Journal of Chromatography, 463 (1989) 375–396
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 077

STUDY OF THE PRIMARY STRUCTURE OF RECOMBINANT TISSUE PLASMINOGEN ACTIVATOR BY REVERSED-PHASE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC TRYPTIC MAPPING

R. C. CHLOUPEK*, R. J. HARRIS, C. K. LEONARD, R. G. KECK, B. A. KEYT, M. W. SPELLMAN, A. J. S. JONES and W. S. HANCOCK

Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080 (U.S.A.)

(Received September 6th, 1988)

SUMMARY

Two high-resolution tryptic maps have been developed for recombinant tissue plasminogen activator (rt-PA) that separate the expected 51 tryptic peptides. The trypsin digestion was performed after reduction and S-carboxymethylation of the protein. The high-performance liquid chromatographic separation of the tryptic peptides used a Nova-Pak C_{18} (5 μm) column with a mobile phase that contained 0.1% aqueous trifluoroacetic acid (TFA) or 50 m*M* sodium phosphate (pH 2.85) and a linear gradient of acetonitrile. A TFA solvent system was also used for re-purification and for characterization of the peptides isolated from the phosphate-based separation. All of the isolated peptides had compositions consistent with the sequence proposed for rt-PA.

The identities of the glycopeptides were confirmed by lectin chromatography on concanavalin A-Sepharose. The mixture of tryptic peptides was also treated with *endo*-β-N-acetylglucosaminidase H and peptide:N-glycosidase F to locate the position of either high mannose or complex oligosaccharides. These studies demonstrated that a high mannose oligosaccharide is attached to Asn-117 while complex carbohydrate side-chains are attached to Asn-184 and Asn-448. The residue Asn-184 is the site of optional glycosylation that results in the formation of two rt-PA variants that contain either two or three oligosaccharides.

INTRODUCTION

Reversed-phase high-performance liquid chromatography (RP-HPLC) has been shown to give a rapid and highly selective separation of tryptic digests of proteins[1] and has been applied to the characterization of several recombinant proteins such as methionyl-human growth hormone[2], bovine growth hormone[3] and interleukin[4]. This analytical technique has become a significant part of quality control procedures for the manufacture of protein pharmaceuticals produced by recombinant-DNA technology[5].

The application of tryptic mapping procedures to recombinant DNA-derived

0021-9673/89/$03.50

tissue plasminogen activator (rt-PA) produced in Chinese hamster ovary cell culture[6] represents a greater challenge. The previous tryptic map applications involved non-glycosylated proteins of molecular weights of up to 22 000 daltons, while rt-PA is a glycoprotein with a polypeptide molecular weight of 59 042 daltons[7] containing glycosylation variants that have either three or two carbohydrate side chains (type I and type II, respectively)[8,9]. This report describes the characterization of the primary structure of rt-PA utilizing two tryptic mapping methods. The tryptic peptides were found to be consistent with the primary sequence derived from the cDNA sequence of rt-PA[7]. The glycopeptides present in the tryptic mixture were further studied by lectin affinity chromatography and by digestion with glycosidases.

EXPERIMENTAL

Materials

HPLC/spectro-grade trifluoroacetic acid (TFA) (Pierce, Rockford, IL, U.S.A.), UV-grade acetonitrile (Burdick & Jackson, Muskegon, MI, U.S.A.), Sephadex G-75 superfine (Pharmacia, Piscataway, NJ, U.S.A.), α-methyl mannoside (Aldrich, Milwaukee, WI, U.S.A.), analytical-grade reagents and distilled, deionized Milli-Q™ water (Millipore, Bedford, MA, U.S.A.) were used. Concanavalin A (Con A) Sepharose, DL-dithiothreitol (DTT) and iodoacetic acid were obtained from Sigma (St. Louis, MO, U.S.A.). The enzymes used were TPCK (L-1-tosylamide-2-phenylethyl chloromethyl ketone) treated trypsin (Cooper Biomedical), peptide:N-glycosidase F (PNGase F) and *endo*-β-N-acetylglucosaminidase H (Endo H) (Genzyme, Boston, MA, U.S.A.). Recombinant tissue plasminogen activator was purified from mammalian cell culture (Chinese hamster ovary cell supernatants)[6] and reconstituted to a concentration of 1 mg/ml.

Sample preparation

A 5 mg-amount of rt-PA was reduced and S-carboxymethylated (RCM) according to a modified method based on Crestfield *et al.*[10]. The rt-PA was first dialyzed into the 8 *M* urea reaction buffer, treated for 4 h at ambient temperature with 10 m*M* DTT, followed by 25 m*M* iodoacetic acid. The resultant RCM rt-PA was then dialyzed into 0.1 *M* ammonium bicarbonate using dialysis tubing with a 3500 molecular weight cut-off.

The two domains of plasmin-treated RCM rt-PA (amino acid residues 1–275 and 276–527) were separated by gel filtration on a Sephadex G-75 superfine column (100 × 1.5 cm I.D.) in 0.1 *M* ammonium bicarbonate as described by Vehar *et al.*[11].

Tryptic digestion

RCM rt-PA was digested in 0.1 *M* ammonium bicarbonate at ambient temperature with an addition of TPCK-trypsin at an enzyme-to-substrate ratio of 1:100 (w/w), followed by a second addition of 1:100 after 8 h. The digestions were stopped after 24 h by freezing (−70°C).

Chromatography

The HPLC separations, unless otherwise indicated, were performed on a Waters gradient liquid chromatograph that included two 6000A pumps, a 720 controller,

a WISP 710B injector and a Nova-Pak 15 × 0.39 cm I.D., 5 μm, C_{18} reversed-phase column. The elution profile was monitored by combined LKB (2138 UVicord) and Perkin-Elmer (LC75) spectrophotometers for dual-wavelength detection at 214 and 280 nm, respectively. The HPLC separations of glycopeptides were performed on a Hewlett-Packard 1090M liquid chromatograph.

TFA tryptic map

The TFA solvent system used 0.1% aqueous TFA with a linear gradient of 0.08% TFA in acetronitile at a rate of 0.5%/min for 50 min, followed by a 1.0%/min linear gradient for 35 min at a flow-rate of 1.0 ml/min. Tryptic peptide peaks from RCM rt-PA 1–275 and RCM rt-PA 276–527 digests were collected manually from semipreparative separations using the TFA solvent system. Approximately 6.5 nmol of starting material was loaded for each separation.

Sodium phosphate tryptic map

The sodium phosphate solvent system used 50 m*M* sodium phosphate, pH 2.85 with a linear gradient of acetonitrile at a rate of 0.3%/min for 90 min, followed by a 1.0%/min linear gradient for 30 min at a flow-rate of 1.0 ml/min.

Tryptic peptide peaks from the RCM rt-PA digest were manually collected from three semi-preparative separations using the sodium phosphate solvent system totaling 18 nmol of starting material. The peptide peaks were vacuum concentrated and the total sample was rechromatographed in the volatile TFA solvent system using 0.1% TFA with a rapid linear gradient of acentonitrile containing 0.08% TFA at a rate of 10%/min to 60% acetonitrile, followed by isocratic elution at 60% acetonitrile for 20 min.

Peak analysis

Peptide peaks collected from both HPLC solvent systems were characterized by either amino acid analysis after acid hydrolysis or N-terminal sequencing with a Beckman 890C sequencer. Hydrolysis was performed by incubation of the peptides in constant boiling hydrochloric acid for 20 h at 110°C *in vacuo* for the trifluoroacetic acid tryptic map peptide peaks, followed by analysis on an LKB 4400 amino acid analyzer. Phosphate tryptic map peptides were hydrolyzed by the batch method of Waters Assoc. before Pico Tag™amino acid analysis. Amino acids were determined as their phenylthiocarbamyl (PTC) derivatives which were prepared by reaction with phenylisothiocyanate (PITC) at pH 9.5. The PTC-amino acids were chromatographed using an RP-HPLC system with detection at 254 nm. In both cases, identification and quantitation were performed by area comparison with an external standard. Edman degradation methodology was used to sequence peptides for identification in the case of ambiguous amino acid compositional data.

Carbohydrate-containing peptides

Lectin chromatography. Tryptic glycopeptides were isolated by lectin chromatography using a Con A-Sepharose column with a bed volume of 5 ml, equilibrated in 100 m*M* ammonium bicarbonate containing 1 *M* sodium chloride, 1 m*M* calcium chloride and 0.02% sodium azide. Trypsin-digested, RCM rt-PA (10 mg) was applied to the column in 100 m*M* ammonium bicarbonate containing 1 m*M* calcium chloride.

The absorbance at 280 nm was monitored and 0.5-ml fractions were collected. The column was washed with the equilibration buffer until the absorbance at 280 nm returned to baseline. Tryptic glycopeptides were eluted with 100 m*M* ammonium bicarbonate containing 0.2 *M* α-methyl mannoside and 0.02% sodium azide.

Treatment with peptide:N-glycosidase F. A 0.4-mg aliquot of RCM rt-PA was reconstituted in 0.08 ml of 250 m*M* sodium phosphate (pH 8.6) containing 10 m*M* EDTA and 0.02% sodium azide. PNGase F (0.5 manufacturer's units) in 0.018 ml of 50% glycerol was added and the sample was incubated overnight at 37°C. The sample was then diluted to 0.4 ml with water and dialyzed against 100 m*M* ammonium bicarbonate prior to trypsin digestion as described above.

Treatment with endo-β-N-acetylglucosaminidase H. A 0.4-mg aliquot of RCM rt-PA was reconstituted in 0.08 ml of 200 m*M* arginine phosphate (pH 6) containing 0.03 m*M* Polysorbate 80 and 0.02% sodium azide. The sample was treated with Endo H (0.04 manufacturer's units) in 0.02 ml of 50 m*M* sodium phosphate (pH 6) and incubated overnight at 37°C. The sample was then diluted to 0.4 ml with water, dialyzed against 100 m*M* ammonium bicarbonate, and digested with trypsin as described above.

RESULTS

TFA tryptic map

The tryptic map of RCM rt-PA obtained using TFA as the ionic modifier is shown in Fig. 1. The theoretical tryptic peptides are designated by "T"and numerically ordered from amino-terminal (T1) to the carboxy-terminal peptide (T51) (Table I). When a peptide elutes in more than one position in the tryptic map, the minor peak is

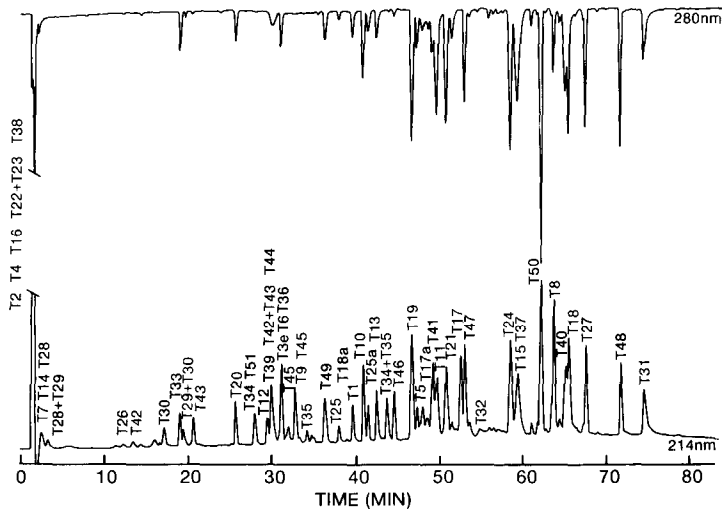

Fig. 1.TFA-based tryptic map of RCM rt-PA. This separation was performed on a 5-μm Nova-Pak C_{18} column (15 × 0.46 cm I.D.). Mobile phase A consisted of aqueous 0.1% TFA and mobile phase B was acetonitrile with 0.08% TFA. A linear gradient of 0–25% mobile phase B was run in 50 min followed by 25–60% mobile phase B in 35 min. The 200-μg sample was loaded in 0.2 ml of 0.1 *M* ammonium bicarbonate and monitored at both 214 and 280 nm.

TABLE I

THEORETICAL TRYPTIC PEPTIDES OF rt-PA

Tryptic Peptide	*Residue No.*		*Sequence**
	Start	*End*	
T1	1	7	SYQVICR
T2	8	10	DEK
T3	11	27	TQMIYQQHQSWLRPVLR
T4	28	30	SNR
T5	31	40	VEYCWCNSGR
T6	41	49	AQCHSVPVK
T7	50	55	SCSEPR
T8	56	82	CFNGGTCQQALYFSDFVCQCPEGFAGK
T9	83	89	CCEIDTR
T10	90	101	ATCYEDQGISYR
T11	102	129	GTWSTAESGAECTNWNSSALAQKPYSGR
T12	130	135	RPDAIR
T13	136	145	LGLGNHNYCR
T14	146	149	NPDR
T15	150	159	DSKPWCYVFK
T16	160	162	AGK
T17	163	189	YSSEFCSTPACSEGNSDCYFGNGSAYR
T18	190	212	GTHSLTESGASCLPWNSMILIGK
T19	213	228	VYTAQNPSAQALGLGK
T20	229	233	HNYCR
T21	234	247	NPDGDAKPWCHVLK
T22	248	249	NR
T23	250	250	R
T24	251	267	LTWEYCDVPSCSTCGLR
T25	268	275	QYSQPQFR
T26	276	277	IK
T27	278	296	GGLFADIASHPWQAAIFAK
T28	297	298	HR
T29	299	299	R
T30	300	304	SPGER
T31	305	327	FLCGGILISSCWILSAAHCFQER
T32	328	339	FPPHHLTVILGR
T33	340	342	TYR
T34	343	351	VVPGEEEQK
T35	352	356	FEVEK
T36	357	361	YIVHK
T37	362	378	EFDDDTYDNDIALLQLK
T38	379	383	SDSSR
T39	384	392	CAQESSVVR
T40	393	416	TVCLPPADLQLPDWTECELSGYGK
T41	417	427	HEALSPFYSER
T42	428	429	LK
T43	430	434	EAHVR
T44	435	440	LYPSSR
T45	441	449	CTSQHLLNR
T46	450	462	TVTDNMLCAGDTR
T47	463	489	SGGPQANLHDACQGDSGGPLVCLNDGR
T48	490	505	MTLVGIISWGLGCGQK
T49	506	513	DVPGVYTK
T50	514	522	VTNYLDWIR
T51	523	527	DNMRP

* Single-letter symbols for amino acids: A = Ala; C = Cys; D = Asp; E = Glu; F = Phe; G = Gly; H = His; I = Ile; K = Lys; L = Leu; M = Met; N = Asn; P = Pro; Q = Gln; R = Arg; S = Ser; T = Thr; V = Val; W = Trp; Y = Tyr.

denoted by the letter of the alphabet in order of relative recovery (T17, T17a, T17b, for example). A peptide generated by an incomplete cleavage is designated by a plus sign, *e.g.*, T29 + T30. The elution profile was monitored at two different wavelengths (214 and 280 nm) and each of the peptides that contained tyrosine and/or tryptophan was observed in the 280-nm profile. Despite the complexity of the mixture of tryptic peptides, the high resolution separation shown in Fig. 1 would suggest minimal co-elutions. This was confirmed by examination of the tryptic maps of the separated domains: RCM 1–275 and RCM 276–527 (Fig. 2).

Peaks were collected for amino acid analysis from both the total TFA tryptic map and the maps of the separated domains. Table II shows the peptide compositions of the peaks collected from the separated domains. If the amino acid analysis was ambiguous, such as in cases of co-elutions, the sample was also sequenced by Edman degradation to establish the identity. The relative contribution of each peptide to the total amino acid analysis of co-eluting peptides was calculated and peptide compositions were then reported separately (for example, T39, T42 + T43 and T44 at 29 min, Fig. 2). In those cases where the observed composition was very close to the deduced

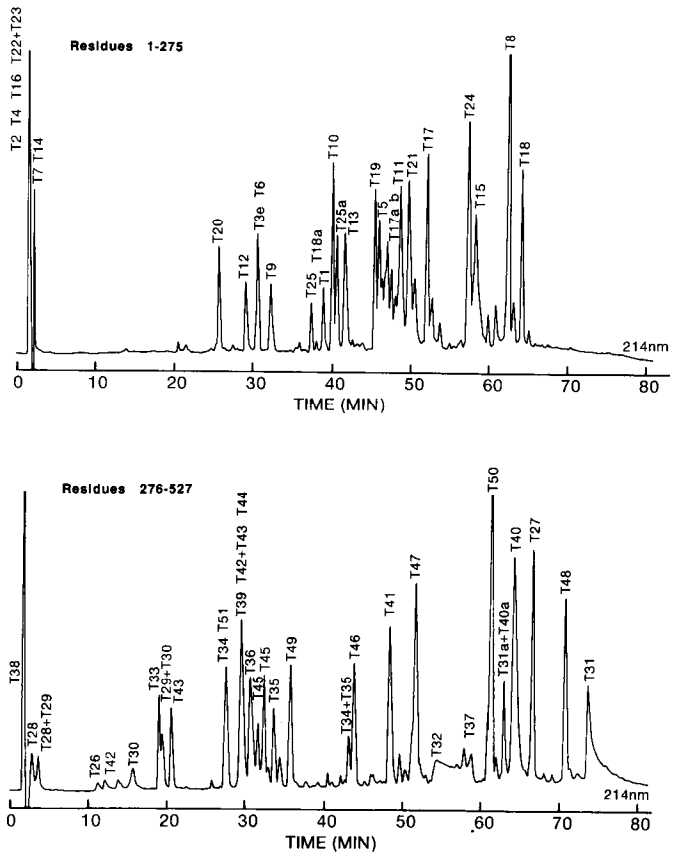

Fig. 2. TFA-based tryptic map comparison of the two separated domains of RCM rt-PA. The 200-μg sample loads were chromatographed as described in Fig. 1.

TABLE II

AMINO ACID COMPOSITIONS OF TFA TRYPTIC PEPTIDE PEAKS
(s) = N-Terminal sequencing; (□) = theoretical values.

	T1	*T3d*	*T5*	*T6(s)*	*T7(s)*	*T8*
Cmc	0.3 (1)	0.0	1.5 (2)	0.6 (1)	0.2 (1)	2.0 (4)
Asx	0.3	0.0	1.1 (1)	0.0	0.0	2.1 (2)
Thr	0.0	0.0	0.0	0.0	0.0	0.9 (1)
Ser	1.1 (1)	0.0	1.0 (1)	0.9 (1)	2.0 (2)	1.0 (1)
Glx	1.2 (1)	0.0	1.0 (1)	1.1 (1)	1.2 (1)	4.0 (4)
Pro	0.0	1.0 (1)	0.0	0.9 (1)	0.8 (1)	1.1 (1)
Gly	0.0	0.0	1.1 (1)	0.0	0.0	4.0 (4)
Ala	0.2	0.0	0.2	1.0 (1)	0.0	2.0 (2)
Cys	0.2	0.0	0.0	0.0	0.0	1.6
Val	0.7 (1)	1.0 (1)	0.8 (1)	2.1 (2)	0.0	1.0 (1)
Met	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.9 (1)	0.0	0.0	0.0	0.0	0.0
Leu	0.1	1.0 (1)	0.1	0.0	0.0	1.1 (1)
Tyr	1.0 (1)	0.0	0.9 (1)	0.0	0.0	1.0 (1)
Phe	0.0	0.0	0.0	0.0	0.0	4.0 (4)
Lys	0.1	0.0	0.0	1.1 (1)	0.0	1.0 (1)
His	0.0	0.0	0.0	1.0 (1)	0.0	0.0
Arg	1.1 (1)	1.0 (1)	1.0 (1)	0.0	1.1 (1)	0.0

	T9	*T10*	*T11*	*T12*	*T13*	*T14(s)*
Cmc	1.3 (2)	0.8 (1)	0.6 (1)	0.0	0.6 (1)	0.0
Asx	1.0 (1)	1.3 (1)	2.0 (2)	1.0 (1)	2.0 (2)	1.5 (2)
Thr	0.9 (1)	1.0 (1)	2.5 (3)	0.0	0.1	0.0
Ser	0.0	1.5 (1)	4.1 (5)	0.1	0.1	0.0
Glx	1.1 (1)	2.2 (2)	3.4 (3)	0.1	0.4	0.0
Pro	0.0	0.0	1.1 (1)	0.9 (1)	0.1	0.8 (1)
Gly	0.1	1.4 (1)	3.0 (3)	0.0	2.0 (2)	0.0
Ala	0.0	1.2 (1)	3.7 (4)	1.0 (1)	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Val	0.0	0.1	0.2	0.0	0.0	0.0
Met	0.6	0.0	0.1	0.0	0.0	0.0
Ile	1.0 (1)	1.0 (1)	0.2	0.9 (1)	0.0	0.0
Leu	0.0	0.2	1.2 (1)	0.0	2.1 (2)	0.0
Tyr	0.0	1.9 (2)	1.5 (1)	0.0	0.9 (1)	0.0
Phe	0.0	0.1	0.2	0.0	0.1	0.0
Lys	0.0	0.4	1.4 (1)	0.0	0.1	0.0
His	0.0	0.0	0.2	0.0	0.9 (1)	0.0
Arg	1.0 (1)	1.0 (1)	1.3 (1)	2.1 (2)	1.0 (1)	1.1 (1)

	T15	*T17*	*T17a*	*T18*	*T18a(s)*
Cmc	0.7 (1)	1.7 (3)	2.6 (3)	1.3 (1)	0.2
Asx	1.0 (1)	2.8 (3)	3.5 (3)	1.2 (1)	0.4
Thr	0.3	1.2 (1)	0.8 (1)	2.0 (2)	0.0
Ser	1.4 (1)	5.1 (6)	5.1 (6)	3.6 (4)	0.1
Glx	0.4	3.5 (2)	2.6 (2)	1.1 (1)	0.0
Pro	1.0 (1)	0.8 (1)	1.1 (1)	0.8 (1)	0.0
Gly	0.4	2.8 (3)	3.2 (3)	3.1 (3)	1.0 (1)
Ala	0.2	1.8 (2)	1.9 (2)	1.0 (1)	0.0

(*Continued on p. 382*)

TABLE II (*continued*)

	T15	*T17*	*T17a*	*T18*	*T18a(s)*	
Cys	0.0	0.0	0.0	0.0	0.0	
Val	0.9 (1)	0.0	0.5	0.0	0.0	
Met	0.1	0.0	0.0	0.3 (1)	0.0	
Ile	0.1	0.0	0.0	1.9 (2)	2.0 (2)	
Leu	0.5	0.0	0.1	3.1 (3)	1.1 (1)	
Tyr	0.9 (1)	2.8 (3)	2.9 (3)	0.0	0.0	
Phe	1.0 (1)	1.7 (2)	1.4 (2)	0.1	0.0	
Lys	2.0 (2)	0.0	0.3	1.1 (1)	0.9 (1)	
His	0.2	0.0	0.0	1.0 (1)	0.0	
Arg	0.1	1.3 (1)	1.6 (1)	0.0	0.0	
	T19	*T20*	*T21*	*T24*	*T25(s)*	*T25a*
Cmc	0.1	0.8 (1)	0.7 (1)	2.5 (3)	0.2	0.1
Asx	1.1 (1)	1.1 (1)	2.7 (3)	1.1 (1)	0.7	0.2
Thr	1.0 (1)	0.1	0.1	1.8 (2)	0.5	0.1
Ser	0.9 (1)	0.0	0.3	1.8 (2)	1.4 (1)	1.0 (1)
Glx	2.1 (2)	0.1	0.2	1.0 (1)	3.0 (3)	3.1 (3)
Pro	0.9 (1)	0.0	1.7 (2)	1.1 (1)	1.1 (1)	0.9 (1)
Gly	2.1 (2)	0.1	1.1 (1)	1.2 (1)	0.8	0.1
Ala	2.9 (3)	0.0	1.2 (1)	0.1	0.9	0.1
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Val	0.8 (1)	0.0	0.8 (1)	0.9 (1)	0.3	0.0
Met	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.0	0.1	0.0	0.0	0.2	0.2
Leu	2.1 (2)	0.0	1.0 (1)	2.0 (2)	0.5	0.1
Tyr	0.9 (1)	0.8 (1)	0.0	0.9 (1)	0.8 (1)	1.0 (1)
Phe	0.1	0.1	0.0	0.2	0.9 (1)	1.0 (1)
Lys	1.0 (1)	0.1	1.9 (2)	0.3	0.2	0.0
His	0.0	1.0 (1)	0.9 (1)	0.0	0.0	0.0
Arg	0.0	1.1 (1)	0.0	1.2 (1)	1.3 (1)	0.1
	T26	*T27(s)*	*T28(s)*	*T28 + T29(s)*	*T29 + T30(s)*	*T30*
Cmc	0.0	0.1	0.0	0.0	0.0	0.0
Asx	0.0	1.0 (1)	0.1	0.0	0.0	0.0
Thr	0.0	0.1	0.0	0.0	0.1	0.0
Ser	0.0	0.9 (1)	0.1	0.1	1.0 (1)	0.9 (1)
Glx	0.0	1.1 (1)	0.0	0.1	0.9 (1)	1.0 (1)
Pro	0.0	1.0 (1)	0.0	0.0	0.9 (1)	1.0 (1)
Gly	0.0	2.2 (2)	0.0	0.1	0.9 (1)	1.0 (1)
Ala	0.0	4.7 (5)	0.0	0.0	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Val	0.0	0.2	0.0	0.0	0.0	0.0
Met	0.0	0.1	0.0	0.0	0.0	0.0
Ile	1.0 (1)	2.3 (2)	0.0	0.0	0.1	0.0
Leu	0.1	1.2 (1)	0.0	0.0	0.0	0.0
Tyr	0.0	0.1	0.0	0.0	0.4	0.0
Phe	0.0	1.9 (2)	0.0	0.0	0.0	0.0
Lys	1.0 (1)	1.5 (1)	0.0	0.0	0.0	0.0
His	0.0	1.0 (1)	0.9 (1)	1.0 (1)	0.0	0.0
Arg	0.0	0.1	1.1 (1)	2.0 (2)	2.2 (2)	1.0 (1)

TABLE II (*continued*)

	T31	*T31a+T40a (s)*	*T32*	*T33*	*T34*	*T35*
Cmc	2.5 (3)	1.8 (4)	0.0	0.0	0.0	0.0
Asx	0.4	2.6 (2)	1.0	0.0	0.0	0.0
Thr	0.4	1.8 (1)	1.1 (1)	1.0 (1)	0.0	0.1
Ser	2.7 (3)	2.6 (3)	0.6	0.0	0.0	0.1
Glx	2.1 (2)	4.2 (5)	0.5	0.0	3.6 (4)	1.0 (2)
Pro	0.0	2.6 (3)	1.9 (2)	0.0	1.0 (1)	0.1
Gly	2.9 (2)	2.8 (2)	1.8 (1)	0.0	1.0 (1)	0.1
Ala	2.3 (2)	3.4 (3)	0.7	0.0	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Val	0.2	0.0	1.0 (1)	0.1	1.4 (2)	1.0 (1)
Met	0.1	0.0	0.0	0.0	0.0	0.0
Ile	3.1 (3)	1.1 (1)	0.9 (1)	0.0	0.0	0.1
Leu	3.3 (3)	3.8 (5)	2.3 (2)	0.0	0.0	0.2
Tyr	0.0	0.6 (1)	0.2	1.0 (1)	0.0	0.0
Phe	1.9 (2)	1.2 (1)	1.1 (1)	0.0	0.0	1.0 (1)
Lys	0.3	1.2 (1)	0.3	0.1	1.0 (1)	1.0 (1)
His	1.0 (1)	0.9 (1)	1.9 (2)	0.1	0.0	0.0
Arg	1.0 (1)	0.9 (1)	1.0 (1)	1.1 (1)	0.0	0.1

	T34+T35	*T36(s)*	*T37*	*T39(s)*	*T40*	*T41*
Cmc	0.0	0.0	0.2	1.1 (1)	1.8 (2)	0.1
Asx	0.0	0.0	6.0 (6)	0.0	1.8 (2)	0.1
Thr	0.0	0.0	1.0 (1)	0.0	1.8 (2)	0.0
Ser	0.0	0.0	0.3	1.9 (2)	1.2 (1)	1.8 (2)
Glx	5.7 (6)	0.0	1.9 (2)	2.0 (2)	2.9 (3)	2.0 (2)
Pro	1.2 (1)	0.0	0.0	0.0	3.2 (3)	1.2 (1)
Gly	1.0 (1)	0.0	0.4	0.0	2.2 (2)	0.2
Ala	0.0	0.0	1.1 (1)	1.0 (1)	1.2 (1)	1.1 (1)
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Val	2.2 (3)	0.5 (1)	0.3	2.0 (2)	0.9 (1)	0.1
Met	0.0	0.0	0.0	0.0	0.1	0.0
Ile	0.0	0.9 (1)	1.2 (1)	0.0	0.3	0.0
Leu	0.0	0.0	2.9 (3)	0.0	4.0 (4)	1.1 (1)
Tyr	0.0	1.0 (1)	1.0 (1)	0.0	0.9 (1)	0.9 (1)
Phe	1.0 (1)	0.0	1.0 (1)	0.0	0.1	1.1 (1)
Lys	2.1 (2)	1.0 (1)	1.0 (1)	0.0	1.1 (1)	0.0
His	0.0	1.1 (1)	0.1	0.0	0.1	1.0 (1)
Arg	0.0	0.0	0.2	1.0 (1)	0.1	1.0 (1)

	T42	*T43*	*T42+T43*	*T44*	*T45*	*T46*
Cmc	0.0	0.1	0.0	0.0	1.0 (1)	1.0 (1)
Asx	0.0	0.1	0.0	0.0	1.1 (1)	2.9 (3)
Thr	0.0	0.0	0.0	0.0	1.0 (1)	2.7 (3)
Ser	0.0	0.2	0.0	1.8 (2)	0.9 (1)	0.0
Glx	0.0	1.1 (1)	1.2 (1)	0.0	1.1 (1)	0.0
Pro	0.0	0.0	0.0	1.2 (1)	0.0	0.0
Gly	0.0	0.1	0.0	0.0	0.2	1.1 (1)
Ala	0.0	1.0 (1)	1.3 (1)	0.0	0.1	1.1 (1)
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Val	0.0	0.9 (1)	0.9 (1)	0.0	0.0	1.1 (1)

(*Continued on p. 384*)

TABLE II (*continued*)

	T42	*T43*	*T42+T43*	*T44*	*T45*	*T46*
Met	0.0	0.0	0.0	0.0	0.0	0.9 (1)
Ile	0.0	0.0	0.0	0.0	0.0	0.0
Leu	1.0 (1)	0.0	1.0 (1)	1.0 (1)	2.0 (2)	1.1 (1)
Tyr	0.0	0.0	0.0	1.0 (1)	0.0	0.0
Phe	0.0	0.0	0.0	0.0	0.0	0.0
Lys	1.0 (1)	0.1	0.7 (1)	0.0	0.0	0.0
His	0.0	1.0 (1)	0.9 (1)	0.0	0.9 (1)	0.0
Arg	0.0	1.1 (1)	1.1 (1)	1.1 (1)	1.0 (1)	1.1 (1)
	T47	*T48*	*T49*	*T50*	*T51*	
Cmc	1.7 (2)	1.0 (1)	0.1	0.0	0.0	
Asx	5.0 (5)	0.2	1.1 (1)	2.0 (2)	2.0 (2)	
Thr	0.1	0.9 (1)	1.0 (1)	0.9 (1)	0.0	
Ser	1.9 (2)	1.0 (1)	0.1	0.0	0.0	
Glx	2.0 (2)	1.0 (1)	0.1	0.0	0.0	
Pro	2.2 (2)	0.3	1.1 (1)	0.0	1.0 (1)	
Gly	6.0 (6)	4.1 (4)	1.0 (1)	0.0	0.0	
Ala	2.1 (2)	1.1 (1)	0.1	0.0	0.0	
Cys	0.0	0.0	0.0	0.0	0.0	
Val	0.9 (1)	0.0	1.9 (2)	1.1 (1)	0.0	
Met	0.0	0.8 (1)	0.0	0.0	0.9 (1)	
Ile	0.1	1.3 (2)	0.0	1.0 (1)	0.0	
Leu	3.0 (3)	2.0 (2)	0.1	1.0 (1)	0.0	
Tyr	0.1	0.0	0.9 (1)	1.0 (1)	0.0	
Phe	0.1	0.0	0.0	0.0	0.0	
Lys	0.1	1.1 (1)	1.0 (1)	0.0	0.0	
His	1.0 (1)	0.0	0.0	0.0	0.0	
Arg	1.0 (1)	0.0	0.1	1.0 (1)	1.0 (1)	

composition no further analysis was performed. Several predicted peptides were sufficiently polar to be unretained by the column and were identified by Edman degradation: Asp-Glu-Lys (T2), Ser-Asn-Arg (T4), Ala-Gly-Lys (T16) and Asn-Arg-Arg (T22 + T23). Peptide T38 was shown to be unretained by the column by chromatography of a synthetic sample (Ser-Asp-Ser-Ser-Arg). All expected peptides were observed in the TFA tryptic maps except the very hydrophobic peptide Gln-Met-Ile-Tyr-Gln-Gln-His-Gln-Ser-Trp-Leu-Arg-Pro-Val-Leu-Arg (T3).

Heterogeneity can be introduced into a tryptic map by the generation of peptides other than those arising directly from the predicted tryptic digestion. In this tryptic map (Fig. 1) additional peptide peaks are the result of various mechanisms such as partial cleavage by trypsin at less favorable sites (adjacent to proline, etc.), chymotrypsin-like cleavages and carbohydrate heterogeneity (Table III). A partial trypsin cleavage product of the arginyl-prolyl bond in peptide T3e was identified at 31.5 min, but the intact peptide T3 was not found in this map. Two chymotryptic-like cleavages were observed: a minor cleavage of the methionyl-isoleucyl bond of peptide T18 at 66 min resulting in peptide T18a (residues 208–212) at 39 min, and cleavage of peptide T25 (38 min) after phenylalanine residue 274 resulting in peptide T25a (residues 268–274) at 41.5 min. A small percentage of peptides T31 and T40 were found

TABLE III

MULTIPLE ELUTION TIMES, NON-TRYPSIN LIKE CLEAVAGES AND INCOMPLETE TRYPSIN CLEAVAGE IN THE rt-PA TRYPTIC DIGEST

Tryptic peptide	*Fragment*	*Sequence*	*TFA map*	*Phosphate map*
T3	11–27	TQMIYQQHQSWLRPVLR		X
T3a	11–23	TQMIYQQHQSWLR		X
T3b	11–22	TQMIYQQHQSWL		X
T3c	22–27	—LRPVLR		X
T3d	24–27	—PVLR	X	X
T11	102–129	GTWSTAESGAECTNWNSSA-LAQKPYSGR	X	X
T11a	102–129	GTWSTAESGAECTNWMSSA-LAQKPYSGR		X
T17	163–189	YSSEFCSTPACSEGNSDCYFGNG-SAYR	X	X
T17a	163–189	YSSEFCSTPACSEGNSDCYFGNG-SAYR	X	X
T18	190–212	GTHSLTESGASCLPWNSMILIGK	X	X
T18a	208–212	—ILIGK	X	X
T25	268–275	QYSQPQFR	X	X
T25a	268–274	QYSQPQF	X	
T27	278–296	GGLFADIASHPWQAAIFAK	X	X
T27a	278–291	GGLFADIASHPWQA		X
T32	328–339	FPPHHLTVILGR	X	X
T32a	328–339	FPPHHLTVILGR		X
T4 + T5	28–30 + 31–40	SNRVEYCWNSGR		X
T23 + T24	250 + 251–267	RLTWEYCDVPSCSTCGLR		X
T28 + T29	297–298 + 299	HRR	X	
T29 + T30	299 + 300–304	RSPGER	X	X
T31a + T40a	314–327 395–416	SCWILSAAHCFQER CLPPADLQLPDWTECELS-GYGK	X X	
T34 + T35	343–351 + 352–356	VVPGEEEQKFEVEK	X	X
T42 + T43	428–429 + 430–434	LKEAHVR	X	

linked by an intact disulfide and containing chymotryptic-like cleavages (see Fig. 2, T31a + T40a, 63 min). Multiplets of the glycopeptides T11, T17, T45 were found (see carbohydrate-containing peptide analysis section), Five examples of incomplete tryptic digestion were also identified: T22 + T23 (void), T28 + T29 (13 min), T29 + T30 (19 min), T34 + T35 (44 min), T42 + T43 (29 min).

Phosphate tryptic map

Identification of the peaks obtained from the sodium phosphate tryptic map required the removal of the non-volatile salt from the isolated peptides. This was accomplished by rechromatography of the collected peaks in an HPLC solvent system using volatile components. Rechromatography also provided an evaluation of the purity of the peaks in the original phosphate-based map by a second dimension separation of peaks containing multiple components.

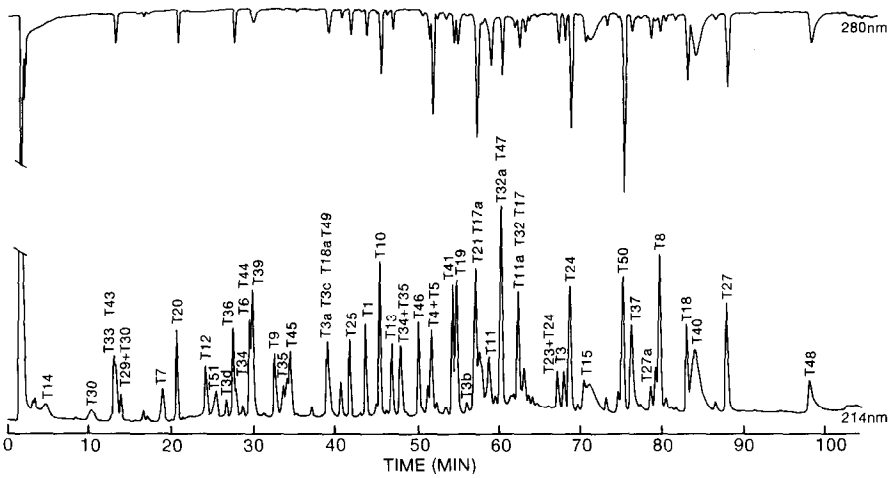

Fig. 3. Sodium phosphate based tryptic map of RCM rt-PA. This separation was performed on a 5-μm Nova-Pak C_{18} column (15 × 0.46 cm I.D.), Mobile phase A consisted of 0.05 *M* sodium phosphate (pH 2.85), and mobile phase B was acetonitrile. A linear gradient of 0–30% mobile phase B was run in 90 min followed by 30–60% mobile phase B in 30 min. The 70-μg sample was loaded in 0.1 ml of 0.1 *M* ammonium bicarbonate and monitored at both 214 and 280 nm.

Fifty-three peaks recovered from the semi-preparative separation of the RCM rt-PA tryptic digest in the sodium phosphate and acetonitrile solvent system (Fig. 3) were rechromatographed in the rapid TFA and acetonitrile system (see Experimental section) then analyzed for amino acid composition (Table IV). The location of the tryptic peptides in the rt-PA sequence was determined from the amino acid analysis, except for very minor peptides and for the polar peptides in the void volume which had been identified in the TFA-based peptide map. In this set of samples, tyrosine recovery was variable while tryptophan was found to be destroyed using the Pico Tag batch hydrolysis system for amino acid analysis. Therefore, the presence of these residues was confirmed by the relative absorbance of each peak monitored at 280 nm. The utility of dual wavelength detection is shown in Fig. 4 where peptide T45 does not contain tyrosine (Cm Cys-Thr-Ser-Gln-His-Leu-Leu-Asn-Arg) and exhibits only 214 nm absorption, whereas peptide T25 (Gln-Tyr-Ser-Gln-Pro-Gln-Phe-Arg) is detected both at 214 and 280 nm. The majority of the peptide peaks eluted as single components in the second dimension chromatography using a TFA–acetonitrile solvent system as in Fig. 4. However, additional purification of some of the fractions isolated from the phosphate-based chromatography was obtained. For example, chromatogram A in Fig. 5 resolved one major peak (T49) and two minor peaks containing the peptides T3a, T3c and T18a. Chromatogram B resolved one major peak (T47) and two minor peaks (T32). Several other peaks clearly contained multiple peptide components in the phosphate tryptic map and peptides were resolved in the TFA-based rechromatography (Fig. 5C, D–F). In most cases, the minor peak represented an insignificant level of peptide by amino acid analysis.

As in the TFA tryptic map, peptide heterogeneity was found by the resolution of other peptides in addition to those arising from the predicted tryptic digestion. Six peptides were found to be reproducibly represented in more than one peak in the

TABLE IV

AMINO ACID COMPOSITIONS OF SODIUM PHOSPHATE TRYPTIC MAP PEPTIDE PEAKS

N.D. = Not detected.

	T1	*T3*	*T3a*	*T3b*	*T3c*	*T3d*
Asp	0.2	0.0	0.2	0.6	0.0	0.0
Glu	1.3 (1)	3.3 (4)	3.7 (4)	3.8 (4)	0.0	0.0
Cm Cys	0.8 (1)	0.0	0.0	0.1	0.0	0.0
Ser	1.1 (1)	0.9 (1)	0.4 (1)	1.2 (1)	0.0	0.0
Gly	0.2	0.2	0.5	0.8	0.0	0.0
His	0.0	1.0 (1)	1.2 (1)	1.0 (1)	0.0	0.0
Arg	1.1 (1)	2.2 (2)	1.1 (1)	0.2	1.9 (2)	0.9 (1)
Thr	0.0	1.1 (1)	1.3 (1)	0.9 (1)	0.0	0.0
Ala	0.0	0.0	0.0	0.4	0.0	0.0
Pro	0.0	1.2 (1)	0.2	0.3	1.0 (1)	1.1 (1)
Tyr	1.0 (1)	N.D (1)	0.1 (1)	0.8 (1)	0.0	0.0
Val	0.8 (1)	1.1 (1)	0.1	0.3	1.1 (1)	1.0 (1)
Met	0.0	0.4 (1)	0.3 (1)	0.2 (1)	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.8 (1)	0.9 (1)	1.1 (1)	1.2 (1)	0.0	0.0
Leu	0.1	2.2 (2)	0.3 (1)	0.5 (1)	1.5 (2)	1.1 (1)
Phe	0.0	0.1	0.0	0.1	0.0	0.0
Lys	0.0	0.1	0.1	0.3	0.0	0.0

	T4 + T5	*T6*	*T7*	*T8*	*T9*	*T10*
Asp	2.1 (2)	0.0	0.1	2.4 (2)	1.1 (1)	1.1 (1)
Glu	1.3 (1)	1.1 (1)	1.2 (1)	4.0 (4)	1.2 (1)	2.1 (2)
Cm Cys	1.2 (2)	0.9 (1)	0.8 (1)	3.6 (4)	1.8 (2)	0.8 (1)
Ser	1.6 (2)	0.7 (1)	1.9 (2)	0.9 (1)	0.0	1.0 (1)
Gly	1.0 (1)	0.0	0.3	3.8 (4)	0.0	1.1 (1)
His	0.0	1.1 (1)	0.0	0.0	0.0	0.0
Arg	1.6 (2)	0.0	1.0 (1)	0.0	1.0 (1)	1.0 (1)
Thr	0.1	0.0	0.0	1.0 (1)	0.9 (1)	1.0 (1)
Ala	0.1	1.0 (1)	0.0	2.0 (2)	0.0	1.0 (1)
Pro	0.0	0.8 (1)	1.1 (1)	1.0 (1)	0.0	0.0
Tyr	0.8 (1)	0.0	0.0	0.9 (1)	0.0	1.5 (2)
Val	0.9 (1)	1.2 (2)	0.0	0.9 (1)	0.0	0.0
Met	0.0	0.0	0.0	0.0	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.1	0.0	0.0	0.1	1.0 (1)	1.1 (1)
Leu	0.1	0.0	0.0	1.2 (1)	0.0	0.1
Phe	0.0	0.0	0.0	3.6 (4)	0.0	0.0
Lys	0.0	0.9 (1)	0.0	1.0 (1)	0.0	0.0

	T11	*T11a*	*T12*	*T13*	*T14*	*T15*
Asp	2.3 (2)	2.3 (2)	1.1 (1)	2.6 (2)	2.1 (2)	1.6 (1)
Glu	3.0 (3)	3.5 (3)	0.0	0.0	0.0	0.0
Cm Cys	1.0 (1)	0.7 (1)	0.0	0.7 (1)	0.0	0.8 (1)
Ser	5.1 (5)	4.8 (5)	0.0	0.0	0.0	1.2 (1)
Gly	3.2 (3)	3.3 (3)	0.0	2.0 (2)	0.0	0.2
His	0.0	0.0	0.1	1.0 (1)	0.0	0.1
Arg	0.9 (1)	0.7 (1)	1.9 (2)	1.0 (1)	0.9 (1)	0.0

(*Continued on p. 388*)

TABLE IV (*continued*)

	T11	*T11a*	*T12*	*T13*	*T14*	*T15*
Thr	2.4 (3)	2.8 (3)	0.0	0.0	0.0	0.1
Ala	3.8 (4)	3.7 (4)	1.1 (1)	0.0	0.0	0.1
Pro	1.1 (1)	1.2 (1)	1.0 (1)	0.0	1.0 (1)	1.2 (1)
Tyr	N.D.(1)	0.9 (1)	0.0	0.9 (1)	0.0	0.9 (1)
Val	0.1	0.2	0.1	0.0	0.0	1.0 (1)
Met	0.0	0.1	0.0	0.0	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.1	0.5	0.8 (1)	0.0	0.0	0.1
Leu	1.2 (1)	1.4 (1)	0.2	2.0 (2)	0.1	0.2
Phe	0.2	0.1	0.0	0.0	0.0	1.1 (1)
Lys	1.0 (1)	0.9 (1)	0.1	0.0	0.1	1.7 (2)
	T17	*T17a*	*T18*	*T18a*	*T19*	*T20*
Asp	3.1 (3)	3.2 (3)	1.1 (1)	0.0	1.2 (1)	1.2 (1)
Glu	2.4 (2)	2.5 (2)	1.1 (1)	0.0	2.1 (2)	0.0
Cm Cys	3.1 (3)	2.6 (3)	0.9 (1)	0.0	0.0	0.8 (1)
Ser	5.6 (6)	4.9 (6)	3.6 (4)	0.0	1.0 (1)	0.0
Gly	3.4 (3)	3.5 (3)	2.9 (3)	1.1 (1)	1.9 (2)	0.0
His	0.1	0.1	1.0 (1)	0.0	0.2	0.9 (1)
Arg	0.9 (1)	1.1 (1)	0.1	0.0	0.4	1.0 (1)
Thr	1.0 (1)	1.1 (1)	1.9 (2)	0.0	1.0 (1)	0.0
Ala	2.3 (2)	2.3 (2)	0.9 (1)	0.0	2.6 (3)	0.0
Pro	1.2 (1)	1.2 (1)	1.0 (1)	0.0	1.0 (1)	0.0
Tyr	N.D.(3)	2.0 (3)	0.0	0.0	0.8 (1)	0.8 (1)
Val	0.1	0.2	0.0	0.0	1.1 (1)	0.0
Met	0.0	0.0	0.9 (1)	0.0	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.1	0.0	2.1 (2)	1.8 (2)	0.3	0.0
Leu	1.3	0.4	3.4 (3)	1.0 (1)	1.8 (2)	0.0
Phe	2.1 (2)	2.0 (2)	0.0	0.0	0.0	0.0
Lys	0.1	0.1	1.3 (1)	1.2 (1)	1.2 (1)	0.0
	T21	*T23+T24*	*T24*	*T25*	*T27*	
Asp	3.0 (3)	1.2 (1)	1.2 (1)	0.1	1.0 (1)	
Glu		1.3 (1)	1.2 (1)	2.4 (3)	1.0 (1)	
Cm Cys	0.9 (1)	1.5 (3)	2.4 (3)	0.0	0.0	
Ser	0.0	1.8 (2)	1.8 (2)	1.0 (1)	1.0 (1)	
Gly	1.0	1.3 (1)	1.0 (1)	1.0	2.1 (2)	
His	0.8 (1)	0.3	0.0	0.0	1.0 (1)	
Arg	0.0	1.7 (2)	0.9 (1)	1.1 (1)	0.1	
Thr	0.0	1.8 (2)	1.8 (2)	0.0	0.0	
Ala	1.1 (1)	0.3	0.0	0.0	5.1 (5)	
Pro	2.0 (2)	1.2 (1)	1.1 (1)	1.1 (1)	1.0 (1)	
Tyr	0.0	0.7 (1)	1.0 (1)	0.9 (1)	0.0	
Val	0.9 (1)	1.0 (1)	1.0 (1)	0.0	0.0	
Met	0.0	0.0	0.0	0.0	0.0	
Cys	0.0	0.0	0.0	0.0	0.0	
Ile	0.0	0.2	0.0	0.1	2.2 (2)	
Leu	1.0 (1)	2.4 (2)	1.8 (2)	0.1	1.2 (1)	
Phe	0.0	0.2	0.0	1.1 (1)	2.3 (2)	
Lys	2.3 (2)	0.1	0.1	0.0	1.4 (1)	

TABLE IV (*continued*)

	T27a	*T29+T30*	*T30*	*T32*	*T32a*
Asp	0.9 (1)	0.0	0.0	0.1	0.2
Glu	1.0 (1)	1.0 (1)	1.0 (1)	0.2	0.4
Cm Cys	0.0	0.0	0.0	0.0	0.0
Ser	1.0 (1)	1.0 (1)	1.0 (1)	0.3	0.3
Gly	1.9 (2)	1.0 (1)	1.1 (1)	1.2 (1)	1.5 (1)
His	0.9 (1)	0.0	0.0	1.8 (2)	1.7 (2)
Arg	0.1	2.0 (2)	1.0 (1)	1.3 (1)	0.8 (1)
Thr	0.1	0.1	0.0	0.9 (1)	0.8 (1)
Ala	3.1 (3)	0.0	0.0	0.0	0.0
Pro	0.9 (1)	1.0 (1)	1.1 (1)	2.2 (2)	1.9 (2)
Tyr	0.1	0.1	0.0	0.0	0.0
Val	0.1	0.0	0.0	1.0 (1)	0.8 (1)
Met	0.0	0.0	0.0	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0
Ile	1.1 (1)	0.0	0.0	0.9 (1)	1.0 (1)
Leu	1.2 (1)	0.0	0.0	2.6 (2)	2.3 (2)
Phe	1.3 (1)	0.0	0.0	1.2 (1)	1.2 (1)
Lys	0.1	0.0	0.0	0.1	0.1

	T33	*T34*	*T34+T35*	*T35*	*T36*	*T37*
Asp	0.0	0.1	0.1	0.1	0.0	4.8 (6)
Glu	0.0	3.2 (4)	5.1 (6)	1.8 (2)	0.1	2.4 (2)
Cm Cys	0.0	0.0	0.1	0.0	0.0	0.0
Ser	0.0	0.1	0.1	0.1	0.0	0.2
Gly	0.0	1.0 (1)	1.1 (1)	0.2	0.1	0.7
His	0.0	0.1	0.0	0.0	1.0 (1)	0.0
Arg	1.0 (1)	0.1	0.1	0.1	0.1	0.1
Thr	1.1 (1)	0.1	0.1	0.0	0.0	1.0 (1)
Ala	0.0	0.2	0.1	0.1	0.0	1.2 (1)
Pro	0.0	1.2 (1)	1.1 (1)	0.0	0.0	0.1
Tyr	1.1 (1)	0.1	0.0	0.0	0.9 (1)	0.8 (1)
Val	0.0	1.2 (2)	2.1 (3)	1.0 (1)	0.7 (1)	0.1
Met	0.0	0.0	0.0	0.0	0.0	0.0
Cys	0.0	0.0	0.1	0.0	0.0	0.0
Ile	0.0	0.2	0.2	0.1	1.7 (1)	1.0 (1)
Leu	0.0	0.1	0.0	0.1	0.0	3.1 (3)
Phe	0.0	0.0	1.2 (1)	1.1 (1)	0.0	1.0 (1)
Lys	0.0	1.1 (1)	2.5 (2)	1.0 (1)	1.1 (1)	1.1 (1)

	T39	*T41*	*T43*	*T44*	*T45*	*T46*
Asp	0.1	0.1	0.1	0.0	1.0 (1)	3.2 (3)
Glu	1.7 (2)	1.8 (2)	1.0 (1)	0.0	1.1 (1)	0.0
Cm Cys	0.9 (1)	0.0	0.0	0.0	1.0 (1)	0.9 (1)
Ser	1.9 (2)	1.5 (2)	0.0	1.5 (2)	2.0 (1)	0.2
Gly	0.0	0.3	0.0	0.0	0.2	1.0 (1)
His	0.1	0.9 (1)	1.0 (1)	0.0	0.9 (1)	0.0
Arg	1.0 (1)	1.0 (1)	1.0 (1)	1.2 (1)	1.0 (1)	1.0 (1)
Thr	0.0	0.1	0.0	0.0	0.9 (1)	2.8 (3)
Ala	1.0 (1)	1.0 (1)	1.1 (1)	0.0	0.0	1.2 (1)
Pro	0.4	1.1 (1)	0.0	0.8 (1)	0.0	0.1
Tyr	0.0	1.0 (1)	0.0	0.8 (1)	0.1	0.0

(*Continued on p. 390*)

TABLE IV (*continued*)

	T39	*T41*	*T43*	*T44*	*T45*	*T46*
Val	1.3 (2)	0.1	1.0 (1)	0.0	0.0	1.1 (1)
Met	0.0	0.1	0.0	0.0	0.0	0.7 (1)
Cys	0.0	0.0	0.0	0.0	0.0	0.0
Ile	0.0	0.0	0.0	0.0	0.0	0.1
Leu	0.2	1.3 (1)	0.0	1.0 (1)	1.5 (2)	1.1 (1)
Phe	0.0	1.0 (1)	0.0	0.0	0.0	0.0
Lys	0.2	0.1	0.0	0.0	0.0	0.1

	T47	*T48*	*T49*	*T50*	*T51*
Asp	4.9 (5)	0.1	1.1 (1)	1.9 (2)	2.2 (2)
Glu	2.2 (2)	1.1 (1)	0.2	0.0	0.0
Cm Cys	1.8 (2)	0.9 (1)	0.0	0.1	0.0
Ser	1.8 (2)	0.9 (1)	0.0	0.2	0.1
Gly	5.7 (6)	3.9 (4)	1.1 (1)	0.0	0.1
His	1.0 (1)	0.0	0.0	0.0	0.0
Arg	1.1 (1)	0.1	0.1	0.9 (1)	1.3 (1)
Thr	0.1	1.1 (1)	1.2 (1)	1.0 (1)	0.0
Ala	2.1 (2)	0.0	0.1	0.2	0.0
Pro	2.0 (2)	0.0	1.2 (1)	0.0	1.0 (1)
Tyr	0.0	0.0	0.9 (1)	1.0 (1)	0.0
Val	1.0 (1)	1.0 (1)	1.2 (2)	1.0 (1)	0.1
Met	0.0	0.8 (1)	0.0	0.0	0.5 (1)
Cys	0.0	0.0	0.0	0.0	0.0
Ile	0.1	0.9 (2)	0.1	0.9 (1)	0.0
Leu	3.4 (3)	2.1 (2)	0.0	1.3 (1)	0.1
Phe	0.1	0.0	0.0	0.1	0.0
Lys	0.0	1.2 (1)	0.9 (1)	0.1	0.1

phosphate map (Table III). Peptide T3 at 68 min (Fig. 3) was found to have significant cleavage by trypsin between the arginyl-prolyl bond resulting in recovery of peptide T3a (residues 11–23) at 39 min and T3d (residues 24–27) at 26 min. A small amount of chymotryptic-like cleavage at residue 22 (Leu) and residue 21 (Trp) of peptide T3 resulted in the recovery of peptide T3b (residues 11–22) at 60 min and T3c (residues 22–27) at 39 min. Chymotryptic-like cleavage was also found for peptide T18 (83 min) at methionine residue 207, giving rise to peptide T18a (residues 208–212) at 39 min. Peptide T27 (residues 278–296) at 88 min was found to be partially cleaved at alanine residues 291 and 292 (Ala,Ala) resulting in T27a (residues 278–291) at 79 min. Peptide T32 eluting at 63 min was also significantly recovered at 60 min (T32a). The lack of deamidation sites in peptide T32 sequence, and the presence of two prolines suggests conformational effects may play a role in the multiplicity of peaks derived from this peptide[12]. Four examples of incomplete tryptic digestion were identified: T4 + T5 at 52 min, T23 + T24 at 67 min, T29 + T30 at 14 min and T34 + T35 at 48 min. Peptides T29 + T30 and T34 + T35 were also observed in the TFA map of the tryptic digest.

Identification and characterization of carbohydrate containing peptides

Tryptic glycopeptides of RCM rt-PA were isolated by Con A-Sepharose lectin

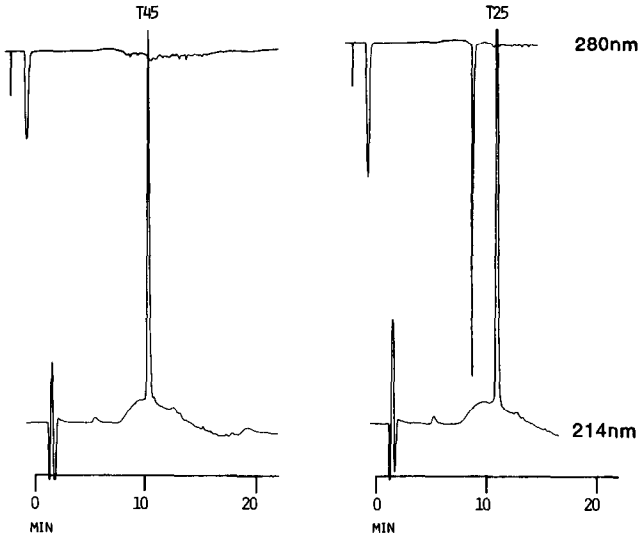

Fig. 4. Rechromatography of peaks from the sodium phosphate tryptic map of RCM rt-PA using the volatile TFA solvent system. Mobile phases A and B were prepared as in Fig. 1. Peaks were eluted by a rapid linear gradient to 60% mobile phase B in 5 min, followed by 20 min of isocratic elution at 60% mobile phase B. The elution profiles at 280 nm demonstrate the presence of aromatic residues tyrosine or tryptophan as in the peptide T25 peak or absence as in the peptide T45 peak.

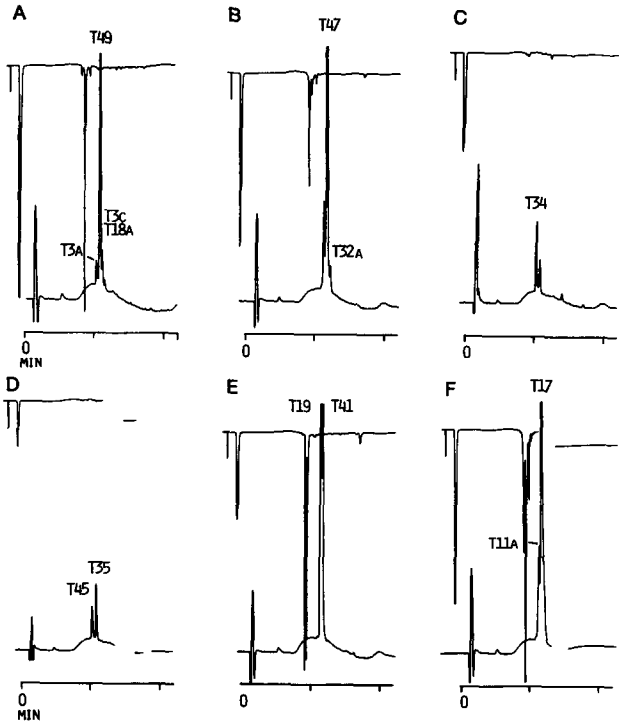

Fig. 5. Examples of the rechromatography of peaks from the sodium phosphate tryptic map of RCM rt-PA using the volatile TFA solvent system described in Fig. 4. For A–F, see text.

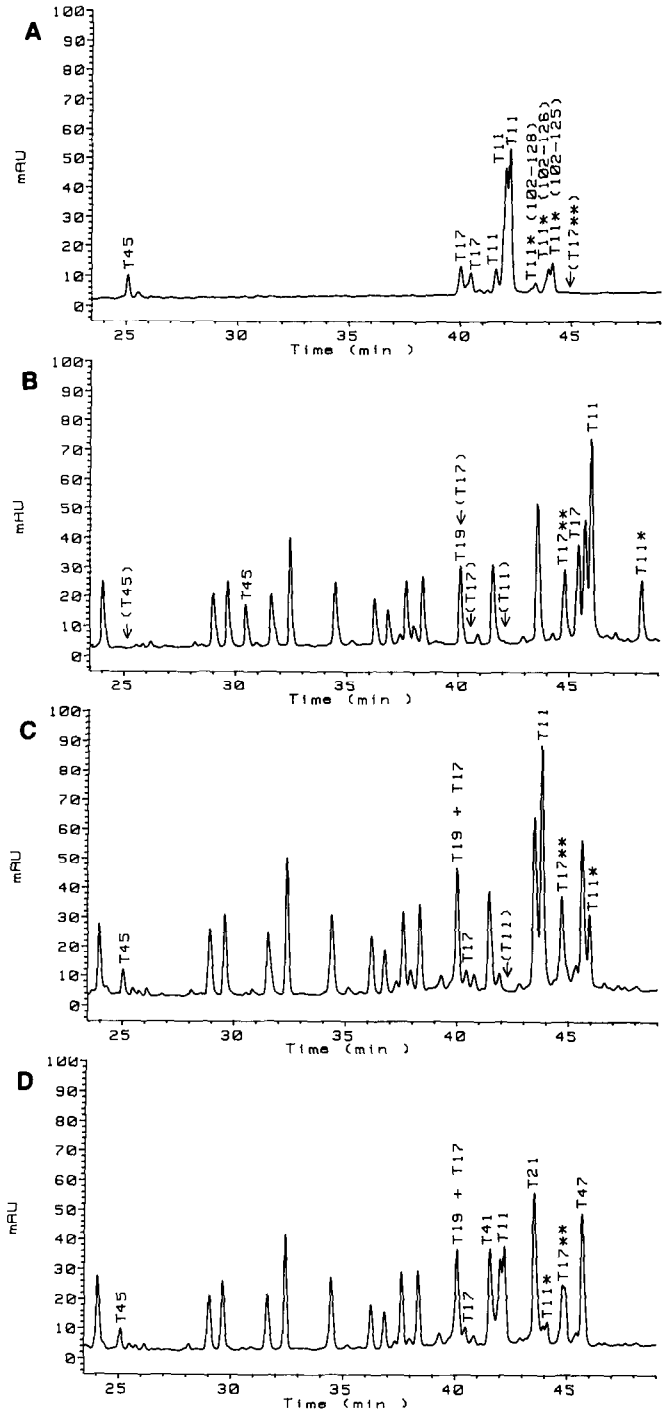

Fig. 6. Comparison of the tryptic peptides of RCM rt-PA recovered from Con A-Sepharose (A), following treatment with PNGase F (B) or following treatment with Endo H (C) to the total RCM rt-PA digest (D). The separation using the TFA solvent system is described in Fig. 1. Approximately 100 μg of sample were loaded for each chromatogram and the eluate was monitored at 220 nm. T11* indicates cleaved forms of peptide T11 as noted on chromatogram (A). T17** indicates the non-glycosylated (Type II) form of peptide T17.

chromatography (Fig. 6A) and compared to a control RCM rt-PA tryptic digest (Fig. 6D) using the TFA tryptic method. Peaks were collected and identified by amino acid analysis. The carbohydrate-containing peptides are T11, T17 and T45, resulting from glycosylations at Asn-117, Asn-184 and Asn-448, respectively. Although peptide T19 contains a potential N-linked glycosylation site, as recognized by the sequence Asn-X-Ser/Thr, it is not glycosylated. These results agree with the earlier report by Pohl *et al.*[9] establishing the identity of the carbohydrate-containing peptides which were further confirmed by mass spectrometry by Carr and Roberts[13]. All of the glycopeptides eluted as multiple peaks, possibly due to carbohydrate microheterogeneity. Peaks labeled T11* arise from carboxy-terminal cleavages of peptide T11. Peaks with amino acid compositions corresponding to T17 are observed at two positions in the control map (40.5 and 44.8 min, Fig. 6D). However, the later-eluting peak (T17**) did not contain detectable amino sugars and was not bound by Con A-Sepharose, indicating that this peptide is probably not glycosylated. Bennet[8] and Pohl *et al.*[9] previously observed that rt-PA can be separated into variants I and II, differing by the optional glycosylation of residue 184 contained in peptide T17.

The type of N-linked oligosaccharide present at each site was determined by TFA tryptic mapping of RCM rt-PA after treatment with Endo H or PNGase F. Endo H removes high-mannose oligosaccharides by cleaving between the two N-acetylglucosamine residues of the core portion[14]. PNGase F hydrolyzes the N-acetylglucosamyl-asparagine bond of both high mannose and complex oligosaccharides[15] resulting in deglycosylation and in conversion of the attachment asparaginyl residue to aspartic acid. Removal of the hydrophilic carbohydrate moiety results in increased retention of a peptide on the RP-HPLC column. Peak identities were confirmed by amino acid analysis.

Fig. 6 shows the portion of the TFA tryptic map that contains the glycopeptide peaks. Digestion with PNGase F (Fig. 6B) caused an increase in the retention time of glycopeptide peaks T11, T17 and T45 consistent with the loss of oligosaccharide side chain. The new peak resulting from deglycosylated peptide T17 does not co-elute with the non-glycosylated peptide T17**, because the former contains aspartic acid while the latter contains asparagine at position 184. The relative elution positions of T17 (Asp-184) and T17** (Asn-184) fits with the observed increase in retention time usually observed with deamidation of a peptide[16]. All of the intact glycopeptides elute as single peaks upon deglycosylation, suggesting that the multiple peaks in the control map are due to carbohydrate microheterogeneity.

Endo H digestion only affected peaks containing glycopeptides T11 and T11* (Fig. 6C). The peaks which result from deglycosylation of peptide T11 by Endo H are of a different retention time than the peaks resulting from PNGase F deglycosylation of peptide T11. This is because Endo H leaves a residue of N-acetylglucosamine attached to asparagine. The altered retention time of glycopeptide peak T11 upon treatment with Endo H demonstrated the presence of a high mannose oligosaccharide at Asn-117. Glycopeptide peaks T17 and T45 were deglycosylated by PNGase F, and not Endo H, further indicating the presence of complex oligosaccharides on Asn-184 and Asn-448.

DISCUSSION

Bacterial clones that contained cDNA sequences corresponding to human tissue plasminogen activator were reported in 1984[6] and from the cDNA data an amino acid sequence of 562 residues was deduced. Amino acid sequencing of purified rt-PA from melanoma cells indicated a protein of 527 amino acids, suggesting post-translational processing that removes a hydrophobic signal peptide and a hydrophilic "pro-peptide" sequence. Pohl *et al.*[9] confirmed the structure of tissue plasminogen activator produced and secreted by a melanoma cell line. These investigators describe an RP-HPLC tryptic map for the N-terminal fragment 1–275 of the protein and of a cyanogen bromide fragment of residues 276–455 both initially formed by plasmin digestion. Subsequently, rt-PA was expressed in Chinese hamster ovary cells and a process was developed for large scale manufacture of the protein for clinical trials as a thrombolytic agent for use in treatment of myocardial infarction[17]. Development of the TFA-based and phosphate-based tryptic maps allowed complete characterization of the protein for FDA submissions for permission to enter clinical trials and for the product licensing application. It was also proposed to use the phosphate-based tryptic map as a release procedure to monitor the consistency of manufactured lots.

The size of rt-PA (66 000 daltons) and the presence of three carbohydrate side chains (including one optional glycosylation site) suggested that the tryptic map would be extremely complex. It was found that other factors such as non-specific cleavages and incomplete tryptic cleavages (Table III) resulted in the formation of more than the expected 51 tryptic peptides (Table I). The presence of 17 disulfide bonds in rt-PA required reduction and S-carboxymethylation of the molecule to achieve satisfactory trypsin digestion, introducing additional concerns for peptide heterogeneity. Under the conditions used for this study, only one example of a small amount of an intact disulfide bond was found (Table III, T31a–T40a) and S-carboxymethylation was otherwise complete as measured by amino acid analysis (Table II and IV).

Because of the complexity of the tryptic map a number of approaches were used to investigate possible co-elutions. Limited digestion of rt-PA with plasmin cleaves the molecule between residue 275 and 276. Two large fragments (1–275 and 276–527) were isolated by gel filtration after reduction and S-carboxymethylation. The TFA tryptic map was optimized using a Nova-Pak C_{18} column with a mobile phase that consisted of 0.1% TFA and a linear gradient of acetonitrile. Comparison of the maps of the separated domains (residues 1–275 and 276–527) with the map of complete molecule (residues 1–527) showed the TFA map contained minimal co-elution. Also, the use of volatile mobile phases allowed direct characterization of the tryptic peptides by amino acid analysis (Table II) and N-terminal sequencing.

A further investigation of potential co-elutions involved the use of 50 m*M* sodium phosphate (pH 2.85) as a mobile phase. It has been shown that the use of different mobile phases produces significant changes in the selectivities of RP-HPLC analyses of peptides[18]. The sodium phosphate tryptic map showed a similar peptide elution order to the TFA tryptic map, as noted for other separations[19]. The peptides that were resolved in the semi-preparative, phosphate-based tryptic map were collected and rechromatographed using a TFA containing mobile phase to allow further characterization. As shown in Fig. 5, some co-elutions were detected by rechromato-

graphy of peaks collected from the phosphate map. Intact peptide T3 and various cleavage products of the peptide were found in the phosphate tryptic map (Table III) whereas only a fragment of T3 was identified in the TFA tryptic map. Peptide T31 (305–327), which contained eight hydrophobic residues and was the last peptide to be eluted in the TFA-based separation, was not found in the phosphate map. The amino acid compositions of all peptides eluted from the two tryptic maps were consistent with the proposed amino acid sequence of rt-PA.

The RCM rt-PA tryptic glycopeptides were characterized by Con A-Sepharose affinity chromatography, specific glycosidases and comparative tryptic mapping. Endo H is specific for high mannose structures, while PNGase F results in removal of all carbohydrate side chains. This combination of techniques confirmed that glycopeptide T11 contained a high mannose oligosaccharide at Asn-177, while T17 and T45 contained complex oligosaccharides at Asn-184 and Asn-448, respectively. Glycosylation of peptide T17 appears to be optional in this expression system and T17 was present in both glycosylated and non-glycosylated forms. Carbohydrate microheterogeneity in the glycopeptides T11, T17 and T45 was demonstrated by the disappearance of multiple peaks for each peptide and the recovery of single peaks after treatment by glycosidases.

In conclusion, these studies have shown that it is possible to develop an RP-HPLC tryptic map for a molecule as complex as rt-PA. The tryptic map has been particularly important in the development of rt-PA as a pharmaceutical product for thrombolytic therapy. The structural information that can be deduced from these high-resolution separations is invaluable in the characterization and quality control of the recombinant DNA-derived product.

ACKNOWLEDGEMENTS

We wish to acknowledge Keri M. Tate, William J. Kohr, Carole Ward and Dr. William F. Bennett for their contributions in many valuable discussions.

REFERENCES

1 W. S. Hancock, C. A. Bishop and M. T. W. Hearn, *Anal. Biochem.*, 92 (1979) 170.
2 W. J. Kohr, R. Keck and R. N. Harkins, *Anal. Biochem.*, 122 (1982) 348.
3 P. A. Hartman, J. D. Stodola, G. C. Harbour and J. G. Hoogerheide, *J. Chromatogr.*, 360 (1986) 385.
4 M. Kunitani, P. Hirtzer, D. Johnson, R. Halenbeck, A. Boosman and K. Koths, *J. Chromatogr.*, 359 (1986) 391.
5 W. S. Hancock. *Chromatogr. Forum*, 1 (1986) 57.
6 D. Collen, J. M. Stassen, B. J. Marifino, Jr., S. E. Builder, F. DeCock, J. Ogez, D. Tagiri, D. Pennica, W. F. Bennett, J. Saliva and C. F. Hoyng, *J. Pharm. Exper. Ther.*, 231 (1984) 146.
7 D. Pennica, W. E. Holmes, W. J. Kohr, R. N. Harkins, G. A. Vehar, C. A. Ward, W. F. Bennett, E. Yelverton, P. H. Seeburg, H. L. Heyneker, D. V. Goeddel and D. Collen, *Nature (London)*, 301 (1983) 214.
8 W. F. Bennett, *Thromb. Res.*, 50 (1983) 106.
9 G. Pohl, M. Kallstrom, N. Bergsdorf, P. Wallen and H. Jornvall. *Biochemistry*, 23 (1984) 3701.
10 A. M. Crestfield, S. Moore and W. H. Stein, *J. Biol. Chem.*, 238 (1963) 622.
11 G. A. Vehar, M. W. Spellman, B. A. Keyt, C. K. Ferguson, R. G. Keck, R. C. Chloupek, R. Harris, W. F. Bennett, S. E. Builder and W. S. Hancock, *Cold Spring Harbor Symp. Quant. Biol.*, 51 (1986) 551.

12 J. Jacobson, W. Melander, G. Vaisnys and Cs. Horváth, *J. Phys. Chem.*, 88 (1984) 4536.
13 S. A. Carr and G. D. Roberts, *Anal. Biochem.*, 157 (1986) 396.
14 A. L. Tarentino and F. Maley, *J. Biol. Chem.*, 249 (1974) 811.
15 A. L. Tarentino, C. M. Gomez and T. H. Plummer, *Biochemistry*, 24 (1985) 4665.
16 D. Guo, C. T. Mant, A. K. Taneja, J. M. R. Parker and R. S. Hodges, *J. Chromatogr.*, 359 (1986) 499.
17 S. E. Builder and E. Grossbard, in *Transfusion Medicine: Recent Technological Advances*, A. R. Liss, New York, NY, 1986, p. 303.
18 W. S. Hancock, C. A. Bishop, J. E. Battersby, D. R. K. Harding and M. T. W. Hearn, *J. Chromatogr.*, 168 (1979) 377.
19 W. S. Hancock and D. R. K. Harding, in W. S. Hancock (Editor), *CRC Handbook of CRC for the Separation of Amino Acids, Peptides and Proteins*, Vol. II, CRC Press, Boca Raton, FL, 1984, p. 3.