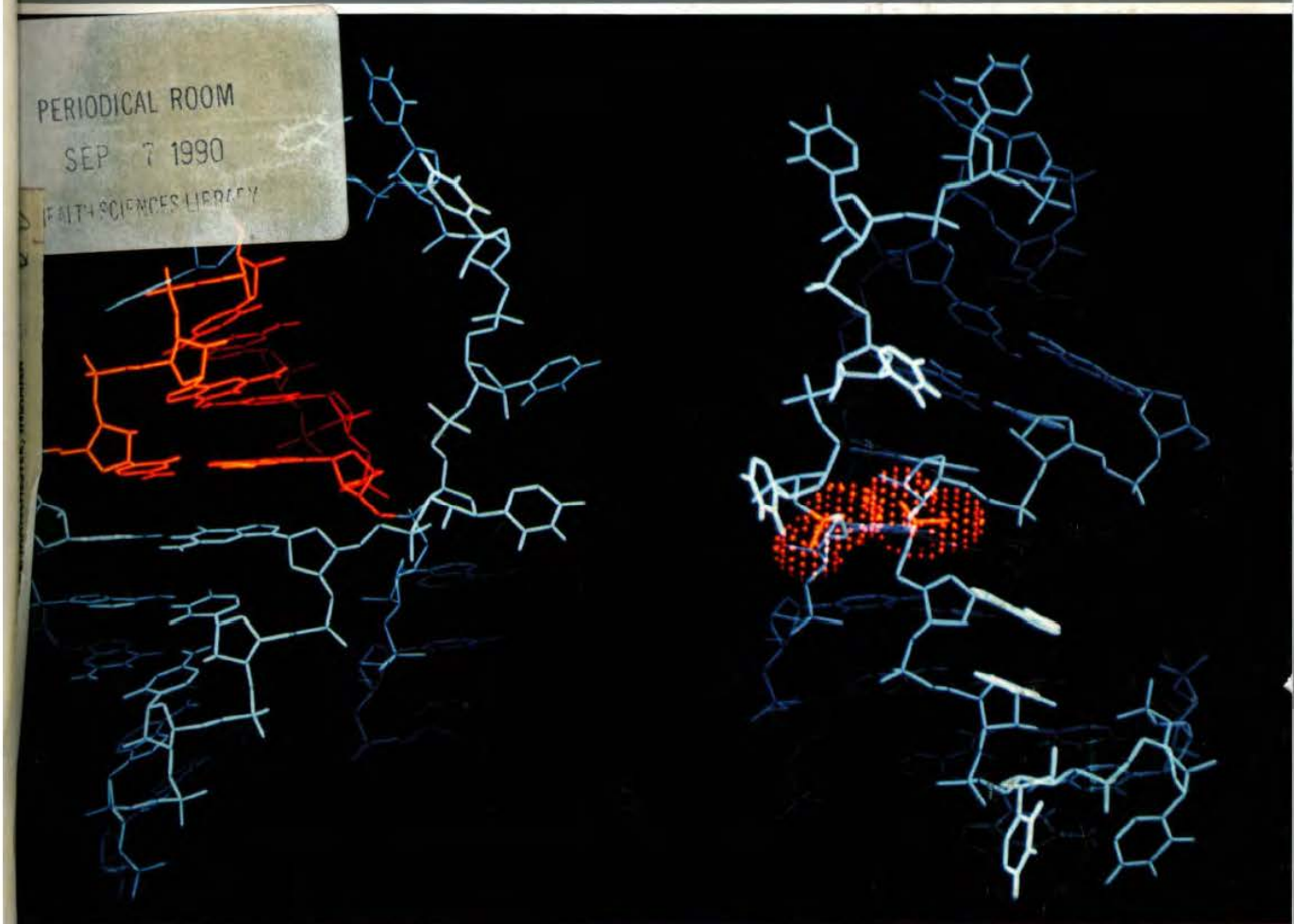
Volume 214 Number 3 5 August 1990

Journal of MOLECULAR BIOLOGY



ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
JMOBAK 214 (3) 613–798 ISSN 0022–2836

Journal of Molecular Biology

JOURNAL OF MOLECULAR BIOLOGY: ISSN 0022–2836. Volumes 211–216, 1990, published twice a month on the 5th and 20th by Academic Press at 24–28 Oval Road, London NW1 7DX, England. Annual subscription price including postage: £768 U.K. and U.S.$1464 overseas. Personal subscription rate: £233 U.K. and U.S. $350 overseas. Subscription orders should be sent to Academic Press Limited, Foots Cray, Sidcup, Kent DA14 5HP, U.K. (Tel: 081–300 3322). Send notices of changes of address to the publisher at least 6–8 weeks in advance, including both old and new addresses.

Second class postage rate paid at Jamaica, NY 11431, U.S.A.

Air freight and mailing in the U.S.A. by Publications Expediting Inc., 200 Meacham Avenue, Elmont, NY 11003, U.S.A.

U.S.A. POSTMASTERS: send change of addresses to JOURNAL OF MOLECULAR BIOLOGY, c/o Publications Expediting, Inc., 200 Meacham Avenue, Elmont, NY 11003, U.S.A.
Printed in U.K.

Volume 214, Number 3

Contents

Cover Illustration: Conformation of an RNA Pseudoknot. Puglisi *et al.*, *J. Mol. Biol.* **214**, 437–453.

Origination by BPCC Whitefriars Ltd, Tunbridge Wells,
and printed and bound in Great Britain by BPCC Wheatons Ltd, Exeter
This journal is printed on acid-free paper

J. Mol. Biol. (1990) **214**, 751–763

Two-domain Structure of the Native and Reactive Centre Cleaved Forms of C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ inhibitor has the same M_r and structure as the native molecule. The length of C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ inhibitor after the scattering experiments had an unaltered specific activity against active-site titrated plasmin.

(b) *Carbohydrate analysis of C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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/

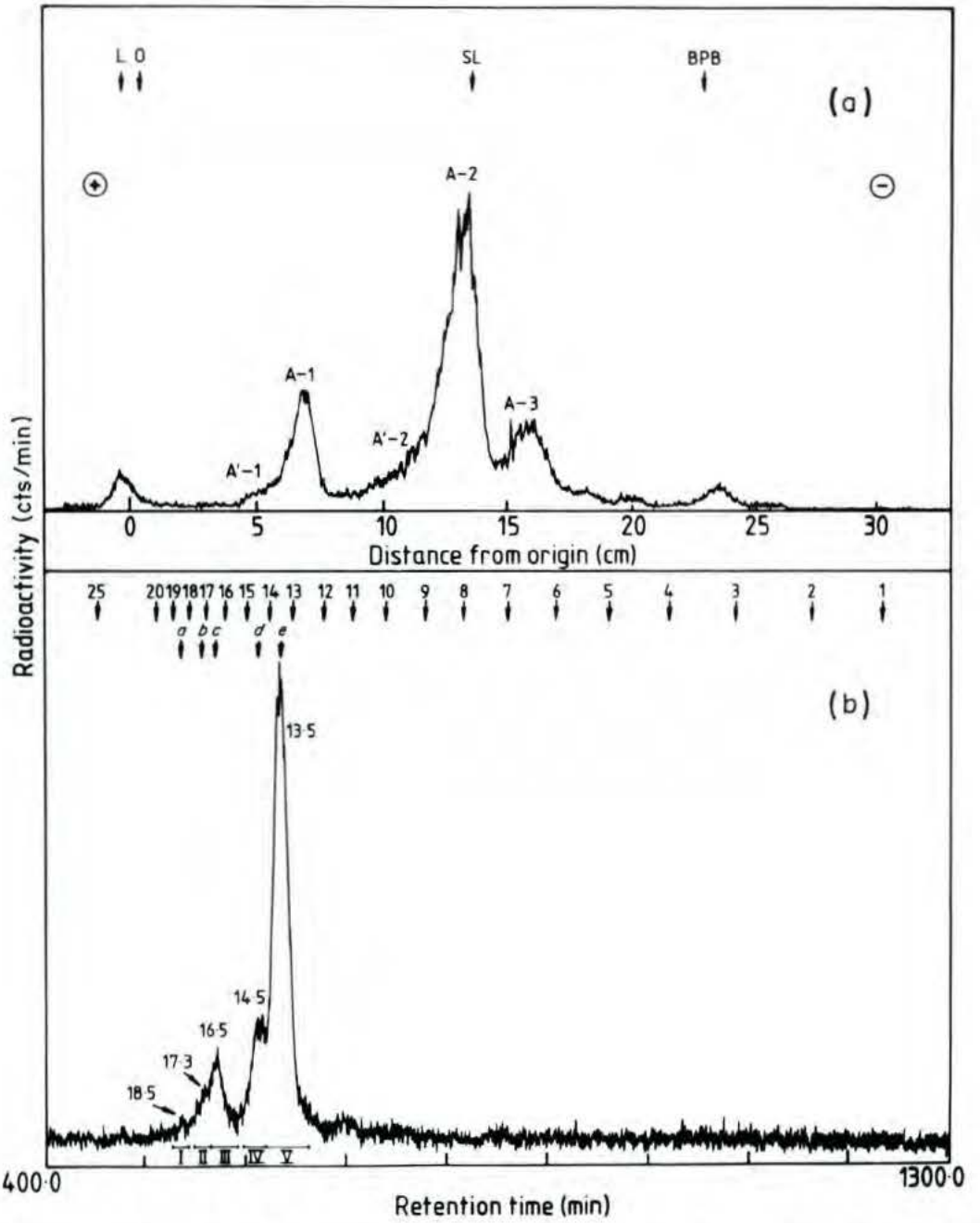


Figure 1. N-linked oligosaccharide analysis of C1̄ inhibitor. (a) High-voltage paper electrophoresis of C1̄ inhibitor N-linked oligosaccharides. Tritium-labelled oligosaccharides were subjected to high-voltage paper electrophoresis (80 V cm^{-1}) in pyridine/acetic acid/water (3 : 1 : 387, by vol.) buffer (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 C1̄ 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 C1̄ inhibitor (Harrison, 1983) or the product of cell-free-translation of C1̄ 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 C1̄ 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 C1̄ inhibitor are calculated for different molecular weights of C1̄ 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 C1̄ inhibitor*

Neutron scattering was applied to both native and split C1̄ 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 C1̄ 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 C1̄ 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-

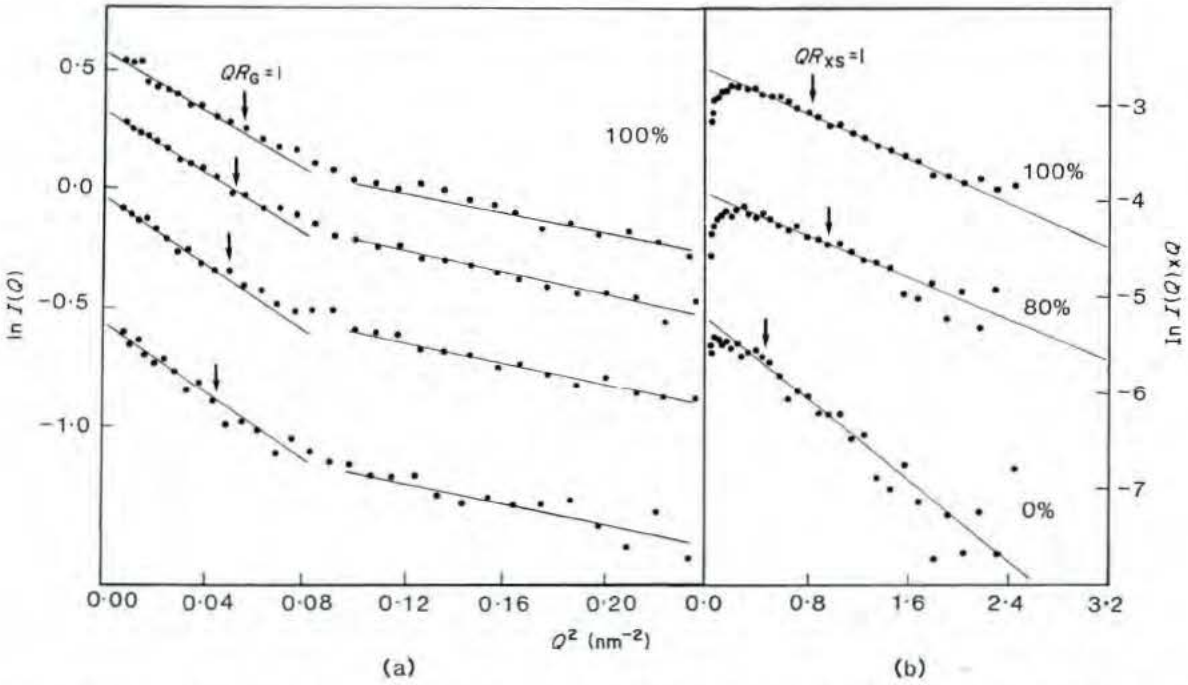


Figure 2. Neutron Guinier analyses for native C1̄ inhibitor. (a) A dilution series in 100% ^2H_2O buffers is depicted. At low Q, the concentration dependence of the Guinier R_G curves is shown. From top to bottom, the sample concentrations are 7·9 mg/ml, 5·9 mg/ml, 4·1 mg/ml and 2·2 mg/ml, with R_G values 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.

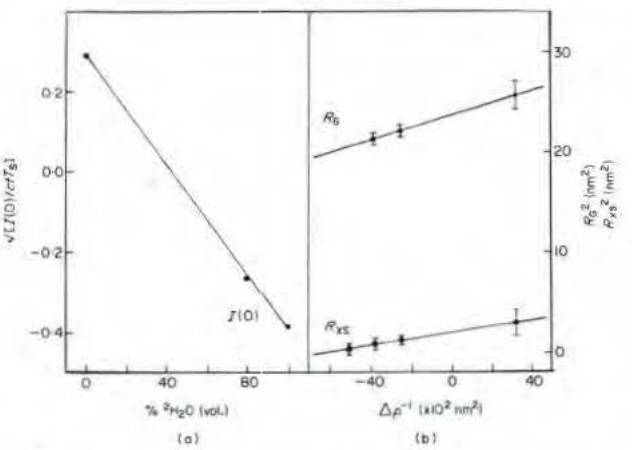


Figure 3. Contrast variation analyses for native and split CĪ inhibitor. (a) The matchpoint determination of CĪ inhibitor, using the $I(0)/cT_s$ data extrapolated to zero concentration, gives 42·3% ^2H_2O. The corresponding matchpoint determination using the mean $[I(Q)\times Q]_{Q\to 0}$ values measured 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 $R_{G\cdot C}$ 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 $R_{XS\cdot C}$ 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 CĪ inhibitor by sedimentation (Haupt *et al.*, 1970). Since the serpin domain of CĪ 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 $A^{1\%}_{1\,cm,280}$ 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 CĪ 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 CĪ 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 CĪ inhibitor. At infinite contrast the $R_{G\cdot C}$ of 4·85(±0·01) nm corresponds to an elongation ratio $R_{G\cdot C}/R_0$ of 2·27, where R_0 is the R_G of the sphere with the same dry volume as CĪ inhibitor. The value of this ratio is 1·35 for the homologous serpin α_1-antitrypsin (Smith *et al.*, 1990). CĪ inhibitor is therefore much more elongated than α_1-antitrypsin. The difference is attributed to the large N-terminal domain of CĪ 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 CĪ 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 $R_{G\cdot C}$, 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 CĪ inhibitor at large* Q

Since CĪ 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 CĪ 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 $R_{XS\cdot C}$ 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 CĪ inhibitor is thus strongly dominated by the carbohydrate component.

Table 2
Properties of the native and split forms of human C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ 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 C1̄ inhibitor was provided by the position of Cys406 and Cys183 of C1̄ inhibitor in the crystal structure of α_1-antitrypsin. These form disulphide bridges with Cys101 and Cys108 in the N-terminal domain. Cys406 and

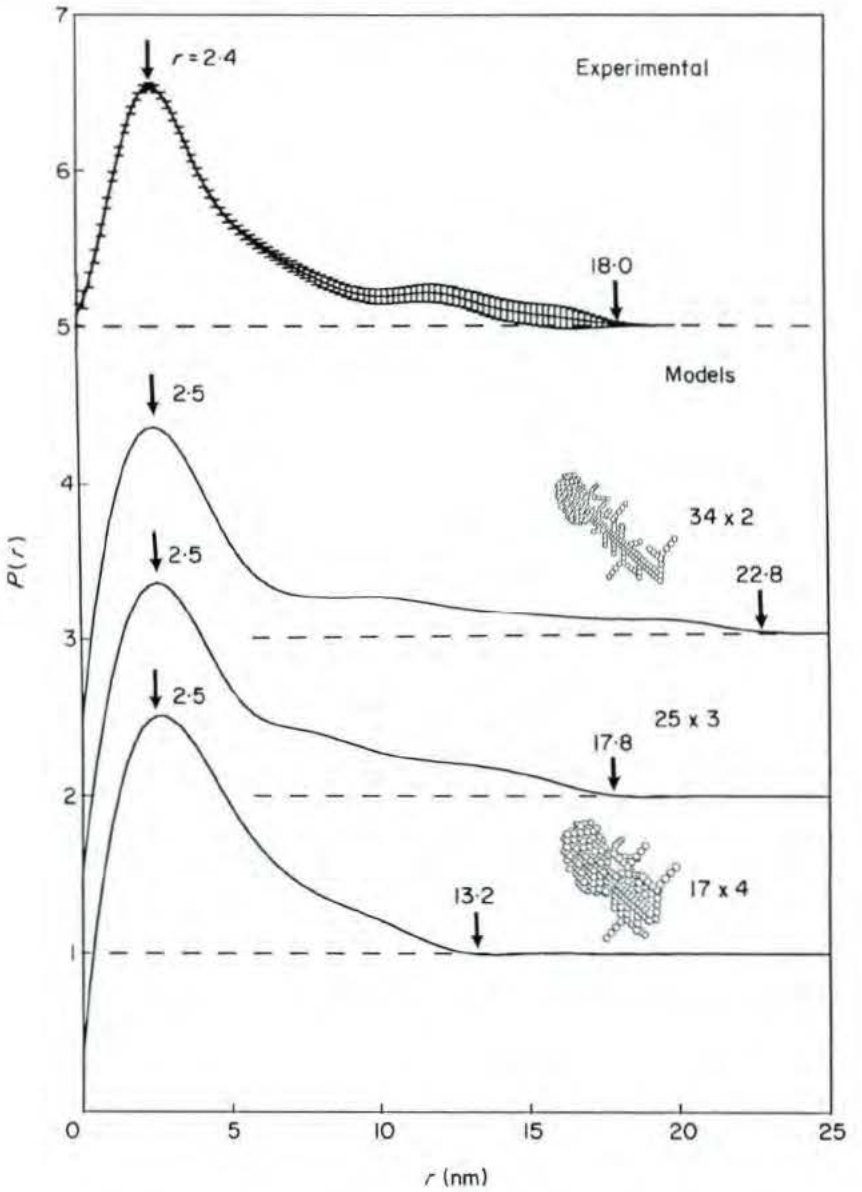


Figure 4. Indirect transformation (Glatter, 1982) of the scattering curves $I(Q)$ for C1 inhibitor. The experimental $P(r)$ curve corresponds to data on native C1 inhibitor at 14 mg/ml in 100% ^2H_2O, based on 69 $I(Q)$ points extending to $Q = 3{\cdot}1\ nm^{-1}$, fitted using 10 B-splines, and transformed 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

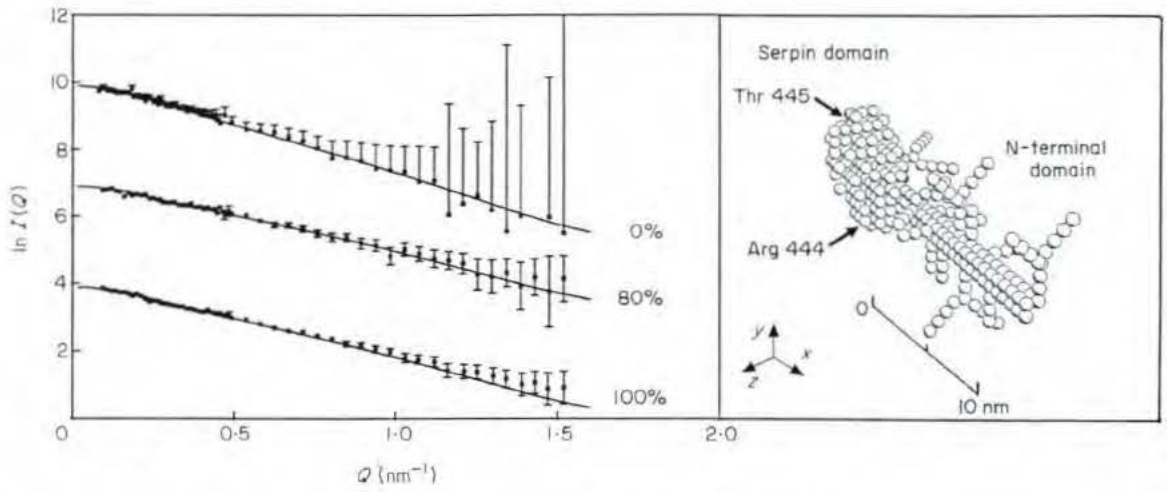


Figure 5. Curve fitting for the Debye sphere models of CI inhibitor. The 395-sphere $+x$ model (shown on the right) leads to simulations that account for the scattering curves of native CI inhibitor in 0%, 80% and 100% ^2H_2O (see the text). Statistical errors in the neutron data are indicated by full (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 $R_{G.C}$ 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.

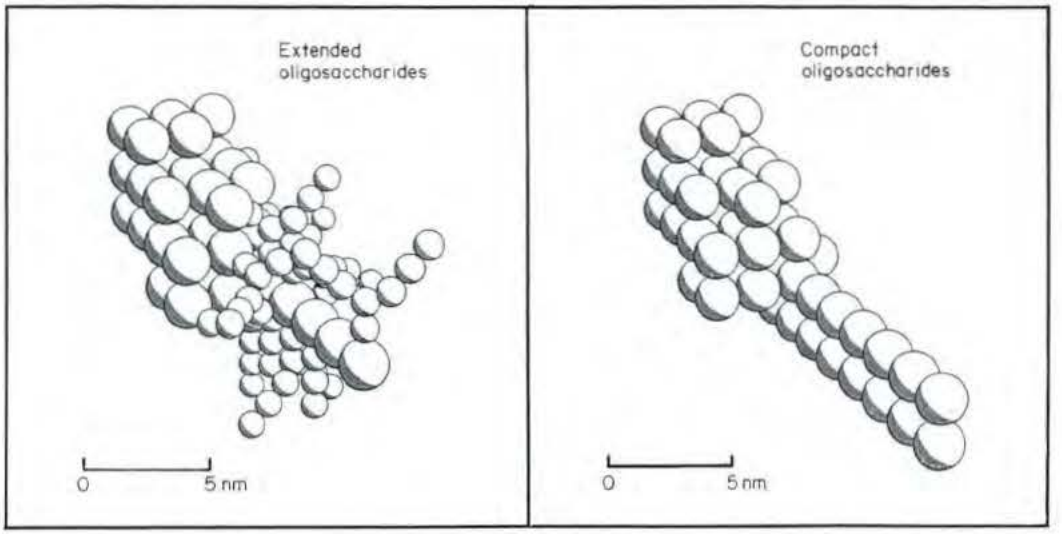


Figure 6. Hydrodynamic models for C1 inhibitor. The oligosaccharides are shown either (a) in an extended or (b) in a compact conformation. Both give $s^0_{20,w}$ values that agree with experiment. Protein spheres: 1·69 nm diameter; carbohydrate spheres: 0·95 nm diameter.

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 ±0·2 S, and the simulations usually agree to within ±0·2 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

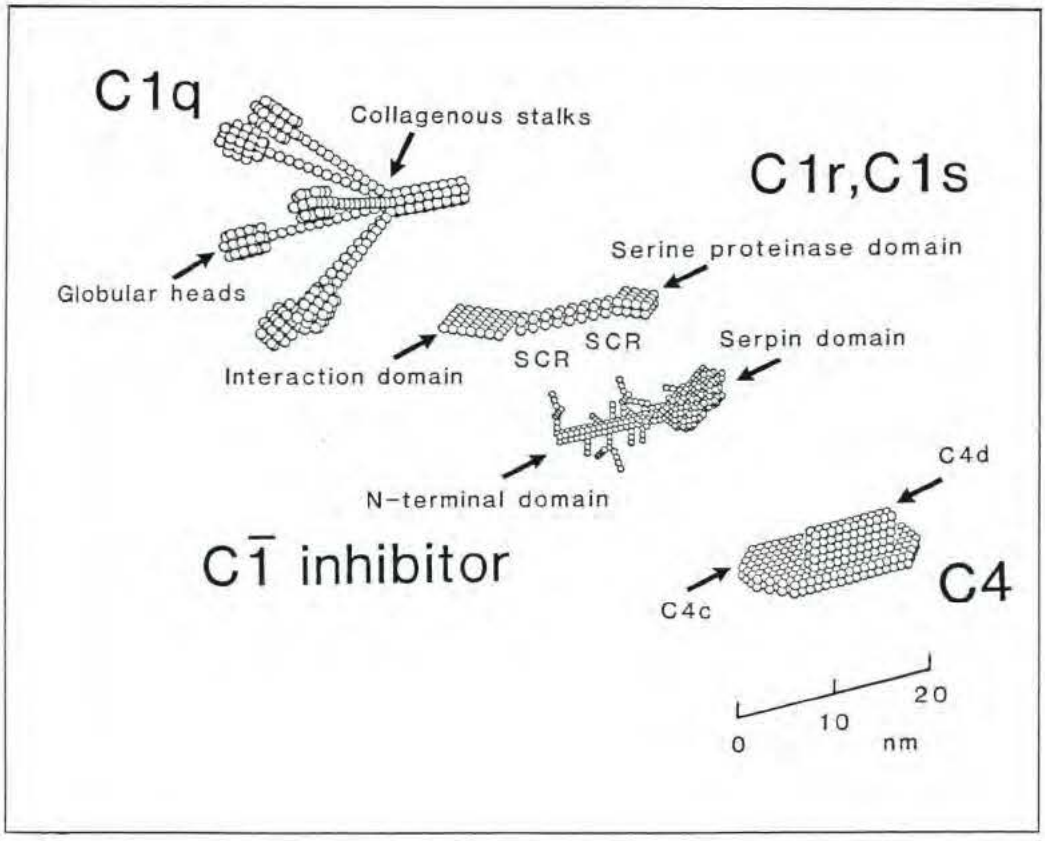


Figure 7. Molecular models of several early components of complement derived from solution scattering. The C1q structure (Perkins, 1985) is hexameric and binds a tetramer of C1r and C1s, of which a monomer is shown (Perkins & Nealis, 1989). The C1̄ inhibitor model from Fig. 5 is also shown. This C1 complex 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); C1̄ 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; C1̄ 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 C1̄ inhibitor is that this blocks the accessibility of the non-proteolytic domains I to V in C1̄r and C1̄s after the serpin domain of C1̄ inhibitor has bound to the proteinase domain VI. Hydrodynamic calculations (S. J. Perkins, unpublished results) show that, in the complexes of C1̄ inhibitor with C1̄r and C1̄s, the N-terminal domain of C1̄ inhibitor is close to the two SCR domains of C1̄r and C1̄s (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 $C1r_2C1s_2$. Once C1̄ inhibitor has reacted with C1̄r and C1̄s, it is reasonable to propose that the N-terminal domain mediates the reduction in the binding affinity of inhibited C1̄r and C1̄s 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 C1̄s after reaction of the latter with C1̄ inhibitor, steric hindrance offered by the N-terminal domain of C1̄ 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 C1̄ 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., Kilehherr, 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