describe a novel method to express SCR-19/20 containing 75 % deuterium in its non-exchangeable positions from *Pichia pastoris*, denoted D-19/20.

Problems encountered in producing deuterated protein include difficulties with adaptation of cells to growth in D_2O, low expression levels due to insufficient growth, and the high costs associated with deuterated media. The level of deuterium in the expressed protein is dependent on the levels of deuterium within the growth media, and on the metabolism of the organism. Hence optimisation of this is required to produce protein which will match out at 100 % D_2O. We present here an adapted method to express 75% deuterated protein in *E. coli* and a novel method to express

75% deuterated protein in *Pichia pastoris.* Both methods use cheap hydrogenated carbon sources and recycled D_2O thereby reducing the cost of producing deuterated protein. 75% D-C3d and 75% D-19/20 were expressed and purified with yields similar to their hydrogenated forms. Match points were confirmed by SANS to be approximately 100% D_2O. Guinier analyses suggested no large size or shape differences were incurred upon deuteration. This means that it is now possible to attempt to carry out contrast variation studies on the C3d and SCR-19/20 interaction.

6.2 Materials and methods

The development of this general approach for match-out labelling in *Pichia pastoris* (Dunne *et al*, in preparation) occurred in the Institut Laue-Langevin's (ILL) Life Sciences Group in Grenoble, France, and has been exploited for numerous protein systems including human serum albumin (HSA) and lysozyme. All expression and purification was carried out at the Deuteration Laboratory within this group and subsequently deployed for the study of the C3d/SCR-19/20 system. Unless otherwise stated all chemicals and media components were obtained from Sigma-Aldrich.

6.2.1 Preparation of deuterated C3d in *E. coli*

Complement C3d was expressed using the *E. coli* BL21(DE3) cells (Guthridge *et al*., 2001; Li *et al*., 2008). C3d was expressed as a glutathione-S-transferase (GST) fusion protein using the pGEX-2T expression plasmid (GE Healthcare) (Hydrogenated C3d expression is described in Chapter Seven, section 7.2.1). The gene conferring ampicillin resistance was replaced with that for kanamycin resistance using the EZ-Tn5^{TM} <KAN-2>Tnp insertion kit (Epicentre®). Colonies were plated on both ampicillin (100 µg/ml) and kanamycin (40 µg/ml) Luria Bertani (LB) agar plates to select for kanamycin-resistant colonies. Minimal media was prepared which contained 6.86 g/L $(NH_4)_2SO_4$, 1.56 g/L KH_2PO_4, 6.48 g/L $Na_2HPO_4{\cdot}2H_2O$, 0.49 g/L $(NH_4)_2HC_6H_5O_7$, 0.25 g/L $MgSO_4{\cdot}7H_2O$, 1.0 ml/L (0.5 g/L $CaCl_2{\cdot}2H_2O$, 16.7 g/L $FeCl_3{\cdot}6H_2O$, 0.18 g/L $ZnSO_4{\cdot}7H_2O$, 0.16 g/L $CuSO_4{\cdot}5H_2O$, 0.15 g/L $MnSO_4{\cdot}4H_2O$, 0.18 g/L $CoCl_2{\cdot}6H_2O$, 20.1 g/L EDTA), 5 g/L glycerol, and 40 mg/L kanamycin. The BL21(DE3) cells containing the DNA construct were firstly adapted to growth in minimal media using a step wise process in which cells were inoculated into minimal media, grown for 36 hours at 37 ºC and then diluted into fresh minimal media. This was repeated five times until cells grew to an OD_{600nm} of 3 (1 cm pathlength) in 24 hours. Adaptation was carried out in sterile plastic Erlenmeyer baffled flasks (Sigma-

Aldrich) using an incubation shaker at 200 rpm (INFORS HT). For the preparation of deuterated minimal media, the mineral salts were prepared as per the recipe above but were dissolved in 85 % D_2O. Cells were then adapted to growth in deuterated media via the same method as described above for the minimal media adaptation. When cell densities reached an OD_{600nm} of 3 after 24 hours of growth, large scale growth and expression were carried out. A 3 L fermenter (Infors, France) was used for high cell density growth. 100 ml of deuterated minimal pre-culture was inoculated into 1.5 L deuterated minimal media. The initial growth step was carried out at 37 ºC. The pD (similar to the pH measurement but with deuterium ions rather than hydrogen and normally taken to be 0.5 units different from pH) was maintained throughout fermentation at 6.9 (by addition of 5 M NaOH which had been prepared in 85% D_2O). The air flow rate was maintained at 0.5 L min^{-1} with stirring adjusted to maintain the oxygen levels (pO_2) at 30 % saturation. When the OD_{600nm} reached 5.1 the fed-batch phase was initiated whereby glycerol from a stock solution at 15% (v/v) was added to the culture to keep the growth rate stable during fermentation. When the OD_{600nm} reached 14.7 (30 h), the temperature was reduced to 16 ºC and expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubation continued for 24 h (Figure 6.2). 72.8 g of cells were then harvested by centrifugation at 12,000 rpm (Avanti-J30I, Beckman Coulter) and stored at - 80 ºC.

6.2.2 Preparation of deuterated SCR-19/20 in the yeast *Pichia pastoris*

SCR-19/20 was expressed in the yeast *Pichia pastoris* using the X-33 cells (hydrogenated SCR-19/20 expression was described in Chapter Four, section 4.2.1). The pPICZαA expression plasmid was used which contained a zeocin resistant gene (Cheng *et al.*, 2005; Herbert *et al*., 2006). Adaptation and large scale expression were carried out using sterile plastic Erlenmeyer baffled flasks using an incubation shaker at 29 ºC with shaking at 220 rpm. X-33 cells which had previously been transformed with the recombinant plasmid coding for SCR-19/20 were adapted to growth in a minimal media. Hydrogenated minimal media consisted of 13.4% yeast nitrogen base, 0.02 % biotin, 1% glycerol and 100 mM potassium phosphate, pH 6.0. Adaptation was carried out by a similar protocol to that of the *E. coli* system with the cells grown for 48 hours before transferral into fresh minimal media. Cell growth was monitored by OD_{600nm} measurements with a 1 cm pathlength. When the OD_{600nm} reached approximately 5 after 48 hours of growth, cells were then adapted to growth in 85%

deuterated minimal media by the same process as for hydrogenated minimal media. Deuterated minimal media was prepared as per the hydrogenated media above in 85% D_2O. Large scale expression was then carried out by sustaining cell growth for 72 hours in 1 L of deuterated minimal media containing 1% glycerol (250 ml per 2 L baffled flask to allow sufficient aeration). After 72 hours when the OD_{600nm} was approximately 5, cells were harvested by centrifugation at 2,000 rpm and re-suspended in 330 ml deuterated minimal media with 0.5% hydrogenated methanol in place of glycerol to induce protein expression. Expression was sustained for 96 hours by adding methanol to a final concentration of 0.5% every 24 hours. The final culture was centrifuged at 9,000 rpm with the media supernatant retained which contained the secreted protein.

6.2.3. Protein purification

D-C3d was purified from the BL21 DE3 cells. Cells were resuspended in 25 mM Tris-Cl, 0.5 M NaCl, 1 mM EDTA, 1 mM DTT, pH 8.0 and sonicated. Cellular debris was removed by centrifugation at 12,000 rpm for 30 minutes. D-C3d was purified via its GST tag using a GSTrap glutathione sepharose column connected to a HiTrap Benzamidine FF column (GE Healthcare). GST tag cleavage was carried out on the column using thrombin. C3d was eluted by washing with the column wash buffer (25 mM Tris, 500mM NaCl, 1mM EDTA, pH8.0) and 1 mM AEBSF hydrochloride protease inhibitor to inactivate the thrombin. GST was eluted from the column using 25 mM Tris, 10 mM reduced glutathione, pH 8.0. Size exclusion chromatography using a Superdex 75 16/600 column (GE Healthcare) was used as a final purification step with a running buffer of 10 mM Hepes, 137 mM NaCl, pH 7.4.

D-19/20 was purified from the *Pichia pastoris* supernatant using a SP FF cation exchange column (GE Healthcare) as per the method described for hydrogenated SCR-19/20 in Chapter Four (Section 4.2.2). Briefly, the supernatant was dialysed into the column wash buffer (50 mM Tris-Cl, 25 mM NaCl, 1 mM EDTA, pH7.4) and eluted using a linear NaCl gradient (25 mM – 1000 mM). Size exclusion chromatography was carried out using a Superdex-75 16/300 column (GE Healthcare) as a final purification step in 10 mM Hepes, 137 mM NaCl, pH7.4. Purity was analysed by SDS-PAGE on a 12 % Tris-Tricine gel under non-reducing conditions. Molecular weight standards were used (PageRuler^{TM} plus, ThermoFisher). Protein concentrations were calculated from their A_{280nm} absorbance readings (Cary® 50 UV-Vis

Spectrophotometer) using the Beer-Lambert law. Protein ε values and molecular weights were calculated from their primary amino acid sequences using SLUV (Perkins, 1986). Protein SLD's and deuterated molecular weights (Figure 6.1) were calculated using the Biomolecular Scattering Length Density Calculator from ISIS (www.psldc.isis.rl.ac.uk/Psldc/). For C3d the molecular weight of the hydrogenated protein was calculated to be 34609 Da with a ε value of 45505 M^{-1} cm^{-1} at 280 nm, which corresponds to an absorption coefficient of 13.2 (1%, 1 cm). The molecular weight of 75% deuterated C3d was calculated to be 36,116 Da. For SCR-19/20 the molecular weight was calculated to be 14699 Da. 75% deuterated SCR-19/20 had a calculated mass of 15,281 Da. The ε value for SCR-19/20 was 27430 M^{-1} cm^{-1} which corresponded to a 280 nm absorption coefficient of 19.1 (1%, 1 cm).

6.2.4. Mass spectroscopy

Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy analysis was carried out on the hydrogenated and deuterated SCR-19/20 fragment through the Mass Spectroscopy platform of the Partnership for Structural Biology (PSB), Grenoble, France operated by Dr. Luca Signor. Mass spectroscopy measurements were carried out at 0.5 mg/ml using a MALDI TOF MS (Autoflex, Bruker Daltonics). Samples in 10 mM Hepes 137 mM NaCl, pH 7.4 buffer were diluted (1:2) into the matrix which consisted of sinapinic acid in acetonitrile/water-0.1% TFA (50:50).

6.2.5. SANS data Collection

Data were collected on the SANS instrument D22 at the ILL, with a wavelength of 6 Å ($\Delta\lambda/\lambda$=10%) at a single detector/collimator configuration (5.6M/5.6M) using a ^{3}He multidetector. To calculate the match point the protein was dialysed against D_2O buffers containing 10 mM Hepes, 137 mM NaCl, pH 7.4. Data for D-19/20 were collected in 0 %, 25 %, 40 %, 75 % and 100 % D_2O buffer. Data for D-C3d were collected in 0 %, 40 %, 50 %, 75 %, and 100 % D_2O. Data was collected in Helma 1mm and 2 mm QS quartz cells (precision 0.001). Scattering was measured for 20 minutes per sample with transmission recorded for one minute. The data were normalised to the incident beam intensity followed by subtraction of the sample holder scattering and instrument background scattering. Data reduction and transmission calculations were carried out using the program GRASP (Dewhurst, 2002). Buffer subtractions and Guinier analyses were carried out using PRIMUS (Konerev *et al.*,

2003). The curves were fitted using the Guinier approximation yielding the radius of gyration (R_G) and the scattering intensity at zero angle (*I(0)*) (Glatter and Kratky, 1982).

The match point was determined by plotting the $\sqrt{I(0)/Ttc}$ *i.e.* the normalised scattering intensity, as a function of the volume % D_2O where *T* is the sample transmission, *t* is the cuvette pathlength (cm) and *c* is the concentration (mg/ml) calculated from the protein absorbance at 280 nm (Perkins, 1986). The contrast match point was taken as the intersect of the x-axis *i.e.* where the normalized *I(0)* or scattering intensity is zero.

6.3 Results

6.3.1 Expression of deuterated C3d in *E. coli*

Colonies which grew only on kanamycin LB agar plates were selected for adaptation to minimal media. The replacement of the ampicillin resistant gene by one for kanamycin resistance was to enable high cell density growth as ampicillin will be degraded in the high density culture by β-lactamase enzymes produced by the bacterial cells. *E. coli* BL21(DE3) cells containing the kanamycin-resistant C3d-pGEX-2t construct were successfully adapted to growth in hydrogenated minimal media and deuterated minimal media. Five changes of media was sufficient for successful adaptation to growth in minimal media followed by five changes for adaptation from minimal media into 85% deuterated minimal media. Successful adaptation was determined by measuring the OD_{600nm} after 24 hours of growth and comparing them to the OD_{600nm} of growth in hydrogenated nutrient rich media (LB) growth. Cells adapted to growth in hydrogenated minimal media had an OD_{600nm} of 2.9, cells adapted to growth in deuterated minimal media had an OD_{600nm} of 2.8, while OD_{600nm} measurements for LB media was 3.1. A fermenter where conditions were tightly regulated was used to obtain a high cell density culture for D-C3d expression. Figure 6.2 shows the fermenter trace monitoring each of the conditions during the growth and expression of the C3d-containing BL21(DE3) cells. The OD_{600nm} was continuously monitored. The temperature was tightly controlled and kept constant. The stirring speed (rpm) was adapted (by the fermenter software) to maintain the O_2 level at 30 % while the pD was maintained at 6.9. Fermenter start cultures had an OD_{600nm} of 0.12, which reached an OD_{600nm} of 14.7 before induction after 24 hours. At completion 72.8

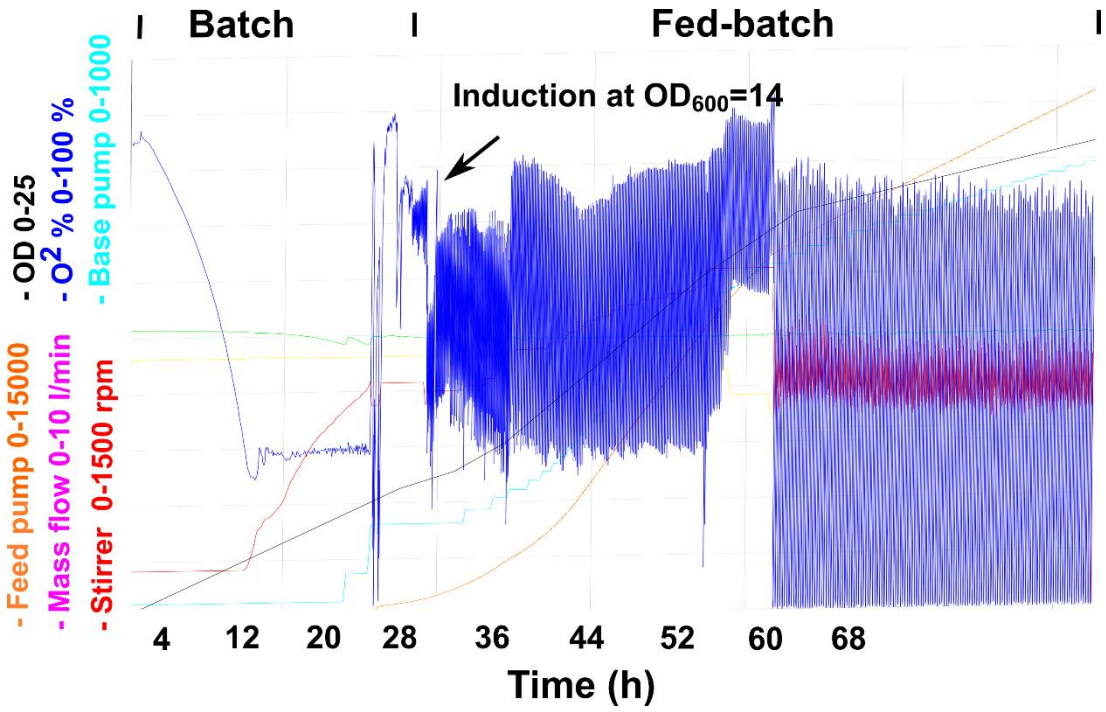


Figure 6.2. The trace of the fermenter parameters for the BL21(DE3) *E. coli* cells. The conditions are tightly regulated to achieve high cell density growth. The black line represents the OD at 600 nm, the blue line represents the percentage of O_2, the pink line represents the mass flow rate and the red line represents the stirring speed. The orange line represents the feed pump which adds glycerol during the fed-batch phase while the base pump shown in green represents the amount of NaOH (in 85% D_2O) added to keep the pD constant at 6.9. Cells were induced at an OD of ~ 14 which was reached after 30 hours.

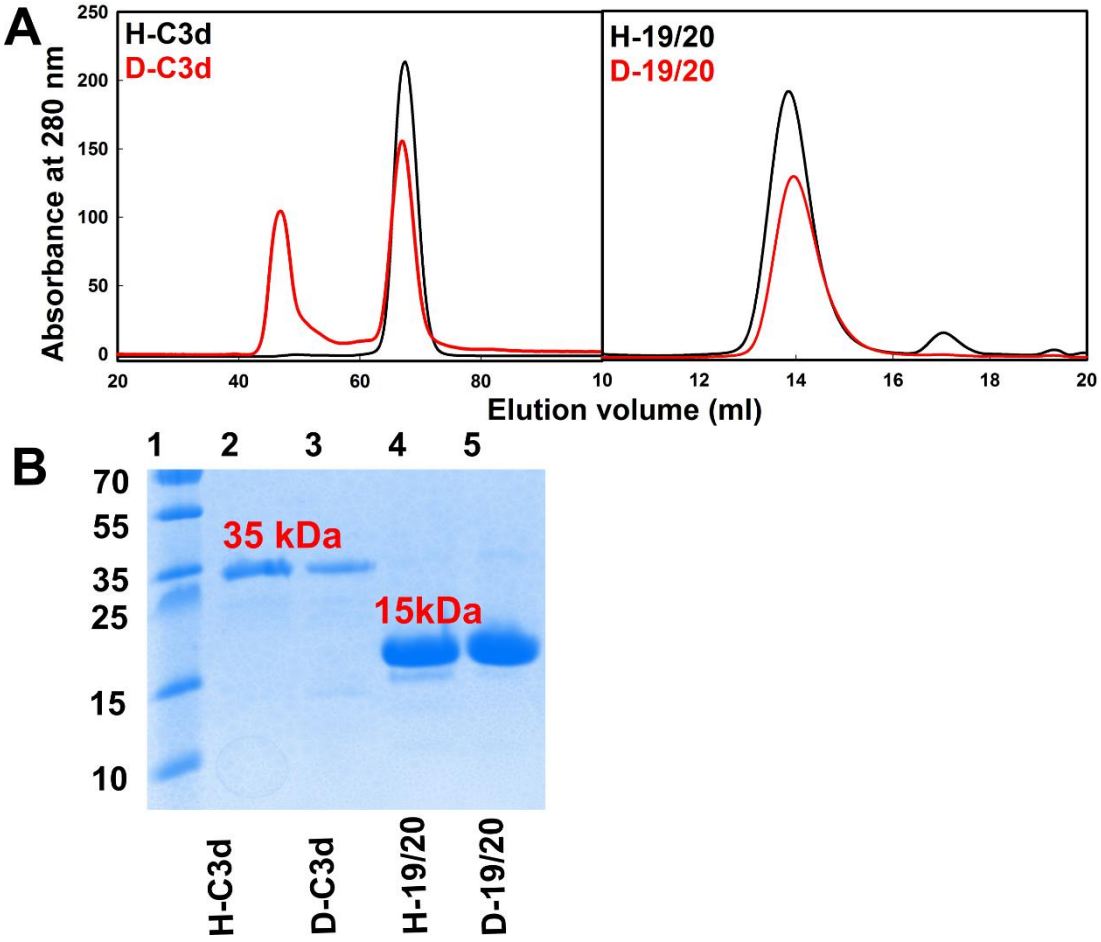


Figure 6.3 Purification of D-C3d and D-19/20 compared to their hydrogenated forms H-C3d and H-19/20. *A*, Left panel, shows the size exclusion elution profiles of H-C3d (black) and D-C3d (red) from a Superdex-75 (16/600). Right panel, H-19/20 (black) and D-19/20 from a Superdex-75 (16/300) column. *B*, the corresponding 12% resolving Tris-Tricine SDS-PAGE analysis in non-reducing conditions. Lane 1, molecular weight standards (PageRuler^{TM} plus); lane 2 H-C3d; lane 3, D-C3d; lane 4, H-19/20; lane 5, D-19/20.

g of cell paste was obtained from 1.5 L 85% deuterated minimal media. The final OD_{600nm} was 24.

6.3.2 D-C3d purification and quantification

D-C3d was purified in batches each starting with approximately 10 g of cell paste. GST purification and cleavage was successful with pooled fractions of D-C3d loaded onto the Superdex 75 size exclusion chromatography column. Size exclusion chromatography revealed that D-C3d was prone to aggregation with two peaks consistently eluting from the column. One peak was close to the void volume (~40 ml) corresponding to aggregated D-C3d while the second peak eluted at a similar position to H-C3d corresponding to monomeric protein (Figure 6.3 A). SDS-PAGE analysis of the monomeric peak revealed a single band corresponding to a molecular weight of 35 kDa similar to that of H-C3d. No significant size change was observed by SDS-PAGE for D-C3d in comparison to H-C3d (Figure 6.3 B). The yield of D-C3d was calculated by measuring the absorbance (A_{280}) of the pooled peak from the size exclusion column corresponding to monomeric D-C3d. The yield for 10g cell paste was 4 mg of soluble D-C3d. Therefore for one 1.5 L fermenter growth and expression, 27.48 mg of D-C3d was obtained. For 1 L growth of hydrogenated nutrient rich media using standard flasks and incubators, 6.7 g of cell paste is obtained with a yield of 6 mg of soluble H-C3d (Figure 6.2 A). Therefore expression yields are similar for the hydrogenated nutrient rich media and deuterated minimal media but D-C3d aggregation results in the loss of approximately 2.5 mg of protein.

6.3.3 Match point calculation for D-C3d

To determine the match point of D-C3d, SANS experiments were carried out to calculate the % of D_2O with respect to H_2O in which D-C3d has a normalised scattering intensity (*I(0)*) of zero (Figure 6.4 A). Measurements were carried out in buffers containing 0%, 40%, 50%, 75% and 100% D_2O with 0.2 mg/ml of D-C3d. Low concentrations were used due to propensity of D-C3d to aggregate. This meant that the scattering curve *I(Q)* suffered from a poor signal-to-noise ratio and consequently it was difficult to carry out Guinier analysis. Nonetheless it was possible to obtain the *I(0)* values but with high error due to the poor quality of data. The *Q* range used for the Guinier fits was 0.16-0.73 nm^{-1}. The *I(0)* was normalised to the sample concentration, cell thickness and transmission which was plotted as a function of D_2O percentage (Figure 6.4 A). The resultant plot showed that D-C3d in buffers

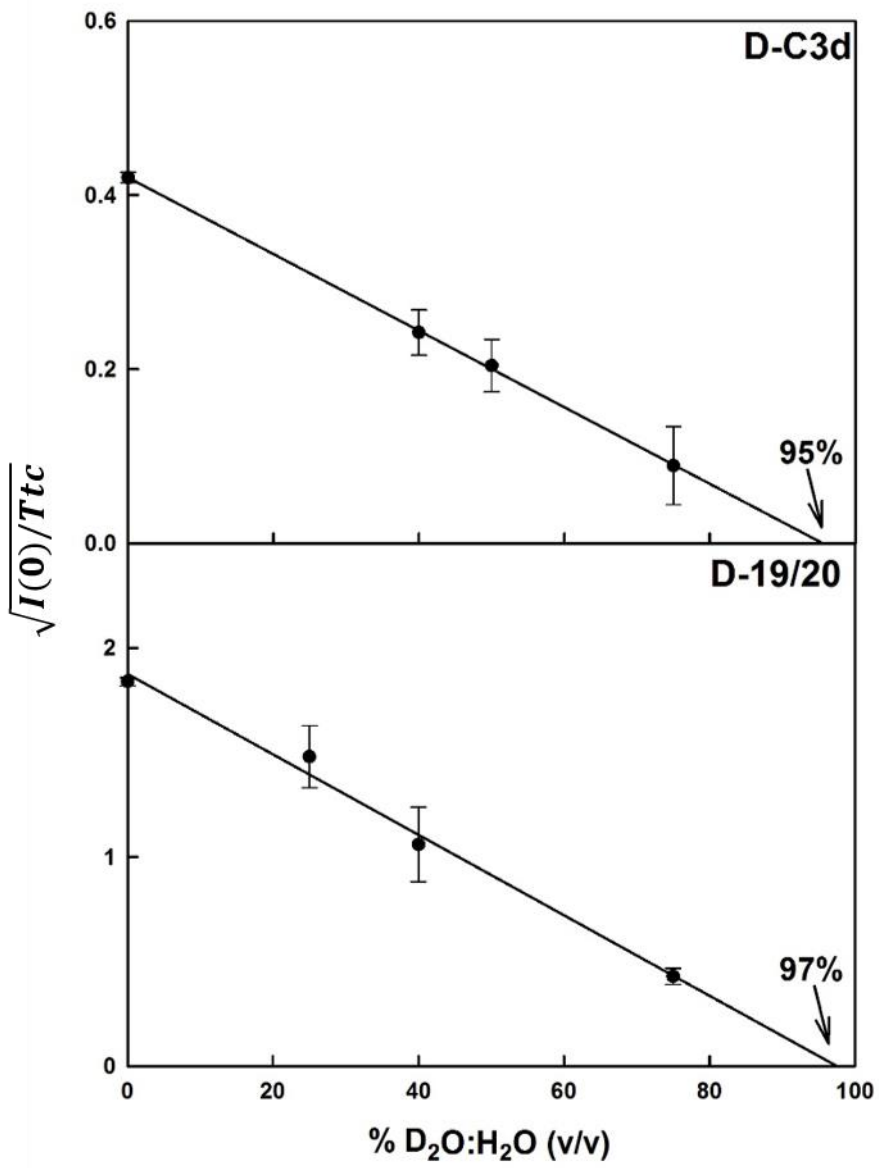


Figure 6.4. Determination of the experimental match point for each of the deuterated proteins by SANS. *A*, The match point for D-C3d expressed in *E.coli* was determined to be 95 ± 6% D_2O. *B*, The match point for D-19/20 expressed in *Pichia pastoris* was determined to be 97 ± 5% D_2O.

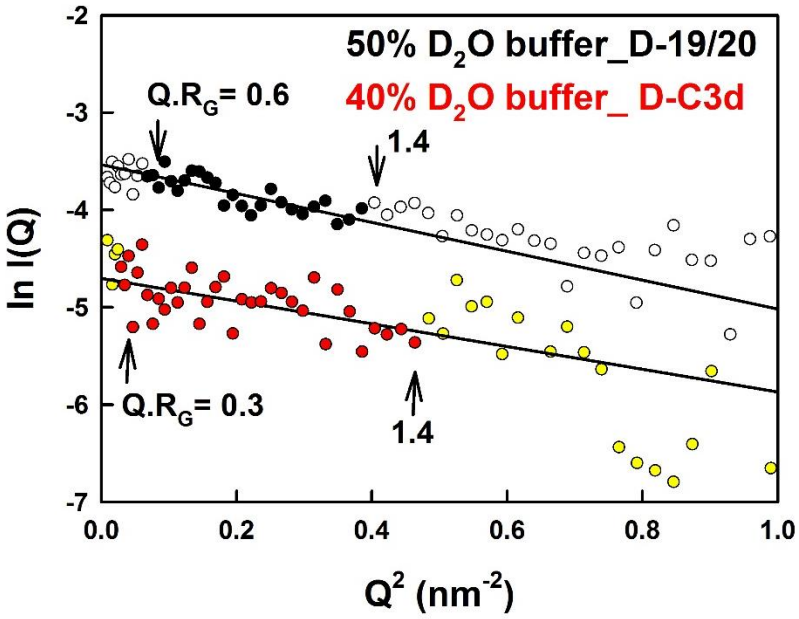


Figure 6.5 Guinier analyses of D-C3d and D-19/20. The Guinier R_G analysis for D-19/20 in 50% D_2O buffer is shown with white circles where the black filled circles highlight the Q range (0.26-0.62 nm^{-1}) used for the Guinier fit. The Guinier R_G analysis for D-C3d in 40% D_2O buffer is shown with yellow circles. The red circles highlight the Q range (0.16-0.73 nm^{-1}) used for the Guinier fit.

containing approximately 95% D_2O had a normalised *I(0)* of zero therefore the match point of D-C3d was 95 ± 6% D_2O (error was calculated from the error on the *I(0)*) confirming that the protein was approximately 75% deuterated at its non-exchangeable hydrogen positions (Figure 6.1). This result also confirmed that 85% deuterated minimal media (as described above) is suitable to produce 75% deuterated protein from *E. coli* for SANS experiments. Figure 6.5 shows the Guinier fit for D-C3d in 40% D_2O. The 40% D_2O buffer sample had a much better signal-to-noise than the other buffers and so is shown while the others are not. The R_G was calculated to be 1.9 ± 0.2 nm which was in agreement with the R_G value for hydrogenated C3d (calculated in Chapter Seven, section 7.3.2.1). This agreement suggested that deuteration did significantly affect the size and shape of C3d although the increase in protein aggregation suggested that it did have a destabilising effect on it.

6.3.4 D-19/20 expression in *Pichia pastoris*

Pichia pastoris X-33 cells were successfully adapted to growth in minimal media and 85% deuterated minimal media. Five changes of hydrogenated media was sufficient to adapt the *Pichia pastoris* to growth in minimal media followed by five changes of media for deuterated minimal media. Each growth was carried out for a longer time period (48 hours) than the *E. coli* adaptation method to account for the slower growth rates of the eukaryotic *Pichia pastoris*. Growth rates were monitored by OD_{600nm} measurements. Adaptation was deemed successful when the OD_{600nm} reached approximately 5 after 48 hours, similar to the OD_{600nm} for hydrogenated nutrient rich media of 3.2. For hydrogenated minimal media the OD_{600nm} was 5.2, while the deuterated minimal media was 4.7. For large scale expression shake flask cultures were used with 250 ml of media per 2L flask to ensure sufficient aeration for the *Pichia pastoris* cells. OD_{600nm} measurements were taken at the end of the glycerol growth step (after 48 hours) which was 5 similar to that for growth in hydrogenated nutrient rich media which was 5.6. The deuterated minimal methanol start culture was prepared to a start OD_{600nm} of 1.0 with expression maintained for 96 hours.

6.3.5 D-19/20 purification and quantification

D-19/20 purification from the supernatant was successfully carried out using cation exchange chromatography. D-19/20 eluted at a similar NaCl concentration as the hydrogenated SCR-19/20 (~700 mM NaCl). Size exclusion chromatography of D-19/20 was similar to hydrogenated SCR-19/20. D-19/20 eluted with a single peak

corresponding to monomeric D-19/20 (Figure 6.3 A). SDS-PAGE analysis of the peak showed a single band corresponding to a protein with a molecular weight of approximately 15 kDa. No significant size change was observed by SDS-PAGE for D-19/20 in comparison to hydrogenated SCR-19/20 (H-19/20) (Figure 6.3 B). The yield of D-19/20 was calculated from the absorbance (A_{280nm}) of the pooled size exclusion chromatography fractions corresponding to the protein peak. From 1 L of glycerol growth media 7.2 mg (on average) of D-19/20 were obtained. This was similar to that obtained for hydrogenated SCR-19/20 in hydrogenated nutrient rich media which on average had yields of 7 mg.

6.3.6 Mass spectroscopy analysis of D-19/20

To confirm that the secretion signalling peptide had been cleaved from D-19/20 in a similar manner to hydrogenated SCR-19/20 mass spectroscopy was carried out. The α-factor secretion signalling peptide is expressed at the N-terminal of SCR-19/20 which signals it for transport to the cell membrane for secretion. It is 9.4 kDa and contains a KEX cleavage site for the peptide to be cleaved as the protein is secreted (Paifer *et al*., 2004). Hydrogenated SCR-19/20 had a mass of 14734 Da determined by mass spectroscopy close to the expected mass of 14699 Da (Figure 6.6 A). Mass spectroscopy showed an increase in mass from 14734 Da for H-19/20 to 15281 Da with a minor peak at 15683 Da also observed for D-19/20 (Figure 6.6 B). The calculated mass of the main peak corresponds to approximately 73 % deuterium in the non-exchangeable positions. This was calculated by calculating the amount of non-exchangeable hydrogens per amino acid of the SCR-19/20 sequence and calculating the change with respect to fully hydrogenated SCR-19/20. The second minor peak is 209 Da larger which could correspond to two additional amino acids being present on D-19/20. This could be from a variation in the cleavage specificity of the signalling peptide by KEX for the deuterated amino acids. However the mass spectroscopy results confirm that the large 9 kDa peptide has been cleaved.

6.3.7 Match point calculation for D-19/20 by SANS

To experimentally calculate the match point of D-19/20, SANS experiments were carried out to calculate the volume percenatge of D_2O in which D-19/20 has a normalised scattering intensity (*I(0)*) of zero (Figure 6.4 B). Measurements were carried out in buffers containing 0%, 25%, 40%, 75% and 100% D_2O with

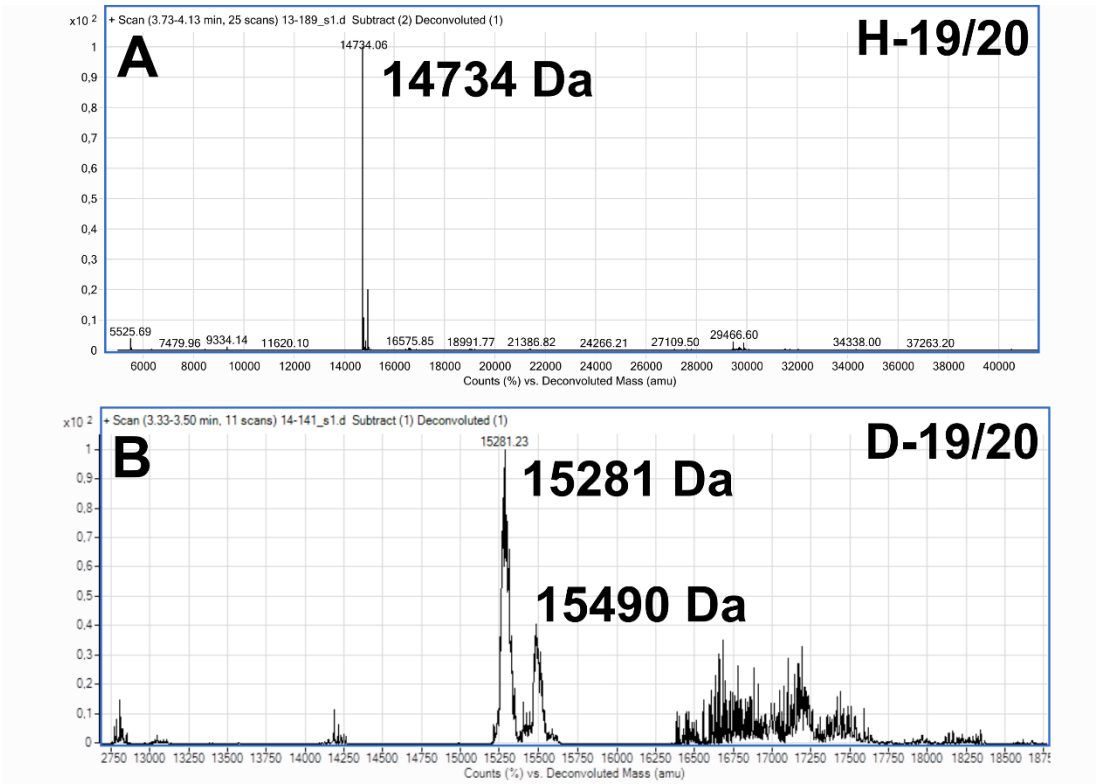


Figure 6.6 MALDI-TOF mass spectroscopy. Mass spectroscopy was carried out on the hydrogenated and deuterated complement Factor H SCR-19/20 fragment. *A*, The spectrum for hydrogenated SCR-19/20 (H-19/20). *B*, The spectrum for deuterated SCR-19/20 (D-19/20). Both are at 0.5 mg/ml in hydrogenated buffer (10 mM Hepes, 137 mM NaCl, pH7.4). The obtained masses are indicated on the spectra.

concentrations of 0.5 mg/ml D-19/20. The small size and low concentration of D-19/20 meant that Guinier analyses were difficult due to a low signal-to-noise. However it was possible to obtain the *I(0)* in each buffer. The *Q* range used was 0.26-0.62 nm^{-1} for the Guinier fits. The normalised *I(0)* against the percentage of D_2O with respect to H_2O in the buffer was plotted and showed that D-19/20 in a buffer containing 97% D_2O had an *I(0)* of zero therefore the match point of D-19/20 was 97 ± 5% D_2O (error was calculated from the error on the *I(0)*) confirming that the protein was approximately 75% deuterated at non-exchangeable positions. This confirms that the 85% deuterated minimal media described above is sufficient to express 75% deuterated protein in *Pichia pastoris*. Figure 6.5 shows the Guinier fit for D-19/20 in 50% D_2O buffer. As can be seen in the graph the data was noisy however the R_G was determined at 2.2 ± 0.2 nm. This value was in excellent agreement with the R_G value for hydrogenated SCR-19/20 (Chapter Four, section 4.3.4.1) indicating that deuteration does not significantly affect the size and shape of SCR-19/20.

6.4 Discussion

We present two methods to express high yields of protein which contain 75% deuterium in non-exchangeable positions that can be used for contrast variation SANS experiments. D-C3d is expressed in *E. coli* using a method which had been adapted from a previous study (Artero *et al.*, 2005) while deuterated SCR-19/20 is expressed in *Pichia pastoris* using a novel method. These two methods are inexpensive to carry out and result in protein yields similar to those in hydrogenated nutrient rich media, both of which are desirable for structural studies which often require large amounts of protein. Guinier analyses of both proteins show that they retain a similar size and shape to their hydrogenated forms suggesting that deuteration has little effect on their structures. The successful expression and purification of D-C3d and D-19/20 will allow us to carry out contrast variation studies on their formed complex enabling us to obtain structural information about each of the individual proteins within the complex.

6.4.1 Expression in *E. coli*

D-C3d was expressed in BL21(DE3) *E. coli* cells using a minimal media which consisted of essential salts, hydrogenated glycerol and 85% recycled D_2O. Methods using *E. coli* as an expression system are available to produce fully and partially deuterated protein however these methods often use expensive minimal medias and deuterated carbon sources for growth and also suffer from low expression levels due

to slower growth rates in minimal medias (Paliy & Gunasekera, 2007; Sugiyama *et al.*, 2011). Deuterated acetate has been used to reduce the costs in comparison to deuterated glycerol. However this results in slower growth rates and low yields of expressed protein when compared to expression in hydrogenated nutrient rich media (Venter *et al*., 1995). The method presented here has been adapted from an existing method used to express fully deuterated protein in *E. Coli* (Artero *et al*., 2005). The method described in this chapter overcomes the expense associated with producing 100% deuterated protein in *E. coli* by using hydrogenated carbon sources and recycled D_2O. By extensively adapting the cells to growth firstly in hydrogenated minimal media and then into deuterated minimal media we obtain cultures that have similar growth rates to those grown in hydrogenated nutrient rich media. By carrying out large scale growth and expression in a fermenter using well adapted cells we obtained a similar yield of protein to the hydrogenated form. The use of high cell density cultures also helps reduce the cost associated with producing large volumes of deuterated protein. Here we obtained approximately six times more cell paste using a fermenter than obtained from the same volume of start culture using regular flask cultures.

Approximately 30% D-C3d is lost during purification due to aggregation which is not observed with H-C3d. Deuteration can affect the stability of proteins due to the weakening of the hydrogen bond networks. This effect is particularly observed in globular proteins such as C3d in which its structure is stabilised through a strong hydrophobic effect (Nagaar *et al*., 1998; Efimova *et al.*, 2006). This deuteration method has been successfully applied to other proteins. The majority of publications which used this method do not discuss protein expression in detail and generally do not give information about protein yields and so it is difficult to compare with our results. However as the publications report successful SANS experiments it is reasonable to assume that they obtained high yields of soluble proteins (Rochel *et al.*, 2011; Appolaire *et al*., 2014). One publication which used this method stated low yields of expression (0.75 mg/L). However they did not state yields for the hydrogenated protein and the low expression may be due to the protein rather than the conditions of expression (Cuypers *et al*., 2013). Another publication described the expression of three deuterated proteins by this same method. All three had similar growth rates to the growth in hydrogenated media, two had similar expression levels, while for the third protein levels were significantly reduced due to the formation of

insoluble inclusion bodies which will be protein specific (Breyton *et al*., 2013). The BL21(DE3) cell line was used here which readily adapted to growth in deuterated minimal media. BL21(DE3) is a commonly used cell line which is compatible with a number of expression plasmids meaning that this deuteration method is suitable for a large number of proteins and can be adapted to the needs of various plasmids and expression systems (Rosano & Ceccarelli, 2014).

6.4.2 Expression in *Pichia pastoris*

D-19/20 was expressed in the *Pichia pastoris* X-33 cells using a minimal media containing nutrients and hydrogenated glycerol for growth and hydrogenated methanol for induction in 85% recycled D_2O. For *Pichia pastoris,* limited methods are available to express 100% deuterated protein and 85% deuterated protein but prior to this study none were available to express 75% deuterated protein for SANS experiments. The available methods use expensive deuterated glycerol and deuterated methanol and obtain similar yields of protein when compared to expression in hydrogenated nutrient rich media (Massou *et al*., 1999; Pickford & O'Leary, 2004). The method we present here uses cheap hydrogenated carbon sources, cheap media components and recycled D_2O which reduces the cost in comparison to existing deuteration strategies. D-19/20 was expressed in a soluble form with no aggregation observed. Yields were similar to those obtained using hydrogenated nutrient rich media. In this study we used regular shake flask cultures as protein yields were similar to yields for hydrogenated proteins. However this method is also applicable to high density fermenter growth for proteins which are not expressed in such high quantities. Our method for expressing match-out deuterated protein in *Pichia pastoris* means that proteins which express in an insoluble manner in *E. coli* or proteins which require post-translational modifications may now be deuterated for SANS experiments.

6.4.3 Conclusion

While the two methods described here have been discussed in the context of contrast variation SANS experiments, deuterium isotope labelling is also employed in other techniques such as nuclear magnetic resonance (NMR) and neutron reflectometry (Venters *et al*., 1995; Heinrich, 2015). Therefore the methods presented which will be easily adapted to other systems can be used to express deuterated protein for a wider range of techniques. In conclusion we have successfully expressed match-out deuterated proteins in both *Escherichia coli* and *Pichia pastoris* which are suitable

for use in contrast variation SANS experiments. We confirmed that the match points were close to 100% D_2O by SANS experiments. The labelling protocol we have described for *Pichia pastoris* is novel and hopefully will have an important impact on future SANS experiments. While the method for *E. coli* greatly reduces the cost of producing high quantities of such deuterated protein. The expression of 75% D-C3d and 75% D-19/20 will allow us to now carry out contrast variation experiments on the interaction between the two proteins to examine their solution structures free and in complex.

Chapter Seven

Small angle scattering studies of C3d with SCR-19/20 of complement Factor H in solution

7.1 Introduction

The alternative pathway of complement is activated through the spontaneous hydrolysis of C3 leading to the formation of the fluid-phase C3 convertase which cleaves C3 into C3a and C3b. C3a cleavage results in a conformational change exposing the C3b reactive thioester domain (TED) which allows C3b to bind to exposed surfaces in a non-discriminative manner. C3b binding acts as an opsonin which targets the surface for immune destruction. Surface bound C3b binds Factor B promoting its cleavage to Ba and Bb. Properdin stabilises the association of C3b and Bb through its recruitment of fluid-phase C3b which then associates with Bb forming the C3 convertase (C3bBb) of the alternative pathway (Smith *et al*., 1984; Hourcade, 2006). This is the amplification loop of the alternative pathway.

Complement Factor H (CFH) is a 154 kDa glycoprotein composed of 20 short complement regulator (SCR) domains consisting of approximately 60 amino acids and joined by variable linker regions of between three to eight amino acids (Rodriguez de Cordoba *et al.*, 2004; Rodriguez de Cordoba & de Jorge, 2008). Structural information on full length CFH is difficult to obtain due to its size, glycosylation and variable flexibility. High resolution information is available for some SCR domains while solution structures and electron microscopy data on full length CFH show that it adopts a partially folded back structure as a result of electrostatic interactions between the SCR domains (Perkins, 1991; Discipio, 1992; Okemefuna *et al*., 2009b). CFH forms dimers through its SCR-6/8 and SCR-17/18 domains (Fernando *et al*., 2007; Okemefuna *et al*., 2008; Chapter Four). CFH regulates the alternative pathway by binding C3b which has been deposited on host cell surfaces, thereby protecting them from immune destruction. It recognises host surfaces through its interactions with surface glycosaminoglycans such as heparin sulphate (Fearon, 1978; Oppermann *et al*., 2006; Blaum *et al*., 2015). CFH which has bound C3b acts as a cofactor for Factor I which cleaves C3c, ultimately leaving the TED C3d moiety bound to the host surface (Whaley & Ruddy, 1976; Pangburn *et al*., 1977).

C3d is the TED of C3 which has been cleaved by Factor I and remains bound to the cell surface. C3d is a 35 kDa protein which has an α-α barrel structure which contains five short segments of 3_{10} helices (Nagar *et al*., 1998). One end of the barrel is acidic while the other end contains the TED and the residues involved in covalent

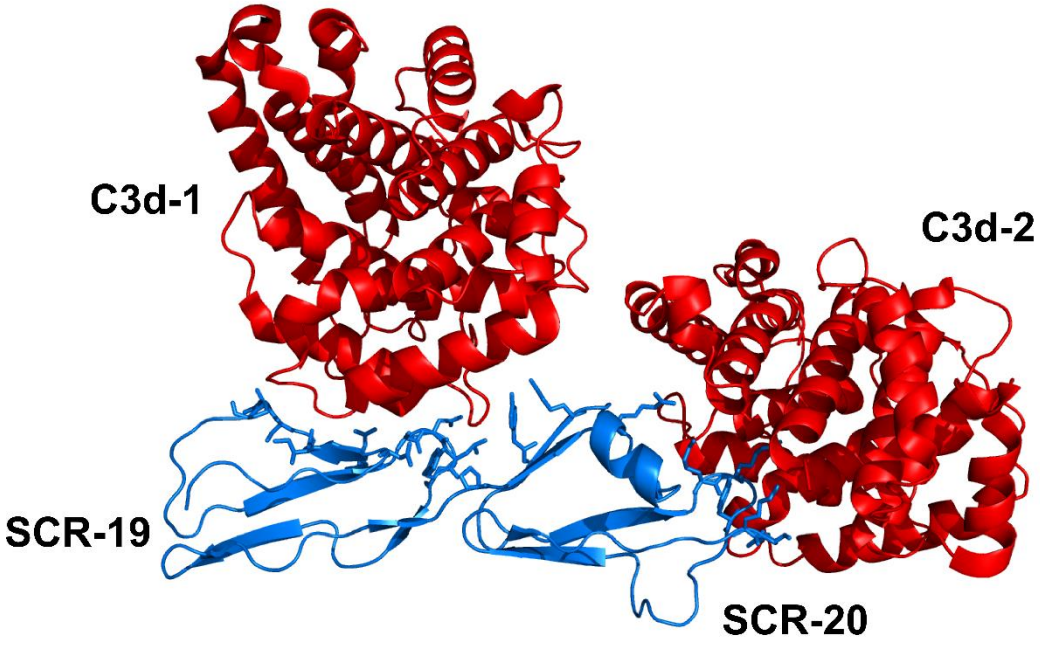


Figure 7.1 The crystal structure of C3d and SCR-19/20. The C3d-1 binding site at SCR-19 was observed in both crystal structures while the C3d-2 binding site at SCR-20 was observed in the second crystal and was deemed biologically relevant. PDB code: 3OXU (1:2) and 2XQW (1:1) (Kajander *et al*., 2011; Morgan *et al*., 2011)

attachment to antigens. C3d functions in bridging the innate and adaptive immune systems through its interaction with complement receptor two (CR2) which is expressed on B-cell surfaces. Surface bound C3d will interact with CR2 thereby eliciting an adaptive immune response (Sahu *et al.*, 2001; Carroll *et al.*, 2004; Toapanta *et al.*, 2006).

Atypical haemolytic uraemic syndrome (aHUS) is a severe disease in which the epithelial cells of the kidney are damaged due to impaired complement regulation. This can lead to kidney failure and is often fatal (Kavanagh *et al.*, 2006). Both SCR-19/20 and C3d contain a number of mutations which are associated with aHUS, suggesting the importance of both proteins in efficient complement regulation (Rodriguez *et al*., 2014).

CFH interacts with C3b through three distinct binding sites, SCR1/4 and SCR-6/10 bind C3b while SCR-19/20 domains bind the C3d portion of C3b (Lambris *et al*., 1988; Sharma & Pangburn 1996; Jokiranta *et al.*, 2000, 2001). In order to understand how CFH regulates C3b and complement, it is necessary to understand how it interacts with C3b. Initial studies on the CFH-C3d interaction were carried out using full length CFH and the five domain SCR-16/20 fragment of CFH. This study revealed multimeric binding which could not be explained by a simple 1:1 interaction (Okemefuna *et al*., 2009a). Subsequently, two crystal structures of the SCR-19/20 and C3d complex were published. Both structures revealed a C3d binding site in SCR-19 with binding contributions from the cleft between SCR-19 and SCR-20 as labelled C3d-1 in figure 7.1. One publication claimed a 1:1 binding stoichiometry for C3d binding to SCR-19, leaving the SCR-20 domain free to bind to surface glycosaminoglycans (Morgan *et al.,* 2011). The second crystal structure showed a second binding site in SCR-20 for C3d, claiming a 1:2 binding stoichiometry (C3d-2 Figure 7.1) (Kajander *et al.,* 2011). Here the thioester is oriented in the same way in both binding sites, indicating that it is structurally feasible for a 1:2 interaction to occur between CFH and surface bound C3d. Given that conditions under which proteins are crystallised may lead to artefacts, it is of crucial importance to examine the interaction with C3d in solution in near physiological conditions that can be related to those relevant to its regulatory role in complement.

Binding studies have clarified that SCR-19/20 interacts with C3d in both 1:1 and 1:2 binding stoichiometries. The SCR-19 binding site is driven by hydrophobic

forces and hydrogen bonding while the SCR-20 binding site is formed by four salt bridges. Two dissociation constants (K_D) were calculated at 1 μM and 8 μM (Rodriguez *et al*., unpublished). The aim of the study presented here was to examine the binding stoichiometry, in solution, of C3d and SCR-19/20, and to obtain structural information of the complex using small angle scattering techniques. Contrast variation with small angle neutron scattering (SANS) indicates that neither C3d nor SCR-19/20 undergo any major structural changes upon binding in comparison to the crystal structures of C3d and SCR-19/20. Small angle X-ray scattering (SAXS) showed that both 1:1 and 1:2 complexes exist in solution with 1:2 as the preferred stoichiometry. This study also confirmed the presence of an electrostatic binding site with the amount of 1:2 complex reduced with increasing ionic strength.

7.2 Materials and Methods

7.2.1 Protein expression and purification

Expression and purification of 75% deuterated C3d (D-C3d) and 75% deuterated SCR-19/20 (D-19/20) is described in detail in Chapter Six. Hydrogenated SCR-19/20 (H-19/20) was expressed and purified from *Pichia pastoris* as described previously (Cheng *et al*., 2005; this thesis, Chapter Four sections 4.2.1, 4.2.2). Hydrogenated C3d (H-C3d) was expressed as a glutathione-s-transferase (GST) fusion protein (Guthridge *et al*., 2001; Li *et al*., 2008). The DNA construct had previously been cloned into the pGEX-2T expression plasmid which was transformed into *E. coli XL*-1 blue cells. Transformed colonies were selected using ampicillin (100 μg/ml) Luria Bertani (LB) agar plates by growth overnight at 37 ºC. Colonies were inoculated in 10 ml LB broth containing ampicillin and grown over night at 37 ºC with shaking at 200 rpm. 5 ml of the overnight culture was then inoculated into 1 L 2xYT media (tryptone and yeast extract) with ampicillin (100 μg/ml) and grown at 37 ºC with shaking at 200 rpm until the OD_{600nm} reached 0.25. The temperature was reduced to 25 ºC and cells were allowed to grow until they reached an OD_{600nm} of 0.4. Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and expression was allowed to proceed overnight at room temperature with shaking at 200 rpm. Cells were harvested by centrifugation and resuspended in lysis buffer (25 mM Tris, 0.5 M NaCl, 1 mM EDTA, 1 mM DTT, pH 8.0). Protein was extracted from the cell by sonication followed by centrifugation. Purification was as described for D-

C3d in Chapter Six (section 6.2.3) via a GSTrap column followed by a benzamidine column with the final purification step of size exclusion chromatography.

7.2.2 Small angle X-ray scattering

To analyse the effects of NaCl concentration on C3d-SCR-19/20 complex, formation data were collected in buffers with varying NaCl concentrations. 10 mM Hepes with 137 mM NaCl, 250 mM NaCl and 350 mM NaCl buffers at pH 7.4 were used. For 1:1 HC3d and H-19/20, molar ratios were prepared with 68 μM of each protein, 1:2 were 68 μM H-19/20 and 136 μM H-C3d. For 1:5 molar ratios of H-C3d and H-19/20, 13 μM of H-19/20 and 68 μM of H-C3d were mixed together.

SAXS data were collected on D-19/20 and on mixtures of D-19/20 and H-C3d in both H_2O and D_2O buffers as a control for the SANS experiments. Experiments were carried out in 10 mM Hepes, 137 mM NaCl, pH 7.4.

Samples were extensively dialysed against the required buffer prior to SAXS data collection. Data collection was carried out on the BioSAXS beamline BM29 at the ESRF, France as described in Chapter Four (section 4.2.4). Briefly, 10 data frames was collected on 50 μl volumes with radiation damaged frames discarded. Buffer was subtracted from the averaged frames yielding the scattering curve *I(Q)*. The radius of gyration (R_G), *i.e.* a measure of the proteins elongation, was calculated from the linear portion of the Guinier plot (ln *I(Q)* v Q^2) within the $Q.R_G$ limit of 1.5. Data analyses were carried out using PRIMUS (Konerev *et al*., 2003). The distance distribution *P(r)* calculated by GNOM (Svergun, 1992) yielded the real space R_G, the maximum dimension of the molecule (L) and the interatomic distance most frequently observed in the molecule (M).

7.2.3 Analysis of SAXS data by the program OLIGOMER

The program OLIGOMER was used to calculate the volume fraction of each of the components in the mixture from the SAXS curves (Konarev *et al*., 2003). OLIGOMER input files were the experimentally obtained scattering curves for monomeric SCR-19/20 and monomeric H-C3d and the theoretical scattering curves of the 1:1 crystal structure (PDB:3OXU) and the 1:2 crystal structure (PDB:2XQW). Theoretical scattering curves were calculated from the crystal structures using the software SCT (Wright & Perkins, 2015). Form factor files for each of the curves were prepared using the FFMAKER tool in the ATSAS package (Konarev *et al*., 2003).

Molecular masses were calculated from the protein primary sequences and the sequences provided with each of the crystal structures.

7.2.4 Dynamic light scattering

Dynamic light scattering (DLS) measurements were carried out on a Malvern Zetasizer instrument with a wavelength of 633 nm. 100 µl quartz cuvettes were used with all measurements at 20 ºC. Measurements were taken in triplicate. Plots of the intensity against size (diameter in nm) were obtained along with a polydisperisty index (PDI) indicating the polydispersity within the sample. The PDI is a measure of the size distribution within the sample.

7.2.5 Small angle Neutron scattering

SANS experiments for H-C3d and D-19/20 alone were carried out as described in Chapter Six with SANS data collected on the instrument D22 at the Institut Laue-Langevin (ILL), France (see section 6.2.6). For the samples containing mixtures of H-C3d and D-19/20, concentrations similar to those for SAXS measurements were prepared as follows; for 1:1 mixtures the concentration was 68 µM H-C3d and 68 µM D-19/20. For 1:2 68 µM D-19/20 and 136 µM of H-C3d were prepared. The mixtures were then dialysed against 0 % D_2O, 40 % D_2O and 100% D_2O buffers - all containing 10 mM Hepes, 137 mM NaCl, pH 7.4. SANS data were collected on D33 at the ILL with a wavelength of 6 Å ($\Delta\lambda/\lambda$ = 10%) at a single detector/collimator configuration (2.0m/5.3m). Data were collected in 1 mm (0% D_2O samples) and 2 mm (40% D_2O and 100% D_2O samples) QS quartz cuvettes with cadmium, empty cell, empty beam, and H_2O measured for data reduction. Data were collected for 30 minutes per sample with buffer data also measured for 30 minutes. Sample transmission and scattering were measured. Transmission was measured via the insertion of beam attenuators with the direct beam passing through the sample recorded.

The 1:1 molar mixture of D-C3d and H-19/20 in 100% D_2O was collected on Sans2d at ISIS, UK. Time-of-flight data were collected with a wavelength range of 0.175 to 1.65 nm, collimation was 4 m and a sample-to-detector distance of 4 m. Data were collected for 3 h in 2 mm quartz banjo cells at a temperature of 20 ºC.

Data reduction and transmission calculations were carried out using the program GRASP (D33) and MANTID (Sans2d) to produce the *I(Q)* scattering curve.

Guinier analysis and distance distribution *P(r)* were carried out by the same method as the SAXS data described above (section 7.2.2).

7.3 Results

7.3.1 Protein expression and purification

H-C3d and H-19/20 were expressed and purified as described. 75% D-C3d and D-19/20 were also expressed and purified as described in Chapter Six with SANS measurements confirming approximately 75% deuterium content in non-exchangeable positions with a match point of approximately 100% D_2O (Chapter Six, Figure 6.4). Each of the proteins were subjected to size exclusion chromatography to remove aggregation and any remaining impurities (Chapter six, Figure 6.3). H-19/20, D-19/20 and H-C3d all eluted with a single peak corresponding to monomeric protein in 137 mM NaCl. D-C3d consistently eluted with two peaks, one for monomeric D-C3d and one which corresponded to aggregated protein in the void volume of the column. SDS-PAGE analysis of the elution from the size exclusion column showed single bands corresponding to 35 kDa for H-C3d and 15 kDa for H-19/20 (Chapter Six, Figure 6.3).

7.3.2 SAXS

7.3.2.1 SAXS of the individual proteins

SAXS data were collected in order to characterise each of the individual proteins. Data were also collected to analyse the effects of deuteration on the SCR-19/20 fragment. This data was collected in 137 mM NaCl buffers. Guinier analysis, within the fit limits of the $Q.R_G \leq 1.3$, was carried out on H-C3d, H-19/20 and D-19/20 to calculate the radius of gyration (R_G) which is a measure of the overall elongation of the molecule (Figure 7.2 A). SAXS data were not obtained for D-C3d alone due to its persistent aggregation. H-C3d had a R_G of 2.2 ± 0.1 nm, H-19/20 had a R_G of 2.4 ± 0.1 nm while its deuterated form had a R_G of 2.5 ± 0.1 nm indicating that deuteration had little significant effect on the shape of SCR-19/20. The distance distribution curve *P(r)* was obtained via an indirect Fourier transform to yield information about each of the individual proteins in real space (Figure 7.2 B). The R_G in real space, the M value or the most frequently observed intermolecular distance, and the maximum dimension or length of the molecule observed (L) were obtained. Each of the R_G values calculated in real space were in excellent agreement with the Guinier analysis. H-C3d had an L value of 7.4 ± 0.3 nm and an M value of 2.5 nm. H-19/20 had an L value of 8.0 ± 0.4 nm and an M value of 1.7 nm. D-19/20 had an L value of 8.2 ± 0.4 nm and an M value

of 1.7 nm, further confirming that deuteration did not affect the shape of the SCR-19/20 fragment. The SAXS results obtained are summarised in Table 7.1.

7.3.2.2 Theoretical SAXS curves for the C3d and SCR-19/20 crystal structure

SAXS data were collected on mixtures containing molar ratios of H-19/20 and H-C3d. The relatively weak binding of C3d and SCR-19/20 meant that the samples measured contained a mixture of monomeric C3d, monomeric SCR-19/20 and the SCR-19/20 and C3d complex (Rodriguez *et al*., unpublished). Therefore the calculated size parameters from the SAXS experiment represent an average of the components in the sample. To obtain insight on the stoichiometry of the binding in the experimental SAXS mixtures, theoretical scattering curves were calculated for the two crystal structures of SCR-19/20 and C3d. By comparing these to the experimental SAXS data of the mixtures it was possible to obtain information about the complexes formed in the sample mixture.

(i) From the 1:1 crystal structure, the theoretical R_G was calculated to be 2.4 nm with an L value of 8.4 nm and an M value of 2.4 nm.

(ii) For the 1:2 complex crystal structure, the theoretical R_G was calculated to be 3.2 nm with an L value of 10.5 nm and an M value of 4.4 nm (Table 7.1).

7.3.2.3 SAXS of the SCR-19/20 and C3d mixtures

Data was collected on 1:1, 1:2 and 1:5 ratios with SCR-19/20 concentration constant in buffers containing 137mM NaCl, 250 mM NaCl and 350 mM NaCl to determine the effects of NaCl on binding. Guiner fits were carried out within the $Q.R_G$ limits (Figure 7.2 A). The results of the SAXS experiments are summarised in Table 7.1.

(i) In physiological NaCl (137 mM) the R_G was 2.8 nm ± 0.01 nm for the 1:1 molar mixture, 3.2 ± 0.01 nm for the 1:2 molar mixture and 2.6 ± 0.3 nm for the 1:5 molar mixture.

(ii) In 250 mM NaCl, the R_G for the 1:1 molar mixture was 2.7 ± 0.01 nm, 2.9 ± 0.01 nm for the 1:2 molar mixture and 2.4 ± 0.2 nm for the 1:5 molar mixture.

(iii) In 350 mM NaCl, the R_G for the 1:1 molar mixture of H-19/20 and H-C3d was 2.6 ± 0.02 nm, 2.8 ± 0.2 nm for the 1:2 molar mixture and 2.4 ± 0.018 nm for the 1:5 molar mixture.

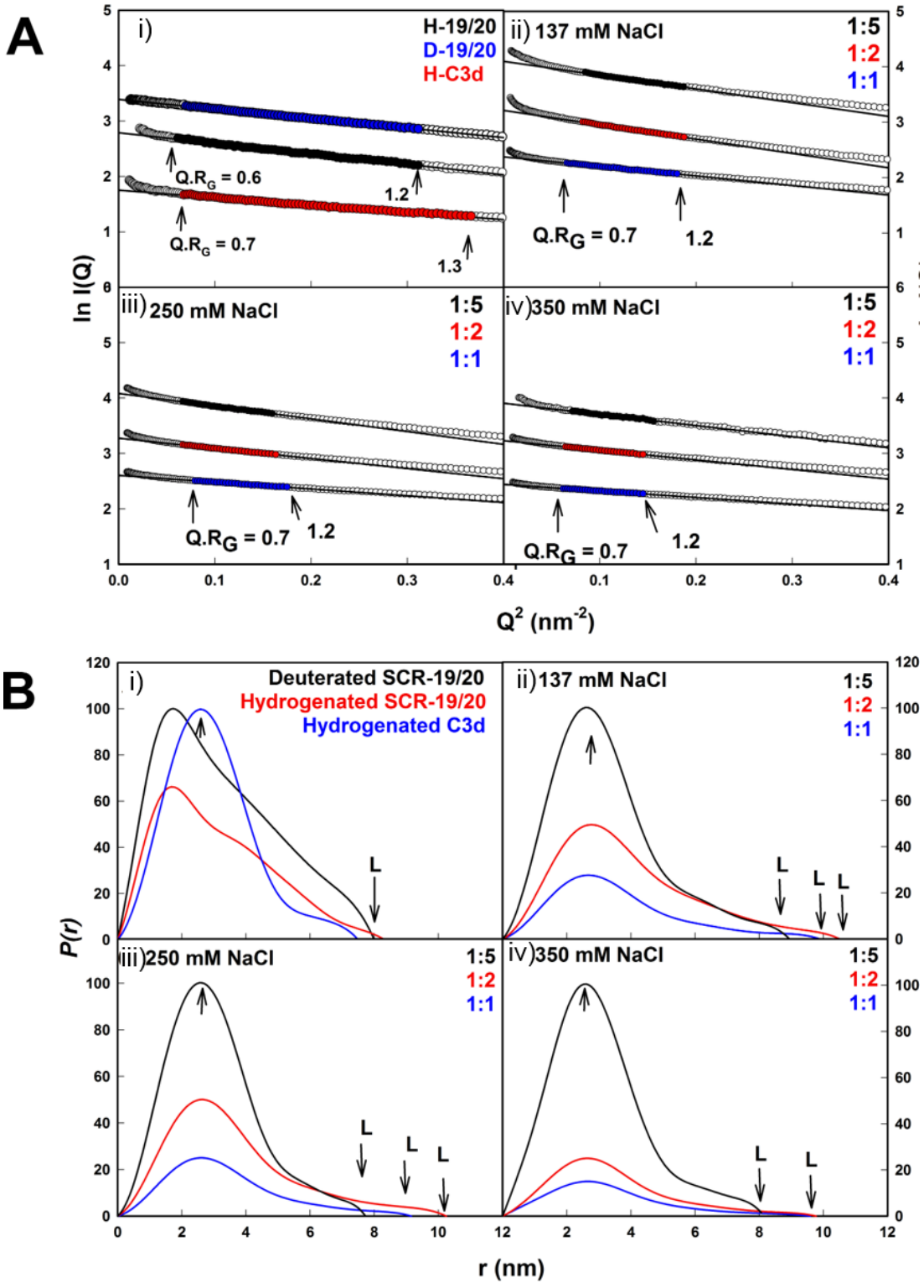


Figure 7.2 SAXS data analyses. A. Guinier R_G analysis of the SAXS data for H-19/20, D-19/20 and H-C3d alone and for mixtures of H-19/20 (SCR-19/20) and H-C3d in molar ratio of 1:1, 1:2 and 1:5. Three different NaCl

(Figure 7.2 continued) concentrations were used. The $Q.R_G$ limits of the linear fit are shown. The *Q* range used was for the Guinier fit was 0.27-0.54 nm^{-1} for H-C3d, and 0.29-0.5 nm^{-1} for H-19/20 and D-19/20, For the mixtures in 137 mM NaCl, the *Q* range was 0.25-0.43 nm^{-1} for the 1:1 mixture, 0.22-0.38 nm^{-1} for the 1:2 mixture and 0.27-0.46 nm^{-1} for the 1:5. In 250 mM NaCl, the *Q* range for 1:1 was 0.25-0.44 nm^{-1}, 1:2 was 0.24-0.41 nm^{-1}, and 1:5 was 0.29-0.5 nm^{-1}. In 350 mM NaCl, the *Q* range for 1:1 was 0.27-0.46 nm^{-1}, 1:2 was 0.25-0.43 nm^{-1} and 1:5 was 0.29-0.5 nm^{-1}. B The distance distribution *P(r)* analyses corresponding to the same samples in A. The arrows indicate the peak maxima which is the interatomic distance (M) most frequently observed while L indicates the maximum observed dimension.

The large R_G indicates that 1:2 complex is present in each of the molar mixtures with the average R_G higher than expected if only monomeric SCR-19/20 and C3d and 1:1 complex were present. For the 1:5 ratio sample, the R_G is lower than that for the 1:1 and 1:2 mixtures. This is due to the lower protein concentration of the 1:5 mixture in which less complex formation will take place. The R_G decreases with increasing NaCl concentration, implying that less complex is being formed at higher salt concentrations (Figure 7.3).

The *P(r)* analysis was carried out with the real space R_G values agreeing with those from the Guinier analysis (Figure 7.2 B).

(iv) In 137 mM NaCl buffer, the 1:1 molar mixture had an L value of 9.9 ± 0.4 nm with an M value of 2.8 nm. The 1:2 mixture had an L value of 10 ± 0.4 nm and an M value of 3.3 nm with the 1:5 molar mixture at 8.9 ± 0.3 nm and an M value of 2.6 nm.

(v) In 250 mM NaCl, the L value for the 1:1 molar mixture was 9.2 ± 0.4 nm and the M value was 2.7 nm. The 1:2 molar mixture gave an L value of 10.3 ± 0.5 nm and an M value of 2.9 nm while the 1:5 ratio had an L value at 7.7 ± 0.3 nm and an M value of 2.4 nm.

(vi) In the highest NaCl concentration of 350 mM, the 1:1 ratio had an L value of 9.6 ± 0.5 nm and an M value of 2.6 nm, the 1:2 mixture had an L value of 9.8 ± 0.5 nm and an M value of 2.8 nm. The 1:5 molar mixture had an L value of 8.2 ± 0.3 nm and an M value of 2.4nm.

Each of the molar mixture samples show a decrease in the R_G and M value with increasing NaCl concentration with no significant change in the maximum observed length (Table 7.1). The decrease in R_G of each of the sample mixtures with increasing NaCl indicates a decrease in complex formation suggesting that binding is driven by electrostatic interactions which are weakened in higher ionic conditions (Figure 7.3).

7.3.3 Calculation of volume fractions

The program OLIGOMER was used to calculate the relative volumes of each species present in the scattering curves and to analyse how the respective volumes change with increasing salt concentration (Konarev *et al.*, 2003). OLIGOMER calculates the volumes based on either experimental scattering curves of the individual components or on theoretical scattering curves of high resolution structures of the

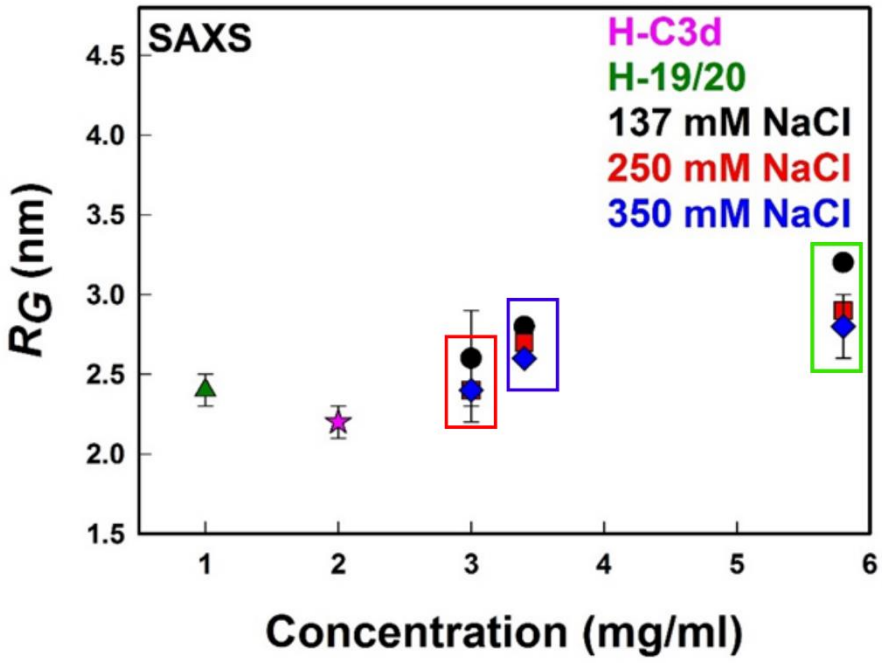


Figure 7.3 The R_G of the H-19/20 and H-C3d complex with respect to NaCl concentration. Shown in the red box are the samples corresponding to 1:5 molar mixtures. Lower concentrations were used in making 1:5 molar ratio mixtures and so the concentration at 3 mg/ml corresponds to 1:5, 1:1 molar mixtures of H-C3d and H-19/20 are shown in the blue box and 1:2 molar mixtures of H-C3d and H-19/20 are shown in the green box.

individual components in the mixture. Four species are hypothesised to be present in the mixture: monomeric C3d, monomeric SCR-19/20, the 1:1 SCR-19/20 and C3d complex, and the 1:2 SCR-19/20 and C3d complex (Rodriguez *et al*., unpublished). Scattering curves for monomeric hydrogenated C3d and monomeric hydrogenated SCR-19/20 and the theoretical scattering curves for the 1:1 and 1:2 crystal structures were used. The volume fractions obtained were considered to be an estimation of the percentage of each species present as the structures of the 1:1 and 1:2 complex may differ in solution in comparison to the crystal structure.

The results obtained from OLIGOMER are summarised in Table 7.2 (A) with the relative volumes converted to percentage for simplicity. Each sample contains SCR-19/20 and C3d alone as expected due to the relatively weak binding. For the 1:1 and 1:2 molar mixture samples, both 1:1 and 1:2 complexes were present with the 1:2 complex estimated to be a higher percentage. 1:5 molar mixture samples consisted mainly of the 1:2 complex.

(i) For the sample containing the 1:1 molar mixture of SCR-19/20 and C3d in 137 mM NaCl, both 1:1 and 1:2 complexes were present. The 1:1 complex accounted for 11% of the total sample while the 1:2 complex was estimated to be 38% of the total sample. The percentage of 1:2 complex present decreased with increasing salt concentration while the percentage of 1:1 complex increased. 1:2 complex decreased to 17% in 350 mM NaCl, while 1:1 increased to 19%. This indicated that the predicted electrostatic binding in SCR-20 site was weakened with increasing salt concentration.

(ii) The 1:2 molar mixture sample was composed of 1:2 complex with no 1:1 complex observed in 137 mM NaCl, suggesting that the higher C3d concentration favoured binding in a 1:2 stoichiometry. In 137 mM NaCl, 67% of the sample existed as 1:2 complex; this percentage decreased with increasing salt concentration. In 350 mM NaCl, 29% was bound in a 1:2 complex. The amount of 1:1 increased from 0% to 12% in 350 mM NaCl. This suggested that, similar to the 1:1 mixture, the electrostatic binding site was weakened with increasing salt concentration.

(iii) For the 1:5 molar mixture, where the concentrations were significantly lower than the 1:1 and 1:2 mixtures, the 1:2 complex was estimated at 40% in 137 mM NaCl with no evidence of a 1:1 complex. The amount of 1:2 complex decreased with salt concentration, although no significant increase was calculated for the 1:1 complex with increasing NaCl; this may be explained by the low concentrations used.

OLIGOMER analyses confirmed that both 1:1 and 1:2 SCR-19/20 and C3d complex formation takes place in solution. It also confirmed that the second binding site was weakened with increasing salt concentration, confirming the presence of an electrostatic binding site for C3d in SCR-20.

7.3.4 Dynamic light scattering

DLS was used to check the quality of the samples for SANS experiments in both H_2O and D_2O buffers containing 137 mM NaCl. The intensity of the light scattered in a DLS experiment is sensitive to aggregation and so was used to determine the effect of solvent on protein aggregation. From the DLS measurement, the polydispersity index (PDI) is calculated. A PDI of less than 0.2 is accepted as being aggregation free.

(i) H-C3d, H-19/20 and D-19/20 were stable and contained no aggregation in H_2O or D_2O buffers with PDI values below 0.2.

(ii) D-C3d persistently aggregated in both H_2O and D_2O buffers with the PDI above 0.5 at concentrations above 0.2 mg/ml. This concentration is too low to carry out SANS experiments that will have good signal-to-noise.

(iii) The presence of H-19/20 had a stabilising effect on D-C3d with a molar mixture of 1:1 having a PDI of 0.16 in 100% D_2O buffer. This meant that it was possible to carry out contrast variation SANS experiments on a mixture of D-C3d and H-19/20.

7.3.5 SAXS as a validation for neutron experiments

SAXS data were collected on samples similar to those to be measured for contrast variation SANS experiments to ensure that the 75% deuterated SCR-19/20 fragment and D_2O buffers did not change the properties of the C3d and SCR-19/20 binding (Figure 7.4). A similar analysis was carried out as described for the SAXS experiments above. 1:1 molar ratios of D-19/20 and H-C3d were measured in 0% D_2O buffer and 100% D_2O buffer in 137 mM NaCl. Concentrations similar to those used for the SAXS experiments on H-19/20 and H-C3d mixtures were prepared to allow a direct comparison. OLIGOMER was also used to calculate the volume fractions of each species present in the D-19/20 and H-C3d sample mixtures (Table 7.2 B).

(i) For a 1:1 H-C3d and D-19/20 molar mixture in 0% D_2O buffer, the R_G was 3.0 ± 0.02 nm, the L value was 10.4 ± 0.5 nm and the M value was 3 nm. These values

Table 7.1 Summary of SAXS results for H-C3d and H-19/20 in different NaCl conditions

Sample	R_G (nm) 137 mM	R_G (nm) 250 mM	R_G (nm) 350 mM	L (nm) 137 mM	L (nm) 250 mM	L (nm) 350 mM	M (nm) 137 mM	M (nm) 250 mM	M (nm) 350 mM
HC3d	2.2 ± 0.1	na	na	7.4 ± 0.3	na	na	2.5	na	na
H-19/20	2.4 ± 0.1	na	na	8.0 ± 0.4	na	na	1.7	na	na
Crystal 1 :1[a]	2.4	na	na	8.4	na	na	2.8	na	na
Crystal 1 :2[a]	3.2	na	na	10.5	na	na	2.9	na	na
ratio 1:1	2.8 ± 0.01	2.7 ± 0.01	2.6 ± 0.02	9.9 ± 0.4	9.2 ± 0.4	9.6 ± 0.5	2.8	2.7	2.6
ratio 1:2	3.2 ± 0.01	2.9 ± 0.01	2.8 ± 0.2	10 ± 0.4	10.3 ± 0.5	9.8 ± 0.5	3.3	2.9	2.7
ratio 1:5*	2.6 ± 0.3	2.4 ± 0.2	2.4 ± 0.01	8.9 ±0.3	7.7 ± 0.3	8.2 ± 0.3	2.6	2.4	2.4

[*]For the 1:5 ratio, a lower total concentration was used in comparison to 1:1 and 1:2 therefore a lower amount of complex is formed

[a] Parameters calculated from the theoretical scattering curves calculated for the 1:1 and 1:2 crystal structures using SCT program

Table 7.2 (A). Volume fractions of the SAXS H-C3d and H-19/20 mixtures calculated by OLIGOMER*

Sample	% volume C3d	% volume SCR-19/20	% volume 1:1 complex	% volume 1:2 complex
1:1 137 mM Nacl	33%	18%	11%	38%
1:1 250 mM NaCl	39%	18%	7%	36%
1:1 350 mM NaCl	13%	51%	19%	17%
1:2 137 mM NaCl	17%	16%	0%	67%
1:2 250 mM NaCl	36%	19%	3%	42%
1:2 350 mM NaCl	29%	30%	12%	29%
1:5 137 mM NaCl	48%	12%	0%	40%
1:5 250 mM NaCl	59%	6%	6%	29%
1:5 350 mM NaCl	47%	27%	0%	26%

Table 7.2 (B). Volume fractions of the SAXS H-C3d and D-19/20 mixtures calculated by OLIGOMER*

Sample	% volume C3d	% volume SCR-19/20	% volume 1:1 complex	% volume 1:2 complex
1:1 0% D2O	21%	23%	11%	45%
1:1 100% D2O	18%	22%	6%	54%

*Volumes converted to percentage from fractions

are slightly increased with respect to the fully hydrogenated samples (from SAXS) and indicate that the H-C3d and D-19/20 complex may have a higher affinity than with H-C3d and H-19/20 with a higher percentage of complex formed at these concentrations. OLIGOMER estimated that 45% of the mixture corresponded to a 1:2 complex while 11% corresponded to a 1:1 complex.

(ii) For the corresponding sample in 100% D_2O, the R_G was 3.0 ± 0.02 nm, the L value was 10.7 ± 0.5 nm and the M value was 3 nm showing that D_2O buffer did not affect the size and shape of the formed complex. In 100% D_2O the percentage of 1:2 complex present increased to 54% with the fraction of the 1:1 complex decreasing to 6% of the mixture, indicating that deuteration of SCR-19/20 may increase the amount of 1:2 complex formed, with D_2O buffer increasing this further.

7.3.6 Contrast variation SANS experiments

Contrast variation experiments carried out on a complex containing a 75% deuterated protein and a hydrogenated protein provide information on the size and shape of each protein individually within the complex (Chapter Six). In 0% D_2O buffer, both proteins will be visible to the neutron beam, in 40% D_2O buffer only the deuterated protein will be visible and in 100% D_2O buffer only the hydrogenated protein will be visible. The D-19/20 and D-C3d match-points were confirmed at approximately 100% D_2O (Chapter Six, Figure 6.4).

7.3.6.1 SANS analysis of H-19/20 and H-C3d in 100% D_2O

H-C3d and H-19/20 alone were measured in 100% D_2O buffer to confirm their scattering at this contrast.

(i) H-C3d alone in 100 % D_2O buffer was measured by SANS. This had a R_G of 1.9 ± 0.03 nm, an L value of 6.5 ± 0.3 nm, and an M value of 2.3 nm (Figure 7.5 *A*, Figure 7.6 *A*). These size parameters are smaller than those observed by SAXS which is expected due to the insensitivity of SANS to the proteins hydration layer.

(ii) H-19/20 was measured in 100% D_2O buffer. H-19/20 in 0% D_2O buffer had a R_G of 2.2 ± 0.1 nm, an L value of 7.5 ± 0.3 nm and an M value of 2.1 nm again showing slightly smaller size parameters than by SAXS (Figure 7.5 *A*, Figure 7.6 *A*).

7.3.6.2 Contrast variation of protein mixtures

Contrast variation SANS experiments were carried out on 1:1 and 1:2 molar ratios of H-C3d and D-19/20 in 0% D_2O, 40% D_2O and 100% D_2O buffers containing

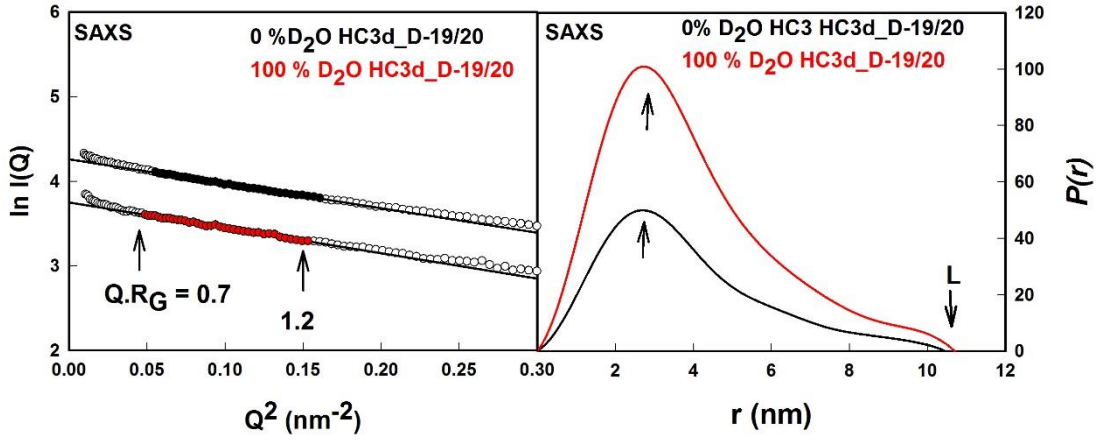


Figure 7.4 SAXS analyses as a control for contrast variation SANS experiments. On the left are the Guinier R_G analyses of the molar mixtures of H-C3d and D-19/20. Coloured circles represent the *I(Q)* values used to determine the RG with the line representing the linear fit to this region. Black is for 0% D_2O samples *i.e.* samples in 100% H_2O buffer and red is for samples in 100% D_2O buffer. The $Q.R_G$ limits are shown. The *Q* range used for the Guinier fits was 0.23-0.4 nm^{-1}.

On the right is the distance distribution *P(r)* analyses for H-C3d and D-19/20 molar mixtures in 0% D_2O (black) and 100% D_2O (red). The arrows indicate the maximum of the peak which is the interatomic distance most frequently observed, while L indicates the maximum dimension observed.

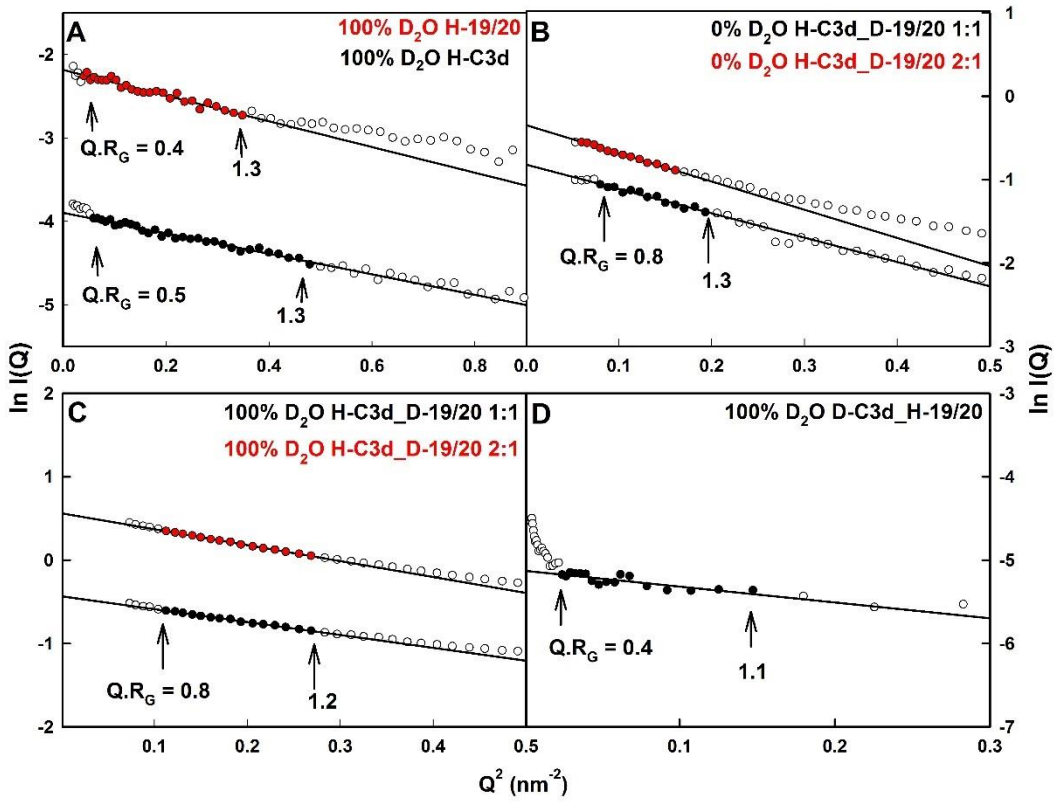


Figure 7.5. Contrast variation with SANS Guinier analyses. *A.* The R_G Guinier analyses for H-C3d and for H-19/20 individually in 100 % D_2O. The *Q* range for the Guinier fit was 0.26-0.68 nm^{-1} for H-C3d and 0.18-0.6 nm^{-1} for H-19/20. The coloured circles represent the *I(Q)* values used to determine the R_G. The $Q.R_G$ limits of the linear fit are show. *B.* The R_G Guinier analyses from the contrast variation SANS experiments of 1:1 (black) and 1:2 (red) molar ratios of H-C3d and D-19/20 in 0 % D_2O. The *Q* range was 0.26-0.43 nm^{-1} for the 1:1 mixture and 0.25-0.41 nm^{-1} for the 1:2 mixture. *C*, The R_G Guinier analysis of 1:1 and 1:2 molar ratios of D-19/20 and H-C3d in 100% D_2O. The *Q* range was 0.36-0.56 nm^{-1} for the 1:1 mixture and 0.33-0.5 nm^{-1} for the 1:2 mixture. *D*, The R_G Guinier analysis for the 1:1 molar mixture of D-C3d and H-19/20 in 100% D_2O. The *Q* range was 0.15-0.48 nm^{-1}. Aggregation is observed at low *Q* (excluded from the Guinier fit).

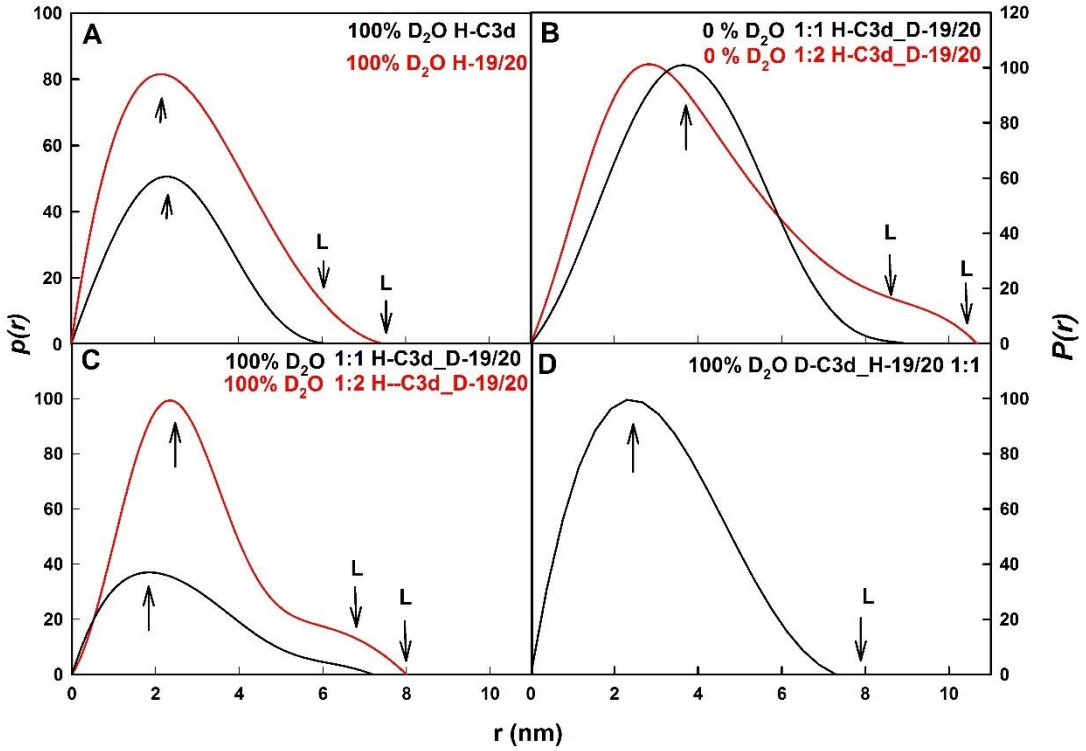


Figure 7.6 The distance distribution function *P(r)* from the contrast variation SANS experiments. The arrow under the peak represents the M value, the most frequently observed interatomic distance, while L represents the maximum observed dimension. *A*, The *P(r)* curve for H-19/20 and H-C3d individually in 100% D_2O. *B*, The *P(r)* curve for 1:1 and 1:2 molar mixtures of H-C3d and D-19/20 in 0% D_2O buffer. *C*, The *P(r)* curve for 1:1 and 1:2 molar mixtures of H-C3d and D-19/20 in 100% D_2O buffer. D, The *P(r)* curve for 1:1 molar mixtures of D-C3d and H-19/20 in 100% D_2O buffer.

137 mM NaCl. Samples in 40% D_2O buffer suffered from poor signal-to-noise ratios, and further analysis was not possible (result not shown). Data were also collected for 1:1 molar mixtures of D-C3d and H-19/20 in 0% D_2O, 40% D_2O and 100% D_2O buffers. The samples in 0% and 40% D_2O buffers suffered from poor signal-to-noise ratios although data analysis was possible for the mixture in 100% D_2O buffer. In common with SAXS data analysis, Guinier analysis and distance distribution *P(r)* function analyses were carried out (Figure 7.5, 7.6).

(i) For H-C3d and D-19/20 samples in 0% D_2O buffer, the 1:1 molar mixture had a R_G of 3.0 ± 0.1 nm, an L value of 9 ± 0.4 nm, and an M value of 2.9 nm, similar to the data obtained by SAXS. For the 1:2 molar mixture, the R_G was 3.2 ± 0.1 nm, the L value was 10.6 ± 0.5 nm, and the M value was 3.2 nm, again similar to those observed using SAXS but with a slight increase in size indicating a higher percentage of complex formation as confirmed by OLIGOMER (Table 7.2 B) (Figure 7.5 B, Figure 7.6 B).

(ii) For the H-C3d and D-19/20 samples in 100% D_2O, the solvent contrast is such that only H-C3d scattering will be observed. For the 1:1 molar mixture in 100% D_2O, the R_G was 2.2 ± 0.1 nm, the L value was 7.2 ± 0.4 nm, and the M value was 2.1 nm. For the 1:2 mixture, the R_G was 2.4 ± 0.05 nm, the L value was 7.2 ± 0.3 nm, and the M value was 2.4 nm. This suggested that no significant changes in size and shape occurred between C3d alone (Table 7.1) and C3d in a complex with D-19/20 (Figure 7.5 C, Figure 7.6 C).

(iii) For the D-C3d and H-19/20 1:1 molar mixture in 100 % D_2O, the solvent contrast is such that only H-19/20 scattering will be observed. The R_G was calculated to be 2.4 ± 0.2 nm, the L value was 7.3 ± 0.4 nm, and the M value was 2.3 nm similar to that obtained for H-19/20 alone in 100% D_2O buffer. This indicated that there were no significant structural changes between SCR-19/20 in complex with C3d and SCR-19/20 alone (Figure 7.5 D, Figure 7.6 D). Aggregation was observed at low *Q*, the Guinier fits excluded this region however the R_G may still be effected by this.

The SANS data was in agreement with the obtained SAXS data in showing that both 1:1 and 1:2 complex formation takes place between SCR-19/20 and C3d. The contrast variation experiments suggest that neither SCR-19/20 nor C3d undergo any major structural changes upon complex formation. However the information obtained is

limited as the sample being studied is a mixture of different components *i.e.* bound and non-bound protein.

7.4 Discussion

This study confirms the presence of both 1:1 and 1:2 binding for SCR-19/20 and C3d in solution. It also confirms the presence of an electrostatic binding site driven by salt bridge formation. Initial studies of the CFH and C3d interaction in solution suggested that multimeric binding took place. However these studies were complicated due to the use of the SCR-16/20 fragment of CFH which can form dimers, making data analysis and interpretation difficult (Okemefuna *et al*., 2009a). The stoichiometry of the C3d and CFH complex has been the subject of some debate following the publication of two different crystal structures. The first structure showed 1:2 binding with two C3d molecules bound to SCR-19/20 (Kajander *et al*., 2011). The second crystal structure showed 1:1 binding between SCR-19/20 and C3d and claimed that the second binding site observed in the Kajander *et al*., structure was an artefact of crystallisation and occurred as a result of crystal packing effects in the crystallographic unit cell (Morgan *et al*., 2011). In order to clarify the correct binding stoichiometry, a study of the SCR-19/20 and C3d interaction in solution was needed. Initial binding studies both in solution and on surfaces showed that both 1:1 and 1:2 binding existed with the primary binding site at SCR-19 stabilised by hydrogen bonds and hydrophobic interactions, while the second SCR-20 binding site was stabilised by the formation of four salt bridges (Rodriguez *et al.*, unpublished). In this study we further confirm using SAXS that both 1:1 and 1:2 binding is present in solution, and that a higher percentage of 1:2 is present than 1:1. It is also shown that the formation of a 1:2 complex is significantly reduced with increasing salt concentration, suggesting that the electrostatic binding site observed in SCR-20 is weakened with increasing ionic strength. Contrast variation SANS experiments confirm that there are no significant size changes of C3d or SCR-19/20 upon complex formation.

Theoretical scattering curves were calculated for the two crystal structures, which show that the size and shape parameters of the 1:1 complex do not change dramatically by comparison with the individual proteins, while for a 1:2 complex a significant increase in the R_G and L values are observed (Table 7.1). In 137 mM NaCl, the size and shape parameters obtained for the 1:1 and 1:2 molar mixtures calculated from the experimental SAXS data indicated that the 1:2 complex was most prominent

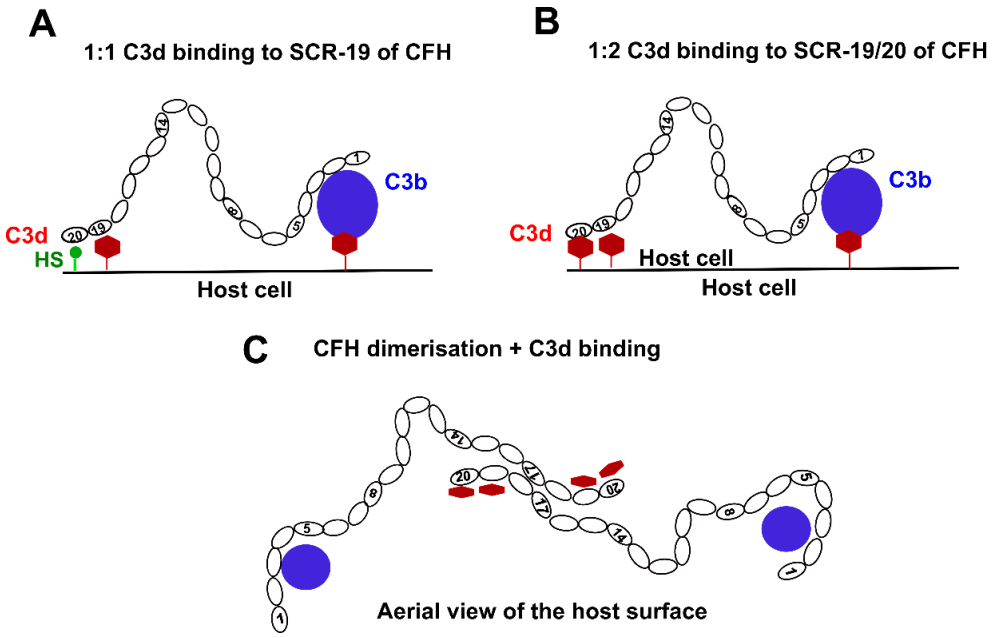


Figure 7.7 A schematic cartoon of how CFH may regulate both C3d and C3b on host surfaces. *A*, CFH binding one C3d molecule through its SCR-19 domain whilst binding the anionic cell surface (heparin sulphate, HS) through its SCR-20 binding site. The SCR-1/4 domains are free to bind surface bound C3b. *B*, CFH binding two C3d molecules on the host surface during a heavy immune response with the SCR-1/4 of CFH bound to C3b. *C*, The possible binding which may occur when CFH forms a dimer through its SCR-17/18 domains. The two SCR-19 domains can bind C3d, with the two SCR-20 domains available to bind either surface glycans or additional C3d molecules. The two SCR-1/4 domains of the CFH dimer can bind two C3b molecules.

with R_G and L values larger than would be expected for a 1:1 complex and closer to those of a 1:2 complex. For the 1:2 molar mixture, the R_G and L values were similar to those calculated for the 1:2 crystal complex indicating that with an increase in C3d concentration, more 1:2 complex will form. Due to differences in concentration the 1:5 molar mixtures cannot be directly compared to 1:1 and 1:2 mixtures; however 1:5 has a R_G of 2.6 nm indicating that more 1:2 complex has been formed, similar to 1:2 where adding an excess of C3d favours 1:2 complex formation. The program OLIGOMER calculated the volume fractions of each component in the scattering curve. This confirmed that the 1:2 complex was most prevalent in each of the molar mixtures, with a larger amount of 1:2 complex formed in the mixtures which have a higher C3d concentration (1:1 and 1:5). 1:5 showed no 1:1 formation in 137 mM NaCl, implying that an excess of C3d favours 1:2 complex formation. The effects of NaCl concentration on the binding was also tested. It was proposed that a reduction in the formation of a 1:2 complex would be observed with increasing NaCl concentration due to the weakening of the proposed SCR-20 electrostatic binding site. The R_G decreases with increasing NaCl concentration, while the L and M values from the *P(r)* do not change significantly. This decrease in R_G suggests that there is a reduction in complex formation which is confirmed by the OLIGOMER results. For both 1:1 and 1:2 mixtures these results reveal that the amount of 1:2 complex formed decreases with increasing NaCl concentration while the amount of 1:1 increases. This strongly suggests that the second C3d binding site which is formed by salt bridges is significantly weakened due to the increasing ionic strength of the buffer, but that a 1:1 complex still exists through the SCR-19 domain interaction with C3d. For the 1:5 molar mixture, this decrease in 1:2 complex with increasing NaCl concentration is also observed However no 1:1 complex is observed for the 1:5 molar mixture which may be as a result of the low concentrations used.

This work illustrates that contrast variation is a powerful technique for the study of protein-protein interactions in solution. It allows individual components within a complex to be studied. To carry out contrast variation experiments on a protein-protein interaction it is necessary to deuterate one of the components so that a full contrast variation series can be carried out as described in Chapter Six. Initial experiments using SEC and DLS showed that D-C3d alone consistently aggregated in both H_2O and D_2O buffers. However, by dialysing D-C3d into D_2O buffers in the

presence of H-19/20 it was possible to collect SANS data on the mixture. To ensure that the deuteration of SCR-19/20 does not affect the size and shape of the protein and complex formation, SAXS experiments were carried out on the deuterated protein alone and in complex with H-C3d. These results revealed no significant change in the size or shape of SCR-19/20 upon deuteration. However the SAXS results on the complex indicated that a higher percentage of complex formed - most probably due to the stabilisation of the hydrogen bonds between H-C3d and D-19/20. SANS data were collected on the individual proteins with H-19/20 in 0% D_2O and H-C3d in 100% D_2O. This confirmed the scattering of both hydrogenated proteins in 100% D_2O while the match points of the deuterated proteins were calculated at approximately 100% D_2O (This thesis, Chapter Six). SANS experiments carried out in 40% D_2O buffers suffered from a very low signal-to-noise ratio, which meant that further analysis was impossible (results not shown). Samples in 0 % D_2O agreed with the SAXS data indicating that a similar amount of complex had formed. The samples in 100% D_2O buffer, where the scattering from only the H-C3d component will be observed indicated that C3d does not undergo any large size or shape changes upon complex formation. An increase was observed in the R_G with respect to the protein alone by SANS which is most likely due to the high percentage of 1:2 complex formed. The R_G increased from 2.2 nm to 2.4 nm from the 1:1 molar mixture to the 1:2 molar mixture indicating a higher percentage of 1:2 complex formation. The absence of any dramatic size or shape changes indicate that C3d does not undergo any dramatic changes upon complex formation. The sample containing D-C3d and H-19/20 shows signs of aggregation at low Q values. However this region was omitted for the Guinier analyses which yielded size parameters similar to those for H-19/20 alone in 100% D_2O. This indicated that, as with C3d, SCR-19/20 does not undergo any large structural changes upon complex formation. The results clearly show that CFH can bind two C3d molecules through its C-terminal SCR-19 and SCR-20 domains in near physiological conditions in solution. The results suggest that 1:2 is the preferred binding stoichiometry but that it can also exist as a 1:1 interaction.

These results also confirm that the SCR-20 domain is driven by the formation of salt bridges which are significantly weakened with increasing salt concentration (Rodriguez *et al*., unpublished). This study was carried out using higher concentrations of protein than at physiological concentrations, and also in the absence of other factors

such as heparin sulphate but nonetheless strongly suggest that both 1:1 and 1:2 binding can occur. The binding affinities for the two binding sites were calculated to be 8 µM and 1 µM suggesting that at physiological concentrations both 1:1 and 1:2 complexes can form (Rodriguez *et al*., unpublished). Additionally the SCR-20 domain is responsible for binding heparin sulphate on host cell surfaces (Opperman *et al*., 2006). The presence of heparin sulphate on host surfaces will allow CFH to bind heparin sulphate through its SCR-20 domain, and bind C3d through its SCR-19 domain. The N-terminal SCR-1/4 domains remain available to bind C3b eliciting immune protection on the host cell (Figure 7.7 A) (Blaum *et al*., 2015) However in times of heavy complement activation and C3b deposition, this second C3d binding site will alleviate the need for CFH to bind to heparin sulphate through its SCR-20 domain and will allow it to bind to two C3d molecules (Figure 7.7 B). Additionally, the identification of the CFH C-terminal dimer site at SCR-17/18 means that the SCR-19/20 domains will remain free to bind either one C3d molecule and heparin sulphate or to bind two surface bound C3d molecules in the CFH dimer. CFH dimer formation at cell surfaces (depending on the orientation) may enable CFH to bind multiple surface bound C3d molecules and two C3b molecules through its SCR-1/4 domains (Figure 7.7 B). Binding of CFH to two C3d molecules which are attached to the host cell will allow CFH to more efficiently protect that host cell from immune destruction. Binding to heparin sulphate can also be mediated through the SCR-7 domain of CFH, allowing CFH to efficiently recognise host cells while interacting with two surface bound C3d molecules and possibly an additional C3b molecule through its SCR-1/4 domains (Blackmore *et al*., 1996). This study, which provides more information on the CFH C-terminal interaction with C3d, will greatly aid the understanding of complement regulation by CFH and subsequently dysregulation by CFH in the disease state, which in the future may aid in the design of potential therapeutic targets for CFH associated diseases.

Chapter Eight

Conclusions

The complement system is a crucial component of the innate immune system which protects its host against invading organisms preventing disease and infection. Complement activation occurs via three pathways; the classical pathway in which C1q interacts with a pathogen or an antibody with a bound antigen, the lectin pathway where the mannose binding lectin protein interacts with specific carbohydrates on pathogenic surfaces and the alternative pathway which is spontaneously activated through the hydrolysis of C3. The three pathways converge at the formation of the C3 convertase which cleaves C3 to its active component C3b and C3a, an anaphylatoxin which initiates the inflammatory response. C3b binds exposed surfaces targeting them for phagocytosis. If the complement cascade is allowed to proceed then the C5 convertase which acts on C5 to produce C5a and C5b, is formed which then brings about the formation of the membrane attack complex. The membrane attack complex inserts into the cell membrane ultimately causing cell lysis.

C3 regulation is crucial to prevent excessive activation of the complement system. Regulation is necessary to ensure that only pathogenic or damaged host cells are destroyed and that healthy host cells remain intact. The spontaneous activation of C3 and the effector molecules produced as a result of complement activation are regulated by a number of fluid-phase and surface bound regulatory molecules. Complement Factor H (CFH) is one of the most important regulators of C3b both in the fluid-phase and on host cell surfaces. CFH can distinguish between host and non-host surfaces through its interactions with surface glycosaminoglycans. It is composed of 20 short complement regulator (SCR) domains. Mutations and polymorphisms within CFH are correlated with immune diseases such as age-related macular degeneration (AMD) and atypical haemolytic uraemic syndrome (aHUS) which are associated with impaired complement regulation. In this thesis the CFH C-terminal functional interactions with its ligands were studied. The aim was to further our understanding of how CFH regulates the complement system and to use this knowledge to understand how dysregulation may lead to disease.

8.1 Complement Factor H self-association

CFH self-associates to form dimers with two dimerisation sites located within SCR-6/8 and SCR-16/20. The C-terminal SCR-16/20 site was initially hypothesised to be located in SCR-20, the region responsible for host cell recognition. SCR-16/20 contains 50 mutations which are associated with aHUS. The majority of these

mutations are located in SCR-19/20 however there are also 19 disease-associated mutations located in SCR-16/18.

In Chapter Four, using a combination of size exclusion chromatography (SEC), analytical ultracentrifugation (AUC) and small angle X-ray scattering (SAXS) it was determined that the C-terminal CFH dimer site is located in domains SCR-17/18. Seven SCR fragments of the CFH C-terminal were expressed in *Pichia pastoris*. SEC and AUC experiments confirmed that each of the fragments with the exception of monomeric SCR-19/20 forms dimers. SCR-17/18 and SCR-17 alone form the strongest dimers with estimated K_D values of 3 μM and 5 μM, respectively. These results strongly suggest that SCR-17 is the main dimer site in the C-terminus of CFH but that SCR-18 is also involved. Examination of the shape of each of the SCR fragments by AUC and SAXS reveal that the dimer is most likely formed by a side-by-side antiparallel arrangement. The existence of this arrangement is further strengthened by the fact that the dimer of the complement factor H related 1 (CFHR1) protein is also formed by a side-by-side antiparallel interaction. A model was constructed for the SCR-17/18 CFH dimer based on the CFHR1 crystal structure which revealed a weak hydrophobic dimer interface. AUC and SAXS experiments in 50 mM NaCl buffers showed no difference in comparison to physiological NaCl buffers. This suggests that electrostatic interactions are not involved, supporting the hypothesis of a hydrophobic interaction.

The presence of the dimer site in SCR-17/18 may explain why aHUS mutations are located in the SCR-17/18 region of CFH. CFH dimerization may be functionally important for CFH providing it with a mechanism by which it can concentrate on host surfaces protecting them from complement activation.

Future experiments are needed to validate the suggested model and to confirm the importance of CFH dimerisation for its function. With the presence of a mixture of components *i.e.* monomer and dimer, it was not possible to construct a model from the SAXS data. Online purification techniques with SAXS, whereby the elution fractions from the size exclusion column pass directly through the X-ray beam, may allow SAXS data to be collected on the dimer alone. This would allow low resolution models to be constructed and compared with the suggested model in Chapter Four. Experiments to determine the effects that the aHUS-associated mutations have on the C-terminal dimer formation would provide an interesting insight into the importance

of CFH dimerisation. These experiments may also yield information about how these mutations lead to a disease state.

8.2 Complement Factor H binds pentameric C-reactive protein

C-reactive protein (CRP) is an acute phase protein which activates complement via the classical pathway through its interaction with C1q. CRP is a pentameric protein composed of five identical monomer subunits. The existence of an interaction between CFH and CRP was initially debated due to the dissociation of CRP into its monomer subunits. However an interaction between CFH and pentameric CRP was confirmed with two CRP binding sites identified in SCR-6/8 and SCR-16/20. The interaction between CRP and CFH has been associated with AMD.

In Chapter Five of this thesis the C-terminal SCR-16/20 CFH binding site for pentameric CRP is investigated using SEC, fluorescence detection AUC (FDS-AUC), surface plasmon resonance (SPR), microscale thermophoresis (MST) and negative stain electron microscopy (EM). Using four recombinantly expressed SCR fragments of the CFH C-terminus, a weak binding site is identified in SCR-19/20 for pentameric CRP. No binding is observed by any method for SCR-16/18H or SCR-17/18H. Weak SCR-16/20 binding is observed by FDS-AUC, SPR and MST. The dissociation constant (K_D) is estimated to be above 100 μM however the accuracy of this is limited due to the propensity of SCR-16/20 to self-associate which may prevent saturation of the SPR and MST binding curves. SCR-19/20 binding is observed by both SPR and MST with an estimated K_D of above 124 μM. The lack of saturation is less easy to explain for SCR-19/20 binding due to its monomeric properties suggesting that CFH C-terminus binding to CRP is weak. Negative stain EM images appear to show two SCR-16/20 molecules bound to CRP which has formed an open shaped decamer. Due to the size limitations associated with negative stain EM, the information obtained about SCR-16/20 is minimal.

From this study a possible mechanism of CFH binding to CRP is proposed whereby the SCR-6/8 domains initiate binding to CRP on the cell surface which brings SCR-19/20 into proximity with the surface allowing it to interact with CRP thereby protecting that surface from immune activation. The distinction between the C-terminal CRP binding site and the C-terminal dimer site may allow CFH to further concentrate on cell surface while retaining its ability to bind CRP on the cell surface. Plasma concentrations of CFH are between 0.2-0.8 mg/L (1.2 - 5 μM) while CRP

levels can rise to 500 mg/L (4.3 μM) in an acute phase response. These concentrations may vary with protein localisation suggesting that it may be possible for CRP and CFH to interact on cell surfaces.

Reversed SPR and MST experiments in which the SCR fragments are bound to the chip surface or are labelled may help to overcome the issues associated with self-association however CRP also self-associates to form decamers and the same issues may be observed. Nonetheless these experiments would provide additional information about the CRP and CFH C-terminal interaction. Negative stain EM experiments using full length CFH and CRP will help to clarify the interaction between CFH and CRP as both CRP and full length CFH are visible by negative stain EM. This experiment may be crucial in determining the binding stoichiometry of the interaction and also the oligomeric state of both CFH and CRP in the formed complex.

8.3 Deuteration of C3d and SCR-19/20 for SANS experiments

Contrast variation with small angle neutron scattering (SANS) is a powerful technique which allows biomolecules within a complex to be selectively observed. Changing the neutron scattering length density of the buffer to that of the biomolecule results in no observed scattering for that molecule. The different classes of biomolecules naturally have different scattering length densities. In order to carry out contrast variation experiments on biomolecules of the same class it is necessary to isotopically label one of the interacting partners with deuterium to change its scattering length density with respect to its hydrogenated partner. In Chapter Six of this thesis two methods are presented to produce deuterated C3d and SCR-19/20 of CFH in which 75% of their non-exchangeable hydrogens are replaced with deuterium. Contrast variation SANS experiments could then be used to examine the interaction between the two proteins.

75% deuterated C3d is expressed in an *E. coli* expression system by adapting an existing method to produce 100% deuterated protein. 75% deuterated C3d is more aggregation prone than its hydrogenated form however the match point at approximately 100% D_2O is confirmed by SANS. This method provides a means to express deuterated protein suitable for contrast variation experiments in the relatively easy and quick *E. coli* expression system. Aggregation is generally protein specific with globular proteins like C3d more effected due to the change in the hydrophobic effect as a result of deuteration.

A novel method is presented for the expression of deuterated protein in *Pichia pastoris* which was successfully used to produce 75% deuterated SCR-19/20. The match point of the deuterated protein was confirmed at approximately 100% by SANS experiments with mass spectroscopy confirming the non-exchangeable deuteration level at 73%. This method means that proteins which require post-translational modifications or need to be secreted for folding purposes can also be deuterated for contrast variation experiments using the *Pichia pastoris* expression system.

Both of the described methods use cheap carbon sources and recycled D_2O which significantly reduces the cost of deuterium isotope labelling. Additionally, both methods resulted in yields similar to those obtained for hydrogenated expression meaning that SANS studies on the C3d and SCR-19/20 interaction could be attempted. The two methods presented here will allow other groups to carry out similar deuterium labelling in either *E. coli* or *Pichia pastoris* expression systems to carry out SANS experiments or other types of experiments which require deuterium isotope labelling.

8.4 C3d interaction with SCR-19/20 in solution

CFH interacts with the thioester domain-containing C3d fragment of C3b which remains on host surfaces upon C3b degradation by Factor I. This interaction prevents the initiation of an adaptive response by preventing C3d binding to complement receptor two. SCR-19/20 of the C-terminal of CFH binds C3d when bound to a non-activator surface. Two crystal structures are available for the C3d and SCR-19/20 complex; one which shows two C3d molecules bound to one SCR-19/20, while the second shows one C3d molecule bound to one SCR-19/20, claiming the second site is observed as a result of crystal packing effects. In order to understand how CFH regulates complement it is necessary to understand how it interacts with C3d on non-activator surfaces. Therefore the aim of Chapter Seven was to clarify the binding stoichiometry between SCR-19/20 and C3d, and to examine the structure of their formed complexes in solution.

SAXS experiments are carried out on mixtures of hydrogenated SCR-19/20 and C3d in 137 mM, 250 mM, and 350 mM NaCl. Using the program OLIGOMER the relative percentage of 1:1 and 1:2 complex in the mixture are calculated. This reveals that both 1:1 and 1:2 complexes form between SCR-19/20 and C3d. It also confirms that with increasing ionic strength the 1:1 binding is significantly reduced

confirming that the second C3d binding site in SCR-19/20 is driven by salt bridge formation.

Contrast variation SANS experiments using the 75% deuterated protein are carried out on deuterated SCR-19/20 (D-19/20) and hydrogenated C3d (H-C3d) and on hydrogenated SCR-19/20 (H-19/20) and deuterated C3d (D-C3d). The amount of information obtained from these experiments is limited as the sample contained a mixture of scattering components. However, it is evident that neither SCR-19/20 nor C3d undergo any major structural changes upon interacting.

From these results it is likely that both crystal structures are representative of the interactions that can occur between C3d and SCR-19/20. In the presence of surface glycosaminoglycans it is likely that SCR-20 will interact with these while SCR-19 will interact with C3d. However, the capability of SCR-19/20 to bind two surface bound C3d molecules will allow CFH to further protect host cells from immune attack especially in times of a heavy immune response.

8.5 Overall conclusion

The work in this thesis provides an insight into the interactions that the C-terminal of CFH undergoes in order to regulate the alternative pathway of complement activation. The three major findings of this thesis are that the CFH C-terminus self-association site is in SCR-17/18, a region which previously had no known function but that contains disease-associated mutations. SCR-19/20 of the CFH C-terminus contains a weak binding site for pentameric CRP which may allow CFH to regulate CRP on host surfaces. The final major finding of this thesis is that SCR-19/20 of CFH can bind to C3d in both a 1:1 and 1:2 binding stoichiometry allowing CFH to efficiently protect host surfaces from excessive complement activation.

Publications in Preparation

Dunne, O., Moulin, Callow, P., Weidenhaupt, Perkins, S., M., Haertlein, M., & Forsyth, V.T., Matchout deuterium labelling of proteins using prokaryotic and eukaryotic expression systems.

Dunne, O., Nan, R., Gor., J., Adamson, P., Gordon, D. L., Moulin, M., Forsyth, V. T., & Perkins, S. J. A novel dimerisation site at SCR-17/18 in Factor H may explain its disease causing mutations and a new mechanism for regulatory control.

Presentations and Abstracts

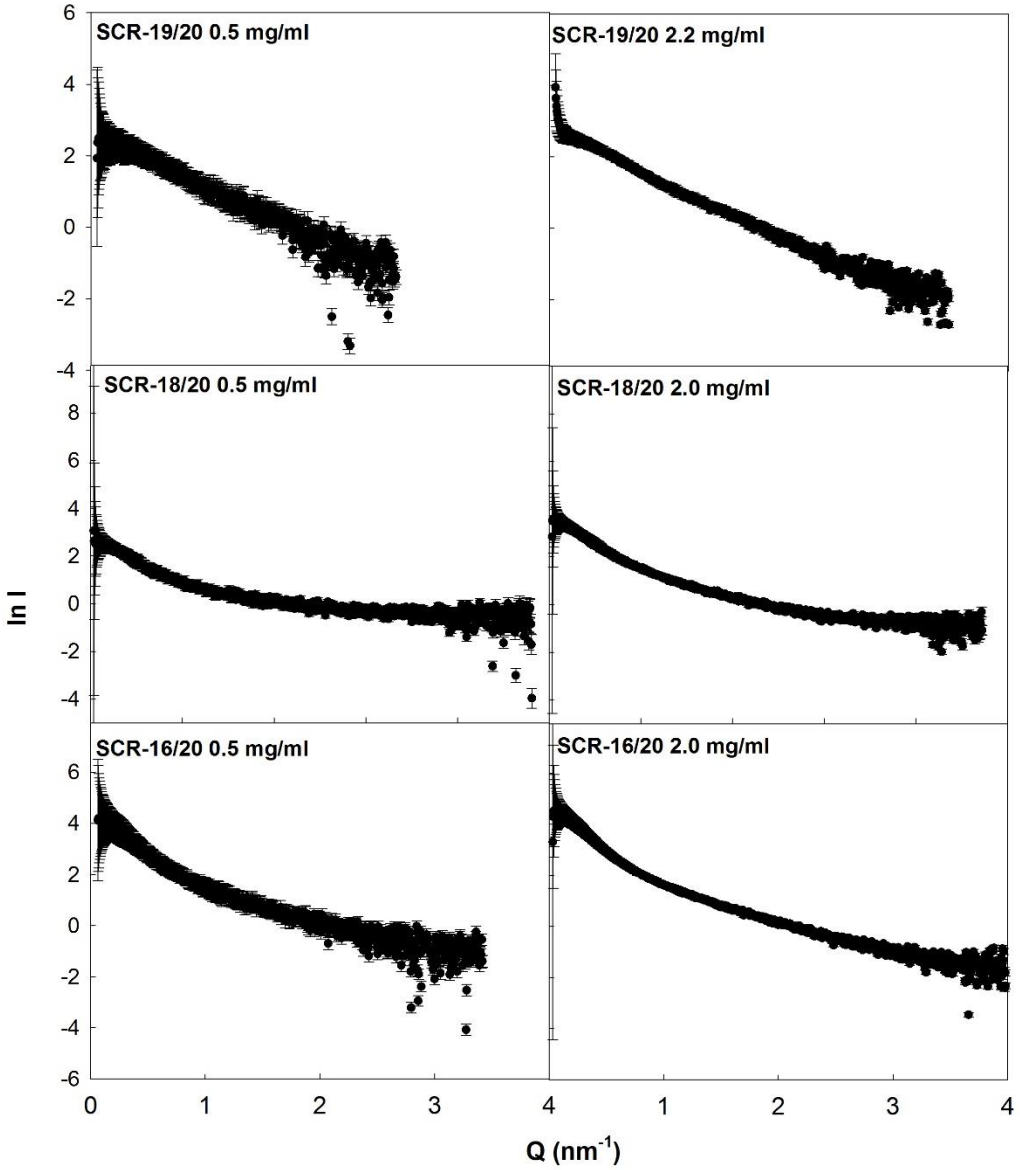
Dunne, O., Moulin, M., Haertlein, M., Forsyth,V. T., & Perkins, S. J. Small angle scattering studies of the complement regulator Factor H with its ligand C3d. European Conference on Neutron Scattering. 29 September - 4th August, 2015. Zaragoza, Spain (Poster and Abstract).

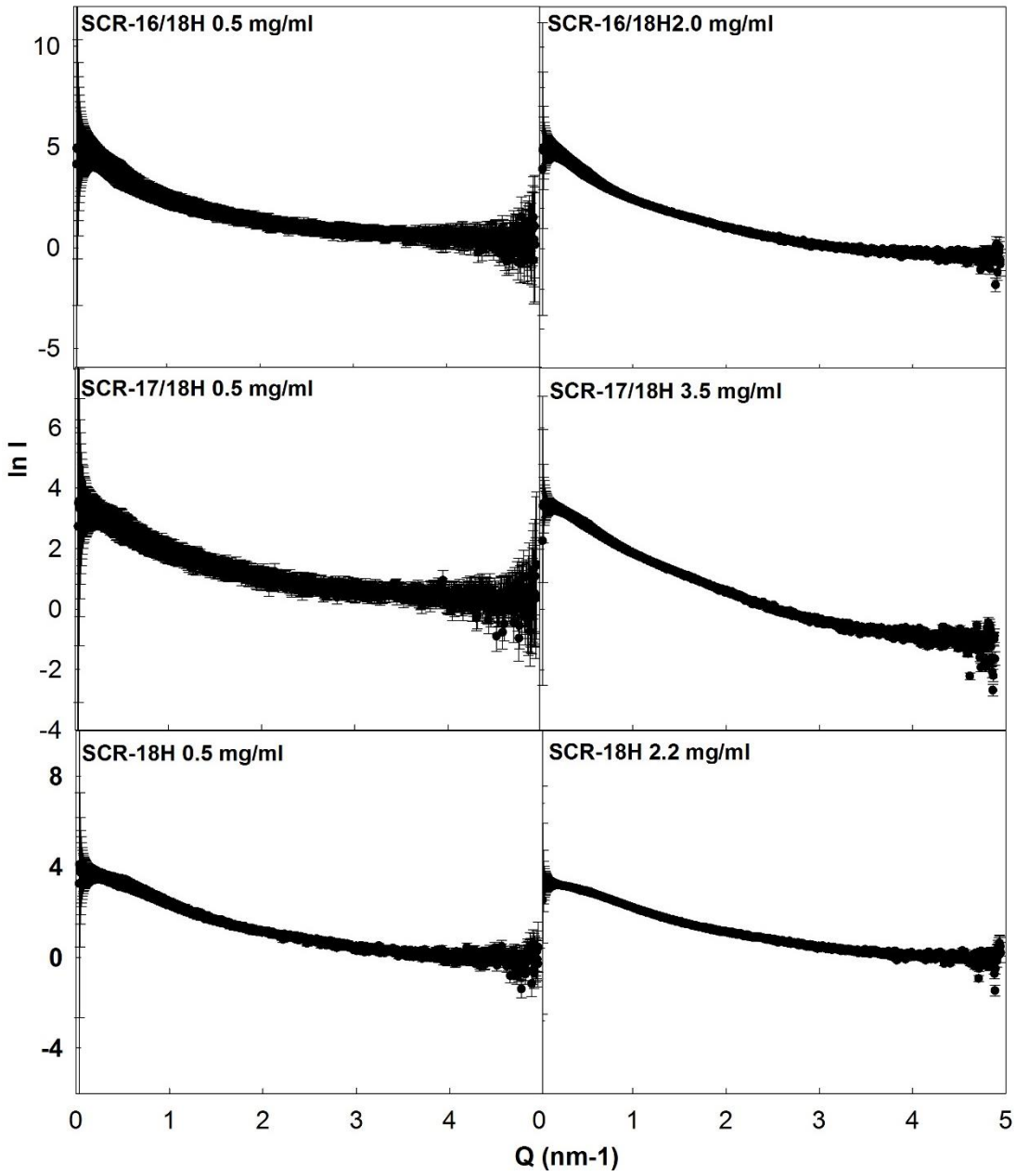
Dunne, O., Nan, R., Gor., J., Adamson, P., Gordon, D. L., Moulin, M., Forsyth, V. T., & Perkins, S. J. A dimerisation site at SCR-17/18 clarifies a new Factor H may explain its disease causing mutations and a new mechanism for regulatory control and its disease causing mutations. 15th European meeting on Complement in Human Disease. 27-30 June 2015. Uppsala, Sweden (Poster and Abstract).

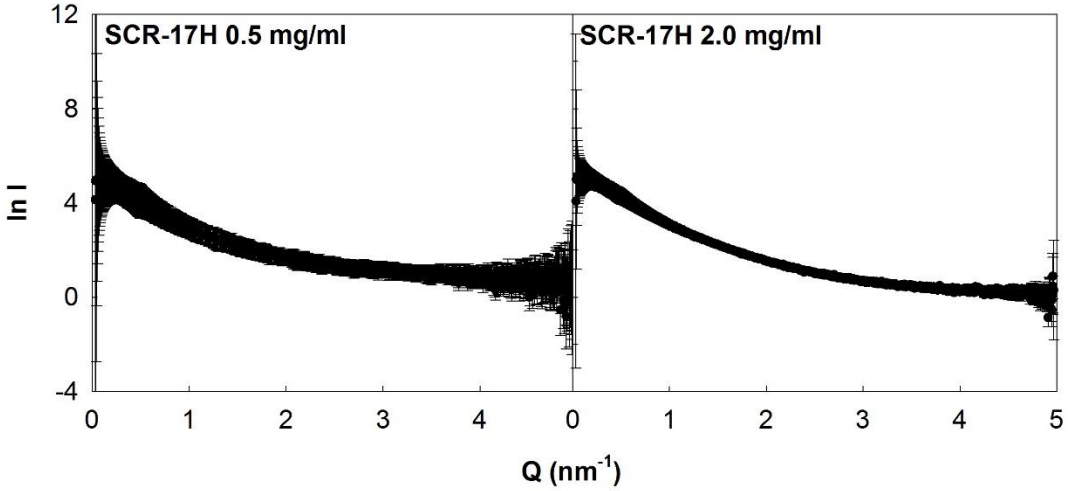
Dunne, O., Moulin, M., Adamson, P., Haertlein, M., Martel, A., Round, A., Forsyth, V. T., & Perkins, S. J. Unravelling the structural interactions of complement Factor H with its C-reactive protein and C3d ligands. Neutrons in Biology and Biotechnology. 19-21 February, 2014. Grenoble, France (Poster and Abstract).

Appendix I

The scattering curves I(Q) with experimental error corresponding to Figure 4.13, Chapter Four, page 103.

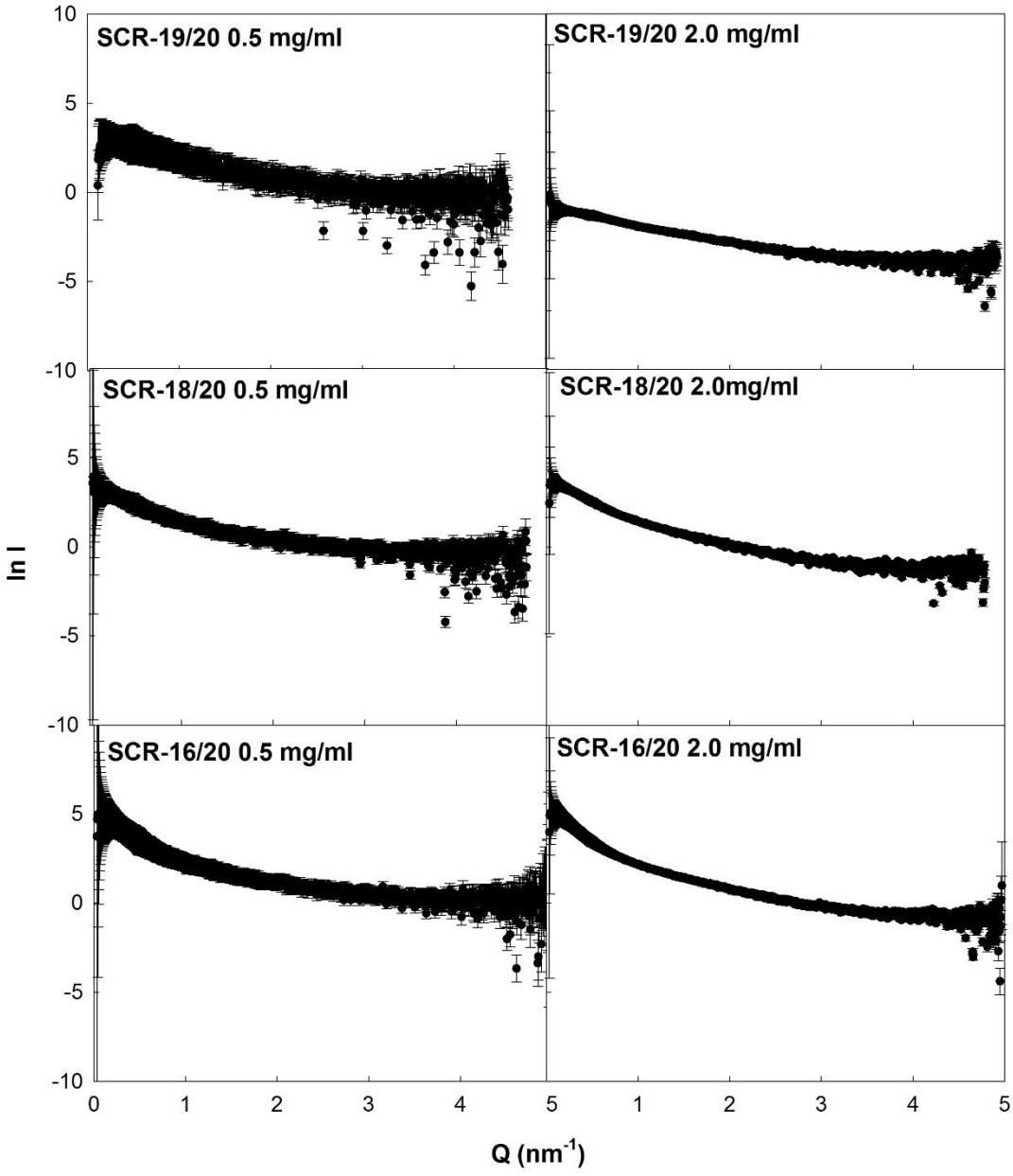


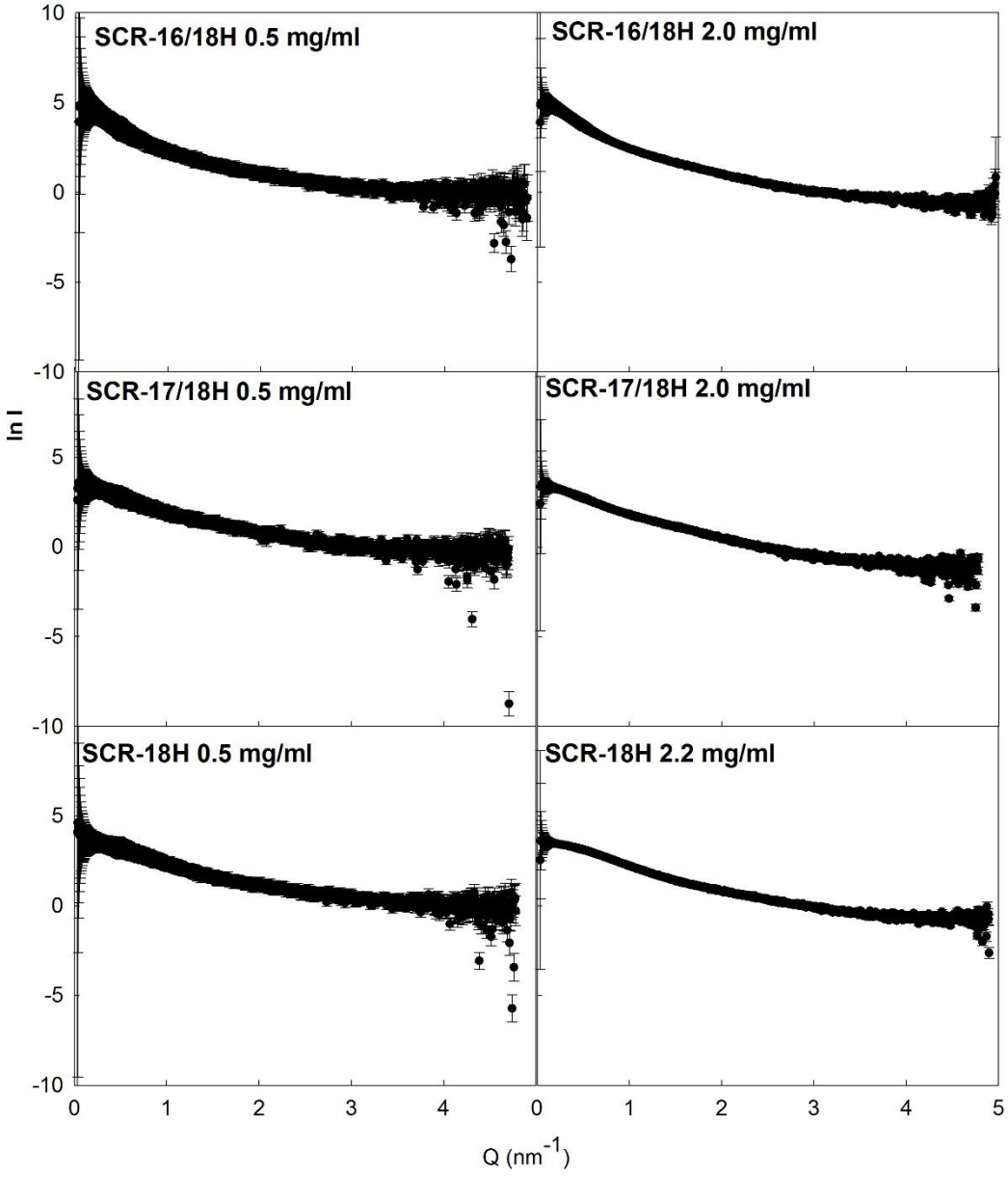

SCR-16/18H 0.5 mg/ml
SCR-16/18H2.0 mg/ml
SCR-17/18H 0.5 mg/ml
SCR-17/18H 3.5 mg/ml
SCR-18H 0.5 mg/ml
SCR-18H 2.2 mg/ml
ln I
Q (nm-1)

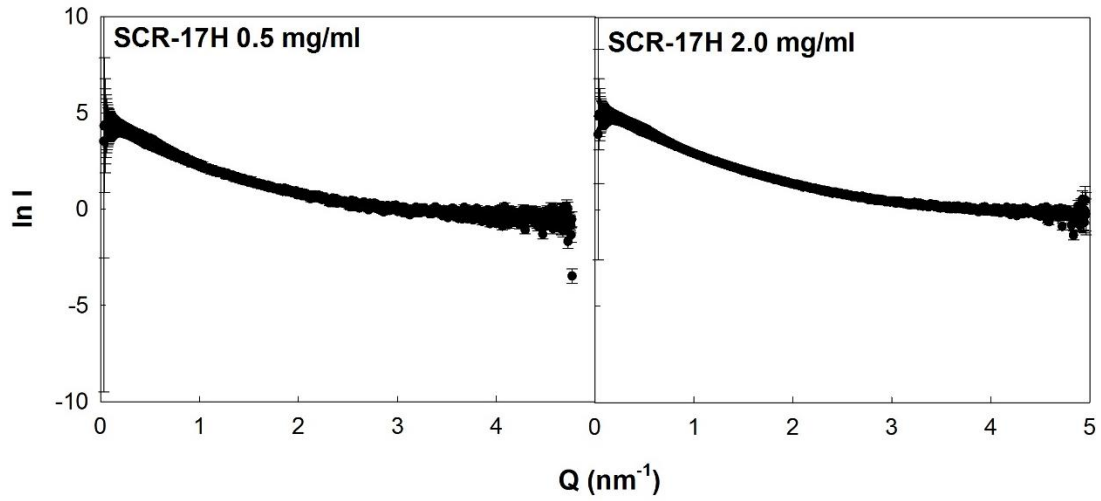

SCR-17H 0.5 mg/ml
SCR-17H 2.0 mg/ml
ln I
Q (nm-1)

Appendix II

The scattering curves I(Q) with experimental error corresponding Figure 4.17, Chapter Four, page 110.

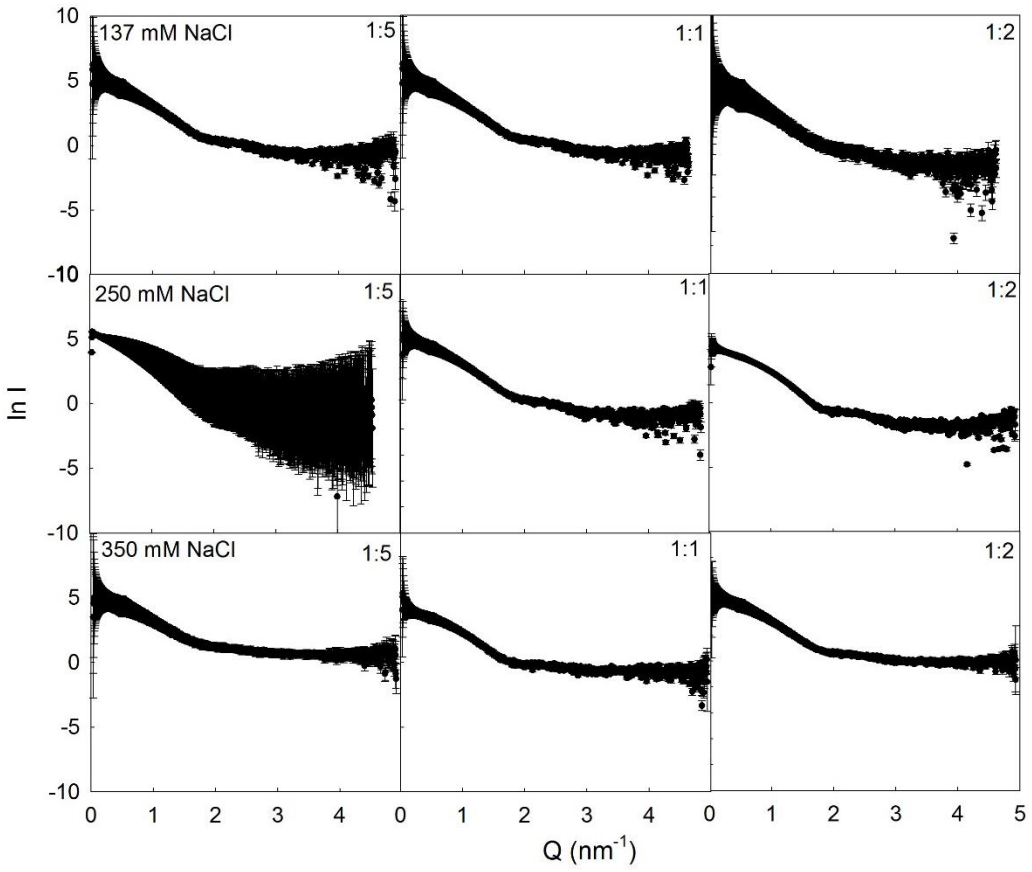



SCR-16/18H 0.5 mg/ml
SCR-16/18H 2.0 mg/ml
SCR-17/18H 0.5 mg/ml
SCR-17/18H 2.0 mg/ml
SCR-18H 0.5 mg/ml
SCR-18H 2.2 mg/ml
ln I
Q (nm^{-1})



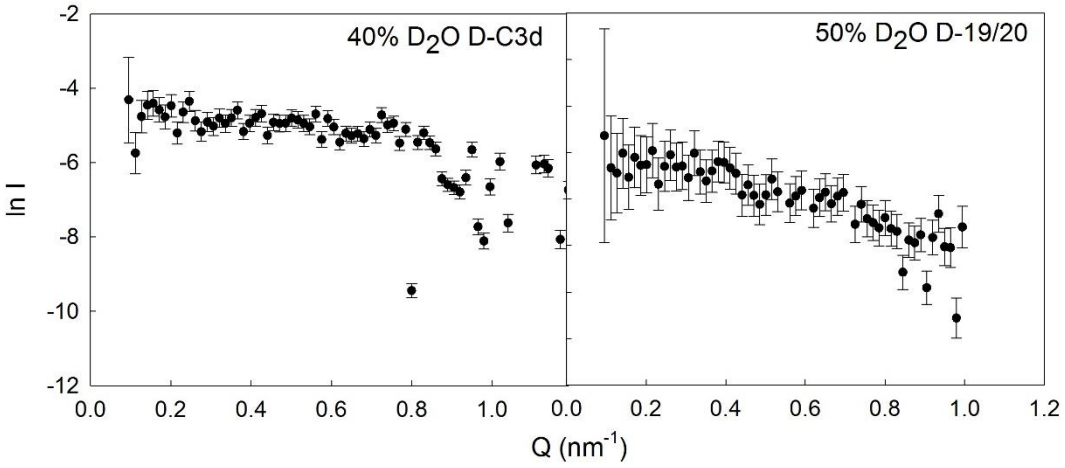
Appendix III

The scattering curves I(Q) with experimental error corresponding Figure 7.2, Chapter Seven, page 193.



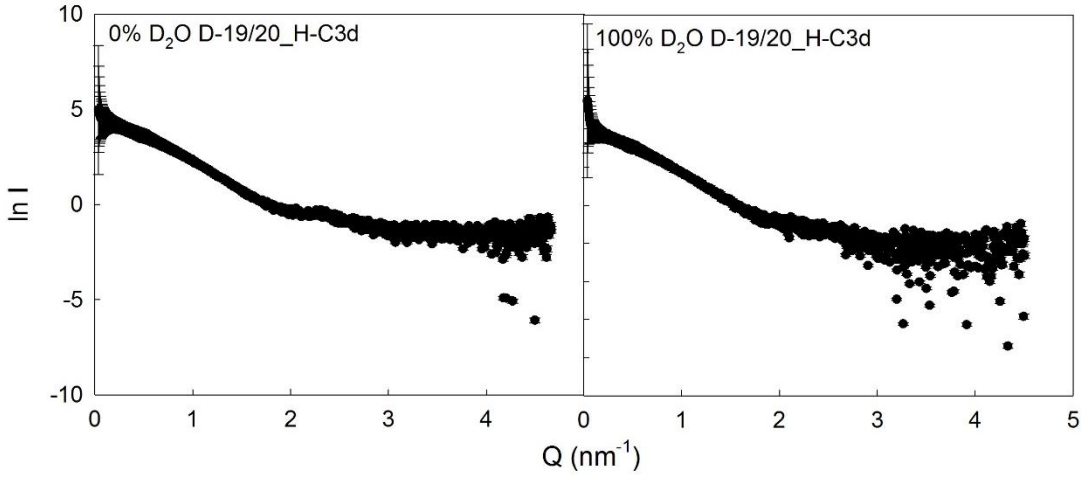
Appendix IV

The scattering curves I(Q) with experimental error corresponding to Figure 6.5 Chapter Six, page 176.



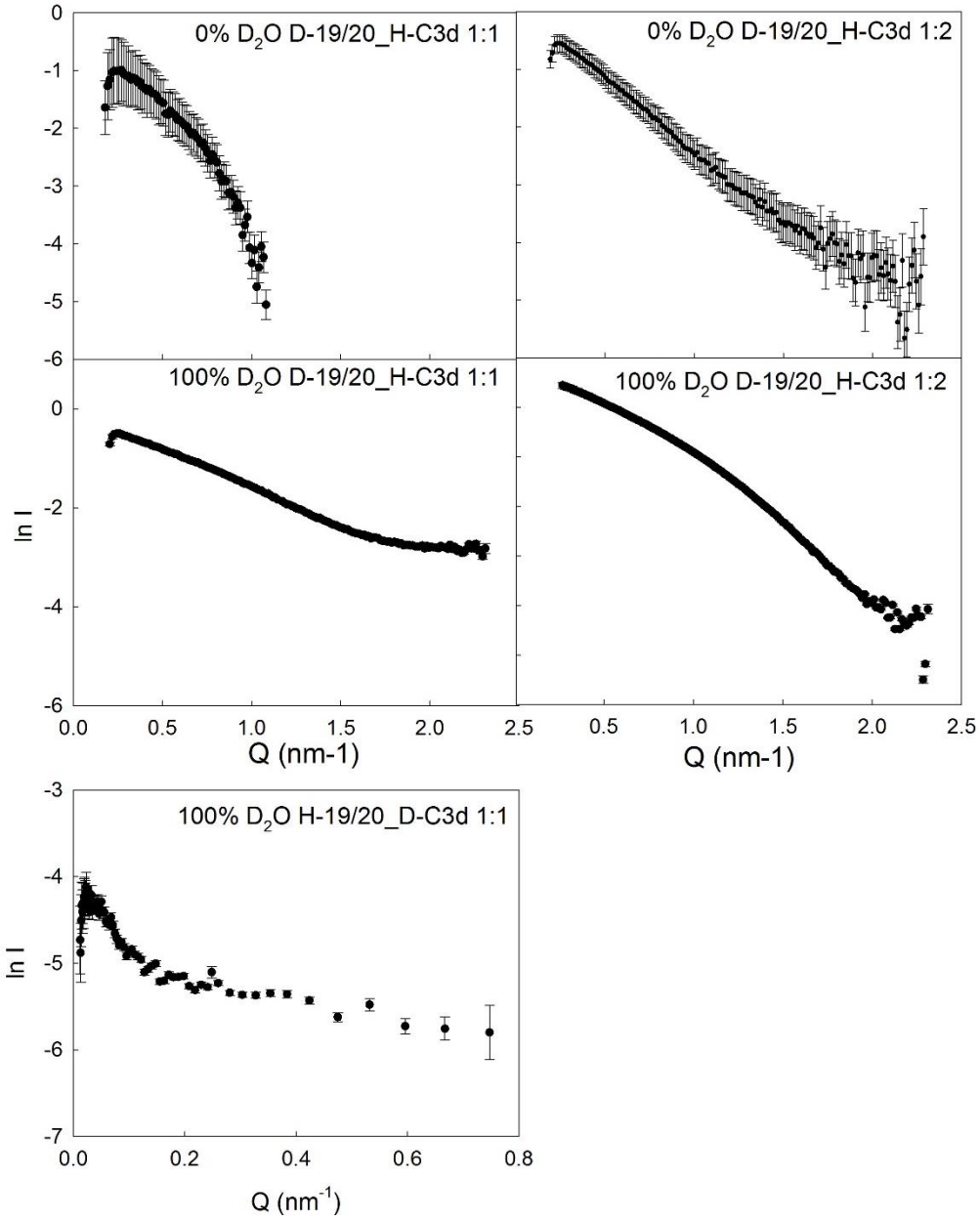
Appendix V

The scattering curves I(Q) with experimental error corresponding to Figure 7.4 Chapter Seven, page 202.



Appendix VI

The scattering curves I(Q) with experimental error corresponding Figure 7.5, Chapter Seven, page 203.



References

Agrawal, A., Shrive, A. K., Greenhough, T. J., & Volanakis, J. E. (2001). Topology and structure of the C1q-binding site on C-reactive protein. *J. Immunol.* **166,** 3998–4004.

Alexander, C. G., Jürgens, M. C., Shepherd, D. A., Freund, S. M. V., Ashcroft, A. E., & Ferguson, N. (2013). Thermodynamic origins of protein folding, allostery, and capsid formation in the human hepatitis B virus core protein. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 2782-2791.

Anderson, D.H., Ozaki, S., Nealon, M., Neitz, J., Mullins, R.F., Hageman, G.S., Johnson, L.V. (2001). Local cellular sources of apolipoprotein E in the human retina and retinal pigmented epithelium: implications for the process of drusen formation. *Am. J. Opthalmol*. **6**, 767-781.

Appolaire, A., Girard, E., Colombo, M., Asunció N Durá, M., Moulin, M., Härtlein, M., Franzetti, B. & Gabel, F. (2014). Small-angle neutron scattering reveals the assembly mode and oligomeric architecture of TET, a large, dodecameric aminopeptidase. *Acta. Crystallogr. D. Biol. Crystallogr.* **70,** 2983–2993.

Arlaud, G.J., Thielens, N.M., Aude, C.A. (1989). Structure and function of C1r and C1s: current concepts. *Behring. Inst. Mitt*. **84**, 56-64.

Artero, J. B., Härtlein, M., McSweeney, S., & Timmins, P. (2005). A comparison of refined X-ray structures of hydrogenated and perdeuterated rat γE-crystallin in H2O and D2O. *Acta. Crystallogr. D. Biol. Crystallogr.* **61,** 1541-1549.

Ascher, D. B., Wielens, J., Nero, T. L., Doughty, L., Morton, C. J., & Parker, M. W. (2014). Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. *Sci. Rep*. **23,** 1-12

Aslam, M., & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **309,** 1117-1138

Balbo, A., & Schuck, P. (2005). Analytical Ultracentrifugation in the Study of Protein Self-association and Heterogeneous Protein-Protein Interactions. In *Protein-Protein Interactions : A Molecular Cloning Manual*.

Ballermann, B. (1998). Endothelial cell activation. *Kidney International*, **53,** 1810–1826.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, **10,** 411–421.

Barlow, P. N., Norman, D. G., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, P. C., Sim, R., & Campbell, I. D. (1992). Solution structure of the fifth repeat of

factor H: a second example of the complement control protein module. *Biochemistry*, **31,** 3626–3634.

Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B., & Campbell, I. D. (1993). Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol*. **232,** 268-284.

Bird, A. C. (1992). Bruch's membrane change with age. *Brit. J. Ophthalmol*. **76**, 166-168.

Bird, A. C., Bressler, N. M., Bressler, S. B., Chisholm, I. H., Coscas, G., Davis, M. D., de Jong, P. T., Klaver, C. C., Klein, B. E., Klein, R., Mitchell, P., Sarks, J. P., Sarks, S. H., Soubrane, G., Taylor, H. R., Vingerling, J. R. & The International ARM Epidemiological study group. (1995). An international classification and grading system for age-related maculopathy and age-related macular degeneration. *Surv. Ophthalmol*. **39**, 367-374.

Bíró, A., Rovó, Z., Papp, D., Cervenak, L., Varga, L., Füst, G., & Prohászka, Z. (2007). Studies on the interactions between C-reactive protein and complement proteins. *Immunology*. **121,** 40-50.

Black, S., Kushner, I., & Samols, D. (2004). C-reactive protein. *J. Biol Chem.* **279,** 48487–48490.

Blackmore, T. K., Hellwage, J., Sadlon, T. A., Higgs, N., Zipfel, P. F., Ward, H. M., & Gordon, D. L. (1998). Identification of the second heparin-binding domain in human complement factor H. *J. Immunol.* **160,** 3342-3348.

Blackmore, T. K., Sadlon, T. A., Ward, H. M., Lublin, D. M., & Gordon, D. L. (1996). Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J. Immunol.* **157,** 5422-5427.

Blanchet, C. E., & Svergun, D. I. (2013). Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution SAXS: small-angle X-ray scattering. *Annu. Rev. Phys. Chem*. **64,** 37–54.

Blom, A. M., Kask, L., & Dahlbäck, B. (2003). CCP1-4 of the C4b-binding protein α-chain are required for factor I mediated cleavage of complement factor C3b. *Mol. Immunol*. **39,** 547-556.

Bordet, J., & Gengou, O. (1901). Sur l'existence de substances sensibilisatrices dans la plupart des serum antimicrobiens. *Ann. Inst. Pasteur*. **15**, 289-302.

Botto, M., Fong, K. Y., So, A. K., Barlow, R., Routier, R., Morley, B. J., & Walport, M. J. (1992). Homozygous hereditary C3 deficiency due to a partial gene deletion. *Proc. Nat. Acad. Sci. U. S. A.* **89,** 4957–4961.

Boussaad, S., Pean, J., & Tao, N. J. (2000). High-resolution multiwavelength surface plasmon resonance spectroscopy for probing conformational and electronic changes in redox proteins. *Anal. Chem.* **72,** 222–226.

Bretthauer, R. K., & Castellino F.J. (1999). Glycosylation of *Pichia pastoris* derived proteins. *Biotechnol. Appl. Biochem.* **30,** 193-200

Breyton, C., Gabel, F., Lethier, M., Flayhan, A., Durand, G., Jault, J.M., Juillan_Binard, C., Imbert, L., Moulin, M., Ravaud, S., Haertlein, M. & Ebel, C. (2013). Small angle neutron scattering for the study of solubilised membrane proteins. *Europ. Phys. J. E. Soft Matter Biol. Phys.* **36,** 71-87.

Brooimans, R. A., Hiemstra, P. S., van der Ark, A. A., Sim, R. B., van Es, L. A., & Daha, M. R. (1989). Biosynthesis of complement factor H by human umbilical vein endothelial cells. Regulation by T cell growth factor and IFN-gamma. *J. Immunol.* **142**, 2024-2030.

Carter, R. H., Tuveson, D. A., Park, D. J., Rhee, S. G., & Fearon, D. T. (1991). The CD19 complex of B lymphocytes. Activation of phospholipase C by a protein tyrosine kinase-dependent pathway that can be enhanced by the membrane IgM complex. *J. Immunol.* **147,** 3663–3671.

Cheng, Z., Jokiranta, T. S., Seeberger, H., Jòzsi, M., Heinen, S., Noris, M., Remuzzi, G., Ormsby, R., Gordon, D., Meri, S., Hellwage, J., & Zipfel, P. F. (2005). Binding of complement factor H to endothelial cells is mediated by the carboxy-terminal glycosaminoglycan binding site. *Am. J. Pathol.* **167,** 1173–1181.

Clark, S., Ridge, LA, Herbert, AP, Hakobyan, S., Mulloy, B., Lennon, R., Wurrzner, B., Morgan, P., Uhrin, D., Bishop, P. & Day, A. (2013). Tissue-Specific Host Recognition by Complement Factor H is mediated by differential activities of its glycosaminoglycans-Binding regions. *J. Immunol.* **190,** 2049–2057.

Cole, J. L., & Hansen, J. C. (1999). Analytical ultracentrifugation as a contemporary biomolecular research tool. *J. Biomol. Tech.* **10,** 163–176.

Cole, J. L., Lary, J. W., P. Moody, T., & Laue, T. M. (2008). Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods. Cell Biol.* **84,** 143-179.

Crowell, R. E., Du Clos, T. W., Montoya, G., Heaphy, E., & Mold, C. (1991). C-reactive Protein Receptors on the Human Monocytic cell line U-937. *J Immunol.*, **147,** 3445–3451.

Curtis, J. E., Raghunandan, S., Nanda, H., & Krueger, S. (2012). SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comput. Phys. Commun*. **183,** 382-389.

Cuypers, M. G., Trubitsyna, M., Callow, P., Forsyth, V. T., & Richardson, J. M. (2013). Solution conformations of early intermediates in Mos1 transposition. *Nucleic Acids Res*. **41,** 2020-2033.

Dam, J., & Schuck, P. (2005). Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions c(s) and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophys. J*. **89,** 651-666.

Davis, A E., & Harrison, R. A. (1982). Structural characterization of factor I mediated cleavage of the third component of complement. *Biochemistry*, **21,** 5745–5749.

Dewhurst, C. D. (2008). D33—a third small-angle neutron scattering instrument at the Institut Laue Langevin. *Meas. Sci. Technol*. **19,** 1-8.

DiScipio, R. G. (1992). Ultrastructures and interactions of complement factors H and I. *J. Immunol.* **149,** 2592-2599.

DiScipio, R. G., Unit, M. R. C. I., Road, S. P., & Ox, O. (1981). The conversion of human complement component C5 into fragment C5b by the alternative-pathway C5 convertase, Biochem. J. **199,** 497–504.

Diaz-Guillen, M.A., Rodriguez de Cordoba, S., Heine-Suner, D. (1999). A radiation hybrid map of complement factor H and factor H-related genes *Immunogenetics*, **49**, 549–552

Dragon-Durey, M. A. & Fremeaux-Bacchi, V. (2005). Atypical haemolytic uraemic syndrome and mutations in complement regulator genes. *Springer Semin. Immunopathol*. **27**, 359-374.

Edwards, A. O., Ritter, R., Abel, K. J., Manning, A., Panhuysen, C., & Farrer, L. A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science* **308,** 421–424.

Efimova, Y. M., Haemers, S., Wierczinski, B., Norde, W., & Van Well, A. (2007). Stability of globular proteins in H_2O and D_2O. *Biopolymers*. **85,** 264-273.

Fearon, D. T. (1978). Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proc. Nat. Acad. Sci. U.S.A.* **75,** 1971-1975.

Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riand, S., Siret, L., & Bihoreau, N. (2007). Site-specific N-glycan characterization of human complement factor H. *Glycobiology*, **17,** 932–944.

Fernando, A. N., Furtado, P. B., Clark, S. J., Gilbert, H. E., Day, A. J., Sim, R. B., & Perkins, S. J. (2007). Associative and Structural Properties of the Region of Complement Factor H Encompassing the Tyr402His Disease-related Polymorphism and its Interactions with Heparin. *J. Mol. Biol.*, **368,** 564–581.

Fersht, A. (1999). Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding 1st Edition. W.H. Freeman, New York.

Forneris, F., Ricklin, D., Wu, J., Tzekou, A., Wallace, R. S., Lambris, J. D., & Gros, P. (2010). Structures of C3b in complex with factors B and D give insight into complement convertase formation. *Science* **330,** 1816-1820.

Fredslund, F., Laursen, N. S., Roversi, P., Jenner, L., Oliveira, C. L. P., Pedersen, J. S., Nunn, M.A., Lea, S.M., Discipio, R., Sottrup-Jensen, L., & Andersen, G. R. (2008). Structure of and influence of a tick complement inhibitor on human complement component 5. *Nat. Immunol.* **9,** 753–60.

Fremeaux-Bacchi, V., Fakhouri, F., Garnier, A., Bienaimé, F., Dragon-Durey, M. A., Ngo, S., Moulin, B., Servais, A., Provot, F., Rostaing, L., Burtey, S., Niaudet, P., Deschenes, G., Lebranchu, Y., Zuber, J. & Loirat, C. (2013). Genetics and outcome of atypical hemolytic uremic syndrome: A nationwide french series comparing children and adults. *Clin. J. Am. Soc. Nephrol*. **8,** 554-562.

Friese, M. a., Hellwage, J., Jokiranta, T. S., Meri, S., Peter, H. H., Eibel, H., & Zipfel, P. F. (1999). FHL-1/reconectin and factor H: Two human complement regulators which are encoded by the same gene are differently expressed and regulated. *Mol. Immunol.* **36,** 809–818.

Ganapathi, M. K., Rzewnicki, D., Samols, D., Jiang, S. L., & Kushner, I. (1991). Effect of combinations of cytokines and hormones on synthesis of serum amyloid A and C-reactive protein in Hep 3B cells. *J.Immunol.* **147,** 1261–1265.

García De La Torre, J., Huertas, M. L., & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J*. **78,** 719-730.

Gershov, D., Kim, S., Brot, N., & Elkon, K. B. (2000). C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity. *J. Exp. Med.* ***192,*** 1353–1364.

Gewurz, H., Ying, S.C., Jiang, H., Lint, T.F. (1993). Nonimmune activation of the classical complement pathway. *Behring. Inst. Mitt*. **93**, 138-147.

Gigli, I., Fujita, T., & Nussenzweig, V. (1979). Modulation of the classical pathway C3 convertase by plasma proteins C4 binding protein and C3b inactivator. *Proc. Natl. Acad. Sci. U.S.A.* **76,** 6596–6600.

Glatter, O. & Kratky, O. (1982). Small-angle X-ray scattering. Academic Press, New York.

Goicoechea De Jorge, E., Caesar, J. J. E., Malik, T. H., Patel, M., Colledge, M., Johnson, S., Hakobyan, S., Morgan, P.B., Harris, C.L., Pickerin, M.C., & Lea, S. M. (2013). Dimerization of complement factor H-related proteins modulates complement activation in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 4685-4690.

Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J., & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *J. Immunol.* **155,** 348-356.

Guillon, C., Bigouagou, U. M., Folio, C., Jeannin, P., Delneste, Y., & Gouet, P. (2015). A staggered decameric assembly of C-reactive protein stabilised by zinc ions revealed by X-ray crystallography. *Protein Pept. Lett.* **22,** 248-255

Guthridge, J. M., Rakstang, J. K., Young, K. A., Hinshelwood, J., Aslam, M., Robertson, A., Gipson, M.G., Sarrias, M-R., Moore, W. T., Meagher, M., Karp, D., Lambris, J.D., Perkins, S.J., & Holers, V. M. (2001). Structural studies in solution of the recombinant N-terminal pair of short consensus/complement repeat domains of complement receptor type 2 (CR2/CD21) and interactions with its ligand C3dg. *Biochemistry*. **40,** 5931-5941.

Hageman, G. S., Anderson, D. H., Johnson, L. V, Hancox, L. S., Taiber, A. J., Hardisty, L. I., Hageman, J. L., Stockman, H. A., Borchardt, J.D., Gehrs, K. M., Smith, R. J. H., Silvestri, G., Russell, S. R., Klaver, C. C., Barbazetto, I., Chang, S., Yannuzzi, L. A., Barile, G. R., Merriam, J. C., Theodore Smith, R., Olsh, A. K., Bergeroni, J., Zernant, J., Merriam, J. E., Gold, B., Dean, M., & Allikmets, R. (2005). A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 7227-7232.

Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., Spencer, K., Kwan, S., Noureddine, M., Gilbert, J., Schnetz-Boutard, N., Agarwal, A., Postel, E., & Pericak-Vance, M. A. (2005). Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308,** 419-421.

Hakobyan, S., Harris, C. L., Van Den Berg, C. W., Fernandez-Alonso, M. C., De Jorge, E. G., De Cordoba, S. R., Rivas, G., Mangione, P., Pepys, M. B., & Morgan, B. P. (2008). Complement factor H binds to denatured rather than to native pentameric C-reactive protein. *J. Biol. Chem*. **283,** 30451-30460.

Haon, S., Auge, S., Tropis, M., Milon, A., & Lindley, N. C. (1993). Low cost production of perdeuterated biomass using methylotrophic yeasts. *J. Labelled Compd. Radiopharm*. **33,** 1053-1063.

Hardy, R. C. and Cottington, R. L. (1949) Viscosity of deuterium oxide and water in the range 5° to 125° C. *Journal of Research of the National Bureau of Standards*, **42**, 573-578.

Heinen, S., Hartmann, A., Lauer, N., Wiehl, U., Dahse, H. M., Schirmer, S., Gropp, K., Enghardt, T., Wallich, R., Halbich, S., Mihlan, M., Schlotzer-Schrehardt, U., Zipfel, P.F., & Skerka, C. (2009). Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation. *Blood*. **114,** 2439-2447.

Heinrich, F. (2015). Deuteration in Biological Neutron Reflectometry. *Methods Enzymol. I*n press.

Herbert, A. P., Deakin, J. a., Schmidt, C. Q., Blaum, B. S., Egan, C., Ferreira, V. P., Pangburn, M., Lyon, M., Uhrin, D., & Barlow, P. N. (2007). Structure shows that a glycosaminoglycan and protein recognition site in factor H is perturbed by age-related macular degeneration-linked single nucleotide polymorphism. *J. Biol. Chem.* **282,** 18960–18968.

Herbert, A. P., Uhrín, D., Lyon, M., Pangburn, M. K., & Barlow, P. N. (2006). Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. *J. Biol. Chem.* **281,** 16512-16520.

Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrin, D., & Barlow, P. N. (2008). Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *J. Biol. Chem.* **283,** 9475–9487.

Hourcade, D. E. (2006). The role of properdin in the assembly of the alternative pathway C3 convertases of complement. *J. Biol. Chem*. **281,** 2128-2132.

Jacrot, B. (1976). The study of biological solutions by neutron scattering from solution. *Rep. Prog. Phys*. **39,** 911-953

Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D., & Gros, P. (2006). Structure of C3b reveals conformational changes that underlie complement activity. *Nature,* **444,** 213–216.

Janssen, B. J. C., Huizinga, E. G., Raaijmakers, H. C. a., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B., & Gros, P. (2005). Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature,* **437,** 505–511.

Jarva, H., Jokiranta, T. S., Hellwage, J., Zipfel, P. F., & Meri, S. (1999). Regulation of complement activation by C-reactive protein: targeting the complement inhibitory activity of factor H by an interaction with short consensus repeat domains 7 and 8-11. *J. Immunol.* **163,** 3957-3962.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., & Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077,** 101–113.

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., & Duhr, S. (2011). Molecular interaction studies using microscale thermophoresis. *Assay and Drug Dev. Technol.* **9,** 342–353.

Jiang, B. H., Robey, F. A., & Gewurz, H. (1992). Localization of sites through which C-reactive protein binds and activates complement to residues 14-26 and 76-92 of the human C1q A chain. *J. Exp. Med.* **175,** 1373–1379.

Johnson, P. T., Betts, K. E., Radeke, M. J., Hageman, G. S., Anderson, D. H., & Johnson, L. V. (2006). Individuals homozygous for the age-related macular degeneration risk-conferring variant of complement factor H have elevated levels of CRP in the choroid. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 17456-17461.

Jokiranta, T. S., Hellwage, J., Koistinen, V., Zipfel, P. F., & Meri, S. (2000). Each of the three binding sites on complement factor H interacts with a distinct site on C3b. *J. Biol. Chem.* **275,** 27657-27662.

Jokiranta, T. S., Westin, J., Nilsson, U. R., Nilsson, B., Hellwage, J., Löfås, S., Gordon, D., Ekdahl, K., & Meri, S. (2001). Complement C3b interactions studied with surface plasmon resonance technique. *Int. Immunopharmacol.* **1,** 495–506.

Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Pä Repalo, M., Meri, S., & Goldman, A. (2006). Structure of complement factor H carboxyl- terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J.* ***25,*** 1784–1794.

Jonsson, U., Fagerstam, L., Lofas, S., Stenberg, E., Karlsson, R., Frostell, A., Markey, F., & Schindler, F. (1993). Introducing a biosensor based technology for real-time biospecific interaction analysis. *Ann. Biol. Clin. (Paris)* **51,** 19-26

Józsi, M., & Zipfel, P. F. (2008). Factor H family proteins and human diseases. *Trends Immunol*. **29,** 380-387.

Kajander, T., Lehtinen, M. J., Hyvärinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E.,Meri, S., Goldman, A., & Jokiranta, T. S. (2011). Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2897–2902.

Kavanagh, D., Goodship, T. H. J., & Richards, A. (2006). Atypical haemolytic uraemic syndrome. *Br. Med. Bull*. **77/78,** 5-22.

Reid, K. B. M. (1977). Amino acid sequence of the N-terminal forty-two amino acid residues of the C chain of subcomponent Clq of the first component of human complement. *Biochem. J.* **161,** 247–251.

Kerr, F. K., Thomas, A. R., Wijeyewickrema, L. C., Whisstock, J. C., Boyd, S. E., Kaiserman, D., Matthews, A.Y., Bird, P. I. Thielens, N. M. Rossi, V., & Pike, R. N. (2008). Elucidation of the substrate specificity of the MASP-2 protease of the lectin complement pathway and identification of the enzyme as a major physiological target of the serpin, C1-inhibitor. *Mol. Immunol.* **45,** 670–677.

Kerr, M. A. (1980). The human complement system: assembly of the classical pathway C3 convertase. *Biochem. J.* **189,** 173–181.

Kinoshita, T., Takata, Y., Kozono, H., Takeda, J., Hong, K., & Inoue, K. O. (1988). C5 convertaste of the alternative complement pathway: covalent linkage between two C3b molecules within the trimolecular complex enzyme. *J. Immunol.* **141,** 3895–3901.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., Hennin, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., & Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* **308,** 385–389.

Knobel, H. R., Villiger, W., & Isliker, H. (1975). Chemical analysis and electron microscopy studies of human C1q prepared by different methods. *Eur. J. Immunol.* **5,** 78–82.

Korkhov, V. M., Mireku, S. A., Veprintsev, D. B., & Locher, K. P. (2014). Structure of AMP-PNP–bound BtuCD and mechanism of ATP-powered vitamin B12 transport by BtuCD–F. *Nat. Struct. Mol. Biol.*, **21.** 1097-1100.

Kristensen, T. (1986). Murine protein H is comprised of 20 repeating units, 61 amino acids in length. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3963–3967.

Krych-Goldberg, M., & Atkinson, J. P. (2001). Structure-function relationships of complement receptor type 1. *Immunol. Rev.* **180,** 112–122.

Kühn, S., & Zipfel, P. F. (1996). Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur. J. Immunol.* **26,** 2383-2387.

Lachmann, P. J. (2002). Microbial subversion of the immune response. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8461-8462.

Laine, M., Jarva, H., Seitsonen, S., Haapasalo, K., Lehtinen, M. J., Lindeman, N., Anderson, D. H., Johnson, P. T., Jarvela, I., Jokiranta, T. S., Hageman, G. S. Immonen, I., & Meri, S. (2007). Y402H polymorphism of complement factor H affects binding affinity to C-reactive protein. *J. Immunol.* **178,** 3831-3836.

Lambris, J. D., Avila, D., Becherer, J. D., & Muller-Eberhard, H. J. (1988). A Discontinuous Factor H Binding Site in the Third Component of Complement as Dealinated by Synthetic Peptides. *J. Biol. Chem.* **263,** 12147–12150.

Lambriss, J. D., Avilas, D., Becherers, J. D., & Muller-Eberhardq, H. J. (1988). A Discontinuous Factor H Binding Site in the Third Component of Complement as Delineated by Synthetic Peptides. J. Biol. Chem. **263,** 12147–12150.

Lamm, O. (1929). Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat. Astr. Fys.* **21B**, 1-4.

Law, S. K., & Levine, R. P. (1977). Interaction between the third complement protein and cell surface macromolecules. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 2701–2705.

Law, S.K. & Reid, K.B.M. (1995). Complement 2nd Edition. IRL press, Oxford.

Lebowitz, J., Lewis, M. S., & Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* **11,** 2067–2079.

Lederer, H., May, R. P., Kjems, J. K., Schaefer, W., Crespi, H. L., & Heumann, H. (1986). Deuterium incorporation into Escherichia coli proteins. A neutron-scattering study of DNA-dependent RNA polymerase. *Eur. J. Bioch.* **156,** 655-659.

Leiting, B., Marsilio, F., & O'Connell, J. F. (1998). Predictable deuteration of recombinant proteins expressed in Escherichia coli. *Anal. Bioch.* **265,** 351-355.

Li, K., Okemefuna, A. I., Gor, J., Hannan, J. P., Asokan, R., Holers, V. M., & Perkins, S. J. (2008). Solution Structure of the Complex Formed between Human Complement C3d and Full-length Complement Receptor Type 2. *J. Mol. Biol.* **384,** 137-150.

Li, S. H., Szmitko, P. E., Weisel, R. D., Wang, C. H., Fedak, P. W. M., Li, R. K., Mickle, D. A. G. & Verma, S. (2004). C-Reactive Protein Upregulates Complement-Inhibitory Factors in Endothelial Cells. *Circulation* **109,** 833–836.

Lian, L.Y. & Roberts, G. (2011). Protein NMR spectroscopy: Practical techniques and applications 1st Edition. Wiley & Sons Ltd, United Kingdom.

Loos, M. (1982). Antibody-independent activation of C1 the first component of complement. *Ann. lmmunol.* **81,** 165–179.

Lu, J., Thiel, S., Wiedemann, H., Timpl, R., & Reid, K. B. M. (1990). Binding of the pentamer.hexamer forms of mannan-binding protein to zymosan activates the proenzyme C1r2C1s2 complex of the classical pathway of complement, without involvement of C1q. *J. Immunol.* **144,** 2287-2294.

Lublin, D. M., & Atkinson, J. P. (1989). Decay -accelerating factor: Biochemistry, molecular biology, and function. *Ann. Rev. Immunol*. **7,** 35-58.

Macauley-Patrick, S., Fazenda, M. L., McNeil, B., & Harvey, L. M. (2005). Heterologous protein production using the Pichia pastoris expression system. *Yeast* **22,** 249-270.

Maga, T. K., Nishimura, C. J., Weaver, A. E., Frees, K. L., & Smith, R. J. H. (2010). Mutations in alternative pathway complement proteins in American patients with atypical hemolytic uremic syndrome. *Hum. Mut.* **31,** 1445-1460.

Makou, E., Mertens, H. D. T., MacIejewski, M., Soares, D. C., Matis, I., Schmidt, C. Q., … Barlow, P. N. (2012). Solution structure of CCP modules 10-12 illuminates functional architecture of the complement regulator, factor H. *J. Mol. Biol.* **424,** 295–312.

Male, D., Ormsby, R., Ranganathan, S., Giannakis, E., & Gordon, D. (2000). Sequence analysis of 222 KB of human genomic DNA containing the entire FH, FHR-1 and FHR-3 genes. *Immunopharm*. **37,** 41-52.

Marnell, L. L., Mold, C., Voizer, M. A, Burlingame, R. W., & Du Clos, T. W. (1995). C-Reactive Protein Binds to FcγRI in transfected COS cells. *J. Immunol*. **155,** 2185-2193.

Massou, S., Puech, V., Talmont, F., Demange, P., Lindley, N. D., Tropis, M., & Milon, A. (1999). Heterologous expression of a deuterated membrane-integrated receptor and partial deuteration in methylotrophic yeasts. *J. Biomol. NMR*. **14,** 231-239.

Matsushita, M., Endo, Y., & Fujita, T. (1998). MASP1 (MBL-Associated Serine Protease 1). *Immunobiology*, **199,** 340–347.

Matsushita, M., Endo, Y., & Fujita, T. (2000). Cutting Edge: Complement-Activating Complex of Ficolin and Mannose- Binding Lectin-Associated Serine Protease 1. *J. Immunol.* **2,** 2–5.

Matsushita, M., & Fujita, T. (1992). Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease. *J. Exp. Med.* **176,** 1497–1502.

Mayer, M.M. (1958). Studies on the mechanism of hemolysis by antibody and complement. *Progr. Allergy*. **5**, 241-270.

Meyers, R.A. (2007). Immunology 1st Edition. Wiley-Blackwell, Weinheim.

Mihlan, M., Stippa, S., Jozsi,M., & Zipfel, P. F. (2009). Monomeric CRP contributes to complement control in fluid phase and on cellular surfaces and increases phagocytosis by recruiting Factor H. *Cell Death Differ.* **16,** 1630-1640

Mihlan, M., Blom, A.M., Kupreishvili, K., Lauer, N., Stelzner, K., Bergström, F., Niessen, H.W., & Zipfel, P.F. (2011). Monomeric C-reactive protein modulates classic complement activation on necrotic cells. *FASEB J.* **25,** 4198-4210

Mindall, J. A., & Grigorieff, N., (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334-347

Miwa, T., & Song, W. C. (2001). Membrane complement regulatory proteins: insight from animal studies and relevance to human diseases. *Int. Immunopharm.* **1,** 445–459.

Mold, C., Gewurz, H., & Du Clos, T. W. (1999). Regulation of complement activation by C-reactive protein. *Immunopharm.* **42,** 23-20.

Morgan, H. P., Mertens, H. D. T., Guariento, M., Schmidt, C. Q., Soares, D. C., Svergun, D. I., Herbert, A., Barlow, P., & Hannan, J. P. (2012). Structural analysis of the C-terminal region (modules 18-20) of complement regulator factor H (FH). *PloS One*, **7,** 32187-32197.

Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Kavanagh, D., Merterns, H. D. T., Svergun, D. I., Johansson, C. M., Uhrin, D., Barlow, P. N., & Hannan, J. P. (2011). Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol.* **18,** 463–470.

Mortensen, S., Kidmose, R. T., Petersen, S. V., Szilagyi, A., Prohaszka, Z., & Andersen, G. R. (2015). Structural Basis for the Function of Complement Component C4 within the Classical and Lectin Pathways of Complement. *J. Immunol*. **194,** 5488-5496.

Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E., & Rini, J. M. (1998). X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* **280,** 1277–1281.

Nan, R., Gor, J., & Perkins, S. J. (2008). Implications of the Progressive Self-association of Wild-type Human Factor H for Complement Regulation and Disease. *J. Mol. Biol*. **375,** 891-900.

Nilsson, U. R., & Muller-eberhard, H. J. (1965). Isolation of β1H Globulin from human serum and its characterisation as the fifth component of complement. *J. Exp. Med.* **122,** 277–298.

Noris, M., Caprioli, J., Bresin, E., Mossali, C., Pianetti, G., Gamba, S., Daina, E., Fenili, C., Castelletti, F., Sorosina, A., Piras, R., Donadelli, R., Maranta, R., van der Meer, I., Conway, E. M., Zipfel, P. F., Goodship, T. H. & Remuzzi, G. (2010). Relative role of genetic complement abnormalities in sporadic and familial aHUS and their impact on clinical phenotype. *Clin. J. Am. Soc. Nephrol*. **5,** 1844-1859.

Okemefuna, A. I., Gilbert, H. E., Griggs, K. M., Ormsby, R. J., Gordon, D. L., & Perkins, S. J. (2008). The Regulatory SCR-1/5 and Cell Surface-binding SCR-16/20 Fragments of Factor H Reveal Partially Folded-back Solution Structures and Different Self-associative Properties. *J. Mol. Biol.* **375,** 80-101.

Okemefuna, A. I., Li, K., Nan, R., Ormsby, R. J., Sadlon, T., Gordon, D. L., & Perkins, S. J. (2009a). Multimeric Interactions between Complement Factor H and Its C3d Ligand Provide New Insight on Complement Regulation. *J. Mol. Biol.* **391,** 119-135.

Okemefuna, A. I., Nan, R., Gor, J., & Perkins, S. J. (2009b). Electrostatic Interactions Contribute to the Folded-back Conformation of Wild Type Human Factor H. *J. Mol. Biol.* **391,** 98–118.

Okemefuna, A.I., Nan., R., Miller, A., Gor, J., & Perkins, S.J. (2010a). Complement factor H binds at two independent sites to C-reactive protein in acute phase concentrations. *J. Biol. Chem.* **285,** 1053-1065.

Okemefuna, A. I., Stach, L., Rana, S., Zlai Buetas, A. J., Gor, J., & Perkins, S. J. (2010b). C-reactive protein exists in an NaCl concentration-dependent pentamer-decamer equilibrium in physiological buffer. *J. Biol. Chem.* **285,** 1041–1052.

Olins, P. O., & Lee, S. C. (1993). Recent advances in heterologous gene expression in Escherichia coil. *Curr.Opin. Biotechnol.* **4,** 520–525.

Oppermann, M., Manuelian, T., Józsi, M., Brandt, E., Jokiranta, T. S., Heinen, S., Meri, S., Skerka, C., Gotze, O. & Zipfel, P. F. (2006). The C-terminus of complement regulator Factor H mediates target recognition: Evidence for a compact conformation of the native protein. *Clin. Exp. Immunol.* **144,** 342–352.

Painter, R.H. (1993). The binding of C1q to immunoglobulins. *Behring. Inst. Mitt*. **93**, 131-137.

Paliy, O., & Gunasekera, T. S. (2007). Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl. Microbiol. Biotechnol*. **73,** 1169-1172.

Pangburn, M. K., Schreiber, R. D., & Müller-Eberhard, H. J. (1977). Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein beta1H for cleavage of C3b and C4b in solution. *J. Exp. Med.* **146,** 257-270.

Pepys, M. B., & Hirschfield, G. M. (2003). C-reactive protein : a critical update. *J. Clin. Invest.* **111,** 1805–1812.

Perkins, S. J. (1988). Structural studies of proteins by high-flux X-ray and neutron solution scattering. *Biochem J*. **254,** 313-327.

Perkins, S. J., Gilbert, H. E., Aslam, M., Hannan, J., Holers, V. M., & Goodship, T. H. (2002a). Solution structures of complement components by X-ray and neutron scattering and analytical ultracentrifugation. *Biochem. Soc. Trans.* **30,** 996–1001.

Perkins, S. J., & Goodship, T. H. J. (2002b). Molecular modelling of the C-terminal domains of factor H of human complement: a correlation between haemolytic uraemic syndrome and a predicted heparin binding site. *J. Mol. Biol.* **316,** 217–224.

Perkins, S. J., Nealis, A., & Sim, R. (1991). Oligomeric Domain Structure of Human Complement Factor H by X-ray and Neutron Solution Scattering. *Biochemistry* **15,** 1591–1594.

Perkins, S. J., Okemefuna, A. I., Fernando, A. N., Bonner, A., Gilbert, H. E., & Furtado, P. B. (2008). X-Ray and Neutron Scattering Data and Their Constrained Molecular Modeling. *Methods Cell Biol*. **84,** 375-423.

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.

Philo, J. S. (2006). Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal. Biochem.* **354,** 238–246.

Pickford, A.R. & O'Leary, J.M. (2004). Isotopic labelling of recombinant proteins from the methylotrophic yeast Pichia pastoris. *Methods. Mol. Biol*. **278,** 17-33.

Prosser, B. E., Johnson, S., Roversi, P., Clark, S. J., Tarelli, E., Sim, R. B., Day, A., & Lea, S. M. (2007). Expression, purification, cocrystallization and preliminary crystallographic analysis of sucrose octasulfate/human complement regulator factor H SCRs 6-8. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **63,** 480–483.

Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J., Jowitt, T. A., Clark, S. J., Tarelli, E., Uhrin, D., Barlow, P. N., Sim, R. B., Day, A. J., & Lea, S. M. (2007). Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med*. **204,** 2277-2283.

Ralston, G. (1993). Introduction to Analytical Ultracentrifugation, Beckman Instruments, Palo Alto.

Ramadan, M.A., Shrive, A.K., Holden, D., Myles, D.A., Volanakis, J.E., DeLucas, L.J., & Greenhough, T.J. (2002). The three-dimensional structure of calcium-depleted human C-reactive protein from perfectly twinned crystals. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 992-1001.

Rhodes, G. (2000). Crystallography made crystal clear 2nd Edition. Academic Press, New York.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European MOlecular Biology Open Software Suite. *Trends Geneti.* **16,** 276–277.

Richard, D., Ferrand, M., & Kearley, G. J. (1996). Analysis and visualisation of neutron-scattering data. *J. Neutron Res*. **4,** 33-39.

Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mely, Y., Svergun, D. I., & Moras, D. (2011). Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.* **18,** 564-571.

Rodríguez De Córdoba, S., Díaz-Guillén, M. A., & Heine-Suñer, D. (1999). An integrated map of the human regulator of complement activation (RCA) gene cluster on 1q32. *Mol. Immunol.* **36,** 803–808.

Rodríguez De Córdoba, S., Esparza-Gordillo, J., Goicoechea De Jorge, E., Lopez-Trascasa, M., & Sánchez-Corral, P. (2004). The human complement factor H: Functional roles, genetic variations and disease associations. *Mol. Immunol.* **41,** 355-367.

Rodriguez, E., Rallapalli, P. M., Osborne, A. J., & Perkins, S. J. (2014). New functional and structural insights from updated mutational databases for complement factor H, Factor I, membrane cofactor protein and C3. *Biosci. Rep*. **34,** 635-649.

Rodriguez E. (2014). Molecular interactions of human complement C3 and its regulation. *PhD Thesis, University College London.*

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5,** 1-17.

Rother, K., Till., G., & Hansch, G.M. (1998). The complement system 2nd Edition. Springer-Verlag Berlin Heidelberg New York.

Saunders, R. E., Goodship, T. H. J., Zipfel, P. F., & Perkins, S. J. (2006). An Interactive Web Database of Factor H- Associated Hemolytic Uremic Syndrome Mutations: Insights Into the Structural Consequences of Disease-Associated Mutations. *Hum. Mut.* **27,** 21–30.

Scharfstein, J., Correa, E. B., & Gallo, G. R. (1978). Human C4b-binding protein. Association with immune complexes in vitro and in vivo. *J. Clin. Invest.*, **63,** 437–442.

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo EM structure determination. *J. Struct. Biol.* **180,** 519-530

Schoenborn, B.P. (1984). Neutrons in Biology 1st Edition. Springer Science + Business Media, New York.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J*. **78,** 1606-1619.

Schwaeble, W., Zwirner, J., Schulz, T. F., Linke, R. P., Dierich, M. P., & Weiss, E. H. (1987). Human complement factor H: expression of an additional truncated gene product of 43 kDa in human liver. *Eur. J. Immunol*. **17,** 1485-1489.

Seddon, J. M., Gensler, S. G., Milton, R. C., Klein, M. L., & Rifai, N. (2004). Association Between C-Reactive Protein and Age-Related Macular Degeneration. *JAMA*. **291,** 704-710.

Serdyuk, I.N., Zaccai, N.R. & Zaccai, J. (2007). Methods in molecular biophysics 1st Edition. Cambridge University press.

Sharma, A. K., & Pangburn, M. K. (1996). Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10996-11001.

Shrive, a K., Cheetham, G. M., Holden, D., Myles, D. a, Turnell, W. G., Volanakis, J. E., Pepys, M., Bloomer, A., & Greenhough, T. J. (1996). Three dimensional structure of human C-reactive protein. *Nat. Struct. Biol.* **3,** 346–354.

Siegel, B. Y. J., Osmand, A. P., Wilson, M. F., & Gewurz, H. (1975). Interactions of C-reactive protein with the complement system II. C-Reactive protein-mediated consumption of complement by poly-L-lysine polymers and other polycations. *J. Exp. Med.* **142,** 709-721.

Siegel, R. C., & Schumaker, V. N. (1983). Measurement of the association constants of the complexes formed between intact C1q or pepsin-treated C1q stalks and the unactivated or activated C1r2C1s2 tetramers. *Mol. Immunol.* **20,** 53–66.

Sim, R. B., & Reid, K. B. (1991). C1: molecular interactions with activating systems. *Immunol. Today* **12,** 307–311.

Sjöberg, A. P., Trouw, L. a., Clark, S. J., Sjölander, J., Heinegård, D., Sim, R. B., Day, A.J., & Blom, A. M. (2007). The factor H variant associated with age-related macular degeneration (His-384) and the non-disease-associated form bind differentially to C-reactive protein, fibromodulin, DNA, and necrotic cells. *J. Biol. Chem.* **282,** 10894–10900.

Skerka, C., Lauer, N., Weinberger, A., Keilhauer, C., Sühnel, J., Smith, R., Schlotzer-Schrehardt, U., Fritsche, L., Heinen, S., Hartmann, A., Weber, B., & Zipfel, P. F. (2007). Defective complement control of Factor H (Y402H) and FHL-1 in age-related macular degeneration. *Mol. Immunol.* **44,** 3398–3406.

Smith, C. A., Pangburn, M. K., Vogelli, C., & Muller-Eberhard, H. (1984). Molecular Architecture of Human Properdin, a Positive Regulator of the Alternative Pathway of Complement. *J. Biol. Chem.* **259,** 4582–4588.

Soares, D. & Barlow, P. N. (2005). Complement control protein modules in the regulators of complement activators. In *Structural Biology of the Complement System* (Morikis, D. & Lambris, J. D., eds), 19-62. Taylor & Francis, Boca Raton, USA.

Stafford, W. F. (2000). [131 [13] Analysis of reversibly interacting macromolecular systems by time derivative sedimentation velocity. Methods. Enzymol. **323,** 302–325.

Suankratay, C., Mold, C., Zhang, Y., Lint, T. F., & Gewurz, H. (1999). Mechanism of complement-dependent haemolysis via the lectin pathway: Role of the complement regulatory proteins. *Clin. Exp. Immunol.* **117,** 442–448.

Sugiyama, M., Kurimoto, E., Yagi, H., Mori, K., Fukunaga, T., Hirai, M., Zaccai, G., & Kato, K. (2011). Kinetic asymmetry of subunit exchange of homooligomeric protein as revealed by deuteration-assisted small-angle neutron scattering. *Biophys. J.* **101,** 2037-2042.

Svedberg, T. and Pedersen, K.O. (1940). The ultracentrifuge. Oxford University Press, London.

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

Svergun, D. I. (1999). Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys. J.* **76,** 2879-2886.

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

Szalai, A. J., van Ginkel, F. W., Wang, Y., McGhee, J. R., & Volanakis, J. E. (2000). Complement-Dependent Acute-Phase Expression of C-Reactive Protein and Serum Amyloid P-Component. *J. Immunol.* **165,** 1030–1035.

Tang, G., Pang, L., Baldwin, P. R., Mann, D. S., Jiang,W., Reese, I., & Ludtke, S. J. (2007). EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38-46

Thiel, S., Vorup-Jensen, T., Stover, C. M., Schwaeble, W., Laursen, S. B., Poulsen, K., Willis, A. C., Eggleton, P., Hansen, S., Holmskov, U., Reid, K. B. M., & Jensenius, J. C. (1997). A second serine protease associated with mannan-binding lectin that activates complement. *Nature*. **386,** 506-510.

Toniatti, C., Arcone, R., Majello, B., Ganter, U., Arpaia, G., & Ciliberto, G. (1990). Regulation of the human C-reactive protein gene, a major marker of inflammation and cancer. *Mol. Biol. Med.* **3**, 199-212.

van den Elsen, J.M.H. & Isenman, D.E. (2011). A crystal structure of the complex between human complement receptor 2 and its ligand C3d. *Science*. **332**, 608-611.

Venter, R., Huang, C., Farmer, B., Trolard, R., Spicer, L., & Fierke, C. (1995). High-level 2H/13C/l5N labeling of proteins for NMR studies. *J. Biomol. NMR.* **5,** 339–334.

Vik, D. P., & Fearon, D. T. (1987). Cellular distribution of complement receptor type 4 (CR4): expression on human platelets. *J. Immunol.* **138,** 254–258.

Volanakis, J. E. & Kaplan, M. H. (1971). Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proc. Soc. Exp. Biol. Med.* **136**, 612-614.

Volanakis, J. E., Clements, W. L. & Schrohenloher, R. E. (1978). C-reactive protein: purification by affinity chromatography and physiochemical characterization. *J. Immunol. Methods.* **23**, 285-295.

Volanakis, J. E. (1982). Complement activation by C-reactive protein complexes. *Ann. N. Y. Acad. Sci.* **389**, 235-250.

Weiler, J. M., Daha, M. R., Austen, K. F., & Fearon, D. T. (1976). Control of the amplification convertase of complement by the plasma protein beta1H. *Proc. Natl. Acad. Sci. U. S. A.* **73,** 3268–3272.

Weis, J. J., Tedder, T. F., & Fearon, D. T. (1984). Identification of a 145,000 Mr membrane protein as the C3d receptor (CR2) of human B lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **81,** 881–885.

Whaley, B. Y. K., & Ruddy, S. (1976). Modulation of the alternative complement pathway by β1H globulin. **144,** 1147-1163.

Wistow, G., Bernstein, S. L., Wyatt, M. K., Ray, S., Behal, A., Touchman, J. W., Bouffard, G., Smith, D., & Peterson, K. (2002). Expressed sequence tag analysis of human retina for the NEIBank Project: retbindin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts. *Mol. Vision*. **8,** 196-204.

Wong, W., Bai, X. C., Brown, A., Fernandez, I. S., Hanssen, E., Condron, M., Tan, Y. H., Baum, J., & Scheres, S. H. W. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife*. **3,** 3080-3100.

Woo, P., Korenberg, J. R., & Whitehead, a S. (1985). Characterization of genomic and complementary DNA sequence of human C-reactive protein, and comparison with the complementary DNA sequence of serum amyloid P component. *J. Biol. Chem.* **260,** 13384–13388.

Wright, D. W., & Perkins, S. J. (2015). *SCT* : a suite of programs for comparing atomistic models with small-angle scattering data. *J. Appl. Crystallogr.* **48,** 953-961.

Wu, J., Wu, Y.-Q., Ricklin, D., Janssen, B. J. C., Lambris, J. D., & Gros, P. (2009). Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat. Immunol*. **10,** 728-734.

Yu, J., Wiita, P., Kawaguchi, R., Honda, J., Jorgensen, A., Zhang, K., Fischetti, V. A., & Sun, H. (2007). Biochemical analysis of a common human polymorphism associated with age-related macular degeneration. *Biochemistry,* **46,** 8451–8461.

Ziccardi, R. J., & Cooper, N. R. (1979). Active Disassembly of the First Complement Component, C1, by C1 inactivator. J. Immunol. **123,** 788-792.

Zipfel, P. F., Heinen, S., Józsi, M., & Skerka, C. (2006). Complement and diseases: Defective alternative pathway control results in kidney and eye diseases. *Mol. Immunol*. **43,** 97-106.