The complete amino acid sequence of human complement factor H

Jean RIPOCHE,* Anthony J. DAY,* Timothy J. R. HARRIS† and Robert B. SIM*

*MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and †Department of Molecular Biology, Celltech Ltd., 244–250 Bath Road, Slough SL1 4DY, Berks., U.K.

The complete amino acid sequence of the human complement system regulatory protein, factor H, has been derived from sequencing three overlapping cDNA clones. The sequence consists of 1213 amino acids arranged in 20 homologous units, each about 60 amino acids long, and an 18-residue leader sequence. The 60-amino-acid-long repetitive units are homologous with those found in a large number of other complement and non-complement proteins. Two basic C-terminal residues, deduced from the cDNA sequence, are absent from factor H isolated from outdated plasma. A tyrosine/histidine polymorphism was observed within the seventh homologous repeat unit of factor H. This is likely to represent a difference between the two major allelic variants of factor H. The nature of the cDNA clones indicates that there is likely to be an alternative splicing mechanism, resulting in the formation of at least two species of factor H mRNA.

INTRODUCTION

Factor H is an abundant plasma glycoprotein of M_r 155000, estimated to contain about 9% carbohydrate (Whaley & Ruddy, 1976; Sim & DiScipio, 1982). It is likely to exist in two different conformational states (ϕ_1 and ϕ_2) that can be separated by hydrophobic chromatography (Ripoche *et al.*, 1984; J. Ripoche, A. Erdei, D. Gilbert, A. Al Salihi & M. Fontaine, unpublished work). There is no evidence to suggest that these forms differ in their primary structure. The structural gene for factor H has been mapped to chromosome 1 in man (E. Solomon, unpublished work) and in mouse (Kristensen & Tack, 1986). Five allelic variants of factor H (FH1–FH5), distinguishable at the protein level by charge differences, have been identified (Rodriguez de Cordoba & Rubinstein, 1987). Factor H also exists as a membrane-bound species, present on several human leukocyte-derived cell lines and on B lymphocytes (Schulz *et al.*, 1984; Malhotra & Sim, 1985; M. J. Demares, J. Ripoche, C. Davrinche, D. Gilbert & M. Fontaine, unpublished work). No significant difference in M_r has been detected between the soluble and membrane-bound forms, and the mechanism by which factor H remains membrane-associated is unknown.

The most clearly identified function of factor H is the regulation of enzymic processing of the major complement component C3 (Whaley & Ruddy, 1976; Weiler *et al.*, 1976). Factor H causes dissociation of the complement system enzyme C3bBb (the alternative-pathway C3 convertase) by binding to the C3b subunit and displacing Bb. This activity is known as ‘decay accelerating activity’. In binding to C3b, factor H renders C3b susceptible to proteolytic inactivation by factor I. This is known as the ‘factor I-cofactor activity’ of factor H. A number of other complement-system proteins, namely complement receptor type 1 (CR1), decay accelerating factor (DAF), membrane cofactor

protein (MCP) and C4b-binding protein (C4bp) also possess decay acceleration and cofactor activities. CR1, C4bp and DAF (Caras *et al.*, 1987) are known to be homologous with factor H (for reviews, see Holers *et al.*, 1985; Reid *et al.*, 1986; Sim *et al.*, 1987). The structural genes for CR1, C4bp and DAF are also located on human chromosome 1, and the genes for factor H, CR1 and C4bp form a closely linked cluster, termed the ‘regulation of complement activation (RCA) cluster’ (Rodriguez de Cordoba & Rubinstein, 1986).

Factor H also has other activities which are associated with binding to specific factor H receptors on leucocytes (Lambris & Ross, 1982; Erdei & Sim, 1987). These activities are not yet well defined, but a protein likely to correspond to the B-lymphoblastoid cell receptor for factor H has been isolated (Erdei & Sim, 1987). The conformational variant of factor H, ϕ_1 , does not appear to interact with cellular receptors, whereas ϕ_2 does (J. Ripoche, A. Erdei, D. Gilbert, A. Al Salihi & M. Fontaine, unpublished work).

The N-terminal amino acid sequence of human factor H has been determined by automated Edman degradation (Sim & DiScipio, 1982; Ripoche *et al.*, 1984; Kristensen *et al.*, 1986) and partial cDNA-derived sequences have been reported (Ripoche *et al.*, 1986; Kristensen *et al.*, 1986; Schulz *et al.*, 1986; Day *et al.*, 1987a). The complete sequence of mouse factor H has also been determined (Kristensen & Tack, 1986). The complete sequence of human factor H, established by characterization of three overlapping cDNA clones, is reported here.

MATERIALS AND METHODS

Materials

Restriction endonucleases and other enzymes for processing of DNA were purchased from Amersham

Abbreviations used: CR1(2), complement receptor type 1(2); C4bp, C4b-binding protein; MCP, membrane cofactor protein; DAF, decay accelerating factor; bp, base-pair(s); poly(A)⁺, polyadenylated; kb, kilobase(s).

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00716.

International (Aylesbury, Bucks., U.K.), New England Biolabs (Beverly, MA, U.S.A.) or Boehringer (Lewes, E. Sussex, U.K.). Radiolabelled compounds, M13 vectors and nitrocellulose (Hybond C) were from Amersham International. Human liver cDNA libraries (Belt *et al.*, 1984; Bentley, 1986) in plasmid pAT 153/*Pvu*II/8 were generously supplied by Dr. K. T. Belt, Dr. R. D. Campbell and Dr. D. R. Bentley of the MRC Immunology Unit, University of Oxford.

Protein sequencing

Factor H was isolated from outdated human plasma (Sim & DiScipio, 1982) pooled from 12 donors (kindly donated by the Regional Blood Transfusion Service, Oxford, U.K.). A final step using zinc-chelate affinity chromatography (Day & Sim, 1986) was done to remove minor contaminants. Factor H (170 nmol) was reduced under denaturing conditions, alkylated with iodo[2-³H]-acetic acid, succinylated and digested with trypsin, as described previously (Ripoche *et al.*, 1986). Tryptic peptides were separated into 17 pools by chromatography on Sephadex G-75 in 0.1 M-NH₄HCO₃. Material from these pools was then further purified by (i) ion exchange on DEAE-Sephacel in 10 mM-NH₄HCO₃, making use of linear gradient elution with 10 mM–1 M-NH₄HCO₃, or (ii) reverse-phase h.p.l.c. using the trifluoroacetic acid/acetonitrile or NH₄HCO₃/acetonitrile solvent systems described by Christie & Gagnon (1982). Peptides were subjected to amino acid analysis after 24 h hydrolysis in constant-boiling HCl. Peptides selected on the basis of their composition and apparent purity in chromatographic procedures were sequenced by automated Edman degradation in a Beckman 890C sequencer (Christie & Gagnon, 1982).

Selection of cDNA clones: general strategy

As described previously, the human liver cDNA library of Belt *et al.* (1984) was initially screened with a 17-base-long mixed oligonucleotide probe complementary to the mRNA for the amino acid sequence MFGDEE (Ripoche *et al.*, 1986). Two types of positive clone, R1a and R2a, were identified, and shown by DNA sequence analysis to be specific for factor H (Ripoche *et al.*, 1986). The insert of R2a was completely sequenced and shown to contain 1972 bp of coding sequence followed by a TAG stop codon and 157 bp of 3' untranslated sequence (Day *et al.*, 1987a). This clone provided the sequence of the C-terminal 653 amino acid residues of factor H (Day *et al.*, 1987a; present Figs. 1 and 2). R1a was contained within R2a. Rescreening of the same library using the R2a insert as a probe resulted in isolation of a further 16 distinct types of clone. None of these, however, was larger than R2a, and none extended beyond the 5' end. No clones providing a significant extension of R2a were detected by screening, with R2a, the liver cDNA library of Bentley (1986) or the Clontech liver cDNA library (type HL 1001a; Clontech, Palo Alto, CA, U.S.A.).

Two unique oligonucleotide probes were synthesized on the basis of a partial factor H cDNA sequence (Kristensen *et al.*, 1986) known to correspond to the N-terminal region of factor H. The two unique oligonucleotide probes were denoted '611' (sequence 5'-CATGATTTTCTTCACA) and '612' (5'-TGTTTCATCACAGTAATA). The cDNA library of Bentley (1986) (100 000 clones) was screened with each of these two probes, and 28 clones which reacted with one

or both probes were identified. The inserts of two such clones, B-15-1 and B-38-1, were characterized further and subjected to partial (B-15-1) or complete (B-38-1) sequencing. The sequences confirmed that B-38-1 corresponded to the N-terminal end of factor H (see Figs. 1 and 2) and that B-15-1 was contained within B-38-1. None of the 28 cDNA clones isolated using 611 and 612 as probes cross-hybridized to R2a, indicating that sequences in the middle of factor H were missing. A further 'primer extension' library was constructed (see below) in pAT 153/*Pvu*II/8 to provide cDNA clones bridging between B-38-1 and R2a. Three clones which hybridized with both R2a and B-38-1 in colony hybridization tests were isolated from this library. Two such clones were sequenced partially (PE-2) or completely (PE-3), allowing the complete factor H cDNA sequence to be established.

Library screening and isolation of cDNA clones

cDNA libraries in pAT 153/*Pvu*II/8 were plated on nitrocellulose and replica filters were made on nitrocellulose (Grosveld *et al.*, 1981). Filters were screened with ³²P-labelled oligonucleotide probes as summarized previously (Ripoche *et al.*, 1986).

Plasmid DNA was prepared from bacterial cultures (Birnbom & Doly, 1979), and cloned cDNA inserts excised by *Msp*I/*Cla*I (R2a) or *Bam*H1/*Cla*I (PE-2, PE-3, B-38-1, B-15-1) double-restriction digestion. Appropriate restriction fragments were separated by electrophoresis on low-gelling-temperature agarose (Sigma, Poole, Dorset, U.K.; Type VII) and eluted from the gel (Maniatis *et al.*, 1982). cDNA fragments used for further library screening were nick-translated (Rigby *et al.*, 1977) to a specific activity of $> 2 \times 10^8$ c.p.m./μg. Hybridization conditions for large cDNA probes were as summarized by Chung *et al.* (1985).

DNA sequence analysis

The complete or partial sequence of relevant cDNA clones was obtained by dideoxy sequencing (Sanger *et al.*, 1980) with deoxyadenosine 5'-[³⁵S]thiotriphosphate. Suitable restriction fragments were first generated at random, then from sites predicted by the derived DNA sequence. Fragments were either blunt-ended using nuclease S1 (Maniatis *et al.*, 1982) and subcloned into *Eco*RV-cut M13 tg 130 and tg 131 by using *E. coli* JM 101, or subcloned directly into compatible sites in M13 tg 130 and 131. Other fragments were filled in and subcloned into *Sma*I-cut M13 mp8 and mp9 by using *E. coli* Tg1. Sequence data were processed using the programs described by Staden (1982).

'Primer extension' library

To obtain cDNA clones which extended upstream of the 5' end of clone R2a, a primer for cDNA synthesis was prepared from R2a. The primer was a 196-bp fragment located at positions defined by nucleotides 1887–2082 (Fig. 2a). This is close to the 5' end of R2a (Fig. 1). The primer was made by digesting R2a with *Eco*RI under conditions (20 mM-NaCl/50 mM-Tris/HCl/1 mM-MgCl₂, pH 8.0) in which 'star' activity is observed (Gardner *et al.*, 1982). In this case, cleavage occurred at NAATC sites. Poly(A)⁺ mRNA from human liver was prepared (Chirgwin *et al.*, 1979), and 2 μg of this material was hybridized to 1 μg of primer for 1 h at 50 °C in 82% (v/v) formamide, 0.4 M-NaCl, 1 mM-EDTA, 0.1 M-Pipes,

Fig. 1. Overlapping cDNA clones

The relative positions of clones B-38-1, PE-3, and R2a are shown. Overall, they span 3926 bp. Below the rectangles representing the clones is shown a map of the restriction digests used in sequencing the clones. Short horizontal arrows show the extent and direction of DNA sequence determination of various restriction-digest fragments. Clone B-38-1 is inverted at the 5' end, and the point of inversion is shown by ●. B-38-1 also contains a 3' region that differs from the sequence of PE-3. The area where B-38-1 differs from PE-3 is shown hatched (▨) on the upper part of the diagram. On the restriction map, the site where the two sequences diverge is shown by ▼. A vertical broken line from this symbol leads downwards to the restriction map and sequencing diagram corresponding to the hatched area of B-38-1. Long horizontal arrows summarize that all sequence has been determined in both directions, apart from a segment close to the point of inversion of B-38-1. Symbols of restriction endonucleases are: a, *AclI*; b, *AclI*; c, *AvaII*; d, *BamHI*; e, *DdeI*; f, *DraI*; g, *EcoRI*; h, *HaeIII*; i, *HindIII*; j, *HinfI*; k, *HinfI*; m, *MspI*; n, *RsaI*; o, *XmnI*.

(a)

-10 -15 -10 -5
M R L L A K I I C L M L W A I C

AATTCTTGGGAAGGAGGAAGTGGACGTTGTGAACAGAGTTAGCTGTGTAATGTCTCTTAAAGATCCAAAAATGAGACTTCTAGCAAGATTATTTGGCTTATGTTATGGGCTATTGT
10 20 30 40 50 60 70 80 90 100 110 120

-1 1
V A E D C N E L P P R R N T E I L T G S W S D Q T Y P E G T Q A I Y K C R P G Y
TGTACCAAGAGTTGCAATGAACTTCTCCCAAGAAAGAAATACAGAAATTTCTGACAGGTTCTGGTCTGACCAACATATCCAGAAAGGCCACCGGCTATCTATAAATGCCGCCCTGGATA
130 140 150 160 170 180 190 200 210 220 230 240

R S L G N V I M V C R K G E W V A L N P L R K C Q K R P C G H P G D T P F G T F
TAGATCTCTTGGAAATGTAATAATGGTATGCAGGAAGGGAGAAATGGGTTGCTCTTAATCCATTAAAGGAAATGTCAGAAAGGCCCTGTGGACATCTCTGGAGATCTCTTTTGGTACTTT
250 260 270 280 290 300 310 320 330 340 350 360

T L T G G N V F E Y G V K A V Y T C N E G Y Q L L G E I N Y R E C D T D G W T N
TACCCCTACAGGAGGAAATGTTTGAATATGGTGTAAAGCTGTGTATACATGTAATGAGGGGTATCAATTGCTAGGTGAGATTAAATTACCGTGAAATGTGACACAGATGGATGGACCAA
370 380 390 400 410 420 430 440 450 460 470 480

D I P I C E V V K C L P V T A P E N G K I V S S A M E P D R E Y H F G Q A V R F
TGATATTCCTATATGTGAAGTTGTGAAGTTGTACCAAGTACAGCAACAGAGAAATGGAAAAATTTGTCAGTATGCAATGGAACCCAGATCGGGAAATACCAATTTGGCAAGCAGTACGGTT
490 500 510 520 530 540 550 560 570 580 590 600

V C N S G Y K I E G D E E M H C S D D G F W S K E K P K C V E I S C K S P D V I
TGTATGTAAGTCAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGATGGAAGGAT
610 620 630 640 650 660 670 680 690 700 710 720

N G S P I S Q K I I Y K E N E R F Q Y K C N M G Y E Y S E R G D A V C T E S G W
AAATGGATCTCTATATCTCAGAAATTAATTAAGGAAATGGAAGGATTTCAATATAAATGTAACATGGGTTATGAAATCAGTGAAGAGGAGATGCTGTATGCACTGAAATCTGGATG
730 740 750 760 770 780 790 800 810 820 830 840

R P L P S C E E K S C D N P Y I P N G D Y S P L R I K H R T G D E I T Y Q C R N
GGTCCGTTGCTTATGTAAGAAAAATCATGTGATAATCTCTATATTTCCAAATGGTGACTACTACCTTTAAGGATTAAACACAGAACTGGAGATGAAATCACCTACCAGTGTAGAAA
850 860 870 880 890 900 910 920 930 940 950 960

G F Y P A T R G N T A K C T S T G W I P A P R C T L K P C D Y P D I K H G G L Y
TGGTTTTTCTCTCAACCCCGGGGAAATACAGCCAAATGCACAACTACTGGCTGGATACCTGCTCCGAGATGATACCTTGAAACCTTGTGATTATCCAGACATTAACCAATGGAGGCTCTATA
970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080

H E N M R R P Y F P V A V G K Y Y S Y Y C D E H F E T P S G S Y W D H I H C T Q
TCATGAGAAATGCGTAGACCTACTTTCCAGTAGCTGTAGGAAAAATATTACTCTCTATGTAAGTGAACATTTTGAGACTCCGTCAGGAAGTTACTGGAGTACATTCATTCACACACA
1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200

D G W S P A V P C L R K C Y F P Y L E N G Y N Q N H G R K F V Q G K S I D V A C
AGATGGATGCTCCAGCAGTACCATGCTCAGAAATGTTATTTCTTATTTGGAAAAATGGATATAATCAAAATCATGGAAGAAAGTTTGTACAGGTTAAATCTATAGACGTTGCCCTG
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320

H P G Y A L P K A Q T T V T C M E N G W S P T P R C I R V K T C S K S S I D I E
CCATCTGCTTACGCTCTTCCAAAAAGCCAGACACAGTACATGTAATGGAGAAATGGCTGGTCTCTACTCCAGATGATCCGTTGCAAAACATGTTCCAAATCAAGTATAGATATGGA
1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440

N G F I S E S Q Y T Y A L K E K A K Y Q C K L G Y V T A D G E T S G S I R C G K
GAATGGGTTTATTTCTGAATCTCAGTATACATATGCCCTTAAAAAGAAAAAGGAAATATCAATGCAAACTAGGATATGTAACAGCAGATGGTGAAACATCAGGATCAATTAGATGTGGAA
1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560

D G W S A Q P T C I K S C D I P V F M N A R T K N D F T W F K L N D T L D Y E C
AGATGGATGGTCACTCAACCCAGCTGCAATTAATCTTGTGATATCCAGTATTTATGAAATGCCAGAACTAAAAATGCACTTCACATGGTTTAAOCTGAAAGACAAATGGAGTATGAAATG
1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680

H D G Y E S N T G S T T G S I V C G Y N G W S D L P I C Y E R E C E L P K I D V
CCATGATGGTTATGAAAGCAATACTGGAAGCACCAGTCTCCATAGTGTGGTTTACAATGGTTGGTCTGATTATACCCATATGTTATGAAAGAGAAATGCCAACTTCTTAAATAGATGT
1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

H L V P D R K K D Q Y K V G E V L K F S C K P G F T I V G P N S V Q C Y H F G L
ACACTTATGTTCTGATCCCAAGAAAGACCACTATAAGTGTGGAGAGGTGTGAAATTCCTCTGCAAAACAGGATTTACAATAGTTGGACCTAATTCGGTTCAGTGTACCACTTTGGATT
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920

S P D L P I C K E Q V Q S C G P P P E L L N G N V K E K T K E E Y G H S E V V E
GTCTCTGACCTCCCAATATGTAAGAGCAAGTACAATCAATGGTCCACCTCTGAACTCTCAATGGGAATGTTAAGGAAAAAACGAAAGAAATATGGACACAGTGAAGTGGTGA
1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040

Y Y C N P R F L M K G P N K I Q C V D G E W T T L P V C I V E E S T C G D I P E
ATATTATGCAATCTAGATTTCTAAAGGAGCACTAATAAAATTTCAATGTTGTGATGGAGAGTGGACAACTTTACCAAGTGTGATTGTGGAGGAGAGTACCTGTGGAGATATACCTGA
2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160

L E H G W A Q L S S P P Y Y Y G D S V E F N C S E S F T M I G H R S I T C I H G
ACTTGAACATGCTGGGCCAGCTTTCTTCCCTCTTATTACTATGGAGATTCAGTGGAAATTCATTTGCTCAGAACTATTACAAATGATTGATGGACAGATCAATTCAGTGTATTCATGG
2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280

V W T G Q L P Q C V A I D K L K K C K S N L I I L E S H L K N K K E F D H N S N
AGTATGACCCAACTTCCCGAGTGTGGCAATAGATAAACTTAAGAAAGTGAATATCAATTAATTTATCTTGAAGAACATTTAAAAACAAGAAAGGAATTCGATCATATTTCTAA
2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400

I R Y R C R G K E G W I H T V C I N G R W D P E V N C S M A Q I Q L C P P P P Q
CATAAAGTACAGATGTAAGAGAAAGAAAGGATGGATACACACAGTCTGCAATAAATGGAAGATGGGATCCAGAAAGTGAATGCTCAATGGCAAAATACAAATTAAGCCCACTCCCACTCA
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520

I P N S H N M T T T L N Y R D G E K V S V L C Q E N Y L I Q E G E E I T C K D G
GATTCGCAATTTCTCACAATATGCAACACCACTGAATATTCGGGATGGAGAAAAAGTATCTGTTCTTTGCCAAGAAAAATTTCTAAATTCAGGAAGGAGAAAGAAATTTACATGCAAGATGG
2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640

R W Q S I P L C V E K I P C S Q P P Q I E H G T I N S S R S S Q E S Y A H G T K
AAGATGCGAGTCAATACCACTCTGTGTTGAAAAATTTCCATGTTTCAACACCACTTCAGATAGAACACGGAAGCAATTAATTCATCCAGTCTTCACAGAAAGTATGACATGCGGACTAA
2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760

L S Y T C E G G F R I S E N E T T C Y M G K W S S P P Q C E G L P C K S P P E
ATTGAAGTTACTTGTGAAGGTTGTTTCCAGGATATCTGAAGAAAAATGAACCAACATGCTACATGGGAAATGGAGTTCTCCACCTCAGTGTGAAGGCCCTTCTTGTAAATCTCCACCTGA
2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880

I S H G V V A H M S D S Y Q Y G E E V T Y K C F E G F G I D G P A I A K C L G E
GATTTCTATGGTGTGTAGCTCAGATGTCAGACAGTATCAGTATGGAAGAAAGTTACGTACAAATGTTTGAAGTTTGGAAATTTGATGGCCCTGCAATTCGCAAAATGCTTGAAGAGA
2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000

K W S H P P S C I K T D C L S L P S F E N A I P M G E K K D V Y K A G E Q V T Y
AAAAATGCTTCCCTCCATCATGCAAAAAAGATTTGCTCAGTTTACCTAGCTTTGAAAAATGCCATACCCATGGAGAGAGAAAGGATGTTATTAAGCGGGGTGAGCAAGTGACTTA
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120

```

T C A T Y Y K M D G A S N V T C I N S R W T G R P T C R D T S C V N P P T V Q N
CACTTGTGCAACATATTACAAATGGATGGAGCCAGTAATGTAACATGCAATTAATAGCAGATGGACAGGAAGGCCAACATGCCAGACACCTCCTGTGTAATCCGCCACAGTACAAAA
3130      3140      3150      3160      3170      3180      3190      3200      3210      3220      3230      3240

A Y I V S R Q M S K Y P S G E R V R Y Q C R S P Y E M F G D E E V M C L N G N W
TGCTTATATAGTGTGCAAGACAGATGAGTAAATATCCATCTGCTGAGAGAGTACGTTATCAATGTAGGAGCCCTTATGAAATGTTTGGGATGAAGAAGTGTGTGTTTAAATGGAAACTG
3250      3260      3270      3280      3290      3300      3310      3320      3330      3340      3350      3360

T E P P Q C K D S T G K C G P P P P I D N G D I T S F P L S V Y A P A S S S V E Y
GACCGAACCCCTCAATGCCAAAGATTCTACAGGAAAAATGTGGGCCCTCCACCTATTGACAAATGGGGACATTACTTCATTCCCGTTGTCAGTATATGCTCCAGCTTCAATCAGTTGAGTA
3370      3380      3390      3400      3410      3420      3430      3440      3450      3460      3470      3480

Q C Q N L Y Q L E G N K R I T C R N G Q W S E P P K C L H P C V I S R E I M E N
CCAATGCCAGAACTTGTATCAACTTGAGGGTAACAAGCGAATAACATGTAGAAATGGACAAATGGTCAGAACCCAAAAATGCTTACATCCGTGTGTAATATCCCGAGAAATATGAAAA
3490      3500      3510      3520      3530      3540      3550      3560      3570      3580      3590      3600

Y N I A L R W T A K Q K L Y S R T G E S V E F V C K R G Y R L S S R S H T L R T
TTATAACATAGCATTAAAGGTGGACAGCCAAACAGAAAGCTTTATTCGAGAAACAGGTGAATCAGTTGAAATTTGTGTGTAACCGGGATATCGTCTTTCATCAGTTCTCACACATTCGGAAC
3610      3620      3630      3640      3650      3660      3670      3680      3690      3700      3710      3720

T C W D G K L E Y P T C A K R *
AACATGTTGGGATGGGAACTGGAGTATCCAACTTGTGCAAAAAGATAGAAATCAATCATAAAGTGCACACCTTTATTTCAGAACTTTAGTATTAAATCAGTTCTCAATTTTATTTATG
3730      3740      3750      3760      3770      3780      3790      3800      3810      3820      3830      3840

TATTGTTTTACTCCTTTTATTTCATACGTAATAATTTGGATTAAATTTGTGAAAAATGTAATTATAAGCTGAGACCGGTGGCTCTCTT
3850      3860      3870      3880      3890      3900      3910      3920

```

(b)

```

      C M E N G W S P T P R C I R V K T C S K
PE3    ....TGTATGGGAGAATGGCGGTCTCCTACTCCAGATGCATCCGTGTCAAAACATGTTCCAAA.....
      =====
B38-1  ....TGTATGGGAGAATGGCGGTCTCCTACTCCAGATGCATCCGTGTGAGCTTTACCTCTGAACCTTCTGATCGA b1-b72
      C M E N G W S P T P R C I R V S F T L *
B38-1  AGGTCATCCCTCTCCAGCTTGAGTGGATCAAAGATGACAAGGGCCAATGGAACCAAGTTTGAGTCTTGCCAG b73-b144
      GTCAATACTTGGGTCTGAGTATGGTGAAGTAGTATCTGTTTTGTTATGTGTGATTATTCCAGCCAGAATGG b145-b216
      GAAATGCTAATTCAGCTCTCCAGGCAGCCAATGGGGCTGGTGGCTTTGAGATTATTAACTCTT b217-b281

```

Fig. 2. Complete cDNA and derived amino acid sequence

(a) The sequence derived from the three overlapping clones. The continuous sequence of 3926 nucleotides is shown. The leader sequence (Met-Arg-Leu---) starts at nucleotide 74, whereas the sequence of the secreted protein (Glu-Asp-Cys---) starts at nucleotide 128. The positions of the three oligonucleotide probes are marked by lines above the relevant nucleotide sequence. (b) The region of B-38-1 which differs from the sequence of PE-3. The sequence of clone PE-3 corresponding to nucleotides 1364-1423 (a) is shown (top line), together with the sequence of the 3' 281 bases of B-38-1. The sequences match in the area denoted by horizontal broken lines. The putative polyadenylation signal in B-38-1 is underlined. Nucleotide numbering is different from that in Fig. (a) to avoid confusion.

pH 6.4 (Hagenbuchle *et al.*, 1980). The liver sample was the same as that used in preparing the library described by Bentley (1986).

The mRNA-primer hybrid and a cDNA-synthesis kit (Amersham) were used to make double-stranded cDNA. Reverse transcriptase was used to form the first strand, followed by *E. coli* DNA polymerase 1 and *E. coli* ribonuclease H for the second strand. The double-stranded cDNA was blunt-ended with phage-T₄ DNA polymerase, ligated into *Pvu*II-cut and dephosphorylated pAT 153, and the recombinant plasmid was used to transform *E. coli* MC1061 made competent by the standard CaCl₂ method. In all, 10000 colonies of the library were screened with R2a and B-38-1 as described above, and three types of positive clone (PE-1, PE-2 and PE-3) were identified.

RESULTS

Overlapping clones

Fig. 1 shows the relative position and extent of overlap of clones R2a, PE-3 and B-38-1. The restriction-endonuclease digestion products used in sequencing are

also summarized in this Figure. The complete cDNA sequence and derived amino acid sequence are shown in Fig. 2(a).

Clone R2a contains 1972 bp of coding sequence followed by a TAG stop codon and 157 bp of 3' untranslated sequence. It represents the C-terminal end of the factor H protein (Day *et al.*, 1987a). The first 13 nucleotides at the 5' end of R2a (Day *et al.*, 1987a) are not in agreement with the sequence of PE-3, and are likely to arise from a cloning artefact. The sequence of these 13 nucleotides is not shown. The sequence of R2a from nucleotide 14 onwards is shown, starting at base no. 1808 on Fig. 2(a). Clone PE-3 contains 1553 bp of coding sequence (nucleotides 455-2006 in Fig. 2a). It overlaps the 5' end of R2a by 212 bases, and is in complete agreement with the R2a sequence (except as noted above) where these overlap. Clone B-38-1 (1644 bp) is less straightforward, and has two unexpected features. Firstly, the clone ends with a 3' untranslated region, although it contains sequences corresponding to the N-terminus of factor H. It overlaps PE-3 and is in complete agreement with the PE-3 sequence from nucleotide 455 (the 5' end of PE-3) up to nucleotide 1409

(Fig. 2a). Thereafter the B-38-1 sequence differs from that of PE-3, as summarized in Fig. 2(b). From position 1410, B-38-1 contains coding sequence for four further amino acids (SFTL), followed by a TGA stop codon and 221 bp of 3' untranslated sequence. Two similar 'truncated' factor H cDNA clones, which code for *N*-terminal regions of factor H but have a 3' untranslated region, have been reported. Clone pH 1050 (Kristensen *et al.*, 1986) contains 672 bp of coding sequence (starting at nucleotide 749 in Fig. 2a), followed by a TGA stop codon, a 232-bp 3' untranslated region containing a polyadenylation signal and a 27-base poly(A) tail. B-38-1 and pH 1050 are identical in sequence where they overlap. H19 (Schulz *et al.*, 1986) contains 1193 bases of coding sequence, the same stop codon and 3' untranslated region, and four bases of the poly(A) tail. H19 starts at a position corresponding to nucleotide 230 in Fig. 2(a) and has only a few minor differences from the sequences of B-38-1 and pH 1050 where it overlaps with them. The differences include the first six bases of H19 (ATCCTG in H19, CCGCCT in B-38-1), the insertion of an A between bases 1158 and 1159 (AG) in Fig. 2(a) and the deletion and insertion of A in the 3' untranslated sequence (deletion of base b102 of Fig. 2b: insertion between positions b124 and b125). The occurrence, in three different cDNA libraries made in different ways, of similar clones coding for *N*-terminal regions of factor H, but having a stop codon, 3' untranslated region and poly(A) tail, strongly suggests that these are not cloning artefacts. Their origin is indicated by Northern-blotting experiments (Ripoche *et al.*, 1987), which show that clone B-38-1 hybridizes to two mRNA species in blots of liver poly(A)⁺ mRNA. These are the normal 4.3-kb mRNA, plus a 1.8-kb species of similar abundance to the 4.3-kb species. The 1.8-kb species does not hybridize with 3' clones, such as R2a. Clones B-38-1, pH 1050 and H19 are likely to be derived from the shorter mRNA. Because of the extended region of identical sequence in the B-38-1 and PE-3, and because the cDNA sequence of B-38-1 is extensively confirmed by sequencing of peptides derived from plasma factor H (Fig. 4 below), we conclude: (i) that the 4.3-kb and 1.8-kb mRNA species are likely to arise from the same gene by an alternative splicing process (Ripoche *et al.*, 1987), and (ii) that the sequence of B-38-1 up to nucleotide 1409 (Fig. 2a) encodes the true *N*-terminal sequence of factor H.

Clone B-38-1 is inverted at the 5' end. Nucleotides 1–(approx.) 375 (Fig. 2a) are inverted and, because of this, our determination of bases 376–379 is uncertain. This region is, however, covered by the H19 clone (Schulz *et al.*, 1986), and the sequence 376–379 (AAAT) shown in Fig. 2(a) is derived from the H19 clone.

Complete cDNA sequence

The cDNA sequence assembled from the three clones is shown in Fig. 2(a). A 73-bp 5' untranslated region is followed by sequence coding for an 18-amino-acid putative leader peptide. The leader sequence is consistent with the general pattern of such sequences (von Heijne, 1984). The coding sequence for the secreted protein begins at base 128 (corresponding to the protein sequence EDC...) and continues to base 3766 (3639 bp; coding for 1213 amino acids). This is followed by a TAG stop codon and 157 bp of 3' untranslated sequence. The 5' and 3' untranslated regions are not highly homologous with the corresponding regions of mouse factor H cDNA

(Kristensen & Tack, 1986). The region encoding the leader peptide is, however, identical with the mouse nucleotide sequence in 44 out of 54 positions.

Two possible polymorphisms were detected in DNA sequencing. Base 1324 (Fig. 2a) is T in clones B-38-1, PE-3 and pH 1050 (Kristensen *et al.*, 1986) but is C in H19 (Schulz *et al.*, 1986). This does not affect the amino acid sequence, but alters a potential *FokI* restriction site. Since *FokI* is a frequent cutter, polymorphism at this position in genomic DNA has not been investigated. Base 1277 is T (coding for a tyrosine residue) in clones H19, pH 1050 and, from the present study, in the partially sequenced clones B-15-1 and PE-2. In clones B-38-1 and PE-3, base 1277 is C, coding for histidine. Both tyrosine and histidine were detected at this position in amino acid sequencing (Day *et al.*, 1987b; Fig. 4 below), and it is likely that this polymorphism may account for the charge difference observed at the protein level (Rodriguez de Cordoba & Rubinstein, 1987), between the two most abundant factor H variants. In the DNA sequence, the T→C change destroys an *NlaIII* restriction site. Since *NlaIII* is difficult to obtain commercially, and is a frequent cutter, it may be impractical to examine this polymorphic site at the genomic DNA level.

In Fig. 2(b) is shown the sequence at the 3' end of clone B-38-1 and the point at which it diverges from the sequence of PE-3. The B-38-1 clone, as well as the pH 1050 (Kristensen *et al.*, 1986) and H19 (Schulz *et al.*, 1986) clones are likely, as discussed above, to be derived from the 1.8-kb mRNA species seen in Northern blotting (Ripoche *et al.*, 1987). The complete coding sequence within this mRNA species is represented by nucleotides 74–1409 in Fig. 2(a), followed by nucleotides b47–b57 (encoding a Ser-Phe-Thr-Leu sequence) in Fig. 2(b). Clone B-38-1 also provides an incomplete 5' untranslated and incomplete 3' untranslated sequence for this mRNA. The complete 3' untranslated sequence is provided by clone pH 1050 (Kristensen *et al.*, 1986) and consists of the 3' untranslated sequence shown in Fig. 2(b) with a further ten bases (TCTCTGCTGC) at the 3' end, and a poly(A) tail.

As noted previously (Day *et al.*, 1987a), a further short mRNA species hybridizing with 3' factor H cDNA probes (e.g. R2a), but not with 5' cDNA probes (e.g. B-38-1), is also detectable in Northern blots of human liver poly(A)⁺ mRNA. This species is 1.2–1.5 kb in length, and is much less abundant than the 4.3- or 1.8-kb species. The origin of this mRNA is unknown, but a number of short cDNA clones (1.2–1.5 kb) have been selected from the library described by Belt *et al.* (1984), using R2a as a probe. These clones may be derived from this third mRNA species and appear to have minor sequence differences from R2a (R. B. Sim & J. Ripoché, unpublished work). Further characterization of these clones is required.

Amino acid sequence of factor H

The signal sequence of human factor H (Fig. 2a) is similar to that of mouse factor H (Kristensen & Tack, 1986), having 11 out of 18 identities and three conservative replacements. The *C*-terminus of the human sequence ends with two basic residues (Lys-Arg) (Figs. 2a and Fig. 3), which are not present in the mouse sequence, although the preceding *C*-terminal sequence is highly homologous between human and mouse. As noted previously (Day *et al.*, 1987a), amino acid analysis and

The entire amino acid sequence of factor H (omitting the leader peptide) is shown arranged in 20 homologous units, each of about 60 amino acids. Gaps are inserted, or residues inserted above the line (indicated by ▼) to maximize alignment. Highly conserved residues (including conservative replacements) are boxed, and a summary of the conserved residues (consensus sequence) is shown at the bottom. Only the cysteine residues (marked with an asterisk) are absolutely conserved in all repeat units. The two residues in parentheses (line 7) indicate the site of the His/Tyr polymorphism. Residues number (top) is approximate only, owing to variation in length of repeat units.

The rest of the factor H amino acid sequence (1211 residues) is arranged in 20 homologous repeating units, each about 60 amino acids long, based on a framework of four invariant cysteine residues and highly conserved

acidic, proline, asparagine, glycine, tyrosine, phenylalanine and tryptophan residues. The entire sequence (excluding the signal sequence) is arranged in Fig. 3 to demonstrate the internal homology. A consensus sequence of highly conserved residues is shown at the bottom of the Figure. Mouse factor H (Kristensen & Tack, 1986) is also arranged into 20 similar homologous segments, and the mouse and human sequences have

The cDNA-derived amino acid sequence of factor H is shown (in the same format as Fig. 3). Sequences of peptides derived as described in the text are shown. Sequences of two peptides derived by trypsin treatment of 'native' factor H (peptide V45, line 15, and V120, line 6) are also shown (A. J. Day, V. Malhotra, N. J. Price & R. B. Sim, unpublished work). Sites of potential *N*-glycosylation are boxed.

The 60-amino-acid repeating unit that makes up the entire sequence of factor H is also found in the functionally related proteins CR1, C4bp and DAF, and also in CR2 and in other complement proteins (C2, factor B, C1r and C1s) and non-complement proteins (β_2 I, factor XIII, the interleukin 2 receptor and haptoglobin). These similarities have been discussed extensively (see, e.g., Reid *et al.*, 1986; Sim *et al.*, 1986, 1987; Caras *et al.*, 1987; Klickstein *et al.*, 1987).

Peptide sequencing

Sequences of peptides obtained by trypsin digestion of reduced, carboxymethylated, succinylated factor H, or by trypsin treatment of 'native' factor H under non-denaturing conditions (A. J. Day, V. Malhotra, N. J. Price & R. B. Sim, unpublished work) confirm 480 out of the 1211–1213 residues of factor H (40 %) (Fig. 4). All peptide sequences obtained were found within the cDNA-derived amino acid sequence. Kristensen *et al.* (1986) also reported peptide sequences. Some of these

overlap with peptide sequences shown on Fig. 4, but non-overlapping sequences confirm a further 163 residues of the cDNA-derived sequence. In all, therefore, 53 % of the sequence is confirmed at the peptide level. One sequence (peptide HT-16-37) reported by Kristensen *et al.* (1986) was not found in the cDNA-derived sequence. This peptide does not align well with the general consensus sequence, and so may not have been derived from factor H. At the *N*-terminus of factor H, the sequence was defined in previous reports (Sim & DiScipio, 1982; Ripoche *et al.*, 1984; Kristensen *et al.*, 1986) as EDQN--. The cDNA-derived sequence shows that the sequence is EDCN--, and this was confirmed by re-sequencing the *N*-terminus of reduced, radioalkylated, factor H. The original assignment of Q in the third position was made by using protein that was not radio-labelled at the alkylation step; this makes distinction of *S*-carboxymethylcysteine from glutamine difficult. Identification of this third residue as cysteine confirms that factor H has an even number of cysteine residues (four per repeat), all of which are likely to be involved in disulphide bridges. Factor H has no free SH groups titratable with iodo[2-³H]acetamide under denaturing conditions (8 M-guanidium chloride, pH 8.0).

Two unusual 'tryptic' cleavages were deduced from the peptide sequences shown in Fig. 4. These were a Y-A cleavage (repeat 8, peptide TR-6-2-II, Fig. 4), and an N-S cleavage (repeat 10, peptide TR-6-2-I). We have no explanation for these.

The total polypeptide M_r of factor H is 136824 (for 1211 residues) or 137108 (for 1213 residues, including the two basic *C*-terminal residues). There are nine potential attachment sites for asparagine-linked sugars (Fig. 4). Peptide sequencing has confirmed (judged by the complete absence of asparagine in the expected position in sequencing) that the single sites in repeat 9, 13, 14 and the two sites in repeat 15 are occupied. Other studies, using carbohydrate staining of fragments produced by proteolysis of undenatured factor H (Sim & DiScipio, 1982; A. J. Day, V. Malhotra, N. J. Price & R. B. Sim, unpublished work), show that the site in repeat 4 is not glycosylated and suggests that, of the sites in repeats 12, 17 and 18, only one is occupied. Assuming that each of the six occupied carbohydrate-linkage sites bears an oligosaccharide of about 3000 M_r , then the overall M_r of the glycoprotein is about 155000, a value in agreement with previous estimates (Sim & DiScipio, 1982).

Mouse factor H has eight potential sites for *N*-linked carbohydrate, but with the exception of the sites in repeats 12, 13 and 17, the location of these sites is not conserved between mouse and human factor H.

DISCUSSION

The complete cDNA-derived amino acid sequence of human factor H has been determined, and the derived sequence is confirmed at 53 % of the residue positions by direct peptide sequencing. The overall sequence for human factor H is 61 % homologous with that of mouse factor H (Kristensen & Tack, 1986) and, as in the mouse protein, the entire amino acid sequence is arranged into 20 internally homologous units, each about 60 amino acids long.

This type of 'cysteine-rich' repeat, as noted above, is present in many complement proteins and also some non-complement ones. Each repeat unit is likely to be

encoded as a single exon [see Reid *et al.* (1986) for discussion], and each is likely to represent a compact domain within the protein. No experimentally derived information is available on the secondary and tertiary structure of each repeat units, but it is clear that proteins which consist of large numbers of contiguous repeats (e.g. factor H, C4bp, β_2 I) are very elongated in structure. CR1 has been likened to a straight string of beads, in which each bead consists of one 60-amino-acid long repeat unit (Klickstein *et al.*, 1987).

Sequencing of cDNA clones for human factor H, combined with Northern-blotting experiments (Ripoche *et al.*, 1987) indicates the high probability that two mRNA species (4.3 and 1.8 kb) are generated from a single gene sequence by alternative splicing. Another explanation, that these two mRNA species are products of different, but closely related, genes cannot yet be ruled out. The shorter mRNA species appears to encode the first 427 amino acids of the full factor H sequence (Fig. 2), followed by four amino acids (Ser-Phe-Thr-Leu) not present in the sequence of 'full-length' factor H. Western-blotting experiments on whole human serum samples separated by SDS/polyacrylamide-gel electrophoresis have been done (J. Ripoche, unpublished work). Monoclonal and polyclonal anti-(factor H) antibodies recognize, on these blots, in addition to the familiar factor H of 155000 M_r , a low-abundance species which has an M_r of 45000–50000 under non-reducing conditions. This species may be the translation product of the 1.8-kb mRNA. However, extensive further experimentation, including in particular the determination of the *C*-terminal sequence, is required to distinguish this material from an unusual minor degradation product of the 155000- M_r factor H. The expected translation product, on the basis of the cDNA sequence, would have an M_r of 49100 and is likely to be non-glycosylated. It is also likely to possess the 'cofactor' activity of factor H, since Alsenz *et al.* (1985) have shown that a shorter *N*-terminal fragment (residues 1–313; M_r 36520) formed by limited trypsin digest of factor H, possesses cofactor activity. Further, the expected translation product would almost certainly be recognized by the eight characterized anti-(factor H) monoclonal antibodies, as all of these react with the 36520- M_r *N*-terminal tryptic fragment (Sim *et al.*, 1983; Alsenz *et al.*, 1985; J. Ripoche, A. Erdei, D. Gilbert, A. Al Salihi & M. Fontaine, unpublished work).

The origin and nature of the other minor factor H-related mRNA (1.2–1.5 kb) (Day *et al.*, 1987) is not yet clear.

The sequence determined for factor H cDNA gives no indication of any difference between soluble and membrane-bound factor H. The functionally related proteins DAF and CR1 both also have membrane-bound and soluble forms (Caras *et al.*, 1987; Yoon & Fearon, 1985). CR1, however, possesses a *C*-terminal hydrophobic transmembrane segment (Klickstein *et al.*, 1987), which adequately explains the association of the protein with the membrane. The mechanism of release of the soluble form, which has an M_r apparently identical with that of the membrane form, is unknown.

In DAF, the membrane-bound form may be attached either via a transmembrane segment (Caras *et al.*, 1987) or be processed to form a glycosylphosphatidylinositol anchorage (Low, 1987). The soluble form, however, is translated from an alternatively spliced mRNA and lacks

a transmembrane segment (Caras *et al.*, 1987). In factor H it is possible that the two C-terminal basic residues may be modified to form covalent anchorages of the types summarized by Low (1987). Different mRNA species encoding additional structures may also be present in cell types that synthesize factor H destined to become membrane-associated. These points require investigation.

Note added in proof (received 23 November 1987)

The amino acid sequence in the region of the inversion of clone B-38-1 has been confirmed from tryptic-peptide data. Peptide TR-8-8-I confirms amino acid residues 61–90, corresponding to bases 308–397 in Fig. 2(a).

We are most grateful to Professor M. P. Dierich, Dr. T. Kristensen, Dr. T. F. Schulz, Dr. W. Schwäble and Dr. B. F. Tack for helpful discussion and exchange of results before their publication. We thank Dr. D. R. Bentley, Dr. E. Sim and Dr. R. D. Campbell for generous help and advice, and Miss B. Moffatt and Mrs. J. Newell for expert technical assistance. J. R. is an EMBO (European Molecular Biology Organization) Fellow, and A. J. D. holds an SERC (Science and Engineering Research Council) CASE (Co-operative Awards in Science and Engineering) studentship.

REFERENCES

- Alsenz, J., Schulz, T. F., Lambris, J. D., Sim, R. B. & Dierich, M. P. (1985) *Biochem. J.* **232**, 841–850.
- Belt, K. T., Carroll, M. C. & Porter, R. R. (1984) *Cell* (Cambridge, Mass.) **36**, 907–914.
- Bentley, D. R. (1986) *Biochem. J.* **239**, 339–345.
- Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1521–1534.
- Caras, I. W., Davitz, M. A., Rhee, L., Weddell, G., Martin, D. W. & Nussenzweig, V. (1987) *Nature* (London) **325**, 545–549.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **24**, 5294–5299.
- Christie, D. L. & Gagnon, J. (1982) *Biochem. J.* **201**, 555–563.
- Chung, L. P., Bentley, D. R. & Reid, K. B. M. (1985) *Biochem. J.* **230**, 133–141.
- Day, A. J. & Sim, R. B. (1986) *Biochem. Soc. Trans.* **14**, 73–74.
- Day, A. J., Ripoche, J., Lyons, A., McIntosh, B., Harris, T. J. R. & Sim, R. B. (1987a) *Biosci. Rep.* **7**, 201–207.
- Day, A. J., Ripoche, J., Willis, A. C. & Sim, R. B. (1987b) *Immunogenetics*, in the press.
- Erdei, A. & Sim, R. B. (1987) *Biochem. J.* **246**, 149–156.
- Gardner, R. C., Howarth, A. J., Messing, J. & Shepherd, R. (1982) *DNA* **1**, 109–115.
- Grosveld, F. G., Dahl, H. M., Boer, E. D. & Flavell, R. A. (1981) *Gene* **31**, 227–237.
- Hagenbuchle, O., Bovey, R. & Young, R. A. (1980) *Cell* (Cambridge, Mass.) **21**, 179–187.
- Holers, V. M., Cole, J. L., Lublin, D. M., Seya, T. & Atkinson, J. P. (1985) *Immunol. Today* **6**, 188–192.
- Klickstein, L. B., Wong, W. W., Smith, J. A., Weis, J. H., Wilson, J. G. & Fearon, D. T. (1987) *J. Exp. Med.* **165**, 1095–1112.
- Kristensen, T. & Tack, B. F. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3963–3967.
- Kristensen, T., Wetsel, R. A. & Tack, B. F. (1986) *J. Immunol.* **136**, 3407–3411.
- Lambris, J. D. & Ross, G. D. (1982) *J. Exp. Med.* **155**, 1400–1411.
- Low, M. G. (1987) *Biochem. J.* **244**, 1–13.
- Malhotra, V. & Sim, R. B. (1985) *Eur. J. Immunol.* **15**, 935–941.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T. & Tack, B. F. (1986) *Immunol. Today* **7**, 230–234.
- Rigby, P. W. J., Dieckman, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
- Ripoche, J., Al Salihi, A., Rousseaux, J. & Fontaine, M. (1984) *Biochem. J.* **221**, 89–96.
- Ripoche, J., Day, A. J., Willis, A. C., Belt, K. T., Campbell, R. D. & Sim, R. B. (1986) *Biosci. Rep.* **6**, 65–72.
- Ripoche, J., Day, A. J., Moffatt, B. & Sim, R. B. (1987) *Biochem. Soc. Trans.* **15**, 651–652.
- Rodriguez de Cordoba, S. & Rubinstein, P. (1986) *J. Exp. Med.* **164**, 1274–1283.
- Rodriguez de Cordoba, S. & Rubinstein, P. (1987) *Immunogenetics* **25**, 267–268.
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
- Schulz, T. F., Schreiner, D., Alsenz, J., Lambris, J. D. & Dierich, M. P. (1984) *J. Immunol.* **132**, 392–398.
- Schulz, T. F., Schwäble, W., Stanley, K. K., Weiss, E. & Dierich, M. P. (1986) *Eur. J. Immunol.* **16**, 1351–1355.
- Sim, E., Palmer, M. S., Puklavec, M. & Sim, R. B. (1983) *Biosci. Rep.* **3**, 1119–1131.
- Sim, R. B. & DiScipio, R. G. (1982) *Biochem. J.* **205**, 285–293.
- Sim, R. B., Malhotra, V., Ripoche, J., Day, A. J., Miklem, K. J. & Sim, E. (1986) *Biochem. Soc. Symp.* **51**, 83–96.
- Sim, R. B., Malhotra, V., Day, A. J. & Erdei, A. (1987) *Immunol. Lett.* **14**, 183–190.
- Staden, R. (1982) *Nucleic Acids Res.* **10**, 4731–4751.
- von Heijne, G. (1984) *EMBO J.* **3**, 2315–2318.
- Whaley, K. & Ruddy, S. (1976) *J. Exp. Med.* **144**, 1147–1163.
- Weiler, J. M., Daha, M. R., Austen, K. F. & Fearon, D. T. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3268–3272.
- Yoon, S. H. & Fearon, D. T. (1985) *J. Immunol.* **134**, 3332–3338.