Human ficolin: cDNA cloning, demonstration of peripheral blood leucocytes as the major site of synthesis and assignment of the gene to chromosome 9

Jinhua LU,*‡ Puei Nam TAY,* Oi Lian KON* and Kenneth B. M. REID†

*Department of Biochemistry, Faculty of Medicine, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511 and †MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, U.K.

Pig ficolins and a number of other proteins contain sequences that are homologous to the C-terminal halves of fibrinogen β- and γ-chains. To clone the cDNA for human ficolin, two degenerate oligonucleotide primers were synthesized, based on two stretches of protein sequence that were highly conserved among those proteins, and used for PCR with cDNA from a human uterus λgt11 library as a template. A PCR product with a predicted size of 300 bp was obtained and this was used to screen a uterus cDNA library. Of the positive clones isolated, two (U1 and U2), containing inserts of 1.7 and 1.1 kb respectively, were found to encode human ficolin. The cDNA-derived amino acid sequence of human ficolin has approx. 75 % identity with, and a similar domain organization to, the two pig ficolin sequences, which are characterized by the presence of a leader peptide, a short N-terminal segment followed by a collagen-like region and then by a C-terminal fibrinogen-like domain. The 1.1 kb insert of clone U2 was used in Northern-blot analysis, and a very strong signal for a 1.4 kb mRNA species was detected in mRNA from human peripheral blood leucocytes. This showed that, despite the initial characterization of pig ficolin as a putative receptor on uterine cells for transforming growth factor β1, blood leucocytes are probably the major site of human ficolin synthesis. Much weaker signals of the same size were also detected in spleen, lung and thymus and may be due to the presence of tissue macrophages or trapped blood in these tissues. An mRNA species of approx. 1.3 kb in human liver also weakly hybridized to the U2 probe, indicating the presence of a sequence that was distinct from, but related to, ficolin. The gene for human ficolin has been mapped to chromosome 9.

INTRODUCTION

Gly-Xaa-Yaa repeating sequences are characteristic of collagen triple helices but they are also found in a number of other proteins, such as the complement protein C1q and the group of C-type lectins known as collectins. C1q is the recognition component of the classical pathway of complement and each of its three types of chain (A, B and C) consists of a short N-terminal segment followed by a Gly-Xaa-Yaa repeating sequence and then a C-terminal globular domain [1]. Formation of triple helices between the collagen-like regions of each of the six A-, six B- and six C-chains in C1q results in assembly of the 18 polypeptide chains into an overall 'tulip-like' structure, as viewed in the electron microscope [1–3]. C1q binds to the Fc regions of IgG and IgM in immune complexes via its C-terminal globular domains, and this binding can activate the $C1r_2C1s_2$ serine protease proenzyme complex, which is associated with the collagen-like stalks of C1q, and subsequently the C4 and C2 components of the complement cascade [1,4], resulting in the lysis and clearance of target micro-organisms. The collagen-like domains of C1q can also bind to the C1q receptor to mediate activation of some effector cell functions, such as phagocytosis and immunoglobulin and cytokine production [5–7].

The collectins represent a subfamily of the animal C-type lectins, which include mannan-binding protein, conglutinin, lung surfactant protein A, lung surfactant protein D and collectin 43 [8,9]. They display an overall structural similarity to C1q and appear to play important roles in innate immune defence and clearance, such as mannan-binding-protein-mediated complement activation, activation of macrophages and phagocytosis [9,10]. Most collectins (except surfactant-protein D) are also considered to bind to the C1q receptor [11,12]. A major structural difference between the collectins and C1q is that they contain globular C-type carbohydrate-recognition domains (CRDs) rather than the immunoglobulin-binding globular domains, found in C1q. This structural distinction is the basis of their difference in target recognition compared with the collectins which bind to carbohydrate structures rather than immune complexes.

Ichijo et al. [13] have characterized a putative receptor, for transforming growth factor β1 (TGF-β1), from the membrane fraction of pig uterus. cDNA cloning revealed two highly homologous sequences, named ficolin-α and ficolin-β, which contain collagen-like sequences and have similar overall domain organizations to C1q and the collectins [14]. However, the ficolins have C-terminal fibrinogen-like globular domains (defined as FBG modules [15]) which are quite distinct from the immunoglobulin-binding globular domains in C1q sequences and the C-type CRDs found in the collectins. Apart from the fibrinogen β- and γ-chains, FBG has been found in a number of other proteins including the tenascins, a human fibrinogen-related protein, a mouse cytotoxic T-lymphocyte-specific protein, a *Drosphila melanogaster* scabrous protein and a protein found in sea cucumber (see Figure 2a for references). The collagen-like Gly-Xaa-Xaa sequences of the pig ficolins probably form triple helices and thus arrange globular clusters of three FBGs at their C-termini. Therefore ficolin may, like C1q and the collectins, bind to certain targets via its globular 'heads' and subsequently mediate effector reactions via its collagen-like triple-helical stalks.

In this paper, we have characterized the cDNA sequence of

Abbreviations used: FBG, fibrinogen-like domain; CRD, carbohydrate-recognition domain; TGF-β1, transforming growth factor β1.

‡ To whom correspondence should be addressed.

human ficolin using the sequence homology of known FBGs. Our results indicate that there is a single human ficolin sequence represented in the uterus cDNA library which probably originates from blood leucocytes or tissue macrophages. We have detected a ficolin-related but distinct mRNA species in human liver. The human ficolin gene has also been assigned to chromosome 9. The demonstration that the major site of synthesis of human ficolin is in peripheral blood leucocytes contrasts with the previous characterization of pig ficolins as a putative uterine receptor for TGF-β1 [13,14] and is especially important for the functional characterization of this protein.

MATERIALS AND METHODS

Materials

A human λgt11 uterus cDNA library and two human multiple-tissue RNA blots were purchased from Clontech Laboratories (Palo Alto, CA, U.S.A.). A DNA panel derived from human/rodent somatic cell hybrids (mapping panel no. 2) was purchased from Coriell Cell Repositories.

PCR amplification of a human ficolin cDNA fragment probe

Two stretches of sequence were identified as being highly conserved among the known FBGs (see Figure 2a) spanning approx. 100 residues. Two degenerate oligonucleotide primers, HCFC-1 (sense, 5′-GAA/GTTT/CTGGT/CTIGGIAAT/CGA-3′) and HCFC-2 (antisense, 5′-TGA/GCAIT/C/GIA/GTTA/GTACCACCA-3′), were synthesized on the basis of two stretches of sequence as highlighted in Figure 2(a). The two primers were designed to include most of the residue variations of the aligned sequences of mammalian origin with the maximum exclusion of residues unique to tenascin sequences which have been detected in most human tissues. Inosine was introduced into positions of high degeneracy. PCR in mixtures (50 μl) containing 0.5 μM HCFC-1, 0.5 μM HCFC-2, 0.2 mM dNTP, approx. 10^6 phage (in 1 μl) from a human uterus cDNA library, 1.5 units of Dynazyme (Finnzymes, Espoo, Finland) and $1\times$PCR buffer containing 1.5 mM $MgCl_2$ was carried out for 35 cycles, each consisting of 30 s at 94 °C, 30 s at 45 °C, and 25 s at 72 °C. A 300 bp PCR product was detected and purified from a 2% (w/v) SeaPlaque low-melting agarose gel and used as a cDNA probe in subsequent library screening.

Screening of the human uterus cDNA library

Approx. 4×10^5 phage from the library were plated out using *Escherichia coli* Y1090 and then screened by standard methods [16] using the ^{32}P-labelled 300 bp PCR product. The inserts of the purified positive clones were subcloned into the pBluescript vector (pBSKS+) and subjected to automated sequencing.

Northern-blot analysis

Two human multiple-tissue Northern blots containing mRNA from human brain, colon, heart, kidney, liver, lung, ovary, pancreas, peripheral blood leucocytes, placenta, prostate, skeletal muscle, small intestine, spleen, testis and thymus were hybridized with the 1.1 kb insert of clone U2. The blots were washed as in the library screening with a final wash of 30 min at 65 °C in $0.1\times$SSC/0.1% (w/v) SDS (where $1\times$SSC is 0.15 M NaCl/0.015 M sodium citrate).

Chromosome localization

A panel of human/rodent hybrid somatic DNAs purified from rodent cell lines, each containing a single human chromosome, were used as templates in PCR amplification using the two primers, FC-3 (5′-GAACTTCTCAGCACAATTCG-3′) and FC-5 (5′-CCGTGTAGACCTGGTGGA-3′) corresponding to residues 261–267 and 197–203 respectively of the human ficolin sequence (Figure 1). Attempts were made to design the two primers with reference to the intron/exon structure of the human tenascin-X gene [17], to avoid intron/exon boundaries in the

```
                                                        GCTGGG-527
TCAATAGTACTTGCCGCAGTACTCTTAAAACTAGCGATGGAGGCTTCGACATGG-473
GCTTTAGGGAGTCATAAGTGGAGTCCGGAAAGAGGTATCTGTACTATAAAAGCT-419
ATTGTGTAAGCTAGTCATATTAAGTTGTTGGCTCAGGAGTTTGAAGTCCTTGAG-365
AGAGGATTATGATGCGACGTGAGTGCGTTCGTAATTTGAGTCTGCTTGGCAGAT-311
AGTAATGAGGATGTAAGCCCGTGGGCGATTATGAGAATGACTGCGCCGGTGAAG-257
CTTCAGGGGGTTTGGATGAGAATGGCTGTTACTACGAGGGCTATGTGGCTGATT-203
GAAGAGTATGCAATGAGCGATTTTAGGTCTGTTTGTCGTAGGCAGATGGAGCTT-149
GTTATAATTATGCCTCATAGGGATAGTACAAGGAAGGGGTAGGCTATGTGTTTT-95
GTCAGGGAGTTGAGAAACTGTGGCACAAGGCGAGAGCTGGTTTCCTCTGCCCTG-41
TTAGAGCTGGGGGACTCTTCAGAGTCAAAGGCCAGAGAGCATGGAGCTGAGTGG 14
                                                M  E  L  S  G  5

AGCCACCATGGCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAA 68
  A  T  M  A  R  G  L  A  V  L  L  V  L  F  L  H  I  K 23

GAACCTGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCT 122
  N  L  P  A  Q  A  A  D  T  C  P  E  V  K  V  V  G  L 41

GGAGGGCTCTGACAAGCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGC 176
  E  G  S  D  K  L  T  I  L  R  G  C  P  G  L  P  G  A 59

CCCAGGGCCAAAGGGAGAGGCAGGTGTCATTGGAGAGAGAGGAGAACGCGGTCT 230
  P  G  P  K  G  E  A  G  V  I  G  E  R  G  E  R  G  L 77

CCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCAAGGGAGACCGAGGAGA 284
  P  G  A  P  G  K  A  G  P  V  G  P  K  G  D  R  G  E 95

GAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTGCGAC 338
  K  G  M  R  G  E  K  G  D  A  G  Q  S  Q  S  C  A  T 113

AGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTG 392
  G  P  R  N  C  K  D  L  L  D  R  G  Y  F  L  S  G  W 131

GCACAATATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGA 446
  H  N  I  Y  L  P  D  C  R  P  L  T  V  L  C  D  M  D 149

CACGGACGGAGGGGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGA 500
  T  D  G  G  G  W  T  V  F  Q  R  R  M  D  G  S  V  D 167

CTTCTATCGGGACTGGGCCGCATACAAGCAGGGCTTCGGCAGTCAGCTGGGGGA 554
  F  Y  R  D  W  A  A  Y  K  Q  G  F  G  S  Q  L  G  E 185

GTTCTGGCTGGGGAACGACAACATCCACGCCCTGACTGCCCAGGGAAGCAGCGA 608
  F  W  L  G  N  D  N  I  H  A  L  T  A  Q  G  S  S  E 203

GCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCTAAGTACAA 662
  L  R  V  D  L  V  D  F  E  G  N  H  Q  F  A  K  Y  K 221

ATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTT 716
  S  F  K  V  A  D  E  A  E  K  Y  K  L  V  L  G  A  F 239

TGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTC 770
  V  G  G  S  A  G  N  S  L  T  G  H  N  N  N  F [F  S 257

CACCAAAGACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCA 824
  T  K  D  Q  D  N  D  V  S  S  S  N  C  A  E  K  F  Q 275

GGGAGCCTGGTGGTACGCCGACTGTCATGCTTCGAGCCTCAATGGTCTCTACCT 878
  G  A  W  W] Y  A  D  C  H  A  S  S  L  N  G  L  Y  L 293

CATGGGACCCCATGAGAGCTATGCCAATGGTATCAACTGGAGTGCGGCGAAGGG 932
  M  G  P  H  E  S  Y  A  N  G  I  N  W  S  A  A  K  G 311

GTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCCTAGACGGG 986
  Y  K  Y  S  Y  K  V  S  E  M  K  V  R  P  A  *        326
CCAGGACCCCTCCACATGCACCTGCTAGTGGGGAGGCCACACCCACAAGCGCTG 1040
CGTCGTGGAAGTCACCCCATTTCCCCAGCCAGACACACTCCCATGACGCCCACA 1094
GCTGCCCCTTTGCCCCCAGCTCAGTCAAGCCGTCACATGCCCACAACCTCACCA 1148
GAGGGAGAATTATGTTTCTAAATATGTTTACTTTTGGGGACAGAAAAAAAAAA 1202
AA                                                      1204
```

Figure 1 cDNA and derived amino acid sequence of human ficolin

The cDNA sequence was obtained from two human ficolin clones (U1 and U2), with the single-underlined sequences being found only in U1. The cDNA sequence is numbered from the 5′ end of clone U2, and the derived protein sequence, in single-letter code, is numbered from the first methionine residue. The first 27 residues, giving the sequence of the potential leader peptide, are indicated by italic letters. The potential asparagine-linked glycosylation site in the FBG is indicated by bold letters. The sequence that interrupts the continuity of the collagen-like region is double underlined. The putative Ca^{2+}-binding motif has been placed in brackets.

primer sequences. The primers were also designed to include a single intron in the PCR product as predicted from the human tenascin-X gene. Mixtures for PCR amplification (50 μl) contained 0.5 μM primer FC-3, 0.5 μM FC-5, 0.2 mM dNTP, 0.2 μg of either the hybrid DNAs or the parental human, mouse or Chinese hamster genomic DNAs or approx. 1000 EMBL-3 phage (in 1 μl) of the human ficolin genomic clone, 1.5 units of Dynazyme and 1 × PCR buffer containing 1.5 mM $MgCl_2$. Amplification was carried out for 35 cycles, each consisting of 30 s at 94 °C, 30 s at 55 °C, and 90 s at 72 °C. PCR products (10 μl) were separated on a 1 % (w/v) agarose gel containing 0.5 μg/ml ethidium bromide, and visualized under UV light.

RESULTS

Generation of a 300 bp human ficolin cDNA fragment

The sequence homologies of known FBGs are characterized by the complete conservation of 26 residues including four cysteines (Figure 2a), with 40 more residues being highly conserved. Two stretches of sequences were identified as being highly conserved among the aligned sequences of mammalian origin, and these were used to design two degenerate primers, HCFC-1 and HCFC-2, with the intention of including most of the residue variations among the considered sequences but to exclude amplification of human tenascins, which are extracellular matrix proteins found in most human tissues (Figure 2a). The inclusion of inosines at highly degenerate positions reduced the degeneracy of the two primers to mixtures of 16 and 24 respectively, which increased the specificity of the primers by narrowing the range of melting temperature of the primers (48–56 °C). PCR was performed for 25 s at the 72 °C extension step to minimize amplification of larger non-specific sequences, and a 300 bp fragment was the only product obtained (results not shown).

Characterization of human ficolin cDNA and derived amino acid sequences

Approx. 20 clones were identified from the human uterus library when the filters were washed at lower stringency (60 °C instead of 65 °C in the last wash). In the rescreening at a higher washing stringency (65 °C), two of the clones (U1 and U2) which had much stronger signals than the rest in the initial library screening were selected for further purification and characterization. Sequence analysis showed that the two positive clones, U1 and U2, contained a cDNA sequence that encoded an amino acid sequence consisting of an apparent leader peptide (27 residues) as predicted by the method of von Heijne [25], a short N-terminal segment (12 residues) followed by a Gly-Xaa-Yaa repeating sequence (69 residues) and then a C-terminal FBG (218 residues) (Figure 1). The derived protein sequence showed a higher degree of similarity (75 % identity) to the pig ficolins than to any other proteins containing FBGs. We therefore consider this sequence to represent human ficolin. The U1 clone contains the complete coding sequence for human ficolin and U2 covers virtually the entire translated region of ficolin mRNA except the first seven residues of the putative signal peptide (Figure 1). Clone U1 contains longer 5′- (546 bp) and 3′- (59 bp) untranslated regions (Figure 1). The cDNA sequence has a potential poly(A) tail of 13 nucleotides at its 3′ end, but no typical polyadenylation signal was found upstream (Figure 1).

Despite the apparent presence of a leader peptide in the derived sequence, the N-terminus of ficolin cannot be clearly defined, as no protein sequence has been generated from the N-termini of ficolin proteins from any species. Therefore the proposed N-terminus for human ficolin, as indicated in Figure 1, will have to be confirmed by protein sequencing. As found in C1q and most of the collectins, human ficolin has an interruption

(a)

```
                                                                  165
SCATGPRNCKDLLDRGYFLSGWHN-IY-LPDCRPLT--VLCDMDTDGGGWTVFQRRMDGS   HFCLN
QCLTGPRTCKELLTRGHILSGWHT-IY-LPDCQPLT--VLCDMDTDGGGWTVFQRRSDGS   PFCLNα
SCATGPRTCKELLTRGHFLSGWHT-IY-LPDCQPLT--VLCDMDTDGGGWTVFQRRSDGS   PFCLNβ
GSKRQYADCSEIFNDGYKLSGFYK-IKPLQSPAEFS--VYCDM-SDGGGWTVIQRRSDGS   HFREP1
LLYPFPKDCSQAMLNGDTTSGLYT-IY-LNGDKAQALEVFCDMTSDGGGWIVFLRRKNGR   HTNC
LRIPFPRDCGEEMQNGAGASRTST-IF-LNGNRERPLNVFCDMETDGGGWLVFQRRMDGQ   HTNX
 SPEYPRDCYDILQSCSGQSPPSGQYY-IQPDGGNLIKVYCDMETDEGGWTVFQRRIDGT   SCFP
IINKLPHDCSEVHTQTDGLHLIAPAGQRH----PLM--THC---TAD-GWTTVQRRFDGS   DROMESP
VQHLIYKDCSDHYVLGRRSSGAYR-VT--PDHRNSSFEVYCDMETMGGGWTVLQARLDGS   MCTLSP
 PVVSGKECEEIIRKGGETSEMYL-IQ--PDSSVKPYRVYCDMNTENGGWTVIQNRQDGS   HFIBGNβ
 HDITGKDCQDIANKGAKQSGLYF-IK--PLKANQQFLVYCEIDGSGNGWTVFQKRLDGS   HFIBGNγ
........C..*.........*...*............**C***...*GW.***.R.*G.

                                                                  211
VDFYRDWAAYKQGFGS-----------QL-GEFWLGNDNIHALTAQGSSE--LRVDLVDF   HFCLN
VDFYRDWAAYKRGFGS-----------QL-GEFWLGNDHIHALTAQGTNE--LRVDLVDF   PFCLNα
VDFYRDWAAYKRGFGS-----------QL-GEFWLGNDHIHALTAQGTSE--LRVDLVDF   PFCLNβ
ENFNRGWKDYENGFGN---------FVQKHGEYWLGNKNLHFLTTQEDYT--LKIDLADF   HFREP1
ENFYQNWKAYAAGFGD------------RREEFWLGLDNLNKITAQGQYE--LRVDLRD-   HTNC
TDFWRDWEDYAHGFGN------------ISGEFWLGNEALHSLTQAGDYS--IRVDLRA-   HTNX
INFYRSWSYYQTGFGN------------LNTEFWLGNDNIHYLTSQGDYE--LRVELNNT   SCFP
ADFNRSWADYAQGFGA------------PGGEFWIGNEQLHHLTLDNCSR--LQVQMQDI   DROMESP
TNFTREWKDYKAGFGN------------LEREFWLGNDKIHLLTKSKEMI--LRIDLEDF   MCTLSP
VDFGRKWDPYKQGFGNVATNTDGKNYCGLPGEYWLGNDKISQLTRMGPTE--LLIEMEDW   HFIBGNβ
VDFKKNWIQYKEGFGHLSPT-------GT-TEFWLGNEKIHLISTQSAIPYALRVELEDW   HFIBGNγ
.*F...W..Y..GFG.....................E*W*G**.*..**........*.***.*.

                                                                  258
FGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTG-------------HNNNFFST   HFCLN
EGNHQFAKYRSFQVADEAEKYMLVLGAFVEGNAGDSLTS-------------HNNSLFTT   PFCLNα
EGNHQFAKYRSFQVAGEAEKYKLVLGGFLEGNAGDSLSS-------------HRDQFFST   PFCLNβ
EKNSRYAQYKNFKVGDEKNFYELNIGEY-SGTAGDSLAGNFHPEVQWWAS--HQRMKFST   HFREP1
HGETAFAVYDKFSVGDAKTRYKLKVEGY-SGTAGDSMAY-------------HNGRSFST   HTNC
GDEAVFAQYDSFHVDSAAEYYRLHLEGY-HGTAGDSMSY-------------HSGSVFSA   HTNX
LGNHYYAKYNKFRIGDSFSEYLLVLGAY-SGTAGDSLAY-------------HNTMRFST   SCFP
YDNVWVAEYKRFYISSRADGYRLHIAEY-SGNASDALNY-------------QQGMQFSA   DROMESP
NGLTLYALYDQFYVANEFLKYRLHIGNY-NGTAGDALRFSRHYN--------HDLRFFTT   MCTLSP
KGDKVKAHYGGFTVQNEANKYQISVNKYRGTAGNALMDGASQLMGENRTMTIHNGMFFST   HFIBGNβ
NGRTSTADYAMFKVGPEADKYRLTYAYFAGGDAGDAFDGFDFGDD-------PSDKFFTS   HFIBGNγ
......A.Y..F.*.....Y.*.*..*..*.*.*.*.*.....................F..

                                                                  304
KDQDNDVS-SSN-------CAEKFQ---GAWWYADCHASSLNGLYLMGPHES-YANGI--   HFCLN
KDQDNDQY-ASN-------CAVLYQ---GAWWYNSCHVSNLNGRYLGGSHGS-FANGV--   PFCLNα
KDQDNDNH-SGN-------CAEQYH---GAWWYNACHSSNLNGRYLRGLHTS-YANGV--   PFCLNβ
WDRDHDNY-EGN-------CAEEDQS--G-WWFNRCHSANLNGVYYSGPYTAKTDNGI--   HFREP1
FDKDTDSA-ITN-------CALST-R--G-FWYRNCHRVNLMGRY--G--DNNMSQGV--   HTNC
RDRDPNSL-LIS-------CAVSY-R--GAWWYRNCHYANLNGLY--G--STVDHQGV--   HTNX
YDNDNDVY-SIN-------CASHSSYGRGAWWYKSCLLSNLNGQYYD--Y-S-GAPSI--   SCFP
IDDDRDIS-QTH-------CAANYE---GGWWFSHCQHANLNGRYNLG--LTWFDAAR--   DROMESP
PDRDNDRYPSGN-------CGL-Y-YSSG-WWFDSCLSANLNGKYYHQKYKG-VRNGIFW   MCTLSP
KDRDNDGWLTSDPRKQ---CSKED--GGG-WWYNRCHAANPNGRYYWGGQYTWDMAKHGT   HFIBGNβ
HNGMQFSTWDNDNDKFEGNCAEQD--GSG-WWMNKCHAGHLNGVYYQGGTYSKASTPNGY   HFIBGNγ
.*.*.*.............C........G.*W*..C..***G.Y..............

                          326
------N-WSAAKGYKYSYKVSEMKVRPA   HFCLN
------N-WSSGKGYNYSYKVSEMKFRAT   PFCLNα
------N-WRSGRGYNYSYQVSEMKVRLT   PFCLNβ
------V-WYTWHGWWYSLKSVVMKIRPN   HFREP1
------N-WFHWKGHEHSIQFAEMKLRPS   HTNC
------S-WYHWKGFEFSVPFTEMKLRPR   HTNX
------Y-WSYLPGDNDQIPFAEMKLRNR   SCFP
------NEWIAVKSSRMLVKRLPAVECQA   DROMESP
GTWPGIN---QAQPGGYKSSFKQAKMMIR   MCTLSP
DDGVVWMNW---KGSWYSMRKMSMKIRPF   HFIBGNβ
DNGIIWATW---KTRWYSMKKTTMKIIPF   HFIBGNγ
........*..............***...
```

(b)

```
                                                                  60
Human   MELSGATMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAP
Pig β   MELSRVAV*L*PTGQ*L***SFQT*A**********************S**********A
Pig α   MDTRGVAA*MRPL****A**---CTA*P*L****************S**********A

                                                                 108
Human   GPKGEAGVIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQS
Pig β   *******AN*PK****S**VV*****A**********A*****EP**L
Pig α   *******AS*PK*GQ*P*****EP**P**********EP*P**ESWET
```

Figure 2 Alignment of FBGs present in 11 different proteins (a) and the N-terminal halves of the three known ficolin sequences (b)

(**a**) The FBG of human ficolin (HFCLN) is aligned with those of pig ficolins (PFCLNα and PFCLNβ [14]), a human fibrinogen-related protein cloned from the liver (HFREP1 [18]), human tenascin-C (HTNC [19]), human tenascin-X (HTNX [17]), a sea cucumber fibrinogen-like protein (SCFP [20]), a ***Drosphila melanogaster*** scabrous protein (DROMESP [21]), a mouse cytotoxic T-lymphocyte-specific protein (MCTLSP [22]) and the human fibrinogen β- and γ-chains (HFIBGNβ and HFIBGNγ [23.24]). –, Spaces introduced to optimize the sequence identities. The two stretches of sequence on which the two degenerate PCR primers used in the initial cloning of the 300 bp ficolin fragment were designed are highlighted by single underlining (residues 185–191 of HCFC-1) and double underlining (residues 278–284 of HCFC-2) respectively. The 26 completely conserved residues are specified under the aligned sequences. The additional 46 highly conserved residues are emphasized by * under the aligned sequences. (**b**) Alignment of the N-terminal halves of human ficolin sequence (Human) with those of pig ficolin-β (Pig β) and -α (Pig α). Residues in pig ficolin-β and -α that are identical with the human ficolin sequence are denoted by *. The putative leader sequence of human ficolin is single-underlined. Numbering is based on the human ficolin sequence as seen in Figure 1.

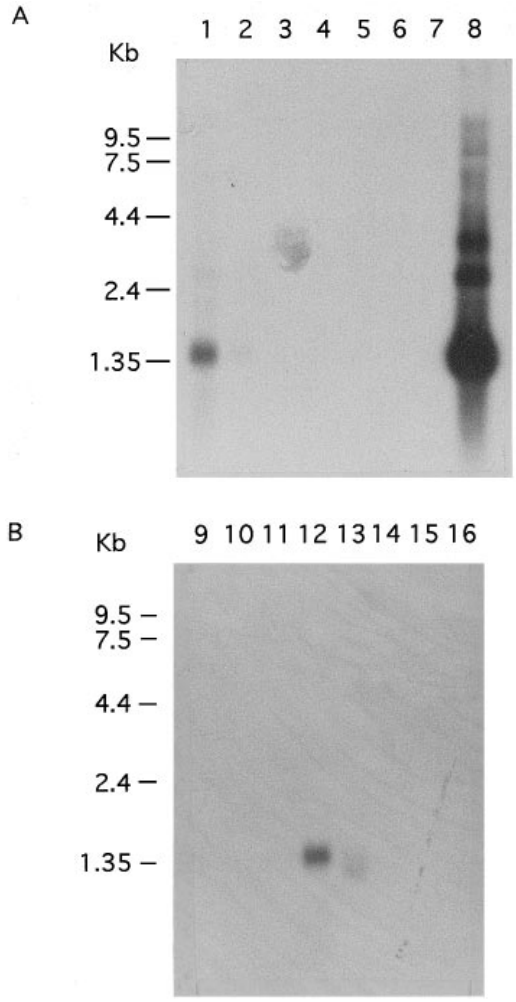

Figure 3 Northern-blot analysis of human ficolin mRNA

Two multiple human tissue Northern blots were probed using the 1.1 kb U2 insert. The mRNA samples on the blots were from the following human tissues: lane 1, spleen; lane 2, thymus; lane 3, prostate; lane 4, testis; lane 5, ovary; lane 6, small intestine; lane 7, colon; lane 8, peripheral blood leucocytes; lane 9, heart; lane 10, brain; lane 11, placenta; lane 12, lung; lane 13, liver; lane 14, skeletal muscle; lane 15, kidney; lane 16, pancreas. The blots were hybridized with the 1.1 kb human ficolin cDNA and exposed to X-ray film for 7 days at −85 °C. Abbreviation: Kb, kb.

in the collagen-like region and this interruption is found after the second Gly-Xaa-Yaa repeat where a sequence of KLTILR is inserted (Figure 1). The corresponding interrupting sequence present in the two pig ficolin sequences is KLSILR (Figure 2b). However, it is unlikely that a triple helix could be formed in ficolin via the two Gly-Xaa-Yaa triplets that precede the six-residue-long interruption. There is a cysteine residue in the predicted N-terminal segment of human ficolin and a cysteine residue was also found in the collagen-like region within the first Gly-Xaa-Yaa repeat (Gly-Cys-Pro) after the interrupting sequence. These two cysteines are conserved in the two pig ficolin sequences, and pig ficolin-α potentially has an additional cysteine residue close to the N-terminus (Figure 2b). These cysteine residues may be involved in interchain disulphide bond linkages and could play an important role in the assembly of the ficolin molecules.

The potential affinity of ficolin FGB for certain targets may require Ca^{2+}, as a putative Ca^{2+}-binding site is conserved in all ficolin sequences. In human ficolin it is found between residue 256 and 279 (Figure 1). A potential asparagine-linked glycosylation site was found in human ficolin at residue 305 which is conserved in the pig ficolin-α, but not the pig ficolin-β, sequence (Figure 2a). Human ficolin has a predicted molecular mass of 33 kDa, but a mature human ficolin polypeptide may have a higher molecular mass as the result of hydroxylation and O-linked glycosylation in the collagen-like region and/or N-linked glycosylation at Asn-305.

Tissue origins of human ficolin

Ichijo et al. [14] have shown that pig ficolin cDNAs hybridize to 1.4 kb RNA in human lung, placenta and skeletal muscle. In the present study, the 1.1 kb insert of clone U2 was labelled and used in Northern-blot analysis. The cDNA probe hybridized very strongly to a 1.4 kb mRNA species in peripheral blood leucocytes (Figure 3a). It also hybridized, very weakly, to similar-sized mRNAs in spleen, lung and thymus (Figure 3). This indicates that peripheral blood leucocytes or tissue macrophages are probably the major site of ficolin synthesis. No human uterus RNA was available on the two blots probed with ficolin cDNA, and therefore the possibility that uterus may also be a major site of ficolin synthesis cannot be ruled out. The lack of the 1.4 kb mRNA in human liver, a blood-rich organ, does not, however, support the view that there is a common cellular origin of human ficolin mRNA found in different tissues. The 1.3 kb mRNA detected in liver presumably encodes a protein closely related to ficolin. The Northern-blot results obtained in this paper are in strong contrast with the detection of mRNAs in human placenta and muscle using pig ficolin cDNAs [14]. This may be partially explained by the use of low-stringency washes in the previous study which may have allowed the ficolin probe to cross-hybridize. The isolation of human ficolin cDNA clones from a human uterus library suggests that human ficolin is expressed in the uterus but the cellular origin of the cloned DNA cannot be defined with certainty. A number of larger mRNA species in the leucocyte RNAs also hybridized to the 1.1 kb U2 insert. These signals were not detected in other tissues. The amounts of the different tissue mRNAs on the blots were very similar, as judged by the similar intensities of β-actin mRNA signals across the blots (results not shown), therefore the strong signals detected in peripheral blood leucocyte RNA for the 1.4 kb human ficolin mRNA and also other larger mRNAs are probably due to their high level of synthesis in the leucocytes. The larger forms of mRNA detected in peripheral blood leucocytes may represent sequences that are highly homologous to human ficolin, but they could also represent human ficolin mRNAs that were differentially polyadenylated.

Chromosomal localization of the human ficolin gene

The human/rodent somatic cell hybrid mapping panel used in the present study consists of DNA samples purified from 24 mouse or Chinese hamster cell lines each containing a single human chromosome. The intron/exon structure of the FBG in the human tenascin-X gene is different from those of the fibrinogen β- and γ-chains. As the FBGs in ficolins appear to be more closely related to human tenascins, as judged by overall sequence homology (Figure 2a), the known structure of the human tenascin-X gene was used to predict the intron/exon boundaries of the human ficolin gene encoding the FBG. Two

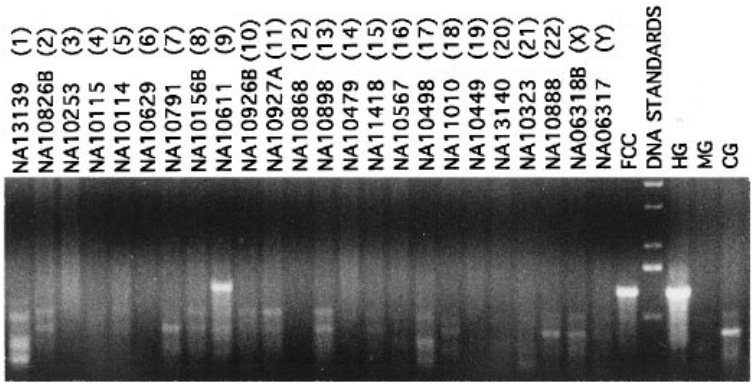

Figure 4 Chromosome localization of the human ficolin gene by PCR

PCR was carried out as described in the Materials and methods section using primers FC-3 and FC-5. The expected PCR product generated from the human ficolin gene is approx. 1.3 kb as indicated by PCR with parental human (HG), mouse (MG) and Chinese hamster (CG) genomic DNAs and a human ficolin genomic clone (FCC) isolated from an EMBL-3 genomic library (P. N. Tay and J. Lu, unpublished work). PCR was carried out with 24 hybrid DNA samples each isolated from a rodent cell line containing a single different human chromosome. The lanes containing PCR products amplified from hybrid DNAs are labelled with the names of the cell lines from which the genomic DNA was isolated followed by the number of the chromosome contained in the cell line in parentheses. PCR products were subjected to electrophoresis on a 1% (w/v) agarose gel containing ethidium bromide (0.5 μg/ml) and visualized under UV light. The DNA standard mixture gave a ladder of bands, 4.0, 3.0, 2.0 and 1.6 kb, running from top to the bottom.

primers, FC-3 and FC-5, were designed that would be expected to flank a single intron of 78 bp as in the tenascin-X gene. PCR on parental human genomic DNA amplified a DNA fragment of approx. 1.3 kb (Figure 4, lane HG) which was not observed in PCR where mouse and Chinese hamster genomic DNAs were used as templates (Figure 4, lanes MG and CG). Several genomic clones from human ficolin have been isolated from an EMBL-3 library using the 1.1 kb inserts of clone U2 (J. Lu, unpublished work). PCR carried out on one of the clones with FC-3 and FC-5, using the isolated phage as template, also yielded a 1.3 kb product (Figure 4, lane FCC). Therefore the 1.3 kb fragment clearly represents part of the human ficolin gene. When PCR was carried out using the hybrid DNA samples, a 1.3 kb product was predominantly amplified in the reaction with DNA purified from cell line NA10611 which contained the human chromosome 9, but no 1.3 kb product was seen in PCR with other hybrid DNAs (Figure 4). This clearly showed that the human ficolin gene is located on chromosome 9.

DISCUSSION

Although the two human ficolin clones isolated from the uterus cDNA library code for identical protein sequences, clone U1 is approx. 500 bp longer than clone U2 at the 5′ end. The U1 insert size of 1.7 kb is in contrast with the result of Northern-blot analysis which showed that human ficolin mRNA is approx. 1.4 kb long as detected in peripheral blood leucocytes, spleen, lung and thymus (Figure 3). Sequences have been obtained from a human ficolin genomic clone (P. N. Tay and J. Lu, unpublished work) and sequence identity between the genomic and cDNA sequences of ficolin stops at −92 in the cDNA sequence (Figure 1). Since, in the genomic sequence, it is preceded by a sequence (5′-……GGGTGTCCC-3′) that shows no homology to the consensus 3′-sequence of introns (5′-……AG-3′), most of the 'extra' 5′-sequence in U1 (−532 to −93) probably represents a cloning artifact unrelated to the human ficolin gene (Figure 1).

Despite the characterization of two ficolin sequences from pig uterus, only one human ficolin sequence has been found in the human uterus cDNA library. The 1.3 kb mRNA detected in human liver probably encodes a protein that is distinct from, but shows a high degree of sequence similarity to, human ficolin. The suggestion that there are two species of pig ficolin was based on Northern-blot studies which indicated that mRNA of pig ficolin-α was present in the lung and placenta whereas that of ficolin-β was found in the skeletal muscle [14]. In the current study, we were unable to detect human ficolin mRNA expression in placenta and muscle under conditions in which the cDNA probe hybridized quite clearly to other tissue RNAs. The isolation of a single human ficolin sequence from the uterus does not completely rule out the possibility that a second human ficolin sequence is present there. On the other hand, the characterization of two pig ficolin sequences in the uterus does not necessarily indicate that there are two types of pig ficolin molecule, as it is possible that these two sequences may represent two highly homologous subunits of a single type of pig ficolin molecule. This situation is seen in C1q which is composed of three slightly different polypeptide chains [1]. Human surfactant protein A also contains at least two slightly different polypeptides, whereas surfactant protein A molecules from other species are composed of a single type of polypeptide chain [26]. Therefore this may also be the case for ficolin: human ficolin consists of a single type of polypeptide chain whereas pig ficolin is assembled from two ficolin sequences, ficolin-α and ficolin-β.

A number of genomic clones have been isolated by using human ficolin cDNA (U2), and, in addition to the human ficolin gene, a ficolin-related gene has been identified which may correspond to the 1.3 kb mRNA detected in the liver (J. Lu and P. N. Tay, unpublished work). All the human collectin genes characterized so far have been localized to chromosome 10q21-23 [27], which indicates the close evolutionary origin of this group of human lectins. The genes for the three highly homologous C1q sequences are located on chromosome 1 on a 10 kb genomic DNA fragment [28]. The finding of the human ficolin gene on chromosome 9 excludes its linkage to the collectin or C1q genes.

FBG has been found in many proteins and is often located at the C-terminus. The other regions of these proteins are diverse in sequence and size. This situation is very similar to that seen with the C-type CRD which is found in proteins with diverse structures and functions [29]. The FBG has also been shown to be a

common module conserved early in evolution. It is not only found in different proteins of human and other animal origin, but also in lower animals such as sea cucumber and fruit fly [20,21]. The domain consists of approx. 220–250 residues and is characterized by the complete conservation of 26 residues, which are mostly hydrophobic, including four cysteine residues. At least 40 more residues are highly conserved. By analogy to C1q and the collectins, the FBG in ficolin probably functions as a primary ligand-binding site of the molecule. The recognition of certain targets by ficolin, via FBGs, may trigger important effector reactions via other domains of the protein, e.g. the collagen-like domains. The FBGs may have similar functions in other FBG-bearing proteins. Elucidation of the structures and binding specificities of the various FBGs will greatly advance our understanding of the properties of proteins containing this type of domain.

The pig ficolin sequences were isolated as a putative receptor for TGF-β1. Its original purification from the membrane fraction of pig uterus also indicated its association with membrane structures [13]. The cDNA-derived amino acid sequences for pig and human ficolins, however, do not predict the presence of typical transmembrane domains. As proteins with similar structures to the ficolins (C1q and the collectins) are soluble secretory proteins, ficolins may also be secreted into the circulation and perhaps also other body fluids.

We thank Dr. U. Holmskov of Odense University, Odense, Denmark and Dr. R. B. Sim of the University of Oxford for advice. This project was supported by the National University of Singapore grant no. RP940311. J. L. is a recipient of the Lee Kuan Yew Fellowship.

REFERENCES

1 Reid, K. B. M. (1983) Biochem. Soc. Trans. **11**, 1–12
2 Knobel, H. R., Illiger, W. and Isliker, H. (1975) Eur. J. Immunol. **5**, 78–82
3 Brodsky-Doyle, B., Leonard, K. R. and Reid, K. B. M. (1976) Biochem. J. **159**, 279–286
4 Duncan, A. R. and Winter, G. (1988) Nature (London) **332**, 738–740
5 Ghebrehiwet, B. and Peerschke, E. I. B. (1993) Behring Inst. Mitt. **93**, 236–240
6 Tenner, A. J. (1993) Behring Inst. Mitt. **93**, 241–253
7 Malhotra, R. (1993) Behring Inst. Mitt. **93**, 254–261
8 Holmskov, U., Malhotra, R., Sim, R. B. and Jensenius, J.-C. (1994) Immunol. Today **15**, 67–74
9 Lu, J., Wiedmann, H., Timpl, R. and Reid, K. B. M. (1993) Behring inst. Mitt. **93**, 6–16
10 Lu, J. and Sim, R. B. (1994) in New Aspects of Complement Structure and Function (Erdei, A., ed.), pp. 85–106, R. G. Lands Co., Austin, TX
11 Malhotra, R., Thiel, S., Reid, K. B. M. and Sim, R. B. (1990) J. Exp. Med. **172**, 955–959
12 Miyamura, K., Leigh, L. E. A., Lu, J., Hopkin, J., Bernal, A. and Reid, K. B. M. (1994) Biochem. J. **300**, 237–242
13 Ichijo, H., Ronnstrand, L., Miyagawa, K., Ohashi, H., Heldin, C.-H. and Miyazono, K. (1991) J. Biol. Chem. **266**, 22459–22464
14 Ichijo, H., Helman, U., Wernstedt, C., Gonez, L. J., Claesson-Welsh, L., Heldin, C-H. and Miyazono, K. (1993) J. Biol. Chem. **268**, 14505–14513
15 Bork, P. and Bairoch, A. (1995) Trends Biochem. Sci. **20**, Appendix
16 Lu, J., Willis, A. C. and Reid, K. B. M. (1992) Biochem. J. **284**, 795–802
17 Bristow, J., Tee, M. K., Gitelman, S. E., Mellon, S. H. and Miller, W. L. (1994) J. Cell Biol. **122**, 265–278
18 Yamamoto, T., Gotoh, M., Sasaki, H., Terada, M., Kitajima, M. and Hirohashi, S. (1993) Biochem. Biophys. Res. Commun. **193**, 681–687
19 Siri, A., Carnemolla, B., Saginati, M., Leprini, A., Casari, G., Baralle, F. and Zardi, L. (1991) Nucleic Acids Res. **19**, 525–531
20 Xu, X. and Doolittle, R. F. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 2097–2101
21 Baker, N. E., Moldzik, M. and Rubin, G. M. (1990) Science **250**, 1370–1377
22 Koyama, T., Hall, L. R., Haser, W. G., Tonegawa, S. and Saito, H. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 1609–1613
23 Chung, D. W., Harris, J. E. and Davie, E. W. (1990) Adv. Exp. Med. Biol. **281**, 39–48
24 Rixon, M. W., Chung, D. W. and Davie, E. W. (1985) Biochemistry **24**, 2077–2086
25 von Heijne, G. (1988) Biochim. Biophys. Acta **947**, 307–333
26 Voss, T., Melchers, K., Scheirle, G. and Schafer, K. P. (1991) Am. J. Respir. Cell Mol. Biol. **4**, 88–94
27 Kolble, K., Lu, J., Mole, S. E., Kaluz, S. and Reid, K. B. M. (1993) Genomics **17**, 294–298
28 Sellar, G. C., Blake, D. J. and Reid, K. B. M. (1991) Biochem. J. **274**, 481–490
29 Bezouska, K., Crichlow, G. V., Rose, J. M., Taylor, M. E. and Drickamer, K. (1991) J. Biol. Chem. **266**, 11604–11609

Received 10 May 1995/23 August 1995; accepted 13 September 1995