Automated C-Terminal Protein Sequence Analysis Using the HP G1009A C-Terminal Protein Sequencing System

The HP G1009A is an automated system for the carboxy-terminal amino acid sequence analysis of protein samples. It detects and sequences through any of the twenty common amino acids. This paper describes a number of applications that demonstrate its capabilities.

by Chad G. Miller and Jerome M. Bailey

Peptide and protein molecules are composed of amino acid molecules covalently coupled end-to-end through peptide bonds, resulting in linear sequences of the amino acid residues. These polypeptide molecules have two chemically distinguishable termini or ends, identified as the amino (N) terminus and the carboxy (C) terminus. The N-terminal and C-terminal amino acid residues define the ends of the amino acid sequence of the polypeptide as depicted in Fig. 1.

The amino acid sequence of the polypeptide molecule imparts its chemical properties, such as the 3D spatial conformation, its biological properties including enzymatic function, its biomolecular recognition (immunologic) properties, and its role in cellular shape and architecture. It is the actual amino acid sequence of the polypeptide (the order in which the amino acid residues occur in the polypeptide from terminus to terminus) rather than the amino acid composition as a random linear arrangement that is crucial in defining the structural and functional attributes of peptides and proteins.

C-terminal amino acid sequence analysis of peptide and protein samples provides crucial structural information for bioscientists. Determination of the amino acid sequence constituting the C-termini of proteins is required to define or verify (validate) the full-length structural integrity of native proteins isolated from natural sources or recombinantly expressed protein products that are genetically engineered.

The cellular processing of proteins at the C-terminus is a relatively common event yielding proteins with varying degrees of amino acid truncations, deletions, and substitutions.[1] The processing of the C-terminal end of the protein results in a family of fragmented protein species terminating with different amino acid residues. The N-terminal amino acid sequence, however, may remain identical for all of these protein species. These *ragged C-terminal* proteins often constitute the family of mature protein forms derived from a single gene expression. Many proteins and peptides are initially biosynthesized as inactive large precursors that are enzymatically cleaved and processed to generate the mature, smaller, bioactive forms. This type of *post-translational* processing also serves to mediate the cellular levels of the biologically active forms of peptides and proteins. Information derived solely from DNA analysis does not permit the prediction of these post-translational proteolytic processing events. The identification of the resultant C-terminal amino acids helps elucidate these cellular biosynthetic processes and control mechanisms.

The polypeptide alpha-amino and alpha-carboxylic acid organic chemical functional groups of the N-terminal and C-terminal amino acid residues, respectively, are sufficiently different to require different chemical methods for the

Fig. 1. *Amino acid sequence of a polypeptide depicting the N-terminal end (indicated by the subscript n+1) and the C-terminal end (indicated by the subscript 1) of the molecule, with the intervening amino acid residues indicated by the subscript n. C-terminal sequence analysis begins at the C-terminal amino acid residue and sequentially degrades the polypeptide toward the N-terminal amino acid residue.*
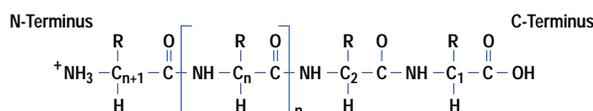
Abbreviations for the Common Amino Acids

| Amino Acid	Three-Letter Code	Single-Letter Code
Alanine	Ala	A
Asparagine	Asn	N
Aspartic Acid	Asp	D
Arginine	Arg	R
Cysteine	Cys	C
Glycine	Gly	G
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tyrosine	Tyr	Y
Tryptophan	Trp	W
Valine	Val	V

determination of the amino acid sequences at these respective ends of the molecule. N-terminal protein sequence analysis has been refined over the course of four decades and is currently an exceptionally efficient automated chemical analysis method. The chemical (nucleophilic) reactivity of the alpha-amino group of the N-terminal amino acid residue permits a facile chemical sequencing process invoking a cyclical degradative scheme, fundamentally based on the reaction scheme first introduced by P. Edman in 1950.[2] Automation of the N-terminal sequencing chemistry in 1967 resulted in a surge of protein amino acid sequence information available for bioscientists.[3]

The much less chemically reactive alpha-carboxylic acid group of the C-terminal amino acid residue proved to be exceedingly more challenging for the development of an efficient chemical sequencing process. A variety of C-terminal chemical sequencing approaches were explored during the period in which N-terminal protein sequencing using the Edman approach was optimized.[4,5] None of the C-terminal chemical reaction schemes described in the literature proved practical for amino acid sequence analysis across a useful range of molecular weight and structural complexity. Carboxypeptidase enzymes were employed to cleave the C-terminal amino acid residues enzymatically from intact proteins or proteolytic peptides. These carboxypeptidase digests were subjected to chromatographic analysis for the identification of the protein C-terminal amino acid residues. This manual process suffered from the inherent enzymatic selectivity of the carboxypeptidases toward the amino acid residue type and exhibited protein sample-to-sample variability and reaction dependencies. The results frequently yielded ambiguous sequence data. The typical results were inconclusive and provided, more often than not, amino acid compositional information rather than unambiguous sequence information. An alternative approach required the proteolytic digestion of a protein sample (typically with the enzyme trypsin), previously chemically labeled (modified) at the protein C-terminal amino acid residue, and the chromatographic fractionation of the resulting peptides to isolate the labeled C-terminal peptide. The isolated peptide was subjected to N-terminal sequence analysis in an attempt to identify the C-terminal amino acid of the peptide. The limited amount and quality of C-terminal amino acid sequence information derived from these considerably tedious, multistep manual methods appeared to apply best to suitable model test peptides and proteins and offered little generality.

HP Thiohydantoin C-Terminal Sequencing Chemistry

The general applicability of a chemical sequencing scheme for the analysis of the protein C-terminus has only very recently been reported and developed into an automated analytical process.[6] The Hewlett-Packard G1009A C-terminal protein sequencing system, introduced in July 1994, automates a C-terminal chemical sequencing process developed by scientists at the Beckman Research Institute of the City of Hope Medical Center in Duarte, California. The overall chemical reaction scheme is depicted in Fig. 2 and consists of two principal reaction events. The alpha-carboxylic acid group of the C-terminal amino acid residue of the protein is modified to a chemical species that differentiates it from any other constituent amino acid residue in the protein sample. The chemical modification involves the chemical *coupling* of the C-terminal amino acid residue with the sequencing coupling reagent. The modified C-terminal amino

Fig. 2. *Overall reaction scheme of the C-terminal sequencing process depicting the initial chemical coupling reaction, modifying the C-terminal amino acid (as indicated by a circle), and the chemical cleavage reaction, generating the C-terminal amino acid thiohydantoin derivative (indicated as a circle) and the shortened peptide.*
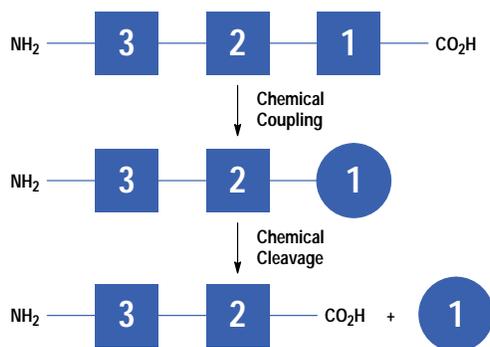
acid residue is in a chemical form that permits its facile chemical *cleavage* from the rest of the protein molecule. The cleavage reaction generates the uniquely modified C-terminal amino acid residue as a thiohydantoin-amino acid (TH-aa) derivative, which is detected and identified using HPLC (high-performance liquid chromatography) analysis. The remaining component of the cleavage reaction is the protein shortened by one amino acid residue (the C-terminal amino acid) and is subjected to the next round of this coupling/cleavage cycle. The overall sequencing process is thus a sequential chemical degradation of the protein yielding successive amino acid residues from the protein C-terminus that are detected and identified by HPLC analysis.

As shown in Fig. 3, the coupling reaction event of the sequencing cycle begins with the activation of the carboxylic acid function of the C-terminal amino acid residue, promoting its reaction with the coupling reagent, diphenyl phosphorylisothiocyanate (DPPITC). The reaction is accelerated in the presence of pyridine and suggests the formation of a peptide phosphoryl anhydride species as a plausible chemical reaction intermediate. The final product of the coupling reaction is the peptide thiohydantoin, formed by chemical cyclization of the intermediate peptide isothiocyanate.

The peptide thiohydantoin, bearing the chemically modified C-terminal amino acid, is cleaved with trimethylsilanolate (a strong nucleophilic base) to release the C-terminal thiohydantoin-amino acid (TH-aa) derivative for subsequent analysis and the shortened peptide for the next cycle of chemical sequencing. The thiohydantoin derivative of the C-terminal amino acid residue is chromatographically detected and identified by HPLC analysis at a UV wavelength of 269 nm.

Data Analysis and Interpretation of Results

The data analysis relies on the interpretation of HPLC chromatographic data for the detection and identification of thiohydantoin-amino acid derivatives. The sequencing system uses the HP ChemStation software for automatic data processing and report generation. By convention, the amino acid sequence is determined from the successive single amino acid residue assignments made for each sequencer cycle of analysis beginning with the C-terminal residue (the first sequencer cycle of analysis). The thiohydantoin-amino acid derivative at any given sequencer cycle is assigned by visually examining the succeeding cycle (n+1) and preceding cycle (n–1) chromatograms with respect to the cycle in question (n). The comparison of peak heights (or widths) across three such cycles typically enables a particular thiohydantoin-amino acid derivative to be detected quantitatively, rising above a background level present in the preceding cycle, maximizing in absorbance in the cycle under examination, and decreasing in absorbance in the succeeding cycle. Exceptions to this most basic algorithm are seen in sequencing cycles in which there are two or more identical amino acid residues in consecutive cycles and for the first sequencer cycle, which has no preceding cycle of analysis.

An HPLC chromatographic reference standard consisting of the twenty common amino acid thiohydantoin synthetic standards defines the chromatographic retention time (elution time) for each particular thiohydantoin-amino acid derivative. The TH-aa derivatives chromatographically detected in a sequencer cycle are identified by comparison of their chromatographic retention times with the retention times of the 20-component TH-aa derivatives of a reference standard mixture. The practical assignment of amino acid residues in sequencing cycles is contingent on an increase in peak absorbance above a background level followed by an absorbance decrease for a chromatographic peak that exhibits the same chromatographic retention time as one of the twenty thiohydantoin-amino acid reference standards. A highly robust chromatographic system is required for this mode of analysis to be reliable and reproducible.

The chromatographic analysis of the online thiohydantoin-amino acid standard mixture is shown in Fig. 4. The peaks are designated by the single-letter codes for each of the 20 common amino acid residues (as shown in the **list** of

Fig. 3. *Detailed reaction scheme of the HP thiohydantoin C-terminal sequencing chemistry. The chemical coupling reactions with diphenyl phosphorylisothiocyanate (DPPITC) generate a mixed anhydride followed by the extrusion of phosphate with ring formation to yield the peptide thiohydantoin. The subsequent chemical cleavage reaction with potassium trimethylsilanolate (KOTMS) releases the C-terminal amino acid thiohydantoin derivative (TH-aa) from the shortened peptide.*
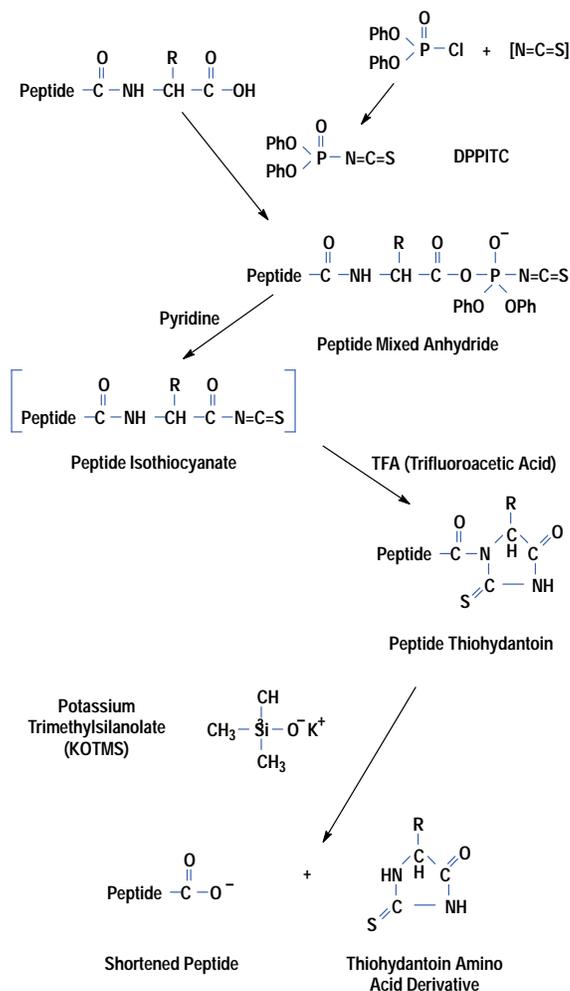
abbreviations for the common amino acids) and correspond to approximately 100 picomoles of each thiohydantoin derivative. The thiohydantoin standard mixture is composed of the sequencing products of each of the amino acids so that Ser and Cys amino acid derivatives each result in the same compound, designated S/C, which is the dehydroalanine thiohydantoin-amino acid derivative. The peaks labeled D′ and E′ signify the methyl esters of the thiohydantoin-amino acid derivatives of Asp and Glu.

The C-terminal protein sequencing system is configured with the HP G1009A protein sequencer which automates the chemical sequencing, the HP 1090M HPLC for the chromatographic detection and identification of the thiohydantoin-amino acids, and an HP Vectra personal computer for instrument control and data processing as shown in Fig. 5. The chemical sequencer consists of assemblies of chemically resistant, electrically actuated diaphragm valves connected through a fluid path of inert tubing that permit the precise delivery of chemical reagents and solvents. The fluid delivery system is based on a timed pressurization of reagent and solvent bottles that directs the fluid flow through delivery valves (including valve manifolds) into the fluid path. The sequencer control software functions on a Microsoft® Windows platform and features an extensive user-interactive graphical interface for instrument control and instrument method editing. The sequencer data analysis software is a modified version of the HP ChemStation software with features specific for the analysis and data reporting of the thiohydantoin-amino acid residues.

Sample Application to Zitex Membranes

The sequencing chemistry occurs on Zitex reaction membranes (inert porous membranes), which are housed in Kel-F (inert perfluorinated plastic) sequencer reaction columns. The Zitex membrane serves as the sequencing support for the coupling and cleavage reactions. The membrane is chemically inert to the sequencing reagents and retains the

Fig. 4. *HPLC chromatogram of the thiohydantoin-amino acid standard mixture (TH-Std). The thiohydantoin-amino acid derivatives of the common amino acids are indicated by their single-letter code designations (see* list*). The peak identified as S/C represents the thiohydantoin derivative of serine and cysteine. The peak designated D' represents the methyl ester sequencing product of aspartic acid and the E' peak represents the methyl ester sequencing product of glutamate. Isoleucine, designated I, elutes as two chromatographically resolved stereoisomeric thiohydantoin derivatives.*
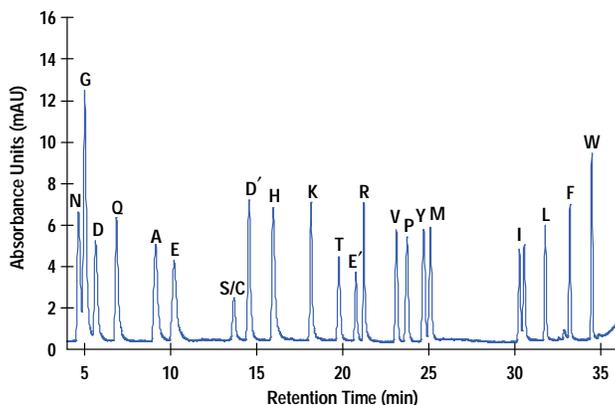
Fig. 5. *The HP G1009A C-terminal protein sequencing system consists of the HP C-terminal sequencer, the HP 1090M HPLC, and an HP Vectra computer.*

sample through multipoint hydrophobic interactions during the sequencing cycle. The chemical methodology performed on the Zitex membrane enables the sequence analysis of proteins and low-molecular-weight peptide samples. The sequencer sample reaction columns are installed in any one of four available sequencing sample positions, which serve as temperature-controlled reaction chambers. In this fashion, four different samples can be independently programmed for automated sequence analysis.

C-terminal sequence analysis is readily accomplished by directly applying the protein or peptide sample to a Zitex reaction membrane as diagrammed in Fig. 6. The process does not require any presequencing attachment or coupling chemistries. The basic, acidic, or hydroxylic side-chain amino acid residues are not necessarily subjected to any presequencing chemical modifications or derivatizations. Consequently, there are no chemically related sequencing ambiguities or chemical inefficiencies. Protein samples that are isolated using routine separation procedures involving various buffer systems, salts, and detergents, as well as samples that are prepared as product formulations, can be directly analyzed using this technique. The chemical method is universal for any of the 20 common amino acid residues and yields thiohydantoin derivatives of serine, cysteine, and threonine—all of which frequently appear in protein C-terminal sequences.

To be successful, the sequence analysis must provide unambiguous and definitive amino acid residue assignments at cycle 1, since all protein forms—whether they result from internal processing, clippings, or single-residue truncations—are available for analysis at that cycle in their relative proportions.

New Sequencing Applications

The Hewlett-Packard G1009A C-terminal protein sequencing system is capable of performing an expanded scope of sequencing applications for the analysis of peptide and protein samples prepared using an array of isolation and

Fig. 6. *Procedure for sample application (loading) onto a Zitex C-terminal sequencing membrane.*

1. Zitex membrane is treated with alcohol.

2. Sample solution is directly applied to membrane and allowed to dry.

3. Membrane is inserted into a sequencer column.
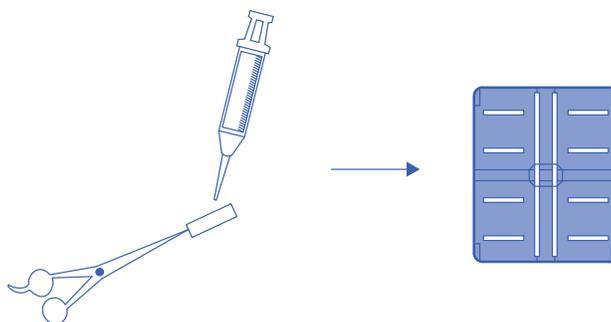
purification techniques. The recently introduced version 2.0 of the HP thiohydantoin sequencing chemistry for the HP G1009A C-terminal sequencer now supports both the "high-sensitivity" sequence analysis of protein samples recovered in 50-to-100-picomole amounts and an increase in the number of sequenceable cycles. The high-sensitivity C-terminal sequence analysis is compatible with a great variety of samples encountered at these amount levels in research laboratories.

C-terminal sequence analysis of protein samples recovered from SDS (sodium dodecylsulfate) gel electrophoresis is an important application enabled by the high-sensitivity sequencing chemistry. SDS gel electrophoresis is a routine analytical technique based on the electrophoretic mobility of proteins in gel matrices. The technique provides information on the degree of overall sample heterogeneity by allowing individual protein species in a sample to be resolved. The sequence analysis is performed on the gel-resolved proteins after the physical transfer of the protein components to an inert membrane, such as Teflon, using an electrophoretic process known as electroblotting. SDS gel electrophoresis of native and expressed proteins frequently exhibits multiple protein bands indicating sample heterogeneity or internal protein processing—both of critical concern for protein characterization and production.

The combined capabilities of high-sensitivity sequence analysis and sequencing from SDS gel electroblots has enabled the development of tandem N-terminal and C-terminal sequence analyses on single samples using the HP G1005A N-terminal sequencer in combination with the HP G1009A C-terminal sequencer. This procedure unequivocally defines the protein N-terminal and C-terminal amino acid residues and the primary structural integrity at both termini of a single sample, thereby eliminating multiple analytical methods and any ambiguities resulting from sample-to-sample variability.

C-Terminal Sequence Analysis Examples

The first five cycles of an automated C-terminal sequence analysis of horse heart apomyoglobin (1 nanomole) are shown in Fig. 7. The unambiguous result observed for cycle 1 confirms the nearly complete homogeneity of the sample, since no significant additional thiohydantoin derivatives can be assigned at that cycle. The analysis continues clearly for five cycles enabling the amino acids to be assigned with high confidence. The results of a sequence analysis of 500 picomoles of recombinant interferon are shown in Fig. 8. The first five cycles show the unambiguous sequencing residue assignments for the C-terminal amino acid sequence of the protein product.

The results of a C-terminal sequence analysis of 1 nanomole, 100 picomoles, and 50 picomoles of bovine beta-lactoglobulin A are shown in Fig. 9 as the cycle-1 HPLC chromatograms obtained for each of the respective sample amounts. The recovery for the first cycle of sequencing is approximately 40% to 50% (based on the amounts applied to the Zitex membranes) and an approximately linear recovery is observed across the range of sample amount. The linear response in the detection of the thiohydantoin-amino acid sequencing products and a sufficiently quiet chemical background permit the high-sensitivity C-terminal sequence analysis.

The automated C-terminal sequencing of the HP G1009A C-terminal sequencer facilitated the detection and identification of a novel C-terminal modification observed for a class of recombinant protein molecules expressed in E. coli that has compelling biological implications. Dr. Richard Simpson and colleagues at the Ludwig Institute for Cancer Research in Melbourne, Australia reported the identification of a population of C-terminal-truncated forms of murine interleukin-6 molecules recombinantly expressed in E. coli that bear a C-terminal peptide extension.[7] Fig. 10 shows the results of the first five cycles of a C-terminal sequence analysis of a purified form of C-terminal-processed interleukin-6 identifying the amino acid sequence of the C-terminal-appended peptide as Ala(1)-Ala(2)-Leu(3)-Ala(4)-Tyr(5).

Fig. 7. *Cycles 1 to 5 of C-terminal sequencing of 1 nanomole of horse apomyoglobin.*
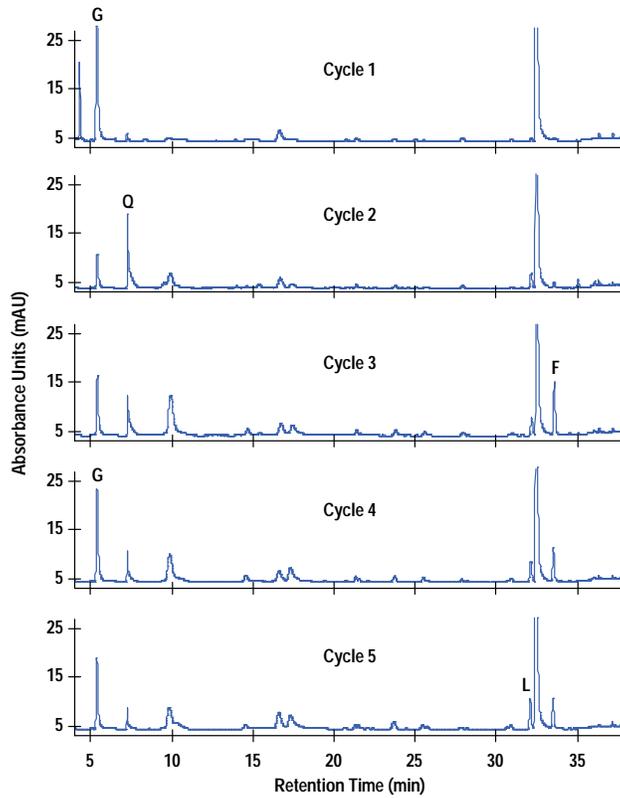

Fig. 8. *Cycles 1 to 5 of C-terminal sequencing of 500 picomoles of recombinant beta-interferon applied to Zitex (230 picomoles of Asn recovered in cycle 1).*
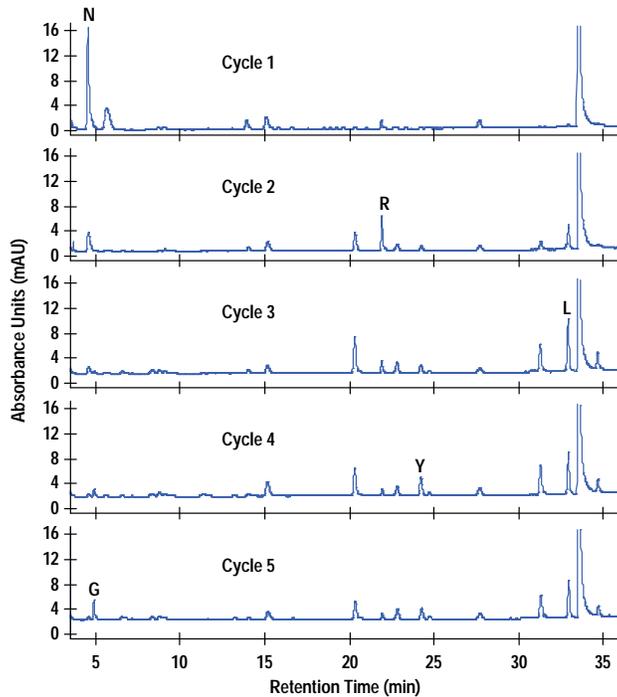
Fig. 9. *C-terminal sequence analysis of bovine beta-lactoglobulin A across a 20-fold range of sample amount.*
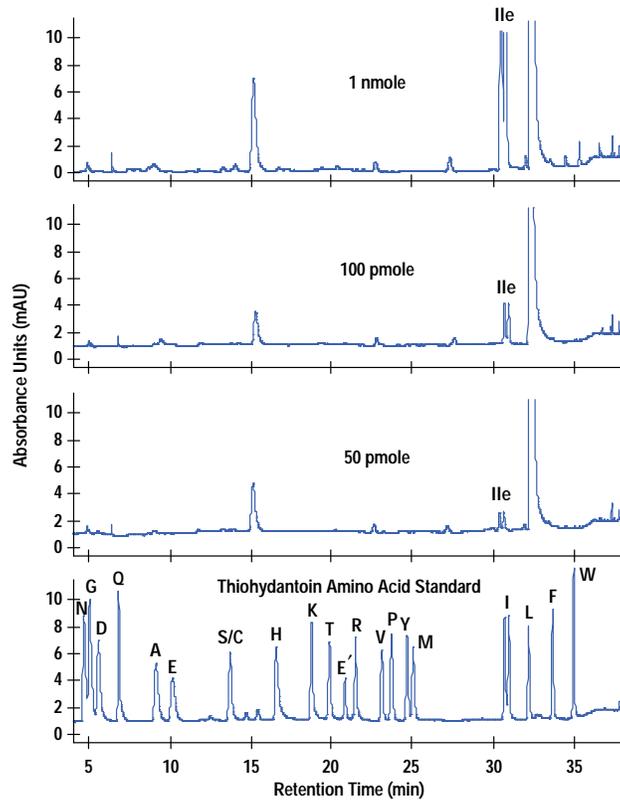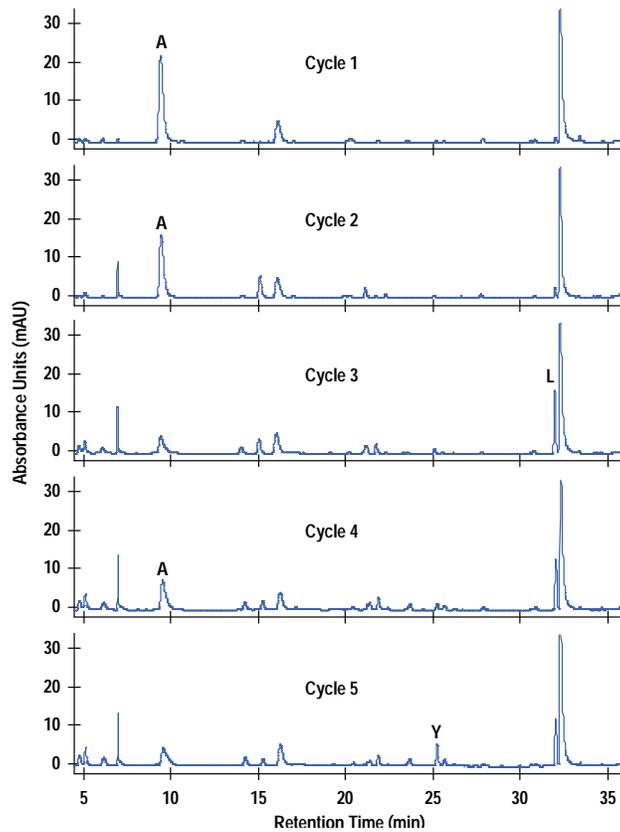

Fig. 10. *Cycles 1 to 5 of C-terminal sequencing of recombinant murine interleukin-6 identifying the C-terminal extension peptide (AALAY) as described in the text. (1 nanomole applied to Zitex: 555 picomoles of Ala recovered in cycle 1.)*

High-Sensitivity C-Terminal Sequence Analysis

Protein samples in amounts of 50 picomoles or more are applied in a single step as liquid aliquots to isopropanol-treated Zitex reaction membranes. The samples disperse over the membrane surface where they are readily absorbed and immobilized. The dry sample membranes do not usually require washing once the sample is applied even in those cases where buffer components and salts are present. As a rule, the matrix components do not interfere with the automated chemistry and are eliminated during the initial steps of the sequencing cycle.

Fig. 11 shows the unambiguous residue assignments for the first three cycles of sequence analysis of a 50-picomole sample of bovine beta-lactoglobulin A applied to a Zitex membrane. Again, the linear response in the detection of the thiohydantoin-amino acid sequencing products coupled with a sufficiently quiet chemical background enable high-sensitivity C-terminal sequence analysis. The sequence analysis of 50 picomoles of human serum albumin is shown in Fig. 12. The chromatograms permit clear residue assignments for the first three cycles of sequencing.

C-Terminal Sequence Analysis of Electroblotted SDS Gel Samples

Analytical gels are routinely used for the analysis of protein samples and recombinant products to assess sample homogeneity and the occurrence of related forms. The common observation of several closely associated protein SDS gel bands is an indication either of cellular processing events or of fragmentations induced by the isolation and purification protocols. Because of its capacity to perform high-sensitivity analysis, C-terminal sequence analysis provides a direct characterization of these various protein species and facilitates the examination of their origin and control.

In collaboration with scientists at Glaxo Wellcome Research Laboratories, Research Triangle Park, North Carolina, SDS gel electroblotting procedures have been developed and applied to protein samples of diverse nature and origin.[8] Typically, 50-picomole (1-to-10-microgram) sample amounts are loaded into individual SDS gel lanes and separated

Fig. 11. *High-sensitivity C-terminal sequencing of 50 picomoles of bovine beta-lactoglobulin A indicating approximately 50% recovery (24 picomoles) of cycle-1 isoleucine.*
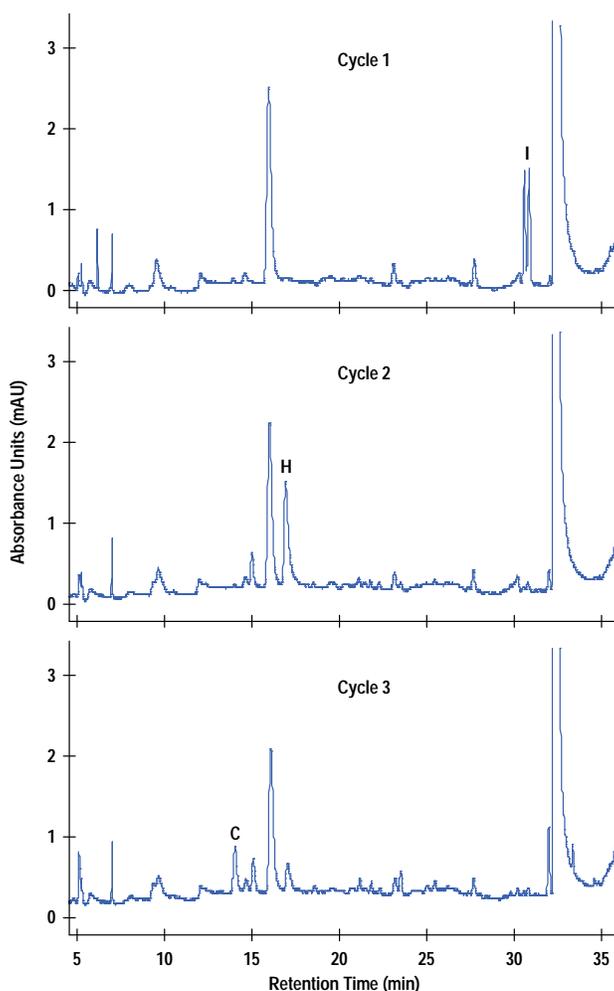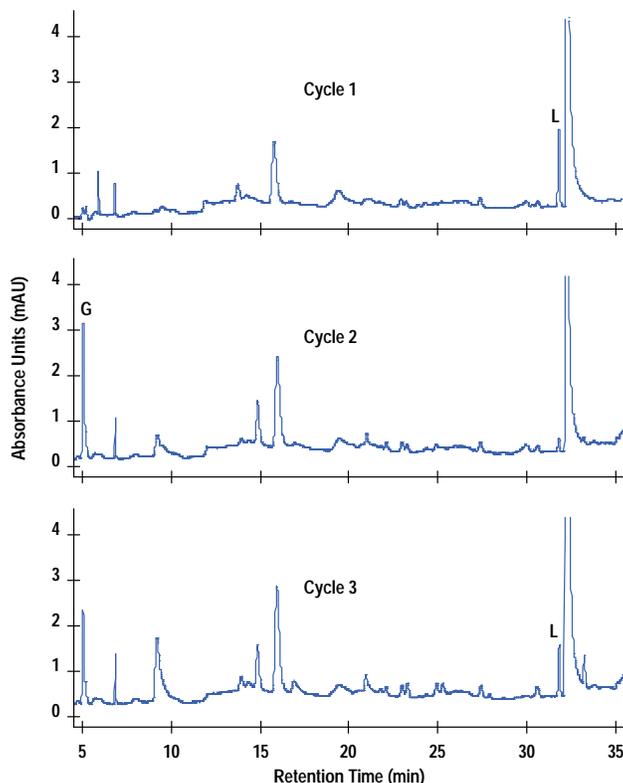
Fig. 12. *High-sensitivity C-terminal sequencing of 50 picomoles of human serum albumin indicating approximately 56% recovery (28 picomoles) of cycle-1 leucine.*

according to molecular weight by an applied voltage. Following electrophoretic resolution of the protein samples, the protein bands on the SDS gel are electroblotted to a Teflon membrane. The electroblotted proteins, visualized on the Teflon membrane after dye-staining procedures, are excised in the proper dimensions for insertion directly into C-terminal sequencer columns as either single bands (1 cm) or multiple bands. The visualization stain does not interfere with the sequence analysis and the excised bands do not have to be treated with any extensive wash protocol before sequencing.

Fig. 13 shows the results of sequence analysis of approximately 250 picomoles of a bovine serum albumin (68 kDa†) sample loaded into one lane of the SDS gel, subjected to electrophoresis, and electroblotted to a Teflon membrane. The chromatograms for cycles 1 to 3 of the single 1-cm protein band sequenced from the Teflon electroblot illustrate the high sensitivity of the sequencing technique. In Fig. 14, approximately 50 picomoles of a phosphodiesterase (43 kDa) sample were applied to an SDS gel lane, subjected to electrophoresis, electroblotted to Teflon, and excised as a single 1-cm sulforhodamine-B-stained band. C-terminal sequence analysis of the Teflon blot enabled the determination of the extent of C-terminal processing as demonstrated by the presence in cycle 1 of the Ser expected from the full-length form, in addition to the His residue arising from a protein form exhibiting the truncation of the C-terminal amino acid, in about a 10:1 ratio. The expected full-length C-terminal sequence continues in cycle 2 (His) and cycle 3 (Gly), with the truncated form yielding Gly in cycle 2 and Glu in cycle 3. Additional processed protein forms also are identifiable.

This application was exploited by researchers at the Department of Biological Chemistry of the University of California at Irvine to investigate the C-terminal amino acid sequences of polio virus capsid proteins and examine additional viral systems.[9] Polio viral capsid proteins VP2 and VP3 were electroblotted to Teflon tape from SDS gels for C-terminal sequence analysis. Figs. 15 and 16 show the results of the sequence analysis, which unambiguously identifies the C-terminal residue in both proteins as Gln, with additional amino acid residue assignments being Gln(1)-Leu(2)-Arg(3) for VP2 and Gln(1)-Ala(2)-Leu(3) for VP3 subunits, respectively.

† kDa = kilodaltons. A dalton is a unit of mass measurement approximately equal to 1.66×10^{-24} gram.

Fig. 13. *C-terminal sequence analysis of 250 picomoles of bovine serum albumin applied to an SDS gel, electroblotted to Teflon tape, and subjected to three cycles of analysis. 70 picomoles of Ala recovered in cycle 1.*
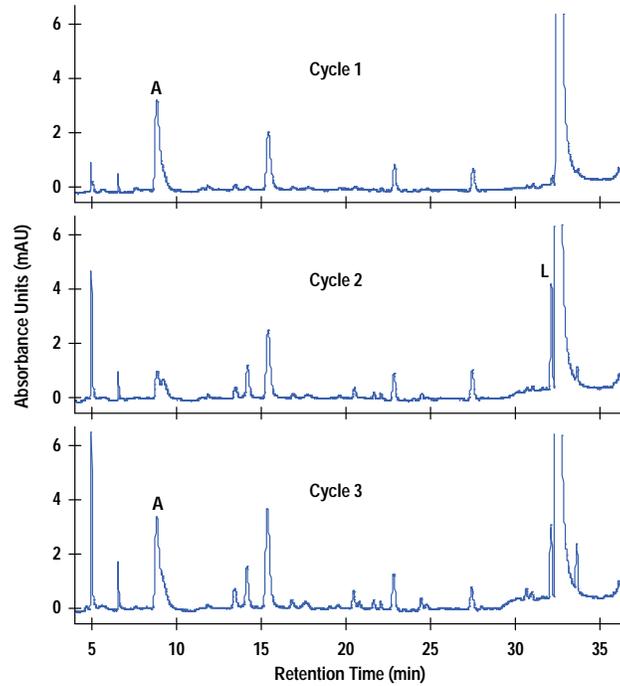

Fig. 14. *Cycles 1 to 3 of a C-terminal sequence analysis of 50 picomoles of phosphodiesterase electroblotted from an SDS gel to Teflon tape. Three protein species are identified at cycle 1 of the sequence analysis indicating protein C-terminal processing of the sample.*
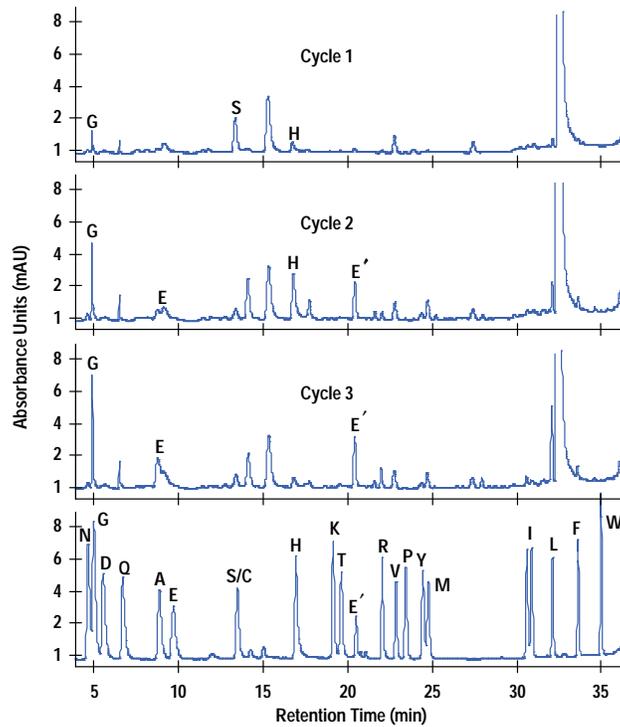
Fig. 15. *Cycles 1 to 3 of a C-terminal sequence analysis of polio viral capsid protein, VP2, electroblotted from an SDS gel to Teflon tape.*
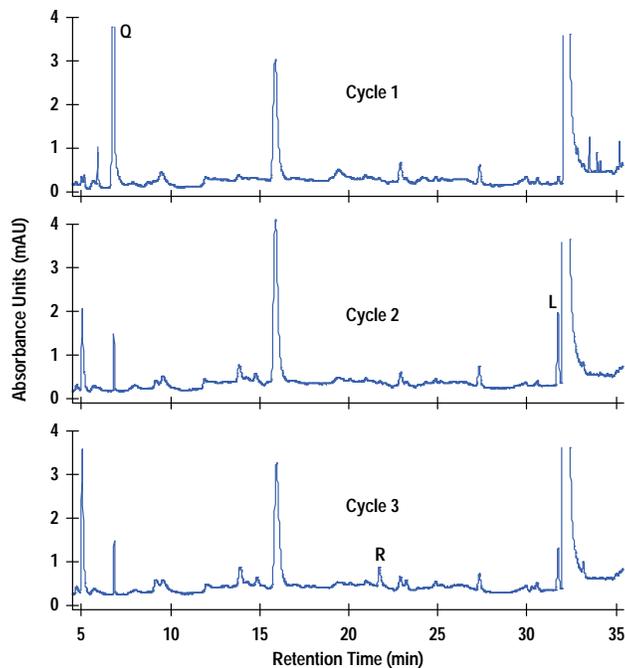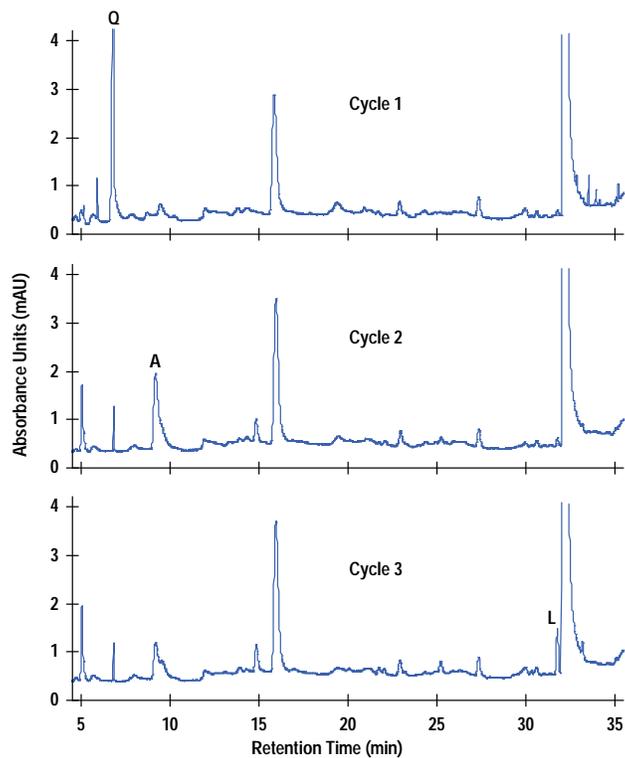

Fig. 16. *Cycles 1 to 3 of a C-terminal sequence analysis of polio viral capsid protein, VP3, electroblotted from an SDS gel to Teflon tape.*

Tandem N-Terminal and C-Terminal Sequence Analysis

The amino acid sequence analysis of both the N-terminus and the C-terminus of single protein samples enables amino acid sequence identification at the respective terminal ends of the protein molecule in addition to a rapid and reliable detection of ragged protein termini. The tandem sequence analysis process of N-terminal protein sequencing followed by C-terminal protein sequencing of the same single sample provides a combined, unequivocal technique for the determination of the structural integrity of expressed proteins and biologically processed precursors. The analyses are performed on a single sample of protein, thereby eliminating any ambiguities that might be attributed to sample-to-sample variability and sample preparation.

The protein sample is either applied as a liquid volume onto a Zitex membrane or electroblotted to Teflon from an SDS gel and inserted into a membrane-compatible N-terminal sequencer column. The sample is subjected to automated N-terminal sequencing using the HP G1005A N-terminal protein sequencer. The sample membrane is subsequently transferred to a C-terminal sequencer column and subjected to automated C-terminal sequence analysis using the HP G1009A C-terminal protein sequencer. In this fashion, a single sample is analyzed by both sequencing protocols, resulting in structural information that pertains precisely to a given population of proteins.[9]

Figs. 17a and 17b demonstrate the tandem sequencing protocol on an SDS gel sample of horse myoglobin. About 250 picomoles of myoglobin were applied to an SDS minigel across five lanes, subjected to electrophoresis, and electroblotted to a Teflon membrane. Five sulforhodamine-B-stained bands were excised and inserted into the N-terminal sequencer column. Five cycles of N-terminal sequencing identified the sequence Gly(1)-Leu(2)-Ser(3)-Asp(4)-Gly(5) with 47 picomoles of Gly recovered at cycle 1 as shown in Fig. 17a. The sequenced Teflon bands were transferred to a C-terminal sequencer column and the results of Fig. 17b show the expected C-terminal sequence Gly(1)-Gln(2)-Phe(3) with 79 picomoles of Gly recovered at cycle 1. The HP tandem sequencing protocol is currently being employed to ascertain the primary structure and sample homogeneity of pharmaceutical and biotechnology protein products in addition to protein systems under active study in academic research laboratories.

Fig. 17. *(a) Cycles 1 to 5 of an N-terminal sequence analysis of 250 picomoles of horse apomyoglobin electroblotted from an SDS gel to Teflon tape. The HP G1005A N-terminal sequencer was used to sequence the sample applied to the Zitex membrane. 47 picomoles of Gly recovered at cycle 1. (b) Cycles 1 to 3 of the C-terminal sequence analysis of the N-terminal-sequenced myoglobin sample of Fig. 17a subjected to C-terminal sequencing using the HP G1009A C-terminal sequencer. The sample membrane was directly transferred from the HP G1005A N-terminal sequencer to the HP G1009A C-terminal sequencer. 79 picomoles of Gly recovered at cycle 1.*
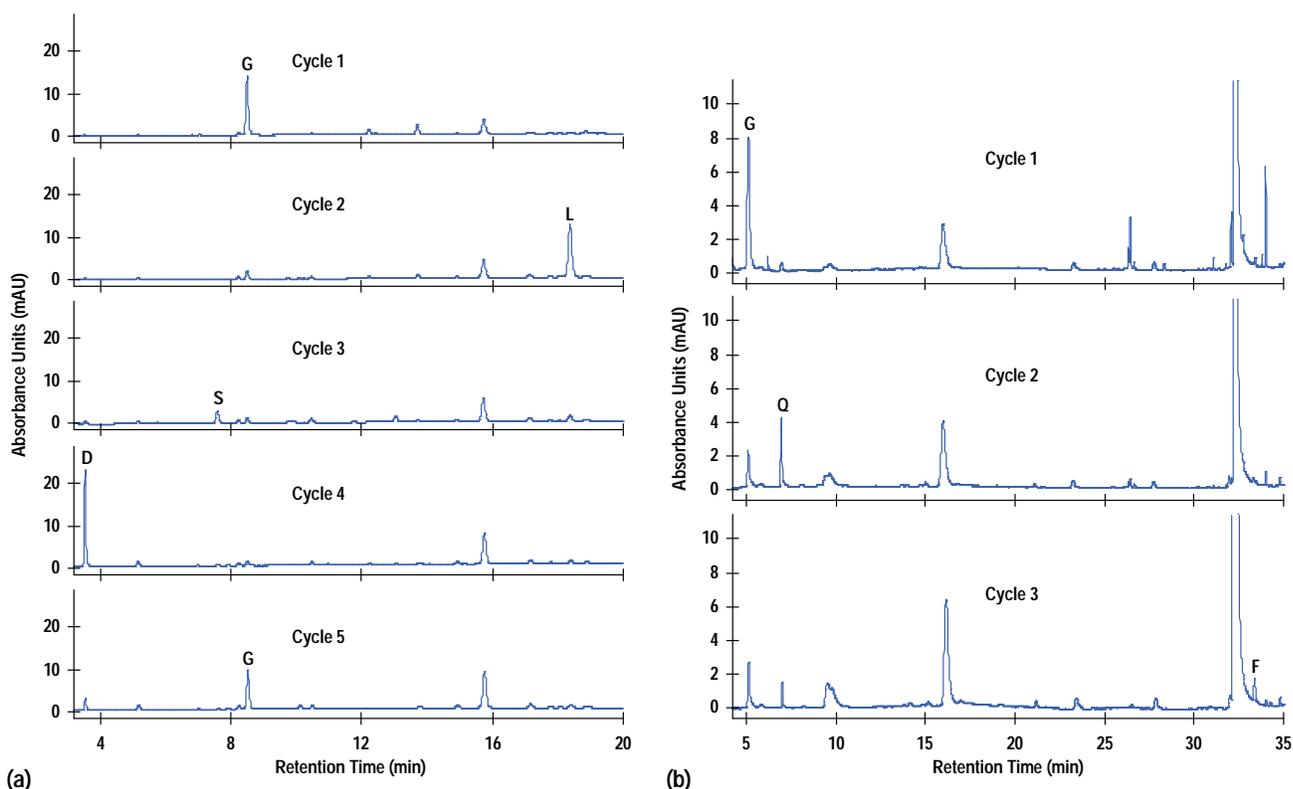
Summary

The HP G1009A C-terminal protein sequencer generates peptide and protein C-terminal amino acid sequence information on a wide variety of sample types and sample preparation techniques, including low-level sample amounts (<100 picomoles), HPLC fractions, SDS gel electroblotted samples, and samples dissolved in various buffer solutions and formulations. The automated sequence analysis provides unambiguous and definitive amino acid residue assignments for the first cycle of sequence analysis and enables additional residue assignments for typically three to five cycles. The ability to sequence through any of the common amino acid residues provides researchers with a critical component for reliable sequence identification. The tandem process of N-terminal and C-terminal sequence analysis of single protein samples provides a significant solution for the evaluation of sample homogeneity and protein processing. The utility of these current applications is being applied to the investigation of many protein systems of importance in biochemistry and in the development and characterization of protein therapeutics.

Acknowledgments

References

1. R. Harris, "Processing of C-terminal lysine and arginine residues of proteins isolated from mammalian cell culture," *Journal of Chromatography A*, Vol. 705, 1995, pp. 129-134.

2. P. Edman, "Method for determination of the amino acid sequence in peptides," *Acta Chemica Scandinavica*, Vol. 4, 1950, pp. 283-293.

3. P. Edman and G. Begg, "A protein sequenator," *European Journal of Biochemistry*, Vol. 1, 1967, pp. 80-91.

4. A.S. Inglis, "Chemical procedures for C-terminal sequencing of peptides and proteins," *Analytical Biochemistry*, Vol. 195, 1991, pp. 183-196.

5. J.M. Bailey, "Chemical methods of protein sequence analysis," *Journal of Chromatography A*, Vol. 705, 1995, pp. 47-65.

6. C.G. Miller and J.M. Bailey, "Automated C-terminal analysis for the determination of protein sequences," *Genetic Engineering News*, Vol. 14, September 1994.

7. G-F Tu, G.E. Reid, J-G Zhang, R.L. Moritz, and R.J. Simpson, "C-terminal extension of truncated recombinant proteins in Escherichia coli with a 10Sa RNA decapeptide," *Journal of Biological Chemistry*, Vol. 270, 1995, pp. 9322-9326.

8. W. Burkhart, M. Moyer, J.M. Bailey, and C.G. Miller, "C-terminal sequence analysis of proteins electroblotted to Teflon tape and membranes," Poster 529M, *Ninth Symposium of the Protein Society*, Boston, Massachusetts, 1995.

9. C.G. Miller and J.M. Bailey, "Expanding the scope of C-terminal sequencing," *American Biotechnology Laboratory*, October 1995.