Bioinformatics studies on sequence, structure and functional relationships of proteins involved in the complement system

Dinesh Christopher Soares

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

January 2007

ABSTRACT

The complement system consists of ~30 proteins involved in a series of cascading reactions ultimately resulting in the destruction of invading pathogens. Central to the pathway are three paralogous proteins, C3, C4 and C5. Furthermore, pivotal to the successful functioning of this system are five proteins belonging to the regulators of complement activation (RCA) family - they ensure that a complement-mediated immune response is proportionate and targeted against infection. RCA proteins are characterised by numerous occurrences of a single module-type; the complement control protein (CCP) module. In this work, comprehensive bioinformatics analyses of sequence and structure of CCP modules was undertaken. Through extensive database and literature searches CCP module sequences and structures were retrieved and large-scale *all-against-all* sequence and structure comparisons performed; analysis of intermodular orientations for pairs of modules and larger fragments was performed. These analyses revealed interesting consensus and novel trends. Based upon optimal use of experimentally determined CCP module structures as templates, an automated large-scale protein structure comparative modelling procedure was implemented for a set of 203 CCP-module sequences. The accuracies of models produced were validated, and the models are publicly available online at "The CCP module model database" (http://www.bru.ed.ac.uk/~dinesh/ccp-db.html), which also serves as a comprehensive resource for information on CCP modules. By way of example, surface electrostatic analyses were undertaken for the first 28 CCP module models of complement receptor type 1 (CR1). This revealed a striking case of adaptive evolution – assignments to clusters based on surfaces differ from assignments to clusters based on sequences and sheds light on mutagenesis studies. Overall, the models provide a rich vein of information for design of mutants, interpretation of phenotypic consequences of polymorphisms, and prediction of function. This wider utility of models is discussed in relation to inferring consequences of several disease-associated sequence variations for complement proteins - CR1, factor H, MCP, and another CCP-containing protein, SRPX2. Finally, homology models of C5 and C5b were created on the basis of the recent landmark publication of C3 and C3b structures. This revealed presence of a novel

putative disulfide bond in C5, and additionally helped revisit previous functional peptide and mutant data and provides insight into the latter stages of complement assembly.

Dedicated to the loving memory of my dad, Joe Soares

ACKNOWLEDGEMENTS

It is difficult to thank all the people, who have in many ways been helpful during the course of my Ph.D.

Family: First and foremost, thanks to my wonderful family. To my dad Joe, to whom this thesis is dedicated – thanks for everything you did for me; for the love, encouragement and support in getting me to where I am today. You are missed! To my mum Vilma, sister Divya, and brothers Jitesh and Rohan – for the years of patient, kindness and friendship you have shown me, and for all the "crazy" fun times. To my (much) better-half, Joanna Sharman, simply for everything! You are the best - I salute you all, hope I have made you proud.

Paul Barlow and Dietlind Gerloff: Thanks to Dietlind, for inviting me to pursue a Ph.D. in Edinburgh in the first instance, and for her ever positive outlook. Extra-special thanks to Paul Barlow, for his constant guidance throughout this journey – and for going over this thesis several times! It would not have been possible without your help and encouragement, during the difficult times, and over the many months (!) of thesis writing.

Edinburgh friends and colleagues: Thanks to the many friends, past and present I have made over the years here in Edinburgh. Thanks to the pub-quiz "dream team": Russell Hamilton, Ralf Schmid, Kirsten Lillie, Graeme Ball, Claudia Blindauer, and Iain McNae. Special thanks to Lindsay Tulloch, Heather Davidson, and Liam Worrall for their friendship (and housing) over the years. Thanks to the Swann third floor (structural biochemistry and bioinformatics groups) and the "complement experts" in the BioNMR group for several useful discussions. Thanks to Paul Taylor (EPIC), Russell Hamilton (NOVA) and Thomas Juettemann (AMBLER/LYCIA) for excellent computing support. Thanks to the amazing Edinburgh University Table Tennis Club – keep on winning! And last, but not least, a big thank you to my many collaborators for making this possible. Thanks to the Edinburgh Protein Interaction Centre (EPIC) for funding. All scientific acknowledgements follow each chapter.

PUBLICATIONS

The following thesis describes contributions contained in the following:

1. Blein, S., Ginham, R., Uhrin, D., Smith, B. O., **Soares, D. C.**, Veltel, S., McIlhinney, R. A., White, J. H., and Barlow, P. N. (2004). Structural analysis of the complement control protein (CCP) modules of GABA(B) receptor 1a: only one of the two CCP modules is compactly folded. J Biol Chem *279*, 48292-48306.

2. Bramham, J., Thai, C. T., **Soares, D. C.**, Uhrin, D., Ogata, R. T., and Barlow, P. N. (2005). Functional insights from the structure of the multifunctional C345C domain of C5 of complement. J Biol Chem *280*, 10636-10645.

3. **Soares, D. C.**, and Barlow, P. N. (2005). Complement Control Protein Modules in the Regulators of Complement Activation. In Structural Biology of the Complement System, D. Morikis, and J. D. Lambris, eds. (Boca Raton, CRC Press, Taylor & Francis Group), pp. 19-62.

4. **Soares, D. C.**, Gerloff, D. L., Syme, N. R., Coulson, A. F., Parkinson, J., and Barlow, P. N. (2005). Large-scale modelling as a route to multiple surface comparisons of the CCP module family. Protein Eng Des Sel *18*, 379-388.

5. Brook, E., Herbert, A. P., Jenkins, H. T., **Soares, D. C.**, and Barlow, P. N. (2005). Opportunities for new therapies based on the natural regulators of complement activation. Ann N Y Acad Sci *1056*, 176-188.

6. Herbert, A. P.*, **Soares, D. C.***, Pangburn, M. K., and Barlow, P. N. (2006). Disease-associated sequence variants in factor H: a structural biology approach. Adv Exp Med Biol *586,* 313-327. *(* = co-first authors).*

7. Royer, B., **Soares, D. C.**, Barlow, P. N., Bontrop, R., Roll, P., Robaglia-Schlupp, A., Blancher, A., Cau, P., Pontarotti, P., and Szepetowski, P. (2007). Molecular evolution of the human *SRPX2* gene that causes brain disorders of the rolandic and sylvian speech areas. *(submitted).*

8. **Soares, D. C.**, Quan, X., Ogata, R. T., and Barlow, P. N. (2007). Structural and functional insights of human complement component C5/C5b through comparative 3D-modelling *(manuscript in preparation).*

CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
Å	Angstroms
aHUS	atypical haemolytic uraemic syndrome
AMD	age-related macular degeneration
ANA	Anaphylatoxin/Anaphylotoxin
BLAST	Basic Local Alignment Search Tool
BM	Brüch's membrane
BSA	Buried surface area
C3, C4, C5	Complement components C3, C4 and C5
C4BP	C4b-binding protein
CA	Cofactor activity
CCP	Complement control protein
CD	Clusters of differentiation
CE	Combinatorial Extension
CR1	Complement receptor type1
CR2	Complement receptor type 2
CSMD	CUB and sushi multiple domain protein 1
C-terminal	Carboxyl-terminal
CUB	domain first found in C1r, C1s, uEGF, and bone morphogenetic protein
CVF	Cobra venom factor
DAA	Decay accelerating activity
DAF	Decay accelerating factor
DSSP	Database of Secondary Structure in Proteins
ECC	Electrostatic correlation coefficient
E-value	Expectancy value
FH	Factor H
FHL	Factor H-like
FHR	Factor H-related
FIMAC/FIM	Factor I-membrane attack complex module/Factor-I module
FRET	Fluorescence resonance energy transfer microscopy
FXIIIb	blood-clotting factor XIII b subunit
GABA-BR1a	Gamma-aminobutyric acid type B receptor, subunit 1 α
GAG	Glycosaminoglycan
GPI-anchor	glycosyl-phosphatidylinositol-anchor
GRASP	Graphical Representation and Analysis of Structural Properties
H-bond	Hydrogen-bond
HCC	Hydrophobic correlation coefficient
HMM	Hidden Markov model
HSQC	Heteronuclear single quantum correlation
HUS	haemolytic uraemic syndrome
HYR	hyalin repeat domain
IL2R and IL15R	Interleukin 2 and 15 receptors (α-chain)
Indels	Insertions or deletions
KCP/Kaposica	Kaposi's sarcoma-associated herpesvirus
LHR	Long homologous repeat
LNK	Linker
MAC	Membrane Attack Complex

MASP	Mannan- or mannose-binding protein-associated serine protease
MBP/MBL	Mannan-binding protein/Mannan-binding lectin
MCP	Membrane cofactor protein
MG	macroglobulin
MPGN II	membranoproliferative glomerulonephritis type II
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
N-terminal	Amino-terminal
NTR	Netrin-like module
PDB	Protein Data Bank
PEG	Polyethylene glycol
PfEMP1	*Plasmodium falciparum* Erythrocyte Membrane Protein 1
pH	Potenz hydrogen
PIPSA	Protein Interaction Property Similarity Analysis
PQR	PDB file, where B-factor column has been replaced by "charge" (Q)
ProbCons	Probabilistic Consistency-based Multiple Alignment of Amino Acid Sequences
r.m.s.d./rmsd	Root mean square deviation
RBC	Red blood cell
RCA	Regulator of Complement Activation
RPE	retinal pigment epithelium
SA	Surface area
SAXS	small-angle X-ray scattering
SCOPPI	Structural Classification of Protein-Protein Interfaces
SCR	Short consensus repeat
sCR1	Soluble complement receptor type 1
SCWRL	Side-Chain replacement With Rotamer Libraries
SMART	Simple Modular Architecture Research Tool
SNP	Single-nucleotide polymorphism
SPICE	Smallpox inhibitor of complement enzymes
SRPX2	sushi-repeat-containing protein, X-linked 2
TED	Thioester-domain
TEP	Thioester-containing protein
TTP	thrombotic thrombocytopenic purpura
UPGMA	Unweighted Paired-Group Method with Arithmetic mean
VCP	Vaccinia virus complement control protein
WT	Wild-type
α2M	α2-macroglobulin
β2-GPI/ApoH	Beta-2-Glycoprotein I/Apolipoprotein H

Names, Abbreviations and Single Letter Codes for Amino Acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid (Aspartate)	Asp	D
Cysteine	Cys	C
Glutamic acid (Glutamate)	Glu	E
Glutamine	Gln	Q
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 1

The Complement System: Function, Regulation and Role in Disease

---

1.1 The Complement System

The complement system is an evolutionary ancient, essential molecular component of innate immunity that consists of a set of approximately 30-plasma and cell-membrane-attached proteins (See Table 1.1 for a complete list). These proteins act in an orchestrated fashion (Figure 1.1), to eliminate infectious organisms and other unwanted particles from the body, elicit inflammatory responses, and facilitate adaptive immune responses (Walport 2001b; a; Carroll 2004b; a).

No.	Complement component protein	Approximate molecular mass (kDa)	No. of chains	Approximate plasma or serum concentrations (μg/ml)	Cleavage fragments
1	C1q	410	18	70-300	
2	C1r	83	1	34-100	
3	C1s	85	1	30-80	
4	C2	102	1	15-30	C2a, C2b
5	C3	190	2	1200-1500	C3a, C3b, C3c, C3d, C3f, C3g, C3dg, C3d-K, iC3b

6	C4	205	3	350-600	C4a, C4b, C4c, C4d, C4dg
7	C5	196	2	70-85	C5a, C5b
8	C6	125	1	60-70	
9	C7	120	1	55-70	
10	C8	150	3	55-80	
11	C9	66	1	50-160	
12	Factor B	100	1	140-240	Ba, Bb
13	Factor D	24	1	1-2	
14	Properdin	224	4	20-30	
15	MBL	540	18	1	
16	MASP-1/ MASP-3	94	1		
17	MASP-2	76	1		
18	C1 inhibitor (C1inh)	105	1	180-275	
19	C4BP	550	7	250	
20	Factor H	150	1	300-560	
21	Factor I	100	2	34-50	
22	CR1 (CD35)	160-250	1		
23	CR2 (CD21)	140	1		
24	CR3 (CD11b/18)	265	2		
25	CR4 (CD11c/18)	245	2		
26	DAF (CD55)	70	1		
27	MCP (CD46)	45-70	1		
28	C1qR	65	1		
29	C3aR	54	1		
30	C5aR (CD88)	45	1		
31	CD59	20	1		
32	Carboxypeptidase N	280	4		

Table 1.1: Proteins involved in the complement system. The table provides a summary of information on the complement proteins. Most values derive from the Merck manual (http://www.merck.com/mrkshared/mmanual/), others from literature. Abbreviations used in table: C (except where stated otherwise) = complement component; MBL = mannose- or mannan-binding lectin; MASP = mannose-binding lectin-associated serine protease; C4BP = C4b-binding protein; CR = complement receptor; DAF = decay accelerating factor; MCP = membrane cofactor protein; CD = clusters of differentiation; cleavage fragments are denoted with a small letter after the component.

Even though its principal function is supposed to be protective, inappropriate complement activation or overzealous complement activity have been linked to a

wide variety of inflammatory conditions – reviewed in Section 1.5. In this sense, the complement system has been referred to as a "double-edged sword" (Colten 1994).

The complement system can be activated via three routes namely, the alternative pathway, classical pathway, and lectin pathway. All three activation pathways result in formation of bimolecular convertase enzymes (Figure 1.1) that drive a cascade of proteolytic events, providing a basis for amplification.

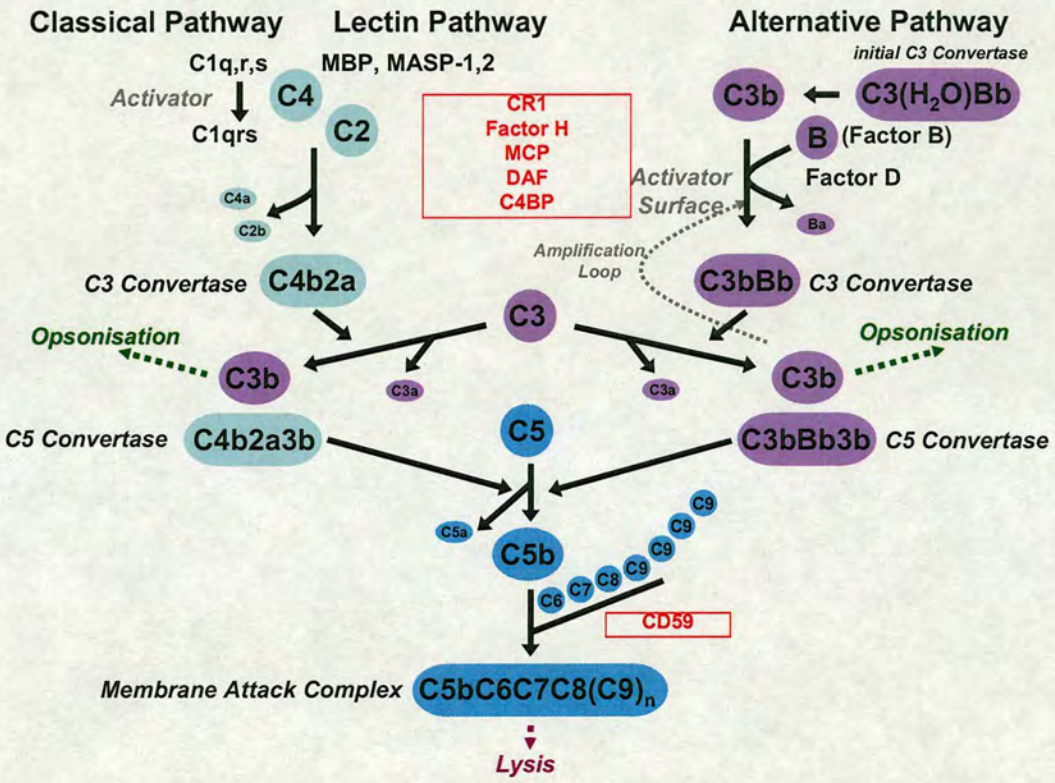


Figure 1.1: Schematic diagram to summarise the complement "cascade" system. The classical and lectin pathways are triggered by binding of antibodies to complement component C1 (consisting of C1q and two copies of C1r and C1s), or specific sugars to the mannose-binding protein/mannose-binding protein-associated serine protease (MBP/MASP) complex. The alternative pathway runs in continuous "tick-over" mode because C3 undergoes spontaneous conversion to C3(H2O). The five RCAs are C4b-binding protein (C4BP), MCP, DAF, CR1, and factor H. All act in subtly different ways upon the convertases. See text for a description of pathway and explanation of abbreviations.

1.1.1 The Alternative pathway

The *alternative pathway* of complement activation provides a first line of defence because it is permanently switched on at a low level (*i.e.* it operates in "tick-over" mode), but is kept in check by regulators that act only on self-surfaces. The presence of any unprotected, "non-self" surface thus induces rapid escalation of the system. In the alternative pathway (Figure 1.1), the complement component protein C3 undergoes slow, spontaneous hydrolysis to yield $C3(H_2O)$. $C3(H_2O)$ associates with a second protein, factor B. Factor B, bound to $C3(H_2O)$, is a substrate for factor D. In this context factor B is cleaved to form Bb, and the resultant $C3(H_2O)$.Bb complex is a proteolytically active entity. It cleaves C3 into C3a and C3b. C3b attaches covalently to surfaces, and like $C3(H_2O)$, associates with factor B, ultimately forming the alternative pathway convertase, C3b.Bb. This surface-associated bimolecular enzyme converts more molecules of C3 to C3b, thus contributing to the formation of extra convertases and the creation of a positive feedback loop.

1.1.2 The Classical and Lectin pathways

The complement system may additionally be activated by antibodies bound to polyvalent antigens. This is referred to as the *classical pathway* (linking the innate and adaptive immune systems). In the lectin pathway, sugars on bacterial surfaces activate complement. The classical and lectin pathways involve cleavage of complement component C4, a homologue of C3, resulting in formation of C4b. C4b also binds to surfaces and it forms a convertase through association with C2, creating the bimolecular, proteolytically active complex C4b.2a. Like the alternative pathway convertase, the classical/lectin pathway convertase thus formed cleaves C3 to C3b and

C3a. The C3b product is available for formation of more alternative pathway convertase complexes (*i.e.*, C3b.Bb).

1.1.3 The common-Lytic pathway

Both alternative and classical pathway convertases can associate with further molecules of C3b to form trimolecular complexes, C3b.Bb.C3b and C4b.2a.C3b, which have C5 convertase activity, *i.e.* they cleave complement component C5 into C5a and C5b. C5b may subsequently act as a nucleation site for assembly of C6, C7, C8 and several molecules of C9 to form a multiprotein membrane attack complex (MAC, or C5-C9) that penetrates cell membranes, causing lysis, leakage and cell death.

A further consequence of complement activation, which is independent of MAC formation, is deposition of multiple molecules of C3b and C4b on the target surface. This is called opsonisation and provides the basis for immune clearance, an important process by which C3b/C4b-coated particles are cleared from the bloodstream (Ross and Medof 1985). Furthermore, the cleavage products of C3b - C3d and C3dg - are ligands for other receptors, able to trigger signalling pathways that bring about, for example, enhanced levels of antibody production by B-cells (Dempsey et al. 1996). Finally, the small fragments, C3a, C4a, and C5a, cleaved from C3, C4, and C5 during complement activation, are the anaphylatoxins (also referred to as the anaphylotoxins) – *i.e.* chemotactic peptides that mediate a series of proinflammatory phenomena. Thus, activation of complement disrupts the membrane of the target cell, removes foreign and other unwanted particles to the liver for destruction, enhances the adaptive immune response, and, via anaphylatoxins, recruits lymphocytes and macrophages to the site of infection.

1.2 The Regulators of Complement Activation

Not only is the complement system effective in immune defence, but it also has potential for damage to self. It is therefore vital that complement activation is both capable of being rapidly triggered, and tightly regulated. These properties are achieved largely as a result of assembly and decay of the aforementioned C3 convertase enzymes, C3b.Bb and C4b.C2a, which occupy pivotal positions within the pathway. It is the positive feedback loop involving C3b production and convertase formation that provides complement with the potential to act in a rapid (explosive) fashion. The bimolecular convertases (C3b.Bb and C4b.C2a) have a half-life in the order of minutes. Once they have decayed into their components, they cannot reassemble; for example, C3b must associate with a fresh molecule of intact factor B to produce a new convertase complex. Consequently, there is the potential for a complement response to die out when the stimulus to activation is destroyed and/or subject to immune clearance.

In addition, a critical set of five proteins are dedicated to regulation of the complement cascade via interactions with the convertases. These are known as the regulators of complement activation (RCA) family, and consist of C4b-binding protein (C4BP), complement receptor type 1 (CR1, CD35), decay accelerating factor (DAF, CD55), factor H (fH) and membrane cofactor protein (MCP, CD46) (Table 1.1 and Figure 1.2). The RCA proteins have two principal activities in terms of their action on the convertases. Some of the RCAs accelerate the decay of the convertases, reducing their half-life from minutes to seconds. This is decay accelerating activity (DAA). The mechanism is unknown. Some RCAs act as cofactors for the protease, factor I (cofactor activity, CA); factor I cleaves C3b and C4b to generate C3dg and C4dg, thus further reducing potential for convertase formation. The five human

RCAs have overlapping functional profiles. Molecules of at least one of DAF, MCP, and CR1 are expressed on the surfaces of every cell in the body where they play a protective role. The soluble regulators, C4BP and fH, are particularly important for preventing activation at self-surfaces that lack membranes, because such surfaces, are unlikely to possess the cell-bound regulators DAF, MCP, and CR1.

The RCA family is exceptional amongst the complement proteins in that each of the five proteins are composed almost entirely of multiple examples of a single type of domain or module known as the complement control protein (CCP) module (analysed in detail in the following chapter). RCA-mediated protein-protein interactions occur via binding sites extending on one or over several adjacent CCP molecules.

As a consequence of its potential for inflicting damage on invaders, a large range of complement evasion strategies have evolved amongst microorganisms (Lindahl et al. 2000). The most widespread anti-complement measure involves "hijacking" of RCAs. Many bacteria express proteins that act as virulence factors through their ability to sequester RCAs. Another approach is to "steal" an RCA gene and adapt it for the purposes of the invader. Thus one or more pox viruses, like Vaccinia virus, produce proteins that closely mimics the mammalian RCAs (VCP) both in sequence and function (Ciulla et al. 2005). The following section briefly describes the main functions of each RCA (and VCP) in terms of their activities/interactions with other complement components, and additionally lists interactions with all other known ligands to date. The structures of these proteins are discussed in depth in Chapter 2.

1.2.1 Complement Receptor type 1

Complement receptor type 1 is a transmembrane, single polypeptide chain glycoprotein containing 30 CCP modules in its most common allotype, expressed primarily on erythrocytes and other blood cells (Klickstein et al. 1987). It is unique in having both DAA and CA and acting on both the alternative and classical pathway convertases. In CR1 (Krych-Goldberg and Atkinson 2001b) there are three functional regions, each extending over three modules, within the extracellular domain: one copy of "site 1" and two near-identical copies of "site 2". Site 1 resides within the N-terminal, membrane-distal, three CCP modules of CR1 (CCPs 1-3). It has convertase DAA along with the ability to bind C4b (and C3b weakly). Each copy of site 2 (CCPs 8-10 and CCPs 15-17) has high affinity for C3b and C4b, and has CA.

Ligand	Binding site	References
Complement component C3b	CCPs 8, 9, 10 + CCPs 15, 16, 17 Binds CCPs 1, 2, 3 weakly	(Krych et al. 1991; Krych et al. 1994; Krych et al. 1998)
Complement component C4b	CCPs 1, 2, 3 + CCPs 8, 9, 10 + CCPs 15, 16, 17	(Krych et al. 1991; Krych et al. 1994; Krych et al. 1998)
Complement component iC3b	CCPs 1, 2, 3 + CCPs 8, 9, 10 + CCPs 15, 16, 17	(Krych et al. 1991; Krych et al. 1994; Krych et al. 1998)
C1q and mannan-binding lectin	CCPs 22-28	(Klickstein et al. 1997; Ghiran et al. 2000)
Plasmodium falciparum adhesin • Erythrocyte Membrane Protein 1 (PfEMP1)	CCPs 8, 9, 10 + CCPs 15, 16, 17	(Rowe et al. 1997; Rowe et al. 2000)
Leishmania major	-	(da Silva et al. 1988; Da Silva et al. 1989)
Mycobacterium tuberculosis	-	(Schlesinger et al. 1990)

Table 1.2: Binding sites on CR1 for various ligands.

1.2.2 Factor H

Factor H is an abundant, soluble single-chain, plasma glycoprotein composed from 20 CCP modules. It has both DAA and CA, but it acts exclusively upon the alternative pathway convertase. The four N-terminal CCP modules (CCPs 1-4) harbour a C3b-binding site and this region is the locus of the CA and DAA of factor H, while the two C-terminal modules (CCPs 19-20) form another C3b-binding site (specifically for the C3d component of C3b) and are crucial for self and non-self discrimination. A third set of modules involving CCPs 12-14 are thought to bind another region of C3b (C3c).

Ligand	Binding site	References
Complement component C3b	CCPs 1, 2, 3, 4 + CCPs 12, 13, 14 (C3c) + CCPs 19, 20 (C3d)	(Jokiranta et al. 2000)
Glycosaminoglycans (GAGs)	CCP 7 + CCP 9 + CCPs 12, 13, 14 + CCPs 19, 20	(Pangburn et al. 1991; Blackmore et al. 1996; Blackmore et al. 1998b; Prodinger et al. 1998; Ormsby et al. 2006)
Sialic acid	CCPs 16 – 20	(Ram et al. 1998b)
L-selectin	-	(Malhotra et al. 1999)
C-Reactive Protein (CRP)	CCPs 7 – 11	(Jarva et al. 1999)
Adrenomedullin (AM)	CCPs 15-20 + CCPs 8-11	(Martinez et al. 2003)
Borrelia afzelii • BaCRASP-1, -2 • BaCRASP-4, -5	 CCPs 1 - 7 CCPs 19, 20	(Kraiczy et al. 2001a)
Borrelia burgdorferi • BbCRASP-1, -2, -3, -4, -5 • OspE • p21 • ErpA (BBL39) • ErpC • ErpP (BBN38)	 CCPs 19, 20 CCPs 15 – 20 CCPs 19, 20 CCPs 19, 20 CCPs 19, 20 CCPs 19, 20 -	(Alitalo et al. 2001; Hellwage et al. 2001; Kraiczy et al. 2001a; Kraiczy et al. 2001b; Alitalo et al. 2002; Stevenson et al. 2002; McDowell et al. 2003; Metts et al. 2003)

• 35kDa protein		
Borrelia hermsii • FhbA (FHBP19) • FHBP28	CCPs 1-7 + CCPs 16-20	(Hovis et al. 2004; Hovis et al. 2006)
Borrelia recurrentis	-	(Meri et al. 2006)
Borrelia duttonii	-	(Meri et al. 2006)
Borrelia parkeri	-	(McDowell et al. 2003; McDowell et al. 2004)
Borrelia garinii	-	(Alitalo et al. 2005)
Candida albicans • CaCRASP-1, -2	CCPs 6, 7 + CCP 19	(Meri et al. 2002a)
Streptococcus pneumoniae • Hic • PspC	CCPs 8 – 11 CCPs 6 – 10 + CCPs 13 - 15	(Duthy et al. 2002; Jarva et al. 2002; Dave et al. 2004a; Dave et al. 2004b; Quin et al. 2005)
Streptococcus pyogenes • M-Proteins • Fba	CCP 7 CCP 7	(Blackmore et al. 1998a; Kotarsky et al. 1998; Pandiripally et al. 2002; Pandiripally et al. 2003; Perez-Caballero et al. 2004; Wei et al. 2005)
Streptococcus agalactiae • β protein	CCP 13 or CCP 20	(Areschoug et al. 2002; Jarva et al. 2004)
Echinococcus granulosus	-	(Diaz et al. 1997)
HIV • gp41, gp120	-	(Stoiber et al. 1996; Stoiber et al. 1997)
Neisseria gonorrhoeae • Porin1A • LPS	CCPs 16 – 20 -	(Ram et al. 1998a; Ram et al. 1998b)
Neisseria meningitidis • Class 3 PorB	-	(Ram et al. 1999)
Onchocerca volvulus • Microfilariae (mf)	CCPs 8 – 20	(Meri et al. 2002b)
Yersinia enterocolitica • YadA	-	(China et al. 1993)

Trypanosoma cruzi	-	(Tomlinson et al. 1994)
Treponema denticula	CCPs 1-7	(McDowell et al. 2005)
Actinobacillus actinomycetemcomitans • Omp100	-	(Asakawa et al. 2003)
Leptospira interrogans • LfhA	CCPs 18 – 20	(Verma et al. 2006)

Table 1.3: Binding sites on Factor H for various ligands

1.2.3 Decay-accelerating factor

Decay accelerating factor (DAF, also called CD55), is a single-chain glycoprotein (Medof et al. 1987a; Medof et al. 1987b) attached to the membrane by a glycosyl-phosphatidylinositol (GPI)-anchor (Medof et al. 1986), and possesses only DAA. It has four CCP modules and its classical pathway convertase regulation has been mapped to CCP modules 2 and 3, while alternative pathway convertase regulation is encompassed by CCPs 2-4.

Ligand	Binding site	References
Complement convertases/components • C3bBb • C4b2a • C3b • Factor B • Bb subunit	 CCPs 2, 3, 4 CCPs 3, 4 - - -	(Brodbeck et al. 1996; Kuttner-Kondo et al. 2001; Harris et al. 2005)
CD97	CCP 1	(Hamann et al. 1996; Hamann et al. 1998; Qian et al. 1999)
Echovirus • *7* • *11* • *12*	 CCPs 2, 3, 4 CCP 3 + 20% contribution from CCPs 1, 2, 4 CCPs 2, 3, 4	(Clarkson et al. 1995; Lea et al. 1998; Bhella et al. 2004; Pettigrew et al. 2006)
Coxsackievirus • *B3*	CCP 2	(Bergelson et al. 1995;

• *A21*	CCP 1	Shafren et al. 1997)
Enterovirus 70	CCP 1	(Karnauchow et al. 1998)
Escherichia coli • Dr adhesins • X adhesins • Afa adhesins • F1845 adhesins	 CCPs 2, 3 CCPs 3, 4 CCPs 2, 3, 4 CCPs 2, 3, 4	 (Nowicki et al. 1993; Pham et al. 1995; Le Bouguenec et al. 2001; Hasan et al. 2002; Anderson et al. 2004)
Helicobacter pylori	CCPs 1 – 4	(O'Brien et al. 2006)

Table 1.4: Binding sites on DAF for various ligands

1.2.4 Membrane cofactor protein

Membrane cofactor protein (MCP, also called CD46) is expressed on the surface of nucleated human cells (Lublin et al. 1988) and contains four CCP modules. MCP exhibits only CA, *i.e.* acts as a cofactor for factor I-catalysed proteolysis of C3b and C4b molecules that are deposited on host cell surfaces and hence regulates both pathways of complement. Domain deletion/swap experiments in MCP indicate that for C4b-binding and CA, CCPs 2-4 are essential, although module 1 also contributes. CCP modules 3 and 4 bind C3b, but the additional presence of module 2 is required for C3b CA.

Ligand	Binding site	References
Complement component C3b	CCPs 2-4	(Adams et al. 1991; Iwata et al. 1995)
Complement component C4b	CCPs 1-4	(Adams et al. 1991; Iwata et al. 1995; Liszewski et al. 2000)
$\beta_1$integrins	-	(Lozahic et al. 2000)
Measles virus • Hemagglutinin	 CCPs 1, 2	 (Dorig et al. 1993; Naniche et al. 1993; Manchester et al. 1997)
Human herpesvirus 6		

• Complex of glycoprotein H, L and Q	CCPs 2, 3	(Santoro et al. 1999; Greenstone et al. 2002; Mori et al. 2003)
Group B and D adenoviruses • Fiber knob domain	CCPs 1, 2	(Gaggar et al. 2003; Segerman et al. 2003; Wu et al. 2004; Gaggar et al. 2005)
Neisseria gonorrhoeae and meningitidis • Type iv pili	CCP 3, 4 + serine-threonine-proline domain	(Kallstrom et al. 1997; Kallstrom et al. 1998; Kallstrom et al. 2001)
Group A Streptococcus • M-Protein	CCPs 3, 4	(Okada et al. 1995; Giannakis et al. 2002)

Table 1.5: Binding sites on MCP for various ligands

1.2.5 C4b-binding protein

C4b-binding protein (C4BP) is a soluble glycoprotein, which performs a parallel role to factor H, but acts predominantly upon the classical pathway convertase. It is the only polymeric member of the RCAs, having seven α-chains, each containing eight CCP modules and a single β-chain containing three CCP modules. In the case of C4BP, the three N-terminal CCP modules of the α-chain are required for C4b-binding and these are also the site for DAA and CA, whereas modules 1-4 have been implicated in weak C3b-binding.

Ligand	Binding site	References
Complement component C4b (C4c and C4dg fragments)	CCPs 1, 2, 3 of α-chain, cluster of charged AAs on interface CCP1-2	(Ogata et al. 1993; Accardo et al. 1996; Hardig et al. 1997; Blom et al. 1999; Blom et al. 2001a; Leung et al. 2006)
Complement component C3b	CCPs 1-4 of α-chain, cluster of charged AAs on interface	(Blom et al. 2003)

	CCP 1-2	
Glycosaminoglycans (GAGs)	CCPs 1, 2, 3 of α-chain, cluster of charged AAs on interface CCP1-2	(Blom et al. 1999; Blom et al. 2001a)
Protein S	CCPs 1, 2 of β-chain	(Hardig et al. 1993; Hardig and Dahlback 1996; Blom et al. 1998; van de Poel et al. 1999; Webb et al. 2001)
CD40	CCP 8 of α-chain + core region	(Brodeur et al. 2003)
Serum amyloid P component (SAP)	CCP 8 of α-chain + core region	(Garcia de Frutos et al. 1995)
DNA, RNA and necrotic cells	CCPs 1, 2 of α-chain, cluster of charged AAs on interface CCP1-2	(Trouw et al. 2005)
Streptococcus pyogenes • M-proteins	CCPs 1, 2 of α-chain	(Thern et al. 1995; Accardo et al. 1996; Blom et al. 2000a; Perez-Caballero et al. 2004; Jenkins et al. 2006)
Bordetella pertussis • Filamentous hemagglutinin	Cluster of charged AAs on interface CCP1-2	(Berggard et al. 2001)
Neisseria gonorrhoeae • Type iv pili • Por1A • Por1B	CCPs 1, 2 of α-chain CCP 1 of α-chain CCP 1 of α-chain	(Blom et al. 2001b; Ram et al. 2001)
Escherichia coli K1 • OmpA	CCP 3 of α-chain	(Prasadarao et al. 2002)
Moraxella catarrhalis • UspA1 • UspA2	CCPs 2, 5, 7 of α-chain CCPs 2, 5, 7 of α-chain	(Nordstrom et al. 2004)
Candida albicans	CCPs 1, 2 of α-chain	(Meri et al. 2004)
Borrelia recurrentis	-	(Meri et al. 2006)
Borrelia duttonii	-	(Meri et al. 2006)

Table 1.6: Binding sites on C4BP for various ligands

1.2.6 Complement receptor type 2

Complement receptor 2 is a glycoprotein of 15 or 16 CCP modules that does not regulate complement activation, but is located in the same gene cluster as the remaining five RCAs. It connects the innate and acquired immune responses (Matsumoto et al. 1991); C3d becomes covalently attached to antigens, and the subsequent recognition of C3d by CR2 on the surface of a B cell enhances induction of the immunoglobulin IgG response by three to four orders of magnitude (Dempsey et al. 1996).

Ligand	Binding site	References
Complement components C3d, C3dg	CCPs 1, 2	(Carel et al. 1990; Martin et al. 1991; Szakonyi et al. 2001)
Complement component iC3b	CCPs 1, 2	(Schwendinger et al. 1997)
CD23	CCPs 1, 2 + CCPs 5-8	(Aubry et al. 1994)
Interferon-α (IFN-α)	CCPs 1-4?	(Delcayre et al. 1991)
DNA	CCPs 1, 2	(Holers and Kulik 2006)
Epstein-Barr virus • Gp350, gp220	CCPs 1, 2	(Fingeroth et al. 1984)

Table 1.7: Binding sites on CR2 for various ligands

1.2.7 Vaccinia virus complement control protein

VCP was the first microbial protein identified to have complement inhibitory activity (Kotwal and Moss 1988; Kotwal et al. 1990), and is the smallest and simplest complement control protein (244 residues, four CCP modules), lacking glycosylation or other modifications. It has been reported to act as a cofactor for factor I cleavage of C3b and C4b (Kotwal et al. 1990; McKenzie et al. 1992) inhibiting both alternative and classical pathways of complement (Sahu et al. 1998) and also weak DAA activity (Mullick et al. 2005). Through this ability, VCP secreted from virally infected cells,

protects the host cell and released viral progeny from host complement attack (Isaacs et al. 1992; Howard et al. 1998).

Ligand	Binding site	References
Human complement component C3b (C3dg)	CCPs 1-4	(Rosengard et al. 1999; Bernet et al. 2004; Mullick et al. 2005)
Human complement component C4b (C4c)	CCPs 1-4	(Bernet et al. 2004; Mullick et al. 2005)
Complement convertases • C3bBb • C4b2a	 CCPs 1- 4 CCPs 1 – 4	(McKenzie et al. 1992; Mullick et al. 2005)
Glycosaminoglycans (GAGs)	CCPs 1,2 + CCP 4 (mainly)	(Smith et al. 2000; Murthy et al. 2001; Smith et al. 2003; Ganesh et al. 2004)
Suramin	CCP 4	(Ganesh et al. 2005)

Table 1.8: Binding sites on VCP for various ligands

1.3 Complement components C3, C4 and C5

Fundamental to the complement pathway are three large paralogous proteins C3, C4 and C5. These proteins emerged over 700 million years ago (Sunyer et al. 1998) and belong to the α2-macroglobulin (α2M) family, whose other members include the proteinase inhibitor α2M, the insect and nematode thioester-containing proteins (TEPs), and CD109 (Chapter 5). These proteins (~1400 – 1800 amino acid-residues) are characterised by homologous sequence features - including a unique thioester motif (not present in C5) - and have important roles in the immune response in metazoans, predating the emergence of immunoglobulins (Budd et al. 2004). These proteins are now recognised as being multiple-domain in nature (discussed further in Chapter 5).

During and following complement activation, C3, C4 and C5 are cut selectively by specific proteases at analogous Arg residues in the α-chain to create the anaphylatoxins, C3a, C4a and C5a, and the larger C3b, C4b and C5b fragments. Otherwise, despite high sequence similarities, these three proteins have distinct functions. The thioester allows activated C3b and C4b molecules to bind covalently to other molecules. Complement components C3, C4 and C5 are synthesised as single chains, but post-translational processing results in disulfide-linked multichain structures (two chains in C3 and C5, α and β; three chains in C4 α, β and γ via an additional tetra-arginine site).

1.4 Terminal components

The membrane attack complex is a large, membrane-bound protein complex that is the ultimate product of complement activation. It is assembled from five distinct soluble plasma proteins (C5b-C9). The terminal components C6 (125 kDa), C7 (120 kDa), C8 (150 kDa) and C9 (66 kDa) have almost entirely modular structures. C6 is composed of nine modules or domains, whereas C7, the C8α and C8β subunits, and C9, all resemble truncated forms of C6, with successively fewer modules (eight, four, four, and three, respectively) (Figure 1.3).

The five proteins are incorporated in a sequential process that begins with proteolytic activation of C5 at the target cell surface (see above, and Chapter 5). This generates a metastable activated C5b that forms the nucleus for subsequent addition of single molecules of C6, C7, C8, and finally, multiple copies of C9 to yield the lytic MAC. In addition to the irreversible post-cleavage complex (*i.e.* C5b.C6), C5 also binds reversibly to C6 forming the pre-activation complex (*i.e.* C5:C6), and to C7 via its FIMAC (Factor I membrane attack complex) domains (Arroyave and Mullereb.Hj 1973; Thai and Ogata 2004).

1.5 Complement and Disease

Despite the actions of the RCA protein family and other inhibitory proteins, such as CD59, that act elsewhere within the complement system, there remain numerous situations in which complement-mediated damage to self-tissues is pathologic. A particularly graphic example of complement's destructive power is provided by the events that follow xenotransplantation. Naturally occurring xenoreactive antibodies trigger a complement-mediated "hyperacute rejection", resulting in decimation of the xenograft within minutes. This phenomenon remains one of the major hurdles to a potentially revolutionary strategy for overcoming the chronic shortage of donor-organs.

Complement has also been implicated in the debilitating symptoms of a very long list of clinical conditions (Morgan and Harris 2003; Mizuno and Morgan 2004). Varying amounts of evidence exists for complement involvement in the following diseases: rheumatoid arthritis, systemic lupus erythematosus, most types of glomerulonephritis, atypical haemolytic uraemic syndrome (see Chapter 4, Section 4.1.1), autoimmune myocarditis, multiple sclerosis, myasthenia gravis, antiphospholipid syndrome, noncardiac myositis, type I diabetes mellitus, age-related macular degeneration (see Chapter 4, Section 4.1.2) and asthma. There is also evidence for an important role of complement in degenerative diseases, such as Alzheimer's disease, and in reperfusion following ischemia during myocardial infarction and stroke. Inappropriate complement activation can also be triggered by burns and by treatments such as cardiopulmonary bypass and haemodialysis. One in five people, according to some estimates, will have a complement-mediated symptom at some stage in their lives. Considerable effort has been invested in the search for

safe and effective agents to use in the therapeutic downregulation of the complement system.

1.5.1 Inhibiting Complement

Several candidate anti-complement drugs have made it as far as clinical trials. Only brief mention will be made here of the majority of the contenders, the focus of this thesis are the natural regulators (and their engineered derivatives).

1.5.1.1 Small molecule inhibitors

An excellent example of this approach is compstatin, a cyclic peptide derivative that is the product of a painstaking design process (Soulika et al. 2003), although the exact mechanism whereby compstatin acts is unknown. It is, however, a good convertase inhibitor, and its efficacy in this respect is manifested in several potential clinical applications – it is beneficial in animal models of cardiac surgery and cardiopulmonary surgery; it also shows promise in *ex vivo* models of xenotransplantation.

The 74-amino acid C5a fragment - produced upon cleavage of C5 by the C5 convertases - is an anaphylatoxin. This very potent proinflammatory peptide is one of the strongest chemotactic substances known. Acting via specialised receptors on their surfaces, it recruits neutrophils and macrophages to a site of complement activation. The potential for antagonising its receptor is obvious, and this has been explored most successfully using novel, small cyclic molecules derived from the C-terminus of C5a. Administration of such compounds to rats or mice by oral and other routes exhibited a broad range of anti-inflammatory activities (Woodruff et al. 2005). This work nicely demonstrates the therapeutic potential of anticomplement therapies.

1.5.1.2 Antibody-based therapies

An alternative approach is to block C5 cleavage production by means of derivatives of antibodies to C5. Two antibody-derived molecules - eculizumab and pexelizumab (Kaplan 2002; Whiss 2002) - are at an advanced stage of development for cardiopulmonary bypass (Fleisig and Verrier 2005), myocardial infarction and stroke, as well as chronic conditions such as rheumatoid arthritis, dermatomyosis and pemphigus. For example, a study showed that patients with the rare, occasionally fatal, blood disease, paroxysmal nocturnal haemoglobinuria (PNH), who received eculizumab maintained statistically significant reductions in red blood cell destruction and blood transfusions (Hillmen et al. 2004).

1.5.2 Exploiting the RCAs

A possible advantage of therapy based on inhibition at the level of C5 is that it leaves intact earlier stages of complement, including the generation of opsonins and immune clearance. Thus a patient taking anti-C5 therapy would not be entirely immunocompromised. If, on the other hand, the goal is to suppress all the principal outcomes of complement activation – immune clearance, the generation of all three anaphylatoxins, and the formation of the membrane attack complex – then the regulators of complement activation represent an excellent starting point.

An early, thoroughly investigated, example is soluble CR1 (sCR1) or TP10 (Rioux 2001). This recombinant protein contains all 30 of CR1's CCP modules encompassing functional site 1 and two copies of site 2. Nanomolar concentrations of TP10 are able to down-regulate C3 and C5 convertases, in both alternative and classical pathways of complement, via decay acceleration and cofactor activities. Thereby it diminishes production of opsonins, anaphylatoxins and membrane attack

complexes. In an exciting initial report (Weisman et al. 1990), TP10 reduced infarction size by 44% in a rat model of reperfusion injury of ischemic myocardium. Seventeen years later, the efficacy of TP10 in ischemia-reperfusion injury associated with cardiopulmonary bypass - the most promising application so far - is still the subject of ongoing investigation. A placebo-controlled, double-blind study involving 564 high-risk patients (Lazar et al. 2004) showed that TP10 significantly inhibited complement activity immediately after cardiopulmonary bypass, and this inhibition persisted for several days. While TP10 was clinically ineffective when considered over the entire population of patients, it did significantly decrease the incidence of mortality and myocardial infarction in male patients. In a much smaller (n = 15), related, phase I/II trial, the possible use of TP10 in infants undergoing cardiopulmonary bypass was investigated and no adverse events were attributable to TP10.

Another possible application of TP10 is in xenotransplantation: in pig-to-primate heart transplantation, a survival time of up to seven days (compared to 30-40 minutes for the control) under continuous treatment with TP10 was reported (Pruitt et al. 1997). On the other hand TP10 was apparently less successful in xenotransplantation when administered in conjunction with the use of transgenic donor pigs expressing human DAF (Azimzadeh et al. 2003; Lam et al. 2005), probably reflecting the multilayered complexity of the immune response and our lack of detailed knowledge of how complement is regulated in various tissues.

The very large size (~190 kDa) of TP10, the expense of its production, and its complexity mean it is unlikely to find wide application in humans. A smaller, more elegant solution might lie in a molecule termed APT070 (Smith 2002). This therapeutic compound consists only of the N-terminal three CCP modules of CR1,

followed by a positively charged peptide sequence (known as an "addressin"), and a myristoyl group. The peptide and fatty acid moieties help to direct the CCP module component to the phospholipid bilayer and hold it there – their addition significantly enhances the complement inhibitory properties of the triple CCP module fragment alone. In some respects therefore, APT070 resembles GPI-anchored DAF. Indeed, the three CCP modules encompass functional "site 1" of CR1. APT070 prevents complement-mediated tissue damage in animal models of rheumatoid arthritis, renal transplant reperfusion injury and vascular shock (Smith 2002). Most recently (Halstead et al. 2005), it has been investigated as a possible therapy in Guillain-Barré syndromes. Despite these promising results, the company that currently possesses the rights to APT070 (Inflazyme Pharmaceuticals, http://www.inflazyme.com) is no longer pursuing its development.

Building on the idea that engineered complement regulators are effective anti-complement agents, the possible uses of hybrids with antibodies have been explored (Morgan and Harris 2003). In some cases, the antibody Fab arms are attached to the complement regulator in the hope that they will target the attached regulator to a specific site. Alternatively, regulator molecules may be attached to the Fc portion of an antibody, thus they form the equivalents to the Fab arms, and the Fc helps to extend the plasma half-life. Some fusion proteins act as prodrugs in that the full inhibitory potential of the attached regulator may only be revealed by proteolysis, thus providing an additional opportunity for targeting specific sites.

Finally, mention must be made of VCP and its relative from the small pox (variola) virus inhibitor of complement enzymes (SPICE). Neither VCP nor SPICE is glycosylated - both consist entirely of four CCPs. Of all the RCAs, VCP and SPICE therefore have the simplest composition. VCP is also the best-characterised RCA

from a structural standpoint, initially from NMR studies (Wiles et al. 1997; Henderson et al. 2001), and latterly by X-ray crystallography (Murthy et al. 2001). Furthermore, 3D structures have been determined of VCP in complex with two ligands – heparin (Ganesh et al. 2004) and suramin (Ganesh et al. 2005). Unfortunately, a commensurate body of site directed mutagenesis data is not available. Nonetheless, domain deletion experiments (Smith et al. 2003) demonstrate that all four CCP modules are required, while module 4 contains the main heparin-binding site. Also of interest is a functional and sequence comparison with SPICE (Rosengard et al. 2002). Neither VCP nor SPICE possess potent decay accelerating activity, but they both have cofactor activity for factor I-catalysed proteolysis of C3b and C4b. SPICE is a better cofactor than VCP, and the cofactor activity of SPICE is ~25-50% that of sCR1. Thus the putative, single functional site in SPICE is comparable in potency with each of the three functional sites in sCR1. There are eleven amino acid residue differences between SPICE and VCP, all located in CCPs 2, 3 and 4. None of the amino acid residues that differ between the two proteins are near intermodular junctions and all are surface exposed, implying no significant structural differences between the two. Clearly, some or all of the eleven substitutions in SPICE compared to VCP are contact residues – either for C3b/C4b or factor I and they warrant further inspection. It is striking that VCP can bind heparin and simultaneously inhibit complement activation via its DAA and CA, implying that it has the ability to protect surfaces (Murthy et al. 2001). SPICE binds heparin equally well. These properties have encouraged the hope that VCP (and SPICE) could be used therapeutically in humans, despite their probable immunogenicity. In support of this, the possible benefits of VCP in numerous animal models of disease has been shown (Jha and Kotwal 2003).

1.5.3 Future of Anti-Complement Therapy

Engineered RCA proteins (or viral proteins used without modification) have great therapeutic potential as a consequence of their efficacy in blocking all the major outcomes of a complement activation event, and their specificity. This is illustrated to some extent by the relative success of sCR1, although it looks as if this agent will only be used in limited and specific circumstances where it can be physically delivered to where it is needed – *e.g.* during cardiac surgery. The ~190 kDa sCR1 was not rationally designed; since the entire extracellular portion of CR1 was used, no knowledge of structure or function was applied. A significant advance was the truncation of sCR1 to three CCP modules and attachment of a targeting entity to create APT070. Although this product is not currently being developed for the clinic, it very nicely illustrates the benefits of rational design based on knowledge of the location of functional sites in CR1, and on understanding of the improved effectiveness of surface associated complement regulators. APT070 additionally exemplifies the advantages of a targeting strategy, as was also evident from the improved efficacy of RCA-antibody fusions.

That SPICE (and to a lesser extent, VCP) can encapsulate much of the (so-far identified) activity of sCR1 in just four CCP modules suggests that a simple truncation, as was done in the creation of APT070, does not go far enough. Indeed, a strategy not yet well explored, is the use of knowledge-based site directed mutagenesis as a means of tailoring the activities of the CCP modules within the regulatory portions of hybrid molecules such as TP10 and APT070. This lack of progress is probably attributable to the dearth of knowledge regarding mechanism at the atomic level despite a wealth of structural information. The results of mutagenesis, however, while they are not easy to interpret for reasons described

earlier, are tantalising. For example, single amino acid residue swaps between functional sites 1 and 2 of CR1 were sufficient to delete the cofactor activity in site 2 and introduce it into site 1 (Krych-Goldberg and Atkinson 2001a). Numerous other mutagenesis studies support the suggestion that significant decreases and increases in inhibitory activity may be brought about with a few judiciously selected amino acid substitutions. Because there appear to be different sites (in CR1 for example) for cofactor activity and decay accelerating activity, and non-identity of C3b- and C4b-interaction sites (for example, in DAF), it is also possible to tailor an RCA to selectively inhibit a specific step in the complement cascade. The fusion of much more powerful and selective complement regulatory components with the elegant solutions for targeting and extending plasma half-lives described above could yield a new generation of powerful complement inhibitors able to act at much lower plasma concentrations than the present ones. Currently, however, there are far too many gaps in our knowledge of how the RCAs work at the atomic level, and this is holding back rational design of highly potent and specific inhibitors.

1.6 Structure determination of proteins in complement

Even though the complement system was discovered more than a century ago, the first three-dimensional (3D) structure of a complement component, that of a fragment of human C3a and desArg-C3a, was determined only in 1980 by Huber et al., by X-ray diffraction (Huber et al. 1980). Since then there has been good progress in determining 3D structures of complement proteins (Table 1.9; RCAs shown in Figure 1.2).

Protein[a]	PDB ID	Species	Reference
C1q	1PK6^	*Homo sapiens*	(Gaboriaud et al. 2003)
C1q	1C28^	*Mus musculus*	(Shapiro and Scherer 1998)
C1r	1GPZ^1,2, 1MD7^2, 1MD8^2, 1APQ*	*Homo sapiens*	(Bersch et al. 1998; Budayova-Spano et al. 2002a; Budayova-Spano et al. 2002b)
C1s	1ELV^2, 1NZI^	*Homo sapiens*	(Gaboriaud et al. 2000; Gregory et al. 2003)
MASP-2	1Q3X^2, 1ZJK^1,2	*Homo sapiens*	(Harmat et al. 2004; Gal et al. 2005)
MASP-2	1NT0^	*Rattus norvegicus*	(Feinberg et al. 2003)
C3	2A73^	*Homo sapiens*	(Janssen et al. 2005)
C3	2B39^	*Bos taurus*	(Fredslund et al. 2006)
C3b	2HR0^, 2I07^, 2ICE^, 2ICF^	*Homo sapiens*	(Abdul Ajees et al. 2006; Janssen et al. 2006; Wiesmann et al. 2006)
C3c	2A74^	*Homo sapiens*	(Janssen et al. 2005)
C3d	1C3D^, 1GHQ^(in complex with CR2)	*Homo sapiens*	(Nagar et al. 1998; Szakonyi et al. 2001)
C3dg	1QQF^, 1QSJ^	*Rattus norvegicus*	(Zanotti et al. 2000)
C4Adg	1HZF^	*Homo sapiens*	(van den Elsen et al. 2002)
C5a[c]	1KJS*	*Homo sapiens*	(Zhang et al. 1997)
C5a-desArg	1C5A*	*Sus scrofa*	(Williamson and Madison 1990)

C5-C345C	1XWE*	*Homo sapiens*	(Bramham et al. 2005a)
C8γ	1IW2^ (pH 7), 1LF7^ (pH 4)	*Homo sapiens*	(Ortlund et al. 2002)
Crry	1NTL#1-5	*Mus musculus*	(Aslam et al. 2003)
Crry	1NTJ#1-5	*Rattus norvegicus*	(Aslam et al. 2003)
Factor B	1DLE^, 1Q0P^	*Homo sapiens*	(Jing et al. 2000; Bhattacharya et al. 2004)
Factor D	1DSU^, 1HFD^, 1BIO^, 1DFP^, 1DIC^, 1DST^, 1FDP^	*Homo sapiens*	(Narayana et al. 1994; Kim et al. 1995a; Cole et al. 1997; Cole et al. 1998; Jing et al. 1998; Jing et al. 1999)
CR1	1GKN*15,16, 1GKG*16,17, 1PPQ*16	*Homo sapiens*	(Smith et al. 2002; O'Leary et al. 2004)
CR2	1GHQ^1,2(in complex with C3d), 1LY2^1,2, 2ATY#1,2, 1W2R#1,2(in complex with C3d)	*Homo sapiens*	(Szakonyi et al. 2001; Prota et al. 2002; Gilbert et al. 2005; Gilbert et al. 2006a)
CR3	1D2L*	*Homo sapiens*	(Dolmer et al. 2000)
Factor H[b]	1HFH*15,16, 1HFI*15, 1HCC*16, 2BZM*19,20, 1HAQ#	*Homo sapiens*	(Norman et al. 1991; Barlow et al. 1993; Aslam and Perkins 2001; Herbert et al. 2006c)
DAF	1H03^3,4, 1H04^3,4, 1H2P^3,4, 1H2Q^3,4, 1UOT^3,4, 1OJV^1-4, 1OJW^1-4, 1OJY^1-4, 1OK1^1-4, 1OK2^1-4, 1OK3^1-4, 1OK9^1-4, 1NWV*2,3, 1M11&1-4(in complex with echovirus), 1UPN&1-4(in complex with echovirus), 2C8I&1-4(in complex with echovirus)	*Homo sapiens*	(He et al. 2002; Uhrinova et al. 2003; Williams et al. 2003; Bhella et al. 2004; Lukacik et al. 2004; Pettigrew et al. 2006)
MCP	1CKL^1,2	*Homo sapiens*	(Casasnovas et al. 1999)
C4BP	2A55*1,2	*Homo sapiens*	(Jenkins et al. 2006)
CD59	1CDQ*, 1CDR*, 1CDS*, 1ERG*, 1ERH*	*Homo sapiens*	(Fletcher et al. 1994; Kieffer et al. 1994)
VCP	1G40^1-4, 1G44^1-4, 1VVC*3,4, 1VVD*3,4, 1VVE*3,4, 1E5G*2,3, 1RID^1-4, 1Y8E^1-4	*Vaccinia virus*	(Wiles et al. 1997; Henderson et al. 2001; Murthy et al. 2001; Ganesh et al. 2004; Ganesh et al. 2005)

Table 1.9: Structures of complement proteins deposited in the PDB (as on Dec, 2006). Superscripts used in table: [a] No discrimination is made between intact whole proteins or fragments. In the case of CCP-module-containing complement proteins, CCP modules solved are indicated as

subscript. [b] Factor H CCP module 5 has been determined by NMR (Barlow et al. 1992), but not deposited into the PDB. [c] C3a and desArg-C3a fragments have been determined by X-ray diffraction (Huber et al. 1980), but not deposited into the PDB. ^ Structures solved by X-ray diffraction, * Structures solved by NMR, # Structure solved by X-ray and neutron scattering, & Structure solved by cryo-electron microscopy.

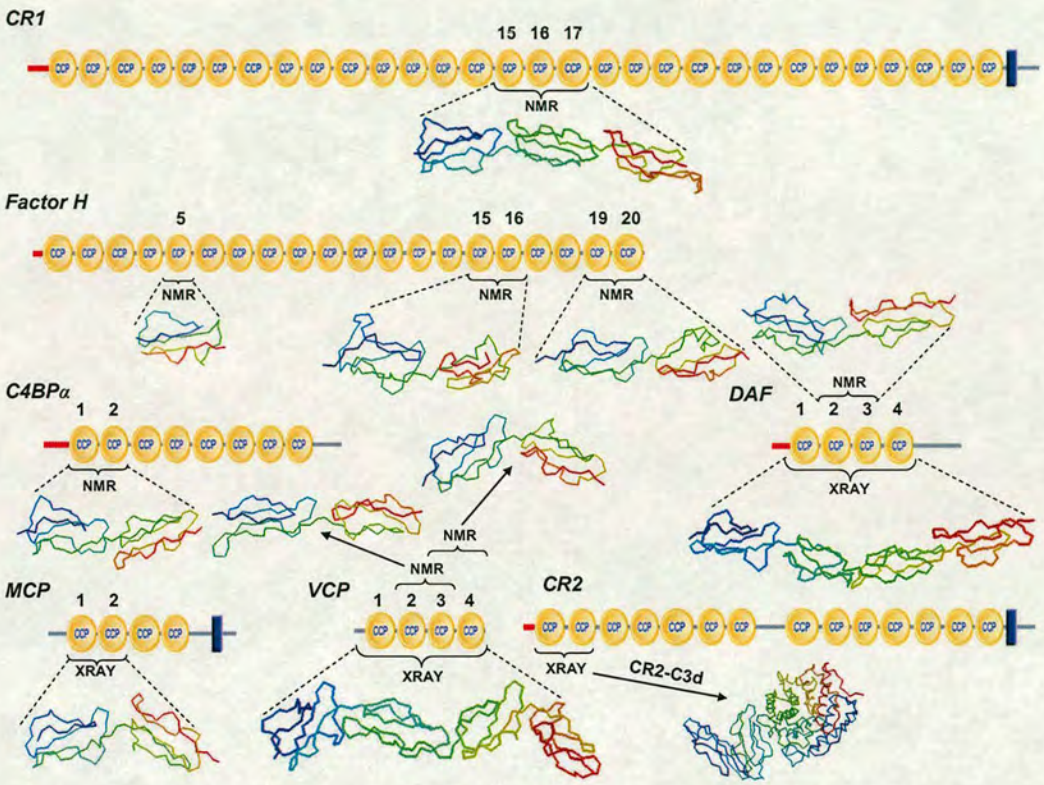


Figure 1.2: Available structural information for the RCA family (including CR2) and the viral mimic, VCP. Where they are known, 3D structures of modules are shown as backbone traces, and their relative positions within the protein are shown diagrammatically. Figure adapted from (Brook et al. 2005).

Most structures have been determined using high resolution techniques, X-ray diffraction and nuclear magnetic resonance (NMR), however very few structures are derived from the complete native intact protein. Many have been determined by more than one technique, and in some cases by the same technique in different conditions. Some structures have been solved as complexes and multimers, while others as smaller fragments, domains or modules, mutants, structures with various bound

ligands or inhibitors, structures with various sample or crystallisation conditions (*e.g.* pH), structures of proteins crystallised in various space groups, proteins from various species, or NMR structures deposited as ensembles of low-energy structures or in one case as an average minimised structure. There have also been low resolution X-ray and neutron solution scattering structures and cryoelectron microscopy structures deposited in the Protein Data Bank (PDB). There still however remain large gaps in our structural knowledge and understanding of various proteins. Modelling can play an important role in filling these gaps. In particular, the work on the structure of RCA proteins over the last 15 years (Figure 1.2), and the very recent structural breakthroughs on C3 and its fragments, afford the opportunity to apply homology modelling to a sizeable proportion of proteins in the complement system. This thesis describes efforts in this direction and discusses the utility of the models thus created.

1.7 Summary aims of Ph.D. project

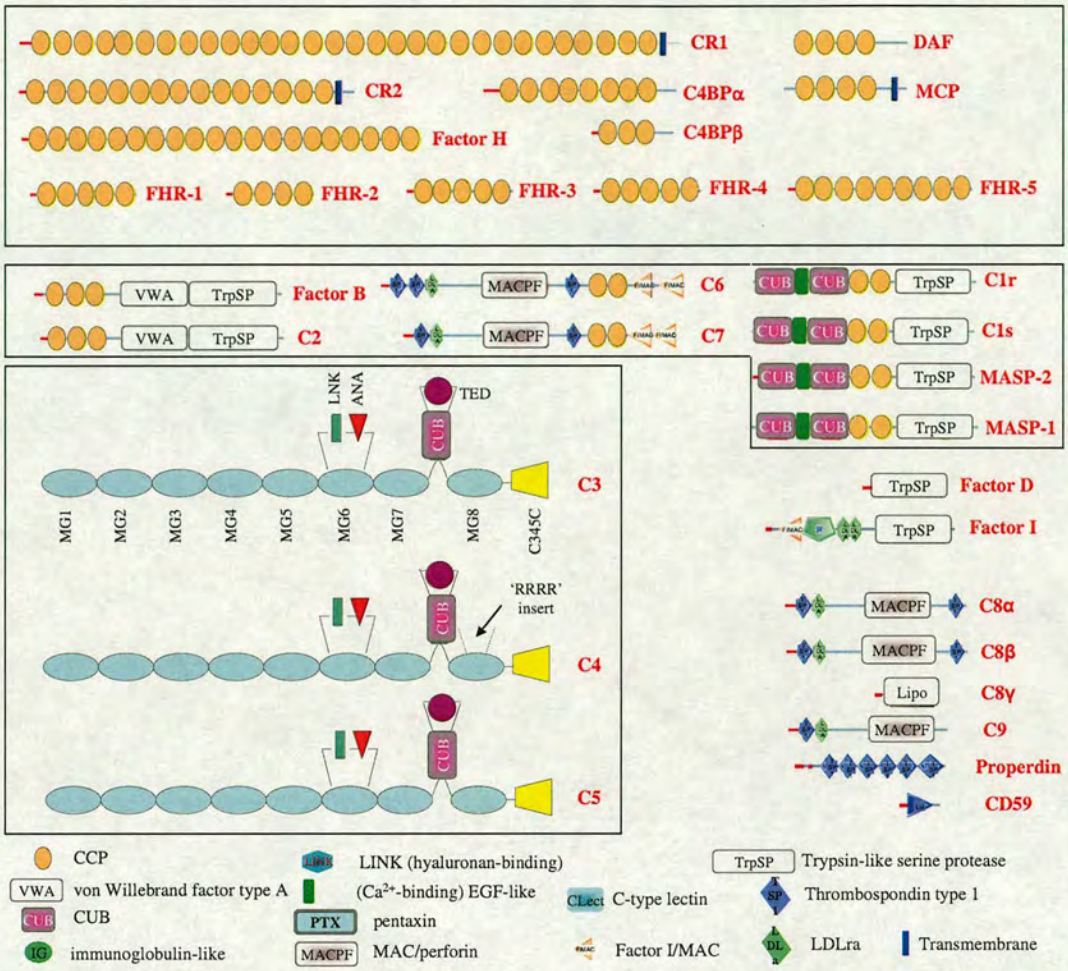


Figure 1.3: Complement protein targets for modelling. Modular architecture of the main complement components are shown and labelled. The target proteins for modelling; *i.e.* the CCP modules within the complement system and the C3, C4 and C5 proteins are shown boxed. Unless labelled on the figure, all domains are shown within key. Abbreviations for the complement components are explained in text.

- To undertake comprehensive bioinformatics analysis on the sequence and structure of CCP modules, including analyses of intermodular orientations among experimentally determined structures; to discover consensus and novel trends (Chapter 2).

- To develop and implement a large-scale comparative modelling protocol for individual human CCP-modules (Chapter 3).

- To undertake surface analyses of CCP module structures, using complement receptor type 1 as an example (Chapter 3).

- To set up and maintain an up-to-date online resource on CCP modules and store and make publicly available the models (Chapter 3).

- To infer the consequence of numerous disease-associated sequence variations on the basis of the models derived from the large-scale protocol, and additionally through other newly created ones (Chapter 4).

- To provide novel insights into the latter stages of complement through comparative models of complement component C5 and C5b (Chapter 5).

1.8 Declarations and Acknowledgements

- Section 1.5 of this introduction appeared in:

Brook, E., Herbert, A. P., Jenkins, H. T., **Soares, D. C.**, and Barlow, P. N. (2005). Opportunities for new therapies based on the natural regulators of complement activation. Ann N Y Acad Sci *1056*, 176-188.

Chapter 2

A bioinformatics analysis of the sequence, structure and dynamics of CCP-modules

---

2.1 Preamble to survey

A very striking feature of proteins that belong to the "regulators of complement activation" (RCA) family (Chapter 1, Section 1.2) is the presence of numerous examples of a repeating motif of ~60 amino acids that was originally called the "short consensus repeat" (SCR) (Morley and Campbell 1984; Reid and Day 1989). We now know that each repeat corresponds to a structural unit called a "complement control protein" (CCP) module. The CCP module has also been called the "sushi domain" reflecting the limited knowledge of its three-dimensional (3D) structure in the past (Ichinose et al. 1990). In addition to its pre-eminence in the RCA family, the CCP module is found in other proteins within the complement system, and in a wide range of non-complement proteins (http://smart.embl-heidelberg.de).

Strings of between four and 30 CCP modules joined by short linking segments are found within all members of the RCA family. These proteins are expressed by a cluster of genes located on the long arm of chromosome 1 (1q32). Their primary role is to ensure a complement-mediated response to infection is targeted and proportionate (as discussed in Chapter 1, Section 1.2). They interact, via binding sites that entail between two and four CCP modules, with components of the C3 and C5

convertase complexes (Chapter 1, Tables 1.2 to 1.6). By blocking formation of new convertases, accelerating the dissociation of existing convertases, and acting as cofactors for proteolytic degradation of the dissociated components, the RCA proteins negatively regulate the complement cascade. All host cells exposed to serum have RCA proteins embedded in, attached to, or associated with, their surfaces to protect themselves against complement. Many of the RCA proteins have numerous additional binding partners (listed in Chapter 1). In some cases these include proteins expressed by pathogens in an effort to anchor themselves to a host cell, to subvert the host's complement system, or to exploit the signal transduction properties of several membrane-associated RCAs.

To date, no large-scale bioinformatics analysis has been performed on this family of proteins and their component CCP modules. Hence, several questions remain unanswered. Just how widespread are CCP modules among humans and non-humans? Are there any functional correlations among the proteins that contain them? How do the "many" experimentally-determined CCP module structures compare with one another? Where do the most significant structural differences among them lie?

It has been established that many RCA proteins are predominantly extended molecules that have a large surface area relative to their molecular weight available for potential interactions with binding partners. In general, each RCA protein has multiple binding sites (Chapter 1), and each binding site encompasses surface elements from two or more neighbouring CCP modules. Hence, simultaneous occupation of two or more binding sites on the same protein molecule, and interplay between them, may be required for biological function of some RCA proteins. In the light of these observations, it is also important to consider the way in which each CCP module is attached to its neighbours. This raises several further important questions -

are there well-structured "junctions" between modules that result in rigid, defined intermodular angles? Or are junctions open and flexible allowing unrestricted intermodular movement? Do the modules that contribute surface elements towards a common binding site rearrange themselves during binding to a ligand? These issues bear upon mechanisms of binding and biological action. They could also contribute to the design of therapeutic versions of the RCAs (see Chapter 1, Section 1.5.2 and 1.5.3).

In the current work, for the first time, the above questions are addressed in the context of an overview of the literature and comprehensive "all-against-all" sequence and structure comparisons of CCP modules. Analyses of the inter-module orientations within module pairs and larger fragments of CCP modules are also performed; these are useful for the design of mutagenesis experiments and the interpretation of functional data. The emphasis of this survey is on CCP modules of the following RCA proteins: decay accelerating factor (DAF, CD55), complement receptor type 1 (CR1, CD35), membrane cofactor protein (MCP, CD46), the factor H (fH) family, C4-binding protein (C4BP); and on a viral mimic of the mammalian RCA proteins, Vaccinia virus complement control protein (VCP). Another CCP-containing complement protein – complement receptor type 2 (CR2, CD21) – is encoded by a gene in the same region, and normally regarded as a member of the RCA family even though it is not involved in regulation of complement activation. The blood-clotting factor XIII b subunit (FXIIIb) is built from CCP modules, and its gene is also located in the 1q32 cluster; but it is not regarded as an RCA since it does not interact with complement proteins, and hence it will not be dealt with in this chapter.

2.2 Materials and methods

2.2.1 CCP module sequence retrieval, analysis and alignment

CCP modules were identified using a combination of literature and database searches. The main database that was queried was the Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de) (Schultz et al. 1998; Letunic et al. 2002; Letunic et al. 2004), which identifies, annotates and presents domain architecture from sequence information. Several hundred sequences were identified as putative CCP modules on the basis of this search. This list of sequences was then manually rendered non-redundant, to discard duplicates, and fragments of modules were also eliminated. Splice variants were identified. The highly curated sequence database, SwissProt and its supplement, TrEMBL (http://www.expasy.org/sprot/) (Boeckmann et al. 2003) were then used to retrieve sequences corresponding to the identified putative CCP-modules from humans and non-humans. Domain boundaries were defined, on the basis of inspection of CCP 3D-structures, as one residue before the first Cys and three residues after the fourth Cys.

The multiple sequence alignment for the 84 individual human RCA CCP-modules was created using ClustalX (Thompson et al. 1997). The resultant alignment was further manually edited. Editing was based upon conservation of amino acid residues and secondary structure elements as predicted by PsiPred (McGuffin et al. 2000), and as determined from 3D-structures of CCP modules in the Protein Data Bank (PDB) (Berman et al. 2000) to place gaps most appropriately.

2.2.2 All-against-all pairwise sequence comparison amongst CCP modules

Two sets of pairwise sequence comparisons were performed. The first set (88 sequences) of comparisons included sequences from the 84 individual CCP modules of the six proteins in the RCA family- CR1 (30 CCPs), CR2 (15 CCPs), C4BPα (8 CCPs) and β (3 CCPs), DAF (4 CCPs), MCP (4 CCPs), factor H (20 CCPs), and additionally four CCPs from Vaccinia virus, VCP. The second set (48 sequences) encompassed sequences from the factor H family, including the 20 individual CCP module sequences from factor H, and additionally, 28 CCP module sequences from factor H-related proteins (FHRs); FHR1 (5 CCPs), FHR2 (4 CCPs), FHR3 (5 CCPs), FHR4 (5 CCPs), and FHR5 (9 CCPs).

The sequences (SwissProt Accession Nos.; CR1: P17927, CR2: P20023, C4BPα: P04003, C4BPβ: Q9BS25, DAF: P08174, MCP: P15529, factor H: P08603, FHR1: Q03591, FHR2: P36980, FHR3: Q02985, FHR4: Q92496, FHR5: Q9BXR6, and VCP: P10998) were downloaded from SwissProt and saved into two files locally in FASTA format. The CCP-module specific sequence files were then formatted using the '*formatdb*' program, which comes as part of the stand-alone BLAST executable (Altschul et al. 1990; Altschul et al. 1997) (http://www.comparative-legumes.org/lis/blast/AboutBLAST2.html#standalone). The '*formatdb*' command must be used in order to format protein source databases before these databases can be searched by '*blastall*', *i.e.* it turns FASTA files into BLAST readable databases.

formatdb –i /usr/progs/BLAST/data/.fsa*
(where '' = any database filename in FASTA format)*

An automated BLAST was run separately for the 88- and the 48-sequence databases such that each sequence was compared against every other sequence in the

respective databases. The *'blastall'* command, which incorporates the *'blastp'* program, was implemented and the output file *'OUTPUT'* saved after the search was performed on the *'INPUT'*.

*blastall –p blastp –d *.fsa –i INPUT –e 10.0 > OUTPUT*
[where '.fsa' = database of 88 or 48 sequences; 'INPUT' = input file (same as database file); 'OUTPUT' = output file;' –e 10.0' = evalue cut-off]*

The results obtained were then parsed and extracted using a PERL script (Appendix X). To enable this PERL script to run, a Boulder-BLAST package, which is an already available program to aid in parsing and reading of BLAST files, was installed. The results were then saved in a Microsoft Excel Workbook (Appendix CD).

2.2.3 Structural comparisons among experimentally determined CCP modules

Combinatorial Extension (CE) was used to compare individual CCP modules against each other. CE is a method for calculating pairwise structure alignments using characteristics of their local geometry as defined by vectors between Cα positions (Shindyalov and Bourne 1998). Thirty-four individual experimentally determined CCP module structures were first parsed from their parent protein using coordinates downloaded from the PDB. In the case of those modules where more than one structure exists, or structures have been solved by more than one technique, the highest resolution crystal structure was used. In the case of NMR ensembles, the closest-to-mean, lowest energy or average minimised structure was used. In cases

where both liganded and unliganded structures existed, the unliganded structure was adopted. For each module, domain boundaries were considered from one residue before the first Cys to three residues after the last Cys (as with the sequence comparisons), where available in the structural co-ordinates. A PERL script (Appendix X) was used to automate CE to compute an all-against-all pairwise structure alignment, among individual CCP module structures.

The global overlay of all 34 individual CCP-modules was performed using the program, Multiprot (Shatsky et al. 2002; 2004). MultiProt superposes the input PDB files based on the common geometrical cores between them, and derives multiple alignments from simultaneous superpositions of input molecules. It is a sequence-independent method of multiple structure alignment.

2.2.4 Calculation of buried surface area

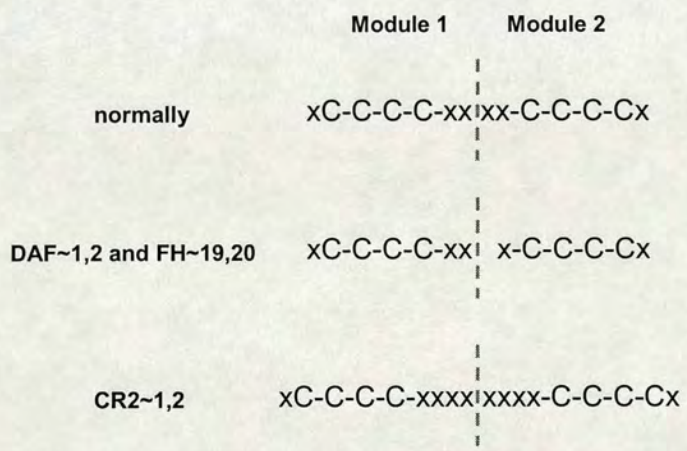


Figure 2.1: Schematic showing module and bimodule boundaries. The schematic indicates the individual module and bimodule pair boundaries considered for surface area calculation, where x = any residue; the four cysteine residues in each module are shown (C); and the dashed line indicates the position within the linker where the modules are separated. The scheme for the exceptions to the general case, *i.e.* DAF~1,2, FH~19,20 and CR2~1,2 are shown.

This was estimated by first calculating the Surface Area (SA) of a CCP module pair 12 (SA Bimodule^{12}), then repeating the calculation for each of the modules in turn (SA Module^{1} and SA Module^{2}). The web server "GETAREA 1.1" (Fraczkiewicz and Braun 1998) (http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl) was used for surface area calculations, using default parameters (radius of water probe in Angstroms = 1.4; output level = only total area). The surface area (SA) that was buried was calculated as: (SA Module^{1} + SA Module^{2}) - SA Bimodule^{12}, and therefore includes contacts between atoms belonging to the two halves of the linker. For Module 1 (in a four-residue linked module pair), module boundaries were considered to encompass one residue before the first CYS to two residues after the last CYS; and for Module 2 (in a four-residue linked module pair), from two residues before the first CYS until one residue after the last CYS. For a module pair, boundaries were considered from one residue before the first CYS of Module 1 to one residue after the last CYS of Module 2 (Figure 2.1). All the linker lengths for the set of solved structures have four residues, except in the cases of CR2~1,2 (where linker length = 8 residues) and DAF~1,2 and factor H~19,20 (where linker length = 3 residues). For CR2~1,2, four residues after the last CYS of Module 1 and four residues before the first CYS of Module 2 were taken as boundaries for individual modules. For DAF~1,2 and factor H~19,20, two residues after the last CYS of Module 1 (and one residue before the first CYS); and one residue before the first CYS of Module 2 (and one residue after the last CYS) were taken as boundaries for individual modules (Figure 2.1).

As with CE (Section 2.2.3), for structures solved by X-ray diffraction, the highest resolution crystal structure was used, or in cases where both unliganded and complexed structures were available, the unliganded form was used for surface area calculation. In the case of structures solved by NMR, the closest to mean, average

minimised structure or the lowest energy structure was used. X-ray structures were protonated before calculations. The IL-2Rα module pair (PDB code: 2B5I) was excluded from this analysis, because of its unique, irregular appearance (discussed in Chapter 3).

2.2.5 *Intermodular angle calculations*

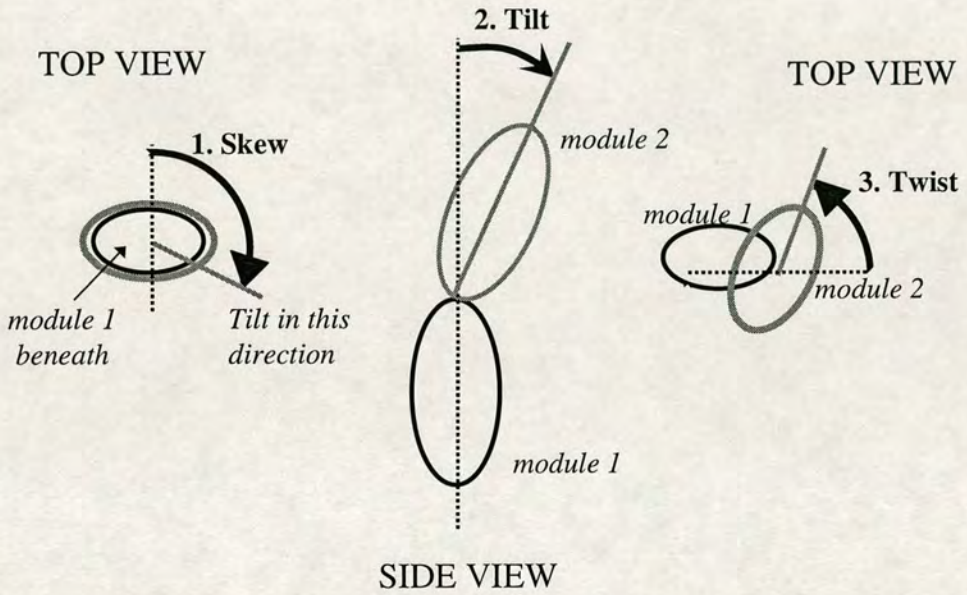


Figure 2.2: Intermodular arrangements. Schematic of skew, tilt and twist angles that define the orientation of module 2 (grey) with respect to module 1 (black). These three angles are necessary and sufficient to reproduce, in three steps, the experimentally determined orientation of module 2 relative to module 1 starting from a theoretical conformation in which the two modules are aligned, with module 2 above and module 1 below (see left-hand panel).

Intermodular orientations may briefly be described by a tilt angle and an angle of twist. These two angles are sufficient to describe how one module may be rotated relative to its neighbour about two defined (with respect to some feature that is common to both modules) orthogonal axes in such a way that the two modules become aligned end-to-end with one another (Bork et al. 1996). These two angles

alone are not, however, sufficient to describe the reverse process – how to restore a unique arrangement of two CCP modules starting from the case where both are aligned end-to-end. This is because it would be impossible to know on the basis of just that information in which direction to tilt the second module with respect to the first even though the angle of the tilt itself is known. The direction may be defined by a third angle – skew (Barlow et al. 1993) (see Figure 2.2). Finally, because the linker varies in length another parameter – distance (between centres of neighbouring modules) - is sometimes helpful in describing an intermodular orientation.

Tilt, twist and skew angles were calculated using the program 'mod22' (written by Bruno Kieffer, University of Strasbourg – first applied in (Barlow et al. 1993)) for all experimentally determined pairs and larger fragments of CCP-module containing proteins, where one, or more structures were available – *i.e.* for an ensemble of NMR-derived structures or where several molecules present in the unit cell of a crystal exist. Where several X-ray structures are in the database for the same protein, the structure of highest resolution was used. This program was first applied to the module pair of factor H~15,16 (Barlow et al. 1993), and was hence used in this exercise to assess intermodular angles in a uniform manner for all module pairs and larger fragments. Module boundaries were considered as Cys 1 and 4 for each module, and the Cα-atom of the consensus Trp in each module. In each case, the z-axis was defined by the principal inertia tensor, and the x-axis was defined with reference to the Cα-atom of the consensus Trp residue in each module (Leu in the case of FH~20).

2.3 Results and discussion

2.3.1 Occurrence of CCP modules

2.3.1.1 Modular composition of the RCA proteins

Many proteins of the complement system are built up from a limited range of module-types. The CCP module is very strongly represented with 69 occurrences amongst the most common allotypes of the five proteins that have been shown to regulate complement and belong to the RCA family: MCP, CR1, fH, DAF, and C4BP (see Figure 2.3). If all 19 CCP module-containing complement proteins arising from unique genes (including the fH-related proteins – see below) are included, the total is 130.

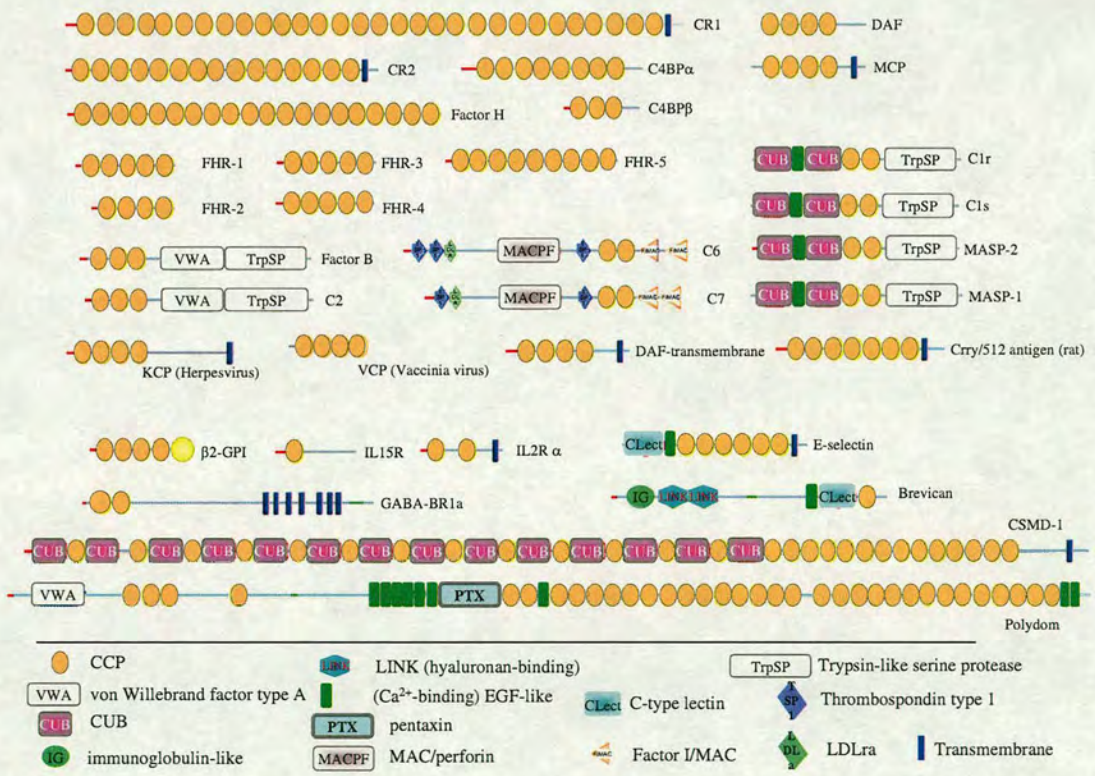


Figure 2.3: Occurrence of CCP modules in complement and other proteins. This figure was compiled on the basis of information from the SMART database (Schultz et al. 1998; Letunic et al. 2002; Letunic et al. 2004) (http://smart.embl-heidelberg.de/). The upper part shows proteins of the human complement system, and a small selection of non-human complement regulatory proteins. The

lower part illustrates just a few of the numerous mammalian non-complement system proteins that contain CCP modules. Abbreviations used: FHR – factor H-related protein; VCP and KCP - Vaccinia and Kaposi's virus complement control proteins; IL2R and IL15R – interleukin 2 and 15 receptors (α-chain); GABA-BR1a – $GABA_B$ receptor type 1 α-chain; CSMD - CUB and sushi multiple domain protein 1; other abbreviations are explained in the text. Symbols are explained in the key (abbreviations used in key: CUB = first found in C1r, C1s, uEGF and bone morphogenetic protein; LDLa = low density lipoprotein receptor domain class A; EGF = epidermal growth factor; MAC = membrane attack complex.)

The five complement regulators listed above, along with the fH-related proteins and CR2, are unusual in that they consist entirely, or almost entirely, of a single module type. Other CCP module-containing complement proteins, such as C1r, C1s, C2, factor B, mannan-binding lectin-associated serine protease (MASP) 1, MASP 2, C6 and C7, are more typical of extracellular multiple-module proteins since they contain mixtures of several module-types (see Figure 2.3) and are said to be "mosaic" (Bork et al. 1996).

2.3.1.2 Splice variants

Figure 2.3 shows only the most common splice-variants of each human protein. A fH-like protein (FHL-1) contains the seven N-terminal CCP modules of fH and a C-terminal extension of four residues (Zipfel et al. 1999). CR1 exists as shorter (23 modules) and longer (37 and 44 modules) versions that differ by the presence or absence of seven-module-long blocks called long homologous repeats (LHRs) – thus in the most common variant there are four such LHRs (LHR-A - LHR-D) plus two additional modules close to the C-terminus and the membrane. CR2 exists as a 15 CCP-module variant (shown in Figure 2.3) and as a 16-module version with an additional module, number 11 (Barel et al. 1998). Splice variants of human MCP and

DAF differ in their serine-threonine-proline extension, and in the nature of their membrane attachment, not in the number of CCP modules (Post et al. 1990; Liszewski et al. 1994). There do not appear to be any splice-variants of human C4BP α- or β-chains, but two minor forms have been observed with different combinations of α- and β-chains (Criado Garcia et al. 1995). The most common form of C4BP exist in the blood as a complex with protein S (Dahlback et al. 1983; Hillarp and Dahlback 1988).

2.3.1.3 Other forms of RCAs in non-humans

There is variation, even between closely related species, in the set of proteins used for immune adherence (a function of erythrocyte-borne CR1 in humans) and to regulate complement. For example baboon and chimp erythrocytes carry versions of CR1 composed of seven and eight CCP modules respectively (Birmingham and Hebert 2001). The first five modules of baboon CR1, and the first six of chimp CR1, are highly similar to the equivalent modules of human CR1; the sequence of the last module in both cases resembles that of human CR1 module 21. Mouse versions of CR1 and CR2 are encoded by the same gene, alternatively spliced to generate a form of CR1 (21 modules) or CR2 (15 modules) (Kurtz et al. 1990). Mouse tissues also express complement receptor-1 related gene/protein Y (Crry) that has decay accelerating and membrane cofactor activities (Kim et al. 1995b) but does not have an immune adherence role (Molina et al. 1992). It has five CCPs that are similar in sequence to the five N-terminal modules of human CR1. Crry-knock out mouse embryos are killed by maternal complement (Xu et al. 2000). The rat equivalent to mouse Crry has six or seven CCP modules (Takizawa et al. 1994) (Figure 2.3). Rodent versions of both MCP and DAF are expressed, but MCP expression is

restricted to testis (Miwa et al. 1998). There are two mouse DAF genes, I and II, producing predominantly transmembrane and GPI-anchored versions, respectively (Miwa et al. 1998). GPI-linked, transmembrane, and secreted versions of rat, guinea pig and pig DAF are probably splice variants (Nonaka et al. 1995; Miwa et al. 2000; Perez de la Lastra et al. 2000). Pig DAF is unusual in having only three CCPs (Perez de la Lastra et al. 2000). Chickens produce a transmembrane complement regulatory protein Cremp (also called Crem) that has five CCP modules and an additional CCP-like module (Inoue et al. 2001). The first two modules resemble modules 2 and 3 of DAF, and the third and fourth modules are very similar to the equivalent modules of MCP. The locus of the first non-mammalian RCA gene cluster to be reported was recently identified in chickens (Oshiumi et al. 2005), revealing two additional proteins; Cres, a secretory protein consisting of 10 CCP modules, modules 2-to-4 displaying homology to the equivalent modules of the α-chain of C4BP in humans, while the remaining modules do not show marked similarity to any other RCA module; and Creg, a GPI-anchored membrane protein, consisting of seven CCP modules, with high homology to modules of an LHR of CR1 in humans (CCPs 1-4, 8-11, 15-18, and 22-25). A protein, SBP1, from the bony fish barred sand bass has 17 CCP modules, which resemble factor H in humans and was demonstrated to bind both C3b and C4b and to act as a cofactor for their enzymatic cleavage by factor I (Kemper et al. 1998). A family of large transmembrane receptors and adhesive proteins, CSMD (CUB and sushi multiple domain) 1, 2 and 3 each contain more than 25 CCP modules (Lau and Scholnick 2003; Shimizu et al. 2003). CSMD1 from rat was recently shown to possess complement regulatory activity, thus raising the possibility that a similar feat is performed by its human orthologue (Kraus et al. 2006).

2.3.1.4 *Viral versions of RCAs*

Several viruses in the Herpesviridae and Poxviridae families have genes for CCP module-containing proteins, presumably acquired through a process of horizontal gene transfer. Pox viruses generally produce a soluble four-CCP module complement inhibiting protein (Figure 2.3), although the monkeypox version has only three modules (Uvarova and Shchelkunov 2001). The best-studied example is VCP, from Vaccinia virus (Kotwal and Moss 1988) that has four CCPs. Vaccinia virus has an additional gene encoding a distinct four CCP-module containing protein – VB05 – that has a transmembrane domain and reportedly plays a role in plaque-size and host-range (Takahashi-Nishimaki et al. 1991). Another virus, variola encodes a smallpox inhibitor of complement enzymes (SPICE) protein, which also functions as a complement regulator (Dunlop et al. 2003). Within the herpes family, some members of the rhadinovirus genus have been found to produce a four CCP module-protein with a transmembrane domain (Albrecht and Fleckenstein 1992). The version expressed by Kaposi's sarcoma-associated herpesvirus ("KCP", also known as "Kaposica") (Figure 2.3) has been shown to inhibit complement effectively and contribute significantly to virulence (Mullick et al. 2003; Spiller et al. 2003).

2.3.1.5 *Other proteins that contain CCP modules*

Several other human proteins that interact with C3b and/or C4b contain two or three CCP modules – these are factor B and C2, C1r and C1s, MASP 1/3 and MASP 2 (Figure 2.3). The late complement components, C6 and C7, contain pairs of tandem CCP modules that may interact with the C5b* fragment generated by C5 convertase (DiScipio et al. 1999). The '*' refers to the unstable, transitory nature of this

fragment that is able to interact with C6 to form C5b.C6 in an initial irreversible step along the pathway to formation of the membrane attack complex (DiScipio et al. 1999) (discussed further in Chapter 5).

There are also a large and very diverse range of non-complement mammalian proteins that contain CCP modules. Illustrative examples of note are: clotting FXIIIb; the adhesive extracellular matrix proteins - aggrecan, neurocan and brevican that each contain a single CCP module; the selectin family – whose stalks are composed of trains of CCPs; a subunit of the heterodimeric metabotropic G-protein linked GABA receptor - $GABA_B$ receptor type 1a (Blein et al. 2000) (type 1b lacks the two CCP modules (Kaupmann et al. 1997)); the α-chain of the IL-2 and IL-15 receptors - the single CCP module of the IL-15 receptor, together with a C-terminal Pro/Thr region, is necessary and sufficient for IL-15 binding (Wei et al. 2001); β-2-glycoprotein I also known as apolipoprotein H (β2GPI, ApoH) for which two crystal structures have been solved (Bouma et al. 1999; Schwarzenbacher et al. 1999); polydom, a 387 kDa placentally expressed protein with 34 CCP modules (Gilges et al. 2000), also referred to as SEL-OB (selectin-like osteoblast derived) and SVEP1 (sushi, von Willebrand factor type A (vWA), EGF, and pentraxin 1) in humans (Shur et al. 2006). Perhaps the most primitive eukaryote known to possess a CCP module to date is *Plasmodium falciparum* (O'Keeffe et al. 2005).

This widespread use of CCP modules underlines the versatility of a structural scaffold that can be adapted by evolution to suit many purposes, both "architectural" – endowing on specific proteins an appropriate reach and level of flexibility or rigidity - and "functional" *i.e.* providing specific surfaces that are directly involved in binding (Kirkitadze and Barlow 2001). In recent years there has been increasing evidence suggesting that complement components, particularly complement component C3

may also participate in non-inflammatory and developmental processes, such as tissue regeneration, neuron development, wound healing and reproduction (Mastellos and Lambris 2002; Mastellos et al. 2003; Mastellos et al. 2005), indicating that it has a broad role in cell differentiation. It is plausible that several of the non-complement CCP modules could interact with C3 in its developmental role.

2.4 Sequences of CCP modules

2.4.1 The consensus sequence

Figure 2.4: Multiple sequence alignment of CCP modules in RCA proteins. The 84 individual CCP modules of the six proteins in the RCA family- CR1 (30 CCPs), CR2 (15 CCPs), C4BPα (8

CCPs) and β (3 CCPs), DAF (4 CCPs), MCP (4 CCPs) and factor H (20 CCPs) aligned using ClustalX (Thompson et al. 1997). Consensus residues generated by WebLogo version 2.8.2 (http://weblogo.berkeley.edu/) (Crooks et al. 2004) are indicated at the bottom of the alignment (the greater the height of the residue letter, the greater the conservation for that residue at that position). This alignment differs slightly from the one published in (Soares and Barlow 2005), in light of new structural information.

Historically, the existence of CCP modules was inferred from inspection of the emerging primary sequence information for RCA proteins during the late 1980s (Campbell et al. 1988; Reid and Day 1989). An imperfectly repeating motif of approximately 60 amino acid residues was noticed and termed the "short consensus repeat". This repeat is characterised (Figure 2.4) by the presence of a cysteine residue at either end (C^{I} and C^{IV}) with two additional cysteines (Cys^{II} and Cys^{III}) distributed between them – it was later shown that these are disulfide linked in the pattern I → III and II → IV. A tryptophan occurs - with two exceptions only among the RCA proteins (the 10^{th} and 20^{th} CCPs of fH, which have a Leu instead) - in the sequence between cysteines III and IV. Proline residues often occur at the third and/or fourth position after the first Cys (*i.e.* $C^{I} + 3/4$) and the second position before the fourth Cys ($C^{IV} - 2$). Additional Pro residues often occur adjacent, or next-door-but-one, to these consensus prolines. A Gly residue is well conserved at ($C^{II} + 3$) and three additional conserved glycines commonly occur at ($C^{I} + 8/10$), ($C^{II} - 7$) and ($C^{II} + 6/8$). Within the ten residues prior to C^{II} there are four positions where hydrophobic residues are present in the sequences of the majority of CCP modules, and additional hydrophobic residues appear beyond C^{II}. Thus the motif "hXhGXXhXhXC^{II}XXG↑hXhXG" (where ↑ is the site of an insertion in the larger CCP modules) occurs commonly, but not in every CCP module. Insertions and

deletions occur most frequently after each of the four consensus glycines, apart from the (C^{II} – 7) Gly; and before the consensus Trp.

CR1	FH	CR2	C4BPα	C4BPβ	MCP	DAF
(57) 1-2 (4)	(60) 1-2 (4)	(60) 1-2 (8)	(59) 1-2 (4)	(54) 1-2 (4)	(59) 1-2 (4)	(59) 1-2 (3)
(58) 2-3 (4)	(57) 2-3 (4)	(56) 2-3 (7)	(58) 2-3 (4)	(54) 2-3 (4)	(59) 2-3 (4)	(61) 2-3 (4)
(67) 3-4 (5)	(60) 3-4 (4)	(57) 3-4 (4)	(60) 3-4 (4)	(53) 3	(62) 3-4 (4)	(58) 3-4 (4)
(56) 4-5 (3)	(53) 4-5 (4)	(57) 4-5 (4)	(56) 4-5 (4)		(56) 4	(59) 4
(57) 5-6 (4)	(54) 5-6 (4)	(67) 5-6 (8)	(62) 5-6 (4)			
(59) 6-7 (4)	(61) 6-7 (3)	(56) 6-7 (3)	(58) 6-7 (3)			
(67) 7-8 (5)	(54) 7-8 (5)	(57) 7-8 (4)	(55) 7-8 (3)			
(57) 8-9 (4)	(58) 8-9 (3)	(52) 8-9 (4)	(55) 8			
(58) 9-10 (4)	(56) 9-10 (4)	(67) 9-10 (8)				
(67) 10-11 (5)	(55) 10-11 (6)	(56) 10-11 (4)				
(56) 11-12 (3)	(55) 11-12 (6)	(53) 11-12 (4)				
(57) 12-13 (4)	(54) 12-13 (8)	(61) 12-13 (8)				
(59) 13-14 (4)	(51) 13-14 (7)	(56) 13-14 (7)				
(67) 14-15 (5)	(54) 14-15 (5)	(57) 14-15 (4)				
(57) 15-16 (4)	(57) 15-16 (4)	(57) 15				
(58) 16-17 (4)	(54) 16-17 (4)					
(67) 17-18 (5)	(55) 17-18 (4)					
(56) 18-19 (3)	(55) 18-19 (6)					
(57) 19-20 (4)	(55) 19-20 (3)					
(59) 20-21 (4)	(62) 20					
(67) 21-22 (8)						
(57) 22-23 (4)						
(58) 23-24 (4)						
(67) 24-25 (5)						
(56) 25-26 (3)						
(57) 26-27 (4)						
(59) 27-28 (4)						
(67) 28-29 (7)						
(57) 29-30 (4)						
(57) 30						
(57)Mean(4.3)	**(56)Mean(4.6)**	**(57.9)Mean(5.5)**	**(57.9)Mean(3.7)**	**(53.7)Mean(4)**	**(59)Mean(4)**	**(59.3)Mean(3.7)**

Table 2.1: Lengths, in residues, of CCP modules (and inter-CCP module linkers) in RCA proteins. The residue-length of a CCP module is defined as the number of residues between the N-terminal cysteine and the C-terminal cysteine, both inclusive (shown in brackets on the left). Residue length of inter-CCP module linkers is defined as the number of residues between the C-terminal cysteine of the preceding CCP module and the N-terminal cysteine residue of the following CCP module (shown in brackets on the right). Mean values are shown below, corresponding to each protein.

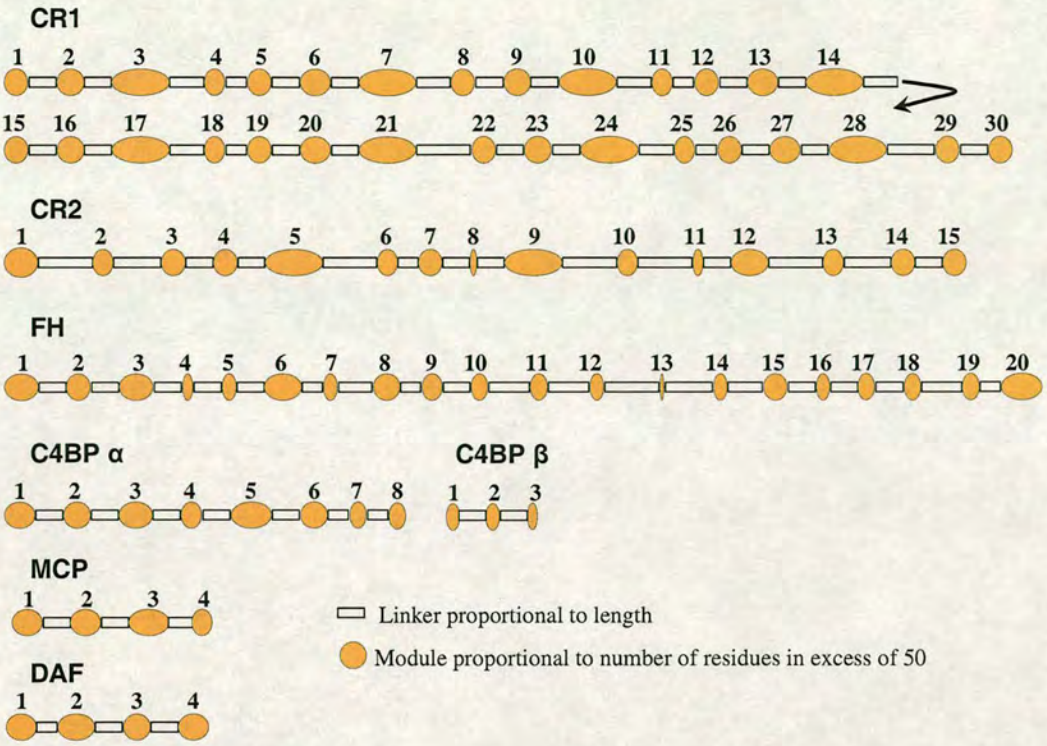


Figure 2.5: Schematic to illustrate module and linker lengths. This figure is drawn on a distorted scale to emphasise the variation in length of CCP modules (51-67 residues) within RCA proteins and of the linkers (3-8 residues) between them. For convenience the first and last residues of a module are defined as C^{I} and C^{IV} of the consensus sequence (see Figure 2.4), and linkers are defined as the sequences lying between C^{IV} of the N-terminal module and C^{I} of the C-terminal module. Modules are numbered. For abbreviations see text. The raw data for this figure were compiled from the Swissprot database (Boeckmann et al. 2003).

Three-dimensional (3D) structure determination of CCP modules indicated that the residue before the first Cys, and the two residues beyond the last Cys commonly contribute to the three-dimensional structure of the module. Strictly speaking therefore, these residues could be considered to belong to the module rather than the connecting linking sequences. For the purposes of the alignment in Figure 2.4, sequences starting at the residue before C^{I} and finishing three residues after C^{IV} (inclusive) were used. It is nonetheless convenient to regard the boundaries of each CCP module as C^{I} and C^{IV}, and the "linker" residues as lying between the C^{IV} of the preceding module and the C^{I} of the following module. According to that definition,

within the RCA proteins, the length of CCP module sequences in RCA proteins varies from 51 (factor H~13) to 67 (eight modules in CR1, two in CR2) (see Table 2.1; Figure 2.5). And the length of linker ranges from three to eight amino acid residues (Table 2.1).

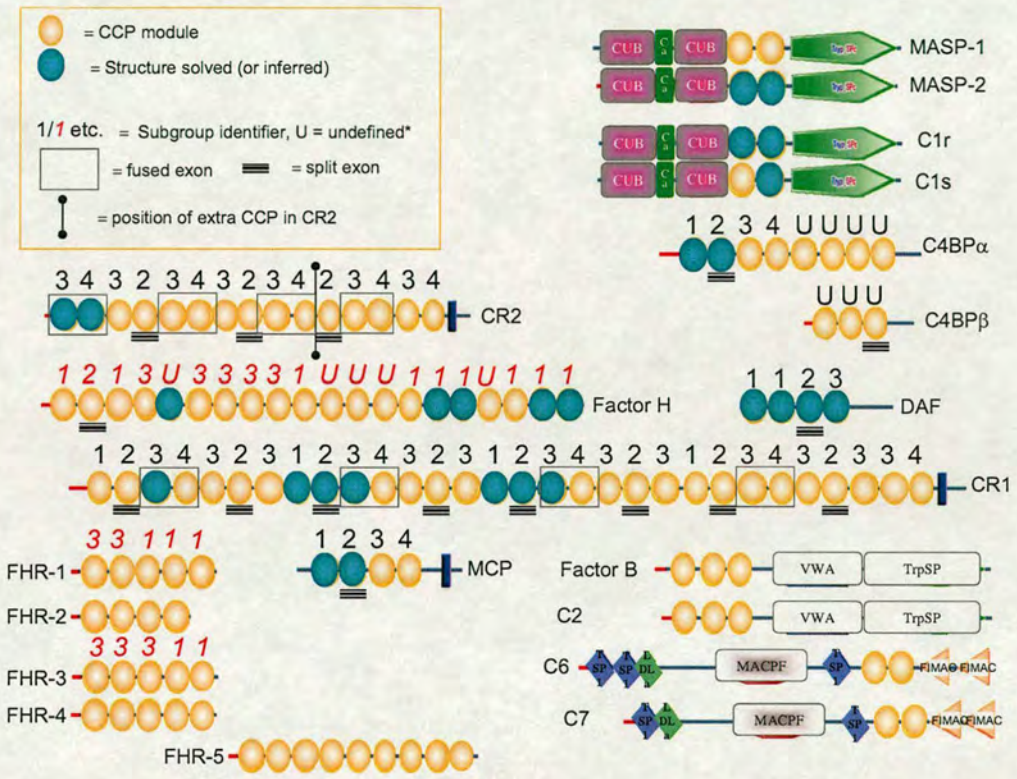


Figure 2.6: Exon boundaries and assignment of the CCP modules of complement proteins into groups. The symbols used for module-types are the same as in figure 2.3. The modules for which experimentally solved (or inferred) 3D structures are available are shaded in blue. Most CCP modules are encoded by single exons; split and fused exons are highlighted in this figure – see key for symbols. The number above each module indicates assignment to sub-groups of CCPs according to a phylogenetic study by Krushkal et al. (2000) discussed in the text. The study divided the RCA proteins into two distinct groups (fH, the fH-related proteins and Factor XIIIb in one group – shown by the use of italics and numbering in red – and the remainder in the second group). Those modules that could not be classified are indicated by a "U".

Many CCP module sequences are encoded by discrete exons and therefore they are class 1 domains. The exon boundaries appear to lie at random positions

within the linkers but nearly all exons in RCA genes have phase 1 introns (*i.e.* they interrupt the coding sequence after the first base-pair of a codon) at both ends, a typical feature of extracellular metazoan proteins and a signature of domain shuffling (Kaessmann et al. 2002). In CR2, however, the fourth, eighth and eleventh CCPs are class 2 domains since they are encoded by a split exon, while the following CCP pairs of CR2 are encoded by a fused exon: 1,2; 5,6; 9,10; and 12,13 (Fujisaku et al. 1989) (see Figure 2.6). The second CCPs of other RCAs - CR1, factor H, MCP and C4BPα (and the third modules of DAF and C4BPβ) are also class 2 domains with a splice site after the second position of the codon for the conserved Gly at position (Cys^{II} + 3) (Hillarp et al. 1993). In the case of CR1, a similarly split exon encodes the second CCPs of LHRs B, C and D, and the sixth module of each LHR (Vik and Wong 1993). The third and fourth modules of each LHR in CR1 are encoded by fused exons (Vik and Wong 1993).

2.4.2 "Evolution" of CCP modules

The RCA gene cluster is thought to have evolved through a series of gene duplication events (Farries and Atkinson 1991), consistent with the symmetrical phases of most introns. Krushkal *et al.* (Krushkal et al. 2000) constructed a phylogenetic tree for 132 individual module sequences of the RCA gene cluster including Factor XIII b subunit (and SBP1 from sand bass). Based on this they inferred a summary tree of relationships for the parent proteins. The tree identified two groups of proteins – factor H and the factor H-related proteins along with Factor XIIIb (and SBP1) formed one group, while MCP, DAF, C4BPα, CR2 and CR1 formed the other (and C4BPβ falls into neither). This division corresponds to two distinct sub-clusters on chromosome 1 separated by 7-22 centiMorgans. One possibility is that SBP1 is the

evolutionary precursor of both factor H and C4BPα, and that the two groups diverged in evolutionary history after the separation of the fish lineage. Within these groups, these authors identified "sub-types" of CCPs (illustrated in Figure 2.6). These maybe compared with the "clusters" described in Chapter 3, performed for a greater number of CCP module sequences.

2.4.3 Comparison of sequences amongst CCP modules

Levels of sequence identity amongst CCP modules in human RCA proteins range from virtually undetectable (except with reference to the four defining Cys residues and the highly conserved Trp) to 100%. Some of the sequences with low similarity in sequence are shown in Figure 2.7a. Even within a protein, CCP sequences may be very dissimilar *e.g.* C4BPα~1 versus C4BPα~6, or fH~13 versus fH~20. In fact, the multiple sequence alignment (Figure 2.4) reveals that in addition to the Trp → Leu substitution in fH~10 and fH~20, there are several further oddities among the CCP-module sequences of RCA proteins – *e.g.* the presence of an additional pair of Cys residues in C4BPα~6; the absence of any consensus Pro or Gly residues after C^I in fH~13, fH~20, CR2~10 and CR2~11, and the presence of five Pro residues within positions $C^I + 8$ in fH~14.

(A)

```
C4BPB_01     1 HCPE-LPPVDNSI---FVAKEV---EGQI-LG---TYVCIKG------YHLVG---KKTLFCNASK----EWD--NTTTECRLG
CR2_08       1 ACEATGRQLL--------TKPQ---HQFV--RPDVNSSCGEG------YKLSG---SVYQECQGTIP----WF--MEIRLCKEI
CR2_11       1 TCQH-VRQSL-------QELPA---GSRV--EL-VNTSCQDG------YQLTG---HAYQMCQDAENG--IWF--KKIPLCKVI
C4BPA_06     1 CCPE--PKLNNGEITQHRKSRPANHCVYF-YGDEISFSCHET----------S---RFSAICQGDG----TWS--PRTPSCGDI
FH_13        1 KCKS-SNLIILEE---HLKNKK----EFD-HNSNIRYRCRGK---------EG---WIHTVCIN-G----RWD---PEVNCSMA
FH_20        1 PCVI-SREIMENYN--IALRWTAKQKLYSRTGESVEFVCKRG------YRLSSRSHTLRTTCWD-G----KLE----YPTCAKR
```

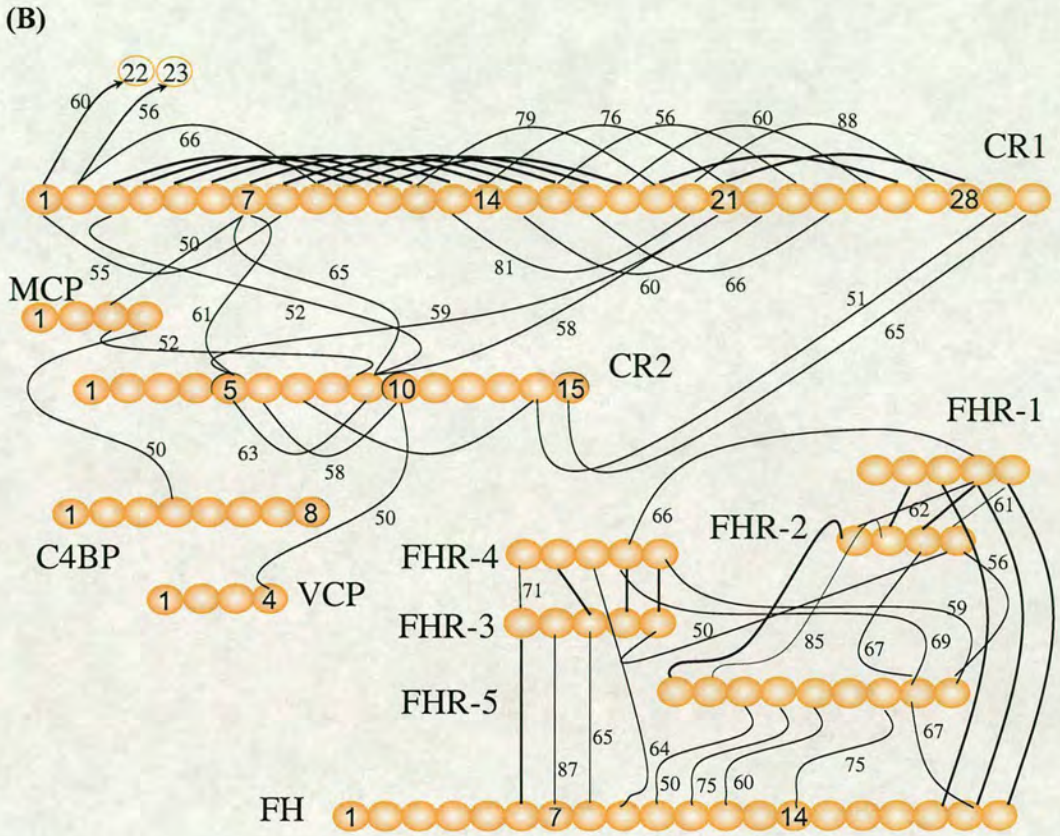


Figure 2.7: Differences and similarities between the primary sequences of CCP modules. (A) Six relatively dissimilar sequences were selected from the sequence alignment and juxtaposed. Levels of pair-wise identity are typically < 20%. Figure produced using BoxShade version 3.21 (http://www.ch.embnet.org/software/BOX_form.html). **(B)** Thick lines connect modules that share more than 90% pair-wise sequence identity; thin lines indicate 50-89% identity, with a label indicating the percentage identity (Altschul et al. 1990; Altschul et al. 1997). For clarity, not all (>90% identical) lines are drawn since the levels of identity can be deduced in some cases (*e.g.* between CR1~3 and CR1~17).

CR1 on the other hand is exceptional in the extent of its internal sequence similarities - several of the CCP sequences within CR1 are exact, or nearly exact, duplicates or triplicates (see Figure 2.7b): modules 3-to-9 are almost identical to CR1 modules 10-to-16; CCP modules 3 (and 10) and 4 (and 11) are nearly identical with modules 17 and 18, respectively; and modules 19-to-21 are ≥ 88% identical with modules 26-to-28. As mentioned previously, this occurrence of similar modules at

intervals of seven led to the division of the N-terminal 28 modules of CR1 into four LHRs, A, B, C and D. Modules in CR1 with divergent sequence lie at the N- and C-termini: modules 29 and 30 each have less than 50% identity with any other CR1 module, while modules 1 and 2 are only 60-to-66% identical at best to other CCP modules. For completeness, it should be noted that modules 22-to-25 of CR1 are also rather divergent. No other complement control protein shows such a high level of internal sequence similarity, although CR2 modules 5 and 6 are 63% and 58% identical to CR2~9 and CR2~10, respectively (see Figure 2.7b).

There is less similarity between modules located in different RCA proteins. The C-terminal modules of CR1 – numbers 29 and 30 - are 51% and 65% identical to the C-terminal two modules of CR2 (14 and 15) respectively. In fact, outside of the fH family, the 65% identity between C-terminal modules of CR1 and CR2 is matched only by CR2~9 versus CR1~7 (or 14). The set of other inter-protein identity-levels ≥ 50% are summarised in Figure 2.7b.

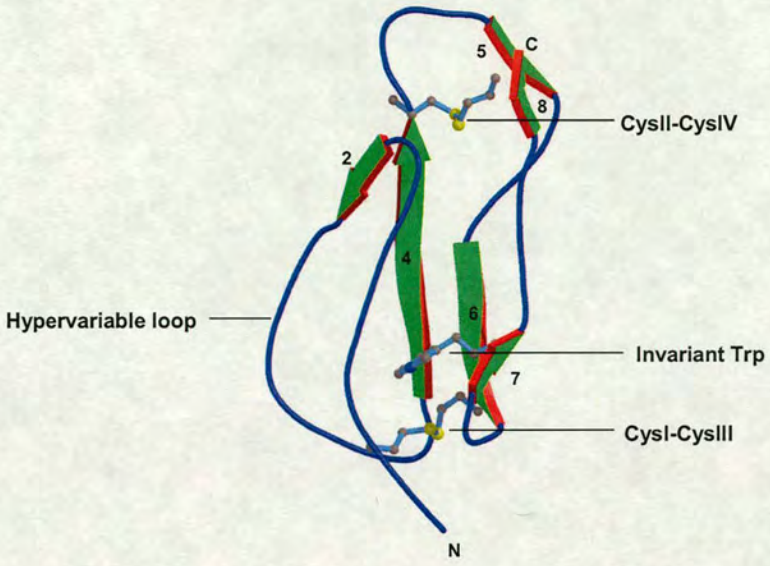
2.5 The 3D structure of CCP module

2.5.1 Early structural work on factor H

The first structural data on CCP modules, derived from circular dichroism (CD) studies of fH, demonstrated an absence of α-helices (Sim and DiScipio 1982). A subsequent study combining Fourier transform-infrared spectroscopy (FT-IR) with secondary structure averaging methods predicted a β-structure for each of the 20 CCP modules of factor H (Perkins et al. 1988).

In order to obtain more detailed structural information (Barlow et al. 1991), a portion of factor H corresponding to what was predicted to be a single CCP module (residues 909-967 of the mature sequence, the 16th CCP within the sequence, known as fH~16) was expressed in *Saccharomyces cerevisiae* as a recombinant entity in isolation from the remainder of the molecule. This sequence was not selected on the basis of its contributions to any known function of the parent protein, but because its sequence was regarded as typical of CCP module sequences in factor H. A pure (no salt or buffer) 1 mM solution of this material (fH~16) at pH 4.0 and 37 °C was analysed by two-dimensional (2D) ^{1}H nuclear magnetic resonance spectroscopy (NMR). The spectra were consistent with those expected of a compactly folded globular protein, and demonstrated that such a sequence is able to fold independently. The NMR data confirmed the presence of β-strands but, in comparison with the outcome of the FT-IR based-study, suggested different numbers and locations of strands.

(A)



(B)

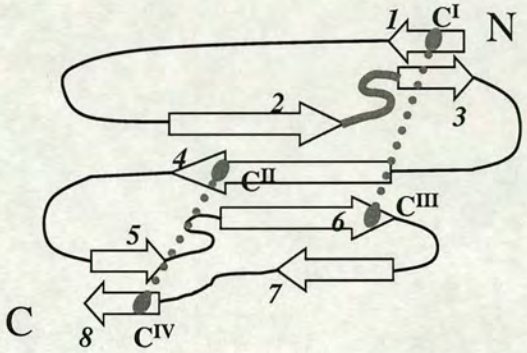


Figure 2.8: Secondary and tertiary structure of a CCP module. **(A)** Molscript (Kraulis 1991) representation of the first CCP module structure [the 16th module of factor H, PDB ID: 1HCC (Norman et al. 1991)] to be revealed by NMR spectroscopy. The consensus "scaffold" of conserved residues, involving the cysteines (forming two disulfide bridges) and a buried tryptophan are drawn as ball-and-stick. β-strands are shown as arrows. **(B)** β-strands drawn as arrows, disulfides indicated by the thick dotted lines, and the hypervariable loop drawn as a thick line between strands 2 and 3. The exact position and length of β-strands, as formally defined according to the Kabsch and Sander criteria (Kabsch and Sander 1983), varies between modules. All structures solved to date, however, contain some or all of the eight strands at the approximate positions relative to the four consensus cysteines indicated here. In this Chapter we use the strand-numbering shown in this diagram even in modules where some of the strands are not formed. Thus the six strands of the 16th CCP of fH are numbered 2, 4, 5, 6, 7 and 8 (see Figure 2.8a and text). In the literature, pairs of strands such as 5 and 6 are sometimes considered to be segments of a common strand and thus often referred to as, for example, D' and D.

A more detailed analysis of the NMR data yielded the first 3D structure of a CCP module (Figure 2.8a) (Norman et al. 1991). The fold was described originally as approximating to a β-sandwich or barrel, but it is smaller and less regular than classical examples of this fold such as found amongst the immunoglobulin (Bork et al. 1994) or fibronectin type III families (Potts and Campbell 1994; 1996). It does not in fact resemble closely any other known protein structure. One face consists of a twisted sheet of three antiparallel strands (numbers 2, 4 and 6 in Figure 2.8a). The other face was originally described as a two-stranded sheet. The longer of the two strands in this other face has two segments separated by a β-bulge – the more N-terminal segment extends the three-stranded sheet and thus participates in both faces, while the more C-terminal part (also corresponding to the C-terminus of the module) is exclusively part of the two-stranded face. In the light of further examples of CCP module structures, these two segments can be regarded as two independent strands 7 and 8 (Figure 2.8) – and it is more convenient to consider fH~16 as consisting of a twisted four-stranded sheet (containing β-strands now numbered 2, 4, 6 and 7 – Figure 2.8a) forming a half-barrel while the separate two-stranded sheet (β-strands 5 and 8) closes off one end of the barrel. The open side of the half-barrel is covered by the N-terminal region of the module. Both sheets contribute hydrophobic side-chains to a compact core and there is a disulfide bond between C^{I} and C^{III}, and another disulfide between C^{II} and C^{IV}, as expected. These two disulfide bridges staple the N-terminus and the C-terminus to opposite ends of the four-stranded sheet, and are far apart from one another, forming the upper and lower boundaries of the hydrophobic core. The module has, overall, an elongated shape with β-strands 2, 4, 6 and 7 aligned very approximately with the long axis while the N- and C-termini are found at opposing ends of the long axis. This is consistent with the extended,

head-to-tail arrangement of CCP modules within RCA proteins suggested by other experimental methods.

Alignments of CCP module sequences analysed in the light of the structure showed that many structurally important residues, such as the tryptophan that lies between C^{III} and C^{IV} in the sequence and is buried in the structure, are conserved or conservatively replaced (Figure 2.4); therefore it appeared likely that other CCP modules would share a similar structure. To illustrate this point, the fH~16 structure was used as a template for modelling of putative CCP modules from the selectin family, which have an extra pair of cysteines. The homology modelling (Norman et al. 1991) was consistent with the additional cysteines being arranged in such a way as to allow formation of the expected third disulfide.

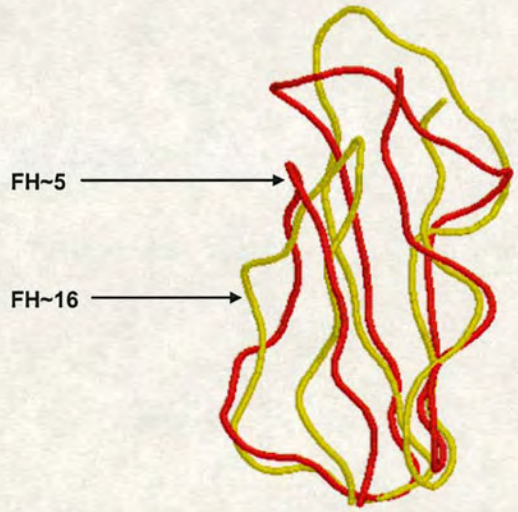


Figure 2.9: Overlay of the 3D solution structures of the 5th and 16th CCP modules of fH. A tube representation of the backbones of the first two CCP module structures to be solved, which illustrates that although sequence identity between these two is low, they share a similar 3D structure.

The fifth module of factor H (fH~5) has only 33% sequence identity with fH~16. It was therefore selected for 3D structure determination by NMR as a second, divergent,

example of a CCP module (Barlow et al. 1992). The two modules may be seen (Figure 2.9) to share the elongated scaffold-like β-structure discussed above. There are, however, eight short β-strands (numbered 1-8 – see Figure 2.8b) in fH~5. Strands 1 and 3 are apparently not present in module 16 (in other words the equivalent residues of fH~16 while strand-like, did not adopt Φ and Ψ dihedral angles within the ranges expected of a β-strand). These two additional strands form a small β-sheet that closes off the other (N-terminal) end of the half-barrel with respect to that occupied by the 5/8 β-sheet. To summarise the topology of the CCP module (Figure 2.8b): strand 1, when present, includes the N-terminus and C^{I}; strand 2 follows the Gly at position (C^{I} + 8/10); strands 3 (when present), 4 and 5 occur (underlined) within the "hXhGXXhXhXC^{II}XXG↑hXhXG" motif; strand 6 precedes (and may include) C^{III}; strand 7 includes the (almost) invariant Trp; and strand 8 includes C^{IV} and the C-terminus of the module.

The 3D structures of fH~5 and fH~16 may be overlaid on the α and β carbons of the equivalent consensus Cys and Trp residues with a root mean square deviation (rmsd) of 1.45 Å; but loops, bulges and turns diverge. The rmsd over the equivalent Cα atoms of 49 residues after superposition using the program Combinatorial Extension (CE) (Shindyalov and Bourne 1998) is 2.6 Å. One region in particular - corresponding to residues 17-22 of fH~16 - was not well defined by the experimental data in either module, is highly variable in sequence amongst CCP modules generally, and corresponds to a prominent loop (Figure 2.8) that projects laterally from the elongated body of the module - this was named the "hypervariable" loop or region.

2.5.2 Further examples of CCP module structures

CCP-containing protein	Module(s) solved	PDB code(s)	References
Complement Receptor 1 (CR1)	15, 16, 17	1GKN*$_{15,16}$, 1GKG*$_{16,17}$, 1PPQ*$_{16}$	(Smith et al. 2002; O'Leary et al. 2004)
Complement Receptor 2 (CR2)	1, 2	1GHQ^$_{1,2}$, 1LY2^$_{1,2}$	(Szakonyi et al. 2001; Prota et al. 2002)
Decay Accelerating Factor (DAF)	1, 2, 3, 4	1H03^$_{3,4}$, 1H04^$_{3,4}$, 1H2P^$_{3,4}$, 1H2Q^$_{3,4}$, 1UOT^$_{3,4}$, 1OJV^$_{1-4}$, 1OJW^$_{1-4}$, 1OJY^$_{1-4}$, 1OK1^$_{1-4}$, 1OK2^$_{1-4}$, 1OK3^$_{1-4}$, 1OK9^$_{1-4}$, 1NWV*$_{2,3}$	(Uhrinova et al. 2003; Williams et al. 2003; Lukacik et al. 2004)
Membrane Cofactor Protein (MCP)	1, 2	1CKL^$_{1,2}$	(Casasnovas et al. 1999)
factor H (fH)	5*, 15, 16, 19, 20	1HFH*$_{15,16}$, 1HFI*$_{15}$, 1HCC*$_{16}$, 2BZM*$_{19,20}$	(Norman et al. 1991; Barlow et al. 1992; Barlow et al. 1993; Herbert et al. 2006c)
C4b-Binding Protein α (C4BPα)	1, 2	2A55*$_{1,2}$	(Jenkins et al. 2006)
Vaccinia Complement Protein (VCP)	1, 2, 3, 4	1G40^$_{1-4}$, 1G44^$_{1-4}$, 1VVC*$_{3,4}$, 1VVD*$_{3,4}$, 1VVE*$_{3,4}$, 1E5G*$_{2,3}$, 1RID^$_{1-4}$, 1Y8E^$_{1-4}$	(Wiles et al. 1997; Henderson et al. 2001; Murthy et al. 2001; Ganesh et al. 2004; Ganesh et al. 2005)
C1r	1, 2	1GPZ^$_{1,2}$, 1MD7^$_{2}$, 1MD8^$_{2}$	(Budayova-Spano et al. 2002a; Budayova-Spano et al. 2002b)
C1s	2	1ELV^$_{2}$	(Gaboriaud et al. 2000)
MASP-2	1, 2	1Q3X^$_{2}$, 1ZJK^$_{1,2}$	(Harmat et al. 2004; Gal et al. 2005)
β-2-glycoprotein I / Apolipoprotein H (β2-GPI)	1, 2, 3, 4	1C1Z^$_{1-4}$, 1QUB^$_{1-4}$	(Bouma et al. 1999; Schwarzenbacher et al. 1999)
$GABA_B$-R1α (rat)	2	1SS2*$_{2\ (cis\text{-}form)}$, 1SRZ*$_{2\ (trans\text{-}form)}$	(Blein et al. 2004)
IL-2Rα	1, 2	1Z92^$_{1,2}$, 2B5I^$_{1,2}$, 2ERJ^$_{1,2}$	(Rickert et al. 2005; Wang et al. 2005; Stauber et al. 2006)
IL-15Rα	1	2ERS*$_{1}$	(Lorenzen et al. 2006)

Table 2.2: Experimentally solved CCP-modules. Table of all experimentally determined CCP modules from human and non-human proteins at 3rd April 2006. Only those structures determined by X-ray diffraction and NMR are listed. Symbols used in table: * CCP-modules solved by NMR, ^ CCP-

modules solved by X-ray diffraction. It should be noted that although, factor H module 5 co-ordinates have not been deposited into the PDB, they are available for download from: http://www.bru.ed.ac.uk/~dinesh/ccp-db.html. Module number or ranges are indicated as subscript alongside their PDB codes. Abbreviations used in table: $GABA_B$-R1a = Gamma-aminobutyric acid type B receptor, subunit 1 α ; IL-2Rα / IL-15Rα = Interleukin-2/-15 receptor α chain; MASP-2 = mannan-binding lectin-associated serine protease-2.

In the decade and a half since the initial structural work that focussed on these two isolated CCP modules from factor H, a wealth of structural data on CCP module 3D structures has been amassed (Table 2.2). To date 35 such structures have been solved – normally in the context of longer fragments of the parent protein. Of these, four derive from the complete structure of β2GPI (Bouma et al. 1999; Schwarzenbacher et al. 1999) - which also contains a fifth very divergent version of a CCP module that is involved in phospholipid binding (Steinkasserer et al. 1992), two from IL2Rα (Rickert et al. 2005; Wang et al. 2005; Stauber et al. 2006) and one each from IL15Rα (Lorenzen et al. 2006) and $GABA_B$R1a (Blein et al. 2004). The remainder are examples found in complement proteins – two from MASP 2 (Harmat et al. 2004; Gal et al. 2005), one module from C1s (Gaboriaud et al. 2000) and two from C1r (Budayova-Spano et al. 2002a; Budayova-Spano et al. 2002b), two from the N-terminus of CR2 (Szakonyi et al. 2001; Prota et al. 2002), and the remainder from the complement regulators: fH (*i.e.* fH~15 (Barlow et al. 1993) and fH~19,20 (Herbert et al. 2006c), in addition to fH~5 and fH~16), C4BP (two) (Jenkins et al. 2006), MCP (two) (Casasnovas et al. 1999), DAF (all four modules) (Uhrinova et al. 2003; Williams et al. 2003; Lukacik et al. 2004), CR1 (three) (Smith et al. 2002; O'Leary et al. 2004) and the viral mimic, VCP (all four modules representing the only complete protein structure) (Wiles et al. 1997; Henderson et al. 2001; Murthy et al. 2001; Ganesh et al. 2004; Ganesh et al. 2005).

Of the 35 CCP modules for which structures have been solved, 24 were determined by X-ray crystallography, 16 by NMR and five by both techniques. Several CCP module structures have been solved in different contexts – *e.g.* alone, or with neighbouring modules attached.

	VCP2-X (1G40)	**VCP3-X (1G40)**	**VCP4-X (1G40)**	**DAF2-X (1OK3)**	**DAF3-X (1H03)**
VCP2-N (1E5G)	3.31 (61)				
VCP3-N (1E5G)		1.70 (57)			
VCP3-N (1VVC)		1.14 (58)			
VCP4-N (1VVC)			1.40 (58)		
DAF2-N (1NWV)				2.00 (65)	
DAF3-N (1NWV)					1.69 (62)

Table 2.3: Comparison of individual CCP module structures solved by both X-ray diffraction and NMR. CE comparison of structures solved by X-ray diffraction (-X) against those solved by NMR (-N), PDB codes provided. All values are in Angstroms (Å); Alignment length is indicated in brackets below.

Where both NMR and X-ray diffraction were used to solve the structure (Table 2.3), agreement varied from good (*e.g.* 1.14 Å for VCP~3) (Wiles et al. 1997; Murthy et al. 2001) to poor (3.31 Å for VCP~2) (Henderson et al. 2001; Murthy et al. 2001). It is rare for solution and crystallographic structures of proteins to differ significantly (Billeter 1992; Garbuzynskiy et al. 2005). In the case of VCP~2 the presence or absence of neighbouring modules could possibly explain some of the discrepancies – *e.g.* the solution structure of VCP~2 was solved in the context of the 2,3 pair, while

the crystal structure derives from a structure of intact protein. On the other hand, the solution structure of CCP module 16 of CR1 was essentially the same when solved as an isolated module (O'Leary et al. 2004) or in the context of the CR1~15,16 and CR1~16,17 pairs (Smith et al. 2002).

	BGPI ~1	BGPI ~2	BGPI ~3	BGPI ~4	C1r ~1	C1r ~2	C1s ~2	C4BP ~1	C4BP ~2	CR1 ~15	CR1 ~16	CR1 ~17	CR2 ~1	CR2 ~2	DAF ~1	DAF ~2	DAF ~3	DAF ~4	GABc ~2	GABt ~2	FH ~5	FH ~15	FH ~16	FH ~19	FH ~20	IL15R ~1	MSP2 ~1	MSP2 ~2	MCP ~1	MCP ~2	VCP ~1	VCP ~2	VCP ~3
BGPI ~1																																	
BGPI ~2	2.59 (57)																																
BGPI ~3	2.11 (59)	1.75 (57)																															
BGPI ~4	2.51 (59)	1.40 (57)	1.48 (60)																														
C1r ~1	1.74 (61)	1.85 (58)	1.70 (59)	1.80 (59)																													
C1r ~2	2.13 (38)	1.77 (36)	3.37 (42)	2.06 (33)	8.45 (52)																												
C1s ~2	2.04 (50)	1.83 (57)	1.59 (58)	1.64 (49)	1.83 (59)	1.61 (37)																											
C4BP ~1	3.10 (60)	3.46 (57)	2.76 (62)	2.99 (60)	3.14 (59)	3.00 (35)	2.71 (51)																										
C4BP ~2	2.36 (59)	1.79 (58)	1.95 (58)	1.93 (58)	1.98 (59)	1.66 (32)	1.87 (58)	3.39 (57)																									
CR1 ~15	2.32 (54)	3.12 (57)	2.20 (54)	2.54 (54)	2.63 (59)	2.70 (39)	2.19 (48)	2.18 (60)	2.58 (51)																								
CR1 ~16	2.81 (57)	1.61 (58)	1.83 (58)	1.87 (58)	1.93 (59)	1.57 (33)	1.54 (57)	3.20 (55)	1.56 (62)	2.88 (50)																							
CR1 ~17	2.18 (58)	3.58 (69)	1.85 (58)	2.27 (59)	2.28 (58)	1.69 (40)	2.17 (58)	2.82 (59)	2.10 (60)	2.61 (55)	2.22 (60)																						
CR2 ~1	2.01 (60)	1.72 (58)	1.49 (59)	1.43 (59)	1.64 (60)	1.84 (38)	3.81 (74)	3.08 (59)	1.61 (62)	2.63 (58)	1.77 (62)	2.02 (62)																					
CR2 ~2	2.23 (59)	1.46 (57)	1.19 (58)	1.68 (58)	1.74 (59)	2.21 (35)	1.90 (59)	2.68 (58)	1.98 (58)	2.52 (58)	1.79 (57)	1.66 (57)	1.57 (59)																				
DAF ~1	2.07 (61)	2.72 (57)	2.41 (60)	2.59 (58)	2.60 (62)	2.13 (39)	2.34 (51)	2.77 (61)	2.30 (59)	2.40 (60)	2.51 (59)	2.12 (59)	2.08 (60)	2.15 (58)																			
DAF ~2	2.31 (57)	2.98 (58)	2.25 (57)	2.65 (56)	2.85 (62)	2.85 (34)	2.29 (51)	2.48 (63)	2.55 (59)	1.99 (55)	2.88 (58)	2.21 (59)	2.52 (60)	2.23 (59)	1.70 (63)																		
DAF ~3	2.49 (58)	1.21 (58)	1.69 (58)	1.51 (58)	1.69 (59)	1.53 (37)	1.40 (57)	3.92 (66)	1.73 (62)	3.02 (58)	1.26 (62)	2.31 (60)	1.75 (62)	1.89 (58)	2.58 (59)	2.86 (59)																	
DAF ~4	1.83 (59)	1.86 (57)	1.46 (58)	2.01 (59)	1.56 (59)	1.93 (38)	1.76 (58)	2.90 (59)	1.55 (61)	2.47 (53)	1.88 (61)	1.83 (61)	1.20 (62)	1.38 (58)	1.74 (59)	2.05 (59)	2.05 (61)																
GABc ~2	2.65 (55)	1.92 (57)	2.47 (57)	2.45 (60)	2.71 (56)	2.96 (32)	1.93 (54)	3.44 (55)	1.86 (54)	3.19 (56)	1.72 (54)	2.47 (57)	1.88 (59)	2.17 (54)	2.69 (56)	3.32 (61)	1.59 (54)	2.31 (54)															
GABt ~2	2.91 (58)	1.57 (57)	2.56 (57)	2.37 (56)	3.19 (66)	2.83 (36)	1.94 (53)	3.72 (59)	1.87 (54)	3.40 (59)	1.78 (58)	2.50 (57)	1.60 (53)	1.98 (54)	2.66 (58)	3.24 (59)	1.49 (52)	2.11 (54)	1.26 (61)														
FH ~5	2.73 (57)	2.30 (54)	2.47 (56)	2.45 (56)	2.37 (57)	3.06 (34)	2.65 (49)	3.02 (54)	2.66 (55)	2.84 (51)	2.41 (54)	2.56 (56)	2.42 (56)	2.18 (56)	2.96 (57)	2.90 (57)	2.52 (54)	2.32 (56)	3.61 (51)	3.86 (57)													
FH ~15	2.51 (58)	3.02 (52)	2.20 (59)	2.79 (59)	2.70 (57)	2.95 (33)	2.65 (49)	2.67 (60)	2.74 (57)	2.43 (58)	3.00 (57)	2.36 (58)	2.69 (58)	2.44 (57)	1.83 (59)	2.08 (61)	3.09 (57)	2.13 (57)	3.18 (55)	3.26 (57)	2.72 (56)											
FH ~16	2.56 (51)	2.77 (50)	1.93 (54)	2.65 (56)	2.57 (52)	2.50 (34)	2.49 (49)	2.05 (56)	2.77 (55)	2.22 (50)	3.09 (55)	2.28 (46)	2.39 (46)	2.19 (55)	2.32 (53)	2.01 (56)	3.02 (55)	2.32 (47)	2.99 (51)	2.98 (51)	2.60 (49)	2.09 (56)										
FH ~19	2.35 (58)	3.02 (53)	2.45 (59)	2.87 (57)	2.83 (59)	3.60 (33)	2.62 (57)	3.37 (59)	2.69 (57)	2.96 (57)	2.89 (57)	2.41 (57)	2.47 (58)	2.39 (57)	2.31 (59)	2.39 (56)	3.11 (57)	1.89 (57)	2.87 (52)	2.86 (53)	2.96 (56)	1.89 (58)	2.55 (53)									
FH ~20	3.00 (58)	2.94 (60)	3.35 (72)	3.45 (56)	3.39 (69)	5.42 (46)	3.62 (77)	4.06 (53)	3.22 (61)	4.91 (56)	2.75 (65)	3.66 (74)	3.17 (73)	3.40 (59)	3.58 (55)	3.58 (59)	3.59 (59)	2.94 (60)	3.38 (56)	3.44 (57)	3.59 (57)	3.38 (50)	3.61 (65)	2.77 (57)								
IL15R ~1	2.44 (57)	3.23 (57)	2.52 (56)	2.94 (56)	3.10 (59)	2.31 (36)	2.52 (50)	3.58 (65)	2.78 (61)	2.96 (55)	2.65 (61)	2.36 (60)	2.70 (60)	2.76 (56)	1.85 (59)	2.25 (58)	3.03 (60)	2.48 (60)	2.65 (53)	2.89 (53)	3.29 (55)	2.46 (55)	2.83 (54)	2.26 (56)	3.39 (72)							
MSP2 ~1	1.78 (61)	1.66 (58)	1.76 (59)	1.66 (59)	0.84 (66)	8.30 (52)	1.78 (59)	3.15 (60)	1.91 (59)	2.67 (59)	1.90 (59)	2.34 (58)	1.59 (60)	1.80 (59)	2.60 (62)	2.80 (62)	1.69 (59)	1.55 (59)	2.47 (59)	2.31 (59)	2.34 (57)	2.65 (58)	2.55 (52)	2.68 (58)	3.32 (70)	3.11 (60)						
MSP2 ~2	2.26 (60)	1.59 (58)	1.62 (59)	1.49 (59)	1.91 (61)	0.91 (41)	1.46 (65)	3.19 (56)	1.75 (59)	2.71 (61)	2.00 (59)	1.91 (60)	1.80 (60)	1.80 (59)	2.62 (62)	3.02 (56)	1.34 (59)	1.73 (59)	2.28 (61)	1.82 (58)	2.57 (57)	2.88 (60)	2.33 (46)	3.35 (66)	4.48 (74)	2.97 (69)	1.87 (61)					
MCP ~1	2.10 (57)	2.77 (56)	2.25 (57)	2.74 (58)	2.63 (61)	2.96 (33)	2.33 (58)	2.60 (56)	2.51 (58)	2.58 (59)	2.82 (58)	2.48 (58)	2.14 (59)	2.11 (58)	1.41 (62)	1.56 (62)	2.74 (58)	1.74 (58)	3.69 (64)	2.61 (54)	3.15 (57)	2.16 (58)	2.25 (52)	2.10 (55)	3.56 (60)	2.22 (58)	2.60 (61)	2.96 (62)				
MCP ~2	2.77 (58)	1.62 (58)	2.12 (58)	1.88 (57)	2.15 (60)	1.68 (33)	1.78 (57)	3.69 (57)	2.00 (62)	3.26 (58)	1.35 (62)	2.72 (61)	2.27 (63)	2.31 (59)	2.94 (59)	3.36 (60)	1.58 (62)	2.36 (62)	2.10 (59)	1.90 (58)	3.02 (56)	3.39 (58)	3.42 (56)	3.20 (56)	3.38 (71)	2.97 (61)	2.27 (60)	1.64 (60)	3.16 (58)			
VCP ~1	3.17 (59)	2.98 (56)	2.71 (60)	2.41 (59)	2.70 (61)	2.56 (35)	3.16 (66)	3.02 (62)	3.59 (62)	2.97 (54)	2.49 (57)	2.78 (58)	2.51 (58)	2.48 (57)	3.14 (61)	3.05 (63)	2.79 (57)	2.53 (58)	3.28 (60)	3.15 (59)	2.54 (57)	3.15 (61)	3.01 (56)	3.17 (58)	3.85 (65)	3.12 (58)	2.74 (60)	2.80 (55)	3.05 (57)	3.31 (58)		
VCP ~2	2.64 (59)	2.20 (58)	2.16 (56)	2.07 (57)	2.32 (58)	2.04 (36)	2.27 (58)	3.45 (58)	2.91 (60)	2.93 (57)	2.64 (60)	3.35 (60)	2.60 (62)	2.32 (57)	2.68 (57)	2.81 (57)	2.77 (61)	2.87 (60)	2.45 (51)	2.47 (57)	2.36 (54)	2.82 (56)	2.69 (51)	2.84 (50)	3.38 (66)	3.54 (60)	2.16 (58)	2.12 (58)	2.96 (58)	2.68 (61)	3.73 (66)		
VCP ~3	2.80 (57)	2.24 (52)	2.00 (57)	2.34 (58)	2.60 (58)	2.05 (34)	2.32 (50)	2.92 (57)	2.24 (57)	2.56 (56)	2.24 (57)	2.27 (57)	2.52 (57)	2.06 (53)	2.41 (58)	2.45 (57)	2.39 (57)	2.14 (57)	2.68 (53)	2.68 (53)	2.74 (56)	2.08 (58)	2.34 (54)	2.68 (58)	3.31 (57)	2.65 (54)	2.65 (58)	2.17 (58)	2.56 (55)	2.78 (58)	2.42 (58)	2.26 (55)	
VCP ~4	2.56 (57)	1.81 (49)	1.67 (57)	1.93 (57)	2.03 (57)	2.54 (33)	2.33 (56)	3.05 (58)	2.32 (56)	2.36 (51)	2.07 (55)	1.91 (57)	1.88 (57)	1.51 (57)	2.62 (57)	2.48 (57)	2.11 (56)	1.88 (57)	3.14 (55)	2.45 (56)	2.09 (56)	2.64 (56)	2.55 (46)	2.55 (56)	3.08 (56)	2.93 (59)	1.97 (57)	2.34 (58)	2.69 (57)	2.68 (57)	2.11 (57)	2.22 (53)	2.17 (55)

Table 2.4: Pairwise comparison of individual CCP module structures based on Cα rmsds using CE. All values are in Angstroms (Å); values in brackets indicate "Alignment length". Colour scheme used: ≤ 2.00 Å (blue); > 2.00 Å and ≤2.50 Å (green); > 2.50 Å and < 3.00 Å (magenta); ≥3.00 Å (black); Alignment length ≤ 40 amino acids (red). Abbreviations as in text except for BGPI = β2GPI; C4BP = C4BPα; GABc/GABt = cis-/trans-$GABA_BR1a$; IL15R = IL15Rα; MSP2 = MASP2.

Using CE, all 34 solved CCP module structures were compared with one another (Table 2.4). The more recently solved module structures may be superimposed by CE on the original mean structure of fH~16 using the Cα atoms of between ~46 and 65 residues (out of ~60). The resulting rmsd values lie between approximately 1.9 and 3.6 Å (see Table 2.4). The most similar to fH~16 is β2GPI~3 (rmsd = 1.93 Å over 54 residues), DAF~2 (rmsd = 2.01 Å over 56 residues) and C4BPα~1 (rmsd = 2.05 Å over 56 residues). The neighbouring module - fH~15 - also overlays relatively well (rmsd = 2.09 Å over 56 residues). Note that the CCP module of known structure with most sequence similarity (42% identity) to fH~16 is VCP~3 (rmsd = 2.34 Å over 54 residues). This is considerably less structural similarity than is generally observed for proteins of such high sequence similarity - *e.g.* 50% sequence identity typically corresponds on average to an rmsd < 1 Å over equivalent backbone atoms (Chothia and Lesk 1986). The most dissimilar modules in structure to fH~16 are IL15Rα~1, MCP~2 and CR1~16 (see Table 2.4). Fewer CCP module structures superimpose well on the structure of fH~5 than on fH~16 – the most similar is VCP~4 (rmsd = 2.09 Å) and the most different are the two forms of $GABA_BR1a$ (rmsd = 3.86 Å (trans) and 3.61 Å (cis)). The second CCP module of C1r was excluded from this analysis since the relevant coordinate files are missing residues in its module making its structural alignment length too short.

Globally, DAF~4, MASP2~2, β2GPI~2 and ~3 appear to be the most conventional out of the set of solved 3D structures – their rmsds (according to CE) are < 2 Å for each of sixteen other modules. DAF~4 can be considered the most "typical" with none of the 33 modules compared > 3 Å, and its rmsd as low as 1.2 Å when compared to CR2~1 (it only deviates by more than 2.5 Å in the case of four modules). The two most similar modules structurally are the first modules of C1r and

MASP2, whose rmsd = 0.84 Å over all residues. These two modules share ~42% sequence identity. The most divergent among all CCP modules are fH~20, C4BPα~1, and VCP~1, with rmsds > 3 Å for each of 29, 19, and 16 other modules, respectively. The two most dissimilar modules structurally are fH~20 and CR1~15 (rmsd = 4.91 Å over 56 residues); fH~20 also deviates by > 4 Å versus MASP2~2 and C4BPα~1. The two conformers of $GABA_BR1a$ deviate from each other by 1.26 Å.

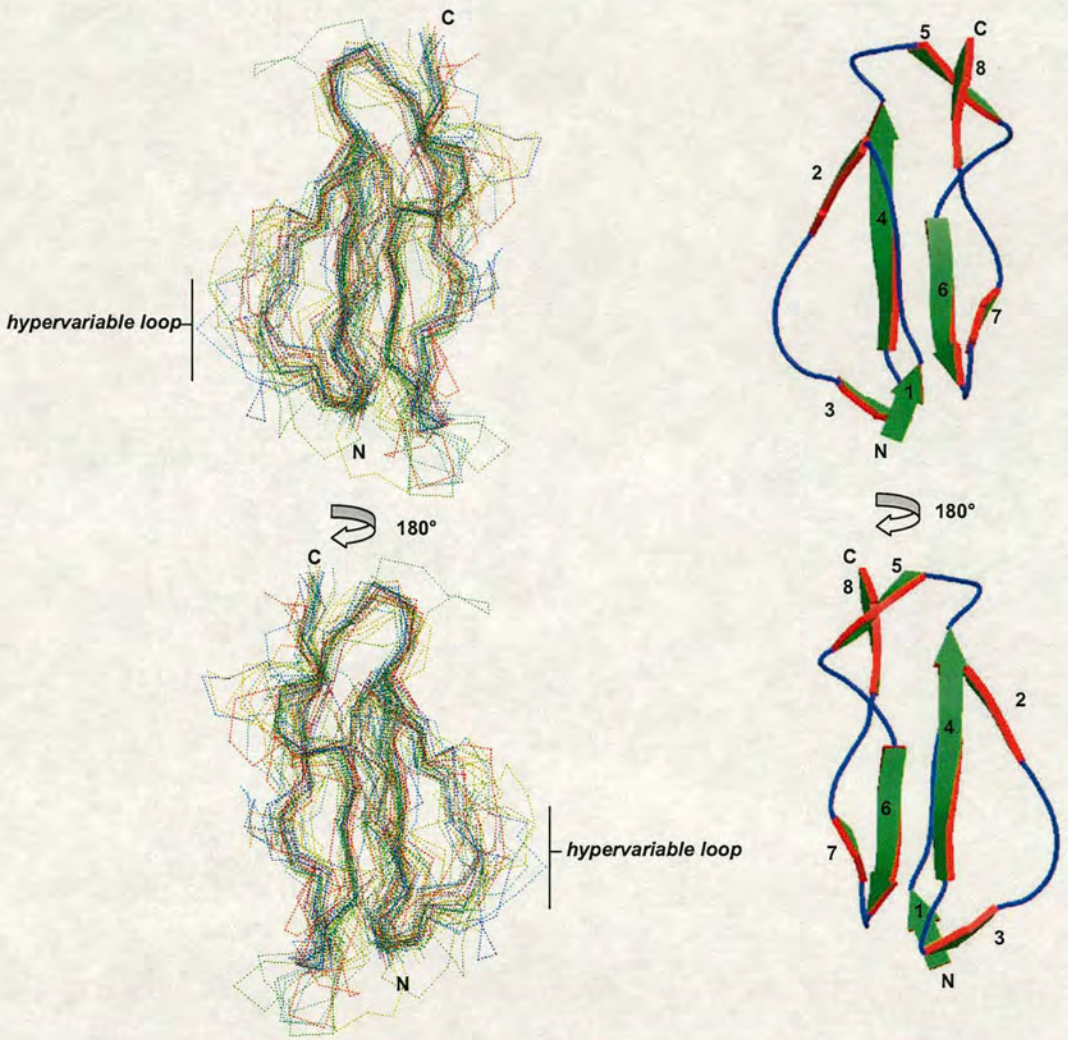


Figure 2.10: Overlay of all 34 CCP structures. A consensus 3D structure emerges from an overlay of all 34 CCP structures solved to date (3rd April 2006). Left-hand side: two views (related by a 180° rotation about the vertical axis) of a superposition (using Multiprot, see text) of all 34 accessible CCP module 3D structures. The structures of the two atypical modules in IL2Rα were excluded from this analysis. To help interpret the overlay, a Molscript representation showing the same views of fH~5 is

drawn on the right-hand side, with strands numbered 1-8 as in Figure 2.8. It is not intended that the viewer try to trace individual structures within the superposition but this figure summarises structural divergence and provides a good representation of the consensus structure; the hypervariable loop is indicated.

Figure 2.10 shows a global overlay of all 34 CCP module structures for which coordinates are currently available. From this superposition, a "consensus module structure" emerges that can be considered in conjunction with the multiple sequence alignment in Figure 2.4 and schematic in Figure 2.8. The best conserved part of the structure corresponds to the four stranded antiparallel β-sheet comprised of strands 2, 4, 6 and 7, with the inner β-strands 4 and 6 being particularly highly conserved. Residues corresponding to β-strands 3 and 5 converge less well and the first and last strands (1 and 8) are the least well-conserved. The sequence connecting β-strands 3 and 4 is never the site of an insertion and is the most highly conserved turn. The 4-5 loop is also structurally conserved – insertions in the 4-5 loop as exemplified in this set of structures by CR1~17, occur only in CR1~7/14/21/28; CR1~3/10/17/24; CR2~5 and 9 (six residue insertions); and MCP~3 (five residue insertion). The 6-7 loop is more variable in size – with CR1~17 illustrating the maximum length it attains amongst the RCA proteins, and IL15Rα~1 having highest deviation among the set. The stretches of residues between strands 1 and 2, are poorly conserved but have the same general appearance in the overlaid structures - an extended region running towards the C-terminus followed by a change of direction so that strand 2 heads towards the N-terminus. The 1-2 region is a site for insertions in only a few modules (*e.g.* C4BPα~6) for which there are no structures as yet. The three regions in the current structure set displaying the highest variability lie between strands 2 and 3 (hypervariable loop), strands 5 and 6, and strands 7 and 8. All these regions lie to the

"sides" of the modules, away from the termini and the junctions with neighbouring modules. Both the 5-6 and the 7-8 regions of the great majority of the solved structures do, nonetheless, converge towards a consensus that is perceptible in the overlay. The 5-6 region is notable for its bulge-like appearance as the main chain leaves strand 5 at the (C^{II} + 8) Gly and makes two sharp bends to direct itself first back into the body of the module and then "down" (in Figure 2.10) to form strand 6 that is aligned with strand 4. Obvious outliers are the C1r, C1s and MASP2 modules that are unusual in having large insertions in both these regions, which form prominent features protruding laterally from the bodies of the module (in the case of C1r~2 and C1s~2 the 7-8 insertions were not visible in the electron density, presumably due to their flexibility). It is only in the 2-3 region that no consensus appears in the overlay – this is the "hypervariable" region referred to previously, and is a site of sequence and length-diversity across the CCP module family (Kirkitadze and Barlow 2001); it is also the site of the *cis/trans* isomerisation in $GABA_BR1a$~2 (Blein et al. 2004).

2.6 Intermodular junctions – structure and mobility

2.6.1 Introduction to junctions

Each intermodular junction may be regarded as being made up from proximal loops and strands of the respective neighbours, plus the "linker". As already discussed, the linker is defined, for convenience, as the sequence of residues between C^{IV} of the preceding module and C^{I} of the next module. The forces that influence the structure and flexibility of such a junction thus consist of: covalent linkages and non-covalent interactions within the linker; interactions between the proximal loops/strands of the neighbouring modules; and interactions between loops/strands of the respective modules and the linker. Approximately 14% of linkers between modules of the RCA proteins have only three residues. This appears to be a minimum with regard to the possibility of both modules being able to fold successfully, consistent with the fact that one or two residues after C^{IV}, and before C^{I}, commonly form part of strands 8 and 1 respectively. The most common linker length (see Figure 2.5) is four residues (~58% of linkers), while linkers of five, six, seven and eight residues also occur (Table 2.1). Four of a total of six eight-residue linkers and two out of a total of four seven-residue linkers are found in CR2, which has on average the longest linkers (5.5 residues) of any of the RCA proteins (cf. factor H – 4.6; CR1 – 4.3). The orientation of one module relative to its neighbour can be characterised by the three angles: tilt, twist and skew (Figure 2.12) described in the Materials and Methods section. Figure 2.12 is a summary of tilt, twist and skew angles (Table 2.6) for all of the CCP junctions that are available in the Protein Data Bank. The surface area (SA) buried in each junction (Table 2.5) is also indicated in Figure 2.12.

2.6.2 Examples of intermodular junction structures - discussion

The initial structural investigation of a pair of CCP modules was performed on the 15th and 16th modules from factor H expressed as a recombinant protein in *S. cerevisiae*. These studies were carried out using two-dimensional ^{1}H-NMR on a pure (no salt or buffer) 2 mM sample of protein at pH 4.7 and 27 °C (Barlow et al. 1993). The basis of an NMR-derived structure is a list of pairs of atoms that are < 5 Å apart in the folded protein – each entry in such a list is referred to as a distance-restraint. While many *intra*modular distance-restraints were inferred from the spectra, *inter*modular distance restraints (*i.e.* between two residues, each residing in a different module) were almost absent. Only one such restraint could be detected. The ensuing structure calculations, based on all distance-restraints, indicated an extended, head-to-tail arrangement of the two neighbouring CCP modules with a very limited intermodular interface (buried surface area = 573 Å^2). The lack of significant contacts between the modules was consistent with the observation that the NMR frequencies of module 16 hydrogen nuclei (protons) were independent of whether module 16 was studied in isolation or in the context of the pair. A preferred intermodular orientation was evident amongst the ensemble of calculated structures, presumably due to distance-restraints between residues of the linker and residues of near-by loops from the two modules. A subsequent structure calculation, however, using a different computational algorithm (Wiles et al. 1997), indicated a somewhat different intermodular conformation. It is highly probable that significant flexibility exists between modules 15 and 16 of factor H. On the other hand, the shortness of the linker (four residues) and the steric bulk of the two modules must limit the freedom of intermodular movement to some extent - disallowing for example a conformation in which the two modules are organised in a side-by-side arrangement.

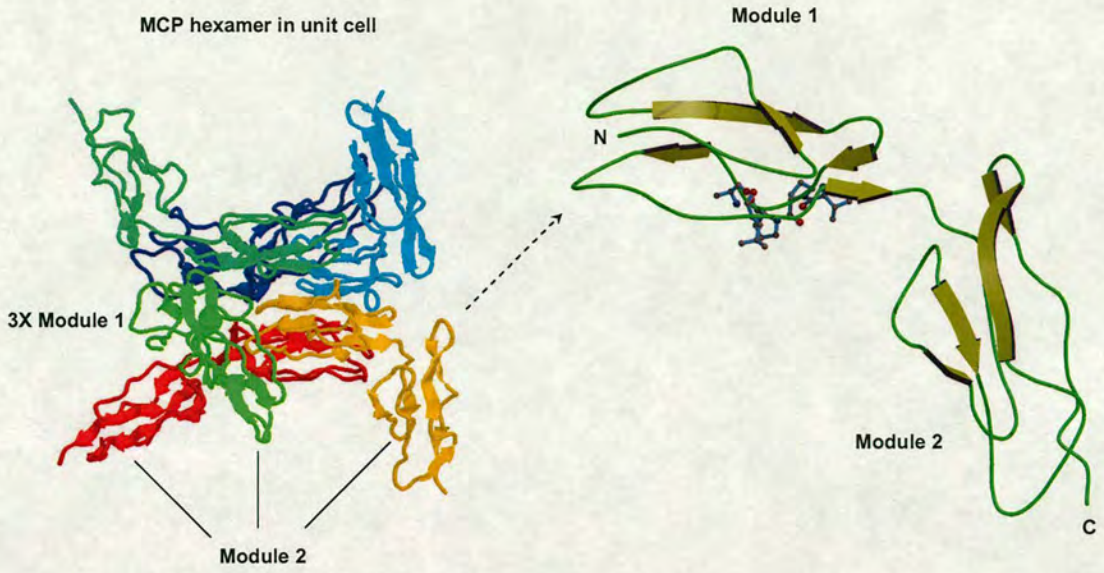


Figure 2.11: The 3D crystal structure of the N-terminal two CCP modules of MCP. The left-hand panel shows the arrangement of the six molecules in the unit cell, with each molecule drawn in a different colour. The right-hand panel shows a cartoon of one molecule. Some residues that were proposed to represent a key interaction site for measles virus binding are drawn as ball-and –stick.

Subsequently the structure of a second module-pair was determined using X-ray crystallography (Casasnovas et al. 1999). The N-terminal two CCP modules of MCP (MCP~1,2) were expressed in Chinese hamster ovary cells as a high-mannose glycoform, and their structure was solved at a resolution of 3.1 Å. Crystals were grown at 20 °C, pH 6.5 [in 16% 8 kDa polyethylene glycol (PEG)] and, notably, 40 mM $CaCl_2$. The asymmetric unit of the crystal contained six molecules of MCP~1,2. Two sets of three molecules each formed a "three-legged table" in which each leg consists of module 2 (Figure 2.11). One table is inverted and it stacks on top of the other creating the hexamer. Extensive contacts between molecules were observed, and these were particularly evident within each trimer. The N-terminal end of the first module contacts residues within the small interface between the two CCP modules in the neighbouring molecule. A Ca^{2+} ion appeared to be ligated by aspartate residues of CCP module 1 lying close to the interface and this ion participates in solvent-mediated contacts with module

2. Bearing in mind the extent of inter*molecular* interactions compared to the more limited intramolecular inter*modular* contacts, it seems likely that the N-terminal two domains of a monomeric version of MCP would be less constrained relative to one another than they are in the crystallographic hexamer. Some evidence for flexibility derives from the range of conformations within the unit cell (see Figure 2.12). The tilt angle between the two CCP modules differs between members of the hexamer over the range 61 to 75 ° (mean 68.3 ± sd 5.7 °, n = 6), and the twist and skew angles are -158±7 ° and –20±2 ° respectively. While Ca^{2+} was needed for crystallisation, the physiological relevance of the Ca^{2+} ion, and its role in stabilising the junction within the crystal are not entirely clear. The crystal structure of MCP~1,2 thus leaves unresolved the issue of the flexibility of the intermodular junction under physiological conditions. It remains possible that in its physiological state it is just as flexible as the factor H 15-16 module pair. An additional interesting aspect of the MCP~1,2 structure is the presence of a glycan moiety on each module (N-linked to N49 and N80). These occupy the concave surface created by the tilt angle between modules, and project towards one another. The N-glycan on N80 (in module 2) could have a stabilising influence on the structure of that module since it has hydrophobic contacts with hydrophobic side-chains that might otherwise be solvent-exposed. This could explain the requirement of N-glycosylation at N80 for virus binding (Figure 2.11) (Maisner et al. 1996).

CCP-containing protein	PDB code(s)	Module 1	Module 2	Bimodule	Buried
C4BPα	2A55*$_{1,2}$	4340.74	4365.78	8194.07	512.45
factor H	1HFH*$_{15,16}$	4161.13	3843.43	7431.17	573.39
	2BZM*$_{19,20}$	3920.47	4388.50	7576.76	732.21
DAF	1H03^$_{3,4}$	4072.09	4303.13	7864.32	510.90
	1NWV*$_{2,3}$	4684.01	4044.11	8275.20	452.92
	1OK3^$_{1,2}$	4314.42	4306.75	8141.22	479.95
	1OK3^$_{2,3}$	4403.66	3992.75	7793.96	602.45
	1OK3^$_{3,4}$	4033.63	4350.62	7848.45	535.80
MCP	1CKL^$_{1,2}$	4576.91	4220.77	8340.89	456.79
CR1	1GKN*$_{15,16}$	4489.55	3976.54	7971.51	494.58
	1GKG*$_{16,17}$	4068.81	4809.27	8380.24	497.84
CR2	1LY2^$_{1,2}$	4700.86	4180.76	7897.61	984.01
VCP	1G40^$_{1,2}$	5341.84	4201.85	8653.74	889.95
	1G40^$_{2,3}$	4103.66	3979.33	7492.39	590.60
	1G40^$_{3,4}$	3988.90	4083.96	7261.93	810.93
	1VVC*$_{3,4}$	3686.33	3785.37	7080.40	391.30
	1E5G*$_{2,3}$	3961.39	3454.97	7146.18	270.18
C1r	1GPZ^$_{1,2}$	4672.76	4759.40	8954.35	477.81
β2-GPI	1QUB^$_{1,2}$	4173.13	3942.93	7619.73	496.33
	1QUB^$_{2,3}$	3928.82	4134.70	7700.87	362.65
	1QUB^$_{3,4}$	4186.09	3874.36	7491.51	568.94
MASP-2	1ZJK^$_{1,2}$	4178.72	4244.20	7987.84	435.08

Table 2.5: Calculated buried surface area among CCP bimodules. All units are in Å^2. Experimentally determined structure pairs up to 17th April 2006 included.

Modules	PDB code	Order	Tilt	Twist	Skew	Tilt Error	Twist Error	Skew Error
VCP~13-X	1G40	1	4.4	-9.9	66.5	1.6	8.6	8.1
DAF~34-X	1OK3	2	7.6	113.7	59.1	1.5	9.8	3.9
B2GPI~12-X	1QUB	3	10.9	133.8	-139.9	na	na	na
FH~1920-N	2BZM	4	11.1	15.5	149.7	6.4	8.4	ne†
DAF~34-X	1H03	5	12	112.6	39	6	5.6	39.5
MASP2~12-X	1ZJK	6	12.5	79.1	119.4	na	na	na
CR1~1617-N	1GKG	7	16.3	110.8	-6.6	9	18.5	41.5
C1r~12-X	1GPZ	8	22.2	97.6	124.9	na	na	na
DAF~24-X	1OK3	9	24.9	-55.8	-54.1	4.8	13.4	1
VCP~23-N	1E5G	10	31.2	144.1	77.6	10.4	11.9	24.7
CR1~1516-N	1GKN	11	31.2	161	-44.9	7.3	13.7	21.5
DAF~23-X	1OK3	12	31.5	-168.5	-63.5	5.3	4.4	2
B2GPI~23-X	1QUB	13	32.9	89.1	60.1	na	na	na
VCP~24-X	1G40	14	35.2	-12.7	-91.5	1.3	3.5	3.6
DAF~13-X	1OK3	15	37	114.7	139.1	2.1	8.2	3.4
CR1~1517-N	1GOP	16	37.6	-81.7	-19.9	na	na	na
C4BPa~12-N	2A55	17	38.4	3.7	-110.2	4.4	5.7	17.9
DAF~14-X	1OK3	18	39.3	-134.1	152.5	0.4	17.7	0.5
DAF~12-X	1OK3	19	41.1	-87	107.4	0.1	2.8	10.4
B2GPI~13-X	1QUB	20	41.9	-139.2	-143.7	na	na	na
DAF~23-N	1NWV	21	45.4	128.7	-127.5	20.9	15.3	24.2
B2GPI~24-X	1QUB	22	48.4	144.1	103.3	na	na	na
FH~1516-N	1HFH	23	50	-130	155	13	17	23
B2GPI~34-X	1QUB	24	53.2	40.2	100.2	na	na	na
B2GPI~14-X	1QUB	25	56.3	-78.5	-121.4	na	na	na
VCP~12-X	1G40	26	62.8	-18.5	69.1	4	9.2	8.5
VCP~23-X	1G40	27	64.6	7.4	-91	1.7	2.6	3.6
VCP~34-N	1VVC	28	67.2	7.2	81.3	6.3	9.2	11
MCP~12-X	1CKL	29	68.3	-158.3	-20.3	5.7	7	2.2
VCP~34-X	1G40	30	97.9	-31.5	66.5	4.1	10	8.2
VCP~14-X	1G40	31	99.7	-19.4	75.4	1.3	2	1.2
CR2~12-X	1LY2	32	142.2	-4.2	43.6	2.6	20.1	2.9

Table 2.6: Intermodular angles for all CCP module pairs and larger fragments. Tilt, twist and skew angles plotted for all experimentally determined structures of multiple CCP-module fragments. The values for pairs of modules and for non-consecutive modules (red), in order of increasing tilt angle (°). Modules solved by X-ray diffraction marked with an "**–X**", while structures solved by NMR marked with an "**-N**" and corresponding PDB codes shown. Error values represent the standard deviation of angles observed, calculated for structures where several molecules are present in the unit cell of a crystal or where an ensemble exists for structures solved by NMR (na = not applicable; ne† = not entered; value seems to be an overestimation of intermodular flexibility, and could be due to local flexibility).

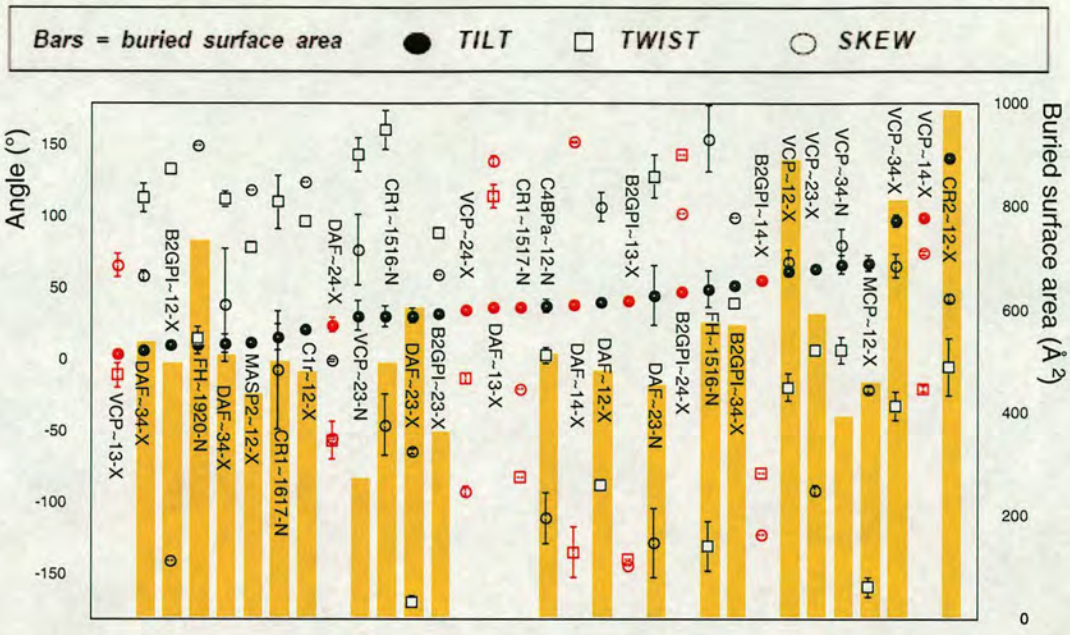


Figure 2.12: Graphical representation of intermodular angles and buried surface area. Skew (transparent dots), tilt (solid dots) and twist (squares) angles plotted (left-hand scale, see key for symbols) for all experimentally determined structures of multiple CCP-module fragments. Values for pairs of modules are plotted in order of increasing tilt angle; the identity of the protein and the two modules involved is marked on the graph; N = solved by NMR; X = solved by X-ray diffraction. Red symbols are used to emphasise that some of the pairs analysed are non-adjacent. Where more than one structure is available – *i.e.* where there is an ensemble of NMR-derived structures or where several molecules are present in the unit cell of a crystal – the data points plotted are average values and error bars represent the standard deviation of angles observed. Buried surface areas for each junction are plotted as yellow bars (right-hand scale). Details of the calculation are given in the text. The raw data is in the tables on previous two pages.

Two papers (Bouma et al. 1999; Schwarzenbacher et al. 1999) subsequently described the crystal structure of human β-2-glycoprotein I (β2GPI) purified from plasma. While this protein is not directly involved in the complement system, this work was of interest since it revealed three new inter-CCP module junctions; the fifth CCP module is aberrant and so the 4-5 junction is excluded from the overall analysis. Crystals were grown at high salt [1.5 M $(NH_4)_2SO_4$ or $(NH_4)_2HPO_4$] and either pH 5.6 or 7.5. Both X-ray diffraction studies of β2GPI produced very similar structures to one another with excellent conservation of the intermodular angles. As in previous examples, the

modules are arranged end-to-end with relatively few intermodular contacts. In particular the 2-3 intermodular junction buries little surface area (362 $Å^2$) and there is a relatively small tilt between these modules (tilt = 33 °; twist = 89 °; skew = 60 °) (see Figure 2.12). The 1,2 junction in both β2GPI structures, although burying a larger SA (496 $Å^2$) than the 2,3 junction exhibits an even shallower tilt – just 10.9 ° on average (twist = 134 ° and skew = -140 °) – thus making it one of the most elongated module pair observed to date. The β2GPI 3,4 junction is more markedly tilted (tilt = 53 °, twist = 40 °; skew = 100 °). In the light of previous comments on intermodular orientations of MCP within the crystal lattice, it is of interest that a subsequent solution study of the structure of β2GPI (Hammel et al. 2002), based on small-angle X-ray scattering (SAXS), suggested that the "J-shaped" conformation observed in both crystal structures does *not* predominate – an alternative "S-shaped" solution conformation of the protein was proposed that is more consistent with the SAXS data. In this conformation the 2-3 junction has a 60 ° tilt rather than the 33 ° tilt observed in the crystal structure. Various explanations for the inconsistency between the crystal structures and the scattering data are possible, including a dynamic interchange of still other conformations. Thus, like the structure of MCP~1,2, the crystal structures of β2GPI may reveal only a very limited set of the possible conformations at the 2-3 junction that occur under physiological conditions. One consideration, advanced by the authors of the SAXS study (Hammel et al. 2002) is that the high ionic strength of conditions used for crystallisation was such that hydrophobic interactions between modules could be artificially strong.

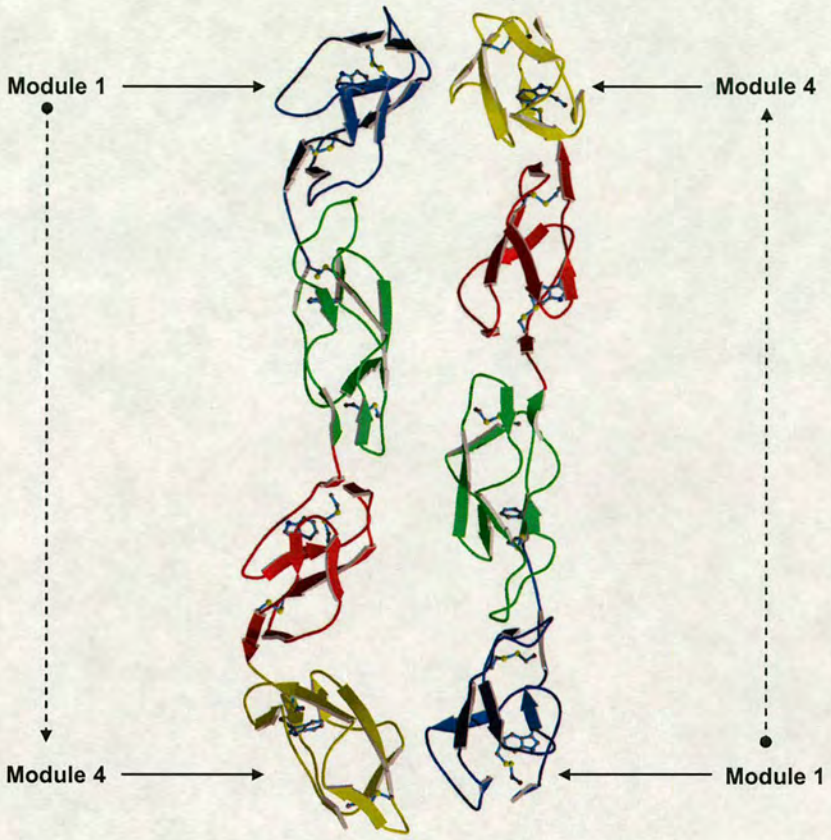


Figure 2.13: The 3D crystal structure of intact VCP. A Molscript representation of both molecules (A and B) of VCP in the unit cell, with each CCP module drawn in a different colour. The relatively open junction between the central modules is evident, as the more extensive junction between modules 1 and 2.

Another example of differences between the conformations of solution and crystal structures came to light with the determination of the first complete complement protein structure – that of VCP (Murthy et al. 2001) (Figure 2.13). Crystals of VCP were obtained at pH 7.5, 20 °C and in ~10% PEG 6000 or 8000. The three protein molecules in the unit cell of one crystal form of intact VCP, and two molecules in another, exhibited essentially the same tilt, twist and skew angles at the three respective intermodular junctions, suggesting a structure (Murthy et al. 2001) that is rigid throughout its length. Within each molecule, the three tilt angles (ranging from 63-100 °) are larger than the tilt

angles observed in β2GPI, but the 2-3 junction buries the least SA (591 $Å^2$). Interestingly, the solution structure of the 2,3 CCP module pair of VCP determined by $^1H,^{15}N$-NMR (Henderson et al. 2001), together with measurements of NMR relaxation and a series of other biophysical studies (Kirkitadze et al. 1999b), strongly indicated flexibility at this junction. The solution structure of a VCP 3,4 module-pair had also been solved using $^1H,^{15}N$-NMR (Wiles et al. 1997). Unlike in the cases of VCP~2,3 and fH~15,16, for VCP~3,4 a useful number of distance-restraints involving hydrogen nuclei on different modules were obtained. In contrast to the VCP~2,3 situation, these allowed the intermodular orientation to be calculated with some confidence, although there were insufficient distance-restraints to define it fully (tilt = 67±6 °; twist = 7±9 °; skew = 81±11 °). The subsequent crystal structure of intact VCP (Murthy et al. 2001) yielded a more tilted but roughly similar 3,4 intermodular orientation (see Figure 2.12, tilt = 100±1 °; twist = -19±2 °; skew = 75±1 °) with a junction involving the same residues. Thus in the case of VCP~3,4, the differences between crystal and NMR structures was less significant – and both methods were consistent with a relatively inflexible junction.

A third example where crystal and solution structures have been interpreted in a complementary fashion is provided by studies of the N-terminal modules of CR2. An unusual side-by-side orientation of these two modules (reflected in a very large tilt angle of 142 °, and buried SA of 984 $Å^2$), made possible by the eight-residue linker, was observed by crystallography for both free and complexed (with C3d) forms (Szakonyi et al. 2001; Prota et al. 2002). On the other hand, biophysical solution studies and sedimentation modelling were consistent with a more extended arrangement (Guthridge et al. 2001). Additionally, another solution study using X-ray scattering and sedimentation modelling (Gilbert et al. 2005) revealed a high degree of flexibility in the

linker between the two modules, that was able to adopt a wide range of different conformations.

A fourth example, the central modules of decay accelerating factor (DAF~2,3), were solved both in solution as a module-pair (Uhrinova et al. 2003), and in its intact four module form, by crystallography (DAF~1-4) (Lukacik et al. 2004). While there is reasonable agreement between NMR and X-ray at the level of the structures of the individual CCP 2 and 3 modules (Table 2.3), the crystal structure solved under conditions used to encourage crystal growth (4°C, 20% polyethylene glycol, pH 4.6) indicated a more closed and rigid CCP2-CCP3 interface (burying the alkyl components of Arg and Lys side chains previously inferred to have direct binding roles) (buried surface = 602.45 $Å^2$; tilt = 31.5±5.3 °; twist = -168.5±4.4 °; skew = -63.5±2 °) than the solution one, which under physiological conditions, suggested an open and semi-flexible junction (buried surface = 452.92 $Å^2$; tilt = 45.4±20.9 °; twist = 128.7±15.3 °; skew = -127.5±24.2 °) exposing positively charged side chains. As a result, the two papers (Uhrinova et al. 2003; Lukacik et al. 2004) suggest different interpretations with respect to functional versus structural roles of residues in and near the junctions between CCP modules, and the "true" nature of the junction arrangement is still a matter of debate and investigation (Brook et al. 2004).

These examples demonstrate that junctions that are flexible in solution under approximately physiological conditions may be rigidified either by the conditions used for growing crystals, or by the process of forming the crystal. In the latter case, the observed conformation might reflect one of several, or even many, possibilities that is energetically favoured in solution. Clearly, caution should be exercised when inferring from these structures the proximity of residues that are functionally important within the same binding site, but are on different modules. It is known, however, that some

junctions between CCP modules are relatively rigid, even in solution. The interface between modules 15 and 16 of CR1 (Smith et al. 2002) is another example of a more defined junction, as indicated by NMR and a range of other biophysical data (Kirkitadze et al. 1999a; Kirkitadze et al. 1999c); a number of NOE-derived distance restraints between the two modules, as well as between the linker and the modules were observed.

Crystal structures for the pairs of CCP modules in classical and lectin pathway complement proteins, C1r (Budayova-Spano et al. 2002b) and MASP2 (Gal et al. 2005), that both possess identical overall domain architecture, revealed highly analogous individual CCP module structures [when compared with each others corresponding modules (rmsd < 1 Å, Table 2.4)], and also revealed comparable intermodular small tilts (Table 2.6). A noticeable feature for these pairs of modules was the linker region of four residues (372-375 in C1r and 362-365 in MASP2) between the two CCPs, which formed part of one continuous β-strand in both C1r and MASP2. These regions in C1r and MASP2 formed backbone-backbone hydrogen bonds and van der Waals contacts with residues from both CCPs. Even though the amino acid sequence of the linker region possessed only 50% similarity in MASP2 and C1r, the segments that contacted this linker (*e.g.* in MASP2 residues 330-335, 414-416, and 387-388) when compared with equivalent segments in C1r, revealed a high degree of similarity (~67%) (Gal et al. 2005), forming a basis of explanation for similar relative orientations observed in the MASP2 and C1r structures.

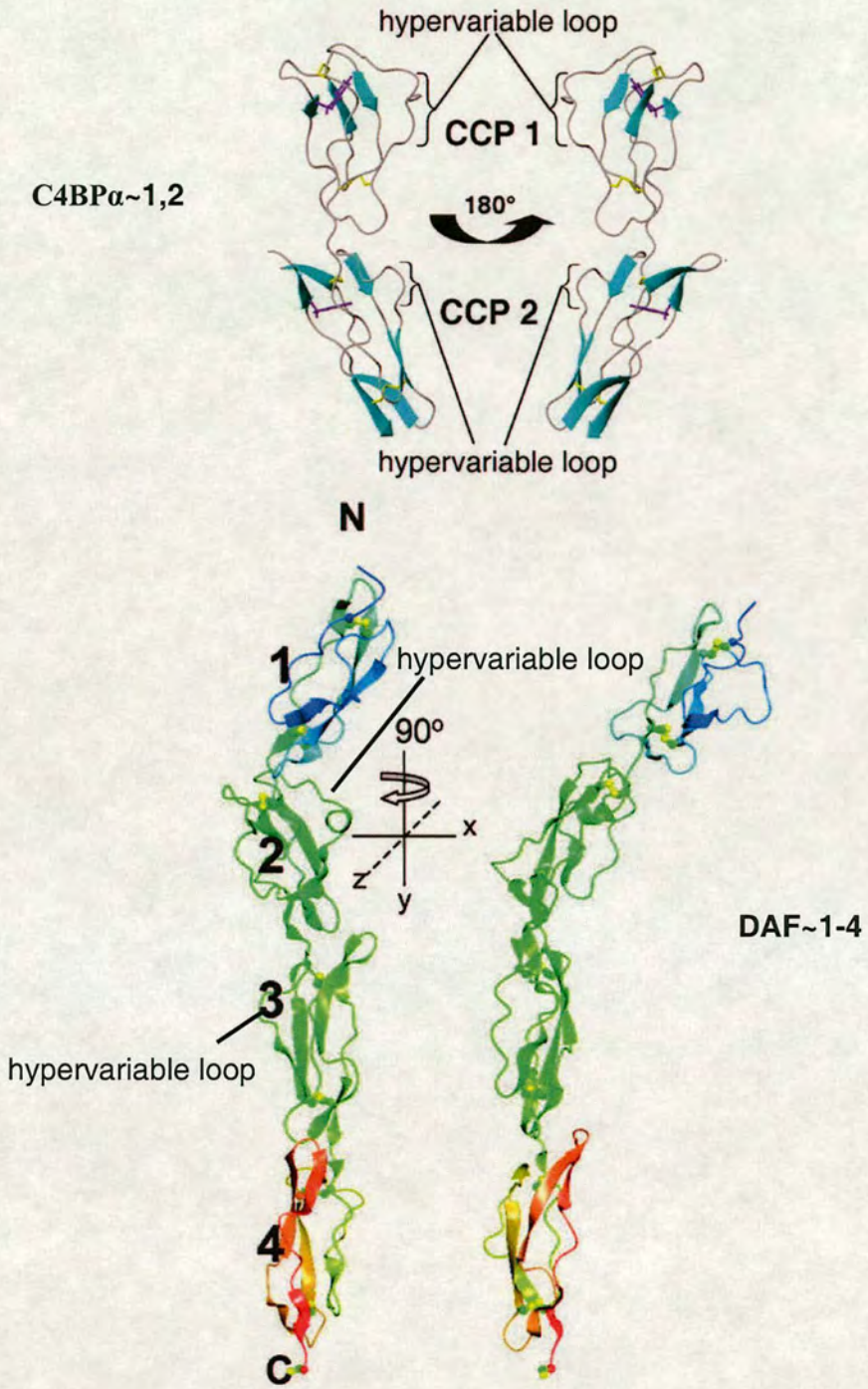


Figure 2.14: Location of hypervariable loops in C4BPα~1,2 and DAF~2,3. The hypervariable loops in consecutive C4BPα modules 1 and 2, lie on the same side of the *y*-axis (Jenkins et al. 2006), whereas in DAF~2,3 they lie on opposite sides (Lukacik et al. 2004). Thus the respective intermodular orientations greatly complicate interpretation of mutagenesis data on the basis of sequence alignments alone; apart from other evidence of flexibility and rigidity among regulators, and the fact that more than one module is involved in a combinatorial fashion for complement regulation (See Table 2.6 for intermodular angles). Figures adapted from (Lukacik et al. 2004; Jenkins et al. 2006).

The latest two structures to be solved were solution structures of the two N-terminal modules in C4BPα (Jenkins et al. 2006), and the C-terminal two modules in factor H (Herbert et al. 2006c). Like CR1~15,16, both these module pairs (C4BPα~1,2 and fH~19,20) appeared to show limited flexibility, consistent with sets of module-to-module and module-to-linker distance restraints that dictate a well defined mutual orientation of the CCP-module pairs. The module pair in C4BP exhibits small twist angles, and a unique arrangement, where both hypervariable loop regions of adjacent modules occur on the same side (Figure 2.14).

In Figure 2.12 intermodular angles of all the available NMR and crystal structures of CCP module-pairs, are plotted in order of increasing tilt angles. This figure also contains data for the orientations of non-neighbouring modules. The uncertainties (reflecting experimental limitations, flexibility or both) are greatest for the NMR-derived structures. Average tilt angles for the neighbouring pairs range from 8 to 142 °, with over half the structures displaying tilts between 31 and 64 °. There is no correlation between tilt angles and twist or skew angles – the latter two angles cover almost the full 0 to +/-180 ° range. There is a trend for higher buried SA with higher tilt-angles to the extent that outliers can be commented upon. The VCP~1,2 junction buries a large SA for its tilt angle but this can be explained by the specific nature of the intermodular interface (see Figure 2.13). On the other hand, MCP~1,2 buries a small SA relative to the intermodular tilt – but as mentioned previously there are complications here due to glycosylation, crystal contacts and the Ca^{2+}-ion present at the intermodular interface.

2.6.3 Some emerging themes

With the exception of the CR2~1,2, DAF~1,2 and fH~19,20 pairs, all linkers within the 3D structures solved so far consist of four residues. Inspection of the various intermodular interfaces within this set of structure reveals the following. ***(i)*** The backbone of the linker participates in H-bonds with residues in the body of the module – the N-terminal half of the linker frequently forms part of β-strand 8 (which has H-bonds to strand 5 and a disulfide bond to strand 4, see Figure 2.8b) and the C-terminal half is sometimes a part of β-strand 1 (H-bonded to strand 3, disulfide linked to strand 6). ***(ii)*** The consequence of these linker residues being involved in β-strand formation is that few covalent bonds within the linker are freely rotatable. Combined with the steric bulk of the modules this limits flexibility at the junction even in the absence of favourable interactions between the bodies of the modules. ***(iii)*** Because of the lack of potential flexibility and the fact that the bodies of the modules are tethered by a short linker, there is inevitably the potential for van der Waals contacts between modules to provide energetically favoured conformations. Due to the generally small size of the interface and possible lack of specificity such conformations may not occupy deep potential energy wells, and may not be unique – but some conformations may be sufficiently favoured to be "frozen out" by crystallisation, especially where crystals are grown in the cold, or under far from physiological conditions (*e.g.* high salt, high PEG). ***(iv)*** Where alkyl and aromatic side-chains (or long polar or charged side-chains with alkyl segments) are present, there is the potential for more specific interactions created by both close packing (van der Waals) and exclusion of water (hydrophobic effect). The linker – which frequently contains residues with large side-chains - and both modules can contribute to such a junction which consequently has the potential to be relatively rigid. An example of such a junction is that of VCP~3-4. ***(v)*** Equivalent residues within

modules participate in most junctions – these are located in: the regions just prior to strand 2, and strand 5 itself (at the N-terminus); and strand 3 and the 3-4 loop, and the 6-7 loop (at the C-terminus). Inspection of these residues in sequence alignments could provide clues as to the sort of junction that might be formed. ***(vi)*** The 1-2, 3-4, 4-5 and 6-7 loops are frequently the sites of insertions and these may create unique junctions. For example in VCP~1,2, the side-chain of Arg40 in the 4-5 loop of module 1 is extended and largely buried; it runs parallel with the linker, making many contacts with linker residues, and its guanido- group participates in an unusual *inter*modular H-bond with the carboxyl group of Asp84. On the other hand, such insertions do not necessarily participate in the junction – an example is CR1~16's 1006-1010 (or C4BPα~2's 108-112) insertion between strands 6 and 7.

In conclusion, a very wide range of intermodular orientations has now been observed and the various junctions between CCP modules may have varying degrees of flexibility. Under the circumstances it is hardly surprising that it has proved impossible to predict, on the basis of the residues likely to be involved, either the structure or the flexibility of the junction. Therefore, for the time being, the exploitation of homology for reliably predicting RCA protein structures is limited to individual modules (Chapter 3). Still more experimental evidence is required before junction structures and intermodular orientations can be predicted by homology. There is clearly a strong case for future extensive mutagenesis and structural studies of intermodular junctions, as well as structural work on further new examples of junctions with native sequences.

2.7 Concluding remarks

The past decade and a half has delivered a wealth of detail for the CCP modules that are the structural and functional building-blocks of the regulators of complement activation. The sequence and structure comparisons undertaken in this chapter provide a basis for further sub-family specific classification of CCP-modules that can be exploited using homology modelling to fill gaps in structural knowledge for other modules (see Chapter 3). There is a need for more high-resolution studies of protein fragments; by NMR, which can deliver information on structure and flexibility under approximately physiological conditions and in solution, and crystallography, which give access to larger fragments or intact proteins, and produce more accurate structures at the level of individual modules. Mutagenesis of junction residues combined with high-resolution structural work will shed much-needed light on the relationship between primary structure and intermodular orientation. Biophysical studies, such as FRET, are needed to obtain data on very long fragments, too large for NMR and too flexible for crystallography. There is a particularly pressing need for more studies of the RCA proteins in complex with their binding partners. The "holy grail" would be a structure of an intact RCA with C3b or C4b (and perhaps with factor I). But more realistic targets such as mapping RCA residues involved in ligand binding using NMR, the use of biophysical techniques to monitor changes in overall conformation upon ligand-binding, or high-resolution structural work on complexes of RCA fragments with smaller ligands (fragments of C3b/C4b, or pathogenic proteins) would also be immensely valuable.

2.8 Declarations and Acknowledgements

- Most of the work presented here appeared in:

  Soares, D.C., and Barlow, P.N. (2005). Complement Control Protein Modules in the Regulators of Complement Activation. In *Structural Biology of the Complement System*. (eds. D. Morikis, and J.D. Lambris), pp. 19-62. CRC Press, Taylor & Francis Group, Boca Raton.

- The program to calculate intermodular angles 'mod22' was written by Bruno Kieffer, University of Strasbourg, and used with permission.

- All scripts are provided in the Appendix and on CD.

Chapter 3

Large-scale modelling as a route to multiple surface comparisons of the CCP module family

---

3.1 Motivation for modelling CCP modules

A sizable fraction of the vertebrate proteome consists of proteins that are composed from a limited repertoire of domain- or module-types (Bork et al. 1996). In the cases of the most commonly occurring types of modules there are now numerous experimentally-derived three-dimensional (3D) structures (Copley et al. 2002). These provide the opportunity to predict structures for the remaining examples of a particular module-type through comparative modelling. The resolution attained with this widely used technique (also known as homology modelling) is high enough to advocate the use of such models as approximations of the true 3D structures (Baker and Sali 2001; Peitsch 2002). In turn, a well-populated database of 3D structures allows, within a class of modules, comparisons based upon structural features rather than sequences. Such an approach should provide a richer source of information for the design and rationalisation of mutagenesis experiments; it might also shed light on phenotypes associated with specific polymorphisms. Ultimately structure-based comparisons could be a useful tool in the prediction of function, identification of functional sites (Nicholls et al. 1991; Blomberg et al. 1999; Pawlowski and Godzik 2001), and design of therapeutic proteins in appropriate cases. The most obvious

features to use in a comparison of module structures of the same type are their surfaces.

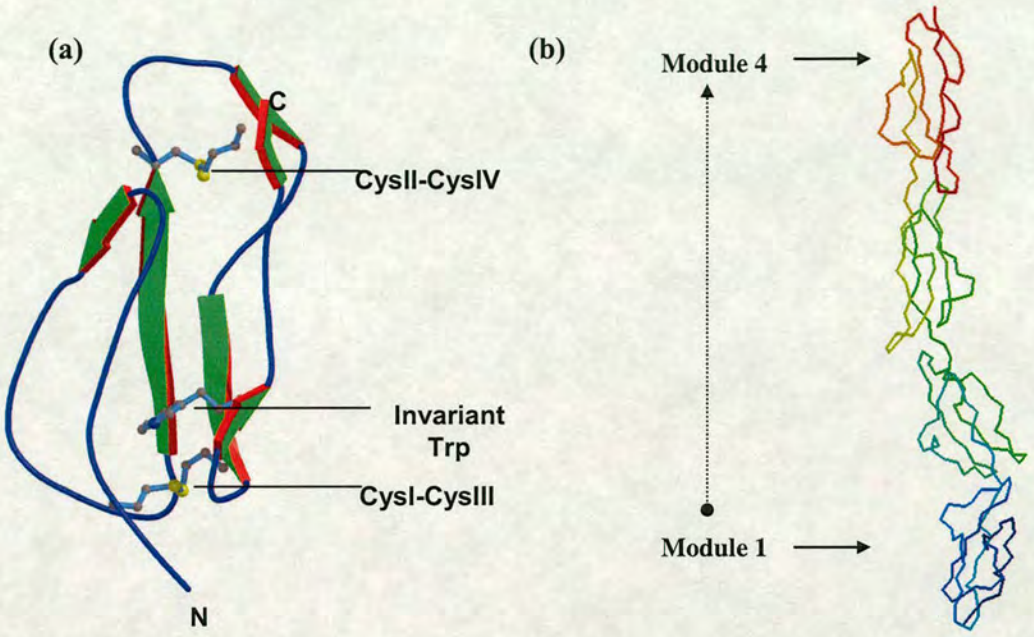


Figure 3.1: The CCP module fold and arrangement. (a) Molscript (Kraulis 1991) representation of the 3D-structure of the first CCP module [the 16th module of factor H, PDB ID: 1HCC (Norman et al. 1991)] revealed by NMR spectroscopy. The consensus "scaffold" of conserved residues, involving the cysteines (forming two disulfide bridges) and a buried tryptophan are drawn as ball-and-stick. β-strands are shown as arrows. (b) Rasmol (Sayle and Milner-White 1995) Cα-trace representation of the 3D-crystal structure of the intact four CCP modules of DAF [PDB ID: 1OK3 (Lukacik et al. 2004)]. The four modules are arranged in a head-to-tail fashion connected by their short linkers.

The 3D structures of a wide range of CCP modules - 25 in total, including 14 CCP modules from human RCA proteins - have been determined experimentally over the last decade [count up to May 2005 (Soares et al. 2005); update provided in Section 3.5] (See Figure 3.4, CCP-module structures solved shown in blue). Each approximately 60-residue CCP module is characterised by a compact hydrophobic core wrapped in a β-sheet framework, held together by two strictly conserved disulfide bridges (Kirkitadze and Barlow 2001) (Figure 3.1, Chapter 2). Because it is time-consuming to solve 3D structures by NMR or X-ray crystallography, many

attempts to model CCP-module structures have been reported (Norman et al. 1991; Kuttner-Kondo et al. 1996; Chou and Heinrikson 1997; Villoutreix et al. 1998; Liszewski et al. 2000; Ranganathan et al. 2000; Aslam and Perkins 2001; Hellwage et al. 2002; Perkins and Goodship 2002; Weyer et al. 2004; Saunders et al. 2006). In most cases, however, the CCP module structures within these models were based on a smaller set of structural templates than are now available, non-optimal templates, or in the case of Chou and Henrikson (1997) – erroneously created, when the module structure was already published. Consequently, there is currently scope for improving the quality of modelled CCP structures.

Module deletion and site-directed mutagenesis approaches have revealed the identity of numerous functional CCP modules (Medof et al. 1982; Krych et al. 1991; Brodbeck et al. 1996; Hardig et al. 1997; Krych-Goldberg et al. 1999; Jokiranta et al. 2000; Liszewski et al. 2000; Blom et al. 2001a; Kuttner-Kondo et al. 2001). This, however, is restricted mainly to the RCAs, and is not exhaustive even for this set of proteins. Moreover, at the level of individual side-chains, functional-site mapping through experiment is still incomplete in most cases.

While computational, sequence-based methods exist that screen for residues important for function on the basis of sub-family specific conservation (Lichtarge et al. 1996; Glaser et al. 2003), such analyses are hampered if the location of functional patches on different modules is not equivalent (with reference to the common structural scaffold). For CCP-modules, such non-equivalence is strongly suggested by the observation that in most cases several neighbouring modules are implicated in interactions with a single asymmetric partner protein. Because protein-protein and protein-ligand interactions are a surface phenomenon, one approach to suggesting the location of functional sites in this case is to examine the properties of CCP-module

surfaces. The concept of surface-to-surface comparison holds considerable promise in analysing models based upon sequence homology (Blomberg et al. 1999). Surface electrostatics play a key role in protein-ligand binding and protein-protein interactions. Electrostatic properties determine the relative orientation during molecular recognition, whereas charge-charge and hydrogen bonding interactions contribute to binding specificity and affinity [see (Honig and Nicholls 1995) for a review]. It has been demonstrated that where similar electrostatic properties are shared by a set of proteins, this may indicate similar behaviour and function (Ullmann et al. 1997; Botti et al. 1998; Wade et al. 1998; Blomberg et al. 1999). There is plenty of evidence for the involvement of electrostatic interactions in the recognition by RCA proteins of C3b and C4b, and glycosaminoglycans (Pangburn et al. 1991; Blackmore et al. 1996; Blackmore et al. 1998b; Krych et al. 1998; Prodinger et al. 1998; Blom et al. 1999; Blom et al. 2000b; Brodbeck et al. 2000; Liszewski et al. 2000; Blom et al. 2001a; Kuttner-Kondo et al. 2001).

In this chapter, a large-scale modelling strategy targeted at individual modules is presented, with the aim of providing surface information for a large number of modules with the highest possible precision. The strategy makes optimal use of the current set of available experimentally-derived CCP module structures. The database of models was subsequently used in a multiple surface comparison of 28 CCP modules from complement receptor-type 1 (CR1, CD35) (Klickstein et al. 1987). Functionally critical surface residues are likely to be under different evolutionary pressures compared to non-critical residues. Despite highly conserved sequences, the two functional sites in CR1 display radically different surface features. This observation is entirely consistent with previous studies of CR1 that show distinct functional profiles for the two sites (Krych et al. 1991; Krych et al. 1994; Krych et al.

1998). Further analysis demonstrated that in this example, surface comparison readily highlighted functionally important sites.

3.2 Introduction to comparative modelling

Of all protein structure prediction methods, *i.e.* comparative modelling, fold recognition/threading and *ab-initio* modelling - comparative or homology-based modelling results in the most accurate models of protein structures (Sanchez and Sali 1998; Peitsch 2002). It is based on the principle that if a protein sequence shows significant sequence similarity to another protein of known three-dimensional structure due to homology, then an approximate 3D model of that protein can be obtained.

The steps involved in comparative protein structure modelling (Figure 3.2) are described in brief, and include:

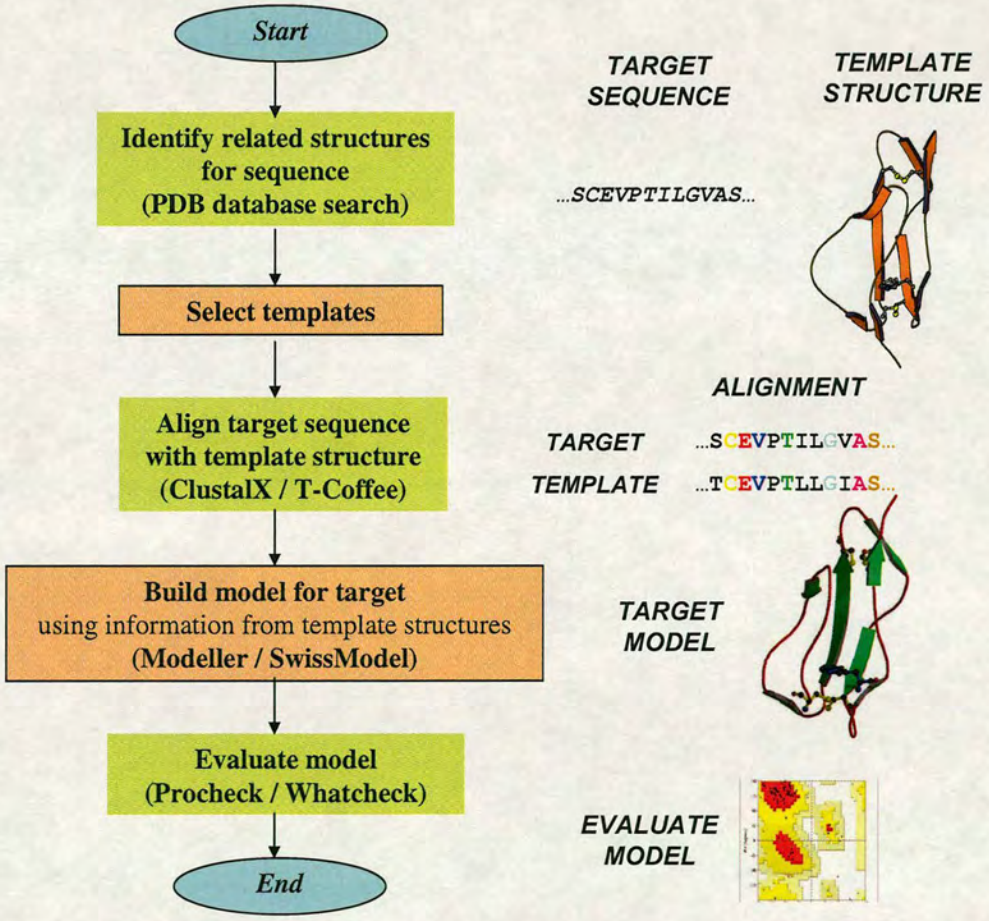


Figure 3.2: Flowchart depicting the various steps involved in comparative modelling.

(1) Identification of known structures (templates) related to the target sequence

(2) Alignment of the templates with the target sequence

(3) Building of a model based on the alignment

(4) Evaluation of the model

3.2.1 Template search and selection

A similarity search such as BLAST (Altschul et al. 1990) or FASTA (Pearson and Lipman 1988) is carried out against the Protein Data Bank (PDB) (Berman et al. 2000) to find a homologue of the query sequence (target) with at least one experimental structure (template). If more than one (different) homologous structure exists, it is not necessary to select only one template for modelling. In fact, the use of several templates approximately equidistant from the target sequence generally increases model accuracy (Srinivasan and Blundell 1993; Sanchez and Sali 1997). However, if more than one crystal structure exists for the same protein sequence; the highest resolution structure should be selected as template. Often a large family of proteins that includes target and templates can be organised into sub-families. In these cases construction of a multiple sequence alignment, the use of a clustering procedure, or creation of a phylogenetic tree can help in selecting the template from the sub-family that is closest to the target sequence (Corpet 1988; Retief 2000).

3.2.2 Target-template alignment

This is an important step in comparative modelling. Once the most suitable template is selected following the template search, maximal effort to obtain the most accurate alignment possible is needed because no comparative modelling method can recover

from an incorrect alignment (Marti-Renom et al. 2002). Various pair-wise alignment algorithms are used to calculate a similarity score on the basis of which to infer sequence homology between two sequences (*e.g.* target and template sequence). Examples include BLAST (based upon the Smith-Waterman algorithm) (Altschul et al. 1990), which searches for both full and partial sequence matches, *i.e.* local and global similarity; and FASTA (based upon the Needleman-Wunsch algorithm) (Pearson and Lipman 1988), which calculates a global similarity score between two sequences in the library. For aligning three or more sequences (multiple sequence alignments), ClustalW (Thompson et al. 1994) or ClustalX (Thompson et al. 1997) are the most widely used programs (Wallace et al. 2005). They are based on a heuristic referred to as 'progressive alignment' (Feng and Doolittle 1987). Progressive alignment allows large alignments of distantly related sequences to be constructed quickly and simply. It is based on building up a full alignment progressively, using branching order of a quick approximate tree (called the guide tree) to guide the alignments. This principle is also used as an optimiser for other programs, such as T-Coffee (Notredame et al. 2000) and ProbCons (Do et al. 2005), which are more accurate (Wallace et al. 2005) and gaining popularity.

All of these programs attempt to align sequences according to their evolutionary relatedness. For closely related protein sequences with sequence identity higher than 40%, the alignment is almost always correct. Regions of low local sequence similarity become common when the overall sequence identity is below 40% (Saqi et al. 1998). The alignment becomes difficult in the "twilight zone" of between 20-30% sequence identity (Rost 1999). As the sequence similarity decreases, alignments contain an increasingly large number of gaps, and alignment errors occur more frequently, especially when prepared automatically. In such cases,

careful manual editing of the alignment (*i.e.* human intervention) is required, to incorporate sequence-specific details such as local secondary and tertiary structure information, to place gaps more appropriately, and render the alignment more accurate.

3.2.3 Model building

Given an optimal target-template alignment, a variety of methods can be used to create a 3D model of a protein. The original and still widely used method is modelling by rigid-body assembly (Blundell et al. 1987; Greer 1990; Schwede et al. 2003), which assembles a model from a small number of rigid bodies obtained from aligned protein template structure. The assembly involves fitting the rigid bodies onto the framework (*i.e.* the core of the aligned regions) and rebuilding the non-conserved parts, *i.e.* loops and side-chains. Another method, modelling by segment matching, relies on the approximate positions of conserved atoms in templates (Claessens et al. 1989; Levitt 1992). It uses a subset of atomic positions, derived from the alignment as a guide to find matching segments in a representative database of all known protein structures. The database contains short segments of protein structures that are selected using energy or geometry rules, or a combination of these criteria. A third method involves modelling by satisfaction of spatial restraints (spatial restraints involve distances, angles, dihedral angles, and pairs of dihedral angles defined by atoms or pseudo atoms) obtained from the template, and its application to the target sequence in the alignment (Sali and Blundell 1993; Aszodi and Taylor 1996; Kolinski et al. 2001). Spatial restraints are derived from the alignment, and the model is obtained by minimising the violations to these restraints. Accuracies of the various

model building methods are comparable when used optimally (Marti-Renom et al. 2002).

Several programs and servers based on the above methods exist for model building (Wallner and Elofsson 2005). For example, SWISS-MODEL (http://swissmodel.expasy.org//SWISS-MODEL.html) (Guex and Peitsch 1997; Schwede et al. 2003) is a fully automated web server where a user can choose from two modelling modes: (1) A *First Approach mode*, where the user simply inputs a sequence and the server performs all the necessary steps automatically and returns a model via email; and (2) An *Optimise mode*, where the user can edit alignments under the Swiss-PdbViewer (Guex and Peitsch 1997) interface before sending out a modelling request to the server. Other modelling programs include Composer (Sutcliffe et al. 1987a; Sutcliffe et al. 1987b) available as part of the SYBYL (Tripos Associates, St. Louis, MO, USA) modelling environment; and Homology (Greer 1990) under the Insight II (Accelrys Inc., San Diego, CA, USA) modelling environment. Another widely used program called Modeller (Sali and Blundell 1993) constructs comparative models given a target-template alignment by satisfaction of spatial restraints. It is available as a stand-alone command line program, although a graphical interface to it does exist as part of the QUANTA (Accelrys Inc., San Diego, CA, USA) suite of programs.

3.2.4 Model evaluation

It is essential for assessing the value of 3D protein models to estimate their overall accuracy. In general, mistakes in comparative modelling include side-chain packing errors, distortions and rigid body shifts in correctly aligned regions, errors in inserted regions (loops), incorrect alignments, and inappropriate templates (Marti-Renom et al.

2000). Another form of inaccuracy could result from the deviation from ideal stereochemical values for bond lengths and angles. The sources of these problems are primarily based on alignment errors, and excessive molecular dynamics attempting to remedy these errors. Such inaccuracies can easily be detected with programs such as Procheck (Laskowski et al. 1993) and WHATCHECK (Hooft et al. 1996), which, like Procheck, runs a series of tests to evaluate model stereochemistry that detect deviations from normal bond lengths, dihedrals, etc.

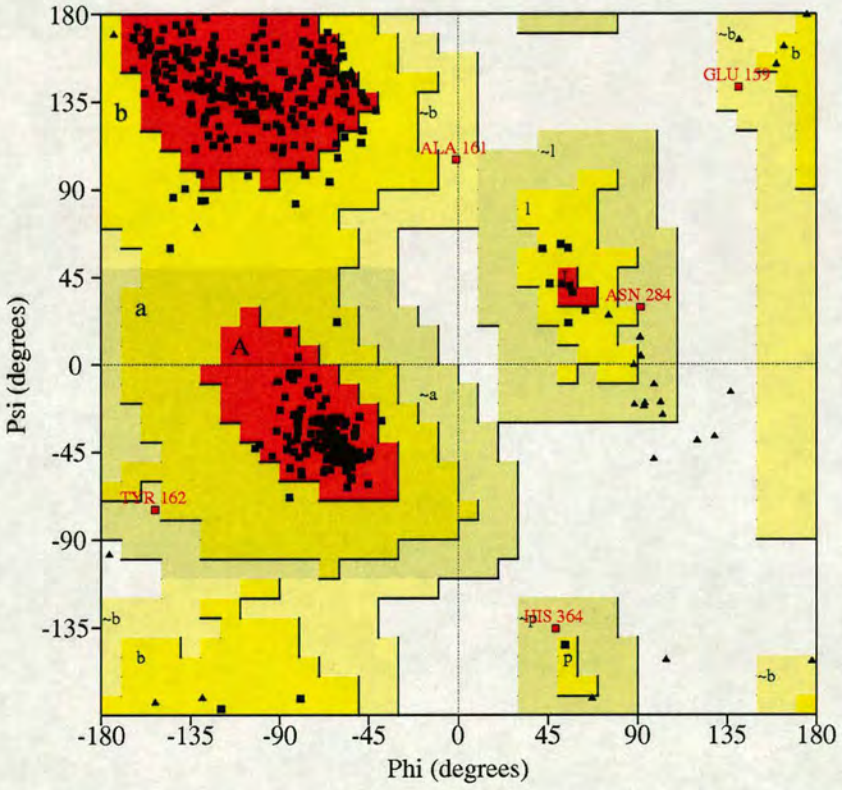


Figure 3.3: Example Ramachandran plot. The phi-psi torsion angles for all residues in the structure (except those at the chain termini) are shown as square dots. Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot that are applicable to the other sidechain types. The colouring/shading on the plot represents the different regions described in (Morris et al. 1992): the darkest areas (here shown in *red*) correspond to the "core" regions representing the "most favoured" combinations of phi-psi values; the darker yellow region correspond to the "additionally allowed regions", while the lighter yellow region correspond to the "generously allowed regions". The disallowed regions lie outside these regions (white). Symbols within the plot as described in Morris et al. 1992: A - Core α; a - Allowed α; ~a - Generous α; B - Core β; b - Allowed β; ~b - Generous β; L -

Core left-handed α; l - Allowed left-handed α; ~l - Generous left-handed α; p - Allowed ε; ~p - Generous ε.

A Ramachandran plot produced as part of a PROCHECK evaluation depicts the phi-psi torsion angles (in a polypeptide, the main chain N-Cα and Cα-C bonds are relatively free to rotate; these rotations are represented by the torsion angles phi and psi, respectively) for all residues in a structure (except those at the chain termini) (Figure 3.3), and thus detects inaccuracies that could arise from the deviation from ideal stereochemical values of bond lengths and bond angles. The different regions, *i.e.* most favoured, additional allowed, generously allowed and disallowed regions on the Ramachandran plot are described in Morris et al. (Morris et al. 1992). Ideally, one would hope to find greater than 90% of the residues for a particular model within the "most favoured or core regions" of the plot, and it serves as one of the better guides to the stereochemical quality of a protein structure (Laskowski et al. 1993).

3.3 Materials and Methods

3.3.1 Sequence-based clustering of CCP modules

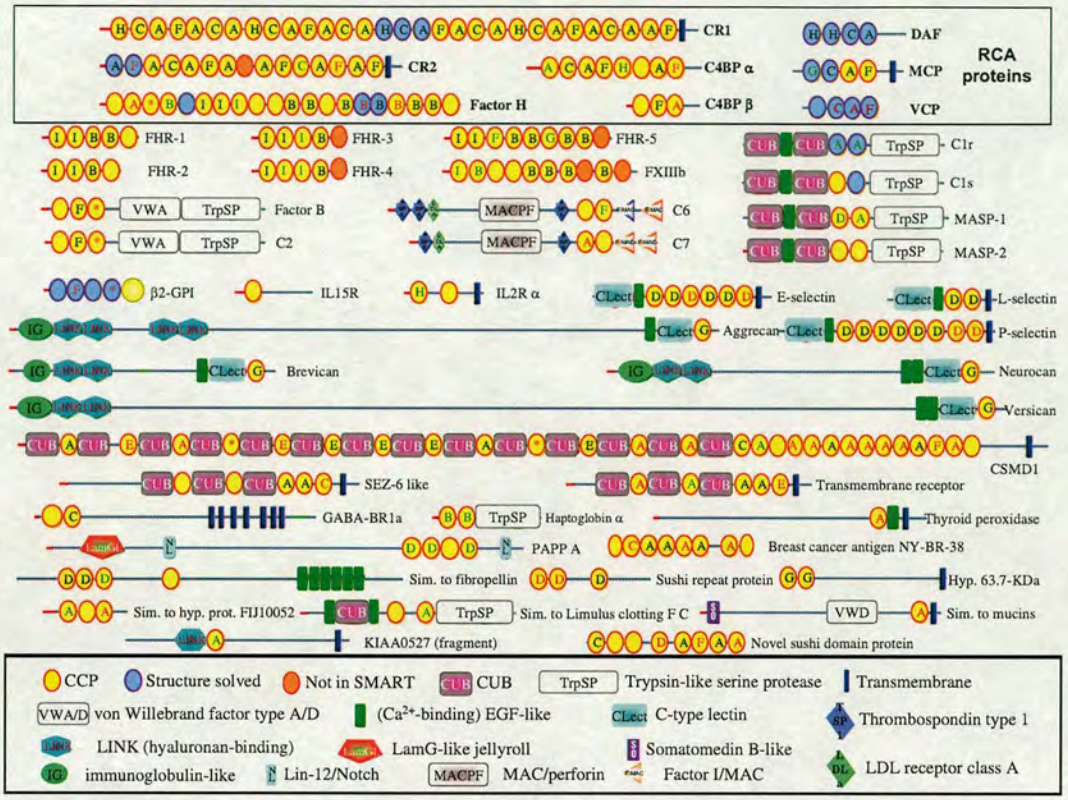


Figure 3.4: Cluster assignment of CCP modules. Cartoon displaying the molecular architectures of a representative set of human CCP-module containing proteins derived from the database SMART (http://smart.embl-heidelberg.de/). The 243 putative CCP-modules are depicted by yellow (blue when the module is of known structure, orange when it is known to be a CCP-module, but not detected by SMART) oval symbols. The letters within the CCP-module symbols denote the cluster to which the sequence was assigned. The colour of the text indicates the level of confidence in the assignment, decreasing from black (initial core assignment) to red (strong) to green (weak), as described in the "Materials and Methods" section. Refer to key for explanation of symbols. Ambiguously assigned modules are marked by red asterisks (A/E: CSMD1~4 and CSMD1~10; A/F: Factor B~3 and C2~3; B/D: Factor H~3; B/F: β2GPI~4). Abbreviations used in the figure: Sim. = similar; EGF = epidermal growth factor; CUB = first found in C1r, C1s, uEGF and bone morphogenic protein; LDL = low density lipoprotein; MAC = membrane attack complex; PAPP A = pregnancy-associated plasma protein-A; SEZ-6 like = Seizure 6-like; β2GPI = β-2-glycoprotein I (also called Apolipoprotein H or ApoH); CSMD1 = CUB and sushi domain-containing protein 1; IL2Rα = Interleukin-2 receptor α chain; IL15R = Interleukin-15 receptor α chain; GABA-BR1a = Gamma-aminobutyric acid type B receptor subunit 1 α. All other abbreviations explained in text.

Previous work (Kirkitadze and Barlow 2001) revealed that the residue before the first cysteine and two or three residues after the fourth and last cysteine of the consensus sequence commonly contribute to the 3D structure of a CCP module. Consequently, in the current study, one residue prior to the first consensus cysteine and three residues following the fourth consensus cysteine were included in the sequence selected to represent each CCP module.

Using this approach, sequences representing 243 CCP modules from 48 proteins [47 human proteins plus one from vaccinia virus (VCP)] were extracted from the SMART database (http://smart.embl-heidelberg.de/) (Schultz et al. 1998; Letunic et al. 2004). The vaccinia virus protein, VCP, with four CCP modules, was included along with the human sequences in this process because its structure has been experimentally determined (Wiles et al. 1997; Henderson et al. 2001; Murthy et al. 2001; Ganesh et al. 2004). It therefore provides valuable additional templates for modelling purposes. As a route to optimal template selection, the sequences were then classified by means of a clustering procedure. Initial cluster assignments were produced according to the "unweighted paired-group method with arithmetic mean" (UPGMA) using a program that implements Corpet's MULTALIN algorithm (Corpet 1988), concomitantly with rounds of multiple sequence alignment and subsequent manual removal of individual sequences that impeded convergence. Nine stable groups or clusters, (labelled *A* through *I*) were identified in this manner.

A hidden Markov model (HMM) was then built for each of these nine clusters using the software package HMMer version 2.0 (http://hmmer.wustl.edu/). HMMs are sensitive at detecting remote relationships among sequences and are hence widely used [see (Durbin 1998) for information on HMM-methodology]. All sequences (including the initial set) were scanned with the HMMs, and module sequences were

assigned to clusters using cut-off expectancy values (E-values) of 10^{-10} for strong, and 10^{-5} for weak assignments. Sequences that failed to be matched to an HMM were labelled as unassigned. Weak HMM-assignments were further investigated by pairwise similarity comparisons of the respective module sequences with the stably assigned set using MPSrch (Collins and Coulson 1990). They were accepted only if corroborated by predominance of the HMM-assignment in the MPSrch list of the five most similar module sequences. If sequences were matched to more than one HMM with E-value $< 10^{-10}$, then they were either: labelled as ambiguous if the E-value of the second match was within one order of magnitude; or assigned to the cluster with the higher score. Of 243 module sequences, 169 were strongly assigned by HMM, 30 were weakly assigned and six were ambiguously assigned, while 38 could not be assigned to a cluster at all (Figure 3.4).

Each cluster was subsequently aligned separately using ClustalX version 1.81 (Thompson et al. 1997). The automated alignments were edited manually placing strong emphasis on conserving the "Cys-Cys-Cys-Trp-Cys" signature pattern (Kirkitadze and Barlow 2001) and positioning alignment gaps plausibly considering the experimentally solved 3D structures within the cluster.

3.3.2 Selecting the most suitable program for CCP module modelling

A test comparison of the various model building programs mentioned earlier resulted in Modeller (Sali and Blundell 1993) being selected as the most appropriate for automated large-scale modelling of complement control protein modules. Modeller is freely available to academic users at: http://salilab.org/modeller/. It is fully command line driven and hence provides more flexibility than both, Composer (Sutcliffe et al. 1987a; Sutcliffe et al. 1987b) under the SYBYL environment (Tripos Associates, St.

Louis, MO, USA), and Homology under the Insight II modelling environment (Accelrys Inc., San Diego, CA, USA), which require visualisation graphical interfaces. A web-server such as SWISS-MODEL (Guex and Peitsch 1997; Schwede et al. 2003) is unsuitable for handling a large volume of requests since this could overload the server. Another major advantage of Modeller was that it allows for direct submission of target-template alignments for model building. Finally, it is ideal for CCP modelling because it automatically restrains disulfide bridges during model construction.

Model building using Modeller was performed by extracting distance and dihedral angle restraints on the target sequence from its alignment with the target structure. These template-derived restraints were combined with most of the CHARMM energy terms (Brooks et al. 1983) to obtain a full *objective function*. Finally, this function was optimised by conjugate gradients and molecular dynamics with simulated annealing to construct a model that satisfies all the spatial restraints as well as possible. Ten models were thus derived for each target sequence. The representative model for each sequence was that which had the lowest value of the *objective function* score.

3.3.3 Automated comparative modelling of each cluster

A sequence of programs to carry out the step-wise modelling task as described below were called from a PERL script as summarised in Figure 3.5. Eighty-three sequences out of the original set of 243 were not modelled because either they were not assigned to a cluster, or they belong to clusters *D*, *E*, *I* for which no templates are yet available. A detailed work-flow of the automation is provided below.

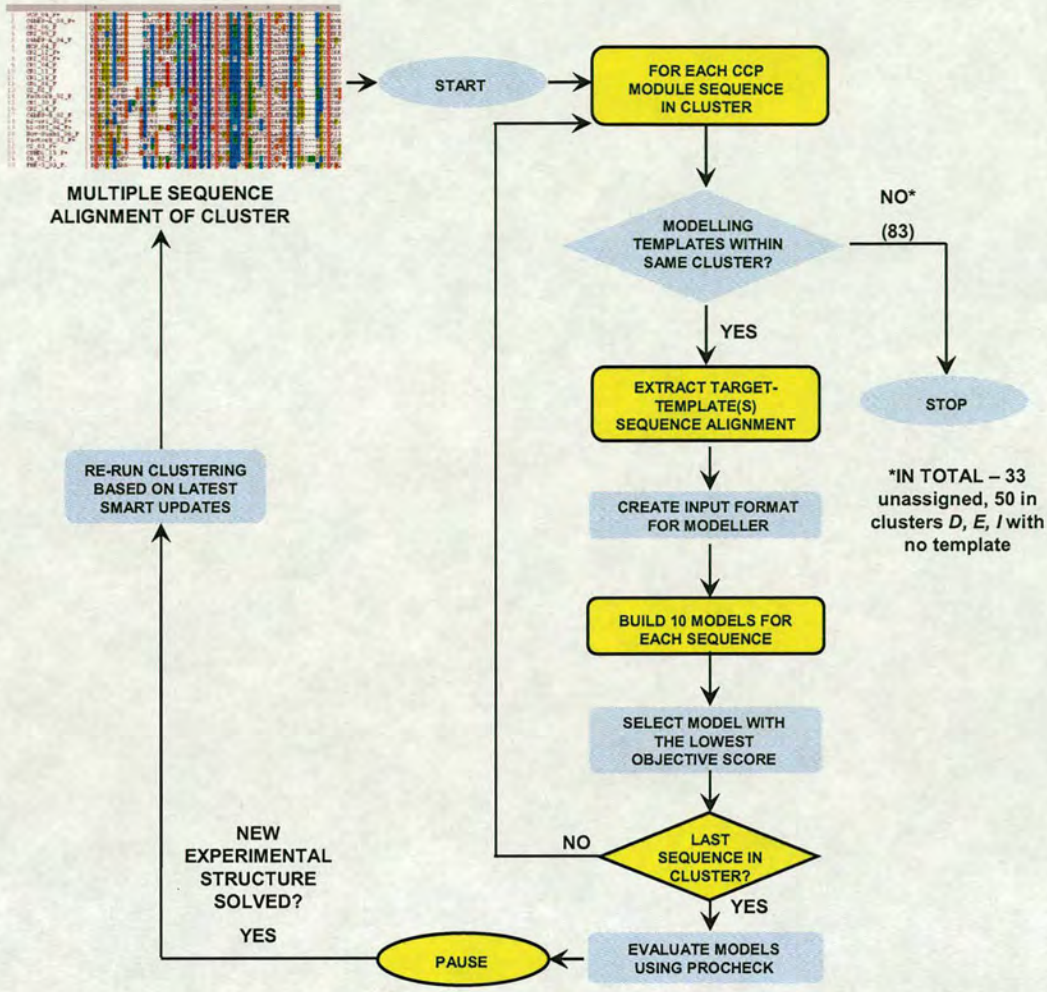


Figure 3.5: Work flow of automated method for each cluster. Flow-chart summarising the tasks implemented by a PERL script, which automatically calls the appropriate programs when presented with a multiple sequence alignment for a given cluster.

3.3.3.1 Identifying modelling templates within each cluster alignment

All sequence names were substituted by their unique Accession Numbers to ease the programming task. The control PERL script first employs an option within ClustalW version 1.83 (Thompson et al. 1994) to convert an entire cluster multiple sequence alignment from FASTA to PIR format, as required for Modeller version 6v2 (Sali and Blundell 1993). All target sequences are indicated by keyword "sequence" and template structures are flagged by keyword "structure" (Figure 3.6).

```
>P1;P02749_04_F+
structure:P02749_04_F+:.:.:.:.:.:.:.:.
KCPFPSRPDNGFVNYPAKPTLYYKDKATFGCHDGYSLDGPEEIECTKLGNWSA-MPSCKAS*

>P1;P17927_04_F
sequence:P17927_04_F:.:.:.:.:.:.:.:.
KCTPP-NVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKCQALNKWEPELPSCSRV*

>P1;P17927_11_F
sequence:P17927_11_F:.:.:.:.:.:.:.:.
KCTPP-NVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKCQALNKWEPELPSCSRV*

>P1;P17927_18_F
sequence:P17927_18_F:.:.:.:.:.:.:.:.
KCTPP-NVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKCQALNKWEPELPSCSRV*

>P1;P17927_25_F
sequence:P17927_25_F:.:.:.:.:.:.:.:.
KCTAP-EVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTVQCQTNGRWGPKLPHCSRV*
```

Figure 3.6: Example of a section of cluster *F* alignment in Modeller-PIR file format. For each sequence in a cluster alignment file, the first line is composed of a unique code containing the SwissProt/TrEMBL (Boeckmann et al. 2003) accession number followed by the module number and then the cluster, in the format sequence ">P1;code" (*e.g.* >P1;P17927_25_F). The second line with ten fields separated by colons generally contains information about the sequence or structure file. Only the first two fields are mandatory indicating whether the file contains a "sequence" or "structure" and providing Modeller with their unique codes. The remaining fields are left blank indicated by a series of ":". The third line contains the respective sequence, with "*" marking its end. All gaps in the alignment are depicted by "-".

3.3.3.2 Extracting target-template alignment

For each Modeller-PIR format cluster alignment, each sequence is considered individually, and the cluster screened for sequences corresponding to modules of known structures, to serve as templates for modelling. If no template is found, no model is created. On the other hand, if one or more template sequences do exist (flagged by keyword "structure"), the script extracts the target sequence along with the template sequence(s) directly from the alignment (Figure 3.7).

P17927_25_F.ali

```
>P1;P02749_04_F+
structure:P02749_04_F+:.:.:.:.:.:.:.:.
```

```
KCPFPSRPDNGFVNYPAKPTLYYKDKATFGCHDGYSLDGPEEIECTKLGNWSA-MPSCKAS*

>P1;P17927_25_F
sequence:P17927_25_F:.:.:.:.:.:.:.:.
KCTAP-EVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTVQCQTNGRWGPKLPHCSRV*
```

Figure 3.7: Example intermediary output for each sequence extracted depicting a typical target-template alignment. The example above displays the alignment for the target sequence P17927_25_F with its template structure P02749_04_F+, saved in a unique file name P17927_25_F.ali

3.3.3.3 Preparing input files for Modeller, model building and evaluation

The script then prepares a standard input format file called a 'top' file for input to Modeller (Figure 3.8). Once a target-template alignment is provided, Modeller calculates a 3D model of the target automatically. The "P17927_25_F.top" routine below generates ten models of target sequence "P17927_25_F" based on the "P02749_04_F+" template structure and the alignment in the file "P17927_25_F.ali".

P17927_25_F.top

```
# Homology modelling by the Modeller6v2 TOP routine 'model'
INCLUDE                                 # Include the predefined TOP routines
SET OUTPUT_CONTROL = 1 1 1 1 1          # uncomment to produce a large log file
SET ALNFILE = 'P17927_25_F.ali'         # Alignment file
SET KNOWNS = 'P02749_04_F+'             # Template structure
SET SEQUENCE = 'P17927_25_F'            # Target sequence
SET ATOM_FILES_DIRECTORY = '/CCP-structures/' # Template PDB file location

SET STARTING_MODEL = 1                  # index of the first model
SET ENDING_MODEL = 10                   # index of the last model (create 10
models)
CALL ROUTINE = 'model'                  # build homology model
```

Figure 3.8: Example of a Modeller TOP routine (P17927_25_F.top). The first line includes Modeller variable and routine definitions. The following lines set parameter values for the "model" routine. ALNFILE names the file that contains the target-template alignment in Modeller-PIR format. KNOWNS define the known template structure(s) in the alignment file. SEQUENCE defines the name of the target sequence in ALNFILE. STARTING_MODEL and ENDING_MODEL define the number of models that are calculated (their indices run from 1 to 10). The last line in the file calls the "model" ROUTINE that actually calculates the model. (# - Prefixed comments explaining each step follow the commands).

Once the ten models for each sequence were generated, the model with the lowest value of the Modeller "*objective function*" (which is reported in the second line of the PDB file) is selected. The general form of the *objective function* and the structure of optimisation are similar to molecular dynamics programs, such as CHARMM (Brooks et al. 1983; Sali and Blundell 1993).

The model with the lowest value of the Modeller "objective function" is then subjected to stereochemical evaluation using the program Procheck version 3.5.4 (Laskowski et al. 1993). For the RCA CCP-modules, out of the 63 modelled structures, 61 had more than 90% of residues in combined "core" and "allowed" regions of the Ramachandran plot, and the other two had more than 85% of residues in these regions.

It should be noted that the PERL script has been implemented to prompt the user to select between the option of directly modelling all sequences in a given cluster alignment, or to model protein sequences individually based on their unique accession number.

3.3.4 Surface electrostatic analysis

(a)

(b)

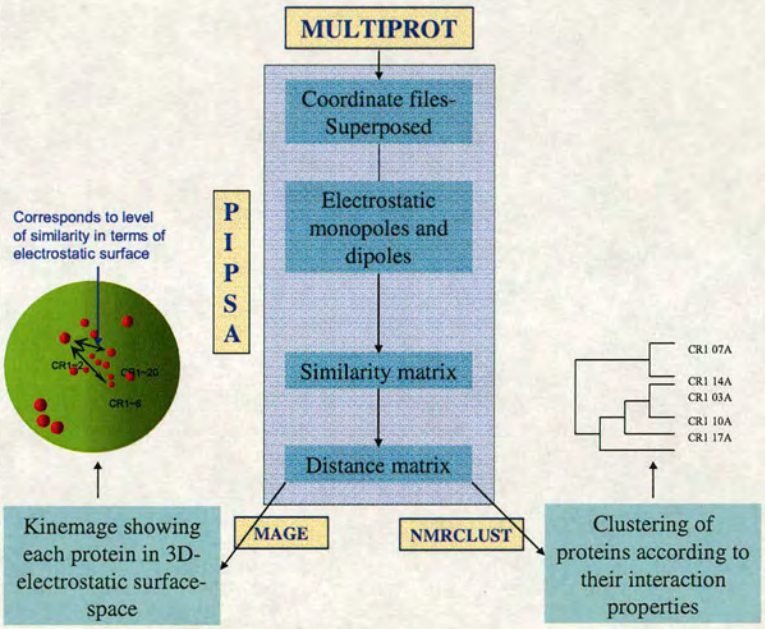


Figure 3.9: Scheme for comparing electrostatics of CR1. **(a)** CR1 superposition obtained using Multiprot (Shatsky et al. 2002; 2004) shown as dashed Cα trace. **(b)** The dissimilarity matrix produced by PIPSA is submitted to NMRClust for clustering, or alternatively can be viewed using program Mage (Richardson and Richardson 1992), to see each protein in 3D-electrostatic surface space.

The CCP modules of CR1 were selected to illustrate the utility of a surface-comparison approach because they have been the subject of extensive functional and

mutagenesis studies in the past (Krych et al. 1991; Krych et al. 1994; Krych et al. 1998; Krych-Goldberg et al. 1999). The set of experimentally determined and modelled 3D structures representing the N-terminal 28 CR1 CCP modules (out of 30 in total) was pulled from the database, and subjected to a comparative analysis of electrostatic properties using a combination of the programs; PIPSA version 1.0 (Blomberg et al. 1999), NMRClust version 1.2 (Kelley et al. 1996) and GRASP (Nicholls et al. 1991).

A structural alignment was obtained using Multiprot (Shatsky et al. 2002; 2004) to ensure that all module orientations are preserved such that surfaces at equivalent locations could be compared (Figure 3.9a). The aligned structures were then analysed using PIPSA, to find similarities within this set of CCP modules based on their surface electrostatic properties. PIPSA computes the molecular potentials of the model surfaces analytically as a multipole expansion that permits comparison of large datasets (Blomberg et al. 1999). The modules were then clustered by submitting the dissimilarity matrix generated by PIPSA to NMRClust. Surface electrostatic property-based cluster diagrams were derived manually by successive joining of closest neighbours based on similarity threshold values provided to NMRClust (Figure 3.9b). To help assess the validity of the clusters derived in this way, electrostatic surface images were also generated by GRASP, and inspected and grouped manually.

3.4 Results and Discussion

3.4.1 Clustering helps to optimise the choice of templates for modelling

A reliable procedure for cluster assignment enhances the value of the large-scale modelling procedure used in this study because it ensures that the most appropriate set of templates will be employed in each case. Using an implementation of the hierarchical cluster assignment method of Corpet (Corpet 1988), which was extended through subsequent sequence comparisons using hidden Markov models and exhaustive similarity comparisons, a total of 205 out of the original set of 243 CCP module sequences were each assigned to one of nine clusters (labelled *A* through *I*). Standard phylogenetic methods have drawbacks when applied to this family as alternative approaches to clustering. This is often observed with shorter protein sequences (Rokas et al. 2003). The main problem is that estimates of evolutionary distances are at lower levels of precision than with longer sequences. This "low signal-to-noise ratio problem" is addressed more satisfactorily by a protocol such as the one we applied here to produce the initial set of clusters, since only stable assignments are reported. The additional HMM-based assignments (shown in Figure 3.4) are more tentative, but all of them are corroborated independently by sequence comparisons.

The clusters to which the modules of the RCA proteins are assigned are shown in Figure 3.4. In such a representation, the four heptad or "long homologous repeats" (LHRs) – *HCAFACA* – that comprise the N-terminal 28 modules of CR1 are readily apparent. The LHRs of CR1 were recognised previously (Klickstein et al. 1987; Hourcade et al. 1988; Klickstein et al. 1988) and are thought to have an evolutionary origin involving exon duplication and shuffling (Hourcade et al. 1990). The triad

CAF additionally appears twice in CR2 and once each in C4BP and MCP for a total of eight occurrences. Also apparent in Figure 3.4 are imperfect tetrad repeats in CR2: from the N-terminus – *AFAC*, *AFA*X, *AF*↑*C*, and *AFAF* [X = unassigned, ↑ is the site of an inserted module in the 16-module splice variants (Moore et al. 1987; Barel et al. 1998), which in independent sequence comparisons using MPSrch, displays strong similarity with other *A*-cluster members]. With the exception of factor H and factor H related proteins, CCP modules from just a few clusters - *A*, *F*, *C* and *H* - account for all but one of the assigned modules of the RCA family (MCP has a *G*-cluster member at its N-terminus).

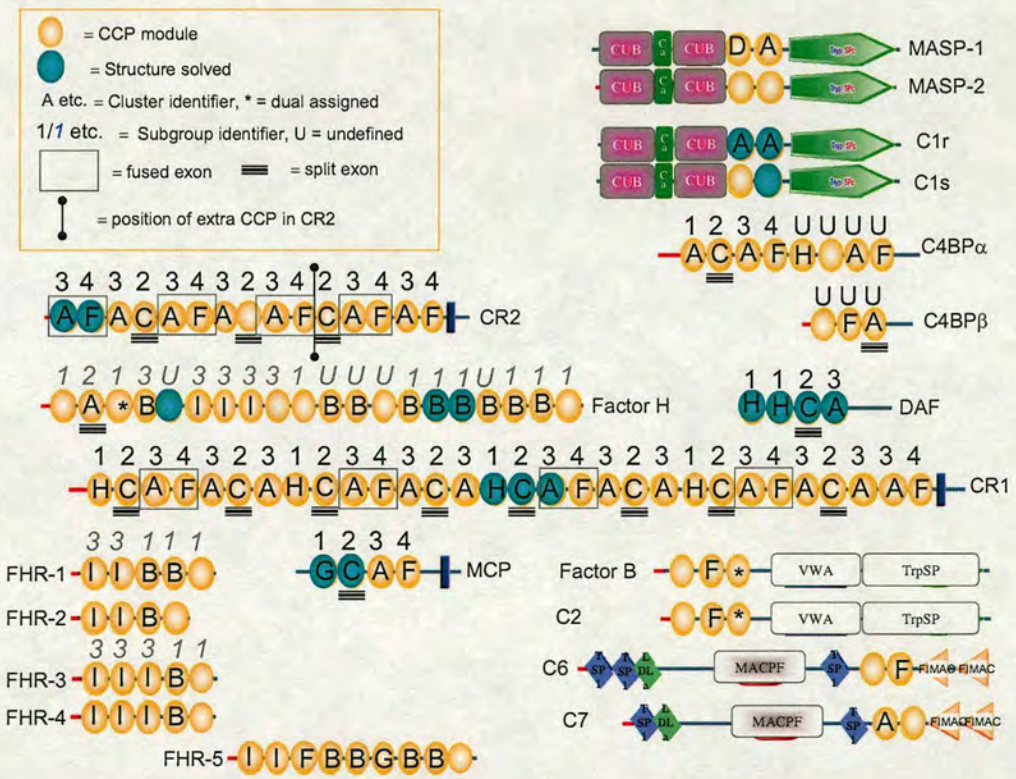


Figure 3.10: Comparison of assignment of the CCP modules of complement proteins into clusters or groups. The symbols used for module-types are the same as in figure 3.4 for a set of human complement proteins. The modules for which experimentally solved 3D structures are available are shaded more strongly (see key). The number above each module indicates assignment to sub-groups of CCPs according to a phylogenetic study by Krushkal *et al.* (Krushkal et al. 2000) (as in the chapter 2) and Soares *et al.* (Soares et al. 2005) clusters are indicated by letters within the CCP symbols. Those

modules that could not be classified by these methods are indicated by a "U" (Krushkal *et al.*) or by an absence of a letter (Soares *et al.*).

Interestingly, the majority of N-terminal CCP modules from amongst the RCA proteins either remained unassigned (fH, VCP and C4BPβ), or were only weakly assigned (MCP and C4BPα) to a cluster. This might imply varying rates of evolution and, intriguingly coincides with a requirement of these modules for functional viability of the respective proteins. Moreover, in the first, second and third LHRs of CR1, and in DAF, C4BPα and MCP, a triad of modules comprise a binding site for components within the convertase enzymes of complement, and facilitates the dissociation or destruction of these proteolytic complexes (Medof et al. 1982; Krych et al. 1991; Brodbeck et al. 1996; Hardig et al. 1997; Liszewski et al. 2000; Blom et al. 2001a; Kuttner-Kondo et al. 2001). The cluster assignments for these functional units all seem to follow a motif X*CA* (where X = *A*, *G*, *H* or unassigned).

The unit X*CA* is not, however, seen in fH even though this protein regulates the convertases in a similar way to the other RCAs, and its N-terminal three CCP modules are critical for this function (Jokiranta et al. 2000). Factor H contains six modules not assigned to clusters, consecutive runs of six *B*-cluster members and three modules of cluster *I*. Members of these clusters are unique to fH and the fH-related proteins amongst the RCA proteins, but blood-clotting factor XIIIb also contains *B*- and *I*-cluster members. These proteins are closely linked and located within 650 kb-2.2 Mb in chromosome 1 (Rey-Campos et al. 1990; Skerka et al. 1995). This observation of a clear distinction between fH and the other RCAs is consistent with one made by Krushkal *et al.* who constructed a phylogenetic tree for 132 individual module sequences of the RCA gene cluster (Figure 3.10). Factor H is thought to have

diverged from CR1, CR2, MCP and DAF at an early point in evolutionary history (Krushkal et al. 2000) (Chapter 2).

3.4.2 Modelling

Three-dimensional (3D) structure modelling was accomplished for 135 individual human CCP module sequences, in each case based on the most similar homologues or set of homologues for which experimentally determined structures were available. At the start of this procedure (outlined schematically in Figure 3.5), each of the nine sequence clusters was aligned separately. The multiple alignments were subsequently used to guide the automated modelling of individual CCP modules of unknown structure using the program Modeller (Sali and Blundell 1993). When the large-scale modelling process was first run, a total of 16 experimentally solved 3D CCP module structures were available as templates. The experimentally determined 3D structures of the four DAF CCP modules were unavailable at the time, and therefore not included. Subsequently, these served as a very useful means of validating the structural models (see Section 3.4.3). A further five experimentally determined module structures - fH~5, β2GPI~1, β2GPI~3, C1s~2 and VCP~1 - were not assigned to any of the nine clusters and therefore not used as templates.

The total of 135 modelled structures resulting from the first run included those of 63 RCA CCP modules (27 from CR1, nine from factor H, four from DAF, two from MCP, nine from C4BPα and β, 12 from CR2). All models are available at the CCP-module model database http://www.bru.ed.ac.uk/~dinesh/ccp-db.html (snapshot shown below, Figure 3.11). This database is updated as necessary, to reflect the

growing number of available CCP module sequences, and as more template structures are deposited.

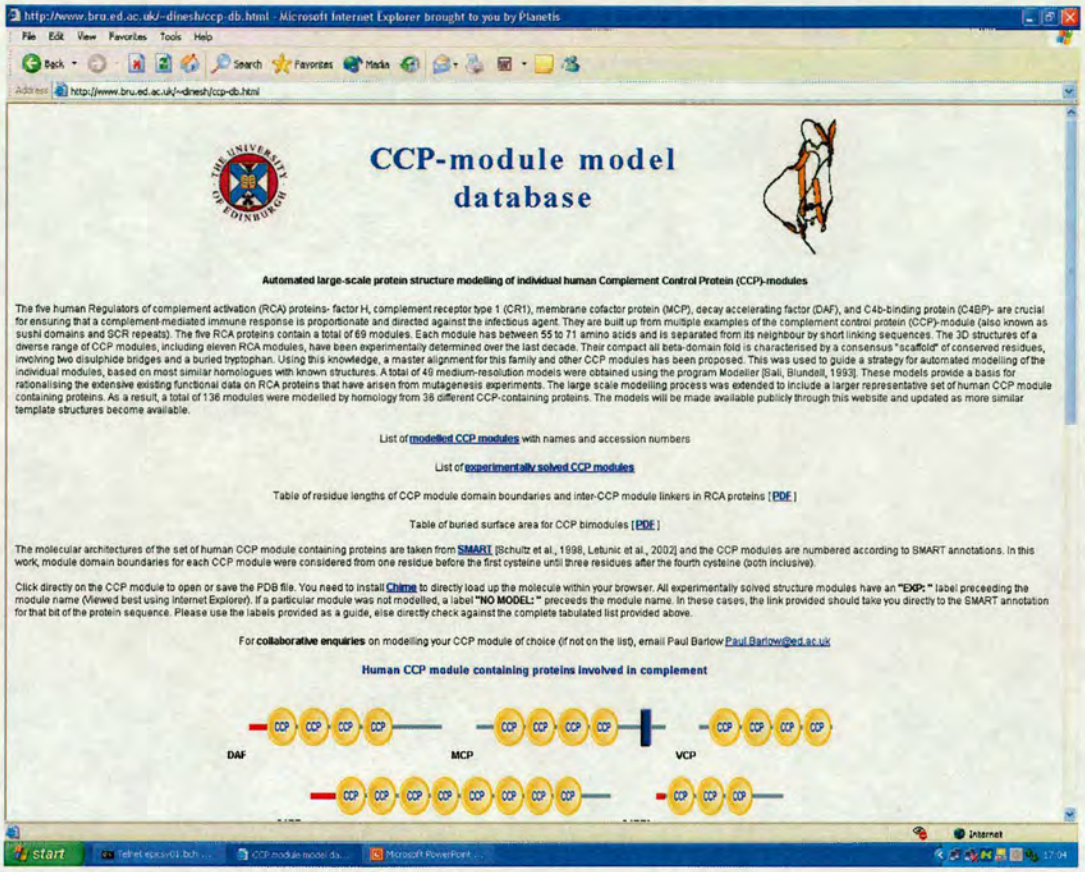
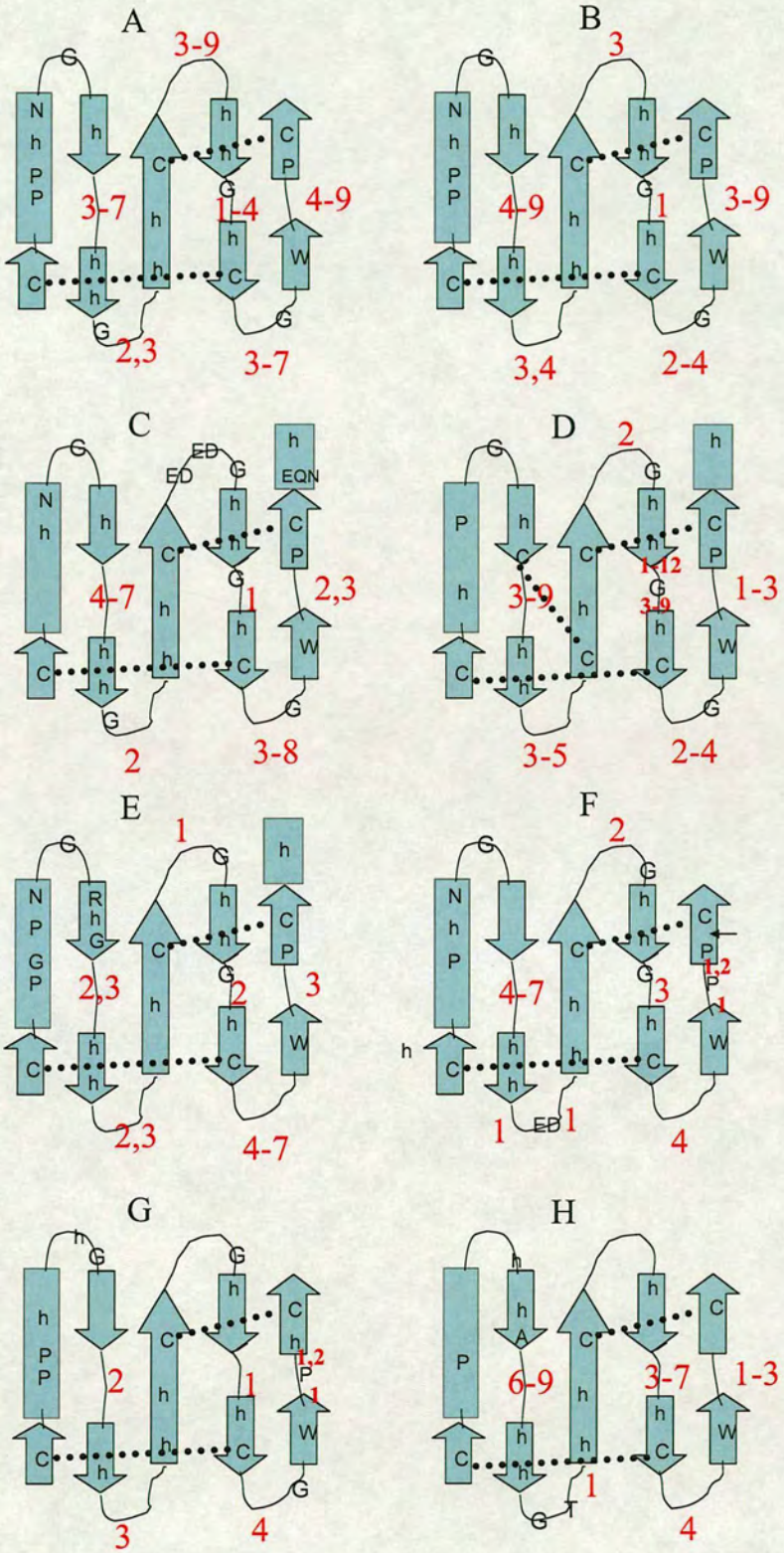


Figure 3.11: Snapshot of the CCP-module model database website. The database is accessible at http://www.bru.ed.ac.uk/~dinesh/ccp-db.html. Models are downloadable (in PDB format) at the click of a mouse-button. The website also serves as a comprehensive CCP module information resource. Since its inception, the website has received a total of 4,291 hits, including 2,460 unique hits (different IP-addresses) worldwide. The website statistics are provided by the CQ-counter (http://www.cqcounter.com/) as of May, 2006.

Fifty sequences, assigned to clusters *D*, *E* and *I*, were not modelled since no template structures were available at the time. While models for these modules could be obtained by modelling based on sequence similarity with the closest template identified using programs such as BLAST (Altschul et al. 1997) or MPSrch, they would be expected to be less reliable. It was decided to refrain from such attempts in

this automated exercise because *I*-cluster sequences are noticeably different, for example, compared with *B*- and *A*-cluster sequences (Figure 3.12), which are hard to distinguish by visual inspection, in that they appear to lack the conserved or conservatively replaced hydrophobic residues in predicted β-strands 2 and 3 (Figure 3.12). Similarly, cluster *E* members are characterised by short hypervariable loops and a lack of insertions between β-strands 5 and 6, or 7 and 8- using the Wiles *et al.*, convention for numbering the strands (Wiles et al. 1997) and *D*-cluster members have an extra pair of cysteines, allowing an additional putative disulfide bond [modelled in (Norman et al. 1991; Weyer et al. 2004), and in Chapter 4].


A
3-9
3-7
1-4
4-9
2,3
3-7
B
3
4-9
1
3-9
3,4
2-4
C
ED
EQN
4-7
1
2,3
2
3-8
D
2
1-12
3-9
3-9
1-3
3-5
2-4
E
1
2,3
2
3
2,3
4-7
F
2
4-7
3
1-2
1
1
ED
1
4
G
2
1
1-2
1
3
4
H
6-9
3-7
1-3
1
4

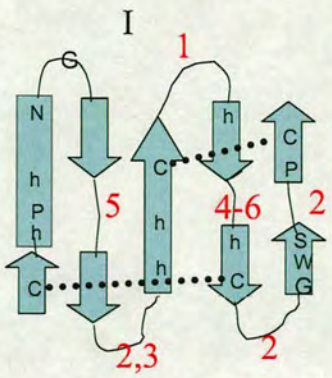


Figure 3.12: Signature topology diagram of clusters. The topology schematic depicts extent of consensus secondary structure and conserved or conservatively replaced residues for each cluster alignment (*A* through to *I*). Full cluster alignments are provided in Appendix. β-strands are depicted by arrows, and other consensus residues out-with the strands shown within cylinder representations on the basis of the structure of factor H module 5. Consensus residues are labelled by their single-letter amino acid codes, and h = hydrophobic residue. Disulfide bonds are indicated by dotted lines between cysteine residues. The lengths of loops are indicated numerically.

The models generated in the current work may be compared with those reported previously from the other RCAs – namely, C4BPα (Villoutreix et al. 1998), MCP (Liszewski et al. 2000), and fH (Aslam and Perkins 2001; Perkins and Goodship 2002) - for which no experimental structures were available at the time. In the case of C4BPα, the r.m.s.d. values over Cα atoms range from 2.6 Å (C4BPα~5) to 3.2 Å (C4BPα~4); for MCP~3 and MCP~4 the values are 2.2 Å and 2.7 Å, respectively; for fH they range from 1.6 Å (fH~12) to 3.4 Å (fH~2). The new models are thus significantly different from previously published models. But how do they compare to experimentally determined structures?

3.4.3 Validation of models against known structures – Part I

The coordinates for the experimentally determined structures (Uhrinova et al. 2003; Williams et al. 2003; Lukacik et al. 2004) of the DAF CCP modules became available

after the first run of the modelling procedure described above had been completed. This allowed a comparison of modelled and experimental structures. A high level of structural similarity - r.m.s.d.s (Cα) of 1.7 Å, 2 Å, 1.2 Å and 1.9 Å for DAF modules 1 to 4, respectively - was observed, supporting the modelling strategy used in this work (Figure 3.13a). Previously reported models of DAF, built when only fH~15 and fH~16 were available as templates (Kuttner-Kondo et al. 1996), exhibit a lower level of similarity with the crystal structure – r.m.s.d.s (Cα) of 2.1 Å, 2.6 Å, 3.3 Å and 2.5 Å for DAF modules 1 to 4, respectively. Models of DAF~2 and DAF~3 that were built (Kuttner-Kondo et al. 2001) after MCP~1 and MCP~2 also became available as templates, display r.m.s.d.s (Cα) of 2 Å and 1.6 Å compared to the equivalent modules of the crystal structure, *i.e.* were no more accurate than those obtained using the automated cluster-based modelling procedure described here. These data suggest that the current set of modelled structures are generally likely to be in the same range of accuracy, or more accurate, than previous ones (also see Section 3.5.1.2). In addition, the advantages of an automated modelling procedure for straightforward updating of the database are evident. The utility of the database has already been put to use as described in Chapter 4.

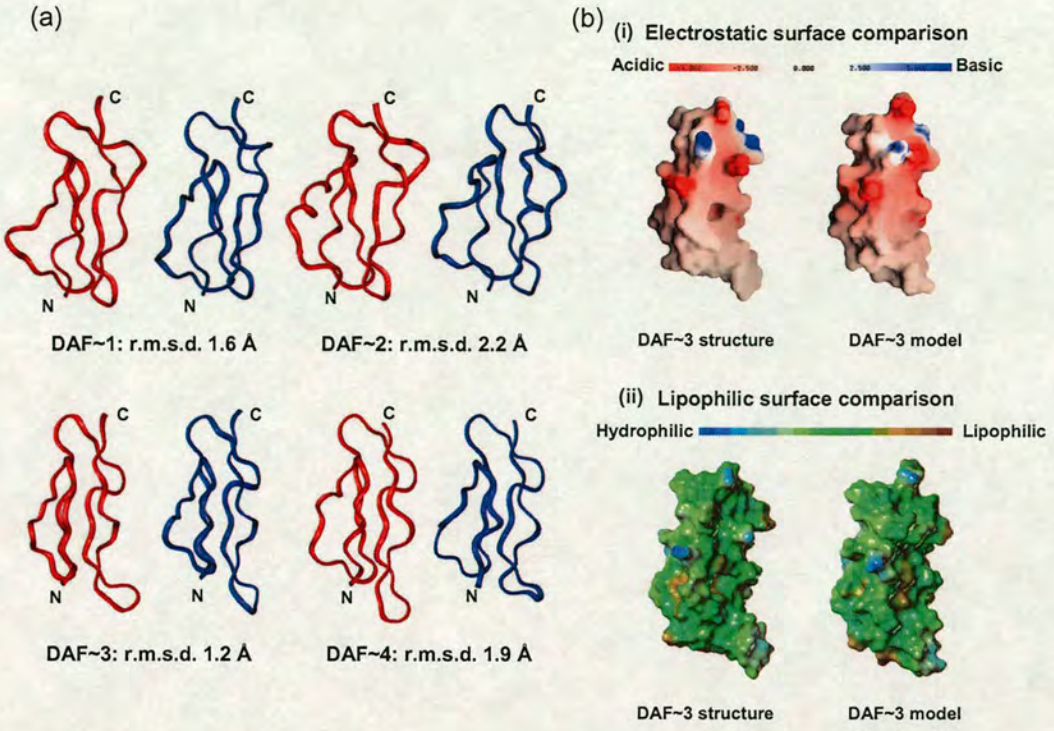


Figure 3.13: DAF~1,2,3,4 test comparison. (a): Ribbon trace of the modelled module structures of DAF~1-4 (blue) created by the automated modelling procedure, compared with corresponding modules from the highest resolved crystal structures (red) (PDB ID: 1OK3 for modules 1 and 2; PDB ID: 1H03 for modules 3 and 4). R.m.s.d. values (Cα atoms) are indicated in each case. **(b):** Comparison of the modelled (on the right) versus the experimental structure of DAF~3 showing (i) electrostatic surfaces (negative charge coloured red, positive charge coloured blue, ranging from –5 kT to +5 kT) generated with GRASP (Nicholls et al. 1991) and (ii) lipophilic surfaces (ranging from hydrophilic coloured blue, to lipophilic or hydrophobic coloured brown) generated using MOLCAD (Heiden et al. 1993).

The feasibility of using the new models in surface comparisons was examined by a comparison of electrostatic and lipophilic surfaces of modelled versus empirical CCP module structures. The two structures of DAF~3, for example, share very similar surface properties (Figure 3.13b). [The DAF~3 model was built using three templates - VCP~2 (Murthy et al. 2001), CR1~16 (Smith et al. 2002) and MCP~2 (Casasnovas et al. 1999) - with pairwise sequence identity percentages of 45%, 38% and 34%, respectively.] The surface representation of our model is also visibly in better agreement with the experimentally determined structure than a previously

published model of DAF~3 that was based on a template from a different cluster (Kuttner-Kondo et al. 1996). Thus this example illustrates that careful template-choice could benefit subsequent application of the models involving surface-comparison. While one cannot be certain that all models generated by this procedure will be as accurate as this verifiable example, the strategy to select the best template, and to update the CCP model database when new templates become available, aims to ensure that the best possible surface representation is available to the user at any given time.

3.4.4 Surface electrostatic analysis of CR1

The program PIPSA (Blomberg et al. 1999) was employed to compare electrostatic properties of the N-terminal 28 CR1 modules (25 modelled structures plus three experimentally determined ones). The PIPSA-generated cluster diagram (Figure 3.14) was compared to the manually grouped GRASP-generated electrostatic surface-images of the CR1 modules. Reassuringly, most module associations made by PIPSA showed good agreement with the groups of GRASP images. This implies that PIPSA could, in future work, be applied to a larger set of CCP modules and used to compare module surfaces between proteins.

The majority of relationships in the PIPSA-generated cluster diagram based on electrostatic surfaces (Figure 3.14) match up both with a sequence-based tree (Figure 3.15), and the cluster-assignments (Figure 3.4); *e.g.* CCP modules 4, 11, 18 and 25 (all in cluster *F*) are in the same surface-cluster. However, PIPSA also suggested associations based on electrostatic surfaces that are different to those based on sequence alone.

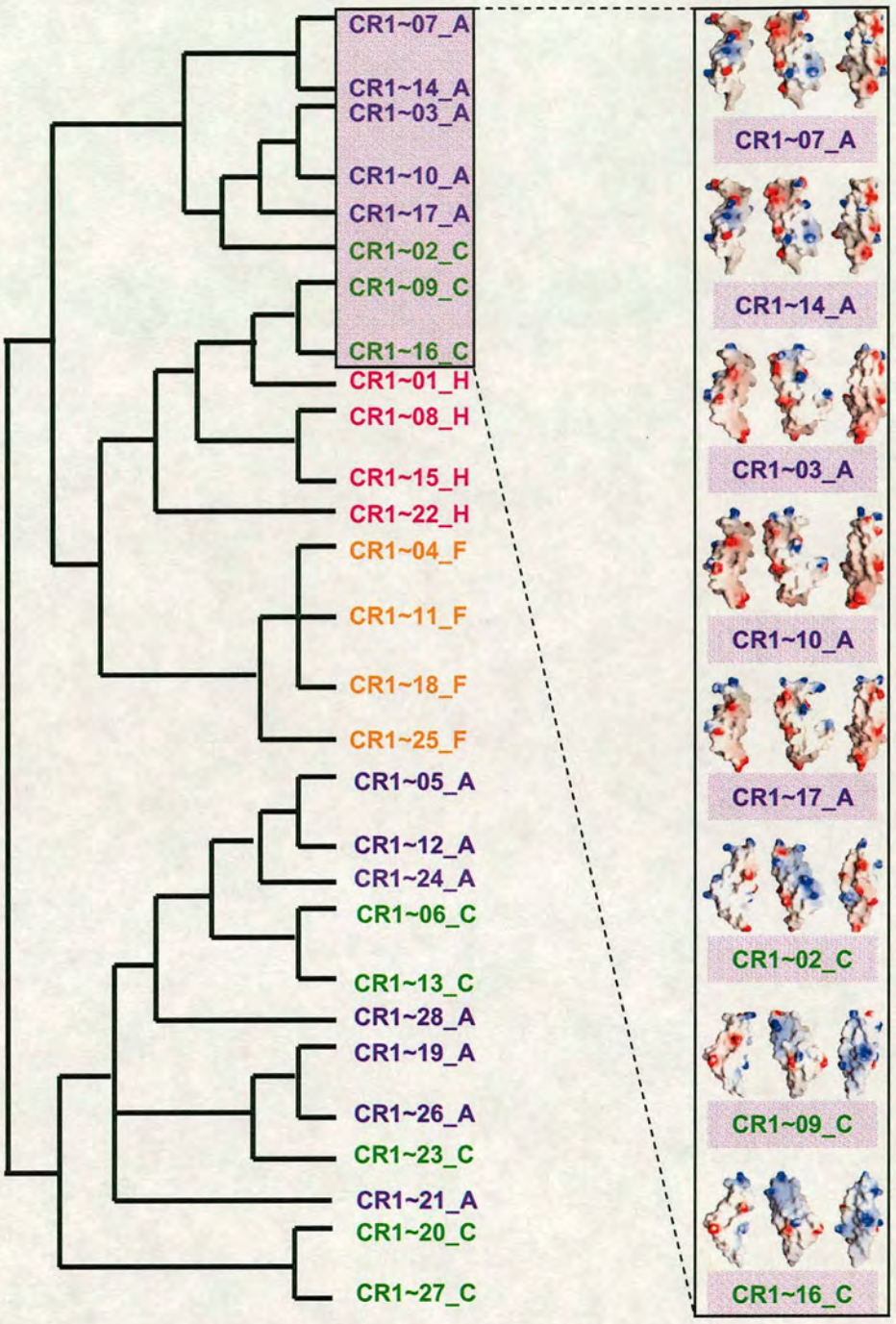


Figure 3.14: Electrostatic surface clustering of CR1 modules 1-28. Surface cluster-diagram depicting electrostatic similarities amongst the N-terminal 28 modules of CR1 according to PIPSA and NMRClust. Labels are coloured according to sequence cluster. GRASP-derived images of electrostatic surfaces for selected modules are also shown (box) to validate the clustering. In each case, three views rotated by 120° around the *y*-axis are provided.

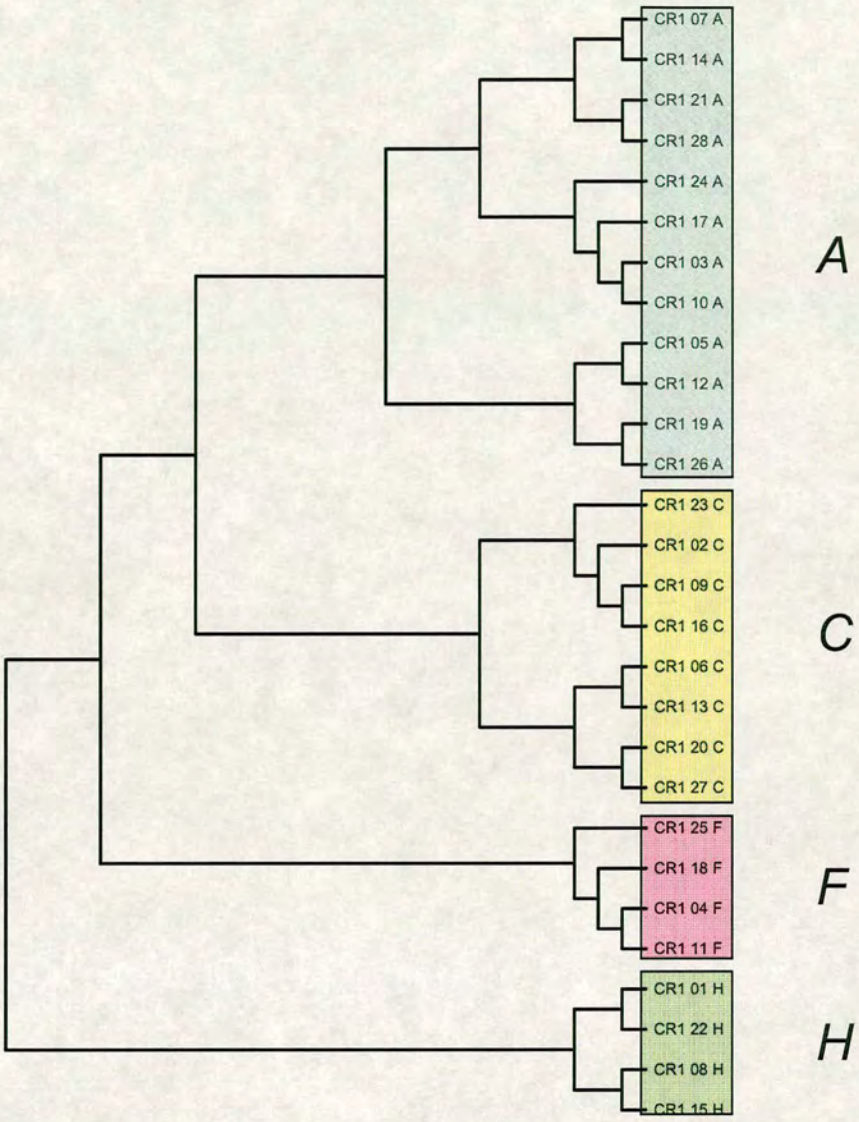


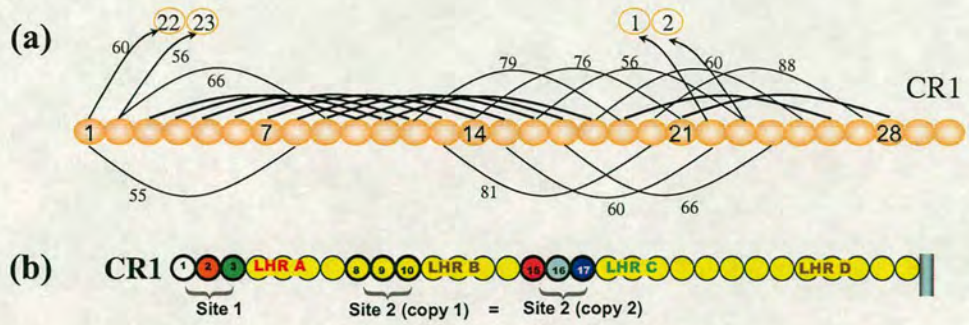
Figure 3.15: Sequence-based CR1 phylogenetic tree. The tree confirms the validity of the clustering, by placing all cluster members in their respective groups/clusters (*A, C, F, H*).

3.4.5 Surface comparisons can identify modules important for function

According to the sequence-based clustering, CR1~2 is most similar to CR1~9, CR1~16 and CR1~23; with all four modules belonging to cluster *C*. For example, CR1~2 has approximately 67% pairwise sequence identity with both CR1~9 and CR1~16 (Figure 3.16a). In the cluster diagram based on electrostatic surface-properties, however, CR1~2 is instead grouped with CR1~3, CR1~10 and CR1~17,

which were all assigned to sequence cluster *A* (Figure 3.14). Note that no significant sequence similarity was found through a pairwise BLAST comparison of CR1~2 and CR1~3. From inspection of Figure 3.14, it is evident that the electrostatic surface of CR1~2 does indeed resemble more closely that of CR1~3 than it does those of CR1~9 or CR1~16 (see below). Thus a comparison of relationships based on surfaces versus those based on sequences readily identified a case where two modules (CR1~2 and CR1~9) had clearly diverged significantly in terms of their surface properties, but not in overall sequence. Given the close evolutionary relatedness of these modules as judged by their equivalent positions within the four LHRs in CR1 (Hourcade et al. 1990) (Figure 3.4, Figure 3.16), this observation implies that some surface-exposed residues have been the subject of adaptive mutation while the remainder of the sequence has undergone predominantly neutral variation. Therefore the modules affected are candidates for being functionally important ones, with their non-conserved residues representing possible sites of interaction with binding partners.

This notion is borne out by previously reported (Klickstein et al. 1988; Krych et al. 1991; Krych et al. 1994) functional studies showing that CR1~2 and CR1~9 (and the nearly identical CR1~16) contribute to two different multimodular binding sites - called functional sites 1 and 2 respectively (Figure 3.16b). These sites have different, although related functional profiles (Klickstein et al. 1988; Krych et al. 1991; Krych et al. 1994). Site 1 was shown to be composed of CCP modules 1-3; it primarily binds C4b, while C3b is only weakly bound, and is the primary locus of convertase decay accelerating activity in CR1. Site 2 – of which there are two copies, one formed by CCP modules 8-10 and the other by CCP modules 15-17 – binds both C4b and C3b and is a key contributor to cofactor activity.



The two sites have different- although related functional profiles

	Site 1	Site 2
C4b-binding	yes	yes
C3b-binding	weak	yes
Decay accelerating activity	yes	weak
Cofactor activity	weak	yes

Figure 3.16: CR1 inter-module sequence relationships. (a) This figure was extracted from Chapter 2 (Figure 2.7). The thick lines connect modules that share more than 90% pairwise sequence identity; thin lines indicate 50% to 89% identity, with a label indicating the percentage identity. For clarity, not all lines are drawn (more than 90% identical) since the level of identity can be deduced in some cases (*e.g.* CR1~3 and CR1~17). Pairwise sequence identities calculated using BLAST (Altschul et al. 1990). **(b)** Modular representation of CR1 showing the functional sites: Site 1, and two copies of Site 2, and their corresponding functional profiles below. The four heptad long homologous repeats (LHRs) are also labelled.

3.4.6 Surface comparisons pinpoint functional sites

While CR1~2 and CR1~9/CR1~16 do not cluster together on the basis of their electrostatic surfaces, the differences between them lie primarily on one face - that presented in the right-hand frame of Figure 3.17. This face of CR1~16 is dominated by electropositive side-chains while the equivalent face of CR1~2 displays significantly less positive charge and a substantial amount of negative charge. The remainder of the module surfaces, on the other hand, appear similar; those charged residues present are conserved or conservatively replaced (R64 is conservatively replaced with K964, D is conserved at positions 68 and 968, and R is conserved at 122 and 1022). Thus further inspection of a pair of modules that have diverged in

terms of surface properties, but not in sequence, immediately suggests a putative binding face that could be explored using mutagenesis.

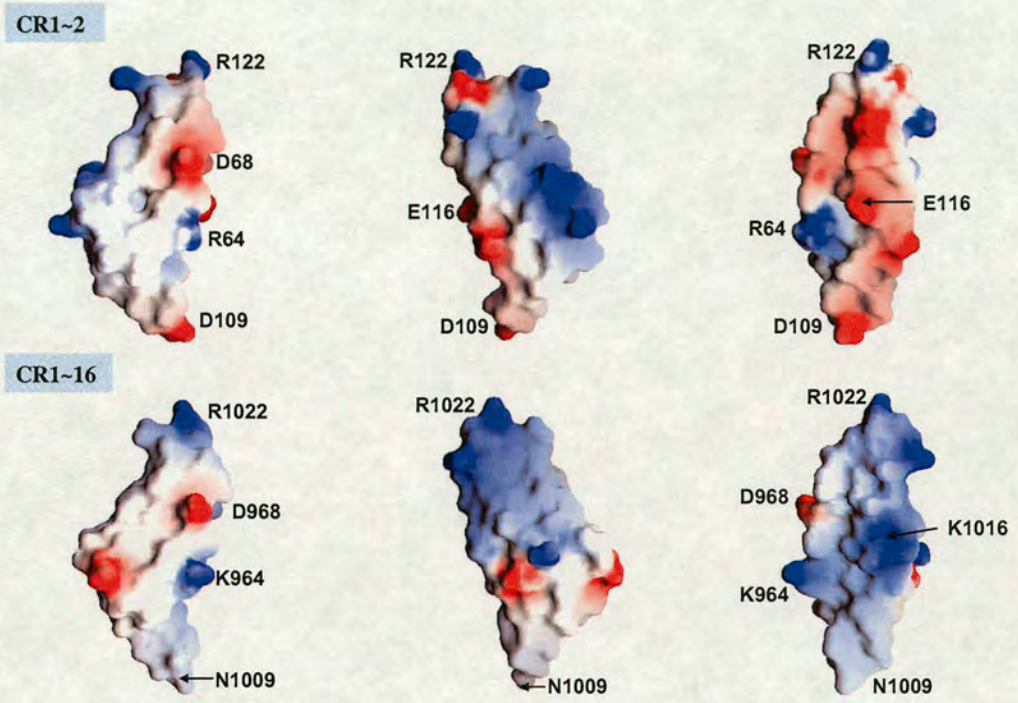


Figure 3.17: A more detailed comparison of surface electrostatics of CR1~2 and CR1~16. Three views, rotated by 120° about the *y*-axis, are shown for each module. The faces of CR1~2 and CR1~16 displayed in the left-hand frame exhibit similarities to one another. In contrast, the faces displayed in the right-hand frame, exhibit clear differences between modules. Those amino acid residues discussed in the text are highlighted.

In fact, residues lying on the divergent faces of these two modules have, in previous work (Krych et al. 1994; Krych et al. 1998), already been substituted for one another. In CR1~16, two such individual substitutions - N1009D and K1016E (Figure 3.17) – brought about, in functional site 2, the loss of iC3 binding [iC3 is a form of C3 that, as a result of hydrolysis of the thioester bond, has a conformation and reactivity similar to that of C3b (Isenman et al. 1981)] and C4b binding, as well as a loss of cofactor activity directed towards both these ligands. Strikingly, the reciprocal

mutation D109N *increased* C3b-binding, while mutation E116K also conferred C3b-binding on site 1 and increased its cofactor activity with respect to C3b and C4b. These experimental results thus fully support the notion that the divergence in surface charge between these modules has occurred amongst functionally critical residues and has arisen through the need for the different sites to perform different functions. Therefore, at least in this example, the approach of multiple surface-comparisons is remarkable in its ability to quickly pin-point important regions. The extent to which such comparisons identify functional sites more generally thus warrants further investigation.

3.5 Further discussion and validation of models in light of new experimentally determined CCP modules

Ten experimentally determined CCP-structure modules (Blein et al. 2004; Harmat et al. 2004; Gal et al. 2005; Rickert et al. 2005; Wang et al. 2005; Herbert et al. 2006c; Jenkins et al. 2006; Lorenzen et al. 2006; Stauber et al. 2006) have been deposited into the PDB since publication of the abovementioned work (Soares et al. 2005) (Figure 3.18; see Chapter 2, Table 2.2 for up-to-date list).

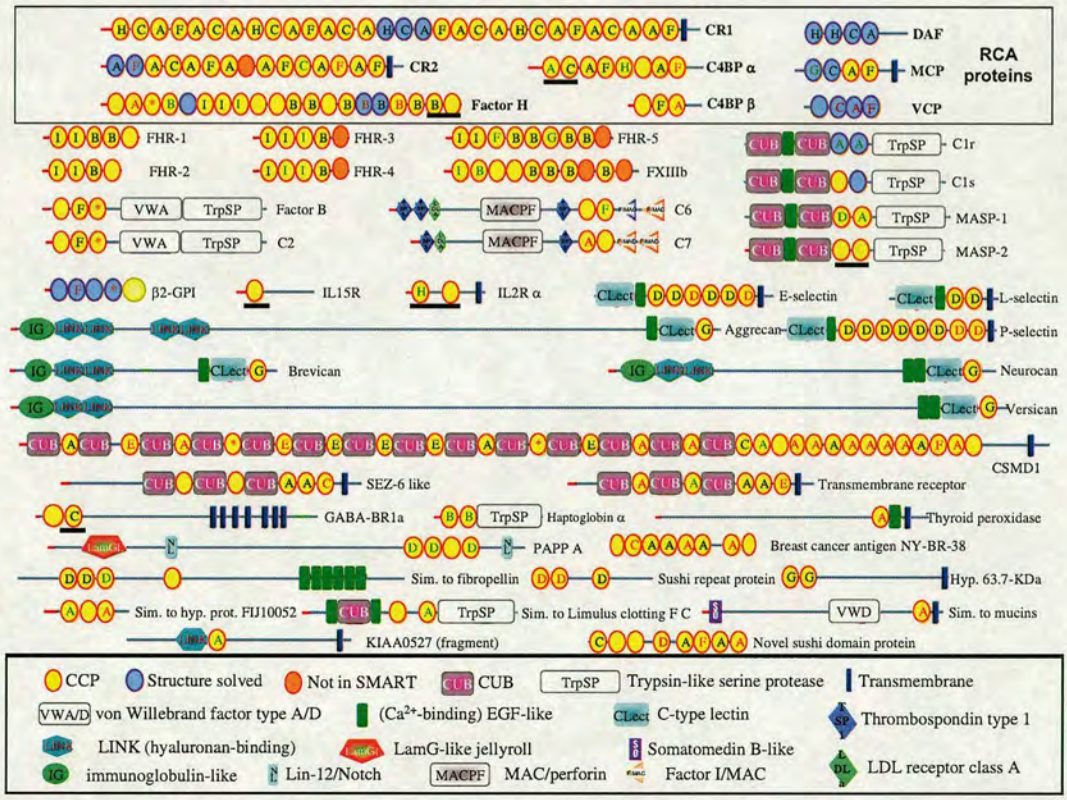


Figure 3.18: Structural status of experimentally determined CCP modules. Revisited cartoon displaying the molecular architectures of a representative set of human CCP-module containing proteins derived from the database SMART (http://smart.embl-heidelberg.de/) as in Figure 3.4. Refer to figure 3.4 and key for explanation of symbols. All previously solved CCP modules at time of submission for publication (Soares et al. 2005) shown in blue. New experimentally determined CCP modules are underlined in bold. These include CCP modules from: Interleukin-2 receptor α chain (IL-2Rα) – the two CCP modules; Interleukin-15 receptor α chain (IL-15R) – the N-terminal module,

Gamma-aminobutyric acid type B receptor, subunit 1 α (GABA-B-R1a) solved for rat – two forms of the second CCP-module in *cis*- and *trans*- conformations (it should be noted that there exists only two conservative substitutions between rat and human for this module); mannan-binding lectin-associated serine protease-2 (MASP-2) - the two CCP-modules; factor H – the C-terminal modules (19 and 20); and C4b-binding protein α (C4BPα) – the N-terminal two CCP-modules.

The newly solved structures (Figure 3.18) include five modules not assigned to any cluster (IL-15Rα~1, IL-2Rα~2, fH~20, MASP2~1,2), two from cluster *C* (C4BPα~2 and $GABA_BR1\alpha$~2), and one each from clusters *A* (C4BPα~1), *B* (fH~19) and *H* (IL2Rα~1). These new structures allow further scrutiny of the accuracy of the 3D-models produced by the automated procedures. Recent experimentally determined module structures that do not fall into any sequence cluster will not be discussed in this chapter.

3.5.1 Model versus structure comparisons – Part II

3.5.1.1 The novelty of experimentally determined CCP module structures in IL2Rα and $GABA_B$-R1α

Perhaps the most interesting of the new structures to be determined, is that of the pair of modules from IL2Rα (Rickert et al. 2005; Wang et al. 2005; Stauber et al. 2006).

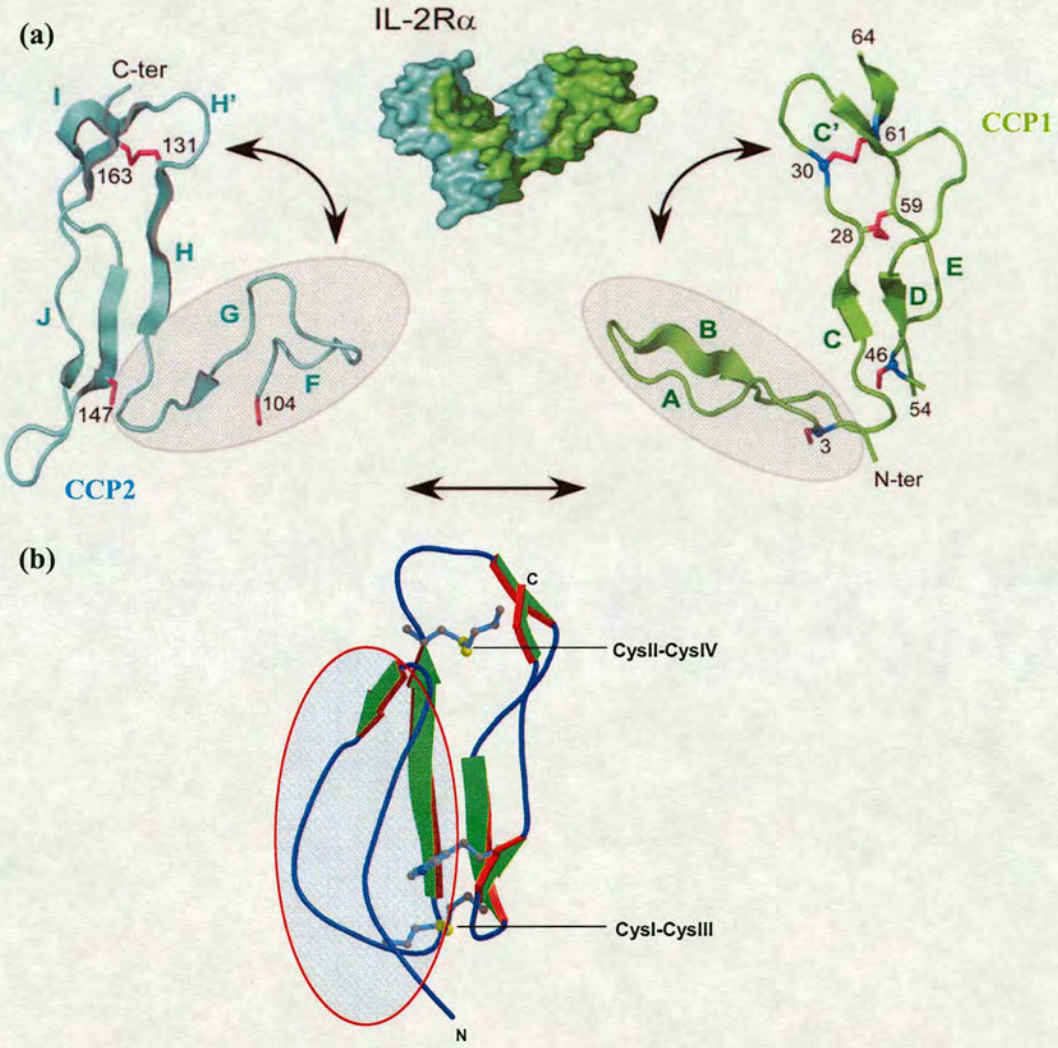


Figure 3.19: Domain swapping in IL-2Rα. **(a)** The IL-2Rα structure (Rickert et al. 2005) is separated into its CCP1 (blue) and CCP2 (green) modules to show the swapped A and B (CCP1), and F and G (CCP2) β-strand locations, respectively. Cysteine residues (six in CCP1 and four in CCP2) are shown in purple as stick representations and labelled. Semitransparent ovals in two diagrams of the separated modules highlight the corresponding strands, which are involved in the strand swap. A surface representation of IL-2Rα in the centre shows the intact pair of modules. The figure is adapted from Rickert *et al.* (Rickert et al. 2005). **(b)** A representative structure of a CCP module [factor H

module 16 (Norman et al. 1991)]. The analogous strand/loop region to those involved in the strand swap between IL-2Rα CCP1 and CCP2 domain are highlighted by a semitransparent oval in fH~16.

According to the sequence clustering, module 1 of IL-2Rα belongs to cluster *H*, while module 2 is not assigned (Figure 3.18). The crystal structure of this module-pair was recently solved (Rickert et al. 2005; Wang et al. 2005; Stauber et al. 2006). It revealed an unusual and unique arrangement for the module-pair, where portions of each module were involved in a structural "swap" with each other. Strands A and B of CCP1 exchange with strands F and G of CCP2, to give the folding topology strand F-on strands-G-C-D-E for CCP1 and strand A-on strands-B-H-I-J for CCP2 (Figure 3.19). The intermodular disulfide bonds between Cys3 and Cys46 of CCP1, and Cys147 and Cys104 of CCP2 enforce the strand exchange by "pinning" strand A of CCP1 to strand I of CCP2 and strand G of CCP2 to strand D in CCP1 to maintain the CCP-module-like fold in both modules.

This unexpected structural aberrance resulted in the model for IL-2Rα~1 being different from other CCP-like module structures, especially for the *swapped strand* portion of the module. The sequence clustering conducted in this work did not pick out this unique arrangement. This begs the question - should this module, IL-2Rα~1, have been placed in cluster-*D*, which are characterised by six cysteine-containing motif sequences (Figure 3.12)? None of the other modules in cluster-*H* contain six cysteines. Furthermore, is there a possibility that consecutive neighbouring pairs of modules in members of cluster-*D* adopt a similar arrangement to IL-2Rα? Both scenarios are unlikely for the following reasons. Even though the sequences of cluster-*D* members contain six cysteine residues per module, these are unlikely to undergo a strand swap, given that the additional pair of cysteines are present at different locations to the ones seen in the sequence of IL-2Rα (Figure 3.20) and

additionally, the extra pair of cysteine residues among cluster-*D* members (example modelled in Chapter 4) are close in space within the folded CCP module - to readily form three stable intramodular disulfides, unlike IL-2Rα~1.

Figure 3.20: Alignment of cluster-*D* and IL-2Rα~1 sequence. ClustalX alignment of a selection of cluster-*D* sequence members with cluster-*H* member, IL-2Rα~1 (P01589_01_H), to bring out their variation in sequence. The alignment reveals that the extra pair of cysteines in IL-2Rα~1 (circled) is located at different positions in comparison with other cluster-*D* members. The sixth cysteine in IL-2Rα~1 is located in the linker region of its sequence. The disulfide pattern for cluster-*D* is indicated above. Region involved in strand swap in IL-2Rα~1 shown with a dashed line below its sequence.

It should be pointed out that in the functionally similar IL-15Rα, which contains a single CCP module, the experimentally determined CCP-module forms a regular CCP-module structure with the expected disulfide pattern formation (Cys I – Cys III, Cys II – Cys IV) (Lorenzen et al. 2006).

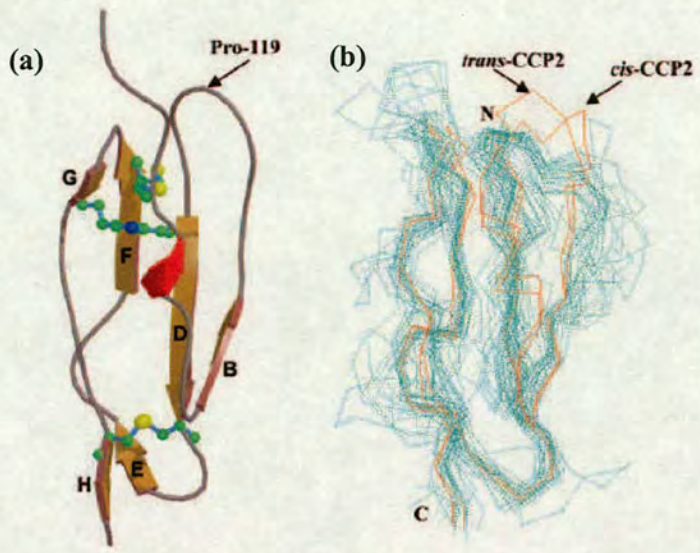


Figure 3.21: Solution structure of rat GABA$_B$-R1α CCP2. **(a)** Molscript (Kraulis 1991) representation of the closest-to-mean structure of the *trans*-form. Assignment of β-strands was based on a consensus among members of the ensemble (Procheck-NMR (Laskowski et al. 1996)) and annotated according to Henderson et al. (Henderson et al. 2001). Cysteine and tryptophan are shown in ball-and-stick representation. The helical segment is seen in red. **(b)** Multiprot (Shatsky et al. 2002; 2004) Cα trace superposition of experimentally solved CCP module structures (*trans*-CCP2 and *cis*-CCP2 in red, others in cyan). The figure is taken from Blein *et al.*, 2004 (Blein et al. 2004).

From the solution structure of the CCP module pair in the heterodimeric G-protein-coupled receptor GABA$_B$-R1α of rat, it was seen that only one of the two modules was compactly folded - module 2 (Blein et al. 2004). Module 1 was experimentally confirmed as containing large regions of disorder. The structure of module 2 in human GABA$_B$-R1α~2 (cluster *C*) displayed two forms - a major form with a *trans*-configuration at the Leu118-Pro119 peptide bond, and a minor form with a *cis*-configuration at the same location; *cis-trans* isomerisation is unique among all solved CCP-modules to date (Figure 3.21). These two forms have a backbone r.m.s.d. value of 1.26 Å between them over all 61 residues (Blein et al. 2004). Although its CCP module has an overall 3D-structure that resembles other CCP modules (Soares and Barlow 2005), an additional unusual feature of its sequence is a lack of proline

residues in the stretch following the first consensus cysteine, which corresponds to a short helical segment seen in the closest-to-mean structure (Figure 3.21a), and 37.5% of structures in the ensemble. This is unusual among CCP modules, and the average number of proline residues within this region is two (Figure 3.22).

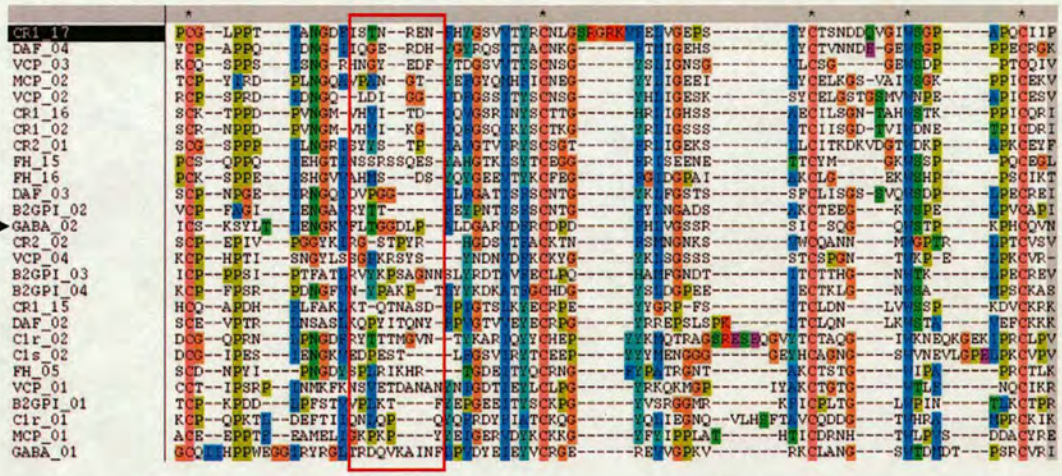


Figure 3.22: Multiple sequence alignment of a number of solved CCP module structures. The hyper-variable loop is marked within the red box. Also included in the alignment are the GABA_01 and GABA_02 sequences (arrow). The lack of proline residues in $GABA_B$-R1α~2 is apparent.

The two conformers were compared with the models produced using the automated procedure, and appear to be rather divergent structurally, with backbone r.m.s.d. values of 2.55 Å for the *cis*-form and 2.92 Å for the *trans*-form. The major sources of deviation occur in the region of the hypervariable loop - the location of peptide bond Leu-Pro isomerisation. It should however be noted, that there exists two conservative substitutions (Val143Ile and Glu127Asp) between the rat structure, and human model.

3.5.1.2 *Comparison of models with new CCP module structures of C4BPα and factor H*

The N-terminal module pair of C4BPα was recently solved by NMR (Jenkins et al. 2006). A structural superposition of individual models of modules and the closest-to-mean NMR structure by Combinatorial Extension (CE) (Shindyalov and Bourne 1998) revealed a Cα r.m.s.d. value of 2.64 Å over 61 residues for C4BPα~1 (a member of cluster *A*), and an r.m.s.d. value of 1.7 Å over 62 residues for C4BPα~2 (belonging to cluster *C*). These values compare favourably with other previously published models by Villoutreix *et al.*, 1998 (Villoutreix et al. 1998) (Table 3.1, Figure 3.23).

	C4BPα~1 NMR structure	**C4BPα~2 NMR structure**
C4BPα~1 model †	2.64 (61)	
C4BPα~2 model †		1.7 (62)
C4BPα~1 model Δ	2.92 (63)	
C4BPα~2 model Δ		3.15 (61)

Table 3.1: Comparison of C4BPα~1 and C4BPα~2 with previously published models. † = Model created by the automated procedure (Soares et al. 2005); Δ = Previously published model by Villoutreix *et al.*, 1998 (Villoutreix et al. 1998). All r.m.s.d. values are in Angstroms (Å). The alignment length is provided below in parenthesis.

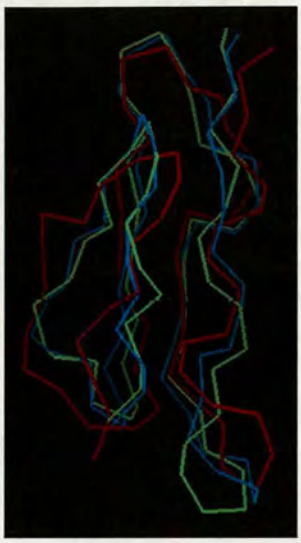


Figure 3.23: Overlay of the experimentally determined C4BPα~2 with models. NMR structure of C4BPα~2 [PDB code: 2A55, (Jenkins et al. 2006)] shown in green; model created by automated procedure (Soares et al. 2005) shown in blue, previously published model by Villoutreix *et al.*, 1998 (Villoutreix et al. 1998) shown in red.

The two C-terminal CCP-modules of factor H (modules 19 and 20) were also solved recently by NMR (Herbert et al. 2006c). CCP-module 19 was assigned to cluster *B*, whereas module 20 was unassigned, consistent with its divergent sequence, lacking the tryptophan residue between cysteines 3 and 4 (Soares and Barlow 2005). While module 20 is structurally divergent to any CCP solved to date, module 19 appears to share features characteristic of *B*-cluster members. A comparison of model versus structure using CE, reveal a backbone r.m.s.d. value of 1.9 Å over all residues. This can be compared with the previously published models by (1) Aslam and Perkins, 2001 (Aslam and Perkins 2001) (model PDB code: 1HAQ, chain A), which has an r.m.s.d. value of 3.46 Å over all 59 residues against the experimental structure; (2) the model published by Hellwage et al., 2002 (Hellwage et al. 2002) (model PDB code: 1FHC), which reveal an r.m.s.d. value of 3.5 Å over 58 residues against the experimental structure, and (3) the model available at the factor H-associated HUS

mutation database (http://www.fh-hus.org/) (Saunders et al. 2006), which has an r.m.s.d. value of 2.8 Å over 55 residues against the experimental structure.

As was the case in the DAF modules, the models created for the two C4BPα modules and factor H module 19, using the automated procedure, are more accurate than the other published ones.

3.6 Declarations and Acknowledgements

- This work has been published:

  Soares, D. C., Gerloff, D. L., Syme, N. R., Coulson, A. F., Parkinson, J., and Barlow, P. N. (2005). Large-scale modelling as a route to multiple surface comparisons of the CCP module family. Protein Eng Des Sel 18, 379-388. Epub 2005 Jun 2023.

- The initial clustering work was done by Dietlind Gerloff, John Parkinson and Andrew Coulson. Neil Syme created all GRASP images for the CR1 electrostatic comparison and assessed the feasibility of using PIPSA for large-scale surface comparison. One PERL script (Appendix 3.8.3.5: pdbres2seq.pl) was obtained from the Protein Data Bank mailing-list. I would like to thank Russell Hamilton (Edinburgh) for system administration support.

- Additional figures, all scripts and protocols are provided in the Appendix and on CD.

Chapter 4

Disease-causing sequence variations in CCP-module-containing proteins: applications of models to medicine

---

4.1 Introduction

A failure to properly control complement underlies the debilitating symptoms of a long list of autoimmune, degenerative and iatrogenic disorders (Chapter 1). Membrane-associated regulators of complement activation (RCA) are present on all host cells, where they have a protective function. Soluble regulators augment the membrane-bound ones and, additionally, act to modulate complement activation on self-surfaces not enclosed by a cell membrane and on apoptotic and necrotic cells. Because there is functional degeneracy amongst the RCA proteins, individuals with variant forms may escape consequences for most of their lives, but they may be predisposed to medical conditions that arise from inappropriate or misdirected complement activation. For example, several mutations in factor H (fH), and a few in membrane cofactor protein (MCP), predispose to atypical haemolytic uraemic syndrome (aHUS) (Warwicker et al. 1998; Richards et al. 2003; Goodship 2004; Goodship et al. 2004), a rare, often fatal kidney disease. A different amino-acid substitution in fH increases the risk of contracting the eye condition, age-related macular degeneration (AMD), a common cause of irreversible sight-loss in people

over 60 (Edwards et al. 2005; Hageman et al. 2005; Haines et al. 2005; Klein et al. 2005). Variations in complement receptor type 1 (CR1) are known to occur with increased frequency in certain malaria-exposed populations in Africa (Moulds et al. 2000), which confer resistance to some pathological symptoms.

Based on the models described in Chapter 3, and new ones described in this chapter, we can predict structural and functional consequences of disease-linked amino acid substitutions. The results of such an exercise are described and discussed below. Disease-causing and human-specific evolutionary mutations that occur in the CCP module of a protein that is probably not a complement participant - sushi-repeat-containing protein, X-linked 2 (SRPX2) are also described (Roll et al. 2006). The sequence data and the 3D-structural models used are all available online at http://www.bru.ed.ac.uk/~dinesh/ccp-db.html. To place this work in context, some background information on diseases caused by sequence variations in the above mentioned proteins is presented first.

4.1.1 Atypical Haemolytic Uraemic Syndrome

Haemolytic uraemic syndrome (HUS) causes the majority of kidney failures in children and is typically associated with bacterial infections. It is characterised by thrombocytopenia, microangiopathic haemolytic anaemia, and acute renal failure (Taylor 2001; Blackall and Marques 2004). A related disorder of the microvasculature is thrombotic thrombocytopenic purpura (TTP). TTP occurs mainly in adults, and is normally accompanied by neurological deficits; but TTP and HUS are regarded by many clinicians as a different manifestation of the same pathophysiological process.

The rare atypical HUS (5-10 % of cases of HUS) is not linked to infection, but is sporadic or familial. While full recovery from typical HUS is the norm, the long-term diagnosis for sufferers of aHUS is unfavourable. A growing list of genetic mutations and SNPs (Table 4.2 and 4.3) within the RCA gene cluster has been associated with aHUS (Warwicker et al. 1998; Buddles et al. 2000; Richards et al. 2001; Caprioli et al. 2003; Dragon-Durey et al. 2004; Sanchez-Corral et al. 2004). These include both missense mutations yielding normal plasma levels of factor H and MCP, and mutations leading to fH and MCP deficiency. Factor H deficiency has also been linked to membranoproliferative glomerulonephritis type II (MPGN II), and (in one case) to type III collagen glomerulopathy (Ault 2000). Additionally, some aHUS patients carry mutations in the protein encoded by the factor I gene (Fremeaux-Bacchi et al. 2004) – a complement, non-CCP-containing protein, which will not be discussed further.

Numerous missense mutations associated with aHUS cluster in fH CCPs 19 and 20 (Perez-Caballero et al. 2001; Esparza-Gordillo et al. 2005), and a haplotype block involving alleles spread across the RCA gene cluster (Esparza-Gordillo et al. 2005) is strongly linked to disease. A reasonable biological model would involve a deficit in fH-mediated complement-regulation on the basement membrane exposed by the large fenestrations of the capillary endothelium within the glomerulus. If accompanied by a deficit of one or more cell-surface complement regulators, the defensive capacity of near-by cells might be overwhelmed. Alternatively, inadequate protection of endothelial cells against complement activation, might lead to compromise of the endothelial cell layer and exposure of the underlying basement membrane that is not protected by a fully functional fH. Environmental triggers, such as immunosuppressive drugs, cancer therapies, and oral contraceptives, are also

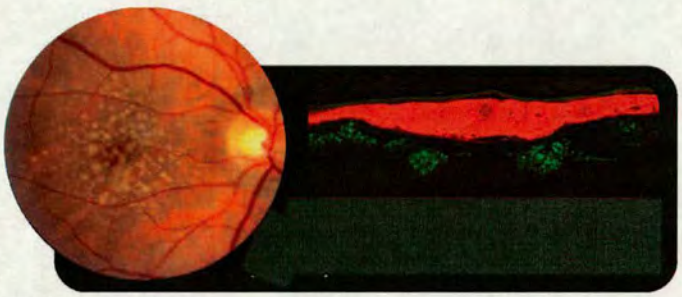
involved. Thus the disease is probably multifactorial with fH dysfunction playing a central role. Intriguingly, a factor H mutation was identified in a TTP patient who developed renal failure – the mutation was absent in a sibling who also had TTP but presented with exclusively neurological symptoms (Noris et al. 2005). Overall, the correlation of mutations/SNPs with disease penetrance, disease severity, prognosis, and choice of treatments, remains obscure in the case of aHUS.

4.1.2 Age-related Macular Degeneration

As a disease of the macula, a region of densely packed photoreceptors towards the centre of the retina, AMD affects central-field vision leading to significant loss in quality of life for up to 50 million sufferers worldwide. AMD is a complex disease, probably involving a combination of genetic factors and environmental triggers, such as diet and (especially) smoking. The characteristic feature of early AMD is the presence of large fatty yellow deposits – drusen - between the retinal pigment epithelium (RPE) that underlies and nourish the photoreceptors, and the layer of extracellular matrix known as Brüch's membrane (BM) that separates the RPE from its blood supply in the choroid (Figure 4.1). The drusen intrude into the RPE and contain cell debris in addition to copious quantities of immune-associated molecules. In particular, they contain both complement regulators and the C5b-C9, membrane attack complex (Johnson et al. 2001). Hageman *et al.* cleverly noticed a connection between AMD and MPGN type II, a rare kidney disease (Mullins et al. 2001). Some MPGN II patients develop ocular drusen that closely resemble those of AMD. Glomerulonephritis had been connected to fH dysfunction or deficiency, leading Hageman *et al.* to perform an SNP analysis of *FH1* in two large cohorts of AMD patients and matched control sets (Hageman et al. 2005). This led to the identification

of both at-risk and protective haplotypes clustered in the *N*-terminal half of factor H (Table 4.2). The most common at-risk haplotype – present at a frequency of 50% in AMD cases and 29% in controls - included the SNP Y402H. Simultaneously, three other groups had independently identified Y402H of fH as a susceptibility-SNP in North American populations (Edwards et al. 2005; Haines et al. 2005; Klein et al. 2005).

(a)



(b)

Figure 4.1: Age-related macular degeneration. **(a)** The retinas of AMD eyes (top left) develop abnormal deposits called drusen (yellow spots); as shown by the red staining (right), drusen contain complement proteins. **(b)** AMD robs it victims of central vision (as illustrated by the simulated picture on the bottom right), and can leave them legally blind even though a rim of peripheral vision may remain. Figure taken from (Marx 2006).

A plausible disease model would involve an instigating, localised, release of debris from necrotic RPE cells that is deposited in the space between the RPE and BM. The deposits would include potent stimulators of the alternative pathway. In the absence of fully functional fH, they could trigger a self-perpetuating cycle of inflammatory events involving bystander cell-damage, further release of cell debris,

and accumulation of immune molecules. This process would lead over time to formation and then expansion of drusen to the point where they compromise the RPE and incite phenomena associated with the later stages of AMD – geographic atrophy and choroidal neovascularisation. Other genes have also been linked to AMD, and it may well be that disease instigation and progression depends upon additional proteins malfunctioning, as well as environmental factors. A SNP in the coding region of the *LOC387715* gene (Rivera et al. 2005), accounts for 40% of AMD cases. Variation in the complement *factor B* gene (Gold et al. 2006) have also been linked to AMD. Studies conducted in Chinese, Finnish, French, Italian, and Russian populations confirmed the existence of the Y402H variant of fH among AMD patients (Souied et al. 2005; Fisher et al. 2006; Lau et al. 2006; Seitsonen et al. 2006; Simonelli et al. 2006). Surprisingly, another study among patients of AMD in a Japanese population, did not make any statistically significant link between factor H variants and AMD (Okamoto et al. 2006).

4.1.3 Role of CR1 in malaria

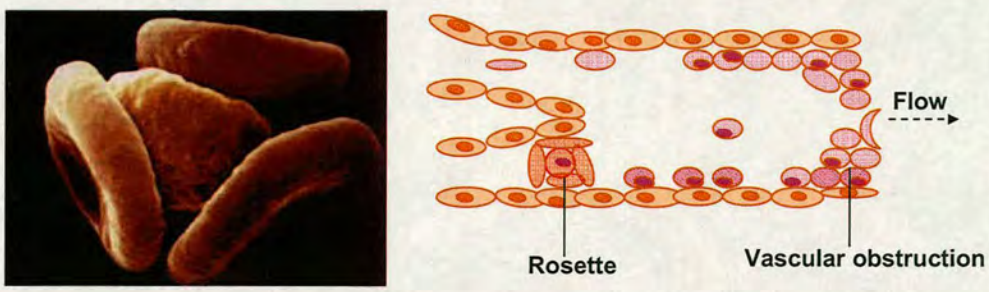


Figure 4.2: Rosette formation. The left panel (image source: Texas Medical Center website) shows an infected RBC (centre) binding healthy RBCs causing clumping or agglutination called "rosetting". The right panel depicts formation of a rosette in the microvasculature. Rosetting is probably a strategy used by the parasite to remain sequestered in the microvasculature to avoid destruction in the spleen and liver. Erythrocyte rosetting causes obstruction of the blood flow in microcapillaries. Formation of rosettes correlates with severity of malaria (Rowe et al. 1995).

Malaria is an infection caused by the parasite genus *Plasmodium*, and is spread from person to person by mosquitoes. Each year, malaria kills over a million people, primarily in Sub-Saharan Africa (source: World Health Organization, 2006), where infection is caused by the *Plasmodium falciparum* strain – the most severe, and life-threatening form of the disease.

Like other complement regulators, CR1 has been hijacked by pathogens, to facilitate entry into host cells or down-regulate complement activation (Chapter 1). In this regard, CR1 has been reported to act as a receptor on the surface of uninfected red blood cells (RBCs) for a protein encoded by the *Plasmodium falciparum* genome – *i.e. P. falciparum* erythrocyte membrane protein 1 (PfEMP1), that is expressed on the surface of infected RBCs (Rowe et al. 1997; Krych-Goldberg et al. 2002). This interaction contributes to agglutination and rosetting – events that are important in malaria pathogenesis (Figure 4.2) (Rowe et al. 1995; Rowe et al. 1997; Rowe et al. 2000).

Phenotype[Δ]	Frequency (%)			Amino Acid	Correlation with malaria
	Caucasians	African-Americans	Africans		
McC(a+)	98	99	82-95	Lys 1590	Unknown
McC(b+)	1	44	44-56	Glu 1590	Unknown
Sl:1	99	65	37	Arg 1601	
Sl:2	1	39	68-72	Gly 1601	Reduced rosetting of Sl:2 RBCs with RBCs infected with *Plasmodium falciparum*

Table 4.1: Knops blood group polymorphisms. [Δ] McC = McCoy; Sl = Swain-Langley. McCa and McCb are one allelic antigen pair. Sl1 and Sl2 are another pair. The corresponding phenotypes for the first pair are McC(a+) and McC(b+), and for the second pair are Sl:1 and Sl:2. Table adapted from (Krych-Goldberg et al. 2002).

CR1 in humans exhibit various genetic polymorphisms, including those represented by the Knops blood-group system (Table 4.1) (Moulds et al. 1991; Rao et al. 1991; Moulds 2002; Moulds et al. 2002), which is of interest in this chapter. In the most common allotype of CR1 (the 30 CCP module size-variant, containing four LHRs), two of the Knops blood-group antigens were traced to point mutations that produce a single amino acid residue change in CCP 25 (Moulds et al. 2001; Moulds et al. 2002). An A to G mutation, at nucleotide 4795, causes a Lys1590Glu change, and defines the McCa/McCb polymorphism. An SNP at nucleotide 4828 causes an Arg1601Gly change, and defines the Sl:1/Sl:2 polymorphism. The frequency of McC(b+) and Sl:2 phenotypes or polymorphisms are greatly increased in malaria-exposed African populations as compared to Caucasians (Table 4.1) (Moulds et al. 2000). This suggests a selective pressure for McC(b+) and Sl:2 in malaria-exposed populations.

The pathological event proposed to be transformed by these epitopes is rosetting. Rosetting of RBCs carrying the Sl:2 variant is lower compared to RBCs carrying the Sl:1 variant (Rowe et al. 1997). Reduced rosette formation would decrease the microcapillary occlusion that impairs blood flow in malaria (Figure 4.2). Further, the number of the parasites sequestered in the bloodstream would be less if rosetting was inefficient (Miller et al. 1994).

4.1.4 *Disease-causing and evolutionary mutations in SRPX 2*

Sushi-repeat protein, X-linked 2 (SRPX2) is a secreted protein expressed in neurons of the adult human brain. It is made up from four major domains; three putative CCP modules, with one hyalin repeat domain (HYR) inserted between CCP modules 2 and 3 (Figure 4.3a). It shares 35% sequence identity with cluster-*D* member SRPX (Figure 4.3b) and displays sequence characteristics of other cluster-D members (Chapter 3), including conservation of the six-cysteine residue motif.

(a)

(b)

```
SRPX2-CCP1     1 CYTLNIQDCEATCYSPKCCNYHSSLGTRCELSCDRCFRPICRRSVCCLPSRRWSGTAYC
SRPX-CCP1      1 CSPIKVKYCDVYCRAEQCGCYYRTALGTRCDIRCQKCYDLHGSSLLICQSNRRWSDKVIC
consensus      1 *   . .  *.  *  * ** *.. *****.. * .*. * *    . *    .***   *
```

Figure 4.3: Modular architecture and sequence of SRPX2. **(a)** Modular architecture of SRPX2. The 465-residue SRPX2 protein is made up of a 23 amino acid signal sequence (shown in red), followed by a 34-residue stretch preceding the three CCP modules and Hyalin repeat domain (HYR), which is inserted between CCPs 2 and 3. Figure is taken from SMART (http://smart.embl-heidelberg.de/). **(b)** Pairwise sequence alignment between CCP1 of SRPX and SRPX2. BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html) output for pairwise comparison between CCP1 of SRPX and SRPX2. Identities = 21/59 (35%) marked by '*', Positives (identities and conservative substitutions) = 36/59 (61%) marked by '.', Gaps (insertions or deletions) = 0/59 (0%).

The *SRPX2* gene was recently identified as being responsible for two related disorders of the rolandic/sylvian speech areas (Roll et al. 2006). The rolandic and sylvian fissures divide the cortex hemispheres of primate brains into their main anatomical structures. In humans, these areas participate in speech production under the control of the Broca's area.

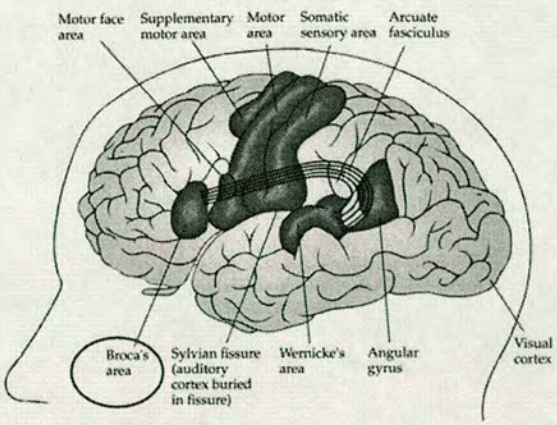


Figure 4.4: Main anatomical structures of the brain involved in linguistic activity. Image taken from "The Brain: A Neuroscience Primer, 3rd edition"; Worth Publishers, New York, 2000, pp. 442 by Richard F. Thompson.

Two mutations; (i) N327S, which results in gain-of-glycosylation in an as yet unidentified region, and (ii) Y72S located in the first CCP module, were implicated in rolandic seizures associated with oral and speech dyspraxia and mental retardation (Roll et al. 2006). Since it is linked to defects in the functioning and the development of such brain regions, *SRPX2* may be one of the specific genes whose evolution may have participated in the recent emergence of higher-order human-specific cognitive functions.

In a study conducted by Royer et al., 2006 (unpublished), the molecular evolution of the SRPX2 gene was examined (see attached preprint in appendix). While population genetics did not detect a selective sweep, adaptive evolution in the human lineage was suggested by a synonymous/non-synonymous analysis (a stochastic effect was not ruled out). One single, fixed amino acid change (R75K) occurred in the first CCP module of SRPX2 after the human-chimpanzee split. The three-dimensional modelling work, discussed in this chapter, examined the effects of both this evolutionary mutation (R75K) and the previously identified disease-causing mutation (Y72S).

4.2 Materials and Methods

Homology-derived models for eight modules in factor H - numbers 2, 3, 4, 11, 12, 14, 17, and 18; and the remaining two and 27 modules in MCP and CR1 respectively - are downloadable from the CCP-module model database (http://www.bru.ed.ac.uk/~dinesh/ccp-db.html). These were created using the large-scale modelling procedure described in Chapter 3 (Soares et al. 2005). Two other modules in factor H - numbers 1 and 7 - were homology-modelled individually for the purposes of this exercise, and are likewise available online.

4.2.1 Modelling method for factor H CCP modules 1 and 7

Modelling of each of these two CCP modules was based upon the structure with the most similar sequence as identified in the PDB (Berman et al. 2000) (http://www.rcsb.org/pdb) with a BLAST search (Altschul et al. 1990), and employed the program Modeller release 8 version 1 (Sali and Blundell 1993). Modules 1 and 7 were based upon templates DAF~2 (PDB ID: 1OK3) and VCP~4 (PDB ID: 1G40) respectively. The sequence alignments between target and template sequences (Figure 4.5) were created based on initial multiple sequence alignments, which were further manually edited to ensure the most plausible alignment of conserved amino acid residues and secondary structure elements between the templates and those strongly predicted by PsiPred version 2.4 (McGuffin et al. 2000) in the target sequences. Disulfide bridges were automatically restrained during model building. Twenty models were generated in each case, and the one with the lowest *objective function score* was selected as the representative model.

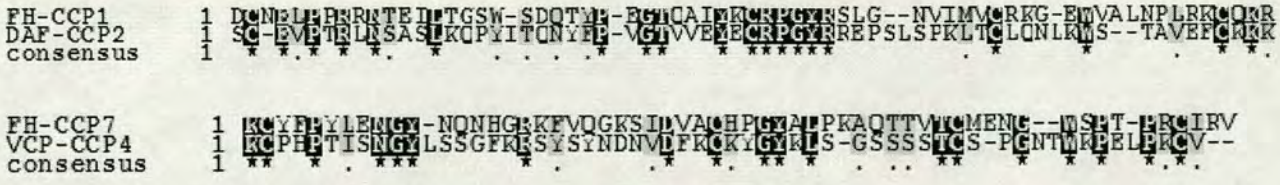


Figure 4.5: Optimal target-template alignment used for modelling fH CCPs 1 and 7. BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html) output of the optimal sequence alignment between target fH-CCP1 and template DAF-CCP2 (above) and fH-CCP7 and VCP-CCP4 9below). Identical residues are marked by '*' and conservative substitutions shown by '.' in the consensus line.

For each representative model, non-identical side-chains between template and target were optimised using the side-chain replacement program, SCWRL version 3 (Bower et al. 1997). The models were then protonated, and subjected to energy minimisation carried out in SYBYL version 6.9 (Tripos Associates, St. Louis, MO, USA) using the Tripos forcefield (30 steps steepest descent followed by 30 steps conjugate gradient). The models were finally checked for valid stereochemistry using PROCHECK version 3.5.4 (Laskowski et al. 1993).

The accuracy of these two models may be inferior relative to those created previously by the large-scale modelling strategy because of lower sequence similarity between targets and templates. They represent, however, the product of the best available modelling techniques and are available for download at the CCP-module model database (http://www.bru.ed.ac.uk/~dinesh/ccp-db.html).

4.2.2 Modelling method of the first CCP module of human SRPX2

Modelling of the first CCP module of human SRPX2 was undertaken based on its closest homologue in the Protein Data Bank (PDB) (http://www.rcsb.org/pdb) (Berman et al. 2000) - the first CCP module of complement receptor type 2 (CR2) (Prota et al. 2002) (PDB ID: 1LY2), identified with a BLAST search (Altschul et al.

1990; Altschul et al. 1997). The target and template share 29% sequence identity from *N*- to *C*-terminal cysteines. The program Modeller release 8 version 1 (Sali and Blundell 1993) was used.

The alignment between the target and template sequences (Figure 4.6) was based on an initial multiple sequence alignment of several individual human CCP-module sequences assigned to cluster-*D* (with the addition of template sequence CR2-CCP1) using the program PROBCONS (Do et al. 2005). Cluster-*D* members are characterised by the presence of a six-cysteine residue motif, forming three putative disulfide bridges (Chapter 3). The target-template sequence alignment was subjected to further minor manual editing, guided by positioning of secondary structure elements more appropriately between the target and template sequences. Secondary structure elements were defined by the server PsiPred version 2.4 (McGuffin et al. 2000) for the target sequence, and DSSP (Kabsch and Sander 1983) for the template structure.

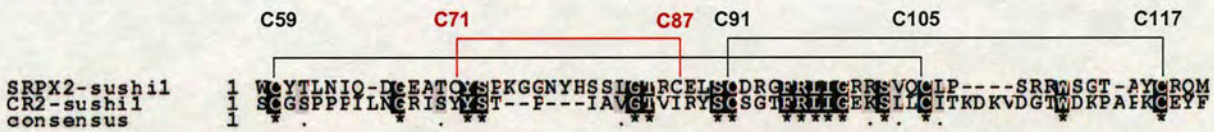


Figure 4.6: Optimal target-template alignment used for modelling SRPX2-CCP1. BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html) output of the optimal sequence alignment between target CCP1 of SRPX2 (SRPX2-sushi1), and template CR2 CCP1 (CR2-sushi1). Cysteine residues involved in disulfide bridge formation are connected by lines and labelled above the SRPX2-CCP1 sequence (the additional disulfide in SRPX2 is shown in red). Identical residues are marked by '*' and conservative substitutions shown by '.' in the consensus line.

It was noticed from a preliminary modelling run, that Cys 71 and Cys 87 lie close in space within the folded structural model. Subsequently, this additional putative disulfide bridge present in the target, absent in its corresponding template, was restrained during model building. Twenty-five models were generated, and the

one with the lowest objective function score (Sali and Blundell 1993) selected as the representative model. The loop '*KGGNY*' in the hypervariable loop region of SRPX2, for which CR2-template-derived restraints were largely absent, was subsequently remodelled under SYBYL version 6.9 (Tripos Associates, St. Louis, MO, USA) after conducting a loop search for that region against a high-resolution subset of the PDB, to select the best root-mean square fitted, matching loop conformation for that region.

Non-identical side-chain residues of the representative model were optimised using the side-chain replacement program SCWRL version 3 (Bower et al. 1997). The model was then protonated and subjected to energy minimisation using the Tripos forcefield (Clark et al. 1989) (20 steps steepest descent followed by 20 steps conjugate gradient) under SYBYL to remove clashes and bad geometries. The model structure was finally checked for valid stereochemistry using PROCHECK version 3.5.4 (Laskowski et al. 1993).

4.3 Results and Discussion

4.3.1 Predicted structural consequences of amino acid substitutions in factor H

Structural predictions (summarised in Table 4.2 below) can be combined with knowledge of phenotype and of the locations of binding sites on factor H (Chapter 1), to help dissect the contribution of the mutation or SNP (Figure 4.8) to the disease process.

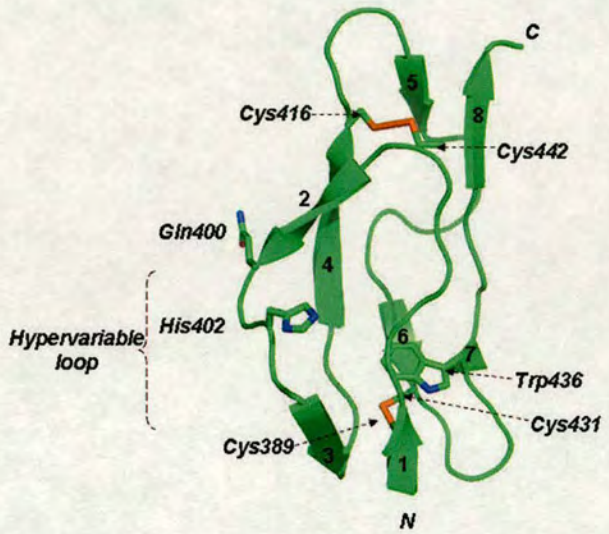


Figure 4.7: 3D-model of factor H CCP module 7. Cartoon schematic PyMol (http://www.pymol.org) representation of module depicting a typical CCP-module fold. β-strands numbered and disulfides and tryptophan are drawn as sticks. The AMD-associated SNP, 402, and aHUS-implicated variant, 400, located in the hypervariable loop are shown as sticks.

For example, the model of module 7 suggests that the SNP involving residue position 402 – a His in the population at-risk of AMD, otherwise a Tyr – is not likely to disrupt the structure of module 7 since its side-chain is largely solvent-exposed (Figure 4.7). Nor is it likely to affect the orientation or flexibility of module 7 with respect to its neighbouring modules. A local effect on the structure or flexibility of the hypervariable loop cannot be ruled out. Module 7 is known to form part of a binding site for polyanions (one of three such sites in factor H), and also to contribute

to a binding site for C-reactive protein (Giannakis et al. 2003). It is therefore likely that its hypervariable loop is involved in direct contact with ligand, and that substitution with His perturbs one or other interaction so as to compromise affinity. Even a minor effect on binding could be critical since incomplete shut-down of the AP, over a time-course of years, could promote the gradual enlargement of ocular drusen. Many individuals with the at-risk allele do not have AMD, so the disease process may require some triggering event – for example compromise of cells in the RPE due to another disease or an environmental insult – or there might be additional gene products involved. The occurrence of another structurally-neutral (according to our predictions) mutation in an adjacent residue, Q400K (Figure 4.7), in an aHUS patient reinforces the hypothesis that there is a binding site in this hypervariable region of module 7.

Table 4.2 summarises the predicted structural consequences for a selection of disease-linked mutations and SNPs of factor H. Many of these involve a consensus Cys; or (not shown) they involve the introduction of stop codons/deletions. Variants in either of these categories are likely to have serious consequences for expression, and structural integrity and stability – indeed many are known to be associated with factor H deficiencies. Another class of mutations are those that are predicted to disrupt the interface between two modules. In this case, it is not likely that expression will be adversely affected, although a more open interface could increase susceptibility to proteolysis. More likely, these mutations illustrate the functional importance of defined levels of flexibility between modules of the intact protein. Of the remaining variations (Figure 4.9), several involve buried hydrophobic residues (V835, W978, W1157, V1197, and F1199) likely to destabilise individual module structures. One of these (W978) affects the stability of module 16, not previously

shown to interact directly with other complement proteins, and therefore having a potential role in the overall protein architecture of fH. On the other hand this CCP module, along with modules 1, 7 (see above), 14, 18, 19, and 20, harbour disease-associated residues that are solvent-exposed suggesting the involvement of ligands, some of which may not have been identified so far.

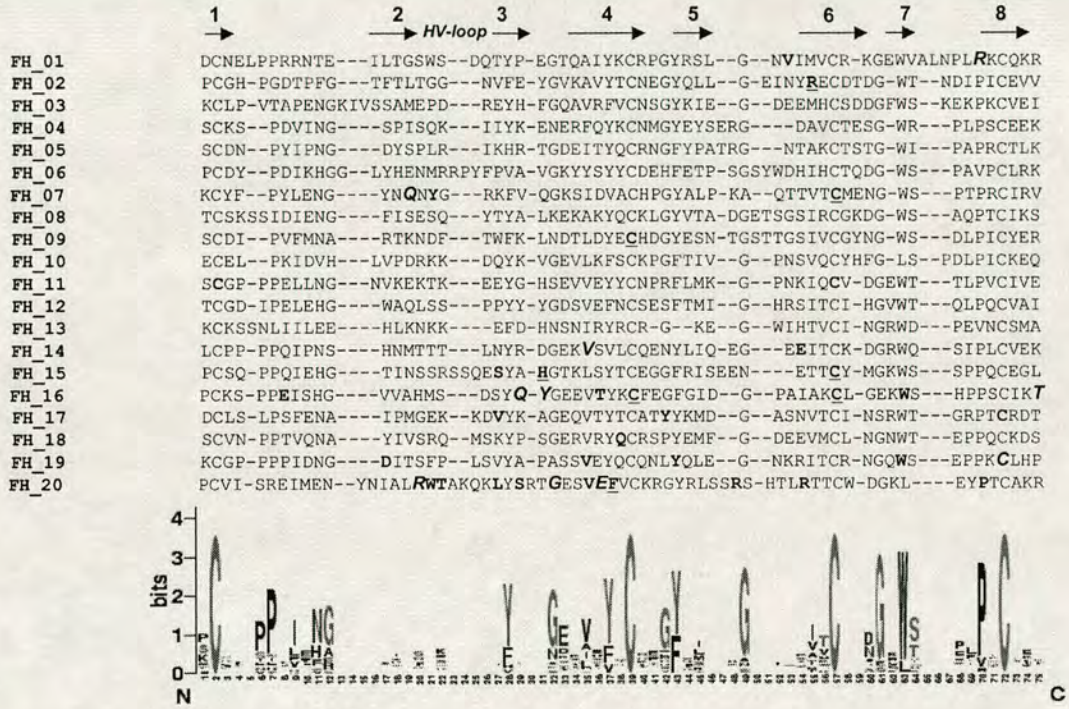


Figure 4.8: Disease-associated variants mapped on the multiple sequence alignment of the 20 factor H CCP modules. β-strands (arrows) indicated above sequences along with hypervariable loop (*HV-loop*). Consensus residues for each position in the alignment (the greater the height of the residue, the greater the conservation) are shown below (Crooks et al. 2004) (http://weblogo.berkeley.edu/). Variants (Table 4.2) highlighted as follows: bold letters indicate the variant is expressed at normal or high levels in plasma; underlined bold residues indicate the variant is expressed at low levels or is absent; italicised bold residues indicate variants where no information is available.

Importantly, the majority of non-structural residues, however, lie in or close to the GAG-binding site. It is interesting that several such mutants are nonetheless pulled out on a heparin column (Lehtinen et al. 2006). On the other hand, heparin is not regarded as a physiological ligand for GAG-binding proteins.

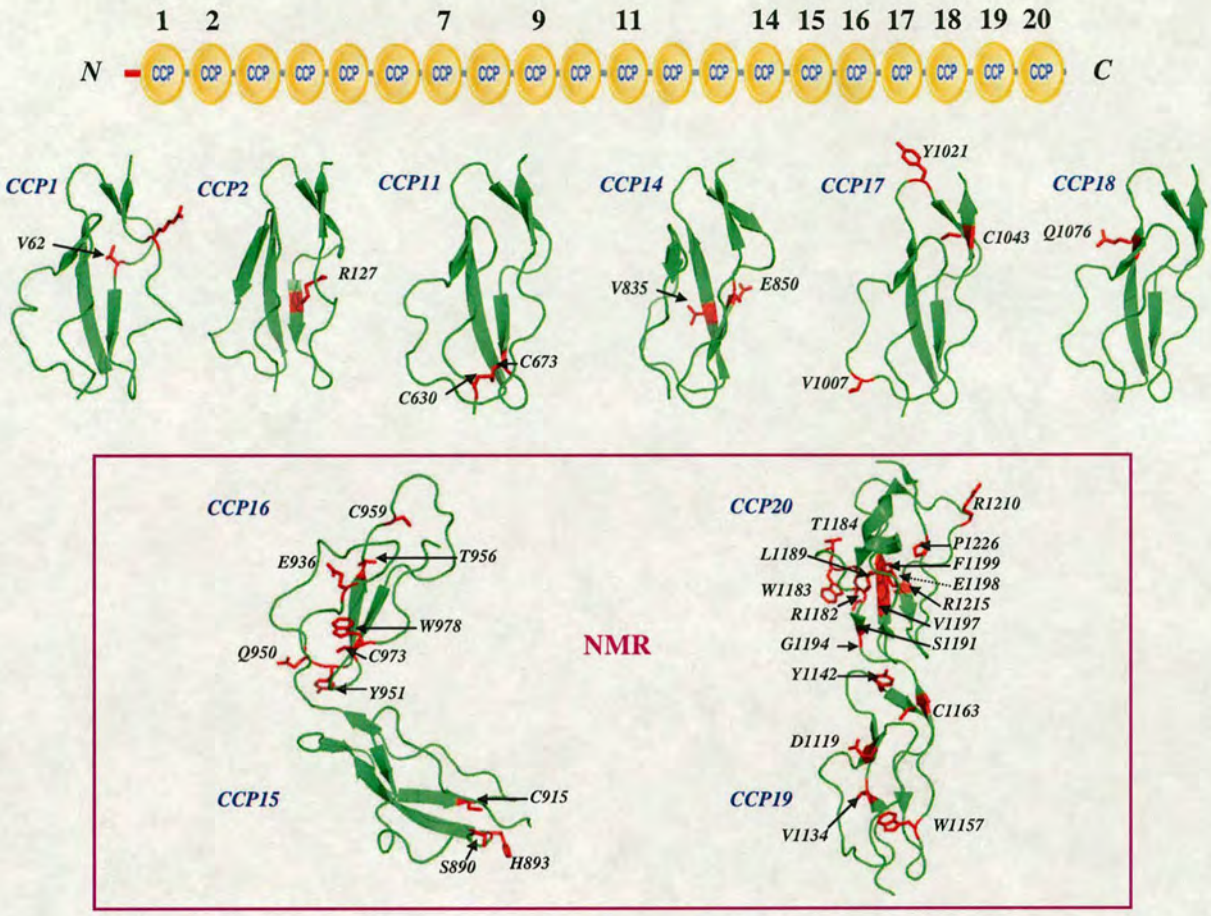


Figure 4.9: Disease-associated factor H variants mapped onto models and NMR structures. SMART representation of factor H, with CCPs involved in disease labelled at the top. Models (CCPs 1, 2, 11, 14, 17 and 18) and NMR structures (CCPs 15 and 16; 19 and 20 – boxed below) shown in cartoon representation, with disease-associated variant residue side-chains shown in stick representation coloured red and labelled. Factor H models for CCP 7 (see Figure 4.7) and CCP 9 are not shown. The collagen type-III glomerulopathy implicated variant in CCP9 (C536R) co-occurs with C959Y in CCP16, and is highly likely to cause structural disruption.

Variant/Reference	CCP	Predicted effect	Pathology	fH level
I62V (Hageman et al. 2005)	1	No effect on structure	AMD	High
R78G (Caprioli et al. 2003)	1	Structural disruption	HUS	n.r.
R127L (Dragon-Durey et al. 2004)	2	Protein more "sticky?"	MPGNII	Low
Q400K (Dragon-Durey et al. 2004)	7	No effect on structure	HUS	n.r.
Y402H (Hageman et al. 2005)	7	No effect on structure	AMD	High
C431S (Dragon-Durey et al. 2004)	7	Structural disruption	MPGNII	Low
C536R,C959Y (Ault et al. 1997; Ault 2000)	9,16	Structural disruption	CIIIG	Absent
C630W (Neumann et al.	11	Structural disruption	HUS	Normal

2003)				
C673Y (Dragon-Durey et al. 2004)	11	Structural disruption	HUS	Low/Normal
C673S (Dragon-Durey et al. 2004)	11	Structural disruption	MPGNII	Low
V835L (Gilges et al. 2000)	14	Structural disruption	HUS	n.r.
E850K (Neumann et al. 2003)	14	No effect on structure	HUS	High
S890I (Noris et al. 2005)	15	Protein more "sticky?"	HUS	Normal
H893R (Dragon-Durey et al. 2004)	15	Disrupts junction	HUS	Low
C915S (Dragon-Durey et al. 2004)	15	Structural disruption	HUS	Low
E936D (Caprioli et al. 2003)	16	No effect on structure	HUS	Normal
Q950H (Caprioli et al. 2003)	16	No effect on structure	HUS	n.r.
Y951H (Caprioli et al. 2003)	16	Disrupts junction?	HUS	n.r.
T956M (Perez-Caballero et al. 2001; Esparza-Gordillo et al. 2005)	16	No effect on structure	HUS	Normal/High
C973Y (Gilges et al. 2000)	16	Structural disruption	HUS	Low/Absent
W978C (Neumann et al. 2003)	16	Structural disruption	HUS	Normal
T987A (Gilges et al. 2000)	16/17	Disrupts junction	HUS	n.r.
V1007C (Gilges et al. 2000)	17	Protein cross-linked?	HUS	Normal
Y1021F (Neumann et al. 2003)	17	Disrupts junction?	HUS	Normal
C1043R (Neumann et al. 2003)	17	Structural disruption	HUS	Normal
Q1076E (Richards et al. 2001; Neumann et al. 2003)	18	No effect on structure	HUS	Normal
D1119G (Richards et al. 2001)	19	Disrupts GAG-binding site†	HUS	Normal
V1134G (Neumann et al. 2003)	19	Structural disruption	HUS	High
Y1142D (Neumann et al. 2003)	19	Disrupts junction	HUS	High
W1157R (Neumann et al. 2003)	19	Structural disruption	HUS	High
C1163W (Caprioli et al. 2003)	19	Structural disruption	HUS	n.r.
R1182S (Gilges et al. 2000)	20	Disrupts GAG-binding site	HUS	n.r.
W1183R (Remuzzi et al. 2002; Neumann et al. 2003)	20	Disrupts GAG-binding site	HUS	Normal
W1183L (Perez-Caballero et al. 2001; Dragon-Durey et al. 2004; Esparza-Gordillo et al. 2005)	20	Disrupts GAG-binding site	HUS	Normal/High
T1184R (Richards et al. 2001)	20	Disrupts GAG-binding site	HUS	Normal

L1189F (Rodriguez de Cordoba et al. 2004; Esparza-Gordillo et al. 2005)	20	C3b-binding?	HUS	Normal
L1189R (Perez-Caballero et al. 2001; Esparza-Gordillo et al. 2005)	20	C3b-binding?	HUS	Normal/High
S1191L (Richards et al. 2001; Neumann et al. 2003)	20	Disrupts GAG-binding site	HUS	n.r.
S1191W (Rodriguez de Cordoba et al. 2004; Esparza-Gordillo et al. 2005)	20	Disrupts GAG-binding site	HUS	Normal
G1194D (Perkins and Goodship 2002; Caprioli et al. 2003)	20	Structural disruption	HUS	n.r.
V1197A (Caprioli et al. 2001; Perez-Caballero et al. 2001; Richards et al. 2001; Caprioli et al. 2003; Esparza-Gordillo et al. 2005)	20	Structural disruption	HUS	Low/Normal
E1198A (Caprioli et al. 2003)	20	Disrupts GAG-binding site	HUS	n.r.
F1199S (Dragon-Durey et al. 2004)	20	Structural disruption	HUS	Low
R1210C (Caprioli et al. 2001; Perkins and Goodship 2002; Sanchez-Corral et al. 2002; Caprioli et al. 2003; Neumann et al. 2003; Esparza-Gordillo et al. 2005)	20	Protein cross-linked?	HUS	Normal/High
R1215G (Warwicker et al. 1998; Richards et al. 2001)	20	Disrupts GAG-binding site	HUS	Normal
R1215Q (Caprioli et al. 2001)	20	Disrupts GAG-binding site	HUS	Normal
P1226S (Neumann et al. 2003)	20	Structural disruption	HUS	High

Table 4.2: Disease-associated factor H variants. † The GAG-binding site on fH~19,20 has been very recently delineated experimentally (Herbert et al. 2006b). Abbreviations used in Table4.2: CIIIG = collagen type III glomerulopathy; n.r. = not reported/known; other abbreviations explained in text.

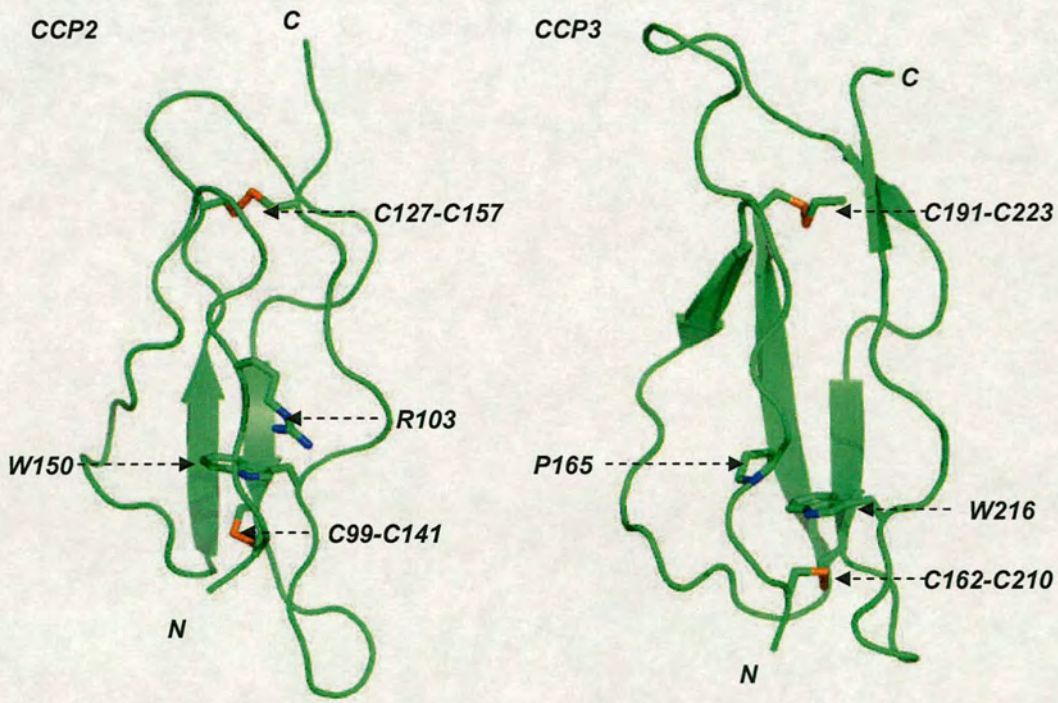
The data in Table 4.2 provides a very clear indication that it is the inability of variant forms of factor H to distinguish, properly, self- from non-self surfaces via their carbohydrates that is the dominant disease-causing factor.

4.3.2 Predicted structural consequences of amino acid substitutions in MCP

To date, only three missense disease-associated mutations have been observed in MCP that predispose individuals to aHUS (Table 4.3, Figure 4.10) – one each, occurring in modules 2, 3 and 4.

Variant/Reference	CCP	Predicted effect	Pathology	MCP level
R103W[#] (Esparza-Gordillo et al. 2005)	2	Affects C3b-binding?	HUS	Normal
P165S (Esparza-Gordillo et al. 2005)	3	Disrupts structure	HUS	Low
S240P[*] (Richards et al. 2003)	4	Affects C3b-binding and cofactor activity toward C4b	HUS	Normal

Table 4.3: Disease-associated MCP variants. [#]R103W was erroneously reported as R103T in (Esparza-Gordillo et al. 2005). A corrigendum was subsequently published in this regard. [*]S240P was reported as S206P in (Richards et al. 2003) – this numbering excluded the signal peptide. For consistency, all numbering in the above table include the 34 amino acid signal peptide.



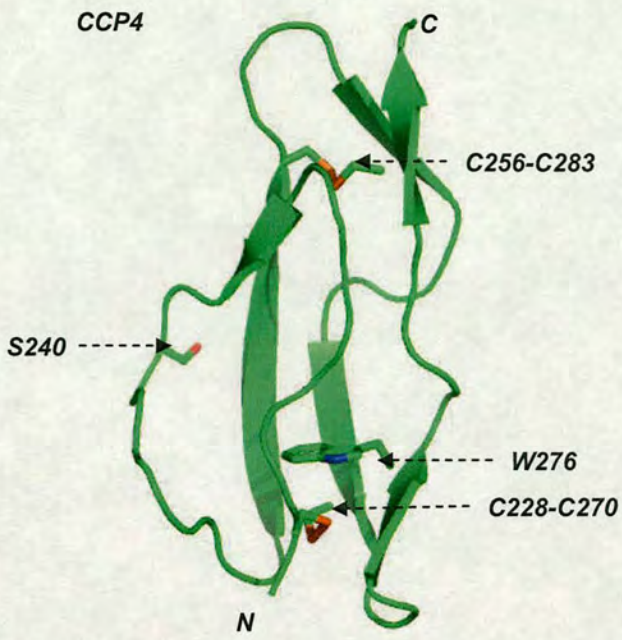


Figure 4.10: Crystal structure (CCP 2) and 3D-models (CCP 3 and 4) of MCP. Cartoon schematic PyMol (http://www.pymol.org) representation of MCP CCP-modules 2, 3 and 4, with disulfides and tryptophan drawn as sticks and labelled. The location of aHUS-associated SNPs; R103 (CCP2), P165 (CCP3) and S240 (CCP4) are shown in stick representation and labelled.

Arg 103 in module 2 is highly surface exposed (Figure 4.10), away from the interface with neighbouring modules. Substitution of a positively charged Arg with a hydrophobic aromatic Trp residue at that position would have a drastic effect on the surface properties in that region and could thus have a functional effect. It is therefore likely that R103 is in direct contact with a ligand. Pro 165 located in module 3, on the other hand is completely buried (Figure 4.10). A proline residue at this position is commonly conserved, as part of the CCP module scaffold (Chapter 2, Section 2.4.1). A change from a conformationally important proline, to a polar serine amino acid could therefore, impact on local structure and disrupt it. Ser 240 is surface exposed and located at the start of the hypervariable loop in CCP4 (Figure 4.10). It has been shown to be important for complement regulation through mutagenesis experiments (Liszewski et al. 2000). A change to a proline residue at this position would thus have

both a structural and a functional effect. In fact, individuals with the S240P variant expressed normal quantities of protein, but demonstrated approximately 50% reduction in C3b-binding in heterozygotes, and a complete lack of C3b-binding in homozygotes (Richards et al. 2003). This suggests that it is the failure of the variant form to inactivate or regulate C3b that mediates development of disease (Richards et al. 2003; Esparza-Gordillo et al. 2005).

4.3.3 Model of CR1 CCP 25 sheds light on haplotypic variants in malaria-exposed populations

From the data provided in Table 4.1, epitopes in CR1-CCP 25 appear to protect against severe malaria. The availability of 3D-models provides a useful means of inferring how such SNPs influence the surfaces properties of the module concerned and how this could affect their functions.

The model of wild-type CCP 25 was downloaded from the CCP module model database (http://www.bru.ed.ac.uk/~dinesh/ccp-db.html) (Soares et al. 2005). The mutant version was modelled using the same alignment as used for the wild-type model from the large-scale procedure, but substituting the two implicated residues and subsequently remodelled using Modeller (Sali and Blundell 1993).

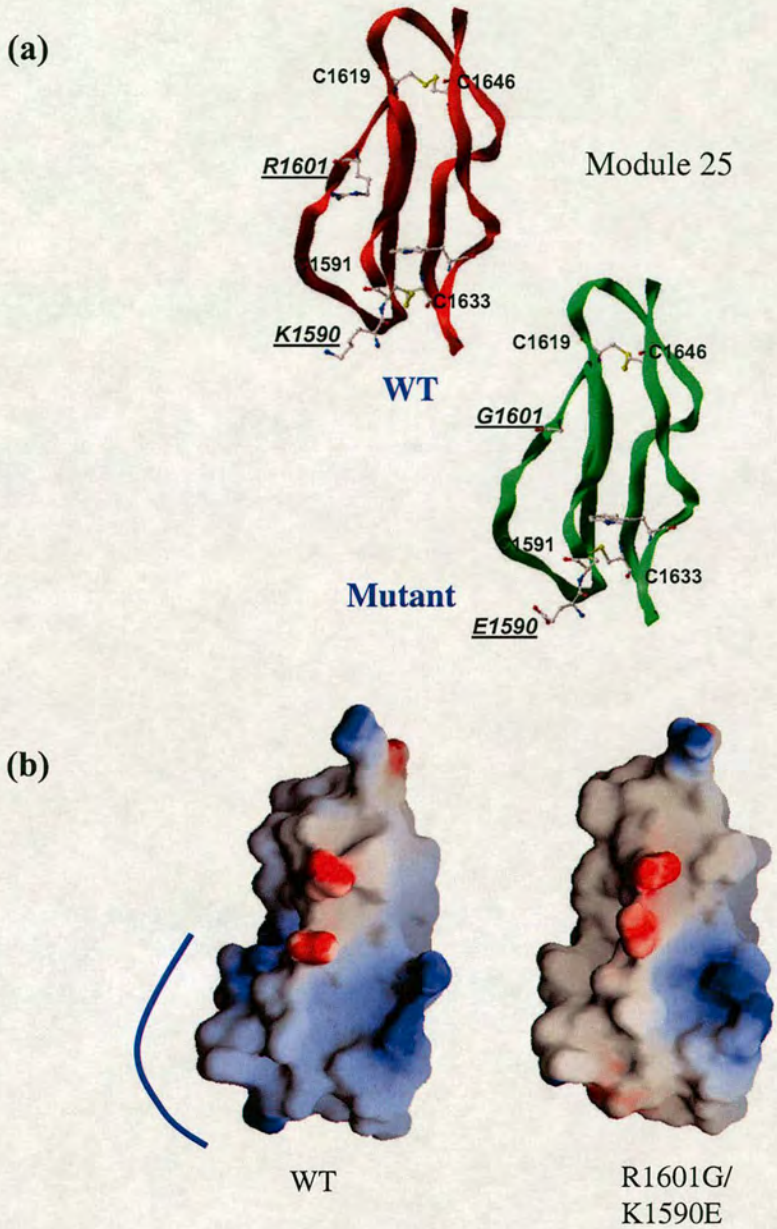


Figure 4.11: Interpreting CR1 SNPs in malaria-exposed populations. (a) The upper panel shows wild-type (WT) (red) and variant (green) models for CR1~25 as backbone ribbons, with disulfides, conserved tryptophan and SNP residues shown in ball-and-stick representation. Figure created using SYBYL (Tripos Associates, St. Louis, MO, USA) **(b)** The lower panel depicts the electrostatic GRASP surfaces (Nicholls et al. 1991) of wild-type and haplotypic variants in CR1~25. The region of corresponding surface-charge change is shown by the blue line.

The modelled structure reveals that the side-chains of R1601 and K1590 are surface-exposed and proximal to one another (Figure 4.11a). R1601 is located at the

start of the hypervariable loop, which commonly provides a large, exposed surface area for protein-protein interactions (Jenkins et al. 2006) (Section 4.6.4). K1590 is located at the N-terminus of the module, one residue before the first Cys (C1591) residue in CCP 25. Substitution with Gly and Glu (respectively) will result in drastic change in electrostatic properties (a net loss of two positively charged residues and a gain of a negative charge) in this region as seen in the modelled variant form (Figure 4.11b). Additionally, any change in the linker residue K1590 could alter the junction orientation between CCPs 24 and 25. This suggests further mutagenesis experiments for understanding the CR1-PfEMP1 interaction that would target surrounding surface residues.

Even though increased frequency of phenotypes in African populations suggest that epitopes in CCP 25 protect against severe malaria, the interaction site between CR1 and PfEMP1 has been mapped directly to functional site 2 – *i.e.* CCPs 8-10 and CCPs 15-17 of CR1, with modules 10 and 17 playing a major role (Rowe et al. 2000). How then do mutations in module 25, affect the interaction of CR1 and PfEMP1 (which lie in CCPs 15-17)? A possible explanation lies in the potential flexibility of the 30 module CR1 structure. It is seen in crystal structures of the intact, shorter, RCA proteins DAF (Lukacik et al. 2004) and VCP (Murthy et al. 2001; Ganesh et al. 2004; Ganesh et al. 2005) that their four CCP modules are arranged linearly in a head-to-tail fashion, with not much flexibility between neighbouring modules (Chapter 2) - their short linker lengths (3 to 4 amino acids) preclude that. On the other hand, however, the existence of a longer eight residue linker between CCPs 1 and 2 in another complement receptor, CR2, conferred on the module pair a unique side-by-side arrangement (Chapter 2) with respect to each other, resulting in a "V-shaped" arrangement seen in the crystal structure (Szakonyi et al. 2001; Prota et al.

2002). This arrangement between two adjacent modules buries the largest surface area for any known module pair (Chapter 2). Other low resolution studies on factor H (Aslam and Perkins 2001), also suggested a "bent-back-upon-itself" folded protein, made possible through the longer linkers present between modules 10 and 15 (6 to 8 amino acids). It is thus plausible that in CR1, CCPs 17 and 25 are brought close in 3D-space facilitated by the long eight-residue linker (longest for any neighbouring pair of CCP modules; see Chapter 2) between modules 21 and 22 (Figure 4.12), ideally located four modules away from CCP 17 in site 2 and three modules prior to CCP 25.

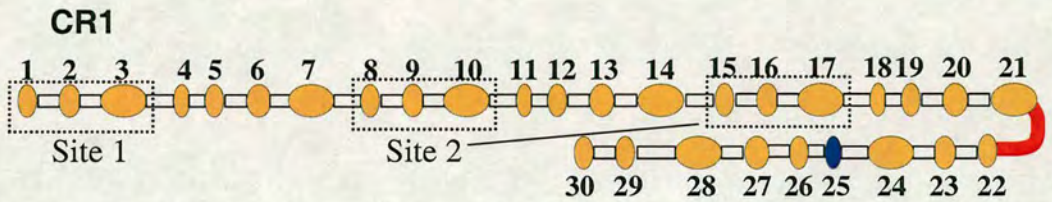


Figure 4.12: Schematic to illustrate module and linker lengths in CR1. This figure is drawn on a distorted scale to emphasise the variation in length among CR1 CCP modules (56-67 residues) and of the linkers (3-8 residues) between them. Figure adapted from chapter 2. For convenience the first and last residues of a module are defined as Cys ^{I} and Cys ^{IV} of the consensus sequence, and linkers are defined as the sequences lying between Cys ^{IV} of the N-terminal module and Cys ^{I} of the C-terminal module. Modules are numbered and functional sites 1 and 2 labelled. Site 2 binds PfEMP1. CCP 25, which contains the antigen seen in some African-specific populations, is coloured blue, and the eight residue linker length between CCPs 21 and 22 proposed to contain the bend coloured red.

However, this conformation cannot explain how CCP 25 can be brought into close proximity with the other copy of Site 2 (*i.e.* CCPs 8-10) as to be able to effect an interaction? The answer could lie in an "avidity" effect as seen in the case of another CCP-containing protein, β2GPI (Cucnik et al. 2004a; Cucnik et al. 2004b; Bozic et al. 2005; Cucnik et al. 2005).

4.3.4 Insights from the model of the first CCP module of human SRPX2

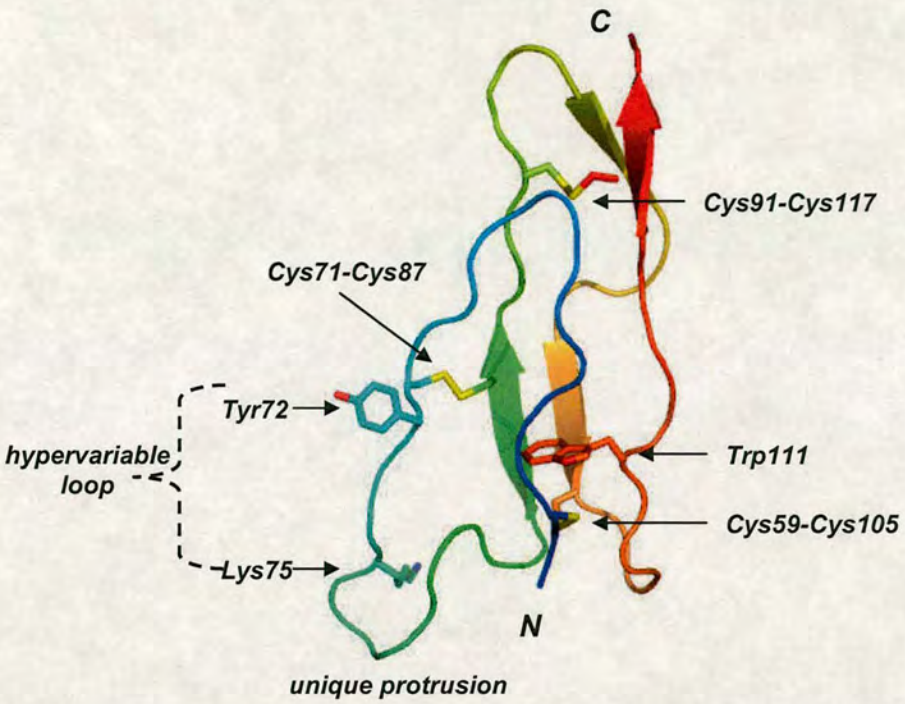


Figure 4.13: 3D-model of SRPX2, CCP 1. Cartoon representation of the model of CCP1 of SRPX2 coloured from blue to red (N- to C-terminus). Secondary structure elements have been assigned by default settings in PyMol (http://www.pymol.org). The location of the side-chains of the three putative disulfide bonds, conserved tryptophan, Y72 and K75 are indicated as stick-representations and labelled.

The model of the first CCP module of human SRPX2 confirms an additional putative disulfide bridge (Cys71-Cys87), atypical of the classical CCP-module fold (Figure 4.13). As discussed in previous chapters, among all CCP modules, a loop, highly variable in length, sequence and conformation called the hypervariable loop is present between the typically occurring strands 2 and 3 (according to the Wiles *et al.* convention for numbering strands; Chapter 2, Section 2.5) (Wiles et al. 1997). Depending on its length, this projects laterally from the module and forms an obvious candidate surface for protein-protein recognition. Indeed, the hypervariable loop has been implicated as a "hot-spot" for several protein-protein interactions and disease-causing SNPs in other CCP-containing complement proteins (Richards et al. 2003;

Herbert et al. 2006a; Herbert et al. 2006b; Jenkins et al. 2006). Cys71 lies at the beginning of the hypervariable loop that is exceptionally long in CCP1. The hypervariable loop is thus forced to form a prominent protruding extension towards the N-terminus of the module (Figure 4.13). This feature has not been seen in any experimentally determined CCP module to date, nor is predicted to occur in other *D*-cluster members (Weyer et al. 2004; Soares et al. 2005) (except SRPX, which contain the same number of amino acids between the two additional cysteine residues). It is noteworthy, that the stretch of sequence between Cys59 and Cys71 does not contain any proline residues, and could correspond to a short helical segment, as seen in a percentage of module structures in the ensemble of $GABA_BR1\alpha$ CCP2 (Blein et al. 2004).

The Y72S disease-causing mutation (Roll et al. 2006) is largely solvent exposed and located within the hypervariable loop, adjacent to Cys71 that participates in the non-typical, third, disulfide (Figure 4.13). This change from a large aromatic side-chain to a small, polar one at position 72 will have a profound effect on the surface properties of this region that is close to the aforementioned prominent protrusion – so an effect on function of such a mutation is plausible. The R75K mutation is located nearby, within the protrusion, and its side-chain is exposed (Figure 4.13). It is reasonable to suggest that the unique structural feature formed by the hypervariable loop of CCP1 performs some role that is specific to the protein. Presumably, it is not a coincidence that it is also the site of a human-specific change. An R75K substitution is a conservative one; the substitution of one exposed, positively charged residue for another can easily be accommodated by small atom shifts in surrounding side-chains, and is not likely to affect the structure of the CCP module. Such a change might have a small but not a dramatic functional effect.

Hence it fits well with positive selection, according to the neo-Darwinian view of evolution, as it would lead to a small step rather than to a giant leap (Wyckoff et al. 2000). Moreover, conservative R-to-K and K-to-R substitutions can result in the altered properties of either secreted proteins (Higuchi et al. 1991) or the extracellular domains of some plasma membrane proteins (Vinson et al. 1996; Rodien et al. 1998), including a member of the selectin family (Erbe et al. 1992).

There exists an *N*-terminal 34 amino acid sequence prior to the first CCP module (Figure 4.3a). Its conformation and location with respect to K75 and Y72 is speculative and hence cannot be commented upon.

4.4 Concluding remarks

The utility of the homology-based models for interpretation of various evolutionary and disease-associated polymorphisms is evident from this exercise. Another example where application of the created here models have proven useful, but in work by others, was in site-directed mutagenesis experiments guided by the models in delineation of the binding site (functional site 1) in CR1 (Krych-Goldberg et al. 2005). More recently, the individual CCP-module models from CR2 deposited on the website were used to discern the potential overall arrangement of its 15 CCP modules in solution, using a combination of X-ray scattering, and analytical ultracentrifugation (Gilbert et al. 2006b). Several potentially interesting cases for use of models among other non-complement CCP-module-containing proteins remain. An example is the interaction of the sole CCP module in neurocan with a glycosaminoglycan and its consequences for L1-binding which was previously reported (Oleszewski et al. 2000). A surface representation of its CCP-module model revealed a large basic patch, with intriguing possibilities for interactions with GAGs. Another demonstration of model-application could involve the prediction of complement regulatory sites within the giant protein CSMD1 in rat (Kraus et al. 2006) – its human homologue has the same domain architecture.

4.5 Declarations and Acknowledgements

Sections of the work described in this chapter have been published in:

1. Herbert, A.P.*, **Soares, D.C.***, Pangburn, M.K., and Barlow, P.N. 2006. Disease-associated sequence variations in factor H: a structural biology approach. *Adv Exp Med Biol* 586**:** 313-327. (* co-first authors).

2. **Soares, D.C.**, Gerloff, D.L., Syme, N.R., Coulson, A.F., Parkinson, J., and Barlow, P.N. 2005. Large-scale modelling as a route to multiple surface comparisons of the CCP module family. *Protein Eng Des Sel* 18: 379-388. Epub 2005 Jun 2023.

3. Royer, B., **Soares, D.C.**, Barlow, P.N., Bontrop, R., Roll, P., Robaglia-Schlupp, A., Blancher, A., Cau, P., Pontarotti, P., Szepetowski, P. Molecular evolution of the human *SRPX2* gene that causes brain disorders of the rolandic and sylvian speech areas. (*submitted).*

Chapter 5

Structural and functional insights from models of C5 and C5b: insight into the latter stages of complement assembly

---

5.1 Introduction

5.1.1 Background

Central to the complement pathway are three large, paralogous, plasma proteins C3, C4 and C5. How the clear functional distinctions amongst this set of paralogues relate to their differences at the structural level is not known. Structural knowledge of these large proteins had, until recently, been limited to the ~8-kDa C3a (Huber et al. 1980) and C5a (Zhang et al. 1997) fragments, the ~38-kDa C3d (Nagar et al. 1998; Szakonyi et al. 2001) and C4d (van den Elsen et al. 2002) fragments (that are cleaved from C3b and C4b respectively), and the ~18-kDa C345C domain of C5 (Bramham et al. 2005b). Even though crystallisation of C5 was reported (Discipio et al. 1998), no structure has been forthcoming. In the then absence of complete 3D structures of any of C3, C4 or C5, some previous attempts to assign specific amino acid residues to the various protein-protein interaction sites were based on sequence comparisons. This approach focussed on the identification of regions of insertion or deletion – *i.e.* "indels" – between C3, C4 and C5 (Section 5.3.3). In addition, several binding sites on C3b for factor B, properdin (Daoudaki et al. 1988) and the RCAs (factor H, CR1)

(Lambris et al. 1988; Oran and Isenman 1999) were suggested on the basis of mutagenesis and peptide studies. Recently, hydrogen/deuterium exchange and mass spectrometry was used to map regions of conformational change that occur upon transition of C3 to C3(H_20) (Winters et al. 2005).

The recent landmark publications of the 3D-structures of human C3 and C3b (Janssen et al. 2005; Abdul Ajees et al. 2006; Janssen et al. 2006; Wiesmann et al. 2006) now shed considerable extra light on these previous peptide-based and mutagenesis results. In particular the new structures provide the opportunity to construct 3D structural models of C5 and C5b that can be used to reanalyse many of the previous functional studies of this protein. C3, C4 and C5 have similarities in their subunit and precursor composition, protease sensitivities, and other properties, which suggest that they possess very similar 3D structures despite having distinct functions and binding specificities (Low et al. 1999). Thus, even though complement C5 lacks the thioester moiety, it might adopt a C3-like overall native fold. This notion is borne out by experiments where mutation of the thioester moiety in C3 resulted in it retaining its native-like C3 conformation (Isaac et al. 1998). Various low-resolution experimental evidence for C5 (DiScipio et al. 1983), and other modelling-based lines of evidence (Janssen et al. 2005; Fredslund et al. 2006; Janssen et al. 2006) suggest that even the more distant homologues, host-defence-like α2-macroglobulin (α2M) proteins, exhibit similar ultrastructure, domain arrangements, and conformations to those of C3. Thus, the C3 and C3b structures would appear to represent good templates for homology modelling of C5 and C5b.

In this chapter, the modelling by homology of C5 and C5b is described along with the structural and functional insight afforded by the models. Note that homology

models of C4 and C4b have also been created, although this are not discussed here. All the models (C4, C4b and C5, C5b) are available online.

5.1.2 The α2-macroglobulin (α2M) family

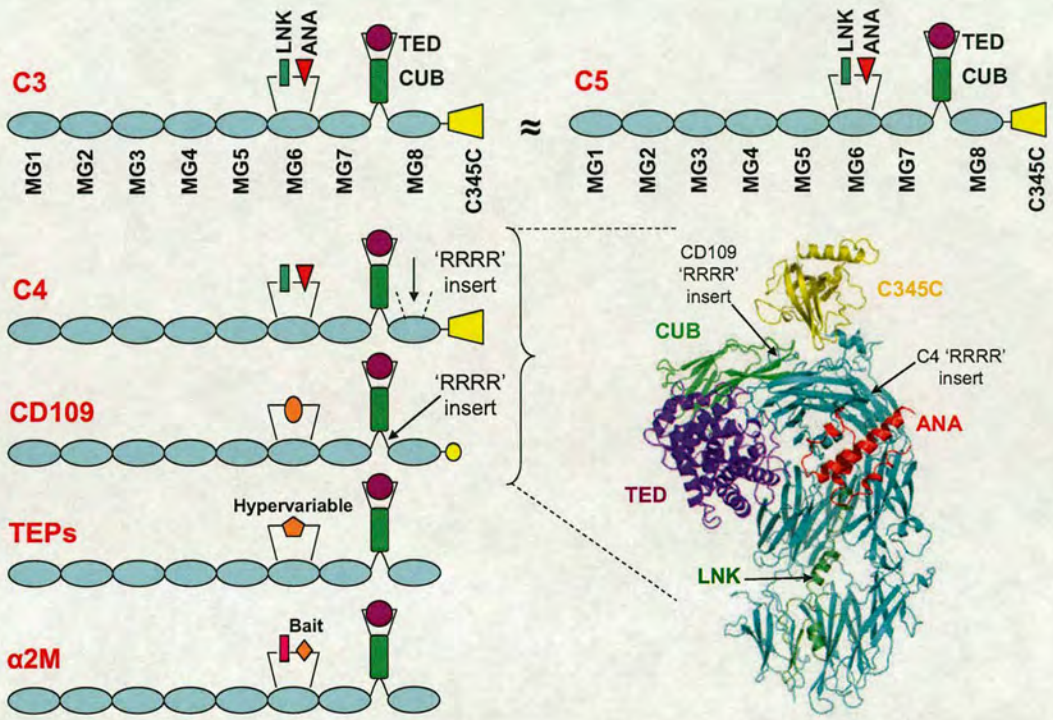


Figure 5.1: Domain organisation of the α2M family deduced from the C3 structure. Members of the α2-macroglobulin (α2M) family are shown schematically; TEPs = thioester-containing proteins; other abbreviations as in Table 5.1. The family is characterised by homologous sequence features, including a unique thioester motif "GCGEQ" (absent in C5 and some TEPs (Blandin and Levashina 2004)). The domain organisation of C3 and C5 are identical. The differences among the remaining members are highlighted. Native C4 has an additional tetra-arginine (RRRR) processing site that occurs in an insert of 46 residues (within MG8, see alignment in Figure 5.4). CD109 (Solomon et al. 2004) has a tetra-arginine site in the linker between CUB and MG8. Only C3, C4 and C5 have the C345C domain since α2M and TEPs lack this domain, and CD109 has a GPI-anchor instead. Figure created on the basis of Janssen et al. (2005).

Complement component C3, C4, and C5 belong to the α2-macroglobulin (α2M) family, whose other members include the proteinase inhibitor α2M, the insect and nematode thioester-containing proteins (TEPs), and CD109 (Figure 5.1). These

proteins (~1400 – 1800 amino acid-residues) are characterised by homologous sequence features, including a unique thioester motif (not present in C5 and some TEPs); and they have important roles in the immune response in metazoans, predating the emergence of immunoglobulins (Budd et al. 2004).

5.1.3 Structural insights gained from C3, C3b and C3c

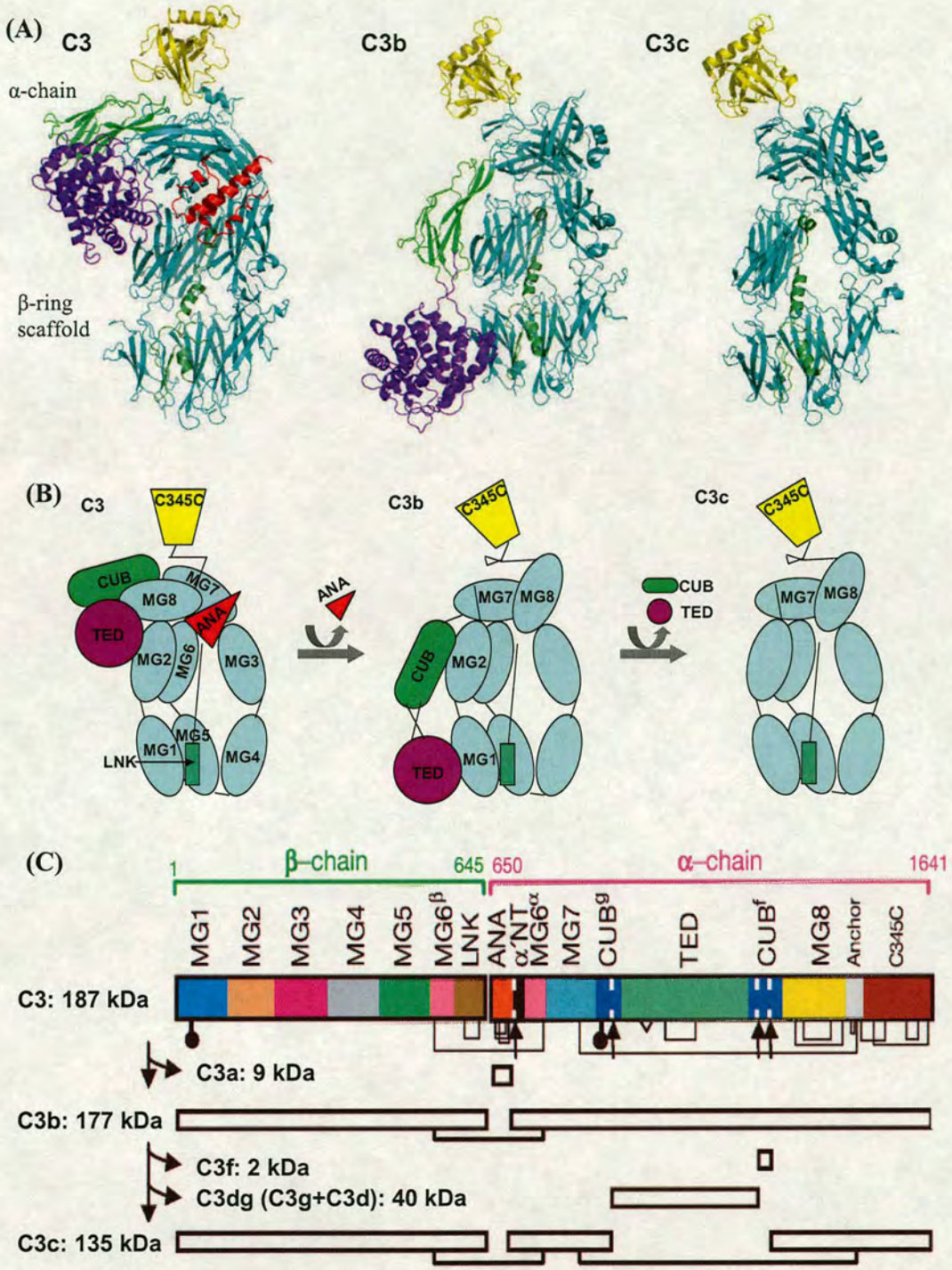


Figure 5.2: Structures of human C3, C3b and C3c. **(a)** Cartoon representations of native C3 (13 domains) (PDB ID: 2A73) (left panel) (Janssen et al. 2005), C3b (12 domains) (PDB ID: 2I07) (middle panel) (Janssen et al. 2006), C3c (10 domains) (PDB ID: 2A74) (right panel) (Janssen et al. 2005) are shown and domains coloured [MG-domains (cyan), ANA (red), LNK (dark green), TED (purple), CUB (light green), C345C (yellow)]. **(b)** The schematic domain arrangements in C3, C3b and C3c are

shown. C3 was found to contain multiple immunoglobulin-like domains that have been called macroglobulin (MG) domains. Upon cleavage, C3 undergoes conformational changes to C3b and C3c. Domains in C3b and C3c exhibit almost identical conformations with each other, when compared with C3 (Wiesmann et al. 2006). **(c)** Shown on the linear sequence/domain arrangement is the thioester site (white triangle) in TED, disulfide bridges, glycan positions and cleavage sites (see text). Sequential proteolysis from C3 → C3b → C3c is indicated. MG6 intertwines the β- and α-chain of mature C3; while CUB is also formed from intertwined regions (named CUB^{g} and CUB^{f}). Figures created on the basis of (Janssen et al. 2005; Janssen et al. 2006).

Domain	PDB Chain	Domain residue range
MG1	A	1 – 104
MG2	A	105 – 209
MG3	A	210 – 328
MG4	A	329 – 426
MG5	A	427 – 534
$MG6^{\beta}$	A	535 – 577
LNK	A	578 – 645
ANA	B	650 – 726
α'NT	B	727 – 745
$MG6^{\alpha}$	B	746 – 806
MG7	B	807 – 911
CUB^{g}	B	912 – 962
TED	B	963 – 1268
CUB^{f}	B	1269 – 1330
MG8	B	1331 – 1474
Anchor	B	1475 – 1495
C345C	B	1496 – 1641

Table 5.1: C3 domains and boundaries. Domain lengths for C3 as deduced by (Janssen et al. 2005) from the structure (PDB ID: 2A73) - Chain A: 1-645, Chain B: 650-1641 (Figure 5.2). Abbreviations used in table: MG = macroglobulin; LNK = linker; ANA = anaphylatoxin; NT = N-terminal; CUB = domain first found in C1r, C1s, uEGF, and bone morphogenetic protein; TED = thioester-containing domain; C345C = domain present in human complement components C3, C4 and C5.

The 3D structures of C3, C3b and C3c were recently determined (Janssen et al. 2005; Abdul Ajees et al. 2006; Janssen et al. 2006; Wiesmann et al. 2006), providing the first insights into the native overall arrangement of α2M family members. The two chains of C3 - a β-chain made up from residues 1-645, and an α-chain made up of residues 650-1641 - together form 13 domains. In C3b, 12 domains are present within a complete β-chain and an α-chain that lacks the ANA domain (ANA gets cleaved off as C3a). C3c on the other hand consists of three chains; the β-chain, and two fragments of the α-chain (TED and CUB are cleaved off from C3b) that form ten domains (Figure 5.2).

From the C3 structure (Janssen et al. 2005) it can be seen that one of the domains is formed by parts of both the β- and α-chains (Figure 5.2, Table 5.1). In addition, each chain forms six domains by itself. Out of these, eight domains (5.5 of β and 2.5 of α) show a fibronectin-type-3-like (immunoglobulin-like) core fold, even though there is no sequence homology apparent amongst them. These domains were termed macroglobulin (MG) domains and form a β-ring like scaffold. Residues 1-534 form five MG domains (MG1-MG5) and residues 535-577 form one half, denoted $MG6^{\beta}$, of a β-chain/α-chain intertwined MG6 domain. Finally, residues 578-645 of the β-chain form a linker domain, termed LNK. The α-chain begins with (residues 650-726) the anaphylatoxin (ANA) domain (which when cleaved off forms the C3a fragment). Loop 727-745 connects ANA to MG6, and forms the N-terminal region of the cleaved (in C3b) α-chain, α'NT. Residues 746-806 form $MG6^{\alpha}$, which complements $MG6^{\beta}$. This is followed by MG7 (residues 807-911). The polypeptide chain that may be excised by factor I to yield the 40-kd fragment C3dg and 2-kd fragment C3f is located between MG7 and MG8 (residues 1331-1474). The cleavage sites that yield C3g, C3d and C3f (Harrison and Lachmann 1980; Lachmann et al.

1982) do not fully coincide with domain boundaries. Residues 912-962 and 1269-1330, *i.e.* 63% of C3g and all of C3f, contribute two halves (approximately) of one CUB domain and have been named CUB^g and CUB^f, respectively. The remaining parts of C3g and C3d (residues 963-1268) together form the thioester-containing domain (TED). Finally, residues 1496-1641 form the carboxy-terminal C345C domain. This domain is covalently linked to MG8 by the polypeptide chain and to MG7 by a disulfide bond (Thomas et al. 1982), in what was called an anchor region (residues 1475-1495). In total, nine of the thirteen domains were unpredicted (MG1-MG7, LNK and CUB). All the domains together formed an irregular shaped C3.

The structures of C3b (Abdul Ajees et al. 2006; Janssen et al. 2006; Wiesmann et al. 2006), revealed the conformational changes that accompany loss of the ANA domain to expose multiple-binding sites for factor B and RCA proteins previously hidden in intact native C3. In these structures, the thioester in TED is completely exposed for covalent attachment to target surfaces. It lies ~80 Å away from the buried site in native C3, since the TED has been relocated to the bottom of the MG-domain β-ring (Figure 5.2). Significant further domain rearrangements (that result in loss and gain of domain-domain interactions in C3b) accompany activation particularly in the α-chain; while the β-ring like scaffold remains almost intact (Figures 5.2a, 5.2b). In one of the published C3b structures (Abdul Ajees et al. 2006), however, there occurs a marked loss of secondary structure and unfolding of the CUB domain (not shown). Following cleavage of CUB and TED, to form the C3c fragment structure (Janssen et al. 2005), no major additional conformational changes are seen (Figure 5.2b), indicating that the main conformational rearrangements occur in the first activation step from C3 to C3b. Apart from the relocation of TED, MG7 and MG8 almost swap places in the overall structure; the C345C domain swivels 32° and

moves 10.1 Å; and the anchor region alters its conformation from an α-helix in C3 to a β-hairpin in C3b and C3c (Janssen et al. 2005).

5.2 Materials and Methods

5.2.1 Template selection and sequence alignments

Three-dimensional modelling of the amino acid sequences of human C5 (SWISS-PROT entry: CO5_HUMAN), was based upon two templates: the intact native human C3 crystal structure [PDB ID: 2A73 (Janssen et al. 2005)]; and the solution structure of the closest-to-mean C5-C345C domain, model 13 from PDB ID: 1XWE (Bramham et al. 2005b). The experimentally solved structure of isolated C5a (Zuiderweg et al. 1989) was not employed on the grounds that the structure of its equivalent, free C3a, did not match to the structure of C3a in the context of intact C3. The C345C domain was inferred to be a more independent unit, exhibiting few interactions with other domains in C3.

For modelling C5b, the crystal structure template PDB ID: 2I07 (Janssen et al. 2006) of unliganded C3b was employed, along with the C5-C345C domain. This unliganded C3b structure, was solved at higher resolution (PDB ID: 2I07; 4 Å) than the liganded structure determined by Wiesmann *et al.* (PDB ID: 2ICF; 4.1 Å) and was hence the preferred template. Both these C3b structures exhibit almost identical conformations (Nishida et al. 2006). The aberrant C3b structure, PDB ID: 2HR0 (Abdul Ajees et al. 2006), which exhibits a different conformation to 2I07 and 2ICF, was ignored for purposes of the modelling work.

The optimal target-template alignments were based upon initial multiple sequence alignments from Ogata and Low, 1998; Levashina et al., 2001 and Bramham et al., 2005 (Ogata and Low 1997; Levashina et al. 2001; Bramham et al. 2005b). The alignment was subjected to further manual editing, guided by secondary structure as defined on the basis of the C3 crystal structure (Janssen et al. 2005) or

predicted by PsiPred (McGuffin et al. 2000) for the target C5 sequence, in order to place conserved amino acid residues, and indels, most appropriately.

5.2.2 Comparative modelling, evaluation and data analysis

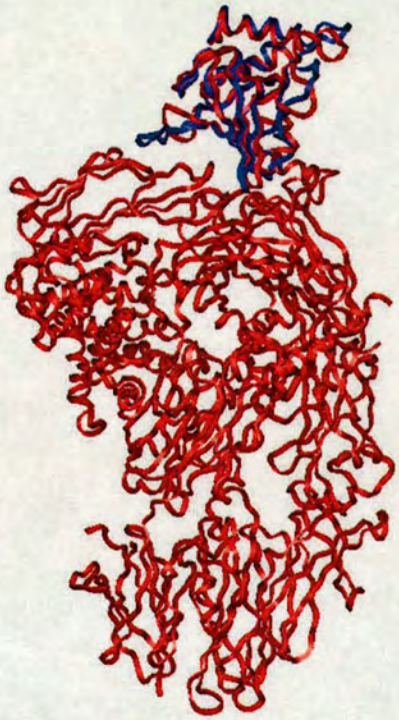


Figure 5.3: Overlay of templates for model building. The experimentally determined C5-C345C domain (blue) superposed on C3-C345C (red) using all equivalent secondary structure elements under INSIGHTII (Accelrys Inc., San Diego, CA, USA). This overlay ensures that during model building, the same inter-domain orientation is preserved in the model of C5. The C5-C345C domain is modelled upon itself and reflects the original experimentally solved structure.

For modelling of C5, the previously solved solution structure of the recombinantly expressed C5-C345C domain was first superposed on C3-C345C (Figure 5.3), as it appeared in the crystal structures of C3, using all equivalent elements of secondary structure, under INSIGHT II (Accelrys Inc., San Diego, CA, USA). An equivalent operation was performed for modelling of C5b. This ensures that the relative orientation of this domain as observed in C3 (or C3b) is potentially preserved (fixed)

during subsequent model building of C5 or C5b. The C5-C345C (C5b-C345C) domain structure was then simply modelled upon its self, *i.e.* the original NMR-derived solution structure. The appropriate inputs, including the PDB template files and the optimal target-template alignments, were provided to the comparative modelling program, Modeller release 8 version 1 (Sali and Blundell 1993).

The twenty-six cysteine residues conserved between C3 and C5 that formed thirteen disulfide bridges in C3 were automatically restrained and created for C5 during model building. Twenty models were generated, and the one with the lowest *objective function score* (Sali and Blundell 1993) selected as the representative model. It was noticed in this representative 3D-model that, apart from the expected thirteen disulfide bridges (as occur in the C3 structure used as a template), two additional cysteine residues (Cys838 and Cys865) occur close in space on neighbouring antiparallel β-strands within the MG7 domain. The orientations of all side-chains in the representative model of the target, that are not identical to equivalents in the template structure (including the two abovementioned cysteine residues) were optimised using the side-chain replacement program SCWRL version 3 (Bower et al. 1997; Canutescu et al. 2003). This automatically detected and created an additional disulfide bridge in the C5 model.

The model of C5 was then protonated under SYBYL version 6.9 (Tripos Associates, St. Louis, MO, USA), and subject to brief energy minimisation employing the Tripos forcefield (Clark et al. 1989) using 100 steps of the Powell algorithm (the C5-C345C domain residues 1514-1658 were excluded from this) to remove clashes and bad geometries. The model structure was finally checked for valid stereochemistry using PROCHECK version 3.5.4 (Laskowski et al. 1993) (Table 5.2a). Additionally, the packing quality of individual domains was checked using

WHATIF (http://swift.cmbi.kun.nl/WIWWWI/) (Vriend 1990; Vriend and Sander 1993) (Table 5.2b). An analogous set of procedures was employed to build and evaluate the model of C5b.

(A)	Most favoured regions	Additional allowed regions	Generously allowed regions	Disallowed regions
C5 model	83.1%	14.1%	1.8%	1%
C5b model	83.5%	13.2%	2%	1.2%

(B) Domain	Average packing quality control score in C5 (C3)	Average packing quality control score in C5b (C3b)
ANA	-1.667 (-1.946)	Absent in C5b
MG1	-2.364 (-1.856)	-1.962 (-0.895)
MG2	-1.476 (-0.970)	-1.258 (-0.626)
MG3	-1.221 (-1.042)	-0.991 (-0.627)
MG4	-2.117 (-1.862)	-2.095 (-1.535)
MG5	-1.746 (-1.352)	-1.558 (-0.861)
MG6	-1.742 (-1.311)	-1.498 (-1.106)
MG7	-1.278 (-0.786)	-1.423 (-0.955)
MG8	-1.207 (-1.090)	-1.585 (-1.162)
CUB	-2.000 (-1.209)	-1.959 (-1.416)
LNK	-2.538 (-2.527)	-2.336 (-2.407)
TED	-1.461 (-0.936)	-1.333 (-1.054)
C345C	-1.324 (-0.889)	-1.354 (-0.972)

Table 5.2: Quality analysis for models. (a) The PROCHECK Ramachandran plot statistics are given for the models of C5 and C5b. **(b)** The average packing quality score for each domain in the models of C5 and C5b, and the corresponding templates (in brackets), are reported. To place the values in context, a molecule is certain to be incorrect if the score is less than -3.0. Poorly refined molecules, extensively energy minimised misthreaded molecules, and low resolution homology models give values between -2.0 and -3.0. The average quality control score obtained by 200 highly refined X-ray structures was -0.5 ± 0.4 (Vriend 1990). The two halves of MG6 and CUB were evaluated together.

Electrostatic surface representations were created using GRASP (Nicholls et al. 1991). Lipophilic surface representations were created using MOLCAD (Heiden et al. 1993).

5.2.3 Domain-domain interaction identification and analysis

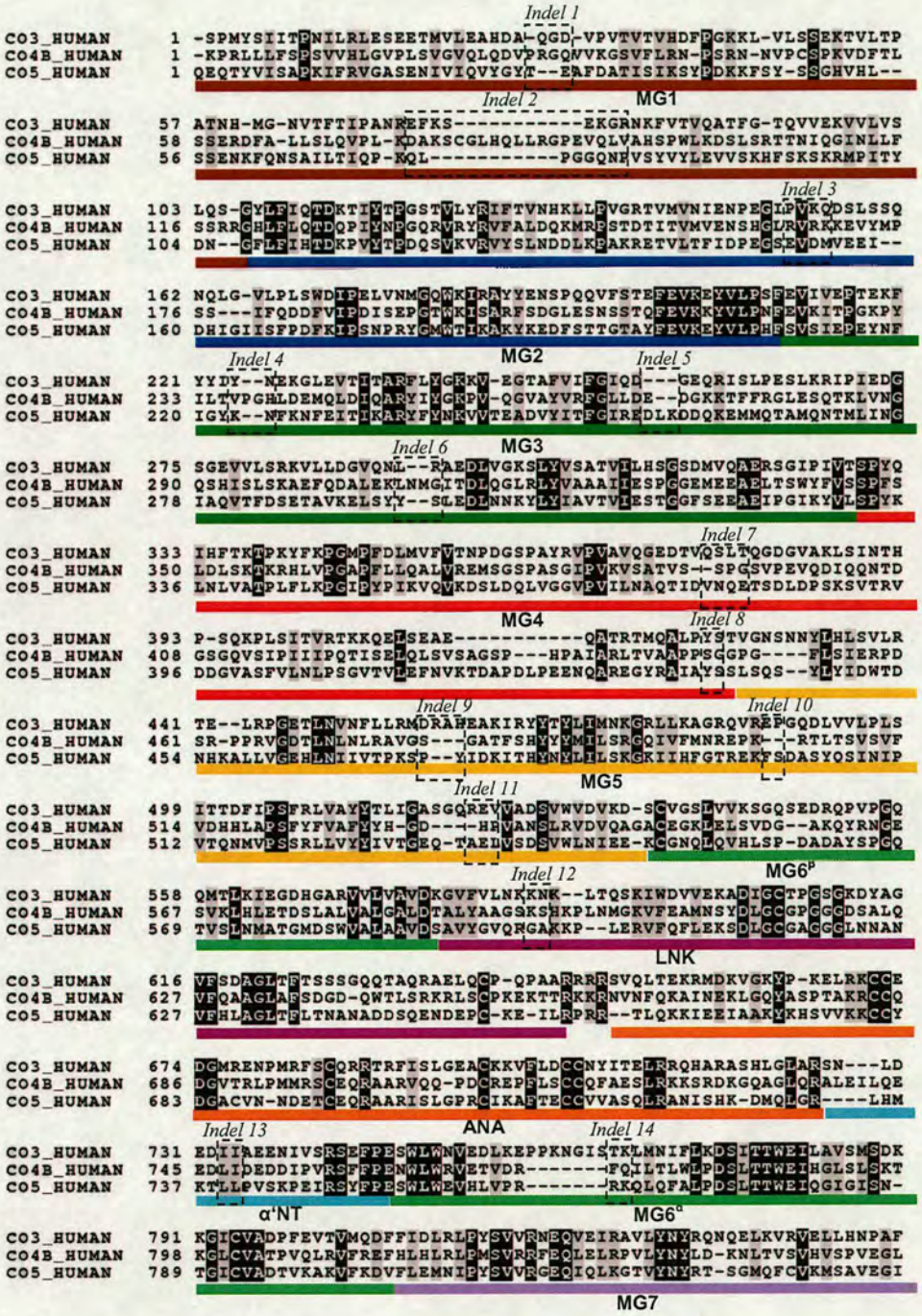
Domain-domain interactions were compared between the crystal structure of C3/C3b and the modelled structures of C5/C5b. First, domain-domain interactions were identified for both the template and model structures using the program, Structural Classification of Protein-Protein Interfaces (SCOPPI) (Winter et al. 2006). Identified interacting domains were subsequently extracted individually, saved into separate PDB files, and then converted into their corresponding PQR format employing the PDB2PQR version 1.0.1 server (Dolinsky et al. 2004) (http://nbcr.sdsc.edu/pdb2pqr/) using default parameters. PQR files are simply PDB files where the occupancy and B-factor columns have been replaced by atom charge and radius. The PQR files for each interacting pair of domains were then submitted to the MolSurfer server (http://projects.villa-bosch.de/dbase/molsurfer/submit-elec.html) (Gabdoulline et al. 1999; 2003) and analysed using default parameters.

MolSurfer links the 3D structure of a protein-protein complex to a 2D projection map using graphical outputs. Interface properties, including electrostatic and hydrophobic interactions, are projected to the two, symmetrical, 2D interface maps, and cross-correlation coefficients of surface properties between these two interface maps are then calculated. Because electrostatic and hydrophobic interactions are considered to be major factors that stabilise complexes, Molsurfer is particularly helpful in exploring complementarities at macromolecular interfaces (Gabdoulline et al. 2003) - cross-correlation coefficients of the mapped interfaces can be computed for electrostatic and hydrophobic components; these are ideal for comparison. Thus MolSurfer was used to calculate the "electrostatic correlation coefficient" (ECC) and "hydrophobic correlation coefficient" (HCC) for each domain-domain interaction in the C5/C5b models, and likewise for the corresponding

interactions in C3/C3b. In addition, the buried surface area (BSA) was calculated for each pair using the GETAREA version 1.1 server (http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl) (Fraczkiewicz and Braun 1998). The BSA was calculated as: $[(SA^{domain1}) + (SA^{domain2})] - (SA^{domain12})$; where SA = surface area; $SA^{domain12}$ = surface area of combined domain-domain pair).

5.3 Results and Discussion

5.3.1 Domain composition and organisation of C5 and C5b



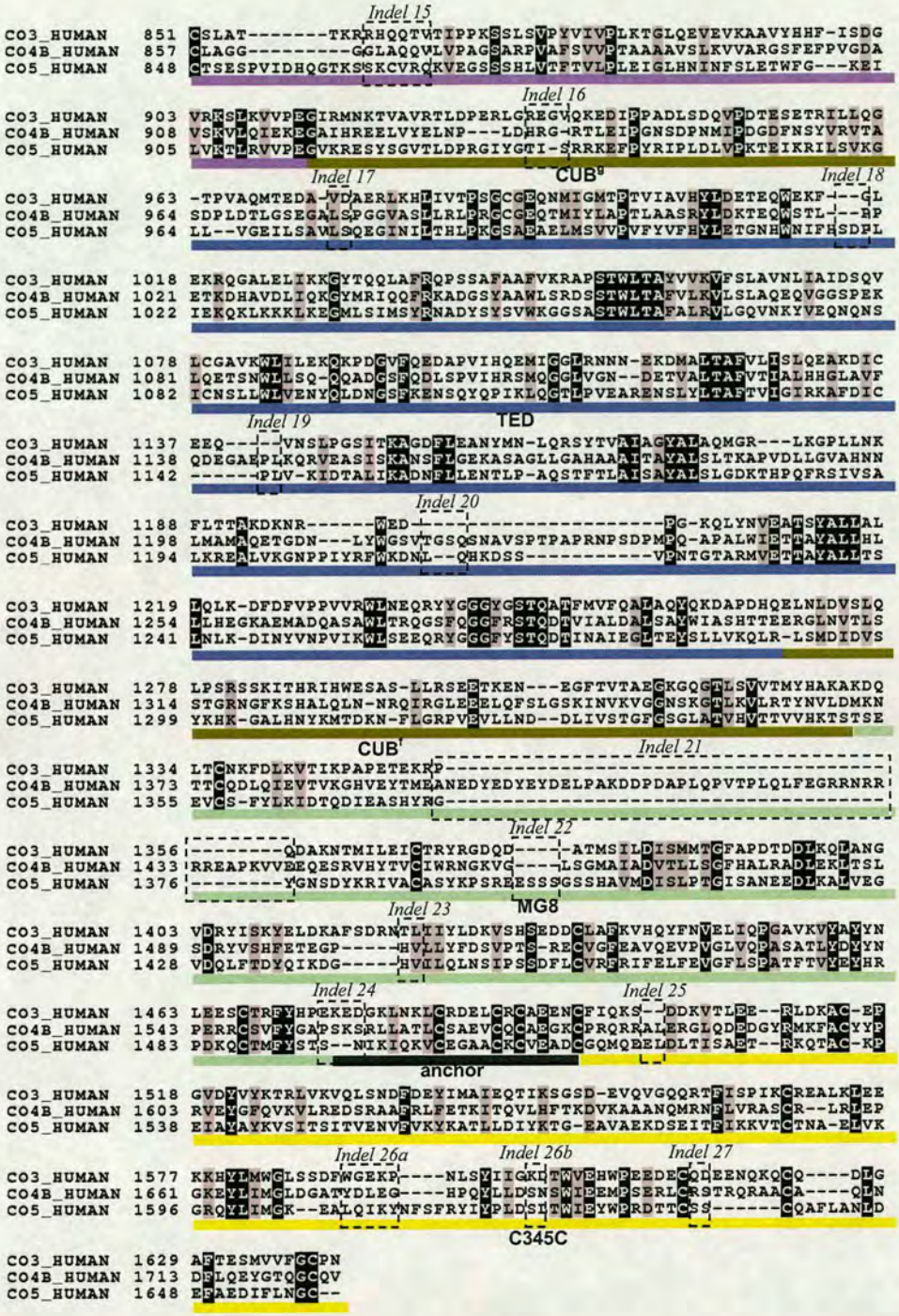


Figure 5.4: Sequence alignment between C3, C4, and C5. The optimal structure-guided sequence alignment between C3 (SwissProt entry name: CO3_HUMAN), C4 (SwissProt entry name: CO4B_HUMAN) and C5 (SwissProt entry name: CO5_HUMAN), as used for modelling, is shown using BOXSHADE version 3.21 (http://www.ch.embnet.org/software/BOX_form.html). It should be noted that two polymorphic C4 sequences exist - CO4A_HUMAN and CO4B_HUMAN (Yu et al. 1986); the latter is more common. Completely conserved (dark shading) and conservatively substituted (light shading) residues are depicted on the alignment; domain boundaries are also indicated and

labelled (Refer to text for abbreviations). The indels proposed on the basis of the original (pre-structure) alignment (Ogata and Low 1997; Ogata et al. 1998; Low et al. 1999) are also indicated.

Domain	Chain	C5 domain residue range	C5 vs. C3 domain-domain % sequence identity
MG1	A	1 – 105	18%
MG2	A	106 – 208	36%
MG3	A	209 – 331	32%
MG4	A	332 – 440	23%
MG5	A	441 – 546	30%
$MG6^{\beta}$	A	547 – 588	29%
LNK	A	589 – 655	31%
ANA	B	660 – 733	34%
α'NT	B	734 – 751	22%
$MG6^{\alpha}$	B	752 – 804	45%
MG7	B	805 – 913	29%
CUB^{g}	B	914 – 963	20%
TED	B	964 – 1290	29%
CUB^{f}	B	1291 – 1351	22%
MG8	B	1352 – 1494	25%
Anchor	B	1495 – 1513	32%
C345C	B	1514 – 1658	29%

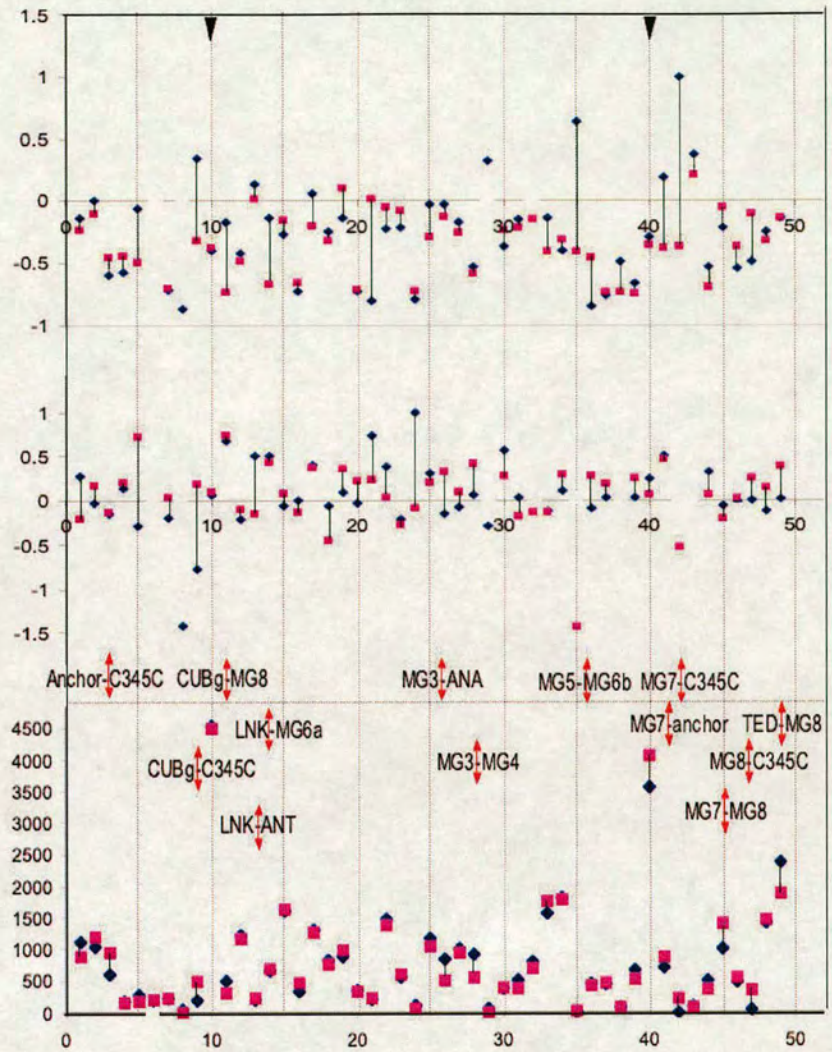
Table 5.3: C5 domain boundaries and comparison with C3. Domain lengths for C5 as determined from the alignment in Figure 5.4 - Chain A: 1-655, Chain B: 660-1658. The percentage sequence identity between individual domains in C5 and C3 was derived using the percentage identity matrix option under ClustalX (Thompson et al. 1997). The overall percentage sequence identity between C5 and C3 is 29%.

That the domain composition of C3 is conserved in C5 is evident from the sequence similarity between the two proteins (sequence identities at the domain-level vary from 18% in MG1 to 45% in $MG6^{\alpha}$) (Table 5.3). C5 consists of two chains, β (residues 1 – 655) and α (residues 660 – 1658) that, unsurprisingly, fold, with error-free packing

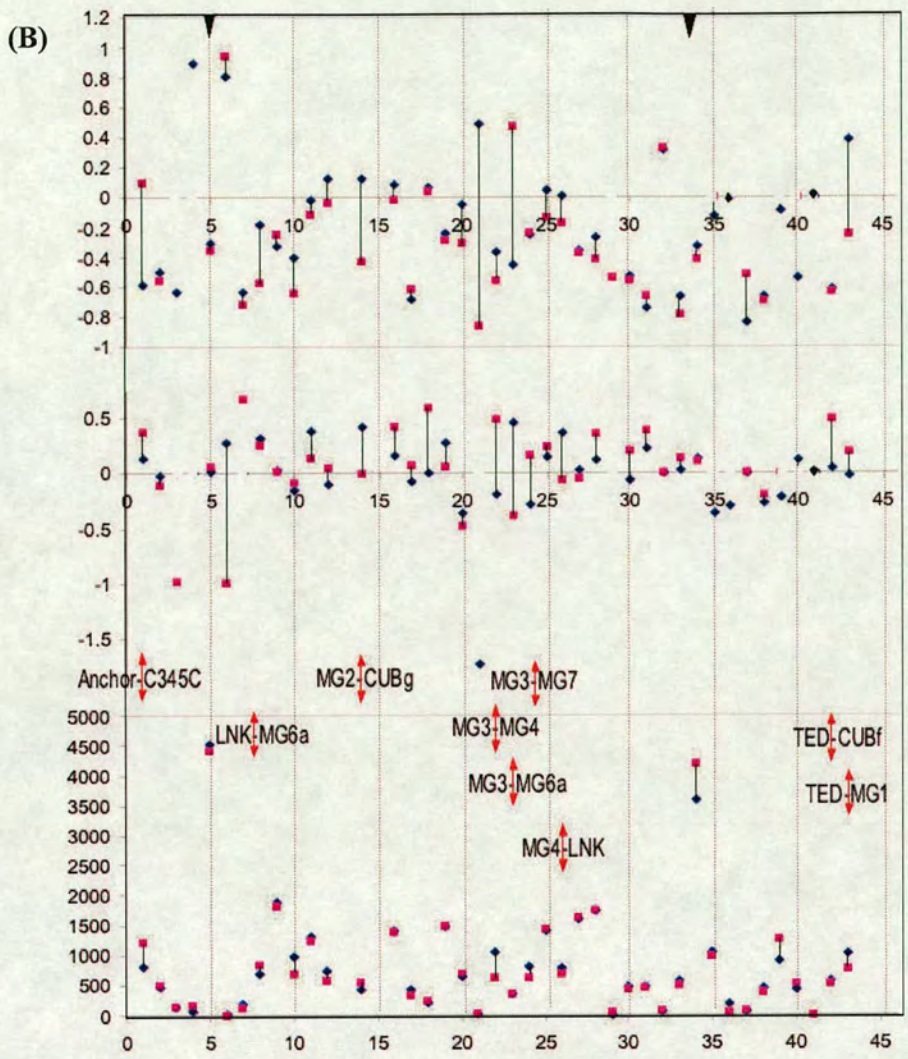
scores (Table 5.2b), into the same 13 domains in the modelled structure as were described for C3 (Table 5.1). The modelling protocol dictates a similar organisation of domains in the model to that of the template. It is noteworthy, however, that the crystal structures showed C3 and its major fragments C3b and C3c to significantly differ in the spatial configuration of the 10 component domains they have in common (Figure 5.2). This evidence for conformational versatility suggests that the spatial arrangement of domains within C5 could, in principal, differ from that of C3. To assess the likelihood of this possibility, and as an additional test of the quality and reliability of the new C5 model, the interfaces between the 13 domains of the modelled structure were compared to the equivalent domain-domain interfaces in C3 with respect to the extent of buried surface area, and conservation of favourable contacts (Figure 5.5).

As in C3, the β-chain is made up from five-and-a-half MG domains and a LNK domain (see Section 5.1.3). The α-chain is made up from the ANA domain (which upon cleavage forms the C5a anaphylatoxin domain), a short α'NT region, two-and-a-half MG domains, a TED and a CUB domain. Its C-terminus consists of an anchor region followed by the C345C domain. Thus, one MG domain (MG6) is formed from parts of both the α- and β-chains; critically, in the C5 model these two parts bury approximately the same large surface area and make contacts that are equally as favourable as those in the C3 crystal structure (Figure 5.5). The CUB domain in the α-chain is also composed from two parts, with the TED domain inserted between them; again, the two parts interact just as favourably as in C3. Many of the other interdomain interactions within the C5 structure likewise match their counterparts in C3, while the exceptions might reveal any differences in the mechanism of action of the two proteins (Figure 5.5).

(A)



	MG1	MG2	MG3	MG4	MG5	MG6β	LNK	ANA	α'NT	MG6α	MG7	CUBg	TED	CUBf	MG8	Anchor
MG1																
MG2	16															
MG3		20														
MG4			28													
MG5	17	21		33												
MG6β		23			36											
LNK	15	19		32	34	39										
ANA			26													
α'NT		18	27			38	13	1								
MG6α		22	29		35	40	14		4							
MG7			30							37						
CUBg											44					
TED		25										12				
CUBf											43	10	48			
MG8		24	31					2	5		45	11	49	7		
Anchor											41	8			46	
C345C											42	9		6	47	3



	MG1	MG2	MG3	MG4	MG5	MG6β	LNK	α'NT	MG6α	MG7	CUBg	TED	CUBf	MG8	Anchor
MG1															
MG2	10														
MG3		17													
MG4			22												
MG5	11	18		27											
MG6β		20			30										
LNK	9	16		26	28	33									
α'NT						32									
MG6α		19	23		29	34	8	2							
MG7			24						31						
CUBg		14								38					
TED	43										7				
CUBf		12								37	5	42			
MG8		21	25							39	6		3		
Anchor										35	4			40	
C345C										36				41	1

Figure 5.5: Domain interface analysis for C5, C3 and C5b, C3b. In each panel (A and B), the electrostatic correlation coefficient (ECC) (top panel), hydrophobic correlation coefficient (HCC) (middle panel) (Gabdoulline et al. 2003) and buried surface area (BSA) (lower panel) (Fraczkiewicz

and Braun 1998) for all interacting domain-domain pairs in **(a)** C5 (blue symbol) vs. C3 (pink rectangle) and **(b)** C5b (blue symbol) vs. C3b (pink rectangle) are plotted. ECC, HCC and BSA values are plotted on the *y*-axis, while domain-domain interacting pairs are shown on the *x*-axis according to the table numbers shown below the respective graphs. Some outliers (but only those with largish BSA) are labelled. Position for the two half domains of MG6 and CUB are also indicated with a solid black arrow at the top of the graph. Raw data may be viewed on Appendix CD, A.5.1 and A.5.2.

In some cases (see Figure 5.5) there are differences in one or other of buried surface area, HCC or ECC values, but in general these compensate for one another. Some major differences in buried surface area and energetically favoured contacts suggest that the arrangements of the anchor (32% identical) and C345C (29% identical) domains at the "top" of C3 are not necessarily emulated in C5 and therefore the model is possibly unreliable with respect to the precise positioning of these domains. It is interesting that these domains are amongst those rearranged upon cleavage of C3 to C3b. Overall, the analysis supports a very similar arrangement of domains in C5 to that in C3, in agreement with other experimental observations (DiScipio et al. 1983) as discussed in the Introduction (Section 5.1.1).

Turning to the C5b model, it is again apparent that most domain-domain interactions are as favourable as their counterparts in C3b. On the other hand, several of the new interdomain contacts that arise from domain rearrangements during the C5 to C5b transition appear less convincing. In comparison with the equivalent interface in C3b, the TED-MG1 interface has poorer electrostatic and hydrophobic contacts but only a slightly larger buried surface area. The nascent TED-CUB^{f} and MG3-MG7 interactions in C5b also appear weaker than those in C3b. Therefore, confidence in the model of C5b is lower and it is possible that the conformational transition upon activation of C5 does not emulate that observed crystallographically for C3 activation.

5.3.2 The model of C5 reveals a novel putative disulfide bridge

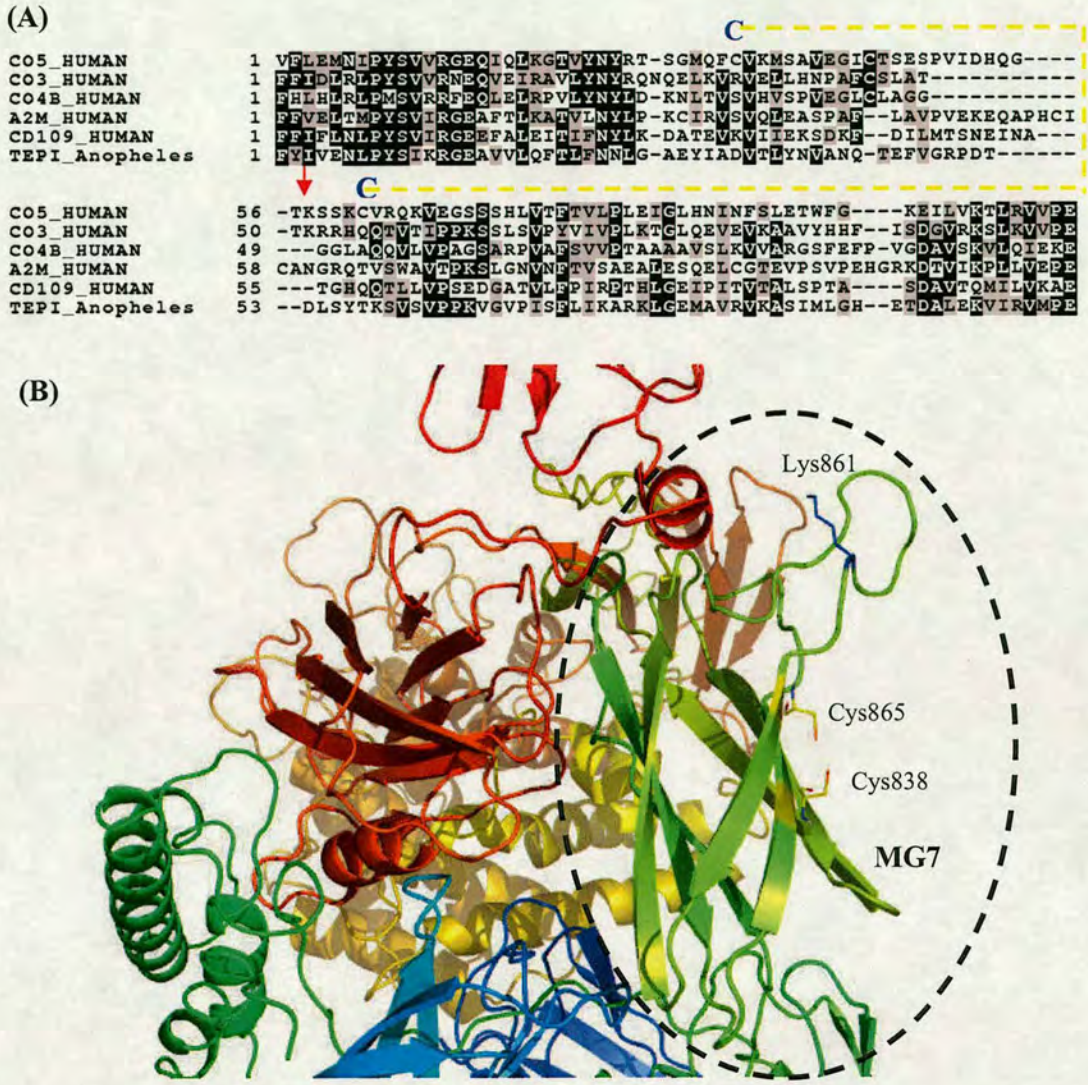


Figure 5.6: Novel disulfide bond in human C5. **(a)** Sequence alignment of MG7 among A2M family members. The MG7 domains from C5 (CO5_HUMAN), C3 (CO3_HUMAN), C4 (CO4_HUMAN), α2-macroglobulin (A2M_HUMAN), CD109 (CD109_HUMAN) from human and thioester-containing protein I (TEPI_Anopheles) from *Anopheles gambiae* are shown using BOXSHADE version 3.21 colouring (http://www.ch.embnet.org/software/BOX_form.html). The two cysteine residues in C5 involved in the extra disulfide bond are marked on the alignment (connected with a yellow dashed line), and the location of the loop containing the trypsin cleavage site (K861) in C5 is indicated (red arrow). **(b)** MG7 domain, highlighting the location of the C5-specific disulfide bond. The MG7 domain (marked within the dashed oval and labelled), depicting the location of the additional disulfide bond in C5 (Cys838-Cys865), on adjacent neighbouring antiparallel β-strands. Also indicated is the location of the trypsin cleavage site at position Lys861, whose side-chain (blue) is highly exposed.

The crystal structure of C3 revealed thirteen disulfide bridges - eleven in the α-chain, one in the β-chain, and one inter-chain disulfide in $MG6^{\beta/\alpha}$ – all of which are predicted by conservation/homology to occur in related components C4 and C5. The model of C5, however, revealed a fourteenth disulfide bridge, not obvious previously, formed by residues Cys838 and Cys865 in its MG7 domain (Figure 5.6). This disulfide was calculated and built by the side-chain replacement program, SCWRL (Bower et al. 1997; Canutescu et al. 2003), which confirmed that these Cys side-chains were in allowed rotamer conformations to form a disulfide bond. The measured Cα-Cα, and Sγ-Sγ atom distances between these Cys residues for the two models are as follows: C5-model: 6.55 Å, 3.10 Å; C5b-model: 6.61 Å, 3.11 Å. Disulfide bonds may be assumed to exist between two cysteine residues if their Sγ-Sγ atoms are < 3.5 Å apart (Cheek et al. 2006) (and Cα-Cα atoms are approximately < 8 Å).

The existence of this additional disulfide, not present in C3, C4 or other members of the α2M family (Figure 5.6a), raises several important questions. Was this disulfide bridge present in an ancestral protein but lost by each of the other A2M-family members, during evolution, due to a lack of importance? Or has C5 evolved a new disulfide because it is needed to fulfil an important specific role? It is interesting to note that a trypsin cleavage site (Lys 861) (Wetsel and Kolb 1982) lies in the loop connecting the two antiparallel β-strands that are held together by this disulfide bond. Additionally, it is worth pointing out absence of an otherwise conserved disulfide bond within the TED domain in C4 and the α2M-family member Anopheles-TEP (Callister, Soares et. al., unpublished). In TED the two cysteine residues involved are strictly conserved in C3d (Cys 1079, Cys 1136) for a range of species (mouse, chicken, cobra, trout, lamprey, and hagfish) (Nagar et al. 1998).

5.3.3 Location of indels and active sites mapped onto the models of C5 and C5b

Previous experimental studies on structure-function relationships in C5 focussed on indels – regions of deletion or insertion - between C3, C4 and C5 (Ogata and Low 1997; Ogata et al. 1998; Low et al. 1999). The authors reasoned that regions of a protein near indels were good candidates for protein-protein interactions, because they would usually be expected to appear on the protein surface as loops. A series of indel-based mutational experiments and indel-derived peptide inhibitor studies were carried out on C5 (Low et al. 1999). This mutagenesis strategy reversed deletion and insertion sites in C5 relative to C3 or C4 at each of 27 indels in the C3/C4/C5 family, based upon inspection of a previously created multiple sequence alignment (Ogata and Low 1997). For example, if the indel were a putative insertion in C5 relative to C3 and or C4, those residues were deleted in the mutant; for a deletion in C5 relative to C3 and or C4, the corresponding residues from C3 or C4 were inserted. To further assess the usefulness of the indel strategy, the authors designed 23 peptides with indel-proximal sequences. Peptide sequences of about 14 residues in length, spanning an indel, were usually chosen. The precise length, sequence, and presence or absence of terminal amidation/acetylation were selected primarily to optimise peptide solubility (Low et al. 1999). The resulting panel of synthetic peptides was tested for their capacity to inhibit specific C5 and C5b-mediated protein-protein interactions.

The availability of the C5 and C5b models now sheds fresh light on these studies. In particular they allow judgement of the likelihood that a particular mutation will cause localised structural perturbation or more general disruption of the structure. Moreover they allow each mutation to be considered in the context of its position on

the 3D structure relative to other domains and or residues implicated in function.

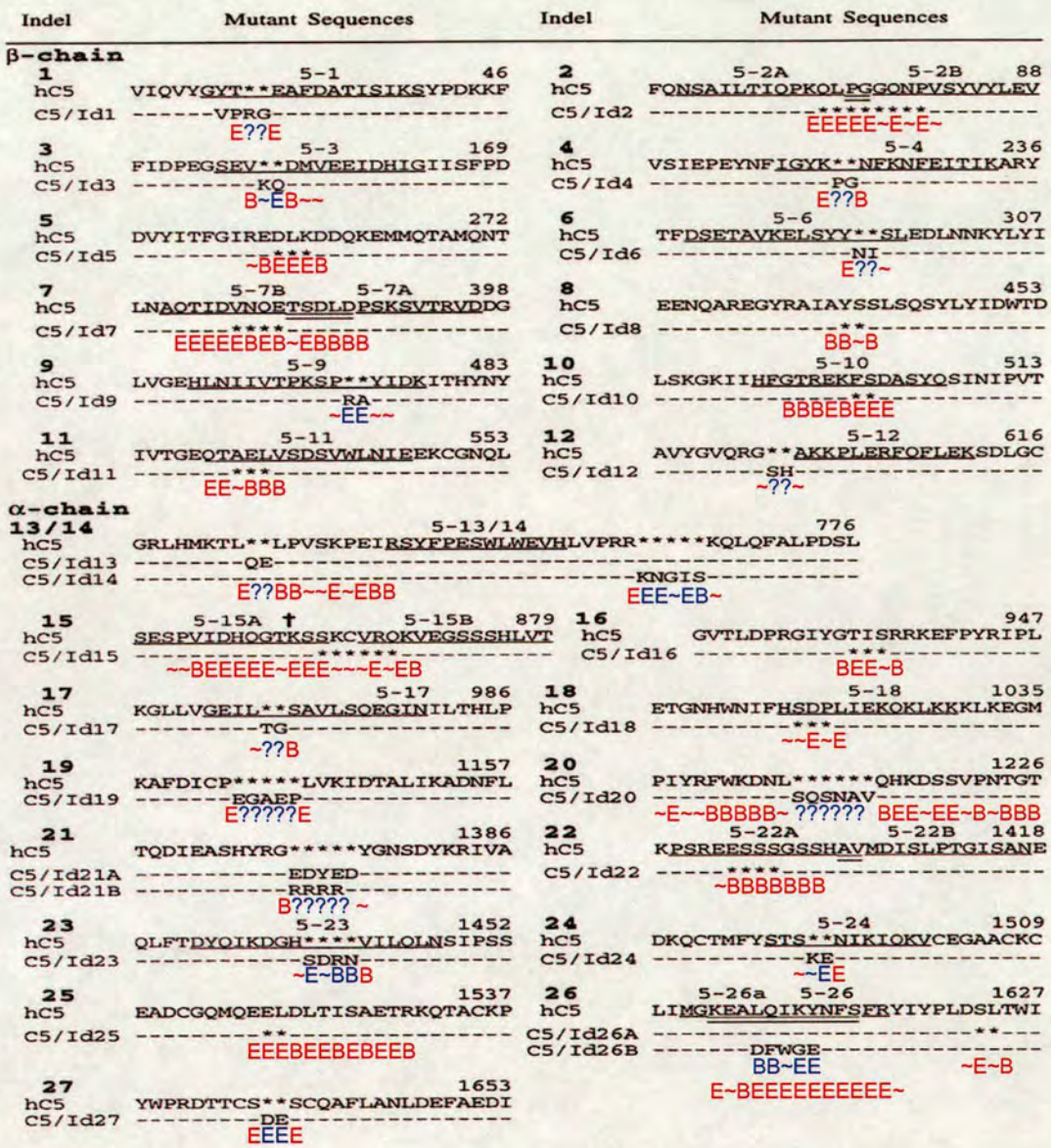


Figure 5.7: Exposure or burial of residues in human C5 or C3 identified on the basis of a previous indel-driven mutagenesis strategy. The figure has been adapted from (Low et al. 1999) to show regions of the 27 originally identified indels (Id1, Id2, Id3 etc.), along with their accessibility according to the 3D-model of C5 and structure of C3. Dashes indicate identity of the indel sequence with the human C5 sequence. Asterisks indicate absent residues (gaps) in the indel relative to the C5 sequence, or *vice versa* (*i.e.* deletion or insertion mutants respectively). Sequences of the 23 synthetic C5-derived peptides are underlined; double underlining indicates overlaps between synthetic peptides. Lys861, which is the primary trypsin cleavage site in C5, is labelled with a '†' near indel 15. The surface accessibility (Fraczkiewicz and Braun 1998) of indel residue side-chains, and some flanking residues, is indicated below the sequence: 'E' = exposed; 'B' = buried; '~' = neither exposed nor

buried; '?' = relative to C4; red labelling indicates C5 model accessibility; blue labelling indicates C3 structure residue accessibility.

Indel	Confirmed indel?	Low expression	Cleaved by convertase?	Category (A, B, C or D)	Poor activity
1	No			D	
2	No	Yes		No data	
3	No			C	Yes
4	Yes	Yes		No data	
5	Yes			B	
6	Yes	Yes		A	Yes
7	No			D	
8	No	Yes		D	
9	Yes	Yes		A	Yes
10	Yes			A	Yes
11	No	Yes		No data	
12	No			D	
13	No			D	
14	Yes		Poor	B	
15	No		Poor	Special	Yes
16	No			C	Yes
17	No			D	
18	No			D	
19	No			D	
20	No			D	
21	Yes			B	
22	Yes			B	
23	Yes			A	Yes
24	Yes			B	
25	Yes			B	
26a	No			D	
26b	No			C	Yes
27	Yes			B	

Table 5.4: Indels categorised on basis of activity: Indels enumerated in the previous study (Low et al. 1999) are categorised according to the new structure-based sequence alignment as "confirmed indels", *i.e.* an appropriate insertion or deletion, expression levels, activity in MAC formation, along with susceptibility to convertase cleavage are indicated. Note (cobra venom factor.Bb) CVF.Bb was used as the convertase, rather than a true C5 convertase; CVF.Bb cleaves both C3 and C5.

The C3 structure allows structure-guided sequence alignment and hence more accurate identification of indels. Many of these match to the previously proposed ones (Figure 5.7, which also summarise whether the residues concerned are exposed or buried within the structure of the parent protein). Thus of the mutations in the paper (identified as "indel 1" through "indel 27"), the following correspond to

deletions of residues that are located in loops within the modelled C5 structure, or insertions of residues corresponding to loops in the C3 structure: indels 4, 5, 6, 9, 10, 14, 21a, 21b, 22, 23, 24, 25 and 27 (Table 5.4). Despite their presence in loops, not all of the residues concerned are in fact exposed (due to the multiple domain nature of the protein), but at the domain level these mutations should have only a localised structural effect. Other deletion mutations performed in the original work correspond to non-loop residues that are generally (but not always) buried in C5, while in some cases insertions were made between a pair of buried residues (Figure 5.7). These could prevent correct folding of a domain or have a wider structural impact. Poor expression levels were obtained for mutants with the following mutations: indels 2, 4 and 11, and no functional data could be obtained for these.

In the light of the C5 structure it is possible to group these mutations into four categories as follows. *A*: appropriate mutations with regard to an indel-based strategy (using the new alignment) that cause reduced haemolytic activity (*i.e.* a reduction in the ability of the mutated C5 to participate in terminal steps of the complement cascade); *B*: appropriate mutations as in category A but resulting in no loss of activity; *C*: mutations that don't match insertions or deletions according to the new alignment and that result in lowered haemolytic activity; and finally *D*: inappropriate mutations as in category C but not associated with significant loss of activity. The positions of residues corresponding to the various mutations were mapped onto the 3D structures of C5 and C5b and are illustrated in Figures 5.8 and 5.9. Note that mutations identified as indel numbers 14 and 15 were found to be poor substrates for C5 (CVF) convertase, while all the other expressed mutants were cleaved nearly as efficiently as the wild type (Table 5.4). Note that there is a trypsin cleavage site in C5 at Lys 861, in the sequence 'QGTKSS' (while the convertase cleavage site is located at Arg 733,

in the sequence 'LGRLH'). Trypsin cleavage is also blocked by Alexion's inhibitory antibodies (Alexion Pharmaceuticals, Cheshire, CT, USA), eculizumab and pexelizumab (Kaplan 2002; Whiss 2002), humanised monoclonal antibodies that prevents the cleavage of human complement component C5 into its pro-inflammatory components (see Chapter 1, Section 1.5.1.2). Within the modelled C5 structure, Lys 861 occurs in the protruding loop in the MG7 domain, in between two disulfide bridges (one of which is unique to C5) (Figure 5.6b). This residue lies approximately 48 Å (Cα-to-Cα) from the convertase cleavage site at Arg 733.

Those mutations that reduce haemolytic activity will be considered first. These might prevent productive interaction with C6 or C7, either in the C5:C6 or C5:C7 pre-activation complexes, or the C5b:C6 post-cleavage complex (Thai and Ogata 2004), although they might also interfere with subsequent steps in assembly of the MAC (see Section 5.3.4). Mutations might also perturb the mechanisms involved in the extensive domain-domain rearrangements that are required for the transition from C5 to C5b. Indels 6, 9, 10 and 23 belong to category A (*i.e.* domain structures are likely retained). Indels 9 and 10, both in MG5, correspond to proximal loops at the base of the molecule (in the conventional view with C345C at the top, Figure 5.8) and together constitute an excellent candidate for a protein-protein recognition patch. Since this part of the molecule is completely unaffected by the conformational changes accompanying activation (cleavage of C5 to C5b) it could be involved in forming the pre-activation complexes. Indel 6 (MG3) lies on the side on the β ring and is also unaffected by activation since it is positioned far from the nascent interaction site for the displaced TED (in C5b). Despite the fact that they are ~51 Å (in C5) and ~55 Å (in C5b) apart, all three of these indels could contribute to the same binding face. Indel 23, corresponding to an insertion in C5 (with respect to C3),

surprisingly involved insertion of several buried residues. Intriguingly, it lies on the face of MG8 close to the interface with the TED domain (with residue Ile1443, buried at the interface). Thus this mutation likely disrupts the arrangement of TED with respect to the rest of the C5 protein.

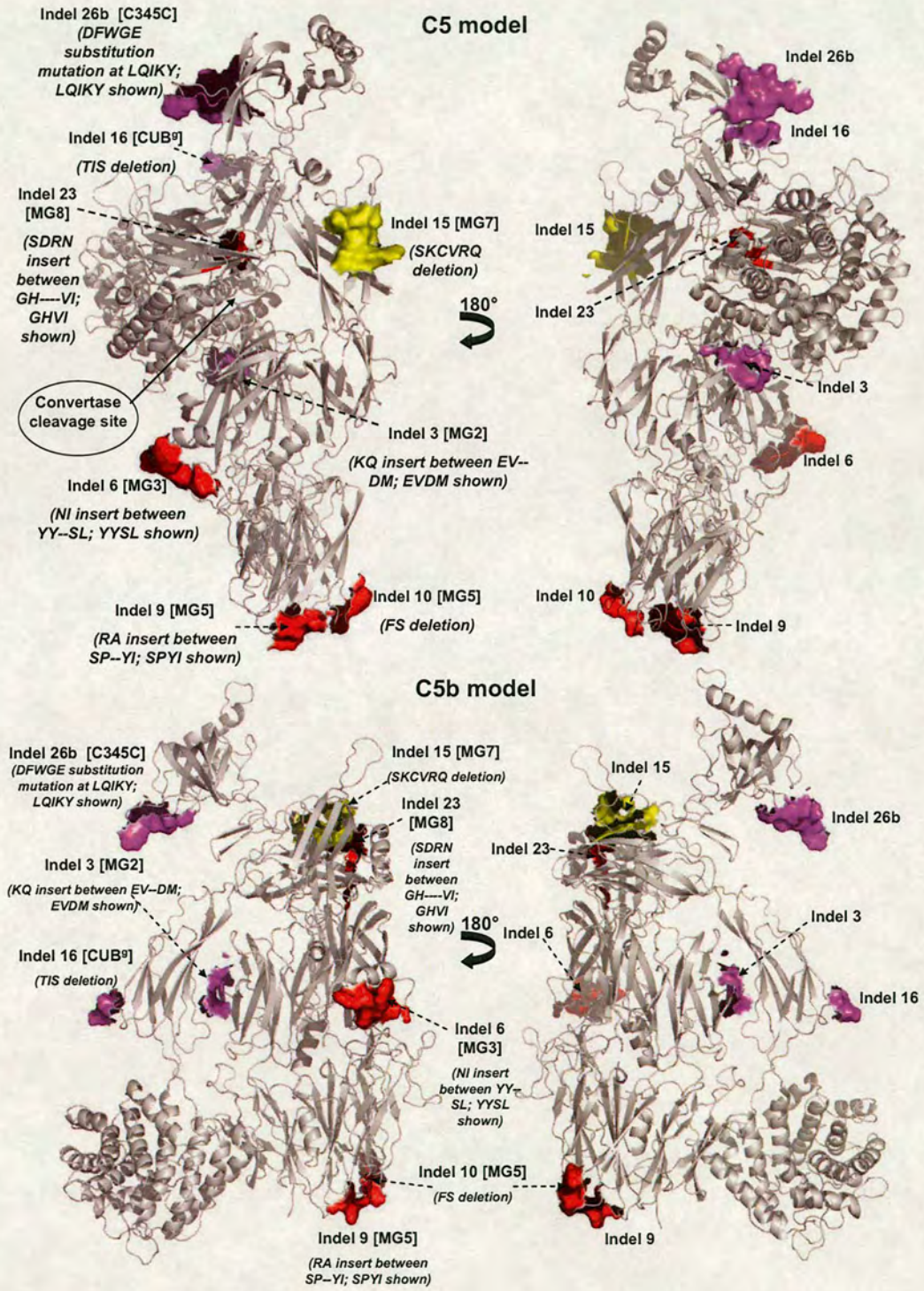


Figure 5.8: Mutant mapping on C5 and C5b models that reduce haemolytic activity. Two views rotated by 180° about the *y*-axis, showing C5/C5b surface mapping of mutant (indel) residues with reduced haemolytic activity (Low et al. 1999), *i.e.* categories A (red) and C (violet); also mapped in yellow is indel 15. Their corresponding domains are shown within square brackets.

In category C, indel 3 also lies in the β-ring, within MG2 and near the interface with TED in C5, and with CUB in C5b. So this mutation might also disrupt the activation mechanism. Indels 16 (CUB) and 26b (C345C), correspond to two regions on the α-chain, and exist in close proximity at the top of the structure (as viewed in Figure 5.8). Indel 16 involves deletion of three residues, Thr-Ile-Ser, which are exposed and lie in a loop of CUB (even though they do not correspond to an indel) and hence this mutation may in fact cause little structural change. The CUB domain undergoes a major orientation upon activation and in C5b, indel 16 appears to be exposed on the far side of CUB where it could form part of a nascent interaction site for C6. Indel 26b corresponds to a prominent feature of C345C that has been discussed elsewhere (Bramham et al. 2005b) (discussed further in Section 5.3.4).

Categories B and D – that have no significant functional consequences – are distributed widely over the structure (Figure 5.9). Note, however, that no such mutants were found in MG5, MG7, or CUB. Although this could be a coincidence, it is interesting to note that the face of MG5 between indels 9 and 10 (which are proximal) and indel 6 remains a very good candidate for C6 interactions and hence for future mutagenesis studies. One obvious difference between C5 and C5b is the prominence of indel 21 in the activated structure. The fact that insertion of a multiple residue sequence in this loop has no functional effects, however, implies that it is not critical for MAC formation.

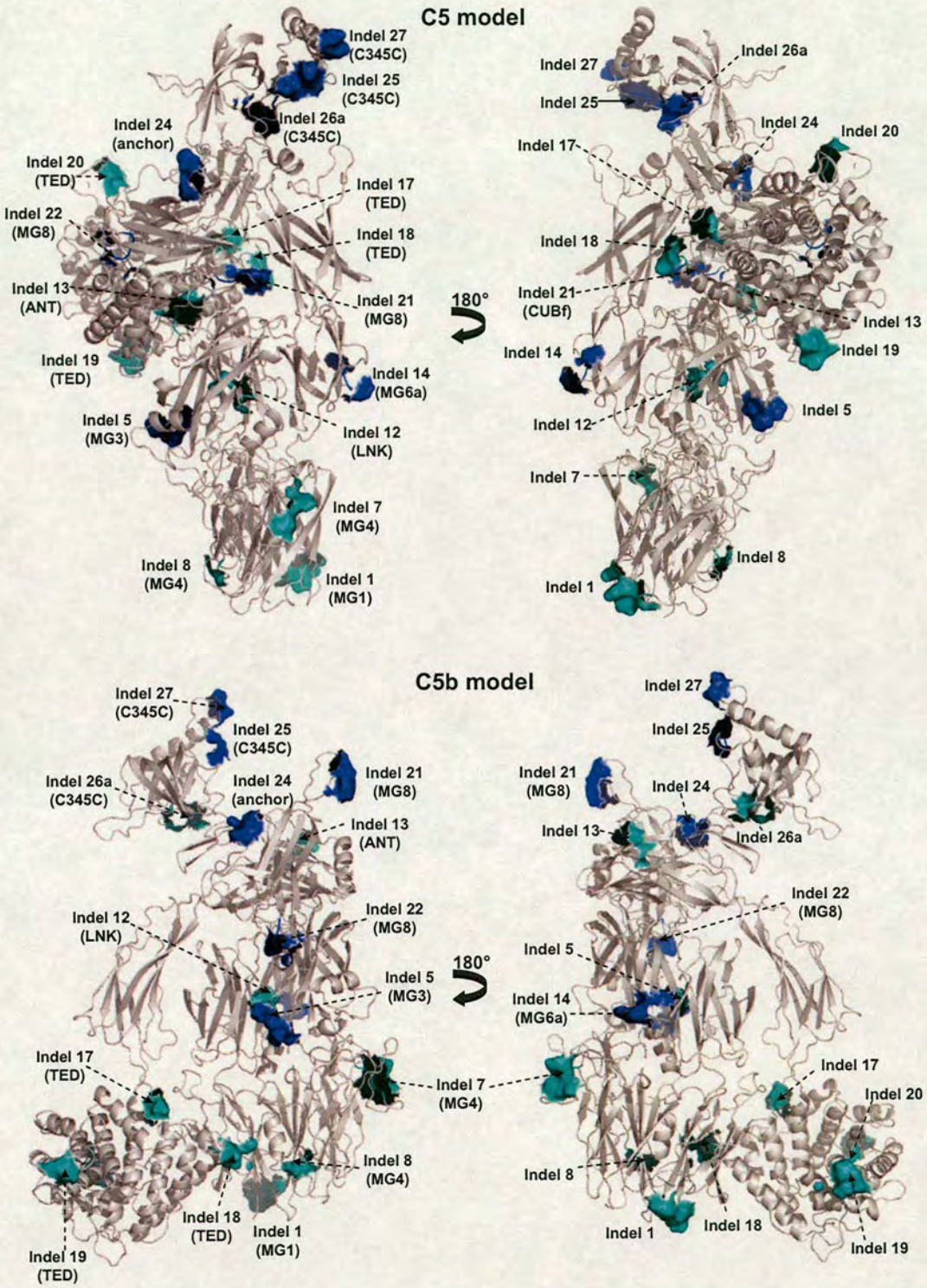


Figure 5.9: Mutant mapping on C5 and C5b models that cause no loss of activity. Two views rotated by 180° about the *y*-axis, showing C5/C5b surface mapping of mutant (indel) residues that cause no loss of haemolytic activity (Low et al. 1999), *i.e.* categories B (blue) and D (cyan). Their corresponding domains are shown within brackets.

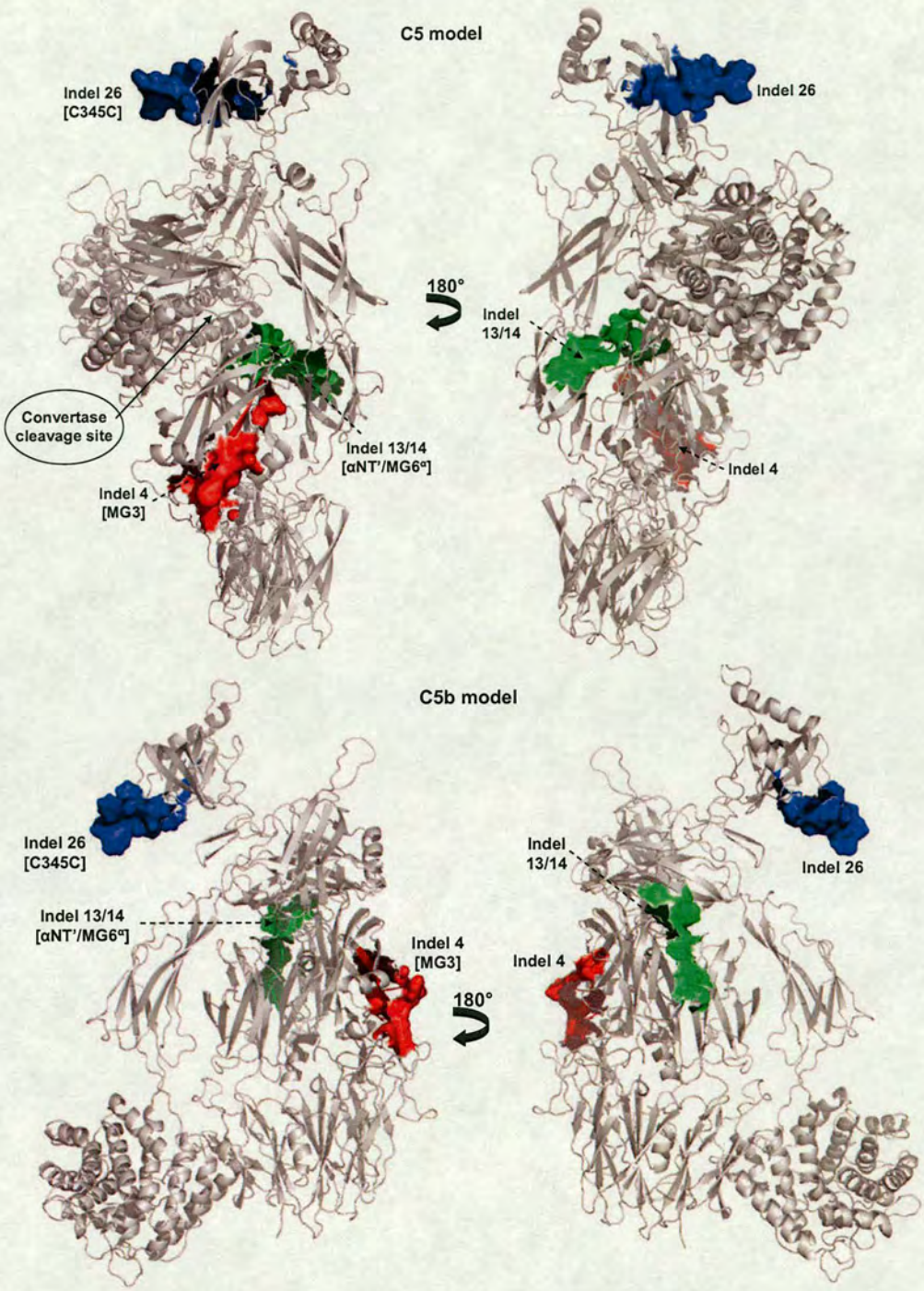


Figure 5.10: Peptide mapping on C5 and C5b models. Two views rotated by 180° about the *y*-axis, showing surface mapping of peptide residues that cause 50% inhibition of both haemolytic and bactericidal activities at <=100μM in C5 and C5b (Low et al. 1999); indels 4, 13/14 and 26 coloured differently. Their corresponding domains are shown within square brackets.

A review of the peptide studies in the 3D structural context is, unfortunately, less illuminating (Figure 5.10). Only peptide 26 coincides with a category A or C

mutation, thus confirming this prominent feature of C345C as of significant functional interest. Peptide 4 corresponds with indel 4, which was not expressed well and therefore not investigated – nonetheless the equivalent sequence of residues (to peptide 4) lies close to indel 6 (Figure 5.10) that was previously identified as potentially important (see above). The lack of any significant functional consequences for indels 13 and 14 is not irreconcilable with the observation that a peptide selected from a region between these two indels (Figure 5.10) inhibited haemolysis. This peptide corresponds to the C-terminal part of α'NT and could participate in protein-protein interactions independently of indels 13 and 14, although none of the other experiments had identified this region as important.

5.3.4 The functional role of the C345C domain revisited in view of the intact C5 model

The C-terminal ~150 amino acid residue portions of the α-chains of C3 and C5 and of the γ-chain of C4 had been anticipated to form independently folded units prior to determination of the C3 structure (Ishii et al. 1992; Banyai and Patthy 1999; Bramham et al. 2005b). These regions not only exhibit high sequence-similarity with one another (Table 5.3) but their sequences also resemble those of the C-terminal segment of the *Caenorhabditis elegans* UNC-6, procollagen C-proteinase enhancer proteins and secreted frizzled-related proteins among others. This family of domains has been named the netrin-like (NTR) module (Banyai and Patthy 1999).

It is notable that C5-C345C, unlike the C345C domain of C3, can bind to both C6 and C7 in surface plasmon resonance-based assays (Thai and Ogata 2003). Furthermore C5-C345C can inhibit recruitment of C7 by C5b.C6 through an interaction between C5-C345C and the pair of factor I membrane attack complex

(FIMAC) domains, (also called Factor-I-modules (FIMs)), at the C-terminus of C7. Thus the C5-C345C domain provides at least part of the interacting surface between C5b and C7 in formation of the MAC. The C345C domain also harbours a region that interacts with the C5 convertase (Sandoval et al. 2000), although the cleavage site itself lies some 800 residues away towards the N-terminus of the α-chain of C5 (~45 Å from the cleavage site). Expression of the segment of C5 corresponding to its C345C domain (C5-C345C) resulted in a full 3D solution structure (Bramham et al. 2005b). The NMR-derived structure helped to identify functionally important patches on C5-C345C but the orientation of C5-C345C relative to the remainder of C5/C5b was not known and this limited the usefulness of structure-function analysis. The new homology models for C5 and C5b partially make good this knowledge deficit (although there is some uncertainty about the reliability of the modelling in terms of placing this domain accurately - see Section 5.3.1).

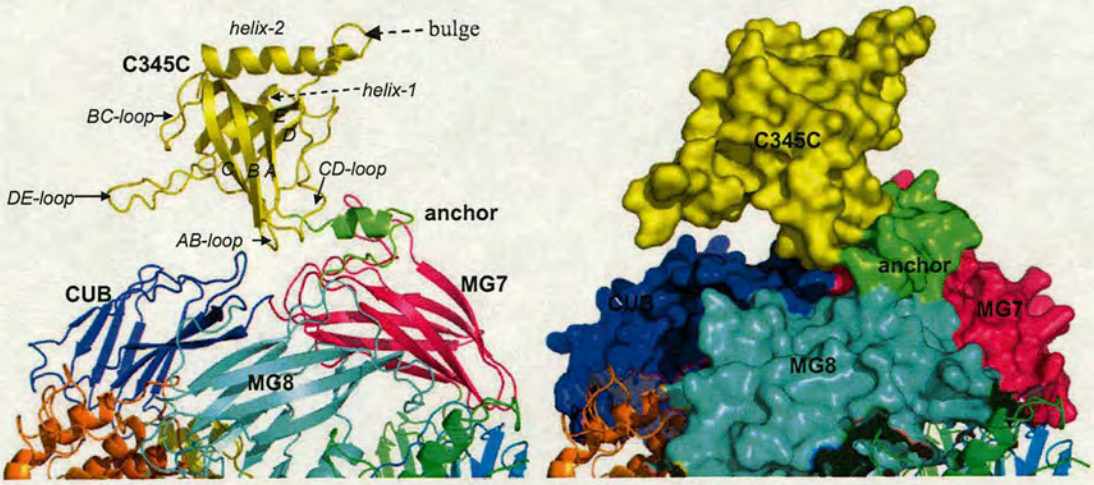


Figure 5.11: Position of C345C within intact C5 model. A cartoon representation of the C345C domain is shown in yellow along with neighbouring and interacting domains - CUB (blue), MG7 (hot pink), MG8 (cyan) and the anchor region (light green). The assigned secondary structure in C345C is labelled according to Bramham et al. (Bramham et al. 2005b). A surface representation of the same view is shown on the right hand panel.

The core of C345C is a barrel composed from two orthogonal, three-stranded, antiparallel twisted β-sheets made up from strands A^C-B-C and strands A^N-D-E (where the superscripts N and C denote the N- and the C- terminal halves of strand A) (Figure 5.11). Two adjacent α-helices – a shorter more N-terminal one (helix-1) and a longer C-terminal helix-2 - are essentially aligned with, and lie against, the convex face of the A^N-D-E sheet. The N-terminal segment of the domain runs above one end of the barrel, from Cys 1514 to the top of helix-1. Cys 1514 is disulfide-linked to Cys 1588, which is located in the long CD-loop; the CD loop crosses over the otherwise open end of the barrel from the A^C-B-C sheet to the A^N-D-E sheet. At the bottom of the short helix-1 the transition to the long strand A contains Cys 1535, which is linked to Cys 1658, the C-terminal residue of the module. The BC-loop caps off the other end of the barrel. The long DE-loop protrudes prominently from the more open side of the barrel-like structure, opposite to the pair of helices. After strand E, a stretch of 13 residues makes the transition to the top of helix-2. From residue 1634 onwards this region forms a bulge above helix-2 (Figure 5.11). Most of the solvent-exposed hydrophobic residues lie in the DE loop, while Phe 1654 and Leu 1655 are exposed near the C-terminus. Adjacent to this exposed pair of side-chains is a patch of negatively charged side-chains (Glu 1528, 1648 and 1651; Asp 1647 and 1652) that dominate the electrostatic surface of C5-C345C (see below) (Bramham et al. 2005b).

In the typical view of C3/C5 (Figures 5.2 and 5.8), the C345C domain is a clearly defined structural entity perched prominently atop the α chain exposing a large surface area for potential protein-protein interactions. This remains true upon transition to C3b/C5b, even though C345C swivels 32° and moves 10.1 Å in C3b (Janssen et al. 2005; Janssen et al. 2006). The face of C345C that contacts the body of C5 is made up from the AB loop, along with the CD loop that is disulfide linked to

the N-terminal residue of the domain (Figure 5.11). This leaves the two α-helices facing away from the rest of C5, and the aforementioned negative patch and adjacent Phe 1654-Leu 1655 pair are well placed to make intermolecular contacts. This part of the structure exhibits structural differences from other NTR modules, and diversity in terms of chemical character amongst C3, C4 and C5. The C345C domain of C3 has many more Asp and Glu residues than either the C4 or C5 equivalents, and consequently has an overall electronegative character - but C3 lacks the patch of five negatively charged side chains that occur next to the exposed hydrophobic side-chains near the C-terminus of C5 (Bramham et al. 2005b). Note that C5-C345C, but not the equivalent domain from C3, binds reversibly to C6 and C7 (the pI of C7-FIMAC-I is 9.5 consistent with a strong ionic component in recognition of the highly acidic C5-C345C domain).

The very long DE loop (14 residues, Ala 1606-to-Tyr 1619) is also exposed and is not involved in the interface with the rest of C5 or C5b according to the models. This is surprising given its high content of hydrophobic residues, and (if the model is accurate) enhances the value of previous mutagenesis work on this region (Low et al. 1999). Substitution of Leu 1607-Tyr 1611 - a poorly structured region of the DE loop with no buried side-chains - by the sequence DFWGE (indel 26b, discussed above) destroyed haemolytic activity. Subsequent substitutions of Ile 1609 or Tyr 1611 by Ala had little effect on haemolytic activity or proteolytic susceptibility. Substitution of the exposed Lys 1610, on the other hand, produced mutant C5 molecules with both low haemolytic activity and decreased sensitivity to proteolytic activation. This pinpoints Lys 1610 as the functionally critical residue in the indel 26b pentapeptide. Substitution of Phe 1613 and Phe 1615 also perturbed haemolytic activity and decreased proteolytic susceptibility to the classical pathway

convertase (but not cobra venom factor, which presumably has a different recognition mechanism). Comparison of HSQC spectra for the wild-type and Phe1613Ala versions of C5-C345C has proved that this mutation exerts no non-local structural effects at the domain level (Bramham et al. 2005b), and the intact model suggests no effects on interdomain orientations; Phe 1615 is also exposed, and mutation to Ala would be equally unlikely to disrupt structure. Therefore these mutagenesis results, considered in the context of the model of C5, clearly identify three exposed side-chains (1610, 1613 and 1615) on the DE extension as being specifically involved in an interaction with the convertase that is critical for function.

This is reinforced (see above) by the observation that a peptide extending from Lys 1604 to Arg 1616 inhibited complement haemolytic activity and activation of C5 by the convertase pathway C5 convertase (but not cobra venom factor) (Sandoval et al. 2000). Furthermore, the consequences for inhibitory activity of alanine substitution within the peptide reflected alanine-scanning mutagenesis results. The peptide studies therefore provide additional evidence for a direct interaction between the DE extension and the classical pathway convertase. Based on the structure of C5, the C345C module is ~45 Å from the convertase cleavage site, implying an extended interaction surface. An extended interaction surface between C5 and its convertase is not surprising, since the convertase is a huge complex (> 400 kDa).

5.3.5 Electrostatic and lipophilic surface analysis of C5 and C5b

Negative charge dominates the surfaces of the models of C5 and C5b (Figure 5.12). This is consistent with their calculated theoretical pI values derived from their sequence (Bjellqvist et al. 1993). The pI of C5 is 6.07, and it is 5.85 for C5b. The ANA domain, which gets cleaved off as C5a on the other hand, is basic (pI = 8.93).

Prominent charged features in the model of C5, involve the β-chain MG-ring-like scaffold that appears to contain a highly acidic patch, while regions on TED remain largely neutral. Upon activation, conformationally different C5b reveals a more negatively charged α-chain, while TED and CUB remain largely apolar (Figure 5.12).

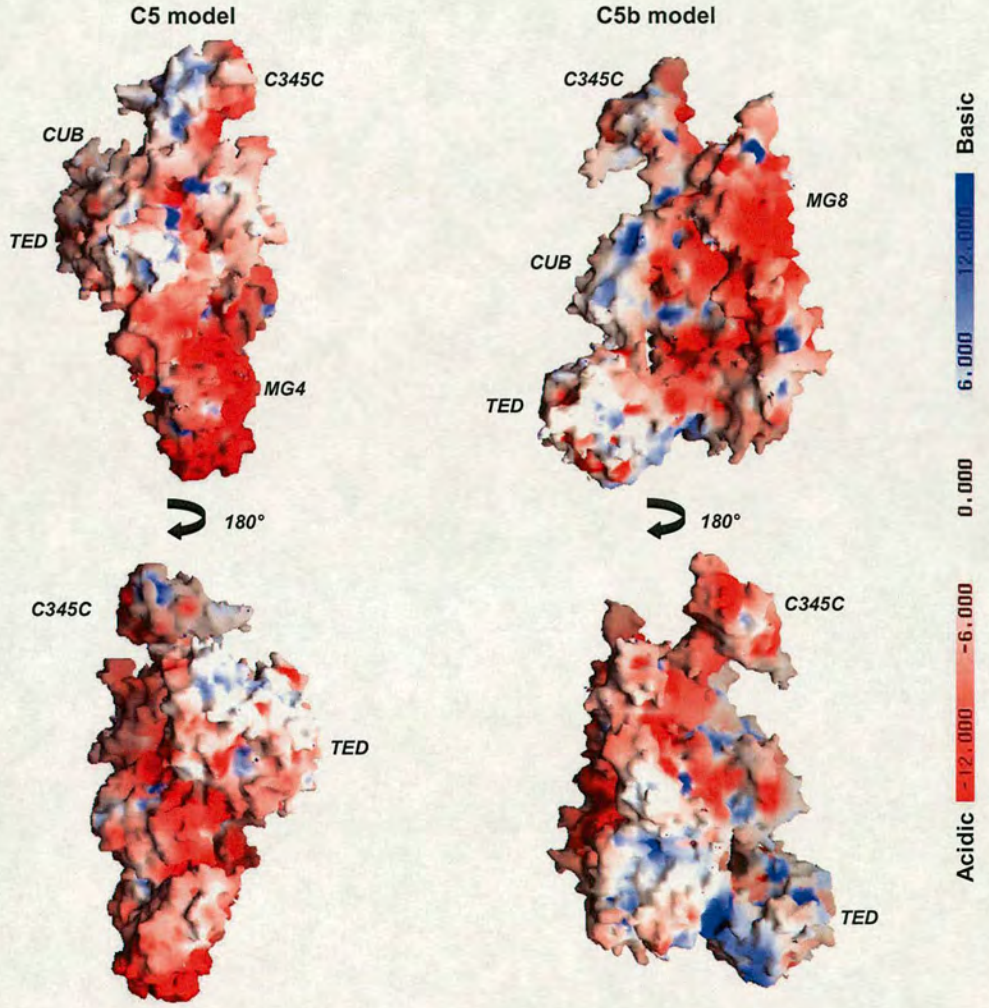


Figure 5.12: GRASP surface representations of C5 and C5b models. Two views rotated by 180° about the *y*-axis depicting the electrostatic surface representation of C5 (left panel) and C5b (right panel) models, respectively. Scale shown on the right-hand side; coloured from red (acidic) to blue (basic) (-12kT to +12 kT). A few domains are labelled to orient the viewer.

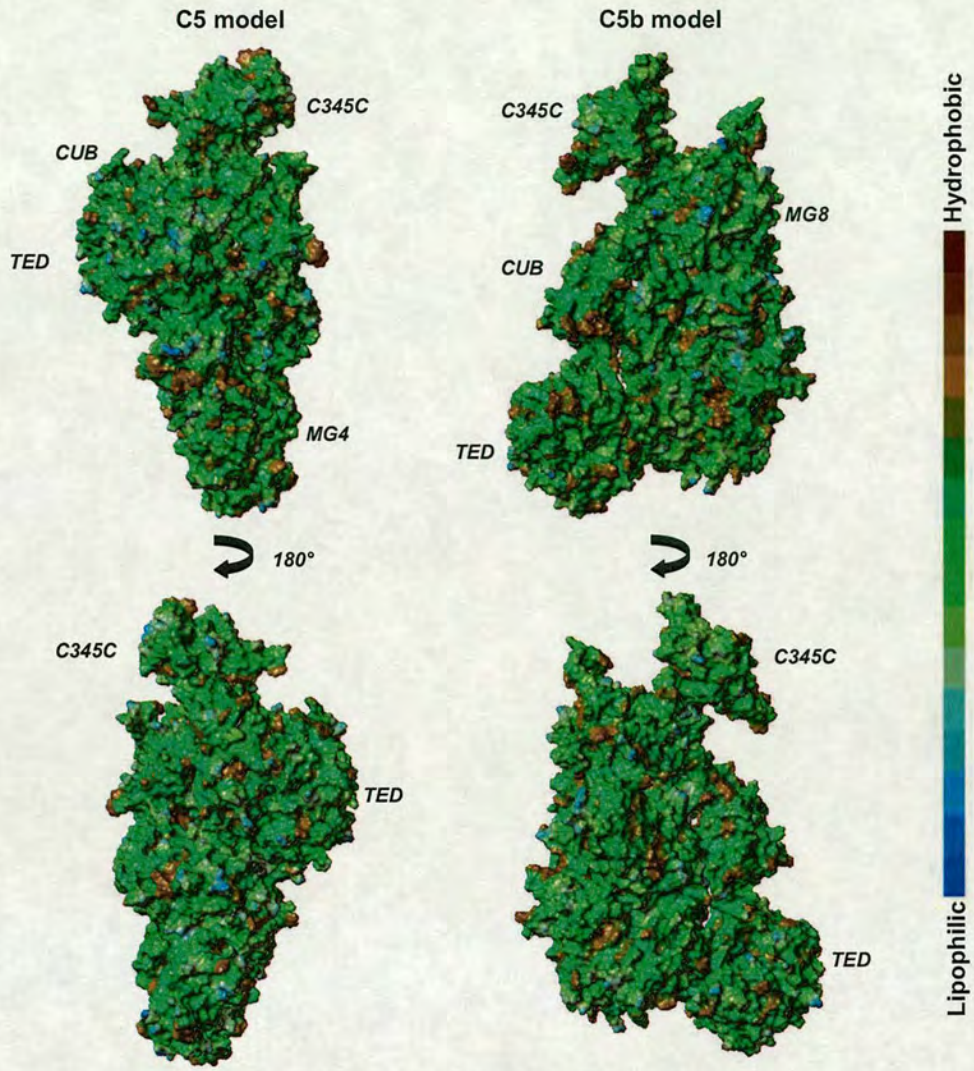


Figure 5.13: MOLCAD surface representations of C5 and C5b models. Two views rotated by 180° about the *y*-axis depicting the lipophilic surface representation of models C5 and C5b, coloured from brown (hydrophobic) to blue (hydrophilic) – scale shown on the right-hand side.

The conformational change that accompanies activation of C5 to C5b is central to the formation of the membrane attack complex. Upon activation, a transient hydrophobic patch appears in C5b that binds irreversibly to C6 (Muller-Eberhard 1985; 1986), creating a nucleation site (C5bC6, the post-activation complex) that proceeds to accrete components C7, C8 and multiple copies of C9, leading to formation of the MAC. The lipophilic surface analysis (Figure 5.13) of the model of C5b now allows proposal of candidate sites. Interesting among them are exposed

patches of hydrophobic residues on CUB, TED - particularly near their interface - and as discussed earlier (Section 5.3.4) in the DE-loop of C345C.

5.4 Concluding remarks

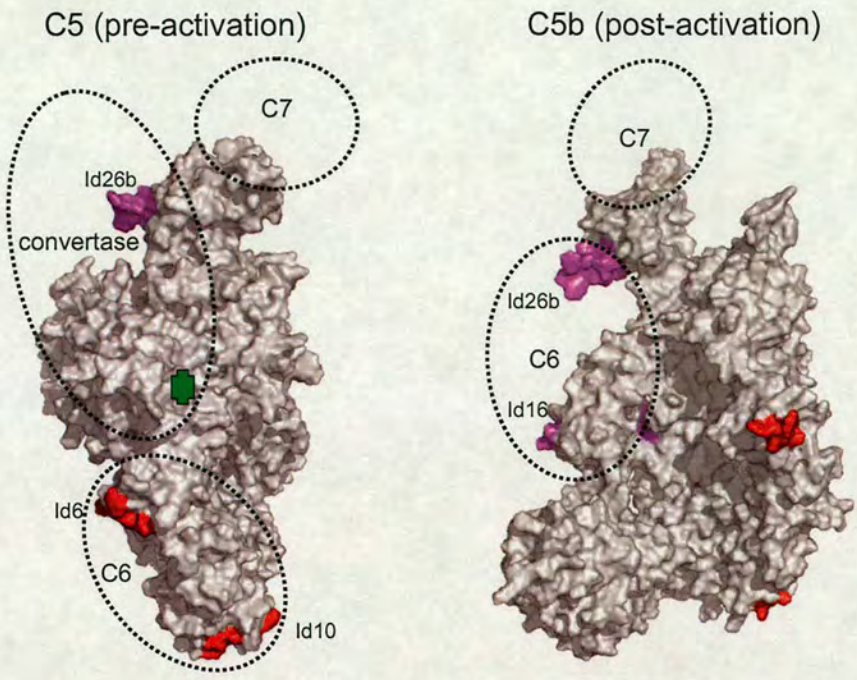


Figure 5.14: Hypothetical model for assembly of terminal pathway interactions. The models of C5 and C5b are shown along with potential regions of interaction with the large convertase (in C5), C6 and C7 (in both) – not drawn to scale. According to this model, in C5, indels 6, 9, 10 and 26b potentially participate in protein-protein interactions along with the region above helix-1 of C345C. The convertase cleavage site (Arg 733) is indicated in green. The right hand panel shows indels 16 and 26b in C5b now contributing to the interaction site with C6, while the interaction with C7 remains unchanged.

In the work described in this chapter, a model of C5 has been built by homology with the C3 structure and evaluated. The C5 model reveals the presence of a previously unrecognised disulfide. A model of C5b was also built that is probably valid at the domain level. These models are not inconsistent with conformational changes upon activation that generally emulate those seen in C3. But analysis of domain-domain interactions revealed uncertainty in the precise orientation of domains in the activated structure. In particular, the model did not provide support for the newly positioned TED in C5b adopting the same orientation as in C3b. That the TED in C5b could be twisted, for example, is not unreasonable since the C5 TED contains no thioester and

must perform a different role to the C3/C4 TEDs. Inspection of previous indel-based mutagenesis data allows construction of a hypothetical model. This is illustrated in Figure 5.14. Models such as this are useful as hypothesis generators and can be tested experimentally by mutagenesis or by biophysical methods such as SAXS or FRET. The model predicts that C6 associates differently with C5 (pre-activation complex) and C5b (post-activation complex), while C7 remains attached to the C345C domain. In the activation complex, C6 and C7 could be brought into proximity creating a novel site for C8.

5.5 Declarations and Acknowledgements

- I would like to thank Ronald T. Ogata (TPIMS, San Diego) for several useful discussions on C5; Christof Winter (TU Dresden) for running his program SCOPPI independently on the model and template structures upon request; and Xueping Quan (Edinburgh) for advice on Molsurfer's ECC and HCC.

- Parts of the discussion in Section 5.3.4 were published in:

  Bramham, J., Thai, C. T., **Soares, D. C.**, Uhrin, D., Ogata, R. T., and Barlow, P. N. (2005). Functional insights from the structure of the multifunctional C345C domain of C5 of complement. J Biol Chem *280*, 10636-10645.

References

Abdul Ajees, A., Gunasekaran, K., Volanakis, J.E., Narayana, S.V., Kotwal, G.J., and Murthy, H.M. 2006. The structure of complement C3b provides insights into complement activation and regulation. *Nature* **444:** 221-225.

Accardo, P., Sanchez-Corral, P., Criado, O., Garcia, E., and Rodriguez de Cordoba, S. 1996. Binding of human complement component C4b-binding protein (C4BP) to Streptococcus pyogenes involves the C4b-binding site. *J Immunol* **157:** 4935-4939.

Adams, E.M., Brown, M.C., Nunge, M., Krych, M., and Atkinson, J.P. 1991. Contribution of the repeating domains of membrane cofactor protein (CD46) of the complement system to ligand binding and cofactor activity. *J Immunol* **147:** 3005-3011.

Albrecht, J.C., and Fleckenstein, B. 1992. New member of the multigene family of complement control proteins in herpesvirus saimiri. *J Virol* **66:** 3937-3940.

Alitalo, A., Meri, T., Comstedt, P., Jeffery, L., Tornberg, J., Strandin, T., Lankinen, H., Bergstrom, S., Cinco, M., Vuppala, S.R., et al. 2005. Expression of complement factor H binding immunoevasion proteins in Borrelia garinii isolated from patients with neuroborreliosis. *Eur J Immunol* **35:** 3043-3053.

Alitalo, A., Meri, T., Lankinen, H., Seppala, I., Lahdenne, P., Hefty, P.S., Akins, D., and Meri, S. 2002. Complement inhibitor factor H binding to Lyme disease spirochetes is mediated by inducible expression of multiple plasmid-encoded outer surface protein E paralogs. *J Immunol* **169:** 3847-3853.

Alitalo, A., Meri, T., Ramo, L., Jokiranta, T.S., Heikkila, T., Seppala, I.J., Oksi, J., Viljanen, M., and Meri, S. 2001. Complement evasion by Borrelia burgdorferi: serum-resistant strains promote C3b inactivation. *Infect Immun* **69:** 3685-3691.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

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

Anderson, K.L., Billington, J., Pettigrew, D., Cota, E., Simpson, P., Roversi, P., Chen, H.A., Urvil, P., du Merle, L., Barlow, P.N., et al. 2004. An atomic resolution model for assembly, architecture, and function of the Dr adhesins. *Mol Cell* **15:** 647-657.

Areschoug, T., Stalhammar-Carlemalm, M., Karlsson, I., and Lindahl, G. 2002. Streptococcal beta protein has separate binding sites for human factor H and IgA-Fc. *J Biol Chem* **277:** 12642-12648.

Arroyave, C.M., and Mullereb.Hj. 1973. Interactions between Human C5, C6, and C7 and Their Functional Significance in Complement-Dependent Cytolysis. *Journal of Immunology* **111:** 536-545.

Asakawa, R., Komatsuzawa, H., Kawai, T., Yamada, S., Goncalves, R.B., Izumi, S., Fujiwara, T., Nakano, Y., Suzuki, N., Uchida, Y., et al. 2003. Outer membrane protein 100, a versatile virulence factor of Actinobacillus actinomycetemcomitans. *Mol Microbiol* **50:** 1125-1139.

Aslam, M., Guthridge, J.M., Hack, B.K., Quigg, R.J., Holers, V.M., and 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.

Aslam, M., and 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.

Aszodi, A., and Taylor, W.R. 1996. Homology modelling by distance geometry. *Fold Des* **1:** 325-334.

Aubry, J.P., Pochon, S., Gauchat, J.F., Nueda-Marin, A., Holers, V.M., Graber, P., Siegfried, C., and Bonnefoy, J.Y. 1994. CD23 interacts with a new functional extracytoplasmic domain involving N-linked oligosaccharides on CD21. *J Immunol* **152:** 5806-5813.

Ault, B.H. 2000. Factor H and the pathogenesis of renal diseases. *Pediatr Nephrol* **14:** 1045-1053.

Ault, B.H., Schmidt, B.Z., Fowler, N.L., Kashtan, C.E., Ahmed, A.E., Vogt, B.A., and 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.

Azimzadeh, A., Zorn, G.L., 3rd, Blair, K.S., Zhang, J.P., Pfeiffer, S., Harrison, R.A., Cozzi, E., White, D.J., and Pierson, R.N., 3rd. 2003. Hyperacute lung rejection in the pig-to-human model. 2.

Synergy between soluble and membrane complement inhibition. *Xenotransplantation* **10:** 120-131.

Baker, D., and Sali, A. 2001. Protein structure prediction and structural genomics. *Science* **294:** 93-96.

Banyai, L., and Patthy, L. 1999. The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases. *Protein Sci* **8:** 1636-1642.

Barel, M., Balbo, M., and Frade, R. 1998. Evidence for a new transcript of the Epstein-Barr virus/C3d receptor (CR2, CD21) which is due to alternative exon usage. *Mol Immunol* **35:** 1025-1031.

Barlow, P.N., Baron, M., Norman, D.G., Day, A.J., Willis, A.C., Sim, R.B., and Campbell, I.D. 1991. Secondary structure of a complement control protein module by two-dimensional 1H NMR. *Biochemistry* **30:** 997-1004.

Barlow, P.N., Norman, D.G., Steinkasserer, A., Horne, T.J., Pearce, J., Driscoll, P.C., Sim, R.B., and 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., and Campbell, I.D. 1993. Solution structure of a pair of complement modules by nuclear magnetic resonance. *J Mol Biol* **232:** 268-284.

Bergelson, J.M., Mohanty, J.G., Crowell, R.L., St John, N.F., Lublin, D.M., and Finberg, R.W. 1995. Coxsackievirus B3 adapted to growth in RD cells binds to decay-accelerating factor (CD55). *J Virol* **69:** 1903-1906.

Berggard, K., Lindahl, G., Dahlback, B., and Blom, A.M. 2001. Bordetella pertussis binds to human C4b-binding protein (C4BP) at a site similar to that used by the natural ligand C4b. *Eur J Immunol* **31:** 2771-2780.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Bernet, J., Mullick, J., Panse, Y., Parab, P.B., and Sahu, A. 2004. Kinetic analysis of the interactions between vaccinia virus complement control protein and human complement proteins C3b and C4b. *J Virol* **78:** 9446-9457.

Bersch, B., Hernandez, J.F., Marion, D., and Arlaud, G.J. 1998. Solution structure of the epidermal growth factor (EGF)-like module of human complement protease C1r, an atypical member of the EGF family. *Biochemistry* **37:** 1204-1214.

Bhattacharya, A.A., Lupher, M.L., Jr., Staunton, D.E., and Liddington, R.C. 2004. Crystal structure of the A domain from complement factor B reveals an integrin-like open conformation. *Structure* **12:** 371-378.

Bhella, D., Goodfellow, I.G., Roversi, P., Pettigrew, D., Chaudhry, Y., Evans, D.J., and Lea, S.M. 2004. The structure of echovirus type 12 bound to a two-domain fragment of its cellular attachment protein decay-accelerating factor (CD 55). *J Biol Chem* **279:** 8325-8332.

Billeter, M. 1992. Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. *Q Rev Biophys* **25:** 325-377.

Birmingham, D.J., and Hebert, L.A. 2001. CR1 and CR1-like: the primate immune adherence receptors. *Immunol Rev* **180:** 100-111.

Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S., and Hochstrasser, D. 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14:** 1023-1031.

Blackall, D.P., and Marques, M.B. 2004. Hemolytic uremic syndrome revisited: Shiga toxin, factor H, and fibrin generation. *Am J Clin Pathol* **121:** S81-88.

Blackmore, T.K., Fischetti, V.A., Sadlon, T.A., Ward, H.M., and Gordon, D.L. 1998a. M protein of the group A Streptococcus binds to the seventh short consensus repeat of human complement factor H. *Infect Immun* **66:** 1427-1431.

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

Blackmore, T.K., Sadlon, T.A., Ward, H.M., Lublin, D.M., and 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.

Blandin, S., and Levashina, E.A. 2004. Thioester-containing proteins and insect immunity. *Mol Immunol* **40:** 903-908.

Blein, S., Ginham, R., Uhrin, D., Smith, B.O., Soares, D.C., Veltel, S., McIlhinney, R.A., White, J.H., and Barlow, P.N. 2004. Structural analysis of the complement control protein (CCP) modules

of GABA(B) receptor 1a: only one of the two CCP modules is compactly folded. *J Biol Chem* **279:** 48292-48306.

Blein, S., Hawrot, E., and Barlow, P. 2000. The metabotropic GABA receptor: molecular insights and their functional consequences. *Cell Mol Life Sci* **57:** 635-650.

Blom, A.M., Berggard, K., Webb, J.H., Lindahl, G., Villoutreix, B.O., and Dahlback, B. 2000a. Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *J Immunol* **164:** 5328-5336.

Blom, A.M., Covell, D.G., Wallqvist, A., Dahlback, B., and Villoutreix, B.O. 1998. The C4b-binding protein-protein S interaction is hydrophobic in nature. *Biochim Biophys Acta* **1388:** 181-189.

Blom, A.M., Kask, L., and Dahlback, B. 2001a. Structural requirements for the complement regulatory activities of C4BP. *J Biol Chem* **276:** 27136-27144.

Blom, A.M., Kask, L., and Dahlback, B. 2003. CCP1-4 of the C4b-binding protein alpha-chain are required for factor I mediated cleavage of complement factor C3b. *Mol Immunol* **39:** 547-556.

Blom, A.M., Rytkonen, A., Vasquez, P., Lindahl, G., Dahlback, B., and Jonsson, A.B. 2001b. A novel interaction between type IV pili of Neisseria gonorrhoeae and the human complement regulator C4B-binding protein. *J Immunol* **166:** 6764-6770.

Blom, A.M., Webb, J., Villoutreix, B.O., and Dahlback, B. 1999. A cluster of positively charged amino acids in the C4BP alpha-chain is crucial for C4b binding and factor I cofactor function. *J Biol Chem* **274:** 19237-19245.

Blom, A.M., Zadura, A.F., Villoutreix, B.O., and Dahlback, B. 2000b. Positively charged amino acids at the interface between alpha-chain CCP1 and CCP2 of C4BP are required for regulation of the classical C3-convertase. *Mol Immunol* **37:** 445-453.

Blomberg, N., Gabdoulline, R.R., Nilges, M., and Wade, R.C. 1999. Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins* **37:** 379-387.

Blundell, T.L., Sibanda, B.L., Sternberg, M.J., and Thornton, J.M. 1987. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326:** 347-352.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31:** 365-370.

Bork, P., Downing, A.K., Kieffer, B., and Campbell, I.D. 1996. Structure and distribution of modules in extracellular proteins. *Q Rev Biophys* **29:** 119-167.

Bork, P., Holm, L., and Sander, C. 1994. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* **242:** 309-320.

Botti, S.A., Felder, C.E., Sussman, J.L., and Silman, I. 1998. Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. *Protein Eng* **11:** 415-420.

Bouma, B., de Groot, P.G., van den Elsen, J.M., Ravelli, R.B., Schouten, A., Simmelink, M.J., Derksen, R.H., Kroon, J., and Gros, P. 1999. Adhesion mechanism of human beta(2)-glycoprotein I to phospholipids based on its crystal structure. *Embo J* **18:** 5166-5174.

Bower, M.J., Cohen, F.E., and Dunbrack, R.L., Jr. 1997. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J Mol Biol* **267:** 1268-1282.

Bozic, B., Cucnik, S., Kveder, T., and Rozman, B. 2005. Avidity of anti-beta-2-glycoprotein I antibodies. *Autoimmun Rev* **4:** 303-308.

Bramham, J., Thai, C.T., Soares, D.C., Uhrin, D., Ogata, R.T., and Barlow, P.N. 2005a. Functional insights from the structure of the multifunctional C345C domain of C5 of complement. *J Biol Chem* **280:** 10636-10645. Epub 12004 Dec 10614.

Bramham, J., Thai, C.T., Soares, D.C., Uhrin, D., Ogata, R.T., and Barlow, P.N. 2005b. Functional insights from the structure of the multifunctional C345C domain of C5 of complement. *J Biol Chem* **280:** 10636-10645.

Brodbeck, W.G., Kuttner-Kondo, L., Mold, C., and Medof, M.E. 2000. Structure/function studies of human decay-accelerating factor. *Immunology* **101:** 104-111.

Brodbeck, W.G., Liu, D., Sperry, J., Mold, C., and Medof, M.E. 1996. Localization of classical and alternative pathway regulatory activity within the decay-accelerating factor. *J Immunol* **156:** 2528-2533.

Brodeur, S.R., Angelini, F., Bacharier, L.B., Blom, A.M., Mizoguchi, E., Fujiwara, H., Plebani, A., Notarangelo, L.D., Dahlback, B., Tsitsikov, E., et al. 2003. C4b-binding protein (C4BP) activates B cells through the CD40 receptor. *Immunity* **18:** 837-848.

Brook, E., Herbert, A.P., Jenkins, H.T., Soares, D.C., and Barlow, P.N. 2005. Opportunities for new therapies based on the natural regulators of complement activation. *Ann N Y Acad Sci* **1056:** 176-188.

Brook, E., Lin, F., Medof, M.E., and Barlow, P.N. 2004. Does the rigid junction between the central modules of DAF that is inferred from crystallography exist under physiological conditions? *Molecular Immunology* **41:** 215-215.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. 1983. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **4:** 187-217.

Budayova-Spano, M., Grabarse, W., Thielens, N.M., Hillen, H., Lacroix, M., Schmidt, M., Fontecilla-Camps, J.C., Arlaud, G.J., and Gaboriaud, C. 2002a. Monomeric structures of the zymogen and active catalytic domain of complement protease c1r: further insights into the c1 activation mechanism. *Structure* **10:** 1509-1519.

Budayova-Spano, M., Lacroix, M., Thielens, N.M., Arlaud, G.J., Fontecilla-Camps, J.C., and Gaboriaud, C. 2002b. The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. *Embo J* **21:** 231-239.

Budd, A., Blandin, S., Levashina, E.A., and Gibson, T.J. 2004. Bacterial alpha2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol* **5:** R38.

Buddles, M.R., Donne, R.L., Richards, A., Goodship, J., and Goodship, T.H. 2000. Complement factor H gene mutation associated with autosomal recessive atypical hemolytic uremic syndrome. *Am J Hum Genet* **66:** 1721-1722.

Campbell, R.D., Law, S.K., Reid, K.B., and Sim, R.B. 1988. Structure, organization, and regulation of the complement genes. *Annu Rev Immunol* **6:** 161-195.

Canutescu, A.A., Shelenkov, A.A., and Dunbrack, R.L., Jr. 2003. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* **12:** 2001-2014.

Caprioli, J., Bettinaglio, P., Zipfel, P.F., Amadei, B., Daina, E., Gamba, S., Skerka, C., Marziliano, N., Remuzzi, G., and Noris, M. 2001. The molecular basis of familial hemolytic uremic syndrome: mutation analysis of factor H gene reveals a hot spot in short consensus repeat 20. *J Am Soc Nephrol* **12:** 297-307.

Caprioli, J., Castelletti, F., Bucchioni, S., Bettinaglio, P., Bresin, E., Pianetti, G., Gamba, S., Brioschi, S., Daina, E., Remuzzi, G., et al. 2003. Complement factor H mutations and gene polymorphisms in haemolytic uraemic syndrome: the C-257T, the A2089G and the G2881T polymorphisms are strongly associated with the disease. *Hum Mol Genet* **12:** 3385-3395. Epub 2003 Oct 3328.

Carel, J.C., Myones, B.L., Frazier, B., and Holers, V.M. 1990. Structural requirements for C3d,g/Epstein-Barr virus receptor (CR2/CD21) ligand binding, internalization, and viral infection. *J Biol Chem* **265:** 12293-12299.

Carroll, M.C. 2004a. The complement system in B cell regulation. *Mol Immunol* **41:** 141-146.

Carroll, M.C. 2004b. The complement system in regulation of adaptive immunity. *Nat Immunol* **5:** 981-986.

Casasnovas, J.M., Larvie, M., and Stehle, T. 1999. Crystal structure of two CD46 domains reveals an extended measles virus-binding surface. *Embo J* **18:** 2911-2922.

Cheek, S., Krishna, S.S., and Grishin, N.V. 2006. Structural classification of small, disulfide-rich protein domains. *J Mol Biol* **359:** 215-237.

China, B., Sory, M.P., N'Guyen, B.T., De Bruyere, M., and Cornelis, G.R. 1993. Role of the YadA protein in prevention of opsonization of Yersinia enterocolitica by C3b molecules. *Infect Immun* **61:** 3129-3136.

Chothia, C., and Lesk, A.M. 1986. The relation between the divergence of sequence and structure in proteins. *Embo J* **5:** 823-826.

Chou, K.C., and Heinrikson, R.L. 1997. Prediction of the tertiary structure of the complement control protein module. *J Protein Chem* **16:** 765-773.

Ciulla, E., Emery, A., Konz, D., and Krushkal, J. 2005. Evolutionary history of orthopoxvirus proteins similar to human complement regulators. *Gene* **355:** 40-47.

Claessens, M., Van Cutsem, E., Lasters, I., and Wodak, S. 1989. Modelling the polypeptide backbone with 'spare parts' from known protein structures. *Protein Eng* **2:** 335-345.

Clark, M., Cramer, R.D., and Vanopdenbosch, N. 1989. Validation of the General-Purpose Tripos 5.2 Force-Field. *J Comput Chem* **10:** 982-1012.

Clarkson, N.A., Kaufman, R., Lublin, D.M., Ward, T., Pipkin, P.A., Minor, P.D., Evans, D.J., and Almond, J.W. 1995. Characterization of the echovirus 7 receptor: domains of CD55 critical for virus binding. *J Virol* **69:** 5497-5501.

Cole, L.B., Chu, N., Kilpatrick, J.M., Volanakis, J.E., Narayana, S.V., and Babu, Y.S. 1997. Structure of diisopropyl fluorophosphate-inhibited factor D. *Acta Crystallogr D Biol Crystallogr* **53:** 143-150.

Cole, L.B., Kilpatrick, J.M., Chu, N., and Babu, Y.S. 1998. Structure of 3,4-dichloroisocoumarin-inhibited factor D. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5):** 711-717.

Collins, J.F., and Coulson, A.F. 1990. Significance of protein sequence similarities. *Methods Enzymol* **183:** 474-487.

Colten, H.R. 1994. Immunology. Drawing a double-edged sword. *Nature* **371:** 474-475.

Copley, R.R., Doerks, T., Letunic, I., and Bork, P. 2002. Protein domain analysis in the era of complete genomes. *FEBS Lett* **513:** 129-134.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16:** 10881-10890.

Criado Garcia, O., Sanchez-Corral, P., and Rodriguez de Cordoba, S. 1995. Isoforms of human C4b-binding protein. II. Differential modulation of the C4BPA and C4BPB genes by acute phase cytokines. *J Immunol* **155:** 4037-4043.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. 2004. WebLogo: a sequence logo generator. *Genome Res* **14:** 1188-1190.

Cucnik, S., Bozic, B., Kveder, T., Tomsic, M., and Rozman, B. 2005. Avidity of Anti-{beta}2-Glycoprotein I and Thrombosis or Pregnancy Loss in Patients with Antiphospholipid Syndrome. *Ann N Y Acad Sci* **1051:** 141-147.

Cucnik, S., Kveder, T., Krizaj, I., Rozman, B., and Bozic, B. 2004a. High avidity anti-beta 2-glycoprotein I antibodies in patients with antiphospholipid syndrome. *Ann Rheum Dis* **63:** 1478-1482.

Cucnik, S., Kveder, T., Rozman, B., and Bozic, B. 2004b. Binding of high-avidity anti-beta2-glycoprotein I antibodies. *Rheumatology (Oxford)* **43:** 1353-1356.

da Silva, R.P., Hall, B.F., Joiner, K.A., and Sacks, D.L. 1988. CR1 mediates binding of L. major metacyclic promastigotes to human macrophages. *Mem Inst Oswaldo Cruz* **83 Suppl 1:** 459-463.

Da Silva, R.P., Hall, B.F., Joiner, K.A., and Sacks, D.L. 1989. CR1, the C3b receptor, mediates binding of infective Leishmania major metacyclic promastigotes to human macrophages. *J Immunol* **143:** 617-622.

Dahlback, B., Smith, C.A., and Muller-Eberhard, H.J. 1983. Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proc Natl Acad Sci U S A* **80:** 3461-3465.

Daoudaki, M.E., Becherer, J.D., and Lambris, J.D. 1988. A 34-amino acid peptide of the third component of complement mediates properdin binding. *J Immunol* **140:** 1577-1580.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K., and McDaniel, L.S. 2004a. Dual roles of PspC, a surface protein of Streptococcus pneumoniae, in binding human secretory IgA and factor H. *J Immunol* **173:** 471-477.

Dave, S., Pangburn, M.K., Pruitt, C., and McDaniel, L.S. 2004b. Interaction of human factor H with PspC of Streptococcus pneumoniae. *Indian J Med Res* **119 Suppl:** 66-73.

Delcayre, A.X., Salas, F., Mathur, S., Kovats, K., Lotz, M., and Lernhardt, W. 1991. Epstein Barr virus/complement C3d receptor is an interferon alpha receptor. *Embo J* **10:** 919-926.

Dempsey, P.W., Allison, M.E., Akkaraju, S., Goodnow, C.C., and Fearon, D.T. 1996. C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science* **271:** 348-350.

Diaz, A., Ferreira, A., and Sim, R.B. 1997. Complement evasion by Echinococcus granulosus: sequestration of host factor H in the hydatid cyst wall. *J Immunol* **158:** 3779-3786.

Discipio, R.G., Jenner, L., Thirup, S., Sottrup-Jensen, L., Nyborg, J., and Stura, E. 1998. Crystallization of human complement component C5. *Acta Crystallogr D Biol Crystallogr* **54:** 643-646.

DiScipio, R.G., Linton, S.M., and Rushmere, N.K. 1999. Function of the factor I modules (FIMS) of human complement component C6. *J Biol Chem* **274:** 31811-31818.

DiScipio, R.G., Smith, C.A., Muller-Eberhard, H.J., and Hugli, T.E. 1983. The activation of human complement component C5 by a fluid phase C5 convertase. *J Biol Chem* **258:** 10629-10636.

Do, C.B., Mahabhashyam, M.S., Brudno, M., and Batzoglou, S. 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res* **15:** 330-340.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32:** W665-667.

Dolmer, K., Huang, W., and Gettins, P.G. 2000. NMR solution structure of complement-like repeat CR3 from the low density lipoprotein receptor-related protein. Evidence for specific binding to the receptor binding domain of human alpha(2)-macroglobulin. *J Biol Chem* **275:** 3264-3269.

Dorig, R.E., Marcil, A., Chopra, A., and Richardson, C.D. 1993. The human CD46 molecule is a receptor for measles virus (Edmonston strain). *Cell* **75:** 295-305.

Dragon-Durey, M.A., Fremeaux-Bacchi, V., Loirat, C., Blouin, J., Niaudet, P., Deschenes, G., Coppo, P., Herman Fridman, W., and Weiss, L. 2004. Heterozygous and homozygous factor h deficiencies associated with hemolytic uremic syndrome or membranoproliferative glomerulonephritis: report and genetic analysis of 16 cases. *J Am Soc Nephrol* **15:** 787-795.

Dunlop, L.R., Oehlberg, K.A., Reid, J.J., Avci, D., and Rosengard, A.M. 2003. Variola virus immune evasion proteins. *Microbes Infect* **5:** 1049-1056.

Durbin, R., Eddy, S., Krogh, A., Mitchison, G. 1998. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge.

Duthy, T.G., Ormsby, R.J., Giannakis, E., Ogunniyi, A.D., Stroeher, U.H., Paton, J.C., and Gordon, D.L. 2002. The human complement regulator factor H binds pneumococcal surface protein PspC via short consensus repeats 13 to 15. *Infect Immun* **70:** 5604-5611.

Edwards, A.O., Ritter, R., 3rd, Abel, K.J., Manning, A., Panhuysen, C., and Farrer, L.A. 2005. Complement factor H polymorphism and age-related macular degeneration. *Science* **308:** 421-424. Epub 2005 Mar 2010.

Erbe, D.V., Wolitzky, B.A., Presta, L.G., Norton, C.R., Ramos, R.J., Burns, D.K., Rumberger, J.M., Rao, B.N., Foxall, C., Brandley, B.K., et al. 1992. Identification of an E-selectin region critical for carbohydrate recognition and cell adhesion. *J Cell Biol* **119:** 215-227.

Esparza-Gordillo, J., Goicoechea de Jorge, E., Buil, A., Carreras Berges, L., Lopez-Trascasa, M., Sanchez-Corral, P., and Rodriguez de Cordoba, S. 2005. Predisposition to atypical hemolytic uremic syndrome involves the concurrence of different susceptibility alleles in the regulators of complement activation gene cluster in 1q32. *Hum Mol Genet* **14:** 703-712. Epub 2005 Jan 2020.

Farries, T.C., and Atkinson, J.P. 1991. Evolution of the complement system. *Immunol Today* **12:** 295-300.

Feinberg, H., Uitdehaag, J.C., Davies, J.M., Wallis, R., Drickamer, K., and Weis, W.I. 2003. Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2. *Embo J* **22:** 2348-2359.

Feng, D.F., and Doolittle, R.F. 1987. Progressive Sequence Alignment as a Prerequisite to Correct Phylogenetic Trees. *Journal of Molecular Evolution* **25:** 351-360.

Fingeroth, J.D., Weis, J.J., Tedder, T.F., Strominger, J.L., Biro, P.A., and Fearon, D.T. 1984. Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. *Proc Natl Acad Sci U S A* **81:** 4510-4514.

Fisher, S.A., Rivera, A., Fritsche, L.G., Babadjanova, G., Petrov, S., and Weber, B.H. 2006. Assessment of the contribution of CFH and chromosome 10q26 AMD susceptibility loci in a Russian population isolate. *Br J Ophthalmol.*

Fleisig, A.J., and Verrier, E.D. 2005. Pexelizumab -- a C5 complement inhibitor for use in both acute myocardial infarction and cardiac surgery with cardiopulmonary bypass. *Expert Opin Biol Ther* **5:** 833-839.

Fletcher, C.M., Harrison, R.A., Lachmann, P.J., and Neuhaus, D. 1994. Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure* **2:** 185-199.

Fraczkiewicz, R., and Braun, W. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* **19:** 319-333.

Fredslund, F., Jenner, L., Husted, L.B., Nyborg, J., Andersen, G.R., and Sottrup-Jensen, L. 2006. The structure of bovine complement component 3 reveals the basis for thioester function. *J Mol Biol* **361:** 115-127.

Fremeaux-Bacchi, V., Dragon-Durey, M.A., Blouin, J., Vigneau, C., Kuypers, D., Boudailliez, B., Loirat, C., Rondeau, E., and Fridman, W.H. 2004. Complement factor I: a susceptibility gene for atypical haemolytic uraemic syndrome. *J Med Genet* **41:** e84.

Fujisaku, A., Harley, J.B., Frank, M.B., Gruner, B.A., Frazier, B., and Holers, V.M. 1989. Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus receptor. *J Biol Chem* **264:** 2118-2125.

Gabdoulline, R.R., Wade, R.C., and Walther, D. 1999. MolSurfer: two-dimensional maps for navigating three-dimensional structures of proteins. *Trends Biochem Sci* **24:** 285-287.

Gabdoulline, R.R., Wade, R.C., and Walther, D. 2003. MolSurfer: A macromolecular interface navigator. *Nucleic Acids Res* **31:** 3349-3351.

Gaboriaud, C., Juanhuix, J., Gruez, A., Lacroix, M., Darnault, C., Pignol, D., Verger, D., Fontecilla-Camps, J.C., and Arlaud, G.J. 2003. The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J Biol Chem* **278:** 46974-46982.

Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G.J., and Fontecilla-Camps, J.C. 2000. Crystal structure of the catalytic domain of human complement c1s: a serine protease with a handle. *Embo J* **19:** 1755-1765.

Gaggar, A., Shayakhmetov, D.M., and Lieber, A. 2003. CD46 is a cellular receptor for group B adenoviruses. *Nat Med* **9:** 1408-1412.

Gaggar, A., Shayakhmetov, D.M., Liszewski, M.K., Atkinson, J.P., and Lieber, A. 2005. Localization of regions in CD46 that interact with adenovirus. *J Virol* **79:** 7503-7513.

Gal, P., Harmat, V., Kocsis, A., Bian, T., Barna, L., Ambrus, G., Vegh, B., Balczer, J., Sim, R.B., Naray-Szabo, G., et al. 2005. A True Autoactivating Enzyme: STRUCTURAL INSIGHT INTO MANNOSE-BINDING LECTIN-ASSOCIATED SERINE PROTEASE-2 ACTIVATIONS. *J Biol Chem* **280:** 33435-33444.

Ganesh, V.K., Muthuvel, S.K., Smith, S.A., Kotwal, G.J., and Murthy, K.H. 2005. Structural Basis for Antagonism by Suramin of Heparin Binding to Vaccinia Complement Protein(,). *Biochemistry* **44:** 10757-10765.

Ganesh, V.K., Smith, S.A., Kotwal, G.J., and Murthy, K.H. 2004. Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation. *Proc Natl Acad Sci U S A* **101:** 8924-8929. Epub 2004 Jun 8923.

Garbuzynskiy, S.O., Melnik, B.S., Lobanov, M.Y., Finkelstein, A.V., and Galzitskaya, O.V. 2005. Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray- and NMR-resolved protein structures? *Proteins* **60:** 139-147.

Garcia de Frutos, P., Hardig, Y., and Dahlback, B. 1995. Serum amyloid P component binding to C4b-binding protein. *J Biol Chem* **270:** 26950-26955.

Ghiran, I., Barbashov, S.F., Klickstein, L.B., Tas, S.W., Jensenius, J.C., and Nicholson-Weller, A. 2000. Complement receptor 1/CD35 is a receptor for mannan-binding lectin. *J Exp Med* **192:** 1797-1808.

Giannakis, E., Jokiranta, T.S., Male, D.A., Ranganathan, S., Ormsby, R.J., Fischetti, V.A., Mold, C., and Gordon, D.L. 2003. A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein. *European Journal of Immunology* **33:** 962-969.

Giannakis, E., Jokiranta, T.S., Ormsby, R.J., Duthy, T.G., Male, D.A., Christiansen, D., Fischetti, V.A., Bagley, C., Loveland, B.E., and Gordon, D.L. 2002. Identification of the streptococcal M protein binding site on membrane cofactor protein (CD46). *J Immunol* **168:** 4585-4592.

Gilbert, H.E., Aslam, M., Guthridge, J.M., Holers, V.M., and Perkins, S.J. 2006a. Extended flexible linker structures in the complement chimaeric conjugate CR2-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. *J Mol Biol* **356:** 397-412.

Gilbert, H.E., Asokan, R., Holers, V.M., and Perkins, S.J. 2006b. The 15 SCR flexible extracellular domains of human complement receptor type 2 can mediate multiple ligand and antigen interactions. *J Mol Biol* **362:** 1132-1147.

Gilbert, H.E., Eaton, J.T., Hannan, J.P., Holers, V.M., and 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.

Gilges, D., Vinit, M.A., Callebaut, I., Coulombel, L., Cacheux, V., Romeo, P.H., and Vigon, I. 2000. Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von Willebrand factor A domains. *Biochem J* **352 Pt 1:** 49-59.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. 2003. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19:** 163-164.

Gold, B., Merriam, J.E., Zernant, J., Hancox, L.S., Taiber, A.J., Gehrs, K., Cramer, K., Neel, J., Bergeron, J., Barile, G.R., et al. 2006. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* **38:** 458-462.

Goodship, T. 2004. Inherited dysregulation of the complement system. *Bull Mem Acad R Med Belg* **159:** 195-198.

Goodship, T.H., Liszewski, M.K., Kemp, E.J., Richards, A., and Atkinson, J.P. 2004. Mutations in CD46, a complement regulatory protein, predispose to atypical HUS. *Trends Mol Med* **10:** 226-231.

Greenstone, H.L., Santoro, F., Lusso, P., and Berger, E.A. 2002. Human Herpesvirus 6 and Measles Virus Employ Distinct CD46 Domains for Receptor Function. *J Biol Chem* **277:** 39112-39118.

Greer, J. 1990. Comparative modeling methods: application to the family of the mammalian serine proteases. *Proteins* **7:** 317-334.

Gregory, L.A., Thielens, N.M., Arlaud, G.J., Fontecilla-Camps, J.C., and Gaboriaud, C. 2003. X-ray structure of the Ca2+-binding interaction domain of C1s. Insights into the assembly of the C1 complex of complement. *J Biol Chem* **278:** 32157-32164.

Guex, N., and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18:** 2714-2723.

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., et al. 2001. Structural studies in solution of the recombinant N-terminal pair of short consensus/complement repeat domains of complement receptor type 2 (CR2/CD21) and interactions with its ligand C3dg. *Biochemistry* **40:** 5931-5941.

Hageman, G.S., Anderson, D.H., Johnson, L.V., Hancox, L.S., Taiber, A.J., Hardisty, L.I., Hageman, J.L., Stockman, H.A., Borchardt, J.D., Gehrs, K.M., et al. 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. Epub 2005 May 7223.

Haines, J.L., Hauser, M.A., Schmidt, S., Scott, W.K., Olson, L.M., Gallins, P., Spencer, K.L., Kwan, S.Y., Noureddine, M., Gilbert, J.R., et al. 2005. Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308:** 419-421. Epub 2005 Mar 2010.

Halstead, S.K., Humphreys, P.D., Goodfellow, J.A., Wagner, E.R., Smith, R.A., and Willison, H.J. 2005. Complement inhibition abrogates nerve terminal injury in Miller Fisher syndrome. *Ann Neurol* **58:** 203-210.

Hamann, J., Stortelers, C., Kiss-Toth, E., Vogel, B., Eichler, W., and van Lier, R.A. 1998. Characterization of the CD55 (DAF)-binding site on the seven-span transmembrane receptor CD97. *Eur J Immunol* **28:** 1701-1707.

Hamann, J., Vogel, B., van Schijndel, G.M., and van Lier, R.A. 1996. The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF). *J Exp Med* **184:** 1185-1189.

Hammel, M., Kriechbaum, M., Gries, A., Kostner, G.M., Laggner, P., and 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.

Hardig, Y., and Dahlback, B. 1996. The amino-terminal module of the C4b-binding protein beta-chain contains the protein S-binding site. *J Biol Chem* **271:** 20861-20867.

Hardig, Y., Hillarp, A., and Dahlback, B. 1997. The amino-terminal module of the C4b-binding protein alpha-chain is crucial for C4b binding and factor I-cofactor function. *Biochem J* **323 (Pt 2):** 469-475.

Hardig, Y., Rezaie, A., and Dahlback, B. 1993. High affinity binding of human vitamin K-dependent protein S to a truncated recombinant beta-chain of C4b-binding protein expressed in Escherichia coli. *J Biol Chem* **268:** 3033-3036.

Harmat, V., Gal, P., Kardos, J., Szilagyi, K., Ambrus, G., Vegh, B., Naray-Szabo, G., and Zavodszky, P. 2004. The structure of MBL-associated serine protease-2 reveals that identical substrate specificities of C1s and MASP-2 are realized through different sets of enzyme-substrate interactions. *J Mol Biol* **342:** 1533-1546.

Harris, C.L., Abbott, R.J., Smith, R.A., Morgan, B.P., and Lea, S.M. 2005. Molecular dissection of interactions between components of the alternative pathway of complement and decay accelerating factor (CD55). *J Biol Chem* **280:** 2569-2578.

Harrison, R.A., and Lachmann, P.J. 1980. Novel cleavage products of the third component of human complement. *Mol Immunol* **17:** 219-228.

Hasan, R.J., Pawelczyk, E., Urvil, P.T., Venkatarajan, M.S., Goluszko, P., Kur, J., Selvarangan, R., Nowicki, S., Braun, W.A., and Nowicki, B.J. 2002. Structure-function analysis of decay-accelerating factor: identification of residues important for binding of the Escherichia coli Dr adhesin and complement regulation. *Infect Immun* **70:** 4485-4493.

He, Y., Lin, F., Chipman, P.R., Bator, C.M., Baker, T.S., Shoham, M., Kuhn, R.J., Medof, M.E., and Rossmann, M.G. 2002. Structure of decay-accelerating factor bound to echovirus 7: a virus-receptor complex. *Proc Natl Acad Sci U S A* **99:** 10325-10329.

Heiden, W., Moeckel, G., and Brickmann, J. 1993. A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J Comput Aided Mol Des* **7:** 503-514.

Hellwage, J., Jokiranta, T.S., Friese, M.A., Wolk, T.U., Kampen, E., Zipfel, P.F., and 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.

Hellwage, J., Meri, T., Heikkila, T., Alitalo, A., Panelius, J., Lahdenne, P., Seppala, I.J., and Meri, S. 2001. The complement regulator factor H binds to the surface protein OspE of Borrelia burgdorferi. *J Biol Chem* **276:** 8427-8435.

Henderson, C.E., Bromek, K., Mullin, N.P., Smith, B.O., Uhrin, D., and Barlow, P.N. 2001. Solution structure and dynamics of the central CCP module pair of a poxvirus complement control protein. *J Mol Biol* **307:** 323-339.

Herbert, A.P., Soares, D.C., Pangburn, M.K., and Barlow, P.N. 2006a. 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., and Barlow, P.N. 2006b. 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., Uhrin, D., Lyon, M., Pangburn, M.K., and Barlow, P.N. 2006c. Disease-associated sequence variations congregate in a polyanion-recognition patch on human factor H revealed in 3D structure. *J Biol Chem.*

Higuchi, M., Kazazian, H.H., Jr., Kasch, L., Warren, T.C., McGinniss, M.J., Phillips, J.A., 3rd, Kasper, C., Janco, R., and Antonarakis, S.E. 1991. Molecular characterization of severe hemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the factor VIII gene. *Proc Natl Acad Sci U S A* **88:** 7405-7409.

Hillarp, A., and Dahlback, B. 1988. Novel subunit in C4b-binding protein required for protein S binding. *J Biol Chem* **263:** 12759-12764.

Hillarp, A., Pardo-Manuel, F., Ruiz, R.R., Rodriguez de Cordoba, S., and Dahlback, B. 1993. The human C4b-binding protein beta-chain gene. *J Biol Chem* **268:** 15017-15023.

Hillmen, P., Hall, C., Marsh, J.C., Elebute, M., Bombara, M.P., Petro, B.E., Cullen, M.J., Richards, S.J., Rollins, S.A., Mojcik, C.F., et al. 2004. Effect of eculizumab on hemolysis and transfusion requirements in patients with paroxysmal nocturnal hemoglobinuria. *N Engl J Med* **350:** 552-559.

Holers, V.M., and Kulik, L. 2006. Complement receptor 2, natural antibodies and innate immunity: Inter-relationships in B cell selection and activation. *Mol Immunol.*

Honig, B., and Nicholls, A. 1995. Classical electrostatics in biology and chemistry. *Science* **268:** 1144-1149.

Hooft, R.W., Vriend, G., Sander, C., and Abola, E.E. 1996. Errors in protein structures. *Nature* **381:** 272.

Hourcade, D., Miesner, D.R., Atkinson, J.P., and Holers, V.M. 1988. Identification of an alternative polyadenylation site in the human C3b/C4b receptor (complement receptor type 1) transcriptional unit and prediction of a secreted form of complement receptor type 1. *J Exp Med* **168:** 1255-1270.

Hourcade, D., Miesner, D.R., Bee, C., Zeldes, W., and Atkinson, J.P. 1990. Duplication and divergence of the amino-terminal coding region of the complement receptor 1 (CR1) gene. An example of concerted (horizontal) evolution within a gene. *J Biol Chem* **265:** 974-980.

Hovis, K.M., Jones, J.P., Sadlon, T., Raval, G., Gordon, D.L., and Marconi, R.T. 2006. Molecular analyses of the interaction of Borrelia hermsii FhbA with the complement regulatory proteins factor H and factor H-like protein 1. *Infect Immun* **74:** 2007-2014.

Hovis, K.M., McDowell, J.V., Griffin, L., and Marconi, R.T. 2004. Identification and characterization of a linear-plasmid-encoded factor H-binding protein (FhbA) of the relapsing fever spirochete Borrelia hermsii. *J Bacteriol* **186:** 2612-2618.

Howard, J., Justus, D.E., Totmenin, A.V., Shchelkunov, S., and Kotwal, G.J. 1998. Molecular mimicry of the inflammation modulatory proteins (IMPs) of poxviruses: evasion of the inflammatory response to preserve viral habitat. *J Leukoc Biol* **64:** 68-71.

Huber, R., Scholze, H., Paques, E.P., and Deisenhofer, J. 1980. Crystal structure analysis and molecular model of human C3a anaphylatoxin. *Hoppe Seylers Z Physiol Chem* **361:** 1389-1399.

Ichinose, A., Bottenus, R.E., and Davie, E.W. 1990. Structure of transglutaminases. *J Biol Chem* **265:** 13411-13414.

Inoue, N., Fukui, A., Nomura, M., Matsumoto, M., Nishizawa, Y., Toyoshima, K., and Seya, T. 2001. A novel chicken membrane-associated complement regulatory protein: molecular cloning and functional characterization. *J Immunol* **166:** 424-431.

Isaac, L., Aivazian, D., Taniguchi-Sidle, A., Ebanks, R.O., Farah, C.S., Florido, M.P., Pangburn, M.K., and Isenman, D.E. 1998. Native conformations of human complement components C3 and C4 show different dependencies on thioester formation. *Biochem J* **329 (Pt 3):** 705-712.

Isaacs, S.N., Kotwal, G.J., and Moss, B. 1992. Vaccinia virus complement-control protein prevents antibody-dependent complement-enhanced neutralization of infectivity and contributes to virulence. *Proc Natl Acad Sci U S A* **89:** 628-632.

Isenman, D.E., Kells, D.I., Cooper, N.R., Muller-Eberhard, H.J., and Pangburn, M.K. 1981. Nucleophilic modification of human complement protein C3: correlation of conformational changes with acquisition of C3b-like functional properties. *Biochemistry* **20:** 4458-4467.

Ishii, N., Wadsworth, W.G., Stern, B.D., Culotti, J.G., and Hedgecock, E.M. 1992. UNC-6, a laminin-related protein, guides cell and pioneer axon migrations in C. elegans. *Neuron* **9:** 873-881.

Iwata, K., Seya, T., Yanagi, Y., Pesando, J.M., Johnson, P.M., Okabe, M., Ueda, S., Ariga, H., and Nagasawa, S. 1995. Diversity of sites for measles virus binding and for inactivation of complement C3b and C4b on membrane cofactor protein CD46. *J Biol Chem* **270:** 15148-15152.

Janssen, B.J., Christodoulidou, A., McCarthy, A., Lambris, J.D., and Gros, P. 2006. Structure of C3b reveals conformational changes that underlie complement activity. *Nature* **444:** 213-216.

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

Jarva, H., Hellwage, J., Jokiranta, T.S., Lehtinen, M.J., Zipfel, P.F., and Meri, S. 2004. The group B streptococcal beta and pneumococcal Hic proteins are structurally related immune evasion molecules that bind the complement inhibitor factor H in an analogous fashion. *J Immunol* **172:** 3111-3118.

Jarva, H., Janulczyk, R., Hellwage, J., Zipfel, P.F., Bjorck, L., and Meri, S. 2002. Streptococcus pneumoniae evades complement attack and opsonophagocytosis by expressing the pspC locus-encoded Hic protein that binds to short consensus repeats 8-11 of factor H. *J Immunol* **168:** 1886-1894.

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

Jenkins, H.T., Mark, L., Ball, G., Persson, J., Lindahl, G., Uhrin, D., Blom, A.M., and Barlow, P.N. 2006. Human C4b-binding Protein, Structural Basis for Interaction with Streptococcal M Protein, a Major Bacterial Virulence Factor. *J Biol Chem* **281:** 3690-3697.

Jha, P., and Kotwal, G.J. 2003. Vaccinia complement control protein: multi-functional protein and a potential wonder drug. *J Biosci* **28:** 265-271.

Jing, H., Babu, Y.S., Moore, D., Kilpatrick, J.M., Liu, X.Y., Volanakis, J.E., and Narayana, S.V. 1998. Structures of native and complexed complement factor D: implications of the atypical His57 conformation and self-inhibitory loop in the regulation of specific serine protease activity. *J Mol Biol* **282:** 1061-1081.

Jing, H., Macon, K.J., Moore, D., DeLucas, L.J., Volanakis, J.E., and Narayana, S.V. 1999. Structural basis of profactor D activation: from a highly flexible zymogen to a novel self-inhibited serine protease, complement factor D. *Embo J* **18:** 804-814.

Jing, H., Xu, Y., Carson, M., Moore, D., Macon, K.J., Volanakis, J.E., and Narayana, S.V. 2000. New structural motifs on the chymotrypsin fold and their potential roles in complement factor B. *Embo J* **19:** 164-173.

Johnson, L.V., Leitner, W.P., Staples, M.K., and Anderson, D.H. 2001. Complement activation and inflammatory processes in Drusen formation and age related macular degeneration. *Exp Eye Res* **73:** 887-896.

Jokiranta, T.S., Hellwage, J., Koistinen, V., Zipfel, P.F., and 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.

Kabsch, W., and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22:** 2577-2637.

Kaessmann, H., Zollner, S., Nekrutenko, A., and Li, W.H. 2002. Signatures of domain shuffling in the human genome. *Genome Res* **12:** 1642-1650.

Kallstrom, H., Blackmer Gill, D., Albiger, B., Liszewski, M.K., Atkinson, J.P., and Jonsson, A.B. 2001. Attachment of Neisseria gonorrhoeae to the cellular pilus receptor CD46: identification of domains important for bacterial adherence. *Cell Microbiol* **3:** 133-143.

Kallstrom, H., Islam, M.S., Berggren, P.O., and Jonsson, A.B. 1998. Cell signaling by the type IV pili of pathogenic Neisseria. *J Biol Chem* **273:** 21777-21782.

Kallstrom, H., Liszewski, M.K., Atkinson, J.P., and Jonsson, A.B. 1997. Membrane cofactor protein (MCP or CD46) is a cellular pilus receptor for pathogenic Neisseria. *Mol Microbiol* **25:** 639-647.

Kaplan, M. 2002. Eculizumab (Alexion). *Curr Opin Investig Drugs* **3:** 1017-1023.

Karnauchow, T.M., Dawe, S., Lublin, D.M., and Dimock, K. 1998. Short consensus repeat domain 1 of decay-accelerating factor is required for enterovirus 70 binding. *J Virol* **72:** 9380-9383.

Kaupmann, K., Huggel, K., Heid, J., Flor, P.J., Bischoff, S., Mickel, S.J., McMaster, G., Angst, C., Bittiger, H., Froestl, W., et al. 1997. Expression cloning of GABA(B) receptors uncovers similarity to metabotropic glutamate receptors. *Nature* **386:** 239-246.

Kelley, L.A., Gardner, S.P., and Sutcliffe, M.J. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng* **9:** 1063-1065.

Kemper, C., Zipfel, P.F., and Gigli, I. 1998. The complement cofactor protein (SBP1) from the barred sand bass (Paralabrax nebulifer) mediates overlapping regulatory activities of both human C4b binding protein and factor H. *J Biol Chem* **273:** 19398-19404.

Kieffer, B., Driscoll, P.C., Campbell, I.D., Willis, A.C., van der Merwe, P.A., and Davis, S.J. 1994. Three-dimensional solution structure of the extracellular region of the complement regulatory protein CD59, a new cell-surface protein domain related to snake venom neurotoxins. *Biochemistry* **33:** 4471-4482.

Kim, S., Narayana, S.V., and Volanakis, J.E. 1995a. Crystal structure of a complement factor D mutant expressing enhanced catalytic activity. *J Biol Chem* **270:** 24399-24405.

Kim, Y.U., Kinoshita, T., Molina, H., Hourcade, D., Seya, T., Wagner, L.M., and Holers, V.M. 1995b. Mouse complement regulatory protein Crry/p65 uses the specific mechanisms of both human decay-accelerating factor and membrane cofactor protein. *J Exp Med* **181:** 151-159.

Kirkitadze, M.D., and Barlow, P.N. 2001. Structure and flexibility of the multiple domain proteins that regulate complement activation. *Immunol Rev* **180:** 146-161.

Kirkitadze, M.D., Dryden, D.T., Kelly, S.M., Price, N.C., Wang, X., Krych, M., Atkinson, J.P., and Barlow, P.N. 1999a. Co-operativity between modules within a C3b-binding site of complement receptor type 1. *FEBS Lett* **459:** 133-138.

Kirkitadze, M.D., Henderson, C., Price, N.C., Kelly, S.M., Mullin, N.P., Parkinson, J., Dryden, D.T., and Barlow, P.N. 1999b. Central modules of the vaccinia virus complement control protein are not in extensive contact. *Biochem J* **344 Pt 1:** 167-175.

Kirkitadze, M.D., Krych, M., Uhrin, D., Dryden, D.T., Smith, B.O., Cooper, A., Wang, X., Hauhart, R., Atkinson, J.P., and Barlow, P.N. 1999c. Independently melting modules and highly structured intermodular junctions within complement receptor type 1. *Biochemistry* **38:** 7019-7031.

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., et al. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science* **308:** 385-389. Epub 2005 Mar 2010.

Klickstein, L.B., Barbashov, S.F., Liu, T., Jack, R.M., and Nicholson-Weller, A. 1997. Complement receptor type 1 (CR1, CD35) is a receptor for C1q. *Immunity* **7:** 345-355.

Klickstein, L.B., Bartow, T.J., Miletic, V., Rabson, L.D., Smith, J.A., and Fearon, D.T. 1988. Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CR1, CD35) by deletion mutagenesis. *J Exp Med* **168:** 1699-1717.

Klickstein, L.B., Wong, W.W., Smith, J.A., Weis, J.H., Wilson, J.G., and Fearon, D.T. 1987. Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristics of C3/C4 binding proteins. *J Exp Med* **165:** 1095-1112.

Kolinski, A., Betancourt, M.R., Kihara, D., Rotkiewicz, P., and Skolnick, J. 2001. Generalized comparative modeling (GENECOMP): a combination of sequence comparison, threading, and lattice modeling for protein structure prediction and refinement. *Proteins* **44:** 133-149.

Kotarsky, H., Hellwage, J., Johnsson, E., Skerka, C., Svensson, H.G., Lindahl, G., Sjobring, U., and 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.

Kotwal, G.J., Isaacs, S.N., McKenzie, R., Frank, M.M., and Moss, B. 1990. Inhibition of the complement cascade by the major secretory protein of vaccinia virus. *Science* **250:** 827-830.

Kotwal, G.J., and Moss, B. 1988. Vaccinia virus encodes a secretory polypeptide structurally related to complement control proteins. *Nature* **335:** 176-178.

Kraiczy, P., Skerka, C., Brade, V., and Zipfel, P.F. 2001a. Further characterization of complement regulator-acquiring surface proteins of Borrelia burgdorferi. *Infect Immun* **69:** 7800-7809.

Kraiczy, P., Skerka, C., Kirschfink, M., Brade, V., and Zipfel, P.F. 2001b. Immune evasion of Borrelia burgdorferi by acquisition of human complement regulators FHL-1/reconectin and Factor H. *Eur J Immunol* **31:** 1674-1684.

Kraulis, P.J. 1991. Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J Appl Crystallogr* **24:** 946-950.

Kraus, D.M., Elliott, G.S., Chute, H., Horan, T., Pfenninger, K.H., Sanford, S.D., Foster, S., Scully, S., Welcher, A.A., and Holers, V.M. 2006. CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J Immunol* **176:** 4419-4430.

Krushkal, J., Bat, O., and Gigli, I. 2000. Evolutionary relationships among proteins encoded by the regulator of complement activation gene cluster. *Mol Biol Evol* **17:** 1718-1730.

Krych, M., Clemenza, L., Howdeshell, D., Hauhart, R., Hourcade, D., and Atkinson, J.P. 1994. Analysis of the functional domains of complement receptor type 1 (C3b/C4b receptor; CD35) by substitution mutagenesis. *J Biol Chem* **269:** 13273-13278.

Krych, M., Hauhart, R., and Atkinson, J.P. 1998. Structure-function analysis of the active sites of complement receptor type 1. *J Biol Chem* **273:** 8623-8629.

Krych, M., Hourcade, D., and Atkinson, J.P. 1991. Sites within the complement C3b/C4b receptor important for the specificity of ligand binding. *Proc Natl Acad Sci U S A* **88:** 4353-4357.

Krych-Goldberg, M., and Atkinson, J.P. 2001a. Structure-function relationships of complement receptor type 1. In *Immunol Rev*, pp. 112-122.

Krych-Goldberg, M., and Atkinson, J.P. 2001b. Structure-function relationships of complement receptor type 1. *Immunol Rev* **180:** 112-122.

Krych-Goldberg, M., Hauhart, R.E., Porzukowiak, T., and Atkinson, J.P. 2005. Synergy between Two Active Sites of Human Complement Receptor Type 1 (CD35) in Complement Regulation: Implications for the Structure of the Classical Pathway C3 Convertase and Generation of More Potent Inhibitors. *J Immunol* **175:** 4528-4535.

Krych-Goldberg, M., Hauhart, R.E., Subramanian, V.B., Yurcisin, B.M., 2nd, Crimmins, D.L., Hourcade, D.E., and Atkinson, J.P. 1999. Decay accelerating activity of complement receptor type 1 (CD35). Two active sites are required for dissociating C5 convertases. *J Biol Chem* **274:** 31160-31168.

Krych-Goldberg, M., Moulds, J.M., and Atkinson, J.P. 2002. Human complement receptor type 1 (CR1) binds to a major malarial adhesin. *Trends Mol Med* **8:** 531-537.

Kurtz, C.B., O'Toole, E., Christensen, S.M., and Weis, J.H. 1990. The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene transcripts predicts two distinct gene products that share homologous domains with both human CR2 and CR1. *J Immunol* **144:** 3581-3591.

Kuttner-Kondo, L., Medof, M.E., Brodbeck, W., and Shoham, M. 1996. Molecular modeling and mechanism of action of human decay-accelerating factor. *Protein Eng* **9:** 1143-1149.

Kuttner-Kondo, L.A., Mitchell, L., Hourcade, D.E., and Medof, M.E. 2001. Characterization of the active sites in decay-accelerating factor. *J Immunol* **167:** 2164-2171.

Lachmann, P.J., Pangburn, M.K., and Oldroyd, R.G. 1982. Breakdown of C3 after complement activation. Identification of a new fragment C3g, using monoclonal antibodies. *J Exp Med* **156:** 205-216.

Lam, T.T., Hausen, B., Hook, L., Lau, M., Higgins, J., Christians, U., Jacobsen, W., Baluom, M., Duthaler, R., Katopodis, A., et al. 2005. The effect of soluble complement receptor type 1 on acute humoral xenograft rejection in hDAF-transgenic pig-to-primate life-supporting kidney xenografts. *Xenotransplantation* **12:** 20-29.

Lambris, J.D., Avila, D., Becherer, J.D., and Muller-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.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. 1993. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr* **26:** 283-291.

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

Lau, L.I., Chen, S.J., Cheng, C.Y., Yen, M.Y., Lee, F.L., Lin, M.W., Hsu, W.M., and Wei, Y.H. 2006. Association of the Y402H polymorphism in complement factor H gene and neovascular age-related macular degeneration in Chinese patients. *Invest Ophthalmol Vis Sci* **47:** 3242-3246.

Lau, W.L., and Scholnick, S.B. 2003. Identification of two new members of the CSMD gene family small star, filled. *Genomics* **82:** 412-415.

Lazar, H.L., Bokesch, P.M., van Lenta, F., Fitzgerald, C., Emmett, C., Marsh, H.C., Jr., and Ryan, U. 2004. Soluble human complement receptor 1 limits ischemic damage in cardiac surgery patients at high risk requiring cardiopulmonary bypass. *Circulation* **110:** II274-279.

Le Bouguenec, C., Lalioui, L., du Merle, L., Jouve, M., Courcoux, P., Bouzari, S., Selvarangan, R., Nowicki, B.J., Germani, Y., Andremont, A., et al. 2001. Characterization of AfaE adhesins produced by extraintestinal and intestinal human Escherichia coli isolates: PCR assays for detection of Afa adhesins that do or do not recognize Dr blood group antigens. *J Clin Microbiol* **39:** 1738-1745.

Lea, S.M., Powell, R.M., McKee, T., Evans, D.J., Brown, D., Stuart, D.I., and van der Merwe, P.A. 1998. Determination of the affinity and kinetic constants for the interaction between the human virus echovirus 11 and its cellular receptor, CD55. *J Biol Chem* **273:** 30443-30447.

Lehtinen, M.J., Hagglund, H., and Jokiranta, T.S. 2006. Reason for the association between Factor H mutations and atypical hemolytic-uremic syndrome is not in all cases impaired heparin binding of the mutant protein. *Molecular Immunology* **43:** 166-166.

Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P., and Bork, P. 2004. SMART 4.0: towards genomic data integration. *Nucleic Acids Res* **32:** D142-144.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., and Bork, P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* **30:** 242-244.

Leung, E., Blom, A.M., Clemenza, L., and Isenman, D.E. 2006. The complement regulator C4b-binding protein (C4BP) interacts with both the C4c and C4dg subfragments of the parent C4b ligand: evidence for synergy in C4BP subsite binding. *Biochemistry* **45:** 8378-8392.

Levashina, E.A., Moita, L.F., Blandin, S., Vriend, G., Lagueux, M., and Kafatos, F.C. 2001. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, Anopheles gambiae. *Cell* **104:** 709-718.

Levitt, M. 1992. Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol* **226:** 507-533.

Lichtarge, O., Bourne, H.R., and Cohen, F.E. 1996. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* **257:** 342-358.

Lindahl, G., Sjobring, U., and Johnsson, E. 2000. Human complement regulators: a major target for pathogenic microorganisms. *Curr Opin Immunol* **12:** 44-51.

Liszewski, M.K., Leung, M., Cui, W., Subramanian, V.B., Parkinson, J., Barlow, P.N., Manchester, M., and Atkinson, J.P. 2000. Dissecting sites important for complement regulatory activity in membrane cofactor protein (MCP; CD46). *J Biol Chem* **275:** 37692-37701.

Liszewski, M.K., Tedja, I., and Atkinson, J.P. 1994. Membrane cofactor protein (CD46) of complement. Processing differences related to alternatively spliced cytoplasmic domains. *J Biol Chem* **269:** 10776-10779.

Lorenzen, I., Dingley, A.J., Jacques, Y., and Grotzinger, J. 2006. The Structure of the Interleukin-15{alpha} Receptor and Its Implications for Ligand Binding. *J Biol Chem* **281:** 6642-6647.

Low, P.J., Ai, R., and Ogata, R.T. 1999. Active sites in complement components C5 and C3 identified by proximity to indels in the C3/4/5 protein family. *J Immunol* **162:** 6580-6588.

Lozahic, S., Christiansen, D., Manie, S., Gerlier, D., Billard, M., Boucheix, C., and Rubinstein, E. 2000. CD46 (membrane cofactor protein) associates with multiple beta1 integrins and tetraspans. *Eur J Immunol* **30:** 900-907.

Lublin, D.M., Liszewski, M.K., Post, T.W., Arce, M.A., Le Beau, M.M., Rebentisch, M.B., Lemons, L.S., Seya, T., and Atkinson, J.P. 1988. Molecular cloning and chromosomal localization of human membrane cofactor protein (MCP). Evidence for inclusion in the multigene family of complement-regulatory proteins. *J Exp Med* **168:** 181-194.

Lukacik, P., Roversi, P., White, J., Esser, D., Smith, G.P., Billington, J., Williams, P.A., Rudd, P.M., Wormald, M.R., Harvey, D.J., et al. 2004. Complement regulation at the molecular level: the

structure of decay-accelerating factor. *Proc Natl Acad Sci U S A* **101:** 1279-1284. Epub 2004 Jan 1220.

Maisner, A., Alvarez, J., Liszewski, M.K., Atkinson, D.J., Atkinson, J.P., Herrler, G. 1996. The N-glycan of the SCR 2 region is essential for membrane cofactor protein (CD46) to function as a measles virus receptor. *J Virol* **70:** 4973-4977.

Malhotra, R., Ward, M., Sim, R.B., and Bird, M.I. 1999. Identification of human complement Factor H as a ligand for L-selectin. *Biochem J* **341:** 61-69.

Manchester, M., Gairin, J.E., Patterson, J.B., Alvarez, J., Liszewski, M.K., Eto, D.S., Atkinson, J.P., and Oldstone, M.B. 1997. Measles virus recognizes its receptor, CD46, via two distinct binding domains within SCR1-2. *Virology* **233:** 174-184.

Martin, D.R., Yuryev, A., Kalli, K.R., Fearon, D.T., and Ahearn, J.M. 1991. Determination of the structural basis for selective binding of Epstein-Barr virus to human complement receptor type 2. *J Exp Med* **174:** 1299-1311.

Martinez, A., Pio, R., Zipfel, P.F., and Cuttitta, F. 2003. Mapping of the adrenomedullin-binding domains in human complement factor H. *Hypertens Res* **26 Suppl:** S55-59.

Marti-Renom, M.A., Madhusudhan, M.S., Fiser, A., Rost, B., and Sali, A. 2002. Reliability of assessment of protein structure prediction methods. *Structure* **10:** 435-440.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **29:** 291-325.

Marx, J. 2006. Genetics. A clearer view of macular degeneration. *Science* **311:** 1704-1705.

Mastellos, D., Germenis, A.E., and Lambris, J.D. 2005. Complement: an inflammatory pathway fulfilling multiple roles at the interface of innate immunity and development. *Curr Drug Targets Inflamm Allergy* **4:** 125-127.

Mastellos, D., and Lambris, J.D. 2002. Complement: more than a 'guard' against invading pathogens? *Trends Immunol* **23:** 485-491.

Mastellos, D., Morikis, D., Isaacs, S.N., Holland, M.C., Strey, C.W., and Lambris, J.D. 2003. Complement: structure, functions, evolution, and viral molecular mimicry. *Immunol Res* **27:** 367-386.

Matsumoto, A.K., Kopicky-Burd, J., Carter, R.H., Tuveson, D.A., Tedder, T.F., and Fearon, D.T. 1991. Intersection of the complement and immune systems: a signal transduction complex of the B lymphocyte-containing complement receptor type 2 and CD19. *J Exp Med* **173:** 55-64.

McDowell, J.V., Lankford, J., Stamm, L., Sadlon, T., Gordon, D.L., and Marconi, R.T. 2005. Demonstration of factor H-like protein 1 binding to Treponema denticola, a pathogen associated with periodontal disease in humans. *Infect Immun* **73:** 7126-7132.

McDowell, J.V., Wolfgang, J., Senty, L., Sundy, C.M., Noto, M.J., and Marconi, R.T. 2004. Demonstration of the involvement of outer surface protein E coiled coil structural domains and higher order structural elements in the binding of infection-induced antibody and the complement-regulatory protein, factor H. *J Immunol* **173:** 7471-7480.

McDowell, J.V., Wolfgang, J., Tran, E., Metts, M.S., Hamilton, D., and Marconi, R.T. 2003. Comprehensive analysis of the factor h binding capabilities of borrelia species associated with lyme disease: delineation of two distinct classes of factor h binding proteins. *Infect Immun* **71:** 3597-3602.

McGuffin, L.J., Bryson, K., and Jones, D.T. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16:** 404-405.

McKenzie, R., Kotwal, G.J., Moss, B., Hammer, C.H., and Frank, M.M. 1992. Regulation of complement activity by vaccinia virus complement-control protein. *J Infect Dis* **166:** 1245-1250.

Medof, M.E., Iida, K., Mold, C., and Nussenzweig, V. 1982. Unique role of the complement receptor CR1 in the degradation of C3b associated with immune complexes. *J Exp Med* **156:** 1739-1754.

Medof, M.E., Lublin, D.M., Holers, V.M., Ayers, D.J., Getty, R.R., Leykam, J.F., Atkinson, J.P., and Tykocinski, M.L. 1987a. Cloning and characterization of cDNAs encoding the complete sequence of decay-accelerating factor of human complement. *Proc Natl Acad Sci U S A* **84:** 2007-2011.

Medof, M.E., Walter, E.I., Roberts, W.L., Haas, R., and Rosenberry, T.L. 1986. Decay accelerating factor of complement is anchored to cells by a C-terminal glycolipid. *Biochemistry* **25:** 6740-6747.

Medof, M.E., Walter, E.I., Rutgers, J.L., Knowles, D.M., and Nussenzweig, V. 1987b. Identification of the complement decay-accelerating factor (DAF) on epithelium and glandular cells and in body fluids. *J Exp Med* **165:** 848-864.

Meri, T., Blom, A.M., Hartmann, A., Lenk, D., Meri, S., and Zipfel, P.F. 2004. The hyphal and yeast forms of Candida albicans bind the complement regulator C4b-binding protein. *Infect Immun* **72:** 6633-6641.

Meri, T., Cutler, S.J., Blom, A.M., Meri, S., and Jokiranta, T.S. 2006. Relapsing fever spirochetes Borrelia recurrentis and B. duttonii acquire complement regulators C4b-binding protein and factor H. *Infect Immun* **74:** 4157-4163.

Meri, T., Hartmann, A., Lenk, D., Eck, R., Wurzner, R., Hellwage, J., Meri, S., and Zipfel, P.F. 2002a. The yeast Candida albicans binds complement regulators factor H and FHL-1. *Infect Immun* **70:** 5185-5192.

Meri, T., Jokiranta, T.S., Hellwage, J., Bialonski, A., Zipfel, P.F., and Meri, S. 2002b. Onchocerca volvulus microfilariae avoid complement attack by direct binding of factor H. *J Infect Dis* **185:** 1786-1793.

Metts, M.S., McDowell, J.V., Theisen, M., Hansen, P.R., and Marconi, R.T. 2003. Analysis of the OspE determinants involved in binding of factor H and OspE-targeting antibodies elicited during Borrelia burgdorferi infection in mice. *Infect Immun* **71:** 3587-3596.

Miller, L.H., Good, M.F., and Milon, G. 1994. Malaria pathogenesis. *Science* **264:** 1878-1883.

Miwa, T., Nonaka, M., Okada, N., Wakana, S., Shiroishi, T., and Okada, H. 1998. Molecular cloning of rat and mouse membrane cofactor protein (MCP, CD46): preferential expression in testis and close linkage between the mouse Mcp and Cr2 genes on distal chromosome 1. *Immunogenetics* **48:** 363-371.

Miwa, T., Okada, N., and Okada, H. 2000. Alternative exon usage in the 3' region of a single gene generates glycosylphosphatidylinositol-anchored and transmembrane forms of rat decay-accelerating factor. *Immunogenetics* **51:** 129-137.

Mizuno, M., and Morgan, B.P. 2004. The possibilities and pitfalls for anti-complement therapies in inflammatory diseases. *Curr Drug Targets Inflamm Allergy* **3:** 87-96.

Molina, H., Wong, W., Kinoshita, T., Brenner, C., Foley, S., and Holers, V.M. 1992. Distinct receptor and regulatory properties of recombinant mouse complement receptor 1 (CR1) and Crry, the two genetic homologues of human CR1. *J Exp Med* **175:** 121-129.

Moore, M.D., Cooper, N.R., Tack, B.F., and Nemerow, G.R. 1987. Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (complement receptor type 2) of human B lymphocytes. *Proc Natl Acad Sci U S A* **84:** 9194-9198.

Morgan, B.P., and Harris, C.L. 2003. Complement therapeutics; history and current progress. *Mol Immunol* **40:** 159-170.

Mori, Y., Yang, X., Akkapaiboon, P., Okuno, T., and Yamanishi, K. 2003. Human herpesvirus 6 variant A glycoprotein H-glycoprotein L-glycoprotein Q complex associates with human CD46. *J Virol* **77:** 4992-4999.

Morley, B.J., and Campbell, R.D. 1984. Internal homologies of the Ba fragment from human complement component Factor B, a class III MHC antigen. *Embo J* **3:** 153-157.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. 1992. Stereochemical quality of protein structure coordinates. *Proteins* **12:** 345-364.

Moulds, J.M. 2002. A review of the Knops blood group: separating fact from fallacy. *Immunohematol* **18:** 1-8.

Moulds, J.M., Kassambara, L., Middleton, J.J., Baby, M., Sagara, I., Guindo, A., Coulibaly, S., Yalcouye, D., Diallo, D.A., Miller, L., et al. 2000. Identification of complement receptor one (CR1) polymorphisms in west Africa. *Genes Immun* **1:** 325-329.

Moulds, J.M., Nickells, M.W., Moulds, J.J., Brown, M.C., and Atkinson, J.P. 1991. The C3b/C4b receptor is recognized by the Knops, McCoy, Swain-langley, and York blood group antisera. *J Exp Med* **173:** 1159-1163.

Moulds, J.M., Zimmerman, P.A., Doumbo, O.K., Diallo, D.A., Atkinson, J.P., Krych-Goldberg, M., Hourcade, D.E., and Moulds, J.J. 2002. Expansion of the Knops blood group system and subdivision of Sl(a). *Transfusion* **42:** 251-256.

Moulds, J.M., Zimmerman, P.A., Doumbo, O.K., Kassambara, L., Sagara, I., Diallo, D.A., Atkinson, J.P., Krych-Goldberg, M., Hauhart, R.E., Hourcade, D.E., et al. 2001. Molecular identification of Knops blood group polymorphisms found in long homologous region D of complement receptor 1. *Blood* **97:** 2879-2885.

Muller-Eberhard, H.J. 1985. Transmembrane channel-formation by five complement proteins. *Biochem Soc Symp* **50:** 235-246.

Muller-Eberhard, H.J. 1986. The membrane attack complex of complement. *Annu Rev Immunol* **4:** 503-528.

Mullick, J., Bernet, J., Panse, Y., Hallihosur, S., Singh, A.K., and Sahu, A. 2005. Identification of complement regulatory domains in vaccinia virus complement control protein. *J Virol* **79:** 12382-12393.

Mullick, J., Bernet, J., Singh, A.K., Lambris, J.D., and Sahu, A. 2003. Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) open reading frame 4 protein (kaposica) is a functional homolog of complement control proteins. *J Virol* **77:** 3878-3881.

Mullins, R.F., Aptsiauri, N., and Hageman, G.S. 2001. Structure and composition of drusen associated with glomerulonephritis: implications for the role of complement activation in drusen biogenesis. *Eye* **15:** 390-395.

Murthy, K.H., Smith, S.A., Ganesh, V.K., Judge, K.W., Mullin, N., Barlow, P.N., Ogata, C.M., and Kotwal, G.J. 2001. Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans. *Cell* **104:** 301-311.

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

Naniche, D., Varior-Krishnan, G., Cervoni, F., Wild, T.F., Rossi, B., Rabourdin-Combe, C., and Gerlier, D. 1993. Human membrane cofactor protein (CD46) acts as a cellular receptor for measles virus. *J Virol* **67:** 6025-6032.

Narayana, S.V., Carson, M., el-Kabbani, O., Kilpatrick, J.M., Moore, D., Chen, X., Bugg, C.E., Volanakis, J.E., and DeLucas, L.J. 1994. Structure of human factor D. A complement system protein at 2.0 A resolution. *J Mol Biol* **235:** 695-708.

Neumann, H.P., Salzmann, M., Bohnert-Iwan, B., Mannuelian, T., Skerka, C., Lenk, D., Bender, B.U., Cybulla, M., Riegler, P., Konigsrainer, A., et al. 2003. Haemolytic uraemic syndrome and mutations of the factor H gene: a registry-based study of German speaking countries. *J Med Genet* **40:** 676-681.

Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11:** 281-296.

Nishida, N., Walz, T., and Springer, T.A. 2006. Structural transitions of complement component C3 and its activation products. *Proc Natl Acad Sci U S A* **103:** 19737-19742.

Nonaka, M., Miwa, T., Okada, N., and Okada, H. 1995. Multiple isoforms of guinea pig decay-accelerating factor (DAF) generated by alternative splicing. *J Immunol* **155:** 3037-3048.

Nordstrom, T., Blom, A.M., Forsgren, A., and Riesbeck, K. 2004. The emerging pathogen Moraxella catarrhalis interacts with complement inhibitor C4b binding protein through ubiquitous surface proteins A1 and A2. *J Immunol* **173:** 4598-4606.

Noris, M., Bucchioni, S., Galbusera, M., Donadelli, R., Bresin, E., Castelletti, F., Caprioli, J., Brioschi, S., Scheiflinger, F., and Remuzzi, G. 2005. Complement factor H mutation in familial thrombotic thrombocytopenic purpura with ADAMTS13 deficiency and renal involvement. *J Am Soc Nephrol* **16:** 1177-1183. Epub 2005 Mar 1130.

Norman, D.G., Barlow, P.N., Baron, M., Day, A.J., Sim, R.B., and Campbell, I.D. 1991. Three-dimensional structure of a complement control protein module in solution. *J Mol Biol* **219:** 717-725.

Notredame, C., Higgins, D.G., and Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302:** 205-217.

Nowicki, B., Hart, A., Coyne, K.E., Lublin, D.M., and Nowicki, S. 1993. Short consensus repeat-3 domain of recombinant decay-accelerating factor is recognized by Escherichia coli recombinant Dr adhesin in a model of a cell-cell interaction. *J Exp Med* **178:** 2115-2121.

O'Brien, D.P., Israel, D.A., Krishna, U., Romero-Gallo, J., Nedrud, J., Medof, M.E., Lin, F., Redline, R., Lublin, D.M., Nowicki, B.J., et al. 2006. The role of decay-accelerating factor as a receptor for Helicobacter pylori and a mediator of gastric inflammation. *J Biol Chem* **281:** 13317-13323.

Ogata, R.T., Ai, R., and Low, P.J. 1998. Active sites in complement component C3 mapped by mutations at indels. *J Immunol* **161:** 4785-4794.

Ogata, R.T., and Low, P.J. 1997. Complement-inhibiting peptides identified by proximity to indels in the C3/4/5 protein family. *J Immunol* **158:** 3852-3860.

Ogata, R.T., Mathias, P., Bradt, B.M., and Cooper, N.R. 1993. Murine C4b-binding protein. Mapping of the ligand binding site and the N-terminus of the pre-protein. *J Immunol* **150:** 2273-2280.

Okada, N., Liszewski, M.K., Atkinson, J.P., and Caparon, M. 1995. Membrane cofactor protein (CD46) is a keratinocyte receptor for the M protein of the group A streptococcus. *Proc Natl Acad Sci U S A* **92:** 2489-2493.

Okamoto, H., Umeda, S., Obazawa, M., Minami, M., Noda, T., Mizota, A., Honda, M., Tanaka, M., Koyama, R., Takagi, I., et al. 2006. Complement factor H polymorphisms in Japanese population with age-related macular degeneration. *Mol Vis* **12:** 156-158.

O'Keeffe, A.H., Green, J.L., Grainger, M., and Holder, A.A. 2005. A novel Sushi domain-containing protein of Plasmodium falciparum. *Mol Biochem Parasitol* **140:** 61-68.

O'Leary, J.M., Bromek, K., Black, G.M., Uhrinova, S., Schmitz, C., Wang, X., Krych, M., Atkinson, J.P., Uhrin, D., and Barlow, P.N. 2004. Backbone dynamics of complement control protein (CCP) modules reveals mobility in binding surfaces. *Protein Sci* **13:** 1238-1250.

Oleszewski, M., Gutwein, P., von der Lieth, W., Rauch, U., and Altevogt, P. 2000. Characterization of the L1-neurocan-binding site. Implications for L1-L1 homophilic binding. *J Biol Chem* **275:** 34478-34485.

Oran, A.E., and Isenman, D.E. 1999. Identification of residues within the 727-767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J Biol Chem* **274:** 5120-5130.

Ormsby, R.J., Jokiranta, T.S., Duthy, T.G., Griggs, K.M., Sadlon, T.A., Giannakis, E., and Gordon, D.L. 2006. Localization of the third heparin-binding site in the human complement regulator factor H1. *Mol Immunol* **43:** 1624-1632.

Ortlund, E., Parker, C.L., Schreck, S.F., Ginell, S., Minor, W., Sodetz, J.M., and Lebioda, L. 2002. Crystal structure of human complement protein C8gamma at 1.2 A resolution reveals a lipocalin fold and a distinct ligand binding site. *Biochemistry* **41:** 7030-7037.

Oshiumi, H., Shida, K., Goitsuka, R., Kimura, Y., Katoh, J., Ohba, S., Tamaki, Y., Hattori, T., Yamada, N., Inoue, N., et al. 2005. Regulator of complement activation (RCA) locus in chicken: identification of chicken RCA gene cluster and functional RCA proteins. *J Immunol* **175:** 1724-1734.

Pandiripally, V., Gregory, E., and Cue, D. 2002. Acquisition of regulators of complement activation by Streptococcus pyogenes serotype M1. *Infect Immun* **70:** 6206-6214.

Pandiripally, V., Wei, L., Skerka, C., Zipfel, P.F., and Cue, D. 2003. Recruitment of complement factor H-like protein 1 promotes intracellular invasion by group A streptococci. *Infect Immun* **71:** 7119-7128.

Pangburn, M.K., Atkinson, M.A., and Meri, S. 1991. Localization of the heparin-binding site on complement factor H. *J Biol Chem* **266:** 16847-16853.

Pawlowski, K., and Godzik, A. 2001. Surface map comparison: studying function diversity of homologous proteins. *J Mol Biol* **309:** 793-806.

Pearson, W.R., and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85:** 2444-2448.

Peitsch, M.C. 2002. About the use of protein models. *Bioinformatics* **18:** 934-938.

Perez de la Lastra, J.M., Harris, C.L., Hinchliffe, S.J., Holt, D.S., Rushmere, N.K., and Morgan, B.P. 2000. Pigs express multiple forms of decay-accelerating factor (CD55), all of which contain only three short consensus repeats. *J Immunol* **165:** 2563-2573.

Perez-Caballero, D., Garcia-Laorden, I., Cortes, G., Wessels, M.R., de Cordoba, S.R., and Alberti, S. 2004. Interaction between complement regulators and Streptococcus pyogenes: binding of C4b-binding protein and factor H/factor H-like protein 1 to M18 strains involves two different cell surface molecules. *J Immunol* **173:** 6899-6904.

Perez-Caballero, D., Gonzalez-Rubio, C., Gallardo, M.E., Vera, M., Lopez-Trascasa, M., Rodriguez de Cordoba, S., and Sanchez-Corral, P. 2001. Clustering of missense mutations in the C-terminal region of factor H in atypical hemolytic uremic syndrome. *Am J Hum Genet* **68:** 478-484.

Perkins, S.J., and 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., Haris, P.I., Sim, R.B., and 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.

Pettigrew, D.M., Williams, D.T., Kerrigan, D., Evans, D.J., Lea, S.M., and Bhella, D. 2006. Structural and functional insights into the interaction of echoviruses and decay-accelerating factor. *J Biol Chem* **281:** 5169-5177.

Pham, T., Kaul, A., Hart, A., Goluszko, P., Moulds, J., Nowicki, S., Lublin, D.M., and Nowicki, B.J. 1995. dra-related X adhesins of gestational pyelonephritis-associated Escherichia coli

recognize SCR-3 and SCR-4 domains of recombinant decay-accelerating factor. *Infect Immun* **63:** 1663-1668.

Post, T.W., Arce, M.A., Liszewski, M.K., Thompson, E.S., Atkinson, J.P., and Lublin, D.M. 1990. Structure of the gene for human complement protein decay accelerating factor. *J Immunol* **144:** 740-744.

Potts, J.R., and Campbell, I.D. 1994. Fibronectin structure and assembly. *Curr Opin Cell Biol* **6:** 648-655.

Potts, J.R., and Campbell, I.D. 1996. Structure and function of fibronectin modules. *Matrix Biol* **15:** 313-320; discussion 321.

Prasadarao, N.V., Blom, A.M., Villoutreix, B.O., and Linsangan, L.C. 2002. A novel interaction of outer membrane protein A with C4b binding protein mediates serum resistance of Escherichia coli K1. *J Immunol* **169:** 6352-6360.

Prodinger, W.M., Hellwage, J., Spruth, M., Dierich, M.P., and Zipfel, P.F. 1998. The C-terminus of factor H: monoclonal antibodies inhibit heparin binding and identify epitopes common to factor H and factor H-related proteins. *Biochem J* **331 (Pt 1):** 41-47.

Prota, A.E., Sage, D.R., Stehle, T., and Fingeroth, J.D. 2002. The crystal structure of human CD21: Implications for Epstein-Barr virus and C3d binding. *Proc Natl Acad Sci U S A* **99:** 10641-10646.

Pruitt, S.K., Bollinger, R.R., Collins, B.H., Marsh, H.C., Jr., Levin, J.L., Rudolph, A.R., Baldwin, W.M., 3rd, and Sanfilippo, F. 1997. Effect of continuous complement inhibition using soluble complement receptor type 1 on survival of pig-to-primate cardiac xenografts. *Transplantation* **63:** 900-902.

Qian, Y.M., Haino, M., Kelly, K., and Song, W.C. 1999. Structural characterization of mouse CD97 and study of its specific interaction with the murine decay-accelerating factor (DAF, CD55). *Immunology* **98:** 303-311.

Quin, L.R., Carmicle, S., Dave, S., Pangburn, M.K., Evenhuis, J.P., and McDaniel, L.S. 2005. In vivo binding of complement regulator factor H by Streptococcus pneumoniae. *J Infect Dis* **192:** 1996-2003.

Ram, S., Cullinane, M., Blom, A.M., Gulati, S., McQuillen, D.P., Monks, B.G., O'Connell, C., Boden, R., Elkins, C., Pangburn, M.K., et al. 2001. Binding of C4b-binding protein to porin: a molecular mechanism of serum resistance of Neisseria gonorrhoeae. *J Exp Med* **193:** 281-295.

Ram, S., Mackinnon, F.G., Gulati, S., McQuillen, D.P., Vogel, U., Frosch, M., Elkins, C., Guttormsen, H.K., Wetzler, L.M., Oppermann, M., et al. 1999. The contrasting mechanisms of serum resistance of Neisseria gonorrhoeae and group B Neisseria meningitidis. *Mol Immunol* **36:** 915-928.

Ram, S., McQuillen, D.P., Gulati, S., Elkins, C., Pangburn, M.K., and Rice, P.A. 1998a. Binding of complement factor H to loop 5 of porin protein 1A: a molecular mechanism of serum resistance of nonsialylated Neisseria gonorrhoeae. *J Exp Med* **188:** 671-680.

Ram, S., Sharma, A.K., Simpson, S.D., Gulati, S., McQuillen, D.P., Pangburn, M.K., and Rice, P.A. 1998b. A novel sialic acid binding site on factor H mediates serum resistance of sialylated Neisseria gonorrhoeae. *J Exp Med* **187:** 743-752.

Ranganathan, S., Male, D.A., Ormsby, R.J., Giannakis, E., and Gordon, D.L. 2000. Pinpointing the putative heparin/sialic acid-binding residues in the 'sushi' domain 7 of factor H: a molecular modeling study. *Pac Symp Biocomput*: 155-167.

Rao, N., Ferguson, D.J., Lee, S.F., and Telen, M.J. 1991. Identification of human erythrocyte blood group antigens on the C3b/C4b receptor. *J Immunol* **146:** 3502-3507.

Reid, K.B., and Day, A.J. 1989. Structure-function relationships of the complement components. *Immunol Today* **10:** 177-180.

Remuzzi, G., Ruggenenti, P., Codazzi, D., Noris, M., Caprioli, J., Locatelli, G., and Gridelli, B. 2002. Combined kidney and liver transplantation for familial haemolytic uraemic syndrome. *Lancet* **359:** 1671-1672.

Retief, J.D. 2000. Phylogenetic analysis using PHYLIP. *Methods Mol Biol* **132:** 243-258.

Rey-Campos, J., Baeza-Sanz, D., and Rodriguez de Cordoba, S. 1990. Physical linkage of the human genes coding for complement factor H and coagulation factor XIII B subunit. *Genomics* **7:** 644-646.

Richards, A., Buddles, M.R., Donne, R.L., Kaplan, B.S., Kirk, E., Venning, M.C., Tielemans, C.L., Goodship, J.A., and 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.

Richards, A., Kemp, E.J., Liszewski, M.K., Goodship, J.A., Lampe, A.K., Decorte, R., Muslumanoglu, M.H., Kavukcu, S., Filler, G., Pirson, Y., et al. 2003. Mutations in human complement regulator, membrane cofactor protein (CD46), predispose to development of familial hemolytic uremic syndrome. *Proc Natl Acad Sci U S A* **100:** 12966-12971.

Richardson, D.C., and Richardson, J.S. 1992. The kinemage: a tool for scientific communication. *Protein Sci* **1:** 3-9.

Rickert, M., Wang, X., Boulanger, M.J., Goriatcheva, N., and Garcia, K.C. 2005. The structure of interleukin-2 complexed with its alpha receptor. *Science* **308:** 1477-1480.

Rioux, P. 2001. TP-10 (AVANT Immunotherapeutics). *Curr Opin Investig Drugs* **2:** 364-371.

Rivera, A., Fisher, S.A., Fritsche, L.G., Keilhauer, C.N., Lichtner, P., Meitinger, T., and Weber, B.H. 2005. Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum Mol Genet* **14:** 3227-3236.

Rodien, P., Bremont, C., Sanson, M.L., Parma, J., Van Sande, J., Costagliola, S., Luton, J.P., Vassart, G., and Duprez, L. 1998. Familial gestational hyperthyroidism caused by a mutant thyrotropin receptor hypersensitive to human chorionic gonadotropin. *N Engl J Med* **339:** 1823-1826.

Rodriguez de Cordoba, S., Esparza-Gordillo, J., Goicoechea de Jorge, E., Lopez-Trascasa, M., and Sanchez-Corral, P. 2004. The human complement factor H: functional roles, genetic variations and disease associations. *Mol Immunol* **41:** 355-367.

Rokas, A., Williams, B.L., King, N., and Carroll, S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425:** 798-804.

Roll, P., Rudolf, G., Pereira, S., Royer, B., Scheffer, I.E., Massacrier, A., Valenti, M.P., Roeckel-Trevisiol, N., Jamali, S., Beclin, C., et al. 2006. SRPX2 mutations in disorders of language cortex and cognition. *Hum Mol Genet* **15:** 1195-1207.

Rosengard, A.M., Alonso, L.C., Korb, L.C., Baldwin, W.M., 3rd, Sanfilippo, F., Turka, L.A., and Ahearn, J.M. 1999. Functional characterization of soluble and membrane-bound forms of vaccinia virus complement control protein (VCP). *Mol Immunol* **36:** 685-697.

Rosengard, A.M., Liu, Y., Nie, Z., and Jimenez, R. 2002. Variola virus immune evasion design: expression of a highly efficient inhibitor of human complement. *Proc Natl Acad Sci U S A* **99:** 8808-8813. Epub 2002 May 8828.

Ross, G.D., and Medof, M.E. 1985. Membrane complement receptors specific for bound fragments of C3. *Adv Immunol* **37:** 217-267.

Rost, B. 1999. Twilight zone of protein sequence alignments. *Protein Eng* **12:** 85-94.

Rowe, A., Obeiro, J., Newbold, C.I., and Marsh, K. 1995. Plasmodium falciparum rosetting is associated with malaria severity in Kenya. *Infect Immun* **63:** 2323-2326.

Rowe, J.A., Moulds, J.M., Newbold, C.I., and Miller, L.H. 1997. P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* **388:** 292-295.

Rowe, J.A., Rogerson, S.J., Raza, A., Moulds, J.M., Kazatchkine, M.D., Marsh, K., Newbold, C.I., Atkinson, J.P., and Miller, L.H. 2000. Mapping of the region of complement receptor (CR) 1 required for Plasmodium falciparum rosetting and demonstration of the importance of CR1 in rosetting in field isolates. *J Immunol* **165:** 6341-6346.

Sahu, A., Isaacs, S.N., Soulika, A.M., and Lambris, J.D. 1998. Interaction of vaccinia virus complement control protein with human complement proteins: factor I-mediated degradation of C3b to iC3b1 inactivates the alternative complement pathway. *J Immunol* **160:** 5596-5604.

Sali, A., and Blundell, T.L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234:** 779-815.

Sanchez, R., and Sali, A. 1997. Advances in comparative protein-structure modelling. *Curr Opin Struct Biol* **7:** 206-214.

Sanchez, R., and Sali, A. 1998. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. *Proc Natl Acad Sci U S A* **95:** 13597-13602.

Sanchez-Corral, P., Gonzalez-Rubio, C., Rodriguez de Cordoba, S., and Lopez-Trascasa, M. 2004. Functional analysis in serum from atypical Hemolytic Uremic Syndrome patients reveals impaired protection of host cells associated with mutations in factor H. *Mol Immunol* **41:** 81-84.

Sanchez-Corral, P., Perez-Caballero, D., Huarte, O., Simckes, A.M., Goicoechea, E., Lopez-Trascasa, M., and de Cordoba, S.R. 2002. Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome. *Am J Hum Genet* **71:** 1285-1295. Epub 2002 Nov 1286.

Sandoval, A., Ai, R., Ostresh, J.M., and Ogata, R.T. 2000. Distal recognition site for classical pathway convertase located in the C345C/netrin module of complement component C5. *J Immunol* **165:** 1066-1073.

Santoro, F., Kennedy, P.E., Locatelli, G., Malnati, M.S., Berger, E.A., and Lusso, P. 1999. CD46 is a cellular receptor for human herpesvirus 6. *Cell* **99:** 817-827.

Saqi, M.A., Russell, R.B., and Sternberg, M.J. 1998. Misleading local sequence alignments: implications for comparative protein modelling. *Protein Eng* **11:** 627-630.

Saunders, R.E., Goodship, T.H., Zipfel, P.F., and 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.

Sayle, R.A., and Milner-White, E.J. 1995. RASMOL: biomolecular graphics for all. *Trends Biochem Sci* **20:** 374.

Schlesinger, L.S., Bellinger-Kawahara, C.G., Payne, N.R., and Horwitz, M.A. 1990. Phagocytosis of Mycobacterium tuberculosis is mediated by human monocyte complement receptors and complement component C3. *J Immunol* **144:** 2771-2780.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95:** 5857-5864.

Schwarzenbacher, R., Zeth, K., Diederichs, K., Gries, A., Kostner, G.M., Laggner, P., and Prassl, R. 1999. Crystal structure of human beta2-glycoprotein I: implications for phospholipid binding and the antiphospholipid syndrome. *Embo J* **18:** 6228-6239.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31:** 3381-3385.

Schwendinger, M.G., Spruth, M., Schoch, J., Dierich, M.P., and Prodinger, W.M. 1997. A novel mechanism of alternative pathway complement activation accounts for the deposition of C3 fragments on CR2-expressing homologous cells. *J Immunol* **158:** 5455-5463.

Segerman, A., Atkinson, J.P., Marttila, M., Dennerquist, V., Wadell, G., and Arnberg, N. 2003. Adenovirus type 11 uses CD46 as a cellular receptor. *J Virol* **77:** 9183-9191.

Seitsonen, S., Lemmela, S., Holopainen, J., Tommila, P., Ranta, P., Kotamies, A., Moilanen, J., Palosaari, T., Kaarniranta, K., Meri, S., et al. 2006. Analysis of variants in the complement factor H, the elongation of very long chain fatty acids-like 4 and the hemicentin 1 genes of age-related macular degeneration in the Finnish population. *Mol Vis* **12:** 796-801.

Shafren, D.R., Dorahy, D.J., Ingham, R.A., Burns, G.F., and Barry, R.D. 1997. Coxsackievirus A21 binds to decay-accelerating factor but requires intercellular adhesion molecule 1 for cell entry. *J Virol* **71:** 4736-4743.

Shapiro, L., and Scherer, P.E. 1998. The crystal structure of a complement-1q family protein suggests an evolutionary link to tumor necrosis factor. *Curr Biol* **8:** 335-338.

Shatsky, M., Nussinov, R., and Wolfson, H.J. 2002. MultiProt - A multiple protein structural alignment algorithm. *Lect Notes Comput Sc* **2452:** 235-250.

Shatsky, M., Nussinov, R., and Wolfson, H.J. 2004. A method for simultaneous alignment of multiple protein structures. *Proteins-Structure Function and Bioinformatics* **56:** 143-156.

Shimizu, A., Asakawa, S., Sasaki, T., Yamazaki, S., Yamagata, H., Kudoh, J., Minoshima, S., Kondo, I., and Shimizu, N. 2003. A novel giant gene CSMD3 encoding a protein with CUB and sushi multiple domains: a candidate gene for benign adult familial myoclonic epilepsy on human chromosome 8q23.3-q24.1. *Biochem Biophys Res Commun* **309:** 143-154.

Shindyalov, I.N., and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* **11:** 739-747.

Shur, I., Socher, R., Hameiri, M., Fried, A., and Benayahu, D. 2006. Molecular and cellular characterization of SEL-OB/SVEP1 in osteogenic cells in vivo and in vitro. *J Cell Physiol* **206:** 420-427.

Sim, R.B., and DiScipio, R.G. 1982. Purification and structural studies on the complement-system control protein beta 1H (Factor H). *Biochem J* **205:** 285-293.

Simonelli, F., Frisso, G., Testa, F., di Fiore, R., Vitale, D.F., Manitto, M.P., Brancato, R., Rinaldi, E., and Sacchetti, L. 2006. Polymorphism p.402Y>H in the complement factor H protein is a risk factor for age related macular degeneration in an Italian population. *Br J Ophthalmol* **90:** 1142-1145.

Skerka, C., Moulds, J.M., Taillon-Miller, P., Hourcade, D., and Zipfel, P.F. 1995. The human factor H-related gene 2 (FHR2): structure and linkage to the coagulation factor XIIIb gene. *Immunogenetics* **42:** 268-274.

Smith, B.O., Mallin, R.L., Krych-Goldberg, M., Wang, X., Hauhart, R.E., Bromek, K., Uhrin, D., Atkinson, J.P., and Barlow, P.N. 2002. Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor. *Cell* **108:** 769-780.

Smith, R.A. 2002. Targeting anticomplement agents. *Biochem Soc Trans* **30:** 1037-1041.

Smith, S.A., Mullin, N.P., Parkinson, J., Shchelkunov, S.N., Totmenin, A.V., Loparev, V.N., Srisatjaluk, R., Reynolds, D.N., Keeling, K.L., Justus, D.E., et al. 2000. Conserved surface-exposed K/R-X-K/R motifs and net positive charge on poxvirus complement control proteins serve as putative heparin binding sites and contribute to inhibition of molecular interactions with human endothelial cells: a novel mechanism for evasion of host defense. *J Virol* **74:** 5659-5666.

Smith, S.A., Sreenivasan, R., Krishnasamy, G., Judge, K.W., Murthy, K.H., Arjunwadkar, S.J., Pugh, D.R., and Kotwal, G.J. 2003. Mapping of regions within the vaccinia virus complement control protein involved in dose-dependent binding to key complement components and heparin using surface plasmon resonance. *Biochim Biophys Acta* **1650:** 30-39.

Soares, D.C., and Barlow, P.N. 2005. Complement Control Protein Modules in the Regulators of Complement Activation. In *Structural Biology of the Complement System.* (eds. D. Morikis, and J.D. Lambris), pp. 19-62. CRC Press, Taylor & Francis Group, Boca Raton.

Soares, D.C., Gerloff, D.L., Syme, N.R., Coulson, A.F., Parkinson, J., and Barlow, P.N. 2005. Large-scale modelling as a route to multiple surface comparisons of the CCP module family. *Protein Eng Des Sel* **18:** 379-388. Epub 2005 Jun 2023.

Solomon, K.R., Sharma, P., Chan, M., Morrison, P.T., and Finberg, R.W. 2004. CD109 represents a novel branch of the alpha2-macroglobulin/complement gene family. *Gene* **327:** 171-183.

Souied, E.H., Leveziel, N., Richard, F., Dragon-Durey, M.A., Coscas, G., Soubrane, G., Benlian, P., and Fremeaux-Bacchi, V. 2005. Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population. *Mol Vis* **11:** 1135-1140.

Soulika, A.M., Morikis, D., Sarrias, M.R., Roy, M., Spruce, L.A., Sahu, A., and Lambris, J.D. 2003. Studies of structure-activity relations of complement inhibitor compstatin. *J Immunol* **171:** 1881-1890.

Spiller, O.B., Blackbourn, D.J., Mark, L., Proctor, D.G., and Blom, A.M. 2003. Functional activity of the complement regulator encoded by Kaposi's sarcoma-associated herpesvirus. *J Biol Chem* **278:** 9283-9289.

Srinivasan, N., and Blundell, T.L. 1993. An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng* **6:** 501-512.

Stauber, D.J., Debler, E.W., Horton, P.A., Smith, K.A., and Wilson, I.A. 2006. Crystal structure of the IL-2 signaling complex: Paradigm for a heterotrimeric cytokine receptor. *Proc Natl Acad Sci U S A.*

Steinkasserer, A., Barlow, P.N., Willis, A.C., Kertesz, Z., Campbell, I.D., Sim, R.B., and Norman, D.G. 1992. Activity, disulphide mapping and structural modelling of the fifth domain of human beta 2-glycoprotein I. *FEBS Lett* **313:** 193-197.

Stevenson, B., El-Hage, N., Hines, M.A., Miller, J.C., and Babb, K. 2002. Differential binding of host complement inhibitor factor H by Borrelia burgdorferi Erp surface proteins: a possible mechanism underlying the expansive host range of Lyme disease spirochetes. *Infect Immun* **70:** 491-497.

Stoiber, H., Clivio, A., and Dierich, M.P. 1997. Role of complement in HIV infection. *Annu Rev Immunol* **15:** 649-674.

Stoiber, H., Pinter, C., Siccardi, A.G., Clivio, A., and Dierich, M.P. 1996. Efficient destruction of human immunodeficiency virus in human serum by inhibiting the protective action of complement factor H and decay accelerating factor (DAF, CD55). *J Exp Med* **183:** 307-310.

Sunyer, J.O., Zarkadis, I.K., and Lambris, J.D. 1998. Complement diversity: a mechanism for generating immune diversity? *Immunol Today* **19:** 519-523.

Sutcliffe, M.J., Haneef, I., Carney, D., and Blundell, T.L. 1987a. Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng* **1:** 377-384.

Sutcliffe, M.J., Hayes, F.R., and Blundell, T.L. 1987b. Knowledge based modelling of homologous proteins, Part II: Rules for the conformations of substituted sidechains. *Protein Eng* **1:** 385-392.

Szakonyi, G., Guthridge, J.M., Li, D., Young, K., Holers, V.M., and Chen, X.S. 2001. Structure of complement receptor 2 in complex with its C3d ligand. *Science* **292:** 1725-1728.

Takahashi-Nishimaki, F., Funahashi, S., Miki, K., Hashizume, S., and Sugimoto, M. 1991. Regulation of plaque size and host range by a vaccinia virus gene related to complement system proteins. *Virology* **181:** 158-164.

Takizawa, H., Okada, N., and Okada, H. 1994. Complement inhibitor of rat cell membrane resembling mouse Crry/p65. *J Immunol* **152:** 3032-3038.

Taylor, C.M. 2001. Hemolytic-uremic syndrome and complement factor H deficiency: clinical aspects. *Semin Thromb Hemost* **27:** 185-190.

Thai, C.T., and Ogata, R.T. 2003. Expression and characterization of the C345C/NTR domains of complement components C3 and C5. *J Immunol* **171:** 6565-6573.

Thai, C.T., and Ogata, R.T. 2004. Complement components C5 and C7: recombinant factor I modules of C7 bind to the C345C domain of C5. *J Immunol* **173:** 4547-4552.

Thern, A., Stenberg, L., Dahlback, B., and Lindahl, G. 1995. Ig-binding surface proteins of Streptococcus pyogenes also bind human C4b-binding protein (C4BP), a regulatory component of the complement system. *J Immunol* **154:** 375-386.

Thomas, M.L., Janatova, J., Gray, W.R., and Tack, B.F. 1982. Third component of human complement: localization of the internal thiolester bond. *Proc Natl Acad Sci U S A* **79:** 1054-1058.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:** 4876-4882.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673-4680.

Tomlinson, S., Pontes de Carvalho, L.C., Vandekerckhove, F., and Nussenzweig, V. 1994. Role of sialic acid in the resistance of Trypanosoma cruzi trypomastigotes to complement. *J Immunol* **153:** 3141-3147.

Trouw, L.A., Nilsson, S.C., Goncalves, I., Landberg, G., and Blom, A.M. 2005. C4b-binding protein binds to necrotic cells and DNA, limiting DNA release and inhibiting complement activation. *J Exp Med* **201:** 1937-1948.

Uhrinova, S., Lin, F., Ball, G., Bromek, K., Uhrin, D., Medof, M.E., and Barlow, P.N. 2003. Solution structure of a functionally active fragment of decay-accelerating factor. *Proc Natl Acad Sci U S A* **100:** 4718-4723.

Ullmann, G.M., Hauswald, M., Jensen, A., Kostic, N.M., and Knapp, E.W. 1997. Comparison of the physiologically equivalent proteins cytochrome c6 and plastocyanin on the basis of their electrostatic potentials. Tryptophan 63 in cytochrome c6 may be isofunctional with tyrosine 83 in plastocyanin. *Biochemistry* **36:** 16187-16196.

Uvarova, E.A., and Shchelkunov, S.N. 2001. Species-specific differences in the structure of orthopoxvirus complement-binding protein. *Virus Res* **81:** 39-45.

van de Poel, R.H., Meijers, J.C., and Bouma, B.N. 1999. Interaction between protein S and complement C4b-binding protein (C4BP). Affinity studies using chimeras containing c4bp beta-chain short consensus repeats. *J Biol Chem* **274:** 15144-15150.

van den Elsen, J.M., Martin, A., Wong, V., Clemenza, L., Rose, D.R., and Isenman, D.E. 2002. X-ray crystal structure of the C4d fragment of human complement component C4. *J Mol Biol* **322:** 1103-1115.

Verma, A., Hellwage, J., Artiushin, S., Zipfel, P.F., Kraiczy, P., Timoney, J.F., and Stevenson, B. 2006. LfhA, a novel factor H-binding protein of Leptospira interrogans. *Infect Immun* **74:** 2659-2666.

Vik, D.P., and Wong, W.W. 1993. Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele. *J Immunol* **151:** 6214-6224.

Villoutreix, B.O., Hardig, Y., Wallqvist, A., Covell, D.G., Garcia de Frutos, P., and Dahlback, B. 1998. Structural investigation of C4b-binding protein by molecular modeling: localization of putative binding sites. *Proteins* **31:** 391-405.

Vinson, M., van der Merwe, P.A., Kelm, S., May, A., Jones, E.Y., and Crocker, P.R. 1996. Characterization of the sialic acid-binding site in sialoadhesin by site-directed mutagenesis. *J Biol Chem* **271:** 9267-9272.

Vriend, G. 1990. What If - a Molecular Modeling and Drug Design Program. *J Mol Graphics* **8:** 52-&.

Vriend, G., and Sander, C. 1993. Quality-Control of Protein Models - Directional Atomic Contact Analysis. *J Appl Crystallogr* **26:** 47-60.

Wade, R.C., Gabdoulline, R.R., and Luty, B.A. 1998. Species dependence of enzyme-substrate encounter rates for triose phosphate isomerases. *Proteins* **31:** 406-416.

Wallace, I.M., Blackshields, G., and Higgins, D.G. 2005. Multiple sequence alignments. *Curr Opin Struct Biol* **15:** 261-266.

Wallner, B., and Elofsson, A. 2005. All are not equal: a benchmark of different homology modeling programs. *Protein Sci* **14:** 1315-1327.

Walport, M.J. 2001a. Complement. First of two parts. *N Engl J Med* **344:** 1058-1066.

Walport, M.J. 2001b. Complement. Second of two parts. *N Engl J Med* **344:** 1140-1144.

Wang, X., Rickert, M., and Garcia, K.C. 2005. Structure of the quaternary complex of interleukin-2 with its alpha, beta, and gammac receptors. *Science* **310:** 1159-1163.

Warwicker, P., Goodship, T.H., Donne, R.L., Pirson, Y., Nicholls, A., Ward, R.M., Turnpenny, P., and Goodship, J.A. 1998. Genetic studies into inherited and sporadic hemolytic uremic syndrome. *Kidney Int* **53:** 836-844.

Webb, J.H., Villoutreix, B.O., Dahlback, B., and Blom, A.M. 2001. Localization of a hydrophobic binding site for anticoagulant protein S on the beta -chain of complement regulator C4b-binding protein. *J Biol Chem* **276:** 4330-4337.

Wei, L., Pandiripally, V., Gregory, E., Clymer, M., and Cue, D. 2005. Impact of the SpeB protease on binding of the complement regulatory proteins factor H and factor H-like protein 1 by Streptococcus pyogenes. *Infect Immun* **73:** 2040-2050.

Wei, X., Orchardson, M., Gracie, J.A., Leung, B.P., Gao, B., Guan, H., Niedbala, W., Paterson, G.K., McInnes, I.B., and Liew, F.Y. 2001. The Sushi domain of soluble IL-15 receptor alpha is essential for binding IL-15 and inhibiting inflammatory and allogenic responses in vitro and in vivo. *J Immunol* **167:** 277-282.

Weisman, H.F., Bartow, T., Leppo, M.K., Marsh, H.C., Jr., Carson, G.R., Concino, M.F., Boyle, M.P., Roux, K.H., Weisfeldt, M.L., and Fearon, D.T. 1990. Soluble human complement receptor type 1: in vivo inhibitor of complement suppressing post-ischemic myocardial inflammation and necrosis. *Science* **249:** 146-151.

Wetsel, R.A., Kolb, W.P. 1982. Complement-independent activation of the fifth component (C5) of human complement: limited trypsin digestion resulting in the expression of biological activity. *J Immunol* **128:** 2209-2216.

Weyer, K., Overgaard, M.T., Laursen, L.S., Nielsen, C.G., Schmitz, A., Christiansen, M., Sottrup-Jensen, L., Giudice, L.C., and Oxvig, C. 2004. Cell surface adhesion of pregnancy-associated plasma protein-A is mediated by four clusters of basic residues located in its third and fourth CCP module. *Eur J Biochem* **271:** 1525-1535.

Whiss, P.A. 2002. Pexelizumab Alexion. *Curr Opin Investig Drugs* **3:** 870-877.

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., et al. 2006. Structure of C3b in complex with CRIg gives insights into regulation of complement activation. *Nature* **444:** 217-220.

Wiles, A.P., Shaw, G., Bright, J., Perczel, A., Campbell, I.D., and Barlow, P.N. 1997. NMR studies of a viral protein that mimics the regulators of complement activation. *J Mol Biol* **272:** 253-265.

Williams, P., Chaudhry, Y., Goodfellow, I.G., Billington, J., Powell, R., Spiller, O.B., Evans, D.J., and Lea, S. 2003. Mapping CD55 function. The structure of two pathogen-binding domains at 1.7 A. *J Biol Chem* **278:** 10691-10696.

Williamson, M.P., and Madison, V.S. 1990. Three-dimensional structure of porcine C5adesArg from 1H nuclear magnetic resonance data. *Biochemistry* **29:** 2895-2905.

Winter, C., Henschel, A., Kim, W.K., and Schroeder, M. 2006. SCOPPI: a structural classification of protein-protein interfaces. *Nucleic Acids Res* **34:** D310-314.

Winters, M.S., Spellman, D.S., and Lambris, J.D. 2005. Solvent accessibility of native and hydrolyzed human complement protein 3 analyzed by hydrogen/deuterium exchange and mass spectrometry. *J Immunol* **174:** 3469-3474.

Woodruff, T.M., Pollitt, S., Proctor, L.M., Stocks, S.Z., Manthey, H.D., Williams, H.M., Mahadevan, I.B., Shiels, I.A., and Taylor, S.M. 2005. Increased potency of a novel complement factor 5a receptor antagonist in a rat model of inflammatory bowel disease. *J Pharmacol Exp Ther* **314:** 811-817. Epub 2005 May 2005.

Wu, E., Trauger, S.A., Pache, L., Mullen, T.M., von Seggern, D.J., Siuzdak, G., and Nemerow, G.R. 2004. Membrane cofactor protein is a receptor for adenoviruses associated with epidemic keratoconjunctivitis. *J Virol* **78:** 3897-3905.

Wyckoff, G.J., Wang, W., and Wu, C.I. 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403:** 304-309.

Xu, C., Mao, D., Holers, V.M., Palanca, B., Cheng, A.M., and Molina, H. 2000. A critical role for murine complement regulator crry in fetomaternal tolerance. *Science* **287:** 498-501.

Yu, C.Y., Belt, K.T., Giles, C.M., Campbell, R.D., and Porter, R.R. 1986. Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity. *Embo J* **5:** 2873-2881.

Zanotti, G., Bassetto, A., Battistutta, R., Folli, C., Arcidiaco, P., Stoppini, M., and Berni, R. 2000. Structure at 1.44 A resolution of an N-terminally truncated form of the rat serum complement C3d fragment. *Biochim Biophys Acta* **1478:** 232-238.

Zhang, X., Boyar, W., Toth, M.J., Wennogle, L., and Gonnella, N.C. 1997. Structural definition of the C5a C terminus by two-dimensional nuclear magnetic resonance spectroscopy. *Proteins* **28:** 261-267.

Zipfel, P.F., Jokiranta, T.S., Hellwage, J., Koistinen, V., and Meri, S. 1999. The factor H protein family. *Immunopharmacology* **42:** 53-60.

Zuiderweg, E.R., Nettesheim, D.G., Mollison, K.W., and Carter, G.W. 1989. Tertiary structure of human complement component C5a in solution from nuclear magnetic resonance data. *Biochemistry* **28:** 172-185.