Protein Science (1992), *1*, 68–80. Cambridge University Press. Printed in the USA.

Automated carboxy-terminal sequence analysis of peptides

JEROME M. BAILEY, NARMADA R. SHENOY, MICHAEL RONK, AND JOHN E. SHIVELY
Beckman Research Institute of the City of Hope, Division of Immunology, 1450 E. Duarte Road, Duarte, California 91010

(RECEIVED May 10, 1991; ACCEPTED August 26, 1991)

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 solution phase conditions for formation of the peptidylthiohydantoins with trimethylsilylisothiocyanate (TMS-ITC) and for hydrolysis of these peptidylthiohydantoins into an amino acid thiohydantoin derivative and a new shortened peptide capable of continued degradation (Bailey, J.M. & Shively, J.E., 1990, *Biochemistry 29*, 3145–3156). The current study is a continuation of this work and describes the construction of an instrument for automated C-terminal sequencing, the application of the thiocyanate chemistry to peptides covalently coupled to a novel polyethylene solid support (Shenoy, N.R., Bailey, J.M., & Shively, J.E., 1992, *Protein Sci. 1*, 58–67), the use of sodium trimethylsilanolate as a novel reagent for the specific cleavage of the derivatized C-terminal amino acid, and the development of methodology to sequence through the difficult amino acid, aspartate. Automated programs are described for the C-terminal sequencing of peptides covalently attached to carboxylic acid-modified polyethylene. The chemistry involves activation with acetic anhydride, derivatization with TMS-ITC, and cleavage of the derivatized C-terminal amino acid with sodium trimethylsilanolate. The thiohydantoin amino acid is identified by on-line high performance liquid chromatography using a Phenomenex Ultracarb 5 ODS(30) column and a triethylamine/phosphoric acid buffer system containing pentanesulfonic acid. 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 were found to sequence in high yield (90% or greater) except for asparagine and aspartate, which could be only partially removed, and proline, which was found not to be capable of derivatization. In spite of these current limitations, the methodology should be a valuable new tool for the C-terminal sequence analysis of peptides.

Keywords: automated sequencing technique; carboxy-terminus; isothiocyanate reagents

The chemical sequential analysis of peptides and proteins from the amino-terminus has been well established (Edman, 1950). On the contrary, a satisfactory chemical sequential analysis of peptides and proteins from the carboxy-terminus has not been demonstrated. Such a procedure, in addition to complementing amino-terminal degradation, would be invaluable for the detection of post-translational processing at the carboxy-terminus of expressed gene products 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), the method based on the derivatization of the carboxy-terminal amino acid to a 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).

The chemical scheme used for the thiocyanate degradation is outlined in Figure 1. Each cycle of degradation

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

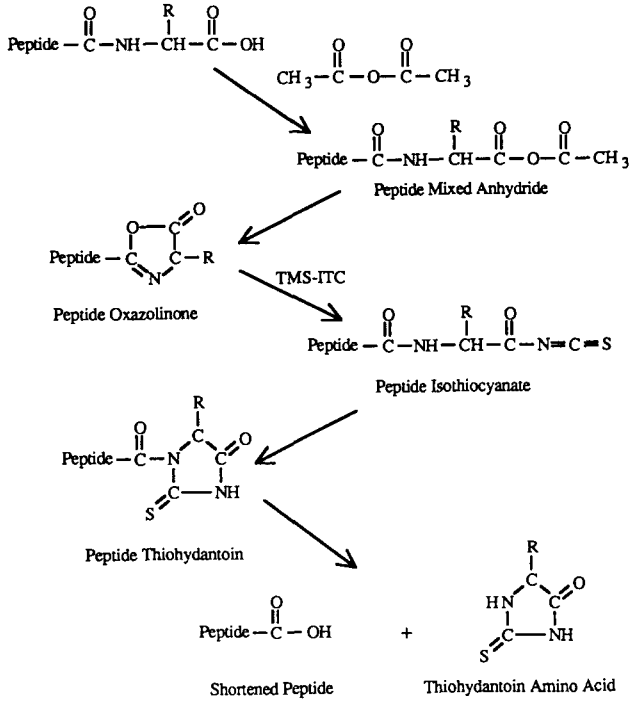

Fig. 1. Reaction scheme for the sequential C-terminal degradation of peptides using the thiocyanate chemistry.

consists of three steps: activation of the carboxy-terminal carboxylic group, derivatization to a thiohydantoin amino acid, and specific hydrolysis of the derivatized C-terminal amino acid to a thiohydantoin amino acid derivative and a shortened polypeptide or protein capable of continued degradation. Current limitations of this method include a limited number of cycles due to sample losses inherent during manual manipulation, a reduced yield when asparagine is encountered during the degradation, and the inability of the method to degrade proline or aspartate when it is at the C-terminus of peptides or proteins. Application of this chemistry to the solid phase and to automation has the potential to permit extended sequential degradation from the C-terminus of peptides and proteins with minimal sample loss. Advantages to the solid phase approach include the ability to use reagents and solvents optimal for sequencing without causing sample washout and the capability to wash the sample efficiently to remove reaction by-products that might otherwise interfere with identification of the released thiohydantoin amino acids.

The need to apply the chemistry for C-terminal sequencing to the solid phase was recognized early on by a number of groups. Solution phase methods required laborious gel filtrations for purification of the peptidylthiohydantoin and evaporations to remove reagents (Stark, 1968). By covalently coupling a polypeptide to a solid support and thereby providing the ability to rinse away excess solvents simply without loss of the sample, the time required for a single cycle of degradation could be reduced to hours rather than the several days required by solution methods. Williams and Kassell (1975) were able to perform one to three cycles on peptides (1 mmol) covalently attached to *N*-hydroxysuccinimide-activated glass beads using ammonium thiocyanate for derivatization and 12 N HCl for cleavage of the peptidylthiohydantoins. Utilizing this same procedure, Rangarajan and Darbre (1976) were able to perform six cycles on ribonuclease (1 μmol) covalently coupled to glass beads with a cycle time of 5–6 h. Three successful cycles, with high performance liquid chromatographic (HPLC) identification of the released amino acid thiohydantoins, were performed by Meuth et al. (1982) on a 22-amino acid polypeptide (350 nmol) covalently linked to carbonyldiimidazole-activated aminopropyl glass. These authors used thiocyanic acid for derivatization to a peptidylthiohydantoin and acetohydroxamate for cleavage, further reducing the time per cycle to 3 h. Inglis et al. (1989) illustrated the sequential degradation of nine residues from a synthetic decapeptide (30 nmol) covalently coupled to glass beads with a cycle time of 48 min, but few experimental details were given.

This report describes the construction of an instrument capable of C-terminal sequence analysis and the application of the thiocyanate chemistry to model peptides covalently coupled to a novel polyethylene membrane support (Shenoy et al., 1992) containing a variety of C-terminal amino acid side chains. The results of analysis of all 20 of the amino acid side chains commonly found in proteins were analyzed in order to optimize conditions for the formation of peptidylthiohydantoins with trimethylsilylisothiocyanate (TMS-ITC) and for cleavage of the peptidylthiohydantoin derivatives into a shortened peptide and thiohydantoin amino acid derivative. Additionally, the use of sodium trimethylsilanolate for the cleavage of the derivatized C-terminal amino acid is described. Previous cleavage reagents such as acetohydroxamate (Stark, 1968), aqueous triethylamine (this report), and 12 N HCl (Hawke et al., 1987) have performed poorly on solid phase supports when used in an automated instrument. The sodium trimethylsilanolate reagent permits, for the first time, an extended sequential degradation of a polypeptide from the C-terminal end in high yield.

Results

High performance liquid chromatographic 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 × 25 mm) (Bailey & Shively, 1991). Although this column provided an excellent sep-

aration of the derivatized amino acids, its instability to the acidic pH (2.1) used for the separation was a problem. In order to increase the stability of the column, a higher pH (3.2) solvent system for the separation of the thiohydantoin amino acids was developed. This new separation system utilizes a triethylamine/phosphoric acid buffer system containing alkylsulfonates of varying chain lengths as ion pair reagents in order to place the positively charged amino acids (thiohydantoin histidine and arginine) optimally in the chromatogram. The separation of the thiohydantoin amino acids (250 pmol) with this new solvent system is shown in Figure 2. The phenylthiohydantoin derivative of glutamate (250 pmol) was included as an internal standard and demonstrates how much more hydrophilic and earlier eluting the thiohydantoin derivatives are as compared to the more hydrophobic phenylthiohydantoin derivatives.

Automated sequencing of peptides covalently coupled to carboxylic acid-modified polyethylene (PE-COOH)

Based on our solution phase studies (Bailey & Shively, 1990) and our initial automated procedure (Bailey & Shively, 1991), the most plausible solid phase sequencing strategy involved the following series of steps: (1) covalent immobilization of the sample to a solid support, (2) activation with acetic anhydride, (3) removal of acetic anhydride, (4) reaction with TMS-ITC in acetonitrile, and (5) cleavage of the derivatized C-terminal amino acid with sodium trimethylsilanolate (Fig. 3). Methods for the covalent attachment of peptides to PE-COOH supports have been optimized and are described by Shenoy et al. (1992). Initial automated sequencing results with peptides covalently immobilized on PE-COOH showed that the success of the cleavage reaction with sodium trimethylsilanolate was highly dependent on both the solvent used and the need to exclude oxygen. Of the solvents tested, only alcohols were found to be able to promote the cleavage reaction. Methanol, isopropanol, and trimethylsilylethanol were all found to be suitable solvents for dissolving the reagent and promoting cleavage of the peptidylthiohydantoin. The use of methanol as a solvent was found to facilitate the breakdown of some of the thiohydantoin amino acids, particularly glycine. This may be due to the formation of sodium methoxide. The thiohydantoin amino acids are known to be unstable when exposed to basic conditions (Scoffone & Turco, 1956). This was also a problem when isopropanol was used as a solvent. The breakdown of thiohydantoin glycine was completely

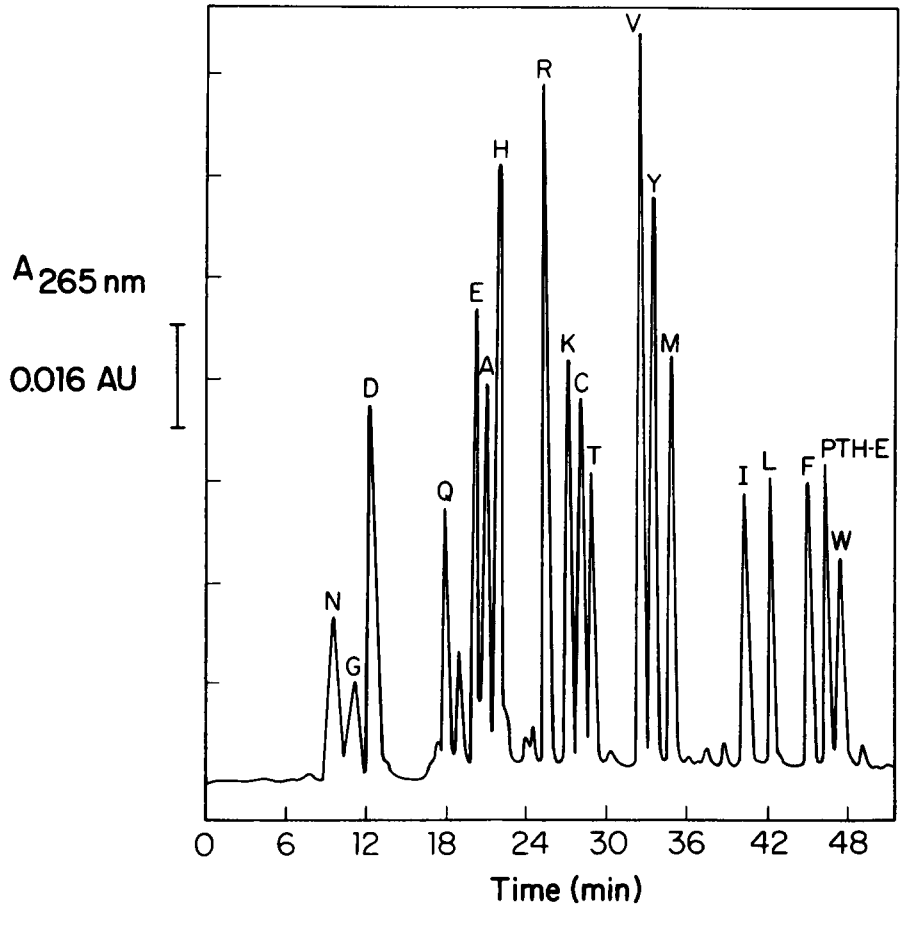

Fig. 2. High performance liquid chromatographic separation of the amino acid thiohydantoin standards. The amino acid thiohydantoin standards (250 pmol) were separated on a Phenomenex Ultracarb 5 ODS(30) column as described under Materials and methods. Absorbance was monitored at 265 nm at 0.16 AUFS. The thiohydantoin derivative of cysteine is S-methylated.

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

R1, Acetic anhydride
R2, 30% TMS-ITC in acetonitrile (v/v)
R3, 0.10 M Sodium trimethylsilanolate in 50% methanol, 50% *t*-butyl alcohol
R4, Methanol
S1, 5% Triethylamine in water
S2, 1.0% TFA in water
S3, 50% Methanol, 50% water
S4, Acetonitrile

eliminated with the use of trimethylsilylethanol as a solvent, presumably due to its lesser tendency to form an alkoxide. The only drawback to this solvent is its high boiling point (71–73 °C/35 mm). Once the cleaved thiohydantoin was transferred over to the conversion flask during automating sequencing, removal of excess solvent by evaporation was difficult and resulted in the formation of a two-phase system when aqueous trifluoroacetic acid (TFA) was added for HPLC injection. This resulted in poor recoveries of the released thiohydantoin amino acid. A more suitable solvent has been found to be a 50% (v/v) solution of *tert*-butyl alcohol and methanol. Both of these solvents have a low boiling temperature and are easily removed prior to injection of the released thiohydantoin amino acid into the HPLC. All of the thiohydantoin amino acids were found to be stable to cleavage in this solvent. A number of different concentrations of sodium trimethylsilanolate have been tested. The optimum concentration, in terms of stability of the released thiohydantoin amino acids and rate of cleavage, has been found to be 0.1 M. The sequence of several peptides, containing a variety of different amino acid side chains, from the C-terminus is shown in Figures 4–6. The composition of the reagents and solvents used and a description of the program used are shown in Tables 1 and 2, respectively. The total time required for each cycle was 120 min. The automated C-terminal sequencing of leucine enkephalin, YGGFL (24.2 nmol), is shown in Figure 4. The initial yield was 30%. All five cycles were easily identified, although the yield of thiohydantoin tyrosine in cycle 5 was reduced. Because this is the amino acid from

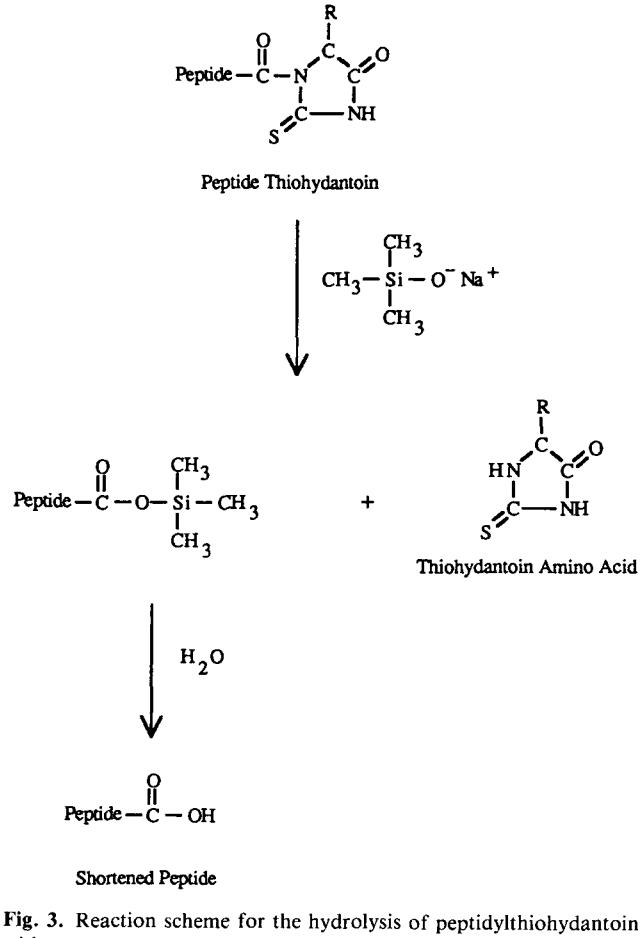

Fig. 3. Reaction scheme for the hydrolysis of peptidylthiohydantoin with sodium trimethylsilanolate.

Table 2. *C-terminal sequencer program summary*

Continuous flow reactor (CFR) (65 °C)	Conversion flask (CF) (55 °C)	Duration (s)
S1 reaction		120
S1 reaction		120
	R4 rinse	120
Dry	Dry	100
R1 reaction		180
R2 reaction		450
S4 rinse		60
R1 reaction		180
R2 reaction		450
S4 rinse		60
R1 reaction		180
R2 reaction		450
S4 rinse		60
R1 reaction		180
R2 reaction		450
S4 rinse		120
S2 rinse		60
S4 rinse		120
S3 rinse		120
S4 rinse		120
R3 reaction		1,200
R3 to CF		45
R3 reaction	Dry	240
R3 reaction	Dry	240
S2 delivery		12
S2 reaction		120
S2 to CF		45
S2 delivery		11
S2 reaction		120
S2 to CF		45
Inject		7
Pause		60

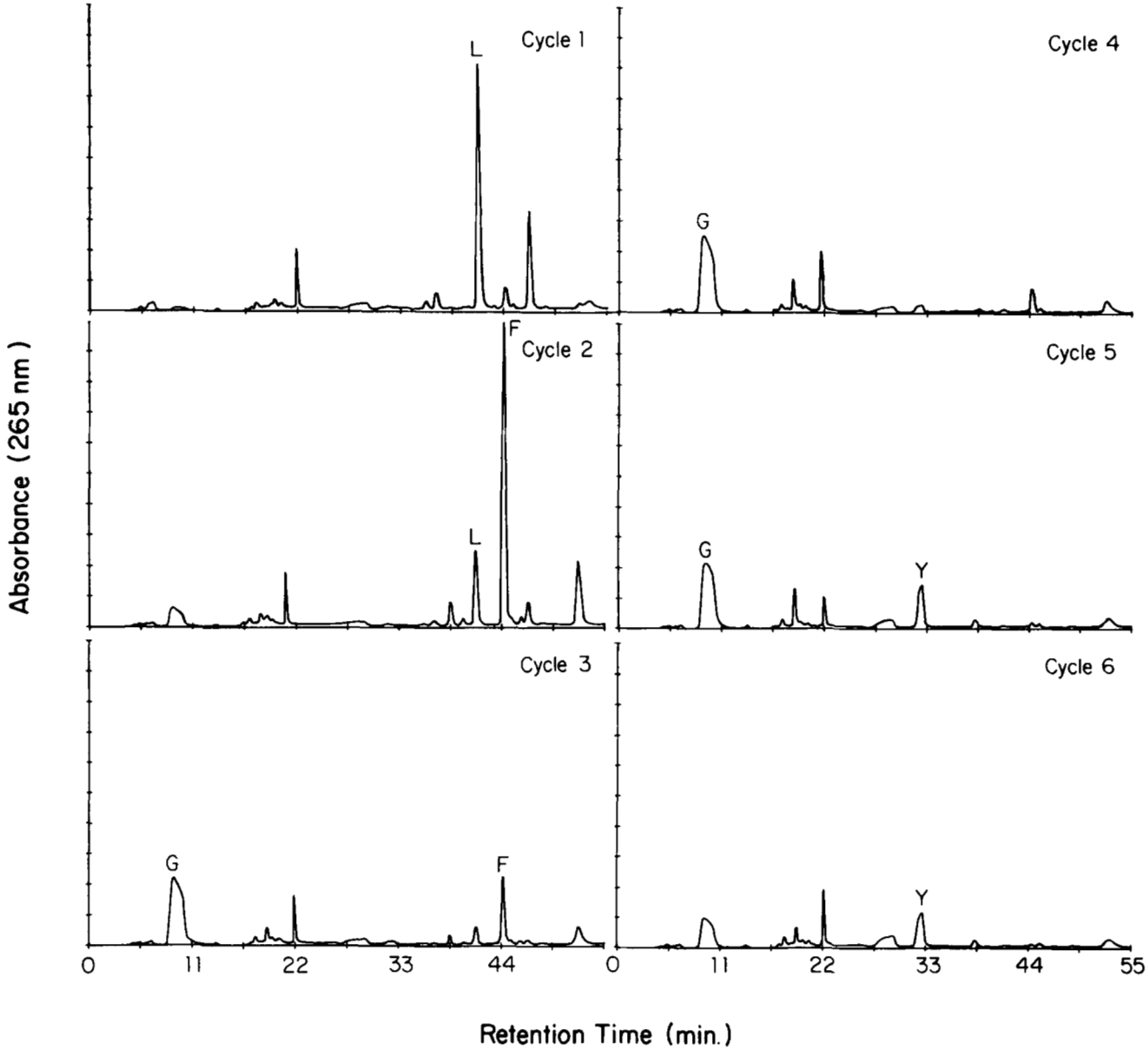

Fig. 4. Automated C-terminal sequencing of YGGFL (24.2 nmol) covalently coupled to PE-COOH. The initial yield of sequencing was approximately 30%.

which the peptide was covalently attached to the solid support, it may be either difficult to derivatize this amino acid or subsequently hydrolyze it due to its proximity to the membrane surface. Figure 5 shows the automated sequencing of the octapeptide, YGGFMRGL (9.6 nmol). The initial yield was 30%. All eight cycles were readily identified, again with low yields of the N-terminal amino acid. Because thiohydantoin-Arg coelutes with a background peak, this cycle was seen as an increase in this background peak at cycle 3. The background peaks are due to products formed by the isothiocyanate reagent. Figure 6 shows automated sequencing from the C-terminus of the hexapeptide, RGYALG (8.1 mol). The initial yield was 50%. As shown in Figures 4–6, the repetitive yield of sequencing begins to decrease as the degradation gets closer to the surface of the solid support. This was confirmed by amino acid analysis of the membrane support both before and after sequencing. In the case of the octapeptide, YGGFMRGL (Figure 5), after nine cycles of sequencing all of the amino acids were removed from the solid support, except for the last two, G and Y. The amount removed of each amino acid decreased as the amino acid became closer to the support. Amino acid analysis of the supports containing the peptides YGGFL

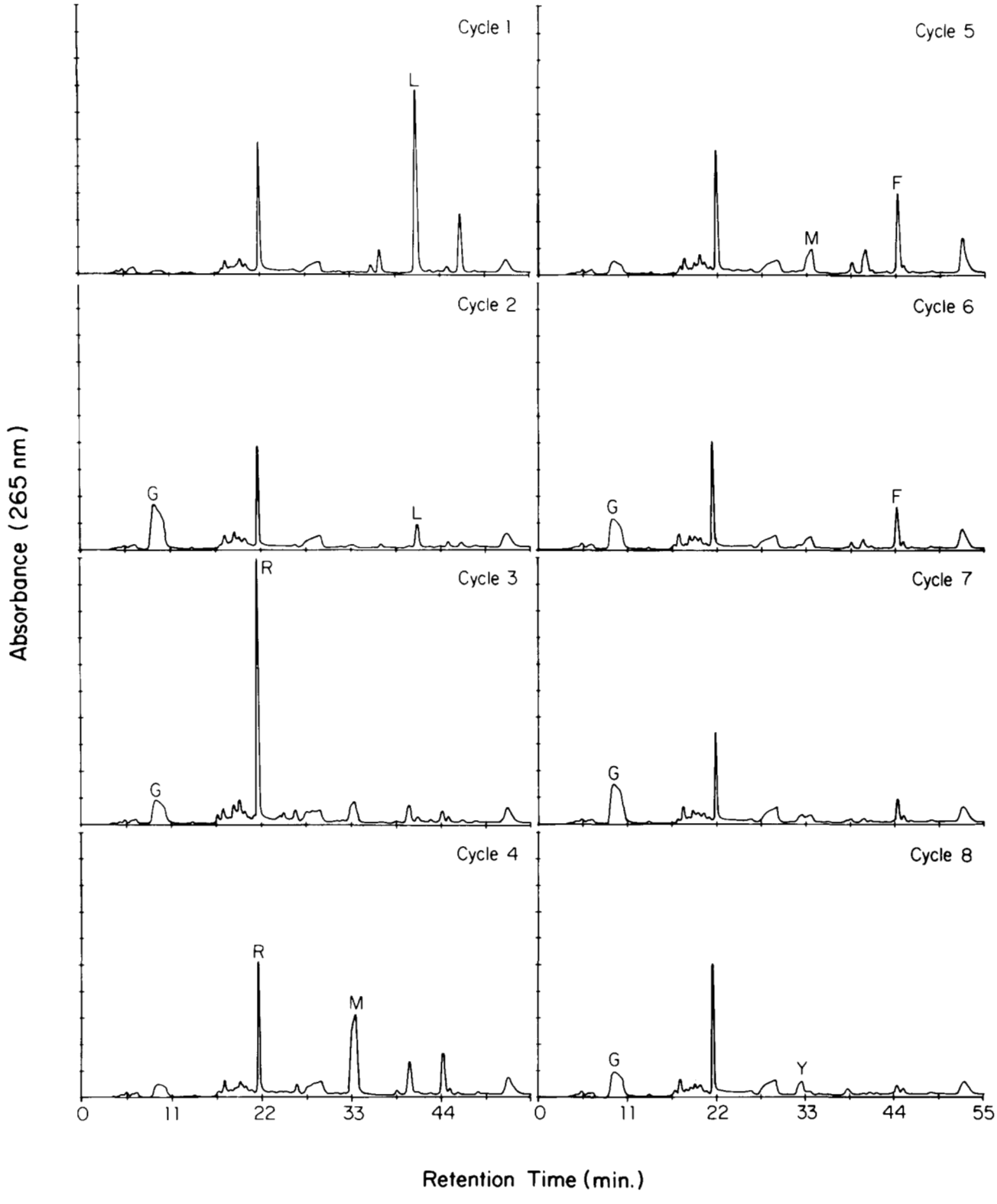

Fig. 5. Automated C-terminal sequencing of YGGFMRGL (9.6 nmol) covalently coupled to PE-COOH. The initial yield of sequencing was approximately 30%.

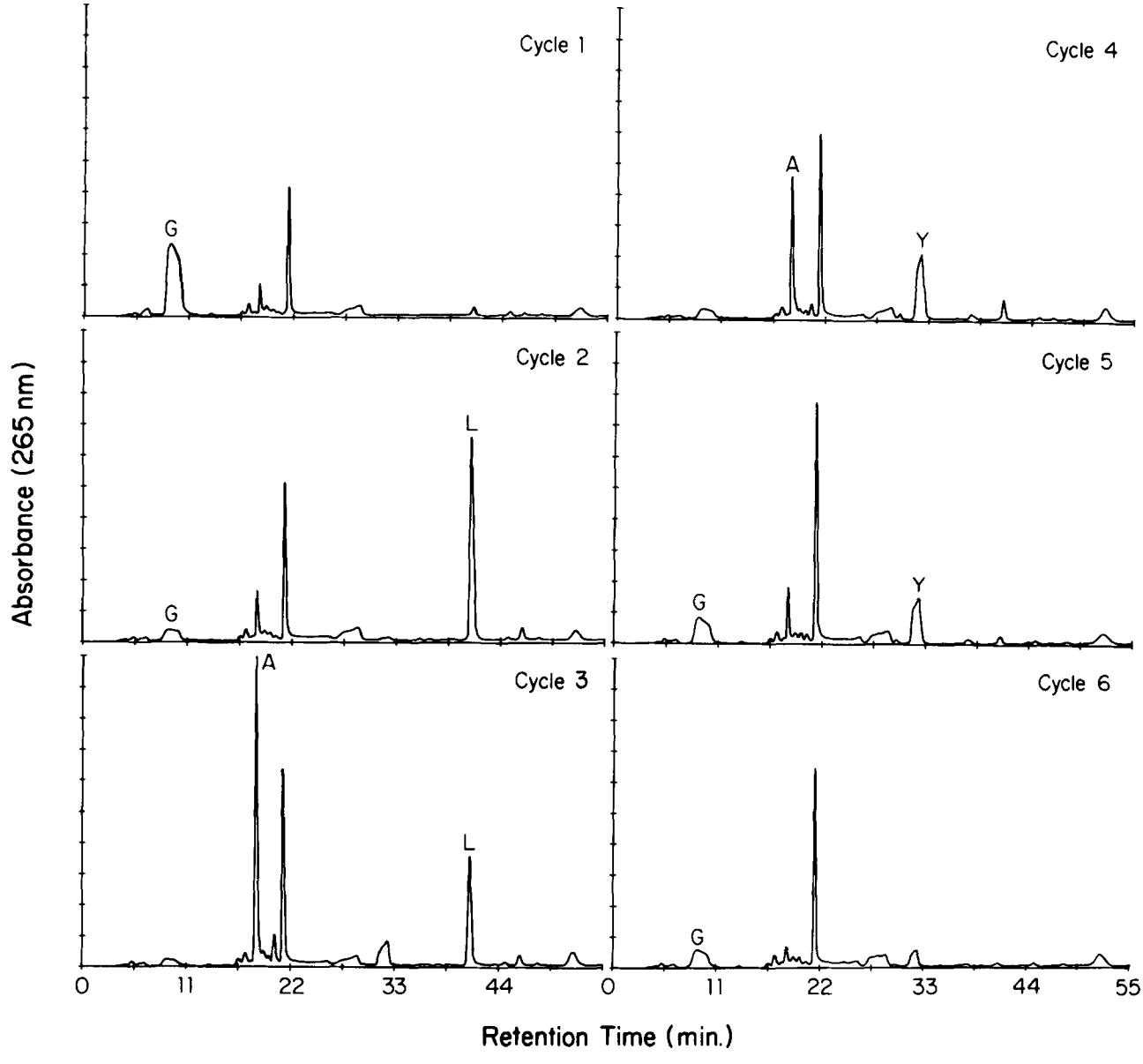

Fig. 6. Automated C-terminal sequencing of RGYALG (8.1 nmol) covalently coupled to PE-COOH. The initial yield of sequencing was approximately 50%.

and YGGFMRGL after six and nine cycles of sequencing, respectively, revealed that 50% of the N-terminal amino acid was not removed.

Unlike the carboxylic acid-modified polyvinylidene difluoride (PVDF) support (Bailey & Shively, 1991), peptidylthiohydantoins covalently coupled to the PE-COOH support can be cleaved by dilute aqueous triethylamine with repetitive yields of approximately 40%. Application of a 2% (v/v) aqueous triethylamine solution to the automated sequence determination of YGGFL (12 nmol) covalently coupled to PE-COOH permitted the sequencing of only three cycles, due to the reduced repetitive yields (data not shown). However, the new cleavage reagent results in much-improved yields ranging from 70 to 95%.

Generality of sequencing

We have applied the above-described automating 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. As discussed earlier (Bailey & Shively, 1990) and again demonstrated in this study, all of the amino acids except for Asn, Asp, and Pro can be sequenced in high yield with the thiocyanate

chemistry. The thiohydantoin amino acid derivatives of Lys and Thr are eluted as two peaks due to partial acetylation of their side chains by acetic anhydride during the activation step. As with N-terminal Edman chemistry, the Ser and Thr derivatives are converted to their corresponding dehydro derivatives. This occurs quantitatively in the case of serine and partially in the case of threonine. C-terminal Cys was also readily sequenced with this methodology. The side-chain sulfhydryl group was readily lost during the cleavage reaction, forming a derivative indistinguishable from that formed with Ser (dehydroalanine). 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.

Discussion

Previous studies in this laboratory (Bailey & Shively, 1990) have focused on solution phase conditions for formation of the peptidylthiohydantoins with TMS-ITC and for hydrolysis of these peptidylthiohydantoins into an amino acid thiohydantoin derivative and a new shortened peptide capable of continued degradation. This investigation 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. The current study continues this work by automating the thiocyanate chemistry and applying it to peptides covalently coupled to solid supports.

High performance liquid chromatographic separation of the amino acid thiohydantoins

The hydrophilic nature of the thiohydantoin amino acids as shown in Figure 2 makes their separation by reverse-phase HPLC more problematic than the phenylthiohydantoin amino acids. Particularly difficult are the most hydrophilic thiohydantoin amino acids, Asp, Asn, and Gly, which are poorly retained on the C18 column, resulting in peak broadening. The separation we described in our most recent publication also utilized the Ultracarb ODS(30) column from Phenomenex, but instead used an aqueous TFA buffer for solvent A (Bailey & Shively, 1991). Although this solvent system resulted in slightly better retention of the thiohydantoin-Gly, -Asn, and -Asp analogues, the lower pH (2.1) of this solvent system resulted in a rapid deterioration of the column. The column lifetime was significantly increased (from 2–3 weeks to 2–3 months) by use of the triethylamine/phosphoric acid system (pH 3.2) described in this report. Pentanesulfonic acid was included as an ion pair reagent in order to optimally position the thiohydantoin derivatives of histidine and arginine in the chromatogram. By using alkylsulfonates of varying chain lengths it is possible to position both thiohydantoin-His and -Arg derivatives almost anywhere within the chromatogram. It is therefore anticipated that the coelution of thiohydantoin-Arg with the only major background peak in this chromatogram (Fig. 5) can be rectified by the inclusion of a different length alkyl sulfonate.

The derivatization reaction

As shown in Figure 1, the thiocyanate chemistry for C-terminal sequencing consists of two separate steps: (1) the activation and derivatization of the C-terminal amino acid to form a thiohydantoin, and (2) the specific cleavage of that C-terminal thiohydantoin to form a shortened peptide capable of continued degradation and a thiohydantoin amino acid. Although the derivatization of amino acids to their corresponding thiohydantoins has been studied since this reaction was first reported in 1911 (Johnson & Nicolet, 1911), the mechanism of peptidylthiohydantoin formation by acetic anhydride and thiocyanate ions is still not well understood. Early experiments confirmed the assumption that thiocyanic acid is first formed when acetic anhydride and acetic acid interact with ammonium thiocyanate, and it is the thiocyanic acid that actually reacts to form the thiohydantoin (Johnson & Scott, 1913). The differing ability of the various salts of thiocyanic acid to form a thiohydantoin was reasoned to result from their ability to form thiocyanic acid on interaction with acetic anhydride and acetic acid (Johnson et al., 1915). When more convenient methods for the preparation of thiocyanic acid became available, thiocyanic acid in the presence of acetic anhydride was found to be more reactive for the formation of 2-thiohydantoins than were the thiocyanate salts (Dwulet & Gurd, 1979). As a result, thiocyanic acid has been used by Kubo et al. (1971) and more recently by Inglis et al. (1989) for the sequential degradation of peptides from the C-terminus. However, one of the principal drawbacks of thiocyanic acid is that it tends to be self-reactive, even at ambient temperature, and quickly loses its ability to derivatize the peptide. Furthermore, these polymeric thiocyanic acid products are intensely UV absorbing at the wavelengths used for thiohydantoin detection and can subsequently interfere with the HPLC identification of the released thiohydantoin amino acid. The instability of the free thiocyanic acid presents difficulties when the chemistry is automated, as the reagent needs to be stable to storage in a reagent bottle at room temperature. As discussed by Inglis et al. (1991), one way to help stabilize the thiocyanic acid is to refrigerate it while in the automated instrument. A less costly and more convenient solution to the problems caused by the use of thiocyanic acid was presented in earlier work from our laboratory with the introduction of TMS-ITC for derivatization of the C-terminal amino acid to a thiohydantoin (Hawke et al., 1987). The

trimethylsilyl group offered two advantages: (1) it stabilized the thiocyanate sufficiently so that self-reaction was no longer a problem, and (2) it did not compromise the ability of the thiocyanate to form thiohydantoins. This is consistent with the observation that silylated amines have often been found to be better nucleophiles than the corresponding unsubstituted amines (Fleming, 1979).

The intermediate involved in thiohydantoin formation has been a subject of study for many years. An oxazolinone was postulated to be a necessary intermediate during the synthesis of amino acid thiohydantoins with acetic anhydride and ammonium thiocyanate when this reaction was first studied (Johnson & Scott, 1913). The racemization of the C-terminal amino acids observed on reaction with acetic anhydride and TMS-ITC (Bailey & Shively, 1990) suggests that reaction of peptides with acetic anhydride forms a peptide oxazolinone (Fig. 1). This is consistent with the above-postulated mechanism. The formation of oxazolinones is known to cause racemization of amino acids (Csonka & Nicolet, 1933; Carter & Stevens, 1940; Goodman & Levine, 1964). Further evidence of an oxazolinone intermediate during the formation of amino acid thiohydantoins was obtained by Csonka and Nicolet (1933). Additional studies described by Cornforth (1949) actually demonstrated the formation of an oxazolinone intermediate in the formation of thiohydantoins by the combined use of absorption spectra and polarimetry to follow the rate of oxazolinone formation. In fact, once the oxazolinone was formed, the reaction with isothiocyanic acid was found to be facile enough to occur readily at 0 °C in the case of 2-phenyl-4-benzyl-5-oxazolinone (Cornforth, 1949).

The cleavage reaction

Although the cleavage reaction has been extensively studied since the thiocyanate chemistry for C-terminal degradation was first proposed by Schlack and Kumpf in 1926, a chemical method has not yet been proposed that is capable of an extended degradation. Cleavage with 1 N sodium hydroxide as first proposed by Schlack and Kumpf (1926) is well known to hydrolyze proteins and peptides at other sites in addition to cleavage of the C-terminal peptidylthiohydantoin. The released thiohydantoin amino acid derivatives are also known to be unstable in hydroxide solutions (Scoffone & Turco, 1956). Recent work by Inglis et al. (1991) employed an aqueous solution of potassium hydroxide containing 33% methanol and dithioerythritol (DTE) or dithiothreitol (DTT) for the cleavage reaction. Although the presence of DTE or DTT has been shown to protect the released thiohydantoin amino acid from degradation under these basic conditions (Inglis et al., 1989; Miller et al., 1989), the mercapto group of these molecules has also been shown to form an adduct at the C-terminus of the shortened peptide resulting in the formation of a percentage of shortened peptide blocked to further degradation at each cycle (Miller et al., 1989). This may partially explain the decrease seen in the repetitive yield with each cycle observed by Inglis et al. (1989).

When cleavage of peptidylthiohydantoins by 12 N HCl was applied to proteins and peptides no more than two or three cycles could be performed (Cromwell & Stark, 1969; Hawke et al., 1987). This was probably due to differences in the rate of hydrolysis of peptidylthiohydantoins containing different amino acid side chains as well as to hydrolysis of other internal amide bonds. Likewise, during the synthesis of the standard amino acid thiohydantoin derivatives corresponding to the naturally occurring amino acids (Bailey & Shively, 1990) it was observed that the rate of deacetylation of the *N*-acetylthiohydantoin amino acids by 12 N HCl depended on the nature of the amino acid side chain.

The alternative to aqueous hydroxide or 12 N HCl was to find another reagent capable of rapid and specific hydrolysis of C-terminal peptidylthiohydantoins in an organic solvent. Stark (1968) studied a number of different cleavage reagents and found that oxygen-containing nucleophiles were the best choice of reagents to effect this reaction. Although acetohydroxamate, originally proposed as a result of the study by Stark (1968), is an excellent cleavage reagent for the first amino acid, it forms a stable peptidyl hydroxamate ester that is difficult to remove and which can partially or completely block (depending on the conditions employed) the shortened peptide from further sequencing (Bailey & Shively, 1990). Primary amines, such as *N*-butylamine in trifluoroethanol, were also found to be excellent cleavage reagents by Inglis et al. (1989), but were recently shown by Hawke and Boyd (1991) and Inglis et al. (1991) to permit only one cycle of sequencing, as they form a stable amide at the C-terminus of the shortened peptide, effectively blocking the shortened peptide to further sequencing. It appears that any carbon-based cleavage reagent that is a good nucleophile will also be a poor leaving group, thereby blocking much of the shortened peptide from further sequencing. Ideally a cleavage reagent should possess the following characteristics: (1) it should be able to remove only the derivatized C-terminal amino acid, (2) the shortened peptide should be capable of continued degradation, (3) the released thiohydantoin amino acid should not be destroyed by this reagent, and (4) this reagent should not absorb light in the same range as is used for detection of the released thiohydantoin amino acid derivatives. Although aqueous triethylamine would seem to answer all of these criteria for the cleavage of peptidylthiohydantoins in the solution phase, when applied to automated sequencing on the solid phase several problems became apparent. The principal drawback was the instability of the aqueous solutions of triethylamine in a reservoir bottle for extended periods of time. Many of the breakdown products formed are UV absorbing and

interfere with the identification of the released thiohydantoin amino acids. Because these UV-absorbing products are introduced during the course of the cleavage reaction, they cannot be washed away and are injected into the HPLC along with the thiohydantoin amino acid. Additionally, some of the products formed during the decomposition, such as primary and secondary amines, may be capable of partially blocking the shortened peptide from further sequencing, contributing to the poor repetitive yields observed during the automated sequencing with this reagent.

The new reagent, sodium trimethylsilanolate, utilized in this study for the cleavage of the derivatized C-terminal amino acid does not suffer from the limitations associated with previous reagents and has permitted the extended sequencing of peptides covalently coupled to polyethylene supports in high yield.

Effect of solid support linkers on sequencing yields

In this work we have concentrated our efforts on small peptides and noticed decreasing yields as we approached the solid support. Work is in progress to minimize this surface effect by the introduction of linker arms to the membrane surface prior to the attachment of peptide samples. This will permit the peptide sample being sequenced to be further from the surface of the solid support. Preliminary results using 11-aminoundecanoic acid as a linker arm support this conclusion and resulted in significantly improved yields of sequencing with the pentapeptide, YGGFL. Amino acid analysis revealed only 12% of the tyrosine remaining on the support after six cycles of sequencing.

Generality of sequencing

All 20 of the amino acid side chains commonly found in proteins were tested in our automated C-terminal sequencing procedure in this study. All could be sequenced in high yield except for Asn, Asp, and Pro. As discussed previously during our solution phase studies (Bailey & Shively, 1990), side reactions with asparagine occur during hydrolysis of the peptidylthiohydantoins and are caused by five-membered cyclic imide formation with the asparagine side chain and the peptide–thiohydantoin amide bond during the basic conditions of the cleavage reaction. Such cyclic imide formation can occur both when asparagine is in the C-terminal and C-1 positions. When asparagine is in the C-terminal position some of the new shortened peptide is derivatized to a C-terminal amide (31%) and therefore incapable of continued degradation and when asparagine is in the C-1 position some of the new shortened peptide forms a succimide derivative (64%) that is also not capable of continued degradation. In the former case, when Asn is in the C-terminal position, it was predicted that both thiohydantoin-Asn and thiohydantoin-Asp would be formed (Bailey & Shively, 1990). This was confirmed in this study. Glutamine, which would be expected to form a six-membered cyclic imide in theory, has been found here to degrade smoothly without any evidence of cyclic imide formation. We have found no evidence of the side reactions reported by Inglis et al. (1991) when sequencing glutamine either in this study or in our previous work (Bailey & Shively, 1990).

The ability to sequence through a C-terminal aspartate with the thiocyanate chemistry has remained a subject of controversy since this technique was first described in 1926 (Schlack & Kumpf, 1926). Although some investigators have reported being able to form the thiohydantoin derivatives of aspartate, either free in solution or at the C-terminus of a peptide (Yamashita & Ishikawa, 1971; Inglis et al., 1989), others have been unable to obtain any experimental evidence for the formation of thiohydantoin aspartate (Nicolet, 1930; Swan, 1952; Barker, 1953; Stark, 1968). Aspartate is known to form a cyclic anhydride, which is not reactive to thiocyanate, on reaction with acetic anhydride (Nicolet, 1930; Swan, 1952; Barker, 1953; Stark, 1968). In our earlier study, C-terminal sequencing by the thiocyanate method in the solution phase came to a halt whenever an aspartate was encountered as the C-terminal amino acid (Bailey & Shively, 1990). The addition of a nucleophile prior to reaction with TMS-ITC was found to permit the partial sequencing of this residue.

Our automated sequencing procedure has enabled us to sequence an aspartate residue successfully with the thiocyanate chemistry and, for the first time, to propose an explanation for the apparent contradictions about the ability to sequence through this residue. As discussed previously (Bailey & Shively, 1990), when glutamate is encountered at the C-terminus of a protein or peptide, activation with acetic anhydride first forms an oxazolinone, which is capable of reaction with TMS-ITC, but rapidly converts to the more thermodynamically stable six-membered cyclic anhydride, which is not capable of reaction with TMS-ITC. Minimization of the activation time with acetic anhydride was found to alleviate this effect. The same argument is applicable to aspartic acid, although on a faster time scale, as formation of a five-membered cyclic anhydride with aspartate is more kinetically favored than the six-membered cyclic anhydride with glutamate. As shown in Figure 7, we postulate that the formation of the oxazolinone of C-terminal aspartate, required for thiohydantoin formation, is the initially favored reaction with acetic anhydride, but once formed, rapidly rearranges to the more thermodynamically stable five-membered cyclic anhydride, which is not capable of reaction with TMS-ITC. The automated solid phase method we have developed permits us to perform short reactions with acetic anhydride, rapidly remove it, and then immediately follow with TMS-ITC. This procedure is repeated several times in order to optimize conversion

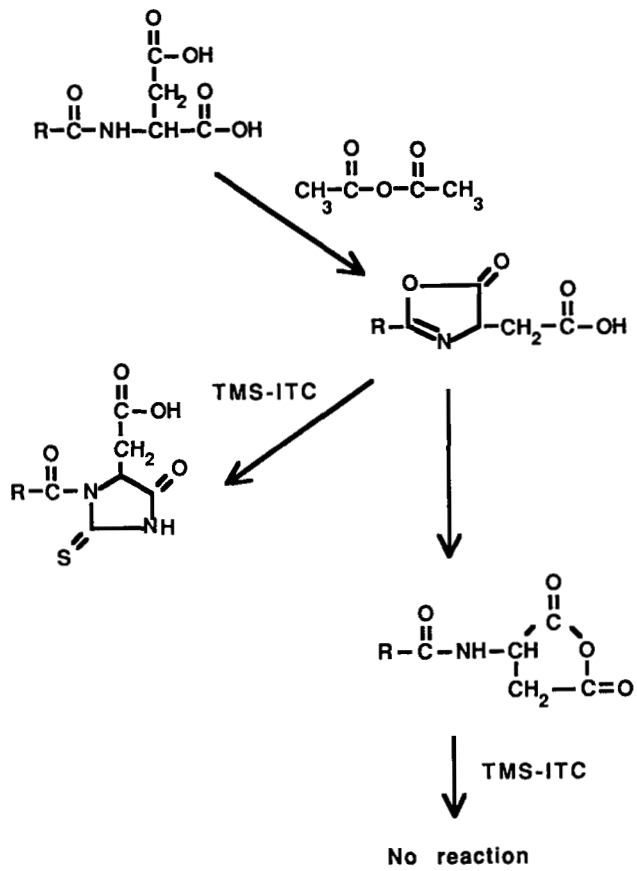

Fig. 7. Reaction scheme for the formation of the thiohydantoin derivative of C-terminal aspartate.

of the C-terminal amino acid to a thiohydantoin (Table 2). As discussed previously (Bailey & Shively, 1990), minimization of the reaction time is necessary to obtain good sequencing yields when Thr, Ser, and Glu are encountered. The automated program described in Table 2 allows us to obtain quantitative derivatization of C-terminal Glu, Thr, and Ser, but only partial derivatization of C-terminal Asp (approximately 50%). This is due to competition of the TMS-ITC reaction with cyclic anhydride formation. The C-terminal aspartate, which is not derivatized to a thiohydantoin during the initial cycle, is converted back to aspartate during the cleavage reaction and is derivatized with TMS-ITC during subsequent cycles. Continuing work is directed toward improving the yields when aspartate is encountered during C-terminal sequencing.

C-terminal proline was found not to be capable of forming a thiohydantoin. This is most likely due to the inability of proline to form an oxazolinone (Matsuo et al., 1966; Holcomb et al., 1968). A report by Inglis et al. (1989) claims to have synthesized proline thiohydantoin, although no characterization data were provided. A more recent report from the same group now acknowledges difficulty in the derivatization of C-terminal proline, but provides little explanation (Inglis et al., 1991). We conclude that a separate procedure may have to be employed for sequencing through proline. Such a procedure could be initiated upon discovering that no thiohydantoins were detected after a given cycle. The current status of the chemistry described here permits sequencing of all the amino acids except proline, which in our opinion is an acceptable first approach to the introduction of automated C-terminal sequence analysis.

In summary, we have constructed an instrument capable of automated C-terminal chemistry, described the use of a novel support for the covalent attachment of polypeptides, optimized the use of a novel reagent for the cleavage of the derivatized C-terminal amino acid, and described methodology for the sequence analysis of all amino acids except proline. In addition, we have optimized conditions for the partial degradation of C-terminal aspartate in yields of about 50%. Work is continuing on the optimization of the automated conditions and toward applying this chemistry to the C-terminal sequence determination of covalently coupled proteins. Work is also in progress toward the development of methodology and reagents capable of derivatizing proline and minimizing the side reactions involved when sequencing Asp and Asn.

Materials and methods

Materials

Acetic anhydride and *O*-phosphoric acid (HPLC grade) were purchased from Fisher. Ammonium thiocyanate, TMS-ITC, anhydrous dimethylformamide (DMF), hexafluoroacetone trihydrate, and pentanesulfonic acid were from Aldrich. Water was purified on a Millipore Milli Q system. Sodium trimethylsilanolate was obtained from Petrarch (Huls). All of the peptides used in this study were either obtained from Bachem or Sigma. Triethylamine (sequanal grade) and 1,3-dicyclohexylcarbodiimide (DCC) were obtained from Pierce. The PE-COOH membranes were from the Pall Corporation (Long Island, New York). The amino acid thiohydantoins used in this study were synthesized as described (Bailey & Shively, 1990).

Covalent coupling of peptides to PE-COOH

Details of the methods used for the covalent attachment of peptides to the PE-COOH are described in the preceding paper (Shenoy et al., 1992).

Instrument design

The C-terminal protein sequencer was constructed (in the cabinet from an out-of-service Beckman 121 amino acid analyzer) by the Biomedical Instrument Facility at the

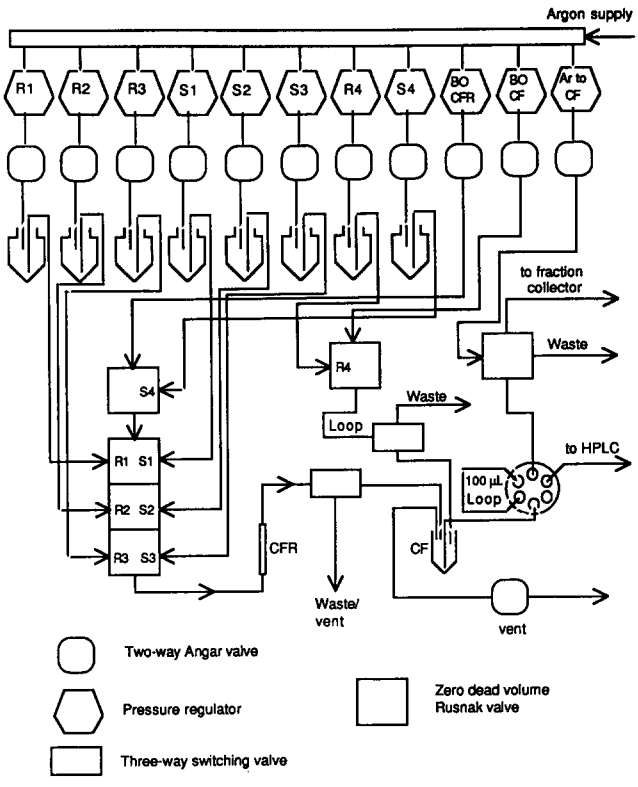

Fig. 8. Schematic of the C-terminal sequencing instrument. Identification of the components is shown at the bottom of the schematic.

City of Hope. A schematic of the instrument is shown in Figure 8. This design is similar to the gas-phase N-terminal sequencer described by Hawke et al. (1985). The instrument is equipped with a temperature-controlled continuous flow reactor (CFR) and conversion flask (CF) as described by Shively et al. (1987). All of the reagents and solvents except for R4 (see Table 1) are directed to the CFR. In order to minimize bubble formation during the delivery of reagents to the CFR, solvent/reagent input is from the bottom of the CFR. Solvent and reagent deliveries to the CFR and conversion flask are achieved by maintaining a low pressure on each bottle (0.5–2.0 p.s.i.) and performing timed deliveries. Unlike the N-terminal sequencers, the conversion flask is used only as a holding vessel for the thiohydantoin amino acid prior to injection onto the on-line HPLC. The cleavage solution containing the C-terminal amino acid as a thiohydantoin derivative is transferred from the CFR to the conversion flask where it is then dried under a stream of argon, taken up in 1% aqueous TFA, and directed to a Rheodyne 7126 pneumatically actuated injector. All the valves shown in Figure 8 are controlled by an Apple II computer and are solenoid actuated. The zero-dead volume Rusnak valve is described by Hawke et al. (1985) and the three-way switching valve is described by Calaycay et al. (1991).

High performance liquid chromatographic separation of the amino acid thiohydantoins

Reverse-phase HPLC separation of the thiohydantoin amino acid derivatives was performed on a Phenomenex Ultracarb 5 ODS(30) column (2.0 × 25 mm) on a Beckman 126 Pump Module with a Shimadzu (SPD-6A) detector. The column was eluted for 5 min with solvent A (0.006% phosphoric acid, 0.006% triethylamine, 0.045% pentanesulfonic acid) and then followed by a discontinuous gradient to solvent B (0.03% phosphoric acid, 0.045% triethylamine, 30% acetonitrile, 0.045% pentanesulfonic acid) at a flow rate of 0.15 mL/min at 22 °C. The gradient used was as follows: 0% B for 5 min, 0–20% B over 7 min, 20–100% B over 25 min, 100% for 23 min, and 100–0% B over 2 min. Absorbance was monitored at 265 nm. The phenylthiohydantoin derivative of glutamate was included as an internal standard.

Acknowledgments

We thank Dr. Harvey Brandwein of the Pall Corp. (Long Island) for providing us with the PE-COOH support and Jim Sligar for performing the amino acid compositions. This work was supported in part by NSF grant BBS-8714266 and by the Tosoh Corporation.

References

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, Inc., New York.

Barker, C.C. (1953). The dehydration and racemisation of N-acyl-L-aspartic acids by acetic anhydride. *J. Chem. Soc.* 453–456.

Calaycay, J., Rusnak, M., & Shively, J.E. (1991). Microsequence analysis of peptides and proteins IX. An improved, compact, automated instrument. *Anal. Biochem. 192*, 23–31.

Carter, H.E. & Stevens, C.M. (1940). Azlactones II. Azlactone formation in glacial and in aqueous acetic acid and preparation of benzoyl-α-aminocrotonic acid azlactone II. *J. Biol. Chem. 133*, 117–128.

Cornforth, J.W. (1949). Oxazoles and oxazolinones. In *The Chemistry of Penicillin* (Clarke, H.T., Johnson, J.R., & Robinson, R.), pp. 688–848. Princeton University Press, Princeton, New Jersey.

Coull, J.M., Dixon, J.D., Laursen, R.A., Koester, H., & Pappin, D. (1989). Development of membrane supports for the solid phase sequence analysis of proteins and peptides. In *Methods in Protein Sequence Analysis* (Wittmann-Liebold, B., Ed.), pp. 69–78. Springer-Verlag, Berlin.

Cromwell, L.D. & Stark, G.R. (1969). Determination of the carboxyl termini of proteins with ammonium thiocyanate and acetic anhydride, with direct identification of the thiohydantoins. *Biochemistry 8*, 4735–4740.

Csonka, F.A. & Nicolet, B.H. (1933). The preparation of optically active thiohydantoins and the racemization of amino acids as their azlactones. *J. Biol. Chem. 99*, 213–216.

Dwulet, F.E. & Gurd, F.R.N. (1979). Use of thiocyanic acid to form 2-thiohydantoins at the carboxyl terminus of proteins. *Int. J. Pept. Protein Res. 13*, 122–129.

Edman, P. (1950). Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand. 4*, 283–293.

Fleming, I. (1979). Organic silicon chemistry. In *Comprehensive Or-*

ganic Chemistry, Vol. 3 (Jones, D.N., Ed.), pp. 541-686. Pergamon Press, New York.

Goodman, M. & Levine, L. (1964). Peptide synthesis via active esters. IV. Racemization and ring-opening reactions of optically active oxazolones. *J. Am. Chem. Soc. 86*, 2918-2922.

Hawke, D. & Boyd, V.L. (1991). Chemical C-terminal sequencing. In *Methods in Protein Sequence Analysis* (Jornvall, H., Hoog, J.O., & Gustavsson, A.M., Eds.), pp. 35-45. Birkhauser-Verlag, Basel.

Hawke, D.H., Harris, D.C., & Shively, J.E. (1985). Microsequence analysis of peptides and proteins V. Design and performance of a novel gas-liquid phase instrument. *Anal. Biochem. 147*, 315-330.

Hawke, D.H., Lahm, H.-W., Shively, J.E., & Todd, C.W. (1987). Microsequence analysis of peptides and proteins: Trimethylsilylisothiocyanate as a reagent for COOH-terminal analysis. *Anal. Biochem. 166*, 298-307.

Holcomb, G.N., James, S.A., & Ward, D.N. (1968). A critical evaluation of the selective tritiation method of determining C-terminal amino acids and its application to luteinizing hormone. *Biochemistry 4*, 1291-1296.

Inglis, A.S., Moritz, R.L., Begg, G.S., Reid, G.E., Simpson, R.J., Graffunder, H., Matschull, L., & Wittman-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.

Johnson, T.B., Hill, A.J., & Bailey, B.H. (1915). Researches on hydantoins. XXXIV. The interaction of hippuric acid with thiocyanates. *J. Am. Chem. Soc. 37*, 2406-2416.

Johnson, T.B. & Nicolet, B.H. (1911). Hydantoins: The synthesis of 2-thiohydantoin. *J. Am. Chem. Soc. 33*, 1973-1978.

Johnson, T.B. & Scott, W.M. (1913). Hydantoins: The action of ammonium thiocyanate on lactone-anhydrides of acyl-amino acids. *J. Am. Chem. Soc. 35*, 1136-1143.

Kubo, H., Nakajima, T., & Tamura, Z. (1971). Formation of thiohydantoin derivative of proline from the C-terminal of peptides. *Chem. Pharm. Bull. 19*, 210-211.

Matsuo, H., Fujimoto, Y., & Tatsuno, T. (1966). A novel method for the determination of C-terminal amino acid in polypeptides by selective tritium labelling. *Biochem. Biophys. Res. Commun. 22*, 69-74.

Meuth, J.L., Harris, D.E., Dwulet, F.E., Crowl-Powers, M.L., & Gurd, F.R.N. (1982). Stepwise sequence determination from the carboxyl terminus of the peptide. *Biochemistry 21*, 3750-3757.

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, Inc., New York.

Nicolet, B.H. (1930). Interpretation of the dehydration of acetylglutamic acid by means of glutamylthiohydantoin derivatives. *J. Am. Chem. Soc. 52*, 1192-1195.

Pappin, D.J.C., Coull, J.M., & Koester, H. (1990). Solid phase sequence analysis of proteins electroblotted or spotted onto polyvinylidene fluoride membranes. In *Current Research in Protein Chemistry* (Villafranca, J.J., Ed.), pp. 191-202. Academic Press, Inc., 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.

Rangarajan, M. & Darbre, A. (1976). Studies on sequencing of peptides from the carboxyl terminus by using the thiocyanate method. *Biochem. J. 157*, 307-316.

Schlack, P. & Kumpf, W. (1926). Uber eine neue methode zur Ermittlung der konstitution von peptiden. *Z. Physiol. Chem. 154*, 125-170.

Scoffone, E. & Turco, A. (1956). Indagini sulla struttura delle proteine mediante determinazione dell'ammino acido terminale. *Ric. Sci. 26*, 865-871.

Shenoy, N.R., 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.

Shively, J.E., Miller, P., & Ronk, M. (1987). Microsequence analysis of peptides and proteins VI. A continuous flow reactor for sample concentration and sequence analysis. *Anal. Biochem. 163*, 517-529.

Stark, G.R. (1968). Sequential degradation of peptides from their carboxyl termini with ammonium thiocyanate and acetic anhydride. *Biochemistry 7*, 1796-1807.

Swan, J.M. (1952). Anhydrides of N-acylaspartic acid N-acylglutamic acids. *Nature 169*, 826-828.

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

Williams, M.J. & Kassell, B. (1975). A solid phase method for peptide sequencing from the carboxyl terminal. *FEBS Lett. 54*, 353-357.

Yamashita, S. (1971). Sequential degradation of polypeptides from the carboxyl ends. *Biochim. Biophys. Acta 229*, 301-309.

Yamashita, S. & Ishikawa, N. (1971). Sequential degradation of polypeptides from the carboxyl ends II. Application to polypeptides. *Proc. Hoshi Coll. Pharm. 13*, 136-138.