Human Complement Factor I: Analysis of cDNA-derived Primary Structure and Assignment of Its Gene to Chromosome 4*

(Received for publication, January 29, 1987)

Gabriel Goldberger‡§, Gail A. P. Bruns‡, Miriam Rits‡, Michael D. Edge‡, and David J. Kwiatkowski**

From the §Laboratory of Human Biochemistry and Genetics Division, Children's Hospital Corporation, Boston, Massachusetts 02115, **Department of Hematology-Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, and Imperial Chemical Industries PLC, Pharmaceuticals Division, Cheshire SK10 4TG, United Kingdom

Factor I is a serine proteinase of complement which together with one of several specific cofactors cleaves activation products of the third and fourth components of complement (C3b and C4b) and modulates the activity of C3 convertase. A heterodimer glycoprotein (M, = 88,000), factor I is synthesized as a single-chain precursor, prepro-I, which undergoes intracellular proteolytic processing. The human hepatoma line HepG2, however, secretes predominantly the single-chain precursor pro-I. In order to determine the molecular basis for this apparent processing defect, factor I cDNA clones were isolated from a HepG2 mRNA-derived library. Sequencing of the largest insert, H11971, revealed that it contains 14 base pairs of 5′ untranslated region, the complete coding sequence for the 583-residue prepro-I (NH₂-signal peptide-heavy chain-linking peptide-light chain-COOH), two polyadenylation signals within the 200-base pair 3′ untranslated region, and a portion of poly(A) tail. Analysis of the derived protein structure 1) reveals a mosaic multidomain structure of the heavy chain; 2) demonstrates structural similarity between intracellular conversion of pro-I and activation of other serine proteinase zymogens; and 3) indicates that the light chain of factor I resembles most closely the active subunit of tissue plasminogen activator among all serine proteinases. Furthermore, this protein sequence was compared to the sequences of factor I cDNA clones isolated from normal human liver libraries and found to be identical. By exclusion, this defines as cellular the basis for the inefficient processing of pro-I by the HepG2 line. Chromosomal localization by the somatic cell hybrid method maps the factor I gene to chromosome 4.

* This work was supported by National Institutes of Health Grants GM 32293, HD 16858, HL 01852, and AI 124836. Grant 1 U41 RR 01685 was used for computer resources provided by the BIONET National Computer Resource for Molecular Biology and MBCR at Dana-Farber Cancer Institute, Boston, MA. Computer resources were provided by the BIONET National Computer Resource for Molecular Biology and Molecular Biomedical Computer Research Resource at Dana-Farber Cancer Institute-Harvard School of Public Health, Boston, MA, supported by Grant 1 U41 RR 01685. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J02770.

‡ To whom correspondence should be addressed: Laboratory of Human Biochemistry, Children's Hospital Corp., 320 Longwood, Boston, MA 02115.

Factor I (C3b/C4b inactivator) is a regulatory serine proteinase of the complement cascade. Together with one of several specific cofactors, it cleaves activation products of C3 (C3b) or C4 (C4b) and modulates (by cleavage of C3b) the amplified activation of C3 in the alternative pathway (1–3). It is thus indispensable in the generation of many C3 and C4 fragments and their activities. As a result, hereditary factor I deficiency in humans is characterized by an increased susceptibility to bacterial infections due to hypercatabolism of C3 and abnormalities of many complement functions (4).

Factor I is a heterodimer composed of disulfide-linked heavy (M, = 50,000) and light (M, = 38,000) chains (5–7) with a carbohydrate content of 11–27% (5, 8). It is present in plasma as an active enzyme at a concentration of ~3.5 mg/dl (5). A partial NH₂-terminal sequence of the heavy chain and the near complete sequence of the light chain have been determined (9, 10). The light chain sequence shows blocks of homology with the active subunits of other serine proteinases. However, unlike other serine proteinases of coagulation, fibrinolysis, and complement (with the exception of complement factor D) it is present in the circulation as an active enzyme (11). Recently, a polymorphism of this protein in the Japanese population has been described (12, 13).

Biosynthesis of human factor I has been reported in liver and primary monocyte cultures (8, 14, 15). A detailed analysis of its biosynthesis under cell-free conditions and in three human hepatoma lines indicates that it is synthesized as a single-chain precursor and undergoes proteolytic processing and N-linked glycosylation (8). Each of three hepatoma lines, HepG2, Hep3B, and PLC/P1/F5, secretes both precursor and mature forms, but unlike many other proproteinases, pro-I has not been identified in plasma (16, 17). Moreover, variation in the proportion of factor I secreted as pro-I was observed among various cell lines, and in one of the lines, HepG2, pro-I is the predominant secreted form. Analysis of the processing of other single-chain precursors by these hepatoma lines showed that the ratios of secreted, completely processed, partially processed, and/or proprotein forms did not vary among them, indicating that this phenomenon is protein-specific. Xenopus oocytes injected with HepG2 mRNA secrete, however, primarily the processed two-chain protein. Thus, the basis for the inefficient processing of pro-I by the HepG2 line has not been determined and may be intrinsic to the produced factor I or the proteolytic processing mechanism itself.

In order to test whether a structural defect is the cause of this phenomenon and to permit a further analysis of factor I
structure, we isolated factor I cDNA clones from HepG2 and normal human liver libraries. These clones were used to determine the nucleotide sequence of the factor I mRNA, derive the complete primary structure of its translation product, and determine the chromosomal localization of the factor I gene.

MATERIALS AND METHODS

Oligonucleotide Preparation—The oligonucleotide mixtures for the library screening and sequence-specific primers were synthesized by automated phosphoramidite chemistry on a DNA synthesizer model 380B (Applied Biosystems), deblocked by incubation with 1 volume of 10 N ammonium hydroxide for 6 h at 55 °C, and then lyophilized. Some of the oligonucleotides were further purified by acrylamide gel electrophoresis, electroelution, desalting, and lyophilization. For the Northern blot hybridization and cDNA library screening the probe was labeled with \( ^{32}P \)dATP (Du Pont-New England Nuclear) by the T4 polynucleotide kinase method (18).

Identification of the Factor I mRNA—Cytoplasmic RNA from the human hepatoma cell line HepG2 was isolated by phenol/chloroform extraction and analyzed by formaldehyde agarose gel electrophoresis and Northern blot hybridization using the \( ^{32}P \)-labeled oligonucleotide mixture as a probe (19).

Isolation of Factor I cDNA Clones—A λgt10 cDNA library prepared from the human hepatoma line HepG2 mRNA (20) was screened to isolate factor I clones (21). The screening was done with a \( ^{32}P \)-labeled oligonucleotide mixture under stringency calculated for the sequence with a minimal G/C content (19). Positive clones were plaque-purified, and the phage DNA were isolated by the plate-lysis method (19). The HI1971 insert was isolated by low-melt agarose gel purification and subcloned into the pUC-18 vector (22).

Nucleotide Sequence Analysis—The HI1971 insert was subcloned into the M13mp18 vector (23), and each strand was sequenced using BIONET.

Chromosomal Localization of Factor I Gene—DNA from human-mouse and human-hamster hybrid cells containing known complements of human chromosomes (26, 27) was analyzed by Southern blotting hybridization (19) for human factor I sequences using the pUC-18-derived insert labeled by the random hexanucleotide method (28).

RESULTS

Isolation of Factor I cDNA Clones and Determination of the Derived mRNA and Protein Structures—Based upon the amino acid sequence of the NH2-terminal region of the light chain (10), a degenerate 23-mer complementary oligonucleotide probe was synthesized (Fig. 1A). When the \( ^{32}P \)-labeled mixture was hybridized to a Northern blot containing HepG2 RNA, a single band corresponding to an mRNA of ~2.3 kilobases was identified (Fig. 1B). (The identical result was obtained using the cDNA probe as well.)

The same \( ^{32}P \)-labeled oligonucleotide probe was used to screen 6 \( \times \) 10\(^5\) phage of the human hepatoma HepG2 cDNA library. Thirteen primary isolates were purified by serial plating and screening, and the phage DNA were isolated. Digestion of the phage DNA with EcoRI permitted assessment of the insert size in the positive clones. The largest insert, derived from the \( \lambda \)cH1971 clone, was subcloned into M13mp18 and analyzed by the dideoxy method using universal and sequence-specific primers (Fig. 1C). Its complete 1971-base pair sequence comprises ~85% of the factor I mRNA and read in the single open reading frame contains 68 bases upstream of the region coding for the single-chain propeptide, pro-I, followed by 200 bases of the 3′ untranslated region and 8 bases of the poly(A) tail of the factor I mRNA (Fig. 2).

Within the 68-base pair 5′ region there is an in-frame methionine codon beginning at base 13 which delineates a peptide with the following features characteristic of a signal peptide (29): (a) length of 18 residues; (b) a high proportion (67%) of hydrophobic residues with an uninterrupted stretch in the middle portion; and (c) a cysteine residue at its end. In addition, the pentanucleotide CCAC preceding the methionine codon is similar to the translation initiation consensus sequence CC-CC (30). These features strongly suggest that this peptide represents the entire signal peptide and that the HI1971 insert, therefore, contains sequence coding for the entire primary translation product of factor I mRNA, prepro-I, consisting of 583 residues. The 200-base pair 3′ untranslated region contains two polyadenylation signals (AATAAA) starting at positions −140 and −22 from the polyadenylation tail. This implies a possible heterogeneity in the length of factor I mRNA as has been observed for complement C1r and other plasma protein mRNAs (31). Comparison of the size of this insert and the estimated size of factor I mRNA (~2.3 kilobases) suggests that the latter contains an ~300-base 5′ untranslated region.

The organization of the factor I subunits within the single chain precursor, NH2-heavy chain-light chain-COOH (Fig. 1C), is similar to that of other proteasezymogens of the coagulation, fibrinolytic, and complement systems (1, 22, 32, 33). However, in contrast to the latter which contain a single basic residue at the site of cleavage/activation, pro-I contains a sequence of four basic amino acids (Arg-Arg-Lys-Arg) between its subunits. This is characteristic of linking peptides in several single-chain complement (pro-C3, pro-C4, and pro-C5) and retroviral (env) protein precursors but not mature proteins (34, 35). While the COOH-terminal amino acid analysis of some of these proteins shows a complete removal of the linking peptide (36), in at least one, human coagulation profactor X, the first arginine residue of an identical tetrapeptide is retained as the carboxyl terminus of the NH2-terminal subunit in the mature protein (37). An unambiguous number of residues (and amino acid composition) could, therefore, be estimated only for the light (244) but not the heavy chain (<321).

The predicted size of pro-I, 63.5 kDa (568 residues), is in good agreement with the estimates for the size of the nonglycosylated pro-I (\( M_r = 65,000 \)) and the carbohydrate content of 27% (8). The determined amino acid composition of the heavy and light chains (heavy/light) is: alanine (17/13), aspartic acid (21/15), glycine (22/24), histidine (7/5), isoleucine (11/20), leucine (21/11), lysine (<27/15), methionine (7/3), phenylalanine (13/9), proline (8/11), serine (22/17), threonine (22/10), tryptophan (2/10), tyrosine (8/14), and valine (19/19).

The ~23-kDa difference between the sizes of unglycosylated and the fully glycosylated pro-I has been determined to be due to N-linked glycosylation (8). Three potential N-linked glycosylation sites (Asn-X-Thr/Ser (38)) are present in each of the chains: heavy, residues 52, 86, and 159, and light, residues 125, 155, and 197. Though generally only a third of the potential sites is glycosylated and the aspartic acid residue following Asn\(^{65} \) in the light chain indicates a less likely site of glycosylation, utilization of most of the acceptor sites would be required to account for the estimated carbohydrate content. Both chains are likely to be glycosylated since the sizes of each observed by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (50 and 38 kDa) are larger than those predicted by the sequence data (36 and 27 kDa).

\(^2\) Communicated by Drs. J.-M. Yuan and J. Gagnon prior to publication.
Mosaic Structure of Complement Factor I

\[
\begin{align*}
\text{A} & \quad \text{NH}_2- W - Q - V - A - I - K - D - A - \text{COOH} \\
\text{B} & \quad \text{Factor I} \\
\text{C} & \quad \text{Prepro I}
\end{align*}
\]

Comparison of the derived amino acid sequence with that reported for 30 NH\textsubscript{2}-terminal residues of heavy chain (9) shows identity with the exception of Gln\textsuperscript{12} which is glutamic acid in our sequence. This discrepancy might represent a post-translational modification of the heavy chain, while the protein data show glutamic acid in our sequence. This discrepancy might represent a post-translational modification of the heavy chain.
the penultimate asparagine residue in that position.

Comparison of the Factor I Structure Derived from the Nucleotide Sequences of HepG2 and Adult Human Liver cDNA Clones—To determine whether the structure of pro-I produced by the HepG2 line is different from the normal, we isolated factor I cDNA clones from a normal human adult liver cDNA library (39). Comparison of the derived amino acid sequence of prepro-I synthesized by HepG2 with partial sequences derived from the clones isolated from this library and a complete sequence derived from clones isolated from another human liver library (56) demonstrates complete identity (data not shown). The identity of the factor I sequences derived from the HepG2 and human liver cDNA clones suggests by exclusion that the inefficient processing of pro-I by the HepG2 line is the result of a cellular processing defect and not a change in the protein. This is in contrast to the known cases of hyperproinsulinemia which are caused by mutations at the cleavage sites (40). The fact than no variation has been observed in the processing of other single-chain precursors by the three hepatoma cell lines (8) further suggests that the proteolytic processing of single-chain precursors might be protein-specific.

Chromosomal Assignment of Human Factor I Gene—To determine the chromosomal localization of the factor I gene, a panel of DNA isolated from human-mouse and human-hamster somatic cell hybrids containing known complements of human chromosomes (26, 27) was analyzed for the presence of factor I sequences. This was accomplished by a Southern blot hybridization analysis for human factor I gene sequences in the somatic cell hybrids. Genomic DNA isolated from human-tissue (lanes 1 and 2) human-hamster, and human-mice hybrids containing (lanes 2 and 5) and lacking human chromosome 4 (lanes 3 and 6) were digested with HindIII (lanes 1–3) and SacI restriction endonucleases (lanes 4–6). Combined patterns of human and rodent factor I sequences are observed for the DNA isolated from hybrids containing human chromosome 4 (lanes 2 and 5). Patterns identical to the DNA hybrids lacking human chromosome 4 were observed also for the DNA isolated from rodent nonhybrid cells. On the left positions of the HindIII digestion fragments of λ phage are shown. kbp, kilobase pairs.

![FIG. 3. Southern blot hybridization analysis for human factor I gene sequences in the somatic cell hybrids. Genomic DNA isolated from human tissue (lanes 1 and 2) human-hamster, and human-mice hybrids containing (lanes 2 and 5) and lacking human chromosome 4 (lanes 3 and 6) were digested with HindIII (lanes 1–3) and SacI restriction endonucleases (lanes 4–6). Combined patterns of human and rodent factor I sequences are observed for the DNA isolated from hybrids containing human chromosome 4 (lanes 2 and 5). Patterns identical to the DNA hybrids lacking human chromosome 4 were observed also for the DNA isolated from rodent nonhybrid cells. On the left positions of the HindIII digestion fragments of λ phage are shown. kbp, kilobase pairs.](image-url)

**TABLE I**

| Chromosome | Hybridization pattern | Discordant fraction |
|------------|-----------------------|--------------------|
| 1          | +/+                  | 0.25               |
| 2          | +/−                  | 0.41               |
| 3          | −/−                  | 0.39               |
| 4          | +/−                  | 0.39               |
| 5          | −/−                  | 0.41               |
| 6          | +/−                  | 0.48               |
| 7          | −/−                  | 0.55               |
| 8          | +/−                  | 0.52               |
| 9          | −/−                  | 0.33               |
| 10         | +/+                  | 0.62               |
| 11         | +/−                  | 0.40               |
| 12         | −/−                  | 0.47               |
| 13 and 13q+ | +/−     | 0.40               |
| 14         | +/−                  | 0.40               |
| 15         | −/−                  | 0.40               |
| 16         | +/−                  | 0.35               |
| 17         | −/−                  | 0.35               |
| 18         | +/−                  | 0.50               |
| 19 and 19q+ | +/−     | 0.60               |
| 20         | +/−                  | 0.58               |
| 21         | +/−                  | 0.57               |
| 22         | +/−                  | 0.42               |
| X and Xq− | +/−                  | 0.30               |
| Y          | −/−                  | 0.21               |

*a The human chromosome complements of the hybrids were determined by isozyme and cytogenetic techniques (26, 27). Hybrid DNAs were also analyzed with DNA probes for each of the human autosomes and the X chromosome.

*b The column designations refer to presence (+/) and absence (−/) of human factor I sequence or presence (+/) and absence (−/) of a human chromosome.

*c For calculation of discordant fractions, hybrids with a rearranged chromosome or in which the chromosome was present in less than 15% of the cells or the characteristic isozyme or DNA probe was weakly positive were excluded.

*d For the single clone derived from fusion with fibroblasts from the X/18 translocation carrier, this category represents the der18 chromosome.

*e For the 17 hybrid clones derived from fusions with leukocytes from the two different X/19 translocation carriers, this category represents the der19 translocation chromosomes.

*f The X category includes hybrids with an intact X and those with derX translocation chromosomes.

![FIG. 4. Comparison of chymotrypsinogen activation (I) and pro-I processing (II). The NH₂-terminal region of the chymotrypsinogen is aligned with a segment of pro-I (COOH terminus of the heavy chain-linking peptide-NH₂ terminus of the light chain). The arrows indicate peptides cleaved off during the activation of chymotrypsinogen and conversion of pro-I to native factor I. Identical residues in the chymotrypsinogen and heavy chain of factor I and the active subunits in both proteins are underlined. (For the sake of comparison the numbering of factor I light chain Cys114 residue presumed to form the disulfide interchain link is given as 435 according to its position in pro-I.)](image-url)
chromosome 1 (0.75). Only a single hybrid was discordant for factor I and chromosome 4, likely due to rearrangement of chromosome 4. No other complement protein nor a serine proteinase has so far been mapped to this chromosome. This indicates that the factor I gene is not a member of a cluster of genes for other regulatory complement proteins (C4BP, CR1, and H) located on chromosome 1 (41).

**DISCUSSION**

Analysis of a human complement factor I cDNA clone isolated from a HepG2 library has resulted in a determination of the structure of factor I mRNA and derivation of a complete amino acid sequence of its translation product prepro-I. It has permitted characterization of the signal peptide, determination of the organization of subunits in the single-chain precursor, and assignment of the factor I gene to chromosome 4. Consistent with biosynthetic studies, structural analysis of prepro-I indicates that factor I is synthesized as a single-chain precursor, which is processed by two proteolytic steps (prepro-I to pro-I to native I) and an N-linked glycosylation. The noncatalytic subunits of serine proteinases are known to contain a variety of domains (32, 33). The heavy chain of factor I contains 29 cysteine residues (11%) and hence is capable of forming 14 intrachain disulfide links which are likely to contribute to domain structure. Accordingly, we analyzed (by BIONET and manually) the primary structure for sequence homology with other proteins and propose six domains in factor I heavy chain (Fig. 5, A and B).

Three of these domains, II (26–60), V (196–238), and VI (239–276) (Fig. 5A), are homologous to the ~40-residue domain found once in another complement protein C9 (44) and repeated seven times in tandem in the binding region of the LDL receptor (45, 46). As defined in these proteins, each domain contains overall 6 conserved cysteines among other consensus residues. Though amino acid sequence homology between complement proteins interacting with factor I and the ligands of the LDL receptor (apoE and apoB) has not been found (47, 48), the presence of these domains in factor I might imply a higher order structural similarity among them.

Domain III (71–136) contains 7 out of 10 conserved residues of an ~60-residue sequence found (primarily) in tandem repeats (from 1 up to 20 times or more) in seven complement (factors B and H, C2, C1r, C4BP, CR1, and CR2) and four noncomplement proteins (β2-glycoprotein I, interleukin-2 receptor, coagulation factor XIII, and haptoglobin), as shown in Fig. 5B (41, 49, 50). In addition, this domain contains 6 residues of a consensus sequence found in the repeats of at least three of these proteins (C4BP, factors B and H, β1 and/or coagulation factor XIII) (57). The presence of this domain among all of these proteins does not necessarily imply an unknown common function. It might be that the conserved residues constitute a structural framework for the unique residues defining the functional specificity.

The heavy chain also contains six basic di- and tripeptides (Fig. 6). The functional significance of these peptides is not certain, but they may contribute to domain delineation or represent sites of cleavage/inactivation or binding.

The disulfide linkages of the heavy chain may be predicted only for domain III (Fig. 6). We presume that disulfide linkages within II, V, and VI are confined within each domain. The remaining 6 cysteine residues we propose form linkages to generate the simplest secondary structure (Fig. 6). Four of these (by analogy with C1r domains I and III) would define distinct domains I and IV (31). The remaining pair then

---

3 The abbreviation used is: LDL, low density lipoprotein.
would link the outer regions of domain III and together with the other two disulfide bridges form a kringle-like module found in other serine proteinases (32). This proposition is supported by the fact that a haptoglobin region containing the consensus residues of domain III (49) is also similar to the kringles of other serine proteinases (52).

The protein sequence was further analyzed for similarities with other proteins and internal homology using BIONET. A remarkable 9-out-of-10-residue match was found between a decapeptide starting at position 78 (domain III) and a region in the G2 protein of Punta Toro Phlebovirus (Fig. 5C) with a respective nucleotide match of 73% (53). The protein similarity suggests a possible function for that region in factor I and/or a pathophysiological role of that region in the virus.

Analysis for internal homology demonstrates three pairs of identical tetrapeptides (Fig. 5, D–F). The significance of these observations is not clear, but the respective locations of the first pair of tetrapeptides in the heavy chain at the beginning and the end of domain III (Figs. 5D and 6) and, additionally, the degree of nucleotide homology (11 out of 12) suggests a remnant of a duplicated domain III.

The complete sequence of the light chain of factor I is, as expected (10), homologous to the catalytic subunits of other serine proteinases with conservation of 26 of 29 invariant residues (42). It is closest in sequence to tissue plasminogen activator (41% identical residues), plasma kallikrein (37%), and coagulation factor XII (34%) among all serine proteinases and to factor D (28%) among the complement proteins (11, 25, 54, 55). The substitutions of Gly\(^6\) and Leu\(^{106}\) (chymotrypsinogen numbering) by Thr\(^6\) and Met\(^{106}\), respectively, might be significant for the higher order structure and hence enzymatic specificity of factor I, while the third substitution, Val\(^{50}\) for Ile\(^{51}\), is a conservative one. Though the light chain structurally and functionally resembles trypsin (9), the placement of its cysteine residues (and hence its disulfide linkage) and other structural features (Fig. 4) demonstrate unique chymotrypsin features as well. The disulfide linkages for the 11 cysteine residues of the light chains are, therefore, proposed by analogy with chymotrypsin and are identical to t-PA (Fig. 6).

In summary, we propose that factor I contains up to six domains within its heavy chain. Three of its domains (II, V, and VI) are homologous to the C9-like domains of the LDL receptor but have not been found in other serine proteinases. This together with other features (synthesis of an active enzyme, solitary presence of its gene on chromosome 4) indicate that factor I is a highly diverged and specialized member of the serine proteinase family. By analogy with the other multidomain proteins that evolved via an exon-shuffling mechanism (46), the domains of factor I gene are expected to be encoded by distinct exons.

Acknowledgments—We thank Drs. H. R. Colten for support, S. B. Dowton and J. Gitlin for assistance in preparing the oligonucleotide probes, D. H. Bing for insights into serine proteinase structure, E. Hoffmann and T. F. Smith for help in the analysis of derived protein sequence, and F. S. Rosen for critical reading of the manuscript.

REFERENCES

1. Reid, K. B. M., and Porter, R. R. (1981) Annu. Rev. Biochem. 50, 433–464
2. Arnaout, M. A., and Colten, H. R. (1984) Mol. Immunol. 21, 1191–1199
3. Lambiris, J. D., and Tsokos, G. C. (1985) Anticancer Res. 5, 515–524
4. Abrahamson, N. L., Alper, C. A., Lachman, P. J., Rosen, F. S., and Jandl, J. H. (1971) J. Immunol. 107, 19–27
5. Pangburn, M. K., Schreiber, R. D., and Muller-Eberhard, H. J. (1977) J. Exp. Med. 146, 257–270
6. Fearon, D. T. (1977) J. Immunol. 119, 1248–1252
7. Crossley, D. L., and Porter, R. R. (1980) Biochem. J. 191, 173–182
8. Goldberger, G., Arnaout, M. A., Kay, R., Rits, M., and Colten, H. R. (1984) J. Biol. Chem. 259, 6492–6497
9. Davis, A. E., III (1981) FEBS Lett. 134, 147–150
10. Yuan, J.-M., Hsu, L.-M., and Gagnon, J. (1986) Biochem. J. 233, 339–345
11. Niemann, M. A., Bhow, A. S., Bennett, J. C., and Volanakis, J. E. (1984) Biochemistry 23, 2482–2486
12. Nakamura, S., and Abe, K. (1985) Hum. Genet. 71, 45–48
13. Nishimukai, H., and Tamaki, Y. (1986) Hum. Hered. 36, 185–197
14. Whaley, K. (1980) J. Exp. Med. 151, 501–516
15. Morris, K. K., Aden, D. P., Knowles, B. E., and Colten, H. R. (1982) J. Clin. Invest. 70, 906–913
16. Steiner, D. F., Docherty, K., and Carroll, R. (1984) J. Cell Biochem. 24, 121–130
17. Minta, J. D., Ngan, B. Y., and Pang, A. S. D. (1979) J. Immunol. 122, 2415–2420
18. Maksam, A., and Gilbert, W. (1980) Methods Enzymol. 65, 497–509
19. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
20. Kwiatkowski, D. J., Stossel, T. P., Orkin, S. H., Mole, J. E., Colten, H. R., and Yin, H. L. (1986) Nature 323, 455–458
21. Ginsburg, D., Handin, R. I., Bonthron, D. T., Donlon, T. A.,
Mosaic Structure of Complement Factor I

10071

Bruno, G. A. P., Latt, S. A., and Orkin, S. H. (1985) *Science* **228**, 1401-1406

22. Vieira, J., and Messing, J. (1982) *Gene* (Amst.) **19**, 259-268

23. Messing, J. (1983) *Methods Enzymol.* **101**, 20-78

24. Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3960-3965

25. McMullen, B., and Fujikawa, K. (1985) *J. Biol. Chem.* **260**, 5328-5341

26. Brunns G. A. P., Mintx, B. J., Leary, A. C., Regina, V. M., and Gerald, P. S. (1979) *Biochem. Genet.* **17**, 1031-1059

27. Kanda, N., Alt, F., Schreck, R. R., Brunns G. A. P., Baltimore, D., and Latt, S. A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 4069-4073

28. Feinberg, A. P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266-267

29. von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17-21

30. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857-872

31. Leitus, S. F., Kurachi, K., Sakariassen, K. S., and Davie, E. W. (1986) *Biochemistry* **25**, 4855-4863

32. Patchy, L. (1985) *Cell* **41**, 657-663

33. Neurath, H. (1985) *Fed. Proc.* **44**, 2907-2913

34. Lundwall, A. B., Wetsel, R. A., Kristensen, T., Whitehead, A. S., Woods, D. E., Ogden, R. C., Colten, H. R., and Tack, B. F. (1985) *J. Biol. Chem.* **260**, 2108-2112

35. Sonigo, P., Alizon, M., Staakse, K., Klettmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Hasse, A., and Wain-Hobin, S. (1986) *Cell* **42**, 369-382

36. Tack, B. F., Janatova, J., Thomas, M. L., Harrison, R. A., and Hammer, C. H. (1981) *Methods Enzymol.* **80**, 64-101

37. Leytus, S. P., Chung, D. W., Kiasel, W., Kurachi, K., and Davie, E. W. (1984) *Proc. Natl. Sci. U. S. A.* **81**, 3699-3702

38. Kornberg, R., and Kornberg, S. (1986) *Annu. Rev. Biochem.* **54**, 631-664

39. Woods, D. W., Markham, A. F., Ricker, A. T., Goldberger, G., and Colten, H. R. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 5661-5665

40. Gabby, H. K. (1980) *N. Engl. J. Med.* **302**, 165-167

41. Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T., and Tack, B. F. (1986) *Immunol. Today* **7**, 230-234

42. Young, C. L., Barker, W. C., Tomaselli, C. M., and Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structures* (Dayhoff, M. O., ed) Vol. 5, Suppl. 3, pp. 77-93, National Biomedical Research Foundation, Wash. D. C.

43. Barnum, S. R., and Volanakis, J. E. (1985) *J. Immunol.* **134**, 1799-1803

44. Stanley, K. K., Kocher, H.-P., Luzio, J. P., Jackson, P., and Tchopp, J. (1988) *EMBO J.* **4**, 375-382

45. Sudhof, T. C., Goldstein, J. L., Brown, M. S., and Russell, D. W. (1985) *Science* **228**, 815-822

46. Sudhof, T. C., Russell, D. W., Goldstein, J. L., Brown, M. S., Sanchez-Pescador, R., and Bell, G. I. (1985) *Science* **228**, 895-896

47. Breslow, J. L., McPherson, J., Nusbaum, A. L., Williams, H. W., Loquast-Kahl, F., Karathanasis, S. K., and Zannis, V. I. (1985) *J. Biol. Chem.* **254**, 14639-14641

48. Cladaras, C., Hadzopoulou-Cladaras, M., Nolte, R. T., Atkinson, D., and Zannis, V. I. (1986) *EMBO J.* **5**, 3455-3507

49. Ichinoe, A., McMullen, B. A., Fujikawa, K., and Davie, E. W. (1986) *Biochemistry* **25**, 4633-4638

50. Weiss, J. J., Fearon, D. T., Klickstein, L. B., Wong, W. W., Richards, S. A., de Bruyn Kops, A., Smith, J. A., and Weiss, J. H. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 5639-5643

51. Lozier, J., Takahashi, N., and Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 3640-3644

52. Hewitt-Emmett, D., Czelusniak, J., and Goodman, M. (1981) *Ann. N. Y. Acad. Sci.* **370**, 511-527

53. Ihara, T., Smith, J., Dalrymple, J. M., and Bishop, D. H. L. (1985) *Virology* **144**, 246-256

54. Chung, D. W., Fujikawa, K., McMullen, B. A., and Davie, E. W. (1986) *Biochemistry* **25**, 2410-2417

55. Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., and Goeddel, D. V. (1983) *Nature* **301**, 214-221

56. Catterall, C. F., Lyons, A., Sim, R. B., Day, A. J., and Harris, T. J. R. (1987) *Biochem. J.* **242**, 849-856

57. Kristensen, T., Ogata, R. T., Chung, L. P., Reid, K. B. M., and Tack, B. F. (1987) *Biochemistry*, in press