- Garibaldi, J. A., & Neilands, J. B. (1955) J. Am. Chem. Soc. 77, 2429-2430.
- Glickson, J. D., Pitner, T. P., Webb, J., & Gams, R. A. (1975) J. Am. Chem. Soc. 97, 1679-1683.
- Hoffer, P. B., Huberty, J. P., & Khayam-Bashi, H. (1977) J. Nucl. Med. 18, 713-717.
- Johnston, D. O., & Cottingham, A. B. (1969) Biochim. Biophys. Acta 177, 113-117.
- Lascelles, J. (1964) Tetrapyrrole Biosynthesis and Its Regulation, p 51, W. A. Benjamin, New York,
- Llinas, M., Klein, M. P., & Neilands, J. B. (1970) J. Mol. Biol. 52, 399-414.

- Monzyk, B., & Crumbliss, A. L. (1983) J. Inorg. Chem. 19, 19-39.
- Moore, R. E., & Emery, T. (1976) Biochemistry 15, 2719-2723.
- Raymond, K. N., Muller, G., & Matzanke, B. F. (1984) Top. Curr. Chem. 123, 49-101.
- Stookey, L. L. (1970) Anal. Chem. 42, 779-781.
- Topham, R. C., Walker, M. C., Calisch, M. P., & Williams, R. W. (1982) *Biochemistry 21*, 4529-4535.
- Tufano, T. P., & Raymond, K. N. (1981) J. Am. Chem. Soc. 103, 6617-6624.

Amino Acid Sequence of the b Subunit of Human Factor XIII, a Protein Composed of Ten Repetitive Segments[†]

Akitada Ichinose, Brad A. McMullen, Kazuo Fujikawa, and Earl W. Davie*

Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received February 28, 1986; Revised Manuscript Received March 28, 1986

ABSTRACT: Factor XIII is a plasma protein that participates in the final stages of blood coagulation. The complete amino acid sequence of the b subunit of human factor XIII was determined by a combination of cDNA cloning and amino acid sequence analysis. A λ gt11 cDNA library prepared from human liver mRNA was screened with an affinity-purified antibody against the b subunit of human factor XIII. Nine positive clones were isolated from 2 × 10⁶ phage and plaque-purified. The largest cDNA insert was sequenced and shown to contain 2180 base pairs coding for a portion of the leader sequence (19 amino acids), the mature protein (641 amino acids), a stop codon (TGA), a 3' noncoding region (187 nucleotides), and a poly(A) tail. When the b subunit of human factor XIII was digested with cyanogen bromide, nine peptides were isolated by gel filtration and reverse-phase high-performance liquid chromatography. Amino acid sequence analyses of these peptides were performed with an automated sequenator, and 299 amino acid residues were identified. These amino acid sequences were in complete agreement with the amino acid sequence predicted from the cDNA. The b subunit of factor XIII contained 10 repetitive homologous segments, each composed of about 60 amino acids and 4 half-cystine residues. Each of these repeated segments is a member of a family of repeats present in human β_2 -glycoprotein I, complement factor B, and haptoglobin α^1 chain. Three potential Asn-linked carbohydrate attachment sites were also identified in the b subunit of factor XIII.

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 polymerization of fibrin monomers through the formation of intermolecular ϵ -(γ -glutamyl)lysine bonds. This reaction occurs in the presence of Ca²⁺ [see Folk and Finlayson (1977) and Lorand et al. (1980) for reviews].

The cross-linking reactions catalyzed by factor XIII_a lead to dimerization of the γ chains of fibrin (γ -dimerization) (Chen & Doolittle, 1970) followed by 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 & Lorand, 1973; Shen et al., 1975) and an increase in its resistance to proteolytic degradation by plasmin (Lorand & Jacobsen, 1962; Gaffney

& Whitaker, 1979). Factor XIII_a also catalyzes a cross-linking between the α chain of fibrin with fibronectin (Mosher, 1975; Tamaki & Aoki, 1981) and the α chain of fibrin with α_2 -plasmin inhibitor (Sakata & Aoki, 1980; Tamaki & Aoki, 1981). The cross-linking of collagen and fibronectin is also catalyzed by factor XIII_a (Mosher & Schad, 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 or α_2 -plasmin inhibitor result in "delayed bleeding", while primary hemostasis in individuals with these traits is normal (Folk & Finlayson, 1977; Aoki et al., 1979; Lorand et al., 1980). This suggests that these proteins also play an important role in the protection of the fibrin clot from digestion by plasmin (Sakata & Aoki, 1980, 1982).

[†]This study was supported by Research Grant HL 16919 from the National Institutes of Health.

4634 BIOCHEMISTRY ICHINOSE ET AL.

Factor XIII (M. 300 000) circulates in blood 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 contains the catalytic site(s) of the enzyme (Chung et al., 1974). The precise function of the b subunit is not known, but it is thought to protect or stabilize the a subunit (Cooke, 1974; Folk & Finlayson, 1977; Lorand et al., 1980) or to regulate the activation of the zymogen (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. The sequence of this peptide has been determined (Takagi & Doolittle, 1974; Nakamura et al., 1975) as well as the amino acid residues around the catalytic site(s) of the a subunit (Holbrook et al., 1973). At present, little is known about the amino acid sequence of the b subunit of factor XIII other than the amino-terminal sequence of Glu-Glx-Lys-Pro (Takagi & Doolittle, 1974).

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

MATERIALS AND METHODS

Materials. 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, adenosine 5'-triphosphate (ATP), deoxynucleotides, dideoxynucleotide triphosphates, M13mp10, M13mp11, M13mp18, and M13mp19 were supplied by Bethesda Research Laboratories. Na¹²⁵I was obtained from New England Nuclear, and $[\alpha$ -35S]dATP α S was purchased from Amersham. Normal human plasma was obtained from Blood Services Pacific Northwest Region, Portland, OR.

Polyclonal antibody against the b subunit of human factor XIII was raised in rabbits, and the immunoglobulin fraction was purified by ammonium sulfate fractionation, chromatography on a DEAE-Sephadex column and an immobilized antigen-Sepharose column (Canfield & Kisiel, 1982).

Preparation of the b Subunit of Factor XIII and the Amino Acid Sequence of the Cyanogen Bromide Fragments. Factor XIII was purified from human plasma according to the method of Curtis and Lorand (1976). After incubation with thrombin, the b subunit of factor XIII was separated from the a subunit by gel filtration (Chung et al., 1974). The purified b subunit (10 mg) was then S-pyridylethylated (Friedman et al., 1970) and digested with cyanogen bromide. The resulting fragments were separated by gel filtration on a Sephadex G-50 superfine column using 5% HCOOH. The peptides obtained in the subfractions from the gel filtration column were further purified on a Waters HPLC system using 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. The eluant was monitored by absorbance at 214 nm. Sequence

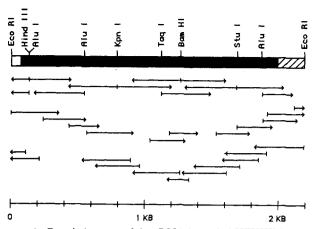


FIGURE 1: Restriction map of the cDNA insert in λ HFXIIIb2.2 that codes for the b subunit of human factor XIII. The open, solid, and hatched bars represent a portion of the leader peptide, the mature protein, and the 3' noncoding region, respectively. The extent of DNA sequencing is shown by the length 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.

analysis of the intact protein and 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).

Screening of the cDNA Library. The λ gt11 cDNA library prepared human liver mRNA (Kwok et al., 1985) was kindly provided by Dr. Savio L. C. Woo. Approximately 2×10^6 phage were screened by the method of Young and Davis (1983) as modified by Foster and Davie (1984). The purified antibody was labeled with Na¹²⁵I to a specific activity of 4×10^6 cpm/ μ g and was used to screen filters containing phage plated at a density of 1.5×10^5 plaques per 150-mm plate. Positive clones were isolated and plaque-purified.

DNA Sequence Analysis. 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). The cDNA insert was isolated by digestion of the phage DNA with EcoRI endonuclease and subcloned into plasmid pUC9 (Vieira & Messing, 1982). Appropriate restriction fragments from the insert were then subcloned into M13mp10 or M13mp18 for sequencing by the dideoxy method (Sanger et al., 1977) using $[\alpha^{-35}S]dATP\alpha S$ and buffer gradient gels (Biggin et al., 1983). Controlled digestions with nuclease BAL-31 were performed to generate suitable fragments, which provided overlapping sequences with the restriction fragments (Poncz et al., 1982). All sequence determinations were performed on both strands of DNA at least 3 times. DNA sequences were analyzed with the computer program of Textco (West Lebanon, NH) using an Apple Macintosh computer (Gross, 1986).

RESULTS AND DISCUSSION

A λ gt11 expression library containing cDNAs prepared from human liver mRNA was screened for the b subunit of human factor XIII by employing a ¹²⁵I-labeled affinity-purified rabbit antibody. Nine positive clones were isolated by screening 2 × 10⁶ phage, and each was plaque-purified. The clone with the largest cDNA insert (approximately 2.2 kilobases) was named λ HFXIIIb2.2 and was selected for further study. The cDNA insert released by treatment of the phage DNA from λ HFXIIIb2.2 with *Eco*RI endonuclease was subcloned into plasmid pUC9, and the nucleotide sequence

¹ Abbreviations: ATP, adenosine 5'-triphosphate; HPLC, high-performance liquid chromatography; PTH, phenylthiohydantoin; SDS, sodium dodecyl sulfate.

-19 R L K N L T F I I I L I I S G E L Y A E E K P C G G AGG TIG AAA AAC CTG ACT TIT ATC ATC ATA TTG ATA ATC TCA GGA GAA CTC TAT GCA GAA GAG AAA CCC TGT GGT TIT CCT CAT GTG ENGRIAQYYYTFKSFYFPM SIDKKLSFFCL 11 89 GAA AAT GGA AGA ATT GCC CAA TAT TAC TAT ACT TIT AAA AGC TIT TAC TIT CCA ATG AGC ATA GAC AAA AAA TIG TCA TIT TTC TGC TTG G R Q E E Q T T C T T E G W S P E P R C F 179 GCT GGT TAT ACC ACT GAA AGT GGA AGA CAA GAA GAG CAA ACC ACG TGT ACA ACA GAA GGC TGG TCT CCA GAG CCA AGG TGC TTC AAA AAA C T K P D L S N G Y I S D V K L L Y K I Q E N M H Y G C A S 71 269 TGC ACT AAG CCT GAC CTG AGT AAT GGT TAC ATC TCT GAT GTA AAG TTA TTG TAT AAA ATT CAA GAG AAC ATG CAT TAT GGT TGC GCT TCA 101 T T G G K D E E V V O C L S D G W S S O P T C R K F H GGG TAC AAA ACC ACT GGA GGG AAG GAT GAA GAA GTG GTT CAA TGT CTC TCT GAT GGA TGG TCT TCT CAA CCA ACC TGT AGG AAA GAA CAT 359 ETCLAPELYNGNYSTTQKTFKVKDKVQYEC 131 GAA ACG TGT TTG GCT CCT GAA TTA TAT AAT GGA AAT TAT TCC ACA ACA CAG AAA ACA TTC AAA GTG AAG GAC AAA GTA CAA TAC GAA TGT 449 A G G K K T E E V E C L T Y G W S L T P K C T K 161 539 GCT ACT GGC TAC TAC ACA GCT GGA GGA AAG ACA GAG GAG GAG GTA GAA TGT CTC ACA TAC GGA TGG TCT CTC ACA CCA AAA TGT ACC AAA 191 K C S S L R L I E N G Y F H P V K Q T Y E E G D V V Q F TTA AAG TGC TCT TCT TTA AGA TTA ATT GAA AAT GGT TAT TTT CAT CCT GTA AAG CAA ACC TAT GAA GGA GGA GGT GTC GTT CAG TTT TTC 629 221 C H E N Y Y L S G S D L I Q C Y N F G W Y P E S P V C E G R TGT CAT GAA AAT TAT TAT CTA AGT GGA TCT GAT TTA ATT CAA TGC TAT AAC TTT GGT TGG TAC CCA GAA TCT CCT GTA TGC GAA GGA AGA 719 251 C P P P P L P I N S K I Q T H S T T Y R H G E I V H I 809 E C E L N F E I H G S A E I R C E D G K W T E P P K C I E G 281 GAA TGT GAA CTT AAT TTT GAG ATC CAT GGG TCA GCA GAA ATA CGT TGT GAA GAT GGA AAA TGG ACA GAA CCT CCA AAA TGC ATT GAA GGA Q E K V A C E E P P F I E N G A A N L H S K I Y Y N G D K V 311 CAG GAG AAG GTA GCC TGT GAG GAA CCA CCC TTC ATT GAA AAT GGT GCA GCA AAT TTA CAC TCT AAG ATT TAT TAC AAT GGG GAT AAA GTG 989 Y A C K S G Y L L H G S N E I T C N R G K W T L P P E C V 341 ACA TAT GCA TGT AAA AGC GGC TAC CTT CTC CAT GGA TCC AAT GAG ATA ACT TGT AAT CGT GGA AAA TGG ACA CTT CCT CCT GAG TGT GTT 1079 ENNENCK H P P V V M N G A V A D G I L A S Y A T G S S 371 GAA AAT AAT GAG AAT TGT AAG CAT CCT CCT GTT GTA ATG AAT GGG GCT GTT GCA GAC GGG ATA TTG GCA AGC TAT GCA ACA GGA TCC TCA 1169 401 L R G S K I S R C F O G K W S S P P V C GTG GAA TAT AGA TGC AAT GAA TAT TAC TTA CTG AGG GGA TCA AAA ATA TCT CGT TGC GAA CAA GGA AAA TGG TCA TCC CCA CCT GTT TGC 1259 431 LEPCTVN V D Y M N R N N I E M K W K Y E G K V L H G D 1349 TTG GAA CCA TGT ACT GTT AAT GTG GAT TAC ATG AAC AGA AAT AAC ATA GAA ATG AAG TGG AAA TAT GAA GGG AAA GTC TTA CAT GGA GAT 461 V C K Q G Y D L S P L T P L S E L S V Q C N R G E TTA ATA GAT ITT GTA TGT AAA CAG GGA TAT GAC ITA TCT CCA TTA ACC CCA TTG TCT GAA TTA TCT GTG CAG TGC AAC AGA GGA GGA GTG 1439 KYPLCTRKESKGMCTSPPLIKHGVI 491 AAA TAT CCT TTA TGT ACT AGA AAA GAA TCT AAA GGA ATG TGC ACA TCT CCT CCT CTT ATT AAA CAT GGA GTC ATT ATT AGT TCA ACA GTA 1529 Y E N G 5 S V E Y R C F D H H F L E G S R E A Y C L D G 521 GAC ACC TAT GAA AAT GGC TCT TCA GTA GAA TAC AGA TGT TTT GAT CAC CAT TTC CTA GAA GGA TCT AGG GAG GCC TAT TGT TTA GAT GGA 1619 C L E P C T L S F T E M E K N N L L L K W D F 551 TPPL ATG TGG ACT ACA CCA CCA TTG TGT TTA GAG CCA TGC ACA TTA TCT TTT ACT GAA ATG GAA AAG AAT AAT TTA CTT CTG AAA TGG GAT TTT 1709 581 D N R P H I L H G E Y I E F I C R G D T Y P A E L Y I T G S 1799 GAC AAT AGA CCA CAC ATT TIG CAT GGT GAA TAT ATT GAG TIT ATT TGT AGA GGA GAT ACT TAT CCA GCT GAA TTA TAT ACT ACT GGA TCT 611 I L R M Q C D R G Q L K Y P R C I P R Q S YQEP ATA CTT AGA ATG CAA TGT GAC AGA GGG CAG TTA AAA TAT CCA AGA TGT ATT CCA AGA CAA AGC ACT CTG TCT TAT CAA GAA CCC TTA AGA 1889 641 1979 ACA TAG AAATGAATGGCAGAAAGAGGGGTCATATTTCAATACATCATGAAATTCCTTATAAAATTTTGAGGAAATAAGTTAAAAACTTCGGAGGTTTTTCCTTGCTTAAATATTTTG 2099

FIGURE 2: Nucleotide sequence of the cDNA insert coding for the b subunit of human factor XIII and predicted amino acid sequence. The amino acids numbered -19 to -1 represent a portion of the leader or signal sequence, while those numbered +1 to 641 represent the amino acids present in the mature b subunit circulating in blood. The amino acid residues that are overlined were also determined by amino acid sequence analysis. (Residues 394, 536, and 537, however, were not identified by this analysis.) Residues marked with solid diamonds are potential Asn-linked glycosylation sites at an Asn-X-Ser sequence, while the residue marked with an open diamond is a potential Asn-linked glycosylation site at an Asn-X-Cys sequence. The nucleotide segments that correspond to the polyadenylation signal and the poly(A) tail are shown in boxes.

was determined by the strategy shown in Figure 1. The cDNA insert was found to be composed of 2180 base pairs coding for the entire amino acid sequence for the b subunit

of factor XIII that circulates in blood (Figure 2). The b subunit is composed to 641 amino acids starting with an amino-terminal Glu (nucleotides 56-58). The amino-terminal

4636 BIOCHEMISTRY ICHINOSE ET AL.

Table I: Amino Acid Sequences of the Intact Molecule and Cyanogen Bromide Peptides of the b Subunit of Factor XIII

-,						
XII	Ιb	EEKP-GFPHV	ENGRIAQYYY			
CNB	1	EEKPCGFPHV	ENGRIAQYYY	TF		
CNB	2	SIDKKLSFFC	LAGYTTESGR	QEEQTTCTTE		
CNB	3	HYGCASGYKT	TGGKDEEVVQ	CLSDGWSSQP	TCR	
CNB	4	NGAVADGILA	SYATGSSVEY	RCNEYYLLRG	SKISRCEQGK	WS
CNB	6	KWKYEGKVLH	GDLIDFVCKQ	GYDLSPLTPL	SELSVQCNRG	EVKYPL
CNB	7	CTSPPLIKHG	VIISSTVDTY	E-GSSVEYRC	FD-HFLEGSR	EAYCLDG
CNB	8	WTTPPLCLEP	CTLSFTE			
CNB	9	EKNNLLLKWD	FDNRPHILHG	EYIEFICRGD	TYPAELYITG	
CNB	10	QCDRGQLKYP	RCIPROSTLS	YQEP		

sequence of Glu-Glx-Lys-Pro for the b subunit of human factor XIII was established earlier by Takagi and Doolittle (1974) and extended in the present studies (see below). The carboxyl-terminal Thr (nucleotides 1979–1981) is followed by a stop codon (TAG), 187 base pairs of noncoding sequence, and a poly(A) tail of 9 base pairs. The polyadenylation or processing signal of AATAAA (Proudfoot & Brownlee, 1976) was identified 19 nucleotides upstream from the poly(A) tail.

The cDNA clone also codes for 19 amino acid residues that constitute a portion of a leader peptide. The cDNA did not, however, extend to the 5' end of the mRNA and, thus, did not include the initiation Met residue. The partial leader sequence of 19 amino acids includes a hydrophobic core and an Ala residue at position -1 and Leu at position -3. These residues are consistent with the "-1 and -3 rule" in which the -1 position is occupied predominantly by Ala in signal sequences (Perlman & Halvorson, 1983; von Heijne, 1984).

Amino acid sequence analysis of the intact protein was then carried out in a Beckman sequenator and 19 amino acids were identified (Table I). This extended the amino-terminal sequence originally reported by Takagi and Doolittle (1974). Amino acid sequence analyses were also performed on the cyanogen bromide fragments of the b subunit of human factor XIII. In these experiments, the b subunit was S-pyridylethylated and digested with cyanogen bromide. The resulting peptide fragments were separated by gel filtration (Figure 3) and further purified by reverse-phase HPLC. Nine of the ten cyanogen bromide fragments expected from the amino acid sequence obtained from the cDNA were isolated and corresponded to cyanogen bromide fragments 1-4 and 6-10 as numbered from the amino-terminal to the carboxyl-terminal end of the protein (Figure 2). Each of these fragments was then subjected to amino acid sequence analysis, and a total of 299 residues were unequivocally identified (Table I). These amino acid sequences were in complete agreement with those predicted by the cDNA and the amino-terminal sequence analysis of the intact protein. The third peak (Figure 3) contained cyanogen bromide peptides 1 and 2, which apparently were aggregated or incompletely digested with cyanogen bromide.

The b subunit of factor XIII consists of 641 amino acid residues and has the following composition: Ala₁₇, Arg₂₆, Asn₃₀, Asp₂₂, ¹/₂Cys₄₀, Gln₂₀, Glu₆₀, Gly₄₈, His₁₈, Ile₂₉, Leu₄₇, Lys₄₄, Met₉, Phe₂₀, Pro₄₁, Ser₄₅, Thr₄₄, Trp₁₀, Tyr₄₂, Val₂₉. The molecular weight of the polypeptide portion was calculated to be 69 973. The addition of 8.5% carbohydrate (Bohn, 1970; Bohn et al., 1972) gives a molecular weight of approximately

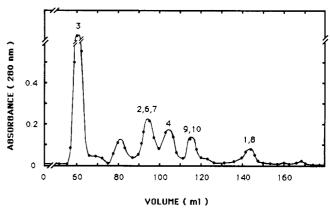


FIGURE 3: Fractionation of the cyanogen bromide fragments of the b subunit of human factor XIII by gel filtration. The b subunit (10 mg) was S-pyridylethylated, digested with cyanogen bromide, dissolved in 2 mL of 5% formic acid, and applied to a Sephadex G-50 column (1.6 \times 90 cm) previously equilibrated with 5% formic acid. The column was eluted with 200 mL of 5% formic acid at approximately 12 mL/h. Eight separate peaks were collected for further fractionation by reverse-phase HPLC as described under Materials and Methods. The numbers above each peak refer to the cyanogen bromide fragments and correspond to their position (1-10) in the intact protein.

76 500 for each of the b subunits of human factor XIII. This is in good agreement with the value of 80 000 estimated by SDS-polyacrylamide gel electrophoresis (Schwartz et al., 1973; Chung et al., 1974).

The protein sequence predicted from the cDNA includes two potential Asn-linked glycosylation sites with the sequence of Asn-Tyr-Ser and Asn-Gly-Ser starting with amino acid residues 142 and 525, respectively. In addition, a third potential carbohydrate attachment site is present in the sequence of Asn-Arg-Cys starting at residue 252. The attachment of carbohydrate chains to Asn residues in a sequence of Asn-X-Cys was first reported in bovine protein C (Stenflo & Fernlund, 1982) and later in human von Willebrand factor (Titani et al., 1986). The differential glycosylation of these Asn residues may be in part responsible for the microheterogeneity of the b subunit of human factor XIII (Curtis et al., 1974; Board, 1980). The 20 potential disulfide bonds identified by the cDNA coding for the b subunit of factor XIII are in reasonably good agreement with the 16-17 disulfide bonds reported earlier (Chung et al., 1974). The b subunit of factor XIII is free of -SH groups (Chung et al., 1974).

The amino acid sequence of the b subunit of factor XIII shows evidence of considerable internal gene duplication, involving 10 repetitive sequences of approximately 60 amino acids. These repetitive sequences were subclassified into four distinct groups (Figure 4). The identities within groups 1, 2, 3, and 4 were 34-42%, 34-42%, 38%, and 41%, respectively. Although the identities among the four groups of repeats are obviously less than each internal identity, high alignment scores employing the Dayhoff program (Dayhoff et al., 1983) indicate that these four groups have diverged from one prototype. For instance, the alignment scores were 7.7 between repeats 3 and 4, 12.1 between repeats 5 and 7, and 3.4 between repeats 6 and 8. The ten repeats (1-10) are aligned consecutively throughout 98% of the molecule and include amino acid residues 1-626. A short sequence of 15 amino acids (residues 627-641) following the last repeat at the carboxyl end of the molecule was not homologous with repeats 1-10.

Computer-assisted analysis using a Dayhoff program (Dayhoff et al., 1983) revealed that the repeated sequences in the b subunit of factor XIII are members of a family of repeats that are very similar to three repeated segments present in the Ba chain of factor B of human complement (Mole et

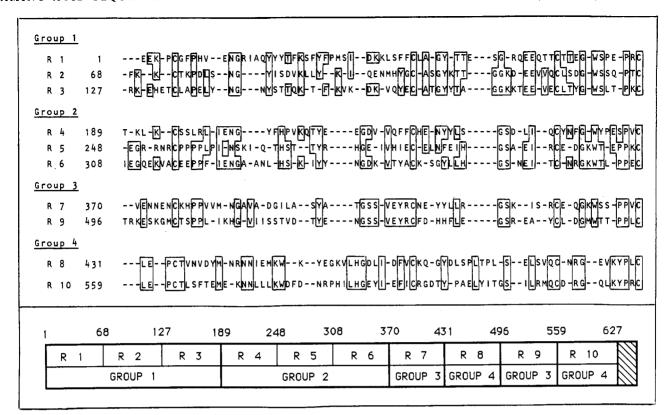


FIGURE 4: Alignment of the ten amino acid repeats present in the b subunit of factor XIII. The repeats were subclassified into four groups, including repeats 1, 2, 3; 4, 5, 6; 7, 9; and 8, 10. The numbers prior to each repeat identify the residue number of the first amino acid in each repeat. Two or more identical residues at the same position in each group of repeats were boxed. Gaps were inserted to obtain maximal alignment between the ten repeated segments.

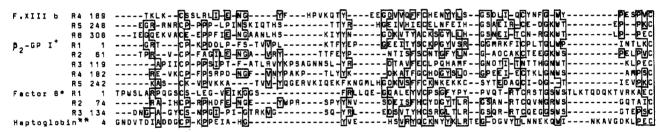


FIGURE 5: Comparison of the amino acid sequence in repeats 4, 5, and 6 of the b subunit of factor XIII with the five repeats in human β_2 -glycoprotein I, the three repeats in the Ba chain of factor B of human complement, and human haptoglobin α^1 chain. Four or more identical residues in the same position were boxed. Gaps were inserted to obtain maximum alignment. The numbers prior to each repeat identify the residue number of the first amino acid in each repeat. *From Mole et al., 1984; *from Lozier et al., 1984; **from Kurosky et al., 1980.

al., 1984), five repeated segments in human β_2 -glycoprotein I (Lozier et al., 1984) [a protein identical with activated protein C binding protein (Canfield & Kisiel, 1982)], and human haptoglobin α^1 chain (Kurosky et al., 1980). This sequence homology is shown in Figure 5. The fourth segment of β_2 -glycoprotein I, the second segment of complement factor Ba, and human haptoglobin α^1 chain show the highest alignment scores (11, 8.0, and 4.2, respectively) with repeat 6 of the b subunit of human factor XIII. These data indicate that these four proteins share a common ancestry and have probably resulted from exon shuffling during evolution. Accordingly, it will be of interest to determine the location of the introns in the gene coding for the b subunit for factor XIII and to determine whether introns are located between the repeated segments.

The location of five disulfide bonds in the five homologous segments in β_2 -glycoprotein I has been established and shown to occur between the first and third and the second and fourth Cys residues in each segment (Lozier et al., 1984). Thus, it seems probable that a similar pairing occurs with the disulfide bonds in the ten repeats in the b subunit of factor XIII.

ACKNOWLEDGMENTS

We thank Dr. Savio L. C. Woo for kindly providing the Agt11 library and Drs. Dominic Chung, Steven Leytus, Donald Foster, Torben Petersen, Benito Que, and Barbara Schach for their help and advice. We also thank Lee Hendrickson for his assistance in the purification of factor XIII.

Registry No. Factor XIII, 9013-56-3; blood coagulation factor XIII (human b subunit reduced), 103067-91-0.

REFERENCES

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

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

Board, P. G. (1980) Am. J. Hum. Genet. 32, 348-353.

Bohn, H. (1970) Thromb. Diath. Haemorrh. 23, 455-468.
Bohn, H., Haupt, H., & Kranz, T. (1972) Blut 25, 235-248.
Canfield, W. M., & Kisiel, W. (1982) J. Clin. Invest. 70, 1260-1272.

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

4638 BIOCHEMISTRY ICHINOSE ET AL.

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

- Chung, S. I., Lewis, M. S., & Folk, J. E. (1974) J. Biol. Chem. 249, 940-950.
- Cooke, R. D. (1974) Biochem. J. 141, 683-691.
- Curtis, C. G., & Lorand, L. (1976) Methods Enzymol. 45, 177-191.
- Curtis, C. G., Brown, K. L., Credo, R. B., Domanik, R. A., Gray, A. Stenberg, P., & Lorand, L. (1974) *Biochemistry* 13, 3774-3780.
- Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) Methods Enzymol. 91, 524-545.
- 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., & 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.
 Gross, R. H. (1986) Nucleic Acids Res. 14, 591-596.
- 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.
- Kurosky, A., Barnett, D., Lee, T.-H., Touchstone, B., Hay,
 R. E., Arnott, M. S., Brown, B. H., & Fieth, W. M. (1980)
 Proc. Natl. Acad. Sci. U.S.A. 77, 3388-3392.
- Kwok, S. C. M., Ledley, F. D., DiLella, A. G., Robson, K.
 J. H., & Woo, S. L. C. (1985) Biochemistry 24, 556-561.
 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.

- Lozier, J., Takahashi, N., & Putnam, F. W. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 3640-3644.
- Mockros, L. F., Roberts, W. W., & Lorand, L. (1974) Biophys. Chem. 2, 164-169.
- Mole, J. E., Anderson, J. K., Davison, E. A., & Woods, D. E. (1984) J. Biol. Chem. 259, 3407-3412.
- Mosher, D. F. (1975) J. Biol. Chem. 250, 6614-6621.
- Mosher, D. F., & Schad, P. E. (1979) J. Clin. Invest. 64, 781-787.
- Nakamura, S., Iwanaga, S., & Suzuki, T. (1975) J. Biochem. (Tokyo) 78, 1247-1266.
- 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. (1973) J. Clin. Invest. 71, 1336-1341. Shen, L. L., Hermans, J., McDonagh, J., McDonagh, R. P.,
- Shen, L. L., Hermans, J., McDonagh, J., McDonagh, R. P. & Carr, M. (1975) *Thromb. Res.* 6, 255–265.
- 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.
- Stenflo, J., & Fernlund, P. (1982) J. Biol. Chem. 257, 12180-12185.
- 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.
- Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., & Fujikawa, K. (1986) *Biochemistry* 25, 3171-3184.
- Vieira, J., & Messing, J. (1982) Gene 19, 259-268.
- von Heijne, G. (1984) J. Mol. Biol. 173, 243-251.
- Young, R. A., & Davis, R. W. (1983) Science (Washington, D.C.) 222, 778-782.