Structure of Transglutaminases

Akitada Ichinose, Ralph E. Bottnen, and Earl W. Davie

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

Transglutaminases are enzymes that catalyze a Ca\(^{2+}\)-dependent acyl transfer reaction, resulting in the formation of an \(\epsilon\)-(\(\gamma\)-glutamyl)lysine bond between the \(\gamma\)-carboxyl group of a glutamine residue in one polypeptide chain and the \(\epsilon\)-amino group of a lysine residue in a second polypeptide chain (for reviews, see Refs. 1 and 2). Transglutaminases are widely distributed in various organs, tissues, and body fluids. These include liver tissue transglutaminase, hair follicle transglutaminase, epidermal transglutaminase, prostate transglutaminase, and coagulation factor XIII from blood (3, 4). They are distinguishable from each other to a large extent by their physical properties and distribution in the body.

Factor XIII (fibrin-stabilizing factor or fibrinoligase) is one of the best characterized transglutaminases, and its physiological role is well established. It is a plasma protein that circulates in blood as a tetramer of \(a_2b_2\) (\(M_r = 320,000\)) and consists of two catalytic \(a\) subunits (\(M_r \approx 75,000\) each) and two noncatalytic \(b\) subunits (\(M_r \approx 80,000\) each) (5, 6). The \(b\) subunit is thought to stabilize the \(a\) subunit. Factor XIII also exists as a dimer of only a subunits in platelets, placenta, uterus, prostate, macrophages, and other tissues and cells. Tissue transglutaminase, however, differs in its sequence from the plasma enzyme and is present as a monomer of \(M_r 75,000-80,000\) in liver and erythrocytes. Epidermal transglutaminase also occurs as a monomer (\(M_r 50,000-55,000\)), while hair follicle transglutaminase exists as a dimer composed of two identical subunits (\(M_r 27,000\)).

In contrast to many transglutaminases, factor XIII is a proenzyme that is converted in blood to the active enzyme by thrombin in the presence of fibrin (7). The active enzyme (factor XIIIa) is generated during the final stages of the blood coagulation cascade by the release of an activation peptide (\(M_r 4,000\)) from the NH\(_2\)-terminus of each of the \(a\) subunits (5, 8). In the presence of calcium ions, the tetramer (\(a_2b_2\)) dissociates into an active dimer (\(a_2'\)) and two \(b\) subunits (6, 9, 10). Calcium ions bind to the \(a\) subunits and unmask the active site(s) in the enzyme (11, 12).

The transglutamination reaction catalyzed by factor XIIIa leads to the cross-linking of a number of proteins in plasma. These include the dimerization of the \(\gamma\) chains of two different fibrin molecules followed by the polymerization of the \(\alpha\) chains of fibrin (5, 13). These reactions are critical to the blood coagulation cascade and result in the formation of a tough insoluble fibrin clot. A second important reaction catalyzed by factor XIIIa is the cross-linking of \(\alpha\)-plasmin inhibitor to the \(\alpha\) chains of fibrin (14, 15). This reaction plays a significant role in the regulation of fibrinolysis. A third important reaction catalyzed by factor XIIIa is the cross-linking of fibronectin to the \(\alpha\) chains of fibrin (16, 17) and to collagen (18), a reaction closely associated with wound healing (1, 2, 10). Accordingly, a deficiency of factor XIII can result in a lifelong bleeding tendency, defective wound healing, and habitual abortion (19).

Additional functions of transglutaminases may include irreversible membrane stiffening of the erythrocyte (20), receptor-mediated endocytosis and regulation of cellular growth and differentiation by tissue transglutaminase (21-24), formation of an insoluble "cornified" envelope in epidermal keratinocyte (25) by epidermal transglutaminase, and formation of a postcervical vaginal plug by prostate transglutaminase in rodents (26).

The primary structures of plasma factor XIII and liver tissue transglutaminase have been established by a combination of cDNA cloning and amino acid sequence analysis (27-31). These studies have made it possible to examine and compare the structures of these proteins and their genes.

The \(a\) subunit of human factor XIII consists of 731 amino acid residues (27, 30) with a molecular weight of 83,150. The \(a\) subunit of plasma factor XIII and the enzyme in several tissues seem to be identical, since the amino acid sequence analyses of corresponding regions of the purified enzymes from plasma (27), platelet (8), and placenta (27, 29, 30) are identical.

The mature protein starts with acetyl-Ser-Glu-Thr- (8), whereas the COOH terminus of the purified \(a\) subunit (placenta) has been reported to be heterogeneous in length (29).

Although, there are 9 Cys residues in the molecule, it is unlikely that any are involved in disulfide bonds.

Several differences have been found when the amino acid sequence deduced from the DNA sequence of the gene for the \(a\) subunit of human factor XIII (32) or from cDNAs (27, 30) was compared with that isolated from plasma (27) or placenta (29) and examined by amino acid sequence analysis. These include Arg-77 to Gly, Arg-78 to Lys, Phe-88 to Leu, Pro-564 to Leu, Val-650 to Ile, and Gln-651 to Glu. Some of these differences have been confirmed by the DNA sequence analysis of the genomic DNAs from individuals (37). Using agarose gel electrophoresis, several variant alleles for the \(a\) subunit of factor XIII have been reported in the normal population (33).

The substitution of a charged residue, such as Arg-77 or Glu-651, for an uncharged residue may contribute to the differential electrophoretic mobility of the gene products and may be related to the heterogeneity of the \(a\) subunit.

In plasma, the activation by thrombin involves the cleavage of the peptide bond following Arg-57 in the \(a\) subunit. This results in the release of an activation peptide of 37 residues from the NH\(_2\)-terminal end of the protein. Thrombin is also known to inactivate factor XIIIa and release a fragment (24 kDa) from the COOH terminus of the molecule (5, 34). The COOH-terminal cleavage by thrombin occurs primarily between Lys-513 and Ser-514 (29), even though this peptide bond is not a particularly good substrate for thrombin.

The active site Cys residue is located at position 314 within the sequence Tyr-Gly-Gln-Cys-Trp. This sequence is identical to that found in tissue transglutaminase (31, 35, 36). The intracellular mechanism for the activation of the \(a\) subunit of factor XIII is not known. Presumably, this occurs by minor proteolysis, although a conformational change in the protein by some other mechanism is possible.

Carbohydrate has not been identified in the six potential N-glycosylation sites (27, 29) suggesting that the \(a\) subunit may not contain 1.5% carbohydrate as originally reported (37). The two regions surrounding Gly-251 and Gly-473 ap-
pear to be good candidates for Ca\textsuperscript{2+}-binding sites in the molecule, as suggested by a comparison of adjacent sequences with a consensus sequence of other Ca\textsuperscript{2+}-binding proteins (38–40).

The cDNA clones coding for the plasma enzyme span about 3.8 kb (27, 30). The cDNAs indicated that the placental enzyme does not contain a typical hydrophobic leader sequence for secretion (27, 30). The a subunit of factor XIII, however, is known to remain in the cytoplasm of placenta (41), macrophages (42, 43), megakaryocytes (44), and platelets (45, 46). An acylated NH\textsubscript{2} terminus and the absence of glycosylation and disulfide bonds are also consistent with the fact that the a subunit is a typical cytoplasmic protein.

Some tissues, however, secrete the a subunit of factor XIII directly into the blood. The plasma enzyme apparently results from the secretion of the protein from a tissue that recognizes an internal secretion signal or some other signal by some unknown mechanism(s).

**Properties of Liver Tissue Transglutaminase**

A cDNA coding for guinea pig liver transglutaminase has been obtained from a liver cDNA library by Ikura et al. (31). Tissue transglutaminase also lacks a typical leader sequence with a hydrophobic core. The protein contains 690 amino acid residues with a calculated molecular weight of 78,620. The mature protein probably contains an NH\textsubscript{2}-terminal Ala that is blocked (31), although Connellan et al. (35) reported that the NH\textsubscript{2}-terminal region is pyro-Glu-Ala-Asp-Leu-. Also, the COOH-terminal sequence as determined by cDNA cloning is not identical with that previously reported (35). These differences in amino acid sequences may be due in part to polymorphisms.

The active site Cys residue in tissue transglutaminase is located at position 276. Although the molecule contains 17 Cys residues, none are involved in disulfide bond formation. Furthermore, tissue transglutaminase does not contain carbohydrate even though it contains six potential N-linked glycosylation sites.

Tissue transglutaminase can be isolated as an active enzyme from the cytoplasm of cells, and minor proteolysis is not required for its activation. Regions for potential Ca\textsuperscript{2+}-binding sites have not been identified in the protein sequence (31), although the enzyme requires Ca\textsuperscript{2+} for its activity. Consequently, two regions rich in Glu residues (around amino acids 450 and 470) have been proposed as possible Ca\textsuperscript{2+}-binding sites in the enzyme (31).

A partial amino acid sequence for human tissue transglutaminase purified from erythrocytes has been determined by Titani, Ando, Zenita, and Kannagi (see Ref. 47). As expected, human and guinea pig tissue transglutaminases as well as the a subunit of factor XIII (Fig. 1). Accordingly, it is likely that these three proteins evolved from a common ancestral gene. The middle portion of the two proteins contains most of the homologous sequences, and these occur in clusters and include the region surrounding the active site Cys residue. The NH\textsubscript{2}- and COOH-terminal regions of the two proteins show less homology. Hydroxylation analysis of human factor XIII and human and guinea pig tissue transglutaminases revealed that the active sites in these enzymes are located at a transition region between the hydrophilic and hydrophobic areas. This is consistent with a previous report that the active site Cys is located near a hydrophobic region of the molecule (48). The predicted secondary structures of this region of the various transglutaminases are also similar. A short amino acid sequence (4 out of 6 residues) in this region is also conserved in the active site Cys region of thiol proteases (29), although its significance, if any, is not clear.

Most recently, the primary structure of human erythrocyte membrane band 4.2 protein has been determined by cDNA cloning and found to be homologous (approximately 30%) to both tissue transglutaminase and the a subunit of factor XIII (49, 50). However, the active site Cys residue is substituted by an Ala residue in the band 4.2 protein, and it does not show transglutaminase activity (50). These three proteins share a common ancestor gene and evolved over a long period of time.

**Gene Coding for the a Subunit of Factor XIII**

The gene for the a subunit of factor XIII is located on chromosome 6 at p24–25 (51). The gene spans more than 160 kb and includes 15 exons interrupted by 14 introns (32). The activation peptide is encoded by the 2nd exon, the first putative Ca\textsuperscript{2+}-binding site by the 8th, the active site Cys region by the 7th, the second putative Ca\textsuperscript{2+}-binding site by the 11th, and a thrombin inactivation site by the 12th exon (32). Accordingly, the introns may separate the a subunit into functional and structural domains.

The organization and structure of the gene coding for tissue transglutaminase have not been reported. Accordingly, it is not known if the location of introns and the intron/exon splice junction types in this gene are similar to those in the a subunit of human factor XIII. The splice junction types and the location of introns in the genes coding for other families of proteins, such as the serine proteases that participate in the blood coagulation cascade, have been shown to be very
similar relative to their polypeptide sequence (52). A similar situation probably exists for the transglutaminases.

**Properties of the b Subunit of Human Factor XIII**

The primary structure of the b subunit has been determined by cDNA cloning and partial amino acid sequence analysis (28). It is composed of 641 amino acids with a molecular weight of 73,183 (Fig. 2). The addition of 8.5% carbohydrate (53) gives a molecular weight of about 79,700 for each subunit. The cDNA clones obtained from the human liver cDNA library span 2.2 kb of DNA. The amino-terminial sequence of the b subunit established by amino acid sequence analysis was Glu-Glu-Lys-Pro-, which is consistent with that deduced by cDNA cloning.

The cDNA also codes for 19 amino acids that constitute nearly the full-length leader sequence of 20 amino acids (Fig. 2). At the NH2-terminal position of the leader sequence, the initiator Met was identified by sequence analysis of the gene coding for the b subunit.

Near the COOH terminus of the b subunit of factor XIII there is an Arg-Gly-Asp sequence which plays a role in the binding of a number of proteins to cells (54). Whether or not this Arg-Gly-Asp sequence in the b subunit is related to cell adhesion is not known.

During the sequence analysis of the b subunit of factor XIII, 10 tandem repeats of about 60 amino acids were observed. These repetitive sequences also contained 4 half-Cys and highly conserved Pro, Gly, Tyr, Phe, and Trp residues (28). The internal identities within four subgroups of these repeats were 34–42%, while the overall identities between all 10 structures were lower.

A significant homology between the b subunit of factor XIII and a number of other proteins has also been observed. As in the b subunit of factor XIII, the homologous segments in these other proteins were composed of approximately 60 amino acids, including 4 half-Cys residues. Furthermore, the Pro, Gly, Tyr, Phe, and Trp residues were highly conserved. These repeats were initially called GP-I structures (52) because they were first identified in &beta-glycoprotein I (55). More recently, they have been called short consensus repeats or sushi structures because of their shape. The sushi structures have characteristic disulfide bonds between the 1st and 3rd, and the 2nd and 4th Cys residues in each repeat (55). It is highly likely that a similar pairing occurs with the disulfide bonds in the other sushi structures in other proteins. The b subunit of factor XIII contains 10 sushi structures that occur as tandem repeats representing 98% of the molecule (Fig. 2). Thus far, 26 proteins and two cDNAs have been found that contain sushi structures. These include complement receptor type 1 [30 sushi structures] and type 2 [16], complement protein H [20], the b subunit of factor XIII [10], &alpha subunit of C4 binding protein [8], endothelial leukocyte adhesion molecule 1 [6], &beta2 glycoprotein I [5], decay accelerating factor [4], membrane cofactor protein [4], &beta subunit of C4 binding protein [3], lymph node homing receptor [2], factor I [1], C6 [2], C7 [2], C2 [3], and factor B [3], Cr [2], and Cls [2], interleukin-2 receptor [2], cartilage proteoglycan core protein [1], fibroblast proteoglycan [1], thyroprotein [1], and haptoglobin [1 or 2] (52, 56).

Johnston et al. (57) identified nine sushi structures in a granule membrane protein (GMP-140) present in platelets and endothelial cells, while Paul et al. (58) reported two mouse cDNAs (called CRXY) which may code for five sushi structures. Furthermore, Tokunaga et al. (59) found that coagulation factor C from horseshoe crab contains five sushi structures. Also, a 35-kDa secretory protein of vaccinia virus has been reported to contain four sushi structures (60). Thus, there are proteins from vertebrates and invertebrates as well as viruses that contain sushi structures. Accordingly, this family of proteins forms one of the largest protein superfamilies known thus far and includes 14 proteins involved in complement and others that participate in diverse systems such as blood coagulation, lymphokines, and oxygen transport. Since most of these proteins bind to other proteins, it appears likely that the sushi structures often function as a protein-binding module. These proteins are usually assembled as multiple sushi structures or in combination with other segments such as growth factor domains.

Perkins et al. (61) proposed an extensive antiparallel &beta-sheet model for the sushi structures in complement factor H using Fourier transition infrared spectroscopy. Secondary structure predictions were also carried out and averaged on the basis of an alignment scheme for 101 sushi structures found in 13 proteins. As a result, a clear prediction of four strands of &beta-structure and four &beta-turns has been shown by both the Robson and the Chou-Fasman methods (61).

**The Gene Structure of the b Subunit for Factor XIII**

Recently, the gene coding for the b subunit of human factor XIII has been sequenced (62). The gene spans 28 kb of DNA and includes 12 exons interrupted by 11 introns. The first and last exons code for the leader sequence and COOH-terminal region, respectively, while the remaining 10 exons code for 10 sushi structures. The introns interrupt exactly the codons for the second amino acid residue from the last Cys in 7 of the 10 sushi structures, while in the remaining 3 sushi structures, the introns occur in the codons for the fourth or fifth amino acid from the last Cys (Fig. 2). Accordingly, each sushi structure is encoded by a single exon in the gene. Also, the first 10 splice junction types dividing the sushi structures were type 1 in that the introns divided codons between the first and second nucleotides. Type 1 splice junctions appear to be typical and favorable for exon shuffling during evolution (63).

The intron/exon boundaries for the genes of several of the other proteins containing sushi structures have also been reported. In the &alpha subunit of C4 binding protein, one intron...
is inserted exactly between the exons coding for the third and fourth sushi structures (64). Complement receptor types II has a combination of complete and divided domains, and two domains encoded by a single exon (65). In factor H, each of the 20 sushi structures is encoded by a single exon, except for the second repeat, which is encoded by two exons (66). In haptoglobin, two introns are located in the DNA corresponding to both ends of a sushi structure (67). In factor B, three sushi structures are divided by four introns (68). Two sushi structures in the interleukin-2 receptor are also encoded by two separate exons, and another exon which corresponds to the extra sequence is inserted between them (69, 70).

These data clearly show that a single exon, which corresponds to a single sushi structure, is the primary unit responsible for genetic shuffling and duplication in the evolution of these proteins.

It is noteworthy that many of the genes coding for the proteins containing multiple sushi structures are clustered on chromosome 1, band q22. These include complement receptor types 1 and 2, factor H, decay-accelerating factor, and the cytosolic and nuclear proteins containing multiple sushi structures in the interleukin-2 receptor are also encoded by two separate exons, and another exon which corresponds to the extra sequence is inserted between them (69, 70).

Three-dimensional Structure of Human Factor XIII

Recently, Carrell et al. (73) have published their studies on the electron microscopy of rotary-shadowed molecules for each of the α and β subunits of human factor XIII, and the αβ2 tetramer. The α dimer consists of two globular particles, and each of the α subunits is about 6 nm across, where the individual subunits are not distinguishable.

These results are consistent with the proposed subunit structure of factor XIII based on the results obtained by sedimentation and gel filtration (5, 6, 37, 75), except that the electron microscopic study indicates a monomeric molecule for the β subunit instead of the dimer.

REFERENCES

1. Folk, J. P., and Finlayson, J. S. (1977) Adv. Protein Chem. 31, 1-133
2. Lorand, L., Losowsky, M. G., and Milosevich, K. J. M. (1986) Prog. Hemostasis Thromb. 5, 245-290
3. Chung, S. L. (1979) Ann. N. Y. Acad. Sci. 202, 240-255
4. Folk, J. P. (1978) Annu. Rev. Biochem. 49, 517-531
5. Schwartz, M. L., Frazzo, S. V., Hill, R. L., and McKee, P. A. (1973) J. Biol. Chem. 248, 487-490
6. Chung, S. L., Lewis, M. S., and Folk, J. B. (1974) J. Biol. Chem. 249, 940-943
7. Lewis, S. D., Janus, T. J., Lorand, L., and Saffier, J. A. (1985) Biochemistry 24, 6772-6777
8. Takag, T., and Doddlin, R. F. (1974) Biochemistry 13, 730-750
9. Cooke, R. D., and Holbrook, J. J. (1974) Biochemistry 13, 41, 79-82
10. Lorand, L., Gray, A. J., Brown, K., Creed, B. M., Curtis, G. C., Domanik, R. A., and Sterenberg, P. (1974) Biochem. Biophys. Res. Commun. 61, 914-922
11. Curran, C. G., Brown, K. L., Creed, B. M., Domanik, R. A., Gray, A., Sterenberg, P., and Lorand, L. (1974) Biochemistry 13, 5745-5769
12. Cooke, R. D. (1974) Biochem. J. 141, 851-859
13. Fjian, J. P., Bromsted, J. T., Petyon, M. P., and Finlayson, J. S. (1972) Annu. Rev. Biochem. 41, 98-121
14. Sakata, Y., and Aoki, N. (1980) J. Clin. Invest. 65, 290-297
15. Sakata, Y., and Aoki, N. (1980) J. Clin. Invest. 65, 539-542
16. Mosher, D. F., and Jacob, F. (1979) J. Biol. Chem. 254, 885-890
17. Tamaki, T., and Aoki, N. (1981) Biochem. Biophys. Acta 681, 280-286
18. Mosher, D. F., Schad, P. E., and Kleinman, H. K. (1979) J. Clin. Invest. 64, 781-787
19. Duckett, R. (1972) Annu. N. Y. Acad. Sci. 202, 150-150