The EMBO Journal vol.8 no.3 pp.711-718, 1989

# Molecular cloning and characterization of the novel, human complement-associated protein, SP-40,40: a link between the complement and reproductive systems

**Louis Kirszbaum, Jennifer A.Sharpe, Brendan Murphy[1], Anthony J.F.d'Apice[1], Brendan Classon[2], Peter Hudson[3] and Ian D.Walker**

Department of Veterinary Preclinical Sciences, the University of Melbourne, [1]St Vincent's Hospital, Melbourne and [3]CSIRO Division of Biotechnology, Parkville, Victoria, Australia

[2]Present address: MRC Cellular Immunology Unit, South Parks Road, Oxford, UK

Communicated by A.Williams

**The cDNA sequence encoding the human complement-associated protein, SP-40,40, is reported. The two chains of SP-40,40 are coded in a single open reading frame on the same mRNA molecule, indicating the existence of a biosynthetic precursor protein which matures post-synthetically by the proteolysis of at least one peptide bond. The precursor is preceded by a signal sequence for vectorial export and contains six N-linked glycosylation sites distributed equally between the two chains of the structure. The sequence of the SP-40,40 precursor bears a 77% identity to a rat sulphated glycoprotein-2 (SGP-2) which is the major secreted product of Sertoli cells. The presence of SP-40,40 within human seminal plasma at levels comparable to those in serum was demonstrated, indicating that SP-40,40 and SGP-2 are serum and seminal forms of the same protein. A sequence of 23 amino acids within the β-chain of SP-40,40 exhibited significant homology to corresponding segments located within complement components C7, C8 and C9. The short cysteine-containing motif represented the only evidence of a possible vestigial relationship between SP-40,40 and other complement components. The precise role of SP-40,40 is not known in either blood or semen but the present findings document an intriguing link between the immune and the reproductive systems.**

*Key words:* complement/cDNA cloning/sequence homology/Sertoli cell/reproduction

## Introduction

The human complement system consists of ~20 serum proteins, the collective role of which is to eliminate cells identified by antibodies as foreign. Several steps in the pathways of complement activation are catalysed by active proteases and these steps serve to amplify the response to a specific immune encounter. An elaborate series of checks and balances has evolved to control several steps of the amplification cascade (Müller-Eberhard, 1986; Reid, 1986).

We recently described a novel human serum protein (SP-40,40) using a series of monoclonal antibodies directed to the immune deposit-containing glomerular basement membranes of a patient with membranous glomerulonephritis (Murphy *et al.*, 1988). This protein was shown to be a normal constituent of human blood. It consists of two chains, α and β, covalently joined by disulphide bonds; neither was related in amino-terminal sequence to any known protein. The histological distribution of SP-40,40 within nephritis kidneys closely resembled that of certain components of the C5b-9 complex of complement (also termed the membrane attack complex, MAC). We formally established that SP-40,40 was a member of the human complement system by directly demonstrating its presence within the SC5b-9 complex, the S-protein-containing soluble variant of the C5b-9 complex (Podack *et al.*, 1978). This previous study suggested a role for SP-40,40 within the complement pathway and prompted the present investigation on the structure of the SP-40,40 protein and its mRNA. The aim of the present study was to investigate the structural homology of SP-40,40 to other members of the complement pathway or to other known structures. The precedent for this approach was the recent demonstration that S-protein, a constituent of the SC5b-9 complex and a regulator of the

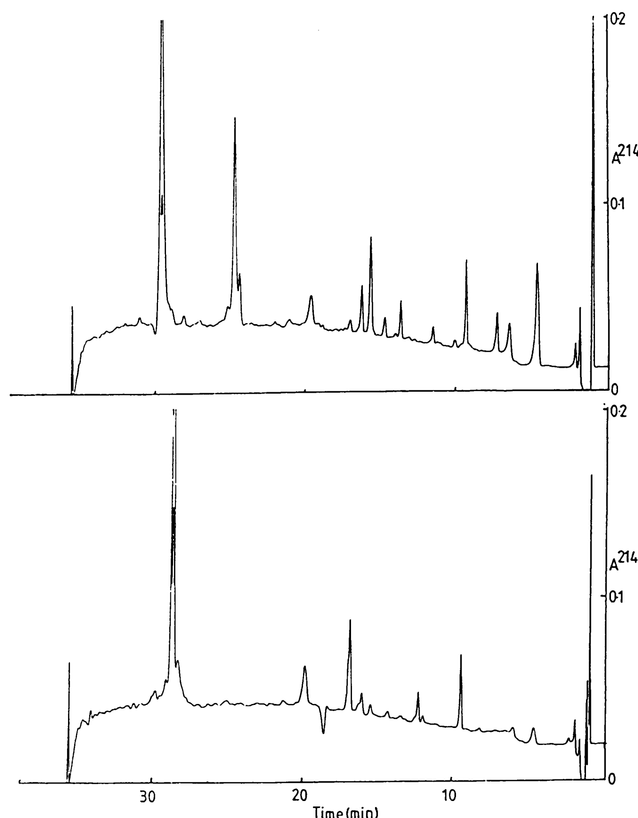


**Fig. 1.** Peptide mapping of α- and β-chains of SP-40,40. Samples of carboxamidomethylated, purified α- and β-chains were digested and subjected to chromatography using a C18 reversed-phase HPLC column and acetonitrile gradient elution. **Upper panel**: α-chain digest. **Lower panel:** β-chain digest.

©IRL Press

**Table I.** Partial amino acid sequence of SP-40,40

| Peptide | Sequence | Location (Figure 3) |
|---|---|---|
| $NH_2$-α | SLMPFSPYEPLNFHAMFQPFLEMI[a]HEA | 206–231 (α) |
| $NH_2$-β | DQTVSDNELQEMSNQG[a]SKYVNKE[b]IQN | 1– 26 (β) |
| C.I.2 | IHEAQQA | 229–235 (α) |
| C.4 | NGDRIDSLLENDRQQT(H) | 132–149 (β) |
| C.5 | FQPFLE | 222–227 (α) |
| C.6 | ETVAEKALQEYRKKHREE | 410–427 (α) |
| C.10.1 | DIHFHSPAFQHPPTEFIRE | 237–255 (α) |
| C.17.1 | SNQGSKYVNKEIQNAVN | 13– 29 (β) |
| C.17.2 | QDHFSRASSIIDELFQDRFFTREPQDTT | 155–182 (β) |
| C.23 | L–TSSLLEQLNEQFN | 331–345 (α) |
| L.4 | SYQWK | 325–329 (α) |
| L.5 | YNELLK | 319–324 (α) |
| L.6 | FMETVAEK | 408–415 (α) |
| L.7 | CREILSVDCSTNNPSQAK | 283–300 (α) |
| L.9 | LFDSDPITVTVPVEVSRK | 387–404 (α) |
| L.12 | LRRELDESLQVAERLTRK | 301–318 (α) |
| L.16 | ML–TSSLLEQLNEQFNWVSRLA–LTQGEDQYYLR | 330–363 (α) |
| L.26 | DQTVSDNELQEMSNQGSK | 1– 18 (β) |
| L.28 | TLLSNLEEAK | 47– 56 (β) |
| V8.16 | IHFHSPAFQHP | 238–248 (α) |
| V8.27 | LLKSYQWKML–T | 322–333 (α) |
| V8.31 | SLMPFSPYEPLNFHA | 206–220 (α) |
| V8.38 | HFSRASSIIDE | 157–167 (β) |
| V8.41 | NDRQQTHMLDVMQ | 143–155 (β) |
| V8.44 | VVVKLFD | 383–389 (α) |
| V8.45 | TMMALWE(E) | 83– 90 (β) |
| V8.53 | FL–QSSPFYFWMNGDRIDSLLE | 121–142 (β) |
| V8.58 | AQQAMDIHFHSPAFQHPPTEFIRE | 232–255 (α) |
| V8.65 | DQYYLRVTTVASHTSD | 358–373 (α) |
| V8.73 | SDPITVTVPVE | 390–400 (α) |
| V8.75 | ILSVDCSTNNPSQAKLRRE | 286–304 (α) |
| V8.77 | IRH–STGCLRMKDQCDKCRE | 266–285 (α) |
| V8.79 | VSRKNPKFME | 401–410 (α) |
| V8.86.1 | YRKKHRE | 420–426 (α) |
| V8.86.2 | KALQE | 415–419 (α) |
| V8.87 | GDDRTVCRE | 256–265 (α) |
| V8.88 | MIHE | 228–231 (α) |
| V8.91 | RLTRKYNE | 314–321 (α) |
| V8.92 | SLQVAE | 308–313 (α) |

– indicates cycles where no PTH-amino acid derivative was assigned (glycosylated asparagine cannot be sequenced as a PTH derivative).
[a]Residues positively assigned by peptide sequencing since publication of original sequence (Murphy *et al.*, 1988).
[b]Residue originally published as Ile but corrected by peptide sequencing.
( ) residue not securely assigned.

coagulation pathway, was identical to vitronectin, a cell adhesion protein (Jenne and Stanley, 1985). Within the complement pathway S-protein binds to nascent C5b-7 and thereby prevents the insertion of the resultant complex into plasma membranes (Podack *et al.*, 1977; Bhakdi and Roth, 1981). This property may serve to protect bystander cells at sites of complement-mediated immune attack. The subsequent addition of complement components C8 and C9 to the SC5b-7 complex is not inhibited, but the resultant SC5b-9 complex is water soluble and non-lytic. The selective occurrence of SP-40,40 within the SC5b-9 complex, along with S-protein suggested a function and possibly a structural link between them (Murphy *et al.*, 1988). Comparison herein of the two sequences did not support a close structural relationship, but detailed data base searches for a homologue of SP-40,40 revealed that a rat sulphated glycoprotein (SGP-2), a recently described product of Sertoli cells (Collard and Griswold, 1987), was the rat counterpart of SP-40,40. The SGP-2 protein is the major secreted product of Sertoli cells and has been suggested to play a critical role in spermatogenesis (Griswold *et al.*, 1986). Sertoli cells are thought to provide the cellular and biochemical support system for sperm development (Yanagimachi, 1988) and indeed SGP-2 is bound to the acrosomal and distal tail portions of mature spermatozoa (Sylvester *et al.*, 1984). The finding that rat SGP-2 and human SP-40,40 are species homologues suggests that SP-40,40 may have at least two roles: one within the reproductive system and the other within the complement system.

## Results

### *The SP-40,40 α- and β-chains are unrelated*

Purification of the SP-40,40 α- and β-chains was achieved by reversed-phase HPLC and samples of each were digested with endoprotease lysine-C and the digests subjected to

**Table II.** SP-40,40 oligonucleotide probes

**SP40,40 $\alpha$-**

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Y | E | P | L | N | F | H | A | M | F | Q | P | F | |
| 5′ | TAT | GAG | CCC | CTG | AAC | TTC | CAT | GCC | ATG | TTC | CAG | CCC | TT | 3′ |
| 3′ | ATA | CTC | GGG | GAC | TTG | AAG | GTA | CGG | TAC | AAG | GTC | GGG | AA | 5′ |

**SP-40,40 $\beta$-**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | F | Y | F | W | M | N | G | D | R | I | D | |
| 5′ | C<br>TT<br>T | TAC | TTC | TGG | ATG | AAC | G<br>GGA<br>T | GAC | G<br>AG<br>A | ATT | GA | 3′ |
| 3′ | G<br>AA<br>A | ATG | AAG | ACC | TAC | TTG | C<br>CCT<br>A | CTG | C<br>TC<br>T | TAA | CT | 5′ |

The single-letter code is used for amino acid sequences (upper lines) and the sequence strands corresponding to the mRNA (middle lines) and the oligonucleotide sequences used as probes (lower lines) are formatted beneath corresponding amino acid residues.

peptide mapping (Figure 1). The two elution profiles contained no common peaks, suggesting that the chains of SP-40,40 were not closely related. Further proteolytic digestion experiments were conducted and individual peptide peaks were sequenced. A summary of the data obtained is presented in Table I. The peptide sequences obtained confirmed the absence of homology between the two chains of SP-40,40.

In order to dictate the synthesis of oligonucleotides for library screening, two runs of amino acid sequence (one for each chain) with minimum codon ambiguity were selected (Table II). Dinucleotide combinations which generated potential methylation sites were eliminated and codon frequency tables were used to minimize the complexity of the final oligonucleotide mixtures (Lathe, 1985).

### *The SP-40,40 $\alpha$- and $\beta$-chains are part of a common precursor molecule*

The $\alpha$-specific oligonucleotide was used to screen the human liver cDNA library. Of $\sim 2.5 \times 10^5$ plaques screened by replicate filter hybridization, 131 contained DNA inserts which annealed strongly to the $\alpha$-probe. The bound $\alpha$-probe was then detached from the filters (confirmed by autoradiography) prior to reprobing with the $\beta$-specific oligonucleotide. Comparison of the distribution of $\alpha$- and $\beta$-reactive phage recombinants revealed an almost perfect concordance between the two sets of autoradiographs. This result strongly suggested that the $\alpha$- and $\beta$-chains of SP-40,40 were encoded by the same mRNA as linked peptide segments on a biosynthetic precursor protein.

### *Complete amino acid sequence of SP-40,40*

One recombinant DNA insert (LK107) was excised from bacteriophage DNA using *Eco*RI endonuclease and shown to consist of two fragments of 0.81 and 0.86 kb respectively, indicating the presence of an internal *Eco*RI restriction site. Both restriction fragments were subcloned into M13 sequencing vectors and the nucleotide sequence of both fragments was determined and confirmed by sequencing each insert on both strands (Figure 2). The fragments were aligned and overlapped using a key segment of peptide sequence which bridged the internal *Eco*RI restriction site junction between the 0.86- and 0.81-kb DNA fragments (see Figure 2). Every peptide sequence from Table I was found within the complete amino acid sequence predicted from the cDNA, confirming that SP-40,40 rather than a closely related variant cDNA had been cloned and sequenced.

Several features are noteworthy from the data of Figure 2. (i) The coding order of the SP-40,40 chains upon the mRNA is $\beta-\alpha$. (ii) The $\beta$-sequence is appended by a run of 22 amino acids which constitute a hydrophobic signal sequence for vectorial export. Two contiguous in-frame methionine codons are present at the start of the signal sequence. (iii) The $\alpha$-chain sequence is read in the same reading frame as the $\beta$-chain which precedes it on the mRNA. The carboxyl-terminal amino acid of the $\beta$-chain is not yet known but the sequence of peptide C17.2 (Table I) indicates that no more than 23 amino acids can be excised from the SP-40,40 precursor protein during its post-synthetic maturation. (iv) There are six potential sites for N-glycosylation on the SP-40,40 precursor protein, three on the $\alpha$-chain and three on the $\beta$-chain. (v) The $\alpha$-chain consists of 222 amino acids and a stop codon is located immediately after the codon for the carboxyl-terminal Glu residue of the mature sequence. (vi) The 3′ untranslated region contains 272 nt including the AATAAA recognition sequence for poly(A) addition (Proudfoot and Brownlee, 1976).

### *SP-40,40 is the human homologue of rat SGP-2*

Computer-assisted homology searches were conducted primarily to determine if SP-40,40 was related in sequence to any known component of the complement system. In the course of these searches SP-40,40 was found to be 76.8% identical to a rat Sertoli cell glycoprotein, SGP-2 (Figure 3). This identity level between SGP-2 and SP-40,40 is well within the normal interspecies variation range between homologous proteins of human and rat (Hudson *et al.*, 1984).

The identity level between SGP-2 and SP-40,40 cDNAs within the coding region is even higher (82%) and the distribution of identities between the two cDNAs is depicted in the Diagon plot of Figure 4. Notably, the distribution of identities indicates that the highest sequence conservation occurs within the coding regions of SGP-2 and SP-40,40,

```
.........................................................................GAATTCCGCCGCTGACCGAGGCGTGCAAAGACTCCAGAATTGGAGGC 47

                                                                                 -1  +1
Met Met Lys Thr Leu Leu Leu Phe Val Gly Leu Leu Leu Thr Trp Glu Ser Gly Gln Val Leu Gly Asp Gln Thr Val Ser Asp Asn Glu Leu Gln Glu 11
ATG ATG AAG ACT CTG CTG CTG TTT GTG GGG CTG CTG CTG ACC TGG GAG AGT GGG CAG GTC CTG GGG GAC CAG ACG GTC TCA GAC AAT GAG CTC CAG GAA 146

Met Ser Asn Gln Gly Ser Lys Tyr Val Asn Lys Glu Ile Gln Asn Ala Val Asn Gly Val Lys Gln Ile Lys Thr Leu Ile Glu Lys Thr Asn Glu Glu 44
ATG TCC AAT CAG GGA AGT AAG TAC GTC AAT AAG GAA ATT CAA AAT GCT GTC AAC GGG GTG AAA CAG ATA AAG ACT CTC ATA GAA AAA ACA AAC GAA GAG 245

Arg Lys Thr Leu Leu Ser Asn Leu Glu Glu Ala Lys Lys Lys Lys Glu Asp Ala Leu Asn Glu Thr Arg Glu Ser Glu Thr Lys Leu Lys Glu Leu Pro 77
CGC AAG ACA CTG CTC AGC AAC CTA GAA GAA GCC AAG AAG AAG AAA GAG GAT GCC CTA AAT GAG ACC AGG GAA TCA GAG ACA AAG CTG AAG GAG CTC CCA 344

Gly Val Cys Asn Glu Thr Met Met Ala Leu Trp Glu Glu Cys Lys Pro Cys Leu Lys Gln Thr Cys Met Lys Phe Tyr Ala Arg Val Cys Arg Ser Gly 110
GGA GTG TGC AAT GAG ACC ATG ATG GCC CTC TGG GAA GAG TGT AAG CCC TGC CTG AAA CAG ACC TGC ATG AAG TTC TAC GCA CGC GTC TGC AGA AGT GGC 443

Ser Gly Leu Val Gly Arg Gln Leu Glu Glu Phe Leu Asn Gln Ser Ser Pro Phe Tyr Phe Trp Met Asn Gly Asp Arg Ile Asp Ser Leu Leu Glu Asn 143
TCA GGC CTG GTT GGC CGC CAG CTT GAG GAG TTC CTG AAC CAG AGC TCG CCC TTC TAC TTC TGG ATG AAT GGT GAC CGC ATC GAC TCC CTG CTG GAG AAC 542

Asp Arg Gln Gln Thr His Met Leu Asp Val Met Gln Asp His Phe Ser Arg Ala Ser Ser Ile Ile Asp Glu Leu Phe Gln Asp Arg Phe Phe Thr Arg 176
GAC CGG CAG CAG ACG CAC ATG CTG GAT GTC ATG CAG GAC CAC TTC AGC CGC GCG TCC AGC ATC ATA GAC GAG CTC TTC CAG GAC AGG TTC TTC ACC CGG 641

Glu Pro Gln Asp Thr Tyr His Tyr Leu Pro Phe Ser Leu Pro His Arg Arg Pro His Phe Phe Phe Pro Lys Ser Arg Ile Val Arg Ser Leu Met Pro 209
GAG CCC CAG GAT ACC TAC CAC TAC CTG CCC TTC AGC CTG CCC CAC CGG AGG CCT CAC TTC TTC TTT CCC AAG TCC CGC ATC GTC CGC AGC TTG ATG CCC 740

Phe Ser Pro Tyr Glu Pro Leu Asn Phe His Ala Met Phe Gln Pro Phe Leu Glu Met Ile His Glu Ala Gln Gln Ala Met Asp Ile His Phe His Ser 242
TTC TCT CCG TAC GAG CCC CTG AAC TTC CAC GCC ATG TTC CAG CCC TTC CTT GAG ATG ATA CAC GAG GCT CAG CAG GCC ATG GAC ATC CAC TTC CAC AGC 839

Pro Ala Phe Gln His Pro Pro Thr Glu Phe Ile Arg Glu Gly Asp Asp Asp Arg Thr Val Cys Arg Glu Ile Arg His Asn Ser Thr Gly Cys Leu Arg 275
CCG GCC TTC CAG CAC CCG CCA ACA GAA TTC ATA CGA GAA GGC GAC GAT GAC CGG ACT GTG TGC CGG GAG ATC CGC CAC AAC TCC ACG GGC TGC CTG CGG 938

Met Lys Asp Gln Cys Asp Lys Cys Arg Glu Ile Leu Ser Val Asp Cys Ser Thr Asn Asn Pro Ser Gln Ala Lys Leu Arg Arg Glu Leu Asp Glu Ser 308
ATG AAG GAC CAG TGT GAC AAG TGC CGG GAG ATC TTG TCT GTG GAC TGT TCC ACC AAC AAC CCC TCC CAG GCT AAG CTG CGG CGG GAG CTC GAC GAA TCC 1037

Leu Gln Val Ala Glu Arg Leu Thr Arg Lys Tyr Asn Glu Leu Leu Lys Ser Tyr Gln Trp Lys Met Leu Asn Thr Ser Ser Leu Leu Glu Gln Leu Asn 341
CTC CAG GTC GCT GAG AGG TTG ACC AGG AAA TAC AAC GAG CTG CTA AAG TCC TAC CAG TGG AAG ATG CTC AAC ACC TCC TCC TTG CTG GAG CAG CTG AAC 1136

Glu Gln Phe Asn Trp Val Ser Arg Leu Ala Asn Leu Thr Gln Gly Glu Asp Gln Tyr Tyr Leu Arg Val Thr Thr Val Ala Ser His Thr Ser Asp Ser 374
GAG CAG TTT AAC TGG GTG TCC CGG CTG GCA AAC CTC ACG CAA GGC GAA GAC CAG TAC TAT CTG CGG GTC ACC ACG GTG GCT TCC CAC ACT TCT GAC TCG 1235

Asp Val Pro Ser Gly Val Thr Glu Val Val Val Lys Leu Phe Asp Ser Asp Pro Ile Thr Val Thr Val Pro Val Glu Val Ser Arg Lys Asn Pro Lys 407
GAC GTT CCT TCC GGT GTC ACT GAG GTG GTC GTG AAG CTC TTT GAC TCT GAT CCC ATC ACT GTG ACG GTC CCT GTA GAA GTC TCC AGG AAG AAC CCT AAA 1334

Phe Met Glu Thr Val Ala Glu Lys Ala Leu Gln Glu Tyr Arg Lys Lys His Arg Glu Glu                                            427
TTT ATG GAG ACC GTG GCG GAG AAA GCG CTG CAG GAA TAC CGC AAA AAG CAC CGG GAG GAG TGAGATGTGGATGTTGCTTTTGCACCTTACGGGGGCATCTTGAGTCCAGCT 1445

CCCCCCAAGATGAGCTGCAGCCCCCCAGAGAGAGCTCTGCACGTCACCAAGTAACCAGGCCCCAGCCTCCAGGCCCCCAACTCCGCCCAGCCTCTCCCCGCTCTGGATCCTGCACTCTAACACTCGACTCT 1576

GCTGCTCATGGGAAGAACAGAATTGCTCCTGCATGCAACTAATTCAATAAAACTGTCTTGTGAGCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAATTC.............................. 1676
```

**Fig. 2.** The nucleotide sequence of clone LK107 and the predicted amino acid sequence of the human SP-40,40 precursor. The two major *Eco*RI fragments of the LK107 cDNA were bidirectionally sequenced in M13 sequencing vectors (DNA overlaps not shown). The information required to overlap the 0.86-kb (5′ end) and 0.81-kb (3′ end) fragments was found in the amino acid sequence of peptides C10.1 and V8.58 (Table I). Segments of the predicted amino acid sequence corresponding to peptide fragments in Table I are overlined. The residues at the junction between the signal peptide and the mature SP-40,40 sequence are designated −1 and +1 respectively. The polyadenylation recognition sequence (AATAAA) is underlined. The internal *Eco*RI site is designated (▽) and the stop codon following the carboxyl-terminal residue is indicated (*).

**Fig. 3.** Comparison of the amino acid sequences of SP-40,40 with SGP-2. The upper line corresponds to the sequence of SP-40,40 and the SGP-2 sequence (lower line) is from Collard and Griswold (1987). The amino-terminal residues of the $\alpha$- and $\beta$-chains of mature SP-40,40 are shown and correspond exactly to their SGP-2 counterparts as determined by direct protein sequencing. Identities between sequences are indicated by solid lines. ● Corresponds to asparagine residue in cDNA sequence not assigned by direct protein sequencing due to presence of attached carbohydrate, i.e. confirmed N-linked glycosylation site. ○ Predicted N-linked glycosylation site. • Deletion of a residue.

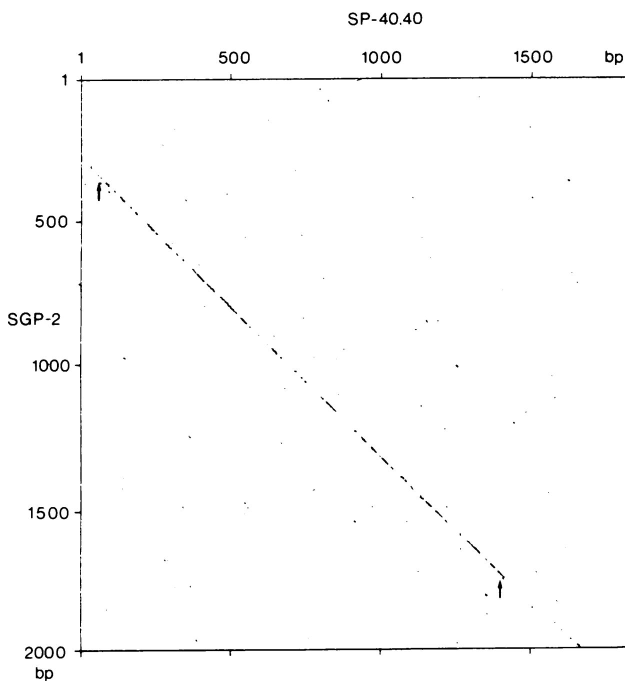


**Fig. 4.** Diagon plot of SP-40,40 (horizontal axis) and SGP-2 (vertical axis) cDNAs. The displacement from the 1,1 position on the vertical axis is due to the fact that the (putative) extreme 5′ terminus of SP-40,40 cDNA was not sequenced in this study. The position of the start and stop codons for SP-40,40 mRNA are indicated by vertical arrows. A span width of 10 and a score of 90% identity were used.

with the 3′ and 5′ untranslated regions exhibiting somewhat lower levels of identity.

SP-40,40 was detected by ELISA in normal human seminal plasma at a concentration significantly higher than that in human serum. In contrast, other human terminal complement components, including C6, C7, C8, C9 and S-protein, were undetectable in seminal plasma (data not shown). The source of the seminal plasma SP-40,40 is likely to be the Sertoli cell which is known to be the major site of synthesis of testicular SGP-2 in the rat (Griswold *et al.*, 1986). Characterization of human seminal plasma using Western immunoblotting revealed no difference in size between the serum and seminal plasma forms of SP-40,40 (Figure 5).

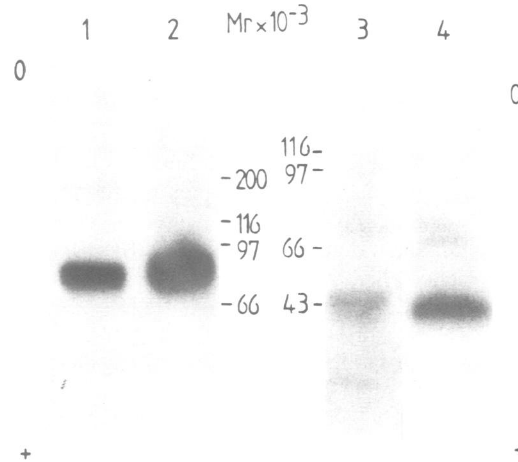


**Fig. 5.** SP-40,40 is a normal constituent of human seminal plasma. Samples of human seminal plasma (**lanes 2, 4**) were subjected to SDS–PAGE along with purified samples of serum SP-40,40 (**lanes 1, 3**). After electrophoresis the gels were subjected to Western immunoblotting using an SP-40,40 monoclonal antibody as described (Murphy *et al.*, 1988). The migration distances of mol. wt marker polypeptides are shown. Samples were run under reducing (**lanes 3, 4**) or non-reducing (**lanes 1, 2**) conditions.

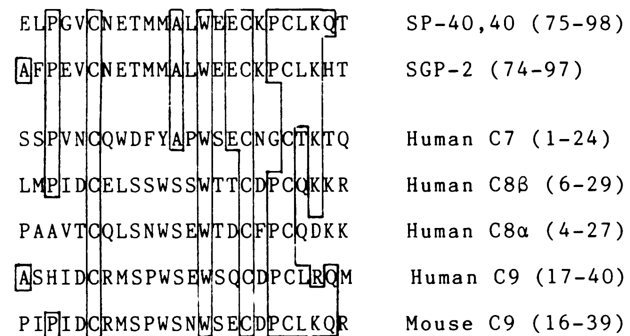


**Fig. 6.** Related cysteine-rich motifs occur in SP-40,40, SGP-2 and several terminal complement components. Residues common to SP-40,40 or SGP-2 and complement components C7, C8$\alpha$, C8$\beta$ and C9 are boxed.

### *Limited homology between SP-40,40 and other complement components*

In Figure 6 a segment within the SP-40,40 $\beta$-chain is aligned to the amino-terminal regions of complement components C7, C8 and C9 (Stanley *et al.*, 1985; Howard *et al.*, 1987; Rao *et al.*, 1987; Di Scipio *et al.*, 1988). These alignments represent the best matches found between SP-40,40 and complement components C1q, C1r and C1s, C2, C3, C4, C5, C7, C8, C9 and S-protein.

## Discussion

This paper describes the molecular cloning of a cDNA for SP-40,40, a recently discovered member of the human complement system (Murphy *et al.*, 1988). The SP-40,40 protein is a normal component of human blood and its concentration in human serum samples ranges from 35 to 105 μg/ml regardless of the sex of serum donors. We previously demonstrated the presence of SP-40,40 in the SC5b-9 complex which consists of complement components C5b, C6, C7, C8, C9 and S-protein.

The nucleotide sequence of SP-40,40 cDNA (Figure 2) revealed that its $\alpha$- and $\beta$-chains were not related in sequence and this accords well with the peptide mapping studies (Figure 1). However, both chains were found to be coded on the same mRNA, the translation of which would produce a precursor protein consisting of an amino-terminal signal sequence of 22 residues and a pro SP-40,40 molecule containing 427 amino acids. Two contiguous, in-frame methionine codons occur at the start of the signal sequence and translation is most likely to begin at the first of these codons (Kozak *et al.*, 1980).

The $\alpha$-chain of the structure contains 222 amino acids and its coding segment is located 3′ to the $\beta$-chain in the mRNA. The $\beta$ chain contains between 182 and 205 amino acids. At least one cleavage event, proteolysis of the Arg-205–Ser-206 peptide bond takes place during the post-synthetic maturation of the pro SP-40,40 structure. The cysteine residues in SP-40,40 are distributed five per chain in two relatively discrete clusters. The mature structure must contain at least one interchain disulphide bond (Murphy *et al.*, 1988).

A 76.8% identity level between the polypeptide sequences of SP-40,40 and SGP-2 was revealed by computer-assisted homology searches (Figure 3). This level of identity establishes that SP-40,40 and SGP-2 are species counterparts in human and rat respectively. Detection of SP-40,40 in human seminal plasma (Figure 5) at levels higher than in blood serum together with the binding of SP-40,40-specific mAb to human sperm (data not shown) provides further evidence that SP-40,40 is the human counterpart of SGP-2. At positions where the SP-40,40 and SGP-2 sequences differ, amino acid interchanges are generally conservative (Figure 3) and this is reflected in the 82% identity level between their cDNA sequences in the coding region. The amino-terminal and cDNA sequences of the mature forms of both proteins indicate that both are derived from single chain precursors (Collard and Griswold, 1987) and that the proteolytic cleavage event(s) which ensue during maturation are analogous. One of the SGP-2 chains is smaller (34 kd) than its SP-40,40 correspondent ($\alpha$, 40 kd), probably due to the addition of carbohydrate to an extra N-linked glycosylation site in SP-40,40 $\alpha$. Conversely the SGP-2 $\beta$-chain (47 kd) contains one more presumptive glycosylation site than SP-40,40 $\beta$ (40 kd), accounting for its increased mol. wt. The locations of all cysteine residues in SP-40,40 and SGP-2 are conserved.

The unexpected feature of this work is that the systems in which SGP-2 and SP-40,40 participate appear to be physiologically remote. SP-40,40 is involved in the terminal phase of the complement pathway as suggested by its histologic association with other components of SC5b-9, the principal form of C5b-9 now known to occur in glomerular immune deposits (Murphy *et al.*, 1988). In addition, its specific incorporation into the SC5b-9 complex, generated by activating the alternative pathway *in vitro*, has been demonstrated. Detailed studies are presently under way to investigate the exact role of SP-40,40 within the human complement pathway.

The role of SGP-2 within the reproductive pathway is unclear. It is the major secreted product of the Sertoli cell (Griswold *et al.*, 1986) and binds to the sperm plasma membrane proximal to the acrosome and to the distal segment of the tail (Sylvester *et al.*, 1984). The presence of high levels of SP-40,40 in human seminal plasma accords well with the conclusion, based on comparison of their sequences, that SGP-2 and SP-40,40 are species homologues. The major forms of SP-40,40 in blood and seminal plasma were indistinguishable as judged by reducing and non-reducing SDS–PAGE followed by Western immunoblotting (Figure 5). Three possibilities may be considered to account for the presence of SP-40,40 in seminal plasma and on sperm membranes.

(i) SP-40,40 modulates endogenous complement activity within the female urogenital tract. A complement inhibitor may be necessary to protect sperm within the vagina and uterus from the complement system of the female.

The role of the immune system within the female reproductive tract is at present controversial. On the one hand high serum levels of sperm-specific antibodies have been suggested as a cause of both male and female infertility (Rümke *et al.*, 1975; Aitken and Paterson, 1988; Clarke, 1988; Tsuiji *et al.*, 1988). In addition, haemolytic activity within cervical mucus has previously been demonstrated, suggesting that complement may play a role in immunological defence within the female reproductive tract (Price and Bouttcher, 1979). On the other hand the correlation between high serum levels of sperm-specific antibodies and infertility is by no means perfect (Hjort and Brogaard Hansen, 1983) and it has never been demonstrated that the levels of haemolytic activity within cervical mucus would suffice to promote the lysis of antibody coated sperm. The role of SP-40,40 within seminal plasma may nevertheless be to inhibit complement-mediated lysis of sperm within the female reproductive tract. It should be of considerable interest in this context to assay the seminal plasma of selected infertile males to determine whether SP-40,40 is absent or diminished in concentration.

(ii) A complement-like mechanism which SP-40,40 modulates is involved in sperm capacitation or in fertilization. Disruption of sperm plasma membranes occurs after delivery into the female genital tract. The acrosome reaction which involves the fusion of the acrosomal membrane to the sperm plasma membrane commences shortly before fertilization (Talbot, 1985; Wassarman, 1987; Yanagimachi, 1988). Vesicles containing degradative enzymes and composed partly of acrosomal membrane and partly of overlying plasma membrane are released during the acrosome reaction. Although the acrosome reaction *in vivo* seems to occur only after sperm has attached to the zona pellucida of the oocyte, sperm can be induced to undergo this reaction *in vitro* without contact with the zona pellucida. Sperm may contain an endogenous complement-like system capable of producing the membrane rupture and vesiculation associated with the acrosome reaction. SP-40,40 might prevent the premature triggering of this reaction by engaging and inhibiting a

complement-like intermediate in this event. To invoke a mechanism for membrane perturbation involving complement-related structures which does not involve classical complement components is not totally without precedent. Perforin, a major protein constituent of the granules within cytotoxic T cells exhibits significant structural homology (27%) to complement component C9 (Shinkai *et al.*, 1988). The C9 component is the main constituent of the complement lytic pore and the homology between C9 and perforin suggests that perforin may be functionally analogous to C9 in the lytic pathway used by cytotoxic T cells to kill target cells.

(iii) SP-40,40 plays a role in spermatogenesis or fertilization unrelated to its role in the complement cascade. It is likely that SP-40,40 and SGP-2 are the species homologues of clusterin, a protein isolated from ram rete testis (Rosenoir *et al.*, 1987). Clusterin, an 80-kd heterodimer which is the major secreted product of ram Sertoli cells, is reported to promote specifically the clustering of several cell types (Fritz *et al.*, 1983). A more detailed understanding of the function of SP-40,40 within the complement system should clarify the physiological significance of the cellular adhesive properties of clusterin.

Another component of the complement system, S-protein, was recently shown to be identical to vitronectin (serum spreading factor, named for its *in vitro* adhesive properties; Jenne and Stanley, 1985). The role played by S-protein within the complement system, to inhibit formation of the membrane attack complex, seems to be mechanistically unrelated to its effect on the coagulation system which is to annul the thrombin–antithrombin III interaction. It will be important to know if SP-40,40, like S-protein, is a multifunctional protein performing unrelated roles in the reproductive system in one instance and the immune system in the other.

Comparison between the amino acid sequences of SP-40,40 and S-protein did not reveal an obvious evolutionary relationship which might reflect a structural as well as functional link between them. However, limited homology between a short segment of SP-40,40 and corresponding segments in complement components C7, C8 and C9 was revealed in parallel comparisons (Figure 6). The homology between this region of the complement components to several other membrane-associated structures has recently been documented (Goundis and Reid, 1988; Robson *et al.*, 1988; Stanley, 1988). In particular, this segment within the amino-terminal sequences of C7, C8 and C9 exhibits evolutionary homology to repetitious sequences within thrombospondin, a platelet-derived polypeptide which acts within the blood coagulation system (Di Scipio *et al.*, 1988). No direct homology was detected between the SP-40,40 segment and the thrombospondin repeats, but the alignment of Figure 6 suggests that key features of an ancient conserved, cysteine-rich motif common to certain terminal components of complement, thrombospondin, properdin (an activator of the alternative pathway of complement activation) and the major surface protein of malaria sporozoites, may occur in SP-40,40 (Goundis and Reid, 1988; Robson *et al.*, 1988). The homology between SP-40,40 and the complement components in Figure 6 relies heavily on the locations of cysteine residues spaced identically within the five sequences compared. The limited homology which the comparisons imply would be further strengthened by the demonstration of a similar disulphide bonding pattern within the structures compared.

Current studies on SP-40,40 are aimed at providing a more detailed understanding of its role within the complement system. It is notable that other proteins containing cysteine-rich motifs related to that present in SP-40,40 are either membrane proteins or are destined to interact with membranes. The elucidation of the exact roles of these cysteine-rich motifs within the specific proteins which contain them should contribute important insights on the possible function of SP-40,40 within the membrane attack complex of the complement pathway and its role within the reproductive system.

## Materials and methods

### Protein chemistry

Samples of human serum were obtained from a healthy male volunteer and were subjected to affinity chromatography as described (Murphy *et al.*, 1988). Purified SP-40,40 was reduced and carboxamidomethylated. The $\alpha$- and $\beta$-chains were purified and subjected separately to cleavage with a number of proteolytic enzymes and with cyanogen bromide as previously described (Classon *et al.*, 1986; Walker *et al.*, 1986; Hibbs *et al.*, 1988). Peptides were isolated using reverse-phase HPLC and sequenced using an Applied Biosystems Inc. model 470A protein sequencer equipped with an on-line injector. PTH-amino acids were detected as previously described (Cole *et al.*, 1986).

### Molecular cloning

Selected segments of amino acid sequence were used to predict the sequences of oligonucleotide probes as described by Lathe (1985). Oligonucleotides were synthesized using an Applied Biosystems Inc. DNA Synthesizer. Oligonucleotides were 5′ end-labelled with $^{32}P$ as described (Hibbs *et al.*, 1986) and used as probes to screen a commercially available human liver cDNA library (Clontech). The library contained $5.5 \times 10^5$ independent cDNA clones inserted into the *Eco*RI restriction site of the cloning vector λ-gt11. Plaques containing probe-reactive inserts were picked and recombinant phage were clonally purified. Insert DNA was excised by *Eco*RI endonuclease digestion and sized by agarose gel electrophoresis using ethidium bromide to visualize bands.

### DNA sequencing

*Eco*RI endonuclease digests of appropriate recombinant phage DNA were ligated into M13 sequencing vectors and sequenced using the dideoxynucleotide chain termination technique of Sanger *et al.* (1977).

### Enzyme-linked immunoassays

Samples of human serum and seminal plasma were analysed by ELISA using two SP-40,40 monoclonal antibodies directed to different epitopes as previously described (Murphy *et al.*, 1988). Western transfer studies were conducted after SDS–PAGE analysis (Laemmli, 1970) as previously described (Murphy *et al.*, 1988).

## Acknowledgements

We wish to thank Drs Harry Ward and Morry Frenkel for constructive comments on the manuscript, Mr Phillip Smith for his photographic work and Mrs Lesley Boseley for kindly typing the manuscript. Dr Tony Kyne provided valuable assistance in the computer homology searches. This work was supported by grants from The Jack Brockhoff Foundation, The William Buckland Foundation, The Ian Potter Foundation and the National Health and Medical Research Council.

## References

Aitken,R.J. and Paterson,M. (1988) *Nature*, **335**, 492–493.

Bhakdi,S. and Roth,M. (1981) *J. Immunol.*, **127**, 576–580.

Clarke,G.N. (1988) *Am. J. Reprod. Immunol.*, **16**, 135–138.

Classon,B.J., Tsagaratos,J., McKenzie,I.F.C. and Walker,I.D. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4499–4503.

Cole,W.G., Chan,D., Chambers,G.W., Walker,I.D. and Bateman,J.-F. (1986) *J. Biol. Chem.*, **261**, 4496–4503.

Collard,M.W. and Griswold,M.D. (1987) *Biochemistry*, **26**, 3297–3303.
Di Scipio,R.J., Chakravarti,D.N., Müller-Eberhard,H.J. and Fey,J.H. (1988) *J. Biol. Chem.*, **263**, 549–560.
Fritz,I.B., Burdzy,K., Setchell,B. and Blaschak,O. (1983) *Biol. Reprod.*, **28**, 1173–1188.
Griswold,M.D., Roberts,K.P. and Bishop,P.D. (1986) *Biochemistry*, **25**, 7265–7270.
Goundis,D. and Reid,K.B.M. (1988) *Nature*, **335**, 82–85.
Hibbs,M.L., Walker,I.D., Kirszbaum,L., Pietersz,G.A., Deacon,N.J., Chambers,G.W., McKenzie,I.F.C. and Hogarth,P.M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6980–6984.
Hibbs,M.L., Classon,B.J., Walker,I.D., Hogarth,P.M. and McKenzie, I.F.C. (1988) *J. Immunol.*, **140**, 544–550.
Hjort,T. and Brogaard Hansen,K. (1983) In Schullman,S. and Dondero,D. (eds), *Immunological Factors in Human Contraception*. Acta Medica Roma, Rome, p. 47.
Howard,O.M.Z., Rao,A.G. and Sodetz,J.M. (1987) *Biochemistry*, **26**, 3565–3570.
Hudson,P., John,M., Crawford,R., Scanlon,D., Gorman,G., Tregear,G., Shine,J. and Niall,H. (1984) *EMBO J.*, **3**, 2333–2339.
Jenne,D. and Stanley,K.K. (1985) *EMBO J.*, **4**, 3153–3157.
Kozak,M. (1980) *Cell*, **22**, 7–8.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Lathe,R. (1985) *J. Mol. Biol.*, **183**, 1–12.
Müller-Eberhard,H.J. (1986) *Annu. Rev. Immunol.*, **4**, 503–528.
Murphy,B.F., Kirszbaum,L., Walker,I.D. and d'Apice,A.J.F. (1988) *J. Clin. Invest.*, **81**, 1858–1864.
Podack,E.R., Kolb,W.P. and Müller-Eberhard,H.J. (1977) *J. Immunol.*, **119**, 2024–2029.
Podack,E.R., Kolb,W.P. and Müller-Eberhard,H.J. (1978) *J. Immunol.*, **120**, 1841–1848.
Price,R.J. and Bouttcher,B. (1979) *Fertil. Steril.*, **32**, 61–69.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Rao,A.G, Howard,O.M.Z., Ng,S.C., Whitehead,A.S., Colton,H.R. and Sodetz,J.M. (1987) *Biochemistry*, **26**, 3556–3564.
Reid,K. (1986) *Essays Biochem.*, **22**, 27–68.
Robson,K.J.H., Hall,J.R.S., Jennings,M.W., Harris,T.J.R., Marsh,K., Newbold,C.I., Tate,V.E. and Weatherall,D.J. (1988) *Nature*, **335**, 79–82.
Rosenoir,J., Tung,P.S. and Fritz,I.B. (1987) *Biol. Reprod.*, **36**, 1313–1320.
Rümke,P.H., Van Amstel,M., Messer,E.N. and Bezemer,P.D. (1974) *Fertil. Steril.*, **25**, 393–398.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Shinkai,Y., Tsakio,K. and Okumura,K. (1988) *Nature*, **334**, 527–529.
Stanley,K.K. (1988) *Nature*, **334**, 475–476.
Stanley,K.K., Kocher,H.-P., Luzio,J.P., Jackson,P. and Tschopp,J. (1985) *EMBO J.*, **4**, 375–382.
Sylvester,S.R., Skinner,M.K. and Griswold,M.D. (1984) *Biol. Reprod.*, **31**, 1087–1101.
Talbot,P. (1985) *Am. J. Anat.*, **174**, 331–346.
Tsuji,Y., Clausen,H., Nudelman,E., Kaizu,T., Hakomori,S.I. and Isojima,S. (1988) *J. Exp. Med.*, **168**, 343–356.
Walker,I.D, Murray,B.J., Kirszbaum,L., Chambers,G.W., Deacon,N.J. and McKenzie,I.F.C. (1986) *Immunogenetics*, **23**, 60–63.
Wassarman,P.M. (1987) *Science*, **235**, 553–560.
Yanagimachi,R. (1988) In Knobil,E. and Neill,J. (eds), *The Physiology of Reproduction*. Raven Press, New York, USA, pp. 135–185.

*Received on November 11, 1988; revised on December 23 1988*