Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5355–5359, September 1984
Biochemistry

The structure of the human tissue-type plasminogen activator gene: Correlation of intron and exon structures to functional and structural domains

(serine protease/genomic cloning/DNA sequence analysis)

TOR NY*, FREDRIK ELGH, AND BJÖRN LUND

Unit for Applied Cell and Molecular Biology, University of Umeå, S-901 87 Umeå, Sweden

Communicated by David M. Prescott, May 14, 1984

ABSTRACT **A genomic clone carrying the human tissue-type plasminogen activator (t-PA) gene was isolated from a cosmid library, and the gene structure was elucidated by restriction mapping, Southern blotting, and DNA sequencing. The cosmid contained all the coding parts of the mRNA, except for the first 58 bases in the 5′ end of the mRNA, and had a total length of >20 kilobases. It was separated into at least 14 exons by at least 13 introns, and the exons seemed to code for structural or functional domains. Thus, the signal peptide, the propeptide, and the domains of the heavy chain, including the regions homologous to growth factors, and to the "finger" structure of fibronectin, are all encoded by separate exons. In addition, the two kringle regions of t-PA were both coded for by two exons and were cleaved by introns at identical positions. The region coding for the light chain, comprising the serine protease part of the molecule was split by four introns, revealing a gene organization similar to other serine proteases.**

Plasminogen activators (PAs), are a class of serine proteases that convert the proenzyme plasminogen into plasmin, which then degrades the fibrin network of blood clots (1, 2). The PAs have been classified into two immunologically unrelated groups, the urokinase-type PA (u-PA) (M_r, 55,000) and the tissue-type PA (t-PA) (M_r, 72,000) (3, 4). The latter activator is believed to be the physiological vascular activator (5) and is composed of a single polypeptide chain (6). However, in the presence of plasmin or trypsin it is cleaved at a single site in the central region of the molecule (7), converting it into a two-chain disulfide-linked form. This latter form is composed of a light and a heavy chain, derived from the COOH-terminal and NH_2-terminal parts, respectively. The light chain of t-PA contains the active site and is highly homologous to other serine proteases (8). The heavy chain exhibits a number of structural features homologous to structures found in other plasma proteins. Thus, the heavy chain contains two kringle structures (8) similar to those found in prothrombin (9, 10), plasminogen (11), and urokinase (12). It also possesses a growth factor-like domain (12, 13) and a domain that shows homology to the fibrin-binding "finger"-like structures of fibronectin (13).

Although it has been postulated that exons represent genetic building blocks coding for discrete structural or functional domains (14, 15), other studies seem to argue against this simple model (16). Herein we describe the isolation and characterization of a cosmid carrying the gene for t-PA. DNA sequence analyses revealed that the structural domains of this protein correlate well with the exon–intron pattern of the gene.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Materials. Restriction enzymes, T4 DNA ligase, and *Escherichia coli* DNA polymerase I were purchased from New England BioLabs. T4 polynucleotide kinase was from Boehringer Mannheim GmbH. Nick-translation kits, [α-^{32}P]dGTP (3000 Ci/mmol; 1 Ci = 37 GBq), and [γ-^{32}P]ATP (5000 Ci/mmol) were from Amersham.

General Methods. Plasmid and cosmid DNA were isolated by a modification of the clear lysate procedure of Birnboim and Doly (17) and Grosveld *et al.* (18), followed by two consecutive ethidium bromide/CsCl equilibrium centrifugations. Enzyme reactions were carried out according to the conditions suggested by the suppliers.

Screening of Human Cosmid Library. A human cosmid library derived from placental DNA (19, 20) (a generous gift from Werner Lindenmaier) was used and screened essentially according to Grosveld *et al.* (18). *In vivo* packaged cosmids were transduced into the *E. coli* K-12 strain HB101 (21). The cosmid-containing bacteria were grown on nitrocellulose filters on selective agar plates. Replica filters were prepared and colony hybridization was carried out.

Analysis of Cloned DNA. The purified DNAs were cleaved with various restriction enzymes either singly or in combination, and the DNA fragments were separated on agarose gels (0.5%–2.0%) or polyacrylamide gels (5%–8%). DNA blotting was carried out according to Southern (22), and the filters were hybridized as described by Moseley *et al.* (23).

DNA Sequence Analysis. DNA fragments were cloned into the M13 cloning vectors mp8 and mp9 (24). Chimeric phages carrying exon DNA were identified either by sequence analysis or by plaque hybridization (25). Either nick-translated cDNA (26) or synthetic [γ-^{32}P]ATP-labeled oligodeoxyribonucleotides (27) covering different parts of the t-PA gene were used as hybridization probes. The synthetic probes were designed using published nucleotide sequence data (8). Single-stranded template DNA was isolated from the phage essentially as described (24). The DNA sequence was determined by the dideoxy-chain-termination method (28) using both a universal M13 primer and specific t-PA primers.

RESULTS AND DISCUSSION

Isolation of a Cosmid Carrying the Human t-PA Gene. A cDNA clone (pPA01) containing sequences from the mRNA for human t-PA (26) was used to isolate the t-PA gene from a human genomic cosmid library (20, 21). This *in vivo* packaged cosmid library is divided into 12 pools. Southern blot analysis using digested DNA from these pools hybridized to

Abbreviations: PA, plasminogen activator; t-PA, tissue-type PA; kb, kilobase(s); bp, base pair(s).

*Present address: Department of Immunology, Scripps Clinic and Research Foundation, La Jolla, CA 92037.

the ^{32}P-labeled pPA01 revealed that the strongest hybridization signal was obtained from cosmid pool 7 (data not shown). This observation suggested that pool 7 had the highest proportion of cosmids carrying the t-PA gene. When 150,000 cosmids from this pool were examined for t-PA sequences by *in situ* hybridization, 15 cosmid clones were found to hybridize strongly to the pPA01 probe. Upon digestion with *Eco*RI and *Bam*HI, all 15 clones showed the same restriction pattern. Therefore, one cosmid, pcosPAU01, was arbitrarily selected for further analyses.

Structural Characterization of the t-PA Gene. The DNA of pcosPAU01 was initially characterized by restriction enzyme analysis with *Bam*HI, *Nru* I, and *Hpa* I, and a gross restriction map was established (Fig. 1). To further characterize the genomic DNA of pcosPAU01, five DNA fragments (outlined in Fig. 1) were subcloned into pBR322 (29). By restriction enzyme analysis of these subclones, a complete restriction map was constructed (Fig. 2). As determined by gel electrophoresis, the length of the chromosomal insert is about 30 kilobases (kb). To determine the location of t-PA sequences in pcosPAU01, Southern blot hybridizations were carried out with restriction fragments of cosmid DNA immobilized to nitrocellulose. The nick-translated cDNA of pPA01, which carries a 334-base-pair (bp) DNA sequence from the middle part of the t-PA gene (26), hybridized to a 2.8-kb *Eco*RI fragment, as outlined in Fig. 2. To determine the 5′ to 3′ orientation of the t-PA gene, specific oligodeoxyribonucleotide probes for the 5′ and 3′ ends were synthesized and used as hybridization probes. A 5′-specific probe, complementary to nucleotides 91–108 of the mRNA (8), hybridized to a 690-bp *Bam*HI/*Cla* I fragment, and a 3′-specific probe, complementary to nucleotides 2168–2185 of the mRNA, hybridized to a 440-bp *Bgl* II/*Eco*RI fragment (Fig. 2). The human genome contains >300,000 copies of related sequences, termed *Alu* I sequences (30), which are 300 bp long and for which no known function has been demonstrated. Using an *Alu* I family hybridization probe (31), we detected *Alu* I sequences at five different regions of the 30-kb t-PA gene region (Fig. 2).

Sequence Analysis of Exons and Flanking Regions. Restriction fragments containing exon sequences were first identified by Southern blot analysis. The hybridization probes used are outlined in Fig. 2. After this initial mapping procedure, many restriction sites present in the full-length cDNA of the t-PA gene (8) were identified in the genomic t-PA clone. To determine the DNA sequence, relevant restriction fragments containing exon DNA were isolated from agarose gels and subcloned into the M13 vectors mp8 or mp9 (24). With this approach, it was possible to determine the positions of the exons relative to known restriction sites in the t-PA gene (Fig. 2). Finally, the exact coding nucleotide sequence of the exons and their flanking regions of the intervening sequences were determined with reference to the known cDNA sequence (8) and the consensus sequence of the exon–intron boundary (32).

Sequence analysis of the 5′ region of the t-PA gene showed that the 5′ end and the first 58 nucleotides of the published cDNA sequence were missing. The genomic DNA sequence differed from the cDNA sequence upstream of nucleotide 58. The dinucleotide A-G found at this position most likely marks the 3′ border of the first intron of the t-PA gene. Using a 5′-specific probe, we isolated 20 new cosmids from the same library, but they all contained exactly the same genomic region as pcosPAU01 (data not shown). Thus, the transcription initiation site of the t-PA mRNA has not been determined.

At the 3′ end of the genomic sequence, the last exon encodes the entire 3′ trailer sequence of the mRNA as well as the coding sequence for the COOH-terminal amino acids (Figs. 2 and 3). It is not possible to define the exact limit of the last exon because the first three adenines of the poly(A) tail may have been transcribed from the genomic DNA (Fig. 3). About 30 bp upstream from the poly(A) attachment site the consensus polyadenylation sequence A-A-T-A-A-A is found (33).

The number of exons is tentative because the cosmid did not contain the first 58 nucleotides of the mRNA. Because exons of higher eukaryotes do not exhibit a marked variation in size (34), we have made the assumption that these 58 nontranslated nucleotides are encoded by exon I (Fig. 2). Consequently, the t-PA gene is divided into at least 14 exons by at least 13 introns and has a total length of >20 kb. All intron boundaries agree with the consensus sequence for such regions (32). Each intron begins with G-T at the 5′ terminus, and ends with A-G at the 3′ terminus.

The t-PA cDNA sequence previously determined (8) was aligned with the genomic sequence as shown in Fig. 3. The genomic sequence is consistent with the cDNA sequence, except for a few indicated substitutions. None of the substitutions affects the amino acid sequence of the protein, because they are either neutral changes at the third position of codons or they are located in the noncoding parts of the gene (Fig. 3).

Two lines of evidence support the conclusion that this is the only t-PA gene in the human genome and that the cosmid pcosPAU01 represents a continuous stretch of the human genome: (*i*) Southern blot hybridizations of restricted genomic DNA probed with the 334-bp cDNA clone revealed the hybrid fragments expected from the physical map of pcosPAU01 (Fig. 2). (*ii*) DNA sequences of the exons were identical to those of the cDNA, except for a few wobble base shifts. These base substitutions are not unexpected, because allelic variations normally are found.

Correlation of Exon Regions with Structural Units of the Protein. It has been postulated that exons represent genetic building blocks that code for discrete structural or functional domains of proteins (14, 15). To determine whether this is the case for the t-PA gene, we have indicated the positions of the introns in the t-PA molecule (Fig. 4). The mRNA of many eukaryotic viruses and some cellular proteins (e.g., ovalbumin) contain nontranslated RNA sequences at the 5′

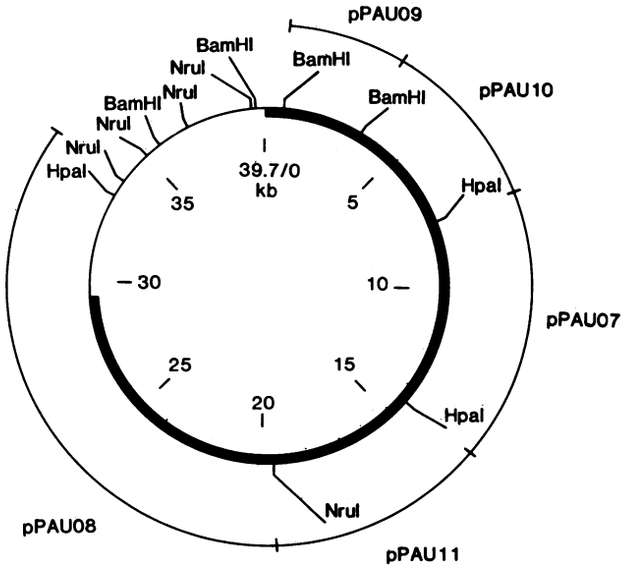

FIG. 1. Gross restriction map of cosmid pcosPAU01. The thin and thick lines of the circle represent the vector pHC79- 2cos/tK and genomic DNA, respectively. The positions of the recognition sites for the enzymes used are indicated. The restriction fragments pPAU07–pPAU11 that were subcloned into pBR322 for further analyses are outlined on the outer semicircle.

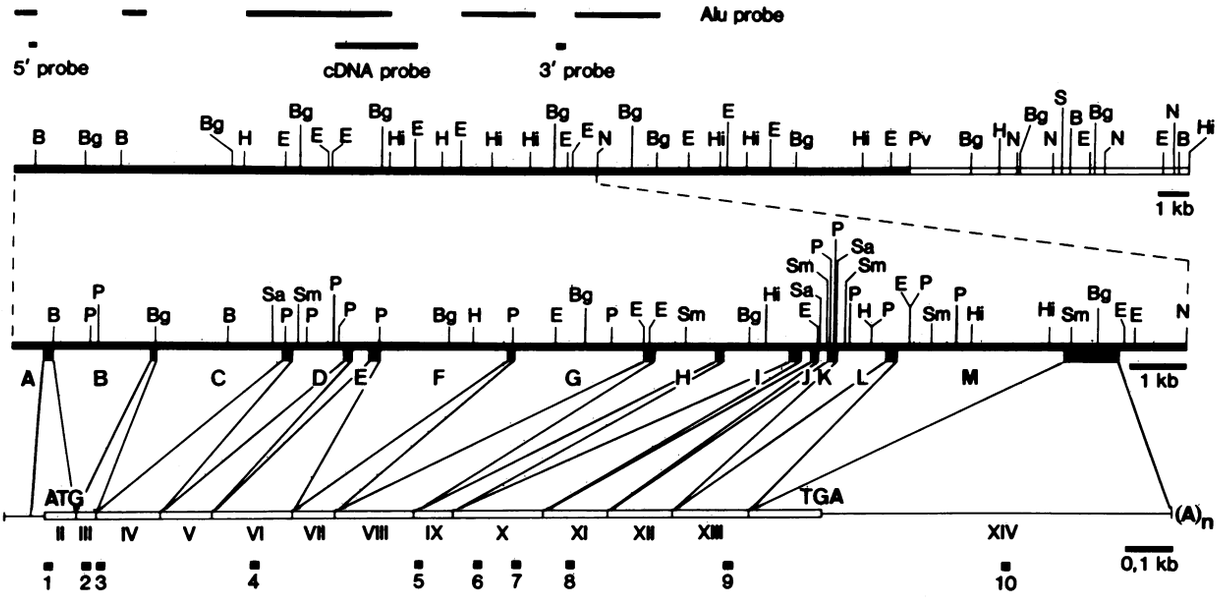

FIG. 2. Physical map of cosmid pcosPAU01 and organization of the human t-PA gene. A restriction map of the cosmid is shown on the top line. Thin and thick lines denote vector and inserted human DNA sequences, respectively. The DNA region encoding the t-PA gene is expanded on the second line. The locations of exons II–XIV (solid boxes) and introns A–L were determined by Southern blotting and DNA sequence analysis. The mRNA structure from Pennica *et al.* (8) is depicted on the third line, and the open box represents the coding nucleotide sequences. (A) indicates the poly(A) addition site of the mRNA. 5′ probe, 3′ probe, cDNA probe, and *Alu* probe were used to locate these regions in the cosmid DNA, and the result is outlined at the top of the figure. The map positions of the oligodeoxyribonucleotides 1–10 used for Southern blotting and sequence analysis are outlined at the bottom of the figure. The different restriction sites are indicated as follows: B, *Bam*HI; Bg, *Bgl* II; E, *Eco*RI; Hi, *Hin*dIII; H, *Hpa* I; N, *Nru* I; P, *Pst* I; Pv, *Pvu* I; Sa, *Sac* I; S, *Sal* I; Sm, *Sma* I. Restriction sites for *Sac* I, *Sma* I, and *Pst* I are not shown on the first line.

end that are encoded by DNA sequences physically separated from those coding for the protein (32). Our results indicate that the t-PA mRNA also has a so-called leader sequence (≧58 nucleotides long) and that this sequence is encoded by the first exon(s). Secreted proteins often contain signal peptide, which is important for their secretion (35). Although the exact size of the signal peptide in t-PA is not known, the signal peptidase likely cleaves at the carboxyl side of either serine number −15 or −13 (Fig. 4) (36). The signal peptide and the following one or three amino acids are encoded by exon II.

Most of the exon III of the t-PA gene codes for a "pro"-sequence-like structure similar to that found for serum albumin (37, 38). The extension of this "pro"-segment cannot be

FIG. 3. Nucleotide sequence of human t-PA gene and comparison with the cDNA sequence. Nucleotide sequences corresponding to 13 exons and their flanking regions in the intervening sequences are shown. The approximate length of introns A–M are shown in kb and intron sequences are underlined. The first 58 bases present at the 5′ end of the cDNA are missing in the genomic DNA. Possible poly(A) attachment sites were determined by comparison with the cDNA sequence (8); they are indicated by arrows. The consensus polyadenylylation signal A-A-T-A-A-A is marked. The genomic sequence is consistent with the cDNA sequence (8) except where indicated. Deviating nucleotides in the cDNA are shown below the genomic sequence enclosed in boxes. * indicates that a nucleotide is absent in one of the sequences in comparison to the other.

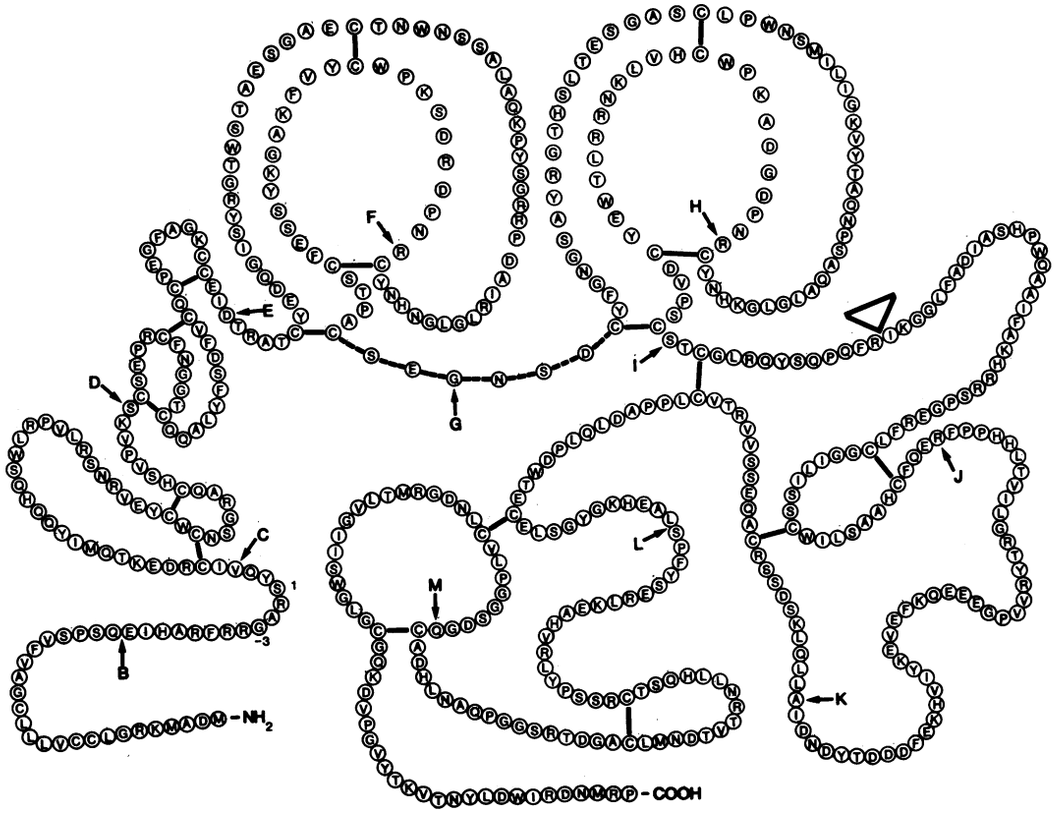

FIG. 4. Schematic two-dimensional model of the potential precursor t-PA protein including signal peptide and pro-sequence. The standard one-letter code for each of the amino acids is given in the open circles. The solid black bars indicate the potential disulfide bridges. The model is a combination of the models presented by Pennica *et al.* (8) and Bányai *et al.* (13). The arrows B–M indicate the map position of the individual introns in the protein. The triangle between intron positions I and J indicates the cleavage site between the heavy and the light chain. The serine designated as number 1 is suggested by Pennica *et al.* (8).

determined because of the NH_2-terminal heterogeneity of the mature t-PA protein (39). Sequence analysis of purified t-PA from melanoma cells has shown that the protein starts at the serine designated number 1 or at the glycine designated number −3 (Fig. 4) (7, 8). The "pro"-segment of serum albumin ends with the amino acids Phe-Arg-Arg (38). This same amino acid sequence immediately precedes the glycine in position −3 of the t-PA sequence (Fig. 4). Assuming that the longer form of t-PA is the cleaved product, these amino acids may be part of a recognition sequence for the enzyme system(s) that cleaves "pro"-sequences (40). The shorter form of t-PA may result from additional proteolytic digestion, presumably by a plasmin-like enzyme in the culture medium (39). In the t-PA gene, the signal peptide and the "pro"-segment are coded by separate exons, whereas both of these segments belong to the same exon in the serum albumin gene (41).

Fibronectin has been implicated in a variety of biological activities, most of which involve adhesive binding functions (42). The fibrin-affinity of fibronectin has been correlated to nine so-called "finger" domains (or type I homologies) (43). t-PA has a segment in the heavy chain that is homologous to these finger structures (13). The fourth exon of the t-PA gene codes exclusively for this "finger-like" domain (Fig. 4). The gene for chicken fibronectin contains at least 48 small exons. One of these has been sequenced (44), and the deduced protein sequence reveals type I homology. This exon and exon IV of the human t-PA are similar in size. Therefore, it is likely that the finger-like domains of t-PA and fibronectin have evolved from the same primordial gene.

A domain shared with growth factors has been found in human t-PA, human high molecular weight u-PA, and bovine clotting factor X (12, 13). Furthermore, homologies to this part of t-PA exist in human and mouse epidermal growth factor, bovine protein C, factor IX (12), bovine prothrombin, and rat transforming growth factor (45, 46). It is not known if this domain is coupled to a function shared by these proteins. In the t-PA gene, the growth-factor-like domain is encoded for exclusively by exon V, suggesting that this domain may have a specific function in t-PA as well as in the other proteins.

The heavy chain of t-PA contains two triple disulfide structures ("kringles"). Such structures have also been found in prothrombin (9, 10), plasminogen (11), and urokinase (12), and they are thought to be important for the binding of these proteins to fibrin (11, 47, 48). In the t-PA gene, introns demarcate each side of the kringles from the surrounding sequences. The two kringles also show an identical intron–exon pattern and are likely to have arisen from a common ancestor. Both introns F and H divide the coding regions for the kringles into two exons by cleaving an arginine codon (Fig. 4). Around these arginine residues there are 6-amino-acid-long homologies between the kringles of plasminogen, prothrombin, and urokinase (9–12). The existence of potential fibrin binding sites in both the finger and kringle regions of t-PA raises the question of whether both of these regions are important for the binding of t-PA to fibrin *in vivo*.

The light chain of t-PA, which contains the active site, is divided into five exons by four introns. The amino acid sequences of t-PA and other serine proteases exhibit a characteristic pattern of structurally conserved regions (49). Since the intron positions for chymotrypsin, trypsin, and elastase were known (50), we compared the map positions of the introns in the protein structure between these proteins and t-PA. All four genes have introns at or close to positions corresponding to the introns J, L, and M of the t-PA gene, but

only the rat elastase gene has an intron at the same position as intron K of the t-PA gene. The fact that there are not only conserved regions of homologous amino acids within this family of proteins, but also a conserved intron–exon structure of the genes coding for these proteins, supports the theory that the serine proteases belong to a gene family derived from a common ancestor (51).

It has been proposed that the positions of splice junctions correspond to positions of length variations between different members of the same gene family, and that these variations can be generated by "sliding" of the intron–exon junctions (50). A comparison of the gene and protein structure of t-PA with those of trypsin, chymotrypsin, and elastase reveals that introns J, K, and M map close to positions of amino acid insertions in the t-PA gene as compared to the other three serine proteases. Intron L maps at a position where the sequence of the different serine proteases varies. This observation reinforces the hypothesis that splice junctions are associated with variations in protein structure between different members of a gene family (50).

In summary, the t-PA molecule consists of several structural domains. Here we have shown that these domains are encoded by separate exons. Although structural domains have been found in many other proteins, it has been difficult to correlate a structure to a specific function. Access to a genomic clone for t-PA provides the opportunity of studying the individual functions of these protein domains by separately expressing them in bacterial systems.

We thank Assar Bäckman, Kerstin Enqvist, Carina Fredrikson, and Seija Järvinen for skillful technical assistance; Staffan Josephson for oligomer synthesis; David Loskutoff for valuable discussions; and Christopher Korch and Gerry Josephs for help with the manuscript. Financial support was obtained from Kabigen AB, The National Swedish Board for Technical Development (Dnr 825463), and Lion's Research Foundation, Umeå (project 242/82). Part of the work was performed while T.N. was supported by an European Molecular Biology Organization fellowship.

1. Christman, J. K., Silverstein, S. C. & Acs, G. (1977) in *Proteinases in Mammalian Cells and Tissues*, ed. Barret, A. J. (Elsevier, Amsterdam), p. 91.
2. Collen, D. (1980) *Thromb. Haemostasis* **43,** 77–89.
3. Williams, J. R. B. (1951) *Br. J. Exp. Pathol.* **32,** 530–539.
4. Astrup, T. & Permin, P. M. (1947) *Nature (London)* **159,** 681–682.
5. Rijken, D. C., Wijngaards, G. & Wellbergen, J. (1980) *Thromb. Res.* **18,** 815–830.
6. Rijken, D. C. & Collen, D. (1981) *J. Biol. Chem.* **256,** 7035–7041.
7. Wallén, P., Pohl, G., Bergsdorf, N., Rånby, M., Ny, T. & Jörnvall, H. (1983) *Eur. J. Biochem.* **132,** 681–686.
8. Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Benett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V. & Collen, D. (1983) *Nature (London)* **301,** 214–221.
9. Sottrup-Jensen, L., Zajdel, M., Claeys, H., Petersen, T. E. & Magnusson, S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 2577–2581.
10. Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Dudek-Wojciechowska, G. & Claeys, H. (1976) in *Proteolysis and Physiological Regulation*, eds. Ribbons, D. W. & Brew, K. (Academic, New York), pp. 203–212.
11. Sottrup-Jensen, L., Claeys, H., Zajdel, M., Petersen, T. E. & Magnusson, S. (1978) in *Progress in Chemical Fibrinolysis and Thrombolysis*, eds. Davidson, J. F., Rowan, R. M., Samama, M. M. & Desuoyers, P. C. (Raven, New York), Vol. 3, pp. 191–209.
12. Günzler, W. A., Steffens, G. J., Ötting, F., Kim, S.-M. A., Frankus, E. & Flohe, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem.* **363,** 1155–1165.
13. Bányai, L., Várdi, A. & Patthy, L. (1983) *FEBS Lett.* **163,** 37–41.
14. Gilbert, W. (1978) *Nature (London)* **271,** 501.
15. Blake, C. C. F. (1978) *Nature (London)* **273,** 267.
16. Quinto, C., Quiroga, M., Swain, W. F., Nikovits Jr., W. C., Strandring, D. N., Pictet, R. L., Valenzuela, P. & Rutter, W. J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 31–35.
17. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1522.
18. Grosveld, F. G., Dahl, H.-H. M., deBohr, E. & Flavell, R. A. (1981) *Gene* **13,** 227–237.
19. Lindenmaier, W., Hauser, H., Greiser de Wilke, I. & Schütz, G. (1982) *Nucleic Acids Res.* **10,** 1243–1256.
20. Lund, B., Edlund, T., Lindenmaier, W., Ny, T., Collins, J., Lundgren, E. & von Gabain, A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2435–2439.
21. Boyer, H. W. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41,** 459–472.
22. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
23. Moseley, S. L., Hug, I., Alim, A. R. M. A., So, M., Sanadpour-Motalebi, M. & Falkow, S. (1980) *J. Infect. Dis.* **142,** 892–898.
24. Messing, J. & Vieiea, J. (1982) *Gene* **19,** 269–276.
25. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
26. Edlund, T., Ny, T., Rånby, M., Heden, L.-O., Palm, G., Holmgren, E. & Josephson, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 349–352.
27. Josephson, S., Palm, G. & Lagerholm, E. (1984) *Acta Chem. Scand.*, in press.
28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5487.
29. Bolivar, F., Rodriquez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2,** 95–113.
30. Houck, C. M., Rinehart, F. P. & Schmid, C. W. (1979) *J. Mol. Biol.* **132,** 289–306.
31. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedman, T. & Schmid, C. W. (1980) *Nature (London)* **284,** 372–375.
32. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4853–4857.
33. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
34. Naora, H. & Deacon, N. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6196–6200.
35. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 852–862.
36. Austen, B. M. (1979) *FEBS Lett.* **103,** 308–312.
37. Lawn, R. M., Adelman, J., Bock, S. C., Franke, A. E., Houck, C. M., Najarian, R. C., Seeburg, P. H. & Wion, K. L. (1981) *Nucleic Acids Res.* **9,** 6103–6114.
38. Patterson, J. E. & Geller, D. M. (1977) *Biochem. Biophys. Res. Commun.* **74,** 1220–1226.
39. Jornvall, H., Pohl, G., Bergsdorf, N. & Wallen, P. (1983) *FEBS Lett.* **156,** 47–50.
40. Pradayrol, L., Jörnvall, H., Mull, V. & Ribet, A. (1980) *FEBS Lett.* **109,** 55–58.
41. Sargent, T. D., Jagodzinski, L. L., Yang, M. & Bonner, J. (1981) *Mol. Cell Biol.* **1,** 871–883.
42. Yamada, K. M. (1983) *Annu. Rev. Biochem.* **52,** 761–799.
43. Petersen, T. E., Thøgersen, H. C., Skorstensgaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. & Magnusson, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 137–141.
44. Hirano, H., Yamada, Y., Sullivan, M., DeCrombrugghe, B., Pastan, I. & Yamada, K. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 46–50.
45. Magnusson, S., Petersen, T. E., Sottrup-Jensen, L. & Claeys, H. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 123–149.
46. Marquardt, H., Hunkapiller, M. W., Hood, L. E., Twardzik, D. R., De Larco, J. E., Stephenson, J. R. & Todaro, G. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4484–4688.
47. Thorsen, S., Glas-Greenwalt, P. & Astrup, T. (1972) *Thromb. Diath. Haemorrh.* **28,** 65–74.
48. Thorsen, S. (1975) *Biochim. Biophys. Acta* **393,** 55–65.
49. Strasburger, W., Wollmer, A., Pitts, J. E., Glover, I. D., Tickle, I. J., Blundell, T. L., Steffens, G. J., Günzler, W. A., Ötting, F. & Flohé, L. (1983) *FEBS Lett.* **157,** 219–223.
50. Craik, C. S., Rutter, W. J. & Fletterick, R. (1983) *Science* **220,** 1125–1129.
51. Hartley, B. S. & Shotton, D. M. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), Vol. 3, pp. 323–373.