Tissue Plasminogen Activator: Peptide Analyses Confirm an Indirectly Derived Amino Acid Sequence, Identify the Active Site Serine Residue, Establish Glycosylation Sites, and Localize Variant Differences†

Gunnar Pohl,* Margareta Källström, Nils Bergsdorf, Per Wallén, and Hans Jörnvall

ABSTRACT: Tissue plasminogen activator, separated into variants I and II (differing in M_r by 2000–3000), was reduced and [^{14}C]carboxymethylated. Fragments from cleavages with enzymes and cyanogen bromide (CNBr) were separated by reverse-phase high-performance liquid chromatography and subjected to sequence degradations. All seven CNBr fragments were purified and found to be compatible with the cDNA-derived amino acid sequence [Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Ylverton, E., Seeburg, P. H., Heynecker, H. L., Goeddel, D. V., & Collen, D. (1983) *Nature (London) 301*, 214–221]. Chemical characterization of 93% of the 527 residues recovered in 50 peptides confirmed the indirectly deduced primary structure of the protein. The tryptic peptide patterns from the two variants were found to differ for one peptide (T15). Since carbohydrate was present in this peptide for variant I and since a marked difference in chromatographic behavior for T15 was observed in variant II, we conclude that carbohydrate differences in this peptide (i.e., Asn-184 in the numbering system of the cDNA-derived amino acid sequence) are the explanation for the size differences between variants I and II. Carbohydrate was also found at two other positions in the protein, corresponding to Asn-117 and Asn-448. However, a fourth potential glycosylation site, Asn-218, is apparently not utilized for carbohydrate attachment. The enzyme is inactivated by diisopropyl phosphorofluoridate, which covalently modifies the serine residue corresponding to position 478, identifying this as the active site serine residue. In addition, firm evidence was found that on plasmin treatment of the 527-residue chain, no additional bond is cleaved apart from that previously known at Arg-275–Ile-276.

The plasminogen activator produced and secreted from a melanoma cell line (Bowes) has been purified by two different methods using several chromatography steps, with (Wallén et al., 1983) or without (Rijken & Collen, 1981) an immunosorbent step. In both cases, the native enzyme obtained in the presence of protease inhibitors is a one-chain carbohydrate-containing molecule different from human urokinase but immunologically identical with tissue activator from human uterus (Rijken & Collen, 1981; Wallén et al., 1981, 1983; Rånby et al., 1982). However, the uterine form is now known to be susceptible to slightly different N-terminal processing (Pohl et al., 1984).

Plasmin treatment causes cleavage of the molecule into two disulfide-connected chains (A chain and B chain from N- and C-terminal parts, respectively). Diisopropyl phosphorofluoridate (DFP)[1] inhibits enzyme activity and is incorporated into the B chain (Wallén et al., 1983).

Bacterial clones containing activator cDNA sequences have been isolated (Pennica et al., 1983; Edlund et al., 1983). They were detected by hybridization against oligonucleotide probes synthesized to match suitable parts of protein regions initially determined by peptide analysis. From the cDNA data, an amino acid sequence of 527 residues has been deduced (Pennica et al., 1983).

Limited sequence analyses of the activator have shown N-terminal processing differences, giving S and L chains differing by the absence or presence of three extra residues (Wallén et al., 1983; Jörnvall et al., 1983a), as well as still shorter U chains (Pohl et al., 1984). A possible residue microheterogeneity apparently coupled with the S/L chain processing difference has also been observed (Jörnvall et al., 1983a), but other sequence analyses of the protein have not been directly reported. In the present report, the numbering system of the S chain is used in order to facilitate direct correlation with the numbers derived from the cDNA-deduced structure.

A further variation is type I and II activator molecules, which can be separated on Sepharose–arginine. A molecular weight difference of 3000 was apparent between the A chains from the two variants, but no size difference was found for corresponding B chains (Rånby et al., 1982).

In this work, we have studied the protein structure of separated A and B chains in order to obtain direct peptide data from all regions of the protein and to determine specific sites. Tryptic digestion and preparative peptide mapping by reverse-phase high-performance liquid chromatography (HPLC) made it possible to compare both fragmentation patterns and compositions from the two variants of A chains. A total of 50 fragments was isolated from the protein, and sequence degradations were performed by a sensitive manual technique or in a liquid-phase sequencer. The structures of the isolated peptides support the cDNA-derived amino acid sequence of the molecule. In addition, the active site and the glycosylated residues were identified. One peptide was isolated that differed between variants I and II, probably in carbohydrate composition. Finally, it was concluded that C-terminal cleavages in A or B chains do not occur on plasmin treatment.

Materials and Methods

Protein. Tissue plasminogen activator was purified from culture medium of a melanoma cell line (Bowes) and separated

† From the Department of Physiological Chemistry, Umeå University, S-901 87 Umeå, Sweden (G.P., N.B., and P.W.), and the Department of Chemistry I, Karolinska Institutet, S-104 01 Stockholm, Sweden (M.K. and H.J.). *Received November 29, 1983.* This work was supported by grants from the Swedish Medical Research Council (Projects 13X-3532 and 13X-3906), the Magn. Bergvall Foundation, the Swedish Cancer Society, the Knut and Alice Wallenberg Foundation, and Kabi Gen AB.

[1] Abbreviations: CNBr, cyanogen bromide; DFP, diisopropyl phosphorofluoridate; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid.

0006-2960/84/0423-3701$01.50/0

Table I: Total Compositions of CNBr Fragments from Separated A and B Chains of [^{14}C]Carboxymethylated Tissue Plasminogen Activator[a]

composition	A-CB1	A-CB2	A-CB3	sum A[d]	B-CB1	B-CB2	B-CB3	B-CB4	sum B[d]
Cys(Cm)	1.1 (1)	21.5 (18)	6.0 (5)	28.6 (24)	6.7 (7)	2.7 (3)	0.9 (1)		10.3 (11)
Asx	1.2 (1)	18.4 (18)	7.6 (7)	27.2 (26)	14.2 (13)	6.5 (6)	4.7 (5)	0.5	25.9 (24)
Thr	1.0 (1)	9.0 (9)	2.9 (3)	12.9 (13)	7.9 (8)	0.9 (1)	2.4 (3)		11.2 (12)
Ser	1.0 (1)	20.4 (24)	4.4 (4)	25.8 (29)	16.1 (16)	1.7 (2)	1.2 (1)	0.4	19.4 (19)
Glx	3.1 (3)	20.5 (19)	6.1 (6)	29.7 (28)	24.2 (21)	2.2 (2)	1.7 (1)	0.4	28.5 (24)
Pro		10.2 (10)	4.6 (5)	14.8 (15)	11.3 (10)	2.6 (2)	1.5 (1)	1.2 (1)	16.6 (14)
Gly	0.6	17.4 (17)	5.7 (5)	23.7 (22)	12.2 (9)	6.1 (7)	5.7 (5)	0.3	24.3 (21)
Ala		13.5 (13)	4.1 (4)	17.6 (17)	9.9 (12)	2.5 (3)			12.4 (15)
Val	0.8 (1)	5.3 (6)	3.2 (3)	9.3 (10)	9.2 (10)	1.1 (1)	2.7 (4)		13.0 (15)
Hsl[b]	+ (1)	+ (1)	-	+ (2)	+ (1)	+ (1)	+ (1)	-	+ (3)
Ile	0.7 (1)	4.8 (4)	1.9 (2)	7.4 (7)	9.3 (9)		2.1 (3)		11.4 (12)
Leu		9.9 (8)	5.5 (6)	15.4 (14)	17.4 (18)	3.1 (4)	2.8 (3)		23.3 (25)
Tyr	0.8 (1)	9.7 (11)	3.5 (4)	14.0 (16)	5.7 (6)		1.7 (2)		7.4 (8)
Phe		6.4 (7)	1.3 (1)	7.7 (8)	6.8 (8)				6.8 (8)
Trp[c]	nd (6)	nd (6)	nd (2)	nd (8)	nd (3)	nd	nd (2)	nd	nd (5)
Lys	1.2 (1)	6.3 (6)	3.5 (4)	11.0 (11)	7.2 (8)	0.3	1.9 (2)	0.4	9.4 (10)
His		4.1 (4)	1.8 (2)	5.9 (6)	7.0 (9)	0.9 (1)			7.9 (10)
Arg	1.0 (1)	12.8 (13)	4.8 (5)	18.6 (19)	12.7 (12)	2.0 (2)	0.7 (1)	0.8 (1)	16.2 (16)
sum	(13)	(194)	(68)	(275)	(180)	(35)	(35)	(2)	(252)

[a] Peptides are listed in the order they occur in the chains and analytical values in molar ratios are shown with cDNA-derived values within parentheses. Results are from acid hydrolysis without correction for destruction, incomplete liberations, or impurities (except for B-CB2 and B-CB3, which are corrected for cross-contamination). Values lower than 0.3 are omitted. [b] Methionine is shown by the presence (+) or absence (-) or homoserine/homoserine lactone. [c] Tryptophan was not determined (nd). [d] The sums of residues from peptides of the A and B chains respectively.

into variants I and II as described (Wallén et al., 1983; Rånby et al., 1982). Plasmin treatment, active site labeling with [^{3}H]DFP, reduction, and [^{14}C]carboxymethylation were performed (Wallén et al., 1983), and the A chain was separated from the B chain on Sepharose–arginine (Rånby et al., 1982).

Peptide Nomenclature. The CNBr fragments from A and B chains are designated A-CB and B-CB and given numbers corresponding to the order in which they appear in the sequence. T and E denote peptides obtained by digestion with trypsin and Glu-specific protease, respectively. The fragments identified are numbered T1–T36 and E1–E7, starting at the N-terminal end of the A chain.

Proteolytic Treatments. The separated variants of A and B chains (0.5–1 mg/mL) were treated with CNBr (20–100 mg/mL) in 70% formic acid for 24 h at room temperature. Reagents were removed by evaporation. L-1-(Tosylamido)-2-phenylethyl chloromethyl ketone treated trypsin (Worthington) and staphylococcal Glu-specific protease (Miles) were used for enzymatic treatment. Protein (0.5–1.5 mg/mL) was digested in 0.1 M ammonium bicarbonate, at an enzyme to protein ratio of 1:10. After incubation for 5 h at 37 °C, reactions were stopped by freeze–drying. However, to initially dissolve the B chain, 8 M urea was used and diluted with ammonium bicarbonate before subsequent digestion, which was performed in 0.8 M urea and 0.1 M ammonium bicarbonate.

Peptide Purification. The lyophilized mixture of CNBr fragments was dissolved in 1.5 mL of 30% acetic acid and fractionated on a 1.6 × 150 cm TSK 55F column (Merck, West Germany) equilibrated in the same buffer. This eluate was monitored at 280 nm on an ISCO UA 5 instrument, and the radioactivity was measured on 50–100-μL samples in a Packard scintillation counter. The lyophilized digests were then dissolved in 0.1% trifluoroacetic acid and fractionated by reverse-phase HPLC by using a Waters system with gradient elution, a μBondapak C_{18} column, and a variable-wavelength detector. The column was equilibrated with 0.1% trifluoroacetic acid (sequencer grade, Fluka, Buchs, Switzerland) before injection of about 250 μL of sample, and the peptides were eluted with a linear gradient of acetonitrile (HPLC grade S, Rathburn Chemical, Great Britain). Absorbance was monitored at 215 nm at a flow rate of 1.5 mL/min.

Sequence Analysis. Amino acid sequences were determined by liquid-phase sequencer degradations (Beckman 890D) in the presence of 0.1 M Quadrol and precycled Polybrene (Jörnvall & Philipson, 1980) or by manual degradations with the [(dimethylamino)azo]benzene–isothiocyanate method (Chang et al., 1978) utilizing byproducts to assist identifications (von Bahr-Lindström et al., 1982). Phenylthiohydantoins from sequencer degradations were quantitated on a Hewlett-Packard high-performance liquid chromatograph (1084 B) with gradient elution (Zimmerman et al., 1977) and a Nucleosil C_{18} column (Macherey-Nagel, Düren, West Germany). Radioactivity of the thiohydantoins from both manual and sequencer-assisted degradations was measured in an LKB-Wallac scintillation counter. Amino acid compositions were determined with a Beckman 121M analyzer, after hydrolysis with 5.7 M HCl–0.5% phenol in evacuated tubes for 24 h at 110 °C.

Results

Preparation of A and B Chains. Native one-chain tissue plasminogen activator variants I and II (Rånby et al., 1982) were separated, cleaved with plasmin, inactivated with [^{3}H]DFP, reduced, [^{14}C]carboxymethylated, and finally separated into the A and B chains. These chains were the starting material for the structural studies, which were based on characterization of all CNBr fragments, most tryptic peptides, and some Glu-specific protease peptides, as detailed below.

CNBr Fragments. The three CNBr fragments from the A chain (A-CB1, A-CB2, and A-CB3) were separated by exclusion chromatography. The overlap between the first two fragments has already been established by direct sequence analysis of the intact protein (Wallén et al., 1983), and the overlap between the last two CNBr fragments was now proven by isolation of a tryptic peptide (T16; cf. Figure 1). Total compositions are given in Table I, and results from sequence analyses are summarized in Figure 1.

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis of separated and CNBr-treated variants of A chains (thus A-CB1, A-CB2, and A-CB3 from variants I and II) showed that the apparent size difference between the variants resides in fragment A-CB2. However, no significant difference in total

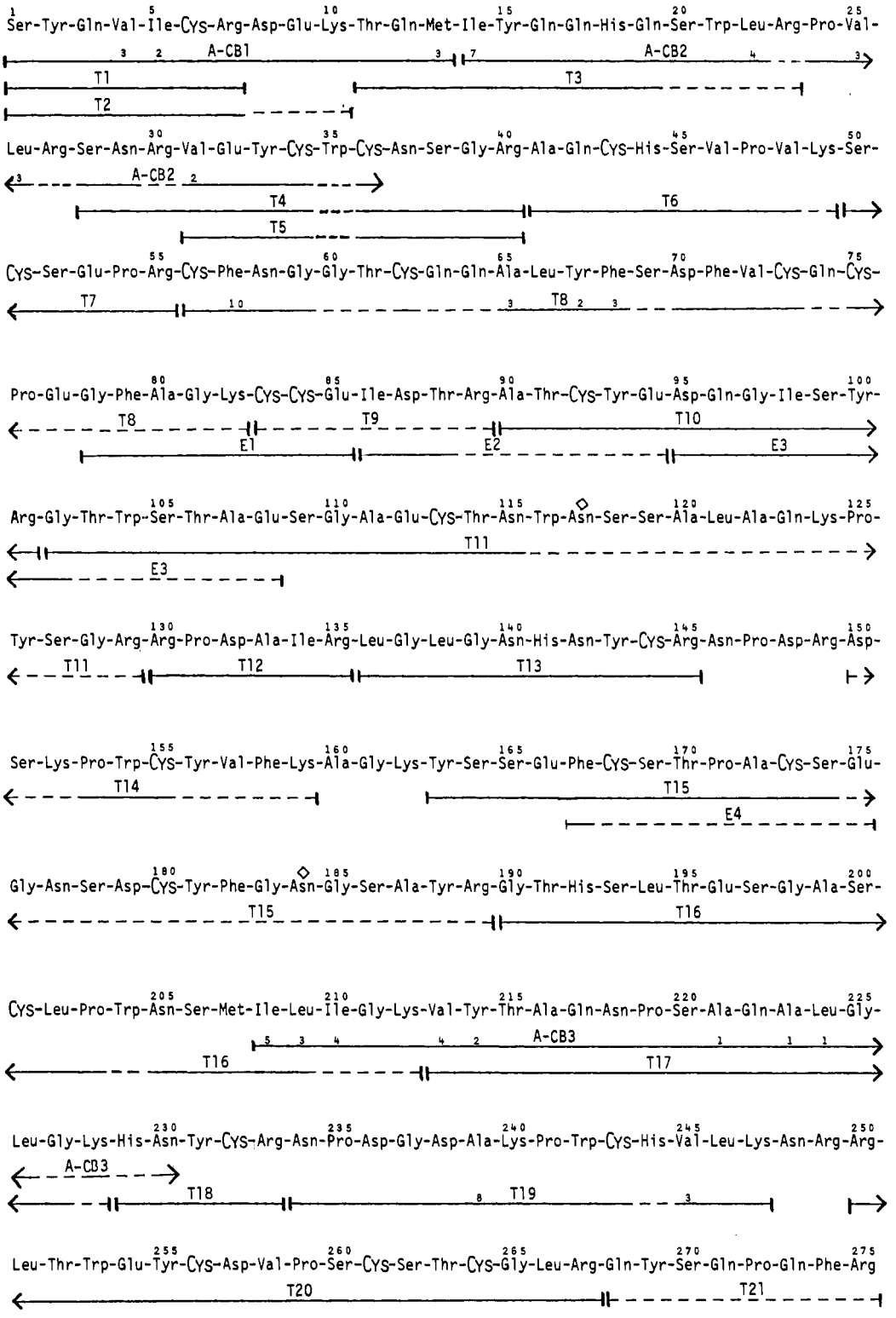

FIGURE 1: Amino acid sequence of the A chain (previously deduced from cDNA) and positions of the constituent peptides analyzed. Solid lines indicate regions of peptides proven by sequence analysis, and dashed lines show regions compatible with total compositions. Small numbers on the solid line show recoveries in nanomoles for those peptides that were degraded in the sequencer. Potential glycosylation sites are marked with (◇) where aminohexoses were detected in the corresponding peptide. The cysteine residues of the molecule were analyzed as (carboxymethyl)cysteine and are shown in capital letters. Details about the sequence analyses and peptide nomenclature are given under Materials and Methods.

composition of this CNBr fragment from the two variants was detectable, suggesting that the difference is not derived from differences in the primary structure of the polypeptide part.

CNBr fragments from the B chain (B-CB1, B-CB2, B-CB3, and B-CB4) were fractionated by exclusion chromatography and reverse-phase HPLC. The identification of the two terminal fragments was established by direct analyses (B-CB1 yielding an amino acid sequence identical with that of the N-terminal part of the whole B chain and B-CB4 lacking homoserine), whereas B-CB2 and B-CB3 were initially given a tentative order on the basis of the homology with other serine proteases. This order was verified when the cDNA sequence became available (Pennica et al., 1983). As with the A-chain fragments, no difference in total composition (Table I) between

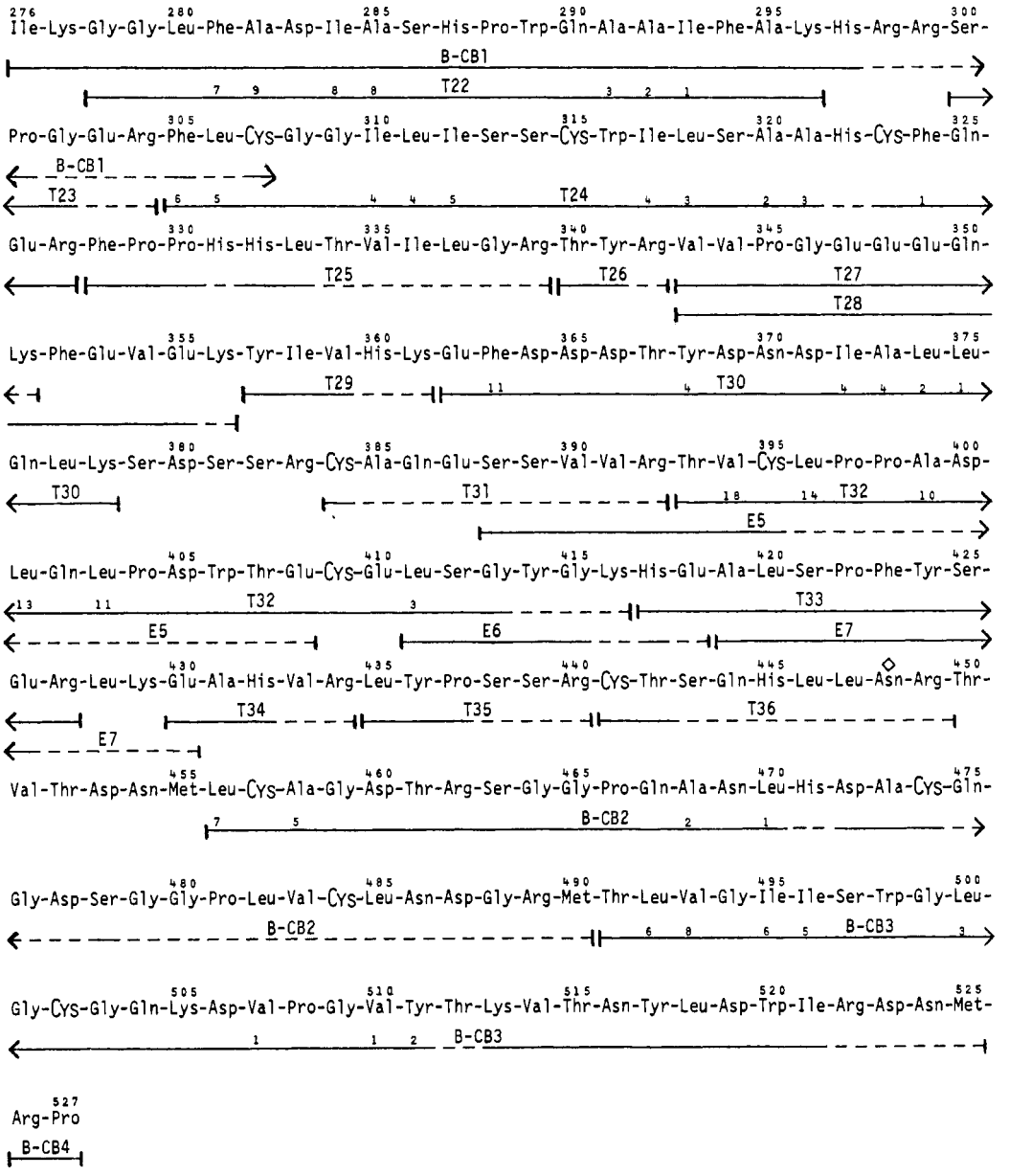

FIGURE 2: The cDNA-derived amino acid sequence of the B chain and positions of the constituent peptides. Solid lines and nomenclature are as in Figure 1.

fragments from variants I and II was found. The big fragment B-CB1 was pure from exclusion chromatography and was redigested with trypsin, while the other fragments were obtained in mixture and then further purified by HPLC. Results from sequence analyses are summarized in Figure 2.

Active Site Serine Residue. Radioactivity from [^{3}H]DFP in preparations specifically labeled at the active site serine residue was found in fragment B-CB2, and the radioactivity was recovered in cycle 23 upon sequencer degradations of that fragment. This result gives positive identification of the active site serine residue as the amino acid corresponding to position 478 in the numbering system of the cDNA-derived protein structure.

Tryptic Fragments. The A chains from the separated variants (about 45 nmol) were digested with trypsin. The fragments were separated by reverse-phase HPLC, giving well-resolved peptide maps of the two variants (Figure 3). All fractions corresponding to the individual absorbance peaks were collected and subjected to structural analyses.

Comparisons of total compositions of 27 fractions from each of the two variants showed that corresponding fractions in most cases were indistinguishable. These fractions were pooled and analyzed by sequence degradations. Results are summarized in Figure 1, and the total compositions are given in Table II. However, in two cases (corresponding to the fractions containing peptide T15; Figure 3) from the two variants, compositions were found to differ between corresponding peaks, and these fractions were therefore degraded separately. Results showed that peptide T15 was responsible for the differences, contaminating peptide T19 in variant I whereas for variant II it coeluted with T11 (cf. Figure 3). The only difference observed between the tryptic digests of variants I and II was this change in chromatographic behavior for T15. The structures of T11 and T19 were separately analyzed in pure form (T11 from variant I and T19 from variant II) and fully confirmed by analyses of the mixtures with T15. Peptide T15 itself was determined from the other sequence in each mixture and then gave identical results from each variant. Also, the total compositions of T15, calculated after corrections for the presence of T11 and T19, respectively, were indistinguishable between the two variants. Thus, the amino acid sequence of this peptide is deduced to be as shown in Figure 1 and to be

Table II: Total Compositions of Tryptic Peptides from the A Chain[a]

composition	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Cys(Cm)	1.1 (1)	1.7 (1)	0.7	2.3 (2)	2.2 (2)	0.8 (1)	0.8 (1)	3.5 (4)	2.0 (2)	1.2 (1)	1.2 (1)
Asx		0.9 (1)	0.9	1.8 (2)	0.9 (1)			2.2 (2)	1.0 (1)	1.1 (1)	2.1 (2)
Thr			1.3 (1)					1.2 (1)	0.9 (1)	0.8 (1)	2.1 (3)
Ser	1.0 (1)	1.1 (1)	1.8 (1)	2.0 (2)	1.0 (1)	1.0 (1)	1.8 (2)	1.6 (1)	0.7	1.0 (1)	3.8 (5)
Glx	0.9 (1)	1.7 (2)	3.9 (4)	1.2 (1)	1.2 (1)	1.2 (1)	1.1 (1)	3.8 (4)	1.4 (1)	2.2 (2)	2.9 (3)
Pro						1.0 (1)	1.0 (1)	1.3 (1)			1.2 (1)
Gly			0.9	1.1 (1)	1.1 (1)			3.8 (4)	0.4	1.1 (1)	3.1 (3)
Ala			0.8			1.0 (1)		1.9 (2)		0.9 (1)	2.8 (4)
Val	1.2 (1)	1.7 (1)	0.5	0.9 (1)	0.9 (1)	1.9 (2)		1.0 (1)			0.4
Met			0.7 (1)								
Ile	0.6 (1)	0.7 (1)	0.7 (1)						0.7 (1)	0.7 (1)	
Leu			0.5 (1)					1.1 (1)			1.4 (1)
Tyr	0.6 (1)	0.8 (1)	1.2 (1)	0.9 (1)	0.9 (1)			0.9 (1)		1.8 (2)	1.2 (1)
Phe								3.0 (4)			
Trp	nd	nd	nd (1)	nd (1)	nd (1)	nd	nd	nd	nd	nd	nd (2)
Lys		0.9 (1)	0.4			1.1 (1)		1.5 (1)			1.1 (1)
His			1.0 (1)			1.1 (1)					
Arg	0.9 (1)	1.0 (1)	0.5 (1)	1.7 (2)	0.8 (1)		1.3 (1)		0.8 (1)	0.8 (1)	1.1 (1)
sum	(7)	(10)	(13)	(13)	(10)	(9)	(6)	(27)	(7)	(12)	(28)

composition	T12	T13	T14	T15	T16	T17	T18	T19	T20	T21
Cys(Cm)		1.0 (1)	1.0 (1)	2.7 (3)	0.7 (1)		1.1 (1)	0.9 (1)	3.4 (3)	
Asx	1.0 (1)	2.0 (2)	1.3 (1)	2.8 (3)	1.5 (1)	1.2 (1)	1.1 (1)	2.8 (3)	1.1 (1)	
Thr				1.6 (1)	1.7 (2)	0.9 (1)			1.5 (2)	
Ser			1.0 (1)	6.0 (6)	3.5 (4)	1.1 (1)			1.7 (2)	1.0 (1)
Glx		0.5		3.1 (2)	1.3 (1)	2.0 (2)			1.1 (1)	3.7 (3)
Pro	1.2 (1)		1.1 (1)	0.9 (1)	1.0 (1)	1.1 (1)		1.8 (2)	0.9 (1)	1.0 (1)
Gly		1.9 (2)		3.1 (3)	3.2 (3)	1.9 (2)		1.1 (1)	1.0 (1)	
Ala	1.0 (1)			2.2 (2)	1.3 (1)	2.7 (3)		1.1 (1)		
Val			1.0 (1)		0.4	1.0 (1)		0.9 (1)	0.9 (1)	
Met					0.5 (1)					
Ile	0.9 (1)				1.6 (2)					
Leu		1.7 (2)			2.7 (3)	1.7 (2)		1.5 (1)	2.1 (2)	
Tyr		1.1 (1)	0.8 (1)	3.0 (3)		0.9 (1)	1.0 (1)		0.8 (1)	0.6 (1)
Phe			0.7 (1)	2.0 (2)						0.8 (1)
Trp	nd	nd	nd (1)	nd	nd (1)	nd	nd	nd (1)	nd (1)	
Lys			2.2 (2)		1.2 (1)	1.0 (1)		2.0 (2)		
His		1.0 (1)			0.8 (1)		0.9 (1)	0.9 (1)		
Arg	2.1 (2)	1.0 (1)		1.1 (1)			0.9 (1)		1.4 (2)	0.9 (1)
sum	(6)	(10)	(10)	(27)	(23)	(16)	(5)	(14)	(18)	(8)

[a] Order of peptides and analytical results are given as in Table I, with T1, T2, and T21, T4 and T5, and T15 corrected for cross-contaminations (Figure 3 and text) and with tryptophan not determined (nd).

identical between the two variants.

In three other cases, peptides were also not resolved in the chromatography, as shown in Figure 3, but identical mixtures were then obtained from both variants.

Methodologically, an apparent N-terminal cyclization of glutamine in position 1 of T21 was noticed since this peptide gave weak results on sequence analysis in a mixture with T1/T2, although composition showed all three fragments to be present in equal yield. Also noticeable, lack of proline in the total composition of peptide T3 indicated unusual tryptic cleavage of an Arg–Pro bond (positions 23–24; cf. Figure 1) whereas the prolyl bonds at potential tryptic cleavage sites in T11, T14, and T19 were resistant as traditionally expected.

Tryptic peptides from the B chain were obtained by digestion of fragment B-CB1 from the two variants, and since the chromatograms were identical (Figure 4, numbers of the identified peptides are given as in Figure 2), corresponding fractions were collected, pooled, and analyzed in a mixture. Results from peptide analyses are summarized in Figure 2, with total compositions given in Tables I and III.

Fragments Obtained by Cleavage with Glu-Specific, Staphylococcal Protease. Separated variants of CNBr fragment A-CB2 were treated with this enzyme, and the peptides were fractionated in the same way as the tryptic fragments. The chromatograms gave highly similar patterns for the two variants, and the positions of four fragments were identified in the structure by manual sequence analyses (Figure 1) and determinations of total compositions (Table III).

The protease was also used for digestion of intact B chain (unseparated variants) and three fragments were similarly identified (Figure 2). The composition of E7 (Table III) indicated that its Glu–Arg bond (positions 426–427 in Figure 2) was not cleaved by the enzyme.

Carbohydrate-Containing Fragments. Peptide T15, which is the peptide that differs in elution position between the two variants when tryptic digests are separated on reverse-phase HPLC (Figure 3) contains the sequence Asn–X–Ser, known to be a signal for N-linked glycosylation. In the case of variant I, T15 is obtained in a mixture with T19 (above), and the total composition shows the presence of carbohydrate in this fraction. The pure T19 is obtained from variant II and does not contain aminohexoses or any sites for N-glycosylation. We therefore conclude that T15 is glycosylated in variant I. From the variant II digest, the corresponding peptide coelutes with T11, which also contains carbohydrate, and has an Asn–X–Ser sequence. Consequently, the carbohydrate content of T15 in variant II could not be established. However, the difference in chromatographic behavior of T15 when obtained from the two variants of the molecule, the carbohydrate content of T15 in variant I, and the indistinguishable results on amino acid analysis of T15 from variants I and II (above) strongly suggest that the apparent molecular weight difference between the variants is due to differences in glycosylation of peptide T15. This is in agreement with a short report elsewhere (Bennett,

Table III: Total Compositions of Tryptic Peptides from the Largest CNBr Fragment of the B Chain (B-CB1) and of Peptides Obtained by Digestion with Glu-Specific Protease of A-CB2 (E1–E4) and Intact B Chain (E5–E7)[a]

composition	T22	T23	T24	T25	T26	T27	T28	T29	T30	T31	T32
Cys(Cm)	0.6		2.6 (3)							0.9 (1)	2.0 (2)
Asx	1.0 (1)	0.5	0.4	0.6					4.8 (6)		1.9 (2)
Thr	0.4			1.2 (1)	1.0 (1)				1.2 (1)		1.6 (2)
Ser	1.1 (1)	1.1 (1)	2.7 (3)	0.4					0.9	1.7 (2)	1.1 (1)
Glx	1.3 (1)	1.1 (1)	2.4 (2)	0.4		4.1 (4)	6.2 (6)		1.9 (2)	2.1 (2)	2.9 (3)
Pro	1.0 (1)	0.7 (1)		1.7 (2)		1.1 (1)	1.1 (1)				2.9 (3)
Gly	1.9 (2)	1.4 (1)	2.4 (2)	1.3 (1)		1.2 (1)	1.2 (1)		1.1		2.2 (2)
Ala	2.8 (5)		2.1 (2)						1.3 (1)	1.1 (1)	1.4 (1)
Val				0.6 (1)		1.6 (2)	2.5 (3)	1.0 (1)	0.6	1.9 (2)	1.0 (1)
Met											
Ile	1.3 (2)		2.9 (3)	0.4 (1)				1.0 (1)	1.2 (1)		
Leu	1.7 (1)		3.0 (3)	1.6 (2)					2.8 (3)		3.5 (4)
Tyr					1.0 (1)			1.0 (1)	1.0 (1)		0.8 (1)
Phe	1.2 (2)		1.7 (2)	0.6 (1)			1.2 (1)		0.6 (1)		
Trp	nd (1)	nd	nd (1)	nd	nd	nd	nd	nd	nd	nd	nd (1)
Lys	0.8 (1)					0.9 (1)	1.8 (2)	1.1 (1)	1.0 (1)		1.0 (1)
His	0.6 (1)		1.1 (1)	1.4 (2)				1.1 (1)			
Arg		0.5 (1)	0.8 (1)	0.7 (1)	0.8 (1)					0.8 (1)	
sum	(19)	(5)	(23)	(12)	(3)	(9)	(14)	(5)	(17)	(9)	(24)

composition	T33	T34	T35	T36	E1	E2	E3	E4	E5	E6	E7
Cys(Cm)				1.0 (1)	1.5 (2)	1.0 (1)		1.6 (2)	(1)		
Asx				1.1 (1)	0.6	1.3 (1)	1.6 (1)	0.5	2.0 (2)		0.4
Thr				0.9 (1)		1.7 (2)	1.6 (2)	0.8 (1)	1.8 (2)		
Ser	1.7 (2)		1.7 (2)	1.0 (1)		0.4	2.0 (2)	1.8 (2)	1.8 (2)	1.0 (1)	1.0 (2)
Glx	2.0 (2)	1.1 (1)		1.7 (1)	1.1 (1)	1.2 (1)	1.8 (2)	0.9 (1)	2.9 (2)	1.1 (1)	2.1 (2)
Pro	1.0 (1)		1.1 (1)					1.3 (1)	2.6 (3)		1.0 (1)
Gly					1.9 (2)		2.0 (2)	0.3	0.6	2.0 (2)	0.4
Ala	1.1 (1)	1.1 (1)			0.7 (1)	1.0 (1)	1.1 (1)	0.6 (1)	1.2 (1)		1.0 (1)
Val		1.0 (1)		0.4			0.4		2.1 (3)		
Met											
Ile						1.0 (1)	0.5 (1)				
Leu	1.0 (1)		0.9 (1)	2.0 (2)			0.5		2.6 (3)	1.0 (1)	1.9 (2)
Tyr	0.9 (1)		1.0 (1)			1.1 (1)	1.0 (1)			1.0 (1)	1.0 (1)
Phe	1.0 (1)				1.0 (1)			0.9 (1)			0.9 (1)
Trp	nd	nd	nd	nd	nd	nd	nd (1)	nd	nd (1)	nd	nd
Lys					1.0 (1)					1.0 (1)	0.9 (1)
His	0.9 (1)	1.0 (1)		1.0 (1)					0.5	1.0 (1)	
Arg	0.7 (1)	0.8 (1)	0.8 (1)	0.8 (1)		1.0 (1)	1.3 (1)		1.0 (1)		1.0 (1)
sum	(11)	(5)	(6)	(9)	(8)	(9)	(14)	(9)	(21)	(8)	(12)

[a] Order of peptides and analytical values are given as in Tables I and II.

1983). Presence of aminohexoses was also apparent for T36; thus, three of the four potential N-glycosylation sites in the molecule were shown to be glycosylated. The fourth site of the type Asn–X–Ser/Thr is found in peptide T17, but no carbohydrate was detected in the composition of this fragment. In addition, sequential degradations of T17 (and A-CB3 containing the same structure, Figure 1) clearly identified intact asparagine at the critical position (corresponding to position 218 in Figure 1), which therefore is deduced not to be glycosylated.

Discussion

Peptide analyses of tissue plasminogen activator have been carried out on a small scale. Essentially, 50-nmol starting aliquots have been used for generation and analysis of complete digests of this large protein. The high resolving power of HPLC plus sensitive and rapid manual methods for sequence analysis were the prerequisites for this work. In this way, it is possible to obtain a high degree of structural information from small amounts of protein. This approach can also be powerful in searching for sequences suitable for synthesis of oligodeoxynucleotide probes intended for hybridizations in cloning experiments, as demonstrated both for this protein and other proteins where complete digests were similarly separated (Jörnvall et al., 1983b). All protein fragments analyzed can then be used for correlation with cDNA structures when clones have been investigated.

A total of 93% of the cDNA-derived amino acid sequence (Pennica et al., 1983) was now covered and was fully confirmed by peptide analyses. In addition, total compositions of all CNBr fragments were compatible with the entire cDNA-deduced structure. Apart from the sequence data of the peptides, we also obtained information about the glycosylation sites and other special regions in the molecule.

The asparagine residues potentially sensitive to N-glycosylation must be part of the sequence Asn–X–Ser/Thr [reviewed in Sharon & Lis (1982)]. Two of the four potential sites, Asn-117 and Asn-448, were now found glycosylated in both variants I and II of tissue plasminogen activator, whereas one site (Asn-218) apparently is not. This is the only site of the four where X is proline, and it is known that proline in that position may inhibit glycosylation (Marshall, 1974). The tryptic peptide maps from HPLC show that the two variants are very similar. The only difference observed was a difference in elution for one peptide. This peptide contains the fourth potential N-glycosylation site in the molecule (Asn-184), and it is concluded that different glycosylation in this peptide is responsible for the apparent size difference between variants I and II of melanoma tissue plasminogen activator. Peptide T8 appeared as two peaks in the HPLC chromatograms (same for both variants). Total compositions showed incomplete carboxymethylation and traces of aminohexoses in one of the two fractions. An additional carbohydrate attachment (in that case presumably O-glycosylation) in this region can therefore

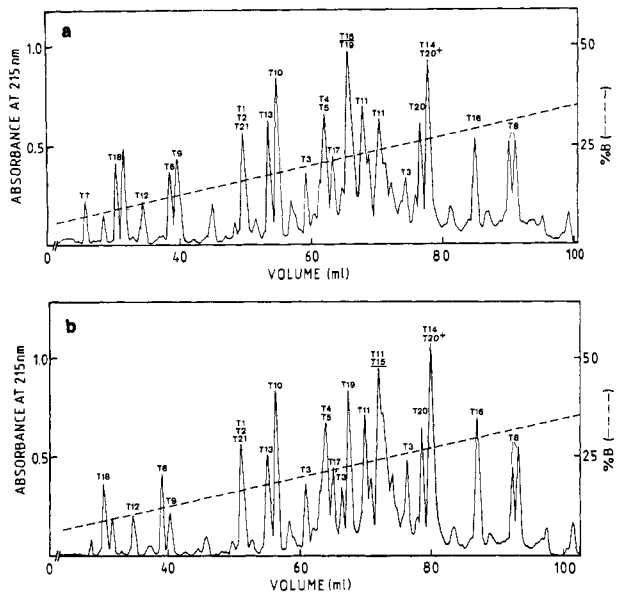

FIGURE 3: Separation of peptides after tryptic cleavages of A chains from variant I (a) and variant II (b) on a reverse-phase HPLC column. About 45 nmol of digested polypeptide was injected. Solvent A was 0.1% TFA and solvent B was acetonitrile. Fractions corresponding to the individual absorbance peaks were collected and analyzed from each of the variants (details under Materials and Methods). The peptides that were identified in the fractions are shown by numbers referring to the structure given in Figure 1. $T20^+$ differs from T20 by the lack of N-terminal Arg, and T15, which eluted differently in the two chromatograms, is underlined.

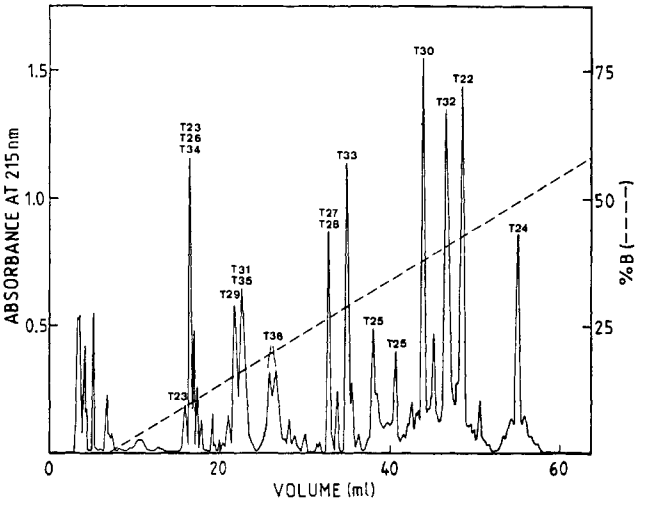

FIGURE 4: Separation of peptides after tryptic digestion of CNBr fragment B-CB1 with a reverse-phase HPLC column. Fractions corresponding to the individual absorbance peaks were collected and analyzed as detailed under Materials and Methods. Solvent A was 0.1% TFA and solvent B was acetonitrile. Identical chromatograms were obtained when tryptic digests of B-CB1 from the separated variants were chromatographed. The numbers of the identified peptides refer to Figure 2.

not be fully excluded, but no evidence for differences between variants I and II was found for peptide T8. Another tryptic peptide (T3) was recovered in two or three different fractions (cf. Figure 3). This heterogeneity may be due to desamidations, since the peptide contains three glutamine residues (Figure 1).

The C-terminal part of the A chain was now identified (T21), and we therefore conclude that plasmin does not cause any additional cleavage in the C-terminal part of the A chain, which has previously not been fully established. Thus, the molecular weight of one-chain [S-form; cf. Jörnvall et al. (1983a)] and two-chain tissue plasminogen activator is identical (59 000 for the polypeptide part). Finally, we have identified the active site serine residue of the enzyme by incorporation of [^{3}H]DFP and have found that the label is attached to residue 23 of CNBr fragment B-CB2, which corresponds to residue 478 of the whole structure (Pennica et al., 1983). In summary, apart from differential N-terminal processings previously detected (Wallén et al., 1983; Jörnvall et al., 1983a; Pohl et al., 1984), the protein studies fully confirm the cDNA-derived amino acid sequence and give direct identifications of glycosylation sites, a differential peptide for two variants of tissue plasminogen activator, an active site serine residue, and a single plasmin cleavage site.

Acknowledgments

The skillful technical assistance by Carina Palmberg, Rut Nyström, and Maria Swärd is gratefully acknowledged.

Registry No. Plasminogen, 9001-91-6; L-serine, 56-45-1.

References

Bennett, W. F. (1983) *Thromb. Haemostasis 50*, 106, Abstr. 315.

Chang, J.Y., Brauer, D., & Wittmann-Liebold, B. (1978) *FEBS Lett. 93*, 205–214.

Edlund, T., Ny, T., Rånby, M., Hedén, L.-O., Palm, G., Holmgren, E., & Josephson, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 349–352.

Jörnvall, H., & Philipson, L. (1980) *Eur. J. Biochem. 104*, 237–247.

Jörnvall, H., Pohl, G., Bergsdorf, N., & Wallén, P. (1983a) *FEBS Lett. 156*, 47–50.

Jörnvall, H., Kalkkinen, N., Luka, J., Kaiser, R., Carlquist, M., & von Bahr-Lindström, H. (1983b) in *Modern Methods in Analytical Protein Chemistry* (Tschesche, H., Ed.) pp 1–19, de Gruyter, Berlin.

Marshall, R. D. (1974) *Biochem. Soc. Symp. 40*, 17–26.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Ylverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., & Collen, D. (1983) *Nature (London) 301*, 214–221.

Pohl, G., Kaplan, L., Einarsson, M., Wallén, P., & Jörnvall, H. (1984) *FEBS Lett. 168*, 29–32.

Rånby, M., Bergsdorf, N., Pohl, G., & Wallén, P. (1982) *FEBS Lett. 146*, 289–292.

Rijken, D. C., & Collen, D. (1981) *J. Biol. Chem. 256*, 7035–7041.

Sharon, N., & Lis, H. (1982) *Proteins 5*, 1–144.

von Bahr-Lindström, H., Hempel, J., & Jörnvall, H. (1982) *J. Protein Chem. 1*, 257–261.

Wallén, P., Rånby, M., Bergsdorf, N., & Kok, P. (1981) *Prog. Chem. Fibrinolysis Thrombolysis 5*, 16–23.

Wallén, P., Pohl, G., Bergsdorf, N., Rånby, M., Ny, T., & Jörnvall, H. (1983) *Eur. J. Biochem. 132*, 681–686.

Zimmerman, C. L., Appella, E., & Pisano, J. J. (1977) *Anal. Biochem. 77*, 569–573.