PROTOCOL

Enzymatic Cleavage and HPLC Peptide Mapping of Proteins

Kenneth R. Williams and Kathryn L. Stone*

Abstract

Detailed procedures are described for successfully digesting reasonably small quantities (i.e., usually >10 pmol) of proteins with a variety of proteases and for then isolating the resulting peptides by reverse-phase HPLC. Since sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) appears to be the current method of choice for final purification of proteins for structural analysis, special attention is given to carrying out in-gel proteolytic digests on SDS-PAGE-separated proteins that have usually been stained with Coomassie Blue. A compilation of data from nearly 200 "unknown" samples is used to help provide realistic expectations with respect to the results that are likely to be obtained from carrying out in-gel proteolytic digests on large numbers of proteins.

Index Entries: SDS-PAGE; HPLC; in-gel digestion; comparative HPLC peptide mapping; pepsin; trypsin; chymotrypsin; lysyl endopeptidase; protease V8.

1. Introduction

The purpose of this article is to provide realistic procedures for enzymatically cleaving reasonably small quantities of proteins and for then fractionating the resulting digests via reverse-phase high-pressure liquid chromatography (HPLC). Because sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is the current method of choice for purifying the >10 pmol amounts of protein that are most often used for HPLC peptide mapping, those procedures for generating internal peptides from proteins that have been separated by SDS-PAGE are often the most useful. Several approaches may be taken to obtain peptides from SDS-PAGE-separated proteins. These include, but are not limited to:

1. Digesting the protein in the gel in the presence *(1)* or absence of SDS *(2–5)* and then diffusing the resulting peptides out of the gel.
2. Blotting the protein onto nitrocellulose *(6,7)* or polyvinylidene difluoride (PVDF) *(7)* and then digesting it *in situ*.
3. Blotting the protein onto PVDF, cleaving it *in situ* with cyanogen bromide, eluting the resulting peptides and then subjecting them to further digestion with trypsin *(8)*.

Each of these procedures has its own advantages and disadvantages with the impetus toward deriving better "in-gel" techniques, deriving from the observation that electroblotting is often not quantitative. Although the emphasis of this article is on enzymatically digesting and HPLC peptide mapping of Coomassie Blue-stained proteins that have been isolated in SDS-PAGE, procedures are also described for digesting proteins in solution. In both instances, particular attention is given to the special requirements imposed by comparative HPLC peptide mapping and high-sensitivity internal sequencing. In the latter case, a compilation of data from nearly 200 "unknown" samples is used to help establish realistic expectations with regard to overall peptide recovery, the fraction of peptides sequenced that provide useful data, and in particular the impact that decreasing sample amount has on these parameters and on the overall probability of success.

*Author to whom all correspondence and reprint requests should be addressed. W. M. Keck Foundation Biotechnology Resource Laboratory and Howard Hughes Medical Institute Biopolymer Facility, Yale University, New Haven, CT 06536, e-mail: kenneth.williams@yale.edu.

Molecular Biotechnology ©1997 Humana Press Inc All rights of any nature whatsoever reserved 1073–6085/1997/8 2/155–167/$11 25

2. Materials

2.1. SDS-PAGE

1. Laemmli gels *(9)*: Prepare using 7–15% polyacrylamide, with the percent polyacrylamide determined by the size of the protein.
2. Gel stain: 0.1% Coomassie Blue in 10% acetic acid, 50% methanol, 40% H_2O (1.0 g Coomassie Blue + 100 mL acetic acid + 500 mL methanol + 400 mL H_2O).
3. Gel destain: 10% acetic acid, 50% methanol, 40% H_2O (100 mL acetic acid + 500 mL methanol + 400 mL H_2O).

2.2. Amino Acid Analysis

Acid for hydrolysis: 6*N* HCL, 0.2% phenol containing 2 nmol/100 µL norleucine as an internal standard (50 mL concentrated HCL + 200 µL phenol + 100 µL 20 m*M* norleucine + H_2O to a total volume of 100 mL).

2.3. Enzymatic Digestion of Proteins

1. Enzymatic digestion is carried out with modified trypsin from Promega, sequencing-grade chymotrypsin, or endoproteinase Glu-C (Protease V8 from *Staphylococcus Aureus*) (Boehringer Mannheim, Indianapolis, IN), lysyl endopeptidase (#129-02541, Achromobacter Protease I from *Achromobacter lyticus*) (Wako Pure Chemical Industries, Ltd, Osaka, Japan) or pepsin from Sigma (St. Louis, MO).
2. 0.1 mg/mL Trypsin stock solutions: Prepare by dissolving a 20-µg aliquot (as purchased from Promega, Madison, WI) in 0.2 mL 1 m*M* HCl. Immediately before use, a 0.0333 mg/mL working solution is prepared by adding 1 volume 0.1 mg/mL stock solution to 2 volumes 200 m*M* NH_4HCO_3, pH 8.0.
3. 0.1 mg/mL Lysyl endopeptidase stock solutions: Prepare by adding the appropriate volume of water to a 2 or 10 AU vial (Wako) after using the stated specific activity, which is usually approx 4.5 AU/mg, to determine the dry weight of protein in the vial. Immediately before use, a 0.0333 mg/mL working solution is prepared by adding 1 volume 0.1 mg/mL stock solution to 2 volumes 200 m*M* NH_4HCO_3, pH 8.0.
4. 0.1 mg/mL Chymotrypsin stock solutions: Prepare by dissolving 100 µg enzyme (Boehringer Mannheim) in 1 mL of 1 m*M* HCl.
5. 0.1 mg/mL Protease V8 stock solutions: Prepare by dissolving 50 µg enzyme (Boehringer Mannheim) in 500 µL of 50 m*M* NH_4HCO_3.
6. Pepsin (Sigma): Dissolve in 5% formic acid (Baker, Phillipsburg, NJ) at a concentration of 0.1 mg/mL.
7. 8 *M* Urea, 0.4 *M* NH_4HCO_3 buffer (for digesting proteins in solution): Prepare by dissolving Sequanal Grade urea (4.8 g) and Baker ammonium bicarbonate (0.316 g) in 10 mL H_2O.
8. 0.1% Trifluoroacetic acid (TFA), 60% CH_3CN (for in-gel digestion): Prepare by mixing 0.5 mL 20% (v/v) TFA with 60 mL CH_3CN (Baker), then add deionized water to a final volume of 100 mL.
9. The 50% CH_3CN, 0.2 *M* NH_4HCO_3, pH 8.0 (for in-gel digestion): Prepare by mixing 1.58 g Baker ammonium bicarbonate (0.316 g) with 40 mL deionized water and 50 mL CH_3CN (Baker), then bring to a final volume of 100 mL with deionized water.
10. 45 m*M* Dithiothreitol (DTT) (for protein reduction): Prepare by dissolving 69.3 mg DTT in 10 mL H_2O.
11. The 100 m*M* Iodoacetic acid (IAA) (for protein alkylation): Prepare by dissolving 185.9 mg IAA in 10 mL H_2O.
12. 0.05% SDS, 5 m*M* NH_4HCO_3, pH 8.0 solution (for protein dialysis): Prepare by adding 5 mL 1% SDS and 39.5 mg NH_4HCO_3 to 95 mL deionized water.

2.4. HPLC Separation of Peptides

1. HPLC system: Peptide separations should be carried out on a Hewlett Packard 1090 *M* HPLC system equipped with a diode array or variable-wavelength detector, a 250-µL injection loop (or comparable system). Fractions should be collected by peak automatically into capless Eppendorf tubes (positioned in 13 × 100-mm test tubes) using an Isco Foxy fraction collector with an Isco Model 2150 Peak Separator.
2. A Vydac C-18 reverse phase (or comparable), 1 × 250 mm column (Separations Group, Hesperia, CA), is recommended.
3. pH 2.0 Buffer system:
   a. Buffer A: 0.06% TFA (3 mL 20% TFA/L of H_2O).

b. Buffer B: 0.052% TFA, 80% acetonitrile (2.7 mL 20% TFA, 800 mL CH_3CN, deionized water to a final volume of 1.0 L).

4. pH 6.0 Buffer system:
   a. Buffer C: 5 m*M* potassium phosphate, pH 6.0 (12 mL 0.5 *M* KH_2PO_4 in a total volume of 1200 mL H_2O).
   b. Buffer D: 1 m*M* potassium phosphate, pH 6.0, 80% (v/v) CH_3CN (200 mL buffer C and 800 mL acetonitrile).

3. Methods

3.1. SDS-PAGE

The percent acrylamide gel used for protein purification is determined by the size of the protein. In general, proteins >100 kDa were electrophoresed in 7–10% polyacrylamide gels, whereas smaller proteins were electrophoresed in 12.5–15% polyacrylamide gels. Staining of the gel should be carried out at room temperature for the minimum time necessary to visualize the bands of interest (typically <60 min). Destaining should be carried out for a minimum of 3 h with at least one solvent change. If the gel still has a Coomassie Blue background, destaining should be continued until the background is nearly clear. After destaining, the protein of interest (along with a blank section of gel approximately equal in size to that containing the protein of interest) is excised from the gel using a razor blade and tweezers and frozen at –20°C in an Eppendorf tube.

3.2. Amino Acid Analysis

1. Remove 10–15% of the sample (or of the gel slice containing the sample) and hydrolyze *in vacuo* in 100 µL (or, in the case of gel samples, 200 µL) 6 *N* HCl, 0.2% phenol (containing 2 nmol/100 µL norleucine as an internal standard) at 115°C for 16 h.
2. Following hydrolysis, solution samples should be dried in a Speedvac and, in the case of gel hydrolysates, the supernatant is transferred to a second tube and then dried in a Speedvac.
3. The dried hydrolysate is then dissolved in Na-S sample dilution buffer and run on a Beckman Model 7300 Amino Acid Analyzer using ion-exchange separation of the amino acids and post-column ninhydrin detection or dissolved in 70% formic acid and run on a phenylthiocarbamyl (PTC) amino acid analyzer.

3.3. Enzymatic Digestion of Proteins

3.3.1. Digestion of Proteins in Solution

3.3.1.1. Trypsin, Chymotrypsin, Lysyl Endopeptidase, or Protease V8 Digestion

Enzymatic digestion of proteins requires that a reasonable level of care be exercised in terms of final sample preparation (*see* **Note 1**). Proteins that are isolated in solution, that contain <~0.1 mmol monovalent salt, and that are free of detergents and glycerol can often simply be dried in a Speedvac. Higher levels of salts and many detergents such as SDS can be removed from the sample by adding 1/9th volume of 100% trichloroacetic acid (TCA), incubating on ice for 30 min, centrifuging, and then washing the protein pellet with 100 µL cold acetone. In general, we recommend that the glycerol concentration be lowered to below 15% and that, if possible, the protein concentration be increased to at least 100 µg/mL prior to TCA precipitation. An alternative approach that may be taken with samples that contain high concentrations of salts and SDS is to first dialyze them vs 0.05% SDS, 5 m*M* NH_4HCO_3 to remove the salt. After dialysis, the sample may then be dried in a Speedvac prior to adding 50 µL water and 450 µL cold acetone. After incubating the sample at –20°C for at least 1 h and centrifuging, the pellet is then washed with 100 µL cold acetone to remove the SDS ***(10)***. If necessary, the acetone extraction may be repeated one or more times to remove even relatively large amounts of detergent.

Following removal of any detergent that might have been present and lowering of the salt concentration, the sample is ready for enzymatic digestion. Trypsin is often the enzyme of choice for comparative mapping because it cleaves with high specificity at the COOH-terminal side of lysine and arginine. Although bonds involving acidic amino acids are cleaved slowly, the only bonds that are extremely resistant to trypsin cleavage are those involving lysine-proline and arginine-proline linkages. Another advantage of trypsin

(and chymotrypsin) is that it can readily digest proteins that are insoluble in the 2 *M* urea, 0.1 *M* NH_4HCO_3 digestion buffer. Chymotrypsin has significantly less specificity than trypsin in that it cleaves at the COOH-terminal side of tryptophan, tyrosine, and phenylalanine residues with additional cleavages occurring after some leucine, methionine; and other amino acids containing hydrophobic amino acid side chains. Another commonly used enzyme is lysyl endopeptidase, which cleaves specifically after lysine. This enzyme has an advantage in that it will produce larger fragments than trypsin. Finally, Protease V8 cleaves after glutamic acid residues in either ammonium bicarbonate (pH 8.0) or ammonium acetate (pH 4.0) buffers, and cleaves after both aspartic acid and glutamic acid residues in phosphate buffer (pH 7.8).

All enzyme stocks should be divided into aliquots and frozen at –20°C. Stocks of trypsin and chymotrypsin are stable under these conditions for at least 6 mo, whereas the lysyl endopeptidase is stable for up to 2 yr (based on the manufacturer's recommendations). According to the manufacturer, endoproteinase Glu-C is stable for approximately 1 mo at –20°C. All enzyme stocks should be discarded once thawed.

The digestion procedure that follows can be used with any of the aforementioned enzymes (assuming the buffer is changed in the case of Protease V8).

1. Dissolve the dried or precipitated protein in 20 µL 8 *M* urea, 0.4 *M* NH_4HCO_3, pH 8.0, and then remove a 10–15% aliquot for amino acid analysis. If the analysis indicates there is sufficient protein to digest (*see* **Note 2**), proceed with **step 2**; otherwise, additional protein should be prepared to pool with the sample.
2. Check the pH of the sample by spotting 1–2 µL on pH paper. If necessary, adjust the pH to between 7.5–8.5.
3. Add 5 µL 45 m*M* DTT and incubate at 50°C for 15 min to reduce the protein (*see* **Note 3** regarding the necessity of this step).
4. After cooling to room temperature, alkylate the protein by adding 5 µL 100 m*M* IAA and incubating at room temperature for 15 min.
5. Dilute the digestion buffer with H_2O so that the final digest will be carried out in 2 *M* urea, 0.1 *M* NH_4HCO_3.
6. Add the enzyme in a 1:25, enzyme:protein (weight:weight) ratio (*see* **Note 4** for exceptions to the 1:25 [w/w] guideline).
7. Incubate at 37°C for 24 h.
8. Stop the digest by freezing, acidifying the sample with TFA, or by injecting onto a reverse phase HPLC system.

3.3.1.2. Pepsin Digestion

Although the broad specificity of pepsin hinders its routine use, it is applicable in the case of otherwise intransigent proteins as well as for further digesting relatively small peptides and, particularly, for studies directed at identifying disulfide bonds (*see* **Note 5**). Under acidic conditions, pepsin cleaves proteins at a wide variety of peptide bonds. Although it cleaves preferentially between adjacent aromatic or leucine residues, pepsin also cleaves at either the NH_2- or COOH-terminal side of any amino acid except proline. A typical digestion procedure follows:

1. Dissolve the dried protein in 100 µL 5% formic acid.
2. Add pepsin at a 1:50, enzyme:protein (w:w) ratio.
3. Incubate the sample at room temperature for 1–24 h, with the time of incubation being dependent upon the desired extent of digestion.
4. Dry the digest in a Speedvac prior to dissolving in 0.05% TFA and injecting onto a reverse-phase HPLC system.

3.3.2. Digestion of Proteins in SDS Polyacrylamide Gels

Although the in-gel digestion procedure has succeeded with as little as 1–5 pmol amounts of "unknown" proteins, in general we recommend that at least 10 pmol protein be subjected to SDS-PAGE and that the density of protein in the gel band that is to be digested be above ~0.05 µg/mm^3. Generally, the latter requires that at least 1–2 µg of the protein of interest be loaded in a single lane of a 0.75-mm thick SDS polyacrylamide gel. At this level of protein, there are usually only a few absorbance peaks in the "blank," control digest

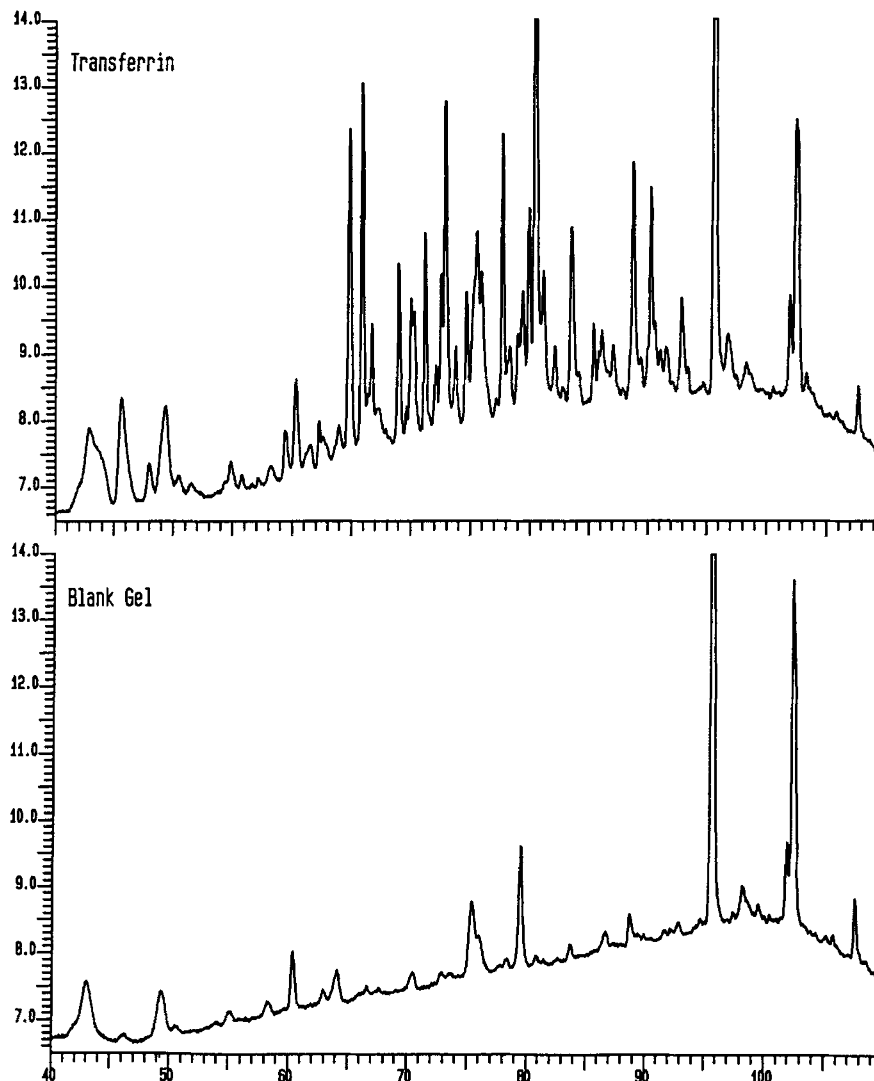

Fig. 1. In-gel tryptic digestion of transferrin. Following SDS-PAGE of 25 pmol of transferrin in a 12.5% polyacrylamide gel, the gel was stained and destained as described in Materials section. The protein band was then excised, along with a control blank, and digested in the gel with trypsin as described in Methods section. Peptides were chromatographed on a Vydac C-18, 1 × 250-mm reverse-phase column that was eluted at a flow rate of 50 μL/min as described in Materials section. A comparison of the transferrin digest (top profile) with the control digest (bottom profile), which was carried out on a blank section of gel, indicates the digest proceeded well.

that are as intense as those in the protein digest (*see* **Fig. 1**) and the overall success rate approaches 100% (*see* **Table 1**). The Coomassie Blue-stained band of interest, along with an equal size, blank section of gel, is excised, and 10–15% quantitated by amino acid analysis as outlined in **Subheading 3.2. Notes 6–9** contain further discussion relating to the amounts and concentrations of protein that are required and of the results that might be expected from the in-gel digestion procedure. The procedure is as follows:

1. Sample and blank gel pieces are cut into approx 1 × 2-mm sections, placed into 1.5-mL Eppendorf tubes (which were prewashed with Buffer E [0.1% TFA, 60% CH_3CN]) and then washed with 250 μL Buffer F (50% CH_3CN, 200 m*M* NH_4HCO_3, pH 8.0) for 30 min at room temperature on a tilt table.

Table 1
Summary of Results Obtained from 191 in Gel Digests

Parameter	Amount of protein digested (pmol)				Total
	<50	51–100	101–200	>200	
Number of proteins digested	28	54	65	44	191
Average mass of protein, kDa	87	62	60	59	64
Median amount digested, pmol	29	78	138	271	100
Avg. density protein band, μg/mm^3	0.10	0.22	0.26	0.49	0.28
Avg. # peptides sequenced/protein	2.2	2.1	2.2	2.0	2.1
% Peptides successfully sequenced	77.4	77.0	82.1	84.3	80.4
Average % initial sequence yield[a]	17.6	10.0	11.2	12.8	12.2
Avg. # residues sequenced/peptide	11.2	11.4	12.8	14.2	12.5
Overall digest success rate, %	100	96.3	96.9	97.7	97.9

[a]Based on the initial peptide sequencing yield divided by the estimated amount of protein digested which is based on hydrolysis/amino acid analysis of the submitted gel slice.

2. After removing the wash, sufficient Buffer F (usually about 100 μL) is added to cover the gel pieces and the approximate total volume estimated by comparing to Eppendorf tubes containing known volumes of water.
3. Sufficient 45 m*M* DTT is then added to bring the final concentration to 1 m*M* before incubating the samples for 20 min at 37°C.
4. Twice the volume (as compared to DTT) of 100 m*M* methyl 4-nitrobenzene sulfonate (or an equal volume of iodoacetic acid or iodoacetamide) is added followed by a 40-min incubation at 37°C.
5. After removing the supernate, the gel slices are washed at room temperature on a tilt table for 30 min and then twice more for 15 min with 250 μL Buffer F.
6. After removing the last wash, the gel pieces are brought to dryness in a Speedvac and then hydrated by adding 1 μL/mm^3 (initial estimated gel volume) of a freshly prepared enzyme solution made by mixing 1 volume 0.1 mg/mL trypsin (Promega modified) or lysyl endopeptidase (Wako) with two volumes 200 m*M* NH_4HCO_3. If necessary, additional enzyme solution (0.0333 mg/mL) is added to totally immerse the gel pieces.
7. After incubating at 37°C for 24 h, peptides are extracted with 100 μL (or a volume equal to the gel volume if that is larger) Buffer E for 1 h at room temperature on a tilt table.
8. After repeating **step 7**, the combined extracts are dried in a Speedvac, dissolved in 20 μL 0.05% TFA, 25% CH_3CN, and diluted with 90 μL 0.05% TFA prior to subjecting 100 μL to HPLC.

3.4. HPLC Separation of Peptides

The TFA/acetonitrile buffer system described in **Subheading 2.4.** is an almost universal reverse-phase solvent system owing to its low ultraviolet (UV) absorbance, high resolution, and excellent solubilizing properties. By adding a slightly higher percentage of TFA (0.06%) to Buffer A than to Buffer B (0.052%), the baseline at 210 nm can be readily balanced to allow very high sensitivity runs ***(11)***. The gradient that we have generally used is:

Time	% B or D
0–60 min	2–37.5%
60–90 min	37.5–75%
90–105 min	75–98%

In the case of extremely complex digests (i.e., tryptic digests of proteins that are above about 100 kDa), the gradient times given may be doubled. Peptides that are incompletely resolved may be rechromatographed under the same conditions on a C-8 Aquapore column (1.0 × 250 mm). This can most conveniently be done by diluting the peak twofold (or more) with 0.02% Tween-20 prior to injecting onto the Aquapore column. Alternatively, peptides isolated from the Vydac C-18 column can be rechromatographed on the same column using the pH 6.0 buffer system. The differing

selectivity of the Aquapore column and of the pH 6.0 buffer system often bring about significant further purification of peptides that were originally isolated from a Vydac C-18 column developed in the TFA buffer system *(11)*. Further discussion regarding optimizing the reproducibility, resolution, and sensitivity of reverse-phase HPLC may be found in **Note 10**. Finally, **Note 5** discusses more general aspects of using comparative HPLC peptide mapping to identify and isolating peptides that contain posttranslational modifications.

4. Notes

1. In many instances, large losses occur during the final purification steps when the protein concentrations are invariably lower. Hence, although ultrafiltration or dialysis of a 5 mg/mL crude solution of a partially purified enzyme may lead to nearly 100% recovery of activity, similar treatment of a 25 µg/mL solution of the purified protein might well lead to significant, if not total loss of activity owing to nonspecific adsorption. Similarly, the effectiveness of organic and acid precipitation procedures often decreases substantially as the final protein concentration is decreased below about 100 µg/mL. Whenever possible, therefore, the final purification step should be arranged such that the resulting protein solution is as concentrated as possible and, ideally, can simply be dried in a Speedvac prior to enzymatic digestion. In this regard, it should be noted that a final NaCl concentration of 1 *M* does not significantly effect the extent of trypsin digestion. When it is necessary to carry out an organic or acid precipitation to remove salts or detergents, the protein should first be dried in a Speedvac (in the 1.5-mL tube in which it will ultimately be digested) prior to redissolving or suspending in either a minimum volume of water (in the case of an acetone precipitation) or in 10% TCA (in the case of an acid precipitation/extraction) so as to increase the protein concentration and minimize loss. Two common contaminants that are extremely deleterious to enzymatic cleavage are detergents (as little as 0.005% SDS will noticeably decrease the rate of tryptic digests carried out in the presence of 2 *M* urea *[10]*) and ampholines. Because detergent removal is often associated with protein precipitation and because many detergents (such as SDS) form large micelles that cannot be effectively dialyzed, it is usually preferable to extract the detergent from the protein (that has been dried in the tube in which it will be digested), rather than to dialyze it away from the protein. In the case of ampholines, our experience is that even prolonged dialysis extending over several days with 15,000 Da cut-off membrane is not sufficient to decrease the ampholine concentration to a level that permits efficient trypsin digestion. Rather, the only effective methods that we have found for complete removal of ampholines are TCA precipitation or hydrophobic chromatography.
2. In our experience, one of the most common causes of "failed" digests is that the amount of protein being subjected to digestion has been over-estimated. Often this is owing to the inaccuracy of dye-binding and colorimetric assays. For this reason, we recommend that an aliquot of the sample be taken for hydrolysis and amino acid analysis prior to digestion. The aliquot for amino acid analysis should be taken either immediately prior to drying the sample in the tube in which it will be digested, or after redissolving the sample in 8 *M* urea, 0.4 *M* NH_4HCO_3. Although up to 10 µL of 8 *M* urea is compatible with ion-exchange amino acid analysis, this amount of urea may not be well-tolerated by PTC amino acid analysis. Hence, in the latter case the amino acid analysis could be carried out prior to drying and redissolving the sample in urea. Although it is possible to succeed with less material, to ensure a high probability of success we recommend that a minimum of 10 pmol protein be digested. Typically, 10–15% of this sample would be taken for amino acid analysis. In the case of a 50-kDa protein, the latter corresponds to only 0.125–0.188 µg protein being analyzed. When such small amounts of protein are being analyzed, it is important to control for the ever-present background of free amino acids that are in buffers, dialysis tubing, plastic tubes and tips, and so on. If sufficient protein is available, aliquots should be analyzed both before (to determine the free amino acid

concentration) and after hydrolysis. Alternatively, an equal volume of sample buffer should be hydrolyzed and analyzed and this concentration of amino acids should then be subtracted from the sample analysis.

3. Because many (probably most) native proteins are resistant to enzymatic cleavage, it is usually best to denature the protein prior to digestion. Although some proteins may be irreversibly denatured by heating in 8 *M* urea (as described in the aforementioned protocol), this treatment is not sufficient to denature transferrin. In this instance, prior carboxymethylation, which irreversibly modifies cysteine residues, brings about a marked improvement in the resulting tryptic peptide map ***(10)***. Another advantage of carboxymethylating the protein is that this procedure enables cysteine residues to be identified during amino acid sequencing. Cysteines have to be modified in some manner prior to sequencing to enable their unambiguous identification. Under the conditions that are described in Methods section, the excess DTT and IAA do not interfere with subsequent digestion. Although carboxymethylated proteins are usually relatively insoluble, the 2 *M* urea that is present throughout the digest is frequently sufficient to maintain their solubility. However, even in those instances where the carboxymethylated protein precipitates following dilution of the 8 *M* urea to 2 *M*, trypsin and chymotrypsin will usually still provide complete digestion. Often, the latter is evidenced by clearing of the solution within a few minutes of adding the enzyme. If carboxymethylation is insufficient to bring about complete denaturation of the substrate, an alternative approach is to cleave the substrate with cyanogen bromide (1000-fold excess over methionine, 24 h at room temperature in 70% formic acid). The resulting peptides can then either be separated by SDS-PAGE (because they usually do not separate well by reverse-phase HPLC) or, preferably, they can be enzymatically digested with trypsin or lysyl endopeptidase and then separated by reverse-phase HPLC. If this approach fails, the protein may be digested with pepsin, which, as previously described, is carried out under very acidic conditions, or be subjected to partial acid cleavage ***(12)***. However, the disadvantage of these latter two approaches is that they produce an extremely complex mixture of overlapping peptides. Finally, extensive glycosylation (i.e., typically >10–20% by weight) can also hinder enzymatic cleavage. In these instances, it is usually best to remove the carbohydrate prior to beginning the digest.

4. Every effort should be made to use as high substrate and enzyme concentration as possible to maximize the extent of cleavage. Although the traditional 1:25 (w/w) ratio of enzyme to substrate provides excellent results with mg amounts of protein, it will often fail to provide complete digestion with low microgram amounts of protein. For instance, using the procedures previously outlined, this w/w ratio is insufficient to provide complete digestion when the substrate concentration falls below about 20 μg/mL ***(13)***. The only reasonable alternative to purifying additional protein is to either decrease the final digestion volume below the 80 μL value previously used or to compensate for the low substrate concentration by increasing the enzyme concentration. The only danger in doing this, of course, is the increasing risk that some peptides may be isolated that are autolysis products of the enzyme. Assuming that only enzymes such as trypsin, chymotrypsin, lysyl endopeptidase, and Protease V8 are used (whose sequences are known), it is usually better to risk sequencing a peptide obtained from the enzyme (which can be quickly identified via a database search) than it is to risk incomplete digestion of the substrate. Often, protease-autolysis products can be identified by comparative HPLC peptide mapping of an enzyme (i.e., no substrate) control and by subjecting candidate HPLC peptide peaks to matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) prior to sequencing. The latter can be extremely beneficial both in ascertaining the purity of candidate peptide peaks as well as in identifying (via their mass) expected protease-autolysis products. In order to promote more extensive digestion, we have sometimes used enzyme:substrate mole ratios that approach unity. If there is any doubt concerning the appropriate enzyme concentration to use with a particular substrate

concentration, it is usually well worth the effort to carry out a control study (using a similar concentration and size-standard protein) where the extent of digestion (as judged by the resulting HPLC profile) is determined as a function of enzyme concentration.

5. Provided that samples of both the modified and unmodified protein are available, comparative HPLC peptide mapping provides an extremely facile means of rapidly identifying peptides that contain posttranslational modifications. In the case of proteins that have been expressed in *Escherichia coli*, the latter can often serve as the unmodified control as relatively few posttranslational modifications occur in this organism. Certainly the first attempt at comparative HPLC tryptic peptide mapping should be with enzymes such as trypsin or lysyl endopeptidase that have high specificity, and the digests should be separated using acetonitrile gradients in 0.05% TFA. Although elution position (as detected by absorbance at 210 nm) provides a sensitive criterion to detect subtle alterations in structure, the value of comparative HPLC peptide mapping can be further enhanced by multi-wavelength monitoring and, especially, by on-line or off-line MS of the resulting peptide fractions. If comparative peptide mapping fails to reveal any significant changes, it is often worthwhile running the same digest in the pH 6.0 phosphate-buffered system that was previously mentioned. At this higher pH, some changes such as deamidation of asparagine and glutamine produce a larger effect on elution position than at pH 2.2, where ionization of the side-chain carboxyl groups would be suppressed. Another possible reason for failing to detect differences upon comparative HPLC is that the peptide(s) containing the modifications are either too hydrophilic to bind or too hydrophobic to elute from reverse-phase supports. Hence, in addition to trying a different HPLC solvent system, another approach that may be taken to expand the capabilities of comparative HPLC peptide mapping is to try a different proteolytic enzyme such as chymotrypsin or Protease V8. Finally, the failure to observe any difference upon comparative HPLC peptide mapping may result from loss of the posttranslational modification during either the cleavage or the subsequent HPLC. Assignment of disulfide bonds is one example where this can be a problem. Even if the reduction and carboxymethylation steps are deleted from the method previously outlined (so that native disulfide bonds are left intact), disulfide interchange may occur during enzymatic cleavage, which is typically carried out at pH 8.0. This problem can be addressed by either going to shorter digestion times ***(14)*** or by carrying out the cleavage under acidic conditions where disulfide interchange is less likely to occur. For this reason, pepsin (which is active in 5% formic acid) digests are often used for isolating disulfide-bonded peptides. Providing that the control sample is reduced, comparative HPLC peptide mapping can be used to rapidly isolate disulfide-linked peptides. If the sequence of the protein of interest is known, then MALDI-MS (before and after reduction of the disulfide-linked peptide) can be used to identify the two peptides that are disulfide bonded.
6. As in the case of in solution digests, care must be exercised to guard against sample loss during final purification. Whenever possible, SDS (0.05%) should simply be added to the sample prior to drying in a Speedvac and subjecting to SDS-PAGE. Often times, however, if the latter procedure is followed, the final salt concentration in the sample will be too high (i.e., >~1 *M*) to enable it to be directly subjected to SDS-PAGE. In this instance the sample may either be concentrated in a Speedvac and then precipitated with TCA (as previously described) or it may first be dialyzed to lower the salt concentration. If dialysis is required, the dialysis tubing should be rinsed with 0.05% SDS prior to adding the sample, which should also be made 0.05% in SDS. After dialysis vs a few m*M* NH_4HCO_3 containing 0.05% SDS, the sample may be concentrated in a Speedvac and then subjected to SDS-PAGE. (Note that samples destined for SDS-PAGE may contain several % SDS.) Finally, an alternative approach to concentrating samples prior to SDS-PAGE is to use a funnel-well gel ***(15)***, where the protein is actually concentrated while it is being subjected to SDS-PAGE.

7. Although most estimates of protein amounts are based on relative staining intensities, our data suggest that, in general, there is at least a two- to threefold range in the relative staining intensity of different proteins. In the case of Coomassie Blue, this range in staining intensity probably results from the differential content of basic amino acids, which appear to represent the primary binding sites for this dye *(16)*. Obviously, when working with limiting amounts of protein, such a two- to threefold range could well mean the difference between success and failure. Hence, we routinely subject an aliquot of the SDS-PAGE gel (usually 10–15% based on the length of the band) to hydrolysis and ion-exchange amino acid analysis prior to proceeding with the digest. As these analyses will often contain less than 0.5 μg protein, it is important that a "blank" section of gel that is about the same size as that containing the sample, be hydrolyzed and analyzed as a control to correct for the background level of free amino acids that are usually present in polyacrylamide gels. Based on samples taken from 156 different gels submitted by users of the W. M. Keck Foundation Biotechnology Resource Laboratory, the background typically ranges up to about 0.2 μg and averages about 0.07 μg in these 10–15% aliquots. Although amino acid analyses on gel slices are complicated by this background level of free amino acids and by the fact that some amino acids (i.e., glycine, histidine, methionine, and arginine) usually cannot be quantitated following hydrolysis of gel slices, these estimates are still considerably more accurate than are estimates based on relative staining intensities. In those instances where amino acid analysis indicates <10 pmol protein, the stained band may be stored frozen while additional material is purified. If it is not possible to carry out high-sensitivity amino acid analyses on aliquots of SDS-PAGE bands submitted for internal sequencing, we suggest that several concentrations of commercial mixtures of known proteins be run on the same SDS polyacrylamide gel as the protein of interest so that the relative staining intensity can be more accurately estimated.
8. In general, the sample should be run in as few SDS-PAGE lanes as possible to maximize the substrate concentration and to minimize the total gel volume present during the digest. Whenever possible, a 0.5–0.75-mm thick gel should be used and at least 1–2 μg of the protein of interest should be run in each gel lane so the density of the protein band is at least 0.050 μg/mm^3. The data in **Table 1** summarizes the results of internal sequencing of 191 "unknown" proteins submitted to the internal sequencing service of the Keck Facility at Yale University. The median amount of protein digested in these studies was 100 pmol and the average number of peptides sequenced per protein was close to two. This number is relatively low because 68.4% of the proteins summarized in **Table 1** were identified based on searching protein databases with the first peptide sequence obtained. In these instances, additional confirmation of the identification was obtained on the basis of the apparent molecular weight of the protein and by matching observed and predicted peptide masses. By "screening" peptides destined for sequencing with MALDI-MS *(17)*, we have been able to maintain an 80% success rate in terms of successfully sequencing peptides obtained from in gel digests. Approximately 10% of peptides subjected to sequencing fail to provide any data either because they derive from the (usually) blocked NH_2-terminus of the protein or perhaps were lost subsequent to HPLC collection, whereas the remaining 10% of peptides that fail to provide usable sequences prove to contain mixtures that were not detected by either HPLC absorbance peak shape or MALDI-MS screening *(17)*. It is important to note that the overall percent initial sequencing yields, which have been calculated based on the average initial peptide sequencing yield divided by the amount of protein digested, are usually near 12–18% with the higher value of about 18% observed for the <50 pmol samples probably resulting from slight underestimation of the amount of protein actually digested in this range (**Table 1**). The problem in this regard is that as the actual amount of protein hydrolyzed in the aliquot of gel matrix is decreased, nonspecific

losses owing to adsorption and other factors become more important. As noted *(4,18)*, the overall success rate of in gel digests (98%) is extremely high and is all the more remarkable in view of the fact that the data summarized in **Table 1** derives from in-gel samples prepared by >150 different principal investigators. Although several enzymes (i.e., trypsin, chymotrypsin, lysyl endopeptidase, and Protease V8) may be used with the in-gel procedure, nearly all of our experience (and all of the data in **Table 1**) has been obtained with trypsin and lysyl endopeptidase.

9. To minimize the HPLC background resulting from reagent peaks, we recommend that the sample be reduced and cysteines be modified prior to in-gel digestion *(18)* and that when extensive (i.e., >10–20% by weight) glycosylation is present, that the sample be deglycosylated prior to cleavage. Often the latter can be accomplished immediately prior to SDS-PAGE, which thus prevents loss (owing to insolubility) of the deglycosylated protein and effectively removes the added glycosidases.
10. Three factors that play a critical role in HPLC peptide mapping are reproducibility, resolution, and—when the amount of protein is limiting—sensitivity. In the following paragraphs, each of these topics will be discussed briefly. For a more complete discussion, the reader is referred to Mant *(19)*. Reproducibility is important both in terms of being able to identify artifact peaks by comparison to the control digest carried out on a blank section of SDS polyacrylamide gel and for comparative peptide mapping, which requires that successive chromatograms of digests of the same protein be sufficiently similar that they can be overlaid onto one another with little or no detectable differences. In general, the latter requires that average peak retention times not vary by more than about ±0.20% *(11)*. Assuming the digests were carried out under identical conditions, problems with regard to reproducibility often relate to the inability of the HPLC pumps to deliver accurate flow rates at the extremes of the gradient range. That is, to accurately deliver a 99% Buffer A/1% Buffer B composition at an overall flow rate of 0.15 mL/min requires that pump B be able to accurately pump at a flow rate of only 1.5 µL/min. The latter is well beyond the capabilities of many conventional HPLC systems. Although reproducibility can be improved somewhat by restricting the gradient range to 2–98%, as opposed to 0–100% Buffer B, the reproducibility of each HPLC system will be inherently limited in this regard by the ability of its pumps to accurately deliver low flow rates. Obviously, some HPLC systems that provide reproducible chromatograms at an overall flow rate of 0.5 mL/min might be unable to do so at 0.15 mL/min *(11)*. Similarly, minor check valve, piston seal, and injection valve leaks that go unnoticed at 0.5 mL/min might well account for reproducibility problems at 0.15 mL/min.

    The ability of HPLC to fractionate complex peptide mixtures and to discriminate peptides that contain minor posttranslational modifications is critically dependent upon resolution, which, in turn, depends on a large number of parameters including the flow rate, gradient time, column packing, and dimensions, as well as the mobile phase *(11,20)*. Studies with tryptic digests of transferrin suggest that, within reasonable limits, gradient time is a more important determinant of resolution than is gradient volume. In general, a total gradient time of ~100 min seems to represent a reasonable compromise between optimizing resolution and maintaining reasonable gradient times *(20)*. Optimal flow rates depend upon the inner diameter of the column being used, which, in turn, is dictated by the amount of protein that has been digested. In general, we find that optimal resolution of protein digests in the 50–250 pmol range is obtained at ~0.15 mL/min on 2.0–2.1-mm ID columns whereas larger amounts are best chromatographed at ~0.5 mL/min on 3.9–4.6-mm ID columns. However, increased sensitivity can be obtained by decreasing the flow rate to 50 µL/min and by using a 1-mm ID column. Under these conditions, it is possible to extend the range down to the less than 25 pmol level and still be able to fractionate amounts of digests that extend up to about 250 pmol. Unless precautions are taken to minimize dead volumes, significant problems may be encountered in terms of automated peak

detection/collection and post-column mixing as flow rates are lowered much below 0.15 mL/min *(11)*. Typically, the use of flow rates in the 25–75 µL/min range requires that fused silica tubing be used between the detector and the fraction collector and that a low-volume flow cell (i.e., 1–2 µL) be substituted for the standard flow cell in the UV detector.

Several commercially available C-18, reverse-phase supports provide high resolution. These include (but certainly are not limited to) Alltech Macrosphere, Vydac, Waters' Delta Pak *(13,20)* and Reliasil. Although under the conditions tested, the resolution obtained on a Brownlee Aquapore C-8 column was somewhat less than that on some other supports *(13,20)*, the Aquapore column appeared to have significantly different selectivity and hence provides a valuable means of further purifying peptides that may have been isolated on one of these other supports *(11)*. In general, best results appear to be obtained on 300 Å pore size, 5 micron particle size supports. In addition, because peptide resolution has been shown to be directly related to column length *(11,13,19)*, whenever possible the 25-cm versions of these columns should be used. One *caveat* in regards to the latter is that we have found that a 15-cm Delta-Pak C-18 column provides similar resolution to that obtained on a 25-cm Vydac C-18 column *(20)*.

Although the low-UV absorbance, high resolution, and excellent solubilizing properties of the 0.05% TFA/acetonitrile, pH 2.2, buffer system have made it the almost universal mobile phase for reverse-phase HPLC, there are occasions when a different mobile phase might be advantageous. Hence, the differing selectivity of the 5 m*M*, pH 6.0, phosphate system *(11)* makes this a valuable mobile phase for detecting posttranslational modifications (such as deamidation) that may be more difficult to detect at the lower pH of the TFA system (where ionization of side-chain carboxyl groups would be suppressed). In addition, changing the mobile phase provides another approach for further purifying peptides that were originally isolated in the TFA system.

The sensitivity of HPLC is dictated primarily by the volume in which each peak is eluted. Although sensitivity can be increased by simply decreasing the flow rate (while maintaining a constant-gradient time program), eventually the linear flow velocity on the column will be reduced to such an extent that optimal resolution will be lost. At this point the column diameter needs to be decreased so that an optimal linear flow velocity can be maintained at a lower flow rate. General guidelines for selecting flow rates and column diameters that optimize both resolution and sensitivity have been given. In general, the sensitivity of detection is increased as the wavelength is decreased with the practical limit being about 210 nm. As noted in **Subheading 3.4.**, high-sensitivity HPLC requires that the baseline be "balanced" by adding a slightly higher percentage of TFA to Buffer A than to Buffer B. Although we generally use 0.06% TFA in Buffer A compared to 0.052% TFA in Buffer B, minor alterations in these compositions can be accurately made following the running of a blank run with each new set of buffers. Finally, an important determinant of sensitivity (that is often overlooked) is the path length of the flow cell. For instance, an HP1090 equipped with a 0.6-cm path length cell provides (at the same flow rate) a threefold increase in sensitivity over that afforded by a Michrom UMA System equipped with a 0.2-cm path length cell.

References

1. Kawasaki, H., Emori, Y., and Suzuki, K. (1990) Production and separation of peptides from proteins stained with Coomassie brilliant blue R-250 after separation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal. Biochem.* **191,** 332–336.
2. Rosenfeld, J., Capdevielle, J., Guillemot, J., and Ferrara, P. (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **203,** 173–179.
3. Hellman, U., Wernstedt, C., Gonez, J., and Heldin, C. (1995) Improvement of an "in-gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal. Biochem.* **224,** 451–455.
4. Williams, K. R. and Stone, K. L. (1995) In gel digestion of SDS PAGE-separated proteins: observations from internal sequencing of 25 proteins, in *Techniques in Protein Chemistry VI* (Angeletti, R., ed.), Academic, New York, pp. 143–152.
5. Stone, K. L. and Williams, K. R. (1996) Enzymatic digestion of proteins in solution and in SDS polyacry-

lamide gels, in *The Protein Protocols Handbook* (Walker, J. M., ed.), Humana, Totowa, NJ, pp. 415–425.

6. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., and Kent, S. B. (1987) Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. USA* **84,** 6970–6974.
7. Fernandez, J., DeMott, M., Atherton, D., and Mische, S. M. (1992) Internal protein sequence analysis: enzymatic digestion for less than 10 micrograms of protein bound to polyvinylidene difluoride or nitrocellulose membranes. *Anal. Biochem.* **201,** 255–264.
8. Stone, K. L., McNulty, D. E., LoPresti, M. B., Crawford, J. M., DeAngelis, R., and Williams, K. R. (1992) Elution and internal amino acid sequencing of PVDF-blotted proteins, in *Techniques in Protein Chemistry III* (Angeletti, R., ed.), Academic, New York, pp. 23–34.
9. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680.
10. Stone, K. L. and Williams, K. R. (1993) Enzymatic digestion of proteins and HPLC peptide isolation, in *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Matsudaira, P. T., ed.), Academic, New York, pp. 45–69.
11. Stone, K. L., Elliott, J. I., Peterson, G., McMurray, W., and Williams, K. R. (1990) Reversed-phase high-performance liquid chromatography for fractionation of enzymatic digests and chemical cleavage products of proteins, in *Methods in Enzymology* (McCloskey, J., ed.), Academic, New York, pp. 389–412.
12. Sun, Y., Zhou, Z., and Smith, D. (1989) Location of disulfide bonds in proteins by partial acid hydrolysis and mass spectrometry, in *Techniques in Protein Chemistry* (Hugli, T., ed.), Academic, New York, pp. 176–185.
13. Stone, K. L., LoPresti, M. B., and Williams, K. R. (1990) Enzymatic digestion of proteins and HPLC peptide isolation in the sub-nanomole range, in *Laboratory Methodology in Biochemistry: Amino Acid Analysis and Protein Sequencing* (Fini, C., Floridi, A., Finelli, V. N., and Wittman-Liebold, B., eds.), CRC, Boca Raton, FL, pp. 181–205.
14. Glocker, M. O., Arbogast, B., Schreurs, J., and Deinzer, M. L. (1993) Assignment of the inter- and intramolecular disulfide linkages in recombinant human macrophage colony stimulating factor using fast atom bombardment mass spectrometry. *Biochemistry* **32,** 482–488.
15. Lombard-Platet, P. and Jalinot, P. (1993) Funnel-well SDS-PAGE: a rapid technique for obtaining sufficient quantities of low-abundance proteins for internal sequence analysis. *BioTechniques* **15,** 669–672.
16. Chial, H. J. and Splittergerber, A. G. (1993) A comparison of the binding of Coomassie brilliant blue to proteins at low and neutral pH. *Anal. Biochem.* **213,** 362–369.
17. Williams, K. R., Samandar, S. M., Stone, K. L., Saylor, M., and Rush, J. (1996) Matrix assisted-laser desorption ionization mass spectrometry as a complement to internal protein sequencing, in *The Protein Protocols Handbook* (Walker, J. M., ed.), Humana Press, Totowa, New Jersey, pp. 541–555.
18. Williams, K. R., LoPresti, M., and Stone, K. L. (1996) Internal protein sequencing of SDS-PAGE-separated proteins: optimization of an in gel digest protocol, in *Techniques in Protein Chemistry VIII* (Marshak, D., ed.), Academic, New York, in press.
19. Williams, K., LoPresti, M., and Stone, K. (1997) Internal protein sequencing of SDS-PAGE-separated proteins: optimization of an in gel digest protocol, in *Techniques in Protein Chemistry VIII* (Marshak, D. R., ed.), Academic, San Diego, CA, pp. 79–90.
20. Stone, K. L., LoPresti, M. B., Crawford, J. M., DeAngelis, R., and Williams, K. R. (1991) Reverse-phase HPLC separation of sub-nanomole amounts of peptides obtained from enzymatic digests, in *High Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation* (Mant, C. T. and Hodges, R. S., eds.), CRC Press, Boca Raton, FL, pp. 669–677.