Cell, Vol. 41, 657–663, July 1985, Copyright © 1985 by MIT 0092-8674/85/070657-07 $02.00/0

Evolution of the Proteases of Blood Coagulation and Fibrinolysis by Assembly from Modules

Review

László Patthy
Institute of Enzymology
Biological Research Center
Hungarian Academy of Sciences
Budapest P.O. Box 7
Hungary, H-1502

The regulatory proteases of the fibrinolytic and blood coagulation systems have large noncatalytic segments attached to the amino-terminal end of the trypsin-homolog region, a feature that distinguishes them from the simple digestive proteases. In general, the function of these noncatalytic segments is to mediate binding of the proteases or their zymogens to other macromolecules or macroscopic structures and, through these interactions, to regulate the cascades of fibrinolysis and blood coagulation.

In principle, the evolutionary relationship between the low molecular weight digestive proteases and the high molecular weight regulatory proteases may be viewed in two different ways. According to one hypothesis, the growth of the noncatalytic region has been accomplished by a series of small tandem duplications (Doolittle, 1979) and subsequent evolution of the different domains of the noncatalytic region. Alternatively, the different domains of the noncatalytic region could have evolved outside the family of serine proteases and been fused to the catalytic region individually.

Some of the modules of the noncatalytic region of serine proteases also occur in fibronectin. We have suggested that the presence of homologous units in these otherwise unrelated proteins may be due to exon-shuffling (Bányai et al., 1983; Patthy et al., 1984). Based on this assumption, a detailed evolutionary scenario is proposed for the assembly of the noncatalytic region of proteases from modules.

Domain Organization of Trypsin-type Serine Proteases

The simple digestive proteases involved in the indiscriminate degradation of proteins or processing of biologically active peptides (pancreatic proteases, proteases of the glandular kallikrein family) are synthesized in the form of small precursors in which only a signal peptide-domain and a short zymogen peptide-domain are attached to the amino-terminal end of the catalytic region (Carne and Scheele, 1982; MacDonald et al., 1982; Mason et al., 1983). In the complex regulatory proteases of the complement activation, blood coagulation, and fibrinolytic cascades very large segments are inserted between the signal peptide- and zymogen peptide-domains (Kurachi and Davie, 1982; Degen et al., 1983; Morley and Campbell, 1984; MacGillivray and Davie, 1984; Fung et al., 1984; Ny et al., 1984; Anson et al., 1984). The large noncatalytic regions in the proteases of blood coagulation and fibrinolysis are essential for the biological specificity of the enzymes. The regulatory regions are organized into domains that belong to four main types: kringle-domains, vitamin K-dependent calcium-binding-domains, finger-domains, and growth factor-domains (Figure 1).

We have pointed out previously that kringle-domains correspond to autonomous structural, functional, and folding units that serve as versatile protein binding modules of the different proteases (Váli and Patthy, 1982; Trexler and Patthy, 1983; Patthy et al., 1984). On the basis of our observation that domains homologous to kringles are found in fibronectin, we proposed that the kringle-modules also possess a certain degree of evolutionary autonomy (Patthy et al., 1984).

Several lines of evidence indicate that the other domains of the noncatalytic regions may also be considered as autonomous units. The vitamin K-dependent calcium-binding-domain corresponds to an independent structural-functional unit (Ploplis et al., 1981; Olsson et al., 1982) that ensures the calcium-mediated binding of proteases to phospholipid membranes. The evolutionary autonomy of the finger-domains and growth factor-domains is strongly suggested by the fact that they also occur outside the family of serine proteases. The finger-domain of tissue-type plasminogen activator is closely related to the finger-structures of fibronectin (Bányai et al., 1983), and structures related to the growth factor-domains of proteases are now known to occur in epidermal growth factor precursor protein (Doolittle et al., 1984), and in low-density lipoprotein receptor (Russell et al., 1984). Further support for the view that all four domain-types correspond to genetically autonomous units comes from a survey of the domain composition of different proteases (Figure 1), which clearly suggests shuffling of these modules.

Accepting the view that the different domains of the noncatalytic regions correspond to autonomous "miniproteins" which have been fused to proteases at different times during evolution, it follows that the dendrograms for the different modules will not necessarily coincide with one another or with the dendrogram for the catalytic region. Detection of such discrepancies should therefore be useful in testing this view and in determining the point at which a given domain has been introduced.

Correct reconstruction of the evolutionary history of the individual modules on the basis of sequence homology assumes the unambiguous definition of the boundaries of the sequences that correspond to independent genetic units. Since the peptide segments separating the neighboring domains are usually not conserved, sequence similarity may drop well before the boundary of the genetic unit has been transgressed. The low degree of sequence similarity observed in connecting regions may thus be due to chance similarity of segments that do not share common ancestry, or to divergence of the same ancestral module. Conversely, if in two proteins homology extends to the adjacent module it may be difficult to determine the boundary of the modules just by sequence comparison. In such instances, localization of the limits of the independent genetic units may be aided by studies on the

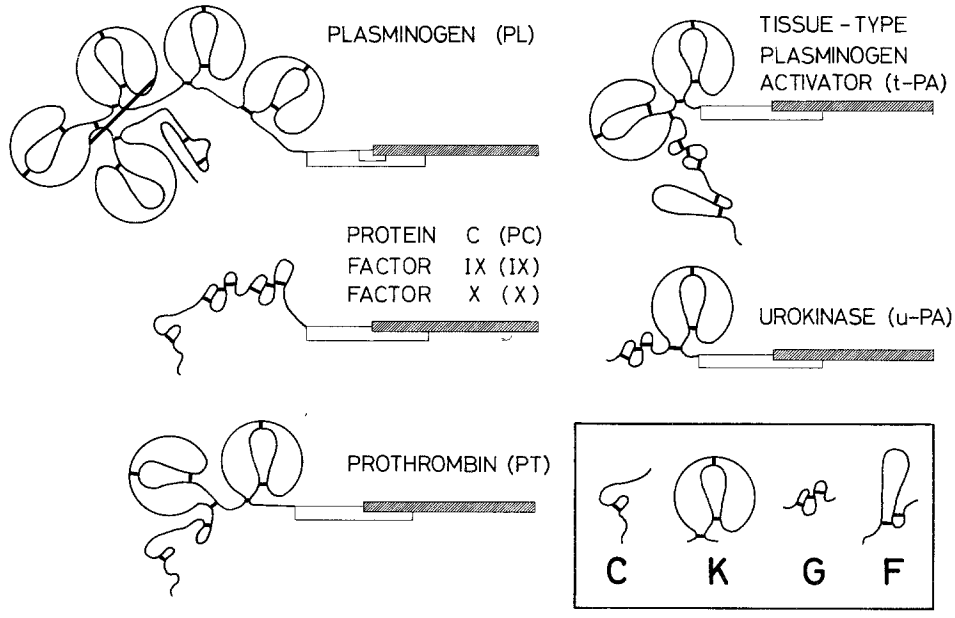

Figure 1. Structures of Plasminogen, Factor IX, Factor X, Protein C, Prothrombin, Urokinase, and Tissue-type Plasminogen Activator

The crosshatched bars represent the protease-regions homologous to trypsin. The inset shows the different modules of the nonprotease regions. (C):vitamin K-dependent calcium-binding-module, (K):kringle-module, (G):growth factor-module, and (F):finger-module.

exon-intron organization of the genes. This information, however, should be used only in conjunction with data on sequence similarity since during evolution introns may be lost or their positions may be shifted by changing the splicing pattern (Gilbert, 1978; Blake, 1983). Using the data obtained from sequence comparisons and from studies of gene structure, the modules were defined as follows:

Vitamin K-dependent calcium-binding-module. The sequences of the vitamin K-dependent proteases are very similar up to position 45 of factor IX. Beyond this point, the similarity between prothrombin and the other vitamin K-dependent proteases falls abruptly, suggesting that this area may be the border of the ancestral calcium-binding-module. This conclusion is supported by the exon-intron organization of the genes coding for factor IX and prothrombin. The intron separating the exons of the vitamin K-dependent calcium-binding domain from the exon of the growth factor-domain of factor IX splits the triplet of residue 47; the intron terminating the calcium-binding-domain of prothrombin lies at the same position (Anson et al., 1984). The evolutionary tree of the calcium-binding-module was therefore constructed using the sequences homologous to residues 1–47 of factor IX. It should be noted that this segment is much shorter than the one used previously to construct an evolutionary tree for the vitamin K-dependent calcium-binding-domains of different proteases (Hewett-Emmett et al., 1981). These authors compared sequences thought to be homologous to residues 1–65 of prothrombin and included segments in their comparison that are now known to belong to the first growth factor-domains of protein C, factor IX, and factor X (see below).

Growth factor-module. The fourth and fifth exons of the human factor IX gene code for residues 47–85 and 85–128 (Anson et al., 1984), sequences that correspond to the two growth factor-domains defined previously on the basis of homology with epidermal growth factor (Young et al., 1978; Doolittle et al., 1984). The fifth exon of tissue-type plasminogen activator codes for residues 50–87 (Ny et al., 1984), a region that is homologous to epidermal growth factor, the growth factor-domains of urokinase, protein C, factor IX, and factor X (Bányai et al., 1983). The limits of the growth factor-domains redefined in view of the limits of the exons differ slightly from those indicated previously (Doolittle et al., 1984) in that the small segments corresponding to residues 45–46 of factor IX are now known to belong to the vitamin K-dependent calcium-binding module (see above).

Kringle-module. All previous sequence comparisons of kringles defined the first and sixth conserved cysteines as the boundaries of the kringle-domains, since sequence similarity drops dramatically outside these regions (Sottrup-Jensen et al., 1978; Young et al., 1978; Hewett-Emmett et al., 1981; Patthy et al., 1984). Studies on the structure of the genes show that the exons coding for kringles always terminate outside, but usually do not extend more than four residues beyond these boundaries, as in the case of the kringles of tissue-type plasminogen activator and urokinase (Ny et al., 1984; Verde et al., 1984). The splice junction at the amino-terminal boundary of some kringles (e.g., the fourth kringle of plasminogen, the S kringle of prothrombin) has been shifted upstream, thereby lengthening the sequences that separate the adjacent modules (Degen et al., 1983; Malinowski et al., 1984). The limits of the kringle-modules were redefined in

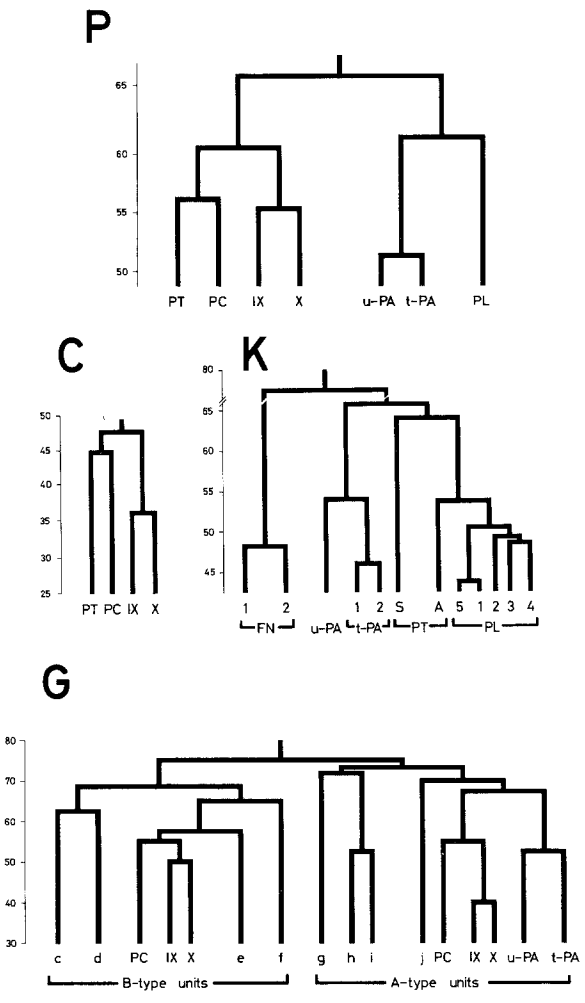

Figure 2. Evolutionary Trees of the Individual Modules

The trees are labeled as (P):protease-module, (C):Vitamin K-dependent calcium-binding-module, (K):kringle-module, (G):growth factor-module. The vertical axes indicate the percent difference between sequences. These values were calculated from the unitary matrix determined by pairwise comparison of individual sequences, assigning a value of +1 to identical residues and 0 to nonidentical ones. The proteins are labeled as (PL):plasminogen, (PT):prothrombin, (PC):protein C, (IX):factor IX, (X):factor X, (u-PA):urokinase, (t-PA):tissue-type plasminogen activator, (FN):fibronectin, and (c,d,e,f,g,h,i,j):epidermal growth factor precursor protein.

view of these findings by extending the limits to the boundaries of the neighboring modules, but not beyond the position of the introns.

Finger-module. The fourth exon of the gene for tissue-type plasminogen activator codes for residues 4–50 (Ny et al., 1984), a segment that is homologous with the finger-domains of fibronectin (Bányai et al., 1983). The boundaries of the finger-domains established on the basis of sequence comparison of fibronectin-fingers (Skorstengaard et al., 1984) agree with the limits of this exon.

Protease-module. The protease-module was defined as the region homologous to bacterial trypsin (Young et al., 1978).

Genealogy of Individual Modules

We have determined independent evolutionary trees for each of the modules. The optimum alignment of the sequences was found using a scoring system in which identities are equal to +10, identity of cysteines is equal to +20, and each gap is equal to −25 (Jue et al., 1980; Doolittle, 1981). Terminal gaps are not included in this calculation. Genealogy of sequences was determined on the basis of percent difference and alignment scores (Schwartz and Dayhoff, 1978), the position and number of gaps, and the position of disulphide bridges (deHaën et al., 1975) using the parsimony principle. The hierarchies of the sequences determined with the different scoring methods were found to be the same.

The sequences of human and bovine prothrombin (Magnusson et al., 1975; Walz et al., 1977; Degen et al., 1983), human and bovine protein C (Fernlund and Stenflo, 1982; Stenflo and Fernlund, 1982; Foster and Davie, 1984), human and bovine factor IX (Katayama et al., 1979; Kurachi and Davie, 1982; Anson et al., 1984), human and bovine factor X (Enfield et al., 1980; Leytus et al., 1984; Fung et al., 1984), human urokinase (Günzler et al., 1982; Steffens et al., 1982; Verde et al., 1984), human tissue-type plasminogen activator (Pennica et al., 1983; Ny et al., 1984), human plasminogen (Sottrup-Jensen et al., 1978), bovine fibronectin (Skorstengaard et al., 1984), murine epidermal growth factor precursor protein (Gray et al., 1983), and bovine low-density lipoprotein receptor (Russell et al., 1984) were taken from the literature.

The dendrogram of the protease-module shows that the proteases of the blood coagulation cascade are closely related and, within this group, prothrombin and protein C form one subgroup, while factor IX and factor X form another. Plasminogen and the two plasminogen activators are more closely related to each other than either is to any of the blood coagulation enzymes (Figure 2P).

The evolutionary tree of the vitamin K-dependent calcium-binding-module is consistent with the genealogy of the catalytic regions of the blood coagulation enzymes (Figures 2P and 2C), suggesting that these two regions were genetically linked before the divergence of prothrombin, protein C, factor IX, and factor X.

The evolutionary tree of the kringle-module, based on the sequence boundaries redefined by the genes' exon-intron structures (Figure 2K), is indistinguishable from that established previously (Patthy et al., 1984). The high degree of sequence similarity of the plasminogen kringles suggests that they arose through internal gene multiplication. The similarity between the two kringles of tissue-type plasminogen activators is also indicative of internal gene duplication. The recent finding that the coding regions of both tissue-type plasminogen activator kringles are divided by introns at identical positions (Ny et al., 1984) underlines the homology of these kringles. The evolutionary tree for the group of plasminogen-kringles, the group of tissue-type plasminogen activator kringles, and the urokinase-kringle is consistent with the genealogy of their protease-modules (Figures 2P and 2K), indicating that the

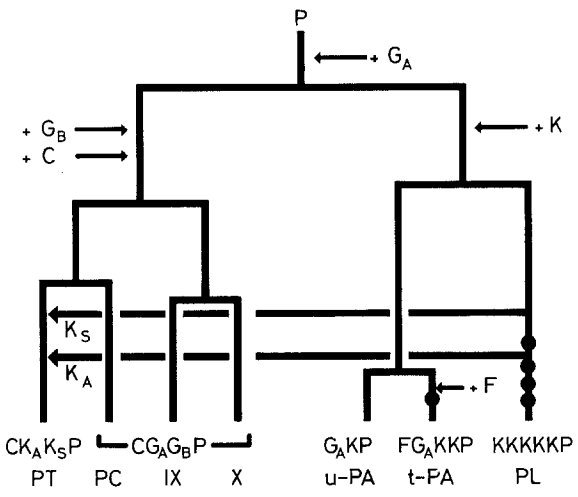

Figure 3. Evolutionary Tree Reconciling the Dendrograms of the Individual Modules

The vertical axis of the tree has a time dimension. The branches are drawn similarly to those in Figure 2P since it is assumed that, on average, the catalytic regions of the different proteases change at the same rate. P:protease-module, C:calcium-binding module, K:kringle-module, F:finger-module, G_A:A-type growth factor-module, G_B:B-type growth factor-module; ● represents an internal gene duplication.

two modules were genetically linked prior to the divergence of plasminogen and the plasminogen activators.

Both kringles of prothrombin, however, are more closely related to plasminogen-kringles than to the plasminogen activator-kringles (Patthy et al., 1984), although the catalytic-modules of the blood coagulation enzymes and fibrinolytic enzymes diverged before the divergence of plasminogen and plasminogen activators. This conflict between the two evolutionary trees (Figures 2P and 2K) can be eliminated if one assumes that both kringles of prothrombin were derived from an ancestor of plasminogen. The fact that kringle S of prothrombin is more distantly related to plasminogen-kringles than are any of the plasminogen-kringles to each other, indicates that the ancestor of kringle S was transferred to prothrombin before the multiplication and divergence of plasminogen-kringles. The close similarity of prothrombin-kringle A to plasminogen-kringles suggests that the transfer of this kringle occurred at about the time of the multiplication and divergence of plasminogen-kringles.

Comparison of the two (A- and B-type growth factor-modules of protein C, factor IX, factor X, and the ten (a–j) growth factor-domains of epidermal growth factor precursor has shown that the A-domains of proteases are more closely related to the g,h,i, and j domains, whereas the B-domains are more similar to the c,d,e, and f domains (Doolittle et al., 1984). The single growth factor-domains of urokinase and tissue-type plasminogen activator belong to the group of A-type-domains (Bányai et al., 1983). An evolutionary tree for the growth factor-modules of proteases and the c,d,e,f,g,h,i,j units of epidermal growth factor precursor is presented in Figure 2G. (In this dendrogram, the a,b units of epidermal growth factor precursor are not shown, since they are only distantly related to the other growth factor units. The single growth factor-domain of low-density lipoprotein receptor, which corresponds to the f unit of epidermal growth factor precursor [Russell et al., 1984], is not indicated.) The evolutionary tree for the A-type growth factor domains of proteases is consistent with the genealogy of their catalytic-modules (Figures 2P and 2G), suggesting that these regions were genetically linked before the divergence of the fibrinolytic and blood coagulation enzymes. The A-type growth factor-domains of proteases are most closely related to the j unit of epidermal growth factor precursor.

The evolutionary tree for the B-type units of proteases is also consistent with the genealogy of the protease-regions, indicating the linkage of these units prior to the divergence of protein C, factor IX, and factor X. The B-type units of proteases are more closely related to the c,d,e, and f units of epidermal growth factor precursor than they are to the A-type units of proteases; this suggests that they were derived from a B-type unit of a protein related to epidermal growth factor precursor, rather than from the A-domains of proteases by gene duplication. The B-domains of proteases are most closely related to the e unit of epidermal growth factor precursor.

Assembly of the Noncatalytic Regions of Proteases from Modules

A hypothetical sequence of events, reconciling the dendrograms of each homology unit is shown in Figure 3. According to this hypothesis, the ancestral trypsin-like protease at the root of this combined evolutionary tree already had a signal peptide- and a zymogen peptide-domain, since the sequences regulating the transport and activity of secretory proteases evolved before the divergence of the blood coagulation and fibrinolytic enzymes from the digestive proteases of the pancreas (Neurath, 1975; James, 1980; MacDonald et al., 1982; Carne and Scheele, 1982; Kurachi and Davie, 1982; Mason et al., 1983; Degen et al., 1983; Neurath, 1984; MacGillivray and Davie, 1984; Fung et al., 1984; Ny et al., 1984; Anson et al., 1984). It is assumed that all other regulatory modules of plasma proteases were inserted later, between the signal peptide and zymogen-peptide domains.

Before the divergence of the blood coagulation and fibrinolytic enzymes a growth factor-domain (G_A), closely related to the j unit of epidermal growth factor precursor, was probably inserted in the ancestral protease. Descendants of this domain are found in urokinase, tissue-type plasminogen activator, and in the A units of protein C, factor IX, and factor X; however, no related structure appears to be preserved in prothrombin or plasminogen.

The hypothesis assumes that after the divergence of the fibrinolytic and blood coagulation enzymes—but before the divergence of prothrombin, protein C, factor IX, and factor X—a second growth factor-domain (G_B), closely related to the e unit of epidermal growth factor precursor, was inserted along with a calcium-binding-domain (C) in an ancestral blood coagulation enzyme. The relative order of these events cannot be determined at present. This ancestral domain organization is preserved in protein C, factor IX, and factor X but, in pro-

thrombin, the two growth factor-domains are replaced by two kringles (see below).

Following the divergence of the fibrinolytic and blood coagulation enzymes, a kringle-domain was inserted in the common ancestor of plasminogen and plasminogen activators. It is proposed that after the separation of the lines of plasminogen and plasminogen activators—at a time preceding the internal multiplication of plasminogen-kringles—a kringle of ancestral plasminogen (K_S) was transferred to prothrombin. The kringle-domain of ancestral plasminogen underwent internal quintuplication and, at about the same time, a second genetic exchange between ancestral plasminogen and prothrombin transferred a second kringle (K_A) to prothrombin. (Because the A kringle of prothrombin is much more similar to the kringles of plasminogen than to the S kringle of prothrombin, it was suggested previously that the A kringle was derived from one of the plasminogen kringles [Kurosky et al., 1980].) It is not known when the hypothetical growth factor-domain was replaced by the amino-terminal peptide in plasminogen or whether this structure is a descendant of the growth factor-domain.

After the divergence of urokinase and tissue-type plasminogen activator a finger-module (F), related to fibronectin-fingers, was inserted into tissue-type plasminogen activator and the kringle-module was duplicated.

Very little is known about the time of the proposed evolutionary events outlined above. The fact that prothrombin, plasminogen, and other components of the blood coagulation and fibrinolytic systems of birds are similar to those of mammalian species (Brockway and Castellino, 1972; Hewett-Emmett et al., 1974; Belleville et al., 1982) indicates that the events leading to the present domain organization of these proteins preceded the divergence of birds and mammals, and probably date back to the time of appearance of vertebrates (Doolittle, 1979).

Mechanisms Underlying the Changes in the Domain Composition of Proteases

Internal duplications resulting from unequal cross-over are obvious in the case of the two kringles of tissue-type plasminogen activator and the five kringles of plasminogen. The most parsimonious combined evolutionary tree excludes internal duplication in the case of the two prothrombin-kringles and the two growth factor-domains of protein C, factor IX, and factor X and suggests independent insertion of these modules.

The genetic events leading to insertion of modules in proteases depend on the origin of the modules. If they existed as independent miniproteins, or were parts of proteins unrelated to proteases, their insertion might be the result of random transposition of genetic material. The available data do not provide any information about the mechanism by which proteases acquired the calcium-binding-module, the A-type growth factor-module, or the first kringle-module.

The transfer of modules between proteins showing sequence homology may be facilitated by mispairing and double cross-over or gene conversion. Recent evidence shows that exchange of segments between genes within the same multigene family by this mechanism has played a significant role in the evolution of neurophysins, globins, immunoglobulins, and major histocompatibility antigens (Liebhaber et al., 1981; Schreier et al., 1981; Schon et al., 1982; Hayashida and Miyata, 1983; Weiss et al., 1983; Ruppert et al., 1984). Introduction of the second, B-type growth factor-domain into the ancestral blood coagulation protease probably occurred by a similar mechanism. Mispairing and genetic exchange would be facilitated by the preexisting homology between the A-type growth factor-domain of this protease and the growth factor units of ancestral epidermal growth factor precursor. Mispairing of the A unit of the ancestral blood coagulation protease (G_AP or CG_AP) with one unit (d) of the cluster of B-type growth factor-domains in ancestral epidermal growth factor precursor protein may have permitted transfer of a neighboring unit (e) into the protease by gene conversion. This would provide a simple explanation for why the order of A- and B-type domains in proteases is the reverse of that in epidermal growth factor precursor (Doolittle et al., 1984).

The transfer of kringle S of prothrombin from an ancestral plasminogen has probably been facilitated by the homology of ancestral prothrombin (CG_AG_BP) and ancestral plasminogen (G_AKP). Mispairing of the two related regions and gene conversion or double cross-over in the region between the two aligned regions might explain the introduction of kringle S and loss of the B-type growth factor-domain. The increased homology between the resulting hybrid (CG_AK_SP) and a more advanced plasminogen ancestor, in which duplication of kringles and obliteration of the growth factor-domain has already occurred (KKP, KKKP, etc.), would increase the probability of further mispairing and genetic exchange between these ancestors of prothrombin and plasminogen. Such an event may have been responsible for the transfer of a second kringle (K_A) and loss of the A-type growth factor domain (G_A) from prothrombin.

Similarly, since the kringle-domains are homologous to the type II units of fibronectin (Patthy et al., 1984), this partial homology probably facilitated the mispairing and further exchange of genetic information between the fibrinolytic proteases and fibronectin. The introduction of finger-domain to the ancestor of tissue-type plasminogen activator from an ancestral fibronectin may have been facilitated by the preexisting homology of the kringle-like regions of proteases and fibronectin.

Irrespective of the actual mechanisms responsible for the shuffling of modules, it seems likely that the introns separating the finger-, growth factor-, kringle-, and calcium-binding-modules would increase the probability of recombination outside the coding regions, and thus aid in the reassortment of the modules (Gilbert, 1978).

Predictions and Possible Tests of the Hypothesis

The model assumes that the changes in the domain composition of proteases were brought about by internal gene duplications, by gene conversion or double cross-over be-

tween partially homologous genes, and by random transposition of genetic material. The latter assumption implicitly predicts that, on an evolutionary time scale, these modules may behave like transposons. After submission of this manuscript, Blomquist et al. (1984) and Brown et al. (1985) published data that are in line with this prediction. These authors found that a 19 kd protein of vaccinia virus possesses a domain closely homologous with epidermal growth factor; the gene for this protein lies in a region that has a transposon-like arrangement. This finding raises the possibility that viruses serve as vehicles for the transfer of modules.

The evolutionary scenario shown in Figure 3 predicts specific intermediate states in the evolution of the proteases of blood coagulation and fibrinolysis. The validity of the hypothesis can be probed by structural studies on other components of the blood coagulation and fibrinolytic cascades. Determination of the structures of plasma proteases from lower vertebrates and invertebrates may help in assigning a time scale to the depicted events.

References

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A., and Brownlee, G. G. (1984). The gene structure of human anti-haemophilic factor IX. EMBO J. *3*, 1053–1060.

Bányai, L., Váradi, A., and Patthy, L. (1983). Common evolutionary origin of the fibrin-binding structures of fibronectin and tissue-type plasminogen activator. FEBS Lett. *163*, 37–41.

Belleville, J., Cornillon, B., Paul, J., Baguet, J., Clendinnen, G., and Eloy, R. (1982). Haemostasis, blood coagulation and fibrinolysis in the Japanese quail. Comp. Biochem. Physiol. *71A*, 219–230.

Blake, C. (1983). Exons—present from the beginning? Nature *306*, 535–537.

Blomquist, M. C., Hunt, L. T., and Barker, W. C. (1984). Vaccinia virus 19-kilodalton protein: relationship to several mammalian proteins, including two growth factors. Proc. Natl. Acad. Sci. USA *81*, 7363–7367.

Brockway, W. J., and Castellino, F. J. (1972). Measurement of the binding of antifibrinolytic amino acids to various plasminogens. Arch. Biochem. Biophys. *151*, 194–199.

Brown, J. P., Twardzik, D. R., Marquardt, H., and Todaro, G. J. (1985). Vaccinia virus encodes a polypeptide homologous to epidermal growth factor and transforming growth factor. Nature *313*, 491–492.

Carne, T., and Scheele, G. (1982). Amino acid sequences of transport peptides associated with canine exocrine pancreatic proteins. J. Biol. Chem. *257*, 4133–4140.

Degen, S. J. F., MacGillivray, R. T. A., and Davie, E. W. (1983). Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin. Biochemistry *22*, 2087–2097.

deHaën, C., Neurath, H., and Teller, D. C. (1975). The phylogeny of trypsin-related serine proteases and their zymogens. New methods for the investigation of distant evolutionary relationships. J. Mol. Biol. *92*, 225–259.

Doolittle, R. F. (1979). Protein evolution. In The Proteins, Vol. 4, H. Neurath and R. L. Hill, eds. (New York: Academic Press), pp. 1–118.

Doolittle, R. F. (1981). Similar amino acid sequences: chance or common ancestry? Science *214*, 149–159.

Doolittle, R. F., Feng, D. F., and Johnson, M. S. (1984). Computer-based characterization of epidermal growth factor precursor. Nature *307*, 558–560.

Enfield, D. L., Ericsson, L. H., Fujikawa, K., Walsh, K. A., Neurath, H., and Titani, K. (1980). Amino acid sequence of the light chain of bovine factor X_1 (Stuart factor). Biochemistry *19*, 659–667.

Fernlund, P., and Stenflo, J. (1982). Amino acid sequence of the light chain of bovine protein C. J. Biol. Chem. *257*, 12170–12179.

Foster, D., and Davie, E. W. (1984). Characterization of a cDNA coding for human protein C. Proc. Natl. Acad. Sci. USA *81*, 4766–4770.

Fung, M. R., Campbell, R. M., and MacGillivray, T. A. (1984). Blood coagulation factor X mRNA encodes a single polypeptide chain containing a prepro leader sequence. Nucl. Acids Res. *12*, 4481–4492.

Gilbert, W. (1978). Why genes in pieces? Nature *271*, 501.

Gray, A., Dull, T. J., and Ullrich, A. (1983). Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor. Nature *303*, 722–725.

Günzler, W. A., Steffens, G. J., Ötting, F., Kim, S-M. A., Frankus, E., and Flohé, L. (1982). The primary structure of high molecular mass urokinase from human urine. The complete amino acid sequence of the A chain. Hoppe-Seyler's Z. Physiol. Chem. *363*, 1155–1165.

Hayashida, H., and Miyata, T. (1983). Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA *80*, 2671–2675.

Hewett-Emmett, D., McCoy, L. E., Hassouna, H. I., Reuterby, J., Walz, D. A., Seegers, W. H. (1974). A partial gene duplication in the evolution of prothrombin? Thrombosis Research *5*, 421–430.

Hewett-Emmett, D., Czelusniak, J., and Goodman, M. (1981). The evolutionary relationships of the enzymes involved in blood coagulation and hemostasis. Ann. NY Acad. Sci. *370*, 511–527.

James, M. N. G. (1980). An X-ray crystallographic approach to enzyme structure and function. Can. J. Biochem. *58*, 251–271.

Jue, R. A., Woodbury, N. W., and Doolittle, R. F. (1980). Sequence homologies among E. coli ribosomal proteins: evidence for evolutionarily related groupings and internal duplications. J. Mol. Evol. *15*, 129–148.

Katayama, K., Ericsson, L. H., Enfield, D. L., Walsh, K. A., Neurath, H., Davie, E. W., and Titani, K. (1979). Comparison of amino acid sequence of bovine coagulation factor IX (Christmas factor) with that of other vitamin K-dependent plasma proteins. Proc. Natl. Acad. Sci. USA *76*, 4990–4994.

Kurachi, K., and Davie, E. W. (1982). Isolation and characterization of a cDNA coding for human factor IX. Proc. Natl. Acad. Sci. USA *79*, 6461–6464.

Kurosky, A., Barnett, D. R., Lee, T-H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H., and Fitch, W. M. (1980). Covalent structure of human haptoglobin: A serine protease homolog. Proc. Natl. Acad. Sci. USA *77*, 3388–3392.

Leytus, S. P., Chung, D. W., Kisiel, W., Kurachi, K., and Davie, E. W. (1984). Characterization of a cDNA coding for human factor X. Proc. Natl. Acad. Sci. USA *81*, 3699–3702.

Liebhaber, S. A., Goossens, M., and Kan, Y. W. (1981). Homology and concerted evolution at the $\alpha 1$ and $\alpha 2$ loci of human α-globin. Nature *290*, 26–29.

MacDonald, R. J., Swift, G. H., Quinto, C., Swain, W., Pictet, R. L., Nikovits, W., and Rutter, W. J. (1982). Primary Structure of two distinct rat pancreatic preproelastases determined by sequence analysis of the complete cloned messenger ribonucleic acid sequences. Biochemistry *21*, 1453–1463.

MacGillivray, R. T. A., and Davie, E. W. (1984). Characterization of bovine prothrombin mRNA and its translation product. Biochemistry *23*, 1626–1634.

Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., and Claeys, H. (1975). Complete primary structure of prothrombin: isolation, structure and reactivity of ten carboxylated glutamic acid residues and regulation of prothrombin activation by thrombin. In Proteases and Biological Control, E. Reich, D. B. Rifkin, and E. Shaw, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 123–147.

Malinowski, D. P., Sadler, J. E., and Davie, E. W. (1984). Characterization of a complementary deoxyribonucleic acid coding for human and bovine plasminogen. Biochemistry *23*, 4243–4250.

Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., and Richards, R. I. (1983). Structure of mouse kallikrein gene family suggests a role in specific processing of biologically active peptides. Nature *303*, 300–307.

Morley, B. J., and Campbell, R. D. (1984). Internal homologies of the Ba fragment from human complement component Factor B, a class III MHC antigen. EMBO J. *3*, 153–157.

Neurath, H. (1975). Limited proteolysis and zymogen activation. In Proteases and Biological Control, E. Reich, D. B. Rifkin, and E. Shaw, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 51–64.

Neurath, H. (1984). Evolution of proteolytic enzymes. Science *224*, 350–357.

Ny, T., Elgh, F., and Lund, B. (1984). The structure of the human tissue-type plasminogen activator gene: correlation of intron and exon structures to functional and structural domains. Proc. Natl. Acad. Sci. USA *81*, 5355–5359.

Olsson, G., Andersen, L., Lindqvist, O., Sjolin, L., Magnusson, S., Petersen, T. E., and Sottrup-Jensen, L. (1982). A low resolution model of fragment 1 from bovine prothrombin. FEBS Lett. *145*, 317–322.

Patthy, L., Trexler, M., Váli, Z., Bányai, L., and Váradi, A. (1984). Kringles: modules specialized for protein binding. Homology of the gelatin-binding region of fibronectin with the kringle structures of proteases. FEBS Lett. *171*, 131–136.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., and Collen, D. (1983). Cloning and expression of human tissue-type plasminogen activator cDNA in E. coli. Nature *301*, 214–221.

Ploplis, V. A., Strickland, D. K., and Castellino, F. J. (1981). Calorimetric evaluation of the existence of separate domains in bovine prothrombin. Biochemistry *20*, 15–21.

Ruppert, S., Scherer, G., and Schutz, G. (1984). Recent gene conversion involving bovine vasopressin and oxytocin precursor genes suggested by nucleotide sequence. Nature *308*, 554–557.

Russell, D. W., Schneider, W. J., Yamamoto, T., Luskey, K. L., Brown, M. S., and Goldstein, J. L. (1984). Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor. Cell *37*, 577–585.

Schon, E. A., Wernke, S. M., and Lingrel, J. B. (1982). Gene conversion of two functional goat α-globin genes preserves only minimal flanking sequences. J. Biol. Chem. *257*, 6825–6835.

Schreier, P. H., Bothwell, A. L. M., Mueller-Hill, B., and Baltimore, D. (1981). Multiple differences between the nucleic acid sequences of the $IgG2a^a$ and $IgG2a^b$ alleles of the mouse. Proc. Natl. Acad. Sci. USA *78*, 4495–4499.

Schwartz, R. M., and Dayhoff, M. O. (1978). In Atlas of Protein Sequence and Structure, M. O. Dayhoff, ed. (Silver Spring, Maryland: National Biomedical Research Foundation), pp. 353–358.

Skorstengaard, K., Thøgersen, H. C., and Petersen, T. E. (1984). Complete primary structure of the collagen-binding domain of bovine fibronectin. Eur. J. Biochem. *140*, 235–243.

Sottrup-Jensen, Claeys, H., Zajdel, M., Petersen, T. E., and Magnusson, S. (1978). The primary structure of human plasminogen: isolation of two lysine-binding fragments and one "mini"-plasminogen (MW 38,000) by elastase-catalyzed-specific limited proteolysis. In Progress in Chemical Fibrinolysis and Thrombolysis, Volume 3, J. F. Davidson, R. M. Rowan, M. M. Samama, and P. C. Desnoyers, eds. (New York: Raven Press), pp. 191–209.

Steffens, G. J., Günzler, W. A., Ötting, F., Frankus, E., and Flohé, L. (1982). The complete amino acid sequence of low molecular mass urokinase from human urine. Hoppe-Seyler's Z. Physiol. Chem. *363*, 1043–1058.

Stenflo, J., and Fernlund, P. (1982). Amino acid sequence of the heavy chain of bovine protein C. J. Biol. Chem. *257*, 12180–12190.

Trexler, M., and Patthy, L. (1983). Folding autonomy of the kringle 4 fragment of human plasminogen. Proc. Natl. Acad. Sci. USA *80*, 2457–2461.

Váli, Z., and Patthy, L. (1982). Location of the intermediate and high affinity ω-aminocarboxylic acid-binding sites in human plasminogen. J. Biol. Chem *257*, 2104-2110.

Verde, P., Stoppelli, M. P., Galeffi, P., Nocera, P. D., and Blasi, F. (1984). Identification and primary sequence of an unspliced human urokinase poly(A)$^+$ RNA. Proc. Natl. Acad. Sci. USA *81*, 4727–4731.

Walz, D. A., Hewett-Emmett, D., and Seegers, W. H. (1977). Amino acid sequence of human prothrombin fragments 1 and 2. Proc. Natl. Acad. Sci. USA *74*, 1969–1972.

Weiss, E. H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J., and Flavell, R. A. (1983). The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion-like events. Nature *301*, 671–674.

Young, C. L., Barker, W. C., Tomaselli, C. M., and Dayhoff, M. O. (1978). Serine proteases. In Atlas of Protein Sequence and Structure, M. O. Dayhoff, ed. (Silver Spring, Maryland: National Biomedical Research Foundation), pp. 73–93.