The EMBO Journal vol.6 no.7 pp.1951–1957, 1987

Topological mapping of complement component C9 by recombinant DNA techniques suggests a novel mechanism for its insertion into target membranes

Keith K.Stanley and Joachim Herz
with technical assistance by John Dickson

European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, D-6900 Heidelberg, FRG

Communicated by K.Simons

cDNA molecules coding for mouse and trout C9 have been isolated and the derived amino acid sequences compared with that of human C9. Regions of high homology between the closely related species (mouse and human) correlate with putative domains in the protein structure supporting a model of C9 having five globular domains. Comparison between the more distant species (trout and human) suggests regions of particular importance to C9 structure and function. In addition the three related sequences allow the secondary structure to be predicted with more confidence and we have tested the prediction by mapping surface features of the protein. Reported here is a recombinant DNA approach to fine mapping of antibody epitopes. Two of the putative domains of C9 are connected by a stretch of about 40 amino acid residues in which features characteristic of individual conformational forms of C9 are concentrated. We suggest that this region might act as a hinge allowing the rearrangement of globular domains necessary for membrane insertion. In the membrane inserting domain one highly conserved sequence has the potential to form an amphipathic α-helix once it is buried in the lipid bilayer. These features suggest a novel mechanism for the irreversible, post-translational insertion of C9 into target membranes.

Key words: complement/epitope mapping/membrane assembly/species homology/protein topology

Introduction

Complement component C9 is a globular plasma protein which has the ability to insert into membranes containing a preformed C5b-8 complex (Hadding and Muller-Eberhard, 1969; Tschopp *et al.*, 1982a). As, or soon after, C9 interacts with the C5b-8 complex it undergoes a major conformational change (Silverman and Nelsestuen, 1986) in which the molecule approximately doubles in length (Podack and Tschopp, 1982) and hydrophobic structures become exposed which can be labelled with membrane restricted photo-affinity probes (Ishida *et al.*, 1982; Amiguet *et al.*, 1985). This conformational change is also associated with the appearance of previously hidden antibody epitopes (Podack *et al.*, 1982). Only small differences in the total secondary structure can be detected with circular dichroism, however (Tschoop *et al.*, 1982b) suggesting that a redistribution of globular domains rather than a wholesale refolding of the protein is involved. We have used a variety of techniques to map residues which are likely to be located at the surface of the C9 protein. These include the sites of N-linked oligosaccharide attachment, sites of proteolytic cleavage, and the position of amino acids whose codon is interrupted by introns in the gene (Stanley *et al.*, 1985; Marazziti,D., Eggertsen,G., Stanley,K.K. and Fey,G., in preparation). Reported here are studies on the conservation of C9 sequences between different species and the mapping of antibody epitopes. Mouse C9 was chosen so that comparisons would eventually be possible with the structurally and functionally related protein, perforin (Tschopp *et al.*, 1986; Lowrey *et al.*, 1987), present in cytoplasmic granules of mouse cytotoxic T-lymphocytes. Trout was taken as an example of a distantly related C9 molecule as normal cytolytic activity has been described in trout (Nonaka *et al.*, 1981) allowing conclusions to be made about the functional significance of conserved sequences. Taken together with published biochemical experiments the inter-species sequence homologies and topological mapping suggests a 5 domain model for C9 structure, and a possible mechanism for the unfolding of the molecule and its interaction with the lipid bilayer of a target cell.

Results

Inter-species comparison of C9 amino acid sequences

cDNA libraries were constructed from mouse and trout liver poly(A)$^+$ RNA using an adaptor cloning strategy (Haymerle *et al.*, 1986) and screened with radioactive cDNA probes (human C9 for isolating the mouse C9 cDNA and mouse C9 for isolating the trout clone). Clones were selected on the basis of size and the largest clone was sequenced. The DNA sequences which are not presented here have been submitted to the EMBL nucleotide sequence data library.

Figure 1 shows a comparison of the amino acid sequences of mouse, trout and human C9 using the matrix comparison method of Argos (1987). In both cases all possible spans of 25 amino acid residues were compared on the basis of the Dayhoff relatedness odds matrix (Barker *et al.*, 1978) and five residue physical parameters (Argos, 1987). The score relating each pair of 25 residue spans was then re-expressed as a function of the distribution of scores in the whole comparison matrix. The points in the upper panels of Figure 1 represent pairs of sequence spans in the comparison matrix with scores at least four standard deviations above the mean. Using this procedure the broad areas of greatest homology are emphasised as a series of diagonal lines with gaps defining the regions of lower homology. It is striking that the large gap at residue 228 in the mouse-human comparison is also found in the trout-human comparison, and that the gaps at 360 and 436 are also common, although of different widths.

Each diagonal line contains further information about the homology of the two sequences since the quality of fit at each point is given by its score in the comparison matrix. These scores are plotted in the lower panels of Figure 1 as a function of the sequence position in human C9. It can be seen from this comparison that human C9 is more closely related to mouse C9 than to trout C9, having wide plateaus of high homology separated by narrow valleys which often correspond to insertions and deletions in the alignment. Trout C9 on the other hand has only narrow peaks of high homology which always fall within the plateaus of high homology seen in the mouse-human comparison. These

© IRL Press Limited, Oxford, England

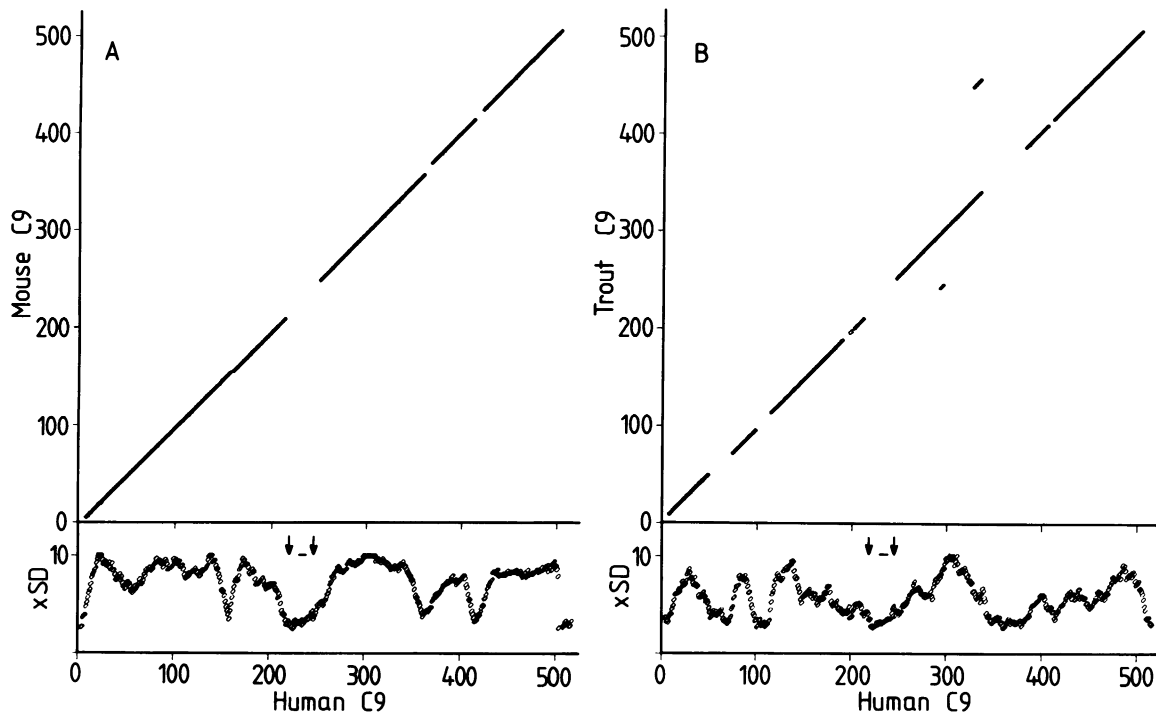

Fig. 1. Comparison of the amino acid sequences of mouse, trout and human C9. Mouse (**A**) and trout (**B**) C9 amino acid sequences were compared with the sequence of human C9 over all possible spans of 25 residues. Points in the comparison matrix 4.0 or more standard deviations above the mean of the matrix are shown in the upper plot. The highest standard deviation value in the comparison matrix for each sequence position of human C9 (corresponding in every case to the diagonal line) are shown below. Arrows indicate the cleavage sites of chymotrypsin and α-thrombin, the bar indicates the position of the neoantigen epitope.

sharp peaks presumably represent regions where particular amino acids are of special importance to C9 structure or function.

Sequence and biochemical considerations suggest that C9 folds into several globular domains, which would also be consistent with its image in negative stained electron micrographs (Podack and Tschopp, 1982). The first domain, containing six cysteine residues in a stretch of 66 amino acid residues in human C9, is found duplicated in the β-chain of complement component C8 (Haefliger *et al.*, 1987) suggesting that it folds as an independent entity. The second domain (residues 78–115) corresponds to the class A cysteine rich homology (Doolittle, 1985; Stanley *et al.*, 1986) which has been found in the low density lipoprotein (LDL) receptor and several other complement components (Haefliger *et al.*, 1987; Rao *et al.*, 1987; Catterall *et al.*, 1987). By analogy with the cysteine rich sequences of wheat germ agglutinin (Stanley *et al.*, 1985; Drenth *et al.*, 1980) it is very likely that this region also folds as a small globular domain. In the central portion of the molecule, between the α-thrombin and trypsin cleavage sites (residues 245 and 392) lies the principal region defined biochemically as interacting with the lipid bilayer when C9 unfolds and inserts into a target membrane (Ishida *et al.*, 1982; Amiguet *et al.*, 1985). It is likely that this region also forms a discrete domain of the protein (Hammer *et al.*, 1977; Podack and Tschopp, 1982). There is less data about the remaining two areas of the protein, but by default they must constitute at least one domain each, giving a protein of five domains. In the trout-human comparison (Figure 1B) these five domains correspond to the five major diagonal lines and five highest peaks of homology.

The five peaks of high homology in the trout-human comparison (Figure 1B) are of special interest since they probably define areas of pivotal significance to C9 structure and function.

Fig. 2. A possible membrane interacting region of C9. (**A**) Conserved amino acid sequence (residues 310–326 in human C9), (**B**) projection of a hypothetical α-helix showing amphipathic structure (shaded area is hydrophobic).

One of these falls within the area which interacts with lipids in the target cell membrane and might therefore be the principal element involved. The sequence of this region is shown in Figure 2. This sequence can be folded into an amphipathic α-helix provided that glycine residues can be tolerated in the helix. Although this is very unusual for α-helices in globular proteins (it may be seen in Figure 3 that this region does not predict as α-helix using parameters for soluble proteins), it has been shown that for membrane spanning helices glycine is tolerated, and even slightly favored (Rao and Argos, 1986). Previous studies on possible membrane interacting structures (Stanley *et al.*, 1986) did not take account of this. The other possible membrane interacting structures predicted in that study have conserved

```
Mouse          QMPIPVSREEQEQH.YPIPIDCRMSPWSNWSECDPCLKQRFRSRSILAFGQFNGKSCVDVLGDRQG.CETTQECEEIQ.ENCGN.DFQCETGRCIKRRLL
Human      1   QYTTSYDPELTESSGSASHIDCRMSPWSEWSQCDPCLRQMFRSRSIEVFGQFNGKRCTDAVGDRRQ.CVPTEPCEDAE.DDCGN.DFQCSTGRCIKMRLR
Trout          QANGTLGRSRWL...PLDPVDCVWSRWSEWTPCNSCTKIRHRSRSVEVFGQFVGKPCQGQPIGEQQACTSDAVCEQALPSECSSIEFTCESGACIKLRLS
                tttttttaaaaattttbbbbbbbtttttttttt     bbbbbbbbbttttttttbbbbttttt        aaaaaaa  tttt     ttt    aaaaa
               <---E1---->            <-E2->                                                <--E3-->

Mouse          CNGDNDCGDYSDENDCDDDPRTPCRDRVAEESELGLRAGYGINILGME....PFDNEFYNGLCDRVRD..EKTYYRKPWNVVSLIYETKADKSFRTENYD
Human     99   CNGDNDCGDFSDEDDCESEPRPPCRDRVVEESELARTAGYGINILGMDPLSTPFDNEFYNGLCNRDRDGNTLTYYRRPWNVASLIYETKGEKNFRTEHYE
Trout          CNGDYDCEDGSDE.DCEP.VRKPCGTKLYDTNEQGRTAGYGINILGMEPRINPFNNDYFNGMCNKVKNINNNEYNRLPWNVGLLNYETIAEETVSKEIYE
               ttttttttttttt   ttttttttt aaaaaaa  tttttbbbb   tttttttttttttt    ttttttttttttttbbbbbbbaaaaaa  aaaaa

Mouse          EHLEVFKAINREKTSNFNADFALKFSPTEVPQKGAGEVSPAEHSSKATNISAKFTISYFMGKNFRRLSSYLSQSKKMFVHLRGVVQLGRFVMRNRDVVLR
Human    199   EQIEAFKSIIQEKTSNFNAAISLKFTPTETNKAEQCCEET..ASSISLHGKGSFRFSYSKNETYQLFLSYSSKKEKMFLHVKGEIHLGRFVMRNRDVVLT
Trout          DTYTLLRELMTETKLTVSAGLNLKFTPTEKSMA.........KSNTTVSGGVGLDAEYDRTQMIKEVSEYTTIKNKSFMRVNGRVQLSTYRMRSRDLQVA
               aaaaaaaaaaaaaattttaaaaaattttaaaaaaaaaaaaaaattt  tttttbbbbttttt     bbbbb aaaaaaaa  bbbbbbbbb   bbbbb
                                    <----E4---->

Mouse          STFLDDVKALPTSYEKGEYFGFLETYGTHYSTSGSLGGQYEIVYVLDKASMKEKGVDLNDVKHCLGFNMDLRIPLQDDLKDASVTASVNADGCIKTDNGK
Human    297   TTFVDDIKALPTTYEKGEYFAFLETYGTHYSSSGSLGGLYELIYVLDKASMKRKGVELKDIKRCLGYHLD....VSLAFSEISVGAEFNKDDCVKRGEGR
Trout          GEFLEHVKSLPLEYEKGQYFSFLEDYGTHYTRNGKSGGEHQLVYVLNQDTIKDKKLTERKLQDCIKVGIS...ANFDTNIGIGGDAHIRPGHCKDTVNKN
               bbaaaaaaaa ttttttt     tttt  ttttttttt  bbbbbaaaaaaaaaaaaaaaaaaaabbbb       aaaaaabbbb     tttt   ttttt

Mouse          TVNITRDNIIDDVISFIRGGTREQAILLKEKILRGDKTFDKTDFANWASSLANAPALISQRMSPIYNLIPLKIKDAYIKKQNLEKAVEDYIDEFSTKRCY
Human    393   AVNITSENLIDDVVSLIRGGTRKYAFELKEKLLRG.TVIDVTDFVNWASSINDAPVLISQKLSPIYNLVPVKMKNAHLKKQNLERAIEDYINEFSVRKCH
Trout          TAEKEGKALVDKVITVVRGGTLEAAVAMRTQITKE.GLMDVKTYQNWARTVGDAPALLSSEPEPIQTLIPLSMPDANTRRLNMQRATQEYEAEYSVCKCK
               t          bbbbbbbbbttttaaaaaaaaaaaaattbbbbbbbtttt   ttt       tttbbbbbbaaaaaaaaaaaaaaaaaaaaaaaaabbbbbb
                                 <------E5------>                                                               <--

Mouse          PCLNGGTIILLDGQCLCSCPMMFRGMACEIHQQI
Human    492   TCQNGGTVILMDGKCLCACPFKFEGIACEISKQKISEGLPALEFPNEK
Trout          PCHNGGSLALLDGKCLCLCLPQFEGLACQDAKADNNKNTKTPVRVFLEKGNWSVGQPGQAAVEENV
                 tttt              aaaaa aaaaaaaaa
               -----E6------><------E7------>
```

Fig. 3. Alignment of the C9 sequences from mouse, trout and human C9. The cDNA derived amino acid sequences of three species of C9 and predicted secondary structure. Solid lines indicate identities, broken lines indicate conserved amino acids according to the groups: (RK), (DENQ), (ST), (GP), (MCHAVLIFYW), and symbols representing probable structures are: **a** = α-helix, **b** = β-pleated sheet, **t** = turn. The position of epitopes E1–E7 are indicated below the sequence.

hydrophobic residues in mouse and trout C9, but do not have the high level of overall homology shown in Figure 2. It is conceivable therefore that this metastable helix serves some special function, for instance to lock C9 into the target membrane by forming a hydrophobic surface after becoming embedded in a hydrophobic environment. The ability of this region of C9 to pass through the endoplasmic reticulum membrane during its secretion on the other hand could be ascribed to the charges present on the other face which make it quite different from a 'Stop transfer' signal. Only by polymerisation to form a membrane pore could a stable interaction with the lipid bilayer result.

An alignment of the three sequences is shown in Figure 3. Although the amino-terminus of mouse and trout C9 is not known a glutamine residue situated adjacent to a hydrophobic leader peptide was chosen in each case as a likely start of the mature protein. In this alignment 22 out of 24 cysteine residues in human C9 are found in the same position as those in mouse and trout C9. The two cysteines found in the hinge region in human C9 (residues 234 and 235) are not found in mouse or trout but are unlikely to be disulphide linked since they are adjacent. Trout C9 has one extra cysteine at a position equivalent to residue 482 in human C9. This lies just before the region with homology to urokinase (Doolittle, 1985; Stanley *et al.*, 1986) and is

A Globular C9

B Extended C9

1 2 5 3 4

chymotrypsin α–thrombin

Epitope 4

1 2 3 5 4

Fig. 4. A domain model for complement component C9. The 5 globular domains suggested by comparison of the species sequences of C9 have been arranged to explain the transition from globular to extended forms. The 'hinge' region is shown as a zigzag line in globular C9 which unfolds in extended C9 giving access to the neoantigen epitope 4 (monoclonal antibody bC5), and occluding the cleavage sites of chymotrypsin and α-thrombin.

```
Apo-B          T T R L T R K R G L K
(3357-3367)        +     + + +     +

Apo-E          H L R K L R K R L L R
(140-150)          + +   + + +     +

Human C9       Y A F E L K E K L L R
(415-425)                +   +     +
```

Fig. 5. Apolipoprotein homology of C9. The receptor binding region of apo-E and apo-B is compared with a sequence in human C9 that shows a high homology on the basis of residue physical parameters (Argos, 1987).

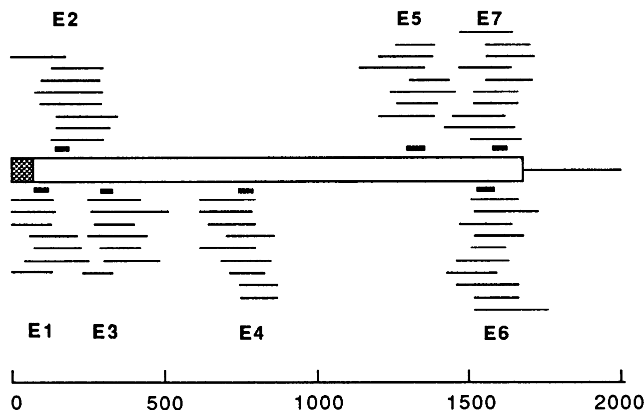

Fig. 6. Linear map of C9 epitope clones. Plasmid DNA was prepared from clones positive with each monoclonal or affinity purified antibody and the ends sequenced by double stranded DNA sequencing. Fragments were also excised wtih *Nco*I and end-labelled to check that only one fragment was present in the clone. The position of each clone is shown on a linear map of the C9 molecule. Heavy bars indicate the minimum areas of overlap of each group of clones. The shaded area shows the leader peptide, and the open box the open reading frame of human C9. The scale is calibrated in base pairs.

presumably not disulphide bonded. Interestingly, an extra cysteine also occurs in this position in C8β (Haefliger *et al.*, 1987). The most variable region of the sequences was at the carboxy terminal end of the molecule, which in human C9 corresponds to a separate exon of the C9 gene (D.Marazziti,, G.Egertsen, K.K.Stanley and G.Fey, in preparation). Only human C9 therefore contains the pentapeptide homology with melittin (Esser, 1987) making it unlikely that this sequence plays any direct role in C9 function.

In the centre of the C9 sequence (residues 228–271 in human C9) is a region showing low homology between all three species. The position of this low homology region is identical in each species. Although lacking a conserved sequence, suggesting that no binding function of C9 can reside in this portion, the region does contain a high concentration of features present in only one of the two conformational forms of C9 suggesting a structural role in the unfolding of C9. Thus two proteolytic cleavage sites only accessible in the globular form of human C9 (chymotrypsin and α-thrombin) and an antibody epitope found only in its unfolded state are located here (Figure 1). As detailed below, a hinge region at this point of the molecule would also be necessary to move the biochemically defined membrane inserting domain to the physical end of the molecule where polymerised C9 interacts with a target membrane (Podack and Tschopp, 1982).

A domain model for C9 unfolding

Based on the above analysis it is proposed that complement component C9 consists of five globular domains. In order to pack into an elipsoidal globular shape, as is observed for monomeric C9 (Podack and Tschopp, 1982), these globular domains are presumably arranged in a cluster similar to that shown in Figure 4a with the hydrophobic faces of membrane inserting structures either hidden from the aqueous environment or destroyed by a different folding of the membrane inserting region (see above). In this conformation the two proteolytic cleavage sites in the 'hinge' region are exposed, but the neoantigen epitope is occluded (Figure 4a). Since in the membrane attack complex the membrane interacting segment of C9 is present on the tip of the cylindrical pore complex, we propose that rearrangement of the globular domains allows the membrane interacting domain to migrate to the end of the molecule even though it is not situated at either physical end of the polypeptide chain. For this reason we postulate in Figure 4b that the 'hinge' region unfolds during the monomer to unfolded C9 transition to expose the neoantigen epitope and occlude the proteolytic cleavage sites. The hinge region contains one predicted α-helix (Figure 3), which if it were unfolded could account for 2% of the α-helix content, consistent with the change in the circular dichroic spectra of the molecule in its two forms (Tschopp *et al.*, 1982b).

An important corollary of the conformational transition in C9 is that specific forces must hold the monomeric C9 in its globular form and a specific mechanism must exist for unfolding the molecule. Since C9 contains a globular domain with a high level of homology to the LDL receptor including the cluster of negative charges responsible for apolipoprotein binding (Yamamoto *et al.*, 1984), an attractive hypothesis would be that an internal ligand-receptor interaction occurs within the C9 molecule. For the LDL receptor the interaction with apolipoproteins E and B (apo-E and apo-B) is mainly electrostatic, involving one or two clusters of basic residues arranged on a predicted α-helical region of the protein (Innerarity *et al.*, 1984; Knott *et al.*, 1986). No overall consensus is found in these sequences although one of the regions in each apolipoprotein shows good homology (Figure 5). When the human C9 sequence was compared with that of apo-E using the sensitive search procedure of Argos (1987) the best homology between the two proteins, having a score five standard deviations above the mean of the comparison matrix, fell in the sequence of apo-E which interacts with the LDL receptor. This region (residues 415–425 in human C9, Figure 5) falls in the fifth putative domain of C9. Two of the positive charges are preserved in mouse and trout C9. Some other sequences with conserved positive charges in a predicted α-helix (e.g. residues 469–489) are present close by. These sequences are therefore of interest for investigation of C9 structure and function by site directed mutagenesis.

Secondary structure prediction of C9

The secondary structure of C9 was predicted using the Chou and Fasman rules (1974a,b) for both the individual sequences and an average of all three sequences. For the human C9 sequence alone the predicted content of α-helix was 24% agreeing with the measured value for monomeric C9 (Tschopp *et al.*, 1982) while the β-pleated sheet was underestimated at 20% (compared with 32% measured). When the average prediction for all three sequences was computed the α-helix increased to 30% and the β-pleated sheet decreased to 18%. The positions of elements of secondary structure were in general not changed but became more

Table I. Sequences of polyclonal anti-human C9 epitopes

Epitope	Position	Sequence	Species/	Charges
1	1–11	QYTTSYDPELT	Human	2
2	29–34	EWSQCD	Human	2
		NWSECD	Mouse	2
3	79–85	DCGNDFQ	Human	2
4	229–238	KAEQCCEETA	Human	4
5	412–426	RKYAFELKEKLLRGT	Human	6
		REQAILLKEKILRGD	Mouse	5
6	487–505	KCHTCQNGGTVILMDGKCL	Human	3
		RCYPCLNGGTIILLDGQCL	Mouse	2
7	505–521	LCACPFKFEGIACEISK	Human	4

Table II. Epitope composition of polyclonal anti-human C9 sera

Epitope	Antibody	Number of positives
1	A.P.	52
2	M47	1
3	A.P.	14
4	bC5	45
5	M42	12
6	M3	21
7	A.P.	32

165 positive clones obtained using a polyclonal antibody were divided into groups using monoclonal antibodies or the polyclonal antibody affinity purified (A.P.) on an individual clone. 16 clones were not classified in this procedure, but these were all very weak with the original polyclonal antibody.

Table III. Size requirement for antibody binding

Size of fragments base pairs	% Positive colonies		
	M47	M3	M42
100–200	0.6	2.3	3.2
200–400	3.1	4.0	2.6
400–600	18.5	10.6	11.6

1000 colonies from three libraries having different DNA insert sizes were transferred from Petri dishes to nitrocellulose filters and screened for binding to each monoclonal antibody using the colony blot procedure (Stanley, 1983).

clearly defined and changed in length. These conserved elements of secondary structure are shown beneath the sequence in Figure 3.

Mapping of antibody binding sites in C9

Libraries of 100–200 base pair random fragments of human C9 cDNA were screened in pEX 627 using a polyclonal antibody. Positive clones were identified as belonging to one of seven epitopes by challenging the polyclonal positive clones with monoclonal antibodies, and where this gave negative results, by repeating the procedure with the polyclonal antibody affinity purified on one of the polyclonal positive clones. Where tested, both procedures yielded the same pattern of positive clones showing that the antibody affinity purified in this way was essentially monospecific. 90% of the polyclonal positive clones could be identified as belonging to one of seven discrete antibody specificities in this manner. The remaining clones were all very weak in the original polyclonal antibody strain and by hybridisation with DNA probes made from the epitope clones in each region they could be assigned as belonging to one of the existing groups. The same 7 groups of positive clones were obtained using three different polyclonal antibodies raised in different laboratories to different preparations of C9, suggesting that they represent particularly immunogenic surface features of the C9 molecule. One epitope however corresponded to the anti-neoantigen monoclonal antibody described by Mollnes *et al.* (1985), and has its epitope situated in the 'hinge' region of C9. The random distribution of positive epitope clones for this antibody about a short common sequence (E4 in Figure 6) suggested that this was one of the most 'continuous' of all 7 epitopes of C9 detected by this method within the polyclonal antiserum. It is not possible however to distinguish between the possibilities that this sequence is either inaccessible in globular C9 or simply folded in a different manner. Clones from each subset of positives were sequenced by double stranded DNA sequencing. Each subset contained clones unique to one region of the C9 cDNA (Figure 6) indicating that specific sequences must be expressed to achieve antibody binding. Only in the case of epitopes 6 and 7 were some clones found that were positive with two different monospecific antibodies, but analysis of the DNA fragments in these colonies showed that this was because these two epitopes were suffficiently close to be present on a single DNA insert. The minimum region of overlap defined by these sequences ranged from 6–18 amino acids (Table I) and in every case contained more than two charged amino acids, suggesting a surface location. When compared with the secondary structural prediction it was found that the antigenic regions nearly always overlapped with high turn or coil probabilities, consistent with a surface location (see Figure 3). One exception to this was the neoantigenic epitope E4, but this region of the polypeptide might unfold in the extended form as conjectured above. Three epitopes contained two cysteine residues and might represent extended turn structures held in a loop by disulphide bond formation. Further data were obtained by examining the cross-reacting epitopes in mouse C9. By screening libraries of mouse C9 DNA fragments with the polyclonal antibody raised against human C9 it was found that three of the epitopes in human C9 were conserved in mouse. Aligning the sequences in these regions (Table I) shows that predominantly conservative changes occur, especially in the polar amino acids.

The seven epitopes unexpectedly were obtained at very different frequencies when the polyclonal positive clones were analysed into groups (Table II). For a random DNA fragment library all epitopes would be expected at the same frequency. Monoclonal antibody M47 in particular was poorly represented. When libraries containing different size cDNA fragments were screened with this monoclonal antibody and two control antibodies (M42 and M3, Table III), it was found that the proportion of positive clones was only different for the library containing small fragments. This suggests that this antibody requires more flanking region in order to define the correct folding of a small continuous epitope, or that the antibody epitope was partially discontinuous in nature, requiring amino acids from distant parts of the polypeptide chain to form an epitope. Indeed only a single clone (Figure 6) allowed the definition of the short epitope in Table I. Without this clone the epitope for M47 extended from residue 29–75 of human C9.

Discussion

The terminal complement component, C9, is capable of an unusual rearrangement of its protein structure which enables it

to interact with, and insert into lipid bilayers. During this transition the molecule unfolds to approximately double its length and cryptic sites capable of interaction with the lipid bilayer are exposed. A similar transition probably occurs in C6, C7 and C8 since the antibodies which bind to neoantigens in C9 also cross react with these earlier components (Tschopp *et al.*, 1986). Elucidation of the molecular mechanism of this transition is of interest not only in complement action, but also in the understanding of membrane protein insertion in general.

Analysis of the C9 amino acid sequence in several species supports the view that the C9 molecule is composed of several small globular domains. Surprisingly, the membrane interacting domain is not located at the physical end of the polypeptide chain, but an unfolding 'hinge' region in the molecule could account for its displacement and the differential access of several surface topographical features in the transition from globular to extended C9.

C9 in its monomeric form is a stable globular protein which shows little tendency to interact with biological membranes. This is in contrast to the related protein, perforin, found in cytotoxic and natural killer cells which spontaneously inserts into biological membranes in the presence of calcium ions once it is released from the effector cell. In this respect C9 most likely represents a later evolutionary form of this killing molecule, having evolved a mechanism which allows its constitutive secretion into the blood stream without damage to normal cells in the blood vessel walls. Two additional properties would be expected in the C9 molecule that are not present in the perforin molecule for this to occur. The first of these is a mechanism to hold the molecule in its non-cytolytic globular conformation. We have shown here that a possible explanation of this stability is an intra-molecular ligand-receptor like interaction which is conserved in C9 molecules but is absent in the perforin sequence (Lowrey *et al.*, 1987). In this respect C9 may be an example of a molecule from which both ligand and receptor have evolved (Pfeffer and Ullrich, 1985). The second necessity for a constitutively secreted killing molecule is the evolution of a highly specific activation mechanism. It is of interest therefore to note that both C8 α and β chains are C9-like (Haefliger *et al.*, 1987; Rao *et al.*, 1987) and that C8α is apparently incorporated into the tubular C9 complex (Podack, 1984) consistent with its close relation to C9. Taken with the fact that C8 addition to the membrane attack complex is essential for C9 binding, these findings suggest that C8 and C9 evolved as a 2-stage killing process from an ancestral perforin-like gene.

At 4°C it has been shown that C9 can reversibly associate with C5b-8 assembled into a target membrane (Boyle *et al.*, 1978). If a monoclonal antibody directed against the amino-terminal portion of C9 is included with a marker inside erythrocyte ghosts before addition of C9, no efflux of marker is observed after the ghosts are subsequently warmed up to 37°C (antibody C9-34, Morgan *et al.*, 1984). This has been interpreted as an inhibition of C9 polymerisation (Stanley *et al.*, 1986), since the amino-terminal portion of C9 polymerisation is unlikely to be accessible to the cytoplasmic face of the membrane after polymerisation, and because the antibody will inhibit zinc ion catalysed polymerisation *in vitro* (D.Maldonado and K.K.Stanley, unpublished). The binding site for globular C9 must therefore lie on a portion of C8 close to, or within the unstable pore created by C5b-8 since otherwise the antibody could not gain access to C9 from the inside of the erythrocyte ghosts. After warming up for a few seconds to 37°C in the absence of antibodies, C9 becomes irreversibly bound to the membrane and trypsin sensitivity, characteristic of the monomeric form of C9 is lost (Boyle *et al.*, 1978), presumably indicating a transition to the extended form of C9. The rearrangement of globular domains in the extended conformation of C9 could be driven by the formation of an amphipathic helix in the hydrophobic environment of the membrane as described here. By forming a membrane interacting element after entry into the lipid bilayer the insertion event would be rendered effectively irreversible.

Analysis of polyclonal antibodies using libraries of random DNA fragments expressed in *E. coli* provides a rapid method for mapping their binding sites. In essence the method is the same as using overlapping synthetic peptides (for review see Atassi, 1984), except that the size of the peptide probe can be made larger, possibly allowing a more native fold of the protein. Synthesis as a fusion protein with *E. coli* β-galactosidase may have an additional stabilizing effect. Using plasmid expression vectors for this purpose has the advantage over λ-vectors (Nunberg *et al.*, 1984; Mehra *et al.*, 1986) that DNA sequencing is more simple, allowing a finer map to be produced. For this method to work the antibody must be capable of binding to SDS-denatured antigen in a Western blot. In the case of C9, antibodies were also detected which had partially discontinuous epitopes, since they required larger lengths of DNA to express the antibody binding site. Furthermore, in the case of the two adjacent epitopes, E6 and E7, some clones expressing the sequence for both binding sites were capable of binding only one antibody. This again suggests some specificity to the folding of the fusion protein even after SDS treatment and transfer to nitrocellulose. Care should perhaps be exercised in calling the regions defined by this method as 'epitopes' since the stretch of polypeptide chain so defined might represent only the major part, rather than necessarily the whole of an antibody binding site. Nevertheless this method provides a quick check on secondary structural predictions by mapping loops of the protein likely to be situated at the surface of the protein. Knowing the major binding sites of antibodies having inhibitory activity allows possible functional surfaces of the protein to be defined for analysis by *in vitro* mutagenesis. Thus, antibodies against epitopes E2 and E5 can inhibit the flux of small molecules through complement lesions (Morgan *et al.*, 1984) suggesting that parts of domains 1 and 5 are close to the lumen of the pore. In addition the fusion protein clones which express these surface located portions of the polypetide chain are particularly suitable as antigens for making antibodies of predetermined specificity which will bind to native protein. The affinity purification of polyclonal antibodies on these 'epitope' clones also gives monospecific reagents (Tooze and Stanley, 1986) which in many cases are more useful than monoclonal antibodies because they have lower background binding and are composed of antibodies having a variety of dissociation constants.

We have described two recombinant DNA approaches which together give a wealth of information about the folding and surface features of a protein. When X-ray crystallographic data is difficult to obtain (as in the case of C9 which has so far resisted attempts to crystallize it) these techniques can give sufficient information about the important structural and functional regions of a protein to allow functional hypotheses to be made and then tested by site-directed mutagenesis.

Materials and methods

Cloning of mouse and trout C9 cDNA

Mouse and trout liver were pulverised in liquid nitrogen and homogenised in 6 M guanidine HCl, 1 mM DTT, 20 mM acetate buffer pH 7.0. RNA was selectively precipitated with 0.5 vols ethanol as described by Cox (1968) and modified

by Deeley *et al.* (1977). After two cycles of solubilisation and precipitation the guanidine HCl solution was extracted with buffered phenol, chloroform and isoamyl alcohol (24:24:1, v/v/v), and precipitated with 0.5 vols of ethanol. The pellet was taken up into water and reprecipitated with sodium acetate (0.3 M) and ethanol (2.5 vols). Poly(A)$^+$-RNA was prepared by the method of Werner *et al.* (1984) and cDNA libraries constructed in pEX1 (Stanley and Luzio, 1984) using adaptor oligonucleotides as described by Haymerle *et al.* (1986). The mouse liver library was screened with a cDNA coding for human C9 radioactively labelled by the method of Feinberg and Vogelstein (1983). After washing the filters at intermediate stringency (1 × SSC buffer at 42°C), areas containing positive colonies were scraped off the agar plates and resuspended in 5 ml of broth. 0.1 ml of this was kept at −80°C in freezing medium while plasmid DNA was amplified and then extracted from the remainder. The mixed plasmid preparations were then cut with *Bam*HI to release the cDNA inserts and analysed by Southern blotting with the human C9 probe. The preparation with the largest hybridising fragment was chosen and the corresponding frozen aliquot of cells used to re-plate and colony purify the clone. A similar procedure was used to screen for trout C9 using mouse C9 as a probe. In both cases clones containing the full open reading frame were obtained. The clones were fragmented by ordered deletion (Labeit *et al.*, 1987) and random shotgun approaches and sequenced using the dideoxy method (Sanger *et al.*, 1977). All sequences were determined on both strands of the DNA.

Construction of epitope libraries

Human C9 cDNA (5 μg) was fragmented by digestion with DNase I in the presence of manganese ions and repaired using the Klenow fragment of *E. coli* DNA polymerase. Adaptor oligonucleotides (Haymerle *et al.*, 1986) were separately ligated to these fragments and to *Bam*HI cut pEX627 plasmid. pEX627 is a derivative of pEX1, having bases 628−3195 deleted. It produces a smaller fusion protein than pEX1 and is easier to use for double stranded plasmid sequencing. The adaptor-ligated fragments were fractionated on a low melting point agarose gel, and fragments of 100−200 base pairs in length were excised, melted and ligated to the adapted vector at 37°C. DNA was then extracted from the agarose gel ligation mix using phenol and chloroform, and precipitated with sodium acetate and ethanol. Competent cells of strain pop 2136 were used to make a library of about 2×10^4 transformants which were screened according to published procedures (Stanley, 1983; Stanley and Luzio, 1984) with a polyclonal antiserum raised against C9 protein purified by affinity purification on a monoclonal antibody column (Morgan *et al.*, 1983) or using commercial antisera against biochemically purified C9. 165 positive clones were picked onto cellulose nitrate filters and, after expression, challenged with monoclonal antibodies and the polyclonal antiserum affinity purified on individual 'epitope' clones. This step was carried out by spreading about 1000 colonies from a colony purified clone onto a nitrocellulose filter and processing as described for the colony blot to obtain a filter with solubilsed fusion protein and *E. coli* proteins bound. This filter was then incubated with polyclonal antiserum diluted 1:100 for 16 h at 4°C, and washed 4 × in phosphate buffered saline containing 0.1% gelatin, 0.1% Triton X-100, 1 mM benzamidine and 40 μg/ml phenylmethylsulphonyl fluoride. Specific antibody was eluted by incubation for 3 min at 0°C in Tris glycine pH 2.5 containing 0.1% gelatin. Eluted antibody was immediately neutralised in an appropriate amount of 2 M Tris base and used immediately to stain filters of the original 165 epitope clones. These procedures are described in detail elsewhere (Stanley, 1987).

Clones selected for sequencing were grown up in 40 ml cultures in order to isolate plasmid DNA. This was then subjected to double stranded sequencing as described (Chen and Seeburg, 1985) using the following primers: 5′ sequencing primer (pEX627); GAATTATTTTTGATGGCGTTAACTCGGCG 3′-sequencing primer: CTAGAGCCGGATCGATCCGGTC.

References

Amiguet,P., Brunner,J. and Tschopp,J. (1985) *Biochemistry*, **24**, 7378−7334.

Argos,P. (1987) *J. Mol. Biol.*, **193**, 385−396.

Atassi,M.Z. (1984) *Eur. J. Biochem.*, **145**, 1−20.

Barker,W.C., Ketcham,L.K. and Dayhoff,M.O. (1978) In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*. Vol. 5, suppl. 3, pp. 359−362, NBRF, Washington DC.

Boyle,M.D.P., Langone,J.J. and Barsos,T. (1978) *J. Immunol.*, **120**, 1721−1725.

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

Chen,E.J. and Seeburg,P.H. (1985) *DNA*, **4**, 165−170.

Chou,P.Y. and Fasman,G.D. (1974a) *Bichemistry*, **13**, 211−221.

Chou,P.Y. and Fasman,G.D. (1974b) *Biochemistry*, **13**, 222−245.

Cox,R.A. (1968) *Methods Enzymol.*, **12**, 120−129.

Deeley,R.G., Gordon,J.I., Burns,A.T.H., Mullinix,K.P., Binastein,M. and Goldberger,R.F. (1977) *J. Biol. Chem.*, **252**, 8310−8319.

Doolittle,R.F. (1985) *Trends Biochem. Sci.*, **10**, 233−237.

Drenth,J., Low,B.W., Richardson,J.S. and Wright,C.S. (1980) *J. Biol. Chem.*, **55**, 2650−2655.

Esser,A.F. (1987) *Fed. Proc.*, in press.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6−13.

Hadding,U. and Müller-Eberhard,H.J. (1969) *Immunology*, **16**, 719−735.

Haefliger,J., Tschopp,J., Nardelli,D., Wahli,W., Kocher,H.-P., Tosi,M. and Stanley,K.K. (1978) *Biochemistry*, in press.

Hammer,C.H., Shin,M.L., Abramovitz,A.S. and Mayer,M.M. (1977) *J. Immunol.*, **119**, 1−8.

Haymerle,H., Herz,J., Bressan,G.M., Frank,R. and Stanley,K.K. (1986) *Nucleic Acids Res.*, **14**, 8615−8624.

Innerarity,T.L., Weisgraber,K.H., Arnold,K.S., Rall,S.C. and Mahley,R.W. (1984) *J. Biol. Chem.*, **259**, 7261−7267.

Ishida,B., Wisnieski,B.J., Lavine,C.H. and Esser,A.F. (1982) *J. Biol. Chem.*, **257**, 10551−10553.

Knott,T.J., Pease,R.J., Powell,L.M., Wallis,S.C., Rall,S.C., Innerarity,T.L., Blackhart,B., Taylor,W.H., Marcel,Y., Milne,R., Johnson,D., Fuller,M., Lusis,A.J., McCarthy,B.J., Mahley,R.W., Levy-Wilson,B. and Scott,J. (1986) *Nature*, **323**, 734−742.

Labeit,S., Goody,R.S. and Lehrach,H. (1987) *Methods Enzymol.*, **5**, 173−177.

Lowrey,D.M., Rupp,F., Aebischer,T., Grey,P., Hengartner,H. and Podack,E.R. (1987) *Fed. Proc.*, **46**, 920.

Mehra,V., Sweetser,D. and Young,R.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7013−7017.

Mollnes,T.E., Lea,T., Harboe,M. and Tschopp,J. (1985) *Scand, J. Immunol.*, **22**, 183−195.

Morgan,B.P., Daw,R.A., Siddle,K., Luzio,J.P. and Campbell,A.K. (1983) *J. Immunol. Method.*, **64**, 269−281.

Morgan,B.P., Luzio,J.P. and Campbell,A.K. (1984) *Biochem. Biophys. Res. Commun.*, **118**, 616−622.

Nonaka,M., Yamaguchi,N., Natsumme-sakai,S. and Takahashi,M. (1981) *J. Immunol.*, **126**, 1489−1494.

Nunberg,J.H., Rodgers,G., Gilbert,J.H. and Snead,R.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3675−3679.

Pfeffer,S. and Ullrich,A. (1985) *Nature*, **313**, 184.

Podack,E.R. and Tschopp,J. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 574−578.

Podack,E.R., Tschopp,J. and Müller-Eberhard,H.J. (1982) *J. Exp. Med.*, **156**, 268−282.

Podack,E.R. (1984) *J. Biol. Chem.*, **259**, 8641−8647.

Rao,A.G., Howard,O.M.Z., Ng,S.C., Whitehead,A.S., Colten,H.R. and Sodetz, J.M. (1987) *Biochemistry*, in press.

Rao,J.K.M .and Argos,P. (1986) *Biochim. Biophys. Acta*, **869**, 197−214.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.

Silverman,R.J. and Nelsestuen,G.L. (1986) *Biochemistry*, **25**, 852−860.

Stanley,K.K. (1983)*Nucleic Acids Res.*, **11**, 4077−4092.

Stanley,K.K. and Luzio,J.P. (1984) *EMBO J.*, **3**, 1429−1434.

Stanley,K.K,. Kocher,H.-P., Luzio,J.P., Jackson,P. and Tschopp,J. (1985) *EMBO J.*, **4**, 375−382.

Stanley,K.K., Page,M., Campbell,A.K. and Luzio,J.P. (1986) *Mol. Immunol.*, **23**, 451−458.

Stanley,K.K. (1987) in Walker,J.M. (ed.), *Methods in Molecular Biology*. Humana Press, New Jersey.

Tooze,S.A. and Stanley,K.K. (1986) *J. Virol.*, **60**, 928−934.

Tschopp,J., Podack,E.R. and Müller-Eberhard,H.J. (1982a) *Proc. Natl. Acad. Sci. USA*, **79**, 7474−7478.

Tschopp,J., Müller-Eberhard,H.J. and Podack,E.R. (1982b) *Nature*, **298**, 534−538.

Tschopp,J., Masson,D. and Stanley,K.K. (1986) *Nature*, **322**, 831−834.

Werner,D., Chemla,Y. and Herzberg,M. (1984) *Anal. Biochem.*, **141**, 329−336.

Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L., Goldstein,J.L. and Russell,D.W. (1984) *Cell*, **39**, 27−38.

Received on March 24, 1987; revised on April 24, 1987