**Dartmouth Molecular Biology and Proteomics Core Facility**

Dartmouth Medical School
HB 7650
Hanover, NH 03755

Main Office/DNA Services Remsen 243
Protein Services Vail 811A
Phone 603-650-6546, 1788
FAX 603-650-1129

mbcore@dartmouth.edu
www.dartmouth.edu/~mbcf

---

## Protein Sequencing Sample Preparation and General information



The major difficulty in protein sequencing is sample preparation. The major difficulty with sample preparation is that every protein is different and behaves differently when removed from its "environment," whether this is an in vivo environment or an in vitro environment in which you are used to handling it. If you have a sequencing project, please consult with us from the start of your project if possible. We may be able help you plan your purification strategy to maximize your chance of having enough protein to sequence. At the very least, we can make suggestions to improve your chances.

### Sample Amount

Most investigators greatly over-estimate the amount of protein they have. This is understandable in light of the difficulty and the amount of labor involved in protein purification. The MB&P Core has an Applied Biosystems (ABI) Model 494 sequencer which will sequence proteins in the low picomolar range. The model 494 sequencer uses standard Edman degradation, which has been modified to accommodate automated sequencing. The 494 will perform both gas phase and pulsed liquid sequencing. It is ideal for a situation where a mixture of sample types, such as gel blots or samples in solution, is encountered.

The manufacturer of our instrument, Applied BioSystems, guarantees that the instrument will sequence 10 picomoles of standard protein with >93% repetitive yield. That means that with 10 pmol of protein, for example, about 0.5 micrograms of a protein with a molecular weight of 50,000, the instrument is able to sequence twenty amino acids and still remain well within the detection limits of the PTH analyzer (see Chemistry below). The protein standard that is used to demonstrate this is beta-lactoglobulin, which is a readily purified protein and easily sequenced.

**Sample Quantitation**

The two most common causes of unsuccessful sequencing are insufficient sample and a blocked N-terminus. The best way to distinguish between the two problems is sample quantitation. If there is sufficient protein present and no sequence is obtained, the N-terminal is probably blocked. Therefore, quantitation is an absolute requirement for sample submission. Blotting to PVDF membrane is the method of choice for most unknown proteins. For blotted samples an easy way to estimate the number of pmoles in your protein sample is to run it with known quantities of at least two concentrations of unstained standards in adjacent lanes of the same gel/blot. Try to choose known proteins whose molecular weights closely approximate your protein(s) of interest. We recommend BioRad unstained molecular weight markers; they are available in several molecular weight ranges to suit the needs of your experiment. Bovine serum albumin is not recommended as a standard since it reacts atypically with many stains. Densitometry should be used to compare the staining intensity of your sample to the intensity of the known quantities of protein standards. There is a laser densitometer and appropriate software in the core facility available 24 hours per day, seven days per week. We like Invitrogen's colloidal Coomassie Blue kit; the sensitivity of staining is about 50ng and it is easy to use.

**Blocked N-terminal**

Many proteins acquire blocked N-termini during purification. A procedure called internal sequencing is usually used on these samples unless the N-terminal blocking group can be easily removed. Internal sequencing requires that the N-terminally blocked protein be subjected to proteolytic digestion. The fragments are then separated by micro-bore or capillary HPLC, then collected and sequenced individually. Fractions containing more than one peptide are difficult to interpret. Even if the sequences of all the species are known, sorting out which amino acids belong to which peptide is sometimes not possible due to "preview" and "lag" phenomena which occur as a result of the sequencing process. Sometimes two proteins that occur in two vastly different quantities - e.g., 100 pmol and 10 pmol - can be co-sequenced and provide usable data.

## Sample Preparation

Samples are loaded onto the sequencer in two ways: in solution or blotted to PVDF membranes. For proteins or peptides in solution, the sample is dried onto a specially treated glass fiber filter. This method is suitable for most proteins and peptides. Usually, samples presented in solution have been isolated by traditional means such as ion exchange and/or size exclusion chromatography, RP-HPLC, etc. Proteins can also be loaded on the sequencer blotted to a PVDF membrane. These can be electro-blotted samples from PAGE gels or samples adsorbed to PVDF (ABI's ProSorb cartridge) membrane. This technique is particularly useful for samples containing buffers that may interfere with the sequencing chemistry. We recommend ABI's specially treated ProBlot PVDF Membrane for electro-blotting samples. You can download the procedure HERE. Other brands of specially treated PVDF membrane work as well. Transferring the sample to PVDF membrane has become the preferred method for sample presentation. This method offers a slightly easier sample preparation and a lower background on the sequencer. The only problem with PVDF blots is that small peptides of <30 amino acids tend to wash off the membrane during the sequencing operation. Samples should be cut out and trimmed as closely as possible. Multiple bands of the same protein can be submitted if conditions prevent having sufficient protein in a single band. Blotted samples should be submitted dry in 1.5mL eppendorf tubes. Please consult the MB&P Staff for help quantitating proteins prepared by chromatographic methods.

## Chemistry

The chemistry for automated protein sequencing is a slight variation of the method developed by Pehr Edman in the 1950's. It consists of three main steps: coupling, cleavage, and conversion. The free N-terminus of the first amino acid in the protein is coupled to phenylisothiocyanide. This complex is subsequently cleaved off the protein with a strong acid, trifluoroacetic acid (TFA), leaving a free N-terminus on the protein. The resulting complex, the anilinothiazolinone (ATZ) amino acid, is extracted and removed to a separate vessel and converted in aqueous TFA to the more stable phenylthiohydantoin (PTH) derivative. The chemistry is then cycled again and additional PTH amino acids are made in the order in which they appear in the protein. The PTH amino acids are then analyzed by reverse phase HPLC. In automated sequencing, one cycle takes approximately thirty minutes. The limit of this chemistry is the amount of PTH amino acid you can "see" with the UV detector. This means that every step of the chemistry must be optimized to give a maximum yield of all PTH amino acids. This, unfortunately, is a major source of variability in protein sequencing. Some amino acids do not exhibit average behavior when subjected to the Edman chemistry. Since it can't be predicted a priori when they will occur in a sequence, changes in the chemistry can't be made to accommodate them.

Typically, the amino acids Cys, Ser, Pro, and Gly are problematic for different reasons. Detection of Cys requires an additional derivitazation step. Other amino acids may present problems to a lesser extent. These problems are usually related to the protein's primary structure.

**Laboratory procedures**

The optimization of the Model 494 will be checked every twenty sequencing runs. Drying times and reagent and solvent deliveries will be checked before each run. The PTH analyzer will be pre-cycled to ensure adequate separation of all PTH-amino acids at the lowest possible concentration of standards. Biweekly, or once in every twenty runs, a control protein of known sequence, beta-lactoglobulin, will be run after the optimization procedure. The standard will be run at a sufficiently low level to ensure sequencer performance at that level. These test runs will be compared to previous runs to determine if the instrument is performing correctly. If it is not performing to our standards, we will troubleshoot and correct the problem.

**Other Terms**

Samples that do not meet the criteria for sequencable samples will be put on the instrument at the normal charge of $250. If there is no discernable sequence or there are unexpected multiple sequences, the instrument will be stopped at the discretion of the operator. If a protein is sequencing particularly well and in the opinion of the operator it will sequence beyond the requested number of amino acids, a reasonable attempt will be made to contact the investigator for permission to continue sequencing. If the investigator cannot be contacted, the sample filter or blot will be preserved with the chance that additional sequence can be obtained if the sample is reloaded on the instrument.