

**Complement Factor H: Solution Structures and  
Interactions with Ligands**

**Thesis Presented for the Degree of  
Doctor of Philosophy**

**by**

**Azubuiké Ifeanyi Okemefuna**

**Institute of Structural and Molecular Biology,  
University College London**

**June 2009**

I, Azubuiké Ifeanyi Okemefuna, confirm that the 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**

Factor H (FH) is a plasma glycoprotein that plays a central role in regulating the alternative pathway of complement. It is composed of 20 short complement regulator (SCR) domains. The SCR-1/5 fragment is required for decay acceleration and cofactor activity, while the SCR-16/20 fragment possesses binding sites for complement C3d and host cell-surface polyanionic ligands. C3d is a 35-kDa fragment of C3b, the activated form of the central complement protein C3. C-reactive protein (CRP) is an acute-phase reactant that activates complement through the classical pathway but inhibits the alternative pathway. In this thesis, X-ray scattering, analytical ultracentrifugation (AUC) and constrained modelling were used to determine solution structures for the SCR-1/5 and SCR-16/20 fragments, as well as for intact FH. Surface plasmon resonance (SPR) was used in conjunction with these methods to investigate the interactions of FH with C3d and CRP in the fluid and solid phases. Structural studies revealed that at physiological concentrations, SCR-1/5 is monomeric and SCR-16/20 exists in a weak monomer-dimer equilibrium. In the best-fit models, both FH fragments adopt a partially folded-back orientation in solution, and the SCR-16/20 dimer is formed by association of two SCR-20 domains. FH exists in a partially folded-back orientation in solution and forms oligomers. Here, these FH oligomers were shown to exist under different buffer conditions of NaCl concentration and pH, and these two factors also influence the conformation of FH. New, significantly improved molecular structures showed that FH orientation is maintained by short-to-middle distance interdomain interactions. There was no evidence of long distance interactions between the N- and C-terminals of FH. Studies on the interaction of FH with C3d identified a number of multimeric complexes formed between C3d and both SCR-16/20 and native FH. Binding to C3d did not significantly increase the overall length of FH, and this interaction is NaCl concentration-dependent. CRP, another ligand of FH, is known to exist physiologically as a pentamer. Here, CRP is shown to exist in a pentamer-decamer equilibrium under physiological buffer conditions. This equilibrium is NaCl concentration dependent and occurs in both fluid and solid phases. SPR binding studies showed that interaction with CRP inhibits FH oligomerisation, and further identified a novel CRP-binding site within SCR-16/20 of FH. Thus ionic strength-dependence may be a general feature of FH interaction with its ligands. These results provide insight into the complement regulatory activity of FH.

## **Acknowledgements**

I would like to thank my supervisor Professor S. J. Perkins for his guidance and encouragement throughout the period of my work with him. I am also grateful to the other two members of my thesis committee, Dr. Andrew Martin and Professor Martin Crompton, for their advice and support.

I would like to thank my collaborators, including Professor David L. Gordon, Dr Rebecca Ormsby, Dr. Tania Sadlon and Ms. Kim Griggs for their expertise and for supplying the proteins and expression systems used in my studies.

I am very grateful to the UCL graduate school and Overseas Research Fund for a postgraduate scholarship.

My thanks also go to Mr. Jayesh Gor for his invaluable help with experiments and lab work. I thank Dr. Patricia Furtado for her helpful and gentle advice, and my colleagues at UCL for their untiring help, understanding and patience.

And lastly, to my parents and family - thank you all for believing.

<b>Contents</b>	<b>Page</b>
<b>Chapter One</b>	
<b>The Complement Cascade</b>	<b>1</b>
<b>1.1 Introduction to complement</b>	<b>2</b>
<b>1.2 Complement activation</b>	<b>3</b>
1.2.1 C3 in complement activation	5
1.2.2 The classical pathway	7
1.2.3 The alternative pathway	9
1.2.4 The lectin pathway	11
1.2.5 The membrane attack complex	12
<b>1.3 Regulation of complement activation</b>	<b>12</b>
<b>1.4 Biosynthesis of complement proteins</b>	<b>16</b>
<b>1.5 Complement in disease</b>	<b>16</b>
<b>Chapter Two</b>	
<b>Factor H of Complement</b>	<b>19</b>
<b>2.1 Factor H expression and biosynthesis</b>	<b>20</b>
2.1.1 The Factor H gene family	20
2.1.2 Factor H expression and sequence	20
2.1.3 Biosynthesis of Factor H	21
<b>2.2 Structure of Factor H</b>	<b>21</b>
2.2.1 The SCR domain structure	21
2.2.2 Structures of Factor H SCR domains	24
2.2.3 Factor H inter-domain arrangements	27
<b>2.3 Function of Factor H in complement</b>	<b>28</b>
2.3.1 Regulation of complement activation	28
2.3.2 Factor H binding to C3b	30
2.3.3 Factor H binding to CRP	31
2.3.3.1 Introduction to human CRP	31
2.3.3.2 Factor H interaction with CRP as a means of complement regulation	33
2.3.3.3 The role of the interaction between Factor H and CRP in the molecular mechanism of age-related macular degeneration	34

2.3.4 Factor H binding to other ligands	35
<b>2.4 Factor H mutations and disease</b>	<b>35</b>
<b>Chapter Three</b>	
<b>Protein Structure and Interaction Measurement</b>	<b>38</b>
<b>3.1 Introduction to protein structure determination</b>	<b>39</b>
<b>3.2 Analytical ultracentrifugation</b>	<b>41</b>
3.2.1 Introduction to the analytical ultracentrifuge	41
3.2.2 Uses of analytical ultracentrifugation	42
3.2.3. Theory of sedimentation	42
3.2.4. Instrumentation	46
3.2.4.1. The rotor	46
3.2.4.2. The cells	46
3.2.4.3. Absorbance optics	48
3.2.4.4. Interference optics	48
3.2.5. Applications of analytical ultracentrifugation	50
3.2.5.1. Sedimentation velocity	50
3.2.5.2. Sedimentation equilibrium	51
3.2.5.3 Analytical ultracentrifugation data analyses	51
<b>3.3 Solution scattering</b>	<b>55</b>
3.3.1 Introduction to solution scattering	55
3.3.2 Types of solution scattering	56
3.3.3 Theory of solution scattering	56
3.3.4 Properties of X-ray scattering experiments	58
3.3.5 Instrumentation	61
3.3.6 X-ray sources	61
3.3.7 Data interpretation	65
3.3.7.1 Guinier analyses	65
3.3.7.2 Distance distribution function analyses	66
<b>3.4 Biomolecular modelling</b>	<b>69</b>
3.4.1 Scattering curve modelling	69
3.4.1.1 Analysis of protein composition	69
3.4.1.2 Creating atomic models	70
3.4.1.3 Molecular model searches	70

3.4.1.4 Sphere modeling	71
3.4.1.5 Debye scattering curve calculation	72
3.4.2 Identification of best-fit models	72
3.4.3 Calculation of hydrodynamic properties	73
<b>3.5 Introduction to protein interaction measurement</b>	<b>73</b>
3.5.1 Surface plasmon resonance	75
3.5.2 Experiments on the Biacore X100	77
3.5.3 Applications of Biacore systems	79
3.5.3.1 Binding analyses	80
3.5.3.2 Equilibrium analyses	80
3.5.3.3 Concentration assays	81
<b>Chapter Four</b>	
<b>The Regulatory SCR-1/5 and Cell-surface-binding SCR-16/20</b>	<b>87</b>
<b>Fragments of Factor H Reveal Partially Folded-back Solution Structures and Different Self-associative Properties</b>	
<b>4.1 Introduction</b>	<b>88</b>
<b>4.2 Results and Discussion</b>	<b>89</b>
4.2.1 X-ray scattering data for SCR-1/5 and SCR-16/20	89
4.2.2 Analytical ultracentrifugation data for SCR-1/5 and SCR-16/20	97
4.2.3 Constrained scattering modelling of SCR-1/5 and monomeric SCR-16/20	105
4.2.4 Trial models for dimeric SCR-16/20	114
<b>4.3 Conclusions</b>	<b>117</b>
<b>4.4 Materials and Methods</b>	<b>125</b>
4.4.1 Expression and preparation of SCR-1/5 and SCR-16/20	125
4.4.2 X-ray scattering data collection and analysis	126
4.4.3. Sedimentation velocity and equilibrium data for SCR-1/5 and SCR-16/20	128
4.4.4. Modelling of the SCR-1/5 and SCR-16/20 solution structures	129
4.4.5 Debye scattering and sedimentation coefficient modelling	130
4.4.6 Protein data bank accession number	131

<b>Chapter Five</b>	
<b>Electrostatic Interactions Contribute to the Folded-back Conformation of Wild-type Human Factor H</b>	<b>132</b>
<b>5.1 Introduction</b>	<b>133</b>
<b>5.2 Results and Discussion</b>	<b>134</b>
5.2.1 Analytical ultracentrifugation of Factor H	134
5.2.2 X-ray scattering data for Factor H	146
5.2.3 Constrained scattering modelling of monomeric Factor H	150
<b>5.3 Conclusions</b>	<b>158</b>
<b>5.4 Materials and Methods</b>	<b>167</b>
5.4.1 Purification of Factor H	167
5.4.2 Sedimentation velocity data for Factor H	169
5.4.3 X-ray scattering data collection and analysis	170
5.4.4 Modelling of the Factor H solution structures	171
5.4.5 Debye scattering and sedimentation coefficient modelling	172
5.4.6 Protein Data Bank accession number	173
<b>Chapter Six</b>	
<b>Multimeric Interactions Between Complement Factor H and its C3d Ligand Provide New Insight on Complement Regulation</b>	<b>174</b>
<b>6.1 Introduction</b>	<b>175</b>
<b>6.2 Results and Discussion</b>	<b>176</b>
6.2.1 Solution structural data for the Factor H-C3d complex	176
6.2.2 Analytical ultracentrifugation of C3d	179
6.2.2.1 Sedimentation velocity data for C3d	179
6.2.2.2 Sedimentation equilibrium data for C3d	181
6.2.3 Analytical ultracentrifugation of the SCR-16/20 and Factor H complexes with C3d	183
6.2.3.1 Sedimentation velocity data for the SCR-16/20 complex with C3d	183
6.2.3.2 Sedimentation velocity data for the complex of native Factor H with C3d	188
6.2.4 Surface plasmon resonance studies of the complexes of Factor H and SCR-16/20 with C3d	190

6.2.5 X-ray scattering data of the SCR-16/20 complex with C3d	197
6.2.6 Molecular modelling of SCR-16/20 complexes with C3d	201
<b>6.3 Conclusions</b>	<b>205</b>
<b>6.4 Materials and methods</b>	<b>211</b>
6.4.1 Protein expression and purification	211
6.4.2 Analytical ultracentrifugation data and analysis	215
6.4.2.1 Sedimentation velocity data for C3d and its complexes with Factor H and SCR-16/20	215
6.4.2.2 Sedimentation equilibrium data for C3d	216
6.4.3 Surface plasmon resonance data for Factor H and C3d	216
6.4.4 X-ray scattering data collection and analysis	217
6.4.5 Molecular modelling of the SCR-16/20:C3d complexes	218
<b>Chapter Seven</b>	
<b>C-reactive Protein Exists in NaCl-concentration dependent pentamer -decamer equilibrium in physiological buffer containing calcium</b>	<b>219</b>
<b>7.1 Introduction</b>	<b>220</b>
<b>7.2 Results and Discussion</b>	<b>221</b>
7.2.1 Size exclusion chromatography of CRP	221
7.2.2 Sedimentation velocity analyses of CRP	223
7.2.3 X-ray scattering analyses of CRP	226
7.2.4 Modelling of pentameric and decameric CRP	234
7.2.5 Surface plasmon resonance analyses of CRP	237
<b>7.3 Conclusions</b>	<b>239</b>
<b>7.4 Materials and Methods</b>	<b>246</b>
7.4.1 Purification of CRP	246
7.4.2 Analytical ultracentrifugation data and analyses	247
7.4.3 X-ray scattering data collection and analysis	247
7.4.4 Modelling of CRP	248
7.4.5 Surface plasmon resonance data	249

<b>Chapter Eight</b>	
<b>Complement Factor H binds at two distinct sites to C-reactive protein in solution and on surfaces in an NaCl concentration-dependent manner</b>	<b>251</b>
<b>8.1 Introduction</b>	<b>252</b>
<b>8.2 Results and Discussion</b>	<b>253</b>
8.2.1 Sedimentation velocity studies of the Factor H-CRP complex	253
8.2.2 Surface plasmon resonance studies of the Factor H-CRP complex	259
8.2.3 X-ray scattering studies of the Factor H-CRP complex	264
<b>8.3 Conclusions</b>	<b>266</b>
<b>8.4 Materials and Methods</b>	<b>272</b>
8.4.1 Purification of Factor H and CRP	272
8.4.2 Analytical ultracentrifugation data for Factor H-CRP	273
8.4.3 Surface plasmon resonance data for Factor H-CRP	273
8.4.4 X-ray scattering data for Factor H-CRP	274
<b>Chapter Nine</b>	
<b>Conclusions</b>	<b>278</b>
<b>9.1 Prologue</b>	<b>279</b>
<b>9.2 Solution structures of SCR-1/5, SCR-16/20 and full-length Factor H</b>	<b>279</b>
<b>9.3 Interactions of Factor H with C3d and C-reactive protein</b>	<b>280</b>
<b>9.2 Implications for disease</b>	<b>282</b>
<b>References</b>	<b>283</b>
<b>Publications</b>	<b>312</b>
<b>Presentations and abstracts</b>	<b>313</b>

<b>Contents – Figures</b>	<b>Page</b>
<b>Chapter One</b>	
Figure 1.1 The central role of C3 in complement activation	4
Figure 1.2 The crystal structure of C3 showing the 13 domains in different colours	6
Figure 1.3 C3 cleavage and activation	8
Figure 1.4 A tentative scheme for C1 activation as seen from the bottom view of the complex	10
Figure 1.5 Schematic diagram summarizing the three activation pathways of the complement cascade	13
Figure 1.6 Schematic diagram summarizing the regulation of the complement system	17
<b>Chapter Two</b>	
Figure 2.1 A schematic representation of the <i>HF1</i> gene showing its position within the RCA gene cluster.	22
Figure 2.2 Sequence alignment of the 20 SCR domains of human complement FH (PDB accession number P08603)	23
Figure 2.3 FH domain structure	25
Figure 2.4 Inactivation of C3b by FH	29
Figure 2.5 Crystal structure of native CRP pentamer complexed with phosphocholine	32
<b>Chapter Three</b>	
Figure 3.1 The forces acting on a solute particle in a gravitational field	45
Figure 3.2 Sedimentation velocity and equilibrium cell designs	47
Figure 3.3 Schematic diagram of the optical system of the Beckman Optima™ XL-A analytical ultracentrifuge	49
Figure 3.4 Boundary movement in sedimentation velocity and equilibrium experiments	52
Figure 3.5 A schematic diagram showing the scattering angle ( $2\theta$ ) from the main beam	57
Figure 3.6 Definition of the scattering vector, $Q$	59
Figure 3.7 General features of a solution scattering curve showing	60

the Guinier region at low and high $Q$ values	
Figure 3.8 Linear relationship between the molecular weight and neutron $I(0)/c$ values for glycoproteins in 100% $^2\text{H}_2\text{O}$ buffer on the LOQ instrument	63
Figure 3.9 Schematic layout of the X-ray ID02 beamline at ESRF	64
Figure 3.10 Guinier $R_G$ and $R_{XS}$ analyses of SCR-1/5 and SCR-16/20 of factor H	67
Figure 3.11 Distance distribution functions $P(r)$ for SCR-1/5 and SCR-16/20 of factor H	68
Figure 3.12 Flow chart of the automated modelling procedure	74
Figure 3.13 Diagram illustrating the SPR principle	76
Figure 3.14 A Biacore sensorgram of response units versus time	78
Figure 3.15 Illustration of direct and indirect immobilisation (capture) of ligand and the regeneration process in Biacore experiments	83
<b>Chapter Four</b>	
Figure 4.1 Schematic cartoon of the FH SCR-1/5 and SCR-16/20 fragments and their sequences	90
Figure 4.2 Size-exclusion gel filtration and SDS-PAGE of SCR-1/5 and SCR-16/20	92
Figure 4.3 Guinier $R_G$ and $R_{XS}$ analyses for SCR-1/5 and SCR-16/20 monomers	94
Figure 4.4 Guinier $R_G$ and $R_{XS}$ analyses for SCR-16/20 dimer	96
Figure 4.5 Distance distribution function $P(r)$ analyses for SCR-1/5 and SCR-16/20	98
Figure 4.6 Sedimentation velocity $g(s^*)$ analyses of SCR-1/5 using DCDT+	100
Figure 4.7 Sedimentation velocity $c(s)$ distribution analyses of SCR-1/5 using SEDFIT	101
Figure 4.8 Concentration dependence of the observed $s_{20,w}$ values for SCR-1/5 and SCR-16/20	102
Figure 4.9 Sedimentation velocity $g(s^*)$ analyses of SCR-16/20 using DCDT+	103
Figure 4.10 Sedimentation velocity $c(s)$ distribution analyses of SCR-16/20 using SEDFIT	104
Figure 4.11 Summary of the global sedimentation equilibrium fits for SCR-1/5 absorbance data using SEDPHAT	106
Figure 4.12 Summary of the global sedimentation equilibrium fits	107

for SCR-16/20 using SEDPHAT	
Figure 4.13 Structural analyses for 2000 trial models for SCR-1/5 and monomeric SCR-16/20	109
Figure 4.14 Results of fitting 2,000 SCR-1/5 models (white circles) to the scattering data at 0.21 mg/ml	111
Figure 4.15 Summary of 2,000 fits to the monomeric SCR-16/20 scattering data at 0.14 mg/ml	113
Figure 4.16 Best-fit and poor-fit X-ray scattering modelling for SCR-1/5	116
Figure 4.17 Best-fit and poor-fit X-ray scattering modelling for monomeric and dimeric SCR-16/20	118
Figure 4.18 Outcome of the constrained modelling fits for SCR-1/5 and SCR-16/20	121
Figure 4.19 The best-fit SCR-1/5 and SCR-16/20 models (green) are shown in comparison with the scattering models for SCR-6/8 (blue: PDB code 2ic4) and intact FH (red: PDB code 1haqA)	123
Figure 4.20 Electrostatic maps of the SCR-1/5 and SCR-16/20 best-fit models are shown in two 180° front and back rotational views about their longest axes	124
 <b>Chapter Five</b>	
Figure 5.1 Charge summation in the 20 SCR domains of FH	135
Figure 5.2 Sedimentation velocity $c(s)$ distribution analyses of FH	138
Figure 5.3 NaCl concentration-dependence of the sedimentation of FH oligomers	140
Figure 5.4 pH-dependence of the sedimentation of FH oligomers	143
Figure 5.5 Screen shots of SEDFIT sedimentation analyses of FH in 137 mM NaCl, pH 7.4 at a rotor speed of 40,000 r. p. m.	144
Figure 5.6 NaCl concentration and pH dependence of the formation of FH dimer and oligomer	145
Figure 5.7 Determination of the Guinier parameters for FH	147
Figure 5.8 NaCl concentration and pH dependence of the Guinier parameters for FH	149
Figure 5.9 Distance distribution function $P(r)$ analyses for FH	151
Figure 5.10 Comparison of the Guinier fits and distance distribution $P(r)$	152

analyses for native (Gly) and deglycosylated (Degly) FH	
Figure 5.11 Outcome of five constrained modelling searches to determine a FH solution structure	154
Figure 5.12 Best-fit FH X-ray scattering curve fit for the five experimental conditions	156
Figure 5.13 Summary of the effects of NaCl concentration, pH and deglycosylation on FH inter-domain orientation	159
Figure 5.14 The best-fit FH models in 50 mM and 250 mM NaCl	165
Figure 5.15 Electrostatic views of the best-fit model of FH in 137 mM NaCl and its comparison with the crystal structure of C3b	168
<b>Chapter Six</b>	
Figure 6.1 Schematic cartoon of the SCR arrangement in the monomer and dimer of SCR-16/20	177
Figure 6.2 Size-exclusion chromatography of unbound SCR-16/20, C3d and their complex	178
Figure 6.3 Concentration dependence of the sedimentation velocity data for unbound C3d in 137 mM NaCl (green) and 50 mM NaCl (red) buffers	180
Figure 6.4 Multispeed global fit sedimentation equilibrium analyses of C3d in 137 mM NaCl (green) and 50 mM NaCl (red) buffers	182
Figure 6.5 Size distribution $c(s)$ analyses of SCR-16/20 and C3d mixtures in 50 mM NaCl HEPES buffer	184
Figure 6.6 Size distribution $c(s)$ analyses of SCR-16/20, FH and C3d mixtures in 137 mM NaCl PBS or HEPES buffer	186
Figure 6.7 Size distribution $c(s)$ analyses of native FH and C3d mixtures in 50 mM NaCl HEPES buffer	189
Figure 6.8 SPR analysis of C3d binding to immobilised FH	191
Figure 6.9 Time-resolved SPR analysis of C3d binding to immobilised FH	194
Figure 6.10 SPR analyses of ternary complexes formed between C3d, SCR-16/20 and immobilised FH in 137 mM NaCl buffer	195
Figure 6.11 SPR analyses of ternary complexes formed between C3d and FH in the presence of immobilised FH in 137 mM NaCl buffer	196
Figure 6.12 X-ray Guinier $R_G$ analyses for SCR-16/20 and C3d mixtures in 50 mM NaCl HEPES buffer	198

Figure 6.13 X-ray distance distribution functions $P(r)$ for SCR-16/20 and C3d mixtures	200
Figure 6.14 Molecular models for unbound SCR-16/20 and C3d and their complexes	202
Figure 6.15 X-ray scattering fits for SCR-16/20 and its C3d complex	203
Figure 6.16 Schematic cartoons of FH multimer binding to cell surfaces in the presence of surface-bound C3d and heparin-like analogues	212
 <b>Chapter Seven</b>	
Figure 7.1 Size-exclusion chromatography of 1 mg/ml CRP in 140 mM NaCl buffers	222
Figure 7.2 NaCl concentration dependence of CRP decamer formation by AUC in the presence of $\text{Ca}^{2+}$ .	224
Figure 7.3 NaCl concentration dependence of CRP decamer formation by AUC in the absence of $\text{Ca}^{2+}$ .	227
Figure 7.4 NaCl concentration and $\text{Ca}^{2+}$ dependence of CRP pentamer-decamer equilibrium by AUC	228
Figure 7.5 X-ray scattering Guinier $R_G$ analyses for CRP	230
Figure 7.6 Concentration dependence of X-ray scattering Guinier $R_G$ analyses for CRP	231
Figure 7.7 Concentration dependence of the X-ray distance distribution $P(r)$ for CRP	233
Figure 7.8 X-ray scattering curve fits for the CRP pentamer and decamer in the presence and absence of $\text{Ca}^{2+}$	235
Figure 7.9 Concentration dependence of X-ray scattering curve fits for the CRP pentamer and decamer in the presence and absence of $\text{Ca}^{2+}$	236
Figure 7.10 Comparison of the X-ray scattering curves for CRP pentamer and decamer	238
Figure 7.11 SPR binding analysis of CRP self-association	240
Figure 7.12 SPR equilibrium analysis of CRP self-association	241
Figure 7.13 Pentamers and decamers of CRP as a function of NaCl concentration and $\text{Ca}^{2+}$	245

## **Chapter Eight**

Figure 8.1 Electrostatic views of FH and CRP	254
Figure 8.2 AUC analysis of FH-CRP mixtures in <b>(a)</b> 50 mM and <b>(b)</b> 137 mM NaCl	256
Figure 8.3 Surface plasmon resonance analysis of the FH-CRP interaction	260
Figure 8.4 Surface plasmon resonance analyses of FH fragments with immobilised CRP	262
Figure 8.5 X-ray scattering analyses for the FH-CRP interaction in 137 mM NaCl	265
Figure 8.6 Cartoon summarising the observed interactions between FH and CRP	269

<b>Contents – Tables</b>	<b>Page</b>
<b>Chapter Two</b>	
Table 2.1 Experimentally-determined FH domain structures available in the protein data bank (March 2009)	26
<b>Chapter Three</b>	
Table 3.1 Examples of scattering lengths and densities	62
Table 3.2 A summary of the Sensor chips available for Biacore systems	84
Table 3.3 A summary of the covalent coupling chemistry for Biacore systems	85
Table 3.4 A summary of techniques for ligand capture in Biacore systems	86
<b>Chapter Four</b>	
Table 4.1 X-ray scattering and sedimentation coefficient fits for the solution structures of SCR-1/5 and SCR-16/20	115
<b>Chapter Five</b>	
Table 5.1 X-ray scattering and sedimentation coefficient data for monomeric FH and their modelling	157
Table 5.2 Summary of inter-SCR domain angles for the best-fit solution structures of monomeric FH	161
<b>Chapter Six</b>	
Table 6.1 Sedimentation coefficient data and their modelling	204
Table 6.2 Electrostatics of FH-C3d interactions	213
Table 6.3 Ionic strength conditions used for C3 binding studies	214
<b>Chapter Seven</b>	
Table 7.1 X-ray scattering and sedimentation coefficient data and Modelling of CRP	250
<b>Chapter Eight</b>	
Table 8.1 Buffer conditions used in nine studies of the Factor H-CRP Interaction	276

## Abbreviations

Ab	antibody
Ag	antigen
aHUS	atypical haemolytic uraemic syndrome
AMD	age-related macular degeneration
AUC	analytical ultracentrifugation
C1INH	C1 inhibitor
C4BP	C4-binding protein
CR1	complement receptor 1
CR2	complement receptor 2
CRP	C-reactive protein
DAF	decay accelerating factor
DSSP	dictionary of secondary structure prediction
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EM	electron microscopy
ESRF	European synchrotron radiation facility
FH	factor H
FHL-1	FH-like protein 1
FHR	FH-related protein
FI	factor I
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HRF20	homologous restriction factor
IgG	immunoglobulin G
IgM	immunoglobulin M
IL	interleukin
K <sub>D</sub>	dissociation constant
L	length
M	maximum
MAC	membrane attack complex
MACIF	MAC inhibiting factor
MASP-1	mBL-associated serine protease 1
MBL	mannan-binding lectin

MCP	membrane cofactor protein
MIRL	membrane inhibitor of reactive lysis
MPGN	membranoproliferative glomerulonephritis
MW	molecular weight
NMR	nuclear magnetic resonance
PBS	phosphate-buffered saline
PDB	protein data bank
RCA	regulators of complement activation
$R_G$	radius of gyration
r.m.s.d.	root mean square deviation
RPE	retinal pigment epithelium
RU	Resonance units
$R_{XS}$	cross-sectional radius of gyration
SAP	serum amyloid P component
SANS	small-angle neutron scattering
SAS	small-angle scattering
SAXS	small-angle X-ray scattering
SCR	short complement regulator/ short consensus repeat
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	sedimentation equilibrium
SLE	systemic lupus erythematosus
SPR	surface plasmon resonance
SV	sedimentation velocity

## **Amino Acid Abbreviations**

<b>Amino Acid</b>	<b>Three Letter Code</b>	<b>One Letter Code</b>
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

## **Chapter One**

### **The Complement cascade**

### **(1.1) Introduction to complement**

The term “complement” refers to a group of proteins that react with each other in a cascade and work with antibody to eliminate infectious agents from the host. The complement system is not antigen-specific, although some complement proteins are activated by interaction with antibodies. Complement can thus be said to be both a part of the innate and adaptive immune systems (Janeway *et al.*, 2005).

“Complement” as a nomenclature was first used by Ehrlich and Morgenroth to describe a heat-labile component of the immune system that was thought to complement the action of the cells of the immune system. However, the earliest record of complement activity was in the late nineteenth century when fresh, cell-free blood serum was found to lyse bacteria (Nuttall, 1888; Buchner, 1889). Buchner named the observed bactericide activity “alexin”. In an attempt to explain this observation, Jules Bordet demonstrated that this property of serum could be classified into heat-stable and heat-labile components. He showed that red blood cell lysis was dependent on the action of the heat-labile component, and that the heat-stable component was ineffective on its own. The heat-stable component was found to confer immunity against specific microbes (this component corresponds to antibodies), while the heat-labile component was responsible for non-specific antimicrobial activity (this component is now known as complement). Bordet proposed that Buchner’s bactericidal component of normal serum, alexin, was identical to complement as described by Ehrlich and Morgenroth (Bordet, 1909). Bordet is therefore recognised as having discovered the complement system as his complement fixation test not only clearly differentiated between complement and antibodies, but also showed that complement was not merely an auxiliary factor of host immunity.

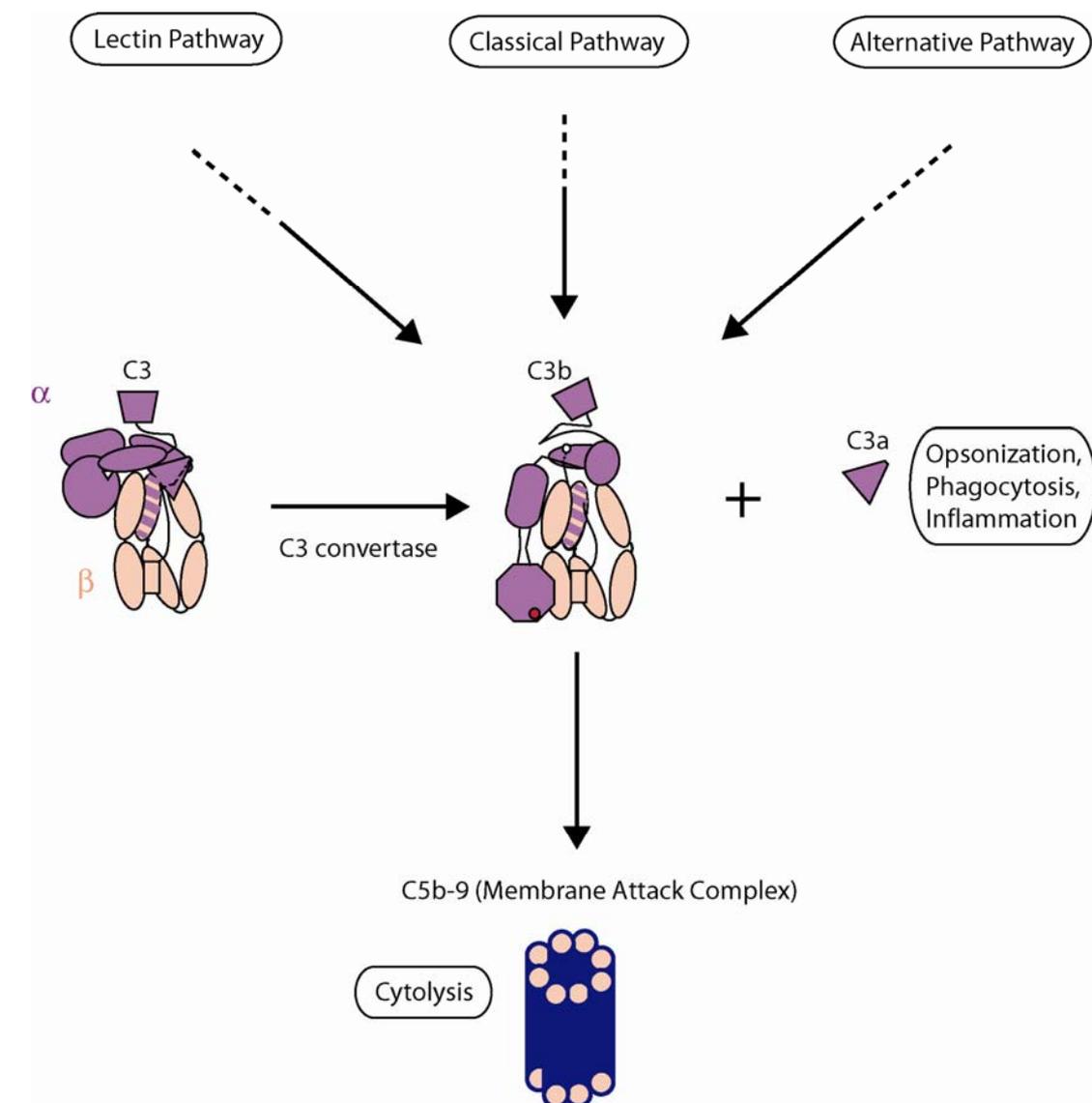
The resolution of the components of complement began in the late nineteenth century with the observation of particulate “globulin” and soluble “albumin” fractions of serum, and that neither of these alone could destroy bacteria (Buchner, 1889). Ferrata (Ferrata, 1907) later reproduced these results with his immune haemolysis experiment, and further showed that the recombined fractions regained lytic activity, and that the “globulin” fraction reacted first with antibody, then the “albumin” fraction reacted second to cause cell lysis. In the early twentieth century, a third component of complement was discovered in two separate studies with cobra venom (Flexner & Noguchi, 1903) and yeast (Coca, 1914), both of which showed the

inactivation of complement by removal of an activity factor. This activity factor was later named the third component ([Whitehead \*et al.\*, 1925](#)), and a fourth component was discovered during the same period ([Gordon \*et al.\*, 1914](#)). Thus within the first half of the twentieth century, it was known that complement consisted of at least four components which were all required to lyse erythrocytes coated with antibody or kill bacteria sensitised with antibody ([Pillemer \*et al.\*, 1941](#)). Through the use of “R” reagents developed during this period for measurement of complement components, it was shown that the order of reactivity of the four components was C1, C4, C2 and C3 ([Bier \*et al.\*, 1945](#)). The development of a large number of new and powerful protein separation techniques during the late-1950s and early-1960s made it possible to purify and characterise most of the complement proteins. At this time Nelson ([Nelson \*et al.\*, 1966](#)) and Müller-Eberhard ([Müller-Eberhard, 1968](#)) characterized the immunologically-activated pathway called the classical pathway. By the late-1960s therefore, complement was known to comprise 11 proteins which reacted in the order of: C1q, C1r, C1s, C4, C2, C3, C5, C6, C7, C8 and C9.

Earlier, in 1954, Pillemer ([Pillemer \*et al.\*, 1954](#)) had described a previously unidentified serum protein named “properdin” which bound to bacteria and yeast, and activated only the latter complement components without appearing to utilize the earlier components. This finding was not generally accepted until the late-1960s and early-1970s, at which time this hypothesis was confirmed by several other independent experiments, ([Ellman \*et al.\*, 1970](#); [Sandberg & Osler, 1971](#); [Götze & Müller-Eberhard, 1971](#)) and the pathway was finally recognised as an “alternative” pathway of complement activation. More recently, the lectin pathway was discovered with the purification of mannan-binding lectin (MBL) from rabbit liver in the late-70s ([Kawasaki \*et al.\*, 1978](#)). Its human homologue was later discovered ([Kawasaki \*et al.\*, 1983](#)), and the lectin pathway is now established as a third pathway of complement activation ([Ikeda \*et al.\*, 1987](#); [Ji \*et al.\*, 1988](#)).

## **(1.2) Complement activation**

Complement activation can take place at the occurrence of any of the following events: (a) Interaction of C1 complement protein with immune complex, aggregated monomeric antibody IgG or pentameric IgM, or C-reactive protein (CRP), (b) spontaneous hydrolysis or molecular perturbation of the C3 complement protein, or (c) Interaction of C4 and C2 with mannan-binding lectin (MBL), a specialised protein



**Figure 1.1**

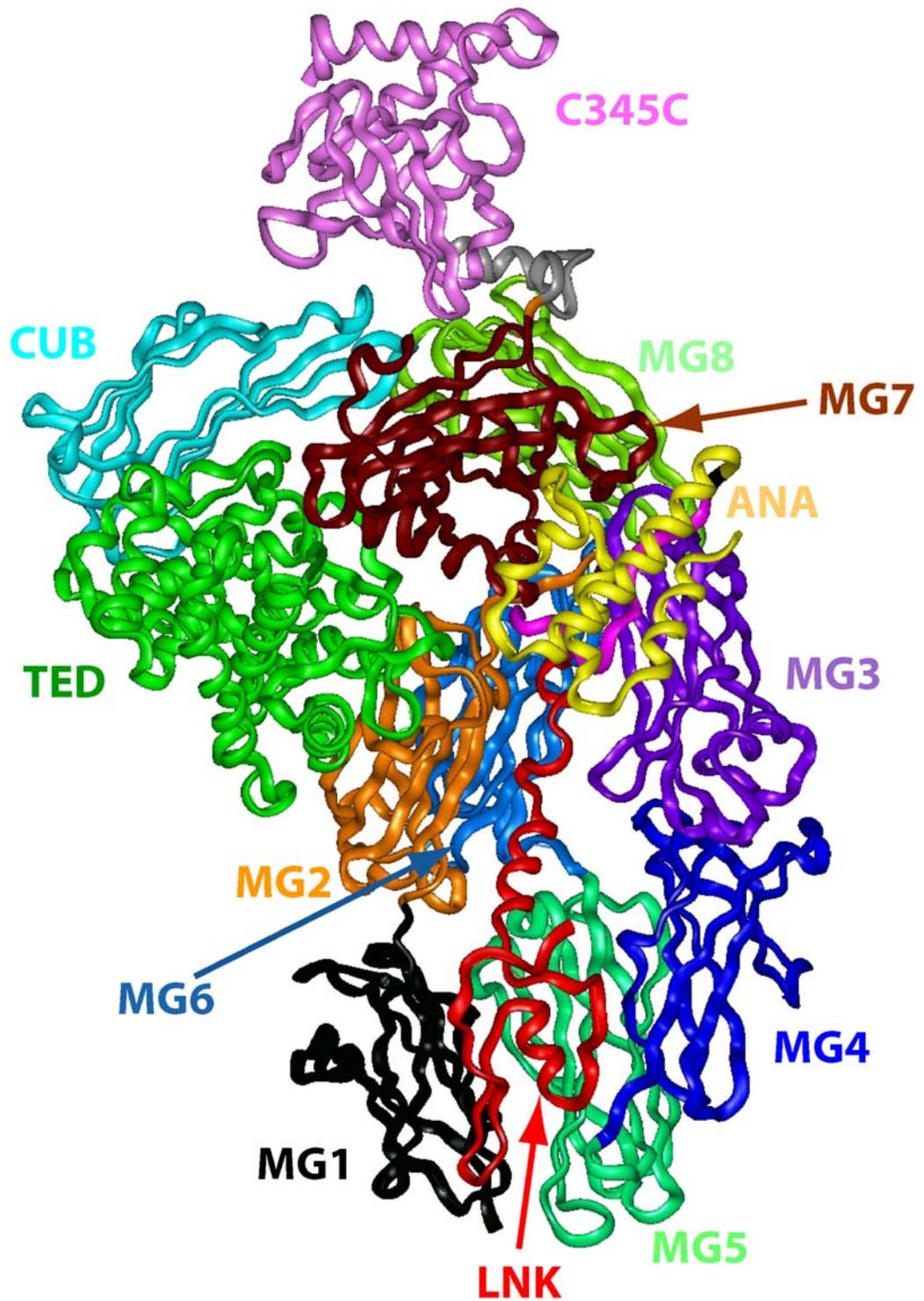
The central role of C3 in complement activation. The  $\alpha$  and  $\beta$  chains of C3 are shown in different colours. Activation of C3 may take place through three pathways leading to the generation of C3b. The endpoint of C3b deposition on the pathogen cell is the formation of the cytolytic membrane attack complex (MAC).

that binds the microbial polysaccharide mannose. These three events represent the classical, alternative and lectin activation pathways of complement respectively (Morley & Walport, 2000). When complement is activated, a cascade of reactions involving other complement proteins, proteases, regulatory proteins and membrane-bound control proteins and receptors takes place, the goal of which is to eliminate foreign particles from the host. Several complement proteins are zymogens. When activated, they become proteases which, in turn, activate other proteins in the complement system by cleaving them to expose reactive groups or active sites. The endpoint of complement activation is the formation of the membrane attack complex (MAC). The complement component C3 is the convergence point for all three pathways of complement activation (Figure 1.1)

### **(1.2.1) C3 in complement activation**

C3 is a 187-kDa protein that belongs to the  $\alpha$ 2-macroglobulin ( $\alpha$ 2M) family. The large proteins (with approximately 1400-1800 amino acid residues) belonging to this family are characterised by homologous sequence features, including a unique thioester site and a central highly variable part (Janssen *et al.*, 2005). Other members of the  $\alpha$ 2M family include the complement components C4 and C5, the proteinase inhibitor  $\alpha$ 2M and the insect and nematode thioester-containing proteins (TEPs). C3 has at least 10 distinct binding sites with which it interacts separately with other complement proteins, receptors and regulators, as well as with non-complement foreign particles such as viral and bacterial proteins (Müller-Eberhard, 1988). C3 is the precursor of a number of biologically active fragments that are released by sequential proteolytic cleavage. Initial proteolytic activation of C3 yields C3a, a small anaphylatoxic fragment, and C3b, a 177-kDa protein and the major active fragment of C3 (Figure 1.1). Deposition of C3b on a target triggers the amplification step of the complement cascade (Section 1.2.3). To prevent complement activation at host surfaces, C3b is degraded by sequential cleavage to iC3b (with the release of C3f), C3c, C3dg and C3d.

C3 consists of two chains  $\beta$  (residues 1-645) and  $\alpha$  (residues 650-1641), which together form a total of 13 domains (Figure 1.2). In the crystal structure of C3 (Janssen *et al.*, 2005), each of the two chains forms six domains, while they are intertwined to form the 13th domain (MG6). Of these 13 domains, there are eight MG



**Figure 1.2**

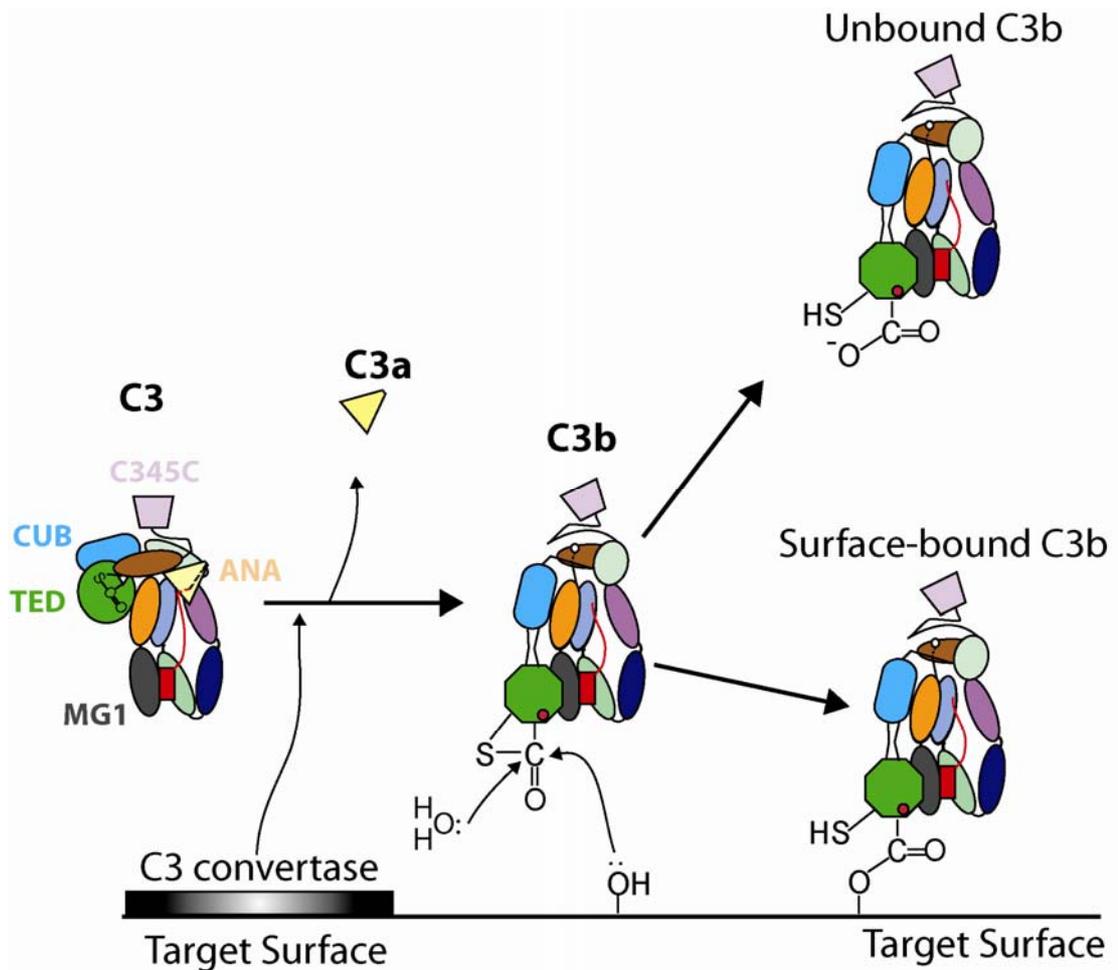
The crystal structure of C3 showing the 13 domains in different colours. INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA) was used to generate a ribbon diagram from PDB entry 2A73 (Janssen *et al.*, 2005).

(macroglobulin) domains and one each of LNK (linker), ANA (anaphylatoxin), CUB, TED (thioester-containing domain) and C345C domains. Domains MG1 - MG5 span residues 1-534, while MG6 spans residues 535-577 ( $\beta$  chain) and 746-806 ( $\alpha$  chain). The two chains within MG6 are interrupted by the LNK (residues 578-645) and ANA (residues 650-726) domains. Two other domains, CUB (residues 912-962; 1269-1330) and TED (residues 963-1268), separate MG7 and MG8 from MG6. MG7 and MG8 occupy residues 807-911 and 1331-1474 respectively. Within the CUB domain is the polypeptide chain that is cleaved by factor I, first at the sites Arg1281-Ser1282 and Arg1298-Ser1299 to release the C3f fragment, then at or around the site Arg932-Glu933 to yield the fragments C3dg and C3c (Rother *et al.*, 1998). The last domain, C345C, is located between residues 1496-1641. The cleavage site (Arg726-Ser727) for C3 convertases is located in a disordered surface loop (residues 720-729), part of which comprises the N-terminal region of C3b after cleavage. The ANA domain forms the small C3a fragment of C3 when excised, and the TED domain corresponds to the C3d fragment of C3. During C3 activation, cleavage of the ANA domain (C3a) triggers a large conformational change involving the TED domain, which reveals the previously hidden reactive thioester group (Figure 1.3).

### (1.2.2) The classical pathway

In the classical pathway, the multiple-subunit protein C1 has long been recognised for its ability to bind pathogens either directly or after their recognition by antibodies or C-reactive protein (CRP) (Jiang *et al.*, 1991; Gaboriaud *et al.*, 2004). C1 has three subunits q, r and s which are products of different genes. Within the C1 complex, the active sites of these subunits are positioned in close proximity, and activation of one of the subunits leads to activation of the others in turn. The C1q subunit is the recognition subunit and makes the first contact with the primary activating body, thus recognising abnormal structures from self.

C1q normally circulates in blood as a  $\text{Ca}^{2+}$ -dependent complex with two molecules each of the serine protease zymogens C1r and C1s. It is composed of three subunits arranged in a collagen-like triple helix with an N-terminal “stalk” and six globular “flower heads”. An activator-binding site is located in each of the six heads. A change in its conformation is stimulated when more than one of these six “flower heads” interacts with an activated surface. C1q interacts readily with antibodies of the



**Figure 1.3**

C3 cleavage and activation. The colour scheme follows that of [Figure 1.2](#). The reactive thioester group is located in the TED domain and hidden in inactive C3. Upon cleavage, conformational changes and domain rearrangement occur within C3 to expose the reactive thioester group, thus forming C3b. Nascent C3b can either be stabilised by reacting with active groups on the target surface or inactivated by reacting with water molecules in the fluid phase.

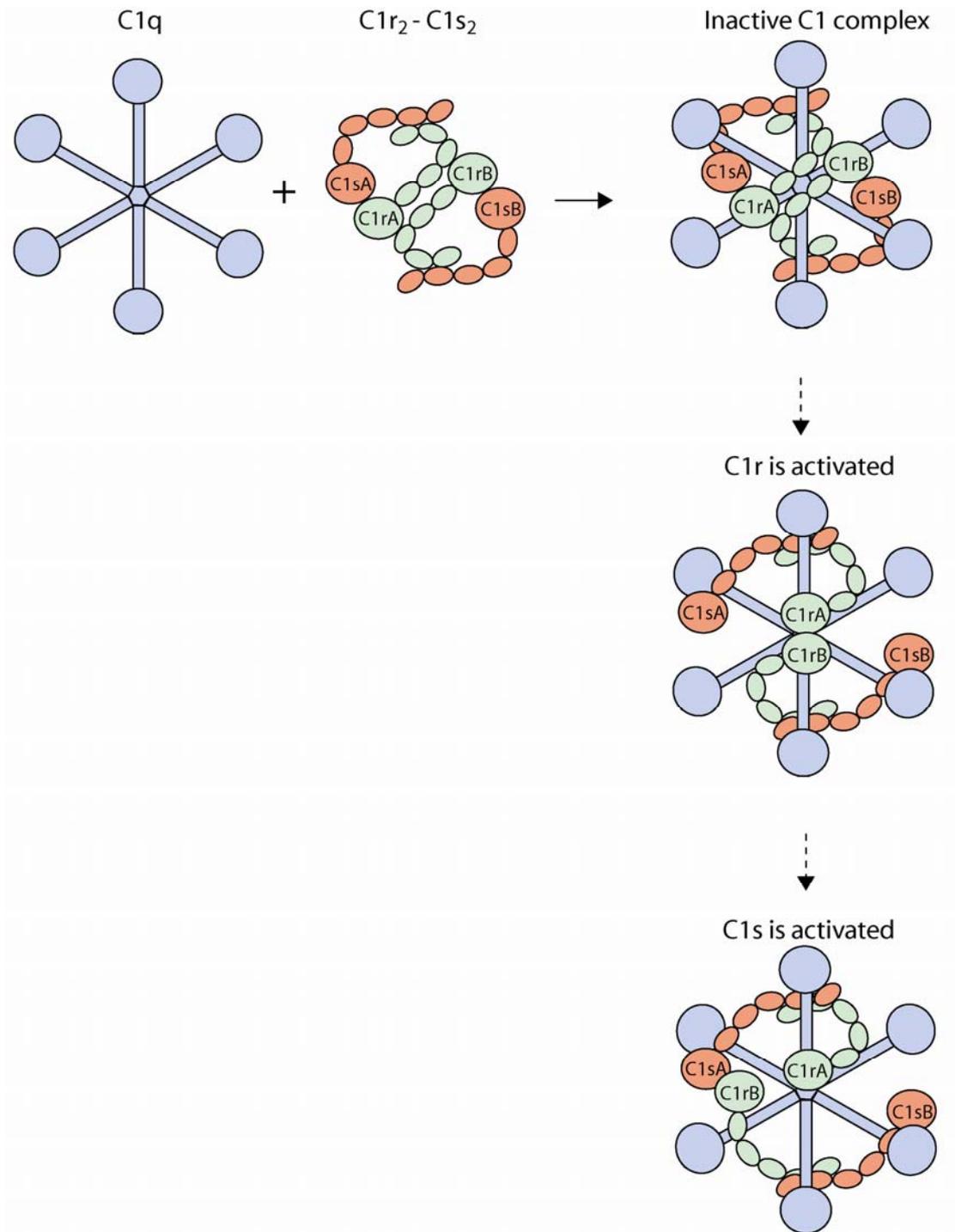
IgG or IgM class (though relatively weakly with single monomeric units of IgG) and CRP. It has also been reported to bind to apoptotic cells (Navratil *et al.*, 2001; Botto & Walport, 2002), with growing evidence of interaction with  $\beta$ -amyloid fibrils (Rogers *et al.*, 1992; Tacnet-Delorme *et al.*, 2001) and the pathological form of the prion protein (Mabbott *et al.*, 2001; Klein *et al.*, 2001). In the case of immune complexes, its “flower heads” readily interact with the C<sub>γ</sub>2 domains of adjacent IgG molecules or with the C<sub>μ</sub>3 domain of pentameric IgM. Occupation of two or more of its activator sites induces a conformational change in C1q, which causes autoactivation of C1r from a single-chain zymogen to a two-chain, disulphide-linked active protease. C1r then cleaves a single peptide bond in the C1s zymogen to form the two-chain active protease C1s (Volanakis & Frank, 1998). The C1 activation scheme is summarized in Figure 1.4.

On activation, C1s cleaves C4 to yield a small C4a and a larger C4b particle. The C4a particle is slightly inflammatory and chemoattractant, and is released into plasma. C4b has an exposed metastable thioester bond with which it binds to hydroxyl and amino groups on the activating surface. Following the deposition of C4b on the activating surface, C2 is covalently attached to it in the presence of magnesium ions. C2 is then cleaved by C1s to yield C2b and C2a. C4b and C2a combine to form C4b2a, which is the C3 convertase of the classical pathway. This convertase cleaves C3 to C3a and C3b. C3b, like C4b, possesses a metastable thioester which enables its deposition on the activating cell surface. Further binding of C3b by C4b2a forms the C5 convertase of the classical pathway, C4b2a3b, which cleaves C5 into C5a and C5b. C3a and C5a are released into plasma to perform inflammatory, anaphylactic and chemoattractive functions.

C1q interacts with CRP within the collagenous region just below its globular heads (Jiang *et al.*, 1991). In a proposed interaction model, electrostatic interactions allow the positively-charged C1q head to fit into the central negatively-charged central pore of pentameric CRP. This interaction involves all three subunits of C1q, in contrast to the model for C1q-IgG interaction, which involves only a single C1q subunit (Gaboriaud *et al.*, 2004).

### **(1.2.3) The alternative pathway**

Activation of complement by the alternative pathway is initiated by a variety of cellular surfaces, including those of various pathogenic bacteria, parasites, viruses,



**Figure 1.4**

A tentative scheme for C1 activation as seen from the bottom view of the complex. In the inactive complex, the zymogen C1r<sub>2</sub>-C1s<sub>2</sub> is in an 8-shaped, closed conformation. Conformational change of C1q triggers autoactivation of C1r, leading to an S-shaped open conformation and subsequent activation of C1s (Gaboriaud *et al.*, 2004).

virally-infected cells and fungi (Pangburn & Müller-Eberhard, 1984) without antibody involvement. C3 is directly activated by either spontaneous hydrolysis, which constantly takes place in plasma at very low levels (C3 tick-over), or so-called non-specific molecular perturbations. These perturbations either allow water into the globular C3 molecule to hydrolyse the otherwise hidden thioester bond between the amino acid residues Cys-1010 and Gln-1013 of the sequence Cys-Gly-Glu-Gln located in the  $\alpha$ -chain, or cause this bond to be exposed to water in plasma (Rother *et al.*, 1998). The contribution of this process to complement activation is minimal unless the activated product (C3\*) is stabilized by proteins such as plasmin and other inflammation products. This is because C3\* is unstable and readily reacts with water molecules to inactivate the reactive thioester group (Law & Reid, 1995).

Hydrolysis of C3 yields active C3.H<sub>2</sub>O (C3\*), whose exposed reactive thioester group readily binds to foreign targets. C3\* binds the serum protein factor B. Following this, C3\*-bound factor B is cleaved by the protease factor D to yield a small Bb fragment. The resulting enzyme complex, C3\*Bb, is the second C3 convertase. C3\*Bb cleaves the  $\alpha$ -chain of inactive C3 between Arg-726 and Ser-727 to yield C3a and C3b. This cleavage induces a conformational change that makes the thioester bond briefly accessible for nucleophilic attack, either by water molecules or hydroxyl and amino groups on the target surface. If the thioester bond is attacked by water, Gln-1013 is converted to Glu-1013 and C3b remains unbound and rapidly becomes inactive. If the thioester bond reacts with an active hydroxyl or amino group on a target surface, C3b is deposited on that surface and retains its active conformation (Figure 1.3). It then combines with factor B, which is cleaved by factor D, to produce more molecules of C3bBb. Whereas initial spontaneous hydrolysis of C3 takes place only very slowly, activation of subsequent C3 molecules by C3bBb takes place very fast and at exponential rates. This is the amplification step of complement activation (Figure 1.5), and it can theoretically proceed indefinitely until one or more of the reactants are exhausted.

#### **(1.2.4) The Lectin pathway**

In this pathway, complement is activated independent of antibody by interaction of C4 and C2 with specialised proteins that bind to mannose, a sugar found in microbial surface polysaccharides. MBL (Mannose-Binding Lectin) is produced by the liver in response to microbial infection. Structural similarities of

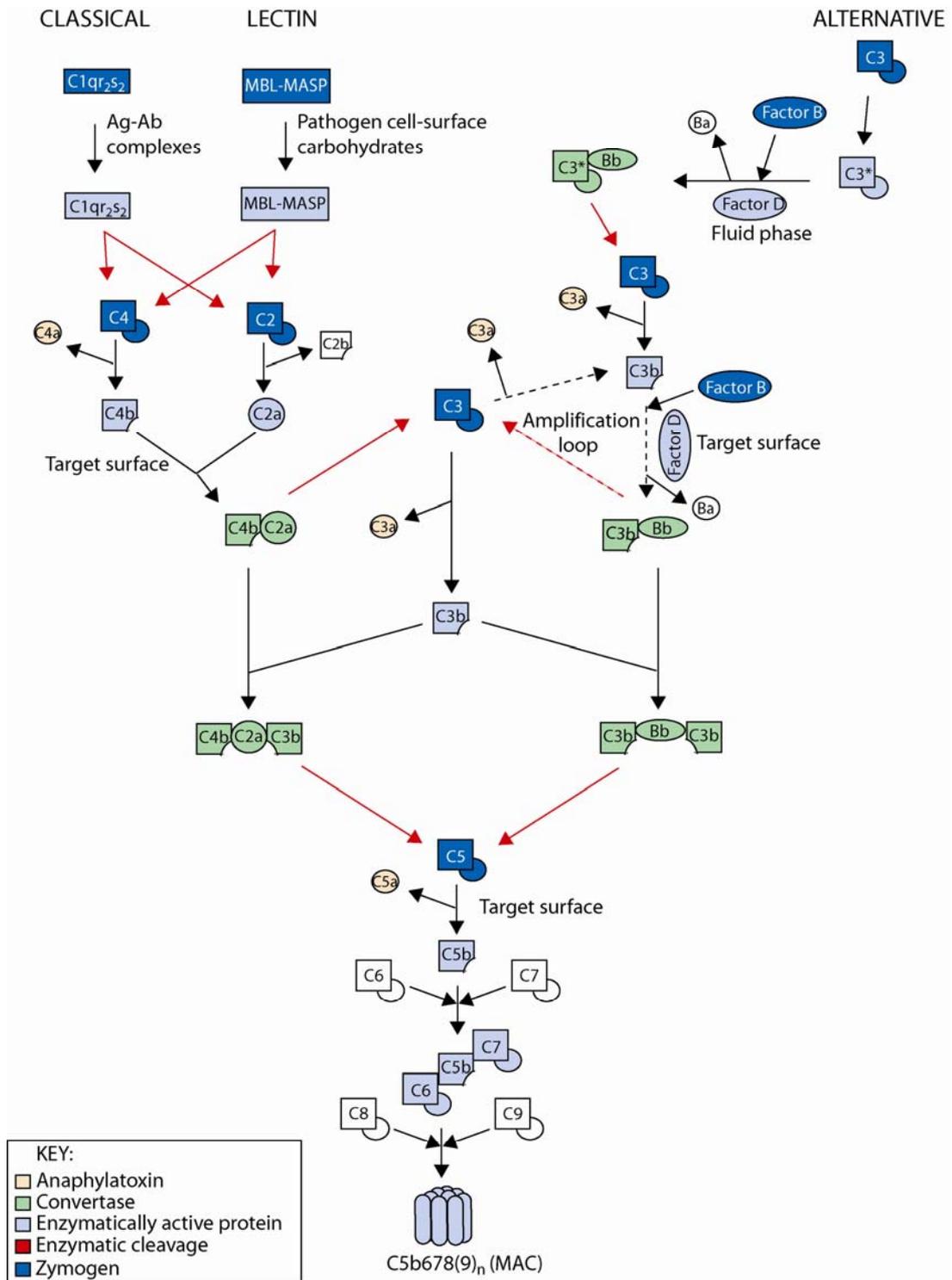
MBL, MASP-1 (MBL-associated serine protease 1) and MASP-2 with C1q, C1r and C1s respectively provide some explanation as to how this occurs. MBL is known to have two to six “heads” in the same way as C1q and is associated with MASP-1 and MASP-2 in the same way as C1q with C1r and C1s.

MBL binds specifically to mannose and certain other sugars such as fucose with a spacing pattern that permits its attachment on the surface of microbial cells. These residues are usually shielded by other sugar groups such as sialic acid in vertebrate cells, and this provides an important way of differentiating between self and non-self. Interaction with these sugar residues triggers a conformational change in MBL, which stimulates a conversion of the zymogen MASP-2 into a protease. MASP-2 cleaves C4 and C2 in turn to produce the C3 convertase. Cells bound by MBL are hence more susceptible to phagocytosis (Reid, 1998).

#### **(1.2.5) The Membrane Attack Complex (MAC)**

The membrane attack complex is a group of soluble hydrophilic proteins C5b, C6, C7, C8 and C9. Through a non-enzymatic process of self assembly, they form a protein complex containing up to 18 molecules of C9 which behaves as an integral membrane protein complex or transmembrane channel.

Binding of the first C9 molecule to the MAC appears to stimulate conformational changes that enhance the binding of subsequent C9 molecules to the complex. The initial C5b6789 complex acts as a core for the formation of a C9 barrel-like polymer that bores through the cell membrane of the target cell. In general, about ten to eighteen C9 molecules are thought to be involved, and it is believed that involvement of greater numbers of C9 molecules in MAC formation improves the lytic reactivity of the complex. The channel thus created allows the transmembrane equilibration of cellular solutes, efflux of cell contents and osmotic lysis by the explosive influx of water. This is referred to as reactive lysis (Rother *et al.*, 1998). The most important function of complement is to facilitate and enhance the uptake and destruction of pathogens by phagocytes. In addition to this, complement can also directly lyse cells and play a role in the clearance of immune complexes and inflammation. The pathways of complement activation are summarized in [Figure 1.5](#).



**Figure 1.5**

Schematic diagram summarizing the three activation pathways of the complement cascade.

### **(1.3) Regulation of Complement Activation**

Complement activation is largely localised and confined to the surface on which it commences. Regulation ensures that complement activation is initiated only on non-self surfaces, otherwise its hugely destructive capability would be turned on host cells and the complement proteins in circulation would be quickly depleted. This scenario is exemplified by deficiencies of complement regulatory proteins, which lead to excessive complement activation and pathologic states. One means of regulating complement activation by the host is selective binding of complement proteins to non-host cell surfaces. Hence surface markers to which complement proteins bind, when also present on host cells, are protected by other groups that prevent their interaction with complement proteins. For example, mannose residues with which MBL interacts are shielded by other sugars on host cell surfaces. In addition, several of the activated components are either unstable and decay easily, or have labile binding sites that display only transient ability to bind to target surfaces. Stabilising proteins are present on target surfaces to perpetuate the effect of such components. For example, the C3 convertase C3bBb decays rapidly and is stabilised on target surfaces by properdin, also known as factor P. Properdin binds to many microbial surfaces and thus may be regarded as a positive regulatory factor of complement.

On the surface of host cells, certain surface-expressed regulatory proteins ensure that complement activation is rapidly brought under control. Complement receptor 1 (CR1), decay activating factor (DAF) and membrane cofactor protein (MCP) are examples of such proteins. CR1 is a receptor for C3b and C4b and competes with factor B in binding to the reactive sites of these complement proteins. On binding C3b and C4b, it serves as a cofactor for their factor I-mediated cleavage and consequent inactivation. Furthermore, CR1 also accelerates the decay of C3 and C5 convertases, effectively ensuring that their activity is inhibited on the host cell surfaces. A degree of redundancy is observed in the action of most complement regulatory proteins, thus DAF also accelerates the decay of C3 and C5 convertases on host cell surfaces and MCP acts as cofactor for the cleavage of C3b and C4b by factor I.

Regulation is not restricted to the activity of proteins expressed at host cell surfaces. A number of soluble plasma proteins also play a role in regulating complement activation. C1 is inhibited by C1 inhibitor (C1INH), a serine protease

inhibitor (serpin) that binds to sites on activated C1r and C1s, inactivating them and causing their dissociation from C1q (Law & Reid, 1995). C1INH limits the length of time C1 stays activated and able to cleave C4 and C2, preventing over-activation of complement through the classical pathway. High plasma levels of cleaved fragments of C4 and C2 have been used in the past as pointers to functional deficiency of C1INH.

The activity of the anaphylatoxins C5a, C3a and C4a is regulated in plasma by the anaphylatoxin inactivator serum carboxypeptidase N. This enzyme inactivates the proteins by cleaving them at the C-terminal arginine residue. C4b-binding protein (C4BP) and Factor H (FH) are plasma proteins that regulate the activity of C4b and C3b respectively. C4BP binds C4b and acts as a cofactor for its cleavage by the serine protease factor I. It also accelerates decay of the C3 convertase C4b2a and the C5 convertase C4b3b2a. FH performs the same functions in the alternative pathway; as a cofactor for the factor I-mediated cleavage of C3b and decay-acceleration of the C3 convertase C3bBb.

Factor I is a plasma serine protease that cleaves and inactivates C4b and C3b. It circulates in plasma in active form but can only cleave C3b and C4b when they are bound by cell receptors or their regulators FH and C4BP. It cleaves C3b to yield iC3b (inactive C3b) with the release of C3f, and then further to yield C3c, C3dg and C3d. These products of C3b breakdown also play a role in opsonisation and clearance of immune complexes when bound to appropriate receptors. C3d remains bound to the pathogen surface and is bound by complement receptor type 2 (CR2) which, when cross-linked to a B cell receptor through a C3d-antigen complex, amplifies a signal transduction cascade. Thus CR2 provides an important link between the innate and adaptive immune systems. Factor I also cleaves C4b to yield C4c and C4d.

The membrane attack complex is regulated by a number of proteins, both membrane-bound and soluble. CD59, also known as homologous restriction factor (HRF20), MAC inhibiting factor (MACIF) or membrane inhibitor of reactive lysis (MIRL), is a protein expressed on the surface of host cells that inhibits the assembly of C9 polymer in the C5b-8 complex and hence protects host cells from reactive lysis (Zalman *et al.*, 1986). The activity of CD59 is thought to be responsible for the significantly reduced lytic activity of MAC on homologous cells. In fluid phase, S-protein (site-specific protein), a serum protein synonymous with vitronectin (serum spreading factor), has been found to bind to a metastable binding site on C5b-7,

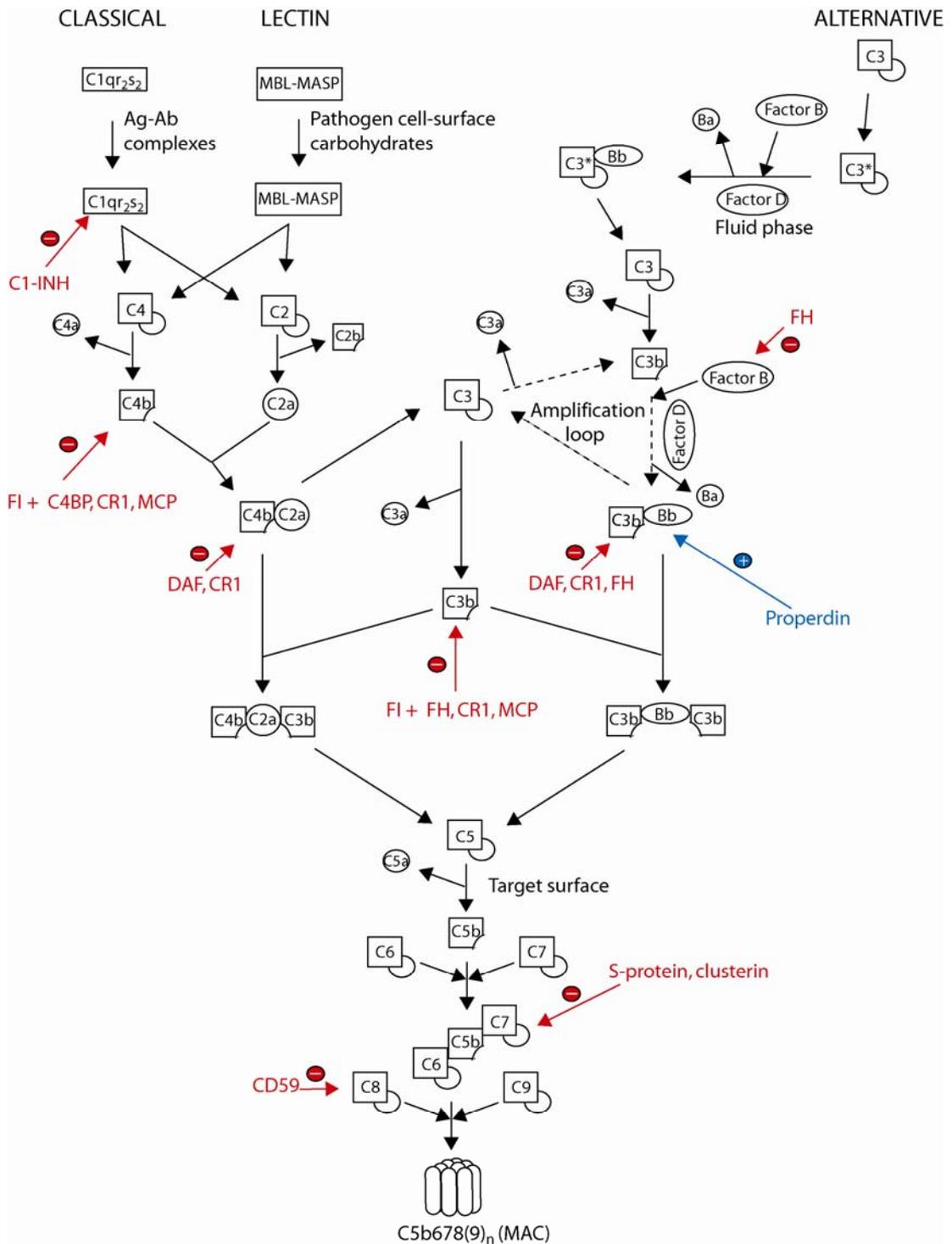
preventing its insertion into cell membranes. Clusterin (also known as apolipoprotein SP40, 40) performs a similar function. In both cases, C8 and C9 binding and subsequent assembly of the MAC may still occur, but the assembled complex is lytically inactive (Liszewski & Atkinson, 1998). Regulation of the complement pathway is summarised in Figure 1.6.

#### **(1.4) Biosynthesis of complement proteins**

Most soluble complement proteins are primarily synthesized in the liver, with varying degrees of extrahepatic biosynthesis. Important exceptions are C1q, factor D, properdin and C7. Extrahepatic synthesis may provide important sources of localised immune response to invasion. The primary sites of C1q synthesis are macrophages (Schwaeble *et al.*, 1995). In addition to secreting C1q, macrophages are thought to also display the molecule on their surfaces where it binds to a variety of molecules to activate complement (Law & Reid, 1995). The major site of synthesis of factor D is adipose tissue (White *et al.*, 1992), and it is thought to probably play a role in fat metabolism distinct from its complement function. Properdin is synthesised in monocytes, T-cells and granulocytes (Wirthmueller *et al.*, 1997) while C7 is mainly synthesised in granulocytes (Hogasen *et al.*, 1995).

#### **(1.5) Complement in disease**

Deficiencies in classical pathway activation components (C1 complex, C4 and C2) are generally associated with immune complex disorders and systemic lupus erythematosus (SLE)-like symptoms. In the case of C4, immune complex disorder is correlated with deficiency of its C4A isotype only (Law & Reid, 1995). C3 deficiency leads to increased susceptibility to pyogenic infections, membranoproliferative glomerulonephritis and SLE-like illness (Morley & Walport, 1999), however deficiencies in the alternative pathway are of relatively low frequency. MBL deficiency is one of the most commonly found in humans and leads to increased risk of suffering from recurrent infections and opsonin deficiency in both children (Super *et al.*, 1989) and adults (Summerfield *et al.*, 1995). MBL mutations and deficiency have been shown to have an increased frequency in cases of SLE (Davies *et al.*, 1995; 1997). Deficiencies in the terminal complex (MAC) components are associated with recurrent *Neisserial* infections, possibly because of the ability of these microorganisms to survive intracellularly in phagocytes. However, individuals



**Figure 1.6**

Schematic diagram summarizing the regulation of the complement system. Positive regulators are shown in blue while negative regulators are shown in red.

deficient in C9 are mostly healthy. MAC activation also plays an important role in ischaemia-reperfusion (I/R) tissue injury (Reidemann & Ward, 2003). Deficiencies in complement regulatory proteins lead to uncontrolled complement activation and disturbed homeostasis. Deficiency in the C1-inhibitor C1INH causes depleted levels of C4 and C2 in serum, often with recurrent localised oedema on the skin, gastrointestinal tract or larynx. Individuals with this deficiency may also develop SLE-like symptoms (Davis, 1998). Factor I deficiency leads to increased C3 activation, increased incidence of recurrent pyogenic infection and SLE-like illness (Vyse *et al.*, 1996). Low levels of FH in plasma are associated with uncontrolled C3 activation leading to membranoproliferative glomerulonephritis (MPGN) and renal failure (Ault *et al.*, 1997), as well as increased susceptibility to meningococcal infections (Nielsen *et al.*, 1989). Deficiencies in FH function due to mutations and polymorphisms are implicated in atypical haemolytic uraemic syndrome (aHUS) (Zipfel *et al.*, 2006), and age-related macular degeneration (AMD) (Klein *et al.*, 2005; Haines *et al.*, 2005; Edwards *et al.*, 2005). Deficiencies in DAF, CD59 and other GPI-anchored proteins result in an inability of the proteins to insert into the cell membrane, and lead to paroxysmal nocturnal haemoglobinuria (Nicholson-Weller *et al.*, 1983; Pangburn *et al.*, 1983).

## **Chapter Two**

### **Factor H of complement**

## **(2.1) Factor H expression and biosynthesis**

### **(2.1.1) The Factor H gene family**

FH is the best-characterised protein of the FH gene family, members of which belong to the regulators of complement activation (RCA) family of proteins. First identified in 1965 as  $\beta$ 1H globulin (Nilsson & Müller-Eberhard, 1965), it is a 154-kDa (Fenaille *et al.*, 2007), soluble glycoprotein that circulates in human plasma at a concentration of 235-810  $\mu$ g/ml (Saunders *et al.*, 2006). The FH gene family provides a prime example of a multidomain multifunctional protein family whose individual members are defined by conserved common structural elements and display diverse but often overlapping functions. There are six identified members of this protein family of secreted plasma proteins that are primarily synthesized in the liver, and they are all thought to have similar roles in complement control (Zipfel *et al.*, 1999). FH-like protein 1 (FHL-1 or reconectin) shares complement regulatory functions with FH and interacts with heparin. It displays cell spreading activity and binds to the N-terminus of the streptococcal M protein (Zipfel & Skerka, 1999). The functions of the FH-related proteins (FHR-1 to FHR-4) are not well defined. FHR-1, FHR-2 (Zipfel *et al.*, 2002) and FHR-4 (Skerka *et al.*, 1997) are constituents of lipoproteins, while FHR-3 is known to interact with heparin (Hellwage *et al.*, 1999). Binding to C3b and C3d has been demonstrated for FHR-3 and FHR-4 and the two proteins display cofactor-related activity (Hellwage *et al.*, 1999). FHR-4 has recently been shown to bind C-reactive protein and may have a role in the opsonisation of necrotic cells (Mihlan *et al.*, 2009).

### **(2.1.2) Factor H expression and sequence**

FH is found at the RCA gene cluster and is encoded by a single gene (*HF1*) located on human chromosome 1q32 (Rodriguez de Córdoba *et al.*, 1999). *HF1* is closely linked to the *FHR1*, *FHR2*, *FHR3*, *FHR4* and *FHR5* genes encoding five FH-related human plasma proteins (Zipfel *et al.*, 1999) (Fig. 2.1). Sequence analyses of the *HF1/FHR1–5* gene region shows the existence of a number of large genomic duplications within this region, including different exons of the *HF1/FHR1–5* genes (Rodriguez de Córdoba *et al.*, 2004). *HF1* comprises 23 exons distributed over 94 kb of genomic DNA (Male *et al.*, 2000) (Fig. 2.1). Each SCR (short complement regulator) domain in FH is encoded by a single exon, except for SCR-2 that is

encoded by exons 3 and 4. Exons 8–9 and 21–23 of *HFI* are duplicated in *FHR3* and *FHR1* (Rodriguez de Córdoba *et al.*, 2004).

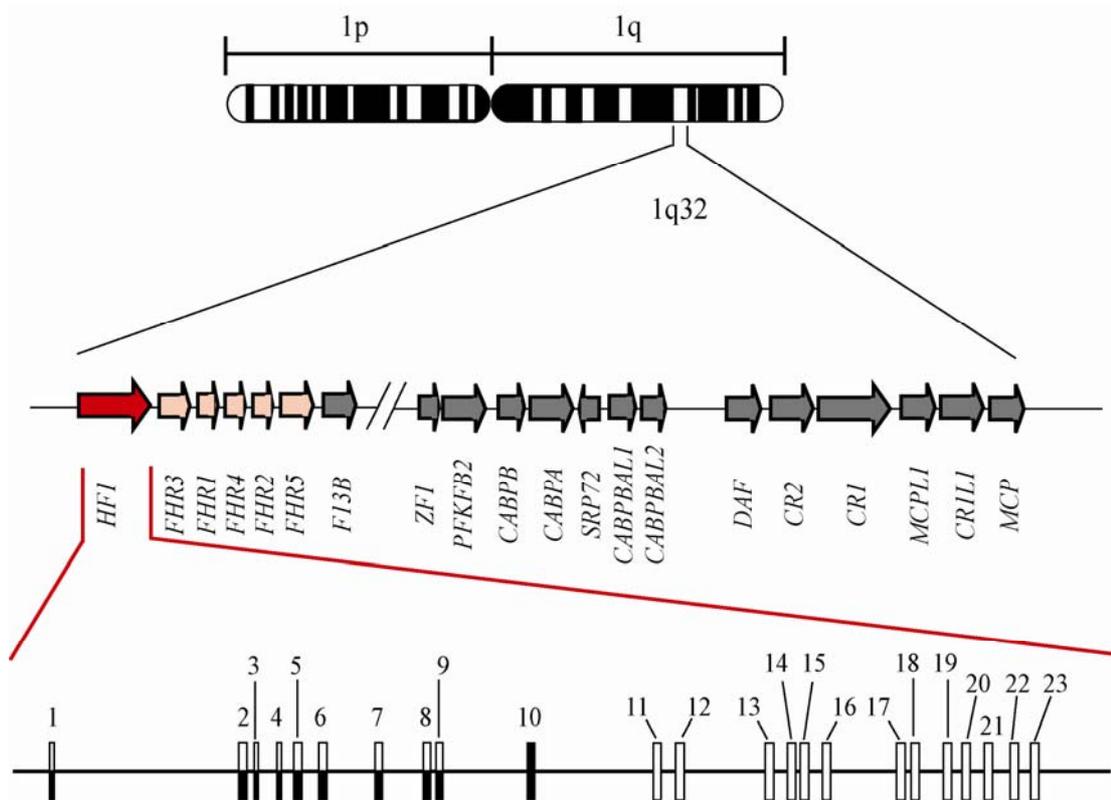
### **(2.1.3) Biosynthesis of Factor H**

FH is synthesized as a single-chain molecule of 1231 amino acid residues, including an 18-residue leader sequence (Ripoche *et al.*, 1988). It is constitutively produced by the liver (Schwaeble *et al.*, 1987), although extrahepatic synthesis also occurs in a wide variety of cell types, such as peripheral blood lymphocytes, myoblasts, rhabdomyosarcoma cells, fibroblasts, umbilical vein endothelial cells, glomerular mesangial cells, neurons, and glia cells (Friese *et al.*, 1999). FH synthesis has also been reported to occur in the human retinal pigment epithelium (RPE) and choroid (Wistow *et al.*, 2002; Klein *et al.*, 2005). The proinflammatory cytokine IFN $\gamma$  and T cell growth factor (Brooimans *et al.*, 1989), as well as the anti-inflammatory glucocorticoid dexamethasone (Friese *et al.*, 2000), have been reported to induce and regulate FH biosynthesis. Particularly strong inductive effects have been reported with IFN $\gamma$  and dexamethasone, both of which may lead to increased local FH biosynthesis in response to inflammation. The inductive effect of dexamethasone is consistent with the presence of a glucocorticoid responsive element in the TATA-less human *HFI* promoter (Ward *et al.*, 1997).

## **(2.2) Structure of Factor H**

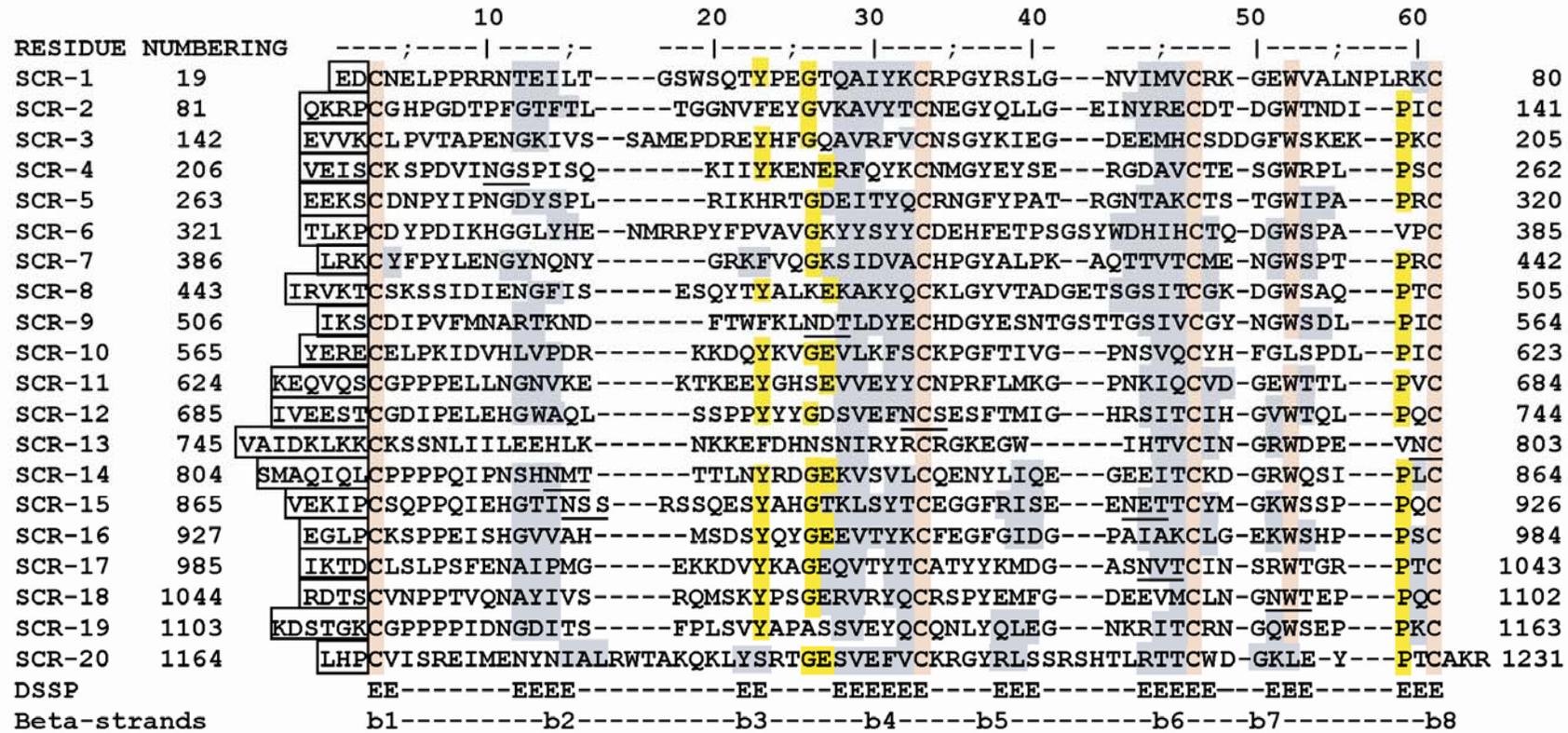
### **(2.2.1) The SCR domain structure**

SCR (short complement regulator/ short consensus repeat) domains are the most abundant domain type in the complement system and are connected by inter-SCR linkers, the length of which generally determines the degree of flexibility between two adjacent SCR domains within the molecule. FH is composed entirely of 20 SCR domains in linear arrangement, each about 61 residues in length. SCR domains exist in other non-RCA family complement proteins, including C2, C6, C7, C1r and C1s, as well the complement receptor proteins CR1 and CR2. SCR domains are also found outside the complement system in proteins such as the  $\beta$ -subunit of factor XIII (Ichinose *et al.*, 1986), haptoglobin (Kurosky *et al.*, 1980) and  $\beta$ 2 glycoprotein I ( $\beta$ 2 GP-I) (Lozier *et al.*, 1984). Within each domain there are four invariant Cys residues that form disulphide bridges in a 1-3, 2-4 pattern, and one conserved Trp residue, in addition to several other conserved hydrophobic residues



**Figure 2.1**

A schematic representation of the *HF1* gene showing its position within the RCA gene cluster. The top part of the figure shows the short and long arms (1p and 1q respectively) of human chromosome 1, with the band 1q32 highlighted. Within this band (middle part of the figure) are found all the RCA genes, as well as the gene encoding the B subunit of the coagulation Factor XIII, *F13B*. The bottom part of the figure illustrates the encoding of both FH and FHL-1 by exons 1-9 (black-and-white), whereas exon 10 (black) encodes only FHL-1, and exons 11-23 (white) encode only FH (adapted from [Rodriguez de Córdoba et al., 2004](#)).



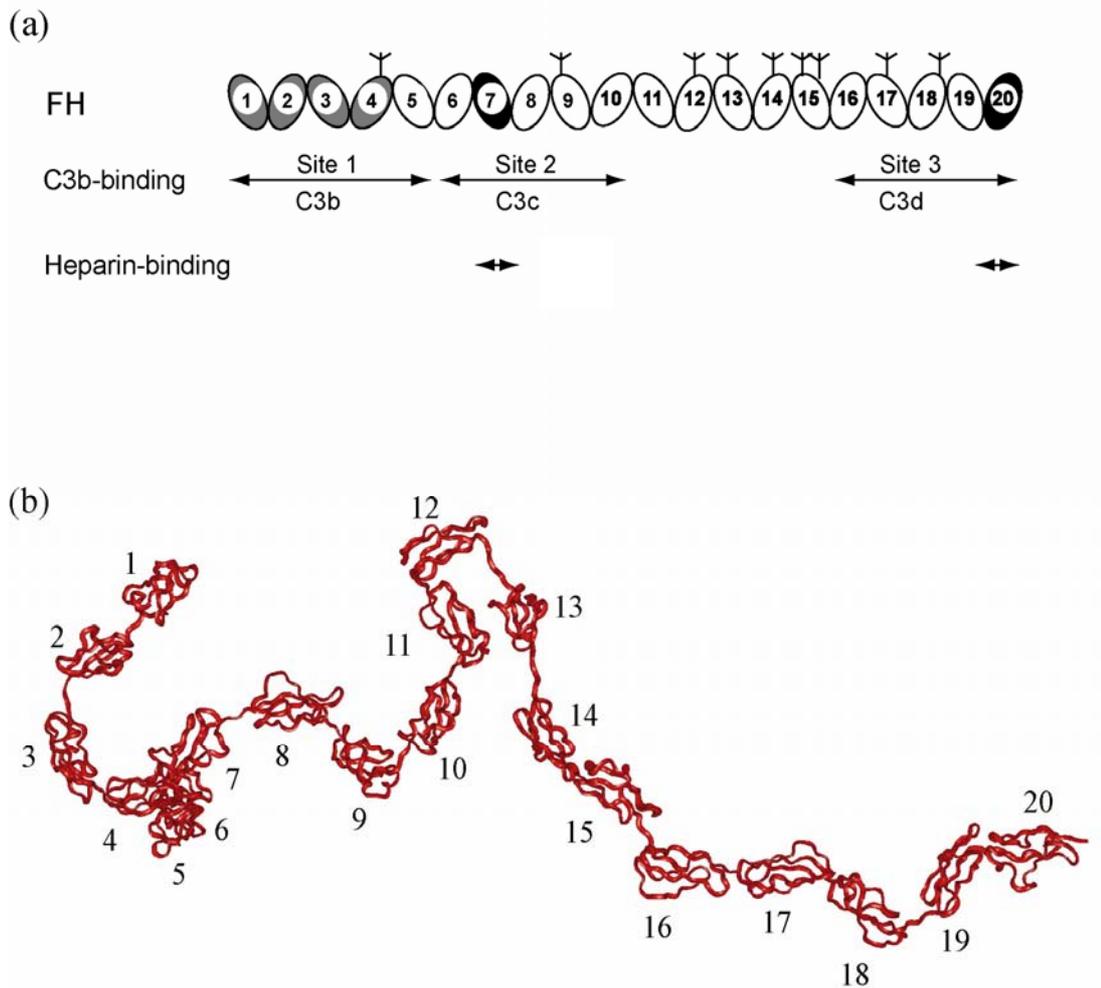
**Figure 2.2** Sequence alignment of the 20 SCR domains of human complement FH (PDB accession number P08603). Conserved cysteine and tryptophan residues are highlighted in light brown. Other conserved residues are in yellow, the putative carbohydrate sites are underlined, and the linker residues are boxed. The Define Secondary Structure Prediction (DSSP)-assigned  $\beta$ -strand secondary structure regions are highlighted in grey, and the position and labelling of the eight consensus  $\beta$ -strands b1 to b8 are shown below the alignment.

(Figure 2.2). Each SCR domain is surrounded by two antiparallel  $\beta$  strands that run approximately along its long axis. These  $\beta$  strands are in turn joined by loops and turns, including the poorly conserved hypervariable loop found between the  $\beta$ -strand secondary structure regions b2 and b3 (Figure 2.2). There are nine potential N-glycosylation sites on FH, one of which is known to be unoccupied in SCR-4 (Morley & Walport, 2000; Aslam & Perkins, 2001; Fenaille *et al.*, 2007).

### (2.2.2) Structures of Factor H SCR domains

The domains of proteins containing several SCR domains are usually arranged contiguously, like beads on a string. The average length of inter-SCR linkers is four amino acid residues, with the shortest being three and the longest eight residues. Inherent flexibility thus presents potential problems in structure determination for multi-SCR proteins such as FH (20 SCRs) and complement receptor 1 protein CR1 (30 SCRs). The structures of several SCR domains have been investigated by X-ray crystallography, nuclear magnetic resonance (NMR), electron microscopy (EM) and small angle scattering (SAS). These studies show that, in general, the orientation between two adjacent SCR domains is not easily predicted (Perkins *et al.*, 2002; Lehtinen *et al.*, 2004). Atomic level structures are often constrained to one or two domains, and the sensitivity of NMR and X-ray crystallography techniques to linker flexibility limits their usefulness in studying molecules such as FH in their native form. EM techniques, while able to directly observe the intact molecule, give only a static picture, and the conditions of study are generally harsh and non-physiological.

Thus, to date, intact FH has not been crystallised for reason of its size, glycosylation and inter-SCR flexibility. However, other methods such as solution scattering, analytical ultracentrifugation and electron microscopy have been applied to demonstrate that FH does not possess a fully-extended SCR structure in solution but is partially folded back (Perkins, 1991; Discipio, 1992; Aslam & Perkins, 2001; Chapter 5). To date therefore, the only molecular structures for intact FH are medium-resolution constrained X-ray scattering models (Aslam & Perkins, 2001; Chapter 5) based on medium-to-high-resolution structures for individual (or groups of) SCR domains. NMR structures have been determined for SCR-5 (Barlow *et al.*, 1992), SCR-15/16 (Barlow *et al.*, 1993) and more recently SCR-1/3 (Hocking *et al.*, 2008), while crystal structures have been determined for SCR-19/20 (Jokiranta *et al.*, 2006) and SCR-6/8 (Prosser *et al.*, 2007) (Table 2.1). Homology models for the



**Figure 2.3**

FH domain structure (modified from [Aslam & Perkins, 2001](#)). **(a)** Schematic view of the 20 SCR domains of FH. The positions of three C3b binding sites, with the first region also having decay acceleration and cofactor activity (grey-coloured domains) are shown. Also shown are two heparin-binding sites on SCR-7 and SCR-20 (black), and nine putative N-linked glycosylation sites (Y). **(b)** 3-D model of FH showing its flexible ball-on-string domain arrangement (PDB code 1 HAQ, [Aslam & Perkins, 2001](#)).

<b>FH domains</b>	<b>PDB entry code</b>	<b>Method of determination <sup>a</sup></b>	<b>Comments/ Reference</b>
SCR-1/2	2RLP	NMR	<a href="#">Hocking <i>et al.</i>, 2008</a>
SCR-2/3	2RLQ		
SCR-1/5	2QFG	SAXS	<a href="#">Chapter 4</a>
SCR-6/8	2IC4	SAXS	<a href="#">Fernando <i>et al.</i>, 2007</a>
SCR-6/8	2UWN	X-ray	402H variant; resolutions
	2V8E	crystallography	2.35 Å (UWN) and 2.50 Å (2V8E). <a href="#">Prosser <i>et al.</i>, 2007</a>
SCR-7	2JGW	NMR	402H variant (2JGW); 402Y
	2JGX		variant (2JGX). <a href="#">Herbert <i>et al.</i>, 2007</a>
SCR-15	1HFI	NMR	<a href="#">Barlow <i>et al.</i>, 1993</a>
SCR-15/16	1HFH	NMR	<a href="#">Barlow <i>et al.</i>, 1993</a>
SCR-16	1HCC	NMR	<a href="#">Barlow <i>et al.</i>, 1993</a>
SCR-16/20	2QFH	SAXS	<a href="#">Chapter 4</a>
SCR-19/20	2BZM	NMR	<a href="#">Herbert <i>et al.</i>, 2006</a>
SCR-19/20	2G7I	X-ray	Resolution 1.75 Å.
		crystallography	<a href="#">Jokiranta <i>et al.</i>, 2006</a>
SCR-1/20	1HAQ	SAXS	<a href="#">Aslam &amp; Perkins, 2001</a>
SCR-1/20	3GAU	SAXS	In 50 mM NaCl (3GAU);
	3GAV		in 137 mM NaCl (3GAV);
	3GAW		in 250 mM NaCl (3GAW)
			<a href="#">Chapter 5</a>

<sup>a</sup>Abbreviations: NMR, nuclear magnetic resonance; SAXS, small angle X-ray scattering.

### **Table 2.1**

Experimentally-determined FH domain structures available in the protein data bank (March 2009).

remaining SCR domains have been described (Aslam & Perkins, 2001; Saunders *et al.*, 2006).

### (2.2.3) Factor H inter-domain arrangements

Since the orientation between two adjacent SCR domains is variable and not easily predicted (Perkins *et al.*, 2002; Lehtinen *et al.*, 2004), methods that study the molecule in solution and under near-physiological conditions may yield structural information not readily available otherwise. X-ray scattering and analytical ultracentrifugation in combination with constrained scattering modelling leads to medium resolution molecular structures (Perkins *et al.*, 1998; 2005; 2008). Using these techniques, solution structures for full-length FH (Aslam & Perkins, 2001; Chapter 5), SCR-6/8 (Fernando *et al.*, 2007) and SCRs 1/5 and 16/20 (Chapter 4) have been described, revealing that FH is bent and has a tendency to dimerise at SCRs 6/8 and 16/20 (Fernando *et al.*, 2007; Chapter 4). This is consistent with early electron micrographs showing its flexibility (Discipio, 1992) and other studies showing the tendency of FH fragments to self-associate (Perkins *et al.*, 1991; Jokiranta *et al.*, 2000; 2006; Oppermann *et al.*, 2006).

The folded-back structure of FH may be important for its regulatory function (Chapter 4). However, the molecular basis of this inter-domain arrangement remains unknown. For example, intramolecular interactions between the N- and C- regions to stabilise an omega-shaped structure for FH have been proposed as a possible mechanism (Oppermann *et al.*, 2006). This proposed mechanism, based on the interaction between SCR-1/7 and SCR-1/20 as observed by surface plasmon resonance, may however be alternatively explained by the tendency of SCR-6/8 to dimerise (Fernando *et al.*, 2007). Shorter-distance intra-molecular interactions between SCR domains in close proximity, stabilised by electrostatic forces, are another possible basis for maintaining a folded-back structure in solution. If such electrostatic interactions contribute towards maintaining a stable folded back FH structure under physiological conditions, then FH conformation will be affected by ionic strength and pH. This would suggest that effect of electrostatics in FH interaction with its ligands is not limited to the residues near the binding surface, but extends to other residues important for stabilising an optimal structure for these interactions. This may also provide a partial explanation for the importance of mutations located outside the known functional regions of FH. For example, a small

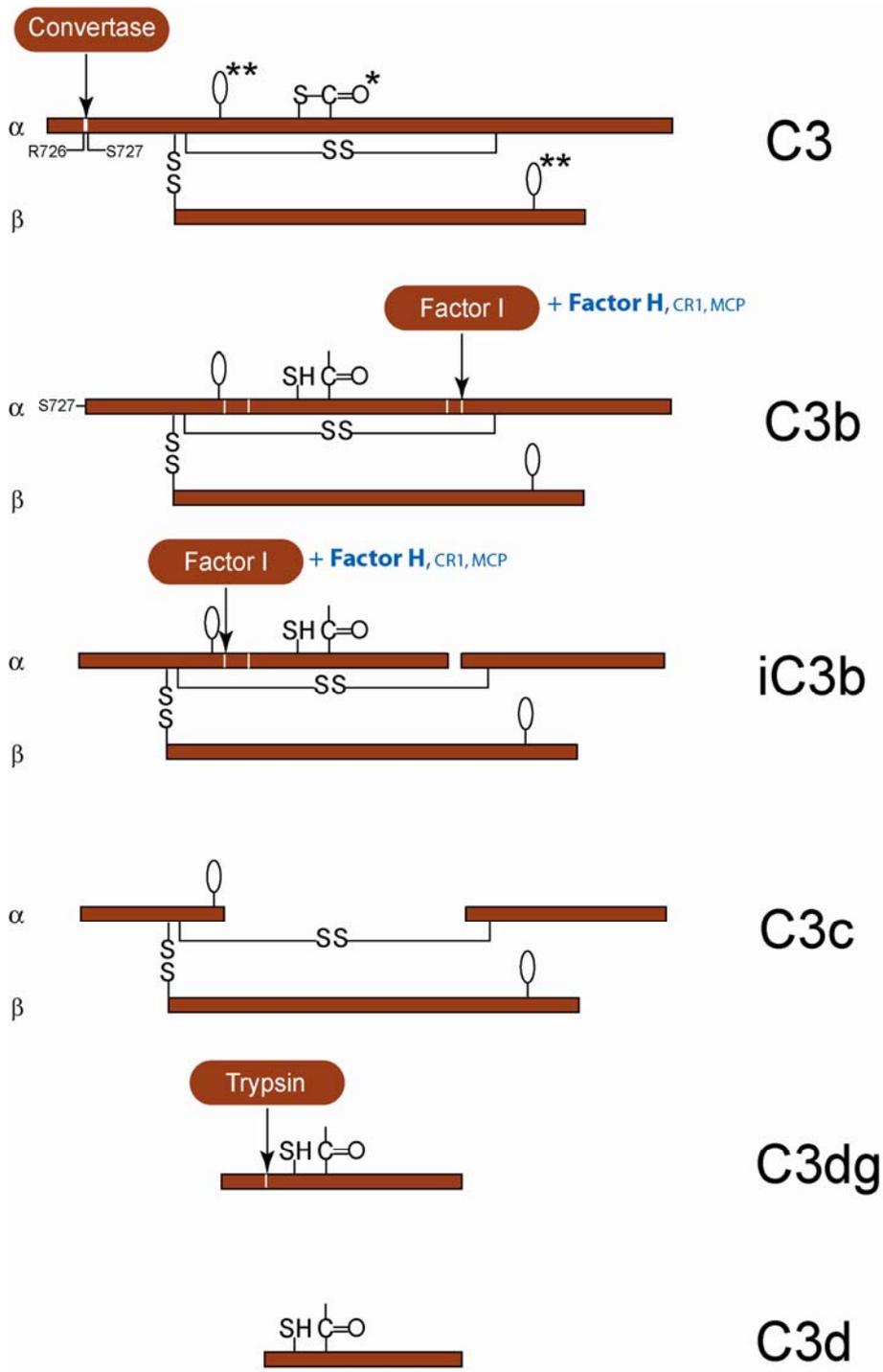
but significant number of aHUS mutations occur outside SCR-19 and SCR-20, and the reason for this is not known (Section 2.4; Saunders *et al.*, 2007).

In this thesis, the interdomain arrangements within the functionally-relevant SCR-1/5 and SCR-16/20 fragments of FH have been studied with a view to explaining the overall structure of FH (Chapter 4). Also, electrostatic interactions have been hypothesized to contribute towards maintaining a stable folded back FH structure under physiological conditions. Accordingly, the effects of ionic strength and pH on the conformational properties of native FH have been investigated by analytical ultracentrifugation. Through the use of improved X-ray scattering instrumentation and aggregate-free FH purifications, new molecular models for FH have been derived for deposition in the Protein Data Bank (Chapter 5). These new models give much improved scattering curve fits compared to the previous (Aslam & Perkins, 2001).

## **(2.3) Function of Factor H in complement**

### **(2.3.1) Regulation of complement activation**

The primary role of FH in complement is the regulation of the alternative pathway C3 convertase C3bBb and its C3b component, both in plasma and on host cell surfaces. FH binds to C3b and acts as a cofactor for the factor I-mediated proteolytic inactivation of C3b (Figure 2.4; Pangburn *et al.*, 1977), while competing with Factor B in binding to C3b to form the alternative pathway C3-convertase C3bBb (Whaley and Ruddy, 1976). It also accelerates the decay of C3bBb (Weiler *et al.*, 1976). However, while FH readily binds and inactivates C3b in the fluid-phase, the inactivation of surface-bound C3b by FH is significantly affected by the nature of the surface to which C3b is bound. On host cell surfaces, sialic acids, glycosaminoglycans or sulphated polysaccharides such as heparin are known to increase the affinity of FH for C3b (Fearon, 1978). The ability of some microorganisms to recruit FH to their surface by mimicking these surface polysaccharides is an important factor in their pathogenicity. For example, FH binds to surface sialylated lipopolysaccharides of *Neisseria gonorrhoeae* (Ram *et al.*, 1999), a membrane lipoprotein of *Borrelia burgdorferi* (Helwage *et al.*, 2001; Brooks *et al.*, 2005) and the M protein of Group A Streptococci (Jarva *et al.*, 2003). Other microbes that recruit FH to their surface include *Echinococcus granulosus*, *Onchocerca volvulus*, *N. meningitidis* and HIV (Schneider *et al.*, 2006).



**Figure 2.4**

Inactivation of C3b by FH. Sites of cleavage by the C3-converstase and factor I in the  $\alpha$ -chain are indicated by arrows. Inactivation of C3b, first to iC3b (with the release of C3f), then C3c and C3dg by factor I requires the cofactor function of FH. The positions of the C3 internal thioester (\*), glycosylation sites (\*\*), and interchain disulphide bonds (S S) are shown (modified from Rother *et al.*, 1998).

### (2.3.2) Factor H binding to C3b

FH regulates surface-bound C3b activity by recognising charge (anionic) clusters on host cell surfaces that are mimicked by heparin. FH makes the initial contact with host cells through the C-terminal SCR-20, and this is followed by N-terminal regulatory activity (Oppermann *et al.*, 2006; Ferreira *et al.*, 2006). FH has three C3b-binding sites (Figure 2.3(a)). The four N-terminal domains SCR-1/4 bind to intact C3b, a second site in SCR-6/10 binds to the C3c region of C3b, and a third site within SCR-16/20 binds to the C3d region of C3b (Lambris *et al.*, 1988; Sharma & Pangburn, 1996; Jokiranta *et al.*, 2000; 2001).

A molecular understanding of the FH-C3b complex requires the dissection of the role of each of these C3b sites on FH. In terms of the C3d fragment, ligand-binding assays in 137 mM NaCl buffer first revealed that a peptide derived from C3d inhibited FH binding to C3b (Lambris *et al.*, 1988). Surface plasmon resonance analysis in 50 mM and 75 mM NaCl indicated that C3d binds weakly to FH and the FH-related proteins FHR-3 and FH-4 with dissociation constants  $K_D$  of 0.09-0.36  $\mu\text{M}$  (Hellwage *et al.*, 2002). The interaction of FH with C3d in 50 mM and 75 mM NaCl was further confirmed, and the binding site was localised to SCR-19/20 and subsequently to SCR-20 (Jokiranta *et al.*, 2000; Hellwage *et al.*, 1999). Mutations in SCR-20 lead to reduced binding of C3d and heparin (Manuelian *et al.*, 2003; Jokiranta *et al.*, 2005). The binding sites for C3d and heparin partially overlap and lead to a ternary complex of FH-C3d-heparin in 70 mM NaCl (Hellwage *et al.*, 2002; Jokiranta *et al.*, 2005). Other related studies showed that dimeric C3dg remains bound to IgG after C3b inactivation, suggesting that dimeric C3d is required for efficient binding to CR2 and that C3d oligomers may be physiologically important (Arnaout *et al.* 1983; Dempsey *et al.*, 1996; Jelezarova *et al.*, 2003).

In this thesis, insight into FH interaction with C3b has been obtained by studying its complex with C3d (Chapter 6). The effect of ionic strength on these interactions has been investigated in an effort to obtain new insight into complement regulation and host protection, as well as a clearer understanding of the molecular mechanism involved in some of the disease conditions in which FH mutations are implicated.

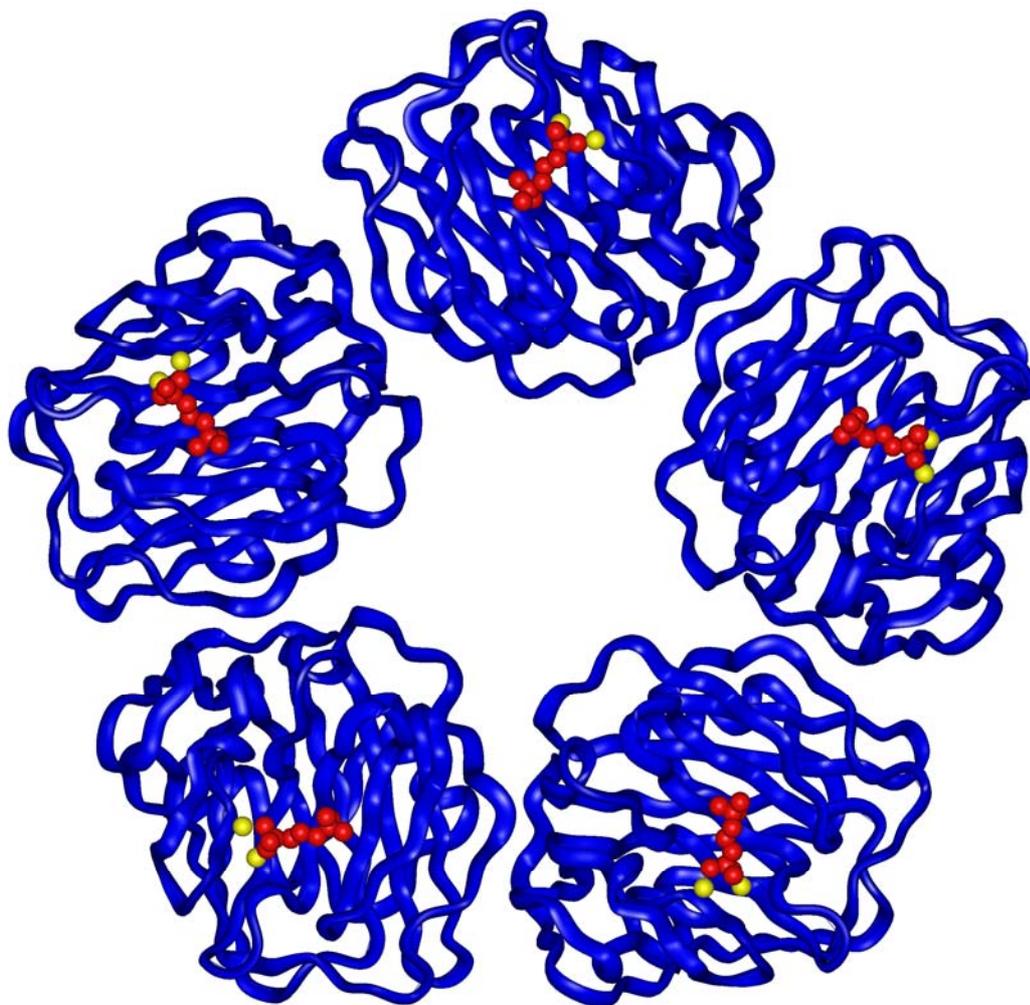
### **(2.3.3) Factor H binding to CRP**

#### **(2.3.3.1) Introduction to human CRP**

C-reactive protein (CRP) is an acute-phase protein of the pentraxin family of  $\text{Ca}^{2+}$ -dependent ligand-binding proteins found in serum. The other member of this family in humans is serum amyloid P component (SAP). CRP was originally discovered in 1930 during studies of patients with *Streptococcus pneumoniae* infection as a substance in serum that reacted with the C-polysaccharide of *Pneumococcus* (Tillett & Francis, 1930), hence its nomenclature. CRP occurs in serum at a median concentration of 0.8 mg/L, up to a maximum of around 10 mg/L (Pepys & Hirschfield, 2003). This concentration may rise up to 10,000 times within six hours of acute inflammation events such as infection, peaking after 48 hours. With a constant half-life of 19 hours, its serum level is mainly determined by the rate of production which is in turn influenced by the severity of infection. This feature of CRP is exploited diagnostically in its use as a marker of inflammation. The gene encoding human CRP is on chromosome 1 (1q21-1q23).

CRP transcription occurs only in hepatocytes and is regulated predominantly by the cytokine interleukine-6. On translation, CRP comprises 224 amino acids, of which residues 1-18 are the signal peptide with 19-224 making up the mature protein. Native CRP is unglycosylated and consists of 5 identical subunits arranged non-covalently as a cyclic pentamer (Figure 2.5), with an overall molecular weight of 115 kDa. Each subunit or protomer has 206 amino acids and a molecular weight of 23 kDa. CRP protomers possess a characteristic “lectin fold” composed of a two-layered  $\beta$ -sheet with flattened jellyroll topology and one concave side (Shrive *et al.*, 1996; Thompson *et al.*, 1999). The  $\beta$ -sheets are linked by loops on the periphery (Figure 2.5). A single  $\alpha$ -helix in the CRP protomer is located on the “A” face opposite the  $\text{Ca}^{2+}$ -binding site and spans amino acid residues 168-176 (Shrive *et al.*, 1996). The concave side, or “B” face, contains the phosphocholine-binding site, with two  $\text{Ca}^{2+}$  bound 4 Å apart by protein side chains (Figure 2.5; Pepys & Hirschfield, 2003).

CRP binds phosphocholine (Volanakis & Kaplan, 1971), as well as phosphoethanolamine, microbial surface proteins, chromatin, histones, fibronectin, small nuclear ribonucleoproteins, laminin and polycations (Black *et al.*, 2004). Ligand-bound or aggregated CRP activates the classical pathway of complement by



**Figure 2.5**

Crystal structure of native CRP pentamer complexed with phosphocholine (Thompson *et al.*, 1999). INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA) was used to generate a ribbon diagram from PDB entry 1B09. The Ca<sup>2+</sup> are yellow and phosphocholine is red. The cyclic pentameric arrangement of its protomers is shown. This structure is viewed with the B face uppermost.

interacting with C1q. Also, CRP interacts with immunoglobulin receptors FcγRI and FcγRII to stimulate phagocytic cells. CRP has been described as producing both pro- and anti-inflammatory effects (Black *et al.*, 2004). Some of its pro-inflammatory effects include up-regulation of adhesion molecule expression in endothelial cells, inhibition of endothelial nitric oxide synthase expression in aortic endothelial cells, stimulation of interleukin-8 release from several cell types and increase of tumour necrosis factor-α release. Examples of its anti-inflammatory effects include the induction of interleukin-1 receptor antagonist expression, increase of the release of the anti-inflammatory interleukin-10 and the repression of interferon-γ synthesis. This apparent contradiction in CRP activity results from the study of CRP in either its native form or as abnormally modified forms (mCRP), with different biological and antigenic properties (Potempa *et al.*, 1983a; 1983b; 1987; Khreiss *et al.*, 2002). mCRP consists of dissociated CRP protomers which are produced *in vitro* under harsh conditions of high acidic pH (2.0), temperature (above 60°C) or 8 M urea. It has been suggested that mCRP may occur naturally in normal or inflamed human and animal tissues (Samols *et al.*, 1985; Rees *et al.*, 1988; Egenhofer *et al.*, 1993; Diehl *et al.*, 2000), and that Ca<sup>2+</sup> depletion of native CRP may lead to spontaneous formation of mCRP (Motie *et al.*, 1996; Wu *et al.*, 2002). mCRP or CRP aggregation or multimerisation is suggested to be important for CRP function at cell surfaces (Jiang *et al.*, 1991; Taylor & van den Berg, 2006). However, once CRP is denatured, it is rapidly catabolised, and there is no firm evidence for the persistence of monomeric, denatured or aggregated CRP *in vivo* (Pepys & Hirschfield, 2003).

In this thesis, a multidisciplinary strategy using analytical ultracentrifugation, X-ray scattering, constrained molecular modelling and surface plasmon resonance was applied to resolve the solution properties of CRP (Chapter 7) to provide insights into the multivalent interaction of CRP with a range of macromolecular ligands.

#### **(2.3.3.2) Factor H interaction with CRP as a means of complement regulation**

CRP activates complement via the classical pathway (Kaplan and Volanakis, 1974), but inhibits activation by the alternative pathway. Evidence of this was found on examining the presence of individual components when complement was activated by CRP. Activation was found to be efficient only at the initial stages involving C1-C4, with little activation of the late proteins C5-C9 (Mold *et al.*, 1999). The net effect of this is the ability of CRP to participate in host defence systems

while limiting the potentially damaging effects of the late stage complement components by acting as a regulator of complement activation. CRP up-regulation of the expression of the complement inhibitory factors decay-accelerating factor (DAF), membrane co-factor protein (MCP) and CD59 has been suggested as one mechanism for this effect (Li *et al.*, 2004). However, evidence for the recruitment of FH by CRP suggests that the inhibitory effect of CRP on alternative pathway activation may be mediated mainly by its interaction with FH. Direct evidence for this was first provided by an ELISA assay of plastic-immobilised CRP binding of FH (Mold *et al.*, 1999). However, in the same study, fluid-phase CRP only bound to immobilised FH with less than 1% efficiency.

### **(2.3.3.3) The role of the interaction between Factor H and CRP in the molecular mechanism of age-related macular degeneration**

In an early surface plasmon resonance study, CRP was immobilised on the sensor chip surface by amine coupling in the absence of  $\text{Ca}^{2+}$ . SCR-7 and SCR-8/11 of FH were found to bind to CRP in a  $\text{Ca}^{2+}$ -dependent manner (Jarva *et al.*, 1999). Reverse binding of CRP to immobilised FH in the presence of  $\text{Ca}^{2+}$  was also reported in the same study. Binding of FH to immobilised CRP has since then been demonstrated repeatedly, with differences in binding affinity between the H402 and Y402 variants of FH (Laine *et al.* 2007; Sjoberg *et al.*, 2007; Skerka *et al.*, 2007; Yu *et al.*, 2007; Herbert *et al.*, 2007). Because the Y402H polymorphism within SCR-7 of FH has been implicated as a risk factor in age-related macular degeneration, AMD (Section 2.4), this interaction has been suggested to be important in the molecular mechanism of AMD. However, a recent study using size exclusion chromatography and analytical ultracentrifugation (sedimentation equilibrium) has suggested that the observed interactions are a result of CRP denaturation resulting from its immobilisation in the absence of  $\text{Ca}^{2+}$ , and that FH does not interact with native pentameric CRP (Hakobyan *et al.*, 2008).

In this thesis, the interaction of FH with CRP was investigated in detail by analytical ultracentrifugation, surface plasmon resonance and X-ray scattering (Chapter 8). The discrepancy between these studies was re-examined in physiological buffer conditions and protein concentrations, being careful to work in conditions when CRP retains its native pentameric arrangement. To ensure that CRP remained in native undenatured form during surface plasmon resonance studies, CRP

immobilisation was carried out in the presence of  $\text{Ca}^{2+}$ , and reverse binding experiments were used to verify the results. Together with improved knowledge of the solution properties of native CRP (Chapter 7), this study will help clarify the role of FH in the molecular mechanism of AMD and aHUS diseases.

#### **(2.3.4) Factor H binding to other ligands**

Two separate binding sites for heparin and cell surface polyanions (Figure 2(a)) have been identified in SCR-7 (Blackmore *et al.*, 1996; Schmidt *et al.*, 2008) and SCR-20 (Blackmore *et al.*, 1998; Schmidt *et al.*, 2008) of FH, with the possibility of a third site in SCR-9 (Ormsby *et al.*, 2006) or SCR-13 (Sharma & Pangburn, 1996). FH can also function as an adhesion ligand enhancing the activation response of neutrophils. The receptor mediating this interaction has been identified as the integrin MAC-1 (CD11b/CD18), which is also a receptor for iC3b (CR3) and ICAM-1 (DiScipio *et al.*, 1998).

#### **(2.4) Factor H mutations and disease**

Over 130 known mutations in FH are associated with disease (Saunders *et al.*, 2007; [www.fh-hus.org](http://www.fh-hus.org)). The majority of these mutations (85 of 134, more than 60%) occur at the C-terminus within SCR-16/20, 25 occur within SCR-11/15 and 24 in the region between SCR-1/10 ([www.fh-hus.org](http://www.fh-hus.org)). The N-terminal SCR-1/4 region of FH is essential for the cofactor activity of FH (Gordon *et al.*, 1995; Sharma & Pangburn, 1996). SCR-7 and SCR-8/11 are important for binding of heparin and CRP, while the C3b/polyanion-binding site located within SCR-16/20 is critical for cell binding and recognition (Ferreira *et al.*, 2006, Opermann *et al.*, 2006). These functional regions participate in the complement-regulatory activity of FH at host cell surfaces. FH ligand-binding sites may possess specificity for multiple ligands. For example, SCR-7 binds to both heparin and CRP (Blackmore *et al.*, 1996; Jarva *et al.*, 1999; Giannakis *et al.*, 2003; Schmidt *et al.*, 2008), the C3b-binding site in SCR-6/10 (Lambris *et al.*, 1988; Sharma & Pangburn, 1996; Jokiranta *et al.*, 2000; 2001) overlaps with this and another CRP-binding site in SCR-8/11 (Jarva *et al.*, 1999), while SCR-20 binds both C3d and heparin (Hellwage *et al.*, 2002).

Defects in plasma proteins are generally classified as Type I or Type II. Type I is defined as when the protein plasma concentration is reduced and indicates a secretory defect or rapid degradation. Type II defects show normal plasma

concentrations but reduced protein function. Mutations within FH can be classified as Type I or Type II according to the same definitions (Perkins & Goodship, 2002). Of the 134 currently known mutations in FH, 27 are classified as Type I and 52 as Type II, while the rest have an undescribed phenotype ([www.fh-hus.org](http://www.fh-hus.org)). Type I mutations in FH are associated mainly with membranoproliferative glomerulonephritis (MPGN), while Type II mutations are principally associated with haemolytic uraemic syndrome (HUS). A single Tyr402His polymorphism within FH has been strongly linked to the development of age-related macular degeneration (AMD) (Klein *et al.*, 2005; Haines *et al.*, 2005; Edwards *et al.*, 2005; Hageman *et al.*, 2005; Saunders *et al.*, 2007).

Type I FH mutations have been shown to strongly correlate with development of MPGN. MPGN is a rare kidney disease characterized by complement deposition within the basement membrane of the glomerular capillary wall. Associated phenotypes of this condition include progressive glomerulonephritis, mesangial cell proliferation and an increase in mesangial matrix size (Saunders *et al.*, 2006). Type II mutations at the C-terminus of FH are mainly associated with atypical HUS (aHUS). HUS is a clinical syndrome that manifests with thrombocytopenia, microangiopathic haemolytic anaemia and progressive renal failure. HUS may be classified as diarrhoeal (D+ or typical HUS) which occurs in association with enterohaemorrhagic *Escherichia coli*, or non-diarrhoeal (D- or atypical HUS). aHUS is a disease of complement dysregulation caused by genetically determined deficiencies in the activity of FH and other complement regulatory proteins including MCP and factor I (Section 1.3; Amirlak & Amirlak, 2006; Kavanagh *et al.*, 2006). Mutations in FH account for up to 30% of all reported cases of aHUS (Kavanagh *et al.*, 2006). The primary event for aHUS is endothelial cell activation in which a procoagulant phenotype is maintained above a noncoagulant phenotype. Damage of the endothelial surface attracts platelets to the site of damage, and where this is enhanced by unregulated complement activation in small vessels, blockage of the small vessel occurs due to the presence of large numbers of platelets at the site (thrombotic microangiopathy). This disease has a poor prognosis and may lead to irreversible renal damage and lifetime dependence on kidney dialysis (Kavanagh *et al.*, 2006).

The Tyr402His polymorphism in FH is strongly associated with AMD, the most common cause of blindness in the Western world (Klein *et al.*, 2005; Haines *et al.*, 2005; Edwards *et al.*, 2005; Hageman *et al.*, 2005; Saunders *et al.*, 2007). One of

the hallmarks of AMD is the appearance of drusen, an amyloid plaque-like deposition in the Bruch's membrane. Drusen contains around 120 oxidised lipids and aggregated proteins, including FH. The Bruch's membrane is a layer interposed between the retinal pigment epithelium and the choroidal vasculature (Bird, 1992; Bird *et al.*, 1995). Both unregulated complement activation at the retina and uncontrolled self-association of FH are thought to be relevant to drusen formation (Nan *et al.*, 2008a).

## **Chapter Three**

### **Protein structure and interaction measurement**

### (3.1) Introduction to protein structure determination

Protein structure determination is often necessary to understand the molecular basis of protein function. Given an amino acid sequence, it is possible to predict the secondary structure of a protein. However, experimental determinations of three-dimensional tertiary structures provide more complete structural information on interactions between amino acid residues not close together in the primary or secondary structure. Protein amino acid sequences can be determined directly by mass spectrometry or using Edman degradation, Sanger's reagent or Dansyl chloride methods. They can also be determined indirectly by sequencing the protein gene or cDNA. Protein secondary structure can be determined by low-resolution methods that analyse the characteristic  $\beta$ -sheet and  $\alpha$ -helix contents of the protein, such as circular dichroism and Fourier transform infrared spectroscopy. Nuclear magnetic resonance (NMR) spectroscopy can also be used to determine protein secondary structure.

For protein tertiary structure determination, there are a variety of low ( $\sim 2$ -4 nm) and high (atomic coordinate level) resolution techniques. Low resolution techniques include analytical ultracentrifugation (AUC) (Section 3.2), small angle X-ray and neutron scattering (SAXS and SANS) (Section 3.3) and electron microscopy (EM), while high resolution techniques include NMR and X-ray crystallography. AUC extracts hydrodynamic and structural information from the sedimentation behaviour of macromolecules under high centrifugal forces. Small angle X-ray or neutron scattering studies the diffraction of X-rays or neutrons by macromolecules in solution. The resolution of SAXS and SANS can be improved to medium levels (0.5-1.0 nm) if combined with biomolecular modelling (Section 3.4).

EM is a low resolution method that has the same operating principle as light microscopy, except that it uses an electron beam and electrostatic and electromagnetic lenses to visualise protein molecular shapes. The greater resolution and magnification of an electron microscope compared to a light microscope is due to the wavelength of an electron being much smaller than that of light. However, not only can the high level of electron radiation potentially damage a protein sample, the relatively harsh conditions of sample preparation can damage or distort the protein structure. For example protein fixation and staining, as well as dehydration and *in vacuo* viewing, are some of the processes that may be involved in sample preparation and analysis. In a typical solution, protein molecules may exist in several

different conformations rather than a single, ordered orientation. Hence structural information obtained under such harsh conditions may not fully represent the protein's native conformation. This is particularly true for flexible molecules with interdomain linkers such as FH. Furthermore, a low signal-to-noise ratio means that structural information from several macromolecules must be averaged, presenting difficulties in analysing highly flexible structures.

High resolution protein structure determination by NMR utilises the magnetic properties of certain nuclei, most notably  $^1\text{H}$  and  $^{13}\text{C}$ . Magnetic nuclei within protein samples are aligned with a constant magnetic field at different frequencies and then exposed to another magnetic field in a perpendicular direction. The perturbation of this alignment results in a characteristic absorption energy which differs according to the chemical environment of different nuclei within the same molecule. This dependence of nuclear magnetic energy levels on the local electronic environment is known as chemical shift. The chemical shift is reported as a relative value measured against a standard reference signal. In two-dimensional NMR (2D-NMR) used for structure determinations, multinuclear NMR and different types of experiments allow chemical shifts to be assigned to specific nuclei. Nuclear Overhauser effects ([Overhauser, 1953](#)) permit the specification of spatial relationships between atoms, thus allowing three-dimensional models to be created for the protein structure ([Wuthrich, 1986](#)). NMR studies proteins in solution, but an upper molecular weight constraint of about 35 kDa for most traditional techniques, beyond which the signal resolution deteriorates rapidly, means that it is unsuitable for studying relatively large proteins in their intact forms.

In X-ray crystallography, positions of individual atoms are determined by interpreting the diffraction of X-rays by protein crystals. The ease of obtaining these crystals is affected by the conformational properties of proteins as well as the presence of post-translational modifications such as glycosylation. It is difficult to grow FH crystals for both reasons. Provided the protein crystals can be grown, this method can yield high resolution structures (0.1-0.3 nm). When these crystals are exposed to a monochromatic, collimated X-ray beam, electrons within the protein scatter the beam, and the resulting diffraction pattern depends on the properties of the scattered waves. Thus only waves that are in phase and exhibit constructive interference will produce a diffraction spot ([Rhodes, 2000](#)). A fundamental limitation in interpreting the diffraction pattern is the loss of phase. This occurs

because the detector measures only the intensity of the diffracted beam, whereas both intensity (amplitude) and phase are needed to determine the electron density within the protein. Calculation of the phase of a diffracted beam can be done by several methods. Two examples of such methods are molecular replacement (if a homologous protein is known) and multiple isomorphous replacement (MIR) using heavy atoms (e.g. Selenomethionine) to replace a number of atoms in the protein, thereby altering the diffraction properties of the crystal. When molecular replacement is done, the resulting replacement molecular model can be fitted to the obtained electron density map. When MIR is done, the strong scattering of the replacement heavy atoms allows the determination of the phase. Electron density mapping of the protein atomic coordinates is done by trial-and-error, and further refinement of the structure is carried out using goodness-of-fit  $R$  factor and  $R$  free values.

### **(3.2) Analytical ultracentrifugation**

#### **(3.2.1) Introduction to the analytical ultracentrifuge**

The analytical ultracentrifuge (AUC)<sup>†</sup> is different from preparative ultracentrifuges. It combines the features of a preparative ultracentrifuge with a sensitive optical detection system that allows the direct measurement of sample concentrations within the cells during sedimentation. This makes possible the monitoring of sedimentation events in real time, and a combination of these features enables the use of AUC to characterise core biophysical properties of molecules, rather than as a purification or preparative step for further analysis (Cole & Hansen, 1999). The most popular analytical ultracentrifuges are the Optima™ or Proteomelab™ XL-A and XL-I systems, both made by Beckman Coulter Instruments. The XL-A uses UV and visible absorbance optics for absorbance measurements, while the XL-I uses integrated absorbance and interference optics for both absorbance and interference measurements (McRorie & Voekler, 1993). More recently, fluorescence optics have been developed for integration into the XL-I optical system (Coleman *et al.*, 2008).

---

<sup>†</sup> The acronym AUC is generally used in reference to both ‘analytical ultracentrifuge’ and ‘analytical ultracentrifugation’.

### (3.2.2) Uses of analytical ultracentrifugation

AUC can be used to assess sample purity, size and shape distributions, polydispersity and conformational changes. It can also be used to determine average sample molecular weights and kinetic constants for associating systems. Sedimentation velocity measurements can provide information on sample heterogeneity and polydispersity (Section 3.2.5.1). Sedimentation and diffusion coefficients obtained from these measurements can be used to analyse the size and shape of molecules and complexes. When this information is combined with average molecular weights, it is possible to detect the presence of relatively small quantities of contaminants. Sedimentation coefficients, in particular, can be used to track changes in the conformation of proteins.

Molecular weights ranging from several hundreds (for example, sucrose) to several millions (for example, virus particles and organelles) can be measured by sedimentation equilibrium (Section 3.2.5.2). Equilibrium constants can be derived by studying the changes in molecular weight when molecules associate to form a more complex structure. Compared to electrophoresis techniques which can provide information on the components of a complex and measure their relative stoichiometry in the complex, sedimentation equilibrium measurements by AUC can also be used to measure the molecular weight of the complex as it exists in solution, independent of its shape.

### (3.2.3) Theory of sedimentation

A particle suspended in a solvent and subjected to a gravitational force is acted upon by three forces (Figure 3.1). The sedimenting force,  $F_s$ , acts in the direction of acceleration of the particle and is proportional to the velocity of angular rotation,  $\omega$ , the radius of rotation,  $r$ , and the mass (in grams, g),  $m$ .

$$(Eq. 3.1) \quad F_s = \omega^2 r m$$

In general, the mass of a material can be determined from the equation:

$$(Eq. 3.2) \quad m = \frac{M}{n}$$

where  $M$  is the molar concentration of the material in g/mol, and  $n$  is the number of moles of the material.

The buoyant force,  $F_b$ , acts in an opposing direction to the sedimenting force. Derived from Archimedes' principle, it is equivalent to the weight of fluid displaced by the particle in solution.

$$(Eq. 3.3) \quad F_b = -\omega^2 r m_o$$

The mass of fluid displaced by the particle is denoted by  $m_o$ , and given by the equation:

$$(Eq. 3.4) \quad m_o = m\bar{v}\rho = \frac{M}{N}\bar{v}\rho$$

where  $\bar{v}$  is the partial specific volume of the particle,  $\rho$  is the density of the solvent and  $N$  is Avogadro's constant ( $6.023 \times 10^{23}$ ). The partial specific volume is the volume in millilitres (ml) that each gram of the solute occupies in solution and is the inverse of its density. The density and viscosity of the solvent can be calculated from its composition using tabulated data, or measured directly using a density or viscosity meter.

Provided that the density of the solute is greater than that of the solvent, the particle begins to move towards the bottom of the cell in a radial direction when gravitational force is applied. As it moves away from the meniscus, its velocity increases due to increasing acceleration ( $a = \omega^2 r$ ; acceleration is proportional to radial distance). A particle moving through a fluid experiences a frictional force  $F_f$  proportional to its velocity, and acting in an opposite direction to the sedimentation force,  $F_s$  (Figure 3.1). As the particle moves further along the radius,  $F_f$  increases proportionally with its velocity.

$$(Eq. 3.5) \quad F_f = -fu$$

$f$  is the frictional coefficient and is dependent on the shape and size of the particle. Bulky, elongated or rough particles will experience a greater frictional force than small, compact or smooth ones. The three forces come into equilibrium within a very short time (usually less than  $10^{-6}$  seconds). From Figure 3.1,

$$(Eq. 3.6) \quad F_s + F_f + F_b = 0$$

Substituting values from Equations 3.1, 3.3 and 3.5,

$$(Eq. 3.7) \quad \omega^2 r m - f u - \omega^2 r m \bar{v} \rho = 0.$$

Rearranging,

$$(Eq. 3.8) \quad \omega^2 r m (1 - \bar{v} \rho) = f u.$$

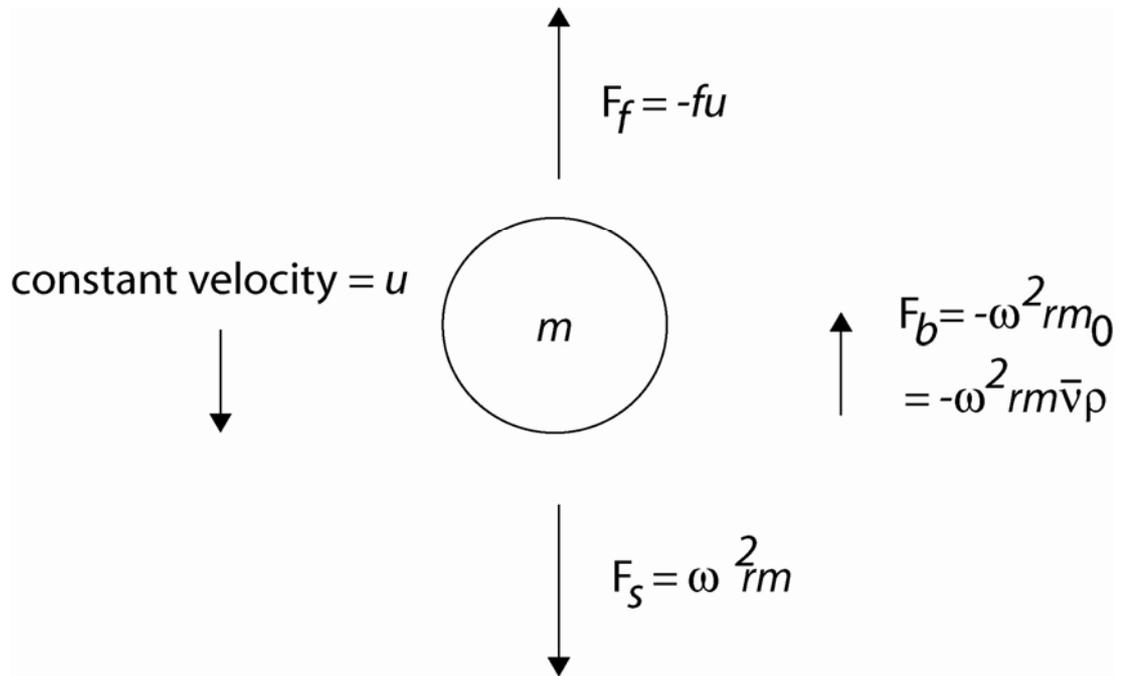
Dividing both sides of Equation 3.8 by the common factor and substituting  $m$ ,

$$(Eq. 3.9) \quad \frac{M(1 - \bar{v} \rho)}{Nf} = \frac{u}{\omega^2 r} \equiv s$$

The quantity  $u/\omega^2 r$  is called the sedimentation coefficient ( $s$ ) and is directly proportional to the buoyant weight of the particle  $M(1 - \bar{v} \rho)$ , and inversely proportional to the frictional coefficient ( $f$ ). It has a unit of seconds (s), but is conventionally expressed in terms of Svedberg units (S) in honour of Thé Svedberg, the original inventor of the AUC. Most sedimentation coefficient values lie between  $1 \times 10^{-13}$  s and  $100 \times 10^{-13}$  s, and one Svedberg unit is equivalent to  $10^{-13}$  s. Therefore a value of 4.5 S would translate to  $4.5 \times 10^{-13}$  s. Equation 3.9 can be alternatively expressed in standard form as the Svedberg equation:

$$(Eq. 3.10) \quad \frac{M(1 - \bar{v} \rho)}{RT} = \frac{s}{D},$$

where  $R$  is the gas constant and  $T$  is the absolute temperature in Kelvin. The sedimentation coefficient of a particle may therefore be defined as the velocity of the particle per unit gravitational acceleration. It is dependent on the following properties of the particle: molecular weight, frictional coefficient (a function of its shape and size), partial specific volume (volume per gram of particle) and density of the solvent. In practice, experimentally-observed  $s$  values are often converted to standard conditions (water at 20°C) by correcting for the buffer and temperature conditions (Balbo & Schuck, 2005). This standard  $s$  value is expressed as  $s_{20,w}^0$ .



**Figure 3.1**

The forces acting on a solute particle in a gravitational field.  $F_f$  and  $F_b$  act in an upward direction, while  $F_s$  and  $u$  act downwards (Ralston, 1993).

### **(3.2.4) Instrumentation**

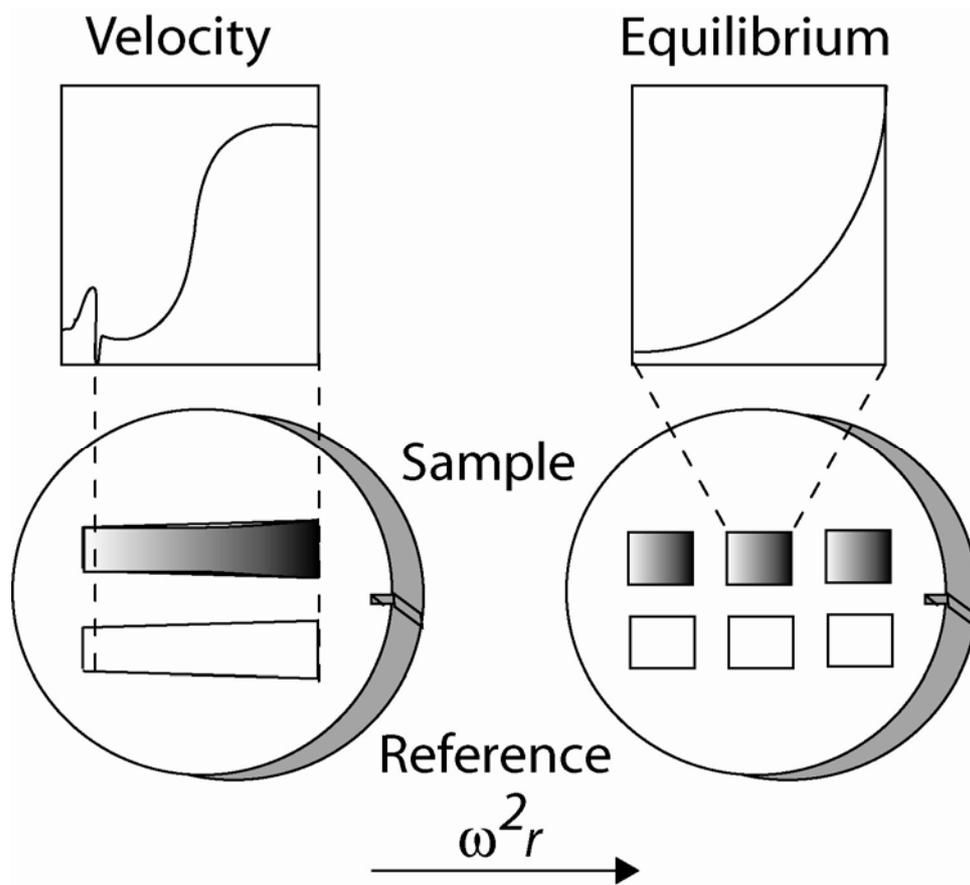
In order to achieve rapid sedimentation and minimize diffusion, the AUC must spin samples at relatively high speeds up to 60,000 r.p.m. To minimize friction and eliminate aerodynamic turbulence, the rotors are spun in an evacuated chamber. It is also important to avoid wobbling or precession (change in direction of axis of rotation). This can lead to stirring of the cell contents and, particularly during studies with low solute concentration and concentration gradient, can lead to uncertainty in the concentration distribution.

#### **(3.2.4.1) The rotor**

At speeds of 60,000 r.p.m., a typical ultracentrifuge rotor generates a gravitational force of about 250,000  $\times g$  in the cell. Under these conditions of enormous gravitational force, a mass of 1g would weigh 250,000 times its weight, 250 kg. It is therefore necessary that the rotor be capable of withstanding great gravitational stress. AUC rotors must also be able to allow the passage of light through the cells. These rotors contain either four or eight holes and are capable of holding three or seven sample cells respectively, together with a reference cell used for radial calibration. The multiple-cell system of operation makes possible the study of more than one sample in a single experiment.

#### **(3.2.4.2) The cells**

An AUC cell consists of sample cavities sealed off by two thick windows made of optical-grade sapphire or quartz ([Figure 3.2](#)) and assembled within a cell housing. The material containing the cavities (called sectors or channels) is known as a centrepiece. The centrepiece is usually made of aluminium-filled or charcoal-filled epon for experiments at speeds up to 42,000 r.p.m., or aluminium for higher speeds. The use of specialised materials ensures that the cell is able withstand high gravitational forces, does not leak or change shape, and allows the passage of light. AUC cells contain two sectors for sedimentation velocity experiments and two, six or eight channels for equilibrium experiments. For sedimentation velocity experiments using absorbance or interference optics, the column height is 12 mm and the cell is sector-shaped because of the radial direction of particle movement during sedimentation. If the cell walls are parallel, sedimenting particles will collide with the cell walls and cause convective disturbances in the solution. If the cell walls diverge



**Figure 3.2**

Sedimentation velocity and equilibrium cell designs. The left panel shows a two-sector cell during a sedimentation velocity experiment. The sample is loaded into the upper sector, and the reference solution (buffer) is loaded into the bottom sector. The sample is centrifuged at high rotational velocity, generating a boundary that moves toward the bottom the cell. The right panel illustrates a six-channel sedimentation equilibrium cell. Three sample-reference pairs are loaded into the cell, which is centrifuged at moderate rotational velocity, resulting in equilibrium concentration gradients in each sample channel (Adapted from [Cole & Hansen, 1999](#)).

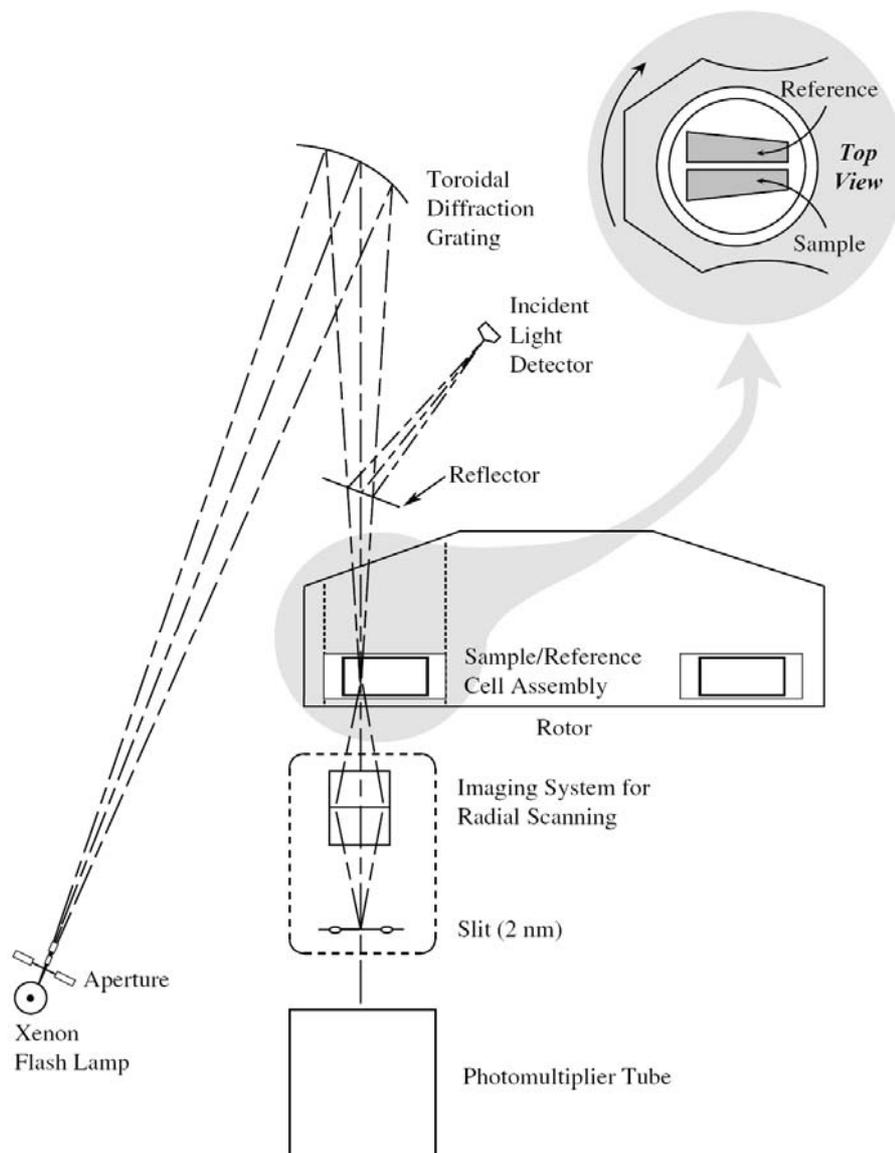
more or less than the radius of sedimentation, disturbances will also occur. In contrast, the sedimentation equilibrium cell is shaped like a regular cuboid and is usually 3mm long with six channels. The use of shorter cells with eight channels enables the quicker attainment of equilibrium and requires less sample volume, but does not provide as much information per data set (Cole & Hansen, 1999).

#### **(3.2.4.3) Absorbance optics**

The absorbance optical system of the AUC makes use of a high intensity xenon flash lamp and a scanning monochromator that is sensitive enough for measurements at wavelengths ranging between 190 and 800 nm (Figure 3.3). The xenon lamp ‘flashes’ briefly as the selected sector passes the detector. Timing information is provided with the aid of a reference magnet located at the base of the rotor. The xenon flash lamp moves radially along the length of the sector, allowing different positions of the solute to be scanned. A slit below the sample moves to allow the light through to the cells. This way, the AUC can perform wavelength scans at a fixed position in the cell, as well as radial scans at a fixed wavelength. Absorbance measurements produce good results for molecular species containing strong chromophores (for example, proteins at 230 nm and nucleic acids at 260 nm), even at relatively low concentrations. In addition, data collection at multiple wavelengths can be carried out within the same experiment for species with different absorption properties.

#### **(3.2.4.4) Interference optics**

Interference optical measurements are based on the refraction of light as it moves from one medium to another. When two beams of light from the same source (and therefore of the same wavelength and phase) pass through different media, they are refracted to different extents. Superposition of the emergent beams of light from the two media will produce a pattern called an interference fringe. The AUC uses the Rayleigh interference optical system. In this system, two beams of light are passed through the sample and reference sectors. Differences in refractive index occur because of different solute concentrations in the two sectors, and an interference fringe is produced when the two emergent beams are superposed. During sedimentation experiments, a displacement of the interference fringe pattern occurs due to changes in solute concentration at the boundary region.



**Figure 3.3**

Schematic diagram of the optical system of the Beckman Optima™ XL-A AUC (Taken from [Ralston, 1993](#)).

Because virtually all materials refract light, interference optics can be used to analyse samples that do not contain natural chromophores. Also, the high sensitivity of interference measurements means they can be used to study solutions which are too dilute for absorbance measurements. This method is useful for samples of high molecular weight or complex associating systems because scans can be collected more rapidly and frequently than for absorbance optics. It can also be used to analyse associating systems since fringe displacement is proportional to sample concentration over a wide range. It is particularly useful for measurements of samples containing strongly absorbing buffer components, as well as for measurements at high sample concentrations (at which the absorbance optics are saturated).

### **(3.2.5) Applications of analytical ultracentrifugation**

The AUC can be used to perform two basic types of experiments: sedimentation velocity and sedimentation equilibrium.

#### **(3.2.5.1) Sedimentation velocity**

In a sedimentation velocity experiment, a sufficiently high angular velocity is applied to a solute suspended in solvent. Initially there is uniform concentration of solute throughout the solution. As the velocity increases, there is relatively rapid movement of solute away from the meniscus toward the bottom of the cell. The movement of solute particles causes a reduction in its concentration near the meniscus (Figure 3.4 (a)). With time, a sharp boundary is formed between the depleted region and the uniform concentration of sedimenting solute. Although movement of individual particles cannot be analysed, the change in concentration at the boundary region causes a gradual shift of the concentration boundary. Measurements of the rate of boundary movement and broadening are taken as a function of time, and this can be analysed to determine the sedimentation coefficient. As the boundary moves down the cell, boundary broadening from diffusion occurs and the concentration in the plateau region reduces due to radial dilution (Figure 3.4 (a)). Solute particles moving from the narrower end toward the bottom are dissolved in an increasingly larger volume of the solvent and a small drop in the solute concentration consequently occurs. Each species of solute in solution will produce a different concentration boundary in a sedimentation velocity experiment. Thus, single or multiple sedimenting boundaries can be taken as evidence of homogeneity or heterogeneity of

the sample. In some cases, two components of the system may have sedimentation coefficients so similar that they cannot be easily separated. Such components will not be well resolved by sedimentation velocity.

### **(3.2.5.2) Sedimentation equilibrium**

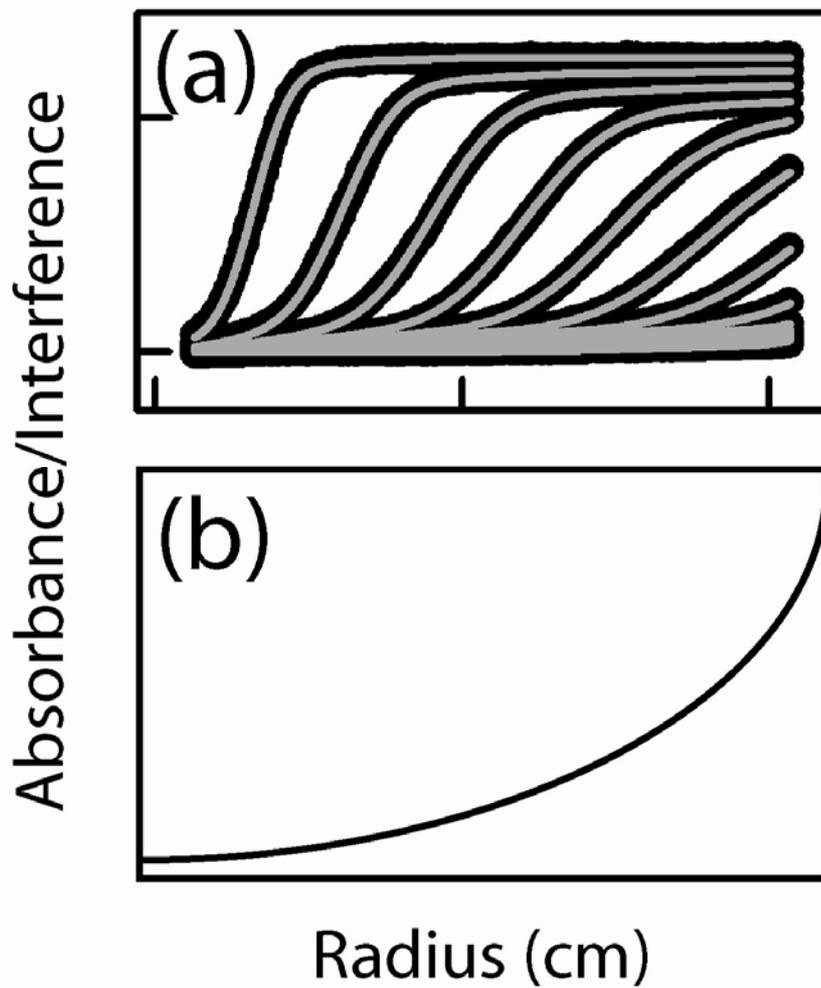
In sedimentation equilibrium experiments, a small volume of an initially uniform solution is centrifuged at a lower angular velocity than required for a sedimentation velocity experiment. As the solute begins to sediment towards the bottom of the cell and the concentration at the bottom increases, the process of diffusion opposes the process of sedimentation. After an appropriate period of time, the two opposing forces reach equilibrium ([Figure 3.4 \(b\)](#)). This technique is sensitive to the mass but not the shape of the molecules. When several species with different molecular weights are present, each of the species will be distributed over the sample cell until it reaches equilibrium. As a consequence, higher molecular weight species will be located towards the bottom of the cell, while the lower molecular weight species will dominate at the top. It provides complementary information to sedimentation velocity measurements and when both techniques are combined, the best results can be obtained. The molecular weight (MW) of the sample is determined according the equation:

$$(Eq. 3.11) \quad c_r = c_{r_0} \exp [(\omega^2/2RT) MW (1 - \bar{v} \rho) (r^2 - r_0^2)],$$

where  $c_r$  is the concentration at radius  $r$ ,  $c_{r_0}$  is the concentration of the protein at the reference radius  $r_0$ ,  $\omega$  is the angular velocity,  $R$  is the gas constant,  $T$  is the temperature (Kelvin), and  $\rho$  is the density of the buffer.

### **(3.2.5.3) Analytical ultracentrifugation data analyses**

There are a number of methods for analysing AUC data, with many alternative approaches and software packages available. For sedimentation velocity experiments, examples of data analysis programmes include the time derivative  $dc/dt$  transformation DCDT+ ([Stafford, 2000](#); [Philo, 2006](#)), the  $c(s)$  analysis (SEDFIT) and global analysis of interacting systems (SEDPHAT) software ([Schuck, 1998](#); [2000](#);



**Figure 3.4**

Boundary movement in sedimentation velocity and equilibrium experiments. (a) Movement of the boundary in a sedimentation velocity experiment as the boundary progresses down the cell. (b) Schematic representation of sedimentation equilibrium. Sedimentation is balanced by diffusion, which increases with concentration gradient. The boundary is shaped like an exponential curve as a result of the balance between sedimentation and diffusional forces.

2003), the Van Holde-Wieschet method in ULTRASCAN (Van Holde & Wieschet, 1978; Demeler & Van Holde, 2004; Demeler, 2005), and an alternative implementation of boundary modelling of reactive systems in time-difference mode, SEDANAL (Stafford & Sherwood, 2004). Other examples are the direct boundary modelling software LAMM (Behlke & Ristau, 1997) and SVEDBERG (Philo, 1997). For sedimentation equilibrium experiments, examples of analysis programmes include WINNONLIN (Johnson & Straume, 1994), HETEROANALYSIS (Cole, 2004) and ULTRASPIN (Cole *et al.*, 2008), all of which are specifically designed to implement a nonlinear least squares algorithm. Other programmes that can also be used for this include SEDPHAT, SEDANAL and ULTRASCAN. Software that implement the molecular weight moment determination method include SEDANAL, ULTRASPIN (Cole *et al.*, 2008) and MSTAR (Harding *et al.*, 1992).

In this thesis, two approaches were used for analysing sedimentation velocity data, namely the  $dc/dt$  and  $c(s)$  methods. In the  $dc/dt$  method, a number of closely-spaced scans are subtracted in pairs to approximate  $(\Delta c/\Delta t)$  the time derivative of the data and determine how much material is sedimenting at various rates (Philo, 2006; Cole *et al.*, 2008). Subtraction removes systematic noise, following which the data is transformed to an apparent sedimentation coefficient distribution  $g(s^*)$ . Several pairs of scans are averaged to improve the signal-to-noise ratio. Individual peaks in the  $g(s^*)$  distribution can be fitted to Gaussian functions to derive the concentration (from the peak area), the sedimentation coefficient (from the centre position) and the diffusion coefficient (from the width of the peak) for homogeneous species or non-interacting mixtures (Philo, 2006). The  $dc/dt$  analysis can provide a simple means to assess homogeneity and stoichiometry (for interacting species). For example, a shift in peak position with increasing concentration is evidence for a mass-action equilibrium where the species interact on the timescale of sedimentation. It is important to restrict the number of scans analysed to values below the peak broadening limit calculated by the software (DCDT+ version 2.06). This is because as the time span increases, the  $\Delta c/\Delta t$  curve calculated from each pair of scans deviates more from the true derivative,  $dc/dt$ . This deviation can lead to errors in diffusion coefficients.

The second approach used in this thesis is the  $c(s)$  analysis implemented in SEDFIT (Schuck, 1998; 2000), which describes a superposition of sedimentation boundaries of many species with different sedimentation coefficient ( $s$ ) values (Balbo

& Schuck, 2005). The sedimentation boundary for each species is fitted according to the Lamm equation (Lamm, 1929), which is derived by combining the law of diffusion with centrifugal migration. Analysis using SEDFIT does not restrict the number of scans that can be loaded, and this is an important difference from the  $dc/dt$  method. In addition, diffusion broadening is deconvoluted from the distribution based on the scaling relationship between  $s$  and  $D$  (Equation 3.10).  $c(s)$  analyses take into account the diffusion of all species by assuming the same frictional ratio for all sedimenting species (Schuck, 2000). The frictional ratio  $f/f_0$  is a measure of macromolecular elongation, where  $f$  is the observed frictional coefficient (Section 3.2.3), and  $f_0$  is the frictional coefficient of a sphere with the same volume as the hydrated macromolecule. The assumption of equivalent frictional ratio is based on the fact that  $f/f_0$  is not very sensitive to shape changes, and hence will be similar for broadly analogous molecules. The  $c(s)$  analysis results in peaks that reflect the populations of different sedimenting species, along with residuals for assessing goodness-of-fit. Since the  $c(s)$  distribution is a calculated curve obtained from modelling the raw data, it is important that the goodness-of-fit parameters (boundary fits, r.m.s.d. values, and residuals bitmaps) are carefully inspected to determine if the model adequately describes the experimental data. In general, the  $c(s)$  analysis will be strictly accurate only for dilute monodisperse samples or non-interacting protein mixtures.

For interacting mixtures, it is more difficult to assign peaks to discrete sedimenting species due to the presence of reaction boundaries (Dam & Schuck, 2005; Brown & Schuck, 2006). Reaction boundaries are peaks that correspond to the co-sedimentation of both free and complexed species. They are distinct from undisturbed boundaries, which correspond to discrete (free, uncomplexed) sedimenting species. Reaction boundaries occur in distinct proportions and at well-defined sedimentation coefficients in-between the  $s$  values of the free and complexed components (Scott & Schuck, 2005). However, if the peaks exhibit moderate or higher stability, the  $c(s)$  distribution and peak positions can give valuable insights into the reaction scheme and stoichiometry. Integration of the  $c(s)$  peaks gives the concentrations of the species. For a distribution containing a single major peak, the  $c(s)$  distribution can be transformed to a molar mass distribution  $c(M)$ . A  $c(M)$  distribution may also be used for heterogeneous and interacting mixtures. In such cases the accuracy will depend on the degree of similarity of the  $f/f_0$  values for the

different species, and the analysis will be valid only for peaks that correspond to sedimenting species and not reaction boundaries.

For sedimentation equilibrium experiments in this thesis, determination of molecular mass and equilibrium (association and dissociation) constants was done using the nonlinear least squares fitting method implemented in SEDPHAT (Schuck, 2003). Generally, interactions in sedimentation equilibrium lead to an increased steepness of the sedimentation profile (Figure 3.4(b)) for reason of the species with higher molecular mass. A model-free analysis of sedimentation equilibrium data of interacting systems can be implemented in the form of weight-average molecular mass as a function of loading concentration (Equation 3.11; Balbo & Schuck, 2005). Interaction is revealed by an increase in average molecular mass with sample concentration. Where the reaction scheme is known, data can be fitted with specific models to determine equilibrium constants. In this thesis, prior knowledge on the reaction scheme was obtained by sedimentation velocity (Chapters 4; 6). Global analysis of several concentrations at different speeds can be used to improve the accuracy of the analysis. As for SEDFIT analysis, the goodness-of-fit parameters provide a good indication as to whether the chosen model adequately describes the experimental data.

### **(3.3) Solution scattering**

#### **(3.3.1) Introduction to solution scattering**

Solution scattering is a low resolution diffraction technique that is used to study the overall structure of biological macromolecules in random orientations (Perkins *et al.*, 2008). X-ray crystallography is the most powerful method of obtaining three-dimensional models of proteins at atomic resolutions of up to 0.1 nm, but only allows molecules to be viewed in a fixed orientation. In contrast, solution scattering allows structures to be viewed at resolutions of about 2-4 nm, but in various orientations. Not only does solution scattering complement crystallographic investigation, it can also provide the only means of analysing multidomain protein structures that are not crystallisable for various reasons, such as heavy glycosylation or the presence of flexible inter-domain linkers. The technique also minimizes the need for non-physiological high concentrations of sample, meaning that proteins can be studied at lower concentrations which often more closely resemble physiological conditions. Scattering experiments performed over time can be used to monitor

biological events such as conformational changes or oligomerisation processes. In comparison to other methods of biological structure determination such as 2D-NMR, EM and AUC, solution scattering provides a useful complementary approach to structural studies (Perkins *et al.*, 2008).

### (3.3.2) Types of solution scattering

Two common types of solution scattering experiments are X-ray and neutron scattering (static and dynamic light scattering experiments are a third type). X-rays are diffracted by electrons, and neutrons by nuclei, but all types of scattering experiments make use of the same physical principle. A sample solution is irradiated by a highly collimated (divergent or convergent rays converted into a bundle of parallel rays with minimum divergence or convergence) beam of X-rays or neutrons. The intensity of the scattering curve  $I(Q)$  is measured as a function of the scattering angle  $2\theta$  (Figure 3.5), where the scattering vector  $Q = 4\pi \sin \theta/\lambda$  and  $\lambda$  is the wavelength.

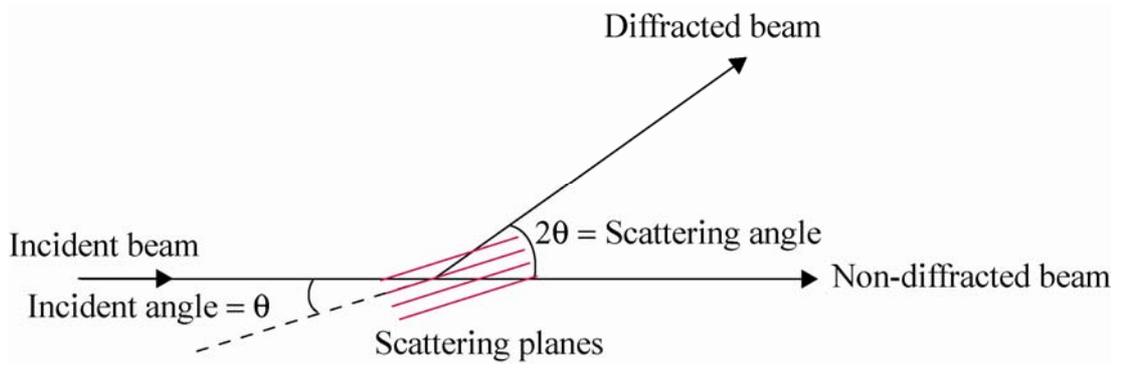
The overall radius of gyration ( $R_G$ ) and the forward scattering at zero angle  $I(0)$  can be obtained at small values of the scattering vector  $Q$  (Figure 3.6), while the cross-sectional radius of gyration ( $R_{XS}$ ) can be determined at higher  $Q$  values. Solution scattering  $Q$  ranges are usually between  $0.05$ - $3 \text{ nm}^{-1}$ , and analysis of the scattering curve by indirect Fourier transform gives the distance distribution function  $P(r)$ , from which the maximum dimension of the molecule and its shape can be deduced. When used in conjunction with information from known protein sequences and crystal structures, the structural resolution of solution scattering is improved to a medium-resolution level of about  $0.5$ - $1.0 \text{ nm}$ , which is the same resolution provided by homology modelling (where a known crystal structure is used to model another related protein).

### (3.3.3) Theory of Solution Scattering

Diffraction is described by Bragg's Law according to the equation

$$\text{(Eq. 3.12)} \quad \lambda = 2d \sin \theta,$$

where  $d$  is the diffraction spacing, and  $2\theta$  is the angle of diffraction (or scattering angle). In solution scattering,  $d$  corresponds to dimensions of  $1$ - $100 \text{ nm}$ . Scattering



**Figure 3.5**

A schematic diagram showing the scattering angle ( $2\theta$ ) from the main beam.

angles are given in terms of the scattering vector  $Q$ . The resolution of a scattering experiment is usually defined by  $2\pi/Q_{max}$  (where  $Q_{max}$  is the largest  $Q$  value measured in the experiment). For any pair of scattering elements (electrons for x-rays, nuclei for neutrons), when  $2\theta$  is zero, the scattered waves are in phase and the intensity of scattering is at a maximum. Therefore, at  $2\theta = 0$ , the intensity of scattering,  $I(0)$ , is a direct measure of molecular mass.

When  $Q$  and  $2\theta$  are non-zero, the intensity of scattering  $I(Q)$  is given by the Debye equation:

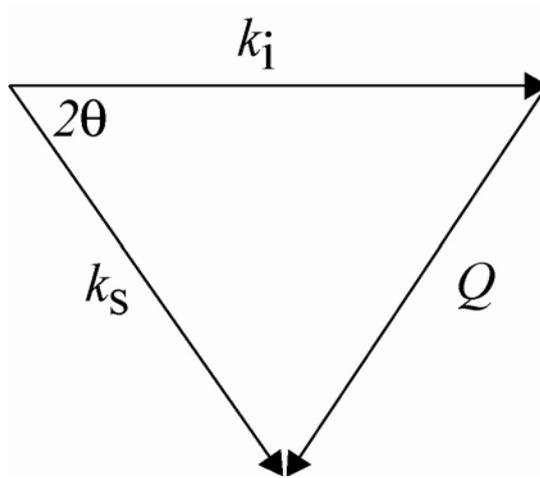
$$\text{(Eq. 3.13)} \quad I(Q) = \sum_p \sum_q f_p f_q \frac{\sin(rQ)}{rQ}$$

where the terms  $f_p$  and  $f_q$  (or  $b_p$  and  $b_q$  for neutrons) are the scattering lengths of the electrons (or nuclei) at points P and Q in the molecule, and  $r$  corresponds to the distance between P and Q. Whereas scattering length,  $f$ , is proportional to atomic number for X-rays,  $b$  for neutrons is roughly similar for different nuclei with the exception of  $^1\text{H}$  (Table 3.1). The scattering density of a molecule is the total of all the scattering lengths within that molecule divided by its molecular volume.

### (3.3.4) Properties of X-ray scattering experiments

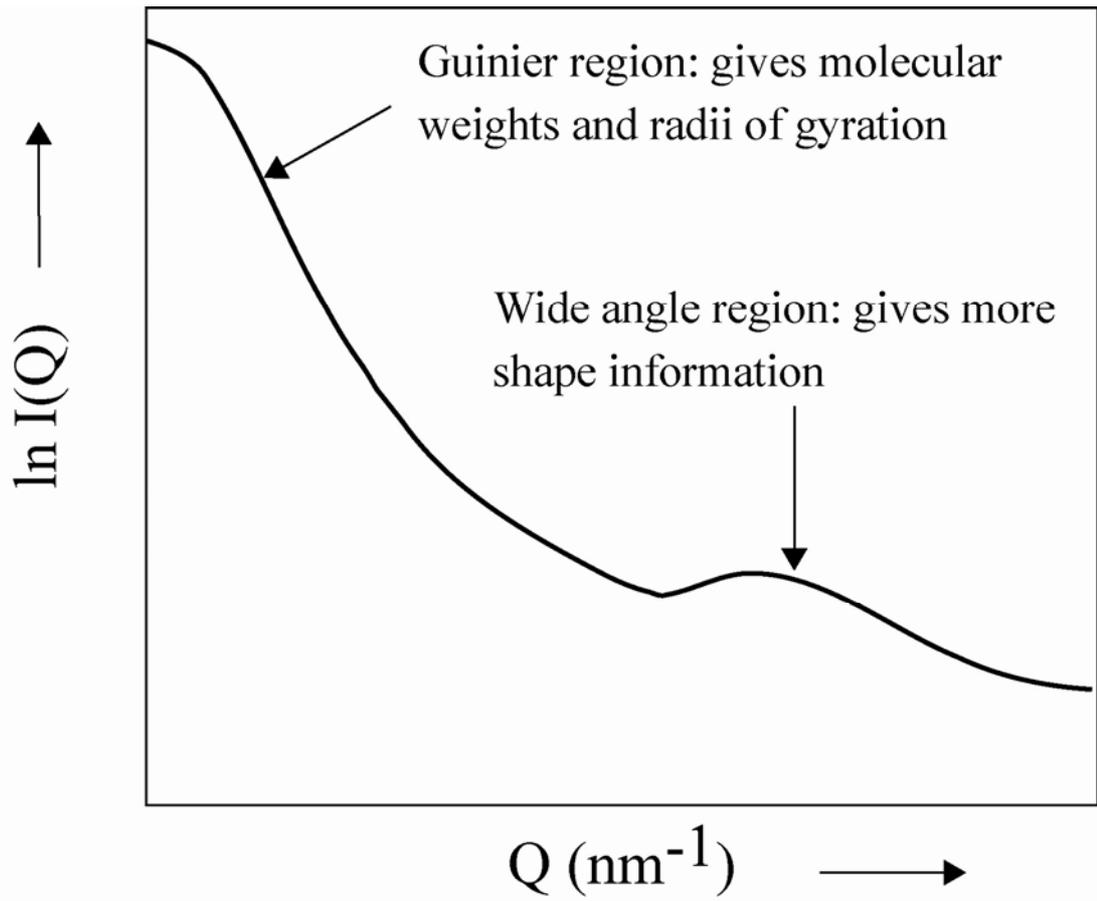
(1) Most biological macromolecules are studied in high positive solute-solvent contrasts; only lipids have an electrical density less than that of water. Because positive contrast usually corresponds to the situation when the scattering density of the whole macromolecule is significantly higher than that of the solvent, the effect of this is to minimize systematic errors in the curve modelling of proteins and glycoproteins as internal density fluctuations can be neglected.

(2) Good counting statistics are obtained in this positive contrast, provided that high background levels from the instrument are minimized. This is not always possible at low  $Q$  values due to high instrumental background scattering levels. In distinction, the positive contrasts obtained using neutron data for samples in  $\text{H}_2\text{O}$  buffers suffer from a high incoherent scattering background, as a result of which neutron signal-noise ratios can be poor for dilute samples.



**Figure 3.6**

Definition of the scattering vector,  $Q$ .  $Q$  is usually measured in  $\text{nm}^{-1}$ , and  $k_i$  and  $k_s$  are the momentum of the incident and scattered rays, respectively.



**Figure 3.7**

General features of a solution scattering curve showing the Guinier region at low and high  $Q$  values.

(3) Instrumental errors caused by wavelength polychromicity and beam divergence are minimal for X-ray scattering experiments. Thus Guinier and wide-angle analyses are not affected by systematic errors caused by the instrument geometry, and curve corrections do not have to be applied.

(4) X-ray scattering reveals the hydrated dimensions of the macromolecule. In solution, a monolayer of water molecules is hydrogen bonded to the protein surface, meaning that the electron density of bound water is increased compared to that of bulk water. The macromolecule is made to appear larger by the thickness of this water monolayer, which is readily detected by X-rays.

### **(3.3.5) Instrumentation**

The probability of a diffraction event when an X-ray photon approaches an electron is very low at  $10^{-25}$ . In distinction to the use of scattering in other disciplines such as chemistry or metallurgy, a drawback of biological experiments lies in low signal-to-noise ratios. This is because biological samples are typically dilute compared to other disciplines. The use of high flux sources overcomes this limitation.

High flux X-ray sources characteristically comprise a synchrotron to accelerate an electron beam to high energies and a storage ring to maintain this beam for several hours. The electrons in the storage ring travel at very high speeds, and are constrained in a circular ring trajectory by electromagnets. This causes them to radiate energy, including X-rays. The lost energy is replenished by an oscillatory radiofrequency electric field. The use of magnetic devices such as undulators or wigglers improves the beam brightness. The brilliance of a synchrotron source (the flux divided by the angle through which the radiation is emitted and by the cross-sectional area of the source) is the most important parameter for solution scattering. It is usually expressed in units of number of photons·s<sup>-1</sup>·mrad<sup>-2</sup>·mm<sup>-2</sup> per 0.1% relative bandwidth, and is often plotted as a function of photon energy or wavelength. Other important properties of synchrotron radiation for solution scattering are the very good intrinsic collimation of the beam and the range of wavelengths available ([Perkins \*et al.\*, 2008](#)).

### **(3.3.6) X-ray sources**

There are about 50 X-ray synchrotron sources in the world, of which the largest and most powerful ones are currently the 8 GeV (Giga electron Volt) super

(a)

		Atomic number	$f(2\theta = 0^\circ)$ (fm)	B (fm)
<b>Hydrogen</b>	$^1\text{H}$	1	2.81	-3.742
	$^2\text{H}$	1	2.81	6.771
<b>Carbon</b>	$^{12}\text{C}$	6	16.9	6.651
<b>Nitrogen</b>	$^{14}\text{N}$	7	19.7	9.40
<b>Oxygen</b>	$^{16}\text{O}$	8	22.5	5.804
<b>Phosphorus</b>	$^{31}\text{P}$	15	42.3	5.1

(b)

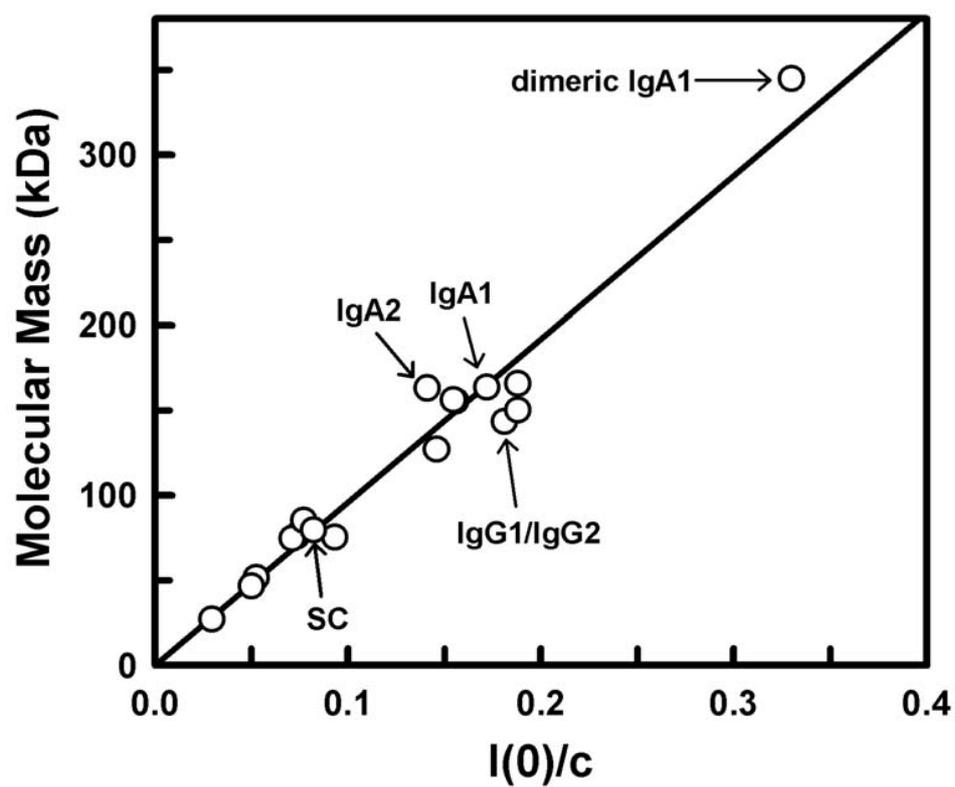
Macromolecule	X-rays ( $\text{e}^{-}\text{nm}^{-3}$ )	Neutrons (% $^2\text{H}_2\text{O}$ )
<b>H<sub>2</sub>O</b>	334	0
<b><math>^2\text{H}_2\text{O}</math></b>	334	100
<b>50 % (w/w) sucrose in H<sub>2</sub>O</b>	402	13
<b>Lipids</b>	310-340	10-14
<b>Detergents</b>	300-430	6-23
<b>Proteins</b>	410-450	40-45
<b>Carbohydrates</b>	490	47
<b>DNA</b>	590	65
<b>RNA</b>	600	72

**Table 3.1**

Examples of scattering lengths and densities (Perkins, 1988a).

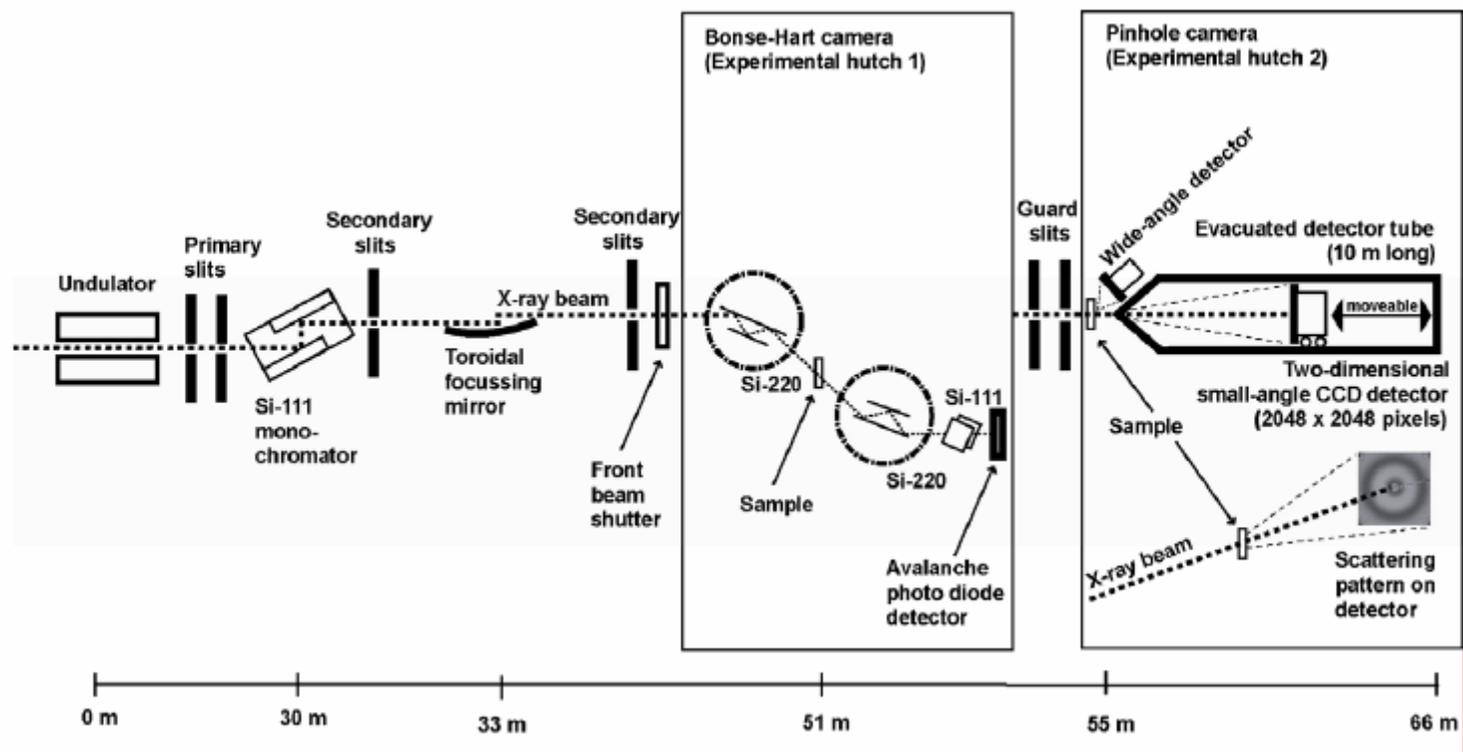
(a) Scattering lengths of biologically important nuclei

(b) Scattering densities of solvents and biological macromolecules.



**Figure 3.8**

Linear relationship between the molecular weight and neutron  $I(0)/c$  values for glycoproteins in 100%  $^2\text{H}_2\text{O}$  buffer on the LOQ instrument (Perkins *et al.*, 2008).



**Figure 3.9**

Schematic layout of the X-ray ID02 beamline at ESRF. The undulator provides a high photon flux with a low divergence. Only the central radiation cone of the undulator is used. The beamline optics consist of a cryogenic- (liquid nitrogen) cooled Si-111 channel-cut monochromator and a focusing toroidal mirror. There are two separate experimental stations, one for combined small/ wide angle scattering and the other for ultra-small angle scattering (not shown) (Perkins *et al.*, 2008).

photon ring (Spring8) near Kyoto, Japan, the 6 GeV European synchrotron radiation facility (ESRF) in France, the 7 GeV advanced photon source (APS) in Chicago, Illinois, USA. Recently, a third-generation 3 GeV synchrotron (Diamond) has commenced full operation in the UK. The first accelerators (cyclotrons) were built by particle physicists in the 1930s, and atomic nuclei were split in these using the collision of high energy particles. Synchrotron radiation was seen for the first time at General Electric (USA) in 1947 in a different type of accelerator (synchrotron). It was first considered a nuisance because it caused the particles to lose energy, but it was then recognised in the 1960s as light with exceptional properties. A schematic diagram of the ID02 beamline at ESRF is shown in [Figure 3.9](#).

### (3.3.7) Data Interpretation

#### (3.3.7.1) Guinier Analyses

At sufficiently low  $Q$  values, the Debye equation becomes the Guinier approximation ([Perkins, 1988a; 1988b](#)). The overall radius of gyration  $R_G$  and the scattering intensity at zero scattering angle  $I(0)$  parameters are obtained from a plot of  $\ln I(Q)$  against  $Q^2$  (where  $Q$  is the scattering vector) ([Figure 3.10](#)):

$$(Eq. 3.14) \quad \ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}$$

The  $R_G$  parameter characterizes the degree of elongation of the structure. It is analogous to the sedimentation or diffusion coefficients of hydrodynamic experiments.  $I(0)$  is proportional to  $M_r^2$  ( $M_r$  = relative molecular mass). Guinier plots are valid only at low  $Q$ , but the lowest  $Q$  values cannot be measured for reason of overlap with the main beam and beam stop. The smallest  $Q$  required to measure  $R_G$  is given by  $Q \cdot D_{max} \leq \pi$ , where  $D_{max}$  is the maximum particle dimension ([Glatter & Kratky, 1982](#)). The largest  $Q$  permitted in Guinier plots depends on the particle shape, but the value of  $Q \cdot R_G$  is generally no larger than 1.5. The cross-sectional radius of gyration  $R_{XS}$  can be calculated for rod-like and ellipsoid shapes according to the general formula:

$$(Eq. 3.15) \quad \ln[I(Q) \cdot Q] = \ln[I(Q)Q]_{Q \rightarrow 0} - \frac{R_{XS}^2 Q^2}{2}$$

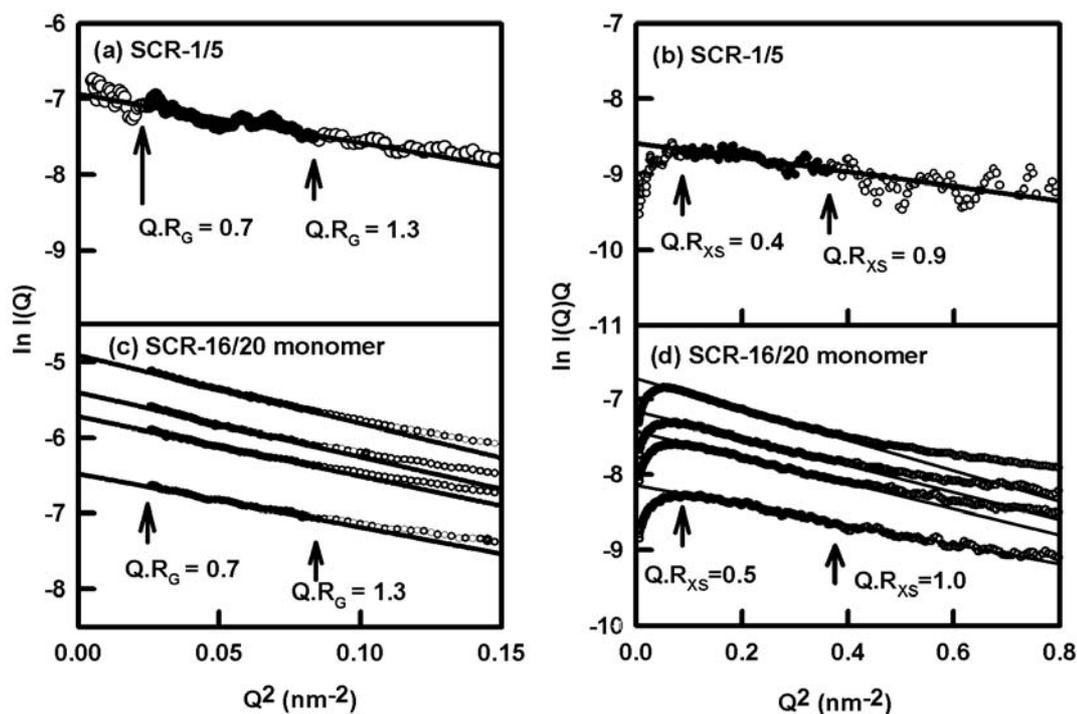
where  $[I(Q)Q]_{Q \rightarrow 0}$  is the cross-sectional intensity at zero angle and curve fit analyses are done in a different  $Q$  range that is larger than, and does not overlap with, the one used for the  $R_G$  determinations (Figure 3.10). The  $R_G$  and  $R_{XS}$  analyses lead to the triaxial dimensions of the macromolecule if the structure can be represented by an elliptical cylinder, where the maximum dimension,  $L$ , is given by the equation

$$\text{(Eq. 3.16)} \quad L = \sqrt{12(R_G^2 - R_{XS}^2)}$$

Concentration series for the samples should yield linear, reproducible Guinier plots. Molecular mass analyses are based on the  $I(0)$  parameter of Guinier plots. Thus after normalisation of the sample concentration,  $c$ , the  $I(0)/c$  values obtained in a given measurement will yield relative  $M_r$  data. It is important to note that absolute molecular mass values cannot be calculated from X-ray experiments, but can be obtained from neutron studies (Figure 3.8). Once validated using the  $I(0)/c$  values, the  $R_G$  values can be analysed to obtain the mean and standard deviation. From the available data, the magnitude of any concentration dependence of the  $R_G$  data can be determined. If the  $R_G$  increases with dilution, interparticle effects may be important (when the protein molecules are too close together). If it decreases with dilution, dissociation may be occurring. In order to interpret the  $R_G$  value properly, it must be compared with that calculated for a sphere of the same hydrated volume ( $R_0$ ).  $R_G/R_0$ , or the anisotropic ratio, will indicate the degree of macromolecular elongation when compared to a set of standard values for different shapes (Perkins, 1988b).

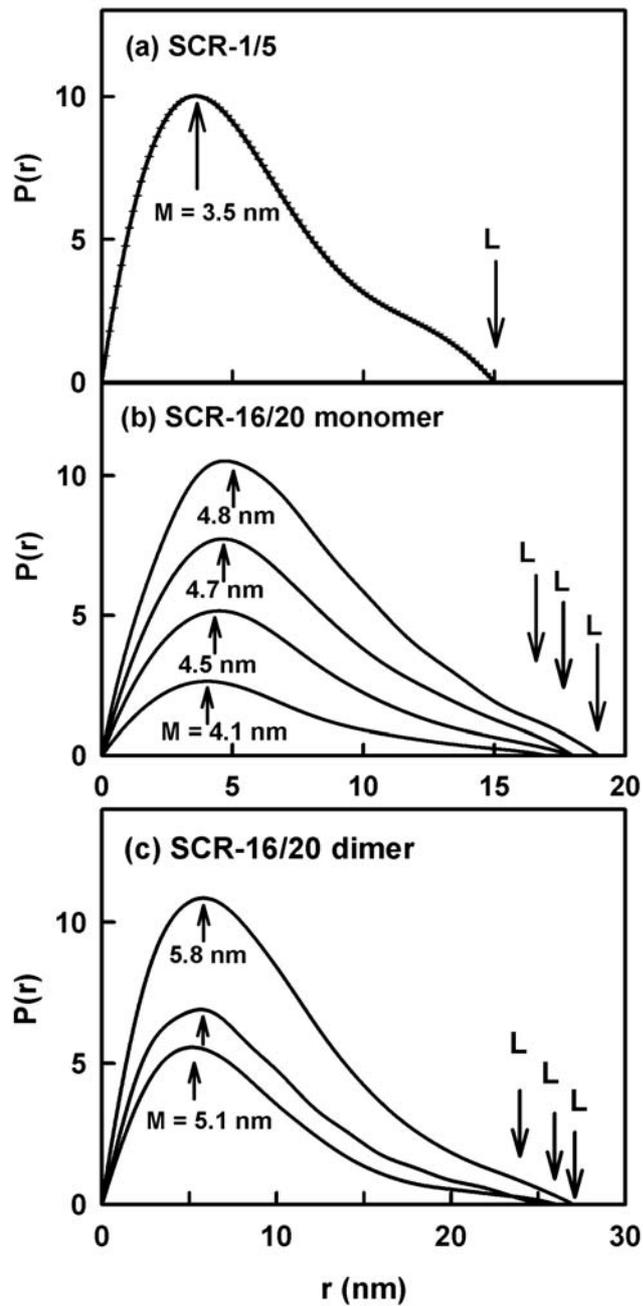
### (3.3.7.2) Distance distribution function analyses

The distance distribution function  $P(r)$  is measured in  $r$  units of nm and reveals the structure of the macromolecule in real space, in contrast to the scattering curve which is measured in reciprocal space with units of  $\text{nm}^{-1}$ . Indirect Fourier transformation of the scattering curve  $I(Q)$  yields the distance distribution function if a value for the maximum dimension  $L$  is assumed. Its main advantage is that  $L$  is readily obtained from the point where  $P(r)$  drops to zero (Figure 3.11). At this point  $r = L$ . This transformation requires the range  $0 \leq Q \leq \infty$ , a range which introduces some difficulty, since data points will be missing at the lowest  $Q$  values due to the beam stop, and are truncated at the highest  $Q$  values at the limit of the active detector region



**Figure 3.10**

Guinier  $R_G$  and  $R_{XS}$  analyses of SCR-1/5 and SCR-16/20 of FH. The  $I(Q)$  data in the Guinier regions used to obtain  $R_G$  values are denoted by filled circles in the range enclosed by arrows. Plots are displaced arbitrarily on the intensity axis for reason of clarity (Chapter 4).



**Figure 3.11**

Distance distribution functions  $P(r)$  for SCR-1/5 and SCR-16/20 of FH. The maximum of the  $P(r)$  curve is denoted by  $M$ , the most frequently occurring intramolecular distance.  $L$  denotes the maximum length, which is taken to be where the  $P(r)$  curve intersects  $P(r) = 0$  (Chapter 4).

This problem is dealt with by the use of an automated procedure, the GNOM program of Svergun ([Semenyuk & Svergun, 1991](#)) which makes use of 10 to 20 B-spline mathematical functions optimally fitted to the scattering curve in order to represent the data as a continuous analytical function, after which the B-splines are transformed to give  $P(r)$ .

### **(3.4) Biomolecular modelling**

#### **(3.4.1) Scattering curve modelling**

After data processing, molecular modelling of the X-ray and neutron scattering curves is performed to verify and extend data interpretation. The fundamental principle of scattering modelling is the generation of a three-dimensional model for a structure that will give rise to a calculated scattering curve corresponding to that observed experimentally. Scattering modelling improves the utility of solution scattering by providing a means of ruling out classes of structures that are incompatible with the experimental scattering curves. It is especially useful for multidomain proteins for which overall structures are unknown, yet molecular structures are available for all of the individual domains. Examples of such cases are large multidomain proteins that are not crystallisable (while intact) for reason of interdomain flexibility or heavy surface glycosylation. It should be noted that unique structures cannot be determined by this method. However it can be used to identify a group of best-fit conformers that meet well-defined selection criteria ([Section 3.4.2](#)). An outline of the modelling procedure is shown in [Figure 3.12](#).

##### **(3.4.1.1) Analysis of protein composition**

The first step towards solution scattering modelling is to determine essential protein parameters from the amino acid composition of the protein using the program SLUV ([Perkins, 1986](#)). SLUV also takes into account the actual or putative carbohydrate content of the protein. The most important parameters are the protein partial specific volume, estimated molecular mass and molecular volumes (hydrated and unhydrated). For concentration measurements, the absorption coefficient (at a wavelength of 280 nm) is calculated from the tryptophan, tyrosine and cysteine content of the protein ([Perkins, 1986](#)).

### **(3.4.1.2) Creating atomic models**

The use of known atomic structures to constrain the modelling of the scattering curve greatly improves the usefulness of the information provided by structural modelling. The first stage of this process is to identify the atomic structures that best represent the subunits of the full structure to be modelled, whether directly obtained from crystal or NMR structures, or indirectly by the application of homology or “comparative” modelling.

### **(3.4.1.3) Molecular model searches**

After identifying the subunit structures, a set of molecular models, typically in the range of 2,000-10,000 is generated in random conformations. The details of the so-called “search” will depend on certain features such as the type of molecule that is being modelled, the quality of the atomic coordinates being used and hence the particular parameter that is being modelled, and its interaction with other molecules. Thus a search can be carried out in a variety of ways aimed at different permutations of the particular feature that is being randomly varied in the models.

For a multidomain or multifragment protein, a translational search optimises the relative positions of the individual fragments within each molecule, while a rotational search is based on an automated set of systematic domain rotations with a fixed interdomain separation. A translational search therefore looks for the interdomain or inter-fragment distances without changing the domain rotational positions, while a rotational search looks for different possible rotations while keeping the interdomain distances fixed. This approach has been used to successfully model factor I of complement ([Chamberlain \*et al\*, 1998](#)).

Symmetry often plays a role in the curve modelling of oligomeric proteins, and these considerations lead to the most precise scattering modelling fits. In this method, the symmetry of the oligomer (where known) is fixed while doing a translational search to optimise the positions of the monomers. One molecule can be held fixed while the other is translated along the symmetry axis in short steps of, say, 0.1 nm. This method was successfully used in modelling AmiC, a two-domain periplasmic binding protein existing in a monomer-trimer equilibrium, and RuvA, a disc-shaped tetrameric molecule that forms face-to-face octamers and binds to a four-stranded DNA helix structure called a Holliday junction ([Chamberlain \*et al\*, 1998](#)).

Additionally for multidomain proteins, it is sometimes better to replace the straightforward rotational and translational searches using a fragmented subunit structure with a full covalently-connected structure. This avoids the arbitrary nature of these searches, as each model is now stereochemically correct before it enters the curve fit process. If the conformationally variable inter-domain linker peptides are modelled using structural libraries calculated by molecular dynamics simulations, full models of the protein are generated by an automated process that assembles randomly selected linker peptides with the individual domains to create the full structure. Curve fits then become a trial-and-error procedure that is left to run until a sufficient number of good-fit solutions are obtained. This method has been used successfully with SCRs 1/20, 1/5, 6/8 and 16/20 of FH ([Aslam & Perkins, 2001](#); [Fernando \*et al.\*, 2007](#); [Chapter 4](#)).

#### **(3.4.1.4) Sphere modelling**

The next stage is to convert the randomly generated three-dimensional protein molecular models into sphere models consisting of small non-overlapping spheres of uniform density and with the same total sphere volume as the dry protein. If a single subunit protein with a known rigid, well-defined and high-resolution structure is being studied, the coordinates of the atomic structure or model are converted directly to spheres by placing all the atoms within a three-dimensional grid of cubes of side about 0.6 nm. A cube is included as a sphere in the model if it contains sufficient atoms above a specified cut-off level such that the total volume of all the cubes that result from this cut-off equals that of the unhydrated protein calculated from the sequence. If residues are missing in the crystal structure or a homologous crystal structure is used, the ensuing volume discrepancy can be compensated by adjustment of the cut-off value for sphere generation in structures. The simplest way to hydrate the cube models is to increase the size of the cubes so that their total volume matches the hydrated volume. However, this can significantly distort the macromolecular structure if, for example, there is a void space at its centre. An improved approach is to add a layer of additional spheres evenly over the surface of the dry sphere model, where the total number of spheres gives the hydrated volume ([Perkins, 1986](#); [Ashton \*et al.\*, 1997](#); [Perkins \*et al.\*, 2008](#)).

### (3.4.1.5) Debye scattering curve calculation

Curves are calculated from Debye's law adapted to small spheres (Perkins *et al*, 1998):

$$(Eq. 3.17) \quad \frac{I(Q)}{I(0)} = g(Q) (n^{-1} + 2n^{-2} \sum_{j=1}^m A_j \frac{\sin Qr_j}{Qr_j})$$

$$(Eq. 3.18) \quad g(Q) = (3(\sin QR - QR \cos QR))^2 / Q^6 R^6,$$

where  $g(Q)$  is the squared form factor for the sphere of radius  $R$ ,  $n$  is the number of spheres filling the body,  $A_j$  is the number of distances  $r_j$  for that value of  $j$ ,  $r_j$  is the distance between the spheres, and  $m$  is the number of different distances  $r_j$ . These spheres have to be sufficiently small (about 0.6 nm) so that their form factor in the Debye equation is almost invariant in the  $Q$  scattering range used. Interference between scattered waves gives rise to diffraction phenomena and consequently the spatial arrangement of electrons in the protein determine the scattering curve. The curves generated from this calculation can be compared directly with experimental curves to determine best-fit models.

### (3.4.2) Identification of best-fit models

The identification of best-fit models and the removal of unsatisfactory models is achieved by the use of four filters:

- (1) The systematic creation of models can result in physically unreasonable steric overlaps between domains or subunits. This is readily detected by the grid transformation which will give too few spheres if significant overlap occurs, and so a requirement of at least 95% of the expected total of spheres is necessary to ensure that only good models are selected.
- (2) The  $R_G$  and  $R_{XS}$  values determined from the calculated curves in the same  $Q$  ranges used in the experimental Guinier fits of the experimental data are required to be within 5% of the experimental values in a good model.
- (3) The remaining good models are then assessed using a goodness-of-fit R-factor, a rank ordering of all the good models to define the best-fit structures. An R factor is calculated for a quantitative comparison of each model scattering curve  $I(Q)_{cal}$  against experimental scattering data  $I(Q)_{exp}$  over the whole  $Q$  range, where  $I(Q)_{cal}$  is set as 1000. R factor is given by the equation:

$$(Eq. 3.19) \quad R = \frac{\sum |I(Q)_{exp} - I(Q)_{cal}|}{\sum |I(Q)_{exp}|} \times 100\%$$

The R factor used here is analogous to that used for the refinement of crystallographic models. A good fit should therefore have an R factor less than 10%. A large spreadsheet is prepared with the geometrical steps used to define each model, the number of spheres in it, the  $R_G$  and  $R_{XS}$  values, and the R factor values. This is used to optimize the best cut-off filters, sort the models, and identify the best family of structures.

(4) Sedimentation coefficients of the best-fit models are calculated from the hydrated sphere models used for X-ray fits for comparison with the experimental values as an independent monitor.

### **(3.4.3) Calculation of hydrodynamic properties**

Sedimentation coefficients are calculated directly from the hydrated scattering sphere models using the HYDRO program, (Garcia de la Torre *et al.*, 1994) and also directly from the atomic coordinates in the HYDROPRO shell modelling program using the default value of 0.31 nm for the atomic element radius for all atoms to represent the hydration shell. (Garcia de la Torre *et al.*, 2000). The sedimentation coefficient values are the final independent filter for the best-fit molecular models.

### **(3.5) Introduction to protein interaction measurement**

Studying the interaction of a protein with ligand can often provide insight into its structure and help to clarify structure-function relationships. There are several methods used to detect these interactions, with varying degrees of specificity and sensitivity. In addition to AUC and solution scattering (Sections 3.2; 3.3), examples include co-immunoprecipitation, enzyme-linked immunosorbent assay (ELISA), label transfer, protein-protein docking, fluorescence resonance energy transfer (FRET) and surface plasmon resonance. Co-immunoprecipitation of endogenous proteins using antibodies is one of the most popular assays for verifying real interactions between suspected interaction partners. In this method, interacting proteins are identified by Western blotting, but since the technique involves specific interactions, it cannot be used for screening assays. In the ELISA technique, a protein is immobilised on a solid

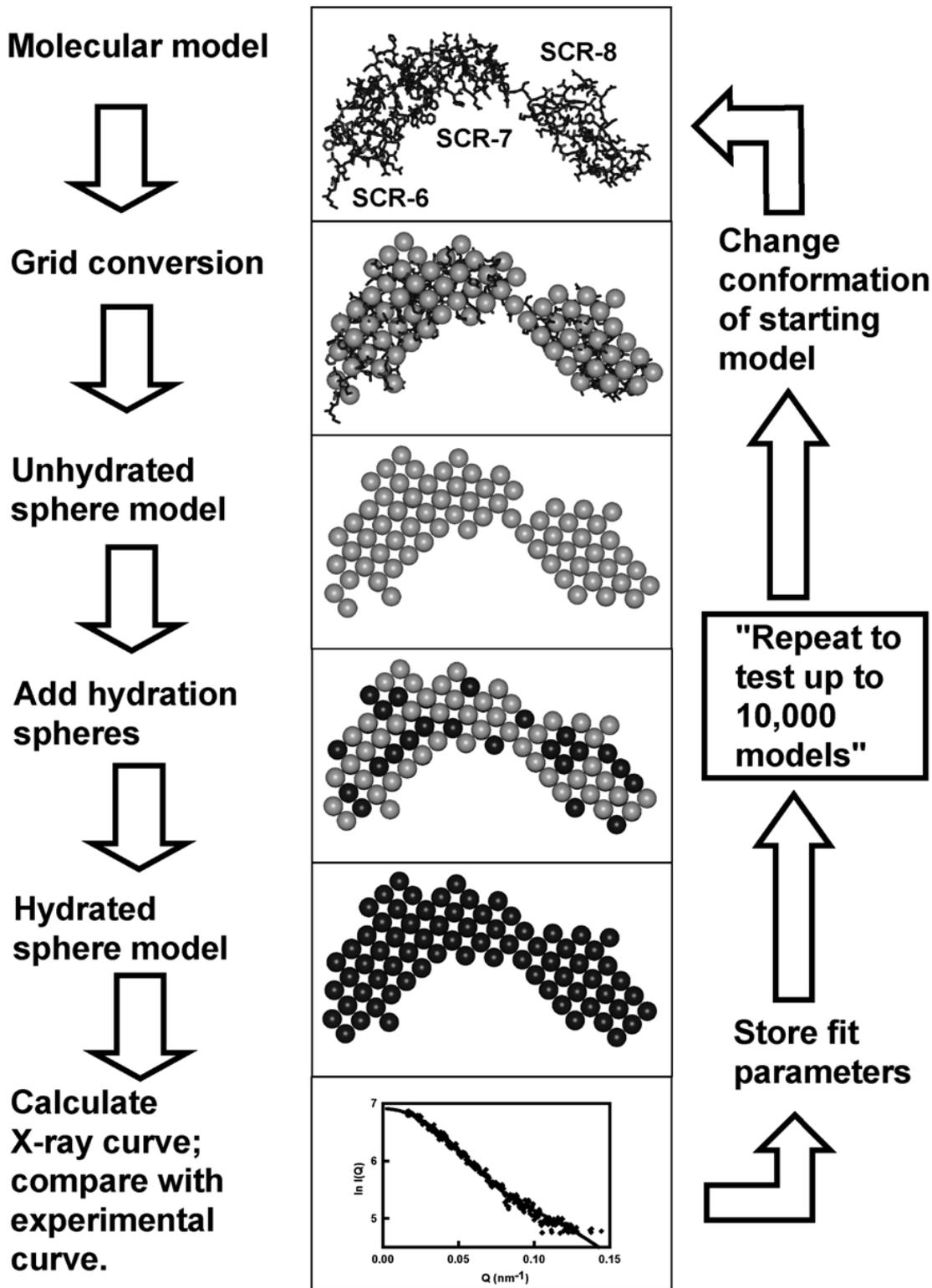


Figure 3.12

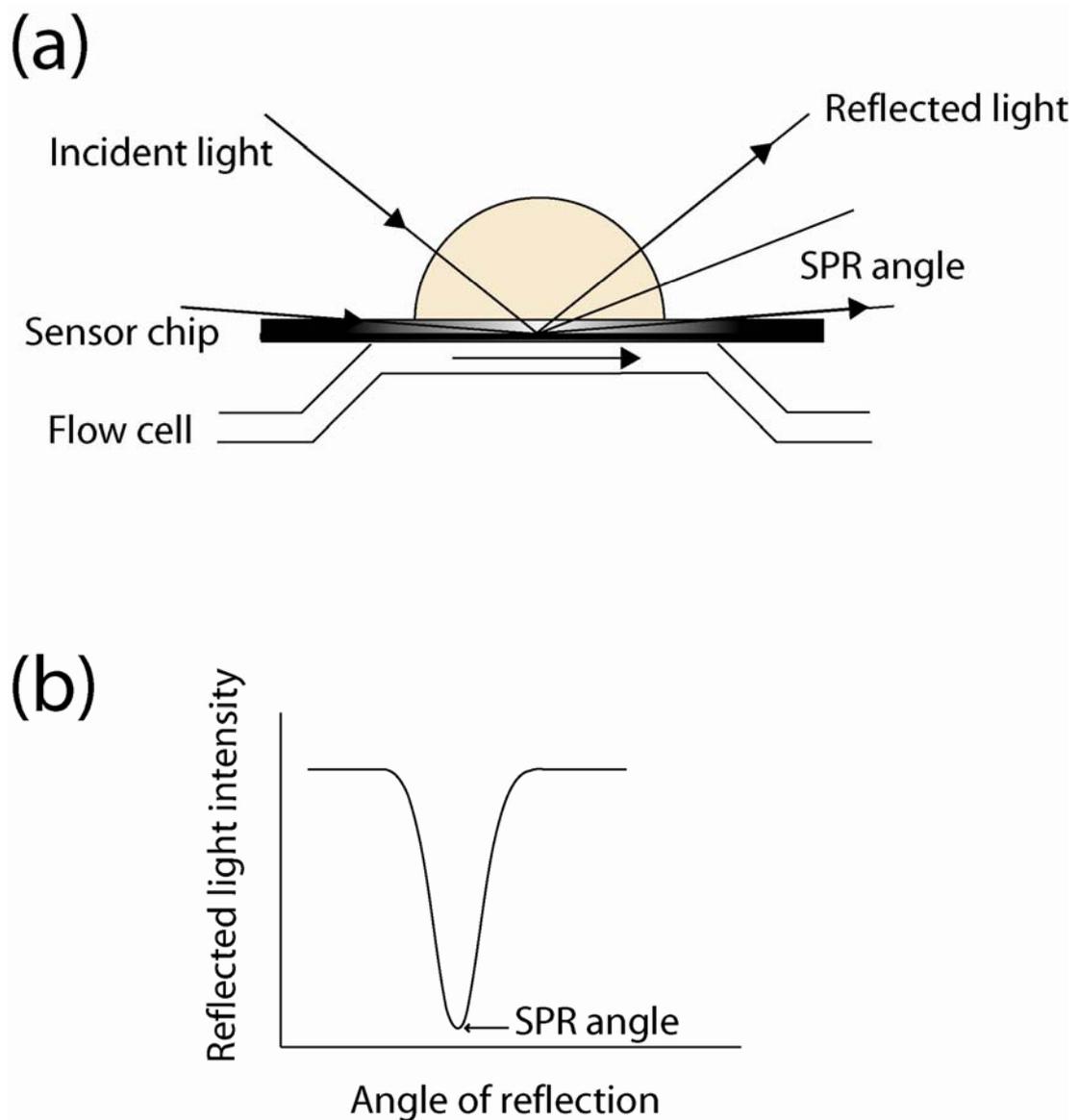
Flow chart of the automated modelling procedure (Taken from Perkins *et al.*, 2008).

support (usually plastic) either directly by adsorption or indirectly by high affinity molecular capture (for example, using an antibody). When the interaction partner is added, interaction is detected using an enzyme-linked antibody specific for the interaction partner. Substrates for the enzymatic detection are usually chromogenic or fluorogenic. The label transfer technique involves transferring a label between two interacting proteins, and this can provide information about reaction sites. It can be used for in-vitro screening of weak or transient reactions that are difficult to study using other methods. Protein-protein docking is a method involving computational modelling based on known structures for two potentially interacting molecules. Since this method is based on physical principles, it can be used to study relatively unknown interaction partners. However, it needs to be complemented and verified by experimental methods. FRET is a method based on the transfer of energy between two chromophores located within a certain distance (typically within 10 nm) of each other. Interactions can be studied if these chromophores are located on different potentially interacting molecules.

### **(3.5.1) Surface plasmon resonance**

Surface plasmon resonance (SPR) is a phenomenon in which electron charge density waves (plasmons) are excited at the surface of a thin conducting film by incident light. It is the basis of measuring the interactions of molecules with material fixed on planar metal surfaces or metal nanoparticles. In its simplest form, SPR measurements can be used to detect changes in refractive index caused by binding of one molecule to another. The first biosensor based on SPR was developed in the early 1990s (Jönsson *et al.*, 1991; Jönsson & Malmqvist, 1992), and this technique is now very widely used for analysing macromolecular interactions. The Biacore™, manufactured by GE Healthcare, is the most widely-used SPR-based system and offers particular advantages for analysing weak macromolecular interactions (Van der Merwe, 2000).

Light passing from a medium with higher refractive index to one with lower refractive index is partly reflected and partly refracted. Above a certain critical angle of incidence, no light is refracted across the refractive index interface, and total internal reflection is observed. Under conditions of total internal reflection, incident light generates an evanescent wave field that passes a short distance (tens to hundreds of nanometers) across the refractive index interface between different media. If this



**Figure 3.13**

Diagram illustrating the SPR principle. (a) shows the Biacore™ detection setup, while (b) illustrates the reduction of reflected light intensity on binding of molecules to the sensor chip surface. (Adapted from the [Biacore X100 handbook, 2008](#))

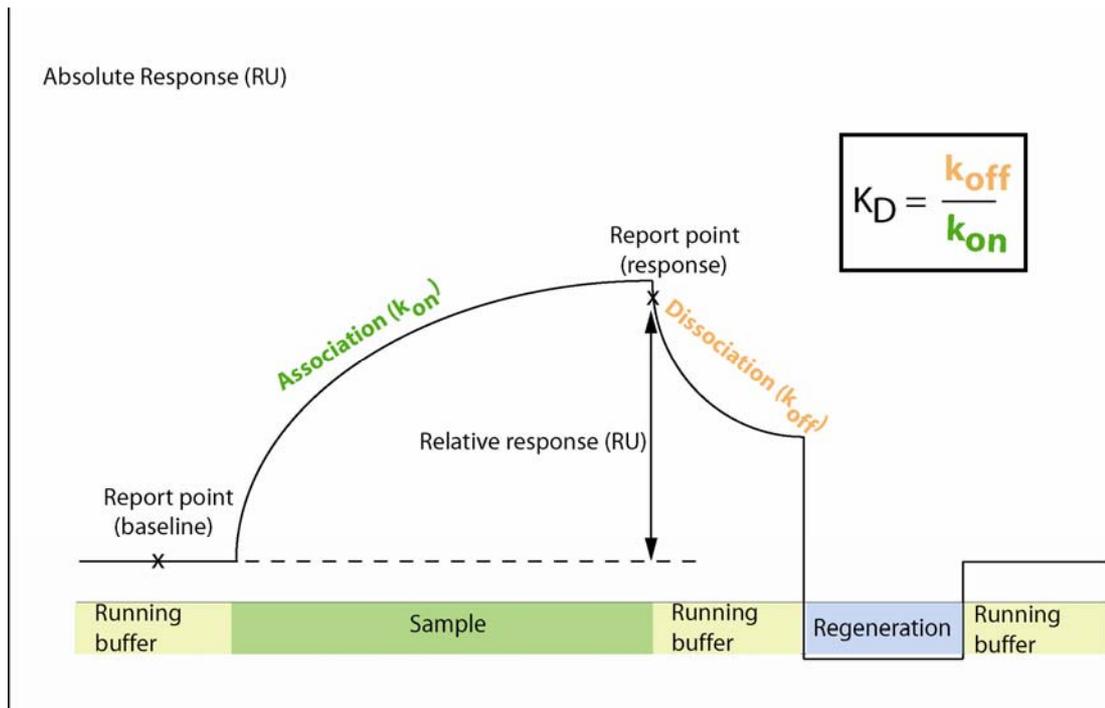
interface is coated with a thin layer of metal, and the light is monochromatic and plane-polarised, the intensity of reflected light (reflectivity) of the incident light drops markedly at an angle called the SPR angle (Figure 3.13(a,b)). In the Biacore detection system, these media are the immobilising surface (sensor chip) and the sample or buffer solution within the flow cell (Figure 3.13(a)). The SPR angle is very sensitive to refractive index changes in the solution at the sensor chip surface. At a fixed wavelength and angle of incidence, these changes can be directly correlated to shifts in the SPR angle. This is the basis of measurement of molecular interactions in the Biacore detection system.

### (3.5.2) Experiments on the Biacore X100

The Biacore X100 is an instrument controlled from a PC running Biacore X100 Control and Evaluation Software. It uses label-free detection for real-time monitoring of molecular interactions involving practically any biological molecule, including larger structures such as cells and viruses.

The sensor chip in a Biacore system is a glass slide covered by a gold film and mounted on a supporting frame. There are several sensor chip types available for use on the Biacore (Table 3.2). Interaction between the molecules of interest occurs on the gold film surface which forms the lower wall of the flow cell (20-60 nanolitres in volume) when the sensor chip is docked (Figure 3.13(a)). Running buffer passes through the flow cell at a continuous flow rate of 1-100  $\mu\text{l}/\text{min}$ . The SPR response is a measure of the refractive index of the solution within 300 nm from the surface of the sensor chip. Since this distance is small in relation to the sample volume, the SPR response effectively measures refractive index changes at the chip surface. This is related to the total mass of material bound to the chip surface, although a difference between the sample and running buffers will also generate a response. One response unit (RU) represents approximately 1 pg bound protein/ $\text{mm}^2$  (Van der Merwe, 2000). The actual binding response for an experiment is obtained by subtracting the background response on a reference flow cell from that on the main flow cell. The real time graph of response units with time is called a sensorgram (Figure 3.14).

When a molecule is fixed on the sensor chip surface, it is said to be immobilised. Immobilisation can be done either directly by covalent coupling, or indirectly via high-affinity capture (such as an antigen-antibody interaction). The ligand, defined in the context of a Biacore application, is the interaction partner



**Figure 3.14**

A Biacore sensorgram of response units versus time. A report point is a specific point at which the response level is averaged over a short time window. The response may be absolute (as determined by the detector) or relative (with respect to another specified report point). The bars below the curve indicate solutions that pass over the sensor surface. Equilibrium dissociation or association constants ( $K_D$  or  $K_A$ ) can be calculated from the association ( $k_{on}$ ) and dissociation ( $k_{off}$ ) phases of the sensorgram (Adapted from the [Biacore X100 handbook, 2008](#)).

immobilised to the chip surface. The analyte is the interaction partner in solution, passed over the ligand. Regeneration is the process of removing bound analyte and/ or ligand without damaging the ligand or capturing molecule (Figure 3.15). A summary of the coupling chemistry and techniques for immobilisation used in Biacore chips is presented in Tables 3.3 and 3.4.

To increase the amount of immobilisation possible, a 100-200 nm layer of carboxymethylated dextran matrix is attached to the general-purpose CM-family Biacore sensor chips (Table 3.2). This has two important drawbacks (Van der Merwe, 2000). Firstly, the rate at which binding occurs at the chip surface may exceed the rate at which mass transport occurs (i.e., the rate at which the binding molecule is delivered to the surface). A consequence of this is that the measured association rate constant ( $k_{on}$ ) is artificially slower than the true value. Secondly, following dissociation, binding can reoccur to unoccupied surface. This means that the measured dissociation rate constant ( $k_{off}$ ) may be artificially slower than its true value. Both problems are addressed by using high flow rates of at least 30  $\mu\text{l}/\text{min}$  for equilibrium experiments and limiting the amount of immobilised ligand to quantities sufficient to reach a maximum response ( $R_{\text{max}}$ ) of 100-200 RU. The required immobilisation level ( $R_L$ ) can be calculated from the equation:

$$\text{(Eq. 3.20)} \quad R_{\text{max}} = \frac{\text{Analyte MW} \times R_L \times S_m}{\text{Ligand MW}}$$

where  $S_m$  is the stoichiometric ratio of interaction.

### (3.5.3) Applications of Biacore systems

Biacore systems are generally unsuited to studying analytes that are very small ( $\text{MW} < 1 \text{ kDa}$ ) compared to ligand, as they give very small responses. Because high amounts of ligand need to be immobilised to achieve measurable responses, mass transport limitations and re-binding make it difficult to measure kinetic parameters for such interactions with confidence. Very large molecules such as viruses and cells should not be immobilised on the chip surface, but rather injected as sample. This is because detection decreases exponentially with distance away from the surface, therefore interactions at their distal ends may be outside the system's detection limit.

### (3.5. 3.1) Binding analysis

Binding analysis on the Biacore can be used to assess the activity of native or recombinant proteins. Since interactions with ligands involve multiple residues and require correct folding to occur, a “yes or no” binding assay can determine the structural integrity of a recombinant protein. Binding specificity, formation of multimolecular complexes and the effects of different experimental conditions (e.g. pH, ionic strength and temperature) on binding can also be assessed by binding analysis. Typically, the results from such studies are based on report point values (Figure 3.14) and not complete association or dissociation phases, and are evaluated via plots and bar charts.

### (3.5. 3.2) Equilibrium analysis

Using multiple sequential injections of analyte at different concentrations, affinity (binding strength) and kinetic (association and dissociation rate-) constants of a ligand-analyte interaction can be determined. Low affinity interactions are easier to study as regeneration is more readily achieved. However, for high affinity interactions for which regeneration is very difficult, single-cycle assays may be used. In these assays, the analyte is injected at increasing concentrations without regenerating the chip surface in-between. Dissociation constants can be determined from equilibrium experiments in two ways, (i) by steady state affinity measurements and (ii) from kinetic rate constants. The steady state affinity model within the Biacore X100 evaluation software (version 1.1) is set to a default stoichiometry of 1:1, but other ratios can be included by the user where necessary. This model calculates the equilibrium dissociation constant  $K_D$  from a plot of steady state binding levels ( $R_{eq}$ ) against analyte concentration ( $C$ ). The equation allows for the contribution of buffer to the signal, known as bulk refractive index (RI), to serve as a baseline offset to the  $R_{eq}$  axis.

$$(Eq. 3.21) \quad R_{eq} = \frac{CR_{max}}{K_D + C} + RI$$

An alternative approach is to export the data to suitable analysis software such as Sigmaplot™ to calculate  $K_D$  values, although this may not make sufficient allowance for the bulk refractive index. For  $K_D$  values determined from affinity measurements to

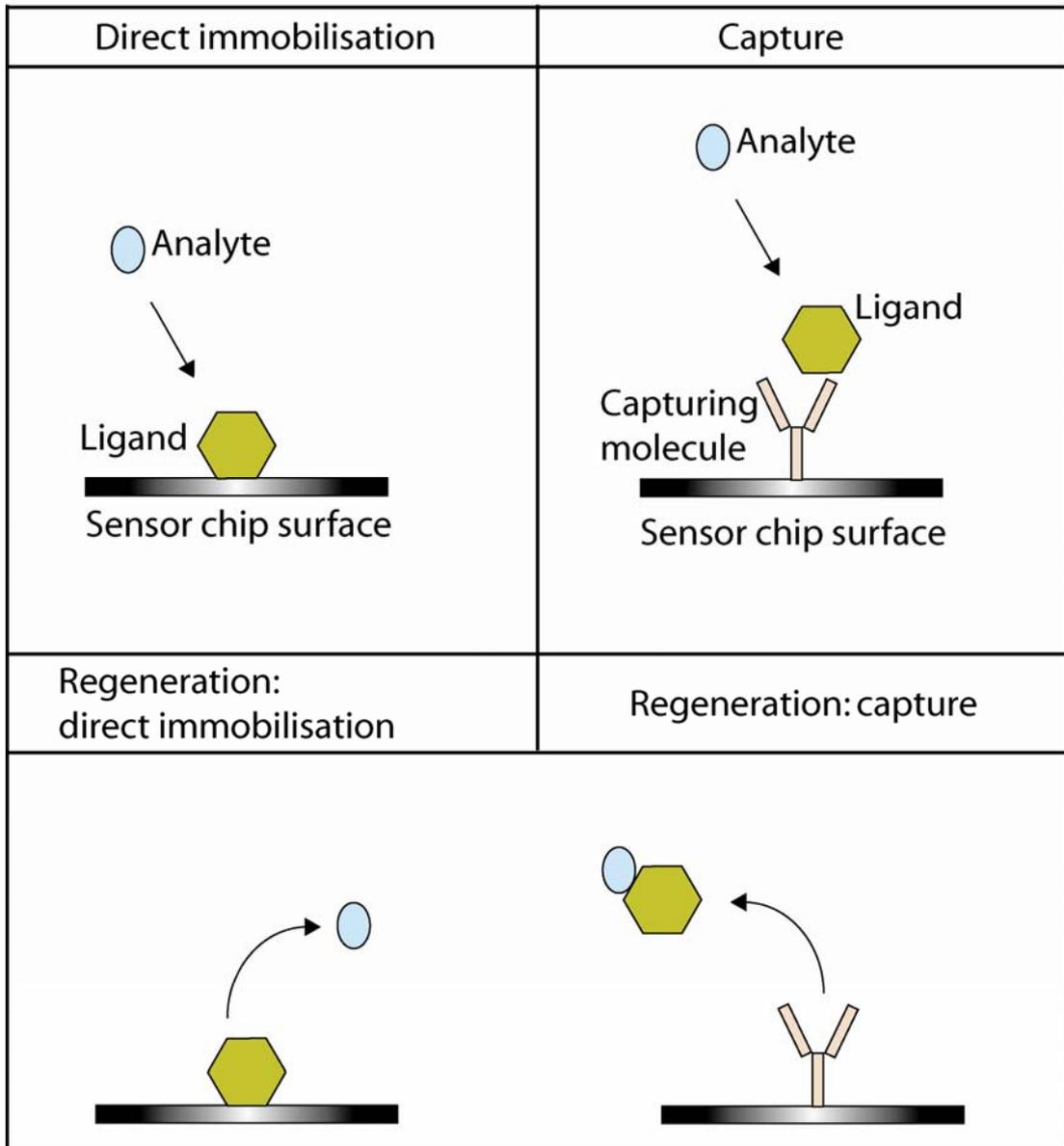
be accurate, majority of the multiple sensorgrams within the experiment must attain steady-state binding levels. Steady-state is defined by the response level at which the net association and dissociation rates are zero, and appears as a sustained plateau on the sensorgram.

$K_D$  can also be determined from calculating the kinetic constants of the interaction ( $K_D = k_{off}/k_{on}$ ; [Figure 3.14](#)). The key utility of kinetic analysis lies in the fact that although two interactions may possess the same affinity, the kinetic rate constants (and hence the nature) of these interactions may be entirely different. Thus kinetic analyses provide insight into molecular interactions beyond simple affinity measurements. The Biacore X100 evaluation software implements a number of predefined mathematical models for data interpretation. All of these models extract kinetic parameters from experimental data using an iterative process that finds the best fit for a set of equations describing the interaction. The simplest model is the 1:1 interaction, which fits for the parameters  $k_a$  (or  $k_{on}$ ; association rate constant in  $M^{-1}s^{-1}$ ),  $k_d$  (or  $k_{off}$ ; dissociation rate constant in  $s^{-1}$ ),  $R_{max}$ ,  $t_c$  (flow rate-independent aspect of the mass transfer constant) and RI ([Biacore X100 handbook, 2008](#)). Kinetic analyses are possible where steady state binding levels are not attained, provided the sensorgrams have sufficient curvature. This is especially useful for interactions in which it is difficult to attain steady-state binding levels. However, for interactions with very fast association and dissociation phases, kinetic analyses may not be possible. The X100 evaluation software (version 1.1) can accurately determine  $k_{on}$  values within the range of  $3 \times 10^2 M^{-1}s^{-1}$  to  $3 \times 10^7 M^{-1}s^{-1}$ , and  $k_{off}$  values between  $10^{-5} s^{-1}$  to  $10^{-1} s^{-1}$ . For interactions with  $k_{on}$  and  $k_{off}$  values outside these limits of quantitation, affinity measurements will provide more accurate values for  $K_D$ .

### **(3.5. 3.3) Concentration analysis**

This method establishes an unknown sample concentration by generating a calibration curve of known samples. This is especially useful for determining the concentration of a specific ingredient within a sample containing impurities, degradation products or matrix components. The principle of concentration measurement with Biacore is similar to that of established methods such as ELISA, with the following important differences: (i) there is no need to use labels for detection, (ii) the extent of interaction is measured directly, allowing the entire interaction process to be observed, and (iii) measurements can be done on a real-time

basis, as opposed to only end-point levels. Setting up a concentration assay involves five steps. These are: selecting an assay format (including choice of analyte, ligand and capture as applicable), preparing the sensor surface, establishing adequate regeneration conditions, determining a calibration curve using known analyte concentrations, and validating the assay by quantitating the relevant performance parameters. Performance parameters include specificity, selectivity and cross-reactivity, accuracy, precision, limit of detection (LOD), limit of quantitation (LOQ), linearity, range, robustness and sensitivity ([Biacore concentration analysis handbook, version AA, 2001](#)).



**Figure 3.15**

Illustration of direct and indirect immobilisation (capture) of ligand and the regeneration process in Biacore experiments (Adapted from the [Biacore X100 handbook, 2008](#)).

<b>Sensor chip type</b>	<b>Surface chemistry</b>	<b>Free carboxymethyl groups</b>	<b>Comments</b>
<b>CM5 (research grade or certified)</b>	Carboxymethylated dextran (CMD) matrix	Yes	Most widely used sensor chip, suitable for most applications.
<b>CM4</b>	Lower degree of carboxymethylation than CM5	Yes	Well suited for negatively charged molecules.
<b>CM3</b>	CMD matrix with shorter dextran chains	Yes	Useful for working with very large molecules.
<b>C1</b>	No dextran matrix, but is carboxymethylated.	Yes	Useful for multivalent interactions and large molecules. Also for interactions affected by the presence of dextran.
<b>L1</b>	Lipophilic groups attached to a CMD matrix.	No	Suitable for direct attachment of lipid vesicles to create a membrane-like environment.
<b>SA</b>	CMD matrix with streptavidin pre-coupled	No	Can be generated by coupling streptavidin to CM5 chip
<b>NTA</b>	CMD matrix with NTP pre-coupled	No	For capturing proteins with Histidine tags
<b>HPA</b>	Flat hydrophobic surface	No	For immobilising lipid monolayers.

**Table 3.2.**

A summary of the Sensor chips available for Biacore systems.

	Surface activation <sup>a</sup>	Group required on protein	Protein preparation on	Comments
<b>Amine</b>	EDC/NHS <sup>b</sup>	Amine groups (lysine unblocked termini)	No preparation needed	Can be used with most proteins. Risk of multiple coupling
<b>Surface-thiol</b>	EDC/NHS followed by cystamine/DTT to introduce a thiol group	Amine or carboxyl groups	Need to introduce reactive disulphide using heterobifunctional reagent such as PDEA (carboxyl) or SPDP (amine).	Not widely used
<b>Ligand-thiol</b>	EDC/NHS followed by PDEA to introduce a reactive disulphide group	Surface-exposed free cysteine or disulphide.	Reduce disulphides under non-denaturing conditions to generate free cysteine.	Useful for proteins with exposed disulphides or free cysteines.
<b>Aldehyde coupling</b>	EDC/NHS followed by hydrazine	Aldehyde groups	Create aldehydes by oxidising <i>cis</i> -diols (in sugars) with periodate. Works especially well with sialic acid.	Useful for polysaccharides and glycoproteins.

<sup>a</sup> All coupling reactions utilise carboxymethyl groups on the sensor surface and can be used with any sensor chip which has free carboxymethyl groups. The first step is to activate these groups with N-hydroxysuccinimide, thus creating a highly reactive succinimide ester which reacts with amine groups on protein, or can be further modified.

<sup>b</sup> EDC, N-ethyl-N'-(dimethylaminopropyl) carbodiimide; NHS, N-hydroxysuccinimide; DTT, dithiothreitol; PDEA, 2-(2-pyridinyldithio) ethane-amine.

### Table 3.3.

A summary of the covalent coupling chemistry for Biacore systems (Adapted from [Van der Merwe, 2000](#)).

	Covalently coupled molecule	Captured ligand	Valency of ligand	Removal of tag
<b>Antibodies</b>	Rabbit anti-mouse Fc polyclonal	Any mouse monoclonal	IgG Divalent	No
	Mouse anti-human Fc, monoclonal	Any protein fused to the Fc portion of human IgG <sub>1</sub> .	Divalent	Not routinely.
	Mouse anti-rat CD4, monoclonal	Any protein fused to rat CD4 domains 3 and 4.	Monovalent	No
	Anti-GST <sup>a</sup>	Any GST-fusion protein	Divalent	yes
<b>Other</b>	Streptavidin	Any biotinylated molecule.	Monovalent or tetravalent	No
	Ni-NTA	Any protein with oligohistidine tag	Monovalent	yes

<sup>a</sup> GST, glutathione-S-transferase, Ni-NTA, nickel-nitriloacetic acid.

**Table 3.4.**

A summary of techniques for ligand capture in Biacore systems (Adapted from [Van der Merwe, 2000](#)).

## **Chapter Four**

**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**

#### (4.1) Introduction

The SCR-1/5 fragment of FH is important for cofactor, C3b-binding and decay-accelerating activity (Gordon *et al.*, 1995; Sharma & Pangburn, 1996). The SCR-16/20 fragment comprises the major location for aHUS mutations that affect the ability of FH to regulate the C3 convertase, for which the distribution of over 100 genetic alterations are summarised at <http://www.fh-hus.org> (Section 2.4; Saunders *et al.*, 2007). The disease-related mutations in the N-terminal region appear to correspond to Type I quantitative reductions in plasma levels of FH, suggesting a distinct role for FH in this region (Saunders *et al.*, 2007). The aHUS related mutations cluster in SCR-20 and to a lesser degree in SCR-16 to SCR-19. As these mutations are generally located at surface-exposed positions and have normal FH plasma levels, they lead to Type II functional defects in FH (Saunders *et al.*, 2007). An analysis of the first 12 C-terminal aHUS mutations to be identified suggested the location of a heparin-binding site in SCR-20, (Perkins & Goodship, 2002) and this prediction was supported by subsequent functional and NMR structural studies (Joszi *et al.*, 2006; Herbert *et al.*, 2006). Hence molecular structures for the SCR-1/5 and SCR-16/20 fragments of FH will reveal key insights into the role of these regions in the overall functional activity of FH.

Methods such as solution scattering and electron microscopy have been applied to demonstrate that FH does not possess a fully-extended SCR structure in solution and may exist in monomeric and dimeric forms (Perkins *et al.*, 1991; DiScipio, 1992; Aslam & Perkins, 2001). NMR and crystal structures have been determined for SCR-1/3, SCR-5, SCR-6/8, SCR-15/16 and SCR-19/20, (Barlow *et al.*, 1992, 1993; Soares & Barlow 2005; Herbert *et al.*, 2006; Jokiranta *et al.*, 2006; Prosser *et al.*, 2007; Hocking *et al.*, 2008) while homology models for the remaining SCR domains have been described (Aslam & Perkins, 2001; Saunders *et al.*, 2006). However the orientation between two adjacent SCR domains is variable and not easily predicted (Perkins *et al.*, 2002). X-ray scattering and AUC in combination with constrained scattering modelling leads to medium resolution molecular structures (Perkins *et al.*, 1998; 2005; 2008). In this study, constrained scattering modelling is applied to the critical functional and recognition five SCR-domain fragments, SCR-1/5 and SCR-16/20. SCR-1/5 is shown to be a monomer in solution, unlike SCR-16/20 which exhibits a monomer-dimer equilibrium. The structural models are compared with the corresponding solution structures for rat Crry and  $\beta_2$ -

glycoprotein I, each of which contains five SCR domains (Aslam *et al.*, 2003; Hammel *et al.*, 2002). Interestingly, both SCR-1/5 and SCR-16/20 show more folded-back bent SCR arrangements in line with that seen in the solution structure of intact FH. The relevance of these solution structures in clarifying several aspects of FH function and the FH-associated diseases aHUS and AMD are discussed.

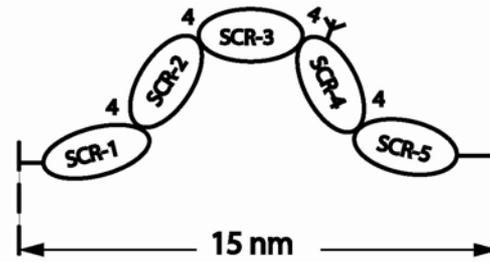
## **(4.2) Results and Discussion**

### **(4.2.1) X-ray scattering data for SCR-1/5 and SCR-16/20**

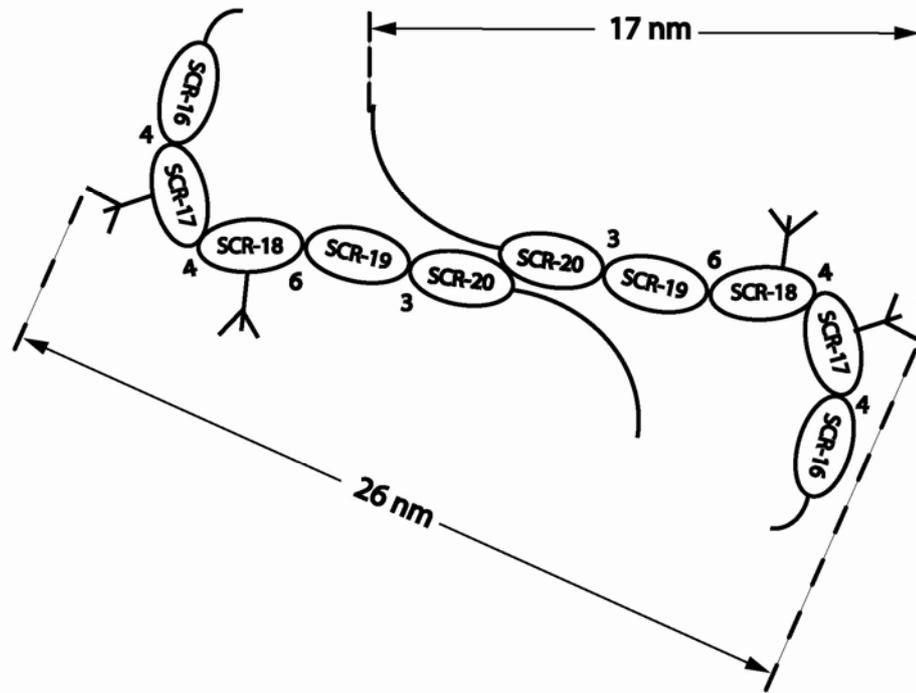
SCR-1/5 and SCR-16/20 were purified from yeast culture medium by virtue of their 6-His C-terminal peptides (Figure 4.1) using nickel affinity chromatography (Section 4.4). SCR-1/5 ran as a single homogeneous peak in size-exclusion gel filtration (Figure 4.2(a)) and as a single band on SDS-PAGE with an apparent molecular weight of 30 kDa under non-reducing conditions (Figure 4.2(c)). SCR-16/20 reproducibly eluted as a broader peak which was occasionally partly-resolved into two components in size-exclusion gel filtration (Figure 4.2(b)). The molecular weight standards based on globular proteins resulted in estimated molecular weights that were larger than the expected values of 35.3 kDa and 39.7 kDa for SCR-1/5 and SCR-16/20 respectively. As the latter possess extended protein structures, a logarithmic plot of the elution volumes showed that the apparent molecular weights of the two peaks for SCR-16/20 were in the approximate ratio of 2:1, suggesting that this existed in an equilibrium between dimeric and monomeric forms. SCR-16/20 ran as a major band on non-reducing SDS-PAGE with an apparent molecular weight of 37 kDa (Figure 4.2(c)). Both FH fragments are potentially glycosylated at one or two positions (Figure 4.1). Treatment with Endo H glycosidase showed no change for SCR-1/5 (Figure 4.2(c)), suggesting that, in agreement with previous work (Aslam & Perkins, 2001), the single site at SCR-4 is not occupied. Under these conditions, Endo H glycosidase resulted in a decrease of 2 kDa in the apparent molecular weight of SCR-16/20, suggesting that at least one of its two glycosylation sites is occupied (Figure 4.2(c)). Both purified proteins were stable upon storage at 4°C.

The domain arrangement and oligomerisation of SCR-1/5 was analysed by synchrotron X-ray scattering. In two distinct Q ranges (where  $Q = 4 \pi \sin \theta / \lambda$ ;  $2\theta$  = scattering angle;  $\lambda$  = wavelength), Guinier analyses to measure the degree of elongation of SCR-1/5 resulted in two linear plots, from which the overall radius of

(a) SCR-1/5



(b) SCR-16/20



(c) SCR-1/5 and SCR-16/20

RESIDUE NUMBERING	10	20	30	40	50
SCR-1	17- <b>EFED</b> QNELPPRRNTEILTGSWSDQT... <b>YPEG</b> TOAIYK <b>CRPGY</b> RLG... <b>NVIMV</b> ORK <b>GEW</b> VALNPLR <b>KG</b> -80				
SCR-2	81- <b>QKR</b> CGHP... <b>GDTP</b> FGT <b>FTLT</b> GGNV... <b>FEY</b> GVKAVY <b>TON</b> EGYQLLGE... <b>IN</b> YR <b>CDT</b> . <b>DG</b> WTNDI... <b>PTC</b> -141				
SCR-3	142- <b>EVV</b> RLPV... <b>TAP</b> ENKIVSSAME <b>PDREY</b> HFG <b>QAVR</b> FV <b>ONS</b> GYKIEG... <b>DEEM</b> CSDD <b>GF</b> NSKEK... <b>PK</b> -205				
SCR-4	206- <b>VEI</b> SKSP... <b>DV</b> INGSPISQKII... <b>YK</b> ENERFQY <b>KON</b> MGY <b>EY</b> SE... <b>RG</b> DAV <b>CTE</b> . <b>SG</b> WRPL... <b>PS</b> -262				
SCR-5	263- <b>EES</b> SDNP... <b>YI</b> PNGDY <b>SPL</b> RIK... <b>HRT</b> GDEI <b>TYO</b> CRNGFY <b>PA</b> TR... <b>GNT</b> AK <b>CTS</b> . <b>TG</b> WIPA... <b>PR</b> -320				
DSSP	EE-----EEEE-----EEEE-----EEEE-----EEEE-----EEE				
Beta-strands	b1-----b2-----b3-----b4-----b5-----b6-----b7-----b8-				
C-terminus	321- <b>FL</b> HHHHHH-328				
SCR-16	922- <b>EAEAE</b> FGL <b>CK</b> SP. <b>PEI</b> SHGVVAHMSDS... <b>YQY</b> GEEV <b>TYK</b> CFEG <b>FG</b> IDG... <b>PAI</b> AK <b>CLG</b> . <b>ER</b> WSHP... <b>PS</b> -984				
SCR-17	985- <b>IKT</b> DL <b>SL</b> . <b>PS</b> FENAI <b>PM</b> GEKKDV... <b>YK</b> AGEQV <b>TYT</b> CA <b>TY</b> YKMDG... <b>AS</b> NVICIN. <b>SR</b> WTGR... <b>PT</b> -1043				
SCR-18	1044- <b>RD</b> TSG <b>V</b> NP. <b>PT</b> VQ <b>NA</b> I <b>VS</b> RQMSK... <b>Y</b> PSGER <b>VR</b> Y <b>CR</b> SP <b>Y</b> EMFG... <b>DE</b> EVN <b>CLN</b> . <b>GN</b> WTEP... <b>PG</b> -1102				
SCR-19	1103- <b>KD</b> STG <b>CG</b> PPPP <b>ID</b> NGDI <b>TS</b> FPLSV... <b>Y</b> APASS <b>VEY</b> Q <b>CQ</b> NLY <b>Q</b> LEG... <b>NK</b> RIT <b>CRN</b> . <b>GO</b> WSEP... <b>PK</b> -1163				
SCR-20	1164- <b>LH</b> CVISREIMEN <b>YN</b> I <b>AL</b> RWTAK <b>Q</b> KLY <b>SRT</b> GES <b>VE</b> FV <b>CR</b> GY <b>RL</b> SSRS <b>HTL</b> RT <b>IC</b> WD. <b>GK</b> LE. <b>Y</b> ... <b>PT</b> -1228				
DSSP	EE-----EEEE-----EE-----EEEEEE-----EEE-----EEEEEE-----EEE				
Beta-strands	b1-----b2-----b3-----b4-----b5-----b6-----b7-----b8-				
C-terminus	1229- <b>AKR</b> PLE <b>QK</b> LISEED <b>LNS</b> AVD <b>HH</b> HHHH-1254				

Figure 4.1

(Figure legend overleaf)

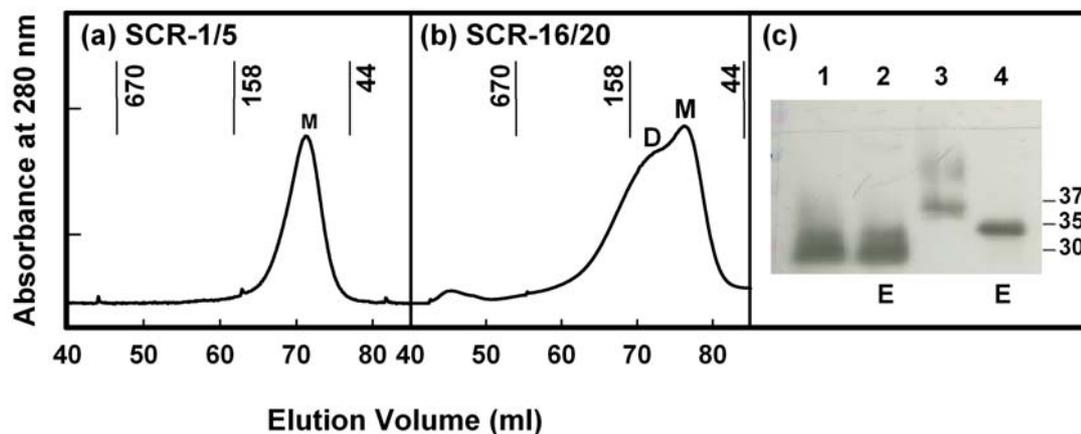
(Figure 4.1 legend)

Schematic cartoon of the FH SCR-1/5 and SCR-16/20 fragments and their sequences

(a) SCR-1/5 is shown together with the residue lengths of the four inter-SCR linkers and its overall length determined by X-ray scattering (below). A putative glycosylation site on SCR-4 is shown, but it is not occupied in the recombinant fragment

(b) A possible SCR-16/20 dimer is shown if it is assumed that the SCR-20 domains are in contact (see text). The putative positions of N-linked carbohydrates on SCR-17 and SCR-18 are highlighted, of which at least one is occupied

(c) The sequences of the recombinant SCR-1/5 and SCR-16/20 proteins are shown, with the five conserved Trp and Cys residues highlighted in black and other conserved residues in yellow. The putative carbohydrate sites are underlined, and the linker residues are boxed. The Define Secondary Structure Prediction-assigned  $\beta$ -strand secondary structure for the ten domains are highlighted in grey, and the position and labelling of the eight consensus  $\beta$ -strands b1 to b8 are shown below the alignment. In SCR-1/5, residues 17-EF-18 correspond to the EcoRI cloning site, and residues 323-HHHHHH-328 correspond to the 6-His tag used for purification. In SCR-16/20, residues 922-EAEAEF-927 correspond to the  $\alpha$ -factor signal and EcoRI site, and residues 1232-PLEQKLISEEDLNSAVDHHHHHH-1254 correspond to the myc epitope and the 6-His tag.



**Figure 4.2**

Size-exclusion gel filtration and SDS-PAGE of SCR-1/5 and SCR-16/20. The eluate from the nickel affinity column was loaded onto a Superdex-200 prep grade XK 16/60 column

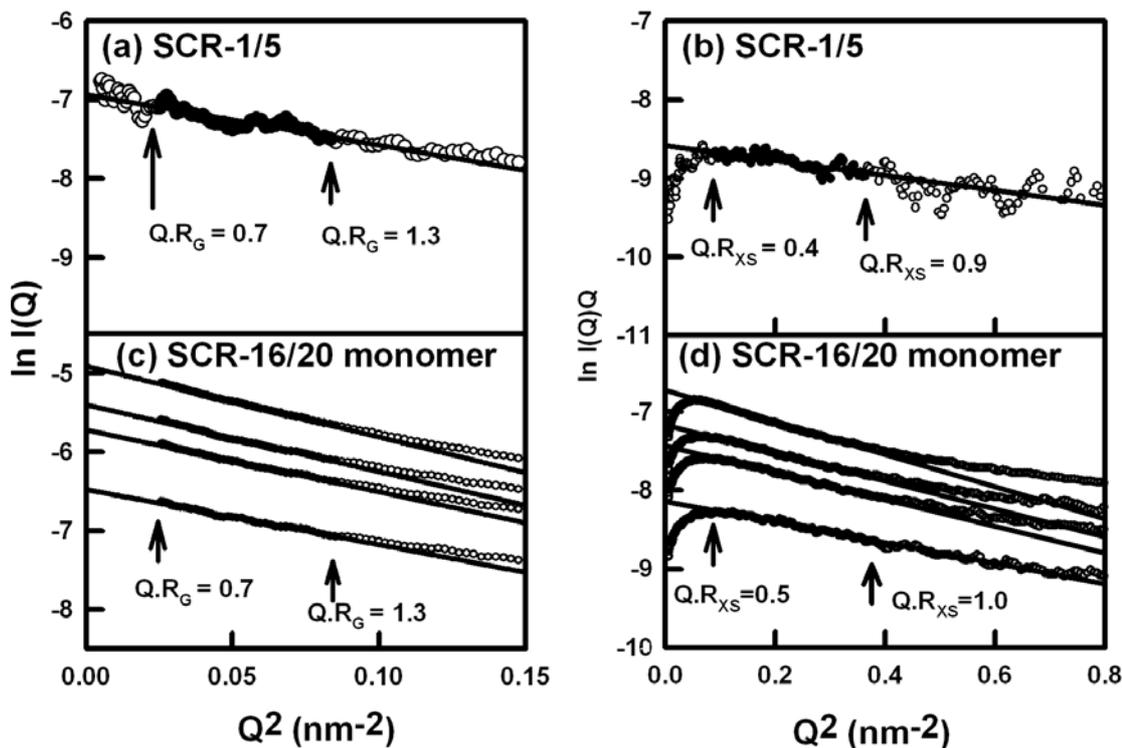
(a) SCR-1/5 eluted as a monomeric peak M. The elution positions of three molecular mass standards are shown

(b) SCR-16/20 eluted as a peak that was partly resolved into two components designated as monomer (M) and dimer (D). The recalibrated elution positions of three molecular mass standards are shown

(c) Non-reducing SDS-PAGE analyses of SCR-1/5 and SCR-16/20 treated with Endo H (lanes 2 and 4; labelled with E) or not treated (lanes 1 and 3) are shown, as carried out by Dr. R. J. Ormsby (Australia).

gyration ( $R_G$ ) and the radius of gyration of the cross-section ( $R_{XS}$ ) values were obtained within satisfactory  $Q.R_G$  and  $Q.R_{XS}$  limits (Figures 4.3(a,b)). Time frame analyses during data acquisition showed that the  $R_G$  values were affected by slight radiation damage only by the longest exposure times of 2 sec per frame in 4-bunch mode. To eliminate radiation damage effects, data were therefore collected at exposures of 1.0 sec, 0.5 sec and 0.2 sec in a flow cell for ten frames, of which only the first time frame was used for Guinier analyses. No concentration dependence of the Guinier parameters was seen between 0.21 mg/ml to 0.84 mg/ml. However at concentrations above 0.21 mg/ml, the Guinier  $R_G$  fits showed that slight non-specific aggregates were present, and this was subsequently confirmed by molecular modelling (see below). Accordingly the final  $R_G$  determination was performed at 0.21 mg/ml. The mean X-ray  $R_G$  and  $R_{XS}$  values for SCR-1/5 were 4.26 ( $\pm$  0.16) nm (3 values) and 1.46 ( $\pm$  0.08) nm (10 values) respectively (Table 4.1).

Compared to rat Crry, a five-SCR domain complement regulatory protein related to complement receptor type 1 with  $R_G$  and  $R_{XS}$  value of 5.0 ( $\pm$  0.4) nm and 1.5 ( $\pm$  0.1) nm respectively, (Aslam *et al.*, 2003) these  $R_G$  and  $R_{XS}$  values indicated that SCR-1/5 showed a slightly less elongated structure than Crry. SCR-1/5 has a similar degree of domain bend or partial folding back when compared to human  $\beta_2$ -glycoprotein I with an  $R_G$  value of 4.30 nm which has an S-shaped arrangement of five SCR domains in solution (Hammel *et al.*, 2002). The anisotropy ratio  $R_G/R_O$  (where  $R_O$  is the  $R_G$  value of the sphere with the same volume as the hydrated glycoprotein) was 2.32 and this is intermediate between that of 2.54 for rat Crry and that of 2.12 for human  $\beta_2$ -glycoprotein I. Hence SCR-1/5 is more elongated than a globular protein, for which the typical  $R_G/R_O$  is 1.28 (Perkins, 1988b). Assuming that SCR-1/5 can be represented by an elliptical cylinder in shape, the combination of the  $R_G$  and  $R_{XS}$  values resulted in an overall length of 14.2 nm for SCR-1/5 (Section 4.4). The corresponding Guinier analyses for SCR-16/20 gave  $R_G$  and  $R_{XS}$  values within satisfactory  $Q.R_G$  and  $Q.R_{XS}$  limits (Figures 4.3(c,d); 4.4(c,d)). The individual time frames during data acquisition showed that the  $R_G$  values were not affected by radiation damage, even with the longest exposure times of 2 sec per frame in 4-bunch mode, and 0.2 sec per frame in uniform bunch mode. Accordingly, data were collected at exposures of mostly 1.0, 0.5, 0.2 and 0.1 sec in 4-bunch mode and for 0.2 and 0.1 sec in uniform bunch mode. Analyses were based on the average of all ten



**Figure 4.3**

Guinier  $R_G$  and  $R_{XS}$  analyses for SCR-1/5 and SCR-16/20 monomers. The filled circles correspond to the  $I(Q)$  values used to determine  $R_G$  and  $R_{XS}$ , while the straight lines correspond to the line of best fit through these points. The  $Q.R_G$  and  $Q.R_{XS}$  values for the fitting ranges are arrowed

(a)  $R_G$  plot for SCR-1/5 at a concentration of 0.21 mg/ml using a  $Q$  range of 0.16-0.29  $\text{nm}^{-1}$

(b)  $R_{XS}$  plot for SCR-1/5 at 0.21 mg/ml using a  $Q$  range of 0.30-0.60  $\text{nm}^{-1}$

(c)  $R_G$  plots for principally monomeric SCR-16/20 from 0.14 mg/ml to 0.62 mg/ml using a  $Q$  range of 0.16-0.29  $\text{nm}^{-1}$

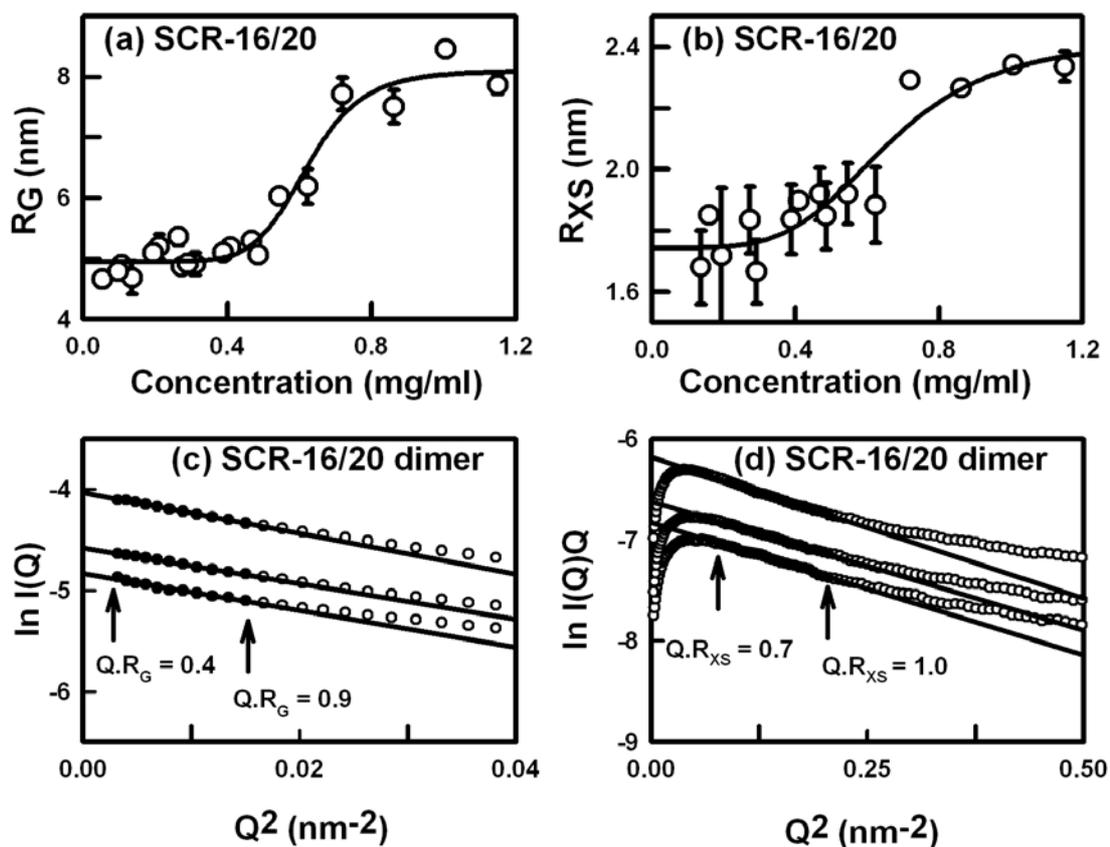
(d) The corresponding  $R_{XS}$  plots for SCR-16/20 in a  $Q$  range of 0.30-0.60  $\text{nm}^{-1}$

1.

time frames. In distinction to SCR-1/5, SCR-16/20 exhibited a concentration dependence of both its  $R_G$  and  $R_{XS}$  values (Figure 4.4(a,b)). At the lowest concentrations, the Guinier fits showed slight non-specific aggregation. At the highest concentrations, the fits were consistent with weak dimer formation. Accordingly the  $Q$  ranges used to determine the  $R_G$  and  $R_{XS}$  values were set to exclude the lowest  $Q$  values at the lowest concentrations to avoid the slight aggregates (Figure 4.3(c)), and to exclude the larger  $Q$  values at the highest concentrations in order to accommodate the larger  $R_G$  and  $R_{XS}$  values (Figure 4.4(c,d)). The Guinier  $I(0)/c$  values doubled from 0.0072 ( $\pm$  0.0002) at low concentrations to 0.0147 ( $\pm$  0.0005) at high concentrations, indicating that a monomer-dimer equilibrium has been observed.

Over a concentration range of 0.05 mg/ml to 1.15 mg/ml, the  $R_G$  values for SCR-16/20 ranged from 4.66 ( $\pm$  0.13) nm to 8.46 ( $\pm$  0.12) nm. A two-state regression fit gave a value of 4.95 ( $\pm$  0.08) nm for the monomer and an estimate of 8.09 ( $\pm$  0.21) nm for the dimer (20 values), noting that the dimer cannot be fully formed for reason of its weak  $K_D$  value (Figure 4.4(a)). A two-state regression fit of the  $R_{XS}$  values ranged from 1.74 ( $\pm$  0.04) nm to 2.42 ( $\pm$  0.14) nm (18 values) (Figure 4.4(b)). Assuming that SCR-16/20 can be represented by an elliptical cylinder in shape, the combination of the  $R_G$  and  $R_{XS}$  values resulted in overall lengths of 15 nm and 27 nm respectively for the monomer and dimer. On the basis of a monomer-dimer equilibrium, the regression analyses of the  $R_G$  and  $R_{XS}$  values resulted in estimated  $K_D$  values of 16 ( $\pm$  2)  $\mu$ M and 17 ( $\pm$  4)  $\mu$ M respectively. The  $R_G/R_O$  ratio of 2.40 for monomeric SCR-16/20 is similar to that of 2.32 for SCR-1/5, suggesting that both FH fragments have similar extended structures, even though at least one oligosaccharide chain is present on SCR-16.20 (Figure 4.2(c)). However the estimated  $R_G/R_O$  ratio of 3.30 for the dimer shows this is significantly more elongated than the monomer. Interestingly the latter value is comparable to that for intact FH SCR-1/20 which is 3.65 (Aslam & Perkins, 2001).

The distance distribution function curve  $P(r)$  provides structural information in real space (Figure 4.5). The maximum dimension  $L$  is determined from  $P(r)$  as the value of  $r$  when  $P(r) = 0$  at large  $r$  and  $L$  is now independent of shape assumptions. The  $P(r)$  curves provide an alternative calculation of the mean  $R_G$  value. For SCR-1/5 this  $R_G$  was similar at 4.42 ( $\pm$  0.13) nm (3 values) to that from the Guinier analyses but is shorter than  $L$  of 18 nm for rat Crry (Aslam *et al.*, 2003). The average



**Figure 4.4**

Guinier  $R_G$  and  $R_{XS}$  analyses for SCR-16/20 dimer

(a,b) Concentration dependence of the  $R_G$  and  $R_{XS}$  values of monomeric SCR-16/20. The data were fitted to an arbitrary four-parameter logistic function of the form  $y = y_0 + (a / (1 + (x/x_0)^b))^{-1}$  from which the monomer values are well determined but the dimer values are not well determined because of the weak association. Statistical error bars are shown where visible

(c)  $R_G$  plots for principally dimeric SCR-16/20 at 0.72 mg/ml, 0.86 mg/ml and 1.15 mg/ml, using a  $Q$  range of 0.05-0.12  $\text{nm}^{-1}$

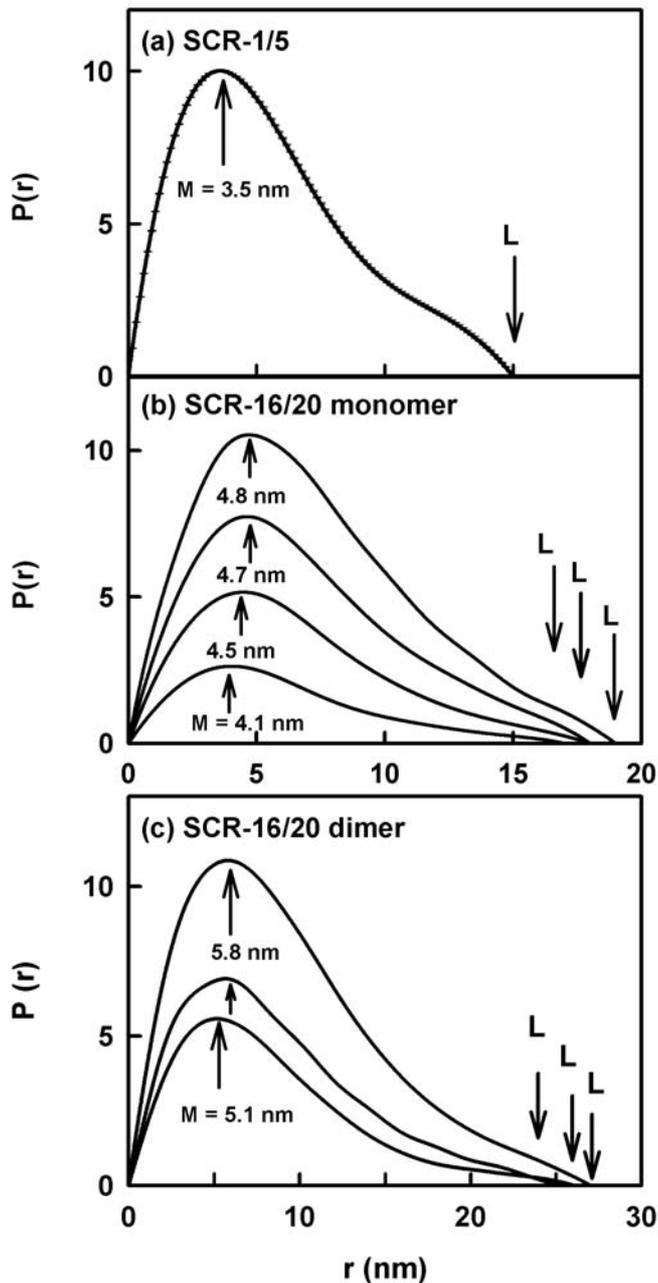
(d) The corresponding  $R_{XS}$  plots for dimeric SCR-16/20 using a  $Q$  range of 0.30-0.45  $\text{nm}^{-1}$ .

length of a single SCR domain is 3.6 nm along its longest axis, (Aslam *et al.*, 2003) thus a maximally-extended set of five SCR domains would be at least 18 nm in length. Hence the  $L$  value for SCR-1/5 corresponds well with a partly folded-back bent arrangement. The maximum in the  $P(r)$  curve corresponds to the most frequently occurring interatomic distance within the structure. A single peak  $M$  was observed at an  $r$  value of 3.5 nm at 0.21 mg/ml (Figure 4.5(a)). In comparison, a single peak was observed at an  $r$  value of 2.7 nm for rat Crry, and this increase in  $r$  value for SCR-1/5 is in agreement with the greater degree of bend in its structure compared to that in Crry (Furtado *et al.*, 2007).

In contrast to SCR-1/5, the  $P(r)$  curve for SCR-16/20 showed a concentration dependence of  $L$ . The  $R_G$  value was determined to be 4.81 ( $\pm$  0.11) nm at the lowest concentrations (Figure 4.5(b)) and as 7.51 ( $\pm$  0.24) nm at the highest concentrations (Figure 4.5(c)), both of which agree well with the corresponding Guinier values of 4.66 ( $\pm$  0.13) nm and 8.09  $\pm$  0.21 nm respectively. At the lowest concentrations,  $L$  was 17 nm for the monomer (Figure 4.5(b)), and at the highest concentrations,  $L$  was 27 nm when dimer was present (Figure 4.5(c)), both of which are comparable with the estimates of 15 nm and 27 nm from the Guinier analyses. The  $L$  value of 17 nm for the SCR-16/20 monomer shows that this is longer than SCR-1/5 ( $L$  value of 15 nm), and this is most readily explained by the presence of a 26-residue C-terminal tail in SCR-16/20 in place of an eight-residue C-terminal tail in SCR-1/5. The increase in  $L$  with concentration reflects the formation of a more elongated structure, in agreement with the increase in the  $R_G/R_O$  value. A single peak maximum  $M$  was observed at all concentrations of SCR-16/20, and its  $r$  values increased from 4.1 nm to 5.8 nm as the concentration was increased. Together with the large increase in the  $R_{XS}$  values with concentration, this shows that the SCR-16/20 dimer is formed by a partial side-by-side association of two monomers that increased the apparent folding back of the overall SCR domain arrangement.

#### (4.2.2) Analytical ultracentrifugation data for SCR-1/5 and SCR-16/20

In sedimentation velocity experiments by AUC, the sedimentation coefficient  $s_{20,w}^0$  provides an independent measure of macromolecular elongation and oligomerisation. For SCR-1/5, sedimentation velocity experiments were performed at two rotor speeds at six concentrations between 0.04 mg/ml and 0.85 mg/ml. The analyses of 10 to 30 scans using DCDT+ time-derivative analyses



**Figure 4.5**

Distance distribution function  $P(r)$  analyses for SCR-1/5 and SCR-16/20

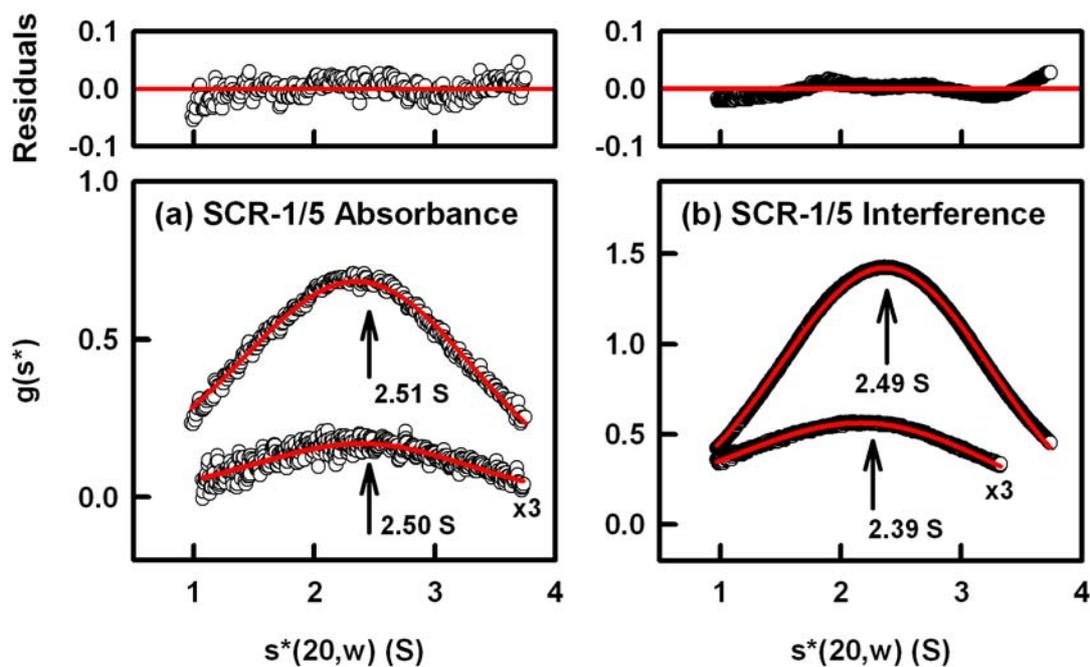
(a) For SCR-1/5,  $M$  represents the most frequent distance within the protein and is 3.5 nm at 0.21 mg/ml, and  $L$  represents its maximum dimension at 15 nm.

(b) For SCR-16/20 at lower concentrations,  $M$  rises from 4.1 nm to 4.8 nm from 0.14 mg/ml to 0.55 mg/ml, and  $L$  increases from 17 nm to 19 nm

(c) For primarily dimeric SCR-16/20 at 0.72 mg/ml, 0.86 mg/ml and 1.15 mg/ml,  $M$  ranges from 5.1 nm to 5.8 nm and  $L$  ranges from 25 nm to 27 nm.

gave excellent fits with small residuals to Gaussian functions (Figure 4.6). The use of Lamm equation fits of up to 200 scans using SEDFIT with the same data sets was used to assess size distribution analyses  $c(s)$  of any species present in the samples. Both the interference and absorbance fits resulted in a single clean peak with an  $s_{20,w}$  value of 2.4 S (Figure 4.7). The conversion of the  $c(s)$  plots to  $c(M)$  plots showed that this peak corresponded to a molecular weight of  $38 (\pm 6)$  kDa, in good agreement with the sequence-derived value of 35.3 kDa. From the DCDT+ analyses, both the absorbance and interference fits showed no concentration dependence of the  $s_{20,w}$  values, and this was determined to be  $2.52 (\pm 0.02)$  S (Figure 4.8(a)). The SEDFIT analyses also showed no concentration dependence (Figure 4.8(b)). Thus the final  $s_{20,w}^0$  values extrapolated to zero concentration for DCDT+ and SEDFIT were in good agreement at 2.52 S and 2.37 S respectively. From these, the frictional coefficient ratio  $f/f_o$  for SCR-1/5 is 1.45, which is larger than the value of 1.2 typically seen for globular-shaped proteins.

For SCR-16/20, the corresponding sedimentation velocity experiments confirmed the presence of a monomer-dimer equilibrium. Data were collected at five different rotor speeds at 20 concentrations between 0.01 mg/ml and 1.48 mg/ml (Materials & Methods). In the DCDT+ time-derivative analyses, the absorbance and interference fits for SCR-16/20 both showed a concentration dependence of the  $g(s^*)$  curves (Figure 4.9(a,b)). This is consistent with a monomer-dimer equilibrium (Stafford, 2000). Extrapolation to zero concentration gave a DCDT+ sedimentation coefficient  $s_{20,w}^0$  value of  $2.78 \pm 0.30$  S (Figure 4.8(a)). In the SEDFIT  $c(s)$  distribution analyses, both the absorbance and interference fits showed that two peaks could be resolved at all concentrations, more clearly so for the more detailed data sets from the interference optics and at the lowest concentrations (Figure 4.10(a,b)). Even with the saturation of the absorbance optics in Figure 4.10(b), two peaks were seen. The conversion of the  $c(s)$  plots to  $c(M)$  plots showed that the two peaks correspond to molecular weights of  $42 \pm 7$  kDa and (depending on the resolution) 60-80 kDa, showing that these peaks correspond respectively to the monomeric and dimeric forms of SCR-16/20. The mean values of these two peaks gave a sedimentation coefficient  $s_{20,w}^0$  value of  $2.87 \pm 0.20$  S for the monomer and  $3.92 \pm 0.23$  S for the dimer (20 values) (Figure 4.8(b)). From these  $s_{20,w}^0$  values, the  $f/f_o$  value is 1.35 for the monomer and 1.58 for the dimer. The first value of 1.35 is similar to that of 1.45 for SCR-1/5, suggesting that both monomeric SCR-16/20 and

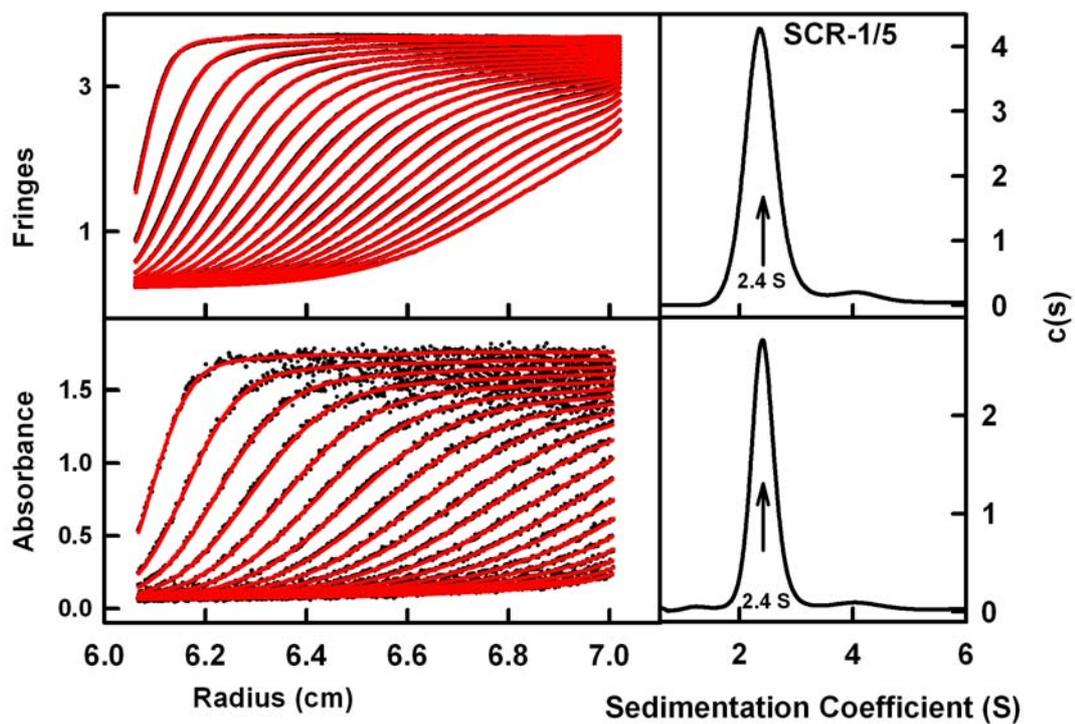


**Figure 4.6**

Sedimentation velocity  $g(s^*)$  analyses of SCR-1/5 using DCDT+.

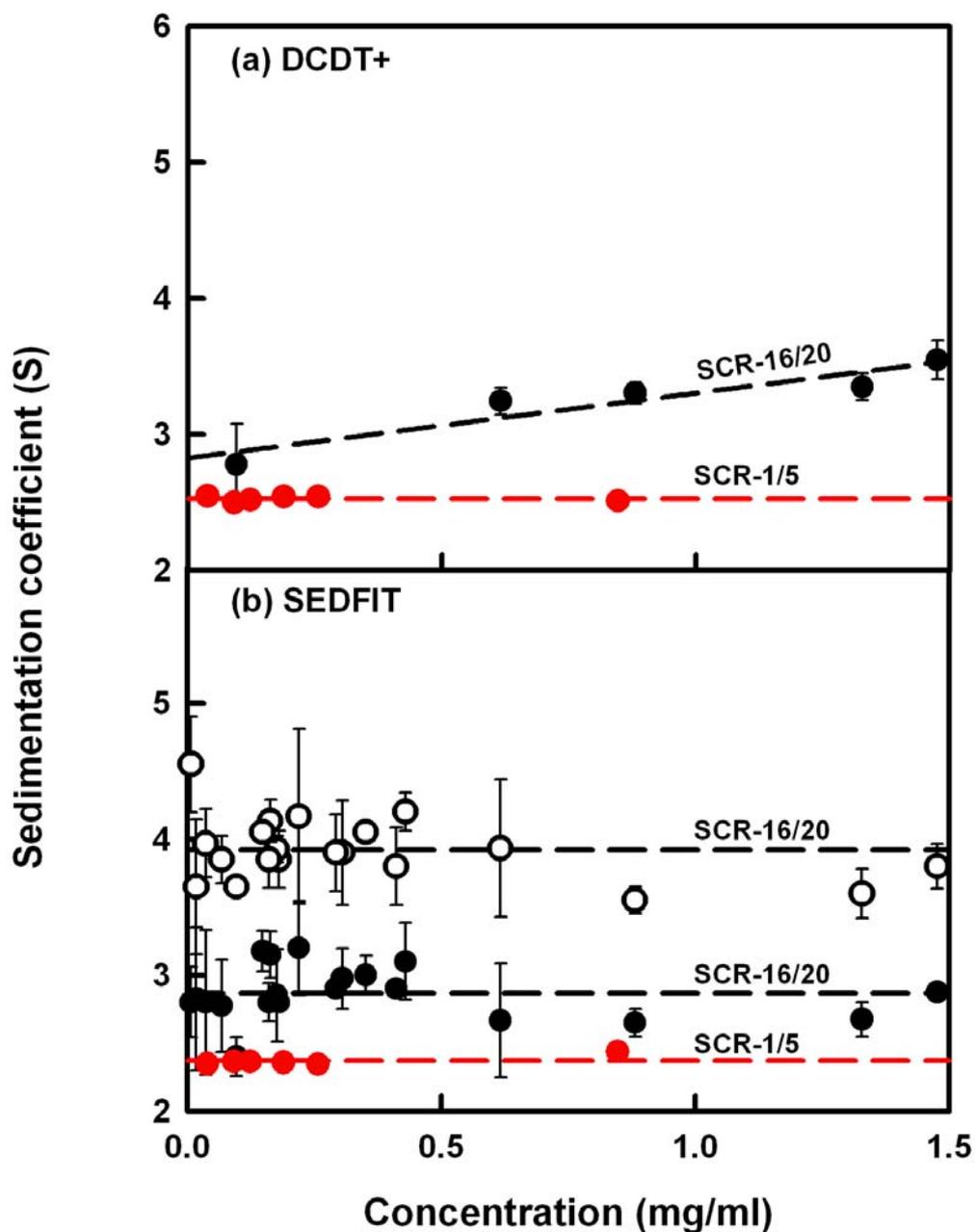
(a) Twenty absorbance scans for SCR-1/5 at 0.09 mg/ml and 0.85 mg/ml and a rotor speed of 35,000 r.p.m. gave  $s_{20,w}^{\circ}$  values of 2.50 S and 2.51 S respectively. The intensity of the  $g(s^*)$  curve at 0.09 mg/ml was increased three-fold for reason of clarity.

(b) The corresponding 14 and 22 interference scans from the same SCR-1/5 experiment gave  $s_{20,w}^{\circ}$  values of 2.39 S and 2.49 S respectively.



**Figure 4.7**

Sedimentation velocity  $c(s)$  distribution analyses of SCR-1/5 using SEDFIT. Every tenth scan in the total of 200 scans is shown in the boundary fits in the left panels, while the  $c(s)$  distributions are shown in the right panels. Fits at 0.85 mg/ml and 35,000 r.p.m. gave a  $s_{20,w}$  value of 2.4 S for both the interference and absorbance data sets (upper and lower panels respectively).

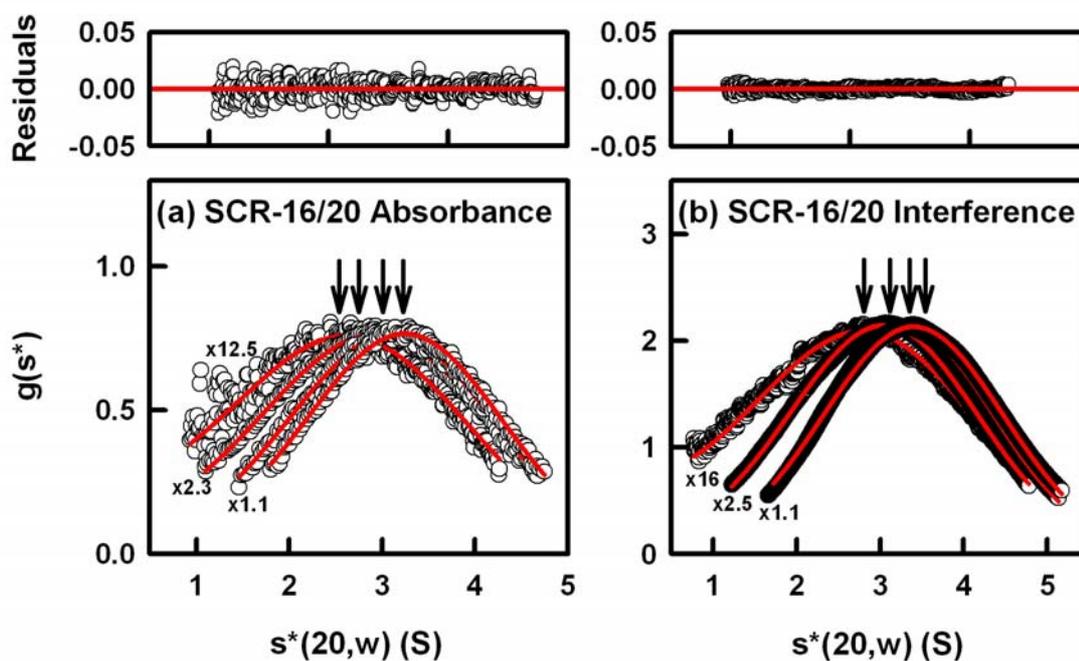


**Figure 4.8**

Concentration dependence of the observed  $s_{20,w}$  values for SCR-1/5 and SCR-16/20. Statistical error bars in these are shown where visible.

(a) The DCDT+ analyses of 14-22 scans per experiment showed single peaks, where the  $s_{20,w}$  values for SCR-1/5 are shown in red and those for part of the SCR-16/20 data is shown in black

(b) The SEDFIT analyses of 200 scans per experiment showed single peaks for SCR-1/5 and two peaks for SCR-16/20 (●, monomer; ○, dimer).

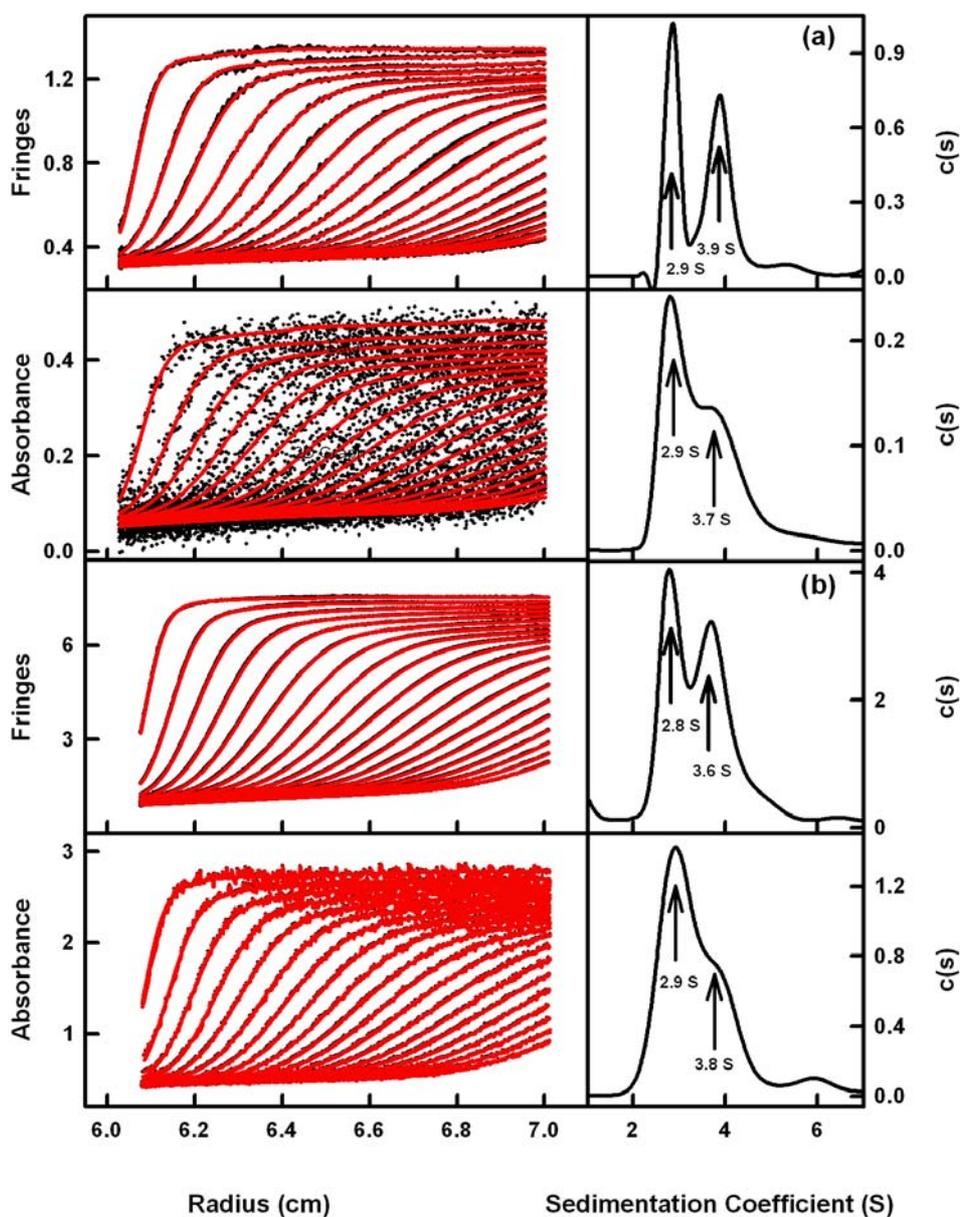


**Figure 4.9**

Sedimentation velocity  $g(s^*)$  analyses of SCR-16/20 using DCDT+.

(a) Absorbance data for SCR-16/20 at 0.10 mg/ml (40,000 r.p.m.) and 0.62 mg/ml, 1.33 mg/ml and 1.48 mg/ml (50,000 r.p.m.) showed that the  $s_{20,w}^o$  values increased from 2.8 S to 3.5 S as the concentration increased. The intensities of the  $g(s^*)$  curves were equalised for reason of clarity

(b) The corresponding interference data sets for SCR-16/20 showed that the  $s_{20,w}^o$  values likewise increased from 2.7 S to 3.5 S as the concentration increased.



**Figure 4.10**

Sedimentation velocity  $c(s)$  distribution analyses of SCR-16/20 using SEDFIT.

(a) The fits for monomeric SCR-16/20 at 0.22 mg/ml and 42,000 r.p.m. gave two peaks at  $s_{20,w}$  values of 2.9 S and 3.9 S for the interference data and 2.9 S and 3.7 S for the absorbance data (upper and lower panels respectively).

(b) The fits for dimeric SCR-16/20 at 1.48 mg/ml and 50,000 r.p.m. gave two peaks at  $s_{20,w}$  values of 2.8 S and 3.6 S for the interference data and 2.9 S and 3.8 S for the absorbance data (upper and lower panels respectively). Note that the absorbance optics are saturated as the absorbance is in excess of 2, nonetheless two peaks are observed.

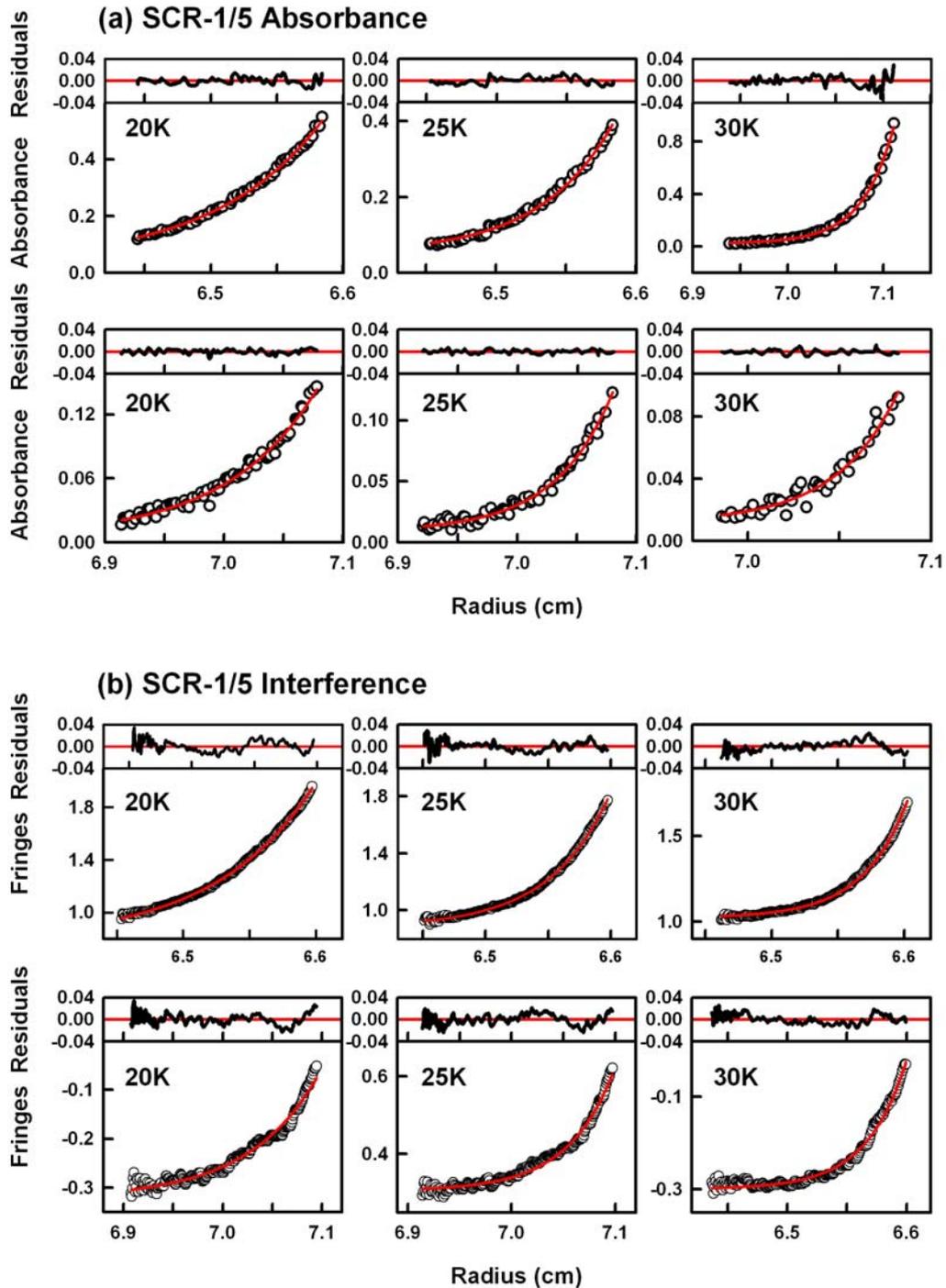
SCR-1/5 have similar degrees of extended shapes. This outcome is consistent with the  $R_G/R_O$  ratios above.

Sedimentation equilibrium experiments resulted in molecular weight ( $MW$ ) determinations for SCR-1/5 and SCR-16/20. For reason of protein stability, the most successful data collection was performed at 4 °C. For SCR-1/5, the  $MW$  was determined using a global fit of 15 curves in SEDPHAT at concentrations between 0.06 mg/ml to 0.39 mg/ml on the assumption of a single species model with both absorbance and interference data. This gave  $MW$  values of  $34.3 \pm 0.4$  kDa and  $36.9 \pm 0.2$  kDa respectively, in good agreement with the expected  $MW$  value of 35.3 kDa (Section 4.4). The analyses in Figure 4.11(a,b) show good visual fits with low random residuals. No concentration dependence was observed when data from each concentration was fitted globally at three speeds (not shown). Hence SCR-1/5 was concluded to be monomeric.

For SCR-16/20, the equilibrium studies using six different rotor speeds with absorbance data also gave good individual fits with random residuals on the assumption of a monomer-dimer association model. A global fit of 18 data sets at eight different concentrations between 0.13 mg/ml to 0.43 mg/ml and five rotor speeds using SEDPHAT gave  $MW$  values of  $40.8 \pm 0.3$  kDa and  $42.0 \pm 0.8$  kDa for the absorbance and interference data respectively, in good agreement with the expected  $MW$  of 39.7 kDa (Section 4.4). Even with the lower concentrations compared to Figure 4.4(a,b), the SEDPHAT analyses resulted in estimated similar weak  $K_D$  values of  $18 \pm 2$   $\mu$ M and  $13 \pm 2$   $\mu$ M respectively from the absorbance and interference data for the monomer-dimer equilibrium at 4 °C, in good agreement with the two Guinier-determined values of  $16 \pm 2$   $\mu$ M and  $17 \pm 4$   $\mu$ M at 22 °C (Figure 4.4(a,b)). The representative fits from the global analyses shown in Figure 4.12(a,b) show good visual agreements with low random residuals. The degree of curvature of the fits in Figure 4.12 at 20k r.p.m. is similar to that shown in Figure 4.11 at 25k r.p.m., indicating that the molecular weight of SCR-16/20 is larger than that of SCR-1/5. Hence SCR-16/20 exists in both monomeric and dimeric forms.

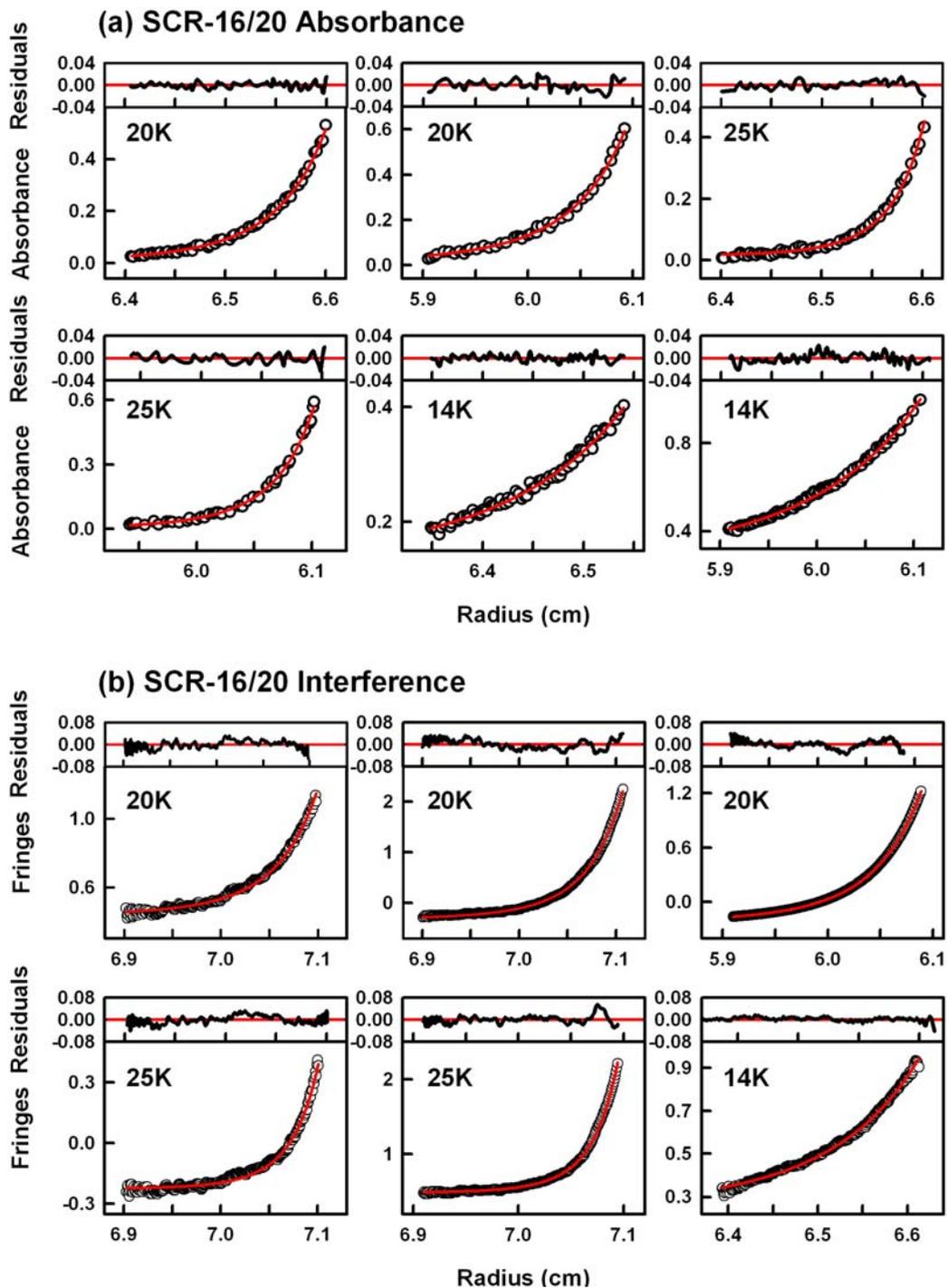
#### **(4.2.3) Constrained scattering modelling of SCR-1/5 and monomeric SCR-16/20**

Solution structures for SCR-1/5 and SCR-16/20 were determined by constrained scattering modelling. The constraints are provided by the use of updated homology models and recent crystal and NMR structures for the ten SCR



**Figure 4.11**

Summary of the global sedimentation equilibrium fits for SCR-1/5 absorbance data using SEDPHAT. The residuals of the fits are shown above each panel (a) The SCR-1/5 absorbance fit at the rotor speeds shown as insets and at concentrations of 0.30 mg/ml (first two upper left panels), 0.10 mg/ml (upper right panel) and 0.06 mg/ml (three lower panels). (b) The SCR-1/5 interference fit at the same rotor speeds and at concentrations of 0.30 mg/ml (three upper panels), 0.06 mg/ml (lower left two panels) and 0.10 mg/ml lower right panel).

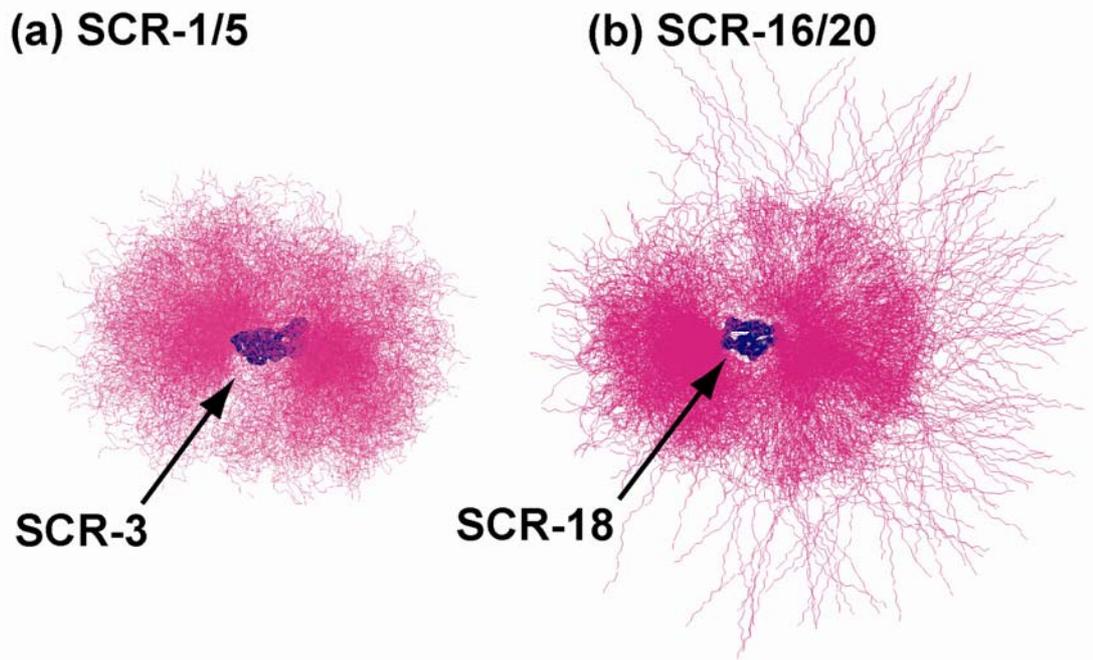


**Figure 4.12**

Summary of the global sedimentation equilibrium fits for SCR-16/20 using SEDPHAT. The residuals of the fits are shown above each panel **(a)** Absorbance fits at the rotor speeds shown as insets correspond to concentrations of 0.13 mg/ml, 0.22 mg/ml, 0.13 mg/ml, 0.22 mg/ml, 0.39 mg/ml and 0.43 mg/ml (from upper left to lower right). **(b)** Interference fits at the rotor speeds shown as insets correspond to concentrations of 0.07 mg/ml, 0.14 mg/ml, 0.22 mg/ml, 0.07 mg/ml, 0.14 mg/ml and 0.39 mg/ml (from upper left to lower right).

domains involved (Saunders *et al.*, 2006; Herbert *et al.*, 2006; Jokiranta *et al.*, 2006). This method is identical to that used in the recent modelling of FH SCR-6/8, (Fernando *et al.*, 2007) and utilises better SCR homology models than those used for the original modelling of intact FH (Aslam & Perkins, 2001). While several *ab initio* curve fit approaches based on shape envelopes, dummy atoms or permutations of spheres have been proposed, these do not utilise the structural constraints provided by the known SCR domain structures, so were not used here. Note that, for the SCR-16/20 searches, the linker between SCR-19 and SCR-20 was held fixed in the slightly bent inter-SCR angle of  $160^\circ$  observed in its crystal structure (Jokiranta *et al.*, 2006). The experimental X-ray curves for the curve fitting utilised those used for the GNOM analyses in Figure 4.5(a,b) at 0.21 mg/ml and 0.84 mg/ml for SCR-1/5 to improve the available Q range, and at 0.05 mg/ml and 0.14 mg/ml for monomeric SCR-16/20 to minimise the contribution of dimer to about 5%.

Models that fitted the X-ray data were identified from the trial-and-error searches. Molecular dynamics were used to produce libraries of 1000 randomised conformations for each of the four linker peptides and the N-terminal and C-terminal peptides (Figure 4.1(c)) by capturing these conformations at set intervals during the simulations (Methods). The lengths of the linkers were held fixed at their theoretical maximum lengths while they were allowed to reorientate, as this gives the best selection of trial models (Fernando *et al.*, 2007). The 6,000 linkers and terminal peptides were randomly selected in order to join up 2,000 trial full models for SCR-1/5 or SCR-16/20 from the five SCR domains within each fragment, and these visibly displayed a full set of randomised conformations as required (Figure 4.13(a,b)). The resulting coordinate models were converted to hydrated Debye sphere models with a sphere volume of  $0.546^3 \text{ nm}^3$  (SCR-1/5) or  $0.544^3 \text{ nm}^3$  (SCR-16/20) in order to calculate their scattering curves. A satisfactory sphere model needed a minimum total number  $N$  of 345 (SCR-1/5) or 391 (SCR-16/20) hydrated spheres in order to correspond to at least 95% of their composition-calculated volumes. Over 99% of the 2,000 models met this “absence of steric overlap” criterion (Table 4.1), and those that did not were discarded. The comparisons of the  $R_{XS}$  and R-factor (goodness of fit) values with the  $R_G$  values from the two sets of models in Figures 4.14(a,b) and 4.15(a-c) showed that the models encompassed the experimental  $R_G$  and  $R_{XS}$  values denoted by dashed lines. This confirmed that a sufficient number of randomised conformations had been generated. Of particular interest in Figures



**Figure 4.13**

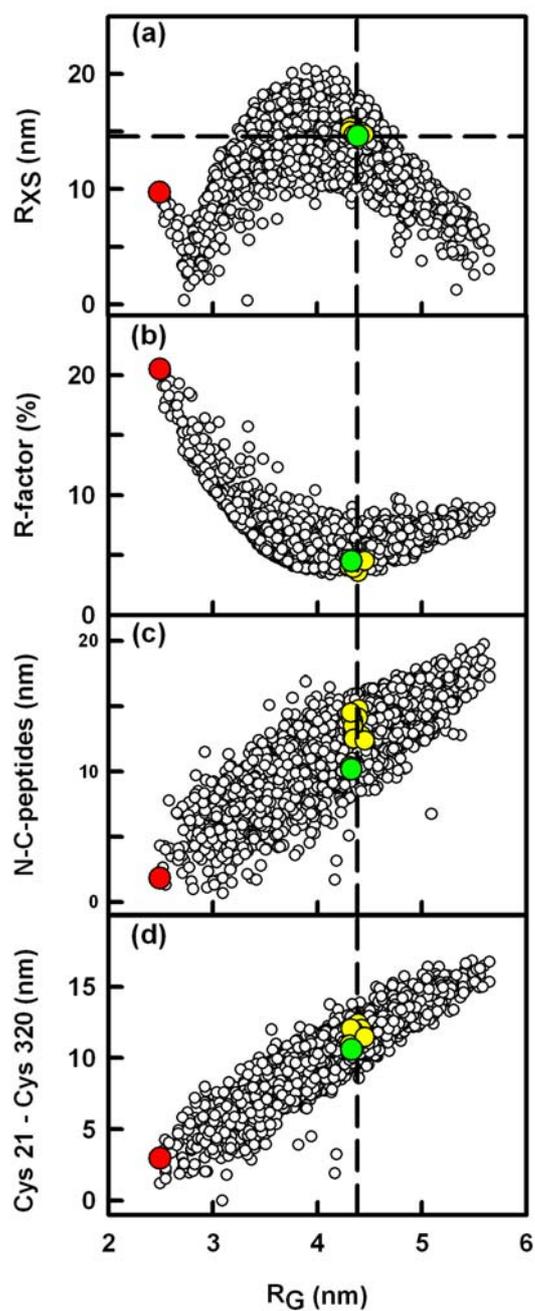
Structural analyses for 2000 trial models for SCR-1/5 and monomeric SCR-16/20.

**(a,b)** Every tenth model (totalling 200) of the 2000 models for each of SCR-1/5 and SCR-16/20 are shown superimposed on SCR-3 and SCR-18 respectively (blue ribbon) in order to show the randomised SCR-1/5 and SCR-16/20 conformations in use.

4.14(a) and 4.15(a) is the correlation between the  $R_G$  and  $R_{XS}$  values, in which an increase in the  $R_G$  values above 4 nm correlates with a straightening of the structure and a reduction in the  $R_{XS}$  values. The lowest  $R$ -factor values indicated the models showing the best agreement with the experimental X-ray curves from the 2,000 models. For both SCR-1/5 and SCR-16/20, this minimum corresponded to an  $R_G$  value that agreed well with the experimental value, provided that the trace aggregates detected in the Guinier plots had been allowed for (Figures 4.14(b); 4.15(b,c)).

The identified best-fit models for both SCR-1/5 and SCR-16/20 were partially folded-back and bent in their appearance. All the models were assessed using the separation between the first and last Cys residues in the first and last SCR domains respectively in order to exclude the N-terminal and C-terminal peptides from this assessment. Thus the 12 best-fit SCR-1/5 models have N- and C-terminal Cys separations of 10.6 nm to 12.4 nm (Table 4.1) in contrast to the maximum possible separation of 18 nm in a linear SCR arrangement and the full range of 1.2 nm to 16.8 nm in all 2,000 models (Figure 4.14(d)). Likewise the ten best-fit SCR-16/20 models have N- and C-terminal Cys separations of 5.5 nm to 9.8 nm in contrast to the full range of 0.9 nm to 16.8 nm in all 2,000 models (Figure 4.15(e)). These comparisons show that both sets of best-fit models have partially folded-back structures. The partially folded-back best fit structures in Figure 4.16(a) reproduced the experimental  $I(Q)$  curve between  $Q$  of  $0.15 \text{ nm}^{-1}$  and  $0.88 \text{ nm}^{-1}$  with an  $R$  factor of 3.6% for SCR-1/5 and in Figure 4.17(a) between  $Q$  of  $0.30 \text{ nm}^{-1}$  and  $0.88 \text{ nm}^{-1}$  with an  $R$  factor of 2.3% for SCR-16/20. The same models fitted the experimental scattering data well at both 0.21 mg/ml and 0.84 mg/ml (SCR-1/5) and 0.05 mg/ml and 0.14 mg/ml (SCR-16/20). The best-fit  $P(r)$  curves (insets in Figures 4.16(a); 4.17(a)) showed maxima  $M$  at  $r$  values close to those experimentally observed (Figure 4.5).

Molecular graphics views showed that the maximum dimension of the best-fit unhydrated SCR-1/5 model is 14.2 nm, which is increased to 15 nm after addition of a hydration shell monolayer (Perkins, 2001). This agrees well with the  $L$  value of 15 nm from the  $P(r)$  curve in Figure 4.5(a). Likewise the maximum dimension of the best-fit SCR-16/20 model is 18.0 nm, this being inclusive of the extended C-terminal 26-residue polypeptide in SCR-16/20 of length 8 nm. This length is estimated to become 19.1 nm on addition of a hydration shell. This will be in good accord with the experimental  $L$  value of 17 nm from the  $P(r)$  analyses in Figure 4.5(b) if the C-

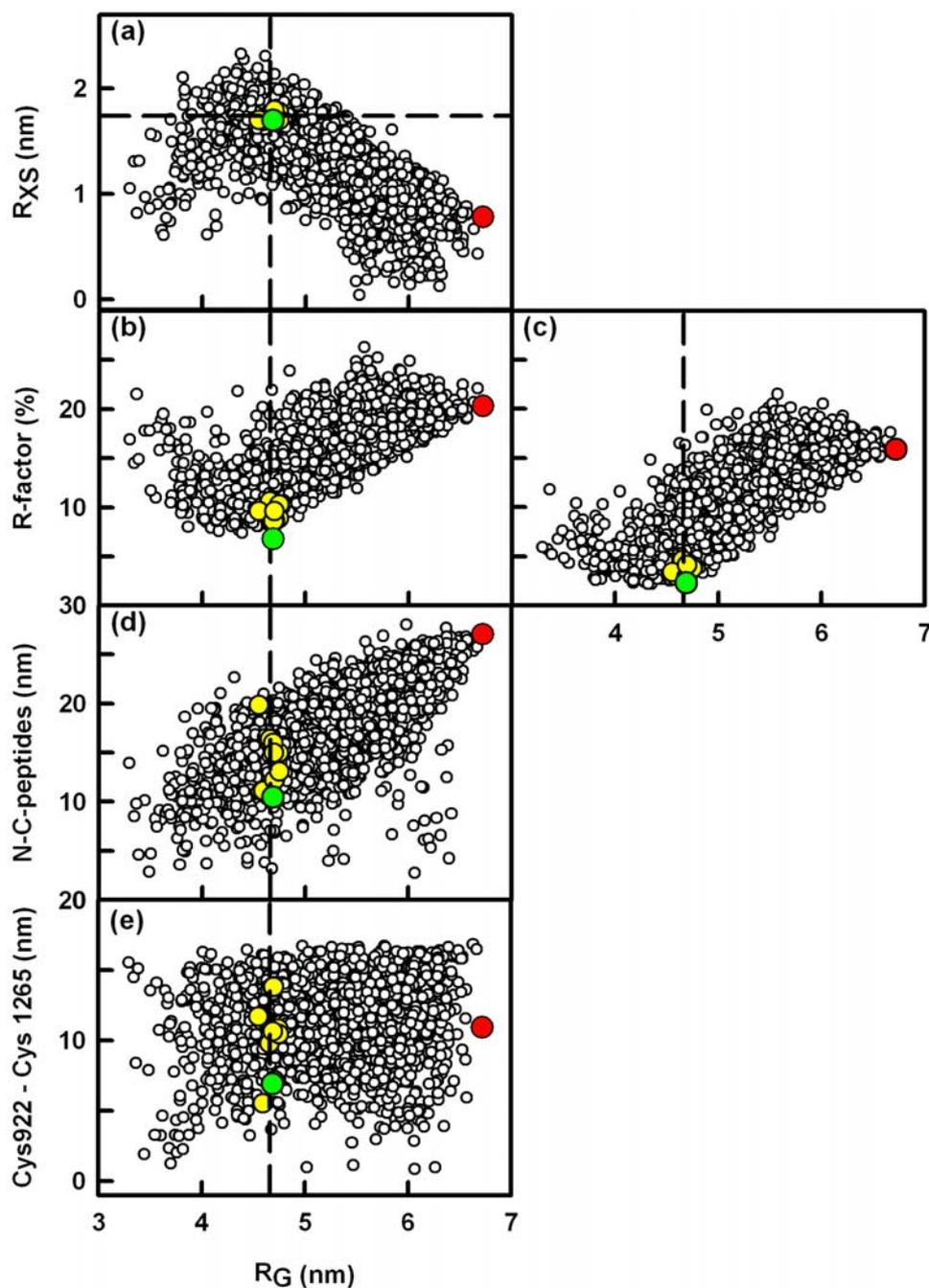


**Figure 4.14**

Results of fitting 2,000 SCR-1/5 models (white circles) to the scattering data at 0.21 mg/ml. Panel (a) compares the  $R_{XS}$  values with the  $R_G$  values. Panel (b) compares the R factors with the  $R_G$  values. Panels (c,d) compare the  $\alpha$ -carbon separation (N- to C- peptides) and (Cys-21 to Cys-320) with the  $R_G$  values. The experimental  $R_G$  and  $R_{XS}$  values of 4.34 nm and 1.46 nm are shown as dashed lines (Table 4.1). The 12 best-fit models on the basis of the  $R_G$ ,  $R_{XS}$  and R-factor filters are shown as yellow and green circles (Table 4.1). The best-fit model (green circle) and the most folded-back model as a representative poor fit (red circle) are also shown.

terminal tail is considered to be sufficiently flexible to adopt multiple conformations, some of which are reoriented towards the centre of the model to reduce its overall length.

Three further sets of calculations clarified the outcome of the scattering modelling for SCR-1/5 and SCR-16/20. Firstly, the ten and 12 best-fit models for SCR-1/5 and SCR-16/20 (yellow and green circles in Figures 4.14 and 4.15) are found in similar locations in the distributions of Figures 4.14 and 4.15, suggesting that all the best-fit scattering models form one related family of structures. The superimposition of these best-fit structures upon the central SCR domain in Figure 4.18 showed that all exhibited similar partially folded-back bent structures. Secondly, the scattering modelling is effective in rejecting poor-fit models. Two such examples are shown in Figures 4.14 and 4.15 (red circles) and the resulting poor fits in Figures 4.16(b) and 4.17(b). One poor curve fit with an  $R$  factor of 20.5% is illustrated by a fully-folded-back SCR-1/5 model in which the five SCR domains form a U-shaped arrangement. This resulted in large deviations in both the  $I(Q)$  and  $P(r)$  curves from the experimental curves (Figure 4.16(b)), and gave an  $R_G$  value of 2.49 nm and an  $R_{XS}$  value of 0.97 nm in large deviation from the experimental values (Table 4.1). Furthermore steric clashes were observed between the SCR domains as the linker peptides are relatively short at four residues in each one. Another poor curve fit with an  $R$  factor of 15.9% is illustrated by an extended SCR-16/20 model (Figure 4.17(b)). Again large deviations in both the  $I(Q)$  and  $P(r)$  curves from the experiment occurred (Figure 4.17(b)), and this gave an  $R_G$  value of 6.72 nm and an  $R_{XS}$  value of 0.78 nm in large deviation from the experimental values. Thirdly, comparison with the four best-fit models A-D for intact FH (PDB code 1haq) showed that the SCR-1/5 and SCR-16/20 fragments extracted from these mostly showed reasonable agreements with the scattering curves for SCR-1/5 and SCR-16/20. Model C had the closest resemblance to the presently-determined structure of SCR-1/5 with an  $R$ -factor of 4.0% compared to that of 3.6% for the fit shown in Figure 4.16(a) and the 12 best fit structures (Table 4.1). The other three models showed  $R$  factors of 5.4% to 7.5%. Model D had the closest resemblance to the presently-determined structures of SCR-16/20 modelled here with an  $R$ -factor of 5.1% compared to that of 2.3% for the fit shown in Figure 4.17(a) (Aslam & Perkins, 2001). The other three models showed too high  $R$  factors of 8.7% to 15.1%; these deviated more because these structures were too extended. Hence the new partially bent back structures of SCR-1/5 and



**Figure 4.15**

Summary of 2,000 fits to the monomeric SCR-16/20 scattering data at 0.14 mg/ml. Panel (a) compares the  $R_{XS}$  values with the  $R_G$  values. Panel (b) compares R factors with  $R_G$  values for experimental scattering curves at 0.05 mg/ml and 0.27 mg/ml, while (c) compares R factors with  $R_G$  values for the experimental scattering curve at 0.14 mg/ml. Panels (d,e) compare the  $\alpha$ -carbon separation for SCR-16/20 (N- to C-peptides) and (Cys-922 to Cys-1228) with the  $R_G$  values. The experimental  $R_G$  and  $R_{XS}$  values of 4.66 nm and 1.74 nm are shown as dashed lines (Table 4.1). The ten best-fit models are shown as yellow and green circles (Table 4.1). The best-fit model is shown in green, and the red circle denotes the most extended model as a representative poor fit.

SCR-16/20 determined in this study resemble those found within the previously determined intact FH solution structure.

As an independent check of the scattering modelling, the  $s_{20,w}^0$  values of the 12 best-fit SCR-1/5 models were calculated. These ranged from 2.28 S to 2.37 S (HYDRO) and 2.06 S to 2.27 S (HYDROPRO) (Table 4.1). The best-fit model gave calculated  $s_{20,w}^0$  values of 2.33 S and 2.22 S, which is in good agreement with the experimental values of 2.37 S to 2.52 S. The  $s_{20,w}^0$  values calculated for the ten best-fit SCR-16/20 models ranged between 2.23 S to 2.30 S (HYDRO) and between 2.06 S to 2.20 S (HYDROPRO) (Table 4.1). The best-fit scattering model gave calculated  $s_{20,w}^0$  values of 2.29 S and 2.17 S. These are about 0.6 S less than the experimental values of 2.78 S to 2.87 S. As the C-terminal tail is extended in all these models, this was reoriented back into the centre of the model. A model of overall length 12 nm in place of 17 nm gave calculated  $s_{20,w}^0$  values of 2.59 S and 2.35 S to show good agreement with experiment. Accordingly this discrepancy is likely to be due to the conformation of the disordered 26-residue C-terminal tail which is inferred to be flexible. In summary, these calculations support the outcome of the scattering modelling which is summarised in Figure 4.18.

#### (4.2.4) Trial models for dimeric SCR-16/20

In order to account for the scattering curve and  $s_{20,w}^0$  value for the SCR-16/20 dimer, molecular graphics was used with the best-fit partially folded-back model for SCR-16/20 to create manually trial conformations for two-fold symmetric SCR-16/20 dimer models. These models assumed that no conformational change occurred in SCR-16/20 upon dimerisation. The weak association of SCR-16/20 meant that about 50% dimerisation had occurred at the 1.15 mg/ml concentration used for curve fits. Dimerisation had previously been observed for the SCR-15/18 and SCR-15/20 fragments in non-reducing SDS PAGE, showing at least one dimerisation site is present in SCR-15/18 (Jokiranta *et al.*, 1998). The earliest scattering modelling of the FH dimer suggested that the dimerisation site may occur at SCR-20, although there was no proof for this (Perkins *et al.*, 1991). While no dimerisation in SCR-19/20 was observed by NMR, (Herbert *et al.*, 2006) a Trp1183Leu mutant in SCR-20 showed enhanced oligomer formation (Jokiranta *et al.*, 2006). Hence the weak association and the variables associated with dimer formation meant that it was not considered feasible to perform constrained modelling searches as above for the two monomers.

**Table 4.1. X-ray scattering and sedimentation coefficient fits for the solution structures of SCR-1/5 and SCR-16/20**

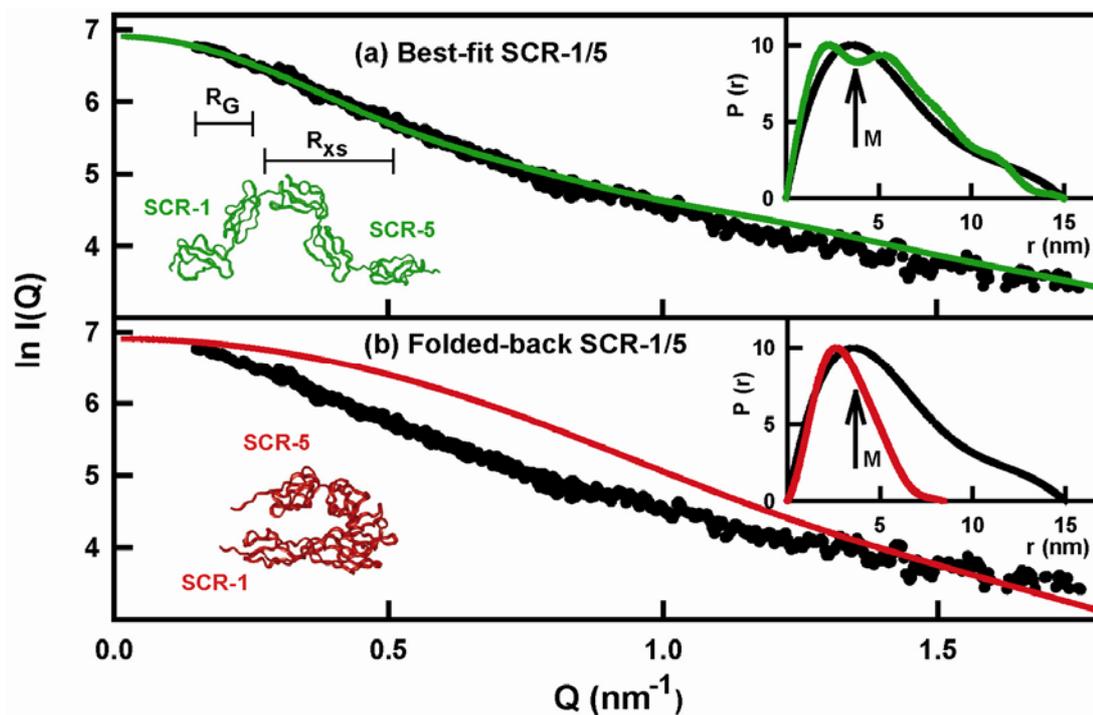
	Filter	Models	Hydrated spheres	$R_G$ (nm)	$R_{XS}$ (nm)	$R$ factor (%)	N-C terminal separation (nm) <sup>a</sup>	$s_{20,w}^0$ (S)
Experimental (SCR-1/5)			363	4.26 ± 0.16; 4.42 ± 0.13 <sup>b</sup>	1.46 ± 0.08			2.52 ± 0.02 2.37 ± 0.03 <sup>c</sup>
Experimental (SCR-16/20 monomer)			412	4.66 ± 0.13; 4.81 ± 0.11 <sup>b</sup>	1.74 ± 0.04			2.78 ± 0.30 2.87 ± 0.20 <sup>c</sup>
Experimental (SCR-16/20 dimer)			824	8.09 ± 0.21 7.51 ± 0.24 <sup>b</sup>	2.42 ± 0.14			3.54 ± 0.14 3.92 ± 0.23 <sup>c</sup>
Modelling (SCR-1/5)	None	2000	319 - 414	2.49 - 5.65	0.04 - 2.03	3.4 - 20.6	1.2 - 16.8	n.a.
Modelling (SCR-1/5)	$R_G$ , $R_{XS}$ and $R$ factor	12	365 - 392	4.31 - 4.45	1.46 - 1.54	3.6 - 4.5	10.6 - 12.4	2.28 - 2.37, 2.06 - 2.27 <sup>d</sup>
Modelling (SCR-1/5)	Best fit	1	392	4.39	1.46	3.6	12.4	2.33, 2.22
Modelling (SCR-16/20 monomer)	None	2000	393 - 487	3.34 - 6.72	0.04 - 2.32	2.2 - 21.5	0.9 - 16.8	n.a.
Modelling (SCR-16/20 monomer)	$R_G$ , $R_{XS}$ and $R$ factor	10	440 - 465	4.55 - 4.75	1.69 - 1.80	2.3 - 4.6	5.5 - 9.8	2.23 - 2.30, 2.06 - 2.20 <sup>d</sup>
Modelling (SCR-16/20 monomer)	Best fit	1	446	4.69	1.70	2.3	6.91	2.29, 2.17
Modelling (SCR-16/20 dimer)	$R$ factor	8	855 - 892	5.57 - 9.10	1.61 - 2.62	5.9 - 13.7	n.a.	3.09 - 3.46, 2.90 - 3.32 <sup>d</sup>
Modelling (SCR-16/20 dimer)	Best fit	1	857	7.11	1.78	5.9	n.a.	3.20, 3.06

<sup>a</sup> Distance between the  $\alpha$ -carbon atoms of the first and last residues in SCR-1/5 between Cys21 and Cys320, or that in SCR-16/20 between Cys931 and Cys1228 (Figure 4.1).

<sup>b</sup> The first value is from the Guinier  $R_G$  analyses (Figures 4.3; 4.4); the second is from the GNOM  $P(r)$  analyses (Figure 4.5). The filtering of the 2,000 models and the  $R_G/R_0$  ratios used  $R_G$  values of 4.34 nm (SCR-1/5) and 4.66 nm (SCR-16/20).

<sup>c</sup> The first value is from the DCDT+ analyses (Figures 4.6; 4.9); the second is from the SEDFIT analyses (Figures 4.7; 4.10).

<sup>d</sup> The first value is calculated from the model using HYDRO; the second is calculated using HYDROPRO



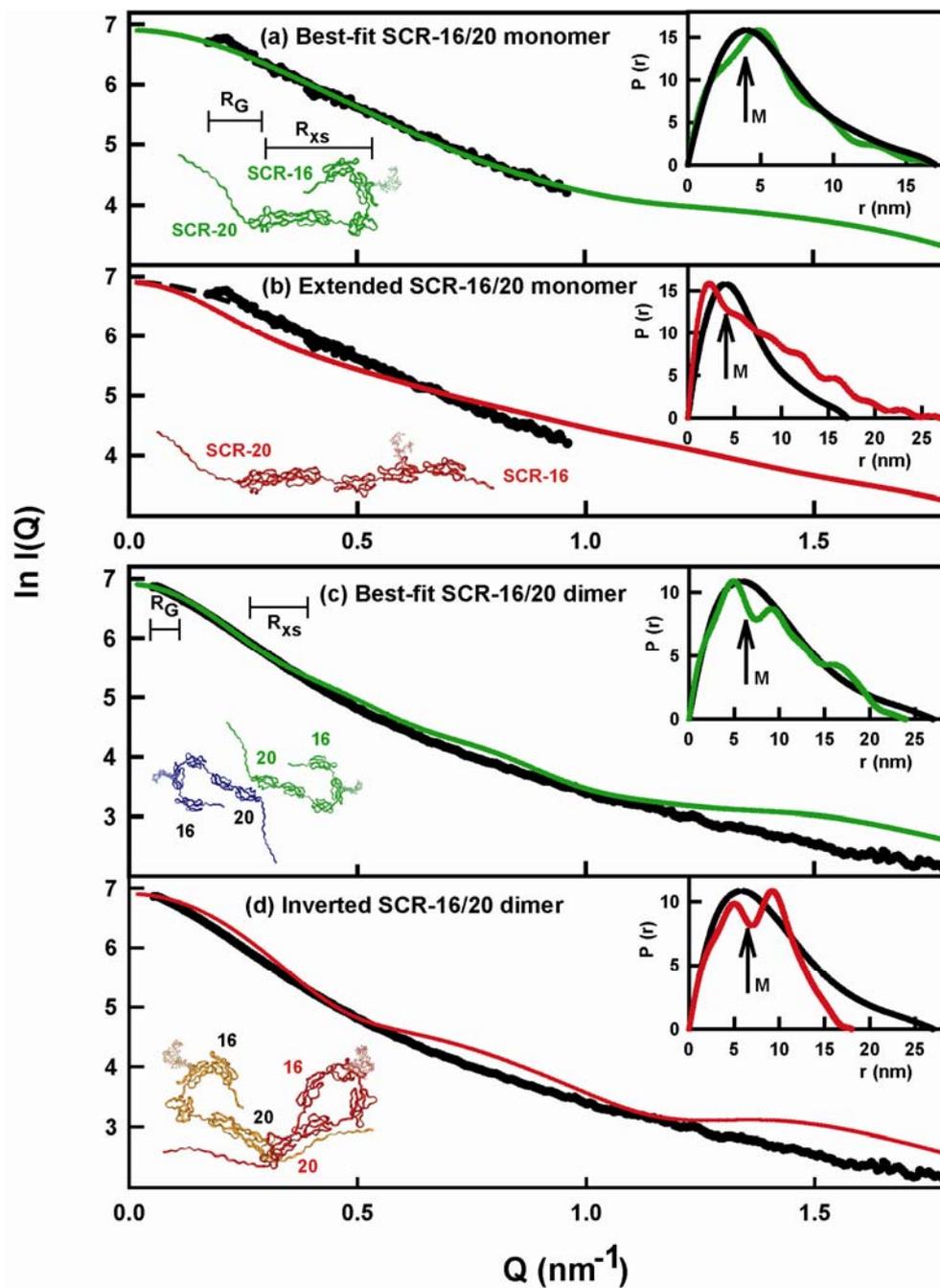
**Figure 4.16**

Best-fit and poor-fit X-ray scattering modelling for SCR-1/5. The fitted experimental scattering data in (a,b) at 0.21 mg/ml and 0.84 mg/ml are denoted by black points. The green and red lines correspond to the best-fit and poor-fit models respectively. The  $Q$  ranges employed in the experimental  $R_G$  and  $R_{XS}$  analyses are denoted by horizontal bars. The insets at the upper right show the modelled  $P(r)$  distance distribution functions in green and red for comparison with the experimental  $P(r)$  curve in black (Figure 4.5). The insets to the lower left show ribbon views of the models used to calculate the curves.

Accordingly a range of conformers involving the arbitrary choice of SCR-20 or SCR-15 as the dimerisation contact point were assessed by using their  $s_{20,w}^0$  and  $R_G$  values. For example, a best-fit dimer model involved anti-parallel side-by-side contacts between the SCR-20 domains (Figure 4.17(c)) and gave  $s_{20,w}^0$  values of 3.20 S and 3.06 S (Table 4.1). These were in fair agreement with the observed experimental values of 3.54 S to 3.92 S. The calculated  $R_G$  value of 7.11 nm was comparable with the experimental  $R_G$  values of 7.51 nm to 8.09 nm. Scattering curve fitting gave a surprisingly reasonable R-factor of 5.9% and good visual agreement with the I(Q) and P(r) curves (Figure 4.17(c)). In contrast to this, a more compact inverted dimer model gave larger  $s_{20,w}^0$  values of 3.43 S and 3.25 S. While both values were in fair agreement with the observed experimental values of 3.54 S to 3.92 S, the calculated  $R_G$  value of 5.92 nm was much less than the experimental  $R_G$  values of 7.51 nm to 8.09 nm. Larger deviations with the I(Q) and P(r) curves were seen (Figure 4.17(d)), and the R-factor was higher at 7.4%. No fits could be obtained on the basis of the overlapped domain arrangements seen in the crystal packing of SCR-19/20 tetramers, (Jokiranta *et al.*, 2006) as these would have significantly reduced the length of the dimer to below that observed in Figure 4.5(c) (see Figure 4.1(b)). In conclusion, these calculations show that the dimer is formed as an extended arrangement of two SCR-16/20 monomers from contacts between two SCR-20 or SCR-15 domains or at a different pair of SCR domains.

### (4.3) Conclusions

This study of the two most functionally-relevant fragments of FH has revealed new properties of this major complement regulator. Firstly, evidence from both X-ray scattering and AUC show that the SCR-16/20 fragment exists in a weak monomer-dimer equilibrium while this is not seen for the SCR-1/5 fragment. These results offer further insights into previous reports of the existence of monomeric or dimeric forms of FH (Perkins *et al.*, 1991; Aslam & Perkins, 2001). Furthermore, conformational similarities in the partially folded-back and bent domain arrangements for both the SCR-1/5 and SCR-16/20 fragments of FH were identified. These are similar to the previously-deduced partially folded-back solution arrangement for the 20 SCR domains in intact FH (Aslam & Perkins, 2001).



**Figure 4.17**

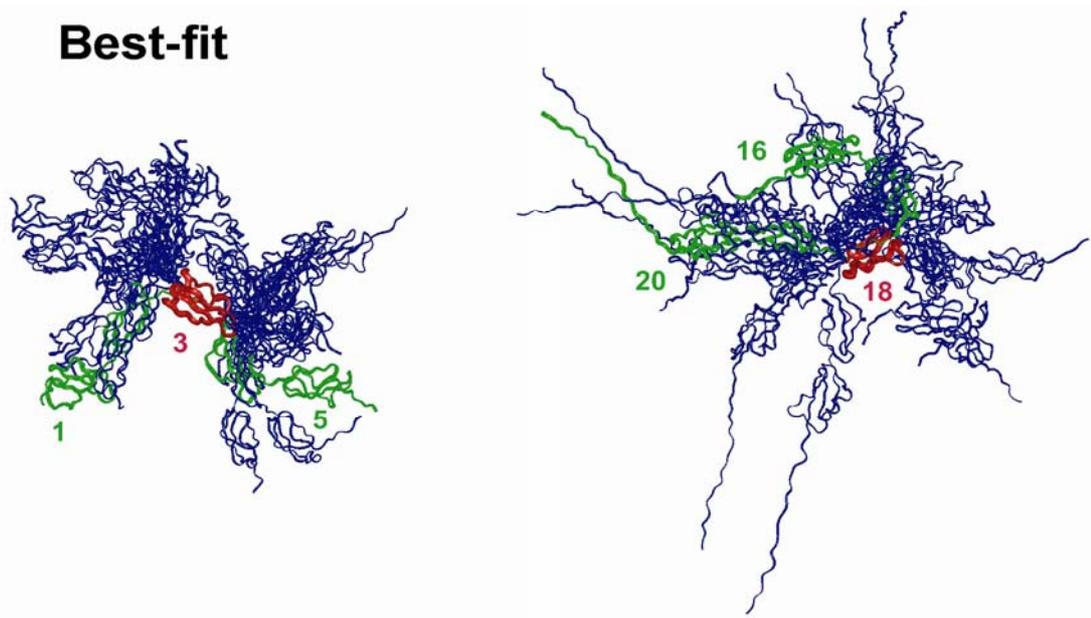
Best-fit and poor-fit X-ray scattering modelling for monomeric and dimeric SCR-16/20. The fitted experimental scattering data in **(a,b)** for monomeric SCR-16/20 at 0.05 mg/ml and 0.14 mg/ml, and **(c,d)** for dimeric SCR-16/20 at 1.15 mg/ml are denoted by black points. The green and red lines correspond to the best-fit and poor-fit models respectively. The  $Q$  ranges employed in the experimental  $R_G$  and  $R_{XS}$  analyses are denoted by horizontal bars. The insets at the upper right show the modelled  $P(r)$  distance distribution functions in green and red for comparison with the experimental  $P(r)$  curve in black (Figure 4.5). The insets to the lower left show ribbon views of the models used to calculate the curves.

The Guinier and P(r) scattering data and the velocity and equilibrium AUC data show that SCR-1/5 does not exhibit self-associative properties, whereas SCR-16/20 does. FH exists in plasma at concentrations of 0.235 mg/ml to 0.810 mg/ml (1.6  $\mu$ M to 5.4  $\mu$ M) (Saunders *et al.*, 2006). The SCR-1/5 concentrations of 0.21-0.84 mg/ml (6.0  $\mu$ M to 24.1  $\mu$ M) and SCR-16/20 concentrations of 0.05-1.15 mg/ml (1.3  $\mu$ M to 29.0  $\mu$ M) used in this study therefore includes and exceeds the physiological plasma concentration range of FH. Four determinations of the  $K_D$  for the monomer-dimer equilibrium from the Guinier analyses and the sedimentation equilibrium fits give a mean  $K_D$  of  $16 \pm 2$   $\mu$ M. If native FH behaves similarly to the SCR-16/20 fragment, then 7% to 21% of native FH in plasma will be dimeric by this mechanism. However, it should be noted that the observed self-association of SCR-16/20 may result from its recombinant nature or its 26-residue C-terminal peptide (Figure 4.1(c)), and therefore it may not apply to native FH. Nonetheless the weak dimerisation of SCR-16/20 is consistent with previous demonstration of monomeric and dimeric forms of FH (Perkins *et al.*, 1991; Aslam & Perkins, 2001). Previously a second weak dimerisation of the SCR-6/8 fragment in FH with a  $K_D$  of 40  $\mu$ M was identified, which corresponds to about 4% to 11% of dimeric FH in plasma (Fernando *et al.*, 2007). FH is found in the drusen deposits of the eye retina that are a hallmark of age-related macular degeneration (Hageman *et al.*, 2005). The potential dimerisation at separate pairs of SCR-6/8 and SCR-16/20 segments between different FH molecules may constitute a cross-linking mechanism by which FH could aggregate continuously to form larger assemblies. The dimerisation of FH may also increase its affinity for binding to heparin-like anionic ligands if the ligand and dimer sites are independent of each other, further promoting aggregation processes. This possibility is supported by the observation of oligomerisation in native FH from plasma (Nan *et al.*, 2008a).

The joint use of scattering and AUC experiments in conjunction with constrained modelling showed that both SCR-1/5 and SCR-16/20 are partially folded-back and bent in solution. The well-resolved peaks in the size distribution analyses of Figures 4.7 and 4.10 indicate no evidence of shape polydispersity. There are indications of disorder in the C-terminal tail of SCR-16/20 from the sedimentation coefficient modelling. The appearance of the model in Figure 4.18 illustrates the degree of SCR arrangements that are possible. However it is stressed

that other partially folded-back structures that fit the scattering data are possible, and specific inter-SCR orientations cannot be identified by constrained modelling. Only the general degree to which this occurs can be elucidated. The coexistence of two conformations such as those shown in [Figures 4.16\(a,b\)](#) or [4.17\(a,b\)](#) is ruled out by the good fits to the available Q ranges in [Figures 4.16\(a\)](#) and [4.17\(a\)](#). The coexistence of multiple conformations had previously been reported for complement receptor type 2 with 15 SCR domains that gave Q-dependent curve fits, ([Gilbert \*et al.\*, 2006](#)) however there were no indications of this in the present study. With two-SCR or three-SCR fragments, it is possible to infer an averaged bend ([Fernando \*et al.\*, 2007](#); [Gilbert \*et al.\*, 2005](#)). Here, with five-SCR fragments, both the SCR-1/5 and SCR-16/20 fragments exhibited an averaged inter-domain SCR angle of  $127 \pm 25^\circ$  and  $125 \pm 31^\circ$  respectively when averaged over the 12 and 10 best-fit models.

These are significantly different from a linear SCR arrangement with an angle of  $180^\circ$ . Comparison with an averaged inter-domain SCR angle of  $85 \pm 43^\circ$  for the SCR-1/5 and SCR-16/20 fragments extracted from the four best-fit models for intact FH shows that the partial folding back of the two fragments is similar to that found in the original constrained scattering modelling of FH ([Aslam & Perkins, 2001](#)). This is well illustrated by the comparative views in [Figure 4.19](#). The advantage of modelling five-SCR fragments is that this reduces the uncertainty implicit in modelling the entire 20-SCR structure. If the central ten SCR domains in FH are arranged similarly to those in SCR-1/5 and SCR-16/20, this will provide an explanation of the shorter overall length of FH when this is compared to that expected from a maximally extended arrangement. This degree of partial folding-back in FH agrees with earlier electron micrographs of FH obtained in relatively harsh conditions ([DiScipio, 1992](#)). The main difference between this present study and the electron micrographs is to demonstrate that it is not necessary for two different SCR domains to interact with each other across the FH structure to form “alpha-shaped” arrangements in order to account for its reduced overall length. The proposed partial folding-back of the SCR domains is sufficient to achieve this with typical inter-SCR angles of about  $85^\circ$  to  $127^\circ$ . In fact, if the alternative hypothesis of “alpha-shaped” arrangements of domains within FH is true, ([Oppermann \*et al.\*, 2006](#); [DiScipio, 1992](#)) the necessary contacts between different SCR domains within the same molecule would have to be sufficiently stable to favour their formation. This may offer the possibility for FH to



**Figure 4.18**

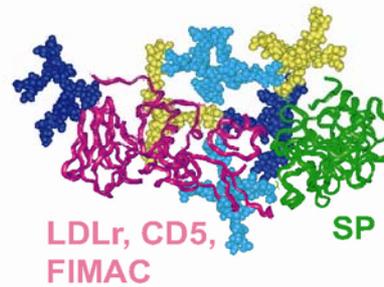
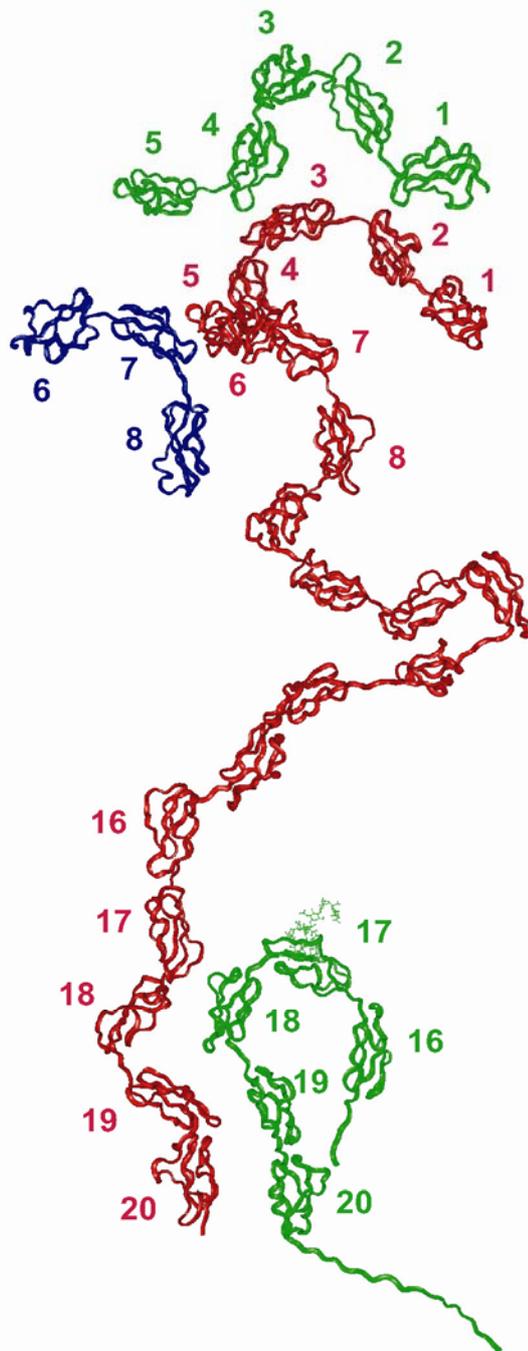
Outcome of the constrained modelling fits for SCR-1/5 and SCR-16/20.

The 12 best-fit models for SCR-1/5 and the ten best-fit models for monomeric SCR-16/20 are shown superimposed on SCR-3 for SCR-1/5 and on SCR-18 for SCR-16/20. SCR-3 and SCR-18 are shown as red ribbons. The overall best-fit models are shown in green, while the remainder is shown as blue ribbons.

self-aggregate continuously through interactions at non-identical SCR binding sites and cause it to precipitate from solution, which is unlikely.

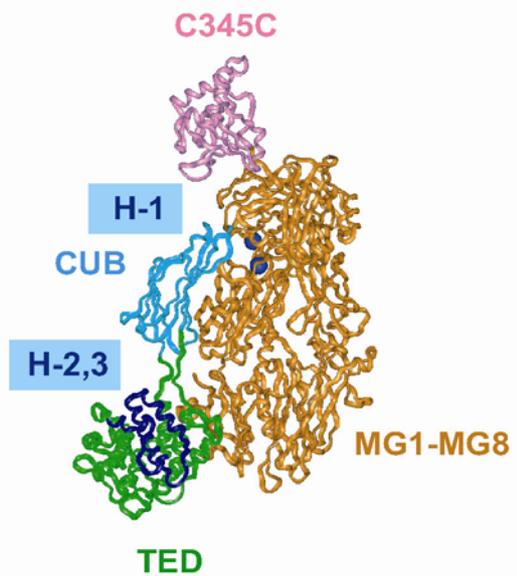
The functional significance of the SCR-1/5 and SCR-16/20 structures is best considered in the context of the macromolecular ligands to which they bind (Figure 4.19). One question is the expected degree of inter-SCR flexibility within these two partially-folded back structures. The four linkers in SCR-1/5 are all four residues long (Figure 4.1(a)), while those in SCR-16/20 are three or four residues long with the interesting exception of the SCR-18/19 linker which is six residues long. It is possible that there may be specific hydrogen bonding networks at the linker regions to stabilise these structures, but a high resolution structure will be required to establish this. If it can be assumed that the partially folded-back structures persist, they appear to be ideal for protein-protein interactions with C3b as illustrated in Figure 4.19 at the three sites denoted H-1 to H-3 by being able to wrap themselves around a localised region of the C3b surface to prepare C3b for its cleavage by Factor I at the CUB domain to release C3d and C3c. However the six-residue SCR-18/19 linker is longer and potentially more flexible. This may permit SCR-19/20 to bind to heparin-like oligosaccharides at host cell surfaces, whilst at the same time permitting the remaining SCR-1/18 segment of FH to orientate itself freely into the solution phase to facilitate contacts with C3b molecules. Poisson-Boltzmann electrostatic surface maps suggest that SCR-16/20 is predominantly basic in its surface regions while SCR-1/5 is predominantly acidic (Figure 4.20). Hence this suggests that SCR-16/20 is well adapted for binding to host cell surfaces as well as to C3d. A further indication of the nature of the interactions of SCR-16/20 with its ligands is the occurrence of Type II mutations within SCR-16/18 as well as in SCR-19/20 that all lead to atypical haemolytic uraemic syndrome (Saunders *et al.*, 2007). Type II mutations are associated with functional binding defects at the surface of the protein, in distinction to Type I mutations that cause protein misfolding to occur and lead to reduced quantities of FH in plasma. Up to now it has been difficult to account for these mutations in SCR-16/18 as well as in SCR-19/20. The comparison of the partially folded-back SCR-16/20 structure with the crystal structures of C3b (Janssen *et al.*, 2006, Weisman *et al.*, 2006; Abdul Ajees *et al.*, 2006) with these mutations suggests that all five domains in SCR-16/20 may have a role to play in mediating functional interactions with C3d or with heparin-like ligands.

## Factor H



## Factor I

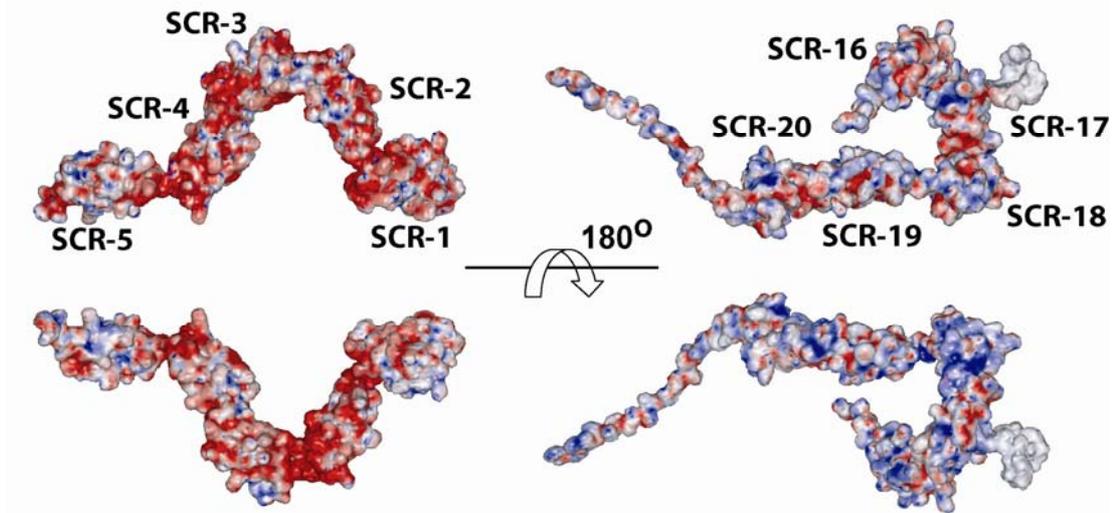
## C3b



**Figure 4.19**

The best-fit SCR-1/5 and SCR-16/20 models (green) are shown in comparison with the scattering models for SCR-6/8 (blue: PDB code 2ic4) and intact FH (red: PDB code 1haqA). The comparison is shown to the same scale with the scattering model for Factor I (green SP domain; magenta: FIMAC, DC5 and LDLr domains). The C3b crystal structure (PDB code: 2i07) is shown in green (TED domain), cyan (CUB domain), orange (the MG1-MG8 domains) and purple (the C345C domain). The FH binding sites on C3b are shown in blue as H-1 to H-3.

## Electrostatic maps



**Figure 4.20**

Electrostatic maps of the SCR-1/5 and SCR-16/20 best-fit models are shown in two 180° front and back rotational views about their longest axes. Acidic surfaces are shown in red, while basic surfaces are shown in blue.



mediated cleavage of C3b (data not shown) (Gordon *et al.*, 1995). The SCR-16/20 fragment likewise possesses activity by binding to heparin and also to glycosaminoglycans on human umbilical vein endothelial cells (Ormsby *et al.*, 2006). The X-ray scattering and AUC experiments were performed separately or jointly on each sample preparation in phosphate buffered saline (PBS) after dialysis into this buffer (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.3).

Compositional data for the data analyses are summarised. The SCR-1/5 and SCR-16/20 sequences were taken from SWISSPROT accession number P08603. SCR-1/5 was taken to be unglycosylated at its single putative N-linked site, while SCR-16/20 was taken to be glycosylated at one of its two putative sites in SCR-17 and SCR-18 (Figure 4.1). SCR-1/5 has a calculated molecular weight of 35,300 Da, an unhydrated volume of 44.7 nm<sup>3</sup>, a hydrated volume of 59.1 nm<sup>3</sup> (based on a hydration of 0.3g H<sub>2</sub>O/g of glycoprotein and an electrostricted volume of 0.0245 nm<sup>3</sup> per bound water molecule), a partial specific volume  $\bar{v}$  of 0.7202 ml/g, and an absorption coefficient at 280 nm (1%, 1 cm path length) of 17.2 (Perkins, 1986). SCR-16/20 has a calculated molecular weight of 39,700 Da, an unhydrated volume of 50.1 nm<sup>3</sup>, a hydrated volume of 66.3 nm<sup>3</sup>, a  $\bar{v}$  of 0.7187 ml/g, and an absorption coefficient of 15.7 (Perkins, 1986). The buffer density was calculated as 1.00534 g/ml from standard tabulations in the program SEDNTERP for use in data processing, (Laue *et al.*, 1992) and confirmed to be 1.00541 g/ml at 20°C by measurement on an Anton-Paar DMA5000 density meter.

#### **(4.4.2) X-ray scattering data collection and analysis**

X-ray scattering data were obtained for SCR-1/5 and SCR-16/20 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France operating with a ring energy of 6.0 GeV on the Beamline ID02 (Narayanan *et al.*, 2001). Storage ring currents ranged from 22 mA to 42 mA for SCR-1/5 (one beam session in four bunch mode). For SCR-16/20, these currents ranged from 23 mA to 42 mA (three beam sessions in four bunch mode) and from 171 mA to 181 mA (one beam session in uniform bunch mode). SCR-1/5 was studied at 0.21 mg/ml to 0.84 mg/ml, while SCR-16/20 was studied at 0.05 mg/ml to 1.15 mg/ml. After one initial session using a static sample holder, all the final samples were measured in flow cells to reduce radiation damage by moving the sample continuously during beam exposure in 10

time frames of duration 0.1 sec, 0.2 sec, 0.5 sec or 1.0 sec each, together with on-line checks for the absence of radiation damage, after which the frames were averaged. Other details including the data reduction procedure are described elsewhere (Gilbert *et al.*, 2005).

In a given solute-solvent contrast, the radius of gyration  $R_G$  is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low  $Q$  gives the  $R_G$ , and the forward scattering at zero angle  $I(0)$  (Glatter & Kratky, 1982):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3.$$

This expression is valid in a  $Q.R_G$  range up to 1.5. If the structure is elongated, the mean radius of gyration of cross-sectional structure  $R_{XS}$  and the mean cross-sectional intensity at zero angle  $[I(Q)Q]_{Q \rightarrow 0}$  is obtained from:

$$\ln [I(Q)Q] = [I(Q)Q]_{Q \rightarrow 0} - R_{XS}^2 Q^2/2.$$

The  $R_G$  and  $R_{XS}$  analyses lead to the triaxial dimensions of the macromolecule if the structure can be represented by an elliptical cylinder, where  $L = \sqrt{12(R_G^2 - R_{XS}^2)}$  and  $L$  is its length (Glatter & Kratky, 1982). The  $R_G$  and  $R_{XS}$  analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics O2 Workstations. Indirect transformation of the scattering data  $I(Q)$  in reciprocal space into real space to give the distance distribution function  $P(r)$  was carried out using the GNOM program (Semenyuk & Svergun, 1991):

$$P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q) Q r \sin(Qr) dQ$$

$P(r)$  corresponds to the distribution of distances  $r$  between volume elements. The SCR-1/5 X-ray curve utilised up to 450 data points in the  $Q$  range between  $0.15 \text{ nm}^{-1}$  and  $0.88 \text{ nm}^{-1}$  at  $0.21 \text{ mg/ml}$  (Figure 4.5(a)), supplemented by that between  $0.16 \text{ nm}^{-1}$  and  $1.76 \text{ nm}^{-1}$  at  $0.84 \text{ mg/ml}$  for Figure 4.16. The SCR-16/20 monomer X-ray curves utilised up to 422 data points in the  $Q$  range between  $0.16 \text{ nm}^{-1}$  and  $0.96 \text{ nm}^{-1}$  at  $0.14$

mg/ml, and between 0.12 nm<sup>-1</sup> and 0.49 nm<sup>-1</sup> at 0.05 mg/ml (Figure 4.17(a,b)). The largest usable Q value increased with concentration. Thus the SCR-16/20 dimer X-ray curve utilised 298 data points in the Q range between 0.06 nm<sup>-1</sup> and 2.01 nm<sup>-1</sup>. Other details are as described previously (Gilbert *et al.*, 2005).

#### (4.4.3) Sedimentation velocity and equilibrium data for SCR-1/5 and SCR-16/20

AUC data for SCR-1/5 and SCR-16/20 were obtained on a Beckman XL-I instrument using AnTi50 or AnTi60 rotors. Sedimentation velocity data was acquired at 20°C at rotor speeds of 35,000 r.p.m and 42,000 r.p.m. for SCR-1/5 and 35,000 r.p.m., 40,000 r.p.m., 42,000 r.p.m., 50,000 r.p.m. and 55,000 r.p.m. for SCR-16/20 in two-sector cells with column heights of 12 mm. The different rotor speeds established the absence of diffusion effects on the sedimentation coefficient. The  $g(s^*)$  time-derivative analyses of between 10 to 30 absorbance and interference scans were analysed using DCDT+ version 2.06, (Philo, 2006) from which the sedimentation and diffusion coefficients (and hence the molecular weight) were determined from the peak position and width. The number of scans selected for the DCDT+ analyses ensured that peak broadening effects were negligible (Philo, 2006). A second independent sedimentation analysis was performed using direct boundary Lamm fits of 200 scans using SEDFIT (version 9.4) (Schuck, 1998; 2000). SEDFIT resulted in size-distribution analyses  $c(s)$  that assumed that all species have the same frictional ratio  $f/f_0$ , and this established the oligomeric species that were present. The final SEDFIT analyses optimised the  $c(s)$  model by floating the meniscus, frictional ratio, and baseline until the overall root mean square deviations and visual appearance of the fits were satisfactory (Figures 4.7; 4.10).

Sedimentation equilibrium data were acquired during a run for 150 h using six-sector cells with solution column heights of 2 mm at rotor speeds of 11,000 r.p.m., 14,000 r.p.m., 20,000 r.p.m., 25,000 r.p.m., 30,000 r.p.m., and 35,000 r.p.m. until equilibrium had been reached at each speed (confirmed by the perfect overlay of runs measured at 3 h intervals). For reason of protein stability, the most successful data collection was performed at 4°C. The molecular weight  $MW$  for SCR-1/5 was initially determined on the assumption of a single species using SEDPHAT (version 4.3; Schuck, 2003):

$$c_r = c_{r_0} \exp [(\omega^2/2RT) MW (1 - \bar{v} \rho) (r^2 - r_0^2)]$$

where  $c_r$  is the concentration at radius  $r$ ,  $c_{r_0}$  is the concentration of the protein at the reference radius  $r_0$ ,  $\omega$  is the angular velocity,  $R$  is the gas constant,  $T$  is the temperature (Kelvin), and  $\rho$  is the solvent density of the buffer. Final molecular weights were obtained using global fits of up to 15 curves in SEDPHAT in the “single species of interacting system” model with both the “M and s” and Marquardt-Levenberg fitting routines, while floating the molar extinction coefficient, the baseline and the bottom of the cell. The molecular weight  $MW$  for SCR-16/20 was initially determined using SEDPHAT on the assumption of a reversible monomer-dimer self-association equilibrium, followed by global fits of up to 18 curves. For a protein in a reversible monomer-dimer self-association, the mass action law  $c_r = c_{r_0}^2 K_{12}$  is fulfilled simultaneously in all positions within the solution column, leading to the distribution: (McRorie & Voelker, 1993)

$$c_r = c_{r_0} \exp [(\omega^2/2RT) MW (1 - \bar{v} \rho) (r^2 - r_0^2)] + 2K_{12} c_{r_0}^2 \exp [(\omega^2/2RT) 2MW (1 - \bar{v} \rho) (r^2 - r_0^2)]$$

where  $K_{12}$  is the association constant for dimer formation. Molecular weights were obtained by fitting on the basis of the “monomer-dimer self-association” model using the “M and s”, simplex and Marquardt-Levenberg fitting routines, while floating the baseline and the bottom of the cell. Statistical error analysis employed the Monte-Carlo approach.

#### (4.4.4) Modelling of the SCR-1/5 and SCR-16/20 solution structures

The modelling of human SCR-1/5 and SCR-16/20 was based on the NMR structure for SCR-16 and the crystal structure for SCR-19/20 (Barlow *et al.*, 1993, Jokiranta *et al.*, 2006), and the updated homology models for the remaining seven SCR domains, which were prepared using MODELLER v7 (Saunders *et al.*, 2006). The FH sequence numbering starts from the N-terminus of the 18-residue signal peptide and not that of the mature protein, to follow recent genetic publications (Klein *et al.*, 2005; Haines *et al.*, 2005; Edwards *et al.*, 2005; Hageman *et al.*, 2005). For SCR-1/5, N-terminal and C-terminal tail sequences 17-EFEDC-21 and 320-CTLHHHHHH-328 and four linker sequences 80-CQKRPC-85, 141-CEVVKC-146, 205-CVEISC-210 and 262-CEEKSC-267 (Figure 4.1) were created in extended  $\beta$ -

strand conformations using INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA) on Silicon Graphics O2 and OCTANE Workstations. For SCR-16/20, N-terminal and C-terminal tail sequences 922-EAEAEFGLPC-931 and 1228-CAKRPLEQKLISEEDLNSAVDHHHHHH-1254 and three linker sequences 984-CIKTDC-989, 1043-CRDTSC-1048 and 1102-CKDSTGKC-1109 (i.e. excluding that between SCR-19 and SCR-20) (Figure 4.1) were created in the same manner. Libraries of 1,000 random linker and 1,000 random tail conformations were generated using molecular dynamics simulations. One structure from each library was randomly selected to assemble 2,000 full SCR-1/5 models and 2,000 full SCR-16/20 monomer models. Further details are given elsewhere (Fernando *et al.*, 2007). In these models, the inter-SCR domain angle is defined as the angle between the two longest axes of two adjacent SCR domains, where each axis is the line between the midpoint of the two  $\alpha$ C atoms in the disulphide bridges at the two ends of the SCR domain (Gilbert *et al.*, 2005). The N-C terminal distance is defined as the distance from Cys-21 to Cys-320 (SCR-1/5) and Cys-931 to Cys-1228 (SCR-16/20).

#### (4.4.5) Debye scattering and sedimentation coefficient modelling

Each SCR-1/5 or SCR-16/20 model was used to calculate X-ray scattering curves for comparison with the experimental curves using Debye sphere models (Gilbert *et al.*, 2005). A cube side length of 0.546 nm (SCR-1/5) and 0.544 nm (SCR-16/20) in combination with a cutoff of 4 atoms for both models consistently produced 275- and 311-sphere unhydrated models respectively, each within 95% of the total volume calculated from its composition. The hydration shell was created using HYPRO, (Ashton *et al.*, 1997) where the optimal total of hydrated spheres for SCR-1/5 is 363 and that for SCR-16/20 is 412. The X-ray scattering curve  $I(Q)$  was calculated using the Debye equation as adapted to spheres (Perkins & Weiss, 1983). Other details are given elsewhere (Perkins, 2001; Boehm *et al.*, 1999). The number of spheres  $N$  in the dry and hydrated models after grid transformation was used to assess steric overlap between the SCR domains, where models showing less than 95% of the optimal total were discarded. The models were assessed by calculation of the X-ray  $R_G$  and  $R_{XS}$  values from the modelled curves in the same  $Q$  ranges used for the experimental Guinier fits in order to allow for any approximations inherent in the use of the  $Q.R_G$  range up to 1.5. Models that passed these three filters were then ranked using a goodness-of-fit  $R$  factor defined by analogy with protein

crystallography and based on the experimental curves, in the Q range extending to  $1.8 \text{ nm}^{-1}$ .

Sedimentation coefficients  $s_{20,w}^0$  were calculated directly from the same hydrated SCR-1/5 and SCR-16/20 scattering sphere models using the HYDRO program, (Garcia de la Torre *et al.*, 1994) and also directly from the atomic coordinates in the HYDROPRO shell modelling program using the default value of 0.31 nm for the atomic element radius for all atoms to represent the hydration shell (Garcia de la Torre *et al.*, 2000).

#### **(4.4.6) Protein Data Bank accession number**

The best-fit SCR-1/5 and SCR-16/20  $\alpha$ -carbon co-ordinate models have been deposited in the Protein Data Bank with the accession codes 2qfg and 2qfh.

## **Chapter Five**

### **Electrostatic interactions contribute to the folded-back conformation of wild-type human Factor H**

## (5.1) Introduction

The FH SCR domain structure is not fully extended in solution but is partially folded back, as shown by solution scattering, AUC and electron microscopy (Perkins *et al.*, 1991; Discipio, 1992; Aslam & Perkins, 2001). Constrained scattering modelling (Perkins *et al.*, 2008) resulted in the first medium resolution molecular structures for FH and its SCR-6/8, SCR-1/5 and SCR-16/20 fragments to show that all four exhibited some degree of domain folding-back (Aslam & Perkins, 2001; Fernando *et al.*, 2007; Chapter 4). This is distinct from the more extended solution structures found for the 30-SCR and 15-SCR domains of complement receptor types 1 and 2 (CR1 and CR2) (Gilbert *et al.*, 2006; Furtado *et al.*, 2008). In addition, FH undergoes self-association and forms a significant percentage of oligomers at physiological concentrations (Perkins *et al.*, 1991; Nan *et al.*, 2008a). Other FH fragments, including SCR-6/8 and SCR-16/20, also self-associate to form dimers (Jokiranta *et al.*, 2000; 2006; Oppermann *et al.*, 2006; Fernando *et al.*, 2007; Chapter 4). The presence of multiple self-dimerisation sites on FH suggested that FH might form indefinite oligomers. Significant amounts of these FH oligomers formed in the presence of at least 60  $\mu\text{M}$  zinc or copper (Nan *et al.*, 2008b).

An understanding of FH regulation requires the elucidation of the conformation and self-association properties of FH in different solution conditions. Currently there are no explanations for its folded-back structure or self-association and its relevance to FH regulation. A significant proportion of aHUS mutations occur outside SCR-20, which contains the binding sites for C3d and heparin. Even though these aHUS mutations appear to affect the function of FH, the reason for their location outside SCR-20 is not yet known (Saunders *et al.*, 2007). It is not clear if an omega-shaped structure for FH proposed from the interaction between SCR-1/7 and SCR-1/20 (Oppermann *et al.*, 2006) may be alternatively explained by the dimerisation of SCR-6/8 (Fernando *et al.*, 2007). The interactions between FH and C3b ligands have been studied by SPR and AUC to show that their affinities are weak with dissociation constants in the micromolar range and depend on salt concentration (Hellwage *et al.*, 2002; Chapter 6; Hellwage *et al.*, 1999). In contrast, FH has a high affinity for heparin (Fernando *et al.*, 2007). Thus it is possible that electrostatic interactions contribute towards the formation of a folded back FH structure. In this chapter, previous AUC data at pH 7.4 and 137 mM NaCl (Nan *et*

*al.*, 2008a) are extended to other NaCl concentrations and pH values to show that the FH conformation depends on both these parameters. Histidine residues are shown to be involved in this. FH self-association also showed a dependence on the NaCl concentration. From new X-ray scattering data collection, improved molecular models for FH compared to those first determined (Aslam & Perkins, 2001) are derived, and these are similar in accuracy to those determined recently for CR1 and CR2 (Furtado *et al.*, 2008; Gilbert *et al.*, 2006).

## **(5.2) Results and Discussion**

### **(5.2.1) Analytical ultracentrifugation of Factor H**

To assess if the folded-back domain structure of FH can result from inter-SCR charge attraction within FH, the charged groups in each SCR domain were summed up (Figure 5.1). There are 1213 residues between Glu19 and Arg1231 in mature FH. In addition, eight of the nine putative oligosaccharide sites are occupied by mostly biantennary sugars (Fenaille *et al.*, 2007). Each oligosaccharide chain was assumed to possess two NeuNAc residues. This gave a total of 326 charged groups in FH with an overall net charge of -2, assuming that the His residues are protonated. The pI value of FH was predicted to be 6.34 using SEDNTERP software (Laue *et al.*, 1993), and 6.12 from the ProtParam tool on the EXPASY website (<http://www.expasy.ch/tools/protparam.html>), neither of which took glycosylation into account. Thus SCR-7, SCR-13 and SCR-20 showed positive net charges of +5 or greater, while SCR-9, SCR-12, SCR-14 and SCR-15 showed negative net charges of -4 or greater (Figure 5.1). The large variations in net charge between individual FH SCR domains make it possible that relatively distant SCR domains might attract each other within FH and contribute to its folded back structure.

For AUC, wild-type FH was freshly purified from a pool of just-outdated anonymised human plasma using an anti-FH monoclonal antibody MRC-OX23 column (Section 5.4.1). This plasma pool is identical to that used before (Nan *et al.*, 2008a; 2008b). Data for FH in 137 mM NaCl, pH 7.4 were collected by

	SCR-1	SCR-2	SCR-3	SCR-4	SCR-5
$\text{NH}_3^+$					
Arg	+6	+2	+2	+3	+5
Lys	+3	+2	+6	+4	+3
His	0	+1	+2	0	+1
Asp	-2	-4	-4	-2	-3
Glu	-5	-4	-8	-6	-3
NeuNAc	0	0	0	0	0
	<u>+3 (17)</u>	<u>-3 (13)</u>	<u>-2 (22)</u>	<u>-1 (15)</u>	<u>+3 (15)</u>

	SCR-6	SCR-7	SCR-8	SCR-9	SCR-10
Arg	+2	+3	+1	+1	+2
Lys	+3	+4	+7	+3	+6
His	+5	+1	0	+1	+2
Asp	-5	-1	-3	-6	-4
Glu	-3	-2	-4	-2	-4
NeuNAc	0	0	0	-2	0
	<u>+2 (18)</u>	<u>+5 (11)</u>	<u>+1 (15)</u>	<u>-5 (15)</u>	<u>+2 (18)</u>

	SCR-11	SCR-12	SCR-13	SCR-14	SCR-15
Arg	+1	+1	+4	+2	+2
Lys	+6	0	+8	+2	+3
His	+1	+3	+3	+1	+2
Asp	-1	-2	-3	-2	0
Glu	-8	-6	-5	-5	-7
NeuNAc	0	-2	-2	-2	-4
	<u>-1 (17)</u>	<u>-6 (14)</u>	<u>+5 (25)</u>	<u>-4 (14)</u>	<u>-4 (18)</u>

	SCR-16	SCR-17	SCR-18	SCR-19	SCR-20
					$\text{COO}^-$
Arg	0	+2	+5	+2	+8
Lys	+4	+5	+1	+4	+5
His	+3	0	0	0	+2
Asp	-2	-3	-2	-3	-1
Glu	-6	-3	-5	-3	-5
NeuNAc	0	-2	-2	0	0
	<u>-1 (15)</u>	<u>-1 (15)</u>	<u>-3 (15)</u>	<u>0 (12)</u>	<u>+8 (22)</u>

**Figure 5.1**

Charge summation in the 20 SCR domains of FH. The total charge at physiological pH is shown under each SCR. It is assumed that each His residue is protonated and there is one additional positive or negative charge at the N- or C-terminus respectively. The linker residues preceding each SCR domain are included within the summation for that domain. The summation for SCR-20 includes the 1229-AKR-1231 C-terminus. The seven glycosylated domains are shaded in grey. SCR-4 is not glycosylated despite the presence of a putative site for this (see text). Each oligosaccharide chain is assumed to contain two N-acetyl-neuraminic acid residues.

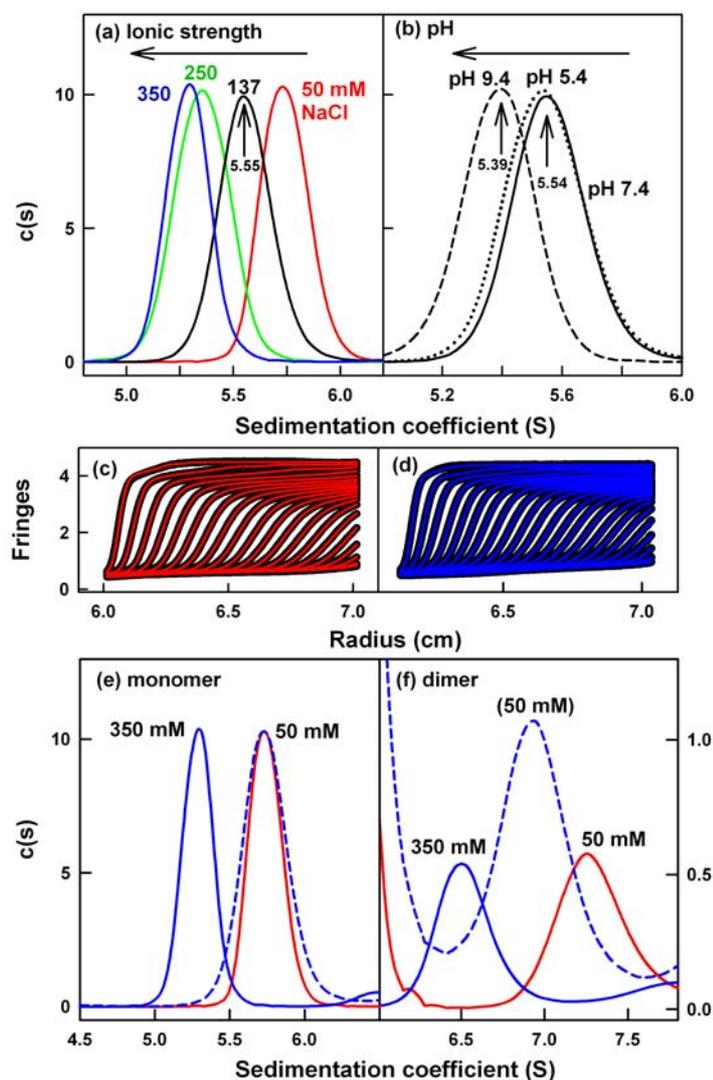
Miss Ruodan Nan (Nan *et al.*, 2008a), and this is acknowledged accordingly in the relevant figure legends.

FH showed similar elution profiles as a single peak (not shown) by gel filtration in buffers containing 50 mM or 137 mM NaCl. Sedimentation velocity runs using interference optics were performed in order to determine sedimentation coefficient  $s_{20,w}$  and molecular weight data for FH. The  $s_{20,w}$  value monitors the degree of macromolecular elongation. Conformational changes in monomeric FH were revealed by performing these runs as a function of three NaCl concentrations between 50 mM and 350 mM NaCl in phosphate-buffered saline (PBS), all at pH 7.4 (Section 5.4.1; Figure 5.2(a)). These were supplemented using a fourth data set previously collected in 137 mM NaCl HEPES buffer (Nan *et al.*, 2008a). The observed sedimentation boundaries were fitted to yield a sedimentation coefficient distribution,  $c(s)$  according to the Lamm equation using SEDFIT (Section 5.4.2). This takes into account all the species present in the sample, although the algorithm assumes that the same frictional ratio applies for all the sedimenting species.

The SEDFIT analyses of up to 300 scans revealed excellent boundary fits (Figure 5.2(c,d)). Concentration series were performed in order to investigate the oligomerisation of FH (Nan *et al.*, 2008a). No concentration dependence in the  $s_{20,w}$  value of FH was observed between 0.1 to 1.0 mg/ml FH in each salt concentration (Figure 5.3). However the  $s_{20,w}$  value decreased significantly as the salt concentration increased with values of 50 mM NaCl, 137 mM, 250 mM and 350 mM NaCl. The  $s_{20,w}$  values were 5.66 S, 5.56 S, 5.41 S and 5.31 S in that order (Table 5.1). Conformational reversibility was demonstrated by dialysis of FH from 350 mM NaCl buffer to 50 mM NaCl buffer, whereupon the  $s_{20,w}$  value reverted from 5.31 S to 5.66 S (Figure 5.2(e)). Additional velocity runs were performed at pH 5.4 and pH 9.4 in PBS with 137 mM NaCl in six concentrations starting from 0.8 mg/ml FH (Figure 5.4). For the FH monomer, the mean  $s_{20,w}$  value of  $5.54 \pm 0.05$  S at pH 5.4 was similar at all rotor speeds to that of  $5.56 \pm 0.08$  S at pH 7.4 (Nan *et al.*, 2008a). However, that at pH 9.4 was consistently lower at  $5.40 \pm 0.05$  S (Figures 5.2(b); 5.4(b)). Conversion of the  $c(s)$  plots to molecular weight size distribution  $c(M)$  plots can provide useful insight on macromolecular sizes, provided that the peaks correspond to discrete sedimenting

species and the individual components have similar frictional ratios  $f/f_o$  (where  $f_o$  is the frictional coefficient of the sphere with the same volume as the hydrated glycoprotein) (Brown & Schuck, 2005; Dam & Schuck, 2005). The  $c(M)$  plots yielded a mean molecular weight of  $157 \pm 9$  kDa for the major peak (four NaCl concentrations and two additional pH values), in good agreement with the value of 154 kDa calculated from the composition of the FH monomer, hence corroborating the assignment of this peak. It is concluded that a NaCl concentration-dependent conformational change occurs in monomeric FH, in which FH becomes more extended in higher salt. Given that His residues deprotonate with a pK close to 7, the change seen at pH 9.4 indicates that the folded-back FH domain structure at low pH becomes more elongated at high pH and involves the release of a charged interaction involving histidine(s).

FH forms dimers (Perkins *et al.*, 1991; Nan *et al.*, 2008a). At FH concentrations below 1 mg/ml, a dissociation constant  $K_D$  of 28  $\mu$ M between the monomer and dimer in 137 mM NaCl (pH 7.4) has been reported (Nan *et al.*, 2008a). The  $c(s)$  plots in Figure 5.3 showed that the most prominent peak after that for the monomer is peak 2 in concentrations up to 1.0 mg/ml. This is assigned as dimeric FH for two reasons: (i) The relative intensity of peak 2 visibly increased with FH concentration in Figure 5.3, showing that peak 2 is in equilibrium with the FH monomer as expected. Integration of peak 2 confirmed this (Figure 5.6(a)). (ii) Assuming that the  $f/f_o$  ratio is unaltered, the  $c(M)$  size distribution plots from the six experiments showed that peak 2 corresponded to a mean molecular weight of  $234 \pm 17$  kDa, which is about one-and-half times that of the monomer at  $157 \pm 9$  kDa. One explanation of the molecular weight difference from the expected value of 309 kDa is that a conformational change in the FH dimer results in a smaller frictional ratio of the FH dimer, and indicative modelling calculations supported this possibility (Nan *et al.*, 2008a). In addition to this, further complexity is indicated by the rotor-speed dependence of the  $s_{20,w}$  values for peak 2 in the  $c(s)$  plots. At 40k r.p.m., 50k r.p.m. and 60k r.p.m., the data sets in 50 mM, 137 mM, 250 mM and 350 mM NaCl buffers showed that the four sets of  $s_{20,w}$  values for peak 2 decreased with increase in



**Figure 5.2**

Sedimentation velocity  $c(s)$  distribution analyses of FH. The curve for 137 mM NaCl, pH 7.4 is used with permission from Miss Ruodan Nan (Nan *et al.*, 2008a). **(a,b)** The effect of NaCl concentration and pH on monomeric Factor H at 1.0 mg/ml. The horizontal arrows show the decrease of the sedimentation coefficient  $s_{20,w}$  values with increase of the NaCl concentration and the pH. The three  $s_{20,w}$  values at pH 5.4, pH 7.4 and pH 9.4 in 137 mM NaCl are arrowed. Here and below, 50 mM NaCl is denoted by red, 137 mM NaCl by black, 250 mM NaCl by green and 350 mM NaCl in blue. **(c,d)** The boundary fits for the sedimentation of FH in 50 mM (red) and 350 mM (blue) NaCl at pH 7.4 are shown in intervals of 10 scans. **(e,f)** The dialysis of FH from 350 mM (blue line) into 50 mM NaCl buffer (dashed blue line) is compared with FH in 50 mM NaCl buffer (red line). The peak for monomeric FH showed positional reversibility in **(c)**, while the peak for dimeric FH shows incomplete positional reversibility in **(d)**.

rotor speed.<sup>I</sup> Similar decreases with increase in rotor speed were seen also at pH 5.4 and pH 9.4. Thus a second explanation of the molecular weight difference is the occurrence of rate exchange processes between the FH monomer and dimer that are comparable to the sedimentation rate, and that peak 2 may correspond to a reaction boundary and not to a dimer (Balbo & Schuck, 2005). Separately from this, an increase in NaCl concentration showed that the  $s_{20,w}$  value for peak 2 in Figure 3 decreased with salt concentration.<sup>II</sup> Interestingly, peak 2 for FH at the highest concentration of 1 mg/ml in 350 mM NaCl did not revert to its original position at 50 mM NaCl after back-dialysis; this may indicate a lack of reversibility in conformation or rate exchange processes (Figure 5.2(d)). The pH dependence of peak 2 showed no clear variation, unlike peak 1.<sup>III</sup> It is concluded that a reversible equilibrium between monomer and dimer occurs over the range of NaCl concentrations that was studied. The FH dimer possesses a structure that becomes more elongated with increase in NaCl concentration like the FH monomer.

FH also forms higher oligomers. This was indicated by the deviations from a uniform sedimentation profile in the uppermost region of the boundary fits, which were most marked in 50 mM NaCl (Figures 5.2(c,d) and 5.3). The corresponding  $c(s)$  plots showed a variable range of small peaks 3 to 8 in a range between 8 S and 20 S in all four NaCl concentrations (Figure 5.3). Similar peaks were seen at pH 7.4 and 137 mM NaCl irrespective of whether the data were acquired using interference or absorption optics, although the peak resolution was weaker with absorption optics (Figure 5.5). Their average  $s_{20,w}$  values agreed with previously reported values, and were previously assigned to higher FH oligomers on the basis of indicative modelling calculations (Nan

---

<sup>I</sup> The  $s_{20,w}$  value of peak 2 decreased from  $7.66 \pm 0.33$  S to  $6.88 \pm 0.24$  S (50 mM NaCl),  $7.88 \pm 0.30$  S to  $7.13 \pm 0.30$  S (137 mM NaCl),  $7.23 \pm 0.37$  S to  $6.80 \pm 0.32$  S (250 mM NaCl) and  $7.07 \pm 0.14$  S to  $6.65 \pm 0.28$  S (350 mM NaCl) with increase in rotor speed.

<sup>II</sup> The four mean  $s_{20,w}$  values of peak 2 were  $7.22 \pm 0.33$  S,  $7.3 \pm 0.2$  S,  $6.79 \pm 0.28$  S and  $6.71 \pm 0.23$  S in that order from 50 mM NaCl to 350 mM NaCl.

<sup>III</sup> The mean  $s_{20,w}$  values of peak 2 were  $6.88 \pm 0.38$  S at pH 5.4,  $7.43 \pm 0.30$  S at pH 7.4, and  $6.80 \pm 0.30$  S at pH 9.4, all at 50,000 r.p.m..

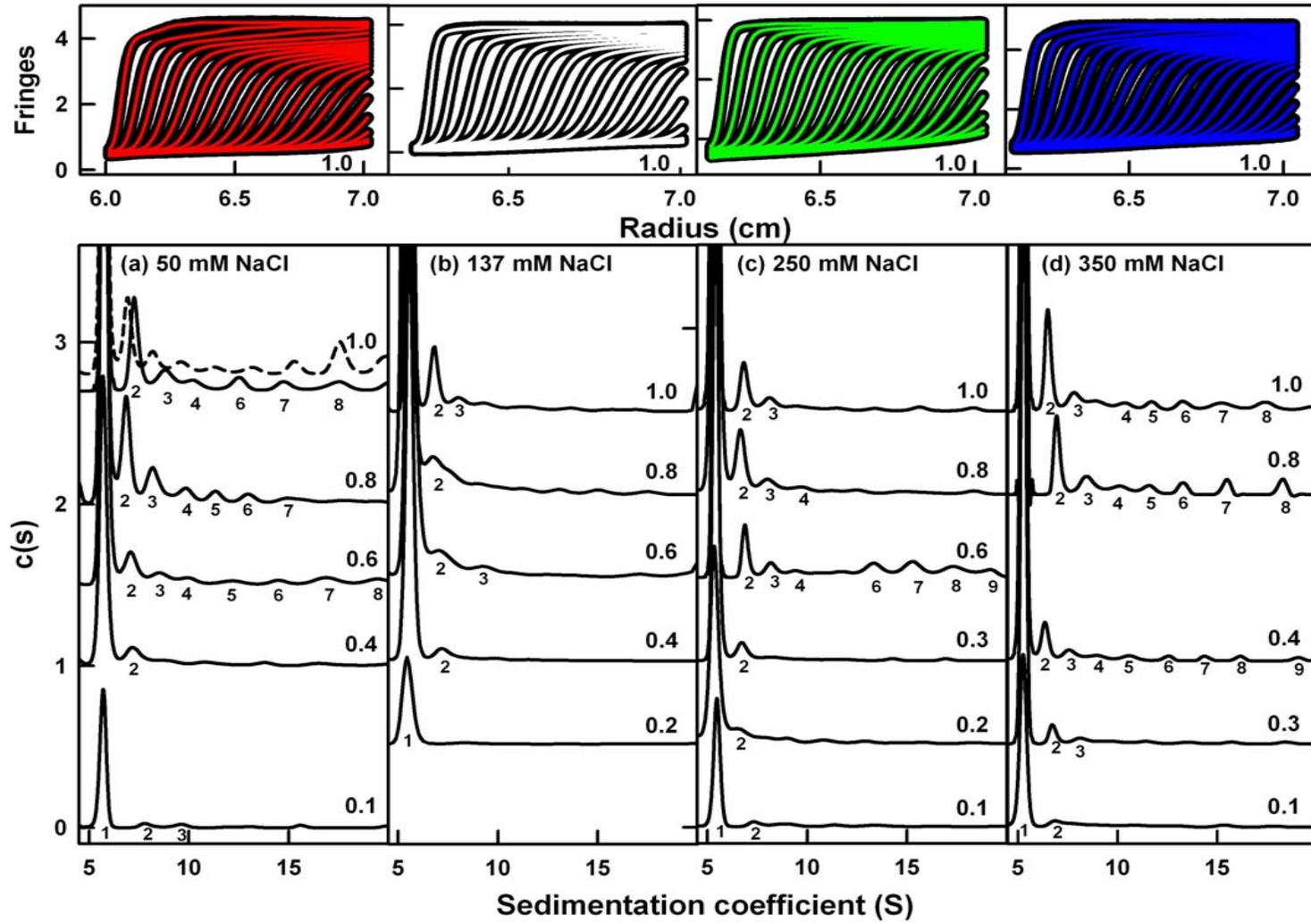


Figure 5.3 (Figure legend overleaf)

(Figure 5.3 legend)

NaCl concentration-dependence of the sedimentation of FH oligomers. Figure (b) is used with permission from Miss Ruodan Nan ([Nan \*et al.\*, 2008a](#)).

Experiments at a rotor speed of 50,000 r.p.m. correspond to buffers containing

(a) 50 mM,

(b) 137 mM,

(c) 250 mM and,

(d) 350 mM NaCl.

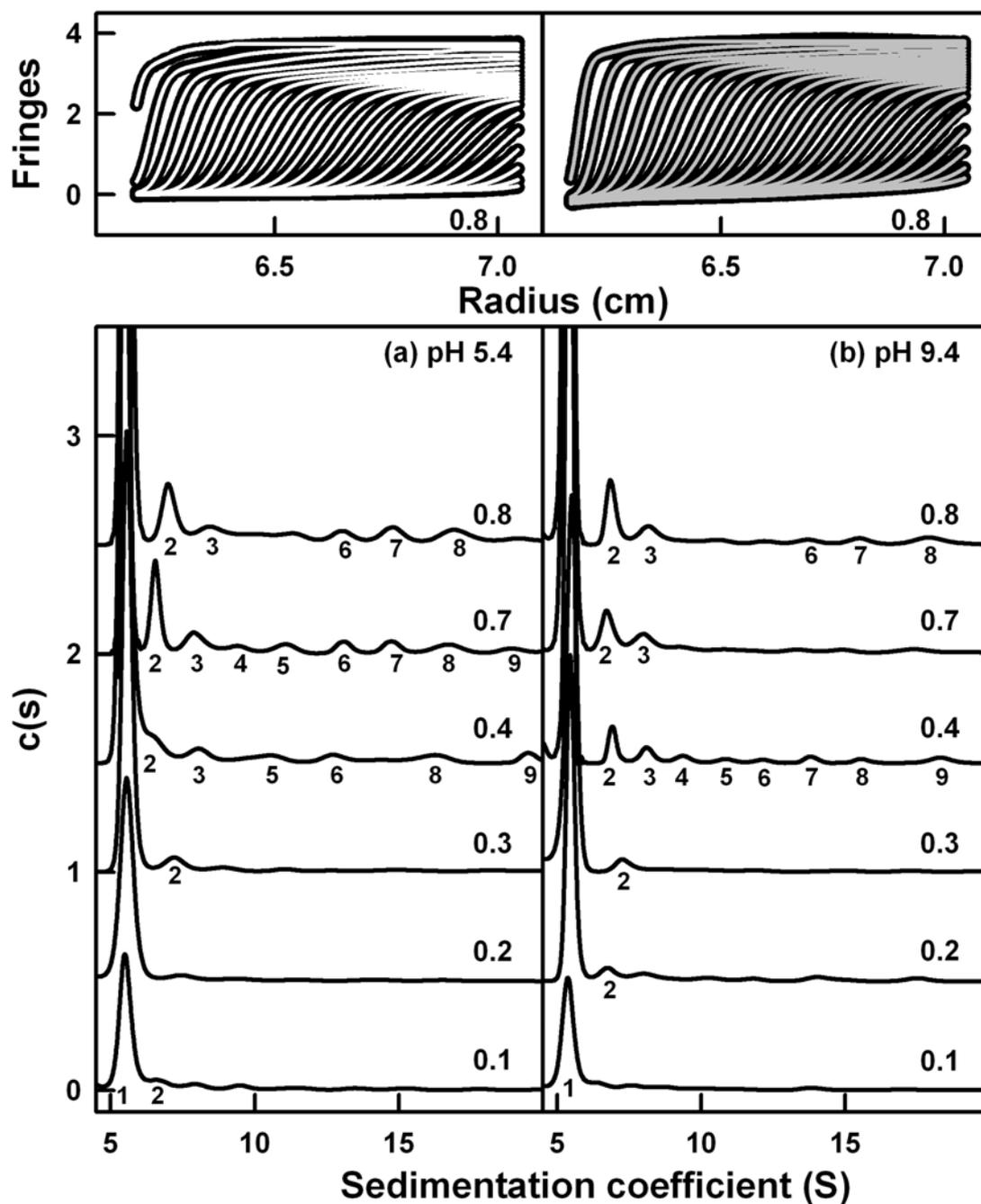
The plots are displaced vertically on the  $c(s)$ -axis and the peak intensities are normalised for clarity. The boundary fits corresponding to 1.0 mg/ml are shown in the upper panels, where every tenth scan is shown. Peak 1 corresponds to monomeric FH, and the peaks attributable to other FH oligomers are labelled from 2 to 9 in order of increasing  $s_{20,w}$  values. Protein concentrations range from 0.1 to 1.0 mg/ml, and are labelled at the right of each  $c(s)$  plot.

In (a), the dashed line at 1.0 mg/ml corresponds to a FH sample that was previously in 350 mM NaCl ([Figure 5.2\(e,f\)](#)). The colour code is given in [Figure 5.2](#).

*et al.*, 2008a). The concentration range at 0.1-1.0 mg/ml FH spanned the physiological FH serum concentration of 0.235-0.81 mg/ml (1.6–5.4  $\mu$ M) observed *in vivo*. However the intensities and resolutions of peaks 3-8 were not consistent at different NaCl concentrations, although more peaks appeared at higher FH concentrations (Figure 5.3; 5.4). This is attributed to the lack of reversibility of oligomer formation, especially at higher FH concentrations (Nan *et al.*, 2008a), and also to possible exchange rate processes involving these peaks.

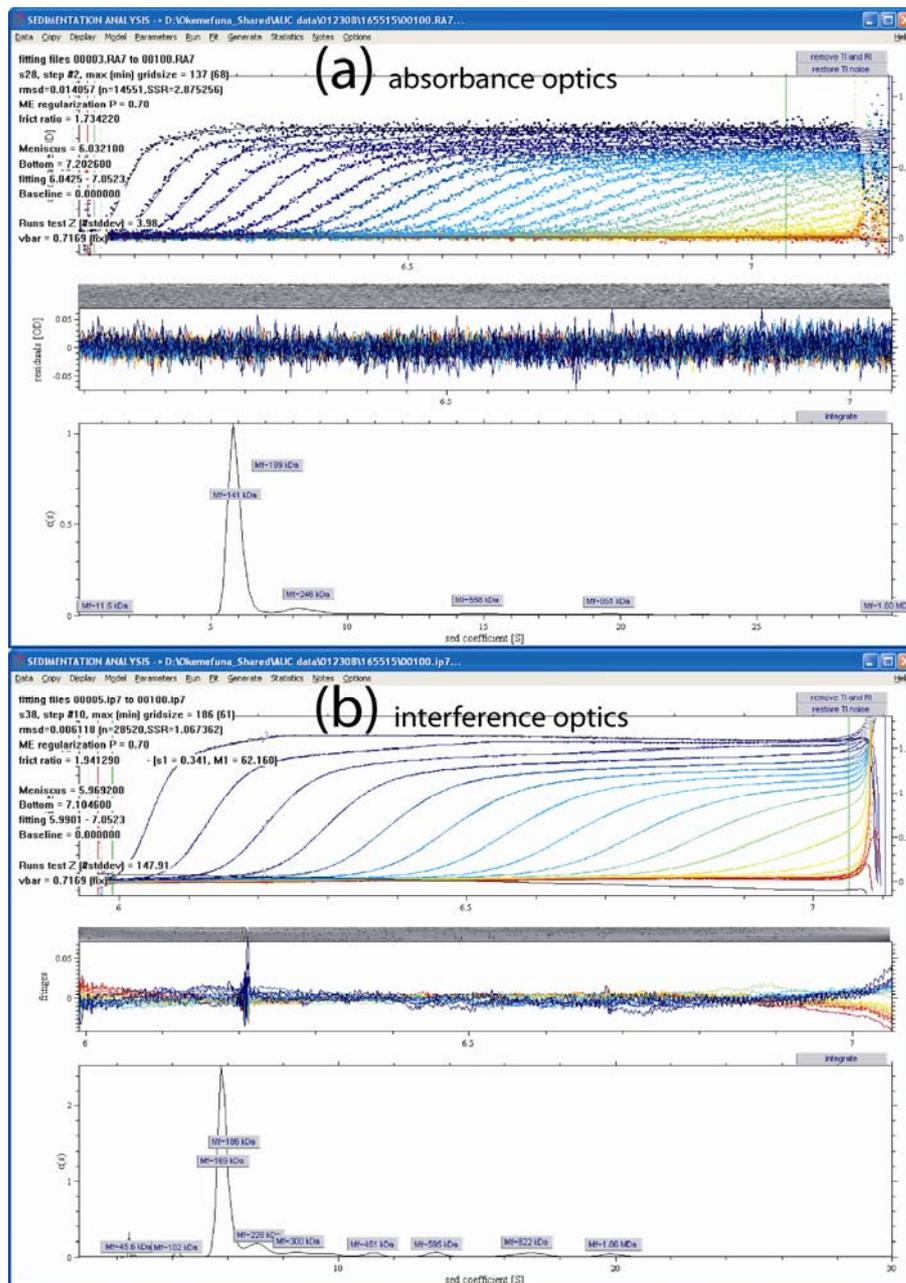
Integration of the  $c(s)$  plots provided the proportion of oligomers corresponding to peaks 2-8 (Figure 5.6). The greatest proportion of oligomer was observed in 50 mM NaCl, ranging from 19% to 34% of the total FH (Figure 5.6(a)). This decreased to a range of 8% to 23% in 137 mM, 250 mM and 350 mM NaCl. The proportion of oligomers increased slightly from 0.1 to 1.0 mg/ml in 50 mM NaCl. No pH dependence of FH oligomerisation was detected, with the proportion of oligomers remaining between 15-17% between pH 5.4 and pH 9.4 (Figure 5.6(b)). These results show that the proportion of FH oligomers is physiologically significant, and that electrostatic interactions promote FH oligomer formation in low salt conditions.

To test whether oligosaccharides affect the FH conformation (Figure 5.1), FH was deglycosylated using PNGase F (Section 5.4.1). SDS-PAGE analysis showed a reduction of the apparent FH molecular weight from 116 kDa to 96 kDa, which is in good accord with the predicted reduction from 154 kDa to 137 kDa from the FH composition. At 0.05 mg/ml and 0.1 mg/ml, velocity runs showed a major FH monomer peak in the  $c(s)$  plot with a  $s_{20,w}$  value of  $5.34 \pm 0.03$  S and an estimated molecular weight of  $128 \pm 14$  kDa. The fitted frictional ratio was determined to be 1.62, which was slightly less than that of 1.75 fitted for native FH. In the  $c(s)$  plots, multiple peaks corresponding to oligomers were observed, although these were shifted to different values compared to those for native FH for reason of the change in mass. These data indicated that carbohydrate has little influence on the FH conformation or its oligomerisation.



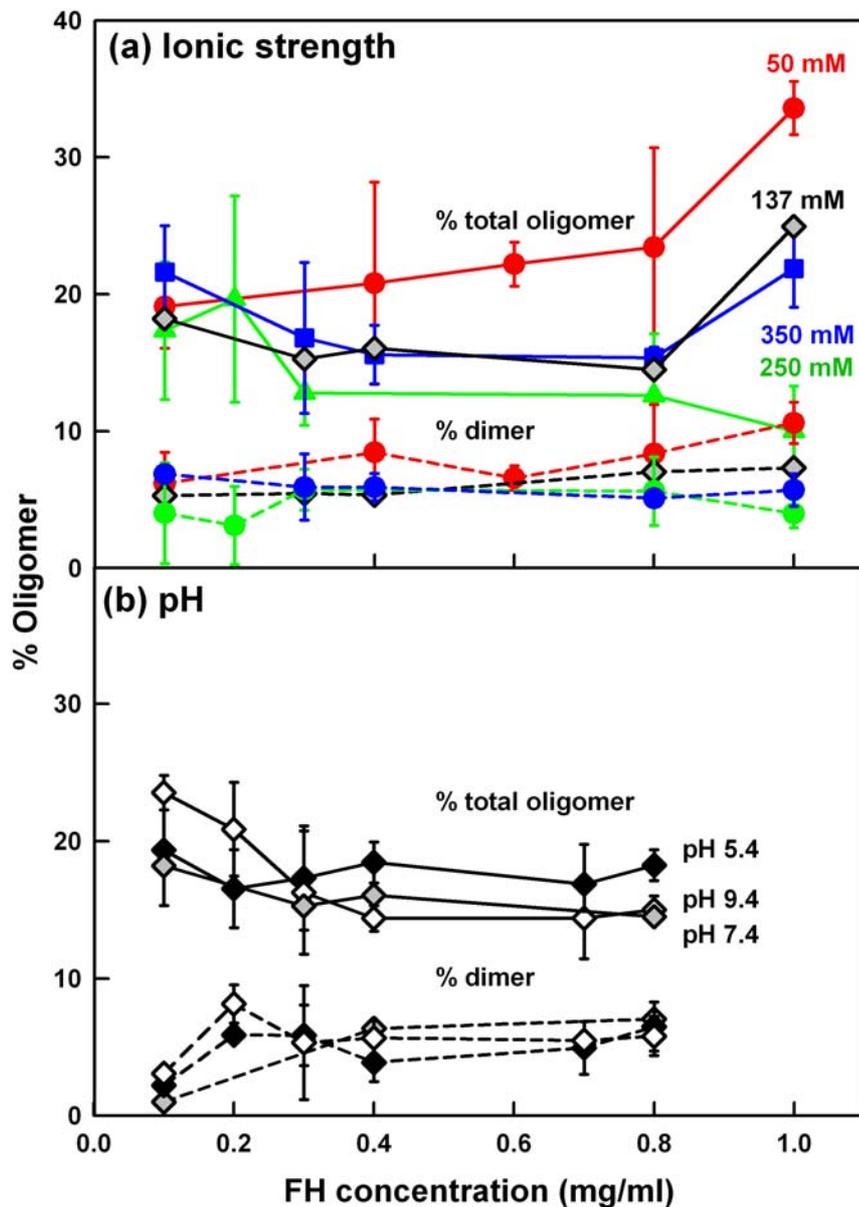
**Figure 5.4**

pH-dependence of the sedimentation of FH oligomers. Experiments are at a rotor speed of 50,000 r.p.m. and correspond to buffers containing 137 mM NaCl at **(a)** pH 5.4 and **(b)** pH 9.4. Protein concentrations range from 0.1 to 0.8mg/ml. Other details follow [Figure 5.3](#).



**Figure 5.5**

Screen shots of SEDFIT sedimentation analyses of FH in 137 mM NaCl, pH 7.4 at a rotor speed of 40,000 r. p. m. for **(a)** absorbance optics and **(b)** interference optics. Boundary fits are shown in the top panel of each screen shot, while  $c(s)$  distributions are shown in the bottom panel. The fit residuals, along with the corresponding bitmap image (grey), are shown in the middle panel. Fit parameters of r.m.s.d. (goodness of fit), frictional ratio, meniscus and bottom, fitting range, baseline and partial specific volume ( $\bar{v}$ ) are shown in the text at the top left corner.



**Figure 5.6**

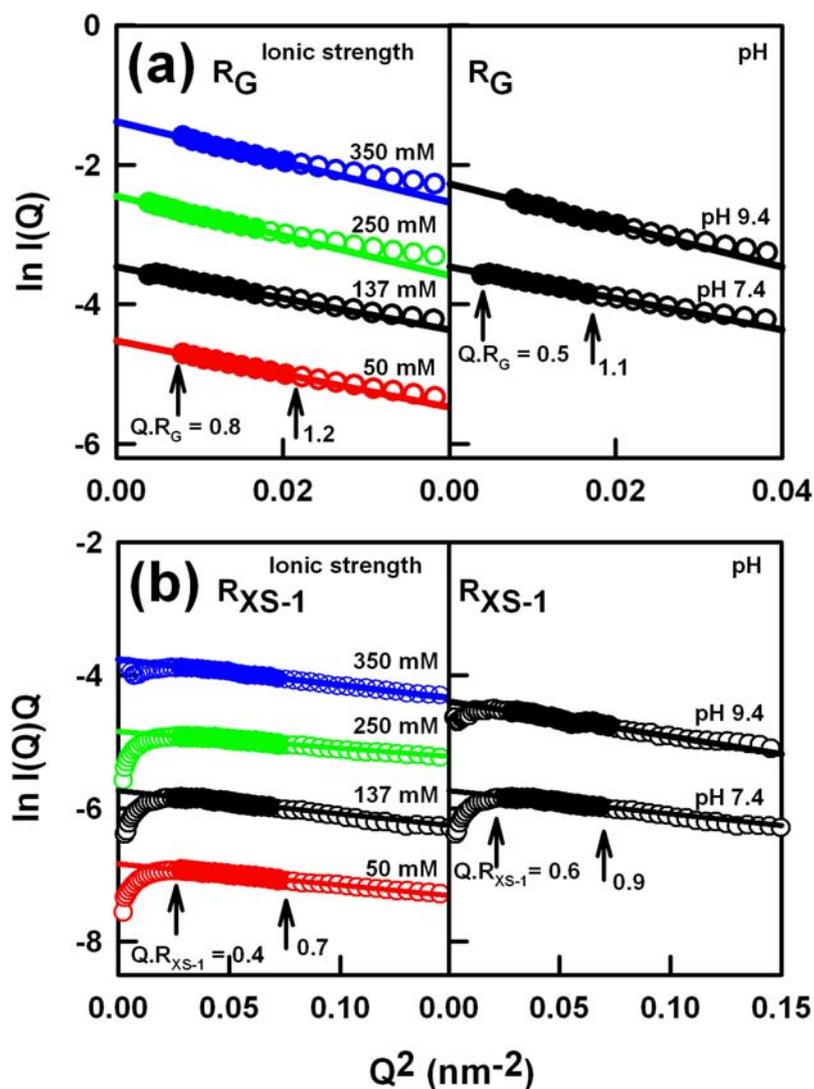
NaCl concentration and pH dependence of the formation of FH dimer and oligomer. At each concentration, the percentage of dimer (dashed lines) and oligomer (solid lines) corresponds to the mean  $\pm$  standard deviation of those obtained at 40,000, 50,000 and 60,000 r.p.m. The dimer corresponds to peak 2 and the oligomer corresponds to peaks 2-8 in [Figure 5.3](#). The colour code is given in [Figure 5.2](#). **(a)** Dependence in four NaCl concentrations with 50 mM, 137 mM, 250 mM and 350 mM NaCl, all at pH 7.4. **(b)** pH dependence in three pH values at pH 5.4 ( $\blacklozenge$ ) and 9.4 ( $\diamond$ ).

### (5.2.2) X-ray scattering data for Factor H

X-ray scattering is a diffraction technique that studies the overall structure of biological macromolecules in random orientations in solution (Perkins *et al.*, 2008). The domain arrangement of FH was assessed by Guinier analyses of the X-ray  $I(Q)$  data in three separate  $Q$  ranges ( $Q = 4 \pi \sin \theta / \lambda$ ;  $2\theta$  = scattering angle;  $\lambda$  = wavelength) to give the radius of gyration  $R_G$  and two cross-sectional radii of gyration  $R_{XS}$  (Perkins *et al.*, 1991; Aslam & Perkins, 2001). The  $R_G$  value monitors the degree of elongation of the 20 SCR domains. The cross-sectional radius of gyration  $R_{XS-1}$  value monitors medium-range spatial relationships between non-neighbouring SCR domains, and this parameter is unique to FH and CR1 amongst the SCR proteins (Furtado *et al.*, 2008). The cross-sectional radius of gyration  $R_{XS-2}$  monitors the degree of bend between adjacent pairs of SCR domains (Fernando *et al.*, 2007).

NaCl concentration effects on FH were studied at 0.4 mg/ml to reduce the effect of reversible dimer formation while obtaining sufficient signal-to-noise ratios (Figure 5.7). The buffers contained 50 mM (PBS), 137 mM (HEPES), 250 mM (PBS) and 350 mM (PBS) NaCl, with the 137 mM data being previously collected by Miss Ruodan Nan (Nan *et al.*, 2008a). Despite the high background in the higher salt buffers, linear Guinier analyses were obtained within satisfactory  $Q.R_G$  and  $Q.R_{XS}$  limits (Figures 5.7(a,b)). Trace amounts of non-specific aggregation were visible in 50 mM and 350 mM NaCl, accordingly a reduced  $Q$  range was used for these  $R_G$  fits (Figure 5.7(a)). In 50 mM, 137 mM, 250 mM and 350 mM NaCl, the mean X-ray  $R_G$  values 8.57 nm, 8.90 nm, 9.22 nm and 9.28 nm (three to six values each) increase in that order (Table 5.1; Figure 5.8(a)). The corresponding  $R_{XS-1}$  values were steadier at 2.54 nm, 2.51 nm, 2.31 nm and 2.71 nm respectively (Table 5.1; Figure 5.8(b)). The  $R_{XS-2}$  values were almost unaffected in a range from 1.43 nm to 1.87 nm (Figure 5.8(c)). Even though the X-ray data correspond to an average of FH monomer, dimer and oligomer, and the  $R_G$  values will be overestimated for that reason, a small increase in the  $R_G$  values occurred with increase in [NaCl] that showed that the FH domain structure becomes more extended at high NaCl concentration. This conformational change agrees with the AUC data.

The X-ray distance distribution function curve  $P(r)$  provides both structural information in real space and an alternative calculation of the Guinier  $R_G$  value following



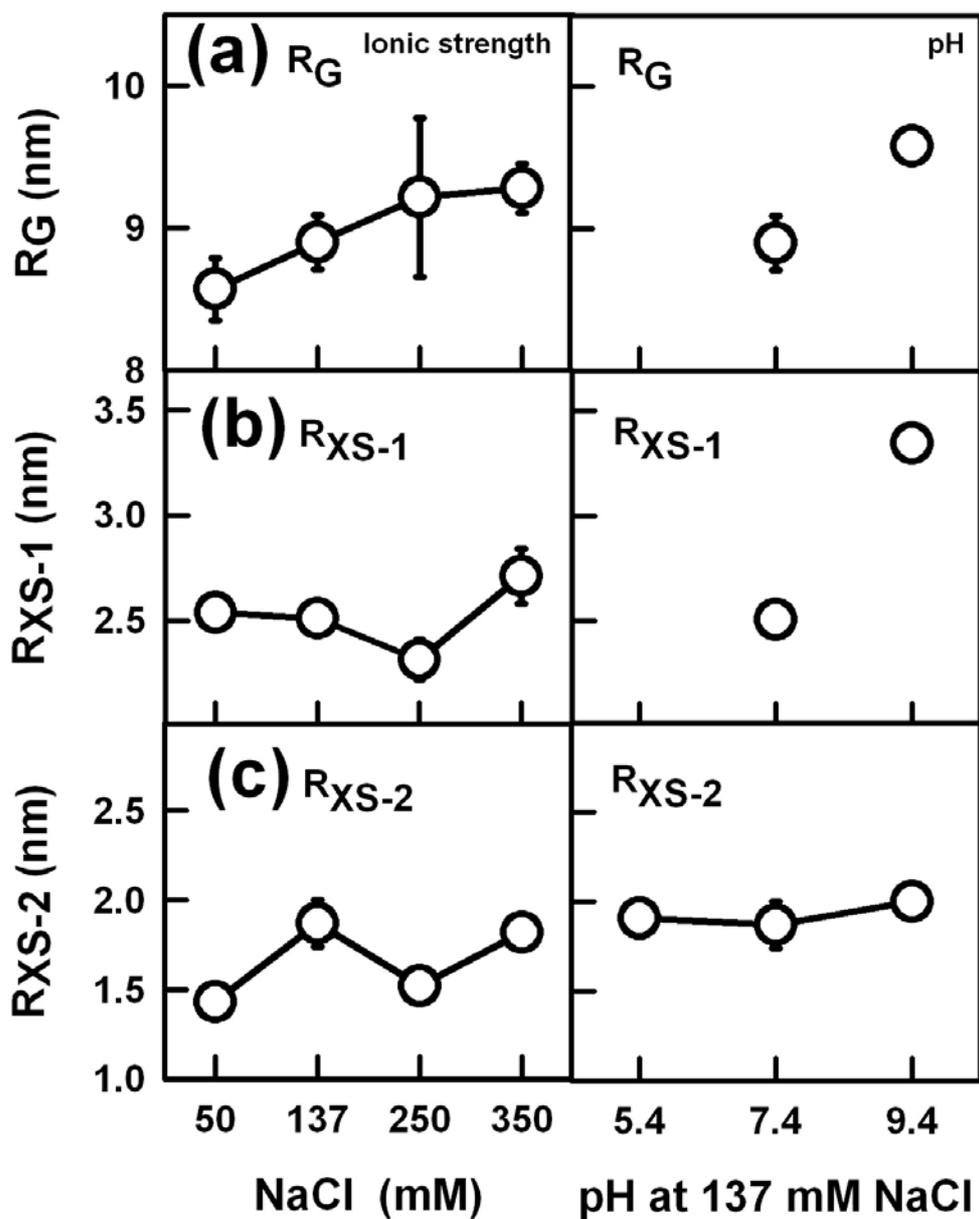
**Figure 5.7**

Determination of the Guinier parameters for FH. The left and right sides of each panel correspond to NaCl concentrations and pH conditions respectively. The curve for 137 mM NaCl and pH 7.4 was used with permission from Miss Ruodan Nan ([Nan et al., 2008a](#)). Filled circles correspond to the  $I(Q)$  values used to determine the best fit lines that give the  $R_G$  and  $R_{XS-1}$  values at 0.4 mg/ml FH. The  $Q \cdot R_G$  and  $Q \cdot R_{XS-1}$  values for the fit ranges are arrowed. The colour code is given in [Figure 5.2](#). (a)  $R_G$  plots using a  $Q$  range of 0.09-0.14 nm<sup>-1</sup> for the 50 mM and 350 mM NaCl buffers and pH 9.4, and 0.06-0.13 nm<sup>-1</sup> for the 137 mM and 250 mM NaCl buffers and pH 7.4. (b)  $R_{XS-1}$  plots using a  $Q$  range of 0.17-0.27 nm<sup>-1</sup>.

an assumption of the value of the maximum dimension  $D_{\max}$ . In buffers containing 50 mM, 137 mM, 250 mM and 350 mM NaCl, the  $R_G$  values were 9.17 nm, 8.57 nm, 9.68 nm and 9.72 nm, which were generally slightly larger but consistent with the Guinier  $R_G$  values, showing that the  $P(r)$  curves refer to the macromolecule being studied (Table 5.1). The maximum  $M$  in the  $P(r)$  curve corresponds to the most frequently occurring interatomic distance within the FH structure. This main peak was almost unaffected in position at  $r$  values of 4.8 nm, 5.4 nm, 4.6 nm and 5.3 nm from 50 mM NaCl to 350 mM NaCl (Figure 5.9). The maximum dimension  $L$  of FH is determined from  $P(r)$  as the value of  $r$  when  $P(r) = 0$  at large  $r$ . The  $L$  value increased from 33 nm in 50 mM NaCl to 38 nm in 250 mM and 350 mM NaCl (Figure 5.9), again showing that the FH domain structure becomes more extended as the NaCl concentration is increased.

pH effects on the X-ray data for FH were studied at pH 5.4 (PBS), pH 7.4 (HEPES) and pH 9.4 (PBS). Limited data were measurable at pH 5.4. The Guinier fits for pH 9.4 showed a slight increase in non-specific aggregation similar to that seen in 50 mM NaCl (Figure 5.7(a)). FH had an  $R_G$  value of 9.58 nm and an  $R_{XS-1}$  value of 3.35 nm (two values), both of which were higher than at pH 7.4 (Table 5.1; Figure 5.8(a,b)). The  $R_{XS-2}$  value remained relatively unchanged at 1.91 nm at pH 5.4 and 2.00 nm at for pH 9.4 (Table 5.1; Figure 5.8(c)). The  $P(r)$  curve resulted in an  $R_G$  value of 9.55 nm at pH 9.4, in good agreement with the Guinier analyses. The  $M$  value increased from 5.4 nm at pH 7.4 to 6.0 nm at pH 9.4, while  $L$  remained unchanged at 32 nm at pH 9.4 compared to pH 7.4 (Figure 5.9). The observed tendency for slight aggregation at the lowest  $Q$  values and the larger  $R_G$  and  $R_{XS-1}$  values may explain the apparent predominantly dimeric FH structure that was first observed by X-ray scattering (Perkins *et al.*, 1991). In the 1991 study, FH was eluted from the MRCOX23 monoclonal column using alkaline conditions. No information on the FH conformation at comparably high pH was available from the present data for analysis.

To assess the effect of carbohydrate on the FH structure, deglycosylated FH was studied in 137 mM NaCl at pH 7.4. Linear Guinier analyses gave  $R_G$ ,  $R_{XS-1}$  and  $R_{XS-2}$  values of 9.48 nm, 2.93 nm and 1.89 nm respectively (4-7 determinations) (Table 1; Figures 5.10(a,b)). The  $P(r)$  curve resulted in an  $R_G$  value of 9.73 nm, and  $M$  and  $L$  values of 5.7 nm and 32 nm respectively (Figure 5.10(e)). The similarity with the



**Figure 5.8**

NaCl concentration and pH dependence of the Guinier parameters for FH. The left and right sides of each panel correspond to NaCl concentrations and pH conditions respectively. Other details follow [Figure 5.7](#).

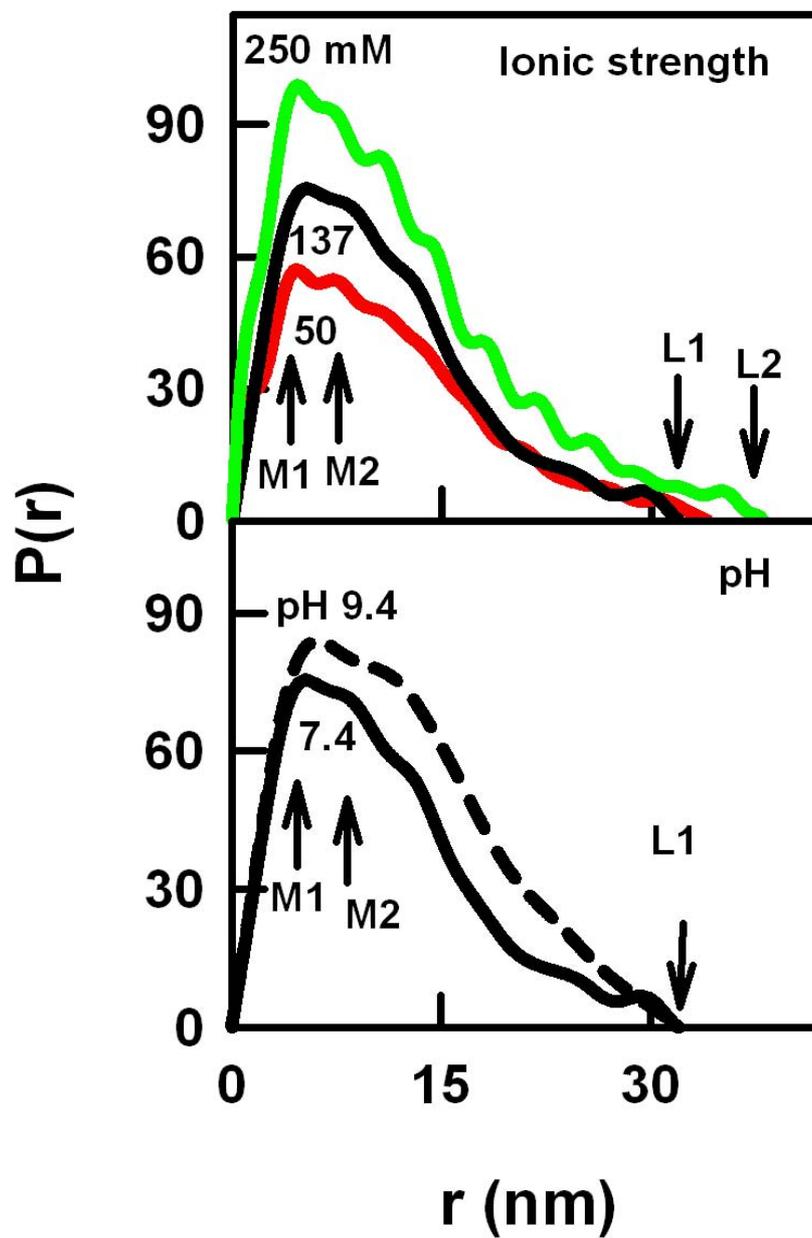
(a) Dependence of the  $R_G$  values on NaCl concentration and pH.

(b,c) Dependence of the  $R_{XS-1}$  and  $R_{XS-2}$  values on NaCl concentration and pH. The  $R_{XS-2}$  values were determined using a  $Q$  range of  $0.50\text{-}0.80\text{ nm}^{-1}$  (not shown).

corresponding Guinier values of 8.90 nm, 2.51 nm and 1.79 nm for FH and the  $M$  and  $L$  parameters of 5.7 nm and 32 nm for wild-type FH in 137 mM NaCl at pH 7.4 (Table 5.1) indicated that FH adopted only a slightly more elongated structure after carbohydrate removal, and that its carbohydrate content plays no major role in determining the folded-back FH structure. This is in accord with the AUC data.

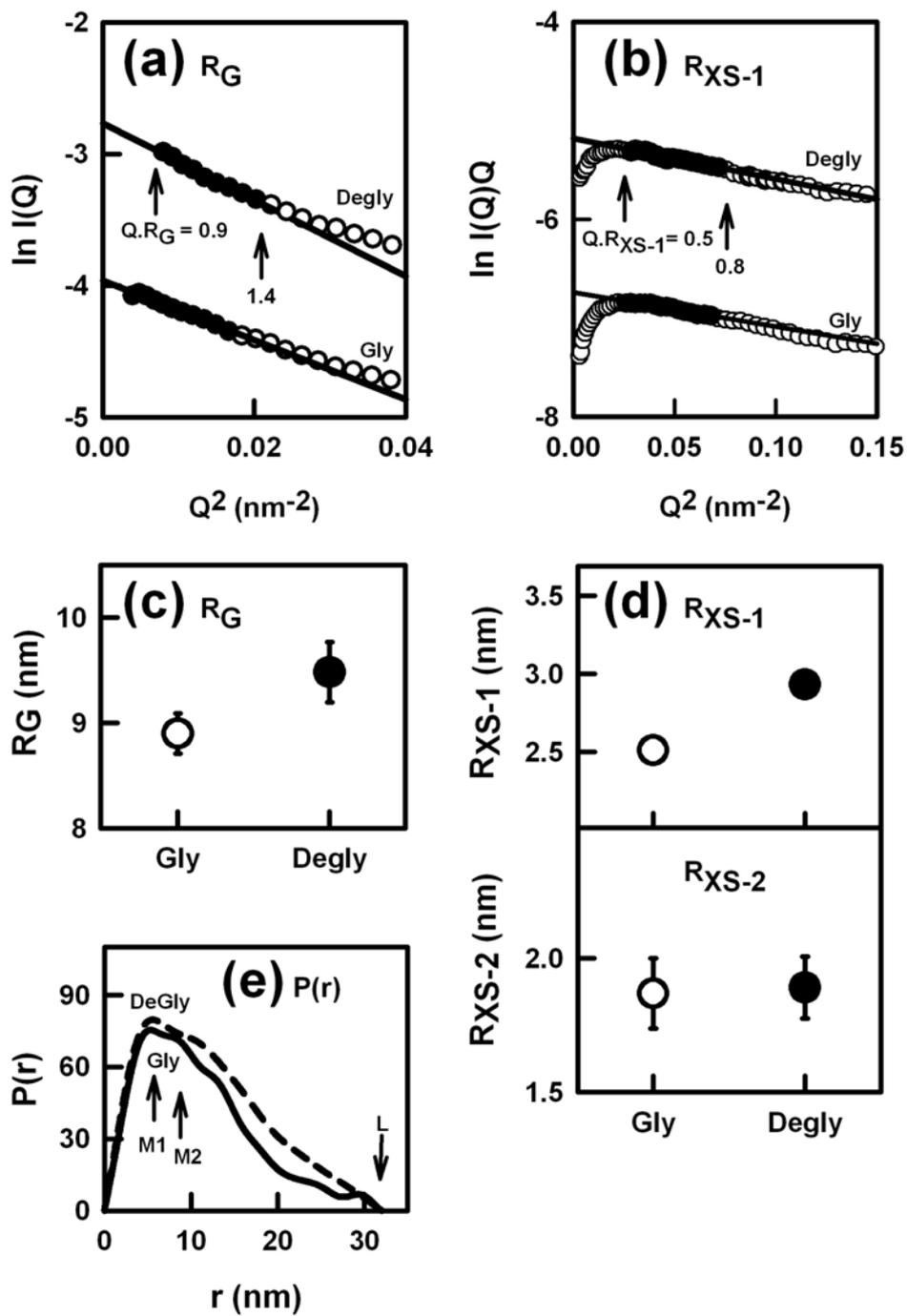
### (5.2.3) Constrained scattering modelling of monomeric Factor H

Solution structures for FH in 50 mM, 137 mM and 250 mM NaCl were determined by constrained scattering modelling, where the constraints were provided by the use of nine updated SCR homology models and 11 crystal and NMR structures for the FH SCR domains (Section 5.4.4). These are the same as those used to model SCR-1/5 and SCR-16/20 (Chapter 4). The original constrained modelling of FH was the first application of this method for any SCR protein (Aslam & Perkins, 2001). At that time, only three SCR structures were known, and the effect of FH oligomer formation and aggregation on the X-ray scattering curve was incompletely understood. Here, the experimental X-ray curves used for curve fitting utilised FH samples in which aggregation effects are absent, although FH oligomers are present and will contribute to the scattering curve. Conformational models for intact FH that best fitted the X-ray data were identified from trial-and-error searches of 5,000 randomised structures. These 5,000 structures were generated using molecular dynamics simulations to create libraries of peptide linker conformations, which were sampled at random and joined with the 20 SCR domain models. Calculation of scattering curves from the 5,000 models showed that a sufficient number of models had been tested (Figure 5.11(a,b)). The comparisons of the  $R$ -factor (goodness of fit) and  $R_{XS-I}$  values with the  $R_G$  values using three experimental curves in 50 mM, 137 mM and 250 mM NaCl buffer, and including two more for a pH 9.4 buffer and deglycosylated FH for comparison (Figure 5.11(a,b)), showed that the models encompassed the experimental  $R_G$  and  $R_{XS-I}$  values. This showed that the 5,000 models were sufficient to determine a best-fit FH solution structure. The best-fit structures were selected on the criteria that they satisfied the  $R_G$  and  $R_{XS-I}$  values to within 5% and showed the lowest  $R$ -factor values. Here, unlike the original fits which



**Figure 5.9**

Distance distribution function  $P(r)$  analyses for FH. The peaks in the  $P(r)$  curves are denoted by  $M1$  and  $M2$ , and the maximum length is denoted by  $L1$  and  $L2$ . Other details follow [Figure 5.7](#).



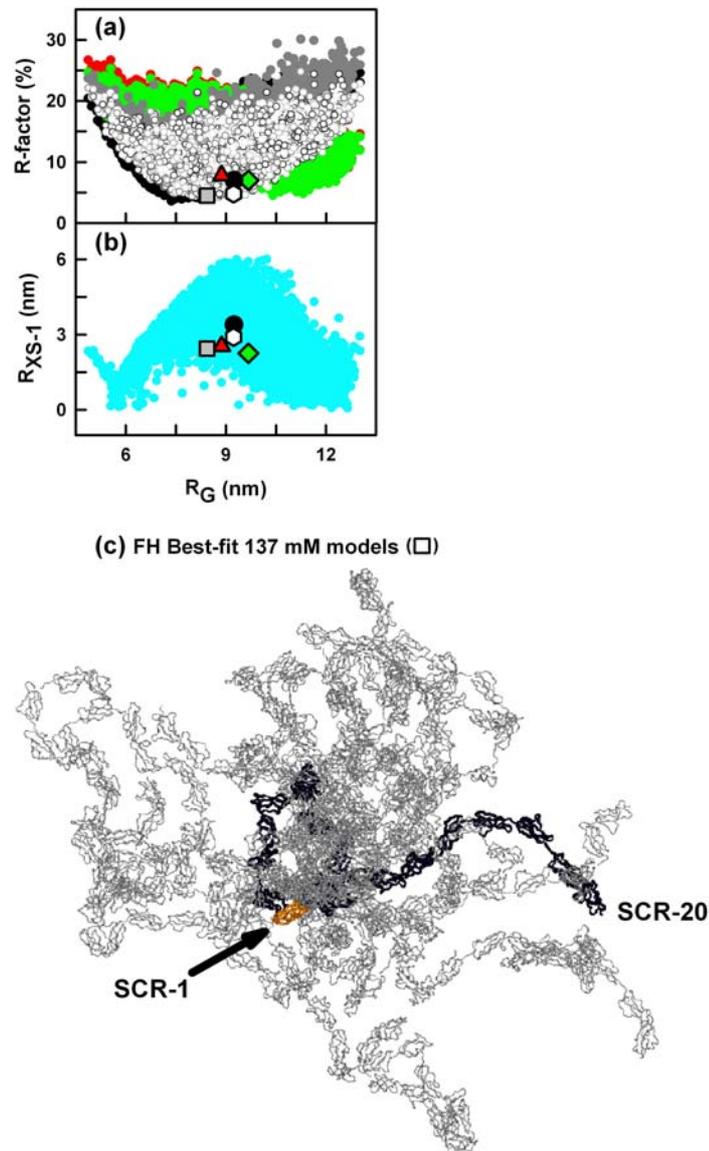
**Figure 5.10**

Comparison of the Guinier fits and distance distribution  $P(r)$  analyses for native (Gly) and deglycosylated (Degly) FH. The FH concentration is 0.4 mg/ml. Details of (a,b) follow Figure 5.7, (c,d) follow Figure 5.8 and (e) follows Figure 5.9.

showed 31 best-fit R-factors of  $19 \pm 3\%$  (Aslam & Perkins, 2001), all best-fit structures resulted in R-factors of less than 10%. This reduction demonstrated the effect of using better experimental data and improved SCR domain models.

The best-fit models for FH are partially folded-back and bent in appearance. The superimposition upon SCR-1 of the 17 best fit structures for 137 mM NaCl (Figure 5.11(c)) showed that this feature is common in all the structures. The best-fit models showed that the  $R_G$  values increased with [NaCl] from 8.79 nm and 8.44 nm to 9.67 nm as this increased from 50 mM and 137 mM to 250 mM, in agreement with experiment. The best-fit model from the five searches were in similar locations in the distributions of Figures 5.11(a,b), accordingly all five experimental curves resulted in similar overall structures for FH. This confirms that the conformational changes in FH between these experimental conditions, though significant, were not large enough to yield distinct groups of best-fit models (Table 5.1). Only the models for 50 mM, 137 mM and 250 mM NaCl were deposited in the Protein Data Bank (PDB), with codes 3gau, 3gav and 3gaw respectively. There is some partial overlap of the best-fit models. For example, five out of the 27 best-fit models for the 50 mM NaCl and 137 mM buffers were identical, while there was no overlap between the 137 mM and 250 mM NaCl models.

For 137 mM NaCl and pH 7.4, the R-factor values for the 17 best-fit models (Figure 5.11(c)) ranged from 4.5-9.6 %. These values are much lower than those for the original four best-fit models in similar buffer that gave R-factor values of 12-15% (Aslam & Perkins, 2001). The visual comparison of the agreement between the experimental and modelled  $I(Q)$  curves showed a good fit from low  $Q$  out to a  $Q$  value of  $1.5 \text{ nm}^{-1}$ , in which the modelled  $I(Q)$  curve decayed uniformly as  $Q$  increased (Figure 8(b)). The  $P(r)$  curves (insets in Figure 5.12(b)) showed maxima  $M$  and maximum dimensions  $L$  at the same  $r$  values, in agreement with the experimental values (Figures 5.9; 5.10(e)). This is much improved over the original fits for which the agreement extended to a  $Q$  value of only  $0.4 \text{ nm}^{-1}$  and the modelled curves did not decay uniformly at large  $Q$  (Aslam & Perkins, 2001). The new FH curve fits show similar agreements in  $Q$  ranges with experiment as those for CR1 and CR2 (Gilbert *et al.*, 2006; Furtado *et al.*, 2008).

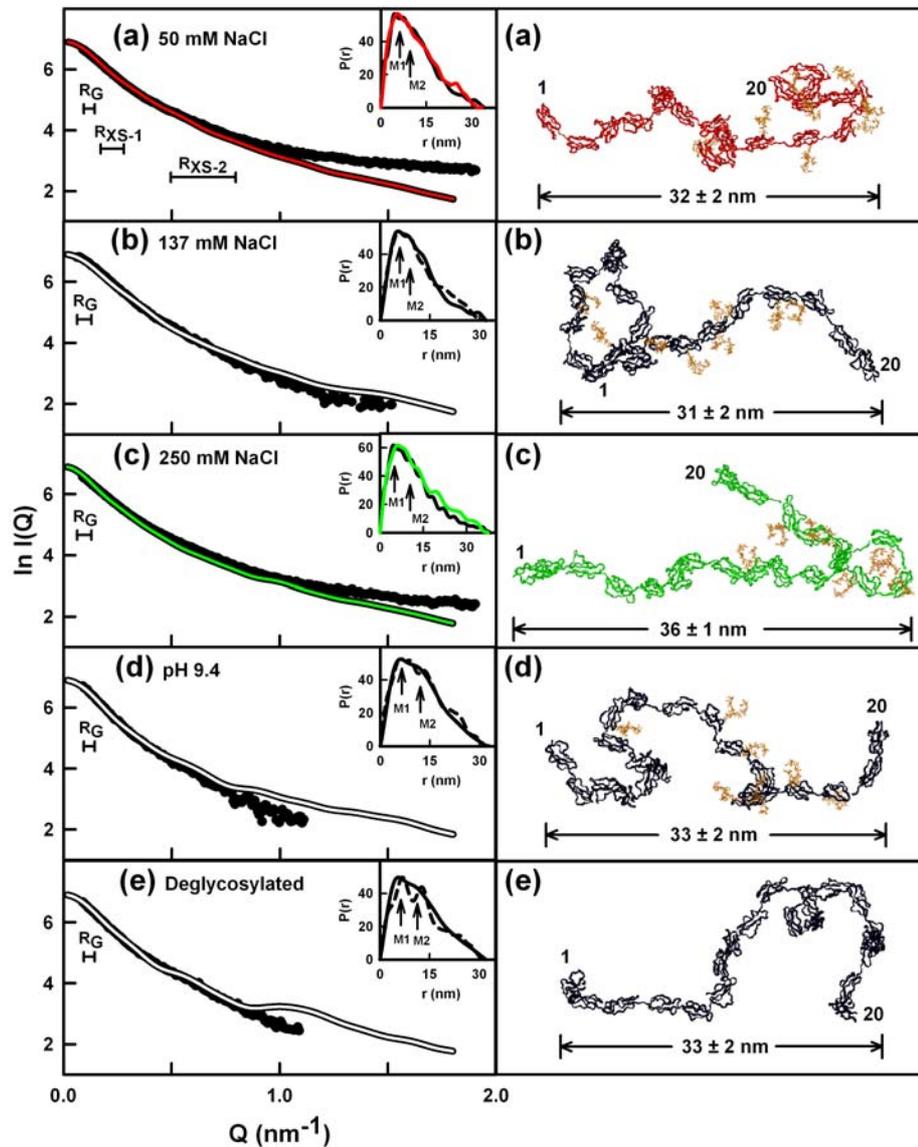


**Figure 5.11**

Outcome of five constrained modelling searches to determine a FH solution structure. In each modelling search, a library of 5,000 conformationally-randomised models were fitted to a given FH scattering curve at 0.4 mg/ml in 50 mM NaCl (red), 137 mM NaCl (grey), 250 mM NaCl (green), pH 9.4 (black), and deglycosylated FH (white). In (a,b), the distribution of the fits to the five experimental curves is shown. The five highlighted symbols correspond to the best-fit model in each fit. (a) Comparison of the  $R$ -factor values with the  $R_G$  values. (b) Comparison of the  $R_{XS}$  values with the  $R_G$  values. (c) The 17 best-fit models for the 137 mM NaCl fit are shown superimposed on SCR-1 (orange). For clarity, the 19 remaining SCR domains of the best-fit model are highlighted in black up to SCR-20 (labelled).

As an independent check of the scattering modelling, the  $s_{20,w}^{\circ}$  values of the best-fit FH models were calculated using HYDROPRO (Table 5.1). For the best-fit models, the calculated  $s_{20,w}^{\circ}$  values ranged from 4.85-5.18 S for 50 mM and 137 mM NaCl buffers, and decreased to a range of 4.59-4.89 S for 250 mM NaCl buffer (Table 5.1). The best-fit models gave calculated  $s_{20,w}^{\circ}$  values of 4.91 S, 5.01 S and 4.59 S for 50 mM, 137 mM and 250 mM NaCl buffers in that order. While the calculated values were 0.6-0.8 S lower than the experimental values of 5.66 S, 5.56 S and 5.41 S respectively, the calculated  $s_{20,w}^{\circ}$  values generally decreased with increase in NaCl concentration in agreement with the experimentally-observed trend. Since good agreements had previously been seen with the calculated and experimental  $s_{20,w}^{\circ}$  values for CR1 and CR2 (Gilbert *et al.*, 2006; Furtado *et al.*, 2008; Li *et al.*, 2008), the discrepancy is most likely due to the influence of FH oligomers on the modelling of the X-ray scattering curves. The presence of these oligomers would result in good fits to FH models that are slightly longer with slightly increased calculated  $R_G$  values. As such, the calculated  $s_{20,w}^{\circ}$  values of models would be slightly reduced. Hence the contribution of oligomers to the X-ray fits accounts for the discrepancy between experimental and calculated  $s_{20,w}^{\circ}$  values in Table 5.1. In this sense, the calculations support the outcome of the scattering modelling (Figure 5.12). The influence of oligomers on the FH curve fits suggests that the actual FH monomer conformation in solution may be more compact with an even lower  $R_G$  value than those reported in Figures 5.7 and 5.8.

The determination of five sets of best-fit FH models that showed partially folded-back SCR arrangements (Table 5.1) permitted the evaluation of the inter-domain orientations that resulted in the formation of these FH structures. This evaluation is limited in scope by the algorithm used to generate the 5,000 randomised FH conformations, because this created the linkers in extended conformations that were permitted to reorientate themselves. The angle between the longest axes of two adjacent SCR domains shows the degree of bend between them (Fernando *et al.*, 2007; Chapter 4). For the 50 mM, 137 mM and 250 mM NaCl buffers, the mean inter-domain angle was similar at  $119^{\circ} \pm 34^{\circ}$ ,  $121^{\circ} \pm 32^{\circ}$  and  $119^{\circ} \pm 30^{\circ}$  respectively in the eight best-fit models for each of these fit searches. As an entirely linear SCR arrangement corresponds to a mean angle of  $180^{\circ}$ , this shows that adjacent pairs of SCR domains in



**Figure 5.12**

Best-fit FH X-ray scattering curve fit for the five experimental conditions.

(a) 50 mM NaCl; (b) 137 mM NaCl; (c) 250 mM NaCl; (d) pH 9.4; (e) deglycosylated FH. The left panels show the fitted experimental scattering data as black points for 0.4 mg/ml FH. The continuous lines correspond to the best-fit modelled curves. The  $Q$  ranges employed in the experimental  $R_G$  and  $R_{XS}$  analyses are denoted by horizontal bars. The insets at the upper right compare the modelled  $P(r)$  curves as a coloured or dashed line with the experimental  $P(r)$  curve in continuous black. The right panels show ribbon views of the best-fit FH model. Oligosaccharides where present are shown in orange. The positions of SCR-1 and SCR-20 are indicated, along with the maximum lengths of the FH models.

**Table 5.1. X-ray scattering and sedimentation coefficient data for monomeric FH and their modelling**

	Filter	Models	Spheres	$R_G$ (nm)	$R_{XS-1}$ (nm)	$R_{XS-2}$ (nm)	$D_{max}$ (nm)	$R$ factor (%)	$s_{20,w}^0$ (S)
<b>Experimental data</b>									
FH in 50 mM NaCl buffer				8.57 ± 0.22; 9.17 ± 0.12	2.54 ± 0.07	1.43 ± 0.05	34		5.66 ± 0.05
FH in 137 mM NaCl buffer				8.90 ± 0.19; 8.57 ± 0.71 <sup>a</sup>	2.51 ± 0.06	1.79 ± 0.01	32		5.56 ± 0.08
FH in 250 mM NaCl buffer				9.22 ± 0.56 9.68 ± 0.49 <sup>a</sup>	2.31 ± 0.09	1.52 ± 0.09	38		5.41 ± 0.06
FH in 350 mM NaCl buffer				9.28 ± 0.17 9.72 ± 0.25 <sup>a</sup>	2.71 ± 0.13	1.82 ± 0.08	38		5.31 ± 0.06
FH in pH 5.4 buffer				n.a	n.a	1.91 ± 0.11	n.a		5.54 ± 0.05
FH in pH 9.4 buffer				9.58 ± 0.08 9.55 ± 0.01 <sup>a</sup>	3.35 ± 0.12	2.00 ± 0.04	32		5.40 ± 0.05
Deglycosylated FH				9.48 ± 0.29 9.73 ± 0.14 <sup>a</sup>	2.93 ± 0.08	1.89 ± 0.11	32		5.34 ± 0.03
<b>Modelling of FH</b>									
Library of 5,000 FH models	None	5,000	1948-2284 <sup>b</sup>	4.88 - 12.83	0.08 - 6.00	0.04 - 2.65	n.a	3.6 - 30.1	n.a.
FH in 50 mM NaCl buffer	$R_G$ , $R_{XS-1}$ and $R$ factor	10	2087-2170	8.18-8.99	2.42-2.66	1.46-1.84	30-34	7.8-9.9	4.86-5.18
FH in 50 mM NaCl buffer	Best fit	1	2170	8.79	2.53	1.65	33	7.8	4.91
FH in 137 mM NaCl buffer	$R_G$ , $R_{XS-1}$ and $R$ factor	17	2087-2220	7.84-8.59	2.41-2.62	0.89-2.19	27-34	4.5-9.6	4.85-5.18
FH in 137 mM NaCl buffer	Best fit	1	2207	8.44	2.44	1.59	32	4.5	5.01
FH in 250 mM NaCl buffer	$R_G$ , $R_{XS-1}$ and $R$ factor	8	2127-2254	9.26-9.67	2.21-2.38	1.40-1.80	35-38	6.4-9.2	4.59-4.89
FH in 250 mM NaCl buffer	Best fit	1	2158	9.67	2.25	1.47	38	7.0	4.59
FH in pH 9.4 buffer	$R_G$ , $R_{XS-1}$ and $R$ factor	4	2191-2212	9.17-9.55	3.26-3.39	1.72-1.77	29-34	7.0-10.0	4.61-4.73
FH in pH 9.4 buffer	Best fit	1	2201	9.17	3.39	1.76	34	7.0	4.73
Deglycosylated FH	$R_G$ , $R_{XS-1}$ and $R$ factor	10	2087-2230	9.05-9.82	2.80-3.06	1.69-2.04	30-36	4.8-9.8	4.07-4.33
Deglycosylated FH	Best fit	1	2191	9.14	2.90	1.82	32	4.8	4.18

<sup>a</sup> The first value is from the Guinier  $R_G$  analyses (Figures 5.7; 5.8 and 5.10(a-d)); the second is from the GNOM  $P(r)$  analyses (Figures 5.9 and 5.10(e)).

<sup>b</sup> The optimum number of hydrated spheres predicted from the sequence is 2074.

n.a. not available

FH will be on average bent in their structure. These mean angles also show that the appearance of more elongated structures with increase in NaCl concentration will result from subtle changes in the inter-SCR angles, potentially involving very few residues (Figure 5.13). This is illustrated by ranking the inter-domain angles for the three best-fit models for each condition studied according to the spread between 0°-180° in contiguous bands of 45° (Figure 5.13; Table 5.2; Section 5.4.4). The six longest linkers with lengths of 5-8 residues occur in the central part of FH between SCR-7/8, SCR-10/15 and SCR-18/19. The best-fit FH models suggest that some of these inter-SCR angles may be much reduced to less than 90° in order to form partially folded-back structures (Table 5.2; Figure 5.13). If so, these would be similar to the low inter-SCR angles seen in the crystal structures for CR2 SCR-1/2 and the three-SCR fragment of complement factor B (Szakonyi *et al.*, 2001; Prota *et al.*, 2001; Mildner *et al.*, 2007).

### (5.3) Conclusions

This study shows that the conformational and oligomeric properties of intact FH depend on NaCl concentration and pH, but less so for its glycosylation. Firstly, the partially folded-back arrangement of 20 SCR domains in FH shows a dependence on electrostatic effects that was previously unknown. This means that the FH conformation will depend on its local environment, and this may be important for the mechanism of its regulatory action through interaction with C3b. Secondly, FH is now shown to be oligomeric in a range of solution conditions. The extent of FH oligomerisation also depends on its local environment, and this may also be important for its role as a complement regulator. These observations appear to be unique to FH, as neither of these two features appears for two other large multi-SCR proteins, CR1 and CR2, both of which are membrane proteins unlike FH. For example, soluble recombinant CR2 shows no difference in conformation between 50 mM and 137 mM NaCl (Li *et al.*, 2008).

Charged groups in FH are able to mediate a slightly more compact arrangement of SCR domains within FH. This may be important for regulation. For example, the polyanionic host cell surface may cause FH to alter its conformation after FH is bound to it. The SCR-1/4 domains of FH interact with its ligand C3b, with a second binding site

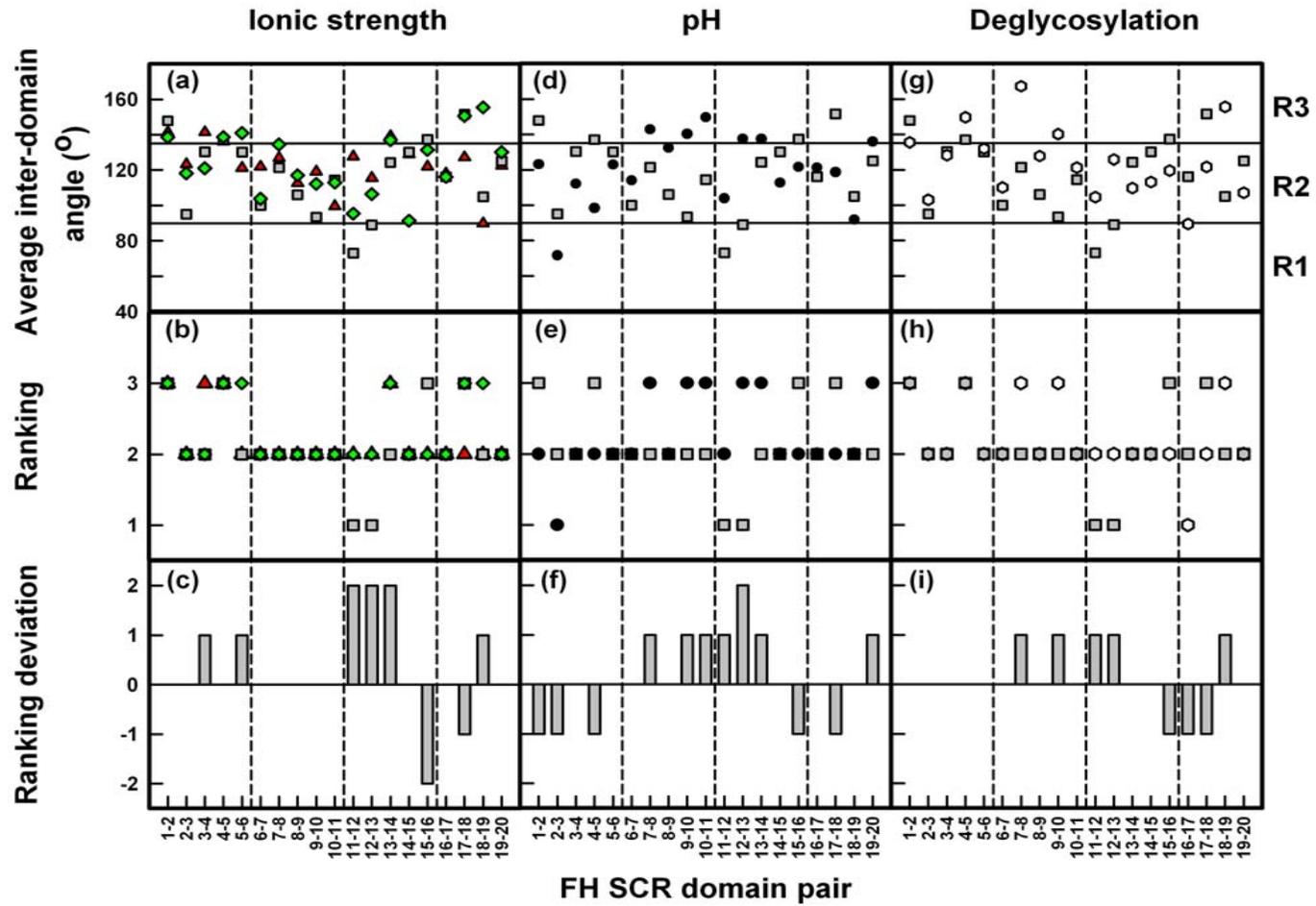


Figure 5.13

(Figure legend overleaf)

(Figure 5.13 legend)

Summary of the effects of NaCl concentration, pH and deglycosylation on FH inter-domain orientation. Average inter-domain angles are shown for the three best-fit FH models at each condition. The angles are ranked 1, 2 or 3 according to the size ranges represented by the areas under the thick black lines R1, R2 and R3.

**(a)** The average inter-domain angle distribution for models in 50 mM NaCl (**▲**), 137 mM NaCl (**■**) and 250 mM NaCl (**◆**) buffers.

**(b)** The corresponding ranking distributions of inter-domain angles for 50-250 mM NaCl. A relative lack of change is indicated by superimposition of the symbols for each condition. The colour scheme follows that of **(a)**.

**(c)** A bar chart summarizing the ranking deviations from a normal orientation in 137 mM NaCl buffer. The height of the bar corresponds to the degree of change. Positive and negative deviations indicate changes towards a more extended or more folded-back orientation respectively.

**(d-f)** Inter-domain angle distributions, ranking distributions and ranking deviations at pH 7.4 (**■**) and 9.4 (**●**). The ranking deviations are shown as deviations from a normal orientation at pH 7.4.

In **(g-i)** for the corresponding ranking analyses for native (**■**) and deglycosylated FH (**○**), native FH is the normal orientation.

**Table 5.2. Summary of inter-SCR domain angles for the best-fit solution structures of monomeric FH**

SCR domains	Average inter-domain angles for best-fit FH models									
	50 mM NaCl buffer		137 mM NaCl buffer		250 mM NaCl buffer		pH 9.4 buffer		Deglycosylated FH	
	Average angle (°)	Ranking <sup>a</sup>	Average angle (°)	Ranking <sup>a</sup>	Average angle (°)	Ranking <sup>a</sup>	Average angle (°)	Ranking <sup>a</sup>	Average angle (°)	Ranking <sup>a</sup>
1-2	141.3 ± 29.6	3	147.9 ± 22.9	3	140.5 ± 10.4	3	123.1 ± 10.2	2	135.6 ± 31.4	3
2-3	123.3 ± 31.8	2	95.1 ± 20.0	2	101.6 ± 50.6	2	71.6 ± 49.5	1	103.0 ± 45.9	2
3-4	141.4 ± 42.8	3	130.4 ± 19.6	2	132.3 ± 33.3	2	112.1 ± 6.6	2	128.3 ± 34.4	2
4-5	136.1 ± 16.5	3	137.0 ± 14.7	3	143.0 ± 20.7	3	98.3 ± 4.1	2	149.7 ± 19.2	3
5-6	121.2 ± 11.6	2	130.2 ± 10.4	2	127.4 ± 26.6	2	122.9 ± 25.4	2	131.9 ± 24.5	2
6-7	121.9 ± 7.2	2	100.0 ± 44.4	2	108.0 ± 3.2	2	113.9 ± 21.5	2	110.0 ± 43.4	2
7-8	126.9 ± 54.9	2	121.4 ± 18.0	2	118.8 ± 28.8	2	142.9 ± 4.2	3	167.2 ± 1.2	3
8-9	112.5 ± 14.2	2	106.1 ± 6.6	2	94.6 ± 36.6	2	132.4 ± 37.1	2	127.8 ± 43.1	2
9-10	119.2 ± 61.1	2	93.4 ± 32.7	2	106.2 ± 32.8	2	140.2 ± 26.5	3	140.2 ± 24.8	3
10-11	99.6 ± 39.0	2	114.4 ± 22.8	2	104.8 ± 9.4	2	149.7 ± 14.2	3	121.2 ± 38.3	2
11-12	127.7 ± 37.7	2	73.1 ± 32.1	1	113.4 ± 20.6	2	103.8 ± 11.2	2	104.5 ± 36.3	2
12-13	115.5 ± 48.2	2	89.0 ± 12.3	1	107.0 ± 6.9	2	137.6 ± 40.7	3	125.8 ± 19.0	2
13-14	139.1 ± 32.9	3	124.3 ± 31.4	2	133.1 ± 40.5	2	137.4 ± 7.0	3	109.7 ± 37.2	2
14-15	128.9 ± 27.3	2	130.1 ± 18.9	2	132.6 ± 36.0	2	112.6 ± 36.5	2	113.1 ± 10.9	2
15-16	122.0 ± 62.6	2	137.4 ± 18.5	3	129.8 ± 9.1	2	121.6 ± 62.4	2	119.4 ± 37.9	2
16-17	118.1 ± 70.2	2	116.2 ± 63.2	2	99.0 ± 35.7	2	121.2 ± 57.8	2	89.3 ± 25.7	1
17-18	127.2 ± 36.0	2	151.5 ± 11.0	3	142.5 ± 12.3	3	118.6 ± 71.9	2	121.7 ± 30.7	2
18-19	89.8 ± 45.1	2	104.8 ± 56.8	2	147.2 ± 19.3	3	91.8 ± 57.9	2	155.5 ± 9.5	3
19-20	122.5 ± 61.5	2	125.0 ± 24.7	2	143.9 ± 8.5	3	135.9 ± 21.3	3	107.0 ± 42.8	2

<sup>a</sup> The ranking is defined as 0 if  $0^\circ \leq \text{inter-SCR angle} \leq 45^\circ$ , 1 if  $45^\circ < \text{inter-SCR angle} \leq 90^\circ$ , 2 if  $90^\circ < \text{inter-SCR angle} \leq 135^\circ$  and 3 if  $135^\circ < \text{inter-SCR angle} \leq 180^\circ$

for C3b at SCR-20 (Jokiranta *et al.*, 2006; Chapter 6), and charge effects at the FH-C3b interface may be important for the binding of FH. By this study, conformational changes with increase of NaCl concentration or pH indicate that the internal charge attractions within FH become weakened as these parameters increase. This conclusion presumes that there are no specific interactions between the protein and Na<sup>+</sup> or Cl<sup>-</sup> ions, i.e. that the effect is due primarily to ionic strength. The constrained modelling of FH shows that the observed conformational changes can be rationalised by relatively minor alterations in a few salt bridges including His residues (given that generally similar SCR arrangements were determined in different NaCl concentrations). It is not known whether the effects are caused within a single inter-SCR linker region, or by interactions between two non-adjacent SCR domains. Electrostatic attractions are however not able to account fully for the folded-back structure for FH, and this implies that other contributions are present. Glycosylation and linker lengths are two possible contributory factors.

The central region of FH is dominated by eight glycosylation sites between SCR-9 to SCR-18, given that the putative site at SCR-4 is not glycosylated (Fenaille *et al.*, 2007). Here, it has been shown that the possible role of carbohydrate in causing FH to adopt a folded-back domain structure is unlikely. The centre is also dominated by long inter-SCR linkers, where six of the eight linkers between SCR-7 and SCR-15 are longer than the usual length of three or four residues (Aslam & Perkins, 2001). Long linkers facilitate the greater folding back of two adjacent SCR domains from an averaged inter-SCR angle of about 120° seen in the best-fit FH models to the lower inter-SCR angles of 37°, 40° and 69° seen in other SCR proteins. These lower values are found in three crystal structures showing folded-back SCR domains, namely CR2 SCR-1/2 with an eight-residue linker (Szakonyi *et al.*, 2001; Prota *et al.*, 2002), and the factor B SCR-2/3 domains with a six-residue linker (Milder *et al.*, 2007). It should be noted that longer linkers can also facilitate flexibility between adjacent SCR domains. This is shown by the comparison of the opened-up solution structure of CR2 SCR-1/2 (inter-SCR angle of 63°-121°) with its folded-back crystal structure (inter-SCR angle of 37°-40°) (Gilbert *et al.*, 2005).

The present conformational studies of FH provide further insight on the conformational properties of long SCR proteins. The original solution structure for FH was the first of its type and was limited by the quality of the scattering data and the SCR domain models (Aslam & Perkins, 2001). The new FH modelling curve fits from this current study are much improved compared to that of 2001. The fits are similar in agreement to those reported for scattering modelling structures for other SCR proteins, including CR1 and CR2 and the fragments of FH, and this shows that the “Randomised linker-3” method is generally useful for these medium resolution solution structure determinations (Fernando *et al.*, 2007; Chapter 4; Gilbert *et al.*, 2006; Furtado *et al.*, 2008). The comparison of SCR proteins in Table 2 of Furtado *et al.*, 2008 can now be updated, in which the  $R_G/R_O$  ratio for FH is now 2.8-2.9 in place of 3.67 previously. This ratio is much lower than those of CR1 with a ratio of 3.76 and CR2 with a ratio of 4.1. Likewise the calculation of the  $ff_o$  ratio from the sedimentation coefficient of 5.56 S (Table 5.1) leads to 1.75 in place of 1.77 previously (Furtado *et al.*, 2008). In distinction, CR1 has a larger  $ff_o$  ratio of 2.29 and CR2 has a larger  $ff_o$  ratio of 1.96. These ratios firmly place FH to possess a folded-back solution structure. Summation of the charged groups in FH, CR1 and CR2 shows small differences, where the composition-predicted pI of FH is lower at 6.1-6.3 compared to those of CR1 and CR2 which are 6.5-6.7 and 7.8-8.5 respectively, and there is a higher proportion of charged groups in FH.<sup>IV</sup> The slight decrease in pI and increase in charged groups for FH are consistent with a role for electrostatic effects in partly determining the FH domain arrangement. Mutations in FH that lead to aHUS occur throughout FH, even though the majority occur within the C-terminal C3d- and heparin-binding SCR-20 domain. Of interest is that about half the known missense aHUS mutations in FH (26 out of 56) involve changes in charged residues, even though charged residues comprise only 27% of the residues in FH (Saunders *et al.*, 2007). Charged residues are often found at protein surfaces and may be involved in direct interactions between FH and its ligands. Differences in residue charge

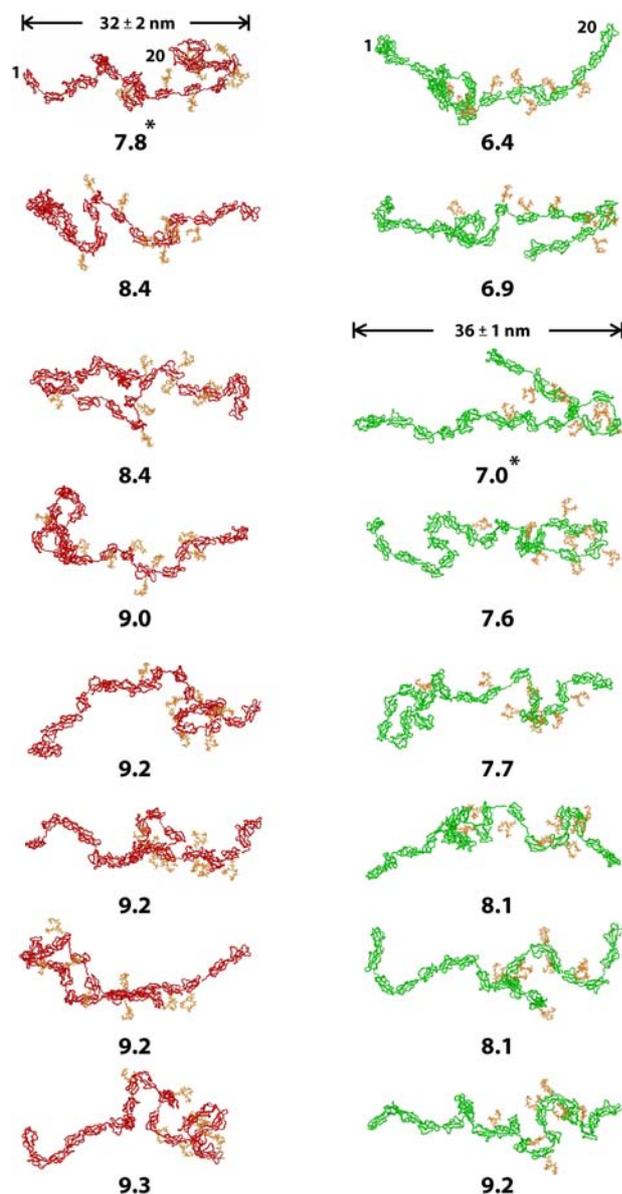
---

<sup>IV</sup> The net charge for FH is -2 out of 326 charged residues (27%) in a total of 1213 amino acid residues. The net charge for CR1 is +11 out of 439 charged residues (23%; not counting the C-terminus) in a total of 1931 residues. The net charge for CR2 is +8 out of 218 charged residues (23%; not counting the C-terminus) in a total of 951 residues.

in aHUS mutations may therefore lead to functional alterations in FH, which in turn will facilitate the onset of aHUS, in addition to direct charged interactions between FH and its ligands. This may account in part for the occurrence of aHUS mutations in SCR domains other than that of SCR-20.

While NaCl concentration-dependent FH conformations have been identified, it is stressed that scattering modelling is unable to specify proximity relationships between individual SCR domains, only that some domains will approach others. It should additionally be noted that, given the effect of electrostatic interaction between separated SCR domains in intact FH, crystal and NMR structural determinations of small FH fragments may give misleading inter-SCR domain conformations when considered in the context of intact FH. A typical range of FH conformations from scattering modelling is depicted in [Figure 5.14](#). The single views of [Figure 5.12](#) are not likely to represent any specific or actual FH conformation in solution.

Previous knowledge of FH binding sites on C3b clarifies the FH interaction with C3b. When three FH-binding peptides in C3b are mapped onto the C3b crystal structure ([Janssen \*et al.\*, 2006](#)), peptide H-1 is found to lie on the opposite face of C3b to those of H-2 and H-3 ([Chapter 4](#)). This indicates that the FH SCR domains will fold themselves around the C3b structure, thus FH is not wholly positioned on one side of C3b. Electrostatic views of FH and C3b can provide further insight into this interaction. Electrostatic surface maps reveal basic surfaces at SCR-7, SCR-13 and SCR-20, while acidic surfaces are revealed at SCR-9, SCR-12 and SCR-15 ([Aslam & Perkins, 2001](#); [Figure 5.15](#)), while a similar varied range of differently-charged surfaces are visible in C3b ([Figure 5.15](#)). For FH, these correspond to the charge summations in [Figure 5.1](#). These indicate that charge interactions will be important for the FH-C3b complex formation, and will facilitate any folding of the FH domains around C3b. The NaCl-dependent conformational behaviour of FH observed in the absence of C3b in the present study is consistent with the folding of FH around C3b. It is therefore not surprising that the SCR domains of unbound FH will display NaCl-dependent conformations. Based on this observation, many more residues in FH outside SCR-19 and SCR-20 are predicted to be important for the C3b-FH interaction. Such an extensive



**Figure 5.14**

The best-fit FH models in 50 mM and 250 mM NaCl. The eight red models in order of R-factor correspond to the FH models in 50 mM NaCl that were deposited in the Protein Data Bank (PDB code 3gau), and the green models correspond likewise to those in 250 mM NaCl (PDB code 3gaw). The best-fit structure is asterisked. The maximum length of these models ( $\pm$  standard deviation) is shown above the best-fit model. The FH conformations shown do not predict any proximity relationships between neighbouring SCR domains; only that the best-fit models are partially folded back and that non-covalent contacts between two non-adjacent SCR domains are possible.

contact model offers a further explanation for the occurrence of aHUS mutations outside SCR-20 (Saunders *et al.*, 2007; Chapter 4).

Knowledge of the oligomeric state of FH is also crucial for understanding its function in complement regulation and its likely roles in AMD and aHUS disease. FH oligomerisation in solution was first identified in 1991 (Perkins *et al.*, 1991), but it is only now that this feature of FH is becoming understood. The most recent study showed that a reversible monomer-dimer equilibrium exists at the lowest concentrations, and higher oligomeric structures that are not in equilibrium with each other exist at concentrations higher than physiological (Nan *et al.*, 2008a). Oligomer formation has been attributed to dimerisation in each of SCR-6/8 and SCR-16/20 in physiologically-relevant concentrations (Fernando *et al.*, 2007, Chapter 4). A repeated alternation of dimerisation events will lead to the formation of an indefinite chain of FH oligomers. The present study extends these results by showing that FH oligomerisation persists in a range of NaCl concentrations and pH values at FH concentrations below and above its physiological concentration in plasma. The proportions of FH monomer, dimer and higher oligomers in the preparations used for this study are approximately 84%, 8% and 8% (Figure 5.6). This is consistent with the previously-determined dissociation constant  $K_D$  value of 28  $\mu\text{M}$  at pH 7.4, 137 mM NaCl, which indicates that about 5% to 15% of FH in plasma will be dimeric, so long as no other factors require consideration (Nan *et al.*, 2008a).

Oligomer formation increases with a reduction in NaCl concentration or an increase in FH concentration. Again this indicates that the regulatory role of FH will be affected by the local environment. Depending on the progression of oligomer formation, ligand binding may be blocked or enhanced. For example, substantially larger FH oligomers form in the presence of at least 60  $\mu\text{M}$  zinc and copper, and these inhibit the regulatory function of FH (Nan *et al.*, 2008b). In addition, given the lack of reversibility of FH oligomer formation at high concentration, it may be important to use freshly gel-filtrated FH samples stored below 1 mg/ml concentrations for complement regulatory functional studies. FH oligomers are potentially relevant for AMD, given that the hall mark of AMD is the formation of sub-RPE deposits (including drusen) at the retinal surface. Drusen are formed from about 120 aggregated proteins and oxidized lipids

adjacent to the Bruch's membrane. Currently there is no detailed knowledge of the mechanism of drusen formation in AMD. It is possible that the local environment at the retina may facilitate the slow formation of protein precipitates, including FH, during the course of a lifetime.

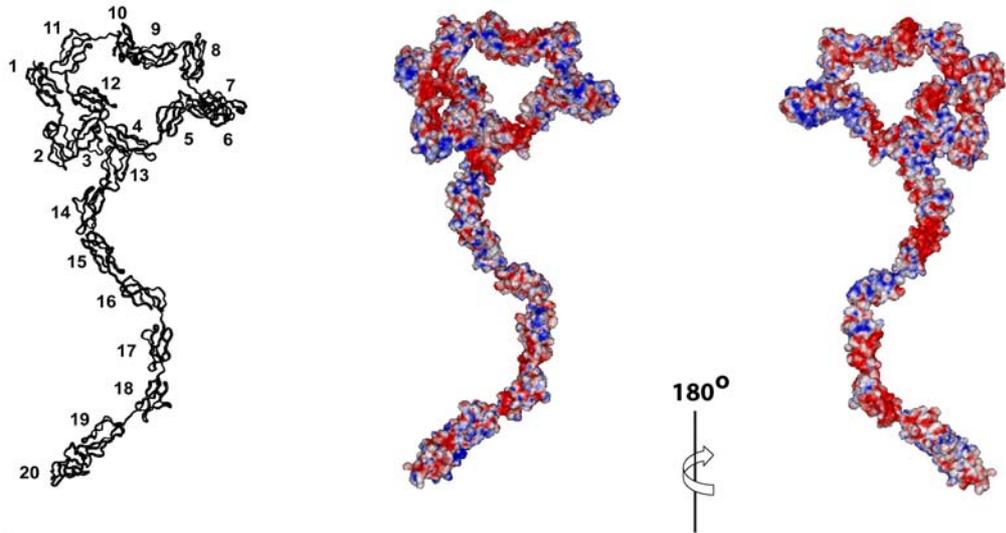
#### **(5.4) Materials and Methods**

##### **(5.4.1) Purification of Factor H**

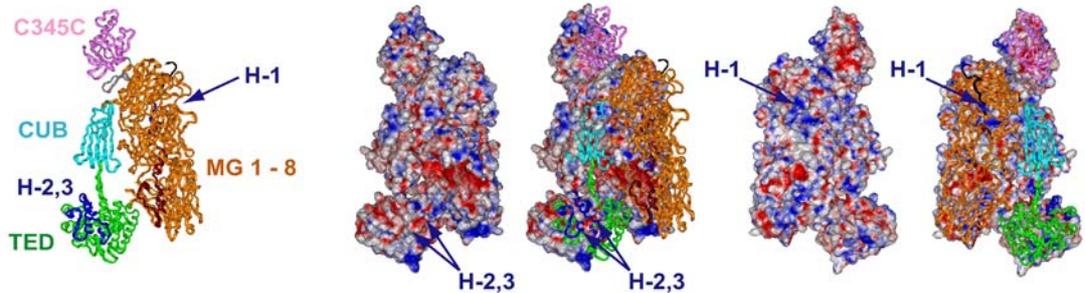
Native FH was purified and concentrated from a pool of just-outdated anonymised human plasma using monoclonal MRC-OX23 Sepharose affinity column by Miss Ruodan Nan ([Sim \*et al.\*, 1993](#); [Nan \*et al.\*, 2008a](#); [2008b](#)). FH samples were checked by SDS-PAGE before and after each X-ray scattering or AUC experiment. Samples were dialysed into a range of buffers by centrifugation at 2,000 g using four buffer changes in Millipore concentrators with a 50 kDa molecular weight cutoff. AUC and X-ray scattering FH data in 137 mM NaCl, pH 7.4 were collected by Miss Ruodan Nan in 10 mM HEPES, 0.5 mM EDTA ([Nan \*et al.\*, 2008a](#)). All other 137 mM NaCl data for this thesis utilised phosphate-buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 8.1 mM NaH<sub>2</sub>PO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>). Because no FH oligomerisation was seen with Zn or Cu below concentrations of 20 µM ([Nan \*et al.\*, 2008b](#)), EDTA was not added to the PBS buffers. According to the manufacturer's specifications, any trace amounts of copper in PBS was below 0.0005%. The NaCl concentration dependence experiments at pH 7.4 were performed for this thesis by replacing 137 mM NaCl with 50 mM NaCl, 250 mM NaCl or 350 mM NaCl. The pH dependence was studied by resetting the pH of PBS to pH 5.4 and 9.4. Deglycosylated FH was prepared by enzymatic digestion using PNGase F in PBS at pH 7.4 ([Fenaille \*et al.\*, 2007](#)). Briefly, 200 µg FH in 90 µl was mixed with 10 µl of the detergent NP-40 (50% solution) and 2 µl of PNGase F solution prepared according to manufacturer's instructions (Prozyme, San Leandro, CA, USA). The resulting mixture was incubated at 37°C for 24 h. Deglycosylation was confirmed by SDS PAGE.

The FH amino acid composition was taken from its sequence (SWISSPROT accession code P08603). The glycosylation of FH was taken to be eight biantennary oligosaccharides located at SCR-9, SCR-12, SCR-13, SCR-14, SCR-15 (two sites),

## Factor H



## C3b



**Figure 5.15**

Electrostatic views of the best-fit model of FH in 137 mM NaCl and its comparison with the crystal structure of C3b. Both FH and C3b are drawn to the same scale. Red denotes acidic regions while blue denotes basic regions.

Upper: The ribbon view for the best fit FH model (PDB code: 3gav; [Figure 5.12\(b\)](#)) is numbered to indicate the positions of the SCR domains. The front and back views of its electrostatic surface are shown. No specific proximity arrangements between non-adjacent SCR domains in FH are implied by this view, only that the SCR domains are folded back.

Lower: The ribbon view of the crystal structure of C3b (PDB code: 2i07) is colour coded to indicate the TED, CUB, C345C and MG1-MG8 domains.

The blue ribbons highlight three previously-mapped peptides that bind to C3b and denoted by H1, H2 and H3. The front and back views of the C3b electrostatic surface are shown, with and without the ribbon view superimposed on this, for clarity. The positions of the H-1, H-2 and H-3 binding peptides in C3b are arrowed.

SCR-17 and SCR-18 from the recent mass spectrometry determination; note that SCR-4 is not glycosylated (Fenaille *et al.*, 2007; Figure 5.1). This resulted in a revised calculated FH molecular weight of 154.4 kDa (in place of 150.0 kDa used in our previous studies), an unhydrated volume of 193.1 nm<sup>3</sup>, a hydrated volume of 256.2 nm<sup>3</sup> (based on a hydration of 0.3g H<sub>2</sub>O/g of glycoprotein and an electrostricted volume of 0.0245 nm<sup>3</sup> per bound water molecule), a partial specific volume  $v$  of 0.715 ml/g, and an absorption coefficient at 280 nm (1%, 1 cm path length) of 16.2 (in place of 16.7 used in our previous studies). The value of 16.2 is based on molar extinction coefficients of 5665 M<sup>-1</sup>cm<sup>-1</sup> for Trp residues, 1535 M<sup>-1</sup>cm<sup>-1</sup> for Tyr residues, and 129 M<sup>-1</sup>cm<sup>-1</sup> for Cys residues (Perkins, 1986). A similar value of 15.7 was calculated based on molar extinction coefficients of 5500 M<sup>-1</sup>cm<sup>-1</sup> for Trp residues, 1490 M<sup>-1</sup>cm<sup>-1</sup> for Tyr residues, and 125 M<sup>-1</sup>cm<sup>-1</sup> for Cys residues (Pace *et al.*, 1995), for which the evaluation of 93 Trp-containing proteins showed a mean standard deviation of 3.167 % from the observed value. Other studies have reported a higher absorption coefficient of 19.5 (Hakobyan *et al.*, 2008), or a lower one of 14.2 and 12.4 (Sim & DiScipio, 1982; Pangburn *et al.*, 2009). The value of 16.2 used here conveniently corresponds to the mean of these experimental determinations. All the buffer densities were measured at 20°C using an Anton-Paar DMA5000 density meter to be 1.00169 g/ml (PBS using 50 mM NaCl), 1.00488 g/ml (HEPES using 137 mM NaCl), 1.00541 g/ml (PBS using 137 mM NaCl), 1.00995 g/ml (PBS using 250 mM NaCl), 1.01394 g/ml (PBS using 350 mM NaCl), 1.00536 g/ml (PBS at pH 5.4) and 1.00554 g/ml (PBS at pH 9.4).

#### (5.4.2) Sedimentation velocity data for Factor H

AUC data for FH were obtained on two Beckman XL-I instruments equipped with AnTi50 and AnTi60 rotors. Sedimentation velocity data was acquired at 20°C at rotor speeds of 40,000 r.p.m, 50,000 r.p.m. and 60,000 r.p.m. in two-sector cells with column heights of 12 mm. Analysis was performed using direct boundary Lamm fits of up to 300 scans using SEDFIT version 11.0 (Schuck, 1998; 2000). SEDFIT resulted in size-distribution analyses  $c(s)$ , the algorithm for which assumes that all species have the same frictional ratio  $f/f_o$ . The final SEDFIT analyses used a fixed resolution of 200 and optimised the  $c(s)$  fit by floating  $f/f_o$  until the overall root mean square deviations and

visual appearance of the fits were satisfactory (Figures 5.2-5.5). The percentage fraction of oligomers in the total loading concentration was derived using the  $c(s)$  integration function.

### (5.4.3) X-ray scattering data collection and analysis

X-ray scattering data were obtained in three beam sessions in either four-bunch or sixteen-bunch mode at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, operating with a ring energy of 6.0 GeV on the Beamline ID02 (Narayanan *et al.*, 2001). For data collection in this thesis, storage ring currents ranged from 32 mA to 75 mA. Data were acquired using a recently installed fibre optically-coupled high sensitivity and dynamic range CCD detector (FReLoN). Together with a smaller beamstop, this allowed a shorter sample-to-detector distance of 2.0 m to be used and resulted in higher signal-to-noise ratios (Li *et al.*, 2008). FH was studied at 0.4 mg/ml to reduce the proportion of reversible oligomers. Samples of volume 100  $\mu$ l were measured in flow cells to reduce radiation damage by moving the sample continuously during beam exposure. Sets of ten time frames each, with frames of duration 0.1 sec, 0.25 sec, 0.5 sec or 1.0 sec each, were acquired. On-line checks during data acquisition confirmed the absence of radiation damage, after which the ten frames were averaged. Other details including the data reduction procedure are described elsewhere (Gilbert *et al.*, 2005).

In a given solute-solvent contrast, the radius of gyration  $R_G$  is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low  $Q$  gives the  $R_G$ , and the forward scattering at zero angle  $I(0)$  (Glatter & Kratky, 1982):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3.$$

This expression is valid in a  $Q.R_G$  range up to 1.5. If the structure is elongated, the mean radius of gyration of cross-sectional structure  $R_{XS}$  and the mean cross-sectional intensity at zero angle  $[I(Q)Q]_{Q \rightarrow 0}$  is obtained from:

$$\ln [I(Q)Q] = [I(Q)Q]_{Q \rightarrow 0} - R_{XS}^2 Q^2/2.$$

The  $R_G$  and  $R_{XS}$  analyses lead to the triaxial dimensions of the macromolecule if the structure can be represented by an elliptical cylinder, where  $L = \sqrt{12(R_G^2 - R_{XS}^2)}$  and  $L$

is its length (Glatter & Kratky, 1982). The  $R_G$  and  $R_{XS}$  analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics O2 Workstations. Indirect transformation of the scattering data  $I(Q)$  in reciprocal space into real space to give the distance distribution function  $P(r)$  was carried out using the program GNOM (Semenyuk & Svergun, 1991):

$$P(r) = \frac{I}{2\pi^2} \int_0^{\infty} I(Q) Qr \sin(Qr) dQ$$

$P(r)$  corresponds to the distribution of distances  $r$  between volume elements. For this, the X-ray  $I(Q)$  curve utilised up to 280 data points in the  $Q$  range between  $0.06 \text{ nm}^{-1}$  and  $1.9 \text{ nm}^{-1}$ .

#### (5.4.4) Modelling of the Factor H solution structures

The first constrained modelling of the 20 SCR domains in FH used NMR structures for three SCR domains and homology models for the remaining 17 SCR domains (Aslam & Perkins, 2001). Here the remodelling of FH was based on NMR and crystal structures for 11 SCR domains and nine improved SCR homology models for the remainder. The homology models were taken from the website <http://www.fh-hus.org> (Saunders *et al.*, 2006). NMR structures were used for SCR-5 (Barlow *et al.*, 1992), SCR-15/16 (PDB code 1hfh; Barlow *et al.*, 1993) and SCR-1/3 (PDB codes 2rlp and 2rlq; Hocking *et al.*, 2008). Crystal structures were used for SCR-19/20 (PDB code 2g7i; Jokiranta *et al.*, 2006) and SCR-6/8 (PDB code 2uwn; Prosser *et al.*, 2007). The FH sequence numbering was initiated from the N-terminus of the 18-residue signal peptide, not that of the mature protein. Twenty peptides that corresponded to the N-terminus (19-ED-20) and the 19 inter-SCR linkers were created in extended  $\beta$ -strand conformations using INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA) on Silicon Graphics OCTANE Workstations (Aslam & Perkins, 2001). Libraries of 1,000 randomised conformations were generated for each peptide using molecular dynamics simulations. This procedure followed the Randomised Method 3 used for intact FH, in which the peptides are constrained to be almost fully extended in length (Aslam & Perkins, 2001).

One conformation was randomly selected from each library for each linker in order to join the 20 SCR domains to assemble 5,000 full FH models. Further details are given elsewhere (Aslam & Perkins, 2001; Gilbert *et al.*, 2005). The inter-SCR domain angle is defined as the angle between the two longest axes of two adjacent SCR domains, where each axis is the line between the midpoint of the two  $\alpha$ C atoms in the disulphide bridges at the two ends of the SCR domain. At the N- and C-termini, the axes were redefined using the  $\alpha$ C atoms at 19-E and C-21, and at 1228-C and 1231-R respectively. In other analyses, the ranking  $R$  for averaged inter-domain angles is defined for the values 0, 1, 2 and 3 as follows:  $R = 0$  if  $0^\circ \leq \text{inter-SCR angle} \leq 45^\circ$ ;  $R = 1$  if  $45^\circ < \text{inter-SCR angle} \leq 90^\circ$ ;  $R = 2$  if  $90^\circ < \text{inter-SCR angle} \leq 135^\circ$ ;  $R = 3$  if  $135^\circ < \text{inter-SCR angle} \leq 180^\circ$  (Table 5.2).

#### (5.4.5) Debye scattering and sedimentation coefficient modelling

Each full FH molecular model was used to calculate its X-ray scattering curve for comparison with five experimental FH curves (Table 5.1) using Debye sphere models (Gilbert *et al.*, 2005). A cube side length of 0.498 nm in combination with a cutoff of 3 atoms was used to convert the coordinates into Debye sphere models with 1485 spheres that corresponded to the unhydrated structure. The hydration shell was created using HYPRO, (Ashton *et al.*, 1997) where the optimal total of hydrated spheres in the FH model is 2074 (Table 5.1). The X-ray scattering curve  $I(Q)$  was calculated using the Debye equation adapted to spheres (Perkins & Weiss, 1983). Other details are given elsewhere (Perkins, 2001). Steric overlap between the SCR domains in the models was assessed using the number of spheres  $N$  in the models after grid transformation, where models showing less than 95% of the required total of 1485 spheres were discarded. Over 99% of the 5,000 models met this “absence of steric overlap” criterion. Next, the X-ray  $R_G$  and  $R_{XS-1}$  values were calculated from the modelled curves in the same  $Q$  ranges used for the experimental Guinier fits. This allowed for any approximations inherent in the use of the  $Q.R_G$  fit range up to 1.5. Models that passed the  $N$ ,  $R_G$  and  $R_{XS-1}$  filters were then ranked using a goodness-of-fit  $R$  factor defined by analogy with protein crystallography and based on the experimental curves in the  $Q$  range extending to  $1.8 \text{ nm}^{-1}$  (Table 5.1).

Sedimentation coefficients  $s_{20,w}^0$  were calculated directly from the atomic coordinates in the HYDROPRO shell modelling program using the default value of 0.31 nm for the atomic element radius for all atoms to represent the hydration shell (Garcia de la Torre *et al.*, 2000).

#### **(5.4.6) Protein Data Bank accession number**

The best-fit FH  $\alpha$ -carbon co-ordinate models have been deposited in the Protein Data Bank with the accession codes 3gau (50 mM NaCl buffer), 3gav (137 mM NaCl buffer), and 3gaw (250 mM NaCl buffer).

## **Chapter Six**

**Multimeric interactions between complement Factor H and its C3d ligand  
provide new insight on complement regulation**

## (6.1) Introduction

The interaction between FH and C3b involves multiple SCR sites within FH. In studies using buffers with 50-75 mM NaCl, the four N-terminal domains SCR-1/4 have been shown to bind to intact C3b, a second site in SCR-6/10 binds to the C3c region of C3b, and a third site within SCR-16/20 binds to the C3d region of C3b (Lambris *et al.*, 1988; Sharma & Pangburn, 1996; Jokiranta *et al.*, 2000; 2001). A molecular understanding of the FH-C3b complex requires the dissection of the role of each of these C3b sites on FH. This chapter focuses on the C3d fragment and its interaction with SCR-16/20 and FH. Ligand-binding assays in 137 mM NaCl buffer first revealed that a peptide derived from C3d inhibits FH binding to C3b (Lambris *et al.*, 1988). SPR analysis in 50 mM and 75 mM NaCl indicate that C3d binds to FH and also to the FH-related proteins FHR-3 and FHR-4 with dissociation constants  $K_D$  of 0.09-0.36  $\mu\text{M}$  (Hellwage *et al.*, 1999; 2002). The FH-C3d interaction in 50 mM and 75 mM NaCl was confirmed, and the binding site was localised to SCR-20 (Jokiranta *et al.*, 2000; Hellwage *et al.*, 1999). Mutations in SCR-20 lead to reduced binding to C3d and heparin (Manuelian *et al.*, 2003; Jokiranta *et al.*, 2005). The FH binding sites for C3d and heparin partially overlap and lead to a ternary complex of FH-C3d-heparin in 70 mM NaCl (Hellwage *et al.*, 2002; Jokiranta *et al.*, 2005). Interestingly, other studies showed that dimeric C3dg remains bound to IgG after C3b inactivation, suggesting that dimeric C3d is required for efficient binding to CR2 and that C3d oligomers may be physiologically important (Arnaout *et al.*, 1983; Dempsey *et al.*, 1996; Jelezarova *et al.*, 2003).

Structural studies on intact FH are hindered by its large size, glycosylation and inter-SCR flexibility, and a multidisciplinary approach has been required to elucidate its structure. Medium resolution solution scattering structures based on homology models and high resolution NMR and crystal structures have shown that FH possesses a partly folded-back SCR arrangement (Aslam & Perkins, 2001; Soares & Barlow, 2005; Jokiranta *et al.*, 2006; Herbert *et al.*, 2006; Prosser *et al.*, 2007; Hocking *et al.*, 2008; Chapter 5). In contrast, structural studies of FH-ligand complexes are more limited. It was previously shown that SCR-6/8 has a bent solution structure, is weakly dimeric, and forms a 1:1 complex with heparin (Fernando *et al.*, 2007). The solution structure for SCR-16/20 (Chapter 4; Figure 6.1) shows that this is partially folded-back and partly dimeric. Native FH exists in oligomeric forms (Chapter 5; Nan *et al.*, 2008a). The solution and crystal structures

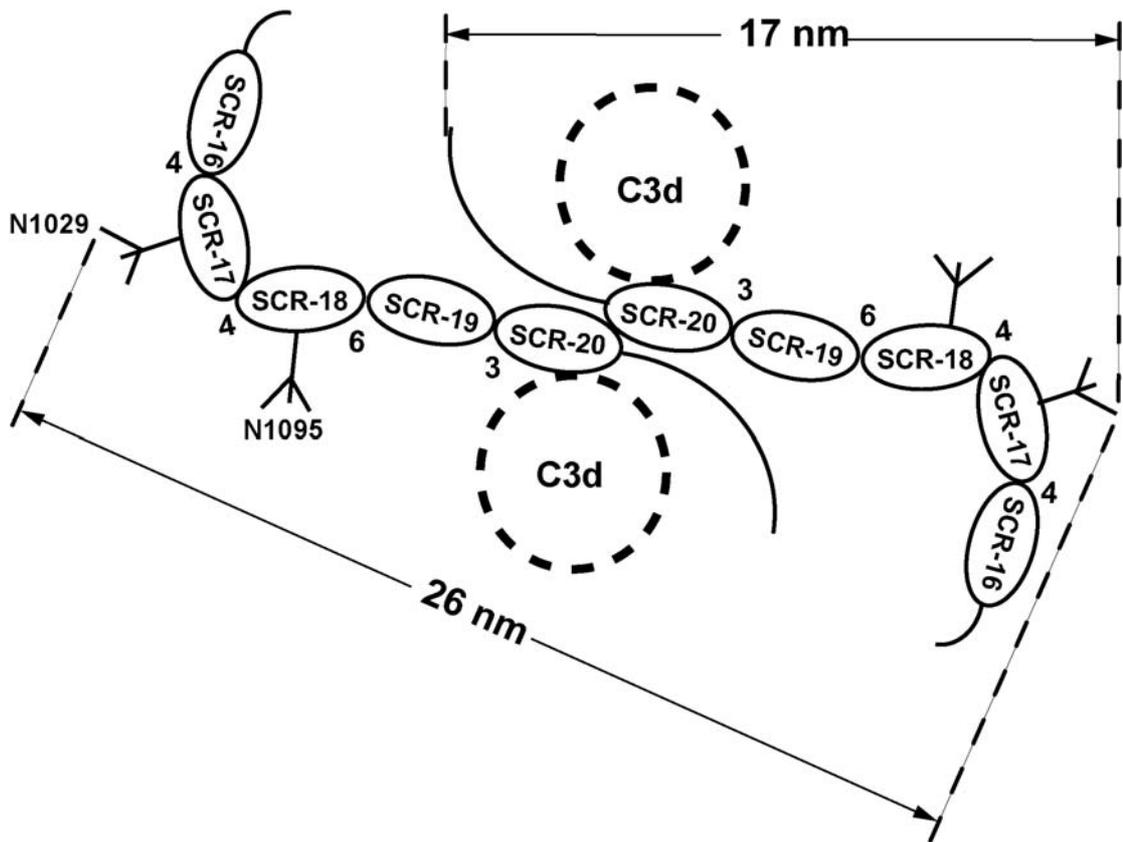
for C3d have been reported (Nagar *et al.*, 1998; Gilbert *et al.*, 2005; Li *et al.*, 2008), and C3d binds to SCR-20 (Jokiranta *et al.*, 2000; Hellwage *et al.*, 1999). Together, these studies now make possible the investigation of the SCR-16/20 and FH complexes with C3d, where it can be predicted that the complexes do not show a 1:1 stoichiometric ratio (Figure 6.1). This chapter presents solution structures for the SCR-16/20- and FH-C3d complexes using AUC, SPR, X-ray scattering and molecular modelling. The hypothesis that the FH-C3d complexes do not exhibit the 1:1 stoichiometry that had been commonly assumed previously is confirmed here. It is also shown that ionic interactions play a major role in FH-C3d complex formation. This observation of C3d-FH multimers provides new insights into the mechanism of complement regulation at host cell surface and contributes to the understanding of aHUS.

## **(6.2) Results and Discussion**

### **(6.2.1) Solution structural data for the Factor H-C3d complex**

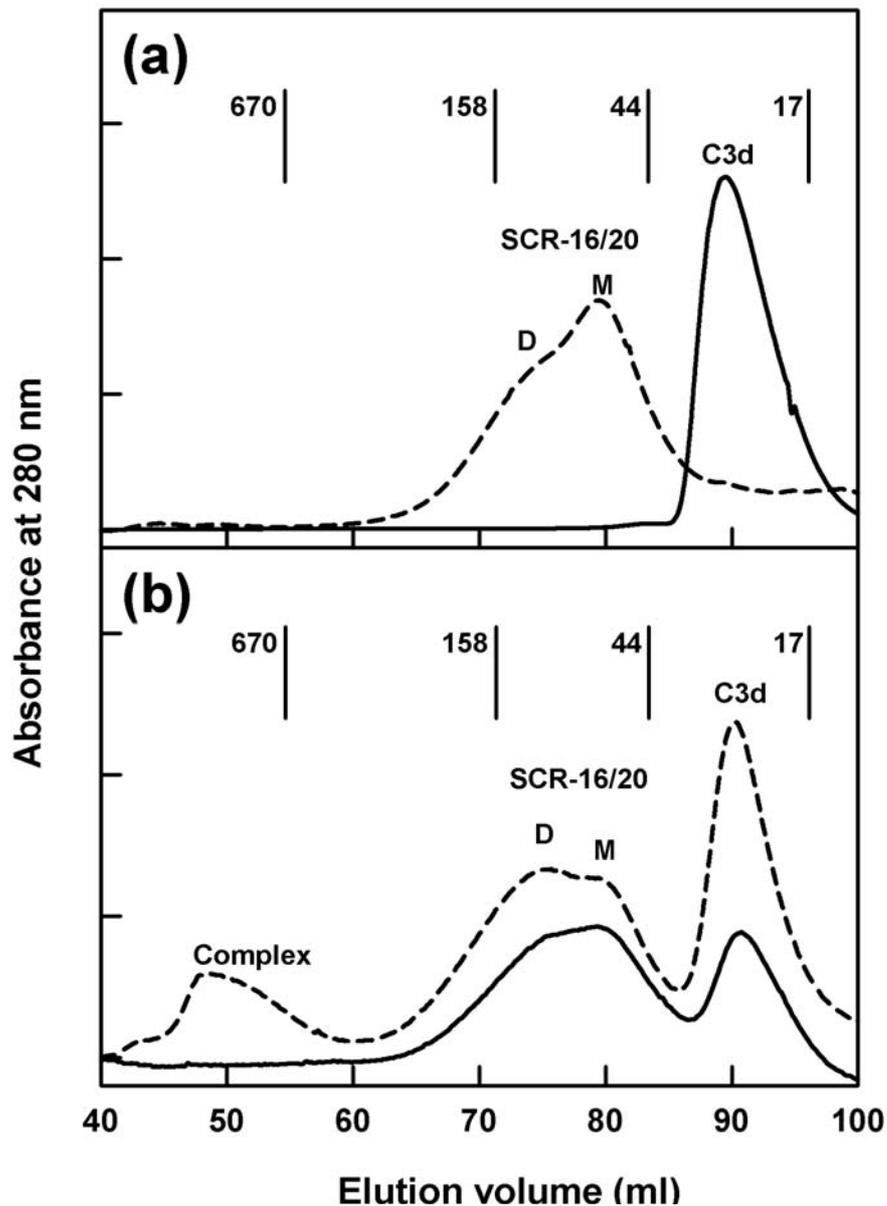
To test the hypothesized spatial arrangement of the individual components within the FH-C3d complexes (Figure 6.1), the approaches used previously to characterize each of unbound SCR-16/20, native FH and C3d in solution (Gilbert *et al.*, 2005; Li *et al.*, 2008; Section 6.2.2; Chapter 4; Chapter 5) were now applied to obtain structural data for their complexes. These data were then fitted using the structural models for SCR-16/20, FH and C3d. Experiments on C3d were performed jointly with Miss Keying Li (Li *et al.*, 2008), and this is acknowledged accordingly in relevant figure legends. The sedimentation velocity data on C3d in HEPES buffer presented here include reanalyses of data previously collected by Dr. Hannah Gilbert (Gilbert *et al.*, 2005).

Native FH, recombinant SCR-16/20, SCR-1/5 and C3d were expressed and purified (Section 6.4.1; Chapter 4) and analysed by size-exclusion chromatography. While C3d and SCR-1/5 are unglycosylated, at least one of the two putative glycosylation sites of SCR-16/20 is occupied (Chapter 4). C3d eluted as a single homogeneous peak by gel filtration, but with a sharp leading edge and a skewed trailing edge that indicated self-association (Figure 6.2(a)), and showed a single band on non-reducing SDS-PAGE with an apparent molecular weight of 30 kDa (not shown). SCR-16/20 reproducibly eluted as a broad peak with a trailing edge that was occasionally partly-resolved into two components (Figure 6.2(a)) and ran as a single



**Figure 6.1**

Schematic cartoon of the SCR arrangement in the monomer and dimer of SCR-16/20. The length of the monomer is 17 nm, while that of the dimer is 26 nm. The location of C3d (dashed outline) on SCR-20 is predicted to lead to no changes in these lengths if no conformational change occurs in SCR-16/20 after C3d binding. The residue lengths of the four inter-SCR linkers of SCR-16/20 are indicated by numbers. The two predicted N-linked oligosaccharide chains at Asn1029 (SCR-18) and Asn1095 (SCR-19), at least one of which is occupied, are indicated by Ψ-like symbols.



**Figure 6.2**

Size-exclusion chromatography of unbound SCR-16/20, C3d and their complex. The elution positions of four standards are shown in kDa. The elution positions of unbound C3d, SCR-16/20 monomer (M) and dimer (D), and complex are labelled as shown.

(a) In 50 mM NaCl buffer, unbound C3d eluted as a single peak (full line), while unbound SCR-16/20 eluted as a peak partly resolved into two components designated as monomer (M) and dimer (D) (dashed line).

(b) A 1:1 mixture of SCR-16/20 and C3d at 1.6 mg/ml was loaded onto a Superdex-200 prep grade XK 16/60 column in 50 mM NaCl HEPES buffer (dashed line) and 137 mM NaCl PBS buffer (full line). The peak for the complex at 50 ml is absent in 137 mM NaCl PBS buffer.

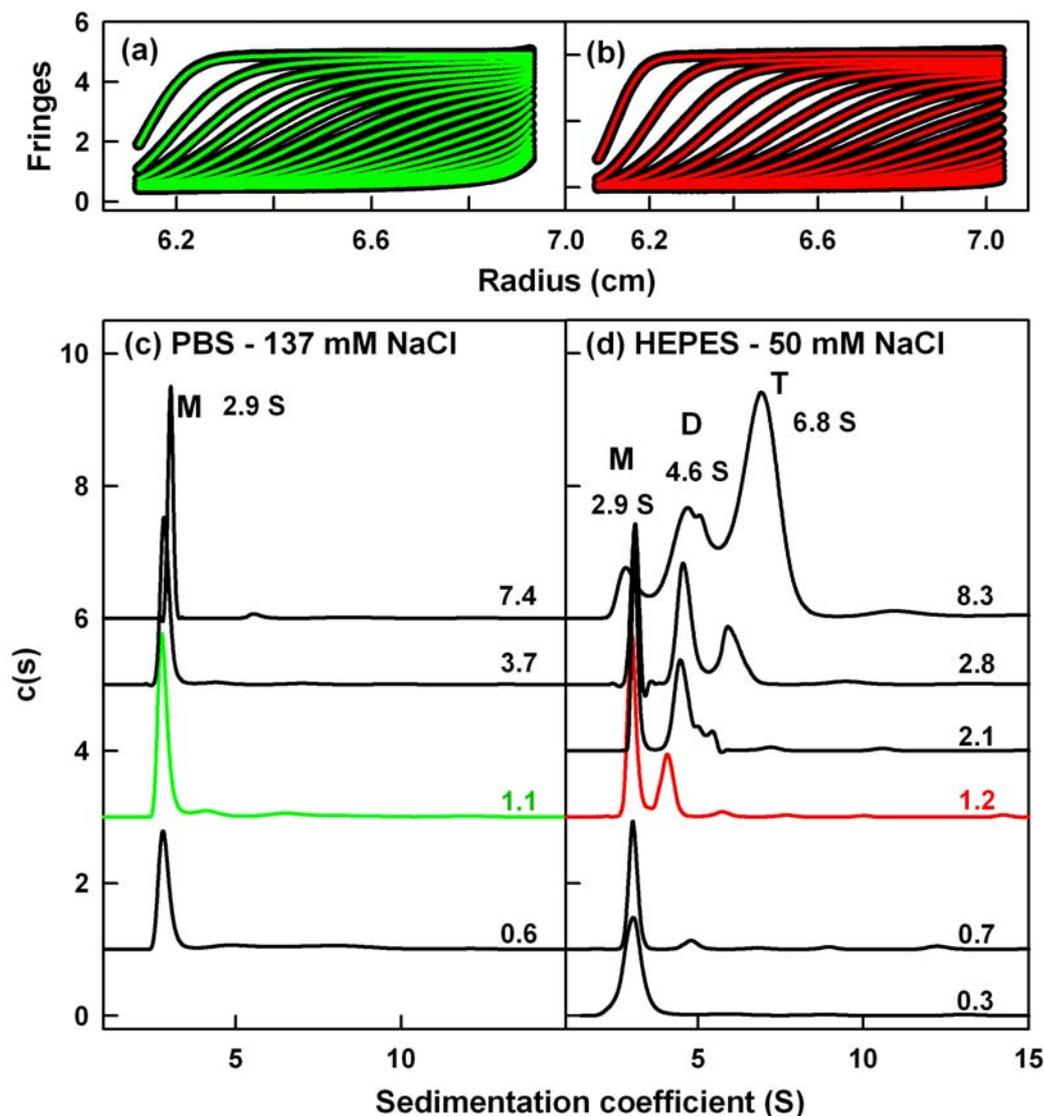
major band on non-reducing SDS-PAGE with an apparent molecular weight of 37 kDa (not shown). Complexes were prepared by mixture of C3d and SCR-16/20 in the appropriate molar ratio. HEPES buffer was used in continuation of the studies with C3d (Gilbert *et al.*, 2005; Li *et al.*, 2008), while PBS was previously used with SCR-16/20 (Chapter 4). No difference was observed in the behaviour of FH in HEPES or PBS buffers at similar NaCl concentrations (Nan *et al.*, 2009b). Hence HEPES buffer was used for SPR, and both HEPES and PBS buffers were used for AUC and X-ray scattering experiments.

By gel filtration, the formation of a weak complex in 50 mM NaCl HEPES buffer was indicated for an equimolar mixture by the appearance of an early peak that eluted at 50 ml in addition to the peaks for SCR-16/20 and C3d (Figure 6.2(b)). The observation of three peaks indicated a micromolar dissociation constant  $K_D$ . Two peaks were observed when 137 mM NaCl PBS buffer was used, indicating that the complex had weakened and that there is a strong ionic strength dependence of complex formation (Figure 6.2(b)).

## (6.2.2) Analytical ultracentrifugation of C3d

### (6.2.2.1) Sedimentation velocity data for C3d

Previously, C3d was observed to undergo what was assumed to be a concentration-dependent monomer-dimer equilibrium with a  $K_D$  value of 40  $\mu$ M in 10 mM HEPES, 50 mM NaCl (Gilbert *et al.*, 2005). Here, this was reinvestigated by sedimentation velocity at concentrations from 0.3 mg/ml to 8.3 mg/ml in both 137 mM (PBS) and 50 mM (HEPES) NaCl buffers (Section 6.4.1). Interference optics data were used throughout, together with absorbance data below 0.8 mg/ml, due to the saturation of the absorbance optics at C3d concentrations above 0.8 mg/ml. Unlike the previous DCDT+ analyses, size-distribution  $c(s)$  analyses using SEDFIT were now used for data analyses, and resulted in good boundary fits (Figure 6.3(a,b)). In 137 mM NaCl, a single predominant species was observed at an  $s_{20,w}^o$  value of  $2.95 \pm 0.05$  S between 0.5-7.5 mg/ml (10 determinations) (Figure 6.3(c)). The  $c(M)$  size distribution plot showed that this corresponded to a molecular mass of  $39 \pm 4$  kDa. This was accordingly assigned to C3d monomers with a molecular mass of 34.6 kDa calculated from its composition. Trace amounts of higher oligomers were visible at higher C3d concentrations (Figure 6.3(c)). In 50 mM NaCl, the  $c(s)$  distributions showed a concentration dependence



**Figure 6.3**

Concentration dependence of the sedimentation velocity data for unbound C3d in 137 mM NaCl (green) and 50 mM NaCl (red) buffers. Figures (a) and (c) are used with permission from Miss Keying Li (Li *et al.*, 2008). (a) The boundary fits for C3d at 1.1 mg/ml in 137 mM NaCl at 32,000 r.p.m. are from interference data. Only every sixth scan in a total of 200 scans is shown for reason of clarity. (b) The boundary fits for C3d at 1.2 mg/ml in 50 mM NaCl at 40,000 r.p.m. are from interference data. Only every fourth scan of the first 80 scans is shown for reason of clarity. (c) The concentration dependence of the c(s) size distributions for 0.6 mg/ml, 1.1 mg/ml (green), 3.6 mg/ml and 7.4 mg/ml of C3d in 137 mM NaCl is shown. A single species is observed at 2.9 S and labelled M (monomer).

(d) The concentration dependence of the c(s) size distributions for 0.3 mg/ml, 0.7 mg/ml, 0.8 mg/ml, 1.2 mg/ml, 2.1 mg/ml, 2.8 mg/ml and 8.3 mg/ml of C3d in 50 mM NaCl is shown. The peaks are labelled as M (monomer), D (dimer) and T (trimer) in increasing order of S values, where the values shown correspond to 8.3 mg/ml.

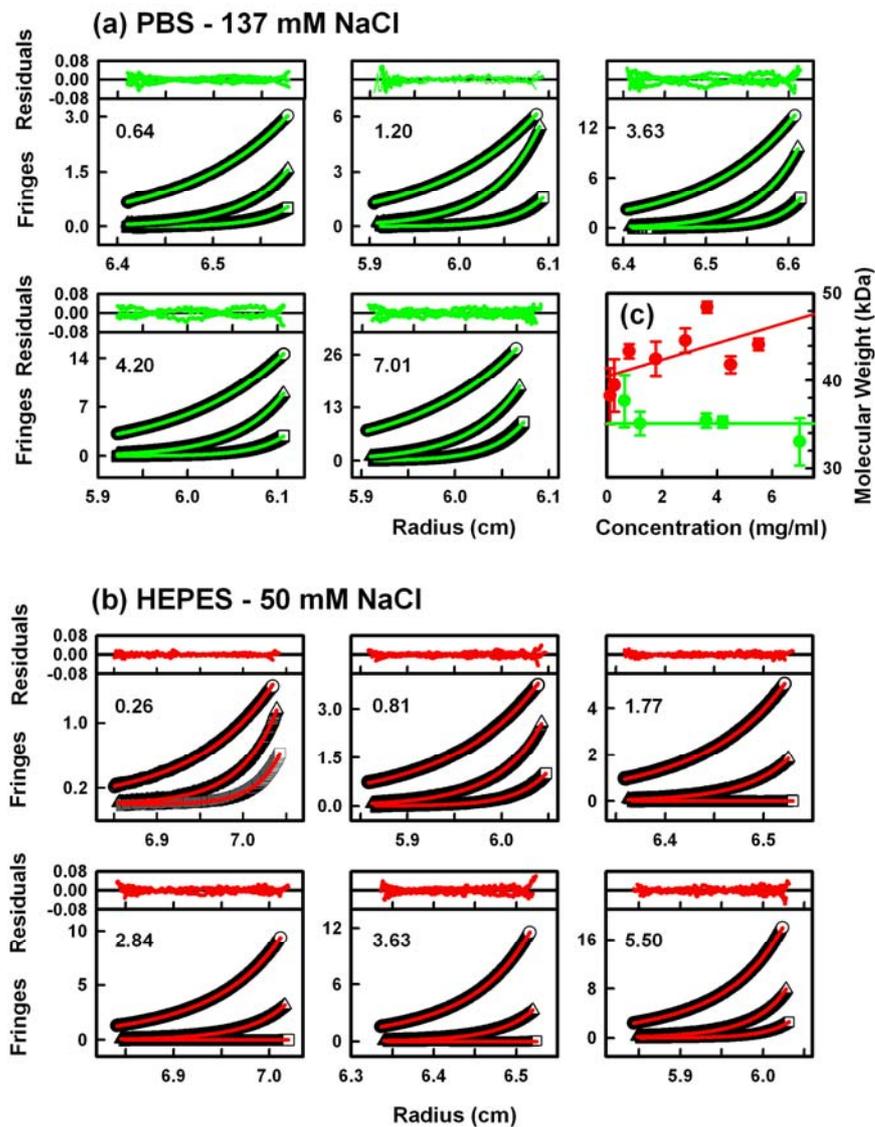
between 0.3-8.3 mg/ml, and three peaks indicating three oligomer species were visible (Figure 6.3(d)). At 0.3 mg/ml, only the monomer (M) was present at 2.9 S.

The mean value is  $2.99 \pm 0.10$  S (21 determinations). Above 0.7 mg/ml, a second peak (D) was observed at  $4.28 \pm 0.29$  S (16 determinations) and its intensity increased with concentration. A third peak (T) appeared above 2.1 mg/ml, whose  $s_{20,w}^{\circ}$  value increased from 5.4 S to 6.8 S as the concentration increased from 2.1 mg/ml to 8.3 mg/ml. The mean value is  $5.99 \pm 0.59$  S (11 determinations). The  $c(M)$  distribution plots showed that peaks M, D and T correspond to molecular masses of  $34 \pm 4$  kDa,  $59 \pm 10$  kDa and  $95 \pm 18$  kDa respectively, assuming that the frictional ratio  $f/fo$  remained unchanged for these species. These masses are consistent with the formation of C3d dimers and trimers of sizes 70 kDa and 105 kDa. The rate of interconversion between the three species was slow enough during the velocity runs for these peaks to be visible in the  $c(s)$  plots. Previous study of C3d in 50 mM buffer using DCDT+ fits of 10-20 scans gave  $s_{20,w}^{\circ}$  values that ranged between 3.26 S to 4.85 S. (Gilbert *et al.*, 2005) This range of  $s_{20,w}^{\circ}$  values is consistent with the range seen in Figure 6.3, and reanalyses of the previous data using SEDFIT confirmed the occurrence of three species in Figure 6.3(d).

#### (6.2.2.2) Sedimentation equilibrium data for C3d

Sedimentation equilibrium experiments in both 137 mM and 50 mM NaCl were performed to explore C3d oligomer formation over a ten-fold concentration range. Absorbance data were not fitted because of the saturation of the optics above 0.8 mg/ml. The individual fits of the equilibrium curves at different speeds and concentrations gave the molecular weights shown in Figure 6.4(c). Thus C3d in 50 mM NaCl showed a concentration dependent mass, while this was not seen in 137 mM NaCl. The final global fit of each multispeed experiment were performed using rotor speeds of 19k, 28k and 35k r.p.m. (Section 6.4.2.2).

For C3d in 137 mM NaCl PBS buffer, the mass was determined using a global fit of 15 curves at five concentrations (Figure 6.4(a)) on the basis of a monomer-dimer-trimer model (Figure 6.3). Low random residuals, a low reduced  $\chi^2$  value of 8.0, and good visual fits were obtained. A C3d monomer mass of  $34.1 \pm 1.6$  kDa was determined, in good agreement with the composition-determined value of 34.6 kDa. The monomer-dimer and monomer-trimer dissociation constants  $K_D$  were



**Figure 6.4**

Multispeed global fit sedimentation equilibrium analyses of C3d in 137 mM NaCl (green) and 50 mM NaCl (red) buffers. The fit residuals are shown above each panel. Figure (a) is used with permission from Miss Keying Li ([Li et al., 2008](#)). (a) The five panels show the multispeed global fit analysis of interference data for C3d (fitted line in green) at five concentrations (0.64 mg/ml, 1.20 mg/ml, 3.63 mg/ml, 4.20 mg/ml and 7.01 mg/ml). Within each panel, the rotor speeds were 19k r.p.m. ( $\circ$ ), 28k r.p.m. ( $\Delta$ ) and 35k r.p.m. ( $\square$ ). The global fits gave a monomer mass of 34.1 kDa. (b) The six panels show the multispeed global fit analysis of interference data for C3d (fitted line in red) at six concentrations (0.26 mg/ml, 0.81 mg/ml, 1.77 mg/ml, 2.84 mg/ml, 3.63 mg/ml and 5.50 mg/ml). The rotor speeds are the same as in (a). (c) Concentration dependence of the molecular mass of C3d in 137 mM (green) and 50 mM (red) NaCl buffers. For 137 mM NaCl, the green line corresponds to a mass of 35 kDa. For 50 mM NaCl, the red line corresponds to the regression line drawn using data from 0.11, 0.26, 0.81, 1.77, 2.84, 3.63, 4.49 and 5.5 mg/ml.

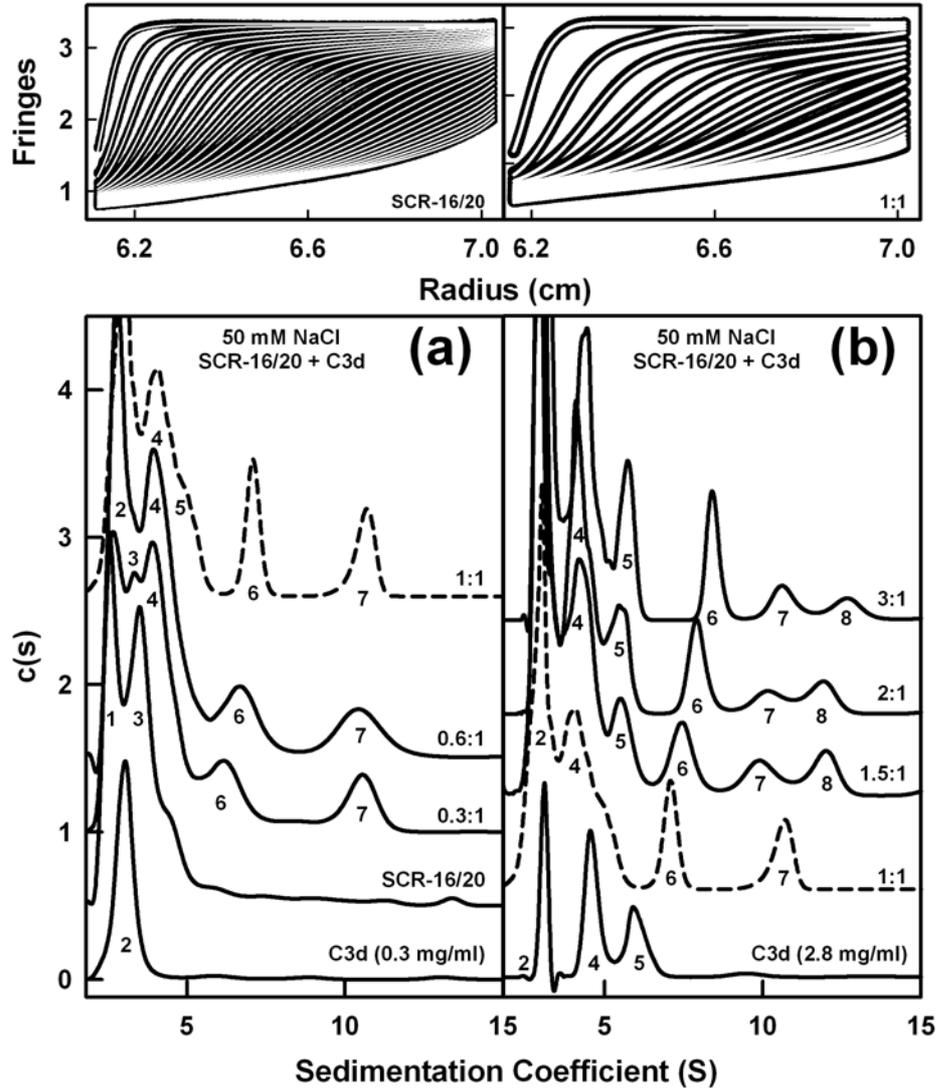
high at 61 mM and 1.4 mM<sup>2</sup> respectively, indicating that no dimer or trimer had formed. This outcome was confirmed by a global fit based on a single species only, which gave a similar mass of 35.1 ± 0.3 kDa and an almost unchanged reduced  $\chi^2$  value of 9.7 (fits not shown). Hence C3d in physiological buffer was concluded to be monomeric.

For C3d in 50 mM NaCl HEPES buffer, the same rotor speeds and concentration range also gave good fits to a monomer-dimer-trimer association model. This fit required that equilibrium had been reached, and this was confirmed using overlays of runs (Section 6.4). The global fit based on a monomer-dimer-trimer association model of 18 curves at six concentrations with a fixed monomer mass of 34.6 kDa gave low residuals (Figure 6.4(b)) and a global reduced  $\chi^2$  value of 3.5. The  $K_D$  values of the monomer-dimer and monomer-trimer equilibria were determined to be 360 ± 0.01 μM and 0.070 ± 0.003 μM<sup>2</sup> respectively. Global fits based on a single species gave much worsened reduced  $\chi^2$  values of 60 to 223. A monomer-dimer model gave a poorer reduced  $\chi^2$  value of 12.9 but gave a  $K_D$  value of 210 μM that was comparable with the previous estimate of 40 μM (Gilbert *et al.*, 2005). Overall, C3d at 8.3 mg/ml in low salt will form more trimer than dimer, and this agrees with the  $c(s)$  analyses of the sedimentation velocity data (Figure 6.3(d)).

### **(6.2.3) Analytical ultracentrifugation of the SCR-16/20 and Factor H complexes with C3d**

#### **(6.2. 3.1) Sedimentation velocity data for the SCR-16/20 complex with C3d**

AUC studies macromolecular structures in solution by following their sedimentation behaviour under a high centrifugal force (Cole *et al.*, 2008). The observed sedimentation boundaries were fitted to yield a sedimentation coefficient distribution,  $c(s)$  according to the Lamm equation using SEDFIT (Section 6.4.2.1). This takes into account all the species present in the sample, although the algorithm assumes that the same frictional ratio applies for all the sedimenting species. Macromolecular elongation is monitored through changes in the sedimentation coefficient  $s_{20,w}^0$  values, and sample polydispersity is determined from the number of peaks observed in the  $c(s)$  plot. AUC sedimentation velocity experiments were carried out for SCR-16/20 and C3d. Only interference optics data were used due to saturation of the absorbance optics at 280 nm in the concentration range used. Free

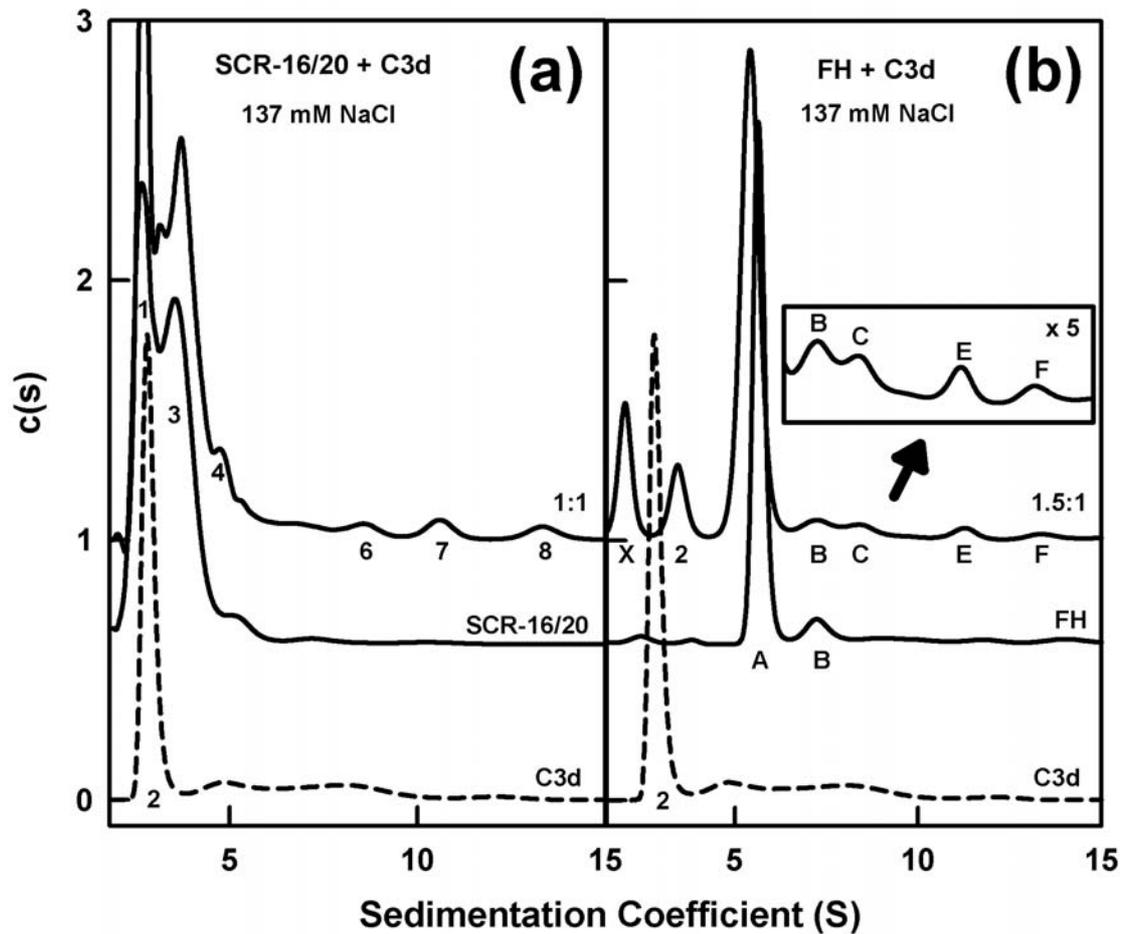


**Figure 6.5**

Size distribution  $c(s)$  analyses of SCR-16/20 and C3d mixtures in 50 mM NaCl HEPES buffer. The SEDFIT analyses correspond to a rotor speed of 50,000 r.p.m. in which C3d was titrated against 1.0 mg/ml SCR-16/20. Representative boundary fits are shown in the two upper panels. The lower panels show nine  $c(s)$  analyses. For clarity, the  $c(s)$  plots were displaced arbitrarily along the  $c(s)$ -axis. Eight peaks were labelled 1-8 in order of increasing  $S$  value. **(a)** The two unbound proteins were assigned as peaks 1 and 3 (SCR-16/20) and peak 2 (C3d at low concentration) at 2.60 S, 2.91 S and 3.30 S in that order. Low molar ratios of C3d (0.3:1, 0.6:1 and 1:1) resulted in complex formation which was suggested by the appearance of peaks 6 and 7. **(b)** C3d at high concentration was assigned to peaks 2, 4 and 5, and complex formation in high molar ratios of C3d (1.5:1, 2:1 and 3:1) was suggested by the additional peaks labelled 6, 7 and 8.

C3d and SCR-16/20 were used as reference (Figure 6.5(a)). Analysis of up to 300 scans revealed excellent boundary fits (Figure 6.5). In the  $c(s)$  distribution of Figure 6.5(a), peak 2 for unbound C3d at 0.3 mg/ml in Figure 6.5(a) shows a  $s_{20,w}$  value of 2.99 ( $\pm 0.08$ ) S, and this single peak is unchanged up to 0.8 mg/ml (Figure 6.3(c,d)). When C3d is in excess of 1 mg/ml in Figure 6.5(b), peaks 4 and 5 correspond to the previously-suggested C3d dimer and trimer respectively (Figure 6.3(d)). It should be noted that the  $c(s)$  analysis will provide peaks for reversibly reacting systems, but these peaks correspond to reaction boundaries and not to distinct species. Accordingly molecular weight analyses may not be justified in the absence of other evidence (Section 3.2.5.3). Figure 6.5(b) thus shows that C3d undergoes an unusual co-operative association to states larger than dimer, where the lack of baseline resolution between peaks 4 and 5 is consistent with C3d undergoing mass action association. Peaks 1 and 3 for SCR-16/20 in Figure 6.5(a) show  $s_{20,w}$  values of 2.69 S and 3.30 S respectively which are consistent with a slow monomer-dimer equilibrium (Chapter 4). The possibility that SCR-16/20 dimerisation occurred as a result of the C-terminal hexahistidine tag was excluded by control AUC experiments. The study of SCR-16/20 using imidazole concentrations of 0.1 M, 0.3 M and 0.5 M showed that dimer was formed in similar amounts to that observed in PBS buffer (not shown).

Because weak complexes were identified by gel filtration (Figure 6.2(b)), the complex was investigated in 50 mM NaCl HEPES buffer by titrating C3d against a fixed concentration of 1.0 mg/ml SCR-16/20 (25.2  $\mu$ M) (Figure 6.5). The C3d molar ratios ranged from 0.3:1 to 3:1. In the  $c(s)$  plots, five additional peaks numbered 4 to 8 were observed that indicated the formation of a series of reversibly reacting SCR-16/20 and C3d complexes (Figures 6.5(a,b)). The underlying model for  $c(s)$  analysis assumes the same frictional ratio for all sedimenting species. In addition, while  $c(s)$  analysis is suitable for studying homogeneous or heterogeneous non-interacting systems, it has not been shown to be appropriate for solutions containing interacting components with multiple assembly states. Thus for the reversibly-interacting system described by this experiment, the  $c(s)$  peaks for the complexes will generally correspond to reaction boundaries and not to distinct molecular species. For the molar ratios up to 1:1 (Figure 6.5(a)), at least four additional peaks (numbered 4 to 7) could be resolved. Peaks 1-5 for unbound C3d and SCR-16/20 did not completely disappear in these mixtures, indicating that SCR-16/20 self-



**Figure 6.6**

Size distribution  $c(s)$  analyses of SCR-16/20, FH and C3d mixtures in 137 mM NaCl PBS or HEPES buffer.

(a) The unbound proteins in PBS buffer were assigned as peaks 1 and 3 (SCR-16/20 at 1.0 mg/ml) and peak 2 (C3d at 0.6 mg/ml) at 2.60 S, 2.91 S and 3.30 S in that order. Their 1:1 mixture at 1.8 mg/ml shows additional weak peaks 6, 7 and 8.

(b) The unbound proteins in PBS buffer were assigned as peak 2 (C3d at 0.6 mg/ml) and peaks A and B (FH monomer and dimer respectively at 1.53 mg/ml) in 137 mM NaCl HEPES buffer. Their 1.5:1 ratio of C3d: FH showed additional weak peaks labelled C, E and F. The presence of Peak X is artefactual.

interaction and C3d oligomerisation occur, even though SCR-16/20 binding to C3d has taken place. The relative stability of peaks 1-5 suggests a slow interaction. For the molar ratios above 1:1 (Figure 6.5(b)), the resolution and positions of peaks 6 and 7 changed, and another peak 8 became visible. This concentration dependence of peaks 6-8 is indicative of a faster equilibrium (Balbo & Schuck, 2005).

While it is impossible to determine exact peak positions for larger sedimenting species in fast-interacting systems, it is nonetheless possible to infer peak positions for smaller species with slow interactions (Balbo & Schuck, 2005). Hence the positions of peaks 1-5 can be deduced since these did not alter in their  $s_{20,w}$  values with concentration. Peak 4 with a mean  $s_{20,w}$  value of 4.25 ( $\pm 0.07$ ) S was seen at all molar ratios, and is assigned to the C3d dimer and possibly to complex formation between SCR-16/20 and C3d. Peak 5 with a mean  $s_{20,w}$  value of 5.55 ( $\pm 0.12$ ) S was seen clearly only at the higher C3d concentrations, and corresponds to the putative C3d trimer and possibly to further complex formation. In contrast, the  $s_{20,w}$  value of peak 6 ranged from 6.23 S to 8.43 S in the titration. Peak 7 with a mean  $s_{20,w}$  value of 10.49 ( $\pm 0.08$ ) S showed no concentration dependence up to a 1:1 molar ratio, and varied between 9.93 S to 10.66 S at higher molar ratios. Peak 8 was seen only at high molar ratios above 1:1, and has an  $s_{20,w}$  value of 11.90 S to 12.66 S. Even though precise  $s_{20,w}$  values could not be determined, the existence of the peaks shows that a series of large multimeric complexes have formed.

The conversion of the observed  $c(s)$  peaks into molecular weights provides some insight into the SCR-16/20 complex with C3d in 50 mM NaCl. Such an analysis assumes that all the sedimentation species possess the same frictional ratio  $f/f_o$  (where  $f_o$  is the frictional coefficient of the sphere with the same volume as the hydrated glycoprotein) and that the peaks do not correspond to reaction boundaries. The  $f/f_o$  ratio is a measure of structural elongation. The  $f/f_o$  values for the elongated SCR-16/20 monomer and dimer are 1.4 and 1.6 respectively (Chapter 4), while that for the globular C3d monomer is 1.1 (Li *et al.*, 2008), showing that such an analysis will be limited. The conversion of  $c(s)$  plots to  $c(M)$  plots (not shown) using a low  $f/f_o$  ratio of 1.185 suggested that peaks 1 and 3 corresponded to putative molecular weights of 29 kDa and 42 kDa. These values are less than the expected molecular weights for the SCR-16/20 monomer (40 kDa) and dimer (79 kDa) because of the low  $f/f_o$  ratio. Peaks 2, 4 and 5 correspond to molecular weights of 31 ( $\pm 3$ ) kDa, 55 ( $\pm 6$ ) kDa and 70 ( $\pm 9$ ) kDa respectively, which are also less than the expected

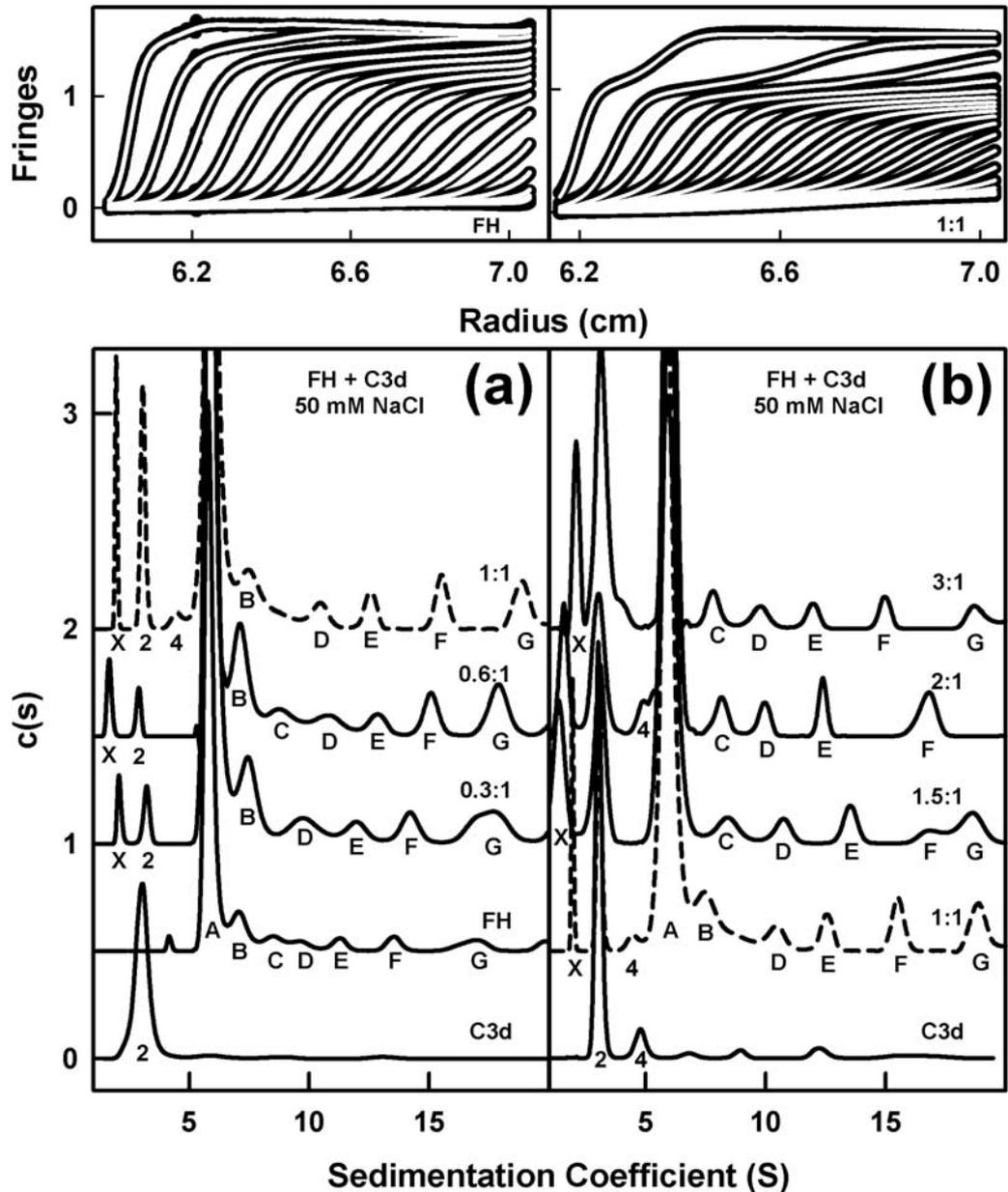
molecular weights for the C3d monomer, dimer and trimer of 35 kDa, 70 kDa and 105 kDa respectively. As the molecular weight for a 1:1 complex is 74 kDa, the discrepancies for peaks 4 and 5 may be explained by complex formation with low stoichiometric ratios. The putative molecular weights of peaks 6-8 are in a range of 111 ( $\pm 9$ ) to 222 ( $\pm 20$ ) kDa with a  $f/f_o$  ratio close to 1.1, while complexes with a 2:2 or 4:4 stoichiometry will have predicted molecular weights of 149 kDa or 297 kDa respectively. If peaks 6-8 are assigned to molecular species, the existence of peaks 6-8 shows that a range of multimeric FH-C3d complexes have been formed, although their stoichiometry cannot be determined because of the uncertainty over the  $f/f_o$  ratio.

The velocity data were repeated using physiological 137 mM NaCl PBS buffer (Figure 6.6(a)). Peaks 1, 2 and 3 were identified again in the  $c(s)$  analyses, although peaks 1 and 3 were slightly shifted in position with  $s_{20,w}$  values of 2.81 S and 3.89 S respectively, compared with those of 2.60 S and 3.30 S in 50 mM NaCl HEPES buffer. The putative molecular weights of 38 kDa (peak 1) and 67 kDa (peak 3) are comparable with the expected molecular weights of 40 kDa, and 79 kDa respectively. For the 1:1 mixture at 1.8 mg/ml, additional peaks 6, 7 and 8 were observed, showing that multimeric complex formation had occurred. Integration of the  $c(s)$  analyses showed that peaks 4-8 comprised 42% of the total signal area for the 1:1 mixture in 50 mM NaCl buffer, while this was reduced to 24% in 137 mM NaCl buffer. This decrease indicated a salt-dependence of complex formation.

### **(6.2.3.2) Sedimentation velocity data for the complex of native Factor H with C3d**

In order to assess the physiological relevance of multimeric SCR-16/20 complexes with C3d, the same AUC titrations were performed with native FH and C3d in 137 mM NaCl PBS buffer. Previously with native FH, oligomers in slow exchange were detected in the  $c(s)$  plot, of which the monomer and dimer were prevalent at 1.0 mg/ml FH (Nan *et al.*, 2008a). The monomer and dimer (peaks A and B respectively) were observed for FH at 1.53 mg/ml (Figure 6.6(b)). The addition of a 50% molar excess of C3d to 1.0 mg/ml FH resulted in multimers as evidenced by the appearance of peaks C, E and F (Figure 6.6(b)).

The FH complex was investigated in 50 mM NaCl HEPES buffer using a titration of C3d against a fixed concentration of 1.0 mg/ml FH (6.7  $\mu$ M). Good



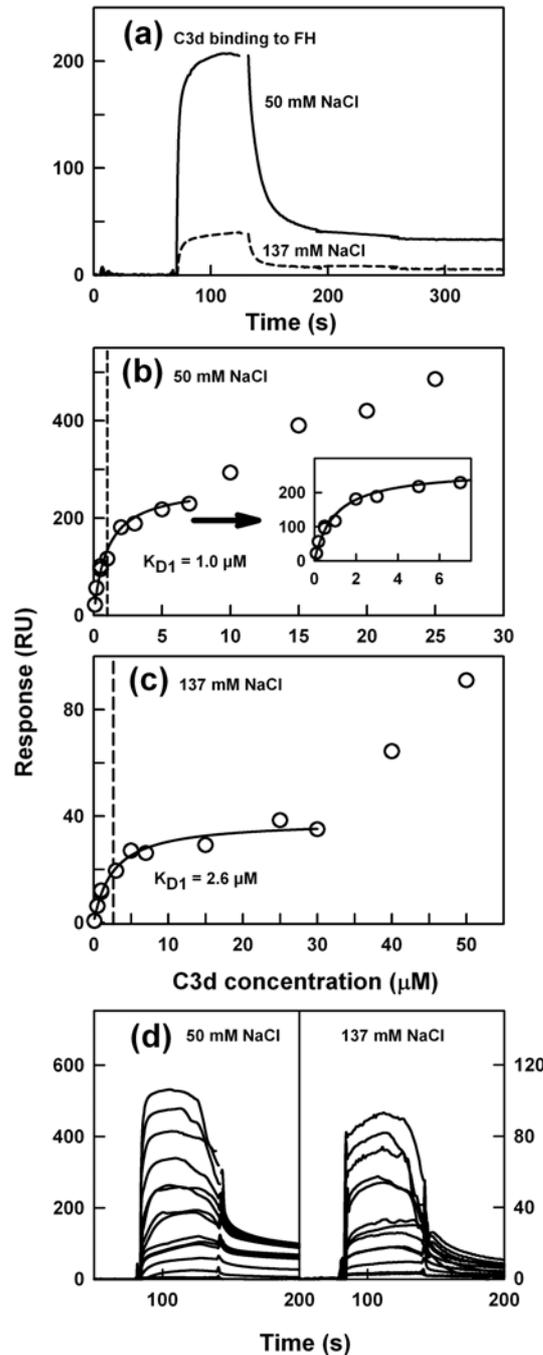
**Figure 6.7**

Size distribution  $c(s)$  analyses of native FH and C3d mixtures in 50 mM NaCl HEPES buffer. The SEDFIT analyses correspond to a rotor speed of 40,000 r.p.m. Representative boundary fits are shown in the two upper panels. The lower panels show nine  $c(s)$  analyses. For clarity, the plots were displaced arbitrarily along the  $c(s)$ -axis. Peaks 2 and 4 were assigned to C3d to follow [Figures 6.5](#) and [6.6](#). Peaks A and B were assigned as the FH monomer and dimer respectively. Other peaks were labelled C to G in order of increasing S value.

boundary fits were obtained in all cases (Figure 6.7). The  $c(s)$  plots revealed the more prominent presence of C3d-induced multimers at lower ionic strength (Figure 6.7). Unbound FH at 1.0 mg/ml exhibited at least seven oligomeric forms labelled A to G in slow exchange, and these have been assigned as predominantly FH monomers (peak A) with small amounts of FH dimers increasing to heptamers (peaks B to G) (Nan *et al.*, 2008a). The addition of C3d up to a 1:1 molar ratio showed that peaks B to G either shift to larger S values or increase in their size. If the peaks correspond to distinct species and/or reaction boundaries, this indicates that complexes had formed. Peak 2 that corresponded to monomeric unbound C3d was also observed in Figure 6.7 (a), indicating that C3d bound incompletely and weakly to FH. Above a 1:1 molar ratio, further changes in the FH peaks B to G positions and intensities occur, together with large increases in peak 2 for C3d (Figure 6.7 (b)). Unlike the case of SCR-16/20, more detailed analyses of the C3d and FH interaction were precluded by the complexity of the system under investigation.

#### **(6.2.4) Surface plasmon resonance studies of the complexes of Factor H and SCR-16/20 with C3d**

SPR monitors the interaction between a binding partner in solution (analyte) and a partner (ligand) attached to the surface of a sensor chip (Tanious *et al.*, 2008). Previous SPR studies of the FH-C3d interaction using immobilised C3d did not report multimer formation (Hellwage *et al.*, 1999, 2002; Jokiranta *et al.*, 2000, 2005; Manuelian *et al.*, 2003). Accordingly C3d binding to immobilised FH was studied in duplicate runs. The binding response decreased approximately five-fold from 50 mM NaCl HEPES to 137 mM NaCl HEPES buffers (Figure 6.8(a)), confirming the ionic strength-dependent nature of this interaction. In both buffers, steady-state affinity studies revealed that C3d bound to immobilised FH in two stages. The equilibrium dissociation constant  $K_{D1}$  for the first stage was well-determined between C3d concentrations of 0.1-7  $\mu$ M in 50 mM NaCl and 0.1-30  $\mu$ M in 137 mM NaCl (Figures 6.8(b,c)). In 50 mM NaCl buffer, the  $K_{D1}$  value was 1.0  $\mu$ M (Figure 6.8(b)), while this was increased to 2.6  $\mu$ M in 137 mM NaCl buffer (Figure 6.8(c)). A second association phase was indicated by a significant increase in the response above 7  $\mu$ M C3d in 50 mM NaCl buffer (Figure 6.8(b)) and above 30  $\mu$ M in 137 mM NaCl buffer (Figure 6.8(c)). However  $K_{D2}$  values could not be well determined for the responses



**Figure 6.8**

SPR analysis of C3d binding to immobilised FH.

(a) Approximately five-fold more C3d binds to FH in 50 mM NaCl buffer compared to 137 mM NaCl buffer. (b) Steady-state affinity analyses for C3d binding to FH in 50 mM NaCl buffer. A secondary increase in the binding response occurred at concentrations above 7 µM C3d (0.2 mg/ml). The equilibrium dissociation constant  $K_{D1}$  was determined for C3d between 0-7 µM (see inset for greater clarity). (c) In 137 mM NaCl buffer, the  $K_{D1}$  value was determined for C3d between 0-30 µM. (d) Sensorgram overlays showing the data points used in the affinity studies of (b) and (c).

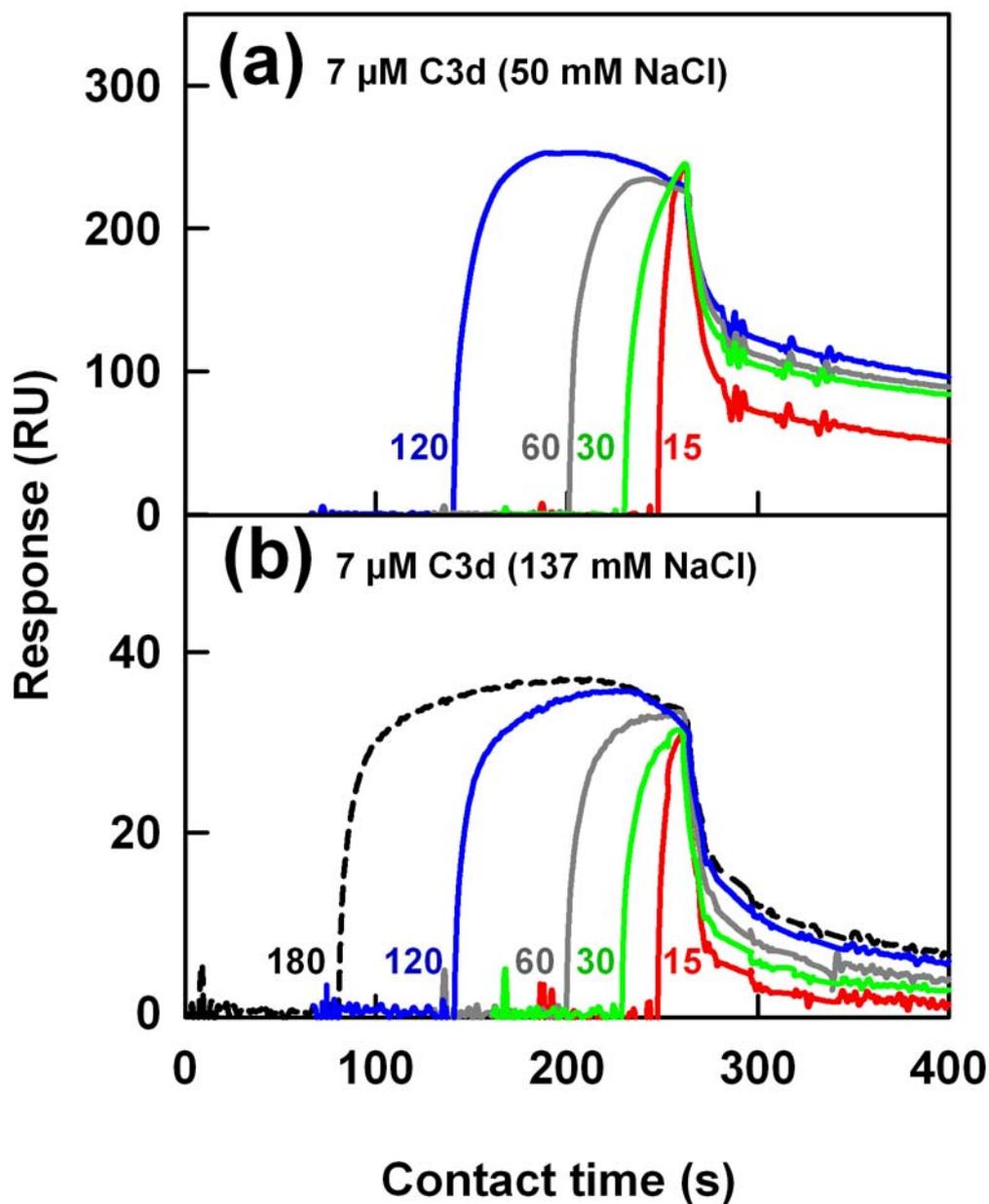
in this second phase because steady-state saturation could not be fully attained. The existence of two  $K_D$  values would be consistent with the presence of both a FH-C3d complex and a multimer complex.

Overlay experiments were performed in order to test further the hypothesis that the C3d interaction with FH is not a simple 1:1 binding reaction. C3d at 7  $\mu\text{M}$  in both 50 mM and 137 mM NaCl buffers was injected onto immobilised FH for increasing amounts of time from 15 sec to 180 sec. The data were overlaid at the start of the dissociation phase (Figure 6.9). A simple binding reaction would be revealed by the dissociation phases lying on top of each other. Instead of this, it was observed in both buffers that, the longer the contact time became, the slower the dissociation time became. This indicated a linked reaction caused by C3d multimer formation on immobilised FH. While alternative explanations based on (i) conformational rearrangements in C3d, or (ii) the presence of a secondary C3d-binding site in immobilised FH are possible, these alternatives are unlikely because (i) C3d possesses a compact domain structure that is unlikely to undergo conformational rearrangement on ligand-binding (Gilbert *et al.*, 2005; Li *et al.*, 2008) and (ii) previous studies show that FH possesses a single C3d-binding site (Sharma & Pangburn, 1996; Jokiranta *et al.*, 2000; Lambris *et al.*, 1988; Jokiranta *et al.*, 2001). The maximum response in this experiment is given by the molecular weight ratio of analyte to ligand (36/154) multiplied by the stoichiometry of 1 and the amount of immobilised FH ligand (700 RU). The observed response of over 200 RU at first-stage saturation in Figures 6.8(b) exceeds the predicted maximum response of 164 RU, which may suggest that C3d multimerisation on immobilised FH has occurred in this buffer. In the light of this observation, the weaker saturation of 40 RU and the changes above 30  $\mu\text{M}$  C3d in 137 mM NaCl (Figures 6.8(c)) are best explained by the occurrence of C3d multimerisation in preference to the FH-C3d interaction in this buffer.

Next, the formation of multimeric FH-C3d complexes in physiological buffer conditions (137 mM NaCl buffer) was studied in more detail using SCR-16/20 or SCR-1/5 fragments with or without C3d as the analyte (Figure 6.10). First, 5  $\mu\text{M}$  of C3d and SCR-16/20 were each passed over immobilised FH (Figure 6.10(a)). Each protein bound to FH. For SCR-16/20, this was in agreement with the monomer-dimer equilibrium for SCR-16/20 shown by AUC (Chapter 4). Next, an equimolar mixture of C3d and SCR-16/20 was flowed over immobilised FH (Figure 6.10 (a)). Note that

the concentration of each protein in the mixture is reduced by half when compared to the unbound proteins in [Figure 6.10\(a\)](#), due to the doubling of the volume on mixing. Given the similar molecular weights of both proteins, the total response of this mixture would be similar to that of either C3d or SCR-16/20 if the proteins compete for binding. The doubling of the response in [Figure 6.10\(a\)](#) therefore indicated the formation of a ternary complex ([Jokiranta et al., 2005](#)). As a control, the corresponding equimolar mixture of SCR-1/5 and C3d showed a reduction in signal response, as expected ([Figure 6.10\(b\)](#)). This shows that SCR-16/20 dimerisation occurs even when C3d is bound, in agreement with the observation of multimeric species in the AUC data of [Figure 6.5](#).

Final experiments in 137 mM NaCl HEPES buffer studied two concentrations of FH in the presence of increasing amounts of C3d (both proteins in the analyte) and using immobilised FH as ligand. Thus C3d in molar ratios ranging from 0:1 to 10:1 was mixed with native FH at two starting concentrations of 0.08 mg/ml (0.5  $\mu$ M) and 0.8 mg/ml (5  $\mu$ M) ([Figure 6.11\(a,b\)](#)). FH self-associates in a concentration-dependent manner through at least two sites localised within SCR-6/8 and SCR-16/20, with the associated forms being more stable at higher FH concentrations ([Fernando et al., 2007](#); [Chapter 4](#); [Nan et al., 2008a](#)). Because of the approximately four-fold mass difference between FH and C3d, a FH-C3d complex will exhibit a lower response than a FH-FH dimer. In the absence of C3d, a high response was observed to show that FH forms at least some FH-FH dimer and possible higher oligomers ([Nan et al., 2008a](#)). On addition of an equal volume of C3d to FH in the six C3d:FH ratios, the halved FH concentration to 2.5  $\mu$ M (from 5  $\mu$ M) and 0.25  $\mu$ M (from 0.5  $\mu$ M) leads to a decreased response compared to the start of [Figures 6.11\(a,b\)](#) for reason of reduced FH-FH association. At these two FH concentrations, the six responses generally decreased with increased C3d concentration. This was more pronounced for 2.5  $\mu$ M FH (from 100 RU to 79 RU) than 0.25  $\mu$ M FH (from 30 RU to 25 RU). This reduction is attributed to the reduced formation of FH-FH associated forms as C3d increases. This shows that C3d partially competes with FH-FH association, although this is more likely to involve higher oligomers given the outcome of [Figure 6.10\(a\)](#). As a non-FH binding benchmark, the 2.5  $\mu$ M FH analyte was replaced by 2.5  $\mu$ M SCR-1/5 in the same C3d titration. This yielded a baseline value for the C3d interaction with immobilised FH that increases with C3d

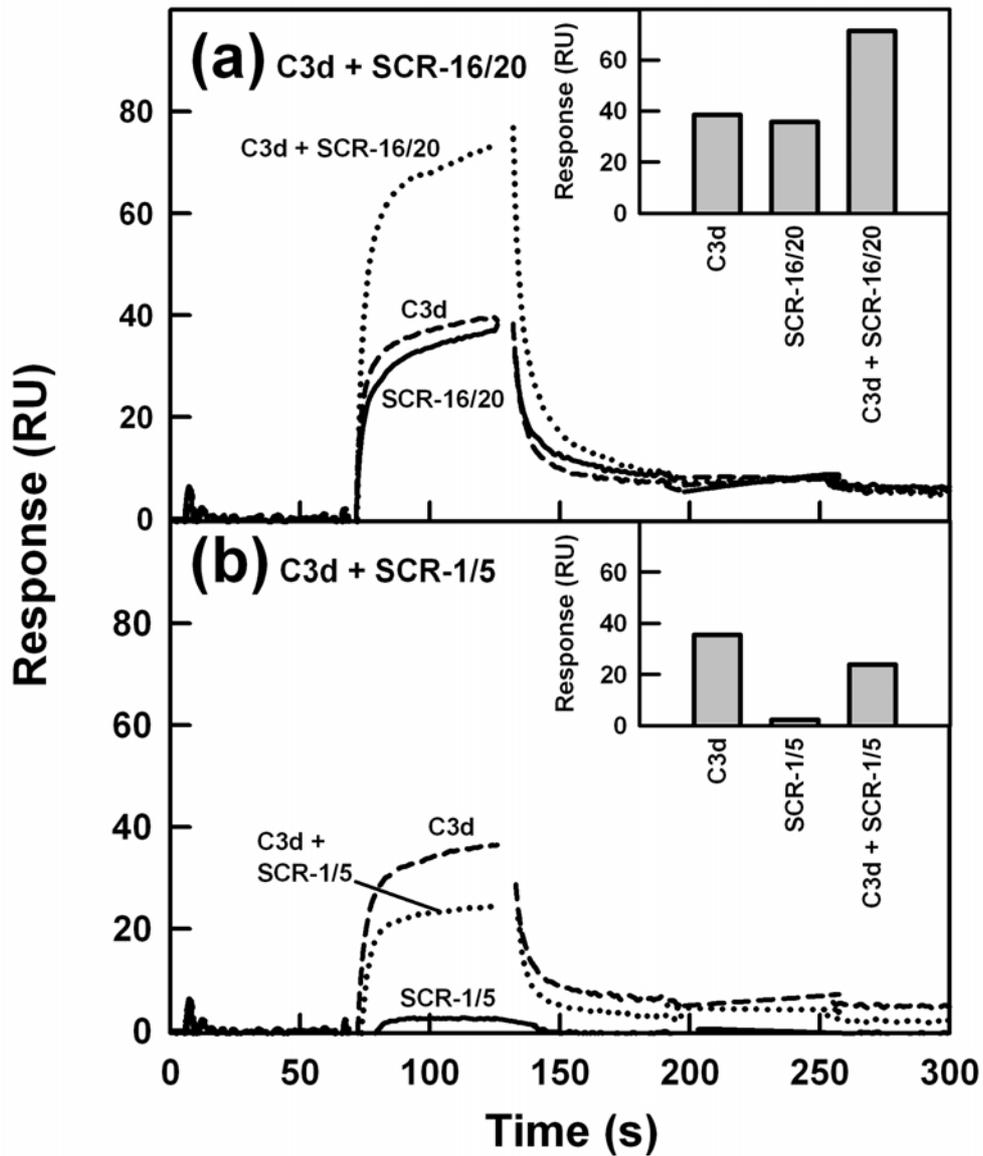


**Figure 6.9**

Time-resolved SPR analysis of C3d binding to immobilised FH.

(a) In 50 mM NaCl buffer, 7  $\mu$ M of C3d analyte was flowed over immobilised FH for 15 sec, 30 sec, 60 sec and 120 sec. The data were overlaid at the start of the dissociation phase.

(b) As for (a) but using 137 mM NaCl buffer.

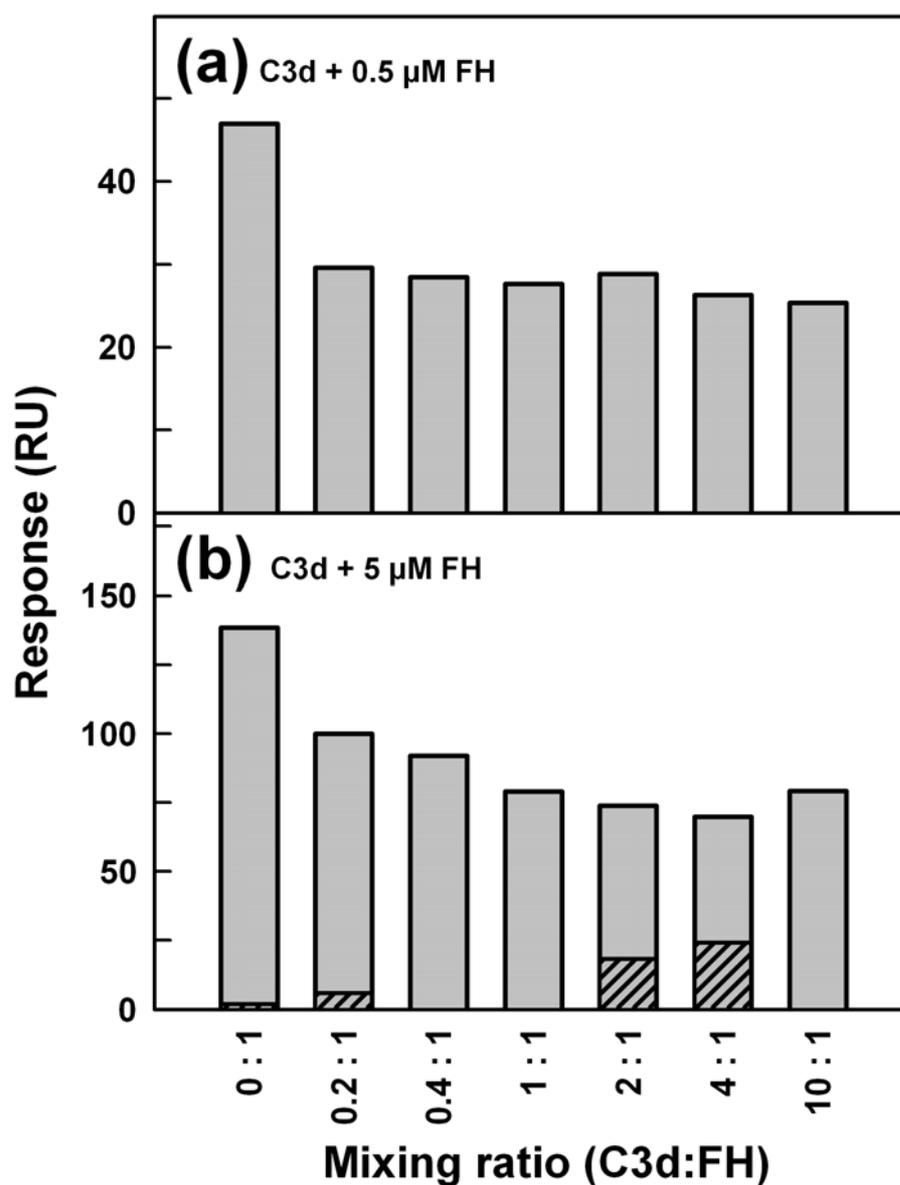


**Figure 6.10**

SPR analyses of ternary complexes formed between C3d, SCR-16/20 and immobilised FH in 137 mM NaCl buffer. Bar charts of the binding response are inset for clarity.

(a) SCR-16/20 and C3d bound to immobilised FH with similar responses, and their equimolar mixture showed a doubled response.

(b) C3d bound to immobilised FH, while SCR-1/5 showed no binding, even when both proteins were present as an equimolar mixture.



**Figure 6.11**

SPR analyses of ternary complexes formed between C3d and FH in the presence of immobilised FH in 137 mM NaCl buffer. Binding responses are shown as bar charts for clarity.

(a) In the presence of a starting concentration of 0.5 μM FH (0.08 mg/ml), the response decreased from 47 to 25 RU as the C3d:FH ratio increased from 0:1 to 10:1.

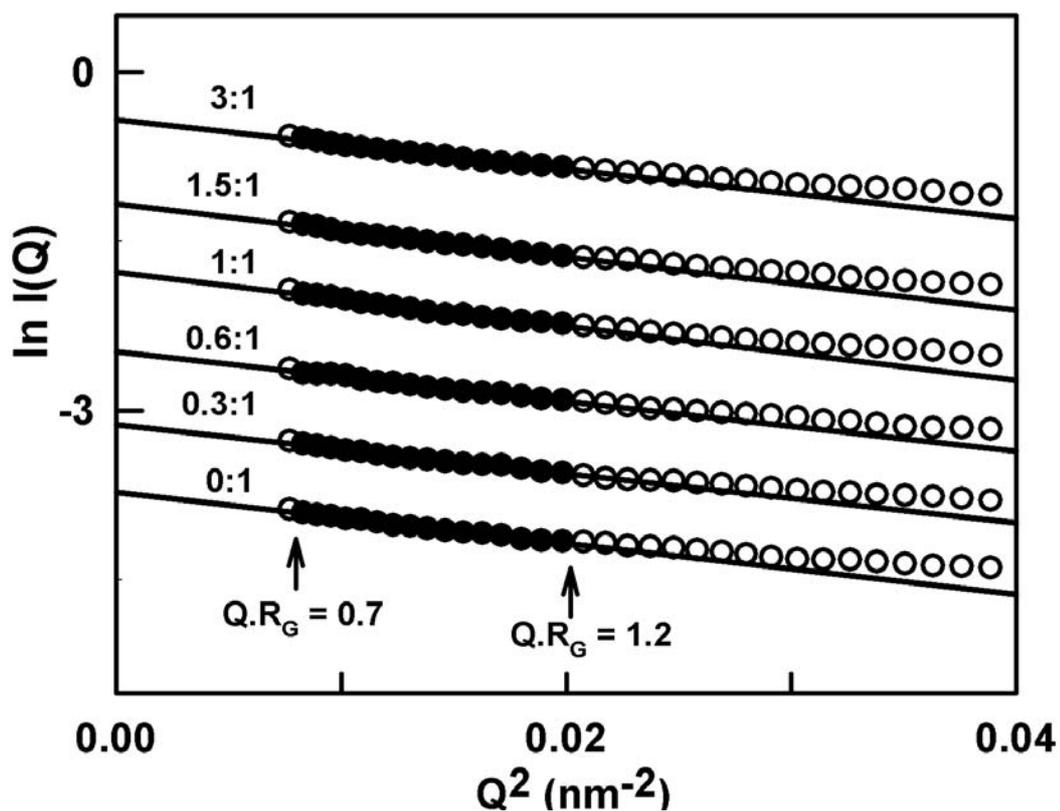
(b) In the presence of a starting concentration of 5 μM FH, the response decreased from 138 to 79 RU as the C3d:FH ratio increased from 0:1 to 10:1. If FH was replaced by SCR-1/5 with a starting concentration of 5 μM, the response is much reduced as the C3d:FH ratio increased (hatched bars).

concentration (hatched in [Figure 6.11\(b\)](#)). Compared to this baseline, the FH-C3d signal is larger than a C3d-only binding event, even at the highest C3d concentrations. This shows that the FH analyte binds to immobilised FH even in the presence of C3d, in accordance with [Figure 6.10\(a\)](#). The higher signal response of 79 RU for the 10:1 mixture in [Figure 6.11\(b\)](#) compared to that of 38 RU for the corresponding C3d concentration of 25  $\mu\text{M}$  in [Figure 6.8\(c\)](#) supports this conclusion. These experiments support the AUC observations of multimeric FH-C3d complexes in 137 mM NaCl ([Figures 6.6\(b\); 6.7](#)).

### **(6.2.5) X-ray scattering data of the SCR-16/20 complex with C3d**

Solution scattering is a diffraction technique that studies the overall structure of biological macromolecules in solution ([Perkins \*et al.\*, 2008](#)). The C3d-SCR-16/20 complexes were analysed by synchrotron X-ray scattering. Time frame analyses showed that the X-ray radii of gyration ( $R_G$ ) were affected by minor radiation damage only by the longest exposure times of 0.5 sec per frame. To eliminate these effects, data were therefore collected using a flow cell with ten exposures of 0.1 sec/frame, and the ten time frames were averaged to improve the signal-noise ratio. In Guinier analyses of  $\ln I(Q)$  vs.  $Q^2$  to determine the  $R_G$  (a measure of elongation) and the intensity of forward scattering  $I(0)/c$  (a measure of molecular weight;  $c$  = concentration), the fits were linear within satisfactory  $Q \cdot R_G$  limits ( $Q = 4 \pi \sin \theta / \lambda$ ;  $2\theta$  = scattering angle;  $\lambda$  = wavelength) ([Figure 6.12](#)). This indicates the monodispersity of the proteins and the absence of non-specific aggregates. SCR-16/20 at 1.0 mg/ml was titrated with C3d at six molar ratios of 0.3:1, 0.6:1, 1:1, 1.5:1, 2:1 and 3:1 in 50 mM NaCl HEPES buffer ([Figures 6.12; 6.13](#)). Unbound SCR-16/20 and C3d were used as controls ([Gilbert \*et al.\*, 2005; Chapter 4](#)). In the titration, the  $I(0)/c$  values decreased linearly from 0.0237 ( $\pm 0.0007$ ) for unbound SCR-16/20 to 0.0167 ( $\pm 0.0002$ ) for the mixture at a molar ratio of 3:1 (not shown). This decrease is consistent with weak complex formation between SCR-16/20 and C3d.

The analysis of changes in the  $R_G$  values indicated complex formation. Unbound SCR-16/20 had a mean  $R_G$  value of 8.05 ( $\pm 0.36$ ) nm (4 determinations), in good agreement with the  $R_G$  of  $8.09 \pm 0.21$  nm for its dimeric form between 0.7-1.2 mg/ml in 137 mM NaCl PBS buffer ([Chapter 4](#)). Unbound C3d showed a concentration-dependent increase in its  $R_G$  value from 1.8 nm to 3.5 nm,



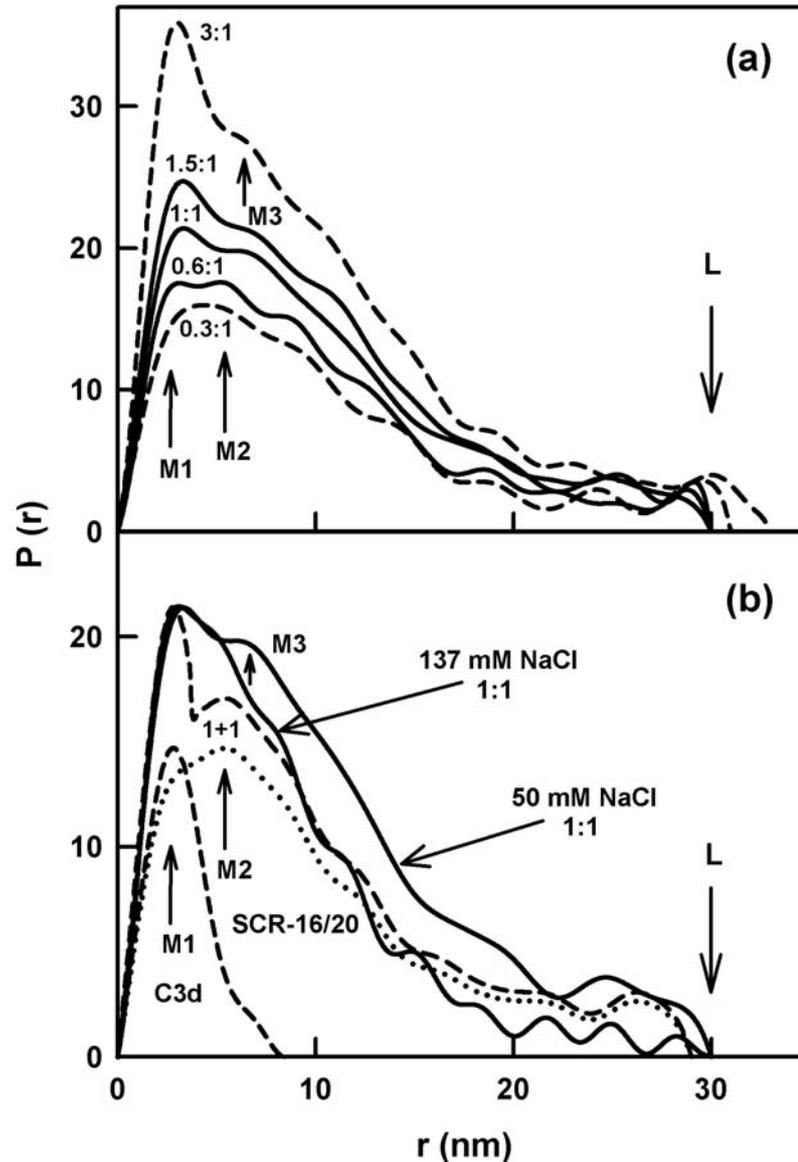
**Figure 6.12**

X-ray Guinier  $R_G$  analyses for SCR-16/20 and C3d mixtures in 50 mM NaCl HEPES buffer. The SCR-16/20 concentration is 1.0 mg/ml. From bottom to top, the molar ratios of C3d: SCR-16/20 are 0:1, 0.3:1, 0.6:1, 1:1, 1.5:1 and 3:1. The Guinier  $R_G$  plots are displaced vertically for clarity. The filled circles between the arrowed  $Q.R_G$  ranges correspond to the  $R_G$  fits using a  $Q$  range of 0.09-0.14 nm<sup>-1</sup>. The straight line corresponds to the best fit through the filled circles.

corresponding to its monomeric, dimeric and trimeric forms (Gilbert *et al.*, 2005; Li *et al.*, 2008). In the titration using 50 mM NaCl buffer, the mean  $R_G$  value increased from 8.05 ( $\pm$  0.36) nm (ratio of 0:1 C3d:SCR-16/20) and 8.26 ( $\pm$  0.25) nm (ratio of 0.3:1) to 8.54 ( $\pm$  0.12) nm (ratio of 1.5:1), after which it decreased to 8.10 ( $\pm$  0.10) nm (ratio of 3:1) (Figure 6.12). The increases indicated complex formation. In the titration with 137 mM NaCl PBS buffer, the  $R_G$  value of the 1:1 mixture was lower at 6.89 ( $\pm$  0.26) nm (3 values), and corresponds to the weighted sum of the  $R_G$  values for unbound monomeric and dimeric SCR-16/20 and monomeric C3d (Perkins *et al.*, 1985). This shows that complex formation is much reduced in 137 mM NaCl buffer.

The distance distribution function  $P(r)$  calculated from the full  $I(Q)$  curve (Section 6.4.4) allows the determination of overall lengths following the assumption of the value of the maximum dimension  $L$  (Figure 6.13). The  $P(r)$  curve provides an independent verification of the Guinier  $R_G$  values and this gave 8.33 ( $\pm$  0.05) nm for unbound SCR-16/20. This is comparable with that of 8.05 ( $\pm$  0.36) nm from the Guinier analysis, showing that the  $P(r)$  curves refer to the macromolecule being studied. In the titration with C3d,  $R_G$  values that increased from 8.32 nm to 8.88 nm were obtained again. The maximum macromolecular dimension  $L$  was determined from the value of  $r$  when  $P(r) = 0$  at large  $r$ . The  $L$  value for unbound SCR-16/20 in 50 mM NaCl was 29 nm (Figure 6.13(a)), again similar to that of 27 nm in 137 mM NaCl (Chapter 4). Even though the  $P(r)$  analyses were not precise at large  $r$  value, the  $L$  values for all molar ratios for the complex ranged between 30 to 33 nm (Figure 6.13(a)). This shows that no major conformational change had occurred on complex formation. Since the maximum length of C3d from its crystal structure is measured to be 7 nm, the lack of change from the unbound  $L$  value of 27-29 nm for the mixtures shows that C3d is incorporated within the length of dimeric SCR-16/20 in its complex. An end-to-end arrangement with SCR-16/20 would have increased the overall length, but this was not observed. As it is known that C3d binds to SCR-20 (Jokiranta *et al.*, 2000; Hellwage *et al.*, 2002), the lack of change in the  $L$  value suggests that the SCR-16/20 dimer is formed by contacts between the two SCR-20 domains, and not between the two SCR-16 domains. The location of the dimer site could not be determined previously (Chapter 4).

Changes in the  $P(r)$  curves demonstrated the occurrence of complex formation (Figure 6.13(a)). The maximum  $M$  corresponds to commonly occurring interatomic distances within the structure. Peak  $MI$  with  $r = 3.3$  nm is unique and



**Figure 6.13**

X-ray distance distribution functions  $P(r)$  for SCR-16/20 and C3d mixtures. The SCR-16/20 concentration is 1.0 mg/ml.

(a) X-ray distance distribution functions for five of the analyses in Figure 6.12. The curves are shown as dashed and full lines for clarity. The maxima  $M1$ ,  $M2$  and  $M3$  represent commonly occurring distances within the macromolecules, and were 2.8 nm, 5.5 nm and 6.6 nm in that order. The maximum dimension is denoted by  $L$  and was 30 nm in all cases.

(b) Comparison of the  $P(r)$  curves for unbound SCR-16/20 and C3d with their 1:1 mixtures in 50 mM NaCl HEPES and 137 mM NaCl PBS buffers. The intensities of  $M1$  and  $M2$  for the two unbound proteins were set equal for clarity, and their sum is labelled 1+1 (dashed line). The  $P(r)$  curves in 50 mM NaCl HEPES and 137 mM NaCl PBS buffers are denoted by full lines. The maximum of the latter three  $P(r)$  curves at  $r = 3$  nm were set equal for clarity.

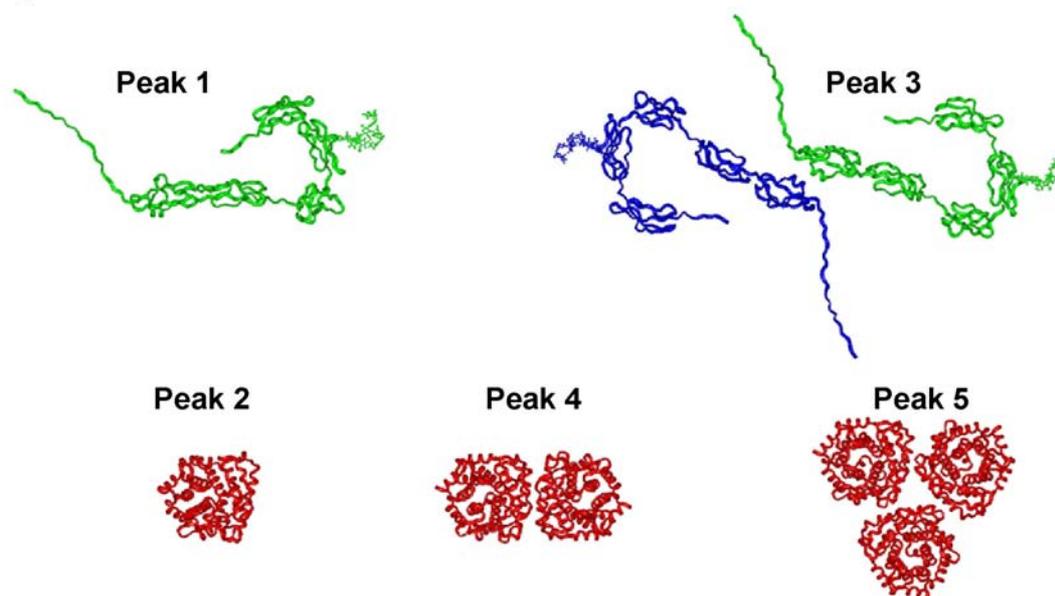
corresponds to unbound C3d (Gilbert *et al.*, 2005). Peak *M1* was visible in mixtures with molar ratios from 0.6:1 to 3:1, indicating that unbound C3d was present and that weak complex formation has occurred in these mixtures (Figure 6.13(a)). Peak *M2* with  $r = 5.5$  nm corresponds to the unbound SCR-16/20 dimer (Chapter 4). Peak *M2* weakened during the titration. In 50 mM NaCl, peak *M3* with  $r = 6.6$  nm indicates complex formation, because this peak was not seen for unbound C3d or SCR-16/20 (Figure 6.13(b)). In 137 mM NaCl, peak *M3* was not visible, showing that complex formation had been reduced. The 50 mM and 137 mM NaCl  $P(r)$  curves can be compared with the sum of the unbound C3d and SCR-16/20  $P(r)$  curves (1+1 in Figure 6.13(b)). The intensities in the  $P(r)$  curve increased more for 50 mM NaCl than for 137 mM NaCl for  $r$  values above 7 nm. This agrees with the greater complex formation in 50 mM NaCl compared to 137 mM NaCl.

#### (6.2.6) Molecular modelling of SCR-16/20 complexes with C3d

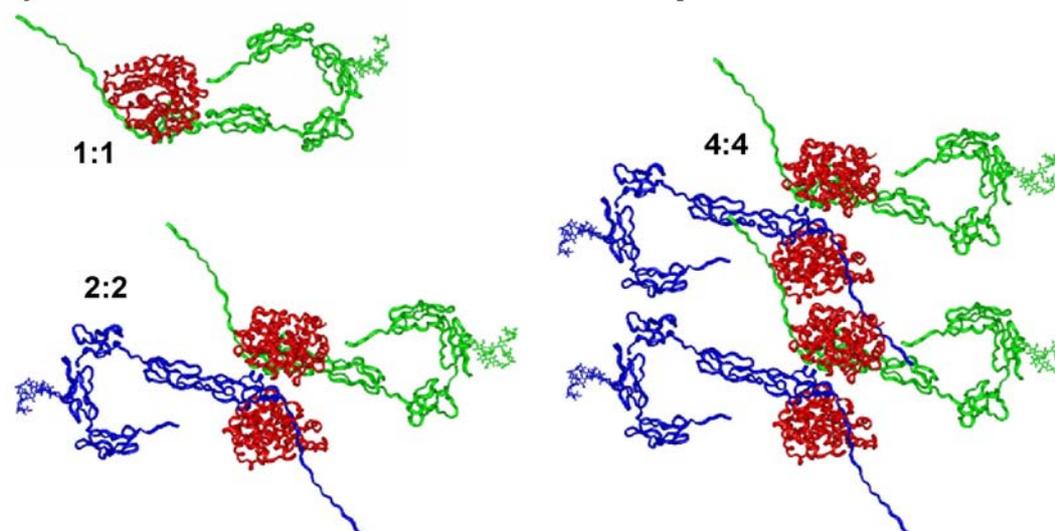
Constrained modelling was performed in order to provide insight into the  $s_{20,w}^0$  values observed in the AUC titrations. The constraints were the crystal structure of C3d (Nagar *et al.*, 1998) and the best-fit model for the SCR-16/20 dimer (Chapter 4). C3d and SCR-16/20 were assumed to form only symmetric dimers in its complexes with SCR-16/20. The calculation of sedimentation coefficients for Peaks 1-5 gave reasonable agreements for these five molecular species to within 0.3 S, and the  $f/f_o$  ratios ranged between 1.1 and 2.0 (Table 6.1(a)). For the additional peaks 6-8, it was assumed that C3d interacted with SCR-20, and that no conformational change occurs in SCR-16/20 during complex formation. This C3d location is similar to two alternative hypothetical docking models for SCR-20 and C3d, in which C3d was located at one side of SCR-20 near its C-terminal tip (Jokiranta *et al.*, 2006; Herbert *et al.*, 2006). A range of stoichiometries was explored.

The simplest scheme was based on equimolar stoichiometries of 1:1, 2:2, 4:4, 6:6 and 8:8 (Figure 6.14). If peaks 4 to 8 were assigned to these, the comparison between the pairs of experimental and mean predicted  $s_{20,w}^0$  values were respectively in reasonable agreement with each other at 4.3 S and 3.6 S (1:1 complex), 5.6 S and 5.4 S (2:2 complex), 6.2-8.4 S and 8.0 S (4:4 complex), 9.9-10.7 S and 9.8 S (6:6 complex), and 11.9-12.7 S and 11.7 S (8:8 complex) (Table 6.1(b)). The comparison between the calculated and the mean predicted  $f/f_o$  ratios likewise gave reasonable agreements at 1.33 and 1.56 (1:1), 1.61 and 1.65 (2:2), 1.68 and 1.77 (4:4), 1.75 and

### (a) Unbound SCR-16/20 and C3d

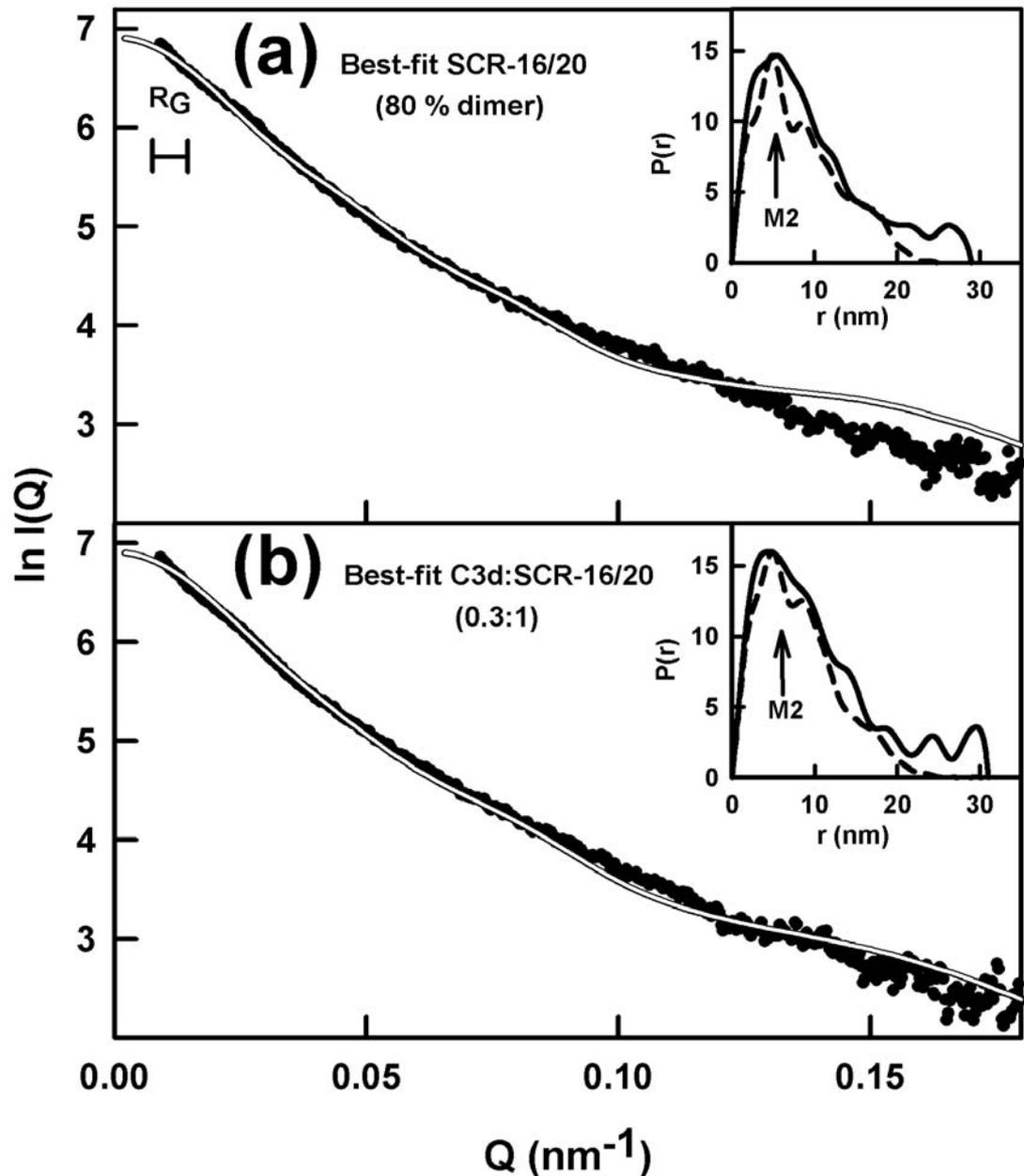


### (b) Putative SCR-16/20-C3d complexes



**Figure 6.14**

Molecular models for unbound SCR-16/20 and C3d and their complexes. C3d is shown in red and bound to SCR-20, and SCR-16/20 is shown in green and blue. (a) Unbound proteins; (b) Models for their putative 1:1, 2:2 and 4:4 complexes. The 6:6 and 8:8 complexes noted in Table 1 were constructed in a similar manner to that of the 2:2 and 4:4 complexes.



**Figure 6.15**

X-ray scattering fits for SCR-16/20 and its C3d complex.

(a) The X-ray curve for SCR-16/20 was fitted to 80% monomer and 20% dimer to follow Figure 6.14. The  $Q$  range for the  $R_G$  analyses is denoted by a horizontal bar. The inset shows the comparison of the modelled (dashed) and experimental (full)  $P(r)$  curves.

(b) The X-ray curve fit for a 0.3:1 mixture of C3d and SCR-16/20 was achieved assuming 10% of a 1:1 complex, 7% of a 2:2 complex, 7% of a 4:4 complex, and 3% of a 6:6 complex to follow Figure 6.14.

**Table 6.1. Sedimentation coefficient data and their modelling**

	Observed $s_{20,w}^0$ (S)	Modelled $s_{20,w}^0$ (S) <sup>a</sup>	Hydrated spheres	$f/fo$ (observed) <sup>b</sup>	$f/fo$ (model)
(a) Unbound proteins					
SCR-16/20 monomer (Peak 1)	2.69	2.29, 2.17	446	1.35	1.69-1.78
C3d monomer (Peak 2)	2.99 ± 0.08	3.14, 2.91	330	1.00	1.01-1.10
SCR-16/20 dimer (Peak 3)	3.30	3.20, 3.06	857	1.58	1.92-2.00
C3d dimer (Peak 4)	4.25 ± 0.07	4.67, 4.54	654		1.08-1.11
C3d trimer (Peak 5)	5.55 ± 0.12	5.88, 5.78	970		1.13-1.15
(b) Putative complexes					
1:1 complex (Peak 4)	4.25 ± 0.07	3.73, 3.49	772	1.33	1.51-1.61
2:2 complex (Peak 5)	5.55 ± 0.12	5.59, 5.29	1565	1.61	1.60-1.69
4:4 complex (Peak 6)	6.23 – 8.43	n.a , 8.03	3005	1.68	1.77
6:6 complex (Peak 7)	9.93 – 10.66	n.a , 9.75	4662	1.75	1.91
8:8 complex (Peak 8)	11.90 – 12.66	n.a , 11.71	6098	1.86	1.93

<sup>a</sup> The first value is calculated from the sphere model using HYDRO; the second is calculated from the coordinate model using HYDROPRO.

<sup>b</sup> These putative values are calculated assuming each peak represents a single sedimenting species.

1.91 (6:6), and 1.86 and 1.93 (8:8) (Table 6.1). While no unambiguous peak assignments could be made, and while it is stressed that the modelling assumed that peaks 4 to 8 correspond to molecular species and not to reaction boundaries, the calculation indicate that a range of multimeric complexes may account for the appearance of additional peaks in the  $c(s)$  fits of Figure 6.5.

Constrained modelling also provided insight into the X-ray scattering analyses. The overall length of the SCR-16/20 complexes was not greatly increased in the C3d titration (Figure 6.13(a)). In agreement with this, the overall length of 25 nm for the model of the SCR-16/20 dimer was only slightly increased to 29 nm for the model of the 6:6 complex. Scattering curve fits for the unbound SCR-16/20 monomer and dimer at 1.0 mg/ml was best achieved with a composition of 20% monomer: 80% dimer, giving a low goodness of fit R-factor of 4.9 % (Figure 6.15(a)). The scattering curve fits for the 0.3:1 molar ratio assumed that all the C3d was bound to SCR-16/20. The relative proportion of unbound SCR-16/20 monomer and dimer was taken to be 9% and 36% respectively, and that for the 1:1, 2:2, 4:4 and 6:6 complexes was taken to be 20%, 14%, 14% and 6% in that order from integration of the  $c(s)$  fits. The weighted summation of the scattering curves from this resulted in a low R-factor of 4.2 % in the curve fit (Figure 6.15(b)). While the agreement of the experimental and predicted  $R_G$  values were not close, the achievement of these X-ray fits indicated that a mixture of multimeric species was likely to be present.

### (6.3) Conclusions

This study of the interaction of the major complement regulator FH with C3d, a functionally relevant cleavage fragment of C3b, has revealed unexpected multimerisation properties. Multimerisation can be significant, in that oligomer formation can enhance or inhibit protein-protein interactions during complement activation. In this context, FH can be either immobilised on the host cell surface or in fluid phase as an abundant soluble plasma protein. One major feature of this study is that physiologically-appropriate protein concentrations were studied in 137 mM NaCl buffer. FH plasma concentrations range between 0.235 and 0.810 mg/ml (1.6  $\mu$ M to 5.4  $\mu$ M) (Saunders *et al.*, 2007). At these concentrations, an estimated 5% to 15% of FH will be dimeric, if no other factors require consideration (Nan *et al.*, 2008a). The molar equivalent physiological concentration of SCR-16/20 would be between 0.06 and 0.2 mg/ml. The normal C3 concentration range in plasma is

between 1 and 1.5 mg/ml (6.3  $\mu$ M to 7.9  $\mu$ M) (Law & Reid, 1995; Morley & Walport, 2000). If all the C3 is fully cleaved, the maximum C3d concentration would be 0.27 mg/ml (7.9  $\mu$ M). However, as C3 is an acute-phase reactant, its concentration can potentially increase much more during inflammation, and any localised production of C3 could also result in high C3 concentrations at inflammatory sites. The other major feature of this study is the use of both 50 mM NaCl and 137 mM NaCl buffers. The studies in 50 mM NaCl produce comparatively strong binding effects, but these are much reduced in physiological 137 mM NaCl. Using both AUC and SPR, studies using 50 mM NaCl show whether or not any interaction occurs, and studies in 137 mM NaCl show whether or not these are biologically significant. In this study, observations of FH-C3d multimers occur in physiologically-relevant concentration ranges in a physiological 137 mM NaCl buffer.

The SCR-16/20 multimeric interaction with C3d was identified by a combination of size-exclusion chromatography, AUC, SPR, X-ray scattering and molecular modelling. Size-exclusion chromatography identified the occurrence of complex formation in 50 mM NaCl buffer but less so in 137 mM NaCl, however this is a qualitative method (Figure 6.2). In distinction, multimerisation events were more clearly detected by AUC. While it was not possible to distinguish between the observation of distinct multimeric species or reaction boundaries from the appearance of peaks 6, 7 and 8 (Figure 6.5), AUC clearly showed that a simple 1:1 complex cannot account for the interaction between SCR-16/20 and C3d. The binding is weak in 50 mM NaCl, and is weaker but observable in physiological 137 mM NaCl (Figures 6.5; 6.6(a)). Structural modelling showed no indication of large conformational changes when SCR-16/20 binds to C3d. This is because the  $s_{20,w}^0$  values of the putative complexes could be replicated starting from previously-determined structures for unbound SCR-16/20 and C3d (Table 6.1). Thus there is no indication of any allosteric mechanism involving conformational changes upon C3d binding to FH. Further insight was obtained from SPR studies using immobilised FH, where ternary complex formation between C3d, SCR-16/20 and FH was detectable, i.e. this was not limited to a 1:1 complex (Figure 6.10(a)). X-ray scattering also showed that complex formation was detectable from the observed large alterations in the  $P(r)$  curves for the mixtures of C3d and SCR-16/20 when compared to the two proteins on their own (Figures 6.13(a,b)). The changes were seen in both 50 mM NaCl and 137 mM NaCl (Figure 6.13(b)). By X-rays, the existence of multimeric

C3d and SCR-16/20 complexes in 50 mM NaCl was also revealed by the inability to fit the scattering curve according to a 1:1 model. Instead, the use of multimeric complexes offered a much improved fit, although the nature of these multimers could not be unambiguously determined (Figure 6.15(b)).

The multimeric interaction between native FH and C3d was identified using AUC and SPR. AUC revealed weak complex formation for FH-C3d in both 50 mM and 137 mM NaCl (Figures 6.6(b) and 6.7). The stronger association in 50 mM NaCl is in agreement with the results for SCR-16/20. However the complexity of the changes seen in the  $c(s)$  plots do not permit a simple interpretation. This aspect was clarified by SPR studies of C3d binding to immobilised FH, in which  $K_D$  fits could not be achieved with a 1:1 binding model for FH-C3d. As the C3d concentration increased, the amount of C3d bound to FH in both 50 mM NaCl and 137 mM NaCl increased beyond that expected from a 1:1 interaction (Figure 6.8(b,c)). This indicated a multimeric interaction. The strength of this binding was again weaker in 137 mM NaCl compared to 50 mM NaCl.

The multimeric interaction between native FH and C3d was identified using AUC and SPR. AUC revealed weak complex formation for FH-C3d in both 50 mM and 137 mM NaCl. The stronger association in 50 mM NaCl is in agreement with the results for SCR-16/20 (Figures 6.6(b) and 6.7). However the complexity of the changes seen in the  $c(s)$  plots do not permit a simple interpretation. This aspect was clarified by SPR studies of C3d binding to immobilised FH, in which  $K_D$  fits could not be achieved with a 1:1 binding model for FH-C3d. As the C3d concentration increased, the amount of C3d bound to FH in both 50 mM NaCl and 137 mM NaCl increased beyond that expected from a 1:1 interaction (Figure 6.8(b,c)). This indicated a multimeric interaction. The strength of this binding was again weaker in 137 mM NaCl compared to 50 mM NaCl.

The SCR-16/20 self-dimerisation site is independent of the SCR-16/20 binding site for C3d. Size-exclusion chromatography showed that the monomer and dimer peaks for SCR-16/20 were still observed, even in the presence of C3d (Figure 6.2(b)). AUC showed that the relative sizes of the double  $c(s)$  peaks for the monomer and dimer of SCR-16/20 were unchanged in the presence of C3d (Figure 6.5). By SPR, the binding of C3d and SCR-16/20 to immobilised FH are seen to occur as independent events (Figure 6.10(a)). The self-association of FH to form dimers and higher oligomers is likewise not significantly perturbed by C3d binding. By AUC,

the  $c(s)$  peaks for these FH oligomers were still mostly observed in the presence of C3d, even though their positions and intensities altered (Figure 6.6). By SPR, FH is seen to self-associate with immobilised FH, and the addition of increasing amounts of C3d did not eliminate this (Figure 6.11).

The observation of FH-C3d multimers in both the solution and surface phases using physiologically-relevant concentrations and buffers adds another layer of complexity to the mechanism of complement activation and control. Unbound FH has often been considered to be a monomer in plasma (Sim & DiScipio, 1982; DiScipio, 1992). Dimeric forms of FH have however been identified (Jokiranta *et al.*, 2000, 2006; Oppermann *et al.*, 2006; Perkins *et al.*, 1991). Weak dimerisation sites have been found in each of the SCR-6/8 and SCR-16/20 fragments but not in the SCR-1/5 fragment (Fernando *et al.*, 2007; Chapter 4) and this observation was confirmed here using FH, SCR-1/5 and SCR-16/20 by SPR experiments (Figure 6.10). The presence of multiple self-association sites explains the ability of FH to form a series of higher oligomers at concentrations above 1 mg/ml (Nan *et al.*, 2008a; Perkins *et al.*, 1991). Unbound C3 itself is also usually considered to be monomeric, with the exception of its disulphide-linked dimers (Perkins & Sim, 1986). In 50 mM NaCl, unbound C3d exhibits monomer, dimer and trimer forms above 2.8 mg/ml (80  $\mu$ M), while in 137 mM NaCl, only the C3d monomer is observed (Gilbert *et al.*, 2005; Li *et al.*, 2008). The  $K_D$  values of 16  $\mu$ M for SCR-16/20 dimerisation and 360  $\mu$ M for C3d dimerisation show that FH dimerisation takes priority over C3d dimerisation in 50 mM NaCl (Chapter 4; Li *et al.*, 2008). These multimers will complicate structural studies of complexes between SCR-20 and C3d by NMR or crystallographic approaches, as these methods require high protein concentrations.

The observations for SCR-16/20 in 50 mM NaCl are most simply explained by the formation of an indefinite series of complexes of alternating dimers of C3d and SCR-16/20 as suggested from the AUC and scattering modelling (Figure 6.14). The interaction sites that lead to these complexes are independent of each other. These multimers are formed more weakly in 137 mM NaCl with possible stoichiometries of 1:1, 2:1 or 2:2, because C3d does not form oligomers in this buffer. The observations of FH and C3d multimers in 50 mM NaCl are likewise most simply explained by the existence of FH oligomers, where each FH monomer is able to bind at least one C3d monomer or C3d dimer in addition to the formation of FH

oligomers. In 137 mM NaCl, these form more weakly. Three previous SPR studies did not report multimeric species of FH and C3d complexes (Hellwage *et al.*, 2002; Manuelian *et al.*, 2003; Jokiranta *et al.*, 2005). This is attributable to (i) the use of low FH concentrations of either 2-3 µg/ml, 100-125 µg/ml or 100-150 µg/ml, that are much less than the physiological range of 235-810 µg/ml, and (ii) the immobilisation of C3d on the chip. Both factors would reduce or eliminate FH or C3d dimerisation. Interestingly, the formation of C3d dimer and trimer in 50 mM NaCl is inhibited by the complex formed with SCR-1/2 of CR2 (Gilbert *et al.*, 2005; Li *et al.*, 2008). This indicates that the C3d surfaces that interact with FH and CR2 are distinct.

The ionic strength dependence of the SCR-16/20-C3d and FH-C3d interaction is explained by the overall negative charge of C3d and the overall positive charge of SCR-20. C3d binding to CR2 also shows a strong ionic strength dependence (Li *et al.*, 2008). Radioactivity assays in 150 mM NaCl indicated that dimeric or oligomeric C3b may form multimeric interactions with CR1, and these interactions were found to be ionic strength dependent (Arnaout *et al.*, 1983). Hence the ionic strength dependence of the interactions between C3 and the complement regulatory proteins appears to be a general phenomenon. Other FH-C3d studies also report similar observations, but did not compare the use of 50-75 mM NaCl buffers with 137 mM NaCl buffers. SPR showed that immobilised C3d did not bind in 75 mM NaCl to a recombinant SCR-15/20 fragment with the SCR-20 substitutions R1203E, R1206E, R1210S, K1230S and R1231A in 75 mM NaCl (Hellwage *et al.*, 2002; Table 6.2). Similarly, an SCR-8/20 fragment with two aHUS mutations R1210C or R1215G in SCR-20 showed reduced binding to immobilised C3d (Manuelian *et al.*, 2003; Table 6.2). The crystal structures for SCR-19/20 and C3d have been used to predict ion-pairs that included residues E1032, R1042, E1047 and K1264 in C3d and residues R1182, K1186, K1188, E1198 and R1215 at the tip of SCR-20 (Jokiranta *et al.*, 2006). Solution studies that utilise low salt buffer (50 mM NaCl) and high protein concentrations in equimolar conditions (ca. 1 mg/ml or 25 µM) amplify the molecular interactions seen between FH and C3d. This accounts for the use of 50-75 mM NaCl buffers in previous SPR studies (Hellwage *et al.*, 2002; Manuelian *et al.*, 2003; Jokiranta *et al.*, 2005; Table 6.3). However observations of binding using both AUC in the solution phase and SPR at a surface in physiological 137 mM NaCl buffer in this thesis confirm the biological relevance of these interactions. The observation of multimeric complexes between 1 mg/ml native FH

with 0.3 mg/ml C3d in 137 mM NaCl PBS buffer occurs in conditions close to physiological (Figure 6.6(b)).

The multimeric SCR-16/20 and C3d interaction provides a fresh insight into the mechanism of C3b regulation by FH. Predominantly monomeric FH together with 5-15% of dimeric and oligomeric FH circulate in plasma (Nan *et al.*, 2008a). FH recognizes the anionic cell surface glycans (glycosaminoglycans and sialic acids) decorating all host cells which present heparin-like structures, but not pathogenic bacteria on which these are absent. This initial contact of FH with host cells is via SCR-20 (Figure 6.16(a)), followed by N-terminal regulatory activity, and this mechanism is the main focus of the recent literature (Oppermann *et al.*, 2006; Ferreira *et al.*, 2006). The steady tick-over activation of C3 to C3b means that C3b activation requires continuous control in the host, as C3b can bind indiscriminately to any cell surface and could, if unregulated, trigger a rapid and uncontrolled inflammatory response. The degradation of cell surface-bound C3b to C3c and C3dg by FH and factor I will leave the host cell surface decorated by C3d covalently linked through its thioester (Figure 6.16(b)). Surface-bound C3d on host cells will present a FH binding site, and this FH-C3d binding will complement the binding of FH to heparin analogues on host cells (Figure 6.16(c)), but not on microbial pathogens. Weak FH dimerisation will potentially maximise FH localisation and complement regulation on host cells.

Furthermore, under inflammatory conditions where high rates of complement activation occur, the observation of FH-C3d multimers offers a mechanism for a further increase in host cell protection, as increased amounts of fluid-phase or surface-bound C3d will lead to the formation of higher multimers of FH-C3d complexes on host cell surfaces (Figure 6.6(b)). The binding of these multimeric FH-C3d complexes to host cells will afford a greater level of protection to the host through the increased concentration of the regulatory FH N-terminal SCR domains at the host cell surface (Figure 6.16(d)). These observations provide further insight into the pathogenic mechanisms underlying aHUS. aHUS mutations occur mostly between SCR-16 and SCR-20 and generally lead to dysfunctional FH molecules (Saunders *et al.*, 2007). Mutations in SCR-20 that affect either C3d or heparin binding or FH dimerisation will directly lead to the impaired protection of host cells. The occurrence of further mutations between SCR-16 to SCR-19 may correspond to

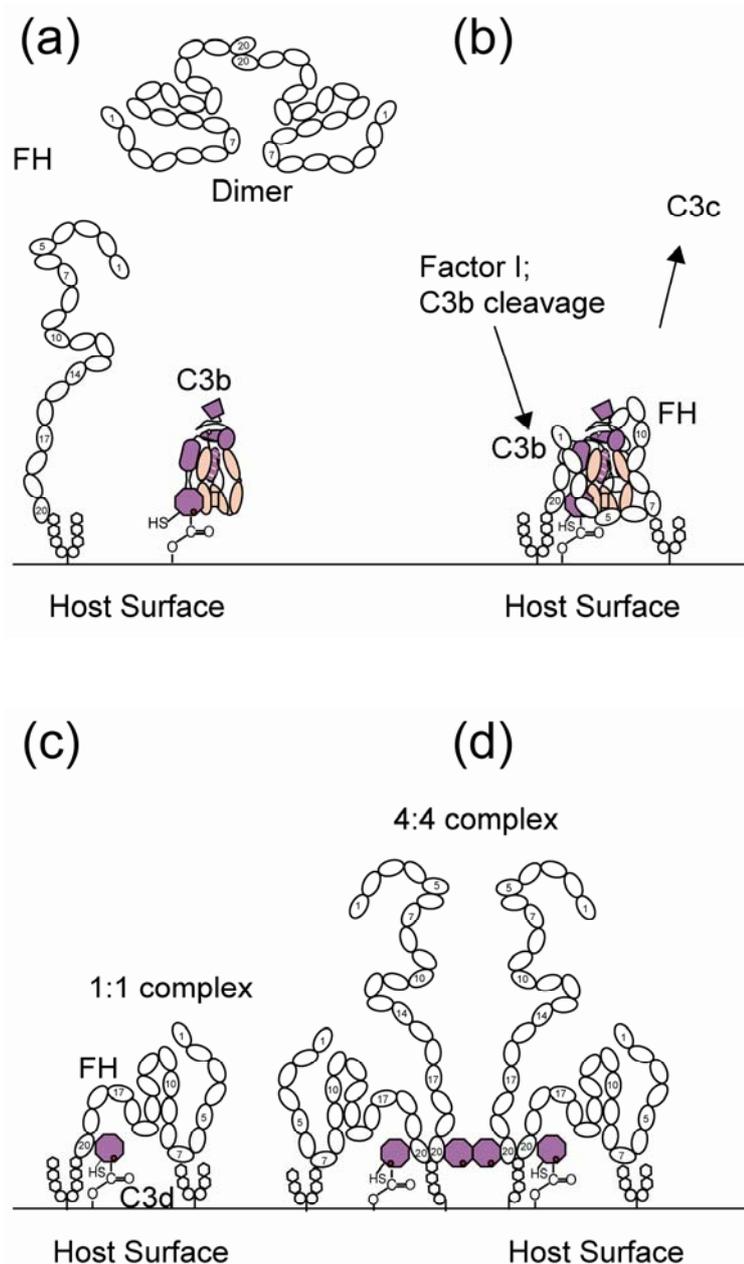
the need for these residues to orientate correctly FH on the host cell surface in order to bind to surface-bound C3b in order to degrade this.

## **(6.4) Materials and Methods**

### **(6.4.1) Protein expression and purification**

SCR-16/20 and SCR-1/5 were expressed as hexaHis-tagged products in *Pichia pastoris* and purified using nickel affinity chromatography (Chapter 4). Native FH was purified from just-outdated pooled plasma using MRC-OX23 affinity chromatography and used while freshly purified to minimise the occurrence of FH aggregates (Sim *et al.*, 1993; Nan *et al.*, 2008a; 2008b; Chapter 5). Human C3d (sequence 994-AVDAE... to ...QLPSR-1303 of the mature protein C3 with SWISSPROT accession code P01024) containing the free cysteine of the thioester was expressed in *Escherichia coli* BL21 transformed with the PGEX-2T vector, as provided by Prof. V. M. Holers and Dr. J. P. Hannan of the University of Colorado School of Medicine, Denver, Colorado, USA. Expression and purification were carried out in conjunction with Miss Keying Li (Li *et al.*, 2008) as described below.

Ampicillin-resistant colonies were used to produce starter cultures, which were then expanded to five litres and grown at 37°C until an absorbance at 600 nm of 0.3 was attained, then cooled to 23°C and incubated to an absorbance of 0.4. Samples were then induced with 0.5 mM IPTG and shaken overnight at 25°C. The culture was lysed in the presence of protease inhibitors by sonication, and after clarification applied to a GSTrap FF column equilibrated in 25 mM Tris, 0.5 M NaCl, 1 mM EDTA, 1 mM DTT, pH 8.0. After washing to remove unbound protein, thrombin was added to the column. After an overnight incubation at 4°C, the thrombin was inactivated and the cleaved C3d was eluted by further washing. Reduced glutathione was added to the buffer in further washes to elute GST and uncleaved GST-C3d. The C3d was concentrated using Amicon® Ultra-15 centrifugal filter devices with a molecular mass cut-off of 10 kDa at 4500 g, and size exclusion chromatography using a Superdex-200 Prep Grade XK 16/60 column resulted in the C3d samples used for data acquisition. Briefly, the C3d-GST cell supernatant was passed over a GSTrap FF column, followed by cleavage using thrombin and size-exclusion chromatography using a Superdex 200 prep grade XK 16/60 column (Figure 6.2). The purified proteins were stable upon storage at 4°C. Experiments were performed using either 10 mM HEPES



**Figure 6.16**

Schematic cartoons of FH multimer binding to cell surfaces in the presence of surface-bound C3d and heparin-like analogues. **(a)** Monomeric FH is bound via SCR-20 via anionic oligosaccharides to the host cell, together with C3b covalently bound through its thioester to the host cell. Small amounts of dimeric FH are also present. **(b)** FH degrades surface-bound C3b through factor I cleavage to leave C3d bound to the host cell surface. SCR-7 also binds to anionic oligosaccharides. **(c)** FH forms a 1:1 complex with host cell-bound C3d. **(d)** A schematic multimeric 4:4 complex of FH:C3d illustrates how high amounts of soluble and cell-bound C3d will lead to FH multimerisation on the host cell surface, thereby enhancing the protection of the host cell.

**Table 6.2. Electrostatics of FH-C3d interactions**

FH protein	Ligand	Conditions	Effect	Reference/ Method
SCR-15/20	C3d	One mutant protein (all residues on SCR-20): R1203E, R1206E, R1210S*, K1230S, R1231A	Mutant showed much reduced binding to C3d. While $K_D$ of wild-type binding was 0.09 $\mu$ M, it was impossible to calculate rate constants for the mutant under the same conditions because binding was too weak to obtain reliable results. Half-life of the complex was reduced to 34.2% of that for wild-type complex.	Hellwage <i>et al.</i> , 2002/ SPR
SCR-8/20	C3d	Two mutant proteins (all residues on SCR-20): <b>1)</b> R1210C*, <b>2)</b> R1215G*	Weak binding of mutants to C3d.	Manuelian <i>et al.</i> , 2003/ SPR

\* Mutation often found in familial HUS.

**Table 6.3. Ionic strength conditions used for C3 binding studies**

C3 protein	Ligand	Buffer	Experimental K <sub>D</sub>	Reference
Low NaCl (~50 mM)				
C3b	CR1	5 mM Veronal, 50 mM NaCl		Arnaout <i>et al.</i> , 1983*
C3b	FH	2.5 mM Veronal, 70 mM NaCl		Sharma & Pangburn, 1996
C3b	FHR-3, FHR-4	1.7 mM Veronal, 50 mM NaCl	0.13-0.34 μM	Hellwage <i>et al.</i> , 1999
C3d	FH, FHR-3, FHR-4	1.7 mM Veronal, 50 mM NaCl	0.09-0.36 μM	Hellwage <i>et al.</i> , 1999
C3b	FH	1.7 mM Veronal, 50 mM NaCl		Jokiranta <i>et al.</i> , 2000
C3b, C3c, C3d	Factor B, FH, C5, sCR1	1.7 mM Veronal, 50 mM NaCl		Jokiranta <i>et al.</i> , 2001
C3b, C3d	FH, SCR-15/20	5 mM Veronal, 75 mM NaCl	0.09 μM***	Hellwage <i>et al.</i> , 2002
C3d	FH	7.5 mM Veronal, 68 mM NaCl		Jokiranta <i>et al.</i> , 2005
C3b	FH	2.5 mM Veronal, 70 mM NaCl		Ferreira <i>et al.</i> , 2006
C3d	CR2	10 mM HEPES, 50 mM NaCl	< 0.28 μM	Li <i>et al.</i> , 2008
C3d	FH	10 mM HEPES, 50 mM NaCl	1.0 μM***	Present study
High NaCl (> 100 mM)				
C3b	CR1	5 mM Veronal, 150 mM NaCl		Arnaout <i>et al.</i> , 1983*
C3b	FH	10 mM Phosphate, 137 mM NaCl		Lambris <i>et al.</i> , 1988
C3d <sub>1-3</sub> -HEL**	CR2	10 mM Phosphate, 137 mM NaCl		Dempsey <i>et al.</i> , 1996
C3b	FH	10 mM Phosphate, 137 mM NaCl		Manuelian <i>et al.</i> , 2003
C3b, C3dg	IgG	4.6 mM Veronal, 150 mM NaCl		Jelesarova <i>et al.</i> , 2003
C3b	FH	10 mM Phosphate, 137 mM NaCl		Opermann <i>et al.</i> , 2006
C3d	CR2	10 mM Phosphate, 137 mM NaCl	> 6.9 μM	Li <i>et al.</i> , 2008
C3d	FH	10 mM Phosphate, 137 mM NaCl	2.6 μM***	Present study

\* Used an ionic strength titration series between high and low NaCl conditions; \*\* HEL – Hen egg lysozyme; C3d<sub>1-3</sub> – C3d monomers to trimers; \*\*\* For C3d binding to the specified fragment of FH.

50-137 mM NaCl, pH 7.4 or Dulbecco's phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.3). Compositional data for the data analyses are reported elsewhere (Perkins, 1986). SCR-16/20 is assumed to be glycosylated at either of two putative sites in SCR-17 and SCR-18, while C3d is unglycosylated (Figure 6.1). In summary, SCR-16/20 has a calculated molecular weight of 39.7 kDa, an unhydrated volume of 50.1 nm<sup>3</sup>, a hydrated volume of 66.3 nm<sup>3</sup>, a partial specific volume  $\bar{v}$  of 0.719 ml/g, and a 280 nm absorption coefficient (1%, 1 cm path length) of 15.7 (Chapter 4). Native FH has a calculated molecular weight of 150.0 kDa (assuming six N-linked oligosaccharide chains), an unhydrated volume of 188.4 nm<sup>3</sup>, a hydrated volume of 249.7 nm<sup>3</sup>, a partial specific volume  $\bar{v}$  of 0.717 ml/g, and a 280 nm absorption coefficient (1%, 1 cm path length) of 16.7 (Aslam & Perkins, 2001). While the recent mass spectroscopic study of FH revealed eight N-linked predominantly biantennary oligosaccharides, the compositional parameters are minimally affected by this (Fenaille *et al.*, 2007). C3d has a calculated molecular weight of 34.6 kDa, an unhydrated volume of 45.5 nm<sup>3</sup>, a hydrated volume of 59.6 nm<sup>3</sup>, a  $\bar{v}$  of 0.746 ml/g, and an absorption coefficient of 13.1. The HEPES and PBS buffer densities were measured as 1.00139 g/ml and 1.00534 g/ml at 20°C using an Anton-Paar DMA 5000 density meter.

## **(6.4.2) Analytical ultracentrifugation data and analysis**

### **(6.4.2.1) Sedimentation velocity data for C3d and its complexes with Factor H and SCR-16/20**

Sedimentation velocity data were acquired on two Beckman XL-I instruments with AnTi50 or AnTi60 rotors at 20°C at a rotor speed of 50,000 r.p.m. in two-sector cells with column heights of 12 mm. The SCR-16/20 and FH concentrations were held fixed at 1.0 mg/ml, into which C3d was added at molar ratios of 0:1, 0.3:1, 0.6:1, 1:1, 1.5:1, 2:1 and 3:1 in 10 mM HEPES buffer (50 mM NaCl), and a molar ratio of 1:1 in PBS buffer (137 mM NaCl). Analyses to determine sedimentation coefficients  $s_{20,w}^0$  were performed using direct boundary Lamm fits of up to 300 scans for size-distribution analyses  $c(s)$  in SEDFIT (version 11.0) that assumed that all species have the same frictional ratio  $f/fo$  (Schuck, 1998; 2000). The  $c(s)$  fits employed a resolution of 200 and were optimised by floating the frictional ratio and baseline until the overall root mean square deviations and visual appearance of the

fits were satisfactory (Figures 6.5 and 6.7). Other analyses utilised the  $c(s, f/fo)$  model to allow for variable frictional ratios and permit more accurate molecular weight determinations.

#### (6.4.2.2) Sedimentation equilibrium data for C3d

Sedimentation equilibrium experiments resulted in molecular weight and dissociation constant  $K_D$  determinations. C3d data were obtained in PBS and HEPES buffers at 20°C over 30 h at each speed using six-sector cells with column heights of 2 mm and rotor speeds of 19k r.p.m., 24k r.p.m., 28k r.p.m., 30k r.p.m. and 35k r.p.m. Experiments in PBS were performed by Miss Keying Li (Li *et al.*, 2008). The existence of equilibrium conditions after at least 10 h in PBS buffer and at the lowest concentrations and after at least 30 h at the highest concentrations in HEPES buffer was confirmed using overlays of runs measured at 3 h intervals in SEDFIT. In PBS buffer, nine concentrations between 0.04 mg/ml to 7.01 mg/ml were used, while in HEPES buffer, nine concentrations between 0.11 mg/ml to 5.50 mg/ml were used. Data analyses utilised SEDPHAT version 5.01 (Schuck, 2003). First, multispeed fits assuming a single species were performed individually. Next, the best interference optics fits from 19k r.p.m., 28k r.p.m., and 35k r.p.m. were used for global fits using the monomer-‘m-mer’-‘n-mer self-association model with  $m = 2$  (dimer) and  $n = 3$  (trimer) in SEDPHAT (Figure 6.4). The global fits utilised a common baseline, meniscus, cell bottom, extinction coefficient, and loading concentration. Fits used the M and s, simplex and Marquardt-Levenberg fitting routines, while floating the baseline and the cell bottom of the cell. Statistical error analyses employed the Monte-Carlo approach.

#### (6.4.3) Surface plasmon resonance data for Factor H and C3d

The interactions between FH, SCR-16/20, SCR-1/5 and C3d were analysed by SPR using Biacore X100 instrument and version 1.1 of its evaluation software (GE Healthcare, Uppsala, Sweden). Native FH was coupled to the flowcell of a carboxylated dextran (CM5) research grade sensor chip via a standard amine coupling procedure according to the manufacturer’s protocol. 10 µg/ml of FH in 10 mM acetate buffer (pH 5.5) was injected over flowcell 2 until the appropriate level of response units for kinetic analyses (~ 700 response units (RU)) was attained. A control surface cell was prepared identically on flow cell 1 but without immobilising

FH. Binding and steady-state analysis with C3d, FH, SCR-16/20 and SCR-1/5 were performed at 25° C using the appropriate Biacore X100 wizards at flow rates of 10-30  $\mu\text{l}/\text{min}$ . Regeneration after each run was achieved by pulsing 10 mM acetate buffer, 2 M NaCl (pH 4.6) across both flow cells twice for 30 sec each.

For binding analyses, the appropriate concentration of each analyte was passed over the chip surface in 50 mM NaCl or 137 mM NaCl HEPES buffers. For steady-state analysis of C3d binding to FH, 0.01-50  $\mu\text{M}$  (0.03-1.7 mg/ml) C3d was passed over the chip surface in 50 mM NaCl or 137 mM NaCl HEPES buffer respectively. For binding assays of C3d-FH ternary complex formation in 137 mM NaCl HEPES buffer, increasing concentrations of C3d were added to 0.5 or 5  $\mu\text{M}$  (0.08 or 0.8 mg/ml) FH in ratios of 0:1, 0.2:1, 0.4:1, 1:1, 2:1, 4:1 and 10:1 before injection. Control experiments using 5  $\mu\text{M}$  (0.2 mg/ml) SCR-1/5 were run at ratios of 0:1, 0.2:1, 2:1 and 4:1.

#### **(6.4.4) X-ray scattering data collection and analysis**

X-ray scattering data were obtained in one beam session at the European Synchrotron Radiation Facility, Grenoble, France operating in sixteen bunch mode with a ring energy of 6.0 GeV on the Beamline ID02 (Narayanan *et al.*, 2001). Storage ring currents ranged from 66 mA to 76 mA. Data were acquired using an improved fibre optically-coupled high sensitivity and dynamic range CCD detector (FReLoN) with a smaller beamstop, which allowed a shorter sample to detector distance of 2.0 m to result in a better signal-to-noise ratio. Sample concentrations were the same as those used in AUC and employed 50 mM NaCl HEPES or 137 mM NaCl PBS buffer. All samples were measured in a flow cell which moved the sample continuously during beam exposure in 10 time frames of duration 0.1 sec each to confirm the absence of radiation damage. This exposure was optimised using on-line checks for the absence of radiation damage, after which the frames were averaged. Other details including the data analysis procedure to determine the radius of gyration  $R_G$  from Guinier plots and the distance distribution function  $P(r)$  are described elsewhere (Gilbert *et al.*, 2005). The X-ray curves contained 541-544 data points in the Q range between 0.09  $\text{nm}^{-1}$  and 1.89  $\text{nm}^{-1}$ .

#### (6.4.5) Molecular modelling of the SCR-16/20:C3d complexes

Modelling was based on the crystal structure for C3d and the scattering solution structure for the SCR-16/20 dimer as constraints (Nagar *et al.*, 1998; Chapter 4). To model the complex, C3d was positioned adjacent to SCR-20 in an arbitrary location that did not lead to steric clashes with other SCR domains. Further symmetry-based manipulations generated compact yet planar models. The calculation of scattering curves for comparison with the experimental X-ray curves used hydrated spheres (Gilbert *et al.*, 2005). Starting from each model, cube side lengths of 0.543 nm (1:1 complex), 0.537 nm (2:2 complex), 0.545 nm (4:4 complex), 0.537 nm (6:6 complex) and 0.541 nm (8:8 complex) in combination with a cutoff of 4 atoms produced 579-, 1192-, 2286-, 3552-, and 4667-sphere models respectively, each within 95% of the total unhydrated volume calculated from its composition. The hydration shell was added using HYPRO (Ashton *et al.*, 1997), and the models were confirmed to be within 95% of the optimal total of hydrated spheres of 786 (1:1 complex), 1626 (2:2 complex), 3111 (4:4 complex), 4879 (6:6 complex) and 6362 (8:8 complex). Each X-ray scattering curve  $I(Q)$  was calculated using the Debye equation adapted to spheres (Perkins & Weiss, 1983; Perkins *et al.*, 2008). The final calculated curve was summed from a linear combination of the individual scattering curve for each molecular species, weighted in the proportions derived from the SEDFIT  $c(s)$  analysis. From this, the  $R_G$  value was calculated in the same  $Q$  range used for the experimental Guinier  $R_G$  fits in order to allow for any approximations inherent in this procedure. The goodness of fit between the calculated and experimental curves was assessed using the  $R$  factor, defined by analogy with protein crystallography from the experimental curve in the  $Q$  range extending to  $1.8 \text{ nm}^{-1}$ . The  $s_{20,w}^0$  values were calculated from the X-ray hydrated sphere models using the HYDRO program (Garcia de la Torre *et al.*, 1994; Ashton *et al.*, 1997), and also from the crystal coordinates using the HYDROPRO shell modelling program using the default value of 0.31 nm for the atomic element radius for all atoms to represent the hydration shell (Garcia de la Torre *et al.*, 2000).

## **Chapter Seven**

**C-reactive protein exists in an NaCl-concentration dependent pentamer-decamer equilibrium in physiological buffer containing calcium**

## (7.1) Introduction

C-reactive protein (CRP) is a 115 kDa acute-phase protein of the pentraxin family of calcium-dependent ligand-binding proteins in serum (Pepys & Hirschfield, 2003; Section 2.3.3.1).

Native CRP is unglycosylated and consists of five identical subunits arranged non-covalently as a cyclic pentamer with a central pore. When co-crystallised with  $\text{Ca}^{2+}$  and phosphorycholine, each subunit has a phosphocholine ligand-binding site proximate to two bound  $\text{Ca}^{2+}$  ions held 0.4 nm apart by acidic side chains on the B (binding) face of the pentamer (Thompson *et al.*, 1999; Pepys & Hirschfield, 2003). The other A (alpha) face of the protomer possesses the single  $\alpha$ -helix of CRP and binds to ligands such as C1q (Gaboriaud *et al.*, 2004). The B face of the pentamer is negatively charged, while the A face showed a more varied charge distribution with a ring of negative charge at the central pore (Thompson *et al.*, 1999; Ramadan *et al.*, 2002). This suggests a role for electrostatic effects in both lipid binding at the B face and C1q binding at the A face (Gaboriaud *et al.*, 2004). Native CRP with  $\text{Ca}^{2+}$  present exists physiologically as a pentamer according to electron microscopy, gel chromatography and crystallography (Osmand *et al.*, 1977; Volanakis *et al.*, 1978; Shrive *et al.*, 1996; Thompson *et al.*, 1999). However, in the presence of  $\text{Ca}^{2+}$ , CRP shows a marked concentration dependence in its scattering and sedimentation properties, showing that self-association had occurred (Perkins & Pepys, 1986; Blizniukov *et al.*, 2003).  $\text{Ca}^{2+}$  depletion leads to monomeric CRP (Motie *et al.*, 1996; Wu *et al.*, 2002). In contradiction to this, crystallographic studies in 100 mM NaCl in the absence of calcium suggested that CRP may be a decamer formed by the A face-A face association of two pentamers (Ramadan *et al.*, 2002). While the unfolding of a large  $\text{Ca}^{2+}$ -binding loop on the B face after  $\text{Ca}^{2+}$  depletion may explain the  $\text{Ca}^{2+}$ -dependent affinity of CRP for lipids, this  $\text{Ca}^{2+}$  site is neither located at the protomer interface nor at the A face. Even though  $\text{Ca}^{2+}$  is present in at least 2 mM in plasma (Hurwitz, 1996), there is much emphasis in the structural literature on the effect of  $\text{Ca}^{2+}$  depletion on CRP, even though  $\text{Ca}^{2+}$  depletion is highly unlikely to occur *in vivo*.

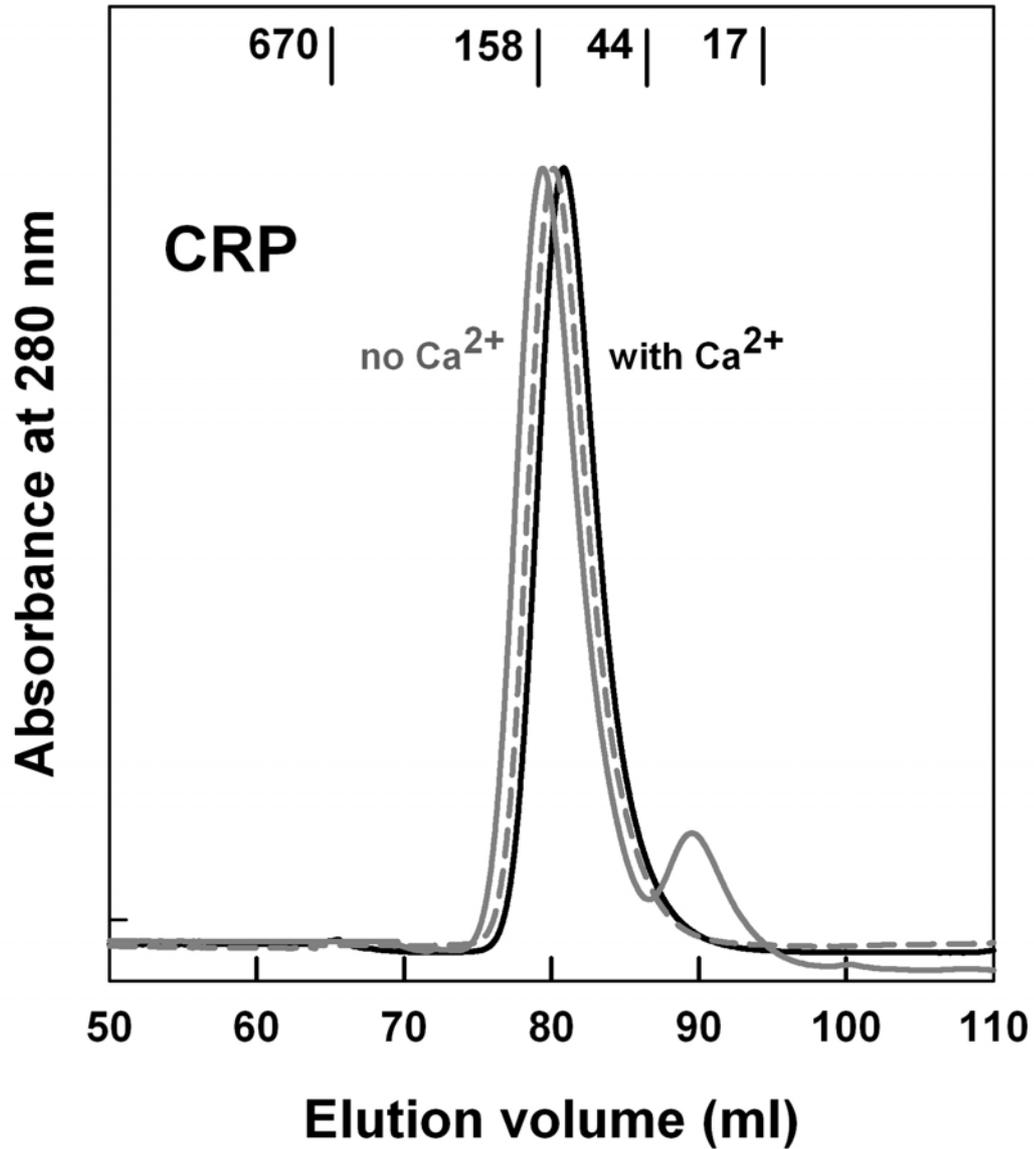
In this thesis, a multidisciplinary strategy based on AUC, X-ray scattering, and SPR established the association state of CRP at concentrations of 0.1-5.4 mg/ml in 2 mM  $\text{Ca}^{2+}$ . CRP normally circulates in blood at 0.0008 mg/ml, but this increases to over 0.5 mg/ml following an acute phase stimulus (Pepys & Hirshfield 2003).

Size-distribution analyses  $c(s)$  from AUC identified pentamers and decamers of CRP, while constrained scattering modelling based on crystal structures quantified the amounts of pentamers and decamers (Cole *et al.*, 2008; Perkins *et al.*, 2008). It is shown here that SPR is able to detect similar degrees of CRP self-association (Tanious *et al.*, 2008). In distinction to previous size-exclusion chromatography studies, the improved resolution of AUC, scattering and SPR shows that under physiological conditions CRP with  $\text{Ca}^{2+}$  present exists in a rapid NaCl-concentration dependent pentamer-decamer equilibrium. Molecular modelling accounted for the formation of decamers. In order to resolve discrepancies in previous CRP functional studies without  $\text{Ca}^{2+}$ , it is shown that this artefactual removal of  $\text{Ca}^{2+}$  leads to the reduced formation of decamers and the dissociation of pentamers into monomers. These studies clarify the molecular interaction of CRP with its ligands, and emphasize the importance of working with CRP in buffers containing  $\text{Ca}^{2+}$ .

## **(7.2) Results and Discussion**

### **(7.2.1) Size exclusion chromatography of CRP**

Size exclusion chromatography measures the effective hydrodynamic volume of a protein (Mori & Barth, 1999). CRP may self-associate into decamers (Ramadan *et al.*, 2002) and/or dissociate into protomers when stored in the absence of  $\text{Ca}^{2+}$  (Motie *et al.*, 1996; Wu *et al.*, 2002). These observations were investigated by size exclusion chromatography of CRP in 140 mM NaCl buffer. CRP stored at 1.0 mg/ml at 4°C in the presence of 2 mM  $\text{Ca}^{2+}$  eluted as a single clean peak in both the presence and absence of  $\text{Ca}^{2+}$  (Figure 7.1). Molecular weight standards indicated that this peak corresponded to an apparent molecular weight of 80-94 kDa, and was therefore assigned to the pentamer. However, when CRP was stored at 4°C in a buffer lacking  $\text{Ca}^{2+}$  for 12 weeks, a second peak was observed at a later elution volume (Figure 7.1), in agreement with earlier observations (Wu *et al.*, 2002). This second peak had an apparent molecular weight of 23 kDa and corresponded to approximately 10% of the total eluted protein. This indicated that partial dissociation into CRP monomers or other lower molecular-weight forms had occurred slowly in the absence of  $\text{Ca}^{2+}$ . Small shifts in the elution position of CRP to earlier elution positions were observed with increased storage periods in  $\text{Ca}^{2+}$ -depleted buffer (Figure 7.1). This may correspond to small increases in the amounts of aggregated



**Figure 7.1.**

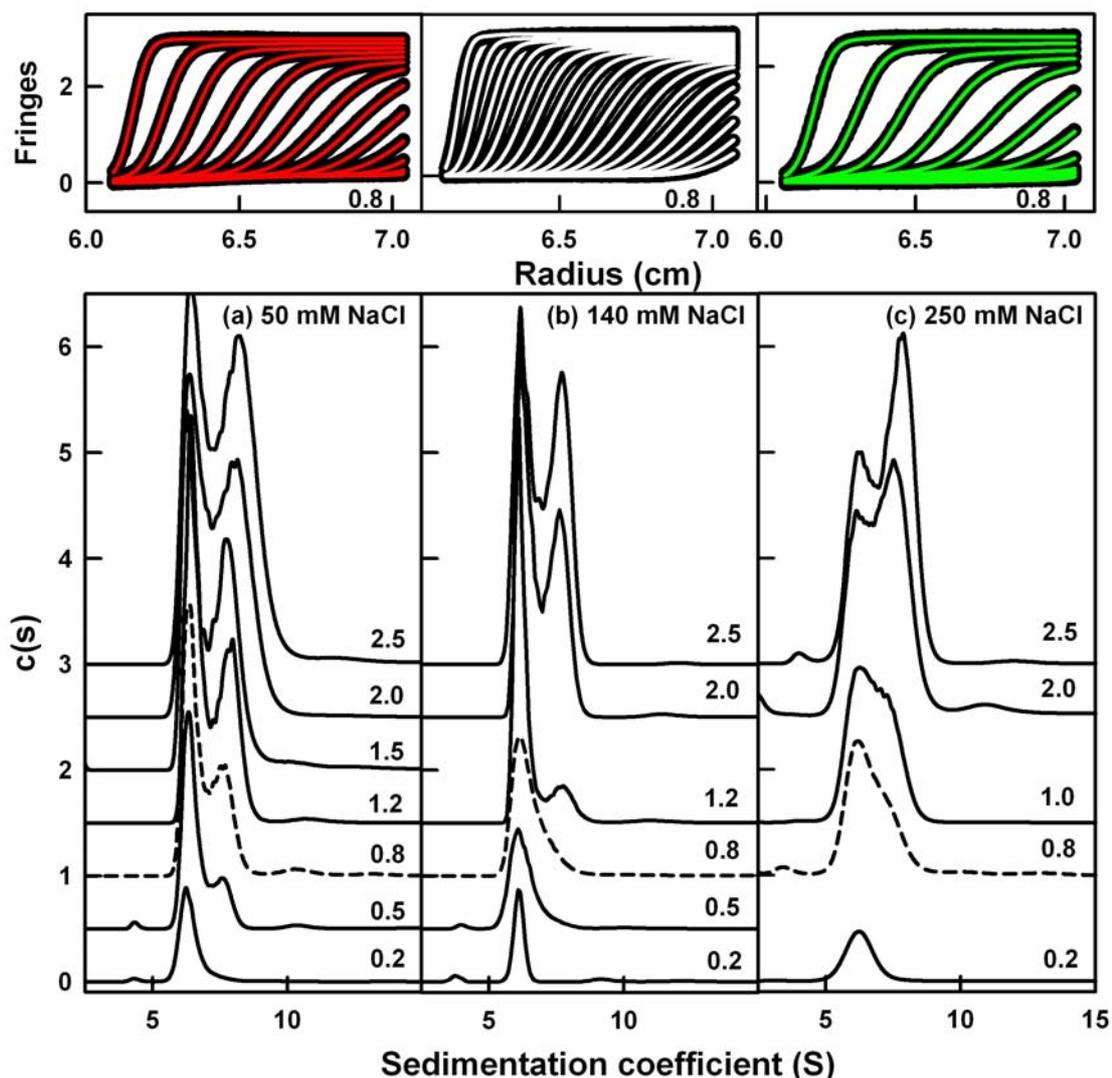
Size-exclusion chromatography of 1 mg/ml CRP in 140 mM NaCl buffers. The elution positions of four globular standards are shown in kDa. CRP stored for over 12 weeks in  $\text{Ca}^{2+}$ -containing buffer eluted as a single peak (solid black line). A fresh sample of CRP eluted with  $\text{Ca}^{2+}$ -depleted buffer eluted earlier (dashed grey line). CRP stored for 12 weeks in  $\text{Ca}^{2+}$ -depleted buffer (solid grey line) eluted earlier still, and showed an additional smaller peak at 90 ml.

forms of CRP.

### (7.2.2) Sedimentation velocity analyses of CRP

AUC studies macromolecular structures in solution by following their sedimentation behaviour in a high centrifugal force (Cole *et al.*, 2008). The much improved resolution of size distribution analyses  $c(s)$  from sedimentation velocity experiments performed on a Beckman Proteomelab XL-I instrument was used to monitor CRP. The observed sedimentation boundaries in sedimentation velocity experiments were fitted to yield a  $c(s)$  plot according to the Lamm equation using SEDFIT (Section 7.4.2). This takes into account all the species present in the sample, although the algorithm assumes that the same frictional ratio applies for all the sedimenting species. Macromolecular elongation is monitored through the sedimentation coefficient  $s_{20,w}^0$  values. The number of species is determined from the peaks observed in the  $c(s)$  plot. The effect of NaCl concentration on CRP was studied in the concentration range of 0.2 mg/ml to 2.5 mg/ml in 50 mM, 140 mM and 250 mM NaCl buffers, all containing 2 mM  $\text{Ca}^{2+}$ . The  $c(s)$  analyses of interference optics data using SEDFIT resulted in good sedimentation boundary fits (top of Figure 7.2). In all three buffers, a single species was observed at the lowest concentration of 0.2 mg/ml, while two species were observed above 1 mg/ml. The  $c(M)$  mass distribution plot showed a molecular weight of  $118 \pm 23$  kDa for the single species. This was accordingly assigned to pentameric CRP which has a sequence-calculated molecular weight of 115.0 kDa. The pentamer was consistently observed at  $s_{20,w}^0$  values of  $6.5 \pm 0.1$  S,  $6.4 \pm 0.2$  S and  $6.4 \pm 0.3$  S at all concentrations in 50 mM, 140 mM and 250 mM NaCl buffers respectively, indicating that the NaCl concentration does not affect the CRP conformation (Figure 7.2; Table 7.1). These values agree with the value of 6.6 S for pentameric CRP in 2 mM  $\text{Ca}^{2+}$  obtained using older Beckman Model E technology (Blizniukov *et al.*, 2003).

A pentamer-decamer equilibrium was identified from the  $c(s)$  plots. Above 0.2 mg/ml in 50 mM NaCl, 0.8 mg/ml in 140 mM NaCl and 1.0 mg/ml in 250 mM NaCl, the  $c(s)$  distributions showed the steady growth of a second species indicating a higher oligomer in equilibrium with the pentamer (Figure 7.2). This peak was attributed to the formation of decameric CRP. However the  $s_{20,w}^0$  value for this peak is variable in ranges from 7.7-8.4 S in 50 mM NaCl, 7.6-8.2 S in 140 mM NaCl and 7.5-7.7 S in 250 mM NaCl. The molecular weight of  $138 \pm 10$  kDa of this peak from  $c(M)$



**Figure 7.2**

NaCl concentration dependence of CRP decamer formation by AUC in the presence of  $\text{Ca}^{2+}$ . Protein concentrations ranged from 0.2 to 2.5 mg/ml, and are labelled at the right of each  $c(s)$  plot. The  $c(s)$  plots are displaced vertically and the peak intensities are normalised for clarity. The boundary fits corresponding to 0.8 mg/ml or 0.6 mg/ml are shown in the upper panels, where only every tenth scan is shown for reason of clarity. Experiments were performed at a rotor speed of 50,000 r.p.m. In Figures 7.2-7.9, data measured in 50 mM NaCl is denoted by red, 140 mM NaCl by black and 250 mM NaCl by green.

(a-c) Buffers containing 2 mM  $\text{Ca}^{2+}$  and 50 mM NaCl, 140 mM NaCl and 250 mM NaCl.

analyses is too low compared to the expected value of 230 kDa. The variation in  $s_{20,w}^o$  values and the low molecular weight indicate that this peak is not a second CRP species but is instead a reaction boundary corresponding to a rapid equilibrium between pentameric and decameric CRP on the time scale of sedimentation (Dam & Schuck, 2005). A  $s_{20,w}^o$  value of 9.5 S has been reported for cross-linked decameric CRP (Blizniukov *et al.*, 2003). Integration of the two  $c(s)$  peak areas resulted in an estimated dissociation constant  $K_D$  value of 19  $\mu\text{M}$  for the CRP pentamer-decamer equilibrium in 50 mM NaCl, 2 mM  $\text{Ca}^{2+}$  buffer (Figure 7.4(a)). Estimates of the  $K_D$  values in 140 mM NaCl and 250 mM NaCl buffers were not possible for reason of the poor resolution of the two peaks at lower concentrations.

The removal of  $\text{Ca}^{2+}$  at CRP concentrations of 0.1 mg/ml to 2.4 mg/ml in 50 mM and 140 mM NaCl buffers revealed the presence of smaller oligomers of CRP, together with pentamers and decamers by AUC. Good boundary fits were obtained (Figure 7.3). At all concentrations, pentamers were clearly observed to dominate the  $c(s)$  analyses with mean  $s_{20,w}^o$  values of  $6.4 \pm 0.1$  S in 50 mM NaCl and  $6.3 \pm 0.1$  S in 140 mM NaCl (Table 7.1). These values were unchanged from those obtained in the presence of  $\text{Ca}^{2+}$  (Table 7.1). The  $c(M)$  analyses yielded a mean molecular weight of  $119 \pm 26$  kDa in agreement with the assignment as pentamers. It was concluded that the loss of  $\text{Ca}^{2+}$  leads to only small effects in pentameric CRP, in agreement with size exclusion chromatography (Figure 7.1). Two further observations were made. Firstly, at all concentrations, one or two additional  $c(s)$  peaks were seen at  $2.8 \pm 0.2$  S and  $4.3 \pm 0.3$  S in 50 mM NaCl, and at  $3.0 \pm 0.2$  S and  $4.0 \pm 0.3$  S in 140 mM NaCl (Figure 7.3). The absence of a concentration dependence showed that these peaks were not in equilibrium with the pentamer. Since the  $c(M)$  analyses yielded mean molecular weights of  $37 \pm 5$  kDa and  $72 \pm 7$  kDa, these peaks are possible dimer (46 kDa) and trimer (69 kDa) forms of CRP. Thus in the absence of  $\text{Ca}^{2+}$ , CRP partially dissociates into lower molecular-weight oligomers. This is consistent with size exclusion chromatography results after storage in the absence of  $\text{Ca}^{2+}$  (Figure 7.1). Secondly, at 1.0 mg/ml and above, the CRP decamer was observable with  $s_{20,w}^o$  values ranging from 7.5-8.9 S and 7.3-7.7 S in 50 mM and 140 mM NaCl buffers respectively, and a mean apparent molecular weight of  $172 \pm 59$  kDa (Figure 7.3). Peak integrations showed that the apparent  $K_D$  for the pentamer-decamer equilibrium is much weakened in higher salt (Figure 7.4(b)). No  $K_D$  values could be estimated for reason of the presence of 2% to 33% of the lower molecular-weight oligomers. It is concluded that  $\text{Ca}^{2+}$  is essential to stabilise the CRP

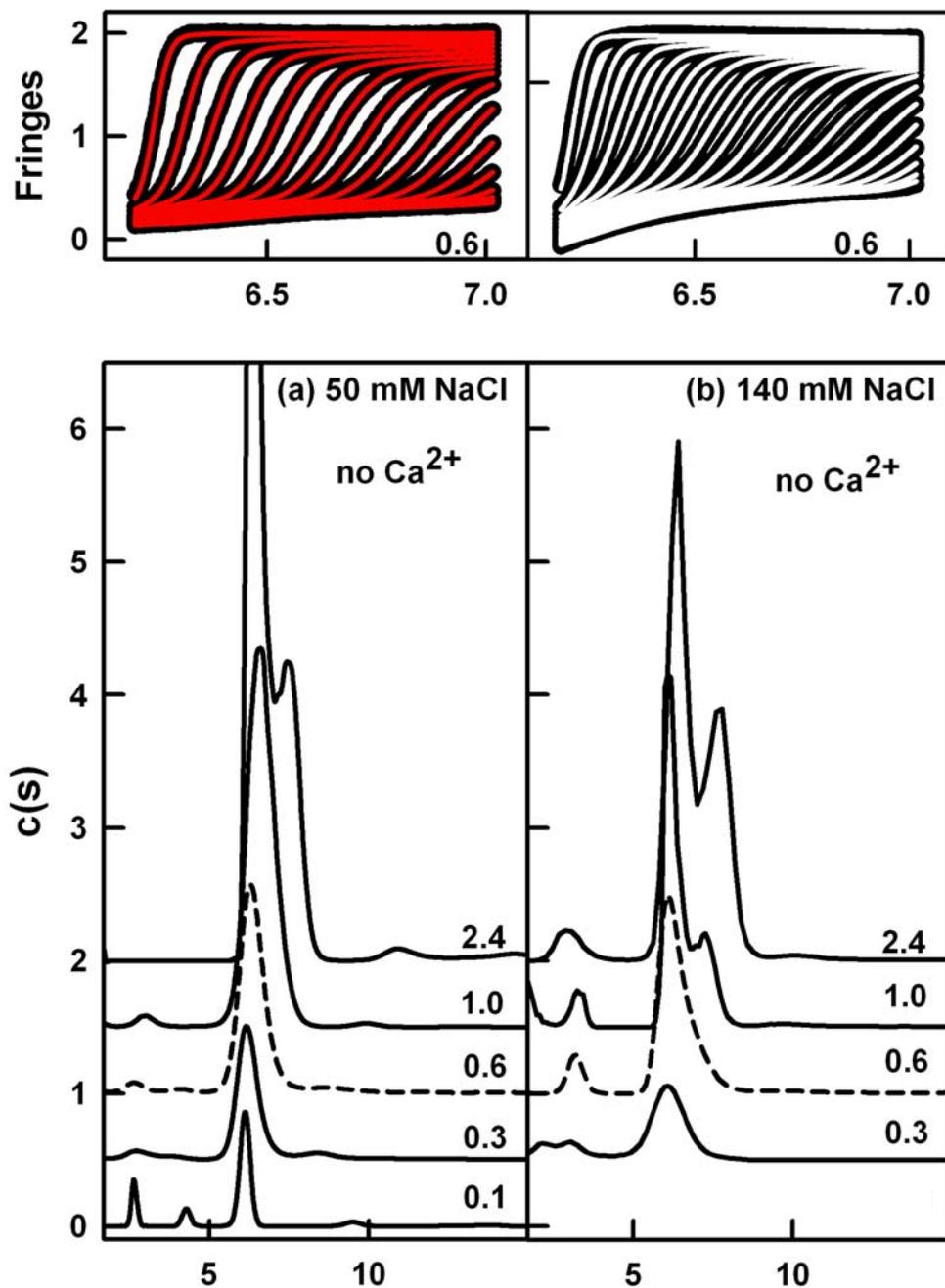
pentamer and the formation of decamers.

### (7.2.3) X-ray scattering analyses of CRP

X-ray scattering is a diffraction technique that studies the overall structure of biological macromolecules in random orientations in solution (Perkins *et al.*, 2008). This was used to observe the self-association and conformation of CRP in the presence and absence of  $\text{Ca}^{2+}$ . Previous X-ray experiments with 2-6 mg/ml CRP were performed on Station 7.3 at the Synchrotron Radiation Source at Daresbury, Cheshire, which utilised a linear detector with a limited  $Q$  range ( $Q = 4\pi \sin \theta / \lambda$ ;  $2\theta =$  scattering angle;  $\lambda =$  wavelength) (Perkins & Pepys, 1986). In comparison, Instrument ID02 at the European Synchrotron Radiation Facility in Grenoble, France enabled much improved experiments. ID02 was equipped with a flow cell to eliminate possible radiation effects and a two-dimensional detector with better sensitivities and larger  $Q$  range (Section 7.4.3). The improvements permitted a detailed study of a more extended CRP concentration range of 0.3-5.4 mg/ml.

The X-ray Guinier analysis confirmed CRP self-association in 140 mM NaCl buffer with 2 mM  $\text{Ca}^{2+}$ , and provided information on the decamer conformation. Guinier analyses of the scattering data  $I(Q)$  at low  $Q$  values yielded the forward scattered intensity  $I(0)$  at zero  $Q$  and the radius of gyration  $R_G$ . The  $I(0)/c$  value is proportional to molecular weight (where  $c$  is the CRP concentration in mg/ml) and  $R_G$  monitors the degree of overall macromolecular elongation. Guinier plots gave linear  $R_G$  fits within appropriate  $Q \cdot R_G$  limits (Figure 7.5(a)). The  $I(0)/c$  values increased from 14.2 to 21.9 units between CRP concentrations from 0.3 to 5.4 mg/ml (Figure 7.6(a)), in agreement with a self-association equilibrium. Assuming that this is a pentamer-decamer equilibrium, extrapolation of  $I(0)/c$  to zero concentration gave a value of 13.1 units for the pentamer, thus assigning 26.2 units for the decamer  $I(0)/c$  value, and leading to a  $K_D$  value of 26  $\mu\text{M}$  (3.0 mg/ml) from the mid-point. The concentration dependence of the  $R_G$  values ranged from  $3.8 \pm 0.1$  nm to  $4.5 \pm 0.2$  nm (Figure 7.6(c)). Extrapolation to zero concentration yielded an  $R_G$  value of 3.7 nm for the pentamer. The  $R_G$  values agreed with the previously-reported values of 4.1-4.5 nm for CRP concentrations of 2-6 mg/ml (Perkins & Pepys, 1986).

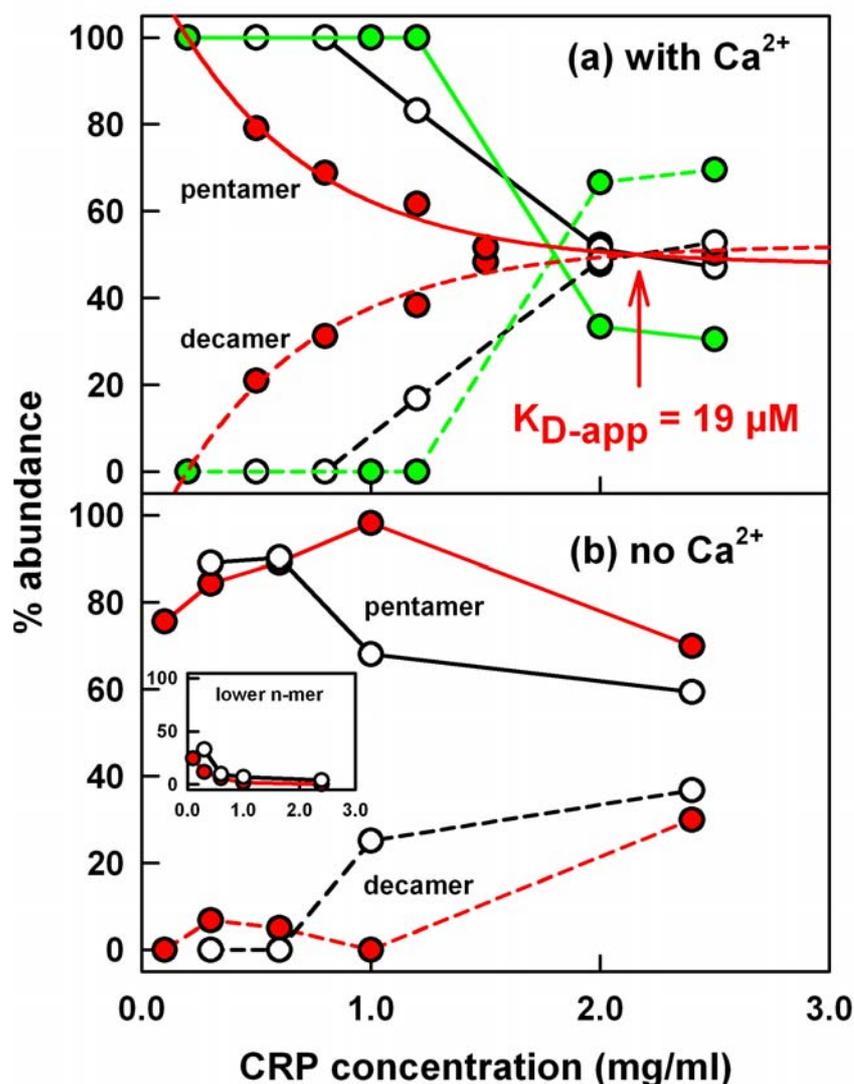
The distance distribution function  $P(r)$  in real space was calculated from the  $I(Q)$  data, following an assumption of the value of the maximum dimension  $D_{\text{max}}$ . The  $P(r)$  curve represents all the distances between pairs of atoms within CRP,



**Figure 7.3**

NaCl concentration dependence of CRP decamer formation by AUC in the absence of  $\text{Ca}^{2+}$ . Protein concentrations ranged from 0.1 to 2.4 mg/ml. All other details follow Figure 7.2.

**(a,b)** Buffers without  $\text{Ca}^{2+}$  and containing 50 mM NaCl and 140 mM NaCl.



**Figure 7.4.**

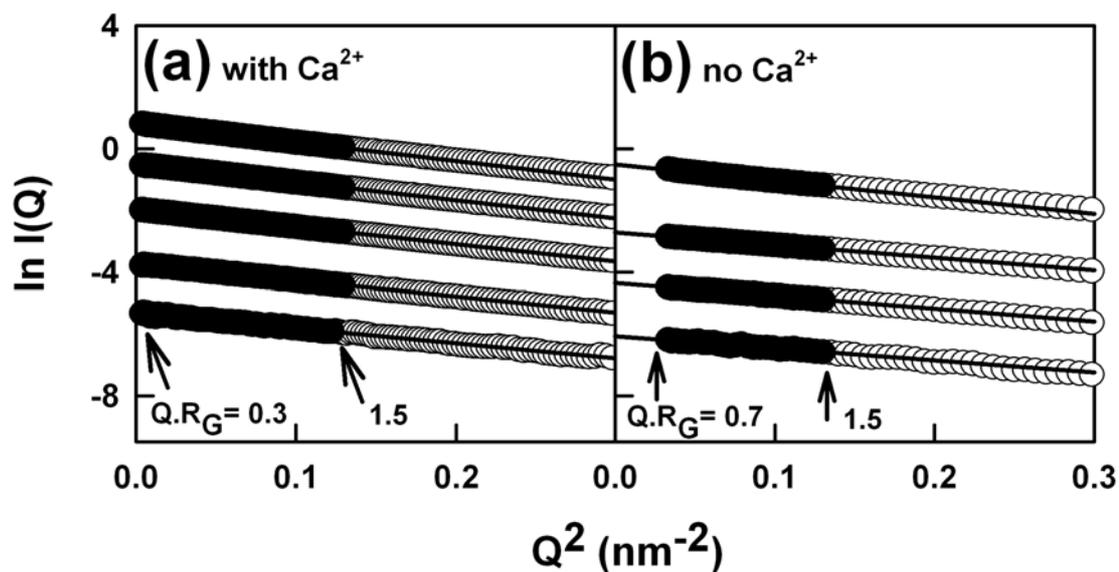
NaCl concentration and  $\text{Ca}^{2+}$  dependence of CRP pentamer-decamer equilibrium by AUC. At each CRP concentration, the relative percentage of pentamer (solid lines) and decamer (dashed lines) corresponds to data obtained at 50,000 r.p.m. The apparent  $K_D$  value ( $K_{D-app}$ ) in 50 mM NaCl is estimated as the point of intersection of the two data sets when fitted to a non-linear regression for an exponential rise (decamer:  $y = y_0 + a(1-b^x)$ ) or decay (pentamer:  $y = y_0 + a(10)^{-bx}$ ) as appropriate. The colours are defined in Figure 7.2.

(a) Data correspond to 50 mM NaCl, 140 mM NaCl and 250 mM NaCl in the presence of 2 mM  $\text{Ca}^{2+}$ .

(b) Data correspond to 50 mM NaCl and 140 mM NaCl in the absence of  $\text{Ca}^{2+}$ . The inset shows the percentage contribution of lower oligomers of CRP, which was not seen with  $\text{Ca}^{2+}$  present.

and provides an independent verification of the Guinier  $R_G$  values. The  $P(r)$  curves from Figure 7.7(a) yielded mean  $R_G$  values of  $3.8 \pm 0.1$  nm to  $4.3 \pm 0.1$  nm, in good agreement with the Guinier values. Two maxima  $M1$  at  $r = 3.1$  nm and  $M2$  at  $r = 5.7$  nm correspond to the most frequently occurring interatomic distances within CRP. Comparison with the pentamer and decamer crystal structures (PDB codes 1b09 and 1lj7) (Thompson *et al.*, 1999; Ramadan *et al.*, 2002) showed that  $M1$  and  $M2$  could be assigned to the CRP pentamer and decamer respectively. While the  $r$  values of  $M1$  and  $M2$  remained unchanged with concentration, the intensity of  $M2$  relative to  $M1$  increased with CRP concentration. This means that the proportion of decameric CRP increases with concentration, which is as expected from the pentamer-decamer equilibrium. The maximum dimension  $L$  of CRP is the  $r$  value when  $P(r)$  decreases to 0 at large  $r$ . The  $L$  values were unchanged with concentration at 11-13 nm (Figure 7.7(a)). The outermost ring diameter of the CRP pentamer is 11-13 nm by electron microscopy (Osmand *et al.*, 1977). Both are in good agreement with the longest dimensions measured from the unhydrated pentamer and decamer crystal structures of 10.5 nm and 11.9 nm respectively, provided that these were increased to 11.1 nm and 12.5 nm respectively by the addition of a uniform shell of water molecules of thickness 0.3 nm on all surfaces (Thompson *et al.*, 1999; Ramadan *et al.*, 2002; Perkins, 2001). The lack of significant change in  $L$  on decamer formation shows that the pentamers are arranged face-to-face within the decamer.

X-ray scattering showed that  $\text{Ca}^{2+}$  removal from CRP promoted some self-aggregation, and reduced the proportion of CRP decamer, but left the CRP conformation largely unchanged.  $\text{Ca}^{2+}$  removal was studied using a CRP concentration range of 0.1-2.5 mg/ml using 50 mM and 140 mM NaCl buffers without  $\text{Ca}^{2+}$ . Trace amounts of non-specific aggregation were readily visible in the Guinier plots as non-linear increases in the  $I(Q)$  values at very low  $Q$  values. Accordingly a reduced  $Q$  range was employed for the  $R_G$  fits (Figure 7.5(b)). The  $I(0)/c$  values showed a reduced concentration dependence compared to that in the presence of  $\text{Ca}^{2+}$ , increasing from 11.7 to 12.9 units and from 8.6 to 11.6 units in 50 mM NaCl and 140 mM NaCl respectively (Figure 7.6(b)). This small increase showed that, even though the pentamer-decamer association is weakened compared to studies in the presence of  $\text{Ca}^{2+}$ , the pentamer-decamer equilibrium is not eliminated by the removal of  $\text{Ca}^{2+}$ . The  $R_G$  values of CRP likewise increased in smaller ranges from  $3.8 \pm 0.0$  nm to  $4.1 \pm 0.0$  nm in 50 mM NaCl, and from  $3.6 \pm$

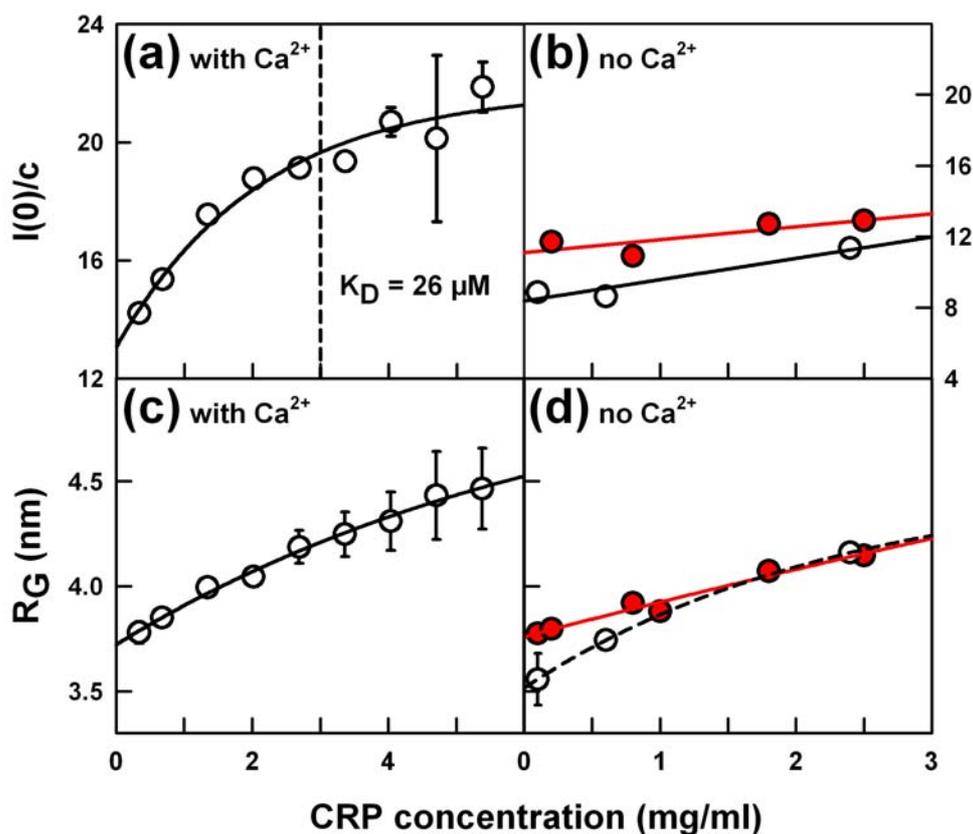


**Figure 7.5**

X-ray scattering Guinier  $R_G$  analyses for CRP.

(a) The filled circles correspond to the data points used to obtain  $R_G$  values using a  $Q$  range of  $0.06\text{-}0.36\text{ nm}^{-1}$  and the straight lines correspond to the best fit through these points. The  $Q.R_G$  fit range is arrowed. In the presence of  $2\text{ mM Ca}^{2+}$ , from bottom to top, concentrations of  $0.3\text{ mg/ml}$ ,  $1.3\text{ mg/ml}$ ,  $2.7\text{ mg/ml}$ ,  $4.0\text{ mg/ml}$  and  $5.4\text{ mg/ml}$  were used.

(b) In the absence of  $\text{Ca}^{2+}$ , the Guinier fits were made in a  $Q$  range of  $0.18\text{ - }0.36\text{ nm}^{-1}$ . From bottom to top, the CRP concentrations were  $0.1\text{ mg/ml}$ ,  $0.3\text{ mg/ml}$ ,  $0.6\text{ mg/ml}$  and  $2.4\text{ mg/ml}$ .



**Figure 7.6**

Concentration dependence of X-ray scattering Guinier  $R_G$  analyses for CRP.

(a) Concentration dependence of the Guinier  $I(0)/c$  parameter in 140 mM NaCl buffer containing 2 mM  $\text{Ca}^{2+}$ . The  $K_D$  value was determined by a non-linear regression fit (solid line) for an exponential rise to a maximum ( $y = y_0 + a(1-b^x)$ ), and the vertical dashed line corresponds to 26  $\mu\text{M}$ . Error bars are shown where visible.

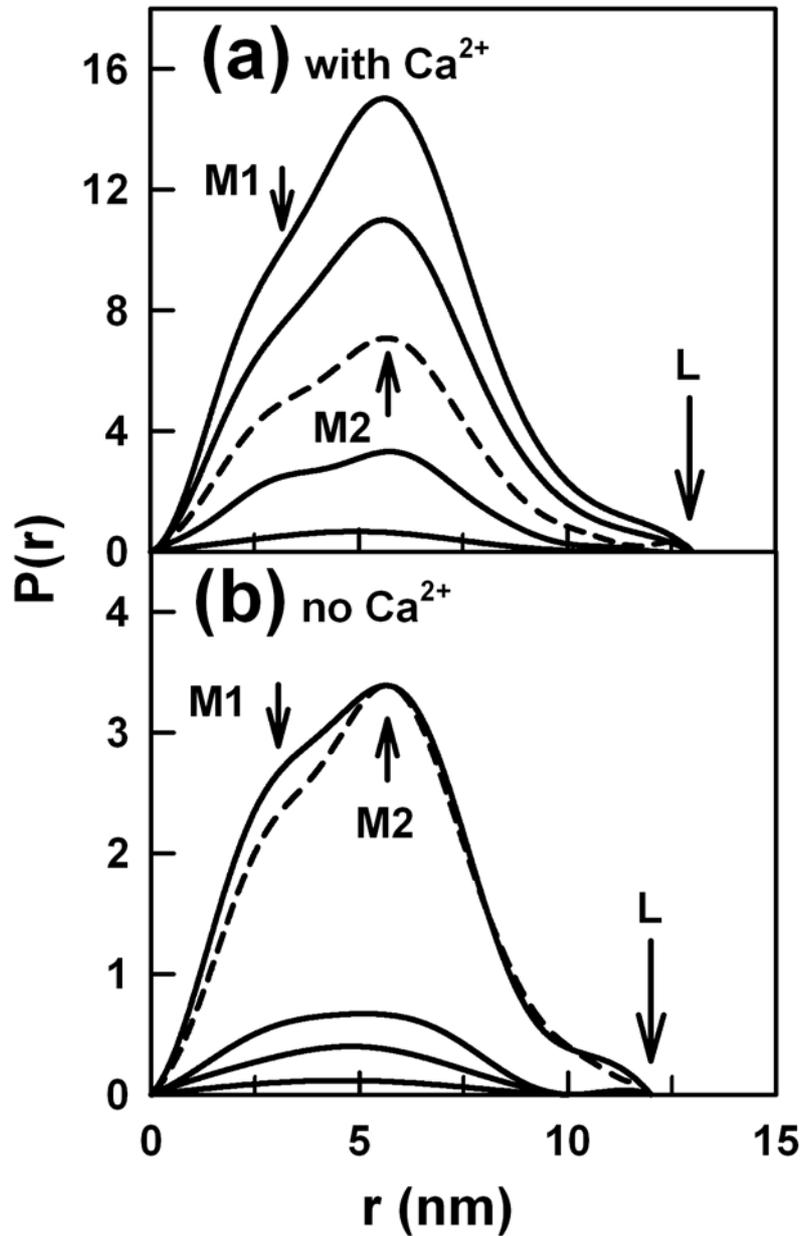
(b) Concentration dependence of the Guinier  $I(0)/c$  parameter in 50 mM NaCl (●) and 140 mM NaCl buffer (○) without  $\text{Ca}^{2+}$ . Note that the  $I(0)/c$  values here were obtained in a different beam session from that of (a), therefore the  $I(0)/c$  values are not comparable. No  $K_D$  value could be determined.

(c,d) The corresponding Guinier  $R_G$  values for CRP in 140 mM NaCl buffer containing 2 mM  $\text{Ca}^{2+}$ , and in 50 mM NaCl (●) and 140 mM NaCl buffer (○) without  $\text{Ca}^{2+}$ .

0.1 nm to  $4.1 \pm 0.0$  nm in 140 mM NaCl buffers (Figure 7.6(d)). Extrapolation to zero concentration yielded an  $R_G$  value of 3.5 nm for the pentamer in 140 mM NaCl (Figure 7.6(d)); this is similar to the  $R_G$  value of 3.7 nm for the pentamer in 140 mM NaCl with  $\text{Ca}^{2+}$  present (Figure 7.6(c)). The  $P(r)$  curves yielded  $R_G$  values of 3.7-4.1 nm in 50 mM NaCl (data not shown) and 3.6-4.0 nm in 140 mM NaCl (Figure 7.7(b)), in agreement with the Guinier analyses. The  $r$  values of the peaks  $M1$  ( $r = 3.1$  nm) and  $M2$  ( $r = 5.7$  nm) were unchanged on  $\text{Ca}^{2+}$  removal, although the increase in the intensity of the  $M2$  peak relative to  $M1$  was now reduced. The  $L$  values of 10-13 nm showed no change from the  $L$  values in buffers containing  $\text{Ca}^{2+}$ . Thus no conformational change in CRP occurred on  $\text{Ca}^{2+}$  removal.

#### (7.2.4) Modelling of pentameric and decameric CRP

Constrained modelling starting from crystal structures permits the molecular interpretation of scattering curves (Perkins *et al.*, 2008). Here, this was used to determine a  $K_D$  value for the CRP pentamer-decamer equilibrium and establish the similarity of the solution structure of the CRP decamer to its crystal structure. The scattering curve for the CRP pentamer was calculated from its crystal structure with  $\text{Ca}^{2+}$  (PDB code 1b09)(Thompson *et al.*, 1999). Because several residues were missing and  $\text{Ca}^{2+}$  was not present in the CRP decamer crystal structure (PDB code 1lj7) (Ramadan *et al.*, 2002), the decamer scattering curve with  $\text{Ca}^{2+}$  was calculated after superimposition of two pentamer crystal structure onto the decamer crystal structure. The modelled pentamer and decamer scattering curves resulted in  $R_G$  values of 3.7 nm and 4.2 nm respectively (Table 7.1). These agree well with the experimental  $R_G$  value of 3.7 nm on extrapolation for the pentamer and the observed range of  $3.8 \pm 0.1$  nm to  $4.5 \pm 0.2$  nm for the pentamer-decamer equilibrium (Figure 7.6(c)). The amounts of pentamer and decamer in each X-ray curve was computed from 101 combinations of the modelled pentamer and decamer curves in 1% increments from 0%-100% to 100%-0% respectively. A total of nine experimental X-ray curves at different concentrations in 140 mM NaCl with  $\text{Ca}^{2+}$  were assessed using a goodness-of-fit  $R$ -factor (Figure 7.8(a)). Excellent  $R$ -factors of 2.0% to 3.7% were obtained for pentamer-decamer ratios that ranged from 74%-26% at 0.3 mg/ml to 28%-72% at 5.4 mg/ml (Figure 7.9(a)). In comparison, the worst possible  $R$ -factor was 15.0%. The pentamer-decamer ratios resulted in a  $K_D$  value of 16  $\mu\text{M}$  (Figure 7.9(a)) which agrees well with the Guinier-determined value of 26  $\mu\text{M}$  (Figure



**Figure 7.7.**

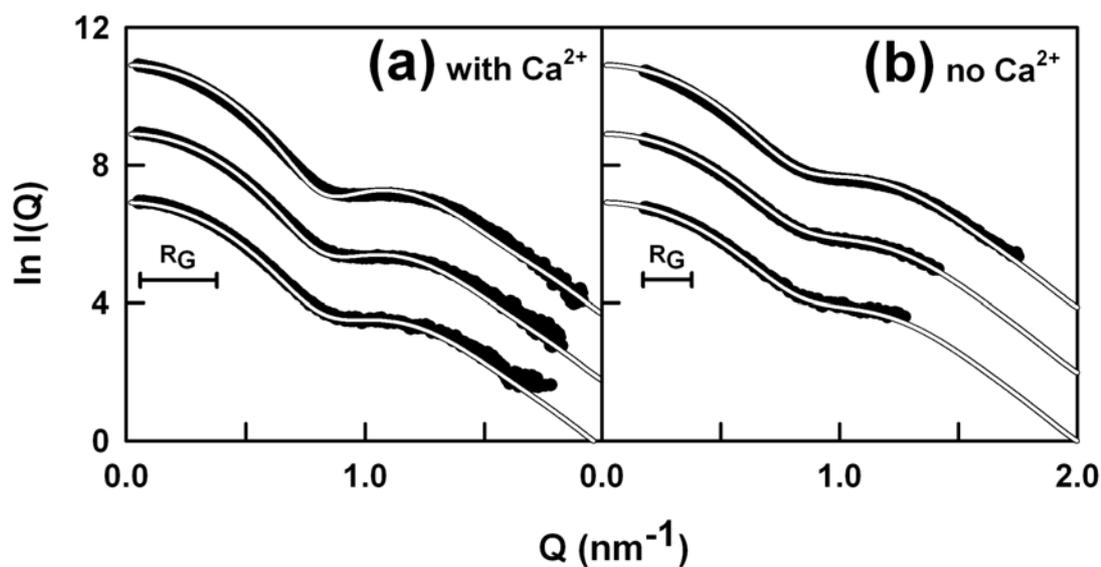
Concentration dependence of the X-ray distance distribution  $P(r)$  for CRP. CRP was studied in 140 mM NaCl buffers (a) containing 2 mM  $\text{Ca}^{2+}$  and (b) without  $\text{Ca}^{2+}$ . The two maxima in  $P(r)$  are denoted by  $M1$  and  $M2$  at 3.1 nm and 5.7 nm respectively, assigned to pentamer and decamer respectively. The maximum dimension is denoted by  $L$ . The CRP concentrations correspond to those in [Figures 7.5\(a,b\)](#). The dashed line in (a) corresponds to 2.7 mg/ml; in (b), this is shown again and normalised to  $M2$  using that at 2.4 mg/ml in order to compare the effect of  $\text{Ca}^{2+}$  removal.

7.6(a)). The modelled and experimental  $P(r)$  curves likewise agree well (Figure 7.9(c)). The success of these curve fits shows that the decamer crystal structure with the two A-faces facing each other at the centre is a good model for its solution structure.

Constrained fits confirmed the NaCl concentration dependence of the pentamer-decamer equilibrium in the absence of  $\text{Ca}^{2+}$  that was seen by AUC and Guinier analyses, and the decreased formation of decamer. The comparison of the 101 pentamer/decamer curves with five experimental X-ray curves in 50 mM NaCl again resulted in excellent best-fit  $R$ -factors of 1.8% to 3.5% (Table 7.1). The good quality of the curve fits indicated the absence of large conformational change in CRP upon  $\text{Ca}^{2+}$  removal. The best-fit ratios ranged from 100%-0% to 56%-44% of pentamer/decamer (Figure 7.9(b)). Likewise the four experimental curves in 140 mM NaCl buffer resulted in excellent best-fit  $R$ -factors of 1.7% to 2.8% (Figure 7.8(b); Table 7.1). The best-fit ratios of 100%-0% to 71%-29% showed that the proportion of decamer had decreased in 140 mM NaCl after  $\text{Ca}^{2+}$  removal (Figure 7.9(b)). No  $K_D$  values were determined because these were too large to be measured. The modelled  $P(r)$  analyses showed good agreement with experimental data (Figure 7.9(d)). Even though lower molecular-weight CRP oligomers will be present according to the  $c(s)$  plots (Figure 7.4(b)), these will have a small effect on curve fits because the scattered intensities are proportional to the square of the molecular weight.

The CRP scattering curves show a minimum at a  $Q$  value of  $0.87 \text{ nm}^{-1}$  (Figure 7.10(a)). This minimum is more prominent in the decamer than the pentamer. The mean diameter of the pentameric ring is given by  $2\pi/Q$  as 7.2 nm, which is in good agreement with the neutron value of 7.4 nm (Perkins & Pepys, 1986) and a value of 7.5 nm measured from the pentamer and decamer crystal structures (Figure 7.10(b)). The agreements support the face-to-face crystallographic model for the CRP decamer, in which one pentamer is rotated relative to the other by about  $20^\circ$ , then brought into contact at their A faces (Ramadan *et al.*, 2002).

Sedimentation coefficients  $s_{20,w}^\circ$  for the pentamer and decamer were calculated from their crystal structure models (Table 7.1). Those for pentameric CRP were 6.0-6.2 S, in good agreement with the experimental value of 6.4 S. That for the decamer was 9.8 S, for which no experimental value from this study was available for comparison. The value of 9.8 S is in good agreement with the experimental value

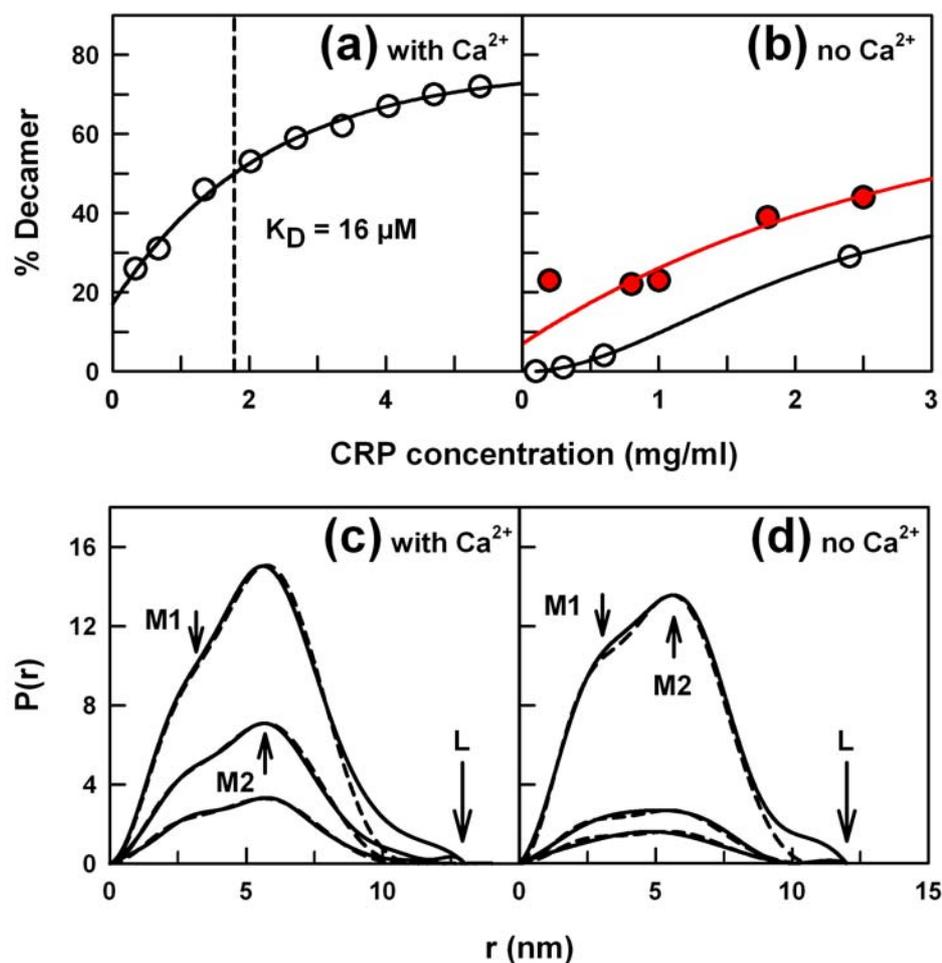


**Figure 7.8.**

X-ray scattering curve fits for the CRP pentamer and decamer in the presence and absence of  $\text{Ca}^{2+}$ . The  $Q$  range for the  $R_G$  analyses is denoted by a horizontal bar.

**(a)** In buffer with 2 mM  $\text{Ca}^{2+}$ , representative best-fit curves are shown for CRP concentrations of 1.3 mg/ml (54% pentamer/ 46% decamer), 2.7 mg/ml (41% pentamer/ 59% decamer) and 5.4 mg/ml (28% pentamer/ 72% decamer) from bottom to top.

**(b)** In  $\text{Ca}^{2+}$ -depleted buffer, representative best-fit curves are shown for CRP concentrations of 0.3 mg/ml (99% pentamer/ 1% decamer), 0.6 mg/ml (96% pentamer/ 4% decamer) and 2.4 mg/ml (71% pentamer/ 29% decamer).



**Figure 7.9.**

Concentration dependence of X-ray scattering curve fits for the CRP pentamer and decamer in the presence and absence of  $\text{Ca}^{2+}$ .

(a)  $K_D$  determination for the CRP pentamer/decamer equilibrium in the presence of 2 mM  $\text{Ca}^{2+}$ . Data points correspond to the best-fit pentamer/decamer ratio for each CRP concentration between 0.3–5.4 mg/ml in 140 mM NaCl buffer. The data were fitted according to Figure 7.6(a), and the vertical dashed line corresponds to a  $K_D$  of 16  $\mu\text{M}$ .

(b) In  $\text{Ca}^{2+}$ -depleted buffer, no  $K_D$  value could be determined as this was outside the concentration range studied.

(c,d) Comparison of the modelled (dashed lines) and experimental (full lines)  $P(r)$  curves in the (c) presence and (d) absence of  $\text{Ca}^{2+}$ .

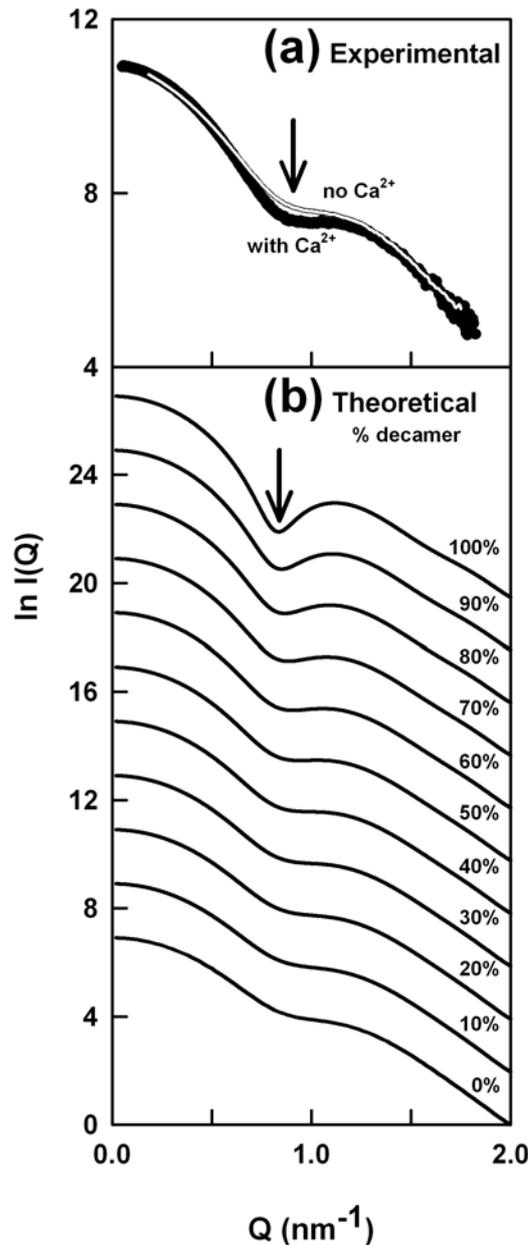
of 9.5 S for the cross-linked CRP decamer (Blizniukov *et al.*, 2003).

### **(7.2.5) Surface plasmon resonance (SPR) analyses of CRP**

SPR monitors the interaction between a binding partner in solution (analyte) and a partner (ligand) immobilised at the surface of a sensor chip (Tanious *et al.*, 2008). CRP accumulates at damaged and apoptotic cells by binding to charged groups at their cell surfaces (Volanakis, 1982; De Beer *et al.*, 1982; Rowe *et al.*, 1984). Accordingly the self-association of CRP in solution with CRP immobilised on a sensor chip can provide insight on this interaction, provided that  $\text{Ca}^{2+}$  is present throughout the SPR experiments.

The NaCl concentration dependence of CRP self-association seen by AUC and scattering was confirmed by SPR. CRP was flowed across a sensor chip surface containing 150 resonance units (RU) of immobilised CRP which was prepared in the presence of  $\text{Ca}^{2+}$  (Section 7.4.5). The binding analysis of 50  $\mu\text{M}$  (5.8 mg/ml) CRP in 50 mM NaCl buffer in the presence of  $\text{Ca}^{2+}$  showed that CRP aggregated on the chip surface, as evidenced by the extremely high binding response to greater than 15,000 RU. This was rapidly followed by CRP dissociation to baseline levels (Figure 7.11(a)). On increasing the salt to 140 mM NaCl, the corresponding analysis of 50  $\mu\text{M}$  CRP resulted in a much-reduced but more rapid binding response of 34 RU, followed by its rapid dissociation (Figure 7.11(b)). This represents 23% saturation of immobilised CRP. At a lower CRP concentration of 20  $\mu\text{M}$  (2.3 mg/ml), the binding responses were reduced to 245 RU in 50 mM NaCl and to 24 RU in 140 mM NaCl (Figures 7.11(c,d)).

Dissociation constants  $K_D$  for CRP self-association were determined. Equilibrium SPR studies in 50 mM NaCl and 140 mM NaCl buffers with  $\text{Ca}^{2+}$  present supported these binding analyses. At concentrations of 1-20  $\mu\text{M}$  in 50 mM NaCl buffer, CRP showed a continuously increasing binding response consistent with aggregate formation (Figure 7.12(a,c)). A  $K_D$  value could not be determined because binding saturation was not attained. In 140 mM NaCl buffer, a  $K_D$  of 23  $\mu\text{M}$  was determined using CRP concentrations between 1-35  $\mu\text{M}$  (Figure 7.12(b,d)). This agrees well with the values of 26  $\mu\text{M}$  and 16  $\mu\text{M}$  in 140 mM NaCl from the X-ray analyses above, showing that immobilised CRP behaves in the same way as CRP in solution.



**Figure 7.10.**

Comparison of the X-ray scattering curves for CRP pentamer and decamer. The arrow corresponds to the position of the minimum in the scattering curve.

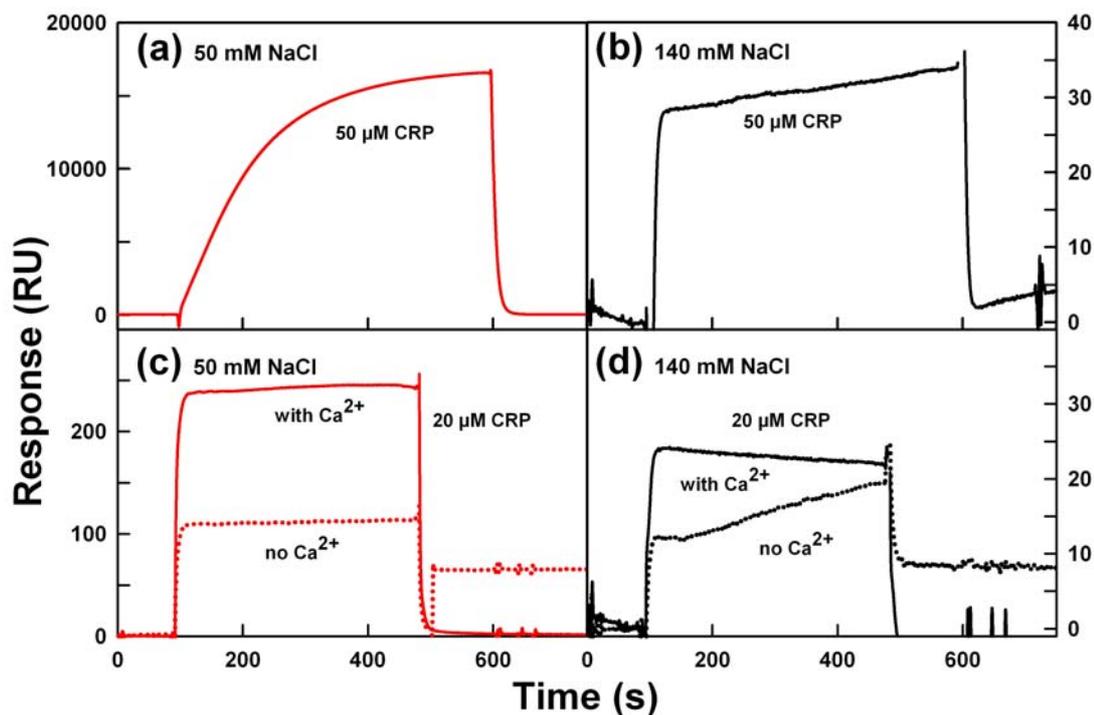
(a) Comparison of the experimental scattering curves for CRP at 2.7 mg/ml in 2 mM  $\text{Ca}^{2+}$  buffer (black) and 2.4 mg/ml in  $\text{Ca}^{2+}$ -depleted buffer (white). The intensity of  $I(Q)$  at the minimum (arrowed) is lower in the presence of  $\text{Ca}^{2+}$ , showing that more decamer has formed

(b) Comparison of the scattering curves for CRP for mixtures of pentamer and decamer calculated from crystal structures. The intensity of the  $I(Q)$  minimum decreases as the proportion of decamer increases from 0% to 100%.

The corresponding experiments in the absence of  $\text{Ca}^{2+}$ , but using the same immobilised CRP sensor chip, showed that CRP binding to immobilised CRP was much reduced. Using 20  $\mu\text{M}$  CRP, the binding responses were reduced from 245 RU to 115 RU in 50 mM NaCl buffer, and from 24 RU to 19 RU in 140 mM buffer (Figures 7.11(c,d)). Interestingly, with both buffers, CRP did not completely dissociate from immobilised CRP on the sensor surface back to baseline levels when this was compared with buffers containing  $\text{Ca}^{2+}$  (dashed lines in Figures 7.11(c,d)). This suggested that, in the absence of  $\text{Ca}^{2+}$ , lower molecular-weight oligomers of CRP bind irreversibly to partially-dissociated immobilised CRP on the sensor chip surface to form aggregates. This is consistent with the AUC results showing reduced CRP decamer formation and partial dissociation of CRP in  $\text{Ca}^{2+}$ -depleted buffers.

### (7.3) Conclusions

In this study, it has been shown that CRP in 2 mM  $\text{Ca}^{2+}$  exists in a rapid pentamer-decamer equilibrium in solution using three different analytical instrumentation methods with improved resolutions compared to previous studies. This equilibrium is dependent on NaCl concentration. As far as could be determined, this is the first evidence for decameric CRP in physiologically-relevant buffer conditions. The CRP concentration in plasma ranges from 0.8  $\mu\text{g}/\text{ml}$  (0.01  $\mu\text{M}$ ) at normal levels to over 0.5 mg/ml (4.4  $\mu\text{M}$ ) during acute phase conditions (Pepys & Hirschfield, 2003). Calcium is present in plasma at 2.5 mM, of which ionic  $\text{Ca}^{2+}$  is present at 1.18 mM (Hurwitz, 1996). Only CRP pentamers and decamers were observed in 2 mM  $\text{Ca}^{2+}$  in CRP concentrations between 0.1-5.4 mg/ml (0.9-47  $\mu\text{M}$ ), where these concentrations include the physiological range of the CRP acute phase response. The formation of decameric CRP with a mean  $K_D$  value of  $22 \pm 5 \mu\text{M}$  can be relevant for the interaction of CRP with its ligands in plasma. However, as plasma contains at least 70 mg/ml of other proteins, more detailed knowledge of how CRP behaves in this milieu will be required before these findings can be extrapolated to CRP in plasma. Most importantly, the AUC data shows that the non-physiological removal of  $\text{Ca}^{2+}$  causes CRP to dissociate into lower oligomers and reduces the proportion of decamers. Many previous functional studies of CRP have reported the existence of denatured or monomeric forms of CRP in the absence of  $\text{Ca}^{2+}$ , and have suggested that these are functionally relevant (Pepys & Hirschfield, 2003). There is

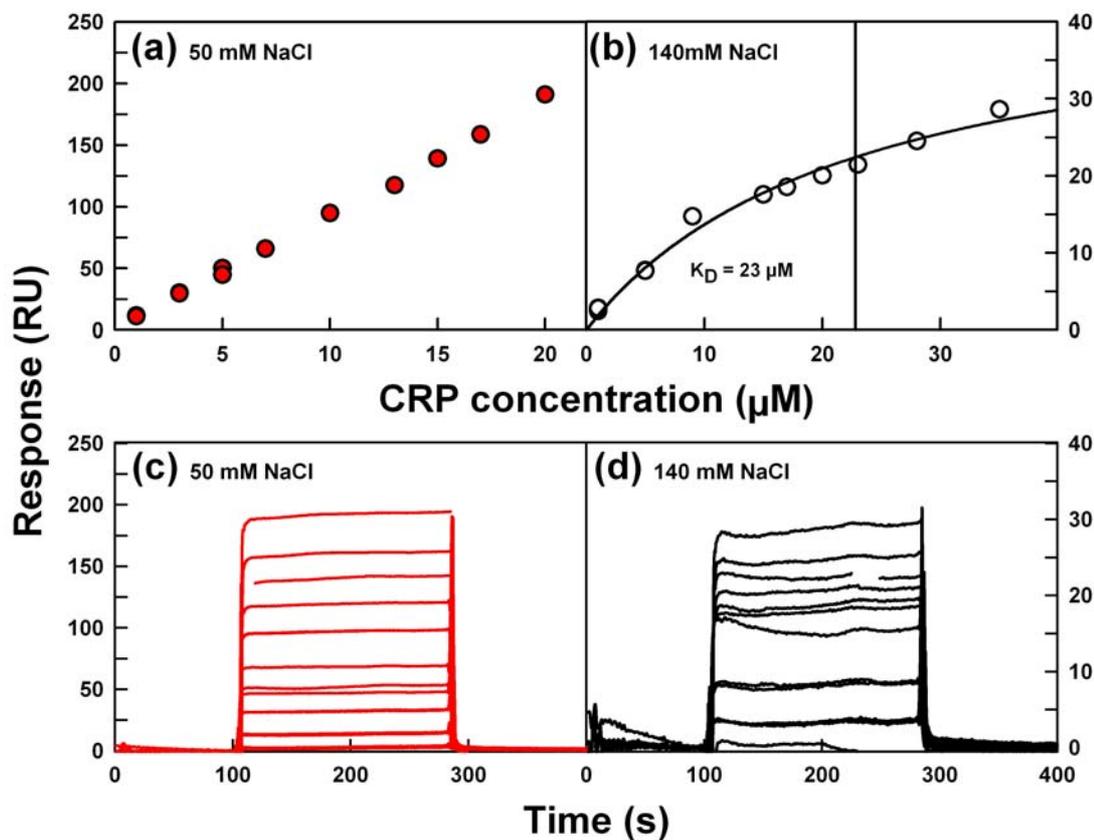


**Figure 7.11.**

SPR binding analysis of CRP self-association. Data correspond to buffers containing 2 mM  $\text{Ca}^{2+}$  unless stated otherwise.

**(a,b)** 50  $\mu\text{M}$  CRP was flowed over immobilised CRP in **(a)** 50 mM NaCl and **(b)** 140 mM NaCl buffer. An abnormally high response ( $> 15000$  RU) was observed in 50 mM NaCl, while that in 140 mM NaCl was normal ( $\sim 35$  RU).

**(c,d)** Comparison of the binding of CRP in the presence of  $\text{Ca}^{2+}$  (solid lines) and absence of  $\text{Ca}^{2+}$  (dotted lines). 20  $\mu\text{M}$  of CRP was flowed over immobilised CRP in buffers containing **(c)** 50 mM NaCl buffer and **(d)** 140 mM NaCl. Approximately ten-fold more response was observed in **(c)** than **(d)**. The dotted lines did not return to baseline levels in **(c)** and **(d)**.



**Figure 7.12.**

SPR equilibrium analysis of CRP self-association. Data correspond to buffers containing 2 mM  $\text{Ca}^{2+}$ .

**(a,b)** Equilibrium analyses for CRP self-association in **(a)** 50 mM NaCl and **(b)** 140 mM NaCl buffer. The binding response increased continuously between 1-20 μM in **(a)** while saturation was approached in **(b)** from which a  $K_D$  value of 23 μM was determined from CRP concentrations between 1-35 μM.

**(c,d)** Sensorgrams used for the equilibrium analyses, where panel **(c)** corresponds to the data in **(a)**, and panel **(d)** corresponds to the data in **(b)**.

no rigorous evidence for the existence of monomeric (i.e. denatured) CRP *in vivo*. Standard preparation protocols for CRP involve 2 mM  $\text{Ca}^{2+}$  throughout (De Beer & Pepys, 1982). Because the AUC data confirm that pentameric CRP is very stable in the presence of  $\text{Ca}^{2+}$ , functional or structural experiments with CRP in which  $\text{Ca}^{2+}$  is omitted have little or no physiological or pathophysiological relevance.

The CRP pentamer-decamer equilibrium was previously unknown. The self-association of pure CRP had been previously reported by X-ray and neutron scattering and by AUC, but was not explained (Perkins & Pepys, 1986; Blizniukov *et al.*, 2003). Here, the molecular basis of this self-association was identified by the combination of the  $c(s)$  analyses from AUC and X-ray scattering modelling. The AUC data revealed a weak pentamer-decamer equilibrium in three NaCl concentrations in the presence of  $\text{Ca}^{2+}$ . While pentamers were unambiguously identified as the 6.4 S  $c(s)$  peak, the second  $c(s)$  peak corresponds to a reaction boundary that could not be assigned to a decamer. This ambiguity was resolved by the excellent X-ray curve fits based on the pentameric and decameric CRP crystal structures. The  $K_D$  values of 19  $\mu\text{M}$  (50 mM NaCl) by AUC, 26  $\mu\text{M}$  and 16  $\mu\text{M}$  (140 mM NaCl) by X-ray Guinier fits and modelling, and 23  $\mu\text{M}$  (140 mM NaCl) by SPR show consistency. The mean  $K_D$  value of  $22 \pm 5 \mu\text{M}$  (140 mM NaCl) suggests that 0% to 15% of CRP will be decameric in its physiological concentration range of 0.01-4.4  $\mu\text{M}$  if no other factors are involved. There were no indications of any indefinite self-association of CRP. The decamer is therefore formed from contacts between either the two A faces or the two B faces of the pentamers. The A faces are preferred as the buried interface because this is observed in the  $\text{Ca}^{2+}$ -free CRP decamer crystal structure, excellent X-ray modelling fits were obtained with the decamer crystal structure (Ramadan *et al.*, 2002), and ionic interactions at the A faces were observed that account for the NaCl concentration dependence of decamer formation (see below). If CRP decamers are formed in plasma, the B faces will remain exposed for lipid binding at damaged host cells or bacterial surfaces and will not be affected. Given that CRP interacts with C1q of complement at its A face, CRP decamer formation may potentially inhibit C1q binding to CRP during a heightened acute phase response. The observation of a pentamer-decamer equilibrium and the absence of any observed 15-mers of CRP indicates that functional studies of chemically cross-linked 15-mers of CRP are not physiologically relevant (Jiang *et al.*, 1991).

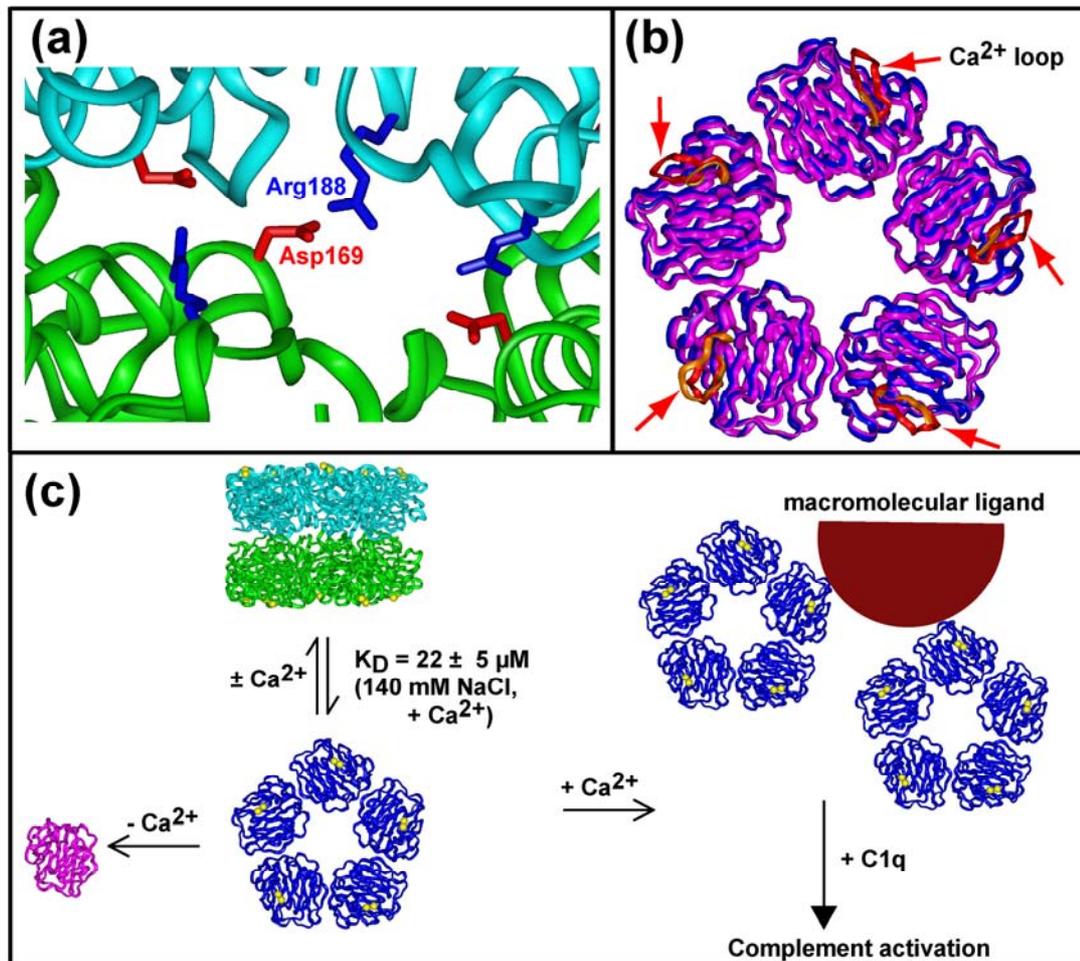
The reduced formation of CRP decamer with increase in NaCl concentration is explained by the CRP crystal structures. This dependence indicates that the A face-A face interface in the decamer is stabilised by ionic interactions. The partially-disordered decamer crystal structure suggested that ten pairs of Thr173 and Pro179 sidechains form the closest contact between the A faces (Ramadan *et al.*, 2002). If the Ca<sup>2+</sup>-bound pentamer crystal structure at higher resolution was superimposed onto the decamer crystal structure (Thompson *et al.*, 1999), it turns out that ten pairs of Asp169 and Arg188 residues on the A faces are well positioned to form the requisite ionic bridges (Figure 7.13(a)). There are no clear candidates for ionic bridges between the B faces. There are Lys69-Asp70 and Glu14-Arg47 pairs within each B face which cancel each other out. Additionally, there is a surplus of acidic residues on the B face (in addition to those involved with Ca<sup>2+</sup> and phosphocholine binding) which may repel each other. The NaCl concentration dependence of decamer formation shows that this may be affected by the local environment. For example, CRP binding to ligands at the exposed B faces may influence the pentamer-decamer association. In this context, CRP is known to interact with a variety of polycations, as well as with polyanion-polycation complexes (Siegel *et al.*, 1974; 1975; Claus *et al.*, 1975).

Previous high resolution CRP crystallographic studies have highlighted the role of Ca<sup>2+</sup> in the formation of its structure. This emphasis is justified by the presence of at least 2 mM Ca<sup>2+</sup> in plasma, which is well in excess of the  $K_D$  for Ca<sup>2+</sup> binding to CRP. Both Ca<sup>2+</sup>-binding sites in CRP possess equal affinity for Ca<sup>2+</sup> in solution with a  $K_D$  of 60  $\mu$ M (Kinoshita *et al.*, 1989). While two CRP pentamer crystal structures were reported in the presence of Ca<sup>2+</sup> (Shrive *et al.*, 1996; Thompson *et al.*, 1999), CRP was crystallised as decamers in 100 mM NaCl buffer without Ca<sup>2+</sup> (Ramadan *et al.*, 2002). The AUC data explains the crystallisation of decamers without Ca<sup>2+</sup> present, simply because decamers were still observable under these conditions. The AUC and gel filtration data also provide insight on the effect of Ca<sup>2+</sup> removal on CRP. In the crystal structure, Ca<sup>2+</sup> binds at a double site on the B face (Figure 7.13(b)). Ca<sup>2+</sup> site 1 is coordinated by amino acids Asp60, Asn61, Glu138, Asp140 and Leu139 (a total of five ligands), while Ca<sup>2+</sup> site 2 is coordinated by Gln138, Asp140, Gln150 and possibly Glu147 (Shrive *et al.*, 1996; Thompson *et al.*, 1999). The two Ca<sup>2+</sup>-binding sites in CRP are therefore differentiated by one coordinating amino acid, even though they are equally occupied in the crystal

structure. Thus  $\text{Ca}^{2+}$  is not structurally involved with decamer formation that probably occurs through A face contacts. In addition,  $\text{Ca}^{2+}$  is not structurally related to the contacts between the five protomers in the pentamer, each of which is stabilised by four ionic bridges that span the interface (Glu42-Lys119; Glu101-Lys201; Asp155-Arg118; Glu197-Lys123). Accordingly the AUC- and gel filtration-observed weakening of decamer formation and the dissociation of smaller oligomers on the removal of  $\text{Ca}^{2+}$  are best explained by a destabilisation of the entire double  $\beta$ -sheet structure of the CRP protomer. Indeed, comparisons of CRP crystal structures with and without  $\text{Ca}^{2+}$  bound (Figure 7.13(b)) show that there is a 0.174 nm root mean square difference in residue positions (Thompson *et al.*, 1999). This conformational sensitivity of the  $\beta$ -sheets to residue change provides an explanation for the notable absence of polymorphic forms of CRP (Pepys & Hirschfield, 2003).

Several CRP studies have proposed that alterations in  $\text{Ca}^{2+}$  binding are important for the mechanism of CRP action, however this view is not supported by this study. In the absence of  $\text{Ca}^{2+}$ , CRP is partially dissociated to form lower molecular-weight oligomers (Figure 7.13(c)). By the current AUC studies, these smaller oligomers were not in equilibrium with pentameric CRP. The existence of these lower oligomers of CRP consisting of monomers to trimers (mCRP) has been reported previously in functional studies under conditions of denaturing pH, temperature (Potempa *et al.*, 1983; 1987) or  $\text{Ca}^{2+}$ -depletion (Motie *et al.*, 1996; Wu *et al.*, 2002). Given the high stability of CRP pentamers in the presence of  $\text{Ca}^{2+}$ , their functional relevance is questionable. AUC and X-ray data from this thesis show that the CRP pentamer does not undergo large conformational changes with NaCl concentration or  $\text{Ca}^{2+}$ -depletion and is highly stable. Given the prevalence of  $\text{Ca}^{2+}$  in plasma, no allosteric mechanism involving the removal or addition of  $\text{Ca}^{2+}$  during CRP function is likely, as proposed in earlier studies (Ramadan *et al.* 2002; Blizniukov *et al.*, 2003). The existence of monomeric CRP with a functional role is likewise questionable (Wu *et al.*, 2002; Wang & Sui, 2001; Taylor & van den Berg, 2006).

The understanding of the solution structural behaviour of CRP from this study is relevant to the elucidation of its function in complement activation and host acute-phase response. The main results show that CRP exists as very stable pentamers in a rapid equilibrium with decamers in 2 mM  $\text{Ca}^{2+}$  (Figure 7.13). On the other hand, the spontaneous formation of lower molecular-weight oligomers of CRP



**Figure 7.13.**

Pentamers and decamers of CRP as a function of NaCl concentration and  $\text{Ca}^{2+}$ .

(a) Crystal structures for pentameric CRP with bound  $\text{Ca}^{2+}$  (cyan and green ribbons) were superimposed onto the crystal structure of decameric CRP without  $\text{Ca}^{2+}$  to reveal ten pairs of Arg188 (blue) -Asp169 (red) ionic bridges that are proposed to stabilise CRP decamer formation between the two A faces. Three pairs of bridges are shown in this view. (b) The B face of pentameric CRP is viewed in blue with bound  $\text{Ca}^{2+}$ , and in magenta without  $\text{Ca}^{2+}$ . The  $\text{Ca}^{2+}$ -binding loop comprising amino acid residues 138-150 in each protomer is coloured red (with  $\text{Ca}^{2+}$ ) or orange (without  $\text{Ca}^{2+}$ ). (c) Pentameric CRP (blue, with bound  $\text{Ca}^{2+}$  shown as yellow spheres) self-associates in a rapid equilibrium in the presence or absence of  $\text{Ca}^{2+}$  to form decamer (green and cyan) through contacts between the A faces in an NaCl concentration-dependent manner. This has a mean dissociation constant  $K_D$  value of  $22 \mu\text{M}$  determined in 140 mM NaCl containing  $\text{Ca}^{2+}$ . In the absence of  $\text{Ca}^{2+}$ , CRP irreversibly dissociates into individual protomers (magenta) and decamer formation is reduced. When  $\text{Ca}^{2+}$  is present, pentameric CRP binds to macromolecular ligands to form complexes that interact with C1q to activate the classical pathway of complement.

(termed mCRP) under conditions of  $\text{Ca}^{2+}$ -depletion was readily seen by AUC. SPR studies indicated that these lower molecular-weight CRP oligomers can bind to immobilised CRP, i.e. these lower molecular-weight oligomers can associate with ligand-bound CRP *in vivo* and may modify its biological activity. These results are cautionary in the sense that experimental work with CRP readily leads to artefactual observations; this most probably results from the ease with which  $\text{Ca}^{2+}$  can be removed from CRP during experimental studies. Indeed CRP becomes denatured and shows subunit rearrangement on binding to membranes (Wang & Sui, 2001). Studies have suggested that mCRP may be produced in normal or inflamed human and animal tissue (Samols *et al.*, 1985; Rees *et al.*, 1988; Egenhofer *et al.*, 1993; Diehl *et al.*, 2000), and that mCRP possesses distinct antigenic properties from native CRP (Potempa *et al.*, 1983; 1983; 1987; Khreiss *et al.*, 2002). There is no robust evidence for these observations (Pepys & Hirschfield, 2003). This present study emphasizes the stability of CRP pentamers and the occurrence of an equilibrium with decamers, and the instability of other different oligomeric forms of CRP. The proper appreciation of this will open the way for a more rational understanding of the role of CRP in the acute phase response and in complement activation.

#### **(7.4) Materials and Methods.**

##### **(7.4.1) Purification of CRP**

Human CRP was isolated and purified as described previously (De Beer & Pepys, 1982). Samples were extensively dialysed into 10 mM Tris buffers (with or without 2 mM  $\text{CaCl}_2$ ) containing 50 mM NaCl, 140 mM NaCl, or 250 mM NaCl, all at pH 8.0, for AUC and X-ray scattering data acquisition. Size exclusion chromatography was carried out using HEPES buffers (10 mM HEPES, 140 mM NaCl, pH 8.0), with or without 2 mM  $\text{CaCl}_2$ . Samples were routinely analysed by SDS-PAGE before and after experiments to confirm their integrity. CRP concentrations were determined using an absorption coefficient of 17.5 (1%, 280 nm, 1 cm path length), which agrees well with a value of 19.2 calculated from its composition (Nelson *et al.*, 1991; Perkins, 1986). CRP is unglycosylated and its composition resulted in a calculated molecular weight of 115.0 kDa, an unhydrated volume of 150.3 nm<sup>3</sup>, a hydrated volume of 197.3 nm<sup>3</sup> and a partial specific volume  $\bar{v}$  of 0.7408 ml/g (Perkins, 1986). The buffer densities for AUC were measured at 20°C using an Anton-Paar DMA5000 density meter to be 1.00096 g/ml or 1.00092

g/ml (50 mM NaCl), 1.00482 g/ml or 1.00455 g/ml (140 mM NaCl), with or without 2 mM CaCl<sub>2</sub> in that order, and 1.00934 g/ml (250 mM NaCl).

#### **(7.4.2) Analytical ultracentrifugation data and analyses**

AUC data for CRP were obtained on two Beckman XL-I instruments equipped with AnTi50 and AnTi60 rotors. Sedimentation velocity data was acquired at 20°C at a rotor speed of 50,000 r.p.m. in two-sector cells with column heights of 12 mm. Sedimentation analysis was performed using direct boundary Lamm fits of up to 300 scans using SEDFIT (version 11.0) (Schuck, 1998; 2000). SEDFIT resulted in size-distribution analyses  $c(s)$  that assume that all macromolecular species have the same frictional ratio  $f/f_o$ . The final SEDFIT analyses used a fixed resolution of 200 and optimised the  $c(s)$  fit by floating  $f/f_o$  and the baseline until the overall root mean square deviations and visual appearance of the fits were satisfactory (Figure 7.2). The percentage fraction of pentamers, decamers or lower oligomers in the total loading concentration was derived using the  $c(s)$  integration function (Figure 7.4).

#### **(7.4.3) X-ray scattering data collection and analysis**

X-ray scattering data were obtained in two beam sessions in sixteen-bunch mode at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, operating with a ring energy of 6.0 GeV on the Beamline ID02 (Narayanan *et al.*, 2001). Storage ring currents ranged from 68 mA to 77 mA (session 1) and from 70 mA to 92 mA (session 2). Data were acquired using a recently installed fibre optically-coupled high sensitivity and dynamic range CCD detector (FReLoN). Together with a smaller beamstop, this allowed a shorter sample-to-detector distance of 2.0 m to be used and resulted in higher signal-to-noise ratios (Li *et al.*, 2008). CRP was studied at concentrations between 0.3 mg/ml and 5.4 mg/ml (with 2 mM Ca<sup>2+</sup>) or between 0.1 mg/ml and 2.5 mg/ml (without Ca<sup>2+</sup>). Samples of volume 100 µl were measured in flow cells to reduce radiation damage by moving the sample continuously during beam exposure. Sets of ten time frames, with frames of duration 0.2 sec, 0.3 sec or 0.5 sec each were acquired. Buffers were measured using the same exposure times in alternation with the samples to eliminate background subtraction errors. On-line checks during data acquisition confirmed the absence of radiation damage, after which the ten frames were averaged. Sample temperature corresponded to ambient conditions at 20°C. Other details including the data

reduction procedure are described elsewhere (Gilbert *et al.*, 2005).

In a given solute-solvent contrast, the radius of gyration  $R_G$  is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low  $Q$  gives the  $R_G$  and the forward scattering at zero angle  $I(0)$  (Glatter & Kratky, 1982):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3.$$

This expression is valid in a  $Q.R_G$  range up to 1.5. The  $R_G$  analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics O2 Workstations. Indirect transformation of the scattering data  $I(Q)$  in reciprocal space into real space to give the distance distribution function  $P(r)$  was carried out using the program GNOM (Semenyuk & Svergun, 1991):

$$P(r) = \frac{I}{2\pi^2} \int_0^\infty I(Q) Qr \sin(Qr) dQ$$

$P(r)$  corresponds to the distribution of distances  $r$  between volume elements. For this, the X-ray  $I(Q)$  curve utilised up to 562 data points in the  $Q$  range between  $0.05 \text{ nm}^{-1}$  and  $1.9 \text{ nm}^{-1}$ .

#### (7.4.4) Modelling of CRP

All the CRP modelling was based on the crystal structure for the pentamer in the presence of  $\text{Ca}^{2+}$  (PDB code 1b09) (Thompson *et al.*, 1999). The CRP decamer was formed by the superimposition of two pentamers face-to-face upon the  $\text{Ca}^{2+}$ -depleted decamer structure (PDB code 1lj7) (Ramadan *et al.*, 2002). The superimposition resulted in an r.m.s.d of  $0.460 \text{ nm}$  in  $\alpha$ -carbon atoms, and overcame the issue of missing residues in the  $\text{Ca}^{2+}$ -depleted crystal structure. Insight II 98.0 (Accelrys, San Diego, CA) software on Silicon Graphics workstations was used for manipulations. A cube side length of  $0.532 \text{ nm}$  in combination with a cutoff of 4 atoms was used to convert the coordinates into Debye sphere models with 998 and 1997 spheres that corresponded to the unhydrated pentamer and decamer structures respectively. The hydration shell was created using HYPRO, (Ashton *et al.*, 1997)

where the optimal total of hydrated spheres in the CRP models were 1310 (pentamer) and 2620 (decamer) (Table 7.1). The X-ray scattering curves  $I(Q)$  for CRP were calculated using the Debye equation adapted to spheres (Perkins & Weiss, 1983). Other details are given elsewhere (Perkins, 2001). The pentamer and decamer curves were merged in increments of 1% from 0% to 100% of the decamer for comparison with nine X-ray curves in 140 mM NaCl buffer containing 2 mM  $\text{Ca}^{2+}$ , and six (50 mM NaCl) or four (140 mM NaCl) curves in buffers without  $\text{Ca}^{2+}$ . Sedimentation coefficients were calculated from the X-ray hydrated sphere models using the HYDRO program (Garcia de la Torre *et al.*, 1994; Ashton *et al.*, 1997), and also from the crystal coordinates using HYDROPRO software (Garcia de la Torre *et al.*, 2000). The default value of 0.31 nm was used to represent the hydration shell. To generate a putative decamer with the B faces now in contact with each other, one pentamer in the decamer model was translated by 7.4 nm relative to the other along its central symmetry axis.

#### **(7.4.5) Surface plasmon resonance data**

The self-association of CRP was analysed by SPR using a Biacore X100 instrument and version 1.1 of its evaluation software (GE Healthcare, Uppsala, Sweden). Native CRP was coupled to the flow cell of a carboxylated dextran (CM5) research grade sensor chip via a standard amine coupling procedure according to the manufacturer's protocol. 10  $\mu\text{g/ml}$  of CRP in 10 mM acetate buffer (pH 4.3) in the presence of  $\text{Ca}^{2+}$  was injected over flowcell 2 until 150 response units (RU) was attained. A control surface cell was prepared identically on flow cell 1 but without protein immobilisation. Binding and equilibrium analyses using these CRP-immobilised chips were performed at 25° C in duplicate or triplicate runs using the appropriate Biacore X100 wizards at flow rates of 10-30  $\mu\text{l/min}$ . Regeneration after each run was achieved by pulsing 10 mM acetate buffer, 2 M NaCl (pH 4.6) across both flow cells once for 30 sec. Running buffer was either TBS (Tris-buffered saline, 10 mM Tris, 140 mM NaCl, 0.005% P20 surfactant, pH 8.0) or HBS (HEPES-buffered saline, 10 mM HEPES, 50 mM NaCl, 0.005% P20 surfactant, pH 8.0), with or without 2 mM  $\text{CaCl}_2$ .

**Table 7.1. X-ray scattering and sedimentation coefficient data and modelling of CRP**

		$R_G$ (nm)	$s_{20,w}^0$ (S)	$s_{20,w}^0$ (S)	$K_D$ ( $\mu$ M)
<b>Experimental buffer</b>			<b>Pentamer</b>	<b>Decamer</b> <sup>a</sup>	
50 mM NaCl, 2 mM Ca <sup>2+</sup>			6.5 $\pm$ 0.1	7.7 – 8.4	19
140 mM NaCl, 2 mM Ca <sup>2+</sup>		3.8 – 4.5; 3.8 – 4.3 <sup>b</sup>	6.4 $\pm$ 0.2	7.6 – 8.2	22 $\pm$ 5
250 mM NaCl, 2 mM Ca <sup>2+</sup>			6.4 $\pm$ 0.3	7.5 – 7.7	n.a.
50 mM NaCl buffer, no Ca <sup>2+</sup>		3.8 – 4.1; 3.7 – 4.1 <sup>b</sup>	6.4 $\pm$ 0.1	7.5 – 8.9	n.a.
140 mM NaCl buffer, no Ca <sup>2+</sup>		3.6 – 4.1; 3.6 – 4.0 <sup>b</sup>	6.3 $\pm$ 0.1	7.3 – 7.7	n.a.
<b>Scattering modelling</b>	<b>Spheres</b>		<b>R factor (%)</b>		
CRP pentamer	1256 <sup>c</sup>	3.7	6.0 – 6.2		
CRP decamer	2492 <sup>c</sup>	4.2		9.8	
Best-fit scattering curves (2 mM Ca <sup>2+</sup> )			2.0 - 3.7		23
Best-fit scattering curves (no Ca <sup>2+</sup> )			1.8 - 3.5; 1.7 - 2.8 <sup>d</sup>		

<sup>a</sup> The experimental value corresponds to the decamer reaction boundary (Figures 7.2; 7.3).

<sup>b</sup> The first value is from the Guinier  $R_G$  analyses (Figures 7.5(a,b)); the second is from the GNOM  $P(r)$  analyses (Figure 7.7).

<sup>c</sup> The optimum number of hydrated spheres predicted from the sequence are 1310 and 2620 for pentamer and decamer respectively.

<sup>d</sup> The first value is for 50 mM NaCl; the second is for 140 mM NaCl.

n.a. not available



## **Chapter Eight**

**Complement Factor H binds at two distinct sites to C-reactive protein in solution and on surfaces in an NaCl concentration-dependent manner**

## (8.1) Introduction

Mutations and polymorphisms in FH are associated with atypical haemolytic uraemic syndrome (aHUS) and age-related macular degeneration (AMD), suggesting that the impaired control of complement activation in the kidney endothelium and the retina respectively are involved in these diseases (Section 2.4). The distribution of over 100 genetic alterations in FH leading to AMD or aHUS are summarised at <http://www.fh-hus.org> (Saunders *et al.*, 2007). In physiological buffer, FH is principally monomeric and co-exists with as much as 15% of dimeric FH in a rapid equilibrium that is NaCl-concentration dependent; in addition FH forms less reversible higher oligomers at high FH concentrations (Nan *et al.*, 2008; Chapter 5).

Native CRP is unglycosylated and is a stable pentamer (Figure 8.1) formed by non-covalently-linked protomers whose assembly is stabilised by 2.5 mM  $\text{Ca}^{2+}$  present in plasma (Hurwitz, 1996; Thompson *et al.*, 1999; Pepys & Hirschfield, 2003; Chapter 7). In physiological buffer, CRP forms a rapid NaCl-concentration dependent pentamer-decamer equilibrium (Chapter 7). CRP binds phosphocholine at its  $\text{Ca}^{2+}$  binding site (Thompson *et al.*, 1999), and to phosphoethanolamine, microbial surface proteins, chromatin, histones, fibronectin, small nuclear ribonucleoproteins, laminin and polycations (Black *et al.*, 2004; Section 2.3.3). These CRP-ligand interactions recognise damaged or apoptotic cells and bacterial pathogens (Black *et al.*, 2004; Thompson *et al.*, 1999). CRP activates complement by the classical pathway (Kaplan & Volanakis, 1974), but inhibits alternative pathway activation by upregulating complement regulators and inhibitors at host cell surfaces (Mold *et al.*, 1999; Li *et al.*, 2004). A common Tyr402His FH polymorphism is an established risk factor for AMD, however the molecular mechanism of the polymorphism in causing AMD remains unclear (Saunders *et al.*, 2007). Individuals who are homozygous for the AMD-risk His402 FH allotype show a 2.5-fold higher level of CRP in the retinal pigmented epithelium (Johnson *et al.*, 2006).

A FH-CRP interaction would explain how CRP inhibits complement activation. Large basic surfaces exist on FH at SCR-7, SCR-13 and SCR-20 that may interact with acidic surfaces on CRP (Figure 8.1). However studies with CRP are readily affected by the ease with which CRP is denatured by the removal of  $\text{Ca}^{2+}$ , given that  $\text{Ca}^{2+}$  binds to CRP with a moderate dissociation constant  $K_D$  of 60  $\mu\text{M}$  (Kinoshita *et al.*, 1989). The assembly of the CRP pentamer is sensitive to small conformational perturbations such as those caused by  $\text{Ca}^{2+}$  removal (Chapter 7). The

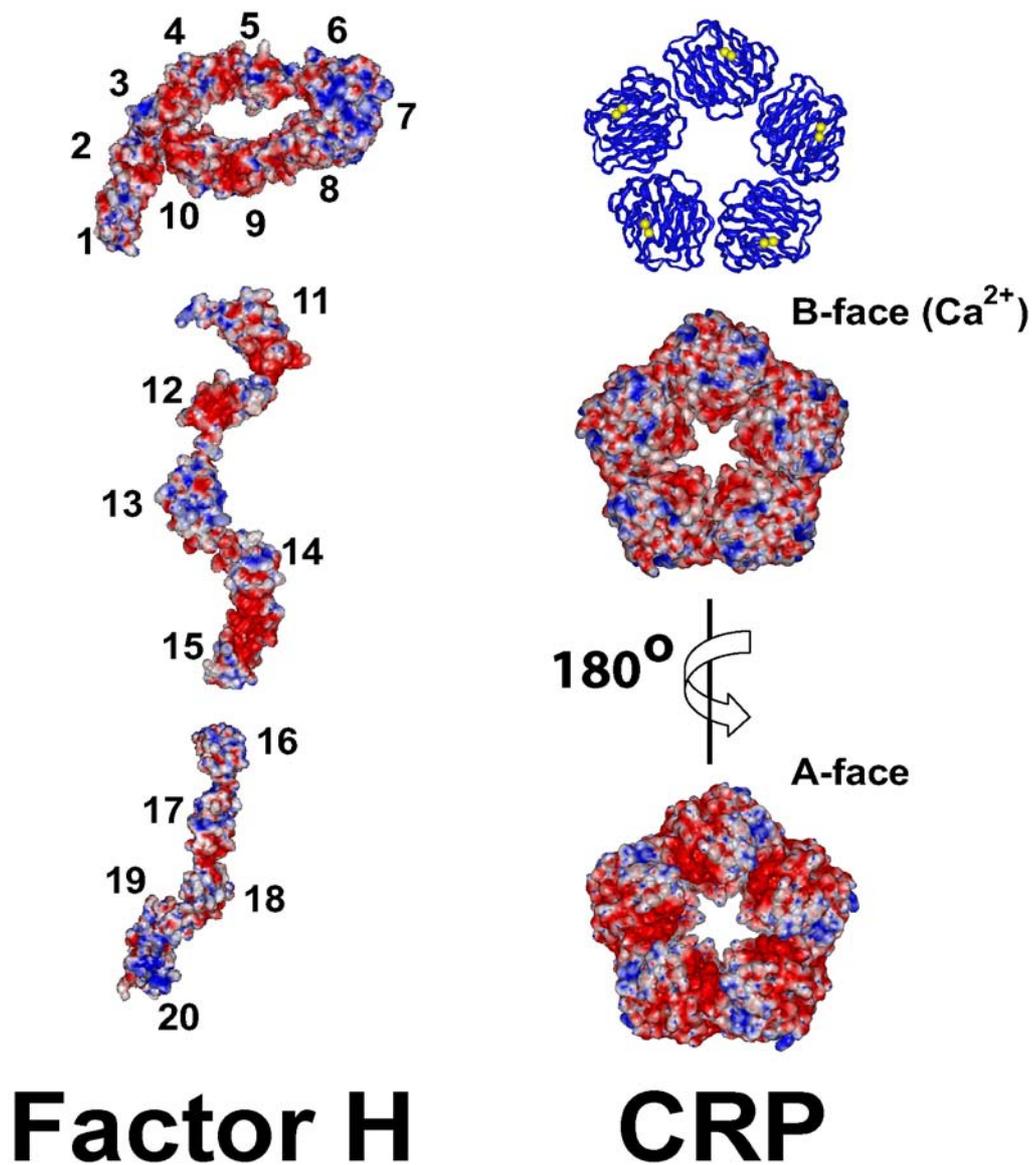
abnormally modified forms of CRP created by  $\text{Ca}^{2+}$  removal have very different functional properties from the highly stable CRP pentamer, and there is no firm evidence for the persistence of any denatured CRP forms *in vivo* (Pepys & Hirschfield, 2003). The first evidence for a FH-CRP interaction was provided by an ELISA assay of FH binding to plastic-immobilised CRP, and SPR analyses showing that FH SCR-7 and SCR-8/11 bind to immobilised CRP (Mold *et al.*, 1999; Jarva *et al.*, 1999). FH binding to immobilised CRP has since then been demonstrated repeatedly, in all of which the Tyr402 allotype binds more strongly to CRP than the His402 allotype (Laine *et al.* 2007; Sjöberg *et al.*, 2007; Skerka *et al.*, 2007; Yu *et al.*, 2007; Herbert *et al.*, 2007). The stronger binding of FH Tyr402 to CRP would be more effective in protecting CRP-decorated host cells, and this may contribute to the molecular mechanism of AMD. However recent ELISA and AUC analyses suggested instead that the observed FH-CRP interactions result from the use of denatured CRP in earlier studies (Bíró *et al.*, 2007; Hakobyan *et al.*, 2008).

In this thesis, the conditions in which the molecular interaction between FH and CRP takes place in solution and on surfaces are defined using a combination of AUC, SPR and X-ray scattering (Cole *et al.*, 2008; Perkins *et al.*, 2008; Tanious *et al.*, 2008). The discrepancy between the studies that report or refute the FH-CRP interaction was re-examined in physiological buffer conditions and protein concentrations, being careful to work in conditions when CRP forms its native stable pentamer. It is shown here that FH binds to CRP in both the solution and surface phases in a NaCl-concentration dependent manner with  $\text{Ca}^{2+}$  present. A previously-unknown second CRP binding site is identified at the C-terminus of FH. These results provide further insight into the multivalent interaction of CRP with ligands at the host cell surface and the involvement of FH in AMD and aHUS diseases.

## **(8.2) Results and Discussion**

### **(8.2.1) Sedimentation velocity studies of the Factor H-CRP complex**

AUC studies macromolecular structures in solution by following their sedimentation behaviour under a high centrifugal force (Cole *et al.*, 2008). The observed sedimentation boundaries are fitted to yield a sedimentation coefficient distribution  $c(s)$  according to the Lamm equation using SEDFIT (Section 8.4.2). This takes into account all the species present in the sample, although the algorithm assumes that the same frictional ratio  $f/f_0$  applies for all the sedimenting species



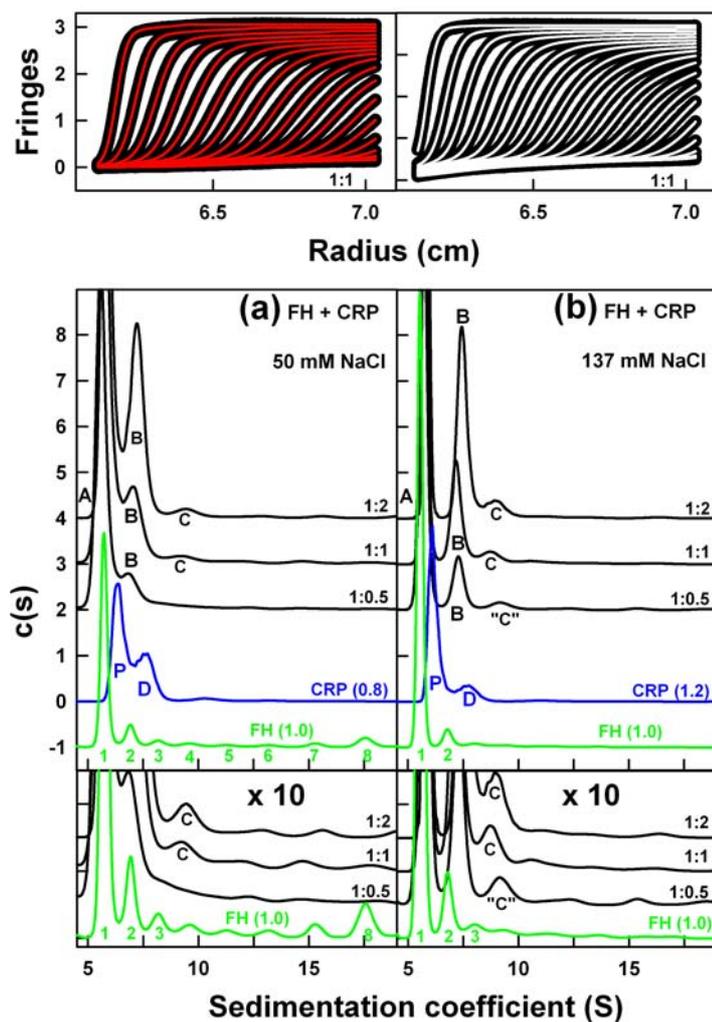
**Figure 8. 1.**

Electrostatic views of FH and CRP. Red denotes acidic regions while blue denotes basic regions. Both FH and CRP are drawn to the same scale. The solution structural model for FH in 137 mM NaCl (PDB code 3gav) is shown as three fragments which are each rotated individually in order to display the large basic surfaces at SCR-7, SCR-13 and SCR-20. All 20 SCR domains are numbered. The front and back views of the crystal structure of pentameric CRP (PDB code 1b09) are shown as the A-face and B-face. The locations of calcium on the B-face are shown in yellow in a ribbon view of CRP. The basic surfaces of FH may interact with a preponderance of acidic surfaces on CRP when this is bound to cell surfaces.

(where  $f_o$  is the frictional coefficient of the sphere with the same volume as the hydrated glycoprotein). Macromolecular elongation is monitored through the sedimentation coefficient  $s_{20,w}^0$  values. For a non-interacting system, sample polydispersity is determined from the number of peaks observed in the  $c(s)$  plot. For an interacting system, the  $c(s)$  peaks may correspond instead to reaction boundaries and not discrete species if rapid exchange occurs during sedimentation. Previous studies by size exclusion chromatography and AUC sedimentation equilibrium in 146 mM NaCl buffer reported that FH does not interact with CRP in solution (Table 8.1) (Hakobyan *et al.*, 2008). However, FH and CRP are reported to interact in 49 mM NaCl buffer by SPR (Jarva *et al.*, 1999). To investigate this discrepancy, sedimentation velocity experiments were performed with FH and CRP mixed in molar ratios of 1:0.5, 1:1 and 1:2, in both 50 mM NaCl and 137 mM NaCl buffers. These correspond to concentrations of 0.1 mg/ml to 1.1 mg/ml (FH) and 0.1 mg/ml to 1.4 mg/ml (CRP). The  $c(s)$  analyses of interference optics data resulted in good boundary fits (Figure 8.2).

AUC velocity experiments on unbound FH and CRP show that both proteins form multimers (Fernando *et al.*, 2007; Nan *et al.*, 2008a; Chapter 5; Chapter 7). Unbound FH forms dimers in a concentration-dependent manner in at least two different sites within SCR-6/8 and SCR-16/20, and the combination of both sites lead to FH dimers and higher FH oligomers in both 50 mM NaCl and 137 mM NaCl. By AUC, these correspond respectively to peak 2 and peaks 3-8 at the bottom of Figure 8.2(a,b). CRP forms pentamers and decamers in a concentration-dependent manner, and these correspond to peaks P and D at the bottom of Figure 8.2(a,b). The absence of  $\text{Ca}^{2+}$  resulted in dissociation of the pentamers into smaller protomers that are visible at low  $s$  values (Chapter 7).

For 50 mM NaCl, AUC studies of nine FH-CRP mixtures revealed complex formation in 2 mM  $\text{Ca}^{2+}$ . In 50 mM NaCl, the boundary fits for  $c(s)$  plots for FH and CRP mixtures were calculated at concentrations of 0.3-1.1 mg/ml of FH (three concentrations) and 0.1-1.4 mg/ml CRP (nine concentrations). These maintained three FH:CRP ratios of 1:0.5, 1:1 and 1:2. The three representative  $c(s)$  plots shown in Figure 8.2(a) consistently showed at least two major peaks denoted A and B in all nine mixtures, together with a weak peak C. Peak A showed an average  $s_{20,w}^0$  value of  $5.7 \pm 0.1$  S similar to that of  $5.7 \pm 0.1$  S for unbound FH in 50 mM NaCl buffer (Chapter 7). Compared to peak 1 for unbound FH monomer (Figure 8.2(a)), the width of peak A



**Figure 8.2.**

AUC analysis of FH-CRP mixtures in (a) 50 mM and (b) 137 mM NaCl. The  $c(s)$  size distribution analyses correspond to sedimentation velocity experiments performed at a rotor speed of 50,000 r.p.m. in which CRP was titrated against 1.0 mg/ml FH at CRP:FH molar ratios of 1:0.5, 1:1 and 1:2 (total concentrations of 1.5 mg/ml, 1.8 mg/ml and 2.5 mg/ml from bottom to top). Representative boundary fits showing every tenth scan are shown in the two upper panels for the 1: 1 mixtures in red (50 mM NaCl) and black (137 mM NaCl). The plots are displaced vertically on the  $c(s)$ -axis and the peak intensities are normalised for clarity. For comparison, the  $c(s)$  analyses for unbound FH (green) and CRP (blue) are shown at the bottom. The insets at the bottom show the peaks at larger  $s$  values in more detail. For FH, peaks 1, 2 and 3-8 have been assigned to monomeric, dimeric and oligomeric FH species in order of increasing  $s$  value. For CRP, peak P corresponds to pentameric CRP and peak D corresponds to the reaction boundary of decameric CRP (Chapter 7). In the FH-CRP mixtures, peaks A, B and C correspond to peaks observed at different positions from those seen for unbound FH and CRP that correspond to either FH-CRP complexes or their reaction boundaries.

was broadened and its intensity increased with CRP concentration at a fixed FH concentration. No separate peak P for pentameric CRP at a  $s_{20,w}^0$  value of  $6.4 \pm 0.1$  S was observed. Both features are consistent with an association between two reacting components (Brown & Schuck, 2006). Thus peak A is attributed to an average of monomeric FH and pentameric CRP that interact in a rapid equilibrium (Figure 8.2(a)). Peak B showed  $s$  values in a range between 6.7-7.3 S, where the  $s$  value increased with increasing CRP concentration, and this indicated a reaction boundary (Dam & Schuck, 2005). This deduction is consistent with the lack of baseline resolution between peaks A and B (Chapter 7). The  $s$  value of peak B is greater than that for the CRP pentamer at 6.4 S, indicating that this peak corresponds to a 1:1 complex between FH and CRP of increased mass. A third peak C corresponding to another sedimentation species was observed at the highest CRP concentrations (Figure 8.2(a)). The intensity of peak C depends on the CRP concentration. The mean  $s$  value of peak C at  $9.2 \pm 0.4$  S was higher than the experimentally observed range of  $s$  values of 7.7-8.4 S for decameric CRP. This increase in  $s$  value is consistent with higher complex formation, thus this is most likely to correspond to a FH-CRP<sub>2</sub> stoichiometry.

For 50 mM NaCl buffer containing 2 mM Ca<sup>2+</sup>, the above assignments of peak A, B and C correspond to the simplest interpretation of the  $c(s)$  plots. This is in terms of unbound FH monomer and CRP pentamer and reaction boundaries for the FH-CRP and the FH-CRP<sub>2</sub> or FH<sub>2</sub>-CRP complexes. Other interpretations cannot be ruled out. To complete the analysis of peaks A, B and C, the  $c(s)$  distributions were firstly converted to molecular weight distributions  $c(M)$ . The  $c(M)$  distributions provide mass information, provided that the  $c(s)$  peaks correspond to non-interacting sedimentation species and the individual components have similar frictional ratios  $f/f_0$ . The  $c(M)$  plots using a mean  $f/f_0$  ratio of  $1.5 \pm 0.2$  showed that peaks A, B and C corresponded to apparent molecular weights of  $122 \pm 16$  kDa,  $169 \pm 20$  kDa and  $239 \pm 22$  kDa respectively. That for peak A is comparable to a mean value of 136 kDa for an equimolar mixture of unbound monomeric FH and pentameric CRP. Those for peaks B and C are lower than those predicted for a FH-CRP complex (269 kDa) and a FH-CRP<sub>2</sub> complex (384 kDa) or even a FH<sub>2</sub>-CRP complex (423 kDa), supporting the deduction that peaks B and C correspond to reaction boundaries. As expected from the reaction boundaries and the very different  $f/f_0$  values of 1.75 for monomeric FH and 1.12 for CRP (Nan *et al.*, 2008a; Chapter 5; Chapter 7), the  $c(M)$  distribution resulted in erroneous molecular weights, although good estimates may be obtained

for the most abundant species (Dam & Schuck, 2005; Brown & Schuck, 2006). Secondly, the  $c(s)$  peak intensities were examined in order to assess the degree of FH oligomerisation and CRP decamer formation. The insets at the bottom of Figure 8.2(a) showed that the intensities of FH oligomer peaks 3-8 were much reduced in the mixtures. The interaction between FH and CRP has clearly suppressed the formation of higher FH oligomers. It is not clear whether FH dimers or CRP decamers still form. Comparison of the ratio of peak sizes of peaks B and C with the ratio of peak sizes for peaks P and D for CRP and peaks 1 and 2 for FH suggests that peak C is smaller in size relative to peak D and similar to peak 2. This would be consistent with the assignment of peak C as a FH-CRP<sub>2</sub> complex, although AUC by itself is unable to distinguish between the possibilities of a FH<sub>2</sub>-CRP or FH-CRP<sub>2</sub> stoichiometry for peak C.

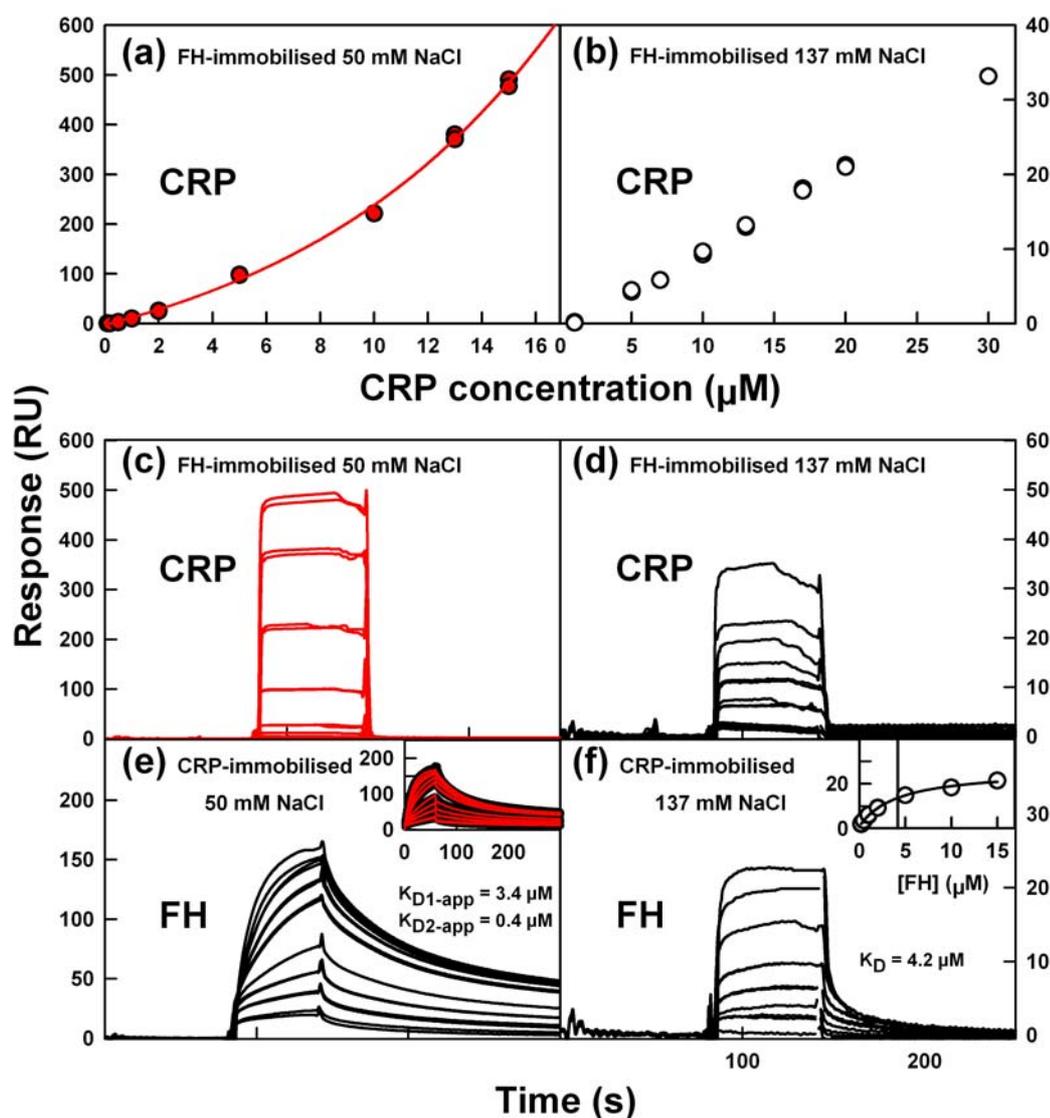
For 137 mM NaCl containing 2 mM Ca<sup>2+</sup>, the corresponding analyses of peaks A, B and C in further velocity experiments showed that the FH-CRP interaction displayed an ionic strength dependence. This is in accord with Figure 8.1. Ten experiments were performed using FH:CRP molar ratios of 1:0.5, 1:1 and 1:2 for FH concentrations between 0.1-1.1 mg/ml and CRP concentrations between 0.2-1.3 mg/ml, of which three experiments are shown in Figure 8.2(b). The  $c(s)$  plots again revealed three peaks A, B and C at all concentrations. Peak A with a  $s_{20,w}^0$  value of  $5.8 \pm 0.1$  S corresponded closely to peak A in 50 mM NaCl buffer with a  $s_{20,w}^0$  value of  $5.7 \pm 0.1$  S (Figure 8.2(a)) and is again attributed as the average of unbound monomeric FH and pentameric CRP interacting with each other. Peak A was less broadened compared to that observed in 50 mM NaCl. The  $c(M)$  plot yielded a molecular weight of  $127 \pm 17$  kDa, which is similar to that of  $122 \pm 16$  kDa in 50 mM NaCl buffer. The  $s$  values for peak B ranged between 6.8-7.5 S in a concentration-dependent manner, which is similar to those of 6.7-7.3 S for peak B in 50 mM NaCl buffer. Peak B in 137 mM NaCl exhibited two important differences from the 50 mM NaCl analyses. Firstly, the baseline resolution between peaks A and B is improved, indicating a reduced association between FH and CRP. Secondly, the width of peaks B and A is also reduced compared to that in 137 mM NaCl. These two features show that complex formation between FH and CRP is reduced in 137 mM NaCl buffer. A third minor peak C was visible at a higher  $s$  value. In contrast to peak C in 50 mM NaCl, peak C in 137 mM NaCl appeared at all concentrations and showed no concentration dependence. Its mean  $s$  value of  $8.8 \pm 0.2$  S was lower than that of  $9.2 \pm 0.4$  S for peak C in 50 mM NaCl.

The most likely assignment of peak C is that at low CRP concentration this corresponds to peak 3 of unbound FH (and is labelled “C” for this reason in [Figure 8.2\(b\)](#)), and at higher CRP concentrations this becomes peak C that is attributed to the FH-CRP<sub>2</sub> complex. In conclusion, both experiments in 50 mM NaCl and 137 mM NaCl buffers show that native FH and CRP interact in solution in an ionic strength-dependent manner.

### **(8.2.2) Surface plasmon resonance studies of the Factor H-CRP complex**

SPR monitors the interaction between a binding partner in solution (analyte) and an immobilised partner (ligand) attached to the surface of a sensor chip ([Tanious \*et al.\*, 2008](#)). Previous SPR studies of the FH-CRP interaction have been questioned on the grounds of the possible denaturation of CRP when immobilised ([Jarva \*et al.\*, 1999](#); [Hakobyan \*et al.\*, 2008](#)). Here, we show that the CRP immobilised on the sensor chip remains active from the similar  $K_D$  value of 22  $\mu\text{M}$  for its pentamer-decamer self-association when studied either in solution or by SPR, both in the presence of 2 mM  $\text{Ca}^{2+}$  ([Chapter 7](#)). Accordingly, the FH-CRP interaction was studied by SPR using either immobilised FH or immobilised CRP in duplicate runs using 50 mM NaCl and 137 mM NaCl buffers containing 2 mM  $\text{Ca}^{2+}$  ([Section 8.4.3](#)).

Native CRP in solution binds to immobilised FH in both 50 mM NaCl and 137 mM NaCl in the presence of 2 mM  $\text{Ca}^{2+}$  ([Figures 8.3\(a-d\)](#)), in agreement with the AUC data. CRP was flowed over a CM5 sensor chip onto which 150 RU of FH had been immobilised ([Section 8.4.3](#)). The sensorgrams show that the on- and off-rates are rapid ([Figures 8.3\(c,d\)](#)). In 50 mM NaCl, the SPR equilibrium studies using 0.1-15  $\mu\text{M}$  CRP showed that CRP aggregated onto immobilised FH ([Figure 8.3\(a\)](#)). This was evidenced by the exponential increase in binding response up to 500 RU, which is five times the predicted maximum binding response ( $R_{\text{max}}$ ) of 112 RU. This large increase is similar to the aggregation of CRP in the same buffer binding to immobilised CRP ([Chapter 7](#)). Here, CRP first binds to immobilised FH, then the FH-bound CRP recruits more CRP. CRP binding was much weakened on progressing from 50 mM NaCl to 137 mM NaCl, in which the binding response of 10  $\mu\text{M}$  CRP decreased 25-fold from 221 RU to 9 RU ([Figure 8.3\(b\)](#)). In 137 mM NaCl, an equilibrium  $K_D$  value could not be determined for 0.1-30  $\mu\text{M}$  CRP because



**Figure 8.3.**

Surface plasmon resonance analysis of the FH-CRP interaction. In plots (a-d), FH was immobilised on the chip surface. In plots (e,f), CRP was immobilised.

(a,b) Equilibrium  $K_D$  analyses for CRP binding to immobilised FH in 50 mM and 137 mM NaCl. In 50 mM NaCl, the binding response increased exponentially to well above the theoretical maximum response of 112 RU at CRP concentrations between 0-15  $\mu\text{M}$ . In 137 mM NaCl, saturation was not attained at CRP concentrations between 0-30  $\mu\text{M}$ .

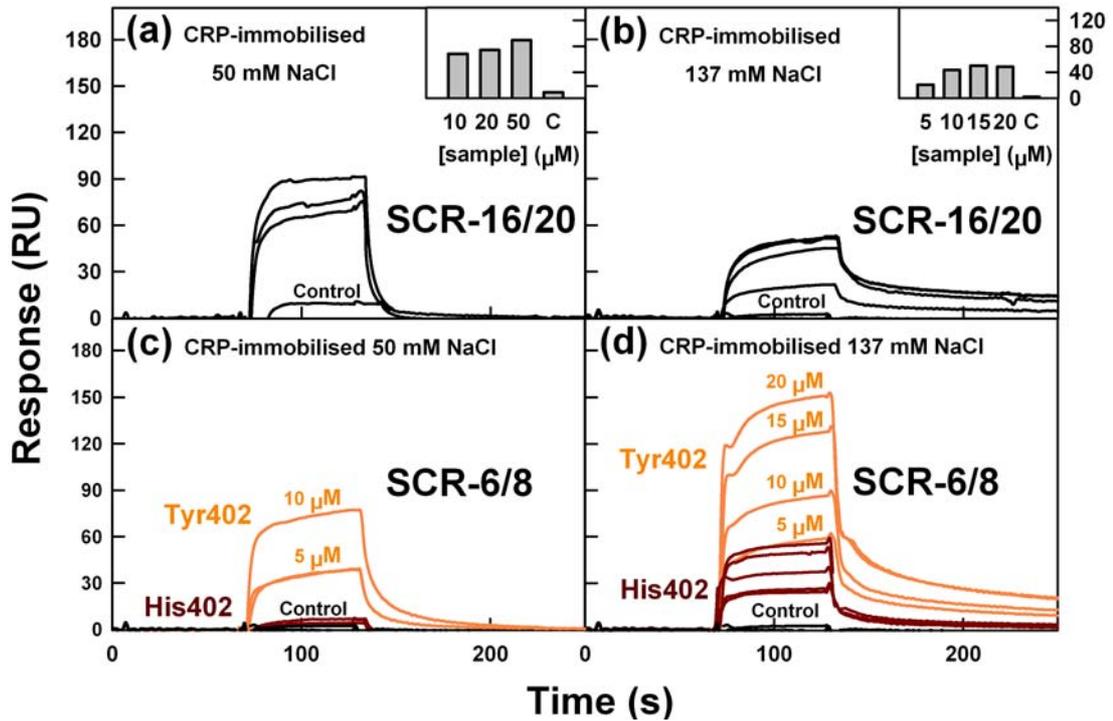
(c,d) Sensorgrams for the equilibrium studies reported in (a,b). The individual responses correspond incrementally to the individual data points shown in (a,b).

(e,f) Equilibrium  $K_D$  analyses for FH binding to immobilised CRP in 50 mM and 137 mM NaCl. In 50 mM NaCl buffer, the binding response approaches the theoretical maximum response of 201 RU at FH concentrations between 0-15  $\mu\text{M}$ . The apparent  $K_D$  value was estimated by fitting to a heterogeneous ligand kinetic model (inset). In 137 mM NaCl, the  $K_D$  value was determined at CRP concentrations between 0-15  $\mu\text{M}$  (inset).

of the non-saturation of CRP binding. The outcome of this experiment is attributed to a combination of the CRP interaction with immobilised FH and the pentamer-decamer equilibrium of CRP.

In the reversed SPR equilibrium experiment, native FH in solution binds to immobilised CRP (Figures 8.3(e,f)). 150 RU of CRP was immobilised on the sensor chip (Chapter 7). In 50 mM NaCl, a much higher response of up to 160 RU was observed compared to 23 RU in 137 mM NaCl. In 50 mM NaCl, steady state binding equilibria were not achieved for the majority of runs for reason of slower association-dissociation rates (Figure 8.3(e)). Hence affinity analyses could not be carried out. In such cases, a  $K_D$  value is determined from the kinetic rate constants for association and dissociation, provided the fitted model describes the experimental data. A simple 1:1 kinetic model did not give a good fit to the experimental data. A much improved fit was obtained with alternative models, such as for a heterogeneous ligand model (Figure 8.3(e), inset; Section 8.4.3). However, this good-fit model did not completely describe the experimental data, and the resulting  $K_D$  values are estimates only. Two apparent equilibrium constants  $K_{D1-app}$  and  $K_{D2-app}$  of 3.4  $\mu\text{M}$  and 0.4  $\mu\text{M}$  were obtained. The complexity of this analysis would be consistent with FH binding at either one of two different sites to one immobilised CRP molecule, or two different sites on FH interact to two immobilised CRP molecules, or the common Tyr402 and His402 FH allotypes in native FH exhibit different CRP binding properties. In 137 mM NaCl, the FH-CRP interaction proceeded to steady state equilibrium (Figure 8.3(f)), and an equilibrium  $K_D$  value of 4.2  $\mu\text{M}$  was determined (Figure 8.3(e), inset). This analysis was well represented by a 1:1 interaction. Thus CRP binds to FH in a NaCl concentration-dependent manner, in agreement with the AUC data.

SCR-7 has the second-highest basic charge density in FH, surpassed only by SCR-20, thus either or both of SCR-7 and SCR-20 may interact with the cationic CRP molecule ( $pI = 5.28$ ) (Figure 8.1) (Chapter 7). Previously by SPR, the binding of CRP to FH in 49 mM NaCl was localised to SCR-7 and SCR-8/11 (Table 8.1) (Jarva *et al.*, 1999). To investigate further the FH binding sites on CRP, SPR binding analyses were carried out using three FH fragments that represent the main functional activities of FH, namely SCR-1/5, SCR-6/8 and SCR-16/20. These were flowed over immobilised CRP at concentrations between 5  $\mu\text{M}$  to 50  $\mu\text{M}$  in 50 mM and 137 mM NaCl buffers containing 2 mM  $\text{Ca}^{2+}$ . Binding of SCR-16/20 to CRP



**Figure 8.4.**

Surface plasmon resonance analyses of FH fragments with immobilised CRP.

(a,b) SCR-16/20 binding in 50 mM and 137 mM NaCl buffer. Here, 10  $\mu\text{M}$ , 20  $\mu\text{M}$  and 50  $\mu\text{M}$  of SCR-16/20 bind to CRP, while 8  $\mu\text{M}$  of C3u was used as a non-binding control.

(c) The Tyr402 and His402 allotypes of SCR-6/8 binding in 50 mM and 137 mM NaCl buffer. Here, 5  $\mu\text{M}$ , 10  $\mu\text{M}$ , 15  $\mu\text{M}$  and 20  $\mu\text{M}$  of the SCR-6/8 allotypes bind differentially to CRP, while 5  $\mu\text{M}$  of SCR-1/5 was used as a non-binding control.

was observed at all three concentrations (10  $\mu\text{M}$ , 20  $\mu\text{M}$  and 50  $\mu\text{M}$ ), while the use of 8  $\mu\text{M}$  C3u as a non-binding control protein showed no binding to CRP (Figure 8.4(a,b)). The binding of SCR-16/20 at four concentrations of 5  $\mu\text{M}$ , 10  $\mu\text{M}$ , 15  $\mu\text{M}$  and 20  $\mu\text{M}$  was consistently weaker in 137 mM NaCl, showing that this interaction is electrostatic. It is concluded that a second binding site for CRP exists in FH within SCR-16/20. The Tyr402 and His402 allotypes of SCR-6/8 bound differentially to immobilised CRP (Figure 8.4(c,d)). Between concentrations of 5  $\mu\text{M}$  to 20  $\mu\text{M}$ , the Tyr402 allotype binds more strongly than the His402 allotype in both 50 mM and 137 mM NaCl. Interestingly, the Tyr402 allotype binds similarly in 50 mM and 137 mM NaCl, while the His402 allotype binding is mostly absent in 50 mM NaCl. This shows that this interaction is driven by hydrophobic forces, in which the presence of the positive charge on His402 in low salt is sufficient to abolish CRP binding. Here, the use of SCR-1/5 as a control showed no binding.

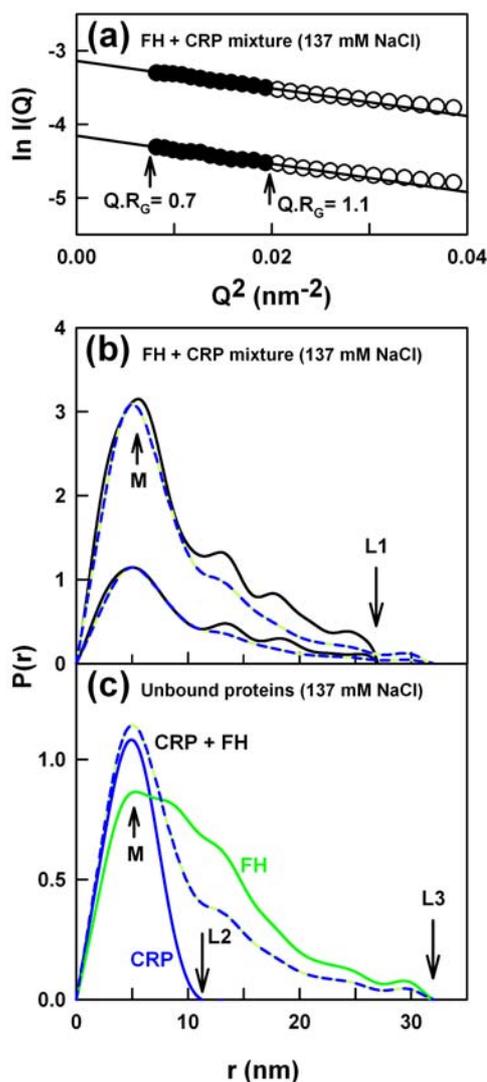
To assess further the FH-CRP interaction, stoichiometric ratios (SR) for the FH-CRP interaction were estimated from the intensity of the binding signals at the highest protein concentrations. The SR is the ratio between the observed maximum binding response and the calculated maximum binding response  $R_{\text{max}}$ .  $R_{\text{max}}$  is given by the molecular weight ratio of FH analyte to CRP ligand multiplied by the stoichiometry of 1 and the amount of immobilised CRP ligand (150 RU). The values of  $R_{\text{max}}$  are 201 RU, 55 RU and 28 RU for FH, SCR-16/20 and SCR-6/8 (both allotypes) respectively. At an FH concentration of 15  $\mu\text{M}$  in 50 mM NaCl buffer containing 2 mM  $\text{Ca}^{2+}$ , the binding response was 160 RU. The SR value was determined as 0.8, which indicated a 1:1 FH-CRP stoichiometry. The observed FH binding saturation at lower signal responses in 137 mM NaCl buffer (approximately 22 RU in Figure 8.3(f)) is attributable to a weakened FH-CRP interaction as a result of increased salt concentration. For SCR-16/20 (Figure 8.4(a,b)), the maximum measured binding responses were 91 RU (50 mM NaCl) and 53 RU (137 mM NaCl), and the corresponding SR values were 1.7 (2:1) and 1.0 (1:1) in that order. SCR-6/8 in 137 mM NaCl buffer containing 2 mM  $\text{Ca}^{2+}$  exhibited a marked difference between Tyr402 and His402 allotypes, with maximum binding responses of 152 RU and 55 RU respectively (Figure 8.4(d)). These corresponded to SR values of 5.5 (5:1) and 2.0 (2:1) in that order. In conclusion, although CRP can bind either of SCR-16/20 or SCR-6/8 with stoichiometries greater than 1:1, a 1:1 ratio is favoured in its

physiologically-relevant interaction with FH.

### (8.2.3) X-ray scattering studies of the FH-CRP complex

X-ray scattering is a diffraction technique performed in solution that studies the overall structure of biological macromolecules in random orientations (Perkins *et al.*, 2008). FH-CRP mixtures were studied in both 50 mM and 137 mM NaCl buffers. X-ray data collection in 50 mM NaCl was not successful for reason of the occurrence of sufficient minor protein aggregation that precluded linear Guinier  $R_G$  analyses. Accordingly the FH-CRP interaction was analysed in 137 mM NaCl in the presence of 2 mM  $\text{Ca}^{2+}$ . A 1:1 mixture of FH and CRP was studied in two dilutions with total protein concentrations of 1.78 mg/ml and 0.71 mg/ml in each (Section 8.4.4). Guinier analyses of the scattering data  $I(Q)$  at low  $Q$  values ( $Q = 4 \pi \sin \theta / \lambda$ ;  $2\theta =$  scattering angle;  $\lambda =$  wavelength) gave linear  $R_G$  fits within appropriate  $Q.R_G$  limits (Section 8.4.4), showing that the samples displayed good monodispersity (Figure 8.5(a)). The mean  $R_G$  value from three measurements was unchanged at both concentrations, with values of  $7.6 \pm 0.1$  nm (1.78 mg/ml total protein) and  $7.9 \pm 0.3$  nm (0.71 mg/ml total protein). If no complex formation occurs, the  $R_G$  value of the mixture is estimated to be  $7.5 \pm 0.3$  nm (Section 8.4.4). This  $R_G$  value is the weighted mean of assumed  $R_G$  values of 8.9 nm for FH, and 3.7 nm and 4.2 nm for the CRP pentamer (80%) and decamer (20%) respectively (Perkins *et al.*, 1985; Nan *et al.*, 2008a; Chapter 5; Chapter 7). The observed and estimated  $R_G$  values are therefore similar to each other in conditions when FH and CRP interact with each other weakly in 137 mM NaCl.

The distance distribution  $P(r)$  curve is calculated from the full  $I(Q)$  curve, and represents all the distances between pairs of atoms within the macromolecule of interest. The  $P(r)$  curve provides an independent verification of Guinier  $R_G$  values, and leads to maximum lengths  $L$  following an assumption of the maximum dimension  $D_{max}$ . The  $P(r)$  curves from Figure 8.5(b) yielded a mean  $R_G$  value of  $8.0 \pm 0.2$  nm. Since this agrees well with the Guinier  $R_G$  value, this shows that the  $P(r)$  curves refer to the macromolecules being studied. The maximum length  $L$  was determined from the value of  $r$  when  $P(r) = 0$ . The  $L1$  value for both  $P(r)$  curves was 27 nm. This is reduced compared to the  $L3$  value of 32 nm for unbound FH in 137 mM NaCl, while it remains greater than the  $L2$  value of 12 nm for unbound CRP pentamer and decamer (Figure 8.5(c)) (Nan *et al.*, 2008a; Chapter 5; Chapter 7). This



**Figure 8.5.**

X-ray scattering analyses for the FH-CRP interaction in 137 mM NaCl. The data correspond to two 1:1 mixtures of FH and CRP.

(a) Guinier  $R_G$  analyses for a total protein concentration of 1.78 mg/ml (upper) and 0.71 mg/ml (lower). The filled circles correspond to the data points that were fitted to obtain  $R_G$  values and the straight lines correspond to the best fit through these points using a  $Q$  range of 0.09-0.14  $\text{nm}^{-1}$ . The  $Q.R_G$  fit range is arrowed. (b) The corresponding distance distribution  $P(r)$  analyses are shown for the two 1:1 mixtures. The total protein concentrations were 1.78 mg/ml (upper) and 0.71 mg/ml (lower). The peak maximum  $M$  is arrowed and the maximum length is denoted as  $L1$ . For comparison, the  $P(r)$  curve is compared with the sum of those for unbound FH and CRP (blue-green dashed line). (c) The  $P(r)$  curve for unbound FH and CRP are shown for comparison. The  $P(r)$  curve for FH at 0.71 mg/ml is shown in green, and that for CRP at 0.31 mg/ml is shown in blue. The sum of the  $P(r)$  curves corresponds to a total concentration of 0.71 mg/ml, and is shown as the blue-green dashed line for the comparisons shown in (b).

reduction in the  $L$  value of the FH-CRP mixture is consistent with the presence of more compact structures in this mixture, even though the  $R_G$  value is similar or slightly increased, i.e. that the length of the FH domain structure is reduced in the presence of CRP. The maximum  $M$  corresponds to the most frequently occurring interatomic distance within the protein structure. A pronounced maximum was observed at  $r = 5.6$  nm for the FH-CRP mixture (Figure 8.4(b)). The corresponding  $M$  values were 4.9 nm for unbound FH, and 3.1 nm and 5.7 nm for unbound pentameric and decameric CRP respectively (Figure 8.4(c)). Thus  $M$  for the FH-CRP mixture is close to the mean  $M$  values of unbound FH and CRP. In contrast to this, the comparison of the two observed  $P(r)$  curves for the FH-CRP mixture with the sum of the  $P(r)$  curves for unbound FH and CRP mixture showed increased intensities between  $r$  values of 10 nm to 25 nm (Figure 8.4(b)). This intensity difference indicates the presence of more compact structures in the FH-CRP mixture. In conclusion, the  $P(r)$  curves show evidence of weak FH-CRP complex formation and the FH-CRP complex may have a more compact structure compared to unbound FH and CRP.

### (8.3) Conclusions

Here, it is established rigorously from a combination of AUC, SPR and X-ray scattering that wild type FH interacts with native CRP in both the fluid and surface phases in conditions when CRP forms stable native pentamers in the presence of  $\text{Ca}^{2+}$ . FH exists in plasma at concentrations between 0.235 and 0.810 mg/ml (1.6  $\mu\text{M}$  to 5.4  $\mu\text{M}$ ) (Saunders *et al.*, 2007). CRP is an acute phase protein, with concentrations that increase from less than 50 ng/ml up to 0.5 mg/ml (0.4 nM to 4.4  $\mu\text{M}$ ) during acute phase responses (Pepys & Hirschfield, 2003). These studies were carried out at FH concentrations of 0.1-1.1 mg/ml and CRP concentrations of 0.3-1.4 mg/ml in physiological buffer containing 137 mM NaCl in the presence of 2 mM  $\text{Ca}^{2+}$ , given that 2.5 mM  $\text{Ca}^{2+}$  is present in plasma (Hurwitz, 1996). Thus these results are of physiological relevance. The overall interaction is ionic strength-dependent, being stronger with decrease in the NaCl concentration to 50 mM. The binding of FH to CRP suppresses the self-association of FH to form dimers and higher oligomers, and the formation of CRP decamer. Because FH self-associates in at least two sites localised within SCR-6/8 and SCR-16/20 (Figure 8.6(a)), it was concluded that CRP interacts with distinct sites in both SCR-6/8 and SCR-16/20.

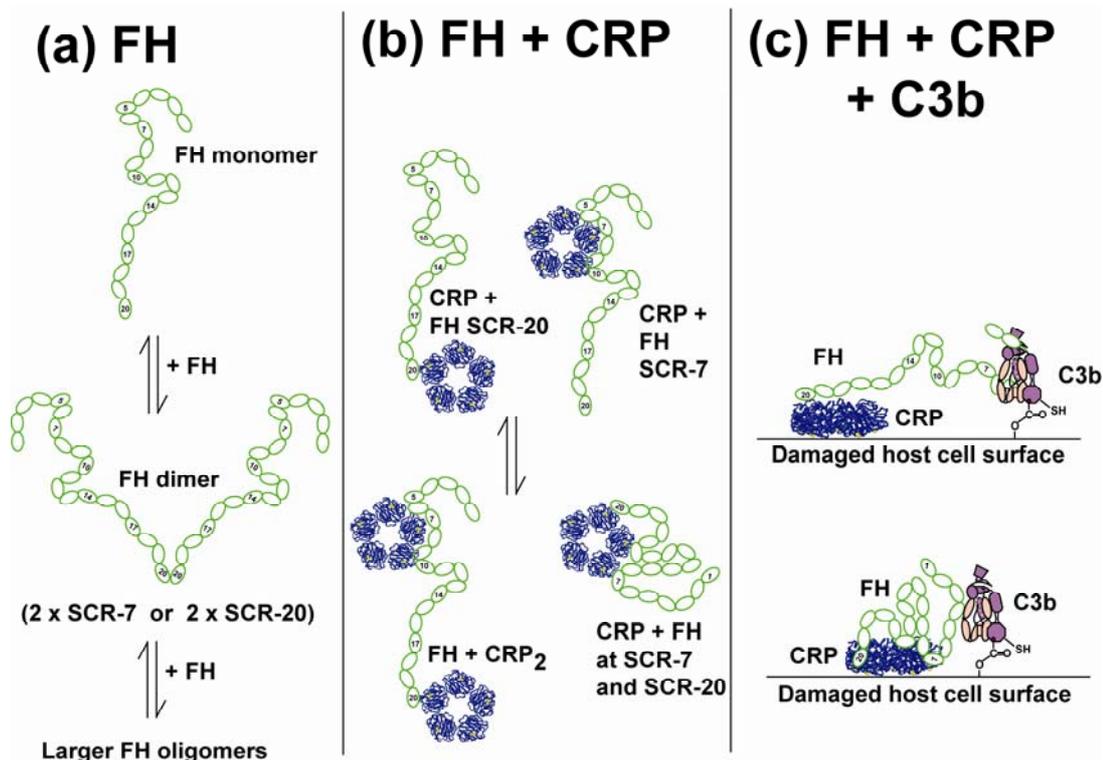
This deduction was verified by additional SPR experiments.

Nine previous reports of the FH-CRP interaction employed a range of buffer conditions that potentially affect the integrity of CRP used in these studies (Table 8.1). CRP is denatured by the removal of  $\text{Ca}^{2+}$  which binds with a  $K_D$  of 60  $\mu\text{M}$ , thus  $\text{Ca}^{2+}$  should be present in at least a ten-fold excess of 60  $\mu\text{M}$  to ensure CRP stability, and 2 mM has proved to be satisfactory (Chapter 7; Kinoshita *et al.*, 1989). The CRP crystal structure shows that  $\text{Ca}^{2+}$  binds to the B-face of CRP (Thompson *et al.* 1999). There is no evidence for denatured CRP *in vivo*, thus it is essential that CRP is correctly folded (Pepys & Hirschfield, 2003). In the studies denoted 1, 2 and 4-8 in Table 8.1, the immobilisation of CRP to ELISA plates or sensor chips for SPR either did not specify the presence of  $\text{Ca}^{2+}$  during the immobilisation process, or  $\text{Ca}^{2+}$  was absent. In four of the studies 1-9 in Table 8.1, no mention was made of the fluid phase  $\text{Ca}^{2+}$  concentration used to study FH binding. In study 9,  $\text{Ca}^{2+}$  was present during CRP immobilisation for ELISA, but was only 0.3 mM in the fluid phase studies (Hakobyan *et al.*, 2008). Another experimental parameter in studies 1-9 has been the use of either 49 mM or 150 mM NaCl concentrations in the buffer with no comparison of its effect on the FH-CRP interaction (Table 8.1). Furthermore, studies 1-9 utilised FH or CRP concentrations that were significantly less than their physiological levels of 1.6-5.4  $\mu\text{M}$  and 0.4-4.4  $\mu\text{M}$  respectively. Given that the  $K_D$  of the FH-CRP interaction is 4  $\mu\text{M}$  in 137 mM NaCl (Figure 8.3(e,f)), significant FH-CRP binding will not be observed if the FH or CRP concentrations are much less than 4  $\mu\text{M}$ . In this context, only one of the eight studies in Table 8.1 employed sufficiently high protein concentrations. In addition, studies 1-9 did not take into consideration the previously-reported self-association of FH and its fragments (Fernando *et al.*, 2007; Nan *et al.*, 2008a; Chapter 5) or that for CRP (Perkins & Pepys, 1986; Blizniukov *et al.*, 2003; Chapter 7), both of which are relevant to the FH-CRP interaction.

Earlier studies had observed FH-CRP interactions at protein concentrations much lower than those used in the present study. Table 8.1 suggests that five earlier FH-CRP studies report interactions that correspond to  $K_D$  values in the region of 0.01  $\mu\text{M}$ , and not that of 4  $\mu\text{M}$  from this study. This apparent anomaly was resolved by the reported strong binding of FH to denatured CRP, but not to native CRP (Bíró *et al.*, 2007; Hakobyan *et al.*, 2008). Evidence that FH binds strongly to denatured CRP but not native FH is provided from the identification of the CRP peptide 86-

SWESASGIVEFWVD-99 as the major linear epitope that binds to FH-like protein-1, whose sequence corresponds to SCR-1/7 of FH (Yu *et al.*, 2007). The CRP crystal structure shows that this peptide corresponds mostly to  $\beta$ -strand I of the double  $\beta$ -sheet structure and part of  $\beta$ -strand H (Shrive *et al.*, 1996).  $\beta$ -strand I is almost completely buried at the interface between the five protomers in CRP. It is not clear how this peptide can interact with FH-like protein-1 unless CRP is first denatured into monomers. In this thesis, SPR studies of the FH-CRP interaction confirm that this occurs in both 50 mM NaCl and 137 mM NaCl buffers containing  $\text{Ca}^{2+}$  (Figures 8.3(a-d)). This agrees with the previous SPR study showing a FH-CRP interaction in 49 mM NaCl (Jarva *et al.*, 1999). It has been suggested that  $\text{Ca}^{2+}$ -depletion leads to CRP denaturation during its immobilisation on the sensor chip (Bíró *et al.*, 2007; Hakobyan *et al.*, 2008). The solution study of CRP showed that its dissociation in the absence of  $\text{Ca}^{2+}$  is slow, and only visible after several days' exposure to such conditions (Chapter 7). Here, CRP immobilisation for SPR was performed in 20  $\mu\text{M}$   $\text{Ca}^{2+}$  using a CRP stock solution stored in 2 mM  $\text{Ca}^{2+}$  (Section 8.4.3), lasted less than 45 mins, and resulted in immobilised CRP that formed decamers with the same  $K_D$  of 22  $\mu\text{M}$  as for CRP in solution (Chapter 7). Therefore it appears unlikely that CRP denaturation occurred on the sensor chip, although other modifications of CRP activity cannot be ruled out.

In this study, the stoichiometry of the FH-CRP interaction in solution and on surfaces was characterised by a combination of AUC and SPR. The AUC sedimentation velocity data used FH concentrations of up to 1.0 mg/ml (6.6  $\mu\text{M}$ ) in 50 mM and 137 mM NaCl and these revealed two distinct FH-CRP complexes in solution (Figure 8.2). A previous AUC sedimentation equilibrium study used protein concentrations of 0.2 mg/ml (1.3  $\mu\text{M}$ ) in 146 mM NaCl, which are below the FH-CRP  $K_D$  value of 4.2  $\mu\text{M}$  from SPR, and would not be able to detect the interaction (Hakobyan *et al.*, 2008). The AUC data also showed that the formation of higher FH oligomers (peaks 3-8) through successive interactions at two different FH sites is suppressed by the FH interaction with CRP. FH self-associates to form higher oligomers at sites located in SCR-6/8 and SCR-16/20 (Fernando *et al.*, 2007; Nan *et al.*, 2008a; 2008b; Chapter 4; Chapter 5). SCR-7 and SCR-8/11 were previously proposed as an interaction site with CRP (Jarva *et al.*, 1999). Steric hindrance effects caused by the FH-CRP interaction would block FH dimerisation at SCR-6/8 (Figure



**Figure 8.6.**

Cartoon summarising the observed interactions between FH and CRP.

(a) Unbound FH (green) exists as 5-15% monomer in equilibrium with dimeric FH with a monomer-dimer  $K_D$  of 28  $\mu\text{M}$ , and with higher oligomers of FH that are formed by contacts between successive pairs of SCR-7 and SCR-20 domains.

(b) Pentameric CRP (blue) binds to FH at either SCR-7 or SCR-20 with a reduced  $K_D$  of 4  $\mu\text{M}$ . The interaction blocks the formation of FH dimers and CRP decamers. Two molecules of CRP can bind simultaneously to one FH molecule. Alternatively one molecule of CRP binds one molecule of FH at SCR-7 and SCR-20 at two different sites on CRP, in which case FH adopts a compact conformation.

(c) A schematic illustration of how CRP can recruit FH onto damaged host cells at either one of SCR-7 or SCR-20 (upper) or at both SCR-7 and SCR-20, thereby regulating the degradation of surface-bound C3b at SCR-1/4, and limiting complement activation at the host cell surface.

8.6(b)), which would then suppress the formation of higher FH oligomers. This is consistent with the FH-CRP  $K_D$  value of 4.2  $\mu\text{M}$  from SPR, which is smaller and takes precedence over the FH-FH  $K_D$  value of 28  $\mu\text{M}$  and the SCR-6/8 dimer  $K_D$  value of 40  $\mu\text{M}$ . The FH-CRP  $K_D$  value of 4.2  $\mu\text{M}$  is also smaller than the CRP-CRP  $K_D$  value of 22  $\mu\text{M}$  from AUC, SPR and X-ray scattering and the SCR-16/20 dimer  $K_D$  value of 16  $\mu\text{M}$  from AUC and X-ray scattering. Thus FH-CRP binding can also block both FH dimerisation at SCR-16/20 and CRP decamer formation. The AUC data in Figure 8.2 do not make it clear whether the formation of CRP decamers or FH dimers or both is inhibited by the FH-CRP interaction, however this appears likely from the above  $K_D$  values. This interpretation is consistent with estimated individual  $K_D$  values for the CRP-SCR-6/8 and CRP-SCR-16/20 interactions of around 10  $\mu\text{M}$  from Figure 8.4. In summary, the FH-CRP interaction is best represented by the formation of both FH-CRP and FH-CRP<sub>2</sub> complexes by CRP pentamers binding to either or both of the two sites on FH (Figure 8.6(b)).

Insight into the solution structure of the FH-CRP complex was obtained only by X-ray scattering. It was not possible to determine sedimentation coefficients for the FH-CRP complexes due to the presence of multiple reaction boundaries. While the X-ray Guinier  $R_G$  analysis was inconclusive, complex formation was detected by comparing the  $P(r)$  curve of the FH-CRP mixture with those of the two proteins on their own (Figure 8.5(b,c)). The  $P(r)$  changes showed that complex formation resulted in the growth of a more compact structure in 137 mM NaCl. This would be consistent with the SPR results showing that each of SCR-6/8 and SCR-16/20 bind to CRP, meaning that one FH molecule is able to interact bivalently with CRP, illustrated in Figure 8.6(b). In solution, AUC readily identified complex formation in 50 mM NaCl, but less so in 137 mM NaCl (Figures 8.2(a,b)). At surfaces, a different set of structures were obtained. This is best explained in terms of further macromolecular interactions noted in the preceding paragraph that are facilitated by the high local concentrations of proteins when bound to surfaces. Even though 50 mM NaCl buffers have clarified the individual roles of SCR-6/8 and SCR-16/20 in binding to CRP by SPR, this buffer resulted in aggregation artefacts in the context of intact FH binding to CRP by SPR. The binding of fluid-phase CRP to immobilised FH was consistently higher in 50 mM NaCl than in 137 mM NaCl (Figures 8.3(a-d)). The reversed binding of fluid-phase FH to immobilised CRP in 50 mM NaCl yielded similar results (Figures 8.3(e,f)). Therefore low salt buffers promote FH-CRP

aggregation at surfaces, and these buffers should only be used with caution for biochemical functional studies. In 137 mM NaCl, the FH-CRP interaction by AUC, SPR and X-ray scattering all reproducibly showed complex formation.

Micromolar affinities and ionic strength-dependence binding are common features of interactions between FH and its ligands and other complement protein-ligand interactions including complement receptor type 2 (Chapter 6; Li *et al.*, 2008). In general, a combination of low ionic strength buffers and high sample concentrations facilitates the observation of weak interactions in solution; however artefacts can result from working in low salt conditions. Thus C3d forms oligomers in 50 mM NaCl, but not in 137 mM NaCl (Li *et al.*, 2008). The NaCl-concentration dependences show that the structural interaction involving SCR-6/8 is principally hydrophobic, while that for SCR-16/20 is principally electrostatic (Figure 8.4). The overall FH-CRP interaction is NaCl-concentration dependent, indicating the dominance of electrostatic forces. Electrostatic maps of CRP show different charge distributions between the A-face and the phosphorylcholine- and Ca<sup>2+</sup>-binding B-face (Figure 8.1). The A-face is generally acidic except for basic charges clustered in the contact region between two adjacent protomers. The B-face that binds to lipid surfaces exhibits more basic charges. FH shows the most basic surfaces at SCR-7, SCR-13 and SCR-20 and it is likely that those at SCR-7 and SCR-20 are involved with CRP binding at the A-face (Figure 8.1). The C1q-binding site in CRP is also at the A-face (Gaboriaud *et al.*, 2004).

Recruitment of FH is the principal mechanism by which CRP down-regulates complement activation by the alternative pathway (Mold *et al.*, 1999; Jarva *et al.*, 1999). This regulatory activity will be particularly useful under conditions of uncontrolled complement activation leading to host damage. In fact, this current study shows that the FH-CRP interaction in solution will only occur during the acute phase response when CRP levels become elevated. The effective local concentration of surface-bound CRP is however increased compared to CRP in solution, and FH will bind to sufficient surface-bound CRP to protect host cells. The suppression of higher FH oligomers by CRP will release FH for regulatory function. The suppression of FH oligomerisation by CRP shows that this interaction is not involved in the molecular mechanism of AMD, given that uncontrolled FH oligomerisation may facilitate the slow formation of sub-RPE deposits (including drusen) at the retinal surface (Nan *et al.*, 2008b). FH binding to CRP does not eliminate the decay-

accelerating and cofactor activities of FH (Jarva *et al.*, 1999). Thus recruitment of FH by CRP bound to damaged host cell surfaces may inhibit complement activation by interacting with bound C3b to regulate this (Figure 8.6(c)). The location of CRP-binding sites in both SCR-6/8 and SCR-16/20 (Figure 8.4) suggests that CRP may play a role in the mechanism of both AMD and aHUS disease. CRP occurs at elevated levels in the retinal layers and drusen deposits of homozygous FH His402 individuals (Johnson *et al.*, 2006). The His402 allotype of FH predisposes to AMD. In this thesis, SPR data showing that the SCR-6/8 His402 variant binds to native CRP more weakly than the native SCR-6/8 Tyr402 variant indicates that FH His402 is less able to protect host cells, which are accordingly more readily damaged (Figure 8.4). Other CRP-binding studies of the His402/Tyr402 FH allotypes show similar trends, but up to now it has not been clear whether native or denatured CRP has been studied (Table 8.1). The binding of SCR-16/20 to CRP is of interest as the majority of aHUS mutations are localised within SCR-16/20. It is possible that the aHUS mutations may affect FH-CRP binding at host renal epithelial cells, thus making these more susceptible to inflammatory attack.

## **(8.4) Materials and Methods**

### **(8.4.1) Purification of Factor H and CRP**

Native FH was purified and concentrated from a pool of just-outdated anonymised human plasma using monoclonal MRC-OX23 Sepharose affinity column (Sim *et al.*, 1993; Nan *et al.*, 2008a; 2008b). SCR-1/5 and SCR-16/20 were expressed as a hexaHis-tagged product in *Pichia pastoris* and purified using nickel affinity and size exclusion chromatography (Chapter 4). SCR-6/8 was expressed and purified following previous procedures (Fernando *et al.*, 2007; A. Miller and S. J. Perkins, unpublished data). Human CRP was isolated and purified as described previously (De Beer & Pepys, 1982). Samples were extensively dialysed into TBS buffers (Tris-buffered saline; 10 mM Tris, 2 mM CaCl<sub>2</sub>, pH 8.0) containing 50 mM NaCl or 137 mM NaCl for AUC. SPR was carried out using HBS buffers (HEPES buffered saline; 10 mM HEPES, 2 mM CaCl<sub>2</sub>, pH 8.0) with 50 mM NaCl or 137 mM NaCl. Samples were routinely analysed by SDS-PAGE before and after experiments to confirm their integrity. The FH amino acid composition was taken from its sequence (SWISSPROT accession code P08603). The glycosylation of FH was taken to be eight biantennary oligosaccharides located at SCR-9, SCR-12, SCR-13, SCR-

14, SCR-15 (two sites), SCR-17 and SCR-18; note that SCR-4 is not glycosylated (Fenaille *et al.*, 2007). This resulted in a calculated FH molecular weight of 154.4 kDa, an unhydrated volume of 193.1 nm<sup>3</sup>, a hydrated volume of 256.2 nm<sup>3</sup> (based on a hydration of 0.3g H<sub>2</sub>O/g of glycoprotein and an electrostricted volume of 0.0245 nm<sup>3</sup> per bound water molecule), a partial specific volume  $\bar{v}$  of 0.715 ml/g, and an absorption coefficient at 280 nm (1%, 1 cm path length) of 16.2 (Perkins, 1986; Chapter 7). The CRP amino acid composition was taken from its sequence (SWISSPROT accession number P02741). The CRP concentration was determined using an absorption coefficient of 17.5 (1%, 280 nm, 1 cm path length) (Nelson *et al.*, 1991). CRP is unglycosylated and has a calculated molecular weight of 115.0 kDa, an unhydrated volume of 150.3 nm<sup>3</sup>, a hydrated volume of 197.3 nm<sup>3</sup> and a partial specific volume  $\bar{v}$  of 0.7408 ml/g (Perkins, 1986). Compositional data for SCR-1/5, SCR-6/8 and SCR-16/20 are reported elsewhere (Chapter 4). For this study, SCR-16/20 was taken to be glycosylated at two sites, with a calculated molecular weight of 41.9 kDa and absorption coefficient of 14.8 (1%, 280 nm, 1 cm path length). Buffer densities for AUC were measured at 20°C to be 1.000958 g/ml (TBS, 50 mM NaCl) and 1.00482 g/ml (TBS, 137 mM NaCl) using an Anton-Paar DMA5000 density meter.

#### **(8.4.2) Analytical ultracentrifugation data for Factor H-CRP**

AUC data for FH and CRP complexes were obtained on two Beckman XL-I instruments equipped with AnTi50 and AnTi60 rotors. Sedimentation velocity data were acquired at 20°C at a rotor speed of 50,000 r.p.m. in two-sector cells with column heights of 12 mm. The sedimentation boundaries were analysed using direct boundary Lamm fits of up to 300 scans using SEDFIT (version 11.7) (Schuck, 1998; 2000). SEDFIT resulted in size-distribution analyses  $c(s)$  for which the fit algorithm assumes that all species have the same frictional ratio  $ff_0$ . The final SEDFIT analyses used a fixed resolution of 200 and optimised the  $c(s)$  fit by floating  $ff_0$  and the baseline until the overall root mean square deviations and visual appearance of the fits were satisfactory (Figure 8.2).

#### **(8.4.3) Surface plasmon resonance data for FH-CRP**

The binding of FH to CRP was analysed by SPR using a Biacore X100 instrument and version 1.1 of its evaluation software (GE Healthcare, Uppsala,

Sweden). Native FH or CRP was coupled to the flowcell of a carboxylated dextran (CM5) research grade sensor chip via a standard amine coupling procedure according to the manufacturer's protocol. FH (20 µg/ml) or CRP (10 µg/ml) in 10 mM acetate buffer, pH 5.5 (FH) or pH 4.3 (CRP) was injected over flow cell 2 until 150 response units (RU) was attained. The CRP sample was maintained in 20 µM CaCl<sub>2</sub> during immobilisation. A control surface cell was prepared identically on flow cell 1 but without protein immobilisation. Binding and equilibrium analyses were performed at 25° C using the appropriate Biacore X100 wizards at flow rates of 10-30 µl/min. Regeneration after each run was achieved by pulsing 10 mM acetate buffer, 2 M NaCl (pH 4.6) across both flow cells once for 30 sec. The running buffer was HBS containing 50 mM or 137 mM NaCl. The kinetic fit for a single set of experiments in 50 mM NaCl (Figure 8.3(e), inset) was performed using the heterogeneous ligand model implemented in the Biacore evaluation software. In this model, the  $k_a$  and  $k_d$  on-rate and off-rate parameters were fitted globally, while two maximum signal response values  $R_{max1}$  and  $R_{max2}$  as well as the bulk refractive index RI were fitted locally to give an overall  $\chi^2$  value of 1.503 RU<sup>2</sup>.

#### **(8.4.4) X-ray scattering data for Factor H-CRP**

X-ray scattering data were obtained in one beam session in four-bunch mode at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, operating with a ring energy of 6.0 GeV on the Beamline ID02 (Narayanan *et al.*, 2001). Storage ring currents ranged from 29 mA to 43 mA. Data were acquired using a recently installed fibre optically-coupled high sensitivity and dynamic range CCD detector (FReLoN) together with a sample-to-detector distance of 3.0 m (Li *et al.*, 2008). FH and CRP were mixed in a molar ratio of 1:1 and studied at total protein concentrations of 1.78 mg/ml (1.00 mg/ml FH and 0.78 mg/ml CRP) and 0.71 mg/ml (0.40 mg/ml FH and 0.31 mg/ml CRP). Samples of volume 100 µl were measured in flow cells in which radiation damage was reduced by moving the sample continuously during beam exposure. Sets of ten time frames, each of duration 0.1 sec, 0.15 sec or 0.2 sec, were acquired. Buffers were measured using the same exposure times in alternation with the samples to eliminate background subtraction errors. On-line checks during data acquisition confirmed the absence of radiation damage, after which the ten frames were averaged. Sample temperature corresponded to ambient conditions at 20°C. Other details including the data

reduction procedure are described elsewhere (Gilbert *et al.*, 2005).

In a given solute-solvent contrast, the radius of gyration  $R_G$  is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low  $Q$  gives the  $R_G$  and the forward scattering at zero angle  $I(0)$ : (Glatter & Kratky, 1982)

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2/3.$$

This expression is valid in a  $Q.R_G$  range up to 1.5. The  $R_G$  analyses were performed using an interactive PERL script program SCTPL7 (J. T. Eaton and S. J. Perkins, unpublished software) on Silicon Graphics O2 Workstations. The  $R_G$  value of a mixture of two proteins A and B is given by the expression (Perkins *et al.*, 1985):

$$R_G^2 = \frac{f_A \sum b_A^2 R_{GA}^2 + f_B \sum b_B^2 R_{GB}^2}{f_A \sum b_A^2 + f_B \sum b_B^2}$$

where  $f_A$  and  $f_B$  are the fractional proportions of A and B,  $b_A$  and  $b_B$  are the total of electrons in A and B, and  $R_{GA}$  and  $R_{GB}$  are the  $R_G$  values of A and B. Indirect transformation of the scattering data  $I(Q)$  in reciprocal space into real space to give the distance distribution function  $P(r)$  was carried out using the program GNOM: (Semenyuk & Svergun, 1991)

$$P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q) Qr \sin(Qr) dQ$$

$P(r)$  corresponds to the distribution of distances  $r$  between volume elements. For this, the X-ray  $I(Q)$  curve utilised up to 392 data points in the  $Q$  range between  $0.09 \text{ nm}^{-1}$  and  $1.8 \text{ nm}^{-1}$ .

**Table 8.1. Buffer conditions used in nine studies of the Factor H-CRP interaction.**

No.	Technique	Immobilised phase	Ca <sup>2+</sup> (yes/no)	Fluid phase	Concentration (mg/ml)	Concentration (µM)	[Ca <sup>2+</sup> ] (mM)	Buffer	Reference
1	ELISA	CRP	n.a.	FH	0.00-0.01	0.0-0.06	n.a.	n.a.	Mold <i>et al.</i> , 1999
2	SPR	CRP	no	FH	0.25	1.6	0.67	1/3 VBS (49 mM NaCl)	Jarva <i>et al.</i> , 1999
		“	“	SCR-1/5	0.027	0.8	“	“	“
		“	“	SCR-1/6	0.105	1.1	“	“	“
		“	“	SCR-1/7	0.05	0.9	“	“	“
		“	“	SCR-8/11	0.05	1.6	“	“	“
		“	“	SCR-8/20	0.045	0.4	“	“	“
		“	“	SCR-15/20	0.045	1.0	“	“	“
		FH		CRP	0.037	0.3	“	“	“
	ELISA	CRP	no	FH	0.001	0.01	1	VBS (147 mM NaCl)	“
3	ELISA	CRP	no	FH	0.0005-0.001	0.003-0.01	5	25 mM Tris (154 mM NaCl)	Bíró <i>et al.</i> , 2007
		PC-KLH-CRP	yes	FH	0.0005-0.001	0.003-0.01	“	“	“
		FH		CRP	0.01	0.1	“	“	“
4	SPR	CRP	no	SCR-6/8 Y402	0.1-0.5	6-20	n.a.	10 mM HEPES, 150 mM NaCl	Herbert <i>et al.</i> , 2007
				SCR-6/8 H402	0.1-0.5	5-21			
5	ELISA	CRP	n.a.	FH	0.00006-0.0015	0.00-0.01	n.a.	VBS (147 mM NaCl)	Laine <i>et al.</i> , 2007
		“	“	SCR-5/7	0.0002-0.005	0.01-0.2	“	“	“
6	ELISA	CRP	no	FH	0-0.001	0.00-0.01	2	50 mM HEPES, 100 mM NaCl	Sjöberg <i>et al.</i> , 2007
		“	“	SCR-6/8	0-0.001	0.0-0.01	“	“	“

7	ELISA	CRP	n.a.	FH	0.001	0.01	n.a.	PBS (137 mM NaCl)	Skerka <i>et al.</i> , 2007
8	ELISA	CRP	no	FH	0.00-0.01	0-0.05	2	20 mM HEPES, 140 mM NaCl	Yu <i>et al.</i> , 2007
9	AUC			CRP, FH	0.4 (0.2, 0.2) <sup>a</sup>	1.7, 1.3 <sup>a</sup>	0.3	VBS (146 mM NaCl)	Hakobyan <i>et al.</i> , 2008
	SEC			CRP, FH	2 (1, 1) <sup>a</sup>	8.7, 6.5 <sup>a</sup>	“	“	“
	ELISA	CRP	yes	FH	0.005-0.1	0.03-0.6	“	“	“
		FH		CRP	0.001-0.1	0.01-0.9	“	“	“
10	AUC			CRP, FH	0.3-2.5 (0.2-1.4, 0.1-1.1) <sup>a</sup>	1.7-12.2, 0.6-7.1 <sup>a</sup>	2	10 mM Tris (137 mM NaCl); 10 mM Tris (50 mM NaCl)	Present study (Chapter 8)
	SPR	FH		CRP	0.01-3.5	0.1-30	“	10 mM Tris (137 mM NaCl); 10 mM HEPES (50 mM NaCl)	”
		CRP	yes	FH	0.02-2.3	0.1-15	“	“	“
				SCR-16/20	0.2-2.1	5-50	“	10 mM HEPES (137 mM NaCl); 10 mM HEPES (50 mM NaCl)	“
				SCR-6/8	0.1-0.4	5-20	“	“	“
	X-ray scattering			CRP, FH	0.7-1.8 (0.3-0.8, 0.4-1.0) <sup>a</sup>	2.6-7.0, 2.6-6.5 <sup>a</sup>	“	10 mM HEPES (137 mM NaCl)	“

n.a. Not available, i.e. either the calcium concentration or the buffer was not stated.

SPR, surface plasmon resonance; AUC, analytical ultracentrifugation; SEC, size exclusion chromatography; VBS, veronal-buffered saline; PC-KLH, phosphorylcholine conjugated to keyhole-limpet haemocyanin.

<sup>a</sup> The first set of values is for CRP, while the second is for FH



**Chapter Nine**  
**Conclusions**

## **(9.1) Prologue**

FH was first discovered as  $\beta$ 1H globulin in 1965 (Nilsson & Müller-Eberhard, 1965). Since then, several important functions have been ascribed to it. The most prominent is the regulation of complement activation by the alternative pathway (Section 2.3.1). The implication of FH in the pathophysiology of diseases such as aHUS and AMD (Section 2.4) further underlines its importance to the host innate immune response. Despite this knowledge, there are currently no high-resolution structures for the full-length protein. The most recent medium resolution structure for FH was published in 2001 (Aslam & Perkins, 2001), at which time only three of the current eleven high resolution structures for individual SCR domains (Section 2.2.2; Table 2.1) were available. Since then, improved data collection and interpretation methods have become available that permit a better understanding of the oligomeric properties of FH. For these reasons, there was a need for updated FH models created using more high-resolution domain structures, and with data interpretation in the light of new information. Structures for shorter fragments could help to provide insight on the overall structural and functional properties of the full-length protein. This thesis presents novel solution structural analyses of the functionally important FH fragments SCR-1/5 and SCR-16/20, updated solution structures for full-length FH, and studies of its interactions with the ligands C3d and CRP. A multidisciplinary approach involving AUC, X-ray scattering and constrained scattering modelling (Sections 3.2; 3.3; 3.4) was used firstly to determine three-dimensional solution structures for FH, C3d and CRP to medium resolution, and secondly these methods in combination with SPR (Section 3.5) were used to elucidate the interactions of FH with its ligands.

## **(9.2) Solution structures of SCR-1/5, SCR-16/20 and full-length Factor H**

The previously reported FH solution structure was updated in this thesis. The three-dimensional structures for SCR-1/5 and SCR-16/20 (Chapter 4) were found to be moderately extended, with relatively folded back but not fully compact domain arrangements. A number of best-fit molecular models with varying orientations were found to meet this criterion (Figure 4.18). These were similar to the structures of the corresponding regions within intact FH (Aslam & Perkins, 2001). Considered in the context of the overall structure of FH, this suggests that the flexibility of the intact molecule is preserved locally between SCR domains within close distances of each

other. A major implication of this finding is that FH can reorientate locally in such a way as to “wrap” itself around its ligands. The interdomain linkers within these two fragments are relatively short, only a maximum of four residues long (except the SCR-18/19 linker, six residues long). The longest interdomain linkers within full-length FH are found in the middle region between SCR-10/15 (Figure 2.2). This central region of FH is heavily glycosylated. Flexibility within this region of FH may allow the functional regions in SCR-1/5 and SCR-16/20 to maintain multiple contacts with the same ligand. SCR-16/20 was found to exist in a monomer-dimer equilibrium in solution. This would have physiological relevance if seen to also occur in the full-length protein, as it could potentially influence the regulatory activity of FH.

New solution structures for full-length FH at ionic strengths of 50 mM NaCl, 137 mM NaCl and 250 mM NaCl (Chapter 5) were used to refine the earlier FH solution structure (Aslam & Perkins, 2001) and confirm the physiological relevance of these results. These new structures showed that the flexible structure of native FH is facilitated by flexible linkers that permit short-to-middle distance interactions between neighbouring SCR domains (Figure 5.11(c)). There was no evidence of a longer-distance interaction between the N- and C-termini within the same FH molecule (Oppermann *et al.*, 2006). Two novel features of FH were revealed by this study. Firstly, these structures showed that the interactions between domains are at least partially electrostatic in nature, and decrease with increased NaCl concentration. Thus FH becomes more elongated with increased ionic strength (Figure 5.14). Secondly, FH forms dimers and oligomers under a range of solution conditions in a NaCl concentration-dependent manner, thus validating the observation of dimerisation in the SCR-16/20 fragment (Chapter 4), as well as in the SCR-6/8 fragment (Fernando *et al.*, 2007). This outcome supports and extends previous observations of FH oligomerisation by examining this in a range of solution conditions (Perkins, 1991; Nan *et al.*, 2008a). The influence of ionic strength on FH conformation suggests that its interaction with ligands may also be partially electrostatic in nature.

### **(9.3) Interactions of Factor H with C3d and C-reactive protein**

The major ligand of FH is C3b of complement which is regulated by FH. Elucidation of the binding of FH to the C3d fragment of C3b (Chapter 6) provided

insights into an aspect of this interaction following the degradation of C3b to C3c and C3d. It has been commonly assumed that the interaction between FH and C3d is a simple 1:1 association. This study revealed, for the first time, that FH forms multimeric complexes with C3d (Figures 6.5-6.7). Dimerisation in SCR-16/20 and oligomerisation in full-length FH were not eliminated by this interaction. Rather, FH forms a ternary complex with C3d and itself (Figures 6.8; 6.10; 6.11). This previously-unknown feature suggests a role for FH oligomers in complement activation where, following C3b cleavage and the deposition of C3d on host cell surfaces, dimeric or oligomeric FH molecules are cross-linked on the host cell surface by C3d (Figure 6.16). This may up-regulate FH in local regions where this is required during inflammation. As hypothesised (Chapter 5; see above), the interaction of FH with this ligand was at least partially electrostatic in nature, thus complex formation was reduced as NaCl concentration increased. This suggests that ionic strength-dependence may be a general feature of the interactions of FH with its ligands.

CRP is an acute-phase protein of the pentraxin protein family that binds ligand in a  $\text{Ca}^{2+}$ -dependent manner. In this thesis, it is shown that CRP exists in a pentamer-decamer equilibrium in physiological buffer (Chapter 7). The pentamer-decamer equilibrium is affected by NaCl concentration and this finding is consistent with that observed for the interaction between FH and C3d. In addition, CRP partially dissociates into lower oligomers, and decamer formation is reduced in the absence of  $\text{Ca}^{2+}$  (Figure 7.4(b)). These two observations show that earlier studies of CRP carried out in  $\text{Ca}^{2+}$ -depleted buffers can yield artefactual results.

Previous studies of the interaction between FH and CRP have been inconclusive and contradictory. Initial reports indicated that FH interacts with CRP (Mold *et al.*, 1999; Jarva *et al.*, 1999; Laine *et al.* 2007; Sjoberg *et al.*, 2007; Skerka *et al.*, 2007; Yu *et al.*, 2007; Herbert *et al.*, 2007). However, several of these studies were performed in non-physiological buffers lacking  $\text{Ca}^{2+}$ , meaning that the results obtained could be artefactual. A more recent study has suggested that the observed interaction is non-physiological and only occurs with denatured CRP (Hakobyan *et al.*, 2008). However, because the 2008 experiments were conducted at below physiological protein concentrations, the biological relevance of this is uncertain. To clarify the details of this interaction, binding of FH to native CRP was studied in both fluid and solid phases in physiological buffers containing  $\text{Ca}^{2+}$  (Chapter 8), and

at physiological concentration ranges for both FH and CRP. This study revealed that FH binds to CRP in both fluid and solid phases in an NaCl concentration-dependent manner, and the interaction suppresses FH oligomerisation (Figure 8.2). In addition, FH binding to CRP leads to a more compact FH domain arrangement (Figure 8.5). It was hypothesised that both known dimerisation sites in FH (SCR-6/8 and SCR-16/20) are involved in binding to CRP. SPR studies confirmed this, identifying for the first time an additional binding site for CRP within SCR-16/20 of FH.

#### **(9.4) Implications for disease**

Determination of the solution structures of the functional FH fragments SCR-1/5 and SCR-16/20, as well as the intact FH molecule considered in the context of their interactions with C3d and CRP, provides new insight on the association of FH with aHUS. The potential role of non C3b- or heparin-binding regions of FH in orientating the molecule for optimal contacts with these ligands may explain the occurrence of Type II mutations (Section 2.4) within SCR-16/19 as well as in SCR-20 (Saunders *et al.*, 2007). FH oligomers are potentially relevant for AMD, given that the hall mark of AMD is the formation of sub-RPE deposits (including drusen, of which FH is a component) at the retinal surface. The suppression of FH oligomerisation by CRP binding provides an insight into how such an interaction at the retinal surface may influence drusen formation, as CRP would hinder drusen growth assuming no other factors are taken into consideration.

## References

- Abdul Ajees, A., Gunasekaran, K., Volanakis, J. E., Narayana, S. V., Kotwal, G. J. & Murthy, H. M. (2006). The structure of complement C3b provides insights into complement activation and regulation. *Nature*, **444**, 221-225.
- Amirlak, I. & Amirlak, B. (2006). Haemolytic uraemic syndrome: an overview. *Nephrology*, **11**, 213-218.
- Arnaout, M. A., Dana, N., Melamed, J., Medicus, R. & Colten, H. R. (1983). Low ionic-strength or chemical cross-linking of monomeric C3b increases its binding affinity to the human complement C3b receptor. *Immunol.* **48**, 229-237.
- 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.
- Aslam, M., Guthridge, J. M., Hack, B. K., Quigg, R. J., Holers, V. M. & Perkins, S. J. (2003). The extended multidomain solution structures of the complement protein Crry and its chimeric conjugate Crry-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. *J. Mol. Biol.* **329**, 525-550.
- Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B. & Perkins, S. J. (1997). Pentameric and decameric structures in solution of the serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J. Mol. Biol.* **272**, 408-422.
- Ault, B. H., Schmidt, B. Z., Fowler, N. L., Kashtan, C. E., Ahmed, A. E., Vogt, B. A. & Colten, H. R. (1997). Human factor H deficiency. Mutations in framework cysteine residues and block in H protein secretion and intracellular catabolism. *J. Biol. Chem.* **272**, 25168-25175.
- Balbo, A. & Schuck, P. (2005). Analytical ultracentrifugation in the study of protein self-association and heterogeneous protein-protein interactions. In *Protein-Protein Interactions* (Golemis, E. & Adams, P. D., eds), pp. 253-277. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY.
- Barlow, P. N., Norman, D. G., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, P. C., Sim, R. B. & 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.
- Behlke, J. & Ristau, O. (1997). Molecular mass determination by sedimentation velocity experiments and direct fitting of concentration profiles. *Biophys. J.* **72**, 428-415.
- Biacore™ concentration analysis handbook, version AA (2001). GE Healthcare Bio-Sciences AB, Uppsala, Sweden.
- Biacore™ X100 handbook, edition AB (2008). GE Healthcare Bio-Sciences AB, Uppsala, Sweden.
- 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., Thielens, N. M., Arlaud, G. J. & 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., 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.
- 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.

- Blizniukov, O. P., Kozmin, L. D., Falikova, V. V., Martynov, A. I. & Tischenko, V. M. (2003). Effect of calcium ions on hydrodynamic properties of pentameric and decameric C-reactive protein in solution. *Mol. Biol.* **37**, 912-918.
- Boehm, M. K., Woof, J. M., Kerr, M. A. & Perkins, S. J. (1999). The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. *J. Mol. Biol.* **286**, 1421-1447.
- Bordet, J. (1909). In *Studies in Immunity*. J. Wiley & Sons, New York.
- Botto, M. & Walport, M. J. (2002). C1q, autoimmunity and apoptosis. *Immunobiology* **205**, 395-406.
- Brooimans, R. A., Hiemstra, P. S., van der Ark, A. A. J., 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.
- Brook, E., Herbert, A. P., Jenkins, H. T., Soares, D. C. & Barlow, P. N. (2005). Opportunities for new therapies based on the natural regulators of complement activation. *Ann. N. Y. Acad. Sci.* **1056**, 176-188.
- Brooks, C. S., Vuppala, S. R., Jett, A. M., Alitalo, A., Meri, S. & Akins, D. R. (2005). Complement regulator-acquiring surface protein 1 imparts resistance to human serum in *Borrelia burgdorferi*. *J. Immunol.* **175**, 3299-3308.
- Buchner, H. (1889). Über die nähere Natur der bakterientötenden Substanz in Blutserum. *Zbl. Bakt.* **6**, 561-565.
- Chamberlain, D., Keeley, A., Aslam, M., Arenas-Licea, J., Brown, T., Tsaneva, I. R. & Perkins, S. J. (1998). A synthetic holliday junction is sandwiched between two tetrameric *Mycobacterium leprae* RuvA structures in solution: new insights from neutron scattering contrast variation and modelling. *J. Mol. Biol.* **284**, 385-400.
- Chamberlain, D., Ullman, C. G. & Perkins, S. J. (1998). Possible arrangement of the five domains in human complement factor I as determined by a combination of X-ray and neutron scattering and homology modeling. *Biochemistry* **37**, 13918-13929.

- Claus, D. R., Siegel, J., Petras, K., Skor, D., Osmand, A. P. & Gewurz, H. (1975). Complement activation by interaction of polyanions and polycations in the presence of C-reactive protein. *J. Immunol.* **118**, 83-87.
- Coca, A.F. (1914). A study of the anticomplementary action of yeast on certain bacteria and of cobra venom. *Z. Immunitätsforsch* **21**, 604–610.
- Cole, J. L. (2004). Analysis of heterogeneous interactions. *Methods Enzymol.* **384**, 212-232.
- 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., Moody, T. P. & Laue, T. M. (2008). Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Meth. Cell Biol.* **84**, 143-211.
- Daha, M. R., van Kooten, C. & Roos, A. (2006). Compliments from complement: a fourth pathway of complement activation? *Nephrol. Dial. Transpl.* **21**, 3374-3376.
- 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.
- Davies, E. J., Snowden, N., Hillarby, M. C., Carthy, D., Grennan, D. M., Thomson, W. & Ollier, W. E. R. (1995). Mannose-binding protein gene polymorphism in systemic lupus erythematosus. *Arthritis Rheum.* **38**, 110-114.
- Davies, E. J., Teh, L. S., Ordi-Ros, J., Snowden, N., Hillarby, M. C., Hajeer, A., Donn, R., Perez-Pemen, P., Villardell-Tarres, M. & Ollier, W. E. (1997). A dysfunctional allele of the mannose binding protein gene associates with systemic lupus erythematosus in a Spanish population. *J. Rheumatol.* **24**, 485-488.
- Davis III, A. E. (1998). C1 Inhibitor Gene and Hereditary Angioedema. In *The Human Complement System in Health and Disease*. (Volanakis, J.E. & Frank, M. M., eds), pp. 455-480. Marcel Dekker Inc., New York.
- De Beer, F. C. & Pepys, M. B. (1982). Isolation of human C-reactive protein and serum amyloid P component. *J. Immunol. Methods* **50**, 17-31.

- De Beer, F. C., Soutar, A. K., Baltz, M. L., Tryner, I. M., Feinstein, A. & Pepys, M. B. (1982). Low density lipoprotein and very low density lipoprotein are selectively bound by aggregated C-reactive protein. *J. Exp. Med.* **156**, 230-242.
- Demeler, B. (2005). UltraScan a comprehensive data analysis software package for analytical ultracentrifugation experiments. In *Modern analytical ultracentrifugation: Techniques and methods* (Scott, D. J., Harding, S. E. & Rowe, A. J., eds), pp. 210-219. Royal Society of Chemistry, Cambridge.
- Demeler, B. & van Holde, K. E. (2004). Sedimentation velocity analysis of highly heterogeneous systems. *Anal. Biochem.* **335**, 279-288.
- Dempsey, P. W., Allison, M. E. D., Akkaraju, C. C. & Fearon, D. T. (1996). C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Nature.* **271**, 348-350.
- Diehl, E. E., Haines, G. K. III, Radosevich, J. A. & Potempa, L. A. (2000). Immunohistochemical localization of modified C-reactive protein antigen in normal vascular tissue. *Am. J. Med. Sci.* **319**, 79-83.
- DiScipio, R. G. (1992). Ultrastructures and interactions of complement factors H and I. *J. Immunol.* **149**, 2592-2599.
- DiScipio, R. G., Daffern, P. J., Schraufstatter, I. U. & Sriramarao, P. (1998). Human polymorphonuclear leukocytes adhere to complement factor H through an interaction that involves alphaMbeta2 (CD11b/CD18). *J. Immunol.* **160**, 4057-4066.
- 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. III, Abel, K. J., Manning, A., Panhuysen, C. & Farrer, L. A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science*, **308**, 421-424.
- Egenhofer, C., Alsdorff, K., Fehsel, K. & Kolb-Bachofen, V. (1993). Membrane-associated C-reactive protein on rat liver macrophages is synthesized within the macrophages, expressed as neo-C-reactive protein and bound through a C-reactive protein-specific membrane receptor. *Hepatology* **18**, 1216-1223.

- Ellman, L., Green, I. & Frank, M. M. (1970). Genetically controlled total deficiency of the fourth component of complement in the guinea pig. *Science*, **170**, 74-75.
- Farries, T. C., Seya, T., Harrison, R. A. & Atkinson, J. P. (1990). Competition for binding sites on C3b by CR1, CR2, MCP, factor B and Factor H. *Complement Inflamm.* **7**, 30-41.
- Fearon, D.T. (1978). Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1971-1975.
- Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riandé, S., Siret, L. & Bihoreau, N. (2007). Site-specific N-glycan characterization of human complement factor H. *Glycobiol.* **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.
- Ferrata, A., (1907). Die unwirksamkeit der komplexen hämolysine in salzfreien lösungen und ihre ursache, 44. Berlin Klin, Wochenschr, p. 366.
- Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N. & Pangburn, M. K. (2006). Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J. Immunol.* **177**, 6308-6316.
- Filep, J. & Földes-Filep, E. (1989). Effects of C-reactive protein on human neutrophil granulocytes challenged with N-formyl-methionyl-leucyl-phenylalanine and platelet-activating factor. *Life Sci.* **44**, 517-524.
- Flexner, S. & Noguchi, H., (1903). Snake venom in relation to haemolysis, bacteriolysis, and toxicity. *J. Exp. Med.* **6**, 277.
- 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.
- Friese, M. A., Hellwage, J., Jokiranta, T. S., Meri, S., Peter, H. H., Eibel, H. & Zipfel, P. F. (2000). Different regulation of factor H and FHL-1/reconectin by inflammatory

- mediators and expression of the two proteins in rheumatoid arthritis (RA). *Clin. Exp. Immunol.* **121**, 406–415.
- Furtado, P. B., Huang, C. Y., Ihyembe, D., Hammond, R. A., Marsh, H. C. & Perkins, S. J. (2007). The partly-folded back solution structure arrangement of the 30 SCR domains in human complement receptor type 1 (CR1) permits access to its C3b and C4b ligands. *J. Mol. Biol.* **375**, 102-118.
- Gaboriaud, C., Thielens, N. M., Gregory, L. A., Rossi, V., Fontecilla-Camps, J. C. & Arlaud, G. J. (2004). Structure and activation of the C1 complex of complement: unravelling the puzzle. *Trends in Immunology* **25**, 368-373.
- Garcia de la Torre, J., Navarro, S., Martinez, M. C. L., Diaz, F. G. & Cascales, J. L. (1994). HYDRO: A computer program for the prediction of hydrodynamic properties of macromolecules. *Biophys. J.* **67**, 530–531.
- Garcia 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.
- Giannakis, E., Jokiranta, T. S., Male, D. A., Ranganathan, S., Ormsby, R. J., Fischetti, V. A., Mold, C. & Gordon, D. L. (2003). A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein. *Eur. J. Immunol.* **33**, 962–969.
- Gilbert, H. E., Eaton, J. T., Hannan, J. P., Holers, V. M. & Perkins, S. J. (2005). Solution structure of the complex between CR2 SCR 1-2 and C3d of human complement: an X-ray scattering and sedimentation modelling study. *J. Mol. Biol.* **346**, 859-873.
- Gilbert, H. E., Asokan, R., Holers, V. M. & Perkins, S. J. (2006). The flexible 15 SCR extracellular domains of human complement receptor type 2 can mediate multiple ligand and antigen interactions. *J. Mol. Biol.* **362**, 1132-1147.
- Glatter, O. & Kratky, O. (1982). Editors of *Small-angle X-ray scattering*. Academic Press, New York.
- Gordon, J., Whitehead, H. R. & Wormall, A. (1914). The action of ammonia on complement. The fourth component. *J. Biochem.* **20**, 1028.

- Gordon, D. L., Kaufman, R., Blackmore, T. K., Kwong, J. & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *J. Immunol.* **155**, 348-356.
- Goodship, T. H. (2006). Atypical HUS and complement dysregulation. *J. Am. Soc. Nephrol.* **17**, 1775-1776.
- Götze, O. & Müller-Eberhard, H. J. (1971). The C3-activator system: An alternative pathway of complement activation. *J. Exp.Med.* **134**, 90-108.
- Ralston, G. (2004). Introduction to Analytical Ultracentrifugation. Beckman.
- 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 its interaction 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., Silvestri, G., Russell, S. R., Klaver, C. C., Barbazetto, I., Chang, S., Yannuzzi, L. A., Barile, G. R., Merriam, J. C., Smith, R. T., Olsh, A. K., Bergeron, 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.
- Hakobyan, S., Harris, C. L., Tortajada, A., Goicochea de Jorge, E., García-Layana, A., Fernández-Robredo, P., Rodríguez de Córdoba, S. & Morgan, B. P. (2008). Measurement of factor H variants in plasma using variant-specific monoclonal antibodies: Application to assessing risk of age-related macular degeneration. *Invest Ophthalmol Vis Sci.* **49**, 1983–1990.
- Hakobyan, S., Harris, C. L., Van den Berg, C. W., Fernandez-Alonso, M. C., Goicochea de Jorge, E., Rodríguez de Córdoba, S., Rivas, G., Mangione, P., Pepys, M. B. & Morgan, B. P. (2008). Complement factor H binds to denatured rather than to pentameric C-reactive protein. *J. Biol. Chem.* **283**, 30451-30460.

- Hammel, M., Kriechbaum, M., Gries, A., Kostner, G. M., Laggner, P. & Prassl, R. (2002). Solution structure of human and bovine  $\beta_2$ -glycoprotein I revealed by small-angle X-ray scattering. *J. Mol. Biol.* **321**, 85-97.
- Hammel, M., Kriechbaum, M., Gries, A., Kostner, G. M., Laggner, P. & Prassl, R. (2002). Solution structure of human and bovine  $\beta_2$ -glycoprotein I revealed by small-angle X-ray scattering. *J. Mol. Biol.* **321**, 85-97.
- Harding, S. E., Horton, J. C., & Morgan, P. J. (1992). MSTAR: A FORTRAN program for the model independent molecular mass analysis of macromolecules using low speed or high speed sedimentation equilibrium. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp 275–294. Royal Society of Chemistry, Cambridge, UK.
- Hellwage, J., Jokiranta, T. S., Koistinen, V., Vaarala, O., Meri, S., & Zipfel, P. F. (1999). Functional properties of complement factor H-related proteins FHR-3 and FHR-4: binding to the C3d region of C3b and differential regulation by heparin. *FEBS Letters*, **462**, 345-352.
- Hellwage, J., Meri, T., Heikkila, T., Alitalo, A., Panelius, J., Lahdenne, P., Seppala, I. J. & Meri, S. (2001). The complement regulator factor H binds to the surface protein OspE of *Borrelia burgdorferi*. *J. Biol. Chem.* **276**, 8427-8435.
- Hellwage, J., Jokiranta, T. S., Friese, M. A., Wolk, T. U., Kampen, E., Zipfel, P. F. & Meri, S. (2002). Complement C3b/C3d and cell surface polyanions are recognized by overlapping binding sites on the most carboxyl-terminal domain of complement factor H. *J. Immunol.* **169**, 6935-6944.
- Herbert, A. P., Soares, D. C., Pangburn, M. K. & Barlow, P. N. (2006). Disease-associated sequence variations in factor H: a structural biology approach. *Adv. Exp. Med. Biol.* **586**, 313-327.
- Herbert, A. P., Uhrin, 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.
- Herbert, A. P., Deakin, J. A., Schmidt, C. Q., Blaum, B. S., Egan, C., Ferreira, V. P., Pangburn, M. K., 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.
- 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.
- Hogasen, A. K. M., Wurzner, R., Abrahamsen, T. G. & Dierich, M. P. (1995). Human polymorphonuclear leukocytes store large amounts of terminal complement components C7 and C6, which may be released on stimulation. *J. Immunol.* **154**, 4734-4740.
- Hurwitz, S. (1996). Homeostatic control of plasma calcium concentration. *Crit. Rev. Biochem. Mol. Biol.* **31**, 41-100.
- Ichinose, A., McMullen, B. A., Fujiwaka, K. & Davie, E. W. (1986). The amino acid sequence of the  $\beta$  subunit of human factor XIII: a protein composed of ten repetitive segments. *Biochemistry*, **25**, 4633-4638.
- Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T. & Yamashina, I. (1987). Serum lectin with known structure activates complement through the classical pathway. *J Biol Chem*, **262**, 7451-7454.
- Jacrot, B. (1976). The study of biological structures by neutron scattering from solution. *Rep. Prog. Phys.* **39**, 911-953.
- Janeway, C. A., Travers, P., Walport, M. & Schlomchik, M. J. (2005). Immunobiology: the immune system in health and disease. Garland Science Publishing, New York. pp. 55-73.
- 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.
- 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.

- Jarva, H., Jokiranta, T. S., Wurzner, R. & Meri, S. (2003). Complement resistance mechanisms of streptococci. *Mol. Immunol.* **40**, 95–107.
- 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 interactin with short consensus repeat domains 7 and 8-11. *J. Immunol.* **163**, 3957–3962.
- Jelezarova, E., Luginbuehl, A. & Lutz, H. U. (2003). C3b<sub>2</sub>-IgG complexes retain dimeric C3 fragments at all levels of inactivation. *J. Biol. Chem.* **278**, 51806-51812.
- Ji, Y. H., Matsushita, M., Okada, H., Fujita, T. & Kawakami, M. (1988). The C4 and C2 but not C1 components of complement are responsible for the complement activation triggered by the Ra-reactive factor. *J Immunol*, **141**, 4271-4275.
- Jiang, H., Siegel, J. N. & Gewurz, H. (1991). Binding and complement activation by C-reactive protein via the collagen-like region of C1q and inhibition of these reactions by monoclonal antibodies to C-reactive protein and C1q. *J. Immunol.* **146**, 2324–2330.
- Johnson, M. L. & Straume, M. (1994). Comments on the analysis of sedimentation equilibrium experiments. In *Modern Analytical Ultracentrifugation* (Schuster, T. M. & Laue, T. M., eds), pp 37-65. Boston, Birkhäuser.
- 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. L., Ekdahl, K. N. & Meri, S. (2001). Complement C3b interactions studied with surface plasmon resonance technique. *International Immunopharmacology.* **1**, 495-506.
- Jokiranta, T. S., Cheng, Z. Z., Seeberger, H., Jozsi, M., Heinen, S., Noris, M., Remuzzi, G., Ormsby, R., Gordon, D. L., 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.
- Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Parepalo, 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.
- Jönsson, U., Fägerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Löfås, S., Persson, B., Roos, H., Rönnberg, I., Sjölander, S., Stenberg, E., Ståhlberg, R., Urbaniczky, C., Östlin, H. & Malmqvist, M. (1991). Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *BioTechniques*, **11**, 620-627.
- Jönsson, U. & Malmqvist, M. (1992). Real-time biospecific interaction. *Advances in Biosensors*, **2**, 291-336.
- Jozsi, M., Heinen, S., Hartmann, A., Ostrowicz, C. W., Halbich, S., Richter, H., Kunert, A., Licht, C., Saunders, R. E., Perkins, S. J., Zipfel, P. F. & Skerka, C. (2006). Factor H and atypical hemolytic uremic syndrome: mutations in the C-terminus cause structural changes and defective recognition functions. *J. Am. Soc. Nephrol.* **17**, 170-177.
- Kaplan, M. H. & Volanakis, J. E. (1974). Interaction of C-reactive protein complexes with the complement system. I. Consumption of human complement associated with the reaction of C-reactive protein with pneumococcal C-polysaccharide and with the choline phosphatides, lecithin and sphingomyelin. *J. Immunol.* **112**, 2135–2147.
- Kavanagh, D., Goodship, T. H. & Richards, A. (2006). Atypical haemolytic uraemic syndrome. *Br. Med. Bull.* **77 & 78**, 5-22.
- Kawasaki, T., Etoh, R. & Yamashina, I. (1978). Isolation and characterization of a mannan-binding protein from rabbit liver. *Biochem Biophys Res Commun*, **81**, 1018-1024.
- Kawasaki, N., Kawasaki, T. & Yamashina, I. (1983). Isolation and characterization of a mannan-binding protein from human serum. *J Biochem (Tokyo)*, **94**, 937-947.
- Khreiss, T., József, L., Hossain, S., Chan, J. S. D., Potempa, L. A & Filep, J. G. (2002). Loss of pentameric symmetry of C-reactive protein is associated with delayed apoptosis of human neutrophils. *J. Biol. Chem.* **277**, 40775-40781.

- Kinoshita, C. M., Ying, S. C., Hugli, T. E., Potempa, L. A., Jiang, H., Houghten, R. A. & Gewurz, H. (1989). Elucidation of a protease sensitive site involved in the binding of calcium to C-reactive protein. *Biochemistry* **28**, 9840-9848.
- Klein, M. A., Kaeser, P. S., Schwarz, P., Weyd, H., Xenarios, I., Zinkernagel, R. M., Carroll, M. C., Verbeek, J. S., Botto, M., Walport, M. J., Molina, H., Kalinke, U., Acha-Orbea, H. & Aguzzi, A. (2001). Complement facilitates early prion pathogenesis. *Nat. Med.* **7**, 488-492.
- Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, 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.
- Kotarsky, H., Hellwage, J., Johnsson, E., Skerka, C., Svensson, H. G., Lindahl, G., Sjobring, U. & Zipfel, P. F. (1998). Identification of a domain in human factor H and factor H-like protein-1 required for the interaction with streptococcal M proteins. *J. Immunol.* **160**, 3349-3354.
- Kratky, O. (1963). X-ray small angle scattering with substances of biological interest in diluted solutions. *Prog. Biophys. Mol. Biol.* **13**, 105-173.
- Kurosky, A., Barnett, D. R., Lee, T. H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H. & Fitch, W. M. (1980). Covalent structure of human haptoglobin: a serine protease homolog. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3388-3392.
- Lamm, O. (1929). Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat. Astr. Fys.* **21B**, 1-4.
- Laine, M., Jarva, H., Seitsonen, S., Haapasalo, K., Lehtinen, M. J., Lindeman, N., Anderson, D. H., Johnson, P. T., Järvelä, 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. & Müller-Eberhard, 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.
- Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical*

- Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 90-125. The Royal Society of Chemistry, Cambridge, U.K.
- Law, S. K. A. & Reid, K. B. M. (1995). Complement (Second edition). IRL Press, Oxford.
- Lebowitz, J., Lewis, M. S. & Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Science* **11**, 2067-2079.
- Lehtinen, M. J., Meri, S. & Jokiranta, T. S. (2004). Interdomain contact regions and angles between adjacent short consensus repeat domains. *J. Mol. Biol.* **344**, 1385-1396.
- Li, S. H., Szmitko, P. E., Weisel, R. D., Wang, C. H., Fedak, P. W., Li, R. K., Mickle, D. A., & Verma, S. (2004). C-Reactive protein upregulates complement-inhibitory factors in endothelial cells. *Circulation* **109**, 833-836.
- 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.
- Liszewski, K. M. & Atkinson, J. P. (1998). Regulatory Proteins of Complement. In *The Human Complement System in Health and Disease* (Volanakis, J. E. & Frank, M. M., eds), pp. 33-48. Marcel Dekker Inc., New York.
- Lozier, J., Takahashi, N. & Putnam, F. W. (1984). Complete amino acid sequence of human plasma  $\beta$ 2-glycoprotein I. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3640-3644.
- Mabbott, N.A., Bruce, M. E., Botto, M., Walport, M. J. & Pepys, M. B. (2001). Temporary depletion of complement component C3 or genetic deficiency of C1q significantly delays onset of scrapie. *Nat. Med.* **7**, 485-487.
- Male, D. E., Ormsby, R. J., Ranganathan, S., Giannakis, E., Gordon, D. L. (2000). Complement factor H: sequence analysis of 221 kb of human genomic DNA containing the entire fH, FHR1 and FHR3 genes. *Mol. Immunol.* **37**, 41-52.
- Manuelian, T., Hellwage, J., Meri, S., Caprioli, J., Noris, M., Heinen, S., Jozsi, M., Neumann, H. P., Remuzzi, G. & Zipfel, P. F. (2003). Mutations in factor H reduce binding affinity to C3b and heparin and surface attachment to endothelial cells in hemolytic uremic syndrome. *J. Clin. Invest.* **111**, 1181-1190.

- Mathieson, P. (2002). Complement factor H and haemolytic uraemic syndrome. *Lancet* **359**, 801-802.
- McRorie, D. K. & Voelker, P. J. (1993). Self-Associating Systems in the Analytical Ultracentrifuge. Beckman Instruments, Fullerton, California.
- Mihlan, M., Hebecker, M., Dahse, H., Hälbich, S., Huber-Lang, M., Dahse, R., Zipfel, P. F. & Józsi, M. (2009). Human complement factor H-related protein 4 binds and recruits native pentameric C-reactive protein to necrotic cells. *Mol. Immun.* **46**, 335-344.
- Milder, F. J., Gomes, L., Schouten, A., Janssen, B. J., Huizinga, E. G., Romijn, R. A., Hemrika, W., Roos, A., Daha, M. R. & Gros, P. (2007). Factor B structure provides insights into activation of the central protease of the complement system. *Nat. Struct. Mol. Biol.* **14**, 224-8.
- Mold, C., Gewurz, H. & Du Clos, T. W. (1999). Regulation of complement activation by C-reactive protein. *Immunopharmacology* **42**, 23-30.
- Mori, S. & Barth, H. G. (1999). Size exclusion chromatography. Springer-Verlag Berlin Heidelberg.
- Morley, B. J. & Walport, M. J. (2000). The Complement Facts Book. Academic Press, London.
- Motie, M., Brockmeister, S., & Potempa, L. A. (1996). Binding of model soluble complexes to modified C-reactive protein. *J. Immunol.* **156**, 4435-4441.
- Müller-Eberhard, H. J. (1968). Chemistry and Reaction Mechanisms of Complement. *Adv. Immunol.* **8**, 1-80.
- Müller-Eberhard, H. J. (1988). Molecular organisation and function of the complement system. *Ann. Rev. Biochem.* **57**, 321-347.
- 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.

- Nan, R., Gor, J., Lengyel, I. & Perkins, S. J. (2008). Uncontrolled zinc- and copper-induced oligomerisation of the human complement regulator factor H and its possible implications for function and disease. *J. Mol. Biol.* **384**, 1341-1352.
- Narayanan, T., Diat, O. & Bosecke, P. (2001). SAXS and USAXS on the high brilliance beamline at the ESRF. *Nucl. Instrum. Methods Phys. Res. A*, **467-468**, 1005-1009.
- Navratil, J. S., Watkins, S. C., Wisnieski, J. J. & Ahearn, J. M. (2001). The globular heads of C1q specifically recognize surface blebs of apoptotic vascular endothelial cells. *J. Immunol.* **166**, 3231-3239.
- Nelson, R. A. Jr., Jensen, J., Gigli, I. & Tamura, N. (1966). Methods for the separation, purification and measurement of nine components of haemolytic complement in guinea-pig serum. *Immunochemistry*, **3**, 111-135.
- Nelson, S. R., Tennent, G. A., Sethi, D., Gower, P. E., Ballardic, F. W., Amatayakul-Chantler, S. & Pepys, M. B. (1991). Serum amyloid P component in chronic renal failure and dialysis. *Clin. Chimica Acta* **200**, 191-200.
- Nicholson-Weller, A., March, J. P., Rosenfeld, S. L. & Austen, K. F. (1983). Affected erythrocytes of patients with paroxysmal nocturnal hemoglobinuria are deficient in the complement regulatory protein, decay accelerating factor. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5066-5070.
- Nielsen, H. E., Christensen, K. C., Koch, C., Thomsen, B. S., Heegaard, N. H. & Tranum-Jensen, J. (1989). Hereditary, complete deficiency of complement factor H associated with recurrent meningococcal disease. *Scand. J. Immunol.* **30**, 711-718.
- Nilsson, U. R. & Müller-Eberhard, H. J. (1965). Isolation of  $\beta$ 1F-globulin from human serum, and its characterization as the fifth component of complement. *J. Exp. Med.* **122**, 277.
- Noris, M. & Remuzzi, G. (2006). Complement Factor H Gene Abnormalities in Haemolytic Uraemic Syndrome: From Point Mutations to Hybrid Gene. *PLoS. Med.* **3**, 1719-1720.
- Nuttall, G. (1888). Experimente über die bacterienfeindlichen einflüsse des thierischen körpers. *Z. Hyg. Infektionskr.* **4**, 353-394.

- 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., Nan, R., Gor, J. & Perkins, S. J. (2009a). Electrostatic interactions contribute to the folded-back conformation of wild-type human Factor H. *J. Mol. Biol.* In press.
- Okemefuna, A. I., Li, K., Nan, R., Ormsby, R. J., Sadlon, T., Gordon, D. L. & Perkins, S. J. (2009b). Multimeric interactions between complement Factor H and its C3d ligand provide new insight on complement regulation. *J. Mol. Biol.* In press.
- Okemefuna, A. I., Stach, L., Rana, S., Ziai Buetas, A. J., Gor, J., Mangione, P., Pepys, M. B. & Perkins, S. J. (2009c). C-reactive protein exists in an NaCl-concentration dependent pentamer-decamer equilibrium in physiological buffer containing calcium. In preparation.
- Oppermann, M., Manuelian, T., Jozsi, 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.
- Ormsby, R. J., Jokiranta, T. S., Duthy, T. G., Griggs, K. M., Sadlon, T. A., Giannakis, E. & Gordon, D. L. (2006). Localization of the third heparin-binding site in the human complement regulator factor H. *Mol. Immunol.* **43**, 1624-1632.
- Osmand, A. P., Friedenson, B., Gewurz, H., Painter, R. H., Hofmann, T. & Shelton, E. (1977). Characterization of C-reactive protein and the complement subcomponent C1t as homologous proteins displaying cyclic pentameric symmetry (pentraxins). *Proc Natl. Acad. Sci. USA.* **74**, 739-743.
- Overhauser, A. W. (1953). Polarisation of nuclei in metals. *Phys. Rev.* **92**, 411-415.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **11**, 2411-2423.
- Pangburn, M. K. & Müller-Eberhard, H. J. (1984). The alternative pathway of complement. *Semin. Immunopathol.* **7**, 163-192.

- Pangburn, M. K., Schreiber, R. D. & Muller-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.
- Pangburn, M. K., Schreiber, R. D. & Müller-Eberhard, H. J. (1983). Deficiency of an erythrocyte membrane protein with complement regulatory activity in paroxysmal nocturnal hemoglobinuria. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5430-5434.
- Pangburn, M. K., Pangburn, K. L., Koistinen, V., Meri, S. & Sharma, A. K. (2000). Molecular mechanisms of target recognition in an innate immune system: interactions among factor H, C3b, and target in the alternative pathway of human complement. *J. Immunol.* **164**, 4742-4751.
- Pangburn, M. K., Rawal, N., Cortes, C., Alam, M. N., Ferreira, V. P. & Atkinson, M. A. (2009). Polyanion-induced self-association of complement factor H. *J. Immunol.* **182**, 1061-1068.
- Pepys, M. B. & Hirschfield, G. M. (2003). C-reactive protein: a critical update. *J. Clin. Invest.* **111**, 1805-1812.
- Perkins, S. J. (1986). Protein volumes and hydration effects: the calculation of partial specific volumes, neutron scattering matchpoints and 280 nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* **157**, 169-180.
- Perkins, S. J. (1988a). Structural studies of proteins by high-flux X-ray and neutron solution scattering. *Biochem. J.* **254**, 313-327.
- Perkins, S. J. (1988b). X-ray and neutron solution scattering. In *New Comprehensive Biochemistry* (Neuberger, A. & Deenen, L. L. M., eds.), Vol 11B, Part II, pp 143-265. Elsevier, Amsterdam.
- Perkins, S. J. (1994). High-flux X-ray and neutron solution scattering. *Methods Mol. Biol.* **22**, 39-60.
- Perkins, S. J. (2001). X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and modelling fits. *Biophys. Chem.* **93**, 129-139.

- Perkins, S. J. & Weiss, H. (1983). Low resolution structural studies of mitochondrial ubiquinol-cytochrome c reductase in detergent solutions by neutron scattering. *J. Mol. Biol.* **168**, 847-866.
- Perkins, S. J. & Sim, R. B. (1986). Molecular modelling of human complement component C3 and its fragments by solution scattering. *Eur. J. Biochem.* **157**, 155-168.
- Perkins, S. J. & Goodship, T. H. (2002). 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., Kerckaert, J. P. & Loucheux-Lefebvre, M. H. (1985). The shapes of biantennary and tri-tetraantennary  $\alpha_1$  acid glycoprotein by small angle neutron and X-ray scattering. *Eur. J. Biochem.* **147**, 525-531.
- Perkins, S. J., Haris, P. I., Sim, R. B. & Chapman, D. (1988). A study of the structure of human complement component factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods. *Biochemistry* **27**, 4004-4012.
- Perkins, S. J., Nealis, A. S. & Sim, R. B. (1991). Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry* **30**, 2847-2857.
- Perkins, S. J., Ashton, A. W., Boehm, M. K. & Chamberlain, D. (1998). Molecular structures from low angle X-ray and neutron scattering studies. *Int. J. Biol. Macromol.* **22**, 1-16.
- Perkins, S. J., Gilbert, H. E., Aslam, M., Hannan, J., Holers, V. M. & Goodship, T. H. (2002). Solution structures of complement components by X-ray and neutron scattering and analytical ultracentrifugation. *Biochem. Soc. Trans.* **30**, 996-1001.
- Perkins, S. J., Gilbert, H. E., Lee, Y. C., Sun, Z. & Furtado, P. B. (2005). Relating small angle scattering and analytical ultracentrifugation in multidomain proteins. In *Modern Analytical Ultracentrifugation: Techniques and Methods* (Scott, D. J., Harding S. E. & Rowe A. J., eds), pp 291-319. Royal Society of Chemistry, London, UK.

- 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 modelling. *Meth. Cell Biol.* **84**, 375-423.
- Philo, J. S. (1997). An improved function for fitting sedimentation velocity data for low-molecular weight solutes. *Biophys. J.* **72**, 435-444.
- Philo, J. S. (2006). Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques *Anal. Biochem.* **354**, 238–246.
- Pillemer, L., Ecker, E. E., Oncley, J. L. & Cohn, E. J. (1941). The preparation and physicochemical characterization of the serum protein components of complement. *J. Exp. Med.* **74**, 297-308.
- Potempa, L. A., Maldonado, B. A., Laurent, P., Zemel, E. S. & Gewurz, H. (1983a). Antigenic, electrophoretic and binding alterations of human C-reactive protein modified selectively in the absence of calcium. *Mol. Immunol.* **20**, 1165–1175.
- Potempa, L. A., Maldonado, B. A., Laurent, P., Zemel, E. S. & Gewurz, H. (1983b). Stimulation of human neutrophils, monocytes and platelets by modified C-reactive protein (CRP) expressing a neoantigenic specificity. *Inflammation.* **12**, 391-405.
- Potempa, L. A., Siegel, J. N., Fedel, B. A., Potempa, R. T. & Gewurz, H. (1987). Expression, detection and assay of a neoantigen (neo-CRP) associated with a free, human C-reactive protein subunit. *Mol. Immunol.* **24**, 531-541.
- 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.
- Prota, A. E., Sage, D. R., Stehle, T. & Fingerroth, J. D. (2002). The crystal structure of human CD21: Implications for Epstein-Barr virus and C3d binding. *Proc. Natl. Acad. Sci. USA* **99**, 10641-10646.
- Ralston, G. (1993). Introduction to Analytical Ultracentrifugation, Beckman Instruments, Palo Alto.
- Ram, S., F. G., Mackinnon, S., Gulati, D. P., McQuillen, U., Vogel, M., Frosch, C., Elkins, H. K., Guttormsen, L. M., Wetzler, M., Oppermann, M., Pangburn, M. K.

- & Rice, P. A. (1999). The contrasting mechanisms of serum resistance of *Neisseria gonorrhoeae* and group B *Neisseria meningitidis*. *Mol. Immunol.* **36**, 915-928.
- Ramadan, M. A. M., Shrive, A. K., Holden, D., Myles, D. A. 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., Sect. D* **58**, 992-1001.
- Rees, R. F., Gewurz, H., Siegel, J. N., Coon, J. & Potempa, L. A. (1988). Expression of a C-reactive protein neoantigen (neo-CRP) in inflamed rabbit liver and muscle. *Clin Immunol Immunopathol.* **48**, 95-107.
- Reid, K. B. M. (1998). C1q and Mannose-Binding Lectin. In *The Human Complement System in Health and Disease* (Volanakis, J. E. & Frank, M. M., eds), pp. 33-48. Marcel Dekker Inc., New York.
- Rhodes, G. (2000). Crystallography made crystal clear 2<sup>nd</sup> Edition. Academic Press, New York.
- Richards, A., Buddles, M. R., Donne, R. L., Kaplan, B. S., Kirk, E., Venning, M. C., Tielemans, C. L., Goodship, J. A. & Goodship, T. H. (2001). Factor H mutations in hemolytic uremic syndrome cluster in exons 18-20, a domain important for host cell recognition. *Am. J. Hum. Genet.* **68**, 485-490.
- Riedemann, N. C. & Ward, P. A. (2003). Complement in ischemia reperfusion injury. *Am. J. Pathol.* **162**, 363-367.
- Ripoche, J., Day, A. J., Harris, T. J. & Sim, R. B. (1988). The complete amino acid sequence of human complement factor H. *Biochem. J.* **249**, 593-602.
- Rodriguez de Córdoba, S., Diaz-Guillén, M. A. & Heine-Suner, D. (1999). An integrated map of the human regulator of complement activation (RCA) gene cluster on 1q32. *Mol. Immunol.* **36**, 803-808.
- Rodriguez 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.

- Rogers, J., Cooper, N. R., Webster, S., Schultz, J., McGeer, P. L., Syren, S. D., Civin, W. H., Brachova, L., Bradt, B. & Ward, P. (1992). Complement activation by  $\beta$ -amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10016-10020.
- Rother, K., Till, G. O. & Hänsch, G. M. (1998). *The Complement System*. Springer, New York.
- Rowe, I. F., Soutar, A. K., Trayner, I. M., Thompson, G. R. & Pepys, M. B. (1984). Circulating human C-reactive protein binds very low density lipoproteins. *Clin. Exp. Immunol.* **58**, 237-244.
- Sambrook J., Fritsch E.F. & Maniatis T. (1989). *Molecular Cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- Samols, D., MacIntyre, S. S. & Kushner, I. (1985). Studies of translatable mRNA for rabbit C-reactive protein. *Biochem. J.* **227**, 759-765.
- Sandberg, A. L. & Osler, A. G. (1941). The preparation and physiochemical characterization of the serum protein components of complement. *J. Exp. Med.* **74**, 297-308.
- Saunders, R. E. & Perkins, S. J. (2006). A user's guide to the interactive Web database of factor H-associated hemolytic uremic syndrome. *Semin. Thromb. Hemost.* **32**, 160-168.
- Saunders, R. E., Goodship, T. H., 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. Mutat.* **27**, 21-30.
- Saunders, R. E., Abarrategui-Garrido, C., Fremeaux-Bacchi, V., Goicoechea, d. J., Goodship, T. H., Lopez, T. M., Noris, M., Ponce, C., I, Remuzzi, G., Rodriguez, d. C., Sanchez-Corral, P., Skerka, C., Zipfel, P. F. & Perkins, S. J. (2007). The interactive Factor H-atypical hemolytic uremic syndrome mutation database and website: update and integration of membrane cofactor protein and Factor I mutations with structural models. *Hum. Mutat.* **28**, 222-234.
- Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S., Lyon, M., Uhrin, D. & Barlow, P. N. (2008). A new map of glycosaminoglycan and C3b binding sites on Factor H. *J. Immunol.* **181**, 2610-2619.

- Schneider, M. C., Exley, R. M., Chan, H., Feavers, I., Kang, Y., Sim, R.B., & Tang, C. M. (2006). Functional significance of factor H binding to *Neisseria meningitidis*. *J. Immunol.* **176**, 7566–7575.
- Schuck, P. (1998). Sedimentation analysis of non-interacting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* **75**, 1503-1512.
- Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606-1619.
- Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104-124.
- Schuck, P. (2007). Protein Interactions: Biophysical Approaches for the Study of Complex Reversible Systems, Springer US. pp. 469-518.
- 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.
- Schwaeble, W., Schäfer, M. K., Petry, F., Fink, T., Knebel, D., Weihe, E. & Loos, M. (1995). Follicular dendritic cells, interdigitating cells, and cells of the monocyte-macrophage lineage are the C1q-producing sources in the spleen. Identification of specific cell types by in situ hybridization and immunohistochemical analysis. *J. Immunol.* **155**, 4971-4978.
- Scott, D. J. & Schuck, P. (2005). A brief introduction to the analytical ultracentrifugation of proteins for beginners. In *Modern analytical ultracentrifugation: Techniques and methods* (Scott, D. J., Harding, S. E. & Rowe, A. J., eds), p.15. Royal Society of Chemistry, Cambridge.
- Semenyuk, A. V. & Svergun, D. I. (1991). GNOM - a program package for small-angle scattering data-processing. *J. Appl. Crystallogr.* **24**, 537-540.
- 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. T., Holden, D., Myles, D. A. A., Turnell, W. G., Volanakis, J. E., Pepys, M. B., Bloomer, A. C. & Greenhough, T. J. (1996). Three dimensional structure of human C-reactive protein. *Nat. struct. biol.* **3**, 346-354.
- Siegel, J., Rent, R. & Gewurz, H. (1974). Interactions of C-reactive protein with the complement system. I. Protamine-induced consumption of complement in acute phase sera. *J. Exp. Med.* **140**, 631-647.
- Siegel, 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-lysine polymers and other polycations. *J. Exp. Med.* **142**, 709-721.
- Sim, R. B. & DiScipio, R. G. (1982). Purification and structural studies on the complement-system control protein beta 1H (Factor H). *Biochem. J.* **205**, 285-293.
- Sim, R. B., Day, A. J., Moffatt, B. E. & Fontaine, M. (1993). Complement factor I and cofactors in control of complement system convertase enzymes. *Meth. Enzymol.* **223**, 13-35.
- 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.
- 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), pp 19-62. Taylor & Francis, Boca Raton, USA.
- Skerka, C., Hellwage, J., Weber, W., Tilkorn, A., Buck, F., Marti, T., Kampen, E., Beisiegel, U. & Zipfel, P. F. (1997). A novel short consensus repeat-containing protein is associated with human triglyceride-rich lipoproteins. *J. Biol. Chem.* **272**, 5627-5634.
- Skerka, C., Lauer, N., Weinberger, A. A., Keilhauer, C. N., Sühnel, J., Smith, R., Schlötzer-Schrehardt, U., Fritsche, L., Heinen, S., Hartmann, A., Weber, B. H. &

- Zipfel, P. F. (2007). Defective complement control of factor H (Y402H) and FHL-1 in age-related macular degeneration. *Mol. Immunol.* **44**, 3398-3406.
- Stafford, W. F. (2000). Analysis of reversibly interacting macromolecular systems by time derivative sedimentation velocity. *Methods Enzymol.* **323**, 302-325.
- Stafford, W. F. & Sherwood, P. J. (2004). Analysis of heterologous interacting systems by time derivative sedimentation velocity. *Methods Enzymol.* **323**, 302-325.
- Summerfield, J. A., Ryder, S., Sumiya, M., Thursz, M., Gorchein, A., Monteil, M. A., & Turner, M. W. (1995). Mannose binding protein gene mutations associated with unusual and severe infections in adults. *Lancet* **345**, 886-889
- Super, M., Thiel, S., Lu, J., Levinsky, R. J. & Turner, M. W. (1989). Association of low levels of mannan-binding protein with a common defect of opsonisation. *Lancet II*, 1236-1239.
- Szakonyi, G., Guthridge, J. M., Li, D., Young, K. A., Holers, V. M. & Chen, X. S. (2001). Structure of complement receptor 2 in complex with its C3d ligand. *Science* **292**, 1725-1728.
- Tacnet-Delorme, P., Chevallier, S. & Arlaud, G. J. (2001).  $\beta$ -amyloid fibrils activate the C1 complex of complement under physiological conditions. Evidence for a binding site for Ab on the C1q globular regions. *J. Immunol.* **167**, 6374-6381.
- Tanious, F. A., Nguyen, B. & Wilson, W. D. (2008). Biosensor-surface plasmon resonance methods for quantitative analysis of biomolecular interactions. *Meth. Cell Biol.* **84**, 53-77.
- Taylor, K. E. & van den Berg, C. W. (2006). Structural and functional comparison of native pentameric, denatured monomeric and biotinylated C-reactive protein. *Immunology*, **120**, 404-411.
- Thompson, D., Pepys, M. B. & Wood, S. P. (1999). The physiological structure of human C-reactive protein and its complex with phosphocholine. *Struct. Fold. Des.* **7**, 169-177.
- Tillett, W. S. & Francis, J. T. (1930). Serological reactions in pneumonia with a nonprotein somatic fraction of pneumococcus. *J. Exp. Med.* **52**: 561-585.

- Van der Merwe, P. A. (2001). Surface plasmon resonance. In *Protein–ligand interactions: hydrodynamics and calorimetry* (Harding, S. & Chowdhry, P. Z., eds), Practical Approach series, pp 137-170. Oxford University Press, New York.
- Van Holde, K. E. & Wieschet, W. O. (1978). Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers* **17**, 1387-1403.
- 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.
- 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.
- Volanakis, J. E. & Frank, M. M. (1998). *The Human Complement System in Health and Disease*. Marcel Dekker Inc., New York.
- Vyse, T. J., Morley, B. J., Bartok, I., Theodoridis, E. L., Davies, K. A., Webster, A. D. & Walport, M. J. (1996). The molecular basis of hereditary complement factor I deficiency. *J. Clin. Invest.* **97**, 925-933.
- Wang, H. W. & Sui, S. F. (2001). Dissociation and subunit rearrangement of membrane-bound human C-reactive proteins. *Biochem. & Biophys. Res. Comm.* **288**, 75-79.
- Ward, H. M., Higgs, N. H., Blackmore, T. K., Sadlon, T. A. & Gordon, D. L. (1997). Cloning and analysis of the human complement factor H gene promoter. *Immunol. Cell. Biol.* **75**, 508–510.
- 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.
- Weiss, W. I. & Drickamer, K. (1994). Trimeric structure of a C-type mannose-binding protein. *Structure.* **2**, 1227-1240.
- Wiesmann, C., Katschke, K. J., Yin, J., Helmy, K. Y., Steffek, M., Fairbrother, W. J., McCallum, S. A., Embuscado, L., DeForge, L., Hass, P. E. & van Lookeren

- Campagne, M. (2006). Structure of C3b in complex with CR1g gives insights into regulation of complement activation. *Nature*, **444**, 217-220.
- Wistow, G., Bernstein, S. L., Wyatt, M. K., Fariss, R. N., Behal, A., Touchman, J. W., Bouffard, G., Smith, D. & Peterson, K. (2002). Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: over 6000 non-redundant transcripts, novel genes and splice variants. *Mol. Vis.* **8**, 205-220.
- Whaley, K. & Ruddy, S. (1976). Modulation of the alternative complement pathways by beta 1 H globulin. *J. Exp. Med.* **144**, 1147-1163.
- White, R. T., Damm, D., Hancock, N., Rosen, B. S., Lowell, B. B., Usher, P., Flier, J. S. & Spiegelman, B. M. (1992). Human adiponin is identical to complement factor D and is expressed at high levels in adipose tissue. *J. Biol. Chem.* **267**, 9210-9213.
- Wirthmueller, U., Dewald, B., Thelen, M., Schäfer, M. K., Stover, C., Whaley, K., North, J., Eggleton, P., Reid, K. B. & Schwaebler, W. J. (1997). Properdin, a positive regulator of complement activation, is released from secondary granules of stimulated peripheral blood neutrophils. *J. Immunol.* **158**, 4444-4451.
- Wright, K. & Catlow, R. (1999). *Microscopic Properties and Processes in Minerals*. Kluwer Academic Publishers, The Netherlands.
- Wu, Y., Li, S. R., Wang, H. W., Sui, S. F. (2002). The study of spontaneous dissociation of C-reactive protein. *Biokhimiya.* **67**, 1377-1382.
- Wuthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.
- 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.
- Zalman, L. S., Wood, L. M. & Müller-Eberhard, H. J. (1986). Isolation of a human erythrocyte membrane protein capable of inhibiting expression of homologous complement transmembrane channels. *Proc. Natl Acad. Sci. USA.* **83**, 6975-6989.
- Zipfel, P. F. (2001). Hemolytic uremic syndrome: how do factor H mutants mediate endothelial damage? *Trends Immunol.* **22**, 345-348.
- Zipfel, P. F. & Skerka, C. (1999). The human factor H-like protein 1: a complement regulatory protein with cell adhesion function. *Immunol. Today* **20**, 135-140.

- Zipfel, P. F., Jokiranta, T. S., Hellwage, J., Koistinen, V. & Meri, S. (1999). The factor H protein family. *Immunopharmacology* **42**, 53-60.
- Zipfel, P. F., Skerka, C., Hellwage, J., Jokiranta, S. T., Meri, S., Brade, V., Kraiczy, P., Noris, M. & Remuzzi, G. (2002). Factor H family proteins: on complement, microbes and human diseases. *Biochem. Soc. Trans.* **30**, 971-978.
- Zipfel, P. F., Misselwitz, J., Licht, C. & Skerka, C. (2006a). The role of defective complement control in hemolytic uremic syndrome. *Semin. Thromb. Hemost.* **32**, 146-154.
- Zipfel, P. F., Heinen, S., Jozsi, M. & Skerka, C. (2006b). Complement and diseases: defective alternative pathway control results in kidney and eye diseases. *Mol. Immunol.* **43**, 97-106.

## Publications

- 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.
- 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.
- 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 modelling. *Meth. Cell Biol.* **84**, 375-423.
- Okemefuna, A. I., Nan, R., Gor, J. & Perkins, S. J. (2009). Electrostatic interactions contribute to the folded-back conformation of wild-type human Factor H. *J. Mol. Biol.* *In press*.
- Okemefuna, A. I., Li, K., Nan, R., Ormsby, R. J., Sadlon, T., Gordon, D. L. & Perkins, S. J. (2009). Multimeric interactions between complement Factor H and its C3d ligand provide new insight on complement regulation. *J. Mol. Biol.* *In press*.
- Perkins, S. J., Okemefuna, A. I., Nan, R., Li, K. & Bonner, A. (2009). Constrained solution scattering modelling of human antibodies and complement proteins reveals novel biological insights. *J. Royal Soc. Int.* *In press*.
- Okemefuna, A. I., Stach, L., Rana, S., Ziai Buetas, A. J., Gor, J., Mangione, P., Pepys, M. B. & Perkins, S. J. (2009). C-reactive protein exists in an NaCl-concentration dependent pentamer-decamer equilibrium in physiological buffer containing calcium. In preparation.
- Okemefuna, A. I., Gor, J., Nan, R., Mangione, P., Pepys, M. B. & Perkins, S. J. (2009). Complement Factor H binds at two distinct sites to C-reactive protein in solution and on surfaces in an NaCl concentration-dependent manner. In preparation.

## **Presentations and abstracts**

Okemefuna, A. I., Gilbert, H. E., Griggs, K., Ormsby, R., Gordon, D. L. & Perkins, S. J.

Solution structures for the SCR 1-5 and SCR 16-20 fragments of Factor H reveal unexpected conformational properties. XXIth International Complement Workshop, Beijing, China, 22-28 October 2006 (Poster).

Okemefuna, A. I., Gilbert, H. E., Griggs, K., Ormsby, R., Gordon, D. L. & Perkins, S. J.

Solution structures for the regulatory SCR-1/5 and cell-surface-binding SCR-16/20 fragments of Factor H reveal different self-associative properties. 11<sup>TH</sup> European Meeting on Complement in Human Diseases, Cardiff, 8-11<sup>th</sup> Sep 2007 (Poster).

Okemefuna, A. I., Nan, R., Gor, J. & Perkins, S. J. Electrostatic interactions contribute to

a folded-back and flexible SCR domain structure of Factor H. XXII International Complement Workshop, Basel, Switzerland, 28 Sep–2 Oct 2008 (Poster).

Okemefuna, A. I., Ormsby, R. J., Sadlon, T., Gordon, D. L. & Perkins, S. J. Multiple

interactions between C3d and the C-terminal SCR-16/20 fragment of complement Factor H. XXII International Complement Workshop, Basel, Switzerland, 28 Sep– 2 Oct 2008 (Poster).