Protein Science (1992), *1*, 1622-1633. Cambridge University Press. Printed in the USA.

Automated carboxy-terminal sequence analysis of peptides and proteins using diphenyl phosphoroisothiocyanatidate

JEROME M. BAILEY, FIROOZEH NIKFARJAM, NARMADA R. SHENOY, AND JOHN E. SHIVELY

Beckman Research Institute of the City of Hope, Division of Immunology, Duarte, California 91010

(RECEIVED May 15, 1992; REVISED MANUSCRIPT RECEIVED July 22, 1992)

Abstract

Proteins and peptides can be sequenced from the carboxy-terminus with isothiocyanate reagents to produce amino acid thiohydantoin derivatives. Previous studies in our laboratory have focused on the automation of the thiocyanate chemistry using acetic anhydride and trimethylsilylisothiocyanate (TMS-ITC) to derivatize the C-terminal amino acid to a thiohydantoin and sodium trimethylsilanolate for specific hydrolysis of the derivatized C-terminal amino acid (Bailey, J.M., Shenoy, N.R., Ronk, M., & Shively, J.E., 1992, *Protein Sci. 1*, 68-80). A major limitation of this approach was the need to activate the C-terminus with acetic anhydride. We now describe the use of a new reagent, diphenyl phosphoroisothiocyanatidate (DPP-ITC) and pyridine, which combines the activation and derivatization steps to produce peptidylthiohydantoins. Previous work by Kenner et al. (Kenner, G.W., Khorana, H.G., & Stedman, R.J., 1953, *Chem. Soc. J.*, 673-678) with this reagent demonstrated slow kinetics. Several days were required for complete reaction. We show here that the inclusion of pyridine was found to promote the formation of C-terminal thiohydantoins by DPP-ITC resulting in complete conversion of the C-terminal amino acid to a thiohydantoin in less than 1 h. Reagents such as imidazole, triazine, and tetrazole were also found to promote the reaction with DPP-ITC as effectively as pyridine. General base catalysts, such as triethylamine, do not promote the reaction, but are required to convert the C-terminal carboxylic acid to a salt prior to the reaction with DPP-ITC and pyridine. By introducing the DPP-ITC reagent and pyridine in separate steps in an automated sequencer, we observed improved sequencing yields for amino acids normally found difficult to derivatize with acetic anhydride/TMS-ITC. This was particularly true for aspartic acid, which now can be sequenced in yields comparable to most of the other amino acids. Automated programs are described for the C-terminal sequencing of peptides covalently attached to carboxylic acid-modified polyethylene and proteins (200 pmol to 5 nmol) noncovalently applied to Zitex (porous Teflon). The generality of our automated C-terminal sequencing methodology was examined by sequencing model peptides containing all 20 of the common amino acids. All of the amino acids tested were found to sequence in good yield except for proline, which was found not to be capable of derivatization. In spite of this limitation, the methodology should be a valuable tool for the C-terminal sequence analysis of peptides and proteins. This work represents the first time an automated, sequential, chemical method for C-terminal sequence analysis has been applied to subnanomolar amounts of protein samples.

Keywords: C-terminal; peptides; proteins; sequencing; thiohydantoins

The chemical sequential analysis of peptides and proteins from the carboxy-terminus is a goal of our laboratory. Such a procedure, in addition to complementing amino-terminal degradation, would be invaluable for the detection of posttranslational processing at the carboxy-terminus of expressed proteins from known DNA sequences, confirmation of the correct placement of initiation codons and reading frames, and assistance in the design of oligonucleotide probes to screen cDNA libraries. Of the chemical methods proposed (for reviews see Ward, 1986; Rangarajan, 1988; Inglis, 1991), the method based on the derivatization of the carboxy-terminal amino acid to a

Reprint requests to: Jerome M. Bailey, Beckman Research Institute of the City of Hope, Division of Immunology, 1450 E. Duarte Road, Duarte, California 91010.

thiohydantoin with thiocyanate-based reagents has been the most widely studied and appears to be the most attractive due to its similarity to the current methods of N-terminal sequencing. Because the thiohydantoin amino acids produced have UV absorption spectra and extinction coefficients similar to the phenylthiohydantoin amino acids formed during the Edman degradation, the sensitivity of the thiocyanate method is expected to be similar to that of current N-terminal methods (10–200 pmol of applied sample).

Preliminary work in our laboratory has examined in detail the generality of this method for the sequential degradation of the 20 naturally occurring amino acids in the solution phase (Bailey & Shively, 1990) and more recently, automation of this chemistry using peptides covalently attached to carboxylic acid-modified polyethylene (Bailey et al., 1992). Each cycle of degradation consisted of three steps: activation of the carboxy-terminal carboxylic group with acetic anhydride to form an oxazolinone, derivatization with trimethylsilylisothiocyanate (TMS-ITC) to form a thiohydantoin amino acid, and specific hydrolysis of the derivatized C-terminal amino acid with sodium trimethylsilanolate, which gave a thiohydantoin amino acid derivative and a shortened peptide capable of continued degradation. The thiohydantoin amino acid was then analyzed by reverse-phase high-performance liquid chromatography (HPLC). Current limitations of this method include a reduced yield when asparagine or aspartate is encountered during the degradation, the inability of the method to degrade proline when it is at the C-terminus of peptides, the difficulty in applying this chemistry to protein samples, and precleavage of the peptide sample during sequencing.

This report describes a reinvestigation of the use of diphenyl phosphoroisothiocyanatidate (DPP-ITC) for C-terminal sequencing. This reagent was first introduced by Kenner et al. (1953), but was abandoned, presumably because the derivatization reaction was found to require several days to go to completion. The proposed mechanism involved the reaction of the C-terminal carboxylate with DPP-ITC, expulsion of the isothiocyanate ion, and reaction of the isothiocyanate ion with the activated carboxylate to form the acyl isothiocyanate. The potential advantage of this type of reagent is that it does not require the use of acetic anhydride for activation because this reagent is capable of both activation and derivatization. Problems caused by the use of acetic anhydride include: cyclic anhydride formation with Glu and Asp, dehydration of Ser and Thr to form an unsaturated oxazolinone, and the fact that the required oxazolinone intermediate for reaction with TMS-ITC cannot form with proline (Bailey et al., 1992). While studying the reaction of DPP-ITC with the C-terminal amino acid of peptide samples covalently coupled to carboxylic acid-modified polyethylene film we found that the mechanism of reaction originally proposed by Kenner et al. (1953) is not correct. Furthermore, we have found that once the peptide had reacted with DPP-ITC, cyclization to a thiohydantoin could be promoted by reagents such as pyridine, imidazole, triazine, and tetrazole. This permitted derivatization of the C-terminal amino acid to a thiohydantoin in less than 1 h at 50 °C rather than the several days previously required. Additionally, we have automated and tested the generality of this chemistry by sequencing model peptides containing all 20 of the naturally occurring amino acids. We also describe the application of this chemistry to the C-terminal sequence analysis of protein samples (200 pmol to 5 nmol) noncovalently applied to a Zitex support.

Results

HPLC separation of the amino acid thiohydantoins

The released thiohydantoin amino acid derivatives from C-terminal sequencing were analyzed by reverse-phase HPLC. Our initial separation of the thiohydantoin amino acids was performed on a Phenomenex Ultracarb 5 ODS (30) column (2.0 mm × 250 mm) (Bailey & Shively, 1991; Bailey et al., 1992). Although this column provided separation of the derivatized amino acids, the instability of this column to the acidic conditions used for the separation and the broad peaks observed during automated sequencing of some amino acids necessitated the development of an alternative separation system. The revised reverse-phase separation system utilizes a C-18 (3 μ, 100 Å) Reliasil column (2.1 × 25 cm). The separation of the thiohydantoin amino acids (600 pmol) with this system is shown in Figure 1.

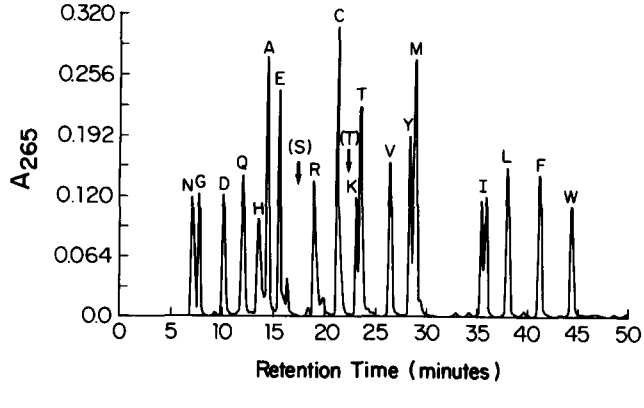

Fig. 1. High performance liquid chromatographic separation of the amino acid thiohydantoin standards. The amino acid thiohydantoin standards (600 pmol) were separated on a C-18 Reliasil column as described in the Materials and methods. Absorbance was monitored at 265 nm at 0.32 AUFS. The thiohydantoin derivative of cysteine is *S*-methylated and the side ε-amino group of the lysine derivative is acetylated. The elution positions of the Thr and Ser analogues formed during automated sequencing are indicated with arrows.

Reaction of N-acetyl methionine and N-acetyl proline with DPP-ITC

N-acetyl methionine and *N*-acetyl proline were reacted with DPP-ITC in the presence of a slight molar excess of pyridine over the concentration of the isothiocyanate reagent. The mixture was deacetylated with 12 N HCl at room temperature and analyzed by fast-atom bombardment mass spectrometry (FAB/MS) and reverse-phase HPLC. The reaction with *N*-acetyl methionine gave exclusively thiohydantoin methionine, and the reaction with *N*-acetyl proline gave exclusively the starting material, *N*-acetyl proline. Diphenyl phosphate ($MH^+ = 251$) was obtained as part of the acid precipitate in both cases.

Reaction of DPP-ITC with peptides in solution

The peptides (60 nmol) were first reacted with acetic anhydride to acetylate the N-terminus, treated with dilute aqueous triethylamine, and then dried. The peptide, *N*-acetyl-Thr-Val-Leu, was treated with DPP-ITC (0.06 mmol) in acetonitrile solution either in the presence of pyridine (0.12 mmol) or triethylamine (0.11 mmol) for 40 min at 50 °C. The reaction products were then analyzed by reverse-phase HPLC and FAB/MS. The expected *N*-acetyl peptide with a C-terminal thiohydantoin leucine ($MH^+ = 415$) was found only in the reaction carried out in the presence of pyridine. The reaction carried out in the presence of triethylamine gave only the *N*-acetylated starting peptide (*N*-acetyl-Thr-Val-Leu ($MH^+ = 374$). Treatment of the peptidylthiohydantoin from the pyridine reaction with dilute aqueous triethylamine for 10 min at 50 °C (Bailey et al., 1990) gave the expected shortened peptide (*N*-acetyl-Thr-Val) and thiohydantoin leucine, as confirmed by reverse-phase HPLC and FAB/MS analysis. Unlike the reaction of Thr-Val-Leu with acetic anhydride and TMS-ITC, which gave two diastereomeric thiohydantoin peptides (Bailey & Shively, 1990), the reaction of Thr-Val-Leu with DPP-ITC in the presence of pyridine gave only a single thiohydantoin peptide. The reaction of either *N*-acetyl-Pro-Phe-Asp or *N*-acetyl-Ala-Phe-Pro with DPP-ITC and pyridine as described above failed to form the expected thiohydantoin peptide. In both cases the starting peptide was recovered. However, the *N*-acetyl-Pro-Phe-Asp peptide could be sequenced under other conditions (see below).

Application of the DPP-ITC chemistry to leucine enkephalin covalently coupled to PE-COOH

Leucine enkephalin (YGGFL) (6–30 nmol) on PE-COOH (2 × 10-mm strip) was used as a model system to study the chemical conditions necessary for automation of this chemistry. It was found that treatment of the covalently coupled peptide with DPP-ITC in the presence of a molar excess of pyridine at 50 °C for 40 min, conditions similar to those employed in the solution phase, formed peptidylthiohydantoin. Cleavage of the derivatized C-terminal amino acid was accomplished with dilute aqueous triethylamine as described above (Bailey & Shively, 1990; Shenoy et al., 1992). The presence of thiohydantoin leucine was assessed by comparison of retention time with standard thiohydantoin leucine on reverse-phase HPLC. It was also found that treatment of the peptide with DPP-ITC together with pyridine was not necessary, in that these steps could be separated in order to reduce background due to excess DPP-ITC. The peptide was first treated with aqueous or methanolic triethylamine for 5 min at 50 °C. The triethylamine solution was removed and the support dried. The peptide then was treated with DPP-ITC in acetonitrile at 50 °C for 20 min, washed with anhydrous dimethylformamide (DMF) or heptane to remove unreacted DPP-ITC, and then treated with pyridine at 50 °C for 20 min. It was found that the concentration of pyridine could range from 1% in acetonitrile to 100% without any effect on the yield (data not shown). When the DPP-ITC step was separated from the pyridine step, the pretreatment with triethylamine was found to be absolutely essential for the subsequent chemistry to work. When the DPP-ITC and pyridine were added together, pretreatment with triethylamine was not required. When DPP-ITC and pyridine were added together the best yields of sequencing were obtained when pyridine was present in molar excess with respect to DPP-ITC. Imidazole, triazine, and tetrazole were found to be equally as capable as pyridine in promoting the reaction with DPP-ITC to form a peptidylthiohydantoin. As with pyridine the best yields when using these alternative reagents were obtained when they were present in molar excess with respect to DPP-ITC.

Automated sequencing of peptides covalently coupled to PE-COOH

Based on the studies described above with YGGFL on the PE-COOH support and our newly developed chemistry for specific hydrolysis of the derivatized C-terminal amino acid (Bailey et al., 1992), we decided on the following sequencing strategy: (1) covalent immobilization of the sample to a solid support, (2) reaction with triethylamine to form the C-terminal carboxylate, (3) removal of the triethylamine solution, (4) reaction with DPP-ITC in acetonitrile, (5) removal of the DPP-ITC, (6) reaction with pyridine (pyridine was chosen over imidazole, triazine, and tetrazole because of its ability to be delivered in the gas phase), and (7) cleavage of the derivatized C-terminal amino acid with sodium trimethylsilanolate. The chemical scheme is shown in Figure 2. Methods for the covalent attachment of peptides to carboxylic acid-modified polyethylene supports have been optimized and described (Shenoy et al., 1992). Leucine enkephalin covalently immobilized on carboxylic acid-modified polyethylene was

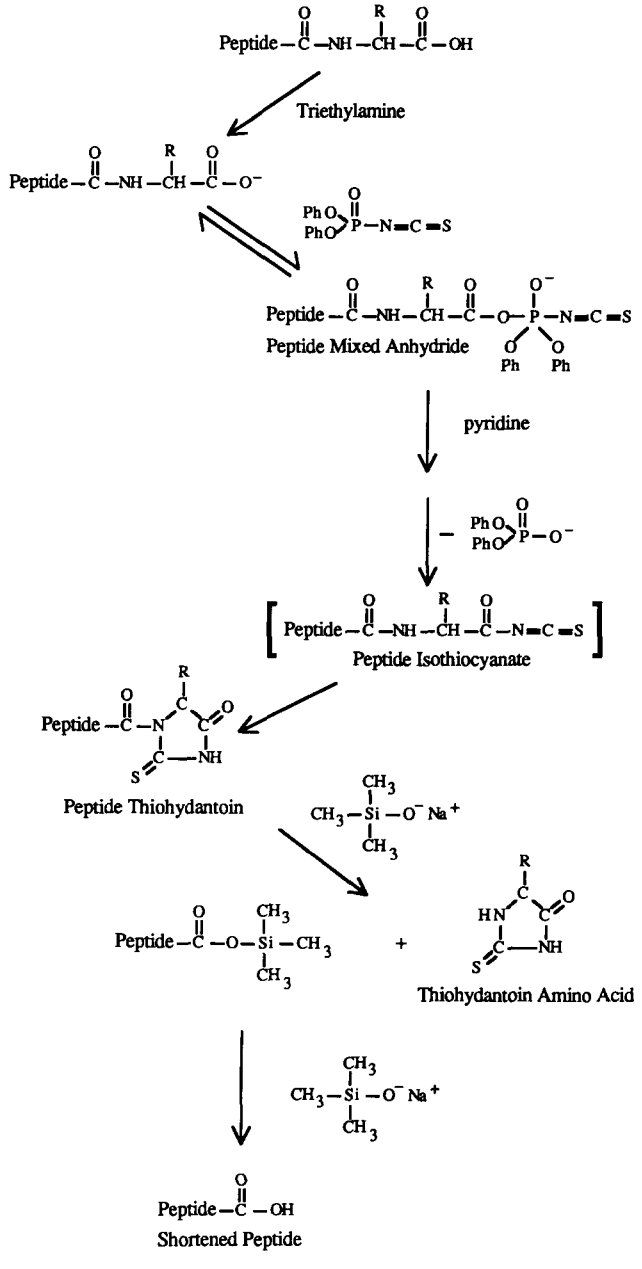

Fig. 2. Reaction scheme for the sequential C-terminal degradation of peptides using the diphenyl phosphoroisothiocyanatidate/pyridine chemistry.

used to optimize automated sequencing by varying the temperature, reaction times, repetition and concentration of DPP-ITC, pyridine, and triethylamine. It was found that the highest concentration (3.0 M) of DPP-ITC used was optimal, the concentration of pyridine could be varied from 1 to 100% in acetonitrile with no change in results, and the best results were obtained with three repetitions of the coupling reaction per cycle. The coupling cycle consists of the triethylamine, DPP-ITC, and pyridine steps. The composition of the reagents and solvents and a description of the program used are shown

Table 1. *Composition of reagents and solvents used in automated C-terminal sequencing*

R1	50% triethylamine in methanol
R2	Diphenyl phosphoroisothiocyanatidate in acetonitrile (3.0 M)
R3	0.10 M sodium trimethylsilanolate in 50% methanol, 50% *t*-butyl alcohol
R4	–
S1	Pyridine
S2	1% trifluoroacetic acid in water
S3	Methanol
S4	Dimethylformamide

in Tables 1 and 2, respectively. The total time required for each cycle was approximately 90 min. Typical initial yields are approximately 50%. Figure 3A shows four cycles of automated C-terminal sequencing for the pentapeptide leucine enkephalin (YGGFL, 6.3 nmol). The concentration of pyridine and DPP-ITC used was 2% (v/v) in acetonitrile and 1.8 M in acetonitrile, respectively. For comparison, Figure 3B shows four cycles of automated C-terminal sequencing of the same peptide (24.2

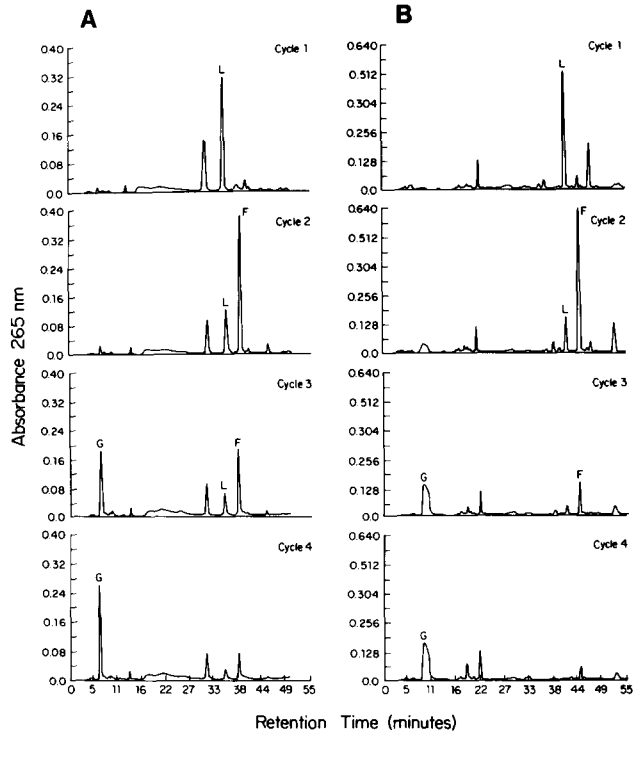

Fig. 3. A: Automated C-terminal sequencing of YGGFL (6.3 nmol) covalently coupled to carboxylic acid modified polyethylene using diphenyl phosphoroisothiocyanatidate (DPP-ITC)/pyridine. The concentration of DPP-ITC used was 1.8 M. Pyridine was delivered as a 2% (v/v) solution in anhydrous acetonitrile. **B**: Automated sequencing of YGGFL (24.2 nmol) covalently coupled to carboxylic acid modified polyethylene using acetic anhydride/trimethylsilylisothiocyanate (Bailey et al., 1992).

Table 2. *C-terminal sequencer program summary used for peptides*

Continuous flow reactor (CFR) (70 °C)	Conversion flask (CF) (40 °C)	Duration (s)
R1 reaction		90
Dry		60
R2 reaction		240
Blow out		15
S1 deliver		90
S1 reaction		30
Blowout		30
R1 reaction		90
Dry		60
R2 reaction		240
Blow out		15
S1 deliver		90
S1 reaction		30
Blowout		30
R1 reaction		90
Dry		60
R2 reaction		240
Blow out		15
S1 deliver		90
S1 reaction		30
Blowout		30
S3 rinse		120
S4 rinse		480
S3 rinse		120
S4 rinse		240
S3 rinse		120
R3 reaction		1,200
R3 to CF		30
R3 reaction	Dry	300
R3 reaction	Dry	300
S2 delivery		10
S2 reaction		5
S2 to CF		10
S2 delivery		9
S2 reaction		5
S2 to CF		10
	Inject	11
Pause	Pause	60
Dry	Dry	100

nmol) using the acetic anhydride/TMS-ITC chemistry (Bailey et al., 1992). A comparison of cycles 2 and 3 of Figure 3A,B suggests the amount of precleavage (preview of the residue expected in the next cycle), measured by the amount of glycine in cycle 2, is reduced with the use of DPP-ITC, as compared to TMS-ITC.

In order to reduce background due to pyridine, this reagent was introduced in the gas phase. Figure 4 shows the automated sequencing of the tripeptide, PFD (32 nmol), with the concentration of DPP-ITC at 3.0 M in acetonitrile and pyridine (neat) in the gas phase. In contrast to the results shown in Figure 3, the pyridine background peak eluting at 20–25 min is greatly diminished. Pyridine delivered in the gas phase was found to be as effective as liquid-phase pyridine for catalyzing thiohydantoin formation. As discussed previously (Bailey et al., 1992), the C-terminal sequencing yields are reduced for the amino acids closest to the solid support.

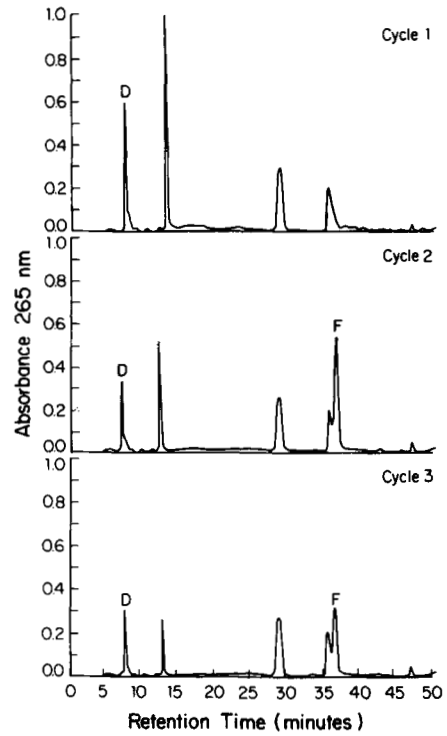

Fig. 4. Automated C-terminal sequencing of PFD (32 nmol) covalently coupled to carboxylic acid-modified polyethylene. The concentration of DPP-ITC used was 3.0 M. Pyridine (neat) was delivered in the gas phase.

Generality of sequencing

We have applied the above-described automated C-terminal sequencing technique to a variety of peptides covalently coupled to PE-COOH. This has enabled us to test the applicability of our sequencing conditions to all 20 of the naturally occurring amino acids (Table 3). All of the amino acids tested, except for proline, could be sequenced in good yield with this chemistry. It has not been possible to sequence through a C-terminal proline. The amino acid immediately before proline was also found to sequence with reduced yields (30–50% of expected). Although Thr and Ser sequence in good yield, the nature of the derivative formed is not well understood. The threonine derivative obtained does not coelute with a dehydro-Thr thiohydantoin standard, but rather elutes 2 min earlier. It has equivalent UV absorbance at 265 and 319 nm, suggesting that it may not be a simple thiohydantoin threonine. The same UV absorbance properties were observed with the serine analogue. The elution position of the Thr and Ser analogues is shown with arrows in Figure 2. We have been unsuccessful in synthesizing thiohydantoin serine starting from the free amino acid,

Table 3. *Peptides used for automated C-terminal sequencing*[a]

Peptide	Number of cycles sequenced
SYSMEHFRWG	5
RVYIHPI	1
GRGD	3
GRGDS	3
YGGFLK[b]	6
FLEEL	5
KCTCCA	2
YGGFMRGL	6
VIHN	4
VIHNL	5
WHWLQL	6
YGGFLR	6
PFD	2
YKW[b]	3
TYS	3
KPT[c]	1
RYLPT[d]	1
AGSE	4
VTCG	4

[a] The amount of peptide sequenced ranged from 5 to 30 nmol. The initial yields were approximately 50%.
[b] Peptides containing Lys (K) show a blank cycle at Lys.
[c] Contains D-Pro.
[d] Contains L-Pro.

so comparison with a standard at this time is not yet possible. Work is continuing on determination of the structure of the Thr and Ser analogues. C-Terminal cysteine was also readily sequenced with this methodology, forming a derivative indistinguishable from that formed with Ser. It is anticipated that this problem can be obviated by alkylating the Cys residues prior to sequencing with reagents such as 4-vinyl-pyridine or iodoacetic acid using established protocols. C-terminal Asn eluted as two peaks with retention times corresponding to thiohydantoin Asn and thiohydantoin Asp. Unlike the acetic anhydride/TMS-ITC chemistry (Bailey et al., 1992), peptides containing C-terminal aspartate can be sequenced in good yield with the DPP-ITC/pyridine chemistry. Peptides containing lysine (YKW and YGGFLK) gave blank cycles where Lys should have been seen. The yields on both sides of the Lys residue were good, indicating the likelihood that the side chain of Lys had become covalently attached during coupling of the peptide to the support. Amino acid analysis of the residual peptide after sequence analysis supported this conclusion. Additionally, automated sequencing of YGGFLK using the acetic anhydride/TMS-ITC chemistry (Bailey et al., 1992), also gave a blank cycle where K should have been seen, supporting the conclusion that lysine is covalently attached to the membrane, and that the blank cycle with lysine is not unique to the DPP-ITC/pyridine chemistry. When the sample to be sequenced is a protein, and therefore does not need to be covalently coupled, lysine was found to be readily sequenced with this chemistry (see below).

Automated C-terminal sequencing of protein samples

Previous efforts in our laboratory to apply our automated C-terminal sequencing procedure to protein samples has met with much difficulty. Chief among these is the insolubility of most proteins in acetic anhydride and the low yields obtained when attempting to covalently couple protein samples to the nonporous PE-COOH support. Because polyvinylidene difluoride is known to degrade under the conditions of our C-terminal sequencing procedure (Shenoy et al., 1992) we have investigated the feasibility of sequencing proteins noncovalently attached to Zitex (porous teflon) using DPP-ITC/pyridine for derivatization. Figures 5 and 6 show four cycles of C-terminal sequencing of β-lactoglobulin A (350 pmol) and superoxide dismutase (5.2 nmol) noncovalently applied onto the Zitex support. The initial yield was approximately 14% for β-lactoglobulin and 4% for superoxide dismutase. Pyridine was used in the gas phase in order to prevent sample washout. In each case two to three cycles could be positively identified. The reason for the low initial and repetitive yields is presently being explored. The initial yield for β-lactoglobulin was found to be the same regardless of the amount of sample loaded (350 pmol to 5 nmol), suggesting that the actual initial yield may be sample dependent. The program used for sequencing proteins (Table 4) was significantly shortened and performed at a lower temperature (55 °C) than that used for peptides due to the ease at which protein samples were found to fragment under the sequencing conditions. How much the program can be further shortened is dependent on the amount of time it takes for the HPLC analysis of the thiohydantoin amino acids (60 min). It should be pointed out that the triethylamine step used to convert the C-terminal carboxylic acid into a carboxylate need only be performed once prior to the first cycle. The reduced repetitive yields observed for protein samples relative to peptide samples cannot be explained by sample washout, since amino acid analysis of the sample support after three cycles of sequencing revealed at least 70% of the protein was still remaining. The ability to sequence through lysine is shown with superoxide dismutase in Figure 6. This sample was treated with acetic anhydride immediately prior to sequencing in order to acetylate the epsilon-amino group of lysine. If this step was not performed, the thiohydantoin lysine derivative formed coeluted with thiohydantoin-Phe. The addition of the acetic anhydride step was found not to affect the sequencing performance and was used as a general procedure for all protein samples.

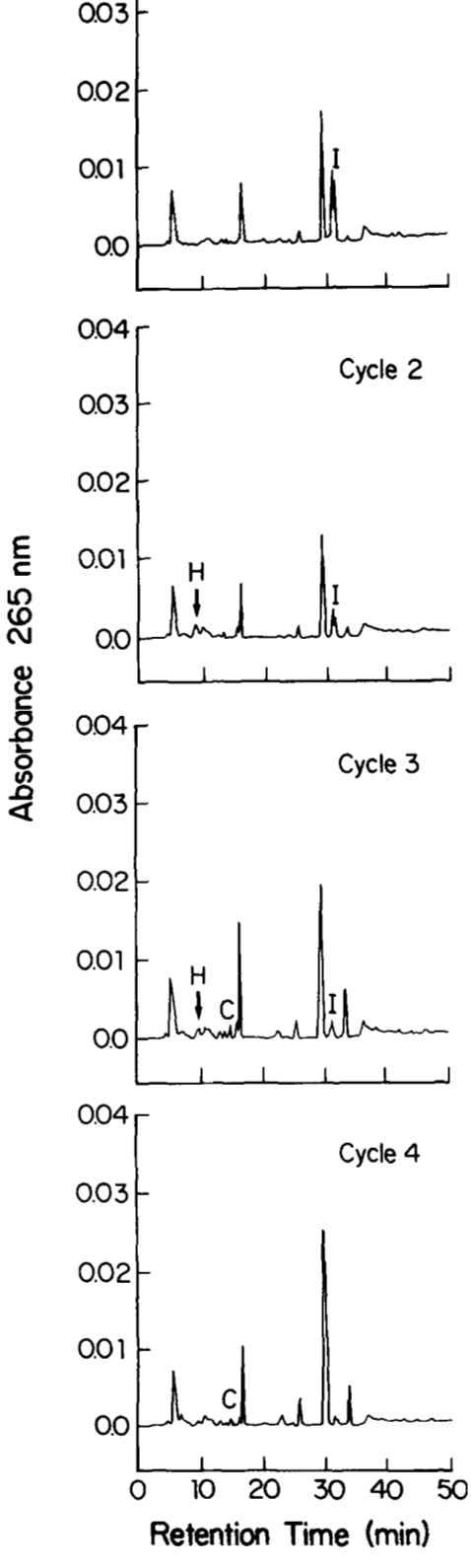

Fig. 5. Automated C-terminal sequencing of β-lactoglobulin A (350 pmol) noncovalently applied to Zitex. The concentration of DPP-ITC used was 3.0 M. Pyridine (neat) was delivered in the gas phase. The sequence of the C-terminus is -Cys-His-Ile.

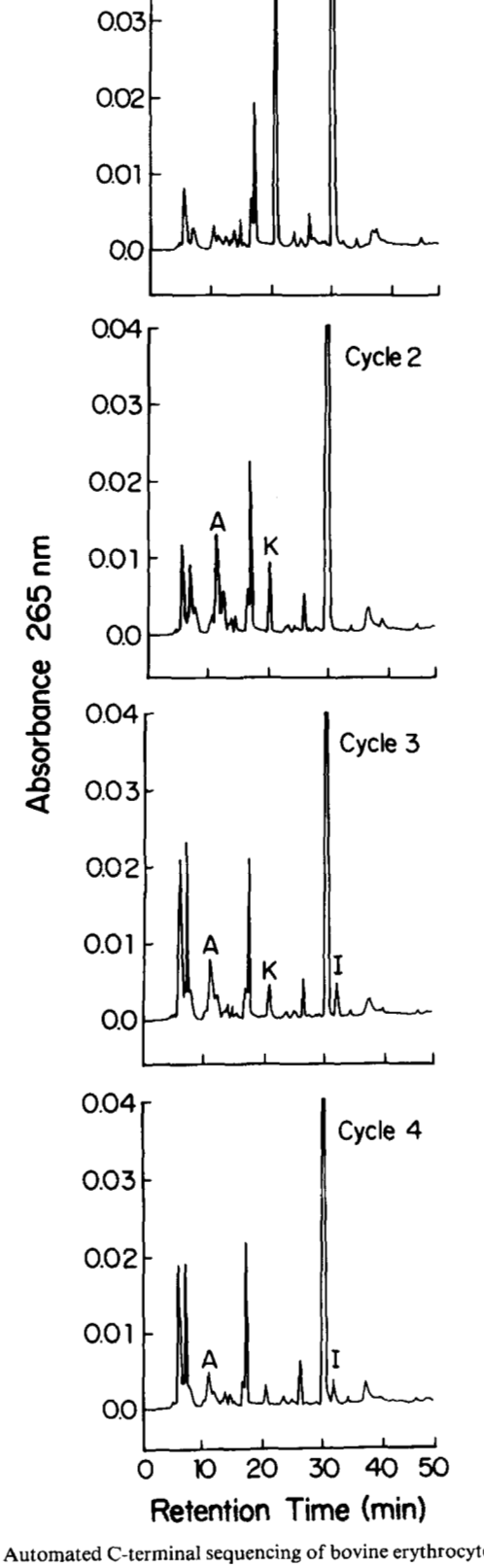

Fig. 6. Automated C-terminal sequencing of bovine erythrocyte superoxide dismutase (5.2 nmol) noncovalently applied to Zitex. The concentration of DPP-ITC used was 3.0 M. Pyridine (neat) was delivered in the gas phase. The sequence of the C-terminus is -Ile-Ala-Lys.

Table 4. *C-terminal sequencer program summary used for proteins*

Continuous flow reactor (CFR) (55 °C)	Conversion flask (CF) (40 °C)	Duration (s)
R1 reaction[a]		90
Dry[a]		60
R2 reaction		120
Blow out		15
S1 deliver		60
Blowout		60
S4 rinse		30
R2 reaction		240
Blow out		15
S1 deliver		60
Blowout		60
S3 rinse		120
Pause		480
S4 rinse		480
S3 rinse		120
S4 rinse		240
S3 rinse		120
R3 reaction		20
R3 to CF		30
	Dry	600
S2 delivery		10
S2 to CF		10
S2 delivery		9
S2 reaction		5
S2 to CF		10
	Inject	11
Pause	Pause	60
Dry	Dry	100

[a] Step used only before cycle 1 in the program.

Discussion

Previous studies in this laboratory have focused on automation of the thiohydantoin chemistry using acetic anhydride/TMS-ITC for formation of the peptidylthiohydantoins and sodium trimethylsilanolate for hydrolysis of these peptidylthiohydantoins into an amino acid thiohydantoin derivative and a new shortened peptide capable of continued degradation (Bailey et al., 1992). These investigations also addressed the generality of sequencing with the thiocyanate chemistry by examining a series of model peptides containing most of the naturally occurring amino acid side chains. Automation of the chemistry provided a means of solving some problems, such as derivatization of C-terminal aspartate; however, it also revealed other problems, such as precleavage, which was not readily apparent with our solution phase analysis (Bailey & Shively, 1990). We believe that precleavage was due to premature hydrolysis of the C-terminal thiohydantoin amino acid during the derivatization reaction, thereby exposing the next in line amino acid to derivatization. Because four repetitions of the derivatization reaction were found to be necessary to derivatize all of the amino acids (Bailey et al., 1992), precleavage could contribute to lower initial yields (the C-terminal amino acid, which was prematurely cleaved, would be washed off to waste during the wash step). Precleavage was also observed by Inglis et al. (1991). In our previous studies, we demonstrated that the TMS-ITC reagent reacted with oxazolinone derivatives to form thiohydantoins and that acetic anhydride was necessary to form the requisite oxazolinone. Most of the amino acids, such as Ser, Thr, Glu, and Asp (excluding Pro which does not degrade), which were found to contribute to reduced yields of sequencing with the acetic anhydride/TMS-ITC chemistry have been shown to form their expected oxazolinone derivatives upon treatment with acetic anhydride (Bailey & Shively, 1990). Competing reactions were found for Glu and Asp, which formed cyclic anhydrides, and for Thr and Ser, which formed unsaturated oxazolinones, all of which were incapable of reacting with TMS-ITC to form the desired thiohydantoin. The lack of reaction of proline with TMS-ITC was reasoned to be due to one of three possible explanations: (1) proline cannot form the required oxazolinone, (2) the amide nitrogen of proline is not nucleophilic enough for attack of the linear isothiocyanate, or (3) the side chain of proline causes steric hindrance preventing formation of the five-membered thiohydantoin ring. It was reasoned that if a reagent could be developed that was capable of forming the necessary linear isothiocyanate (Fig. 1) by a mechanism that did not require the formation of an oxazolinone, many of the above-cited problems could be solved. In our search for such a reagent we reinvestigated the use of DPP-ITC for derivatization of the C-terminal amino acid to a thiohydantoin, because this reagent potentially eliminated the need to form oxazolinones.

Mechanism of thiohydantoin formation with DPP-ITC

The use of DPP-ITC for C-terminal sequencing was first described by Kenner et al. (1953). These authors proposed a reaction mechanism whereby isothiocyanate ion was liberated by exchange with the C-terminal carboxylate forming an acyl phosphate. The acyl phosphate was in turn postulated to be attacked by the isothiocyanate ion to form the desired acyl isothiocyanate and diphenyl phosphate ion. Although this chemistry successfully formed a C-terminal thiohydantoin it was never actively pursued, most likely because it took several days for the reaction to go to completion. We have found that inclusion of pyridine, imidazole, triazine, or tetrazole permitted the reaction of the C-terminal carboxylate with DPP-ITC, to form a thiohydantoin, to be complete in less than 30 min at 50 °C. As discussed in the Results section, reaction of *N*-acetyl-Thr-Val-Leu with DPP-ITC/pyridine formed a single peptidylthiohydantoin. This is in contrast to the reaction of this peptide with acetic anhydride/TMS-ITC for

which we observed the formation of two diastereomeric peptidylthiohydantoins (Bailey & Shively, 1990, Fig. 2). It is likely that racemization was the result of oxazolinone formation. The single peptidylthiohydantoin formed on reaction with DPP-ITC/pyridine suggested that an oxazolinone was not an intermediate in the formation of the C-terminal thiohydantoin by this chemistry. Application of this chemistry to peptides covalently coupled to a solid support revealed that the DPP-ITC reaction could be separated from the pyridine reaction. Because excess DPP-ITC was completely removed by a solvent wash step, we concluded that pyridine does not act as a catalyst to form the acyl phosphate. Indeed this derivative must form without the elimination of -NCS as suggested by Kenner et al. (1953), otherwise the -NCS would be lost in the wash step preventing formation of the thiohydantoin. Because the addition of pyridine after removal of DPP-ITC is required to complete the reaction, we conclude that pyridine promotes the rearrangement of the acylphosphorylisothiocyanate to the acylisothiocyanate with concomitant release of diphenylphosphate. Additionally, when the pyridine and DPP-ITC steps were separated, treatment of the covalently coupled peptide with a base, such as triethylamine, was found to be absolutely essential for this reaction to proceed. Presumably, this converts the C-terminal carboxylic acid group to a carboxylate, a required form for attack on DPP-ITC. When DPP-ITC and pyridine were added together, pretreatment with triethylamine was not required; thus, pyridine most likely performed this function. Nonetheless, the use of liquid pyridine contributed greatly to the background peaks on the HPLC and was not a desirable step. Thus, our optimal protocol involved first treatment of the immobilized peptide with aqueous triethylamine for a few minutes, removal of the triethylamine solution by drying the support, reaction of the peptide with DPP-ITC in acetonitrile, removal of the DPP-ITC solution by washing the support with either anhydrous DMF or heptane, and then treatment of the peptide with pyridine in the gas phase.

The fact that the DPP-ITC and pyridine steps could be separated led us to propose the reaction scheme shown in Figure 2. The C-terminal carboxylic acid is first converted to a carboxylate with triethylamine, the carboxylate can then react with DPP-ITC via an equilibrium mechanism. Pyridine can then promote removal of diphenylphosphate, presumably through a concerted mechanism to form an acyl isothiocyanate, which can then rapidly form the desired thiohydantoin. Pyridine has recently been shown to promote a similar type of reaction by removing carbon dioxide from the mixed anhydride formed on reaction of an amino acid carboxylate with ethyl chloroformate to form an amino acid ester (Husek, 1991). The success of this reaction was dependent on pyridine being present in a molar excess with respect to the ethyl chloroformate. The ability to wash the covalently coupled peptide with a solvent just after reaction with DPP-ITC and before pyridine treatment without affecting the production of peptidylthiohydantoin supports the conclusion that the isothiocyanate ion was not liberated on the reaction of DPP-ITC with the C-terminal carboxylate. If it were, the isothiocyanate ion would have been washed away before introduction of pyridine and a C-terminal thiohydantoin would not have been able to be formed. Separation of the DPP-ITC and pyridine steps was found to be essential for the successful derivatization of C-terminal aspartate to a thiohydantoin. This was most likely due to catalysis of cyclic anhydride formation by pyridine when present with an excess of DPP-ITC. Pyridine was found to catalyze cyclic anhydride formation of C-terminal glutamate when present during peptidylthiohydantoin formation by acetic anhydride/TMS-ITC (Bailey & Shively, 1990).

Automation of the DPP-ITC/pyridine chemistry

The composition of the reagents and solvents used for sequencing and a summary of the program used are shown in Tables 1 and 2, respectively. The reaction sequence, triethylamine, DPP-ITC, pyridine, was repeated three times per cycle in order to insure that the equilibrium reaction between DPP-ITC and the C-terminal carboxylate proceeds to completion. Pyridine was chosen for our automation experiments because it could be delivered in the gas phase and was found to give the best repetitive yields. Methanolic triethylamine was used rather than aqueous triethylamine in order to prevent any unwanted hydrolysis of the already derivatized C-terminal amino acid during the multiple derivatization reaction. Aqueous triethylamine was shown to rapidly hydrolyze C-terminal thiohydantoins, whereas triethylamine in methanol was shown to hydrolyze C-terminal thiohydantoins very slowly if at all (Bailey & Shively, 1991). Both aqueous and methanolic triethylamine were found to be equally effective in converting the C-terminal carboxylic acid to a carboxylate. The sample was then extensively washed with methanol and DMF. Specific cleavage of the derivatized amino acid was performed with sodium trimethylsilanolate to generate a thiohydantoin amino acid, which was subsequently analyzed by online reverse-phase HPLC, and a shortened peptide with a silylated carboxylic acid. The free carboxylic acid was regenerated by a second treatment with sodium trimethylsilanolate. The cleavage reaction worked equally as well with potassium trimethylsilanolate. The successful use of either reagent for the cleavage reaction was dependent on an oxygen-free environment. A possible reason for this may be instability of the released thiohydantoin amino acids under basic conditions in the presence of oxygen. We have observed breakdown of the released thiohydantoin amino acid when cleavage was performed by sodium trimethylsilanolate in the absence of an inert atmosphere. Reducing agents such as dithiothreitol or dithioerythritol have been

shown to protect the thiohydantoin amino acids from degradation under basic conditions (Inglis et al., 1989; Miller et al., 1989).

Comparison of the acetic anhydride/TMS-ITC and DPP-ITC/pyridine chemistry

As discussed above, the advantages of DPP-ITC over TMS-ITC primarily involve the amino acids (Glu, Ser, Thr, Asp), which have decreased yields due to side reactions with acetic anhydride. Asp gives two to three times higher sequencing yields, and Thr, Ser, and Glu all sequence in good yield. Unfortunately, the DPP-ITC chemistry is not capable of derivatizing Pro. Because it is likely that this chemistry does not require an oxazolinone intermediate, this leaves only two explanations for the unreactivity of proline: the proline amide is not nucleophilic enough to attack the isothiocyanate, or the side chain of proline causes steric hindrance to cyclization of the five-membered ring. The lack of reaction of proline with DPP-ITC to form a thiohydantoin leads to the conclusion that derivatization of C-terminal proline to a thiohydantoin is not possible and that the development of a new reagent or special proline step will be required to overcome this problem. Also achieved was a reduction in the amount of precleavage. This is reflected by the average initial yields (peptide samples) of the acetic anhydride/TMS-ITC chemistry and the DPP-ITC/pyridine chemistry, which are 30% and 50%, respectively. The current cycle time for the DPP-ITC chemistry is 90 min for peptides and 60 min for proteins. It is likely that with further optimization the cycle time for peptide samples can be reduced to that used for proteins. The cycle time for the TMS-ITC/acetic anhydride chemistry was 120 min (Bailey et al., 1992). Another potential advantage of the DPP-ITC/pyridine chemistry is the ability to apply certain steps, in particular the pyridine step, to the gas phase. This will permit this chemistry to mimic current methods of N-terminal sequencing chemistry. As shown by Figures 5 and 6, the use of gas-phase pyridine can permit C-terminal sequencing of protein samples without sample washout. Optimization of this technique will routinely permit multiple cycles of C-terminal sequence analysis on noncovalently applied samples, similar to what is currently employed in N-terminal sequence analysis. Elimination of the need to covalently couple protein samples to a solid support has removed the potential for sample loss, which typically occurs during covalent coupling procedures and therefore permits all of any particular protein sample to be used for the analysis. Further work with this chemistry is aimed at sequencing peptides, which are also noncovalently applied to Zitex.

In summary, we have developed a more effective derivatization chemistry for C-terminal sequence analysis of proteins and peptides using the thiohydantoin method. In particular, this method offers increased yields of derivatization with C-terminal aspartic acid and faster cycle times as compared to previous methodology. This is also the first example of an automated method for C-terminal sequence analysis, which can be applied to subnanomolar amounts of protein samples. It is anticipated that this chemistry can be performed on instruments that are already commercially available for protein sequencing. Work is continuing on the optimization of the automated conditions, such as improved washing conditions to lower backgrounds, and on the construction of a C-terminal sequencing instrument with enhanced capabilities to optimize this chemistry. Work is also in progress on the optimization of this chemistry with protein samples and toward the development of methodology and reagents capable of derivatizing proline.

Materials and methods

Materials

Acetic anhydride was purchased from Fisher. TMS-ITC, diphenyl chlorophosphate, anhydrous DMF, anhydrous acetonitrile, and anhydrous pyridine were from Aldrich. Water was purified on a Millipore Milli Q system. Sodium trimethylsilanolate was obtained from Fluka. DPP-ITC was synthesized as described by Kenner et al. (1953). All of the peptides used in this study were either obtained from Bachem or Sigma. Triethylamine (sequenal grade) and 1,3-dicyclohexylcarbodiimide (DCC) were obtained from Pierce. The carboxylic acid-modified polyethylene membranes were from the Pall Corporation (Long Island, New York). Zitex G-110 was from Norton Performance Plastics. The amino acid thiohydantoins used in this study were synthesized as described (Bailey & Shively, 1990). The Reliasil HPLC columns used in this study were obtained from Column Engineering (Ontario, California).

Covalent coupling of peptides to carboxylic acid-modified polyethylene

Peptides were covalently coupled to carboxylic acid-modified polyethylene and quantitated as described (Shenoy et al., 1992).

Application of protein samples to Zitex

The Zitex support (2 × 10 mm) was prewet with methanol, and protein samples (2–5 μL) dissolved in water were applied. In some cases (superoxide dismutase), 4 μL of acetic anhydride were applied to the strip and allowed to dry completely before sequencing.

HPLC separation of the amino acid thiohydantoins

Reverse-phase HPLC separation of the thiohydantoin amino acid derivatives was performed on a C-18 (3 μ,

100 Å) Reliasil column (2.0 mm × 250 mm) on a Beckman 126 pump module with a Shimadzu (SPD-6A) detector. The column was eluted for 2 min with solvent A (0.1% trifluoroacetic acid [TFA] in water) and then followed by a discontinuous gradient to solvent B (10% methanol, 10% water, 80% acetonitrile) at a flow rate of 0.15 mL/min at 35 °C. The gradient used was as follows: 0% B for 2 min, 0–4% B over 3 min, 4–35% B over 35 min, 35–50% B for 10 min, and 50–0% B over 2 min. Absorbance was monitored at 265 nm.

HPLC separation of peptides

Reverse-phase HPLC was performed on a Vydac C18 column (2.1 mm × 250 mm) on a Beckman System Gold with a Shimadzu (SPD-6A) detector. Peptide samples were monitored by absorbance at 214 nm. A linear gradient from 2% solvent B to 60% solvent B over 30 min at a flow rate of 0.25 mL/min was employed at room temperature. Solvents A and B, respectively, were 0.1% TFA in water and 0.1% TFA in 90% acetonitrile/9.9% H_2O.

Reaction of N-acetyl methionine and N-acetyl proline with DPP-ITC

The *N*-acetyl amino acid (0.0034 mol) was added to 10 mL of acetonitrile containing 1 mL of pyridine (0.0125 mol). DPP-ITC (0.008 mol) in acetonitrile was added. The mixture was stirred at 40 °C for 1 h. The solution was then poured onto 100 g of cracked ice with stirring. The resulting product was either collected by filtration or evaporated to an oil. Deacetylation was performed by adding 12 N HCl (20 mL) and stirring at room temperature for 2–6 h. The acidified residue was evaporated to dryness under vacuum and recrystallized from water.

Formation of peptidylthiohydantoins with DPP-ITC in the solution phase

The sample peptide (60 nmol) was placed in a 1.5-mL polypropylene tube and dried in a vacuum centrifuge. On addition of acetic anhydride (100 μL) the peptide solution was allowed to incubate at 50 °C for 10 min The reaction was then taken to dryness in a vacuum centrifuge. The dried peptide was then dissolved with 100 μL of a 2% solution of triethylamine in water. After incubation for 5 min at 25 °C, the solution was again taken to dryness in a vacuum centrifuge. The peptide was then dissolved in 40 μL of anhydrous acetonitrile. Pyridine (10 μL, 0.012 mmol) was added. This was followed by the addition of 60 μL of DPP-ITC (0.06 mmol) in acetonitrile. The solution was allowed to incubate at 50 °C for 30 min. The reaction was then dried in a vacuum centrifuge, dissolved in water (50 μL), and taken to dryness in a vacuum centrifuge. Dried samples were taken up in 0.1% TFA in water (100 μL) and analyzed by reverse-phase HPLC.

Mass spectrometry

Positive ion FAB/MS was performed as described by Hefta et al. (1988). Peptide samples from HPLC were collected in 1.5-mL polypropylene tubes, dried in a vacuum centrifuge, and redissolved in 2–3 μL of 5% aqueous acetic acid. An aliquot (1–2 μL) of the peptide was added directly to 2–3 μL of glycerol on the sample stage. The amino acid thiohydantoins were dissolved in methanol (typically 2–3 mg in 200 μL methanol), and 2 μL of this solution was added directly to the glycerol on the sample stage. Data were collected with a JEOL HX-100HF mass spectrometer operating at a 5-kV accelerating potential.

Automation of the C-terminal sequencing chemistry

The instrument used for automation of the chemistry described in this manuscript has been described previously (Bailey et al., 1992). The solvents and reagents for automated sequencing are shown in Table 1, and a summary of the program used for automated sequencing is provided in Tables 2 and 4.

Acknowledgments

We thank Dr. Harvey Brandwein of the Pall Corp. (Long Island, New York) for providing us with the carboxylic acid-modified polyethylene support, Jim Sligar for performing the amino acid compositions, and John Wieser of Column Engineering (Ontario, California) for providing the Reliasil HPLC columns. This work was supported in part by NIH grant GM46022 (to J.M.B.) and by the Tosoh Corporation.

References

Bailey, J.M., Shenoy, N.S., Ronk, M., & Shively, J.E. (1992). Automated carboxy-terminal sequence analysis of peptides. *Protein Sci. 1*, 68–80.

Bailey, J.M. & Shively, J.E. (1990). Carboxy-terminal sequencing: Formation and hydrolysis of C-terminal peptidylthiohydantoins. *Biochemistry 29*, 3145–3156.

Bailey, J.M. & Shively, J.E. (1991). Carboxy terminal sequencing: Automation and application to the solid phase. In *Techniques in Protein Chemistry: II* (Villafranca, J.J., Ed.), pp. 115–129. Academic Press, New York.

Hefta, S.A., Hefta, L.J.F., Lee, T.D., Paxton, R.J., & Shively, J.E. (1988). Carcinoembryonic antigen is anchored to membranes by covalent attachment to a glycosylphosphatidylinositol moiety: Identification of the ethanolamine linkage site. *Proc. Natl. Acad. Sci. USA 85*, 4648–4652.

Husek, P. (1991). Rapid derivatization and gas chromatographic determination of amino acids. *J. Chromatogr. 552*, 289–299.

Inglis, A.S. (1991). Chemical procedures for C-terminal sequencing of peptides and proteins. *Anal. Biochem. 195*, 183–196.

Inglis, A.S., Moritz, R.L., Begg, G.S., Reid, G.E., Simpson, R.J., Graffunder, H., Matschull, L., & Wittmann-Liebold, B. (1991). C-terminal sequence analysis. In *Methods in Protein Sequence Analysis* (Jornvall, H., Hoog, J.O., & Gustavsson, A.M., Eds.), pp. 23–34. Birkhauser-Verlag, Basel.

Inglis, A.S., Wilshire, J.F.K., Casagranda, F., & Laslett, R.L. (1989). C-terminal sequencing: A new look at the Schlack-Kumpf thiocyanate degradation. In *Methods in Protein Sequence Analysis* (Wittmann-Liebold, B., Ed.), pp. 137–144. Springer-Verlag, Berlin.

Kenner, G.W., Khorana, H.G., & Stedman, R.J. (1953). Peptides. Part IV. Selective removal of the C-terminal residue as a thiohydantoin. The use of diphenylphosphoroisothiocyanatidate. *Chem. Soc. J.*, 673–678.

Miller, C.G., Kong, C.-T., & Shively, J.E. (1989). Carboxy terminal sequence analysis of proteins and peptides by chemical methods. In *Techniques in Protein Chemistry* (Hugli, T.E., Ed.), pp. 67–78. Academic Press, New York.

Rangarajan, M. (1988). Chemical methods of amino acid sequence analysis from carboxyterminal end. In *Protein/Peptide Sequence Analysis: Current Methodologies* (Brown, A.S., Ed.), pp. 135–144. CRC Press, Boca Raton, Florida.

Shenoy, N.S., Bailey, J.M., & Shively, J.E. (1992). Carboxylic acid modified polyethylene: A novel support for the covalent immobilization of polypeptides for C-terminal sequencing. *Protein Sci. 1*, 58–67.

Ward, C.W. (1986). Carboxyl terminal sequence analysis. In *Practical Protein Chemistry—A Handbook* (Darbre, A., Ed.), pp. 491–525. John Wiley and Sons, New York.