The EMBO Journal vol.11 no.3 pp.829-837, 1992

Isolation of a hagfish gene that encodes a complement component

Hiroshi Ishiguro, Kunihiko Kobayashi[1], Masami Suzuki, Koiti Titani, Susumu Tomonaga[2] and Yoshikazu Kurosawa

Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-11, [1]School of Medicine, Hokkaido University, Sapporo 060, and [2]School of Allied Health Sciences, Yamaguchi University, Ube 755, Japan

Communicated by A.F.Williams

It has been widely accepted that cyclostomes are the most primitive vertebrates extant with the ability to produce antibodies. We isolated cDNA clones that encode a putative 'antibody' from one of the cyclostomes, *Eptatretus burgeri*. The amino acid sequence predicted from the nucleotide sequences of the cDNA clones indicated that this gene does actually encode the proteins isolated as hagfish 'antibodies' by various investigators. However, these proteins are not similar to mammalian immunoglobulins but have some characteristics common to complements C3, C4 and C5 in higher vertebrates. We discuss the relationships of the isolated gene for hagfish complement with the mammalian genes for complements C3, C4 and C5. We also discuss the possibility of the presence of antibodies in cyclostomes.
Key words: complement/cyclostome/evolution/immunoglobulin

Introduction

The most intriguing question in the evolution of the adaptive immune system is how did the immunoglobulins (Igs), T-cell receptors (TcRs) and antigens of the major histocompatibility complex (MHC) evolve? There is a consensus of opinion that Ig and TcR genes are derived from a single primordial gene (Williams and Barclay, 1988). It remains to be determined which gene system appeared first, and which species acquired these gene systems during evolution. It is widely accepted that cyclostomes are the most primitive vertebrates extant with the ability to produce antibodies (Abs) (for review see Du Pasquier, 1989). Ig genes in elasmobranchs have been well characterized by Litman's group (Hinds and Litman, 1986), but no one has isolated Ig genes of cyclostome. Marchalonis and Edelman (1968) reported that sea lamprey, *Petromyzon marinus*, one of the cyclostomes, produces a specific Ab in response to immunization with bacteriophage f2. Subsequently, Raison *et al.* (1978) showed that hagfish, *Eptatretus stoutii*, also produces an Ab against group A streptococcal carbohydrate after prolonged immunization with the whole cell vaccine. Judging from the molecular weights and immunoelectrophoretic properties of the two Abs, the authors concluded that the heavy (H) chains of the Abs are μ-like.

Hagfish Abs have been purified by three groups of investigators (Raison *et al.*, 1978; Kobayashi *et al.*, 1985; Varner *et al.*, 1991). Raison *et al.* (1978) reported that the isolated Ab has a high binding affinity for the streptococcal carbohydrate used for immunization. The N-terminal partial amino acid sequences of the H chains of this molecule were reported by Raison's group (Hanley *et al.*, 1990). Litman and his coworkers also reported the partial amino acid sequences of a putative Ig isolated as a reduction-sensitive heterodimer from hagfish serum (Varner *et al.*, 1991). Based on the physical properties, serological cross-reactivity with antisera against Ig of horned shark (*Heterodontus francisci*) and the peptide sequences of the isolated molecules, they suggested that the isolated protein was an Ig. In both studies, the H chains were split into two polypeptides of different sizes, amino acid sequences and composition. In this study we first determined the N-terminal amino acid sequence of the H chains of a putative Ig isolated from another kind of hagfish, *Eptatretus burgeri*, by Kobayashi *et al.* (1985) and then isolated cDNA clones that encoded this molecule. The amino acid sequence predicted from their nucleotide sequences indicated that the three proteins isolated by three independent groups as hagfish Ig (Kobayashi *et al.*, 1985; Hanley *et al.*, 1990; Varner *et al.*, 1991) are identical to one another. Contrary to previous hypotheses, however, these molecules are not similar to mammalian Ig but are similar to complements C3, C4 and C5.

A complement system consists of a set of plasma and cell surface proteins playing an important role in amplification of the recognition and elimination of foreign substances (for review, see Frank and Fries, 1989). This process occurs through an activation cascade which results in the formation of bioactive fragments by proteolytic cleavage and the assembly of a protein complex that is capable of lysing cells. Two pathways have been shown to trigger the activation cascade. The classical pathway of complement is initiated by the formation of an antigen–Ab complex. Recognition of the antigen–Ab complex by the proteins in the classical pathway leads to sequential formation of enzymes with serine protease activity. These enzymes cleave and activate C3. The alternative complement pathway is initiated without Ab (for review see Muller-Eberhard and Schreiber, 1980). Six plasma proteins, C3, factors D, B, H and I and properdine, participate in the initial activation step of this pathway. Among the components of complement, C3, C4 and C5 exhibit functional as well as chemical homologies. C3 plays a key role in both the classical and alternative pathways of complement by interacting with various other plasma and cell surface components of complements (for review see Becherer *et al.*, 1989). Activation of the classical pathway is initiated by the binding of C1 to an antigen–Ab complex and the subsequent activation of the Ab-bound C1. The activated C1 further activates C4 by proteolysis. C4 is not involved in the alternative pathway. C5 is activated by proteolysis with enzymes formed during activation of both the classical and alternative pathways. C3, C4 and C5 were

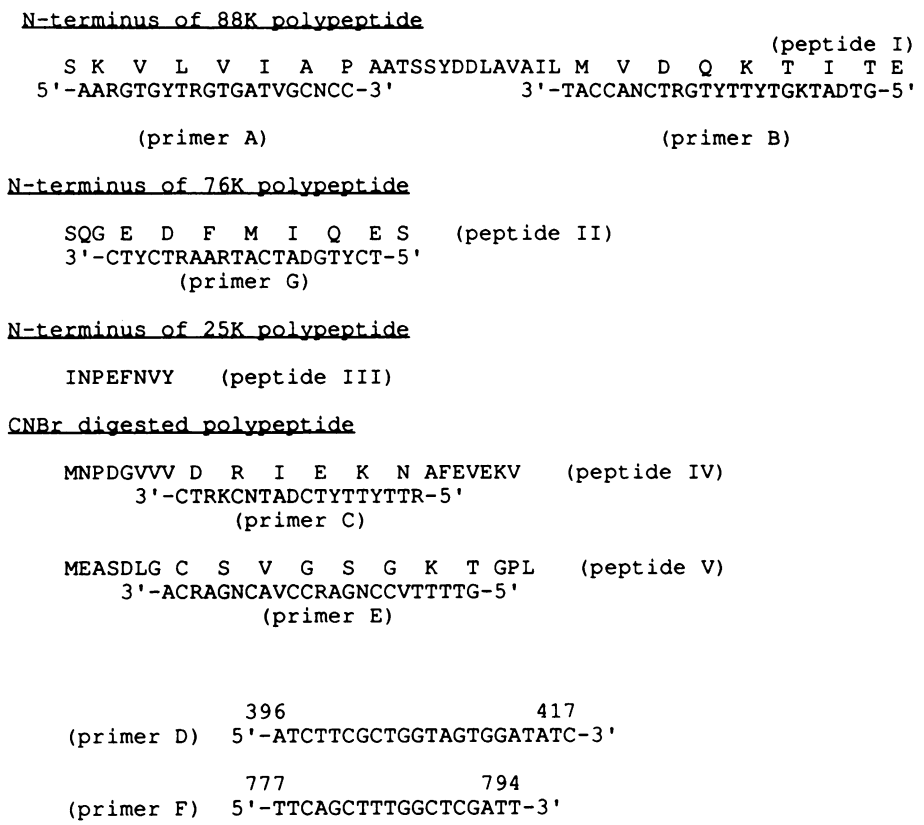


Fig. 1. Partial amino acid sequences of hagfish 'antibody' and primers for PCR. Amino acid sequences of five peptides (peptides I–V) were determined. Based on the peptide sequences, five kinds of mixed oligonucleotide were designed as primers for PCR. Two other oligonucleotides were based on the determined nucleotide sequences. Nucleotide usages are as follows: R: G,A; Y: C,T; K: G,T; V: A,C,T; D: G,A,T; N: G,A,C,T.

shown to have subunit structures. C3 is composed of two disulfide-linked chains: α, 120 kDa and β, 75 kDa (De Bruijn and Fey, 1985). C4 is composed of three disulfide-linked chains, termed α, β and γ, of 93, 78 and 33 kDa, respectively (Belt *et al.*, 1984). C5 consists of two disulfide-linked chains: α, 115 kDa and β, 75 kDa (Wetsel *et al.*, 1987). C3, C4 and C5 are synthesized as a single-chain precursor molecule with the β chain at the N-terminus and each subunit structure is formed as a post-translational event.

Xenopus is the lowest vertebrate from which a gene for complement has been isolated to date (Grossberger *et al.*, 1989). The presence of complements in cyclostomes has already been demonstrated (Day *et al.*, 1970). Nonaka *et al.* (1984) purified a complement protein from lamprey that was functionally homologous to mammalian C3 and structurally homologous to mammalian C4. It consisted of α, β and γ (84, 74 and 32 kDa, respectively) linked by disulfide bonds. Fujii *et al.* (1992) purified a complement protein from hagfish that was homologous to mammalian C3. In the case of the hagfish, however, the protein was composed of two subunits of 115 and 77 kDa linked by disulfide bonds. The hagfish gene isolated in this study encoded the complement protein purified by Fujii *et al.* (1992) although the subunit structure was similar to that of the lamprey complement.

Results

Purification of the putative Ig of the hagfish, *E.burgeri*, was described previously (Kobayashi *et al.*, 1985). The purified protein was subjected to SDS–PAGE under reducing conditions. Two H chain bands, 88 and 76 kDa, were identified although Kobayashi *et al.* (1985) described the presence of only a single H chain band. This discrepancy may be due to differences in resolution during SDS–PAGE. Two light (L) chains were also identified, a major (25 kDa) and a minor (22 kDa) band, in agreement with the previous observation (Kobayashi *et al.*, 1985). After separation by SDS–PAGE, the polypeptides were electroblotted onto a polyvinylidene difluoride (PVDF) membrane, and the N-terminal sequences of the 76 kDa and 25 kDa polypeptides were determined as described by Matsudaira (1987). In the case of the 88 kDa polypeptide, the N-terminal sequence of 31 amino acids was determined after elution of the band from the gel. A mixture of two H chains was treated with cyanogen bromide (CNBr) and separated by reverse phase HPLC. The amino acid sequences of the two fragments were determined. Figure 1 summarizes all the partial amino acid sequences that were determined.

To prepare the probe for hybridization, we used the polymerase chain reaction (PCR) (Saiki *et al.*, 1988). Two primers, A and B, corresponding to the N- and C-terminal sequences of peptide I, respectively, were synthesized for PCR. Since the specific cell type(s) expressing these molecules were unknown, hagfish genomic DNA was subjected to PCR. Amplified products of 86 nucleotides in length were cloned and sequenced. One clone, corresponding to peptide I, was selected as a hybridization probe. A hagfish genomic DNA library was constructed in λ phage, screened with this probe, and one of the positive clones, λhag-1, was characterized (Figure 2). Figure 3 shows the nucleotide sequence of the fragment that contained the probe-positive

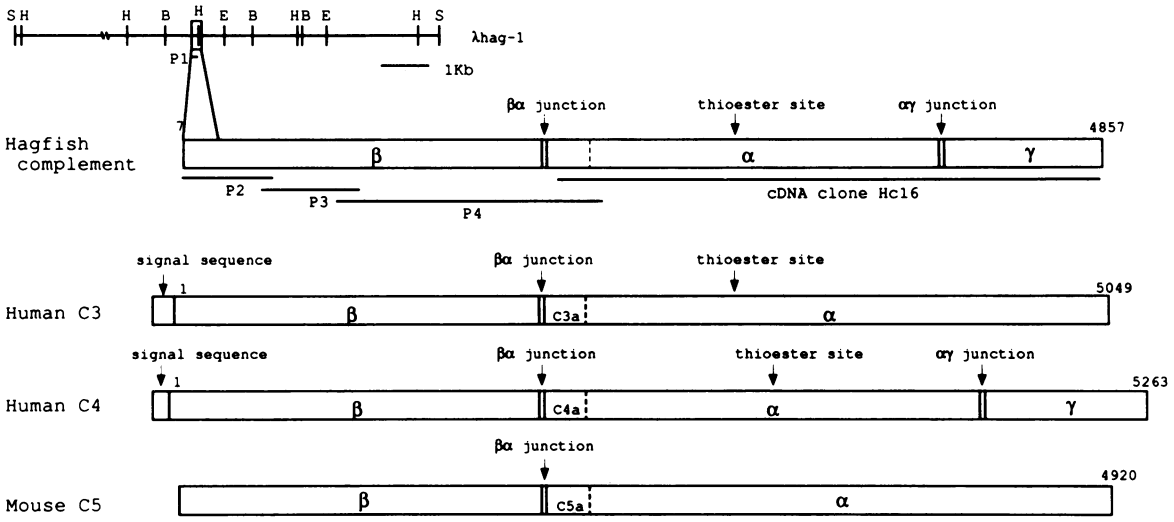


Fig. 2. Gross structures of hagfish complement and mammalian compléments C3, C4 and C5. Restriction map of λhag-1 (top line). Restriction sites: S, *Sau*3A (partially digested ends); H, *Hin*dIII; B, *Bam*HI; E, *Eco*RI. P1 corresponds to fragment I. P2 corresponds to the region amplified by PCR using primers A and C. P3 corresponds to the region amplified using primers D and E. P4 corresponds to the region amplified using primers F and G. Hc16 is a cDNA clone. The structures of human C3, C4 and mouse C5 were taken from De Bruijn *et al.* (1985); Belt *et al.* (1984) and Wetsel *et al.* (1987), respectively.

```
aatacgcctttcctgcaaagccggacaaattatgccagcaaatacaacagatgtgcatgg
tgaatttaatacttttgctcatgataagatgatgataaatatcaccattttgtttgctat
                          V  L  V  I  A  P  A  A  T  S  S
gtctacctgatgaacccccccactagGGTGCTTGTGATTGCGCCGGCTGCAACCTCTTCG
 AccI
 Y  D  D  L  A  V  A  I  L  M  V  D  Q  K  K  I  T  E  V  H
TACGATGACCTCGCAGTTGCAATTTTAATGGTGGATCAGAAAAAAATCACCGAAGTACAT
 V  L  L  V  N  P  H  T  G  A  T  L  D  E  K  K  V  K  L  Q
GTTCTTTTGGTGAACCCACACACTGGCGCTACTCTGGATGAGAAAAAAGTGAAATTGCAA
 W  D  N  K  F  I  A  F  T  K  L  Q
TGGGATAACAAATTCATTGCCTTCACTAAGCTTCAGgtaagtagtttgtttca
                          HindIII
```

Fig. 3. Partial nucleotide sequence of λhag-1. A region encoding peptide I was sequenced. Two amino acid residues at the N-terminus of peptide I were not encoded by this region. This *Acc*I−*Hin*dIII fragment was used as a probe (probe A) for further analysis.

region. This region encodes peptide I with the exception of two amino acids that are missing from the N-terminus and one amino acid difference. Two amino acids at the N-terminus of peptide I should be encoded by another exon, since a pyrimidine-rich stretch plus an AG sequence was observed in the upstream region. Using the 206 bp *Acc*I−*Hin*dIII fragment indicated in Figure 3 as probe (probe A), Southern hybridization was carried out. As shown in Figure 4A, one band was detected in the *Eco*RI-, *Hin*dIII- and *Bam*HI-digested hagfish DNA. This gene seems not to be part of a large family, as expected for genes for Ig V_H. In order to determine the organ in which this gene is expressed, various mRNAs derived from different tissues, including liver, buffy coat fraction, pronephros and intestine, were subjected to Northern hybridization with probe A. A weak band was detected at 5 kb only in the case of liver-derived mRNA (Figure 4B). Since the amount of mRNA encoding this gene seemed to be very small and the size of the mRNA was relatively large, a PCR strategy was adopted to obtain cDNAs, instead of construction of a cDNA library, as follows. Two combinations of primers A and C, as well as A and E, described in Figure 1, were tested for PCR using the cDNA prepared from hagfish liver mRNA. A combination of primer A and primer C complementary to the RNA sequence that encodes the amino acid sequence DRIEKN in peptide IV yielded a 462 bp amplified fragment which was positive to probe A. This 462 bp fragment was cloned

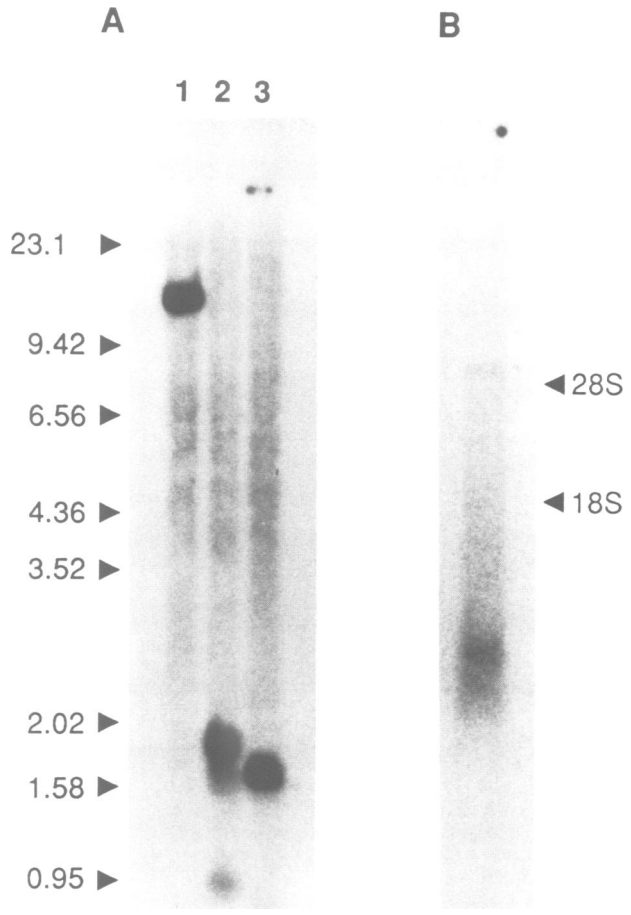
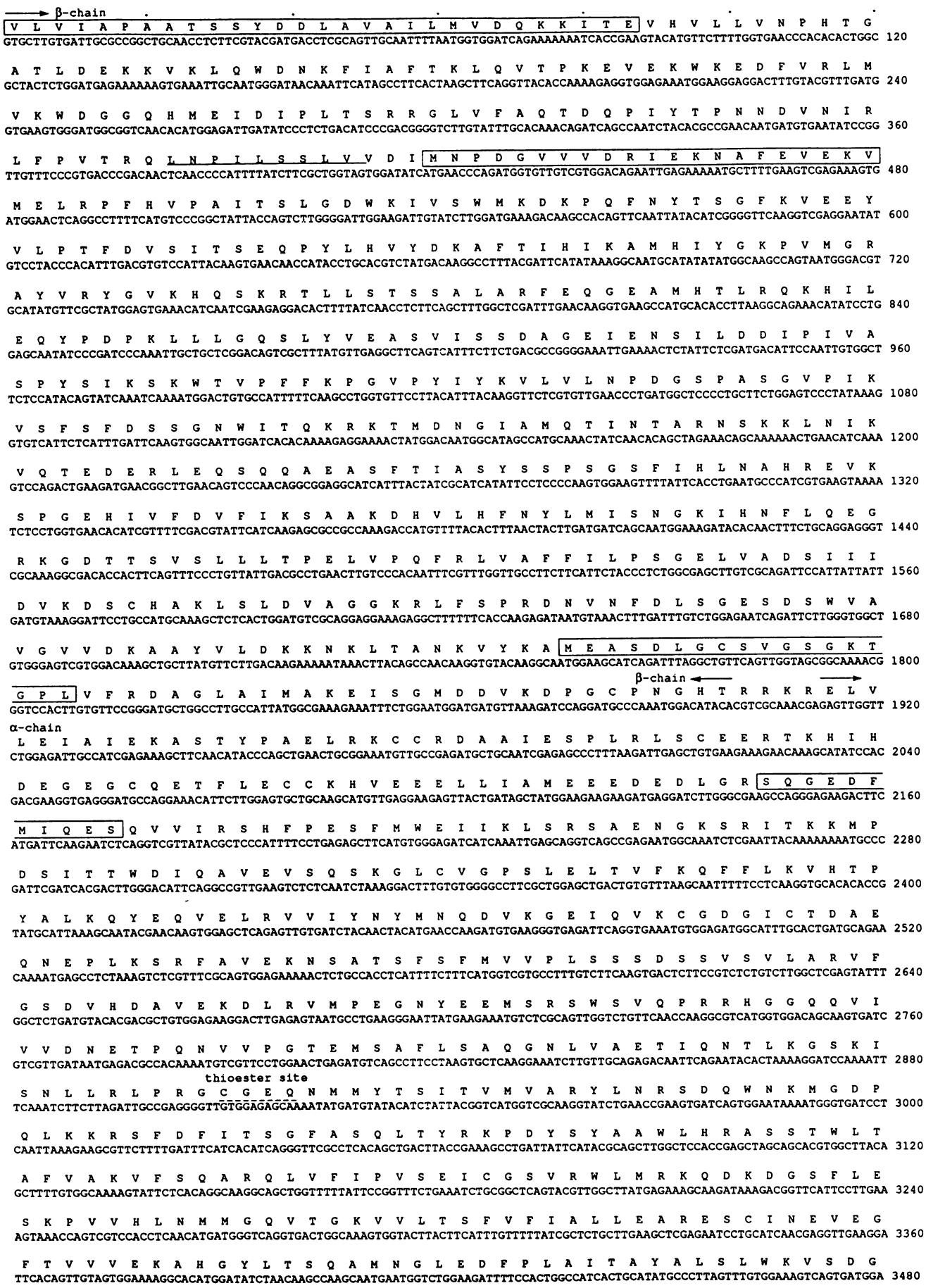


Fig. 4. Southern and Northern hybridizations of hagfish genomic DNA and mRNA. (**A**) Five μg of hagfish genomic DNA were digested with *Eco*RI (lane 1), *Bam*HI (lane 2) or *Hin*dIII (lane 3), and fractionated on a 1% agarose gel. *Hin*dIII and *Hin*dIII−*Eco*RI-digested λ phage DNA was used as a size marker and sizes of fragments are shown in kb. (**B**) Two μg of poly(A)-containing mRNA from liver were subjected to electrophoresis on a 1% agarose gel. Probe A was used for both experiments.


β-chain
V L V I A P A A T S S Y D D L A V A I L M V D Q K K I T E V H V L L V N P H T G
GTGCTTGTGATTGCGCCGGCTGCAACCTCTTCGTACGATGACCTCGCAGTTGCAATTTTAATGGTGGATCAGAAAAAAATCACCGAAGTACATGTTCTTTTGGTGAACCCACACACTGGC 120
A T L D E K K V K L Q W D N K F I A F T K L Q V T P K E V E K W K E D F V R L M
GCTACTCTGGATGAGAAAAAAGTGAAATTGCAATGGGATAACAAATTCATAGCCTTCACTAAGCTTCAGGTTACACCAAAAGAGGTGGAGAAATGGAAGGAGGACTTTGTACGTTTGATG 240
V K W D G G Q H M E I D I P L T S R R G L V F A Q T D Q P I Y T P N N D V N I R
GTGAAGTGGGATGGCGGTCAACACATGGAGATTGATATCCCTCTGACATCCCGACGGGGTCTTGTATTTGCACAAACAGATCAGCCAATCTACACGCCGAACAATGATGTGAATATCCGG 360
L F P V T R Q L N P I L S S L V V D I M N P D G V V V D R I E K N A F E V E K V
TTGTTTCCCGTGACCCGACAACTCAACCCCATTTTATCTTCGCTGGTAGTGGATATCATGAACCCAGATGGTGTTGTCGTGGACAGAATTGAGAAAAATGCTTTTGAAGTCGAGAAAGTG 480
M E L R P F H V P A I T S L G D W K I V S W M K D K P Q F N Y T S G F K V E E Y
ATGGAACTCAGGCCTTTTCATGTCCCGGCTATTACCAGTCTTGGGGATTGGAAGATTGTATCTTGGATGAAAGACAAGCCACAGTTCAATTATACATCGGGGTTCAAGGTCGAGGAATAT 600
V L P T F D V S I T S E Q P Y L H V Y D K A F T I H I K A M H I Y G K P V M G R
GTCCTACCCACATTTGACGTGTCCATTACAAGTGAACAACCATACCTGCACGTCTATGACAAGGCCTTTACGATTCATATAAAGGCAATGCATATATATGGCAAGCCAGTAATGGGACGT 720
A Y V R Y G V K H Q S K R T L L S T S S A L A R F E Q G E A M H T L R Q K H I L
GCATATGTTCGCTATGGAGTGAAACATCAATCGAAGAGGACACTTTTATCAACCTCTTCAGCTTTGGCTCGATTTGAACAAGGTGAAGCCATGCACACCTTAAGGCAGAAACATATCCTG 840
E Q Y P D P K L L L G Q S L Y V E A S V I S S D A G E I E N S I L D D I P I V A
GAGCAATATCCCGATCCCAAATTGCTGCTCGGACAGTCGCTTTATGTTGAGGCTTCAGTCATTTCTTCTGACGCCGGGGAAATTGAAAACTCTATTCTCGATGACATTCCAATTGTGGCT 960
S P Y S I K S K W T V P F F K P G V P Y I Y K V L V L N P D G S P A S G V P I K
TCTCCATACAGTATCAAATCAAAATGGACTGTGCCATTTTTCAAGCCTGGTGTTCCTTACATTTACAAGGTTCTCGTGTTGAACCCTGATGGCTCCCCTGCTTCTGGAGTCCCTATAAAG 1080
V S F S F D S S G N W I T Q K R K T M D N G I A M Q T I N T A R N S K K L N I K
GTGTCATTCTCATTTGATTCAAGTGGCAATTGGATCACACAAAAGAGGAAAACTATGGACAATGGCATAGCCATGCAAACTATCAACACAGCTAGAAACAGCAAAAAACTGAACATCAAA 1200
V Q T E D E R L E Q S Q Q A E A S F T I A S Y S S P S G S F I H L N A H R E V K
GTCCAGACTGAAGATGAACGGCTTGAACAGTCCCAACAGGCGGAGGCATCATTTACTATCGCATCATATTCCTCCCCAAGTGGAAGTTTTATTCACCTGAATGCCCATCGTGAAGTAAAA 1320
S P G E H I V F D V F I K S A A K D H V L H F N Y L M I S N G K I H N F L Q E G
TCTCCTGGTGAACACATCGTTTTCGACGTATTCATCAAGAGCGCCGCCAAAGACCATGTTTTACACTTTAACTACTTGATGATCAGCAATGGAAAGATACACAACTTTCTGCAGGAGGGT 1440
R K G D T T S V S L L L T P E L V P Q F R L V A F F I L P S G E L V A D S I I I
CGCAAAGGCGACACCACTTCAGTTTCCCTGTTATTGACGCCTGAACTTGTCCCACAATTTCGTTTGGTTGCCTTCTTCATTCTACCCTCTGGCGAGCTTGTCGCAGATTCCATTATTATT 1560
D V K D S C H A K L S L D V A G G K R L F S P R D N V N F D L S G E S D S W V A
GATGTAAAGGATTCCTGCCATGCAAAGCTCTCACTGGATGTCGCAGGAGGAAAGAGGCTTTTTTCACCAAGAGATAATGTAAACTTTGATTTGTCTGGAGAATCAGATTCTTGGGTGGCT 1680
V G V V D K A A Y V L D K K N K L T A N K V Y K A M E A S D L G C S V G S G K T
GTGGGAGTCGTGGACAAAGCTGCTTATGTTCTTGACAAGAAAAATAAACTTACAGCCAACAAGGTGTACAAGGCAATGGAAGCATCAGATTTAGGCTGTTCAGTTGGTAGCGGCAAAACG 1800
β-chain
G P L V F R D A G L A I M A K E I S G M D D V K D P G C P N G H T R R K R E L V
GGTCCACTTGTGTTCCGGGATGCTGGCCTTGCCATTATGGCGAAAGAAATTTCTGGAATGGATGATGTTAAAGATCCAGGATGCCCAAATGGACATACACGTCGCAAACGAGAGTTGGTT 1920
α-chain
L E I A I E K A S T Y P A E L R K C C R D A A I E S P L R L S C E E R T K H I H
CTGGAGATTGCCATCGAGAAAGCTTCAACATACCCAGCTGAACTGCGGAAATGTTGCCGAGATGCTGCAATCGAGAGCCCTTTAAGATTGAGCTGTGAAGAAAGAACAAAGCATATCCAC 2040
D E G E G C Q E T F L E C C K H V E E E L L I A M E E E D E D L G R S Q G E D F
GACGAAGGTGAGGGATGCCAGGAAACATTCTTGGAGTGCTGCAAGCATGTTGAGGAAGAGTTACTGATAGCTATGGAAGAAGAAGATGAGGATCTTGGGCGAAGCCAGGGAGAAGACTTC 2160
M I Q E S Q V V I R S H F P E S F M W E I I K L S R S A E N G K S R I T K K M P
ATGATTCAAGAATCTCAGGTCGTTATACGCTCCCATTTTCCTGAGAGCTTCATGTGGGAGATCATCAAATTGAGCAGGTCAGCCGAGAATGGCAAATCTCGAATTACAAAAAAAATGCCC 2280
D S I T T W D I Q A V E V S Q S K G L C V G P S L E L T V F K Q F F L K V H T P
GATTCGATCACGACTTGGGACATTCAGGCCGTTGAAGTCTCTCAATCTAAAGGACTTTGTGTGGGGCCTTCGCTGGAGCTGACTGTGTTTAAGCAATTTTTCCTCAAGGTGCACACACCG 2400
Y A L K Q Y E Q V E L R V V I Y N Y M N Q D V K G E I Q V K C G D G I C T D A E
TATGCATTAAAGCAATACGAACAAGTGGAGCTCAGAGTTGTGATCTACAACTACATGAACCAAGATGTGAAGGGTGAGATTCAGGTGAAATGTGGAGATGGCATTTGCACTGATGCAGAA 2520
Q N E P L K S R F A V E K N S A T S F S F M V V P L S S S D S S V S V L A R V F
CAAAATGAGCCTCTAAAGTCTCGTTTCGCAGTGGAGAAAAACTCTGCCACCTCATTTTCTTTCATGGTCGTGCCTTTGTCTTCAAGTGACTCTTCCGTCTCTGTCTTGGCTCGAGTATTT 2640
G S D V H D A V E K D L R V M P E G N Y E E M S R S W S V Q P R R H G G Q Q V I
GGCTCTGATGTACACGACGCTGTGGAGAAGGACTTGAGAGTAATGCCTGAAGGGAATTATGAAGAAATGTCTCGCAGTTGGTCTGTTCAACCAAGGCGTCATGGTGGACAGCAAGTGATC 2760
V V D N E T P Q N V V P G T E M S A F L S A Q G N L V A E T I Q N T L K G S K I
GTCGTTGATAATGAGACGCCACAAAATGTCGTTCCTGGAACTGAGATGTCAGCCTTCCTAAGTGCTCAAGGAAATCTTGTTGCAGAGACAATTCAGAATACACTAAAAGGATCCAAAATT 2880
thioester site
S N L L R L P R G C G E Q N M M Y T S I T V M V A R Y L N R S D Q W N K M G D P
TCAAATCTTCTTAGATTGCCGAGGGGTTGTGGAGAGCAAAATATGATGTATACATCTATTACGGTCATGGTCGCAAGGTATCTGAACCGAAGTGATCAGTGGAATAAAATGGGTGATCCT 3000
Q L K K R S F D F I T S G F A S Q L T Y R K P D Y S Y A A W L H R A S S T W L T
CAATTAAAGAAGCGTTCTTTTGATTTCATCACATCAGGGTTCGCCTCACAGCTGACTTACCGAAAGCCTGATTATTCATACGCAGCTTGGCTCCACCGAGCTAGCAGCACGTGGCTTACA 3120
A F V A K V F S Q A R Q L V F I P V S E I C G S V R W L M R K Q D K D G S F L E
GCTTTTGTGGCAAAAGTATTCTCACAGGCAAGGCAGCTGGTTTTTATTCCGGTTTCTGAAATCTGCGGCTCAGTACGTTGGCTTATGAGAAAGCAAGATAAAGACGGTTCATTCCTTGAA 3240
S K P V V H L N M M G Q V T G K V V L T S F V F I A L L E A R E S C I N E V E G
AGTAAACCAGTCGTCCACCTCAACATGATGGGTCAGGTGACTGGCAAAGTGGTACTTACTTCATTTGTTTTTATCGCTCTGCTTGAAGCTCGAGAATCCTGCATCAACGAGGTTGAAGGA 3360
F T V V V E K A H G Y L T S Q A M N G L E D F P L A I T A Y A L S L W K V S D G
TTCACAGTTGTAGTGGAAAAGGCACATGGATATCTAACAAGCCAAGCAATGAATGGTCTGGAAGATTTTCCACTGGCCATCACTGCATATGCCCTTAGTTTGTGGAAAGTCAGTGATGGA 3480

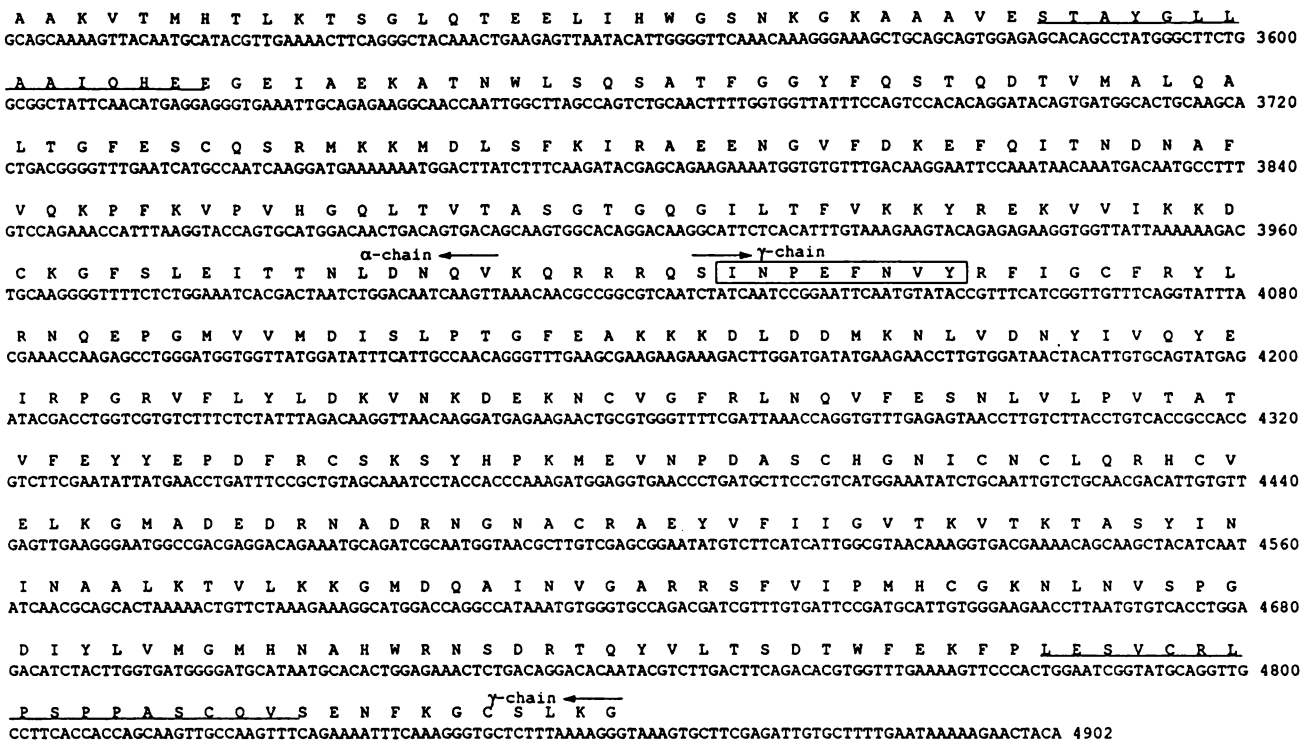


Fig. 5. Total nucleotide and amino acid sequences of the hagfish complement. The isolated hagfish complement consists of three polypeptide, β, α and γ chains. Seventy-seven amino acid residues at the N-terminus of the α chain were missing from the isolated hagfish complement. The five peptides listed in Figure 1 are boxed. Three peptides, presumably corresponding to the sequences reported by Varner *et al.* (1991), are underlined. The possible thioester site is underlined by a broken line.

and sequenced. In the next PCR, primer D, indicated in Figure 1, and primer E, complementary to the RNA sequence that encodes the amino acid sequence CSVGSGKT in peptide V, provided information about the downstream sequence. In this case, six nucleotides at the 3′ end of the sequence of primer E coincidentally corresponded to the amino acid sequence of LLGQ with reading out-of-frame and a 493 bp long fragment of DNA was amplified. After the N-terminal sequence of 290 amino acids had been determined, we noticed from comparisons of sequences that this gene does not encode a mammalian Ig-like molecule but encodes a complement-like molecule. We also noticed that peptide V, indicated in Figure 1, might correspond to the region with the amino acid sequence DLGCXXGXG which is located in close proximity to the C-terminal end of complement β chains and conserved in complements C3, C4 and C5. Since this hagfish gene seemed to encode a complement, we judged that all three polypeptides, of 88, 76 and 25 kDa, should be encoded by a single mRNA. We continued to implement the PCR strategy, using primer F and primer G which is complementary to the RNA sequence that encodes the N-terminus of the 76 kDa polypeptide. As we expected, a single band was observed at 1400 bp in the amplified products. By cloning of the material in this amplified band, the total nucleotide sequence was elucidated. In order to obtain information about sequences further downstream, we constructed a cDNA library from hagfish liver mRNA and screened it with the p4 fragment indicated in Figure 2 as probe. One positive clone, Hc16, had a 2.2 kb insert. Figure 5 shows the nucleotide sequence of the entire coding region of this hagfish gene, apart from the N-terminal signal sequence and two amino acids which should be encoded by another exon. Peptide III was located at amino acid residues 1344−1351. Thus, all five peptides listed in Figure 1 were encoded by this gene.

The N-terminal amino acid sequences of the H chains reported by Raison and coworkers (Hanley *et al.*, 1990), SKVLVIAPAAT and SQGEDFMIQET, were the same as those determined by us except for a difference of one amino acid (underlined). In the report by Varner *et al.* (1991), the N-termini of H and L chains were blocked. They described two partial amino acid sequences of H chains, STAYGNLAAIQPEE and LNPISSPLVV, and part of an L chain, LESIQNLPSPPASCQVS. The hagfish gene shown in Figure 5 also encoded these three peptides, with a few differences (underlined). These discrepancies may be due to differences between the species analyzed. *E.stoutii* (Raison and Litman) and *E.burgeri* (this report). Thus, this hagfish gene does actually encode the proteins so far isolated as hagfish 'antibodies' by three independent groups of investigators.

Discussion

Comparison of hagfish complement with mammalian complements C3, C4 and C5

In Figure 6, the amino acid sequence of the hagfish gene isolated in this study is compared with those of human C3 and C4, as well as murine C5 (De Bruijn *et al.*, 1985: Belt *et al.*, 1984; Wetsel *et al.*, 1987). Based on these alignments of sequences, the extent of sequence identity between the hagfish complement, human C3 and C4, and murine C5 was calculated (Table I). While the sequence of *Xenopus* C3 shows 49% amino acid identity to human C3 (Grossberger

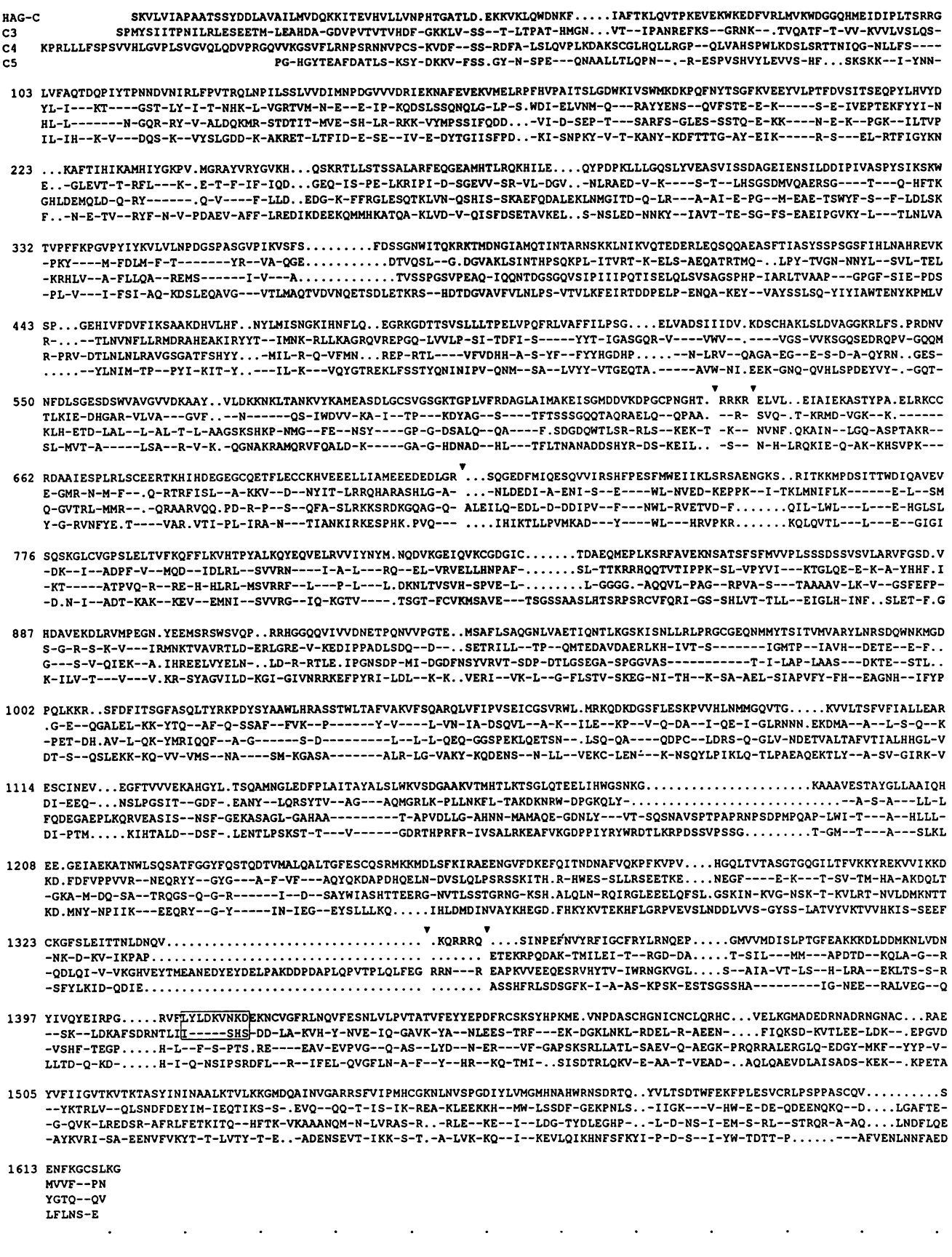


Fig. 6. Comparison of the amino acid sequence of the hagfish complement with those of mammalian complements, C3, C4 and C5. Human C3 and C4 sequences were taken from De Bruijn *et al.* (1985) and Belt *et al.* (1984), respectively. The mouse C5 sequence was taken from Wetsel *et al.* (1987). Bars indicate the same amino acids as those in the hagfish complement. Dots indicate missing amino acids. A possible properdine binding site is boxed (Grossberger *et al.*, 1989).

et al., 1989), the hagfish complement has 26–30% identity to mammalian C3, C4 and C5. Among these three proteins, C3 had the highest degree of sequence identity to the hagfish complement. The extent of the sequence identity between the hagfish complement and either C4 or C5 is roughly equal to that between C3 and either C4 or C5, respectively. In higher vertebrates after precursor molecules are synthesized, they are split into two polypeptides, β chains and α chains, by removal of four basic amino acid residues. In the hagfish complement, the sequence RRKR is observed at the corresponding region. The 77 residue fragments are removed from the N-terminal region of the α chains of C3, C4 and C5 and show anaphylatoxic activity (Hugli, 1986). The hagfish complement gene can encode a similar fragment and the N-terminus of the isolated hagfish complement lacks this region (Figures 1 and 5). This sequence of 77 amino acids contains six cysteine residues which are 100% conserved among all the complements. Mammalian C3 and C4 are different from C5 with respect to the presence of thioester structure in the α chains. This thioester in C3 and C4 is highly reactive and plays a key role in both the classical and alternative pathways of complement activation (Tack, 1983). Since the hagfish complement has an identical structure in the corresponding region, it is likely that the function of the hagfish complement is similar to that of either C3 or C4. The subunit structures differ between mammalian C3 and C4. In mammalian C4, a cluster of basic amino acids is observed between the α and γ chains. The hagfish complement contained an amino acid sequence similar to that of C4 in the corresponding region and, in fact, it was split to yield α and γ chains. Fujii *et al.* (1992) purified a complement composed of two subunits from hagfish and determined the N-terminal ten residue sequence of the β chain, which is identical to our results. The hagfish complement might be present in two different stable forms. Salient characteristics of C3 are its ability to interact with several plasma proteins, cell-surface receptors and foreign proteins (Becherer *et al.*, 1989). Among them properdine is noteworthy. Properdine is a plasma protein that led to the discovery of the alternative pathway (Pillemer *et al.*, 1956). The properdine binding site of human C3 was first localized to a peptide of 34 amino acids in the α chain of C3 (Daoudaki *et al.*, 1988) and was further localized to a nine residue sequence (Grossberger *et al.*, 1989). Figure 7 shows the comparison of the amino acid sequence of the hagfish complement with those of *Xenopus* and mammalian complement C3. The RGD sequence of human complement C3 has been reported to be involved in iC3b binding to CR3 (Wright *et al.*, 1987). However, the RGD sequence is absent in the hagfish complement as in the *Xenopus* complement C3. On the other hand, the properdine binding site corresponded to one of the most conserved regions among them. Thus, it is likely that the function of the hagfish complement is similar to that of mammalian C3 in the alternative pathway. The alternative pathway might be more ancient than the classical pathway (Farries and Atkinson, 1991) and may have been established at the cyclostome level of evolution. Although presence of another complement in cyclostome functionally equivalent to mammalian C4 has not been excluded, the degree of sequence homology of the hagfish complement with mammalian C4 is similar to that with mammalian C3 (Table I). It is possible that diversification into C3 and C4 did not occur at the cyclostome level. C5 could have already been established as an independent molecule at the cyclostome level, judged from its role in the complement cascade. Sequence homologies have been found not only among C3, C4 and C5, but also among these three proteins, α2-macroglobulin (Sottrup-Jensen *et al.*, 1985), α1-macroglobulin (Gordon, 1976), α1-inhibitor III (Braciak *et al.*, 1988) and pregnancy zone protein (Devriendt *et al.*, 1991). Evolutionary relationships between these proteins remain to be elucidated.

Table I. Amino acid sequence identities between hagfish complement, human C3 and C4, and mouse C5

	Hagfish complement	Human C3	Human C4	Mouse C5
Hagfish complement	–	494/1622	461/1622	424/1622
		30.5%	28.4%	26.1%
Human C3	494/1641	–	481/1641	429/1641
	30.1%		29.3%	26.1%
Human C4	461/1722	481/1722	–	408/1722
	26.8%	27.9%		23.7%
Mouse C5	424/1640	429/1640	408/1640	–
	25.9%	26.2%	24.9%	

Amino acid sequence identities between two species of complement were calculated from the sequence alignments in Figure 6. Since the lengths of polypeptides were different, the lengths in amino acids of the polypeptides in the left column were used as the denominators to calculate the percentage of identical residues for each comparison.

```
hagfish  GCFRYLRNQEPGMVVMDISLPTGFEAKKKDLDDMKNLVDNYIVQYEIRPG•••••RVFLYLDKVNKDEKNCVGFRLNQVFESNLVLPVTATVFEYYEPDFRCSKSYHPKM
Xenopus  A-A-H-K-VDAT-SII---MM---SPDTDS--RLMKG--K--SK--VNK-ANDKGTLI------SHIDEE--K-YAH-Y--VGFIQ-ASV--YD--T--N--T-F--VEE
mouse    I-TK--GDVDAT-SIL---MM---APDT---ELLASG--R--SK--MNKAFSNKNTLII--E-ISHT-ED-LT-KVH-Y-NVG-IQ-GSVK-YS--NLEES-TRF---EK
human    I-T--RGD-DAT-SIL---MM---APDTD--KQLA-G--R--SK--LDKAFSDRNTLII-----SHS-DD-LA-KVH-Y-NVE-IQ-GAVK-YA--NLEES-TRF---EK
```

Fig. 7. Comparison of the amino acid sequence of the hagfish complement with those of *Xenopus* and mammalian complements C3. Only the properdine binding site and its surrounding regions are shown. *Xenopus* C3 sequence were taken from Grossberger *et al.* (1989), and mouse and human C3 sequences were from Wetsel *et al.* (1984) and De Bruijn *et al.* (1985). Bars indicate the same amino acids as those in the hagfish complement. Dots indicate missing amino acids. The RGD region and the properdine binding site are marked.

Possibility of the presence of immunoglobulins in cyclostomes

In this study we clearly showed that the hagfish gene encoding the proteins isolated as hagfish 'antibodies' encodes the complement equivalent to mammalian C3. Do cyclostomes really have the ability to produce Abs? Abs should fulfill the following criteria: (i) production of Abs should be induced by immunization with a certain antigen; (ii) Abs should bind to the antigen specifically; (iii) a memory system should exist for the production of Abs. Many classical studies using lamprey and hagfish, carried out by several investigators, seem to support the presence of Abs in cyclostomes (Litman and Marchalonis, 1982). However, the presence of naturally occurring agglutinins, such as lectin, sometimes interfered with the detection of Ab-like activity (Linthicum and Hildemann, 1970) and, moreover, the affinities of Abs isolated from lamprey and hagfish for specific antigens were several orders of magnitude lower than those of Abs from higher vertebrates (Marchalonis and Edelman, 1968; Raison *et al.*, 1978). Although it is not clear whether all the physiological observations that apparently revealed the presence of Ab in cyclostomes were due to the complement system, there is no doubt that the hagfish gene isolated in this study is totally different from mammalian genes for Igs. Marchalonis and Schluter (1989) reported purification of lamprey antibody. However, since the subunit structure of the lamprey H chains consisting of two chains, 68 and 70 kDa, was similar to that of hagfish, it is possible that the lamprey antibody may also be a complement. On the other hand, several reports showed the presence of lymphocytes in cyclostomes (Hagen *et al.*, 1983; Fujii, 1982; Ardavin and Zapata, 1987), and for example, Hagen *et al.* (1983) have visualized them with specific Ab to Ig from O'RBC immunized lampreys. The fact that we found the message for the hagfish complement only in the liver would argue that it is not produced by lymphocytes, and therefore there still remains something to be looked for.

Recently, Hashimoto *et al.* (1990) devised a PCR strategy for the isolation of MHC genes from lower vertebrates. By utilizing two conserved blocks of sequence that surround cysteine residues in one Ig-fold domain, carp genes that encoded class I and IIβ MHC antigens were successfully isolated. We have applied this strategy to the isolation of Ig genes from lamprey and hagfish. However, extensive trials have given only negative results. Although it is very difficult to draw a definitive conclusion from negative results, there appears to be no convincing evidence at protein level for the presence of Abs in cyclostomes. Previous researchers seem to have assumed *a priori* that Ab of cyclostomes should be heterodimers consisting of H and L chains (Raison *et al.*, 1978; Kobayashi *et al.*, 1985; Varner *et al.*, 1991). The molecular nature of Abs in cyclostomes, if they exist, could be different from that in mammals. If TcR appeared earlier during evolution, and if Ig derived from the ancestral TcR, the ancient Igs might have been L chain heterodimers like Fab. If this is the case for Abs of cyclostome, it may explain why previous workers have overlooked them.

Materials and methods

Determination of polypeptide sequences

The hagfish protein previously described as an Ig by Kobayashi *et al.* (1985) was composed of three polypeptides, of 88, 76 and 25 kDa. They were separated by SDS–PAGE under reducing conditions, electroblotted onto a PVDF membrane, and their N-terminal sequences were determined directly by the published method (Matsudaira, 1987). A mixture of 88 and 76 kDa polypeptides was separated by HPLC on TSK gel G3000SW_{XL} and G4000 SW_{XL} columns, and digested with CNBr. The products were separated by reverse phase HPLC on a Synchropack RP-8 column. The amino acid sequence was determined on an Applied Biosystems 470A protein sequencer with an on-line 120A PTH analyzer.

Preparation of DNA and RNA

Hagfish, *E.burgeri*, were captured in the Sea of Japan. The genomic DNA was prepared from the liver by the published method (Gross-Bellard *et al.*, 1973). RNA was prepared from various tissues including liver, buffy coat, pronephros and intestine, by the method described by Chomczynski and Sacchi (1987). Polyadenylated RNA was further purified on an oligo(dT)–cellulose column (Maniatis *et al.*, 1982).

Cloning of the hagfish gene by PCR

Oligonucleotides were synthesized chemically on a DNA synthesizer (Applied Biosystems). PCR was carried out as described by Saiki *et al.* (1988). In brief, a 100 μl reaction mixture contained 1 μg of genomic DNA or a portion of cDNA prepared from hagfish liver RNA using COPYKIT (Invitrogen, CA), 2 μM primers, 200 μM dNTPs, 2.5 U *Taq* DNA polymerase (BIOTECH, WA) in 10 mM Tris–HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$ and gelatin at 200 μg/ml. The mixture was subjected to 35 cycles of amplification in a Perkin-Elmer/Cetus Thermocycler: 1 min at 94°C, 3 min at 55°C and 20 s–3 min at 72°C. After fractionation by electrophoresis on either acrylamide gel or agarose gel, the amplified products were cloned into Bluescript vector (Stratagene) and sequenced.

Preparation of DNA libraries

Hagfish genomic DNA was partially digested with *Sau*3AI, ligated with λdash vector and packaged into phages (Maniatis *et al.*, 1982). A cDNA library was constructed from hagfish liver mRNA by the method described by Gubler and Hoffman (1983). Screening of libraries was performed according to the method of Benton and Davis (1977).

Miscellaneous

Nucleotide sequences were determined by the chain termination method using Bluescript vectors. Southern and Northern hybridizations were carried out using the multi-prime-labeling system and nylon membranes (Maniatis *et al.*, 1982).

Acknowledgements

We are grateful to Dr L.Du Pasquier for critical reading of this manuscript and valuable suggestions, and Dr T.Fujii for information on the N-terminal amino acid sequence of the hagfish complement. We also thank Drs Y.Takagi, I.Ishiguro and K.Fujita for their encouragement, and Ms A.Nagata for preparation of the manuscript. This work was supported in part by grants from the Ministries of Education, Science and Culture; Health and Welfare; and Agriculture, Forestry and Fisheries in Japan, and by Fujita Health University.

References

Ardavin,C. and Zapata,A. (1987) *Dev. Comp. Immunol.*, **11**, 79–93.
Becherer,J.D., Alsenz,J. and Lambris,J.D. (1989) *Curr. Topics Microbiol. Immunol.*, **153**, 45–72.
Belt,K.T, Carroll,M.C. and Porter,R.R. (1984) *Cell*, **36**, 907–914.
Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.
Braciak,T.A., Northemann,W., Hudson,G.O., Shiels,B.R., Gehring,M.R. and Fey,G.H. (1988) *J. Biol. Chem.*, **263**, 3999–4012.
Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156–159.
Daoudaki,M.E., Becherer,J.D. and Lambris,J.D. (1988) *J. Immunol.*, **140**, 1577–1580.
Day,N.K.B., Gewurz,H., Johannsen,R., Finstad,J. and Good,R.A. (1970) *J. Exp. Med.*, **132**, 941–950.
De Bruijn,M.H.L. and Fey,G.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 708–712.
Devriendt,K., van den Berghe,H., Cassiman,J.J. and Marynen,P. (1991) *Biochim. Biophys. Acta*, **1088**, 95–103.
Du Pasquier,L. (1989) In Paul,W.E. (ed.), *Fundamental Immunology*. Raven Press, New York, pp. 139–165.
Farries,T.C. and Atkinson,J.P. (1991) *Immunol. Today*, **12**, 295–300.
Frank,M.M. and Fries,L.F. (1989) In Paul,W.E. (ed.), *Fundamental Immunology*. Raven Press, New York, pp. 679–701.

Fujii,T. (1982) *J. Morphol.*, **173**, 87–100.
Fujii,T., Nakamura,T., Sekizawa,A. and Tomonaga,S. (1992) *J. Immunol.*, in press.
Gordon,A.H. (1976) *Biochem. J.*, **159**, 643–650.
Gross-Bellard,M., Oudet,P. and Chambon,P. (1973) *Eur. J. Biochem.*, **36**, 32–38.
Grossberger,D., Marcuz,A., Du Pasquier,L. and Lambris,J.D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1323–1327.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Hagen,M., Filosa,M.F. and Youson,J.H. (1983) *Immunol. Lett.*, **6**, 87–92.
Hanley,P.J., Seppelt,I.M., Goeley,A.A., Hook,J.W. and Raison,R.L. (1990) *J. Immunol.*, **145**, 3823–3828.
Hashimoto,K., Nakanishi,T. and Kurosawa,Y. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6863–6867.
Hinds,K.R. and Litman,G.W. (1986) *Nature*, **320**, 546–549.
Hugli,T.E. (1986) *Complement*, **3**, 137–151.
Kobayashi,K., Tomonaga,S. and Hagiwara,K. (1985) *Mol. Immunol.*, **22**, 1091–1097.
Linthicum,D.S. and Hildemann,W.H. (1970) *J. Immunol.*, **105**, 912–918.
Litman,G.W. and Marchalonis,J.J. (1982) In Ruben,L.N. and Gershwin,H.E. (eds), *Immune Regulation*. Marcel Decker, New York, pp. 29–60.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marchalonis,J.J. and Edelman,G.M. (1968) *J. Exp. Med.*, **127**, 891–914.
Marchalonis,J.J. and Schluter,S.F. (1989) *Dev. Comp. Immunol.*, **13**, 285–301.
Matsudaira,P. (1987) *J. Biol. Chem.*, **262**, 10035–10038.
Muller-Eberhard,H.J. and Schreiber,R.D. (1980) *Adv. Immunol.*, **29**, 1–53.
Nonaka,M., Fukii,T., Kaidoh,T., Natsuume-Sakai,S., Nonaka,M., Yamaguchi,N. and Takahashi,M. (1984) *J. Immunol.*, **133**, 3242–3249.
Pillemer,L., Blum,L., Lepow,I.H., Wurz,L. and Todd,E.W. (1956) *J. Exp. Med.*, **103**, 1–13.
Raison,R.L., Hull,C.J. and Hildemann,W.H. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 5679–5682.
Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 482–491.
Sottrup-Jensen,L., Stepanik,T.M., Kristensen,T., Lønblad,P.B., Jones,C.M., Wierzbick,D.M., Magnusson,S., Domdey,H., Wetsel,R.A., Lundwall,A., Tack,B.F. and Fey,G.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 9–13.
Tack,B.F. (1983) *Springer Semin. Immunopathol.*, **6**, 259–282.
Varner,J., Neame,P. and Litman,G.W. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1746–1750.
Wetsel,R.A., Lundwall,A., Davidson,F., Gibson,T., Tack,B.F. and Fey,G.H. (1984) *J. Biol. Chem.*, **259**, 13857–13862.
Wetsel,R.A., Ogata,R.T. and Tack,B.F. (1987) *Biochemistry*, **26**, 737–743.
Williams,A.F. and Barclay,A.N. (1988) *Annu. Rev. Immunol.*, **6**, 381–405.
Wright,S.D., Reddy,P.A., Jong,M.T.C. and Erickson,B.W. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1965–1968.

Received on November 8, 1991

Note added in proof

The sequence data reported here have been submitted to the EMBL/GenBank/DDBJ nucleotide sequence databases and are available under the accession numbers Z11595 (cDNA sequence in Figure 5) and Z11596 (genomic sequence in Figure 3).