Functional Insights from the Structure of the Multifunctional C345C Domain of C5 of Complement*

Received for publication, November 22, 2004, and in revised form, December 9, 2004
Published, JBC Papers in Press, December 14, 2004, DOI 10.1074/jbc.M413126200

Janice Bramham‡, Chuong-Thu Thai§, Dinesh C. Soares‡, Dusan Uhrín‡, Ronald T. Ogata§, and Paul N. Barlow‡¶

From the ‡Schools of Chemistry and Biological Sciences, University of Edinburgh, Edinburgh EH9 3JJ, Scotland, United Kingdom and §Torrey Pines Institute for Molecular Studies, San Diego, California 92121-1122

The complement protein C5 initiates assembly of the membrane attack complex. This remarkable process results in lysis of target cells and is fundamental to mammalian defense against infection. The 150-amino acid residue domain at the C terminus of C5 (C5-C345C) is pivotal to C5 function. It interacts with enzymes that convert C5 to C5b, the first step in the assembly of the membrane attack complex; it also binds to the membrane attack complex components C6 and C7 with high affinity. Here a recombinant version of this C5-C345C domain is shown to adopt the oligosaccharide/oligonucleotide binding fold, with two helices packed against a five-stranded β-barrel. The structure is compared with those from the netrin-like module family that have a similar fold. Residues critical to the interaction with C5-convertase cluster on a mobile, hydrophobic interstrand loop that protrudes from the open face of the β-barrel. The opposite, helix-dominated face of C5-C345C carries a pair of exposed hydrophobic side chains adjacent to a striking negatively charged patch, consistent with affinity for positively charged factor I modules in C6 and C7. Modeling of homologous domains from complement proteins C3 and C4, which do not participate in membrane attack complex assembly, suggests that this provisionally identified C6/C7-interacting face is indeed specific to C5.

A complement-mediated response to infection is fundamental to good health, but inappropriate complement activity underlies the symptoms of numerous inflammatory disorders (1). Activation of complement, and the ensuing attack on pathogens, entails a sequence of intermolecular recognition events, enzymatic cleavages, and assemblies of multiprotein complexes. The ~30 fluid-phase and membrane-associated proteins participating in the complement system have been well characterized at the sequence level, and their respective roles are broadly understood (2, 3). There is, however, little understanding at atomic resolution of the interplay between the components. In particular, the sequence in which the five soluble, terminal components of complement (C5, C6, C7, C8, and C9) assemble to form the remarkable lipid bilayer-penetrating membrane attack complex (MAC)1 has been known for many years (4). But the network of protein-protein interactions entailed in forming this stable lytic complex and the involvement of specific amino acids remain a mystery. The key to progress in this area will be more three-dimensional structural information.

Assembly of the MAC is initiated by proteolytic cleavage of C5 by the trimeric enzyme, C5 convertase, at the target cell surface to generate C5a and a metastable species, C5b. C5b has the transient ability to interact tightly with C6 (5). The C5bC6 complex subsequently serves as a nucleation site for sequential assembly of C7, C8, and n molecules of C9 to create the MAC. Mature C5 is a heterodimer consisting of α and β chains of 115 and 75 kDa, respectively. It is evolutionarily related to the earlier complement components C3 and C4. Unlike the majority of proteins of the complement system, C3, C4, and C5 are not entirely modular in their composition. Nonetheless it is surprising, given the intense interest in this family of proteins that has persisted over several decades, that little is known of their structure. For example, although the solution structure of the much smaller C5αa fragment has been solved (6), in the case of C5β there is currently no three-dimensional structural information available.

An opportunity to address this lack of structural information arose from the suggestion that the C-terminal ~150-amino acid residue portions of the α chains of C3 and C5 and of the γ chain of C4 are independently folded units (7, 8). Not only do these regions exhibit high sequence similarity with one another but they also display an ~19% sequence identity to the C-terminal segment of the Caenorhabditis elegans UNC-6, a laminin-related netrin protein involved in axonal path finding (7). Furthermore, homologies with C-terminal segments of procollagen C-proteinase enhancer proteins (PCOLCEs) and of secreted frizzled-related proteins have been noted (8), and this family of domains has been named the NTR (netrin-like) module. The more recently solved three-dimensional structure of the PCOLCE-1 NTR module (9) confirmed structural similarities with the N-terminal domains of tissue inhibitor of metalloproteinases (TIMP)-1 and -2 (10, 11), the laminin-binding domain of agrin (12), and the oligosaccharide/oligonucleotide-binding (OB) domains of some single-stranded DNA-binding proteins (13).

1 The abbreviations used are: MAC, membrane attack complex; C5-C345C, residues Ala1512 to the C-terminal Cys1669 of human C5; FIMAC, factor I membrane attack complex; OB, oligosaccharide/oligonucleotide binding; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; NTR, netrin-like; PCOLCE, type I procollagen C-proteinase enhancer protein; r.m.s.d., root mean square deviation; SPR, surface plasmon resonance; TIMP, tissue inhibitor of metalloproteinase; WT, wild type; PDB, Protein Data Bank.

* This work was supported by the Medical Research Council of the UK, the Wellcome Trust, and National Institutes of Health Grant GM28831. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Schools of Chemistry and Biological Sciences, Joseph Black Chemistry Bldg., University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland, UK. Tel.: 44-131-650-4727; Fax: 44-131-650-7055; E-mail: Paul.Barlow@ed.ac.uk.

‡ The atomic coordinates (code 1XWE) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ © 2005 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in U.S.A.

This paper is available online at http://www.jbc.org
Expression of the segment of C5 corresponding to its C345C domain (14) followed by analysis using CD and NMR confirmed that these amino acid residues fold to form a compact three-dimensional structure (15). Furthermore, C5-C345C, unlike the C345C domain of C3, is able to bind to both C6 and C7 in surface plasmon resonance (SPR)-based assays (14). In further work, C5-C345C was shown to inhibit recruitment of C7 by C5bC6 through an interaction between C5-C345C and the pair of factor I membrane attack complex (FIMAC) domains, also called factor I modules (FIMs), at the C terminus of C7 (16). Thus the C5-C345C domain provides at least part of the interacting surface between C5b and C7 in formation of the MAC. The C345C domain also harbors a region that interacts with the C5 convertase (17), although the cleavage site itself lies some 800 residues away toward the N terminus of the a chain of C5.

Although the fold of C5-C345C might be anticipated to resemble the fold of the NTR module from PCOLCE-1 (9), the of C5. actions are now revealed. surface patches likely to be involved in protein-protein interactions were recently identified as being functionally critical and the location of the C345C domain of C3, is able to bind to both C6 and C7 in surface plasmon resonance (SPR)-based assays (14). In further work, C5-C345C was shown to inhibit recruitment of C7 by C5bC6 through an interaction between C5-C345C and the pair of factor I membrane attack complex (FIMAC) domains, also called factor I modules (FIMs), at the C terminus of C7 (16). Thus the C5-C345C domain provides at least part of the interacting surface between C5b and C7 in formation of the MAC. The C345C domain also harbors a region that interacts with the C5 convertase (17), although the cleavage site itself lies some 800 residues away toward the N terminus of the a chain of C5. Although the fold of C5-C345C might be anticipated to resemble the fold of the NTR module from PCOLCE-1 (9), the sequence identity is low (see Fig. 1A) and disulfide bonding patterns are different, 1–4, 2–5, and 3–6 in the PCOLCE-1 NTR module compared with 1–3, 2–6, and 4–5 in the C3 equivalent (and therefore by inference in the C5 example). The C5-C345C sequence is longer and contains fewer prolines (147 residues including three prolines) compared with the PCOLCE-1 NTR module sequence (119 residues including 11 prolines). An experimentally determined three-dimensional structure of C5-C345C would therefore represent an important advance in understanding the basis, at atomic resolution, for the case of C5, since the anaphylatoxic C5a fragment structure was determined in 1989 (6). The new structure allows the advance in understanding the basis, at atomic resolution, for the at least in part, from a paucity of assignments for the residues in this region of the sequence. No covariant linkage was therefore defined between these residues. As the calculations progressed, the ambiguously assigned distance restraints were “filtered” iteratively to eliminate assignment possibilities contributing less than 1% to the total NOE, and redundant restraints (duplicates) were also removed. A total of 100 structures were calculated from which a representative ensemble of 40 structures, with the lowest NOE-derived energies, was selected. The quality of the ensemble of structures was checked with PROCHECK (24). The NOE-derived distance restraints used for the structure calculations and the coordinates of the ensemble of 40 structures of C5-C345C have been deposited in the Protein Data Bank under accession number 1XWE.

Structure of the C345C Domain of Complement C5

Here we report the use of solution NMR to solve the structure of C5-C345C. We thus provide the first new structural information for the C3/C4/C5 family of proteins since the structures of the C3d and C4d fragments were solved (18, 19) and, in the case of C5, since the anaphylatoxic C5a fragment structure was determined in 1989 (6). The new structure allows the construction of useful models of the C345C domains from C3 and C4. The positions within the structure of residues previously identified as being functionally critical and the location of surface patches likely to be involved in protein-protein interactions are now revealed.

EXPERIMENTAL PROCEDURES

Protein Preparation—PET15b vectors encoding the amino acid residues of C5 from Ala1512 to the C-terminal residue Cys1658 both with and without the point mutation F1613A were constructed as described previously (14). The isotopically enriched recombinant proteins were overexpressed in the Escherichia coli strain Origami (Novagen, Madison, WI) and purified as described previously (14). For NMR studies, 15N- and 13C,15N-protein samples (0.5–1.0 mM) were prepared in buffer containing 20 mM sodium phosphate, 100 mM NaCl, 5 μM EDTA, 0.02% NaN3, pH 6.0, in 95% H2O, 5% D2O.

Binding Studies—Affinities of the recombinant wild-type and F1613A versions of C5-C345C for C6 and C7 were measured using SPR (26, 27) and the GenBankTM nonredundant databases, using the program MUSCLE (28, 29). The multiple sequence alignment (Fig. 1B) was manually edited to ensure the most plausible alignment of conserved amino acid residues and of secondary structure elements as predicted by PsiPred (30) between the target and template. The three putative disulfide bridges and the longer predicted C-terminal α-helix of both the C3 and C4 C345C domains were restrained during model building. Twenty models were generated in each case, and the one with the lowest objective function score (25) was selected as the representative model. The representative model structures were protonated using the program REDUCE (31) and were checked for valid stereochemistry using PROCHECK (24).

RESULTS

Recombinant F1613A Mutant Binds C6/C7—The protein fragment, C5-C345C (residues Ala1512 to the C-terminal Cys1658 of human C5), with an N-terminal His tag was overexpressed in the E. coli strain Origami. The use of a bacterial expression system facilitated isotopic enrichment, and the Origami strain was selected because its oxidizing intracellular...
Structure of the C345C Domain of Complement C5

**A.** Structure-based sequence alignment of C345C/NTR domains. The alignment of C5-C345C with the C-terminal domain of PCOLCE-1, the N-terminal domain of agrin, and the N-terminal domains of TIMP-1 and TIMP-2 was generated using the program MATRAS (38). The extent of average secondary structure is represented above the sequences (solid, strand; hatched, helix), whereas residues within the C5 secondary structure are boxed. Buried (>95%) and surface-exposed (<30%) hydrophobic residues are indicated by filled squares and open squares, respectively. The disulfide linkages of C5 are shown.

**B.** Multiple sequence alignment of C345C from C3, C4, and C5 in a range of species. An initial multiple sequence alignment was derived using the program MUSCLE (28, 29) and edited manually. Residues in C5-C345C that are >95% buried are indicated by yellow highlighting. White letters against black indicate hydrophobic residues that are >30% surface-exposed. Gaps in chicken and rat sequences of C5 correspond to sequence information either missing from the database or erroneously deposited/translated.

**FIG. 1.** Sequence alignments. A, structure-based sequence alignment of C345C/NTR domains. The alignment of C5-C345C with the C-terminal domain of PCOLCE-1, the N-terminal domain of agrin, and the N-terminal domains of TIMP-1 and TIMP-2 was generated using the program MATRAS (38). The extent of average secondary structure is represented above the sequences (solid, strand; hatched, helix), whereas residues within the C5 secondary structure are boxed. Buried (>95%) and surface-exposed (<30%) hydrophobic residues are indicated by filled squares and open squares, respectively. The disulfide linkages of C5 are shown. B, multiple sequence alignment of C345C from C3, C4, and C5 in a range of species. An initial multiple sequence alignment was derived using the program MUSCLE (28, 29) and edited manually. Residues in C5-C345C that are >95% buried are indicated by yellow highlighting. White letters against black indicate hydrophobic residues that are >30% surface-exposed. Gaps in chicken and rat sequences of C5 correspond to sequence information either missing from the database or erroneously deposited/translated.
environment is conducive to formation of disulfide bonds (32). After thrombin cleavage of the His tag, four extra residues (Gly-Ser-His-Met) remained at the N terminus of the C5-C345C sequence. Protein expression levels in rich media were typically 4 mg liter^{-1} but only 0.5 mg liter^{-1} in Martek 9-labeled media. Yields were improved 4–5-fold using a construct with the point mutation F1613A. To assess any structural perturbations that might be introduced by such a mutation, $^{15}$N,$^{13}$C,$^{1}$H-HSQC spectra of $^{15}$N-labeled wild-type and F1613A C5-C345C samples were compared (data not shown). Nearly all the resonances coincide. Significant chemical shift differences were noted only for those peaks corresponding to residues located close in sequence to the mutation, namely Ile$^{1609}$–Tyr$^{1617}$; of these only Asn$^{1612}$, Phe$/$Ala$^{1613}$, and Ser$^{1614}$ show major differences. This observation demonstrates the F1613A mutant of C5-C345C has a near identical structure to that of the native domain. To assay for functional activity, binding to C6 and C7 was measured (Fig. 2). As may be judged from the SPR-derived binding parameters (Table I), the affinities of the WT module, Pro$^{1557}$, Pro$^{1620}$, and Pro$^{1631}$, respectively, for the WT and mutant, respectively, Dashed lines in all cases are theoretical global fit SPR responses calculated for the kinetic parameters given in Table I. The $\chi^2$ values for the fits were 2.5, 11, 0.7, and 11 for the WT/C6, WT/C7, mutant/C6, and mutant/C7 sensograms, respectively. All data were obtained on different sensor chips.

### Table I

|        | C6       | C7       |
|--------|----------|----------|
|        | $K_D$    | $k_{on}$ | $k_{off}$ | $K_D$    | $k_{on}$ | $k_{off}$ |
| WT C5-C345C | 10 $2 \times 10^4$ | 2 $\times 10^{-4}$ | 3 $3 \times 10^4$ | 9 $10^{-5}$ |
| F1613A C5-C345C | 9 $2 \times 10^4$ | 2 $\times 10^{-4}$ | 2 $3 \times 10^4$ | 6 $10^{-5}$ |

(C*, C$, and H$). Tyr$^{1541}$ is unusual in that its H$^\delta$ and its H$^a$ nuclei have nondegenerate chemical shifts; a strong chemical exchange peak between the resonances of H$^\delta$ and H$^a$ and between H$^\gamma$ and H$^\delta$, indicates restricted rotation of its aromatic side chain (subsequently, the structure reveals that this side chain is indeed well buried within the core of the protein). The H$^a$ atom of Leu$^{1521}$ exhibits an unusually low chemical shift of 1.58 ppm (cf. average shift is 4.32 ppm). All three proline residues are in the trans conformation as evidenced by the differences in the chemical shifts, $\delta C^\alpha$–$\delta C^\gamma$ of 4.03, 5.00, and 4.88 ppm for Pro$^{1557}$, Pro$^{1620}$, and Pro$^{1631}$, respectively ($\delta C^\alpha$–$\delta C^\gamma$ is 4.51 ± 1.17 ppm for trans and 9.64 ± 1.27 ppm for cis (33)), as well as strong NOE cross-peaks between the H's of the prolines and the H$^a$ of the preceding residues.

Subsequently, a structure calculation was undertaken using a total of 3544 NMR-derived distance restraints as detailed in Table II. Two disulfide (Cys$^{1513}$–Cys$^{1588}$ and Cys$^{1535}$–Cys$^{1658}$) bonds were added only after NOE-based calculations had established beyond a doubt the proximity and orientation of the contributing cysteine side chains. A third potential disulfide was not invoked because, although the remaining two cysteine residues are close in space, there is insufficient NOE-derived evidence to judge whether their side chains are appropriately juxtaposed. Similarly, distance restraints based on 26 inferred inter-$\beta$-strand H-bonds were not added until the later stages of the structure calculation. A total of 40 structures, selected on the basis of lowest NOE-derived energy from 100 calculated ones, converged well in most regions as may be judged from a backbone overlay (Fig. 3A) and the values for r.m.s.d. in Table II. The r.m.s.d. of the C$\alpha$ coordinates of the 40 selected structures from those of the mean structure are plotted in Fig. 4A as a function of residue number and compared with the distribution of H$^\alpha$–H NOEs (Fig. 4B). Significantly fewer than average NOEs are exhibited by two stretches of residues within the sequence (Ile$^{1609}$–Phe$^{1615}$ and Thr$^{1637}$–Cys$^{1639}$) and by the N-terminal residues Ala$^{1512}$ and Asp$^{1513}$. This is reflected in the elevated r.m.s.d. values of their C$\alpha$’s and is also evident from inspection of the overlay in Fig. 3A. In the case of Ser$^{1637}$ and Ser$^{1638}$, the aforementioned lack of detectable amide signals would account in part for the dearth of NOEs.

**Description of the Structure**—For the purposes of the description below, and unless stated otherwise, a residue is designated as belonging to an $\alpha$-helix or $\beta$-strand in C5-C345C if it is so defined in the majority of the 40 members of the ensemble according to the Kabsch and Sander (34) criteria, as implemented in MolMol (35). Two views of the fold of the closest-to-the-mean C5-C345C structure are shown in Fig. 3B. The core of the structure is an OB-class fold that is most easily thought of as two orthogonal three-stranded, antiparallel twisted $\beta$-sheets composed from strands A$\beta$–B–C and strands A$\alpha$–D–E, where the superscripts N and C denote the N- and the C-terminal halves of strand A (Tyr$^{1541}$–Val$^{1552}$). There are two adjacent helices as follows: a short one (helix-1, Arg$^{1530}$–Ala$^{1534}$) composed of residues from near the N terminus of the module, and a longer and irregular one (helix-2, Leu$^{1643}$–Leu$^{1655}$) close to the C terminus of the module (and of the full-length protein). The two helices

**Description of the Structure**—For the purposes of the description below, and unless stated otherwise, a residue is designated as belonging to an $\alpha$-helix or $\beta$-strand in C5-C345C if it is so defined in the majority of the 40 members of the ensemble according to the Kabsch and Sander (34) criteria, as implemented in MolMol (35). Two views of the fold of the closest-to-the-mean C5-C345C structure are shown in Fig. 3B. The core of the structure is an OB-class fold that is most easily thought of as two orthogonal three-stranded, antiparallel twisted $\beta$-sheets composed from strands A$\beta$–B–C and strands A$\alpha$–D–E, where the superscripts N and C denote the N- and the C-terminal halves of strand A (Tyr$^{1541}$–Val$^{1552}$). There are two adjacent helices as follows: a short one (helix-1, Arg$^{1530}$–Ala$^{1534}$) composed of residues from near the N terminus of the module, and a longer and irregular one (helix-2, Leu$^{1643}$–Leu$^{1655}$) close to the C terminus of the module (and of the full-length protein). The two helices
are tilted with respect to one another but are essentially aligned with, and lie against, the convex face of the A_3^N-D-E sheet. Strand B (Val_1557-Lys_1568) extends beyond the A_3^C-B-C sheet so that its C-terminal part participates in a four-stranded anti-parallel sheet B_2^-A_3^N-D-E. In a small proportion of calculated structures, there are two segments to strand E, E_1^N(Ile_1616-Pro_1629) and E_2^- (Tyr_1626-Tyr_1629), interrupted by coil. Strand E_2^- which is assigned (within MolMol) in only a few structures, forms a small parallel β-sheet with strand C (Glu_1579-Lys_1584). In all of the C5-C345C structures, there is potential for H-bonds between the CO of Tyr_1619 and the NH of Thr_1581, and between the NH of residue Tyr_1619 and CO of Ile_1583, thus completing the hydrogen bond network that forms the barrel-like structure. Strand E_2^- which appears in all structures, is antiparallel to strand D (Gln_1586-Gly_1603). Thus the barrel has a “closed” side made up from the β-strands, and a more “open” side (to the right of the view in the left-hand panel of Fig. 3B) occupied by Tyr_1619, and the residues prior to E_2^-.

The N-terminal segment of the domain runs above one end of the barrel, from Cys_1512 to the top of helix-1. Cys_1512 is disulfide-linked to Cys_1568, which is located in the long CD loop; the CD loop crosses over the otherwise open end of the barrel from the A_3^N-D-E sheet to the A_3^N-D-E sheet (Fig. 3B). The 15N T_1/J_2 ratios (Fig. 4C) in some residues of the N-terminal segment (but not in the CD loop) suggest chemical exchange (i.e. microsecond to millisecond scale conformational fluctuations), whereas in both the N-terminal segment and the CD loop there is also some evidence of rapid (i.e. nanosecond and faster) motion from the heteronuclear NOE plot (Fig. 4D). At the bottom of the short helix-1 the transition to the long strand A contains Cys_1535 which is linked to Cys_1568, the C-terminal residue of the module. The BC loop caps off the other end of the barrel and corresponds to a dip in the heteronuclear NOE plot consistent with rapid motion, but there is no evidence of chemical exchange among these residues. Following strand D, there is a 14-residue loop that in a few members of the ensemble contains antiparallel β-strands from Leu_1607-Ile_1609 and Arg_1618-Ile_1618. This long DE loop protrudes prominently from the open side of the barrel-like structure, opposite to the pair of helices (Fig. 3B, left-hand panel). Although locally defined by °H, °H NOEs, the position of the tip of the loop relative to the remainder of the module is not experimentally defined as indicated by the very high r.m.s.d. values for these residues (Fig. 4A). The residues concerned (1610–1614) are highly mobile on both fast and chemical exchange time scales. After strand E_2^- a stretch of 13 residues makes the transition to the top of helix-2. From residue 1634 onward, this region (which includes the sequence Cys_1636-Ser-Ser-Cys_1639) is mobile on slow and fast time scales, is ill-defined by °H, °H NOEs, and forms a bulge above helix-2 (evident in Fig. 3B, right-hand panel).

Examination of the ensemble of calculated structures indicates that helix-2 is not a straight, regular α-helix over its entire length, a situation reminiscent of the equivalent regions of TIMP-1 and -2 (discussed below). Therefore, no inferred hydrogen bond-based restraints were used in this region so as to ensure the calculation was not biased. In some members of the ensemble, the first and last residues are classified as turns; and more significantly, in nearly half of the structures the α-helix (as assigned in MolMol) is broken by a bend or turn toward the middle. This is reflected in the broken nature of the helix as drawn by PyMol for the closest-to-the-mean structure (Fig. 3B).

The pattern of disulfide bond formation thus agrees with that predicted on the basis of disulfide mapping in C3 (36). The first Cys (1514) is disulfide-bonded to the third Cys (1588); and the second Cys (1535) is linked to the sixth Cys (1658). The third disulfide, involving the remaining fourth and fifth Cys residues (1636 and 1639), has presumably formed because biochemical analysis suggested that no free sulfhydryl groups are present in C5-C345C, but its presence could not be supported by the NOE data. It is possible that this bond exists only transiently due to the constraints placed by there being only two residues between these two cysteines.

Of 36 residues in C5-C345C that are >95% buried (on average in the ensemble) (see Fig. 1), three are likely to be charged in the solution conditions used. The alkyl chain of Lys_1584 is buried, but its amino group is exposed to solvent. However, Arg_1536 and Glu_1628 are completely buried and proximal, indicating the likelihood of an unusual ion pair or salt bridge connecting helix-1 with strand E of the barrel. Ala_1524 of helix-1, a stack of four residues located along one side of the second helical region (Phe_1642, Leu_1644, Phe_1649, and Ile_1653), two buried residues from the start of strand A (Ile_1539 and Ala_1543), two residues from strand D (Leu_1600 and Met_1606), and Pro_1631 from beyond strand E are all deeply buried in a hydrophobic core between the helices and the barrel along with the Arg/Glu salt bridge. Of the remaining deeply buried residues, all contribute to the hydrophobic core of the barrel.

Most of the solvent-exposed (>30%) hydrophobic residues lie in the DE loop, whereas Phe_1654 and Leu_1665 are exposed near the C terminus. Adjacent to this exposed pair of side chains is a patch of negatively charged side chains (glutamates 1528, 1648, and 1651; aspartates 1647 and 1652) that dominate the electrostatic surface of C5-C345C (see below). This surface (to the left in the left-hand panel of Fig. 3B) is likely to be exposed in full-length C5 because it is distal to the N terminus of the C345C domain.

**DISCUSSION**

Comparison with Other Structures—As predicted (8), the lowest NOE energy structure of the C5-C345C domain resembles the N-terminal domains of TIMP-1 (PDB ID 1UEA, chain B) and TIMP-2 (PDB ID 1B9R) (10, 11) (C° r.m.s.d., over 107 and 106 residues of 2.9 and 3.1 Å, respectively), the NTR domain of PCOLCE1 (PDB ID 1UAP) (9) (C° r.m.s.d. over 107 residues = 2.8 Å), and the laminin-binding domain of agrin (PDB ID 1JBC7) (12) (C° r.m.s.d. over 111 residues = 3.5 Å). A comparison of the structures is presented in Fig. 5, and a structure-based alignment of these domains is shown in Fig.
FIG. 3. Solution structure of C5-C345C. A, stereo-view (cross-eyed) showing backbone overlay of 40 lowest NOE energy structures from a total of 100 calculated. Structures are overlaid on all backbone atoms except for those in residues prior to Cys1514, 1610–1615 and 1635–1639. B, two orthogonal views of a PyMol schematic (www.pymol.org), with secondary structure assigned by the standard settings within PyMol, of the closest-to-mean structure from the ensemble in A, with annotated strands, loops, helices and cysteine residues (drawn as sticks). The schematics are colored sequentially from blue (N terminus) to red (C terminus).

1A. This work therefore confirms that the C345C domain of C5 is an example of an NTR module. For the purposes of further discussion, all four domains represented in Fig. 5 (and the equivalent TIMP-1 domain) will henceforth be referred to as NTR modules.

Many of the buried residues of C5-C354C that lie in strands are conserved or conservatively substituted in the other modules. These are drawn in Fig. 5; examples include the following: the first, second, third, fifth, and seventh residues of strand A; the first, third, fifth, and seventh residues of strand B; residues in positions 2–6 of strand D; and three residues (equivalent to Tyr1619, Leu1621, and Ile1627) in the strand E region. On the other hand, many of the side chains that make up the hydrophobic core between the β-barrel-like (β-) subdomain and the helix-rich (α-) subdomain are not well conserved, consistent with a higher degree of structural divergence in the helical subdomain. The buried partners that comprise the putative salt bridge, Arg1530 and Glu1628 of C5, are replaced by hydrophobic residues in other domains. Many of the exposed hydrophobic residues of C5-C345C lie in insertions, and examples include Leu1521, Pro1537, and Val1573 (that lies in the BC loop and is one of the most exposed residues in the protein) and five residues (including Ala1513 that replaces the wild-type Phe) in the DE protuberance. The conspicuous tandem pair of exposed hydrophobic residues near the C terminus (Phe1654 and Leu1655) is not conserved.

As would be expected from the conservation of buried residues, the β-subdomains of the known NTR module structures are minor variations on a common theme of an OB-fold. In TIMPs, strand E has two segments. In the NTR modules of PCOLCE-1, both strands D and E have two segments. In these cases, strand E is antiparallel to the C-terminal segment of strand D and runs parallel with strand C for a short way. In the agrin NTR module, strand E consists of only one segment, equivalent to E1, that is sandwiched between strand C (with which it is parallel) and strand D (anti-parallel). Thus C5-C345C is the only NTR module in which there is no evidence for a small mixed sheet formed by part of strand D, strand E1, and part of strand C, and although there is some hydrogen bonding between residues in equivalent parts of the sequence, the NOEs do not support regular secondary structure. The overall effect is that this side of the barrel is more open in C5-C345C (Fig. 5), and the DE loop is probably more flexible than in the other NTR modules.

In both the TIMP and PCOLCE-1 NTR modules, two disulfides staple the N-terminal region to the rest of the β-subdomain; the first cysteine connects to the crossover CD loop, the second disulfide connects the N-terminal segment either to the short sequence that interrupts strand E (in TIMP), or to the CD loop (in PCOLCE-1). The agrin NTR module and C5-C345C each have only one cysteine in this N-terminal region, which in both cases is disulfide-linked to the CD loop. Compared with C5-C345C, the other NTR modules have shorter N-terminal segments that take a more direct course to reach helix-1. None of them has an equivalent to the exposed leucine (1523) of C5.

In TIMP, the N-terminal stretch of the NTR module corresponds to the N terminus of the full-length protein and harbors its metalloprotein-binding site (11). Given its very different conformation, it seems unlikely that the N-terminal segment of C5-C345C plays a comparable role.

TIMP-2 lacks the prominent loop seen in C5 between strands D and E3, but its A and B strands are more extended. The PCOLCE-1 NTR module also lacks the long loop between strands D and E that forms such a prominent feature of C5-C345C. Between strands B and C of the agrin NTR module, there are two turns of α-helix, absent in C5-C345C and the other NTR modules, that lie across one end of the barrel. As in C5, the agrin loop between strands D and E is extended, although it is not as long or as prominent as in C5.

In all the NTR modules there are helical regions, packed against the convex surface of the A-D-E sheet, forming the α-subdomain. The NTR modules of TIMP-2 and agrin are at the
respectively N termini of the full-length proteins; in the case of TIMP-2, the α-subdomain abuts against the C-terminal domain. By contrast, in PCOLCE-1 and C5-C345C, the NTR modules are located at the C terminus and one surface of the α-subdomain is likely exposed to solvent. The more N-terminal helix of each of the four NTR modules is structurally conserved,
and with the exception of agrin, a disulfide connects a cysteine near the C terminus of helix-1 to a cysteine at or near the C terminus of the module. There is greater structural variation elsewhere in the α-subdomain. PCOLCE-1 and agrin lack the long sequence between strand E and helix-2 that in C5-C345C comprises the prominent bulge above helix-2. In TIMP-2, the C-terminal portion of the NTR module is composed of a helical turn and two separate segments of α-helix; this is reminiscent of the equivalent region of C5 in which there is also evidence of an interrupted α-helix. Helix-2 of PCOLCE-1 is the shortest among this set of modules. In both PCOLCE-1 and agrin, the C-terminal helix is straighter and more regular compared with helix-2 of C5-C345C with no evidence of any interruptions.

**Comparison with Other C345C Domains**—The C345C domains of C3 (residues 1518–1663) and C4 (residues 1595–1744) are both 26% identical to C5-C345C (Fig. 1B). The Arg/Glu salt bridge is conserved as are many other buried residues. Exceptions include the following: Ala1540 at the start of strand A that is replaced by an Asp or a Glu in C3 and C4, respectively; Val1545 that is conserved in C4 but replaced with a Thr or a Gly in some examples of C3; Cys1567 at the beginning of helix-2 is strongly predicted to begin well before the fifth Cys of C3 and C4 (this region forms a turn but is not classified as a helix in the majority of the C5-C345C NMR-derived structures) and to extend to the C terminus. In human C3, the beginning of the predicted helix and the preceding region (following strand E") are rich in negatively charged residues (seven within a 10-residue stretch), and this feature dominates the electrostatic representation of the C3-C345C surface (Fig. 6). By contrast, the equivalent region of the human C5-C345C surface is neutral, whereas in human C4 it is positively charged (Fig. 6). In all three proteins, the middle part of helix-2 is negatively charged, with human C5 having the most charge and human C4 the least. Another feature common to C3, C4, and C5 is that the open face of the β-subdomain is neutral or positively charged.

**Interpretation of Mutagenesis Data**—In previous work, the DE region of C5-C345C had been investigated as a site of possible functional significance on the basis that it is close to an “indel” (indels are evolutionary insertions or deletions of amino acid residues that result in length polymorphisms among members of a protein family). Deletion of the putative insertion Ser1628 and Leu1624 resulted in significant loss of hemolytic activity (40% of wild type, with normal expression levels) in full-length C5 (37). In light of the three-dimensional structure it seems likely that such a deletion, just prior to strand E", would affect the structural integrity of the β-subdomain or at least disrupt the open side of the barrel and the DE loop. Substitution of Leu1607→Tyr1611 by the sequence DFWE resulted in loss of all detectable hemolytic activity (37). In this case, the original sequence corresponds to a poorly structured region of the DE loop with no buried side chains, and the substitution would be unlikely to disrupt the structure of the domain. These observations therefore imply the DE loop in function.

Subsequently, alanine substitutions of residues from Gly1603 to Pro1624 were carried out in full-length C5 (17). Most substitutions had little or no effect on the hemolytic activity of C5 or its susceptibility to proteolytic cleavage. From the structure it can be seen that Tyr1619 is >95% buried within the β-subdomain, yet replacement with alanine resulted in only ~50% loss of activity. This implies that the structural framework provided by the β-barrel is able to accommodate such a substitution, which is intriguing because a bulky hydrophobic residue is conserved at this position in the other NTR modules of known structure.

Substitution of Ile1609 or Tyr1611 by alanine would not be expected to disrupt any local structure within the DE protuberance because their side chains are exposed, and indeed these mutations had little effect on hemolytic activity or proteolytic susceptibility. Substitution of the exposed Lys1616 on the other hand, produced mutant C5 molecules with both low hemolytic activity and decreased sensitivity to proteolytic activation. This is consistent with the pentapeptide substitution experiment and pinpoints Lys1616 as the functionally critical residue in that peptide.

Substitution of Phe1613 and Phe1615 also perturbed hemolytic activity and decreased proteolytic susceptibility to the classical pathway convertase (but not cobra venom factor, which is able to cleave both C3 and C5 and therefore presumably has a different recognition mechanism). Comparison of H5SC spectra for the wild-type and F1613A versions of C5-C345C proved that this mutation has no nonlocal structural effects, and indeed the F1613A mutant was used for structure determination in the current study. Phe1615 is also exposed, and mutation to...
Ala would be equally unlikely to disrupt structure. Therefore these mutagenesis results clearly identify three exposed side chains (1610, 1613, and 1615) as being specifically involved in an interaction, either with the convertase or within the full-length C5, that is critical for function. It is striking that these three residues, whose Ala substitutions cause 80–90% loss of C5 activity, are at the tip of the DE extension and that their side chains, as well as those of two of the four residues whose substitutions cause 50% loss of activity (Arg^{1616} and Tyr^{1617}), are located on the same side of the protuberance.

In the absence of a three-dimensional structure for C3, C4, or C5, the physical distance of the C345C module from the cleavage site (some 800 residues in terms of primary structure) is unknown. From the structure of the module, however, it can be seen that the DE loop exposes hydrophobic side chains (including those of the two critical Phe residues) and lies close to the N terminus of the module. Just prior to Cys^{1514} in the C5 sequence is Cys^{1509} that is (by extrapolation from disulfide-mapping in C3) disulfide-linked to Cys^{458}. Thus the C345C domain is likely closely coupled to further structured domains, and therefore, the DE extension could be buried in the interface between the C345C domain and the remainder of the C5 protein. In that case mutations of the DE loop might disrupt the arrangement of domains within the full-length protein and exert their functional influence indirectly. Arguing against this, however, is the observation that a peptide extending from Lys^{1504} to Arg^{1516} inhibited complement hemolytic activity and activation of C5 by the convertase pathway C5 convertase (but not cobra venom factor) (17). Furthermore, the consequences for inhibitory activity of alanine substitution within the peptide reflected the results of alanine-scanning mutagenesis in C5. The peptide studies are therefore more consistent with a direct interaction between the DE extension and the classical pathway convertase.

C5-C345C, but not the equivalent domain from C3, binds reversibly to C6 and C7, with a preference for C7 (14) (and Fig. 2), and the interaction with C7, but not C6, appears to be essential for the nonreversible formation of the MAC. The F1613A mutant used in the current structure determination retained the ability to bind C6 and C7 (Table I); indeed, none of several DE loop mutations in C5 influenced binding to C6 (17). Therefore, the C6/C7-binding site of C5-C345C is likely to lie elsewhere. Another set of mutants was therefore constructed in which deletions or insertions were made in the N-terminal region, the AB, BC, and CD loops, and the Cys-Ser-Ser-Cys bulge. These changes removed several of the exposed hydrophobic side chains evident in the C5-C345C structure (Fig. 1A). Nonetheless, all the mutants in this set showed full affinity for C6 and C7. One region that has not so far been explored by mutagenesis, however, is that made up from the exposed face of the α-subdomain. This part of the structure exhibits unexpected structural differences from the other NTR modules and diversity in terms of chemical character among C3, C4, and C5. As described above (see Fig. 6), in C5-C345C the region has electronegative and hydrophobic features not seen in the C4-C345C domain. The C345C domain of C3 has many more Asp and Glu residues than either the C4 or C5 equivalents and consequently has an overall electronegative character, but C5 lacks the patch of five negatively charged side chains that occur next to the exposed hydrophobic side chains near the C terminus of C5. C5-C345C has been shown (16) to bind to the FITMAC domain pair of C7. C7-FITMAC-II has been expressed alone and found not to interact with C5-C345C, thus implicating FITMAC-I in the interaction. There is no experimental structural information for FITMAC domains, but it is noteworthy that the

---

2 C.-T. Thai and R. T. Ogata, unpublished data.

---

REFERENCES

1. Mizuno, M., and Morgan, B. P. (2004) Curr. Drug Targets Inflamm. Allergy 3, 87–96
2. Walport, M. J. (2001) N. Engl. J. Med. 344, 1140–1144
3. Walport, M. J. (2001) N. Engl. J. Med. 344, 1058–1066
4. Muller-Eberhard, H. J. (1985) Biochem. Soc. Symp. 50, 255–256
5. DiScipio, R. G., Smith, C. A., Muller-Eberhard, H. J., and Hugli, T. E. (1983) J. Biol. Chem. 258, 10629–10636
6. Banyai, L., and Patthy, L. (1999) Protein Sci. 8, 1636–1642
7. Liepinsh, E., Banyai, L., Piccardo, G., Tressler, M., Pathy, L., and Otting, G. (2003) J. Biol. Chem. 278, 25882–25889
8. Banyai, L., and Patthy, L. (1999) Protein Sci. 8, 1636–1642
9. Liepinsh, E., Banyai, L., Piccardo, G., Tressler, M., Pathy, L., and Otting, G. (2003) J. Biol. Chem. 278, 25882–25889
10. Mollison, D. L., and Patthy, L. (1999) EMBO J. 18, 5236–5248
11. Banyai, L., Pathy, L. (1999) Protein Sci. 8, 1636–1642
12. Liepinsh, E., Banyai, L., Piccardo, G., Tressler, M., Pathy, L., and Otting, G. (2003) J. Biol. Chem. 278, 25882–25889
13. Liepinsh, E., Banyai, L., Piccardo, G., Tressler, M., Pathy, L., and Otting, G. (2003) J. Biol. Chem. 278, 25882–25889
14. Thai, C. T., and Ogata, R. T. (2003) J. Immunol. 171, 6565–6573
15. Brumham, J., Rance, M., Thai, C. T., Uhrin, D., Assa-Munt, N., Ogata, R. T., and Barlow, P. N. (2004) J. Biol. Chem. 279, 217–218
16. Thai, C. T., and Ogata, R. T. (2004) J. Immunol. 173, 4547–4552
17. Sandoval, A., Ri, R., Ostresh, J. M., and Ogata, R. T. (2000) J. Immunol. 165, 1066–1073
18. Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E., and Rini, J. M. (1998) Science 280, 1277–1281
19. van den Elen, J. M., Martin, A., Wong, V., Clemenza, L., Rose, D. R., and Isenman, D. E. (2002) J. Mol. Biol. 322, 1103–1115
20. Kay, L. E., Nicholson, L. A., Delaglin, F., Bax, A., and Torchia, D. E. (1992) J. Magn. Reson. 97, 359–375
21. Grezesz, S., and Bax, A. (1993) J. Am. Chem. Soc. 115, 12583–12584
22. Bragg, A. T., Adams, P. D., Clere, G. M., Delano, W. L., Gross, P., Grose, K., Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, C. J. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
23. Linge, J. P., and Nilges, M. (1999) J. Biomol. NMR 15, 51–59
24. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
25. Sali, A., and Blundell, T. L. (1993) J. Mol. Biol. 234, 779–815
26. Poeckmann, B., Biaoric, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O’Donovan, C., Phan, I., Pihlout, S., and Schneider, M. (2003) Nucleic Acids Res. 31, 365–370
27. O'Donovan, C., Martin, M. J., Gattiker, A., Gasteiger, E., Bairoch, A., and Apweiler, R. (2002) Brief. Bioinform. 3, 275–284
28. Edgar, R. C. (2004) BMC Central Bioinformatics 5, 113
29. Edgar, R. C. (2004) Nucleic Acids Res. 32, 1792–1797
30. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) Bioinformatics 16, 404–405
31. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) J. Mol. Biol. 285, 1735–1747
32. Bessette, P. H., Aaslund, F., Beckwith, J., and Georgiou, G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 13703–13708
33. Schubert, M., Labudde, D., Oschkinat, H., and Schnieder, P. (2002) J. Biomol. NMR 24, 149–154
34. Kabsch, W., and Sander, C. (1983) Biopolymers 22, 2577–2637
35. Koradi, R., Billeter, M., and Wuthrich, K. (1996) J. Mol. Graphics 14, 51–55
36. Dolmer, K., and Sottrup-Jensen, L. (1993) FEBS Lett. 315, 85–90
37. Low, P. J., Ai, R., and Ogata, R. T. (1999) J. Mol. Biol. 162, 6580–6588
38. Kawabata, T. (2003) Nucleic Acids Res. 31, 3367–3369
39. Shatsky, M., Nussinov, R., and Wolfson, H. J. (2002) Lect. Notes Comput. Sci. 2452, 235–250
40. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 10037–10041