Journal of Pharmaceutical & Biomedical Analysis
Vol. 11, No. 8, pp. 737–744, 1993
Printed in Great Britain

0731–7085/93 $6.00 + 0.00
© 1993 Pergamon Press Ltd

RP-HPLC peptide mapping methods for the analysis of recombinant human pro-urokinase*

I. FACCHETTI,† A. VALLA and A. VIGEVANI

Farmitalia Carlo Erba, R & D, Analytical Chemistry, Erbamont Group, Via Papa Giovanni XXIII 23, 20014 Nerviano, Italy

Abstract: Among the techniques available for the detection of protein structure variants such as single point mutations, RP-HPLC peptide mapping plays a key role owing to the high reproducibility of peptide retention times, determined as identity indexes. Because of the possible co-elution of some proteolytic fragments, an improvement of the array of information given by the technique can be achieved by setting up a series of experiments under hydrolytic conditions with different enzymes, followed by appropriate RP-HPLC gradient elutions. Such an experimental approach appears to be particularly useful in the examination of proteins with a high molecular weight, where the resulting RP-HPLC maps are complex. Therefore different RP-HPLC peptide mapping methods have been studied for recombinant human pro-urokinase (r-h-proUK), a thrombolytic agent of apparent molecular weight of 46 kD. The RP-HPLC maps indicate that the methods developed are not only suitable for the qualitative control of the amino acid sequence and arrangement of disulphide bonds but also represent the first demonstration of the identity of the primary structure of the recombinant and of the native species, within the limits of the technique.

Keywords: *Recombinant human pro-urokinase; extractive human pro-urokinase; primary structure; reversed-phase liquid chromatography; enzymatic hydrolysis; peptide mapping.*

Introduction

According to the existing regulations the structure of a recombinant protein should be fully elucidated and compared with that of the native species; in addition its identity should be confirmed for all the subsequent production cycles. Among the wide variety of complementary techniques that can be used to these purposes, RP-HPLC peptide mapping plays a key role [1, 2].

An RP-HPLC map is achieved by hydrolysis of the protein with a proteolytic enzyme and by subsequent gradient elution of the resulting peptides. Even if RP-HPLC peptide mapping is considered a specific method of identification because of the high reproducibility of the proteolytic fragment retention times, the detection limits of variants or other modified products still represent a challenge, being equal to 5% (mole:mole) for a new peak appearing in a map and 15% (mole:mole) for coeluting peaks [3]. Although careful method validation is suggested to cope with problems such as reproducibility of the enzymatic digestion and of the chromatographic conditions, the identity of each sample can be reliably proved by comparison with the reference standard preparation analysed at the same time.

Although RP-HPLC mapping cannot provide information about the identity of each peak, the technique is also applied with the aim of demonstrating the structural identity of the rDNA product with its naturally occurring counterpart. It can be done either for comparison purposes or for structure demonstration by applying appropriate enzymatic cleavage strategies followed by the isolation and sequence determination of the protein fragments.

In this paper the development of different peptide mapping procedures is described with the aim of comparing different production batches of recombinant human pro-urokinase (r-h-proUK) obtained by expression in *Escherichia coli*. The protein, refolded after extraction from *E. coli* and purified to a clinical grade product, is as biologically active as the naturally occurring human pro-Urokinase (h-proUK) and unglycosylated because of the features of the host organism.

*Presented in a preliminary form at the '10th International Symposium on HPLC of Proteins, Peptides and Polynucleotides', Wiesbaden, 29–31 October 1990.

†Author to whom correspondence should be addressed.

H-proUK, analysed for purposes of comparison with the recombinant form, is a serine protease enzyme, consisting of 411 amino acid residues, glycosylated at Asn[302], with an apparent molecular weight equal to 54 kD; the protein has been fully sequenced [4, 5] and the active site has been described [6] and shown to be within the 159–411 AA region, known as the 'serine protease domain' (Fig. 1). In both proteins 12 disulphide bonds are present, whose position should be the same as evidence of a correct refolding of the recombinant form.

The biological properties of h-proUK can be summarized by considering it to be the proenzymatic form of urokinase, endowed with fibrin-selective thrombolytic properties [7, 8].

For the purpose of verifying both the amino acid sequence and the arrangement of disulphide bonds of r-h-proUK batches and their identity in comparison with h-proUK, appropriate methods for peptide mapping analysis have been developed. Trypsin, which specifically cleaves proteins at the carboxy-terminal side of arginine and lysine residues, and pepsin, which specifically cleaves proteins at the phenylalanine, lysine and non-polar residues, have been used as proteolytic enzymes. Pepsin is recommended for digesting proteins in their native form because the low pH of the digestion buffer does not usually allow rearrangement of disulphide bonds.

Tryptic hydrolysis has been carried out both before and after reduction of the disulphide bonds. Under non-reducing conditions the

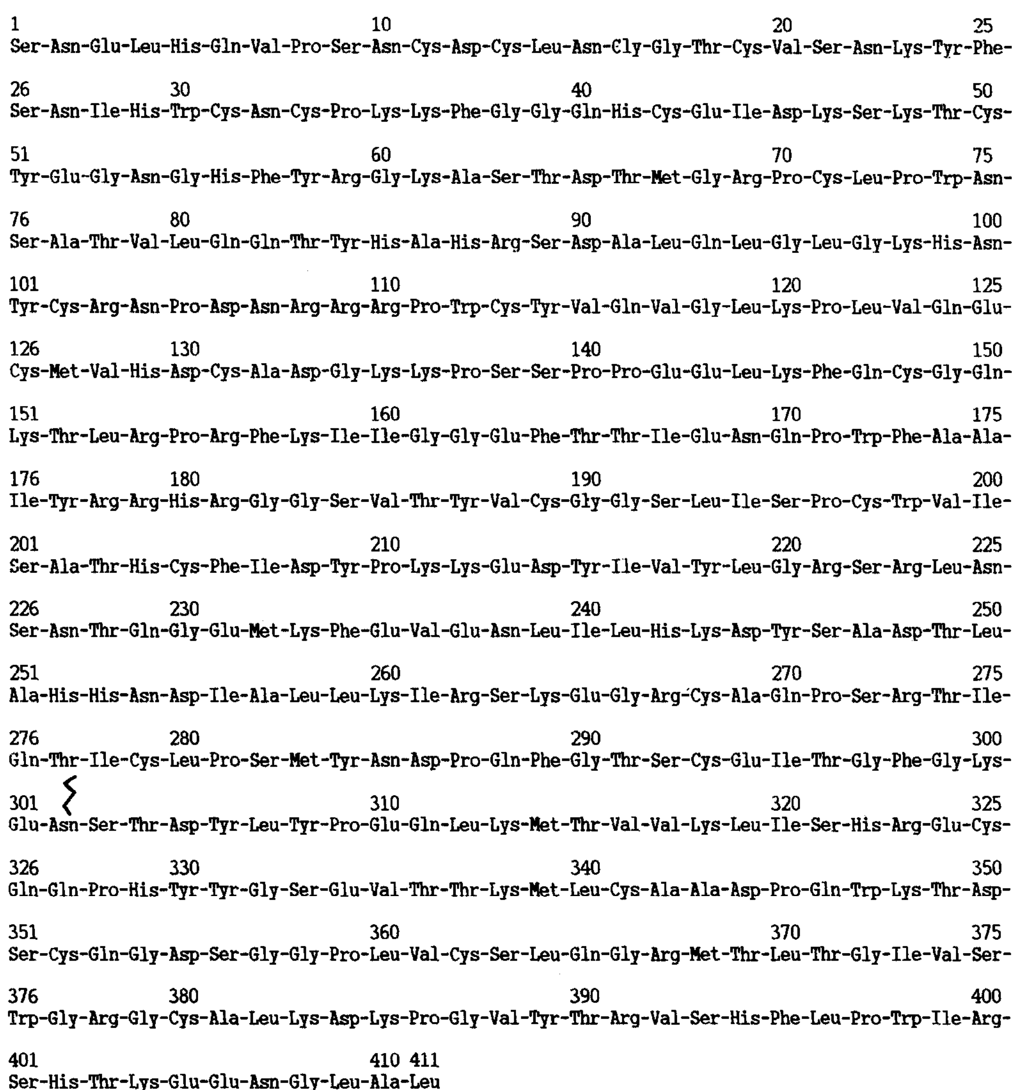

Figure 1
Primary structure of human pro-urokinase. The putative location of the disulphide bonds has been described [6]; the position of the glycosidic chain is indicated by a zig-zag line.

fragmentation pattern of the protein is influenced by its conformation, as only the amino acid residues exposed on the protein surface are accessible to the enzyme active site; in contrast, reduction of the disulphide bonds and the subsequent protection of the thiol groups by a derivatizing agent allows trypsin to cleave at all predictable sites, originating a chromatographic pattern with a characteristic number of fragments.

Preliminary evaluation of the reproducibility of the method demonstrates that the highest variability is associated with the enzymatic digestion step.

In general, the enzymatic activity can vary from batch to batch and, particularly in a tryptic map, additional peaks may be introduced because of chimotrypsin-like cleavages and of the partial hydrolysis of trypsin to ψ-trypsin. Therefore each map presented is related to that of a r-h-proUK preparation simultaneously analysed with a sample of h-proUK for the purpose of showing that under each experimental condition the identity of the two species is verified.

Experimental

Materials

Recombinant human pro-urokinase was a product of clinical grade purity (Farmitalia Carlo Erba); h-proUK was derived from mammalian cells (Sandoz, Vienna). β-Mercaptoethanol (electrophoresis purity grade — Biorad), 4-vinylpyridine (Sigma), guanidine hydrochloride (Sigma), Tris (electrophoresis purity grade — Biorad), urea (electrophoresis purity grade — Biorad), ammonium acetate (RPE — Carlo-Erba), HPLC-grade trifluoroacetic acid (TFA, Pierce), calcium chloride (RPE — Carlo-Erba) analytical grade reagents and HPLC-grade water purified by use of a Milli-Q System (Millipore) were used. The enzymes were diphenylcarbamyl chloride (DPCC)-treated trypsin from bovine pancreas (Sigma) and pepsin from porcine gastric mucosa (Boehringer).

Sample pretreatment

r-h-proUK has been enzymatically digested both in the native and in the reduced and derivatized forms. The derivatized form was obtained by addition of an excess of β-mercaptoethanol (10% v/v in water) followed by incubation for 2 h in the dark; 4-vinylpyridine was subsequently added under a stream of nitrogen and incubated for 2 h in the dark at room temperature, according to the method described by Friedman *et al.* [9]. The reaction reagents were removed by RP-HPLC purification on a Hewlett–Packard 1090M liquid chromatograph, using a C_4-Vydac column (250 × 4.6 mm i.d.; pore size 300 Å) under linear gradient elution conditions. Eluent A was 0.1% v/v TFA in H_2O and eluent B was acetonitrile–water (95:5, v/v, 0.07% TFA); the elution gradient was linear from 10 to 40% of eluent B in 15 min. The collected fraction, corresponding to the pyridylethylated proUK peak, was freeze-dried using Speed Vac A-160 Savant equipment before dissolution in the digestion buffer.

Tryptic digestion

Two different digestion procedures were applied using different enzyme:substrate ratios and hydrolysis times. In both cases approximately 200 μg of pyridylethylated or native protein were completely dissolved in a solution of 1% w/v ammonium bicarbonate, 2 M urea and 0.1 mM $CaCl_2$ (pH 8). In the first experiment trypsin was added at an enzyme:substrate (*E*:*S*) ratio of 1:25 and the hydrolysis was stopped after 18 h.

A second experimental procedure comprised the addition of trypsin at a four-fold higher *E*:*S* ratio (1:6.25) and digestion for 48 h.

Peptic digestion

The same amount of r-h-proUK in its native form was digested in 5% acetic acid at 37°C for 24 h, by adding the enzyme twice, at an *E*:*S* ratio of 1:16.5.

RP-HPLC peptide maps

RP-HPLC separation of the peptides formed by enzymatic digestion was performed on a Hewlett–Packard 1090M liquid chromatograph equipped with a 79994A data station. A 10-μm Micropack C18 column (Varian) was used (300 × 4.6 mm i.d.; pore size 300 Å) under linear gradient elution. Eluent A was 0.1% TFA v/v in water and eluent B was acetonitrile–water (95:5, v/v, 0.07% TFA); the elution gradient was linear from 5 to 30% of eluent B in 130 min (tryptic maps) and 140 min (peptic maps). The maps were recorded at 220 nm by diode-array UV detection.

Reduction of protein digests

For the purpose of reducing intact disulphide bridges within the peptides formed by trypsin and pepsin digestion of the recombinant protein under non-reducing conditions, 1 μl of diluted β-mercaptoethanol (10% v/v in water) was added to approximately 200 μg of protein digest; the samples were then incubated for 1 h at 37°C.

Results

Peptide mapping is considered to be one of the most specific methods for identifying changes in protein structure [1]; additional techniques such as carboxy-terminal sequencing, amino-terminal sequencing and mass spectrometry are needed to identify each peptide. However if the native protein has been fully sequenced, the RP-HPLC maps can be successfully used for structural identification, with detection limits for impurities in the range of 5–15% (mole:mole) [3]. The results of different enzymatic digestion protocols carried out with the aim of comparing r-h-proUK and h-proUK are described in the sections below.

Tryptic maps under reducing conditions

Figures 2 and 3 show comparisons of r-h-proUK and h-proUK tryptic maps, obtained after 18 and 48 h, respectively, of enzymatic hydrolysis. In Fig. 2 the only difference be-

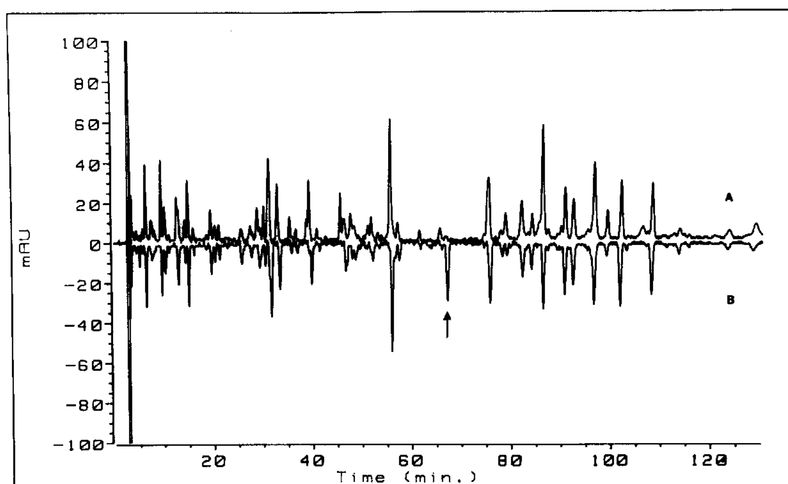

Figure 2
Tryptic patterns of human pro-urokinase (A) in comparison with recombinant human pro-urokinase (B) under reducing conditions (E:S = 1:25, 18 h, 37°C).

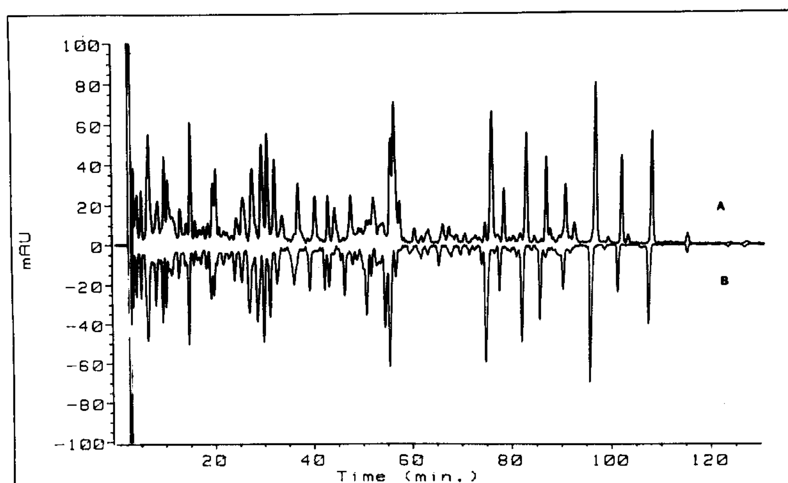

Figure 3
Tryptic patterns of recombinant human pro-urokinase (A) in comparison with human pro-urokinase (B) under reducing conditions (E:S = 1:6.25, 48 h, 37°C).

tween the two maps is represented by a missing peak for h-proUK at a retention time equal to 68 min; in Fig. 3 no relevant differences can be observed between the two RP-HPLC patterns.

Tryptic maps under non-reducing conditions

In Figs 4 and 5 the RP-HPLC chromatograms resulting from 18 and 48 h digestions under non-reducing conditions show different profiles only if limited proteolysis is carried out. An additional peak with retention time equal to 68 min is present in the r-h-proUK patterns both under reducing and under non-reducing conditions. This behaviour may be related to a fragment which does not contain cysteine residues as no differences in the retention time values are observed between the two chromatograms; moreover, as this peptide is generated also under non-reducing conditions, its trypsin cleavage sites are exposed on the protein surface. On the basis of this result the expected peptide should be the Glu^{301}–Lys^{313}, which is absent in the pattern of h-proUK because of the presence of the sugar portion, linked at Asn^{302}, which could influence trypsin cleavage at Lys^{300}.

To gain further structural evidence, amino-terminal sequence analysis has been performed on the collected peaks, with retention time 68 min, derived from the maps under reducing the non-reducing conditions (Figs 2 and 4). An automated Edman degradation followed by

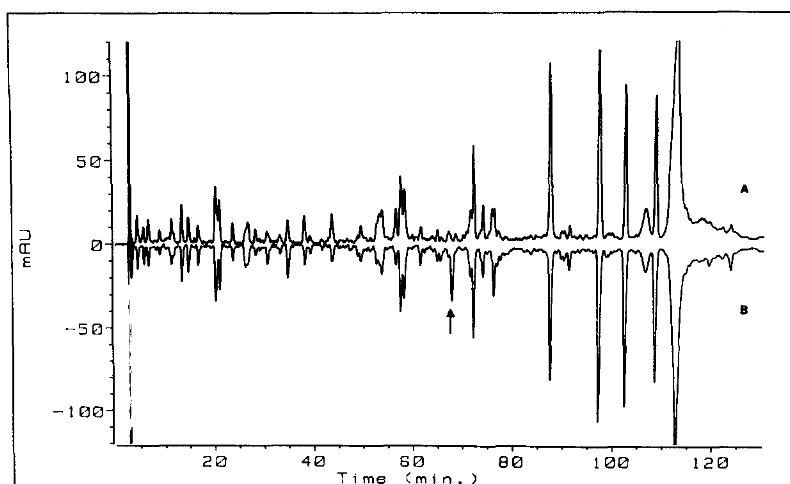

Figure 4
Tryptic patterns of human pro-urokinase (A) in comparison with recombinant human pro-urokinase (B) under non-reducing conditions (E:S = 1:25, 18 h, 37°C).

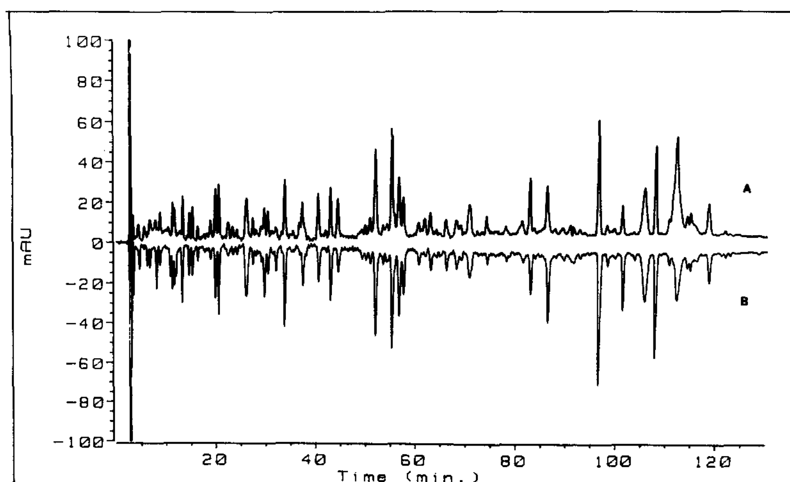

Figure 5
Tryptic patterns of human pro-urokinase (A) in comparison with recombinant human pro-urokinase (B) under non-reducing conditions (E:S = 1:6.25, 48 h, 37°C).

RP-HPLC gradient elution of the resulting PTH-amino acids was carried out on an Applied Biosystem 470 A model gas-phase sequencer. As expected, the results indicate Glu^{301}–Lys^{313} as the main sequence in both fragments.

The presence of intact disulphide bonds has been checked by β-mercaptoethanol reduction of the 48 h digest; in the resulting pattern (Fig. 6) an increase in the number of chromatographic peaks and substantial changes in the relative retention times, as a consequence of disulphide bridges reduction, are apparent. Although all the peaks should be identified to gain definitive evidence of the correct refolding of all the disulphide bridges, the qualitative power of this experiment is confirmed by the pepsin digestion described in the next section.

Peptic maps under non-reducing conditions

The digestion of proteins containing a high number of disulphide bridges such as r-h-proUK can be successfully obtained by the action of pepsin.

The peptic maps obtained for r-h-proUK and the extractive protein are comparable (Fig. 7) and the behaviour upon reduction of the digest by β-mercaptoethanol (Fig. 8) is the same as that encountered for the tryptic mapping. Therefore, the patterns obtained qualitatively indicate that most disulphide bonds are correctly arranged, although more complete evidence can only be obtained by sequencing or mass spectrometric examination of the isolated fragments.

Discussion and Conclusions

The results of this work are not only a first demonstration of the correct primary structure of recombinant human pro-urokinase, but also indicate that peptide mapping is a powerful tool for structural studies. Although the sensitivity of the method in respect of each impurity that might be present is in the range 5–15% (mole:mole), the equivalence of the peptide patterns is considered to be a specific index of structural identity. Nevertheless amino-terminal sequencing was needed to obtain conclusive evidence about the identity of the additional peak found for r-h-proUK under the mildest hydrolytic conditions (Figs 2 and 4). The amino acid sequence of the protein fragment corresponding to Glu^{301}–Lys^{313} for both peaks collected under reducing and non-reducing conditions confirms the role exerted by the glycosidic chain, which sterically hinders the trypsin action; therefore longer hydrolysis times are needed to obtain the same pattern for both proteins. Additional information can be gained, however, by considering the tryptic maps; the peptide Glu^{301}–Lys^{313} is located on the protein surface, as can be deduced by observing the peptide maps obtained under non-reducing conditions (Fig. 4). In h-proUK this peptide too is probably exposed as it contains the glycosidic chain, which is linked at Asn^{302}. This consideration supports once again the correct refolding of the serine protease domain of r-h-proUK, the 159–411 AA region, which contains the catalytically active residues His 204, Asp 255 and Ser 356.

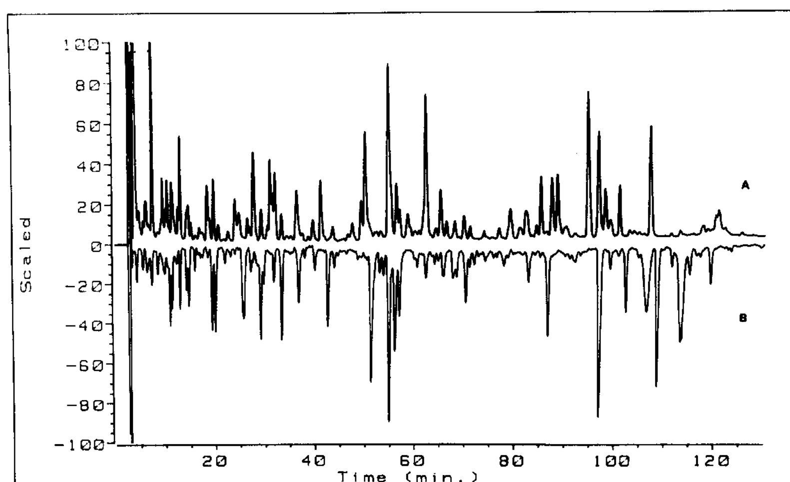

Figure 6
Tryptic patterns of recombinant human pro-urokinase (A) digested under non-reducing conditions (E:S = 1:6.25, 48 h, 37°C) in comparison with a sample (B) treated in the same way, but finally reduced by β-mercaptoethanol.

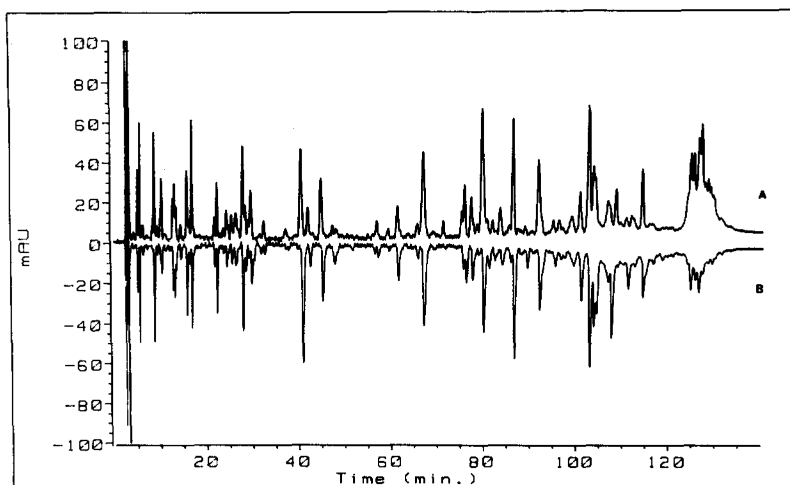

Figure 7
Peptic patterns of human pro-urokinase (A) in comparison with recombinant human pro-urokinase (B) under non-reducing conditions (E:S = 1:16.5, 24 h, 37°C).

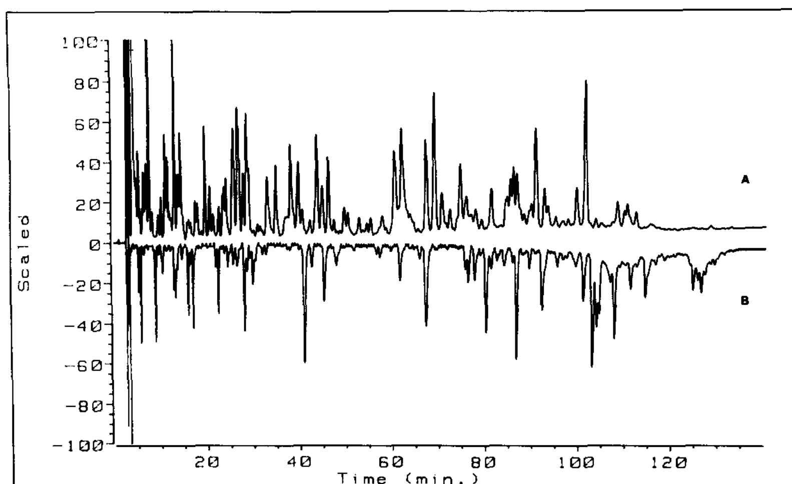

Figure 8
Peptic patterns of recombinant human pro-urokinase (A) digested under non-reducing conditions (E:S = 1:16.5, 24 h, 37°C) in comparison with a sample (B) treated in the same way, but finally reduced by β-mercaptoethanol.

Additional differences due to the slower hydrolysis of h-proUK compared to that of r-h-proUK could be expected such as the presence of a larger fragment for h-proUK (containing not only the glycosylated 301–313 sequence but also the adjacent uncleaved one) and of the peaks of the peptides with the sequences adjacent to the 301–313 sequence for r-h-proUK. However it cannot be excluded that both under reducing and under non-reducing conditions, the expected larger peak containing the uncleaved glycosylated peptide could have the same retention time of the smaller, but unglycosylated corresponding peptide derived from r-h-proUK. On the other hand, identification of the sequences adjacent to the fragment 310–313 for r-h-proUK is not possible on only the basis of the peak retention times and areas.

A possible reason for these apparently missing peaks could be the incomplete digestion of the two proteins after 18 h, the r-h-proUK small fragments being still under the detection limit of the technique; elution of these fragments in close proximity of peaks having similar polarity could be an alternative explanation. However any hypothesis cannot be verified without the identification of each chromatographic peak by sequencing methods.

In conclusion, application of the RP-HPLC

peptide mapping technique to the structural studies of r-h-proUK shows that the use of different enzymes and experimental conditions of hydrolysis is extremely important to provide evidence of the primary structure identity between proteins. In addition, differences between patterns derived from species having variable glycosylation features can arise if the glycosidic chain is located in close proximity to an enzymatic hydrolysis site; in this event, such as the example presented in this paper, a careful evaluation of the cleavage strategy plays a key role. The methods developed provide the basis for further structural studies aimed at the complete demonstration of the structure of the recombinant protein.

Acknowledgement — Thanks are due to G. Nitti for the *N*-terminal sequence determinations.

References

[1] R.L. Garnick, N.J. Solli and P.A. Papa, *Anal. Chem.* **60**, 2546–2557 (1988).
[2] F.E. Regnier, *BioPharm*, 19–21, January (1988).
[3] R. Chloupek, J.E. Battersby and S. Hancock, in *High Performance Liquid Chromatography of Peptides and Proteins* (C.T. Mant and R.S. Hodges), pp. 825–833. CRC Press, London (1991).
[4] W.A. Gunzler, G.J. Steffens, F. Otting, S.M.A. Kim, E. Frankus and L. Flohè, *Hoppe Seyler's Z. Physiol. Chem.* **363**, 1155–1165 (1982).
[5] G.J. Steffens, W.A. Gunzler, F. Otting, E. Frankus and L. Flohè, *Hoppe Seyler's Z. Physiol. Chem.* **363**, 1043–1058 (1982).
[6] T.J.R. Harris, *Protein Engng* **1**, 449–458 (1987).
[7] V. Gurewich, R. Pannell, S. Louie, P. Kelley, R.L. Suddith and R. Greenlee, *J. Clin. Invest.* **73**, 1731–1739 (1984).
[8] C. Zamarron, H.R. Lijnen, B. Van Hoef and D. Collen, *Thromb. Haemostas.* **52**, 19–23 (1984).
[9] M. Friedman, M. Krull and J.F. Cavins, *J. Biol. Chem.* **245**, 3868–3871 (1970).

[*Received for review 18 September 1992; revised manuscript received 7 January 1993*]