Published in final edited form as:

Curr Protoc Protein Sci. 2001 May ; CHAPTER: Unit-11.10. doi:10.1002/0471140864.ps1110s08.

UNIT 11.10 N-Terminal Sequence Analysis of Proteins and Peptides

Kaye D. Speicher, Nicole Gorman, and David W. Speicher

Abstract

Automated N-terminal sequence analysis involves a series of chemical reactions that derivatize and remove one amino acid at a time from the N-terminal of purified peptides or intact proteins. At least several pmoles of a purified protein or 10 to 20 pmoles of a purified peptide with an unmodified N-terminal is required in order to obtain useful sequence information. In recent years the demand for N-terminal sequencing has decreased substantially as some applications for protein identification and characterization can now be more effectively performed using mass spectrometry. However, N-terminal sequencing remains the method of choice for verifying the N-terminal boundary of recombinant proteins, determining the N-terminal of protease-resistant domains, identifying proteins isolated from species where most of the genome has not yet been sequenced, and mapping modified or crosslinked sites in proteins that prove to be refractory to analysis by mass spectrometry.

3-7 key words for indexing

N-terminal sequencing; Edman sequencing; protein structure; blocked N-terminal; protein modifications

INTRODUCTION

Prior to the late 1970s amino-terminal (N-terminal) sequence analysis was used to identify the complete sequences of modest-sized proteins and bioactive peptides. During the 1980's and 1990's N-terminal sequencing was primarily used to determine short stretches of peptide sequences to identify proteins of interest isolated from biological samples, and if the sequence had not been determined at the cDNA level, the partial peptide sequences were used to assist in cloning the gene. Starting in the mid 1990s, the use of N-terminal sequencing began to decrease as mass spectrometry approaches were developed that were superior for most applications (see Units 16.1, 16.4, 16.10) where N-terminal sequencing was previously used. However, N-terminal sequencing remains the method of choice for several types of protein structural analysis applications, including: 1) verifying the N-terminal boundary of recombinant proteins or determining the N-terminal of protease-resistant domains, particularly when the protein or domain is greater than 40–80 kDa or can not be readily purified; 2) identifying proteins isolated from species where most of the genome has not yet been sequenced; and 3) mapping modified or crosslinked sites in proteins that prove to be refractory to analysis by mass spectrometry. The complexity of the chemical reactions involved in N-terminal sequencing (see Fig. 11.10.1) resulted in

Author for correspondence: David W. Speicher, The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, Phone 215-898-3972, Fax 215-898-0664, Speicher@wistar.org.

Contributors

Contributed by Kaye D. Speicher, Nicole Gorman, and David W. Speicher, The Wistar Institute, Philadelphia, Pennsylvania

development of complex instruments (Edman N-terminal sequencers) that process the protein or peptide in an automated manner. In order to undergo this cyclic process, an unmodified α -amino group is required at the N-terminal end of the molecule. After modification with phenylisothiocyanate (PITC), the derivatized terminal amino acid is removed by acid cleavage as its phenylthiohydantoin (PTH) derivative and a new α -amino group on the next amino acid is now available to react with PITC. The series of sequencer reactions shown in Fig. 11.10.1 represents a sequencing cycle that results in identification of the N-terminal amino acid present on the peptide or protein at the beginning of that cycle. If each step were 100% efficient, it would be possible to sequence an entire protein in a single sequencer run. In practice, multiple factors limit the amount of sequence information that can be obtained, and only rarely is it feasible to obtain >30 residues from a single sequencer run even when the amount of available protein is not limiting. With current technology, it is fairly routine to obtain at least 20 to 25 residues of sequence from the N-terminus of proteins and large peptides in the low picomole range (<50 pmol). However, a complication when working with intact proteins is that between 50% and 80% of all proteins from eukaryotic species are estimated to have a modified (blocked) N-terminal amino group and hence cannot be sequenced directly (Brown and Roberts, 1976). Several methods are available to deblock modified N-termini (see [UNIT 11.7](#)), although these methods usually require relatively large amounts of protein and are not consistently successful, especially if the nature of the blocking group is not known.

Because most intact proteins have blocked N-termini, the preferred first step in analyzing an unknown protein is to proteolytically fragment the protein and obtain internal peptide sequences (see [Strategic Planning](#)). This multistep process usually involves separating the protein of interest on either a one-dimensional ([UNIT 10.1](#)) or two-dimensional ([UNIT 10.4](#)) SDS polyacrylamide gel. This is followed by cleavage in the gel ([UNIT 11.3](#)) with a specific protease such as trypsin or endoproteinase Lys-C ([UNIT 11.2](#)). Separation of the peptides is then accomplished by reversed-phase HPLC and the peptides are analyzed by MALDI mass spectrometry ([UNIT 11.6](#)). Finally, sequence analysis of one or more of the resulting peptides is performed (see Basic Protocol 1). Most peptides produced by this method are <30 residues in length and generally the entire peptide sequence can be determined in a single sequence run.

As noted above, currently the most common applications for protein sequence analysis include the following.

1. Verification of the N-terminal boundary of recombinant proteins, particularly large proteins (e.g., larger than 40–80 kDa) where highly accurate masses can not be obtained by ESI MS ([UNIT 23.3](#)).
2. Determining the N-terminal boundary of protease-resistant domains, particularly when the protein or domain is greater than 40–80 kDa or when the domain of interest can not be readily purified. Determining the sequence of protease-resistant domains is particularly useful for designing recombinant fragments for crystallography ([UNIT 17.4](#)). A related application is analysis of physiological proteolytic processing.
3. Identifying proteins isolated from species where most of the genome has not yet been sequenced (Beyer et al., 2002, Kaji et al., 2009). This is because automated interpretation of MS/MS spectra requires that the exact sequence of the protein being analyzed should be present in the database used ([UNIT 16.10 and 25.2](#)). In the increasingly rare cases where the target protein is not in available databases, the peptide sequence may be used either to design oligonucleotide probes or to confirm putative cDNA clones.

4. Mapping modified residues or crosslinked sites in proteins that prove to be refractory to analysis by mass spectrometry.

This unit describes the sequence analysis of protein or peptide samples in solution (see [Basic Protocol 1](#)) or bound to PVDF membranes (see [Basic Protocol 2](#)) using an Applied Biosystems Procise Sequencer. (See [SUPPLIERS APPENDIX](#) for contact information). Methods are provided for optimizing separation of PTH amino acid derivatives on Applied Biosystems instruments (see Support Protocol 1). The data obtained from a single sequencer run is complex, and careful interpretation of this data by an experienced scientist familiar with the current operation performance of the instrument used for this analysis is critically important. A discussion of data interpretation is therefore provided (see Support Protocol 2). Finally, discussion of optimization of sequencer performance as well as possible solutions to frequently encountered problems is included (see Support Protocol 3).

STRATEGIC PLANNING

Instrumentation

Early N-terminal protein/peptide sequencers had micromole level sensitivity. Subsequent improvements in instrument design and introduction of HPLC for separation and quantitation of PTH-amino acids increased sensitivity to the low picomole level. The prototype picomole level sequencer was the Model 470A gas-phase sequencer manufactured by Applied Biosystems in the early 1980s (Hewick et al., 1981). The gas phase designation indicates the mode of TFA (cleavage acid) delivery. Subsequently, several companies manufactured competitive N-terminal sequences, but reduced demand for instruments resulted in phasing out alternatives to the Applied Biosystems instruments. The most recent Applied Biosystems instruments are the Procise Models 491, 492, and 494, with one, two, or four sample cartridges respectively. These systems use liquid phase TFA delivery and include a conventional narrowbore HPLC for PTH-amino acid analysis. High sensitivity versions of the Procise Models have a cLC designation and include a smaller scale reaction cartridge and a capillary HPLC for higher sensitivity (mid femtomole level PTH amino acid detection capacity). Nearly all the instruments currently in use are Applied Biosystems Procise sequencers, and therefore this unit will focus on this instrument family.

The availability of multiple sample cartridges allows automated tandem analysis of multiple sequence samples, which can substantially increase sample throughput and flexibility. Applied Biosystems has sample cartridges designed for liquid samples (Fig. 11.10.2) and proteins bound to PVDF membranes (Sheer et al., 1991; Reim and Speicher, 1994). Automated sequence analysis involves integration and coordination of three components: a computer controller, the sequencer module, and a dedicated in-line HPLC for analysis of the PTH amino acid derivatives obtained from each sequencer cycle (PTH analyzer). The computer controls the overall operation of all components and handles data storage and analysis from the PTH amino acid analyzer. Figure 11.10.3, which is a schematic of the reagents, solvents, and flow paths for a Applied Biosystems Procise Model 494 sequencer (four sample cartridges) highlights the complexity of the instrument and the three major stages of sequence analysis: the Edman chemistry conducted on the sample support (sequencer reactions); sample conversion to a stable PTH derivative (flask reactions); and HPLC analysis (PTH analyzer). Each of these three stages takes a similar amount of time (~30 to 40 min), and to minimize total sample-analysis times, each of these three stages is performed in parallel. Therefore, while the first residue is being separated on the PTH analyzer, the second residue is being converted in the flask, and derivatization and cleavage of the third residue is occurring on the sample column.

Sample Preparation

To ensure the success of a sequencing experiment, the following considerations must be taken into account.

Should internal peptides or intact protein be sequenced when the goal is to identify an unknown?—For applications where knowledge of the N-terminal is essential, one has the choice of either attempting N-terminal sequencing of the intact protein or cleaving the protein into peptides to determine one or more internal sequences. One consideration is that sequencing the N-terminal of the intact protein, which is often done after electrotransferring the protein from a 1D or 2D gel to a PVDF membrane, typically requires about 10-fold less protein than sequencing of internal peptides. That is, for an optimally operated instrument useful sequence information can usually be obtained from an intact protein starting with a few pmoles (conventional HPLC system) to high femtomole amounts (capillary HPLC system). However, as noted above, between 50% and 80% of all proteins from eukaryotes are naturally blocked and cannot be sequenced directly. In contrast, it is very challenging to produce, isolate, and sequence proteolytic fragments when less than 10- to 20-pmol of the protein of interest is analyzed. This detection limitation is one of the reasons why LC-MS/MS analysis of tryptic peptides has replaced N-terminal sequencing for routine protein identifications when the genome of the species being studied is well represented in sequence databases. This also means that if one can not feasibly isolate low pmole amounts of proteins from species poorly represented in sequence data, the only practical alternative may be to use LC-MS/MS and manual de novo interpretation of MS/MS data (*UNIT 16.11*) despite the limitations of this approach. If low pmole amounts of the protein of interest can be isolated, then N-terminal sequencing of intact proteins from prokaryotes, and internal sequencing of proteins from eukaryotes provide the best probability of successfully obtaining the needed sequence information with the minimum investment of sample, effort, and expense.

Matching biological function with the right protein (the right band on a gel)—

The most challenging protein isolation problems are those that arise where only a very limited amount of a protein can be isolated from a species poorly represented in sequence databases. Typically, a biological event or process is being studied and the goal is to identify a specific protein or proteins associated with a particular biological observation. In addition to isolating sufficient quantities of what is often a relatively low-abundance protein within a cell line or tissue, it is critical to ensure that the correct protein is isolated and sequenced. Although this may seem obvious, a substantial portion of samples submitted to sequencing laboratories actually are contaminants and the desired protein is below the detection limit used to visualize the protein of interest. Three helpful steps to ensure that the right protein is being sequenced are: (1) make an order of magnitude estimate of the amount of protein expected, (2) consider likely contaminants, and (3) run appropriate controls to test for likely contaminants.

Order of magnitude estimate: An order of magnitude estimate should be made to determine how much protein can feasibly be isolated from a given amount of source material. Even if little is known about the target protein, making a rough estimate about the protein's likely abundance may avoid effort wasted in attempting to isolate the protein from a source that can not possibly contain enough of the desired protein. Similarly, if the protein yields during purification dramatically exceed expectations, the observed protein band is probably a contaminant rather than the target protein.

As noted above, for those laboratories that have optimized the sensitivity of their procedures, the minimum amount of protein required for obtaining some sequence

information from internal peptide sequencing is ~10 to 20 pmol. Although not every sequencing laboratory can analyze samples at this level, a 20-pmol minimum value will be assumed in this unit; the minimum value should be adjusted accordingly depending upon the anticipated sequencing sensitivity of the laboratory that will be used for the sequence analysis. There are 6×10^{23} molecules/mol or 6×10^{11} molecules/pmol $\times$ 10 pmol desired = 6×10^{12} molecules of purified protein required. To obtain the number of molecules that must be present in the initial sample, this number is divided by an estimated overall purification yield—e.g., 0.25 (25% yield) for a high-yield purification involving only a few efficient steps such as an antibody affinity column followed by SDS-PAGE and electroblotting (lower overall yields would be likely for more complex purifications). With a 25% overall purification yield, 4.8×10^{13} molecules of the desired protein must be in the starting tissue. If the protein is to be purified from tissue culture cells, it is simple to estimate the number of cells required if some idea of the copy number per cell can be made (Table 11.10.1). These estimates illustrate that in most cases it is only feasible to use tissue culture cells as source material for major cellular proteins (>100,000 copies/cell). Purification of low-abundance proteins (100 to 1000 copies/cell) usually requires a large amount of a solid tissue and additional purification steps, which result in lower overall recoveries and further increase the required amount of starting sample.

Contaminants and controls: The four most likely groups of protein contaminants are: (1) major cell proteins; (2) proteins used as reagents in the purification, such as antibodies; (3) proteins from culture media or serum proteins; and (4) keratins shed from skin and hair into the environment or transferred directly to the sample by contact with exposed skin. Residual trace contamination with an abundant cell protein such as myosin or actin in a purified fraction often greatly exceeds the amount of a lower-abundance target protein that has been recovered in high yield. A number of commonly observed protein contaminants are listed in Table 11.10.2. The best guard against focusing on a contaminant protein during isolation of a low-abundance protein is to analyze a parallel control sample that has been exposed to the purification conditions used, but which lacks the target protein. For example, if a cell extract is purified on a monoclonal antibody column, two useful controls can be run in parallel with the protein eluted from the specific antibody column. First, pre-elute the specific antibody column prior to the preparative purification with the same elution conditions as used to elute the target protein. Second, one can pass the cell extract through an unrelated monoclonal antibody column prior to the specific column and elute the nonspecific column using the same elution conditions. The controls and purified protein can then be analyzed in parallel on an SDS gel.

Avoiding chemical modification of the protein—It is especially critical to avoid the use of any reagent during the purification that might react with amino groups if the N-terminal amino group is not naturally blocked and sequencing of the intact protein is planned (e.g., to map a protease cleavage site or if the N-terminus is known to be unblocked). However, even if the planned strategy is to immediately pursue internal peptide sequences, it is very advisable to avoid any chemical modification of reactive groups on the protein. Modification of side-chain amino groups on lysines will interfere with trypsin or endoproteinase Lys-C digestion, and modification of *any* reactive groups on the protein will introduce heterogeneity that may reduce yields during protein purification steps. This microheterogeneity will reduce the yields of specific peptides and increase the complexity of HPLC peptide maps, as the different chemical forms of each peptide may be separated from each other at the reversed-phase HPLC step.

The reagent most frequently used in protein purifications that leads to protein modification is urea (also see [APPENDIX 3A](#)). Although urea itself is uncharged and does not directly react with proteins, it decomposes rapidly to form cyanate, which reacts with amino groups.

If possible, urea should be eliminated from the purification procedure. If its use cannot be eliminated, the following precautions should be taken: (1) purchase an “ultrapure” grade of urea (the dry powder is reasonably stable at room temperature; solutions are *not* stable); (2) prepare urea-containing solutions using ultrapure urea immediately before use and keep the temperature as low as possible when preparing and using the solution; (3) include a scavenger in the urea solution that will react with cyanate as it forms, e.g., 200 mM Tris-Cl or 50 mM glycine; (4) keep the pH at 7.0 or lower (if feasible), as amino groups are ~10-fold more reactive at pH 8.0 than at pH 7.0 and cyanate is less stable at acidic pHs. For example, it should be safe to expose a protein to a 7 M urea solution containing 50 mM glycine at pH 7.0 for at least 24 hr at 4°C or for several hours at room temperature. Note that an appropriate buffer such as sodium phosphate should be included in this solution, as glycine does not have much buffering capacity at pH 7.

Avoiding adsorptive losses—It is well known that small amounts of proteins and peptides adsorb to glass and plastic surfaces and various strategies are commonly employed to minimize such losses. Adsorptive losses are especially problematic in late stages of purifications where low pmol amounts (low ug amounts) of proteins are being isolated for sequence analysis. Addition of a blocking protein such as 0.1% BSA to solutions is not feasible, as minor contaminants and cross-linked forms of the blocking protein are almost certain to interfere with sequence analysis of the target protein even if the target protein has a very different molecular weight and the last step of the isolation involves a 1D gel.

The severity of protein adsorptive losses is often underestimated. As shown in Table 11.10.3, appreciable protein adsorption to both glass and plastic tubes occurs within seconds. The adsorbed protein is usually tightly bound to the container surface and can only be removed with very harsh conditions, e.g., 1% SDS. Because the values in Table 11.10.3 represent brief exposure of a dilute protein to a single tube, the actual adsorptive losses that can occur during late stages of a purification would be expected to be far higher than those shown here. The best method of minimizing such losses is to add a detergent, preferably SDS, to the protein sample in late purification steps; this is likely to be especially important for any step where the total protein concentration of the sample is <0.1 µg/µl. As discussed below, the final purification step should be an SDS gel, therefore the addition of SDS to the sample during late steps of most purifications should be feasible. Of course, biological activity will be irreversibly lost in most cases upon addition of SDS, and if maintaining biological activity is critical for monitoring purification steps, an alternative blocking reagent such as a nonionic detergent (e.g., Triton X-100 or Tween 20) should be considered.

Final purification step—SDS-PAGE gels are the preferred final purification step for most protein sequencing applications. SDS gels are ideally suited for handling low microgram to submicrogram amounts (low picomole levels) of protein. They have high resolving power, and most buffer contaminants that would interfere with direct sequencing or with a protease digestion are separated from the target protein in the gel. Protein losses resulting from adsorption are negligible in comparison to those incurred with virtually any other method, and SDS gels are compatible with addition of SDS or another detergent to samples in preceding purification steps to minimize adsorptive losses (see above). After SDS-PAGE, the protein can be electroblotted to a high-retention PVDF membrane (*UNIT 10.7*) for direct N-terminal sequencing (see Basic Protocols 2) or the protein can be digested with a protease directly in the gel (*UNIT 11.3*). Another advantage of isolating proteins in this manner is that the protein in the gel is in a denatured form, which facilitates proteolytic digestion.

Because SDS gels are usually used as the final purification step for sequence analysis, the protein of interest does not need to be purified to homogeneity by other methods. In some

cases, quite crude extracts have been applied to a one-dimensional gel and the protein of interest was identified or the N-terminal sequence was determined. The limitations to this approach are that: (1) the migration position of the protein of interest on the gel must be known, (2) the target protein must be resolved from other proteins in the sample by the gel system used, and (3) the target protein must be present in the sample in a sufficiently high proportion so that contaminating proteins do not cause a gel-overload condition when sufficient target protein for sequence analysis is applied to the gel. More complex samples that may contain multiple proteins at the target-protein migration position on a 1D gel can usually be adequately separated by 2D gel electrophoresis (*UNIT 10.4*) provided that the migration position of the protein of interest can be determined. If needed, the target protein spot from several replicate 2D gels can be combined and digested with trypsin. Frequently, high-abundance proteins can be isolated for internal sequencing by loading a whole-cell homogenate directly to a series of replicate two-dimensional gels. However, despite the very high resolving power of two-dimensional gels, a single spot on a two-dimensional gel may contain multiple proteins, especially when very complex samples such as whole-cell extracts are used.

A useful guideline is to purify proteins larger than ~6 kDa using SDS-PAGE as the last purification step; peptides smaller than 6 kDa from either in gel protease digests or from other sources should be purified using reversed-phase HPLC (*UNIT 11.6*). If <100 pmol of peptide is being isolated, use of a 1.0- or 2.1-mm-diameter C18 column with 0.1% TFA and acetonitrile as the reversed-phase solvents is recommended.

Amount of protein used for sequence analysis and observed sequencer yields

—Occasionally, enthusiastic sequencer manufacturers as well as research scientists claim that 1 pmol or less of a protein can be sequenced. A critical point that is often not emphasized is the cited sensitivity usually refers to the *observed* sequence yield in early sequencer cycles, not the *total* amount of protein or peptide actually used in the experiment. The average initial sequence yield for proteins or peptides (i.e., the signal observed in early sequencer cycles divided by the amount loaded into the sequencer multiplied by 100) averages ~50% with a typical range of 20% to 80% for samples in the mid pmole range. Because adsorptive losses and artifactual modification of the N-terminus are likely to increase as the quantity of sample decreases, where 1 pmol of a protein or peptide is loaded into a sequencer, a realistic initial yield would be 0.05 to 0.5 pmol. Only a few laboratories that have carefully optimized all aspects of the sequence analysis procedure can obtain significant N-terminal sequence information at this level. Use of a cLC version of a Procise sequencer is especially helpful when attempting to obtain useful sequence information from samples with initial yields below 1 to 2 pmoles.

Obtaining internal sequences involves multiple steps that are individually reasonably efficient but that have the effect of further reducing the observed sequence yield relative to the starting protein. A reasonable guideline is that initial sequencing yields of internal peptides relative to the amount of a protein loaded onto a gel is ~15% (with a typical range of 5% to 30%). This means that 20 pmoles of a target protein applied to a gel will yield internal sequences with initial yields between 1 and 6 pmoles. Although these overall yields may appear to be quite poor, the yields at each step of the procedure must be quite good to achieve such overall yields. An example of yields at each step that might be obtained in a laboratory that has carefully optimized the procedure might be as follows:

- recovered in the major gel band relative to the amount applied to the top of the gel—90%
- electroblotting efficiency—70%

- trypsin digestion efficiency and recovery of cleaved peptides after in-gel digestion—70%
- recovery from the HPLC—90%
- recovery from the collection tube—80%
- initial yield in the sequencer—50%.

The cumulative yield in this example would be: $0.9 \times 0.7 \times 0.7 \times 0.9 \times 0.8 \times 0.5 = 0.16$, or 16%.

Estimating amount of protein used for sequence analysis—Estimating the amount of target protein that has been isolated and that will be used for either obtaining a N-terminal sequence or internal sequences contributes to proper interpretation of results. For example, if a protein band that has been electroblotted to a PVDF membrane does not produce a detectable sequence, knowing the amount loaded into the sequencer should indicate whether the protein was blocked or insufficient protein was present. Alternatively, a low-level sequence may arise from a minor contaminant rather than the target protein, which might be blocked. For example, if ~4 pmoles were present on a PVDF membrane that yielded a 2-pmol sequence, the observed sequence clearly represented the major component in that band. But if 40 pmoles were present on the PVDF membrane that yielded a 2-pmol sequence, the observed sequence might have arisen from a minor contaminant in the gel band and the major component has a blocked N-terminal. To help distinguish between such critically important alternative interpretations, a portion of each protein should be used for accurate quantitation prior to either N-terminal sequencing or protease digestion. Unfortunately, when only low picomole amounts of protein are available, a substantial portion of the total sample is required simply to obtain a reasonable estimate of protein concentration by amino acid analysis (AAA), which is the most reliable method of quantitating protein concentrations in solution, in gels, or on PVDF membranes (see *UNITS 3.2, 11.3 & 11.9* for AAA and related techniques). Therefore, if an unacceptable proportion of the sample would be consumed by quantitative AAA, the amount of protein on the gel or PVDF membrane can be estimated by comparing its staining intensity to the staining intensity of a series of standard protein lanes prepared using 2-fold dilutions—e.g., adjacent lanes containing 2 µg/band, 1 µg/band, 0.5 µg/band, 0.25 µg/band, and 0.125 µg/band. If the protein standards are run on the same gel as the experimental sample, one can readily estimate the amount of target protein available for sequencing to within a factor of ~2 to 4, with the important limitation that there is some variability in staining response for different proteins. However, Coomassie blue or colloidal coomassie, the preferred stains for in-gel digestions (*UNIT 11.3*), and Amido black, the preferred stain for PVDF membranes (*UNIT 11.2*) show much less protein-specific variability in staining than silver stains.

MALDI mass spectrometry—It is highly advantageous to prescreen all peptides prior to sequence analysis using MALDI mass spectrometry (*UNITS 11.6 & 16.2*) to complement peptide sequence analysis. Mass analysis of ~10 to 15 peaks from HPLC separation of an in situ trypsin digest is recommended. The longest peptides (i.e., with the largest masses) should be selected for sequencing, as longer sequences provide more informative database-search results, especially when related proteins, but not the identical protein, are in the sequence database searched. Similarly, if the protein is not known and from a poorly sequenced species, longer sequences provide a wider range of options for oligonucleotide probe design. After the sequence run is complete, comparison of the assigned sequence with the previously determined mass provides confirmation of the residue assignments and can extend interpretation of the sequencer data if most, but not all, of the residues in the peptide are clearly assigned. For example, reconciliation of the mass with the sequence can confirm

one or two tentative assignments, suggest possibilities for an unassigned residue, or indicate the presence of a post-translational modification. In theory, prescreening HPLC peptides by MALDI mass analysis can also potentially identify peptide mixtures. In practice, only some mixtures are detected by this method, because MALDI MS signals are not quantitative and some peptides suppress the signals of other peptides.

BASIC PROTOCOL 1: SEQUENCING LIQUID SAMPLES ON GLASS FIBER FILTERS

This protocol describes application of a highly purified protein or peptide in solution to an Applied Biosystems Procise sequencer. The sample should ideally be in a small volume of high-purity water or a volatile buffer/solvent.

Materials

Liquid peptide or protein sample(s) to be sequenced

Volatile solvents for reversed-phase chromatography: e.g., 0.1% trifluoroacetic acid (TFA) and acetonitrile (*UNIT 11.6*)

Methanol

Argon or nitrogen gas source

1% SDS

5% Acetonitrile/0.1% TFA

Polybrene solution (Biobrene Plus from Applied Biosystems; store in original container up to 3 months at 4°C; discard if contamination is suspected; alternatively store 30- to 100-μl aliquots in clean microcentrifuge tubes up to 1 year at -20°C)

Sequencer solvent/reagent kit (Applied Biosystems) including:

- R1 (phenylisothiocyanate)
- R2C (*n*-methylpiperidine)
- R3 (trifluoroacetic acid)
- R4A (25% trifluoroacetic acid)
- R5 (PTH sequencing standard)
- S2B (ethyl acetate)
- S3 (*n*-butyl chloride)
- S4B (20% acetonitrile)

Premix PTH analyzer kit (Applied Biosystems) including:

- HPLC Solvent A3 (3.5% tetrahydrofuran)
- HPLC Solvent B2 (isopropanol/acetonitrile)
- Premix Buffer Concentrate

1.0- or 2.1-mm-i.d. reversed-phase column (*UNIT 11.6*)

TFA-treated glass fiber filters (GFF; Applied Biosystems)

Cartridge seals (Applied Biosystems)

Applied Biosystems Procise sequencing system, e.g. Model 494 including:

- Sequencer module with glass sample cartridge blocks
- On-line PTH analyzer (dedicated HPLC and detector)
- Computer controller

Powder-free gloves

Stainless steel or Teflon-coated forceps

Additional reagents and equipment for reversed-phase purification of peptides ([UNIT 11.6](#)), concentration of proteins and microdialysis ([UNIT 4.4](#)), quantitation of proteins/peptides by amino acid analysis ([UNITS 3.2 & 11.9](#)), and spectrophotometric quantitation of protein ([UNIT 3.4](#))

NOTE—Prepare solutions with Milli-Q water or equivalent taken directly from the purification unit. Water stored for any length of time in glass or plastic containers supports microbial and algal growth, which results in high amino acid background in early sequencer cycles.

Prepare Sample

1. Prepare peptide or protein samples in a small volume (preferably <50 μ l) of high-purity water or volatile buffer/solvent.

Small amounts (<10 μ mol) of nonvolatile, non-amine reactive salts such as NaCl can usually be tolerated, but at higher levels, salts may interfere with binding of the protein or peptide to the GFF and can interfere with retention times of His and Arg in early cycles. Nonvolatile buffers can interfere with sequencer performance by altering the actual pHs achieved during the coupling (basic pH) and cleavage steps (acid pH). Any nonvolatile components or contaminants such as Tris buffer that react with amine-specific reagents are especially problematic.

Larger volumes can be reduced in a Speedvac evaporator, but this will concentrate nonvolatile impurities and it is very likely to increase adsorptive losses onto the tube holding the sample.

Purify and/or concentrate sample

2a. *If analyzing a peptide sample:* Purify the peptide on a 1- or 2.1-mm-i.d. reversed-phase column using volatile solvents. Typically 0.1% TFA and acetonitrile are used ([UNIT 11.6](#)).

2b. *If analyzing a protein sample:* Concentrate the sample to ~100 μ l or less, then desalt the sample by exhaustive microdialysis ([UNIT 3.4](#)) against water taken directly from a Milli-Q purification unit or equivalent, or by repetitive concentration on a microscale ultra-filtration device.

Precipitation with organic solvents or strong acids is typically not recommended as large sample losses can readily occur and are hard to detect. Furthermore, the resulting precipitate can be difficult to redissolve.

Estimate amount of protein or peptide sample

3a. If >100 pmol of sample is available: Use 20% of the protein or peptide sample for quantitative amino acid analysis using a high-sensitivity method such as precolumn phenylthiocarbamyl (PTC) derivatization ([UNIT 11.9](#)).

This will accurately define the amount to be loaded and detect potential contaminants in the sample that can react with amine reactive reagents.

3b. If <100 pmol of sample is available: Estimate the quantity of sample to be loaded by a non-destructive method such as absorbance at 280 nm (for proteins) or 215 nm (for peptides) as described in [UNIT 3.4](#).

Precycle glass fiber filter (GFF)

4. Sonicate cartridge blocks and cartridge seal with 1% SDS for 15 minutes. Follow with a rinse with MilliQ water and then rinse with 50% acetonitrile/0.1% TFA. Dry with a stream of argon or nitrogen while wearing powder-free gloves that have been prerinsed with Milli-Q water.

Refer to manufacturer's instructions for the Procise system for this and all subsequent steps.

Do not touch with bare hands any surfaces that are part of the sequencer chemistry flow path or that will contact the samples. These include, e.g., the glass sample cartridge, the GFF, solvent pickup tubes, and bottle seals. Also avoid exposing these surfaces to airborne contamination.

5. Place a TFA-treated GFF on the top half of the cartridge block with a pair of stainless steel or Teflon-coated forceps cleaned with methanol. Saturate filter by applying 15 μ l (for 9-mm filters) or 7 μ l (for 6-mm filters used with cLC cartridges) of Polybrene solution.

6. Dry Polybrene solution completely using a gentle argon or nitrogen stream, or air dry GFF under a piece of aluminum foil to prevent dust from settling on it.

7. Assemble the two cartridge halves using a new cartridge seal and insert the assembled unit into the sequencer. Pressure test the cartridge to detect leaks.

8. Run at least four cycles using the appropriate filter-conditioning program.

Samples can be loaded immediately after completing the conditioning cycles or the GFF can be stored in a gas-tight bottle up to 3 days. However, the increased handling involved in removing the filter from the sample cartridge and reinserting it after storage can result in an increased background in early sequencer cycles, especially for Ser and Gly.

The filter-conditioning program can also include a flask program that injects the PTH standard so that the quality of the PTH amino acids can be evaluated. Alternatively, a flask blank can be injected on the HPLC during filter-conditioning cycles.

Optionally, after the conditioning cycles have been completed and prior to loading the sample, one complete sequencing cycle can be run to evaluate any background that is contributed by the Edman sequencing chemistry.

Apply sample and perform sequence analysis

9. Remove the sample cartridge from the sequencer (wearing clean powder-free gloves), place the upper half of the sample cartridge on to the sample loading station, and apply 7 to

15 µl (depending upon filter diameter) of the sample to the GFF using a pipettor. Dry the filter using the gentle stream of argon from the loading station system, then repeat application and drying until the entire sample has been applied. Alternatively, air dry the sample between repetitive applications using an aluminum foil “tent” to prevent airborne contaminants from settling on the GFF.

Do not overload the filter with more liquid than it can hold. Most sample volumes are larger than the filter capacity, hence the need for repetitive applications. The number of applications will depend on the total volume of the sample being loaded (volumes >100 µl are quite time-consuming to load and increase the risk of contamination).

If HPLC-purified peptide samples containing <20 pmol are used for sequence analysis, addition of trifluoroacetic acid (25% final concentration) to the sample may increase the amount of sample actually transferred into the sequencer (Erdjument-Bromage et al., 1993). Estimate the sample volume by comparing the liquid height to an identical tube calibrated by marking volumes bracketing the expected volume on the outside of the tube. Add a volume of neat TFA (reagent R3 from Applied Biosystems) equal to one-third the sample volume (25% TFA final concentration). It is very important not to vortex the sample! Use a pipettor to mix the TFA/peptide solution completely. After loading the entire sample, wash the sides of the empty sample tube with 20 µl neat TFA and apply to the sample filter.

10. After the sample has been completely loaded and dried, place a new cartridge seal on the bottom glass block, reassemble the sample cartridge, and insert the assembly into the sequencer. Pressure test the cartridge assembly to detect leaks.

11. Check the levels of all needed items—e.g., sequencing reagents, PTH analyzer solvents, argon, and printer paper—to ensure that sufficient quantities are available to complete the projected number of sequence cycles. Program the computer appropriately using manufacturer’s recommendations and carry out the sequence run.

At least one PTH standard or a PTH analyzer blank run followed by a PTH standard should be injected immediately prior to each sequence run to verify retention times and to determine calibration factors for each amino acid. The number of Edman cycles that should be run depends upon the characteristics of the sample and the purpose of the analysis. For most applications, it is recommended that peptides be sequenced through their C-terminus plus two cycles (to ensure that the entire sequence has been obtained and to assess background levels of amino acids; see Support Protocol 3). If the peptide’s mass has been determined prior to the sequence run, the number of cycles to run can be estimated by: $\text{cycles} = (\text{peptide mass}/100) + 3$ (rounded up to next integer). For example, if the observed mass is 1644.5 Da, the sequencer can be set to complete 20 sample cycles. When intact proteins or large peptides are being analyzed and a maximum amount of sequence data is desired, set a number of cycles greater than the number that will be completed in an overnight run; the next morning review the data and continue the sequence until the decreasing signal-to-noise level limits sequence assignments (see Support Protocol 3). If the protein sequence is known and the purpose is to verify an intact N-terminal sequence or to define a cleavage site, 7 to 8 cycles should be sufficient in most cases. If the protein is to be searched against sequence databases to identify the protein, analysis of at least 20 cycles is recommended.

Procise Model 494 sequencers can be programmed to automatically analyze up to four samples in tandem. Therefore, care must be taken to ensure that sufficient reagents/solvents are available for the total cycles programmed and that the appropriate sequencing programs have been selected for each sample to be analyzed.

BASIC PROTOCOL 2: SEQUENCING PVDF-BOUND SAMPLES USING A BLOTT CARTRIDGE

For most applications where N-terminal sequence analysis of intact proteins or large peptides (>6 kDa) is desired, the preferred final purification method is to use one-dimensional ([UNIT 10.1](#)) or two-dimensional ([UNIT 10.4](#)) gels followed by electroblotting to a high-retention PVDF membrane ([UNIT 10.7](#)) and staining with Amido black ([UNIT 10.8](#)). For more detailed discussion of these preliminary steps, see [Strategic Planning](#). Care should be taken to use a PVDF membrane designed for protein sequencing as PVDF membranes designed for Western blotting have lower binding capacity and bind proteins with lower affinity. In contrast, sequence grade PVDF membranes have a high protein-binding capacity, bind large peptides and proteins with high affinity, and are inert to the harsh chemical environment in a protein sequencer. However, the wettability characteristics of these hydrophobic membranes are quite different from those of the hydrophilic glass fiber filters (see [Basic Protocol 1](#)). These differences have necessitated development of specific sequencer programs and sample holders for PVDF membranes (see Fig. 11.10.2). The preferred method for loading proteins onto a PVDF membrane is by electroblotting from a gel. Alternatively, dilute samples can be concentrated and/or separated from an incompatible buffer by direct adsorption to PVDF membranes. If detergents are not present in the protein solution, the protein can be adsorbed by immersing a small piece of PVDF membrane (which must be prewetted in 100% methanol and rinsed with water) into the sample solution and incubating it with agitation in a cold room for several hours to overnight. Alternatively, the sample can be loaded onto a PVDF disc in the bottom of a ProSorb device (Applied Biosystems). Directly spotting a protein sample onto a prewetted PVDF membrane is not recommended as this is an unreliable loading method because the liquid volume that can be applied to a PVDF membrane is quite small, and adsorption to the PVDF membrane competes with sample drying. After the liquid has dried, further adsorption does not occur. When the sample is rewet, either prior to loading in the sequencer or during sequence analysis, the dried protein that did not bind is usually washed from the membrane, which can result in greatly reduced sequencing yields.

Materials

PVDF-bound sample(s) to be sequenced

Methanol in wash bottle

Argon or nitrogen gas source

1% SDS

50% acetonitrile/0.1% TFA

Glass gel plate

Stainless steel scalpel and stainless steel or Teflon-coated forceps

Bath sonicator

Blott cartridge for Applied Biosystems Procise sequencer (Applied Biosystems)

Additional reagents and equipment for sequencing on glass fiber filters (see [Basic Protocol 1](#))

NOTE—Prepare solutions with Milli-Q water or equivalent taken directly from the purification unit. Water stored for any length of time in glass or plastic containers supports microbial and algal growth, which results in high amino acid background in early sequencer cycles.

Prepare sample

NOTE—Prepare cartridges first by cleaning with 1% SDS, sonicate for 15 minutes, rinse with MilliQ, rinse with 50% acetonitrile/0.1% TFA. Dry with argon

1. Using a scalpel, excise the desired stained protein band from PVDF membrane, keeping the membrane on a clean dry surface such as a glass gel plate and handling the membrane with a forceps.

The glass plate, scalpel, and forceps should be thoroughly cleaned before use with Milli-Q water and methanol. Wear powder-free gloves rinsed in Milli-Q water. Do not touch the membrane or surfaces that contact the membrane with bare hands.

Dry PVDF membranes are electrostatic and can be difficult to handle. The electrostatic nature of the membranes also attracts airborne contamination that invariably contains amino acid and protein contaminants. Special caution must be exercised to minimize contamination of the sample when working with small amounts of protein (<20 pmoles)

Two to three replicate 3-mm-wide one-dimensional gel bands on PVDF (50 to 75 mm² in area) or an equivalent area of membrane containing replicate two-dimensional gel spots can typically be loaded into a Procise cartridge.

2. While holding the membrane(s) with stainless steel or Teflon-coated forceps over a clean beaker (to prevent loss if dropped), thoroughly rinse with a stream of methanol from a wash bottle, immediately rinse with a wash bottle just filled with **fresh** MilliQ water.

PVDF membranes cannot be solvated directly with aqueous solutions and must initially be wetted in a strong organic solvent such as methanol. After the membranes are wetted with methanol in the above step, do not allow them to dry out until drying is indicated in step 3.

3. Place the membrane(s) in a clean beaker containing 5 ml Milli-Q water and sonicate 5 min in a bath sonicator. Remove the membrane(s) with forceps, rinse with methanol. **DO NOT DRY.**

Load sample and perform sequence analysis

4. Insert the wetted membrane(s) in the slot in the top half of the Blott cartridge.

Refer to the manufacturer's instructions for this and subsequent steps.

If needed, an additional piece of membrane can be placed protein-side-up within the recess on the top half of the cartridge.

5. Reassemble the Blott cartridge using a new cartridge seal and insert the assembly into the sequencer. Pressure test the cartridge assembly to detect leaks.

6. Check the levels of all needed items—e.g., sequencing reagents, HPLC solvents, argon, and printer paper to ensure that sufficient quantities are available to complete the projected number of sequence cycles. Program the computer appropriately using manufacturer's recommendations and carry out the sequence run.

At least one PTH standard or a PTH analyzer blank run followed by a PTH standard should be injected immediately prior to each sequence run to verify retention times and to determine calibration factors for each amino acid. The number of Edman cycles that should be run depends upon the purpose of the analysis. When intact proteins are sequenced and a maximum amount of sequence data is desired, set a number of cycles greater than the number that will be completed in an overnight run; the next morning review the data and continue the sequence until the decreasing signal-to-noise level limits sequence assignments (see Support Protocol 2). If the protein sequence is known and the purpose is to verify an intact N-terminal sequence or to define a cleavage site, 7 to 8 cycles should be sufficient. If the protein is to be searched against sequence databases to identify the protein, analysis of at least 20 cycles is recommended.

SUPPORT PROTOCOL 1: OPTIMIZING SEPARATION OF PTH AMINO ACIDS

Successful N-terminal sequencing is dependent on both the N-terminal sequencer and the in-line HPLC system that separates the PTH amino acids from each sequencer cycle. Both systems must be operating optimally to obtain a maximum amount of sequencing information, especially when analyzing samples in the low picomole range. The HPLC system should provide complete separation of all the commonly occurring PTH-derivative amino acids as well as sequencer reagent-related peaks. Also, retention times of all peaks should be highly consistent over the course of a sequence run. For high-sensitivity work (<10 pmoles), baseline shifts must be minimal and reagent purity must be adequate so that labile amino acid derivatives are not destroyed.

The following guidelines are for an HPLC system connected to the Applied Biosystems Procise Model 491, 492, or 494 sequencer. A typical gradient program is shown in Table 11.10.4 and a summary of typical solvent and gradient adjustments is illustrated in Table 11.10.5.

Additional Materials (also see Basic Protocol 1)

1% (v/v) acetone in Milli-Q water

1.0 M KH₂PO₄ in Milli-Q water

PTH Analyzer Standard mixture (standard mixture of PTH amino acids; Applied Biosystems)

1-liter bottles and three-valve caps (Rainin)

NOTE—Prepare solutions with Milli-Q water or equivalent taken directly from the purification unit. Water stored for any length of time in glass or plastic containers supports microbial and algal growth, which results in high amino acid background in early sequencer cycles.

1. Prepare the “A solvent” by adding 15 ml Premix Buffer Concentrate to 1 liter of solvent A3.

Unopened bottles of solvent A3 should be stored at room temperature but the Premix Buffer Concentrate should be stored at 4°C; partially used bottles of solvent A3 with additives can be stored at room temperature after displacing the air in the bottle with argon. Solvent A3 should be replaced after being in use for 7 to 10 days at room temperature. Peroxides can form with time in this solvent at room temperature and will adversely affect the yield of labile amino acids, especially lysine. The Premix Buffer Concentrate contains an ion-pairing

additive that affects the peak shape and elution position of PTH-Arg, PTH-His, and the pyridylethyl derivative of cysteine. Addition of 15 ml Premix Buffer Concentrate to 1 liter A3 should position the peak for His in front of Ala and the peak for Arg in front of Tyr but after the dehydroalanine derivative of Ser.

2. Transfer the freshly prepared A3 solvent from step 1 and the B2 solvent to two separate clean, dry 1-liter bottles, each sealed with a three-valve cap.

The caps from Rainin are more robust than those supplied by the sequencer manufacturer and are easily replaced if leaking of the argon used to pressurize the bottles occurs.

3. Attach the bottle containing the A solvent to the A pump of the sequencer and the bottle containing the B solvent to the B pump. Purge the A and B pumps with the fresh solvents to replace old solvents and to eliminate any air bubbles in the pumps or solvent supply lines.

Optimize baseline

4. Repeatedly run the (blank) gradient illustrated in Table 11.10.4 using the “Run Gradient” program, each time adjusting the composition of the A solvent to render the baseline as flat as possible (steps 6, 7, and 8). Each time adjustments are made to the solvent composition, purge the pump with the adjusted solvent and run several blank gradients to observe the effects of the additions on the baseline.

5. Flatten the baseline rise typically observed in the latter part of the gradient by adding acetone in Milli-Q water to the A solvent.

Small amounts of 1% acetone in 100- μ l increments can be added to the A solvent to produce an absorbance match between the A and B solvents. This results in a flat baseline when a linear gradient is used.

6. Correct early negative baseline slope by adding 1.0 M KH_2PO_4 to the A solvent.

Frequently, a negative slope is observed in the beginning of the chromatogram between the DTT and Glu peaks (see Fig. 11.10.4). Addition of up to 1 ml of 1.0 M KH_2PO_4 in 10- μ l increments will flatten the baseline in this area.

7. For high-sensitivity sequence analysis (<10 pmoles), continue to adjust the A solvent as necessary with 1% acetone and 1.0 M KH_2PO_4 until the rise in baseline is no more than 0.1 mAU between 0 to 21 min.

Optimize positions of PTH amino acid peaks

8. After the baseline has been optimized, inject the PTH Analyzer Standard mixture by running the “PTH-Standards” program and evaluate the separation. If needed, repeat the run adjusting the composition of the A solvent to optimize the separation (steps 10 and 11). Each time adjustments are made to the solvent composition, purge the pump with the adjusted solvent and carry out another standard run to observe the effects of the additions on the chromatogram.

The chromatogram illustrated in Figure 11.10.4 shows good separation of the common PTH amino acids and the Edman degradation byproducts DMPTU, DPTU, and DPU.

9. Add additional Premix Buffer Concentrate to the A solvent to move the His and Arg peaks earlier in the chromatogram if needed.

Aging of the PTH analyzer column may necessitate increasing the amount of Premix Buffer Concentrate in order to maintain the position of His and Arg. Add Premix Buffer Concentrate to solvent A in 2-ml increments to position His before Ala and Arg before Tyr. The ability of Premix Buffer Concentrate to shift His and Arg retention times is lost if >25 ml are added to the A solvent. In this case, the A solvent can be prepared again as in step 1, but with the amount of Premix Buffer Concentrate added to a liter of A3 reduced to 12 ml, which should position His after Ala and Arg after Tyr. Alternatively, the PTH analyzer column can be replaced with a new one.

10. If further optimization of the separation is necessary for Ile, Lys, and Leu, reposition these peaks by making changes to the gradient as noted in Table 11.10.5.

Ideally, Lys should be centered between Ile and Leu. Consult the manufacturer's instructions if any other separation problems are observed.

SUPPORT PROTOCOL 2: SEQUENCE DATA INTERPRETATION

Introduction

After a sequencer run is completed, the next step is to assign a sequence by analyzing the data from all cycles. Although assigning some sequences can be straightforward – e.g., from a short, pure peptide at the 100-pmol level – accurate sequence assignment is often complex and usually requires an experienced scientist who is familiar with the recent sequencing performance history of the instrument used. Although Applied Biosystems instruments have a software feature that automatically assigns sequences, an independent study that analyzed sequence assignment accuracy on an unknown sample showed that software assignments were much less accurate than those obtained by the average experienced sequencer operator using manual assignments (Yuksel et al., 1991). This same study showed that operators using the software-assigned sequence to assist them in their manual assignments achieved less accurate results than those operators who did not use software assignments at all. Care must be particularly exercised when interpreting the sequence of larger proteins, peptides containing contaminating (secondary) sequences, long peptide or protein sequence runs, and analyses where the signal is below the 10-pmol level.

Factors that must be considered when assigning sequence include the expected recovery of amino acids, the repetitive yield, signal carryover (lag) in subsequent cycles, and increases in background from nonspecific acid cleavage of peptide bonds. In addition, accurate assignment of early cycles in a sequence can be problematic because of high amino acid background resulting from contamination of the sample (from many sources) prior to loading into the sequencer. In general, careful manual examination of the tabulated data from all cycles of the analysis, together with a comparison of the corresponding chromatograms, will produce the most accurate sequence assignment.

Expected Recoveries of PTH Amino Acids

Many PTH amino acids are essentially quantitatively recovered during the sequencing process, whereas others are partially destroyed or incompletely extracted (see Table 11.10.6). However, the recoveries of these problematic amino acids are usually fairly consistent for a given instrument, sequencer method, and set of reagents; they should normally be within the ranges indicated in Table 11.10.6. Lower-than-normal or variable recoveries are usually indicative either of chemical modification of the sample during isolation or problematic sequencer performance (resulting from hardware problems or reagent quality). When assigning sequences, the amino acid recoveries recently observed on the instrument used must be considered, especially when analyzing low-level or difficult sequence data sets.

Effects of Repetitive Yield on Number of Cycles Assigned

The background-corrected signal observed at each cycle of a sequence relative to the background-corrected signal observed in the previous cycle is the **repetitive yield**. Repetitive yield can be most easily calculated from the slope of the best-fit line when the logarithm of the net sequence signal observed (in pmoles) is plotted versus cycle number. The size of the sequence signal always steadily decreases as a sequence run progresses, as a result of two factors. First, the efficiency of the Edman chemistry for each sequencer cycle is high but slightly less than 100%. A reasonable estimate for the overall efficiency of the reactions in each cycle when using an optimized sequencer and high-purity reagents would be ~96%. The second factor that decreases the yield is loss of sample from the sample support (sample washout). This second factor is very dependent upon the specific characteristics of individual samples, the type of sequencer support used, optimization of solvent deliveries in the sequencer, and related factors.

The relationship between repetitive yield and the amount of sequence that can be obtained is illustrated in Figure 11.10.5. In this example, an initial yield of 10 pmoles and a sequencing detection threshold of 0.5 pmoles is used. As shown, a repetitive yield of 95% would allow sequence assignment of >40 residues in this case, whereas lower repetitive yields decrease the maximum amount of sequence information that can be obtained. The observed repetitive yield for any sequencer run is related to both sample characteristics and sequencer operation. Optimization of all sequencer parameters to yield the highest possible repetitive yields, using an appropriate standard protein or peptide, will insure that the maximum amount of sequence can be assigned when sequencing experimental samples.

Carryover (Lag) Can Complicate Sequence Assignments in Later Cycles

Incomplete coupling and incomplete cleavage of the N-terminal residue during each sequencer cycle results in detection of a given amino acid residue in subsequent cycles. This effect is cumulative, so that carryover is usually, but not always, low in early cycles and progressively increases throughout the sequence. Cleavage at prolines is especially difficult, and an obvious jump in carryover at prolines is often observed. The increase in carryover in later cycles and its influence on assigning sequences in later cycles is shown in Figure 11.10.6. The alanine signal in cycle 3 is very clear, with a small amount of carryover into cycle 4. The alanine signal at cycle 13 is much smaller, and the alanine value in cycle 14 is nearly equal to the value in cycle 13. The alanine in cycle 14 results from carryover and not an Ala-Ala sequence. Cycle 14 is a threonine, as indicated by the increase in signal of this residue in this cycle. Another alanine occurs in cycle 23 and the further increase in alanine signal in cycle 24 is due to a progressive further increase in carryover. In this worse-than-average sequence, by cycles 13 and 14 carryover of the alanine and threonine signals, respectively, extend into the $n + 1$, $n + 2$, $n + 3$, and $n + 4$ cycles before the value of the amino acid returns to background levels.

Background in Early Sequencer Cycles

Assigning the first several residues in a sequence can often be difficult because of contaminants introduced during sample loading or from the sample itself, especially when the initial sequence signal is <10 pmoles. These contaminants are primarily free amino acids and small peptides from such sources as airborne contamination, ungloved fingers, and impurities in buffers used in the purification. They result in a high level of multiple amino acids (background), especially glycine and serine, in the first one or more sequencer cycles.

The first four cycles from a sequence run using 10 pmoles of β -lactoglobulin loaded to a GFF are shown in Figure 11.10.7A as an example of a very clean sample with minimal initial background. The first residue, leucine, as well as subsequent cycles can be easily

assigned. The sequence assignment for the four cycles shown is LIVT (i.e., Leu-Ile-Val-Thr); see [Table A.1A.1](#) for one-letter amino acid abbreviations). The sequence in Figure 11.10.7B is shown at the same sensitivity as the β -lactoglobulin sequence; however this experimental sample has substantial free amino acid and small-peptide contamination that results in a high background of several amino acids in early cycles. The multiple signals present in the first three cycles make assignment of sequence in these cycles difficult, but a clear glutamate (E) signal in cycle 4 and additional signals in subsequent cycles show that this protein is not blocked; hence a definitive sequence assignment from cycle 4 on was made. Very tentative assignments for cycles 2 and 3 could be made based on knowledge of how rapidly background normally declined in analogous samples in this sequencer; however, the more conservative assignment for the cycles shown would be XXXE (where X represents an unassigned residue).

Calculation of Background-Corrected Signals

In addition to early-cycle background from contaminants, most sequences exhibit a detectable level of most amino acids in most cycles of the sequence. The major source of this background is “nonspecific” acid cleavage of peptide bonds that free new internal sequences during the cleavage step of each sequencer cycle. The extent of this effect is roughly related to the total number of peptide bonds in the sample being sequenced. Hence, background is usually minimal for peptides, moderate for small proteins and increasingly higher for larger proteins. The new internal N-termini start to produce many very low-level sequences in subsequent cycles. As the degree of cleavage is extremely low and somewhat random, a fairly uniform background results, which steadily increases throughout the sequence as more peptide bonds are cleaved at each cycle. Peptide bonds involving serine, threonine, and aspartic acid are the most labile, and therefore proteins or peptides that have a high content of these residues usually show higher background levels than other proteins of the same size. Table 11.10.7 illustrates data from 10 pmoles of β -lactoglobulin loaded to a GFF and sequenced for 15 cycles with the Procise 494. The picomole values listed for each amino acid/cycle are the uncorrected or raw values automatically calculated versus the PTH standard (see Fig. 11.10.4) injected at the beginning of the run. Because β -lactoglobulin is a fairly small protein, there is detectable background for most amino acids that increases slowly throughout the sequence run, but it does not interfere with sequence assignment.

The accuracy of the reported values in this table should be verified by examining each chromatogram and correcting any problems such as incorrect peak identification or inconsistent peak integration. In some cases, it may be necessary to reintegrate and reprint the entire sequence run prior to manually computing the net or background-corrected amino acid yields of assigned residues in the sequence. After the accuracy of reported values is verified, an appropriate background level can be estimated for each observed sequence signal and subtracted from the sequence signal observed in cycle n . In general, the most appropriate background value is the amount observed in the cycle before the sequence signal ($n - 1$). For example, in Table 11.10.7 the background-corrected value for Val in cycle 3 would be 5.5 pmol ($5.95 - 0.46 = 5.49$). An alternative to evaluating a table of raw data values as shown in Table 11.10.7 is to plot the values for each residue versus cycle number—i.e., the type of plots shown in Figure 11.10.7.

Careful evaluation of net sequence yields for each cycle is especially important for complex sequence data sets and can be useful in identifying potential errors in initial sequence assignments. Especially when one or more lower-level (secondary) sequences are present in addition to a major (primary) sequence, evaluation of sequence yields minimizes the risk of confusing signals from the secondary sequence with positions in the primary sequence, where labile or post-translationally modified residues might occur.

Sequence Assignment

The simplest data set from a sequenced sample is one signal per cycle, as illustrated by the data set in Table 11.10.7. In this case the sequence is easily assigned—e.g., LIVTQ TMKGL DIQKV. Assignments are typically reported using the following conventions: uppercase letters for positive calls (confidence level should be >99%); lowercase letter or letter enclosed by parentheses for a tentative call (confidence level between 50% and 99%); and X for no assignment possible.

When two or more signals are present in each cycle from the sample sequenced and the sequence is unknown, assignment becomes more challenging. A single major sequence in the presence of one or more sequences at much lower levels (less than a third the major sequence level) can usually be assigned with substantial confidence. Sorting of multiple sequences must rely heavily on quantitative yields and must consider corrections for residues not recovered quantitatively (Table 11.10.6). In some cases, a minor sequence can also be assigned, but assignment of some residues in this sequence may be complicated by interference that is due to carryover of the major sequence.

Mass Analysis Can Aid Sequence Assignments

As briefly discussed in Strategic Planning, obtaining a mass for HPLC-purified peptides by MALDI mass spectrometry (*UNITS 11.6 & 16.2*) can aid in assigning sequences. When the entire sequence is assigned by Edman sequencing, comparing the calculated and observed masses provides a useful check that the correct sequence assignment has been made. Similarly, when one or two residues have been assigned tentatively and the remainder of the peptide sequence has been positively assigned, the confidence level of the tentative calls can often be increased or decreased by considering the observed mass. Finally, when a single residue cannot be assigned and the remainder of the sequence has been positively assigned, the difference between the observed and calculated masses will yield the mass of the unassigned residue, which frequently suggests a single likely residue for this position. Examples where this approach is useful include: an X (unassignable residue) in the first cycle resulting from initial background, a missing last residue in a tryptic peptide resulting from washout of the last residue, an unmodified cysteine, or a post-translationally modified residue not detected by Edman sequencing. Mass analysis of peptides is therefore a very valuable complementary method when used in conjunction with Edman sequencing. However, caution must be exercised when using mass analysis to assign residues where no or multiple Edman signals are present. In some cases, more than one possible interpretation of tentative or unassigned residues may fit within the error limits of the mass analysis.

SUPPORT PROTOCOL 3: OPTIMIZING SEQUENCER PERFORMANCE

Introduction

Optimal performance of a protein sequencer depends on careful and frequent evaluation of multiple parameters including proper instrument delivery of solvents and reagents, quality of sequencer reagents and solvents, sequencer performance, and yields of nonquantitatively recovered amino acids when analyzing appropriate peptide and protein standards (Tempst and Riviere, 1989; Atherton et al., 1993; Tempst et al., 1994). For additional information on problems in sequencer performance and their solutions, see [Troubleshooting](#).

Sequencer Standards

Useful protein sequencing standards include: human serum albumin, β -lactoglobulin, and horse apomyoglobin. Peptide standards include melittin or the sequencing test peptide from Sigma. A sequence of a standard sample should be evaluated at least once for every 10 to 20 experimental samples sequenced, to detect non-catastrophic sequence-performance

problems. The standard sample should be analyzed at a level similar to the average levels of experimental samples that are being analyzed.

Standard protein or peptide solutions should be prepared in adequate quantity for multiple analyses, then quantified by amino acid analysis, divided into multiple aliquots, and stored at -20°C . A fresh aliquot can be used for every third standard run to minimize possible variation resulting from degradation or contamination of the aliquot during storage and handling. It is good practice to tabulate results from all sequences for each type of standard in an Excel spreadsheet that includes such items as background corrected yield for each cycle, repetitive yield, and carryover at a specific reference cycle. Sequences that exhibit optimal initial yields, maximum repetitive yields, low carryover, and good recoveries of all amino acids can then be used as reference point for evaluating sequencer performance in future sequence runs.

Proper Delivery of Solvents and Reagents During Sequencer Cycles

The manufacturer's guidelines for evaluating and adjusting deliveries of all solvents and reagents should be used and direct observations of all solvent and reagent deliveries for a complete Edman cycle (both reaction and conversion portions of the cycle) should be made every three to six sequencer runs. These observations should be kept in a log, which will be useful for identifying trends, such as a gradual reduction in wash solvent volume delivery, before sequencer performance is significantly affected. If significant adjustments to delivery volumes and pressures need to be made, sequencer performance should be evaluated using an appropriate peptide or protein standard prior to sequencing experimental samples.

Quality of Solvents and Reagents

All solvents and reagents should be marked with the date of receipt prior to being stored according to the manufacturer's recommendations. Solvents and reagents should be free of any kind of particulate matter, haze, or cloudiness. The lot numbers should be changed in the sequencer-bottle menus whenever a bottle is refilled or replaced on the sequencer. The actual date that the bottle is placed on the sequencer should also be written on the label of the bottle and the appropriate information should be recorded in the sequencer log. When analyzing samples at the $<10\text{-pmol}$ level, many changes in sequencer performance that are not due to hardware problems can quickly be correlated with a lot change for a specific sequencer reagent by careful tracking of reagent lot numbers and installation dates.

In general, dramatic changes in sequencer performance are related to hardware failures or a marked decrease in the quality of the most recently changed solvent or reagent.

COMMENTARY

Background Information

The chemical reactions used in sequencers involve the cyclic removal of the N-terminal residue from a protein or peptide after modification with phenylisothiocyanate (PITC). This basic chemistry remains largely unchanged since it was first described in 1949 by Pehr Edman (Edman, 1949). However, the amount of protein required for routine sequence analysis has steadily declined from micromole quantities using early automated sequencers (Edman and Begg, 1967) to the current low pmol level. Although most sequencing laboratories can effectively sequence samples at the 100-pmol level, substantial optimization of instrumentation and associated methods must be invested to ensure that a sequencing laboratory can routinely obtain sequence data at the 10-pmol or lower level.

In general, N-terminal sequence analyses of intact proteins and large peptide fragments are most effectively performed by electroblotting the sample from an appropriate one- or two-dimensional gel onto a high-retention PVDF membrane ([UNIT 10.7](#)). Electrotransfer efficiencies of 50% to 80% can usually be obtained for most proteins, and total losses including artifactual blocking of free N-terminals and transfer losses are usually <35% even at the <10-pmol level when appropriate precautions are taken (Mozdzanowski and Speicher, 1992; Speicher, 1994). Substantial N-terminal sequence information can be obtained from an electroblotted protein when as little as 2 to 5 pmol of protein is present on the blot and extensive blocking of the N-terminus has been avoided during sample preparation. When 2 pmoles of a desired unblocked protein is electroblotted onto a PVDF membrane, the initial signal observed in the sequencer is expected to be between 0.4 and 1.4 pmoles. As noted above, sequence analysis at this level is currently feasible, but certainly not routine. In some cases, it may be preferable to sequence a soluble, highly purified protein that is already in a solution compatible with direct sequence analysis (see [Basic Protocol 1](#)) instead of using SDS gels and electroblotting.

Because 50% to 80% of proteins from eukaryotic sources are physiologically blocked (Brown and Roberts, 1976), the preferred strategy for obtaining partial sequence data from an unknown protein is to immediately attempt internal sequencing rather than N-terminal sequencing of the intact protein. This approach involves four sequential steps: in situ digestion with a specific protease such as trypsin either in solution or in a gel slice, microbore or narrow-bore reversed-phase HPLC separation of the tryptic peptides, MALDI mass analysis of selected peak fractions, and finally N-terminal sequencing of one or more of the largest peptides as indicated by the mass analysis.

Starting in the 1990's protein identification by LC-MS/MS analysis began to emerge as the preferred method of identifying unknown proteins isolated from species where the genome had been largely determined. In this method, peptides from in-solution or in-gel digestion of a protein band using trypsin is followed by separation of peptides by nanocapillary HPLC interfaced directly with a high sensitivity mass spectrometer capable of tandem MS analysis (Units 16.10 and 23.1). Protein identifications can be performed at far higher sensitivity, many proteins in complex protein mixtures can be identified without purifying the individual proteins, and analyses are far faster and less costly than Edman sequencing. However, if the exact peptide sequence that produces a given MS/MS spectra is not in the database, a correct identification will not result, at least for that peptide.

Critical Parameters

When isolating proteins from natural sources where the genome is largely undetermined, the most difficult part of the project is usually the isolation of a sufficient amount of the protein of interest in a form that is free of contaminating proteins, amino acids, and incompatible inorganic compounds. Purifying relatively low-abundance proteins from natural sources is particularly challenging. In such situations it is advisable to keep the purification scheme to the fewest possible steps and to use either one-dimensional SDS gels or two-dimensional gels for the final purification step. Detailed guidelines for sample isolation are described above (see [Strategic Planning](#)).

Troubleshooting

The first step in troubleshooting sequencer problems is to distinguish between sequence analysis problems and problems arising from the sample – i.e., sample heterogeneity, blocked N-terminal sequence, or insufficient sample loaded into the sequencer. Suspected sequence analysis problems can be readily evaluated by immediately analyzing either a

protein or peptide standard, depending upon the nature of the experimental sample where a problem was encountered.

Table 11.10.8 lists some of the more common problems encountered in sequencer operation and optimization; Table 11.10.9 lists common problems for the HPLC analyzer. In general, dramatic changes in sequencer performance are related to hardware failures or a marked decrease in the quality of the most recently changed solvent or reagent.

Anticipated Results

When the N-terminus of the protein or peptide loaded to the sequencer is not blocked, an initial sequence yield should be observed equal to ~20% to 80% of the amount of sample loaded into the sequencer. The first several cycles of any sequence may have signals from multiple amino acids, which can complicate making positive identifications for these cycles; however, tentative assignments can often be confirmed if the mass of the peptide has been determined prior to sequence analysis. The high initial background, if observed, should disappear by cycle 2 or 3 unless contamination is extremely severe. Some peptide and protein samples will contain minor contaminating sequences, and these secondary sequences can often be distinguished from the major sequence if a large quantitative difference exists. However, caution must be exercised in this case, as the repetitive yields of the major and minor sequences may be different because of varying degrees of sample washout. For example, a short major peptide may exhibit a lower repetitive yield than a minor longer peptide, and similar signals might be observed near the end of the major sequence, resulting in potential misassignment to the major sequence of subsequent residues in the minor sequence. When homogeneous peptides are being sequenced, the entire sequence can usually be assigned unless the peptide exhibits unusually high washout. When proteins or large peptides (>6 kDa) are being sequenced, several factors can limit the amount of sequence that can be obtained, including repetitive yield, progressive increase in carryover as the sequence proceeds, and increased background from nonspecific acid cleavage. Typically, most proteins exhibit high repetitive yields (>92%); hence repetitive yield is rarely the limiting factor when sequencing proteins except when the sequencer is not optimized or the initial sequence yield is near the detection limit (usually 0.1 to 1 pmol). As a general rule, when sequencing proteins >100 kDa, increased background is likely to be the most important limiting factor. For proteins <100 kDa, the limiting factor is usually the rising carryover.

Time Considerations

The time required for preparation of the sequencer and HPLC prior to loading the samples can range from ~5 min (if sufficient sequencer reagents/solvents and HPLC solvents are on the instrument for the projected number of cycles) to up to 2 hr when multiple reagent replacements are needed. Loading PVDF-bound samples requires ~15 min, which includes excising the sample from the membrane(s), washing it, drying it, and loading it into a clean Blott cartridge. Loading liquid samples using a glass fiber filter (GFF) sample support requires the most time. It takes ~15 min to load Polybrene on a GFF and set up four filter precycles that take ~35 min each to precondition the GFF on the sequencer prior to sample application. The time required for sample application will vary widely depending upon the volume to be loaded. It takes ~5 min to apply and dry 15 μ l of sample on a preconditioned 9-mm GFF. Loading large volumes using repetitive application can be quite time-consuming; for example, it would take about 1 hr to load 150 μ l in 15- μ l aliquots.

After the run is complete, the time required for analyzing a sequence is related to the length of the sequence and the complexity of the data. A straightforward 20-cycle peptide sequence may require as little as 30 min to analyze. In contrast a very complex 20-cycle multiple-

sequence data set with integration problems may require at least 2 hr for careful data analysis. It is also advisable for two experienced sequencer operators to independently assign very complex sequences, to improve the accuracy of the assignments.

Literature Cited

- Atherton, D.; Fernandez, J.; DeMott, M.; Andrews, L.; Mische, SM. Routine protein sequence analysis below ten picomoles: One sequencing facility's approach. In: Angelletti, R., editor. *Techniques in Protein Chemistry IV*. Academic Press; San Diego: 1993. p. 409-418.
- Beyer K, Bardina L, Grishina G, Sampson HA. Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: see storage proteins as common food antigens. *J Allergy Clin Immunol*. 2002; 110:154-159. [PubMed: 12110835]
- Brown JL, Roberts WK. Evidence that ~80% of the soluble proteins from Ehrlich ascites cells are N-alpha acetylated. *J Biol Chem*. 1976; 251:1009-1014. [PubMed: 1249063]
- Edman P. A method for the determination of the amino acid sequence in peptides. *Arch Biochem Biophys*. 1949; 22:475-480.
- Edman P, Begg G. A protein sequenator. *Eur J Biochem*. 1967; 1:80-91. [PubMed: 6059350]
- Erdjument-Bromage, H.; Geromanos, S.; Chodera, A.; Tempst, P. Successful peptide sequencing with femtomole level PTH-analysis: A commentary. In: Angelletti, R., editor. *Techniques in Protein Chemistry IV*. Academic Press; San Diego: 1993. p. 419-426.
- Hewick RM, Hunkapiller MW, Hood LE, Dryer WJ. A gas-liquid solid phase peptide and protein sequencer. *J Biol Chem*. 1981; 256:7990-7997. [PubMed: 7263636]
- Kaju K, Tomino S, Asano T. A serine protease in the midgut of the silkworm, *Bombyx mori*: Protein sequencing, identification of cDNA, demonstration of its synthesis as zymogen form and activation during midgut remodeling. *Insect Biochem Mol Biol*. 2009; 39:207-217. [PubMed: 19114104]
- Mozdzanowski J, Speicher DW. Microsequence analysis of electroblotted proteins I. Comparison of electroblotting recoveries using different types of PVDF membranes. *Anal Biochem*. 1992; 207:11-18. [PubMed: 1489083]
- Reim DF, Speicher DW. A method for high-performance sequence analysis using polyvinylidene difluoride membranes with a biphasic reaction column sequencer. *Anal Biochem*. 1994; 216:213-222. [PubMed: 8135354]
- Sheer DG, Yuen S, Wong J, Wasson J, Yuan PM. A modified reaction cartridge for direct sequencing on polymeric membranes. *Biotechniques*. 1991; 11:526-534. [PubMed: 1793588]
- Speicher DW. Methods and strategies for the sequence analysis of proteins on PVDF membranes. *Methods*. 1994; 6:262-273.
- Tempst P, Riviere L. Examination of automated polypeptide sequencing using standard phenylisothiocyanate reagent and subpicomole high-performance liquid chromatographic analysis. *Anal Biochem*. 1989; 183:290-300. [PubMed: 2624317]
- Tempst P, Geromanos S, Elicone C, Erdjument-Bromage H. Improvements in microsequencer performance for low picomole sequence analysis. *Methods*. 1994; 6:248-261.
- Yuksel, KU.; Grant, GA.; Mende-Muller, L.; Niece, RL.; Williams, KR.; Speicher, DW. Protein sequencing from polyvinylidene fluoride membranes: Design and characteristics of a test sample (ABRF-90SEQ) and evaluation of results. In: Villafranca, JJ., editor. *Techniques in Protein Chemistry II*. Academic Press; San Diego: 1991. p. 151-162.

Edman chemistry and automated N-terminal sequence analysis. Modification of the free N-terminus of a protein or peptide sample with phenylisothiocyanate (PITC) at high pH, followed by acid cleavage of the modified terminal residue, results in release of a peptide one residue shorter in length and with a free N-terminus. This process represents a single sequencer cycle; subsequent residues can be removed by repeating this series of reactions (sequencer reactions). The cleaved 2-anilino-5-thiazolinone derivative (ATZ amino acid) is extracted from the sample support and transferred to a “flask” for conversion to the phenylthiohydantoin (PTH) amino acid (conversion reactions). The sample is then injected onto an HPLC column connected in line to the sequencer for separation analysis (PTH amino acid analysis).

Figure 11.10.2. (modified to match modified figure) Sequencer sample cartridges. **(A)** Glass fiber filter (GFF) sample cartridge for the Applied Biosystems Procise sequencer, used to analyze samples in solution. The sample is loaded onto a Polybrene-treated GFF, inserted into the top reaction cartridge block, and the two halves of the cartridge are sealed with a Zitex (Teflon) cartridge seal. **(B)** Blott cartridge for an Applied Biosystems Procise sequencer, used to analyze proteins bound to PVDF membranes, which are inserted into the semicircular slot in the top of the Blott cartridge.

Figure 11.10.3.

Applied Biosystems Procise 494 reagent schematic, illustrating the current complexity of instrumentation used for automated sequence analysis. Bottles for chemistry involved in sequencer reactions include: base (R2) and PITC (R1); trifluoroacetic acid (R3) for cleavage, which can be delivered in either the gas phase or as a small pulse of liquid reagent; solvents for removing reaction byproducts (S1 and S2) and extraction of the cleaved amino acid derivative to the conversion flask (S3); and additional reagent positions for addition of optional chemistries or solvents (X1 and X3). The conversion reactions section includes: aqueous acid (R4) for conversion of the amino acid derivative to PTH amino acids; acetonitrile/water (S4) to redissolve the PTH amino acids for HPLC analysis; PTH standard (R5) used for calibration; and additional positions for user-designated chemistries or solvents (X2 and X3). Delivery of appropriate chemicals or solvents to the sample cartridges

or conversion flask are regulated by the related reagent or solvent valve blocks. The HPLC injector transfers the PTH amino acids from the conversion flask to an HPLC column. The HPLC pumps and detector are controlled through the sequencer computer during the sequencing process in the automated mode. Adapted with permission from Applied Biosystems Procise User's Manual.

Figure 11.10.4.

Chromatogram of PTH standard, illustrating an example of a typical separation of PTH amino acids and Edman degradation byproducts for the Applied Biosystems Procise Models 491, 492, and 494 sequencers. The common amino acids recovered during the sequencing process are well separated, as are the normal byproducts of the Edman process: DMPTU (dimethylphenylthiourea), DPTU (diphenylthiourea), and DPU (diphenylurea). Unmodified cysteine is completely destroyed. Dithiothreitol (DTT) is typically present in the R4 reagent. S' is an adduct between DTT and PTH-dehydroalanine, the latter of which is a degradation product of PTH-serine. Unmodified cysteine also forms S', and can be distinguished from a serine residue because only the S' peak increases in cycles containing unmodified cysteine. However, assigning cysteine based solely on this minor, often somewhat variable peak is risky. Abbreviations: mAU, milli-absorbance units; see [Table A.1A.1](#) for one-letter amino acid abbreviations.

Figure 11.10.5.

Effects of repetitive yield on number of cycles assigned. Theoretical amino acid yields for several different repetitive yields that are frequently encountered in automated sequence analysis illustrate the importance of this parameter on the amount of sequence that can be assigned. An initial yield of 10 pmol and a sequence detection limit of 0.5 pmol were used for this example.

Figure 11.10.6.

Yields of alanine and threonine from a sequence exhibiting somewhat higher-than-normal carryover. Uncorrected values of Ala (crosses) and Thr (triangles) in each cycle of a sequence are shown. The amount of carryover in this sequence was larger than the average for this instrument, but within the range of carryover seen with some samples. The high carryover in later cycles complicates sequence interpretation. This sequence contained Ala at cycles 3, 13, and 23, and Thr at cycle 14.

Figure 11.10.7.

Chromatograms of the first four cycles from two different low-pmol-level sequences. **(A)** 10 pmol β -lactoglobulin were loaded onto a Procise 494. The first four residues can be clearly identified above a moderate initial background. **(B)** A tryptic peptide loaded to the same sequencer. In this case, the high background starting in cycle 1 interferes with sequence assignment of early cycles and the first unambiguous assignment that can be made is E, in cycle 4. Abbreviations: mAU, milli-absorbance units; see [Table A.1A.1](#) for one-letter amino acid abbreviations. All other tables unchanged from original 1997 pub.

Table 11.10.1

Estimating Cell Amounts Needed To Obtain Sequence Information^a

Copies/cell	Cells needed	Comments
10 ⁶	5 × 10 ⁷	Only the most abundant proteins including major cytoskeletal and structural proteins
10 ⁵	5 × 10 ⁸	Abundant proteins usually easy to detect on 2-dimensional gels of whole-cell homogenates
10 ⁴	5 × 10 ⁹	Generally below the detection limit on 2-dimensional gels of whole-cell homogenates
10 ³ –10 ²	Not practical	Examples include many enzymes and regulatory proteins

^aBased on a 25% overall purification yield and analysis by a sequencing laboratory capable of obtaining internal sequence information on 20 pmol of a target protein.

Table 11.10.2
Probable Protein Contaminants When Purifying Low-Abundance Proteins

Mol. wt. (kDa) ^a	Probable contaminants
200	Myosin
68	Serum albumin from source tissue or serum in tissue culture media
50–60	Intermediate filament proteins from source cells or human skin tissue keratins from airborne or direct-contact contamination
55	Antibody heavy chain ^b
43	Actin (usually the most abundant cell protein)
25–28	Antibody light chain ^b
Varied	Any added protein including enzymes such as DNase, RNase, soybean trypsin inhibitor, pancreatic trypsin inhibitor, or other protein-type protease inhibitors

^a Approximate apparent size on an SDS gel run using reducing conditions.

^b Definite major contaminant in immunoprecipitates and a likely contaminant when using an immunoaffinity column (column bleed).

Table 11.10.3

Adsorptive Losses of Proteins from Dilute Solutions^a

Protein	Glass test tube		Siliconized glass test tube		Polypropylene microcentrifuge tube	
	µg	pmol	µg	pmol	µg	pmol
Phosphorylase B	0.4	4.3	0.7	7.5	0.6	6.4
BSA	1.2	18	2.0	30	0.2	3.0
Ovalbumin	0.9	21	1.3	30	0.3	6.9
Myoglobin	1.9	112	2.5	147	1.0	58

^a Amount of protein lost by adsorption to the indicated container surface within 15 sec of addition of a 300-µl aliquot of dilute protein solutions in phosphate buffer, pH 7. Samples were pipetted directly into the bottoms of the tubes to minimize exposure to the container surfaces. Losses were determined by quantitative amino acid analysis of the protein left in solution. Protein adsorbed to the container surface could be extracted using 1% SDS. Adsorptive losses were higher than the values shown here when the solution was vortexed or when protein solutions were applied to the top of the tube and allowed to run down the side.

Table 11.10.4

Gradient for Separation of PTH amino acids^a

Step no.	Time (min)	%B
1	0.0	6
2	0.3	6
3	0.4	16
4	18.0	45
5	18.5	90
6	21.5	90
7	22.0	50

^aFlow rate of 325 µl/min used throughout.

Table 11.10.5

Optimizing the Separation of PTH Amino Acids

Problem	Corrective action
Rise in baseline late in chromatogram	Add small aliquots of 1% acetone to buffer A
Negative baseline between DTT and Glu	Add small aliquots of 1.0 M KH_2PO_4 to buffer A
His too close to Ala or eluting after Ala	Increase Premix in buffer A
Arg too close to Tyr or eluting after Tyr	Increase Premix in buffer A
Ile/Lys not well separated (Lys early)	Decrease %B at 18-min point in gradient
Lys/Leu not well separated (Lys late)	Increase %B at 18-min point in gradient

Table 11.10.6

Expected Recoveries of PTH Amino Acids^a

Amino acid	Recovery (%)
	Procise
Arg	50–75
Asn	95 ^b
Gln	90 ^c
His	50
Ile	100
Lys	75–90
Ser	30–50
Thr	50
Trp	10–40

^aThe common amino acids not listed are essentially quantitatively recovered (except cysteine which is completely destroyed during sequencing unless chemically modified to form a stable derivative).

^bAsparagine is partially deamidated to aspartic acid.

^cGlutamine is partially deamidated (~10%) to glutamic acid.

Table 11.10.7
PTH Amino Acid Yields (Uncorrected) from 15-Cycle Sequence Using 10 pmol β -Lactoglobulin^a

	A	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
1	1.81	0.45	0.79	0.45	2.41	0.36	0.80	1.95	[6.86]	0.32	0.48	0.41	1.03	0.48	3.21 ^d	0.93	0.83	0.10	0.65
2	0.92	0.47	0.60	0.34	1.91	0.15	[7.76]	0.86	0.75	0.15	0.45	0.48	0.59	0.62	1.73	0.45	0.46	0.00	0.44
3	0.70	1.20 ^b	0.57	0.28	1.54	0.16	0.76	0.58	0.66	0.13	0.42	0.60	0.51	0.47	0.94	0.37	[5.95]	0.00	0.42
4	0.66	0.78	0.59	0.24	1.52	0.20	0.53	0.53	0.67	0.13	0.28	0.54	0.61	0.66	0.79	[4.00]	0.60	0.13	0.43
5	0.76	0.98	1.15 ^c	0.25	1.49	0.16	0.51	0.52	0.83	0.10	0.38	0.56	[5.24]	0.60	0.67	0.48	0.47	0.00	0.42
6	0.78	0.56	0.71	0.17	1.48	0.11	0.41	0.42	1.03	0.13	0.38	0.60	0.91	0.72	0.65	[3.47]	0.58	0.00	0.50
7	0.74	0.88	0.76	0.30	1.53	0.19	0.44	0.43	1.02	[4.36]	0.38	0.56	0.64	0.68	0.64	0.53	0.55	0.00	0.50
8	0.77	0.63	0.80	0.26	1.68	0.16	0.30	[3.59]	0.98	0.51	0.38	0.49	0.66	0.66	0.72	0.40	0.52	0.09	0.52
9	0.74	0.66	0.76	0.34	[5.05]	0.14	0.60	0.90	1.02	0.23	0.37	0.47	0.69	0.70	0.67	0.35	0.52	0.03	0.55
10	0.84	0.62	0.75	0.36	2.05	0.09	0.53	0.55	[4.99]	0.20	0.35	0.46	0.69	0.80	0.72	0.32	0.60	0.04	0.56
11	0.87	[3.66]	0.75	0.37	1.83	0.09	0.00 ^b	0.51	1.61	0.26	0.35	0.45	0.68	0.69	0.78	0.27	0.61	0.00	0.59
12	0.88	0.99	0.76	0.37	1.83	0.13	[4.85]	0.56	1.29	0.26	0.26	0.42	0.67	0.73	0.77	0.27	0.62	0.00	0.54
13	0.85	0.72	1.16 ^c	0.35	1.87	0.14	1.10	0.49	1.26	0.24	0.37	0.45	[4.07]	0.68	0.73	0.26	0.58	0.05	0.62
14	0.87	0.67	0.83	0.42	1.89	0.14	0.74	[2.97]	1.21	0.23	0.39	0.22	1.15	0.79	0.81	0.30	0.57	0.03	0.58
15	0.90	0.64	0.76	0.41	1.92	0.08	0.67	0.94	1.27	0.23	0.42	0.51	0.80	0.79	0.79	0.32	[3.71]	0.04	0.57

^aThe assigned sequence is indicated by brackets.

^bValue incorrect due to incorrect baseline.

^cGlu value has increased due to partial deamidation of Gln in cycle.

^dSer contamination is common in cycle 1.

Table 11.10.8

Troubleshooting and Optimizing Sequencer Performance^{a,b}

Problem	Possible cause(s)	Suggested solution(s)
No sequence and no increase in background in later cycles	Insufficient sample	Load larger amount of sample and try to quantify amount loaded
No sequence; background does increase in later cycles	Sample is N-terminally blocked	Cleave protein using trypsin and sequence internal peptides and/or evaluate isolation procedure for conditions that might block N- terminus
No sequence; DMPTU, DPTU, and DPU peaks present	R4 bottle empty or reagent not delivered	Replace with full bottle of R4 and/or check delivery
No sequence and no injection solvent front	Extraction solvent bottle empty or solvent not delivered; injector clogged or not working	Replace extraction solvent with full bottle; check delivery of extraction solvent; fix injector
High DMPTU, DPTU, DPU	Poor washing of sample after coupling; poor R4 drydowns	Observe washes and optimize per manufacturer's recommendations; observe dry times and adjust as needed
PTC analogs of PTH amino acids present	Poor R4 drydowns	Observe dry times and adjust as needed
Low PTH amino acid yields in all cycles	Insufficient transfer of ATZs from sample support	Optimize extractions
High background in initial cycles	Sample excessively contaminated	Clean up sample, buffers, buffer water source, etc.
Low initial background increases too rapidly	Acid cleavage time is excessive	Reduce delivery of TFA (R3)
High lag, low RY, low DMPTU, DPTU, DPU	R2 depleted or too old; R2 delivery too short	Replace R2; check R2 delivery and adjust if needed
High lag after Pro	Typical after Pro residue	Can increase R3 reaction time for Pro-rich samples (but will increase background)
Low initial yield	R1 bottle empty or poor delivery; R1 too old; argon gas contaminated	Replace R1 and/or check delivery; replace if >30 days old; replace argon tank with higher- purity gas
Large peak early in chromatogram (air injected onto HPLC column)	Low delivery of S4; injection loop over/under filled; clogged flask pickup tube; S4 clinging to flask walls	Check S4 delivery and adjust if needed; adjust loop fill time; check values and sensors; clean pickup tube or replace; clean or replace flask

^a If dramatic changes in delivery times are observed or if delivery is variable, possible hardware problems should be considered instead of simply adjusting delivery times for solvents and reagents.

^b Abbreviations: ATZ, 2-anilino-5-thiazolinone; DMPTU, dimethylphenylthiourea; DPTU, diphenylthiourea; DPU, diphenylurea; PTC, phenylthiocarbonyl; PTH, phenylthiohydantoin; RY, repetitive yield; TFA, trifluoroacetic acid.

Table 11.10.9

Troubleshooting and Optimizing HPLC Performance^{a,b}

Problem	Possible cause(s)	Suggested solution(s)
No peaks and no changes in baseline	HPLC pumps not running; detector lamp not on	Evaluate and correct HPLC problem(s); turn on lamp or replace if needed
Retention times for all PTH amino acids are variable	HPLC pump seals worn; A or B pump delivery erratic	Replace pump seals; identify problem pump and repair or replace
Retention times for early PTH amino acids are variable	Insufficient reequilibration prior to injection	Increase reequilibration time
PTH amino acid peaks become broader with time	Contaminants on guard column; PTH column aging	Replace guard column at least every 30 days; replace PTH column
Noisy baseline; erratic small sharp spikes in baseline	HPLC lamp aging	Replace lamp

^a Also see Table 11.10.5 for baseline adjustments and optimization of PTH amino acid separation.

^b Abbreviations: HPLC, high-performance liquid chromatography, PTH, phenylthiohydantoin.