Cloning of cDNAs Coding for the Heavy Chain Region and Connecting Region of Human Factor V, a Blood Coagulation Factor with Four Types of Internal Repeats†

William H. Kane,‡,§ Akitada Ichinose,‡ Frederick S. Hagen,∥ and Earl W. Davie*,‡
Department of Biochemistry, University of Washington, Seattle, Washington 98195, and ZymoGenetics, Inc., Seattle, Washington 98103

Received June 23, 1987; Revised Manuscript Received July 29, 1987

ABSTRACT: Human factor V is a high molecular weight plasma glycoprotein that participates as a cofactor in the conversion of prothrombin to thrombin by factor X_a. Prior to its participation in the coagulation cascade, factor V is converted to factor V_a by thrombin generating a heavy chain and a light chain, and these two chains are held together by calcium ions. A connecting region originally located between the heavy and light chains is liberated during the activation reaction. In a previous study, a cDNA of 2970 nucleotides that codes for the carboxyl-terminal 938 amino acids of factor V was isolated and characterized from a Hep G2 cDNA library [Kane, W. H., & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6800–6804]. This cDNA has been used to obtain additional clones from Hep G2 and human liver cDNA libraries. Furthermore, a Hep G2 cDNA library prepared with an oligonucleotide from the 5′ end of these cDNAs was screened to obtain overlapping cDNA clones that code for the amino-terminal region of the molecule. The composite sequence of these clones spans 6911 nucleotides and is consistent with the size of the factor V message present in Hep G2 cells (approximately 7 kilobases). The cDNA codes for a leader sequence of 28 amino acids and a mature protein of 2196 amino acids. The amino acid sequence predicted from the cDNA was in complete agreement with 139 amino acid residues that were identified by Edman degradation of cyanogen bromide peptides isolated from the heavy chain region and connecting region of plasma factor V. The domain structure of human factor V is similar to that previously reported for human coagulation factor VIII. Both proteins contain an amino-terminal heavy chain region, a connecting region, and a carboxyl-terminal light chain region. Also, each protein contains two A domains present in the heavy chain region and one A domain located in the light chain region. These A domains are ~30% identical with the triplicated A domains of ceruloplasmin, a copper binding protein present in plasma. Factor V and factor VIII also contain a large connecting region rich in carbohydrate, but there is no significant amino acid sequence homology between the two proteins in this portion of the two molecules. Two types of tandem repeats (17 and 9 amino acids) have also been identified in the connecting region of factor V. The light chain regions of factor V and factor VIII also contain two C domains that follow the third A domain. The present data indicate that the amino acid sequence in the heavy and light chain regions of factor V is ~40% identical with the corresponding regions of factor VIII. These experiments provide additional support for the concept that these two proteins as well as ceruloplasmin share a common ancestry during their evolution.

Human coagulation factor V is a high molecular weight glycoprotein that is present in blood plasma and platelets and is essential for normal hemostasis. Factor V circulates in plasma as a large single-chain polypeptide (M_r 330 000) with little or no coagulant activity (Nesheim et al., 1979; Esmon, 1979; Kane & Majerus, 1981; Katzmann et al., 1981; Suzuki et al., 1982). It is converted to an activated species, factor V_a, through limited proteolysis by thrombin (Nesheim et al., 1979; Esmon, 1979; Kane & Majerus, 1981; Katzmann et al., 1981; Suzuki et al., 1982) and factor X_a (Smith & Hanahan, 1976; Foster et al., 1983). Factor V_a is composed of a heavy chain (M_r 110 000) and a light chain (M_r 78 000) derived from the amino- and carboxyl-terminal regions of factor V, respectively. Two carbohydrate-rich fragments from the central connecting region are released during the activation of the molecule by thrombin (Esmon, 1979; Kane & Majerus, 1982; Suzuki et al., 1982). The heavy and light chains remain as a noncovalent complex held together by calcium ions (Esmon, 1979; Suzuki et al., 1983).

Activation of factor V unmasks binding sites in the molecule for prothrombin and factor X_a (Esmon et al., 1973; Suzuki et al., 1982). The prothrombin binding site appears to be present in the heavy chain (Guinto & Esmon, 1984), while the factor X_a binding site appears to be located, at least in part, within the light chain (Tucker et al., 1983). The light chain of factor V_a also contains the binding site for cell and anionic phospholipid surfaces (Tracy & Mann, 1983). Surface-bound factor V_a accelerates prothrombin activation by factor X_a approximately 4 orders of magnitude by shifting the reaction pathway to an ordered sequential reaction in which meizothrombin is the sole intermediate (Krishnaswamy et al., 1986). Factor V_a is inactivated by activated protein C (Kisiel et al., 1977), which makes several cleavages within the heavy chain and light chain. These reactions destroy the prothrombin and factor X_a binding sites in factor V_a (Suzuki et al., 1983; Guinto & Esmon, 1984).

Coagulation factor VIII is a cofactor that is required for the activation of factor X by factor IX_a. Although factor VIII

† This work was supported by Research Grant HL 16919 from the National Institutes of Health. W.H.K. completed this during the tenure of a Clinician Scientist Award (86-415) from the American Heart Association and E. R. Squibb and Sons Inc. with funds contributed in part by the Washington Affiliate of the American Heart Association.

‡ University of Washington.
§ AHA-Squibb Clinician Scientist Awardee.
∥ ZymoGenetics, Inc.

0006-2960/87/0426-6508$01.50/0

has been more difficult to characterize, the role of factor $VIII_a$ in the factor IX_a–factor $VIII_a$ complex appears to be analogous to the role of factor V_a in the prothrombinase complex (van Dieijen et al., 1981). In addition, both cofactors are activated by thrombin and inactivated by activated by protein C (Kisiel et al., 1977; Vehar & Davie, 1980). Fass and co-workers (Fass et al., 1985) found amino acid sequence homology in the amino-terminal regions of the heavy and light chains of bovine factor V_a and the corresponding fragments of porcine factor $VIII_a$. These sequences were also homologous to sequences found in human ceruloplasmin (Church et al., 1984). The complete amino acid sequence of human factor VIII has been determined by cDNA cloning (Gitschier et al., 1984; Vehar et al., 1984; Toole et al., 1984). It contains three A domains (~350 amino acids each), and these domains are ~30% identical with the triplicated A domains present in human ceruloplasmin (Takahashi et al., 1984). In addition, factor VIII contains a carbohydrate-rich connecting region of ~980 amino acids that is located between the second and third A domains and two C domains (~150 amino acids each) at the carboxyl terminus of the molecule.

We have recently reported the sequence of a cDNA coding for the entire light chain region of human factor V and a portion of the connecting region (Kane & Davie, 1986). The amino acid sequence of the light chain region of factor V is ~40% identical with the corresponding carboxyl-terminal fragment of human factor VIII. The carboxyl terminus of the connecting region of factor V, however, showed no significant homology to factor VIII and contained at least 20 tandem repeats of nine amino acids.

In this paper, the sequence of additional overlapping cDNA clones that code for the entire heavy chain region and the connecting region of factor V are presented. These data indicate that human factor V consists of 2196 amino acids that include a heavy chain region of 709 amino acids, a connecting region of 836 amino acids, and a light chain region of 651 amino acids.

Materials and Methods

Restriction endonucleases, nuclease BAL-31, the Klenow fragment of *Escherichia coli* DNA polymerase, and T4 ligase were purchased from Bethesda Research Laboratories, New England Biolabs, International Biochemicals Inc., Boehringer-Mannheim Inc., or Promega Biotech. RNase H, bacterial alkaline phosphatase, M13mp18, M13mp19, PUC18, and PUC18 were supplied by Bethesda Research Laboratories. λgt11 and M13 sequencing primers and DNA sequencing kits were obtained from New England Biolabs. The Sequenase DNA sequencing kit was obtained from United States Biochemical Corp., and ^{32}P-labeled nucleotides were obtained from New England Nuclear. The [α-^{35}S]dATP was purchased from Amersham, while the AMV reverse transcriptase XL was obtained from Life Sciences Inc. RNasin and Packagene extract were purchased from Promega Biotech. *Eco*RI linkers (10 nucleotides in length) were obtained from Pharmacia P-L Biochemicals. Normal human plasma was kindly provided by the Pacific Northwest Red Cross Blood Service, Portland, OR.

The preparation of the oligo(dT)-primed λgt11 Hep G2 cDNA library has been described (Hagen et al., 1986). The oligo(dT)-primed human liver λgt11 cDNA library was kindly provided by Dr. Savio L. C. Woo (Kwok et al., 1985). An oligo(dT)-primed human umbilical vein endothelial cell λgt11 cDNA library was the gift of Dr. J. Evan Sadler (Shelton-Inloes et al., 1986). In order to obtain cDNAs coding for the amino terminus of human factor V, a specifically primed λgt11 cDNA library was constructed from Hep G2 mRNA with a modification of the RNase H method (Gubler & Hoffman, 1983; Toole et al., 1984; Frischer et al., 1986). Poly(A)-selected Hep G2 mRNA was isolated according to standard methods (Chirgwin et al., 1979; Maniatis et al., 1982). The nucleotide primer 5′-GTCATTTGAGGAATTC-3′, which is complementary to nucleotides 3382–3397 in Figure 2, was synthesized on an oligonucleotide synthesizer (Applied Biosystems, Foster City, CA). Hep G2 mRNA (5 μg) and 1 μg of the oligonucleotide primer were mixed, denatured at 65 °C for 3 min, and then cooled on ice. The first-strand synthesis of the cDNA was accomplished with AMV reverse transcriptase XL. Second-strand synthesis was accomplished with RNase H and DNA polymerase I. The cDNA was methylated with *Eco*RI methylase, chromatographed on a Sepharose 6B column, and blunted with T4 DNA polymerase. *Eco*RI linkers were ligated to the cDNA with T4 DNA ligase and then digested with *Eco*RI. After chromatography on a Sepharose 2B column, the cDNA was then ligated to λgt11 arms and packaged with Packagene mix. The packaged cDNA was used to infect *E. coli* strain Y 1088. Approximately 90% of the 8×10^5 phage were recombinants on the basis of β-galactosidase activity. The library was screened without further amplification.

Restriction fragments corresponding to nucleotides 3637–4261 and 3217–3381 (Figure 2) were prepared from λHV2.97 and λHV0.85, respectively. These fragments were nick translated to a specific activity of $\geq 10^8$ cpm/mg (Maniatis et al., 1975) and used to screen the cDNA libraries (Benton & Davis, 1977). Positive clones were plaque purified, and phage DNA was isolated by the liquid culture lysis method (Silhavy et al., 1984) or the miniprep plate lysate method (Helms et al., 1985). The cDNA inserts were isolated following *Eco*RI digestion of the phage DNA and then subcloned into plasmid PUC18 or PUC19 (Vieira & Messing, 1982). Restriction fragments were subcloned into M13mp18 and/or M13mp19 for sequencing. Additional fragments were generated by digestion with the nuclease BAL-31 (Poncz et al., 1982) and subcloned into M13mp18 for sequencing. In some cases, cDNA inserts were sequenced directly in λgt11 with λgt11 sequencing primers. The cDNA inserts were sequenced by the dideoxy method with [α-^{35}S]dATP and either the Klenow fragment of *E. coli* DNA polymerase (Sanger et al., 1977) or Sequenase (United States Biochemical Corp.). Sequencing reactions were run on buffer gradient gels (Biggin et al., 1983) or wedge gradient gels (Ansorge & Labeit, 1984). DNA sequences were analyzed on an Apple MacIntosh computer using the program DNA INSPECTOR II (Gross, 1986) and on a Compaq Deskpro computer using GENEPRO version 4.0 (Riverside Scientific Enterprises, Seattle, WA).

The size of the message for factor V in liver was determined by Northern blotting (Maniatis et al., 1982) employing a nick-translated cDNA probe prepared from the 1.9-kb *Eco*RI insert in λHV2.97. Five micrograms of poly(A)-selected mRNA from Hep G2 cells was electrophoresed on a 1.0% agarose gel containing 6% formaldehyde, transferred to nitrocellulose, and hybridized in 0.25 M NaCl, 7% sodium dodecyl sulfate, and 0.25 M NaH_2PO_4, pH 7.2, at 65 °C overnight with the nick-translated cDNA probe containing 2×10^6 cpm/mL. The filters were then washed in 0.3 M NaCl, 0.3 M trisodium citrate, pH 7.0, and 0.1% sodium dodecyl sulfate for 40 min at room temperature and then 20 min at 65 °C.

Human factor V was isolated as previously described (Kane & Majerus, 1981). The fragment containing the heavy chain

region and the connecting region was isolated from 10 mg of human factor V after incubation with the factor V activator from Russell's viper venom (Suzuki et al., 1982). This fragment was then reduced, carboxymethylated (McMullen & Fujikawa, 1985), and digested with cyanogen bromide (Titani et al., 1972). The cyanogen bromide peptides were purified by gel filtration on Sephadex G-50, followed by HPLC with an Altex Ultrapore C3 reverse-phase separation column (McMullen & Fujikawa, 1985). Nine peptides were isolated in sufficient yield to allow automated amino acid sequence analysis with a Beckman 890C sequencer. Amino acid sequences were analyzed on a VAX computer with the SEARCH and ALIGN programs of Dayhoff (Dayhoff et al., 1983) and on a Compaq Deskpro computer using GENEPRO version 4.0 (Riverside Scientific Enterprises, Seattle, WA). Protein sequences were compared with the National Biomedical Research Foundation Protein Sequence Database of the Protein Identification Resource (National Biomedical Research Foundation, Washington, DC 20007), Release 10 (Aug 1986). In this paper, percent identities were reported by employing the ALIGN program, which inserts gaps for maximal alignment score.

Results and Discussion

In previous studies, the nucleotide sequence for a cDNA insert (λHV2.97) that codes for the carboxyl terminus of human factor V was reported (Kane & Davie 1986). This insert coded for the entire light chain region and a portion of the connecting region of the molecule. A 328-nucleotide *Eco*RI/*Bgl*II fragment from the 5′ end of λHV2.97 was isolated and nick translated to a specific activity of $>10^8$ cpm/μg. This probe was then utilized to screen 2×10^6 phage from the oligo(dT)-primed Hep G2 cDNA library (Hagen et al., 1986). Ten positive clones were plaque purified and characterized. The phage containing the largest inserts were λHV2.77 and λHV0.85 (Figure 1). In addition, 3×10^6 phage from an oligo(dT)-primed normal human liver cDNA library were screened with the nick-translated probe, and a single positive clone (λHV1.66) was isolated and characterized. No positive clones were identified when 2×10^6 phage from a human umbilical vein endothelial cell cDNA library were screened with the probe. λHV1.66 was sequenced completely on both strands, and λHV0.85 was analyzed by restriction mapping and also partially sequenced. The sequence of these clones extends ~400 bp from the 5′ end of λHV2.97. These data indicated that λHV2.97 contains a 297-nucleotide deletion.

In order to obtain clones that code for the amino terminus of human factor V, a specifically primed cDNA library was constructed from purified Hep G2 mRNA with the synthetic oligonucleotide primer 5′-GTCATTTGAGGAATTC-3′. This sequence is complementary to nucleotides 3382–3397 shown in Figure 2. The *Eco*RI fragment (165 bp) from the 5′ end of λHV0.85 was nick translated and used to screen 8×10^5 phage from the oligonucleotide-primed cDNA library. Fifty-three positive clones were plaque purified, and the phage DNA was isolated and analyzed by *Eco*RI digestion. Two clones designated λHV3.37 and λHV3.14 contained the largest inserts and were amplified and characterized by restriction mapping. There was a 155-bp overlap between the sequence of the 3′ ends of λHV3.37 and λHV3.14 and the 5′ end of λHV0.85. The composite DNA sequence for these clones was determined on both strands and on at least one strand for each isolate. The nucleotide and predicted amino acid sequences for the composite cDNA inserts are shown in Figure 2. The sequence of these overlapping cDNA clones, together with the sequence of λHV2.97 which we have previously reported (Kane & Davie, 1986), comprises a total cDNA sequence of 6911 nucleotides for human factor V. When Northern blots of Hep G2 mRNA were probed with a ^{32}P-labeled nick-translated fragment from the insert in λHV2.97, a single band of approximately 7 kb was obtained. This suggests that the cDNA sequence of 6911 nucleotides corresponds to a full-length or nearly full length factor V cDNA. The cDNA sequence includes 76 nucleotides of 5′ untranslated sequence, 84 nucleotides that code for a leader peptide of 28 amino acids, 6588 nucleotides that code for a mature protein of 2196 amino acids, a stop codon, and 155 nucleotides of 3′ noncoding sequence.

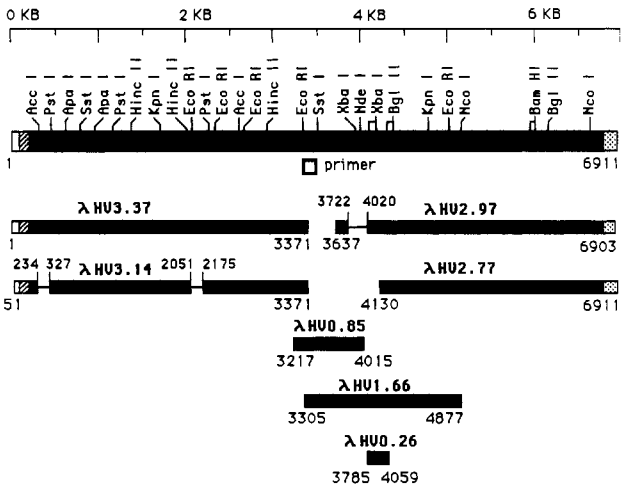

FIGURE 1: Partial restriction map for the various cDNA inserts coding for human coagulation factor V. The open and dotted bars indicate 5′ and 3′ untranslated sequence, respectively. The region coding for the leader sequence is shown by the hashed bar, and the region coding for the mature protein is shown by a solid bar. Restriction sites used in subcloning and DNA sequencing are identified. The first and last nucleotides at the end of each cDNA insert are indicated by the appropriate number. λHV2.97 contained a 297 base pair deletion beginning at nucleotide 3723. λHV3.14 contained two deletions of 92 and 93 base pairs beginning at nucleotides 235 and 2052, respectively. The position of the 16-mer oligonucleotide primer used to construct the cDNA library is shown by the open box.

The predicted amino acid sequence from the cDNA is in complete agreement with 139 amino acid residues determined by Edman degradation of fragments of factor V obtained following cyanogen bromide digestion. These amino acid sequences are overlined in Figure 2 and were obtained from the heavy chain region and the connecting region of plasma factor V. The cDNA codes for a mature protein with a M_r of 249 000 without carbohydrate. The amino acid composition for the mature protein is as follows: $Ala_{96}Arg_{93}Asn_{111}Asp_{133}{}^1/_2\text{-}Cys_{19}Gln_{102}Glu_{143}Gly_{111}His_{68}Ile_{124}Leu_{198}Lys_{125}Met_{52}Phe_{77}Pro_{149}Ser_{245}Thr_{140}Trp_{41}Tyr_{82}Val_{87}$. The molecular weight of single-chain bovine factor V has been determined to be 330 000 by ultracentrifugation (Mann et al., 1981). This suggests that the carbohydrate content of factor V may be as high as 25%, which is twice as high as that previously reported (Nesheim et al., 1979; Kane & Majerus, 1981).

The nucleotide sequence flanking each side of the proposed ATG initiator codon is AGCATGT, which differs from the consensus sequence of ACCATGG proposed by Kozak (1986a,b) at the −2- and +4-positions. However, the proposed flanking sequence agrees with the consensus sequence at the more important −3-position. The proposed leader sequence is 28 amino acids in length, which is a little longer than average for typical eukaryotic signal sequences (von Heijne, 1983, 1985). The leader sequence contains a typical central hydrophobic region of 14 amino acids. The predicted signal

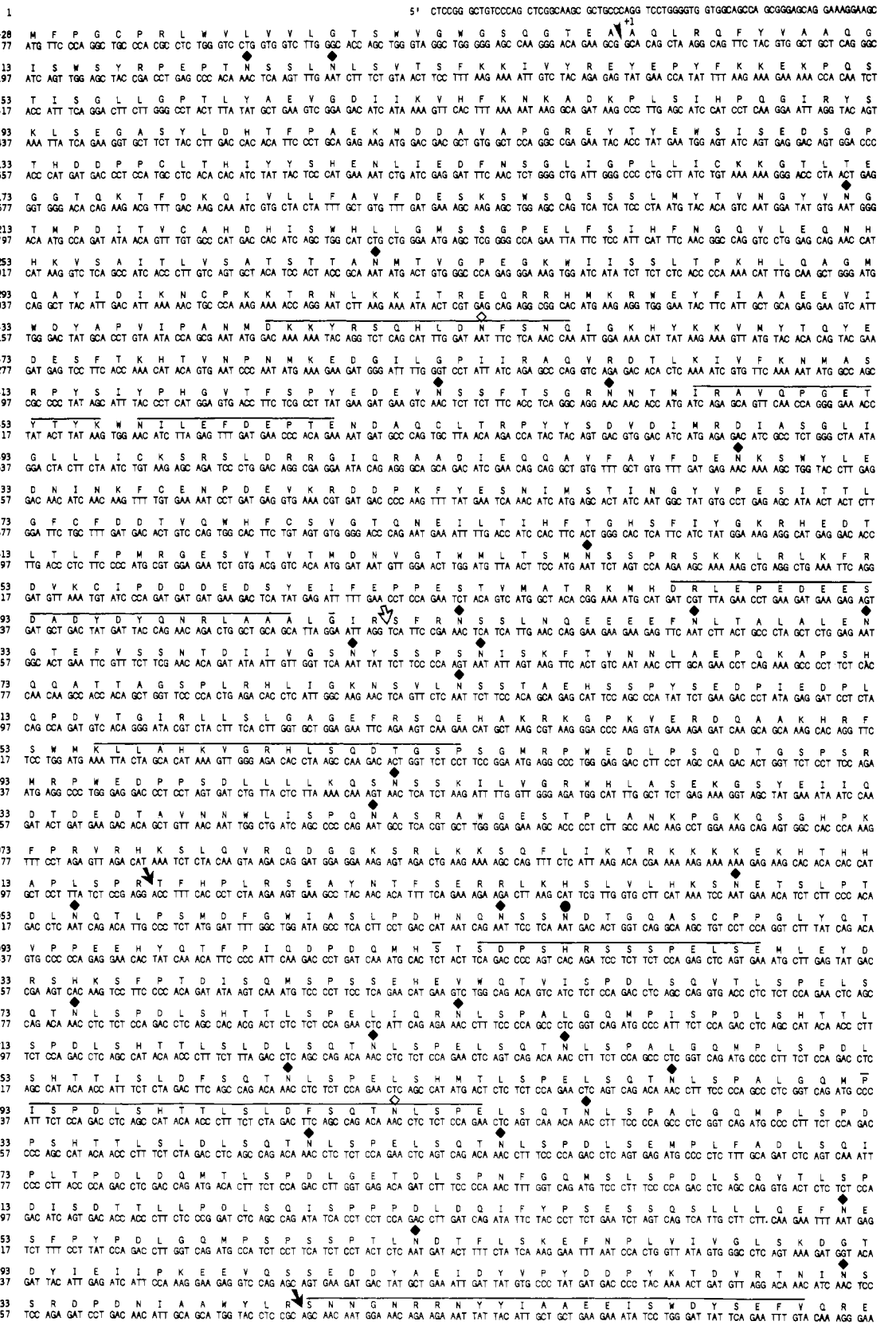

FIGURE 2: Nucleotide sequence of the overlapping cDNAs that code for the heavy chain region and connecting region of human factor V. Nucleotides 1–76 represent the 5′ untranslated sequence. Amino acids −28 to −1 indicate the leader sequence. Amino acids 1–1545 constitute the heavy chain region and connecting region of factor V. The amino acid residues determined by Edman degradation are overlined. Potential N-linked oligosaccharide attachment sites are indicated by the solid diamonds. Protein sequence analysis indicated that Asn-354 and Asn-1310 (open diamonds) are not glycosylated. The proposed thrombin cleavage site between the heavy chain and the connecting region is indicated by the open arrow after Arg-709. The thrombin cleavage sites within the connecting region (after Arg-1018) and between the connecting region (after Arg-1545) and the light chain have been confirmed by protein sequencing and are indicated by the solid arrows (see text). The cDNA sequence coding for the light chain of factor V begins at nucleotide 4797. This corresponds to nucleotide 863 in the clone λHV2.97, which was reported in our previous paper (Kane & Davie, 1986) (Figure 2). The light chain corresponds to amino acids 1546–2196. These were numbered 288–936 in our previous paper.

peptidase cleavage site conforms with the −3, −1 rule of von Heijne (1983) with Thr at position −3 and Ala at position −1. The amino-terminal sequence of human factor V was reported by Suzuki and co-workers as AQLGQFYV (Suzuki et al., 1982). This amino acid sequence is in agreement with the amino-terminal sequence predicted from the cDNA except for a Gly at position +4. This discrepancy could be explained either by sequence polymorphism with a single A to G substitution at position 170, by a reverse transcriptase error during the preparation of the cDNA library, or by an error in protein sequencing.

The nucleotide sequences for λHV3.37 and λHV3.14 are identical except that λHV3.37 contains an additional 50 nucleotides of 5′ sequence and λHV3.14 has two deletions of 92 and 93 nucleotides beginning at positions 235 and 2052, respectively (Figures 1 and 2). Similar deletions were noted in the sequences for some of the cDNAs coding for factor VII, which were also isolated from a Hep G2 cDNA library (Hagen et al., 1986), and may result from abnormal alternative splicing events that occur in Hep G2 cells. Each of the sequences deleted in λHV2.97 and λHV3.14 code for sequences that are homologous to sequences present in other domains within the factor V molecule, making it unlikely that they represent residual intron sequences.

Activation of factor V by thrombin results in the formation of a heavy chain from the amino terminus of the intact factor V molecule and a light chain originating from the carboxyl end of the molecule (Esmon, 1979; Suzuki et al., 1982). Fass et al. (1985) have reported amino acid homology between the N-terminal sequences of the heavy chains of bovine factor V_a and porcine factor $VIII_a$ (Fass et al., 1985). In addition, these data showed amino acid sequence homology with ceruloplasmin, a plasma copper binding protein (Church et al., 1984). The predicted amino acid sequence for human factor V indicates that the amino terminus of factor V is ~40% identical with the amino terminus of human factor VIII. The amino-terminal regions of both proteins have similar domain structures consisting of two ceruloplasmin-like A domains. These data together with that previously published for the light chain (Kane & Davie, 1986) establish that factor V as well as factor VIII and ceruloplasmin contains three A domains. In the case of factor V and factor VIII, the second and third A domains are separated by a large carbohydrate-rich connecting region (see below). On the basis of analogy to factor VIII, the probable thrombin cleavage site that liberates the factor V heavy chain from the connecting region is located between Arg-709 and Ser-710. The molecular weight of the heavy chain of factor V_a was calculated to be about 80 000 without carbohydrate. Asn-354 is not glycosylated as determined by protein sequencing. The addition of carbohydrate to eight of the nine potential N-linked oligosaccharide chains (M_r ~2000 each) would increase the molecular weight of the glycoprotein to approximately 96 000. This is similar to the values of 93 000–110 000, as determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (Katzman et al., 1981; Kane & Majerus, 1982; Suzuki et al., 1982).

Two clusters of Glu, Asp, and Tyr residues are present approximately 10 and 45 amino acids from the proposed thrombin cleavage site between the heavy chain region and the connecting region in factor V (Figure 3). Factor VIII contains a similar cluster of these amino acids adjacent to the thrombin cleavage site between the heavy chain region and the connecting region. Thrombin activation of factor VIII also results in the cleavage of an Arg–Ser bond between the first and second A domains (Eaton et al., 1986). There is also a

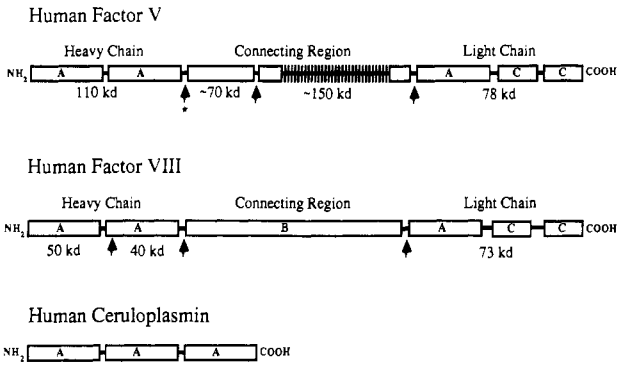

FIGURE 3: Structural domains in factor V, factor VIII, and ceruloplasmin. Thrombin cleavages sites are indicated by the arrows. Molecular weights are estimates based on sodium dodecyl sulfate–polyacrylamide gel electrophoresis and may not be completely accurate (see text). The thrombin cleavage sites in human factor V occur after Arg-709 [(*) tentative], Arg-1018, and Arg-1545. The thrombin cleavage sites in factor VIII occur after Arg-372, Arg-740, and Arg-1689 (Eaton et al., 1986). The identity of the domains is indicated by the letters inside the boxes. The 31 tandem repeats in factor V with the consensus sequence [T,N,P]LSPDLSQT are indicated by the vertical bars. The regions corresponding to the heavy chain, connecting region, and light chain of factors V_a and $VIII_a$ are indicated.

cluster of Glu, Asp, and Tyr residues adjacent to this cleavage site. Similar clusters of these amino acids are also present adjacent to the cleavage site between the connecting region and the light chain in both factor V and factor VIII. Thrombin, however, does not cleave between the first and second A domains in factor V. Also, factor V, unlike factor VIII, contains a cluster of basic amino acids between the first and second A domains with 10 of 19 residues being Arg or Lys. The sites of sulfation of a number of plasma proteins including the fourth component of human complement, heparin cofactor II, and α_2-antiplasmin have been identified by Strauss and co-workers (Hortin et al., 1986a,b,1987). In these cases, tyrosine sulfate occurs adjacent to clusters of Glu and Asp residues similar to those adjacent to some of the thrombin cleavage sites in factor V and factor VIII. The functional significance of tyrosine sulfation in proteins is not yet known. Whether the clusters of acidic amino acids adjacent to thrombin cleavage sites are required for thrombin activation of factor V and whether factor V contains tyrosine sulfate remain to be determined.

Activated protein C makes several cleavages in the heavy chain of human factor V_a that abolish its prothrombin binding activity. This results in an inactivation of the cofactor (Guinto & Esmon, 1984; Suzuki et al., 1983). Four smaller fragments (M_r 50 000, 30 000 24 000 and 21 000) are formed by the cleavage of the heavy chain of factor V_a by activated protein C, indicating that there are at least three cleavage sites in this chain. Furthermore, there are several Arg–X bonds in the heavy chain of factor V_a that could serve as potential cleavage sites for activated protein C. The actual cleavage sites for activated protein C within the heavy chain region of factor V will require amino acid sequencing of the proteolytic fragments.

Factor V binds a number of divalent metal ions including calcium and copper (Hibbard & Mann, 1979; Mann et al., 1984). On the basis of analogy with the copper binding protein plastocyanin, the ligands for three type I copper ions in ceruloplasmin have been proposed to be two His residues, a $^1/_2$-Cys residue, and a Met residue located near the carboxyl terminus of each A domain (Ryden & Bjork, 1976; Ryden, 1982). These potential copper binding sites are not conserved in the three A domains of factor V. This is consistent with

the visible light spectrum of factor V, which does not indicate the presence of type I copper. The potential type I copper binding sites have been conserved in the first and third A domains of factor VIII (Vehar et al., 1984; Toole et al., 1984), although it is not known whether factor VIII contains type I copper.

The second A domain in factor V is separated from the third A domain by 836 amino acids located between residues 710 and 1545. This connecting region is analogous to a similar region of 980 amino acids in factor VIIII. The connecting region of both proteins is characterized by a high content of Ser and Thr residues and a large number of potential N-linked oligosaccharide attachment sites. There is no amino acid sequence homology, however, between these regions in factor V and factor VIII or any other known protein. The connecting region of factor V, however, contains two types of tandem repeats of 17 and 9 amino acids (see below). During the activation of factor V by thrombin, the connecting region is released as two large fragments with apparent M_r of 70 000 and 150 000 (Suzuki et al., 1982; Kane & Majerus, 1981). Thus, a single $^1/_2$-Cys residue is present in the connecting region within the 150-kDa fragment and apparently does not participate in a disulfide bond. These fragments from the connecting region are not essential for the procoagulant activity of factor V, and their function, if any, is not known. The amino-terminal sequence of the purified 150-kDa fragment was found to be XFHXLRSEA (D. B. Wilson, personal communication), indicating that this fragment begins at Thr-1019. The amino-terminal sequence of the 70-kDa fragment, however, has not been established. A potential thrombin cleavage site at Arg-709 would generate an activation fragment beginning with Ser-710. This would give rise to a 70-kDa fragment consisting of 309 amino acids and 8 potential N-linked oligosaccharide attachment sites as well as two tandem repeats of 17 amino acids with a sequence of SQDTGSPSXMRPWEDXP.

We have previously reported an amino acid sequence for the carboxyl terminus of the 150-kDa activation fragment (Kane & Davie, 1986). Our present data complete the predicted amino acid sequence of this fragment and indicate that it is composed of 527 amino acids, including 17 potential N-linked oligosaccharide sites and 31 tandem repeats of 9 amino acids. The consensus sequence for these 31 repeats is [T,N,P]LSPDLSQT.

The 150-kDa activation fragment is a substrate for the plasma transglutaminase factor XIII (Francis et al., 1986). This suggests that the glutamines present within the tandem repeats in this fragment may be the glutamine donors for this reaction. The calculated molecular weights for the 70-kDa and 150-kDa activation fragments are 34 400 and 58 000, respectively, excluding carbohydrate. Asn-1310 is not glycosylated as determined by protein sequencing. The addition of 8 carbohydrate chains from the 70-kDa activation fragment plus 16 of the 17 potential N-linked oligosaccharide chains from the 150-kDa fragment would increase the molecular weights of these glycoprotein fragments to about 50 000 and 92 000, respectively. These fragments contain 19.3 and 25.1% Ser and Thr residues, respectively. Because factor V contains O-linked carbohydrate in addition to N-linked carbohydrate (Kane & Majerus, 1981), the total carbohydrate content of the connecting region of factor V may be over 50%.

The cDNA clones characterized in these studies together with that previously described (Kane & Davie, 1986) establish the complete primary amino acid sequence of human factor V. These findings confirm the proposal first made by Mann and co-workers that factor V, factor VIII, and ceruloplasmin comprise a family of related proteins that have arisen through gene duplication of the A domain (Church et al., 1984; Fass et al., 1985) (Figure 3). The availability of clones coding for factor V will allow for a detailed analysis of the structure–function relationships within the factor V molecule by site-directed mutagenesis as well as the characterization of the factor V gene.

While this paper was in preparation, we received a preprint of a paper by Jenny et al. (1987) that also describes the isolation of cDNA clones coding for the entire factor V molecule. Their results are in good agreement with the present data except for four silent single-base changes at positions 481, 628, 3880, and 4024 and a single-base change at 3929 that predicts a Leu at residues 1257 rather than an Ile (Figure 2). These minor differences may represent cloning artifacts or polymorphisms. Our sequence also differed from theirs in the first 13 nucleotides of 5′ untranslated sequence. Their protein sequence data confirm our proposed thrombin cleavage site at Ser-709.

ACKNOWLEDGMENTS

We thank Drs. Dominic Chung, Kazuo Fujikawa, Steven Leytus, and Donald Foster for helpful discussions and assistance. We also thank Brad McMullen for assistance in the amino acid sequence analysis, Dr. Kjell Sakariassen for assistance in the preparation of the Northern blots, Drs. J. Evan Sadler and S. L. C. Woo for kindly providing their cDNA libraries, Dr. David Wilson for communicating unpublished amino acid sequence data, Dr. Frans Peetoom for making available human plasma for the preparation of factor V, Dr. Kenneth Mann for kindly providing us with a preprint of their studies on the cloning of human factor V, and Lois Swenson for her assistance in the preparation of the manuscript.

REFERENCES

Ansorge, W., & Labeit, S. (1984) *J. Biochem. Biophys. Methods 10*, 237–243.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 188–192.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

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. U.S.A. 81*, 6934–6937.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.

Eaton, D., Rodriguez, H., & Vehar, G. A. (1986) *Biochemistry 25*, 505–512.

Esmon, C. T. (1979) *J. Biol. Chem. 254*, 964–973.

Esmon, C. T., Owen, W. G., Duiguid, O. L., & Jackson, C. M. (1973) *Biochim. Biophys. Acta 310*, 289–294.

Fass, D. N., Hewick, R. M., Knutson, G. J., Nesheim, M. E., & Mann, K. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1688–1691.

Foster, W. B., Nesheim, M. E., & Mann, K. G. (1983) *J. Biol. Chem. 258*, 13970–13977.

Francis, R. T., McDonagh, J., & Mann, K. G. (1986) *J. Biol. Chem. 261*, 9787–9792.

Frischer, L. E., Hagen, F. S., & Garber, R. L. (1986) *Cell (Cambridge, Mass.) 47*, 1017–1023.

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.
Gross, R. H. (1986) *Nucleic Acids Res. 14*, 591–596.
Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–269.
Guinto, E. R., & Esmon, C. T. (1984) *J. Biol. Chem. 259*, 13986–13992.
Hagen, F. S., Gray, C. L., O'Hara, P., Grant, F. J., Saari, G. C., Woodbury, R. G., Hart, C. E., Insley, M., Kisiel, W., Kurachi, K., & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2412–2416.
Helms, C., Graham, M. Y., Dutchik, J. E., & Olson, M. V. (1985) *Gene 4*, 39–49.
Hibbard, L. S., & Mann, K. G. (1979) *J. Biol. Chem. 255*, 638–645.
Hortin, G., Sims, H., & Strauss, A. W. (1986a) *J. Biol. Chem. 261*, 1786–1793.
Hortin, G., Tollefsen, D. M., & Strauss, A. W. (1986b) *J. Biol. Chem. 261*, 15827–15830.
Hortin, G., Fok, K. F., Toren, P. C., & Strauss, A. W. (1987) *J. Biol. Chem. 262*, 3082–3085.
Jenny, R. J., Pittman, D. D., Toole, J. J., Kriz, R. W., Aldape, R. A., Hewick, R. M., Kaufman, R. J., & Mann, K. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4846–4850.
Kane, W. H., & Majerus, P. W. (1981) *J. Biol. Chem. 256*, 1002–1007.
Kane, W. H., & Majerus, P. W. (1982) *J. Biol. Chem. 257*, 3963–3969.
Kane, W. H., & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6800–6804.
Katzmann, J. A., Nesheim, M. E., Hibbard, L. S., & Mann, K. G. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 162–166.
Kisiel, W., Canfield, W. M., Ericsson, L. H., & Davie, E. W. (1977) *Biochemistry 16*, 5824–5831.
Kozak, M. (1986a) *Cell (Cambridge, Mass.) 44*, 283–292.
Kozak, M. (1986b) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2850–2854.
Krishnaswamy, S., Mann, K. G., & Nesheim, M. E. (1986) *J. Biol. Chem. 261*, 8977–8984.
Kwok, S. C., Ledley, F. D., DiLella, A. G., Robson, K. J., & Woo, S. L. C. (1985) *Biochemistry 24*, 556–561.
Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Mann, K. G., Nesheim, M. E., & Tracy, P. B. (1981) *Biochemistry 20*, 28–33.
Mann, K. G., Lawler, C. M., Vehar, G. A., & Church, W. R. (1984) *J. Biol. Chem. 259*, 12949–12951.
McMullen, B. A., & Fujikawa, K. (1985) *J. Biol. Chem. 260*, 5328–5341.
Nesheim, M. E., Myrmel, K. H., Hibbard, L., & Mann, K. G. (1979) *J. Biol. Chem. 254*, 508–517.
Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., & Surrey, S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4298–4302.
Ryden, L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6767–6771.
Ryden, L., & Bjork, I. (1976) *Biochemistry 15*, 3411–3417.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Shelton-Inloes, B. B., Titani, K., & Sadler, J. E. (1986) *Biochemistry 25*, 3164–3171.
Silhavy, T. J., Verman, W. L., & Enquist, L. W. (1984) *Experiments with Gene Fusions*, pp 140–141, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Smith, C. M., & Hanahan, D. J. (1976) *Biochemistry 15*, 1830–1838.
Suzuki, K., Dahlback, B., & Stenflo, J. (1982) *J. Biol. Chem. 257*, 6556–6564.
Suzuki, K., Stenflo, J., Dahlback, B., & Teodorsson, B. (1983) *J. Biol. Chem. 258*, 1914–1920.
Takahashi, N., Ortel, T. L., & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 390–394.
Titani, K., Hermodson, M. A., Fujikawa, K., Ericsson, L. H., Walsh, K. A., Neurath, H., & Davie, E. W. (1972) *Biochemistry 11*, 4899–4903.
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. (1984) *Nature (London) 312*, 342–347.
Tracy, P. B., & Mann, K. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2380–2384.
Tucker, M. M., Foster, W. B., Katzmann, J. A., & Mann, K. G. (1983) *J. Biol. Chem. 258*, 1210–1214.
Van Dieijen, G., Tans, G., Rosing, J., & Hemker, H. (1981) *J. Biol. Chem. 256*, 3233–3441.
Vehar, G. A., & Davie, E. W. (1980) *Biochemistry 19*, 401–410.
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.
Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.
Von Heijne, G. (1983) *Eur. J. Biochem. 133*, 17–21.
Von Heijne, G. (1985) *J. Mol. Biol. 184*, 99–105.