Characterization of the primary amino acid sequence of human complement control protein Factor I from an analysis of cDNA clones

Catherine F. CATTERALL,* Alan LYONS,* Robert B. SIM,† Anthony J. DAY† and Timothy J. R. HARRIS*‡

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

A cDNA clone of the mRNA coding for the human complement system control protein Factor I has been isolated. The predicted amino acid sequence obtained from the DNA sequence demonstrates a protein consisting of a heavy chain (M_r 35400) linked to a light chain (M_r 27600), both of which contain three sites for *N*-linked glycosylation. The light chain has clear homology with other serine proteinases, most notably in the region of the catalytically active and structurally important amino acids and shares some of the features characteristic of the plasminogen activators. The heavy chain has a clear 'mosaic' nature typical of the plasma serine proteinases; in particular it contains class A and class B LDL (low-density lipoprotein) receptor repeats with conserved cysteine residues similar to those found in other complement proteins.

INTRODUCTION

The C3 convertase enzyme, the formation of which is essential in both the classical and alternative complement pathways, is modulated by the action of three plasma control proteins: Factor I, Factor H and C4-binding protein. Human Factor I is a serine proteinase responsible for cleaving the α' chains of C4b and C3b in the presence of the cofactors C4-binding protein and Factor H respectively, to produce C4c and C4d or iC3b. Further proteolysis of iC3b, probably by other serine proteinases, yields C3c and C3dg (Reid, 1986). C3b is susceptible to Factor I proteolysis only after formation of a C3-cofactor complex. Factor I is a plasma glycoprotein (concn. approx. 35 μ g/ml in plasma) of apparent M_r 88000 composed of two polypeptide chains (M_r 50000 and 38000) linked by disulphide bonds.

Limited *N*-terminal sequence analysis of Factor I light chain (Davis, 1981), indicated that Factor I is a serine proteinase, although it is not inhibited by di-isopropyl-fluorophosphate, a general inhibitor of this class of enzymes. Also, unlike the majority of plasma serine proteinases, such as the coagulation enzymes and complement proteinases Clr, Cls, C2 and Factor B, Factor I does not appear to have a circulating proenzyme form. The circulating active form is stable and is not inhibited by any of the known plasma proteinase inhibitors (Crossley, 1981). A single chain precursor form of Factor I has been detected in cell-free systems, and in labelling studies *in vitro* with intact Hep G2 and other human hepatoma cell lines (Goldberger *et al.*, 1983). Recently, the light chain has been purified and partial amino acid sequence data obtained by CNBr cleavage and *N*-terminal amino acid sequencing (Yuan *et al.*, 1986). From this analysis the proposed homology of Factor I to other serine proteinases (Davis, 1981) was confirmed, and potential glycosylation sites identified (Yuan *et al.*, 1986). Little information has been presented concerning the amino acid sequence of the

heavy chain of Factor I apart from the *N*-terminal 32 amino acids (Davis, 1981). Access to this amino acid sequence data has allowed the design and utilization of oligonucleotides to screen human liver cDNA libraries for Factor I specific clones. The isolation, nucleotide sequencing and analysis of the predicted amino acid sequence of both heavy and light chains of Factor I is the subject of this paper.

MATERIALS AND METHODS

The λ gt10 human liver cDNA library was purchased from Clontech Laboratories Inc. (Palo Alto, CA, U.S.A.). Hybond nylon filters, M13 vectors, radioisotopes and most of the enzymes were from Amersham International. Human liver cDNA libraries (Belt *et al.*, 1984; Bentley, 1986) in pAT 153/*Pvu*II/8 were supplied by Drs. K. T. Belt, R. D. Campbell and D. R. Bentley (MRC Immunochemistry Unit).

Oligonucleotide synthesis

Oligonucleotides were synthesized on an Applied Biosystems model 380A DNA synthesizer using phosphoramidate precursors. Deblocked oligonucleotides were purified by reverse-phase h.p.l.c. and integrity of the product peak was checked by 32 P end-labelling and polyacrylamide-gel electrophoresis.

Isolation of plasmid cDNA clones

Plasmid clones from the Belt library (Belt *et al.*, 1984) were screened with the 32 P-labelled long oligonucleotide essentially as described (Docherty *et al.*, 1985) the final wash being in 2 $\times$ SSC at 60 °C (1 $\times$ SSC is 0.15 M-NaCl/0.015 M-sodium citrate). Screening of the Bentley library (Bentley, 1986) with the mixed oligonucleotide (sp. radioactivity 3 $\times$ 10⁷ c.p.m./ μ g) followed the general method of Wallace *et al.* (1981). Filters were washed once in 6 $\times$ SSC at room temperature for 10 min, followed by

‡ To whom correspondence and reprint requests should be sent.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00318.

Vol. 242

codon bias or making short mixed oligonucleotides from amino acid sequences with relatively non-degenerate codons (Lathe, 1985). The initial approach followed the former strategy. A human liver cDNA library in pAT 153/*Pvu*II/8 (Belt *et al.*, 1984) of 60000 clones was screened with a 32 P-labelled long oligonucleotide predicted from amino acids 154–173 of the Factor I light chain (Yuan *et al.*, 1986), (Fig. 1*b*). Seven positive colonies were identified and four single reactive clones were isolated on rescreeing. Further characterization of the plasmid inserts in these clones by restriction enzyme mapping revealed two classes of insert each of about 3 kb but which did not appear to be related. Nucleotide sequence analysis of a restriction fragment hybridizing to the probe from each class of insert (detected by Southern blotting) indicated that these inserts were not Factor I specific and they were not analysed further.

A mixed oligonucleotide probe based on amino acids 160–165 (Yuan *et al.*, 1986) was synthesized therefore (Fig. 1*a*) and used to screen a different human liver plasmid cDNA library (Bentley, 1986). A single positive clone (pFI 29) was isolated from approx. 20000 colonies. The insert in the plasmid was about 800 bp (Fig. 2*a*) and was shown by nucleotide sequence analysis to be Factor I specific. This fragment was then used to screen a human liver library made in λ gt10 to obtain a full-length clone. One of two positive phages obtained (λ FI) was found to contain DNA with an insert about 1200 bp longer than that in pFI 29 by restriction enzyme mapping (Fig. 2*a*). Both inserts were subcloned into M13 vectors and subjected to nucleotide sequence analysis according to the strategy outlined in Fig. 2*a*).

Nucleotide sequence of Factor I cDNA

A DNA sequence of about 2000 bp was obtained, predicting a mRNA with an open reading frame (from the first methionine codon) of 1749 nucleotides (583 amino acids), a 3' untranslated region of 187 nucleotides and a 5' leader sequence of at least 30 nucleotides (Fig. 2*b*). Northern blot analysis of human liver mRNA probed with Factor I cDNA indicated that the mRNA was about 2.4 kb (results not shown). The presence of a polyadenylation signal AATAAA near the 3' end of the DNA sequence (nucleotide 1839) but some distance from the 3' end itself suggests that this clone is missing an additional polyadenylation signal next to the poly(A). It is possible therefore that both the 3' and 5' end of the mRNA extends further than that represented in the clones. No attempts have been made to analyse this in more detail.

The deduced sequence for amino acid residues 1–32 shows complete correspondence to the *N*-terminal sequence of the Factor I heavy chain presented by Davis (1981), except that residue 12 appears to be Glu rather than Gln. The first 18 residues of the coding sequence (–18 to –1) thus appear to constitute a signal sequence and this has the expected hydrophobic character (Watson, 1984). The light chain *N*-terminal sequence (Yuan *et al.*, 1986) is found in the predicted amino acid sequence at residue 322 (Ile) after four basic residues, providing evidence that Factor I is synthesized as a single chain polypeptide containing both heavy and light chains, which is subsequently cleaved. It is likely that the basic residues are removed in the course of the processing, as occurs with other two-chain serine proteinases (e.g. see Long *et al.*, 1984). The first 25

residues of the predicted *N*-terminal light chain sequence differ from that reported by Davis (1981) in 10 positions but are in complete agreement with the sequence of Yuan *et al.* (1986) over this region. Further comparison of the derived light chain sequence with that of Yuan *et al.* (1986) shows that 43 amino acids were missing in the *N*-terminal part of their sequence plus one amino acid at the C-terminus.

The rest of the light chain sequence is identical with that determined by Yuan *et al.* (1986) except for a surprising transposition of a glutamic acid residue. In the sequence of Yuan *et al.* (1986), there is a Glu at residue 82 (equivalent to being between positions 406 and 407 in the cDNA-derived sequence); this residue is not present in the cDNA-derived sequence but there is a Glu at our position 416. Although Factor I is known to have at least two common allelic variants of different charge (Nishimukai & Tamaki, 1986), no evidence has been found for such polymorphism in the sequences available for study.

The amino acid sequence confirms the presence of three potential sites for *N*-linked glycosylation in the light chain (Asn-446, 476 and 518) (Yuan *et al.*, 1986) and also reveals three sites in the heavy chain (Asn-52, 85 and 159). The amino acid composition of Factor I derived from the sequence is in good agreement with previous analyses (Pangburn *et al.*, 1977; Davis, 1981) except that the predicted cysteine content (7.1%) is higher than was derived experimentally.

The overall polypeptide M_r of Factor I is about 63000 (heavy chain 35400, light chain 27600), and this is in good agreement with an estimate of 65000 for the single-chain non-glycosylated precursor form synthesized in a rabbit reticulocyte lysate (Goldberger *et al.*, 1983). The polypeptide M_r values of the heavy and light chains are considerably lower than the M_r estimates for those chains made by SDS/polyacrylamide-gel electrophoresis (50000 and 38000) (Crossley, 1981). If all six potential glycosylation sites are occupied (e.g. by complex *N*-linked sugars of M_r 3000–4000), then the total M_r for the glycosylated heavy and light chains would be about 44000–47000 and 37000–40000 respectively, in reasonable agreement with previous estimates.

The nucleotide sequence coding for the amino acids on which the original long oligonucleotide was based shows considerable divergence from the chosen sequence, mostly because the DNA sequence is more A·T-rich than expected (13/59 choices wrong, 22.0% mismatch; Figs. 1*b* and 1*c*). The fact that these differences are evenly distributed throughout the probe probably explains the lack of specificity achieved at the stringency used.

Analysis of the protein sequence

As pointed out previously (Davis, 1981; Yuan *et al.*, 1986), the light chain has homology with other serine proteinases or with the serine proteinase domains of other proteins. This homology is shown in Fig. 3. The most notable blocks of homology surround the catalytically important residues, e.g. Ile-322, His-362, Asp-411 and Ser-507, including several cysteine, glycine and proline residues. The presence of an aspartate residue at 501 was confirmed and is consistent with the known 'trypsin'-like cleavage specificity of Factor I for its natural substrates C4b and C3b, in which it cleaves only arginyl bonds (Belt *et al.*, 1984; de Bruijn & Fey, 1985).

Apart from the general serine proteinase homology

Fig. 3. Homology of Factor I with eight other serine proteinases or serine proteinase domains

Structurally conserved regions are underlined and residues that are in homologous positions are joined by dotted lines. The numbers at the top of the Figure refer to chymotrypsin and those above the Factor I sequence to those residues specifically referred to in the text. The 'additional' cysteines in tPA, urokinase and Factor I are underlined. Catalytically important residues are marked with an asterisk, as is the aspartate in the specificity pocket. CT, bovine chymotrypsin; T, bovine trypsin; E, porcine elastase; tPA, human tissue plasminogen activator; UK, human urokinase; FI, Factor I; NGF, murine nerve growth factor, α subunit; PC, human protein C; FX, human Factor X. CT, T, E, tPA and UK sequences are from Straßburger *et al.* (1983), NGF from Isackson *et al.* (1984), protein C from Foster & Davie (1984), and Factor X from Leytus *et al.* (1984).

(a)

LDL receptor (295-333)	N E C L D - N N G G C S H - - - V C N D L K I G - - Y E C L C P D G F Q L V A - Q R R C E D
EGF (4-44)	S E C P L S H D G Y C L H D G - V C M Y I E A L D K Y A C N C V V G Y I - - - G E R C Q Y
Phospholipase A ₂ (78-107)	Y S C S G S - A I T C S S K N K E C - - - E A - - - F I C N C D R - - - - - A A I C F S
Complement C9 (486-519)	R K C - - - - H T C Q N G G T V - I L M D G - - K C L C A C P F K F - - - - E G I A C E I
C1r A chain (127-174)	D E C A S R A E D P Q A C - - - Q H L C H β Y V G G - - Y F C S C R P G Y E L Q E D R H S C Q A
Factor I, heavy chain (12-59)	K K C L A K K Y T H A C D K - - V F C Q P W Q R C I E G T C V C K L P Y Q C P K N G T A C A T

(b)

LDL receptor	4-44	D R C E R N E F Q C Q D G K C I S G K W V C D G S A E C Q D G S D E S Q A C L S
	45-85	V T C K S G D F S C G G E A R C I P Q F W R C D G Q V D C D M G S D E Q G C P P
	86-124	K T C S Q D E F R C H D G K C I S R Q F V C D S D R D C L D G S D E A S C P V
	125-165	L T C G P A S F Q C N S S T C I P Q L W A C D N D P D C E D G S D E W F C R G
	174-212	S P C S A F E F H C L S G E C I H S S W R C D G G P D C K D K S D E E N C A V
	213-251	A T C R P D E F H C S D G N C I H G S R Q C D R E Y D C K D M S D E V G C V N
	253-294	T L C A G P N K F H C H S G E C I T L D K V C N M A R D C R D W S D E P I C G T
Complement C9	78-115	D D C - G N D F Q C S T G R C I K M R L R C N G D N D C G D F S D E D D C E S
Factor I	204-238	D S P - M D D A F Q C V N G K Y I S M K A - C D G I N D C G D Q S D E L - C -
	239-275	K A C Q G K G F H C K S G V C I P S Q Y Q C N G E V D C I T G E D E V G C A G

(c)

Fig. 4. Homology of Factor I with other human proteins

(a) A comparison of residues 12-59 of Factor I with the LDL receptor class B repeat structure in other human proteins. Sequence of human LDL receptor, EGF and C9 are from Stanley *et al.* (1986). Phospholipase A₂ sequence is from Verheij *et al.* (1983), and C1r A chain from Arlaud *et al.* (1987). The conserved cysteines and arrangement of disulphide bonds in EGF are indicated and the two extra cysteines found in Factor I are underlined. ' β ' in the C1r A chain indicates β -hydroxyaspartic acid. (b) A comparison of the Class A repeats in human LDL receptor, complement component C9, and Factor I. Sequences are from Stanley *et al.* (1986). Highly conserved residues are indicated and the underlined portion of C9 represents an antigenic sequence to which cross-reacting monoclonal antibodies have been raised (see text and Tschopp & Mollnes, 1986). (c) The arrangement of recognized repetitive units in human LDL receptor, C8 α chain, C9 and Factor I. The alignment of the Class A (■) and Class B (▼) repeats in the LDL receptor is indicated.

Table 1. Miscellaneous homologies of Factor I with other proteins

Comparison	Region		Homology		Sequence
	Protein	Factor I	Identity	Replacements	
α -Foetoprotein precursor (mouse) (A)	267–287	68–89	45%	54%	I CQKSLECLHPGTFKFLNNGTCT :: : : : : : : : A CQGNSLECLQDGEKVMTY-ICS
HSV 2 glycoprotein D (D)	216–236	123–144	39%	56%	I CKSSWSMREANVACL-DLGFQQG :: : : : : : : D CKYALPLRIPPAACLTSKAYQQG
Glycoprotein precursor phlebovirus (punta-toro) (G)	996–1005	78–87	90%	90%	I PGTKFLNNGT : : : : : : : : G PGTKFLNWGT

blocks noted above, the Factor I light chain sequence does not bear a particularly close relationship to the serine proteinase region of any of the other human plasma proteinases, such as Cls, Clr, Factor D, Factor B, plasminogen, Factors IX or X or protein C. Other groups of serine proteinases, such as group-specific protease (Woodbury & Neurath, 1980) nerve growth factor α -subunit (Isackson *et al.*, 1984) and the recently described cytotoxic T cell proteases (Gershenfeld & Weissman, 1986; Lobe *et al.*, 1986) are also not strikingly above the average homology level. The Factor I sequence has some unusual features, such as the additional Cys residues at positions 355 and 426. These are likely to be disulphide-linked, and are absent from most serine proteinases, with the exceptions of urokinase and tissue plasminogen activator (tPA) (Fig. 3). Other cysteine residues at positions 347, 363, 435, 449, 477, 492, 503, 513 and 532 are very highly conserved in other serine proteinases. By comparison with chymotrypsin and plasminogen, Cys residues 347 and 363, 449 and 513, 477 and 492 and 503 and 532 are likely to form disulphide linkages, while Cys-435 is probably involved in disulphide linkage to the heavy chain (Dayhoff, 1978). The methionine at position 417 is unusual, as this position is almost invariably leucine in serine proteinases (Hewett-Emmett *et al.*, 1980).

There are less precise but nevertheless interesting homologies within the heavy chain sequence of Factor I. One region, residues 12–59, has homology with the sequence pattern designated as the class B type repeat of the LDL (low-density lipoprotein) receptor (for discussion see Stanley *et al.*, 1986). A comparison of this region in Factor I with regions of LDL receptor, EGF (epidermal growth factor), phospholipase A₂, human complement component C9 and complement subcomponent Clr A chain is shown in Fig. 4(a). This homology pattern is based on six conserved cysteines, and is also considered to be present in coagulation Factors IX and X, tPA, urokinase, protein C, rat TGF-1 and the 19K protein from vaccinia virus (Stanley *et al.*, 1986). The disulphide bonding pattern in this region of EGF is shown in Fig. 4(a); it is possible that the other proteins

have the same arrangement of disulphides. Factor I, however, does have two extra cysteines in this position.

Two other regions of the Factor I heavy chain (204–238 and 239–275) have homology with the class A repeats of the LDL receptor (Stanley *et al.*, 1986). This is a further type of disulphide-rich repetitive unit found in the LDL receptor and in complement component C9. As with the class B repeat, there is limited sequence identity, other than conservation of cysteines, between these units (Fig. 4b). Complement component C8 (α -chain) also contains a structure of this type (Rao *et al.*, 1986). Recently a monoclonal antibody raised against a synthetic peptide of C9 (region underlined in Fig. 4b) was shown to recognize C8 α chain (Tschopp & Mollnes, 1986; Tschopp *et al.*, 1986). It is of interest that the corresponding region of the first Factor I repeat is very closely related to the structure of the synthetic peptide (Fig. 4b).

Several other homologies with the Factor I heavy chain sequence were found by searching the protein database. These are shown in Table 1. The homology of mouse α -foetoprotein with Factor I is quite extensive over a stretch of 22 amino acids. A similar homology to α -foetoprotein has been reported in the heavy chain of tPA (Baker, 1985). Strong homologies over short segments were also observed between Factor I, a phlebovirus protein and herpes simplex virus (HSV) type 2 glycoprotein D. The latter homology is interesting in view of the known interaction of glycoprotein C of HSV type 2 with the complement system (Friedman *et al.*, 1984).

The significance of the various homologous segments in Factor I heavy chain are not yet interpretable in terms of the functions of this part of the protein. However, Factor I is a good example of a 'mosaic' protein, as discussed in general by Doolittle (1985) and specifically for the serine proteinases by Patthy (1985). The arrangement of recognized homologous segments in Factor I is summarized in Fig. 4(c), and compared with the arrangement in other proteins.

The complete amino acid sequence of the two cofactors, Factor H and C4-binding protein, are now

available, as are the complete sequences of the Factor I substrates (Reid, 1986). With the amino acid sequence now available for Factor I it should be possible to probe more deeply into the specificity of the interactions between the enzyme, the cofactor proteins and the substrates.

We would like to thank Thakor Patel for making the mixed oligonucleotides and Bob Lamb for supplying the long oligonucleotide probe. Drs. Tertia Belt, Duncan Campbell and David Bentley kindly allowed access to the human liver cDNA libraries. We are also grateful to many colleagues for helpful discussion and advice and to Gill Alcock and Jacquie Butler for typing the manuscript. A.J.D. is a recipient of an SERC CASE Award.

REFERENCES

- Arlaud, G.-J., Willis, A. C. & Gagnon, J. (1987) *Biochem. J.* **241**, 711–720
- Baker, M. E. (1985) *FEBS Lett.* **182**, 47–52
- Belt, K. T., Carroll, M. C. & Porter, R. R. (1984) *Cell* **36**, 907–914
- Bentley, D. R. (1986) *Biochem. J.* **239**, 339–345
- Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids. Res.* **7**, 1513–1523
- Crossley, L. G. (1981) *Methods Enzymol.* **80**, 112–124
- Davis, A. E. (1981) *FEBS Lett.* **134**, 147–150
- Davis, R. W., Botstein, D. & Roth, J. R. (1980) *Advances in Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Dayhoff, M. O. (ed.) (1978) *Atlas of Protein Sequence and Structure*, vol. 5, Suppl. 3, pp. 86–87, National Biomedical Research Foundation, Washington DC
- de Bruijn, M. H. L. & Fey, G. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 708–712
- Docherty, A. J. P., Lyons, A., Smith, B. J., Wright, E. M., Stephens, P. E. & Harris, T. J. R. (1985) *Nature (London)* **318**, 66–69
- Doolittle, R. F. (1985) *Trends Biochem. Sci.* **10**, 233–237
- Foster, D. & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4766–4770
- Friedman, H. M., Cohen, G. H., Eisenberg, R. J., Seidel, C. A. & Cines, D. B. (1984) *Nature (London)* **309**, 633–635
- Gershenfeld, H. K. & Weissman, I. L. (1986) *Science* **232**, 854–858
- Goldberger, G., Aden, D., Kay, R. & Colton, H. R. (1983) *Immunobiology* **164**, 247 (abstr.)
- Hewett-Emmett, D., Czelusniak, J. & Goodman, M. (1980) *Ann. N.Y. Acad. Sci.* **370**, 511–527
- Hayward, B. E., Hussain, A., Wilson, R. H., Lyons, A., Woodcock, V., McIntosh, B. & Harris, T. J. R. (1986) *Nucleic Acids. Res.* **14**, 999–1008
- Isackson, P. J., Ullrich, A. & Bradshaw, R. A. (1984) *Biochemistry* **23**, 5997–6002
- Lathe, R. (1985) *J. Mol. Biol.* **183**, 1–12
- Leytus, S. P., Chung, D. W., Kisiel, W., Kurachi, K. & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3699–3702
- Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. H. & Bleackley, R. C. (1986) *Science* **232**, 858–861
- Long, G. L., Belagaje, R. M. & MacGillivray, R. T. A. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5653–5656
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning – A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Nishimukai, H. & Tamaki, Y. (1986) *Hum. Hered.* **36**, 195–197
- Orcutt, B. C. & Dayhoff, M. O. (1984) *PIR Report PRO-0484*, National Biomedical Research Foundation, Washington DC
- Orcutt, B. C. & Dayhoff, M. O. (1985) *PIR Report MAT-0285*, National Biomedical Research Foundation, Washington DC
- Pangburn, M. K., Shreiber, R. D. & Muller-Eberhard, H. J. (1977) *J. Exp. Med.* **146**, 257–270
- Patthy, L. (1985) *Cell* **41**, 657–663
- Rao, A. G., Howard, M. Z., Ng, S. N., Snider, J. V., Whitehead, A. S., Colten, H. R. & Sodetz, J. M. (1986) *Abstr. Int. Congr. Immunol.* 6th, abstr. no. 2.54.44 p.197
- Reid, K. B. M. (1986) *Essays Biochem.* **29**, 27–54
- Rigby, P. W. J., Deickmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251
- Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
- Southern, E. (1975) *J. Mol. Biol.* **98**, 503
- Staden, R. (1982) *Nucleic Acids Res.* **10**, 4731
- Stanley, K. K., Page, M., Campbell, A. K. & Luzio, J. P. (1986) *Mol. Immunol.* **23**, 451–458
- Straßburger, W., Wollmer, A., Pitts, J. E., Glover, I. D., Tickle, I. J., Blundell, T. L., Steffens, G. J., Gunzler, W. A., Otting, F. & Flohe, L. (1983) *FEBS Lett.* **157**, 219–223
- Tschopp, J. & Mollnes, T. E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4223–4227
- Tschopp, J., Mansson, D. & Stanley, K. K. (1986) *Nature (London)* **322**, 831–834
- Verheij, H. M., Westerman, J., Sternby, B. & de Haas, G. H. (1983) *Biochim. Biophys. Acta* **747**, 93–99
- Wallace, R. B., Schold, M., Johnson, M. J., Demberk, P. & Itakura, K. (1981) *Nucleic Acids Res.* **9**, 3647–3656
- Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164
- Woodbury, R. G. & Neurath, H. (1980) *FEBS Lett.* **114**, 189–196
- Yuan, J. M., Hsiung, L. & Gagnon, J. (1986) *Biochem. J.* **233**, 339–345