Journal of Chromatography, 447 (1988) 351–364
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 20 594

NARROW-BORE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY OF PHENYLTHIOCARBAMYL AMINO ACIDS AND CARBOXYPEPTIDASE P DIGESTION FOR PROTEIN C-TERMINAL SEQUENCE ANALYSIS

HSIENG S. LU*, MICHAEL L. KLEIN and POR-HSIUNG LAI
Amgen, 1900 Oak Terrace Lane, Thousand Oaks, CA 91320 (U.S.A.)
(First received March 25th, 1988; revised manuscript received April 29th, 1988)

SUMMARY

Carboxypeptidase P digestion followed by narrow-bore high-performance liquid chromatography of phenylthiocarbamyl amino acids is employed for polypeptide C-terminal end group and sequence determination. Carboxypeptidase P digestion of polypeptides provides specific cleavage of protein C-terminal amino acids. The digestion offers the advantage that it can be carried out in either 10 m*M* sodium acetate or water at pH 4.0 in the presence of an enzyme activator, Brij-35. The narrow-bore high-performance liquid chromatography of all 20 phenylthiocarbamyl-amino acids has provided quantitative analysis at low picomole levels. This efficient and sensitive procedure is particularly useful for examining *in vivo* excision of protein C-termini and for verifying the integrity of various protein products produced by recombinant DNA techniques.

INTRODUCTION

Analyses of polypeptide C-terminal groups and sequences provide important information on identity and purity of a polypeptide, and on posttranslational processing at the C-terminal end. Methods involving digestion of polypeptides by carboxypeptidases (Cpases) and subsequent identification of the released amino acids by ion-exchange chromatography have often been employed for such analyses[1,2]. This method offers advantages over other approaches which require the use of a specific instrument or multi-step analyses, such as fast atom bombardment (FAB) mapping by mass spectrometry and isolation and characterization of the C-terminal peptide[3,4], in that the Cpase digestion method is performed directly with the intact proteins and the digestion kinetics are determined by time course studies. Chemical degradations, such as hydrazinolysis[5] or the tritium labeling method[6], can also be employed for the C-terminal amino acid determination of proteins; however, these procedures are less sensitive, laborious, and can only release one single C-terminal end group. Although sequential chemical degradations of proteins from the C-terminus using different reactions were reported previously[7,8], these degradations require considerably larger amounts of polypeptide and have not been made routine.

0021-9673/88/$03.50

Cpases A (EC 3.4.12.2) and B (EC 3.4.12.3) from bovine or porcine pancreas and Cpase Y (EC 3.4.12.–) from Baker's yeast have been widely used in the C-terminal amino acid determination[9–11]. However, these enzymes have limited uses due to their narrow specificity for end groups or poor ability to digest a broader range of protein substrates. Recently, an acid carboxypeptidase, Cpase C (EC 3.4.12.1), which was isolated from citrus fruits[13,14], was shown to exhibit broader specificity to sequentially remove various C-terminal residues at various cleavage rates. One drawback of using this Cpase, however, is endopeptidase contamination in the commercial preparations; their further purification may become necessary. On the other hand, Cpase P (EC 3.4.12.–). has been purified to homogeneity from *Penicillium janthinellum*[15,16] and has been observed to exhibit broader specificity for various C-terminal amino acids, including those resistant to Cpase A, B, or Y cleavage. However, Cpase P is not widely used for C-terminal sequence analysis since the specificity of the enzyme is not well understood. The lower detection sensitivity of conventional analytical methods for amino acid analysis also hampers the application of the enzymatic methods for C-terminal analysis.

In this communication, we describe the use of narrow-bore high-performance liquid chromatography (HPLC) of phenylthiocarbamyl (PTC)-amino acids to determine the composition of Cpase P digestion products for protein C-terminal end group and sequence determinations. The narrow-bore HPLC of PTC-amino acids allows resolution of all twenty PTC-amino acids including asparagine and glutamine such that the released end groups can be identified and quantified. The present procedure is able to obtain useful C-terminal sequence information by high sensitivity analysis of amino acids released from limited Cpase P digestion of various proteins and peptides.

MATERIALS AND METHODS

Materials

Cpase P (Cat. No. C-5396), hen egg lysozyme and Brij-35 were obtained from Sigma. Cpase P can also be purchased from Boehringer Mannheim. Recombinant interleukin-2 (IL-2) (Ala-125), granulocyte colony stimulating factor (G-CSF), granulocyte macrophage colony stimulation factor (GM-CSF), erythropoietin and consensus α-interferon were research products of Amgen and produced by genetically modified *Escherichia coli* except that recombinant erythropoietin was produced by genetically engineered Chinese hamster ovary cells. Bradykinin was a synthetic peptide purchased from Serva. PD-10 columns, containing prepacked Sephadex G-25, were obtained from Pharmacia. The amino acid sequences of these proteins and peptides at the carboxyl terminus are listed in Table I.

Methods

HPLC apparatus. A Hewlett-Packard microbore LC system (HP1090) was used for the identification of PTC-amino acids. The system is equipped with an HP autosampler and a diode array detector for automatic analysis. A Nelson Analytical 4400 system using XTRA-CHROM software is connected to the chromatographic system for data acquisition and processing.

Phenylisothiocyanate (PITC) derivatization of amino acids and separation of

TABLE I

C-TERMINAL AMINO ACID SEQUENCES OF POLYPEPTIDES STUDIED IN THIS EXPERIMENT

The one-letter code for amino acids is as follows: A = alanine; C = cysteine; D = aspartate; E = glutamate; F = phenylalanine; G = glycine; H = histidine; I = isoleucine; K = lysine; L = leusine; M = methionine; N = asparagine; P = proline; Q = glutamine; R = arginine; S = serine; T = threonine; V = valine; W = tryptophan; Y = tyrosine.

Polypeptides	*Sequences**
	160 174
Human rG-CSF	···F-L-E-V-S-Y-R-V-L-R-H-L-A-Q-P
	120 127
Human rGM-CSF	···D-C‡-W-E-P-V-Q-E
	125 133
Human rIL-2(Ala-125)	···A-Q-S-I-I-S-T-L-T
	158 165
Human recombinant erythropoietin	···G-E-A-C‡-R-T-G-D-[R]**
	157 166
Consensus α-interferon	···N-L-Q-E-R-L-R-R-K-E
Henn egg lysozyme	···C‡-R-L
	1 9
Bradykinin	R-P-P-G-F-S-P-F-R

* C‡ stands for cysteine involved in disulfide bond formation.

** Arg in the bracket is predicted from gene sequence of erythropoietin but not detected by C-terminal amino acid analysis of the hormone.

PTC-amino acids. PITC coupling of standard amino acids or samples obtained from Cpase P digestion was performed according to our previously reported procedures[17]. Separation of the derivatized PTC-amino acids was then carried out by reversed-phase HPLC using an Altex narrow-bore C_{18} column (25 × 0.2 cm) and a gradient of solvent B in solvent A. Solvent A was 25 m*M* sodium acetate (pH adjusted to 5.0 with phosphoric acid). Solvent B was 25 m*M* sodium acetate (pH adjusted to 6.5 with phosphoric acid)–acetonitrile–methanol (40:50:10). The column was equilibrated in A–B (90:10) and operated at 40°C using a flow-rate of 0.25 ml/min. Separation was achieved using a linear gradient of 10 to 11% B in 1 min, 11 to 47% B in 14 min, 47 to 55% B in 5 min, 55 to 85% B in 1 min, 85 to 100% B in 1 min and followed by an isocratic elution at 100% B for 3 min. Triethylamine (100–200 μl) can be added to solvent A if PTC-Arg is difficult to resolve from PTC-Ala. The PTC-amino acid derivatives were detected at 265 nm.

Cpase A, B and Y digestion. A suspension of diisopropyl phosphorofluoridate (DFP)-treated Cpase A (Sigma) was repeatedly washed with distilled water to remove contaminating free amino acids[9]. The enzyme was then collected by centrifugation and redissolved in 0.2 *M* sodium bicarbonate before use. DFP-treated Cpase B (Sigma) and Cpase Y (Pierce) were used without further purification. Cpase A and/or B digestion was performed in 0.1 *M* N-methylmorpholine (pH 7.6) using an enzyme-to-substrate ratio of 1:100 to 1:50 at 37°C. Cpase Y digestion was carried out in 0.1

M ammonium acetate (pH 5.5) at 37°C using an enzyme-to-substrate ratio of 1:100. A combined digestion by Cpases Y and B was performed at two different pH values (pH 5.5 and 7.6) according to Hayashi[12].

Cpase P digestion. Protein substrates (1–4 nmol) were concentrated by centrifugation using a Centricon-10 (Amicon) microconcentrator with a 10-kilodalton (kD) molecular weight cut-off membrane, and buffer salts were removed by gel filtration over a Pharmacia PD-10 column equilibrated in 10 m*M* sodium acetate, pH 4.0 or distilled water. In the latter condition, the pH of the desalted sample solution was adjusted to 4.0 by diluted hydrochloric acid. Before adding enzyme, a concentrated solution of Brij-35 was added to the protein sample to a final concentration of 0.05% (w/v). The Cpase P digestion conditions varied according to the rate of amino acid release obtained from digestion of different substrates. The normal digestion was performed using an enzyme-to-substrate ratio of 1:400 (w/w) at 25°C in 1 h. When such digestion conditions generated a rapid release rate, the reaction could be slowed down by incubating the sample at 4°C. For samples which were difficult to digest, an enzyme-to-substrate ratio of 1:100 for digestion at 37°C in several hours was used. After digestion, samples at various time points including zero time were taken and immediately acidified with trifluoroacetic acid at a final concentration of 10% to stop the digestion. Acidified samples were dried *in vacuo* by a Speedvac centrifuge and subjected to PITC derivatization and subsequent HPLC analysis as described above. Norleucine (1–2 nmol) was usually added to sample solutions prior to digestion as an internal standard to calibrate recovery of amino acids.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of Cpase P digestion products. Aliquots of Cpase P digestion products at specific time intervals were added to SDS-PAGE sample buffer. The mixtures were immediately incubated at 65°C for 30 min or at 100°C for 5 min to stop Cpase P digestion. The samples were then subjected to SDS-PAGE according to Laemmli[18].

RESULTS AND DISCUSSION

Narrow-bore HPLC of PTC-amino acids

Analysis of amino acids released from Cpase digestion of polypeptides requires a chromatographic procedure that can offer a complete separation of all 20 common amino acids including asparagine, glutamine and tryptophan. The conventional ion-exchange chromatography[19] or HPLC of the precolumn derivatized amino acids[20,21] is usually employed for analysis of samples derived from acid hydrolysis of polypeptides and thus excludes the separation of asparagine and glutamine. Although the ion-exchange chromatography used for the separation of physiological amino acids[22] is well established for C-terminal end group and sequence analysis using Cpase digestion, this method is limited by its lower sensitivity of detection. An improved HPLC procedure using precolumn PITC derivatization and subsequent narrow-bore HPLC was developed to enhance the sensitivity of analysis. Fig. 1 illustrates results of the chromatographic analysis of a standard amino acid mixture at levels of 4 pmol (chromatogram A) and 40 pmol (chromatogram B). The chromatographic conditions allow the separation of all 20 common amino acids. The use of mobile phase A at lower pH effectively resolves Asn from Ser, and Gln from Gly, which otherwise coelute using conditions reported previously[17,20,21]. The addition

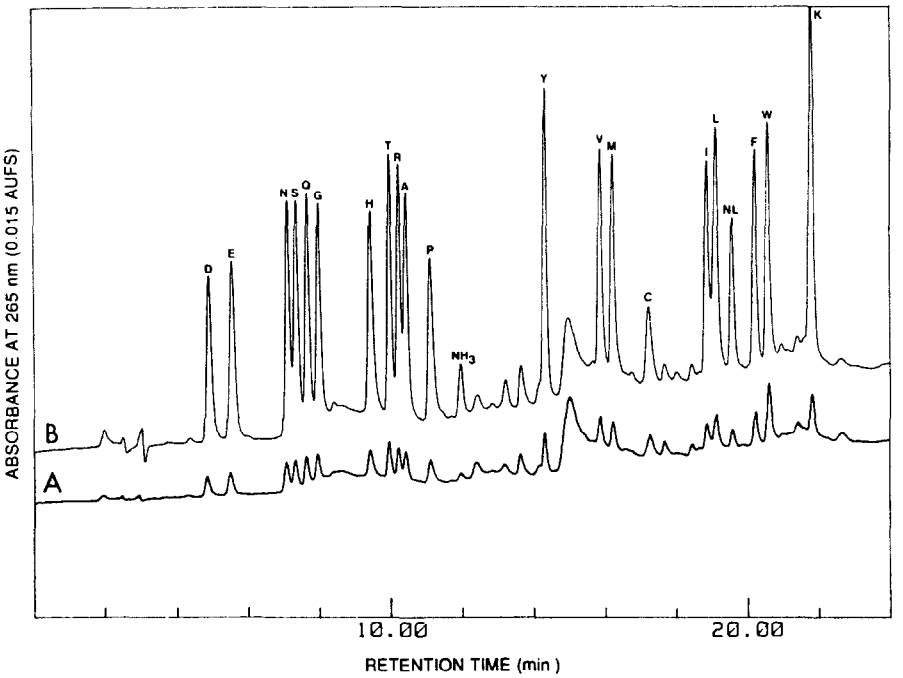

Fig. 1. HPLC separation of 21 PTC-amino acids using an Altex C_{18} column (25 × 0.2 cm, 5 μm). (A) 4 pmol each of standard amino acids including Asn, Gln, Trp and norleucine. (B) 40 pmol of standard.

of triethylamine in mobile phase A facilitates the separation of Arg from Ala. Norleucine completely separates from Leu and Phe, and can be used as an internal standard to evaluate accuracy and reproducibility of PITC coupling and HPLC.

Cpase digestion and amino acid analysis of the digestion products

Cpase A and B as well as Cpase Y digestion have been mostly cited for polypeptide C-terminal analysis[9–11]. Their application in analysis of small peptide sequences was frequently successful. However, it may become more difficult to cleave some of the higher molecular weight protein substrates by these enzymes.

Many lymphokines and hemopoietic factors have been reproduced for human therapeutic uses by the newly developed recombinant DNA technologies. The structural analysis of these highly purified proteins is usually undertaken as a requirement for product application in human clinical use. The purpose of such studies is to confirm the intactness of the molecule, especially at both amino and carboxyl terminal ends. In spite of the successful analysis of protein amino terminal sequences by automated Edman degradation chemistry, we had experienced unsuccessful uses of Cpases A, B and Y to perform C-terminal analyses of many recombinant proteins such as human rIL-2(Ala-125), consensus α-interferon, and erythropoietin (for details on their C-terminal sequences, see Table I) due to specificity and lower activity of these enzymes. Very little information can be obtained even with a digestion using two Cpases together (*i.e.*, Cpases A and B, Cpases B and Y). Moreover, the digestion conditions recommended for these enzymes include buffers containing high concen-

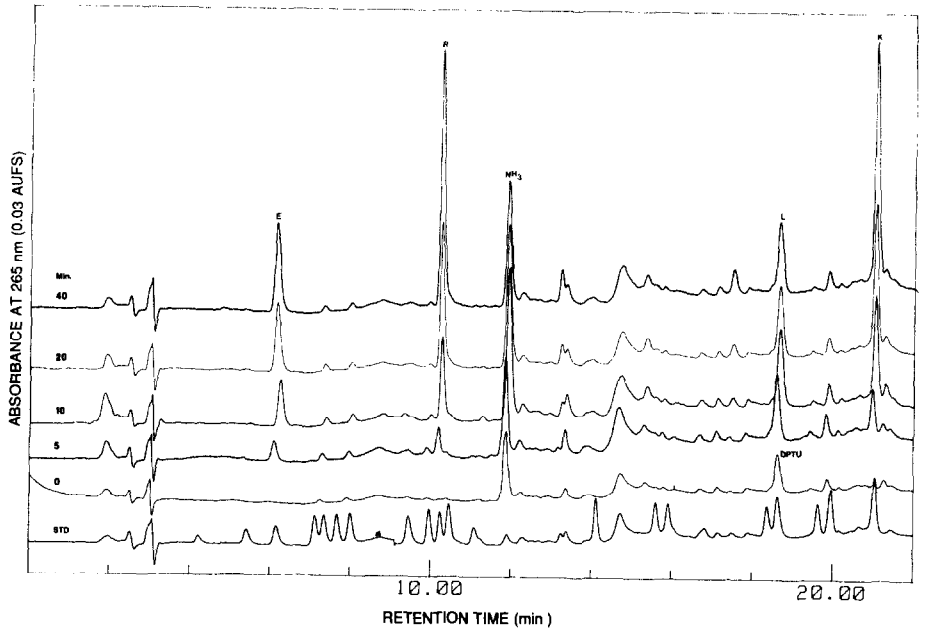

Fig. 2. HPLC analysis of amino acids released from Cpase P digestion of r-consensus α-interferon at various time intervals. Sample (100 pmol) at each time point was subject to analysis after PITC derivatization.

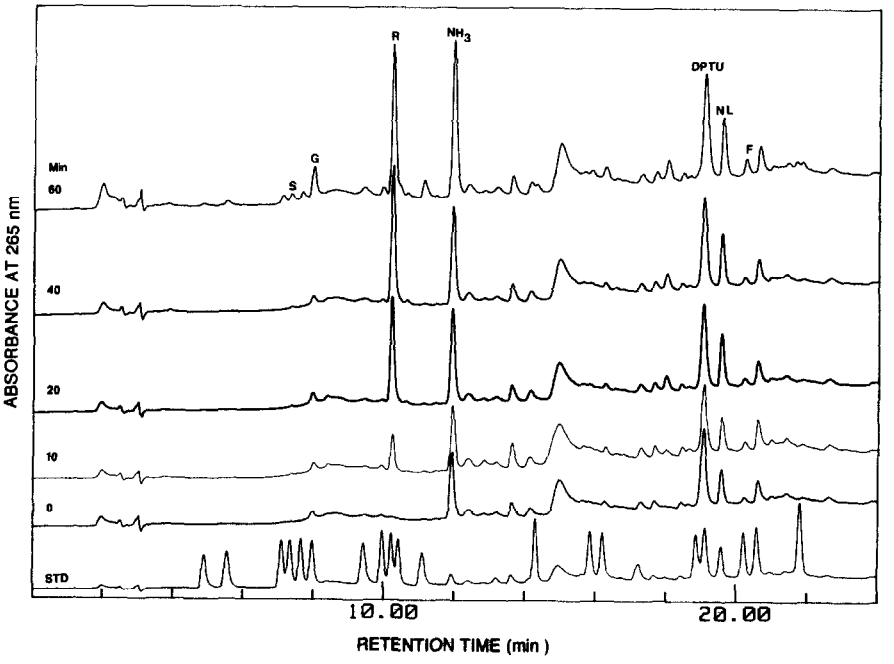

Fig. 3. HPLC analysis of amino acids released from Cpase P digestion of bradykinin at various time intervals. For details, see legend of Fig. 2.

tration of salts, and are not compatible for subsequent PITC derivatization, as buffer salts seriously interfere with the derivatization efficiency and removal of side products during the drying step. These difficulties led us to carry out protein C-terminal analysis using Cpase P digestion in conjunction with narrow bore HPLC.

Fig. 2 shows chromatograms obtained from HPLC of PITC-derivatized digests of r-consensus α-interferon taken at 0, 5, 10, 20 and 40 min. The reaction was carried out using an enzyme-to-substrate ratio of 1:400 at 4°C. At zero time, no amino acid was released by the enzyme, indicating that the enzyme preparation is free from any contaminating amino acids, and that the trifluoroacetic acid (TFA) acidification effectively terminated the digestion. It is apparent that four amino acids, *i.e.*, Glu, Arg, Leu and Lys could be detected and quantified from a cleavage of five consecutive C-terminal amino acids within 40 min. Fig. 3 illustrates the analysis obtained from the digestion of bradykinin. Only one C-terminal Arg was released from digestion of this 9-residue synthetic peptide.

Table II summarizes the quantitative recovery of released amino acids at different time points for proteins and peptides which were treated with Cpase P. All of the samples used in this study are clearly digested by the enzyme at various rates. It is apparent that Cpase P is a very efficient enzyme capable of cleaving most peptide bonds. We found that the appropiate conditions for a successful digestion are to use normal digestion conditions for several time intervals within 1 h, as illustrated in the digestion of human recombinant G-CSF, GM-CSF and erythropoietin (Table II). In the example for the digestion of r-consensus α-interferon, such conditions generated very fast kinetics (results not shown). The experiment was thus conducted at 4°C within 60 min. Bradykinin, human rIL-2(Ala-125), hen egg lysozyme and human r-erythropoietin are more resistant to Cpase P digestion; thus, the enzyme-to-substrate ratio was adjusted to 1:100 and the digestion performed at 37°C for several hours to obtain a good cleavage rate.

For digestion of samples which only release one or two amino acid residue(s), the C-terminal amino acids can be directly identified, as is the case in the digestion of lysozyme and bradykinin for single amino acid release (Leu and Arg, respectively), and rIL-2(Ala-125) as well as rGM-CSF for release of two amino acids (–Leu–Thr and –Gln–Glu, respectively). However, the assignment of the C-terminal sequence for samples with rapid digestion requires quantitative kinetic analysis. Fig. 4 (top) illustrates the plot of amino acid recovery at various time points for r-consensus α-interferon. It is clear that the C-terminal sequence of the recombinant lymphokine can be assigned as ···Leu–Arg–Arg–Lys–Glu. Arginine is released in two residues and its initial release rate is slower than glutamate but faster than Leu. In the case of recombinant G-CSF (Fig. 4, bottom), the first four amino acids can be assigned at 5–10 min digestion (···Leu–Ala–Gln–Pro). At the 30-min time point, more amino acids are released with no further sequence assignment, since histidine at the 5th position is released much more slowly than arginine at the 6th position. However, at least 10 amino acids from the C-terminus up to Tyr-165 are clearly cleaved. The released C-terminal amino acids up to the 10th residue is consistent with the sequence predicted from the cDNA sequence. The mis-assignment at the 5th and 6th positions for histidine and arginine is due to the slower release rate of the His residue. The above analyses indicate that fast kinetic release of amino acids for an unknown protein digestion should be carefully interpreted to prevent mis-assignment.

TABLE II

RECOVERY OF RELEASED AMINO ACIDS FROM CARBOXYPEPTIDASE P DIGESTION OF VARIOUS PROTEIN AND PEPTIDE SUBSTRATES*

(1) r-Consensus α-interferon (enzyme-to-substrate = 1:400 at 4°C)
C-terminal sequence: ···N-L-Q-E-R-L-R-R-K-E (157 … 166)

	Time (min)				
	0	*5*	*10*	*20*	*40*
Glu	0	0.18	0.48	0.75	0.86
Arg	0	0.17	0.57	1.15	1.80
Leu	0	0.11	0.12	0.19	0.23
Lys	0	0.14	0.38	0.53	0.78

(2) Human r-G-CSF (enzyme-to-substrate = 1:400 at 25°C)
C-terminal sequence: ···F-L-E-V-S-Y-R-V-L-R-H-L-A-Q-P (160 … 174)

	Time (min)				
	0	*5*	*10*	*15*	*30*
Gln	0	0.06	0.32	0.74	0.93
His	0	0	0.13	0.41	0.57
Arg	0	0	0.24	0.88	1.45
Ala	0	0.04	0.31	0.91	1.02
Pro	0	0.12	0.55	0.96	1.17
Tyr	0	0	0.02	0.06	0.20
Val	0	0	0.02	0.27	0.59
Leu	0	0.02	0.32	1.30	1.60

(3) Bradkinin (enzyme-to-substrate = 1:100 at 37°C)
Amino acid sequence: R-P-P-G-F-S-P-F-R

	Time (min)				
	0	*10*	*20*	*40*	*60*
Arg	0	0.23	0.72	0.85	0.88
Gly	Trace	0.04	0.04	0.05	0.14
Phe	Trace	0.02	0.02	0.04	0.05

(4) Hen egg lysozyme (enzyme-to-substrate = 1:100 at 37°C)
*C-terminal sequence: ···*C-R-L**

	Time (h)			
	0	*1*	*2*	*4*
Leu	0	0.38	0.39	0.44

TABLE II (*continued*)

(5) Human r-interleukin-2(Ala-125) (A: enzyme-to-substrate = 1:400 at 25°C) (B: enzyme-to-substrate = 1:100 at 37°C)

125 133

C-terminal sequence: ···A-Q-S-I-I-S-T-L-T

(A)	*Time (min)*				
	0	*5*	*20*	*40*	*60*
Thr	0	0.09	0.11	0.13	0.18
Leu	0	0	Trace	Trace	Trace

(B)	*Time (h)*			
	0	*1*	*2*	*4*
Thr	0	0.64	0.85	0.91
Leu	0	0.61	0.81	0.90

(6) Human rGM-CSF (enzyme-to-substrate = 1:400 at 25°C)

120 127

C-terminal sequence: ···D-C‡-W-E-P-V-Q-E

	Time (min)				
	0	*10*	*20*	*40*	*60*
Glu	0	0.18	0.75	0.86	0.92
Gln	0	0.15	0.13	0.29	0.40

(7) Human r-erythropoietin (enzyme-to-substrate = 1:00 at 37°C)

158 165

*C-terminal sequence: ···G-E-A-C‡-R-T-G-D-[R]****

	Time (min)				
	0	*10*	*20*	*40*	*60*
Asp	0	0.09	0.20	0.43	0.57
Gly	0	0.07	0.18	0.34	0.51
Thr	0	0.05	0.08	0.20	0.30
Arg	0	0.01	0.02	0.08	0.12
Ala	0	0.01	0.02	0.08	0.10

* The numbers are expressed in molar recovery of released amino acids. Digestion of all samples was performed in 10 m*M* sodium acetate, pH 4.0, except recombinant consensus α-interferon and r-erythropoietin, which were done in water, pH 4.0.

** C‡ stands for cysteine involved in disulfide bond formation.

*** See footnotes in Table I.

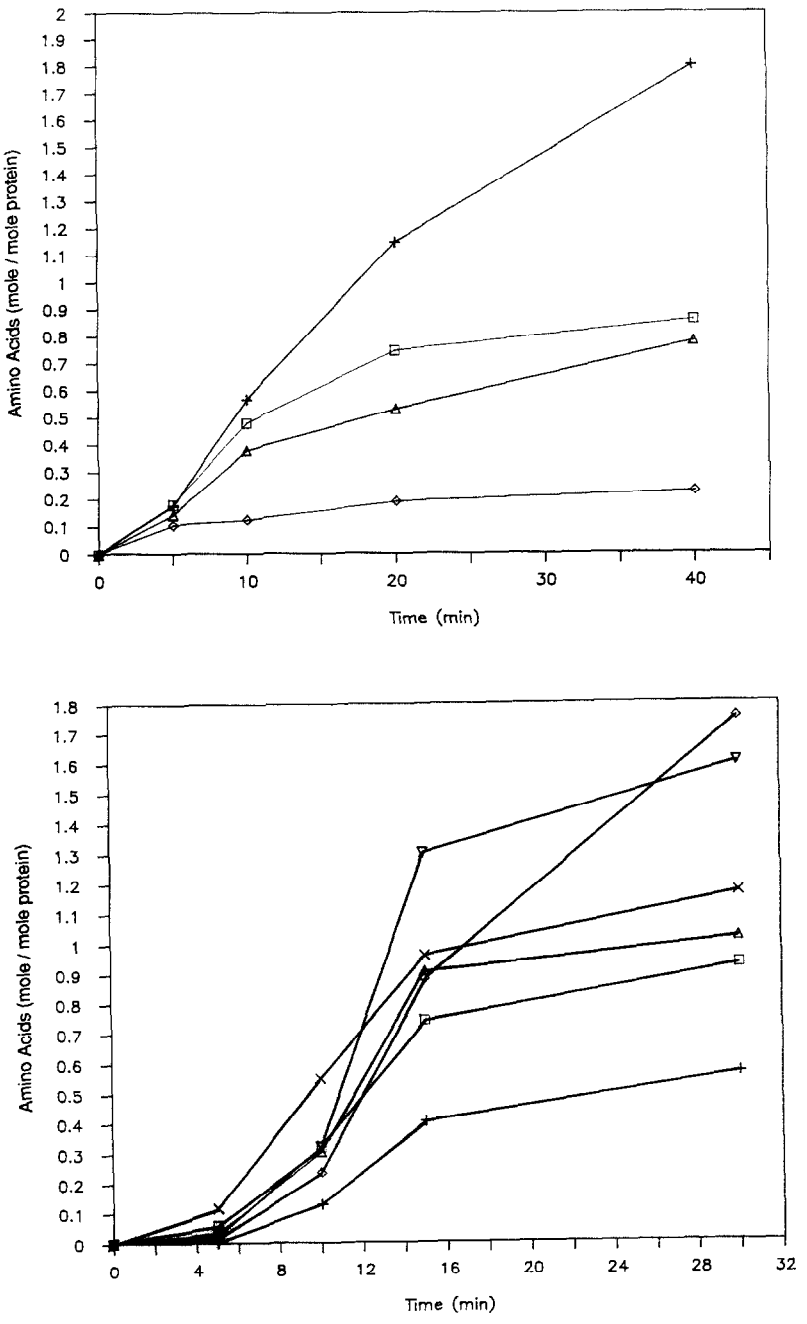

Fig. 4. Kinetic analysis of released amino acids from Cpase P digestion of human recombinant consensus α-interferon (top) and G-CSF (bottom). Top: (□) Glu; (+) Arg; (◇) Leu; (△) Lys. Bottom: (□) Gln; (+) His; (◇) Arg; (△) Ala; (×) Pro; (▽) Leu.

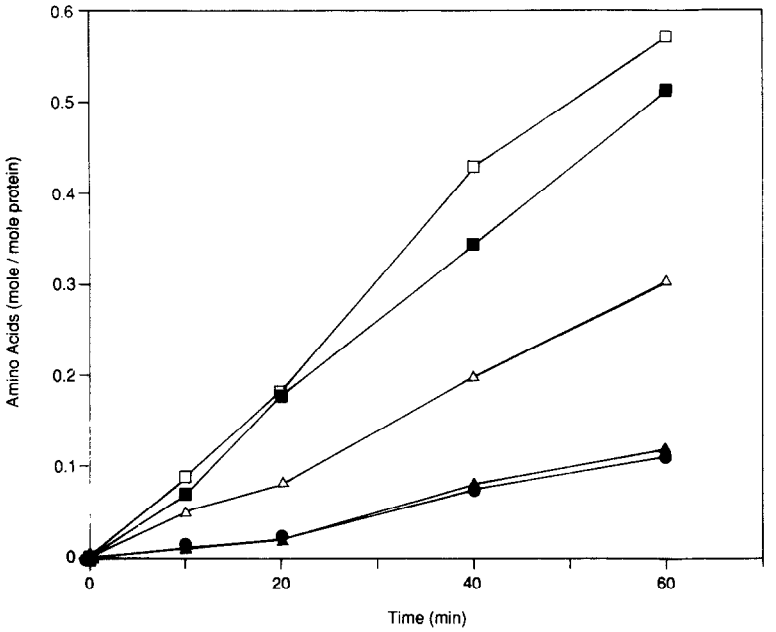

Fig. 5. Kinetic analysis of released amino acids from Cpase P digestion of human r-erythropoietin. (□) Asp; (■) Gly; (△) Thr; (▲) Arg; (●) Ala.

Very slow release of amino acids was observed when the digestion of human r-erythropoietin was done at 25°C with a lower enzyme-to-substrate ratio (1:400). Better results were thus obtained from digestions using an enzyme-to-substrate ratio of 1:100 at 37°C. Fig. 5 illustrates the kinetic plot for the release of amino acids after digestion of erythropoietin, yielding a C-terminal sequence of ···Arg–Thr–Gly–Asp. Much slower release of Arg and Ala was observed, indicating that the Ala–Cys–Arg··· sequence is more resistant to Cpase P attack as found in the digestion of lysozyme. The determined C-terminal sequence indicates that the arginyl residue predicted to be at the C-terminus according to the gene sequence[23] is missing from the recombinant protein. This result is confirmed by the isolation of des-Arg C-terminal peptide and is consistent with the data obtained from Cpase P digestion of human urinary erythropoietin (our unpublished data). It is intriguing to reveal that recombinant erythropoietin expressed in Chinese hamster ovary cells is processed by an endogenous carboxypeptidase in a manner similar to the carboxyl terminal processing of the natural hormone.

The broad specificity of the purified acid carboxypeptidase P has been described previously[16]. The enzymatic properties of this protease are similar to those of Cpase C from citrus fruit; however, their apparent differences were reported previously[15,16]. Cpase C might be contaminated with other amino- or endo-peptidase activity. On the contrary, the commercially available Cpase P does not require further purification prior to use.

As pointed out by Allen[24], endoproteolysis is inherent with all the methods employing Cpases such as Cpases A, B, C or Y. Such non-specific cleavage can be

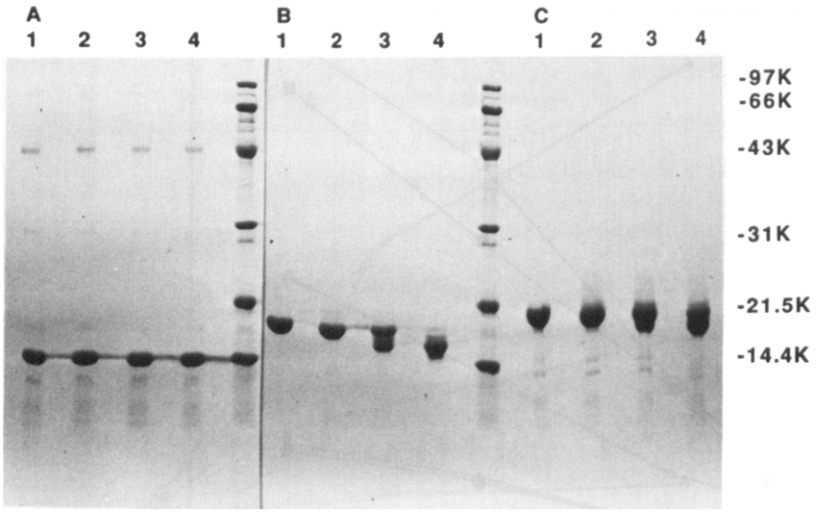

Fig. 6. SDS-PAGE of digested products derived from Cpase P digestion of hen egg lysozyme, human recombinant consensus α-interferon and human rG-CSF. For digestion conditions, see Table I. (A) Lysozyme: lane 1, control: lane 2, 0 time; lane 3, 1 h; lane 4, 4 h. (B) rG-CSF: lane 1, control; lane 2, 0 time; lane 3, 20 min; lane 4, 60 min. (C) Recombinant consensus α-interferon: lane 1, control; lane 2, 0 time; lane 3, 20 min; lane 4, 60 min.

demonstrated by SDS-PAGE of the final digestion product in that significant diminution in the intensity of the original protein band and the appearance of additional bands of lower molecular weight would occur. The non-specific proteolysis may result in interference of sequence assignment as is frequently seen in the literature[25]. In the subsequent studies, Cpase P digestion products obtained from several proteins at various time intervals were examined by SDS-PAGE to examine if the enzyme preparation contained endopeptidase activity. As shown in Fig. 6A, the mobility of protein bands for lysozyme remain unchanged from authentic sample, 0 h, 1 h and 4 h digestion products, indicating that the removal of a single C-terminal amino acid from lysozyme does not alter the apparent molecular weight of the protein. As indicated in Fig. 6B, recombinant G-CSF can be digested by Cpase P in a specific and sequential manner. After 1 h digestion, a new 17.6-kD band is generated by sequential degradation of the original 18.8-kD protein band from the C-terminal end. When the digestion products of recombinant consensus α-interferon were analyzed (Fig. 6C), a protein band at 19 kD appeared together with the 20-kD band within 1 h incubation. The intensity of the 19-kD band increases with the concomitant decrease of the original 20-kD band. It is important to note that no other lower molecular weight degradation bands were observed in the digests of these three proteins. These results also agree with the number of released amino acid residues determined by amino acid analysis (Table I). More importantly, these data indicate that the Cpase P preparation is devoid of endopeptidase activity. Since the PTC-amino acid analysis of Cpase P-treated proteins did not reveal the presence of N-terminal amino acids, especially methionine, it is thus concluded that limited C-terminal exoproteolysis by Cpase P had occurred during a controlled time course digestion.

It has been observed that Cpase P is stable and more reactive in the presence of various non-ionic detergents including Brij-35. The reaction can thus be performed in water containing 0.05% Brij-35 at lower pH. Brij-35 is advantageous to the digestion as it acts as a stabilizer for both Cpase P and protein substrate. The digestion solution containing no buffer or low buffer salt (*i.e.*, 10 m*M* sodium acetate) has great advantage when the analysis is performed using PITC derivatization and HPLC. Salt was found to interfere with subsequent amino acid analysis using narrow-bore HPLC as described previously.

Drawbacks observed during the use of Cpase A and/or B or Y digestion, such as varying cleavage efficiency and resistance of many bonds, have hindered their successful application in C-terminal sequence analysis. For example, recombinant consensus α-interferon, IL-2, and erythropoietin are resistant to cleavage by these enzymes even in combined digestion. On the contrary, Cpase P itself effectively releases C-terminal amino acids from these proteins. Although Cpase P cleaves almost every terminal amino acid, several specific peptide bonds are still resistant to the enzyme. Resistance of Gly–X and Pro–X bonds to Cpase P cleavage has been reported[16]. Similar results are also observed during the digestion of recombinant GM-CSF and bradykinin (Table II). Other peptide bonds such as the Cys–X bond, in which Cys is involved in disulfide linkage, are also resistant to Cpase P digestion. This is evident from the fact that the enzyme can only effectively cleave a single terminal amino acid for lysozyme, since a Cys–Arg bond was found right next to the cleaved Leu. This suggests that S-carboxymethylated or performic acid oxidized samples should be prepared for extended C-terminal analysis.

It is interesting to note that the whole rIL-2(Ala-125) molecule is quite resistant to Cpase P digestion under normal digestion conditions (Table II). The digestion of rIL-2(Ala-125) only releases two amino acids from the C-terminal end of the protein using a 1:100 enzyme-to-substrate ratio at 37°C. The remaining sequence (···Ala–Gln–Ser–Ile–Ile–Ser–Thr) preceding the C-terminal Leu–Thr stays uncleaved for hours. This result seems to suggest that Cpase P is unable to further cleave the Ser–Thr bond.

In summary, the described procedure involving the use of narrow-bore HPLC of PTC-amino acids derived from Cpase P digestion offers an effective and sensitive method for protein and peptide C-terminal sequence analysis. Since no other efficient sequential degradation methods are available for C-terminal analysis, this procedure should be valuable for the determination of the C-terminal sequence of proteins and peptides. The utilization of narrow-bore HPLC for microanalysis of all 20 amino acids greatly enhances the sensitivity required for evaluation of samples available only in minute quantities. Using the described C-terminal analysis methods in conjunction with automated N-terminal sequencing, both amino- and carboxy-terminal sequences can be established for assigning the protein coding region during molecular cloning of genomes. From this study, Cpase P is also best used to compare a native and recombinant protein, given the complexities of its C-terminal exopeptidase activity from one protein to another. Moreover, compared to the more established Cpases A, B, C and Y, Cpase P is a better exopeptidase and has wider applications for limited modification of proteins. It should gain broader application to the study of the relationship between polypeptide structure and function[26].

ACKNOWLEDGEMENTS

We thank Joan Bennett for typing the manuscript and Julie Heuston for preparing the artwork.

REFERENCES

1 K. Narita, H. Matsuo and T. Nakajima in S. B. Needleman (Editor), *Protein Sequence Determination*, Springer-Verlag, New York, Heidelberg, Berlin, 1975, p. 30.
2 H. Fraenkel-Conrat, H. Harris and J. L. Levy, *Methods Biochem. Anal.*, 2 (1955) 359.
3 K. Biemann, in M. Elzinga (Editor), *Methods in Protein Sequence Analysis*, Humana Press, Clifton, NJ, 1982, pp. 279–288.
4 H. R. Morris, *Phil. Trans. R. Soc., London*, A293 (1979) 39–51.
5 W. A. Schroeder, *Methods Enzymol.*, 25 (1972) 138–143.
6 H. Matsuo and K. Narita in S. B. Needleman (Editor), *Protein Sequence Determination*, Springer-Verlag, New York, Heidelberg, Berlin, 1979, p. 104.
7 G. R. Stark, *Methods Enzymol.*, 25 (1972) 369–384.
8 D. H. Hawke, H. W. Lahm, J. E. Shively and C. W. Todd, *Anal. Biochem.*, 166 (1987) 298–307.
9 R. P. Ambler, *Methods Enzymol.*, 11 (1967) 155–166.
10 R. P. Ambler, *Methods Enzymol.*, 25 (1972) 262–272.
11 H. Neurath, in P. D. Boyer, H. Lardy and M. Myrbäck (Editors), *The Enzymes*, Vol. IV, Academic Press, New York, 1960, p. 11.
12 R. Hayashi, *Methods Enzymol.*, 47 (1977) 84–93.
13 H. Teschesche, *Methods Enzymol.*, 47 (1977) 73–83.
14 H. Teschesche and S. Kaufer, *Eur. J. Biochem.*, 26 (1972) 33.
15 S. Yokoyama, A. Oobayashi, O. Tanabe and E. Ichishima, *Appl. Microbiol.*, 28 (1974) 742–747.
16 S. Yokoyama, A. Oobayashi, O. Tanabe and E. Ichishima, *Biochim. Biophys. Acta*, 397 (1975) 443–448.
17 H.-S. Lu and P.-H. Lai, *J. Chromatogr.*, 368 (1986) 215–231.
18 U. K. Laemmli, *Nature (London)*, 227 (1970) 680–695.
19 P. E. Hare, in S. B. Needleman (Editor), *Protein Sequence Determination*, Springer-Verlag, New York, Heidelberg, Berlin, 1975, pp. 204–231.
20 M. W. Hunkapiller, J. E. Strickler and K. J. Wilson, *Science (Washington, D.C.)*, 226 (1985) 304–311.
21 R. Heinrickson and S. C. Meredith, *Anal. Biochem.*, 136 (1984) 65–74.
22 A. Mondino, G. Bongiovanni and S. Fumero, *J. Chromatogr.*, 104 (1975) 297.
23 F. K. Lin, S. Suggs, C. H. Lin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 7580–7585.
24 G. Allen in T. S. Work and R. H. Burton (Editors), *Sequencing of Proteins and Peptides*, Elsevier/North Holland, Amsterdam, New York, Oxford, 1981, pp. 17–42.
25 L. Rask, H. Anudi and P. A. Peterson, *FEBS Lett.*, 164 (1979) 55–58.
26 J. T. Potts, Jr., *Methods Enzymol.*, 11 (1967) 648–663.