Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 2799–2803, April 1989
Immunology

Structural homology of complement protein C6 with other channel-forming proteins of complement

(membrane attack complex/nucleotide sequence/amino acid sequence/cytolytic proteins)

DEB N. CHAKRAVARTI*, BULBUL CHAKRAVARTI*, CARLOS A. PARRA*, AND HANS J. MULLER-EBERHARD†

*Department of Immunology, Research Institute of Scripps Clinic, La Jolla, CA 92037; and †Bernhard Nocht Institute for Tropical Medicine, Hamburg, Federal Republic of Germany

Contributed by Hans J. Muller-Eberhard, December 8, 1988

ABSTRACT **The amino acid sequence of the amino-terminal half of the complement protein C6 has been found to show overall structural homology with the homologous regions of the channel-forming proteins C7, C8α, C8β, and C9. In addition, two specific cysteine-rich segments common to the amino-terminal regions of C7, C8α, C8β, and C9 also occur in their expected positions in C6, suggesting functional significance. Two cDNA clones encoding C6 were isolated from a human liver library in the bacteriophage vector λgt11. The predicted protein sequence contains an apparent initiation methionine and a putative signal peptide of 21 residues, as well as a site for N-glycosylation at residue 303. The sequence of the C6 protein reported here has 47–52% similarity with C7, C8α, C8β, and C9, as well as 31–38% similarity with thrombospondin, thrombomodulin, and low density lipoprotein receptor. The sequence data have been interpreted by using computer algorithms for estimation of average hydrophobicity and secondary structure.**

The membrane attack complex (MAC) of complement forms transmembrane channels causing membrane damage and cytolysis (for review, see ref. 1). The MAC has an approximate M_r of 1.7×10^6 and constitutes a supramolecular organization containing the five precursor proteins C5, C6, C7, C8, and C9. Structural interrelationships among the proteins participating in transmembrane channel formation including the C9-related protein (perforin) of cytotoxic lymphocytes have been suggested by shared antigenic properties (2). From cDNA-derived amino acid sequences, the existence of strong structural homologies has been established between C7 (3), C8α (4), C8β (5, 6), C9 (7, 8), and perforin (9, 10). In addition, these proteins share several specific cysteine-rich homologous segments or modules, which are found in other proteins as well.

The structure of C5 is distinctly different from that of these proteins and is instead related to complement proteins C3 and C4 (11). The only MAC precursor for which the chemical structure has not been known heretofore is C6. Although human C6 was proposed to be a serine protease (12), we have recently established that C6 does not function in the membrane attack pathway of complement as an enzyme (13). Human C6 is a single-chain glycoprotein with an estimated M_r of 95,000–128,000 and a carbohydrate content of 3.8–11.3% (12, 14–17). C6 is larger than C7, C8α, C8β, and C9. However, the sequence of 470 amino acid residues reported here, which spans approximately the amino-terminal half of the protein, represents the size of corresponding homologous regions of C7, C8α, C8β, and C9.‡ C6 is therefore a member of the channel-forming complement proteins.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Materials. Restriction endonucleases and DNA modification enzymes were from Pharmacia, Promega Biotec, Bethesda Research Laboratories, or Boehringer Mannheim. Deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate (dATP [α-^{35}S]) was from Amersham, 5′-[γ-^{32}P]ATP (Crude) was from ICN, and iodo-[2-^{3}H]acetic acid was from NEN. Sequenase DNA sequencing kit was from United States Biochemical. The phagemid vector pBS M13+, *Escherichia coli* XL1-Blue, and sequencing primers were from Stratagene. Colony/plaque screen hybridization transfer membrane obtained from NEN was used as hybridization solid supports. The mixture of synthetic oligodeoxynucleotides used to screen the phage cDNA library was synthesized in an Applied Biosystems model 380A DNA synthesizer using cyanoethyl phosphoramidite chemistry. Electrophoresis reagents were from Bio-Rad. The Vydac C-18 (end-capped) reversed-phase HPLC column was obtained from The Separations Group. L-1-Chloro-3-tosylamido-4-phenylbutan-2-one (TPCK)-treated trypsin was from Cooper Biomedical.

Methods. *Protein/peptide isolation and amino acid sequence analysis.* Isolation of human C6, reduction with dithiothreitol, and subsequent carboxymethylation with iodo-[2-^{3}H]acetic acid was carried out as described (13). Reduced and ^{3}H carboxymethylated C6 (2.6 nmol) in 300 μl of 50 mM NH_4HCO_3 was digested with 1% (wt/wt) TPCK-treated trypsin at 25°C for 1.5 hr. After addition of the same amount of enzyme, it was further incubated at 25°C for 1.5 hr followed by inactivation of trypsin with 2 mM diisopropyl phosphorofluoridate (Sigma). The freeze-dried digest was redissolved in 1 ml of 10% (vol/vol) trifluoroacetic acid (TFA), loaded onto a Vydac reversed-phase C-18 HPLC column, and washed for 5 min with solvent A [0.1% (vol/vol) TFA] followed by a linear gradient of solvent B [acetonitrile/water/TFA, 95:5:0.1 (vol/vol/vol)] for 2 hr to give 100% solvent B. Suitable peptide peaks collected were used in automated amino acid sequence determination by Edman degradation (18) in a Beckman 890M microsequencer as described (13).

Screening of the λgt11 cDNA library. The human liver cDNA library used was a generous gift of S. L. C. Woo (Baylor College of Medicine, Houston, TX). The mixed oligodeoxynucleotide hybridization probe was radiolabeled using T4 polynucleotide 5′-hydroxykinase and [γ-^{32}P]ATP. The cDNA library was propagated in *E. coli* strain Y1090 and $\approx 5.5 \times 10^5$ phages were screened by the plaque-hybridization procedure (19) using the radiolabeled probe. Hybridizations were performed at 42°C for 24 hr in a solution containing 0.1 M Tris·HCl (pH 8.0), 1.0 M NaCl, 5 mM EDTA, 0.2%

Abbreviations: C3, C4, C5, C6, C7, C8α, C8β, and C9, third, fourth, fifth, sixth, seventh, α-chain of the eighth, β-chain of the eighth, and ninth component of complement, respectively; MAC, membrane attack complex of complement; LDL, low density lipoprotein.

‡The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04506).

Ficoll, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin, 0.1% SDS, and 1×10^6 cpm per ml of the probe. After hybridization, the filters were washed twice, in each case for 15 min, in the following order: 2× SSC/0.1% SDS at 23°C; 1× SSC/0.1% SDS at 37°C (1× SSC buffer = 15 mM sodium citrate/150 mM NaCl, pH 7.0) and subsequently autoradiographed. Positive clones were plaque-purified by successive rounds of screening at progressively lower plaque densities.

DNA sequence analysis. DNA was purified from recombinant phages as described (20). cDNA inserts contained in the *Eco*RI site of the vector λgt11 were released by digestion with *Eco*RI. These or their restriction fragments were subcloned into suitable sites in the polylinker of bacteriophage M13mp18 vector or a phagemid vector pBS M13+, both of which were propagated in *E. coli* XL1-Blue. DNA sequencing was carried out by the dideoxynucleotide chain-termination method (21) using modified T7 DNA polymerase (22) as described in the United States Biochemical sequencing manual on a Sequenase DNA sequencing kit. In the case of double-stranded sequencing using pBS M13+, the DNA was initially denatured by alkali (23). Sequencing reactions were carried out with dATP [α-^{35}S] and the reaction products were subjected to electrophoresis in polyacrylamide gels [6% (wt/vol) acrylamide] with multiple loadings. The largest cDNA insert was also sequenced by shot-gun cloning in M13mp8 (24).

Computer analyses. An overall consensus nucleotide sequence was obtained from the DNA sequence data using the computer programs DBAUTO and DBUTIL (25, 26). Search of the National Biomedical Research Foundation (NBRF) protein sequence data base (December 1986 release) for protein sequence homology was carried out by using the program FASTP (27). The program GAP for optimal alignment of protein sequences and the programs PEPPLOT, PEPTIDESTRUCTURE, and PLOTSTRUCTURE for secondary structure prediction and average hydrophobicity calculations were contained in the sequence analysis software package (version 5, June 1987) of the Genetics Computer Group (University of Wisconsin Biotechnology Center, Madison, WI) (28).

RESULTS

Partial Amino Acid Sequence Determination of C6. Trypsin was used to generate tryptic peptides from reduced and ^{3}H

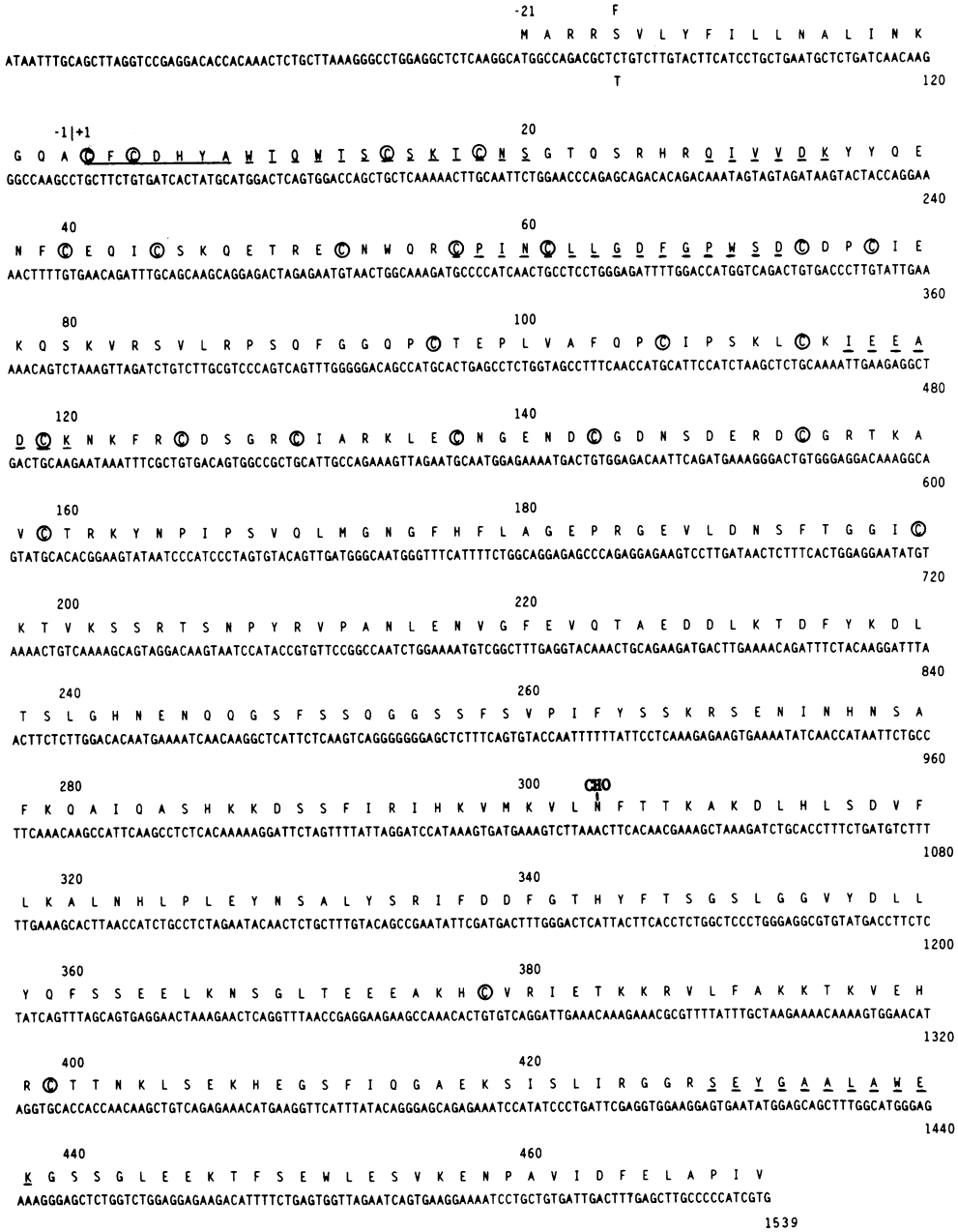

FIG. 1. Nucleotide sequence of λHL C6/11.1 and λHL C6/2.1. The derived amino acid sequence (single-letter code) represents approximately the amino-terminal half of human C6. The amino acid sequence used for construction of the oligonucleotide probe is underlined continuously. Amino acids identified at both the DNA and the protein levels are individually underlined. CHO denotes possible site for N-glycosylation. Cysteine residues are circled. The negatively numbered amino acid sequence represents the probable hydrophobic leader peptide. Each strand of the insert DNA was sequenced in its entirety and each base was sequenced an average of 10.9 times. Polymorphism was detected at position 80 in the nucleotide sequence where λHL C6/2.1 contained C and λHL C6/11.1 contained T, giving rise to a corresponding amino acid sequence polymorphism of either serine or phenylalanine, respectively, in the putative signal peptide.

carboxymethylated C6 and the fragments were separated by HPLC. The amino-terminal sequences of certain peptides were obtained by automated Edman degradation. To utilize the amino acid sequence data for the synthesis of oligonucleotide hybridization probes, we looked for amino acid sequences that are encoded at the nucleotide level by sequences of low redundancy. Inspection of the protein sequence data obtained here and the sequence of 20 amino-terminal residues of human C6 recently reported by us (13) revealed that the most suitable sequence was the amino-terminal seven residues, Cys-Phe-Cys-Asp-His-Tyr-Ala. The oligodeoxynucleotide sequences complementary to this amino acid sequence were synthesized and contained the sequences 5′-GC(A/G) TA(A/G) TG(A/G) TC(A/G) CA(A/G) AA(A/G) CA-3′.

Plaque Hybridization and DNA Sequence Analysis. Approximately 5.5×10^5 phage plaques from the human liver cDNA library constructed in the bacteriophage λgt11 vector were screened with the above oligonucleotide probe. Thirty-five positive clones were identified in the initial screening, 16 of which were plaque-purified. The phage DNAs were prepared from 4 of these clones having insert sizes in the approximate range of 0.6 to 2.0 kilobases (kb) and were subjected to DNA sequence analysis. The nucleotide-derived amino acid sequence of 2 of them—λHL C6/11.1 and λHL C6/2.1, containing 0.6-kb and 1.5-kb inserts, respectively—were found to contain the amino-terminal sequence of human C6. Thus, the cDNA inserts in λHL C6/11.1 and λHL C6/2.1 represented the mRNA for the human complement protein C6. The 2 other clones did not encode C6. Since the cDNA inserts contained in the other 12 positive clones were <1.5 kb, they were not analyzed further.

Initial double-stranded DNA sequence analysis of the cDNA inserts released from the phage vector λgt11 by *Eco*RI digestion followed by subcloning into pBS M13+ showed that the 5′ end of the inserts in λHL C6/11.1 and λHL C6/2.1 represented identical sequences. The portion of the λHL C6/2.1 insert extending beyond the 3′ end of λHL C6/11.1 insert contained unique internal restriction sites for cleavage with *Pst* I or *Bam*HI or *Xba* I. These sites were absent in λHL C6/11.1. Two subclones of the cDNA insert of λHL C6/2.1 in pBS M13+, representing opposite orientations, were digested separately with each of these restriction enzymes and the fragment released from the vector in each case was isolated and subcloned into M13mp18 linearized with the corresponding enzyme. From a combination of nucleotide sequences of the enzymatically obtained fragments and random fragments obtained by shotgun-cloning of the insert in λgt11, the cDNA sequence was deduced as shown in Fig. 1. This contained the coding sequence for the amino-terminal half of the C6 protein as well as additional sequences at the 5′ end possibly encoding a signal peptide. The 470 residues of the cDNA-derived amino acid sequence of C6 represents ≈50% of C6.

Structural Analysis. C6 shows extensive amino acid sequence homology with C7, C8α, C8β, and C9 as shown in Fig. 2. The optimal alignments between C6 and each of C7, C8α, C8β, and C9 were obtained by using the program GAP and showed, respectively, 52.4%, 51.5%, 47.5%, and 47.1% similarity when conserved substitutions are allowed. In addition to the proteins of the complement system, the following proteins showed considerable similarity with C6: low density

```
   C7     .......... .......... .......... .......... .......... .....SSPVN COWDFYAPWS ECNGCTKTQT RRRSV.AVYG QYGGQPCVGN 44
   C8α    .......... .......... .......... .......... .......... ..AATPAAVT COLSNWSEWT DCFPCQDKKY RHRSLLOPNK FGGTICSGDI 48
   C8β    .......... .......... .......... .......... .......... SVDVTLMPID CELSSWSSWT TCDPCQKKRY RYAYLLQP.S QFHGEPCNFS 49
   C9     .......... .......... .......... .........Q YTTSYDPELT ESSGSASHID CRMSPWSEWS QCDPCLROMF RSRSI.EVFG QFNGKRCTDA 60
   C6     CFCDHYAWTQ WTSCSKTCNS GTQSRHRQIV VDKYYQENFC EQICSKQETR ECNWQRCPIN CLLGDFGPWS DCDPCIEKQS KVRSVLRP.S QFGGQPCTEP 99
Consensus ---------- ---------- ---------- ---------- ---------- -------PI- C-LS-WS-WS DCDPC--K-Y R-RSVL-P-- QFGG-PC-D-

   C7     AFETQSCEPT RGCPTE...E GCGERFRCFS G..QCISKSL VC...NGDSD CDEDSADEDR CEDSERRPSC DIDKPPPNIE LTGNGYNELT GQFRNRVINT 136
   C8α    WDQASCSSST TCVRQAQCGQ DFQCK..... ETGRCLKRHL VC...NGDQD CLDGSDEDDC EDVRAIDEDC SQYEPIPGSQ KAALGYNILT QEDAQSVYDA 140
   C8β    DKEVEDCVTN RPCRSQVRCE GFVC.....A QTGRCVNRRL LC...NGDND CGDQSDEA.. NCRRIYKKCQ HEMDQYWGIG SLASGIMLFT NSFEGPVLDH 139
   C9     VGDRRQCVPT EPCE....DA EDDCGNDFQC STGRCIKMRL RC...NGDND CGDFSDEDDC ESEPRPPCRD RVVEESELAR TAGYGINILG MDPLSTPFDN 153
   C6     LVAFQPCIPS KLCKIE...E A.DCKNKFRC DSGRCIARKL EC...NGEND CGDNSDE..R DCGRTKAVCT RKYNPIPSVQ LMGNGFHFLA GEPRGEVLDN 190
Consensus --E---CVPT --C------E ---C------ -TGRCI-R-L -C---NGDND CGD-SDEDD- E--R------ ---EP-P-I- --G-GYN-LT -E----VLD-

   C7     KSFGGQCRKV .......... .........F SGDGKDFYRL SGNVLS.... YTFQVKINND FNYEFYNSTW SYVKHTSTEH TSSSRKRSFF .........R 204
   C8α    SYYGGQCETV ..YNGEWREL RYDSTCERLY YGDDEKYFRK PYNFLKYH.F EALADTGISS EFYDNANDLL SKVKKDKSDS FGVTIGIGPA GSPLLVGVGV 237
   C8β    RYYAGGCSP. .......... ......HYIL NTRFRKPYNV ESYTPQTQGK YEFILKE... .....YESYS DFERNVTEKM ASKSGFSFGF KIPGIFELGI 214
   C9     EFYNGLCNRD RDGNTLTYYR RPWNVASLIY ETKGEKNFRT EHYEEQIEAF KSIIQEKTSN FNAAISLKFT PTETNKAEQC CEETASSISL HGKGSFRFSY 253
   C6     SFTGGICKTV K..S...... .......... .SRTSNPYRV PANLENV.GF .EV.QTAEDD LKTDFYKDLT SLGHNENQQG SFSSQGGSSF SVPIFYSSKR 268
Consensus --YGG-C--V ---------- ---------Y -----K-YR- --N------F --F------- F--D-Y--L- S---N----- ---S-----F --P--F----

   C7     SSSSSSRSYT SHTNEIHKGK SYQLLVVENT VEVAQFINNN PEF..LQLAE PFWKELSHLP SLYDYSA.YR RLIDQYGTHY LQSGSLGGEY RVLFYVDSEK 301
   C8α    SH..SQDTSF LNELNKYNEK KFIFTRIFTK VQTAHFKMRK DDIM...LDE GMLQSLMELP DQYNYG.MYA KFINDYGTHY ITSGSMGGIY EYILVIDKAK 331
   C8β    SSQSDRGKHY IRRTKRFSHT KSVFLHARSD LEVAHYKLKP RSLM...LHY EFLQRVKRLP LEY.SYGEYR DLFRDFGTHY ITEAVLGGIY EYTLVMNKEA 310
   C9     SK....NETY QLFLSYSSKK EKMFLHVKGE IHLGRFVMRN RDV....LTT TFVDDIKALP TTYEKGEYFA .FLETYGTHY SSSGSLGGLY ELIYVLDKAS 344
   C6     SENINHNSAF KQAIQASHKK DSSFIRIHKV MKVLNFTTKA KDL...HLSD VFLKALNHLP LEYNSA.LYS RIFDDFGTHY FTSGSLGGVY DLLYQFSSEE 364
Consensus S--------- ---------K ---FL----- V-VA-F-M-- -DL----L-E -FL--L--LP --Y-----Y- --FDDYGTHY -TSGSLGGIY E---V-DKE-

   C7     LKQNDFNSVE EKKCKSSGWH FVVK.FSSHG CKELENALKA ASGTQNNVLR GEPFIRGGGA GFISG....L SYLELDNPAG NKRRYSAWAE SVTNLPQVIK 396
   C8α    MESLGITSRD ITTCFGGSLG IQYEDKINVG GGLSGDHCKK .......... .FGGGKTERA RKAMAVEDII SRVRG..... GSSGWSGGLA QNRSTI.... 411
   C8β    MERGDYTLNN VHACAKNDFK IGGAIEEVYV SLGVSVGKCR GILNEIKDRN KRDTMVEDLV VLVRG..... .......... GASEHITTLA YQELPTA... 392
   C9     MKRKGVELKD IKRCLGYHLD VSLAFSEISV GAEFNKDDCV KRGEGRAVNI TSENLIDDVV SLIRGGTRKY AFELKEKLLR GTVIDVTDFV .NWASSINDA 443
   C6     LKNSGLTEEE AKHC..VRIE TKKRVLFAK. KTKVEHRCTT NKLSEK.... .......HEG SFIQGAEKSI SLIRG..... GRSEYGAALA WEKGSSGLEE 445
Consensus MK--G-T--- IK-C---L- I--------- ---------- ---------- -------D-- --I-G----- S--------- G-S-Y---LA ----------

   C7     QKLTPLYELV KEVPCASVKK LYLKWALE 424
   C8α    .TYRSWGRSL KYNPVVIDFE MQPIHECC 438
   C8β    DLMQEWGDAV QYNPAIIKVK VEPLYELV 420
   C9     PVLISQKLSP IYNLVPVKMK NAHLKKQN 471
   C6     KTFSEWLESV KENPAVIDFE LAPI...V 470
Consensus --L--W-ESV KYNP-VI--K L-P-----
```

FIG. 2. Amino acid sequence (single-letter code) homology among C6, C7, C8α, C8β, and C9. The sequences were optimally aligned by inserting gaps (periods) to maximize the number of matches using the programs GAP and PRETTY in the sequence analysis software package of the University of Wisconsin Genetics Computer Group (28). GAP implements the alignment method of Needleman and Wunsch (29) using a symbol comparison table in which identical residues score 1.5 and nonidentical residues receive a score derived from the evolutionary distance between the amino acids (mutational difference matrix) as measured by Schwartz and Dayhoff (30) and normalized by Gribskov and Burgess (31). Scores for nonidentical comparisons range from 1.491 for Phe-Tyr to −0.677 for Ala-Trp. PRETTY was used to display multiple sequence alignments and to calculate any possible consensus sequence representing either identity or similarity (i.e., allowing for conserved substitutions) of the residue at that particular position in at least three of five sequences. To be considered similar, the threshold for symbol comparison was set at 1.00.

lipoprotein (LDL) receptor, 38.3%; thrombospondin, 35.4%; thrombomodulin, 30.9%. These scores were based on a gap weight of 5.0, a gap length weight of 0.3, and a symbol comparison match threshold of ≥0.5. A site of N-glycosylation of C6 contained in the sequence Asn-Phe-Thr was located at position 303. Further interpretations of the sequence data using computer algorithm for hydropathy and secondary structure are summarized in Fig. 3. The amino-terminal half of C6 is largely hydrophilic but contains several segments with a strong hydrophobic tendency, especially in the regions of the following residues: 55–62, 94–107, 163–175, 187–195, 254–259, 290–298, 310–318, 344–353, and 458–462. These sequences represent potential membrane surface-seeking segments. β-Sheets and β-turns are predicted to be predominant structures in the amino-terminal half of the sequence of C6 up to residue 218. α-Helices are more predominant in the other half of the reported sequence—i.e., from residue 218 to 470.

DISCUSSION

We have recently identified five different types of cysteine-rich segments containing 35–77 amino acids in the primary structure of complement protein C7 (3). These have been named types I–V based on the order of their appearance in the structure of C7. Of these, types I, II, and III exhibit amino acid sequence homology with the complement proteins C8α, C8β, and C9, whereas types IV and V have been found so far only in C7 among the MAC proteins. The type I segment occurs three and a half times in the platelet adhesive protein thrombospondin; two times each in C7, C8α, and C8β; and once in C9. The type II segment is repeated eight times in the LDL receptor and once each in C7, C8α, C8β, and C9. We have now identified the presence of at least two type I and one type II cysteine-rich homology units in their expected positions toward the amino-terminal end of C6 (Fig. 4). The "modular fusion hypothesis" (3) suggests that such homologous cysteine-rich segments serve as intermolecular fasteners in the assembly of the MAC—i.e., the modules of each subunit bind to the corresponding ones of the other subunits in the complex. The type I unit in thrombospondin is known to constitute a domain that binds fibrinogen, plasminogen, laminin, fibronectin, and collagen (34). Similarly, the collective role of the type II units in the LDL receptor is to bind apoproteins B and E (35). One of the predictions of the hypothesis was the expected presence of these homologous cysteine-rich modules in the primary structure of C6. The present findings on the modular arrangement of the structure of C6 have provided further support in favor of this concept. Secondary structure prediction studies suggest that the cysteine-rich homology segments are rich in β-sheet, and particularly in β-turns. It has been pointed out (3) that β-turns would give rise to stably folded structures present on the surface of proteins that usually participate in recognition and protein–protein interaction (36, 37). Nothing is yet known about the exact function of the cysteine-rich segments in the MAC. The specific arrangement of cysteine residues in these segments may have a similar function in all of the MAC

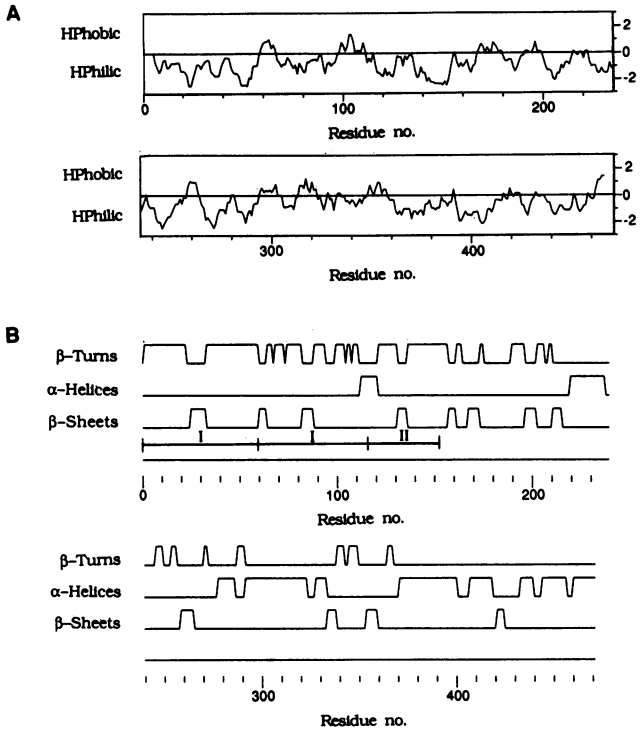

FIG. 3. (*A*) Hydrophobicity profile of residues 1–470 of C6 protein as a function of sequence number calculated according to Kyte and Doolittle (32) using the program PEPPLOT in the sequence analysis software package of the University of Wisconsin Genetics Computer Group (28). The curve is the average of a residue-specific hydrophobicity index over a window of nine residues. When the line is in the upper half of the frame, it indicates a hydrophobic region (HPhobic), and when it is in the lower half, it indicates a hydrophilic region (HPhilic). (*B*) Schematic representation of β-turns, α-helices, and β-sheets in the secondary structure of C6 predicted as a function of sequence number by the method of Garnier *et al.* (33) using the programs PEPTIDESTRUCTURE and PLOTSTRUCTURE in the software package described above. I and II represent the positions of the two type I and one type II cysteine-rich homology units. β-Turns and β-sheets constitute the predominant predicted secondary structure in these regions.

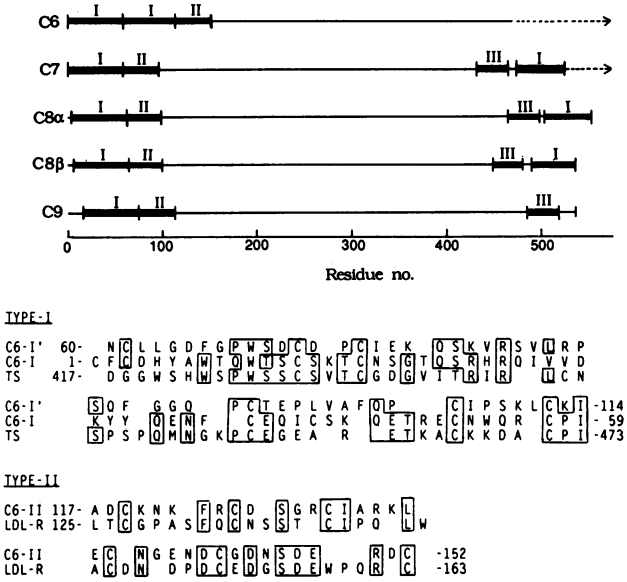

FIG. 4. (*Upper*) Distribution of types I, II, and III cysteine-rich homology units in the primary structure of the complement proteins C6, C7, C8α, C8β, and C9. Broken lines in C6 and C7 denote representation of amino acid sequences not determined and not shown, respectively. Note that the type I unit is represented twice in the amino-terminal end of C6 in contrast to C7, C8α, C8β, and C9, where it occurs only once. (*Lower*) Amino acid sequence (single-letter code) similarity of cysteine-rich homology units of types I and II of C6 with those of thrombospondin (TS) and LDL receptor (LDL-R), respectively. Comparisons with similar units of C7, C8α, C8β, and C9 have not been included here since they appear in Fig. 2. Type I segment occurs three and a half times in thrombospondin and the type II segment is repeated eight times in LDL receptor. However, for brevity, only one of the units has been shown. Type I' unit denotes the second type I unit in C6. Identical residues are boxed and gaps have been introduced to display maximum sequence homology.

proteins and may serve as initiation sites for possible disulfide exchange or intermolecular disulfide bond formation during the assembly of the MAC. The presence of these similar segments in the MAC proteins implies a common ancestry of these proteins and suggests that they are products of mosaic genes (38). The type I cysteine-rich homology unit present in thrombospondin and the MAC proteins was recently pointed out to be present six times in properdin of the alternative complement pathway and once in the circumsporozoite protein of malaria parasites as well as in a protein derived from the asexual blood stages of this parasite (34, 39, 40). This conserved sequence motif has been suggested to play a key role in mechanisms by which malaria parasites avoid host defenses mediated by complement (39). The presence of another cysteine-rich stretch of ≈60 amino acids, found in several complement proteins that share the property of interaction with C3b or C4b (for review, see ref. 41), was pointed out recently for C7 (42). The general consensus sequence of this cysteine-rich motif (41) is indeed present twice in the carboxyl-terminal end of C7 between residues 546 and 606, and 607 and 668. These are the approximate locations of the types IV and IV′ cysteine-rich units described in our previous report (3). This motif is expected also to be present in C6, which binds to C5b, an evolutionary relative of C3b and C4b.

Both thrombospondin and thrombomodulin form complexes with thrombin (43, 44). Interestingly, C6, which shares strong structural similarities with thrombospondin and thrombomodulin, also binds thrombin (13).

The presence of structural homology among C6, C7, C8, and C9 of complement and perforin of T killer and natural killer cells was suggested in recent studies based on antibody cross-reactivity (2). The recently reported sequences of mouse (9) and human (10) perforin have demonstrated distinct homology with the MAC proteins. Approximately 370 of the 534 amino acid residues of human perforin (10) encompassing the putative membrane binding region and the cysteine-rich epidermal growth factor type domain are homologous to the MAC proteins. Thrombospondin and LDL receptor-type cysteine-rich domains are not present in perforin. It is possible that these domains are not essential for transmembrane channel formation or for self-polymerization but are required for the heteropolymeric assembly of the MAC. The homology of the MAC proteins with perforin confirms predictions on the common genetic origin of the two effector systems (45). From the shared homology among this family of mosaic proteins involved in humoral and cellular cytolysis, it is apparent that the properties of membrane attachment, membrane insertion, and polymerization in these proteins have evolved by acquiring different protein domains to give rise to specific physicochemical properties. Based on sequence homology, the present investigation adds C6 as an important member of this family of channel-forming cytolytic proteins.

The authors wish to thank Dr. Savio L. C. Woo (Baylor College of Medicine) for the generous gift of the human liver cDNA library. The authors are grateful to Dr. Dan Noonan (Immunetech, Inc.) for various helpful suggestions during the course of this work and to Dr. Jasper Rees (Salk Institute) for help with the Staden programs. C.A.P. held a Fogarthy International Fellowship and would like to thank Dr. Manuel E. Patarroyo (Instituto de Immunologia, Hospital San Juan de Dios, Universidad Nacional de Colombia, Bogota) for encouragement. This is publication no. 5427-IMM from the Department of Immunology, Research Institute of Scripps Clinic. This work was supported by U.S. Public Health Service Grant AI 17354. D.N.C. and B.C. are equal first authors.

1. Muller-Eberhard, H. J. (1988) *Annu. Rev. Biochem.* **57**, 321–347.
2. Tschopp, J., Masson, D. & Stanley, K. K. (1986) *Nature (London)* **322**, 831–834.
3. DiScipio, R. G., Chakravarti, D. N., Muller-Eberhard, H. J. & Fey, G. H. (1988) *J. Biol. Chem.* **263**, 549–560.
4. Rao, A. G., Howard, O. M. Z., Ng, S. C., Whitehead, A. S., Colten, H. R. & Sodetz, J. M. (1987) *Biochemistry* **26**, 3556–3564.
5. Howard, O. M. Z., Rao, A. G. & Sodetz, J. M. (1987) *Biochemistry* **26**, 3565–3570.
6. Haefliger, J.-A., Tschopp, J., Nardelli, D., Wahli, W., Kocher, H.-P., Tosi, M. & Stanley, K. K. (1987) *Biochemistry* **26**, 3551–3556.
7. Stanley, K. K., Kocher, H.-P., Luzio, J. P., Jackson, P. & Tschopp, J. (1985) *EMBO J.* **4**, 375–382.
8. DiScipio, R. G., Gehring, M. R., Podack, E. R., Kan, C. C., Hugli, T. E. & Fey, G. H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7298–7302.
9. Shinkai, Y., Takio, K. & Okumura, K. (1988) *Nature (London)* **334**, 525–527.
10. Lichtenheld, M. G., Olsen, K. J., Lu, P., Lowery, D. M., Hameed, A., Hengartner, H. & Podack, E. R. (1988) *Nature (London)* **335**, 448–451.
11. Wetsel, R. A., Lemons, R. S., Le Beau, M. M., Barnum, S. R., Noack, D. & Tack, B. F. (1988) *Biochemistry* **27**, 1474–1482.
12. Kolb, W. P., Kolb, L. M. & Savary, J. R. (1982) *Biochemistry* **21**, 294–301.
13. Chakravarti, D. N. & Muller-Eberhard, H. J. (1988) *J. Biol. Chem.* **263**, 18306–18312.
14. DiScipio, R. G. & Gagnon, J. (1982) *Mol. Immunol.* **19**, 1425–1431.
15. Arroyave, C. M. & Muller-Eberhard, H. J. (1971) *Immunochemistry* **8**, 995–1006.
16. Podack, E. R., Kolb, W. P. & Muller-Eberhard, H. J. (1976) *J. Immunol.* **116**, 263–269.
17. Podack, E. R., Kolb, W. P., Esser, A. F. & Muller-Eberhard, H. J. (1979) *J. Immunol.* **123**, 1071–1077.
18. Edman, P. & Begg, P. (1967) *Eur. J. Biochem.* **1**, 80–91.
19. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
20. Helms, C., Graham, M. Y., Dutchick, J. E. & Olson, M. V. (1985) *DNA* **4**, 39–49.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
22. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
23. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4**, 165–170.
24. Bankier, A. T. & Barrell, B. G. (1983) in *Techniques in Nucleic Acid Biochemistry*, ed. Flavell, R. A. (Elsevier Ireland, Limerick), Vol. B5, pp. 1–73.
25. Staden, R. (1982) *Nucleic Acids Res.* **10**, 4731–4751.
26. Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961.
27. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
28. Devereux, J. R., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
29. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
30. Schwartz, R. M. & Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), pp. 353–358.
31. Gribskov, M. & Burgess, R. R. (1986) *Nucleic Acids Res.* **14**, 6745–6763.
32. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
33. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
34. Lawler, J. & Hynes, R. O. (1986) *J. Cell Biol.* **103**, 1635–1648.
35. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russell, D. W. (1984) *Cell* **39**, 27–38.
36. Rossmann, M. G. & Argos, P. (1981) *Annu. Rev. Biochem.* **50**, 497–532.
37. Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1–109.
38. Gilbert, W. (1985) *Science* **228**, 823–824.
39. Goundis, D. & Reid, K. B. M. (1988) *Nature (London)* **335**, 82–85.
40. Robson, K. J. H., Hall, J. R. S., Jennings, M. W., Harris, T. J. R., Marsh, K., Newbold, C. I., Tate, V. E. & Weatherall, D. J. (1988) *Nature (London)* **335**, 79–82.
41. Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T. & Tack, B. F. (1986) *Immunol. Today* **7**, 230–234.
42. Stanley, K. & Luzio, P. (1988) *Nature (London)* **334**, 475–476.
43. Danishefsky, K. J., Alexander, R. J. & Detwiler, T. C. (1984) *Biochemistry* **23**, 4984–4990.
44. Esmon, N. L., Owen, W. G. & Esmon, C. T. (1982) *J. Biol. Chem.* **257**, 859–864.
45. Lachmann, P. J. (1983) *Nature (London)* **305**, 473–474.