PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Structure and Activity of C1r and C1s

P. E. Carter, B. Dunbar and J. E. Fothergill

Phil. Trans. R. Soc. Lond. B 1984 **306**, 293-299
doi: 10.1098/rstb.1984.0090

Phil. Trans. R. Soc. Lond. B **306**, 293–299 (1984)
Printed in Great Britain

Structure and activity of C1r and C1s

By P. E. Carter, B. Dunbar and J. E. Fothergill

Department of Biochemistry, University of Aberdeen, Marischal College, Aberdeen, AB9 1AS, U.K.

During activation of the first component of the classical complement pathway the two zymogen subcomponents, C1r and C1s are converted to active proteolytic enzymes. Activated C1r cleaves C1s which then becomes the activator of C4 and C2. Amino acid sequence studies of the proteolytic chains of C1r and C1s, carried out in Oxford and Aberdeen respectively, have shown that they belong to the serine proteinase family. Modelling of these sequences to the three-dimensional coordinates of chymotrypsin (Birktoft & Blow 1972) reveals that both molecules have a conserved structural core, and that most of the differences lie in the external loops. Catalytically functional residues (Ile-16, His-57, Asp-102, Ser-195) are conserved, and residue 189 is aspartic acid, consistent with the known trypsin-like specificity of cleavage. Examination of the amino acid sequences of C4a, and comparison with those of the homologous molecules C3a and C5a, shows that there is a marked difference in the distribution of basic residues near the C-terminal arginine residue which is the site of action of $C\overline{1}s$. When these amino acid sequences are modelled to the co-ordinates of C3a (Huber *et al.* 1980) and docked to the active site of $C\overline{1}s$, the basic residues of C4a appear to interact with two glutamate residues peculiar to $C\overline{1}s$, suggesting that this interaction may contribute to the ability of $C\overline{1}s$ to discriminate C4 from C3 and C5.

Introduction

The complement system consists of more than a dozen plasma proteins that can be activated by either the classical or the alternative pathway to produce cell lysis as well as inflammatory and other reactions important in the body's defence mechanisms. Activation of the classical pathway is initiated when antigen–antibody complexes or antibodies bound to cell surface antigens interact with the first component of complement, C1. The first component consists of three subcomponents C1q, C1r and C1s that in plasma exist as a Ca^{2+}-dependent complex (Lepow *et al.* 1963). Binding of C1q to immunoglobulin leads to the conversion from zymogen into active enzyme of the two other subcomponents C1r and C1s. The proteolytic activity of $C\overline{1}s$ then leads to activation of C4 and C2, generating the C3 convertase.

Amino acid sequence studies of C1r (Arlaud *et al.* 1982; Arlaud & Gagnon 1983) and C1s (Carter *et al.* 1983, 1984) have shown that they are enzymes homologous to the classical serine proteinases. It has been possible to model the structures of the serine proteinase chains of $C\overline{1}r$ and $C\overline{1}s$ into chymotrypsin coordinates (Birktoft & Blow 1972), and to examine the structural basis of their activity and specificity.

Molecular properties and biological activities of C1r and C1s

In their zymogen form, both C1r and C1s are single polypeptide chains of $M_r \approx 83000$ (Sim *et al.* 1977) that readily dimerize, particularly in the presence of Ca^{2+} ions. They apparently occur as dimers in the whole C1 complex in which the molar ratios are 1:2:2 for C1q:C1r:C1s. On activation the single zymogen chain is split to give two disulphide-bridged chains ($M_r \approx 55000$ and ≈ 27000). Incorporation of diisopropylphosphofluoridate shows that the serine proteinase active site is on the light chain (Barkas *et al.* 1973; Sim & Porter 1976).

The biological activity of C1r is confined to its activity within the C1 complex, initially as a means of generating the active enzyme $C\overline{1r}$, and then as the activator of C1s to $C\overline{1s}$. The mechanism of C1r activation is poorly understood, but it is thought to be an intramolecular autocatalytic process (Dodds *et al.* 1978; Ziccardi 1982). Once activated, $C\overline{1r}$ then activates C1s by proteolysis. This activity appears to be remarkably specific for zymogen C1s, since $C\overline{1r}$ is unable to cause significant proteolysis of other substrates.

Activated C1s cleaves both C2 and C4, leading to the generation of the C3 convertase. However, $C\overline{1s}$ does not cleave C3 or C5, both homologous in structure to C4, nor Factor B, a structural homologue of C2. The amino acid sequences of the activation fragments of both human (Moon *et al.* 1981) and bovine (Smith *et al.* 1982) C4a indicate that cleavage takes place at an arginine residue. Arginine residues are also found at the carboxyl terminus of C3a and C5a (Hugli & Müller-Eberhard 1978) but $C\overline{1s}$ is not active in cleaving these bonds.

The naturally occurring plasma inhibitor $C\overline{1}$-INH forms a covalent stoichiometric complex with both $C\overline{1r}$ and $C\overline{1s}$ (Harpel & Cooper 1975; Reboul *et al.* 1977). Reduction of the complex shows that the inhibitor is bound to the light chains of $C\overline{1r}$ and $C\overline{1s}$, confirming the location of the enzyme active site.

Sequence alignment and homology of C1r, C1s and chymotrypsin

Conservation of amino acid sequence between C1r, C1s and chymotrypsin is sufficiently extensive to allow the three sequences to be aligned using the usual concept of introducing deletions to maximise homology. This is shown in figure 1 which also indicates the chymotrypsinogen numbering system (Hartley 1964) which will be used throughout. The functionally important residues Ile-16, His-57, Asp-102 and Ser-195 are conserved as well as a large number of other residues. Pairwise comparison of the sequences shows 45% identity between C1r and C1s, 38% identity between C1s and chymotrypsin, and 30% identity between C1r and chymotrypsin. Thus C1s and C1r are more similar to one another than either is to chymotrypsin, although C1s is more like chymotrypsin than is C1r. These degrees of relatedness are similar to the usual range (29–44% identity) found for the mammalian serine proteinases (Young *et al.* 1978).

It is apparent that C1r and C1s share a number of common features. The sequence insertions at residues 94 and 177 are quite extensive, and both show deletions in the 33–41 region. Both C1r and C1s lack the proline residue at position 198 that is found in all the other serine proteinases. Similarly Cys-42 and Cys-58 which form the 'histidine-loop' disulphide in most serine proteinases are missing from C1r and C1s (Arlaud & Gagnon 1981) but the disulphide bridges at 168–182 and 191–220 are conserved. The cysteine residue at position 122, which is connected to the A chain in chymotrypsin, is conserved and probably forms the disulphide bridge to the heavy chain, as its analogue in prothrombin and plasminogen does.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

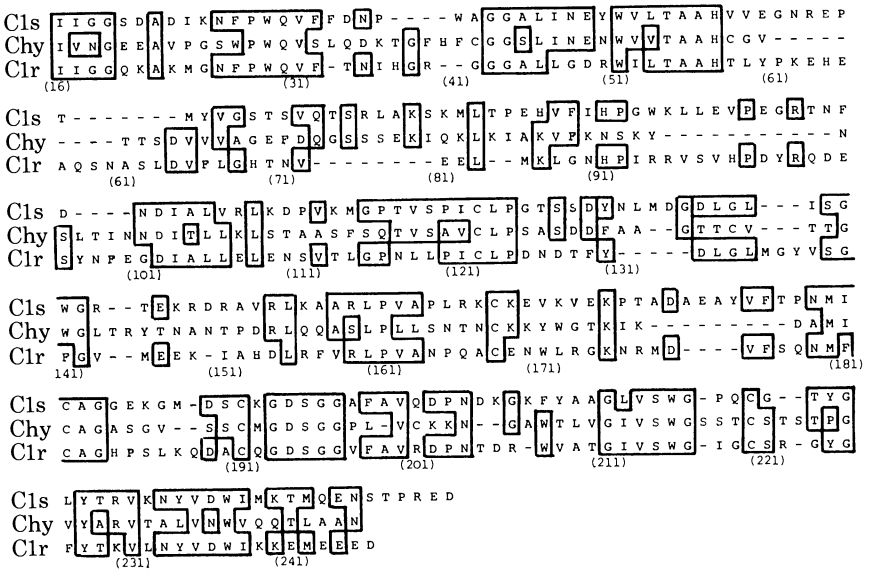

FIGURE 1. The amino acid sequences of chymotrypsin (Hartley 1964), and the light chains of $C\overline{1r}$ (Arlaud & Gagnon 1983) and $C\overline{1s}$ (Carter *et al.* 1984) have been aligned by introducing gaps to maximize the homology among the sequences. Residues that are identical in at least two of the sequences have been enclosed. The chymotrypsinogen numbering system has been used (Hartley 1964).

MODELLING OF C1r AND C1s TO THE CHYMOTRYPSIN COORDINATES

The relatively high degree of homology shown by the amino acid sequences of C1r, C1s and chymotrypsin suggest that it should be possible to build the C1r and C1s structures into the coordinates established by X-ray diffraction for chymotrypsin (Birktoft & Blow 1972). Two main features become apparent. Firstly, the identical residues and the conservative replacements are largely concentrated in the centre of the molecule. Figure 2 shows the conserved residues in C1s and C1r, and it is obvious that they form the structural core of the molecule, providing the framework that allows the catalytic residues to be held in their appropriate positions. Secondly, the replaced residues (figure 2) and the insertions and deletions (figure 3) are mainly associated with external loop regions of the molecule. There are several insertions including particularly extensive ones at residues 94 and 177. These have not been included in the α-carbon skeleton, but from model-building it appears that all the insertions and deletions can be adequately accommodated without disturbing the main structural core. As with the other serine proteinases, the positions of some of these insertions and deletions correlate with intron–exon junctions, but others do not (Craik *et al.* 1983).

A model has been built of the $C\overline{1s}$ light chain structure using a scale of 1 cm = 0.1 nm. The basic assumptions are that identical residues in the amino acid sequence will occupy identical positions in the tertiary structure and that dihedral angles will be maintained as far as possible at replaced residues. The model building proved quite satisfactory in most positions, although the proline replacements were somewhat more difficult. The overall impression is that the shapes of C1r and C1s are quite similar to chymotrypsin but that they have two large extra external loops close together on one side. The disulphide link to the heavy chain at residue 122 is in a position that would be quite consistent with such a function. The physical model is particularly helpful when considering the interaction with the substrates.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

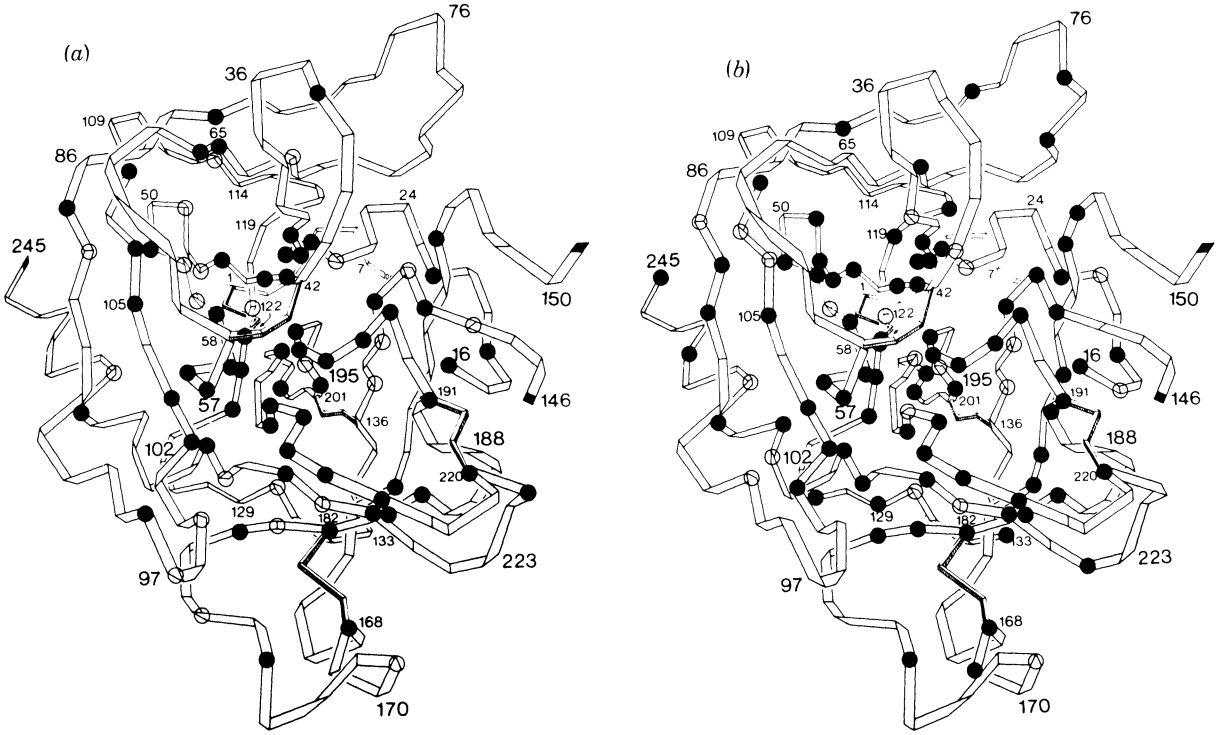

FIGURE 2. The α-carbon ribbon backbone of chymotrypsin (Birktoft & Blow 1972) modified according to the alignment of figure 1 to show identical residues (●) and conservative replacements (○). (*a*) C1r, (*b*) C1s.

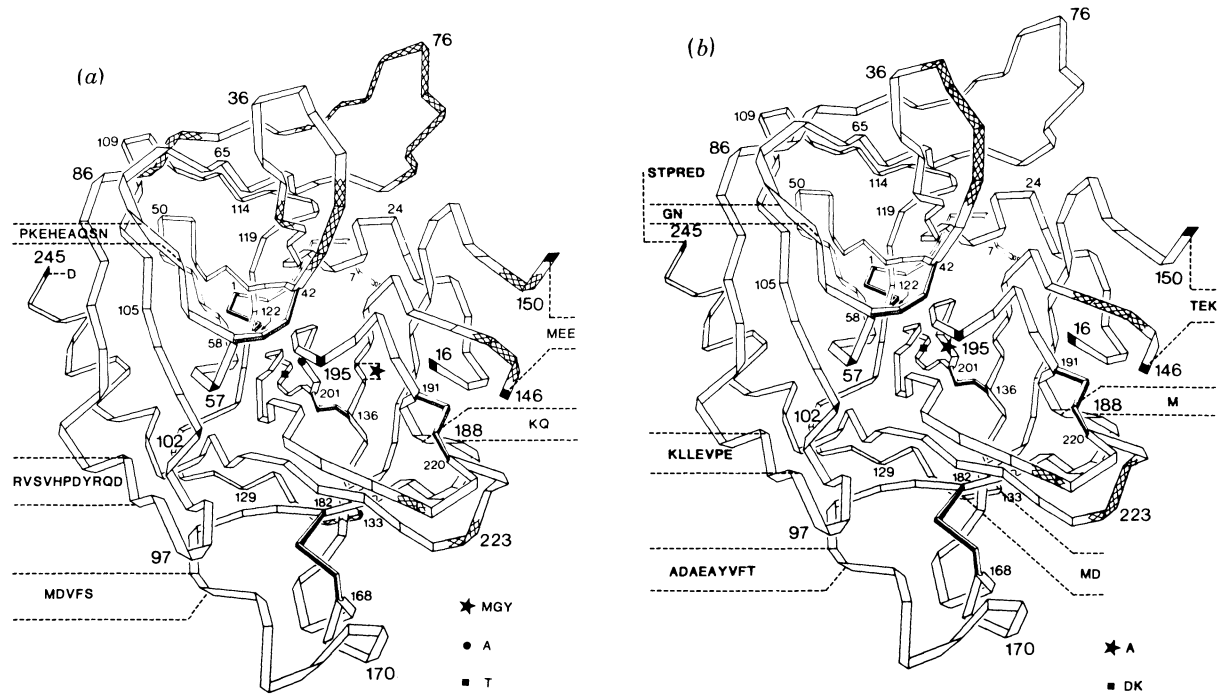

FIGURE 3. The α-carbon ribbon backbone of chymotrypsin (Birktoft & Blow 1972) modified according to the alignment of figure 1 to show deletions by cross-hatching and insertions by dotted line or symbol. The amino acid sequences to be inserted are shown in single letter code reading in the N- to C-terminal direction. (*a*) C1r, (*b*) C1s.

Amino acid sequences and modelling of substrates

The three complement components C3, C4 and C5 are known to be homologous in several respects. From the point of view of substrate specificity for $\overline{C1s}$ the important regions are those containing the site of cleavage around the arginine residue near position 77 in the α-chain (Human $\overline{C1s}$ cleaves both human and bovine C4 (Booth *et al.* 1979)). A comparison (Smith *et al.* 1982) of the amino acid sequences of human and porcine C3a (Hugli & Müller-Eberhard 1978), human (Moon *et al.* 1981) and bovine (Smith *et al.* 1982) C4a, and human (Hugli & Müller-Eberhard 1978) and porcine (Gerard & Hugli 1980) C5a shows that all of these molecules contain the unusual 'disulphide knot' arrangement of six cysteine residues as well as a number of other conserved residues. When the sequences are aligned from the carboxy-terminal arginine residue (figure 4) it is apparent that there is a significant difference in the

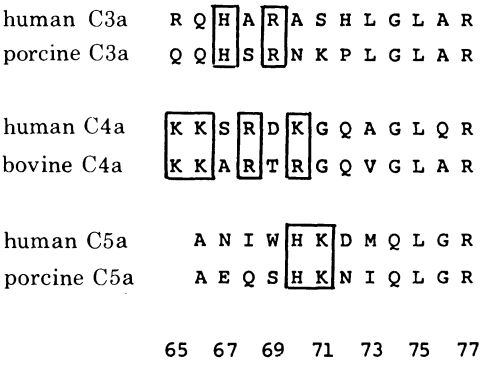

Figure 4. An alignment of the C-terminal sequences of human and porcine C3a (Hugli & Müller-Eberhard 1978), human (Moon *et al.* 1981) and bovine (Smith *et al.* 1982) C4a and human (Hugli & Müller-Eberhard 1978) and porcine (Gerard & Hugli 1980) C5a. Basic residues that are conserved in both sequences have been enclosed. The preceding residue in all sequences is the totally conserved arginine residue (for general alignment see Smith *et al.* 1982). The numbering system is based on C3a.

distribution of basic side chains. The sequences of C3a, C4a and C5a are sufficiently similar to allow them all to be built into the X-ray diffraction structure for C3a (Huber *et al.* 1980). The relevant features of this structure are the central core of the three disulphide bridges making the 'disulphide knot', and the α-helix extending through this central knot towards the carboxy-terminal arginine residue. The regularity of the α-helix is somewhat distorted within the last few residues of the structure. This then provides a structural model for 'docking' against the active site cleft of $\overline{C1s}$ to examine more closely the enzymic specificity.

Structural basis of the enzymic specificity of $\overline{C1s}$

The most obvious relevant point about the structures of $\overline{C1r}$ and $\overline{C1s}$ is that they both have an aspartate residue at position 189 which is characteristic of trypsin and the other serine proteinases that are specific for basic side chains. This fits well with the observation that the carboxy-terminal residue of the heavy chain of $\overline{C1s}$ is thought to be arginine, which on the zymogen occurs immediately before the isoleucine-16, the amino terminal activating residue of the light chain. This specificity can clearly give rise to active $\overline{C1s}$ in the classical manner

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

described for chymotrypsin. However, the very restricted activity of $C\overline{1}r$ in cleaving only zymogen C1s at this position cannot yet be explained.

The occurrence of arginine residues at the carboxy-termini of the two C4a sequences is also consistent with an aspartate residue at position 189. However, other activation fragments, C3a and C5a also have an arginine residue at the corresponding position, yet are not generated by the action of C1s.

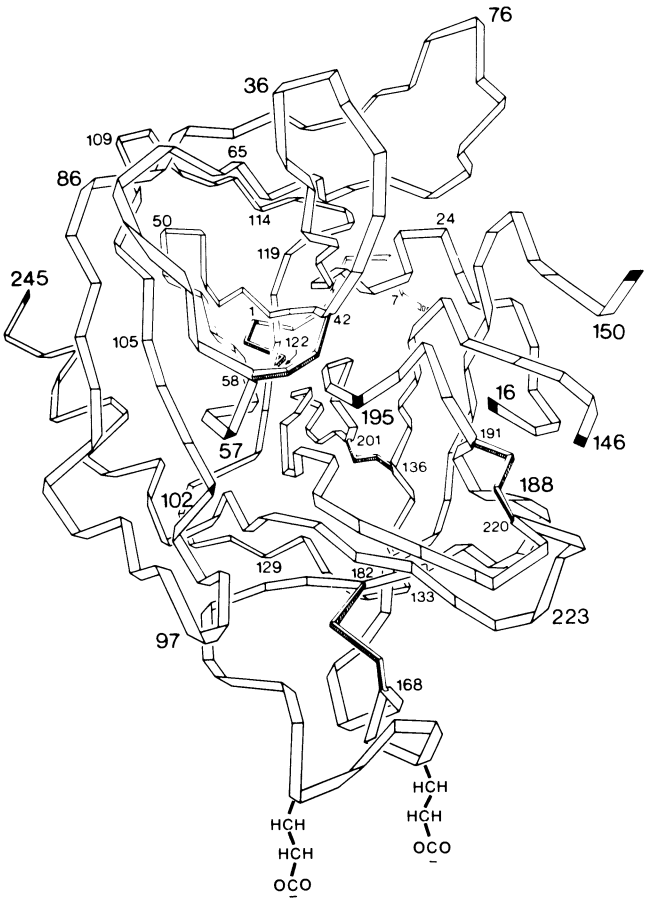

FIGURE 5. The α-carbon ribbon backbone of chymotrypsin (Birktoft & Blow 1972) modified to show the approximate positions of Glu-170 and Glu-174 of the $C\overline{1}s$ light chain, which are thought to interact with the basic side chains of the C4a C-terminal sequences (see figure 4).

We can now examine the interaction of the C4a, C3a and C5a molecules, built into the C3a X-ray structure, with the model of $C\overline{1}s$. (We are not yet able to examine the sequence on the carboxyl side of this arginine residue, but the gene sequences should soon make this possible.) We can then 'dock' the C4a, C3a and C5a structures against the active-site cleft of $C\overline{1}s$, by setting the carboxy-terminal arginine residue so that the guanidinium group interacts at Asp-189 and the carboxyl group at Ser-195. If we now inspect the interaction of the side chains of this near helical structure with the $C\overline{1}s$ molecule we see that there are two glutamate residues at positions 170 and 174 of $C\overline{1}s$ (figure 5) that could interact by salt bridge formation with the basic residues of C4a. If water is adequately excluded, these salt bridges could contribute to the specificity of interaction of $C\overline{1}s$ with C4. Similar comparisons with C3a and C5a do not

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

show any significant interactions of this type. None of the sequences of the other serine proteinases contain two carboxylate residues in the positions corresponding to Glu-170 and Glu-174 of $C\bar{1}s$.

We should like to thank Dr Linda Fothergill, Dr Jean Gagnon, Dr Herman Watson, Professor Tom Blundell and Professor David Blow, for helpful discussions, and the Medical Research Council and the Science and Engineering Research Council for financial support.

References

Arlaud, G. J. & Gagnon, J. 1981 $C\bar{1}r$ and $C\bar{1}s$ subcomponents of human complement: two serine proteinases lacking the 'histidine-loop' disulphide bridge. *Biosci. Rep.* **1**, 779–784.

Arlaud, G. J. & Gagnon, J. 1983 Complete amino acid sequence of the catalytic chain of human complement subcomponent $C\bar{1}r$. *Biochemistry, Wash.* **22**, 1758–1764.

Arlaud, G. J., Gagnon, J. & Porter, R. R. 1982 The catalytic chain of human complement component $C\bar{1}r$. Purification and N-terminal amino acid sequences of the major cyanogen bromide-cleavage fragments. *Biochem. J.* **201**, 49–59.

Barkas, T., Scott, G. K. & Fothergill, J. E. 1973 Purification, characterization and active-site studies on human serum complement subcomponent C1s. *Biochem. Soc. Trans.* **1**, 1219–1220.

Birktoft, J. J. & Blow, D. M. 1972 Structure of crystalline α-chymotrypsin. V. The atomic structure of tosyl-α-chymotrypsin at 2 Å resolution. *J. mol. Biol.* **68**, 187–240.

Booth, N. A., Campbell, R. D. & Fothergill, J. E. 1979 The purification and characterization of bovine C4, the fourth component of complement. *Biochem. J.* **177**, 959–965.

Carter, P. E., Dunbar, B. & Fothergill, J. E. 1983 The serine proteinase chain of human complement component $C\bar{1}s$. Cyanogen bromide cleavage and N-terminal sequences of the fragments. *Biochem. J.* **215**, 565–571.

Carter, P. E., Dunbar, B. & Fothergill, J. E. 1984 The complete amino acid sequence of the serine proteinase chain of human complement component $C\bar{1}s$. (In preparation.)

Craik, C. S., Rutter, W. J. & Fletterick, R. 1983 Splice junctions: association with variation in protein structure. *Science, Wash.* **220**, 1125–1129.

Dodds, A. W., Sim, R. B., Porter, R. R. & Kerr, M. A. 1978 Activation of the first component of human complement (C1) by antigen–antibody aggregates. *Biochem. J.* **175**, 383–390.

Gerard, G. & Hugli, T. E. 1980 Amino acid sequence of the anaphylatoxin from the fifth component of porcine complement. *J. biol. Chem.* **255**, 4710–4715.

Harpel, P. C. & Cooper, N. R. 1975 Studies on human plasma $C\bar{1}$ inactivator-enzyme interactions. *J. clin. Invest.* **55**, 593–604.

Hartley, B. S. 1964 Amino-acid sequence of bovine chymotrypsinogen-A. *Nature, Lond.* **201**, 1284–1287.

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

Hugli, T. E. & Müller-Eberhard, H. J. 1978 Anaphylatoxins: C3a and C5a. *Adv. Immunol.* **26**, 1–53.

Lepow, I. H., Naff, G. B., Todd, E. W., Pensky, J. & Hinz, C. F. 1963 Chromatographic resolution of the first component of human complement into three activities. *J. exp. Med.* **117**, 983–1008.

Moon, K. E., Gorski, J. P. & Hugli, T. E. 1981 Complete primary structure of human C4a anaphylatoxin. *J. biol. Chem.* **256**, 8685–8692.

Reboul, A., Arlaud, G. J., Sim, R. B. & Colomb, M. G. 1977 A simplified procedure for the purification of $C\bar{1}$-inactivator from human plasma. Interaction with complement subcomponents $C\bar{1}r$ and $C\bar{1}s$. *FEBS Lett.* **79**, 45–50.

Sim, R. B. & Porter, R. R. 1976 Isolation and comparison of the proenzyme and activated forms of the human serum complement subcomponents C1r and C1s. *Biochem. Soc. Trans.* **4**, 127–129.

Sim, R. B., Porter, R. R., Reid, K. B. M. & Gigli, I. 1977 The structure and enzymic activities of the C1r and C1s subcomponents of C1, the first component of human serum complement. *Biochem. J.* **163**, 219–227.

Smith, M. A., Gerrie, L. M., Dunbar, B. & Fothergill, J. E. 1982 Primary structure of bovine complement activation fragment C4a, the third anaphylatoxin. Purification and complete amino acid sequence. *Biochem. J.* **207**, 253–260.

Young, C. L., Barker, W. C., Tomaselli, C. M. & Dayhoff, M. O. 1978 Serine proteases. In *Atlas of protein sequence and structure* (ed. M. O. Dayhoff), vol. 5, suppl. 3, pp. 73–93. Washington: National Biomedical Research Foundation.

Ziccardi, R. J. 1982 Spontaneous activation of the first component of human complement (C1) by an intramolecular autocatalytic mechanism. *J. Immunol.* **128**, 2500–2508.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES