

**Dartmouth MB Core Protein Services:  
Protein Identification, Mass Spectroscopy**

**Dartmouth Molecular Biology Core Facility  
Main Office/DNA Services Rem. 243  
Dartmouth Medical School  
HB 7650  
Hanover, NH 03755**

**Protein Services Vail 811A  
mbcore@dartmouth.edu  
[www.dartmouth.edu/~mbcf](http://www.dartmouth.edu/~mbcf)  
Phone 603-650-6546 Protein 1788  
FAX 603-650-1129**

## **Introduction**

The Dartmouth Molecular Biology Core Facility has a Thermoquest LCQ Deca XP electrospray ionization-ion trap mass spectrometer (ESI-IT) and the ABI Voyager matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometer. These instruments give the Dartmouth community up to date mass analysis capabilities, providing not only standard biomolecular mass analysis, but protein identification through sophisticated database searching capabilities. Although many interesting and sophisticated analyses can be performed with these instruments, we anticipate that most investigators' primary interest in these tools will be the identification of proteins from 1D and 2D gel bands. Accordingly, this document and the sample preparation instructions are directed at this method of protein identification. If you are interested in or would like more information about specific assays not covered in this document please contact the Molecular Biology Core facility. Please remember that we have limited personnel in our proteomics section at the present time. Users must perform a large part of the sample preparation in their own laboratories or, if they lack access to specific pieces of equipment, in the core facility with our guidance.

## **Background**

Mass Spectrometry is a method for determining the mass of molecules by producing and analyzing charged species (ions) by measuring the mass to charge ration ( $m/z$ ) of ions. It is important to remember that mass is not directly measured it is derived from the  $m/z$  value. In the late 1980s and early 1990s improvements in ionization techniques facilitated the ionization of large biomolecules and the accurate mass determination of these species in the range of 0.01%-0.05%. Earlier techniques such as sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) produce a mass accuracy in the range  $\pm 5\%$  and in some cases errors are much greater. Mass determinations are used in both the characterization and identification of biomolecules. Our mass spectrometers each have a different mechanism for creating ions and a different method for analyzing the ions as well as different levels of sensitivity. While the individual instruments have their own strengths and weaknesses, together they complement each other very well.

### **ABI Voyager DE Pro (MALDI-TOF)**

The ABI Voyager DE Pro utilizes a laser whose output is matched to the absorption maxima of a matrix compound. The energy absorbed by the matrix results in a plume of volatile ions. If biomolecules are co-crystallized with the matrix they are also ionized in the plume. Ions from the plume are propelled down a long flight tube under the influence of an electric field. Since the time an ion takes to reach the detector at the end of the flight tube is inversely proportional to the mass to charge ratio ( $m/z$ ), accurate measurements over a very wide range ( $>300,000$   $m/z$ ) can be obtained. The advantages of MALDI-TOF include good sensitivity (into the picomole range), and good tolerance to biochemical buffers and salts. In addition, since there is usually only a singly charged ion formed it is a good choice for the analysis of heterogeneous samples.

Access to our MALDI-TOF instrument is provided to qualified investigators who, after a short training session can use this instrument on a sign up basis. Training and advice is provided by the MB&P staff at no cost. Individual users are responsible for methods development, any calibration procedure needed for their research, and must purchase their own sample plate. Standard core facility procedures are not guaranteed to produce publishable data except in the hands of core facility staff.

Other uses of this instrument include obtaining molecular weights of purified molecules such as proteins, peptides, synthetic polymers and oligonucleotides.

### **ThermoFinnigan LCQ Deca XP**

ThermoFinnigan LCQ Deca XP is an electrospray ionization-ion trap mass spectrometer (ESI-IT). As its name implies, ions formed during electrospray ionization are generated by a very small nozzle which produces micro-droplets. Under the influence of a potential and pressure gradient, these droplets are desolvated resulting in the formation of multiply-charged ions that are focused and eventually captured in an ion trap. The captured ions are then ejected from the trap and detected according to their  $m/z$  ratio. Because the ions are multiply-charged, only a limited  $m/z$  range is usually scanned, greatly increasing the mass accuracy. Advantages of ESI-IT are gentle ionization, the ability to adjust conditions to evaluate “native molecules” and in nanospray configuration it is more sensitive than MALDI-TOF. The greatest advantage of ESI-IT MS is the ability to easily select an ion, fragment it and reanalyze it. This is known as MS/MS and gives important structural/sequence information that greatly aids identification of peptides.

The Core Facilities' Thermo Finnigan LCQ Deca XP is equipped with an Advion Nanomate nanospray device. The NanoMate is a “chip”-based automated nanospray source. With this device we can spray for extended periods of time allowing for a “gas phase” separation of components. This method can be used for relatively simple mixtures of digested proteins. LC/MS is used for more complex mixtures, ESI-IT is “coupled” to capillary HPLC (High performance Liquid Chromatography). Two LC techniques used for more complex digests are reverse phase HPLC and MudPIT (Multidimensional Protein Identification Technology), a two dimensional LC technique. The disadvantages of ESI-IT include: interference by common salts and buffers, complex spectra, and inability to handle large molecules ( $m/z >2000\text{Da}$ , enzymatically digested proteins result in peptides in its effective range (200-2000Da)) at the present time.

## **Protein identification**

### **Current Processes**

We anticipate that most investigators' primary interest in MS will be the identification of proteins from 1D and 2D gels. Accordingly, this document and the sample preparation instructions are directed at this method of protein identification. However, the core process, digestion, fractionation and MS analysis has multiple entry points and will accommodate a variety of sample types. Fractions from the Beckman-Coulter PF2D Protein fractionation system are handled in lab according to established protocols and require participation by users only in the selection of fractions for further analysis. If you are interested in or would like more information about specific assays not covered in this document please contact the Molecular Biology and Proteomics Core facility. Please remember that we have limited personnel in our proteomics section at the present time. Users must perform a large part of the sample preparation in their own laboratories or, if they lack access to specific pieces of equipment, in the core facility with our guidance.

## **In Gel Digestion**

Trypsin digestion of a protein and subsequent mass analysis of the tryptic peptides is the generally accepted method for identifying unknown proteins. Two methods are used to identify the protein from its constituent peptides. A simple mass map, a list of masses acquired usually using MALDI-TOF on the whole digest, can be used to search a database. Identification is based on matching masses to a protein database that has been digested "in silico". The second method is the separation of the peptides and subsequent MS/MS analysis. In addition to a mass map, MS/MS provides information about the primary structure of the peptide. The primary structure information supplements the mass map by providing more definitive data to the database search engine

When using mass analysis to identify proteins, a crucial step in the process is the chemical or proteolytic digestion of the sample to produce peptides with predictable characteristics. In-gel tryptic digestion is a technique that was developed for the determination of primary amino acid sequence of N-terminally blocked proteins, but is commonly used for protein identification and proteomics. Digestion is usually preceded by reduction of the protein's disulfide bonds followed by alkylation of the sulfhydryl groups to prevent further reaction. Most commonly trypsin is then used to digest the protein while it is still in the gel slice. This technique typically results in a number of peptides with free amino termini and lysine or arginine at the C-terminal. Peptides from a tryptic digest can be directly applied to a MALDI sample plate and/or can be subjected to gas phase fractionation, RP-HPLC or MudPIT analysis with the column effluent sprayed directly into the ESI-IT instrument. The in-gel digest is the most crucial step in the process of identifying in-gel proteins, therefore this multi-step procedure needs to be performed in a prescribed manner with positive control provided by the Core Facility and a negative control provided by the investigator. It is also important for us to monitor the results from our positive control from sample batch to sample batch to insure consistency over time.

### **MALDI-TOF (ABI Voyager- DE-Pro)**

The advantages of MALDI-TOF MS for peptide analysis and protein identification are speed, ease of analysis, and accuracy of the molecular weights obtained by the TOF detector. The tryptic digest is spotted on the sample holder along with matrix (usually alpha-cyano-4-hydroxycinnamic acid), dried, then placed in the instrument. A laser provides the energy to volatilize and ionize the peptides. A mixture of known peptides is placed on the sample holder near the sample for calibration purposes. Data from this known mixture are used to derive a calibration curve and the curve is then applied to the sample mixture. If there is sufficient high quality data the corrected peptide masses obtained from the sample are used to search databases. The assumption is that an individual protein will contain a distribution of peptide masses that will prove unique to a single protein in a database of "in silico" digested proteins. The NCBI non-redundant or the GenPep are examples of two such databases that are commonly used. A species specific database will be searched if one is available. Provided there are enough data (strong signals from a sufficient number of peptides) and the protein is in a database, this technique will identify a protein over 90% of the time.

### **ESI-IT (ThermoFinnigan LCQ Deca XP)**

ESI-IT is the mass spectrometer most used for protein identification from very small amounts of sample. This type of analysis is more complex, more difficult and therefore more expensive than MALDI-TOF analysis. The electrospray interface allows samples to be directly introduced into the mass spectrometer. This has the advantage of allowing sample application in a variety of buffers and solvents.

Equipped with the NanoMate source, the ThermoFinnigan LCQ Deca XP can perform a “gas phase” separation of a protein digest. The spray is initiated, the mass spectrometer is programmed to collect data across narrow  $m/z$  ranges. In each  $m/z$  range the mass spectrometer is programmed to exclude ions that have already been analyzed starting with the most abundant ions to a set limit. The advantage of this analysis is that it's rapid and relatively easy to do.

An additional advantage of this instrument is the ability to interface it with HPLC. A peptide mixture is partitioned on a column in one or two dimensions then directed into the ESI-IT instrument, allowing peptides to be analyzed as they are eluted from the LC column. The peptides can then be trapped and further degraded (MS/MS). Peptides degrade in a predictable manner;  $m/z$  values can be found in the MS/MS spectra corresponding to the peptide and the peptide missing one, two, three etc. amino acids from both C and N termini. MS/MS provides important compositional data that is not obtained with MALDI-TOF. The accompanying Sequest software performs database searches by querying not only the molecular weight of peptides, as with MALDI-TOF searches, but also compositional data obtained by MS/MS. This analysis can be performed on extremely small amounts of material; low femtomole amounts of peptide, at least two orders of magnitude lower than the MALDI-TOF, can be detected. This is about the same level of detection as the most sensitive silver stained gel.