Methods in Enzymology
Volume 289 (B)

Methods in Enzymology

Volume 289

Solid-Phase Peptide Synthesis

EDITED BY

Gregg B. Fields

DEPARTMENT OF LABORATORY MEDICINE AND PATHOLOGY
BIOMEDICAL ENGINEERING CENTER
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA

ACADEMIC PRESS
San Diego London Boston New York Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

Academic Press
15 East 26th Street, 15th Floor, New York, New York 10010, USA
http://www.apnet.com

Academic Press Limited
24-28 Oval Road, London NW1 7DX, UK
http://www.hbuk.co.uk/ap/

International Standard Book Number: 0-12-182190-0

PRINTED IN THE UNITED STATES OF AMERICA
99 00 01 02 MM 9 8 7 6 5 4 3 2

Methods in Enzymology

Volume 289
SOLID-PHASE PEPTIDE SYNTHESIS

METHODS IN ENZYMOLOGY

EDITORS-IN-CHIEF

John N. Abelson Melvin I. Simon

DIVISION OF BIOLOGY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

FOUNDING EDITORS

Sidney P. Colowick and Nathan O. Kaplan

Section II

Analytical Techniques

[17] Edman Sequencing as Tool for Characterization of Synthetic Peptides

By GREGORY A. GRANT, MARK W. CRANKSHAW, and JOHN GORKA

Introduction

Sequence analysis by Edman chemistry is a useful tool in the characterization of synthetic peptides. As the name implies, its purpose is to determine the sequence of amino acid residues in a peptide. Thus, it is particularly useful for the determination of residue deletions in the sequence of a synthetic peptide. It is also useful for the determination of the location of stable modified amino acids in a sequence and to determine the identity of amino acid residues with equal or similar mass. As with any technique, it has its limitations, and thus a familiarity with these is necessary to avoid misleading or incorrect conclusions. Because peptides are most commonly synthesized on solid supports, most analytical methods can be applied only after the peptide is cleaved from the resin. Edman sequence analysis, however, can be performed either before or after cleavage. It can, in fact, be used to monitor the progress of the synthesis as it is occurring by analyzing a very small amount of resin removed from the reaction vessel of the synthesizer while the synthesis itself continues on.

Overview of Edman Sequencing

Edman Chemistry

Repetitive degradation of proteins and peptides from their amino terminus by reaction of the free α-amino group with phenyl isothiocyanate (PITC) (Fig. 1) was first reported by Edman in 1950[1] and automated by Edman and Begg in 1967.[2] Coupling of PITC to the peptide amino groups takes place in the presence of base at pH 9–10. A primary or secondary amino group is required for coupling to take place, so most modifications of the amino terminus, such as by acylation, will render the peptide refractory to the Edman chemistry. Subsequent to coupling, cleavage of the first amino acid from the peptide is produced by treatment with an anhydrous acid, usually trifluoroacetic acid, which promotes the attack of the PITC sulfur on the carbonyl carbon of the amino acid. This reaction is particularly

[1] P. Edman, *Acta Chem. Scand.* **4**, 283 (1950).
[2] P. Edman and G. Begg, *Eur. J. Biochem.* **1**, 80 (1967).

0076-6879/97 $25.00

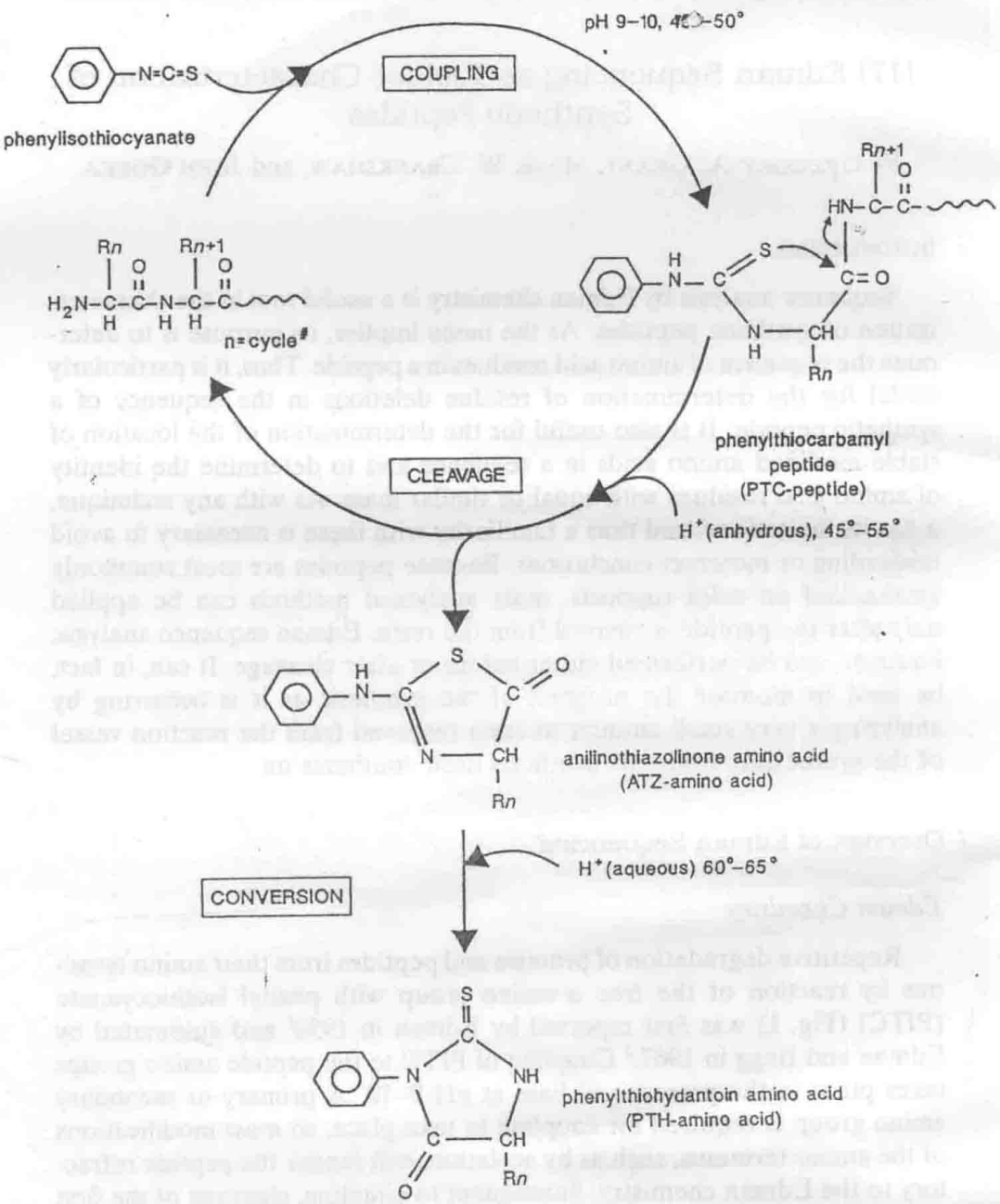

FIG. 1. Edman chemistry for the sequential degradation of proteins and peptides. The reaction is carried out in successive cycles with the product of each cycle being a PTH-amino acid. The PTH-amino acid identified at each cycle corresponds to the position from the amino terminus that the amino acid occupied in the peptide.

dependent on the appropriate chain length leading to a five-membered ring intermediate and thus proceeds efficiently only for α-amino acids. The attack on the carbonyl carbon and formation of the anilinothiazolinone derivative lead to peptide bond cleavage and the generation of a new amino terminus on the peptide that can subsequently react with PITC to repeat the process. Because excess PITC is removed before treatment with acid, only one amino acid is cleaved at a time. This provides the basis for sequence determination through repetitive reaction with PITC. Conversion of the cleaved anilinothiazolinone derivative to the more stable phenylthiohydantoin derivative takes place in aqueous acid, usually 25% (v/v) trifluoroacetic acid. During this conversion, the anilinothiazolinone ring opens to produce a phenylthiocarbamyl (PTC) derivative and then closes to a rearranged ring structure called the phenylthiohydantoin (PTH)-amino acid. If the conversion conditions are not optimized, some PTC-amino acid may persist and will subsequently be seen on high-performance liquid chromatography (HPLC). It is the generation of one PTH-amino acid per peptide chain per reaction cycle that, when subsequently identified, yields the sequence of the peptide.

Both acid treatments of the Edman chemistry occur under conditions that are more extreme than those used to remove many of the protecting groups used in 9-fluorenylmethyloxycarbonyl (Fmoc) synthesis. Thus, these groups will be removed to varying degrees during sequencing if they are present (see Table III and the section on the Analysis of Resin-Bound Peptides).

Identification of Phenylthiohydantoin-Amino Acids

Identification of the released amino acids from repetitive Edman degradation has been a major undertaking since development of the chemistry. The introduction of HPLC in the late 1970s revolutionized this aspect by providing a technique that was relatively rapid and sensitive and that could identify all of the commonly encountered amino acids in a single analysis.[3] As of this writing, all commercially available protein/peptide sequencers use reversed-phase chromatographic methods for the identification of the PTH-amino acids released during sequencing. Figure 2 shows a typical chromatogram of the common PTH-amino acids from one such instrument.

Analysis by HPLC, however, does not unequivocally identify the cleaved amino acid residue. Rather, it relies on the comparison of the elution position of the PTH-amino acid generated during sequencing from a reversed-phase HPLC column to the elution position of standard PTH-amino

[3] M. W. Hunkapiller and L. E. Hood, *Methods Enzymol.* **91**, 486 (1983).

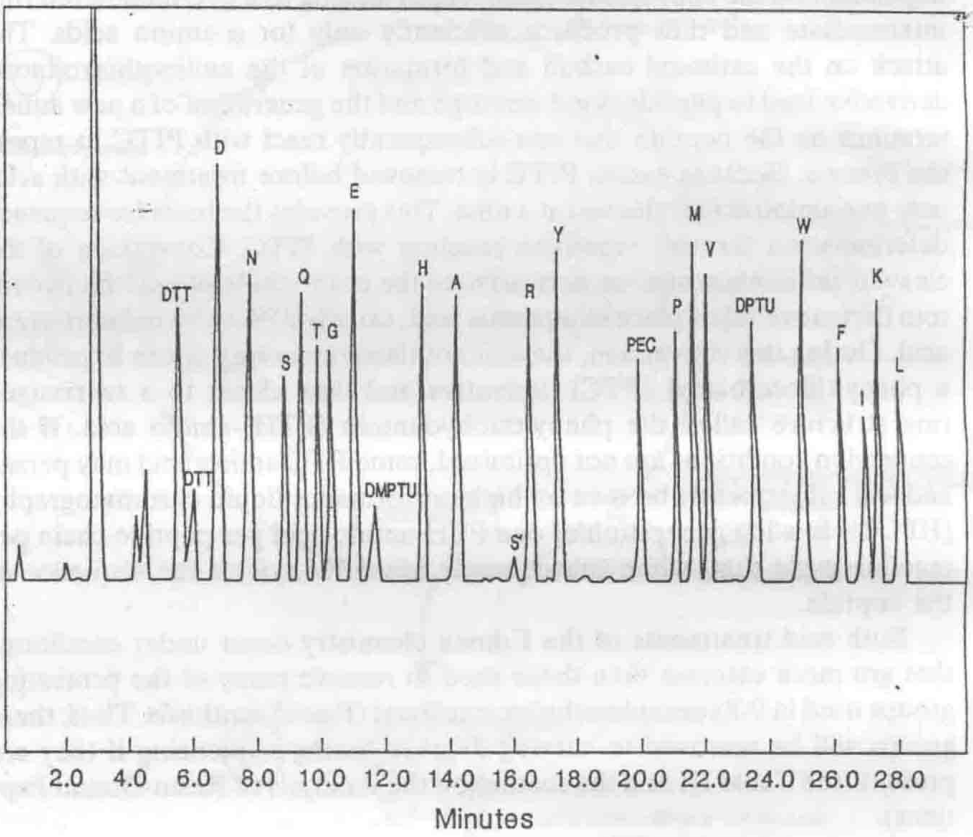

FIG. 2. A reversed-phase HPLC chromatogram of a standard mixture of PTH-amino acids from a Perkin-Elmer Applied Biosystems Division Model 477 protein sequencer. The abbreviations used are the one-letter code for the amino acids. Also DTT, dithiothreitol; DMPTU, dimethylphenylthiourea; PEC, pyridylethyl cysteine; and DPTU, diphenylthiourea. The DTT doublet is reduced and oxidized DTT, and the prime (S′) indicates a DTT derivative of a Ser degradation product. Note that the elution times vary somewhat from those presented in Table III. This is normal, and Table III should be used only as a guide of relative position. The large elution differences for His and Arg seen here and as listed in Table III represent normal variability for these residues arising from differences in elution buffer composition from one sequencer to another.

acids run on the same column under the same conditions. Thus, although it is critical that appropriate standards be available, this does not usually pose a problem for synthetic peptides because their putative sequence is known and standards of stable side-chain protected PTH-amino acids can be easily formed using the Edman chemistry and the appropriate α-amino deprotected amino acid derivative. These standards can be produced manu-

ally[4,5] or by simply subjecting the amino acid derivative to a single cycle of Edman chemistry on the sequencer.[6] Because modern sequencers can sequence sample loads in the low picomole to femtomole range, quantity is seldom a problem for synthetic peptides. However, it is important not to overload the sequencer, particularly for samples containing side-chain protected amino acids, because high levels of PTH-amino acids that may have low solubility in the sequencer solvents can precipitate and occlude the lines of the sequencer or the in-line HPLC. Given the fact that a typical bead from solid-phase peptide synthesis may produce an initial yield in the sequencer of 50–150 pmol of PTH-amino acid,[6] it is very easy to overload the sequencer.

Peptide Washout and Quantitation

Unless a peptide is covalently attached to a support during Edman sequencing, it may wash out of the reaction vessel before the end is reached. This is mainly due to the repeated organic extractions that the peptide is subjected to as the sequencing proceeds. In general, the level of washout tends to increase with the hydrophobicity of the peptide and decrease with increasing peptide length. In addition, as the peptide is shortened during sequencing, the level of washout may increase. Nonetheless, except for long peptides (>15–20 residues), it is usually possible to identify the penultimate residue and often the C-terminal residue. However, this is not guaranteed, and even with peptides that are still attached to the synthetic resin it may be difficult to identify the C-terminal residue. Thus, if a modification or deletion has occurred at the C-terminal residue for relatively short peptides, or close to the C terminus in the case of longer peptides, Edman degradation in itself may not produce usable data in the C-terminal region. Furthermore, Edman degradation does not produce highly quantitative results because not all amino acids are equally stable to the chemistry or extracted from the reaction vessel with the same efficiency.[7] In addition, the initial yield of the sequencing is unpredictable, and the repetitive yield of even the same amino acid at different points in the sequence can vary to some degree.

The common amino acid residues that are usually recovered without appreciable degradation or loss as a consequence of the Edman chemistry

[4] D. H. Schlesinger, *Methods Enzymol.* **91,** 494 (1983).
[5] D. M. Steiman, R. J. Ridge, and G. R. Matsueda, *Anal. Biochem.* **145,** 91 (1985).
[6] J. Pohl, *in* "Methods in Molecular Biology, Peptide Analysis Protocols" (B. M. Dunn and M. W. Pennington, eds.), p. 107. Humana Press, Totowa, New Jersey, 1994.
[7] M. W. Hunkapiller, *in* "Proteins, Structure and Function" (J. J. L'Italien, ed.), p. 363. Plenum, New York and London, 1987.

are Asp, Glu, Gly, Ala, Tyr, Pro, Val, Ile, Leu, Phe, and Lys. Thr, Ser, and Cys are subject to degradation by β-elimination. Although low levels of undegraded Thr and Ser can usually be seen, Cys usually is not seen at all unless it is derivatized to a stable adduct prior to sequencing. Some of the β-elimination products can be identified as relatively stable adducts of dithiothreitol (DTT) by including it in the appropriate sequencer solvents. Note that because the elimination product of Ser and Cys are the same, they cannot be distinguished on this basis. Met and Trp are particularly sensitive to oxidation. Inclusion of DTT in the sequencer solvents will convert any Met sulfoxide formed back to Met so that its yield is usually quite high. Trp can be converted to a number of oxidation products that are variably stable and ill-defined such that its identification can be difficult. Asn and Gln usually undergo a small degree of deamidation (see Table I, cycle 9) which does not interfere with their identification if this is recognized. His and Arg, whose PTH derivatives still retain charge character, are more difficult to extract from the reaction vessel because of this. Because there can be a significant variation in residue yield, Edman sequencing can be very poor at determining the presence of modifications or deletions that occur at low levels and particularly if they are heterogeneously dispersed throughout the peptide. This problem is usually compounded by the fact that the Edman chromatograms almost always show low to moderate levels of background and carryover from previous cycles.

Background and Carryover

Background amino acids are introduced into Edman sequencing either through amino acid contamination from the laboratory environment or more commonly as a consequence of the sequencing chemistry itself. Amino acid contamination from the environment manifests itself as a high level of background at the beginning of a sequence run that gradually diminishes or washes out. In addition, the internal peptide bonds of a peptide will undergo some low level of cleavage at each cycle producing free amino termini that can couple with PITC, resulting in a gradually increasing level of background signal. Depending on the sequence of the peptide, this has been estimated to normally be as much as 0.1% at each cycle.[8] Asp–Pro bonds, and to a lesser extent Asp–X bonds, are more acid labile than most other amino acid amide linkages and may produce higher levels of premature bond cleavage during sequencing, leading to higher background levels.

[8] W. F. Brandt, A. Henschen, and C. von Holt, *in* "Methods in Protein Sequence Analysis" (M. Elzinga, ed.), p. 101. Humana Press, Clifton, New Jersey, 1982.

TABLE I
SEQUENCE ANALYSIS OF PEPTIDE MIXTURE CONTAINING DELETION

Cycle	Sequence	Yield (pmol)[c]								Calculated % preview	
		Met	Asp	Arg	Asn	Glu	Ile	Pro	Gln	Eq. (1)	Eq. (2)
1	Met	**2338**	117	4	24	36	8	25	14	5.7	—[a]
2	Asp	303	**1948**	13	34	47	50	32	11	—	—
3	Asp	16	**1941**	38	62	44	69	60	10	3.0	2.1
4	Arg	4	417	**1240**	***270***	60	61	65	32	22.6	18.7
5	Asn		168	530	**926**	***191***	73	58	37	25.1	21.8
6	Glu		61	81	246	**569**	***240***	58	32	29.9	25.2
7	Ile		26	11	29	118	**564**	***127***	29	27.9	22.4
8	Pro		25	8	10	36	147	**329**	***92***	25.3	20.4
9	Gln		***63***	4	6	51	32	111	**271**	25.5	—[b]
10	Asp		**184**	4	5	19	12	43	87		

[a] Cycle $j + 1$ (Asp-3) is not included because Asp is also the expected residue.
[b] No carryover value; cycle 11 not analyzed.
[c] Numbers that are bold and underlined are yields of amino acids expected at that cycle. Numbers that are bold and italic are the preview yields at that cycle.

Carryover or lag results from the fact that some PITC-coupled peptide will be left behind at each cycle because the chemistry of removal of the amino-terminal residue is never 100% complete.[9] When cleaved in subsequent cycles, it produces a signal that is out of phase or lags behind the major signal. This is usually most pronounced in the cycle immediately following the occurrence of the incomplete cleavage, but it can persist at some level throughout the rest of the sequence. Carryover is often seen most dramatically following proline residues because cleavage of the imino acid is relatively inefficient. Examples of carryover can be seen for most residues in the sequence presented in Table I.

Edman Analysis of Synthetic Peptides

Edman sequencing can be used to analyze both cleaved peptides[10–12] and peptides that have not yet been cleaved from the synthesis resin.[4–6,12–16] Although analysis of both are similar, on-resin sequencing presents some additional features that are discussed separately. In both cases, the major utility of the procedure is the detection of deletion sequences that have resulted from incomplete couplings or incomplete deprotection of the α-amino group. For unpurified cleavage products and for on-resin sequencing, the process has come to be known as preview analysis[13,14] because, when a mixture is present, residues expected at one cycle will show up in the previous cycle owing to the presence of some percentage of the deletion sequence in the mixture (see Table I).

Analysis of Cleaved Peptides

Sequence analysis of cleaved peptides is done in exactly the same manner as sequence analysis of any protein or peptide regardless of origin. For

[9] M. W. Hunkapiller, *in* "Methods in Protein Sequence Analysis" (K. A. Walsh, ed.), p. 367. Humana Press, Clifton, New Jersey, 1982.

[10] M. J. Geisow and A. Aitken, *in* "Protein Sequencing, A Practical Approach" (J. B. C. Findlay and M. J. Geisow, eds.), p. 85. IRL Press, Oxford, 1989.

[11] G. A. Grant, *in* "Synthetic Peptides: A User's Guide" (G. A. Grant, ed.), p. 185. Freeman, New York, 1992.

[12] G. A. Grant and M. W. Crankshaw, *in* "Methods in Molecular Biology, Protein Sequencing Protocols" (B. J. Smith, ed.), pp. 197–215. Humana Press, Totowa, New Jersey, 1996.

[13] G. R. Matsueda, E. Haber, and M. N. Margolies, *Biochemistry* **20**, 2571 (1981).

[14] S. B. H. Kent, M. Riemen, M. LeDoux, and R. B. Merrifield, *in* "Methods in Protein Sequence Analysis" (M. Elzinga, ed.), p. 205. Humana Press, Clifton, New Jersey, 1982.

[15] D. J. McCormick, B. J. Madden, and R. J. Ryan, *in* "Proteins, Structure and Function" (J. J. L'Italien, ed.), p. 403. Plenum, New York and London, 1987.

[16] C. G. Fields, V. L. VanDrisse, and G. B. Fields, *Pept. Res.* **6**, 39 (1993).

that reason, the specific methods for doing so are not repeated here because they are provided by the instrument manufacturer as standard procedure and are routinely available in all resource laboratories offering such services. Because cleaved peptides are not expected to retain side-chain protecting groups under normal circumstances, analysis for these are not usually a concern at this stage. However, for the occasional reticent protecting group, the presence of which is usually first suggested by mass spectrometric (MS) analysis, the same procedure as for on-resin sequencing, discussed below, can be used.

In practice, mainly due to time and expense considerations, purification of the cleaved synthetic peptide mixture is often done first. If, as a result of purification, a major homogeneous product is obtained that has the correct expected mass, there is often nothing to be gained by sequencing the peptide. Even if more than one peptide is obtained in good yield, the correct peptide can usually be identified by mass analysis. However, unless MS/MS capabilities are available and appropriate daughter ions can be formed, the location of the deleted residue in the sequence of the peptide may not be clear if the same residue appears more than once in the sequence or if it involves the presence of equal mass ions such as Ile and Leu or Gln and Lys. If the location of the deletion is in question, perhaps because the synthesis proves to be particularly problematic and attempts need to be made to increase the synthetic efficiency at the appropriate cycle, sequence analysis can provide useful information.

If a homogeneous deletion peptide is sequenced, the deletion will simply show up as an absence of the expected residue. If the crude cleavage product is sequenced before any purification is attempted, the results will more fully reflect what has occurred in the synthesis, and all areas involving deleted residues will show up as a preview of the next residue. Thus, multiple occurrences can be monitored in a single sequence analysis.

Analysis of Resin-Bound Peptides

Sequencing of resin-bound peptides can be approached in much the same manner as cleaved peptides. The major difference encountered with resin-bound peptides is the presence of protecting groups still on the amino acid side chains. Because these protected amino acids are usually more hydrophobic in nature than their deprotected counterparts, they will tend to elute later on reversed-phase HPLC. Hence, the gradient used to analyze the PTH-amino acids must be extended. Similarly, appropriate cycles for the reaction cartridge and conversion flask must be used that provide for appropriate cleavage and extraction parameters for resin-bound peptides.[17]

[17] Applied Biosystems User Bulletin No. 13, pp. 1–18. (1985).

The specific changes to be made will differ depending on the type of sequencer used but generally follow the same pattern of extending cleavage and extraction time and lengthening the HPLC gradient. An example of this for a Perkin-Elmer Applied Biosystems (Foster City, CA) Model 477 sequencer is summarized in Table II. Although the HPLC gradient programs presented in Table II do not produce exactly the same gradient at the beginning of the run where the native PTH-amino acids will elute, the elution profiles are very similar for both programs.

The side-chain protecting groups usually encountered in *tert*-butyloxycarbonyl (Boc) synthesis are designed to be stable to trifluoroacetic acid (TFA) treatment whereas those used for Fmoc synthesis are designed to be labile to TFA. Thus, the stability of the protecting group to the conditions of the sequencing chemistry will affect their analysis. Table III lists the elution positions of many of the common side-chain protected amino acids used in peptide synthesis and provides notes on their relative stability. They usually fall into three classes: those that are completely stable to the chemistry, those that are only partially stable, and those that are completely or nearly completely unstable. In most instances, the unstable derivatives revert back to the native amino acid and will often show signals for both native amino acid and derivative. For example, PTH-His(Bum), which elutes at 27.8 min, converts to PTH-His relatively quickly with successive Edman cycles. So, early in a run it may be identified predominately as PTH-His(Bum) but later in the run, predominately as PTH-His. Because the elution conditions used on one sequencer are never exactly the same as another, the elution positions of the PTH-amino acids will vary to some extent. Thus, Table III should be used only as a guide to relative elution position.

In addition, the linkage of the peptide to the resin is usually much less stable when employing Fmoc chemistry than for Boc chemistry. This may result in better retention of the peptide in the sequencer (less washout) when acid-stable linkages are used, but otherwise does not essentially change the analysis. Polybrene may be included with labile resin linkages to try to increase retention of the peptide, but in practice its effect is variable.

One may approach the loading of resin-bound peptides either qualitatively or quantitatively. That is, resin can either be loaded by applying a fixed volume of suspended beads, or individual beads can be counted under magnification. In either case, the resin beads are suspended at approximately 1 mg/ml in methanol, acetonitrile, or 40% (v/v) methanol in dichloromethane. Immediately following vigorous vortexing, 20 μl of suspension is loaded onto a TFA-treated glass fiber filter that is placed in the top glass cartridge block of a Perkin-Elmer Applied Biosystems sequencer. A cartridge seal is placed over this, followed by the bottom glass cartridge block.