STRUCTURAL INVESTIGATIONS OF THE COMPONENTS

OF THE COMPLEMENT CASCADE

THESIS PRESENTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

KATHRYN FRANCES SMITH

DEPARTMENT OF BIOCHEMISTRY AND CHEMISTRY
ROYAL FREE HOSPITAL SCHOOL OF MEDICINE
UNIVERSITY OF LONDON

1992

ABSTRACT

The complement cascade is a multi-component protein system found in plasma and plays an important role in immune defence. The molecular structures of five proteins were studied using a combination of neutron and X-ray small-angle solution scattering and secondary structure prediction techniques.

α_1-antitrypsin is the best characterized member of the serpin superfamily. It was shown to have a moderately elongated structure. Models of native and reactive-centre cleaved α_1-antitrypsin were created based on the crystal co-ordinates of the cleaved protein. Since these were very similar, the gross conformational change between the two structures is minor. $C\bar{1}$ inhibitor controls the activation of the classical pathway. It consists of a serpin domain and a highly glycosylated N-terminal domain. The structure of the serpin domain was based upon the crystal structure of α_1-antitrypsin. The N-terminal domain was modelled as a highly elongated structure with extended carbohydrates.

The activation of C5 initiates the formation of the membrane attack complex. Solution scattering showed that this had an elongated structure and was similar to the homologous proteins C3 and C4.

Properdin is a regulatory protein of the alternative pathway. The properdin trimer was successfully modelled as a triangle of sides 26 nm in agreement with published electron micrographs. Monomeric properdin contains 6 thrombospondin repeats (TSR). Sequence alignments and secondary structure predictions of the TSR showed this to be comprised of two β-turn regions and one β-strand region.

C9 is the most abundant member of the membrane attack complex. It consists of four domains: thrombospondin repeat (TSR), low density lipoprotein receptor (LDLr), perforin- like region (PLR) and an epidermal growth factor-like (EGF) domain. Again, solution scattering experiments showed C9 was elongated. Models for C9 suggest that the domains are arranged in a V shape with each arm of length 11.1 nm and inter-arm angle of 10^{o}. A total of 74 LDLr, 10 PLR and 99 EGF sequences were aligned and their secondary structures predicted. The LDLr and EGF domains were found to consist of β-strand and β-turn conformations whilst the PLR was predominantly α-helix.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

C4BP	C4b binding protein
CD	circular dichroism
CR1, CR2, CR3, CR4	complement receptor types 1, 2, 3, 4
DAF	decay accelerating factor
EGF	epidermal growth factor
FIM	factor I module
FT-IR	Fourier transform infrared
HRF	homologous restriction factor
ICCC	Imperial College Computer Centre
ILL	Institut Laue-Langevin
ITP	indirect transform procedure
LDL	low density lipoprotein
LDLr	low density lipoprotein receptor
MAC	membrane attack complex
MCP	membrane cofactor protein
M_r	molecular weight
NMR	nuclear magnetic resonance
P_2, P_3, P_4	properdin dimers, trimers, tetramers
PLR	perforin-like region
RCA	regulation of complement activation
RFHSM	Royal Free Hospital School of Medicine
R_G	radius of gyration
R_{XS}	radius of gyration of the cross-section
SCR	short consensus repeat
SERPIN	serine protease inhibitor
SP	serine protease
SRS	synchrotron radiation source
TSR	thrombospondin repeat
vWF	von Willebrand Factor

ACKNOWLEDGEMENTS

I wish to express my appreciation to the following:

Dr S.J. Perkins, my supervisor, for his guidance and continual help and support.

Mr A.S. Nealis for his help and advice regarding computer hardware and software and for his assistance with the photography.

Dr W. Bras at the Synchrotron Research Source (Daresbury, U.K) and Dr A. de Geyer at the Institut Laue-Langevin (Grenoble, France) for their assistance with the use of the small-angle scattering instruments.

Dr R.A. Harrison and Prof. P.J. Lachmann at the MRC Molecular Immunopathology Unit, Cambridge and Dr K.F. Nolan and Dr K.B.M. Reid at the MRC Immunochemistry Unit, Oxford for the protein preparations.

Other members of the Department of Biochemistry who have given help and advice.

My parents for their continual support and encouragement.

The Wellcome Trust for support.

The Science and Educational Research Council for the use of the Synchrotron Research Source, Daresbury and the Institut Laue-Langevin, Grenoble.

CHAPTER 1.

THE COMPLEMENT CASCADE

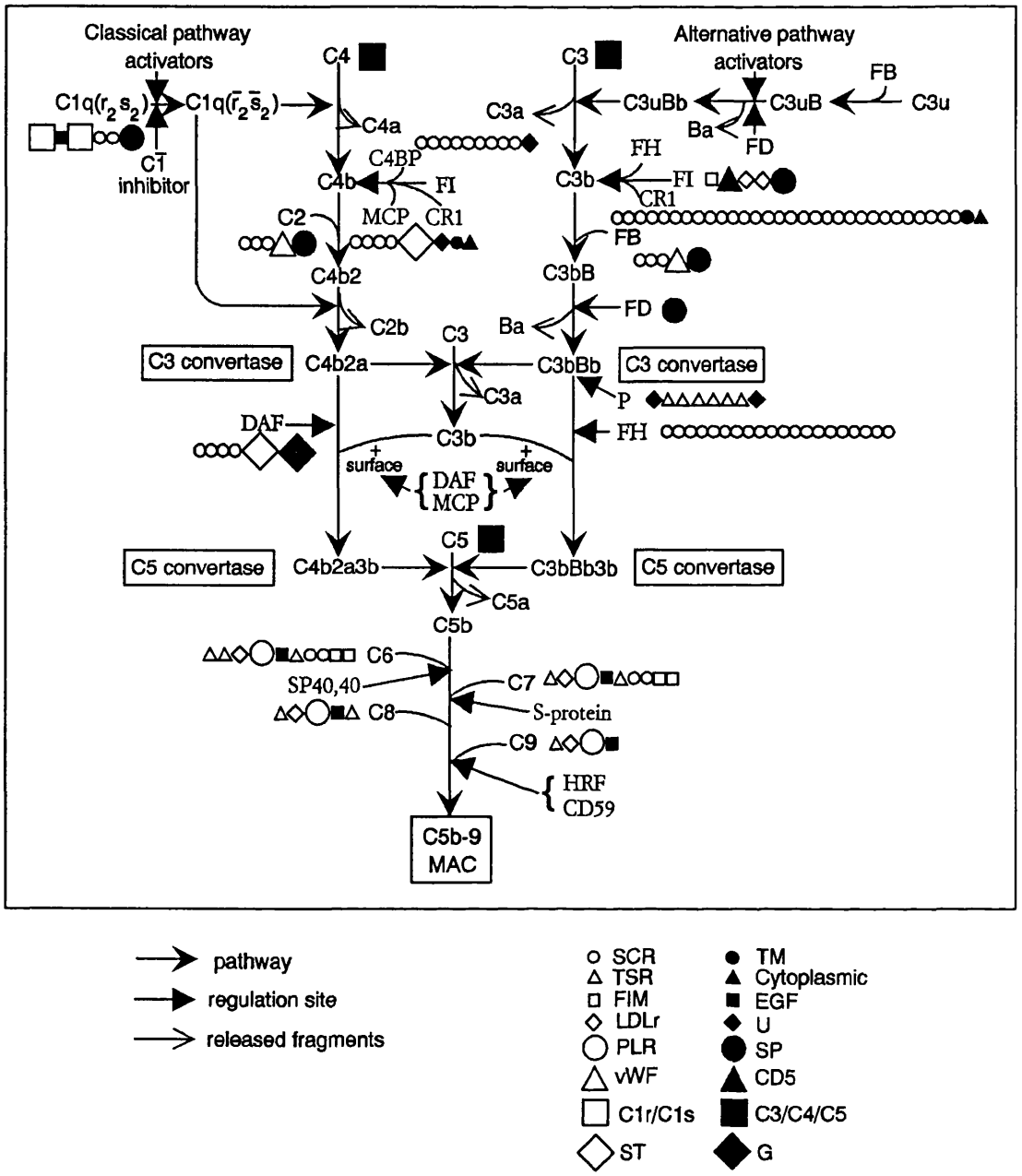


Figure 1.1. Diagrammatic representation of the complement cascade.

The classical pathway is shown on the left, the alternative on the right. The major components, control proteins and proteins involved in host protection mechanisms are shown. Known domain structures are indicated. (See Table 1.1 for abbreviations).

(1.1) INTRODUCTION

The complement cascade is a multi-component protein system found in plasma. It plays an essential role in the immune defence mechanism and in the mediation of inflammation. Complement has three major functions: (a) opsonization and removal of complement-coated antigens via complement receptors on phagocytic cells; (b) recruitment of phagocytic cells into the site of complement activation; (c) destruction of membrane integrity of target organisms. Over 30 proteins are involved (Figure 1.1; Table 1.1), which function either as enzymes, binding proteins, regulators or membrane-bound receptors. The complement cascade proteins are a major constituent in plasma with a total concentration of around 4 g/l. C3 and C4 are the most abundant complement proteins at 1.3 and 0.6 g/l, respectively, consistent with their central role in the cascade whereas factor D is the least abundant (1 mg/l) (Table 1.1).

The cascade is activated by one of two distinct routes, the classical or the alternative pathway (Figure 1.1). Both lead to the formation of complex proteases which activate the central components C3 and C4. Fragments of C3 or C4 covalently attach to neighbouring targets, marking them for clearance by cells of the immune system carrying complement receptors. Activation of C5 initiates the terminal or lytic pathway, ending with the formation of the cytolytic membrane attack complex. Release of the anaphylatoxins, C3a, C4a and C5a during activation induce the host's inflammatory response. The cascade is tightly regulated by a number of proteins which will either enhance or inhibit the progress of the cascade.

	M_r (kDa)	Approx. serum conc. (mg l^{-1})	Domains
Classical			
C1q	457	80	stalks, head
C1r	172	50	2 RS, 2 SCR, EGF, SP
C1s	158	50	2 RS, 2 SCR, EGF, SP
C4	197	600	C3/C4/C5
C2	102	20	vWF, SP, 3 SCR
C3	187	1300	C3/C4/C5
Alternative			
Factor D	24	1	SP
Factor B	89	210	3 SCR, vWF, SP
Terminal			
C5	194	70	C3/C4/C5
C6	107	64	3 TSR, LDLr, PLR, EGF, 2 SCR, 2 FIM
C7	95	56	2 TSR, LDLr, PLR, EGF, 2 SCR, 2 FIM
C8	154	55	2 TSR, LDLr, PLR, EGF
C9	66	59	TSR, LDLr, PLR, EGF
Control proteins			
Plasma			
C1 inhibitor	71	200	N-terminus, serpin
Factor J	20	5	
Factor I	74	35	FIM, CD5, 2 LDLr, SP
Properdin (trimer)	162	20	3 x 6 TSR
C4BP	491	250	7 x 8 SCR + 3 SCR
Factor H	150	480	20 SCR
S-Protein	83	505	
SP-40,40	70	100	coiled-coil
Carboxy-peptidase N	310	35	
Membrane bound			
MCP	45-70		4 SCR, ST, U, TM, CYT
DAF	70		4 SCR, ST, G
HRF	65		
CD59	18-20		murine LY-6 Antigen
Receptors			
CR1	160, 190 220, 250		30 SCR, TM, CYT
CR2	140		16 SCR, TM, CYT
CR3	265		vWF, 3 MB, TM, CYT
CR4	245		vWF, 3 MB, TM, CYT
C5a receptor	39		7 TM
C1q receptor	56		

Abbreviations

CYT	cytoplasmic domain	SCR	short consensus repeat
EGF	epidermal growth factor	SP	serine protease domain
G	glycolipid anchor	ST	serine/threonine-enriched area
FIM	factor I module	TM	transmembrane domain
LDLr	LDL receptor	TSR	thrombospondin repeat
MB	metal binding domain	U	unknown functional significance
PLR	perforin like region	vWF	von Willebrand Factor
RS	C1r/C1s domain		

Table 1.1. Summary of the physiological concentrations and domain structures of the complement components.

(1.2) CLASSICAL PATHWAY

(1.2.1) The structure of C1q, C1r and C1s

The C1 enzyme complex is composed of the subunits C1q, C1r and C1s. C1q is made from 18 polypeptide chains of 3 types, 6A, 6B and 6C. Three of these chains (1A, 1B and 1C) form a collagen-like triple helix at the N-terminal end of the molecule (Reid & Porter, 1976). The 6 triple helices are aligned in parallel for half the collagen-like length and then diverge to terminate at the C-terminus in 6 globular heads (Knobel *et al.*, 1975; Brodsky-Doyle *et al.*, 1976). The subcomponents C1r and C1s have homologous amino acid sequences and similar tertiary structures (Tosi *et al.*, 1987; Kusumoto *et al.*, 1988). C1r and C1s interact to form a tetrameric chain, C1s-C1r-C1r-C1s in which the 2 catalytic domains (Table 1.1) of C1r are orientated in a head to tail manner (Arlaud *et al.*, 1986). C1s interacts with C1r via the C1r/C1s specific domain. The activated $C\bar{\bar{1}}r_2C\bar{\bar{1}}s_2$ complex has been found to have an asymmetric X structure (Weiss *et al.*, 1986; Perkins & Nealis, 1989). A variety of models have been proposed for the interaction of the $C\bar{1}r_2C\bar{1}s_2$ tetramer with C1q to form the C1 complex (Perkins, 1989a; Schumaker *et al.*, 1987). These can be divided into two groups, those which propose that the $C\bar{1}r_2C\bar{1}s_2$ tetramer is positioned on the outside of C1q, summarized in the "W-model" (Perkins, 1989a; Perkins & Nealis, 1989) and the "O_2-model" (Cooper, 1985) and those proposing the $C\bar{1}r_2C\bar{1}s_2$ subunits are interwoven between the arms of C1q, summarized in the "S-model" (Poon *et al.*, 1983; Schumaker *et al.*, 1986) and the "8-model" (Colomb *et al.*, 1989; Arlaud *et al.*, 1986). More structural studies on the individual components and their complexes need to be performed to rectify this discrepancy.

(1.2.2) Biological activities

Activation of the classical pathway is initiated by the binding of C1q to a variety of substances (Sim & Reid, 1991). The most familiar is the formation of immune complexes of either IgG or IgM with antigens. C1q binds via the globular heads to multiple Fc regions of IgG or IgM. The interaction between C1q and the

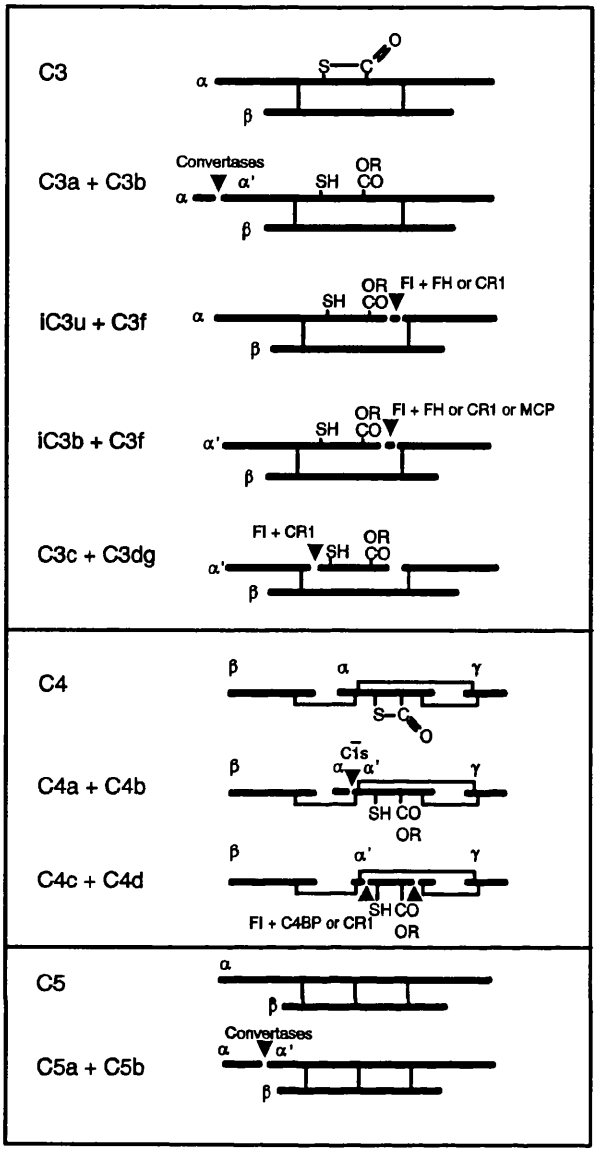


Figure 1.2. Proteolytic fragments of C3, C4 and C5.

The α chain of C3 is cleaved by the C3 convertase enzyme complexes to produce C3a and C3b. C3u is an inactive form of C3 where the internal thiol ester bond has been hydrolysed. C3b or C3u is cleaved by factor I in the presence of cofactors (factor H, CR1 or MCP) to produce iC3b or iC3u with the release of a small fragment, C3f (M_r 2,000). iC3b is further cleaved by factor I in the presence of cofactor CR1 to form C3c and C3dg. C4 is cleaved by $\overline{C1}s$ to give fragments C4a and C4b. C4b is cleaved by factor I with cofactors of C4BP or CR1 to produce C4c and C4d. C5 is cleaved by the C5 convertases to form C5a and C5b. The interchain disulphide bonds are shown as thin lines.

immunoglobulin is facilitated by the close proximity of many IgG molecules (Borsos, 1989) or a conformational change in the Fab arms of IgM (Perkins *et al.*, 1991b) to expose the C1q binding sites. This induces a conformational change in the collagenous stalks of C1q (Heinz, 1989; Hanson *et al.*, 1985; Schumaker *et al.*, 1986; Hoekzema *et al.*, 1988) which increases the affinity of C1q for $\overline{C1r}_2\overline{C1s}_2$. $\overline{C1r}_2\overline{C1s}_2$ will only interact with C1q, in the presence of calcium, once C1q has bound to an efficient activator. The catalytic domain of one C1r autoactivates and cleaves the corresponding domain in the neighbouring C1r molecule (Dodds *et al.*, 1978). This in turn activates C1s by a proteolytic cleavage (Ziccardi, 1976). The activated $\overline{C1s}$ is then able to split the classical pathway components C2 and C4 (Thielens *et al.*, 1984).

C4 has sequence homology to the complement components C3 and C5 (Chapter 5) and the non-complement proteins, α_2-macroglobulin and pregnancy zone protein. The proenzyme form of C4 is composed of 3 chains, α, β and γ. C1s cleaves a single peptide bond in the α-chain of C4 to produce a 9 kDa fragment C4a and a metastable fragment C4b (Figure 1.2). The C-terminal portion of the α-chain (α'-chain) of C4b remains disulphide linked to the β-chain. C4b is able to bind covalently to a variety of surfaces via an internal thiolester group in the α' chain, forming an ester or amide bond with respective groups on the neighbouring surface (Harrison *et al.*, 1981). Proenzyme C2 is split by $\overline{C1s}$ in the $\overline{C1}$ complex to produce fragments C2a and C2b. C2a associates with C4b in a Mg^{2+} dependent interaction to produce the C3 convertase C4b,2a (Horiuchi *et al.*, 1991; Müller-Eberhard *et al.*, 1967).

(1.2.3) Control proteins

Regulation of the classical pathway is controlled at two stages, formation of the $\overline{C1}$ enzyme complex and the C3 convertase. $\overline{C1}$ inhibitor controls $\overline{C1r}_2$ activation through non-covalent interactions (Arlaud *et al.*, 1987; Ziccardi & Cooper, 1979). $\overline{C1}$ inhibitor can also covalently interact with the 4 serine protease active sites of the $\overline{C1}$ complex (Table 1.1) resulting in the dissociation of the complex to form 2 molecules of $\overline{C1}$ inhibitor - $\overline{C1r}$ - $\overline{C1s}$ - $\overline{C1}$ inhibitor (Ziccardi

& Cooper, 1979). The free collagen-like stalks of C1q are then free to bind to cell-surface C1q receptors which may then initiate phagocytosis (Malhotra *et al.*, 1990; Malhotra & Sim, 1989; Ghebrehiwet, 1989). Factor J also controls the formation of the $\overline{C1}$ complex by inhibiting the association of $\overline{C1}r_2\overline{C1}s_2$ with C1q (Lopez-Trascasa *et al.*, 1989).

Control of the C3 convertase is governed by its inherently unstable nature and cleavage by factor I. C2a spontaneously dissociates from the complex. This is enhanced by C4 binding protein (C4BP) (Gigli *et al.*, 1979) and decay accelerating factor (DAF) (Nicholson-Weller *et al.*, 1982). Factor I cleaves C4b in two positions to give fragments C4c and C4d (Figure 1.2). C4BP (Fujita *et al.*, 1978; Gigli *et al.*, 1979), membrane cofactor protein (MCP) (Seya *et al.*, 1986) or complement receptor 1 (CR1) act as cofactors for this factor I mediated cleavage. In most physiological circumstances, the concentration of C4BP exceeds that of C4b generated by activation and therefore further activation of the complement components is prevented.

(1.3) ALTERNATIVE PATHWAY

(1.3.1) Biological activities

The alternative pathway is in a permanently activated state due to the spontaneous reaction of C3 with water. Native C3, like C4, contains an internal thiolester in the α-chain which hydrolyses at a slow rate to give the initial C3 convertase, C3u (also known as C3(H_2O)) which is capable of attaching to membranes (Pangburn & Müller-Eberhard, 1980; Pangburn *et al.*, 1981). Although C3u is structurally different from C3b in that it has an intact α-chain (Figure 1.2), they are functionally similar. C3u interacts with factor B in a Mg^{2+} dependent manner to produce C3uB (Figure 1.1). Factor B is then cleaved by factor D, to produce fragments Ba and Bb which are homologous to the classical pathway fragments C2b and C2a respectively. Fragment Ba dissociates from the complex, leaving the initial fluid-phase C3 convertase, C3uBb. Cleavage of C3 by the convertase releases the fragment C3a and exposes the highly reactive thiolester in C3b, enabling it to bind covalently to surfaces of neighbouring particles. C3b then will bind fragment Bb to produce a covalently linked C3 convertase (C3b,Bb) on the surface of target cells. The ability to bind factor B is thought to precede the susceptibility of C3b to proteolytic cleavage by factor I and factor H (Isenman *et al.*, 1981). The function of this convertase is to increase the number of C3b-bound cells in the immediate environment of activation.

Further activation of the alternative pathway is antibody independent. Activators include polysaccharides, fungi, bacteria, viruses, parasites and certain mammalian cells. It is not known what structure all the activators have in common but it is possible that C3b deposited on activators is protected from proteolytic degradation, enabling the formation of the C3b,Bb complex (Fearon, 1978; Pangburn & Müller-Eberhard, 1978; Pangburn *et al.*, 1980; Fearon & Austen, 1977). C4b bound to a surface can also activate the alternative pathway by binding to C3bBbP, which may be important in cases of C2 deficiency (Farries *et al.*, 1990; Matsusita & Okada, 1986).

(1.3.2) Control proteins

Control of the alternative pathway is comparable to the classical pathway in that homologous proteins fulfil similar roles. Both positive and negative control mechanisms exist. The C3b,Bb convertase is stabilized by the binding of the control protein properdin (P) to form a C3b,Bb,P complex (Fearon & Austen, 1975; Medicus *et al.*, 1976; Farries *et al.*, 1987), which facilitates the binding of factor B and prevents cleavage of the convertase by factor I (Farries *et al.*, 1988a). Negative control of activation is governed by factor H which acts in a similar way to C4BP in the classical pathway by acting as a cofactor for factor I mediated cleavage of C3b and C3u to form the fragments iC3b + C3f and iC3u + C3f (Figure 1.2) (Pangburn & Müller-Eberhard, 1978; Pangburn & Müller-Eberhard, 1983).

(1.3.3) Protection mechanisms

The majority of C3b and C4b (70 to 90%) formed as a result of complement activation do not become covalently linked to surfaces but remain in the fluid-phase (Hourcade *et al.*, 1989). Of the percentage that does bind, some will do so to host cells. These are protected against complement mediated cytolysis by the membrane proteins DAF and MCP (Atkinson & Farries, 1987). DAF prevents assembly of C3 convertase on membranes and will dissociate enzymes already bound. MCP has cofactor activity for factor I mediated cleavage (Liszewski *et al.*, 1991; Fearon, 1979). The other C3b receptors, CR1, CR2, CR3 and CR4 have a number of regulatory and receptor functions (reviews Sim *et al.*, 1987; Ross, 1989). CR1 will inhibit complement activation by either accelerating the decay or dissociation of the C3 convertases or acting as a cofactor for factor I mediated cleavage of C3b. It is involved in processing of immune complexes and in the mediation of binding and phagocytosis of C3b-coated particles by phagocytic cells. CR2 bound to C3b-bearing particles enhances B-cell proliferation in the presence of T-cell derived factors (Lambris, 1988). CR3 adheres to and promotes phagocytic uptake of iC3b-coated particles into circulating phagocytic cells. CR4 mediates binding of iC3b and C3dg (Gaither *et al.*, 1987).

(1.4) TERMINAL PATHWAY

(1.4.1) C5 convertase

The binding of one or more C3b molecules to either the classical or the alternative C3 convertase (C4b,2a or C3b,Bb) changes the specificity of the complex to a C5 convertase, that is, C4b,C2a,C3b or C3b,Bb,C3b respectively (Medicus *et al.*, 1976; Daha *et al.*, 1976). The alternative pathway C5 convertase is regulated by MCP and factor I by the conversion of one C3b dimer to C3bu (Seya *et al.*, 1991) whereas the classical C5 convertase is controlled by factor H (Ito & Tamura, 1983). C5 binds to the C3b fragment in the C5 convertase complex. This convertase cleaves a single peptide bond in the α-chain of C5 to yield the fragments C5a and metastable C5b (Nilsson *et al.*, 1975) (Figures 1.1 & 1.2). Although C5 is very similar in structure to C3, it does not have an internal thiolester bond and therefore does not bind covalently to surfaces after activation (Wetsel *et al.*, 1988; Law *et al.*, 1980).

(1.4.2) Biological activity

The C5b fragment contains a labile binding site with specificity for component C6 (Figure 1.1). The complex C5b,C6 remains loosely bound to C3b on target cell surface until interaction with C7. The C5b,6,7 complex undergoes a conformational change, dissociates from C3b and a metastable membrane binding site is formed on the C7 subunit (Preissner *et al.*, 1985). C5b,6,7 binds to a membrane to become an integral membrane protein. C8 is composed of 3 chains (α, β and γ) and binds to C5b,6,7 via the β-chain (Monahan & Sodetz, 1981; Stewart *et al.*, 1987). C8 is then able to undergo a conformational change which allows penetration into the membrane of the disulphide linked α- and γ-chains. C9 binds to the C5b,6,7,8 complex via the α-subunit on C8 (Stewart *et al.*, 1987). This catalyses the polymerisation of up to 18 molecules of C9. C5b,6,7,8 forms small functional channels of approximately 3 nm in diameter (Zalman & Müller-Eberhard, 1985) which can enlarge to 10 nm upon incorporation of C9 molecules (Ramm *et al.*, 1985). The composition of the membrane attack complex (MAC)

depends on the availability of monomeric C9. If sufficient C9 is present, binding will continue to form a typical cylindrical membrane lesion as seen in electron micrographs (see Chapter 7 for details).

The assembly of a functional MAC on a cell membrane leads to cytolysis. Two major hypothesis exist to explain cell death, the "doughnut" hypothesis where the MAC components form a protein lined channel or the "leaky patch" hypothesis where there is a destabilization of the lipid bilayer due to the insertion of the MAC (reviews Bhakdi & Tranum-Jensen, 1991 and Esser, 1991). The MAC can destroy the membrane of enveloped viruses and can kill many types of bacteria. The formation of a MAC on the surface of nucleated cells will stimulate a rapid increase in the concentration of intracellular Ca^{2+} which in turn stimulates the recovery processes of the cell. This allows a cell to survive a mild complement attack and also initiates in inflammatory response. If a much stronger attack occurs, the recovery processes are overwhelmed and the cell dies (Morgan *et al.*, 1986).

(1.4.3) Control proteins

Control of the terminal components occurs at two points (Figure 1.1). The complement associated protein, SP40,40 binds to C5,6 to prevent the formation of the complete MAC (Choi *et al.*, 1989). The S-protein and serum low density lipoprotein (LDL) will compete with membrane lipid for the binding site on C5b,C7 and hence can prevent attachment to the surface of bystander cells (Podack *et al.*, 1978).

(1.4.4) Protection mechanisms

Complement mediated cell lysis is dependent on 'homologous restriction' where complement from one species is inefficient at lysing cells of the same species. This ensures that the self-inflicted complement damage is kept to a minimum (Ojcius *et al.*, 1990). This is due to the interaction of specific membrane proteins with the terminal complement components (Lachmann, 1990). Two

proteins are involved:

1) homologous restriction factor (HRF) or C8 binding protein (C8BP) (Zalman *et al.*, 1986; Schönermark *et al.*, 1986). HRF/C8BP prevents the binding of C9 to C8$\alpha\gamma$ and its insertion into membranes (Hänsch *et al.*, 1987; Schönermark *et al.*, 1988).

2) P-18 (Sugita *et al.*, 1988), HRF20 (Okada *et al.*, 1989), membrane inhibitor of reactive lysis (MIRL) (Holguin *et al.*, 1989) or CD59 (Davies *et al.*, 1989). Bound CD59 to the C5b-8 complex is thought to prevent the binding of the first molecule of C9 which is required to initiate polymerization and MAC formation (Lachmann, 1991; Rollins & Sims, 1990).

(1.5) THE PHYSIOLOGICAL EFFECTS OF COMPLEMENT

(1.5.1) The inflammatory response

The cleaved peptides, C3a, C4a and C5a (anaphylatoxins) are functionally and structurally homologous (Greer, 1986). These can bind to specific receptors, C3a and C5a receptors (Lambris, 1988) to elicit a variety of cellular responses (Damerau, 1987; Rollins & Springer, 1985; Huey & Hugli, 1985; Johnson & Chenweth, 1985). Signal transduction of the C5a receptor is mediated by a GTP-binding protein (G protein) (Siciliano *et al.*, 1990). The C5a receptor has recently been sequenced (Boulay *et al.*, 1991; Gerard & Gerard, 1991) and has significant homology to other G protein coupled receptors (Boulay *et al.*, 1991). C5a is the most potent mediator of the inflammation response. This leads to chemotactic migration of neutrophilic granulocytes and monocytes, cell adhesion, enzyme release and the formation of arachidonic metabolites and active oxygen species. C3a and C5a can induce histamine release from mast cells and prostaglandins from macrophages which have vasodilator activity and increase the efflux of macromolecules from the plasma. The C-terminal arginine of the anaphylatoxins has been found to be essential for activity. Control of the inflammatory response is mediated by the anaphylatoxin inactivator (carboxypeptidase N) which selectively removes this arginine (Bokisch & Müller-Eberhard, 1970; Plummer & Hurwitz, 1978).

(1.5.2) Defence against infection

Complement can defend against infection either in combination with antibodies or independently via the alternative pathway. C3b produced after alternative pathway activation binds to the surface of foreign cells. Since many host phagocytic cells have receptors for C3b (Fearon & Wong, 1983), these are brought into close contact with the foreign organism. This opsonizes the bacterium and encourages endocytosis by polymorphonuclear leukocytes and monocytes. The Fc, CR1 and CR3 receptors on the phagocytes recognize the Fc portion on the antibody, C3b and iC3b respectively (Micklem & Sim, 1985;

Gresham & Volanakis, 1986). If the amount of antibody attached via the Fc receptor is high enough, phagocytosis of the immune complex will commence. If there is insufficient antibody present, activation of the complement cascade may occur. This will increase efficiency of the binding of phagocytes due to the deposition of C3b on foreign cell surfaces. It is generally considered that opsonization is more important than lysis in defence against infections.

(1.5.3) Involvement in other systems

The complement cascade does not function in isolation but affects both the kallikrein/kinin and the coagulation/fibrinolysis systems (Kalter *et al.*, 1985; Ziccardi, 1983). Activated factor XII (Hageman factor) splits prekallikrein to kallikrein and plasminogen to plasmin and all these enzymes are able to cleave $\overline{C1}$ and initiate the complement cascade. Kallikrein and plasmin both split C3 and plasmin will also split factor B. C1q can bind to fibrinogen and fibrin to localize C1q at the site of a wound to provide efficient opsonization (Entwhistle & Furcht, 1988). Control over the 3 systems is exerted by $\overline{C1}$ inhibitor which is active against $\overline{C1}$r and $\overline{C1}$s, factor XII, plasmin and kinin-generating kallikrein and also by carboxypeptidase N which inactivates C3a, C5a and kinin by splitting off the terminal arginine. Although protein S is a vitamin K dependent cofactor in the coagulation system (Comp *et al.*, 1984), it also can form a complex with the complement protein C4BP (review Hessing, 1991), although it does not affect the function of C4BP as a regulator of the C3 convertase (Dahlbäck *et al.*, 1983).

(1.5.4) The role of complement in disease

Although some complement component deficiencies have no obvious deleterious effects, many produce considerable morbidity and mortality. The analysis of disorders involving deficiencies in the complement cascade can highlight the functions of the components concerned. Deficiencies in the components of the classical pathway, C1, C2 and C4, lead to immune complex disorders such as systemic lupus erythematosus (SLE) (reviews Davies, 1983;

Morgan & Walport, 1991), leading to the conclusion that perhaps a major role of complement *in vivo* is in the prevention of immune complex disease (Lachmann, 1990). $C\bar{1}$ inhibitor deficiency leads to the disease hereditary angioedema (HANE) which is characterised by the swelling of the mucous membranes. This is due to the increased activity of C4, C2 and kallikrein to produce the anaphylatoxin C4a, C2-kinin fragment and bradykinin. These fragments increase vascular permeability and hence oedema (Mollnes & Lachmann, 1988).

There have been no cases reported of homozygous factor B deficiencies. Factor D deficiencies are associated with recurrent bacterial infections (Kluin-Nelemans *et al.*, 1984). Properdin deficiencies are inherited as an X-linked trait and lead to *Neisseria* infections in affected males (Nusinow *et al.*, 1985). C3 deficiencies are characterized by recurrent pyogenic infections thus confirming the essential role C3 plays in opsonizing bacteria for removal by phagocytic cells (Alper *et al.*, 1972; Berger *et al.*, 1983). Deficiency in factor I causes an unrestricted amplification of the alternative pathway which depletes factor B and C3 (Nusinow *et al.*, 1985). Factor H deficiencies have the same characteristics as for factor I (Thompson & Winterborn, 1981) due to the fact that factor H is an obligate cofactor for factor I.

Deficiencies in the terminal pathway are linked with a susceptibility to recurrent infections by the bacteria *Neisseria.* This is rare in Europe and the U.S.A. but relatively common in developing countries (Orren *et al.*, 1987; Schlesinger *et al.*, 1990). This may be because of the contribution of MAC to produce endotoxin shock after infantile gastroenteritis (Orren *et al.*, 1987). If there is a deficiency in the terminal complement pathway, the child may survive (Lachmann, 1990).

Deficiencies of the complement receptors, CR1 (Walport & Lachmann, 1988), CR2, CR3 and CR4 (Kishimoto *et al.*, 1989) lead to a number of disease states, including immune complex diseases, leukocyte adhesion and deficiency syndrome, hypertrophied gums and tendency to have indolent staphylococcal skin infections. The lack of a inflammatory response is because the emigration of

polymorphs and monocytes into the inflammatory region does not occur (Lachmann, 1990).

Patients with paroxysmal nocturnal haemoglobinuria (PNH) lack the ability to synthesize glycosylphosphatidylinositol (GPI) anchors which link a variety of host protection proteins to the cell membrane. These include DAF, HRF and CD59. This makes the cells highly susceptible to complement mediated haemolysis (Lublin & Atkinson, 1989).

(1.6) DOMAIN STRUCTURES AND PROTEIN FAMILIES IN COMPLEMENT

(1.6.1) The definition of a domain

It seems that during evolution regions of homologous sequences have arisen in different proteins (Doolittle, 1985). It is assumed that all these will have similar tertiary structures, although the exact function of the domain may differ. These domains, modules or motifs often correspond to single exons and have the same phase at both the 5' and 3' intron/exon boundaries. It has been proposed that the movement of these domains from the original gene was due to exon insertion and duplication (Doolittle, 1985, 1989; Patthy, 1987, 1988). Although contemporary genes show evidence of gene duplication and significant amino acid divergence, domains often can be recognized by their invariant cysteine residues (Doolittle, 1989). Wetlaufer (1973) defined a domain as an independently folding region of the molecule from a contiguous chain of amino acids. Several authors have proposed that a domain is a stable intermediate in the folding process and functions as an individual entity within the whole protein (Richardson, 1981; Rossman & Argos 1981; Wetlaufer, 1981; Kim & Baldwin, 1982). The minimum size of a domain has been estimated at around 40 residues which reflects the smallest stable structure whereas the largest domain is around 400 residues and represents the largest domain that can fold up efficiently as a single unit (Wetlaufer, 1981; Richardson, 1981). Since domains usually have a stable structure isolated from the protein, the tertiary structure can be determined by 2D-NMR or crystallography and the structure of homologous domains in other proteins can be predicted (Sali *et al.*, 1990).

Brief descriptions of the main domains found in the complement components are given below (Figure 1.1; Table 1.1):

(1.6.2) Short Consensus Repeat (SCR)

The SCR is the most abundant domain found in the complement proteins. The SCR occurs in a total of 12 complement proteins (Figure 1.1) and is also

found in a large number of non-complement proteins (Reid & Day, 1989; Hessing, 1991). The SCR is found in large numbers in the complement proteins involved in the regulation of complement activation (RCA) gene cluster on human chromosome 1q and hence an alternative name was proposed, the Complement Control Protein (CCP) (Reid & Day, 1989). The RCA contains the genes for 6 complement proteins, CR1, CR2, MCP, DAF, C4BP and factor H. CR1 and CR2 contain 30 (Klickstein *et al.*, 1987) and 15-16 SCRs (Weis *et al.*, 1988) respectively. MCP (Lublin *et al.*, 1988) and DAF (Medof *et al.*, 1987; Caras *et al.*, 1987) both contain 4 SCRs each. C4BP is composed of 7 identical α chains and one β chain. The 7 α chains contain 8 SCRs each whereas the β chain has been found to contain 3 SCRs (Hillarp & Dahlbäck, 1990). The entire sequence of factor H is composed of a total of 20 SCRs (Ripoche *et al.*, 1988). All these proteins are involved in decay-acceleration of the C3 convertases, cofactor activity for factor I and have overlapping binding sites on C3b (Holers *et al.*, 1985; Lambris, 1988). The SCR is also found in several complement proteins not involved in regulation. The classical pathway components C1r (Arlaud & Gagnon, 1983), C1s (Carter *et al.*, 1983) and C2 (Bentley, 1986) contain 2, 2 and 3 SCRs respectively. Others include the alternative pathway component factor B (3 SCRs) (Mole *et al.*, 1984), and also the terminal pathway components C6 (DiScipio & Hugli, 1989) and C7 (DiScipio *et al.*, 1988) (2 SCRs each).

The SCR is characterized as a 60 amino acid domain containing 4 half-cysteines which are bonded in the fashion 1-3 and 2-4, and has several other highly conserved residues (Perkins *et al.*, 1988). The 2D NMR structure of 2 SCR domains in factor H have been determined (Barlow *et al.*, 1991a, b; Norman *et al.*, 1991). The overall tertiary structures of the two domains were very similar suggesting that perhaps all SCR domains not only share homologous primary structures but also tertiary structures. The secondary structure consists of β-sheets connected by β-turns (Perkins *et al.*, 1988; Barlow *et al.*, 1991a). It has been proposed that the function of the SCRs found in complement proteins are to bind C3b (or the homologous proteins C4b and C5b) (Lambris, 1988; Reid *et al.*, 1986). Since no complement protein contains only one SCR, it is likely that the C3b-binding site spans two SCR units (Klickstein *et al.*, 1988; Krych *et al.*, 1991).

(1.6.3) Thrombospondin Repeat (TSR)

The TSR occurs as 6 repeats in the complement control protein properdin, three times in the terminal component C6, twice in C7, C8α (Rao *et al.*, 1987), C8β (Haefliger *et al.*, 1987) and once in human C9 (DiScipio *et al.*, 1984). There is also limited homology between the N-terminal region of the TSR and SP-40,40 (Kirszbaum *et al.*, 1989). The domain was first characterized in thrombospondin which contains 3 repeats (Lawler & Hynes, 1986). The TSR can be characterised by a WSxWSxWSxC motif and has 6 Cys residues over a length of approximately 60 residues (Goundis & Reid, 1988; Nolan *et al.*, 1991) (Chapter 6). The TSR may act as a binding domain for protein - protein interactions (Holt *et al.*, 1990; Prater *et al.*, 1991).

(1.6.4) Epidermal Growth Factor (EGF)

The EGF can be found in a wide range of proteins in large numbers of contiguous [e.g. the LDL related protein contains 22 repeats (Herz *et al.*, 1988)] or single repeats (e.g. the EGF itself). In the complement proteins it is found as only one repeat in C1r, C1s, C6, C7, C8α, C8β and C9. The domain is approximately 40 residues in length with 6 conserved cysteine residues forming 3 disulphide bridges, bonded in the manner 1 to 3, 2 to 4 and 5 to 6. There is otherwise comparatively little sequence conservation (Chapter 8). The 2D NMR structure of 5 EGFs have been determined (Cooke *et al.*, 1987, 1990; Montelione *et al.*, 1987; Kohda *et al.*, 1988; Mayo *et al.*, 1989; Brown, S.C. *et al.*, 1989; Selander *et al.*, 1990; Kline *et al.*, 1990). All five have similar tertiary structures, characterized by a core composed of the 6 cysteines joined by 4 loops. The secondary structure consists predominantly of β-strands and β-turns. EGF domains containing a β-hydroxyasparagine or β-hydroxyaspatic acid have a high affinity Ca^{2+} binding site (Öhlin *et al.*, 1988). The binding of Ca^{2+} to the EGFs domain of $\bar{C}1r_2$ is thought to be an essential step in the assembly of the $\bar{C}1r_2\bar{C}1s_2$ complex (Aude *et al.*, 1988). The homologous EGF domain in $\bar{C}1s$ however, requires other structural regions in the molecule for Ca^{2+}-dependent interactions (Thielens *et al.*, 1990).

(1.6.5) LDL receptor (LDLr)

The LDLr domain is found as one repeat in terminal components, C6, C7, C8α, C8β and C9 and also as 2 repeats in factor I (Goldberger *et al.*, 1987; Catterall *et al.*, 1987). It also occurs in LDLr as 7 repeats, LDLr-related protein as 31 repeats and GP330 as 15 repeats (Chapter 8). The LDLr is also a cysteine-rich domain, containing 6 conserved half-cysteines. The disulphide connectivity has not yet been determined. It is approximately 40 amino acids in length and has substantially more residue conservation than the other complement domains (Chapter 8). The homologous domain in the LDLr acts as a binding site and it has been proposed that it has the same function in the complement proteins (Scott, 1989; Tschopp & Masson, 1987).

(1.6.6) Serine Protease (SP)

The serine protease domain found in the complement components C1r, C1s, factor I, C2, factor B and factor D (Johnson *et al.*, 1984). All these proteins show sequence homology to the classic serine proteases, trypsin and chymotrypsin. The catalytic triad (His, Asp, Ser) involved in the active site is conserved amongst all serine proteases. Disulphide bridge connectivity is not fully conserved. One disulphide bridge in C1r and C1s, constituting the 'histidine loop' is absent although this is present in all other mammalian serine proteases (Arlaud *et al.*, 1982). The tertiary structures for seven serine proteases have been elucidated (Greer, 1990). The predominant secondary structure is that of β-sheet along with a characteristic C-terminal α-helix (Read & James, 1988). The β-sheets are organized into two sub-domains of similar topology and are arranged as orthogonal sheets (Read & James, 1988). It is the functional domain involved in enzymatic cleavage of peptide bonds.

(1.6.7) von Willebrand Factor (vWF)

The type A domain of the vWF is found in the complement proteins C2, factor B, CR3 (Corbi *et al.*, 1988) and CR4 (Corbi *et al.*, 1987) and also in the

integrins (Takada & Hemler, 1989), cartilage matrix protein (Kiss *et al.*, 1989) and collagen type VI (Chu *et al.*, 1990). It is approximately 202 residues long. There are no conserved Cys residues. The domain was found to contain significant amounts of α-helices and β-strands (Perkins & Smith, unpublished results). The function of the domain in the non-complement proteins is to mediate the binding to collagen and although there is no evidence for the binding of complement components to collagen, factor B will bind to the cell surface of macrophages (Götze *et al.*, 1979).

(1.6.8) Perforin-like region (PLR)

The terminal components of complement, C6, C7, C8α, C8β and C9 have sequence homology to the pore forming protein perforin (Lichtenheld & Podack, 1989) in the central region of the amino acid sequence. The mean length of the domain is around 360 residues and there are two regions of high homology (Chapter 8). Perforin is a 70 kDa glycoprotein present in the granules of cytotoxic T cells and natural killer cells and is responsible for the formation of pore-like lesions in target cells in a similar way to the MAC of complement.

(1.6.9) Factor I Module (FIM)

This repeat is found once in factor I and twice in C6 and C7. The function of this domain is unknown.

(1.6.10) Serine protease inhibitor (SERPIN)

$C\bar{1}$ inhibitor is a member of the SERine Protease INhibitor (SERPIN) superfamily (Carrell & Travis, 1985). Structural comparisons with the crystal structure of the homologous protein α_1-AT show conservation of those features important in inhibitory function and maintaining tertiary structure (Bock *et al.*, 1986) (Chapter 4).

CHAPTER 2.

STRUCTURAL INVESTIGATIVE TECHNIQUES

(2.1) IMPORTANCE OF STRUCTURE DETERMINATION

All biological macromolecules are composed of a specific arrangement of amino acids, carbohydrates, nucleic acids and lipids. These are organized to produce a unique three dimensional structure which has a particular functional role in the living organism. The spatial arrangement of the composite molecules are very important to the function of the macromolecule and a change in conformation may result in loss of activity. In order to understand the function of a macromolecule it is therefore necessary to have a detailed knowledge of its structure.

(2.2) METHODS OF STRUCTURE DETERMINATION

Biological tertiary structure determination can be divided into two categories, those of high and low resolution. High resolution implies a structural determination down to the atomic level, as exemplified by X-ray crystallography and more recently, neutron crystallography and nuclear magnetic resonance (NMR). Over 790 sets of protein atomic coordinates are currently listed in the Brookhaven Protein Data Bank (October 1991) and resolutions of 0.15 nm have been achieved using X-ray crystallography (e.g. lysozyme; Artymiuk & Blake, 1981). Neutron crystallography is complementary to X-ray crystallography and is used to refine X-ray diffraction maps (reviews: Jacrot, 1987a; Moore, 1985). The major problem with all crystallographic techniques is the initial requirement of a crystal. Good crystals are grown in a protein solution under a narrow range of conditions which are empirically determined. Once a suitable crystal has been obtained, basic diffraction experiments will only determine the intensity of the scattered waves. Further preparative techniques are required to elucidate the phase of the waves. The only complement components that have been crystallized so far are C3a (Huber *et al.*, 1980) and more recently factor D (Narayana *et al.*, 1991a, b). This is due to the high degree of glycosylation of complement proteins, for example $C\bar{1}$ inhibitor has a carbohydrate content of 26% by weight (Perkins *et al.*, 1990d), and also because the complement proteins are composed of many mobile domain structures (Chapter 1) (Reid & Day, 1989). The concentration in

human plasma of complement proteins is low making it difficult to obtain sufficient yields necessary for crystallography.

Two-dimensional NMR can be used to study low molecular weight (M_r) proteins (up to 12 kDa) to atomic resolutions. NMR is the only high resolution technique to determine the three dimensional structure in solution i.e. in an environment near to physiological. Problems in resolving the NMR signals arise with larger proteins. However, new techniques including three dimensional NMR and isotopically labelling protein samples have been used to determine structural information on proteins of M_r up to 27.5 kDa (Fesik *et al.*, 1989). Although the structures of proteins studied by both crystallography and NMR tend to be similar, significant differences have been identified between the solution and crystal states, attributable in part to the effects of crystal packing and the use of unphysiological buffers to promote crystallization (Chazin *et al.*, 1988; Nettesheim *et al.*, 1988). By 1989 the solution structure of more than 20 different globular proteins of 33 to 108 residues have been determined (Wright, 1989). Large multi-domain proteins can also be studied using NMR by isolating each independently folded domain. This has been achieved for the 5^{th} and 16^{th} SCR of factor H (Barlow *et al.*, 1991a, b; Norman *et al.*, 1991) and is therefore of use in studying other complement proteins since most consist of small domains of less than 100 residues.

Low resolution techniques include electron microscopy, hydrodynamic measurements and small-angle solution scattering. Electron microscopy can visualize the whole macromolecule *in vacuo* down to a resolution of around 2 nm. Hydrodynamic measurements give information on M_r and shape of a macromolecule in solution. Small-angle solution scattering also is used to study gross shape parameters of macromolecules in solution to a resolution of 2 to 4 nm, and is discussed in some detail below.

The secondary structure of proteins can determined by the high resolution techniques indicated above but also by spectroscopic and predictive methods. Fourier Transform Infrared (FT-IR) spectroscopy (reviews Surewicz & Mansch,

1988; Byler & Susi, 1986; Lee & Chapman, 1986) allows qualitative analysis of secondary structures by deconvolution of the amide I band (1600 to 1700 cm^{-1}). Secondary structure conformations in ^2H_2O have typical frequencies. Percentages of α-helices, β-sheets, β-turns and random coil can be estimated by adding together all the areas under the bands assigned to these conformations and expressing the sum as a fraction of the total area under the amide I band. Many methods have been developed to predict the protein secondary structure from the primary structure and have an accuracy of around 60% on a residue by residue basis. The methods of Robson (Garnier *et al.*, 1978) and Chou and Fasman (Chou & Fasman, 1978) are discussed in detail below.

(2.3) SMALL-ANGLE SOLUTION SCATTERING

(2.3.1) Coherent and incoherent scattering

Diffraction occurs as the result of interference between scattered waves when radiation, for example, X-rays and neutrons, interacts with matter. X-rays are scattered by electrons and neutrons by atomic nuclei. Scattering can either be coherent or incoherent. Coherent scattering implies no change in wavelength but a phase-shift of π with respect to the incident wave. This produces the interference effect typical of solution scattering (Engelman & Moore, 1975). Incoherent scattering produces random scattering over a solid angle of 4π and is negligible at small angle for all atoms by X-ray scattering. Incoherent neutron scattering only occurs in nuclei which have spin and is only significant for ^{1}H. This does not contribute to the interference effect but produces a high background level (Engelman & Moore, 1975). Scattering by H_2O is therefore essentially isotropic and is used in instrument calibration (see Chapter 3).

(2.3.2) Scattering lengths

The coherent scattering length, or scattering amplitude of an atom is a measure of the probability of a coherent scattering event occurring. In X-ray scattering the scattering length (f) is proportional to the number of electrons in the atom. The scattering length of 1 electron (f_e) is 2.81 x 10^{-15} m (2.81 fm). Hydrogen, with just one electron is therefore very difficult to detect by X-ray diffraction. The neutron coherent scattering length (b) has no simple relationship to the atomic number and is similar for all atoms (Figure 2.1) (Schoenborn & Nunes, 1972). The large difference in scattering lengths of ^{1}H (-3.742 fm) and its isotope ^{2}H (6.671 fm) is very important in neutron solution scattering. The relatively large value for ^{1}H infers that it is comparatively easy to detect by neutron scattering and the minus sign means that there is no phase-shift of π with respect to the incident wave.

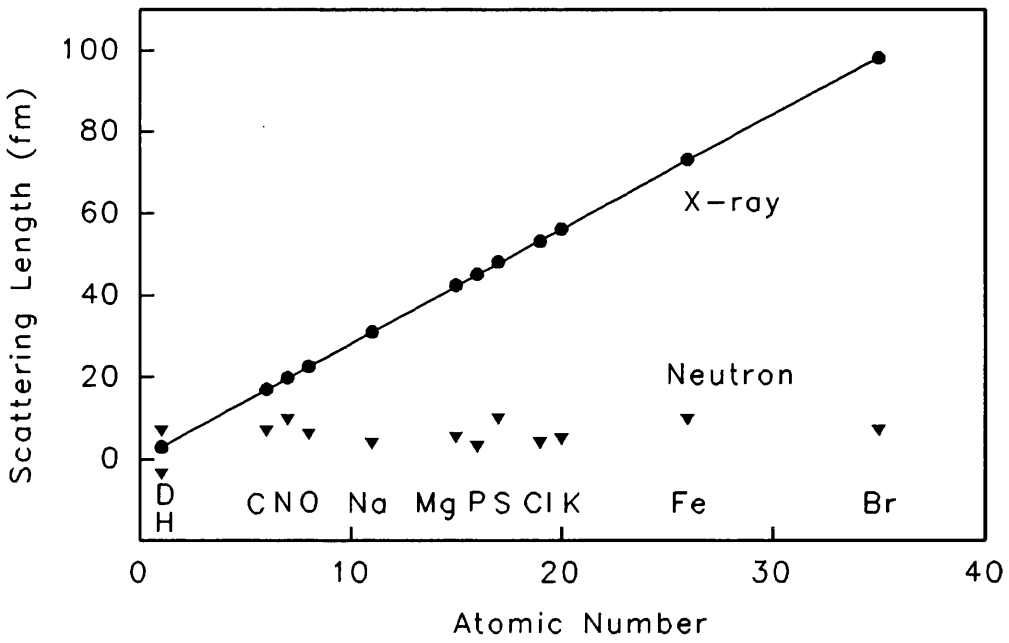


Figure 2.1. The relationship between scattering length and atomic number. The X-ray (●) and neutron (▾) scattering lengths for biologically important atoms are plotted against their atomic number. ^{2}H (D) and ^{1}H have the same scattering length by X-rays but by neutrons ^{2}H has a positive scattering length whilst ^{1}H has a negative scattering length.

(2.3.3) Small-angle scattering

Waves scattered from two points within a macromolecule will cancel out when the path-difference between them is $\lambda/2$. The larger the macromolecule, the smaller the angle required to achieve a path-difference of $\lambda/2$ and the scattering curve will occupy a narrower angular range. Hence the shape of the scattering curve reflects the particle shape in reverse. This is related to the general principle of reciprocity between the dimension of the scattering particle and the scattering angle and is represented by λ/D where D is the size of the particle in solution.

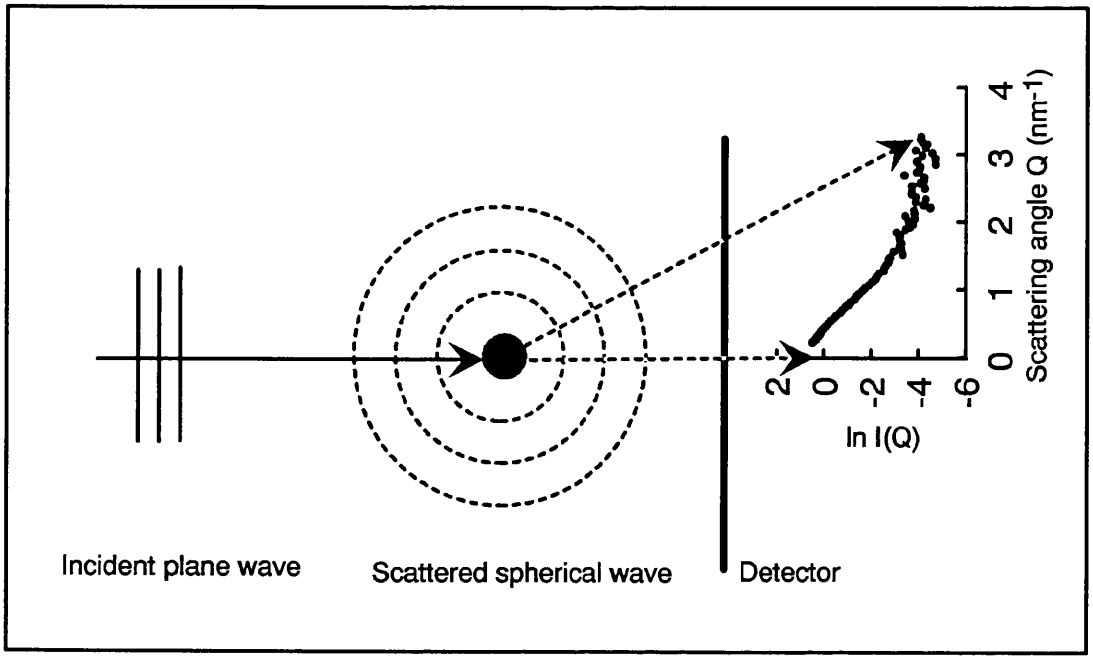


Figure 2.2. Diffraction of electromagnetic radiation.

The incident planar wave, denoted as $\psi(z) = \exp(ikz)$ where $k = 2\pi/\lambda$, λ is the wavelength and r is the distance of the scattering particle from the observer, excites the scattering particle to radiate a spherical wave:

$$\psi(r) = \frac{b}{r}\exp(ikr)$$

where b is the scattering length of the particle. The intensity of scattering as a function of scattering angle Q, is measured at the detector. The plot of ln $I(Q)$ against Q gives a scattering curve characteristic of the protein.

(2.3.4) Derivation of the Debye equation

An incident planar wave, of wavelength λ, is scattered by a macromolecule in the form of a spherical wave (Figure 2.2). The intensity of the radiation is measured with a planar detector as a function of the scattered angle, 2θ relative to the direction of the incident wave. The directions of the incident and scattered waves from the origin O are defined by the unit vectors $\boldsymbol{s}_o$ and $\boldsymbol{s}$ respectively, both of amplitude $2\pi/\lambda$ (Figure 2.3). The scattering vector $\boldsymbol{Q}$ is defined as $(\boldsymbol{s} - \boldsymbol{s}_o)$ and has amplitude $|Q| = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle. A scattered wave from a second identical scattering point P will interfere with that from O. Both waves are exactly in phase when $2\theta = 0$ and the intensity of the scattered wave is at a maximum. At angles greater than zero the interference between the waves due to the phase-difference reduce the intensity dramatically (Figure 2.2). If P and O are separated by a vector $\boldsymbol{r}$, the phase-difference will be $\boldsymbol{r}.\boldsymbol{Q}$. The amplitude of scattered radiation from P along vector $\boldsymbol{s}$ is given by:

$$A_p = A_e f_p \exp(-ir.Q) \qquad (1)$$

where f_p is the X-ray scattering factor at P and A_e is the amplitude of the wave scattered by one electron (Guinier & Fournet, 1955). The terms f_p and A_e are replaced by the analogous terms b_p and A_n in neutron scattering. The total amplitude $A(Q)$ scattered by a macromolecule is simply the sum of all the individual amplitudes. The intensity of scattering $I(Q)$ is the real part of the product of $A(Q)$ and its complex conjugate $A(Q)^*$:

$$I(Q) = A_e^2 \Sigma\Sigma f_p f_q \cos(rQ) = I_e F^2(Q) \qquad (2)$$

where $F(Q)$ is the structure factor of the macromolecule and is defined as the ratio of the total scattered amplitude to the amplitude of radiation scattered by one electron at the origin (Guinier & Fournet, 1955). $I(Q)$ depends only on $\boldsymbol{Q}$ and not on the orientation of the sample and is spherically symmetrical about the origin of reciprocal space. $I(Q)$ is averaged over all possible orientations to give the Debye equation:

$$\langle I(Q) \rangle = \Sigma\Sigma f_p f_q \frac{\sin(rQ)}{rQ} \qquad (3)$$

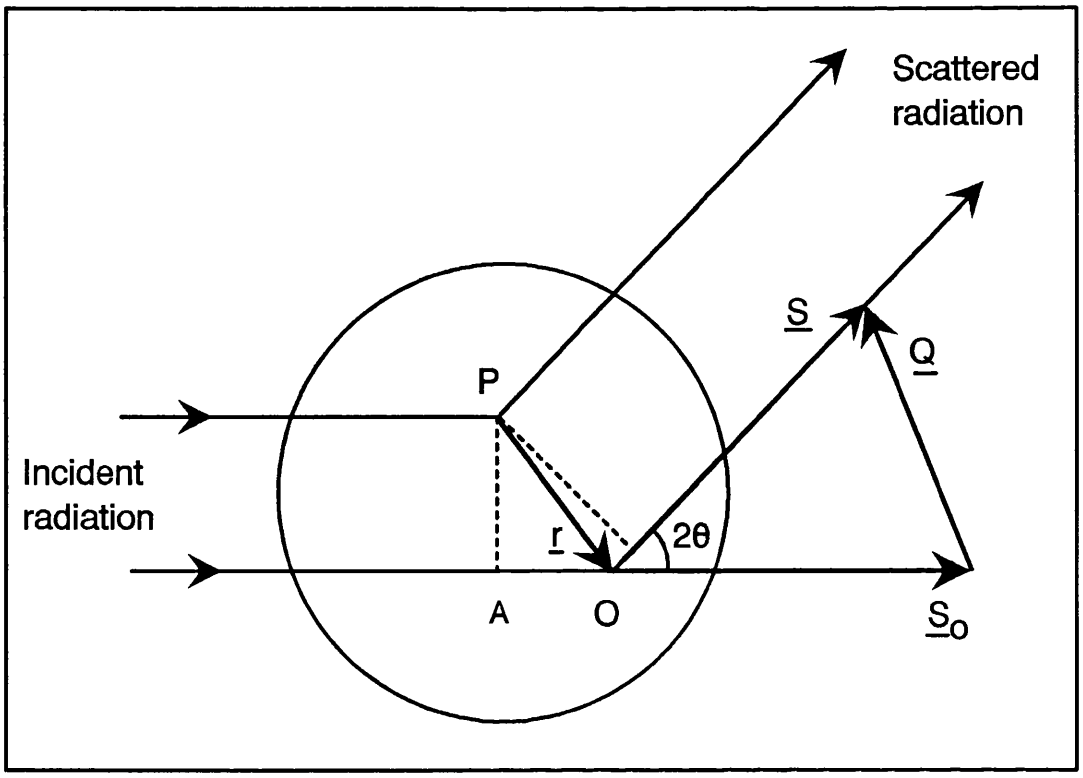


Figure 2.3. Derivation of the scattering angle Q.

Incident radiation is scattered from two points *P* and *O* separated by a distance *r* through an angle 2θ. The phase-difference is $AO + OB$. The unit vectors s_o and s represent the directions of the incident and scattered radiation respectively and Q is the scattering vector $(s - s_o)$.

(the angled brackets denoted the average of the enclosed term) with the following assumptions (Jacrot, 1982; Zaccai & Jacrot, 1983):

1) the solution is infinitely dilute and the waves scattered from different particles do not interact;

2) macromolecules are in random motion and the vector $\boldsymbol{r}$ occurs in all orientations equally;

3) thermodynamically the solution is a two component system and the scattering densities of the solute and solvent are homogenous;

4) all the macromolecules are identical;

5) the scattering is only coherent;

6) point-shaped primary beam cross-section (no collimation effect);

7) the measurements are performed *in vacuo*.

The scattering curve from a dilute system of randomly moving macromolecules is therefore the sum of individual particle scattering intensities averaged over all spatial orientations. Because the intensity is averaged, $F(Q)$ cannot be determined directly as is possible by crystallography and therefore there is a decrease in resolution.

(2.3.5) Scattering densities

The Debye equation above (3) assumes the macromolecule to be *in vacuo* but a macromolecule in solution has a modified scattering intensity and the equation must therefore be adapted. At low scattering angle, with a wavelength between 0.01 to 0.1 nm, the atom-to-atom variations in scattering length do not contribute to the overall scattering. Therefore a macromolecule can be considered to consist of one or a few large regions of average scattering length per unit volume (scattering density) (Engelman & Moore, 1975). At a point P in the macromolecule, the scattering length f_p (or b_p for neutrons) is equal to the scattering length density $\rho(r_p)$ in a small volume dv_p surrounding the point, hence $f_p = \rho(r_p)dv_p$. The average scattering density of the whole macromolecule is:

$$\rho_V = \frac{1}{V}\int \rho(r)\, d^3r \qquad (4)$$

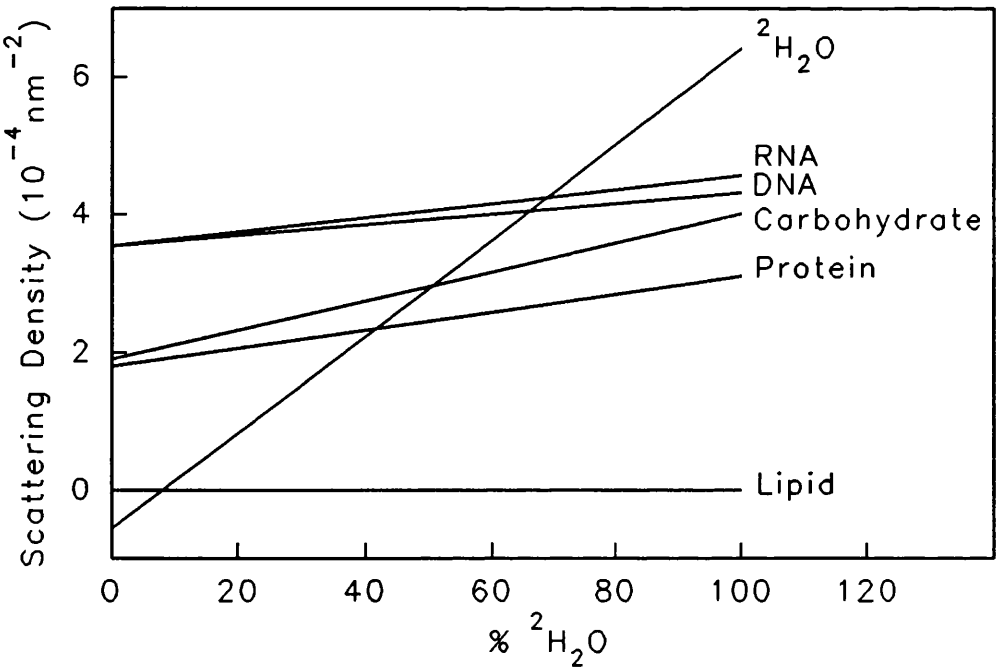


Figure 2.4. The average scattering density of the major biological macromolecules as a function of the ^2H_2O content of the solvent (Jacrot, 1987). The % ^2H_2O at the point at which the scattering density of the constituent intersects the ^2H_2O line is called the matchpoint. The matchpoint value for lipids is 10 to 14 % ^2H_2O, for proteins, 40 to 45 % ^2H_2O, for carbohydrates, 47 % ^2H_2O, for DNA, 65 % ^2H_2O and for RNA, 72 % ^2H_2O (Perkins, 1988).

where V is the macromolecule volume impermeable to solvent. Assuming the solvent to be of uniform density, it can be represented as ρ_S. Scattering occurs when there is a difference in scattering density between the solute and solvent:

$$\Delta\rho = \rho_v - \rho_s \qquad (5)$$

where $\Delta\rho$ is defined as the excess scattering density of the solute above that of the solvent. This difference is analogous to the use of staining techniques used to visualize specimens in both light and electron microscopy. The dependence of the scattered intensity on the contrast is determined by substituting into the Debye equation and integrating over the macromolecular volume:

$$\langle I\ (Q) \rangle = \int\int (\rho\,(r_p) - \rho_S)\, \frac{\sin(rQ)}{rQ}\, d^3r_p d^3r_q \qquad (6)$$

The contrast $\Delta\rho$ and therefore the scattering intensity can be varied by altering either the solute or solvent scattering density. The solvent scattering density can be altered easily in the case of neutron scattering because the densities of H_2O and ^2H_2O are very different (-0.562×10^{-4} nm^{-2} and 6.404×10^{-4} nm^{-2} respectively). Changing the solute scattering density is less straight forward and is achieved by either chemical labelling and/or *in vivo* deuteration of non-labile hydrogen atoms (Worcester, 1988).

Equation (6) above assumes that the scattering density of the solute is homogenous. However, the major biological macromolecular constituents have significantly different scattering densities (Figure 2.4) (Jacrot, 1987b). This is due to the different amounts of non-exchangeable protons found in each class. Lipids have a very high ^1H content whereas nucleic acids have a low content. Amino acids can be further subdivided into two groups, those with high scattering densities and those with lower scattering densities (Table 2.1) (Jacrot, 1976; Perkins, 1988). The hydrophilic amino acids tend to have higher scattering densities than the hydrophobic ones. The scattering density of H_2O : ^2H_2O mixtures varies linearly with % ^2H_2O and spans the range of scattering densities of the major biological constituents. The scattering curve can therefore be considered to be composed from the scattering due to the contrast between the

Name	Composition	Exchangeable protons	Electrons	Σb (fm) (H_2O)	(^2H_2O)	Volume ($\times 10^{-3}$ nm^3)	electron density ($e\ nm^{-3}$)	Neutron mtpt (% ^2H_2O)
Ile	$C_6H_{11}NO$	1	62	13.958	24.361	168.8	367	21.9
Phe	C_9H_9NO	1	78	41.385	51.798	203.4	383	40.2
Val	C_5H_9NO	1	54	14.781	25.194	141.7	381	25.8
Leu	$C_6H_{11}NO$	1	62	13.948	24.361	167.9	369	21.9
Trp	$C_{11}H_{10}N_2O$	2	98	60.345	81.171	237.6	412	50.9
Met	$C_5H_9NO_5$	1	70	17.628	28.041	170.8	410	25.1
Ala	C_3H_5NO	1	38	16.447	26.860	91.5	415	40.5
Gly	C_2H_3NO	1	30	17.280	27.693	66.4	452	58.6
Cys	C_3H_4NOS	1	53	23.036	33.449	105.6	502	45.9
Tyr	$C_9H_9NO_2$	2	86	47.189	68.015	203.6	422	48.5
Pro	C_5H_7NO	0	52	22.265	22.265	129.3	402	32.8
Thr	$C_4H_7NO_2$	2	54	21.418	42.244	122.1	442	44.0
Ser	$C_3H_5NO_2$	2	46	22.251	43.077	99.1	464	57.7
His	$C_6H_8N_3O$	3	73	43.974	75.213	167.3	436	62.6
Glu	$C_5H_6NO_3$	1	67	37.615	48.028	155.1	432	47.5
Asn	$C_4H_6N_2O_2$	3	60	34.560	65.799	135.2	444	67.0
Gln	$C_5H_8NO_3$	3	68	33.727	64.966	161.1	422	52.8
Asp	$C_4H_4NO_3$	1	63	38.448	48.861	124.5	506	59.6
Lys	$C_6H_{13}N_2O$	4	71	15.864	57.516	171.3	414	32.8
Arg	$C_6H_{13}N_4O$	6	85	34.664	97.142	202.1	421	58.8
Glc	$C_6H_{10}O_5$	3	86	31.506	62.745	171.9	500	46.5
Gal	$C_6H_{10}O_5$	3	86	31.506	62.745	166.8	516	48.1
Man	$C_6H_{10}O_5$	3	86	31.506	62.745	170.8	504	46.9
Gln	$C_8H_{13}NO_5$	3	108	42.982	74.221	222.0	486	44.9
Gan	$C_8H_{13}NO_5$	3	108	42.982	74.221	232.0	464	42.8
Fuc	$C_6H_{10}O_4$	3	78	25.702	56.941	160.8	485	43.0
Neu	$C_{11}H_{16}NO_8$	5	153	69.121	121.186	326.3	469	49.9

Table 2.1. Scattering properties of the twenty amino acids and carbohydrate residues. The residues are listed in increasing hydrophilicity, according the scale of Eisenberg (1984). The electron densities are calculated assuming an unhydrated volume and the matchpoint assumes full solvent ^1H - ^2H exchange (Perkins, 1986; Perkins, 1988; Chapter 3).

solute and solvent and also that due to internal contrast variations within the macromolecule. However, because amino acids contain labile protons (N-H and O-H) which are able to exchange with ^2H in the solvent, the scattering density of a macromolecule is not constant but depends on the H_2O : ^2H_2O composition of the solvent. One further complication is that the proton exchange is not uniform over the macromolecular volume because the distribution of labile protons within the macromolecule is not uniform and hence the internal structure is itself a function of contrast. If the amino acid composition of the macromolecule is

known the number of labile protons can be determined. However even in 100% ^2H_2O, only 90% of these will be accessible to the solvent and exchange, the others being involved in secondary structure interactions (Perkins, 1986).

Contrast variation by X-rays is more problematic since the solvent scattering density must be altered by adding small molecules such as sucrose. Since biological molecules are sensitive to the surrounding solvent, X-ray contrast variation experiments are rarely done (Zaccaï, 1978; Moore, 1988).

(2.3.6) Differences between neutron and X-ray scattering

(a) Neutrons are scattered by atomic nuclei, X-rays by electrons.

This has a major implication in the solute volume measured by these two techniques. A protein in solution has associated with it a layer of water molecules of different density compared to the bulk (free) water. This water layer causes X-rays to scatter but does not affect neutrons. Although the Guinier parameters calculated by X-ray scattering should be comparable with those calculated in 0% ^2H_2O buffers by neutrons, often the R_G is slightly larger by X-ray scattering because of this hydration shell. It is important to incorporate this extra volume when modelling (Chapter 3). Since these water molecules are in close proximity to charged groups on the protein surface, the volume occupied by a water molecule will be smaller (i.e. electrostriction). The hydration shell surrounding the protein is calculated assuming 0.3 g of water / g glycoprotein and an electrostricted volume of 0.0245 nm^3 per water molecule, compared to a volume of 0.0299 nm^3 of free water (Perkins, 1986).

(b) The instrumental effects on the incident beam.

The synchrotron X-ray incident radiation is monochromatic and point collimated whereas the neutron radiation is typically subject to a wavelength spread on 10% and slit width and slit height smearing errors.

(c) The effect of beam on the protein.

X-rays produce free radicals and therefore are destructive to protein and samples cannot be reused. Some proteins are very sensitive to X-ray radiation and will aggregate very quickly. Neutron radiation is not destructive and samples can

be remeasured.

(d) Counting statistics in 0% ^2H_2O.

The counting statistics are much better in 0% ^2H_2O by X-ray scattering than by neutron scattering. This is due to the high amount of incoherent scattering and decreased intensity of the neutron beam passing through water.

(e) Contrast variation.

It is very easy to change the scattering density of the solvent by neutron scattering, by simply altering the ratio of ^2H_2O to H_2O. The corresponding procedure by X-ray scattering requires the addition of sucrose or ethanol which may affect the conformation of the macromolecule. The scattering density of the solute can be changed in neutron scattering by selective deuteration of a particular functional group to enhance spatial resolution.

(2.4) COMPUTATIONAL TECHNIQUES

(2.4.1) Analytical techniques

(2.4.1.1) Guinier plot

When Q is small, the Debye equation approximates to a Gaussian function. This can be simplified to give the Guinier equation:

$$\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3} \qquad (7)$$

where the R_G (radius of gyration) is defined as the root-mean-square distance of all scattering centres (electrons or neutrons) from their centre of gravity and $I(0)$ is the intensity of scattering in the direction of the incident beam. These are easily determined by plotting ln $I(Q)$ against Q^2 to get a straight line of negative slope $R_G^2/3$ and y-axis intercept ln $I(0)$. The R_G is a measure of the elongation of a molecule and is independent on shape assumptions. The most compact shape, a sphere, of volume V has the lowest R_G (R_0). The ratio R_G/R_0 is a measure of the anisotropy of a macromolecule and is approximately 1.28 for most glycoproteins (Perkins, 1988). The R_G values for several simple geometric shapes are given in Table 2.2.

Sphere (radius r)	$R_G^2 = 3/5\ r^2$
Ellipsoid (semi-axes a,b,c)	$R_G^2 = (a^2 + b^2 + c^2)/5$
Elliptical cylinder (height h, semi-axes a, b)	$R_G^2 = (3a^2 + 3b^2 + h^2)/12$

Table 2.2. The radii of gyration of simple geometrical bodies.

The smallest Q to yield an R_G is $Q_{min}D_{max} \leq \pi$ where D_{max} is the maximum dimension of the particle. The largest Q used depends upon the macromolecular shape. The Guinier equation approximation breaks down when $Q.R_G$ is greater than 1. For mildly elliptical particles such as most globular proteins, a $Q.R_G$

range out to 1.5 is acceptable (Perkins, 1988). The R_G can be used to follow changes of molecular shape with time for example, dissociation and association effects, conformational changes by denaturation, binding of coenzymes, temperature effects, etc.

(2.4.1.2) Cross-sectional plot

If one macromolecular dimension is large compared to the other two, the analysis of the scattering curve at larger Q will give the radius of gyration of the cross-section R_{XS}:

$$\ln[I(Q).Q] = [\ln(I(Q).Q)]_{Q\to 0} - \frac{R_{XS}^2 Q^2}{2} \qquad (8)$$

A plot of $\ln[I(Q).Q]$ against Q^2 will have a slope of $R_{XS}^2/2$ and y-axis intercept of $[\ln(I(Q).Q)]_{Q\to 0}$. The length L can be determined from the Guinier and cross-sectional parameters by assuming a macromolecular shape of an elliptical cylinder:

$$L=\sqrt{[12(R_G^2-R_{XS}^2)]} \qquad (9)$$

The two shorter dimensions A and B can also be calculated from the dry (neutrons) or hydrated (X-rays) volume V = πABL and $R_{XS}^2 = (A^2 + B^2)/4$. The length can also be calculated assuming a shape of an ellipsoid:

$$L=\frac{\sqrt{(5R_G^2-4R_{XS}^2)}}{2} \qquad (10)$$

The length can be calculated with no assumptions made on the protein shape from the intensity ratios:

$$L=\frac{\pi . I(0)}{[I(Q).Q]_{Q\to 0}} \qquad (11)$$

(iii) Importance of $I(0)$ values

At zero scattering angle, the waves are exactly in phase and the coherent scattering intensity, $I(0)$ depends only on the number and nature of the nuclei and not their spatial arrangement:

$$I(0) = N(\Sigma b_p - \rho_s V)^2 G \qquad (12)$$

where N is the number of particles and G is a factor which represents the geometrical parameters of the experiment (Jacrot & Zaccai, 1981). To a first approximation Σb_p and V are proportional to the relative molecular weight M_r and therefore $I(0)$ is proportional to M_r^2. The factor G is equal to

$$G = T_s I_0 \Omega \qquad (13)$$

where T_S is the transmission of the sample, I_0 is the intensity of the incident neutron beam and Ω is a geometric term accounting for the angular aperture of the incident beam and the solid angle used for detection. The incoherent scattering of H_2O at wavelengths of 1 nm or more is isotropic and the uniform scattering of neutrons over 4π by a sample of thickness 1 mm is approximately 55% of the incident beam. Since other processes which remove neutrons from the main beam are very small, the transmission (T_W) of neutrons through the sample is approximately (1 - 0.55). The number of neutrons scattered per solid angle is:

$$I_{inc} = \frac{1}{4\pi} (1 - T_w) I_0 S \Omega \ (neutrons/s) \qquad (14)$$

where I_{inc} is the intensity of incoherent scattering. I_{inc} can be determined experimentally by measuring the transmission of a water sample subtracted from the transmission due to an empty container. Providing the samples are also measured under identical conditions,

$$I_0 \Omega = \frac{I_{inc}(0)\, 4\pi}{(1 - T_w)\, S} \qquad (15)$$

where $I_{inc}(0)$ is the incoherent scattering of H_2O at zero scattering angle and S is the cross-sectional area of the sample. The term G can now be written as

$$G = \frac{T_s I_{inc}(0)\, 4\pi}{(1 - T_w)\, S} \qquad (16)$$

The number of particles N, is determined from:

$$N = \frac{c N_a S t}{M_r} \times 10^{-3} \qquad (17)$$

where c is the concentration, t is the pathlength and N_a is Avogadro's number. The combination of these equations gives:

$$\frac{I(0)}{cI_{inc}(0)} = N_a t M_r \times 10^{-3} \left(\frac{1}{M_r}(\Sigma b - \rho_s V)\right)^2 \frac{4\pi T_s}{1-T_w} \qquad (18)$$

$I(0)/c$ is hence proportional to the M_r. $I(0)$ can be calculated from the Guinier plot and c from the absorbance at 280 nm. The parameter $\Sigma b/M_r$ can determined from the amino acid composition. The average value of $\Sigma b/M_r$ for a protein is 2.27 x 10^{-14} cm/Da and the deviation from this mean is less than 3% in 80% of cases (Jacrot & Zaccai, 1981). The particle volume V, is the dry volume and is related to the M_r by:

$$V = \frac{M_r \bar{V}}{N_a} \qquad (19)$$

where $\bar{v}$ is the partial specific volume of the particle. In H_2O the scattering density of the solvent ρ_s, is very small and the difficulties in measuring $\bar{v}$ accurately become less significant. The mean value of $\bar{v}$ for a protein is 0.74 cm^2g^{-1}. An error of 5% in $\bar{v}$ gives an error of 1.5% in the M_r (Jacrot & Zaccai, 1981). The determination of the M_r by X-rays is more problematic. The sample intensity is normalized with respect to the detector response rather than H_2O as by neutrons and the detector repsonse varies from session to session. Also by X-rays, an error in $\bar{v}$ leads to a much larger error in M_r, so in X-ray scattering the relative M_r is determined by comparison with a sample of known M_r in the same buffer measured in the same experimental session. The M_r is also obtainable from the amino acid and carbohydrate composition of the macromolecule and can be compared to the experimentally derived value. The accuracy should be within 5% (Jacrot, 1987b) with the limiting factor being the calculation of the macromolecular concentration. Since aggregation is immediately detectable, the determination of a correct M_r is a good control for a good Guinier plot.

(2.4.1.4) Matchpoint determination

From equation (12) a plot of $\sqrt{}$ $[I(0)/cT_st]$, where $I(0)$ is normalised to concentration (c), transmission (T_s) and path length (t), against the solvent

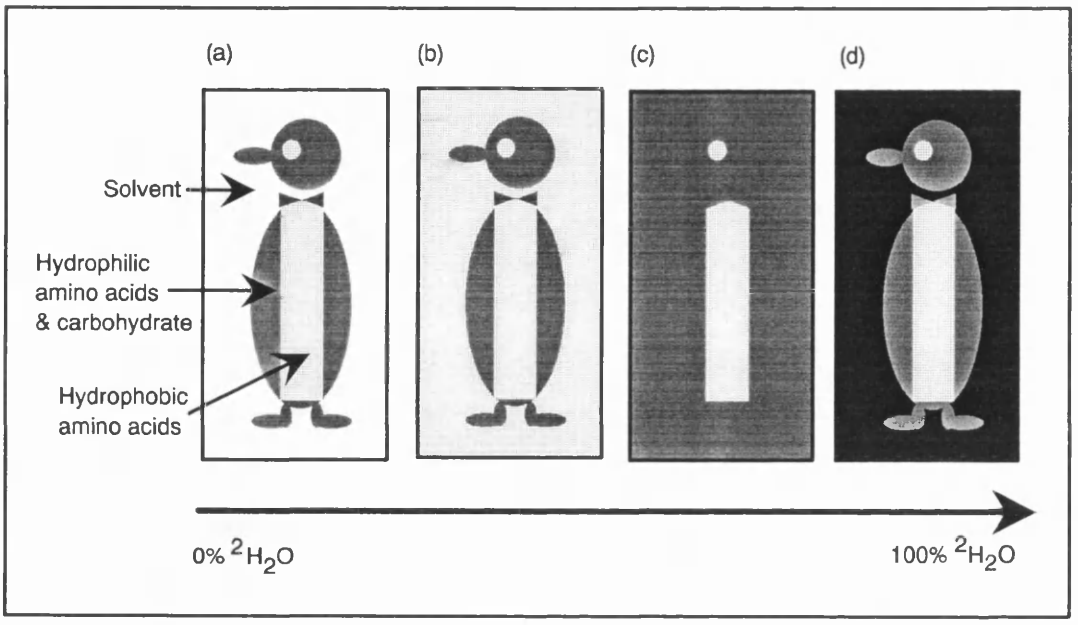


Figure 2.5. The effects of contrast variation on an idealized protein.

In a solvent of 0% ^2H_2O (Penguin (a)) both the hydrophobic and hydrophilic regions of the protein have a higher scattering density than the solvent (positive contrast). Penguin (b) has a hydrophobic core which has the same scattering density as the solvent and the protein is 'visualized' as being hollow. When the solvent ^2H_2O to H_2O ratio is increased further, (Penguin (c)) the hydrophilic region becomes 'invisible' and only the smaller hydrophobic core is seen. Finally in 100% ^2H_2O (Penguin (d)) both hydrophilic and hydrophobic regions are visible against the solvent but have a lower scattering density than the solvent (negative contrast).

scattering density ρ_S produces a straight line. When the solvent scattering density equals the scattering density of the solute, i.e. $\rho_S V = \Sigma b_p$, the contrast $\Delta\rho$ and $I(0)$ are zero. This is termed the solute matchpoint. At this point, the solute will become indistinguishable from the solvent and effectively invisible since the forward scattering intensity is zero (Figure 2.5). This is analogous to the case in light scattering where a particle is not seen if its refractive index is the same as that of the surrounding medium. One constituent of, for example, a 2 component macromolecule can be matched out, leaving only the second 'visible'. Therefore by altering the contrast over the range encompassed by the major groups of biological constituents, the internal composition of the macromolecule can be established. The matchpoint for a protein can also be derived theoretically (Perkins, 1986) assuming 10% of the main-chain peptide protons do not exchange (Chapter 3).

(2.4.1.5) Stuhrmann plot

For most biological macromolecules, the scattering density is not uniform. The scattering intensity varies with contrast (Table 2.1) and hence the particle shape parameters (R_G) will also be dependent on contrast. This is accounted for in the Stuhrmann equation (Ibel & Stuhrmann, 1975):

$$R_G^2 = R_{G\text{-}C}^2 + \frac{\alpha}{\Delta\rho} - \frac{\beta}{\Delta\rho^2} \qquad (20)$$

The Stuhrmann graph is the plot of the experimental $R_G{}^2$ against reciprocal contrast ($1/\Delta\rho$). The $R_{G\text{-}C}{}^2$ (y-axis intercept) is the radius of gyration at infinite contrast (i.e when $1/\Delta\rho = 0$) and corresponds to a particle of homogenous scattering density with the same shape and volume as the one in question. The term α is calculated from the tangent to the curve at $1/\Delta\rho = 0$ and reflects the radial distribution of internal scattering density fluctuations. The slope will be positive if there is an outer region of high scattering density surrounding a lower scattering density core and negative if the reverse is true. The R_G of glycoproteins is larger in positive contrasts (0% ^2H_2O) than negative contrasts (80% ^2H_2O and 100% ^2H_2O) due to surface regions of carbohydrate and hydrophilic amino acids which have a higher scattering density than that of the

hydrophobic core (approximately 50% ^2H_2O and 40% ^2H_2O, respectively). The value for α is therefore positive. β is a measure of the curvature of the parabola and corresponds to the distance between centres of different scattering densities. In practice, for globular proteins this value is too small to be measured unless deuteration is used and the Stuhrmann plot becomes simply a straight line of intercept $R_{G\text{-}C}^2$ and gradient α. The R_G by X-rays, which is measured in H_2O buffers is therefore comparable to the R_G given by neutrons in 0% ^2H_2O.

(2.4.1.6) Distance distribution function

The scattering curve $I(Q)$ against Q represents the macromolecular structure in reciprocal space. This can be converted into real space by a Fourier transform of the experimental $I(Q)$ for $0 \leq Q \leq$ infinity.

$$P(r) = \frac{1}{2\pi^2}\int_0^\infty I(Q) \,.\, Qr \,.\, \sin Qr \,.\, dQ \qquad (21)$$

where $P(r)$ is the distance distribution function and corresponds to the number of distances of length r that connect a volume element i with another element k within the same macromolecule. D is the maximum macromolecular dimension and is easily obtained from $P(r)$ because there can be no distances longer than the maximum dimension, i.e. $P(r) = 0$ at $r = D$ (Figure 4.10). $P(r)$ reaches a maximum at the most probable intramolecular distance. However, $I(Q)$ cannot be measured at zero angle because of the beam stop and at large angles there are high signal to noise ratios. To overcome these termination effects, the Indirect Transform Procedure (ITP) of Glatter (1977a, b, 1982) was formulated. The only requirement is an estimation of D_{max}, which may be 1.5 times the actual value of D. An underestimation of D_{max} leads to strong oscillations of the $P(r)$ curve ('Fourier ripples') whereas an overestimation also leads to errors, especially at high r in the $P(r)$ curve. This is due to errors at low Q in the experimental scattering curve which influence $P(r)$ at high r. These errors in the $P(r)$ curve increase the more D_{max} is overestimated (Müller & Glatter, 1982). The experimental curve is smoothed by approximating to a mathematical function by fitting a specified number of B-splines to the data using a least squares fitting routine. If too few splines are used the approximation to the data is poor whereas the use of too

many splines corresponds to an overinterpretation of the experimental data and produces 'Fourier ripples' in the $P(r)$ curve. These numerical instabilities are dampened by the use of a stabilization parameter, the Lagrange multiplier, λ (Glatter, 1977a). The instrumental smearing effects are also removed i.e. wavelength spread, slit length and slit width and the curve is extrapolated to zero Q. The R_G and $I(0)$ (Guinier & Fournet, 1955) can be determined from $P(r)$, using the whole scattering curve rather than just the Guinier region.

$$I(0) = 4\pi const. \int_0^D P(r)\, dr \qquad (22)$$

$$R_G^2 = \frac{\int_0^D P(r)\, r^2 dr}{2\int_0^D P(r)\, dr} \qquad (23)$$

Experimental data typically have errors due to the subtraction of the scattering background and any concentration dependence. These are characterized in the $P(r)$ curve by a small increase in $P(r)$ at zero r and a negative oscillation around D due to interparticle interference effects, respectively. However the influence of background scattering does not cause any problems in practice and the position of the minimum value of $P(r)$ of the oscillation due to concentration effects remains virtually constant and is in good agreement with the maximum dimension (Müller & Glatter, 1982). Obviously the R_G and $I(0)$ values will still be affected by the concentration effect and the points at lowest Q of the experimental scattering curve may have to be deleted. Providing this is not too broad, it is equivalent to extrapolating the data to zero concentration.

(2.4.2) Modelling procedures

(2.4.2.1) Debye spheres

Simulated scattering curves are calculated from models made from small overlapping spheres. These do not represent chemically defined subunits but are used only to approximate a simple structure. Crystal co-ordinates of either the

whole protein structure or a single domain can be converted into a sphere model. If no tertiary structure is available, simple elliptical cylinder models are created based on molecular dimensions derived from the Guinier parameters or from electron micrographs. The scattering density of the protein can be incorporated into the model by either using a single uniform scattering density or dividing the spheres into two regions of different scattering densities. For a model based on a single scattering density the R_G is calculated from Damaschun *et al.* (1979):

$$R_G^2 = \sum_{k=1}^{n} \frac{1}{n} (R_k^2 + r_k^2) \qquad (24)$$

for n spheres, where R_k is the R_G of a sphere and r_k is the distance between the sphere and centre of mass of the model. The scattering curve is calculated according to the equation (Rol'bin *et al.*, 1974):

$$\frac{I(Q)}{I(0)} = g(Q) \frac{n + 2\sum_{j=1}^{m} A_j \frac{\sin Q.r_j}{Q.r_j}}{n^2} \qquad (25)$$

where n is the number of spheres, A_j is the number of distances r_j and m represents the maximum length of the model.

If two scattering densities are used in the model, the $R_{G\text{-}C}$ and Stuhrmann α can be determined. The Stuhrmann α value is calculated from:

$$\alpha_G = n^{-1} \sum_{k=1}^{n} (\rho_k - \rho_m) r_k^2 \qquad (26)$$

where ρ_k is the scattering density of the sphere and ρ_m is the mean scattering density. The scattering curve is calculated from (Rol'bin *et al.*, 1974):

$$\frac{I(Q)}{I(0)} = g(Q) \frac{n_1\rho_1^2 + 2\rho_1^2 \sum_{j=1}^{m} A_j^{11} \frac{\sin Qr_j}{Qr_j} + 2\rho_2^2 \sum_{j=1}^{m} A_j^{22} \frac{\sin Qr_j}{Qr_j} + 2\rho_1\rho_2 \sum_{j=1}^{m} A_j^{12} \frac{\sin Qr_j}{Qr_j}}{(n_1\rho_1 + n_2\rho_2)^2}$$

(27) where the model is constructed from n_1 and n_2 spheres of different scattering densities ρ_1 and ρ_2; $g(Q) = 3(\sin QR - QR \cos QR)^2/Q^6R^6$ (the squared form factor of the spheres of radius R); A_j^{11}, A_j^{22} and A_j^{12} are the number of distances r_j for that increment of j between the spheres 1 and 1, 2 and 2 and 1 and 2

respectively; m is the number of different distances r_j.

(2.4.2.2) Hydrodynamic modelling

Frictional coefficients f were calculated theoretically from:

$$f=\frac{M_r(1-\bar{v}.\rho_{20,w})}{N_a s^o_{20,w}} \qquad (28)$$

where $\bar{v}$ is the partial specific volume, $\rho_{20,w}$ is the density of water at 20°C, N_a is Avogadro's constant, and $s^o_{20,w}$ is the sedimentation coefficient. Frictional ratios f/f_0 were calculated using $f_0 = 6\pi\eta r$ in which the macromolecule was assumed to be a hydrated sphere of radius r. Frictional coefficients were calculated for hydrodynamic models by the modified Oseen tensor procedure of Bloomfield (Garcia de la Torre & Bloomfield, 1977a, b; Perkins, 1985). The Stokes-law frictional coefficient for a given sphere in a multi-subunit structure is dependent upon the hydrodynamic interactions of the other spheres and is accounted for in the Oseen tensor. The Oseen tensor assumes each element to be a point source, and is hence modified to allow for the finite size of each of the n elements. The frictional force $\mathbf{F}_i$ associated with the i^{th} element of hydrodynamic radius r_i is:

$$F_i=6\pi\eta r_i(u_i-v_i)\ (i=i,n) \qquad (29)$$

where $\boldsymbol{u}_i$ is the velocity of the i_{th} element and $\boldsymbol{v}_i$ is the velocity of the solvent would have at the same point if the element was absent. The modified hydrodynamic interaction tensor, $\boldsymbol{T}_{ij}$ for spheres of finite sizes and of different hydrodynamic radii of r_i and r_j is:

$$T_{ij}=\frac{1}{8\pi\eta R_{ij}}\left(I+\frac{R_{ij}R_{ij}}{R^2_{ij}}+\frac{(r_i^2+r_j^2)}{R^2_{ij}}\frac{1}{3}I-\frac{R_{ij}R_{ij}}{R^2_{ij}}\right) \qquad (30)$$

where $\boldsymbol{I}$ is the unit vector and $\boldsymbol{R}_{ij}$ is the vector between i and j. The $\boldsymbol{T}_{ij}$ and $\boldsymbol{F}_j$ terms are related:

$$v_i=v_i^0-\sum_{j=1}^{n}T_{ij}F_j \qquad (31)$$

(excluding terms $i = j$), where $\boldsymbol{v}_i^0$ is the unperturbed solvent velocity when the other elements are absent.

(2.5) PREDICTIVE TECHNIQUES

The amino acid sequence contains sufficient information to define the final folded conformation of a protein in a particular environment. This was demonstrated by the classic experiment by Anfinsen which showed that denatured ribonuclease was able to refold without loss of enzymatic activity (Anfinsen, 1973; Anfinsen *et al.* 1961). Proteins with similar primary structures tend to have homologous tertiary structures but this can also be true of proteins with different primary structures suggesting that there is some redundancy in the rules determining tertiary from primary structure.

It is theoretically possible to determine the final protein structure using the rules governing protein folding. However the number of possible structures is very large. The prediction of the secondary structure therefore is likely to yield better results. Secondary structure prediction implies the assignment of individual amino acid residues to specific conformational states, α-helices, β-sheets, β-turns and random coils, but the tertiary structure must first be determined before the secondary structures can be finalized. The criteria for assigning secondary structures to known tertiary structures must be unambiguous and physically meaningful. Kabsch and Sander produced a set of criteria for assigning secondary structures to crystal coordinates based on hydrogen bonding patterns (Kabsch & Sander, 1983).

At least 20 methods have been published for the prediction of secondary structure from the amino acid sequence (Fasman, 1989). These can be divided into two main groups, those based on statistical methods, using a database of known sequences and structures (for example, Garnier *et al.*, 1978) and those based on stereochemical methods (for example, Lim, 1974). It is the statistical approach to secondary structure prediction which is of interest here. The secondary structure prediction for the whole protein depends on the prediction of the secondary structure for each individual residue. The particular conformational state of one residue depends ultimately on every other residue in the protein. The two most commonly used procedures are discussed in detail below.

(2.5.1) Robson

The Robson secondary structure prediction method is an information theory technique based initially on 26 proteins of known sequence and conformation (Garnier *et al.*, 1978) and this has been updated using a database of 75 proteins (Gibrat *et al.*, 1987). Each residue was unambiguously assigned one of 4 possible conformations, α-helix (H), β-sheet (extended, E), β-turn (T) or random coil (C). The H and E conformations imply runs of at least 4 residues with ϕ, ψ angles approximately representative of a right-handed α-helix and a β-sheet respectively. The T conformation implies a 2-residue turn and the C state is anything that is not H, E or T (Robson & Garnier, 1986). Although in theory the conformation of any particular residue is dependent on every other amino acid in the protein, in practice, it has been established that the conformation of 1 residue is primarily governed by the 8 residues either side of it i.e. $-8 \leq m \leq 8$ to give 17 separations, m (Robson & Pain, 1974; Robson & Suzuki, 1976). There are 20 common amino acid types and 4 conformations for each separation. The total number of parameters is therefore 20 x 4 x 17 = 1360. The information I on the conformation S of the j^{th} residue R, can be approximated as:

$$I(S_j; R_1, R_2 \ldots R_{last}) \approx \sum I(S_j; R_{j+m}) \qquad (32)$$

Information is expressed in terms of the logarithm of probabilities or probability ratios. Equation (32) is evaluated for each conformational state (H, E, C and T) and the conformation with the highest information content is chosen. A large positive value for a α-helix conformation will indicate a strong bias towards helix formation whereas a large negative value will be bias towards non-helix. A small or zero value indicates impartiality. The information values used in equation (32) are calculated from the frequencies observed in the database for each amino acid R and for each conformation by scanning every position j. The information values for the 1978 and 1987 databases are compared for each amino acid and each conformation in Figure 2.6. Although in most cases the two parameter sets give similar values, notable differences are α-helical propensities for Arg, Cys and Trp; the β-sheet propensities for Gln and Pro; the β-turn propensities for Cys, Gln, Ile and Trp and the random coil propensities for Trp.

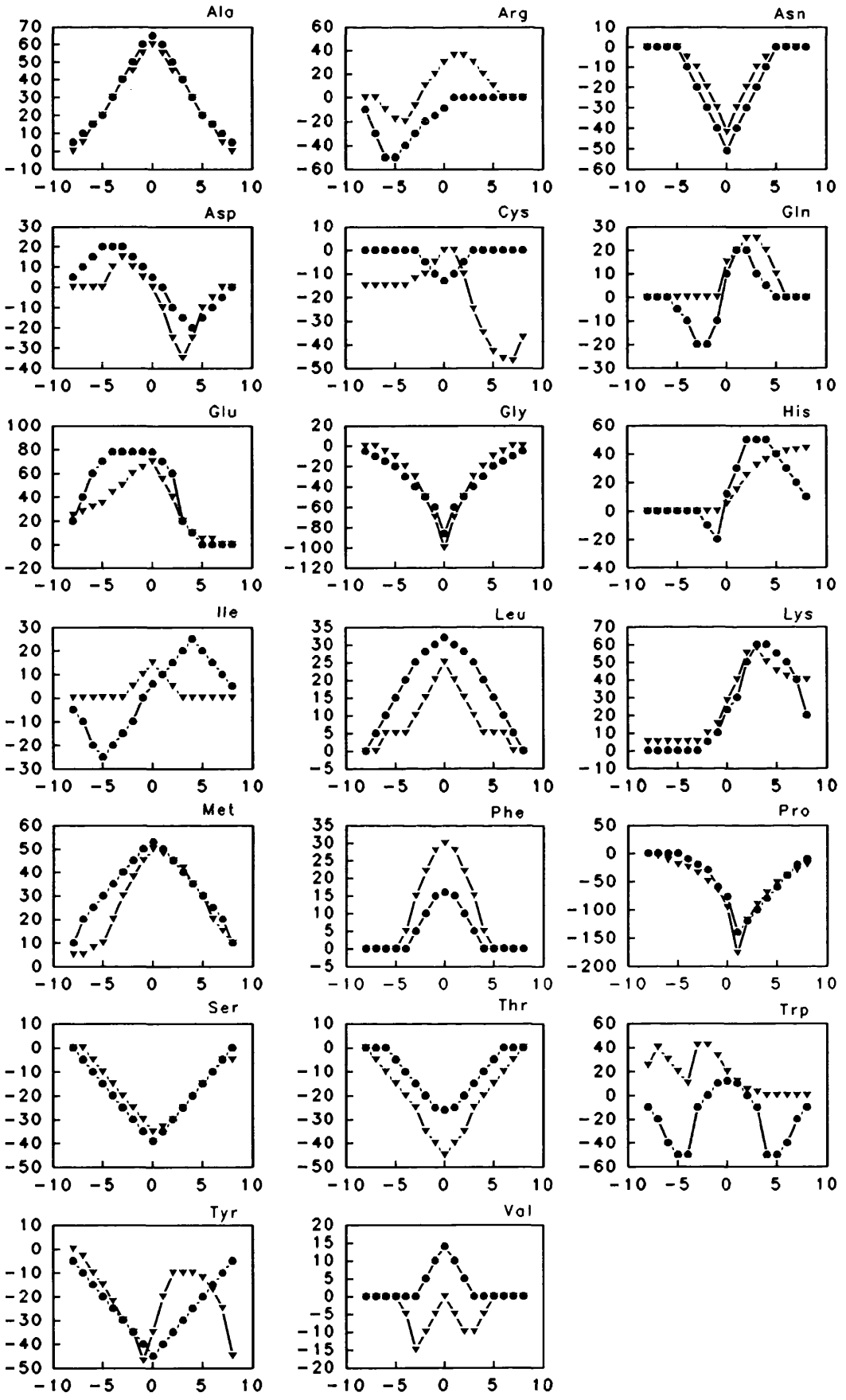


Figure 2.6(a) Alpha propensities

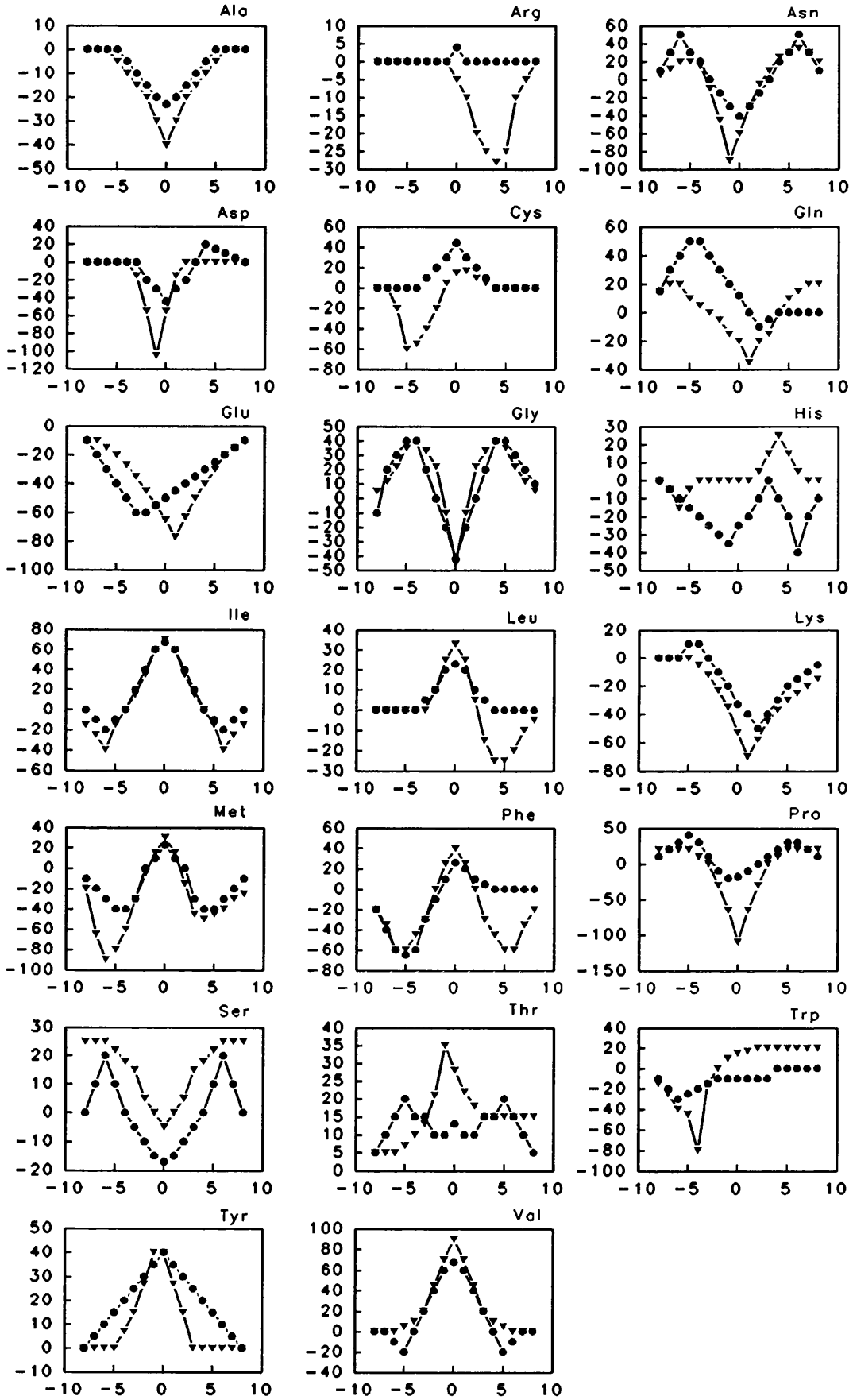


Figure 2.6(b) Beta propensities

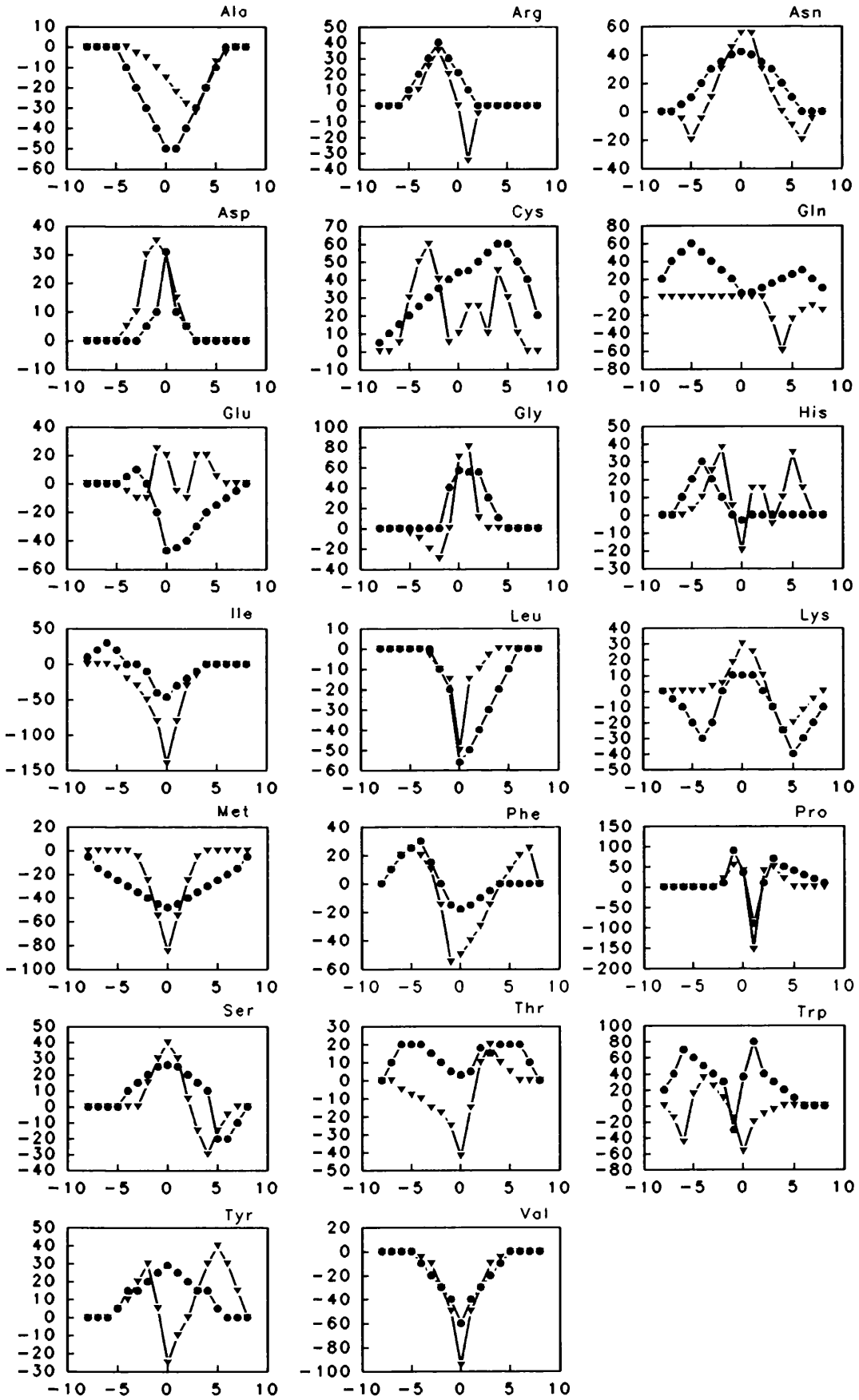


Figure 2.6(c) Turn propensities

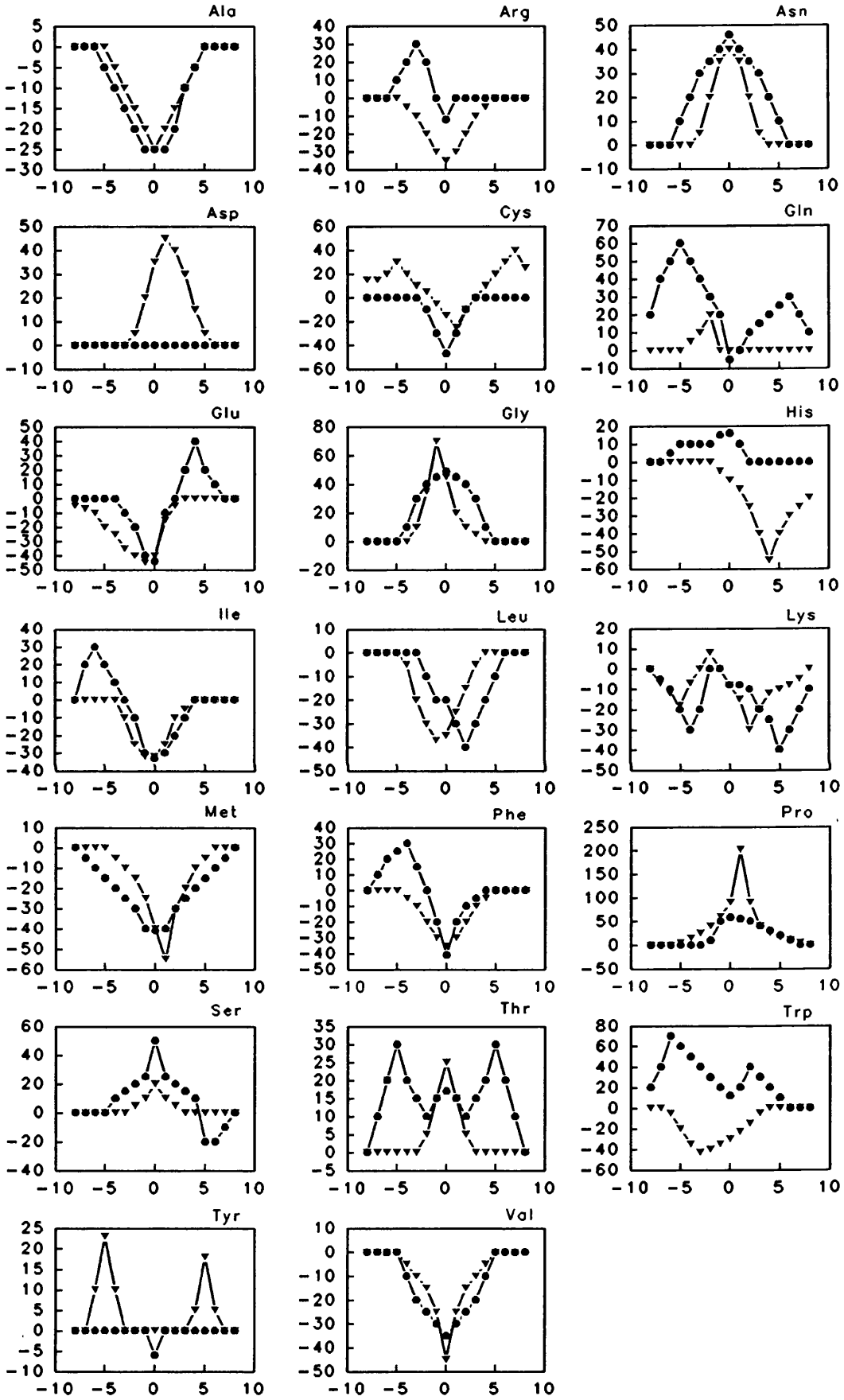


Figure 2.6(d) Coil propensities

	DC_H	DC_E	DC_T	DC_C
β biased	158	-88	0	0
α biased	-10	50	0	0
unbiased	0	0	0	0

Table 2.3. Decision constants for α biased, β biased or unbiased conformation.

Many residues have strong tendencies to form α-helical and β-sheet conformations and these may govern the conformation of less partial neighbouring residues. A region of residues with a strong β-sheet forming tendency may override a nearby region of weak α-helix. The predictions can be improved if experimentally derived information on the secondary structure content is available, for example from circular dichroism (CD) or FT-IR. A constant value (Decision Constant) can be subtracted to favour a particular conformation determined experimentally (Table 2.3) (Garnier *et al.*, 1978). Decision constants are measured in centinats and correspond to units of information content. Only decision constants for α-helix and β-sheet conformations are used because the contribution of turns to the CD spectrum is not understood (Robson & Garnier, 1986). A decision constant of 0 implies that no external information has been used

Figure 2.6. Comparison of the information values for the Robson 1978 and 1987 parameter sets.

The information content for each amino acid (cnats) in each conformation is plotted against residue position *j*, where $-8 \leq j \leq 8$.

The Robson 1978 parameter set is represented by the symbol (●) and the 1987 parameter set by a (▾). A measure of the difference between the old and new parameter sets was calculated as:

$$\sum_{j=-8}^{8} |j_{old} - j_{new}|$$

to weight the raw prediction. If the protein is known to contain less than 20% α-helix and greater than 20% β-sheet, a β-biased prediction is used whereas an α-biased prediction is used when the α-helix content is over 50%.

Using the database of 26 proteins, the overall percentage of correctly predicted conformations for each residue was 60%. Of the four conformations, α-helix was correctly predicted 66%, extended 55%, turn 42% and coil 29% (Garnier *et al.*, 1978). The poor prediction of random coil was due to the low content of this conformation in the proteins tested by Garnier.

(2.5.2) Chou-Fasman

The Chou-Fasman predictive method is based upon a database of 29 known protein structures. The number of times an amino acid was found in a particular conformation was tabulated. Then all the amino acids were classified as being either helix formers, sheet formers, helix breakers, sheet breakers or indifferent (Chou & Fasman, 1978). The α-helix and β-sheet conformational parameters, P_a and P_b were calculated where:

$$P_a = \frac{f_a}{\langle f_a \rangle} \qquad (33)$$

$$P_b = \frac{f_b}{\langle f_b \rangle} \qquad (34)$$

f_a and f_b were the frequency of residues in α-helix and β-sheet regions and $\langle f_a \rangle$ and $\langle f_b \rangle$ were the average frequency of residues in α-helix and β-sheet regions. The α-helical or β-sheet potential for any residue in a protein segment was calculated by averaging the P_a and P_b values for that residue.

To initiate α-helix and β-sheet formation there must be clusters of helix ($\langle P_a \rangle \geq 1.03$) or sheet ($\langle P_b \rangle \geq 1.05$) formers. Extension of the conformation takes place in both directions providing there are no helix or sheet breakers. Termination occurs by a tetrapeptide with a $\langle P_a \rangle$ or $\langle P_b \rangle < 1.00$ or an adjacent region of opposite conformation with a larger potential i.e. if $\langle P_b \rangle > \langle P_a \rangle$, will

terminate an α-helix.

Secondary structure predictions by either method can be enhanced by aligning many homologous proteins (Garnier *et al.*, 1978, Perkins *et al.*, 1988). Interactions that are distant from a particular residue can have substantial effects in determining its conformation but these are not considered in the predictive techniques highlighted above. To predict structures it will always be better to search for a protein of known tertiary structure that is homologous to the protein of interest, since this is likely to have a similar conformation (Creighton, 1984).

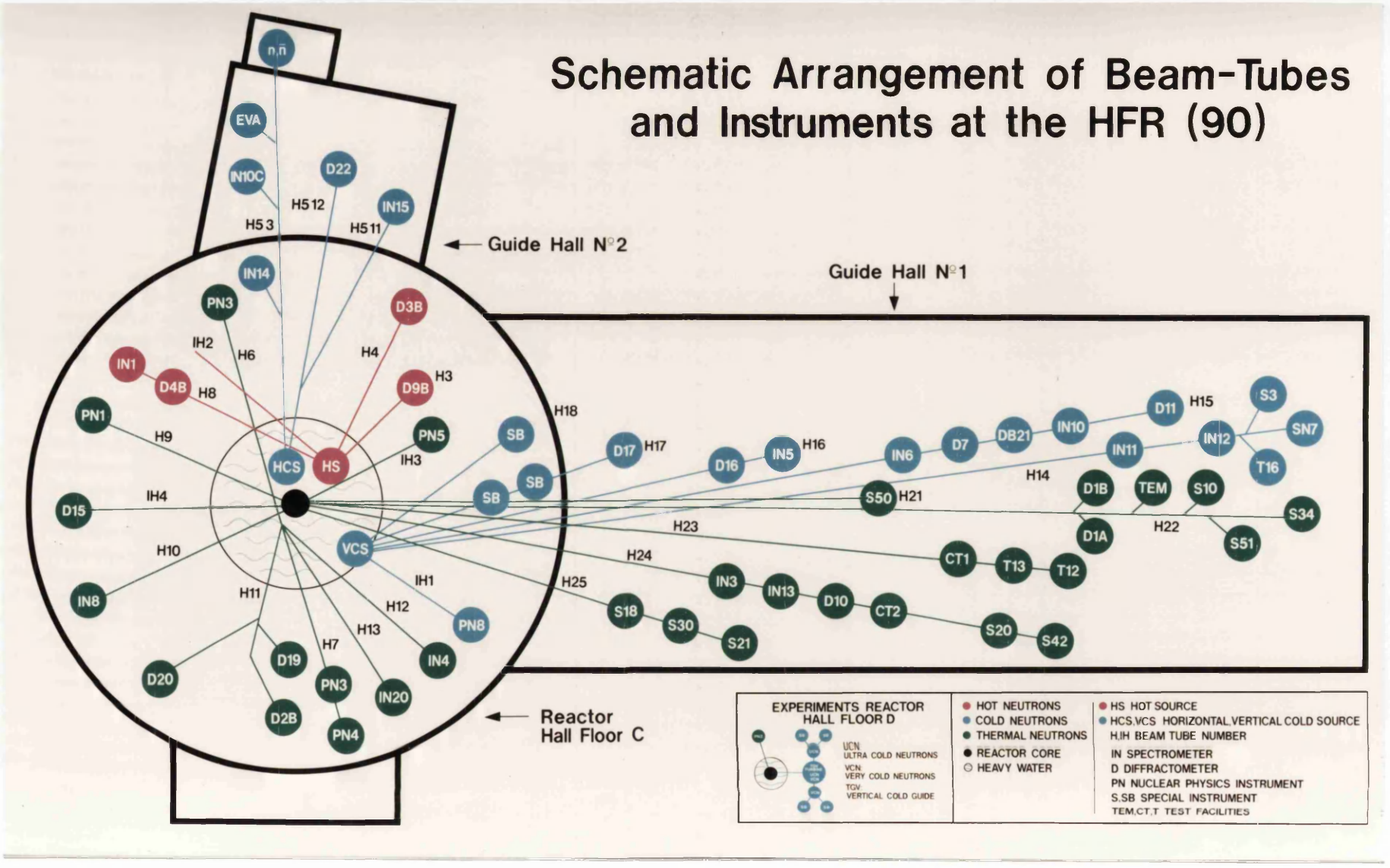
(2.5.3) Hydropathy profiles

Hydrophobicity is a measure of the unfavourable ordering of water molecules around solvent accessible non-polar atoms. A large number of hydrophobicity scales have been derived for amino acids but these are often contradictory and have determined using different criteria. The scales are either solution or empirically derived. Solution measurements are based on distribution coefficients between an aqueous and an organic phase whilst empirical measurements are based on the partitioning between the solvent-accessible surface and the buried interior of proteins of known structure (Fasman, 1989). A comparative analysis of 10 such scales has been reviewed by van Regenmortel & Daney de Marcillac (1988). This found that none of the methods tested had a very high success rate. This was also found by Fasman & Gilbert (1990) when they compared 9 scales on their accuracy to predict the trans-membrane reactive centre protein from *Rhodopseudomonas viridis*. Four scales were used here. Hopp and Woods (1981) use a method based on hydrophilicity scales which select charged or polar residues to identify antigenic determinants on proteins (review Hopp, 1986). The Kyte and Doolittle (1982) method emphasizes the positively charged Arg and Lys residues at the detriment of the negatively charged residues. The hydrophilicity scale of Parker *et al.* (1986) is based on peptide retention times during high performance liquid chromatography and also includes the charged groups at the N- and C-termini of the peptide. The Eisenberg *et al.* (1984) scale

is based on the hydrophobic moment of a protein segment and is used to predict hydrophobic periodicities of protein structures, e.g. α-helices (periodicity of 3.6 residues), 3_{10} helices (periodicity of 2.5 residues) and β-sheets (periodicity of 2.3 residues). Eisenberg (1984) also devised a consensus scale which was the average of four scales in order to alleviate the effects of anomalous values in any one scale.

CHAPTER 3.

METHODS


Schematic Arrangement of Beam-Tubes and Instruments at the HFR (90)
Guide Hall N°2
Guide Hall N°1
Reactor Hall Floor C
n,n̄
EVA
IN10C
D22
IN15
H5 12
H5 3
H5 11
IN14
PN3
D3B
IH2
H6
H4
IN1
D4B
H8
D9B
H3
PN1
H9
PN5
IH3
HCS
HS
SB
SB
SB
H18
D17
H17
D16
IN5
H16
IN6
D7
DB21
IN10
D11
H15
S3
SN7
IN11
IN12
T16
H14
D1B
TEM
S10
S50
H21
S34
D15
IH4
H23
D1A
H22
S51
H10
VCS
H24
CT1
T13
T12
IN3
IN13
D10
CT2
S20
S42
IN8
H11
IH1
H25
S18
S30
S21
H12
PN8
H13
H7
D19
IN4
D20
PN3
IN20
D2B
PN4
EXPERIMENTS REACTOR HALL FLOOR D
UCN ULTRA COLD NEUTRONS
VCN VERY COLD NEUTRONS
TGV VERTICAL COLD GUIDE
HOT NEUTRONS
COLD NEUTRONS
THERMAL NEUTRONS
REACTOR CORE
HEAVY WATER
HS HOT SOURCE
HCS,VCS HORIZONTAL,VERTICAL COLD SOURCE
H,IH BEAM TUBE NUMBER
IN SPECTROMETER
D DIFFRACTOMETER
PN NUCLEAR PHYSICS INSTRUMENT
S,SB SPECIAL INSTRUMENT
TEM,CT,T TEST FACILITIES

(3.1) INSTRUMENTATION

(3.1.1) Small-angle solution scattering by neutrons

(3.1.1.1) Description of the instrument

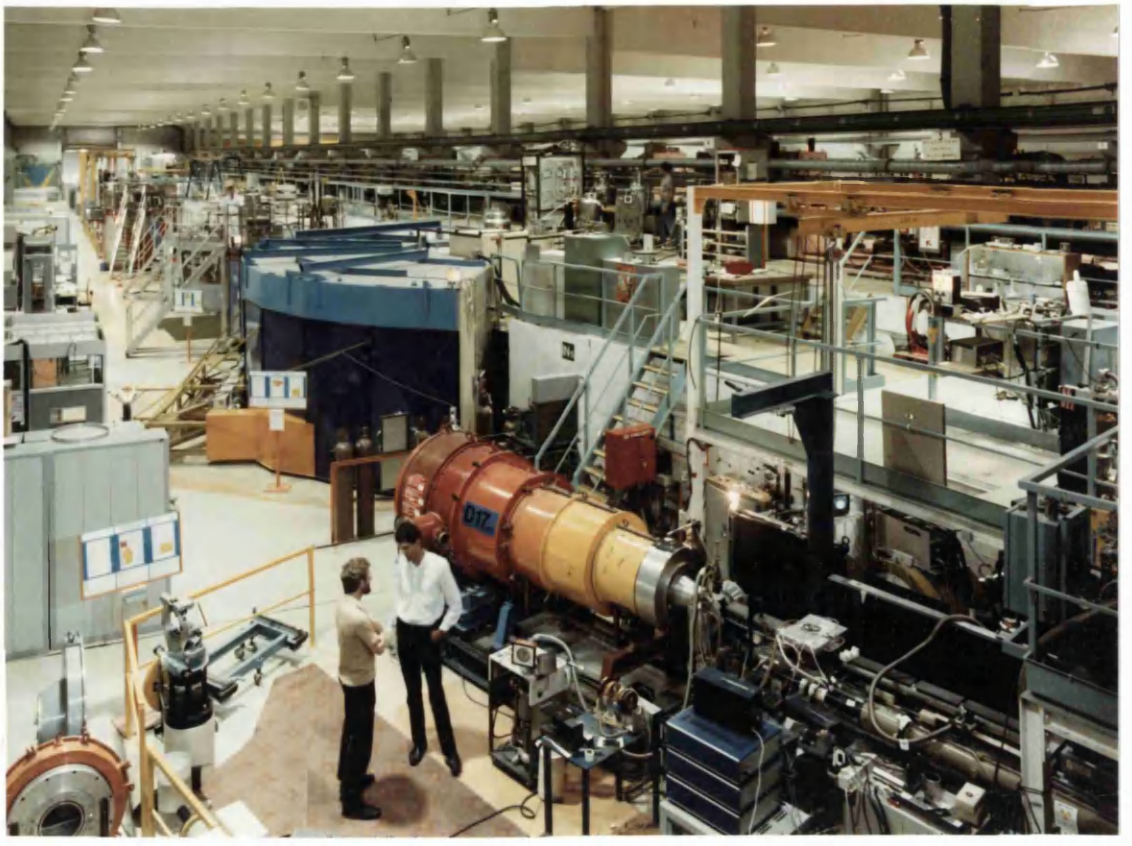
Neutron solution scattering studies were carried out using the high flux reactor at the Institut Laue-Langevin (ILL) in Grenoble, France. High energy neutrons were produced by the fission of U^{235}. The thermal neutron flux was moderated either by the inclusion of a hot (graphite) or a cold (deuterium at 25 K) source to enhance neutron intensities at low wavelengths ($0.04 < \lambda < 0.08$ nm) or at the longer wavelengths used for solution scattering of $\lambda > 0.4$ nm, respectively. Neutron guides transferred the neutrons from the reactor core to the external instruments (Figure 3.1). The small-angle scattering instruments used were D11 on beamline H15 which had a sample-to-detector distance of up to 35 m and a Q range of 5×10^{-2} to 5 nm^{-1} and D17 (Figure 3.2) on beamline H17 with a maximum sample-to-detector distance of 3.5 m and a Q range of 3×10^{-2} to 10 nm^{-1}.

At the instrument, the incident beam was made monochromatic using a velocity selector (Figure 3.2). This consisted of a rotating drum with a helical slot which allowed only those neutrons with a specified velocity and hence wavelength range, to pass through. The beam flux for D17 is shown in Figure 3.3 where the maximum neutron flux occurred around 1.1 nm. The triangular wavelength

---

Figure 3.1. Schematic arrangement of beam-tubes and instruments at the High Flux Reactor, ILL, Grenoble, 1990.

High energy neutrons are produced by the fission of U^{235} and the flux is modified either by the inclusion of a hot source (HS) or a cold source (CS). Neutron guides transfer the neutrons from the reactor core to the external instruments. The small-angle solution scattering stations D11 and D17 are found at the ends of cold neutron guides H15 and H17 respectively.



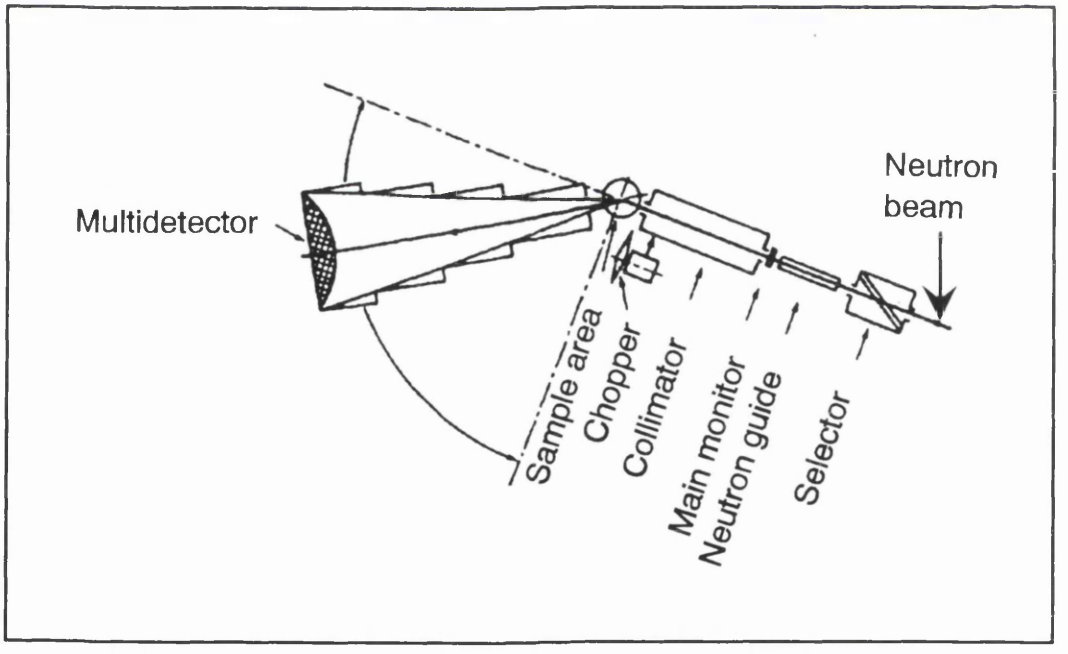


Figure 3.2. Photograph and schematic diagram of D17 set up for small-angle solution scattering, ILL, Grenoble.

The sample-to-detector distance can be set to be 0.8m, 1.4m or 2.8 m to 3.5 m. The detector can be rotated horizontally about the sample from 0^{o} to 90^{o}.

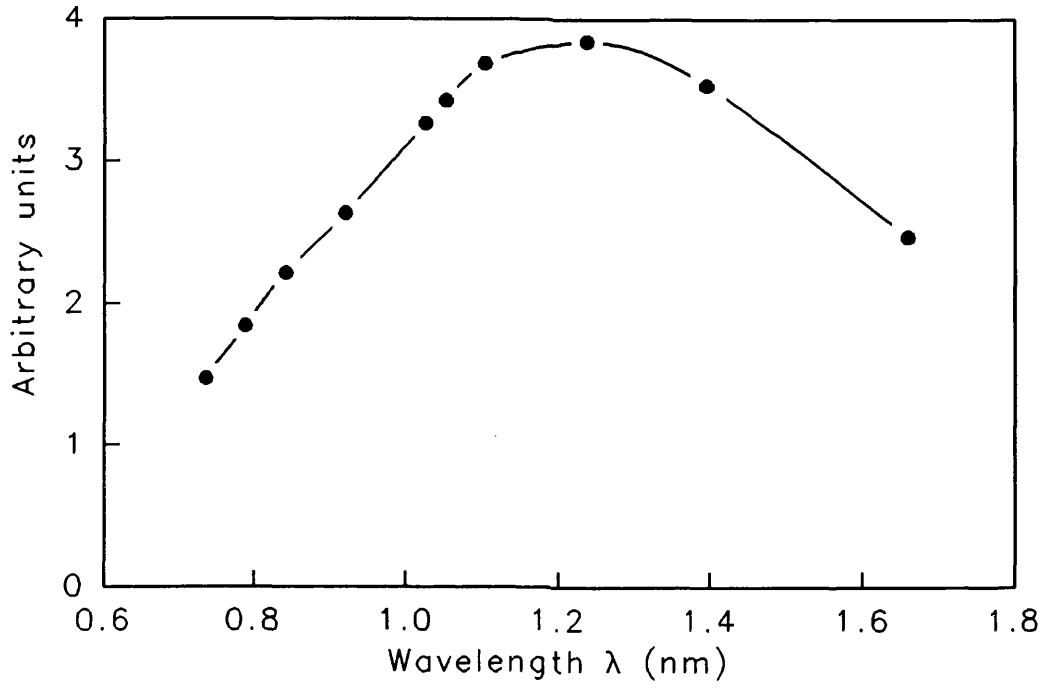


Figure 3.3. Beamflux at the sample on D17, ILL, Grenoble.
The data are taken from 'Guide to neutron research facilities at the ILL', 1988. The maximum flux occurs using around a wavelength of 1.1 nm.

distribution gives a FWHM (full width half maximum) $\Delta\lambda/\lambda$ of 10%. Collimation was achieved by changing the apertures at either end of a 2.5 m boron carbide composite tube (Figure 3.2). Diaphragms of size 20 mm at the reactor end and 14 mm at the detector end of the collimator were used in all experiments. Although ideally the two diaphragms should be of the same size (14 mm) to limit beam divergence, the use of different sizes had the advantage that substantially more neutrons were allowed through at the cost of a slight deterioration in the collimation of the main beam.

Prior to measurement the samples were extensively dialysed against their buffer for at least 36 hours and the final dialysate was used for buffer background subtraction. For neutron scattering studies, proteins were usually measured in buffers containing 0%, 80% and 100% ^2H_2O for matchpoint and Stuhrmann analyses (Chapter 2). Adequate dialysis was especially important in neutron studies because both the sample and buffer must have the same exchangeable

proton content. Samples were measured in quartz Hellma cuvettes of path thickness 1 mm (volume 200 μl) for 0% ^2H_2O samples and 2 mm (volume 400 μl) for 80% and 100% ^2H_2O samples. A shorter path length for H_2O samples was used because of the high incoherent (background) scattering. Up to 12 cuvettes were placed in an automatic and temperature controlled cell holder. This was maintained at 20°C for all measurements. Protein concentrations were ascertained by measuring their optical densities at 280 nm in the same cuvettes and either calculating the absorption coefficient from the amino acid and carbohydrate composition (Perkins, 1986) (Section 3.2.7) or using published values. In accordance with the Debye equation (Chapter 2), the sample of macromolecules must be infinitely dilute to eliminate interparticle interference effects. Interparticle interference effects at low concentrations lead to a decrease in scattering intensity at small angles. High protein concentrations could induce aggregation which was characterized as an increase in intensity at low Q due to the increase in molecular weight. In practice this meant that a concentration series was always performed, and Guinier parameters were extrapolated to zero concentration if a concentration effect was found. The sample counting time was dependent on the sample concentration, the buffer composition and the sample-to-detector distance. This could range from a few minutes at high protein concentration, in 100% ^2H_2O buffers and a sample-to-detector distance of 3.4 m to several hours for all samples in 0% ^2H_2O because of the high degree of incoherent scattering and the decreased intensity of scattering. Samples to be measured for long periods were counted in several separate measurements in case of sudden reactor or instrument failure, and these were subsequently averaged together in the final data reduction stage. The 0% ^2H_2O calibration samples especially were measured throughout the experimental session to ensure reproducibility of the detector response was maintained.

The transmission of the neutron beam through the sample was dependent upon the %^2H_2O present. This varied from approximately 0.45 for samples in 0% ^2H_2O buffers to 0.9 in 100% ^2H_2O buffers. Accurate transmission measurements were necessary for H_2O samples for normalization and calculation of the M_r. The transmission was calculated as the ratio of the intensity of the attenuated direct

beam through the sample to the intensity through the empty cell. The beam was attenuated with a 5 mm polymethylmethacrylate sheet (Rennie, 1988) and the beam stop was moved out to the edge of the detector. All transmission measurements were counted for 2.6 minutes to get sufficient counts for statistically significant results. The transmission from at least one sample from each dilution series was obtained to ensure the dialyses ran correctly. Since transmissions of different samples in the same $\%^2H_2O$ buffer gave approximately the same value, these were averaged together in the final analyses.

Allowable sample-to-detector distances for D17 ranged from 0.8 m to 3.5 m and the detector could be rotated from 0^o to 90^o about the sample (Figure 3.2). Typical configurations of 3.4 m 0^o, 1.4 m 0^o and 1.4 m 20^o gave a total Q range of approximately 0.05 to 3.25 nm^{-1}. The highest nominal structural resolution of the scattering curve was around 2 to 4 nm, according to the equation $2\pi/Q_{max}$.

The BF_3 multi-detector on D17 has dimensions of 64 x 64 cm^2 and is composed of 4,096 cells of area 1 x 1 cm^2. Both the incident and diffracted beam were *in vacuo*. The direct beam was prevented from hitting the detector by a 8 cm x 6 cm cadmium beam stop. The position of the beam stop was determined using a 1 mm thick strip of Teflon inserted in the sample holder, which was a very high scatterer of neutrons. CAMAC standard electronics and a VAX 11-730 mini-computer controlled the instrument and data acquisition.

(3.1.1.2) Primary data reduction

A number of calibration and normalization measurements were performed during the processing of any sample (Ghosh, 1989) (Figure 3.4). DETEC listed the raw counts from the detector cell by cell and was used to calculate the position of the beam stop. RNILS listed and stored the radial distribution function $I(Q)$ of the detector. Individual cells were averaged from a given radial step length of 1 cm to give a mean Q and intensity, $I(Q)$. The error on $I(Q)$ is calculated as:

error $I_{er}(Q)$ = $\sqrt{}$(total counts)/(number of cells)

SPOLLY normalised and combined individual spectra and listed and stored the

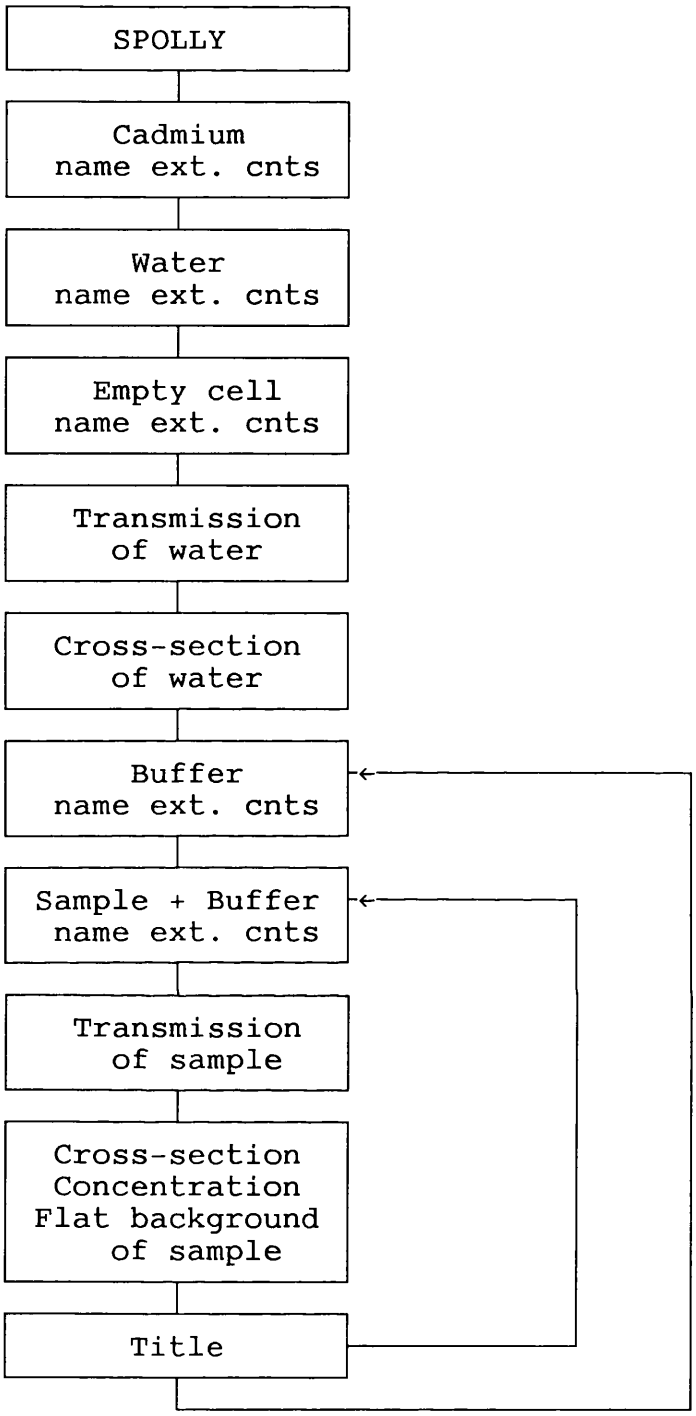


Figure 3.5. Flowchart showing in detail the dialogue for SPOLLY.

Name, ext. and cnts. refer to the individual run number, its extension number and the monitor counts (proportional to the time in the beam) respectively. The cross-section of water and the sample is always set as 1. The transmission of the sample plus buffer and buffer alone are assumed to be equal and the default value of 1 is always used. The sample flat-background default is always taken, which is 0.

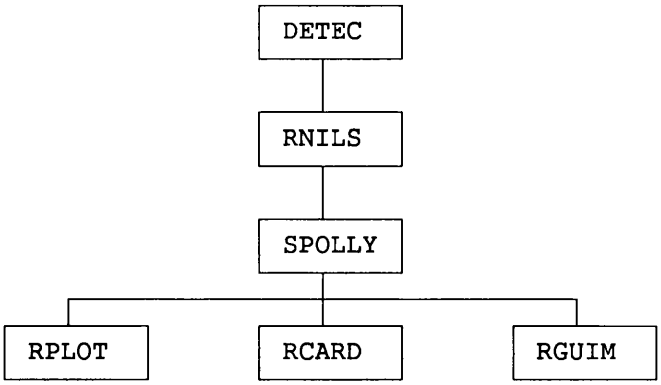
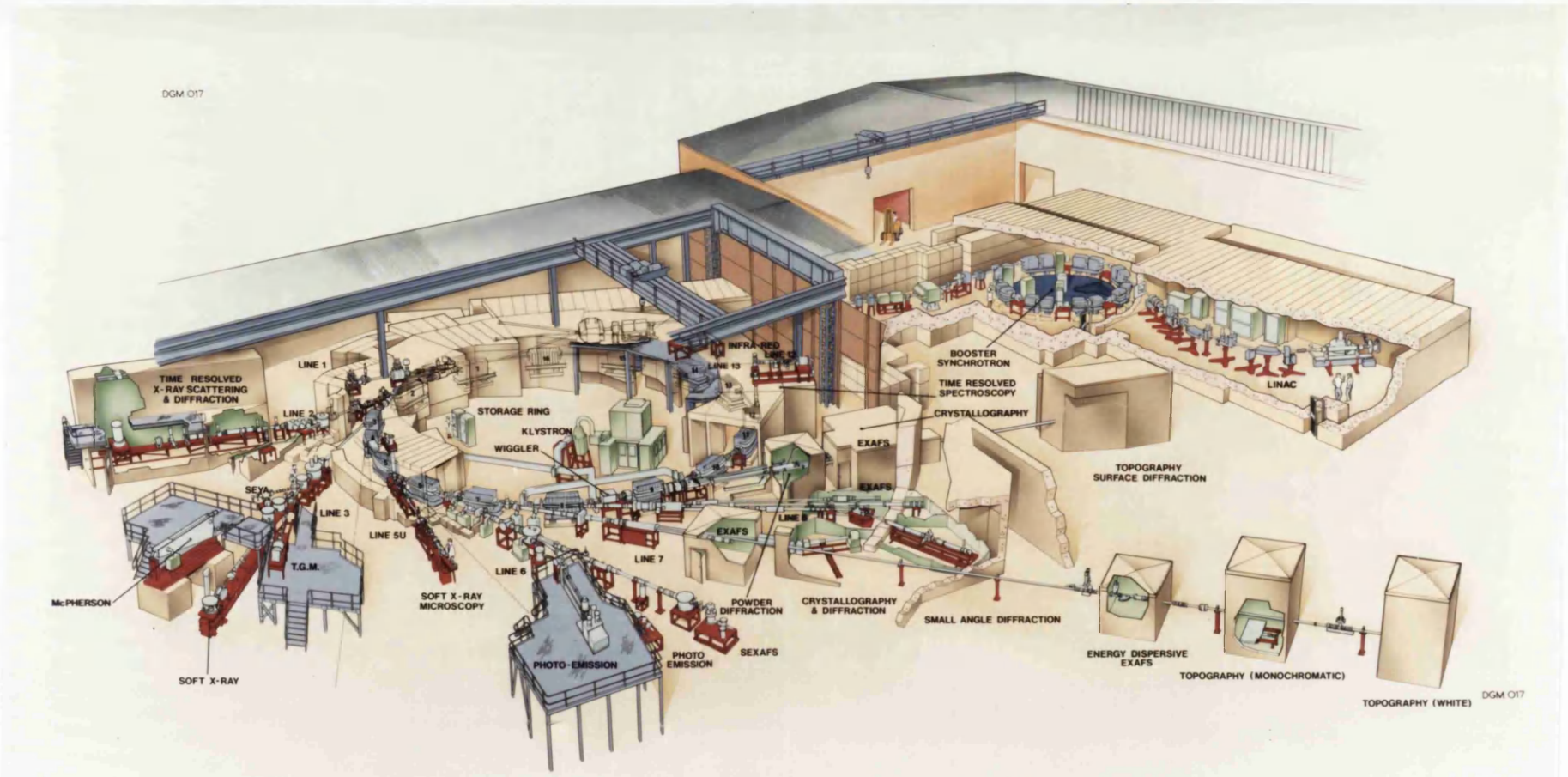


Figure 3.4. Flowchart showing the data reduction procedures for neutron scattering.

DETEC lists the raw counts from the detector cell by cell, RNILS lists and stores the radial distribution function $I(Q)$ of the detector, SPOLLY corrects and normalizes the sample spectrum $I(Q)$, RPLOT plots out 2 spectra on the same axes, RCARD produces a card image disk file of formatted data and RGUIM is used to calculate the Guinier and cross-sectional parameters. See text for details.

output (Figure 3.5). All spectra were normalised with respect to the monitor counts (counting time). The role of the cadmium sample was to block the direct beam of neutrons and hence to provide an estimation of the neutron and electronic background counts in the Guide Hall which was subtracted from each spectrum. Pure H_2O and 0% ^2H_2O buffers (assuming these were very dilute) were used as a standard to produce an absolute scale to enable molecular weight determinations. These were corrected for the transmission of H_2O and the scattering due to the empty cell. Each sample spectrum was normalised by subtracting the buffer background and dividing by the corrected H_2O sample. Preliminary R_G and $I(0)$ values were calculated using RGUIM to check the samples were not aggregated and that the samples had been counted for a sufficient amount of time. RPLOT was a general graphics program and could plot two spectra together. This was used to ensure spectra of the same sample from different instrument configurations overlap in regions of similar Q. RCARD produced output containing Q, $I(Q)$ and error on $I(Q)$ on magnetic tape for transfer to London. The spectra were reanalysed more carefully in London using the FORTRAN interactive graphics program, SCTPL.



THE SYNCHROTRON RADIATION FACILITIES AT DARESBURY LABORATORY

DGM 017

(3.1.2) Small-angle solution scattering by X-rays

(3.1.2.1) Description of the instrument

X-ray studies were carried out at the Synchrotron Radiation Source (SRS) at Daresbury, Warrington, U.K. Electrons were emitted from a hot cathode and accelerated to 12 million electron volts (12 MeV) in a linear accelerator (LINAC). The electrons were further accelerated in a booster synchrotron to 600 MeV and from there, injected into the storage ring and finally accelerated to 2,000 MeV (2 GeV, 99.999997% of the speed of light) by a high power radio-frequency accelerating system. This also maintained the energy at this level. The average starting current was 205 mA with an average overall current of 169 mA (Annual Report, 1990/91). The beam was refilled generally every 24 hours. The main beam intensity continually decreased between beam-fills. The electrons travelled round the 96 m circumference ring 3.12 million times a second and remained in orbit for approximately 30 hours. The path followed by the electrons was bent into a circle by 16 dipole magnets. Synchrotron radiation was emitted by the electrons when they were deflected by the magnetic field. The radiation was transmitted tangentially down beamlines to different experimental stations (Figure 3.6). The small angle X-ray scattering instruments were situated at stations 2.1, 7.3 (now discontinued) and 8.2.

Figure 3.6. Schematic representation of the beamlines at the SRS, Daresbury. Electrons are emitted from a hot cathode and accelerated in a linear accelerator (LINAC). The electrons are further accelerated in a booster synchrotron and then injected into the storage ring. The path followed by the electrons in the storage ring is bent into a circle by 16 dipole magnets and has a 96 m circumference. Synchrotron radiation is emitted by the electrons when they are deflected by the magnetic field and is transmitted tangentially down beamlines to the different experimental stations. The small-angle scattering instruments are located on beamlines 2, 7 and 8.

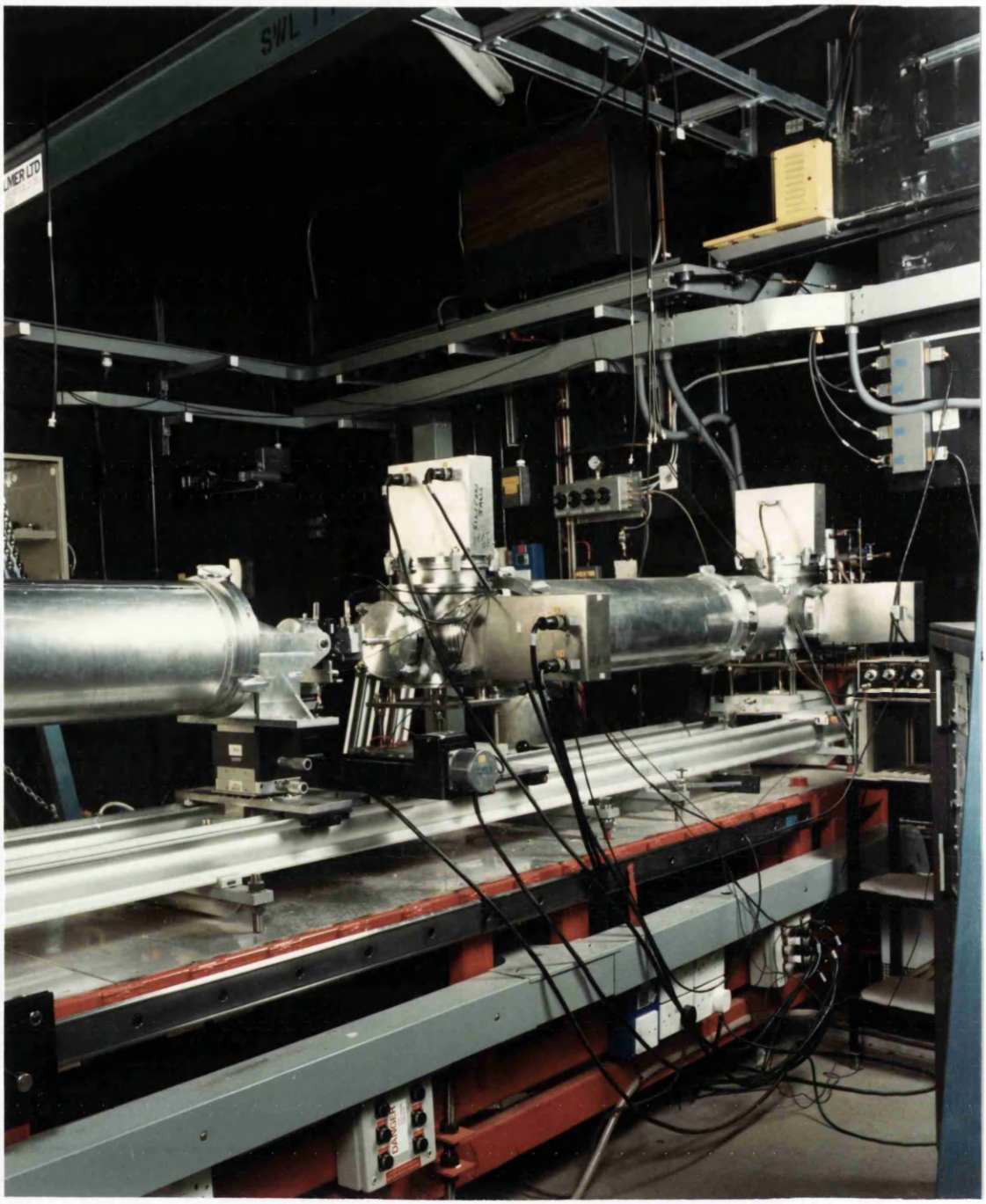


Figure 3.7. Photograph of the sample area of Station 8.2, SRS, Daresbury.
The X-ray beam from the storage ring is monochromated and collimated and passes through a front ion chamber before reaching the sample. The scattered beam goes through a back ion chamber onto the detector. The air gap between the sample and the front and back ion chambers should be kept as small as possible. The sample-to-detector distance can be varied from 0.3 to 4.5 m.

The beam was monochromated to a wavelength of 0.154 nm and focused horizontaly by a Ge(111) crystal and focused vertically with a curved mirror. Sets of slits positioned before and after the monochromator and mirror were used to collimate the beam and reduce parasitic scatter (Towns-Andrews *et al.*, 1988). Wavelength spread is negligible as the result of monochromatization. The beam passed through a front ion chamber, used for qualitative monitoring, through the sample and then through a back ion chamber, used for normalization and equivalent to transmission measurements by neutron scattering (Figure 3.7).

As for neutron experiments, the samples were dialysed against their buffer for at least 36 hours and the final dialysate was used for buffer subtraction procedures. The buffers used were based on a low concentration of salts in H_2O which kept the transmission high. Samples were placed in the beam in round cells with mica window thicknesses of approximately 10 μm, held in place with Teflon inserts. The sample volume was 20 μl and the path length was 1 mm. The cell holder was either made of brass and temperature controlled via a waterbath (maintained at 20°C for all measurements) or Perspex where measurements were made at room temperature ($\approx$20°C). The sample holder was aligned in the beam by using 'green paper' which turns red on exposure to X-rays. Protein concentrations were determined by the optical density at 280 nm, measured in 1 mm path-length quartz Hellma cells and a concentration series was always measured as for neutron scattering. The counting times were dependent upon the sample concentration and were limited in some cases by the susceptibility to radiation damage which caused the sample to aggregate (Figure 3.8). The counting time was split into 10 equal timeframes which were plotted individually to assess if any radiation damage had occurred and then were averaged together. The protein samples were measured alternately with their buffer using equal collection times to reduce background subtraction errors.

The sample-to-detector distances on station 8.2 range from 0.5 to 4.5 m. Sample-to-detector distances from 2.2 to 4.2 m were used here to give a Q range of 0.1 to 1.8 nm^{-1} (linear detector) (0.1 to 3.8 nm^{-1} quadrant detector) or 0.1 to 1.0 nm^{-1} (linear detector) (0.1 to 1.8 nm^{-1} quadrant detector), respectively.

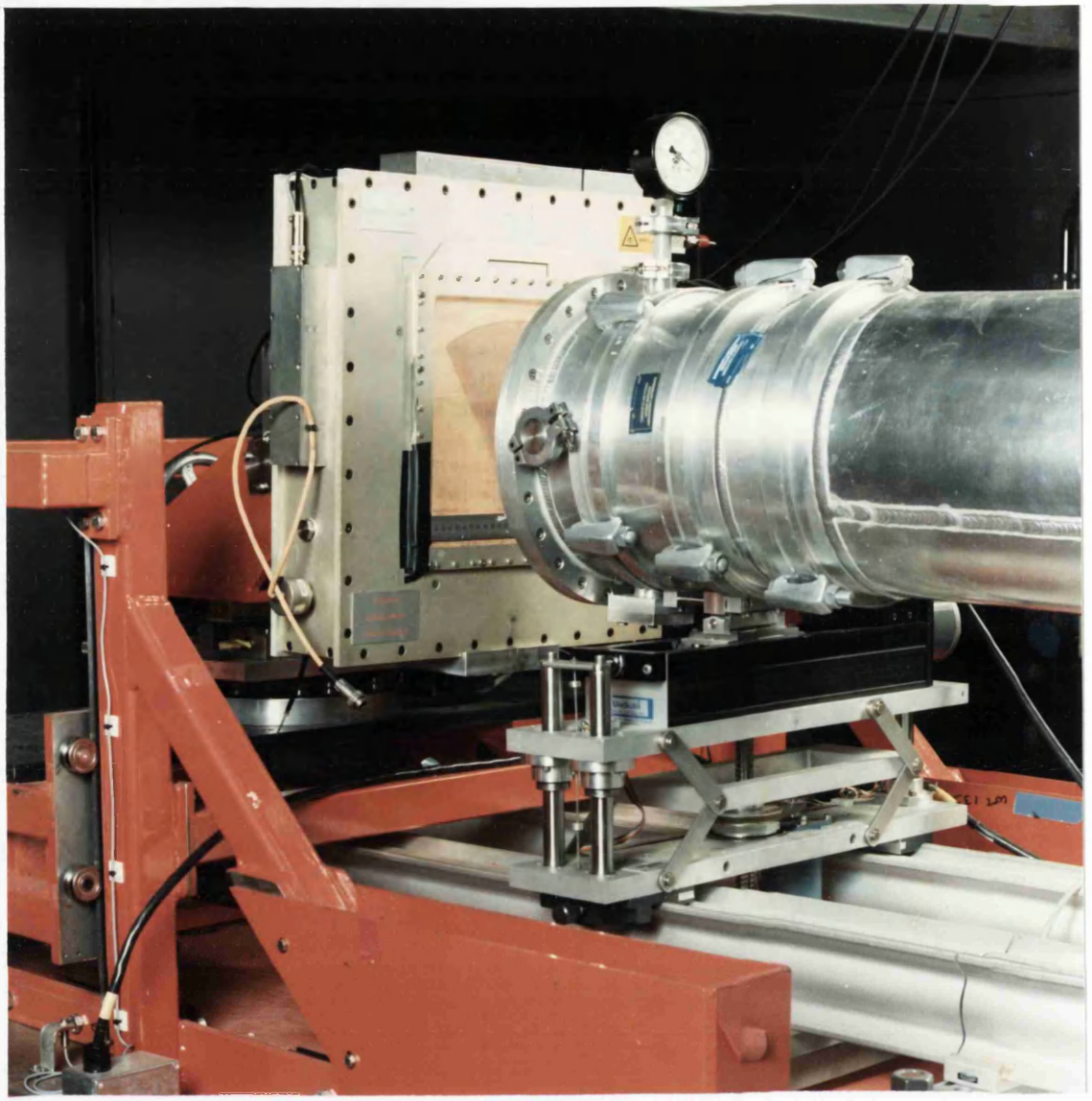


Figure 3.9. Photograph of the quadrant detector at Station 8.2, SRS, Daresbury. The beam is guided to the detector along an evacuated beamline. The quadrant detector measures the scattering intensity in a 76^{o} arc over 150 mm. The beam is aligned with the centre of the circle.

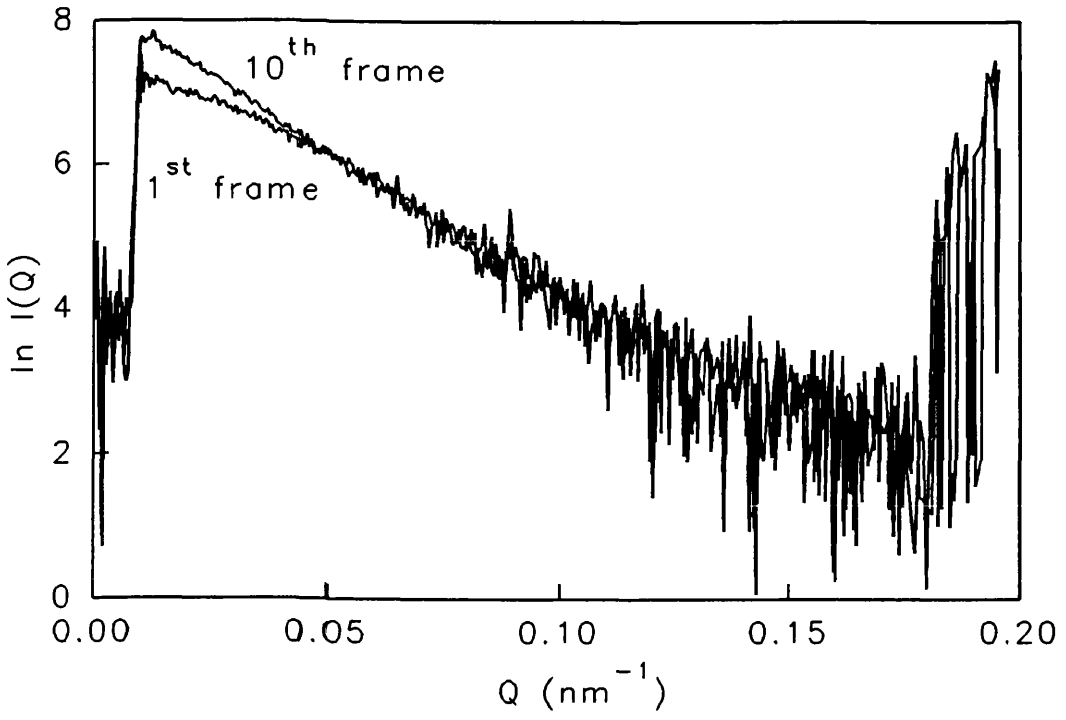


Figure 3.8. Aggregation of complement component C9.

The scattering curves from the first and tenth minute of exposure of C9 to X-rays are very different at low *Q*. This is attributed to protein aggregation as a result of radiation damage.

Two detector types were used, a linear and a quadrant detector. The linear detector measured scattering intensity in one dimension whereas the quadrant detector measured the intensity in a 76^{o} arc over a radial length of 150 mm (Towns-Andrews et al., 1988) (Figure 3.9). The main beam was aligned with the centre of this arc and the increase in circumference with increase of scattering angle compensated for the decrease in scattering intensity at higher angles, hence achieving better counting statistics at high *Q*. To characterise the detector response, a constant radioactive source of ^{55}Fe was measured over several hours.

(3.1.2.2) Preliminary data reduction

Data reduction at Daresbury utilized the FORTRAN software OTOKO (Bendell, P., Bordas, J., Koch, M.H.C. & Mant, G.R., EMBL Hamburg and SERC Daresbury Laboratory, unpublished software) running on a VAX 11-750 (Figure

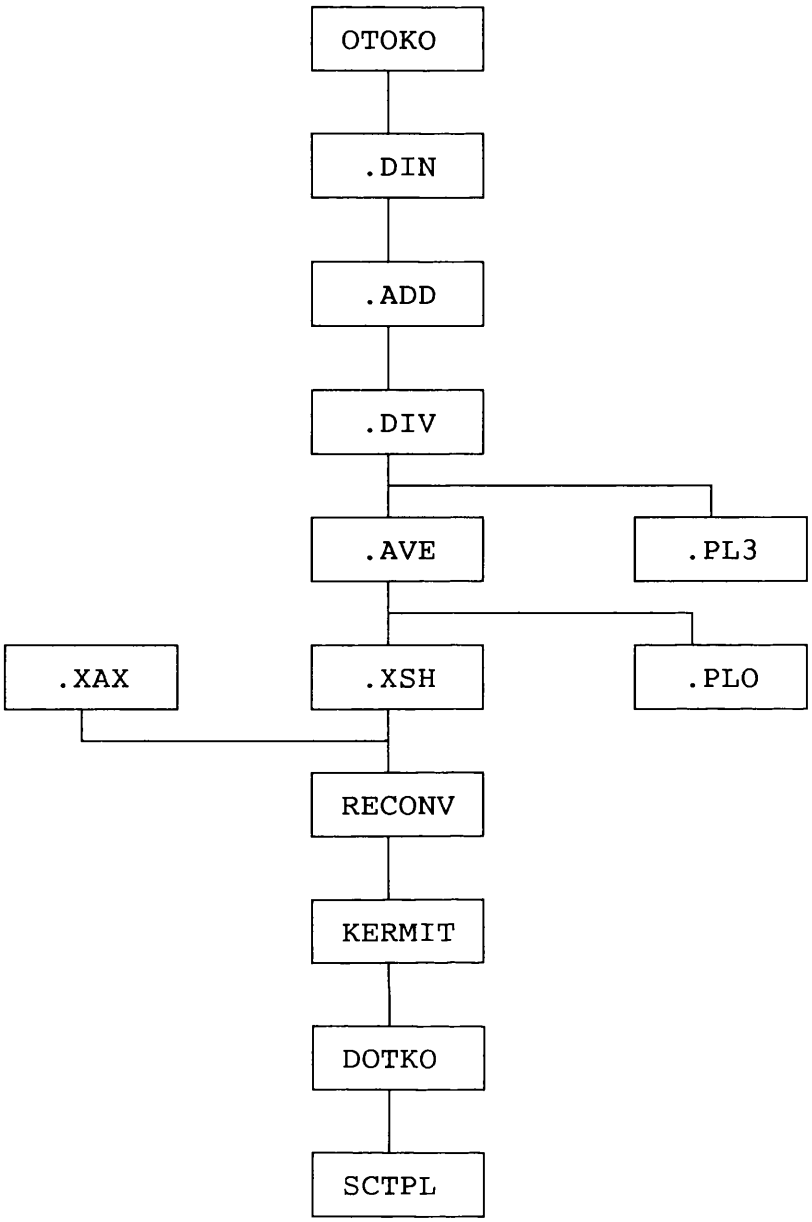


Figure 3.10. Flowchart of the data reduction procedures for X-ray scattering. The .DIN procedure normalises all the spectra to the counts measured in the back ion chamber. The buffer background is subtracted from the sample plus buffer spectra using .ADD. The resultant spectra are normalised to the detector response using .DIV and the 10 individual time-frames are averaged together with .AVE. The spectra are plotted either with .PLO (averaged spectrum) or .PL3 (individual time-frames). The detector gap is removed using .XSH (Figure 3.11). The *Q* axis is calculated using .XAX from the diffraction pattern of wet, slightly stretched rat tail collagen. RECONV converts the binary OTOKO files to card image files for transfer via KERMIT to London. DOTKO combines the *Q* axis file with the spectrum intensity files for final analyses using SCTPL.

3.10). The .DIN procedure normalised all the spectra to the counts measured in the back ion chamber. This corrected for beam flux, transmission of the sample and exposure times. The buffer background was subtracted from the sample spectra using .ADD to give the scattering curve due to the sample only. The resultant spectra were normalised using the detector response using .DIV and finally the 10 individual time-frames were averaged together with .AVE. The spectra could be plotted either with .PLO (plots out the averaged spectrum) or .PL3 (plots out the individual time-frames). The x-axis contained an artificially produced gap due to the electronics (Figure 3.11). This was removed using the program .XSH by specifying the position of beginning of the gap and the size of the gap in pixels. The Q axis was calibrated using .XAX from the diffraction pattern of wet, slightly stretched rat tail collagen (Figure 3.11). The spectrum consisted of a series of peaks of diffraction spacing 67 nm and the major peaks

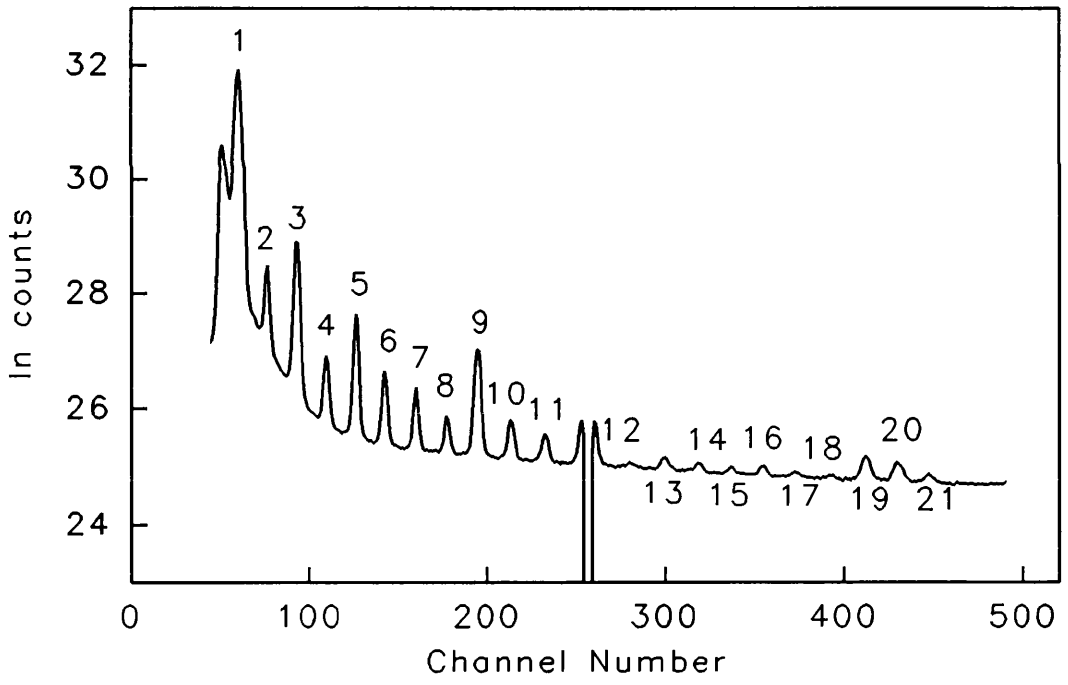


Figure 3.11. Collagen diffraction pattern.

A total of 21 diffraction peaks are visible here, each of which is $2\pi/67$ nm^{-1} apart. The channel numbers are converted into Q values using .XAX. The gap between peaks 11 and 12 is artificial and is removed during data reduction by determining the position of beginning of the gap and the size of the gap in pixels.

were the 1^{st}, 3^{rd}, 5^{th} and 9^{th} peak. The position in pixels of all the peaks was determined. Since the number of pixels between any two peaks was approximately the same, the position of the 0^{th} order peak ($Q = 0$) could be extrapolated. The pixel position and Q value ($Q = 2\pi$ x order / 67) of another peak, usually the 9^{th} order, was calculated. The pixel position and Q value of these two peaks were used by .XAX to produce a Q axis file. RECONV converted the binary OTOKO files to card image files for transfer via KERMIT to London. DOTKO combined the Q axis file with the spectrum intensity files for final analyses using SCTPL.

(3.2) COMPUTATIONAL ANALYSES

(3.2.1) Analyses of scattering data

(3.2.1.1) Guinier and cross-sectional plots

Plots of ln $I(Q)$ against Q^2 (called Guinier plots) (Chapter 2) were performed on neutron data measured at 3.4 m on D17 and 10 m on D11 and all X-ray data using the FORTRAN program SCTPL. The Q range selected for line-fitting went to as low a Q as possible and out to a $Q.R_G$ of approximately 1.5. Curves which were slightly aggregated increased at very low Q and these points were not included in the Guinier analyses. The program calculated the R_G from the gradient of the line and the $I(0)$ from the y-axis intercept (Chapter 2). The Q and $Q.R_G$ range for the fitted points was also given. Cross-sectional plots of ln $[I(Q).Q]$ against Q^2 were performed on neutron data measured at 1.4 m 0^o D17 or 2 m on D11 and all X-ray data again using SCTPL. The Q range was constrained out to a $Q.R_{XS}$ of 1.5 and the lowest Q used in line-fitting was slightly larger than the largest Q used in the Guinier plots. Aggregates were not a problem over this Q range. The R_{XS} was calculated from the gradient and the $[I(Q).Q]_{Q\to 0}$ from the y-axis intercept. The Q and $Q.R_{XS}$ range were also given.

The $I(0)$ and $[I(Q).Q]_{Q\to 0}$ values were dependent on protein concentration and therefore were normalized to concentration. The R_G, $I(0)/c$, R_{XS} and $[I(Q).Q]_{Q\to 0}/c$ parameters were checked for concentration dependences by plotting each against protein concentration using the software SIGMAPLOT. If a concentration dependence was observed, the line was regressed to zero concentration using MINITAB and these values used in further analyses. The length and triaxial dimensions of macromolecules were determined from the Guinier and cross-sectional parameters (Chapter 2). The M_r was calculated (Chapter 2) from the neutron $I(0)/c$ values measured in H_2O only using the spreadsheet, LOTUS 1-2-3.

(3.2.1.2) Matchpoint and Stuhrmann analyses

The experimental matchpoint was determined from a plot of $\sqrt{[I(0)/cT_s t]}$ against % ^2H_2O. The plot should be linear and the errors small. The protein matchpoint corresponded to the % ^2H_2O when the abscissa was zero and was calculated by linear regression analyses using MINITAB. A theoretical matchpoint was calculated for comparison from the amino acid and carbohydrate composition, using the software SLUV (Perkins, 1986). The Stuhrmann parameters were determined from plots of R_G^2 or R_{XS}^2 against $1/\Delta\rho$. $1/\Delta\rho$ corresponds to the difference in scattering density between the solute and solvent. The solute scattering density was calculated from the matchpoint and the solvent scattering density is dependent on the % ^2H_2O composition. The R_G or R_{XS} at infinite contrast, $R_{G\text{-}C}$ or $R_{XS\text{-}C}$ were obtained from the square root of the ordinate when $1/\Delta\rho$ was zero. The gradient of the plots gave α_G or α_{XS}. These values were determined from linear regression analyses using MINITAB.

(3.2.1.3) Distance distribution function

The Indirect Transformation Procedure was performed on neutron, X-ray and simulated curves out to a Q of at least 1.6 nm^{-1}. In the case of the neutron data, this meant joining together the scattering curves from the 3 instrumental set-ups (3.4 m 0^o, 1.4m 0^o and 1.4m 20^o), ensuring that the difference in Q values (ΔQ) at the intercept of each domain corresponded to the ΔQ of either one of the domains. Each domain must have an even number of data points (bearing in mind that the last point of one domain was also the first point of the next domain). The experimental X-ray and simulated curves consisted of a single domain only, of constant ΔQ. The lowest Q value used was the same as that used in the Guinier analyses, i.e. points considered to be indicative of aggregation were not included in the ITP analyses. A flowchart of the typical input statements for ITP is shown in Figure 3.12. The maximum particle dimension, D was estimated from the macromolecular length determined by Guinier analyses (Chapter 2) and D_{max} was calculated as 1.5 * D. The number of splines used was a trial-and-error procedure since the recommended number of 15 to 30 splines (Müller & Glatter, 1982) was

Title:	C9x001
Data on disk: Disdat (I) Graz (D), Neither (N):	N
Normal correlation function (N) Thickness correlation function (L) Cross-section (C) Size distribution (P):	N
Complete listing (Y or N):	N
P(r) output (Y or N):	N
Simulation with theoretical curves:	N
Line printer (default 1):	1
Wavelength spread:	N
Slit length:	N
Slit width:	N
Number of subdomains:	3
Transformation factor (default 1):	1
Constant background (A1):	0
Linear function (A2 * Q):	0
Number of intervals in subdomain 1:	12
Number of intervals in subdomain 2:	16
Number of intervals in last subdomain:	54
Number of continuation steps:	0
Known spline coefficients (Y or N):	N
Number of points neglected at innermost Q:	0
First Q value in subdomain 1:	0.02138
Last Q value in subdomain 1:	0.04109
Last Q value in subdomain 2:	0.1054
Last Q value in last subdomain:	0.3252
ΔQ of the last subdomain:	0.003983
Integration constant (default 1):	1
Intensity from disk (Y or N):	Y
Filename:	G10715P
Data format Basic(0), Hamburg(1), Grenoble(2) Daresbury(3), Argonne(4), Simulated(5):	2
Number of lines:	42
Number of splines:	10
Number of points in real space:	101
Increment in real space:	1.0
Identical abscissa as previous (Y or N):	N
Start Lagrange parameter:	10
Further number of multipliers:	19
Decrement:	1
Filename for card output:	ITP
Write file on teletype (Y or N):	N
Write file on line printer (Y or N):	N

usually too high and overinterpreted the experimental data. Around 10 splines was usually found to be satisfactory. The correct D_{max} was indicated by stable R_G and $I(0)$ values over a suitable range of Lagrange multipliers whereas the wrong choice of D_{max} would not give a stable range (Müller & Glatter, 1982). Criteria for an acceptable calculation of $P(r)$ for glycoproteins were: (i) $P(r)$ should exhibit positive values; (ii) the R_G from ITP should agree with the R_G from Guinier analyses; (iii) $P(r)$ is zero when r is zero; (iv) $P(r)$ is stable and reproducible for different experimental $I(Q)$ curves when the number of splines and D_{max} is varied over a reasonable range. The maximum particle dimension was determined from $P(r)$ when this became zero at large r (Chapter 2) although errors in D can be significant as a result of the low intensity of $P(r)$ in this region.

(3.2.2) Small Debye sphere modelling

Molecular modelling of the neutron and X-ray scattering curves used small spheres based on the Debye equation (Perkins & Weiss, 1983; Perkins, 1985) (Chapter 2). If the crystal structure of the protein was available, the crystal

---

Figure 3.12. Flowchart of the Indirect Transform Procedure.
The typical procedure and input parameters for an experimental scattering curve from Grenoble is shown. The three sub-curves are added together to give a full scattering curve out to a Q of 3.3 nm^{-1}. Although background subtraction errors can be corrected by applying a constant to the experimental curve (values A1 and A2), in practice, if this is deemed necessary, the data is rejected. The number of intervals in a sub-domain is one less the number of points and must be an even number. Points at the innermost part of the curve can be rejected, this is important if the data is suspected of being aggregated. The program extrapolates the experimental curve back to zero Q using the Q values requested. The number of splines used is requested and the number of points in real space multiplied by the increment in real space gives D_{max} (see text for details). The range of Lagrange multipliers printed can be altered but typically, for a good curve, a parameter around 0 gives stable R_G and $I(0)$ values.

co-ordinates were subdivided into small cubes using a program COTOS ("CO-ordinates TO Spheres"). A 3-dimensional grid was defined and each atom was assigned to a particular cube. A cutoff was chosen such that only cubes containing a certain number of atoms were used. The total volume of the cubes should equal the dry volume of the crystal structure (Chothia, 1975). Cubes of side 0.608 nm were used to facilitate the representation of a carbohydrate residue as a single cube. If no atomic resolution structure existed, models were made based on simple geometric bodies using the triaxial dimensions calculated from the Guinier parameters or on electron micrographs, using the program BY444. The dry protein volume was simply divided into a convenient number of small cubes generally of side 0.8 nm. The scattering curve simulations were based on overlapping spheres of the same volume as the cube, positioned at the centre of each cube. The sphere diameter should be much less than the nominal resolution of $2\pi/Q_{max}$ of the scattering curves, i.e. 4.3 nm for Q_{max} = 1.45 nm^{-1}.

If there was a significant dependence of the R_G on contrast (i.e. a large Stuhrmann α value) this was incorporated into the model by subdividing the spheres into an outer hydrophilic surface region and an inner core on the basis of the volumes of these two components. Each component was assigned a different scattering density calculated from the matchpoint. The volumes and matchpoints were determined for the hydrophilic amino acids plus carbohydrates and the hydrophobic amino acids of the glycoprotein using SLUV output.

The 'dry' model was compared with the neutron experimental data (Chapter 2). This was converted for comparison with X-ray data by adding an hydration shell performed by rescaling the co-ordinates by a factor derived from the cube root of the ratio of the hydrated and dry volumes. The hydrated volume was calculated assuming 0.3 g water/g glycoprotein and an electrostricted water volume of 0.0245 nm^3/water molecule (Chapter 2). For a two density model the outermost spheres now corresponded to the hydration shell as well as the carbohydrate and hydrophilic amino acid residues. Electron densities instead of scattering lengths per volume were calculated from SLUV for each component.

If a single scattering density was used, the R_G and full Debye scattering curve were calculated using the programs APS and SCT respectively. If two scattering densities were used, the R_{G-C} and Stuhrmann α values were calculated using APSNEW and the scattering curve was calculated using SCTNWW. Prior to comparison with the experimental neutron scattering curves, the calculated curve was smeared to account for a wavelength spread of 10% on D17 and the beam divergence in radians (Cusack, 1981; Perkins & Weiss, 1983) based on the empirical parameters of 16% and 0.016 radians (Cusack, 1981). Since the X-ray cameras utilise point collimation, the calculated and experimental $I(Q)$ curves were directly comparable without corrections. To facilitate quantitative comparisons between the experimental and model scattering curves, a residual R was employed by analogy with protein crystallography (Blundell & Johnson, 1976):

$$R = \frac{\sum | I(Q)_{exp} - I(Q)_{cal} |}{\sum | I(Q)_{exp} |} \qquad (1)$$

This was generally less than 0.02 for a good curve fit between $0 < Q < 1.6$ nm^{-1}.

(3.2.3) Hydrodynamic modelling

The models were made in similar way to the Debye sphere models, using either COTOS if a crystal structure was available or BY444. The hydrodynamic model differed from the Debye sphere model by having fewer, non-overlapping spheres and a larger volume to include the hydration shell surrounding the protein. Fewer spheres were used due to constraints on computing time. Non-overlapping spheres were used because $\mathbf{T}_{ij}$ is only valid when $r_i + r_j \leq R_{ij}$ (Chapter 2). Since a relatively low number of spheres were used, allowance for the void spaces between the spheres was necessary. Accordingly a hypothetical hydration of 0.39 g H_2O/g glycoprotein in place of 0.3 g H_2O was used to calculate the total hydrated volume. This increased the glycoprotein volume in order to compensate for these voids (Perkins, 1989b). The frictional coefficient was determined from the program FRCC (Garcia de la Torre & Bloomfield, 1977a, b). From this, the sedimentation coefficient was determined and compared with values given in the literature.

(3.2.4) Sequence alignments and secondary structure predictions

Amino acid sequences were aligned both manually and with the guidelines provided with the automatic sequence alignment program, MULTAL (Taylor, 1990). Sequences were downloaded from the databases at Daresbury; and the OWL software was used to find a particular protein sequence from different species. The MULTAL software was used to produce an initial alignment which was then finalized manually. MULTAL aligns the most homologous sequences into groups, all relative to the first sequence which may be a consensus sequence. The sequences then were positioned manually to maximize the alignment of conserved or chemically homologous residues and to minimize gaps. Chemically homologous residues were grouped as:

Tiny: G, A, S

Aliphatic: I, L, V

Aromatic: F, H, Y, W

Positive: R, H, K

Negative: D, E

Exon structures where available, were used to define the beginning and end of the protein domain.

Averaged secondary structure predictions were based on the sequence alignment. The Robson or Chou-Fasman prediction methods were performed on each sequence individually using the programs ROBSCDC and CHOUCDC respectively. For the Robson prediction, if the content of α-helix or β-sheet of the protein was known from other sources, a relevant decision constant was used (Chapter 2), otherwise an unbiased prediction was performed. The Robson algorithm forced a choice between one of four conformational states (α-helix, β-sheet, β-turn and coil), so the sum of all four states was 100%. The Chou-Fasman prediction was unrestrained in this respect, and any propensity for α-helix, β-sheet or β-turn in the sequences was indicated. Residue positions in the alignment which were occupied in over 50% of the sequences were averaged together using the programs PREDRB and PREDCF for the Robson and Chou-Fasman predictions respectively (Perkins *et al.*, 1988). The average total score in centinats and

percentage propensity for each position in the alignment to be in a particular conformation was given as graphical output by PREDRB whereas just the percentage propensities were given in PREDCF. The consensus sequence length was defined as the number of residues occupied in over 50% of the sequences. The mean sequence length was calculated from the sum of the number of residues in each sequence in the alignment divided by the number of sequences. The mean dry volume was similarly determined, by calculating the dry volume for each sequence using SLUV (Chothia, 1975; Perkins, 1986) and dividing by the number of sequences.

(3.2.5) Hydropathy profiles

The averaged hydropathy was calculated using the Kyte & Doolittle (1982), Eisenberg *et al.* (1984), Hopp & Woods (1981) and Parker *et al.* (1986) hydrophilicity or hydrophobicity scales. In contrast to the general use of windows of several residues in length to specify hydropathy profiles, a window size of 1 was found necessary to give an undistorted averaged profile at each residue position in the sequences. The most frequently occurring maxima in the four hydropathy profiles were assigned as hydrophilic or hydrophobic.

(3.2.6) Computer hardware

The principal computers used in London were the Cyber 960 at the Imperial College Computer Centre (ICCC) and the Silicon Graphics 4D35S Datastation at the Royal Free Hospital School of Medicine (RFHSM). All the base small-angle scattering analyses and secondary structure predictions were done on these machines. An Opus PC V and an ELONEX PC at the RFHSM was used for wordprocessing, graphical representation of the data and statistical analyses and a Silicon Graphics Personal Iris 4D25TG Workstation, also at the RFHSM was used for crystal structure and Debye sphere model inspection and diagrams.

(3.2.7) Computer software

The software used was either commercially available or base written in FORTRAN. The commercial programs on the OPUS were WordPerfect (WordPerfect Corporation), Lotus 1-2-3 (Lotus Development, U.K.), Sigmaplot (Jandel Scientific) and Minitab (Ryan *et al.*, 1985) and on the ELONEX, Charisma (Micrografx); and on Silicon Graphics, Insight II version 2 (Biosym Technologies, San Diego, CA, U.S.A.) and Showcase (Silicon Graphics). Many programs were written on the Cyber and the datastation by Dr S.J. Perkins and Mr A.S. Nealis for data analyses. Brief descriptions of the major programs are given below:

General

SLUV (Scattering Length per Unit Volume). Produces a table of protein characteristics calculated from the amino acid and carbohydrate composition (Perkins, 1986). These include:

Amino acid and carbohydrate volumes. These were taken from protein (Chothia, 1975) and carbohydrate (Perkins, 1986) crystal structures.

Absorption coefficient x 1.03. This is based on the Trp, Tyr and Cys content in 23 proteins and calibrated over a range of experimentally calculated A_{280} of between 5 and 27 (Perkins, 1986).

Matchpoint. This first requires the calculation of the neutron scattering length density from the summation of scattering lengths divided by the volume of the macromolecule, $\Sigma b/V$. Since the matchpoint is very sensitive to the amount of hydrogen exchange of the protein with the solvent (Chapter 2), a value of 10% nonexchange of the main-chain peptide NH protons was assumed.

Partial specific volume ($\bar{v}$). This is derived from the consensus volumes data set of Perkins (1986) and is based on the sum of the dry volume and the electrostricted water shell volume as calculated from the amino acid and carbohydrate composition.

ITP. Produces the Fourier transform of either experimental or model data by the Indirect Transform Procedure (Glatter, 1982).

Plotting

SCTPL. An interactive plotting program for graphical representation of experimental scattering curves with error bars (neutron data only), calculation of R_G, $I(0)$, R_{XS} and $[I(Q).Q]_{Q\to 0}$ parameters from Guinier and cross-sectional plots, calculates smeared curves for simulated data prior to comparison with neutron experimental data, comparison of simulated model curves with experimental scattering curves and automatically finds the lowest R factor and writes output in a standard ASCII format for importing into SIGMAPLOT (A.S. Nealis & S.J. Perkins, unpublished software).

Modelling

BY444. Creates a sphere model based on a simple geometric shape and attaches extended carbohydrate structures to this. Input parameters include the cube diameter; number of spheres in the x, y, and z directions; spheres to be deleted at the corners to create simple ellipsoids and carbohydrate structures (S.J. Perkins, unpublished software).

APS. Calculates the R_G and volume of a Debye sphere model for a single density model (Chapter 2) (S.J. Perkins, unpublished software).

APSNEW. Calculates the $R_{G\text{-}C}$, α_G and volume of a Debye sphere model for 2 different scattering densities (Chapter 2) (S.J. Perkins, unpublished software).

COTOS. Converts a set of crystal co-ordinates into an array of cubes of a specified size (S.J. Perkins, unpublished software).

FRCC. Based on the hydrodynamic simulation program, GENDIA (Garcia de la Torre & Bloomfield, 1977a,b), it calculates the frictional coefficient <f> (Chapter 2). The program utilizes the diagonal approximation, where the off-diagonal components of the Oseen tensor are usually much lower than the diagonal ones with little effect on the precision of the calculation.

SCT. Calculates the theoretical scattering curve from the cube co-ordinates of

a specified model (Chapter 2) and then creates a composite file of model and experimental scattering curves for plotting with SCTPL (S.J. Perkins, unpublished software).

SCTNWW. Calculates the complete Debye scattering curve for a model of two scattering densities (Chapter 2) and then creates a composite file of model and experimental scattering curves for plotting with SCTPL (S.J. Perkins, unpublished software).

SLICER. Takes the model co-ordinates and depicts them pictorially on a 2 dimensional grid. Additional co-ordinates can be added (A.S. Nealis, unpublished software).

Secondary structure prediction

The Robson and Chou-Fasman programs were kindly made available through Dr M.J.E. Sternberg (Birkbeck College, London; ICRF, London). The versions implemented on the Cyber were:

ROBSCDC. The Robson secondary structure prediction of a given primary structure is calculated for an amino acid sequence in the 3-letter code format. The decision constants for the four conformations (α-helix, β-sheet, β-turn and random coil) can be varied to correspond to an unbiased, α-biased or β-biased structure (Chapter 2).

CHOUCDC. Similarly, the secondary structure is predicted using the Chou-Fasman method and the prediction is unbiased.

PREDRB. Produces graphical output of averaged Robson predictions of aligned sequences. Plots of either the percentage propensities or the average total score are drawn for α-helix, β-strand, β-turn and random coil conformations (Perkins *et al.*, 1988; A.S. Nealis & K.F. Smith, unpublished software).

PREDCF. Produces graphical output of averaged Chou-Fasman predictions of aligned sequences. Plots of percentage propensities of α-helix, β-strand and

random coil are drawn (Perkins *et al.*, 1988; A.S. Nealis & K.F. Smith, unpublished software).

Databases

The protein sequence database OWL (SERC, Daresbury Laboratory) is a composite of three major sequence databases (National Biomedical Research Foundation, SWISSPROT and GenBank) and four minor ones which were maintained by Dr A.J. Bleasby, who is thanked for his support. Sequences were downloaded onto either the Cyber or the datastation using KERMIT.

CHAPTER 4.

STRUCTURAL STUDIES ON α_1-ANTITRYPSIN AND $C\bar{1}$ INHIBITOR

(4.1) INTRODUCTION

The human serine protease inhibitors of plasma show a remarkable degree of sequence homology and have been grouped together in a superfamily termed SERPINs (SERine Protease INhibitor) (Carrell & Travis, 1985). Many have important roles in the physiological regulation of certain serine proteases, although some have no known function. The most abundant and well characterised member of this superfamily is α_1-antitrypsin (M_r 51,500), sometimes termed α_1-proteinase inhibitor (Carrell *et al.*, 1982). At least 50 other members have been identified (Huber & Carrell, 1989) including antithrombin III (Carrell *et al.*, 1981), α_1-antichymotrypsin (Chandra *et al.*, 1983; Bao *et al.*, 1987), the C1 inhibitor component of complement (Davis *et al.*, 1986; Bock *et al.*, 1986, Tosi *et al.*, 1986) and the non-inhibitory serpin, ovalbumin (Kurachi *et al.*, 1981). In a healthy individual, α_1-antitrypsin is present at a plasma concentration of 1.3 mg/ml (Jeppsson *et al.*, 1978), but this substantially increases under inflammatory conditions. Two common mutants of α_1-antitrypsin are the Z-variant where Glu342 is replaced by a lysyl residue (Owen & Carrell, 1977) and the S-variant where Glu264 is replaced by valine (Owen *et al.*, 1976). Patients homozygous for Z have much decreased levels of α_1-antitrypsin and suffer from frequent and severe bouts of pulmonary emphysema.

α_1-antitrypsin exhibits its highest affinity towards leukocyte elastase released by neutrophil leukocytes (Travis & Salvesen, 1983). It acts by presenting the enzymatic reactive site as an ideal substrate to the proteolytic enzyme to form a stable stoichiometric 1:1 complex, thereby inhibiting it. The reactive site is cleaved if the inhibitor-protease complex is dissociated and the serpin is irreversibly inhibited. The reactive site was identified as being the Met358 - Ser359 peptide bond, which is situated 37 residues from the C- terminus of α_1-antitrypsin (Carrell *et al.*, 1980; Johnson & Travis, 1978). In the intact inhibitor the reactive site is thought to reside on a loop on the surface of the molecule (Engh *et al.*, 1990; Stein & Chothia, 1991).

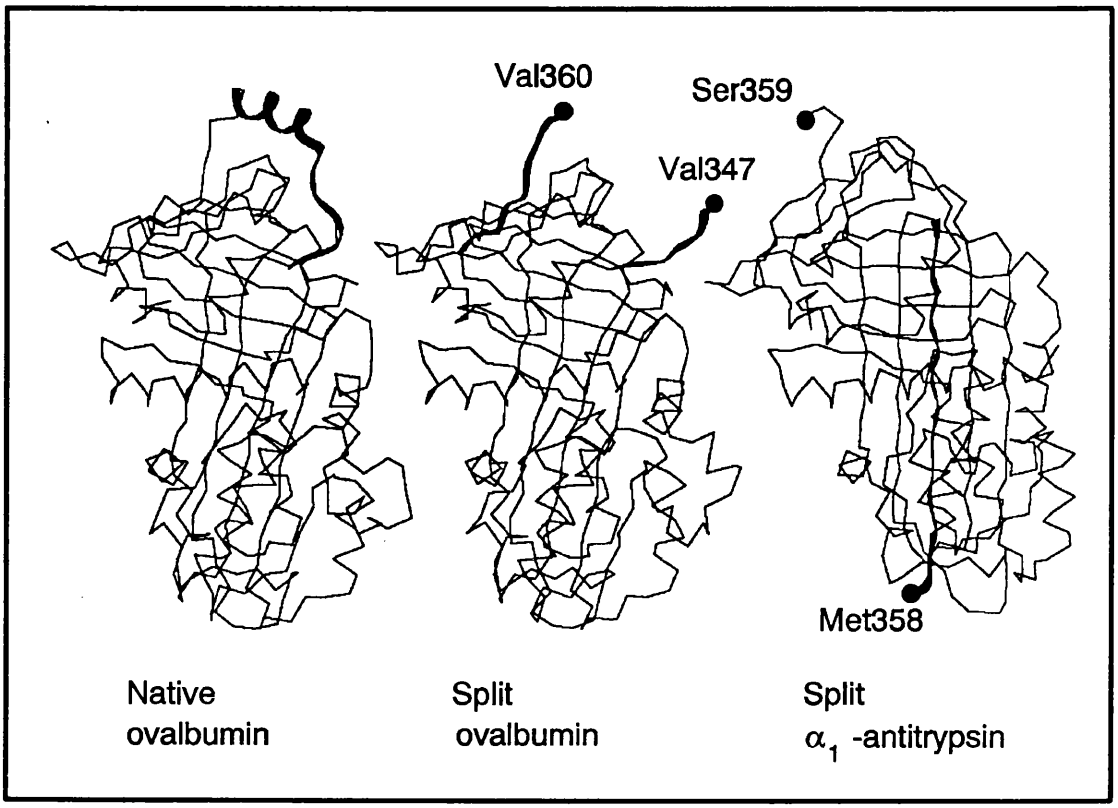


Figure 4.1. Crystal structures of native ovalbumin and the split forms of ovalbumin and α_1-antitrypsin.

The crystal structures of native and split ovalbumin are compared with split α_1-antitrypsin. The co-ordinates for native and split ovalbumin were kindly provided by Dr P.E. Stein (Stein *et al.*, 1990) and Dr R. Engh and Prof. R. Huber (Wright *et al.*, 1990). The co-ordinates for α_1-antitrypsin were taken from the Brookhaven database, reference 7API. The structures were positioned by superimposing α-helices A and B and β-sheet B (Wright *et al.*, 1990). The residues involved in the formation of the reactive loop α-helix in the native form are highlighted in black. The cleavage of ovalbumin resulted in the loss of 12 residues (Val348-Ser359) (α_1-antitrypsin numbering). The residues at either end of this cleavage are indicated. The residues at the reactive site of α_1-antitrypsin are indicated.

The activity of α_1-antitrypsin is controlled either by oxidation of the methionine residue at the reactive site to the sulphoxide form (Schechter *et al.*, 1977) or by proteolytic cleavage of the reactive site (Carrell *et al.*, 1982; Carrell

& Owen, 1985). Oxidation is important physiologically during phagocytosis because leukocytes release myeloperoxidase which oxidizes the methionine residue (Maier *et al.*, 1989). At the site of infection, other proteolytic enzymes also released by leukocytes are no longer inhibited by α_1-antitrypsin. These then break down the surrounding connective tissue as a precursory event to expulsion of the foreign body (Carrell *et al.*, 1982). The spread of this attack on tissue is constrained by α_1-antitrypsin which is not inactivated by leukocytes further away from the site of infection. Unfortunately, other oxidizing agents such as cigarette smoking (Gadek *et al.*, 1979; Carp *et al.*, 1982) are also capable of inhibiting α_1-antitrypsin resulting in pulmonary emphysema (Beatty *et al.*, 1980). The reactive site of α_1-antitrypsin can be cleaved by many proteases, including those secreted by invading pathogens. This "split" α_1-antitrypsin is no longer able to inhibit proteases and the unrestricted neutrophil enzymes may provide nutrients for the pathogens (Travis *et al.*, 1990).

The crystal structures for five serpins have been determined, native ovalbumin (resolution 0.195 nm) (Stein *et al.*, 1990) and the split forms of α_1-antitrypsin (resolution 0.30 nm) (Löbermann *et al.*, 1984), ovalbumin (also known as plakalbumin) (resolution 0.28 nm) (Wright *et al.*, 1990), antithrombin III (resolution 0.32 nm) (Mourey *et al.*, 1990) and α_1-antichymotrypsin (resolution 0.27 nm) (Baumann *et al.*, 1991). The structures of three serpins, native ovalbumin, and the split forms of ovalbumin and α_1-antitrypsin are shown in Figure 4.1. All have similar and secondary and tertiary structures, containing three β-sheets and nine α-helices. The three β-sheets (called A, B and C) contain 6, 6 and 4 strands respectively. The exceptions to this are the native and split forms of ovalbumin where sheet A only contains 5 strands (Wright *et al.*, 1990). The native form of ovalbumin has an intact reactive loop (Gettins, 1989) which comprises of 19 residues (Gly349 to Glu368, α_1-antitrypsin numbering) in an α-helical structure (Stein *et al.*, 1990). In split α_1-antitrypsin, antithrombin III and α_1-antichymotrypsin, the newly-formed N- and C-termini at the cleavage site (Met358 and Ser359; Arg393-Ser394; Leu358-Ser359 respectively) lie at opposite ends of the molecule and are separated by 6 to 7 nm. This is suggestive that a significant conformational change must have occurred on cleavage. It is

commonly assumed that the central strand, s4A, of β-sheet A is removed in the intact protein (Löbermann *et al.*, 1984; Stein & Chothia, 1991) and by analogy with ovalbumin, may form an α-helix containing the reactive site (Stein *et al.*, 1990). The similarity of the rest of the ovalbumin structure with split α_1-antitrypsin is very high, with the differences in sequence being accommodated at a few surface loops (Figure 4.1) (Stein *et al.*, 1990; Stein & Chothia, 1991). The mean difference between the α-carbons was 0.069 nm (Wright *et al.*, 1990). Since the separation of the reactive site peptides was not found in the non-inhibitory serpin, ovalbumin, it may be linked to the inhibitory activity of the other serpins. Insertion of the reactive loop into β-strand A seems necessary for inhibitory activity (Schulze *et al.*, 1990). Residues in the base of the reactive centre loop, the "hinge" region 10 to 14 residues from the reactive site (Figure 4.12) are typically small in the inhibitory serpins. In the non-inhibitory serpins, like ovalbumin and angiotensiogen, the residues have large side-chains and these may prevent the conformational change from occurring (Carrell *et al.*, 1991). This has been substantiated by the lack of complex formation with substrates of inhibitory serpins if the residues in this region were substituted for large-chain amino acids (Skriver *et al.*, 1991).

Split α_1-antitrypsin is a more stable form of the protein. Using circular dichroism and fluorescence, Bruch *et al.* (1988) confirmed that the secondary structures of the native and split forms of four serpins (including α_1-antitrypsin and $C\bar{1}$ inhibitor) were significantly different. The split forms of α_1-antitrypsin and antithrombin III are more resistant to denaturation by heat (Hervé & Ghélis, 1990; Gettins & Harten, 1988), suggesting the transition from a stressed native form to a relaxed cleaved form of the protein (Carrell & Owen, 1985). The non-inhibitory serpin, ovalbumin does not show the characteristic spectral changes or resistance to heat upon cleavage. Although the crystal structure suggests a major conformational change upon cleavage, the glycosylation sites and heparin binding sites on the split forms of serpins are not in conflict with what would be expected in the native form (Huber & Carrell, 1989). This puts a constraint as to the extent of any conformational change. However no experimental data are available on whether the conformational transition of α_1-antitrypsin is a localised or a more

global rearrangement of the structure.

$C\bar{1}$ inhibitor, previously known as $C\bar{1}$ inactivator or $C\bar{1}$ esterase inhibitor, controls both the activation of C1 to $C\bar{1}$ and the activity of $C\bar{1}$ (Harpel, 1976; Sim & Reboul, 1981; Cooper 1985; Davis, 1988). It is present at 0.20 mg/ml in plasma and forms covalent 1:1 complexes at the reactive sites of the $C\bar{1}r$ and $C\bar{1}s$ subcomponents of $C\bar{1}$, preventing further complement activation. Like α_1-antitrypsin, the reactive site of $C\bar{1}$ inhibitor (Arg444-Thr445) is irreversibly cleaved upon complex formation (Salvesen *et al.*, 1985) resulting in a much more stable protein structure (Bruch *et al.*, 1988). Physiological control of the activity of $C\bar{1}$ inhibitor is brought about by cleavage around the reactive centre by neutrophil proteases (Pemberton *et al.*, 1989). While $C\bar{1}$ inhibitor is the only known inhibitor of $C\bar{1}r$ and $C\bar{1}s$, it is a critical regulatory component of the coagulation, fibrinolytic and kinin releasing systems. Much decreased levels of $C\bar{1}$ inhibitor lead to hereditary angioneurotic edema (Donaldson & Evans, 1963; Rosen *et al.*, 1971; Davis, 1988). This is associated with either inherited deficiencies in $C\bar{1}$ inhibitor (type I) or dysfunctional forms of $C\bar{1}$ inhibitor (type II) which are characterised by a single point mutation at the reactive site (Aulak *et al.*, 1988; Aulak & Harrison, 1990).

Compositional and structural studies on the $C\bar{1}$ inhibitor are required to understand the mechanism of control of C1 activation in serum. Its molecular weight has been estimated to fall in a range of 98,000 to 116,000 by SDS gel electrophoresis and ultracentrifugation analyses (Pensky *et al.*, 1961; Haupt *et al.*, 1970; Reboul *et al.*, 1977; Nilsson & Wiman, 1982; Harrison, 1983). However the sequence shows that the protein mass is only 52,300 (Bock *et al.*, 1986). The commonly assumed total molecular weight of 104,000 implies a carbohydrate content as high as 49% (Bock *et al.*, 1986). It is thus strongly glycosylated, but previous determinations of its carbohydrate content have ranged from 33% to 35% by mass (Haupt *et al.*, 1970; Harrison, 1983). The structure of the N-linked and O-linked oligosaccharides have been determined (Strecker *et al.*, 1985). The amino acid sequence indicated the presence of 6 N-linked oligosaccharide binding sites (Asn3, Asn47, Asn59, Asn216, Asn231 and Asn330) (Bock *et al.*, 1986). These and

7 O-linked oligosaccharides were identified by high voltage electrophoresis (Perkins *et al.*, 1990d). These corresponded to the seven carbohydrate attachment sites at Ser42, Thr26, Thr49, Thr61, Thr66, Thr70 and Thr74 (Bock *et al.*, 1986).

The first structural investigations on $C\bar{1}$ inhibitor were based on sedimentation data which reported extremely small $s^{o}_{20,w}$ values at around 3.7 S (Schultze *et al.*, 1962; Haupt *et al.*, 1970; Reboul *et al.*, 1977; Chesne *et al.*, 1982), indicating a highly elongated structure (Odermatt *et al.*, 1981; Perkins, 1985). Electron microscopy suggested that $C\bar{1}$ inhibitor had an elongated two-domain "head-and-tail" structure of total length of 33 to 36 nm and a head region of approximately 4 nm (Odermatt *et al.*, 1981). The existence of a two-domain structure was supported by microcalorimetry fluorescence (Lennick *et al.*, 1985). Because of the high degree of homology of the $C\bar{1}$ inhibitor amino acid sequence with that of α_1-antitrypsin, the crystal structure of α_1-antitrypsin can be used to model the C-terminal 365 residues, corresponding to the "head" of $C\bar{1}$ inhibitor (Löbermann *et al.*, 1984; Bock *et al.*, 1986; Huber & Carroll, 1989). The "tail" therefore corresponds to the heavily glycosylated N-terminal 113 residues of presumed length anywhere between 26 nm to 53 nm. The function of the N-terminal domain has not been clearly established but may be involved in the binding to leukocytes since the removal of sialic acid decreased dramatically the binding of $C\bar{1}$ inhibitor to leukocytes (Chang *et al.*, 1991). The terminal sialic acid also plays a role in the *in vivo* survival of $C\bar{1}$ inhibitor since its removal results in rapid clearance $C\bar{1}$ inhibitor from the blood (Minta, 1981). The structure of the N-terminal domain, however, is unknown (Davis, 1988).

In principle, small-angle solution scattering should detect the large conformational change between the native and split forms of α_1-antitrypsin and $C\bar{1}$ inhibitor. Furthermore, the physiological structure of α_1-antitrypsin can be compared with that observed in higher phosphate concentrations approaching that in which split form was crystallised (2.5 M). Neutron contrast variation in mixtures of H_2O and ^2H_2O provides information on the arrangement of hydrophobic and hydrophilic amino acids and the conformation of the carbohydrate structures. The crystal co-ordinates of split α_1-antitrypsin

(Löbermann *et al.*, 1984) can be used to calculate the scattering curves of native and split α_1-antitrypsin. This permits both a careful comparison of the two structures, and also a critical assessment of the methods used for these calculations. The two domain model for $C\bar{1}$ inhibitor can be compared quantitatively with the crystal structure of α_1-antitrypsin and molecular models for $C\bar{1}$ inhibitor can be developed in order to assess its function.

(4.2) MATERIALS AND METHODS

(4.2.1) Preparation of α_1-antitrypsin and $C\bar{1}$ inhibitor

α_1-antitrypsin and $C\bar{1}$ inhibitor were prepared by Dr R. A. Harrison (Smith *et al.*, 1990; Perkins *et al.*, 1990d). Native α_1-antitrypsin and $C\bar{1}$ inhibitor were purified using methods described by Laurell *et al.* (1975) and Harrison & Lachmann (1986), respectively. Split α_1-antitrypsin was formed by papain cleavage (Johnson & Travis, 1977). This not only cleaved at the reactive-centre Met358-Ser359 peptide bond but also at Gly5-Asp6 and Asp12-Thr13 peptide bonds. Sequence analyses of the N-terminus showed that the small peptides Glu-Asp-Pro-Gln-Gly and Asp-Ala-Gln-Lys-Thr-Asp were lost upon dialysis against scattering buffers whereas native α_1-antitrypsin remained intact, as expected. Split $C\bar{1}$ inhibitor was cleaved by an unidentified protease which produced fragments indistinguishable from those produced by *Pseudomonas aeruginosa* elastase (Pemberton *et al.*, 1989). As with α_1-antitrypsin, cleavage not only occurred at the reactive-centre peptide bond, Ser441-Val442, but also in approximately equimolar amounts at Met31-Leu32 or Leu32-Phe33 peptide bonds. Although after dialysis the C-terminal peptide (M_r 4400) remained associated with the major $C\bar{1}$ inhibitor fragment, up to 30% of the N-terminal 31 or 32 residues (M_r 3200) may have been lost.

Initial neutron scattering experiments for native α_1-antitrypsin and $C\bar{1}$ inhibitor were performed in 0.012 M-sodium/potassium phosphate buffer, pH 7.0 containing 1 mM EDTA and 0.2 M-NaCl. Neutron and X-ray scattering data were also obtained for both native and split α_1-antitrypsin over a range of concentrations of sodium/potassium phosphate buffer from 0.05 M to a maximum of 2.5 M, pH 8.0, used in the crystallization of split α_1-antitrypsin (Löbermann *et al.*, 1982). $K_2HPO_4.3H_2O$ was dissolved to 3.0 M in H_2O or ^2H_2O, back-titrated with stirring at room temperature to pH 8.0 by additions of known amounts of $NaH_2PO_4.2H_2O$, and diluted to the final desired PO_4 concentration. The α_1-antitrypsin samples were dialysed with stirring at 4°C into 0%, 70%, 80% and 100% ^2H_2O buffers for neutron measurements or into 0% ^2H_2O buffers for X-ray

measurements, with four buffer changes over at least 36 hours. Similarly, $C\bar{1}$ inhibitor was dialysed into 0%, 80% or 100% ^2H_2O buffers for neutron measurements. Optical densities for concentration determinations were recorded at 280 nm immediately afterwards, and scattering studies were performed within the following 24-48 hours. SDS/PAGE analysis of samples before and after scattering analysis showed that they had the anticipated band patterns, and that no radiation induced damage or non-specific proteolysis had occurred.

(4.2.2) Neutron and X-ray scattering experiments

(4.2.2.1) α_1-antitrypsin

Neutron data were obtained on Instrument D17 at the high flux reactor at the ILL, Grenoble. Guinier data were obtained in three sessions using sample-to-detector distances of 3.44 m with a wavelength of 1.095 nm (corresponding to a Q range of 0.016 to 0.615 nm^{-1}) and 3.396 m with a wavelength of 1.105 nm (Q range of 0.07 to 0.60 nm^{-1}). Data at large Q were obtained in one session using a sample-to-detector distance of 1.41 m and a wavelength of 1.095 nm (Q range of 0.045 to 1.451 nm^{-1}). Samples were measured for 3 to 16 min in ^2H_2O buffers and 1 to 6 h in H_2O buffers.

X-ray data were obtained at the SRS at Daresbury, operating at 2.00 GeV. The low angle solution scattering cameras at Station 2.1 (preliminary data was measured by Dr S.J. Perkins in 1985 - 1986) (beam currents of 61 mA to 52 mA) and subsequently at Station 8.2 (261 mA to 75 mA) were used in a total of four sessions. At Station 2.1, the detector-to-sample distance was 2.52 m, with a wavelength of 0.155 nm (Q range of 0.189 to 3.206 nm^{-1}). At Station 8.2, the sample-to-detector distances of 2.605 m, 2.608 m and 3.83 m and a wavelength of 0.159 nm resulted in Q ranges of 0.051 to 1.534 nm^{-1} and 0.039 to 1.012 nm^{-1} respectively. Samples were measured for 10-20 min in a protein dilution series.

(4.2.2.2) $C\bar{1}$ inhibitor

Neutron samples were measured on Instrument D17 at the ILL, Grenoble by Dr S.J. Perkins in 1986. Guinier R_G data were collected using a sample-to-detector distance of 3.46 m and neutron wavelengths of 1.385 to 1.395 nm or 1.600 nm, giving a Q range of 0.033 to 0.486 nm^{-1}. Scattering curves at larger Q ranges were recorded with a sample-to-detector distance of 1.40 m and wavelengths of 1.001 to 1.004 nm, to give a Q range of 0.13 to 1.60 nm^{-1}, and also with a detector rotation of 19.89° relative to the main beam direction to give a larger Q range of 0.8 to 3.6 nm^{-1}. Some data were collected on instrument D11 with a sample-to-detector distance of 10.5 m and wavelength of 1.00 nm to achieve a lower Q range of 0.0019 to 0.220 nm^{-1}.

The experimental or theoretical (from Debye sphere models) scattering curve $I(Q)$, measured in reciprocal space can be transformed into real space by the Indirect Transformation Procedure of Glatter (1982). This provides an alternative method of determining the R_G, $I(0)$ and the length of $C\bar{1}$ inhibitor (Chapters 2 & 3). A full experimental scattering curve is created from the 3 domains (or sub-curves) corresponding to the 3 instrumental setups of D17 (see above). A total of 10 splines were used to fit the 75 experimental data points, using a specified D_{max} of 20 nm. ITP analyses of model scattering curves of 99 points out to a Q of 1.6 nm^{-1} were based on a D_{max} of 25 nm and 10 splines.

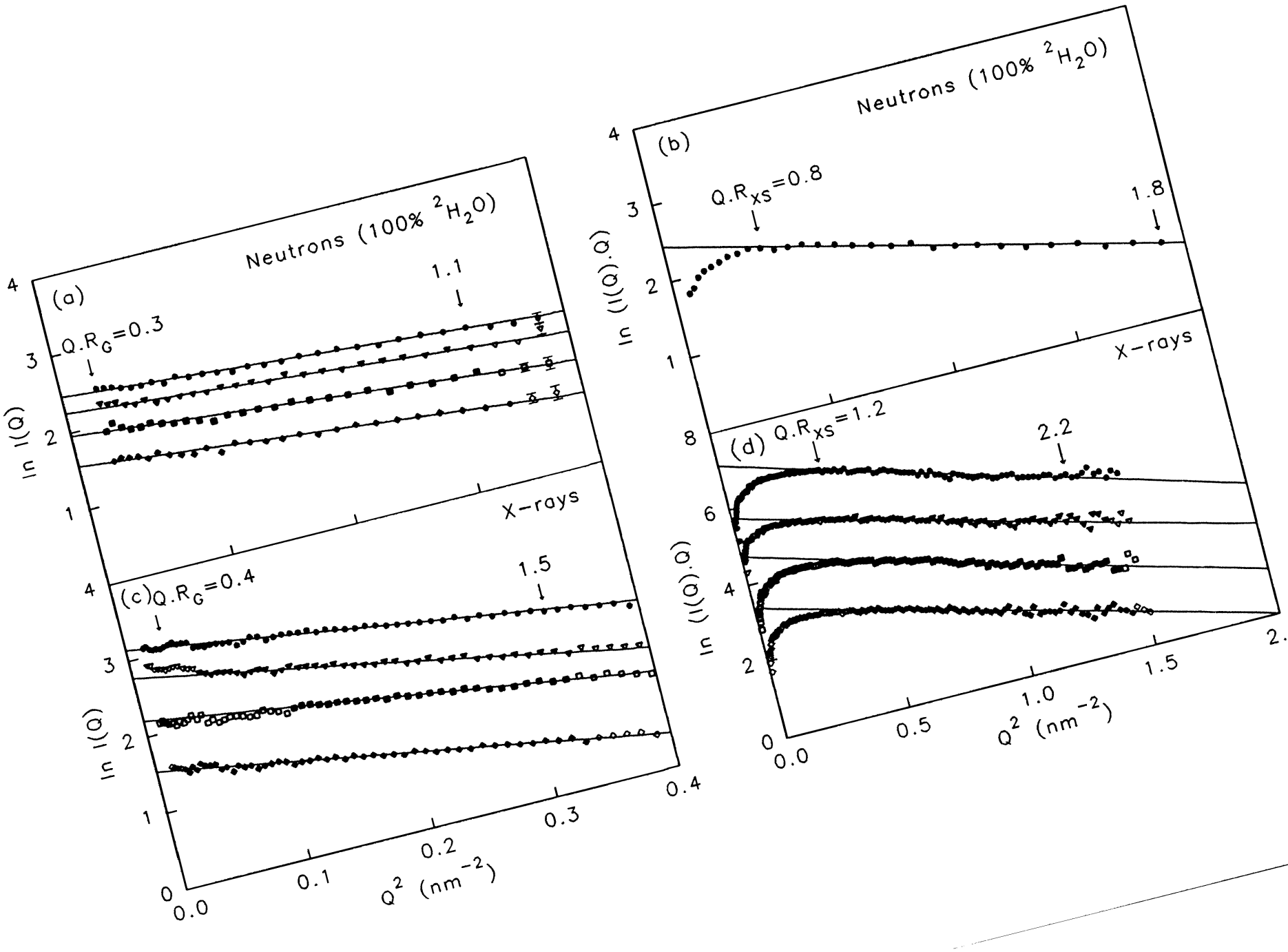
(4.2.3) Debye sphere modelling of α_1-antitrypsin and $C\bar{1}$ inhibitor

Molecular modelling of the neutron and X-ray scattering curves of α_1-antitrypsin and the neutron curves of $C\bar{1}$ inhibitor used small spheres based on the Debye equation (Chapter 2). The crystal co-ordinates of split α_1-antitrypsin (Protein Data Base ref. 7API) (Löbermann *et al.*, 1984; Engh *et al.*, 1989) were subdivided into cubes of side 0.608 nm using the program COTOS (Chapter 3) applying a cut-off such that the total volume of the cubes equalled that of the dry volume of the residues with crystal co-ordinates (Chothia, 1975). The scattering curve simulations were based on the use of overlapping spheres of volume 0.608^3

nm^3 that were positioned at the centre of each cube. The spheres were thus of diameter 0.754 nm, which was much less than the nominal resolution of $2\pi/Q_{max}$ of the scattering curves, i.e. 4.3 nm for Q_{max} = 1.45 nm^{-1}. To account for the contrast dependence of the neutron scattering curves, the 291 spheres of the model were subdivided into an outer hydrophilic surface region and an inner hydrophobic core on the basis of the volume of each component, each of which was assigned a different scattering density. Strictly the neutron scattering densities should be assigned as 49.4 % ^2H_2O and 31.6 % ^2H_2O respectively (Perkins, 1986), but in view of the simplicity of this model, absolute densities of 60 % ^2H_2O and 30 % ^2H_2O respectively were used. These gave net scattering densities of 60% and 30% ^2H_2O in 0% ^2H_2O, 20% and 50% ^2H_2O in 80% ^2H_2O and 40% and 70% ^2H_2O in 100% ^2H_2O contrasts. The neutron model was converted into the hydrated X-ray model (Chapter 2) by rescaling the co-ordinates by a factor of 1.10 derived from the cube root of the ratio of the hydrated and dry volumes, to give a cube side of 0.668 nm. Since X-rays were affected by the layer of water surrounding the protein, the outermost spheres now corresponded to the hydration shell as well as the carbohydrate and hydrophilic amino acid residues. Electron densities were taken as 334 e nm^{-3} for solvent, 435 e nm^{-3} for the hydrophilic spheres of the model, and 389 e nm^{-3} for the hydrophobic spheres.

(4.2.4) Hydrodynamic modelling of α_1-antitrypsin

The crystal structure was subdivided into 31 dry cubes of side 1.216 nm using COTOS (Chapter 3) with a cutoff chosen to make the overall dimensions of the cubes close to those of the co-ordinates. The 31 "dry" co-ordinates were then rescaled by a factor of 1.39 to allow for hydration and non-overlap to give a final sphere diameter of 1.69 nm (Chapter 3) (Figure 4.14). If it is assumed that the protein surface is uniformly covered by a single layer of water molecules of thickness 0.36 nm (which corresponds to a hydration of 0.3 g H_2O /g protein, as seen by crystallography; Perkins, 1986, 1989b), good agreement with the dimensions of the crystal co-ordinates can be obtained.


(a)
Neutrons (100% ^2H_2O)
$Q.R_G$=0.3
1.1
X-rays
(c) $Q.R_G$=0.4
1.5
$I(Q)/I(Q).\underline{\underline{S}}(Q)$
Q^2 (nm^{-2})
(b)
Neutrons (100% ^2H_2O)
$Q.R_{XS}$=0.8
1.8
X-rays
(d) $Q.R_{XS}$=1.2
2.2

(4.3) RESULTS AND DISCUSSION

(4.3.1) Neutron scattering of native α_1-antitrypsin in physiological buffers.

Good linear Guinier plots were obtained in an acceptable $Q.R_G$ range of 0.3 to 1.1 for native α_1-antitrypsin in 100% ^2H_2O under physiological conditions (Figure 4.2(a)). A dilution series of the R_G data for α_1-antitrypsin over concentrations of 2.4 to 13.7 mg/ml (Figure 4.3(a)) in 3 different sodium/potassium phosphate concentrations showed that the R_G was dependent upon both the protein and phosphate concentration. The protein concentration dependence was also found by sedimentation experiments on α_1-antitrypsin (Kress & Laskowski, 1973). Extrapolation to zero concentration by regression gave an R_G

Figure 4.2. Guinier R_G and R_{XS} plots of neutron and X-ray data for native and split α_1-antitrypsin.

The buffer in all cases was 0.05 M-sodium/potassium phosphate buffer, pH 8.0. The $Q.R_G$ used for line fitting is arrowed and indicated by the filled symbols.

(a) Neutron Guinier plots of the dilution series for native α_1-antitrypsin in 100% ^2H_2O at concentrations of 12.4 mg/ml (o), 9.8 mg/ml (▽), 7.0 mg/ml (□) and 4.8 mg/ml (◇). The Q range used for fitting is 0.17 to 0.57 nm^{-1}. The slopes increase slightly with increasing dilution; the relative vertical displacements reflect the dilutions.

(b) Neutron cross-sectional plot of native α_1-antitrypsin at 8.0 mg/ml. The Q range used for fitting is 0.61 to 1.42 nm^{-1}.

(c) X-ray Guinier plots of native (o and ▽) and split (□ and ◇) α_1-antitrypsin at concentrations of 16.8 mg/ml (o), 12.6 mg/ml (▽), 18.9 mg/ml (□) and 14.2 mg/ml (◇). The Q range used for fitting the data is 0.15 to 0.57 nm^{-1}. The curves were displaced arbitrarily for clarity. A protein concentration dependence of the R_G values is visible but not significant.

(d) X-ray cross-sectional plot of the samples in (c). The $Q.R_{XS}$ values are based on a Q range of 0.60 to 1.20 nm^{-1}. Note that the vertical scale is halved in (d) relative to (a), (b) and (c).

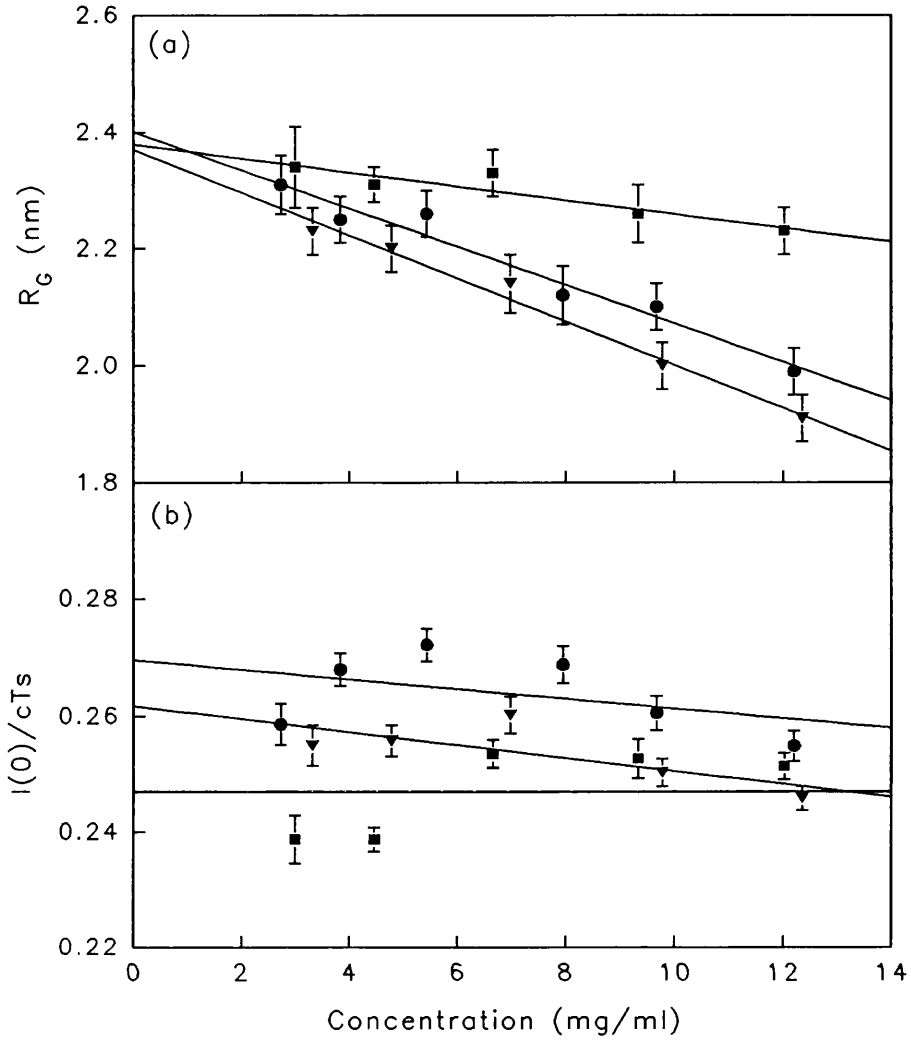


Figure 4.3. Concentration dependence of the R_G and $I(0)/c$ parameters of native α_1-antitrypsin by neutrons.

(a) The R_G values were measured by neutron scattering in 100% ^2H_2O in sodium/potassium phosphate buffers at concentrations of 0.012 M (●), 0.05 M (▾), 0.50 M (■). The R_G at zero α_1-antitrypsin concentration was obtained from linear regression.

(b) The plot $I(0)/cT_s$ of the samples in (a).

of 2.40 ± 0.03 nm for samples in the 0.012 M-sodium/potassium phosphate buffer, 2.37 ± 0.03 nm for the 0.05 M-sodium/potassium phosphate buffer and 2.38 ± 0.03 nm for the 0.5 M-sodium/potassium phosphate buffer (Figure 4.3(a)). The R_G parameter became significantly less dependent on the α_1-antitrypsin concentration with increasing sodium/potassium phosphate concentration. This

	α_1AT	$C\bar{1}$ inhibitor
Residues		
total amino acids	394	478
total carbohydrate	36	89
% mass carbohydrate	14.1	25.7
M_r		
protein	44,300	52,800
carbohydrate	7,300	18,300
total	51,500	71,100
neutron scattering	51,700 ± 2,000	76,000 ± 4,000
Dry volume (nm^3)[1]	65.3	88.2
Overall $\bar{v}$ (ml/g)[2]	0.732	0.7478
Overall $\Sigma b/M_r$ in H_2O (fm)	0.2248	0.2192
in ^2H_2O (fm)	0.3813	0.3846
Predicted A_{280} (1%, 1cm)[3]	3.86	3.87
Literature A_{280}	5.0 - 5.4	3.6 - 4.5
Pred. matchpoint (% ^2H_2O)[4]	40.3	41.0
Exp. matchpoint (% ^2H_2O)		
from $I(0)$	41.3 ± 1.1	42.3 ± 0.5
from $[I(Q).Q]$	42.1 ± 1.2	42.2 ± 0.5
Stuhrmann analysis		
R_{G-C} (nm)	2.61 ± 0.02	4.85 ± 0.01
α_G (x 10^{-5})	29 ± 5	59 ± 3
R_{XS-C} (nm)	1.34 ± 0.01	1.43 ± 0.03
α_{XS} (x 10^{-5})	13 ± 1	29 ± 2
R_{G-C}/R_0	1.35	2.27
Models R_{G-C} (nm)	2.58	4.87
α_G (x 10^{-5})	29	54

Table 4.1. Compositional and scattering data for α_1-antitrypsin and $C\bar{1}$ inhibitor

Footnotes to Table 4.1

[1] The dry volume was taken from Chothia (1975) (amino acids) and Perkins (1986) (carbohydrates) and used in matchpoint and modelling calculations.

[2] Partial specific volumes $\bar{v}$ were calculated from the consensus volume set of Perkins (1986). These values were used in sedimentation simulations.

[3] Calculated from the content of Tyr, Cys and Trp residues by the Wetlaufer (1962) procedure and rescaled using a factor of 1.03 (Perkins, 1986) (Chapter 3). Literature determinations of A_{280} are reported in Bundy & Mehl (1953), Kress & Laskowski (1973) and Bruch *et al.* (1988) for α_1-antitrypsin and Harrison, 1983; Salvesen *et al.*, 1985; Haupt *et al.*, 1970 and Bruch *et al.*, 1988 for $C\bar{1}$ inhibitor.

[4] Calculated using the assumption of the non-exchange of 10% of the mainchain peptide hydrogens (Perkins, 1986).

may have been due to the increased interactions of the dipole moments between different molecules in lower sodium/potassium phosphate buffers. No such dependence was observed for the $I(0)/cT_s$ parameter (Figure 4.3(b)). At larger scattering angles, good cross-sectional Guinier plots were also obtained. Figure 4.2(b) shows that the mean R_{XS} of 1.22 nm was obtained in a satisfactory $Q.R_{XS}$ range of 0.8 to 1.8. The R_{XS} and $I(Q).Q/cT_s$ parameters were independent of sodium/potassium phosphate concentration.

This protein concentration dependence is unusual for plasma glycoproteins although $C\bar{1}$ inhibitor also showed this effect. Analysis of the crystal structure of α_1-antitrypsin showed an asymmetric distribution of acidic and basic charged groups, the S-359 end was positively charged and the M-358 end negatively charged therefore producing a large dipole moment (Löbermann *et al.*, 1984; Engh *et al.*, 1989). $C\bar{1}$ inhibitor also had an asymmetric distribution of charged groups. The serpin domain had a net charge of +6 and the N-terminal domain had a net charge of -20. It was likely that the concentration dependence was due to interparticle interference effects (Guinier & Fournet, 1955).

Full contrast variation using 2.7 to 12.2 mg/ml concentrations of native α_1-antitrypsin was carried out in 0.012 M-sodium/potassium phosphate buffers containing 0%, 70%, 80% and 100% ^2H_2O. The protein concentration dependence was eliminated by regressing the R_G and $I(0)/c$ values to zero concentration. To verify that the neutron data corresponded to monomeric α_1-antitrypsin, the M_r was determined from the $I(0)/c$ values at 0% ^2H_2O (Chapter 3). This required the calculation of the concentration from the absorption coefficient A_{280} (1%, 1 cm). Published values ranged between 5.0 to 5.4 (Bundy & Mehl, 1959; Kress & Laskowski, 1973; Bruch *et al.*, 1988). By using an M_r of 51,500 calculated from the amino acid and carbohydrate composition, the $I(0)/c$ data gave an A_{280} of 5.4 ± 0.2 (Table 4.1). This was in good agreement with Bruch *et al.* (1988) and also showed that α_1-antitrypsin was monomeric in solution. The theoretical calculation of the A_{280} from the molecular composition of α_1-antitrypsin gave a value of 3.86 (Chapter 3) (Wetlaufer, 1962; Edelhoch, 1967; Perkins, 1986) which gave M_r values that are too low. However, it was known that deviations occurred from the linear

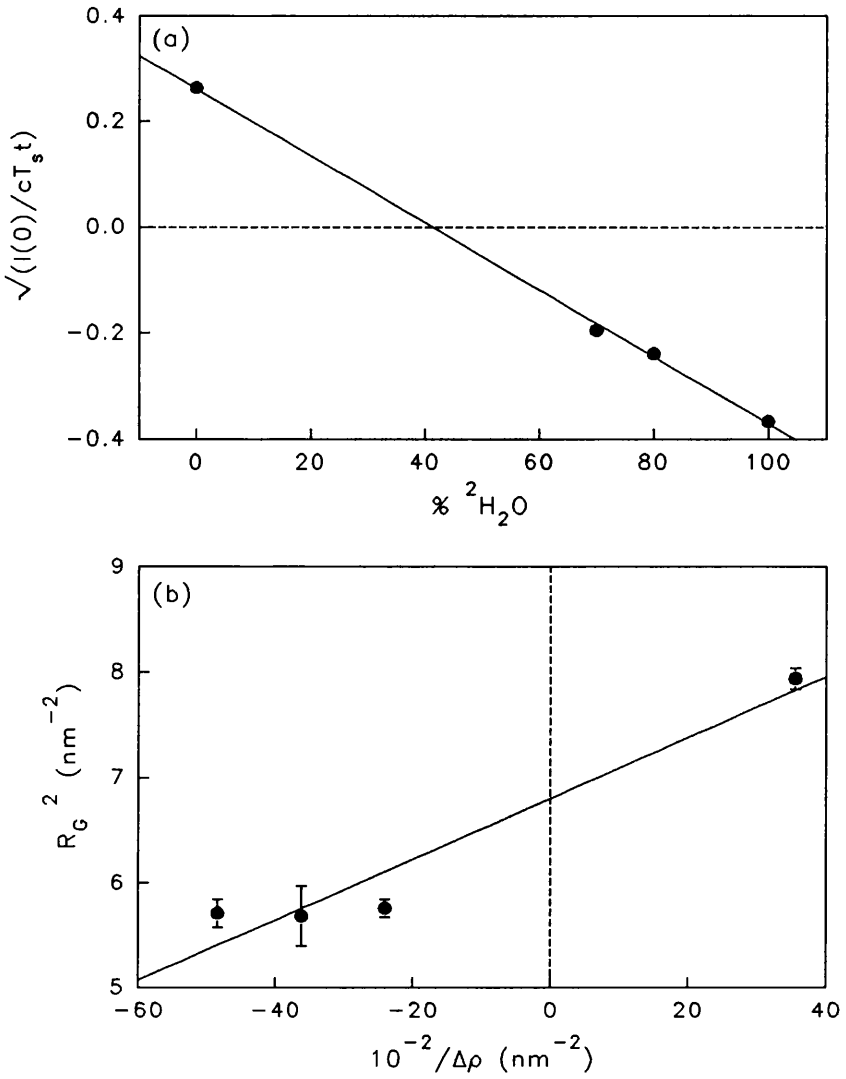


Figure 4.4. Neutron contrast variation analysis of native α_1-antitrypsin.

(a) The matchpoint determination was based on extrapolation of the $I(0)/c$ values to zero concentration, and calculating $\sqrt{I(0)/cT_s t}$ from these. The error in the matchpoint is thus higher than usual.

(b) The Stuhrmann plot of the $R_G{}^2$ against the reciprocal solute-solvent contrast difference $\Delta\rho^{-1}$ is also based on the R_G data extrapolated to zero concentration (Figure 4.3). The regressed line gave an $R_{G\text{-}C}$ of 2.61 ± 0.02 nm and the slope gave a Stuhrmann α of 29 ± 5 x 10^{-5}.

relationship for A_{280} values less than 10 (Wetlaufer, 1962; Perkins, 1986). The neutron matchpoint was determined for comparison with that predicted from the sequence. Figure 4.4(a) gave the matchpoint as 41.3% ± 1.0 % ^2H_2O (16 points in 4 contrasts) which agreed well with a predicted value of 40.3% ^2H_2O (Table 4.1).

The limited cross-sectional data of $\sqrt{[I(Q).Q]_{Q\to 0}/cT_s t}$ (4 points in 2 contrasts) gave a comparable matchpoint of 42.1% ± 1.2 % ^2H_2O. These results showed that by neutron scattering, α_1-antitrypsin was an unhydrated monomeric macromolecule.

The contrast variation R_G data were analyzed in the Stuhrmann plot (Figure 4.4(b)). The R_G at infinite contrast $R_{G\text{-}C}$ was determined to be 2.61 ± 0.02 nm. The elongation ratio, $R_{G\text{-}C}/R_O$ is 1.35 (where R_O is the R_G of the sphere of the same dry volume as α_1-antitrypsin). This is close to the average of 1.28 ± 0.10 found for many globular proteins (Perkins, 1988). The corresponding cross-sectional analysis gave an $R_{XS\text{-}C}$ value of 1.34 ± 0.01 nm. In both cases, the slope α was strongly positive, α_G was 29 ± 5 x 10^{-5} and α_{XS} was 13 ± 1 x 10^{-5}. This indicated that the surface regions of α_1-antitrypsin had a higher scattering density than those near the core. Since the hydrophilic amino acid and carbohydrate residues had a mean matchpoint of 49.4% ^2H_2O and hydrophobic amino acid residues had one of 31.6% ^2H_2O, hydrophilic residues were seen to be predominant in the outer regions of the α_1-antitrypsin structure, as expected.

The triaxial dimensions of α_1-antitrypsin could be estimated from the Guinier analyses (Chapter 2) and compared with those from the crystal structure. The mean length L calculated from the ratio of $\pi.I(0)/[I(Q).Q]_{Q\to 0}$ was 7.6 ± 0.5 nm (4 values). The $R_{G\text{-}C}$ and $R_{XS\text{-}C}$ gave a similar length of 7.8 ± 0.1 nm if α_1-antitrypsin was approximated as an elliptical cylinder. With the length as 7.8 nm, the two shorter axes 2A and 2B were 4.9 ± 0.1 nm and 2.2 ± 0.1 nm (Chapter 2). These dimensions can be compared with those of the crystal structure of α_1-antitrypsin of 6.7 nm x 3.2 nm x 3.2 nm reported in Löbermann *et al.* (1984). The latter should be uniformly increased to 7.2 nm x 3.4 nm x 3.4 nm to give the correct total volume of α_1-antitrypsin. Even though the shorter axes of the protein component were more similar in the crystal structure, molecular graphics showed that the three oligosaccharide sites were positioned in such a way that one of the two shorter axes would be significantly longer if the oligosaccharide chains were extended into solution. The Guinier data therefore suggested that the carbohydrates had extended conformations. This was confirmed by use of the

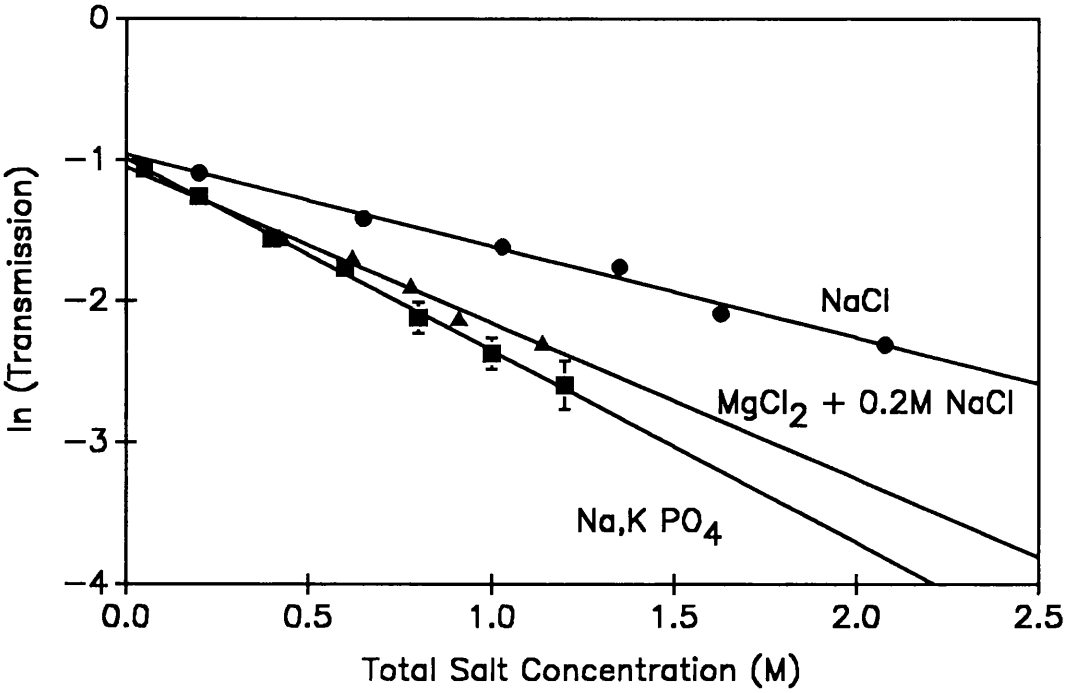


Figure 4.5. X-ray transmissions of salt solutions.

The sample thickness was 1 mm and was contained within mica walls of thickness 10 to 20 μm. The mean transmission of the sodium/potassium phosphate buffers (■) from three independent sessions were compared with a NaCl (●) and a $MgCl_2$ (▲)buffer (Perkins *et al.*, 1985).

neutron R_G and R_{XS} values in 100% ^2H_2O, which emphasize the hydrophobic structure of α_1-antitrypsin more than that of carbohydrate. These gave triaxial dimensions of 7.0 nm x 3.8 nm x 3.1 nm which are now within 0.4 nm of the corrected crystallographic dimensions.

(4.3.2) X-ray and neutron scattering of native and split α_1-antitrypsin in high phosphate buffers

Synchrotron X-ray scattering was used to compare:
(a) the physiological protein structure with that seen in the high phosphate conditions used in X-ray crystallography;
(b) the hydrated structure with the dry structure seen by neutrons and
(c) the native and split structures of α_1-antitrypsin.
As shown in Figures 4.2(c) and 4.2(d), excellent R_G and R_{XS} Guinier plots were obtained for both native and split α_1-antitrypsin in 0.05 M sodium/potassium phosphate buffer. A slight time-dependent aggregation of α_1-antitrypsin was occasionally seen at very low Q values. This was especially prevalent with increased beam currents (especially just after beam refills), increased phosphate concentration, the prior presence of trace aggregates, or the use of native α_1-antitrypsin. The ten individual time-slices of each run were examined to check that the R_G and the $I(0)$ parameters remained constant with time.

A full dilution series was performed with concentrations of native and split α_1-antitrypsin ranging from 2.3 mg/ml to 25.2 mg/ml in sodium/potassium phosphate buffers of phosphate concentrations between 0.05 M and 1.2 M. Sample transmissions decreased with increasing phosphate (Figure 4.5) to the extent that satisfactory signal-to-noise ratios were achieved only up to 0.6 M sodium/potassium phosphate buffer (Figure 4.6). The X-ray analyses showed that, on increasing dilution of α_1-antitrypsin, the $I(0)/c$ values increase (not shown), unlike the corresponding neutron data in 100% ^2H_2O buffers. This was probably due to the different solute-solvent scattering density and radiation in use. The protein concentration dependence decreased with increasing phosphate concentration and the $I(0)/cT_s$ and $[I(Q).Q]_{Q\to 0}/cT_s$ values after extrapolation to

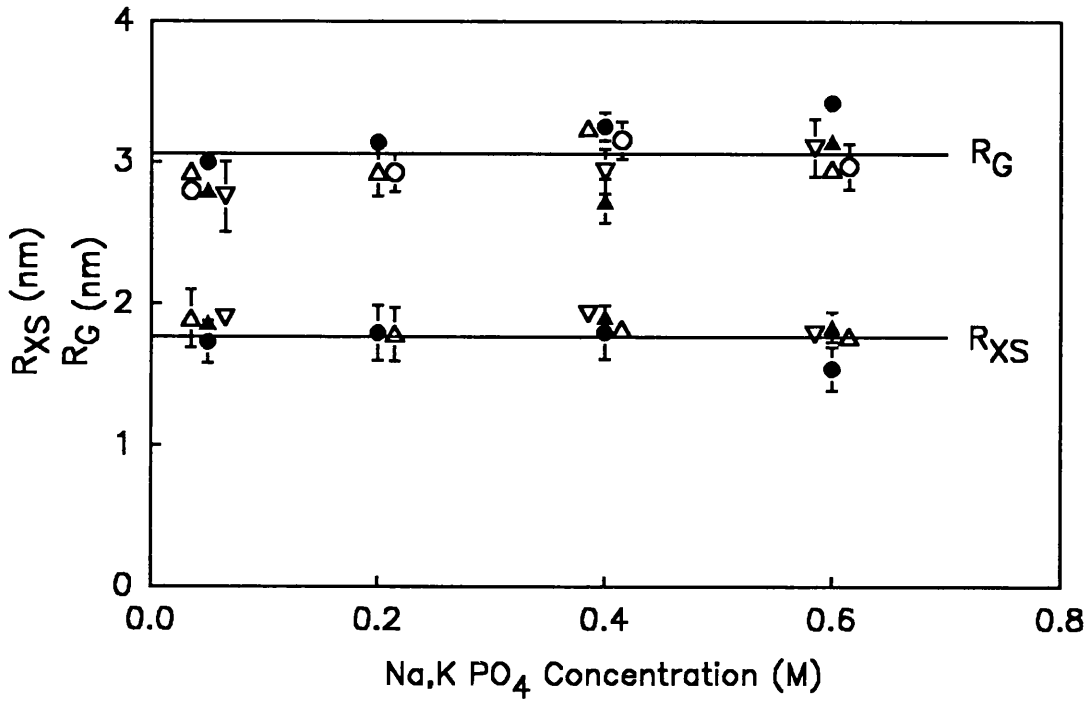


Figure 4.6. Dependence of X-ray Guinier parameters on the phosphate concentration.

The upper plot shows the R_G dependence upon sodium/potassium phosphate concentration for three sessions of X-ray data for native (△, ▽ and o) and split α_1-antitrypsin (▲) and 67% split, 33% native α_1-antitrypsin (●). Error bars are shown only when significant and increase at higher phosphate concentrations. The lower plot shows the corresponding R_{XS} dependence using the same symbols.

zero concentration were constant within error in the different phosphate buffers.

The R_G values did not exhibit a statistically significant dependence on protein concentration, hence the mean R_G from each experiment was shown in Figure 4.6. No concentration dependence is seen for the R_{XS} data and the mean values were given in Figure 4.6. As the phosphate concentration was increased to 0.6 M, the mean R_G of native α_1-antitrypsin remained unchanged. In 0.05 M sodium/potassium phosphate buffer, the mean R_G of native α_1-antitrypsin from three different sessions were 2.80 ± 0.09 nm, 2.93 ± 0.06 nm and 3.00 ± 0.04 nm. These were comparable to the neutron R_G of 2.82 ± 0.03 nm in H_2O as expected, since both experiments corresponded to positive solute-solvent contrasts. The X-ray R_G values tended to be larger than the neutron R_G value due to the

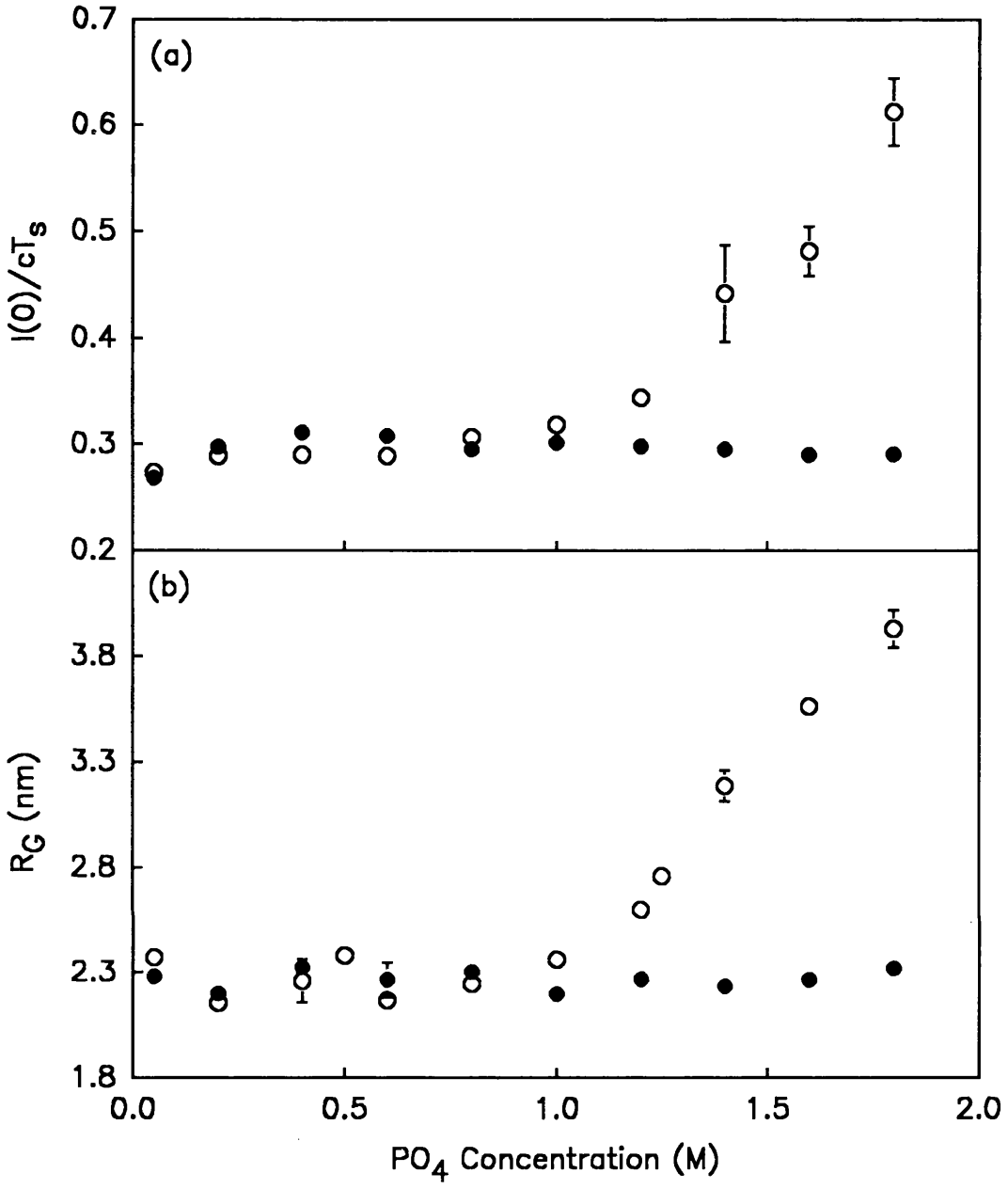


Figure 4.7. Dependence of the neutron Guinier parameters of native (o) and split (●) α_1-antitrypsin in 100% ^2H_2O on phosphate concentration.

(a) Dependence of $I(0)/cT_s$ on sodium/potassium phosphate concentration after extrapolation to zero protein concentration. Between three and six measurements were made at each phosphate concentration in a range of 3.0 to 13.9 mg/ml.

(b) The corresponding R_G values after extrapolation to zero protein concentration.

Error bars are shown only where these are significant.

hydration shell surrounding α_1-antitrypsin visible only by X-rays. Likewise, the mean "hydrated" R_{XS} values of 1.73 ± 0.19 nm (19 values) and 1.82 ± 0.13 nm (7 values) from two experimental sessions were slightly larger than the neutron "dry" R_{XS} value of 1.53 ± 0.10 measured in H_2O in 0.012 M-sodium/potassium phosphate buffer. The corresponding data for split α_1-antitrypsin is also shown in Figure 4.6, and no difference in R_G or R_{XS} was observed. This suggested that the conformational transition between the native and split forms of α_1-antitrypsin involved only a localised rearrangement of the protein structure, and that the loss of the N-terminal peptides (Section 4.2) had a negligible effect on the structure.

The tridimensional axes L x 2A x 2B of native α_1-antitrypsin were determined as 7.9 nm x 7.0 nm x 2.0 nm from the X-ray R_G data, assuming an elliptical cylinder for its structure. These compared well with the corresponding dimensions of 8.3 nm x 5.7 nm x 1.8 nm calculated from the neutron R_G data in H_2O. The two shorter axes were more asymmetric in high positive solute-solvent contrasts than in the case of the ^2H_2O data. This was attributed to the structure of the carbohydrate moiety which scattered more strongly than the protein in this contrast by both X-rays and neutrons. The effect of this was clear from the different slopes of Figures 4.2(a) and 4.2(c).

The effect of high phosphate concentrations (up to 1.8 M phosphate) on native and split α_1-antitrypsin was studied by neutron scattering. This technique was not affected by poor sample transmissions as by X-rays. The R_G and $I(0)/c$ Guinier parameters were dependent on protein concentration between phosphate concentrations of 0.05 M and 0.5 M. No concentration dependence was observed at phosphate concentrations higher than 0.5 M and the mean values were calculated. These were regressed to zero concentration and plotted in Figure 4.7. No differences in the Guinier parameters for native and split α_1-antitrypsin could be detected up to a phosphate concentration of 0.8 M. The conformational changes upon ligation of proteins such as haemoglobin (Conrad *et al.*, 1969), aspartate transcarbamylase (Moody *et al.*, 1979) and hexokinase (McDonald *et al.*, 1979) were detectable from Guinier analyses. It therefore can be concluded that the tertiary structure of native and split α_1-antitrypsin under physiological

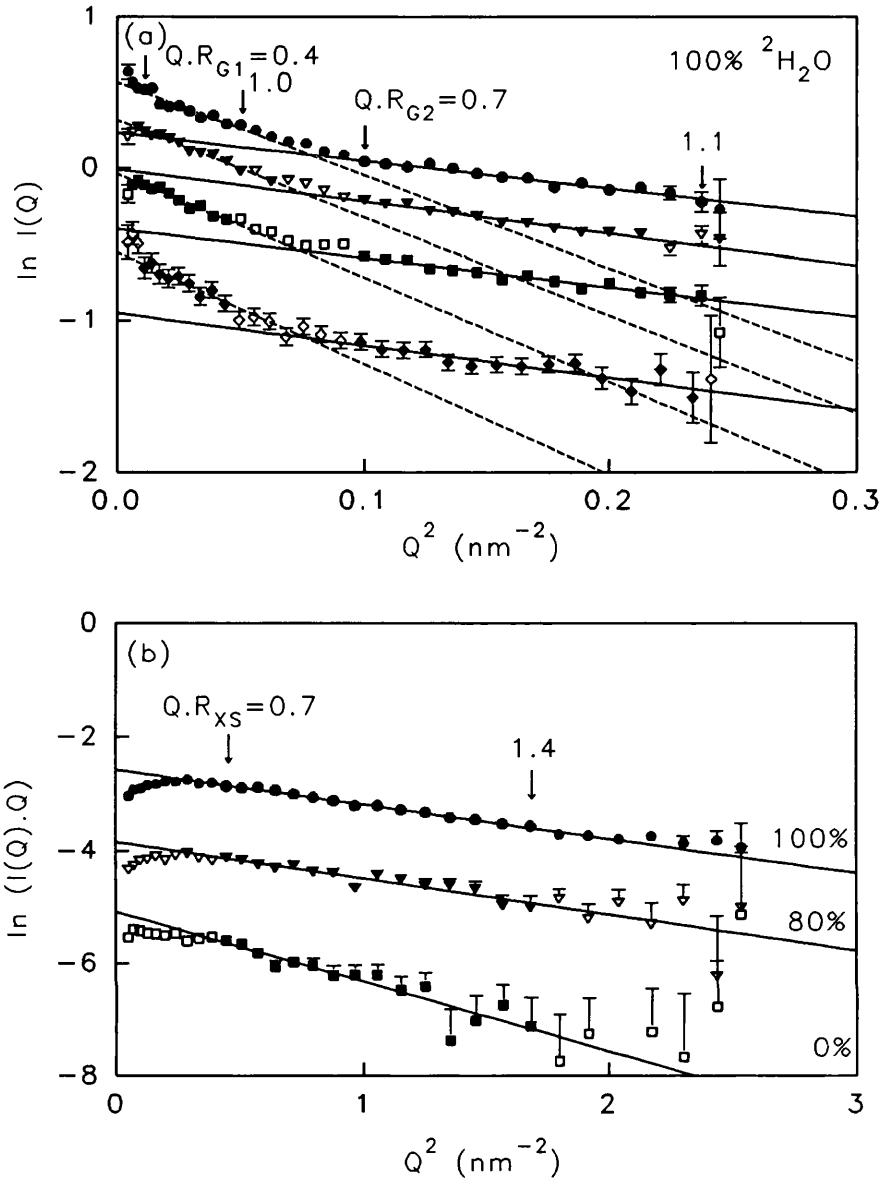


Figure 4.8. Neutron Guinier analyses for native $C\bar{l}$ inhibitor.

The $Q.R_{G1}$, $Q.R_{G2}$ and $Q.R_{XS}$ ranges are arrowed and indicated by filled symbols.

(a) A dilution series in 100% ^2H_2O buffers is shown. At low Q, the concentration dependence of the Guinier R_{G1} curves is visible. From top to bottom, the sample concentrations are 7.9 mg/ml (o), 5.9 mg/ml (▽), 4.1 mg/ml (□) and 2.2 mg/ml (◇). The R_{G1} values were analyzed over a Q range of 0.08 to 0.22 nm^{-1} and the R_{G2} values over a Q range of 0.32 to 0.48 nm^{-1}.

(b) Cross-sectional analyses are shown in 100%, 80% and 0% ^2H_2O buffers. The sample concentrations are 14.0 mg/ml (o), 12.7 mg/ml (▽) and 8.5 mg/ml (□) respectively. The R_{XS} dependence on contrast is clearly visible.

conditions were similar and no gross conformational change occurred upon cleavage.

At phosphate concentrations above 1.0 M, both the R_G and $I(0)/c$ parameters increased for native α_1-antitrypsin but remained unchanged for the split form. The $I(0)/c$ value for native α_1-antitrypsin had doubled by 1.8 M phosphate, suggesting the formation of dimers. The aggregation of native α_1-antitrypsin suggested that the properties of several surface amino acids had changed under high phosphate concentrations. This may be due to the insertion of the reactive loop into β-sheet A (Schulze *et al.*, 1990).

(4.3.3) Neutron analyses of $C\bar{1}$ inhibitor

Neutron Guinier and cross-sectional data for $C\bar{1}$ inhibitor were analyzed jointly with Dr S.J. Perkins and are summarized in Table 4.1. Both native and split data were analyzed (Section 4.2) from three independent sessions over a concentration range of 2 to 14 mg/ml. All the Guinier plots had a biphasic nature, and two R_G ranges were studied, R_{G1} and R_{G2} (Figure 4.8(a)). The region at lowest Q, R_{G1} corresponded to the R_G of the whole protein and was calculated between Q values of 0.08 and 0.22 nm^{-1} on D17. This was confirmed using a lower Q range on D11. The R_{G1} and $I(0)_1/c$ values for $C\bar{1}$ inhibitor showed a similar concentration dependence as for α_1-antitrypsin. This was also seen by sedimentation studies (Haupt *et al.*, 1970). To eliminate this concentration dependence the R_G and $I(0)_1/c$ were regressed to zero concentration. The M_r calculated from the $I(0)_1/c$ extrapolated to zero concentration in H_2O for both forms of $C\bar{1}$ inhibitor was 76,000 ± 4,000 which was compatible with value derived from the sequence of 71,100. The A_{280} used to calculate the concentration was 3.6 (Harrison, 1983; Salvesen *et al.*, 1985) which was in agreement with that determined from the amino acid and carbohydrate composition of 3.86 (Perkins, 1986) (Chapter 3). An A_{280} of 4.5 as reported by Haupt *et al.* (1970) and Bruch *et al.* (1988) would give a M_r (95,000 ± 5,000) much larger than the sequence value. The matchpoint calculated from Figure 4.9(a) was 42.3 ± 0.5 $\%^2H_2O$ which compared well with that determined from the composition of 41.0 $\%^2H_2O$.

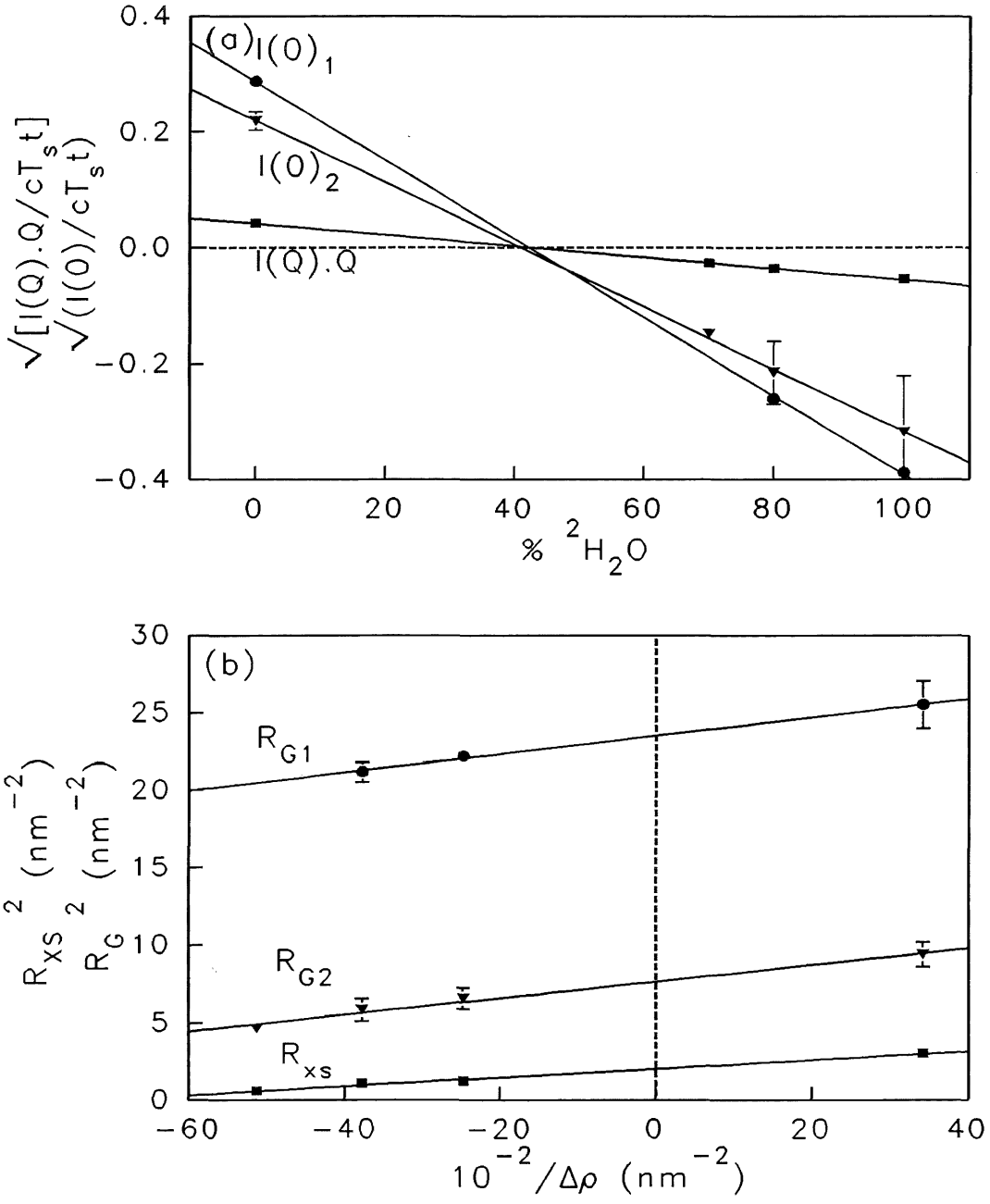


Figure 4.9. Contrast variation analyses for native and split $C\bar{1}$ inhibitor.

(a) The matchpoint determination of $C\bar{1}$ inhibitor, using the $I(0)/cT_st$ data extrapolated to zero concentration, gives 42.3% ^2H_2O and 41.8% ^2H_2O for $I(0)_1$ and $I(0)_1$ respectively. The matchpoint using the mean $[I(Q).Q]_{Q\to 0}$ values gives 42.2% ^2H_2O.

(b) The Stuhrmann plot gives an R_{G1} and R_{G2} values of 4.85 ± 0.01 nm and 2.78 ± 0.03 nm and slopes α of 59 ± 3 x 10^{-5} and 52 ± 5 x 10^{-5} respectively. The Stuhrmann plot of the mean R_{XS}^2 values gives an $R_{XS\text{-}C}$ of 1.43 ± 0.02 nm and an α of 29 ± 2 x 10^{-5}.

The R_{G1} values for the native and split forms of $C\bar{1}$ inhibitor were similar and indicate that, like α_1-antitrypsin, no large conformation change occurred upon proteolytic cleavage. The $R_{G1\text{-}C}$ determined from the Stuhrmann plot gave an elongation ratio $R_{G1\text{-}C}/R_O$ of 2.27 (Figure 4.9(b)). This was much larger than the value of 1.35 for α_1-antitrypsin and the average of 1.28 for globular proteins (Perkins, 1986) and indicated that $C\bar{1}$ inhibitor was highly elongated in solution. This elongation was attributable to the N-terminal domain since the serpin domain was assumed to have similar tertiary structure as α_1-antitrypsin. The large positive slope of the Stuhrmann plot, α_{G1} of 59 ± 3 x 10^{-5} was indicative of a surface region dominated by hydrophilic amino acids and carbohydrates and an inner core of hydrophobic amino acids.

The R_{G2} data were measured over a Q range of either 0.32 to 0.48 nm^{-1} at 3.46m on instrument D17, or a Q range of 0.32 to 0.63 nm^{-1} at 1.40m on instrument D17 (Figure 4.8(a)). No difference was found for the native or split forms and no dependence on protein concentration was observed for either the R_{G2} or $I(0)_2/c$ parameters. The mean M_r calculated from the $I(0)_2/c$ value in H_2O (13 values) was 46,000 ± 6,000. The matchpoint was 41.8 ± 0.3 $\%^2H_2O$ (Figure 4.9(a)) and the Stuhrmann $R_{G2\text{-}C}$ and α_2 (Figure 4.9(b)) were determined as 2.78 ± 0.03 nm and 52 ± 5 x 10^{-5} respectively. Although these values were comparable with those determined for α_1-antitrypsin (M_r of 51,500, matchpoint of 41.3 ± 1.0 $\%^2H_2O$, $R_{G\text{-}C}$ of 2.61 ± 0.02 nm and α of 29 ± 5 x 10^{-5}, respectively), they cannot be attributed to solely the serpin domain of $C\bar{1}$ inhibitor but represent a convolution of scattering from the serpin and N-terminal domains.

Since $C\bar{1}$ inhibitor was elongated, cross-sectional R_{XS} analyses could be performed (Pilz, 1982). Linear plots of ln $I(Q).Q$ against Q^2 were obtained over the Q range of 0.67 to 1.30 nm^{-1} (Figure 4.8(b)). Again, no protein concentration dependence was observed for the R_{XS} or $[I(Q).Q]_{Q\to 0}/c$ parameters and there was no difference in these values for the native and split forms. The matchpoint determined from Figure 4.9(a) was 42.2 ± 0.5 $\%^2H_2O$ which compares well with that of 41.0 $\%^2H_2O$ predicted from the sequence. The Stuhrmann plot of R_{XS}^2 against $1/\Delta\rho$ (Figure 4.9(b)) gave an $R_{XS\text{-}C}$ of 1.43 ± 0.02 nm and an α_{XS} of

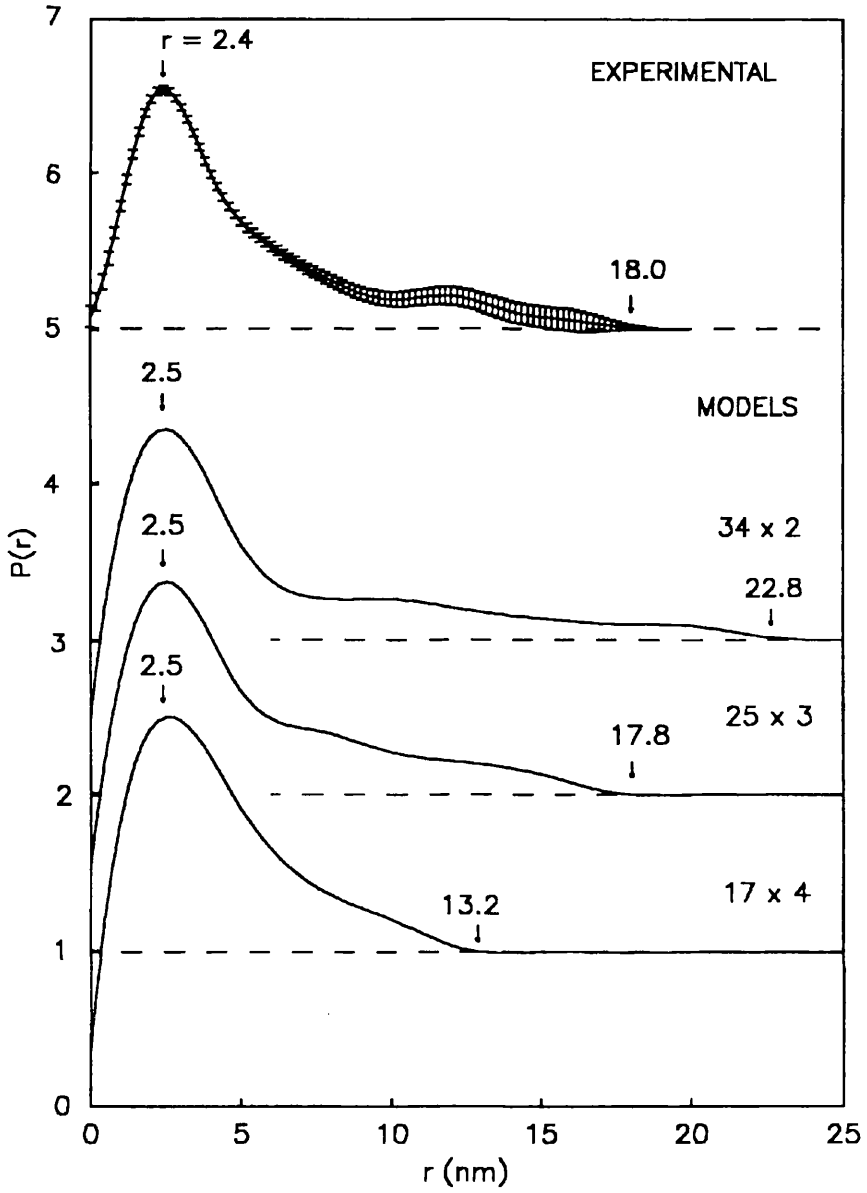


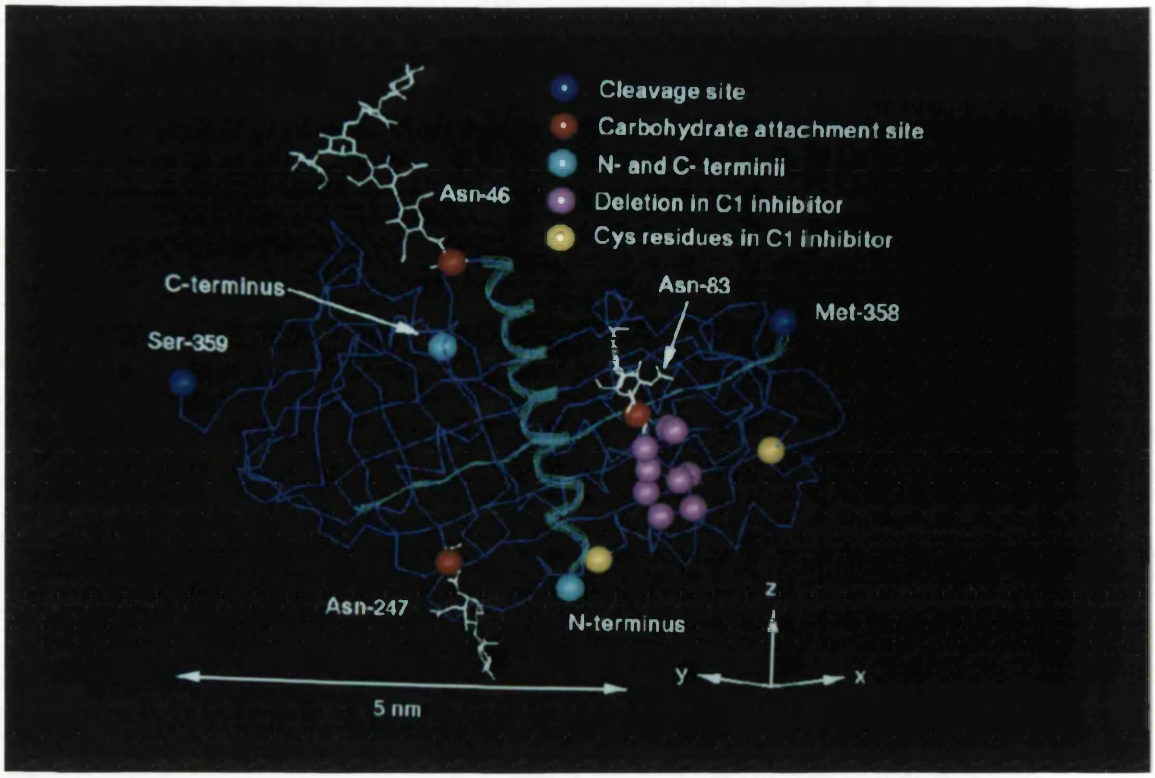
Figure 4.10. Indirect transformation of the scattering curves for $C\bar{1}$ inhibitor. The experimental $P(r)$ curve corresponds to data on native $C\bar{1}$ inhibitor at 14 mg/ml in 100% ^2H_2O, based on 69 $I(Q)$ points extending to Q = 3.1 nm^{-1}, fitted using 10 B-splines, and transformed using a D_{max} of 20.2 nm. The errors of the transformation are shown for the experimental curve. The three model curves are based on those of Figure 4.15 (N-terminus domain length 25 x 3 spheres) and Figure 4.17(a) (34 x 2 and 17 x 4) as calculated for the 100% ^2H_2O contrast, based on 99 I(Q) points extending to Q = 1.6 nm^{-1}, fitted using 10 B-splines, and transformed using a D_{max} of 25.25 nm. The r values corresponding to the positions of the main peak in $P(r)$ and where $P(r)$ becomes zero are arrowed.

$29 \pm 2 \times 10^{-5}$. This unusually high contrast dependence was clearly visible in Figure 4.8(b). The cross-sectional structure of $C\bar{1}$ inhibitor was also strongly influenced by the carbohydrate component on the surface of the molecule.

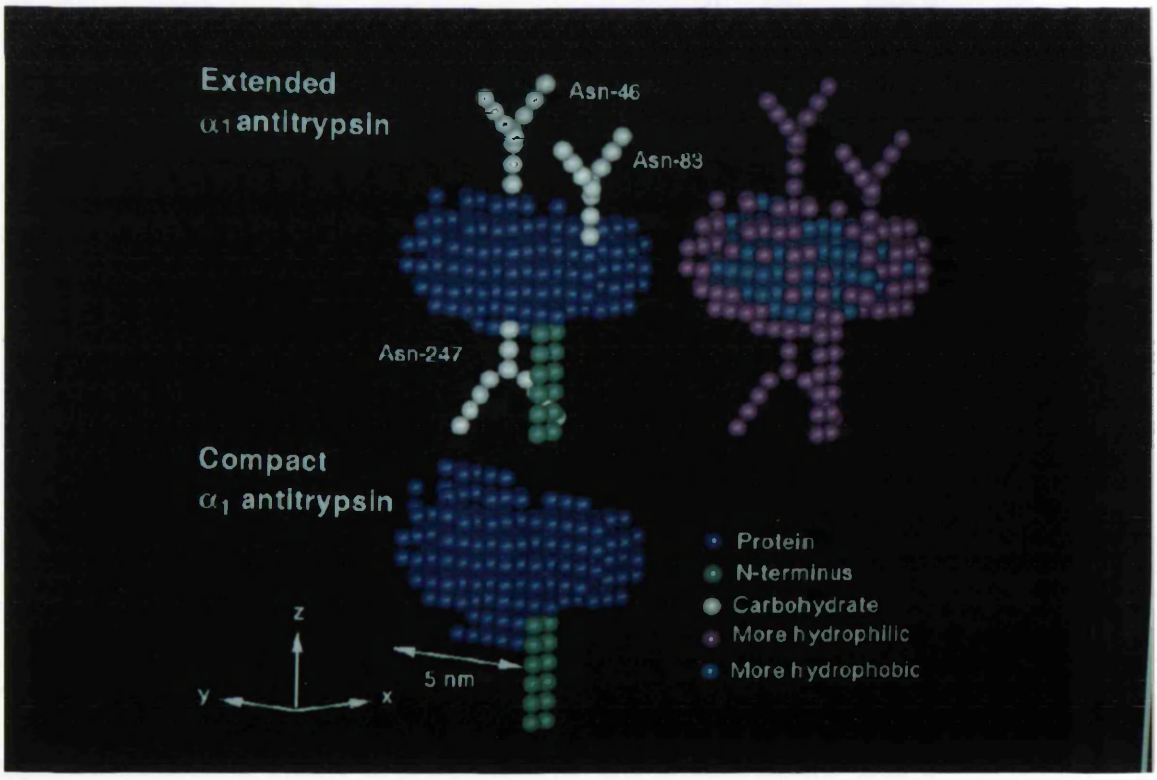
The combination of the R_{GI} data and the R_{XS} values gave estimates of the particle length (Chapter 3). The length calculated from the intensities ratio $\pi.I(0)/[I(Q).Q]_{Q\to 0}$ in 0%, 80% and 100% ^2H_2O gave values of 15.7 nm, 16.1 nm and 15.0 nm respectively. The length calculated assuming an elliptical cylinder, $L^2 = 12(R_G^2 - R_{XS}^2)$ for 0%, 80% and 100% ^2H_2O data gave lengths of 16.5 nm, 15.4 nm and 15.8 nm. The mean length was 15.8 ± 0.5 nm (Table 4.1). The indirect transformation of the scattering curve $I(Q)$ into the distance distribution function $P(r)$ in real space gave another calculation of the length since $P(r)$ becomes zero at the maximum distance within the macromolecule (Chapter 2). This had the advantage over the Guinier analyses in that the full curve to $Q = 3.1$ nm^{-1} was used, although the low intensities at high Q had the highest errors. Experimental scattering curves for both the native and the split forms in 100 % ^2H_2O were analyzed. The most satisfactory $P(r)$ calculations had a maximum $P(r)$ at r of 2.4 nm (Figure 4.10). The R_G value for native $C\bar{1}$ inhibitor (concentration 14.1 mg/ml) was 4.35 nm (stable over a range of Lagrange multipliers of 1.00 to -8.00) and that for split $C\bar{1}$ inhibitor (concentration 8.47 mg/ml) was 4.67 nm (stable over a range of Lagrange multipliers of 0.00 to -8.00). These compare reasonably to the experimental values of corresponding concentration in 100% ^2H_2O of 4.02 nm and 4.29 nm. The maximum distance was determined as 16 to 19 nm which is in agreement with the Guinier estimate of L of 15.8 ± 0.5 nm (Table 4.1).

(4.3.4) Debye sphere modelling of α_1-antitrypsin

Molecular models were constructed for α_1-antitrypsin and $C\bar{1}$ inhibitor to interpret the scattering curve analyses. The crystal structure of split α_1-antitrypsin (Löbermann *et al.*, 1984) was converted into a Debye sphere model for scattering curve simulations (Section 4.2, Chapters 2 & 3) and compared with the solution scattering data. Only 374 of the 394 amino acid residues were present in the crystal structure, since Gly1 to Asp19 and Lys394 are not visible in the


Cleavage site
Carbohydrate attachment site
N- and C- terminii
Deletion in C1 inhibitor
Cys residues in C1 inhibitor
Asn-46
C-terminus
Ser-359
Asn-83
Met-358
Asn-247
N-terminus
z
y
x
5 nm

(a)


Extended
α1 antitrypsin
Asn-46
Asn-83
Asn-247
Compact
α1 antitrypsin
Protein
N-terminus
Carbohydrate
More hydrophilic
More hydrophobic
z
y
x
5 nm

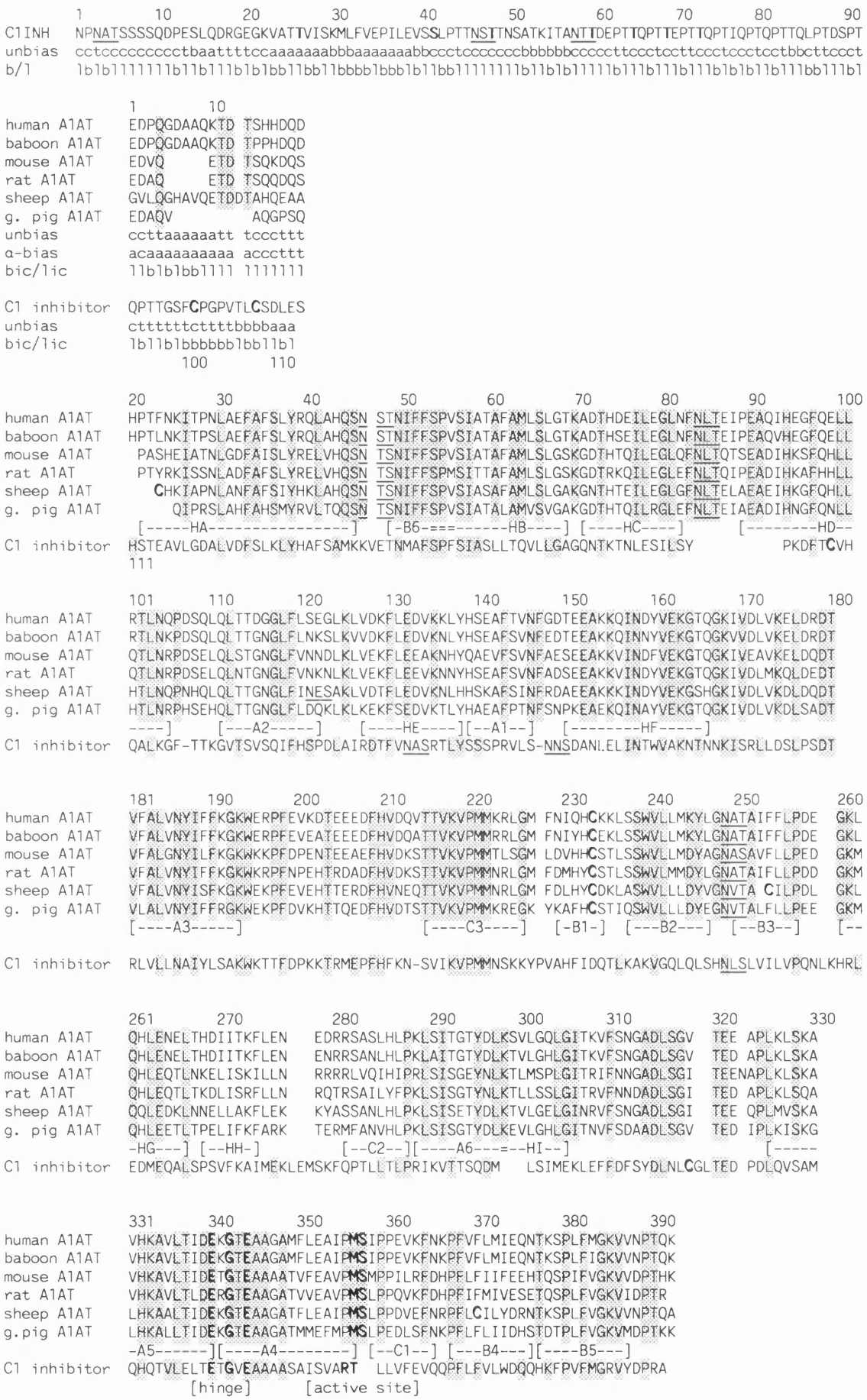
(b)

electron density map. The crystal co-ordinates were subdivided into 246 cubes of side 0.608 nm to obtain an equivalent unhydrated protein volume. A cube of this size represents a single oligosaccharide residue (Perkins *et al.*, 1985). The outermost 163 spheres (hydrophilic) were assigned a scattering density of 60% ^2H_2O and the innermost 128 spheres (hydrophobic) one of 30% ^2H_2O (Figure 11(b)) (Section 4.2). Both tetragonal and hexagonal split α_1-antitrypsin co-ordinates were taken from the Protein Data Bank. After adding hydrogen atom co-ordinates (Perkins, 1982), and using a unit weighting of atoms, the R_G of these structures was calculated as 2.12 nm and 2.13 nm (Chapter 3). This agreed with the R_G of 2.15 nm calculated from the sphere model. The two co-ordinate sets gave very similar results and only the tetragonal model was used further.

Figure 4.11. Crystal structure and sphere models of α_1-antitrypsin.

(a) The α-carbon co-ordinates of tetragonal split α_1-antitrypsin (Löbermann *et al.*, 1984) are shown. The three N-linked carbohydrate attachment sites (Asn-46, Asn-83, Asn-247) are highlighted in red and the carbohydrates are shown in white. The N- and C-termini are highlighted in cyan. The cleaved reactive-site, Met-358 and Ser-359 are coloured in dark blue and the proposed 16 residues of the β-strand involved in the transition from native to split forms are shown as a green ribbon. The N-terminal residues 21 to 45 form an α-helical structure, coloured as a green ribbon. The 10 residue deletion in $C\bar{1}$ inhibitor compared to α_1-antitrypsin are indicated in purple and includes the Asn-83 carbohydrate attachment site. The residues homologous to the two cysteine residues in $C\bar{1}$ inhibitor which bind to the N-terminal domain are highlighted in yellow.

(b) The extended and compact Debye sphere models for native α_1-antitrypsin in the same orientation as the crystal structure. These give rise to the scattering curves (1) and (2) in Figure 4.13(a), 4.13(b) and 4.13(c). The three carbohydrates (coloured white) are assumed to be biantennary and either extended away from the surface of the protein or compact, alongside it. The N-terminus (coloured green) is positioned as an extension of the previous 26 residues in the $-z$ direction. The assignment of hydrophilic (magenta) and hydrophobic (cyan) spheres to the extended carbohydrate model is also shown.


C1INH
unbias
b/l
human A1AT
baboon A1AT
mouse A1AT
rat A1AT
sheep A1AT
g. pig A1AT
unbias
α-bias
bic/lic
C1 inhibitor
unbias
bic/lic
human A1AT
baboon A1AT
mouse A1AT
rat A1AT
sheep A1AT
g. pig A1AT
C1 inhibitor
human A1AT
baboon A1AT
mouse A1AT
rat A1AT
sheep A1AT
g. pig A1AT
C1 inhibitor
human A1AT
baboon A1AT
mouse A1AT
rat A1AT
sheep A1AT
g. pig A1AT
C1 inhibitor
human A1AT
baboon A1AT
mouse A1AT
rat A1AT
sheep A1AT
g.pig A1AT
C1 inhibitor
[hinge]
[active site]

N-terminal and carbohydrate residues were added to the model. The 19 N-terminal residues were represented by 12 cubes positioned close to His20. Examination of the crystal structure showed that residues 21 to 45 form an α-helix running in the $-z$ direction (Figure 4.11(a)) (Löbermann *et al.*, 1984). Robson and Chou-Fasman secondary structure predictions (Garnier *et al.*, 1978; Chou & Fasman, 1978) suggested that 6 to 11 of the missing residues also constituted an α-helix, of length 1 to 2 nm (Figure 4.12). The 19 residues were assumed to be fully solvent exposed since they contain up to 10 charged groups and this amino-terminal section was readily cleaved by a number of proteolytic enzymes (Section 4.2). The N-terminus was thus modelled as a 6 x 2 x 1 array of spheres of length 3.6 nm extending away from the protein surface in the $-z$ direction. This increased the R_G by 0.04 nm. The three N-linked oligosaccharides at Asn46, Asn83 and Asn247 were assumed to be biantennary with 11 residues each (Vaughan

---

Figure 4.12. Alignment of the sequences of α_1-antitrypsin and $C\bar{1}$ inhibitor. The amino acid sequences for human (Carrell *et al.*, 1982), mouse (Hill *et al.*, 1984), baboon (Kurachi *et al.*, 1981), rat (Chao *et al.*, 1990), sheep (Brown *et al.*, 1989) and guinea pig (Suzuki *et al.*, 1991) α_1-antitrypsin are aligned, together with human $C\bar{1}$ inhibitor (Bock *et al.*, 1986). The α-helices A to I (Löbermann *et al.*, 1984) in the crystal structure of α_1-antitrypsin are denoted HA to HI, while the strands in the three β-sheets A, B and C are denoted A1 to A6, B1 to B6, and C1 to C3. N-linked carbohydrate sites are located at Asn46, Asn83 and Asn247 in α_1-antitrypsin, and at Asn216, Asn231 and Asn330 in $C\bar{1}$ inhibitor and are underlined. Residues which are 100% conserved in all six α_1-antitrypsin sequences are shaded. Residues which are conserved between human α_1-antitrypsin and C1 inhibitor are also shaded. Cys residues, the active site (Met358 - Ser359) for the α_1-antitrypsin and (Arg44 - Thr445) for $C\bar{1}$ inhibitor and the conserved "hinge" region (E-X-G-X-E) are highlighted. The secondary structure predictions for the 112 N-terminal residues of $C\bar{1}$ inhibitor (unbiased) and N-terminal 20 residues of α_1-antitrypsin (unbiased and α-biased) are shown (a, α-helix; b, β-strand; c, coil; t, turn). The hydrophobicity (b) or hydrophilicity (l) propensity of these residues is also indicated according to the Eisenberg (1984) scale.

et al., 1982). The carbohydrate residues were modelled in an extended configuration as indicated by the Guinier results and the crystal structure (Figure 4.11(a)). The carbohydrate residues were only partially visible in the electron density map. This was due to packing disorder in the crystal structure and only residues close to the protein surface had a sufficient electron density to be defined (Engh *et al.*, 1989). Each oligosaccharide was thus modelled by 11 cubes in an extended arrangement (Figure 4.11(b)). An alternative configuration is that the oligosaccharide chains are folded against the protein surface. Such a model (Figure 4.11(b)) has an R_G that is 0.22 nm smaller than that of the extended structure.

A model for native α_1-antitrypsin required the rejoining of Met358 and Ser359 of split α_1-antitrypsin (Figures 4.1 & 4.11). Löbermann *et al.* (1984) proposed that residues 359-394 were fixed in the two structures, and residues 343-358 have been relocated (Figure 4.1). This was supported by the susceptibility of peptide bonds around the reactive site to proteolytic attack by several proteases of α_1-antitrypsin and $C\bar{1}$ inhibitor. Inspection of the α_1-antitrypsin crystal structure suggested that the minimum conformational change required to form native α_1-antitrypsin occurred between residues 348-358. Since these residues corresponded to 7 cubes those cubes closest to residues 348-358 were relocated to rejoin Met358 and Ser359 on the protein surface. This relocation was supported by the subsequent crystal structure for native ovalbumin which showed the reactive site to be situated on a surface loop (Stein *et al.*, 1990). The R_G of the native α_1-antitrypsin model was increased by 0.03 nm, which was not detectable experimentally by this technique.

The curve fitting to the neutron and X-ray data for native α_1-antitrypsin in three contrasts is summarised by Curves 1 and 2 in Figure 4.13. The model with extended carbohydrate chains gave the three good Curve 1 fits for native α_1-antitrypsin. The residual R were 0.006, 0.017 and 0.011 respectively for (a), (b) and (c) in Figure 4.13. The calculations based on the compact carbohydrate model gave worsened curve fits as shown by Curves 2 in Figure 4.13 and the R of 0.010, 0.019 and 0.028. This was clearest in Figure 4.13(c) since a positive solute-solvent

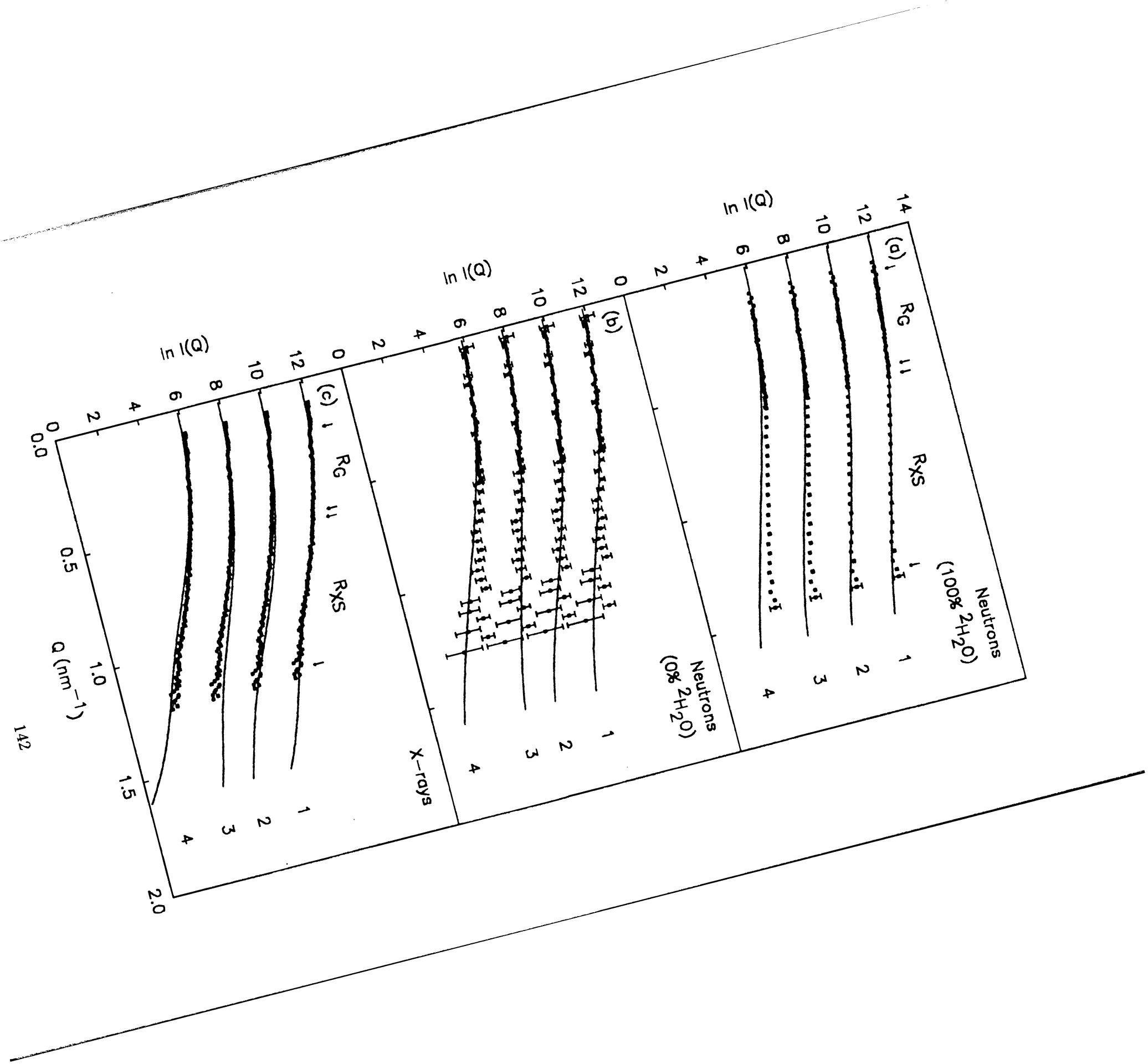

(a)
(b)
(c)
R_G I I
R_XS
Neutrons
(100% ^2H_2O)
Neutrons
(0% ^2H_2O)
X-rays
ln I(Q)
Q (nm^{-1})
0.0
0.5
1.0
1.5
2.0
0
2
4
6
8
10
12
14
1
2
3
4

Figure 4.13. Comparisons of the curve simulations with the α_1-antitrypsin models.

(a) Neutron data in 0.05 M sodium/potassium phosphate buffer in 100% ^2H_2O for native α_1-antitrypsin are compared with the smeared simulated curves (see Chapter 3 for empirical smearing parameters). The vertical arrows indicate the Q ranges of 0.17 to 0.57 nm^{-1} and 0.61 to 1.42 nm^{-1} used for the R_G and R_{XS} determinations (Figures 4.2(a) and 4.2(b)) respectively.

Curve 1: the unhydrated native α_1-antitrypsin structure with extended oligosaccharide and the use of two densities; R of 0.006 (Section 4.2).

Curve 2: unhydrated native α_1-antitrypsin structure with compact oligosaccharide and two densities; R of 0.010.

Curve 3: unhydrated native α_1-antitrypsin with extended oligosaccharide and use of a single density; R of 0.020.

Curve 4: native α_1-antitrypsin with two densities and a hydration of 0.3 g H_2O / g α_1-antitrypsin; R of 0.024.

(b) Neutron data in 0.012 M-sodium/potassium phosphate buffer in H_2O for native α_1-antitrypsin are compared with the four simulated curves numbered as in (a) but recalculated for this contrast. The R for curves 1, 2, 3 and 4 are 0.017, 0.019, 0.017 and 0.029 respectively.

(c) X-ray data in 0.05 M-sodium/potassium phosphate buffer for native α_1-antitrypsin are compared with unsmeared simulated curves. The vertical arrows indicate the Q ranges of 0.33-0.57 nm^{-1} and 0.60-1.20 nm^{-1} used for the R_G and R_{XS} determinations (Figures 4.2(c) and 4.2(d)).

Curve 1: native α_1-antitrypsin, hydrated to 0.3 g H_2O /g α_1-antitrypsin, with extended oligosaccharide and two densities; R of 0.011.

Curve 2: native α_1-antitrypsin, hydrated to 0.3 g H_2O /g α_1-antitrypsin, with compact oligosaccharide and two densities; R of 0.028.

Curve 3: unhydrated native α_1-antitrypsin with extended oligosaccharide and two densities; R of 0.035.

Curve 4: native α_1-antitrypsin, hydrated to 0.67 g H_2O /g α_1-antitrypsin, with extended oligosaccharide and two densities; R of 0.028.

contrast corresponded to the strongest contribution of carbohydrate to the scattering curve. X-ray scattering curves measured in H_2O resulted in better signal-noise ratios at large Q than neutrons in the same contrast (Figure 4.13(c)). The R_{G-C} of the extended carbohydrate model (Curve 1) was 2.58 nm which agreed well with 2.61 nm ± 0.02 nm from the experimental determination (Table 4.1). The compact carbohydrate model (Curve 2) gave worsened agreement with an R_{G-C} of 2.36 nm. The α of the extended model was 29 x 10^{-5} which was in better agreement with the α of 29 ± 5 x 10^{-5} by experiment (Table 4.1) than that of the compact model with an α of 20 x 10^{-5}. The density-weighted X-ray R_G of the extended model was 2.97 nm in good agreement with the experimental X-ray values of 2.80 to 3.00 nm (Figure 4.6). It was concluded that the solution structure of native α_1-antitrypsin was similar to that seen in the crystal structure if the carbohydrate and N-terminal moieties were in extended conformations.

Two sets of control calculations were performed to indicate the importance of incorporating the internal scattering density and hydration into these models.

(a) Given the large α values, it was necessary to assign different scattering densities to the hydrophilic amino acid/carbohydrate and hydrophobic amino acid components in the neutron simulations (Section 4.2). If this was neglected the simulated curves at large Q would underestimate the neutron $I(Q)$ data in negative contrasts in ^2H_2O (Curve 3 of Figure 4.13(a); R of 0.025) and overestimate those in positive contrasts in H_2O (Curve 3 of Figure 4.13(b); R of 0.017).

(b) X-ray simulations required an hydration shell of 0.3 g H_2O/g α_1-antitrypsin in addition to different internal electron densities. Calculations for a model without hydration gave curves that generally overestimated the X-ray $I(Q)$ data (Curve 3 of Figure 4.13(c); R of 0.036). Too great a hydration of 0.67 g H_2O/g α_1-antitrypsin lead to worsened agreements beyond Q = 1 nm^{-1} (Curve 4 of Figure 4.13(c); R of 0.030). The poorer fits of Curves 4 in Figures 4.13(a) and 4.13(b) (R of 0.037 and 0.026 respectively) showed that the accommodation of an hydration shell was not appropriate for neutron curve modelling.

As an independent test of the α_1-antitrypsin solution models, hydrodynamic models were derived from the crystal structure in order to calculate the frictional

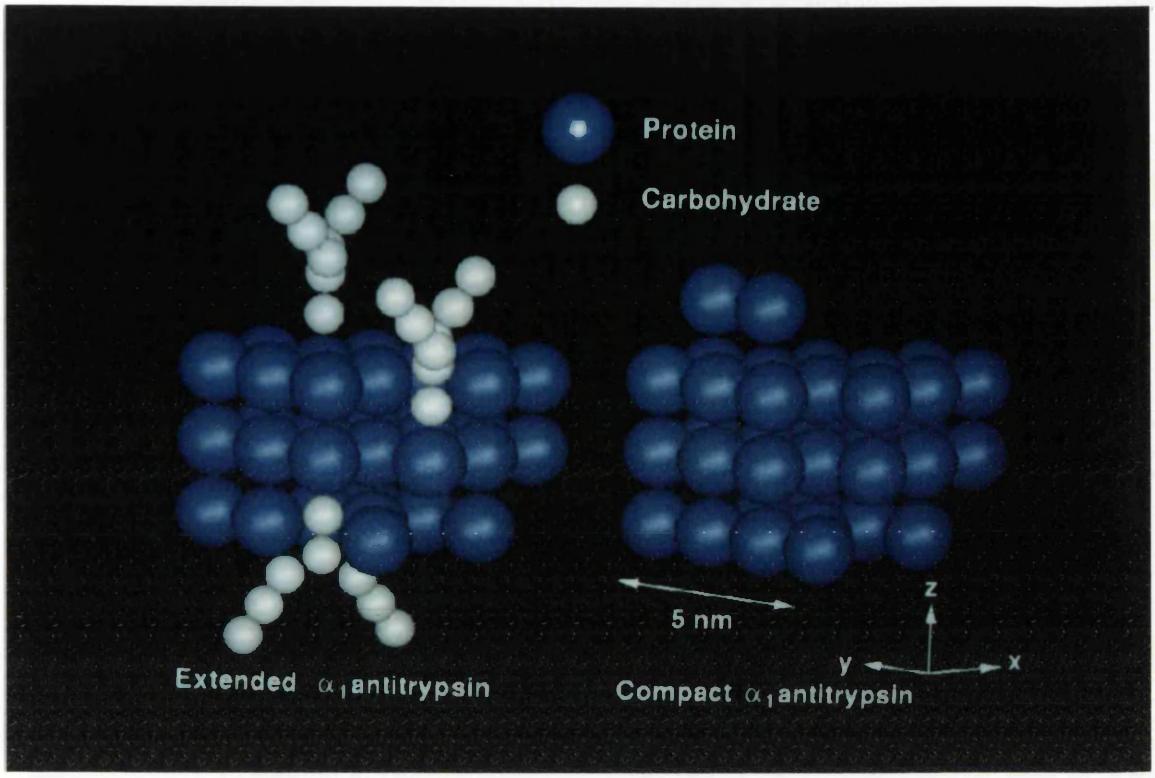


Figure 4.14. Hydrodynamic models for α_1-antitrypsin

The hydrodynamic models of native α_1-antitrypsin with carbohydrates (white) in extended and compact conformation are shown. The protein spheres are 1.69 nm in diameter and the carbohydrate spheres are of diameter 0.95 nm.

coefficient (Section 4.2). The sedimentation coefficient $s^o_{20,w}$ of α_1-antitrypsin had been reported as 3.41 S, 3.3 S and 3.45 S (Bundy & Mehl, 1959; Shamash & Rimon, 1966; Kress & Laskowski, 1973). The frictional coefficient was thus 6.7 to 7.0 x 10^{-8} g/cm/s. From this the frictional ratio f/f_o was calculated as 1.29 to 1.35 to show that this had a moderately elongated structure, in agreement with the conclusion from the R_G/R_O ratio of 1.35. The 31 hydrodynamic spheres for the protein core were of diameter 1.69 nm. One more was added to allow for the N-terminus. Each carbohydrate moiety was represented by 8 spheres of diameter 0.95 nm in the extended model (Figure 4.14). In the compact model, these were replaced by 4 spheres of diameter 1.69 nm. The frictional coefficients were 7.2

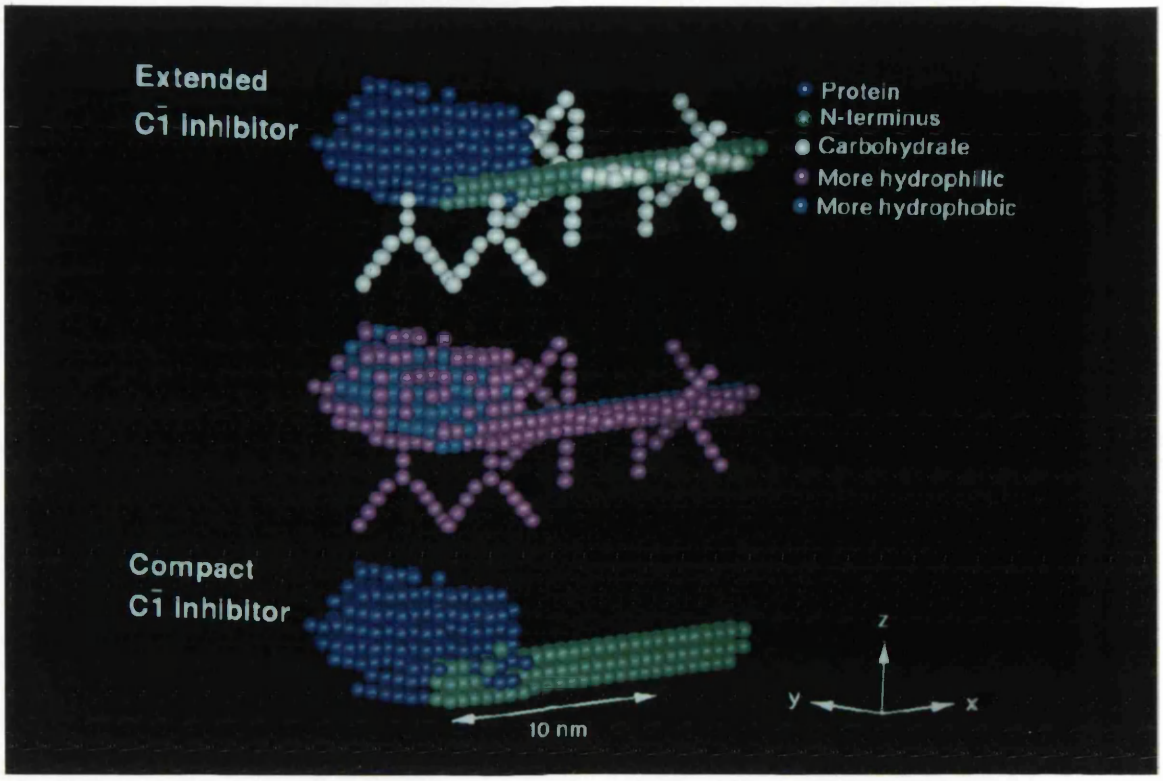


Figure 4.15. Final Debye sphere models of $C\bar{1}$ inhibitor.

Models of $C\bar{1}$ inhibitor with extended and compact carbohydrate structures (white). The N-terminus domain is extended in the $+x$ direction and is 25 spheres in length and 3 spheres in cross-section (green). The assignment of hydrophilic (magenta) and hydrophobic (cyan) spheres to the extended carbohydrate model is also shown. The orientation is the same as that for the crystal co-ordinates of α_1-antitrypsin.

and 6.4×10^{-8} g/cm/s for the extended and compact models, which were greater than and less than the estimated experimental range, respectively. This suggested that the carbohydrate of α_1-antitrypsin existed in a range of extended conformations, and was consistent with the Debye modelling.

(4.3.5) Debye sphere modelling of $C\bar{1}$ inhibitor

Debye sphere models of $C\bar{1}$ inhibitor were based on the crystal structure of split α_1-antitrypsin (Löbermann *et al.*, 1984; Bernstein *et al.*, 1977). The α_1-antitrypsin secondary structure was homologous to the $C\bar{1}$ inhibitor sequence if residues Asn83 to Ile92 were deleted (Figure 4.12; Huber & Carrell, 1989). The deletion corresponded to a surface loop between Helix C and D, and included one of the three oligosaccharide sites in α_1-antitrypsin at Asn83. The 264 residues visible in the crystal structure were converted into 237 spheres of diameter 0.754 nm based on cubes of side 0.608 nm (Figure 4.15). In a similar way as the α_1-antitrypsin model, the model was subdivided into 243 surface hydrophilic spheres and 152 core hydrophobic spheres. These were assigned the same matchpoints as for α_1-antitrypsin of 60% and 30% ^2H_2O respectively.

The N-terminal 19 residues missing in the α_1-antitrypsin crystal structure were replaced by the N-terminal domain (Asn1-Ser113) in $C\bar{1}$ inhibitor (Figure 4.12). This domain was modelled as an extended structure of 69 spheres with an initial total length of 15.2 nm, corresponding to 25 spheres. The reasons for this are summarized below:

(a) Scattering showed that the overall length of α_1-antitrypsin of 7.0 to 7.8 nm was increased to 16 to 19 nm in $C\bar{1}$ inhibitor;

(b) An unbiased Robson secondary structure prediction (Garnier *et al.*, 1978) of the N-terminal 112 residues indicated relatively low amounts of α-helix (17%) and β-sheet (16%), and high amounts of β-turn (25%) and coil (41%) (Figure 4.12) which is consistent with an extended structure;

(c) Of the 112 residues, 61% are hydrophilic (Eisenberg, 1984) and there are 10 glycosylation sites, which was compatible with a high exposure to solvent;

(d) Of the 112 residues 15 are proline (13.4%), mostly between Pro62 and Pro104,

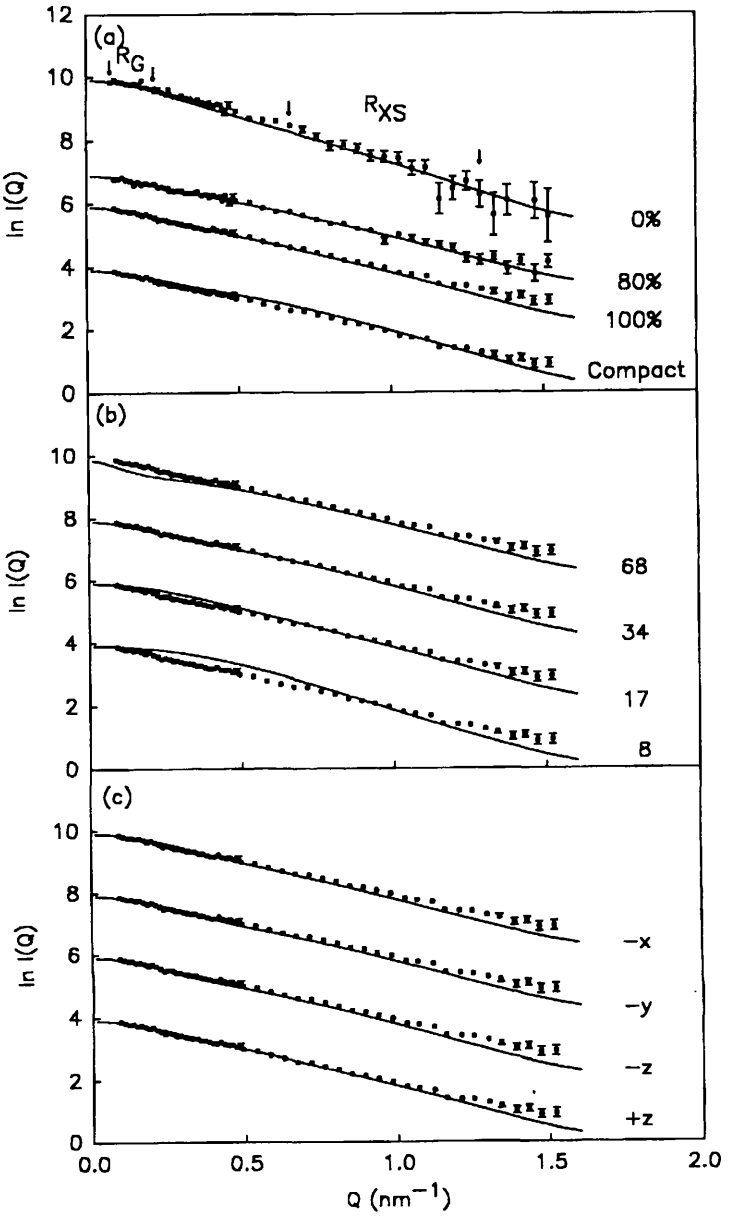


Figure 4.16. Curve fitting for the Debye sphere models of $C\bar{1}$ inhibitor.

(a) Comparison of the simulated scattering curves for the final $C\bar{1}$ inhibitor model with extended carbohydrates shown in Figure 4.15 with scattering curves of native $C\bar{1}$ inhibitor in 0%, 80% and 100% ^2H_2O (see text) and the compact model with the experimental scattering curve in 100% ^2H_2O.

(b) Simulated scattering curves for four $+x$ models in which the N-terminal domain is of length 68, 34, 17 and 8 spheres (models shown in Figure 4.17(a)). These are compared with the native $C\bar{1}$ inhibitor scattering curve in 100% ^2H_2O.

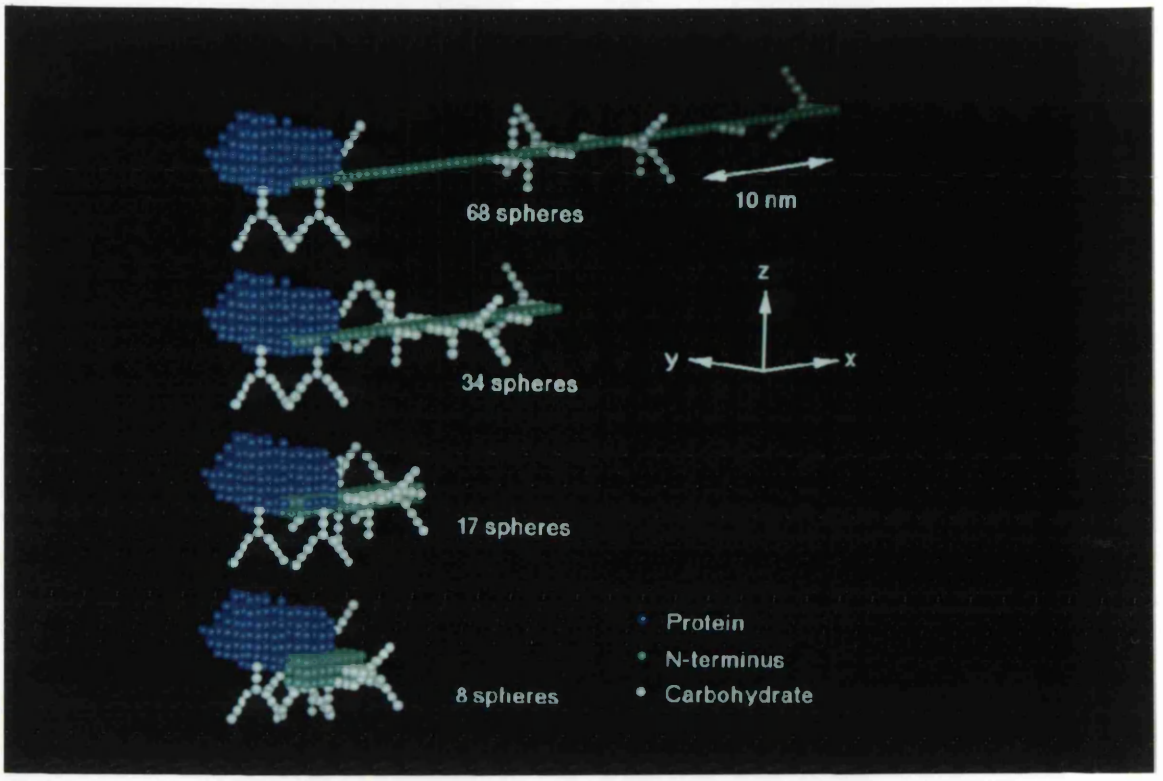
(c) Simulated scattering curves for models in which the N-terminal domain is in the $-x$, $-y$, $-z$ and $+z$ orientations (models shown in Figure 4.17(b)).

which is high compared to an average of 5.2% in proteins (Dayhoff, 1978). The N-terminus may adopt a polyproline II helix, which with a rise of 0.31 nm per residue, would have a length of 13.3 nm over 43 residues (Huber & Carrell, 1989). (e) Several proteases cleave peptide bonds in the N-terminus domain, indicating that these residues lie on the surface (Section 4.2).

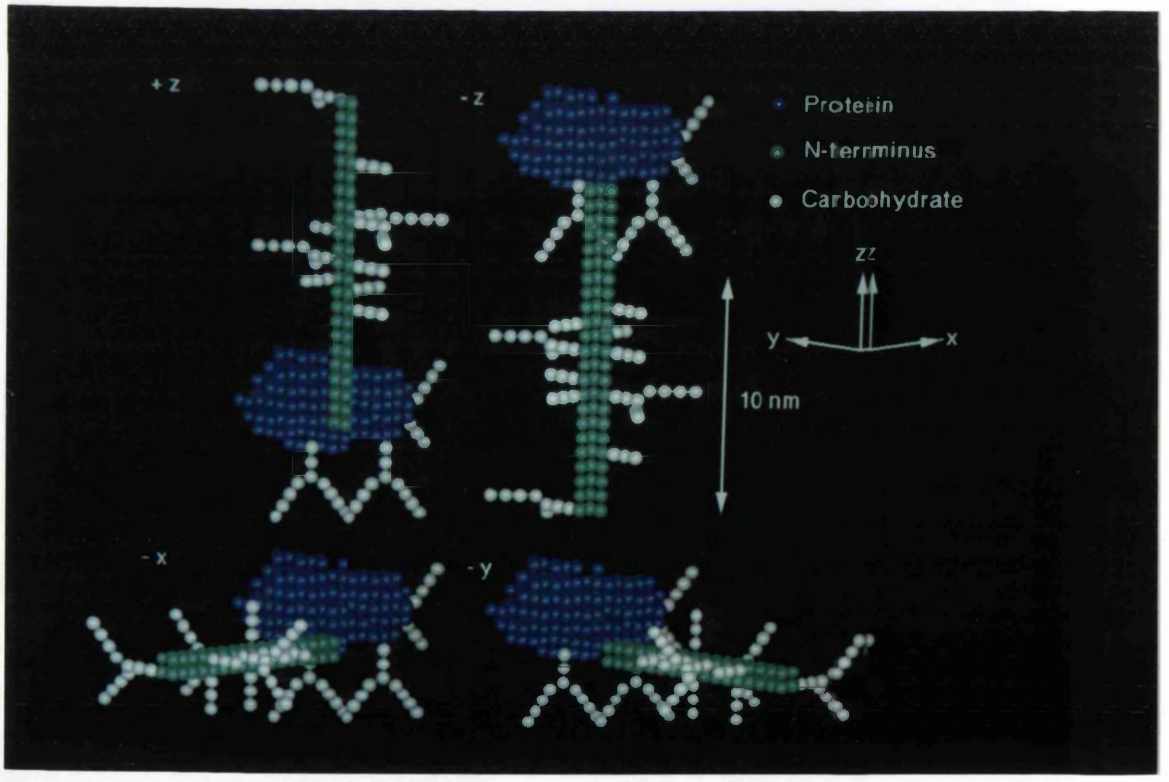
The thirteen oligosaccharides were represented by extended structures, as justified by the large Stuhrmann α. The six N-linked oligosaccharide groups were modelled by biantennary structures of 11 spheres each (Strecker *et al.*, 1985). The seven O-linked oligosaccharides were represented by 3 spheres each. The three N-linked and seven O-linked oligosaccharides of the N-terminal domain were arbitrarily positioned along the 25-sphere length in the same ratio as their occurrence in the primary sequence.

The relative orientation of the two domains in $C\bar{1}$ inhibitor was determined from the crystal structure of α_1-antitrypsin. Cys101 and Cys108 in the N-terminal domain of $C\bar{1}$ inhibitor form disulphide bridges with Cys406 and Cys183 in the serpin domain (Bock *et al.*, 1986). These correspond to residues 7, 14, 321, and 98 respectively of α_1-antitrypsin. This disulphide bridging pattern constrained the positioning of the N-terminus domain to run alongside the serpin domain in the $+x$ direction (Figure 4.11(a)), starting from His20 (α_1-antitrypsin numbering) (the $+x$ model). Some rearrangement of the spheres of the protein core of the N-terminal domain was required to avoid steric conflict with the serpin domain, including the deletion of 6 spheres to conserve the correct volume. The long axes of the two domains lie in similar directions.

The curve simulations are compared with the experimental data in Figure 4.16(a) and the corresponding models are shown in Figure 4.15. The simulations agreed well with experiment. The residuals R were low at 0.020 (0% ^2H_2O), 0.011 (80%) and 0.012 (100%). The R_G in the three contrasts were 5.03 nm, 4.61 nm and 4.71 nm, in good agreement with the experimental values of 5.05 nm, 4.60 nm and 4.71 nm (Table 4.1). The Stuhrmann parameters of this model, $R_{G1\text{-}C}$ = 4.87 nm and $\alpha_{G1} = 54 \times 10^{-5}$ agreed well with the experimental values of 4.85 nm and 59


68 spheres
10 nm
z
y
x
34 spheres
17 spheres
8 spheres
Protein
N-terminus
Carbohydrate

(a)


+ z
- z
Protein
N-terminus
Carbohydrate
z
y
x
10 nm
- x
- y

(b)

x 10^{-5} respectively. Transformation of $I(Q)$ into $P(r)$ gave a maximum dimension of 17.8 nm for the 100% ^2H_2O model, together with a major peak at 2.5 nm and R_G of 4.70 nm, in good agreement with the experimental $P(r)$ which gave a maximum dimension of 18.0 nm, peak at 2.4 nm and R_G of 4.35 nm (Figure 4.10). A second model with oligosaccharide chains repositioned closely against the protein core in a compact conformation gave worsened agreements, as exemplified in Figure 4.15. The R_G values were lower by 0.3 nm and the residual R was 0.0161. This showed that the extended oligosaccharide conformations in $C\bar{1}$ inhibitor could be identified by scattering.

Other models for $C\bar{1}$ inhibitor were produced (Figure 17). Four alternative $+x$ models with different N-termini were constructed, based on sphere lengths and cross-sections of 68 x 1, 34 x 2, 17 x 4 and 8 x 9 (lengths 41 nm, 21 nm, 10 nm and 5 nm) (Figure 4.16(b)). The R_G values of these models were 12.1 nm, 6.2 nm, 3.7 nm and 2.9 nm respectively, and were clearly different from the experimental $R_{G\text{-}C}$ value of 4.85 nm. The deviations from the experimental scattering curve in 100% ^2H_2O seen in Figure 4.16(b) were reflected by the larger residuals R were 0.031, 0.015, 0.023 and 0.040 respectively than the R of 0.012 from Figure 4.15. The $P(r)$ calculated for the 34 x 2 and 17 x 4 models resulted in maximum lengths of 22.8 nm and 13.2 nm and $R_G$values of 5.70 nm and 3.54 nm respectively, which clearly differ from experiment (Figure 4.10). Scattering was thus able to distinguish the length of $C\bar{1}$ inhibitor in solution.

The N-terminus domain was repositioned in the $-x$, $-y$, $-z$ and $+z$ directions relative to the serpin domain. The curve simulations are shown in Figure 4.16(c) and the corresponding models in Figure 4.17(b). The $R_{G\text{-}C}$ of the models was

---

Figure 4.17. Alternative Debye sphere models for $C\bar{1}$ inhibitor

(a) Four models of $C\bar{1}$ inhibitor with the N-terminus domain in the $+x$ direction of lengths 68, 34, 17 and 8 spheres.

(b) Four models of $C\bar{1}$ inhibitor with the N-terminus domain in the $-x$, $-y$, $-z$ and $+z$ directions.

adjusted to be close to the experimental value of 4.85 nm by altering the length of the N-terminus domain. In the $-x$ and $-y$ directions, this was 17 spheres long, with $R_{G\text{-}C}$ of 4.47 nm and 4.68 respectively. Models with the N-terminus domain in the $-z$ and $+z$ directions were 23 spheres long and had $R_{G\text{-}C}$ of 5.27 nm and 4.64 nm, respectively. All gave slightly worsened curve fits compared to those of Figure 4.15 with R values of 0.016, 0.014, 0.016 and 0.013 respectively. While the $+x$ direction was favoured by the modelling, this was not unambiguous.

(4.4) CONCLUSIONS

Large conformational and stability changes had been reported between the native and split forms of α_1-antitrypsin and C1 inhibitor (Löbermann *et al.*, 1984; Carrell & Owen, 1985; Bruch *et al.*, 1988). Solution scattering however showed no evidence for a gross conformational change between the two forms. Although it was assumed from the crystal structure of α_1-antitrypsin that a large conformational change must have occurred on proteolytic cleavage, Löbermann *et al.* (1984) stated that native α_1-antitrypsin could be reformed by the minimal movement of β-strand 4A (residues 343 - 358) from the centre of β-sheet A. This was in agreement with the results of Huber and Carrell (1989) who compared 20 serpin amino acid sequences with the known crystal structure of α_1-antitrypsin. Since then, the crystal structures of four more serpins have been established, including the native form of ovalbumin (Stein *et al.*, 1990). This structure had an intact reactive loop and one less strand in β-sheet A, again confirming the view that after cleavage, the stressed reactive loop of the inhibitory serpins folded back into β-sheet A with a minimum of a conformational change. The models presented here of native α_1-antitrypsin involved a minor relocation of 7 spheres out of the total of 291 in the split α_1-antitrypsin model (Figure 4.11(b)). This placed an upper structural limit on the extent of the changes seen by spectroscopy (Bruch *et al.*, 1988).

More sensitive techniques are required to examine the conformational transition between the native and split forms. FT-IR (Haris *et al.*, 1990), NMR (Perkins *et al.* unpublished data) and CD (Bruch *et al.*, 1988; Gettins & Harten, 1988) spectroscopy and fluorescence quenching (Koloczek *et al.*, 1991) studies on several members of the serpin family have highlighted differences in secondary and tertiary structures between the native and cleaved forms. These data were supported by monoclonal antibodies with different epitopic specificities which bind either to native or to split α_1-antitrypsin (Zhu *et al.*, 1987; Zhu & Chan, 1987). FT-IR spectroscopy indicated the presence of well-defined hydrogen bonds in α-helices and β-sheets in split α_1-antitrypsin that were not present in the native form. Both FT-IR and NMR studies showed that the native form underwent

greater ^{1}H-^{2}H exchange within the peptide backbone than the split form (Haris *et al.*, 1990). The increased stability of the split form of α_1-antitrypsin was probably due to the more complete hydrogen bonding in the secondary structure due to the insertion of s4A into the β-sheet.

Conformational changes in a variety of proteins have been observed by solution scattering. In deoxyhaemoglobin, the neutron and X-ray R_G increased by 0.05 - 0.15 nm on binding oxygen (Conrad *et al.*, 1969; Schneider *et al.*, 1969) which corresponded to α-helix movements of 0.05-0.15 nm and movement between the four subunits of 0.6 nm (Baldwin & Chothia, 1979). The X-ray R_G of aspartate transcarbamylase increased by 0.25 nm after ligation to a substrate analogue (Moody *et al.*, 1979; Hervé *et al.*, 1985). Crystallography showed that the catalytic trimers move apart by 1.2 nm and that the regulatory and catalytic subunits reorientate by rotations of 10^{o} to 15^{o} (Krause *et al.*, 1987; Ke *et al.*, 1988). In X-ray scattering of glucose-binding hexokinase and L-arabinose binding protein (McDonald *et al.*, 1979; Newcomer *et al.*, 1981), the R_G values decrease by 0.09 nm to 0.13 nm after substrate binding. Sometimes no changes are seen by scattering, such as the activation of chymotrypsinogen to α-chymotrypsin (Krigbaum & Godwin, 1968) or the binding of pyridoxal 5'-phosphate to aspartate aminotransferase, (Vergé *et al.*, 1983). In conclusion, the observation of a conformational change depends on the protein system being measured.

The characterization of the serpin structure of α_1-antitrypsin facilitated the modelling of the homologous protein, $C\bar{1}$ inhibitor. Structural evidence for the distinct existence of the serpin and N-terminus domains was obtained from the successful curve fitting (Figure 4.16) and possibly from the biphasic Guinier plots (Figure 4.8). The N-terminal domain was extended, as deduced from the length calculations and comparison of the maximum length of $C\bar{1}$ inhibitor (16 to 19 nm) with α_1-antitrypsin (7.0 to 7.8 nm) and heavily glycosylated as shown by contrast variation experiments (Table 4.1). Electron microscopy had suggested that $C\bar{1}$ inhibitor was much longer at 33 to 36 nm (Odermatt *et al.*, 1981) but inherent difficulties exist in EM work such as the incorporation of artifacts and magnification errors. The final length of 16 to 19 nm was similar to lengths of 17

to 20 nm determined for the substrates of $C\bar{1}$ inhibitor, $C\bar{1}r$ and $C\bar{1}s$ (Figure 9.1) (Villiers *et al.*, 1985; Weiss *et al.*, 1986; Perkins & Nealis, 1989).

The conformational properties of oligosaccharides in glycoproteins are not readily accessed by physical techniques. Solution scattering was potentially of value since the R_G corresponded to the mean square distance of scattering vectors from the centre of scattering. Oligosaccharides are usually on the protein surface and therefore contribute maximally to the R_G or $R_{G\text{-}C}$. Even though the Stuhrmann α is less accurately measured than the $R_{G\text{-}C}$, it can be used to monitor oligosaccharide conformations. In this study, the oligosaccharide chains of α_1-antitrypsin and $C\bar{1}$ inhibitor were found to be mostly extended into solution. These were mostly biantennary oligosaccharides, but may be replaced at Asn83 in α_1-antitrypsin and to a lesser extent at Asn46 by triantennary oligosaccharides (Mega *et al.*, 1980; Vaughan & Carrell, 1981). The carbohydrate of both α_1-antitrypsin and $C\bar{1}$ inhibitor have been found not to be important for inhibitory activity (Guzdek *et al.*, 1990; Prandini *et al.*, 1986; Reboul *et al.*, 1987; Avron *et al.*, 1991) but may have functions in secretion, removal from circulation and the recognition by receptors for complex uptake (Travis *et al.*, 1990).

The present study has tested approaches for the modelling of scattering curves starting from a known crystal structure. The modelling simulations had to include the different scattering densities of the hydrophilic and hydrophobic residues. This however may be less important for proteins with a low Stuhrmann α for its R_G value. Modelling of neutron scattering data does not require the incorporation of an hydration shell which was consistent with a survey of neutron scattering densities of glycoproteins (Perkins, 1986). In contrast, X-ray modelling required hydrated structures. This was consistent with the calculation of partial specific volumes from the sum of the unhydrated volume and the electrostricted water shell surrounding the protein (Perkins, 1986), and also with the calculation of electron densities for X-ray scattering studies by contrast variation (Perkins, 1988).

CHAPTER 5.

STRUCTURAL STUDIES ON C5

(5.1) INTRODUCTION

Activation of the early components of the classical and alternative pathways by proteolytic cleavages leads to the formation of the C5 convertases, C4b2a3b or C3bBb3b respectively. The convertases activate C5 to form C5b and C5a by a single proteolytic cleavage. The generation of C5b is the initial step in the formation of the membrane attack complex, which is assembled from the complement components C6, C7, C8 and C9 in association with C5b. Insertion of the complex into membranes leads to the formation of transmembrane channels in the target cell and ultimately to cell death due to osmotic imbalance.

Human C5 is a glycoprotein of M_r 194,000 with 4 putative N-glycosylation sites. The primary sequences of the complement components C3, C4 and C5 show significant homologies (Belt *et al.*, 1984; de Bruijn & Fey, 1985; Wetsel *et al.*, 1988; Haviland *et al.*, 1991a) although the genes are found on different chromosomes (Carroll *et al.*, 1984; Whitehead *et al.*, 1982; Wetsel *et al.*, 1988). Comparison of the genomic organization of human C3, murine C4 and murine C5 showed that around 75% of the positions of the intron-exon junctions shared between the genes are identical and the remaining are close in position (Haviland *et al.*, 1991b) (Figure 5.5). The exon sizes in the C3, C4 and C5 genes were either the same or very similar but the intron sizes were very different (Haviland *et al.*, 1991b). While not as abundant as C3 and C4 (Table 1.1), C5 at 0.15 mg/ml is still one of the more abundant complement components in plasma.

C5, like C3, is composed of two disulphide-linked polypeptide chains α and β. In contrast C4 has three chains α, β and γ (Figure 1.2). Activation of C5 is brought about by either cleavage of the peptide bond between residues 74 and 75 of the α chain, releasing C5a (Nilsson *et al.*, 1975) or by reaction with oxygen radicals ($OH^{\cdot}$) to form C5(H_2O_2) (Vogt *et al.*, 1989). C5a is the most potent of the three anaphylatoxins C3a, C4a and C5a formed by the complement cascade. C3 and C4, but not C5, have a highly reactive internal β-cysteinyl-γ-glutamyl thiolester bond in a region of high sequence homology. During activation, this enables C3b and C4b to form covalent bonds to target surfaces (Law *et al.*, 1980).

Since there is no thiolester bond in C5, C5b is unable to form a covalent bond to a surface (DiScipio, 1981). Thus C5b in the presence of C6 forms a fluid-phase C5b6 bimolecular complex, and this in association with C7 expresses a membrane-binding site (Podack *et al.*,1978).

In this study, comparisons between the solution structures of C3, C4 and C5 are made. Electron microscopy of C5 depicts a multilobal, irregular ultrastructure of dimensions of 10.4 nm x 14.0 nm x 16.8 nm (DiScipio *et al.*, 1983); different dimensions have however been reported for C3 and C4 (Smith *et al.*, 1982, 1984a; Dahlbäck *et al.*, 1983). The sedimentation coefficients of C3, C4 and C5 are similar at 9.5 S (Pâques, 1980), 8.9 S (Müller-Eberhard *et al.*, 1966) and 9 S (Isenman *et al.*, 1980), respectively. Limited solution scattering data on C3, C4 and C5 suggested that despite their sequence homologies they had dissimilar R_G values (Österberg *et al.*, 1984, 1988). In this study, extensive neutron scattering experiments on the external and internal structure of C5 are critically compared with similar studies on C3 and C4 (Perkins & Sim, 1986; Perkins *et al.*, 1990a). Together, the neutron data show that C3, C4 and C5 are more homologous in their physical structures than previously believed.

(5.2) MATERIALS AND METHODS

(5.2.1) Preparation of C5

C5 was prepared by Dr R.A. Harrison (Perkins *et al.*, 1990b) from 3.5 l of human plasma from normal, healthy donors essentially as described in Harrison & Lachmann (1986). SDS-PAGE analysis showed the C5 to be greater than 95% pure and to have the anticipated two-chain structure. Neutron scattering studies were performed on freshly-prepared C5 or C5 that was stored frozen at -70°C. Samples were re-analyzed by SDS-PAGE subsequent to solution scattering studies, and no differences between pre- and post-scattering were detected.

Four different buffers were used in neutron data collection: A (200 mM NaCl, 12 mM Na phosphate, 1 mM EDTA, pH 7.0); B (50 mM NaCl, 12 mM Na phosphate, 1 mM EDTA, pH 7.0); C (3.1 mM diethyl barbituric acid, 0.9 mM sodium barbitone, 145 mM NaCl, 0.83 mM $MgCl_2$, 0.25 mM $CaCl_2$, 20 μM $ZnSO_4$); D (buffer C diluted 1 in 4, except for $ZnSO_4$ which remained at 20 μM). These had ionic strengths of 0.22 M, 0.07 M, 0.15 M and 0.04 M respectively. Buffer A was predominantly used. Samples were dialysed against the appropriate amounts of ^2H_2O buffers at 6°C with four changes of buffer over at least 36 hours. The final dialysate was used for the neutron background buffer runs.

(5.2.2) Neutron and X-ray scattering experiments

Neutron experiments were performed on instrument D17 at the ILL, Grenoble in five independent sessions by Dr S.J. Perkins. Guinier data were obtained using a sample-to-detector distance of 3.45 to 3.46 m and wavelengths of 1.30 nm to 1.60 nm, giving a Q range of 0.0053 to 0.049 nm^{-1}. Data at larger Q were obtained with a sample-to-detector distance of 1.40 m, wavelengths of 0.99 nm to 1.00 nm and two main beam-to-detector axis angles of 0° to give a Q range of 0.18 to 1.60 nm^{-1} or 20° to give a Q range of 1.03 to 3.41 nm^{-1}. Samples were counted using times of 4 min in ^2H_2O to 62 min in H_2O.

The Indirect Transformation Procedure (ITP) method (Glatter, 1982) was used to transform the neutron scattering curves $I(Q)$ into real space $P(r)$ (Chapters 2 & 3). A total of 10 splines were used to fit the 75 experimental points. The number of splines was limited because of poor counting statistics at large Q. D_{max} was specified as 20 nm on the basis of the length calculations from the Guinier R_G analyses; control calculations based on maximum lengths of 17 nm gave similar results. The number of points used to define $P(r)$ in real space was 101.

Synchrotron X-ray experiments were performed by Dr S.J. Perkins on station 7.3 (Nave *et al.*, 1985) at the SRS, Daresbury. The synchrotron operated at 1.8 GeV with beam currents of 143 to 132 mA. The sample-to-detector distance was 2.185 m and the wavelength was 0.1608 nm which gave a Q range of 0.1 to 1.3 nm^{-1}. The samples were counted for 10 minutes in alternation with their buffers.

(5.2.3) Sequence alignments

The seven C3, C4 and C5 sequences were aligned by hand as shown in Figure 5.5 in the order: human and mouse C5 (Haviland *et al.*, 1991a; Wetsel *et al.*, 1987), human, mouse and rat C3 (de Bruijn & Fey, 1985; Lundwall *et al.*, 1984; Wetsel *et al.*, 1984; Misumi *et al.*, 1990) and human and mouse C4 (Belt *et al.*, 1984; Nonaka *et al.*, 1985; Sepich *et al.*, 1985). The sequences were aligned to maximize the occurrence of identical or chemically homologous residues and to minimize gaps (Chapter 3). The amino acids occurring at intron/exon boundaries were marked for human C3 (Vik *et al.*, 1991) human C4 (Yu, 1991) and murine C5 (Haviland *et al.*, 1991b).

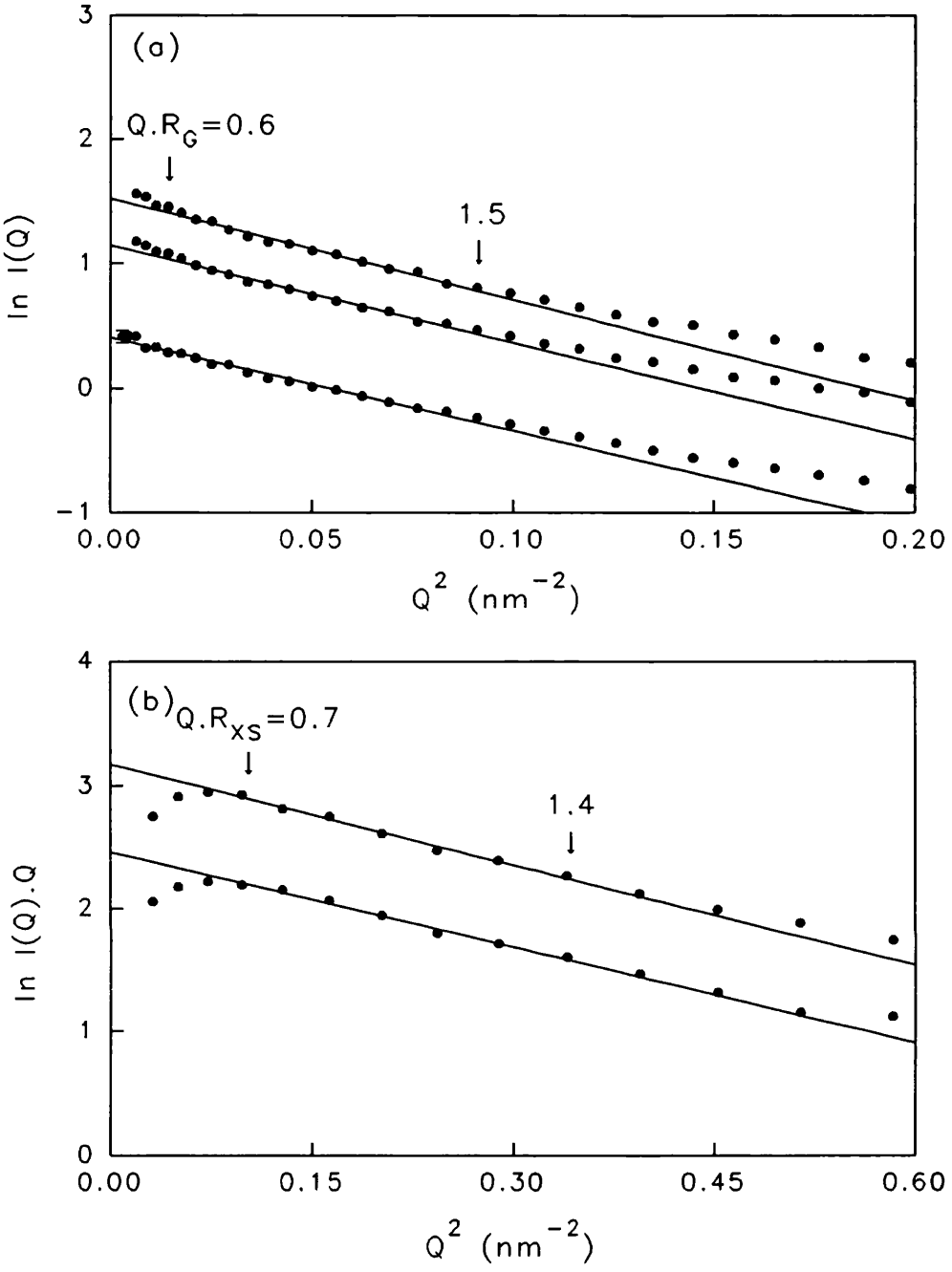


Figure 5.1. Neutron Guinier plots of human C5 in 100% ^2H_2O buffers.

(a) Guinier plots and (b) cross-sectional Guinier plots of a dilution series for C5. The vertical displacements of the curves reflect sample concentrations at 5.4, 3.9 and 1.9 mg/ml in (a) and 3.9 and 1.9 mg/ml in (b). The Q range used for line fitting was 0.12 to 0.30 nm^{-1} in (a) and 0.31 to 0.58 nm^{-1} in (b). A slight decrease of the R_G with concentration is observable in (a) and in other % ^2H_2O buffers, but was not significant in terms of the error of the $R_{G\text{-}C}$determination in Table 5.2.

(5.3) RESULTS AND DISCUSSION

(5.3.1) Neutron scattering of C5

Neutron contrast variation experiments were performed on human C5 in 0%, 70%, 80% and 100% ^2H_2O buffers. Linear Guinier analyses were successfully obtained in 100% ^2H_2O over an acceptable $Q.R_G$ range of 0.6 to 1.5 (Figure 5.1(a)). Similarly linear cross-sectional analyses were obtained in the $Q.R_{XS}$ range of 0.7 to 1.4 (Figure 5.1(b)). Comparisons of the Guinier data from all four buffers A, B, C and D showed that although the two higher salt buffers A and C gave similar good results, the two lower salt buffers showed evidence of aggregation, and these data were not considered further. A dilution series of C5 of 1.8 to 7.9 mg/ml in the high salt buffers indicated that while the $I(0)/c$ and R_G values exhibit minor decreases with concentration (results not shown), this was not significant.

The concentration of C5 was calculated from the absorption coefficient A_{280} (1%, 1 cm) for human C5 which has been estimated as 10.8 (DiScipio *et al.*, 1983) or 10.2 (Österberg *et al.*, 1988). The mean of these two values (10.5) was used. The absorption coefficient calculated from the amino acid and carbohydrate composition (Chapter 3), assuming all four carbohydrate attachment sites were occupied by biantennary carbohydrates was 10.9 (Table 5.1), in good agreement with the value derived by DiScipio *et al.* (1983). The mean M_r of four measurements was found to be 211,000 ± 10,000. This was in reasonable agreement with the calculated M_r of 194,000 from the human C5 sequence (Haviland *et al.*, 1991a) (Table 5.1) and showed that the absorption coefficient of 10.5 was satisfactory. It also confirmed that C5 was monomeric in solution. The 4.5% carbohydrate content by weight determined from the sequence composition was in good agreement with the value of 3.8% determined by Tomana *et al.* (1985). The neutron matchpoint graph (Figure 5.2(a)) gave an excellent straight line with a matchpoint of 40.9 ± 0.4 % ^2H_2O by linear regression analyses of 33 values in 4 contrasts. The corresponding cross-sectional plot (Figure 5.2(b)) yielded a matchpoint of 41.5 ± 0.4 % ^2H_2O (33 values in 4 contrasts). Both of

Residues	
total amino acids	1698
total carbohydrate	44
% mass carbohydrate	4.6
M_r	
protein	186,000
carbohydrate	9,000
total	194,000
neutron scattering	201,000 ± 9000
Dry volume (nm^3)[1]	250
Overall $\bar{v}$ (ml/g)[2]	0.738
Overall $\Sigma b/M_r$ in H_2O (fm)	0.2247
in ^2H_2O (fm)	0.3876
Predicted A_{280} (1%, 1cm)[3]	10.9
Literature A_{280}	10.5
Pred. matchpoint (% ^2H_2O)[4]	40.2
Exp. matchpoint (% ^2H_2O)	
from $I(0)$	40.9 ± 0.4
from $[I(Q).Q]$	41.5 ± 0.4
Stuhrmann analysis	
R_{G-C} (nm)	4.87 ± 0.03
α_G ($\times 10^{-5}$)	13 ± 8
R_{XS-C} (nm)	2.35 ± 0.03
α_{XS} ($\times 10^{-5}$)	13 ± 4
R_{G-C}/R_O	1.61
Models R_{G-C} (nm)	4.97

Table 5.1. Compositional and scattering data for C5

Footnotes to Table 5.1

1, 2, 3 & 4: see Table 4.1.

[3] Literature determinations of A_{280} are reported as 10.8 (DiScipio *et al.*, 1983) or 10.2 (Österberg *et al.*, 1988) and the mean of these two values was used here.

these agreed well with the matchpoint of 40.2% ^2H_2O predicted (Perkins, 1986; Chapter 3) from the human C5 sequence and confirmed the good quality of the scattering data.

To characterise the shape of C5 and its internal structure, the contrast dependence of the structural data was analyzed using Stuhrmann plots of R_G^2 and R_{XS}^2 against the reciprocal solvent-solute contrast difference $\Delta\rho^{-1}$ (Figures 5.2(c) & (d)). The radii of gyration at infinite contrast R_{G-C} and R_{XS-C} were found to be 4.87 nm and 2.35 nm respectively. The elongation ratio R_{G-C}/R_O was

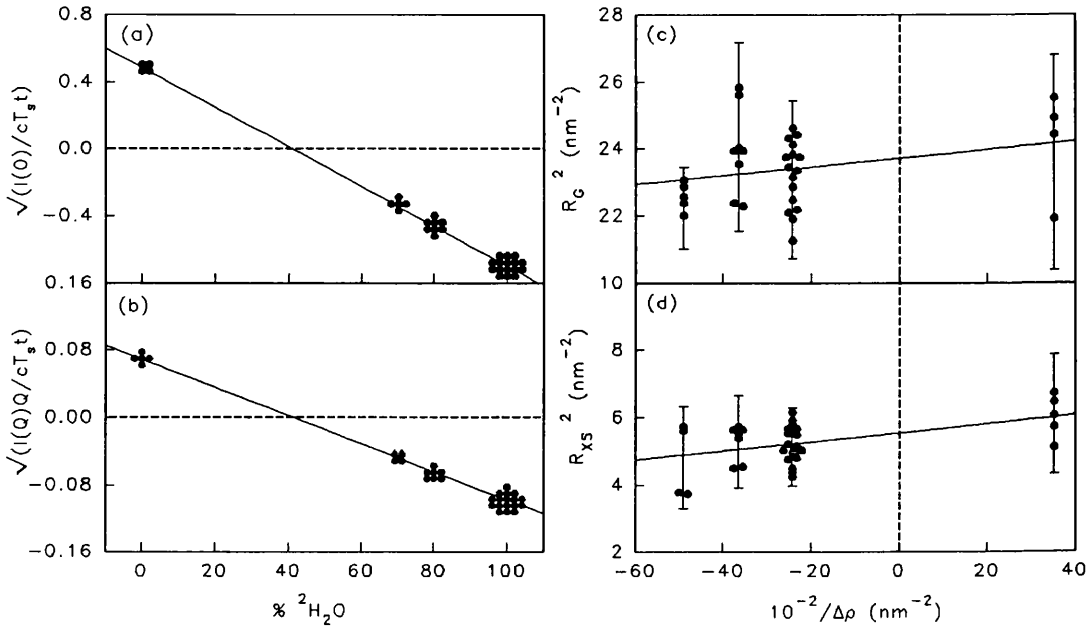


Figure 5.2. Neutron contrast variation analyses of human C5.

(a) and (b) show the matchpoint graphs of the Guinier and cross-sectional data respectively in 0%, 70%, 80% and 100% ^2H_2O buffers. The symbols are displaced for clarity. The matchpoints from linear regression are 40.9 ± 0.4 % ^2H_2O and 41.5 ± 0.4 % ^2H_2O respectively.

(c) and (d) show the R_G^2 and R_{XS}^2 data plotted against the reciprocal solvent-solute contrast difference $\Delta\rho^{-1}$. The regressed line based on the R_G^2values gives an $R_{G\text{-}C}$ of 4.87 ± 0.03 nm and a slope α of 13 ± 8 x 10^{-5}. For the R_{XS}^2 data, $R_{G\text{-}C}$ is 2.35 ± 0.03 nm and α is 13 ± 4 x 10^{-5}.

calculated as 1.61 (where R_O is the R_G of the sphere of the same dry volume as C5) (Table 5.1). This was significantly larger than the average value of 1.28 ± 0.10 for globular proteins (Perkins, 1988), and showed that C5 is elongated in shape. The radial inhomogeneity of scattering density α was positive at 13 x 10^{-5} in both Stuhrmann plots (Table 5.1). On average, the outermost scattering densities in C5 were higher than those nearer the centre. This suggested that C5 had a surface of hydrophilic amino acids and carbohydrates surrounding a hydrophobic core, typical of soluble globular glycoproteins.

The dimensions of C5 can be calculated from the R_G and R_{XS} data and also from the $I(0)$ and $[I(Q).Q]_{Q\to 0}$ data (Chapter 3). Using the data in 100% ^2H_2O, the length L of the longest axis of C5 (assuming an elliptical cylinder) was determined as 14.9 ± 0.5 nm. From the ratio of $\pi . I(0) / [I(Q).Q]_{Q\ 0}$, L was found to be 15.7 ± 1.1 nm. Combination of the L values with the total dry volume and the R_{XS} values gave axial dimensions of 14.9 nm x 8.7 nm x 2.5 nm as a first approximation to the structure of C5.

(5.3.2) X-ray scattering of C5

X-ray scattering of C5 were performed to substantiate the neutron experiments. In Guinier analyses, the final R_G and R_{XS} values were based on $Q.R_G$ and $Q.R_{XS}$ ranges of 1.0 to 1.6 and 0.7 to 1.4, i.e. Q ranges of 0.18 to 0.30 nm^{-1} and 0.31 to 0.60 nm^{-1} respectively, which were slightly more restricted than those of Figure 5.1. Many of the spectra were found to be aggregated and therefore only those spectra which gave linear Guinier plots were used. The mean R_G and R_{XS} values were 5.3 ± 0.1 nm (2 values) and 2.3 ± 0.1 nm (1 value) respectively. These were in reasonable agreement with the neutron scattering data in the same positive contrast buffer (0% ^2H_2O) (Table 5.2).

	C5	C4, C4u, C4(a+b)	C3, C3u, C3(a+b)
X-rays			
R_G (nm)	5.3 ± 0.1	5.3 ± 0.1	5.2 ± 0.1
R_{XS} (nm)	2.3 ± 0.1	2.5 ± 0.1	2.6 ± 0.1
Neutrons			
R_{G-C} (nm)	4.9 ± 0.1	4.9 ± 0.1	5.1 ± 0.1
R_{XS-C} (nm)	2.4 ± 0.1	2.2 ± 0.1	2.4 ± 0.1
α_G(x 10^{-5})	13 ± 8	2 ± 10	30 ± 10
α_{XS}(x 10^{-5})	13 ± 4	20 ± 10	10 ± 10

Table 5.2. Summary of Guinier analyses for C5, C4 and C3 of complement
The data for C4 and C3 are taken from Perkins *et al.* (1990a) and Perkins & Sim (1986), and mean values are shown.

(5.3.3) Comparisons of C5 with C3 and C4 of complement

The present X-ray and neutron data on C5 were compared with data previously obtained for C3 and C4 and their fragments (Perkins & Sim, 1986; Perkins *et al.*, 1990a) to assess the degree of structural homology between the three proteins. C3u and C4u correspond to C3 and C4 with hydrolysed thiolester bonds, while C3(a+b) and C4(a+b) correspond to activated C3 and C4 in which the two fragments continue to be associated (Figure 1.2). By both X-ray and neutron scattering, the R_G and R_{XS} values of C3, C3u, C3(a+b), C4, C4u, C4(a+b) and C5 were found to be the same within error (Table 5.2). Therefore C3, C4 and C5 have similar overall shapes in solution from the scattering curve analyses out to Q of 0.6 nm^{-1}. The values for the Stuhrmann α of C3, C4 and C5 were similar (Table 5.2) although α is less accurately measured than the R_{G-C} values. From the sequences, the ratios of hydrophilic to hydrophobic residues (Eisenberg, 1984) for C3, C4 and C5 were similar at 50:50, 47:53 and 49:51 respectively. The similarity of the α values was thus consistent with compositional data and indicate that C3, C4 and C5 had similar internal structures.

(5.3.4) Large angle curve analyses of C5

Indirect transformation of the neutron curves $I(Q)$ into real space $P(r)$ using ITP (Section 5.2 & Chapter 3) was carried out for C5, C4, C4u and C3u to assess further the structural homologies between C3, C4 and C5. The whole scattering curve was used over a Q range of 0.2 to 3.6 nm^{-1} as obtained in 100 % ^2H_2O in three detector positions at the 3.4 m and 1.4 m detector position and mainbeam-to-detector angles of 0^o and 20^o. The R_G and R_{XS} data above were based on the smaller Q range out to 0.6 nm^{-1}. At large Q, the $P(r)$ curves for C5, C4, C4u and C3u were all similar, as illustrated for C5 and C3u in Figure 5.3. This agreement confirmed the structural homologies in solution to a nominal structural resolution of 1.7 nm. The maximum lengths of C5, C4u and C3u were determined to be 14.0 nm, 15.2 nm and 15.0 nm, respectively, from the value of r where $P(r)$ diminishes to zero intensity. These were compatible with the length of 14.9 to 15.7 nm determined from the Guinier analyses for C5. The most frequently occurring

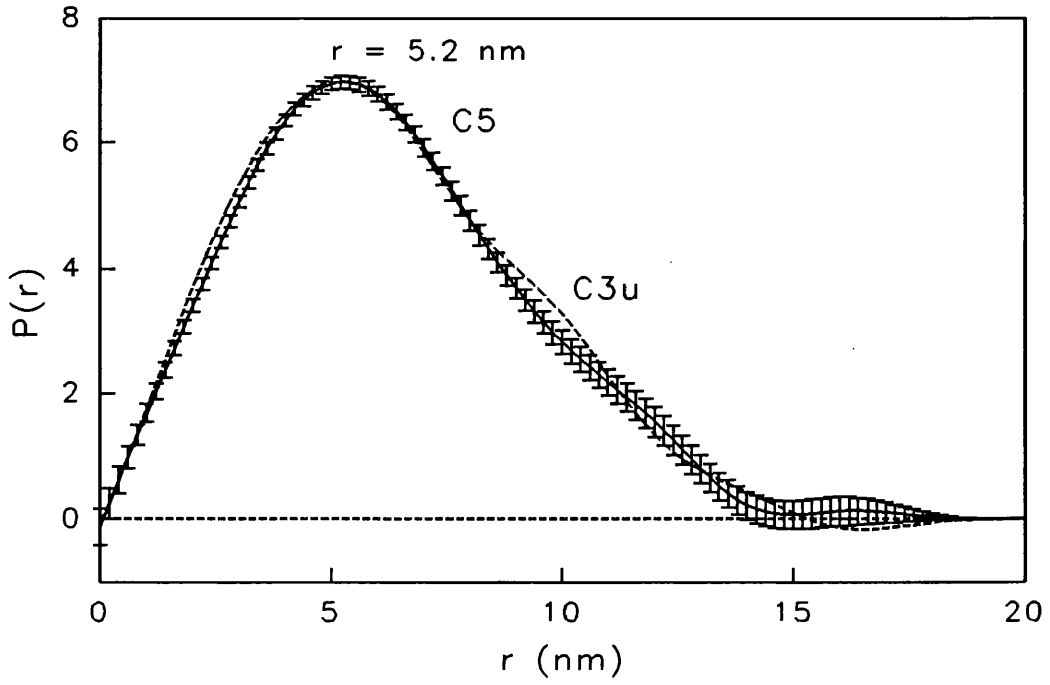


Figure 5.3 Indirect Transformation into real space of the C5 and C3u neutron scattering curves in 100 % ^2H_2O buffer.

The intraparticle distance distribution function $P(r)$ for C5 (——) and C3u (- - -) are shown after normalisation for concentration. The 75 experimental data points extending to a Q of 3.41 nm^{-1} were fitted using 10 splines and transformed assuming a maximum length of 20 nm. The errors of the transformation are shown for C5. The $P(r)$ curves for C5 and C3u are seen to be very similar.

length (the maximum in $P(r)$) for C5, C4u and C3u were all 5.2 nm. The R_G for C5 from the indirect transformation was determined as 4.89 nm over the range of $-8.0 \leq \lambda_{opt} \leq 0.0$, where λ_{opt} is the Lagrange stabilization parameter (Chapter 2). This was in good agreement with the mean value of 4.82 ± 0.10 nm for the 16 R_G values from the Guinier analyses in 100 % ^2H_2O buffers. The R_Gs for C3u and C4u were 4.77 nm and 4.71 nm over a similar range of λ_{opt}. These also compare well with the mean R_G values in 100% ^2H_2O of 5.00 ± 0.03 nm (4 values) and 4.92 ± 0.10 nm.

Further comparisons of the scattering curves out to a Q value of 1.6 nm^{-1} were performed using computer models made from Debye spheres (Chapters 2 &

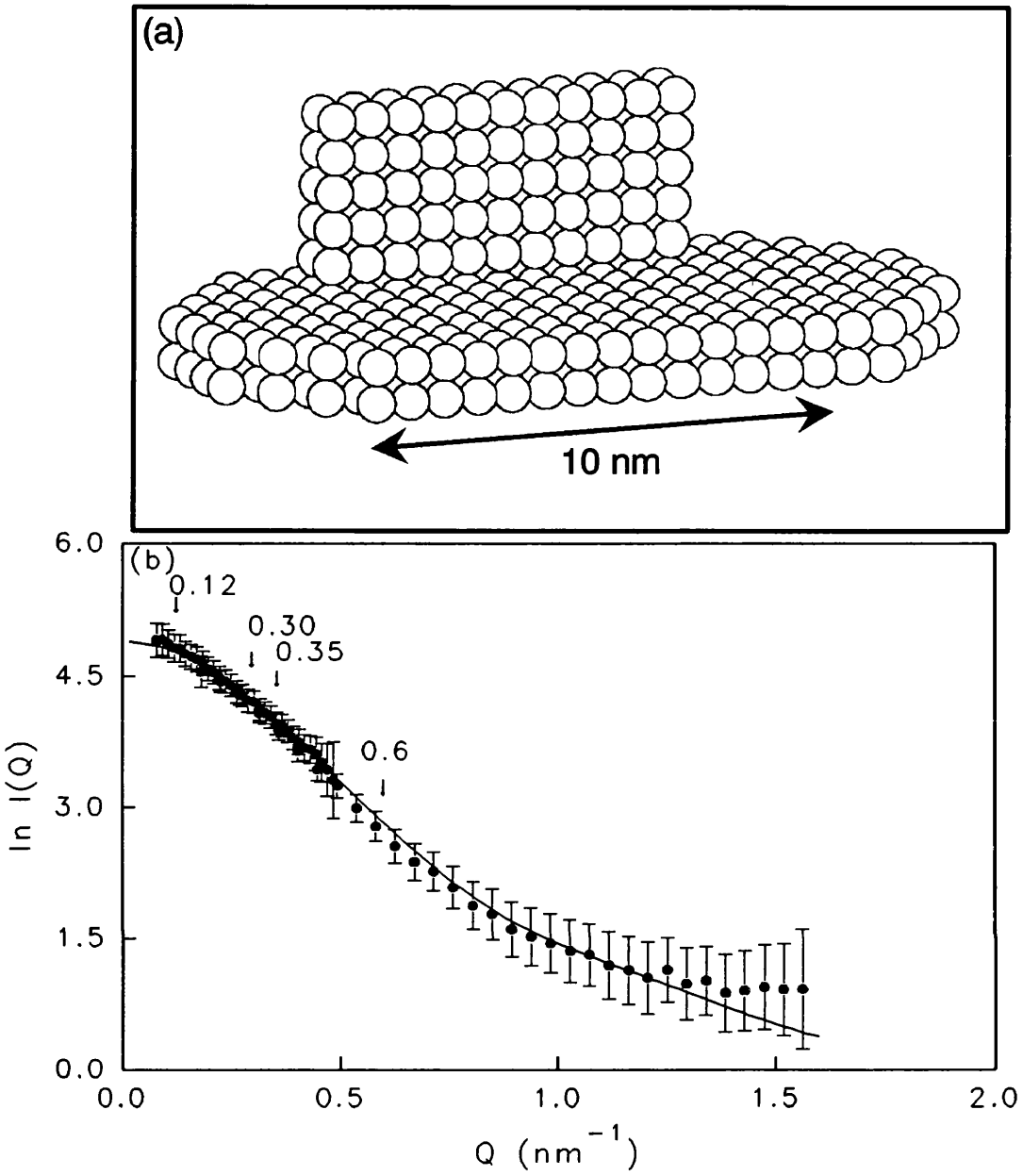


Figure 5.4. Comparison of the neutron scattering curve of C5 in ^2H_2O with Debye scattering curve simulations.

(a) The two domain model for C4 (taken from Perkins *et al.*, 1990a). The precise arrangement for the two domains should not be over interpreted.

(b) The experimental data (●) are compared with the calculated Debye scattering curve (———) using the possible two-domain model for C4 and C4u shown above. This sphere model gives a satisfactory curve fit. The R_G of the model is 4.97 nm, in good agreement with the values of 4.82 nm and 4.89 nm from the Guinier and ITP analyses of the data in 100% ^2H_2O (see text).

3). To test whether C5 and C4 were similar in structure, the experimental scattering curve for C5 was compared with the simulated scattering curve of the C4 and C4u model (Figure 5.4(a)) (Perkins *et al.*, 1990a). The simplistic model was constructed from two distinct domains (Table 5.3) which corresponded to C4d and C4c, and which were set at an angle to each other. Their dimensions are given in Table 5.3; the length of 18 nm was slightly longer than the three calculations of 15 to 16 nm above but was within error of the modelling technique. The R_G of the C4 model (4.97 nm) is in good agreement with the $R_{G\text{-}C}$ (4.87 nm) for the experimental C5 data. The reasonable agreement of the C4 model with the C5 experimental curve (R factor 0.015) (Figure 5.4(b)) suggested that C5 also has a substructure with at least two domains as found for C3 and C4, even though this structure had not been revealed in physiological inactivation events.

		Triaxial dimensions (nm)	Reference
Cobra venom factor*		8.2 x 13.7	Smith *et al.* (1982)
C4b		9 x 17	Dahlbäck *et al.* (1983)
C3b, C4b		6.5 x 7.5 x 12.5	Smith *et al.* (1984a)
C5		10.4 x 14.0 x 16.8	DiScipio *et al.* (1983)
Solution scattering model+			
C3, C4, C5	C3c, C4c	2 x 8 x 18	Perkins *et al.* (1990a)
	C3dg, C4d	2 x 4 x 9	

Table 5.3. Comparisons of electron microscopy studies on C3, C4 and C5 with the Debye solution scattering model

* The cobra venom factor is a structural analogue of C3c. The reported standard deviations are ± 1.1 to ± 1.3 nm.

+ See Figure 5.4(a) for the possible arrangement of the two domains.

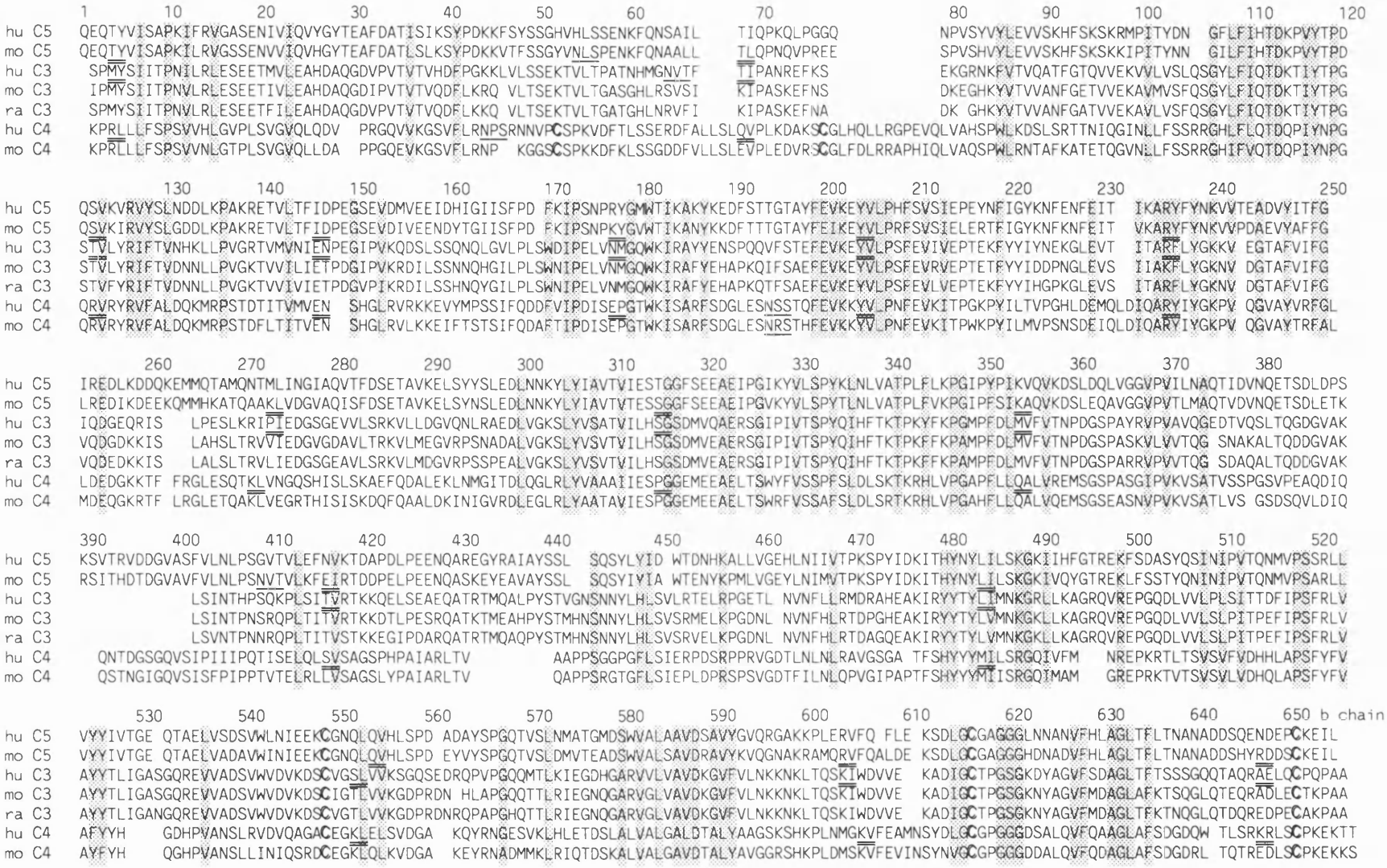
(5.3.5) Sequence alignments of C3, C4 and C5

Seven sequences of C3, C4 and C5 were aligned manually as shown in Figure 5.5 in order to assess the degree of sequence similarity between the three

complement components. The number and size of inserted gaps were minimized. A total of 381 residues, or homologous residues (defined in Chapter 2), were conserved over all 7 sequences and total of 23 (68%) cysteine residues were conserved in all 7 sequences. Comparison of the 1658 residues in human C5 with human C3 and C4 showed that 26% and 22% of residues were identical, respectively. These compare favourably with the values of 27% and 25% determined by the alignment of Wetsel *et al.* (1988). It should be noted that only the last 1639 residues of human C5 were available to Wetsel and a more liberal use of gaps were used in his alignment. Carbohydrate attachment sites (NxS or NxT) were fairly well conserved within species. One site (residue 1612) was conserved over the C3 and C5 sequences. Human C3 contained an RGD sequence (residue 1391, C5 numbering) which is thought to act as a recognition signal in ligand-receptor interactions. Since this sequence is not conserved amongst species, there must be other factors required (Wetsel *et al.*, 1984). The thiolester region in C3 and C4 was well conserved but the critical cysteine and glutamine residues were replaced by serine and alanine respectively in C5.

The amino acids at the splice junctions for murine C5, human C3 and human C4 were compared (Figure 5.5). Each protein contained 40 intron/exon boundaries and of these 75% were found in identical positions for all three proteins. The other 25% were no more than 4 amino acids distant. This indicated that the three genes were very similar in structural organization.

Properdin only binds to C3, not C5 or C4 and the binding site has been localized to a 34 residue fragment at the C-terminal end of the α-chain (Daoudaki *et al.*, 1988). These authors argued that the lack of sequence homology between C3 with C4 and C5 in the region 1402 to 1435 (C3 numbering) and the high degree of homology between C3 species was evidence that this sequence of 34 residues formed part of the binding site. However, in the alignment shown in Figure 5.5, which differs from the Daoudaki *et al.* alignment, a relatively high degree of homology is seen, with C5 and C4 being 35% and 47% homologous to C3 in this region. There is a region of insertion in the C3 sequences relative to C4 and C5 and this may be significant, as was also pointed out by Daoudaki *et al.* (1988).


1 10 20 30 40 50 60 70 80 90 100 110 120
hu C5 QEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKFQNSAIL TIQPKQLPGGQ NPVSYVYLEVVSKHFSKSKRMPITYDN GFLFIHTDKPVYTPD
mo C5 QEQTYVISAPKILRVGSSENVVIQVHGYTEAFDATLSLKSYPDKKVTFSSGYVNLSPENKFQNAALL TLQPNQVPREE SPVSHVYLEVVSKHFSKSKKIPITYNN GILFIHTDKPVYTPD
hu C3 SPMYSIITPNILRLESEETMVLEAHDAQGDVPVTVTVHDFPGKKLVLSSEKTVLTPATNHMGNVTF TIPANREFKS EKGRNKFVTVQATFGTQVVEKVVLVSLQSGYLFIQTDKTIYTPG
mo C3 IPMYSIITPNVLRLESEETIVLEAHDAQGDIPVTVTVQDFLKRQ VLTSEKTVLTGASGHLRSVSI KIPASKEFNS DKEGHKYVTVVANFGETVVEKAVMVSFQSGYLFIQTDKTIYTPG
ra C3 SPMYSIITPNVLRLESEETFILEAHDAQGDVPVTVTVQDFLKKQ VLTSEKTVLTGATGHLNRVFI KIPASKEFNA DK GHKYVTVVANFGATVVEKAVLVSFQSGYLFIQTDKTIYTPG
hu C4 KPRLLLFSPSVVHLGVPLSVGVQLQDV PRGQVVKGSVFLRNPSRNNVPCSPKVDFTLSSERDFALLSLQVPLKDAKSCGLHQLLRGPEVQLVAHSPWLKDSLSRTTNIQGINLLFSSRRGHLFLQTDQPIYNPG
mo C4 KPRLLLFSPSVVNLGTPLSVGVQLLDA PPGQEVKGSVFLRNP KGGSCSPKKDFKLSSGDDFVLLSLEVPLEDVRSCGLFDLRRAPHIQLVAQSPWLRNTAFKATETQGVNLLFSSRRGHIFVQTDQPIYNPG
130 140 150 160 170 180 190 200 210 220 230 240 250
hu C5 QSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHIGIISFPD FKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKNFENFEIT IKARYFYNKVVTEADVYITFG
mo C5 QSVKIRVYSLGDDLKPAKRETVLTFIDPEGSEVDIVEENDYTGIISFPD FKIPSNPKYGVWTIKANYKKDFTTTGTAYFEIKEYVLPRFSVSIELERTFIGYKNFKNFEIT VKARYFYNKVVPDAEVYAFFG
hu C3 STVLYRIFTVNHKLLPVGRTVMVNIENPEGIPVKQDSLSSQNQLGVLPLSWDIPELVNMGQWKIRAYYENSPQQVFSTEFEVKEYVLPSFEVIVEPTEKFYYIYNEKGLEVT ITARFLYGKKV EGTAFVIFG
mo C3 STVLYRIFTVDNNLLPVGKTVVILIETPDGIPVKRDILSSNNQHGILPLSWNIPELVNMGQWKIRAFYEHAPKQIFSAEFEVKEYVLPSFEVRVEPTETFYYIDDPNGLEVS IIAKFLYGKNV DGTAFVIFG
ra C3 STVFYRIFTVDNNLLPVGKTVVIVIETPDGVPIKRDILSSHNQYGILPLSWNIPELVNMGQWKIRAFYEHAPKQTFSAEFEVKEYVLPSFEVLVEPTEKFYYIHGPKGLEVS ITARFLYGKNV DGTAFVIFG
hu C4 QRVRYRVFALDQKMRPSTDTITVMVENSHGLRVRKKEVYMPSSIFQDDFVIPDISEPGTWKISARFSDGLESNSSTQFEVKKYVLPNFEVKITPGKPYILTVPGHLDEMQLDIQARYIYGKPV QGVAYVRFGL
mo C4 QRVRYRVFALDQKMRPSTDFLTITVENSHGLRVLKKEIFTSTSIFQDAFTIPDISEPGTWKISARFSDGLESNRSTHFEVKKYVLPNFEVKITPWKPYILMVPSNSDEIQLDIQARYIYGKPV QGVAYTRFAL
260 270 280 290 300 310 320 330 340 350 360 370 380
hu C5 IREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPS
mo C5 LREDIKDEEKQMMHKATQAAKLVDGVAQISFDSETAVKELSYNSLEDLNNKYLYIAVTVTESSGGFSEEAEIPGVKYVLSPYTLNLVATPLFVKPGIPFSIKAQVKDSLEQAVGGVPVTLMAQTVDVNQETSDLETK
hu C3 IQDGEQRIS LPESLKRIPIEDGSGEVVLSRKVLLDGVQNLRAEDLVGKSLYVSATVILHSGSDMVQAERSGIPIVTSPYQIHFTKTPKYFKPGMPFDLMVFVTNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAK
mo C3 VQDGDKKIS LAHSLTRVVTEDGVGDAVLTRKVLMEGVRPSNADALVGKSLYVSVTVILHSGSDMVEAERSGIPIVTSPYQIHFTKTPKFFKPAMPFDLMVFVTNPDGSPASKVLVVTQG SNAKALTQDDGVAK
ra C3 VQDEDKKIS LALSLTRVLIEDGSGEAVLSRKVLMDGVRPSSPEALVGKSLYVSVTVILHSGSDMVEAERSGIPIVTSPYQIHFTKTPKFFKPAMPFDLMVFVTNPDGSPARRVPVVTQG SDAQALTQDDGVAK
hu C4 LDEDGKKTF FRGLESQTKLVNGQSHISLSKAEFQDALEKLNMGITDLQGLRLYVAAAIIESPGGEMEEAELTSWYFVSSPFSLDLSKTKRHLVPGAPFLLQALVREMSGSPASGIPVKVSATVSSPGSVPEAQDIQ
mo C4 MDEQGKRTF LRGLETQAKLVEGRTHISISKDQFQAALDKINIGVRDLEGLRLYAATAVIESPGGEMEEAELTSWRFVSSAFSLDLSRTKRHLVPGAHFLLQALVQEMSGSEASNVPVKVSATLVS GSDSQVLDIQ
390 400 410 420 430 440 450 460 470 480 490 500 510 520
hu C5 KSVTRVDDGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSL SQSYLYID WTDNHKALLVGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLL
mo C5 RSITHDTDGVAVFVLNLPSNVTVLKFEIRTDDPELPEENQASKEYEAVAYSSL SQSYIYIA WTENYKPMLVGEYLNIMVTPKSPYIDKITHYNYLILSKGKIVQYGTREKLFSSTYQNINIPVTQNMVPSARLL
hu C3 LSINTHPSQKPLSITVRTKKQELSEAEQATRTMQALPYSTVGNSNNYLHLSVLRTELRPGETL NVNFLLRMDRAHEAKIRYYTYLIMNKGRLLKAGRQVREPGQDLVVLPLSITTDFIPSFRLV
mo C3 LSINTPNSRQPLTITVRTKKDTLPESRQATKTMEAHPYSTMHNSNNYLHLSVSRMELKPGDNL NVNFHLRTDPGHEAKIRYYTYLVMNKGKLLKAGRQVREPGQDLVVLSLPITPEFIPSFRLV
ra C3 LSVNTPNNRQPLTITVSTKKEGIPDARQATRTMQAQPYSTMHNSNNYLHLSVSRVELKPGDNL NVNFHLRTDAGQEAKIRYYTYLVMNKGKLLKAGRQVREPGQDLVVLSLPITPEFIPSFRLV
hu C4 QNTDGSGQVSIPIIIPQTISELQLSVSAGSPHPAIARLTV AAPPSGGPGFLSIERPDSRPPRVGDTLNLNLRAVGSGA TFSHYYYMILSRGQIVFM NREPKRTLTSVSVFVDHHLAPSFYFV
mo C4 QSTNGIGQVSISFPIPPTVTELRLLVSAGSLYPAIARLTV QAPPSRGTGFLSIEPLDPRSPSVGDTFILNLQPVGIPAPTFSHYYYMIISRGQIMAM GREPRKTVTSVSVLVDHQLAPSFYFV
530 540 550 560 570 580 590 600 610 620 630 640 650 b chain
hu C5 VYYIVTGE QTAELVSDSVWLNIEEKCGNQLQVHLSPD ADAYSPGQTVSLNMATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQ FLE KSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQENDEPCKEIL
mo C5 VYYIVTGE QTAELVADAVWINIEEKCGNQLQVHLSPD EYVYSPGQTVSLDMVTEADSWVALSAVDRAVYKVQGNAKRAMQRVFQALDE KSDLGCGAGGGHDNADVFHLAGLTFLTNANADDSHYRDDSCKEIL
hu C3 AYYTLIGASGQREVVADSVWVDVKDSCVGSLVVKSGQSEDRQPVPGQQMTLKIEGDHGARVVLVAVDKGVFVLNKKNKLTQSKIWDVVE KADIGCTPGSGKDYAGVFSDAGLTFTSSSGQQTAQRAELQCPQPAA
mo C3 AYYTLIGASGQREVVADSVWVDVKDSCIGTLVVKGDPRDN HLAPGQQTTLRIEGNQGARVGLVAVDKGVFVLNKKNKLTQSKIWDVVE KADIGCTPGSGKNYAGVFMDAGLAFKTSQGLQTEQRADLECTKPAA
ra C3 AYYTLIGANGQREVVADSVWVDVKDSCVGTLVVKGDPRDNRQPAPGHQTTLRIEGNQGARVGLVAVDKGVFVLNKKNKLTQSKIWDVVE KADIGCTPGSGKNYAGVFMDAGLTFKTNQGLQTDQREDPECAKPAA
hu C4 AFYYH GDHPVANSLRVDVQAGACEGKLELSVDGA KQYRNGESVKLHLETDSLALVALGALDTALYAAGSKSHKPLNMGKVFEAMNSYDLGCGPGGGDSALQVFQAAGLAFSDGDQW TLSRKRLSCPKEKTT
mo C4 AYFYH QGHPVANSLLINIQSRDCEGKLQLKVDGA KEYRNADMMKLRIQTDSKALVALGAVDTALYAVGGRSHKPLDMSKVFEVINSYNVGCGPGGGDDALQVFQDAGLAFSDGDRL TQTREDLSCPKEKKS


α chain
α'chain
[thiol ester]
hu C5
mo C5
hu C3
mo C3
ra C3
hu C4
mo C4

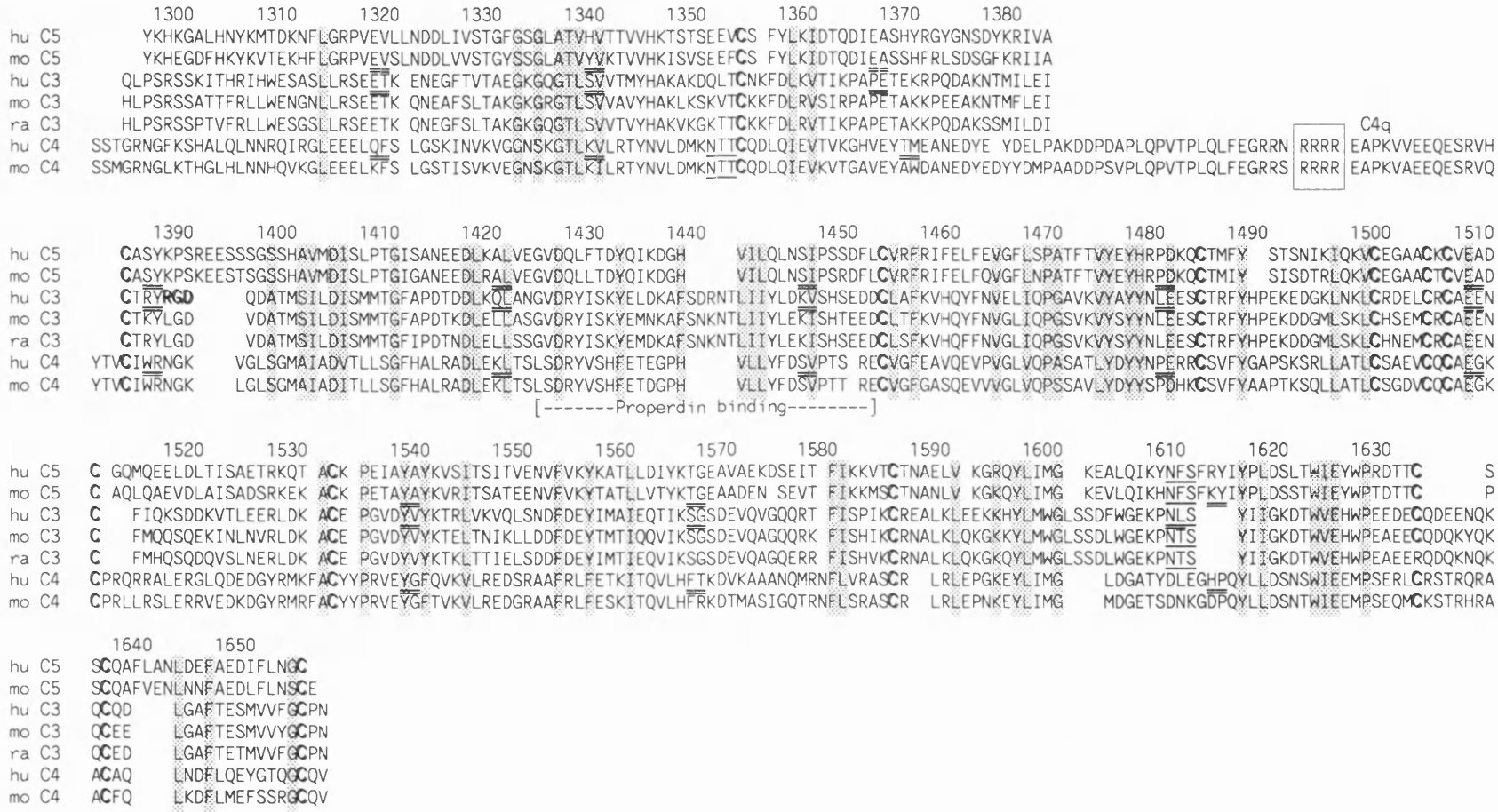


Figure 5.5. Alignment of amino acid sequences of C3, C4 and C5.

Residues or homologous residues conserved over all 7 sequences are shaded. Putative N-carbohydrate attachment sites are underlined and cysteine residues are highlighted. The thiol ester region and the C3 properdin binding site are indicated. The 4 arginine residues lost during formation of the α and β chains in C3, C4 and C5 and the γ chain in C4, are boxed. The amino acids involved at the intron/exon boundaries for murine C5, human C3 and human C4 are double underlined.

(5.4) CONCLUSIONS

Scattering studies on C5 correlated well with similar studies on the homologous complement components C3 and C4 (Perkins & Sim, 1986; Perkins *et al.*, 1990a). The data of Table 5.2 showed that C5 was very similar in gross structural features to those of C3, C3u, C3(a+b), C4, C4u and C4(a+b). Comparisons of the full scattering curves out to a structural resolution of 2 nm (Q = 3.6 nm^{-1}) showed that these three glycoproteins were structurally closely homologous. This was further verified by comparisons between the complete amino acid sequences that have been recently published for human C3, C4 and C5 (de Bruijn & Fey, 1985; Belt *et al.*, 1984; Wetsel *et al.*, 1987, 1988; Haviland *et al.*, 1991a). The C3, C4 and C5 sequences can be readily aligned, and the location of cysteinyl residues and the hydropathy plots for C3, C4 and C5 were very similar (Nonaka *et al.*, 1985; Wetsel *et al.*, 1987).

This result was in contrast to previous electron microscopy studies on these complement components, as summarised in Table 5.3 (Smith *et al.*, 1982, 1984a; Dahlbäck *et al.*, 1983). It was reported that C3, C4 and C5 possessed globular structures which could be subdivided into at least two and at most four or five domains and with dimensions that differ by several nanometres. Calculation of the volume of the C3b and C4b images was comparable with values of 241 to 244 nm^3 calculated from their sequences, while that of C5, assuming a shape of an elliptical cylinder, gave a volume that was five times larger than the sequence volume of 248 nm^3 (Table 5.1). At least one of the dimensions of C5 may be overestimated. These differences were most likely to result from the non-physiological conditions of the electron microscopy measurements. In relation to the limited scattering data reported by Österberg *et al.* (1984, 1988) on C3, C4 and C5, their R_G values of C3u, C4u and C5 were found to be similar at 5.0 to 5.3 nm, in agreement with the data presented here. However their R_G values for C3 and C4 were as low as 4.25 nm, and their α values for C3, C4 and C5 varied between -100 to 55 x 10^{-5}. These differences with the present study were best explained by the limited data that were available in the previous studies. For example, the range of the data points in Figure 5.2 show that repeated

measurements of the scattering curves preferably in different buffers and protein concentrations were necessary for a safer interpretation of the experiments.

Structural information on the three small homologous anaphylatoxins C3a, C4a and C5a (Chapter 1) has been obtained using protein crystallography (Huber *et al.*, 1980), model building (Greer, 1985, 1986) and proton NMR (Muto *et al.*, 1985, 1987; Nettesheim *et al.*, 1988; Zuiderweg *et al.*, 1988a, b; Zarbock *et al.*, 1988). The model building showed that C3a, C4a and C5a all have conserved interior residues but very different external surfaces. The crystal structure of C3a has also proved to be a good model for the secondary structure of C3a and C5a in solution by NMR except at the N- and C-terminal ends of the structure. These results extend the above observations of structural homologies.

C5b initiates the formation of the membrane attack complex. Since this complex is inserted into membranes, the degree to which the C5 structure is hydrophobic becomes relevant. Al Salihi *et al.* (1982, 1988) interpreted observations of the behaviour of C5 on phenyl-Sepharose chromatography and the effects of trypsin digestion as evidence for hydrophobic surface regions on C5. The loss of C5a resulted in an increased hydrophobicity in the C5b fragment, suggesting that a conformational change had occurred (Al Salihi *et al.*, 1988). Furthermore, DiScipio *et al.* (1983) suggested that the aggregation of C5b seen on activation of C5 was the result of the exposure of hydrophobic sites; the observation of such aggregates was supported by our preliminary neutron studies on C5b. It should however be noted that activation of C5 in the absence of C6 is nonphysiological and that neither C5b nor C5b,6 binds to membranes. The formation of a ternary complex with C7 is required before membrane binding occurs, and it is likely that the membrane binding site of the complex lies on C7. Neutron and X-ray scattering provided information about the average internal disposition of hydrophilic and hydrophobic residues. The present data on the Stuhrmann α showed that within C5 this distribution was similar to that of native C3 and C4 (Table 5.2) with hydrophilic residues found on average furthest from the core, as indeed found for globular proteins (Perkins, 1988). This was expected since all three were water soluble and had similar ratios of hydrophobic and

hydrophilic residues. The scattering data thus limits the extent to which the surface of C5 can be considered hydrophobic, and in contrast to deductions made from other biochemical investigations, rules out any significant region of surface hydrophobicity.

CHAPTER 6.

STRUCTURAL STUDIES ON PROPERDIN

```
        1         10         20         30         40
human   DPVLCFTQYEESSGKCKGLLGGGVSVEDCCLNTAFAYQKRSGGLCQPCR
mouse       CFTQYEESSGRCKGLLGRDIRVEDCCLNAAYAFQKEHDGLCQACR

                          N-terminus

        50         60         70         80         90         100
human   SPRWSLWSTWAPCSVTCSEGSQLRYRRCVGWNGQCSGKVAPGTLEWQLQACEDQQCCP
mouse   SPQWSAWSLWGPCSVTCSEGSQLRHRRCVGRGGQCSENVAPGTLEWQLQACEDQPCCP

                                TSR 1

          110        120        130        140        150        160
human   EMGGWSGWGPWEPCSVTCSKGTRTRRRACNHPAPKCGGHCPGQAQESEACDTQQVCP
mouse   EMGGWSEWGPWGPCSVTCSKGTQIRQRVCDNPAPKCGGHCPGEAQQSQACDTQKTCP

                                TSR 2

              170        180        190        200        210        220
human   THGAWATWGPWTPCSASCHGGPHEPKETRSRKCSAPEPSQKPPGKPCPGLAYEQRRCTGLPPCP
mouse   THGAWASWGPWSPRSGSCLGGAQEPKETRSRSCSAPAPSHQPPGKPCSGPAYEHKACSGLPPCP

                                TSR 3

            230        240        250        260        270        280
human   VAGGWGPWGPVSPCPVTCGLGQTMEQRTCNHPVPQHGGPFCAGDATRTHICNTAVPCP
mouse   VAGGWGPWSPLSPCSVTCGLGQTLEQRTCDHPAPRHGGPFCAGDATRNQMCNKAVPCP

                                TSR 4

               290        300        310        320        330        340
human   VDGEWDSWGEWSPCIRRNMKSISCQEIPGQQSRGRTCRGRKFDGHRCAGQQQDIRHCYSIQHCP
mouse   VNGEWEAWGKWSDCSRLRMSINCEGTPGQQSRSRSCGDRKFNGKPCAGKLQDIRHCYNIHNCI

                                TSR 5

                    360        370        380        390        400        410
human   LKGSWSEWSTWGLCMPPCGPNPTRARQRLCTPLLPKYPPTVSMVEGQGEKNVTFWGRPLP
mouse   MKGSWSQWSTWSLCTPPCSPNATRVRQRLCTPLLPKYPPTVSMVEGQGEKNVTFWGTPRP

                                TSR 6

                    420        430        440
human   RCEELQGQKLVVEEKRPCLRVPACKDPEEEEL
mouse   LCEALQGQKLVVEEKRSCLHVPVCKDPEEKKP

                        C-terminus
```

Figure 6.1. Alignment of human and mouse properdin sequences.

Human (Nolan, 1991) and mouse (Goundis & Reid, 1988) properdin are aligned and conserved residues are shaded. Cysteine residues are highlighted. There is a high degree of homology (76%) between the two species. There is a 49 residue N-terminal region followed by 6 tandemly repeated TSR motifs and ending with a 32 residue long C-terminus.

(6.1) INTRODUCTION

Properdin is involved in the regulation of the alternative pathway by binding to and stabilizing the C3/C5 convertase complexes, C3b.Bb and $C3b_n$.Bb (Fearon & Austen, 1975; Medicus *et al.*, 1976; Farries *et al.*, 1987). It also inhibits the cleavage of C3b by factor I in the presence of factor H as cofactor by interfering with the ability of factor I to bind to C3b (Farries *et al.*, 1988a). Properdin binds to the C-terminal region of the α-chain of C3 between residues 1402 - 1435 (Lambris *et al.*, 1984; Daoudaki *et al.*, 1988) (Figure 5.5) and this binding site is expressed preferentially on surface-bound C3b rather than fluid-phase C3b (Farries *et al.*, 1988a). There is also evidence that properdin binds to the Ba domain of factor B in order to facilitate the cleavage by factor D to form the C3 convertase complex, with the concomitant release of Ba (Farries *et al.*, 1988b).

Mouse and human properdin each consist of a single polypeptide chain of 441-442 amino acids (Goundis & Reid, 1988; Nolan & Reid, 1990; Nolan *et al.*, 1991) (Figure 6.1) with an apparent M_r of 55,000 when examined under dissociating conditions with or without reduction of disulphide bridges. The percentage homology between the two sequences is high with a value of around 76% (Nolan *et al.*, 1991). Human properdin also contains up to 10% by weight of complex-type N-linked carbohydrate, corresponding to 1 or 2 oligosaccharides per monomer (Minta & Lepow, 1974; Farries & Atkinson, 1989). Six structural motifs homologous to a domain found in thrombospondin (the thrombospondin repeat; TSR) have been identified in properdin, together with an N-terminal and C-terminal region (Goundis & Reid, 1988) (Figure 6.1). In human properdin, each TSR is coded by a separate exon, except that the C-terminal segment also has half the sixth TSR (Nolan *et al.*, 1991). Each TSR is about 58 residues in size. While circular dichroism experiments did not identify any α helix or β sheet structures (Smith *et al.*, 1984b), FT-IR spectroscopy indicated evidence for β-sheet and β-turn structures in the TSRs, and secondary structure predictions suggested that the β-turns could be correlated with the high proportions of Gly, Pro, Cys and Ser residues in properdin (Perkins *et al.*, 1989).

In plasma, human properdin exists as dimers (P_2; 26% of the total), trimers (P_3; 54%) and tetramers (P_4; 20%) at a total concentration of 5.7 ± 1.0 μg/ml, making it one of the least abundant complement components (Pangburn, 1989). Their specific activities decrease in the order $P_4 > P_3 > P_2$. A polymer of high molecular weight (P_n) had been identified as the "activated" form of properdin, however this has now been shown to be an artefact of purification (Farries *et al.*, 1987; Pangburn, 1989). Electron microscopy on properdin has shown that the monomer is a flexible, rod-like structure of 26 nm in length and 2.5 nm in diameter (Smith *et al.*, 1984b). The quaternary structure of the P_2, P_3 and P_4 oligomers is formed by the association of these monomers into cyclic structures, probably by strong noncovalent interactions between the N-terminal and C-terminal regions of each monomer.

The TSR is the second most abundant domain found in the complement components. Since about 80% of properdin is constructed of TSRs, properdin is ideal for structural studies of the TSR. These data are required for structural studies of the late complement components, C6, C7, C8$\alpha\gamma$, C8β and C9 (Chapter 8). The availability of TSR sequences in these proteins now permits the computation of an optimal sequence alignment and the classification of the TSRs and their physical properties. Small-angle scattering is a powerful technique for determining the arrangement of domains in multi-domain proteins, especially if the analyses can be constrained by other sequence or structural information (Perkins, 1988). For this, high flux beam sources are required in view of the low solubility of properdin. Since the structure is studied under conditions close to physiological, this offers the important advantage that the dimensions of electron microscopy images of properdin (measured *in vacuo*) can be quantitatively tested. This analysis is independently verified by the use of hydrodynamic spheres to simulate experimental sedimentation coefficients. By these means, the arrangements of TSR domains in properdin P_2 and P_3 is determined, together with an assessment of the TSR structure in the late complement components.

(6.2) MATERIALS AND METHODS

(6.2.1) Preparation of properdin

The dimeric P_2, trimeric P_3 and tetrameric P_4 forms of properdin were prepared by Dr K.B.M. Reid as described in Perkins *et al.* (1989). Neutron data were collected on Instrument D11 using either a buffer containing 0.15 M-NaCl, 0.2 mM-EDTA and 12 mM-phosphate, pH 7.4, or 0.5 M-NaCl, 0.2 M-glycine and 12 mM phosphate, pH 7.4, or on Instrument D17 using a buffer containing 0.2 M-NaCl and 12 mM-phosphate at pH 7.0. X-ray data were collected using a buffer containing 0.5 M-NaCl, 0.2 M-glycine and 12 mM phosphate, pH 7.4. For the neutron experiments, samples were dialysed at 6°C into their buffers containing 0%, 80% or 100% ^2H_2O for at least 36 hours with four buffer changes. The properdin samples showed a single band corresponding to a M_r of approximately 55,000 on SDS-PAGE both before and after the solution scattering experiments. The functional activity of the properdin samples was checked after solution scattering in an assay involving the lysis of rabbit red blood cells by a reagent dependent upon the addition of purified properdin (Reid, 1981).

(6.2.2) Neutron and X-ray scattering experiments

Neutron scattering data were obtained in four independent sessions on Instruments D11 and D17 at the ILL, Grenoble. Data were collected by Dr S.J. Perkins on Instrument D11 in two sessions with sample-to-detector distances of 2.00 m, 10.00 m and 11.00 m and wavelengths of 1.001 to 1.002 nm which resulted in a Q range of 0.05 nm^{-1} to 1.1 nm^{-1}. Counting times ranged between 2.5 to 4 hours in 0% ^2H_2O buffers at 10 m and 11 m for Guinier data to 5 to 8 min in 100% ^2H_2O buffers at 2 m for data at large Q, over a concentration range of 0.43 to 1.16 mg/ml. In two sessions on Instrument D17, the Q range used was 0.07 to 1.4 nm^{-1}, based on sample-to-detector distances of 1.40 m and 3.46 m with a wavelength of 1.106 nm and a main beam-to-detector angle of 0°. Samples in 100% ^2H_2O buffers were counted for 1.6 hours at 3.46 m and 0.8 hours at 1.40 m over concentrations of 0.34 to 0.42 mg/ml.

X-ray scattering data were obtained on station 8.2 at the SRS, Daresbury. Experiments were performed with beam currents of 118-133 mA and a ring energy of 2.00 GeV in one session. Samples were measured for 20 minutes at concentrations between 0.28 and 0.85 mg/ml. A 500-channel linear detector at a sample-to-detector distance of 3.83 m was used to give a Q range of 0.05 to 0.97 nm^{-1}. Because of the low protein concentrations the data were prone to buffer subtraction errors and curves were only accepted for subsequent analyses if their Guinier plots were linear and led to reproducible data in repeated measurements. Curves were recorded in 10 time frames which were checked individually for radiation damage and then were averaged.

(6.2.3) Debye sphere modelling of properdin

Small Debye sphere modelling was used to interpret the experimental scattering curve out to a Q of 1.6 nm^{-1}. To constrain the simulations, the unhydrated volume (66.8 nm^3) of properdin (Table 6.1) was calculated using standard crystallographic volumes for amino acids and carbohydrates (Chapters 2 & 3) for the 442 amino acid residues (Nolan *et al.*, 1991) and 28 presumed carbohydrate residues at two sites per monomer. This volume was subdivided into 128 cubes of side 0.8 nm per monomer in an array of 32 x 2 x 2 cubes for the simulations (total volume 65.5 nm^3). The total length of the monomer was 25.6 nm in agreement with electron microscopy. Overlapping spheres of the same volume as the cubes (radius 0.496 nm) were positioned at the centre of each cube. Neutron curve fitting was based on data in 0% ^2H_2O and in 80% and 100% $^2H_20$ to correspond to high positive and high negative solute-solvent contrasts respectively. A trial-and-error procedure was employed to test a wide range of different sphere models in order to determine which ones could be ruled out by the scattering data. Note that this method does not lead to the determination of a unique structure, although all the possible models were constrained by the total volume. The simulated neutron curves were smeared to correct for the contribution of wavelength spread and beam divergence in the experimental data (Chapter 3).

X-ray curve fitting is based on the hydrated volume. The total volume of 88.8 nm^3 per monomer was calculated as the sum of the dry volume and a hydration shell surrounding the protein based on a standard hydration of 0.3 g H_2O / g glycoprotein and an electrostricted water molecule volume of 0.0245 nm^3 in place of the free water volume of 0.0299 nm^3 (Chapter 3). This model was subdivided into 128 spheres as above, and the final Debye models were generated on the basis of cubes of side 0.885 nm and spheres of radius 0.549 nm^{-1}.

(6.2.4) Hydrodynamic modelling of properdin

Modelling of the sedimentation coefficient $s^o_{20,w}$ was based on the hydrated volume of 88.8 nm^3 as in the X-ray curve fitting. The $\bar{v}$ of properdin was calculated as 0.710 ml/g from its composition (Table 6.1), using the consensus volumes in Perkins (1986) which allow for glycoprotein hydration. Calculations of *f* were performed using a small number of non-overlapping spheres (Chapter 2, 3). The properdin monomer was formed from a line of 10 spheres with a sphere diameter of 2.644 nm, giving a total length of 26.44 nm, in agreement with the length from electron microscopy.

(6.2.5) Sequence alignments and secondary structure predictions

A total of 31 TSR sequences from 11 proteins were aligned manually and also with guidelines provided by the automatic multiple sequence alignment program MULTAL (Taylor, 1990). In the order shown in Figure 6.7 these were: human thrombospondin (3 TSRs) (Lawler & Hynes, 1986; Wolf *et al.*, 1990); mouse thrombospondin (2 of 3 TSRs) (Bornstein *et al.*, 1990); mouse properdin (6 TSRs) (Goundis & Reid, 1988); human properdin (6 TSRs) (Nolan *et al.*, 1991); thrombospondin-related anonymous protein (TRAP) from *Plasmodium falciparum* (1 TSR) (Robson *et al.*, 1988); the late complement components human C6 (3 TSRs) (DiScipio & Hugli, 1989; Chakravarti *et al.*, 1989; Haefliger *et al.*, 1989); human C7 (2 TSRs) (DiScipio *et al*, 1988), human C8α (2 TSRs) (Rao *et al.*, 1987); human C8β (2 TSRs) (Haefliger *et al.*, 1987; Howard *et al.*, 1987); human C9 (1 TSR) (DiScipio *et al.*, 1984; Stanley *et al.*, 1985); mouse C9 (1 TSR) and trout C9

(2 TSRs) (Stanley & Herz, 1987).

Secondary structure predictions based on the alignment of Figure 6.7 utilized the Robson and Chou-Fasman methods (Chapter 2, 3). The 31 predictions were averaged using the FORTRAN programs PREDRB7 and PREDCF (Chapter 3).

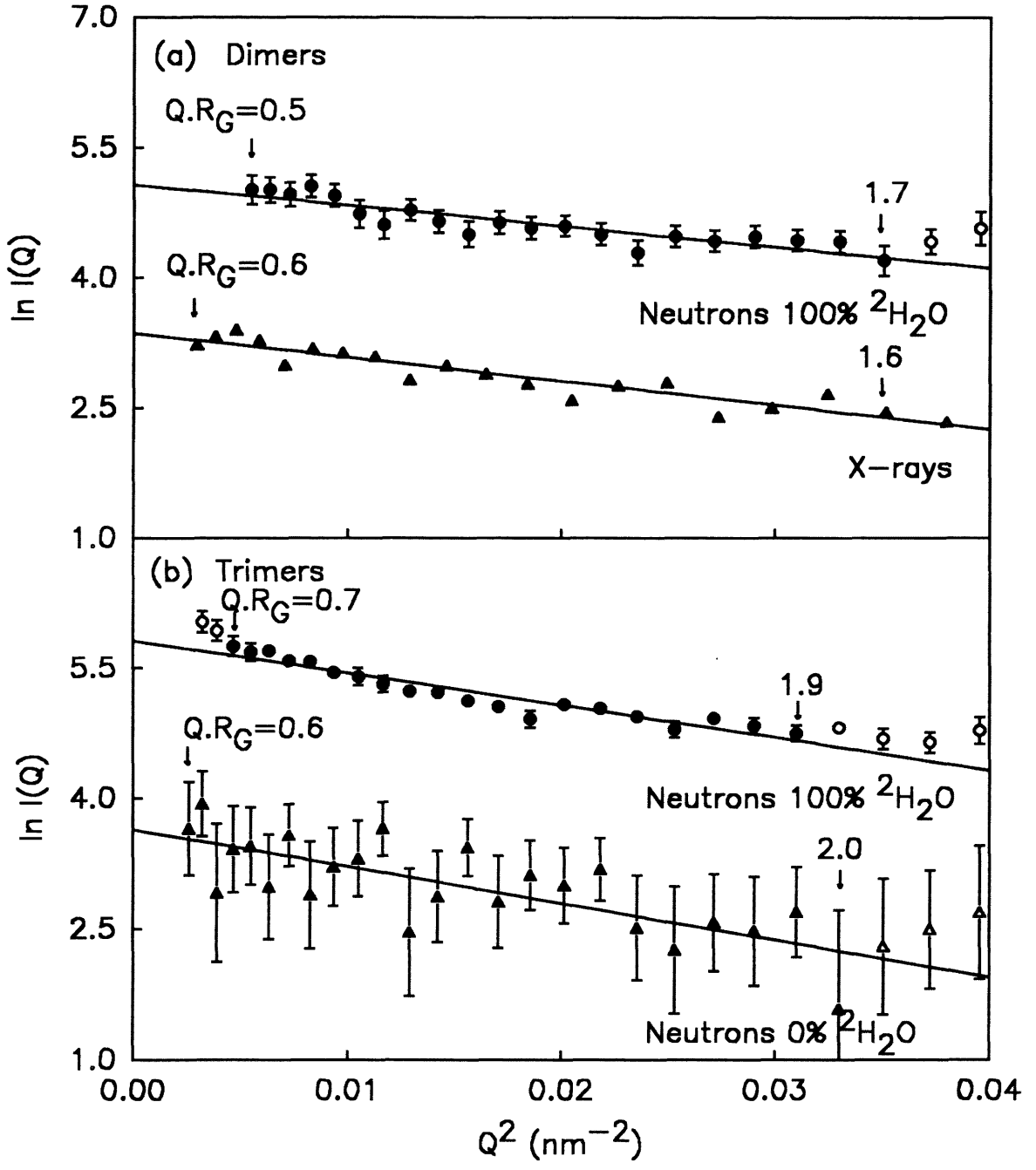


Figure 6.2. Guinier plots of properdin P_2 and P_3.

The regressed $Q.R_G$ range is arrowed and indicated by filled symbols.

(a) Data for P_2 in ^2H_2O buffers (o) by neutron scattering and in H_2O buffers (▵) by X-ray scattering, at concentrations of 0.51 mg/ml and 0.95 mg/ml respectively.

(b) Neutron data for P_3 in 100% (o) and 0% ^2H_2O (▵) at concentrations of 1.21 mg/ml and 1.14 mg/ml respectively.

(6.3) RESULTS AND DISCUSSION

(6.3.1) Neutron and X-ray scattering experiments of P_2 and P_3

In order to obtain solution structure data on the arrangement of the six TSRs in monomeric properdin, P_2, P_3 and P_4 were studied by neutron scattering in three contrasts at concentrations between 0.3 to 1.2 mg/ml. Scattering data were however only accessible for P_2 and P_3 on Instruments D11 and D17, since P_4 was found to possess too elongated a structure. In addition, P_3 could only be studied in 100% ^2H_2O buffers on D17 for reason of the lower neutron flux on D17, since only this buffer has a sufficiently low background. Guinier plots for P_2 and P_3 were linear over a Q range of 0.07 to 0.19 nm^{-1} (Figure 6.2) which gave acceptable $Q.R_G$ ranges of 0.5 to 1.4 (P_2) and 0.7 to 2.0 (P_3). Sample concentrations were determined using an A_{280} (1%, 1 cm) coefficient of 23.9 calculated (Chapter 3) from the human properdin amino acid and carbohydrate composition. It was assumed that there were two triantennary complex-type carbohydrate chains (28 residues) per properdin monomer, giving a total M_r of 54,100 and a 10.6% carbohydrate content by weight (Table 6.1). The latter was compatible with a biochemical determination of 9.8% carbohydrate (Minta & Lepow, 1974). By comparison, if the carbohydrates were biantennary, the M_r was 52,800 (8.3% carbohydrate by weight) and the A_{280} coefficient was 24.5. Both these A_{280} coefficients were higher than other estimates of 18 (Reid, 1981) and 17.1 (DiScipio, 1982), but were similar to a value of 23.4 (Nolan & Reid, 1990).

Analyses of the Guinier neutron $I(0)/c$ or R_G parameters showed that no dependences on protein concentrations or buffer compositions was found. As a further check of the Guinier data, the M_r for P_2 and P_3 were calculated from the $I(0)/c$ values measured in H_2O buffers. The M_r values of 80,000 ± 20,000 (P_2: 1 value) and 150,000 ± 30,000 (P_3: 2 values) were in satisfactory agreement with the expected values from the primary sequence of 108,200 and 162,300 respectively (Table 6.1). The calculated A_{280} coefficient from the primary structure was one-third higher than previously reported by biochemical determinations but this new value led to satisfactory M_r values by solution scattering. This revision also

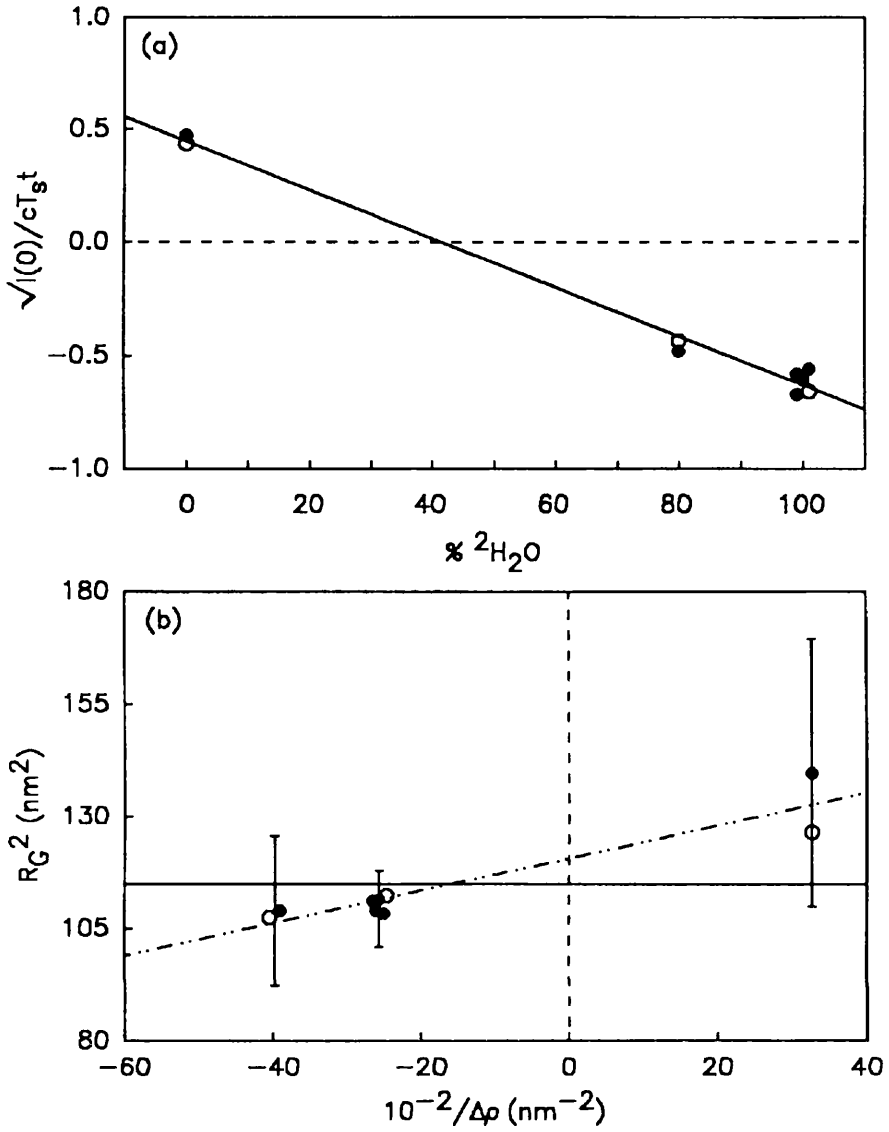


Figure 6.3. Neutron contrast variation analysis of P_3.

A concentration range of 0.5 to 1.2 mg/ml was used. Buffers were either 0.15 M-NaCl or 0.2 M-NaCl in phosphate (●) (Section 6.2), or 0.5 M-NaCl, 0.2 M-glycine and 12 mM phosphate, pH 7.4 (o).

(a) Matchpoint determination of P_3 from a graph of $\sqrt{I(0)}/cT_s$ against the volume percentage of ^2H_2O. Regression gave a matchpoint of 41.8 ± 1.9 % ^2H_2O.

(b) Stuhrmann plot of R_G^2 against the reciprocal solute-solvent contrast difference $\Delta\rho^{-1}$. The solid line corresponds to the mean R_G of 10.7 ± 0.5 nm. The dashed-dotted line corresponds to the regressed line and give an $R_{G\text{-}C}$ of 11.0 nm at zero $\Delta\rho^{-1}$ (dashed line) and a slope α of 4 x 10^{-3}. The error bars indicate one standard deviation of the Guinier analyses.

Residues			
	total amino acids		442
	total carbohydrate		28
	% mass carbohydrate		10.5
M_r Monomer			
	protein		48,400
	carbohydrate		5,700
	total		54,100
	neutron scattering		
		Dimer	80,000 ± 20000
		Trimer	150,000 ± 30000
Dry volume $(nm^3)^2$			66.8
Overall $\bar{v}$ $(ml/g)^2$			0.710
Overall $\Sigma b/M_r$	in H_2O (fm)		0.2377
	in ^2H_2O (fm)		0.4058
Predicted A_{280} (1%, 1cm)[3]			23.92
Literature A_{280}			17.1 to 23.4
Pred. matchpoint (% ^2H_2O)[4]			43.87
Exp. matchpoint (% ^2H_2O)			
		Dimer	39.3 ± 1.2
		Trimer	41.8 ± 1.9
Stuhrmann analysis mean R_G (nm)			
		Dimer	7.3 ± 1.6
		Trimer	10.7 ± 0.5
	R_{G-C}/R_O	Dimer	3.0 ± 0.7
		Trimer	3.8 ± 0.2
Models	R_G (nm)	Trimer	10.9

Table 6.1. Compositional and scattering data for properdin

Footnotes to Table 6.1

1, 2, 3 & 4: see Table 4.1

[3] Literature determinations of A_{280} are reported in Reid, 1981; DiScipio, 1982 and Nolan & Reid, 1990.

implies that previous estimates of the concentration of properdin in plasma, initially 20-25 μg/ml (Minta *et al.*, 1973), then 5.7 μg/ml (Pangburn, 1989), should now be 4.3 μg/ml. For P_3, contrast variation showed that its matchpoint was 41.8 ± 1.9% ^2H_2O (nine data points: Figure 6.3(a)). This agreed well with the prediction of 43.9% ^2H_2O from the amino acid and carbohydrate composition. For P_2, the limited availability of three data points in 0% and 100% buffers gave a comparable matchpoint of 39.3 ± 1.2% ^2H_2O. The linear Guinier plots and the above agreements with known compositions showed that the scattering curves corresponded to monodisperse P_2 and P_3 samples required for structural analyses.

The Stuhrmann plot of P_3 is shown in Figure 6.3(b). As a result of low concentrations, the statistical errors of the R_G data were high. Because of the large errors, the final R_G analysis was based on the mean R_G of 10.7 ± 0.5 nm, even though an $R_{G\text{-}C}$ of 11.0 ± 0.1 nm and a positive slope α of $4 \pm 1 \times 10^{-3}$ for P_3 could be determined from the nine values in Figure 6.3(b). For P_2, the mean R_G value was 7.3 ± 1.6 nm from three measurements. Calculation of the elongation ratio R_G/R_O (where R_O is the R_G of the sphere with the same dry volume as P_2 or P_3) gave values of 3.0 ± 0.7 and 3.8 ± 0.2 for P_2 and P_3 respectively. Since the R_G/R_O values for typical globular proteins are close to 1.28 (Perkins, 1988), both P_2 and P_3 possess highly extended structures in solution, in common with many other complement components (Perkins *et al.*, 1990c). If one molecular dimension is much larger than the other two as here, cross-sectional plots of ln $[I(Q).Q]$ against Q^2 should provide $[I(Q).Q]_{Q\to 0}$ and R_{XS} data. However, such plots were affected by poor signal-noise or buffer subtraction problems. The R_{XS} was estimated to be of the order of 0.6 ± 0.2 nm from data measured in 100% ^2H_2O buffers.

Synchrotron X-ray experiments were performed with P_2, P_3 and P_4 at concentrations of 0.29-0.89 mg/ml in order to extend the neutron analyses. Scattering curves were accepted for analyses if linear Guinier plots were obtained in Q ranges of 0.05-0.10 to 0.18-0.20 nm^{-1} (Figure 6.2(a)). As before, data for P_4 could not be analyzed. The mean R_G values were 9.1 ± 0.6 nm for P_2 (2 values) and 11.6 ± 0.9 nm for P_3 (4 values). The ratio of $I(0)/c$ values for P_2 : P_3 was 1.00 : 1.49 (±17%), which corresponded to the ratio of their relative molecular weights (1 : 1.5). These data were measured in high positive solute-solvent contrasts (Chapter 2), and therefore were comparable with the neutron R_G values in 0% ^2H_2O buffers. Good agreements were found with the neutron R_G of 9 ± 2 nm for P_2 (1 value) and 11.5 ± 0.4 nm for P_3 (2 values) in H_2O, thus corroborating these analyses. From time frame analyses of the P_3 data, the $I(0)$ values increased indicating on-going aggregation during beam exposure, but the R_G values remained constant within error. These data therefore were not utilised in the curve fitting analyses below.

(6.3.2) Debye sphere modelling of P_3

Scattering data for P_3 were compared with molecular models made from small Debye spheres. The length and possible arrangements of the TSRs and the properdin monomers in P_3 therefore could be deduced. Electron micrographs (Smith *et al.*, 1984b) suggested that properdin monomers were 26 nm in length and

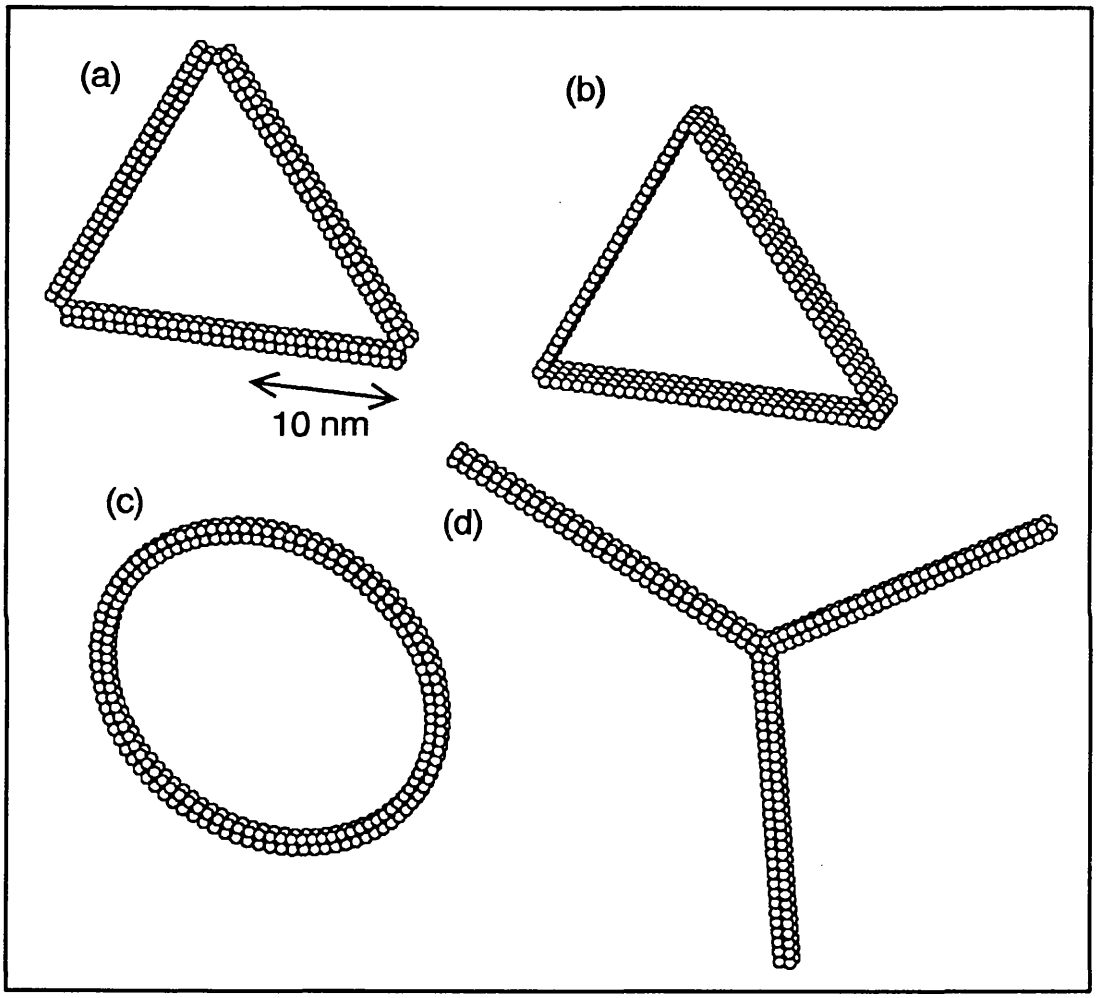


Figure 6.4. Low resolution models of properdin.

(a) The final Debye model for trimeric properdin with a cross-section of 2 x 2 spheres is shown (sphere diameter 0.8 nm, monomer of 32 x 2 x 2 spheres).

(b) This model has a similar triangular shape to (a) but with a cross-section of 1 x 4 spheres

(c) A ring of circumference of 96 spheres and a cross-section of 2 x 2 spheres

(d) A Y shape, again with a cross-section of 2 x 2 spheres.

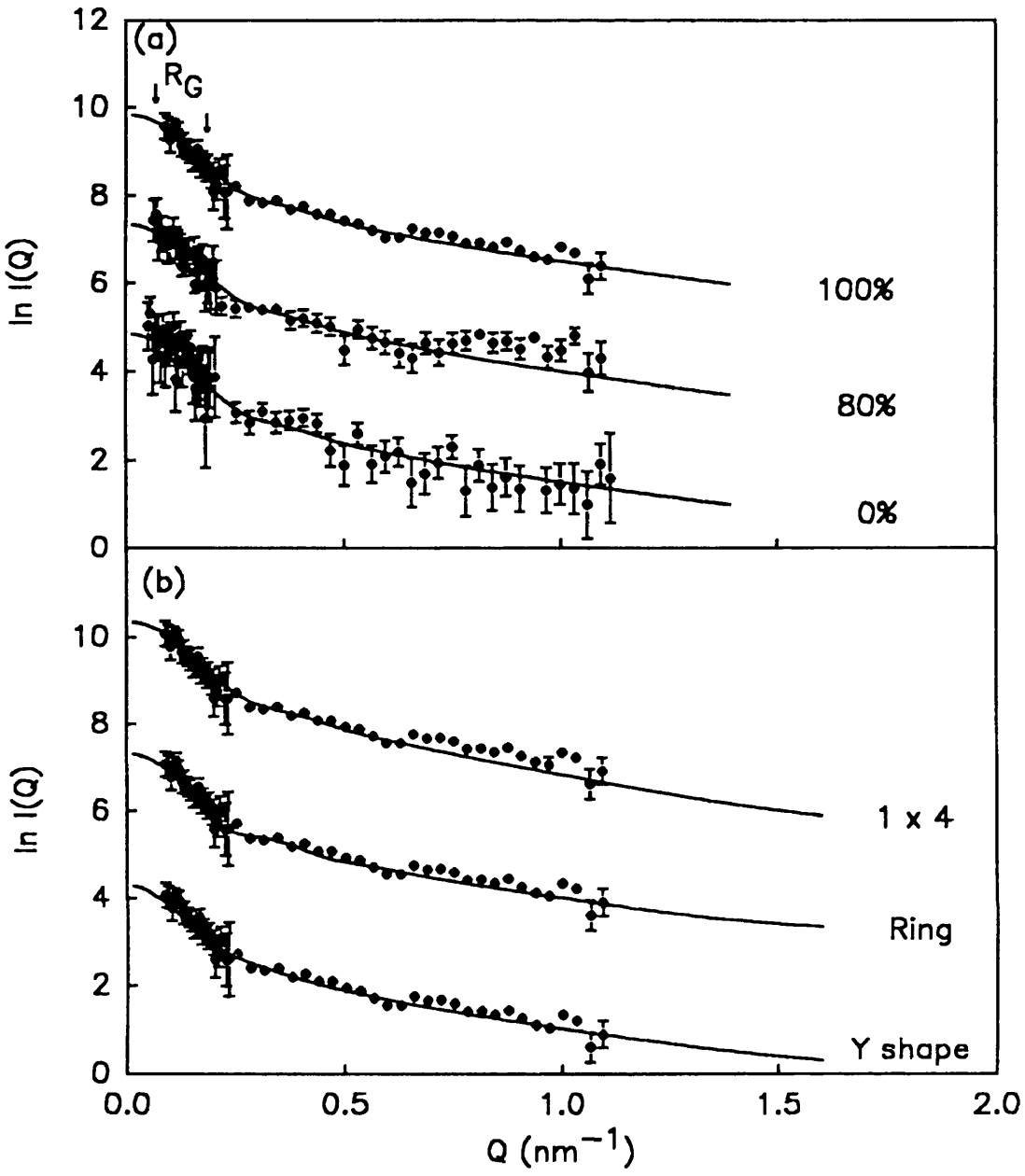


Figure 6.5. Neutron scattering curve simulations for P_3.

(a) The curve calculated from the model shown in Figure 6.4(a) is compared with the experimental curves measured in 100%, 80% and 0% ^2H_2O buffers. The simulated curves were corrected for beam divergence and wavelength spread. The sample concentrations are 0.43, 0.40 and 1.1 mg/ml respectively. Guinier R_G analyses were carried out in the Q range of 0.05 to 0.19 nm^{-1}. Error bars due to data counting statistics are shown.

(b) The experimental curve in 100% ^2H_2O is compared with three alternative models (models (b), (c) and (d) in Figure 6.4).

2.5 to 3 nm in width and associated as cyclic structures. Assuming the monomer to be cylindrical in shape the volume from these dimensions would be 128 to 184 nm^3 (π x (radius)^2 x length) which is 2 to 3 times larger than that calculated from the composition. The Debye sphere model for the monomer was 26 nm in length and the width was constrained by a dry volume close to 66.8 nm^3 (Section 6.2; Table 6.1).

P_3 was most successfully modelled as three linear rods, each of length 32 spheres and circular cross-section 2 x 2 spheres, that were arranged as the sides of an equilateral triangle (Figure 6.4(a)). This dry model had an R_G of 10.9 nm, in very good agreement with the observed mean neutron R_G of 10.7 ± 0.5 nm. The simulated neutron curves agreed well with the neutron data in 100%, 80% and 0% ^2H_2O (Figure 6.5(a)), with satisfactory R values of 0.026, 0.047 and 0.059 respectively. The structure of this model agreed well with the observed electron micrographs of P_3.

Three other cyclic and extended models for P_3 were tested against the solution data (Figure 6.5(b)). The use of an asymmetric cross-section of 1 x 4 spheres in the triangular model (Figure 6.4(b)) led to worsened curve fits at large Q values (R of 0.034 in place of 0.026) although the R_G was comparable at 10.7 nm. The use of a circular ring of circumference 96 spheres and cross-section 2 x 2 spheres (Figure 6.4(c)) led to an R_G of 12.6 nm that was significantly larger than the experimental value, and worsened curve fits in the Q range between 0.2 to 0.4 nm^{-1}, although the curve fits at large Q were similar (R of 0.030). Y-shaped models were tested in which the sides of the triangle were brought into contact with each other to result in a three-armed structure, such that the TSRs within each monomer were positioned in contact with TSRs in the neighbouring monomer (Figure 6.4(d)). Here, however, the R_G value was much increased to 15.0 nm, but the curve fits were similar (R of 0.031). These poorer curve fits and altered R_G values with alternative arrangements of properdin monomers supported the proposal of a triangle model as the likely solution structure for P_3.

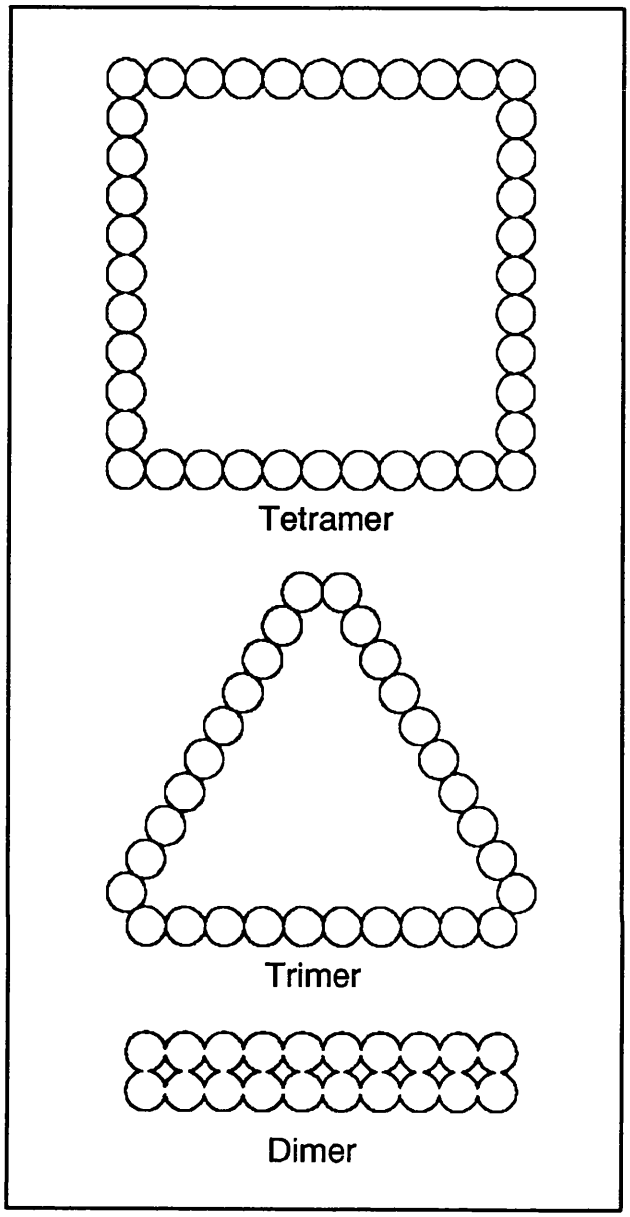


Figure 6.6. Hydrodynamic simulations of properdin.

Hydrodynamic models for tetrameric, trimeric and dimeric properdin are shown, using spheres of diameter 26.44 nm. These are based on a monomer of 1 x 10 spheres.

(6.3.3) Hydrodynamic modelling of properdin

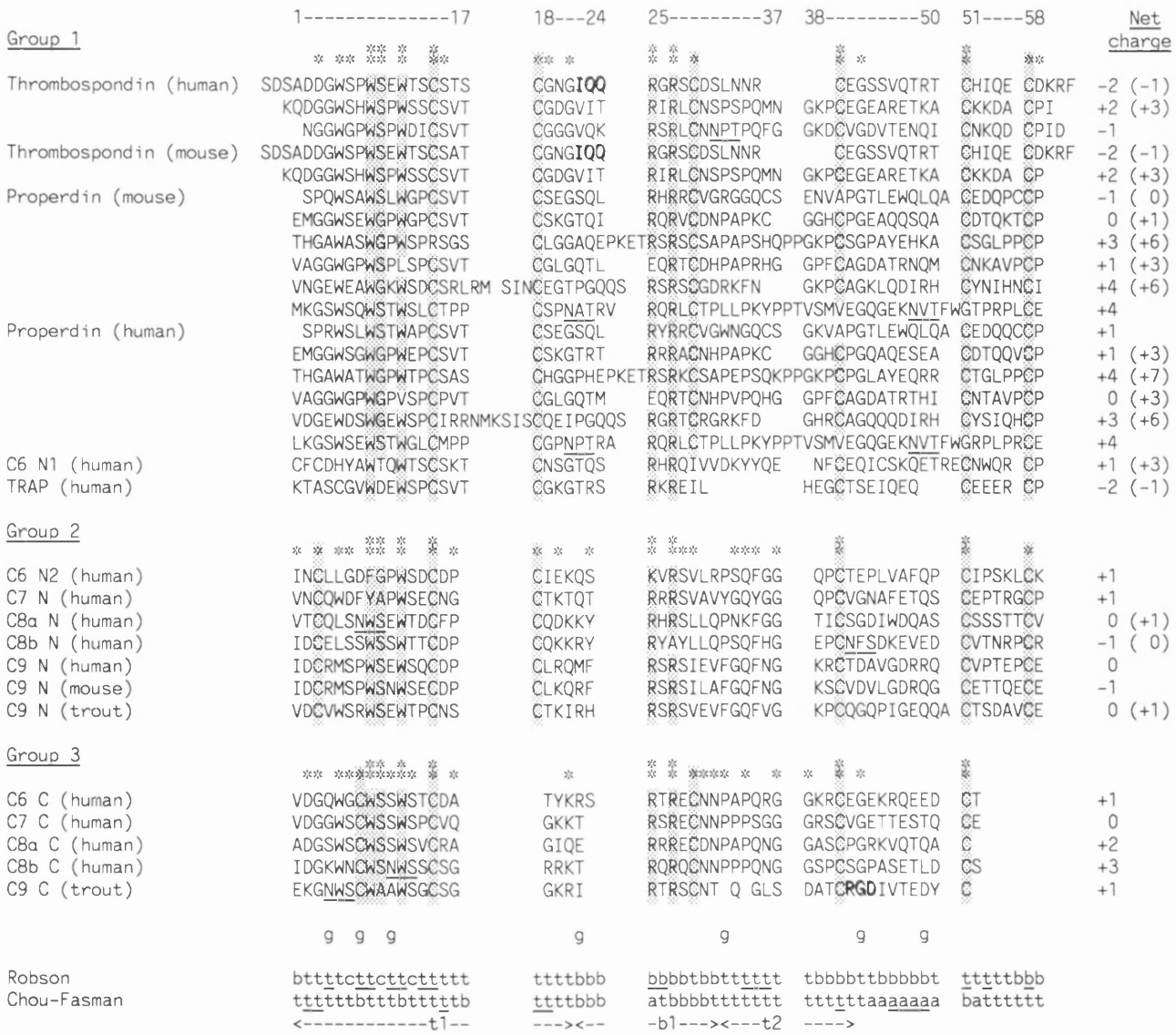
For properdin, sedimentation coefficients $s^{o}_{20,w}$ of 5.0 S (DiScipio, 1982), 5.2 S (Minta & Lepow, 1974; Minta & Kunar, 1976) and 5.1-5.3 S (Pensky *et al.*, 1968) have been reported. These authors were not able to resolve P_2, P_3 and P_4 by ultracentrifugation. These $s^{o}_{20,w}$ data provide independent structural data of the degree of elongation of the P_2, P_3 and P_4 structures in solution. Simulations were performed using hydrodynamic sphere models to calculate the $s^{o}_{20,w}$ values of the models determined by solution scattering. The best agreements were based on a monomer length of 26 nm, constructed from a line of 10 spheres of diameter 26.44 nm. The P_2, P_3 and P_4 models were represented by two parallel lines, an equilateral triangle, and a square respectively (Figure 6.6). The predicted $s^{o}_{20,w}$ values were 5.08 S, 4.98 S and 5.21 S respectively. The similarity of these values explained why the three forms cannot be distinguished by ultracentrifugation. These agreements were evidence that the length of the monomer was very close to that of 26 nm determined by solution scattering and electron microscopy. Calculations based on other lengths (Table 6.2) for the monomer showed that a typical error of ± 0.2 S in the simulation corresponded to changes in the monomer length of ± 2 nm.

Monomer length	$s^{o}_{20,w}$ Dimer (S)	$s^{o}_{20,w}$ Trimer (S)	$s^{o}_{20,w}$ Tetramer (S)
19 nm	5.86	6.14	6.58
26 nm	5.08	4.98	5.21
28 nm	4.96	4.80	5.01
30 nm	4.83	4.63	4.81

Table 6.2. Calculated sedimention coefficients for P_2, P_3 and P_4

(6.3.4) Classification of TSR sequences

In order to characterize the properties of a TSR within the properdin monomer, 31 full TSR sequences were aligned in Figure 6.7. This analysis was improved in comparison to that given in Perkins *et al.* (1989) because the intron/exon boundaries of the human and mouse thrombospondin genes (Wolf


1----------17
18---24
25---------37
38---------50
51----58
Net charge
Group 1
Thrombospondin (human) SDSADDGWSPWSEWTSCSTS CGNGIQQ RGRSCDSLNNR CEGSSVQTRT CHIQE CDKRF -2 (-1)
KQDGGWSHWSPWSSCSVT CGDGVIT RIRLCNSPSPQMN GKPCEGEARETKA CKKDA CPI +2 (+3)
NGGWGPWSPWDICSVT CGGGVQK RSRLCNNPTPQFG GKDCVGDVTENQI CNKQD CPID -1
Thrombospondin (mouse) SDSADDGWSPWSEWTSCSAT CGNGIQQ RGRSCDSLNNR CEGSSVQTRT CHIQE CDKRF -2 (-1)
KQDGGWSHWSPWSSCSVT CGDGVIT RIRLCNSPSPQMN GKPCEGEARETKA CKKDA CP +2 (+3)
Properdin (mouse) SPQWSAWSLWGPCSVT CSEGSQL RHRRCVGRGGQCS ENVAPGTLEWQLQA CEDQPCCP -1 (0)
EMGGWSEWGPWGPCSVT CSKGTQI RQRVCDNPAPKC GGHCPGEAQQSQA CDTQKTCP 0 (+1)
THGAWASWGPWSPRSGS CLGGAQEPKETRSRSCSAPAPSHQPPGKPCSGPAYEHKA CSGLPPCP +3 (+6)
VAGGWGPWSPLSPCSVT CGLGQTL EQRTCDHPAPRHG GPFCAGDATRNQM CNKAVPCP +1 (+3)
VNGEWEAWGKWSDCSRLRM SINCEGTPGQQS RSRSCGDRKFN GKPCAGKLQDIRH CYNIHNCI +4 (+6)
MKGSWSQWSTWSLCTPP CSPNATRV RQRLCTPLLPKYPPTVSMVEGQGEKNVTFWGTPRPLCE +4
Properdin (human) SPRWSLWSTWAPCSVT CSEGSQL RYRRCVGWNGQCS GKVAPGTLEWQLQA CEDQQCCP +1
EMGGWSGWGPWEPCSVT CSKGTRT RRRACNHPAPKC GGHCPGQAQESEA CDTQQVCP +1 (+3)
THGAWATWGPWTPCSAS CHGGPHEPKETRSRKCSAPEPSQKPPGKPCPGLAYEQRR CTGLPPCP +4 (+7)
VAGGWGPWGPVSPCPVT CGLGQTM EQRTCNHPVPQHG GPFCAGDATRTHI CNTAVPCP 0 (+3)
VDGEWDSWGEWSPCIRRNMKSISCQEIPGQQS RGRTCRGRKFD GHRCAGQQQDIRH CYSIQHCP +3 (+6)
LKGSWSEWSTWGLCMPP CGPNPTRA RQRLCTPLLPKYPPTVSMVEGQGEKNVTFWGRPLPRCE +4
C6 N1 (human) CFCDHYAWTQWTSCSKT CNSGTQS RHRQIVVDKYYQE NFCEQICSKQETRECNWQR CP +1 (+3)
TRAP (human) KTASCGVWDEWSPCSVT CGKGTRS RKREIL HEGCTSEIQEQ CEEER CP -2 (-1)
Group 2
C6 N2 (human) INCLLGDFGPWSDCDP CIEKQS KVRSVLRPSQFGG QPCTEPLVAFQP CIPSKLCK +1
C7 N (human) VNCQWDFYAPWSECNG CTKTQT RRRSVAVYGQYGG QPCVGNAFETQS CEPTRGCP +1
C8α N (human) VTCQLSNWSEWTDCFP CQDKKY RHRSLLQPNKFGG TICSGDIWDQAS CSSSTTCV 0 (+1)
C8b N (human) IDCELSSWSSWTTCDP CQKKRY RYAYLLQPSQFHG EPCNFSDKEVED CVTNRPCR -1 (0)
C9 N (human) IDCRMSPWSEWSQCDP CLRQMF RSRSIEVFGQFNG KRCTDAVGDRRQ CVPTEPCE 0
C9 N (mouse) IDCRMSPWSNWSECDP CLKQRF RSRSILAFGQFNG KSCVDVLGDRQG CETTQECE -1
C9 N (trout) VDCVWSRWSEWTPCNS CTKIRH RSRSVEVFGQFVG KPCQGQPIGEQQA CTSDAVCE 0 (+1)
Group 3
C6 C (human) VDGQWGCWSSWSTCDA TYKRS RTRECNNPAPQRG GKRCEGEKRQEED CT +1
C7 C (human) VDGGWSCWSSWSPCVQ GKKT RSRECNNPPPSGG GRSCVGETTESTQ CE 0
C8α C (human) ADGSWSCWSSWSVCRA GIQE RRRECDNPAPQNG GASCPGRKVQTQA C +2
C8b C (human) IDGKWNCWSNWSSCSG RRKT RQRQCNNPPPQNG GSPCSGPASETLD CS +3
C9 C (trout) EKGNWSCWAAWSGCSG GKRI RTRSCNT Q GLS DATCRGDIVTEDY C +1
g g g
g
g
g
g
Robson btttttcttcttctttttt ttttbbb bbbbtbbttttttt tbbbbttbbbbbt ttttttbbb
Chou-Fasman ttttttbtttbttttttb ttttbbb atbbbbttttttt ttttttaaaaaaaa battttttt
<----------t1--
--->
<--
-b1--->
<---t2
--->

et al., 1990; Bornstein *et al.*, 1990) and the human properdin gene (Nolan *et al.*, 1991) have been determined. These define the start and end of each TSR (Craik *et al.*, 1982; Traut, 1988). The start of the TSR has now been displaced by 10 residues compared to that in Perkins *et al.* (1989). The alignment was also improved by the use of the automated multiple sequence alignment program MULTAL (Taylor, 1990). In order to facilitate the averaging of structural properties (Taylor, 1986; Perkins *et al.*, 1988), gaps were minimized and homology was maximized between conserved or conservatively replaced residues. Figure 6.7 shows that 8 residues were greater than 90% conserved in the 31 TSRs (Trp-8, Ser-9, Trp-11, Cys-14, Arg-25, Arg-27, Cys-41, Cys-51).

The TSRs were readily subdivided into 3 groups on the basis of different Cys residue locations and different TSR lengths, in an extension of the initial survey of Patthy (1988). The closer similarities of the sequences within each group was confirmed by the MULTAL analyses using a range of sequence gap penalties. The principal characteristics of the three groups are summarized:

(i) Group 1, which included the thrombospondin and properdin sequences,

---

Figure 6.7. Alignment of 31 sequences for the thrombospondin repeat.
Residue conservation greater than 70% within a group of sequences is denoted by * above the sequences, and are shaded if these are Cys residues. Residues conserved to better than 90% in all 31 sequences are indicated by two * above the sequences, and all are shaded. Putative glycosylation sites are underlined and indicated by g beneath the sequences. IQQ and RDG peptides are shown in the first and fourth sequences, and the last sequence respectively. The averaged secondary structure prediction by the Robson and Chou-Fasman methods for the 31 sequences is shown beneath the sequences (b, β-sheet; a, α-helix; t, turn; c, coil). The letter is underlined if the prediction occurs in over 75% of the 31 sequences. To the right, the net charge found in each TSR is given assuming that the charged residues are D, E, R and K; the bracketed figures correspond to the assumption that H is charged also, which, at a blood pH of 7.4, is not likely to be significant.

and interestingly the first N-terminal TSR of C6, contained complex arrangements of between 4 to 7 Cys residues. These occurred most frequently at Cys-14, -18, -29, -41, -51 and -57. The consensus length of the TSR was 58 residues, of which 19 residues were conserved to better than 70%. The mean length of 19 sequences was 59.2 ± 3.8 residues. The properdin TSRs generally displayed 2 to 4 positive charges on average, depending on the ionisation state of about two His residues per TSR. Insertions or deletions in the Group 1 sequences led to five contiguous peptide segments and four gap/insertion regions as indicated in Figure 6.7, of which the first segment (residues 1-17) exhibited the highest residue conservation.

(ii) Group 2 corresponded to the N-terminal TSRs of the late complement components except for that above in C6. These contained 6 Cys as in Group 1, except that Cys-29 is replaced by Cys-3. Most notably, the well conserved Trp-5, Cys-29 and Gly-43 residues in Group 1 were poorly conserved in Group 2, while several well-conserved residues 33-37 in Group 2 do not have their counterpart in Group 1. The consensus length was 55 residues and the mean length (7 sequences) was 55.1 ± 0.4 residues, 23 of which were conserved to better than 70%. These TSRs were generally uncharged.

(iii) Group 3 corresponded to the C-terminal TSRs in the late components. This group contained 5 Cys residues at Cys-7, -14, -29, -41 and -51. This group exhibited the highest residue conservation and the shortest sequences; the consensus length was 48 residues and mean length (5 sequences) was 47.4 ± 1.5 residues, 25 of which were conserved to better than 70%. While Trp-5, Cys-29 and Gly-43 were as well-conserved as in Group 1, many conserved residues in Group 3 have no counterpart in Group 1. These TSRs exhibited low positive charges.

A possible connectivity scheme for the three principal disulphide bridges in TSRs may be proposed from the alignment. It is likely that Cys-18 is linked to Cys-57 since both were missing in Group 3 when compared to Groups 1 and 2. Inspection of the four remaining Cys residues showed suggestively that Cys-41 and Cys-51 were both missing in the last TSR of human and mouse properdin, and therefore these appeared to be bridged in other sequences. Accordingly Cys-14 may be linked either to Cys-29 (Groups 1 and 3) or Cys-3 (Group 2; and similar

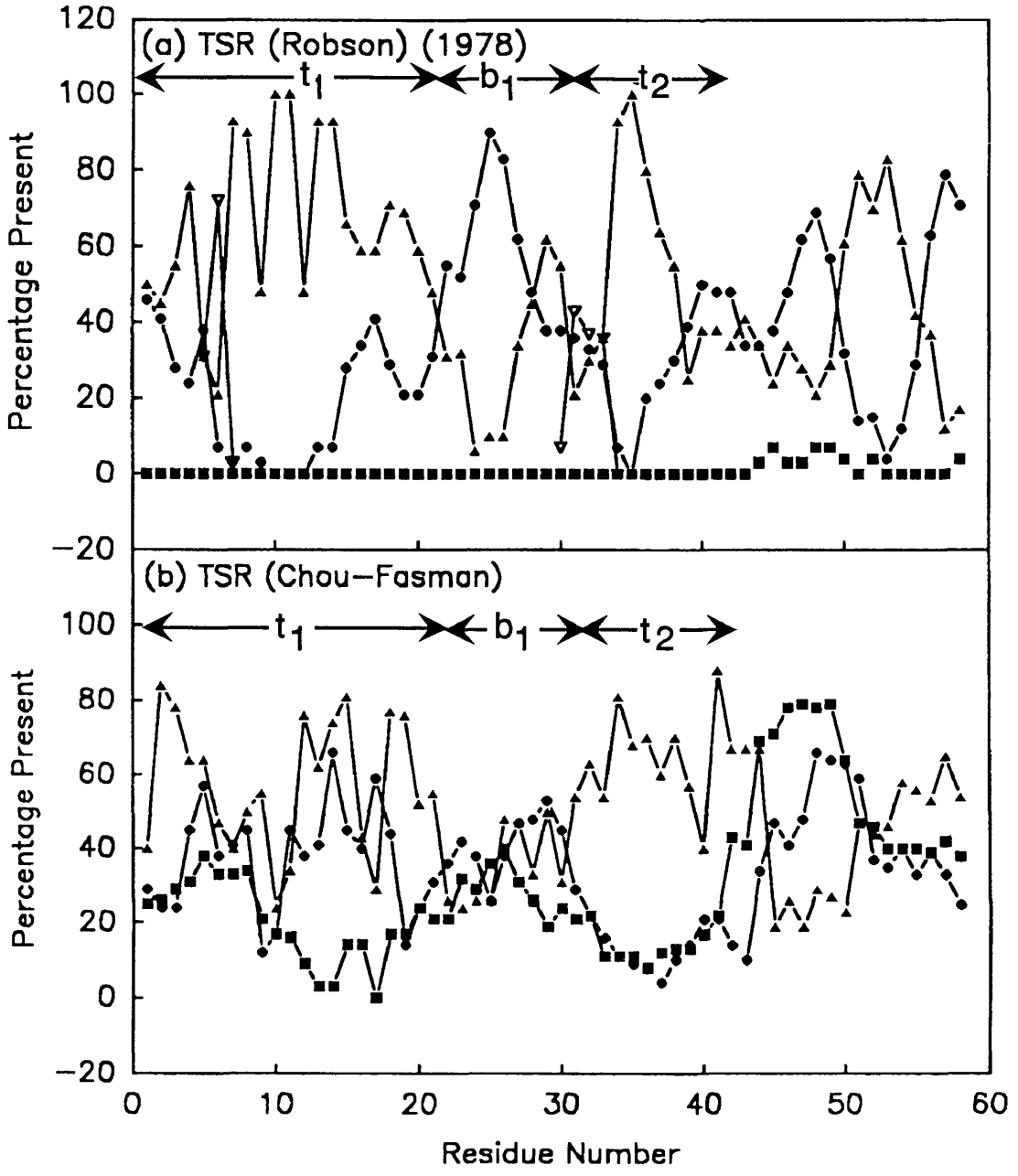


Figure 6.8. Secondary structure predictions of the TSR domain.

(a) The β-biased Robson prediction based on the 1978 (Garnier *et al.*, 1978) propensity tables. The coil predictions are only shown when they are the dominant structure.

(b) The Chou-Fasman prediction.

In both prediction methods, two regions of turn (t_1 and t_2; residues 1 to 22 and 32 to 42) and one region of β-sheet (b_1; residues 23 to 31) could be identified. The symbols are: α-helix (■), β-sheet (●), β-turn (▲) and coil (▽).

Cys residues in the C6 N1 and TRAP sequences of Group 1). This scheme implied that Cys-7 in Group 3 was unlinked within the TSR.

A total of 9 putative NxT or NxS glycosylation sites were identified. An IQQ motif (factor $XIII_a$ binding; Bale *et al.*, 1985) was found in human and mouse thrombospondin. An RGD motif (cell surface receptor binding; Yamada, 1989) was found in trout C9 (Stanley & Herz, 1987). The structural motif WSxWS was found in 14 TSRs, occurring once or twice in all the protein sequences of Groups 1, 2 and 3 except for those of TRAP and trout C9. This motif exists also in the cytokine haematopoietin receptor superfamily but the function of this motif has not been determined (Cosman *et al.*, 1990). Within several TSRs in Group 1 (but not in Groups 2 or 3), a motif CSVTCGxGxxxRxR was found which is thought to function as a sulphated glycoconjugate-binding domain in properdin (Holt *et al.*, 1990), although this motif was only conserved in full in the TRAP sequence and in three of the thrombospondin sequences. Prater *et al.* (1991) suggested that the VTCG peptide was a recognition site for a receptor and a sequence further downstream was the glycoconjugate-binding site. The 2^{nd} and 3^{rd} TSR peptides in thrombospondin contained the VTCG sequence and will attach to melanoma cells whilst the 1^{st} peptide, which does not contain this sequence, will not attach to these cells. The presence of sulphated glycoconjugates inhibited cell attachment and further indicated the presence of a two domain-binding site.

The average of the 31 secondary structure predictions was computed, based on the Robson (β biased) (Figure 6.8(a)) and Chou-Fasman methods (Figure 6.8(b)). Very similar predictions to those reported previously (Perkins *et al.*, 1989) were obtained. The total percentages of the four or three conformations present were identical. By the Robson method, these were: α-helix, 0%; β-sheet, 38%; turn, 57%; coil, 5%. The Chou-Fasman method gave: α-helix, 16%; β-sheet, 18%; turn, 66%. Two regions of high turn propensities could be identified (t_1 and t_2; residues 1-22 and 32-42) and one region of β-sheet (b_1; residues 23-31). While the two turn regions contained both hydrophilic and hydrophobic residues, nine out of the ten residues in the b_1 region were found to be hydrophilic. Further examination of the individual structure predictions supported the subdivision of

TSRs into three groups. Unlike Groups 1 and 3, residues 1 to 5 in Group 2 were predicted as β-sheet. The predictions suggested that the b_1 region had different lengths, namely residues 23 to 28 in Group 1, residues 19 to 33 in Group 2, and residues 22 to 26 in Group 3.

(6.4) CONCLUSIONS

(6.4.1) Structure of monomers in properdin

Despite the low properdin concentrations, the use of high flux beam sources (in particular at the ILL) led to successful measurements of solution scattering curves. These were the most fruitful for the trimer P_3, which is the most abundant oligomer. Curve fitting was most successfully carried out in terms of open triangular structures in solution conditions close to physiological, with overall dimensions that were highly similar to the length of 26 nm reported by electron microscopy. Flexibility between the TSRs in each properdin monomer cannot however be ruled out by the scattering data.

Properdin had a high isoelectric point greater than 9.5. This was readily correlated with the excess of positive charges (R, K) over negative charges (D, E) in five of the six TSRs in the human sequence (Figure 6.7). It is possible that these charges may offer an explanation of these open cyclic structures, since the TSRs may repel each other. The three forms of properdin P_2, P_3 and P_4 once isolated do not re-equilibrate into a mixture. The structural studies suggest that the monomers associate between alternate N-terminal and C-terminal peptides at each end of each monomer. Inspection of their sequences suggest that the stability of these oligomers may be mediated by ionic interactions between up to 22 opposite charges in these regions.

(6.4.2) Structure of TSRs in properdin

Structural information on TSRs have been deduced from this study. Since TSRs exist in single copies in the late complement components, and since the multiple TSRs in thrombospondin and five out of the six TSRs in properdin are encoded by separate exons, it can be inferred that each TSR constitutes an independently folded protein structure. While TSRs are the second most abundant complement sequence motif, analysis shows that these may be subdivided into three groups.

The TSRs in these three groups exhibit distinct differences in Cys contents, lengths and overall charges. Since similar locations for gaps and insertions could be proposed in the 31 TSR sequences (Figure 6.7), the protein backbone structure is expected to be similar in all three groups, even though the TSR in the late complement components is smaller. In support of this premise, only minor differences were seen between the Groups in the structure predictions.

To interpret the solution scattering analyses, the dry volume of each TSR can be calculated from amino acid compositions (Perkins, 1986). The average volume of the TSRs of Group 1 (excepting TRAP), Group 2 and Group 3 without carbohydrates were 8.2 ± 0.6 nm^3, 7.6 ± 0.6 nm^3 and 6.3 ± 0.3 nm^3 respectively. If the TSR were spherical, its diameter would be 2.5 nm, 2.4 nm and 2.3 nm respectively in the three Groups. The average volume for human and mouse properdin of the N- and C-terminal polypeptides was found to be 6.64 ± 0.18 nm^3 and 4.16 ± 0.08 nm^3 respectively. If spherical, these would have diameters of 2.4 and 2.0 nm respectively. The length of a properdin monomer is 26 nm. If each of the six Group 1 TSRs found in properdin are assumed to be spherical in shape, together with the N-terminal and C-terminal peptides, the sum of the eight diameters gives a total length of 19 nm if all eight structures were positioned in a linear arrangement. Since this length is notably shorter than the observed length, this shows that the TSR has an elongated structure. Its length can range between 3.25 nm to 4.3 nm, depending on the presumed structure of the N-terminal and C-terminal regions (Perkins *et al.*, 1989). On the basis of the scattering curve fits at large Q for properdin, the cross-sectional dimensions of the TSR are compact at average values of 1.7 nm x 1.7 nm. On this scale, each TSR is equivalent to 16 spheres (Figure 6.4). Similar overall dimensions are expected for the TSRs found in the late complement components.

CHAPTER 7.

STRUCTURAL STUDIES ON C9

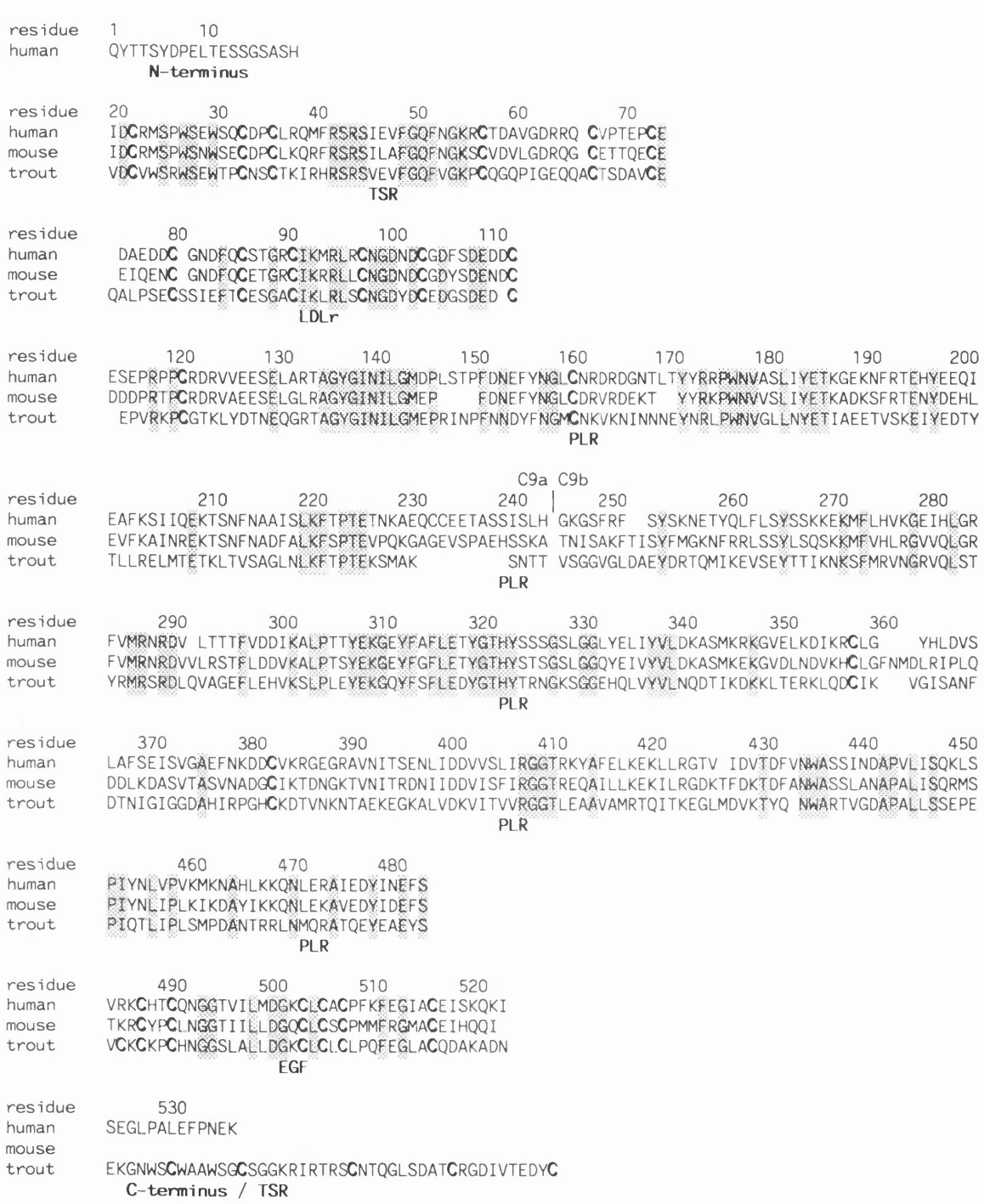


Figure 7.1. Alignment and domain structure of human, mouse and trout C9. Human (DiScipio *et al.*, 1984; Stanley *et al.*, 1985), mouse and trout (Stanley & Herz, 1987) sequences are aligned and residues homologous to all three sequences are shaded. A total of 29% of the residues are conserved in all three sequences with reference to the human sequence. The residue numbering follows the human sequence. The 4 domains of C9 as determined from sequence homologies i.e. the TSR, LDLr, PLR and the EGF are indicated. Full sequence and structural analyses of these domains can be found in Chapters 6 and 8.

(7.1) INTRODUCTION

The complement component C9 is the final protein involved in the lytic pathway of the cascade. The sequential addition of the late complement components C6, C7, C8 and several molecules of C9 (n = 1 to 18) to C5b leads to the formation of the membrane attack complex (MAC). Although the binding of a single C9 molecule to the C5b-8 complex is sufficient for lysis of the simple membrane bilayer of erythrocytes, at least two molecules of C9 are necessary to span the bacterial outer and inner membranes of *Escherichia coli* to cause cell lysis (MacKay & Dankert, 1990; Bloch *et al.*, 1987). Electron micrographs show that the MAC appears as a ring-like or cylinder-like pore structure penetrating the membrane. This creates open lesions through which cellular components can leak, thereby causing cell death (Borsos *et al.*, 1964; Podack *et al.*, 1982a, b; Tschopp *et al.*, 1982, 1984; Tschopp, 1984a; Dankert & Esser, 1985; Bhakdi & Tranum-Jensen, 1986; Biesecker *et al.*, 1991). It is thus clear that C9 is the major protein involved in the terminal stage of complement-mediated cell lysis.

C9 consists of a 537-residue polypeptide chain to which are attached two N-linked oligosaccharides (DiScipio *et al.*, 1984; Stanley *et al.*, 1985; Kontermann & Rauterberg, 1989). C9 contains three cysteine-rich domains. There is a thrombospondin repeat (TSR) and a low density lipoprotein receptor (LDLr) domain at the N-terminus, and an epidermal growth factor-like (EGF) domain at the C-terminus (Figure 7.1). The TSR and LDLr domains form a large part of the hydrophilic C9a fragment (M_r 28,000) formed by the cleavage of C9 into C9a and C9b by α-thrombin (Biesecker *et al.*, 1982: DiScipio *et al.*, 1984; Stanley *et al.*, 1985; Shiver *et al.*, 1986) (Figure 7.1). The central region of the C9 sequence has strong sequence homology with perforin (perforin-like region, PLR), a pore-forming protein secreted from cytotoxic T lymphocytes (Figure 7.1) (Shinkai *et al.*, 1988) and appears to possess some specific hydrophobic properties. Both C9 and perforin form large aqueous channels in membranes (internal diameter 10 nm for C9 and 16 nm for perforin) with little ion selectivity (Young *et al.*, 1986).

Other constituents of the MAC (C6, C7, C8α and C8β) have structural,

functional and antigenical similarities to C9. Comparison of the sequences show that, while C6, C7, C8α and C8β all contain the TSR, LDLr, PLR and EGF domains also found in C9, they also contain between one to six extra TSR, SCR and FIM domains at their amino and/or carboxy terminal ends (Chapter 1) (Haefliger *et al.*, 1989; Chakravarti *et al.*, 1989; DiScipio & Hugli, 1989). C9 is thus the archetypal protein structure whose molecular properties are likely to be present in other components of the MAC.

C9 undergoes a dramatic structural change upon polymerisation and insertion into lipid bilayers in which the N-terminal and C-terminal segments appear to unfold, doubling the length of the molecule (Podack & Tschopp, 1982a; Tschopp *et al.*, 1982; DiScipio & Hugli, 1985, Marazziti *et al.*, 1989) and forming neoantigenic sites (Dalmasso *et al.*, 1984). This unfolding appears to be catalysed by the interaction of C9 with the C5b-8 complex (Tschopp *et al.*, 1984). The insertion of C9 into the MAC is thought to occur in a two stage process. First, C9 binds to C8α in the C5b-8 complex, projecting partly into the membrane. The N-terminus of C9 projects into the cytoplasmic space whereas the C-terminus is not involved in either haemolysis or polymerization (Kontermann *et al.*, 1990). C9 then unfolds in a temperature dependent step (Marazziti *et al.*, 1989) and exposes more binding sites for C8α and C9, allowing C9 polymerization (Kontermann *et al.*, 1990). The sequential association of between 12 and 18 of these extended molecules forms a circular tubular polymer of M_r in excess of 10^6. Electron micrographs of tubular poly-C9 suggest that this assembly has a length of 16 to 21 nm, an outer diameter of 15 to 22 nm and an inner diameter of 9 to 11 nm (Podack & Tschopp, 1982a; DiScipio & Hugli, 1985) whereas monomeric, free C9 has dimensions of 7-8 nm x 5-5.5 nm (Podack & Tschopp, 1982a). An estimation of the channel diameter of poly-C9 by the liposome swelling assay gave a comparable value of 9 to 10 nm (Zalman & Müller-Eberhard, 1990) and the diameter of the MAC in erythrocyte ghosts by confocal laser-scanning microscopy gave a value 7 nm, corresponding to around 8 C9 molecules (Sauer *et al.*, 1991). The diameter of pores created by C5b-8 only were much smaller at 3 nm. The incorporation of 12 molecules of C9 into the MAC gives a similar pore diameter as poly-C9 (Zalman & Müller-Eberhard, 1990).

Three major hypotheses for the action of C9 on lipid bilayers currently exist. In one, it is proposed that C5b-9 complexes form a stable hollow polymeric structure in the bilayer and that leakage of cellular components through this causes cell death (the "doughnut" hypothesis) (Young & Young, 1990). Alternatively, the insertion of C9 destabilises the lipid bilayer surrounding the inserted polymer by a detergent-like mechanism (the "leaky patch" hypothesis) (Mayer, 1972; Esser *et al.*, 1979). Arguments both for and against these two hypothesis have recently been published (Esser, 1991; Bhakdi & Tranum-Jensen, 1991). A third possible mechanism for complement-mediated cell death is the dissipation of the membrane potential across the inner membrane. The insertion of the C5b-8 complex leads to a transient collapse in potential which becomes irreversible upon the incorporation of C9 (Dankert & Esser, 1986; Dankert & Esser, 1987). The presence of the C5b-8 complex also increases the permeability of the membrane. The binding of C8 to the C5b-7 complex is thought to initiate the redistribution of lipid between the inner and outer monolayers, affecting the physical properties of the membrane surface (Van der Meer *et al.*, 1989; Young & Young, 1990).

Small-angle scattering can give structural data on the arrangement of domains in multi-domain proteins, especially if the analyses are constrained by known volumes or known structures (Perkins, 1988). Here, the structure of C9 and its overall dimensions could be defined under conditions close to physiological. Interpretations of the scattering data were constrained by analyses of the sequences for the four domains in C9 (Chapters 6 & 8). Together with similar studies on C5 of complement (Chapter 5; Perkins *et al.*, 1990b), and complementary hydrodynamic modelling studies of C6, C7, C8 and C9, the two principal protein structures involved in MAC formation have been investigated by scattering. This opens the way for more detailed studies of the proteins and their complexes in the MAC.

(7.2) MATERIALS AND METHODS

(7.2.1) Preparation of C9

The C9 samples were prepared by Dr R.A. Harrison (Smith *et al.*, 1992). The final purified preparation of C9 contained a low level (< 5% by weight) of transthyretin (previously known as prealbumin) but no high molecular weight polymers of C9 could be detected.

Neutron scattering data were collected using buffers containing 0.2 M-NaCl, 20 mM-Na,K phosphate, 1 mM-EDTA, pH 7.0, or 0.5 M-NaCl, 20 mM-Na,K phosphate, 1 mM EDTA, pH 7.0. The neutron samples of C9 were dialysed at 6°C into their buffers containing 0%, 80% or 100% ^2H_2O for at least 36 hours with four buffer changes. X-ray data were collected based on C9 in buffers containing either 0.2 M-NaCl, 0.02 M-borate, 1 mM $CaCl_2$, 5 mM EDTA, pH 7.5 or 0.15 M-NaCl, 12.5 mM-phosphate, 0.5 mM-EDTA, pH 7.0.

(7.2.2) Neutron and X-ray scattering experiments

Neutron scattering data on C9 were principally obtained in two independent sessions on Instrument D17 at the ILL, Grenoble. Two further sessions on D17 were used to complete data collection on C9 in 100% ^2H_2O. Sample-to-detector distances of 3.46 m and 1.40 m and a main beam-detector angle of 0° and 20° were used with a neutron wavelength of 1.106 nm. The resulting Q range was 0.05 to 3.25 nm^{-1}. Individual sample counting times typically ranged between 3 to 21 min for C9 concentrations between 2.4 to 7.6 mg/ml in 100% ^2H_2O buffers, and between 26 min to 1.7 h for concentrations between 2.3 to 6.8 mg/ml in 0% ^2H_2O buffers.

X-ray scattering data were obtained using Station 8.2 at the SRS, Daresbury. Experiments were performed with beam currents of 99 to 130 mA and a ring energy of 2.00 GeV in two sessions. Samples were measured for 10 minutes at concentrations between 2 to 23 mg/ml. The use of a 500-channel quadrant

detector at sample-to-detector distances of 4.16 m and 3.45 m resulted in Q ranges of 0.11 to 1.61 nm^{-1} and 0.08 to 2.40 nm^{-1}. Data were only accepted if the subsequent Guinier plots were linear and led to reproducible data in repeated measurements.

Indirect transformation of the scattering data $I(Q)$ into real space $P(r)$ was carried out for both neutron and X-ray data. The R_G and length of C9 can be determined from the whole scattering curve measured out to Q of 3.25 nm^{-1} (neutrons) or 1.61 nm^{-1} or 2.40 nm^{-1} (X-rays). A total of 8 splines (neutrons and X-rays) were used to fit the experimental curves (83 and 113 data points respectively). The maximum dimension D_{max} of C9 for ITP was specified as 15 nm and 20 nm respectively (Chapter 2, 3). The number of points used to define $P(r)$ in real space was 101. The molecular length was determined from $P(r)$ when this became zero at large r.

(7.2.3) Debye sphere modelling of C9

The neutron and X-ray scattering curves for C9 were modelled to large Q using small Debye spheres (Chapters 2 & 3). To constrain the simulations, the unhydrated volume (83.4 nm^3) of C9 (Table 7.1) was calculated from the amino acid composition (Perkins, 1986). This volume was arbitrarily subdivided into 168 cubes of side 0.792 nm. Overlapping spheres of the same volume as a cube were placed at the centre of each cube. Neutron curve fitting was based on data in 0%, 80% and 100% ^2H_2O. The dependence of the scattering on the solute-solvent contrast was accounted for by dividing the spheres into an outer hydrophilic region and an inner hydrophobic core. Strictly, the hydrophilic and hydrophobic components should have been attributed matchpoints of 51% ^2H_2O and 34% ^2H_2O respectively but because of the simplicity of the model, values of 55% ^2H_2O and 35% ^2H_2O were used. The number of spheres assigned to each density was determined from the volume of the hydrophilic amino acids and carbohydrate, and hydrophobic amino acids in the sequence (Chapter 3). A wide range of models were tested on a trial-and-error basis to determine which ones were incompatible with the solution scattering data. Note that this method does not lead to the

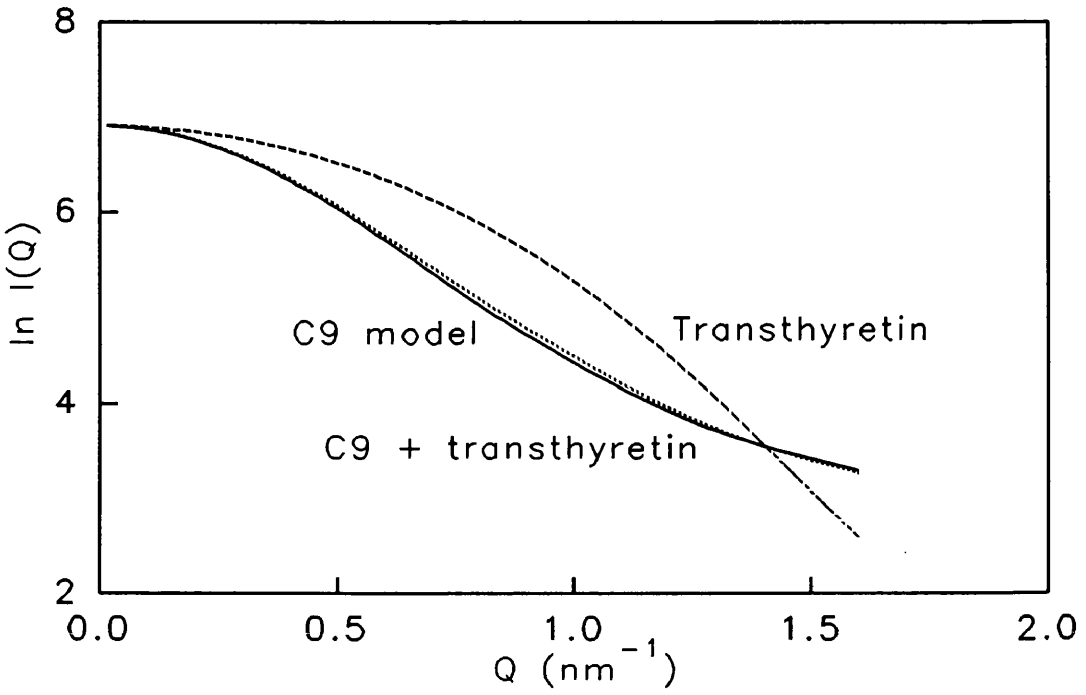


Figure 7.2. Effect of transthyretin on the C9 scattering curve.
The scattering curve for the C9 model (———) is compared with the scattering curve derived from the crystal co-ordinates of transthyretin (– – –) and a composite scattering curve of C9 and transthyretin, assuming a 5% contribution by transthyretin and 95% from the C9 model (·····). The small deviation of the C9 plus transthyretin curve from C9 model curve suggests that a contamination of 5% has no significant effect on the C9 scattering curve.

determination of a unique structure as the result of spherical averaging. Standard wavelength spread and beam divergence corrections were applied to the simulated neutron curves before these were compared with the experimental curves (Chapter 3). Since the C9 samples contained low amounts (< 5%) of transthyretin, control calculations using the X-ray crystallographic co-ordinates of transthyretin (Blake *et al.*, 1978; Brookhaven protein database reference 2PAB) were carried out to show that its effects on the scattering curve simulations of C9 were negligible (Figure 7.2).

X-ray curve fitting for C9 (in a high positive solute-solvent contrast) was based on the hydrated volume. The total volume of 110.5 nm^3 for the 168 Debye

spheres was obtained using a hydration of 0.3 g H_2O / g glycoprotein and an electrostricted water molecule volume of 0.0245 nm^3 (Perkins, 1986) (Chapter 2). The final Debye models were generated with spheres of diameter 1.079 nm (equivalent in volume to cubes of side 0.870 nm).

(7.2.4) Hydrodynamic modelling of C6, C7, C8 and C9

Modelling of the sedimentation coefficient $s^o_{20,w}$ for C6, C7, C8 and C9 was based on the hydrated volume as in the X-ray curve fitting. From sequences, the dry volumes of C6, C7 and C8 were calculated as 134.1 nm^3, 119.8 nm^3 and 193.1 nm^3 respectively, assuming that all the potential glycosylation sites were occupied with biantennary complex-type oligosaccharides. The calculated hydrated $\bar{v}$ for C6, C7, C8 and C9 were 0.720 ml/g, 0.719 ml/g, 0.721 ml/g and 0.722 ml/g respectively (Perkins, 1986). If a small number of hydrodynamic spheres were used, this hydration was replaced by a notional value of 0.39 g H_2O/ g glycoprotein in order to compensate for the volume lost in the void spaces between the non-overlapping spheres required for the simulations of *f* (Chapter 3). Alternatively, the hydrated X-ray model for C9 (168 non-overlapping spheres of diameter 0.870 nm) was employed directly in the simulations. Calculations were performed by the modified Oseen tensor procedure using the program FRCC (Chapters 2 & 3).

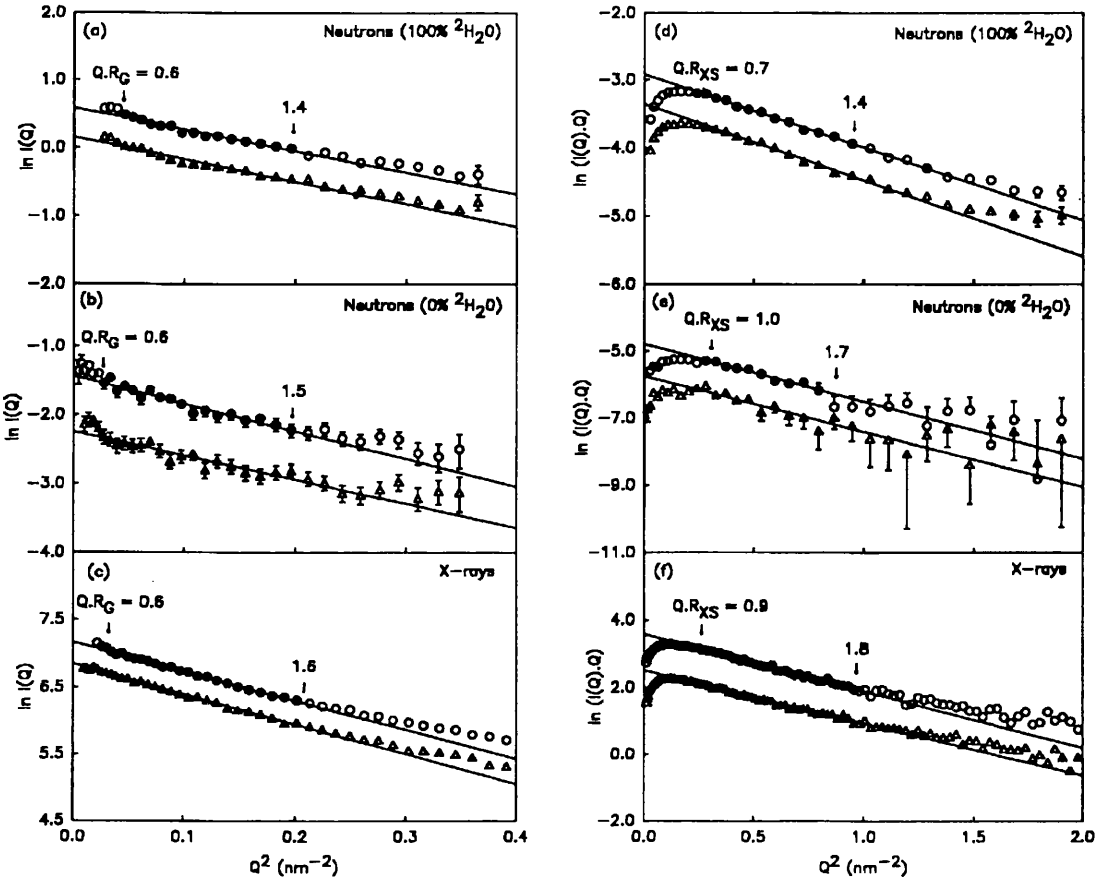


Figure 7.3. Guinier analyses of C9 by neutron and X-ray scattering.

The $Q.R_G$ and $Q.R_{XS}$ ranges used for the R_G and R_{XS} determinations are arrowed and indicated by filled symbols. The neutron data were obtained in 0.5 M NaCl buffers, while the X-ray data were obtained in 0.2 M NaCl buffers (Section 7.2). The X-ray curves in (c) and (f) are displaced on the vertical axis for clarity.

(a) and (d). Neutron R_G and R_{XS} Guinier plots respectively for C9 in 100% ^2H_2O buffers at sample concentrations of 7.3 (o) and 4.7 mg/ml (Δ).

(b) and (e). Neutron R_G and R_{XS} Guinier plots respectively for C9 in 0% ^2H_2O buffers at sample concentrations 6.9 (o) and 4.6 mg/ml (Δ).

(c) and (f). X-ray R_G and R_{XS} Guinier plots respectively for C9 in H_2O buffers (after exposure times of 3 min) at a sample concentration of 7.7 mg/ml.

(7.3) RESULTS AND DISCUSSION

(7.3.1) Neutron scattering studies of C9

Neutron scattering was used to determine the gross and internal solution structure of C9 from data in three contrasts at concentrations between 2.4 to 7.6 mg/ml. Some of the earlier C9 samples contained small amounts of non-specific aggregates, identifiable in the scattering analyses by non-linear Guinier plots. These were successfully eliminated by:

(a) leaving the final gel-filtration stage until just prior to dialysis, which was then followed immediately by neutron data collection;

(b) the use of phosphate buffers in 0.5 M NaCl rather than 0.2 M NaCl and

(c) the use of 0% ^2H_2O rather than 100% ^2H_2O buffers.

Figures 7.3(a) and 7.3(b) show that the resulting neutron Guinier plots for C9 in 100% ^2H_2O and 0% ^2H_2O with 0.5 M NaCl were linear over an acceptable $Q.R_G$ range of 0.6 to 1.5.

Where measurable, the Guinier intensity ($I(0)/c$) and R_G parameters showed no dependence on C9 concentrations between 2.2 to 7.7 mg/ml or buffer compositions in 0.2 M NaCl or 0.5 M NaCl (data not shown). The C9 concentrations were determined using an A_{280} coefficient of 8.37 (1%, 1 cm) (Table 7.1) calculated (Perkins, 1986) from the human sequence (DiScipio *et al.*, 1984; Stanley *et al.*, 1985), assuming that there were 2 triantennary complex-type carbohydrate chains per C9 molecule (Kontermann & Rauterberg, 1989). This composition resulted in an M_r of 66,400, which was comparable with the M_r of 71,000 from SDS-PAGE (Biesecker & Müller-Eberhard, 1980). The 8.6% carbohydrate content by weight was in good agreement with experimental values of 7.8-8.0% (Biesecker & Müller-Eberhard, 1980; DiScipio & Hugli, 1985). The predicted A_{280} value of 8.37 was however 11%-18% lower than literature values of 9.6 (Podack & Tschopp, 1982a) and 9.88 (Monahan *et al.*, 1983; Stewart *et al.*, 1984) based on quantitative amino acid analyses. The M_r for C9 was also calculated from the $I(0)/c$ values measured in H_2O buffers to check that the C9 preparations were monomeric (Jacrot & Zaccai, 1981). The M_r of C9 was

Residues	
total amino acids	537
total carbohydrate	28
% mass carbohydrate	8.6
M_r	
protein	60,700
carbohydrate	5,700
total	66,400
neutron scattering	63,500 ± 3,000
Dry volume (nm^3)[1]	83.4
Overall $\bar{v}$ (ml/g)[2]	0.722
Overall $\Sigma b/M_r$ in H_2O (fm)	0.231
in ^2H_2O (fm)	0.640
Predicted A_{280} (1%, 1cm)[3]	8.37
Literature A_{280}	9.6 to 9.88
Pred. matchpoint (% ^2H_2O)[4]	42.3
Exp. matchpoint (% ^2H_2O	
from $I(0)$	42.9 ± 0.3
from $[I(Q).Q]$	43.6 ± 0.5
Stuhrmann analysis	
R_{G-C} (nm)	3.33 ± 0.03
α_G (x 10^{-5})	35 ± 7
R_{XS-C} (nm)	1.66 ± 0.02
α_{XS} (x 10^{-5})	16 ± 2
Models R_{G-C} (nm)	3.37
α_G (x 10^{-5})	35

Table 7.1. Compositional and scattering data for C9.

Footnotes to Table 7.1

1, 2, 3 & 4: see Table 4.1.

[3] Literature determinations of A_{280} are reported in (Podack & Tschopp, 1982a) and (Monahan *et al.*, 1983; Stewart *et al.*, 1984).

determined to be 63,500 ± 3,000 (5 values) from the neutron data, which was in excellent agreement with the value of 66,400 from the composition (Table 7.1). Had the A_{280} values from experimental amino acid analyses been used, the M_r would have been overestimated as 71,000 to 76,000. The neutron contrast variation experiments shown in Figure 7.4(a) resulted in a matchpoint of 42.9 ± 0.3% ^2H_2O. This was in good agreement with the prediction of 42.3% ^2H_2O from the amino acid and carbohydrate composition (Table 7.1). In summary, the linear Guinier plots and the agreements with M_r and matchpoint data showed that the scattering curves corresponded to monodisperse preparations of C9, as required for structural analyses.

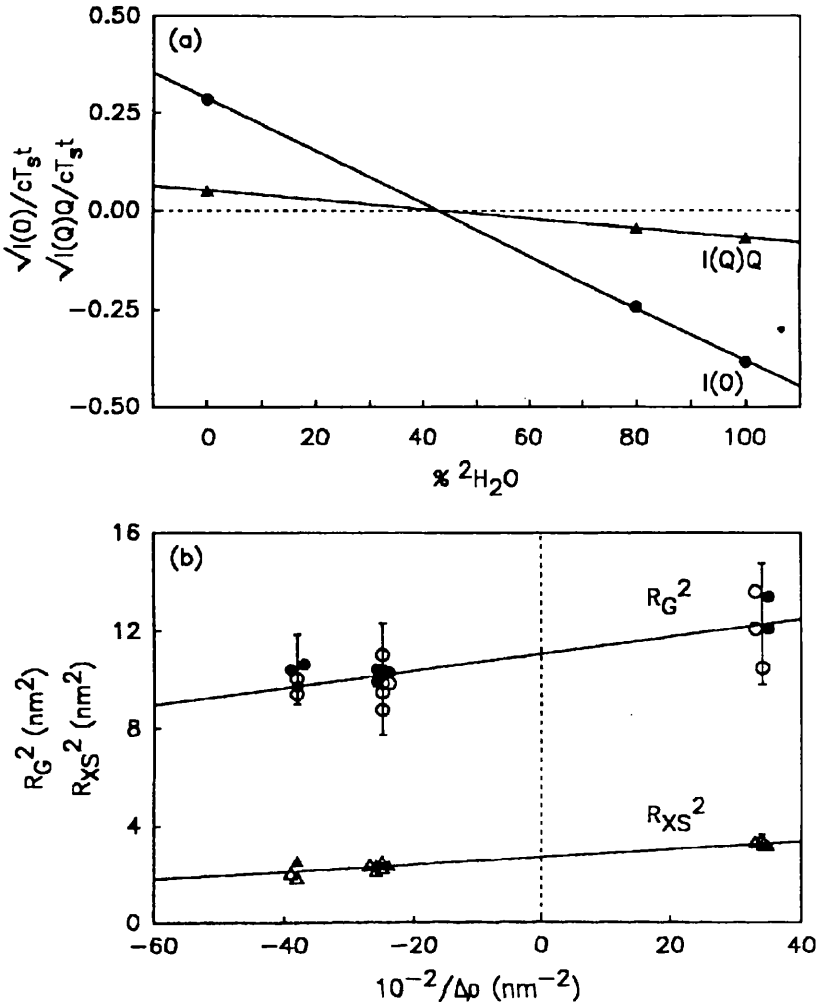


Figure 7.4. Neutron contrast variation analyses of C9.

A concentration range of 7.6 to 2.4 mg/ml was used. Buffers were either 0.2 M-NaCl or 0.5 M-NaCl in 12 mM phosphate, pH 7.4. In (a) and (b), statistical error bars were shown only when these were large enough to be visible.

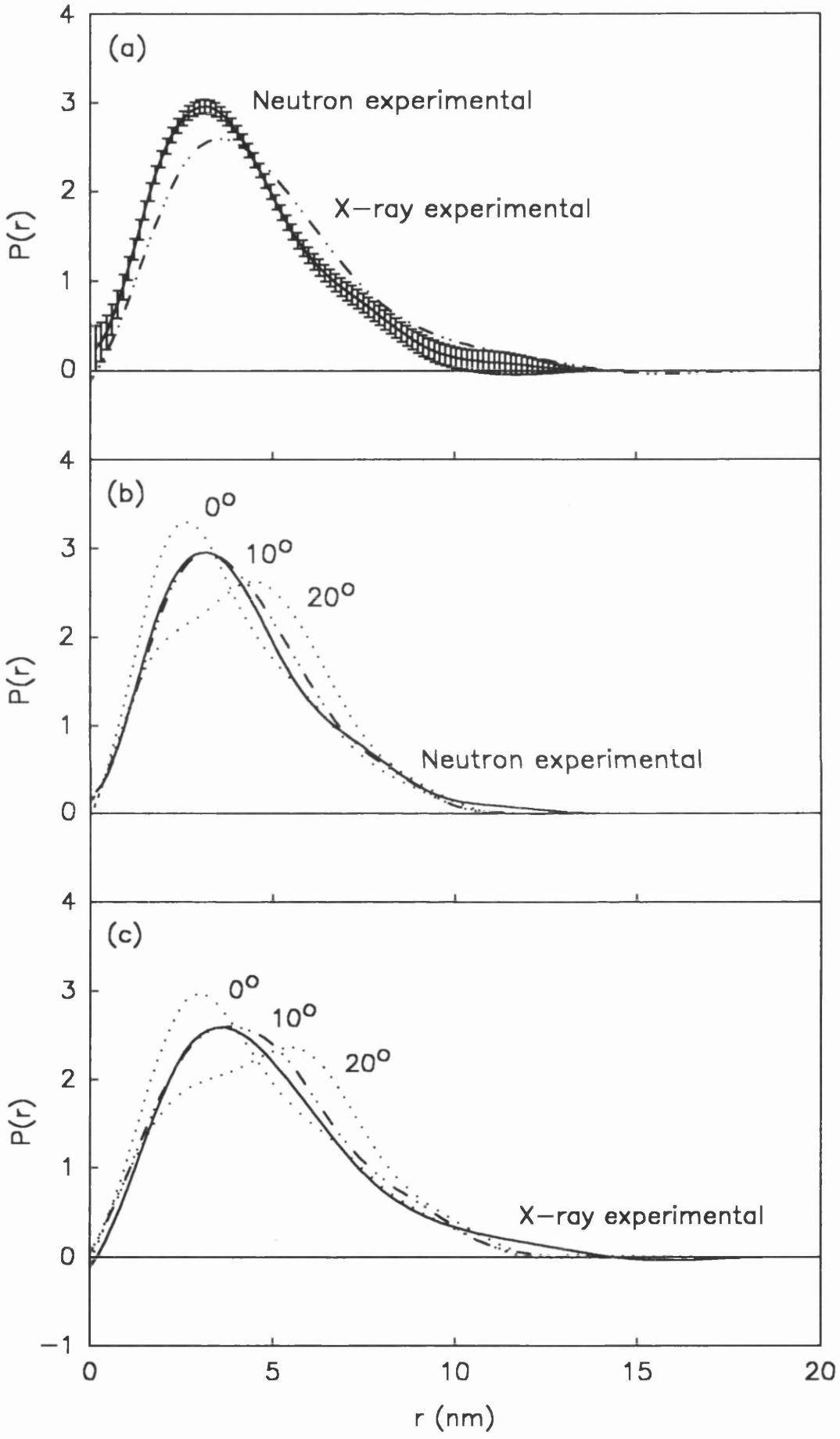
(a) Matchpoint determinations of C9 from graphs of $\sqrt{I(0)/ctT_s}$ (●) and $\sqrt{[I(Q).Q]_{Q\to 0}/ctT_s}$ (▲) against the volume percentage of ^2H_2O. The matchpoints were 42.9 ± 0.3 % ^2H_2O (17 points) and 43.8 ± 0.5 ^2H_2O (11 points) respectively. The mean data points only are shown for reason of clarity.

(b) Stuhrmann plot of R_G^2 against the reciprocal solute-solvent contrast difference $\Delta\rho^{-1}$. The two buffers used are indicated in the Guinier and cross-sectional plots by o and ▵ (0.2 M NaCl) or ● and ▲ (0.5 M NaCl) respectively. The $R_{G\text{-}C}$ at zero $\Delta\rho^{-1}$ was 3.33 ± 0.03 nm and the slope α_G was 35 (± 7) x 10^{-5} (19 points). The $R_{XS\text{-}C}$ was 1.66 ± 0.02 nm at zero $\Delta\rho^{-1}$ and the slope α_{XS} was 16 (± 2) x 10^{-5} (13 points).

The Stuhrmann analysis of the dependence of the neutron R_G data in 0.2 M and 0.5 M NaCl on the solute-solvent contrast resulted in an $R_{G\text{-}C}$ of 3.33 ± 0.03 nm at infinite contrast and a positive slope α_G of 35 (± 7) x 10^{-5} for C9 (Figure 7.4(b)). Calculation of the elongation ratio $R_{G\text{-}C}/R_O$ (where R_O is the R_G of the sphere with the same dry volume as C9) gave a value of 1.59 ± 0.01. Since the $R_{G\text{-}C}/R_O$ values for typical globular proteins are close to 1.28 (Perkins, 1988), C9 possesses an extended solution structure. Because C9 is elongated, the cross-sectional parameters $[I(Q).Q]_{Q\to 0}$ and R_{XS} can be obtained from Guinier analyses of ln $[I(Q).Q]$ against Q^2 (Pilz, 1982). As shown in Figures 7.3(d) and 7.3(e), linear cross-sectional plots were readily obtained in an acceptable $Q.R_{XS}$ range of 0.7 to 1.8. Regression of the data in Figure 7.4(a) gave a matchpoint of 43.6 ± 0.5% ^2H_2O, in agreement with the theoretical value of 42.5% ^2H_2O. The cross-sectional radius of gyration at infinite contrast $R_{C\text{-}XS}$ was 1.66 ± 0.02 nm and the slope α_{XS} was again positive at 16 (± 2) x 10^{-5} (Table 7.1).

Approximate molecular dimensions can be determined using simple shape assumptions. By assuming C9 to resemble an elliptical cylinder, the length was calculated as 10.0 ± 0.2 nm from the $R_{G\text{-}C}$ and $R_{XS\text{-}C}$ values (Chapter 3). Use of the dry volume of 83.4 nm^3 as a constraint resulted in axial dimensions of 10.0 nm x 6.4 nm x 1.7 nm for C9. The cross-section of C9 was therefore elongated also. The determination of the length from the intensity ratio, with no shape assumptions, gave a value of 10.1 ± 0.8 nm (13 values), in good agreement with that above.

The values of α_G and α_{XS} can be compared with those published for other glycoproteins once these have been normalised on the basis of its proportionality to R_G^2 and R_{XS}^2. Using α_G values for lysozyme, myoglobin, trypsin, α_1-acid glycoprotein, α_1-antitrypsin and $C\bar{1}$ inhibitor (Stuhrmann & Fuess, 1976; Ibel & Stuhrmann, 1975; S.J. Perkins, unpublished data; Perkins *et al.*, 1985, 1990d; Smith *et al.*, 1990), α_G for C9 was predicted to be of the order 32 ± 17 x 10^{-5}. This was in good agreement with the observed α_G of 35 ± 7 x 10^{-5} (Figure 7.4(b)). Using α_{XS} values for α_1-antitrypsin, C4, C4u, C4b, C5 and factor H ($R_{XS\text{-}2}$) (Smith *et al.*, 1990; Perkins *et al.*, 1990b, 1990a, 1991a), the predicted value of α_{XS} for C9


(a)
Neutron experimental
X−ray experimental
P(r)
(b)
0°
10°
20°
Neutron experimental
P(r)
(c)
0°
10°
20°
X−ray experimental
P(r)
r (nm)

would be close to 13 ± 5 x 10^{-5}. This was also in good agreement with the observed α_{XS} for C9 of 16 ± 2 x 10^{-5}. $C\bar{1}$ inhibitor (Perkins *et al.*, 1990d) has an exceptionally high α_{XS}, however this has an unusually high content (26% by mass) of extended O-linked and N-linked oligosaccharides located along the long axis of this structure. These comparisons show that C9 contains a similar radial distribution of hydrophilic surface residues and hydrophobic core residues as that found in other soluble glycoproteins.

Use of ITP to convert the scattering curve $I(Q)$ in reciprocal space into its distance distribution function $P(r)$ in real space was applied to three full curves for C9 at concentrations of 7.3, 4.7 and 2.2 mg/ml in 0.5 M NaCl buffers in 100% ^2H_2O. Using the ITP parameters specified in Figure 7.5(a), the R_G and $I(0)$ values were stable over a Lagrange multiplier range of 0 to -8 for all three curves. The

---

Figure 7.5. Distance distribution functions $P(r)$ calculations based on the neutron and X-ray $I(Q)$ data for C9.

(a) The neutron $P(r)$ curve corresponds to the solid line with statistical error bars, using a C9 sample at 7.3 mg/ml in a 100% ^2H_2O buffer with 0.5 M NaCl. The neutron maximum length of C9 lies between 10.5-14 nm, based on a calculation with 8 splines and D_{max} of 15 nm. The X-ray $P(r)$ curve is depicted by the dashed line; this corresponds to C9 at 7.4 mg/ml in a buffer with 0.2 M NaCl. The X-ray maximum length of C9 is estimated as 14 ± 2 nm.

(b) Comparison of the neutron experimental $P(r)$ curve (solid line) with three simulated $P(r)$ curves for two-density V-shaped models of C9 in which the angle between the arms is set as 0°, 10° and 20° (dotted or dashed lines), and the densities are set to correspond to negative solute-solvent contrasts. The maximum of $P(r)$ is located at r = 2.6 nm, 3.2 nm and 4.5 nm for angles of 0°, 10° and 20° respectively.

(c) Comparison of the X-ray experimental $P(r)$ curve (solid line) with the same models for C9 as in (b), now hydrated and density weighted for positive solute-solvent contrasts. The maximum of $P(r)$ is located at r = 3.0 nm, 3.6 nm and 5.4 nm for angles of 0°, 10° and 20° respectively.

R_G and $I(0)$ values agreed well with the Guinier plots for the respective curves (Table 7.2). The maximum $P(r)$ in the three $P(r)$, which corresponded to the most frequently occurring distance r within C9 was 3.00 to 3.15 nm. The error bars in the calculation of $P(r)$ at large r showed that the maximum dimension of C9 was in a range between 10.5 and 14 nm. These dimensions were comparable to but larger than the two estimates of the length of 10.1 to 10.2 nm above from the Guinier analyses. The difference was attributed to the difficulty in determining a precise cut-off when $P(r)$ approached zero at large r.

Curve	Guinier		ITP	
	R_G (nm)	$I(0)$	R_G (nm)	$I(0)$
1	3.08	1.791	3.32	1.837
2	3.14	1.168	3.35	1.190
3	2.96	0.526	3.15	0.524

Table 7.2. Comparison of ITP and Guinier parameters

(7.3.2) X-ray scattering experiments of C9

Synchrotron X-ray experiments were performed with C9 to extend the neutron analyses. C9 was measured at concentrations of 1.8-22.9 mg/ml. However it was found that irradiation by X-rays induced substantial aggregate formation during the 10 min exposure time as indicated by the increase in R_G and $I(0)$ (Figure 7.6). Besides X-rays, it is possible that this aggregation may have been catalyzed by the presence of Zn^{2+} and Cu^{2+} from the brass sample cells used for part of the data collection (Tschopp, 1984b). The mean R_G (4 values) increased from 3.51 ± 0.16 nm after 1 min to 4.61 ± 0.18 nm after 10 min (beam currents of 112-130 mA). In another session, the mean R_G (8 values) rose from 3.64 ± 0.25 nm to 4.03 ± 0.22 nm after 10 min (beam current 99-105 mA). The non-uniform increase in R_G may be due to the different beam currents, sample buffers and cells (Perspex as opposed to brass) in use. Analyses of irradiated C9 samples by SDS-PAGE showed no difference from the starting material. This suggested that

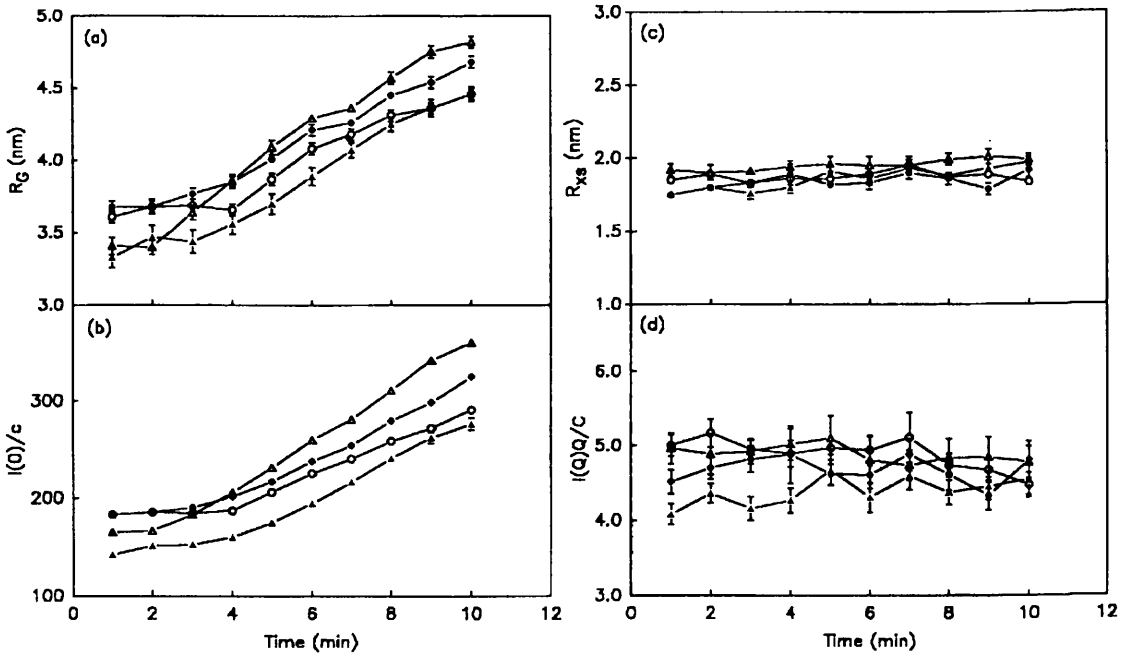


Figure 7.6. Synchrotron X-ray radiation damage effects observed in the R_G and R_{XS} Guinier analyses of C9.

The R_G, $I(0)/c$, R_{XS} and $I(Q)Q.Q/c$ data from each of ten consecutively-recorded time-slices for four C9 samples all at 7.4 mg/ml in 0.2 M NaCl, 0.02 M borate, 1 mM $CaCl_2$ and 5 mM EDTA, pH 7.5 denoted by o, ●, ▵ and ▲ are shown in (a), (b), (c) and (d) respectively.

some form of non-covalent association was induced by the X-ray beam. Since the R_G and $I(0)$ values were constant in the first three min of exposure (Figures 7.6(a) and 7.6(b)), only these were used for Guinier analyses. The R_G Guinier analyses of Figure 7.3(c) were linear. No concentration dependence was observed, and the final mean R_G value was 3.66 ± 0.13 nm (8 values). These data compare well with the neutron R_G value in 0% ^2H_2O buffers (high positive solute-solvent contrast) of 3.51 ± 0.18 nm (5 values) (Figure 7.4(b)).

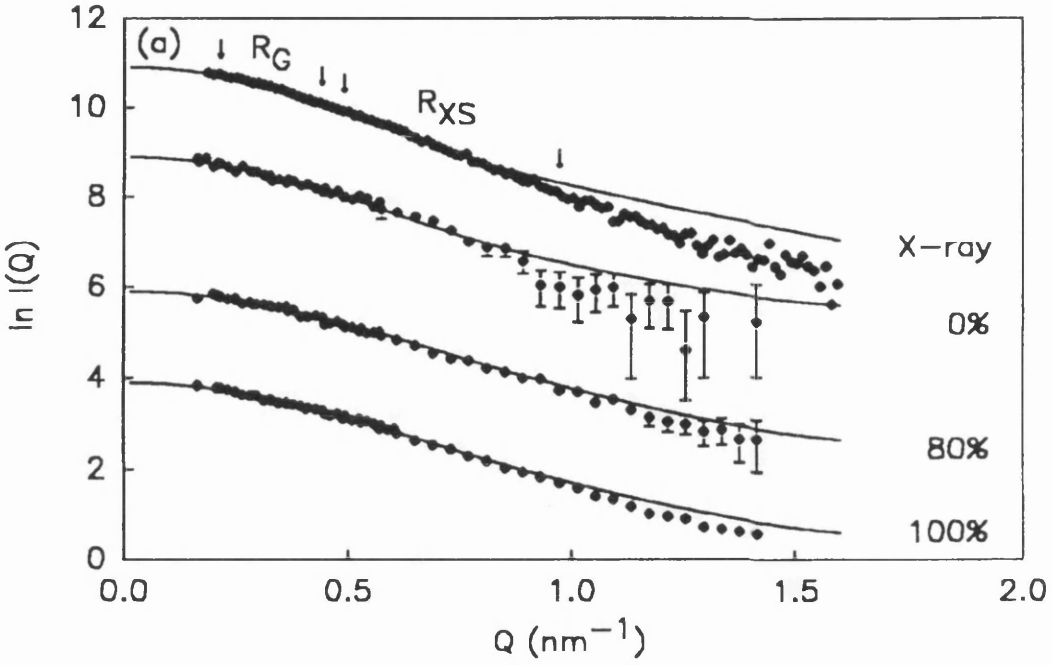
The X-ray cross-sectional Guinier analyses were linear also (Figure 7.3(f)). The R_{XS} and $[I(Q).Q]_{Q\to 0}$ parameters were not affected over time (Figure 7.6(c) & 7.6(d)). The mean R_{XS} (4 values) was unchanged from 1.82 ± 0.08 nm after 1 min to 1.93 ± 0.07 nm after 10 min. The mean R_{XS} was 1.84 ± 0.07 nm (15 values), in excellent agreement with the mean neutron R_{XS} of 1.81 ± 0.03 nm in

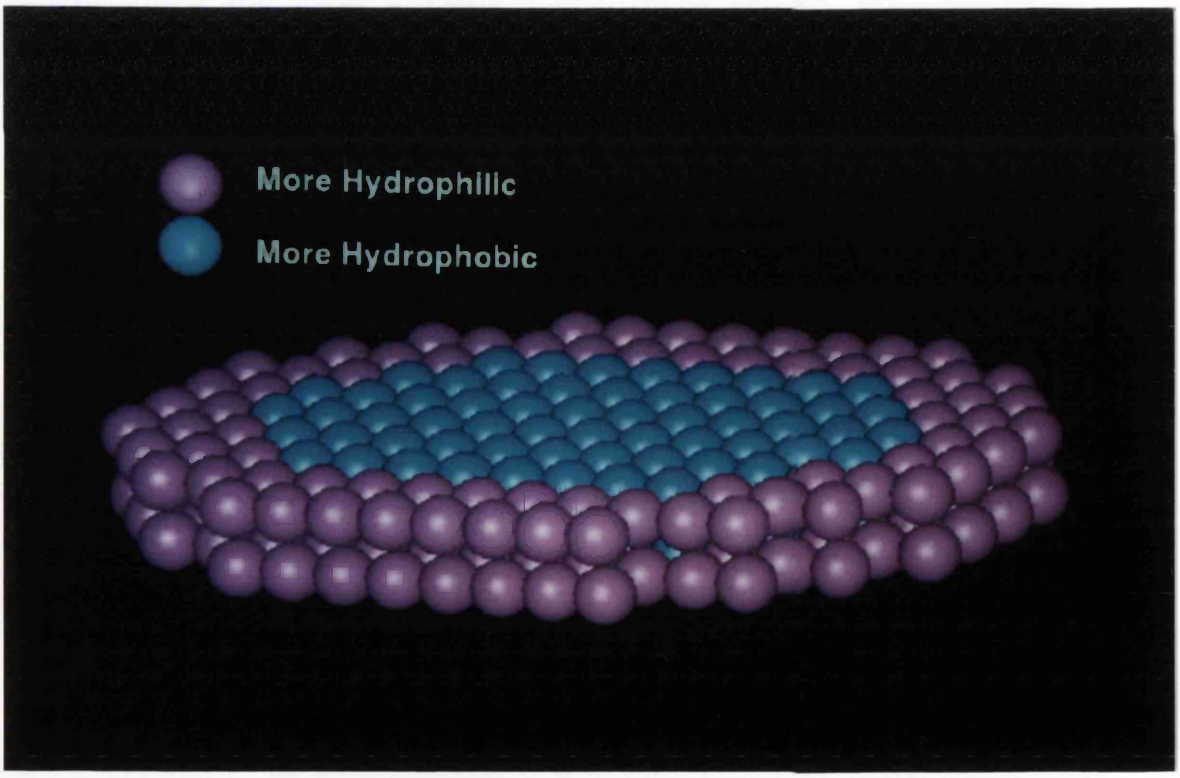
0% ^2H_2O (4 values) (Figure 7.4(b)). The length of C9 from the intensities ratio was 12.3 ± 0.9 nm (8 values), while that from the mean R_G and R_{XS} data assuming the shape of an elliptical cylinder was 11.0 ± 0.7 nm. The molecular dimensions of C9 were determined as 11.0 nm x 7.1 nm x 1.8 nm assuming the shape of an elliptical cylinder of hydrated volume 110.6 nm^3.

The ITP calculation of $P(r)$ in H_2O (high positive solute-solvent contrast) by X-rays was noticeably different from that of the neutron curve in 100% ^2H_2O (high negative solute-solvent contrast) (Figure 7.5(a)). The whole of the $P(r)$ curve was shifted to larger r, and the maximum in $P(r)$ was larger at 3.40 to 3.60 nm. The difference between the neutron and X-ray curves was attributed to the hydrophilic nature of C9. As is typical of globular proteins, the more hydrophilic residues in C9 were positioned further away from the centre of scattering densities than the more hydrophobic ones. Hydrophilic amino acids have a higher scattering density than hydrophobic ones and therefore make a stronger contribution to the X-ray $P(r)$ at large r than the neutron $P(r)$. The ITP calculation also showed that the largest dimension of C9 was estimated to fall in a range of 14 ± 2 nm by X-rays, although again a precise cut-off cannot be determined. This dimension was comparable with that estimated from the neutron $P(r)$ curve.

(7.3.3) Scattering curve simulations for C9

The use of small Debye spheres to model the full scattering curve in the Q range between 0.2 to 1.6 nm^{-1} enabled molecular models for C9 to be constructed. These permitted the assessment of possible arrangements of the TSR, LDL and EGF-like domains and PFL region in C9. Initial calculations were based on the assumption that C9 has a structure resembling an elliptical cylinder, using the dimensions determined from Guinier analyses. Good fits for the neutron scattering curves in 0%, 80% and 100% ^2H_2O buffers and X-ray curves were obtained with a sphere model based on overall dimensions of 10.9 x 6.8 x 1.4 nm (Figure 7.7(a)). This was constructed as an array of 16 x 10 x 2 cubes of side 0.681 nm, from which 14 were removed at each corner, to give a total of 264 spheres (Figure 7.7(b)). The innermost 120 spheres were assigned a scattering densities of


(a)
R_G
R_{XS}
X-ray
0%
80%
100%
ln I(Q)
Q (nm^{-1})
0
2
4
6
8
10
12
0.0
0.5
1.0
1.5
2.0


More Hydrophilic
More Hydrophobic

(b)

35% ^2H_2O and the outermost 144 spheres one of 55% ^2H_2O. This C9 model gave an R_G of 3.31 nm and a Stuhrmann α of 34 x 10^{-5} which were in good agreement with the observed neutron $R_{G\text{-}C}$ of 3.33 nm and α of 35 x 10^{-5}. The goodness-of-fit parameter R was 0.009 (0% ^2H_2O, out to a Q of 0.09 nm^{-1}), 0.011 (80% ^2H_2O) and 0.011 (100% ^2H_2O) [where the poor counting statistics in 0% ^2H_2O precluded the inclusion of data beyond Q of 0.9 nm^{-1} in that contrast]. As this model was only 1.4 nm thick it was considered an unrealistic model for a globular protein.

More satisfactory neutron curve fits were obtained by the use of V-shaped models for C9. Sphere models based on 168 cubes of side 0.792 nm were arranged into a two-armed structure, with each arm of length 10.3 nm, 11.1 nm or 11.9 nm (length 13, 14 or 15 spheres) and cross-sectional dimensions 2.4 nm x 2.4 nm (up to 3 x 3 spheres). The two identical arms were joined together at one end, and a series of models were created by separating the arms by angles from 0° to 20°. The internal distribution of hydrophilic and hydrophobic residues was considered, 96 spheres were allocated a scattering density of 55% ^2H_2O and 72 spheres one of 35% ^2H_2O (Figure 7.8(a)) (Section 7.2). Using neutron data in 0%, 80% and 100% ^2H_2O, the testing of these nine models showed that of arm length 11.1 nm (14 spheres) with an inter-arm angle of 10° (Figure 7.8(b), Figure 7.9(a), (b) and (c)) gave the best curve fits. This model had a $R_{G\text{-}C}$ of 3.37 nm, with individual R_G values of 3.53 nm (0% ^2H_2O), 3.14 nm (80% ^2H_2O) and 3.23 nm (100% ^2H_2O)

Figure 7.7. An elliptical cylinder model for C9.

(a) Comparison of simulated scattering curves for an elliptical cylinder model with experimental data by X-rays, and neutrons in 0%, 80% and 100% ^2H_2O buffers at C9 concentrations of 7.3, 7.3, 7.3 and 6.9 mg/ml respectively. The Q ranges used to calculate the Guinier R_G and R_{XS} analyses specified in Figure 7.3 are arrowed. Error bars due to data counting statistics are shown when these are large enough to be seen. No error bars are available for X-ray data.

(b) A flat elliptical model for C9 (sphere diameter 0.681 nm) showing the 120 hydrophobic spheres in blue and the 144 hydrophilic spheres in pink. This model has dimensions of 10.9 x 6.8 x 1.4 nm.

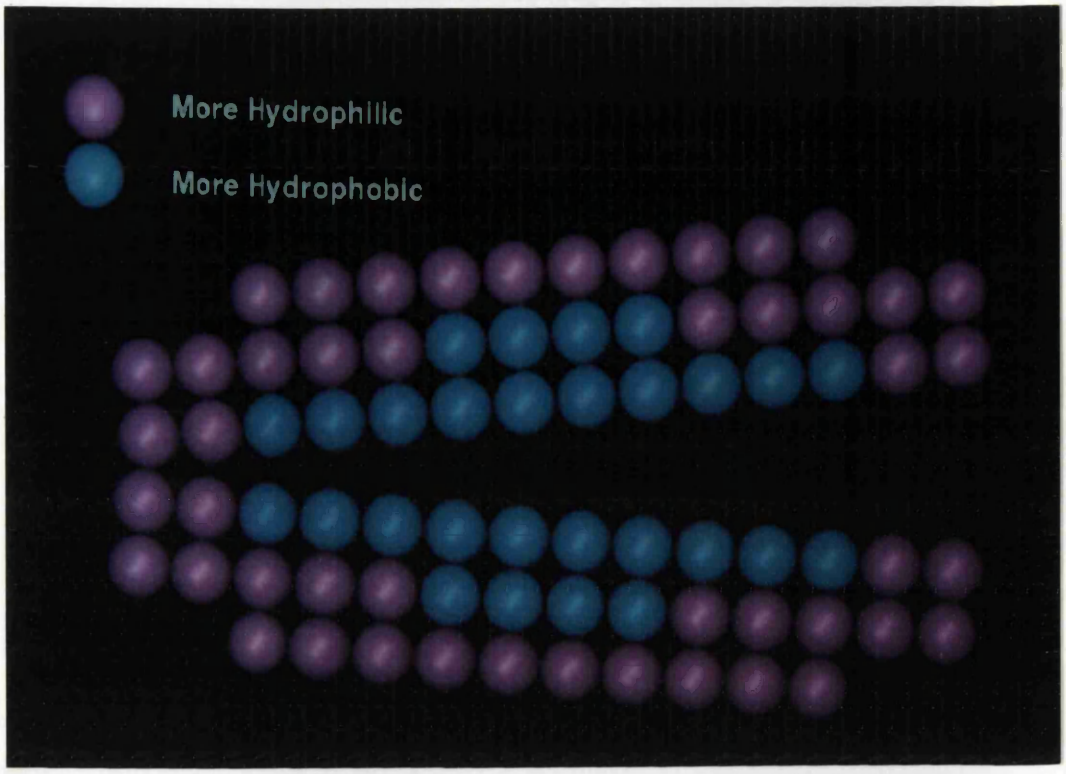
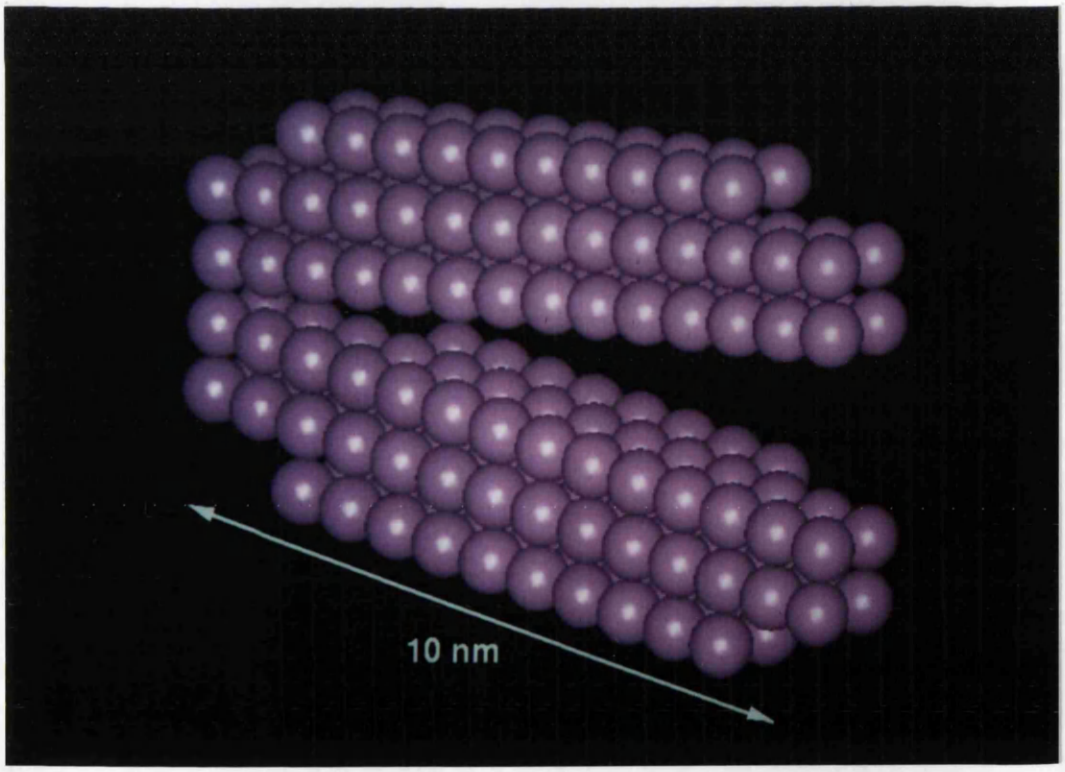


Figure 7.8. Sphere modelling of C9.

(a) Face-on view of the two-density neutron model for C9. The 72 hydrophobic and 96 hydrophilic spheres are coloured magenta and cyan respectively.



(b) The scattering model of C9 (sphere diameter 0.792 nm) is shown.

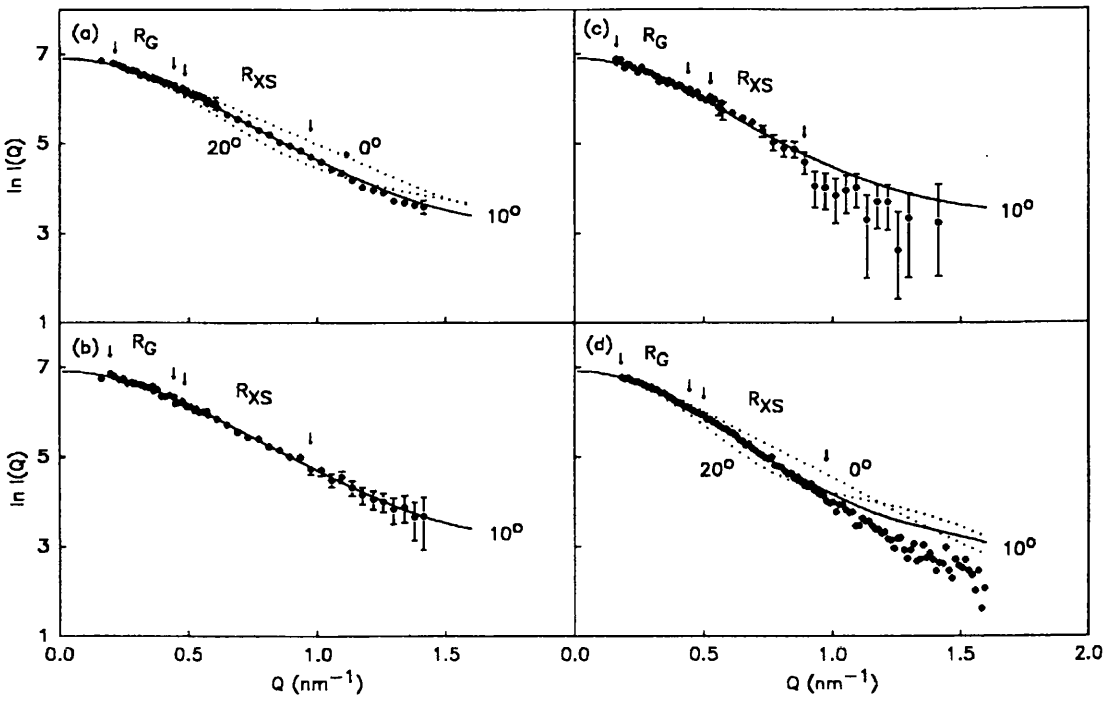


Figure 7.9. Comparisons of the curve simulations for C9 with experimental data. The C9 curve (solid line) calculated from the two-density V-shaped model with an inter-arm angle of 10° shown in Figure 8.8(b) is compared with the experimental curves measured in (a) 100%, (b) 80% and (c) 0% ^2H_2O buffers at C9 concentrations of 7.3, 7.3 and 6.9 mg/ml respectively. In (a), scattering curve simulations are shown for V-shaped models of C9 in which the inter-arm angle was set also as 0° and 20° (dotted lines). The Q ranges used to calculate the Guinier R_G and R_{XS} analyses specified in Figure 8.3 are arrowed. Error bars due to data counting statistics are shown when these are large enough to be seen. In (d), the curve fit to the X-ray data is shown (sample concentration 7.3 mg/ml), based on a hydrated one-density V-shaped model with inter-arm angles of 0°, 10° and 20° to follow (a).

and an α_G of 35 x 10^{-5}. These data agreed very well with the observed values reported in Figures 7.3 and 7.4. The goodness-of-fit parameter R was 0.009 (out to Q of 0.9 nm^{-1}), 0.007 and 0.007 respectively for the 0%, 80% and 100% ^2H_2O curve fits. The length of the arms agreed with the maximum distance calculated from $P(r)$ (Figure 7.5(b)). The quality of the curve fits for the final V-shaped model was thus slightly improved over that for the oblate model. This was particularly so in the Q range above 1.0 - 1.1 nm^{-1} in 80% and 100% ^2H_2O buffers (Figure 7.9(a) and (b)); note that these buffer contrasts accentuate the hydrophobic structure of C9. Angles of 0^o (corresponding to an elliptical cylinder) and 20^o clearly deviated from the experimental $P(r)$ and $I(Q)$ curves (Figures 7.5(b) & 7.9(a)). The use of different arrangements of hydrophobic and hydrophilic spheres within the C9 model resulted in poorer agreements with R_G, for example if an internal hydrophobic core was located along the central axis of each arm of C9 the α values in 0% ^2H_2O were smaller than those in 100% ^2H_2O.

This neutron model was hydrated and compared with the synchrotron X-ray data. A total of 168 cubes of side 0.870 nm were now used. Since the inclusion of both hydration and two densities (401 e/nm^3 and 433 e/nm^3 for the hydrophobic and hydrophilic components, relative to 334 e/nm^3 for the solvent) have similar effects on the scattering curves (Perkins *et al.*, 1991b), only the hydration level of a single density C9 model was used in the curve fits. The best curve fit for the V-shaped model corresponded to the neutron model of Figure 7.8(b), with hydrated arms of length 12.2 nm and an inter-arm angle of 10^o. Other models with angles of 0^o and 20^o gave large deviations from the experimental $P(r)$ and $I(Q)$ X-ray data (Figures 7.5(c) & 7.9(d)). The R_G and R_{XS} data were fully explained by this modelling. The R was however high at 0.070. This was attributed to the deviations seen beyond Q of 0.9 nm^{-1}, which was seen also with the neutron 0% ^2H_2O curve fit of Figure 7.9(c), and suggested that the hydrophilic regions of C9 is not as well modelled as the hydrophobic regions.

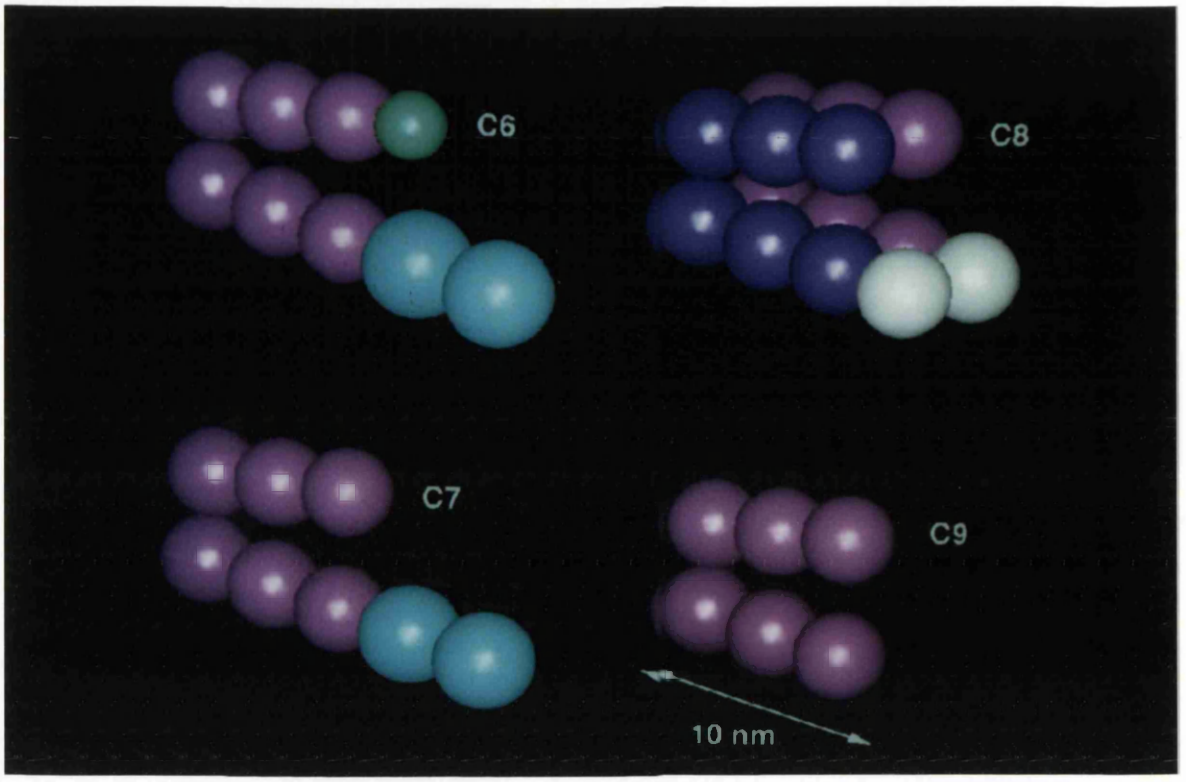


Figure 7.10. Comparison of hydrodynamic models for C6, C7, C8 and C9 of complement.

C9 is modelled as 6 magenta spheres of diameter 3.34 nm. The additional sphere at the N-terminus of C6 is green and 2.54 nm in diameter. The 2 additional spheres at the C-terminus of C6 and C7 are cyan and of diameters 3.96 nm and 3.70 nm respectively. C8 is depicted as a dimer of C9 and the 2 extra white spheres are of diameter 3.34 nm.

(7.3.4) Simulations of hydrodynamic data for C6, C7, C8 and C9

Sedimentation coefficients provide an independent measurement of the degree of elongation of C9 in solution. Published values for C9 of $s^{o}_{20,w}$ are 4.5 S (Podack & Tschopp, 1982b; Tschopp *et al.*, 1984) and 4.7 S (Tschopp, 1984b). A frictional ratio f/f_o of 1.22 ± 0.05 was calculated assuming an hydrated volume of 110.5 nm^3 (hydration 0.39 g H_2O / g protein). This is one of the lowest ratios found among the complement proteins (Perkins *et al.*, 1990c).

Hydrodynamic models made from non-overlapping spheres were used to calculate $s^{o}_{20,w}$ values for C9. Simple models based on 6 spheres of diameter 3.34 nm arranged in a 3 x 2 array (dimensions 10.0 x 6.7 x 3.3 nm) with an inter-arm angles of 0°, 5°, 10°, 15° and 20° gave $s^{o}_{20,w}$ values of 4.70, 4.60, 4.52, 4.45 and 4.40 S respectively. The best agreement with the published value of 4.5 S was obtained with an angle of 10° between the arms (Figure 7.10). A typical error of ± 0.2 S in the simulation (Chapter 6) corresponded to changes in the length of ± 2 nm. In a second approach, the direct use of the final hydrated X-ray sphere model gave $s^{o}_{20,w}$ values of 4.77 S, 4.49 S and 4.40 S for angles of 0°, 10° and 20° in that order. Again, there was excellent agreement between the experimental $s^{o}_{20,w}$ of 4.5 S and the 10° X-ray model. In summary, both sets of calculations showed that the hydrodynamic length of C9 agreed well with the length of 11.1 nm determined by solution scattering modelling.

The sedimentation coefficient for C9 was compared with those for the other homologous proteins which compose the MAC (C6, C7 and C8). The mean of the published $s^{o}_{20,w}$ values for C6 was 5.7 ± 0.2 S (5.5 S, 5.5 S, 5.6 S, 6.0 S and 5.7 S) (Podack *et al.*, 1976; Podack & Tschopp, 1982a, 1984; Priessner *et al.*, 1985; DiScipio & Gagnon, 1982). That for C7 was taken as 5.4 ± 0.2 S (the mean of 5.3 S, 5.3 S, 5.1 S, 5.6 S and 5.6S). That for C8 was taken as 8.1 ± 0.1 S (the mean of 8.0 S and 8.1 S). The frictional ratios f/f_o for C6, C7 and C8 were calculated to be 1.34 ± 0.05, 1.31 ± 0.05 and 1.19 ± 0.01 respectively. The frictional ratios for C6 and C7 were slightly larger than that for C9 (1.22 ± 0.05) indicating these proteins to be more elongated than C9. This was as expected from the presence

of 6 and 5 additional domains in C6 and C7 respectively when compared with C9. The f/f_o of the C8αγ-C8β trimeric complex was similar to C9. This suggested that the C8αγ and C8β subunits of C8 associate in a side-by-side interaction (rather than in an end-to-end one), in which the long axes of C8αγ and C8β were parallel and close to each other.

The six-sphere model for C9 was used as a basis for hydrodynamic modelling of the homologous proteins C6, C7 and C8 (Figure 7.10). All the extra domains in C6, C7 and C8 (except the first TSR in C6) were found at the C-terminus in the amino acid sequences and therefore were positioned at the end of one of the arms of the C9 model. The additional TSR that was found at the N-terminus of C6 was represented by a single sphere at the end of the other arm of the C9 model. The location of the C8γ subunit in C8 was unknown and the volume of this was combined with those of the additional TSR domains in C8 relative to C9. The volume of the additional spheres corresponded to the volume difference between C9 and each of C6, C7 and C8. The $s^o_{20,w}$ values of the models shown in Figure 7.10 were calculated as 5.6 S for C6, 5.2 S for C7 and 8.1 S for C8. These were in excellent agreement with the above experimental $s^o_{20,w}$ values. The length of the longest dimension of the C6 model was 18.0 nm; this was in satisfactory agreement with the lengths of 14.4 nm and 18 ± 2 nm for C6 that were deduced from electron microscopy (DiScipio & Hugli, 1989; Haefliger *et al.*, 1989). The longest dimension of the C7 model was 17.5 nm; this was also close to the length of 15.1 nm for C7 from electron microscopy (DiScipio *et al.*, 1988). These calculations showed that the solution structure of the domains of C9 were compatible with the hydrodynamic data of C6, C7 and C8 and electron micrographs of C6 and C7 taking into account the extra domains.

(7.4) CONCLUSIONS

C9 was best visualised as a typical mosaic protein found in the complement system, but possessing unique structural features. The two most distinctive properties of C9 are its ability to polymerise (either on its own or in association with C5b-8), and its ability to penetrate lipid bilayers (including cell membranes). Evidence from electron microscopy, tryptic cleavage experiments, and the analysis of monoclonal antibody binding sites (Kusunoki *et al.*, 1990) showed that C9 undergoes a drastic unfolding when polymerised and inserted into membranes [summarised in Podack & Tschopp (1984) and Stanley *et al.* (1986)]. Differential scanning calorimetry indicated that C9 was already partially unfolded at body temperature and upon binding to the C5b-8 complex on the membrane, C9 could undergo further refolding into a structure compatible with a typical amphipathic membrane protein (Lohner & Esser, 1991).

By the use of solution scattering techniques, free C9 was found to be fully monomeric in solution. An A_{280} of C9 of 8.37 rather than 9.6 to 9.88 gave a correct determination of the M_r of C9 by neutron scattering. It should also be noted that the conventionally-assumed M_r of C9 is reduced by about 11% to 66,400 when calculated from the sequence. In other words, these revised values may affect determinations of the stoichiometry of C9 within the MAC (generally estimated as 12 to 18 copies of C9).

Both the R_G and $P(r)$ scattering calculations as well as the $s^o_{20,w}$ hydrodynamic calculation showed that electron microscopy underestimated the length of free C9. These calculations presented here showed this to be 11 to 12 nm. Electron micrographs of free C9 (Podack & Tschopp, 1982a; DiScipio & Hugli, 1985) had shown that C9 had dimensions of 8.0 x 5.5 nm or 7.0 x 5.0 (± 0.7) nm. The fact that these dimensions were less than those found here may result from artefacts caused by electron beam damage, the non-physiological conditions of measurement *in vacuo* (using stains), or magnification errors. The dimensions of a tubular ring of poly-C9 by electron microscopy were 15 to 16 nm in length with an internal diameter of 9 to 11 nm and an outer diameter of 15 to 22 nm (Podack

& Tschopp, 1982a; DiScipio & Hugli, 1985), i.e. a monomer of C9 was increased in length from around 11 nm to 16 nm in poly-C9. A ring of 12 to 18 C9 molecules based on the model presented here (diameter 2.38 nm), would give a inner pore diameter of 9.0 to 13.6 nm, in agreement with the electron microscopy and liposome swelling assay (Zalman & Müller-Eberhard, 1990). The solution structure of C9 thus remains compatible with hypotheses that C9 unfolds to form a longer structure in poly-C9.

Of interest is that C9 has a relatively compact structure for a multi-domain protein, as evidenced by the R_G/R_O and f/f_o ratios from scattering and hydrodynamic data. The 11 to 12 nm length of C9 from these solution analyses can be compared with the sizes of the TSR, LDLr, perforin-like and EGF-like moieties in C9. The TSR is of approximate size 4 nm x 1.7 nm x 1.7 nm (Chapter 6; Smith *et al.*, 1991), while the EGF-like domain may be approximately 3.5 nm x 2 nm x 2 nm (Cooke *et al.*, 1987). If domains were spherical (the most compact shape possible), the LDLr domain would be of diameter 2.2 nm, and the perforin-like region would be of diameter 4.6 nm (Chapter 8). The total length of a linear arrangement of these four structures will therefore be between 10.5 to 14.3 nm, depending on the orientation of the TSR and EGF-like domains. While it is possible that these four domains can be contained within an elliptical cylinder of dimensions 10.9 x 6.8 x 1.4 nm (Section 7.3), such a structure appears to be too thin. It is more likely that the LDLr domain and the perforin-like region are non-spherical in shape, and that the long axes of the TSR and LDLr domains are colinear with that of C9. The modelling of C9 as a compact V-shaped structure for C9 becomes plausible, with the perforin-like region located at the angle of the V, and the TSR, LDLr and EGF-like domains placed in the order predicted from the C9 sequence at the ends of the arms. This is further substantiated by antibody binding analyses which suggest that C9 changes conformation around a hinge region located between residues 296 and 528, i.e. in the perforin-like region (Laine & Esser, 1989; Laine *et al.*, 1988).

The initial structural modelling of C6, C7 and C8 (Figure 7.10) suggested that C6, C7, C8 and C9 possess similar structural morphologies of lengths

between 11 nm to 18 nm. The additional length of 7 nm for C6 and C7 when compared to C9 was be attributed to the presence of two SCRs and two FIMs at the C-terminus of these components. The length of an SCR was estimated to be 2.6 to 4.6 nm (Perkins *et al.*, 1991a). If these two adjacent SCRs were positioned lengthwise as found in C1r, C1s, C4BP and factor H, this would account for the increased length of C6 and C7 relative to C9. This means that, even though the length of the FIM is unknown, the two FIMs of C6 and C7 may be located in a compact conformation relative to the two SCRs within these proteins such that the overall length of C6 and C7 is not increased further.

In electron micrographs, a crevice on one face of C9 has been reported by DiScipio & Hugli (1985), even though these authors were not able to detect a bilobal V-shaped structure for C9. This would be consistent with the C9 model shown in Figure 7.8(b). Some experimental evidence for a V-shaped structure was found in the analyses of Figures 7.5 and 7.9. Even though the scattering analyses were constrained by the volume calculated from the primary structure, and were compatible with hydrodynamic modelling, it should be noted that other scattering-equivalent models may offer equally satisfactory curve fits (e.g. models with two arms of different lengths). Relatively few parameters are uniquely determined from solution scattering data, and this is due to the fact that the data are spherically averaged because of the random orientations of the scattering macromolecules.

Neutron scattering probed the internal structure of C9, and interestingly showed that this had a radial distribution of hydrophilic and hydrophobic residues that is typical of soluble glycoproteins. This accounts for its solubility in aqueous media, i.e. the three cysteine-rich TSR, LDLr and EGF-like domains are all in contact with the aqueous phase. This internal structure is similar to C5 (Perkins *et al.*, 1990b), which also participates in the MAC after activation to C5b. The good curve fits for C9 in Figure 7.9 were obtained with the distribution of hydrophilic and hydrophobic residues as shown in Figure 7.8(a). The implication of this distribution is that the unfolding of the C9 structure would expose the more hydrophobic region of this structure, thus facilitating contact with lipids.

CHAPTER 8.

STRUCTURAL STUDIES ON THE DOMAINS OF C9

(8.1) INTRODUCTION

C9 is the major constituent of the membrane attack complex formed in the terminal stages of complement-mediated cell lysis. Free C9 is a water-soluble, hydrophilic protein, while polymerized C9 is amphipathic and binds to detergent (Biesecker & Müller-Eberhard, 1980; Tschopp *et al.*, 1982, 1984; Chapter 7). C9 and the four other late complement components C6, C7, C8α and C8β exhibit structural, functional and antigenic similarities (DiScipio *et al.*, 1984, 1988; Stanley *et al.*, 1985; Rao *et al.*, 1987; Howard *et al.*, 1987; DiScipio & Hugli, 1989; Chakravarti *et al.*, 1989; Haefliger *et al.*, 1987, 1989). All these proteins possess multi-domain structures based on 4 to 12 distinct structural motifs (Figure 1.1).

All of the late complement components have a common region of cysteine-rich independently-folded domains, the TSR, the LDLr domain and the EGF-like domain. The TSR and LDLr domains comprise most of the N-terminal C9a fragment of C9. C9a is known to be hydrophilic and has no membrane-binding properties (Biesecker *et al.*, 1982; Ishida *et al.*, 1982; DiScipio *et al.*, 1984). The TSR is found also in thrombospondin and properdin (Lawler & Hynes, 1986; Goundis & Reid, 1988), and its structural properties have been analyzed in Chapter 6. The LDLr repeat type A was originally identified in the LDL receptor protein (Yamamoto *et al.*, 1984) but no structural studies on this domain have been reported to date. The EGF motif is found as a single entity and also in several mosaic proteins (Carpenter & Cohen, 1990; Carpenter & Wahl, 1990). The putative EGF-like domain in C9 is located at the C-terminus (Stanley *et al.*, 1985; Stanley & Herz, 1987). The molecular structure of human EGF and several analogues have been determined by 2D-NMR spectroscopy (Cooke *et al.*, 1987, 1990; Montelione *et al.*, 1987; Kohda *et al.*, 1988; Mayo *et al.*, 1989; Brown, S.C. *et al.*, 1989; Selander *et al.*, 1990; Kline *et al.*, 1990). Since human EGF is water-soluble, it is possible that the EGF domain in C9 is also hydrophilic, however no general structural analyses on the EGFs found in the complement components have been reported. The EGF is very stable to denaturation by heat and acid (Carpenter & Wahl, 1990; Holladay *et al.*, 1976) and in some proteins may act to stabilize neighbouring domains (Pigott *et al.*, 1991).

Between the LDLr and EGF domains in these five complement proteins, there is a region which is virtually devoid of Cys residues. This region shows strong sequence similarities to perforin and the C9b fragment has similar membrane-binding activities (Ishida *et al.*, 1982; Shiver *et al.*, 1986). C9 and perforin share many functional similarities, for example, the cross-reaction of certain antibodies between the two proteins, both polymerize in the presence of Zn^{2+}, the lytic activity of both proteins is suppressed by lipoproteins and the polymerization of C9 and the lytic activity of perforin is abolished at low pH (Young *et al.*, 1986; Tschopp & Jongeneel, 1988; Tschopp & Nabholz, 1990; Tschopp & Masson, 1987). In this perforin-like region, possible α-helix structures have been identified (Stanley & Herz, 1987; Peitsch *et al.*, 1990) that may be implicated in the formation of trans-membrane channels but there is little experimental evidence for this.

The consensus averaging of structural properties for the EGF, LDLr and PLR sequences permits the most prominent structural features to be identified. Together with similar analyses of the TSR (Chapter 6) the EGF, LDLr and TSR domains in C9 and other related proteins are shown to be hydrophilic and dominated by β-sheet structures. Analyses of the PLR showed that this is hydrophilic also, and contains both α-helix and β-sheet structures. The results of this study will facilitate further structural and functional investigations of the late complement components.

(8.2) MATERIALS AND METHODS

(8.2.1) Sequence alignments

The 99 EGF sequences shown in Figure 8.1 are in the order: human, rat and mouse EGF or EGF precursor (Bell *et al.*, 1986; Simpson *et al.*, 1985; Scott et al, 1983), human and rat transforming growth factor-α (Derynck *et al.*, 1984; Lee *et al.*, 1985), human, mouse, pig and baboon urokinase (Nagai *et al.*, 1985; Degen *et al.*, 1987; Nagamine *et al.*, 1984; Au *et al.*, 1990), human tissue plasminogen activator (Degen *et al.*, 1986), human factor VII (O'Hara *et al.*, 1987), human and bovine factor IX (Yoshitake *et al.*, 1985; Katayama *et al.*, 1979; McMullen *et al.*, 1983), human and bovine factor X (Kaul *et al.*, 1986; Enfield *et al.*, 1980, McMullen *et al.*, 1983), human factor XII (Cool *et al.*, 1985), human and bovine protein C (Plutzky *et al.*, 1986; Long *et al.*, 1984), human and bovine protein S precursor (Hoskins *et al.*, 1987; Dahlbäck *et al.*, 1986), bovine protein Z (Hojrup *et al.*, 1985), human EGF-like (Ciccodicola *et al.*, 1989), human thrombospondin precursor (Hennessy *et al.*, 1989), human uromodulin (Pennica *et al.*, 1987), human, mouse and bovine thrombomodulin (Suzuki *et al.*, 1987; Dittman & Majerus, 1989; Jackman *et al.*, 1986), human complement components C1r and Cls (Leytus *et al.*, 1986a; McKinnon *et al.*, 1987), C6 (DiScipio & Hugli, 1989; Chakravarti *et al.*, 1989; Haefliger *et al.*, 1989), C7 (DiScipio *et al.*, 1988), C8α (Rao *et al.*, 1987), C8β (Haefliger *et al.*, 1987; Howard *et al.*, 1987) and C9 (DiScipio *et al.*, 1984; Stanley *et al.*, 1985), mouse and trout C9 (Stanley & Herz, 1987), human and mouse perforin (Lichtenheld & Podack, 1989; Shinkai *et al.*, 1988; Lowrey *et al.*, 1989; Trapani *et al.*, 1990) and rat cytolysin (Ishikawa *et al.*, 1989).

In the order shown in Figure 8.5, the 74 LDLr sequences correspond to human, rabbit and rat LDLr (Yamamoto *et al.*, 1984, 1986; Lee *et al.*, 1989), human LDLr-related protein (Herz *et al.*, 1988), human GP330 (pathogenic autoantigen in Heymann nephritis) (Raychowdhury *et al.*, 1989) and the complement component factor I (Catterall *et al.*, 1987). Other proteins with LDLr sequences are cited above. Likewise, in Figure 8.7, the 10 PLR sequences were aligned, based on the above protein sequences.

(8.2.2) Structural analyses of sequences

Sequences were aligned both manually and also with the use of guidelines provided by the automatic multiple sequence alignment program MULTAL (Taylor, 1990). Alignments were achieved by maximising the occurrence of identical or chemically homologous residues in these sequences (Chapter 3). Gaps were minimised in order to facilitate the averaging of structural properties. The consensus length was determined from the number of residues occurring in more than 50% of the sequences in the alignment. Exon structures where available, were used to define domain boundaries (Craik *et al.*, 1982; Traut, 1988). The exon structures for EGF domains in porcine urokinase, human tpa, human factor IX, human factor X, human protein C, human EGF precursor, and human thrombospondin have been determined (Nagamine *et al.*, 1984; Ny *et al.*, 1984; Yoshitake *et al.*, 1985; Leytus *et al.*, 1986b; Plutzky *et al.*, 1986; Bell *et al.*, 1986; Wolf *et al.*, 1990). Only one exon structure was available for the LDLr domain which was human LDLr (Südhof *et al.*, 1985).

Consensus secondary structure predictions using the Robson and Chou-Fasman methods (Chapter 3) were based on the alignments of Figures 8.1, 8.5 and 8.7. The original and updated Robson parameter sets for the α-helix, β-sheet, β-turn and random coil conformations were compared for the EGF and LDLr alignments (Garnier *et al.*, 1978; Gibrat *et al.*, 1987; Garnier & Robson, 1989). The 99 EGF, 74 LDLr and 10 PLR predictions were averaged using FORTRAN software (PREDRB and PREDCF: Chapter 3, Perkins *et al.*, 1988). The averaged hydropathy was calculated using the Kyte & Doolittle (1982), Eisenberg *et al.* (1984), Hopp & Woods (1981) and Parker *et al.* (1986) hydrophilicity or hydrophobicity scales (Chapter 3).

(8.3) RESULTS AND DISCUSSION

(8.3.1) Consensus analysis of the mammalian EGF domain

The EGF domain was generally defined by the positioning and connectivities of the six Cys residues. The fourth and fifth Cys were almost always separated by just one residue with the exceptions of the first and second putative EGF-like domains of human thrombospondin, where the separation was by two residues. Disulphide bridge connectivities were known only for the five structures determined by 2D-NMR (human EGF, rat EGF, mouse EGF, bovine factor X, human tgf-α) and C1s (Hess *et al.*, 1991) where the Cys residues were connected in the order 1 to 3; 2 to 4 and 5 to 6.

A total of 99 mammalian EGF sequences from 5 single domain proteins and 40 multi-domain proteins were aligned (Figure 8.1). Their lengths were defined by known exon structures for 7 proteins (20 domains) (Section 8.2). The exon boundaries of EGF domains in these proteins from other species were assumed to be homologous. The other domains were truncated by removal of flanking Cys residues or excessive length. Other mammalian EGF sequences have been reported in laminin, proteoglycans, cytotactin, entactin, GMP-140, and other multi-domain proteins, and also from non-mammalian sources (Carpenter & Wahl, 1990). The aim of this analysis was to determine the putative EGF domain structures in the complement components C1r, C1s, C6, C7, C8α, C8β and C9 by comparison with EGF structures determined by 2D-NMR spectroscopy.

The lengths of the EGF domains were found to be very variable and ranged from 59 residues (the third EGF domain in LDLr) to 36 residues (the fifth EGF domain in thrombomodulin). The lengths were split into 6 regions of deletions or insertions (segments a to f). Three EGF groups could be defined on the basis of the size of the four large loops 1 to 4 located between the six Cys residues (Table 8.1 & Figure 8.4). In Group A (34 sequences) the loops corresponded approximately in size to those of the five EGF sequences for which a 2D-NMR molecular structure was available (Figures 8.1 & 8.2). MULTAL analyses also grouped these

```
GROUP A              1---5        6--10       11-------21      22--27  28-32     33----39
                     ←-a-→        ←-b-→       ←----c-----→     ←-d-→   ←-e-→    ←--f---→
                        ←---loop 1---→ ←--loop 2--→ ←---loop 3---→   ←----loop 4---→
Hydropathy             --LLB         ----B       ----LB-L---      -B-B-B  -LLB-    -L-B-L-L

hu egf                NSDSEC       PLSHDGYC      LHDGVCMYIEA     LDKYACNC  VVGYI    GERCQYRDLKWWELR
rat egf               NSNTGC       PPSYDGYC      LNGGVCMYVES     VDRYVCNC  VIGYI    GERCQHRDLR
mo egf                NSYPGC       PSSYDGYC      LNGGVCMHIES     LDSYTCNC  VIGYS    GDRCQTRDLRWWELR
hu tgf-a             VVSHFNDC      PDSHTQFC       FHGTCRFLVQ     EDKPACVC  HSGYV    GARCEHADLLA
rat tgf-a            VVSHFNKC      PDSHTQYC       FHGTCRFLVQ     EEKPACVC  HSGYV    GVRCEHADLLA
hu urokinase            SNC           DC       LNGGTCVSNKY    FSNIHWCNC  PKKFG    GQHCEID
mo urokinase            SNC           GC       QNGGVCVSYKY    FSRIRRCSC  PRKFQ    GEHCEID
pig urokinase           SNC           GC       LNGGKCVSYKY    FSNIQRCSC  PKKFQ    GEHCEID
baboon urokinase        SDC           GC       LNGGTCMSNKY    FSSIHWCNC  PKKFG    GQHCEID
hu tpa                   SC        SEPRC       FNGGTCQQALY     FSDFVCQC  PEGFA    GKCCEID
hu F VII_1           SYSDGDQC       ASSPC       QNGGSCKDQLQ       SYICFC  LPAFE    GRNCETHK
hu F IX_1               DGDQC       ESNPC       LNGGSCKDDIN       SYECWC  PFGFE    GKNCELD
bo F IX_1               DGDQC       ESNPC       LNGGMCKDDIN       SYECWC  QAGFE    GTNCELD
hu F X_1                DGDQC       ETSPC       QNQGKCKDGLG       EYTCTC  LEGFE    GKNCELF
bo F X_1                DGDQC       EGHPC       LNQGHCKDGIG       DYTCTC  AEGFE    GKNCEFS
hu F XII_1           PKKVKDHC       SKHSPC       QKGGTCVNMPS       GPHCLC  PQHLT    GNHCQKEK
hu F XII_2           QRLASQAC       RTNPC       LHGGRCLEVEG       HRLCHC  PVGYT    GPFCDVDTKAS
hu protein C_1          DGDQC  LVLPLEHPCASLC       CGHGTCIDGIG       SFSCDC  RSGWE    GRFCQR
bo protein C_1          DGDQC  EDRPSGSPCDLPC       CGRGKCIDGLG       GFRCDC  AEGWE    GRFCLH
hu protein S_1       VNAIPDQC          SPLPC    NEDGYMSCKDGKA       SFTCTC  KPGWQ    GEKCEFDIN
bo protein S_1       VNAISDQC          NPLPC    NEDGFMTCKDGQA       TFTCIC  KSGWQ    GEKCESDIN
bo protein Z_1       TYMGGSPC          ASQPC      LNNGSCQDSIR       GYACTC  APGYE    GPNCA
hu egf_like          SKELNRTC             C      LNGGTCMLGSF          CAC  PPSFY    GRNCEHDVRKE
hu egf_pr_6             DQDDC         APVGC       SMYARCISEGE       DATCQC  LKGFA   GDGKLCS
hu egf_pr_7             DIDEC       EMGVPVC      PPASSKCINTEG       GYVCRC  SEGYQ    GDGIHCL
hu egf_pr_8             DIDEC       QLGVHSC       GENASCTNTEG       GYTCMC  AGRLS    EPGLICP
mo egf_pr_6             YEDDC         GPGGC       GSHARCVSDGE       TAECQC  LKGFA    RDGNLCS
mo egf_pr_7             DIDEC       VLARSDC      PSTSSRCINTEG       GYVCRC  SEGYE    GDGISCF
mo egf_pr_8             DIDEC       QRGAHNC       AENAACTNTEG       GYNCTC  AGRPS    SPGRSCP
hu 'spondin_1              GC         LSNPC       FAGVKCTSYPD    GSWKCGAC  PPGYS   GNGIQCTDVDE
hu 'spondin_3        VCKPRNPC       TDGTHDC       NKNAKCNYLGH  YSDPMYRCEC  KPGYA   GNGIICGEDTDLD
hu uromodulin_1      DTSEARWC          SEC       HSNATCTEDEA       VTTCTC  QEGFT    GDGLTCVDL
hu uromodulin_2         DEC       AIPGAHNC       SANSSCVNTPG       SFSCVC  PEGFR   LSPGLGCTD
hu uromodulin_3         VDEC      AEPGLSHC       HALATCVNVVG       SYLCVC  PAGYR    GDGWHCE
-------------------------------------------------------------------------------------------
GROUP B              1--4         5----11         12-----20       21--26  27-31 32--------43
                     ←-a→         ←--b-→          ←---c---→       ←--d-→  ←-e-→ ←-----f-----→
                        ←---loop 1---→ ←--loop 2--→ ←---loop 3---→   ←----loop 4---→
Hydropathy             -LLB         --LLL-B         LL-B-L---       -BLB-B  LL-B- B--L-L-BLL-L

hu F VII_2           DDQLIC         VNENGGC         EQYCSDHTG       TKRSCRC  HEGYS LLADGVSCTPTVEYP
hu F IX_2               VTC         NIKNGRC         EQFCKNSAD       NKVVCSC  TEGYR LAENQKSCEPAV
bo F IX_2               ATC         SIKNGRC         KQFCKRDTD       NKVVCSC  TDGYR LAEDQKSCEPAV
hu F X_2             TRKLC          SLDNGDC         DQFCHEEQN        SVVCSC  ARGYT LADNGKACIPTG
bo F X_2             TREIC          SLDNGGC         DQFCREERS        EVRCSC  AHGYV LGDDSKSCVSTE
hu protein C_2     EVSFLNC          SLDNGGC         THYCLEEVG        WRRCSC  APGYK LGDDLLQCHPA
bo protein C_2     EVRFSNC          SAENGGC         AHYCMEEEG        RRHCSC  APGYR LEDDHQLCVSK
hu protein S_2          EC      KDPSNINGGC         SQICDNTPG        SYHCSC  KNGFV MLSNKKDCKDVD
hu protein S_3          EC         SLKPSIC        GTAVCKNIPG        DFECEC  PEGYR YNLKSKSCEDID
hu protein S_4          EC          SENMC         AQLCVNYPG       GYTCYCDGKKGFK LAQDQKSCEVVSV
bo protein S_2          EC      KDPVNINGGC         SQICENTPG        SYHCSC  KNGFV MLSNKKDCKDVD
bo protein S_3          EC         VLKPSIC        GTAVCKNIPG        DFECEC  AEGYK YNPVSKSCDDVD
bo protein S_4          EC          AENLC         AQLCVNYPG       GYSCYCDGKKGFK LAQDQKSCEAVPV
bo protein Z_2      FAESEC        HPLRLDGC         QHFCYPGPE        SYTCSC  ARGHK LGQDRRSCLPHDR
hu ldlr_1              GTNEC         LDNNGGC         SHVCNDLKI        GYECLC  PDGFQ  LVAQRRCEDID
hu ldlr_2                EC          QDPDTC         SQLCVNLEG        GYKCQC  EEGFQ LDPHTKACKAVGSIA
hu ldlr_3           QPRGVNWC      ERTTLSNGGC         QYLCLPAPQINPHSPKFTCAC  PDGML LARDMRSCLTEAEAA
rab ldlr_1             ATNEC         MRGNGGC         SHTCFDLRI        GHECHC  PKGYR  LVDQRRCEDIN
rab ldlr_2               EC          EDPDIC         SQLCVNLAG        SYKCEC  RAGFQ LDPHSQACKAVDSIA
rab ldlr_3          QPRGVNWC      EKTALPNGGC         QYLCLPAPQINSHSPKFTCAC  PDGTL LAADMRSCRTEADVI
rat ldlr_1             KTNEC         LDNNGGC         SHICKDLKI        GYECLC  PSGFR  LVDGHQCEDID
rat ldlr_2               EC          QEPDTC         SQLCVNLEG        SFKCEC  RAGFH MDPHTRVCKAVGSIG
rat ldlr_3          QPRGVNWC      EATVLPNGGC         QYMCLPAPQISAHSPKFTCAC  PDGML LAKDMRSCLPEVDTV
bo ldlr_3           QPRGVNWC      ERTALRNGGC         QYLCLPAPQINPRSPKFTCAC  PDGML LAKDMRSCLTESESA
hu egf_pr_1            EQKLC         KLRKGNC        SSTVCGQDLQ       SHLCMC  AEGYA LSRDRKYCE
hu egf_pr_2            DVNEC         AFWNHGC         TLGCKNTPG        SYYCTC  PVGFV LLPDGKRCH
hu egf_pr_3            QLVSC         PRNVSEC         SHDCVLTSE        GPLCFC  PEGSV LERDGKTCS
hu egf_pr_4              GC        SSPDNGGC         SQLCVPLSP       VSWECDC  FPGYD LQLDEKSCAAS
hu egf_pr_5            GADPC         LYQNGGC         EHICKKRLG        TAWCSC  REGFM KASDGKTCLALDGHQLLA
mo egf_pr_1            PDPEL        LKQRGRPC        RFGLCERDPK        SHSSAC  AEGYT LSRDRKYCE
mo egf_pr_2            DVNEC         ATQNHGC         TLGCENTPG        SYHCTC  PTGFV LLPDGKQCH
mo egf_pr_3            ELVSC         PGNVSKC         SHGCVLTSD        GPRCIC  PAGSV LGRDGKTCT
mo egf_pr_4              GC        SSPDNGGC         SQICLPLRP       GSWECDC  FPGYD LQSDRKSCAAS
mo egf_pr_5            GADPC         LYRNGGC         EHICQESLG        TARCLC  REGFV KAWDGKMCLPQDYPILSG
hu 'spondin_2            C          KEVPDAC    FNHNGEHRCENTDP       GYNCLPC  PPRFTGSQPFGQGVEHATANKQ
```

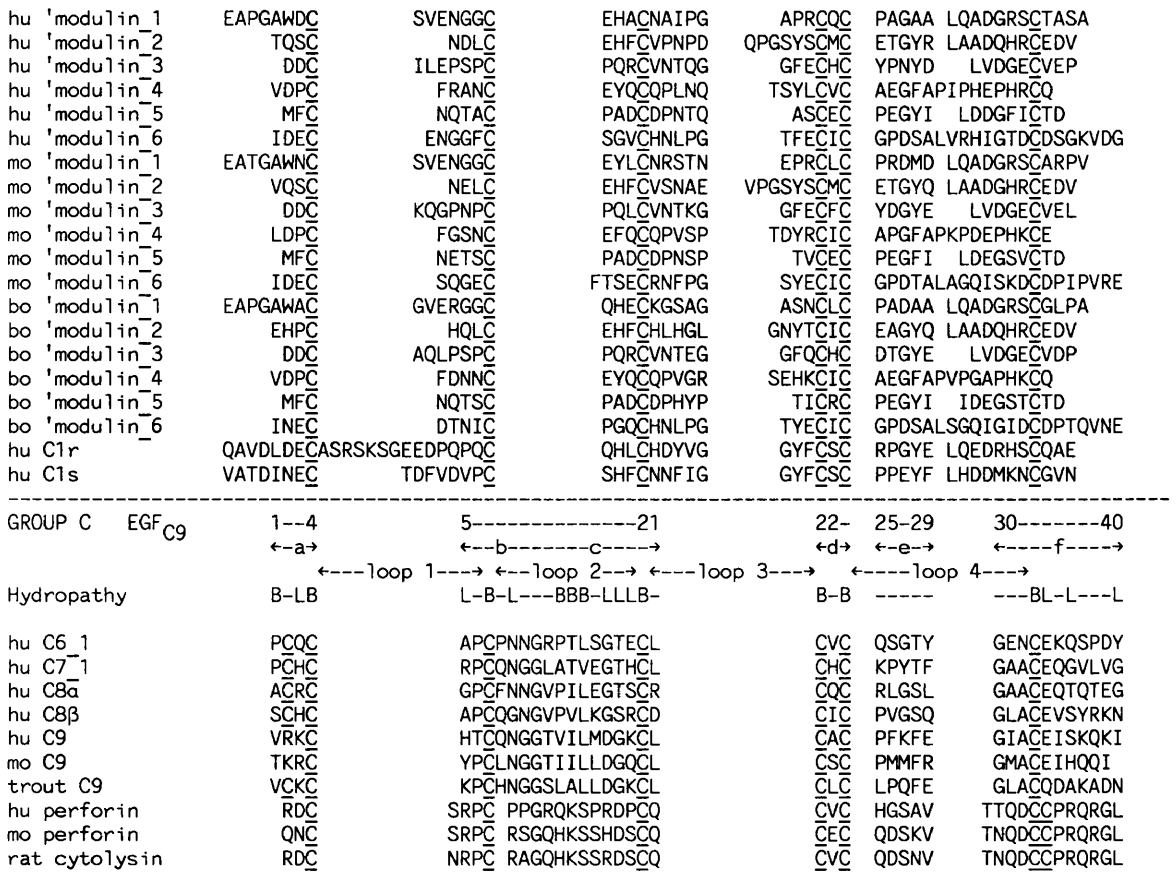


Figure 8.1. Alignment of 99 mammalian EGF sequences found in the complement components and other proteins.

These are divided into six segments labelled a to f (Table 8.1). The 99 sequences are subdivided into three Groups A, B and C (see text). Here and in Figures 8.5 and 8.7, Cys residues are underlined. Four regions of polypeptide between four Cys residues and the CXC tripeptide are designated loops 1 to 4. Repeated EGF domains within one multidomain protein are numbered sequentially starting from the N-terminus, but are located in the Figure according to the lengths of the loops 1 to 4. A summary of the consensus hydropathy of the residue (B, hydrophobic; L, hydrophilic) is shown above the sequences for each Group. Hu 'spondin is human thrombospondin and hu 'modulin is human thrombomodulin. (mo = mouse, bo = bovine, rab = rabbit).

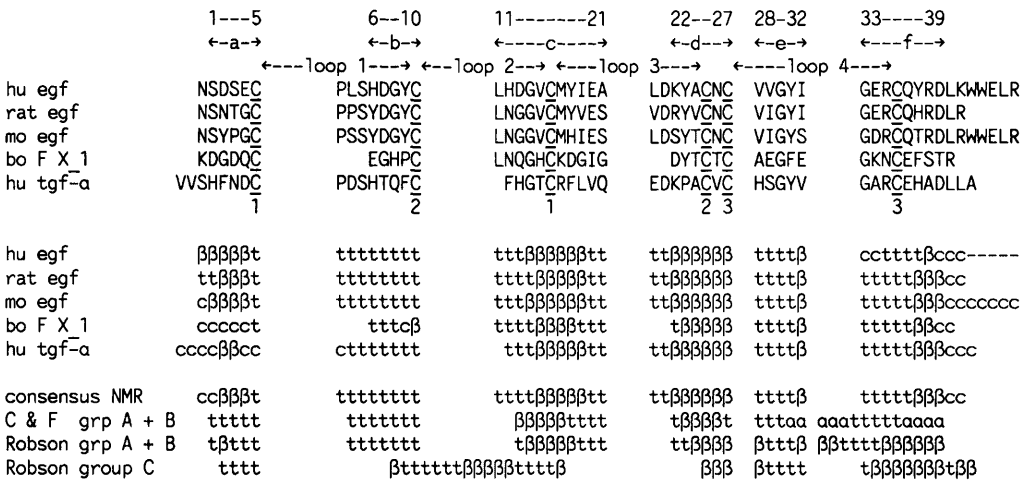


Figure 8.2. Summary of five EGF or TGF-α secondary structures determined by 2D-NMR spectroscopy.

The Group A sequences (upper) are aligned to permit identification of the corresponding secondary structures beneath these. The three disulphide bridges are numbered 1 to 3 beneath the sequences to indicate their connectivity. β, β-sheet; α, α-helix; t, β-turn; c, coil. Structures were taken as follows: human EGF, Cooke *et al.*, 1987, 1990; rat EGF, Mayo *et al.*, 1989; mouse EGF, Montelione *et al.* 1987; factor X_1 EGF, Selander et al, 1990; human TGF-α, Brown *et al.*, 1989, and Kline *et al.*, 1990. Another mouse EGF NMR structure was reported by Kohda *et al.* (1988) in which three fewer residues were assigned as β. While several structures have been reported for α-TGF by different groups, that in Kline *et al.* (1990) has been refined by distance geometry methods. The consensus NMR structure is compared with the two predictions from the average of 89 Groups A and B sequences (Figures 3(a) and 3(b)) and with Group C.

sequences together which suggested that these were closely related. Group B (55 sequences) included the EGFs from C1r and C1s, and formed the other major group of sequences sorted by MULTAL. Group C (10 sequences) corresponded to the putative EGF found in perforin and the late complement components C6, C7, C8α, C8β and C9. MULTAL showed that Group C formed two subsets based on the complement and perforin sequences but no clear correlation with either of Group A and Group B was found.

	Group A	Group B	Group C
loop 1	7	6	3
loop 2	5	3	12
loop 3	10	9	1
loop 4	8	12	8

Table 8.1. The average number of residues in loops 1 to 4 of the EGF domain for Groups A, B and C.

Analyses of Groups A and B show that these were of similar sizes but the consensus length (39 residues compared to 43 residues), mean length (42.9 ± 4.8 residues compared to 44.4 ± 5.4 residues) and dry volume (Chothia, 1975) (5.7 ± 0.8 nm^3 compared to 6.0 ± 0.8 nm^3) of these two groups were slightly different. The EGFs of Group C had a consensus length of 40 residues, mean length of 39.9 ± 0.3 residues and the smallest dry volume of 5.3 ± 0.2 nm^3. Table 8.1 lists the lengths of the four loops for each group. The major difference between Groups A and B lies in the length of loop 4. Loops 2 and 3 in Group C are of very different lengths to either of the other groups. Depending on the protonation state of His residues (0 or +1), the EGF has a negative mean net charge of -1.3 to -2.6 (± 3.3) (Group A) or -1.5 to -2.8 (± 2.8) (Group B). However the EGFs of Group C are positively charged with a net charge of 1.3 to 2.4 (± 2.0). The 58-65% conserved charged residues 36 and 38 primarily occur in Group B, not Group A where they are merged at position 38, and this may be correlated with the increased size of loop 4 in Group B.

Since Groups A and B exhibited the closest similarities, residue conservation over a consensus length of 44 residues averaged over 89 sequences

Segment	Position	Residue	Frequency	Percent
EGF				
a (1-5)	5	Cys	88	(99%)
b (6-12)	10	Gly, Ala, Ser	30 + 1 + 14	(51%)
	12	Cys	89	(100%)
c (13-21)	16	Cys	89	(100%)
	21	Gly, Ala, Ser	37 + 4 + 4	(51%)
d (22-27)	22	Gly, Ala, Ser	20 + 2 + 24	(52%)
	23	Phe, His, Tyr, Trp	17 + 6 + 35 + 2	(67%)
	25	Cys	86	(97%)
	27	Cys	89	(100%)
e (28-32)	30	Gly, Ala, Ser	72 + 1 + 1	(83%)
	31	Phe, His, Tyr, Trp	33 + 1 + 36 + 4	(83%)
f (33-44)	33	Ile, Leu, Val	1 + 32 + 2	(76%)
	36	Asp, Glu	39 + 4	(65%)
	37	Gly, Ser	53 + 5	(65%)
	38	Arg, His, Lys	16 + 8 + 28	(58%)
	40	Cys	88	(99%)
LDLr				
a (1-5)	5	Cys	73	(99%)
b (6-14)	10	Phe, His, Tyr, Trp	68 + 2 + 1 + 1	(97%)
	12	Cys	74	(100%)
c (15-38)	15	Gly, Ala, Ser	43 + 1 + 11	(74%)
	17	Cys	73	(99%)
	18	Ile, Leu, Val	55 + 5 + 11	(96%)
	22	Phe, Tyr, Trp	7 + 4 + 38	(66%)
	24	Cys	74	(100%)
	25	Asp	56	(76%)
	26	Gly, Ala, Ser	51 + 1 + 3	(74%)
	27	Asp, Glu	28 + 14	(57%)
	29	Asp, Glu	68 + 3	(96%)
	30	Cys	74	(100%)
	32	Asp, Glu	66 + 1	(91%)
	33	Gly, Ala, Ser	38 + 2 + 3	(58%)
	34	Ala, Ser	1 + 67	(92%)
	35	Asp	74	(100%)
	36	Glu	74	(100%)
d (39)	39	Cys	74	(100%)

Table 8.2: Survey of the Frequency of Residues in the EGF and LDLr domains[a]

[a] All residues or residue groups present in over 50% of the 89 EGF (Groups A and B) or 74 LDLr sequences are noted here. The numbering of the EGF residues corresponds to the consensus of Groups A and B in Figure 8.2. Chemically homologous residues are defined (Taylor, 1986) as follows: Tiny, G, A, S; Aliphatic, I, L, V; Aromatic, F, H, Y, W; Positive, R, H, K; Negative, D, E.

were summarized in Table 8.2. Sixteen residues were conserved to 50% or better, although only the six Cys residues were conserved to 97% or better. Comparison of Groups A and B with Group C shows that of these sixteen conserved residues only the six Cys residues and residues 30 and 37 were similarly conserved in Group

C. Tyr31 (or other aromatic substitutions) was highly conserved throughout Groups A and B and is thought to be involved in hydrogen bonding between β-sheets. This was poorly conserved in the complement components, only C9 contained an aromatic group at this position. This shows again that the EGFs of Group C are distinct from those of Groups A and B. Interestingly the tripeptide NGG (and chemically homologous versions of this) was found in segment c in Group A, in segment b in Group B and in the merged b/c segment in Group C.

Two other independent classifications of EGF sequences have been proposed by Appella *et al.* (1988) and Rees *et al.* (1988). The EGF Types 1 and 2 of Appella *et al.* (1988) correspond to Group A here. Their EGF Type 3 corresponds partly to Group B here, in which Appella *et al.* (1988) have assumed a different disulphide bridge connectivity within the EGF, the existence of which is not favoured by the results of Hess *et al.* (1991). The three EGF types distinguished by Rees *et al.* (1988) is based on the presence or absence of β-hydroxyaspartate or β-hydroxyasparagine and other conserved residues, however here these are distributed in both Groups A and B. Campbell *et al.* (1990) and Dudgeon *et al.* (1991) have suggested that an EGF is defined by the existence of the six Cys residues, together with three almost completely conserved Gly residues (positions 14, 30 and 37) and an aromatic residue (position 31). The present analyses suggest that residue conservation in the EGF is weaker than previously thought (e.g. Gly-14 is not conserved in the EGF sequences of Group B). In summary, an EGF structure is most generally defined by three conserved disulphide bridges connected according to the scheme 1-3, 2-4 and 5-6, and located within a sequence of about 50 residues in length, and such a structure can be subclassified in a variety of ways.

Given the similarities between the Groups A and B sequences, the Robson (using both the 1978 and 1987 parameter sets) and Chou-Fasman consensus structure predictions of Groups A and B together were summarised in Figures 8.3(a) to 8.3(c). Only β-biased Robson predictions were used since all five NMR structures were very similar and exhibited β-strands connected by turns (Figure 8.2). The averaged secondary structure from the 5 NMR structures was found to

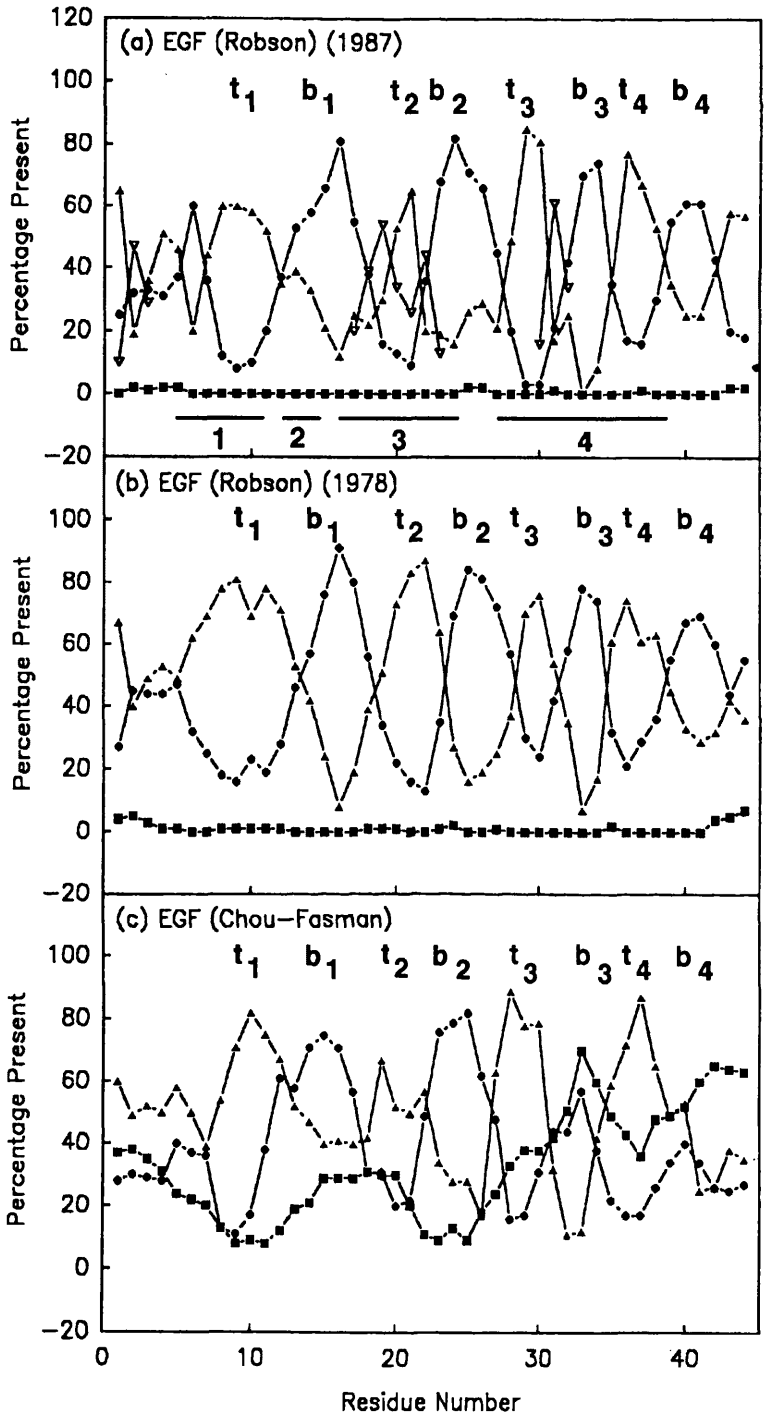


Figure 8.3. Consensus secondary structure predictions for the EGF motif.

(a), (b) and (c) Robson 1987 and 1978 propensity set and Chou-Fasman analyses for the 89 EGF sequences of Groups A and B respectively (■; α-helix; ●; β-sheet; ▲; β-turn). The coil predictions in the Robson analysis are only shown [▽ in (a)] when significant for reason of clarity.

consist of 38% β-strand, 54% turn and 8% coil. The structure predictions for Groups A + B resulted in 45% β-strand and 55% turn (Robson 1978); 43% β-strand, 45% β-turn and 11% coil (Robson 1987); and 20% β-strand; 20% α-helix; 60% turn (Chou-Fasman). The consensus NMR structure exhibited five β-strands interspersed by turns. All three predictions (Figure 8.3(a) to 8.3(c)) suggest that there are four clear segments of turn structures t_1-t_4 in 60-80% of the sequences, interspersed with β-strands b_1-b_4 (Robson) or two β-strands b_1-b_2 and two α-helices (Chou-Fasman). The 1978 Robson predictions gave the best agreement with the NMR structures, while the Chou-Fasman prediction qualitatively differs from the observed structure. In Group C only two of the four main NMR-observed β-strands were predicted, in segments d and f. These differences suggest that the EGF of Group C may be structurally distinct from the 2D-NMR structure of the EGF.

Hydropathy analyses of the EGF in Groups A, B and C were carried out using an average value from four different hydrophilicity or hydrophobicity scales (Section 8.2). The most frequently occurring maxima in the four hydropathy profiles were assigned as hydrophilic or hydrophobic (Figure 8.1). The hydropathy profiles for Groups A and B were similar suggesting that both Groups possess similar surfaces in contact with water. In contrast, different hydropathy patterns were observed for Group C. This was particularly evident in the β-strand of the large loop 2 region which had a hydrophobic character.

The 2D-NMR structure for human EGF can be used to rationalise the appearance of gaps and insertions in the EGF sequences. The EGF motif can be most generally defined as a central core of four Cys residues (Cys-12, -16, -25 and -27 in Table 8.2) whose α-carbon atoms are positioned in a near-tetrahedral arrangement with sides of length 0.4 - 0.8 nm (Figure 8.4). Cys-5 and Cys-40 are attached to this core to define the remainder of the EGF structure. This well-defined disulphide-linked core structure forms attachment points for the sequence segment loops 1 to 4 (Figure 8.1). The main sequence restriction imposed by the existence of this central core appears to be that Cys-25 and Cys-27 are almost always separated by one residue. Although number of consensus sequences for the

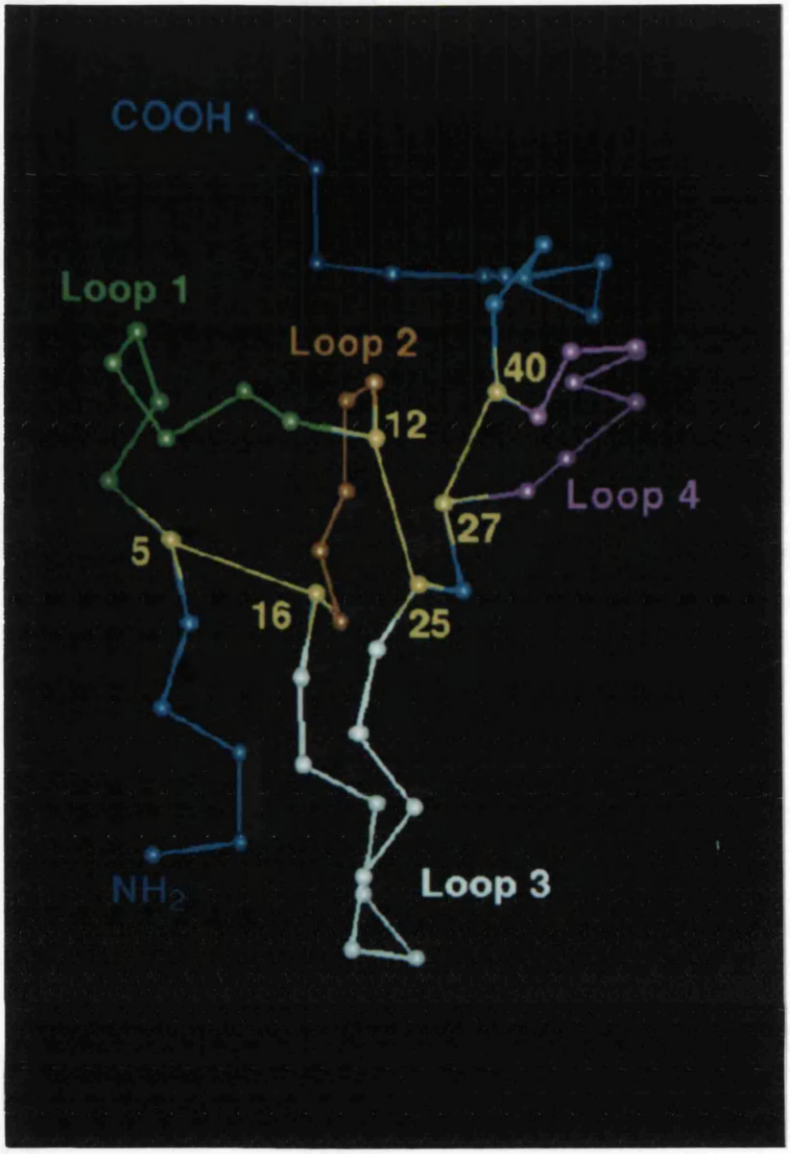


Figure 8.4. Schematic view of the α-carbon co-ordinates of human EGF (Cooke *et al.*, 1987).

Residue numbering follows that for Table 8.2. The six cysteine residues are numbered in yellow and the three disulphide bridges are drawn as a single yellow bond. The central tetrahedral core of Cys-12, Cys-16, Cys-25 and Cys-27 can be seen. The sequence loops are coloured as follows: loop 1, green; loop 2, brown; loop 3, white; loop 4, magenta.

EGF domain have been proposed, the loop structures between the Cys residues are all of variable length. For the EGF family of mitogens, a consensus sequence of $CX_7CX_{4-5}CX_{10-13}CXCX_8C$ of length 36-40 residues was proposed (Carpenter & Wahl, 1990) or more specifically, $X_nCX_7CX_{2-3}GXCX_{10-13}CXCX_3YXGXRCX_4LX_n$ (Carpenter & Cohen, 1990). For other EGF sequences, the consensus sequence was even more variable, $CX_{4-14}CX_{3-8}CX_{4-14}CXCX_{8-14}C$ (Carpenter & Wahl, 1990). Examination of the 2D-NMR structure shows that, while loops 1 and 2 extend into a similar region of space, no steric restrictions apply for loops 3 and 4. In the sequences of Figure 8.1, this offers an explanation of how loop 1 can vary in size between 1 to 14 residues, loop 2 from 3 to 12 residues, loop 3 from 1 to 14 residues, and loop 4 from 8 to 14 residues. In Group C, for steric reasons, the large size of loop 2 may be correlated with the small size of loops 1 and 3 (Table 8.1). The range of conformational possibilities offered by such a structure makes it possible to appreciate how the two EGF Groups A and B can arise.

(8.3.2) Consensus analysis of the LDLr domain

A total of 74 LDLr sequences from 13 proteins were aligned (Figure 8.5). The mean length was 41.2 ± 2.6 and the consensus length was 39 residues (Figure 8.5). These were split into three significant regions of deletions or insertions (segments a to d). Of these 39 residues, 19 positions are conservatively occupied in over 50% of the 74 sequences (Table 8.2); 13 of these are conserved to better than 90%. All six Cys residues Cys-5, Cys-12, Cys-17, Cys-24, Cys-30 and Cys-39 were 99-100% conserved. These hydrophobic Cys residues together with Phe-10, Ile/Val-18, and Trp-22 suggested that all the LDLr domains had a well-defined hydrophobic core. Since the first LDLr domain in Factor I contained only four Cys residues, this was suggestive that Cys-5 and Cys-17 were bridged in the LDLr domain.

The averaged physical properties of the LDLr domain were computed from Figure 8.5. The dry volume of 68 LDLr domains (Chothia, 1975; Perkins, 1986) was determined to be 5.38 ± 0.31 nm^3. This excluded the domains containing large insertions (5^{th} LDLr domain of the LDL receptor and the 8^{th}, 14^{th} and 31^{st}

	←-a-→	←---b---→	←-----	-------c-------	---→	d
Robson prediction	βtttt	tttββtttt	ttt βββ	ββββttttttttttt	tttt	t
Chou-Fasman prediction	ttttt	tttαββttt	tβββαβ	ββββttttttttttt	tttt	t
Hydropathy analysis	----B	---LBLB--	-LBB--	-B-BL-L-LB-L--	LL--	B

ldlr_1 human	AVGDRC	ERNEFQCQD	GKCISY	KWVCDGSAECQDGS	DESQET	C
ldlr_1 rabbit	AAGDKC	GRNEFQCRN	GKCISY	KWVCDGSSECQDGS	DEWEQT	C
ldlr_1 rat	AAEDSC	GKNEFQCRD	GKCIVS	KWVCDGSRECPDGS	DESPET	C
ldlr_2 human	LSVTC	KSGDFSCGGRVNRCIPQ	FWRCDGQVDCDNGS	DEQG	C
ldlr_2 rabbit	MSLTC	KSDDFSCGGRLNRCIPG	HWKCDGQQDCEDGS	DELG	C
ldlr_2 rat	MSVTC	RSGEFSCGGRVSRCIPD	SWRCDGRTDCENGS	DELD	C
ldlr_3 human	PPKTC	SQDEFRCHD	GKCISR	QFVCDSDRDCLDGS	DEAS	C
ldlr_3 rabbit	APKTC	SQDEFRCAE	GACISR	LFACDGEPDCPDGS	DEAS	C
ldlr_3 rat	SPKTC	SLDEFRCQD	GKCISR	QFVCDQDWDCLDGS	DEAH	C
ldlr_4 human	PVLTC	GPASFQCNS	STCIPQ	LWACDNDPDCEDGS	DEWPQR	C
ldlr_4 rabbit	APSTC	GPAHFRCNS	SSCVPA	LWACDGEPDCDDGS	DEWPAR	C
ldlr_4 rat	AATTC	GPAHFRCNS	SSCIPS	LWACDGDRDCDDGS	DEWPQN	C
ldlr_5 human	RGLYVFQGDSSPC	SAFEFHCLS	GECIHS	SWRCDGGPDCKDKS	DEEN	C
ldlr_5 rabbit	GARPSPQPGRGPC	SRHEFHCGS	GECVHA	SWRCDGDADCRDGS	DERD	C
ldlr_5 rat	GAEDTAAEVVSSPC	SSLEFHCGS	SECIHR	SWVCDGAADCKDKS	DEEN	C
ldlr_6 human	AVATC	RPDEFQCSD	GNCIHG	SRQCDREYDCKDMS	DEVG	C
ldlr_6 rabbit	AAATC	RPDEFQCSD	GTCIHG	SRQCDQQQDCGDMS	DEVG	C
ldlr_6 rat	AVTTC	RPDEFQCAD	GSCIHG	SRQCDREHDCKDMS	DELG	C
ldlr_7 human	VNVTLC	EGPNKFKCHS	GECITL	DKVCNMARDCRDWS	DEPIKE	C
ldlr_7 rabbit	VNVTLC	EGPDKFKCHS	GECISL	DKVCNSARDCQDWS	DEPIKE	C
ldlr_7 rat	INVTQC	DGPNKFKCHS	GECISL	DKVCNSARDCRDWS	DEPIKE	C
ldlr_related_1	IDAPKTC	SPKQFACRD	QITCISK	GWRCDGERDCPDGS	DEAPEI	C
ldlr_related_2	PQSKAQRC	QPNEHNCLG	TELCVPM	SRLCNGVQDCMDGS	DEGPH	C
ldlr_related_3	PPPQC	QPGEFACAN	SRCIQE	RWKCDGDNDCLDNS	DEAPAL	C
ldlr_related_4	HQHTC	PSDRFKCEN	NRCIPN	RWLCDGDNDCGNSE	DESNAT	C
ldlr_related_5	SARTC	PPNQFSCAS	GRCIPI	SWTCDLDDDCGDRS	DESAS	C
ldlr_related_6	AYPTC	FPLTQFTCNN	GRCINI	NWRCDNDNDCGDNS	DEAG	C
ldlr_related_7	SHSC	SSTQFKCNS	GRCIPE	HWTCDGDNDCGDYS	DETHAN	C
ldlr_related_8	TNQATRPPGGC	HTDEFQCRL	DGLCIPL	RWRCDGDTDCMDSS	DEKS	C
ldlr_related_9	EGVTHVC	DPSVKFGCKD	SARCISK	AWVCDGDNDCEDNS	DEEN	C
ldlr_related_10	ESLAC	RPPSHPCANNTSVCLPP	DKLCDGNDDCGDGS	DEGEL	C
ldlr_related_11	VNSSC	RAQDEFECAN	GECINF	SLTCDGVPHCKDKS	DEKPSY	C
ldlr_related_12	NSRRC	KKTFRQCSN	GRCVSN	MLWCNGADDCGDGS	DEIP	C
ldlr_related_13	NKTAC	GVGEFRCRD	GTCIGN	SSRCNQFVDCEDAS	DEMN	C
ldlr_related_14	SATDCSSYFRLGVKGVLFQPCER	TSLCYAP	SWVCDGANDCGDYS	DERD	C
ldlr_related_15	PGVKRPRC	PLNYFACPS	GRCIPM	SWTCDKEDDCEHGE	DETH	C
ldlr_related_16	NKFC	SEAQFECQN	HRCISK	QWLCDGSDDCGDGS	DEAAH	C
ldlr_related_17	EGKTC	GPSSFSCPG	THVCVPE	RWLCDGDKDCADGA	DESIAAG	C
ldlr_related_18	LYNSTC	DDREFMCQN	RQCIPK	HFVCDHDRDCADGS	DESPE	C
ldlr_related_19	EYPTC	GPSEFRCAN	GRCLSSRQWECDGENDCHDQS	DEAPKNPH	C
ldlr_related_20	TSPEHKC	NASSQFLCSS	GRCVAE	ALLCNGQDDCGDSS	DERG	C
ldlr_related_21	VSNC	TASQFVCKN	DKCIPF	WWKCDTEDDCGDHS	DEPPD	C
ldlr_related_22	PEFKC	RPGQFQCST	GICTNP	AFICDGDNDCQDNS	DEAN	C
ldlr_related_23	DIHVC	LPSQFKCTN	TNRCIPG	IFRCNGQDNCGDGE	DERD	C
ldlr_related_24	PEVTC	APNQFQCSI	TKRCIPR	VWVCDRDNDCVDGS	DEPAN	C
ldlr_related_25	TQMTC	GVDEFRCKD	SGRCIPA	RWKCDGEDDCGDGS	DEPKEE	C
ldlr_related_26	DERTC	EPYQFRCKN	NRCVPG	RWQCDYDNDCGDNS	DEES	C
ldlr_related_27	TPRPC	SESEFSCAN	GRCIAG	RWKCDGDHDCADGS	DEKD	C
ldlr_related_28	TPRC	DMDQFQCKS	GHCIPL	RWRCDADADCMDGS	DEEA	C
ldlr_related_29	GTGVRTC	PLDEFQCNN	TLCKPL	AWKCDGEDDCGDNS	DENPEE	C
ldlr_related_30	ARFVC	PPNRPFRCKN	DRVCLWI	GRQCDGTDNCGDGT	DEED	C
ldlr_related_31	EPPTAHTTHC	KDKKEFLCRN	QRCLSS	SLRCNMFDDCGDGS	DEED	C
GP330_1	C	GSLSFPCNN	GKCVPS	FFRCDGVDDCHDNS	DEXQ	C
GP330_2	GVFNNTC	SPSAFACVR	GGQCIPG	QWHCDRQNDCLDGS	DEXK	C
GP330_3	PTHATSSTC	PSTSFTCDN	HVCIPK	DWVCDTDNDCSDGS	DEXN	C
GP330_4	QASGTC	QPTQFRCPD	HRCISP	LYVCDGDKDCADGS	DEXG	C
GP330_5	VLNC	TSAQFKCAD	GSSCINS	RYRCDGVYDCRDNS	DEXG	C
GP330_6	PTRPPGMC	HPDEFQCQG	DGTCIPN	TWECDGHPDCIHGS	DEXXG	C
GP330_7	VPKTC	SPTHFLCDN	GNCIYK	AWICDGDNDCRDMS	DEXD	C
GP330_8	PTQPFHC	PSTQWQCPG	YSTCINL	SALCDGVFDCPNGT	DEXXL	C
GP330_9	RNCVPREC	SESEFRCAD	QQCIPS	RWVCDQENDCGDNS	DEXD	C
GP330_10	EMKTC	HPEHFQCTS	GHCVAK	ALACDGRADCLDAS	DEXA	C
GP330_11	PTRFPNGTYC	PAAMFECKN	HVCIQS	FWICDGENDCVDGS	DEXXXL	C
GP330_12	FNIPC	ESPQRFRCDN	SRCVYG	HQLCNGVDDCGDGS	DEXXXH	C
GP330_13	RKPTHKPC	TDTEYKCSN	GNCISQ	HYVCDNVNDCGDLS	DEXG	C
C6 human	IEEADC	KNKFRCDS	GRCIAR	KLECNGENDCGDNS	DERD	C
C7 human	TEEGC	GERFRCFS	GQCISK	SLVCNGDSDCDEDSA	DEDR	C
C8α human	RQAQC	GQDFQCKE	TGRCLKR	HLVCNGDQDCLDGS	DEDD	C
C8β human	SQVRC	EGFVCAQ	TGRCVNR	RLLCNGDNDCGDQS	DEAN	C
C9 human	DAEDDC	GNDFQCST	GRCIKM	RLRCNGDNDCGDFS	DEDD	C
C9 mouse	EIQENC	GNDFQCET	GRCIKR	RLLCNGDNDCGDYS	DEND	C
C9 trout	QALPSEC	SSIEFTCES	GACIKL	RLSCNGDYDCEDGS	DED	C
Factor I_1	QKADS	PMDDFFQCVN	GKYISQ	MKACDGINDCGDQS	DEL	C
Factor I_2	CKAC	QGKGFHCKS	GVCIPS	QYQCNGEVDCITGE	DEVG	C

domains of the LDLr related protein). The subset of the 9 LDLr domains in the complement proteins had a similar volume of 5.17 ± 0.16 nm^3. The cluster of negatively charged residues at the C-terminus of the domain (^{29}Asp-Cys-X-Asp-Gly-Ser-Asp-Glu^{36}) implicated in ligand binding (Yamamoto *et al.*, 1984) was well conserved. Summation of the charged groups showed that 73 of the 74 LDLr domains were negatively charged. Depending on whether the His residues were protonated or not, the average net charge ranged from -3.6 to -4.5 (± 1.8). In the complement LDLr domains, the average net charge was higher at -4.9 to -5.1. In the LDLr protein, this highly negatively charged domain interacts electrostatically with the positively charged ligands apoprotein B and apoprotein E of LDL (Südhof *et al.*, 1985). It was proposed that the homologous domain in the complement components also acted as a binding site (Scott, 1989; Tschopp & Masson, 1987). Since C9 acts as both a receptor and a ligand during polymerisation, presumably there must be a positively charged domain analogous to Apo-B/E. Stanley & Herz (1987) found that residues 415 to 425 in human C9 had homology with the apo-E binding site for the LDL receptor and there were several neighbouring positively charged residues.

The averaged Robson and Chou-Fasman structure predictions are shown in Figures 8.6(a) to 8.6(c). Robson calculations were made with decision constants corresponding to unbiased, α-biased and β-biased secondary structures (Chapter 2). All three predictions show very high amounts of turn conformations, 99% β-turn, 1% β-sheet; 82% β-turn, 18% α-helix; and 77% β-turn, 23% β-sheet respectively. The Chou-Fasman calculation gave a prediction of 72% β-turn, 5%

Figure 8.5. Alignment of 74 LDLr sequences found in the complement components and other proteins.

The sequence is divided into four segments labelled a to d (Table 8.2). The consensus Robson and Chou-Fasman secondary structure predictions are shown above the sequences (β, β-sheet; α, α-helix; t, β-turn), together with the consensus hydropathy of the residue (B, hydrophobic; L, hydrophilic). X: residues unreported in the GP330 protein sequence (Raychowdhury *et al.*, 1989).

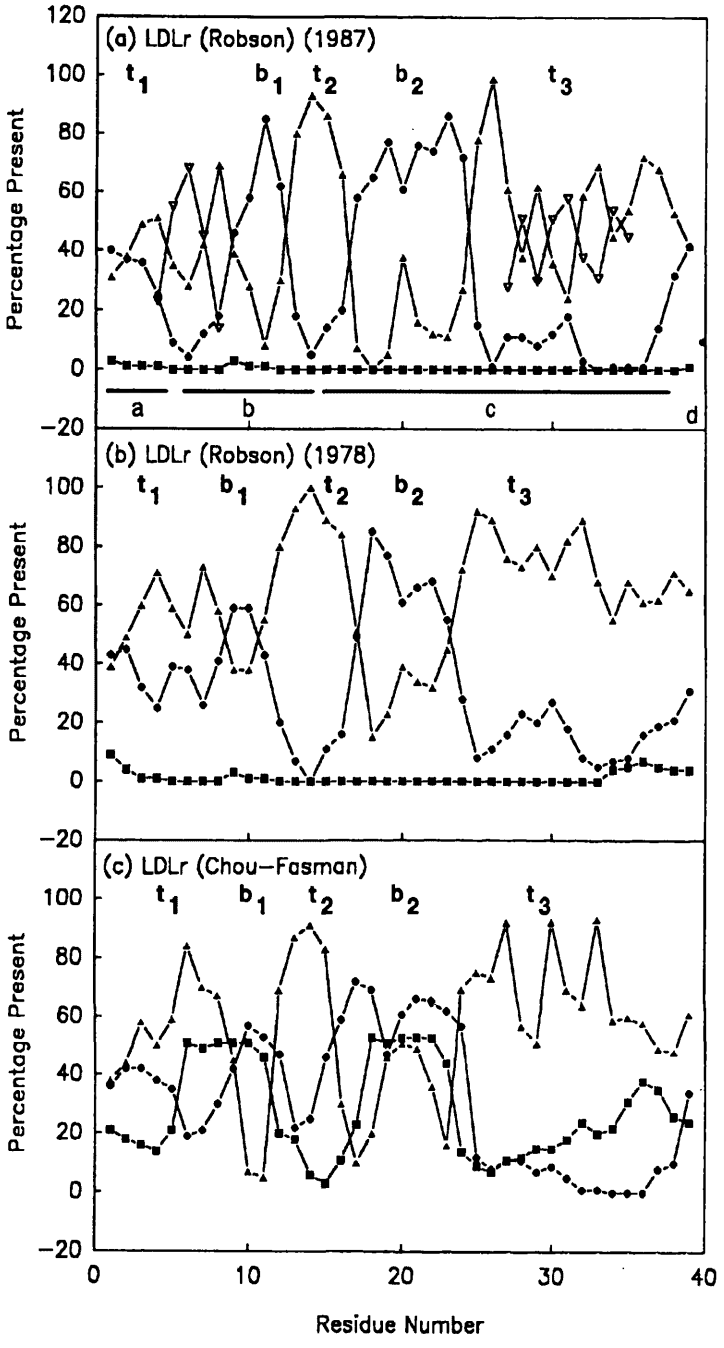


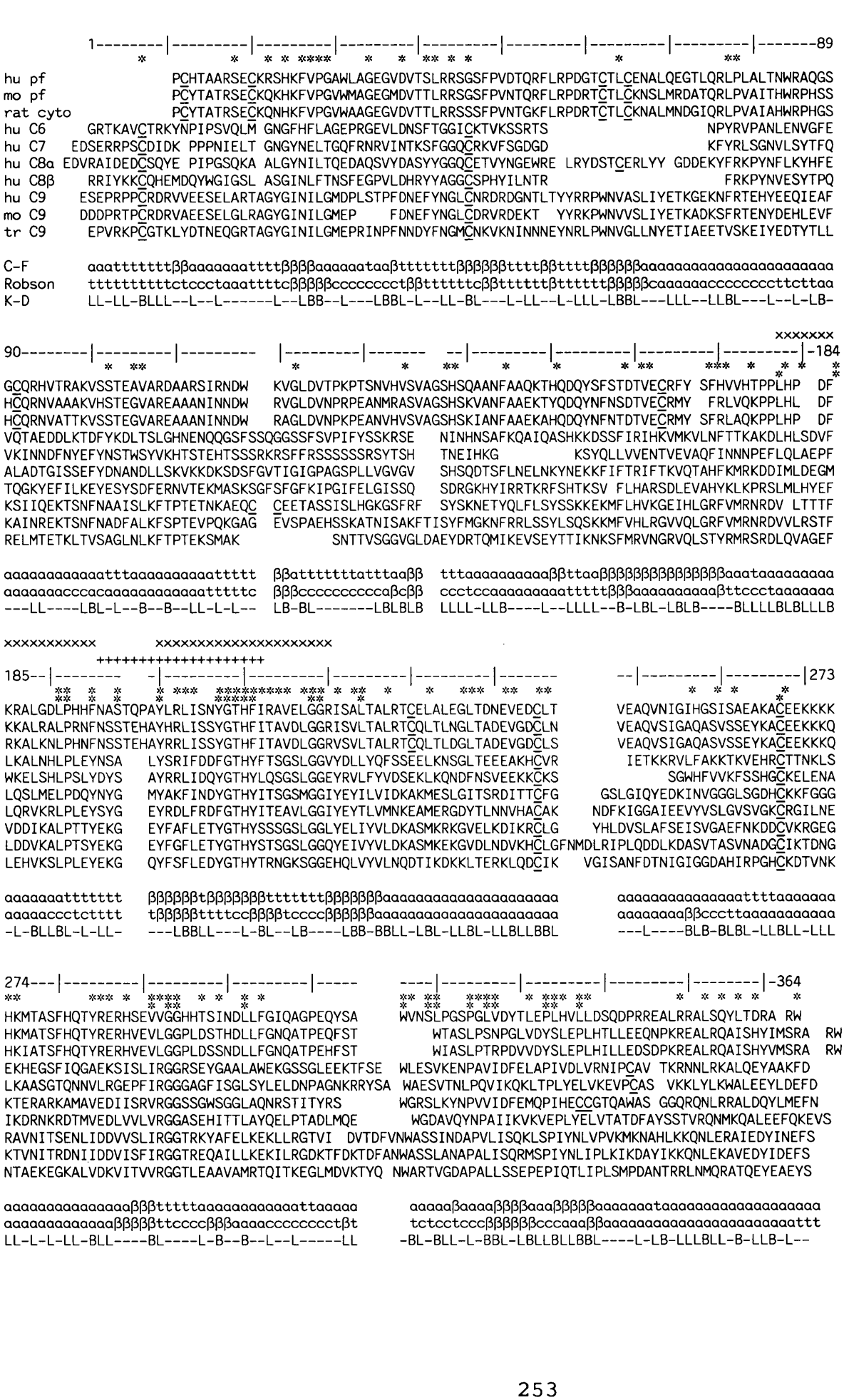
Figure 8.6. Consensus secondary structure predictions for the LDLr motif.

(a), (b) and (c) Robson 1987 and 1978 propensity set and Chou-Fasman analyses for the 74 LDLr sequences (■; α-helix; ●; β-sheet; ▲; β-turn). The coil predictions in the Robson analysis are only shown [▽ in (a)] when significant for reason of clarity.

α-helix and 23% β-sheet (Figure 8.6(c)) which compared well with the β-biased Robson prediction. Predictions of α-helix of 1 to 3 residues long were discounted since the minimum length of an α-helix is 4 residues (Kabsch & Sander, 1983). The α-biased Robson prediction suggested a single α-helix between residues 35 to 38 (data not shown), while the Chou-Fasman prediction did not indicate α-helices longer than 1 residue (Figure 8.6(c)). The occurrence of α-helices in the LDLr was not favoured by these predictions.

Figures 8.6(a) to 8.6(c) show that over 50%-60% of the LDLr predictions favour regions of β-turn or loop conformations between residues 1 to 8 (t_1), 12 to 15 (t_2), and 24 to 39 (t_3), interspersed with two β-strands which occur approximately between residues 9 to 11 (b_1), and 16 to 23 (b_2). The three regions of sequence gaps and insertions occur in the regions of turn predictions as expected. The total amino acid composition of the 3008 residues in Figure 8.5 was computed and showed that the eight most abundant residues were Cys (14.6%), Asp (13.1%), Ser (8.5%), Gly (8.1%), Glu (6.7%), Arg (5.5%), Asn (5.2%) and Pro (5.1%). This composition was consistent with the high amount of turns predicted for this domain. Wilmot & Thornton (1988) reported that in a type I turn Asp, Asn, Ser and Cys were preferred at the first position *i*, Asp, Ser, Thr and Pro at *i+1*, Asp, Ser, Asn and Arg at *i+2*, and Gly, Trp and Met at *i+3*, while in a type II turn Pro was preferred at position *i+1*, Gly and Asn at *i+1*, and Gln and Arg at *i+3*.

The averaged hydropathy profiles were calculated for the 74 LDLr sequences. Use of the above four hydrophilicity or hydrophobicity scales indicated that, of the 39 residue positions, 9 were predominantly hydrophobic (this includes all six Cys residues), and at least 9 were predominantly hydrophilic (this includes the five most conserved Asp/Glu positions: Table 8.2). The three non-Cys hydrophobic residues were all located in the two predicted β-strands. The mean hydropathy was separately calculated for the 9 complement LDLr sequences to show that just these 9 hydrophobic positions were principally conserved, while variability was observed in the positions of the most hydrophilic residues. This suggests that while the LDLr domains have a common hydrophobic core, they possess a range of hydrophilic surfaces that determines its functional role.



(8.3.3) Consensus analysis of the PLR

Ten sequences for the PLR in the late complement components were aligned (Figure 8.7) manually and using MULTAL. The mean length of the PLR was 355.7 ± 12.4 residues, the consensus length was 364 residues and the mean dry volume was 51.4 ± 1.9 nm^3. Of the 364 residues, 107 were conserved in 60-100% of the sequences (Figure 8.7). The perforin and cytolysin sequences generally differed from the complement ones in the first 90 residues (Tschopp & Nabholz, 1990). The highest conservation occurred in two 47-residue segments between residues 178 and 224 and between residues 291 to 337. These two regions contain 29 of the 30 residues or residue groups that are 90-100% conserved in the 10 sequences. The first region has been proposed to contain the possible membrane-spanning regions of C9 (Figure 8.7; Stanley & Herz, 1987; Peitsch *et al.*, 1990).

The average of the secondary structure predictions was computed, based on the Robson (unbiased) and Chou-Fasman methods (Figure 8.7). The 1978 Robson prediction gave 40% α-helix, 16% β-sheet, 21% β-turn and 23% coil, while the Chou-Fasman prediction gave 58% α-helix, 21 % β-sheet and 21% β-turn. FT-IR spectroscopy indicated that C9 has appreciable amounts of α-helix as well as β-sheet (S.J. Perkins, unpublished results) and circular dichroism indicated the presence of 24% α-helix and 32% β-sheet in C9 (Tschopp *et al.*, 1982). Since

---

Figure 8.7. Alignment of 10 PLR sequences found in the complement components and other proteins.

The secondary structure prediction by the Robson and Chou-Fasman methods are shown beneath the sequences (β, β-sheet; α, α-helix; t, β-turn; c, coil), together with the mean Kyte-Doolittle calculation of the most hydrophobic or hydrophilic residues. This is denoted B or L respectively when this index is less than -1 or greater than +1 respectively, according to the Kyte & Doolittle (1982) scale. Possible membrane-interacting regions of the PLR are marked by + (Stanley & Herz, 1987) and x (Peitsch *et al.*, 1990) above the sequences.

secondary structure predications of the TSR, LDLr and EGF domains suggest the presence of only β-sheet and β-turn, this α-helical component of C9 was assigned to the PLR.

Hydropathy analyses by all four methods resulted in the identification of large numbers of hydrophilic and hydrophobic residues. For example, the Kyte-Doolittle analysis (Figure 8.7) showed that, of the 364 residues, 40% were markedly hydrophilic while 17% were markedly hydrophobic. No extensive hydrophobic regions that might form membrane-spanning helices were identified (DiScipio *et al.*, 1984; Stanley *et al.*, 1985) which should be of at least 20 residues (Lear *et al.*, 1988). The two conserved segments (residues 178-224 and 291-337) also exhibited similar amounts, i.e. 34-38% hydrophilic and 21-23% hydrophobic residues (with one conserved charged group at His210). The total net charge of the 10 PLRs varies between -2 to +25 (or between -8 to +9 if the protonation of His is discounted).

(8.4) CONCLUSIONS

The four most abundant domain types found in the principal complement components are the SCR, the TSR, the EGF and the LDLr, respectively (Perkins *et al.*, 1990c). Averaged secondary structure predictions have now been carried out for 101 SCR, 31 TSR, 99 EGF and 74 LDLr sequences (Perkins *et al.*, 1988, 1989; Chapter 6; Section 8.3). All four domains contain β-sheet and β-turn structures. Such β-structures are also found in the immunoglobulin (Ig) folds of IgG and IgM antibodies (length ca: 110 residues) which are involved in the classical pathway activation of complement. Tertiary structures are known for the SCR and EGF from 2D-NMR studies (Barlow *et al.*, 1991a,b), and for variable- and constant-type Ig folds from crystallography. Provided that β-structure had already been identified by spectroscopic means or other independent evidence, the 1978 β-biased Robson predictions were the most successful in identifying the β-strands in the averaged structure prediction. On a residue-for-residue basis, the agreement of the prediction with experimental structures was 69% for the SCR, 69-71% for the EGF, 56-58% for the variable Ig fold of IgG (Perkins *et al.*, 1989), and 71-78% for the constant Ig fold in the α_3 domain of MHC Class I (Tysoe-Calnon *et al.*, 1991). These figures can be compared with the 50-60% accuracies of such predictions for single sequences (Argos, 1989). Summation of the averaged predicted secondary structures for the TSR, LDLr, PLR and EGF in C9 gives 29% α-helix, 22% β-sheet, 32% β-turn and 17% coil by the 1978 Robson method, and 47% α-helix, 21% β-sheet and 32% β-turn by the Chou-Fasman method. It is however necessary to confirm these predictive analyses by experimental techniques such as FT-IR, and such analyses are in progress. The circular dichroism result of 24% α-helix and 32% β-sheet for C9 (Tschopp *et al.*, 1982) are in reasonable agreement with these results. The main advantage of consensus averaging is to smoothen and group together the most significant predicted structures. All of the predictions bear an overall resemblance to the observed secondary structure.

Limited structure predictions for human C6, C7, C8α, C8β and C9 have been reported, in many cases for single sequences only (Stanley *et al.*, 1986;

Shiver *et al.*, 1986; Peitsch *et al.*, 1990). The secondary structure of residues 1-470 of C6 was predicted to be predominantly β-sheet and β-turn in the TSR and LDLr domains and more α-helical in the PLR. A Kyte and Doolittle hydrophobicity profile suggested that although residues 1-470 were predominantly hydrophilic in nature, regions which were predicted to be hydrophobic occurred in the PLR (Chakravarti *et al.*, 1989). C7 was found to be also dominated by β-sheets and β-turns and a small amount of α-helix was found in the middle and the C-terminus of the sequence and also was predominantly hydrophilic (DiScipio *et al.*, 1988). C8α is mainly hydrophilic but contains a prominent hydrophobic region in the PLR which may act as an α-helical transmembrane domain (Rao *et al.*, 1987). C8β is also largely hydrophilic although a number of regions were identified as forming putative amphipathic β-structures (Howard *et al.*, 1987). The average secondary structures for human, mouse and trout C9 were: α-helix, 30% and β-sheet, 18% by the Chou-Fasman method (Stanley & Herz, 1987).

Other structural properties of the TSR, EGF and LDLr domains can be summarised. The LDLr sequences are well-defined in length and can be readily analysed as a single group of sequences, while the TSR and EGF sequences are variable in length and each can be subdivided into 3 distinct subgroups. Residue conservation is highest in the LDLr sequences. Generally, the Groups 1 + 3 TSR and Group C EGF motifs are positively charged, while the Groups A + B EGF and the LDLr motifs are negatively charged, and the SCR and Group 2 TSR motifs are uncharged.

In terms of the structural properties of C9, two further general results were obtained. Firstly, despite widely-held assumptions that the late complement components and perforin contain a region which is homologous to a EGF-like domain, it is clear that the EGF motifs in these proteins are structurally distinct from the classical EGF structure. Evidence for this was found in much reduced sequence homology and residue conservation in Group C, as well as in differences in the net charges, structure predictions and hydropathy analyses. In order to determine whether or not this group of putative EGFs belong to this superfamily, it will be necessary to verify their disulphide bridge connectivities, as well as to

embark on 3-D structural studies of these EGFs. Secondly, all regions of C9 possess hydrophilic structures. Apart from a hydrophobic segment in the putative EGF of C9, there are no clear indications of hydrophobic α-helices or β-strands which will interact with or span membranes.

The role of the three cysteine-rich domains in C9 may be two-fold. Firstly, cysteine-rich domains are likely to be very stable due to the disulphide bridges and are therefore ideal as rigid, binding domains. It has been proposed that during the assembly of the MAC, poly-C9 and polyperforin, domains with a negative and positive charge distribution are of functional importance in the binding of the molecules (Tschopp & Masson, 1987). Secondly, the cysteine-rich domains may confer a certain amount of stability of the perforin domain. It is thought that the EGF domain in the endothelial leukocyte adhesion molecule-1 acts to stabilize the lectin domain which mediates cell adhesion (Pigott *et al.*, 1991).

CHAPTER 9.

CONCLUSIONS

The majority of the complement components are multi-domain proteins. The determination the complete protein structure requires the detailed knowledge of the structure of each individual domain and also the arrangement of these domains in space. High resolution techniques such as X-ray crystallography and 2 dimensional NMR have been used to ascertain the structures of the C3a and C5a fragments (Huber *et al.*, 1980; Nettesheim *et al.*, 1988; Zuiderweg *et al.*, 1989), factor D (Narayana *et al.*, 1991a, b), several serine proteases (e.g. trypsin (Bode & Schwager, 1975)), α_1-antitrypsin (Löbermann *et al.*, 1984) (homologous to the serpin domain of $C\bar{1}$ inhibitor), 2 SCR domains in factor H (Barlow *et al.*, 1991a, b; Norman *et al.*, 1991) and the EGF protein and its homologs (Cooke *et al.*, 1987, 1990) (Chapters 1 & 2). Obviously much more work is needed if the tertiary structures of all the domains found in the complement components are to be determined.

The most powerful techniques used to study the arrangement of the domains at the present time are those of neutron and X-ray small-angle solution scattering. The M_r, degree of oligermization and the degree of elongation can be determined. Large conformational changes can also be detected. Simple structural models can be built based on known tertiary structures or electron micrographs. Secondary structure prediction methods (Garnier *et al.*, 1978; Chou & Fasman, 1978) are useful to ascertain the degree of homology of a domain of unknown structure with one of known structure and allow physical attributes such as the average volume to be determined.

The R_G is a measure of molecular elongation, such that the larger the R_G, the more elongated the molecule. The ratio R_G/R_O compares the R_G of a molecule with the R_G of a sphere of the same volume. The average value for this ratio from the analysis of 47 proteins was found to be 1.28 ± 0.10 (Perkins, 1988). As can be seen from Table 9.1, this ratio for the complement components and associated proteins (IgG and IgM) and their fragments, is approaching 2. This indicates that in solution, complement proteins have highly elongated structures. An independent verification of this is shown by the sedimentation coefficients and the frictional ratios f/f_o also shown in Table 9.1.

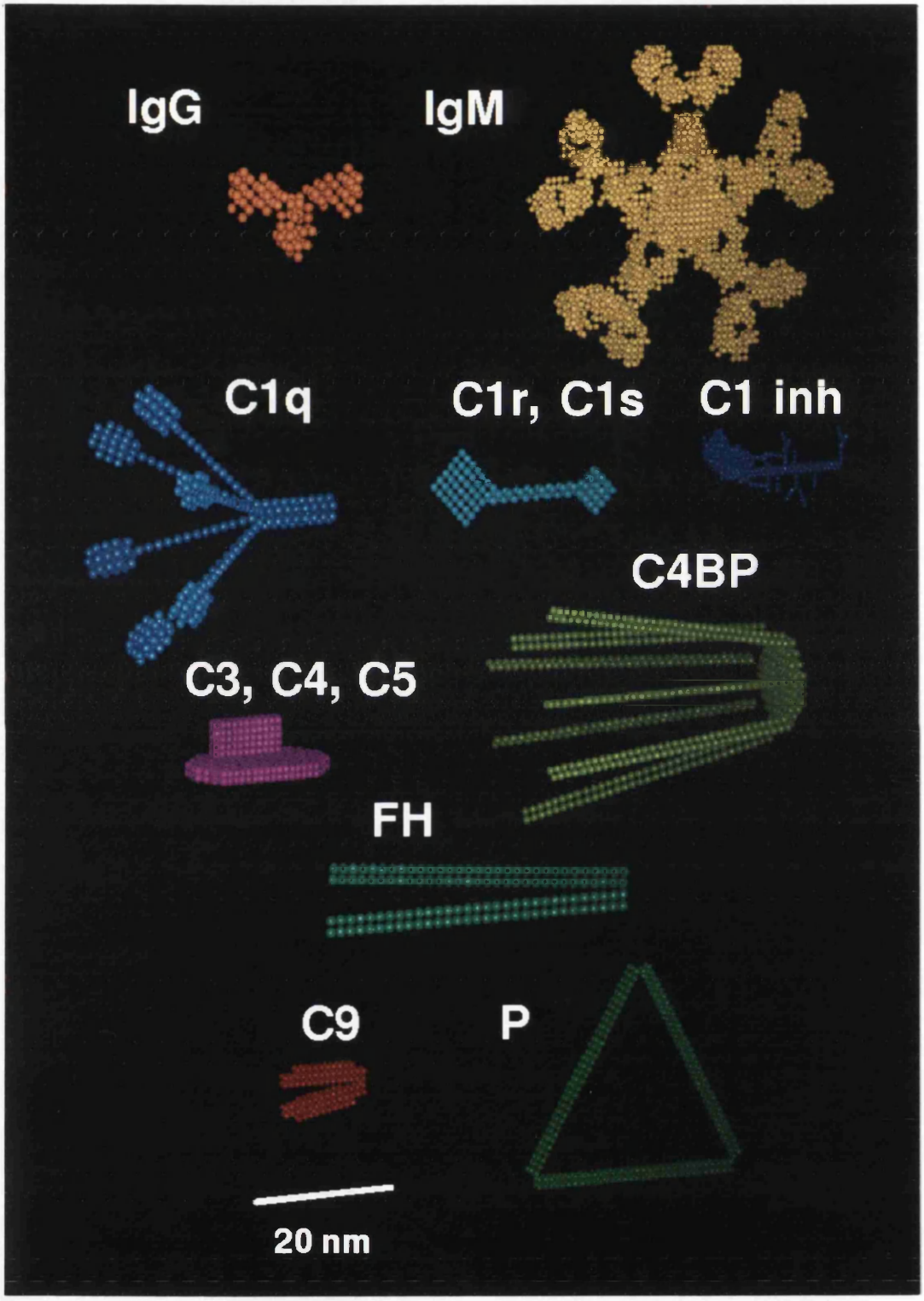
Name	R_G (nm)	R_G/R_O	$S^0_{20,w}$	f/f_o
C1q	12.8	3.2	10.2	1.91
$C1r_2$	8 ± 1	2.8	7.1	1.45
$C1s_2$	8 ± 1	3.6	4.5	1.36
$C1r_2C1s_2$	17	4.7	8.7	1.82
C1	12.6	2.6	15.2-16.2	1.79
IgM human	12.17 ± 0.34	2.4	17.7	1.78
IgM mouse	12.47 ± 0.35	2.4	(17.7)	1.77
Fc_5 human	6.15 ± 0.26	1.6	11.4	1.64
IgM-S human	6.10 ± 0.36		7.4	1.47
Fab'_2 human	4.93 ± 0.09	1.9	6.2	1.36
Fab'_2 mouse	4.82 ± 0.12	1.8	(6.2)	1.41
Fab human	2.94 ± 0.09	1.6	3.7	1.14
Fab mouse	2.90 ± 0.11	1.5	(3.7)	1.17
α_1-antitrypsin	2.61 ± 0.02	1.3	3.3-3.45	1.32
C1 inhibitor	4.85 ± 0.01	2.3	3.67-3.7	1.54
C3	5.1 ± 0.1	1.7	7.0-7.3	1.40
C3c	4.8 ± 0.1	1.8	5.5-6.7	1.37
C3dg	2.7 ± 0.2	1.5	2.6	1.35
C4	4.9 ± 0.04	1.6		
$C4c_2$	4.9 ± 0.03	1.4		
$C4d_2$	2.7 ± 0.01	1.5		
C5	4.9 ± 0.1	1.6		
C4BP	13	3.2	10.7-11.2	1.89
Properdin (dimer)	7.3 ± 1.6	3.0		
(trimer)	10.7 ± 0.5	3.8	5.2-5.5	1.95
C6			5.5-6.0	1.34
C7			5.1-5.6	1.31
C8α, C8β			8.0-8.1	1.19
C9	3.33 ± 0.03	1.6	4.5-4.7	1.22
Factor H	12.6 ± 0.3	3.6	5.5-5.6	2.7

Table 9.1. Elongation parameters of some of the complement components and associated proteins.

R_G/R_O ratios and f/f_o ratios are calculated from sequences.

The table is based upon Table 1 in Perkins *et al.*, 1990c. The values for IgM and its fragments are from Perkins *et al.*, 1991a and factor H (assumed to be a dimer) from Perkins *et al.*, 1991b. Properdin and the late complement components are published in Chapters 6 & 7.

Theoretical scattering curves can be calculated from models created by a trial-and-error procedure from small Debye spheres (Chapters 2 & 3) and compared with the experimentally derived scattering curves. The Debye sphere models for the complement proteins so far determined are shown in Figure 1. Although these models are not unique, they are constrained by the dry volume calculated from the primary structure, and the R_G and molecular dimensions determined from the experimental scattering curves.


IgG
IgM
C1q
C1r, C1s
C1 inh
C4BP
C3, C4, C5
FH
C9
P
20 nm

Secondary structure predictions have been performed on the SCR, EGF, TSR, LDLr, PLR, vWF and the SP domains (Chapters 6 & 8; Perkins *et al.*, 1988; Perkins & Smith, unpublished data). All these domains have large amounts of β-strand and β-turn. The PLR and the vWF also contain significant amounts of α-helix.

In multi-domain mosaic proteins, it is generally presumed that each domain is associated with a specific function. When the serpin domain of $C\bar{1}$ inhibitor binds to the serine protease domain of $C\bar{1}r$ and $C\bar{1}s$, the N-terminal domain of $C\bar{1}$ inhibitor may block the accessibility of the SCR domains of $C\bar{1}r$ and $C\bar{1}s$. Hydrodynamic modelling methods show that in complexes formed between $C\bar{1}$ inhibitor and $C\bar{1}r$ and $C\bar{1}s$, the N-terminal domain is found close to the two SCR domains (Perkins, 1990). In the W-model, S-model or 8-model for $C\bar{1}$ (Chapter 1), the two SCR domains in each of $C\bar{1}r$ and $C\bar{1}s$ interact with the collagenous stalks of C1q. The steric hindrance offered by the N-terminal domain of $C\bar{1}$ inhibitor in the region of these SCR domains in these complexes will prevent their binding to C1q and also prevent the binding of C2 and C4 binding to $C\bar{1}r$ and $C\bar{1}s$ after these have reacted with $C\bar{1}$ inhibitor.

By analogy with structural studies on the formation of C3b from C3 and C4b from C4, C5b will be of similar length to C5 which is 18 nm long (Figure 9.1) (Perkins & Sim, 1986; Perkins *et al*, 1990b). This length is similar to the overall

---

Figure 9.1 Debye sphere models of components of the complement cascade. The structures of $C\bar{1}$ inhibitor (Chapter 4) (Smith *et al.*, 1990; Perkins *et al.*, 1990d), C5 (Chapter 5) (Perkins *et al.*, 1990b), properdin (Chapter 6) (Smith *et al.*, 1991) and C9 (Chapter 7) (Smith *et al.*, 1992) are presented in this thesis. The structure of and interactions between IgG, IgM and C1r, C1s with C1q were proposed in Perkins (1985), Perkins *et al.* (1991b) and Perkins & Nealis (1989). The structures of C3 and C4 were determined by Perkins & Sim (1986) and Perkins *et al.* (1990a). C4BP and monomeric factor H models were presented in Perkins *et al.* (1986) and Perkins *et al.* (1991a).

length of C6 and C7. The function of C5b may be to act as an extended template for complex formation with C6. The length of C5b is sufficient to interact with the two SCR and FIM domains of length 7 nm at the C-terminus of C6 (DiScipio & Hugli, 1989; Haefliger *et al.*, 1989) and also with the C9-like segment of C6 in order to facilitate the unfolding of C6. The unfolded C6 structure would in turn be able to unfold sequentially the homologous proteins C7, C8 and *n* copies of C9, leading to MAC formation and insertion into membranes. Electron microscopy data on complexes formed between C5b6 and C7 have been discussed by DiScipio *et al.* (1988) and these authors propose that the final form of these complexes possess highly extended structures. Solution scattering data on C6, C7 and C8 and their complexes with C5b in conditions close to physiological will extend structural knowledge of these complexes.

Properdin has a monomer length is 26 nm and associates with C3b, which is of maximum length 18 nm (Figure 9.1) (Perkins & Sim, 1986; Perkins *et al.*, 1990a; Ikai *et al.*, 1990). C3b interacts with properdin within a 34-residue segment of C3b (Figure 5.5) (Daoudaki *et al.*, 1988), which if in a maximally extended conformation would be of length 12 nm. These considerations show that it is reasonable to expect that for steric reasons more than one C3b molecule can interact with each of P_2, P_3 and P_4 in agreement with evidence from Pangburn (1989) who showed that the specific activity of properdin decreases in the order $P_4 > P_3 > P_2$.

C9 contains 4 domains, the TSR, LDLr, PLR and EGF-like domain. C9 was successfully modelled as a V-shaped structure. It is presumed that the TSR, LDLr and the EGF domains lie at the ends to the V, whilst the PLR occurs at the angle between the two arms. Such V-shaped models can explain the unfolding of C9 when this is polymerised or inserted into membranes (Stanley *et al.*, 1986). The comparison of C9 and perforin is suggestive that the N-terminal TSR and LDLr domains protect the EGF domain and PLR of free C9 when this is secreted or circulating in plasma, which otherwise would be inserted into membranes. This is the case with perforin in the presence of Ca^{2+} when it is secreted by cytotoxic cells (Tschopp & Nabholz, 1990). This suggests that the EGF domain and PLR are

implicated with the insertion of C9 into membranes, and this only occurs when the TSR and LDLr domain structure of C9 is unfolded to expose a more hydrophobic surface. Such an unfolding has been shown from electron microscopy (Podack & Tschopp, 1982a, b). The affinity of the C9b fragment for lipids also argues in favour of a membrane attachment role for the EGF domain and PLR (Biesecker *et al.*, 1982; Ishida *et al.*, 1982). In addition to this, it has been proposed that the LDLr domain facilitates the polymerization of C9, as evidenced from the use of the monoclonal antibody M34 and the use of LDLr inhibitors (Stanley *et al.*, 1986; Haefliger *et al.*, 1987). In fact, the existence of multiple copies of TSR and LDLr domains in properdin and factor I of the complement activation pathways is suggestive that these play a role in protein-protein interactions in the formation of the MAC from C6, C7, C8 and C9.

Solution structure analyses of the other members of the complement cascade are underway. More high resolution studies on individual domains need to be established to enable an accurate determination of the domain arrangement within a protein structure. Once complete protein structures have been elucidated, solution scattering can be used to study complexes formed between proteins, which are the crux of the complement cascade.

REFERENCES

Alper, C.A., Colten, H.R., Rosen, F.S., Rabson, A.R., MacNab, G.M., Gear, J.S.S. (1972) Lancet *ii*, 1179-1181.

Al Salihi, A.A., Ripoche, J., Pruvost, L. & Fontaine, M. (1982) FEBS Lett. **150**, 238-242.

Al Salihi, A.A., Ripoche, J. & Fontaine, M. (1988) Mol. Immunol. **25**, 367-377.

Anfinsen, C.B. (1973) Science **181**, 233-239.

Anfinsen, C.B., Haber, E., Sela, M. & White, F.H. (1961) Proc. Natl. Acad. Sci. U.S.A. **47**, 1309-1314.

Appella, E., Weber, I.T., & Blasi, F. (1988) FEBS Lett. **231**, 1-4.

Argos, P. (1989) in Protein structure: a practical approach (Creighton, T. E., Ed.) pp 169-189, IRL Press, Oxford, U.K.

Arlaud, G.J. & Gagnon, J. (1983) Biochemistry **22**, 1758-1764.

Arlaud, G.J., Gagnon, J. & Porter, R.R. (1982) Biochem. J. **201**, 49-59.

Arlaud, G.J., Gagnon, J., Villiers, C.L. & Colomb, M.G. (1986) Biochemistry **25**, 5177-5182.

Arlaud, G.J., Colomb, M.G. & Gagnon, J. (1987) Immunol. Today **8**, 106-111.

Artymiuk, P.J. & Blake, C.C.F. (1981) J. Mol. Biol. **152**, 737-762.

Atkinson, J.P. & Farries, T. (1987) Immunol. Today **8**, 212-215.

Au, Y.P.T., Wang, T.W., & Clowes, A.W. (1990) Nucl. Acids Res. **18**, 3411-3411.

Aude, C.A., Lacroix, M.B., Arlaud, G.J., Gagnon, J. & Colomb, M.G. (1988) Biochemistry **27**, 8641-8648.

Aulak, K.S. & Harrison, R.A. (1990) Biochem. J. **271**, 565-569.

Aulak, K.S., Pemberton, P.A., Rosen, F.S., Carrell, R.W., Lachmann, P.J. & Harrison, R.A. (1988) Biochem. J. **253**, 615-618.

Avron, A., Reeve, F.H., Lickorish, J.M. & Carrell, R.W. (1991) FEBS Lett. **280**, 41-43.

Baldwin, J. & Chothia, C. (1979) J. Mol. Biol. **129**, 175-220.

Bale, M.D., Westrick, L G., & Mosher, D.F. (1985) J. Biol. Chem. **260**, 7502-7508.

Bao, J-J., Sifers, R.N., Kidd, V.J., Ledley, F.D. & Woo, S.L.C. (1987) Biochemistry **26**, 7755-7759.

Barlow, P.N., Baron, M., Norman, D.G., Day, A.J., Willis, A.C., Sim, R.B. & Campbell, I.D. (1991a) Biochemistry **30**, 997-1004.

Barlow, P.N., Norman, D.G., Pearce, J.M., Horne, T.J., Steinkasserer, A., Sim,

R.B. & Campbell, I.D. (1991b) Complement Inflamm. **8**, 126 (Abst.).

Baron, M., Norman, D.G. & Campbell, I.D. (1991) Trends Biochem. Sci. **16**, 13-17.

Baumann, U., Huber, R., Bode, W., Grosse, D., Lesjak, M. & Laurell, C.B. (1991) J. Mol. Biol. **218**, 595-606.

Beatty, K., Bieth, J. & Travis, J. (1980) J. Biol. Chem. **255**, 3931-3934.

Bell, G.I., Fong, N.M., Stempien, M.M., Wormsted, M.A., Caput, D., Ku, L., Urdea, M.S., Rall, L.B., & Sanchez-Pescador, R. (1986) Nucl. Acids Res. **14**, 8427-8446.

Belt, K.T., Carroll, M.C. & Porter, R.R. (1984) Cell **36**, 907-914.

Bentley, D.R. (1986) Biochem. J. **239**, 339-345.

Berger, M., Balow, J.E., Wilson, C.B. & Frank, M.M. (1983) New Engl. J. Med. **308**, 1009-1012.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) J. Mol. Biol. **112**, 535-542.

Bhakdi, S., & Tranum-Jensen J. (1986) J. Immunol. **136**, 2999-3005.

Bhakdi, S. & Tranum-Jensen, J. (1991) Immunol. Today **12**, 318-320.

Biesecker, G., & Müller-Eberhard, H.J. (1980) J. Immunol. **124**, 1291-1296.

Biesecker, G., Gerard, C., & Hugli, T.E. (1982) J. Biol. Chem. **261**, 2584-2590.

Biesecker, G., Lachmann, P.J. & Henderson, R. (1991) Complement Inflamm. **8**, 128. (Abstr.)

Blake, C.C.F., Geisow, M.J., Oatley, S.J., Rerat, B., & Rerat, C. (1978) J. Mol. Biol. **121**, 339-356.

Bloch, E.F., Schmetz, M.A., Foulds, J., Hammer, C.H., Frank, M.M. & Joiner, K.A. (1987) J. Immunol. **138**, 842-848.

Blundell, T.L. & Johnson, L.N. (1976) Protein Crystallography, pp. 428-429, Academic Press, London.

Bock, S-C., Skriver, K., Nielsen, E., Thøgersen, H-C., Wiman, B., Donaldson, V.H., Eddy, R.L., Marrian, J., Radziejewska, E., Huber, R., Shows, T.B. & Magnusson, S. (1986) Biochemistry **25**, 4292-4301.

Bode, W. & Schwager, P. (1975) J. Mol. Biol. **98**, 693-717.

Bokisch, V.A. & Müller-Eberhard, H.J. (1970) J. Clin. Invest. **49**, 2427-2434.

Bornstein, P., Alfi, D., Devarayalu, S., Framson, P., & Li, P. (1990) J. Biol. Chem.

265, 16691-16698.

Borsos, T. (1989) Behring Inst. Mitt. **84**, 93-101.

Borsos, T., Dourmashkin, R.R., & Humphrey, J.H. (1964) Nature (London) **202**, 251-252.

Boulay, F., Mery, L., Tardif, M., Brouchon, L. & Vignais, P. (1991) Biochemistry **30**, 2993-2999.

Brodsky-Doyle, B., Leonard, K.R. & Reid, K.B.M. (1976) Biochem. J. **159**, 279-286.

Brown, S.C., Mueller, L., & Jeffs, P.W. (1989) Biochemistry **28**, 593-599.

Brown, W.M., Dziegielewska, K.M., Foreman, R.C., Saunders, N.R. & Wu, Y. (1989) Nucl. Acids Res. **17**, 6398.

Bruch, M., Weiss, V. & Engel, J. (1988) J. Biol. Chem. **263**, 16626-16630.

Bundy, H.F. & Mehl, J.W. (1959) J. Biol. Chem. **234**, 1124-1128.

Byler, D.M. & Susi, H. (1986) Methods Enzymol. **130**, 290-311.

Campbell, I.D., Baron, M., Cooke, R.M., Dudgeon, T.J., Fallon, A.F., Harvey, T.S., & Tappin, M.J. (1990) Biochem. Pharmaco. **40**, 35-40.

Caras, I.W., Davitz, M.A., Rhee, L., Weddell, G., Martin, D.W. & Nussenzweig, V. (1987) Nature (London) **325**, 545-549.

Carp, H., Miller, F., Hoidal, J.R. & Janoff, A. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 2041-2045.

Carpenter, G., & Cohen, S. (1990) J. Biol. Chem. **265**, 7709-7712.

Carpenter, G., & Wahl, M. I. (1990) Handbook Exp. Pharmacol. **95/I**, 69-171.

Carrell, L. & Owen, M.C. (1985) Nature (London) **317**, 730-732.

Carrell, R. & Travis, J. (1985) Trends Biochem. Sci. **10**, 20-24.

Carrell, L., Boswell, D.R., Brennan, S.O. & Owen, M.C. (1980) Biochem. Biophys. Res. Commun. **93**, 399-402.

Carrell, L., Jeppsson, J-O, Vaughan, L., Brennan, S.O., Owen, M.C. & Boswell, D.R. (1981) FEBS Lett. **135**a1-, 301-305.

Carrell, L., Jeppsson, J-O, Laurell, C-B., Brennan, S.O., Owen, M.C., Vaughan, L. & Boswell, D.R. (1982) Nature (London) **298**, 329-334.

Carrell, R.W., Evans, D.L. & Stein, P.E. (1991) Nature (London) **353**, 576-578.

Carroll, M.C., Campbell, R.D., Bentley, P.R. & Porter, R.R. (1984) Nature (London) **307**, 237-241.

Carter, P.E., Dunbar, B. & Fothergill, J.E. (1983) Biochem. J. **215**, 565-571.

Catterall, C.F., Lyons, A., Sim, R.B., Day, A.J. & Harris, T.J.R. (1987) Biochem. J. **242**, 849-856.

Chakravarti, D.N., Chakravarti, B., Parra, C.A., & Müller-Eberhard, H.J. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 2799-2803.

Chandra, T., Stackhouse, R., Kidd, V.J., Robson, K.J.A. & Woo, S.L.C. (1983) Biochemistry **22**, 5056-5061.

Chang, N.-S., Boackle, R.J. & Leu, R.W. (1991) Immunology **73**, 95-101.

Chazin, W.J., Hugli, T.E. & Wright, P.E. (1988) Biochemistry **27**, 9139-9148.

Chesne, S., Villiers, C.L., Arlaud, G.J., Lacroix, M.B. & Colomb, M.G. (1982) Biochem. J. **201**, 61-70.

Choi N-H, Mazda, T. & Tomita, M. (1989) Mol. Immunol. **26**, 835-840.

Chothia, C. (1975) Nature (London) **254**, 304-308.

Chou, P.Y. & Fasman, G.D. (1978) Adv. Enzymol. **47**, 45-148.

Chu, M.-L., Zhang, R.-Z., Pan, T., Stokes, D., Conway, D., Kuo, H.-J., Glanville, R., Mayer, U., Mann, K., Deutzmann, R. & Timpl, R. (1990) EMBO J. **9**, 385-393.

Ciccodicola, A., Dono, R., Obici, S., Simeone, A., Zollo, M., & Persico, M.G. (1989) EMBO J. **8**, 1987-1991.

Cohen, A.B., Geczy, D. & James, H.L. (1978) Biochemistry **17**, 392-400.

Colomb, M.G., Arlaud, G.J. & Chesne, S. (1989) Behring Inst. Mitt. **84**, 121-128.

Comp, P.C., Nixon, R.R., Cooper, M.R. & Esmon, C.T. (1984) J. Clin. Invest. **74**, 2082-2088.

Conrad, H., Mayer, A., Thomas, H.P. & Vogel, H. (1969) J. Mol. Biol. **41**, 225-229.

Cooper, N.R. (1985) Adv. Immunol. **37**, 151-216.

Cooke, R.M., Wilkinson, A.J., Baron, M., Pastore, A., Tappin, M.J., Campbell, I.D., Gregory, H., & Sheard, B. (1987) Nature (London) **327**, 339-341.

Cooke, R.M., Tappin, M.J., Campbell, I.D., Kohda, D., Miyake, T., Fuwa, T., Miyazawa, T., & Inagaki, F. (1990) Eur. J. Biochem. **193**, 807-815

Cool, D.E., Edgell, C.J.S., Louie, G.V., Zoller, M.J., Brayer, G.D., & MacGillivray, R.T.A. (1985) J. Biol. Chem. **260**, 1366-13676.

Corbi, A.L., Miller, L.J., O'Connor, K., Larson, R.S. & Springer, T.A. (1987) EMBO J. **6**, 4023-4028.

Corbi, A.L., Kishimoto, T.K., Miller, L.J. & Springer, T.A. (1988) J. Biol. Chem.

263, 12403-12411.

Cosman, D., Lyman, S.D., Idzerda, R.L., Beckmann, M.P., Park, L.S., Goodwin, R.G. & March, C.J. (1990) Trends Biochem. Sci. **15**, 1951-1957.

Craik, C.S., Sprang, S., Fletterick, R. & Rutter, W.J. (1982) Nature (London) **299**, 180-182.

Creighton, T.E. (1984) Proteins: Structures and Molecular Properties. W.H. Freeman and Company, New York.

Cusack, S. (1981) J. Mol. Biol. **145**, 539-541.

Daha, M.R., Fearon, D.T. & Austen, K.F. (1976) J. Immunol. **117**, 630-634.

Dahlbäck, B., Smith, C.A. Müller-Eberhard, H.J. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3461-3465.

Dahlbäck, B., Lundwall, A. & Stenflo, J. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 4199-4203.

Dalmasso, A.P., Benson, B.A. & Falk, R.J. (1984) Biochem. Biophys. Res. Commun. **125**, 1013-1019.

Damaschun, G, Müller, J.J. & Bielka, H. (1979) Methods Enzymol. **59**, 706-750.

Damerau, B. (1987) Rev. Physiol. Biochem. Pharmacol. **108**, 151-206.

Dankert, J.R., & Esser, A.F. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 2128-2132.

Dankert, J.R., & Esser, A.F. (1986) Biochemistry **25**, 1094-1100.

Dankert, J.R., & Esser, A.F. (1987) Biochem. J. **244**, 393-399.

Daoudaki, M.E., Becherer, J.D. & Lambris, J.D. (1988) J. Immunol. **140**, 1577-1580.

Davies, A.E. (1983) Immunol. Today **9**, 250-252.

Davis, A.E. (1988) Ann. Rev. Immunol. **6**, 595-628.

Davis, A.E., Whitehead, A.S., Harrison, R.A., Dauphinais, A., Bruns, G.A.P., Cicardi, M. & Rosen, F.S. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 3161-3165.

Davies, A., Simmons, D.L., Hale, G., Harrison, R.A., Tighe, H., Lachmann, P.J. & Waldmann, H. (1989) J. Exp. Med. **170**, 637-654.

Dayhoff, M.O. (1978). In Atlas of Protein Sequence and Structure, Vol. 5, p.36, National Biochemical Research Foundation, Silver Springs, MD.

de Bruijn, M.H.L. & Fey, G.H. (1985) Proc. Natl. Acad. Sci., U.S.A. **82**, 708-712.

Degen, S.J.F., Rajput, B., & Reich, E. (1986) J. Biol. Chem. **261**, 6972-6985.

Degen, S.J.F., Heckel, J.L., Reich, E., & Degen, J.L. (1987) Biochemistry **26**, 8270-8279.

Derynck, R., Roberts, A.B., Winkler, M.E., Chen, E.Y., & Goeddel, D.V. (1984) Cell **38**, 287-297.

DiScipio, R.G. (1981) Biochem. J. **199**, 497-504.

DiScipio, R.G. (1982) Mol. Immunol. **19**, 631-635.

DiScipio, R.G., & Gagnon, J. (1982) Mol. Immunol. **19**, 1425-1431.

DiScipio, R.G., & Hugli, T.E. (1985) J. Biol. Chem. **260**, 14802-14809.

DiScipio, R.G. & Hugli, T.E. (1989) J. Biol. Chem. **264**, 16197-16206.

DiScipio, R.G., Smith, C.A., Müller-Eberhard, H.J. & Hugli, T.E. (1983) J. Biol. Chem. **258**, 10629-10636.

DiScipio, R.G., Gehring, M.R., Podack, E.R., Kan, C.C., Hugli, T.E., & Fey, G.H. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 7298-7302.

DiScipio, R. G., Chakravarti, D. N., Müller-Eberhard, H. J., & Fey, G. H. (1988) J. Biol. Chem. **263**, 549-560.

Dittman, W.A., & Majerus, P.W. (1989) Nucl. Acids Res. **17**, 802-802.

Dodds, A.W., Sim, R.B., Porter, R.R. & Kerr, M.A. (1978) Biochem. J. **175**, 383-390.

Donaldson, V.H. & Evans, R.R. (1963) Am. J. Med. **35**, 37-44.

Doolittle, R.F. (1985) Trends Biochem. Sci. **10**, 233-237.

Doolittle, R.F. (1989) Trends Biochem. Sci. **14**, 244-245.

Dudgeon, T.J., Hommel, U., Cooke, R.M., Fallon, A., & Campbell, I.D. (1991) Techniques in Protein Chemistry II (Villafranca, J. J., Ed.) pp 251-261, Academic Press, San Diego.

Edelhoch, H. (1967) Biochemistry **6**, 1948-1954.

Eisenberg, D. (1984) Annu. Rev. Biochem. **53**, 595-623.

Eisenberg, D., Weiss, R.M. & Terwilliger, T.C. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 140-144.

Enfield, D.L., Ericsson, L.H., Fujikawa, K., Walsh, K.A., Neurath, H., & Titani, K. (1980) Biochemistry **19**, 659-667.

Engelman, D.M. & Moore, P.B. (1975) Ann. Rev. Biophys. Bioeng. **4**, 219-241.

Engh, R., Löbermann, H., Schneider, M., Wiegand, G., Huber, R. & Laurell, C-B. (1989) Prot. Eng. **2**, 407-415.

Engh, R.A., Wright, H.T. & Huber, R. (1990) Prot. Eng. **3**, 469-477.

Entwhistle, R.A. & Furcht, L.T. (1988) Biochemistry **27**, 507-512.

Esser, A.F. (1991) Immunol. Today **12**, 316-318.

Esser, A.F., Kolb, W.P., Podack, E.R., & Müller-Eberhard, H.J. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 1410-1414.

Farries, T.C. & Atkinson, J.P. (1989) J. Immunol. **142**, 842-847.

Farries, T.C., Finch, J.T., Lachmann, P.J. & Harrison, R.A. (1987) Biochem. J. **243**, 507-517.

Farries, T.C., Lachmann, P.J. & Harrison, R.A. (1988a) Biochem. J. **252**, 47-54.

Farries, T.C., Lachmann, P.J., & Harrison, R.A. (1988b) Biochem. J. **253**, 667-675.

Farries, T.C., Steuer, K.L. & Atkinson, J.P. (1990) Immunol. Today **11**, 78-80.

Fasman, G.D. (1989) Trends Biochem. Sci. **14**, 295-299.

Fasman, G.D. & Gilbert, W.A. (1990) Trends Biochem. Sci. **15**, 89-92.

Fearon, D.T. (1978) Proc. Natl. Acad. Sci. U.S.A. **75**, 1971-1975.

Fearon, D.T. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 5867-5871.

Fearon, D.T. & Austen, K.F. (1975) J. Exp. Med. **142**, 856-863.

Fearon, D.T. & Austen, K.F. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 1683-1687.

Fearon, D.T. & Wong, W.W. (1983) Ann Rev Immunol 1, 243-271.

Fesik, S.W., Gampe, R.T., Zuiderweg, E.R.P., Kohlbrenner, W.E. & Weigl D. (1989) Biochem. Biophys. Res. Comm. **159**, 842-847.

Fujita, T., Gigli, I. & Nussenzweig, V. (1978) J. Exp. Med. **148**, 1044-1051.

Gadek, J.E., Fells, G.A, Crystal, R.G. (1979) Science **206**, 1315-1316.

Gaither, T.A., Vargas, I., Inada, S. & Frank, M.M. (1987) Immunology **62**, 405-411.

Garcia de la Torre, J. & Bloomfield, V.A. (1977a) Biopolymers **16**, 1747-1761.

Garcia de la Torre, J. & Bloomfield, V.A. (1977b) Biopolymers **16**, 1779-1793.

Garnier, J., & Robson, B. (1989) In "Prediction of Protein Structure and the Principles of Protein Conformation" (Fasman, G. D., Ed.) pp. 417-465, Plenum Press, New York.

Garnier, J., Osguthorpe, D.J. & Robson, B. (1978) J. Mol. Biol. **120**, 97-120.

Gerard, N.P. & Gerard, C. (1991) Nature (London) **349**, 614-617.

Gettins, P. (1989) J. Biol. Chem. **264**, 3781-3785.

Gettins, P. & Harten, B. (1988) Biochemistry **27**, 3634-3639.

Ghebrehiwet, B. (1989) Behring Inst. Mitt. **84**, 204-215.

Ghosh, R.E. (1989) A Computing Guide for Small Angle Scattering Experiments, Institut Laue Langevin Internal Publication 89GH02T.

Gibrat, J.-F., Garnier, J. & Robson, B. (1987) J. Mol. Biol. **198**, 425-443.

Gigli, I., Fujita, T. & Nussenzweig, V. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 6596-6600.

Glatter, O. (1977a) Acta Phys. Austriaca **47**, 83-102.

Glatter, O. (1977b) J. Appl. Crystal. **10**, 415-421.

Glatter, O. (1982) In "Small Angle X-ray Scattering" (Glatter, O. & Kratky, O., eds.) pp 119-196, Academic Press, London.

Goldberger, G., Bruns, G.A.P., Rits, M., Edge, M.D. & Kwiatkowski, D.J. (1987) J. Biol. Chem. **262**, 10065-10071.

Götze, O., Bianco, C. & Cohn, Z.A. (1979) J. Exp. Med. **149**, 372-386.

Goundis, D. & Reid, K.B.M. (1988) Nature (London) **335**, 82-85.

Greer, J. (1985) Science **228**, 1055-1060.

Greer, J. (1986) Enzyme **36**, 150-163.

Greer, J. (1990) Proteins **7**, 317-334.

Gresham, H.D. & Volanakis, J.E. (1986) The Year in Immunol. **2**, 177-186.

Guinier, A. & Fournet, G. (1955) Small Angle Scattering of X-rays, Wiley, New York.

Guzdek, A., Potempa, J., Dubin, A. & Travis, J. (1990) FEBS Lett. **272**, 125-127.

Haefliger, J.-A., Tschopp, J., Nardelli, D., Wahli, W., Kocher, H.P., Tosi, M., & Stanley, K. K. (1987) Biochemistry **26**, 3551-3556.

Haefliger, J.-A., Tschopp, J., Vial, N., & Jenne, D.E. (1989) J. Biol. Chem. **264**, 18041-18051.

Hänsch, G.M., Schönermark, S. & Roelcke, D. (1987) J. Clin. Invest. **80**, 7-12.

Hanson, D.C., Siegel, R.C. & Schumaker, V.N. (1985) J. Biol. Chem. **260**, 3576-3583.

Haris, P.I., Chapman, D., Harrison, R.A., Smith, K.F. & Perkins, S.J. (1990) Biochemistry **29**, 1377-1380.

Harpel, P.C. (1976) Meths. Enzymol. **45**, 751-760.

Harrison, R.A. (1983) Biochemistry **22**, 5001-5007.

Harrison, R.A., Thomas, M.L. & Tack, B.F. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 7388-7392.

Harrison, R.A. & Lachmann, P.J. (1986) Complement Technology. In Handbook of Experimental Immunology (eds., Weir, D.M., Herzenberg, L.A., Blackwell, C. & Herzenberg, L.A.) pp 39.1-49.

Haupt, H., Heimburger, N., Kranz, T. & Schwick, H.G. (1970) Eur. J. Biochem. **17**, 254-261.

Haviland, D.L., Haviland, J.C., Fleischer, D.T., Hunt, A. & Wetsel, R.A. (1991a) J. Immunol. **146**, 362-368.

Haviland, D.L., Haviland, J.C., Fleischer, D.T. & Wetsel, R.A. (1991b) J. Biol. Chem. **266**, 11818-11825.

Heinz, H.P. (1989) Behring Inst. Mitt. **84**, 20-31.

Hennessy, S.W., Frazier, B.A., Kim, D.D., Deckwerth, T.L., Baumgartel, D.M. Rotwein, P., & Frazier, W. A. (1989) J. Cell Biol. **108**, 729-736.

Hervé, G., Moody, M.F., Tauc, P., Vachette, P. & Jones, P.T. (1985) J. Mol. Biol. **185**, 189-199.

Hervé, M. & Ghélis, C. (1990) Eur. J. Biochem. **191**, 653-658.

Herz, J., Hamann, U., Rogne, S., Myklebost, O., Gausepohl, H. & Stanley, K.K. (1988) EMBO J. **7**, 4119-4127.

Hess, D., Schaller, J., & Rickli, E.E. (1991) Biochemistry **30**, 2827-2833.

Hessing, M. (1991) Biochem. J. **277**, 581-592.

Hillarp, A. & Dahlbäck, B. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 1183-1187.

Hoekzema, R., Martens, M., Brouwer, M.C. & Hack, C.E. (1988) Mol. Immunol. **25**, 485-494.

Hojrup, P., Jensen, M.S., & Petersen, T.E. (1985) FEBS Lett. **184**, 333-338.

Holers, V.M., Cole, J.L., Lublin, D.M., Seya, T. & Atkinson, J.P. (1985) Immunol. Today **6**, 188-192.

Holguin, M.H., Fredrick, L.R., Bernshaw, N.J., Wilcox, L.A. & Parker, C.J. (1989) J. Clin. Invest. **84**, 7-17.

Holladay, L.A., Savage, C.R., Jr., Cohen, S., & Puett, D. (1976) Biochemistry **15**, 2624-2633.

Holt, G.D., Pangburn, M.K., & Ginsburg, V. (1990) J. Biol. Chem. **265**, 2852-2855.

Hopp, T.P. (1986) J. Immunol. Methods **88**, 1-18.

Hopp, T.P. & Woods, K.R. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 3824-3828.

Horiuchi, T., Macon, K.J., Engler, J.A. & Volanakis, J.E. (1991) J. Immunol. **147**,

584-589.

Hoskins, J., Norman, D.K., Beckmann, R.J., & Long, G.L. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 349-353.

Hourcade, D., Holers, V.M. & Atkinson, J.P. (1989) Adv. Immunol. **45**, 381-416.

Howard, O.M.Z., Rao, A.G., & Sodetz, J.M. (1987) Biochemistry **26**, 3565-3570.

Huber, R. & Carrell, R.W. (1989) Biochemistry **28**, 8951-8966.

Huber, R., Scholze, H., Pâques, E.P., & Deisenhofer, J. (1980) Hoppe-Seyler's Z. Physiol. Chem. **361**, 1389-1399.

Huey, R. & Hugli, T.E. (1985) J. Immunol. **135**, 2063-2068.

Ibel, K., & Stuhrmann, H.B. (1975) J. Mol. Biol. **93**, 225-266.

Ikai, A., Nishigai, M., Saito, A., Sinohara, H., Muto, Y. & Arata, Y. (1990) FEBS Lett. **260**, 291-293.

Isenman, D.E., Podack, E.R. & Cooper, N.R. (1980) J. Immunol. **124**, 326-331.

Isenman, D.E., Kells, D.I.C., Cooper, N.R., Müller-Eberhard, H.J. & Pangburn, M.K. (1981) Biochemistry **20**, 4458-4467.

Ishida, B., Wisnieski, B.J., Lavine, C.H., & Esser, A.F. (1982) J. Biol. Chem. **257**, 10551-10553.

Ishikawa, H., Shinkai, Y.-I., Yagita, H., Yue, C.C., Henkart, P.A., Sawada, S., Young, H.A., Reynolds, C.W. & Okumura, K. (1989) J. Immunol. **143**, 3069-3073.

Ito, S. & Tamura, N. (1983) Immunology **50**, 631-635.

Jackman, R.W., Beeler, D.L., Vandewater, L., & Ronberg, R.D. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 8834-8838.

Jacrot, B. (1976) Rep. Prog. Phys. **39**, 911-953.

Jacrot, B. (1982) Structural Molecular Biology. Plenum Publishing Corporation. pp 63-85.

Jacrot, B. (1987a) Crystallography in Molecular Biology. Plenum Publishing Corporation.

Jacot, B. (1987b) Animal Virus Structure. Elsevier Science Publishers B.V. (Biomedical Division).

Jacrot, B. & Zaccai, G. (1981) Biopolymers **20**, 2413-2426.

Jeppsson, J.O., Laurell, C.B. & Fagerhol, M.K. (1978) Eur. J. Biochem. **83**, 143-153.

Johnson, D.A. & Travis, J. (1977) Biochem. J. **163**, 639-641.

Johnson, D.A. & Travis, J. (1978) J. Biol. Chem. **253**, 7142-7144.

Johnson, D.M.A., Gagnon, J. & Reid, K.B.M. (1984) FEBS Lett. **166**, 347-351.

Johnson, R.J. & Chenweth, D.E. (1985) J. Biol. Chem. **260**, 7161-7164.

Kabsch, W. & Sander, C. (1983) Biopolymers **27**, 2577-2637.

Kalter, E.S., Daha, M.R., ten Cate, J.W., Verhoef, J. & Bouma, B.N. (1985) J. Infect. Dis. **151**, 1019-1027.

Katayama, K., Ericsson, L.H., Enfield, D.L., Walsh, K.A., Neurath, H., Davie, E.W., & Titani, K. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4990-4994.

Kaul, R.K., Hildebrand, B., Roberts, S., & Jagadeeswaran, P. (1986) Gene **41**, 311-314.

Ke, H., Lipscomb, W.N., Cho, Y. & Honzatko, R.B. (1988) J. Mol. Biol. **204**, 725-747.

Kim, P.S. & Baldwin, R.L. (1982) Annu. Rev. Biochem. **51**, 459-489.

Kirszbaum, L., Sharpe, J.A., Murphy, B., d'Apice, A.J.F., Classon, B., Hudson, P. & Walker, I.D. (1989) EMBO J. **8**, 711-718.

Kishimoto, T.K., Larson, R.S., Corbi, A.L., Dustin, M.l., Staunton, D.E. & Springer, T.A. (1989) Adv. Immunol. **46**, 149-182.

Kiss, I., Deák, F., Holloway, R.G., Delius, H., Mebust, K.A., Frimberger, E., Argraves, W.S., Tsonis, P.A., Winterbottom, N. & Goetinck, P.F. (1989) J. Biol. Chem. **264**, 8126-8134.

Klickstein, L.B., Wong, W.W., Smith, J.A., Weis, J.H., Wilson, J.G. & Fearon, D.T. (1987) J. Exp. Med. **165**, 1095-1112.

Klickstein, L.B., Bartow, T.J., Miletic, V., Rabson, L.D., Smith, J.A. & Fearon, D.T. (1988) J. Exp. Med. **168**, 1699-1717.

Kline, T.P., Brown, F.K., Brown, S.C., Jeffs, P.W., Kopple, K.D., & Mueller, L. (1990) Biochemistry **29**, 7805-7813.

Kluin-Nelemans, H.C., van Velzen-Blad, H., van Helden, H.P.T. & Daha, M.R. (1984) Cli. Exp. Immunol. 1984, 58, 724-730.

Knäuper, V., Triebel, S., Reinke, H. & Tschesche, H. (1991) FEBS Lett. **290**, 99-102.

Knobel, H.R., Villiger, W. & Isliker, H. (1975) Eur. J. Immunol. **5**, 78-81.

Kohda, D., Go, N., Hayashi, K., & Inagaki, F. (1988) J. Biochem. **103**, 741-743.

Koloczek, H., Waśniowska, Potempa, J. & Wasylewski, Z. (1991) Biochim. Biophys. Acta **1073**, 619-625.

Kontermann, R., & Rauterberg, E.W. (1989) Mol. Immunol. **26**, 1125-1132.

Kontermann, R., Deppisch, R. & Rauterberg, E.W. (1990) Eur. J. Immunol. **20**, 623-628.

Kratky, O. (1983) The World of Neglected Dimensions, Small-Angle Scattering of X-rays and Neutrons of Biological Macromolecules.

Krause, K.L., Volz, K.W. & Lipscomb, W.N. (1987) J. Mol. Biol. **193**, 527-553.

Kress, L.F. & Laskowski, M. (1973) Prep. Biochem. **3**, 541-552.

Krigbaum, W.R. & Godwin, R.W. (1968) Biochemistry **7**, 3126-3131.

Krych, M., Hourcade, D. & Atkinson, J.P. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 4353-4357.

Kurachi, K., Chandra, T., Friezner Deger, S.J., White, T.T., Marchioro, T.L., Woo, S.L.C. & Davie, E.W. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 6826-6830.

Kusumoto, H., Hirosawa, S., Salier, J.P., Hagen, F.S. & Kurachi, K. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 7307-7311.

Kusunoki, Y., Takekoshi, Y. & Nagasawa, S. (1990) J. Pharmacobio-Dyn. **13**, 454-460.

Kyte, J. & Doolittle, R.F. (1982) J. Mol. Biol. **157**, 105-132.

Lachmann, P.J. (1990) Biochem. Soc. Trans. **18**, 1143-1145.

Lachmann, P.J. (1991) Immunol. Today **12**, 312-315.

Laine, R.O. & Esser, A.F. (1989) Nature (London) **341**, 63-65.

Laine, R.O., Morgan, B.P. & Esser, A.F. (1988) Biochemistry **27**, 5308-5314.

Lambris, J.D. (1988) Immunol. Today **9**, 387-393.

Lambris, J.D., Alsenz, J., Schulz, T.F., & Dierich, M.P. (1984) Biochem. J. **217**, 323-326.

Laurell, C.-B. Pierce, J., Persson, U. & Thulin, E. (1975) Eur. J. Biochem. **57**, 107-113.

Law, S.K., Lichtenberg, N.A., Holcombe, F.H. & Levine, R.P. (1980) J. Immunol. **125**, 634-639.

Lawler, J. & Hynes, R.O. (1986) J. Cell Biol. **263**, 5495-5498.

Lear, J.D., Wasserman, Z.R. & DeGrado, W.F. (1988) Science **240**, 1177-1181.

Lee, D.C., Rose, T.M., Webb, N.R., & Todaro, G.J. (1985) Nature (London) **313**,

489-491.

Lee, L.Y., Mohler, W.A., Schafter, B. ., Freudenberger, J.S., Byrne-Connolly, N., Mosley, S.T., Enger, K.B., Leighton, J.R., Davids, R.A., & Tanaka, R.D. (1989) Nucl. Acids Res. **17**, 1259-1260.

Lennick, M., Brew, S.A. & Ingham, K.C. (1985) Biochemistry **25**, 2561-2568.

Leytus S.P., Kurachi K., Sakariassen K.S., & Davie E.W. (1986a) Biochemistry **25**, 4855-4863.

Leytus S.P., Foster, D.C., Kurachi K., & Davie E.W. (1986b) Biochemistry **25**, 5098-5102.

Lichtenheld, M.G. & Podack, E.R. (1989) J. Immunol. **143**, 4267-4274.

Lim, V.I. (1974) J. Mol. Biol. **88**, 857-872.

Liszewski, M.K., Post, T.W. & Atkinson, J.P. (1991) Annu. Rev. Immunol. **9**, 431-455.

Löbermann, H., Lottspeich, F., Bode, W. & Huber, R. (1982) Hoppe-Seyler's Z. Physiol. Chem. **363**, 1377-1388.

Löbermann, D., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984) J. Mol. Biol. **177**, 531-556.

Lohner, K. & Esser, A.F. (1991) Biochemistry **30**, 6620-6625.

Long, G.L., Balagaje, R.M. & MacGillivray, R.T.A. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 5653-5656.

Lopez-Trascasa, M., Bing, D.H., Rivard, M. & Nicholson-Weller, A. (1989) J. Biol. Chem. **264**, 16214-16221.

Lowrey, D.M., Aebischer, T., Olsen, K., Lichtenheld, M., Rupp, F., Hengartner, H., & Podack, E.R. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 247-251.

Lublin, D.M. & Atkinson, J.P. (1989) Ann. Rev. Immunol. **7**, 35-58.

Lublin, D.M., Liszewski, M.K., Post, T.W., Arce, M.A., Le Beau, M.M., Rebentisch, M.B., Lemons, R.S., Seya, T. & Atkinson, J.P. (1988) J. Exp. Med. **168**, 181-194.

Lundwall, A., Wetsel, R.A., Domdey, H., Tack, B.F. & Fey, G.H. (1984) J. Biol. Chem. **259**, 13851-13856.

MacKay, S.L.D. & Dankert, J.R. (1990) J. Immunol. **145**, 3367-3371.

Malhotra, R. & Sim, R.B. (1989) Biochem. J. **262**, 625-631.

Malhotra, R., Thiel, S., Reid, K.B.M. & Sim R.B. (1990) J. Exp. Med. **172**, 955-959.

Maier, K.L., Matejkova, E., Hinze, H., Leuschel, L., Weber, H. & Beck-Speier, I. (1989) **250**, 221-226.

Marazziti, D., Luzio, J.P. & Stanley, K.K. (1989) **243**, 347-350.

Matsusita, M. & Okada, H. (1986) J. Immunol. **136**, 2994-2998.

Mayer, M.M. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 2954-2958.

Mayo, K.H., Cavalli, R.C., Peters, A.R., Boelens, R., & Kaptein, R. (1989) Biochem. J. **257**, 197-205.

McDonald, R.C., Steiz, T.A. & Engelman, D.M. (1979) Biochemistry **18**, 338-342.

McMullen, B.A., Fujikawa, K., & Kisiel, W. (1983) Biochem. Biophys. Res. Commun. **115**, 8-14.

McKinnon C.M., Carter P.E., Smyth S.J., Dunbar B., & Fothergill J.E. (1987) Eur. J. Biochem. **169**, 547-553.

Medicus, R.G., Götze, O. & Müller-Eberhard, H.J. (1976) J. Exp. Med. **144**, 1076-1093.

Medof, M.E., Lublin, D.M., Holers, V.M., Ayers, D.J., Getty, R.R., Leykam, J.F., Atkinson, J.P. & Tykocinski, M.L. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 2007-2011.

Mega, T., Lujan, E. & Yoshida, A. (1980) J. Biol. Chem. **255**, 4053-4056.

Micklem, K.J. & Sim, R.B. (1985) Biochem. J. **231**, 233-236.

Minta, J.O. (1981) J. Immunol. **126**, 245-249.

Minta, J.O., & Kunar, E.S. (1976) J. Immunol. **116**, 1099-1104.

Minta, J.O., & Lepow, I.H. (1974) Immunochemistry **11**, 361-368.

Minta, J.O., Goodkofsky, I., & Lepow, I.H. (1973) Immunochemistry **10**, 341-350.

Misumi, Y., Sohda, M. & Ikehara, Y. (1990) Nucl. Acids Res. **18**, 2178-2178.

Mole, J.E., Anderson, J.K., Davison, E.A. & Woods, D.E. (1984) J. Biol. Chem. **259**, 3407-3412.

Mollnes, T.E. & Lachmann, P.J. (1988) Scan. J. Immunol. **27**, 127-142.

Monahan, J.B. & Sodetz, J.M. (1981) J. Biol. Chem. **256**, 3258-3262.

Monahan, J.B., Stewart, J.L., & Sodetz, J.M. (1983) J. Biol. Chem. **258**, 5056-5062.

Montelione, G.T., Wüthrich, K., Nice, E.C., Burgess, A.W., & Scheraga, H.A. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 5226-5230.

Moody, M.F., Vachette, P. & Foote, A.M. (1979) J. Mol. Biol. **133**, 517-532.

Moore, P.B. (1985) Phys. Today **38**, 63-72.

Moore, P.B. (1988) J. Appl. Cryst. **21**, 675-680.

Morgan, B.P. & Walport, M.J. (1991) **12**, 301-306.

Morgan, B.P., Luzio, J.P. & Campbell, A.K. (1986) Cell Calcium **7**, 399-411.

Mourey, L., Samama, J.P., Delarue, M., Choay, J., Lormeau, J.C., Petitou, M. & Moras, D. (1990) Biochimie **72**, 599-608.

Muto, Y., Fukumoto, Y. & Arata, Y. (1985) Biochemistry **24**, 6659-6665.

Muto, Y., Fukumoto, Y. & Arata, Y. (1987) J. Biochem. **102**, 635-641.

Müller, K. & Glatter, O. (1982) Makromol. Chem. **183**, 465-479.

Müller-Eberhard, H.J., Dalmasso, A.P. & Calcott, M.A. (1966) J. Exp. Med. **123**, 33-54.

Müller-Eberhard, H.J., Polley, M.J. & Calcott, M.A. (1967) J. Exp. Med. **125**, 359-380.

Nagai, M., Hiramatsu, R., Kaneda, T., Hayasuke, N., Arimura, H., Nishida, M., & Suyama, T. (1985) Gene **36**, 183-188.

Nagamine, Y., Pearson, D., Altus, M.S., & Reich, E. (1984) Nucl. Acids Res. 12, 9525-9541.

Narayana, S.V.L., Kilpatrick, J.M., El-Kabbani, O., Babu, Y.S., Bugg, C.E., Volanakis, J.E. & DeLucas, L.J. (1991a) J. Mol. Biol. **219**, 1-3.

Narayana, S.V.L., Carson, M., DeLucas, L., Moore, D., Bugg, C.F., Kilpatrick, J.M. & Volanakis, J.E. (1991b) Complement Inflamm. **8**, 198 (Abstr.).

Nave, C., Helliwell, J.R., Moore, P.R., Thompson, A.W., Worgan, J.S., Greenall, R.J., Miller, A., Burley, S.K., Bradshaw, J., Pigram, W.J., Fuller, W., Siddons, D.P., Deutsch, M. & Tregear, R.T. (1985) J. Appl. Crystallogr. **18**, 396-403.

Nettsheim, D.G., Edalji, R.P., Mollison, K.W., Greer, J. & Zuiderweg, E.R.P. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 5036-5040.

Newcomer, M.E., Lewis, B.A. & Quiocho, F.A. (1981) J. Biol. Chem. **256**, 13218-13222.

Nicholson-Weller, A., Burge, J., Fearon, D.T., Weller, P.F. & Austen, K.F. (1982) J. Immunol. **129**, 184-189.

Nilsson, T. & Wiman, B. (1982) Biochim. Biophys. Acta **705**, 271-276.

Nilsson, U.R., Mandle, R.J. & McConnell-Mapes, J.A. (1975) J. Immunol. **114**, 815-822.

Nolan, K. F., & Reid, K. B. M. (1990) Biochem. Soc. Transact. **18**, 1161-1162.

Nolan, K.F., Schwaeble, W., Kaluz, S., Dierich, M.P. & Reid, K.B.M. (1991) Eur. J. Immunol. **21**, 771-776.

Nonaka, M., Nakayama, K., Yeul, Y. D. & Takahashi, M. (1985) J. Biol. Chem. **260**, 10936-10943.

Norman, D.G., Barlow, P.N., Baron, M., Day, A.J., Sim, R.B. & Campbell, I.D. (1991) J. Mol. Biol. **219**, 717-725.

Nusinow, S.R., Zuraw, B.L. & Curd, J.G. (1985) Med. Clin. North Am. **69**, 487-504.

Ny, T., Elgh, F., & Lund, B. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 5355-5359.

Odermatt, E., Berger, H. & Sano, Y. (1981) FEBS Lett. **131**, 283-285.

O'Hara, P.J., Grant, F.J., Haldeman, B.A., Gray, C.L., Insley, M.Y., Hagen, F.S. & Murray, M.J. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**,5158-5162.

Öhlin, A-K., Linse, S. & Stenflo, J. (1988) J. Biol.Chem. **263**, 7411-7417.

Ojcius, D.M., Jiang, S. & Young, J. D-E. (1990) Immunol. Today **11**, 47-49.

Okada, H., Nagami, Y., Takahashi, K., Okada, N., Hideshima, T., Takizawa, H. & Kondo, J. (1989) Biochem. Biophys. Res. Comm. **162**, 1553-1559.

Orren, A. Potter, P.C., Cooper, R.C. & du Toit, E. (1987) Immunology **62**, 249-253.

Österberg, R., Eggertsen, G., Lundwall, A. & Sjoquist, J. (1984) Int. J. Biol. Macromol. **6**, 195-198.

Österberg, R., Malmensten, B., Nilsson, U., Eggertsen, G. & Kjems, J. (1988) Int. J. Biol. Macromol. **10**, 15-20.

Owen, M.C. & Carrell, R.W. (1977) FEBS Lett. **79**, 245-247.

Owen, M.C., Carrell, R.W. & Brennan, S.O. (1976) Biochim. Biophys. Acta **453**, 257-261.

Pangburn, M. K. (1989) J. Immunol. **142**, 202-207.

Pangburn, M.K. & Müller-Eberhard, H.J. (1978) PNAS **75**, 2416-2420.

Pangburn, M.K. & Müller-Eberhard, H.J. (1980) J. Exp. Med. **152**, 1102-1114.

Pangburn, M.K. & Müller-Eberhard, H.J. (1983) Biochemistry **22**, 178-185.

Pangburn, M.K., Morrison, D.C., Schreiber, R.D. & Müller-Eberhard, H.J. (1980) J. Immunol. **124**, 977-782.

Pangburn, M.K., Schreiber, R.D. & Müller-Eberhard, H.J. (1981) J. Exp. Med. **154**, 856-862.

Pâques, E.P. (1980) Hoppe-Seyler's Z. Physiol. Chem. **361**, 445-456.

Parker, J.M.R., Guo, D. & Hodges, R.S. (1986) Biochemistry **25**, 5425-5432.

Patthy, L. (1987) FEBS Lett. **214**, 1-7.

Patthy, L. (1988) J. Mol. Biol. **202**, 689-696.

Peitsch, M.C., Aminguet, P., Guy, R., Brunner, J., Maizel, J.V. & Tschopp, J. (1990) Mol. Immunol. **27**, 589-602.

Pemberton, P.A., Harrison, R.A., Lachmann, R.J. & Carrell, R.W. (1989) Biochem. J. **258**, 193-198.

Pennica, D., Kohr, W.J., Kuang W.-J., Glaister, D., Aggarwal, B.B., Chen, E.Y. & Goeddel, D. V. (1987) Science **236**, 83-87.

Pensky, J., Levy, L.R. & Lepow, I.H. (1961) J. Biol. Chem. **236**, 1674-1679.

Pensky, J., Hinz, C.F., Todd, E.W., Wedgwood, R.J., Boyer, J.T., & Lepow, I.H. (1968) J. Immunol. **100**, 142-158.

Perkins, S.J. (1982) Biol. Magn. Reson. **4**, 193-336.

Perkins, S.J. (1985) Biochem. J. **228**, 13-26

Perkins, S.J. (1986) Eur. J. Biochem. **157**, 169-180.

Perkins, S.J. (1988) in New Comprehensive Biochemistry (Neuberger, A. & Van Deenen, L.L.M., (eds) Vol. 18B Part II, pp 143 - 264.

Perkins, S.J. (1989a) Behring Institute Mitt. **84**, 129-141.

Perkins, S.J. (1989b) In "Dynamic Properties of Biomolecular Assemblies" (Harding, S.E. & Rowe, A.J., eds.), pp. 226-245, Royal Society of Chemistry, London.

Perkins, S.J. & Nealis, A.S. (1989) Biochem. J. **263**, 463-469.

Perkins, S.J. & Sim, R.B. (1986) Eur. J. Biochem. **157**, 155-168.

Perkins, S.J. & Weiss, H. (1983) J. Mol. Biol. **168**, 847-866.

Perkins, S.J., Kerckaert, J.P. & Loucheux-Lefebvre, M.H. (1985) Eur. J. Biochem. **147**, 525-531.

Perkins, S.J., Chung, L.P. & Reid, K.B.M. (1986) Biochem. J. **233**, 799-807.

Perkins, S.J., Haris, P.I., Sim, R.B., Chapman, D. (1988) Biochemistry **27**, 4004-4012.

Perkins, S.J., Nealis, A.S., Haris, P.I., Chapman, D., Goundis, D., & Reid, K.B.M. (1989) Biochemistry **28**, 7176-7182.

Perkins, S.J., Nealis, A.S. & Sim, R.B. (1990a) Biochemistry **29**, 1167-1175.

Perkins, S.J., Smith, K.F., Nealis, A.S., Lachmann, P.J. & Harrison, R.A. (1990b) **29**, 1175-1180.

Perkins, S.J., Smith, K.F. & Nealis, A.S. (1990c) Biochem. Soc. Transact. **18**, 1151-1154.

Perkins, S.J., Smith, K.F., Amatayakul, S., Ashford, D., Rademacher, T.W., Dwek, R.A., Lachmann, P.J. & Harrison, R.A. (1990d) J. Mol. Biol. **214**, 751-763.

Perkins, S.J., Nealis, A.S. & Sim, R. B. (1991a) Biochemistry **30**, 2847-2857.

Perkins, S.J., Nealis, A.S., Sutton, B.J. & Feinstein, A. (1991b) J. Mol. Biol. **221**, 1345-1366.

Pigott, R., Needham, L.A., Edwards, R.M., Walker, C. & Power, C. (1991) J. Immunol. **147**, 130-135.

Pilz, I. (1982) In "Small Angle X-ray Scattering" (Glatter, O. & Kratky, O., eds.) pp 239-293, Academic Press, London.

Plummer, T.H. & Hurwitz, M.Y. (1978) J. Biol. Chem. **253**, 3907-3916.

Plutzky, J., Hoskins, J.A., Long, G.L., & Crabtree, G.R. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 546-550.

Podack, E.R. & Tschopp, J. (1982a) Proc. Natl. Acad. Sci. U.S.A. **79**, 574-578.

Podack, E.R. & Tschopp, J. (1982b) J. Biol. Chem. **257**, 15204-15212.

Podack, E.R. & Tschopp, J. (1984) Mol. Immunol. **21**, 589-603.

Podack, E.R., Kolb, W.P., & Müller-Eberhard, H.J. (1976) J. Immunol. **116**, 263-269.

Podack, E.R, Kolb, W.P. & Müller-Eberhard, H.J (1978) J. Immunol. **120**, 1841-1848.

Podack, E.R., Tschopp, J. & Müller-Eberhard, H.J. (1982) J. Exp. Med. **156**, 268-282.

Poon, P.H., Schmaker, V.N., Phillips, M.L. & Strang C.J. (1983) J. Mol. Biol. **168**, 563-577.

Prandini, M-H., Reboul, A. & Colomb, M.G. (1986) Biochem. J. **237**, 93-98.

Prater, C.A., Plotkin, J., Jaye, D. & Frazier, W.A. (1991) J. Cell Biol. **112**, 1031-1040.

Preissner, K.T., Podack, E.R. & Müller-Eberhard, H.J. (1985) J. Immunol. **135**, 445-451.

Ramm, L.E., Whitlow, M.B. & Mayer, M.M. (1985) J. Immunol. **134**, 2594-2599.

Rao, A.G., Howard, O.M.Z., Ng, S.C., Whitehead, A.S., Colten, H.R., & Sodetz, J.M. (1987) Biochemistry **26**, 3556-3564.

Raychowdhury, R., Niles, J.L., McCluskey, R.T., & Smith, J.A. (1989) Science **244**, 1163-1165.

Read, R.J. & James, M.N.G. (1988) J. Mol. Biol. **200**, 523-551.

Reboul, A., Arlaud, G.J., Sim, R.B. & Colomb, M.G. (1977) FEBS Lett. **79**, 45-50.

Reboul, A., Prandini, M.-H. & Colomb, M.G. (1987) Biochem. J. **244**, 117-121.

Rees, D.J.G., Jones, I.M., Handford, P.A., Walter, S.J., Esnouf, M.P., Smith, K.J., & Brownlee, G.G. (1988) EMBO J. **7**, 2053-2061.

Reid, K.B.M. (1981) Methods Enzymol. **80**, 143-150.

Reid, K.B.M. & Day, A.J. (1989) Immunology Today **10**, 177-180.

Reid, K.B.M. & Porter, R.R. (1976) Biochem. J. **155**, 19-23.

Reid, K.B.M., Bentley, D.R., Campbell, R.D., Chung, L.P., Sim, R.B., Kristensen, T. & Tack, B.F. (1986) Immunol. Today **7**, 230-234.

Rennie, A.R. (1988) A Users Guide to D17.

Richardson, J.S. (1981) Adv. Protein Chem. **34**, 167-339.

Ripoche, J., Day, A.J., Harris, T.J.R. & Sim, R.B. (1988) Biochem. J. **249**, 593-602.

Robson, B. & Garnier, J. (1986) Introduction to Proteins and Protein Engineering. Elsevier, Amsterdam.

Robson, B. & Pain, R.H. (1974) Biochem. J. **141**, 883-897.

Robson, B. & Suzuki, E. (1976) J. Mol. Biol. **107**, 327-356.

Robson, K.B.H., Hall, J.R.S., Jennings, M.W., Harris, T.J.R., Marsh, K., Newbold, C.I., Tate, V.E., & Weatherall, D.J. (1988) Nature (London) **335**, 79-82.

Rol'bin, Y.A., Kayushina, R.L., Feigin, L.A. & Shchedrin, B.M. (1974) Sov. Phys. Crystallogr. **18**, 442-444.

Rollins, S.A. & Sims, P.J. (1990) J. Immunol. **144**, 3478-3483.

Rollins, S.A. & Springer, M.S. (1985) J. Biol. Chem. **260**, 7157-7160.

Rosen, F.S., Alper, C.A., Pensky, J., Klemperer, M.R. & Donaldson, V.H. (1971) J. Clin. Invest. **50**, 2143-2149.

Ross, G.D. (1989) Curr. Opin. Immunol. **2**, 50-62.

Rossman, M.G. & Argos, P. (1981) Annu. Rev. Biochem. **50**, 497-533.

Ryan, B.F., Joiner, B.L. & Ryan, T.A. (1985) Minitab Handbook, 2nd ed., PWS-

Kent Publishing Co., Boston.

Sali, A., Overington, J.P., Johnson, M.S. & Blundell, T.L. (1990) Trends Biochem. Sci. **15**, 235-240.

Salvesen, G.S., Catanese, J.J., Kress, L.F. & Travis, J. (1985) J. Biol. Chem. **260**, 2432-2436.

Sauer, H., Pratsch, L., Fritzsch, G., Bhakd, S. & Peters, R. (1991) Biochem. J. **276**, 395-399.

Schechter, Y., Burstein, Y. & Gertler, A. (1977) Biochemistry **16**, 992-997.

Schindler, R., Gelfand, J.A. & Dinarello, C.A. (1990) Blood **76**, 1631-1638.

Schlesinger, M., Nave, Z., levy, Y., Slater, P.E. & Fishelson, Z. (1990) Clin. Exp. Immunol. **81**, 423-427.

Schneider, R., Mayer, A., Schmatz, W., Kaiser, B. & Scherm, R. (1969) J. Mol. Biol. **41**, 231-235.

Schoenborn, B.P. & Nunes, A.C. (1972) Ann. Rev. Biophys. Bioeng. **1**, 529-552.

Schönermark, S., Rauterberg, E.W., Shin, M.L., Löke, S., Roelcke, D. & Hänsch, G.M. (1986) J. Immunol. **136**, 1772-1776.

Schönermark, S., Filsinger, S., Berger, B. & Hänsch, R.H. (1988), Immunology **63**, 585-590.

Schulze, A.J., Baumann, U., Knof, S., Jaeger, E., Huber, R. & Laurell, C-B. (1990) Eur. J. Biochem. **194**, 51-56.

Schultze, H.E., Heide, K. & Haupt, H. (1962) Naturwissenschaften **6**, 133-134.

Schumaker, V.N., Hanson, D.C., Kilchherr, E., Phillips, M.L. & Poon, P.H. (1986) Mol. Immunol. **23**, 557-565.

Schumaker, V.N., Zavodszky, P. & Poon, P.H. (1987) Ann. Rev. Immunol. **5**, 21-42.

Scott, J. (1989) Nature (London) **338**, 118-119.

Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W.J. & Bell, G. I. (1983) Science **221**, 236-240.

Selander, M., Persson, E., Stenflo, J., & Drakenberg, T. (1990) Biochemistry, **29**, 8111-8118.

Sepich, D.S., Noonan, D.J. & Ogata, R.T. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 5895-5899.

Seya, T., Turner, J.R. & Atkinson, J.P. (1986) J. Exp. Med. **163**, 837-855.

Seya, T., Okada, M., Matsumoto, M., Hong, K., Kinoshita, T. & Atkinson, J.P.

(1991) Mol. Immunol. **28**, 1137-1147.

Shamash, Y. & Rimon, A. (1966) Biochim. Biophys. Acta **121**, 35-41.

Shinkai, Y., Takio, K., & Okumura, K. (1988) Nature (London) **334**, 525-527.

Shiver, J.W., Dankert, J.R., Donovan, J.J., & Esser, A.F. (1986) J. Biol. Chem. **261**, 9629-9636.

Siciliano, S.J., Rollins, T.E. & Springer, M.S. (1990) J. Biol. Chem. **265**, 19568-19574.

Sim, R.B. & Reboul, A. (1981) Meths. Enzymol. **80**, 43-54.

Sim, R.B. & Reid, K.B.M. (1991) Immunol. Today **12**, 307-311.

Sim, R.B., Malhotra, V., Day, A.J. & Erdei, A. (1987) Imunol. Lett. **14**, 183-190.

Simpson, R.J., Smith, J.A., Moritz, R.L., O'Hare, M.J., Rudland, P.S., Morrison, J.R., Lloyd, C.J., Grego, B., Burgess, A.W., & Nice, E.C. (1985) Eur. J. Biochem. **153**, 629-637.

Smith, C.A., Vogel, C.W. & Müller-Eberhard, H.J. (1982) J. Biol. Chem. **257**, 9879-9882.

Smith, C.A., Vogel, C.W. & Müller-Eberhard, H.J. (1984a) J. Exp. Med. **159**, 324-329.

Smith, C.A., Pangburn, M.K., Vogel, C.-W., & Müller-Eberhard, H.J. (1984b) J. Biol. Chem. **259**, 4582-4588.

Smith, K.F., Harrison, R.A. & Perkins, S.J. (1990) Biochem. J. **267**, 203-212.

Smith, K.F., Nolan, K.F., Reid, K.B.M., & Perkins, S.J. (1991) Biochemistry **30**, 8000-8008.

Smith, K.F., Harrison, R.A., & Perkins, S.J. (1992) Biochemistry in press.

Skriver, K., Wikoff, W.R., Patson, P.A., Tausk, F., Schapira, M., Kaplan, A.P. & Bock, S.C. (1991) J. Biol. Chem. **266**, 9216-9221.

Stanley, K.K., & Herz, J. (1987) EMBO J. **6**, 1951-1957.

Stanley, K.K., Kocher, H.P., Luzio, J.P., Jackson, P., & Tschopp, J. (1985) EMBO J. **4**, 375-382.

Stanley, K.K., Page, M., Campbell, A.K., & Luzio, J.P. (1986) Mol. Immunol. **23**, 451-458.

Stein, P.E. & Chothia, C. (1991) J. Mol. Biol. **221**, 615-621.

Stein, P.E., Leslie, A.G.W., Finch, J.T., Turnell, W.G., McLaughlin, P.J. & Carrell, R.W. (1990) Nature (London) **347**, 99-102.

Stewart, J.L., Monahan, J.B., Brickner, A. & Sodetz, J.M. (1984) Biochemistry **23**, 4016-4022.

Stewart, J.L., Kolb, W.P. & Sodetz, J.M. (1987) J. Immunol. **139**, 1960-1964.

Strecker, G., Ollier-Hartmann, M.P., van Halbeek, H., Friederik, J., Vliegenhart, G., Montreuil, J. & Hartmann, L. (1985) C.R. Acad. Sc. Paris **301**, 571-576.

Stuhrmann, H.B., & Fuess, H. (1976) Acta Cryst. **A32**, 67-74.

Südhof, T.C., Goldstein, J.L., Brown, M.S., & Russell, D.W. (1985) Science **228**, 815-822.

Sugita, Y., Nakano, Y. & Tomita, M. (1988) J. Biochem. **104**, 633-637.

Surewicz, W.K. & Mantsch, H.H. (1988) Biochim. Biophys. Acta **952**, 115-130.

Suzuki, K., Kusumoto, H., Deyashiki, Y., Nishioka, J., Maruyama, I., Zushi, M., Kawahara, S., Honda, G., Yamamoto, S., & Horiguchi, S. (1987) EMBO J. **6**, 1891-1897.

Suzuki, Y., Yoshida, K., Honda, E. & Sinohara, H. (1991) J. Biol. Chem. **266**, 928-932.

Takada, Y. & Hemler, M.E. (1989) J. Cell Biol. **109**, 397-407.

Taylor, W.R. (1986) J. Mol. Biol. **188**, 233-258.

Taylor, W.R. (1990) Methods Enzymol. **183**, 456-474.

Thielens, N.M., Villiers, C.L., Villiers, M-B. & Colomb, M.G. (1984) FEBS Lett. **165**, 111-116.

Thielens, N.M., Dorsselaer, A.V., Gagnon, J. & Arlaud, G.J. (1990) Biochemistry **29**, 3570-3578.

Thompson, R.A. & Winterborn, M.H. (1981) Clin. Exp. Immunol. **46**, 110-119.

Tomana, M., Niemann, M., Garner, C. & Volanakis, J.E. (1985) Mol. Immunol. **22**, 107-111.

Tosi, M., Duponchel, C., Bourgarel, P., Colomb, M.G. & Meo, T. (1986) Gene **42**, 265-272.

Tosi, M., Duponchel, C., Meo, T. & Julier, C. (1987) Biochemistry **26**, 8516-8524.

Towns-Andrews, E., Berry, A., Bordas, J., Mant, G.R., Murray, P.K., Roberts, K., Sumner, I., Worgan, J.S., Lewis, R. & Gabriel, A. (1989) Rev. Sci. Instrum. **60**, 2346-2349.

Trapani, J.A., Kwon, B.S., Kozak, C.A., Chintamaneni, C., Young, J.D., & Dupont,

B. (1990) J. Exp. Med. **171**, 545-557.

Traut, T.W. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2944-2948.

Travis, J. & Salvesen, G.S. (1983) Ann. Rev. Biochem. **52**, 655-709.

Travis, J., Guzdek, A., Potempa, J. & Watorek, W. (1990) Biol. Chem. Hoppe-Seyler **371**, Suppl. 3-11.

Tschopp, J. (1984a) J. Biol. Chem. **259**, 7857-7863.

Tschopp, J. (1984b) J. Biol. Chem. **259**, 10569-10573.

Tschopp, J., & Jongeneel, C.V. (1988) Biochemistry **27**, 2641-2646.

Tschopp, J. & Masson, D. (1987) Mol. Immunol. **24**, 907-913.

Tschopp, J., & Nabholz, M. (1990) Annu. Rev. Immunol. **8**, 279-302.

Tschopp, J., Müller-Eberhard, H.J., & Podack, E.R. (1982) Nature (London) **298**, 534-538.

Tschopp, J., Engel, A., & Podack, E. R. (1984) J. Biol. Chem. **259**, 1922-1928.

Tysoe-Calnon, V.A., Grundy, J., & Perkins, S.J. (1991) Biochem. J. **277**, 359-369.

Van Der Meer, B.W., Fugate, R.D. & Sims, P.J. (1989) Biophys. J. **56**, 935-946.

van Regenmortel, M.H.V. & Daney de Marcillac, G. (1988) Immunol. Lett. **17**, 95-108.

Vaughan, L., Lorier, M.A. & Carrell, R.W. (1982) Biochim. Biophys. Acta **701**, 339-345.

Vaughan, L. & Carrell, R.W. (1981) Biochem. Int. **2**, 461-467.

Vergé, D., Tardieu, A. & Arrio-Dupont, M. (1983) FEBS Lett. **154**, 277-281.

Vik, D.P., Amiguet, P., Moffat, G.J., Fey, M., Amiguet-Barras, F., Wetsel, R.A. & Tack, B.F. (1991) Biochemistry **30**, 1080-1085.

Villiers, C.L., Arlaud, G.J. & Colomb, M.G. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4477-4481.

Virca, G.D., Lyerly, D., Kreger, A. & Travis, J. (1982) Biochim. Biophys. Acta **704**, 267-271.

Vogt, W., Damerau, B., von Zabern, I., Nolte, R. & Brunahl, D. (1989) Mol. Immunol. **26**, 1133-1142.

Walport, M.J. & Lachmann, P.J. (1988) Arthritis Rheum. **31**, 153-158.

Weis, J.J., Toothaker, L.E., Smith, J.A., Weis, J.H. & Fearon, D.T. (1988) J. Exp. Med. **167**, 1047-1066.

Weiss, V., Fauser, C. & Engel, J. (1986) J. Mol. Biol. **189**, 573-581.

Wetlaufer, D.B. (1962) Adv. Protein. Chem. **17**, 303-390.

Wetlaufer, D.B. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 697-701.

Wetlaufer, D.B. (1981) Adv. Protein Chem. **34**, 61-92.

Wetsel, R.A., Lundwall, A., Davidson, F., Gibson, T., Tack, B.F. & Fey, G.H. (1984) J. Biol. Chem. **259**, 13857-13862.

Wetsel, R.A., Ogata, R.T. & Tack, B.F. (1987) Biochemistry **26**, 737-743.

Wetsel, R.A., Lemons, R.S., Le Beau, M.M., Barnum, S.R., Noack, D. & Tack, B.F. (1988) Biochemistry **27**, 1474-1482.

Whitehead, A.S., Soloman, E., Chambers, S., Bodmer, W.F., Povey, S. & Fey, G. (1982) Proc. Natl. Acad. Sci., U.S.A. **79**, 5021-5025.

Wilmot, C.M., & Thornton, J.M. (1988) J. Mol. Biol. **203**, 221-232.

Wolf, F.W., Eddy, R.L., Shows, T.B., & Dixit, V.M. (1990) Genomics **6**, 685-691.

Worcester, D.L. (1988) J. Appl. Cryst. **21**, 669-674.

Wright, H.T., Qian, H.X. & Huber, R. (1990) J. Mol. Biol. **213**, 513-528.

Wright, P.E. (1989) Trends Biochem. Sci. **14**, 255-260.

Yamada, K.M. (1989) Curr. Opin. Cell Biol. **1**, 956-963.

Yamamoto, T., Davies, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L., & Russell, D.W. (1984) Cell **39**, 27-38.

Yamamoto, T., Bishop, R.W., Brown, M.S., Goldstein, J.L., & Russell, D.W. (1986) Science **232**, 1230-1237.

Yoshitake, S., Schach, B.G., Foster, D.C., Davie, E.W. & Kurachi, K., (1985) Biochemistry **24**, 3736-3750.

Young, J. D-E., Young, T.M. (1990) Mol. Immunol. **27**, 1001-1007.

Young, J. D-E., Cohn, Z. A., & Podack, E. R. (1986) Science **233**, 184-190.

Yu, C.Y. (1991) J. Immunol. **146**, 1057-1066.

Zaccai, G. (1978) Topics Curr. Phys. **6**, 243-270.

Zaccai, G. & Jacrot, B. (1983) Ann. Rev. Biophys. Bioeng. **12**, 139-157.

Zalman, S. & Müller-Eberhard, H.J. (1985) Fed. Proc. **44**, 551.

Zalman, S. & Müller-Eberhard, H.J. (1990) Mol. Immunol. **27**, 533-537.

Zalman, S., Wood, L.M. & Müller-Eberhard, H.J. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 6975-6979.

Zarbock, J., Gennaro, R., Romeo, D., Clore, G.M. & Gronenborn, A.M. (1988) FEBS Lett. **238**, 289-294.

Zhu, X-J. & Chan, S.K. (1987) Biochem. J. **246**, 19-23.

Zhu, X-J., Kang, S.S., Hargrove, K., Shochat, D., Jarrells, M., Mojesky, M. & Chan, S.K. (1987) Biochem. J. **246**, 25-36.

Ziccardi, R.J. (1976) J. Immunol. **116**, 504-509.

Ziccardi, R.J. (1983) Springer Semin. Immunopathol. **6**, 213-230.

Ziccardi, R.J. & Cooper, N.R. (1979) J. Immunol. **123**, 788-792.

Zuiderweg, E.R.P., Henkin, J., Mollison, K.W., Carter, G.W. & Greer, J. (1988a) Proteins: Structure, Function and Genetics **3**, 139-145.

Zuiderweg, E.R.P., Mollison, K.W., Henkin, J. & Carter, G.W. (1988b) Biochemistry **27**, 3568-3580.

Zuiderweg, E.R.P., Nettesheim, D.G., Mollison, K.W. & Carter, G.W. (1989) Biochemistry **28**, 172-185.

Structural comparisons of the native and reactive-centre-cleaved forms of α_1-antitrypsin by neutron- and X-ray-scattering in solution

Kathryn F. SMITH,*† Richard A. HARRISON‡ and Stephen J. PERKINS*†§

*Department of Biochemistry and Chemistry and †Department of Protein and Molecular Biology, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K., and ‡Molecular Immunopathology Unit, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH, U.K.

α_1-Antitrypsin is the best-characterized member of the serpin (serine-proteinase inhibitor) superfamily. Its solution structure was studied by high-flux neutron-scattering and synchrotron X-ray-scattering. Neutron data show that its absorption coefficient $A^{1\%}_{280,1cm}$ is 5.4. The neutron radius of gyration R_G at infinite contrast for native α_1-antitrypsin is 2.61 nm, characteristic of a moderately elongated structure, and its cross-sectional R_G is 1.34 nm. The internal inhomogeneity of scattering densities within α_1-antitrypsin is high at 29×10^{-5}. The X-ray R_G is 2.91 nm, in good agreement with the neutron R_G of 2.82 nm in ^1H_2O. This R_G is unchanged in reactive-centre-cleaved α_1-antitrypsin. These parameters are also unchanged at pH 8 in sodium/potassium phosphate buffers up to 0.6 M. The neutron and X-ray curves for native α_1-antitrypsin were compared with Debye simulations based on the crystal structure of reactive-centre-cleaved (papain) α_1-antitrypsin. After allowance for residues not visible in the crystallographic electron-density map, and rejoining the proteolysed site between Met-358 and Ser-359 by means of a relatively minor conformational re-arrangement, good agreement to a structural resolution of 4 nm is obtained with the neutron data in two contrasts and with the X-ray data. The structures of the native and cleaved forms of α_1-antitrypsin are thus similar within the resolution of solution scattering. This places an upper limit on the magnitude of the presumed conformational changes that occur in α_1-antitrypsin on reactive-centre cleavage, as indicated in earlier spectroscopic investigations of the Met-358–Ser-359 peptide-bond cleavage. Methods for scattering-curve simulations from crystal structures are critically assessed. The R_G data lead to dimensions of 7.8 nm × 4.9 nm × 2.2 nm for native α_1-antitrypsin. The high internal inhomogeneity and the asymmetric shorter semi-axes of 4.9 nm and 2.2 nm suggest that the three oligosaccharide chains of α_1-antitrypsin are essentially freely extended into solvent in physiological conditions. This conclusion is also supported by the Debye simulations, and by modelling based on hydrodynamic parameters.

INTRODUCTION

The human serine-proteinase inhibitors of plasma show a remarkable degree of sequence homology and have been grouped together in a superfamily termed serpins (serine-proteinase inhibitor). The most abundant member of this superfamily in plasma is α_1-antitrypsin (α_1-AT) (M_r 51500), sometimes termed α_1-proteinase inhibitor (Carrell *et al.*, 1982). Other members include antithrombin III (Carrell *et al.*, 1981), ovalbumin (Kurachi *et al.*, 1981), α_1-antichymotrypsin (Chandra *et al.*, 1983; Bao *et al.*, 1987), β-plasminogen activator (Ny *et al.*, 1986), α_2-antiplasmin (Lijnen *et al.*, 1987) and the C1 inhibitor component of complement (Bock *et al.*, 1986). α_1-AT exhibits its highest affinity towards neutrophil elastase (Travis & Salvesen, 1983). In a healthy individual, α_1-AT is present at a plasma concentration of 1.3 mg/ml (Jeppsson *et al.*, 1978), but this substantially increases under inflammatory conditions, and is sometimes much decreased in disease. It acts by presenting a reactive site as an ideal substrate to the proteolytic enzyme, and a stable stoichiometric 1:1 complex is formed with the enzyme to inhibit it. This reactive site is the Met-358–Ser-359 peptide bond, which is situated 37 residues from the *C*-terminus of α_1-AT (Carrell *et al.*, 1980; Johnson & Travis, 1978). Under certain conditions *in vivo*, a proteolytically modified form of α_1-AT, in which the reactive site is cleaved, is produced by abortive interaction with many proteinases. This 'split' α_1-AT is no longer able to inhibit proteinases. Inhibition by proteolytic cleavage may be important for the regulation of the inhibitor molecule (Carrell *et al.*, 1982; Carrell & Owen, 1985).

Split α_1-AT is a more stable form of the protein and has a known crystal structure (Löbermann *et al.*, 1982, 1984). Its secondary structure contains three β-sheets and eight α-helices. Since the newly formed *C*- and *N*-termini at Met-358 and Ser-359 lie at opposite ends of the molecule and are separated by 7 nm, a significant conformational change must have occurred on cleavage. Using c.d. and fluorescence, Bruch *et al.* (1988) confirmed that the secondary structures of the native and split forms of four serpins (included α_1-AT) were significantly different. However, no experimental data are available on whether the conformational transition is a localized or a more global re-arrangement of the structure.

Neutron-scattering and synchrotron X-ray-scattering in solution are techniques that are able to provide structural information under physiological conditions to nominal resolutions of 2–4 nm (Perkins, 1988*a*,*b*). In principle, therefore, any large conformational change between the native and split forms of α_1-AT should be detectable by both techniques. In addition, they enable comparisons of the physiological structure with those observed in higher phosphate concentrations, including 2.5 M, that from which split α_1-AT was crystallized. The scattering curves determined experimentally under physiological conditions can also be compared quantitatively with scattering curves

Abbreviation used: α_1-AT, α_1-antitrypsin.

§ To whom correspondence and requests for reprints should be addressed, at the Department of Biochemistry and Chemistry.

calculated from the crystal co-ordinates of split α_1-AT. This permits both a careful comparison of the two structures and a critical assessment of the methods used for these calculations. In particular, the comparison allows the average conformation of the three oligosaccharide chains in α_1-AT relative to the protein surface to be deduced. This is not possible by using other physical measurements of solution structures.

MATERIALS AND METHODS

Purification of α_1-AT

α_1-AT purification was based on the thiol–disulphide interchange procedure described by Laurell *et al.* (1975). Briefly, in a typical preparation, 500 ml of fresh frozen plasma was thawed and made 25 mM in 2-mercaptoethanol. An equal volume of saturated $(NH_4)_2SO_4$ was then added with stirring, and the resulting precipitate was removed by centrifugation and discarded. A further addition of $(NH_4)_2SO_4$ was made (160 g/l; final saturation 72%), and the precipitate was collected by centrifugation. This second precipitate was redissolved in 0.1 M-Tris/HCl buffer, pH 8.1, containing 5 mM-EDTA and 5 mM-NaN_3, then dialysed exhaustively against the same buffer to remove any residual 2-mercaptoethanol. The dialysed material was loaded on to a GSH–Sepharose column [170 ml bed volume, previously activated by charging with 5,5′-dithiobis-(2-nitrobenzoate), and washed and equilibrated with 0.1 M-Tris/HCl buffer, pH 8.1, containing 5 mM-EDTA and 5 mM-NaN_3], and the column was washed with the same buffer, then buffer containing 1.0 M-NaCl. (The 1.0 M-NaCl wash was necessary to remove a proteinase that otherwise co-purified with α_1-AT and inactivated the purified inhibitor by specific cleavage within the reactive-centre loop.) α_1-AT was eluted from the column with 0.1 M-Tris/HCl buffer, pH 8.1, containing 0.5 M-NaCl, 5 mM-EDTA, 5 mM-NaN_3, 1 mg of 5,5′-dithiobis-(2-nitrobenzoate)/ml and 0.2 mg of dithiothreitol/ml and dialysed, initially against 0.1 M-Tris/HCl buffer, pH 8.1, containing 50 mM-NaCl, 5 mM-benzamidine and 10 mM-2-mercaptoethanol [to effect the total removal of the protein-bound derivative of 5,5′-dithiobis-(2-nitrobenzoate), 3-carboxy-4-nitrobenzenethiol], then against the same buffer in the absence of 2-mercaptoethanol. The α_1-AT pool was loaded on to a DEAE-Sephacel column (230 ml bed volume) equilibrated in the above final dialysis buffer, and, after a washing with 1 column volume of equilibration buffer, α_1-AT was eluted with a gradient (8 column volumes) to equilibration buffer containing 0.4 M-NaCl. α_1-AT-containing fractions were again pooled, concentrated to 20–25 ml and applied to a Sephadex G-150 (superfine grade) column (2 litres bed volume) equilibrated with 20 mM-sodium/potassium phosphate buffer, pH 7.0, containing 0.5 M-NaCl and 10 mM-NaN_3. The peak fractions only of α_1-AT were pooled, treated with 5 mM-2-mercaptoethanol then 10 mM-iodoacetamide to block the single thiol group of α_1-AT and hence the potential for dimerization, dialysed to remove excess reactants and concentrated to approx. 20 mg/ml. Throughout the procedure α_1-AT was detected by single-dimension crossed immunoelectrophoresis (Laurell, 1966) and its purity was assessed by SDS/PAGE (on 10% gels; buffer system of Laemmli, 1970). The final product gave a single band only.

α_1-AT was either held at 4 °C (for imminent use) or stored frozen at −70 °C until required. Samples that had been frozen were prepared for scattering studies by repeating the gel filtration on Sephadex G-150 to remove any aggregates that had formed. Only the peak fraction(s) were retained. These were pooled and reconcentrated to approx. 20 mg/ml. Immediately before scattering analysis all samples were filtered through a 0.2 μm-pore filter (Gelman) into sterile glass or plastic vials, then held at 4 °C until dialysis against the appropriate scattering buffer.

Reactive-centre-cleaved α_1-AT was generated by papain cleavage (Johnson & Travis, 1977). Although this is reported to react abortively at the reactive-centre Met-358–Ser-359 peptide bond only, SDS/PAGE and h.p.l.c. analysis of the digestion indicated that at least two proteolytic events occurred during generation of the stable cleaved product. Before use, papain (Worthington Corp.) was activated by incubation in 20 mM-sodium/potassium phosphate buffer, pH 7.0, containing 5 mM-cysteine and 10 mM-EDTA. Under the reaction conditions used, papain/α_1-AT ratio 1:1000 (w/w) at 37 °C in 20 mM-sodium/potassium phosphate buffer, pH 7.0, containing 10 mM-EDTA, 5 mM-cysteine and 5 mM-NaCl, an initial fast cleavage was essentially complete after 10 min. However, at this time a second, slower, cleavage was also evident. As the first cleavage could not be effected in the total absence of the second, digestion was continued for 6 h, at which time both cleavages were complete. Papain digestion was terminated by the addition of iodoacetamide (5 mM). *N*-Terminal sequence analysis (performed by Dr. L. Packman at the Protein Sequencing Facility, Department of Biochemistry, University of Cambridge) shows that, in addition to cleavage between Met-358 and Ser-359 at the reactive site, cleavages at Gly-5–Asp-6 and Asp-12–Thr-13 also occur. Sequence analysis of papain-digested α_1-AT before and after dialysis against scattering buffers demonstrated that, as expected, the *C*-terminal peptide (residues 359–394) was retained after dialysis, but the small *N*-terminal peptides Glu-Asp-Pro-Gln-Gly and Asp-Ala-Ala-Gln-Lys-Thr-Asp were lost. The sequence analysis also showed that native α_1-AT remains intact, as expected.

Collection of neutron data for α_1-AT was initially performed in 12 mM-sodium phosphate buffer, pH 7.0, containing 1 mM-EDTA and 0.2 M-NaCl. X-ray data were measured in a range of concentrations from 50 mM-phosphate up to the crystallization buffer 2.5 M-sodium/potassium phosphate buffer, pH 8.0 (Löbermann *et al.*, 1982). $K_2HPO_4,3H_2O$ was dissolved to 3.0 M in ^1H_2O or ^2H_2O, back-titrated with stirring at room temperature to pH 8.0 by additions of known amounts of $NaH_2PO_4,2H_2O$, and diluted to its final desired phosphate concentration. The α_1-AT samples were dialysed against the appropriate buffers with stirring at 4 °C and with four changes over at least 36 h. Optical absorbances were recorded at 280 nm immediately afterwards, and the scattering data were obtained within the following 24–48 h. SDS/PAGE analysis of samples before and after scattering analysis showed that they had the expected band patterns, and that no radiation-induced damage or non-specific proteolysis had occurred.

Neutron- and X-ray-scattering experiments

Neutron data were obtained on Instrument D17 at the Institut Laue Langevin in Grenoble, France. Guinier data were based on a sample-to-detector distance of 3.44 m and a wavelength λ of 1.095 nm (Q range of 0.016–0.615 nm^{-1}, where $Q = 4\pi \cdot \sin\theta/\lambda$ and 2θ is the scattering angle). Data at large Q were obtained by using a sample-to-detector distance of 1.41 m and λ of 1.095 nm (Q range of 0.045–1.451 nm^{-1}). Samples were measured in rectangular quartz Hellma cells of path thickness 1 mm for 0% ^2H_2O buffers (200 μl) and 2 mm for 70%, 80% and 100% ^2H_2O (400 μl) buffers at 20 °C. Counting times were 3–16 min in ^2H_2O buffers and 1–6 h in ^1H_2O buffers. Data reduction in Grenoble utilized standard software and procedures (RNILS, SPOLLY, RGUIM and RPLOT; Ghosh, 1981; Perkins *et al.*, 1990), and the final analyses in London employed the interactive Guinier and large-angle plotting program SCTPL2 (Perkins *et al.*, 1990).

X-ray data were obtained on the Synchrotron Radiation

Source at Daresbury, Warrington, U.K., operating at 2.00 GeV. The low-angle solution scattering cameras at Station 2.1 (beam currents of 61 mA to 52 mA) and Station 8.2 (261 mA to 75 mA) were used (Towns-Andrews *et al.*, 1989) in a total of four sessions. At Station 2.1, the detector-to-sample distance was 2.52 m, with a wavelength of 0.155 nm (Q range of 0.189–3.206 nm^{-1}). At Station 8.2 the detector-to-sample distances of 2.605 m, 2.608 m and 3.83 m and the wavelength of 0.159 nm resulted in Q ranges of 0.051–1.534 nm^{-1} and 0.039–1.012 nm^{-1} respectively. Fresh wet slightly stretched rat tail collagen (diffraction spacing of 67 nm) was used to calibrate the detector Q axis. Samples were contained in round Perspex [polyl(methyl methacrylate)] cells with a volume of 20 μl, a path thickness of 1 mm and mica window thicknesses close to 10 μm, held in position by Teflon inserts, and measured nominally at 20 °C. Samples were measured for 10–20 min in dilution series, in alternation with buffer in the same sample cell and with equal data collection times to reduce the occurrence of background subtraction errors. Data were measured in ten separate time slices as a check for radiation damage, after which the ten curves were averaged. Data reduction was performed by using OTOKO at Daresbury to subtract the buffer background and correct for the detector response measured with a uniform radioactive source, and final data analyses utilized SCTPL2 conjointly with the neutron data.

Analyses of Guinier data

At small Q, the Guiner equation gives the radius of gyration R_G and the forward scattering at zero scattering angle $I(0)$ (Perkins, 1988*a*,*b*):

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

In a given solute–solvent contrast, R_G measures the degree of elongation of a glycoprotein. If this structure is sufficiently elongated, the R_G of the cross-sectional structure R_{XS} and the cross-sectional intensity at zero angle $[I(Q)\cdot Q]_{Q\to 0}$ are obtained from:

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

The matchpoint is obtained from a graph of $\sqrt{[I(0)/cT_s t]}$ against % ^2H_2O (T_s is the sample transmission; t is the sample thickness), and from the cross-sectional analyses by using $\sqrt{\{[I(Q)\cdot Q]_{Q\to 0}/cT_s t\}}$. Experimental matchpoints can be compared with the amino acid sequence on the basis of the unhydrated shape of α_1-AT, the use of crystallographic volumes (non-hydrated) and the 10% non-exchange of the main-chain amide protons (Perkins, 1986). The dependence of the R_G on the internal scattering densities within α_1-AT is given by the Stuhrmann equation (Ibel & Stuhrmann, 1975):

$$R_G^2 = R_C^2 + \alpha/\Delta\rho - \beta/\Delta\rho^2$$

where $\Delta\rho$ is the difference in scattering density ρ between the solute and solvent. In Fig. 3(*b*), R_C is the R_G at infinite contrast and corresponds to the intercept at zero $\Delta\rho^{-1}$, while the slope α reflects the radial distribution of scattering-density inhomogeneities. The subscripts G and XS correspond to the overall and cross-sectional Guinier analyses. If the outermost scattering densities are higher than those nearer the centre, α will be positive. β corresponds to the distance between the centres of two distinct components of very different scattering densities and is neglible for glycoproteins. All linear-regression analyses of the R_G, $I(0)/c$, matchpoint and Stuhrmann graphs used the OXSTAT and MINITAB statistics programs. The Guinier analyses lead to the triaxial dimensions of α_1-AT. If this resembles an elliptical cylinder, its length $L = \sqrt{[12(R_G^2 - R_{XS}^2)]}$. From the intensity parameters, $L = \pi\cdot I(0)/[I(Q)\cdot Q]_{Q\to 0}$. The two shorter semi-axes A and B are obtained from the total dry (neutrons) or hydrated (X-rays) volume $V = \pi ABL$ and $R_{XS}^2 = (A^2+B^2)/4$. For X-ray data, the volume of the hydration shell is calculated assuming 0.3 g of water/g of glycoprotein and 0.0245 nm^3 per water molecule (Perkins, 1986).

Debye sphere modelling

Small spheres were used to model the scattering curves with the use of the Debye equation (Perkins & Weiss, 1983; Perkins, 1985). The crystal co-ordinates of α_1-AT were subdivided into cubes of side 0.608 nm such that their total volume is the dry volume of the crystallographically visible residues. The curve simulations were based on the overlapping spheres of volume 0.608^3 nm^3 set at the centre of each cube. The sphere diameter is 0.754 nm, which is much less than the highest nominal resolution $2\pi/Q_{max.}$ of the curve, i.e. 4.3 nm for $Q_{max.} = 1.45$ nm^{-1} (Fig. 5). The contrast-dependence was modelled by subdividing the 291 spheres into an outer hydrophilic surface region of 163 spheres and an inner core of 128 spheres on the basis of the volumes of these two components. From compositions, the neutron-scattering densities of the two regions should be assigned as 49.4% ^2H_2O and 31.6% ^2H_2O respectively (Perkins, 1986). It was found more effective to use instead 60% ^2H_2O and 30% ^2H_2O respectively. This model was converted into the hydrated X-ray model by expanding the co-ordinates by 1.10, which was derived from the cube root of the ratio of the hydrated and dry volumes. The cube size is now 0.668 nm. For X-rays, the outermost spheres correspond to the hydration shell as well as the carbohydrate and hydrophilic amino acid residues. Electron densities were taken as 334 e/nm^3 for solvent, 435 e/nm^3 for the outermost 216 spheres of the model, and 389 e/nm^3 for the innermost 75 spheres.

The R_{G-C} value is calculated from:

$$R_G^2 = \sum_{k=1}^{n} (R_k^2 + r_k^2)$$

for n spheres, each with an R_G denoted R_k and located at a distance r_k from the centre of mass of the model. The Stuhrmann α value is calculated from:

$$\alpha_G = n^{-1} \sum_{k=1}^{n} (\rho_k - \rho_m) r_k^2$$

where ρ_k is the scattering density of the sphere and ρ_m is the mean scattering density. The full Debye curve is calculated from Rolbin *et al.* (1973):

$$\frac{I(Q)}{I(0)} = g(Q)\left(n_1\cdot\rho_1^2 + n_2\cdot\rho_2^2 + 2\rho_1^2 \sum_{j=1}^{m} A_j^{11}\frac{\sin Q\cdot r_j}{Q\cdot r_j} + 2\rho_2^2 \sum_{j=1}^{m} A_j^{22}\frac{\sin Q\cdot r_j}{Q\cdot r_j} + 2\rho_1\cdot\rho_2 \sum_{j=1}^{m} A_j^{12}\frac{\sin Q\cdot r_j}{Q\cdot r_j}\right)(n_1\cdot\rho_1 + n_2\cdot\rho_2)^{-2}$$

where the model is constructed from n_1 and n_2 spheres of different scattering densities ρ_1 and ρ_2; $g(Q) = 3(\sin Q\cdot R - Q\cdot R\cdot\cos Q\cdot R)^2/Q^6\cdot R^6$ (the squared-form factor of the spheres of radius R); A_j^{11}, A_j^{22} and A_j^{12} are the number of distances r_j for that increment of j between the spheres 1 and 1, 2 and 2 and 1 and 2 in that order; m is the number of different distances r_j. Before comparison with the experimental neutron-scattering curves, the calculated curve incorporated corrections for the wavelength spread of 10% on Instrument D17 and the beam divergence in radians (Cusack, 1981; Perkins & Weiss, 1983), based on the empirical parameters of 16% and 0.016 radians (Cusack, 1981). Since the X-ray cameras utilize point collimation, the calculated and experimental $I(Q)$ curves are directly comparable without corrections. To facilitate quantitative com-

parisons, a residual R was employed, which is defined by analogy with protein crystallography (Blundell & Johnson, 1976):

$$\mathbf{R} = \Sigma|I(Q)_{\text{expt.}} - I(Q)_{\text{calc.}}|/\Sigma|I(Q)_{\text{expt.}}|$$

Hydrodynamic modelling of α_1-AT

Frictional coefficients f are calculated from:

$$f = \frac{M_r(1-\bar{v}\cdot\rho_{20,w})}{N_a \cdot s^0_{20,w}}$$

where $\bar{v}$ is the partial specific volume, $\rho_{20,w}$ is the density of water at 20 °C, N_a is Avogadro's constant, and $s^0_{20,w}$ is the sedimentation coefficient. Frictional ratios f/f_0 are calculated by using $f_0 = 6\pi\eta r$, in which the macromolecule is assumed to be a hydrated sphere of radius r. The value of $\bar{v}$ is derived from the consensus volumes data set of Perkins (1986), which is based on the sum of the electrostricted water shell volume and the dry volume. Frictional coefficients were calculated by using non-overlapping spheres by the modified Oseen tensor procedure of Bloomfield (Garcia de la Torre & Bloomfield, 1977*a*,*b*; Perkins, 1989). The crystal structure was subdivided into 31 dry cubes of side 1.216 nm such that the overall dimensions are close to those of the co-ordinates. Since relatively few spheres are used, allowance for the void spaces between the non-overlapping spheres is necessary. The effective triaxial lengths of the model should be reduced by $(1-\pi/4)d$ (d is the sphere diameter) before comparison with the crystallographic dimensions. To allow for this, a hypothetical hydration of 0.39 g of water/g of glycoprotein in place of 0.30 g of water/g is used to calculate the total hydrated volume of the 31 spheres (Perkins, 1989), i.e. a hydration factor of 1.12. The 31 'dry' co-ordinates were thus rescaled by a factor of 1.39 to give a final sphere diameter of 1.69 nm (Fig. 6). If the protein surface is uniformly covered by a single layer of water molecules of thickness 0.36 nm (which corresponds to a hydration of 0.3 g of water/g of protein; Perkins, 1986, 1989), good agreement with the dimensions of the crystal co-ordinates can be obtained.

RESULTS AND DISCUSSION

Neutron-scattering of native α_1-AT in physiological buffers

Fig. 1(*a*) shows that, under physiological conditions, neutron-scattering in 100% ^2H_2O gave good linear Guinier plots for native α_1-AT in an acceptable $Q\cdot R_G$ range of 0.3–1.1. Dilution series for α_1-AT in the range $c = 2.4$ to 13.7 mg/ml (Fig. 2) showed that the R_G increased by 20%; sedimentation experiments also gave a similar result (Kress & Laskowski, 1973). The extrapolation to zero c gave an R_G of 2.40 ± 0.03 nm for 12 mM-sodium phosphate buffer, pH 7.0, containing 1 mM-EDTA and 0.2 M-NaCl and 2.37 ± 0.03 nm for 50 mM-sodium/potassium phosphate buffer, pH 8.0. No dependence

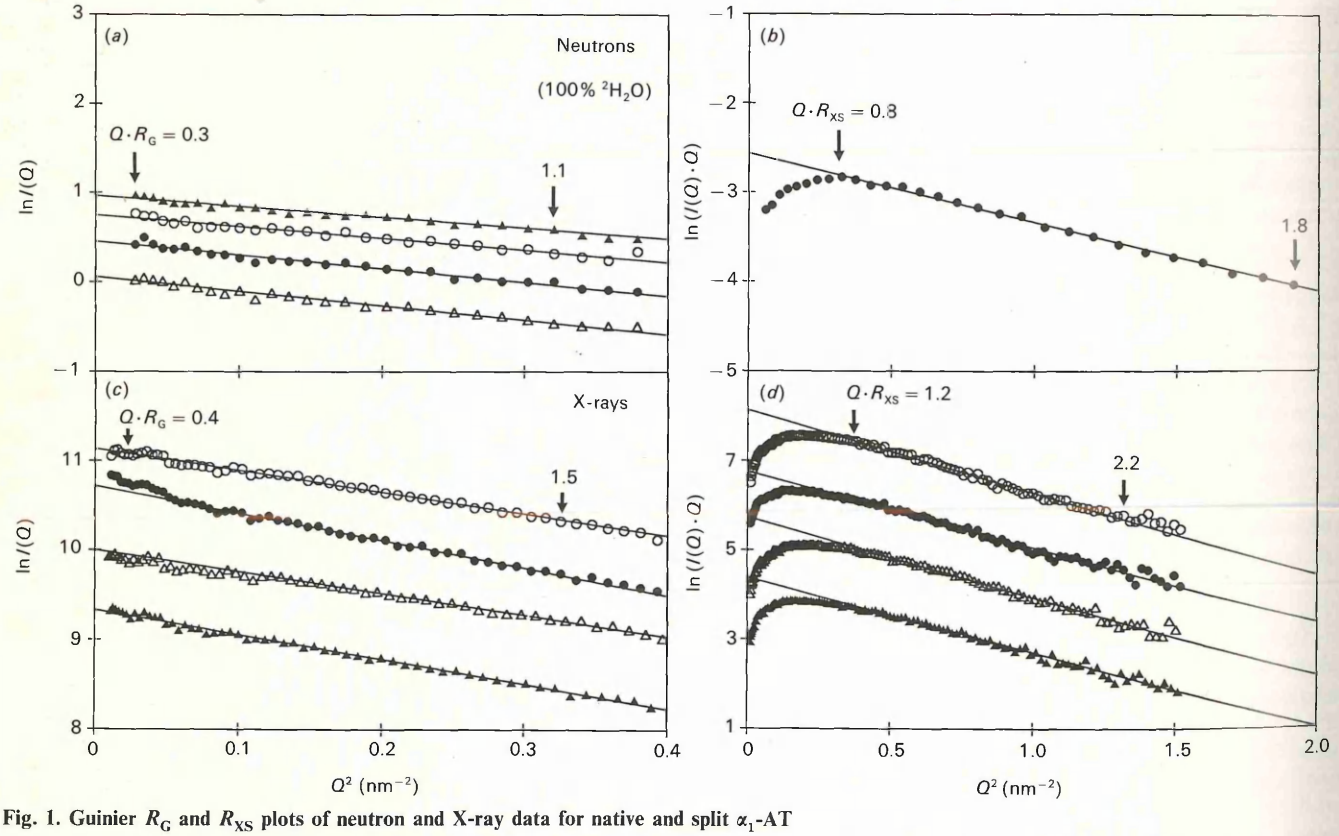


Fig. 1. Guinier R_G and R_{XS} plots of neutron and X-ray data for native and split α_1-AT

The buffer in both cases is 50 mM-sodium/potassium phosphate buffer, pH 8.0. (*a*) Neutron Guinier plots of the dilution series for native α_1-AT in 100% ^2H_2O at concentrations of 12.3 mg/ml (▲), 9.8 mg/ml (○), 7.0 mg/ml (●) and 4.8 mg/ml (△). The Q range employed for fitting is 0.17–0.57 nm^{-1} between the $Q\cdot R_G$ range as arrowed. The slopes increase slightly with increasing dilution; the relative vertical displacements reflect the dilutions. (*b*) Neutron cross-sectional plot of native α_1-AT at 8.0 mg/ml. The Q range used for fitting is 0.61–1.42 nm^{-1}, as indicated by the arrowed $Q\cdot R_{XS}$ range. (*c*) X-ray Guinier plots of native (○ and ●) and split (△ and ▲) α_1-AT at concentrations of 16.8 mg/ml (○), 12.6 mg/ml (●), 18.9 mg/ml (△) and 14.2 mg/ml (▲). The Q range used for fitting the data is 0.15–0.57 nm^{-1} between the $Q\cdot R_G$ range as shown. The curves were displaced arbitrarily for clarity. (*d*) X-ray cross-sectional plot of the samples in (*c*) by X-ray-scattering. The $Q\cdot R_{XS}$ values are based on a Q range of 0.60–1.20 nm^{-1}. Note that the vertical scale is halved in (*d*) relative to (*a*), (*b*) and (*c*).

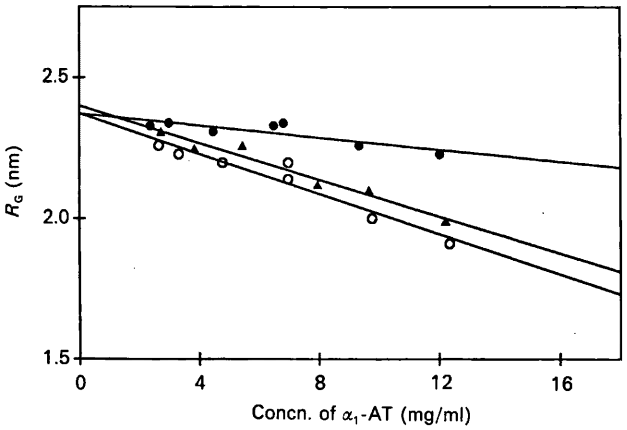


Fig. 2. Concentration-dependence of the R_G of native α_1-AT

The R_G values were measured by neutron-scattering in 100% ^2H_2O in 50 mM-sodium/potassium phosphate buffer, pH 8.0 (○), in 0.50 M-sodium/potassium phosphate buffer, pH 8.0 (●), and in 12 mM-sodium phosphate buffer, pH 7.0, containing 1 mM-EDTA and 0.2 M-NaCl (▲). The R_G at zero α_1-AT concentration was obtained from linear regression with the use of MINITAB.

was observed for the $I(0)/c$ parameter (results not shown). At larger scattering angles, good cross-sectional Guinier plots were also obtained. Fig. 1(*b*) shows that the mean R_{XS} of 1.22 nm was obtained in a satisfactory $Q \cdot R_{XS}$ range of 0.8–1.8.

This concentration-dependence is unusual for plasma glycoproteins. From their analysis of the crystal structure, Löbermann *et al.* (1984) have proposed that α_1-AT probably has a large dipole moment, since the acidic and basic charged groups are positioned asymmetrically on the structure. To test whether this was the source of the concentration-dependence, a further dilution series of α_1-AT was studied in 0.50 M-sodium/potassium phosphate buffer, pH 8.0. As a control, the values of $I(0)/cT_s$ and $[I(Q) \cdot Q]_{Q \to 0}/cT_s$ were calculated and found to be independent of phosphate concentration, as expected. Fig. 2 shows a diminished dependence of the R_G on phosphate concentration, and the extrapolated R_G of 2.39 ± 0.03 nm is within error of those for 12 mM-sodium phosphate buffer, pH 7.0, containing 1 mM-EDTA and 0.2 M-NaCl and 50 mM-sodium/potassium phosphate buffer, pH 8.0. The concentration effect on R_G is thus attributed to the interactions of the dipole moments of different molecules.

Full contrast variation of native α_1-AT was carried out in 0%, 70%, 80% and 100% ^2H_2O. To verify that the neutron data correspond to monomeric α_1-AT, its relative molecular mass M_r was determined from the $I(0)/c$ value in 0% ^2H_2O (Jacrot & Zaccai, 1981), where c was calculated from the absorption coefficient $A^{1\%}_{280,1cm}$ of 5.0–5.4 (Bundy & Mehl, 1959; Kress & Laskowski, 1973; Bruch *et al.*, 1988). By using an M_r of 51500 (calculated from sequence data), $I(0)/c$ gave an A_{280} determination of 5.4 ± 0.2, in good agreement with Bruch *et al.* (1988), and showing also that α_1-AT is monomeric in solution. It may be noted that the molecular composition of α_1-AT leads to too low a predicted A_{280} of 3.86; it is, however, known that deviations from the calculation can occur when the A_{280} is less than 10 (Wetlaufer, 1962; Edelhoch, 1967; Perkins, 1986). The neutron matchpoint was determined for comparison with that predicted from the sequence. In Fig.3(*a*) the matchpoint was determined as 40.3 ± 1.1% ^2H_2O, which agrees well with a prediction of 40.3% ^2H_2O (Table 1). The limited cross-sectional data of $\sqrt{\{[I(Q) \cdot Q]_{Q \to 0}/cT_s t\}}$ (four points in two contrasts; results not shown) gave a comparable matchpoint of 42.1 ± 1.2% ^2H_2O. It is concluded that neutron-scattering 'visualizes' α_1-AT as a non-hydrated monomeric macromolecule.

The contrast-variation R_G data were analysed in the Stuhrmann

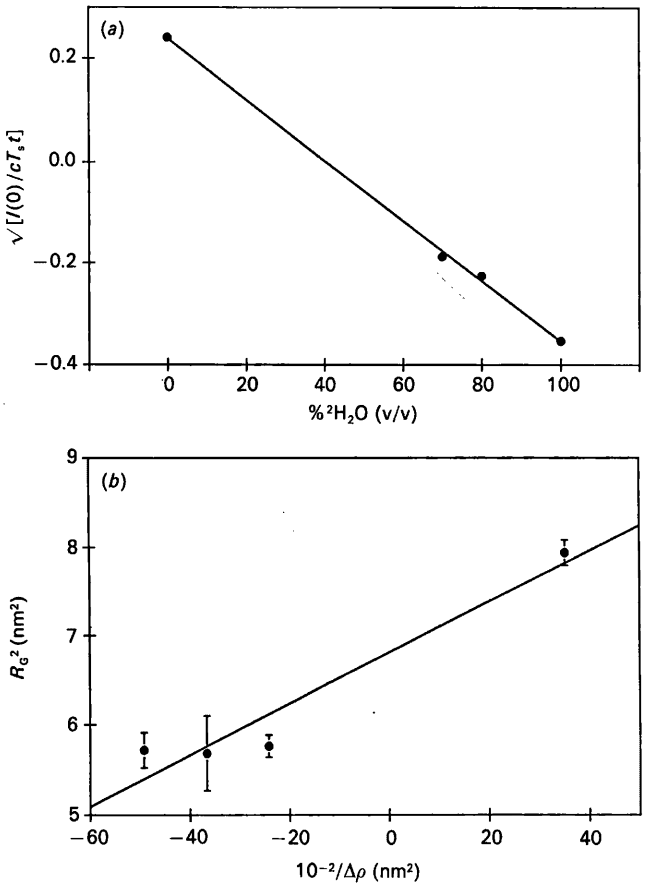


Fig. 3. Neutron contrast-variation analysis of native α_1-AT, with 2.7 to 12.2 mg/ml concentrations in 12 mM-sodium phosphate buffer, pH 7.0, containing 1 mM-EDTA and 0.2 M-NaCl

(*a*) The matchpoint determination was based on extrapolations of the $I(0)/c$ values to zero c (16 points in four contrasts), and calculating $\sqrt{[I(0)/cT_s t]}$ from these (c is the α_1-AT concentration; T_s is the sample transmission; t is the path length). The error in the matchpoint is thus higher than usual. The matchpoint from regression analysis is 40.3 ± 1.1% ^2H_2O. (*b*) The Stuhrmann plot of the R_G^2 against the reciprocal solute–solvent contrast difference $\Delta\rho^{-1}$ is also based on the R_G data extrapolated to zero c (Fig. 2). The regressed line gives an R_{G-C} of 2.61 ± 0.02 nm and the slope gives a Stuhrmann α of $29 \times 10^{-5} \pm 5 \times 10^{-5}$.

plot of Fig. 3(*b*). At infinite contrast R_{G-C} was determined to be 2.61 ± 0.02 nm. The elongation ratio R_{G-C}/R_0 is thus 1.35 (where R_0 is the R_G of the sphere of the same dry volume as α_1-AT), which is close to the average of 1.28 ± 0.10 found for many globular proteins (Perkins, 1988*b*). The corresponding R_{XS} analysis gave an R_{XS-C} of 1.34 ± 0.01 nm. The slopes α were strongly positive, where α_G is $29 \times 10^{-5} \pm 5 \times 10^{-5}$ and α_{XS} is $13 \times 10^{-5} \pm 1 \times 10^{-5}$. It is concluded that the surface regions in α_1-AT have a higher scattering density than those near the core. Since the hydrophilic amino acid and carbohydrate residues have a mean matchpoint of 49.4% ^2H_2O, while hydrophobic amino acid residues have one of 31.6% ^2H_2O, hydrophilic residues are seen to be predominant in the outer regions of the α_1-AT structure, as expected.

Some notable differences are seen on comparison of α_1-AT with α_1-acid glycoprotein (M_r 38000; 44% carbohydrate; Perkins *et al.*, 1985). α_1-Acid glycoprotein has a compact conformation in low-salt buffers with an R_G/R_0 ratio of 1.32 and an α_G of $7 \times 10^{-5} \pm 2 \times 10^{-5}$, whereas its expanded form in high-salt buffers (where the oligosaccharide chains are extended into solution) has an R_G/R_0 ratio of 1.67 and an α_G of $43 \times 10^{-5} \pm 4 \times 10^{-5}$. The α_G of $29 \times 10^{-5} \pm 5 \times 10^{-5}$ for α_1-AT is considerably larger than that for the compact form of α_1-acid

glycoprotein, even when allowance is made for the difference in R_{G-C}, in the precision of measurement of α or in the ratio of hydrophilic to hydrophobic residues (54:46 for α_1-AT and 68:32 for α_1-acid glycoprotein). Since α is proportional to r_k^2 (see the Materials and methods section), these data suggest that the oligosaccharide chains on α_1-AT are extended into solution under physiological conditions.

The Guinier analyses at infinite contrast lead to estimates of the triaxial dimensions of α_1-AT (see under 'Analyses of Guinier data' in the Materials and methods section). The ratio of intensities give a length L of 7.6 ± 0.5 nm (four values), in agreement with an L of 7.8 ± 0.1 nm from the R_G data. α_1-AT is thus determined to be 7.8 nm × 4.9 ± 0.1 nm × 2.2 ± 0.1 nm in size. The reported crystallographic dimensions of 6.7 nm × 3.2 nm × 3.2 nm (Löbermann *et al.*, 1984) should be increased to 7.2 nm × 3.4 nm × 3.4 nm to correspond to the correct total volume of α_1-AT. Even though the shorter axes of the protein component are more similar in the crystal structure, molecular graphics showed that the three oligosaccharide sites of α_1-AT are located in such a way that one short axis will be significantly longer if the oligosaccharide chains are extended into solution. The Guinier data thus support the suggestion of extended oligosaccharide conformations.

X-ray scattering of native and split α_1-AT in high-phosphate buffers

Synchrotron X-ray-scattering was employed to compare (i) the physiological structure with that in high concentrations of phosphate, (ii) the native and split structures of α_1-AT, and (iii) the hydrated structure with the dry one seen by neutron-scattering. Figs. 1(*c*) and 1(*d*) show that excellent R_G and R_{XS} Guinier plots were obtained for both native and split α_1-AT in 50 mM-sodium/potassium phosphate buffer, pH 8.0. The concentration-dependence of the R_G values is visible also in Fig. 1(*c*). Occasionally a slight time-dependent aggregation of α_1-AT was detected by an increase in $I(Q)$ at a very low Q. Its occurrence was facilitated by increased beam currents just after beam refills, increased phosphate concentration, the prior presence of trace aggregates or the use of native α_1-AT. Accordingly the ten time-slices of each run were examined to check that the Guinier parameters were unaffected.

Data collection in sodium/potassium phosphate buffers up to 1.2 M were affected by worsening sample transmissions with increased phosphate concentrations, and satisfactory signal-to-noise ratios could be obtained only up to 0.6 M-sodium/potassium phosphate buffer, pH 8.0 (Fig. 4). Data analyses showed that, unlike the corresponding neutron data in 100% ^2H_2O, the $I(0)/c$ values increase (results not shown). This appears to reflect the different solute–solvent scattering density and radiation in use. This dependence decreased with increase in the phosphate concentration, and the $I(0)/cT_s$ and $[I(Q) \cdot Q]_{Q \to 0}/cT_s$ data after extrapolation to zero c were constant within error in the different phosphate buffers (results not shown), thus confirming the good quality of the Guinier data. The R_G values exhibited a statistical spread comparable in magnitude with the observable R_G-dependence on α_1-AT concentration, which is attributed to fluctuations in the background subtraction. Accordingly the mean R_G from each experiment is reported in Fig. 4. No concentration-dependence is seen for the R_{XS} data, and mean values are again given in Fig. 4.

As the phosphate concentration is increased to 0.6 M, no change is seen in the R_G or R_{XS} values of native α_1-AT, showing that the structure is stable under these conditions. Three different experiments in 50 mM-sodium/potassium phosphate buffer, pH 8.0, gave R_G values of 2.80 ± 0.09 nm, 2.93 ± 0.06 nm and 3.00 ± 0.04 nm. These are similar to but slightly larger than the neutron R_G of 2.82 ± 0.03 nm in water, since both experiments correspond to positive solute–solvent contrasts, but the X-ray data now reflect a hydration shell surrounding α_1-AT. Likewise, two experiments gave 'hydrated' R_{XS} values of 1.73 ± 0.19 nm (19 values) and 1.82 ± 0.13 nm (seven values) that are slightly larger than the neutron 'dry' R_{XS} value of 1.53 ± 0.10 nm measured in ^1H_2O in 12 mM-sodium phosphate buffer, pH 7.0, containing 1 mM-EDTA and 0.2 M-NaCl. The corresponding data for split α_1-AT are also shown in Fig. 4, and no difference in R_G or R_{XS} from the native form is observed. This suggests that the conformational transition between the native and split forms of α_1-AT involves only a localized re-arrangement of the protein structure, and that the loss of the *N*-terminal peptides (see the Materials and methods section) has a negligible effect on the structure.

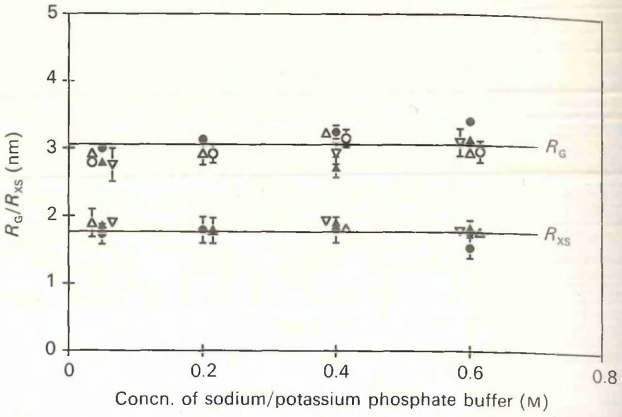


Fig. 4. Dependence of the Guinier parameters on the phosphate concentration

The upper plot shows the R_G-dependence upon sodium/potassium phosphate buffer concentration for the X-ray data (△, ▽ and ○) for native α_1-AT. X-ray data for split α_1-AT (▲) and 67% split 33% native α_1-AT (●) are also shown. Error bars are shown only when significant in the X-ray data and increase at higher phosphate concentrations. The lower plot shows the corresponding R_{XS}-dependence on phosphate concentrations by X-rays, the same symbols being used.

Molecular modelling of α_1-AT

The scattering data were compared directly with the crystal structure of split α_1-AT (Löbermann *et al.*, 1984) by converting the latter into a Debye sphere model for scattering-curve simulations (see the Materials and methods section). Only 374 of the 394 amino acid residues are present in the structure, since Gly-1–Asp-19 and Lys-394 are not visible in the electron-density map. This protein structure was subdivided into the equivalent unhydrated volume of 246 cubes of side 0.608 nm, which facilitates the representation of the oligosaccharide chains as a single cube for each residue (Perkins *et al.*, 1985). Both tetragonal and hexagonal split α_1-AT co-ordinates were taken from the Protein Data Bank (Bernstein *et al.*, 1977). After addition of hydrogen atom co-ordinates (Perkins, 1982), and with the use of a unit weighting of atoms, the R_G of these structures was calculated as 2.12 nm and 2.13 nm respectively by using $R_G^2 = \Sigma_n r_i^2/n$, where r_i is the distance of each atom from the centre of gravity. This agrees with the R_G of 2.15 nm calculated from the sphere model. Since the two co-ordinate sets gave very similar results, only the tetragonal model was used further.

Carbohydrate and the *N*-terminal residues were added to the model. The three *N*-linked oligosaccharides at Asn-46, Asn-83 and Asn-247 were assumed to be biantennary with 11 residues each (Vaughan *et al.*, 1982). Although only partly visible, the

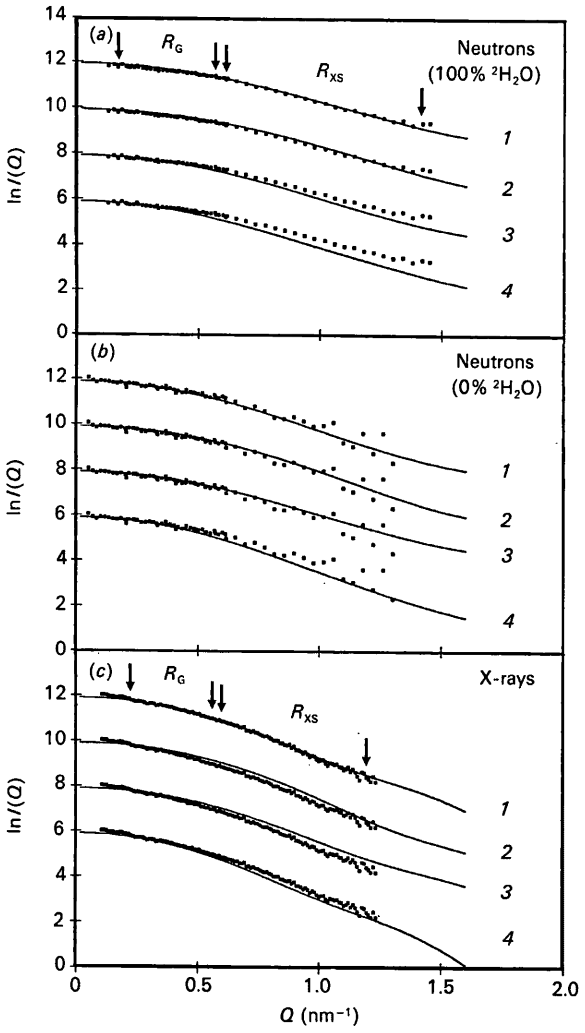


Fig. 5. Comparisons of the curve simulations with the α_1-AT models

(*a*) Neutron data in ^2H_2O in 50 mM-sodium/potassium phosphate buffer, pH 8.0 (high negative solute–solvent contrast difference), for native α_1-AT are compared with the smeared simulated curves. The vertical arrows indicate the Q ranges of 0.17–0.57 nm^{-1} and 0.61–1.42 nm^{-1} used for the R_G and R_{XS} determinations respectively (Figs. 1*a* and 1*b*). Curve *1*: the unhydrated native α_1-AT structure with extended oligosaccharide and the use of two densities; **R** of 0.006 (see the Materials and methods section). Curve *2*: non-hydrated native α_1-AT structure with compact oligosaccharide and two densities; **R** of 0.010. Curve *3*: non-hydrated native α_1-AT with extended oligosaccharide and use of a single density; **R** of 0.020. Curve *4*: native α_1-AT with two densities and a hydration of 0.3 g of water/g of α_1-AT; **R** of 0.024. (*b*) Neutron data in ^1H_2O in 12 mM-sodium phosphate buffer, pH 7.0, containing 1 mM-EDTA and 0.2 M-NaCl (high positive solute–solvent contrast difference) for native α_1-AT are compared with the four simulated curves numbered as in (*a*) but recalculated for this contrast. The **R** values for curves *1*, *2*, *3* and *4* are 0.017, 0.019, 0.017 and 0.029 in that order. (*c*) X-ray data in 50 mM-sodium/potassium phosphate buffer, pH 8.0 (high positive solute–solvent contrast difference), for native α_1-AT are compared with models based on a hydration of 0.3 g of water/g of α_1-AT in curves *1* and *2*. The vertical arrows indicate the Q ranges of 0.33–0.57 nm^{-1} and 0.60–1.20 nm^{-1} used for the R_G and R_{XS} determinations respectively (Figs. 1*c* and 1*d*). Curve *1*: native α_1-AT, hydrated to 0.3 g of water/g of α_1-AT, with extended oligosaccharide and two densities; **R** of 0.011. Curve *2*: native α_1-AT, hydrated to 0.3 g of water/g of α_1-AT, with compact oligosaccharide and two densities; **R** of 0.028. Curve *3*: non-hydrated native α_1-AT with extended oligosaccharide and two densities; **R** of 0.035. Curve *4*: native α_1-AT, hydrated to 0.67 g of water/g of α_1-AT, with extended oligosaccharide and two densities; **R** of 0.028.

Table 1. Compositional and scattering data for α_1-AT

Amino acid residues:	
Total	394
Volume (nm^3)*	57.4
Carbohydrate residues:	
Total	36
Volume (nm^3)*	7.9
Total M_r from sequence	51 500
Percentage carbohydrate by mass	14.1
Overall $\bar{v}$ (ml/g)†	0.732
Overall $\Sigma b/M_r$	
In ^1H_2O (fm)	0.2248
In ^2H_2O (fm)	0.3813
Predicted $A^{1\%}_{280,1cm}$‡	3.86
Literature A_{280}	5.0–5.4
Neutron A_{280} determination	5.4±0.2
Predicted neutron matchpoint (% ^2H_2O)§	40.3
Experimental matchpoint (% ^2H_2O)	
From $I(0)$	40.3±1.1
From $[I(Q)\cdot Q]_{Q\to 0}$	42.1±1.2
Stuhrmann analysis	
R_{G-C} (nm)	2.61±0.02
α_G	$29\times10^{-5}\pm5\times10^{-5}$
R_{XS-C} (nm)	1.34±0.01
α_{XS}	$13\times10^{-5}\pm1\times10^{-5}$

* The dry volume is taken from Chothia (1975) (amino acids) and Perkins (1986) (carbohydrates) and used in matchpoint and modelling calculations. The carbohydrate composition is calculated assuming two biantennary and one triantennary oligosaccharides.

† Partial specific volumes $\bar{v}$ were calculated from the consensus volume set of Perkins (1986) and are close to those determined by the Cohn & Edsall (1943) compilation. These values were used in sedimentation simulations.

‡ Calculated from the content of tyrosine, cysteine and tryptophan residues by the Wetlaufer (1962) procedure and rescaled by using a factor of 1.03 (Perkins, 1986). Literature determinations of A_{280} are reported in Bundy & Mehl (1959), Kress & Laskowski (1973) and Bruch *et al.* (1988).

§ Calculated by using the assumption of the non-exchange of 10% of the main-chain peptide hydrogen atoms (Perkins, 1986).

crystal structure shows that all three are seen in an extended configuration (Fig. 6*a*). Each oligosaccharide was thus modelled by 11 cubes in an extended arrangement (Fig. 6*b*). It is possible that the oligosaccharide chains may be folded against the protein surface. Such a model (Fig. 6*b*) has an R_G that is 0.22 nm smaller than that of the extended structure. The 19 *N*-terminal residues were represented by a further 12 cubes close to His-20. Robson and Chou–Fasman secondary-structure predictions (Garnier *et al.*, 1978; Chou & Fasman, 1978) suggest that six to 11 of these residues constitute an α-helix of length 1–2 nm. The 19 residues were taken to be fully exposed to solvent since they contain up to ten charged groups and since this *N*-terminal section is readily cleaved by proteolysis (see the Materials and methods section). The *N*-terminus in native α_1-AT was thus modelled as a $6\times2\times1$ array of spheres of length 3.6 nm extending away from the protein surface, which increases the R_G by 0.04 nm.

A model for native α_1-AT requires also the rejoining of Met-358 and Ser-359 in split α_1-AT (Fig. 6*a*). Löbermann *et al.* (1984) proposed that residues 359–394 are fixed in the two structures, and residues 343–358 have been relocated. This is supported by the susceptibility of residues 351–358 to proteolytic attack (Kress *et al.*, 1979), and also by the ability of the *C*-terminal residues 359–394 to bind tightly to split α_1-AT by non-covalent forces (Löbermann *et al.*, 1982). Inspection of the α_1-AT crystal structure suggested that the minimum conformational change required to form native α_1-AT involves residues 348–358. Since these correspond to seven cubes, those closest to residues 348–358

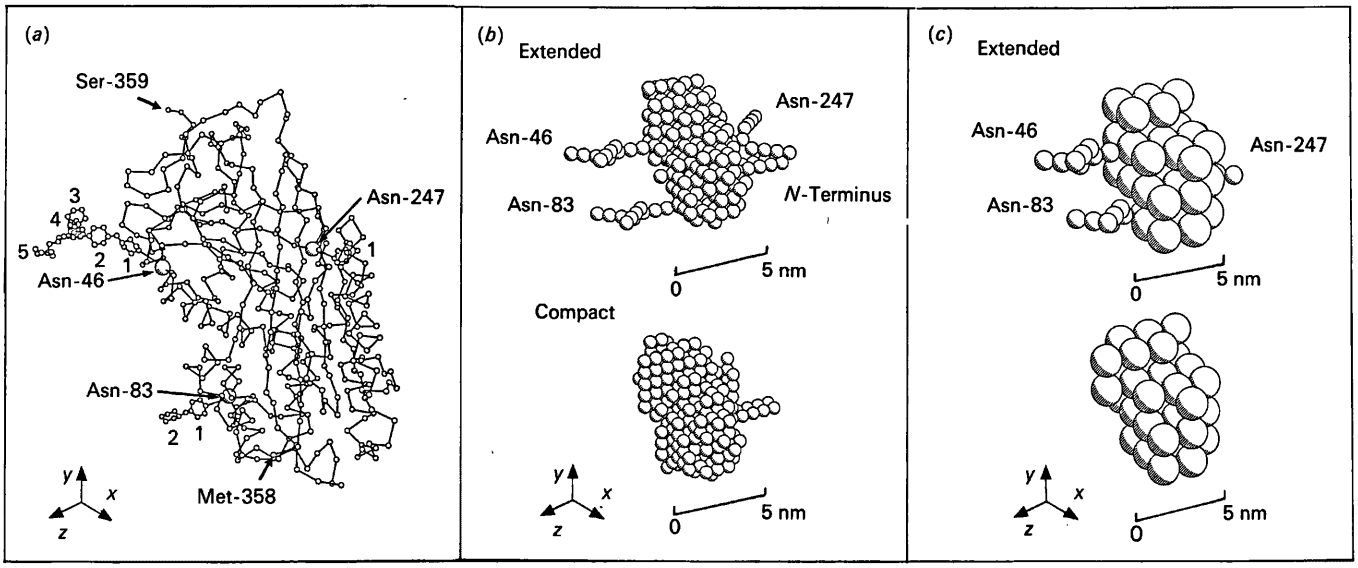


Fig. 6. Comparative views of the crystal structure and models for α_1-AT

(*a*) The α-carbon co-ordinates of tetragonal split α_1-AT are shown (Löbermann *et al.*, 1984). The α-carbon atoms of Asn-46 and Asn-83 are on the $+z$ face of the structure and are indicated by larger spheres, and their associated carbohydrate residues where visible in the electron-density map are numbered as shown. That for Asn-247 is on the $-z$ face. Residues Met-358 and Ser-359 are arrowed. (*b*) The final extended Debye sphere model for native α_1-AT is shown in the same orientation as in (*a*), as employed for curves *1* and *2* of Figs. 5(*a*), 5(*b*) and 5(*c*) (see the text). The three oligosaccharide groups are shown in an extended conformation (curve *1*) and are assumed to possess biantennary structures. For comparison, the compact conformation (curve *2*) is shown also. (*c*) The hydrodynamic model for native α_1-AT is depicted also in the same orientation. The protein core is represented by 32 spheres of diameter 1.69 nm. In the extended model, each of the three carbohydrate chains is represented by eight spheres of diameter 0.95 nm. The compact model is shown as 36 spheres of diameter 1.69 nm. It should be noted that (i) the mean distance between the α_1-AT molecules in a 20 mg/ml solution of α_1-AT is 16 nm; (ii) the maximum length of the extended model shown in (*b*) is 11 nm; (iii) the oligosaccharide chains extend by 3.5 to 3.9 nm from the protein surface; (iv) the upper and lower structures in (*b*) and (*c*) as shown are displaced by 13 nm.

were relocated to rejoin Met-358 and Ser-359 on the protein surface. The R_G of the native α_1-AT model is increased by 0.03 nm, which is not detectable experimentally.

The curve-fitting to the neutron and X-ray data for native α_1-AT in three contrasts is summarized by curves *1* and *2* in Fig. 5. The model with extended carbohydrate chains gives the three good curve *1* fits for native α_1-AT. The residual **R** is 0.006, 0.017 and 0.011 in that order for (*a*), (*b*) and (*c*) in Fig. 5. The calculations based on the compact carbohydrate model give worsened curve fits, as shown by visual inspection of the curves *2* in Fig. 5 and the **R** of 0.010, 0.019 and 0.028. This is clearest in Fig. 5(*c*), since this contrast corresponds to the strongest contribution of carbohydrate to the scattering curve. In positive contrasts, X-rays result in better signal-to-noise ratios at large Q than neutrons. The R_{G-C} of the extended carbohydrate model (curve *1*) is 2.58 nm, which agrees well with 2.61 nm from the experimental determination (Table 1). The compact carbohydrate model (curve *2*) gives worsened agreement, with an R_{G-C} of 2.36 nm. The α of the extended model is 29×10^{-5}, which is in better agreement with the α of $29 \times 10^{-5} \pm 5 \times 10^{-5}$ by experiment (Table 1) than that of the compact model with an α of 20×10^{-5}. The density-weighted X-ray R_G of the extended model is 2.97 nm, in good agreement with the experimental X-ray values of 2.80 to 3.00 nm (Fig. 4). It is concluded that the solution structure of native α_1-AT is similar to that seen in the crystal structure if the carbohydrate and *N*-terminal moieties are in extended conformations.

Two sets of control calculations indicate the importance of incorporating the internal scattering density and hydration into these models.

(*a*) Given the large α values, it is necessary to incorporate the internal inhomogeneity of scattering densities in the neutron simulations (see the Materials and methods section). If this is neglected, it leads to simulated curves at large Q that underestimate the neutron $I(Q)$ data in negative contrasts in ^2H_2O (curve *3* of Fig. 5*a*; **R** of 0.020) and overestimate those in positive contrasts in ^1H_2O (curve *3* of Fig. 5*b*; **R** of 0.017).

(*b*) Likewise the X-ray simulations require an adequate hydration shell of 0.3 g of water/g of α_1-AT in addition to the internal inhomogeneity of electron densities. Calculations for the dry model gave curves that generally overestimated the X-ray $I(Q)$ data (curve *3* of Fig. 5*c*; **R** of 0.035). Too great a hydration of 0.67 g of water/g of α_1-AT leads to worsened agreements beyond $Q = 1$ nm^{-1} (curve *4* of Fig. 5*c*; **R** of 0.028). The poorer fits of curves *4* in Figs. 5(*a*) and 5(*b*) (**R** of 0.024 and 0.029 respectively) show that hydration is not appropriate for neutron curve modelling.

As an independent test of the α_1-AT solution models, hydrodynamic models were derived from the crystal structure in order to calculate the frictional coefficient (see the Materials and methods section). The sedimentation coefficient $s^0_{20,w}$ of α_1-AT has been reported as 3.3 S, 3.41 S and 3.45 S (Bundy & Mehl, 1959; Shamash & Rimon, 1966; Kress & Laskowski, 1973). The frictional coefficient is thus 6.7×10^{-8}–7.0×10^{-8} g/cm per s. From this the frictional ratio f/f_0 is calculated as 1.29–1.35, to show that this has a moderately elongated structure, in agreement with the conclusion from the R_{G-C}/R_0 ratio of 1.35. The 31 hydrodynamic spheres for the protein core are of diameter 1.69 nm. One more was added to allow for the *N*-terminus. Each carbohydrate moiety was represented by eight spheres of diameter 0.95 nm in the extended model (Fig. 6*c*). In the compact model, these were replaced by four spheres of diameter 1.69 nm. The frictional coefficients are 7.2×10^{-8} and 6.4×10^{-8} g/cm per s respectively for the extended and compact models, which are greater than and less than the estimated experimental range. This suggests that the carbohydrate of α_1-AT exists in a range of

extended conformations, and is consistent with the Debye modelling.

Conclusions

Large conformational and stability changes have been reported between the native and split forms of α_1-AT (Löbermann *et al.*, 1984; Carrell & Owen, 1985; Bruch *et al.*, 1988). No evidence for a gross conformational change was found by solution scattering. This is explicable in terms of the observation by Löbermann *et al.* (1984) that, if β-strand 4A (residues 343–358) is removed from the centre of β-sheet A to re-form native α_1-AT, the structural integrity may largely be conserved. In the modelling, this change corresponds to the relocation of just seven spheres out of the total of 291 in the split α_1-AT model (Fig. 6*b*). This places an upper structural limit on the extent of the changes seen by spectroscopy (Bruch *et al.*, 1988).

How large is a conformational change before it is detectable by scattering? The neutron and X-ray R_G of deoxyhaemoglobin increased by 0.05–0.15 nm on the binding of O_2 (Conrad *et al.*, 1969; Schneider *et al.*, 1969); this corresponds to α-helix movements of 0.05–0.15 nm and subunit movements of 0.6 nm (Baldwin & Chothia, 1979). The X-ray R_G of aspartate transcarbamoylase increased by 0.25 nm after ligation to a substrate analogue (Moody *et al.*, 1979; Hervé *et al.*, 1985) and corresponds to 1.2 nm movements of the catalytic trimers and rotational reorientations of 10° to 15° between the regulatory and catalytic subunits (Krause *et al.*, 1987; Ke *et al.*, 1988). Sometimes no changes are seen by scattering, such as in the activation of chymotrypsinogen to α-chymotrypsin (Krigbaum & Godwin, 1968) or in the binding of pyridoxal 5′-phosphate to aspartate aminotransferase (Vergé *et al.*, 1983). The observation of change by solution scattering thus depends on the protein in question.

The conformational properties of oligosaccharides in glycoproteins are not readily accessed by physical techniques. The R_G parameter in solution scattering is of potential value, since this corresponds to the mean square distance of scattering vectors from the centre of scattering. Oligosaccharides usually lie on the protein surface, and will contribute maximally to the R_G. Even though the Stuhrmann α is less accurately measured than the R_C, this may sometimes be of use. In the present study, the oligosaccharide chains of α_1-AT were found to be largely extended into solution. These are mostly biantennary oligosaccharides, but may be replaced at Asn-83 and to a lesser extent at Asn-46 by triantennary oligosaccharides (Mega *et al.*, 1980; Vaughan & Carrell, 1981).

The present study has tested approaches for the modelling of scattering curves starting from a known crystal structure. It was found necessary in the simulations to allow for the different scattering densities of the hydrophilic and hydrophobic residues. This requirement can, however, be unnecessary if the protein has a low Stuhrmann α for its R_G value. Neutron modelling is best performed with structures that are largely or fully non-hydrated; this is consistent with experience of neutron-scattering densities for glycoproteins (Perkins, 1986). In distinction, X-ray modelling requires hydrated structures. This is consistent with the calculation of partial specific volumes from the sum of the unhydrated volume and the electrostricted water shell surrounding the protein (Perkins, 1986), and also with the calculation of electron densities for X-ray-scattering studies by contrast variation (Perkins, 1988*b*).

We thank the Wellcome Trust for support, and the Science and Engineering Research Council for access to the facilities at the Synchrotron Radiation Source (Daresbury) and the Institut Laue Langevin (Grenoble). Dr. W. Bras, Dr. H. Gerritsen, Dr. C. Nave, Dr. J. Torbet and Dr. E. Towns-Andrews are thanked for generous instrumental support, and Mr. A. S. Nealis is thanked for computational support. The technical assistance of Mr. D. J. Seilly in protein preparations is gratefully acknowledged.

REFERENCES

Baldwin, J. & Chothia, C. (1979) J. Mol. Biol. **129**, 175–220
Bao, J.-J., Sifers, R. N., Kidd, V. J., Ledley, F. D. & Woo, S. L. C. (1987) Biochemistry **26**, 7755–7759
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) J. Mol. Biol. **112**, 535–542
Blundell, T. L. & Johnson, L. N. (1976) Protein Crystallography, pp. 428–429, Academic Press, London
Bock, S.-C., Skriver, K., Nielsen, E., Thøgersen, H.-C., Wiman, B., Donaldson, W. H., Eddy, R. L., Marrian, J., Radziejewska, E., Huber, R., Shows, T. B. & Magnusson, S. (1986) Biochemistry **25**, 4292–4301
Bruch, M., Weiss, V. & Engel, J. (1988) J. Biol. Chem. **263**, 16626–16630
Bundy, H. F. & Mehl, J. W. (1959) J. Biol. Chem. **234**, 1124–1128
Carrell, L. & Owen, M. C. (1985) Nature (London) **317**, 730–732
Carrell, L., Boswell, D. R., Brennan, S. O. & Owen, M. C. (1980) Biochem. Biophys. Res. Commun. **93**, 399–402
Carrell, L., Jeppsson, J.-O., Vaughan, L., Brennan, S. O., Owen, M. C. & Boswell, D. R. (1981) FEBS Lett. **135**, 301–305
Carrell, L., Jeppsson, J.-O., Laurell, C.-B., Brennan, S. O., Owen, M. C., Vaughan, L. & Boswell, D. R. (1982) Nature (London) **298**, 329–334
Chandra, T., Stackhouse, R., Kidd, V. J., Robson, K. J. A. & Woo, S. L. C. (1983) Biochemistry **22**, 5056–5061
Chothia, C. (1975) Nature (London) **254**, 304–308
Chou, P. Y. & Fasman, G. D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. **47**, 45–148
Cohn, E. J. & Edsall, J. T. (1943) Proteins, Amino Acids and Peptides, pp. 155–176 and 370–381, Reinhold Publishing Corp., New York
Conrad, H., Mayer, A., Thomas, H. P. & Vogel, H. (1969) J. Mol. Biol. **41**, 225–229
Cusack, S. (1981) J. Mol. Biol. **145**, 539–541
Edelhoch, H. (1967) Biochemistry **6**, 1948–1954
Garcia de la Torre, J. & Bloomfield, V. A. (1977*a*) Biopolymers **16**, 1747–1761
Garcia de la Torre, J. & Bloomfield, V. A. (1977*b*) Biopolymers **16**, 1779–1793
Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) J. Mol. Biol. **120**, 97–120
Ghosh, R. E. (1981) Institut Laue Langevin, Internal Publication 81GH29T
Hervé, G., Moody, M. F., Tauc, P., Vachette, P. & Jones, P. T. (1985) J. Mol. Biol. **185**, 189–199
Ibel, K. & Stuhrmann, H. B. (1975) J. Mol. Biol. **93**, 255–266
Jacrot, B. & Zaccai, G. (1981) Biopolymers **20**, 2413–2426
Jeppsson, J. O., Laurell, C. B. & Fagerhol, M. K. (1978) Eur. J. Biochem. **83**, 143–153
Johnson, D. A. & Travis, J. (1977) Biochem. J. **163**, 639–641
Johnson, D. A. & Travis, J. (1978) J. Biol. Chem. **253**, 7142–7144
Ke, H., Lipscomb, W. N., Cho, Y. & Honzatko, R. B. (1988) J. Mol. Biol. **204**, 725–747
Krause, K. L., Volz, K. W. & Lipscomb, W. N. (1987) J. Mol. Biol. **193**, 527–553
Kress, L. F. & Laskowski, M. (1973) Prep. Biochem. **3**, 541–552
Kress, L. F., Kurecki, T., Shung, K. C. & Laskowski, M. (1979) J. Biol. Chem. **254**, 5317–5320
Krigbaum, W. R. & Godwin, R. W. (1968) Biochemistry **7**, 3126–3131
Kurachi, K., Chandra, T., Friezner Deger, S. J., White, T. T., Marchioro, T. L., Woo, S. L. C. & Davie, E. W. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 6826–6830
Laemmli, U. K. (1970) Nature (London) **277**, 680–685
Laurell, C.-B. (1966) Anal. Biochem. **15**, 45–52
Laurell, C.-B., Pierce. J., Persson, U. & Thulin, E. (1975) Eur. J. Biochem. **57**, 107–113
Lijnen, H. R., Holmes, W. E., van Hoef, B., Wiman, B., Rodiguez, H. & Collen, D. (1987) Eur. J. Biochem. **166**, 565–574
Löbermann, H., Lottspeich, F., Bode, W. & Huber, R. (1982) Hoppe-Seyler's Z. Physiol. Chem. **363**, 1377–1388
Löbermann, D., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984) J. Mol. Biol. **177**, 531–556

Mega, T., Lujan, E. & Yoshida, A. (1980) J. Biol. Chem. **255**, 4053–4056
Moody, M. F., Vachette, P. & Foote, A. M. (1979) J. Mol. Biol. **133**, 517–532
Ny, T., Sawdey, M., Lawrence, D., Millan, J. L. & Loskutoff, D. J. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 6776–6780
Perkins, S. J. (1982) Biol. Magn. Reson. **4**, 193–336
Perkins, S. J. (1985) Biochem. J. **228**, 13–26
Perkins, S. J. (1986) Eur. J. Biochem. **157**, 169–180
Perkins, S. J. (1988*a*) Biochem. J. **254**, 313–327
Perkins, S. J. (1988*b*) New Compr. Biochem. **18B**, Part 2, 143–264
Perkins, S. J. (1989) in Dynamic Properties of Biomolecular Assemblies, pp. 226–245 (Harding, S. E. & Rowe, A. J., eds.), Royal Society of Chemistry, London
Perkins, S. J. & Weiss, H. (1983) J. Mol. Biol. **168**, 847–866
Perkins, S. J., Kerckaert, J. P. & Loucheux-Lefebvre, M. H. (1985) Eur. J. Biochem. **147**, 525–531
Perkins, S. J., Nealis, A. S. & Sim, R. B. (1990) Biochemistry, in the press
Rolbin, Y. A., Kayushina, R. L., Feigin, L. A. & Schredin, B. M. (1973) Kristallografiya **18**, 701–705
Schneider, R., Mayer, A., Schmatz, W., Kaiser, B. & Scherm, R. (1969) J. Mol. Biol. **41**, 231–235
Shamash, Y. & Rimon, A. (1966) Biochim. Biophys. Acta **121**, 35–41
Towns-Andrews, E., Berry, A., Bordas, J., Mant, G. R., Murray, P. K., Roberts, K., Sumner, I., Worgan, J. S., Lewis, R. & Gabriel, A. (1989) Rev. Sci. Instrum. **60**, 2346–2349
Travis, J. & Salvesen, G. S. (1983) Annu. Rev. Biochem. **52**, 655–709
Vaughan, L. & Carrell, R. W. (1981) Biochem. Int. **2**, 461–467
Vaughan, L., Lorier, M. A. & Carrell, R. W. (1982) Biochim. Biophys. Acta **701**, 339–345
Vergé, D., Tardieu, A. & Arrio-Dupont, M. (1983) FEBS Lett. **154**, 277–281
Wetlaufer, D. B. (1962) Adv. Protein Chem. **17**, 303–390

Received 24 August 1989/19 October 1989; accepted 24 October 1989

Conformational Transition between Native and Reactive Center Cleaved Forms of α_1-Antitrypsin by Fourier Transform Infrared Spectroscopy and Small-Angle Neutron Scattering

Parvez I. Haris,[‡] Dennis Chapman,[‡] Richard A. Harrison,[§] Kathryn F. Smith,[‡,∥] and Stephen J. Perkins*[,‡,∥]

Departments of Protein and Molecular Biology and of Biochemistry and Chemistry, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K., and Molecular Immunopathology Unit, MRC Centre, Hills Road, Cambridge CB2 2QH, U.K.

Received November 1, 1989; Revised Manuscript Received December 12, 1989

ABSTRACT: α_1-Antitrypsin (α_1-AT) is the best-characterized member of the serpin superfamily of plasma proteins. Protease inhibitor members of this family undergo a characteristic reactive-center cleavage during expression of their inhibitory activity. The physical basis of this transition in α_1-AT from the stressed native conformation to the more stable reactive center cleaved (split) form was studied by Fourier transform infrared (FT-IR) spectroscopy and neutron scattering. The FT-IR spectra show that, while split α_1-AT has three intense well-resolved components associated with the presence of antiparallel β-sheet and α-helix conformations, the amide I band of native α_1-AT has only one intense component, associated with the presence of β-sheet structure. ^1H–^2H exchange within the polypeptide backbone, studied by FT-IR and NMR spectroscopy, shows that the native form undergoes greater exchange than the split form. Under the same conditions, neutron scattering shows no differences in the radius of gyration R_G of the native and the split forms. In contrast, in high concentrations of phosphate approaching those used for crystallization, the native form (unlike the split form) undergoes dimerization. These data indicate that the conformational transition largely involves localized secondary and tertiary structure rearrangements. We propose that the energetically stressed native α_1-AT structure is the consequence of a significantly reduced number of hydrogen bonds in secondary structure components and that reactive-site cleavage between Met358 and Ser359 is the key for the development of the fully hydrogen bonded more stable serpin structure.

Members of the serpin superfamily, which includes many plasma proteinase inhibitors, present a reactive site as an ideal substrate to the target proteinase (Carrell & Travis, 1985). After reaction, a stable stoichiometric 1:1 complex between proteinase and reactive center cleaved (split) inhibitor is formed. The crystal structure of a reactive center cleaved form of α_1-AT shows that on cleavage a significant conformational change must occur, since the newly formed termini at Met358 and Ser359 are 7 nm apart (Löbermann et al., 1984). Despite this, details of the physicochemical basis of the improved stability of the split form (Carrell & Owen, 1985) remain unclear.

We report here Fourier transform infrared (FT-IR) spectroscopy (Lee & Chapman, 1986; Susi & Byler, 1986; Surewicz & Mantsch, 1988) and small-angle neutron scattering (Perkins, 1988a,b) studies of α_1-AT. These are complementary techniques that examine the polypeptide backbone conformation and the gross structure of α_1-AT in solution, both under conditions close to physiological and to those of the crystal structure (2.6 M phosphate). FT-IR spectroscopy has advantages over circular dichroism for the study of secondary structures since there is no problem in defining a base line in the FT-IR spectrum, different parts of the FT-IR spectrum can be clearly resolved, and information is provided on amide proton exchange rates in proteins. Neutron scattering rather than X-ray scattering is more advantageous, since the sample transmission in 100% ^2H_2O (2-mm path length) is reduced by only 3% in 1.8 M phosphate relative to 0.05 M phosphate, while the reduction for X-rays (1-mm path length) is 81% in 1.2 M phosphate. The combined use of both techniques places limits on the structural changes seen between the two forms of α_1-AT and leads to the proposal of a mechanism by which the split form of α_1-AT is stabilized.

Materials and Methods

The preparation of α_1-AT and the characterization of the samples is described in Smith et al. (1990). The FT-IR, NMR, and neutron samples were dialyzed over 36 h into sodium potassium phosphate buffers ranging from 0.05 to 1.80 M in 100% ^2H_2O at 6 °C with four changes of buffers.

Infrared spectra were recorded on a Perkin-Elmer 1750 Fourier transform infrared spectrometer equipped with a TGS detector. A Perkin-Elmer Model 7300 data station was used for data acquisition, storage, and analysis. Spectral conditions [defined more fully in Haris et al. (1986) and Perkins et al. (1988)] were as follows: α_1-AT concentration 23–25 mg/mL on the basis of an A_{280} (1%, 1 cm) of 5.4; 400 scans (1-h accumulation); spectral resolution 4 cm^{-1}; sample thickness of 50 μm using a Teflon spacer. Spectral deconvolution was performed with the Perkin-Elmer ENHANCE function (Haris et al., 1986), which is analogous to the method developed by Kauppinen et al. (1981), using deconvolution parameters of $\sigma = 8$ (half-width at half-height) and $k = 2.25$. The second-derivative spectra were calculated over a 13-cm^{-1} range.

Neutron scattering data were measured at 20 °C on the small-angle scattering instrument D17 at the Institut Laue Langevin, Grenoble, with a wavelength λ of 1.105 nm

* To whom correspondence should be addressed at the Department of Biochemistry and Chemistry.

‡ Department of Protein and Molecular Biology, Royal Free Hospital School of Medicine.

§ MRC Centre.

∥ Department of Biochemistry and Chemistry, Royal Free Hospital School of Medicine.

0006-2960/90/0429-1377$02.50/0

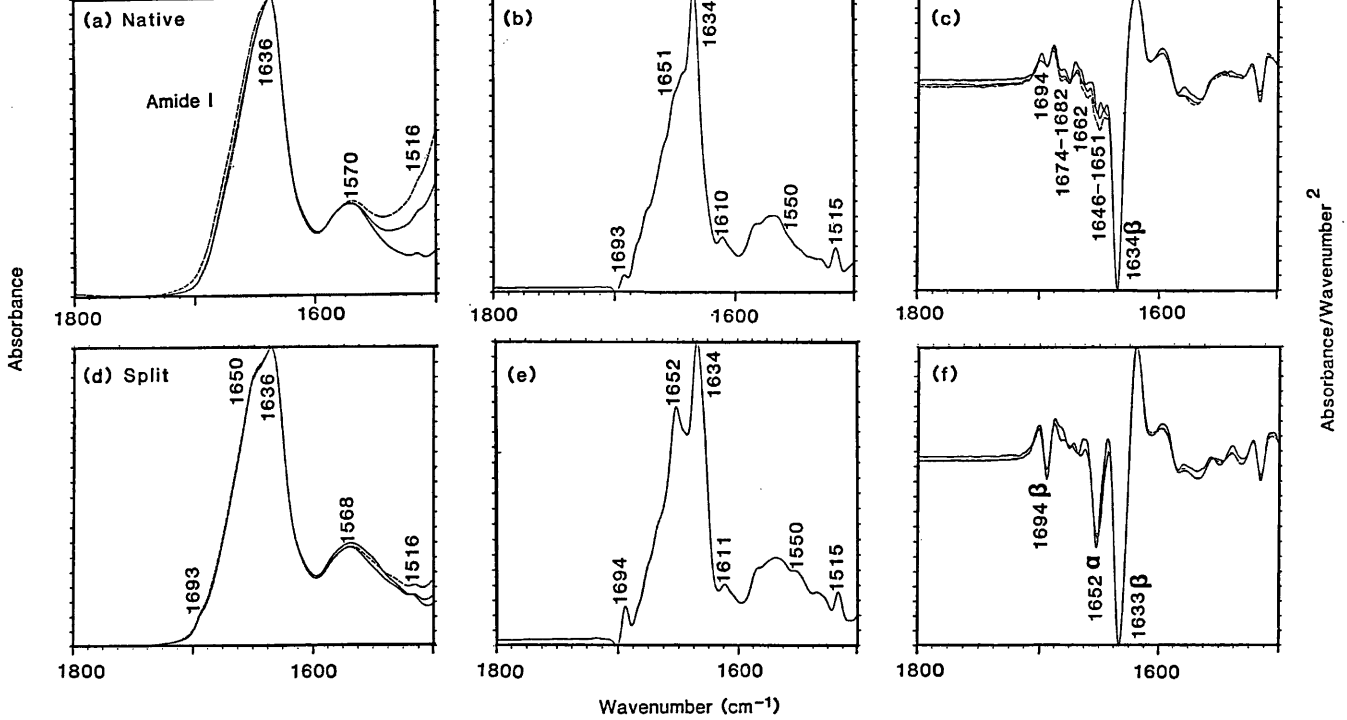


FIGURE 1: Fourier transform infrared spectroscopy of native and split α_1-AT. Panels a, b, and c correspond to the absorbance, deconvoluted, and second-derivative spectra of native α_1-AT in that order; panels d, e, and f correspond likewise to split α_1-AT. The absorbance and second-derivative spectra are shown for samples recorded in 0.05 (—), 0.50 (—), and 1.25 M (- - -) sodium potassium phosphate buffers in ^2H_2O at 20 °C. The increase in signal intensity at 1500 cm^{-1} for native α_1-AT in 0.50 and 1.25 M phosphate buffers is attributed to the presence of trace amounts of $^1HO^2H$. The deconvoluted spectra were recorded in 0.05 M sodium potassium phosphate buffers; the peaks are not fully labeled for reason of clarity but are similar to those in the second-derivative spectra.

(wavelength spread $\Delta\lambda/\lambda$ of 10%) and a sample-detector distance of 3.396 m to correspond to a Q range of 0.07–0.60 nm^{-1} ($Q = 4\pi \sin\theta/\lambda$; 2θ = scattering angle). Data collection employed 2 mm thick Hellma cells. Analyses followed standard procedures, which utilized backgrounds based on the scattering of the buffer dialyzate, cadmium, H_2O, and an empty Hellma quartz cell and measurements of sample and buffer transmissions (Ghosh, 1989; Perkins et al., 1990; Smith et al., 1990). The radius of gyration R_G and the forward scattering at zero scattering angle $I(0)$ were determined from plots of ln $I(Q)$ vs Q^2 with the Guinier equation (Glatter & Kratky, 1982): $\ln I(Q) = \ln I(0) - R_G^2Q^2/3$ in the Q range 0.15–0.58 nm^{-1} except for native α_1-AT above 1.2 M phosphate, where it was 0.15–0.41 nm^{-1}.

Results and Discussion

Panels a and d of Figure 1 show the Fourier transform infrared spectra of native and split α_1-AT. The amide I maxima in both cases is centered at 1636 cm^{-1}. The feature at 1568–1570 cm^{-1} corresponds to absorption from carboxylate side chains. The 1516-cm^{-1} component is normally attributed to tyrosine side chain absorption. The frequency of the main amide I band for both samples is consistent with the presence of a predominantly β-sheet structure in α_1-AT (Surewicz & Mantsch, 1988; Haris et al., 1986). However, native and split α_1-AT differ in that additional features near 1650 and 1693 cm^{-1} are present in the split form. These were reproducibly observed in 0.05, 0.50, and 1.25 M phosphate (and also in 0.80, 1.60, and 1.80 M phosphate; not shown). The second-derivative spectra of the two forms in panels c and f of Figure 1 emphasize the differences more clearly. The main β-sheet feature at 1633–1634 cm^{-1} is visible in both forms. The split form shows considerably more intense β-sheet and α-helix bands at 1694 and 1652 cm^{-1}, and the band at 1633 cm^{-1} is broader. The bands near 1650 and 1690 cm^{-1} have been attributed to α-helix and antiparallel β-sheet structures, respectively (Haris et al., 1986; Olinger et al., 1986). For both native and split α_1-AT, similar weaker features were generally observed near 1646, 1662, and 1674–1682 cm^{-1}. The feature visible in the native form near 1646 cm^{-1} most likely corresponds to disordered structures. Those near 1662 and 1674–1682 cm^{-1} probably correspond to turn structures, although overlap from β-sheet and 3_{10} helices may also occur. The use of deconvolution for resolution enhancement led to very similar results (Figure 1b,e).

The FT-IR spectroscopic results clearly show that the α-helical and antiparallel β-sheet content has significantly increased after reactive-site cleavage of α_1-AT. Methods for the quantitative assessment of the differences seen between the two forms of α_1-AT by curve fitting of the amide I band are however insufficiently accurate at the present time (Mantsch et al., 1989). Inspection of the crystal structure of split α_1-AT (32% α-helix; 40% β-sheet) suggests that in native α_1-AT strand A4 is withdrawn from β-sheet A (Löbermann et al., 1984). The formation of about 20–30 additional hydrogen bonds in the largely antiparallel β-sheet A of split α_1-AT would account for the appearance of the band at 1694 cm^{-1}. The increase in α-helix content after cleavage is however unexpected. In the crystal structure, in linear sequence, helices A and B flank strand B6, helices D, E, and F alternate with strands A2, A1 and A3, and helices H and I flank strand A6. The infrared data suggest that in the native molecule either these helices are not formed or they are considerably distorted along their length at the same time that β-sheet A is incompletely formed. Thus, after reactive-site cleavage, a concerted drive toward the formation of stable, fully hydrogen bonded α-helix and β-sheet structures results in a molecule (split α_1-AT) of increased stability.

The intensity of the amide II band at 1552 cm^{-1} monitors the residual ^1H–^2H exchange of backbone NH protons. Panels

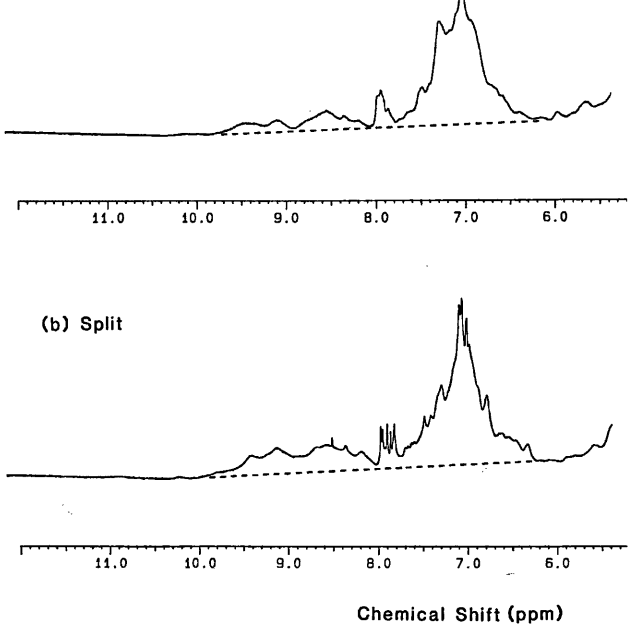


FIGURE 2: 400-MHz ^1H NMR spectra of native and split α_1-AT at 30 °C in 0.05 M sodium potassium phosphate buffers in ^2H_2O (concentrations 23–25 mg/mL). The downfield region is shown, with the chemical shift scale set relative to the $^1HO^2H$ signal at 4.695 ppm downfield of (trimethylsilyl)propionic acid. The aromatic proton signals from 27 Phe, 13 His, 6 Tyr, and 2 Trp residues were assigned to the envelope between 6.3 and 8.0 ppm and the amide proton signals to that between 8.0 and 9.9 ppm. 256 scans were recorded with water decoupling and 3 s between pulses. The spectral resolution was improved with a sine bell squared routine.

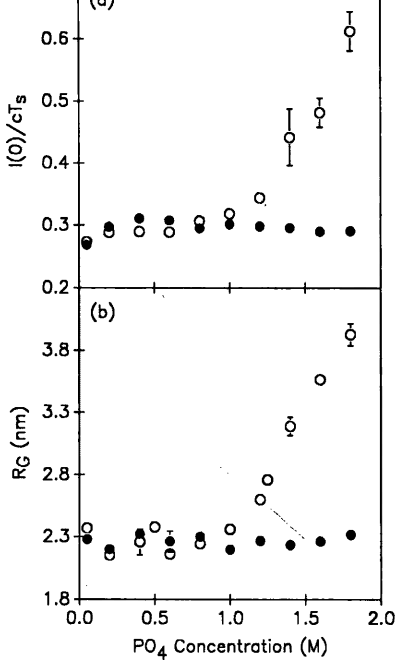


FIGURE 3: Summary of neutron Guinier R_G and $I(0)$ analyses of native (O) and split (●) α_1-AT in ^2H_2O buffers. (a) Dependence of $I(0)/cT_s$ on phosphate concentration after extrapolation of $I(0)/cT_s$ to zero concentration in each buffer (c, α_1-AT concentration; T_s, sample transmission). Between three and six measurements were made at each phosphate concentration in a range of c of 3–12 mg/mL. (b) Corresponding R_G values after extrapolation to zero concentration.

a and d of Figure 1 show that, after identical, simultaneous dialysis into ^2H_2O, this is 15% lower for the native form than for the split form. In the deconvoluted spectra of Figure 1b,e, a peak at 1550 cm^{-1} is visible for split α_1-AT but not for native α_1-AT. These indicate that ^1H–^2H exchange is more extensive in the native form, as might be expected for a more open structure. This was confirmed by 400-MHz ^1H NMR spectroscopy (Figure 2) on the same samples used in Figure 1b,e. Integration of the spectra downfield of the 195 aromatic proton signals shows that while 11% of the 377 NH protons are nonexchanged in native α_1-AT this increases to 19% in split α_1-AT. The greater exchange in native α_1-AT is consistent with a reduced content of ordered secondary structures. [Other NMR spectral differences between the two forms include the observation of narrower line widths in split α_1-AT (Figure 2) and rearranged patterns in the ring current shifted methyl proton signals at the high-field end of the spectra. Both of these are consistent with the existence of structural and dynamic differences between the two forms of α_1-AT.]

The magnitude of the observed changes in backbone structure (Figure 1; Bruch et al., 1988) suggests that native and split α_1-AT may have different gross structures. This was investigated by small-angle neutron scattering. The R_G and $I(0)/c$ Guinier parameters from dilution series are dependent on α_1-AT concentrations c between 0.05 and 0.50 M phosphate. The values after extrapolation to zero c are shown in Figure 3. No difference could be detected between the native and split forms of α_1-AT up to 0.80 M phosphate, even though for other proteins such as hemoglobin (Conrad et al., 1969), aspartate transcarbamylase (Moody et al., 1979), and hexokinase (McDonald et al., 1979) the appropriate conformational changes after ligation are detectable from Guinier analyses. The folded tertiary structures of native and split α_1-AT are thus essentially similar, and this places an upper limit on the structural changes in the polypeptide backbone indicated by FT-IR, circular dichroism (CD), and NMR spectroscopy (Smith et al., 1990).

Final experiments were performed at high phosphate concentrations approaching those used for crystallization. The FT-IR data in 1.25 M (Figure 1), 1.60 M, and 1.80 M phosphate show no significant difference from those in 0.05 M, 0.50 M (Figure 1), and 0.80 M phosphate. However, the neutron data (Figure 3) show that above 1 M phosphate the $I(0)/c$ data rise to double the low-phosphate values for native α_1-AT but are unchanged for the split form. This shows that native α_1-AT forms dimers in 1.8 M phosphate. The corresponding R_G data for native α_1-AT rise in parallel. While the structure of split α_1-AT is unaffected by high phosphate, the association of native α_1-AT indicates that the surface properties of several amino acid residues are significantly different.

CONCLUSIONS

Neutron scattering shows that under physiological conditions the gross folded structures of native and split α_1-AT are similar. However, within the secondary and tertiary structures of both forms of α_1-AT, subtle but marked structural differences were identified by use of FT-IR spectroscopy and neutron scattering. CD and NMR spectroscopic studies (Figure 2; Bruch et al., 1988; Gettins & Harten, 1988) have also shown that there are conformational differences between the native and split forms of other members of the serpin superfamily. These data are supported by monoclonal antibodies with different epitopic specificities which bind either to native or to split α_1-AT (Zhu & Chan, 1987).

In earlier studies, the stabilization of split α_1-AT has generally been discussed in terms of the localized relief of strain caused by cleavage of the Met358–Ser359 bond within a stressed surface-exposed polypeptide loop (Löbermann et al., 1984; Carrell & Travis, 1985; Carrell & Owen, 1985; Bruch et al., 1988), although Löbermann et al. (1984) have noted that strain in a molecule is generally thought to be distributed

changes in the spectral properties of the polypeptide backbone that are observed by FT-IR spectroscopy can be directly assigned to the formation of well-defined hydrogen bonds in α-helices and β-sheets in split α_1-AT. The magnitude of these changes for two forms of a water-soluble protein measured under physiological conditions is unusual. Our data suggest that split α_1-AT is stabilized relative to the native form by a more complete hydrogen bonding of its secondary structure (particularly α-helical). The lability of the native serpin structure is consequent on a far more extensive degree of stress than has previously been supposed.

Acknowledgments

We thank the Wellcome Trust for support and the Science and Engineering Research Council for beamtime at the Institut Laue Langevin, Grenoble. We are grateful to Dr. J. Feeney and Dr. C. Bauer for use of the 400-MHz ^{1}H NMR spectrometer at the National Institute of Medical Research, Mill Hill, London, and Dr. A. de Geyer for support at ILL.

References

Bruch, M., Weiss, V., & Engel, J. (1988) *J. Biol. Chem. 263*, 16626–16630.

Carrell, R. W., & Owen, M. C. (1985) *Nature 317*, 730–732.

Carrell, R., & Travis, J. (1985) *Trends Biochem. Sci. 10*, 20–24.

Conrad, H., Mayer, A., Thomas, H. P., & Vogel, H. (1969) *J. Mol. Biol. 41*, 225–229.

Gettins, P., & Harten, B. (1988) *Biochemistry 27*, 3634–3639.

Ghosh, R. E. (1989) Internal Publication 89GH02T, Institut Laue Langevin, Grenoble, France.

Scattering, Academic Press, London.

Haris, P. I., Lee, D. C., & Chapman, D. (1986) *Biochim. Biophys. Acta 874*, 255–265.

Kauppinen, J. K., Moffatt, D. J., Mantsch, H. H., & Cameron, D. G. (1981) *Appl. Spectrosc. 35*, 271–276.

Lee, D. C., & Chapman, D. (1986) *Biosci. Rep. 6*, 235–256.

Löbermann, D., Tokuoka, R., Deisenhofer, J., & Huber, R. (1984) *J. Mol. Biol. 177*, 531–556.

Mantsch, H. H., Surewicz, W. K., Muga, A., Moffatt, D. J., & Casal, H. L. (1989) Proceedings of the 7th International Conference on FT-IR Spectroscopy, Fairfax, Canada.

McDonald, R. C., Steiz, T. A., & Engelman, D. M. (1979) *Biochemistry 18*, 338–342.

Moody, M. F., Vachette, P., & Foote, A. M. (1979) *J. Mol. Biol. 133*, 517–532.

Olinger, J. M., Hill, D. M., Jakobsen, R. J., & Broody, R. S. (1986) *Biochim. Biophys. Acta 869*, 89–98.

Perkins, S. J. (1988a) *Biochem. J. 254*, 313–327.

Perkins, S. J. (1988b) *New Compr. Biochem. 18B* (Part II), 143–264.

Perkins, S. J., Haris, P. I., Sim, R. B., & Chapman, D. (1988) *Biochemistry 27*, 4004–4012.

Perkins, S. J., Nealis, A. S., & Sim, R. B. (1990) *Biochemistry 29*, 1167–1175.

Smith, K. F., Harrison, R. A., & Perkins, S. J. (1990) *Biochem. J.* (in press).

Surewicz, W. K., & Mantsch, H. H. (1988) *Biochim. Biophys. Acta 952*, 115–130.

Susi, H., & Byler, D. M. (1986) *Methods Enzymol. 130*, 290–311.

Zhu, X.-J., & Chan, S. K. (1987) *Biochem. J. 246*, 19–23.

Reprinted from *J. Mol. Biol.* (1990) **214**, 751–763

Two-domain Structure of the Native and Reactive Centre Cleaved Forms of C$\bar{1}$ Inhibitor of Human Complement by Neutron Scattering

Stephen J. Perkins, Kathryn F. Smith, Supavadee Amatayakul David Ashford, Thomas W. Rademacher, Raymond A. Dwek Peter J. Lachmann and Richard A. Harrison

J. Mol. Biol. (1990) 214, 751–763

Two-domain Structure of the Native and Reactive Centre Cleaved Forms of C1 Inhibitor of Human Complement by Neutron Scattering

Stephen J. Perkins†, Kathryn F. Smith

Departments of Biochemistry and Chemistry and Protein and Molecular Biology Royal Free Hospital School of Medicine Rowland Hill Street, London NW3 2PF, U.K.

Supavadee Amatayakul, David Ashford, Thomas W. Rademacher Raymond A. Dwek

Glycobiology Unit, Department of Biochemistry University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Peter J. Lachmann and Richard A. Harrison

MRC Molecular Immunopathology Unit, MRC Centre Hills Road, Cambridge CB2 2QH, U.K.

(Received 14 November 1989; accepted 20 April 1990)

The C1 inhibitor component of human complement is a member of the serpin superfamily, and controls C1 activation. Carbohydrate analyses showed that there are seven O-linked oligosaccharides in C1 inhibitor. Together with six N-linked complex-type oligosaccharides, the carbohydrate content is therefore 26% by weight and the molecular weight (M_r) is calculated as 71,100. Neutron scattering gives an M_r of 76,000 ($\pm$4000) and a matchpoint of 41·8 to 42·3% ^2H_2O, in agreement with this carbohydrate and amino acid composition. Guinier plots to determine the radius of gyration R_G were biphasic. Neutron contrast variation of C1 inhibitor in H_2O–^2H_2O mixtures gave an overall radius of gyration R_G at infinite contrast of 4·85 nm, from analyses at low Q, and a cross-sectional R_G of 1·43 nm. The reactive centre cleaved form of C1 inhibitor has the same M_r and structure as the native molecule. The length of C1 inhibitor, 16 to 19 nm, is far greater than that of the putative serpin domain. This is attributed to an elongated structure for the carbohydrate-rich 113-residue N-terminal domain. The radial inhomogeneity of scattering density, α, is large at 59×10^{-5} from the R_G data and 28×10^{-5} from the cross-sectional analysis, and this is accounted for by the high oligosaccharide content of C1 inhibitor. The scattering data were modelled using small spheres. A two-domain structure of length 18 nm based on two distinct scattering densities accounted for all the contrast variation data. One domain is based on the crystal structure of α_1 antitrypsin (7 nm × 3 nm × 3 nm). The other corresponds to an extended heavily glycosylated N-terminal domain of length 15 nm, whose long axis is close to the longest axis of the serpin domain. Calculation of the sedimentation coefficient $s^0_{20,w}$ for C1 inhibitor using the hydrodynamic sphere approach showed that a two-domain head-and-tail structure with an M_r of 71,000 and longest axis of 16 to 19 nm successfully reproduced the $s^0_{20,w}$ of 3·7 S. Possible roles of the N-terminal domain in the function of C1 inhibitor are discussed.

† Author to whom correspondence should be addressed, in the Department of Biochemistry and Chemistry, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K.

0022-2836/90/150751-13 $03.00

1. Introduction

The complement system comprises a family of plasma and cell-surface glycoproteins that operate as a cascade and play a major role in immune defence mechanisms. The classical pathway of complement activation is initiated by the first component, C1, which reacts largely with immune complexes of IgM or IgG immunoglobulins bound to foreign material (Cooper, 1985; Reid, 1986; Schumaker *et al.*, 1987). The unrestrained activation of C1 would cause excessive consumption of C2 and C4, thus C1 is regulated by C1 inhibitor (Harpel, 1976; Sim & Reboul, 1981; Cooper 1985; Davis, 1988). This protease inhibitor, present at about 0·20 mg/ml in plasma, forms stable 1:1 complexes with the C1r and C1s subcomponents of C1 to prevent further complement activation. Complex formation is characterized by the irreversible cleavage of the peptide bond at Arg444-Thr445 in the reactive centre of C1 inhibitor (Salvesen *et al.*, 1985). Conformational changes after cleavage lead to a much more stable protein structure (Bruch *et al.*, 1988; Pemberton *et al.*, 1989). While C1 inhibitor is the only known inhibitor of C1r and C1s, it is also a critical regulatory component of the coagulation, fibrinolytic and kinin-releasing systems, and its absence or inactivation by "non-productive" cleavage of the Arg444-Thr445 peptide bond is important in certain disease states (Davis, 1988).

Compositional studies of C1 inhibitor are required to understand the mechanism of control of C1 activation in serum. C1 inhibitor is a member of the serpin superfamily (serine protease inhibitor) of serum inhibitors (Davis *et al.*, 1986; Bock *et al.*, 1986). Its M_r has been estimated as 98,000 to 116,000 by gel electrophoresis and ultracentrifugation analyses (Pensky *et al.*, 1961; Haupt *et al.*, 1970; Reboul *et al.*, 1977; Nilsson & Wiman, 1982; Harrison, 1983). However, the sequence shows that the mature protein M_r is only 52,300 and it has been suggested that this implies a carbohydrate content as high as 49% (Bock *et al.*, 1986). This is not supported by previous carbohydrate analyses, which found a total of 33% to 35% by mass (Haupt *et al.*, 1970; Harrison, 1983). The structure of the N-linked and O-linked oligosaccharides have been determined, but not their total amounts (Strecker *et al.*, 1985). Accordingly, further carbohydrate analyses were performed in order to define the composition of C1 inhibitor.

Structural data are required also to understand the function of C1 inhibitor. Sedimentation data for C1 inhibitor give $s^0_{20,w}$ values mostly at 3·7 S (Schultze *et al.*, 1962; Haupt *et al.*, 1970; Reboul *et al.*, 1977; Chesne *et al.*, 1982) from which highly elongated structures have been deduced (Odermatt *et al.*, 1981; Perkins, 1985). Electron microscopy suggested that C1 inhibitor has an elongated two-domain "head-and-tail" structure (Odermatt *et al.*, 1981), and microcalorimetry experiments supported this two-domain structure (Lennick *et al.*, 1985). The crystal structure of the homologous serpin α_1-antitrypsin can be used to model the C-terminal 365 amino acid residues (Löbermann *et al.*, 1984; Bock *et al.*, 1986; Harrison, 1989), which probably corresponds to the "head" of C1 inhibitor. The "tail" is therefore the heavily glycosylated N-terminal 113 amino acid residues, with a presumed length of anywhere between 26 nm and 53 nm. The structure of the N-terminal domain is poorly understood; indeed, the functional consequences of this domain are wholly unknown (Davis, 1988).

Neutron scattering is an effective multiparameter tool for structural studies of glycoproteins (Perkins *et al.*, 1985; Smith *et al.*, 1990). Contrast variation in H_2O–^2H_2O mixtures will identify the overall structure under conditions close to physiological, in which the internal arrangement of protein and carbohydrate can be allowed for (Perkins, 1988*a*, *b*). The conformational transition between the native and reactive centre cleaved proteins can be monitored. The two-domain model for C1 inhibitor can be compared quantitatively with the crystal structure of α_1-antitrypsin (Löbermann *et al.*, 1984; Smith *et al.*, 1990), where molecular models for C1 inhibitor can be developed. Possible roles for the heavily glycosylated N-terminal domain for the function of C1 inhibitor are discussed.

2. Materials and Methods

(a) *Preparations of the native and reactive centre cleaved forms of C1 inhibitor*

C1 inhibitor was prepared from human plasma according to the method of Harrison & Lachmann (1986) and either used fresh or stored frozen at −70°C until required. If frozen, samples were prepared for scattering studies by gel filtration on Sepharose 6B in 12 mM-sodium phosphate, 200 mM-NaCl, 1 mM-EDTA, pH 7·0. The peak fraction(s) at about 10 mg/ml were then dialysed against the same buffer, filtered through a 0·2 μm filter (Gelman) into sterile glass or plastic vials, and held at 4°C until required for analysis. Two forms of C1 inhibitor were used in analysis. The native form showed a single band on SDS/polyacrylamide gel electrophoresis (Laemmli, 1970), and was fully active against C1s and plasmin. The split form of C1 inhibitor was generated subsequent to isolation by the action of an unidentified protease(s). Gel electrophoresis (Laemmli, 1970;) and high-pressure liquid chromatography analyses of this showed it to have been split at 2 sites, one at or close to the reactive centre exposed loop, and the other close to the amino terminus of the protein, generating fragments indistinguishable from those generated by *Pseudomonas aeruginosa* elastase (Pemberton *et al.*, 1989). The sites of cleavage were confirmed by amino-terminal sequence analysis (performed by Dr L. Packman at the Protein Sequencing Facility, Department of Biochemistry, University of Cambridge) of the cleaved material both before and after dialysis against scattering buffers. This indicated that cleavage had occurred between Ser441 and Val442 in the reactive centre and, in approximately equimolar amounts, either between Met31 and Leu32 or between Leu32 and Phe33 in the amino-terminal region. While the C-terminal peptide, containing the reactive centre residues, was fully retained in the dialysed cleaved molecule, partial (up to 30%) loss of the amino-terminal 31/32 residues may have

occurred, although the nature of the residues at the amino terminus of the protein (N-P-N(CHO)-A-T-S-S-S-S) make this difficult to quantify precisely. Samples were reanalysed by gel electrophoresis subsequent to solution scattering studies, and no differences before or after the scattering experiments were detected. In addition, native $C\bar{1}$ inhibitor after the scattering experiments had an unaltered specific activity against active-site titrated plasmin.

(b) *Carbohydrate analysis of $C\bar{1}$ inhibitor*

The methods for the release of N-linked oligosaccharides by hydrazinolysis and their subsequent analysis have been described (Ashford *et al.*, 1987). Enzyme digestion of oligosaccharides with neuraminidase (*Arthrobacter ureafaciens*) and β-*N*-acetylhexosaminidase (Jack bean) have also been described (Parekh *et al.*, 1987).

O-linked oligosaccharides were released and reduced by alkaline sodium borohydride treatment using a modification of the method described by Mizuochi *et al.* (1980). Typically, 2·5 mg of $C\bar{1}$ inhibitor was treated with 250 μl of 0·6 M-sodium [^{3}H]borohydride (360 mCi/mmol, New England Nuclear) in 0·05 M-sodium hydroxide for 20 h at 50 °C with shaking. The mixture was cooled to 4 °C and 12 μl of acetic anhydride was added. After 10 min, an additional 12 μl of acetic anhydride was added and the reaction allowed to proceed at room temperature for a further 10 min. This addition and incubation was repeated twice more. The reaction mixture was adjusted to approximately pH 6 by dropwise addition of 1 M-acetic acid, and then applied to a 1 ml column of Dowex AG50 × 12 (H^+ form, Bio-Rad). The column was eluted with 5 column volumes of distilled water. The eluate was filtered through a 0·5 μm Teflon filter and evaporated to dryness. Borate was removed by repeated evaporation (5 ×) with methanol (300 μl). The sample was then subjected to descending paper chromatography for 48 h with butan-1-ol/ethanol/water (4 : 1 : 1, by vol).

For quantification of the number of O-linked chains on $C\bar{1}$ inhibitor, lactose was added prior to the alkaline borohydride treatment. After isolation of the reduced, radio-labelled oligosaccharides from paper chromatography, they were treated with neuraminidase and subjected to high-voltage electrophoresis in pyridine/acetic acid/water (3 : 1 : 378, by vol.) buffer, pH 5·4 at 80 V cm^{-1} for 45 min. The neutral oligosaccharides remaining at the origin were eluted with water and applied to a Bio-Gel P-4 high-resolution gel filtration system. Radioactive fractions eluting from 1 to 5 glucose units, by comparison with the internal isomaltooligosaccharide standards, were pooled together. Lactitol was separated from reduced O-glycans by high-voltage electrophoresis in borate buffer (Ashford *et al.*, 1987). The radioactive areas were eluted from the electrophoretogram and quantified by liquid scintillation counting. A portion of the Bio-Gel P-4 pool was also subjected to the procedure for reducing-terminal monosaccharide determination (Ashford *et al.*, 1987). The radioactive areas of the resulting electrophoretogram corresponding to glucitol and *N*-acetylgalactosaminitol were quantified by integration of the radioactive peaks detected by the linear analyser. Corrections in the quantification were made for trace contaminants in the lactitol and for background.

(c) *Neutron data collection and analyses*

Neutron data were collected on Instrument D17 at the Institut-Laue-Langevin, Grenoble. Guinier radius of gyration R_G data were based on a sample to detector distance of 3·46 m and neutron wavelengths of 1·385 to 1·395 nm or 1·600 nm, corresponding to a Q range ($Q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle) of 0·033 to 0·486 nm^{-1}. Data at larger Q were obtained with a sample to detector distance of 1·40 m, wavelengths of 1·001 to 1·004 nm, and main beam to detector angles of 0° and 19·89° to give Q ranges of 0·13 to 1·60 nm^{-1} and 0·8 to 3·6 nm^{-1}. Instrument D11 with a sample to detector distance of 10·5 m and wavelength of 1·00 nm was used to obtain a lower Q range of 0·0019 to 0·220 nm^{-1}. Samples were dialysed at 6 °C with stirring into 12 mM-sodium phosphate, 200 mM-NaCl, 1 mM-EDTA (pH 7·0 in 0%, 80% or 100% ^2H_2O solutions with 4 changes over 36 to 48 h). All neutron data were recorded at 20 °C. Concentrations were measured using an absorption coefficient $A^{1\%}_{1\,cm,280}$ of 3·6 (Harrison, 1983; Salvesen *et al.*, 1985). Data reduction was based on standard Grenoble software (Ghosh, 1981), and the final analyses in London were based on SCTPL (Perkins & Sim, 1986). Statistical analyses were performed using MINITAB (version 6.1) on a microcomputer (Ryan *et al.*, 1985).

At small Q, the Guinier equation gives the radius of gyration R_G and the forward scattering at zero scattering angle $I(0)$ (Guinier & Fournet, 1955):

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

In a given solute–solvent contrast, R_G measures the degree of elongation of a glycoprotein. If this structure is sufficiently elongated, the R_G of the cross-sectional structure R_{XS} and the cross-sectional intensity at zero angle $[I(Q) \times Q]_{Q\to 0}$ are obtained from:

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

The matchpoint is obtained from a graph of $\sqrt{I(0)/cT_s t}$ against % ^2H_2O (T_s is sample transmission, t is sample thickness, c is concentration), and from the cross-sectional analyses using $\sqrt{[I(Q) \times Q]_{Q\to 0}/cT_s t}$. Experimental matchpoints can be compared with the amino acid sequence on the basis of the unhydrated shape of α_1-AT, the use of crystallographic volumes (unhydrated), and the 10% non-exchange of the main-chain amide protons (Perkins, 1986). The contrast variation analysis of the arrangement of carbohydrate and protein is based on the Stuhrmann equation (Ibel & Stuhrmann, 1975):

$$R_G^2 = R_{G\text{-}C}^2 + \alpha_G \times \Delta\rho^{-1} - \beta_G \times \Delta\rho^{-2}$$

$$R_{XS}^2 = R_{XS\text{-}C}^2 + \alpha_{XS} \times \Delta\rho^{-1} - \beta_{XS} \times \Delta\rho^{-2},$$

where $R_{G\text{-}C}$ and $R_{XS\text{-}C}$ are the radii of gyration of the macromolecule and its cross-section at infinite contrast, α_G and α_{XS} measure the corresponding radial inhomogeneity of scattering densities, and $\Delta\rho^{-1}$ is the reciprocal solvent–solute contrast difference. The terms in β_G and β_{XS} measure the displacement of the centre of scattering density as the contrast is varied. From the term in β, the distance Δ between the centres of 2 components of distinct scattering densities ρ_1 and ρ_2 and volumes V_1 and V_2, respectively, can be calculated; however, 3 parameter-weighted least-squares fitting of the data in Fig. 3(b) showed that this was not measurable.

The neutron-scattering curve $I(Q)$ in reciprocal space can be transformed into real space $P(r)$ by use of the indirect transformation procedure (ITP) method (Glatter, 1982). This required the full scattering curve to 3·6 nm^{-1}. It offers an alternative calculation of the R_G and the length L for $C\bar{1}$ inhibitor. The 75 experimental $I(Q)$ data points (in 3 subdomains to correspond to the 3 D17 configurations) were best fitted using 10 to 15 cubic splines, which were transformed into 101 points of the

$P(r)$ function based on a maximum length of 20 nm. Desmearing of the neutron curve was performed on the basis of a D17 wavelength spread of 10% full-width to half-maximum. The slit width and length corrections were set to be equal and based on a beam divergence of 0·0112 rad at Instrument D17.

(d) *Modelling of the structure of C1 inhibitor*

Modelling of the neutron-scattering curves was based on the dry volume of the glycoprotein (Chothia, 1975; Perkins, 1986), and used Debye simulations based on spheres of 2 distinct scattering densities to follow described procedures (Perkins & Weiss, 1983; Perkins, 1985; Smith *et al.*, 1990). Modelling of the sedimentation coefficient $s^0_{20,w}$ was based on the hydrated volume, which corresponds to the sum of the volumes of the dry glycoprotein and the hydration shell (assuming a hydration of 0·3 g H_2O/g glycoprotein and an electrostricted water molecule volume of 0.0245 nm^3), and this leads to a partial specific volume $\bar{v}$ of 0·721 ml/g for C1 inhibitor (Perkins, 1986). When hydrodynamic spheres are used, this hydration is increased to 0·39 g in order to compensate for the void spaces between the non-overlapping spheres (Perkins, 1989*b*). Calculations were performed using the program GENDIA (Garcia de la Torre & Bloomfield, 1977*a*, *b*).

3. Results and Discussion

(a) *Carbohydrate analyses of C1 inhibitor*

Carbohydrate analyses were carried out to confirm the structure(s) of the N-linked oligosaccharides and to quantify the number of O-linked oligosaccharides bound to C1 inhibitor. While the protein sequence (Bock *et al.*, 1986) indicates six occupied N-linked glycosylation sites, the situation for the O-linked chains is less clear.

High-voltage electrophoresis of the N-linked oligosaccharides of C1 inhibitor, released by hydrazinolysis and reduced with NaB^3H_4, indicated that 96·6% of the oligosaccharides carried at least one negative charge (Fig. 1(a)). The distribution of charged species was consistent with a mixture of mono- and di-sialylated biantennary oligosaccharides (A-1 and A-2) and mono-, di- and trisialylated triantennary oligosaccharides (A′-1, A′-2 and A-3). The relative proportions of these components were: A-1, 14·9%; A-2, 58·1%; A′-1, 2·1%; A′-2, 5·1%; A-3, 16·4%. Electrophoresis of the charged N-linked oligosaccharides after treatment with neuraminidase showed that all the charged components were converted to neutral species and remained at the origin. Therefore the only charged moiety was sialic acid. Bio-Gel P-4 gel filtration chromatography of the neuraminidase-treated oligosaccharides gave the elution profile shown in Figure 1(b). The major peak eluted at 13·5 glucose units and minor components eluted at 14·5, 16·5, 17·3 and 18·5 glucose units. These peaks were pooled as shown in Figure 1(b). The relative proportions of each pool were as follows: I, 2·0%; II, 6·5%; III, 9·3%; IV, 20·6%; V, 61·6%. The characteristic elution profile positions on Bio-Gel P-4 (Yamashita *et al.*, 1982) indicated that peak *a* was a tetra-antennary oligosaccharide, peaks *b* and *c* were tri-antennary oligosaccharides (peak *b* core fucosylated), and peaks *d* and *e* were bi-antennary oligosaccharides (peak *d* core fucosylated). These oligosaccharide assignments were confirmed using sequential exoglycosidase sequencing (data not shown).

High-voltage electrophoresis of a portion of the radioactive oligosaccharides obtained by treatment of C1 inhibitor with alkaline sodium [^{3}H]borohydride in the presence of lactose, showed negatively charged material with a similar mobility to sialyl-lactose and authentic NeuNAc2-3(6)Galβ1-3GalNAc in addition to lactitol. When the charged material was treated with neuraminidase and resubjected to electrophoresis, the radioactivity remained at the origin and was therefore neutral, confirming that only sialylated charged species were present (data not shown). The bulk of the radioactive oligosaccharides were then treated with neuraminidase, subjected to electrophoresis and the resulting neutral oligosaccharides were subjected to Bio-Gel P-4 gel filtration. The chromatogram showed a series of peaks eluting at 3·5 and 2·5 glucose units and a minor peak eluting at 4·4 glucose units (data not shown). The peak at 2·5 glucose units corresponds to lactitol and that at 3·5 to Galβ1-3GalNAc$_{ol}$. Insignificant amounts of N-linked oligosaccharides were present. Acid hydrolysis of the pooled fractions from one to five glucose units confirmed that the only reducing-terminal monosaccharides present were *N*-acetylgalactosaminitol and glucitol. These data taken together are consistent with the finding of Strecker *et al.* (1985) that the O-linked oligosaccharides of C1 inhibitor are predominantly NeuNAc2-3(6)Galβ1-3GalNAc.

Inclusion of lactose in the alkaline borohydride treatment allowed measurement of the ratio of this standard to the released O-linked sugars (see Table 1). Two methods of quantification both gave a value of approximately seven O-linked oligosaccharides per molecule of C1 inhibitor.

The addition of six N-linked and seven O-linked oligosaccharides with the structures determined above to the peptide M_r of 52,800 leads to an M_r of 71,000 (Table 2), of which 26% is carbohydrate. This content is less than, but is comparable with, the determinations of 33% to 35% by Haupt *et al.* (1970) and Harrison (1983). The presence of seven O-linked oligosaccharides is consistent with the seven glycosylation sites reported at Ser42, Thr26, Thr49, Thr61, Thr66, Thr70 and Thr74 during sequencing (Bock *et al.*, 1986). It is also clear that the total M_r is 32% less than the generally accepted value of 104,000. Most M_r determinations have been based on SDS/polyacrylamide gel electrophoresis, and values ranging from 98,000 to 116,000 have been reported (Pensky *et al.*, 1961; Haupt *et al.*, 1970; Harpel & Cooper, 1975; Reboul *et al.*, 1977; Nilsson & Wiman, 1982; Harrison, 1983). It is well known that the presence of carbohydrate causes glycoproteins to migrate abnormally in SDS/

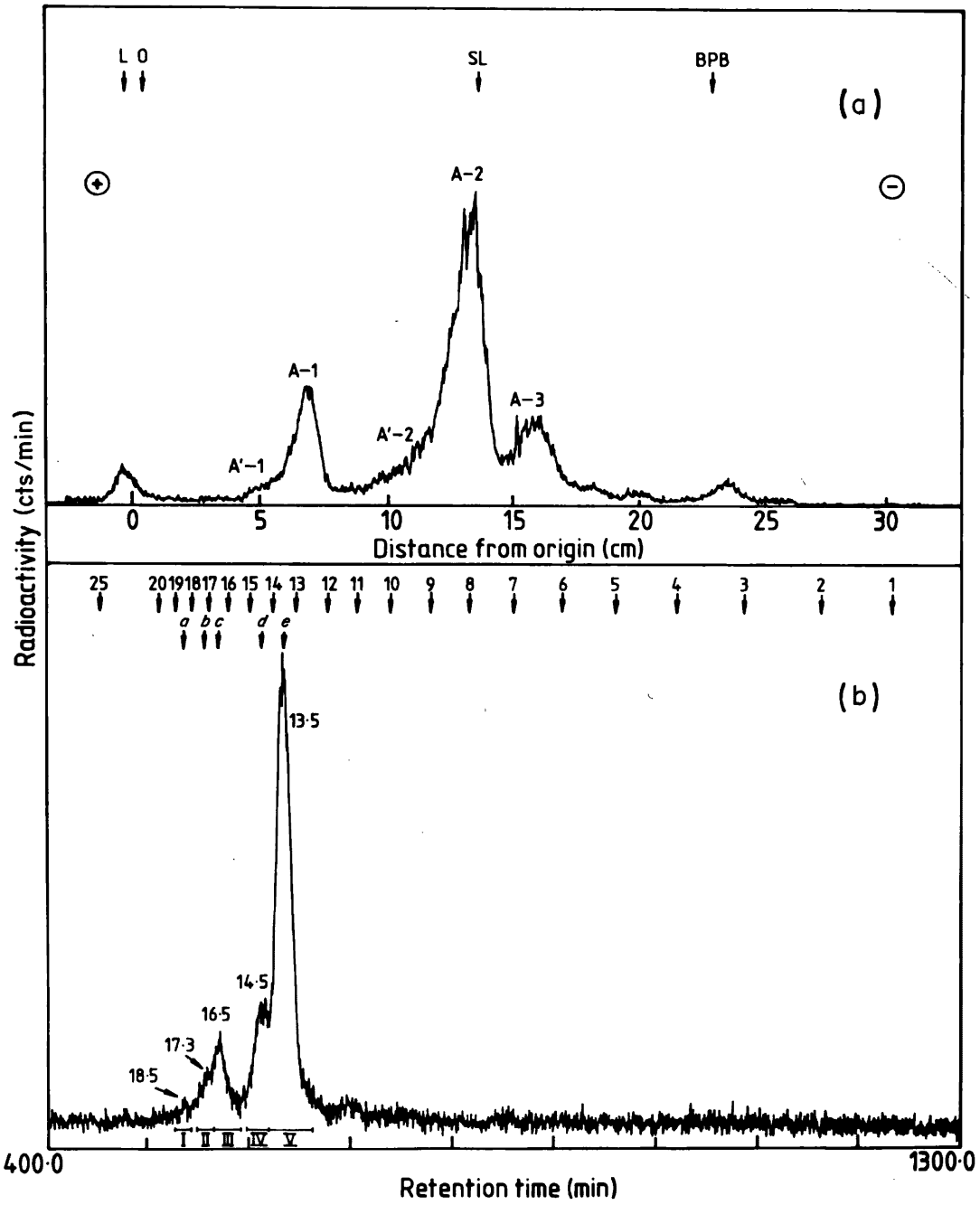


Figure 1. N-linked oligosaccharide analysis of CĪ inhibitor. (a) High-voltage paper electrophoresis of CĪ inhibitor N-linked oligosaccharides. Tritium-labelled oligosaccharides were subjected to high-voltage paper electrophoresis (80 V cm^{-1}) in pyridine/acetic acid/water (3 : 1 : 387, by vol.) buffer (pH 5·4). The arrows indicate the positions of lactitol (L), sialyl-lactitol (SL) and bromophenol blue (BPB) markers. Peaks A-1 and A-2 represent mono- and di-sialylated bi-antennary oligosaccharides. Peaks A'-1, A'-2 and A-3 represent the mono-, di- and tri-silylated oligosaccharides, respectively. O, origin. (b) Gel filtration of CĪ inhibitor desialylated N-linked oligosaccharides. Negatively charged tritium-labelled oligosaccharides from high-voltage electrophoresis were exhaustively digested with neuraminidase then separated by high-resolution gel filtration on Bio-Gel P-4. The arrows indicate the elution position of isomalto-oligosaccharides containing the corresponding number of glucose units. The numbers over each peak indicate the elution position of that peak in glucose units. The time axis is marked at 100 min intervals and the bars indicate the areas that were pooled. Pools I, III and V correspond to the elution positions of tetra-antennary, tri-antennary and bi-antennary oligosaccharides respectively. Pools II and IV correspond to the elution positions of fucosylated tri-antennary and bi-antennary oligosaccharides respectively.

polyacrylamide gel electrophoresis and therefore give apparently higher M_r values (Gordon, 1975). This could account for the discrepancy with the SDS/polyacrylamide gel results. However, SDS/polyacrylamide gel electrophoresis of deglycosylated CĪ inhibitor (Harrison, 1983) or the product of cell-free-translation of CĪ inhibitor mRNA (Tosi *et al.*, 1986; Reboul *et al.*, 1987) gives apparent M_r

Table 1
Quantification of the number of O-linked oligosaccharides released from C1 inhibitor by alkaline sodium borohydride

	Radioactivity† (cts/min)	Radioactivity/ mole (cts/min)	Molar ratio
Oligosaccharides	9.52×10^4	2.74×10^{12}§	6.03 (72,000)
		2.89×10^{12}	6.36 (76,000)
		3.05×10^{12}	6.70 (80,000)
Lactitol	2.85×10^5	4.55×10^{11}	1.00
	Area (μV × min)‡	Area/mole (μV × min)	Molar ratio
N-acetylgalcatos-aminitol	8.02×10^2	2.31×10^{10}§	6.35 (72,000)
		2.44×10^{10}	6.74 (76,000)
		2.57×10^{10}	7.10 (80,000)
Glucitol	1.13×10^3	1.81×10^9	1.00

† After separation of the oligosaccharides by borate electrophoresis, the radioactive areas were eluted and the radioactivity measured by liquid scintillation counting.

‡ After acid hydrolysis and separation of the reducing terminal monosaccharides by borate electrophoresis, the radioactive peak areas were obtained by integration.

§ The number of mol of O-linked glycan/mol of C1 inhibitor are calculated for different molecular weights of C1 inhibitor and are given in parentheses.

values of 78,000, 64,000 and 65,000 in that order. These are higher than the protein mass of 52,800 calculated from the sequence, and may reflect restricted SDS binding and unfolding of the unusual repeating tetrapeptide sequence Glx-Pro-Thr-Thr in the N-terminal domain of the protein.

(b) *Neutron* R_G *Guinier analyses of C1 inhibitor*

Neutron scattering was applied to both native and split C1 inhibitor. The split form is inactive, and differs from the native form by two main-chain breaks, either between Met31 and Leu32 or Leu32 and Phe33, and Ser441 and Val442; the last pair is in the reactive site loop, close to the physiological cleavage site at Arg444-Thr445 (Pemberton *et al.*, 1989). While the C-terminal peptide (M_r 4400) remains associated with the major C1 inhibitor fragment, a partial loss of up to 30% of those at the N terminus (M_r 3200) has probably occurred (see Materials and Methods).

The neutron Guinier analyses on C1 inhibitor were based on three independent preparations of each of the native and split forms studied between concentrations of 2 and 14 mg/ml. All experiments showed biphasic Guinier R_G plots (Fig. 2(a)). The R_G could be analysed in the lowest linear Q^2 range in Figure 2(a) between Q values 0.08 and 0.22 nm^{-1} on Instrument D17, and the data were confirmed using Instrument D11 at lower Q (see Materials and Methods). The R_G and $I(0)/c$ values increase by approximately 15% and 30%, in that order, on dilution from 14 mg/ml in all contrasts, and this is similar to that observed for another serpin, α_1-antitrypsin (Fig. 2 of Smith *et al.*, 1990). A concen-

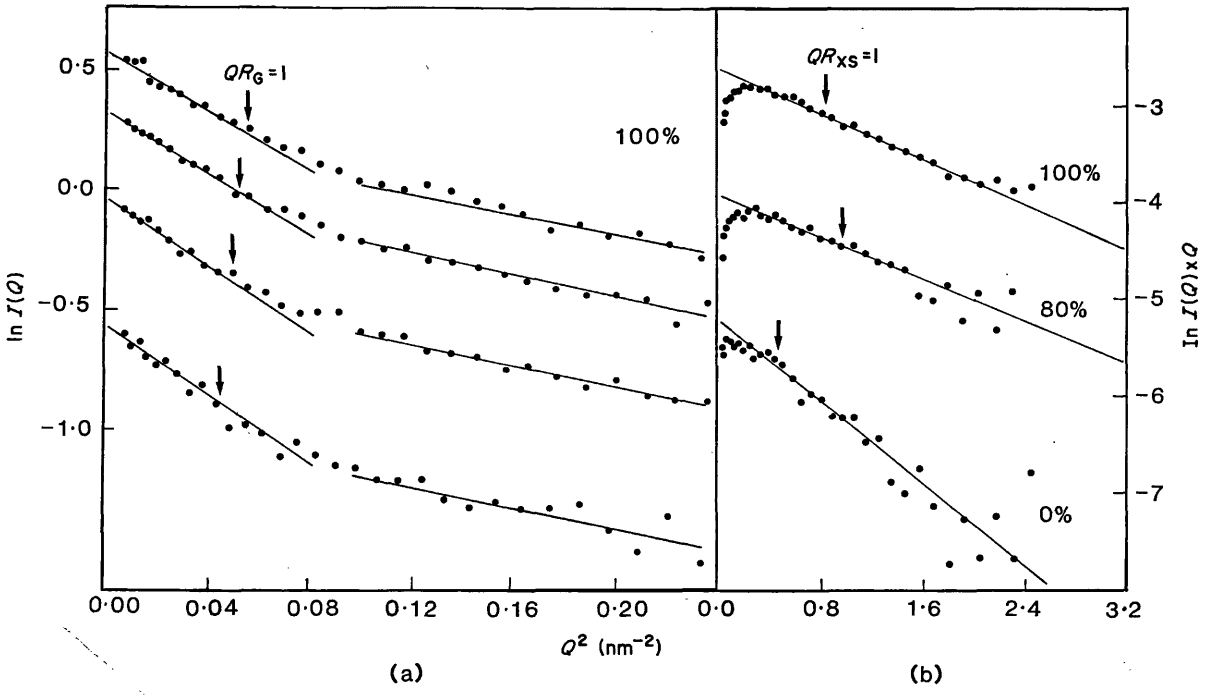


Figure 2. Neutron Guinier analyses for native C1 inhibitor. (a) A dilution series in 100% ^2H_2O buffers is depicted. At low Q, the concentration dependence of the Guinier R_G curves is shown. From top to bottom, the sample concentrations are 7.9 mg/ml, 5.9 mg/ml, 4.1 mg/ml and 2.2 mg/ml, with R_G values of 4.3 nm, 4.4 nm, 4 5 nm and 4.7 nm (errors between ±0.1 and ±0.4 nm: Ghosh, 1981). The condition $QR_G = 1$ is arrowed in each curve. (b) Cross-sectional R_{XS} analyses are shown in 0%, 80% and 100% ^2H_2O buffers. The sample concentrations are 8.5 mg/ml, 12.7 mg/ml and 14.0 mg/ml, in that order. R_{XS} errors are ±0.1 nm and the condition $QR_{XS} = 1$ arrowed.

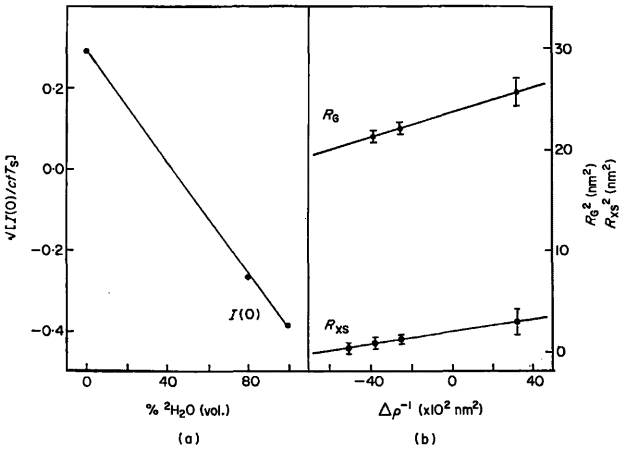


Figure 3. Contrast variation analyses for native and split C1 inhibitor. (a) The matchpoint determination of C1 inhibitor, using the $I(0)/cT_s$ data extrapolated to zero concentration, gives 42·3% ^2H_2O. The corresponding matchpoint determination using the mean $[I(Q) \times Q]_{Q \to 0}$ values measured in 0%, 70%, 80% and 100% buffers (not shown) gives 42·2% ^2H_2O. (b) The Stuhrmann plot of R_G^2 against $\Delta\rho^{-1}$ gives an $R_{G\text{-}C}$ of 4·85(±0·01) nm and a slope α_G of $59(\pm 3) \times 10^{-5}$ on the basis of a weighted 2-parameter least-squares fit. The R_G data had been extrapolated to zero concentration. Analysis of the mean R_{XS}^2 values in 4 contrasts gives an $R_{XS\text{-}C}$ of 1·43(±0·02) nm and an α_{XS} of $29(\pm 2) \times 10^{-5}$. Error bars in (a) and (b) are shown only when large enough to be seen.

tration dependence was also reported for C1 inhibitor by sedimentation (Haupt *et al.*, 1970). Since the serpin domain of C1 inhibitor has a net charge of +6, while the N-terminal domain has a net charge of −20, it is likely that the molecule contains a dipole and that the concentration dependence is the result of interparticle interference phenomena (Guinier & Fournet, 1955).

$I(0)/c$ provides compositional information on M_r and is a control of the measurements. For the native and split forms, their values were the same within error in all contrasts. For the split form, it is inferred that both the C-terminal cleaved peptide and the bulk of the N-terminal peptides (up to 9% of the total M_r) remain associated with the parent structure, in agreement with sequence analyses (see Materials and Methods). Since the C-terminal peptide contains a Trp residue, significantly altered calculations of c from optical densities would have resulted had this peptide been lost during dialysis. Absolute M_r calculations based on $I(0)/c$ data in H_2O extrapolated to zero c gave an M_r of 76,000 (±4000) (Jacrot & Zaccai, 1981). This is consistent with an M_r of 71,100 from the sequence. The M_r calculation is based on an absorption coefficient $A^{1\%}_{1cm,280}$ of 3·6 (Harrison, 1983; Salvesen *et al.*, 1985), which is supported by a calculated A_{280} of 3·86 from the composition (Wetlaufer, 1962; Perkins, 1986). An alternative A_{280} value of 4·5 (Haupt *et al.*, 1970; Bruch *et al.*, 1988) would have given an M_r of 95,000 (±5000). Figure 3(a) shows the linear dependence of $\sqrt{I(0)/c}$ on the contrast. The matchpoint is 42·3±0·5% ^2H_2O, which is compatible with the value of 41·0% ^2H_2O calculated from the composition (Perkins, 1986).

R_G provides overall structural data. For the two forms of C1 inhibitor, the R_G values were similar (and also for the R_{XS} data, see below). This shows that the structural rearrangement between the two forms is relatively localized, as already discussed for α_1-antitrypsin (Smith *et al.*, 1990), even though crystallographic studies show that a significant conformational change must have occurred on cleavage of the reactive centre (Löbermann *et al.*, 1984). Evidence that this change occurs too in C1 inhibitor was demonstrated by circular dichroism (Bruch *et al.*, 1988) and heat stability (Pemberton *et al.*, 1989).

The Stuhrmann dependence of R_G on the contrast (Fig. 3(b)) provides information on the external and internal structure of C1 inhibitor. At infinite contrast the $R_{G\text{-}C}$ of 4·85(±0·01) nm corresponds to an elongation ratio $R_{G\text{-}C}/R_0$ of 2·27, where R_0 is the R_G of the sphere with the same dry volume as C1 inhibitor. The value of this ratio is 1·35 for the homologous serpin α_1-antitrypsin (Smith *et al.*, 1990). C1 inhibitor is therefore much more elongated than α_1-antitrypsin. The difference is attributed to the large N-terminal domain of C1 inhibitor, the analogue of which is not found in the sequence of α_1-antitrypsin. The positive slope α_G of the Stuhrmann plot shows that the internal structure of C1 inhibitor is dominated by a higher scattering density on its surface than in its inner core. Calculations show that this corresponds to hydrophilic surface regions of protein and carbohydrate and a hydrophobic core of protein (Perkins *et al.*, 1981; Perkins, 1986). Even though α_G is less accurately determined than $R_{G\text{-}C}$, the α_G of $59(\pm 3) \times 10^{-5}$ is comparable with those measured for α_1 acid glycoprotein (Perkins *et al.*, 1985) and α_1-antitrypsin (Smith *et al.*, 1990) after rescaling these data on the basis that α is proportional to R_C^2.

(c) *Neutron analyses of C1 inhibitor at large* Q

Since C1 inhibitor is elongated, cross-sectional R_{XS} analyses could be performed (Pilz, 1982). Linear plots of $\ln(I(Q) \times Q)$ *versus* Q^2 were found in the Q range 0·67 to 1·30 nm^{-1} (Fig. 2(b)). No concentration dependence of the R_{XS} or $[I(Q) \times Q]_{Q \to 0}$ parameters was observed, nor was any difference seen between the native and split forms of C1 inhibitor (data not shown). Contrast variation gave a matchpoint of 42·2(±0·5)% ^2H_2O, which is consistent with the predicted value of 41·0% ^2H_2O (Table 2). The Stuhrmann plot of R_{XS}^2 *versus* $\Delta\rho^{-1}$ gave a $R_{XS\text{-}C}$ of 1·43(±0·02) nm and an α_{XS} of $29(\pm 2) \times 10^{-5}$ (Fig. 3(b)). The predicted value of α_{XS} is in the range 3×10^{-5} to 13×10^{-5}, which is considerably less than the observed α_{XS} of 29×10^{-5}. (Note that the contrast dependence of R_{XS} is clearly visible in Fig. 2(b).) The cross-sectional structure of C1 inhibitor is thus strongly dominated by the carbohydrate component.

Table 2
Properties of the native and split forms of human CĪ inhibitor

A. *Compositional*		
M_r (protein)	52,800	(478 residues)
(carbohydrate)	18,300 (26% weight)	(90 residues)
(total)	71,100	(568 residues)
M_r (neutron scattering)	76,000±4000	
Dry volume (nm^3)	88·2	
$\Sigma b/M_r$ in H_2O (fm)	0·2192	
Matchpoint (%^2H_2O)		
(predicted)	41·0 (protein 39·5; CHO 47·1)	
(experimental)	42·3 42·2	
$A^{1\%}_{1\,cm,280}$		
(calculated)	3·86	
(experimental)	3·6	
B. *Structural*		
$R_{G\text{-}C}$ (nm)	4·85±0·01	
$R_{XS\text{-}C}$ (nm)	1·43±0·03	
α_G ($\times 10^{-5}$)	59±3	
α_{XS} ($\times 10^{-5}$)	29±2	
β_G ($\times 10^{-14}$ nm^{-2})	360±30	
L (nm)	15·8±0·5 (R_G and $I(0)$ data)	
	16–19 (ITP analysis)	
$R_{G\text{-}C}/R_0$	2·27	
$R_{XS\text{-}C}/R_0$	1·52	
C. *Models*		
$R_{G\text{-}C}$ (nm)	4·87	
α_G ($\times 10^{-5}$)	54	

The dry volume is calculated from the crystallographic residue volumes of Chothia (1975) and Perkins (1986). The matchpoint is calculated assuming 10% of non-exchange of the peptide NH protons (Perkins, 1986). The N-linked carbohydrate composition is calculated as $NeuNAc_{12}Gal_{13}GalNAc_{25}Man_{18}Fuc_1$, and that of the O-linked oligosaccharides as $NeuNAc_7Gal_7GalNAc_7$.

The scattering data lead to estimates of the length L of the particle. From the Guinier analyses (Perkins *et al.*, 1990), the intensities ratio $\pi \times I(0)/[I(Q) \times Q]_{Q\to 0}$ in each of 0%, 80% and 100% ^2H_2O leads to L values of 15·7 nm, 16·1 nm and 15·0 nm, in that order. From the R_G and R_{XS} values, L values of 16·5 nm, 15·4 nm and 15·8 nm were obtained from $L^2 = 12\ (R_G^2 - R_{XS}^2)$. The mean value of L is 15·8(±0·5) nm (Table 2). The indirect transformation of the scattering curve $I(Q)$ into the distance distribution function $P(r)$ in real space also gives L, since $P(r)$ becomes zero at the maximum distance within the macromolecule (Glatter, 1982). This has the advantage that the full curve to $Q = 3{\cdot}1\ nm^{-1}$ is used, even though intensity errors are greater at large Q when $I(Q)$ is small. Data for both the native and the split forms in 100% ^2H_2O were analysed. The most satisfactory $P(r)$ calculations showed a strong peak at r of 2·4 nm (Fig. 4), and L was determined as 16 to 19 nm, which is compatible with the Guinier estimate of L as 15·8(±0·5) nm.

The scattering curves between the R_G and R_{XS} regions (Q of 0·22 to 0·67 nm^{-1}) have not yet been considered. Even though Figure 2(a) suggests that linear Guinier plots could be obtained in this range, this part of the scattering curve consists of three contributions, one from the serpin domain, one from the N-terminal domain, and one from the scattering interference term of both domains. Even though putative Guinier analyses suggest that the mean $I(0)/c$ value in H_2O corresponds to an M_r of 46,000 (±6000) (which is close to the M_r of 47,700 for the serpin domain of CĪ inhibitor) and the $R_{G\text{-}C}$ of 2·78(±0·03) nm is close to the $R_{G\text{-}C}$ of 2·61±0·02 nm for α_1-antitrypsin (Smith *et al.*, 1990), the deconvolution of the different contributions is not straightforward and requires information about the spatial arrangement of the two domains, which is not available *a priori*.

(d) *Debye sphere modelling of CĪ inhibitor*

Molecular models were constructed to interpret the scattering curve analyses. The crystal structure of split α_1-antitrypsin was represented by Debye spheres (Löbermann *et al.*, 1984; Smith *et al.*, 1990). The secondary structure of α_1-antitrypsin can be fully matched with the CĪ inhibitor sequence if Asn83 to Ile92 in α_1-antitrypsin are deleted (Harrison, 1989). This corresponds to part of helix D and a surface loop between helices C and D, and includes the oligosaccharide site at Asn83. After removing these residues, the crystal structure was converted into 237 spheres of diameter 0·754 nm based on cubes of side 0·608 nm.

Gly1 to Asp19 in the α_1-antitrypsin crystal structure are replaced by the N-terminal domain (Asn1 to Ser113) in CĪ inhibitor. The N-terminal domain was modelled as an extended protein structure of 69 spheres, initially with a total length of 23 spheres and 14·0 nm, for four reasons.

(1) Scattering shows that the overall length of α_1-antitrypsin of 7·0 to 7·8 nm (Smith *et al.*, 1990) is increased to 16 to 19 nm in CĪ inhibitor.

(2) A Robson secondary structure prediction (Garnier *et al.*, 1978) of the N-terminal 113 residues indicates relatively low amounts of α-helix (17%) and β-sheet (16%), and high amounts of β-turn (25%) and coil (41%).

(3) Of the 113 residues, 61% are hydrophilic and there are ten glycosylation sites, which is compatible with a high exposure to solvent.

(4) Of the 113 residues, 15 are proline, mostly between Pro62 to Pro104, which is high compared to an average of 5·2% in proteins (Dayhoff, 1978).

The six N-linked oligosaccharides were modelled as extended bi-antennary structures (Strecker *et al.*, 1985) of 11 spheres each, even though Figure 1 shows that one-sixth of the structures are tri-antennary oligosaccharides, and that possibly one site is occupied uniquely with this tri-antennary class. The seven O-linked oligosaccharides were represented by three spheres each. In the N-terminal domain the oligosaccharides were arbitrarily positioned along the protein core according to their occurrence in the primary sequence (Fig. 5).

Information on the relative orientation of the serpin and N-terminal domains in CĪ inhibitor was provided by the position of Cys406 and Cys183 of CĪ inhibitor in the crystal structure of α_1-antitrypsin. These form disulphide bridges with Cys101 and Cys108 in the N-terminal domain. Cys406 and

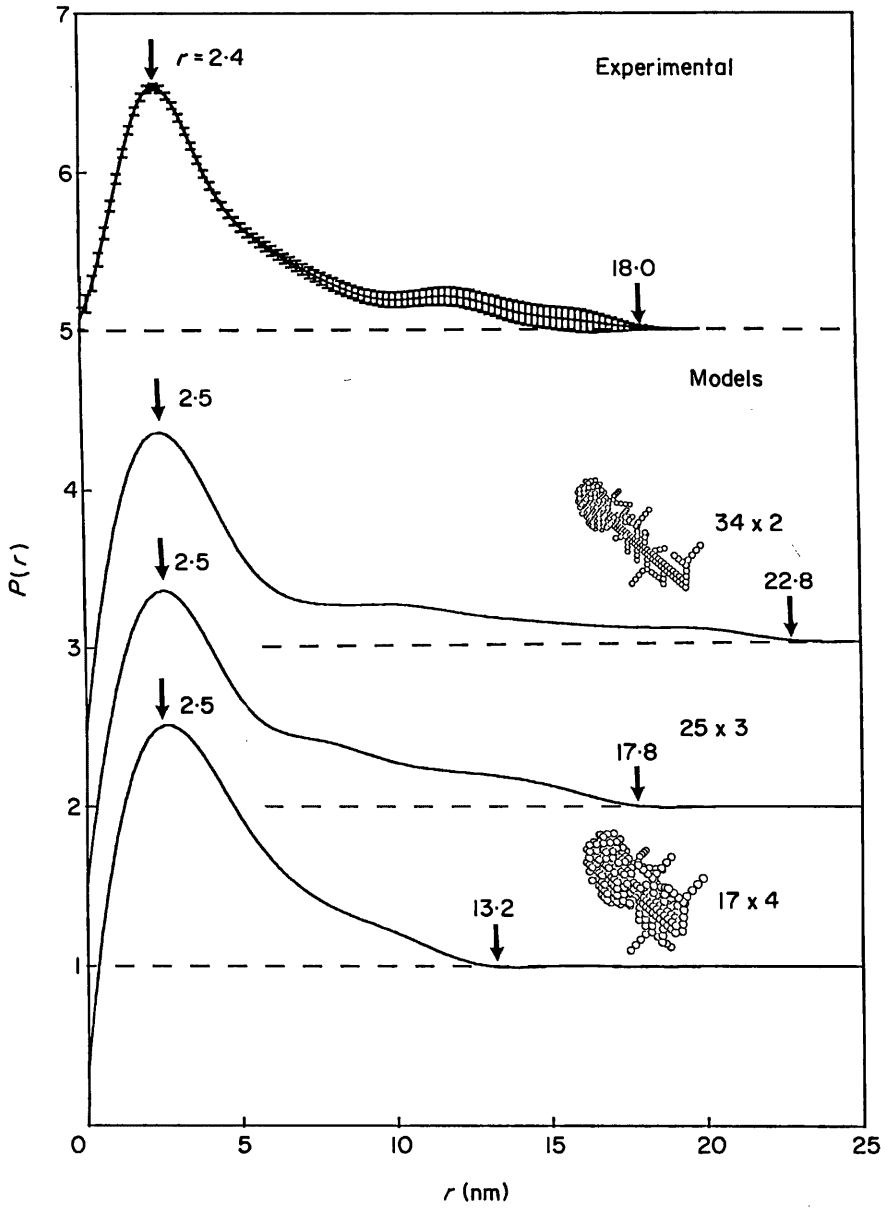


Figure 4. Indirect transformation (Glatter, 1982) of the scattering curves $I(Q)$ for $C\bar{1}$ inhibitor. The experimental $P(r)$ curve corresponds to data on native $C\bar{1}$ inhibitor at 14 mg/ml in 100% ^2H_2O, based on 69 $I(Q)$ points extending to $Q = 3{\cdot}1\ nm^{-1}$, fitted using 10 B-splines, and transformed using an assumed maximum length of 20·2 nm. The error margins of the transformation are shown for the experimental curve. The 25×3 model curve is based on the model shown in Fig. 5, as calculated for the 100% ^2H_2O contrast, based on 99 $I(Q)$ points extending to $Q = 1{\cdot}6\ nm^{-1}$, fitted using 10 B-splines, and transformed using a maximum length of 25·25 nm. The r values corresponding to the positions of the main peak in $P(r)$ and where $P(r)$ becomes zero are arrowed. The 34×2 and 17×4 models are inset.

Cys183 are located on the $-z$ face of the crystal structure, on helix D and on the loop connecting helix I and strand A5, and constrain the N terminus domain to lie along the $+x$ direction of the crystal structure starting from His20 (the $+x$ model). The long axes of the two domains lie in similar directions.

The contrast dependence of the scattering was incorporated by subdividing the model into 243 surface hydrophilic spheres and 152 core hydrophobic spheres, which were assigned matchpoints of 60% and 30% ^2H_2O, respectively (Smith *et al.*, 1990). Figure 5 shows that good agreements with the experimental curves in three contrasts were obtained. The residual R (defined by Smith *et al.* (1990), by analogy with crystallography) is low at 0·020 (0% ^2H_2O), 0·011 (80%) and 0·012 (100%). The R_G in the three contrasts are 5·03 nm, 4·61 nm and 4·71 nm, in good agreement with the experimental values of 5·05 nm, 4·60 nm and 4·71 nm (Fig. 3(b)). The modelled $R_{G\text{-}C}$ and α_G of 4·87 nm and 54×10^{-5} agree well with the experimental values of 4·85 nm and 59×10^{-5}. Transformation of $I(Q)$ into $P(r)$ gave a maximum dimension of 17·8 nm for the 100% ^2H_2O model, together with a major peak at 2·5 nm, in good agreement with the

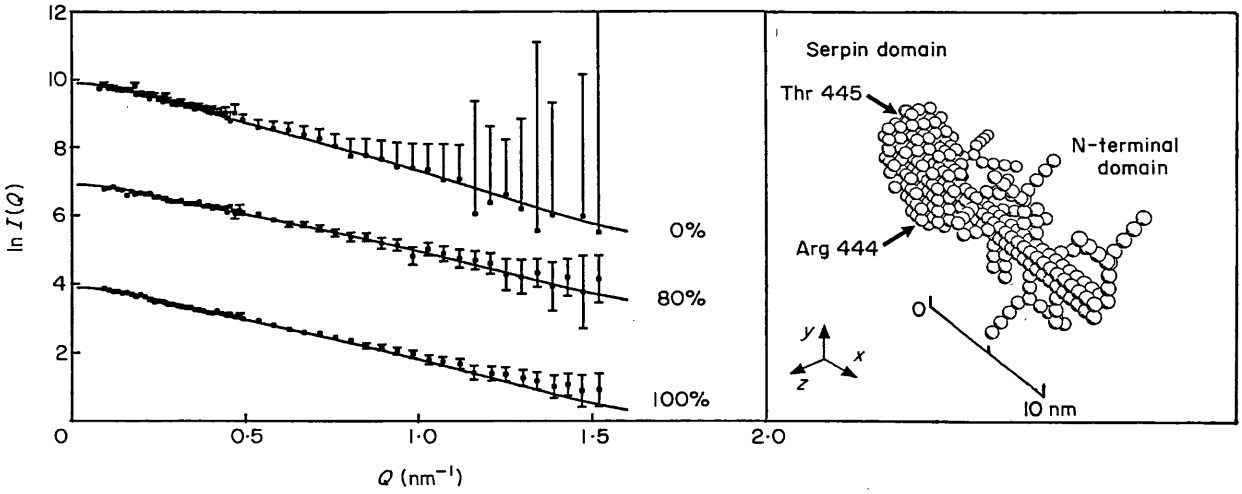


Figure 5. Curve fitting for the Debye sphere models of CI inhibitor. The 395-sphere $+x$ model (shown on the right) leads to simulations that account for the scattering curves of native CI inhibitor in 0%, 80% and 100% ^2H_2O (see the text). Statistical errors in the neutron data are indicated by full (80%, 100% ^2H_2O) or half (0% ^2H_2O) error bars. The N-terminal domain is 23 spheres in length and 3 spheres in cross-section. Since 6 spheres in the N-terminal domain overlapped with the serpin domain in this orientation, and had to be deleted, the N-terminal domain was increased in length from 23 to 25 spheres to conserve the total volume. The sphere model is shown in the same orientation as that of the crystal structure and the Debye scattering model for α_1-antitrypsin in Smith *et al.* (1990).

experimental $P(r)$ curve (Fig. 4). A two-domain structure for CI inhibitor therefore accounts for its scattering properties.

Other models for CI inhibitor were tested. Firstly, four alternative $+x$ models for the protein core were constructed, based on sphere lengths and cross-sections of 68×1, 34×2, 17×4 and 8×9, and lengths 41 nm, 21 nm, 10 nm and 5 nm. Their R_G values were 12·1 nm, 6·2 nm, 3·7 nm and 2·9 nm, in that order, and are clearly different from the experimental $R_{G\text{-}C}$ value of 4·85 nm. The $P(r)$ calculated for the 34×2 and 17×4 models resulted in maximum lengths of 22·8 nm and 13·2 nm, respectively, which clearly differ from the error margin of the experimental $P(r)$ in Figure 4, and show that these models can be excluded. The length of CI inhibitor in solution is therefore well identified by scattering. Finally, in other models, the N-terminal domain was repositioned relative to the serpin domain in the $-x$, $-y$ and $-z$ directions. Even when the R_G was adjusted to be close to 4·85 nm, slightly worse curve fits compared to those of Figure 5 were obtained. While the $+x$ direction is favoured by the modelling, this is not unambiguous.

(e) *Hydrodynamic simulations of CI inhibitor*

For CI inhibitor, the sedimentation coefficient $s^0_{20,w}$ has been reported as 3·67, 3·7, 3·8 and 4·3 S (Schultze *et al.*, 1962; Haupt *et al.*, 1970; Reboul *et al.*, 1977; Chesne *et al.*, 1982). This leads to a moderately high frictional ratio f/f_0 of 1·54, which shows that CI inhibitor is elongated in solution, as already inferred from electron microscopy (Odermatt *et al.*, 1981) and the present neutron data.

Given the M_r and the $s^0_{20,w}$, the length L of the structure is determined as the only unknown if this can be approximated by a rod-like cylinder (eqn (77) of Garcia de la Torre & Bloomfield, 1981). For an M_r of 71,100 and an $s^0_{20,w}$ of 3·7 S, L was found to be 26 nm. An alternative length calculation using eight hydrodynamic spheres of diameter 3·1 nm arranged in a straight line gave 25 nm (Garcia de la Torre & Bloomfield, 1981; Perkins, 1989*b*). Both are substantially less than the first estimates of 55 to 60 nm and 38 to 44 nm from hydrodynamic data also using the cylinder approach (Odermatt *et al.*, 1981; Perkins, 1985). This results from the decrease in M_r to 71,100 from the previously accepted value of 98,000 to 116,000. However, the revised length of 25 to 26 nm is still significantly greater than the neutron length of 16 to 19 nm. This suggests that the use of hydrodynamic cylinders is not satisfactory.

The hydrodynamic structure of α_1-antitrypsin has been successfully modelled (Smith *et al.*, 1990) using 32 spheres of diameter 1·69 nm for the protein and 24 spheres of diameter 0·95 nm for the oligosaccharide chains. This model was adapted to the CI inhibitor structure by (1) relocating the extended N-linked oligosaccharides found at Asn46 and Asn83 in α_1-antitrypsin to Glu132 and Gly147 to correspond to the positions of Asn216 and Asn231 in CI inhibitor; (2) removing one sphere to correspond to the Asn83–Ile93 loop in α_1-antitrypsin that is deleted in CI inhibitor; (3) adding seven spheres of diameter 1·69 nm to the N-terminal sphere of α_1-antitrypsin to give a $1 \times 1 \times 8$ array aligned in the $+x$ direction to represent the protein core of the N-terminal domain; (4) adding three extended N-linked and seven O-linked oligosaccharides to the N-terminal domain using spheres of diameter 0·95 nm. The 107 spheres (Fig. 6(a)) are of overall length 15 nm, and give a calculated $s^0_{20,w}$ of 3·5 S.

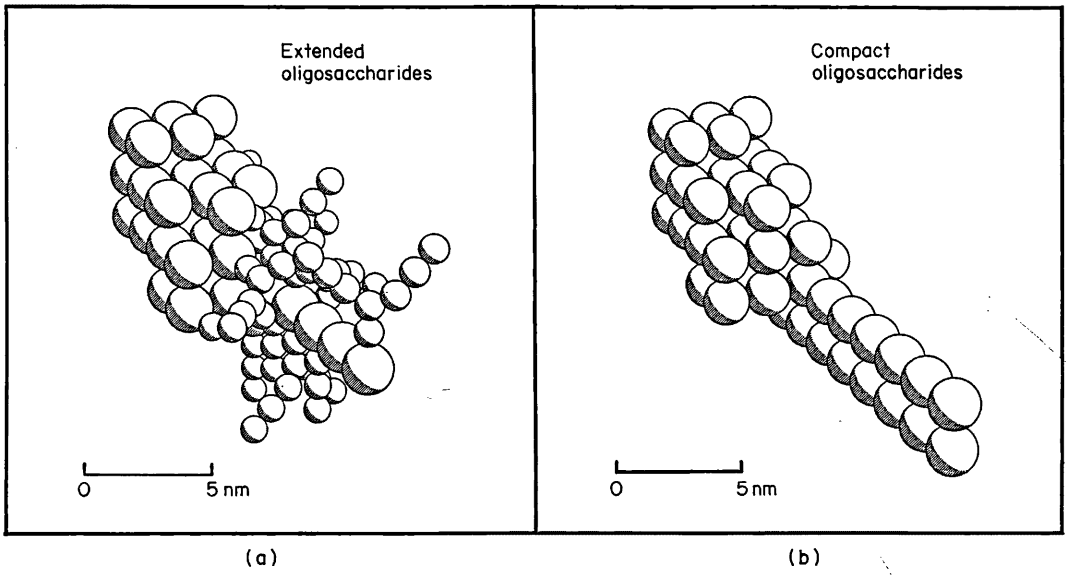


Figure 6. Hydrodynamic models for C1 inhibitor. The oligosaccharides are shown either (a) in an extended or (b) in a compact conformation. Both give $s^0_{20,w}$ values that agree with experiment. Protein spheres: 1·69 nm diameter; carbohydrate spheres: 0·95 nm diameter.

Another model with 50 spheres of diameter 1·69 nm (Fig. 6(b)) and length 18 nm in which the oligosaccharides are merged with the protein gave an $s^0_{20,w}$ of 3·8 S. Both models agree with the experimental values of 3·67 to 3·8 S and the neutron modelling. Since experimental $s^0_{20,w}$ values are usually known to within $\pm 0{\cdot}2$ S, and the simulations usually agree to within $\pm 0{\cdot}2$ S (Perkins, 1989), the oligosaccharide conformation cannot be identified. In conclusion, the hydrodynamic properties of C1 inhibitor are well explained in terms of a two-domain head-and-tail structure, but not by a cylinder.

4. Conclusions

The combination of the neutron and hydrodynamic solution data on C1 inhibitor with the crystal structure of α_1-antitrypsin (Löbermann *et al.*, 1984) and the corresponding solution studies on α_1-antitrypsin (Smith *et al.*, 1990) has led to an improved characterization of its two-domain structure. Structural evidence for the distinct existence of the serpin domain was obtained from the successful curve fitting of Figure 5, the hydrodynamic modelling, and possibly also from R_G in the biphasic Guinier R_G analyses of Figure 2. The structure of the N-terminal domain is heavily glycosylated and extended, as deduced from the overall length of 16 to 19 nm of C1 inhibitor in relation to α_1-antitrypsin, as well as the contrast variation data (Figs 2(b) and 3(b)). Electron microscopy had suggested that C1 inhibitor is much longer at 33 to 36 nm (Odermatt *et al.*, 1981); however, it is possible that artifacts were introduced in the course of that study, for example, through sample denaturation or magnification errors. Previously, lengths of 55 to 60 nm and 38 to 44 nm had been estimated from hydrodynamic data (Odermatt *et al.*, 1981; Perkins, 1985). Here it has been shown that these large values are the result of an erroneous M_r and the unjustified assumption that C1 inhibitor has a cylindrical structure. Figure 7 shows that the final length of 16 to 19 nm for C1 inhibitor is now similar to lengths of 17 to 20 nm determined for its substrates C1r and C1s (Villiers *et al.*, 1985; Weiss *et al.*, 1986; Perkins & Nealis, 1989).

C1 is a complex of two subcomponents C1q and $C1r_2C1s_2$. C1q recognizes the Fc domains of aggregated IgG and binds to them. C1q has a hexameric structure formed from six part-collagenous, part-globular subunits with the appearance of a "bunch of tulips" (Fig. 7). The enzymic activity of C1 resides in the tetramer $C1r_2C1s_2$. It is thought that $C1r_2C1s_2$ has an elongated asymmetric X-shape structure (Weiss *et al.*, 1986; Perkins & Nealis, 1989). Several models for their association to form C1 have been proposed (Schumaker *et al.*, 1987; Perkins, 1989*a*), where $C1r_2C1s_2$ either is intertwined between the six stalks of the C1q structure or is placed fully on the outside of the six C1q stalks. For reason of the requirement of steric accessibility of the serine proteinase and short consensus repeat (Fig. 7, SCR) domains in C1s and C1r to C4, the substrate of C1, the second model has been preferred (Cooper, 1985; Perkins, 1985; Perkins *et al.*, 1990). The comparisons in Figure 7 show that the interaction of the serpin domain with the serine protease domain is most readily understood in models where $C1r_2C1s_2$ is placed outside the six C1q stalks, such as in the W-model (Perkins, 1985, 1989*a*). Once C1q is attached *via* its heads to immune aggregates, it is likely that the size of up to four serpin domains would be too bulky to penetrate within the cone of C1q stalks. This access is required by the intertwined S- and 8-models for the C1

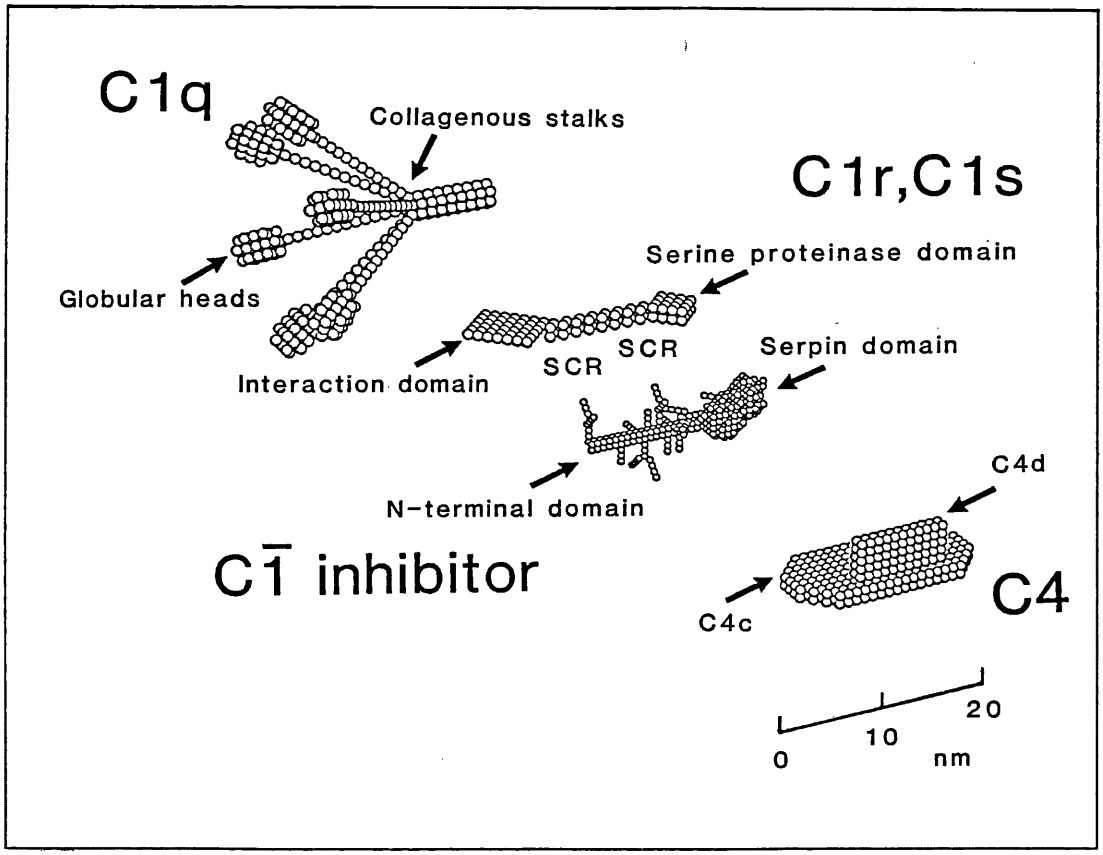


Figure 7. Molecular models of several early components of complement derived from solution scattering. The C1q structure (Perkins, 1985) is hexameric and binds a tetramer of C1r and C1s, of which a monomer is shown (Perkins & Nealis, 1989). The C$\bar{1}$ inhibitor model from Fig. 5 is also shown. This C1 complex is able to activate C4 of complement, for which a schematic model of its 2-domain structure is shown, to indicate the size of C4 (Perkins *et al.*, 1990). The domain structures of these components are indicated: C1q, 6 collagenous stalks and globular heads; C1r, C1s: the C domains I to VI are the interaction domain (I, II and III), 2 "short consensus repeat" (SCR) domains (IV and V), and a serine proteinase domain (VI); C$\bar{1}$ inhibitor: serpin domain and N-terminal domain; C4 can be represented by the juxtaposition of the C4c and C4d domains. Sphere diameters: C1q, 1·19 nm; C1r, C1s, 1·00 nm; C$\bar{1}$ inhibitor, 0·608 nm; C4, 0·80 nm.

complex, since these postulate that the four protease domains in C1 reside within the C1q cone (Colomb *et al.*, 1984; Schumaker *et al.*, 1986).

One possible function for the N-terminal domain of C$\bar{1}$ inhibitor is that this blocks the accessibility of the non-proteolytic domains I to V in C$\bar{1}$r and C$\bar{1}$s after the serpin domain of C$\bar{1}$ inhibitor has bound to the proteinase domain VI. Hydrodynamic calculations (S. J. Perkins, unpublished results) show that, in the complexes of C$\bar{1}$ inhibitor with C$\bar{1}$r and C$\bar{1}$s, the N-terminal domain of C$\bar{1}$ inhibitor is close to the two SCR domains of C$\bar{1}$r and C$\bar{1}$s (Fig. 7). It is possible from the W-model for C1 (or in the other models proposed for C1) that the two SCR domains IV and V in each of C1r and C1s and/or the interaction domains I to III interact with the collagenous stalks of C1q in the C1q complex with $C\bar{1}r_2C\bar{1}s_2$. Once C$\bar{1}$ inhibitor has reacted with C$\bar{1}$r and C$\bar{1}$s, it is reasonable to propose that the N-terminal domain mediates the reduction in the binding affinity of inhibited C$\bar{1}$r and C$\bar{1}$s for C1q. This frees C1q for reaction with the C1q receptor (Reid, 1986). It is also possible that the short consensus repeat domains in C1s interact also with C4. Since it would be necessary to reduce the affinity of C4 for C$\bar{1}$s after reaction of the latter with C$\bar{1}$ inhibitor, steric hindrance offered by the N-terminal domain of C$\bar{1}$ inhibitor would provide a requisite mechanism.

Financial support from the Wellcome Trust is gratefully acknowledged. The Oxford Glycobiology Unit is supported by the Monsanto Company. We thank Dr J. Torbet for generous instrumental support at the Institute-Laue-Langevin, Grenoble, Mr A. S. Nealis for computational support, Dr P. A. Pemberton for the assays with active-site titrated plasmin, and Dr K. S. Aulak for the high-pressure liquid chromatography analysis of split C$\bar{1}$ inhibitor.

References

Ashford, D., Dwek, R. A., Welply, J. K., Amatayakul, S., Homans, S. W., Lis, H., Taylor, G. N., Sharon, N. & Rademacher, T. W. (1987). *Eur. J. Biochem.* **166**, 311–320.

Bock, S. C., Skriver, K., Nielsen, E., Thøgersen, H. C., Wiman, B., Donaldson, V. H., Eddy, R. L., Marrinan, J., Radziejewska, E., Huber, R., Shows,

T. B. & Magnusson, S. (1986). *Biochemistry,* **25**, 4292–4301.
Bruch, M., Weiss, V. & Engel, J. (1988). *J. Biol. Chem.* **263**, 16626–16630.
Chesne, S., Villiers, C. L., Arlaud, G. J., Lacroix, M. B. & Colomb, M. G. (1982). *Biochem. J.* **201**, 61–70.
Chothia, C. (1975). *Nature (London)*, **254**, 304–308.
Colomb, M. G., Arlaud, G. J. & Villiers, C. L. (1984). *Phil. Trans. Roy. Soc. ser. B,* **306**, 283–292.
Cooper, N. R. (1985). *Advan. Immunol.* **37**, 151–216.
Davis, A. E. (1988). *Annu. Rev. Immunol.* **6**, 595–628.
Davis, A. E., Whitehead, A. S., Harrison, R. A., Dauphinais, A., Bruns, G. A. P., Cicardi, M. & Rosen, F. S. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3161–3165.
Dayhoff, M. O. (1978). In *Atlas of Protein Sequence and Structure,* vol. 5, p. 36, National Biochemical Research Foundation, Silver Springs, MD.
Garcia de la Torre, J. & Bloomfield, V. A. (1977*a*). *Biopolymers,* **16,** 1747–1761.
Garcia de la Torre, J. & Bloomfield, V. A. (1977*b*). *Biopolymers,* **16,** 1779–1791.
Garcia de la Torre, J. & Bloomfield, V. A. (1981). *Quart. Rev. Biophys.* **14**, 87–139.
Garnier, J., Osguthorpe, D. J. & Robson, B. (1978). *J. Mol. Biol.* **120**, 97–120.
Ghosh, R. E. (1981). *I.L.L. Internal Publication* 81GH29T.
Glatter, O. (1982). In *Small Angle X-ray Scattering* (Glatter, O. & Kratky, O., eds), pp. 119–196, Academic Press, London.
Gordon, A. H. (1975). In *Electrophoresis of Proteins in Polyacrylamide and Starch Gels* (Work, T. S. & Work, E., eds), pp. 153s–164s, North-Holland Publ. Co., Amsterdam.
Guinier, A. & Fournet, G. (1955). *Small Angle Scattering of X-rays,* Wiley, New York.
Harpel, P. C. (1976). *Methods Enzymol.* **45**, 751–760.
Harpel, P. C. & Cooper, N. R. (1975). *J. Clin. Invest.* **55**, 593–604.
Harrison, R. A. (1983). *Biochemistry,* **22**, 5001–5007.
Harrison, R. A. (1989). *Compl. Inflamm.* **6**, 341 (abstract).
Harrison, R. A. & Lachmann, P. J. (1986). In *Handbook of Experimental Immunology* (Weir, D. M., Herzenberg, L. A., Blackwell, C. & Herzenberg, L. A., eds), pp. 39.1–39.49, Blackwells, Oxford.
Haupt, H., Heimburger, N., Kranz, T. & Schwick, H. G. (1970). *Eur. J. Biochem.* **17**, 254–261.
Ibel, K. & Stuhrmann, H. B. (1975). *J. Mol. Biol.* **93**, 225–266.
Jacrot, B. & Zaccai, G. (1981). *Biopolymers,* **20**, 2413–2426.
Laemmli, U. K. (1970). *Nature (London),* **277**, 680–685.
Lennick, M., Brew, S. A. & Ingham, K. C. (1985). *Biochemistry,* **25**, 2561–2568.
Löbermann, D., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984). *J. Mol. Biol.* **177**, 531–556.
Mizuochi, T., Yamashita, K., Fujikawa, K., Tiatini, K. & Kobata, A. (1980). *J. Biol. Chem.* **255**, 3526–3531.
Nilsson, T. & Wiman, B. (1982). *Biochim. Biophys. Acta,* **705**, 271–276.
Odermatt, E., Berger, H. & Sano, Y. (1981). *FEBS Letters,* **131**, 283–285.
Parekh, R. B., Tse, A. G. C., Dwek, R. A., Williams, A. F. & Rademacher, T. W. (1987). *EMBO J.* **6**, 1233–1244.
Pemberton, A. A., Harrison, R. A., Lachmann, R. J. & Carrell, R. W. (1989). *Biochem. J.* **258**, 193–198.
Pensky, J., Levy, L. R. & Lepow, I. H. (1961). *J. Biol. Chem.* **236**, 1674–1679.
Perkins, S. J. (1985). *Biochem. J.* **228**, 13–26.
Perkins, S. J. (1986). *Eur. J. Biochem.* **157**, 169–180.
Perkins, S. J. (1988*a*). *Biochem. J.* **254**, 313–327.
Perkins, S. J. (1988*b*). In *New Comprehensive Biochemistry* (Neuberger, A. & Van Deenen, L. L. M., eds), vol. 18B, part II, pp. 143–264, Elsevier, Amsterdam.
Perkins, S. J. (1989*a*). *Behring Institute Mitt.* **84**, 129–141.
Perkins, S. J. (1989*b*). In *Dynamic Properties of Biomolecular Assemblies* (Harding, S. E. & Rowe, A. J., eds), pp. 226–245, Royal Society of Chemistry, London.
Perkins, S. J. & Nealis, A. S. (1989). *Biochem. J.* **263**, 463–469.
Perkins, S. J. & Sim, R. B. (1986). *Eur. J. Biochem.* **157**, 155–168.
Perkins, S. J. & Weiss, H. (1983). *J. Mol. Biol.* **168**, 847–866.
Perkins, S. J., Miller, A., Hardingham, T. E. & Muir, H. (1981). *J. Mol. Biol.* **150**, 69–95.
Perkins, S. J., Kerckaert, J. P. & Loucheux-Lefebvre, M. H. (1985). *Eur. J. Biochem.* **147**, 525–531.
Perkins, S. J., Nealis, A. S. & Sim, R. B. (1990). *Biochemistry,* **29**, 1167–1175.
Pilz, I. (1982). In *Small Angle X-ray Scattering* (Glatter, O. & Kratky, O., eds), pp. 239–293, Academic Press, London.
Reboul, A., Arlaud, G. J., Sim, R. B. & Colomb, M. G. (1977). *FEBS Letters,* **79**, 45–50.
Reboul, A., Prandini, M.-H. & Colomb, M. G. (1987). *Biochem. J.* **244**, 117–121.
Reid, K. B. M. (1986). *Essays Biochem.* **22**, 27–68.
Ryan, B. F., Joiner, B. L. & Ryan, T. A. (1985). Minitab Handbook, 2nd edit., PWS-Kent Publishing Company, Boston.
Salvesen, G. S., Catanese, J. J., Kress, L. F. & Travis, J. (1985). *J. Biol. Chem.* **260**, 2432–2436.
Schultze, H. E., Heide, K. & Haupt, H. (1962). *Naturwissenschaften,* **6**, 133–134.
Schumaker, V. N., Hanson, D. C., Kilchherr, E., Phillips, M. L. & Poon, P. H. (1986). *Mol. Immunol.* **23**, 557–565.
Schumaker, V. N., Zavodsky, P. & Poon, P. H. (1987). *Annu. Rev. Immunol.* **5**, 21–42.
Sim, R. B. & Reboul, A. (1981). *Methods Enzymol.* **80**, 43–54.
Smith, K. F., Harrison, R. A. & Perkins, S. J. (1990). *Biochem. J.* **267**, 203–212.
Strecker, G., Ollier-Hartmann, M. P., van Halbeek, H., Friederik, J., Vliegenhart, G., Montreuil, J. & Hartmann, L. (1985). *C.R. Acad. Sci., Paris,* **301**, 571–576.
Tosi, M., Duponchel, C., Bourgarel, P., Colomb, M. G. & Meo, T. (1986). *Gene,* **42**, 265–272.
Villiers, C. L., Arlaud, G. J. & Colomb, M. G. (1985). *Proc. Nat. Acad. Sci. U.S.A.* **82**, 4477–4481.
Weiss, V., Fauser, C. & Engel, J. (1986). *J. Mol. Biol.* **189**, 573–581.
Wetlaufer, D. B. (1962). *Advan. Protein Chem.* **17**, 303–390.
Yamashita, K., Mizouchi, T. & Kobata, A. (1982). *Methods Enzymol.* **83**, 105–126.

Edited by R. Huber

Reprinted from Biochemistry, 1990, 29, 1175.

Structural Homologies of Component C5 of Human Complement with Components C3 and C4 by Neutron Scattering†

Stephen J. Perkins,*,‡ Kathryn F. Smith,‡ Adam S. Nealis,‡ Peter J. Lachmann,§ and Richard A. Harrison§

Department of Biochemistry and Chemistry and Department of Protein and Molecular Biology, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K., and Molecular Immunopathology Unit, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH, U.K.

Received April 11, 1989; Revised Manuscript Received September 14, 1989

ABSTRACT: The complement component C5 is one of a family of structurally related plasma proteins that includes components C3 and C4. Activation of C5 is the initial step in the formation of the membrane attack complex of complement. Analysis of the solution structure of C5 and comparisons with similar analyses of the structures of C3 and C4 are reported here. Neutron solution scattering gave an M_r for C5 of 201 000, which demonstrates that C5 is monomeric in solution. The radius of gyration R_G of C5 at infinite contrast is 4.87 nm and corresponds to an elongated structure. The longest length of C5 was determined to be at least 15–16 nm from three calculations on the basis of the R_G, the scattering intensity at zero angle $I(0)$, and the indirect transformation of the scattering curve into real space. Comparison of the R_G and contrast variation data and indirect transformations of the scattering curves for C3, C4, and C5 show that these have very similar structures. Comparisons of the C5 scattering curve with Debye small-sphere models previously employed for C4 and C3 show that good curve fits could be obtained. Unlike previous studies that have suggested significant differences, these experiments indicate that, while C5 differs from C3 and C4 in its activation and inactivation pathways, significant structural homology exists between the native proteins, as might be predicted from their high (and similar) sequence homology.

Complement, a multicomponent cascade system found in plasma, plays a major role in the humoral immune system (Reid, 1986). Activation of the early components of the classical and alternative pathways by proteolytic cleavages generates C5 convertase activity. The convertases activate C5 to form C5b and C5a. The generation of C5b is the initial step in the formation of the membrane attack complex, which is assembled from the complement components C6, C7, C8, and C9 in association with C5b. Insertion of the complex into membranes leads to the formation of transmembrane channels in the target cell and ultimately to cell death due to osmotic imbalance.

The primary sequences of the complement components C3, C4, and C5 show significant homologies (Belt et al., 1984; de Bruijn & Fey, 1985; Wetsel et al., 1987, 1988). C5, like C3, is composed of two disulfide-linked polypeptide chains α and β. In contrast, C4 has three chains, α, β, and γ. Activation of C5 is brought about by cleavage of the peptide bond between residues 74 and 75 of the α chain, releasing C5a (Nilsson et al., 1975). C5a is the most potent of the three anaphylatoxins C3a, C4a, and C5a formed by the complement cascade. Despite the sequence homology, the genes for C3, C4, and C5 are found on different chromosomes (Carroll et al., 1984; Whitehead et al., 1982; Wetsel et al., 1988). While not as abundant as C3 and C4, C5 at 0.15 mg/mL is still one of the more abundant complement components in plasma. C3 and C4, but not C5, have a highly reactive internal β-cysteinyl-γ-glutamyl thiolester bond in a region of high sequence homology. During activation, this enables C3b and C4b to form covalent bonds to target surfaces (Law et al., 1980). Since there is no thiolester bond in C5, C5b is unable to form a covalent bond to a surface. Thus, C5b in the presence of C6 forms a fluid-phase C5b6 bimolecular complex, and this in association with C7 expresses a membrane-binding site (Podack et al., 1978).

In this study, comparisons between the solution structures of C3, C4, and C5 are made. Electron microscopy of C5 depicts a multilobal, irregular ultrastructure of dimensions of 10.4 nm × 14.0 nm × 16.8 nm (DiScipio et al., 1983); different

† We thank the Wellcome Trust for support, and the Science and Engineering Research Council for access to the facilities at SRS Daresbury and ILL Grenoble.

* To whom correspondence should be addressed at the Department of Biochemistry and Chemistry.

‡ Royal Free Hospital School of Medicine.

§ Medical Research Council Centre.

0006-2960/90/0429-1175$02.50/0

dimensions have, however, been reported for C3 and C4 (Smith et al., 1982, 1984; Dahlbäck et al., 1983). Limited solution scattering data on C3, C4, and C5 suggested that despite their sequence homologies they had dissimilar R_G values (Österberg et al., 1984, 1988). In this study, extensive neutron scattering experiments on the external and internal structure of C5 are critically compared with similar studies on C3 and C4 (Perkins & Sim, 1986; Perkins et al., 1990). Together, the neutron data show that C3, C4, and C5 are more homologous in their physical structures than previously believed.

Materials and Methods

(*a*) *Sample Preparations.* C5 was prepared from 3.5 L of human plasma from normal, healthy donors essentially as described in Harrison and Lachmann (1986), with the following modifications. Plasma was taken from donors over a period of 2 weeks prior to C5 preparation, EDTA and benzamidine were added to final concentrations of 5 mM, and the plasmas were stored frozen at −70 °C until required. After passage though lysine–Sepharose, the plasma pool was dialyzed [by using a Pellicon ultrafiltration cassette (Millipore) equipped with a 5 ft^2, 10000 nominal molecular weight limit, polysulfone membrane] against 5 mM sodium potassium phosphate, 5 mM EDTA, 5 mM benzamidine, and 50 mM NaCl, pH 7.0, and loaded onto a 10-L DEAE-Sephacel column equilibrated in the same buffer. The protein was eluted with a gradient (25 L each side) to 300 mM NaCl, and C5, eluting early in the gradient, was detected by double diffusion in agarose against a rabbit anti-human C5 antiserum. C5-containing fractions were pooled, concentrated to 150 mL, dialyzed against 20 mM sodium potassium phosphate, 10 mM EDTA, 5 mM benzamidine, and 300 mM NaCl, pH 7.0, and gel-filtered on Sepharose CL-6B (Pharmacia; 6-L column). C5-containing fractions were detected as described above, pooled, and dialyzed against 20 mM sodium potassium phosphate and 100 mM KCl, pH 7.0, before being loaded onto a hydroxylapatite column (Bio-Rad HTP; 600 mL bed volume, 30 × 5.5 cm column) equilibrated in the same buffer. C5 was eluted with a gradient (2.5 L each side) to 2.0 M KCl. Antigenic analysis showed that the small amount of C3 that was contained in the DEAE-Sephacel C5 pool and which cochromatographed with C5 on Sepharose CL-6B was not eluted from hydroxylapatite by KCl. SDS–PAGE analysis showed the C5 to be greater than 95% pure and to have the anticipated two-chain structure. The C5 pool from hydroxylapatite was concentrated to approximately 10 mg/mL and dialyzed against PBS/5 mM sodium azide. Initial neutron scattering studies were performed on freshly prepared C5. Prior to analysis, these were dialyzed against buffer A (see below), filtered through a 0.2-μm filter (ACRO LC13; Gelman Sciences) into 3.0- or 4.8-mL sterile evacuated glass vials (Mallinckrodt Diagnostica Holland B.V., Petten, Holland) and held at 4 °C. Subsequent studies were performed on C5 that had been stored frozen at −70 °C in 1-mL (10 mg) aliquots. These samples were prepared for scattering studies by gel filtration on Sepharose 6B in 12 mM sodium phosphate, 200 mM NaCl, and 1 mM EDTA, pH 7.0. The peak fraction(s) only were reconcentrated to about 10 mg/mL and then dialyzed against the same buffer before (ACRO LC13) filtration and storage as described above. Samples were reanalyzed by SDS–PAGE subsequent to solution scattering studies, and no differences pre- and postscattering were detected.

Four different buffers were used in neutron data collection: A (200 mM NaCl, 12 mM sodium phosphate, 1 mM EDTA, pH 7.0); B (50 mM NaCl, 12 mM sodium phosphate, 1 mM EDTA, pH 7.0); C (3.1 mM diethylbarbituric acid, 0.9 mM sodium barbitone, 145 mM NaCl, 0.83 mM $MgCl_2$, or 0.25 mM $CaCl_2$, 20 μM $ZnSO_4$); D (buffer C diluted 1 in 4, except for $ZnSO_4$, which remained at 20 μM). Buffer A was predominantly used, and dialyses were performed as in Perkins et al. (1990). The composition is calculated from the full 1654-residue sequence of human C5 (Wetsel et al., 1988; R. A. Wetsel, unpublished results). All four putative carbohydrate binding sites at Asn-705, -875, -1079, and -1593 in the human sequence (following the mouse sequence numbering; Wetsel et al., 1987) are occupied with biantennary complex type oligosaccharides, giving a 4.5% oligosaccharide content by mass. DiScipio et al. (1983) reported this content to be 2.7% by mass. The absorption coefficient $A_{280}^{1\%,1cm}$ for human C5 has been estimated as 10.8 (DiScipio et al., 1983) or 10.2 (Österberg et al., 1988), and their mean is used here. Calculation of the absorption coefficient (Perkins, 1986) gives 10.9, in good agreement with DiScipio et al. (1983).

(*b*) *Neutron Scattering Measurements.* Neutron experiments were performed on instrument D17 at the Institut Laue-Langevin, Grenoble, in five independent sessions. Guinier data were obtained by using a sample-to-detector distance of 3.45–3.46 m and wavelengths λ of 1.30–1.60 nm, giving a Q range of 0.0053–0.049 nm^{-1} (where $Q = 4\pi \sin\theta/\lambda$ and 2θ is the scattering angle). Data at larger Q were obtained with a sample-to-detector distance of 1.40 m, a main beam–detector axis angle of 0° or 20°, and wavlengths of 0.99–1.00 nm, giving a Q range of 0.23–3.36 nm^{-1}. Data reduction followed standard procedures (Ghosh, 1981; Perkins et al., 1990). Linear regressions of the match point and Stuhrmann graphs were analyzed by microcomputer using the statistics program MINITAB (version 6.1). The molecular modeling of the C5 neutron curves utilized the methods in Perkins and Weiss (1983), Perkins and Sim (1986), and Perkins et al. (1990).

The use of the indirect transformation procedure (ITP) method (Glatter, 1982) enables the neutron scattering curves $I(Q)$ to be transformed into real space $P(r)$. This calculation requires measurement of the whole scattering curve to large Q, ca. 4 nm^{-1}, and offers an alternative calculation of the radius of gyration R_G and the length L for C5. While 20–25 splines should be used to fit the 75 experimental data points of Figure 4a, only 10 splines were used here because of the limited counting statistics at large Q. The maximum length (which can be up to 50% larger than the expected length L) was specified as 20 nm on the basis of the L calculations from the Guinier R_G analyses; control calculations on the basis of maximum lengths of 17 nm gave similar results. The number of points used to define $P(r)$ in real space was 101.

Results and Discussion

(*a*) *Neutron Scattering Data on C5.* Neutron contrast variation experiments were performed on human C5 at concentrations between 1.8 and 7.9 mg/mL in 0%, 70%, 80%, and 100% ^2H_2O buffers. Linear Guinier analyses were successfully obtained (Figure 1a), from which the intensity $I(0)$ and the radius of gyration R_G were derived in a suitable QR_G range of 0.6–1.5. Linear cross-sectional analyses to obtain $[I(Q)\text{-}Q]_{Q\to 0}$ and R_{XS} values were obtained in the QR_{XS} range 0.7–1.4 (Figure 1b). Buffers A–D were used (Materials and Methods) with ionic strengths of 0.22, 0.07, 0.15, or 0.04 M. The two higher salt buffers A and C did not yield data that differed significantly from the average. In the two lower salt buffers, C5 showed evidence of aggregation, so these data were not considered further. In the higher salt buffers, while the $I(0)/c$ (c is the concentration) and R_G values exhibit minor decreases with concentration (results not shown), this was not

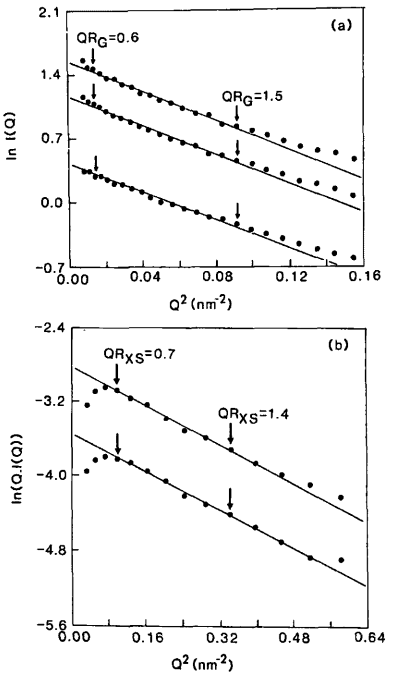


FIGURE 1: Neutron Guinier plots of human complement C5 in 100% ^2H_2O buffers: (a) Guinier plots and (b) cross-sectional Guinier plots of a dilution series for C5. The vertical displacements of the curves reflect sample concentrations at 5.4, 3.9, and 1.9 mg/mL in (a) and 3.9 and 1.9 mg/mL in (b). The R_G fits and QR_G values are reported by using an R_G value of 4.8 nm (Q range of 0.12–0.30 nm^{-1}) and those for QR_{XS} with an R_{XS} value of 2.34 nm (Q range of 0.31–0.58 nm^{-1}). A slight decrease of the R_G with concentration is observable in (a) and in other percent ^2H_2O buffers, but was not significant in terms of the error of the R_{G-C} determination in Table I.

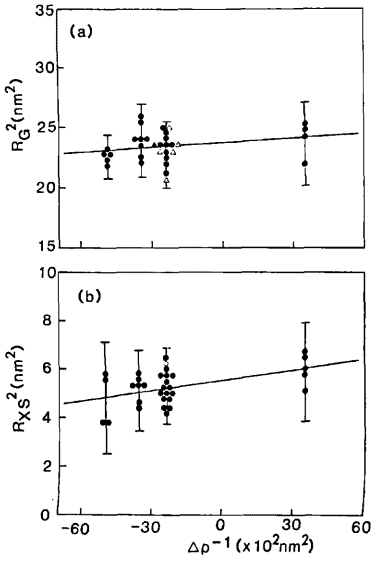


FIGURE 2: Stuhrmann analyses of the contrast variation experiments for human C5. The R_G^2 and R_{XS}^2 data are plotted against the reciprocal solvent–solute contrast difference $\Delta\rho^{-1}$. The symbols correspond to the four buffers used (Materials and Methods) as follows: A, ●; B, ○; C, △; D, ▲. The regressed line based on the R_G^2 values gives an R_C of 4.87 ± 0.03 nm and a slope α of (13 ± 8) × 10^{-5}. For the R_{XS}^2 data, R_C is 2.35 ± 0.03 nm and α is (13 ± 4) × 10^{-5}.

significant, so all the R_G data are given in Figure 2.

Molecular masses M_r were determined from the $I(0)/c$ data for the samples in 0% ^2H_2O in high-salt buffers. The mean of four measurements is 201 000 ± 9000, which is within error of 194000 calculated from the human C5 sequence (Materials and Methods). This shows that the absorption coefficient of 10.5 used to derive concentrations is satisfactory and that C5 is monomeric in solution. A plot of $[I(0)/ctT_s]^{1/2}$ against volume percent ^2H_2O gives the neutron match point (t is the sample thickness; T_s is the sample transmission). An excellent straight line was obtained, which give a match point of 40.9 ± 0.4% ^2H_2O by linear regression analyses of 33 values in 4 contrasts. The corresponding cross-sectional plot of $[[I(Q)-Q]_{Q\to 0}/ctT_s]^{1/2}$ against percent ^2H_2O yielded a match point of 41.5 ± 0.4% ^2H_2O. Both of these are in reasonable agreement with the match point of 40.6% ^2H_2O predicted (Perkins, 1986) from the human C5 sequence. They support the validity of the R_G and R_{XS} analyses in Figure 2.

Table I: Summary of Guinier Analyses for C5, C4, and C3 of Complement[a]

	C5	C4, C4u, C4(a+b)	C3, C3u, C3(a+b)
X-rays			
R_G (nm)	5.3 ± 0.1	5.3 ± 0.1	5.2 ± 0.1
R_{XS} (nm)	2.3 ± 0.1	2.5 ± 0.1	2.6 ± 0.1
neutrons			
R_{G-C} (nm)	4.9 ± 0.1	4.9 ± 0.1	5.1 ± 0.1
R_{XS-C} (nm)	2.4 ± 0.1	2.2 ± 0.1	2.4 ± 0.1
α_G (×10^{-5})	13 ± 8	2 ± 10	30 ± 10
α_{XS} (×10^{-5})	13 ± 4	20 ± 10	10 ± 10

[a] The data for C4 and C3 are taken from Perkins et al. (1990) and Perkins and Sim (1986), and mean values are shown.

To characterize the shape of C5 and its internal structure, the contrast dependence of the structural data was analyzed by using Stuhrmann plots of R_G^2 and R_{XS}^2 against the reciprocal solvent–solute contrast difference $\Delta\rho^{-1}$ (Figure 2). The radii of gyration at infinite contrast R_{G-C} and R_{XS-C} were found to be 4.87 and 2.35 nm. The elongation ratio R_{G-C}/R_0 was calculated as 1.61 (where R_0 is the R_G of the shere of the same dry volume as C5: 248 nm^3). This is significantly larger than the value of 1.28 ± 0.10 found in a survey of many globular proteins by X-ray scattering (Perkins, 1988b) and shows that C5 is elongated in shape with one triaxial dimension that is longer than the other two. The radial inhomogeneity of scattering density α is positive at 13 × 10^{-5} in both Stuhrmann plots. This is evidence that C5 has the spatial distribution of amino acids that is typical of soluble globular proteins.

The dimensions of C5 can be calculated from the R_G and R_{XS} data and also from the $I(0)$ and $[I(Q)Q]_{Q\to 0}$ data (Perkins et al., 1990). By use of the data in 100% ^2H_2O, the length L of the longest axis of C5 (assuming an elliptical cylinder) was determined as 14.9 ± 0.5 nm from the R_G and R_{XS} data. From the ratio $\pi I(0)/[I(Q)Q]_{Q\to 0}$, L is found to be 15.7 ± 1.1 nm.

X-ray scattering curves for C5 were obtained by using the synchrotron radiation source at Daresbury (Perkins et al., 1990). In Guinier analyses, the final R_G and R_{XS} values were based on Q ranges of 0.18–0.30 and 0.31–0.60 nm^{-1}, respectively, which are slightly more restricted than those of Figure 1. Only those spectra that could be analyzed by the criteria of linear Guinier plots were used. These resulted in data that were fully consistent within error with those from neutron scattering (Table I).

(b) Comparisons of C5 with C3 and C4 of Complement. Scattering data analyses relating to the presumed homology between C3, C4, and C5 are presented in Table I. The present X-ray and neutron data on C5 are compared with data previously obtained for C3 and C4 and their fragments (Perkins & Sim, 1986; Perkins et al., 1990). C3u and C4u correspond to C3 and C4 with hydrolyzed thiolester bonds, while C3(a+b) and C4(a+b) correspond to activated C3 and C4 in which the two fragments continue to be associated. By both X-ray and neutron scattering, the R_G and R_{XS} values of C3, C3u, C3-(a+b), C4, C4u, C4(a+b), and C5 were found to be the same

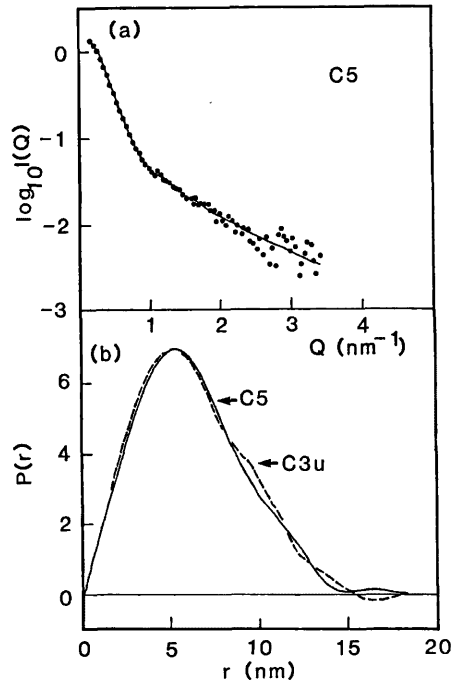


FIGURE 3: Indirect transformation into real space of the C5 and C3u neutron scattering curves in 100% ^2H_2O buffer. (a) The experimental neutron curve of C5 (●) is shown. The continuous line is a polynomial function generated by the ITP program. That for the C3u scattering curve is very similar and is not shown. (b) The intraparticle distance distribution function $P(r)$ for C5 (—) and C3u (---) is shown after normalization for concentration. These were calculated from the curve fits in (a). The $P(r)$ curves for C5 and C3u are seen to be very similar. For C5, a R_G of 4.89 nm was calculated over the range of $-8.0 \leq \lambda_{opt} \leq 0.0$, where λ_{opt} is the Lagrange stabilization parameter used in ITP. For C3u, R_G is 4.77 nm over a similar range of λ_{opt}.

within error (Table I). The R_G values from X-ray scattering correspond to measurements in a positive solute–solvent contrast difference. They are higher than the neutron R_{G-C} values at the infinite contrast condition, thus corroborating the positive Stuhrmann α values obtained from the neutron analyses. It is concluded that C3, C4, and C5 are closely similar in their overall shapes in solution, as judged from the scattering curve analyses out to $Q = 0.6$ nm^{-1}. All three proteins show similar, positive values for α (Table I). From the sequences, the ratios of hydrophilic to hydrophobic residues [defined in Perkins (1986)] are similar at 50:50, 47:53, and 49:51 in that order. The similarity of the α values is thus consistent with compositional data.

(*c*) *Large-Angle Curve Analyses of C5.* Indirect transformation of the neutron curves $I(Q)$ into real space $P(r)$ using ITP (Materials and Methods) was carried out for C5, C4, C4u, and C3u to assess further the structural homologies between C3, C4, and C5. The data on C4 and C4u originated from Perkins et al. (1990). This calculation utilizes the whole scattering curve in the Q range from 0.2 to 3.4 nm^{-1}. The above R_G and R_{XS} data were based on the smaller Q range out to 0.6 nm^{-1}. Sufficient spectra for comparisons were obtained in three sessions. At large Q, the $I(Q)$ and $P(r)$ curves for C5, C4, C4u, and C3u were all similar, as illustrated for C5 and C3u in Figure 3. This agreement confirmed the structural homologies in solution to a nominal structural resolution of 1.7 nm. In Figure 3b, the maximum length of C5 and C3u was determined to be at least 15 nm and could be 18 nm from the value of r, where $P(r)$ diminishes to zero intensity. This is compatible wiith the L of 14.9–15.7 nm determined from the R_G analyses for C5. The R_G from indirect transformation was determined (Glatter, 1982) as 4.89 nm. This is in good agreement with the mean value of 4.82

Table II: Comparisons of Electron Microscopy Studies on C3, C4, and C5 with the Debye Solution Scattering Model

		triaxial dimensions (nm)	reference
cobra venom factor[a]		8.2 × 13.7	Smith et al. (1982)
C4b		9 × 17	Dahlbäck et al. (1983)
C3b, C4b		6.5 × 7.5 × 12.5	Smith et al. (1984)
C5		10.4 × 14.0 × 16.8	DiScipio et al. (1983)
solution scattering model[b]			
C3, C4, C5	C3c, C4c	2 × 8 × 18	Perkins et al. (1990)
	C3dg, C4d	2 × 4 × 9	

[a] The cobra venom factor is a structural analogue of C3c. The reported standard deviations are ±1.1 to ±1.3 nm. [b] See Figure 4 for the possible arrangement of the two domains.

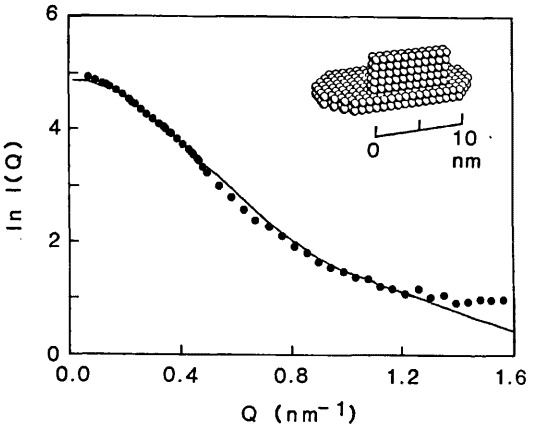


FIGURE 4: Comparison of the neutron scattering curve of C5 in ^2H_2O with Debye scattering curve simulations. The experimental data (●) are compared with the calculated Debye scattering curve (—) by using the possible two-domain model for C4 and C4u from Perkins et al. (1990). This sphere model gives a reasonable curve fit and is shown as the inset in the upper right corner. The precise arrangement of the two domains as shown should not be overinterpreted. The R_G of the model is 4.97 nm, in good agreement with the values of 4.82 and 4.89 nm from the Guinier and ITP analyses of the data in 100% ^2H_2O (see text).

± 0.10 nm for 16 R_G values from the Guinier analyses in 100% ^2H_2O buffers (cf. the Stuhrmann plot of Figure 2a).

A second description of the neutron curve for Q out to 1.6 nm^{-1} is obtained by computer models using Debye spheres. To test whether C5 and C4 are similar in structure, the experimental scattering curve for C5 was compared with the simulation from the C4 and C4u model (Perkins et al., 1990). The model is constructed from two distinct domains (Table II) which correspond to C4d and C4c and which are set at an angle to each other. Their dimensions are given in Table II; the length of 18 nm is slightly longer than the three calculations of 15–16 nm above but is compatible with the anticipated precision of the model. Several neutron scattering curves for C5 in 100% ^2H_2O were measured out to 1.6 nm^{-1} (cf. Figure 2b). The reasonable agreement in Figure 4 makes it likely that C5 also has a substructure with at least two domains as found for C3 and C4, even though this structure is not revealed in physiological inactivation events.

Conclusions

The scattering data from this study on C5 can be correlated with those on the homologous complement components C3 and C4 (Perkins & Sim, 1986; Perkins et al., 1990). The data of Table I show that C5 is very similar in structure to C3, C3u, C3(a+b), C4, C4u and C4(a+b). Comparisons of the full scattering curves out to a structural resolution of 2 nm (Q = 3.6 nm^{-1}) show that these three glycoproteins are structurally

closely homologous. This is fully consistent with the complete sequences that have been recently published for C3, C4, and C5 (de Bruijn & Fey, 1985; Belt et al., 1984; Wetsel et al., 1987, 1988). The comparisons of the C3, C4, and C5 sequences show that these can be readily aligned, and the location of cysteinyl residues and the hydropathy plots for C3, C4, and C5 are very similar (Nonaka et al., 1985; Wetsel et al., 1987).

This result is in contrast to previous electron microscopy studies on these complement components, as summarized in Table II (Smith et al., 1982, 1984; Dahlbäck et al., 1983). It was reported that C3, C4, and C5 possess globular structures which could be subdivided into at least two and at most four or five domains and with dimensions that differ by several nanometers. Calculation of the volume of the C3b and C4b images is comparable with values of 241 to 244 nm^3 calculated from their sequences, while that of C5 gives a volume that is 5 times larger than the sequence volume of 250 nm^3. At least one of the dimensions of C5 may be overestimated. These differences are most likely to result from the nonphysiological conditions of the measurements and perhaps magnification errors. In relation to the limited scattering data reported by Österberg et al. (1984, 1988) on C3, C4, and C5, their R_G values of C3u, C4u, and C5 were found to be similar at 5.0–5.3 nm, in agreement with our data. However, their R_G values for C3 and C4 were as low as 4.25 nm, and their α values for C3, C4, and C5 varied between -100×10^{-5} and 55×10^{-5}. These differences are best explained by the limited data that were available in these studies. For example, the range of the R_G data in Figure 2 shows that repeated measurements of the scattering curves, preferably in different buffers and protein concentrations, are necessary for a safer interpretation of the experiments.

Structural information on the three small homologous anaphylatoxins C3a, C4a, and C5a has been obtained by using protein crystallography, model building, and proton NMR (Huber et al., 1980; Greer, 1985, 1986; Nettesheim et al., 1988; Zuiderweg et al., 1988, 1989). The model building showed that C3a, C4a, and C5a all have conserved interior residues but very different external surfaces. The crystal structure of C3a has also proved to be a good model for the secondary structure of C3a and C5a in solution by NMR except at the N- and C-terminal ends of the structure. Our results extend the above observations of structural homologies.

C5 is the component upon which the membrane attack complex of complement is assembled. Since this complex is inserted into membranes, the degree to which the C5 structure is hydrophobic becomes relevant. Al Salihi et al. (1982, 1988) interpreted observations of the behavior of C5 on phenyl-Sepharose chromatography and the effects of trypsin digestion as evidence for hydrophobic surface regions on C5. Futhermore, DiScipio et al. (1983) suggested that the aggregation of C5b seen on activation of C5 was the result of the exposure of hydrophobic sites; the observation of such aggregates is supported by our preliminary neutron studies on C5b. It should, however, be noted that activation of C5 in the absence of C6 is nonphysiological and that neither C5b nor C5b6 binds to membranes. The ternary complex with C7 is required before membrane binding occurs, and it is likely that the membrane binding site of the complex lies on C7. The contrast variation data on C5 show that this has a similar internal structure to those of C3 and C4 (Table I) with hydrophilic residues found on average furthest from the core, as indeed found for globular proteins (Perkins, 1988a,b). This is to be expected since all three are water soluble and have similar ratios of hydrophobic and hydrophilic residues. The scattering data thus limit the extent to which the surface of C5 can be considered hydrophobic and, in contrast to deductions made from other biochemical investigations, rule out any significant region of surface hydrophobicity.

Acknowledgments

Dr. C. Nave and Dr. J. Torbet are thanked for generous instrumental support, Dr. O. Glatter and Dr. R. P. May are thanked for the use of the ITP and MAKITP programs, and Dr. R. A. Wetsel is thanked for the use of the human C5 sequence.

Registry No. C5, 80295-53-0; C3, 80295-41-6; C4, 80295-48-3.

References

Al Salihi, A. A., Ripoche, J., Pruvost, L., & Fontaine, M. (1982) *FEBS Lett. 150*, 238–242.

Al Salihi, A. A., Ripoche, J., & Fontaine, M. (1988) *Mol. Immunol. 25*, 367–377.

Belt, K. T., Carroll, M. C., & Porter, R. R. (1984) *Cell 36*, 907–914.

Carroll, M. C., Campbell, R. D., Bentley, P. R., & Porter, R. R. (1984) *Nature 307*, 237–241.

Dahlbäck, B., Smith, C. A., & Müller-Eberhard, H. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3461–3465.

de Bruijn, M. H. L., & Fey, G. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 708–712.

DiScipio, R. G., Smith, C. A., Müller-Eberhard, H. J., & Hugli, T. E. (1983) *J. Biol. Chem. 258*, 10629–10636.

Ghosh, R. (1981) ILL Internal Publication 81GH29T.

Glatter, O. (1982) in *Small Angle X-Ray Scattering* (Glatter, O., & Kratky, O., Eds.) pp 119–196, Academic Press, London.

Greer, J. (1985) *Science 228*, 1055–1060.

Greer, J. (1986) *Enzyme 36*, 150–163.

Harrison, R. A., & Lachmann, P. J. (1986) Complement Technology, in *Handbook of Experimental Immunology* (Weir, D. M., Herzenberg, L. A., Blackwell, C., & Herzenberg, L. A., Eds.) pp 39.1–49, Blackwell Scientific Publications, Edinburgh.

Huber, R., Scholze, H., Pâques, E. P., & Deisenhofer, J. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1389–1399.

Law, S. K., Lichtenberg, N. A., Holcombe, F. H., & Levine, R. P. (1980) *J. Immunol. 125*, 634–639.

Nettesheim, D. G., Edalji, R. P., Mollison, K. W., Greer, J., & Zuiderweg, E. R. P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5036–5040.

Nilsson, U. R., Mandle, R. J., & McConnell-Mapes, J. A. (1975) *J. Immunol. 114*, 815–822.

Nonaka, M., Nakayama, K., Yeul, Y. D., & Takahasi, M. (1985) *J. Biol. Chem. 260*, 10936–10943.

Österberg, R., Eggertsen, G., Lundwall, A., & Sjoquist, J. (1984) *Int. J. Biol. Macromol. 6*, 195–198.

Österberg, R., Malmensten, B., Nilsson, U., Eggertsen, G., & Kjems, J. (1988) *Int. J. Biol. Macromol. 10*, 15–20.

Perkins, S. J. (1986) *Eur. J. Biochem. 157*, 169–180.

Perkins, S. J. (1988a) *Biochem. J. 254*, 313–327.

Perkins, S. J. (1988b) *New Compr. Biochem. 18B* (Part II), 143–264.

Perkins, S. J., & Weiss, H. (1983) *J. Mol. Biol. 168*, 847–866.

Perkins, S. J., & Sim, R. B. (1986) *Eur. J. Biochem. 157*, 155–168.

Perkins, S. J., Nealis, A. S., & Sim, R. B. (1990) *Biochemistry* (preceding paper in this issue).

Podack, E. R., Kolb, W. P., & Müller-Eberhard, H. J. (1978) *J. Immunol. 120*, 1841–1848.

Reid, K. B. M. (1986) *Essays Biochem. 22*, 27–68.
Smith, C. A., Vogel, C. W., & Müller-Eberhard, H. J. (1982) *J. Biol. Chem. 257*, 9879–9882.
Smith, C. A., Vogel, C. W., & Müller-Eberhard, H. J. (1984) *J. Exp. Med. 159*, 324–329.
Wetsel, R. A., Ogata, R. T., & Tack, B. F. (1987) *Biochemistry 26*, 737–743.
Wetsel, R. A., Lemons, R. S., Le Beau, M., Barnum, S. R., Noack, D., & Tack, B. F. (1988) *Biochemistry 27*, 1474–1482.
Whitehead, A. S., Soloman, E., Chambers, S., Bodmer, W. F., Povey, S., & Fey, G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5021–5025.
Zarbock, J., Gennaro, R., Romeo. D., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 238*, 289–294.
Zuiderweg, E. R. P., Henkin, J., Mollison, K. W., Carter, G. W., & Greer, J. (1988a) *Proteins: Struct., Funct. Genet. 3*, 139–145.
Zuiderweg, E. R. P., Mollison, K. W., Henkin, J., & Carter, G. W. (1988b) *Biochemistry 27*, 3568–3580.

Reprinted from Biochemistry, 1991, *30*.

Neutron and X-ray Scattering Studies on the Human Complement Protein Properdin Provide an Analysis of the Thrombospondin Repeat

Kathryn F. Smith,[‡] Kathleen F. Nolan,[§] Kenneth B. M. Reid,[§] and Stephen J. Perkins*[,‡]

Department of Biochemistry and Chemistry, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K., and MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Received February 15, 1991; Revised Manuscript Received May 22, 1991

ABSTRACT: Properdin is a regulatory glycoprotein of the alternative pathway of the complement system of immune defense. It is responsible for the stabilization of the C3 convertase complex formed between C3b and the Bb fragment of factor B. Neutron and X-ray solution scattering experiments were performed on the dimeric and trimeric forms of properdin. These have R_G values of 9.1 and 10.7 nm, respectively. The scattering curves were compared with Debye sphere modeling simulations for properdin. Good agreements were obtained for models similar to published electron micrographs showing that the properdin trimer has a triangular structure with sides of 26 nm. Such a structure also accounted for sedimentation coefficient data on properdin. Primary structure analyses for mouse and human properdin have shown that this contains six homologous motifs known as the thrombospondin repeat (TSR), which is the second most abundant domain type found in the complement proteins. Sequences for these 12 TSRs were aligned with 19 others found in thrombospondin and the late complement components. Three distinct groups of TSRs were identified, namely, the TSRs found in thrombospondin and properdin, the TSRs mostly found at the N-terminus of the late complement components, and the TSRs found at the C-terminus of the late components. Averaged secondary structure predictions suggested that all three groups contain similar backbone structures with two amphipathic turn regions and one hydrophilic β-strand region. The mean dimensions of the TSRs of properdin in solution were determined to be approximately 4 nm × 1.7 nm × 1.7 nm, showing that these are elongated in structure.

The components of the complement system provide a major nonadaptive immune defense mechanism for its host (Reid, 1986; Law & Reid, 1988). These are activated in response to the challenge of foreign material in plasma. Complement activation proceeds through a series of limited proteolytic steps in one of two largely independent pathways, the classical and alternative pathways. The latter operates via a C3 "tick-over" mechanism, where the molecular structure of the target cell initiates an amplification loop (Lambris, 1990). Properdin is involved in the regulation of the alternative pathway by binding to and stabilizing the C3/C5 convertase complexes, C3b·Bb and $C3b_n$·Bb (Fearon & Austen, 1975; Medicus et al., 1976; Farries et al., 1987), and also by inhibiting the cleavage of C3b by factor I in the presence of factor H as a cofactor by interfering with the ability of factor I to bind C3b (Farries et al., 1988a). Properdin binds to the C-terminal region of the α-chain of C3 between residues 1402 and 1435 (Lambris et al., 1984; Daoudaki et al., 1988). There is also evidence that properdin binds to the Ba domain of factor B in order to facilitate the cleavage by factor D to form the C3 convertase complex, with the concomitent release of Ba (Farries et al., 1988b).

Mouse and human properdin each consist of a single polypeptide chain of 441–442 amino acids (Goundis & Reid, 1988; Nolan & Reid, 1990; Nolan et al., 1991a,b) with an apparent molecular weight of 55 000 when examined under dissociating conditions with or without reduction of disulfide bridges. Human properdin also contains up to 10% by weight of complex-type N-linked carbohydrate, corresponding to one or two oligosaccharides per monomer (Minta & Lepow, 1974; Farries

* To whom correspondence should be addressed.
‡ Royal Free Hospital School of Medicine.
§ University of Oxford.

0006-2960/91/0430-8000$02.50/0

domain found in thrombospondin (the thrombospondin repeat; TSR)[1] have been identified in properdin, together with an N-terminal and C-terminal region (Goundis & Reid, 1988). In human properdin, each TSR is coded by a separate exon, except that the C-terminal segment also has half of the sixth TSR (Nolan et al., 1991b). Each TSR is about 58 residues in size. While circular dichroism experiments did not identify any α-helix or β-sheet structures (Smith et al., 1984), Fourier transform infrared spectroscopy indicated evidence for β-sheet and β-turn structures in the TSRs, and secondary structure predictions suggested that the β-turns could be correlated with the high proportions of Gly, Pro, Cys, and Ser residues in properdin (Perkins et al., 1989).

In plasma, human properdin exists as dimers (P_2; 26% of the total), trimers (P_3; 54%), and tetramers (P_4; 20%) at a total concentration of 5.7 ± 1.0 μg/mL, making it one of the least abundant complement components (Pangburn, 1989). Their specific activities decrease in the order $P_4 > P_3 > P_2$. A polymer of high molecular weight (P_n) had been identified as the "activated" form of properdin; however, this has now been shown to be an artifact of purification (Farries et al., 1987; Pangburn, 1989). Electron microscopy on properdin has shown that the monomer is a flexible rod-like structure of 26 nm in length and 2.5 nm in diameter (Smith et al., 1984). The quaternary structure of the P_2, P_3, and P_4 oligomers is formed by the association of these monomers into cyclic structures, probably by strong noncovalent interactions between the N-terminal and C-terminal regions of each monomer.

The TSR is the second most abundant domain found in the complement components. Since about 80% of properdin is constructed of TSRs, properdin is ideal for structural studies of the TSR. These data are required for structural studies of the late complement components, C6, C7, C8, and C9 (K. F. Smith, R. A. Harrison, and S. J. Perkins, unpublished data). The availability of the TSR sequences in these proteins now permits the computation of an optimal sequence alignment and the classifications of the TSRs and their physical properties. Small-angle scattering is a powerful technique for determining the arrangement of domains in multidomain proteins, especially if the analyses can be constrained by other sequence or structural information (Perkins, 1988a,b). For this, high-flux beam sources are required in view of the low solubility of properdin. Since the structure is studied under conditions close to physiological, this offers the important advantage that the dimensions of electron microscopy images of properdin (measured in vacuo) can be quantitatively tested. This analysis is independently verified by the use of hydrodynamic spheres to simulate experimental sedimentation coefficients. By these means, the arrangements of TSR domains in properdin P_2 and P_3 is determined, together with an assessment of its structure in the late complement components.

Materials and Methods

Solution Scattering Measurements on Properdin. Neutron-scattering data were obtained in four independent sessions on instruments D11 (Ibel, 1976) and D17 at the Institut Laue Langevin, Grenoble, France. Samples were measured at 20 °C in rectangular quartz Hellma cuvettes of 2-mm path length for samples in 80% and 100% ^2H_2O buffers and 1-mm path length for samples in H_2O buffers. These were also used for 280-nm absorbance measurements. Sample and buffer transmissions were measured in all cases for use in data reduction and analysis. In two sessions at instrument D11, the use of sample-to-detector distances of 2.00, 10.00, and 11.00 m with wavelengths λ of 1.001–1.002 nm resulted in a Q range of 0.05–1.1 nm^{-1} (where $Q = 4\pi \sin\theta/\lambda$; 2θ is the scattering angle; λ is the wavelength). Counting times ranged from between 2.5 and 4 h in 0% ^2H_2O buffers at 10 and 11 m for Guinier data (below) to between 5 and 8 min in 100% ^2H_2O buffers at 2 m for data at large Q (concentration = 0.43–1.16 mg/mL). In two sessions on instrument D17, the Q range used was 0.07–1.4 nm^{-1}, based on sample-to-detector distances of 1.40 and 3.46 m, with λ of 1.106 nm and a main beam-to-detector angle of 0°. Samples in 100% ^2H_2O buffers were counted for 1.6 h at 3.46 m and 0.8 h at 1.40 m (c = 0.34–0.42 mg/mL). Data reduction was performed by using standard Grenoble software (RNILS, SPOLLY, RGUIM, and RPLOT; Ghosh, 1989). A cadmium background is first subtracted from each scattering curve. The buffer background run was subtracted from that of the sample run, and the result was normalized for the detector response by using a water run from which an empty cell background had been subtracted.

X-ray scattering data were obtained with the low-angle solution scattering camera at station 8.2 (Towns-Andrews et al., 1989) at the SRS Daresbury, Warrington, U.K. Experiments were performed with beam currents of 118–133 mA and a ring energy of 2.00 GeV in one session. Samples were measured for 20 min at concentrations between 0.28 and 0.85 mg/mL. A 500-channel linear detector at a sample-to-detector distance of 3.83 m was used to give a Q range of 0.05–0.97 nm^{-1}. The Q range was calibrated using fresh, wet, slightly stretched rat tail collagen with a diffraction spacing of 67 nm. Samples were held in Perspex cells with a sample volume of 20 μL, contained within mica windows with a thickness of between 10 and 15 μm, in a temperature-controlled sample holder connected to a water bath at 20 °C. Buffers and samples were measured in alternation for equal times to minimize background subtraction errors. Since the data were prone to such errors, curves were only accepted for subsequent analyses if their Guinier plots were linear and led to reproducible data in repeated measurements. Curves were recorded in 10 time frames for subsequent checks for radiation damage, after which they were averaged. Data reduction was performed by using standard Daresbury software OTOKO (P. Bendall, J. Bordas, M. H. C. Koch, and G. R. Mant, EMBL Hamburg and SERC Daresbury Laboratory, unpublished software) to subtract the buffer runs from those of the samples. The scattering curves were normalized on the basis of an ion counter monitor positioned after the sample, and a detector response was measured for 6 h by using a uniform ^{55}Fe radioactive source.

The dimeric P_2, trimeric P_3, and tetrameric P_4 forms of properdin were prepared as described in Perkins et al. (1989). Neutron data were collected with either a buffer containing 0.15 M NaCl, 0.2 mM EDTA, and 12 mM phosphate, pH 7.4, or 0.5 M NaCl, 0.2 M glycine, and 12 mM phosphate, pH 7.4, on instrument D11, or a buffer containing 0.2 M NaCl and 12 mM phosphate at pH 7.0 on instrument D17. X-ray data were collected with a buffer containing 0.5 M NaCl, 0.2 M glycine, and 12 mM phosphate, pH 7.4. For the neutron experiments, samples were dialyzed at 6 °C into their buffers containing 0%, 80%, or 100% ^2H_2O for at least 36 h with four buffer changes. Sample concentrations c were determined by using an A_{280} coefficient (1%, 1 cm) of 23.9 calculated (Perkins, 1986) from the human sequence. This A_{280} assumed that there were two triantennary complex-type carbohydrate chains (28 residues) per properdin monomer, giving a total

[1] Abbreviations: TSR, thrombospondin repeat.

content by weight. The latter is compatible with a biochemical determination of 9.8% carbohydrate (Minta & Lepow, 1974). By comparison, if the carbohydrates were biantennary, the M_r is 52 800 (8.3% carbohydrate by weight) and the A_{280} coefficient is 24.5. This A_{280} is higher than other estimates of 18 (Reid, 1981) and 17.1 (DiScipio, 1982) and is similar to a value of 23.4 calculated by Nolan and Reid (1990). The properdin samples showed a single band of approximately 55 000 in molecular weight on SDS-PAGE, in reducing conditions both before and after the solution scattering experiments. After solution scattering, the properdin samples did not exhibit significant changes in their original functional activity in an assay involving the lysis of rabbit red blood cells by a reagent dependent upon the addition of purified properdin (Reid, 1981).

Scattering Data Analyses. At small Q values, analyses of the scattering curves by the Guinier equation give R_G, the radius of gyration (which is a measure of macromolecular elongation), and $I(0)$, the forward-scattered intensity at zero Q (Glatter & Kratky, 1982):

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

Absolute M_r values are obtained from the neutron $I(0)/c$ values measured in H_2O buffers, in which systematic errors of measurements are minimized (Jacrot & Zaccai, 1981). Relative M_r values are obtained from the X-ray $I(0)/c$ data provided that the buffer has the same electron density (Kratky, 1963). If one of the molecular dimensions is much larger than the other two, the radius of gyration of the cross section R_{XS} and the cross-sectional intensity at zero angle $[I(Q)Q]_{Q\to 0}$ are obtained in a larger Q range than that used in the R_G analyses:

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

In contrast variation studies, the neutron scattering match point of the protein [at which it is invisible: $I(0)$ is zero] is obtained experimentally from a plot of $\sqrt{I(0)}/cT_s t$ against percent ^2H_2O (where T_s is the sample transmission and t is the path length). This can also be calculated from amino acid and carbohydrate compositions from sequence data, assuming that the protein volume is unhydrated with 10% of nonexchange of main-chain NH protons (Perkins, 1986). Generally, R_G values depend on the scattering density of the solvent, since different constituents within the protein have different scattering densities. This is described by a simplified form of the Stuhrmann equation (Ibel & Stuhrmann, 1976; Perkins, 1988a,b):

$$R_G^2 = R_C^2 + \alpha/\Delta\rho$$

where R_C is the R_G at infinite contrast, $\Delta\rho$ is the difference in scattering density ρ between the solute and solvent, and α is the radial inhomogeneity of scattering density fluctuations within the protein. α is positive if the outermost scattering densities are higher than those nearer the center.

In London, the neutron and X-ray scattering analyses were performed jointly by using a common interactive analysis and plotting program SCTPL3 (Perkins & Sim, 1986; A. S. Nealis and S. J. Perkins, unpublished software). Results were processed as a spreadsheet with Lotus 1-2-3 version 2.01 (Lotus Development, U.K.). Statistical analyses were performed with MINITAB version 6.1 (Ryan et al., 1985). Graphs were created with SIGMAPLOT version 3.1 (Jandel Scientific).

Modeling of Scattering and Hydrodynamic Data for Properdin. Calculations of scattering curves employed small overlapping Debye spheres to follow standard procedures (Glatter & Kratky, 1982) in application to neutron and synchrotron X-ray scattering data (Perkins & Weiss, 1983; Perkins, 1985; Perkins & Sim, 1986). To constrain the simulations, the unhydrated volume (66.8 nm^3) of properdin was calculated by using standard crystallographic volumes for amino acids and carbohydrates (Chothia, 1975; Perkins, 1986) for the 442 amino acid residues (Nolan et al., 1991a) and 28 presumed carbohydrate residues per monomer. This volume was subdivided into 128 cubes with sides of 0.8 nm per monomer in an array of 32 × 2 × 2 cubes for the simulations, and the total length of the monomer was 25.6 nm, in agreement with electron microscopy. The overlapping spheres of the same volume as the cubes were positioned at the centers of the 128 cubes. Neutron data corresponded largely to this unhydrated volume, since the hydration shell that surrounds the macromolecule is invisible by neutron contrast variation (Perkins, 1986). This means that partial specific volumes $\bar{v}$ are larger if calculated from neutron match points instead of by densitometric means. Neutron curve fitting was based on data in 0% ^2H_2O and in 80% and 100% ^2H_2O to correspond to high positive and high negative solute–solvent contrasts, respectively. A trial-and-error procedure was employed to test a wide range of different sphere models in order to determine which ones could be ruled out by the scattering data. Note that this method does not lead to the determination of a unique structure, although all the possible models are constrained by the total volume. The simulated neutron curves were corrected for wavelength spread and beam divergence (Cusack, 1981; Perkins & Weiss, 1983).

X-ray curve fitting (in a high positive solute–solvent contrast) is based on the hydrated volume. The total volume of 88.8 nm^3 per monomer was calculated as the sum of the dry volume and a hydration volume based on a standard hydration of 0.3 g of H_2O/gram of glycoprotein and an electrostricted water molecule volume of 0.0245 nm^3 in place of the free water volume of 0.0299 nm^3 (Perkins, 1986). This model was subdivided into 128 spheres as above, and the final Debye models were generated on the basis of cubes of side 0.885 nm.

Modeling of the sedimentation coefficient $s^\circ_{20,w}$ was based on the hydrated volume of 88.8 nm^3 as in the X-ray curve fitting. The frictional coefficient f was calculated from

$$f = M_r(1 - \bar{v}\rho_{20,w})/(N_a s^\circ_{20,w})$$

where $\rho_{20,w}$ is the density of water at 20 °C and N_a is Avogadro's constant. The $\bar{v}$ of properdin is calculated as 0.710 mL/g from its composition, with the consensus volumes in Perkins (1986) that allow for glycoprotein hydration. Calculations of f were performed by the modified Oseen tensor procedure using the program GENDIA (Garcia de la Torre & Bloomfield, 1977a, 1977b) and a low number of spheres. Hydrodynamic spheres required for the simulations of f do not overlap. In order to compensate for the volume of the void spaces between these nonoverlapping spheres, it was necessary to increase the total volume by resetting the hydration of 0.3 g of H_2O/gram of glycoprotein to 0.39 g (Perkins, 1989). The properdin monomer was formed from a line of 10 spheres with a sphere diameter of 2.644 nm, giving a total length of 26.44 nm, in accord with the length from electron microscopy.

Sequence Alignments and Secondary Structure Predictions. A total of 31 TSR sequences from 11 proteins were aligned manually and also with guidelines provided by the automatic multiple sequence alignment program MULTAL (Taylor, 1990). In the order shown in Figure 5, these were human thrombospondin (3 TSRs) (Lawler & Hynes, 1986; Wolf et al., 1990), mouse thrombospondin (2 of 3 TSRs) (Bornstein et al., 1990), mouse properdin (6 TSRs) (Goundis & Reid, 1988),

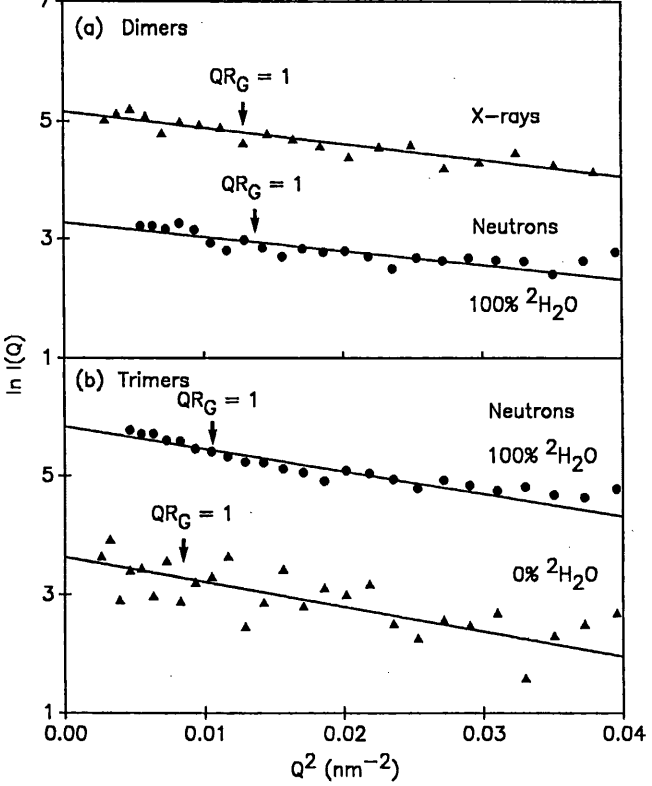


FIGURE 1: Guinier plots of properdin P_2 and P_3. In all cases, the condition $QR_G = 1$ is indicated by an arrow. (a) Data for P_2 in H_2O buffers by X-ray scattering and in ^2H_2O buffers by neutron scattering, at concentrations of 0.95 and 0.51 mg/mL, respectively. (b) Neutron data for P_3 in 100% and 0% ^2H_2O at concentrations of 1.21 and 1.14 mg/mL, respectively.

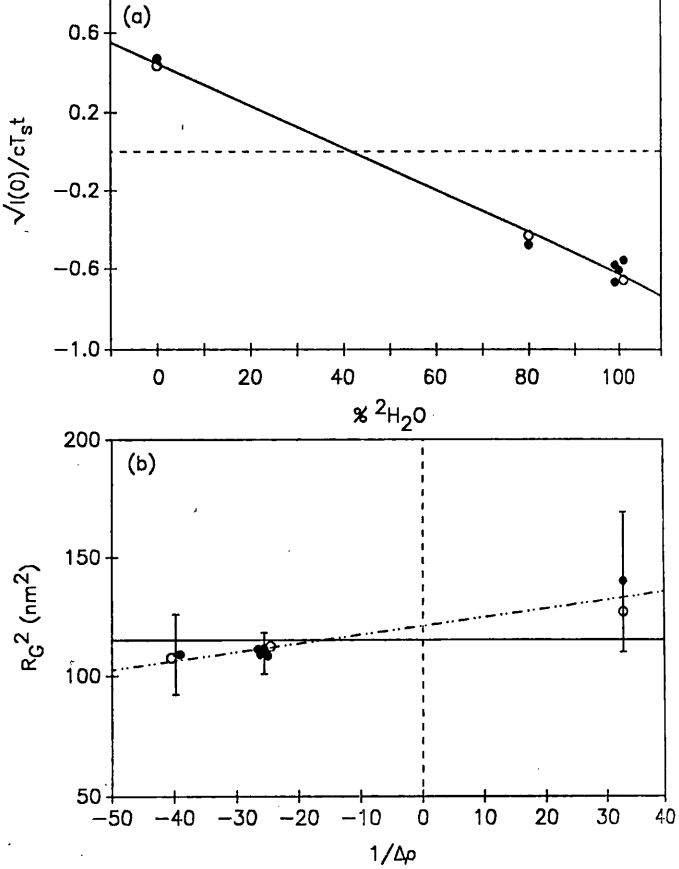


FIGURE 2: Neutron contrast variation analysis of P_3. A concentration range of 0.5–1.2 mg/mL was used. Buffers were either 0.15 M NaCl or 0.2 M NaCl in phosphate (●) (see Materials and Methods) or 0.5 M NaCl, 0.2 M glycine and 12 mM phosphate, pH 7.4 (○). (a) Match point determination of P_3 from a graph of $\sqrt{I(0)/cT_s}$ against the volume percentage of ^2H_2O. Regression gave a match point of 41.8 ± 1.9% ^2H_2O, corresponding to the intersection with the dashed line. (b) Stuhrmann plot of R_G^2 against the reciprocal solute–solvent contrast difference $\Delta\rho^{-1}$. The solid line corresponds to the mean R_G of 10.7 ± 0.5 nm. The dashed–dotted line corresponds to the regressed line and gives an R_C of 11.0 nm at zero $\Delta\rho^{-1}$ (dashed line) and a slope α of 4×10^{-3}. Statistical error bars from the Guinier analyses are shown.

human properdin (6 TSRs) (Nolan et al., 1991a), thrombospondin-related anonymous protein (TRAP) from *Plasmodium falciparum* (1 TSR) (Robson et al., 1988), the late complement components human C6 (3 TSRs) (DiScipio & Hugli, 1989; Chakravarti et al., 1989; Haefliger et al., 1989), human C7 (2 TSRs) (DiScipio et al., 1988), human C8α (2 TSRs) (Rao et al., 1987), human C8β (2 TSRs) (Haefliger et al., 1987; Howard et al., 1987), human C9 (1 TSR) (DiScipio et al., 1984; Stanley et al., 1985), mouse C9 (1 TSR), and trout C9 (2 TSRs) (Stanley & Herz, 1987).

Secondary structure predictions based on the alignment of Figure 5 utilized the Robson and Chou–Fasman methods (Garnier et al., 1978; Chou & Fasman, 1978). The 31 predictions were averaged with the FORTRAN programs PREDRB7 and PREDCF (Perkins et al., 1988). Both programs produced the percentage frequency at which each residue was found in the helix, extended, turn, or coil conformation and the corresponding graphical output, while PREDRB7 also gave the averaged total score in centinats.

RESULTS

Small-Angle Solution Scattering of P_2 and P_3. In order to obtain solution structure data on the arrangement of the six TSRs in properdin, P_2, P_3, and P_4 were studied by neutron scattering in three contrasts at concentrations between 0.3 and 1.2 mg/mL. Scattering data were, however, only accessible for P_2 and P_3 on instruments D11 and D17, since P_4 was found to possess too elongated a structure. In addition, P_3 could only be studied in 100% ^2H_2O buffers on D17 for reason of the lower neutron flux on D17, since only this buffer has a sufficiently low background. Guinier plots for P_2 and P_3 were linear over a Q range of 0.07–0.19 nm^{-1} (Figure 1). This Q range was thus employed for Guinier fits in acceptable QR_G ranges of 0.5–1.4 (P_2) and 0.7–2.0 (P_3).

Analyses of the Guinier neutron intensity $I(0)/c$ or radius of gyration R_G parameters showed that no dependences on protein concentrations or buffer compositions were observed (Figure 2). As a further check of the Guinier data, the molecular weights M_r for P_2 and P_3 were calculated from the $I(0)/c$ values measured in H_2O buffers, on the basis of an A_{280} coefficient of 23.9 (see Materials and Methods). The M_r values of 80000 ± 20000 (P_2, one value) and 150000 ± 30000 (P_3, two values) were in satisfactory agreement with the expected values from the primary sequence of 108 200 and 162 300, respectively. This calculation required the total of scattering lengths $\sum b$ in properdin normalized by the M_r, and this gave a $\sum b/M_r$ of 0.2237 fm in H_2O solutions, which is a typical value for globular proteins. For P_3, contrast variation showed that its match point was 41.8 ± 1.9% ^2H_2O (nine data points; Figure 2a). This agreed well with the prediction of 43.9% ^2H_2O from the amino acid and carbohydrate composition. For P_2, the limited availability of three data points in 0% and 100% buffers gave a comparable match point of 39.3 ± 1.2% ^2H_2O. The linear Guinier plots and the above agreements with known compositions showed that the scattering curves corresponded to monodisperse preparations of P_2 and P_3, as required for structural analyses.

The overall structure of P_3 is analyzed in Figure 2b, which shows the dependence of the neutron R_G data on the contrast.

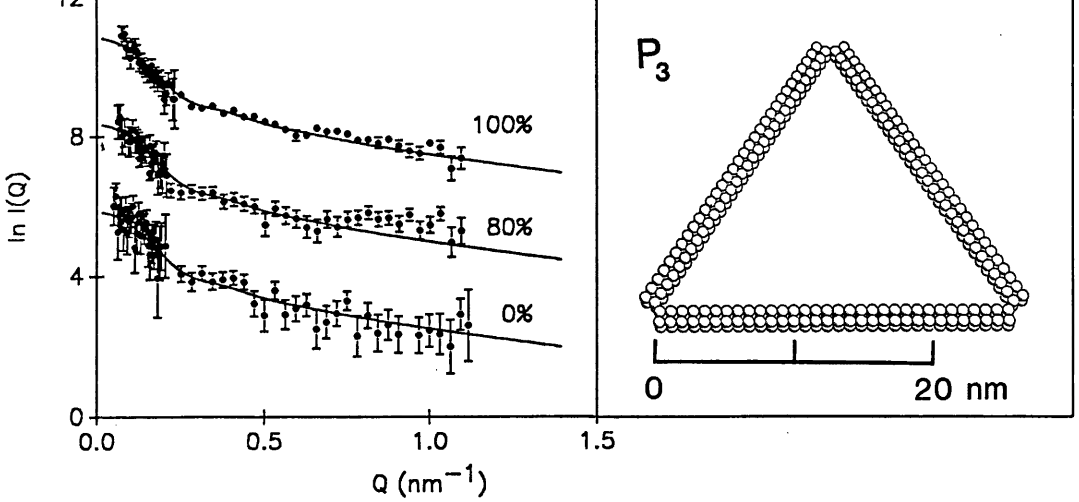


FIGURE 3: Neutron scattering curve simulations for P_3. The curve calculated from the model shown to the right (sphere diameter 0.8 nm; monomer of 32 × 2 × 2 spheres) is compared with the experimental curves measured in 100%, 80%, and 0% ^2H_2O buffers on instrument D11. The simulated curves were corrected for beam divergence and wavelength spread. The samples were at concentrations of 0.43, 0.40, and 1.1 mg/mL, respectively. Guinier R_G analyses were carried out in the Q range of 0.05–0.19 nm^{-1}. Error bars due to data counting statistics are shown.

As a result of low concentrations, the statistical errors of the R_G data were high. Even though an R_C of 11.0 ± 0.1 nm and a positive slope α of (4 ± 1) × 10^{-3} for P_3 could be determined from the nine values in Figure 2b by MINITAB analyses, the final R_G analysis was based on the mean R_G of 10.7 ± 0.5 nm. For P_2, the mean R_G value was 7.3 ± 1.6 nm from three measurements. Calculation of the elongation ratio R_G/R_0 (where R_0 is the R_G of the sphere with the same dry volume as P_2 or P_3) gave values of 3.0 ± 0.7 and 3.8 ± 0.2 for P_2 and P_3, respectively. Since the R_G/R_0 values for typical globular proteins are close to 1.28 (Perkins, 1988b), both P_2 and P_3 possess highly extended structures in solution, in common with many other complement components (Perkins et al., 1990a). In such cases, Guinier analyses of ln $I(Q)Q$ against Q^2 should provide the corresponding cross-sectional parameters $[I(Q)Q]_{Q\to 0}$ and R_{XS}. However, such plots were affected by poor signal–noise or buffer subtraction problems. The R_{XS} was estimated to be of the order of 0.6 ± 0.2 nm from data measured in 100% ^2H_2O buffers.

Synchrotron X-ray experiments were performed with P_2, P_3, and P_4 at concentrations of 0.29–0.89 mg/mL in order to extend the neutron analyses. Scattering curves were accepted for analyses if linear Guinier plots were obtained in Q ranges of 0.05–0.10 to 0.18–0.20 nm^{-1} (Figure 1a). As before, data for P_4 could not be analyzed. The mean R_G values were 9.1 ± 0.6 nm for P_2 (two values) and 11.6 ± 0.9 nm for P_3 (four values). The ratio of $I(0)/c$ values for P_2:P_3 was 1.00:1.49 (±17%), which is the value expected from their relative molecular weights. These data corresponded to measurements in high positive solute–solvent contrasts and as such could be compared with the neutron R_G values in 0% ^2H_2O buffers. Good agreements were found with the neutron R_G of 9 ± 2 nm for P_2 (one value) and 11.5 ± 0.4 nm for P_3 (two values) in H_2O, thus corroborating these analyses. However, there was evidence from time frame analyses of the P_3 data for on-going aggregation during beam exposure (to which the R_G data were insensitive), and these data were therefore not utilized in the curve-fitting analyses below.

Simulations of Scattering Curves for P_3. The use of small Debye spheres in molecular models permitted the interpretation of the scattering data for P_3 that were measured in the larger Q range between 0.2 and 1.1 nm^{-1}. From this, possible arrangements of the TSRs and the properdin monomers in P_3 could be deduced. Electron micrographs (Smith et al., 1984) had reported that properdin existed as cyclic structures, on the basis of a monomer that is 26 nm in length and 2.5–3 nm in width. These dimensions, however, led to a volume that is 2–3 times larger than that calculated from the composition. Here, a model for the monomer was taken to be 26 nm in length, with a width constrained by a dry volume close to 66.8 nm^3 (see Materials and Methods).

P_3 was most successfully modeled as three linear rods (32 spheres spaced 0.80 nm apart) with a circular cross section (2 × 2 spheres) that were arranged as the sides of an equilateral triangle. The dry volume of each rod was 65.5 nm^3. This dry model had an R_G of 10.9 nm, in very good agreement with the observed mean neutron R_G of 10.7 ± 0.5 nm. The simulated neutron curves agreed well with the neutron data in 100%, 80%, and 0% ^2H_2O (Figure 3), with satisfactory R values of 0.026, 0.047, and 0.059, respectively. R is a goodness-of-fit parameter that was defined in Smith et al. (1990). The view of this model is in good accord with observed electron micrographs of P_3.

Three other cyclic and extended models for P_3 were tested against the solution data. The use of an asymmetric cross section of 1 × 4 spheres in the triangular model led to worsened curve fits at large Q values (R of 0.036 in place of 0.026). The use of a circular ring of circumference 96 spheres and cross-section 2 × 2 spheres led to an R_G of 12.6 nm that was significantly larger than the experimental value and to worsened curve fits in the Q range between 0.2 and 0.4 nm^{-1}, although the curve fits at large Q were similar. Y-shaped models were tested in which the sides of the triangle were brought into contact with each other to result in a three-armed structure, such that the TSRs within each monomer were positioned in contact with TSRs in the neighbouring monomer. Here, however, the R_G value was much reduced to 8.1 nm, and the curve fits deteriorated (R of 0.032 in place of 0.026). These poorer curve fits with alternative arrangements of properdin monomers supported the proposal of a triangle model as the likely solution structure for P_3.

Simulations of Sedimentation Data for Properdin. For properdin, sedimentation coefficients $s^\circ_{20,w}$ of 5.0 S (DiScipio, 1982), 5.2 S (Minta & Kunar, 1976), and 5.1–5.3 S (Pensky et al., 1968) have been reported. These authors were not able to resolve P_2, P_3, and P_4 by ultracentrifugation. These $s^\circ_{20,w}$

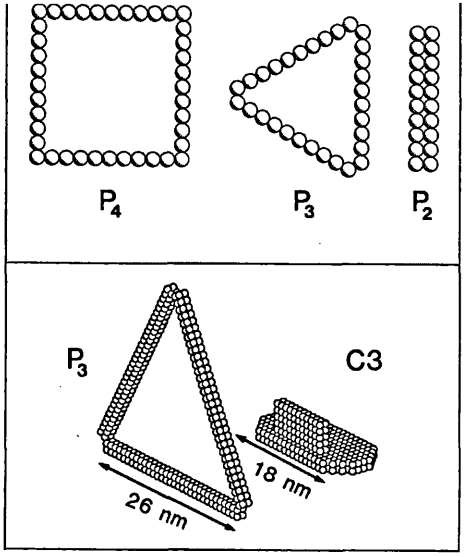


FIGURE 4: Low-resolution models of properdin. (Upper) Hydrodynamic models for tetrameric, trimeric, and dimeric properdin, based on spheres of diameter 2.664 nm. (Lower) The Debye model for trimeric properdin is shown in comparison with a simple two-domain model for complement component C3 from solution scattering analyses (Perkins et al., 1990b). It is stressed that such scattering curve models are only equivalent to the experimentally observed curves. While they do not correspond to a unique structure determination, the models will indicate the relative sizes of P_3 and C3.

elongation of the P_2, P_3, and P_4 structures in solution. Simulations were performed with hydrodynamic sphere models to calculate the $s°_{20,w}$ values of the models determined by solution scattering. The best agreements were based on a monomer length of 26 nm, constructed from a line of 10 spheres of diameter 2.644 nm. The P_2, P_3, and P_4 models were represented by two parallel lines, an equilateral triangle, and a square, respectively (Figure 4a). The predicted $s°_{20,w}$ values were 5.08, 4.98, and 5.21 S in that order. The similarity of these values explained why the three forms cannot be distinguished by ultracentrifugation. These agreements are evidence that the length of the monomer was very close to that of 26 nm determined by solution scattering and electron microscopy. Calculations based on other lengths for the monomer showed that a typical error of ±0.2 S in the simulation (Perkins, 1989) corresponded to changes in the monomer length of ±2 nm.

Classification of TSR Sequences. In order to characterize the properties of a TSR within the properdin monomer, 31 full TSR sequences were aligned in Figure 5. This analysis was improved in comparison to that given in Perkins et al. (1989) because the intron/exon boundaries of the human and mouse thrombospondin genes (Wolf et al., 1990; Bornstein et al., 1990) and the human properdin gene (Nolan et al., 1991b) have been determined. These define the start and end of each TSR (Craik et al., 1982; Traut, 1988). The start of the TSR

```
                    1--------------17      18---24    25---------37  38---------50  51----58        Net
Group 1                                                                                           charge
                     * ** ** *  *  *       *  *       * *  *          *  *    .      *      **
Thrombospondin (human)  SDSADDGWSPWSEWTSCSTS      CGNGIQQ     RGRSCDSLNNR        CEGSSVQTRT  CHIQE CDKRF  -2 (-1)
                          KQDGGWSHWSPWSSCSVT      CGDGVIT     RIRLCNSPSPQMN  GKPCEGEARETKA  CKKDA CPI    +2 (+3)
                            NGGWGPWSPWDICSVT      CGGGVQK     RSRLCNNPTPQFG  GKDCVGDVTENQI  CNKQD CPID   -1
Thrombospondin (mouse)  SDSADDGWSPWSEWTSCSAT      CGNGIQQ     RGRSCDSLNNR        CEGSSVQTRT  CHIQE CDKRF  -2 (-1)
                          KQDGGWSHWSPWSSCSVT      CGDGVIT     RIRLCNSPSPQMN  GKPCEGEARETKA  CKKDA CP     +2 (+3)
Properdin (mouse)           SPQWSAWSLWGPCSVT      CSEGSQL     RHRRCVGRGGQCS  ENVAPGTLEWQLQA CEDQPCCP     -1 ( 0)
                          EMGGWSEWGPWGPCSVT       CSKGTQI     RQRVCDNPAPKC   GGHCPGEAQQSQA  CDTQKTCP      0 (+1)
                          THGAWASWGPWSPRSGS       CLGGAQEPKETRSRSCSAPAPSHQPPGKPCSGPAYEHKA  CSGLPPCP     +3 (+6)
                          VAGGWGPWSPLSPCSVT       CGLGQTL     EQRTCDHPAPRHG  GPFCAGDATRNQM  CNKAVPCP     +1 (+3)
                          VNGEWEAWGKWSDCSRLRM SINCEGTPGQQS    RSRSCGDRKFN    GKPCAGKLQDIRH  CYNIHNCI     +4 (+6)
                          MKGSWSQWSTWSLCTPP       CSPNATRV    RQRLCTPLLPKYPPTVSMVEGQGEKNVTFWGTPRPLCE      +4
Properdin (human)           SPRWSLWSTWAPCSVT      CSEGSQL     RYRRCVGWNGQCS  GKVAPGTLEWQLQA CEDQQCCP     +1
                          EMGGWSGWGPWEPCSVT       CSKGTRT     RRRACNHPAPKC   GGHCPGQAQESEA  CDTQQVCP     +1 (+3)
                          THGAWATWGPWTPCSAS       CHGGPHEPKETRSRKCSAPEPSQKPPGKPCPGLAYEQRR  CTGLPPCP     +4 (+7)
                          VAGGWGPWGPVSPCPVT       CGLGQTM     EQRTCNHPVPQHG  GPFCAGDATRTHI  CNTAVPCP     0 (+3)
                          VDGEWDSWGEWSPCIRRNMKSISCQEIPGQQS    RGRTCRGRKFD    GHRCAGQQQDIRH  CYSIQHCP     +3 (+6)
                          LKGSWSEWSTWGLCMPP       CGPNPTRA    RQRLCTPLLPKYPPTVSMVEGQGEKNVTFWGRPLPRCE      +4
C6 N1 (human)             CFCDHYAWTQWTSCSKT       CNSGTQS     RHRQIVVDKYYQE    NFCEQICSKQETRECNWQR CP     +1 (+3)
TRAP (human)              KTASCGVWDEWSPCSVT       CGKGTRS     RKREIL           HEGCTSEIQEQ     CEEER CP     -2 (-1)

Group 2
                          * * ** ** *  * *       *  *  *     * ***   *** *      *          *       *
C6 N2 (human)             INCLLGDFGPWSDCDP        CIEKQS      KVRSVLRPSQFGG    QPCTEPLVAFQP    CIPSKLCK     +1
C7 N (human)              VNCQWDFYAPWSECNG        CTKTQT      RRRSVAVYGQYGG    QPCVGNAFETQS    CEPTRGCP     +1
C8α N (human)             VTCQLSNWSEWTDCFP        CQDKKY      RHRSLLQPNKFGG    TICSGDIWDQAS    CSSSTTCV      0 (+1)
C8b N (human)             IDCELSSWSSWTTCDP        CQKKRY      RYAYLLQPSQFHG    EPCNFSDKEVED    CVTNRPCR     -1 ( 0)
C9 N (human)              IDCRMSPWSEWSQCDP        CLRQMF      RSRSIEVFGQFNG    KRCTDAVGDRRQ    CVPTEPCE      0
C9 N (mouse)              IDCRMSPWSNWSECDP        CLKQRF      RSRSILAFGQFNG    KSCVDVLGDRQG    CETTQECE     -1
C9 N (trout)              VDCVWSRWSEWTPCNS        CTKIRH      RSRSVEVFGQFVG    KPCQGQPIGEQQA   CTSDAVCE      0 (+1)

Group 3
                          ** ***** * * *            *        * *  **** *  *  *  *  *  *        *
C6 C (human)              VDGQWGCWSSWSTCDA        TYKRS       RTRECNNPAPQRG    GKRCEGEKRQEED   CT           +1
C7 C (human)              VDGGWSCWSSWSPCVQ        GKKT        RSRECNNPPPSGG    GRSCVGETTESTQ   CE           0
C8α C (human)             ADGSWSCWSSWSVCRA        GIQE        RRRECDNPAPQNG    GASCPGRKVQTQA   C            +2
C8b C (human)             IDGKWNCWSNWSSCSG        RRKT        RQRQCNNPPPQNG    GSPCSGPASETLD   CS           +3
C9 C (trout)              EKGNWSCWAAWSGCSG        GKRI        RTRSCNT Q GLS    DATCRGDIVTEDY   C            +1

                            g  g  g                 g           g          g     g

Robson                    bttttcttcttctttttt      ttttbbb     bbbbtbbttttttt   tbbbbttbbbbbbt  ttttttbbb
Chou-Fasman               ttttttbtttbttttttb      ttttbbb     atbbbbttttttt    ttttttaaaaaaa   battttttt
                          <-------------t1--      --->&lt;--      -b1--->&lt;---t2    ---->
```

FIGURE 5: Alignment of 31 sequences for the thrombospondin repeat. A residue conservation greater than 70% within a group of sequences is denoted by ★ above the sequences, which are shaded if these are Cys residues. Residues conserved to better than 90% in all 31 sequences are indicated by two ★ above the sequences, and all are shaded. Putative glycosylation sites are underlined and indicated by g beneath the sequences. IQQ and RDG peptides are shown in the first and fourth sequences, and the last sequence, respectively. The averaged secondary structure prediction by the Robson and Chou–Fasman methods for the 31 sequences is shown beneath the sequences (b, β-sheet; a, α-helix; t, turn; c, coil). The letter is underlined if the prediction occurs in over 75% of the 31 sequences. To the right, the net charge found in each TSR is given assuming that the charged residues are D, E, R, and K; the bracketed figures correspond to the assumption that H is charged also, which, at a blood pH of 7.4, is not likely to be significant.

has now been displaced by 10 residues compared to that in Perkins et al. (1989). The alignment was also improved by the use of the automated multiple sequence alignment program MULTAL (Taylor, 1990). In order to facilitate the averaging of structural properties (Taylor, 1986; Perkins et al., 1988), gaps were minimized and homology was maximized between conserved or conservatively replaced residues. Figure 5 shows that eight residues were greater than 90% conserved in the 31 TSRs (Trp-8, Ser-9, Trp-11, Cys-14, Arg-25, Arg-27, Cys-41, and Cys-51).

The TSRs were readily subdivided into three groups on the basis of different Cys residue locations and different TSR lengths, in an extension of the initial survey of Patthy (1988). The closer similarities of the sequences within each group was confirmed by the MULTAL analyses using a range of sequence gap penalties. Groups 2 and 3 constituted the better defined families of sequences. Insertions or deletions in the group 1 sequences led to five contiguous peptide segments and four gap/insertion regions as indicated in Figure 5, of which the first segment (residues 1–17) exhibited the highest residue conservation in group 1.

A possible connectivity scheme for the three principal disulfide bridges in TSRs may be proposed from the alignment. It is likely that Cys-18 is linked to Cys-57 since both were missing in group 3 when compared to groups 1 and 2. Inspection of the four remaining Cys residues showed suggestively that Cys-41 and Cys-51 were both missing in the last TSR of human and mouse properdin, and therefore these appeared to be bridged in other sequences. Accordingly, Cys-14 may be linked either to Cys-29 (groups 1 and 3) or Cys-3 (group 2 and similar Cys residues in the C6 N1 and TRAP sequences of group 1). This scheme implied that Cys-7 in group 3 is unlinked within the TSR.

A total of nine putative NxT or NxS glycosylation sites were identified. An IQQ motif [factor $XIII_a$ binding; Bale et al. (1985)] was found in human and mouse thrombospondin. An RGD motif [cell surface receptor binding; Yamada (1989)] was found in trout C9 (Stanley & Herz, 1987). The structural motif WSxWS was found in 14 TSRs, occurring once or twice in all the protein sequences of groups 1, 2, and 3 except for those of TRAP and trout C9; this motif exists also in the cytokine hematopoietin receptor superfamily (Cosman et al., 1990). Within several TSRs in group 1 (but not in groups 2 or 3), a motif CSVTCGxGxxxRxR is found that is thought to function as a sulfated glycoconjugate-binding domain in properdin (Holt et al., 1990), although this motif is only conserved in full in the TRAP sequence and in three of the thrombospondin sequences.

The average of the 31 secondary structure predictions was computed, based on the Robson (β biased) and Chou–Fasman methods. Closely similar predictions to those reported previously (Perkins et al., 1989) were obtained. The total percentages of the four or three conformations present were identical. By the Robson method, these were α-helix, 0%; β-sheet, 38%; turn, 57%; and coil, 5%. The Chou–Fasman method gave α-helix, 16%; β-sheet, 18%; and turn, 66%. Again, two regions of high turn propensities could be identified (t_1 and t_2; residues 1–22 and 32–42) and one region of β-sheet (b_1; residues 23–31). While the two turn regions contained both hydrophilic and hydrophobic residues, nine out of the ten residues in the b_1 region were found to be hydrophilic. Further examination of the individual structure predictions supported the subdivision of TSRs into three groups. Unlike groups 1 and 3, residues 1–5 in group 2 were predicted as β-sheet. The predictions suggested that the b_1 region had different lengths, namely residues 23–28 in group 1, residues 19–33 in group 2, and residues 22–26 in group 3.

Discussion

Structure of Monomers in Properdin. Despite the low properdin concentrations, the use of high-flux beam sources (in particular at the Institut Laue Langevin) led to successful measurements of solution scattering curves. These were the most fruitful for the trimer P_3, which is the most abundant oligomer. Curve fitting was most successfully carried out in terms of open triangular structures in solution conditions close to physiological, with overall dimensions that were highly similar to the length of 26 nm reported on the basis of electron microscopy. Flexibility between the TSRs in each properdin monomer cannot, however, be ruled out by the scattering data.

One biochemical insight into the properties of properdin has been gained in this study. The calculated A_{280} coefficient of 23.9 from the primary structure is one-third higher than previously reported by biochemical determinations, as also reported in the calculation of Nolan and Reid (1990). Tests of this revised value led to satisfactory M_r values by solution scattering. This revision also implies that previous estimates of the concentration of properdin in plasma, initially 20–25 μg/mL (Minta et al., 1973) and then 5.7 μg/mL (Pangburn, 1989), should now be revised to 4.2 μg/mL.

Properdin has a high isoelectric point greater than 9.5. This is readily correlated with the excess of positive charges (R, K) over negative charges (D, E) in five of the six TSRs in the human sequence (Figure 5). It is possible that these charges may offer an explanation of these open cyclic structures, since the TSRs may repel each other. The three forms of properdin, P_2, P_3, and P_4, once isolated do not reequilibrate into a mixture. The structural studies suggest that the monomers associate between alternate N-terminal and C-terminal peptides at each end of each monomer. Inspection of their sequences suggest that the stability of these oligomers may be mediated by ionic interactions between up to 22 opposite charges in these regions.

The significance of the P_3 structure is shown by comparisons with the sizes of other proteins with which it interacts (Figure 4). The monomer length is 26 nm. This associates with C3b, which is of maximum length 18 nm (Perkins & Sim, 1986; Perkins et al., 1990b; Ikai et al., 1990). C3b interacts with properdin within a 34-residue segment of C3b (Daoudaki et al., 1988), which if in a maximally extended conformation would be of length 12 nm. These considerations show that it is reasonable to expect that for steric reasons more than one C3b molecule can interact with each of P_2, P_3, and P_4. In fact, Pangburn (1989) showed that the specific activity of properdin decreases in the order $P_4 > P_3 > P_2$.

Structure of TSRs in Properdin. Structural information on TSRs have been deduced from this study. Since TSRs exist in single copies in the late complement components, and since the multiple TSRs in thrombospondin and five out of the six TSRs in properdin are encoded by separate exons, it can be inferred that each TSR constitutes an independently folded protein structure. While TSRs are the second most abundant complement sequence motif, analysis shows that these may be subdivided into three groups. In contrast, the most abundant complement motif, the short complement (or consensus) repeat, is readily considered as a single group of sequences (Perkins et al., 1988).

The TSRs in these three groups exhibit distinct differences in Cys contents, lengths, and overall charges. Since similar locations for gaps and insertions could be proposed in the 31 TSR sequences (Figure 5), the protein backbone structure is expected to be similar in all three groups, even though the TSR

this premise, only minor differences were seen between the groups in the structure predictions. The principal characteristics of the three groups are summarized:

(i) Group 1, which included the thrombospondin and properdin sequences and interestingly the first N-terminal TSR of C6, contained complex arrangements of between four and seven Cys. These occurred most frequently at Cys-14, -18, -29, -41, -51, and -57. The TSR was 58 residues in length, 19 of which were conserved to better than 70%. The properdin TSRs generally displayed two to four positive charges on average, depending on the ionization state of about two His residues per TSR.

(ii) Group 2 corresponded to the N-terminal TSRs of the late complement components except for that above in C6. These contained six Cys as in group 1, except that Cys-29 is replaced by Cys-3. Most notably, the well-conserved Trp-5, Cys-29, and Gly-43 residues in group 1 were poorly conserved in group 2, while several well-conserved residues, 33–37, in group 2 do not have their counterpart in group 1. The average length of the group 2 TSR was 55 residues, 23 of which were conserved to better than 70%. These TSRs were generally uncharged.

(iii) Group 3 corresponded to the C-terminal TSRs in the late components. This group contained five Cys residues at Cys-7, -14, -29, -41, and -51. This group exhibited the highest residue conservation and the shortest sequences; the average length was 47 residues, 25 of which were conserved to better than 70%. While Trp-5, Cys-29, and Gly-43 are now well conserved as in group 1, many conserved residues in group 3 have no counterpart in group 1. These TSRs exhibited low positive charges.

To interpret the solution scattering analyses, the dry volume of each TSR can be calculated from amino acid compositions (Perkins, 1986). The average volume of the TSRs of group (excepting TRAP), group 2, and group 3 without carbohydrates were 8.2 ± 0.6, 7.6 ± 0.6, and 6.3 ± 0.3 nm^3, respectively. If the TSR were spherical, its diameter would be 2.5, 2.4, and 2.3 nm, respectively, in the three groups. The average volume for human and mouse properdin of the N- and C-terminal polypeptides was found to be 6.64 ± 0.18 and 4.16 ± 0.08 nm^3, respectively. If spherical, these would have diameters of 2.4 and 2.0 nm, respectively.

The length of a properdin monomer is 26 nm. If each of the six group 1 TSRs found in properdin are assumed to be spherical in shape, together with the N-terminal and C-terminal peptides, the sum of the eight diameters gives a total length of 19 nm if all eight structures were positioned in a linear arrangement. Since this length is notably shorter than the observed length, this shows that the TSR has an elongated structure. Its length can range between 3.25 and 4.3 nm, depending on the presumed structure of the N-terminal and C-terminal regions (Perkins et al., 1989). On the basis of the scattering curve fits at large Q for properdin, the cross-sectional dimensions of the TSR are compact at average values of 1.7 nm × 1.7 nm. On this scale, each TSR is equivalent to 16 spheres (Figure 3). Similar overall dimensions are expected for the TSRs found in the late complement components.

ACKNOWLEDGMENTS

We thank the Wellcome Trust for support and the Science and Engineering Research Council for use of the Synchrotron Research Source (Daresbury) and the Institut Laue Langevin (Grenoble). Mr. A. S. Nealis, Ms. J. Parsons, and Drs. W. Bras, A. de Geyer, D. Goundis, and J. Torbet are gratefully thanked for support and assistance. A preliminary account of this work was given at the 655th Biochemical Society Meeting, July 17–19, 1990, Aberdeen, U.K.

Registry No. Properdin, 11016-39-0.

REFERENCES

Bale, M. D., Westrick, L. G., & Mosher, D. F. (1985) *J. Biol. Chem. 260*, 7502–7508.

Bornstein, P., Alfi, D., Devarayalu, S., Framson, P., & Li, P. (1990) *J. Biol. Chem. 265*, 16691–16698.

Chakravarti, D. N., Chakravarti, B., Parra, C. A., & Müller-Eberhard, H. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2799–2803.

Chothia, C. (1975) *Nature (London) 254*, 304–308.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Cosman, D., Lyman, S. D., Idzerda, R. L., Beckmann, M. P., Park, L. S., Goodwin, R. G., & March, C. J. (1990) *Trends Biochem. Sci. 15*, 265–270.

Craik, C. S., Sprang, S., Fletterick, R., & Rutter, W. J. (1982) *Nature (London) 299*, 180–182.

Cusack, S. (1981) *J. Mol. Biol. 145*, 539–541.

Daoudaki, M. A., Becherer, J. D., & Lambris, J. D. (1988) *J. Immunol. 140*, 1577–1580.

DiScipio, R. G. (1982) *Mol. Immunol. 19*, 631–635.

DiScipio, R. G., & Hugli, T. E. (1989) *J. Biol. Chem. 264*, 16197–16206.

DiScipio, R. G., Gehring, M. R., Podack, E. R., Kan, C. C., Hugli, T. E., & Fey, G. H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7298–7302.

DiScipio, R. G., Chakravarti, D. N., Müller-Eberhard, H. J., & Fey, G. H. (1988) *J. Biol. Chem. 263*, 549–560.

Farries, T. C., & Atkinson, J. P. (1989) *J. Immunol. 142*, 842–847.

Farries, T. C., Finch, J. T., Lachmann, P. J., & Harrison, R. A. (1987) *Biochem. J. 243*, 507–517.

Farries, T. C., Lachmann, P. J., & Harrison, R. A. (1988a) *Biochem. J. 252*, 47–54.

Farries, T. C., Lachmann, P. J., & Harrison, R. A. (1988b) *Biochem. J. 253*, 667–675.

Fearon, D. T., & Austen, K. F. (1975) *J. Exp. Med. 142*, 856–863.

Garcia de la Torre, J., & Bloomfield, V. A. (1977a) *Biopolymers 16*, 1747–1761.

Garcia de la Torre, J., & Bloomfield, V. A. (1977b) *Biopolymers 16*, 1779–1793.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Ghosh, R. E. (1989) *Internal Publication 89GH02T*, Institut Laue Langevin, Grenoble, France.

Glatter, O., & Kratky, O. (Eds.) (1982) *Small-Angle X-ray Scattering*, Academic Press, New York.

Goundis, D., & Reid, K. B. M. (1988) *Nature (London) 335*, 82–85.

Haefliger, J.-A., Tschopp, J., Nardelli, D., Wahli, W., Kocher, H. P., Tosi, M., & Stanley, K. K. (1987) *Biochemistry 26*, 3551–3556.

Haefliger, J.-A., Tschopp, J., Vial, N., & Jenne, D. E., (1989) *J. Biol. Chem. 264*, 18041–18051.

Holt, G. D., Pangburn, M. K., & Ginsburg, V. (1990) *J. Biol. Chem. 265*, 2852–2855.

Howard, O. M. Z., Rao, A. G., & Sodetz, J. M. (1987) *Biochemistry 26*, 3565–3570.

Ibel, K. (1976) *J. Appl. Crystallogr. 9*, 269–309.

Ibel, K., & Stuhrmann, H. B. (1975) *J. Mol. Biol. 93*, 225–266.

Ikai, A., Nishigai, M., Saito, A., Sinohara, H., Muto, Y., & Arata, Y. (1990) *FEBS Lett. 260*, 291–293.
Jacrot, B., & Zaccai, G. (1981) *Biopolymers 20*, 2413–2426.
Kratky, O. (1963) *Prog. Biophys. Chem. 13*, 105–173.
Lambris, J. D. (Ed.) (1990) *Curr. Top. Microbiol. Immunol. 153*, 1–251.
Lambris, J. D., Alsenz, J., Schulz, T. F., & Dierich, M. P. (1984) *Biochem. J. 217*, 323–326.
Law, S. K. A., & Reid, K. B. M. (1988) *Complement*, IRL Press, Oxford, U.K.
Lawler, J., & Hynes, R. O. (1986) *J. Cell Biol. 263*, 5495–5498.
Medicus, R. G., Götze, O., & Müller-Eberhard, H. J. (1976) *J. Exp. Med. 144*, 1076–1093.
Minta, J. O., & Lepow, I. H. (1974) *Immunochemistry 11*, 361–368.
Minta, J. O., & Kunar, E. S. (1976) *J. Immunol. 116*, 1099–1104.
Minta, J. O., Goodkofsky, I., & Lepow, I. H. (1973) *Immunochemistry 10*, 341–350.
Nolan, K. F., & Reid, K. B. M. (1990) *Biochem. Soc. Trans. 18*, 1161–1162.
Nolan, K. F., Schwaeble, W., Kaluz, S., Dierich, M. P., & Reid, K. B. M. (1991a) *Eur. J. Immunol. 21*, 771–776.
Nolan, K. F., Goundis, D., Kaluz, S., & Reid, K. B. M. (1991b) *Biochem. J.* (submitted).
Pangburn, M. K. (1989) *J. Immunol. 142*, 202–207.
Patthy, L. (1988) *J. Mol. Biol. 202*, 689–696.
Pensky, J., Hinz, C. F., Todd, E. W., Wedgwood, R. J., Boyer, J. T., & Lepow, I. H. (1968) *J. Immunol. 100*, 142–158.
Perkins, S. J. (1985) *Biochem. J. 228*, 13–26.
Perkins, S. J. (1986) *Eur. J. Biochem. 157*, 169–180.
Perkins, S. J. (1988a) *Biochem. J. 254*, 313–327.
Perkins, S. J. (1988b) *New Compr. Biochem. 18B*, 143–264.
Perkins, S. J. (1989) In *Dynamic Properties of Biomolecular Assemblies* (Harding, S. E. & Rowe, A. J., Eds.) Chapter 15, pp 226–245, Royal Society of Chemistry, London.
Perkins, S. J., & Weiss, H. (1983) *J. Mol. Biol. 168*, 847–866.
Perkins, S. J., & Sim, R. B. (1986) *Eur. J. Biochem. 157*, 155–168.
Perkins, S. J., Haris, P. I., Sim, R. B., & Chapman, D. (1988) *Biochemistry 27*, 4004–4012.
Perkins, S. J., Nealis, A. S., Haris, P. I., Chapman, D. Goundis, D., & Reid, K. B. M. (1989) *Biochemistry 28*, 7176–7182.
Perkins, S. J., Smith, K. F., & Nealis, A. S. (1990a) *Biochem. Soc. Trans. 18*, 1151–1154.
Perkins, S. J., Nealis, A. S., & Sim, R. B. (1990b) *Biochemistry 29*, 1167–1175.
Rao, A. G., Howard, O. M. Z., Ng, S. C., Whitehead, A. S., Colten, H. R., & Sodetz, J. M. (1987) *Biochemistry 26*, 3556–3564.
Reid, K. B. M. (1981) *Methods Enzymol. 80*, 143–150.
Reid, K. B. M. (1986) *Essays Biochem. 27*, 27–68.
Robson, K. B. H., Hall, J. R. S., Jennings, M. W., Harris, T. J. R., Marsh, K., Newbold, C. I., Tate, V. E., & Weatherall, D. J. (1988) *Nature (London) 335*, 79–82.
Ryan, B. F., Joiner, B. L., & Ryan, T. A. (1985) *Minitab Handbook*, 2nd ed., PWS-Kent Publishing Company, Boston.
Smith, C. A., Pangburn, M. K., Vogel, C.-W., & Müller-Eberhard, H. J. (1984) *J. Biol. Chem. 259*, 4582–4588.
Smith, K. F., Harrison, R. A., & Perkins, S. J. (1990) *Biochem. J. 267*, 203–212.
Stanley, K. K., & Herz, J. (1987) *EMBO J. 6*, 1951–1957.
Stanley, K. K., Kocher, H. P., Luzio, J. P., Jackson, P., & Tschopp, J. (1985) *EMBO J. 4*, 375–382.
Taylor, W. R. (1986) *J. Mol. Biol. 188*, 233–258.
Taylor, W. R. (1990) *Methods Enzymol. 183*, 456–474.
Towns-Andrews, E., Berry, A., Bordas, J., Mant, G. R., Murray, P. K., Roberts, K., Sumner, I., Worgan, J. S., Lewis, R., & Gabriel, A. (1989) *Rev. Sci. Instrum. 60*, 2346–2349.
Traut, T. W. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2944–2948.
Wolf, F. W., Eddy, R. L., Shows, T. B., & Dixit, V. M. (1990) *Genomics 6*, 685–691.
Yamada, K. M. (1989) *Curr. Opin. Cell Biol. 1*, 956–963.

13. Busby, T. F. & Ingham, K. C. (1987) *Biochemistry* **26**, 5564–5571
14. Busby, T. F. & Ingham, K. C. (1988) *Biochemistry* **27**, 6127–6135
15. Thielens, N. M., Aude, C. A., Lacroix, M. B., Gagnon, J. & Arlaud, G. J. (1990) *J. Biol. Chem.* in the press
16. Busby, T. F. & Ingham, K. C. (1989) *Complement Inflammation* **6**, 318–319
17. Medved, L. V., Busby, T. F. & Ingham, K. C. (1989) *Biochemistry* **28**, 5408–5414
18. Zaccai, G., Aude, C. A., Thielens, N. M. & Arlaud, G. J. (1990) *FEBS Lett.* **269**, 19–22
19. Matsumoto, M., Nagaki, K., Kitamura, H., Kuramitsu, S., Nagasawa, S.& Seya, T. (1989) *J. Immunol.* **142**, 2743–2750
20. Lacroix, M. B., Aude, C. A., Arlaud, G. J. & Colomb, M. G. (1989) *Biochem. J.* **257**, 885–891
21. Arlaud, G. J., Gagnon, J., Villiers, C. L. & Colomb, M. G. (1986) *Biochemistry* **25**, 5177–5182
22. Perkins, S. J. & Nealis, A. S. (1989) *Biochem. J.* **263**, 463–469
23. Boyd, J., Burton, D. R., Perkins, S. J., Villiers, C. L., Dwek, R. A. & Arlaud, G. J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3769–3773
24. Tschopp, J., Villiger, W., Fuchs, H., Kilchherr, E. & Engel, J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7014–7018
25. Ziccardi, R. J. & Cooper, N. R. (1977) *J. Immunol.* **118**, 2047–2052
26. Siegel, R. C. & Schumaker, V. N. (1983) *Mol. Immunol.* **20**, 53–66
27. Lakatos, S. (1987) *Biochem. Biophys. Res. Commun.* **149**, 378–384
28. Strang, C. J., Siegel, R. C., Phillips, M. L., Poon, P. H. & Schumaker, V. N. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 586–590
29. Perkins, S. J., Villiers, C. L., Arlaud, G. J., Boyd, J., Burton, D. R., Colomb, M. G. & Dwek, R. A. (1984) *J. Mol. Biol.* **179**, 547–557
30. Perkins, S. J. (1985) *Biochem. J.* **228**, 13–26
31. Ziccardi, R. J. (1985) *Mol. Immunol.* **22**, 489–494
32. Villiers, C. L., Arlaud, G. J. & Colomb, M. G. (1984) *Eur. J. Biochem.* **140**, 421–426
33. Colomb, M. G., Arlaud, G. J. & Villiers, C. L. (1984) *Philos. Trans. R. Soc. London B* **306**, 282–292
34. Poon, P. H., Schumaker, V. N., Phillips, M. L. & Strang, C. J. (1983) *J. Mol. Biol.* **168**, 563–577
35. Schumaker, V. N., Hanson, D. C., Kilchher, E., Phillips, M. L. & Poon, P. H. (1986) *Mol. Immunol.* **23**, 557–565
36. Perkins, S. J. (1989) *Behring Inst. Mitt.* **84**, 129–141

Received 15 June 1990

Molecular modelling strategies in application to complement

STEPHEN J. PERKINS,* KATHYRN F. SMITH and ADAM S. NEALIS

Departments of Biochemistry and Chemistry, and Protein and Molecular Biology, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K.

Introduction

The complement cascade is composed of a family of plasma proteins whose role is to eliminate foreign particles or organisms [1]. The classical pathway of complement is activated via C1 by antibody recognition, while the alternative pathway is triggered if the foreign object has a suitable recognition site to assist the amplified self-activation of complement. In all, the cascade contains about 20 plasma glycoproteins. The cascade operates through a series of sequential activations by limited proteolysis in response to recognition signals. These ultimately lead to the formation of the membrane attack complex and cellular lysis. Cell surface proteins also participate in the cascade. These activations require a high degree of biological control. Both the activation and control processes involve highly specific protein–protein interactions. Now that all the complement components have been sequenced, structural studies can be pursued in order to understand their functional properties.

Almost all the complement components are constructed from multiple domain structures. Protein structural studies therefore have to proceed at two levels. The overall arrangement of domains within the whole structure has to be determined, as well as the detailed molecular structure within each domain. Protein crystallography is ordinarily the most powerful method of choice. That very few of the complement components have been crystallized to date is thought to be the result of their flexible multi-domain structures and their heavy glycosylation. In addition, the crystallization buffers can often be non-physiological, and the complement components can be susceptible to random proteolysis during crystallization.

The most powerful method for the study of overall domain structures at present is X-ray and neutron solution scattering [2, 3]. This multiparameter technique investigates structures under physiological conditions. Molecular masses and the degree of oligomerization can be determined, large conformational changes can be identified and the results can be compared quantitatively with those from other physical methods (Table 1 and Fig. 1). It is more advantageous than hydrodynamic sedimentation measurements which offer only a measure of macromolecular elongation. Even though electron microscopy has the ability to view structures directly, it has disadvantages in being prone to artefacts caused either by the need to work *in vacuo*, the use of stains, electron beam damage, magnification errors or possible difficulties in interpretation.

The study of individual molecular structures within each domain is dependent on the domain in question. For example, crystal structures are known for the serine proteases (trypsin [4], chymotrypsin, etc.), the serine protease inhibitors (α_1-antitrypsin, [5]) and the anaphylatoxins (C3a, [6]). These can be used as the starting point for atomic models of other structurally homologous domains in the complement components. If the domain is small enough, two-dimensional n.m.r. in combination with distance geometry, energy and molecular dynamics calculations can yield a three-dimensional structure. This has been successfully applied to the epidermal growth factor [7] and the short consensus repeat or short complement repeat (SCR) [8].

Both these high-resolution methods can be powerfully complemented by the use of protein structure prediction methods [9] and Fourier transform infrared (f.t.i.r.) spectroscopy [10]. These methods are more general in that they rely only on available abundant sequence data and standard preparative-scale quantities of the complement components in solution without limitations of molecular mass.

(i) Secondary structure predictions based on primary sequences and the Robson or Chou–Fasman methods are accurate to only 50–60% on a residue-by-residue basis. Both methods are based on statistical analyses of known protein

Abbreviations used: SCR, short consensus/complement repeat; TSR, thrombospondin repeat; vWF, von Willebrand factor; f.t.i.r., Fourier transform infrared.

*To whom correspondence and requests for reprints should be addressed, in the Department of Biochemistry and Chemistry.

Table 1. *Structures of some of the complement components*

Name	Molecular mass*	Number of domains	Radius of gyration, R_G (nm)†	R_G/R_0 ratio	Sedimentation coefficient, $s^0_{20,w}$ (S)	Frictional ratio, f/f_o
C1q	457000	12: head, stalk	12.8	3.2	10.2	1.91
$C1r_2$	172000	12: SP, SCR, EGF, RS	8	2.8	7.1	1.45
C1s	79000	6: SP, SCR, EGF, RS	8	3.6	4.5	1.36
$C1r_2C1s_2$	330000	24: SP, SCR, EGF, RS	17	4.7	8.7	1.82
C1	787000	36: C1q + $C1r_2C1s_2$	12.6	2.6	15.2–16.2	1.79
α_1-antitrypsin	52000	1: Serpin	2.61	1.3	3.3–3.45	1.32
$C\bar{1}$ inhibitor	71000	2: Serpin, *N*-terminus	4.85	2.3	3.67–3.8	1.54
C3, C3u, C3(a + b)	187000	3: C3a, C3c, C3dg	5.1	1.7	7.0–7.3	1.40
C3c	137000	1: C3c	4.8	1.8	5.5–6.7	1.37
C3dg	39000	1: C3dg	2.7	1.5	2.6	1.35
C4, C4u, C4(a + b)	197000	3: C4a, C4c, C4d	4.9	1.6	—	—
$C4c_2$	291000	2: C4c	4.9	1.4	—	—
C4d	41000	1: C4d	2.7	1.5	—	—
C5	194000	3: C5a, 'C5c', 'C5d'	4.9	1.6	—	—
C4BP	491000	56‡: SCR	13	3.2	10.7–11.2	1.89
Properdin (dimer)	108000	12: TSR	9§	3.7§	—	—
(trimer)	162000	18: TSR	11§	3.9§	5.2–5.5	1.95

*Molecular masses, R_G/R_0 ratios and f/f_o ratios are calculated from sequences.
†From neutron scattering in conditions of infinite contrast except for C4BP and properdin, X-ray data for several of these proteins support these determinations.
‡Studied in the form assuming seven α-chains.
§Provisional values.

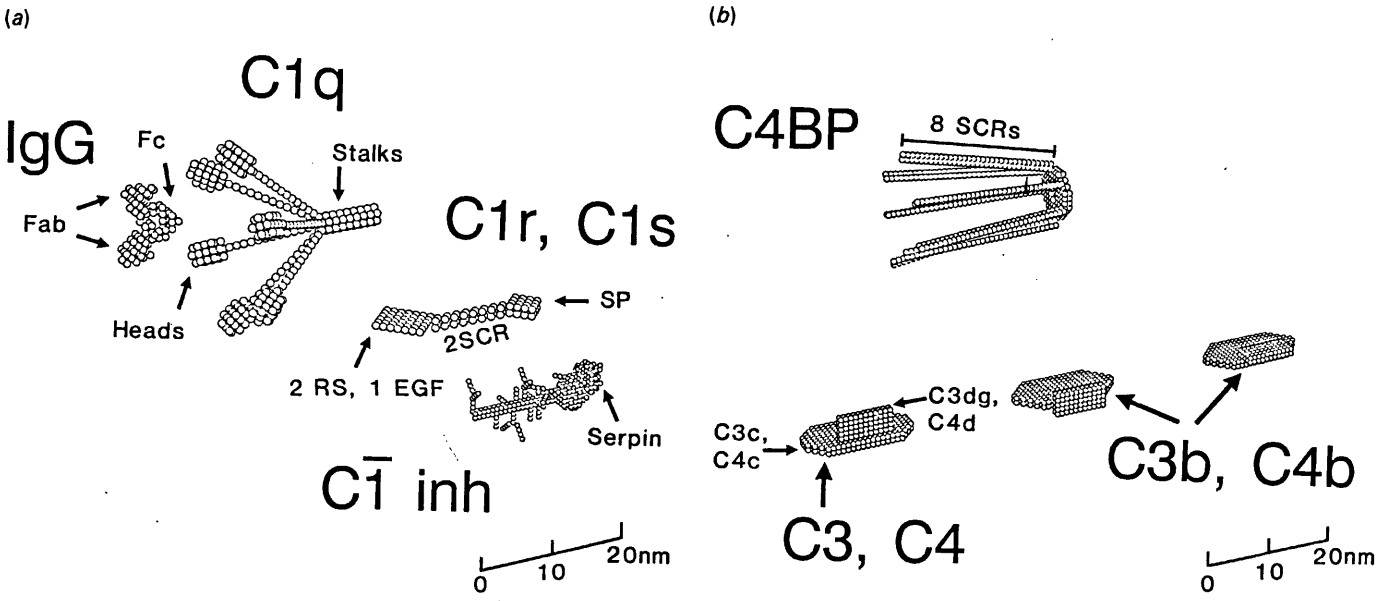


Fig. 1. *Comparison of models from X-ray and neutron scattering curve analyses*

(*a*) Proteins involved in the activation and control of C1. Scattering models for C1q, C1r, C1s and $C\bar{1}$ inhibitor [11] are compared with a schematic model for the Fab and Fc domains in IgG. The Fab domains interact with antigenic surfaces. The Fc domains in aggregates of IgG interact with the C1q heads. $C1r_2C1s_2$ binds to the C1q stalks to give the C1 complex, and is autoactivated to $C\bar{1}r_2C\bar{1}s_2$; this leads to the cleavage of C2 and C4 of complement. Control of $C\bar{1}$ is mediated by the $C\bar{1}$ inhibitor, in which the serpin domain binds to the serine proteinase domain. (*b*) Activation and control of C3 and C4. C3 and C4 are the two most abundant complement components and scattering shows that they are structurally homologous to each other as expected from sequence data [13, 14]. These structures can be represented by two large domains C3c/C3dg and C4c/C4d that can be modelled by ellipsoids as shown (although the accuracy of these models should not be overrated). These form C3b and C4b in that order on removal of the small C3a and C4a anaphylatoxins. A large conformational change relative to C3 and C4 is detectable by solution scattering, the magnitude of which is represented by the two possible models shown for C3b and C4b. Factor I cleaves C3 (in the presence of factor H as cofactor) and C4 (in the presence of the C4b binding protein: C4BP) into C3c and C3dg, and C4c and C4d. Both factor H and C4BP are constructed from highly extended arrangements of SCR domains, as shown in the solution model for C4BP [12].

crystal structures. Improvements can therefore be obtained by averaging many predictions based on the alignment of a large number of sequences. This average will emphasize significant features that arise in the predictions, and will minimize errors from sequence misalignments.

(ii) The prediction analyses can be tested by f.t.i.r. spectroscopy. Study of the f.t.i.r. amide I band reflects primarily the C=O stretching vibration of amide groups coupled to the in-plane NH bending and CN stretching modes, and is therefore sensitive to the local environment of N—H hydrogen bonding with C=O groups, i.e. the protein secondary structure. The advantages of this technique over c.d. are (a) the absence of problems in defining a spectral base line; (b) different bands in the f.t.i.r. spectrum can be clearly resolved; (c) it is not affected by the presence of high amounts of disulphide bridges found in the SCR and the thrombospondin repeat (TSR); (d) amide proton exchange rates and large conformational changes can be followed.

Neutron and X-ray solution scattering

At the smallest scattering angles, Guinier plots give the radius of gyration, R_G. This is a measure of macromolecular elongation. The R_G/R_0 ratio compares the R_G with the R_G of a sphere of the same volume. For 47 globular proteins with molecular masses between 10^4 and 10^6 kDa, the average R_G/R_0 is approximately constant at 1.28 ± 0.10 [2]. In distinction, Table 1 shows that large deviations with R_G/R_0 values close to 2 or more occur in the complement components, even in those with as few as 2 or 3 domains. This shows that, in conditions close to physiological, the complement components possess highly extended domain structures. Independent confirmation of this is obtained from sedimentation coefficients and the frictional ratios, f/f_o, calculated from these, even though these are less sensitive than the corresponding R_G/R_0 ratios (Table 1).

Guinier analyses in combination with the sample concentration lead to molecular masses, M_r. Neutron scattering was used to confirm, that the M_r of the heavily glycosylated $C\bar{1}$ inhibitor structure is closer to 71 000 kDa than the value of about 100 000 kDa observed by SDS/PAGE [11]. In the case of C4BP, an erroneous absorption coefficient at 280 nm was identified by X-ray scattering [12]. The C4c fragment of C4 was shown to be dimeric and not monomeric as believed originally [13].

Analyses of the scattering curves in the full range of scattering angles accessible by experiment is carried out by modelling using small spheres in the Debye equation (Fig. 1). By a trial-and-error procedure, these models are used to calculate curves that are equivalent to the experimental scattering curves. Comparison of the sphere models with the sizes of the individual domains identifies the disposition of the domains within each complement component. Even though solution scattering does not identify unique structures, the modelling is in practise much constrained by the knowledge of which domains are present in the complement component under study and their sizes, and is thus a highly useful method.

Gross conformational changes can be identified by scattering, as exemplified by the conversion of C3 and C4 to C3b and C4b respectively in Fig. 1(*b*), where the two domains in the latter form a more compact arrangement [13, 14].

Sequence analyses and Fourier transform infrared spectroscopy

The most abundant domain found in the complement components is the SCR. Twelve of the soluble and membrane-bound complement components contain 97 SCRs, totalling 145 copies in all on allowing for the presence of multiple copies in heptameric C4BP. The alignment of 101 sequences found in complement and non-complement proteins showed that its average length is 61 residues. Of these, 19 are totally or conservatively conserved in over half the sequences. Secondary structure predictions [15] proposed a clear alternation of four β-strands and four β-turns between residues 21–51 in the 61-residue repeat in the majority of sequences. Factor H contains wholly 20 SCRs. That these predictions are significant was indicated by the observation of two clear bands at 1638 cm^{-1} and 1680 cm^{-1} in the second derivative f.t.i.r. spectrum of 15–30 mg/ml factor H in ^2H_2O which can be assigned to anti-parallel β-strands. Likewise a third band at 1661 cm^{-1} could be assigned to turn structures.

The second most abundant complement domain is the TSR. Six complement components are found to contain 15 TSRs, of which multiple copies exist in the dimeric, trimeric and tetrameric forms of properdin. The first analysis of up to 23 aligned sequences shows that this is on average 58 residues in length, with 21 positions being totally or conservatively conserved in 60% or more of sequences. In properdin, about 80% of the structure contains TSRs. F.t.i.r. spectroscopy of properdin in ^2H_2O shows that there are two major bands at 1676 cm^{-1} and 1633 cm^{-1}, which can be assigned in that order to substantial amounts of β-turn and β-strand structures, even though the protein concentrations are low at 0.7–1.2 mg/ml. The predictions suggest that there are 57–66% β-turn and 18–38% β-sheet structures in a typical TSR, which is consistent with the f.t.i.r. spectra [16].

Similar experiments are in progress on the von Willebrand factor (vWF) domain, which corresponds to the type A domain found in the plasma protein with the same name, and which is present in four complement components. Here the alignment of 46 available similar sequences (average length 202 residues) revealed that residue conservation is generally low, but that a clear pattern of alternating amphiphatic α-helices and hydrophobic β-strands could be detected (S. J. Perkins & K. F. Smith, unpublished work). The alignment of four sequences found in a domain specific to C1r and C1s (RS) shows that this is composed of 13% α-helix, 30% β-sheet and 37% turn [17].

F.t.i.r. spectroscopy can also be used to detect large conformational changes. It was shown that the less stable native form of the serpin domain has a much reduced content of α-helix and β-sheet than the more stable but physiologically inactive reactive-centre-split form. It was proposed that the stabilization of the split form is achieved by a more complete hydrogen bonding in its secondary structure than the native form [18].

Conclusions

At the present time, over half of the soluble complement components have been successfully analysed by solution scattering. Work is in progress with the remaining components. Molecular masses range from 39 000 to 787 000 kDa, and the number of domains range between 1 and 56. Fig. 1 and Table 1 illustrate that these generally have highly extended domain structures.

The joint use of both the predictive and f.t.i.r. techniques make it possible to identify secondary structures in the SCR, TSR and vWF domains. This will assist in the location of the strategic amino acid residues that may usefully be targeted for site-specific mutagenesis and antigenic peptide experiments designed to explore the function of each domain. Other domains are likewise under investigation.

We gratefully acknowledge support from the Wellcome Trust and our collaborations with Professor D. Chapman FRS, Dr P. I. Haris, Dr R. A. Harrison, Dr K. B. M. Reid and Dr R. B. Sim. Solution

scattering beamtime sessions were taken at the S.E.R.C.-supported facilities at the Synchrotron Radiation Source, Daresbury, U.K. and the Institut-Laue-Langevin, Grenoble, France.

1. Law, S. K. A. & Reid, K. B. M. (1988) *Complement*, IRL Press, Oxford
2. Perkins, S. J. (1988) in *New Comprehensive Biochemistry* (Neuberger, A. & Van Deenen, L. L. M., eds.), vol. 18B, part 2, pp. 143–264, Elsevier, Amsterdam
3. Perkins, S. J. (1988) *Biochem. J.* **254**, 313–327
4. Bode, W. & Schwager, P. (1975) *J. Mol. Biol.* **98**, 693–717
5. Löbermann, D., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984) *J. Mol. Biol.* **177**, 531–556
6. Huber, R., Scholze, H., Pâques, E. P. & Deisenhofer, J. (1980) *Hoppe-Seyler's Z. Physiol. Chem.* **261**, 1389–1399
7. Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H. & Sheard, B. (1987) *Nature (London)* **327**, 339–341
8. Baron, M., Willis, A. C., Bazar, K., Campbell, I. D., Sim, R. B. & Day, A. J. (1989) *Complement Inflamm.* **6**, 311
9. Argos, P. (1989) in *Protein Structure: a Practical Approach* (Creighton, E. T., ed.), pp. 169–189, IRL Press, Oxford
10. Lee, D. C. & Chapman, D. (1986) *Biosci. Rep.* **6**, 235–256
11. Perkins, S. J., Smith, K. F., Amatayakul, S., Ashford, D., Rademacher, T. W., Dwek, R. A., Lachmann, P. J. & Harrison, R. A. (1990) *J. Mol. Biol.* **214**, 751–763
12. Perkins, S. J., Chung, L. P. & Reid, K. B. M (1986) *Biochem. J.* **223**, 779–807
13. Perkins, S. J., Nealis, A. S. & Sim, R. B. (1990) *Biochemistry* **29**, 1167–1175
14. Perkins, S. J. & Sim, R. B. (1986) *Eur. J. Biochem.* **157**, 155–168
15. Perkins, S. J., Haris, P. I., Sim, R. B. & Chapman, D. (1988) *Biochemistry* **27**, 4004–4012
16. Perkins, S. J., Nealis, A. S., Haris, P. I., Chapman, D., Goundis, D. & Reid, K. B. M. (1989) *Biochemistry* **28**, 7176–7182
17. Perkins, S. J. (1989) *Behring Inst. Mitt.* **84**, 129–141
18. Haris, P. I., Chapman, D., Harrison, R. A., Smith, K. F. & Perkins, S. J. (1990) *Biochemistry* **29**, 1377–1380

Received 23 July 1990

A new approach to designing active analogues of proteins

JULIA A. EMBER,* NILS L. JOHANSEN† and TONY E. HUGLI*

**Department of Immunology, Research Institute of Scripps Clinic, La Jolla, CA 92037, U.S.A. and †Novo Industries A/S, Bagsvaerd, Denmark DK-2880*

Over the past ten years numerous peptide analogues of C3a have been designed with the purpose of mimicking C3a function and understanding the receptor interactions at the molecular level. The initial observation identified a synthetic octapeptide (Ala-Ser-His-Leu-Gly-Leu-Ala-Arg) that exhibited 1% of the activity of C3a (human) [1]. The analogue peptide derived from the *C*-terminus of C3a (C3a 70–77) defined the effector site on the natural factor. The biological specificity of the octapeptide was exactly that of C3a. It was later found that elongation of the *C*-terminal fragment C3a from 8 to 21 residues led to marked enhancement of activity [2]. Based on the tendency of the 21-residue synthetic C3a peptide to assume a helical conformation in dilute trifluoroethanol, it was concluded that the conformation of this analogue contributed to the high activity. In fact, synthetic C3a (57–77) exhibited nearly equivalent spasmogenic activity to the intact natural factor C3a.

Strategic design of the C3a analogue peptides to enhance helix content and amphipathy served to maintain activity while the incorporation of helix-breaking residues, such as proline, significantly reduced activity [3]. These results led to the conclusion that regular conformation in the *C*-terminal region of C3a was requisite for optimal activity. At the time, we surmised that the *C*-terminal helix may contribute to a configuration of the amino acid side chains at the local L-G-L-A-R effector site leading to a proper presentation of these active site residues to the C3a receptor. This conclusion was supported by the fact that many of the residues in the 21-residue fragment of C3a could be substituted without major consequence to activity.

A more recent observation by Gerardy-Schahn *et al.* [4] was that attachment of hydrophobic groups to the *N*-terminus of short *C*-terminal analogue peptides of C3a led to marked enhancement in activity. This unexpected phenomenon was interpreted as being caused by a membrane effect. It was proposed that the hydrophobic group of the C3a peptide analogues insert themselves into the lipid phase of the biomembrane and thereby alter the association between peptide and receptor from a three-dimensional to a two-dimensional reaction, as has been proposed for supra-agonists luteinizing hormone-releasing hormone by Schwyzer *et al.* [5]. Insertion of the C3a analogue peptides in the biomembrane would increase the effective concentration of the molecule and would result in only an apparent enhanced activity. This was a reasonable explanation for the enhanced activity of a peptide analogue of C3a (Nap-Ahx-Y-R-R-G-R-A-A-A-L-G-L-A-R) that has little resemblance to the fully active natural C3a fragment C3a 57–77 (C-N-Y-I-T-E-L-R-R-Q-H-A-R-A-S-H-L-G-L-A-R).

We questioned whether the hydrophobic group at the *N*-terminus of C3a analogues enhanced activity in a non-specific manner as proposed by Gerardy-Schahn *et al.* [4] or perhaps this effect had a specific component. When a series of peptides was synthesized based on the *C*-terminal C3a sequence and the hydrophobic tryptophan group was substituted for Ahx, Nap or Fmoc, we observed a progressive (10- to 15-fold) increase in the activity as these peptides were lengthened from a decapeptide to a pentadecapeptide. We also noted that two *N*-terminal tryptophanyl groups were more effective than just one. Biological data corresponding to approximately 50 C3a analogue peptides has recently been reported [6].

Peptides longer than 15 or 16 residues exhibited decreased activity indicating an optimal length for maximally active peptides. This result was inconsistent with a non-specific membrane effect. We concluded that C3a analogue peptides with an *N*-terminal hydrophobic group of optimal size represented a unique structure that mimics the essential receptor interactions of C3a, and that the hydrophobic group contributes directly to receptor binding. The model peptide W-W-G-K-K-Y-R-A-S-K-L-G-L-A-R has several residue replacements (underlined) compared with the C3a sequence (see Fig. 1), but the main feature is the *N*-terminal Trp-Trp-sequence. This model peptide is 10–15 times as active as human C3a and exhibits only C3a specific functions. When the conformation of the *C*-terminal region of C3a was elucidated by X-ray analysis [7], helical structure was evident from residue 49 to 72 in C3a. Analysis of C3a in solution by n.m.r. [8] indicated that the *C*-terminal helix terminates earlier than in the crystalline structure, becoming irregular at about residue 65 to 66. The helical region just *N*-terminal to the residue 65 defines an amphiphilic structure