Val, Pro, Ser, Asx, and Ser has the same amino acid composition as that of the chymopapain portion encompassing residues 111-117; an octapeptide containing Gly, Pro, Cys, Gly, Thr, Lys, Leu, and Asx fits exactly the polypeptide segment holding residues 151-158 (papain numbering).

The occurrence of two or three nonessential thiol groups in chymopapain with one of them close to the active site can account for some of the heterogeneity of the enzyme and the contradictory results on thiol reactivity (Brocklehurst et al., 1984; Polgár, 1984). In the different enzyme preparations the nonessential thiol groups can be oxidized to various extents. Indeed, before FPLC purification we have also found chymopapain containing 1.0 SH_C and 2.0 SH_T, which did not show a complete disulfide interchange presumably because the necessary thiol group was partially oxidized. It is remarkable that the thiol group involved in the disulfide bond formation with the active-site thiol is present in both our peak 1 and peak 2 chymopapains. Structural analysis is in progress to determine the cysteine residue implicated in the disulfide bond formation. It may be of regulatory significance that the disulfide form of chymopapain can serve as an inactive "proenzyme" that becomes active on reduction.

ACKNOWLEDGMENTS

We thank Dr. K. Lynn for his information about his studies on the amino sequence of chymopapain.

Registry No. PDS, 2127-03-9; Nbs_2, 69-78-3; Z-glycine *p*-nitrophenyl ester, 1738-86-9; *N*-benzoyl-L-arginine ethyl ester, 971-21-1; *N*-Z-phenylalanylglycine ethyl ester, 2778-34-9; chymopapain, 9001-09-6.

REFERENCES

Asbóth, B., & Polgár, L. (1977) *Acta Biochim. Biophys. Acad. Sci. Hung. 12*, 223-230.
Barrett, A. J., & Buttle, D. J. (1985) *Biochem. J. 228*, 527.
Bender, M. L., & Kézdy, J. F. (1965) *Annu. Rev. Biochem. 34*, 49-76.
Brocklehurst, K. (1982) *Methods Enzymol. 87*, 427-469.
Brocklehurst, K., & Salih, E. (1983) *Biochem. J. 213*, 559-560.
Brocklehurst, K., & Salih, E. (1985) *Biochem. J. 228*, 525-527.
Brocklehurst, K., Baines, B. S., Salih, E., & Hatzoulis, C. (1984) *Biochem. J. 221*, 553-554.
Buttle, D. J., & Barrett, A. J. (1984) *Biochem. J. 223*, 81-88.
Khan, I. U., & Polgár, L. (1983) *Biochim. Biophys. Acta 760*, 350-356.
Laemmli, U. K. (1970) *Nature (London) 227*, 680-685.
Lynn, K. R., Watson, D. C., & Yaguchi, M. (1985) presented at the 17th FEBS Meeting, Amsterdam.
Polgár, L. (1984) *Biochem. J. 221*, 555-556.
Polgár, L., & Sajgó, M. (1981) *Biochim. Biophys. Acta 667*, 351-354.
Robinson, G. W. (1975) *Biochemistry 14*, 3695-3700.
Sluyterman, L. A. AE., Wijdenes, J., & Voorn, G. (1977) *Eur. J. Biochem. 77*, 107-111.
Schwert, G. W., & Takenaka, Y. (1955) *Biochim. Biophys. Acta 16*, 570-575.
Zucker, S., Buttle, D. J., Nicklin, M. J. H., & Barrett, A. J. (1985) *Biochim. Biophys. Acta 828*, 196-204.

Amino Acid Sequence of the a Subunit of Human Factor XIII†

Akitada Ichinose, Lee E. Hendrickson, Kazuo Fujikawa, and Earl W. Davie*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received June 23, 1986

ABSTRACT: Factor XIII is a plasma protein that plays an important role in the final stages of blood coagulation and fibrinolysis. The complete amino acid sequence of the a subunit of human factor XIII was determined by a combination of cDNA cloning and amino acid sequence analysis. A λgtll cDNA library prepared from human placenta mRNA was screened with an affinity-purified antibody against the a subunit of human factor XIII and then with a synthetic oligonucleotide probe that coded for a portion of the amino acid sequence present in the activation peptide of the a subunit. Six positive clones were identified and shown to code for the a subunit of factor XIII by DNA sequence analysis. A total of 3831 base pairs was determined by sequencing six overlapping cDNA clones. This DNA sequence contains a 5′ noncoding region or a region coding for a portion of a pro-piece or leader sequence, the mature protein (731 amino acids), a stop codon (TGA), a 3′ noncoding region (1535 nucleotides), and a poly(A) tail (10 nucleotides). When the a subunit of human factor XIII was digested with cyanogen bromide, 11 peptides were isolated by gel filtration and reverse-phase HPLC. Amino acid sequence analyses of these peptides were performed with an automated sequenator, and 363 amino acid residues were identified. These amino acid sequences were in complete agreement with those predicted from the cDNA. The a subunit of factor XIII contained the active site sequence of Tyr-Gly-Gln-Cys-Trp, which is identical with that of tissue translguaminase. Six potential Asn-linked carbohydrate attachment sites are present in the a subunit. At least three of these sites have little or no carbohydrate as determined by amino acid sequence analysis. Little or no significant homology to other proteins was observed by computer-assisted amino acid sequence analysis employing the Dayhoff protein sequence data base.

Factor XIII (fibrin stabilizing factor, fibrinoligase, or plasma transglutaminase) is a plasma glycoprotein that circulates in blood as a proenzyme. During the final stages of blood coagulation, thrombin converts the proenzyme to an active form called factor $XIII_a$. Factor $XIII_a$ is a transglutaminase that catalyzes the cross-linking of fibrin monomers through the formation of intermolecular ε-(γ-glutamyl)lysine bonds. This

†This work was supported in part by Research Grant HL 16919 from the National Institutes of Health.

0006-2960/86/0425-6900$01.50/0

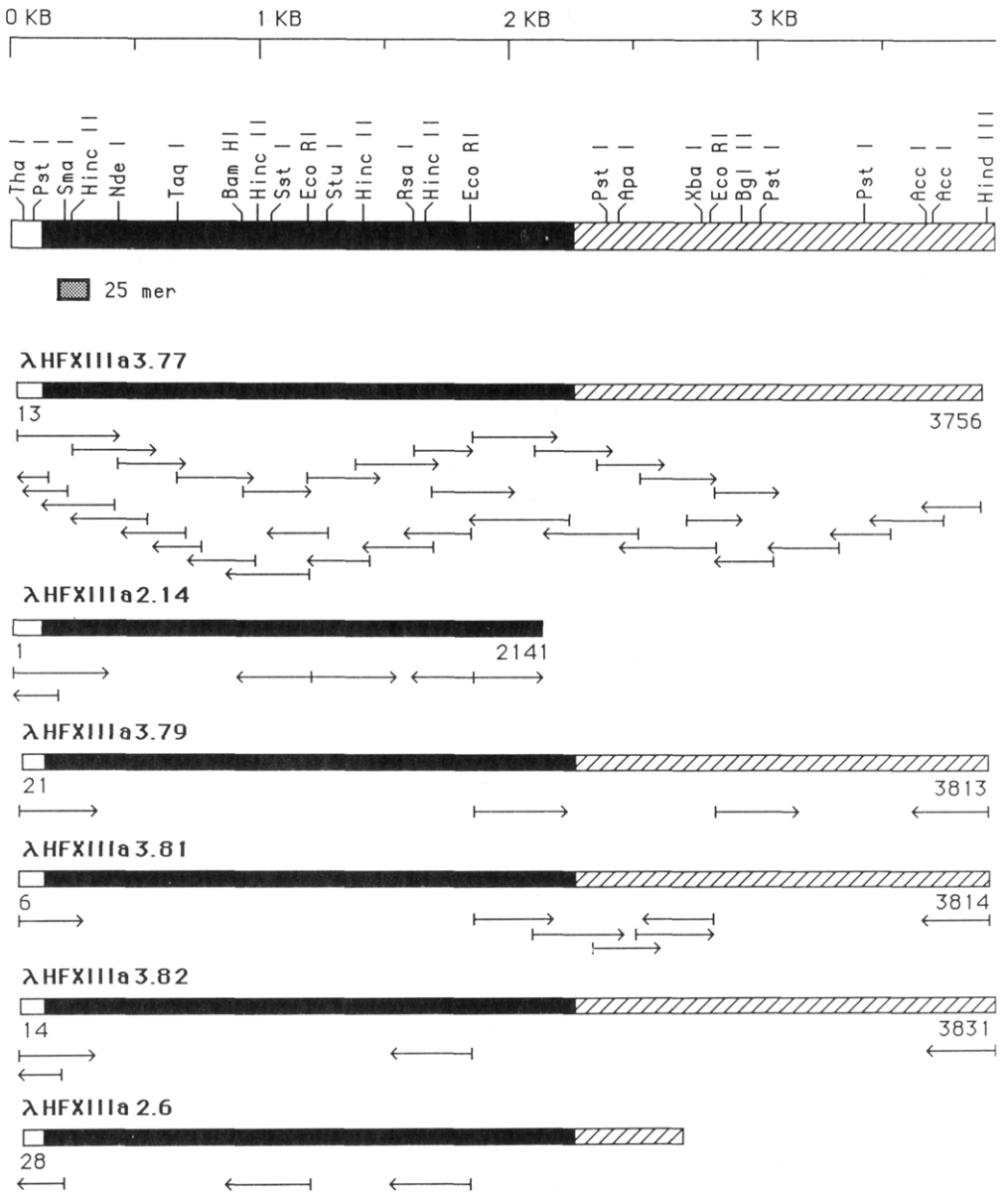


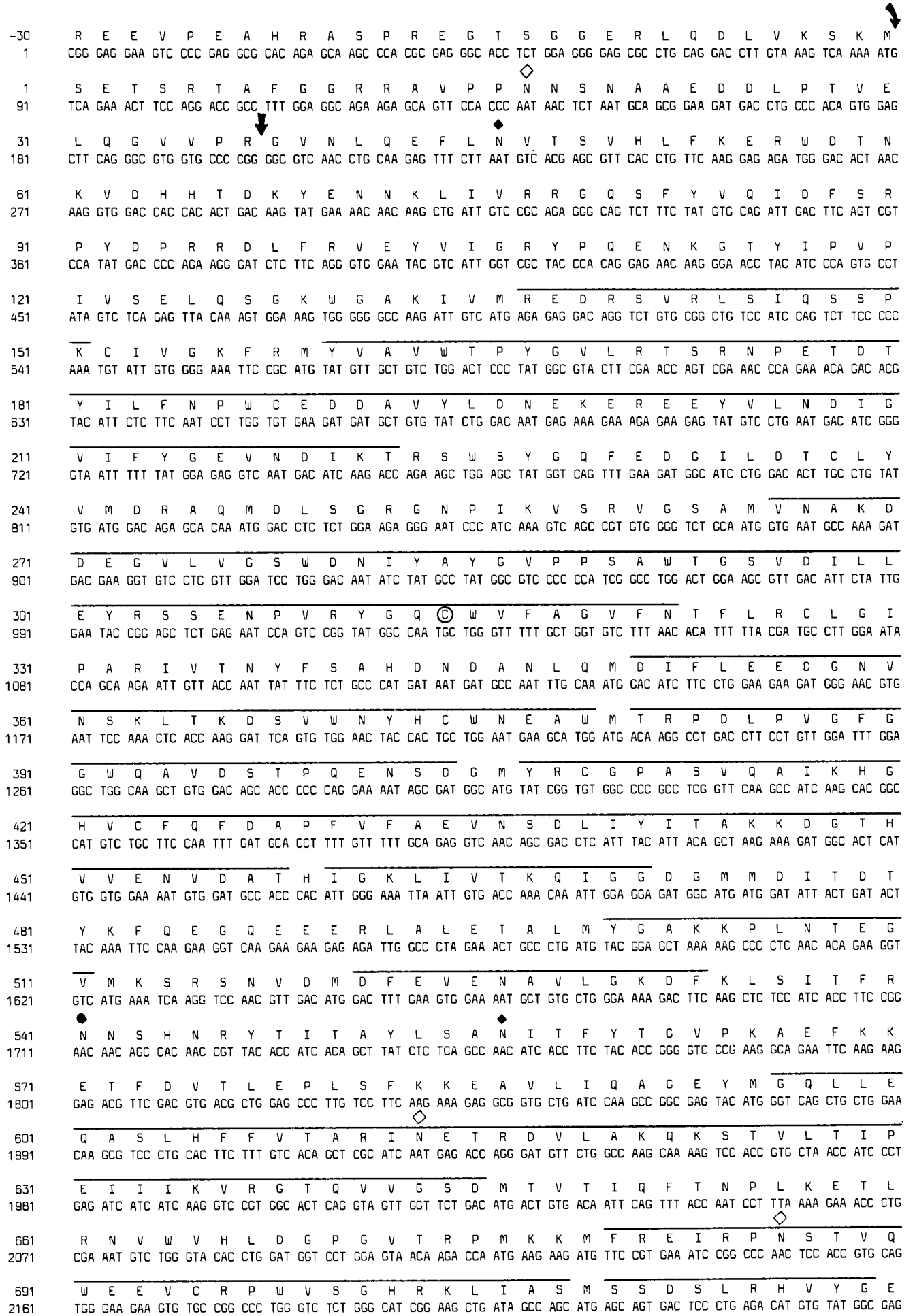
FIGURE 1: Partial restriction map and summary of sequencing strategy for cDNA inserts that code for the a subunit of human factor XIII. The open, solid, and hatched bars represent a portion of the leader peptide or 5′ noncoding region, the mature protein, and the 3′ noncoding region, respectively. The extent of DNA sequencing is shown by the length of each arrow, and the direction of the arrow indicates the strand that was sequenced. Restriction sites used in subcloning and DNA sequencing are also identified. The first and last nucleotides at the ends of each cDNA insert are indicated by the appropriate number. The dotted box indicates the position of the oligonucleotide probe of 25 base pairs.

reaction occurs in the presence of Ca^{2+} [see Folk and Finlayson (1977) and Lorand et al. (1980) for reviews].

The cross-linking reactions catalyzed by factor $XIII_a$ lead to a dimerization of the γ-chains of fibrin (γ-dimerization) (Chen & Doolittle, 1970) followed by a polymerization of the α-chains of fibrin (α-polymerization) (Pisano et al., 1972). The γ-dimerization and α-polymerization reactions result in a fibrin with considerable mechanical strength (Lorand, 1972; Roberts et al., 1973; Mockros et al., 1974; Shen et al., 1975; Shen & Lorand, 1983) and an increase in resistance to proteolytic degradation by plasmin (Lorand & Jacobsen, 1962; Gaffney & Whitaker, 1979). Factor $XIII_a$ catalyzes the cross-linking of the α-chain of fibrin to α_2-plasmin inhibitor (Sakata & Aoki, 1980; Tamaki & Aoki, 1981) and fibronectin (Mosher, 1975). The cross-linking between collagen and fibronectin is also catalyzed by factor $XIII_a$ (Mosher et al., 1979), and this reaction appears to be related to wound healing (Duckert, 1972; Folk & Finlayson, 1977; Lorand et al., 1980).

The cross-linking of α_2-plasmin inhibitor to fibrin (Tamaki & Aoki, 1981) or fibrinogen (Ichinose & Aoki, 1982) in the presence of factor $XIII_a$ occurs at a faster rate than with other proteins. Accordingly, in plasma, α_2-plasmin inhibitor and fibrin are considered to be the best amino acceptor and amino donor, respectively, for factor $XIII_a$ (Tamaki & Aoki, 1982; Carmassi & Chung, 1983). Deficiencies of either factor XIII (Folk & Finlayson, 1977; Lorand et al., 1980) or α_2-plasmin inhibitor (Aoki et al., 1979) result in "delayed bleeding" while primary hemostasis in individuals with these traits is normal. This suggests that α_2-plasmin inhibitor also plays an important role in the protection of the fibrin clot from digestion by plasmin (Sakata & Aoki, 1980, 1982).

Factor XIII (M_r 300 000) circulates in blood as a complex with fibrinogen (Greenberg & Shuman, 1982). The molecule occurs as a tetramer (a_2b_2) consisting of two a subunits (M_r 75 000 each) and two b subunits (M_r 80 000 each) (Schwartz et al., 1973; Chung et al., 1974). The a subunit of factor XIII


-30 R E E V P E A H R A S P R E G T S G G E R L Q D L V K S K M
1 CGG GAG GAA GTC CCC GAG GCG CAC AGA GCA AGC CCA CGC GAG GGC ACC TCT GGA GGG GAG CGC CTG CAG GAC CTT GTA AAG TCA AAA ATG
1 S E T S R T A F G G R R A V P P N N S N A A E D D L P T V E
91 TCA GAA ACT TCC AGG ACC GCC TTT GGA GGC AGA AGA GCA GTT CCA CCC AAT AAC TCT AAT GCA GCG GAA GAT GAC CTG CCC ACA GTG GAG
31 L Q G V V P R G V N L Q E F L N V T S V H L F K E R W D T N
181 CTT CAG GGC GTG GTG CCC CGG GGC GTC AAC CTG CAA GAG TTT CTT AAT GTC ACG AGC GTT CAC CTG TTC AAG GAG AGA TGG GAC ACT AAC
61 K V D H H T D K Y E N N K L I V R R G Q S F Y V Q I D F S R
271 AAG GTG GAC CAC CAC ACT GAC AAG TAT GAA AAC AAC AAG CTG ATT GTC CGC AGA GGG CAG TCT TTC TAT GTG CAG ATT GAC TTC AGT CGT
91 P Y D P R R D L F R V E Y V I G R Y P Q E N K G T Y I P V P
361 CCA TAT GAC CCC AGA AGG GAT CTC TTC AGG GTG GAA TAC GTC ATT GGT CGC TAC CCA CAG GAG AAC AAG GGA ACC TAC ATC CCA GTG CCT
121 I V S E L Q S G K W G A K I V M R E D R S V R L S I Q S S P
451 ATA GTC TCA GAG TTA CAA AGT GGA AAG TGG GGG GCC AAG ATT GTC ATG AGA GAG GAC AGG TCT GTG CGG CTG TCC ATC CAG TCT TCC CCC
151 K C I V G K F R M Y V A V W T P Y G V L R T S R N P E T D T
541 AAA TGT ATT GTG GGG AAA TTC CGC ATG TAT GTT GCT GTC TGG ACT CCC TAT GGC GTA CTT CGA ACC AGT CGA AAC CCA GAA ACA GAC ACG
181 Y I L F N P W C E D D A V Y L D N E K E R E E Y V L N D I G
631 TAC ATT CTC TTC AAT CCT TGG TGT GAA GAT GAT GCT GTG TAT CTG GAC AAT GAG AAA GAA AGA GAA GAG TAT GTC CTG AAT GAC ATC GGG
211 V I F Y G E V N D I K T R S W S Y G Q F E D G I L D T C L Y
721 GTA ATT TTT TAT GGA GAG GTC AAT GAC ATC AAG ACC AGA AGC TGG AGC TAT GGT CAG TTT GAA GAT GGC ATC CTG GAC ACT TGC CTG TAT
241 V M D R A Q M D L S G R G N P I K V S R V G S A M V N A K D
811 GTG ATG GAC AGA GCA CAA ATG GAC CTC TCT GGA AGA GGG AAT CCC ATC AAA GTC AGC CGT GTG GGG TCT GCA ATG GTG AAT GCC AAA GAT
271 D E G V L V G S W D N I Y A Y G V P P S A W T G S V D I L L
901 GAC GAA GGT GTC CTC GTT GGA TCC TGG GAC AAT ATC TAT GCC TAT GGC GTC CCC CCA TCG GCC TGG ACT GGA AGC GTT GAC ATT CTA TTG
301 E Y R S S E N P V R Y G Q C W V F A G V F N T F L R C L G I
991 GAA TAC CGG AGC TCT GAG AAT CCA GTC CGG TAT GGC CAA TGC TGG GTT TTT GCT GGT GTC TTT AAC ACA TTT TTA CGA TGC CTT GGA ATA
331 P A R I V T N Y F S A H D N D A N L Q M D I F L E E D G N V
1081 CCA GCA AGA ATT GTT ACC AAT TAT TTC TCT GCC CAT GAT AAT GAT GCC AAT TTG CAA ATG GAC ATC TTC CTG GAA GAA GAT GGG AAC GTG
361 N S K L T K D S V W N Y H C W N E A W M T R P D L P V G F G
1171 AAT TCC AAA CTC ACC AAG GAT TCA GTG TGG AAC TAC CAC TGC TGG AAT GAA GCA TGG ATG ACA AGG CCT GAC CTT CCT GTT GGA TTT GGA
391 G W Q A V D S T P Q E N S D G M Y R C G P A S V Q A I K H G
1261 GGC TGG CAA GCT GTG GAC AGC ACC CCC CAG GAA AAT AGC GAT GGC ATG TAT CGG TGT GGC CCC GCC TCG GTT CAA GCC ATC AAG CAC GGC
421 H V C F Q F D A P F V F A E V N S D L I Y I T A K K D G T H
1351 CAT GTC TGC TTC CAA TTT GAT GCA CCT TTT GTT TTT GCA GAG GTC AAC AGC GAC CTC ATT TAC ATT ACA GCT AAG AAA GAT GGC ACT CAT
451 V V E N V D A T H I G K L I V T K Q I G G D G M M D I T D T
1441 GTG GTG GAA AAT GTG GAT GCC ACC CAC ATT GGG AAA TTA ATT GTG ACC AAA CAA ATT GGA GGA GAT GGC ATG ATG GAT ATT ACT GAT ACT
481 Y K F Q E G Q E E E R L A L E T A L M Y G A K K P L N T E G
1531 TAC AAA TTC CAA GAA GGT CAA GAA GAA GAG AGA TTG GCC CTA GAA ACT GCC CTG ATG TAC GGA GCT AAA AAG CCC CTC AAC ACA GAA GGT
511 V M K S R S N V D M D F E V E N A V L G K D F K L S I T F R
1621 GTC ATG AAA TCA AGG TCC AAC GTT GAC ATG GAC TTT GAA GTG GAA AAT GCT GTG CTG GGA AAA GAC TTC AAG CTC TCC ATC ACC TTC CGG
541 N N S H N R Y T I T A Y L S A N I T F Y T G V P K A E F K K
1711 AAC AAC AGC CAC AAC CGT TAC ACC ATC ACA GCT TAT CTC TCA GCC AAC ATC ACC TTC TAC ACC GGG GTC CCG AAG GCA GAA TTC AAG AAG
571 E T F D V T L E P L S F K K E A V L I Q A G E Y M G Q L L E
1801 GAG ACG TTC GAC GTG ACG CTG GAG CCC TTG TCC TTC AAG AAA GAG GCG GTG CTG ATC CAA GCC GGC GAG TAC ATG GGT CAG CTG CTG GAA
601 Q A S L H F F V T A R I N E T R D V L A K Q K S T V L T I P
1891 CAA GCG TCC CTG CAC TTC TTT GTC ACA GCT CGC ATC AAT GAG ACC AGG GAT GTT CTG GCC AAG CAA AAG TCC ACC GTG CTA ACC ATC CCT
631 E I I I K V R G T Q V V G S D M T V T I Q F T N P L K E T L
1981 GAG ATC ATC ATC AAG GTC CGT GGC ACT CAG GTA GTT GGT TCT GAC ATG ACT GTG ACA ATT CAG TTT ACC AAT CCT TTA AAA GAA ACC CTG
661 R N V W V H L D G P G V T R P M K K M F R E I R P N S T V Q
2071 CGA AAT GTC TGG GTA CAC CTG GAT GGT CCT GGA GTA ACA AGA CCA ATG AAG AAG ATG TTC CGT GAA ATC CGG CCC AAC TCC ACC GTG CAG
691 W E E V C R P W V S G H R K L I A S M S S D S L R H V Y G E
2161 TGG GAA GAA GTG TGC CGG CCC TGG GTC TCT GGG CAT CGG AAG CTG ATA GCC AGC ATG AGC AGT GAC TCC CTG AGA CAT GTG TAT GGC GAG

```
 721     L  D  V  Q  I  Q  R  R  P  S  M STP
2251     CTG GAC GTG CAG ATT CAA AGA CGA CCT TCC ATG TGA ATGC ACAGGAAGCT GAGATGAACC CTGGCATTTG GCCTCTTGTA GTCTTGGCTA AGGAAATTCT

2351     AACGCAAAAA TAGCTCTTGC TTTGACTTAG GTGTGAAGAC CCAGACAGGA CTGCAGAGGG CCCCAGAGTG GAGATCCCAC ATATTTCAAA AACATGCTTT TCCAAACCCA

2461     GGCTATTCGG CAAGGAAGTT AGTTTTTAAT CTCTCCACCT TCCAAAGAGT GCTAAGCATT AGCTTTAATT AAGCTCTCAT AGCTCATAAG AGTAACAGTC ATCATTTATC

2571     ATCACAAATG GCTACATCTC CAAATATCAG TGGGCTCTCT TACCAGGGAG ATTTGCTCAA TACCTGGCCT CATTTAAAAC AAGACTTCAG ATTCCCCACT CAGCCTTTTG

2681     GGAATAATAG CACATGATTT GGGCTCTAGA ATTCCAGTCC CCTTTCTCGG GGTCAGGTTC TACCCTCCAT GTGAGAATAT TTTTCCCAGG ACTAGAGCAC AACATAATTT

2791     TTATTTTTGG CAAAGCCAGA AAAAGATCTT TCATTTTGCA CCTGCAGCCA AGCAAATGCC TGCCAAATTT TAGATTTACC TTGTTAGAAG AGGTGGCCCC ATATTAACAA

2901     ATTGCATTTG TGGGAAACTT AACCACCTAC AAGGAGATAA GAAAGCAGGT GCAACACTCA AGTCTATTGA ATAATGTAGT TTTGTGATGC ATTTTATAGA ATGTGTCACA

3011     CTGTGGCCTG ATCAGCAGGA GCCAATATCC CTTACTTTAA CCCTTTCTGG GATGCAATAC TAGGAAGTAA AGTGGAAGAA TTTATCTCTT TAGTTAGTGA TTATATTTCA

3121     CCCATCTCTC AGGAATCATC TCCTTTGCAG AATGATGCAG GTTCAGGTCC CCTTTCAGAG ATATAATAAG CCCAACAAGT TGAAGAAGCT GGCGGATCTA GTGACCAGAT

3231     ATATAGAAGG ACTGCAGCCA CTGATTCTCT CTTGTCCTTC ACATCACCCA TGTTGAGACC TCAGCTTGGC ACTCAGGTGC TGAAGGGTAA TATGGACTCA GCCTTGCAAA

3341     TAGCCAGTGC TAGTTCTGAC CCAACCACAG AGGATGCTGA CATCATTTGT ATTATGTTCC AAGGCTACTA CAGAGAAGGC TGCCTGCTAT GTATTTGCAA GGCTGATTTA

3451     TGGTCAGAAT TTCCCTCTGA TAAGTCTAGG GTGTGATTTA GGTCAGTAGA CTGTGATTCT TAGCAAAAAA TGAACAGTGA TAAGTATACT GGGGGCAAAA TCAGAATGGA

3561     ATGCTCTGGT CTATATAACC ACATTTCTAA GCCTTTGAGA CTGTTCCTGA GCCTTCAGCA CTAACCTATG AGGGTGAGCT GGTCCCCTCT ATATATACAT CATACTTAAC

3671     TTTACTAAGT AATCTCACAG CATTTGCCAA GTCTCCCAAT ATCCAATTTT AAAATGAAAT GCATTTTGCT AGACAGTTAA ACTGGCTTAA CTTAGTATAT TATTATTAAT

3781     TACAATGTAA TAGAAGCTTA A AATAAA GTT AAACTGATTA TAAAAAAAAA A
```

FIGURE 2: Nucleotide sequence of cDNA coding for the a subunit of human factor XIII and predicted amino acid sequence. Nucleotides 1–87 represent the 5′ noncoding region of the cDNA or a DNA sequence coding for a portion of a leader sequence. The amino acids numbered 1 to 731 represent the mature a subunit circulating in blood. The amino acid residues that are overlined were determined by amino acid sequence analysis, with the exception of residues 459 and 720. The active site Cys at residue 314 is circled. The curved arrow shows the site of cleavage by a processing protease (such as Met amino peptidase) generating the mature protein with an N-terminal Ser. This Ser residue is acetylated in the mature molecule. The straight arrow identifies the cleavage site for the conversion of factor XIII to factor $XIII_a$ by thrombin. Residues marked with solid diamonds are potential Asn-linked glycosylation sites at Asn-X-Ser or Asn-X-Thr sequences. The open diamonds identify Asn residues with little or no carbohydrate. The nucleotide sequence that corresponds to a potential polyadenylation signal is shown in a box, while the poly(A) tail has been underlined.

is also found in megakaryocytes, platelets, placenta, uterus, spleen, prostate, and macrophage (Chung, 1972; Kiesselbach & Wagner, 1972; Henriksson et al., 1985; Muszbek et al., 1985) and contains the catalytic site of the enzyme (Chung et al., 1974). The precise function of the b subunit in the plasma protein is not known, but it is thought to protect or stabilize the a subunit (Cooke, 1974; Folk & Finlayson, 1977; Lorand et al., 1974). During the conversion of factor XIII to factor $XIII_a$, an activation peptide (M_r 4000) is released by thrombin from the amino terminus of each of the a subunits (Schwartz et al. 1973). This reaction is promoted by non-cross-linked fibrin I and II (Lewis et al., 1985). The sequence of the activation peptide has been reported (Takagi & Doolittle, 1974; Nakamura et al., 1975) as well as the amino acid residues at the catalytic site of the a subunit (Holbrook et al., 1973).

Recently, we have determined the amino acid sequence of the b subunit of human factor XIII and found that it is composed of 10 highly homologous tandem repeats (Ichinose et al., 1986). These repeats are also homologous to at least 10 other protein-binding proteins (Davie et al., 1986).

As a step toward understanding the structure–function relationship of factor XIII and the organization of the genes for the a and b subunits, clones coding for the a subunit of factor XIII have been isolated from a human placenta cDNA library and the complete nucleotide sequence determined. Also, approximately 50% of the amino acid sequence for the a subunit was established by amino acid sequence analysis. These experiments have made it possible to establish the complete amino acid sequence for the a subunit of human factor XIII.

Materials and Methods

Restriction endonucleases, nuclease BAL-31, and T4 DNA ligase were purchased from Bethesda Research Laboratories or New England Biolabs. The Klenow fragment of *Escherichia coli* DNA polymerase, bacterial alkaline phosphatase, ATP, deoxynucleotides, dideoxynucleoside triphosphates, M13mp10, M13mp11, M13mp18, M13mp19, pUC9, and pUC19 were supplied by Bethesda Research Laboratories. A human placenta cDNA library (Millan, 1986) was purchased from Clontech Lab., Inc. (Palo Alto, CA). $Na^{125}I$ and ^{32}P-labeled nucleotides were obtained from New England Nuclear, and [α-^{35}S]dATP was provided by Amersham. Normal human plasma was kindly provided by the Pacific Northwest Red Cross Blood Service, Portland, OR.

Factor XIII was purified from human plasma according to the method of Curtis and Lorand (1976). It was converted to factor $XIII_a$ in the presence of thrombin, and the thrombin was then inactivated by hirudin (Sigma). The a′ and b subunits were fractionated by gel filtration on a Bio-Gel A-5m column (Bio-Rad) according to the method of Chung et al. (1974).

Polyclonal antibody against the a′ subunit of factor XIII was raised in rabbits, and the immunoglobulin fraction was purified by ammonium sulfate fractionation, DEAE-Sephadex column chromatography, and affinity chromatography (Canfield & Kisiel, 1982) employing a column containing Sepharose covalently linked to the a′ subunit of factor XIII.

Ten milligrams of factor XIII was dialyzed overnight against 5% HCOOH, and the dialyzate was adjusted to pH 4.0 with

ammonium hydroxide. The precipitate that was enriched with the a subunit (83% a and 17% b subunits) was S-carboxymethylated (Crestfield et al., 1963) and digested with cyanogen bromide. The resulting fragments were separated by gel filtration on a Sephadex G-50 superfine column with 5% HCOOH and further purified on a Waters HPLC system with an Ultrapore C3 reverse-phase column (Altex). The gradient employed consisted of 0.1% trifluoroacetic acid as a mobile phase and 0.8% trifluoroacetic acid in 80% acetonitrile as a mobile-phase modifier. The column was run at a flow rate of 1.5 mL/min, and the eluant was monitored by absorbance at 214 nm. Sequence analysis of the cyanogen bromide peptides was performed with a Beckman automated sequenator, Model 890C, by the method of Edman and Begg (1967). PTH-amino acids were identified by two complementary reverse-phase column systems (Ericsson et al., 1977; Glajch et al., 1985). Peptides arising from the b subunit (Ichinose et al., 1986) were readily identified and discarded from the remaining peptides originating from the a subunit.

Phage from a human placenta cDNA library were screened by the method of Young and Davis (1983) as modified by Foster and Davie (1984). The purified antibody was labeled with $Na^{125}I$ to a specific activity of 6×10^6 cpm/μg and used to screen filters containing phage plated at a density of 1.5×10^5 plaques per 150-mm plate. Positive clones were plaque-purified and then screened with the following oligonucleotide probe: CTCCACGGTGGGCAGGTCGTCCTCG. This probe coded for the amino acid sequence of Ala-Glu-Asp-Asp-Leu-Pro-Thr-Val-Glu that is present in the activation peptide of the a subunit (Takagi & Doolittle, 1974). The nucleotide sequence for the probe was selected by employing the most common codon usage for amino acids for a number of different human proteins (Chen & Barker, 1985). The oligonucleotide was synthesized on an Applied Biosystems oligonucleotide synthesizer (Foster City, CA) and labeled with ^{32}P to a specific activity of 1.1×10^8 cpm/μg.

Phage DNA was prepared from positive clones by the liquid culture lysis method (Silhavy et al., 1984), followed by centrifugation and banding on a cesium chloride step gradient (Degen et al., 1983). cDNA inserts were isolated by digestion of the phage DNA with *Eco*RI endonuclease and then subcloned into plasmid pUC9 or pUC19 (Vieira & Messing, 1982). Additional restriction fragments from the inserts were also subcloned into M13mp10, -11, -18, or -19 in order to obtain overlapping sequences. The cDNA inserts were then sequenced by the dideoxy method (Sanger et al., 1977) with [α-^{35}S]dATP and buffer gradient gels (Biggin et al., 1983). Digestions with nuclease BAL-31 were performed to generate fragments that provided overlapping sequences with restriction fragments (Poncz et al., 1982). DNA sequences were analyzed by the computer program of Textco (W. Lebanon, NH) with an Apple MacIntosh computer (Gross, 1986).

Results and Discussion

A λgtll expression library containing cDNAs prepared from human placenta mRNA (Millan, 1986) was screened for the a subunit of human factor XIII by employing an ^{125}I-labeled affinity-purified rabbit antibody and a ^{32}P-labeled oligonucleotide probe that coded for a portion of the activation peptide. Six positive clones were isolated by screening approximately 3×10^6 phage, and each positive phage was plaque-purified. The cDNA insertes were isolated from the DNA from each clone following *Eco*RI digestion, and the 5′ and 3′ ends of each insert were sequenced. One of the clones with a large cDNA insert (λHFXIIIa3.77) was selected for further sequence analaysis. This insert contained three internal *Eco*RI sites, giving rise to four cDNA fragments upon digestion of the phage DNA with *Eco*RI. These fragments and several additional restriction fragments were subcloned into plasmid pUC9 or pUC19, and their nucleotide sequences were determined by the strategy shown in Figure 1. Five additional overlapping cDNA inserts were also isolated, subcloned, and partially sequenced (Figure 1). The sequence of 3831 nucleotides from these overlapping clones and the predicted amino acid sequence are shown in Figure 2. This DNA sequence codes for the entire amino acid sequence of the mature a subunit of human factor XIII that circulates in blood. The a subunit is composed of 731 amino acids starting with an amino-terminal sequence of Ser-Glu-Thr-Ser. This amino-terminal sequence was reported earlier by Takagi and Doolittle (1974). The carboxyl-terminal Met (nucleotides 2281–2283) is followed by a stop codon (TGA), 1535 base pairs of noncoding sequence, and a potential polyadenylation or processing signal of AATAAA (Proudfoot & Brownlee, 1976). The polyadenylation sequence was located 14 nucleotides upstream from the poly(A) tail of 10 nucleotides. The poly(A) tail was present only in λHFXIIIa3.82.

At present, it is not known whether the a subunit of factor XIII synthesized in placenta remains as a cytoplasmic enzyme or is secreted into plasma. If the placental enzyme is a cytoplasmic protein, the Met at position −1 would funciton as the initiatior Met for biosynthesis. The removal of the initiator Met would be followed by an acetylation reaction leading to the formation of acetylserine at the amino-terminal end of the mature protein (Mikuni et al., 1973; Takagi & Doolittle, 1974; Nakamura et al., 1975). Also, a cytoplasmic protease with a spcificity similar to thrombin would be required for its conversion to factor $XIII_a$. If, however, the a subunit of factor XIII synthesized in placenta is secreted into plasma, then the cDNA would normally code for a hydrophobic leader sequence. The cDNA, as shown in Figure 2, could code for 30 amino acid residues that constitute a portion of a leader peptide. The cDNA, however, does not code for an initiator Met prior to a typical hydrophobic core. Also, the bond between Met and Ser at positions −1 and 1 is not a very favorable cleavage site for signal peptidase (Perlman & Halvorson, 1983; von Heijne, 1983, 1984). Thus, if the a subunit of factor XIII from placenta is secreted into plasma, it is likely that it is synthesized with a prepro-leader sequence. In this case, a processing protease(s) in addition to a signal peptidase would be required to yield the mature protein with an amino-terminal Ser, which is then converted to acetylserine.

A difference in the nucleotide sequence for the a subunit of factor XIII was found at three positions when a comparison of the cDNA inserts was made in regions where overlapping sequences were obtained. Nucleotides 2038, 2041, and 2727 contained A, C, and T, respectively, in λHFXIIIa3.77, λHFXIIIa2.14, and λHFXIIIa3.81 (Figure 2), while λHFXIIIa3.82 contained G, G, and A in the same positions. These differences result in a change in two amino acids (Ile-650 and Gln-651 to Val and Glu). It is likely that these differences are due to a cloning artifact during the preparation of λHFXIIIa3.82 but also could represent a polymorphism that contributes to the microheterogeneity in the a subunit of factor XIII (Board, 1979; Board & Coggan, 1981).

Amino acid sequence analyses were also performed on the cyanogen bromide fragments of the a subunits of human factor XIII isolated from plasma. In these experiments, the a subunit of factor XIII was S-carboxymethylated and digested with cyanogen bromide. The resulting peptide fragments were separated by gel filtration and purified by reverse-phase

HPLC. Eleven of the eighteen cyanogen bromide fragments expected from the cDNA were isolated. Each of the purified cyanogen bromide fragments was subjected to amino acid sequence analysis, and a total of 363 residues were unequivocally identified (residues overlined in Figure 2). These amino acid sequences were in complete agreement with those predicted by the cDNA isolated from placenta. These data support the conclusion that the a subunits of factor XIII from placenta and plasma are identical.

The protein sequence predicted from the cDNA includes six potential Asn-linked glycosylation sites with a sequence of Asn-X-Ser or Asn-X-Thr. These Asn residues are located at positions 17, 46, 541, 556, 613, and 686. Two of these Asn residues have little or no carbohydrate since Asn was readily identified in positions 613 and 686 by the amino acid sequence analysis. Furthermore, carbohydrate was not reported in Asn-17 by Takagi and Doolittle (1974). Since the a subunit of factor XIII contains only 1.5% carbohydrate (Bohn, 1972), it is also possible that positions 46, 541, and 556 may contain little or no carbohydrate. A partial glycosylation of some of these Asn residues, however, could contribute to the microheterogeneity of the a subunit of human factor XIII (Board, 1979; Board & Coggan, 1981).

The a subunit of factor XIII consists of 731 amino acid residues with the following composition: Ala_{37}, Arg_{45}, Asn_{40}, Asp_{47}, $^1/_2$-Cys_9, Gln_{27}, Glu_{48}, Gly_{50}, His_{14}, Ile_{39}, Leu_{48}, Lys_{38}, Met_{19}, Phe_{32}, Pro_{33}, Ser_{45}, acetyl-Ser_1, Thr_{45}, Trp_{15}, Tyr_{29}, Val_{70}. The molecular weight of the polypeptide portion of the molecule was calculated to be 80 488. The addition of 1.5% carbohydrate (Bohn, 1972) gives a molecular weight of approximately 81 700 for each of the a subunits of human factor XIII. This is in good agreement with the value of 75 000 estimated by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (Schwartz et al., 1973; Chung et al., 1974).

Activation of factor XIII by thrombin is due to the cleavage of an activation peptide from the amino terminus of each of the a subunits of the molecule (Schwartz et al., 1973; Mikuni et al., 1973). The amino-terminal sequence of the a subunit of factor XIII deduced from all six cDNA clones was the same as that reported by Takagi and Doolittle (1974), except for an additional Val at residue 34. Accordingly, the cDNA data predict an activation peptide of 37 rather than 36 amino acids. The a subunit of bovine factor XIII contains a Leu residue in position 34 instead of a Val and is also 37 amino acids in length (Nakamura et al., 1976).

The carboxyl-terminal residue of the a subunit of factor XIII was identified as Met, which is the same as that of the bovine molecule (Nakamura et al., 1976). Both the amino- and carboxyl-terminal sequences of the a subunit of factor XIII, however, are totally different from those of tissue transglutaminase (Connellan et al., 1972).

The amino acid sequences at the active sites of plasma factor XIII and tissue transglutaminase have been identified as Gly-Gln-Cys-Trp and Tyr-Gly-Gln-Cys-Trp, respectively (Cooke, 1976; Folk & Cole, 1966). The present results from cDNA and amino acid sequence analyses indicate that the active site sequence of Tyr-Gly-Gln-Cys-Trp starts at residue Tyr-311.

Loss of biological activity for factor $XIII_a$ in the presence of thrombin was found to occur in parallel with the generation of fragments of M_r 56 000 and 24 000. These fragments originate from the a subunits of the molecule (Schwartz et al., 1973). The amino-terminal residue of the smaller fragment has been reported as Ser (Folk & Finlayson, 1977). Thus, the Arg-515–Ser-516 peptide bond in the a subunit is probably the site of the thrombin cleavage forming the smaller carboxyl-terminal fragment of 216 amino acids.

The nine $^1/_2$-Cys residues identified by the cDNA coding for the a subunit of factor XIII are in fair agreement with the six –SH groups reported earlier (Chung et al., 1974). These residues occur with free –SH groups in the molecule (Chung et al., 1974).

A computer-assisted analysis using a Dayhoff program (Dayhoff et al., 1983) revealed that the amino acid sequence for the a subunit of human factor XIII is unique, and little significant homology was found with any other protein other than the active site of transglutaminase and minor sequence homology with the γ-subunit of acetylcholine receptor (Nef et al., 1984; Shibahara et al., 1985).

A crude placenta factor XIII concentrate has been used in replacement therapy for hemorrhage in both inherited and acquired factor XIII deficiency (Trobisch & Egbring, 1972; Kuratsuji et al., 1982). Recently, the application of a factor XIII concentrate has been extended to disorders in postoperative wound healing (Baer et al., 1980; Mishima et al., 1983) and scleroderma (Delbarre et al., 1981). Accordingly, the cloning of human factor XIII and its preparation by techniques of recombinant DNA offer a potential source of this protein that is free of viral contamination for clinical use.

ACKNOWLEDGMENTS

We thank Drs. Patrick S. H. Chou and Yim Foon Lee for kindly providing the oligonucleotide and Drs. Donald Foster, Steven Leytus, and Barbara Schach for their help and advice. We also thank Brad A. McMullen for his assistance in the amino acid sequence analysis, Lois Swenson for her assistance in the preparation of the manuscript, and Dr. Frans Peetoom for kindly making available human plasma for the preparation of factor XIII.

Registry No. Blood coagulation factor XIII, 9013-56-3; blood coagulation factor XIII (human subunit a protein moiety reduced), 104641-94-3; blood coagulation factor XIIIa (human subunit a protein moiety reduced), 104641-95-4; blood coagulation factor XIIIa, 9067-75-8.

REFERENCES

Aoki, N., Saito, H., Kamiya, T., Koie, K., Sakata, Y., & Kobakura, M. (1979) *J. Clin. Invest. 63*, 877–884.

Baer, V. U., Bauknecht, J., Stangl, Th., & Haring, R. (1980) *Zentralbl. Chir. 105*, 642–651.

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

Board, P. G. (1979) *Am. J. Hum. Genet. 31*, 116–124.

Board, P. G., & Coggan, M. (1981) *Hum. Genet. 59*, 135–136.

Bohn, H. (1970) *Thromb. Diath. Haemorrh. 23*, 455–468.

Canfield, W. M., & Kisiel, W. (1982) *J. Clin. Invest. 70*, 1260–1272.

Carmassi, F., & Chung, S. I. (1983) *Prog. Fibrinolysis 6*, 281–285.

Chen, R., & Doolittle, R. F. (1970) *Proc. Natl. Acad. Sci. U.S.A. 66*, 472–479.

Chen, H. R., & Barker, W. C. (1985) *Trends Genet. 1*, 221–223.

Chung, S. I. (1972) *Ann. N.Y. Acad. Sci. 202*, 240–255.

Chung, S. I., Lewis, M. S., & Folk, J. E. (1974) *J. Biol. Chem. 249*, 940–950.

Connellan, J. M., Chung, S. I., Whetzel, N. K., Bradley, L. M., & Folk, J. E. (1971) *J. Biol. Chem. 246*, 1093–1098.

Cooke, R. D. (1974) *Biochem. J. 141*, 683–691.

Crestfield, A. M., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622–627.

Curtis, C. G., & Lorand, L. (1976) *Methods Enzymol. 45*, 177–191.
Davie, E. W., Ichinose, A., & Leytus, S. (1986) *Cold Spring Harbor Symp. Quant. Biol.* (in press).
Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.
Delbarre, F., Godeau, P., & Thivolet, J. (1981) *Lancet 2*, 204.
Degen, S. J. F., MacGillivray, R. T. A., & Davie, E. W. (1983) *Biochemistry 22*, 2087–2097.
Duckert, F. (1972) *Ann. N.Y. Acad. Sci. 202*, 190–199.
Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80–91.
Ericsson, L. H., Wade, R. D., Gagnon, J., McDonald, R. M., Granberg, R., & Walsh, K. A. (1977) in *Solid Phase Methods in Protein Sequence Analysis* (Previero, A., & Coletti-Previero, M.-A., Eds.) pp 137–142, Elsevier/North-Holland, Amsterdam.
Folk, J. E., & Cole, P. W. (1966) *J. Biol. Chem. 241*, 3238–3240.
Folk, J. E., & Finlayson, J. S. (1977) *Adv. Protein Chem. 31*, 1–133.
Foster, D., & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4766–4770.
Friedman, M., Krull, L. H., & Cavins, J. F. (1970) *J. Biol. Chem. 245*, 3868–3871.
Gaffney, P. J., & Whitaker, A. N. (1979) *Thromb. Res. 14*, 85–94.
Glajch, J. L., Gluckman, J. C., Charikofsky, J. G., Minor, J. M., & Kirkland, J. J. (1985) *J. Chromatogr. 318*, 23–39.
Greenberg, C. S., & Shuman, M. A. (1982) *J. Biol. Chem. 257*, 6096–6101.
Gross, R. H. (1986) *Nucleic Acids Res. 14*, 591–596.
Henriksson, P., Becker, S., Lynch, G., & McDonagh, J. (1985) *J. Clin. Invest. 76*, 528–534.
Holbrook, J. J., Cooke, R. D., & Kingston, I. B. (1973) *Biochem. J. 135*, 901–903.
Ichinose, A., & Aoki, N. (1982) *Biochim. Biophys. Acta 706*, 158–164.
Ichinose, A., McMullen, B. A., Fujikawa, K., & Davie, E. W. (1986) *Biochemistry 25*, 4633–4638.
Kiesselbach, T. H., & Wagner, R. H. (1972) *Ann. N.Y. Acad. Sci. 202*, 318–328.
Kuratsuji, T., Oikawa, T., Fukumoto, T., Shimizu, S., Iwasaki, Y., Tomita, Y., Meguro, T., & Yamada, K. (1982) *Haemostasis 11*, 229–234.
Lewis, S. D., Janus, T. J., Lorand, L., & Shafer, J. A. (1985) *Biochemistry 24*, 6772–6777.
Lorand, L. (1972) *Ann. N.Y. Acad. Sci. 202*, 6–30.
Lorand, L., & Jacobsen, A. (1962) *Nature (London) 195*, 911–912.
Lorand, L., Gray, A. J., Brown, K., Credo, R. B., Curtis, C. G., Domanik, R. A., & Stenberg, P. (1974) *Biochem. Biophys. Res. Commun. 56*, 914–922.
Lorand, L., Losowsky, M. S., & Miloszewski, K. J. M. (1980) *Prog. Hemostasis Thromb. 5*, 245–290.
Mikuni, Y., Iwanaga, S., & Konishi, K. (1973) *Biochem. Biophys. Res. Commun. 54*, 1393–1402.
Millan, J. L. (1986) *J. Biol. Chem. 261*, 3112–3115.
Mishima, Y., Nagao, F., Ishibiki, K., Matsuda, M., & Nakamura, N. (1984) *Chirurg 55*, 803–808.
Mockros, L. F., Roberts, W. W., & Lorand, L. (1974) *Biophys. Chem. 2*, 164–169.
Mosher, D. F. (1975) *J. Biol. Chem. 250*, 6614–6621.
Mosher, D. F., Schad, P. E., & Kleinman, H. K. (1979) *J. Clin. Invest. 64*, 781–787.
Muszbek, L., Adany, R., Szegedi, G., Polgar, J., & Kavai, M. (1985) *Thromb. Res. 37*, 401–410.
Nakamura, S., Iwanaga, S., & Suzuki, T. (1975) *J. Biochem. (Tokyo) 78*, 1247–1266.
Nef, P., Mauron, A., Stalder, R., Alliod, C., & Ballivet, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7975–7979.
Perlman, D., & Halvorson, H. O. (1983) *J. Mol. Biol. 167*, 391–409.
Pisano, J. J., Bronzert, T. J., & Peyton, M. P. (1972) *Ann. N.Y. Acad. Sci. 202*, 98–113.
Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., & Surrey, S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4298–4302.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Roberts, W. W., Lorand, L., & Mockros, L. F. (1973) *Biorheology 10*, 29–42.
Sakata, Y., & Aoki, N. (1980) *J. Clin. Invest. 65*, 290–297.
Sakata, Y., & Aoki, N. (1982) *J. Clin. Invest. 69*, 536–542.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schwartz, M. L., Pizzo, S. V., Hill, R. L., & McKee, P. A. (1973) *J. Biol. Chem. 248*, 1395–1407.
Shen, L., & Lorand, L. (1983) *J. Clin. Invest. 71*, 1336–1341.
Shen, L. L., Hermans, J., McDonagh, J., McDonagh, R. P., & Carr, M. (1975) *Thromb. Res. 6*, 255–265.
Shibahara, S., Kubo, T., Perski, H. J., Takahashi, H., Noda, M., & Numa, S. (1985) *Eur. J. Biochem. 146*, 15–22.
Silhavy, T. J., Berman, W. L., & Enquist, L. W. (1984) *Experiments with Gene Fusions*, pp 140–141, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Takagi, T., & Doolittle, R. F. (1974) *Biochemistry 13*, 750–756.
Tamaki, T., & Aoki, N. (1981) *Biochim. Biophys. Acta 661*, 280–286.
Tamaki, T., & Aoki, N. (1982) *J. Biol. Chem. 257*, 14767–14772.
Trobisch, H., & Egbring, R. (1972) *Dtsch. Med. Wochenschr. 97*, 449–502.
Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.
von Heijne, G. (1983) *Eur. J. Biochem. 133*, 17–21.
von Heijne, G. (1984) *J. Mol. Biol. 173*, 243–251.
Young, R. A., & Davis, R. W. (1983) *Science (Washington, D.C.) 222*, 778–782.