

# Complete cDNA and derived amino acid sequence of human factor V

(cDNA cloning/sequence homology/blood coagulation/gene evolution)

RICHARD J. JENNY\*, DEBRA D. PITTMAN‡, JOHN J. TOOLE‡, RONALD W. KRIZ‡, ROBERT A. ALDAPE‡, RODNEY M. HEWICK‡, RANDAL J. KAUFMAN‡, AND KENNETH G. MANN\*†

\*Department of Biochemistry, College of Medicine, University of Vermont, Burlington, VT 05405; and †Genetics Institute, Cambridge, MA 02140

Communicated by Russell F. Doolittle, March 27, 1987

**ABSTRACT** cDNA clones encoding human factor V have been isolated from an oligo(dT)-primed human fetal liver cDNA library prepared with vector Charon 21A. The cDNA sequence of factor V from three overlapping clones includes a 6672-base-pair (bp) coding region, a 90-bp 5' untranslated region, and a 163-bp 3' untranslated region within which is a poly(A) tail. The deduced amino acid sequence consists of 2224 amino acids inclusive of a 28-amino acid leader peptide. Direct comparison with human factor VIII reveals considerable homology between proteins in amino acid sequence and domain structure: a triplicated A domain and duplicated C domain show ≈40% identity with the corresponding domains in factor VIII. As in factor VIII, the A domains of factor V share ≈40% amino acid-sequence homology with the three highly conserved domains in ceruloplasmin. The B domain of factor V contains 35 tandem and ≈9 additional semiconserved repeats of nine amino acids of the form Asp-Leu-Ser-Gln-Thr-Thr/Asn-Leu-Ser-Pro and 2 additional semiconserved repeats of 17 amino acids. Factor V contains 37 potential N-linked glycosylation sites, 25 of which are in the B domain, and a total of 19 cysteine residues.

Factor V is a large and asymmetric glycoprotein that circulates in plasma and is an essential component of the blood coagulation cascade (1, 2). During coagulation, the procofactor factor V is converted to the active cofactor, factor Va, via limited proteolysis by  $\alpha$ -thrombin (3–5). Factor Va is a cofactor for the serine protease factor Xa, and together, factors Va and Xa assemble on a cellular or phospholipid surface with divalent metal ions to form the prothrombinase complex (1, 6–11). This complex enhances factor Xa activity ≈350,000 fold. The prothrombinase complex is analogous to another complex that proteolytically cleaves zymogen factor X to active enzyme factor Xa—this other “ten-ase” complex is composed of a serine protease (factor IXa), a cofactor (factor VIIIa), phospholipid, and calcium (13, 14).

In addition to the similarities between serine proteases (factors Xa and IXa) and in overall enzyme complex architecture, the cofactors (factor Va and factor VIIIa) are very similar proteins structurally and functionally (13–16). Heavy and light chains of bovine factor Va and porcine factor VIIIa possess amino acid-sequence homology at the amino-terminal portion of the chains—regions of homology that are also homologous to regions in the triplicated domain structure of ceruloplasmin, the primary transport protein for copper in plasma. Available data therefore suggest that factor V, factor VIII, and ceruloplasmin are members of a family of structurally related proteins (15).

The molecular cloning and sequencing of human factor VIII and human ceruloplasmin gives evidence for a common

domain structure and has enabled detailed comparison of their structures (16, 17–20). Recently Kane and Davie (21) published a partial cDNA sequence for human factor V that coded for ≈40% of the molecule. This cDNA coded for the light chain and a small portion of the heavily glycosylated connecting region. We present the complete cDNA and deduced amino acid sequence of human factor V and compare this sequence with the primary structures of factor VIII and ceruloplasmin.

## MATERIALS AND METHODS

**Materials and Reagents.** All DNA-modifying enzymes and cloning vectors were obtained from either New England Biolabs, Bethesda Research Laboratories, or Promega Biotec; all other reagents and supplies were of high quality and are commercially available.

**Screening of Human Fetal Liver cDNA Libraries.** Using amino acid sequence from the amino-terminal portion of human factor Va light chain (15, 22), we synthesized a 39-mer oligonucleotide (AACTAYTAYATTGCTGCTGAGGAGATCACCTGGGACTAT, where: Y = T or C) on an Applied Biosystems DNA synthesizer and subsequently end-labeled it with T4 polynucleotide kinase and [ $\gamma$ <sup>32</sup>-P]ATP (ref. 23, pp. 122–123). This probe was then used to screen an oligo(dT)-primed human fetal liver Charon 21A library (16); clone V1 obtained on the initial screening was used to generate two new probes. Restriction fragments obtained were labeled with [ $\alpha$ -<sup>32</sup>P]ATP by nick translation (ref. 23, pp. 109–112) or random priming (24) and were used to screen for additional clones.

**Isolation, Subcloning, and Sequencing of Insert cDNA.** Phage DNA was prepared from positive clones and digested with appropriate restriction enzymes to isolate cDNA inserts. Inserts were subcloned into the plasmid vector SP65 (Promega Biotec) or M13 phage cloning vector mp18 or mp19 (New England Biolabs) as described by Maniatis *et al.* (ref. 23, pp. 390–401). Nucleotide sequence analysis was done as described by Poncz *et al.* (25), Messing and Vieira (26), and Sanger *et al.* (27). cDNA and protein homologies were studied by computer-assisted searches of Genbank and the National Biomedical Research Foundation protein sequence data base.

**RNA and Southern Blot Analysis.** RNA blot analysis was used to estimate the size of the mRNA and the relative abundance of the message in various cell types and to help confirm full-length clones. mRNA was isolated from cultured cells using the guanidine thiocyanate method (28), and both mRNA and poly(A)<sup>+</sup> mRNA were isolated from human liver as described (29). mRNA was fractionated on a 0.8% agarose-formaldehyde gel and was subsequently transferred to nitrocellulose as described (30). The 1614-bp *Eco*RI restriction

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed.

enzyme fragment was isolated from clone V1 and was used to probe the RNA blots. Similar analysis with other restriction fragments verified factor V clones. Common restriction enzyme fragments in other clones were identified by Southern blot analysis, which was done as described (ref. 23, pp. 383–386).

### RESULTS AND DISCUSSION

Approximately  $10^5$  recombinant  $\lambda$  phage plaques from a human fetal liver cDNA library were screened with the 39-mer oligonucleotide probe derived from amino acid sequencing of the human factor Va light chain, and 14 positive clones were identified. A phage designated V1 was chosen for further analysis. cDNA sequence encoding the amino acids from which the 39-mer was designed confirmed that the clone contained factor V DNA. The cDNA insert in phage V1 was too short to encode the entire factor V molecule. To obtain cDNA encoding the amino terminus of factor V, restriction enzyme fragments from the 5' end of clone V1 were isolated, labeled with  $^{32}\text{P}$ , and used to screen additional recombinant  $\lambda$  phage. Approximately  $10^6$  recombinant phage were screened, and 92 positive clones were identified; two phage designated V401 and V402 were selected for further analysis. Phage V401 contained an *EcoRI* fragment common to both clones as shown by cross-hybridization analysis, but which was longer than the equivalent fragment in phage V402 (as determined by restriction enzyme digestion and subsequent fractionation in agarose gels). Further analysis of this fragment revealed an open reading frame with cDNA sequence encoding the amino-terminal portion of the intact factor V molecule, thus confirming a clone that encoded amino acid sequence beyond the amino terminus of the mature plasma protein.

A schematic representation of the three overlapping clones that were used to complete the cDNA sequence of factor V is shown (Fig. 1); Fig. 2 illustrates cDNA and the deduced amino acid sequence. The cDNA sequence includes a 6672-bp coding region, a 90-bp 5' untranslated region, and a 163-bp

3' untranslated region including a poly(A) tail. The coding region begins at nucleotide 91 with the initiator codon ATG. The 3' untranslated region contains the putative polyadenylation signal sequence AATAAA (31), located 12 nucleotides 5' of the poly(A) tail. The deduced amino acid sequence of factor V consists of 2224 amino acids that include a 28-amino acid leader peptide. The leader peptide contains a cluster of hydrophobic amino acids and an alanine residue at position -1, which is consistent with the structure of most leader peptides and with known specificities of cleavage by leader peptidases (32–34).

Size and abundance of the factor V mRNA from various tissue sources were analyzed by RNA blot hybridization. The results (Fig. 3) indicate that a factor V message of  $\approx 7000$  bp is quite abundant in both liver and HepG2 cells. Prolonged exposure of the RNA blots did not indicate any factor V mRNA contained within the RNA isolated from human umbilical-vein endothelial cells, peripheral blood leukocytes, or U297 cells.

The complete sequence illustrated in Fig. 2 verifies and extends the partial cDNA sequence encoding the 3' half of factor V recently published by Kane and Davie (21). These data indicate that factor V has a domain structure similar to that of factor VIII and that these molecules share  $\approx 40\%$  amino acid sequence identity in the A and C domains. The domains within factor V are schematically illustrated in Fig. 1 and correspond to the following regions: A1, residues 1–317; A2, residues 318–663; A3, residues 1546–1883; B, residues 664–1545; C1, residues 1884–2036; and C2, residues 2037–2196.

A comparison of factor V, factor VIII, and ceruloplasmin by the Dayhoff protein alignment method (35) is represented in Fig. 4. These data indicate highly conserved domains within and between these proteins. In addition, it has also been pointed out that the C domains of factor V and factor VIII share a 20% amino acid-sequence homology with the discoidins, which are phospholipid-binding lectins from *Dicystostelium Discoideum* (21, 36). The B domains of factors V and VIII share a 14% amino acid sequence identity, in

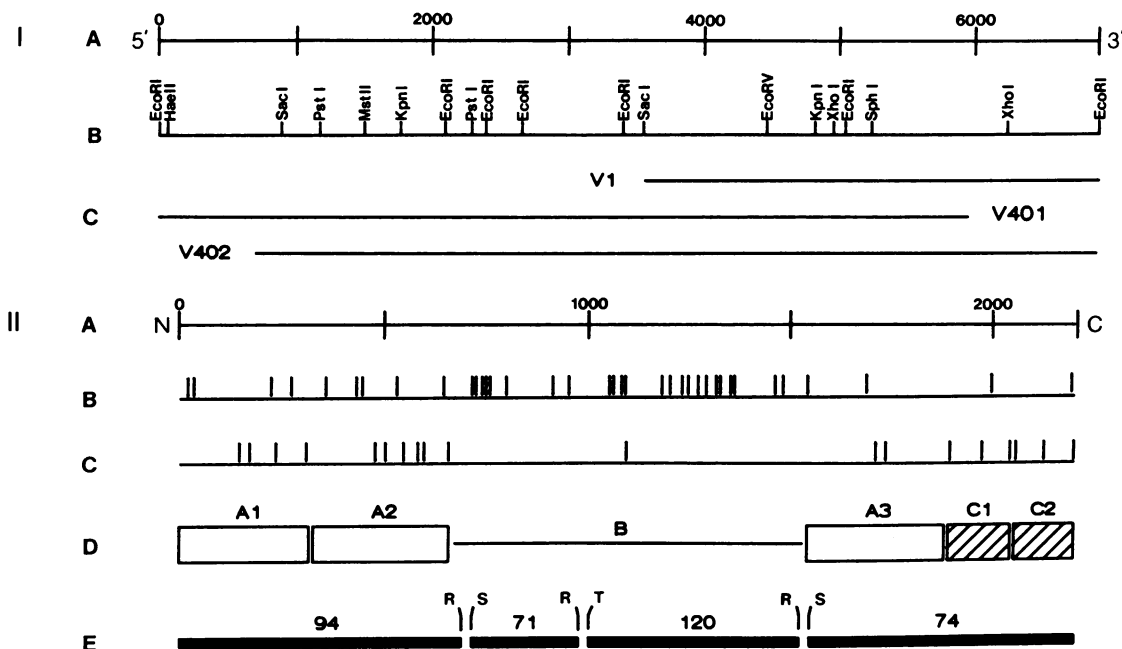


FIG. 1. (I) Line diagram of human factor V cDNA: (A) cDNA number line, (B) partial restriction enzyme map, and (C) three overlapping cDNA clones used to obtain the complete cDNA sequence. (II) Line diagram of human factor V protein: (A) amino acid number line, N, amino terminus, C, carboxyl terminus, (B) potential N-linked glycosylation sites, (C) cysteine residues, (D) domain structure based upon internal homologies and homologies with factor VIII and ceruloplasmin, and (E) thrombin cleavage products labeled according to their putative  $M_r$  values with carbohydrate included. R, Arg; S, Ser; and T, Thr.



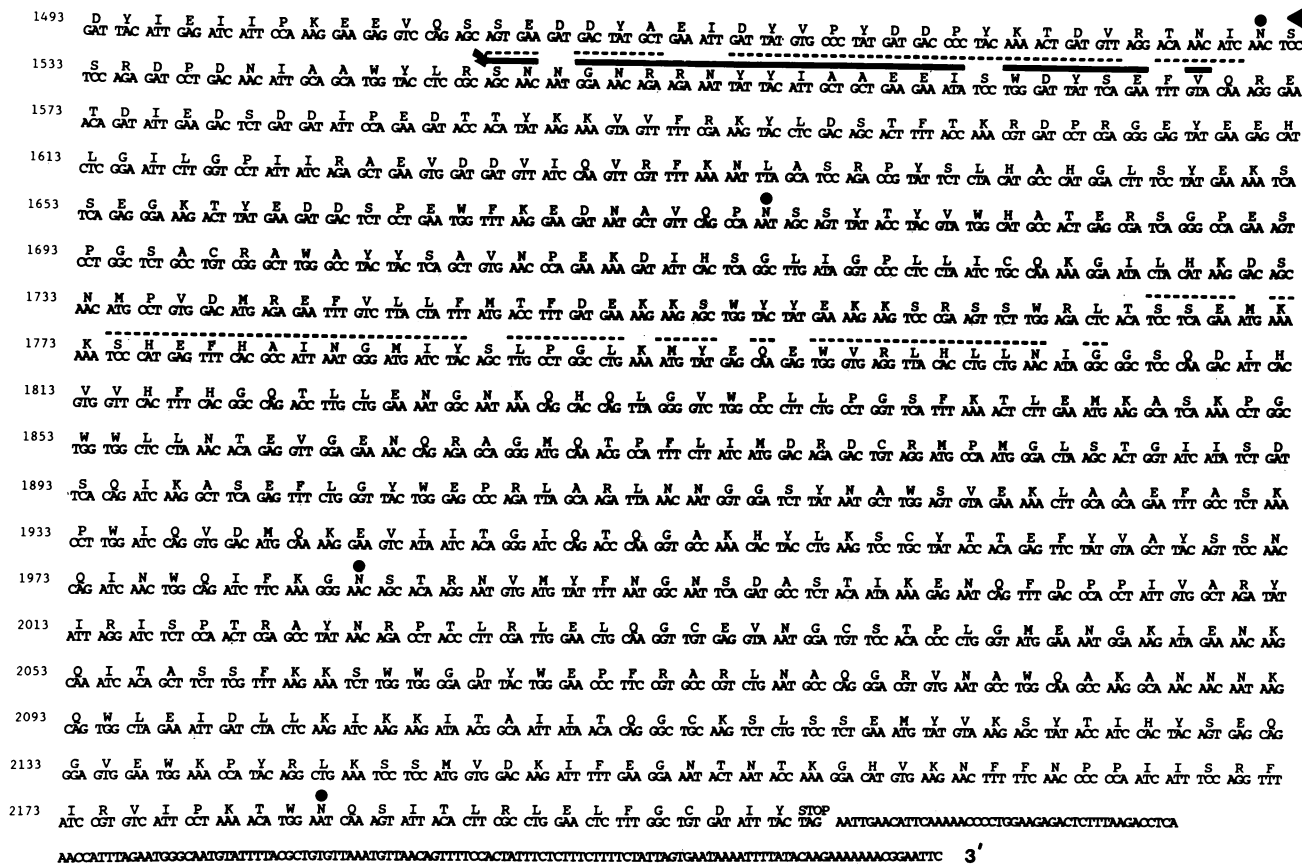


FIG. 2. Complete cDNA sequence and deduced amino acid sequence of human factor V. Sequence is numbered by amino acids with position 1 corresponding to the amino-terminal residue of the plasma protein. Heavy vertical arrow, leader peptide cleavage site; filled circles, potential N-linked glycosylation sites; curved arrows, thrombin cleavage sites. Amino acid sequence determined by amino acid sequencing is indicated by the solid overbar for sequence obtained from human factor V and by a dashed overbar for sequence obtained from bovine factor V. The region containing the 44 semiconserved repeats of nine amino acids is bordered by the solid triangles at the right.

contrast to an  $\approx 40\%$  identity shared by the other domains (Fig. 4). A search of the National Biomedical Research Foundation protein sequence data bank revealed significant homologies with only factor VIII, ceruloplasmin, and discoidin I.

The B domain of factor V is structurally unique relative to the rest of the molecule. Factor V contains 37 potential N-glycosylation sites of the form Asn-Xaa-Ser/Thr of which 25 are located in the B domain (Fig. 1)—consistent with the reports of heavily glycosylated activation products (5, 37–39). Because of thrombin activation of factor V to factor Va, two peptides of estimated  $M_r$  71,000 and 120,000 are released as activation peptides (3–5). From amino acid composition we have determined the  $M_r$  (excluding carbohydrate) of these peptides as  $\approx 34,000$  and  $58,000$ , respectively. With the  $M_r$  values estimated for these peptides, the data indicate a carbohydrate composition of  $\approx 50\%$  by weight. Factor V contains 19 cysteine residues, only one of which is located in the B domain in the 120,000 fragment (Fig. 1). The products from thrombin activation of factor V are not covalently associated, thus indicating a free sulfhydryl group in the 120,000 fragment. The B domain of factor V contains a region of 35 tandem and  $\approx 9$  additional semiconserved repeats of 9 amino acids of the form Asp-Leu-Ser-Gln-Thr/Asn-Leu-Ser-Pro (D-L-S-Q-T-T/N-L-S-P) as shown in Fig. 2. This is larger than the 20 repeats seen by Kane and Davie (21) and may be due to a 297-bp in-frame deletion within this repetitive region. In addition to these repeated sequences, the B domain also contains two highly conserved repeats of 17 amino acids upstream from the tandem repeat region.

A highly reactive free sulfhydryl group and the clustered repeats that contain Glu/Asp residues may contribute to the capability of the 120,000 fragment as a substrate for the

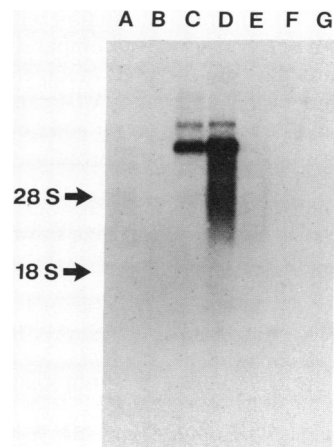


FIG. 3. RNA blot analysis of factor V mRNA. The RNA blot shown was hybridized with the 1614-bp *EcoRI* fragment of clone V1 as described. The 28S and 18S ribosomal RNA band positions are included for reference. (Lane A) human liver total RNA (5  $\mu$ g); (lane B) human umbilical-vein endothelial cell total RNA (5  $\mu$ g); (lane C) human HepG2 cell total RNA (5  $\mu$ g); (lane D) human liver poly(A)<sup>+</sup> selected RNA (4  $\mu$ g); (lane E) human liver poly(A)<sup>+</sup> selected RNA (2  $\mu$ g); (lane F) human peripheral blood leukocyte total RNA (2  $\mu$ g); and (lane G) human U297 (transformed monocyte) cell total RNA (2  $\mu$ g).

	MATCHES/LENGTH (%)												
	V-A2	V-A3	VIII-A1	VIII-A2	VIII-A3	CER-A1	CER-A2	CER-A3	V-C2	VIII-C1	VIII-C2	VIII-B	
V-A1	31	32	36	30	27	31	32	21					
V-A2		29	29	44	31	35	34	35					
V-A3			33	31	39	32	34	39					
VIII-A1				30	32	34	32	33					
VIII-A2					33	38	35	36					
VIII-A3						28	34	36					
CER-A1							37	39					
CER-A2								40					
V-C1									37	44	34		
V-C2										41	42		
VIII-C1												38	
V-B													14

FIG. 4. Comparison of the homologous domains in factor V, factor VIII, and ceruloplasmin. Values in the table represent the total identical amino acid matches divided by the overlapping lengths (including gaps) expressed as percentages.

transglutaminase activity of factor XIIIa seen by Francis *et al.* (40). In addition, modification of these reactive groups may produce effects that will elucidate the function of B domain. The B domain of factor VIII has been mapped to a single exon of  $\approx 3000$  bp (19). From current theories of gene evolution and RNA processing, one would predict multiple exons for the repeat regions of the factor V B domain.

Activation of factor V to factor Va involves three specific enzymatic cleavages catalyzed by thrombin (3–5). Amino acid sequence from  $\alpha$ -thrombin cleavage products shows thrombin cleavage sites at the following positions: Arg-709/Ser-710, Arg-1018/Thr-1019, and Arg-1545/Ser-1546. A schematic of the thrombin activation products is shown in Fig. 1, and the cleavage sites are further illustrated in Fig. 2.

The primary structure of factor V will contribute to better understanding the functions of regions in factor V and factor VIII in membrane binding and interaction with serine proteases factor Xa and factor IXa. By comparative analysis of structural domains and site-directed mutagenesis, functionality of regions in factor V and factor VIII can be detailed. In addition, much is to be learned about gene evolution within the family of ceruloplasmin-related proteins.

The authors thank John Knopf and Lisa Sultzman for helpful discussions and Darlene Vanstone and John Brown for oligonucleotide synthesis. This work was supported in part by National Institutes of Health Grant HL-34575.

- Nesheim, M. E., Katzman, J. A., Tracy, P. B. & Mann, K. G. (1981) *Methods Enzymol.* **80**, 249–274.
- Jackson, C. M. & Nemerson, Y. (1980) *Annu. Rev. Biochem.* **49**, 765–811.
- Esmon, C. T. (1979) *J. Biol. Chem.* **254**, 964–973.
- Nesheim, M. E. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 1326–1334.
- Suzuki, K., Dahlback, B. & Stenflo, J. (1982) *J. Biol. Chem.* **257**, 6556–6564.
- Miletich, J. P., Jackson, C. M. & Majerus, P. W. (1978) *J. Biol. Chem.* **253**, 6908–6916.
- Dahlback, B. & Stenflo, J. (1978) *Biochemistry* **17**, 4938–4945.

- Tracy, P. B., Nesheim, M. E. & Mann, K. G. (1981) *J. Biol. Chem.* **256**, 743–751.
- Tracy, P. B., Eide, L. L. & Mann, K. G. (1985) *J. Biol. Chem.* **260**, 2119–2124.
- Nesheim, M. E., Taswell, J. B. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 10952–10962.
- Nesheim, M. E., Kettner, C., Shaw, E. & Mann, K. G. (1981) *J. Biol. Chem.* **256**, 6537–6540.
- Krishnaswamy, S., Church, W. R., Nesheim, M. E. & Mann, K. G. (1987) *J. Biol. Chem.* **262**, 3291–3299.
- Mann, K. & Fass, D. (1983) in *Hematology*, ed. Fairbanks, V. F. (Wiley, New York), Vol. 2, pp. 347–374.
- Mann, K. G. (1984) in *Progress in Hemostasis and Thrombosis*, ed. Spaet, T. H. (Grune & Stratton, Orlando, FL), Vol. 7, pp. 1–23.
- Church, W. R., Jernigan, R. L., Toole, J., Hewick, R. M., Knopf, J., Knutson, G. J., Nesheim, M. E., Mann, K. G. & Fass, D. N. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6934–6937.
- Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, W. B., Coe, M. L., Knutson, G. J., Fass, D. N. & Hewick, R. M. (1981) *Nature (London)* **312**, 342–347.
- Takahashi, N., Ortel, T. L. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 390–394.
- Koschinsky, M. L., Funk, W. D., van Oost, B. A. & MacGillivray, R. T. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5086–5090.
- Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Capon, D. J. & Lawn, R. M. (1984) *Nature (London)* **312**, 326–330.
- Vehar, G. A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkins, R. N., Tuddenham, E. G. D., Lawn, R. M. & Capon, D. J. (1984) *Nature (London)* **312**, 337–342.
- Kane, W. H. & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6800–6804.
- Katzman, J. A., Nesheim, M. E., Hibbard, L. S. & Mann, K. G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 162–166.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13, and addendum (1984) **137**, 266–267.
- Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4298–4302.
- Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
- Hastie, N. D., Held, W. A. & Toole, J. J. (1979) *Cell* **17**, 449–457.
- Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M. & Darnell, J. E. (1981) *Cell* **23**, 731–739.
- Proudfoot, N. H. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
- von Heijne, G. (1982) *J. Mol. Biol.* **159**, 537–541.
- Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164.
- von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21.
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
- Poole, S., Firtel, R. A., Lamar, E. & Rowekamp, W. (1981) *J. Mol. Biol.* **153**, 273–289.
- Kane, W. H. & Majerus, P. W. (1981) *J. Biol. Chem.* **256**, 1002–1007.
- Esmon, C. T. (1980) in *The Regulation of Coagulation*, ed. Mann, K. G. & Taylor, F. B. (Elsevier, New York), pp. 137–143.
- Lindhout, T., Grovers-Riemslog, J. W. P., van de Waart, P., Hemker, H. C. & Rosing, J. (1982) *Biochemistry* **21**, 5494–5502.
- Francis, R. T., McDonagh, J. & Mann, K. G. (1986) *J. Biol. Chem.* **261**, 9787–9792.