Characterization of cDNA coding for human factor XIIIa

(DNA sequence analysis/amino acid sequence identity/blood coagulation)

Ulrich Grundmann[†], Egon Amann, Gerd Zettlmeissl, and Hans A. Küpper

Department of Molecular Biology, Research Institutes, Behringwerke AG, Postfach 1140, 3550 Marburg, Federal Republic of Germany

Communicated by Frank W. Putnam, July 15, 1986

ABSTRACT A cDNA library prepared from human placenta has been screened for sequences coding for factor XIIIa, the enzymatically active subunit of the factor XIII complex that stabilizes blood clots through crosslinking of fibrin molecules. Two oligonucleotides, based on the amino acid sequences of tryptic peptides of factor XIIIa, were used as hybridization probes. Of 0.36×10^6 independent recombinants, 1 clone was identified that hybridized to both probes. The insert of 1704 base pairs coded for the amino-terminal 541 amino acid residues of the mature factor XIIIa molecule. Blot-hybridization analysis using this cDNA as a probe showed that the factor XIIIa mRNA from placenta has a size of approximately 4000 bases. The insert was used to rescreen cDNA libraries and to identify further factor XIIIa-specific sequences. The total length of the isolated factor XIIIa cDNA is 3905 bases, and it codes for a protein of 732 amino acids. In spite of the presence of factor XIII in blood plasma, we could not identify a leader sequence typical for secreted proteins.

Factor XIII (FXIII) is a blood coagulation protein that is distributed both extracellularly (in plasma) and intracellularly (in megakaryocytes and platelets) (1, 2). In placenta the exact cellular location of FXIII is not known. Data obtained by indirect immunofluorescence suggest, however, that enzymatically active FXIII is located both intracellulary and extracellulary in the stroma of placental villi (3). FXIII is the last zymogen to become activated in the coagulation cascade during the blood-clotting process of vertebrates (4). FXIII molecules from blood platelets and megakaryocytes and from placental tissue are indistinguishable by several criteria and are each dimers composed of two identical a subunits (a_2) with a monomeric molecular weight of about 80,000 (3). Blood plasma FXIII, on the other hand, is a tetramer (a_2b_2) of molecular weight 300,000-350,000. It contains two noncatalytic b subunits ($M_r \approx 90,000$) in addition to its two catalytic a subunits, which are believed to be identical to those of platelet, megakaryocyte, and placental origin (3). Upon activation by thrombin and in the presence of Ca^{2+} , the plasma-derived FXIII dissociates its noncatalytic b subunits and yields the same active enzyme, FXIIIa*, as zymogens derived from megakaryocytes, platelets, or placenta. The enzymatic active form of FXIII, the FXIIIa* [also called activated fibrin-stabilizing factor, fibrinoligase, or plasma transglutaminase, EC 2.3.2.13 (protein-glutamine:amine yglutamyltransferase)] contains a cysteine in its active center and acts as a transglutaminase in the blood coagulation process. It catalyzes the formation of a γ -glutamyl- ε -lysyl bond between fibrin molecules (5), fibrin and α_2 -plasmin inhibitor (6), fibronectin molecules (7), fibronectin and fibrin (7), and fibronectin and collagen (8) during the formation of an insoluble clot. The formation of covalent crosslinks by FXIII is important not only for normal blood coagulation but also for wound healing and placental retention (3).

Normal concentrations of FXIII in plasma are 10–20 mg/liter (9). Deficiency of FXIII results in inefficient wound and bone healing, and women with FXIII deficiency have a high risk of early abortions. The hereditary deficiency of FXIII is rare, with a frequency of 1 in 2×10^6 (10). Hereditary and acquired FXIII deficiencies are treated by substitution therapy.

Inherited deficiency of FXIII usually results from a failure to produce active subunit a or from the production of functionally abnormal forms (11, 12). Family studies of genetic variants of both a and b subunits have established that the structural loci for the two subunits are on different autosomes (13, 14). Moreover, it has been reported recently that the gene for the a subunit is distal to *HLA* on chromosome 6 (15).

The activation of FXIII is a multistep process involving (i) thrombin cleavage of the bond between Arg-37 and Gly-38 and release of the NH₂-terminal 37-residue peptide (16), (ii) a conformational change induced by calcium ions, and (iii) in the case of the plasma zymogen, weakening of the strong noncovalent interaction between subunits a and b (17).

Prior to this study, two groups have reported partial amino acid sequence data for enzymatically active human FXIII: (i)the amino acid sequence of the 37-residue peptide released during the activation of FXIII by thrombin (16) and (ii) a Gly-Gln-Cys-Trp sequence exposed at the active site of the molecule (18).

Here we report the isolation and DNA sequence determination of a cDNA coding for the placental a subunit, which we refer to as FXIIIa.

MATERIALS AND METHODS

Reagents. Restriction endonucleases were purchased from New England Biolabs or Boehringer Mannheim. Reverse transcriptase and RNase H were from P. H. Stehelin (Basel, Switzerland). Phage λ gt10 DNA and an *in vitro* packaging system were obtained from Vector Cloning Systems (San Diego, CA). *Eco*RI methylase was purchased from Bethesda Research Laboratories.

Isolation of RNA and Preparation of cDNA Libraries. RNA was prepared from a mature human placenta by the standard guanidinium thiocyanate extraction procedure (19). $Poly(A)^+$ RNA was selected by oligo(dT)-cellulose chromatography (20). $Poly(A)^+$ RNA ($\approx 4 \mu g$) was copied into cDNA with reverse transcriptase, and double-stranded cDNA was prepared by the method of Gubler and Hoffman (21). After treatment with T4 polymerase, the cDNA was methylated with *Eco*RI methylase and ligated to *Eco*RI linkers. The products were cleaved with *Eco*RI and separated in a 5–20% KOAc gradient in 1 mM EDTA containing 1 μg of ethidium bromide per ml at 50,000 rpm for 3 hr in a Beckman SW 65 rotor. cDNA larger than 1000 base pairs (bp) was pooled and precipitated by ethanol. An aliquot of 50 ng was ligated to

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.

Abbreviations: bp, base pair(s); kb, kilobase(s); FXIII, factor XIII; FXIIIa, a subunit of FXIII. [†]To whom reprint requests should be addressed.

*Eco*RI-cleaved λ gt10 DNA (1 μ g) and packaged *in vitro*. Phages were plated on *Escherichia coli* C600HFL. About 1 × 10⁶ clear plaques were obtained.

Screening of a Human Placenta Library. The cDNA library was screened with two oligonucleotides derived from FXIIIa peptide sequences. Information on different FXIIIa peptides was made available to us by N. Takahashi and F. W. Putnam (personal communication). We chose the sequence Met-Met-Asp-Ile-Thr-Asp-Thr for the synthesis of a "short" 20-mer, degenerated probe with the sequence: ATG ATG GAY ATH ACN GAY AC, in which Y = T or C, H = T or A or C, and N = T or A or C or G. In addition, a "long" 66-mer, undegenerated probe was constructed according to the rules of Lathe (22) by using the following strategy. First, two overlapping oligonucleotides corresponding to the sequences Tyr-Gly-Gln-Phe-Glu-Asp-Gly-Ile-Leu-Asp-Thr-Cys-Leu and Asp-Thr-Cys-Leu-Tyr-Val-Met-Asp-Arg-Ala-Gln-Met-Asp were synthesized. The sequences of the two oligonucleotides A and B are:

(A) 5' TATGGCCAGTTTGAGGATGGCATCCTG<u>GACA</u>-<u>CCTGCCTG</u> 3'

and

(B) 3' <u>CTGTGGACGGACATACACTACCTGGCCCGGG-</u> TCTACCTG 5',

with an overlapping complementary sequence of 12 nucleotides corresponding to the amino acid sequence Asp-Thr-Cys-Leu (underlined).

Next, the two oligonucleotides were separately labeled with T4 polynucleotide kinase and $[\gamma^{-32}P]ATP$. The labeled oligonucleotides were hybridized, and the double-strand was completed by incubation with DNA polymerase I, large fragment, using $[\alpha^{-32}P]$ ATP and three unlabeled deoxynucleotides. The resulting double-stranded probe was purified by electrophoresis on a 15% polyacrylamide gel, the specific activity being 2×10^8 Bq/µg. Both the 20-mer and the 66-mer probe were used for screening of the placenta cDNA library. For hybridization, the filters were washed in 0.45 M NaCl/0.045 M sodium citrate, pH 7/0.1% NaDodSO4 at 65°C overnight and hybridized after prehybridization in 0.9 M NaCl/0.006 M EDTA/0.09 M Tris, pH 8.3/0.2% polyvinylpyrrolidone/0.2% Ficoll/0.2% bovine serum albumin at calculated temperatures (e.g., 42°C for the 20-mer and 65°C for the 66-mer or nick-translated cDNA probes). All washes were performed in 0.90 M NaCl/0.09 M sodium citrate, pH 7/0.5% sodium pyrophosphate (23).

Characterization of cDNA Inserts. Positive phages were plaque-purified and grown in liquid culture. After lysis the bacterial debris was sedimented, and bacteriophage DNA was extracted as described (24) but without centrifugation in cesium chloride. cDNA inserts were mapped by suitable restriction endonuclease double digests and limited digests of ³²P-end-labeled fragments (25). Overlaps were specified by Southern blot analysis (26) and sequence analysis. Nucleotide acid sequence was determined by the chemical method of Maxam and Gilbert (27) as well as by the dideoxy method of Sanger (28). DNA sequence data were analyzed by University of Minnesota Apple II sequence analysis programs.

RESULTS

Isolation of Human FXIIIa cDNA. The amino acid sequence of human placental FXIIIa has been determined by classical methods involving Edman degradation (29). From the available data (N. Takahashi and F. W. Putnam, personal communication), we selected suitable amino acid sequences to design oligonucleotide probes (see *Materials and Methods* and Fig. 3). A 66-mer unique probe was constructed according to rules published by Lathe (22), and a 20-mer was synthesized that was 48-fold degenerate.

Clear plaques (3.6×10^5) of the human placenta cDNA library were screened with either the 20-mer or the 66-mer oligonucleotide as probe. Seven plaques hybridized to either one of the probes, but only one (λ gt10-12) hybridized specifically when rescreened with both probes.

Phage DNA from this positive plaque was isolated, digested with EcoRI, and analyzed by Southern blotting with the same two probes. The analysis resulted in a large (1161 bp) EcoRI fragment hybridizing to the 66-mer probe and a small (543 bp) EcoRI fragment hybridizing to the 20-mer probe. As the 66-mer unique probe was constructed from partially complementary 39-mer oligonucleotides A and B, the λ gt10-12 insert was also probed with A and B separately. Both 39-mer probes hybridized specifically to the large 1161-bp fragment, which was another strong indication that the cDNA insert indeed carries FXIIIa-specific sequences; this subsequently was confirmed by DNA sequencing.

The 1704-bp insert of λ gt10-12 was used to screen the total cDNA library of 1 × 10⁶ clear plaques for further FXIIIaspecific sequences. A total of 16 recombinants showed positive signals and were further characterized. Four of the positive signals showed DNA inserts located at the 5' end of the λ gt10-12 insert, but none of them were found to extend more than six bases from the 5' end of the λ gt10-12 insert sequence. Several cDNA inserts were identified from the 3' end of the FXIIIa coding sequence, of which only the insert of λ gt10-11 extends more than 2 kilobases (kb) beyond λ gt10-12, the overlap between both inserts being 237 bp. Nucleotide analysis showed that this clone specifies the 3' end of the FXIIIa mRNA.

Attempts using the primer extension approach to obtain clones from the 5' end of the mRNA that extend further upstream into the nontranslated region were without success.

Identification of FXIIIa-Specific mRNA. The size of mRNA coding for FXIIIa was determined for $poly(A)^+$ RNA from human placenta and liver by electrophoresis through gels containing formaldehyde and by subsequent blot-hybridization analysis with the 1704-bp insert of λ gt10-12 as a hybridization probe. One mRNA species of about 4 kb hybridized to the FXIIIa cDNA probe. This size of the FXIIIa mRNA agrees well with the size of the mRNA deduced from isolated cDNA inserts shown in Fig. 1. Only the placental poly(A)⁺ RNA showed a signal, while no such signal could be detected in the case of liver poly(A)⁺ RNA (Fig. 2). This finding agrees with our observation that screening of liver cDNA libraries containing a total of 5×10^5 independent recombinants with



FIG. 1. Restriction map of human placental FXIIIa cDNA and the location of coding sequence. The figure shows the restriction map of FXIIIa cDNA and the location of the coding sequence (heavy bar). N and C indicate the amino and carboxyl terminus of the FXIIIa protein molecule, respectively. The overlapping cDNA inserts of clones λ gt10-11, λ gt10-12, and λ gt10-20 cover the complete FXIIIa cDNA sequence.



FIG. 2. Blot-hybridization analysis of mRNA. Poly(A)⁺ RNA (4 μ g) of human placenta (lane 2) and liver (lane 3) were transferred after electrophoresis through a denaturing agarose gel containing formaldehyde to a nitrocellulose filter. The filter was hybridized with the nick-translated cDNA insert of λ gt10-12 at 65°C, washed as described, and exposed to Kodak XAR-5 films in DuPont cassettes with intensifying screens "Quanta Fast Detail." Lane 1 shows the positions of human and bacterial (*Escherichia coli*) ribosomal RNAs.

the oligonucleotide probes and even with the λ gt10-12 insert did not result in isolation of positive clones.

Analysis of cDNA and Deduced Amino Acid Sequence. Nucleic acid sequencing was carried out by the chemical method of Maxam and Gilbert as well as by the dideoxy method of Sanger. This way 90% of the total cDNA was sequenced by two independent methods. The entire nucleotide sequence of the FXIIIa cDNA as determined from clones $\lambda gt10-11$, $\lambda gt10-12$, and $\lambda gt10-20$ is displayed in Fig. 3. The deduced amino acid sequence is also shown. The cDNA shows an open reading frame up to nucleotide 2280 that carries the total information for 732 amino acids, including the initiator methionine for FXIIIa. The 5' nontranslated sequence extends 84 bases upstream from the first AUG codon. Nucleotides 44 to 53 in the 5' nontranslated region are partially complementary to the 3'-terminal region of the eukaryotic 18S RNA (30) and thus could represent a ribosome binding site:

The stop codon UGA is followed by about 1.6 kb of noncoding region, which ends in a poly(A) tail of 89 residues. At the 3' end of the mRNA, the putative poly(A) addition signal AATAAA is located 15 nucleotides upstream of the beginning of the poly(A) tail. The codon usage for the FXIIIa mRNA is in good agreement with the data for eukaryotic proteins presented by Lathe (22). In fact, the 66-mer that was used as a hybridization probe shows differences in only seven positions to the actual sequence (*Materials and Methods* and Fig. 3).

From the deduced amino acid sequence, the mature FXIIIa, after removal of the initiator methionine, is a protein of 731 amino acids and has a molecular weight of 82,996. It has six potential glycosylation sites (asparagine in positions 17, 46, 541, 556, 613, and 686) located at the NH₂-terminal

third and the COOH-terminal third of the molecule and nine cysteine residues clustered with one exception (Cys-695) at the middle third of the molecule. Through activation by thrombin, the Arg-Gly bond at amino acid positions 37-38 is cleaved. The sequence of the 37 residues released in the activation step agrees with data published by Nakamura *et al.* (31) and Takahashi *et al.* (29) but shows an additional value in position 35 when compared to the data of Takagi and Doolittle (16). The sequence Gly-Gln-Cys-Trp at the active site was located at positions 312-315 of the FXIIIa molecule.

DISCUSSION

The amino acid sequence predicted here from the cDNA sequence is in agreement with the sequence of FXIIIa completed by classical protein sequence analysis (29). Minor changes (Leu⁸⁸ \rightarrow Phe⁸⁸, Gln⁶⁵¹ \rightarrow Glu⁶⁵¹), which have been verified after exchange of data prior to publication, might be due to a polymorphism of the FXIIIa gene within the population. While the purified protein represents a pool collected from many individuals, the mRNA was isolated from an individual mature placenta. A heterogeneity also was observed at positions 77 and 78, where the residues Gly-Lys could be detected in addition to the residues Arg-Arg that are specified by the cDNA-derived sequence. The fact that the extra methionine at the 3' end was not found by protein sequencing might be due to protease activity and heterogeneity at the 3' end of the isolated FXIIIa protein.

It was surprising, however, that the deduced amino acid sequence of the cloned FXIIIa mRNA does not contain a leader peptide, which is common in secreted proteins. The translational initiation codon was assigned to the first inframe AUG, which is located 84 bases downstream from the 5' end of the isolated cDNA sequence.

As the open reading frame extends to the 5' very end of the sequence, in theory one could assign another 28 amino acids 5' to the methionine. There are, however, several reasons that make this methionine very likely to be the starting point of translation of the FXIIIa molecule. First, of four cDNA inserts that carry the 5' end of the mRNA molecule, all stopped within six bases at the same place. This was taken as an indication that the 5' end of the mRNA had been reached. Second, the preceding 28 amino acids that could be assigned show no resemblance to a typical leader peptide-e.g., have no stretches of hydrophobic amino acids. Third, the first amino acid of the mature FXIIIa molecule is a serine, which is acetylated (16, 29, 31). This is in agreement with an analysis of 1764 eukaryotic protein sequences that showed that serine. alanine, or glycine in the second position favor removal of the initiator methionine and N-acetylation of the NH2-terminal amino acid (32). Fourth, the third base preceding the codon for the first methionine is an adenine, which is typical for the initiation of translation in eukaryotes (33, 34). Fifth, 41 nucleotides upstream from the initiation codon appears a putative ribosome-binding consensus sequence CCTCTG, which is frequently found in eukaryotic mRNAs (35). The above points indicate that FXIIIa does not have an NH2terminal signal sequence, which is typical for an intracellular protein. However, FXIIIa is obtained as an extracellular protein from plasma in association with the b subunit. The mechanism that promotes the release from the corresponding tissue has yet to be established.

FXIIIa contains six potential glycosylation sites (Fig. 3). Therefore, it was surprising that recent protein sequence analysis did not detect carbohydrates (29) in placental FXIIIa. In the literature, glycosylation of placental FXIIIa has been reported to be 1.5% (3).

The deduced amino acid sequence shows nine cysteine residues. However, from protein analysis there is evidence for only one of the cysteines being involved in a disulfide 0

| GAGGAAGTCCCCCGAGGCGCACAGAGCAAGCCCACGCGAGGGCACCTCTGGAGGGGAGCGCCTGCAGGACCTTGTAAAGTCAAAA 84 | |
|---|---|
| m_{ef}^{-1} ser glu the ser arg the alla phe gly gly arg arg alla val pro pro as as as ser at the tag acc sec the gra gra arg arg alla val pro pro as as an set to take the tag arg arg arg arg arg arg arg arg arg a | 520 MET ASP PHE GLU VAL GLU ASN ALA VAL LEU GLY LYS ASP PHE LYS LEU SER ILE THR PHE ATG GAC THE GAA GTG GAA ANT GET GTG CTG GGA AMA GAC THE ANG CTC THE ATC ATC ATC ATC 1704 |
| AN ALA ALA GLU ASP ASP LEU PRO THR VAL GLU LEU GLN GLY VAL VAL PRO ARG GLY VAL AAT GCA GCG GAA GAT GAC CTG CCC ACA GTG GAG CTT CAG GGC GTG GTG CCC CGG GGC GTC 204 | ARG ANN ANN SER HIS ANN ARG THE THE HEE THE ALA THE LEU SER ALA ANN HEE THE ARE HEE AND ACT AND AND AND HEE THE ARE AND ACT AND |
| 40 ASN LEU GLN GLU PHE LEU ASN VAL THR SER VAL HIS LEU PHE LYS GLU ARG TRP ASP THR AAC CTG CAA GAG TTT CTT AAT GTC ACG AGC GTT CAC CTG TTC AAG GAG AGA TGG GAC ACT 264 | TR THR GLY VAL PRO LYS ALA GLU PHE LYS LYS GLU THR PHE ASP VAL THR LEU GLU PHO TAC ACC GGG GTC COG ANG GCA GAG TTC ANG ANG GAG ACG TTC GAC GTG ACG CTG GAG GCA 1824 |
| 60 Asin Lys val, asp his his thr asp lys tyr qlu asin asin lys leu ile val, arg arg qly Aac aag gtg gac cac cac act gac aag tat gaa aac aac aag ctg att gtc ogc ag ggg 324 | SAO LEU SER PHE LYS LYS QUI ALA VAL LEU ILE QLN ALA QLY QLU TYR MET QLY QLN LEU LEU TTG TCC TTC AAG AAA GAG GGG GTG CTG ATC CAA GGC GGC GAG TAC ATG GGT CAG CTG CTG |
| 80 Gun ser phe tyr val gun ile asp leu ser ang pro tyr asp pro arg ang asp leu phe Cag ict itc tat gtg cag att gac ctc agt cgt cca tat gac ccc aga agg gat ctc ttc 384 | GUU GUN ALA SER LEU HIS PHE PHE VAL THR ALA ARG ILE ASN GLU THR ARG ASP VAL LEU GAA CAA GCG TCC CTG CAC TTC TTT GTC ACA GCT CGC ATC AST GAG ACC AGG GAT GTT CTG |
| 100 Arg val glu tyr val ile gly arg tyr pro gln glu asn lys gly thr tyr ile pro val agg gtg gaa tac gtc att ggt ggc tac cca cag gag aac aag gga acc tac atc cca gtg | 620 ALA LYS GLN LYS SER THR VAL LEU THR ILE PRO GLU ILE ILE LYS VAL ANG GLU THR GCC AAG CAA AAG TCC ACC GTG CTA ACC ATC CTC GAG ATC ATC ATC AAG GTC CGT GGC ACT. |
| 120 PRO ILLE VAL SER QLU LEU QLN SER QLY LYS TRP QLY ALA LYS ILE VAL MET ARG QLU ASP CCT ATA GTC TCA GAG TTA CAA AGT GGA AAG TGG GGG GCC AAG ATT GTC ATG AGA GAG GAC | 640 GLN VAL, VAL, GLY SER ASP MET THR VAL, THR VAL, GLN PHE THR ASN PRO LEU LYS, GLU THR CAG GTA GTT GGT TCT GAC ATG ACT GTG ACA GTT CAG TTT ACC AAT CCT TTA AAA GAA ACC. |
| 140 ARG SER VAL ARG LEU SER ILE GLN SER SER PRO LYS CYS ILE VAL GLY LYS PHE ARG MET AGG TCT GTG CGG CTG TCC ATC CAG TCT TCC CCC AAA TGT ATT GTG GGG AAA TTC CGC AEG | 660 Leu arg asn val trp val his leu asp gly pro gly val thr arg pro net lys lys net CTG cca aat gtc tgg gta cac ctg gat ggt cct gga gta aca aga cca atg ang ang atg |
| 160 TYR VAL ALA VAL TRP THR PRO TYR GLY VAL LEU ARG THR SER ARG ASN PRO GLU THR ASP TAT GTT GCT GTC TGG ACT CCC TAT GGC GTA CTT CGA ACC AGT CGA AAC CCA GAA ACA GAC | 680 PHE ARG GLU ILE ARG PRO ASN SER THR VAL GLN TRP GLU GLU VAL CVS ARG PRO TRP VAL TTC CCT GAA ATC CCG CCC AAC TCC ACC GTG CAG TGG GAA GAA GTG TGC CGG CCC TGG GTC |
| 180 Thr tyr ile leu phe asn pro trp cys glu asp asp ala val tyr leu asp asn glu lys Agg tac att ctc ttc aat gct tgg tgt gaa gat gat gat gt gtg tat ctg gac aat gag aaa | 2184 700 Ser Gly his Arg Lys Leu ile ala ser met ser ser asp ser Leu arg his val tyr Gly TCT gog cat cog arg cit ata goc agc atg acc atg gac to ctg aca cat gtg tat goc |
| 200 Gu arg gu gu tyr val leu asn asp ile gly val ile phe tyr gly gu val asn asp Gaa aga gag tat gtc ctg aat gac atc ggg gta att ttt tat gga gag gtc aat <u>ga</u> c | 720 GLU LEU ASP VAL GLN ILE GLN ARG ARG PRO SER MET \$\$\$ GAG CTG GAC GTG CAG ATT CAA AGA CGA CCT TCC ATG TGA ATGCACAGGAAGCTGAGATGAAQCCTG |
| 220 Lie lys thr arg ser trop ser tyr gly gln phe glu asp gly ile leu asp thr cys leu Atc arg acc rag arg trop ser tyr gly gln phe glu asp gly ile leu asp thr cys leu | 2510 GCATTTGGCCTCTTGTAGTCTTGGCTAAGGAAATTCTAACGCAAAAATAGCTCTTGCTTTGACTTAGGTGTGAAGACCCA 2590 |
| 240 TYR VAL MET ASP ARG ALA GUN MET ASP LEU SER GLY ARG GLY ASN PRO ILE LYS VAL SER | GACAGGACTGCAGAGGGCCCCAGAGTGGAGATCCCACATATTTCAAAAACATACTTTTCCAAACCCAGGCTATTCGGCAA |
| TAT GTG ATG GAC AGA GCA CAA ATG GAC CTC TCT GGA AGA GGG AAT CCC ATC AAA GTC AGC 864 | GGAAGTTAGTTTTTAATCTCTCCACCTTCCAAAGAGTGCTAAGCATTAGCTTTAATTAA |
| ĀRĞ VAL GLY SER ALA MET VAL ASN ALA LYS ASP ASP GLU GLY VAL LEU VAL GLY SER TRP CGT GTG GGG TCT GCA ATG GTG AAT GCC AAA GAT GAC GAA GGT GTC CTC GTT GGA TCC TGG | AACAGTCATCATTATCATCACAAATGGCTACATCTCCAAATATCAGTGGGCTCTCTTACCAGGGAGATTTGCTCAA <u>TAC</u> |
| 280 ASP ASN ILE TYR ALA TYR GLY VAL PRO PRO SER ALA TRP THR GLY SER VAL ASP ILE LEU GAC AAT ATC TAT GCC TAT GCC GTC CCC CCA TGG GCC TGG ACT GGA AGC GTT GAC ATT CTA | 2650 CTGGCCTCATTTAAAACAAGACTTCAGATTCCCCACTCAGCCTTTTGGGAATAATAGCACATGATTTGGGCTCTAGAATT 2710 |
| 300 Ley glu tyr arg ser ser glu asn proval arg tyr gly gln cys trp val phe ala gly | CCAGTCCCCTTTCTCGGGGTCAGGTTCTACCCTCCATGTGAGAATATTTTTCCCAGGACTAGAGCACAACATAATTTTTA |
| TTG GAA TAC COGG AGC TCT GAG AAT CCA GTC COGG TAT GGC CAA TGC TGG GTT TTT GCT GGT 1044 | TTTTTGGCAAAGCCAGAAAAAGATCTTTCATTTTGCACCTGCAGCCAAGCAAATGCCTGCC |
| VÃL PHE ASN THR PHE LEU ARG CYS LEU GLY ILE PRO ALA ARG ILE VAL THR ASN TYR PHE GTC TTT AAC ACA TTT TTA CGA TGC CTT GGA ATA CCA GCA AGA ATT GTT ACC AAT TAT TTC 1104 | 2870 TTAGAAGAGGTGGCCCCATATTAACAAATTGCATTTGTGGGAAACTTAACCACCTACAAGGAGATAAGAAGCAGGTGCA 2950 |
| SER ALA HIS ASP ANA ASP ALA ASN LEU GUN MET ASP ILE PHE LEU GUU GUN ASP GLY ASN TET GEC CAT GAT MAT GAT GEC MAT FIG CAA ATG GAC ATC TIC CTG GAA GAA GAT GEG AAC | ACACTCAAGTCTATTGAATAATGTAGTTTTGTGATGCATTTTATAGAATGTGTCACACTGTGGCCTGATCAGCAGGAGCC 3030 |
| 350 YAL ASM SER LYS LEV THE LYS ASP SER VAL THE ASM TYR HIS CYS THE ASM GLU ALA THE | AATATCCCTTACTTTAACCCTTTCTGGGATGCAATACTAGGAAGTAAAGTGAAGAATTTATCTCTTTAGTTAG |
| GIG AAT TOO AMA CTO AGO GAT TOA GTG TGG AAC TAC CAC TGC TGG AAT GAA GOA TGG 1224 | ATTTCACCCATCTCTCAGGAATCATCTCCTTTGCAGAATGATGCAGGTTCAGGTCCCCTTTCAGAGATATAATAAGQCCA |
| MET THR ANG PRO ASP LEU PRO VAL GLY PHE GLY GLY TRP GLN ALA VAL ASP SER THR PRO ATG ACA AGG CCT GAC CTT CCT GTT GGA TTT GGA GGC TGG CAA GCT GTG GAC AGC ACC CCC 1284 | ACAAGTTGAAGAAGCTGGCGGATCTAGTGACCAGATATATAGAAGGACTGCAGCCACTGATTCTCTCTTGTCCTTCACAT 3270 |
| GUN GLU ASN SER ASP GLY MET TYR ARG CYS GLY FRO ALA SER VAL GLN ALA ILE LYS HIS CAG GAA AAT AGC GAT GGC ATG TAT COG TGT GGC CCC GCC TCG GTT CAA GCC ATC AAG CAC | CACCATTTTGAGACCTCAGCTTGGCACTCAGGTGCTGAAGGGTAATATGGACTCAGCCTTGCAAATAGCCAGTGCTAGTT 5350 |
| 420 GLY HIS VAL CYS PHE GLN PHE ASP ALA PRO PHE VAL PHE ALA GUI VAL ASN SER ASP LEU | CTGACCCAACCACAGAGGATGCTGACATCATTTGTATTATGTTCCAAGGCTACTACAGAGAAGGCTGCCTGC |
| GEC CAT GTC TEC TTC CAA TTT GAT GCA CCT TTT GTT TTT GCA GAG GTC AAC AGC GAC CTC 1404 | |
| ILE THR ALA LYS LYS ASP OLY THR HIS VAL VAL OLU ASN VAL ASP ALA THR HIS ATT TAC ATT ACA OCT AAG AAA GAT GOC ACT CAT GTG GTG GAA AAT GTG GAT GCC ACC CAC 1464 | 2510 AMAAATGAACAGTGATAAGTATAACTGGGGGGCAAAATCAGAATGGAATGCTACTGGTCTATATAACCACATTTCTGAGQCTT 3590 |
| THE GEV LAS HEN THE ARE THE LAS GUN THE GEV GEV ASP GUY MET MET ASP THE THE ASP | TGAGACTGTTCCTGAGCCTTCAGCACTAACCTATGAGGGTGAGCTGGTCCCCTCTATATATA |
| 480 The Lys Phe Gun Gui Guy Gun Gui Gui Gui Arg Leu Ala Leu Gui Thr an a Leu Net | TANGTANTCTCACAGCATTTGCCAAGTCTCCCAATATCCAATTTTAAAATGAAATGCATTTTGCTAGACAGTTAMACTGG 3750 |
| ALL TAC ANA TTC CAA GAA GGT CAA GAA GAA GAA GAĞ ÀĞĂ TTĞ GCC CTĂ GAĂ ĂCT GCC CTĞ ÂTĞ 500 | СТТААСТТАБТАТТАТТАТТААТТАСААТБТААТАБААВСТТААААБТТАААСТБАТТАТАААА |
| TYR QLY ALA LYS LYS PRO LEU ASN THR QLU QLY VAL MET LYS SER ARG SER ASN VAL ASP TAC GGA GCT AAA AAG CCC CTC AAC ACA GAA GGT GTC ATG AAA TCA AGG TCC AAC GTT GAC 1644 | 300 44444444444444444444444444444444444 |

FIG. 3. Nucleotide sequence of human placental FXIIIa cDNA. The nucleotide sequence was determined by analysis of the overlapping clones shown in Fig. 1. The deduced amino acid sequence is displayed above the DNA sequence. Numbering of amino acids is relative to the amino-terminal serine (position 1). The solid arrow indicates the thrombin-activation cleavage site of FXIIIa. Potential attachment sites for carbohydrates are shown by solid diamonds. A putative ribosome binding site 41 bp upstream of the initiation codon AUG is marked by an open circle, and the poly(A) addition site 15 bp upstream of the poly(A) sequence is marked by a closed circle. Positions of the 20-mer and 66-mer oligonucleotides are underlined.

bridge, with four of the cysteines being in the free SH form (29). The fact that most cysteine residues are in a reduced form can also be taken as an indication for the intracellular biological role of FXIIIa.

Blot-hybridization analysis of $poly(A)^+$ RNA from placental tissue showed a mRNA for FXIIIa that is about 4000 bases in length. This is in good agreement with the length of the cloned cDNA sequence of 3905 bases. Thus, the FXIIIa mRNA carries 2196 bases of protein coding sequence, 1625 bases of 3' noncoding sequence, and at least 84 bases of 5' noncoding sequence. The unusually long 3' noncoding sequence carries several open reading frames. Three of them, starting with a methionine, have potential coding capacities for >50 amino acids. The longest could code for a peptide of 67 amino acids (nucleotides 2691–2891), and the others, for peptides of 55 (nucleotides 3348–3512) and 54 (nucleotides 3317–3478) residues, respectively. Whether these open reading frames have any biological significance remains to be shown.

Comparison of blot-hybridization analysis of $poly(A)^+$ RNA from mature placenta and from liver showed that FXIIIa mRNA concentration is at least 20 times higher in placenta compared to liver. Up to now the liver was considered to be the major place for FXIIIa synthesis. The normal concentration of FXIIIa in plasma is about 10-20 mg/liter, while about 30 mg are extracted from a mature placenta (36). The concentrations of FXIIIa during placental development have not yet been reported in the literature, but FXIIIa might be specifically induced at the late stage of placental development where stabilization of blood clots is of high importance. In that respect it is noteworthy that the other examples of secreted proteins that do not carry a leader peptide are induced by hormones: ovalbumin and lipocortin (37, 38). In both cases the NH₂-terminal amino acid of the mature protein molecules (glycine and alanine, respectively) is acetylated.

The cloning of FXIIIa cDNA now enables questions to be answered regarding the glycosylation pattern and the mechanism of secretion by introducing the coding sequence into different host cells.

We thank Prof. H. G. Schwick for initiation of the project, Drs. H. Karges and H. Bohn for helpful discussion, Drs. K. J. Abel and U. Hein for synthesis of oligonucleotides, and T. Rein, T. Schwarz, and M. Cieplik for excellent technical assistance.

- Chung, S. J. (1975) in *Isoenzymes 1*, ed. Boyer, H. (Academic, New York), pp. 259–273.
- Sixma, J. J., Van den Berg, A., Schiphorst, M., Geuze, H. J. & McDonagh, J. (1984) Thromb. Haemostasis Gen. Inf. 51, 388-391.
- 3. Bohn, H. (1978) Mol. Cell. Biochem. 20, 67-75.
- 4. Curtis, C. G. & Lorand, L. (1976) Methods Enzymol. 45, 177-191.
- 5. Folk, J. E. & Finlayson, J. S. (1977) Adv. Protein Chem. 31, 1-133.

- 6. Sakata, Y. & Aoki, N. (1980) J. Clin. Invest. 65, 290-297.
- 7. Mosher, D. F. (1975) J. Biol. Chem. 250, 6614-6621.
- Mosher, D. F., Schad, P. E. & Kleinmann, H. K. (1979) J. Clin. Invest. 64, 781-787.
- 9. Bohn, H., Becker, W. & Trobisch, H. (1973) Blut 26, 303-311.
- Schimpf, K. (1985) Plasmatherapie, eds. Lutz, H. & Rother, K. (Medizinische Verlagsgesellschaft Marburg, F.R.G.), pp. 237-254.
- 11. Castle, S. L. & Board, P. G. (1983) Clin. Chim. Acta 133, 141-151.
- Castle, S. L., Board, P. G. & Anderson, R. A. M. (1981) Br. J. Haematol. 48, 337-342.
- 13. Board, P. G. (1979) Am. J. Hum. Genet. 31, 116-124.
- 14. Board, P. G. (1980) Am. J. Hum. Genet. 32, 348-353.
- Board, P. G., Reid, M. & Serjeantson, S. (1984) Hum. Genet. 67, 406-408.
- 16. Takagi, T. & Doolittle, R. F. (1974) Biochemistry 13, 750-756.
- Lorand, L., Credo, R. B. & Janus, T. J. (1981) Methods Enzymol. 80, 333-341.
 Horachen L. L. Cochen D. D. & Kingerten J. B. (1077)
- Holbrook, J. J., Cooke, R. D. & Kingston, I. B. (1973) Biochem. J. 135, 901-903.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* 18, 5294-5299.
- Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
- 21. Gubler, U. & Hoffman, B. J. (1983) Gene 25, 263-269.
- 22. Lathe, R. (1985) J. Mol. Biol. 183, 1-12.
- Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colten, H. R. (1982) Proc. Natl. Acad. Sci. USA 79, 5661-5665.
- 24. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 25. Smith, H. O. & Birnstiel, M. L. (1976) Nucleic Acids Res. 63, 2387-2398.
- 26. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
- 27. Maxam, A. M. & Gilbert, W. (1977) Proc. Natl. Acad. Sci.
- USA 74, 560-564.
 28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Takahashi, N., Takahashi, Y. & Putnam, F. W. (1986) Proc. Natl. Acad. Sci. USA 83, 8019-8023.
- Chan, Y.-L., Gutell, R., Noller, H. F. & Wool, I. G. (1984) J. Biol. Chem. 259, 224-230.
- 31. Nakamura, S., Iwanaga, S. & Suzuki, T. (1974) Biochem. Biophys. Res. Commun. 58, 250-256.
- Boissel, J. P., Kasper, T. J., Shah, S. C., Malone, J. & Bunn, H. F. (1985) Proc. Natl. Acad. Sci. USA 82, 8448-8452.
- 33. Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.
- 34. Kozak, M. (1986) Cell 44, 283-292.
- 35. Baralle, F. E. & Brownlee, G. G. (1978) Nature (London) 274, 84-87.
- 36. Bohn, H. & Schwick, H. G. (1971) Arzneim.-Forsch. 21, 1432-1439.
- Palmiter, R. D., Gagnon, J. & Walsh, K. A. (1978) Proc. Natl. Acad. Sci. USA 75, 94–98.
- Wallner, B., Mattaliano, R. J., Hession, C., Cate, R. L., Tizard, R., Sinclair, L. K., Foeller, C., Chow, E. P., Browning, J. L., Ramachandran, K. L. & Pepinsky, B. (1986) Nature (London) 320, 77-81.